Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Semantic Technologies in Idea Management Systems: A Model for Interoperability, Linking and Filtering

Tesis Doctoral

Adam Westerski
Ingeniero de Informática

2012

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros de Telecomunicación

Semantic Technologies in Idea Management Systems: A Model for Interoperability, Linking and Filtering

Tesis Doctoral

Adam Westerski
Ingeniero de Informática

2012

Departamento de Ingeniería de Sistemas Telemáticos

Escuela Técnica Superior de Ingenieros de Telecomunicación

Universidad Politécnica de Madrid

Semantic Technologies in Idea Management Systems: A Model for Interoperability, Linking and Filtering

Autor:

Adam Westerski

Ingeniero de Informática

Tutor:

Carlos Ángel Iglesias Fernández

Doctor Ingeniero de Telecomunicación

2012

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día ____ de ________ de ______.

Presidente: ____

Vocal: ____

Vocal: ____

Vocal: ____

Secretario: ____

Suplente: ____

Suplente: ____

Realizado el acto de defensa y lectura de la Tesis el día ____ de ________ de ______ en la E.T.S.I.T. habiendo obtenido la calificación de ____.

EL PRESIDENTE LOS VOCALES

EL SECRETARIO

Acknowledgements

I would like to express my gratitude to all those who supported me during the road leading to the creation of this thesis.

First and foremost, I would like to thank my supervisor, Carlos A. Iglesias for inviting me to collaboration, teaching about doing research and giving me the freedom to pursue and implement new ideas my own way.

Secondly, Mercedes Garijo for admitting me the GSI group and allowing to realise my dream of living and doing research in Spain.

Aside of Mercedes, I would like to thank the remaining members of PhD Committee that agreed to review my thesis and participate in the thesis defence, namely: Juan Botía, Gregorio Convertino, Gregorio Fernández and Felix Larrinaga. Additionally, I would like to express my thanks to external reviewers who helped to assess the value of the thesis: Jacek Jankowski and Giovanni Tummarello.

Furthermore, I would like to thank a great deal of people who collaborated with me on research that in some form stimulated or contributed to this thesis. Among those, I would like to express special thanks to: Theodore Dalamagas, George Anadiotis, Mylene Leitzelman, Felix Larrinaga and Osane Lizarralde.

Finally, I would like to thank all students of the GSI group who contributed to the Gi2MO project, especially: Fernando Tapia Rico and Javier Espinosa Garcia. Their effort put into development of prototypes and participation in experiments for the thesis was invaluable.

Additionally, I would like to thank all my friends from the GSI group for the years spent together on sharing ideas, problems and working together on solving them: José, Laura, Jorge, Paloma, Paco, Miguel, Álvaro, Jota, Geovanny, Vicente, Jota, Emilio and Ganggao.

Last but not the least, I would like to thank the people whom I collaborated with during my academic career prior to arriving in UPM, namely, Sebastian Kruk and Stefan Decker who invited me to DERI and introduced to academic research in the first place; secondly, Giovanni Tummarello who invited me to work on his intense projects crossing over from academia to industry.

Abstract

Idea Management Systems are web applications that implement the notion of open innovation though crowdsourcing. Typically, organizations use those kind of systems to connect to large communities in order to gather ideas for improvement of products or services. Originating from simple suggestion boxes, Idea Management Systems advanced beyond collecting ideas and aspire to be a knowledge management solution capable to select best ideas via collaborative as well as expert assessment methods.

In practice, however, the contemporary systems still face a number of problems usually related to information overflow and recognizing questionable quality of submissions with reasonable time and effort allocation. This thesis focuses on idea assessment problem area and contributes a number of solutions that allow to filter, compare and evaluate ideas submitted into an Idea Management System.

With respect to Idea Management System interoperability the thesis proposes theoretical model of Idea Life Cycle and formalizes it as the Gi2MO ontology which enables to go beyond the boundaries of a single system to compare and assess innovation in an organization wide or market wide context.

Furthermore, based on the ontology, the thesis builds a number of solutions for improving idea assessment via: community opinion analysis (MARL), annotation of idea characteristics (Gi2MO Types) and study of idea relationships (Gi2MO Links).

The main achievements of the thesis are: application of theoretical innovation models for practice of Idea Management to successfully recognize the differentiation between communities, opinion metrics and their recognition as a new tool for idea assessment, discovery of new relationship types between ideas and their impact on idea clustering.

Finally, the thesis outcome is establishment of Gi2MO Project that serves as an incubator for Idea Management solutions and mature open-source software alternatives for the widely available commercial suites. From the academic point of view the project delivers resources to undertake experiments in the Idea Management Systems area and managed to become a forum that gathered a number of academic and industrial partners.

Resumen

Los Sistemas de Gestión de Ideas son aplicaciones Web que implementan el concepto de innovación abierta con técnicas de crowdsourcing. Típicamente, las organizaciones utilizan ese tipo de sistemas para conectar con comunidades grandes y así recoger ideas sobre cómo mejorar productos o servicios. Los Sistemas de Gestión de Ideas han avanzado más allá de recoger simplemente ideas de buzones de sugerencias y ahora aspiran a ser una solución de gestión de conocimiento capaz de seleccionar las mejores ideas por medio de técnicas colaborativas, así como métodos de evaluación llevados a cabo por expertos.

Sin embargo, en la práctica, los sistemas actuales todavía se enfrentan a una serie de problemas, que, por lo general, están relacionados con la sobrecarga de información y el reconocimiento de las ideas de dudosa calidad con la asignación de un tiempo y un esfuerzo razonables. Esta tesis se centra en el área de la evaluación de ideas y aporta una serie de soluciones que permiten filtrar, comparar y evaluar las ideas publicadas en un Sistema de Gestión de Ideas.

Con respecto a la interoperabilidad de los Sistemas de Gestión de Ideas, la tesis propone un modelo teórico del Ciclo de Vida de la Idea y lo formaliza con la ontología Gi2MO que permite enlazar sistemas distribuidos y heterogéneos, lo que facilita comparar y evaluar su innovación asociada, en un contexto más amplio dentro de cualquier organización o mercado.

Por otra parte, basado en la ontología Gi2MO anteriormente presentada, la tesis desarrolla una serie de soluciones para mejorar la evaluación de las ideas a través de: análisis de las opiniones de la comunidad (MARL), la anotación de las características de las ideas (Gi2MO Types) y el estudio de las relaciones entre ideas (Gi2MO Links).

Las aportaciones principales de la tesis son: la aplicación de los modelos teóricos de innovación para la explotación en Sistemas de Gestión de Ideas para identificar las diferencias entre comunidades, la definición de métricas de opiniones de comunidad y su validación como una nueva herramienta para la evaluación de ideas, el descubrimiento de nuevos tipos de relaciones entre ideas y su impacto en la agrupación de éstas.

Por último, el resultado de está tesis es el desarrollo del proyecto Gi2MO que sirve como incubadora de soluciones para Gestión de Ideas y ofrece una alternativa de código abierto madura a las soluciones comerciales. Desde el punto de vista académico, el proyecto ha proporcionado una base común para el desarrollo de experimentos en el área de Sistemas de Gestión de Ideas y ha logrado convertirse en un foro que reunión para un número considerable de socios tanto académicos como industriales.

Contents

Chapter 1

Introduction

Idea Management Systems are software platforms for collecting ideas for innovation from large communities. With the constantly growing popularity of Internet as a communication medium, the Idea Management Systems have become an important element in the innovation management practices of organizations. This thesis looks into the problems that those systems face in terms of management of knowledge and its assessment. In particular, we investigate data interoperability and characteristics of ideas in order to contribute new techniques for comparison of ideas and idea datasets.

In this chapter, the thesis motivation, objectives, research questions and solution architecture are outlined for the reader. We summarize the problems of the contemporary Idea Management Systems and by relating to them we describe our motivations to pursue the later defined thesis objectives. In relation to the objectives, we formulate the specific research questions that this thesis set out to answer. The goal of the research questions is to focus down the research on specific problems that are later evaluated using clear research methodologies. Finally, we summarize the solution architecture that is proposed by the thesis to obtain the answers for the stated research questions.

1.1 Contribution area and scope

The contribution area of this thesis is Idea Management Systems: computer software developed to collect ideas, enable their improvement and provide various idea assessment capabilities.

The notion central to design of Idea Management Systems (IMS) is the participatory role of communities gathered around an organization. Those kind of systems are deployed for collecting feedback from communities gathered outside of the organization (e.g. clients) or internal staff. In particular, the goal is to enable creation of ideas by any member of the community and sharing those ideas in a single space where other community members can see them.

Typically, Idea Management Systems are aimed for interacting with large communities that can submit thousands of ideas within the time frame of months. Therefore, some additional means of information organization are required. This is addressed by the contemporary solutions in two ways: collaborative assessment of ideas by communities; and internal reviews and moderation activities performed by designated staff of the organization.

In the first case, aside of acting as inventors the members of community are also given tools to review each others ideas, e.g. via comments or various ranking mechanisms. Through those collaborative assessment techniques, the organizations are able to discover what kind of innovations are desired by majority of a particular community. Furthermore, recognition of ideas in a public forum serves as an incentive for the community members to improve their ideas and keep on submitting new proposals for innovation.

In the second case, aside of providing support for those collaborative activities, the Idea Management Systems deliver back-end facilities locked away from the public and available for internal staff of the organization in order to moderate, organize and review the gathered information. The goal of those facilities is to analyse ideas under various criteria (e.g. financial, implementation effort etc.), filter out the ideas irrelevant to the organization and help in finding the best picks for ideas which implementation would deliver most benefits. The notion behind this part of Idea Management Systems is delivering information based on expertise of organization staff knowledgeable with organization strategy, objectives, financial capabilities, technical know-how etc.

The analysis and improvement of methods connected to processing idea information in the IMS back-end are within the contribution scope of the thesis. In particular, the progress beyond the state of the art of Idea Management Systems research is achieved by proposing new idea assessment methodologies and data modelling approaches to information already gathered by state of the art IMSes as well as new information that is proposed by the thesis as an addition to the contemporary model. Furthermore, those contributions of the thesis are delivered by introducing the use of various technologies and modelling techniques detailed throughout the following chapters.

1.2 Problem and Motivation

In the era of globalization the markets become more competitive and the organisations seek new ways of innovating. Among those attempts, are Idea Management Systems that employ Information Technology and crowd-sourcing principals to support innovation processes in the organizations. In particular, the notion behind those systems originates from simple suggestion boxes but is transformed into a more sophisticated process (Turrell, 2002). During the last decade of their evolution Idea Management Systems have extended their coverage from collecting ideas from large communities via computer networks to collaborative improvement of those ideas, the assessment of ideas and idea management in synergy with other enterprise processes (Westerski et al., 2011a).

Currently, Idea Management Systems are considered a very promising branch of computer software market (Fenn and LeHong, 2011) and various analyses of the vendor landscape (Rozwell et al., 2010; Brown et al., 2009) show rapid adoption growth in many enterprises in recent years. Nevertheless, current state of the art Idea Management Systems still face key problems related to the large amount of human effort needed during the idea management process.

Based on the testimonials of Idea Management Systems vendors (Baumgartne, 2008) and case studies of various companies (Jouret, 2009; Belecheanu, 2009) a common problem is **idea assessment**. That is: how to judge which ideas are useful for the organization and which are not. Idea assessment rises to a magnitude of a serious challenge due to a number of factors typical for Idea Management System deployments:

- **(1) data overflow**. Idea Management Systems collect ideas from large, distributed communities. Therefore, the amount of input can be very big ranging up to thousands or tens of thousands of ideas. For idea assessment this delivers a problem of how to review and judge all this information in an efficient way and delivering sufficient results so that later managers can make accurate decisions with regard to which ideas should be implemented.

- **(2) noisy data**. Ideally, Idea Contest organizers expect to get diversified and original ideas that otherwise would not be obtained with closed-innovation practices. Nevertheless, in practice of Idea Management Systems innovation proposals born as an effect of community collaboration are small, incremental ideas that often intersect or even duplicate each other. Furthermore, many of the submissions are often trivial innovations that do not offer a new point of view for the organization. This creates a challenge for idea assessment to detect idea intersections and remove all of the unwanted or redundant ideas.

- **(3) large peaks of data over short period of time**. Submission of ideas is often stimulated by organized events for collection of ideas (Idea Contests) or by other events related to activities in the organizations environment. All of those events have certain time boundaries and result in increased activity of innovators during that time.

This creates a problem for idea assessment because ideas need to be moderated and given feedback in a timely manner otherwise innovators as well as the community that supports them loose interest in the platform.

- **(4) difficulties in rating innovation**. Even if all the aforementioned problems would be solved, still judgement of innovation and impact of ideas remains a difficult task on its own. The currently utilized metrics for assessment of idea value are: (1) automatically generated statistics of community activity in the Idea Management System (e.g. amount of comments per idea); (2) expert reviews run in reference to the organization demands (e.g. return of investment, time to market etc.). With regard to idea assessment those metrics are either too generic to give an answer about innovation usefulness (community statistics) or take too much time to generate, require a lot of expertise and consume too much human resources (expert reviews). Studies have shown that although managers use both of the solutions, their final decision has a very weak correlation with the choices suggested by any of the aforementioned indicators (Gangi and Wasko, 2009).

Our motivation when preparing the following thesis was to improve the current situation with relation to aforementioned **idea assessment problems**. In particular, we are motivated by the fact that there is still a lack of solution for indicators and methods for data summarization that allow moderators to compare new ideas or idea contests with older innovations and make decisions about their usefulness for the organization.

1.3 Objectives

The primary objective of the thesis is to deliver a solution that will aid Idea Management System managers in judgement whether particular ideas are good or bad for their organization. By referring to the previous section, it can be observed that issues of idea assessment originate from a number of causes. Therefore, we have decomposed the thesis global objective into a number of more specific ones in order to build the final solution step by step:

- **(1) propose a conceptual model for Idea Management System**. Our objective is to identify the processes that create and transform Idea Management System data to deliver a solution that would leverage information contained in any Idea Management System into a common level. The model should classify and categorize the data to give a clear understanding of the domain so that further works can be conducted to address the problems of Idea Management Systems regardless of particular vendor or case study specifics. The formalization of the model should put impact on data interoperability and portability to enable comparison of ideas, idea datasets or entire Idea Management System across domains or multiple deployments in the same domain (e.g. multi-lingual instances).

- **(2) summarize data of Idea Management Systems**. Our objective is to research on a solution that would allow to describe idea datasets and the communities involved in their creation. This goal involves to discover the typical characteristics of ideas and relationships between them. Furthermore, formalize this information and experiment with application of those new annotations by different kinds of idea reviewers as well as automatic annotation solutions. Additionally, we investigate the significance and correctness of the proposed idea summarization in relation to the full dataset information as well as previous practical and theoretical achievements in the area.

- **(3) deliver metrics for idea assessment**. Our objective is to research on more descriptive metrics than the currently utilized community statistics but also metrics less demanding in terms of measurement effort than expert reviews. In particular, the goal is to propose a method that would quantify aforementioned characteristics for idea summarization so that ideas can be compared to each other.

1.4 Research questions

Apart of identifying the main problems and stating the objectives for the thesis, we have formulated a number of research questions that guided us during the research described in this thesis. The goal of those research questions is to narrow down the problem area within the stated objectives and clearly present the issues that we attempt to solve so that later it would be possible to verify the validity of contributions of the thesis. The questions that we defined for this thesis are:

- **(1) can the contemporary Idea Management Systems be generalized into a single model?** With this research question, we put forth a hypothesis that Idea Management Systems could be captured in a single model that describes all the processes conducted during interaction with the system and the evolution of data that is the output of those interactions. With regard to earlier presented objectives and problems, this research question aims to lead us to a solution that would create a common level of understanding for data concepts related to community innovation in an organization, so that later ideas could be compared and judged regardless of the system or deployment they originate from.

- **(2) can community activity related to ideas be modelled, summarized and measured regardless of the system or domain?** With this research question, we hypothesise that community actions manifested with idea comments can be modelled and described through their relation to the innovation that is being commented. Furthermore, we put forth a hypothesis that the nature of relation between comments and ideas can be measured and used for idea assessment.

- **(3) can ideas be modelled, summarized, measured and compared independently of the domain or used IMS system?** With this research question, we hypothesise that apart of idea information that is individual per use case of the Idea Management System, there are a number of indicators that are independent of the domain and could be used as a tool for idea comparison and to determine if the characteristics of the idea are in-line with the expectations of the IMS managers.

- **(4) can the content of Idea Management Systems be summarized on the basis of relationships between ideas?** With the last research question, we hypothesise that the currently utilised duplicate relationship in the Idea Management Systems does not cover fully the semantics of relationships between ideas. Therefore, we put forth a question if it would make a significant change to introduce more relationship types, would those be applicable in the existing use cases and to what degree would new relationship types improve the capabilities to summarize idea datasets.

Each of the above four research questions are addressed by the thesis. In relation to each other, with the first research question (1) we attempt to create a base for further analysis and comparison of Idea Management Systems. The next three questions (2, 3, 4) leverage those foundations and look into various aspects of data summarization and delivering indicators for Idea Management Systems which together would enable comparison of ideas. To address all of those questions we have proposed a number of solutions connected together in a single architecture that is outlined in the next section.

1.5 Solution architecture outline

In order to fulfil the stated objectives and answer the research questions we propose a number of solutions put together in a single framework that facilitates idea comparison and aims to aid Idea Management System moderators and Idea Competition managers to assess the idea usefulness in relation to their organization (see Fig. 1.1).

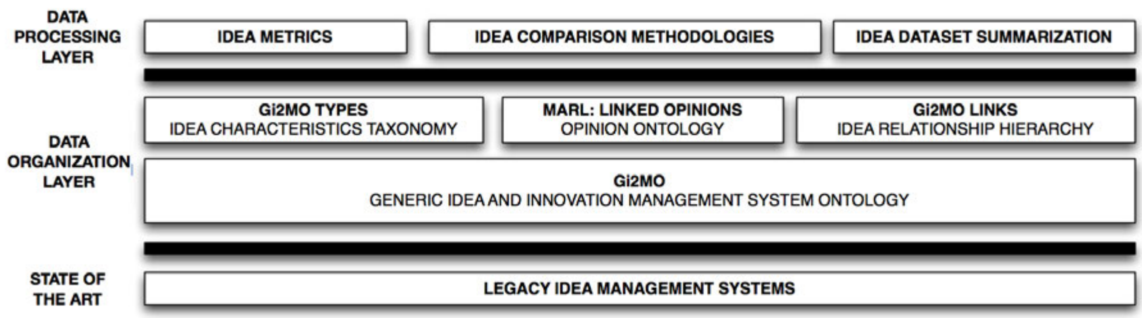


Figure 1.1: Components of the solution architecture

On top of the contemporary state of the art Idea Management Systems we propose a number of enhancements to improve data organization (Data Organization Layer). Based on the proposed extensions, we introduce a number of methodologies for calculating indicators for idea or idea datasets and evaluate their use for comparison of data (Data Processing Layer). The proposed Data Organization Layer consists of four main elements (see Fig. 1.2):

- **(1) Generic model for Idea Management Systems**. We propose a notion of Idea Life Cycle and detail the data evolution within. As a formalization of the life cycle, we propose Gi2MO ontology for describing information inside the contemporary Idea Management Systems. This contribution is described in Chapter 4.

- **(2) Community Opinion Model**. We to extend the base model for Idea Management System with an ontology for describing user opinions (Marl). Additionally, we propose a method for summarization of such opinions and a metric that expresses the sentiment of users discussing an idea. The research on community opinion modelling is described in Chapter 5.

- **(3) Idea Characteristics Model**. We extend the generic IMS model with idea characteristics independent of the domain where system is applied. We deliver a model for characteristics in a form of Gi2MO Types taxonomy and present a solution for transforming idea annotations made with this taxonomy into metrics that are capable of summarizing idea datasets. The research on idea characteristics is detailed in Chapter 6.

- **(4) Idea Relationships Model**. We propose to extend the current relationship model utilized in the state of the art Idea Management Systems. In particular, we deliver a hierarchy of relationships present in an Idea Management System (Gi2MO Links), propose the dependencies between those relationships and calculate an indicator that expresses the summarization ratio that can be achieved by aggregation of ideas based on their relationships and various logic rules. The investigative work done in modelling of idea relationships is described in Chapter 7.

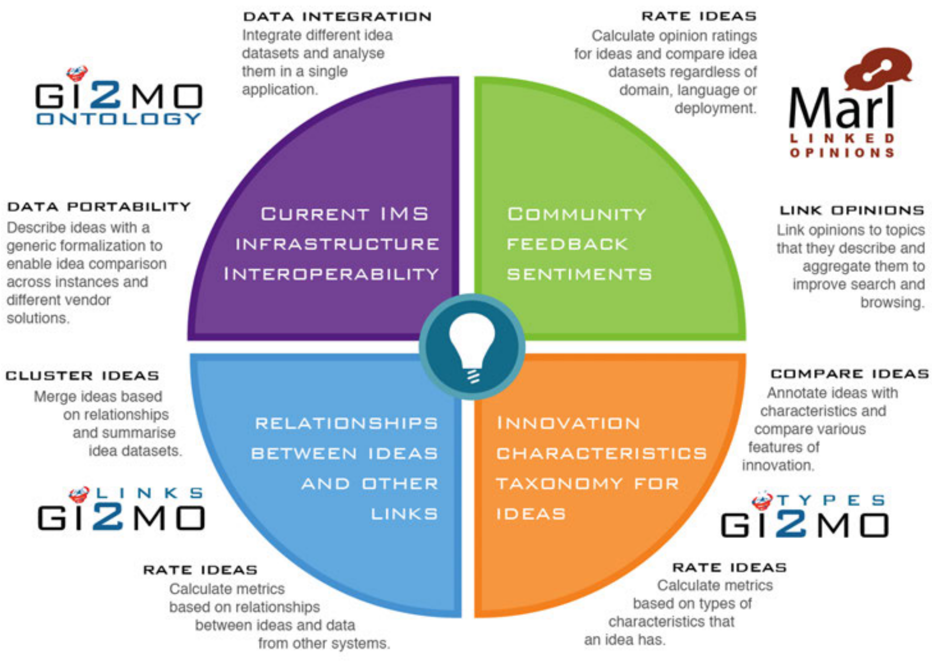


Figure 1.2: Overview of thesis solutions scope and contributions delivered by each

In the presented architecture, the first element of our solution (i.e. 1) addresses a theoretical model of the processes that exist when operating an Idea Management System instance, as well as the practical implementation of that model in a form of ontology that formally defines the data of contemporary Idea Management Systems. The three following elements of the proposed solution (i.e. 2, 3, 4) extend this ontological model by adding new data constructs that do not exist in the contemporary systems and methodologies that facilitate addressing the aforementioned problems of idea assessment and enable to answer the stated research questions. In terms of metadata modelling, each of those three solutions are delivered as extensions for the base Gi2MO ontology.

The research on all proposed solutions has been based on observation of a number of Idea Management System deployments, work in Spanish scientific projects and analysis of theoretical foundations. The project context of the thesis is described in detail in Chapter 2, while theoretical foundations in the contribution areas are summarized in Chapter 3. Each of the solutions presented has been evaluated separately in a series of experiments that test its significance and prove the usefulness of the proposed approach. Therefore, evaluation tasks are described individually in every chapter related to a respectable contribution. Following a detailed presentation of solution elements, we conclude the thesis by summing up the contributions, pointing out the deviation from the anticipated thesis results and suggesting the future research that could be conducted in order to continue the line of work presented in the thesis (see Chapter 8). Finally, the thesis is supplemented with an appendix which describes the impact that the thesis has made on other research projects and practical deployments of the presented research up until the moment of thesis publication (see Chapter A).

Chapter 2

Context

The research of this thesis has been a part of a number of projects that stimulated and influenced its shape. Those projects delivered use cases and evaluation scenarios that we used to test our contributions. Furthermore, throughout the course of the thesis we have gathered a number of publicly available data from various Idea Management instances. Those datasets have also served as a tool for the evaluation. Since effects of all those activities were utilized in each of the solution architecture elements we describe them prior to introducing our contributions.

The main aspects of thesis context discussed in this chapter are:

- Funded projects that influenced thesis creation.
- Datasets and research materials used during the course of the thesis

2.1 Spanish R&D Projects

The research for this thesis has been partially funded by two Spanish projects: THOFU (THOFU, 2012) and RESULTA (Resulta, 2011). Therefore, the objectives and contributions of the thesis are aligned with the context of those projects. Furthermore, each of those projects had particular setting, stakeholders and impact objectives to fulfil. Although the thesis has been built to satisfy generic needs of Idea Management Systems deployed in any environment, the mentioned projects have influenced how we selected our goals and later validated our contributions. In particular, as discussed in the following subsections, each of the projects involved communication with a different kind of communities. This stimulated our research by revealing problems of deploying Idea Management Systems for different scenarios.

2.1.1 THOFU Project

THOFU is a big, long term multidisciplinary project, involving over 30 partners and funded by the Spanish Ministry of Science and Innovation under the CENIT programme. The goal of the project is research and development on innovative technologies for the hotels of the future. The project foundations are established based on participation of major and medium technological companies from throughout the Spain that lead the development activities and subcontract universities or research centres for investigation on new technologies. Furthermore, in terms of impact and dissemination, the project involves a number of partnerships with hotel industry as well as organizations that support development of tourism in Spain.

In the context of this project, the research of the thesis contributed to use of Idea Management Systems for collecting and managing idea proposals from hotel staff and clients. The hotel environment and data obtained from ideas about improving hotels have been used in the thesis as one of the case studies for testing our research contributions. The use cases of THOFU project are examples of situation where Idea Management System is opened to the public and collects a large number of input from a variety of difficult to identify customers. The data is gathered continuously with possible incentives and contests for attracting attention of clients to submit their ideas. The volume of incoming data is often related to a particular period of time when clients visit the hotel or plan their trips (e.g.. tourist season, hotel promotions, travel agency partnerships, local events etc.).

2.1.2 RESULTA Project

RESULTA is a project funded by the Spanish Ministry of Industry, Tourism and Commerce under the Avanza plan, grant number TSI-020301-2009-31. The project involves 12 partners and aims to research on technologies for management of the processes and relationships in the environment of consulting companies (e.g. relationships between companies and clients or companies and subcontractors etc.). The principal directions in the research conducted within this project are: integration of IT systems, social networking platforms for maintaining collaboration and awareness in the consulting company environment, management of enterprise knowledge.

In this context, the thesis has investigated Idea Management Systems integrated together with a number of enterprise platforms and gathering input from employees of organizations. Whereas the previously described THOFU project confronted our research with the problems of idea submitted by crowds from outside the organization, in the RESULTA project we were stimulated to look at problem of a smaller amount of ideas that had more contextual information referring to different repositories and systems across the enterprise. The use cases of RESULTA are about deployment of Idea Management in a controlled environment where all actors can be easily identified. Therefore, the participation in this project stimulated the elements of our research that referred to data interoperability in the enterprise and usage of Idea Management Systems in synergy with other enterprise data silos.

2.2 Idea Management System Deployments and Datasets Used

The participation in the aforementioned funded projects delivered resources and use cases for gathering new data and performing experiments for the need of this thesis. However, during the investigation of the state of the art and the commercial achievements in the area of Idea Management Systems, we discovered a number of publicly available systems that were gathering ideas through a significant amount of time and delivered interesting data. We decided to use those resources during the evaluation experiments for the contributions presented in the thesis.

The final choice of datasets was primarily determined by the availability of the Idea Management System instances to the public during the course of the thesis. However, out of the identified systems that we had access to, we have chosen a set of particular ones based on a number of criteria, such as:

- underlying software capabilities: the progress in the Idea Management Systems domain is mainly stimulated by the achievements of the industry. There is a big number of vendors that offer different software suits and present user with a different workflow in order to submit ideas or collaborate with other members of the community. Those differences result is different kind of ideas gathered.

- domain of application: Idea Management Systems find use in many different use cases ranging from computer software market to toys for children. Depending on the domain ideas are formulated in different ways and focus on communicating different feedback.

- type of community: the identified instances gather ideas from different kinds of innovators. Some communities are more closed and specialist while other systems connect to the mass consumer. Depending on the type of community ideas can be more or less elaborate, contain different amount of references, gain different type of community support etc.

- goals of the deployment: depending on the organizer sometimes the goals why the system has been opened vary a lot. In the case studies that we identified sometimes ideas were collected indefinitely without any particular goal presented to the public, while in other cases the time boundaries and goals where explicitly announced to the audience. The goal and the incentives that the organizer proposes often impact quite significantly the number of ideas and amount of feedback gathered.

- amount of data gathered: depending on all the aforementioned factors public Idea Management Systems attract audience with different successfulness. As a result some instances gather significantly more ideas than others. Additionally, the number of comments and community activity also varies and does not always go in line with the influx of idea amount. The size of data influences the amount of idea relationships that can be identified, the amount of references that users make to each other ideas.

Using the above indicators, we selected the instances that would allow to test our research in conditions of different deployments as well as similar cases to note how different conditions of Idea Management appliance impact our solutions. Furthermore, since one of the goals of the thesis is to provide generic solutions and interoperable data it was crucial to be able to evaluate if indeed our contributions meet those requirements when confronted with real case studies of organizations independent to thesis author.

All instances described below have been open to public and available for access as Internet websites. The data of those instances has been published as HTML and unavailable in any other form. We collected the data of those instances based on the analysis of the entire lifetime of the respectable instances since their start until the time our experiments were conducted (February 2011). For the goals of the thesis, to obtain the data, we have scraped the HTML of the mentioned instances with a number of tools and transformed the information for further analysis to a serialization format for Idea Management Systems that is one of the contributions of this thesis. The details of serialization, data scraping process and data interlinking are described in chapter 4 during the analysis of evaluation activities for the proposed IMS ontology. In the subsections below we only discuss the characteristics of the used datasets according to the aforementioned criteria (see Table 2.1). Later, throughout the particular solution architecture chapters, we refer to those datasets and detail some of their characteristics or compare the datasets depending on the requirements of the particular evaluation activity.

2.2.1 Dell IdeaStorm and MyStarBucks Ideas

Two of the chosen instances are based on the same SaleForce Idea Management System (SalesForce, 2009). Both are administered in a similar manner as indefinite idea competitions: Dell IdeaStorm (Dell, 2009) exists since February 2007, while the myStarbucks (Starbucks, 2009) system is running since March 2008. In both cases, the organizations that own the systems are large multinational corporations with huge user base (e.g. Dell sold 44 million PC units just in 2009 (Dell, 2011), while Starbucks claimed to serve 60 million customers

Table 2.1: List of Idea Management datasets used for throughout the thesis.

System name	#Ideas/ #Comments/ #Users	Area	Case Characteristic
Dell IdeaStorm	15.000/ 90.000/ 2.000	Computers, telecommunication devices and related services.	Focused on collecting ideas for existing products over indefinite amount of time with periodically organized focus sessions
myStarbucks Ideas	8.000/ 80.000/ 3.000	Coffee and related products sold in a coffeehouse chain.	Focused on collecting ideas for existing products and changes in services over indefinite amount of time
Cisco i-Prize	1.000/ 4.000/ 1.000	Computer, networking and communications equipment.	Viewable only after registration and available only during a set amount of time. Focused on collecting very abstract ideas for new area of activity. Introduces considerable money incentives for best inventors.
Adobe Acrobat Ideas	500/ 2.000/ 600	Computer software	Collects ideas on a software product for viewing and editing electronic documents. Opened for an indefinite amount of time and gathering input from a variety of users of different operating systems of personal computers.
Ubuntu Brainstorm	27.000/ 90.000/ 2.000	Open-source operating system and related software.	Very collaborative, computer literate community gathered around open-source software products. Apart of ideas system enables submission of proposed implementation methods for ideas.

weekly in 2011 (Starbucks, 2011)). Up until the time the data was mined both instances presented similar user interface and workflow for the innovators as well as participants of the community. We have chosen those two instances to see if systems deployed in the same way from the perspective of infrastructure as well as idea management practices would diversify due to the fact that ideas are collected for different kinds of products.

2.2.2 Cisco i-Prize

The IMS instance called i-Prize (i-Prize, 2010) is operated by multinational corporation called Cisco. The instance is based on an Idea Management platform by Spigit (Spigit, 2012), started running in February 2010 and was only open for three months. Apart of setting a limited time frame for the collection of ideas, Cisco also offered considerable money incentives for the winners that proposed the best ideas. In contrast other instances do not have any incentives apart of public mentions of the winning ideas. Additionally, the goal of i-Prize contest was to collect ideas for a new major future Cisco business while in all other instances there were no precise goals other than gathering feedback from clients on current products and services.

2.2.3 Adobe Acrobat Ideas

The Acrobat Ideas instance (Adobe, 2012a) belongs to Adobe Systems - an American multinational computer software company (Adobe, 2012b). The system is running based on BrightIdea platform called WebStorm 5.0 (WebStorm, 2009) and has been collecting ideas in an ongoing call since the beginning of 2009. In comparison to all the previous instances, this system was designed to collect ideas only for a single service: Acrobat.com portal - set of online services that allow to operate on PDF documents (like file sharing and storage, PDF converter, online word processor etc.). Acrobat.com is a supplementary solution to the traditional PDF toolkit of Adobe called Adobe Acrobat.

From the point of view of our evaluations, Acrobat Ideas was included as an instance that gathers a community of users of computer equipment similarly like Dell but focuses only on a single product thus allowing users to inspect and follow each others ideas more carefully. Furthermore, Acrobat user base is significantly smaller than in case of Dell (according to BrightIdea Acrobat.com gathered 5 million registered users (Greeley, 2009)), which had an impact on the amount of people eager to participate in the idea competition (over 3 times less than in Dell).

2.2.4 Ubuntu Brainstorm

The final dataset included in our tests came from Canonical's Ubuntu Brainstorm (Ubuntu Brainstorm, 2012) that was opened in February 2008 and is based on an open-source IdeaTorrent platform (IdeaTorrent, 2012). In comparison to the previous instances, the idea submission rules are different and force innovators to deliver solutions for their ideas. Another major difference is that Canonical user base is smaller in comparison to Dell or Starbucks (20 million users total as estimated by Canonical (Ubuntu, 2011)) but also very collaborative (Lakhani and von Hippel, 2003; Feldstein, 2007) and only focused on a single type of an open-source product. The implementation process of ideas is significantly more transparent due to the fact that Ubuntu is an open-source project and all its production infrastructures are available to public and linked to Brainstorm. We have chosen to analyse this instance to see if the computer technology literate audience of Canonical that is used to giving contributions for free would propose ideas that differ in comparison to mass consumer customer base of Dell and Starbucks.

Chapter 3

Foundations

In terms of an extensive introduction to our research achievements, we describe the theoretical background of the thesis. The goal of this chapter is to make the reader accustomed with the terms mentioned throughout the thesis and present the state of the art in the areas where the thesis contributes new solutions. The sections of this chapter are organized to reflect the areas that are gradually introduced by subsequent chapters that describe the thesis contributions.

Firstly, we define Idea Management Systems, their scope as it is today and summarize the past and current research done in the area. Secondly, we note that Idea Management Systems are just a single solution in a more broad domain on Innovation Management. Therefore, we present the state of the art in Innovation Management from the angle that is investigated by the thesis - innovation modelling. Finally, since all models proposed by thesis are evaluated in the context of interoperable IT systems, we present how knowledge models have been applied in the past for: systems integration, data interoperability and information assessment - which are the ultimate goals of this thesis.

The information presented in this chapter describes the state of the art as it is without comparison to our particular contributions. The recap of related work, including critical review and advances made by the thesis, is presented later individually per every chapter, in their respective 'Related Work' subsections.

The chapter presents state of the art in four major areas related to thesis contributions:

- Idea Management Systems research and development
- Research on organizational innovation models
- Opinion Mining and its use cases
- Use of Semantic Web for knowledge formalization and interoperability

3.1 Idea Management Systems research and development

3.1.1 Brief History of Idea Management Systems

Innovation management practices are not new and have been introduced in various organizations much before the burst of IT systems (Gorski and Heinekamp, 2002). For instance, Toyota has a history of over 30 years of innovation management oriented towards the capture of ideas (Baumgartner, 2004). Aside of such practical uses the research on notion of employee ideas in business context can be tracked back as far as to 1930s when it was proposed by Bower (Bower, 1930). The term 'idea management', as used today in relation to the IT market, has been created in reference to systems that emerged in the late 90ies (Rozwell et al., 2002). The coining of term 'Idea Management System' is not widely attributed to a single person but originates from the industry and started to be used by a number of companies offering tools for employee idea collection. Some of the early ones were Imaginatik founded in 1994 (Imaginatik, 2012) and later Bright Idea created in 1999 from General Ideas Inc (BrightIdea, 2012). With respect to existing software at the time, the novelty of those platforms was to aid practices of innovation management by allowing organizations track community generated ideas as they progress through enterprise procedures. The way this vision was implemented as well as goals and scope of those tools have been continuously evolving ever since their origins.

Historically, the precursors of Idea Management Systems were simple suggestion boxes maintained as part of internal corporate systems or with the advent of Internet – company homepage. However, this approach did not introduce any software facilities that would actually aid the management of captured community ideas. These suggestion boxes were just an additional input mechanism. The progress came with connecting the technology with dedicated back-end facilities. The abilities to store, display and organize the submitted ideas gave birth to Idea Management Systems. One of the drawbacks at the time, that limited the software capabilities, was simple user input structure. This has changed along with the huge popularity burst of the Internet and the rise of the so called Social Web. The Idea Management Systems have taken advantage of the Web 2.0 techniques to extend the original submission boxes as idea capture methods. As a consequence, more rich and better organized user input data brought new opportunities to develop management backends towards better data presentation and selection.

While the initial period of Idea Management Systems evolution was about harnessing basic technologies and setting directions, the contemporary systems focus on defining a formalized software-aided idea management process that is well defined, traceable and most importantly repeatable. On top of that, in search of new methodologies, some additional practices are proposed to extend the existing phases towards other areas of innovation management, e.g. the idea generation towards creativity studies (implemented in Ingenuity Bank (IngenuityBank, 2009)) or idea assessment and status monitoring towards market studies and strategic planning (e.g. in Accept Ideas (AcceptIdeas, 2009)).

3.1.2 Classification Scheme for Idea Management Applications

Along with a large amount of Idea Management System vendors comes a big number of case studies. In the following section we shall describe the typical characteristics of situations in which Idea Management Systems have been applied with success (see Fig. 3.1).

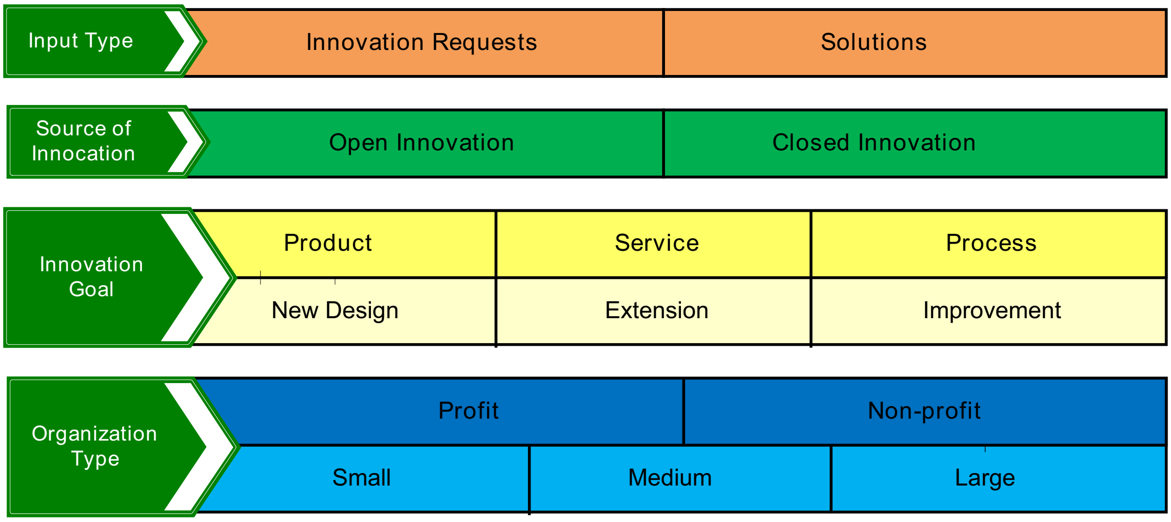


Figure 3.1: Classification of Idea Management Applications

Organization Type. The discussed systems find applications in organizations of different sizes and business models. Typically large enterprises are known to be the main customers for Idea Management Systems (Driver, 2003), however some studies of small/medium enterprises are also available (Nielsen, 2006; Brem and Voigt, 2007). In addition, while the most often use case of Idea Management is to increase revenue in private companies, it is also known to be used by non-profit organizations (e.g. International Olympic Committee using IdeaLink (BrainBank, 2009) or cure for cancer research by Cancer Research UK supported with BrightIdea platform (BrightIdea, 2009a)).

Goals of Innovation. The way Idea Management Systems are applied depends on what the organization wishes to innovate. The Idea Management Systems have been claimed (Imaginatik, 2009) to give best results when the user community is driven to provide input for a specific and very narrow area. In many systems this is achieved by organizing time limited idea collection events that are supposed to reach a specific goal tied to the of organization activity and services. The majority of use cases published define goals that are either product or service oriented (e.g. Storm Sessions in Dell IdeaStorm (Dell, 2009)). However, Idea Management Systems can also be used to introduce innovation in internal organization processes (e.g. Toyota corporation is often quoted to employ a long standing tradition of such activities (Warner, 2002; Das and Puri, 2003)). Furthermore in both cases, the innovation goal can be pushed towards a completely new design (e.g. new product or a company processes), extension (e.g. adding new product features) or improvement (e.g. changing existing product or optimizing processes).

Source of Innovation. Depending on the organization profile and innovation goals put ahead, it is also needed to adjust the group of people that contribute ideas. Idea Management Systems are known to be used both internally in companies as well as to communicate with customers, thus relating to practices of Closed Innovation and Open Innovation (Chesbrough, 2003). In the first case, the often quoted benefit is that the low level organization employees (e.g. engineers) are closest to product development process and may have a large number of valuable ideas that otherwise would never reach the decision level management if not for the idea management practices. On the other hand, Open Innovation is also perceived to provide huge benefits since customers directly communicate what kind of product they wish to buy and share their desires. The choice of innovators community is an important factor for the Idea Management process since it drives the volume and characteristics of data submitted.

Input Type. Finally, the type of idea management system used and results produced are also dependent on what kind of input is gathered. Among the most popular situations, the contributed ideas are formulated either as requests for certain innovation or as solutions to specific problems (or in same cases both). While some systems deliver generic front-ends to capture and assess ideas (e.g. BrightIdea WebStorm (WebStorm, 2009), Imaginatik IdeaCentral (IdeaCentral, 2009)) others are strictly oriented for submitting solutions to advertised problems (those are most commonly referred as Idea Marketplaces and have a very big representation among commercial portals that gather inventors from specific areas (OpenInnovators, 2009)).

An example of the proposed classification for earlier described IMS instances used in the thesis evaluation can be seen in Table 3.1. The analysis of those case studies as well as the aforementioned ones as examples for particular classification categories, shows that Idea Management can be used in many situations. However, regardless of the method and area of application, all case studies prove that Idea Management solutions can not be expected to achieve results when deployed and left unattended. In every case reporting companies emphasize the need for constant support from management. The amount of effort dedicated to facilitate innovation process varies depending upon each of the factors that we enumerated. In most cases this includes both full time employees dedicated to idea management process and also part time staff from the different locations in the organization structure (e.g. engineers, department level managers, high level management etc.).

3.1.3 Idea Management Systems Research so far

Regardless of the domain of appliance and progress in the industrial solutions, the use of Idea Management Systems still faces a number of problems. Case studies across the categorizations presented in the previous section report that those problems are often related to the high amount of feedback that is being gathered (Jouret, 2009) and the small amount of valuable ideas in comparison to total submitted (Voigt and Brem, 2006). As a result a lot of effort is needed during the generation phase (to support the motivation of innovators) as well as evaluation phase (to review and select ideas) so that the initial investment in idea management does not go in vain (van Dijk and van de Ende, 2002). Furthermore, studies of

Table 3.1: Example of case study classification for Idea Management System instances used in the thesis

Case Study	Request	Solution	Open Innovation	Closed Innovation	Product	Service	Process	New Design	Extensions	Improvement	Profit	Non-Profit	Small	Medium	Large
Dell IdeaStorm	x		x		x	x	x	x	x	x	x				x
My Starbucks	x		x		x	x	x		x	x	x				x
Cisco i-Prize	x	x	x		x	x		x			x				x
Acrobat Ideas	x		x		x					x	x				x
Ubuntu Brainstorm		x	x		x	x	x	x	x	x	x			x	

Nortel (Montoya-Weiss and O'Driscoll, 2000) show that problems of idea management are not only connected to efficiency but also quality assurance measures for the entire process, i.e. guidance for inventors on how to formulate ideas and support for evaluators regarding review criteria aligned with company objectives.

Taking notice of those problems there have been a number of solutions that propose a shift in management of organizations to affect how people use Idea Management Systems (e.g. related to new management processes (Glassman, 2009)) as well as changing the software itself. Among the software related approaches that are the interest of this thesis are: 1) solutions that add new activities for the Idea Management System user or propose a shift in the interaction method to produce new data that could be later used to address some of the aforementioned problems; or 2) solutions that introduce new technologies for automatic or semi-automatic data analysis in the back-end to aid knowledge management regardless of user interaction during the idea generation phase.

In the first group, Bothos et al. (Bothos et al., 2008) experiments with prediction markets where innovators buy and sell shares of ideas thus setting their value via mechanics similar to stock markets. Similarly, Witt et al. (Witt et al., 2011) proposes a shift in interaction methods but achieved though game mechanics that are applied to stimulate user motivation for submitting ideas and encourage participation in idea evaluation. In relation to those two contributions, Klein (Klein, 2012) fundamentally changes the interaction by introducing argumentation maps that impact both user attention and allow to generate metrics for idea assessment. Finally, a number of works propose different approach to all aforementioned where new interaction methods help in improvement of quality of the ideas as they are submitted rather than their filtering or assessment in a post processing stage (Ford and Mohapatra, 2011; Baez. and Convertino, 2012).

On the other hand, some solutions claim that the road to improvement of Idea Management Systems leads though better knowledge management of the already available data. Kornish and Ulrich (Kornish and Ulrich, 2011) propose use of various clustering techniques to downsize the amount of ideas for assessment phase. Furthermore, there have been a number proposals related to use of ontologies, namely: Ning el al. (Ning et al., 2006) propose a general vision of an enterprise connected with the use of Semantic Web technologies, Bullinger (Bullinger, 2008) delivers a framework called OntoGate that uses ontologies for modelling details of enterprise structure, goals, strategy etc. for idea screening, while Rield et al. (Riedl et al., 2009a) analyses service innovation and proposes ontologies for data integration.

As it can be seen the Idea Management system research is rather scattered between solving various problems and there has not been a single strong direction or area of research. One of the contributions of this thesis is to organize the processes, user activities and problems related to interaction with Idea Management Systems into a single framework. Therefore, as we define the Idea Life Cycle in Chapter 4, the aforementioned contributions to the state of the art on Idea Management Systems are further investigated in-depth together with comparison to industrial solutions. Similarly, the selected aspects of certain solutions are compared in more detail to the contributions of the thesis as the related work in presented in following chapters.

3.1.4 Related research areas

In the following section we shall describe concepts and disciplines related to Idea Management Systems. Some of them are direct predecessors of idea management while others exist as side reach or businesses. In addition, non of the concepts described below has clear borders, therefore quite often an intersection can be depicted. The following descriptions shall focus on brief characteristics of areas and pointing out differences or similarities and relationship to Idea Management Systems.

Knowledge Management

Knowledge Management in general practice of organizations refers to identification of sources of information, extracting knowledge from those sources, processing it, and delivering to relevant people as they need it (Gamble and Blackwell, 2002). Among other, the purpose of Knowledge Management is re-use information that is already in organizations possession as it is crucial not to redo the same work twice. However, as the organization structure grows, it becomes harder to manage the increasing knowledge. Therefore, often it happens that time and money is wasted on the same task twice because it is easier to repeat the process rather than find the needed information in the company repository or get through to the correct person. The research in Knowledge Management area attempts to find a remedy to searching vast information silos. The key goals are improving the accessibility of information with biggest impact on the accessibility speed and interface simplicity. A typical process for Knowledge Management is where a person presents a query and he is given a precise answer with references to available resources.

In terms of computer systems that support those activities, Mika et al. (Mika and Akkermans, 2003) categorizes solutions into static and dynamic. The static solutions focus on development of models to organize and structure information in enterprise systems, which is often done though classification of knowledge with natural language processing techniques according to some predefined taxonomies. On the other hand, the dynamic solutions refer to moving the knowledge to and from repositories in order to deliver it as needed.

Although some practices of Idea Management Systems could be classified under the above categories, there are also some differences. The key concept of idea management practices is to obtain and organize the information that has not been produced yet in the enterprise. Therefore, the goals are: stimulation of knowledge generation, cataloguing the knowledge and most importantly knowledge evaluation. As a result, the discipline research is about finding ways to provoke a selected community to produce innovative ideas in a formalized manner. Once this is available the key task for idea management is to extract the best ideas through ranking and other techniques.

The intersection with Knowledge Management is in the interest in knowledge organization. However, Idea Management Systems are much more precise and focused. The goals are more clear and the discipline is much more narrow. The key difference is that a user approaching a typical corporate knowledge management system does exactly know what he wants, while the beneficiary (or operator) of the Idea Management System does not have a precise idea what type of knowledge he wants to extract from the system. The search is much more profile and rank based.

Open Innovation

Open Innovation (Chesbrough, 2003) is a term that assumes to allow interaction between organization and the community gathered outside of the organization. Specifically, the aim is to gather input about everything relevant to organization ecosystem that can introduce innovation. However, following the extensive reviews of the state of the art in the area (Dahlander and Gann, 2010), this openness can be perceived and understood in many different ways. For instance, open innovation can be inbound or outbound (Huizingh, 2011), that is: implemented as internal use of external knowledge or giving the ability for external use of internal organization resources (e.g. through publishing research results and engaging in discussion).

In this context, Idea Management Systems are a technology that supports Open Innovation concept with the respect to idea generation and selection. The Idea Management Systems provide a front-end that allows for the company to reach its community through the Web technologies. However, the same systems are also often used for closed innovation within companies to gather ideas from employees.

Furthermore Idea Management Systems are mostly using submitting brief ideas, feature requests and desires of the clients. Open Innovation supported by IT does not have to limit to mere management of ideas, it can also mean accepting entire solutions or even stimulating or subcontracting external parties to provide a particular solution to a given problem.

Crowdsourcing

Crowdsourcing (Howe, 2004) is a term that connects to a situation that inevitably happens when Open Innovation idea is employed by a big organization in the context of its clients. The input is provided by many people and it is needed for a company to harness that data. Crowdsourcing not only declares the need for cooperation with individuals outside the company, it goes a step further and points out that the real value is in reaching community to perform some task difficult or costly to achieve by a single individual or organization (e.g. analyse large amount of information like in SETI@Home project (Anderson et al., 2002)).

From the idea management perspective, by applying the aforementioned notion of Open Innovation the organizations connect to crowds in order to obtain ideas. Therefore, the Idea Management Systems implement crowdsourcing principals during the idea generation phase. Furthermore, during the evolution of Idea Management Systems it became apparent that one of the key problems is to cope with large volumes of data. One of the solutions to this problem is to apply crowdsourcing to evaluate ideas via voting mechanisms and comments.

Brainstorming

Brainstorming is a technique for stimulating creativity though spontaneous generation of ideas with a goal to solve a particular problem (Osborn, 1957). Although brainstorming can be performed by individual, often it is implemented and investigated in terms of group work (Zainol et al., 2012; Aiken et al., 1994; Barki and Pinsonneault, 2001).

In relation to idea management practices in general, brainstorming can be modelled as one of the input processes that deliver ideas into the system of an organization (Applegate, 1986). Bozios et al. (Theodoros Bozios, 2009) classify brainstorming as one of the first activities within (1) creativity support in innovation support systems followed by: (2) discovery, transfer and customization of ideas; (3) idea management; (4) knowledge management and collaborative work. In context of the contemporary Idea Management Systems, Lorenzo et al. (Lorenzo et al., 2011) explores this connection in detail though the contribution of brainstorming ontology for on-line communities and analyses its appliance in relation to already existing solutions for Idea Management Systems. Aside of those similarities in knowledge flow and data modelling, there a number of differences can be observed in terms of scale (typically Idea Management Systems are for large communities, while brainstorming is more often implemented for smaller groups) and control (in brainstorming ideas are created spontaneously but there has to be strict control over the process including withholding criticism (Osborn, 1957), which is often not the case for idea management).

Computer-supported cooperative work (CSCW)

The CSCW is a research area related to systems that support collaborative work in a much wider scope of than Idea Management Systems. With respect to all aforementioned related areas, the CSCW is focused on computer systems development and research while the other terms originate from organization management studies. According to Gurdin (Grudin, 1994), CSCW started as an effort by the technologies to connect to economists and social psychologists in order to understand how group activity works. Therefore, most of the contributions of CSCW are located in the area of computer systems that improve human communication as well as capture organizational knowledge (Kraut, 1996). In particular, the state of the art in the area can be divided into support for: communications, information sharing and coordination (Poltrock and Grudin, 1999).

In relation to Idea Management there have been a number of works related to idea generation though groupware (Lu and Mantei, 1991; Vivacqua et al., 2010; Garcia et al., 2010; Applegate et al., 1986). However, contrary to Idea Management Systems as used today, the majority of works in the CSCW area focus on coordination of small and medium size groups within an organization rather than large scale deliberation (Beaudouin-Lafon, 1999; Grudin, 1994).

Innovation Management

Similarly as CSCW can be labelled as research area encapsulating Idea Management Systems research from the computer science and IT systems point of view, the Innovation Management is a domain that collects all research on innovation aspects and relates them to organization, including idea management practices (Trott, 2008). More precisely, Innovation Management research focuses on recognition how innovation is created within an organization and furthermore how can it be stimulated, controlled and transferred. According to definition of Afuah (Afuah, 1998) the ultimate goal of innovation management is to improve the organizations for gaining and maintaining a competitive advantage in the business world.

In this domain, idea management is viewed as a organization management process that is a part of a bigger framework for a systematic approach to the strategies and processes for organizing innovation practices in the company (Afuah, 1998). On the other hand, Idea Management Systems are a tool that provides support for this process. In comparison to the presented earlier literature on Idea Management Systems perceived from the computer science point of view, the research on the topics of idea management processes in enterprise (Gorski and Heinekamp, 2002) and idea generation (often referred to as the fuzzy front end) (Koen et al., 2001; Husig and Kohn, 2003) is significantly more explored. Among those, Glassman studies the relationships between the use of support systems and idea management processes. His studies show the necessity for control of the idea generation phase (Glassman, 2009) and management of the associated activities in a form of organizational processes. Furthermore, Bakker et al. (Bakker et al., 2006) link those idea management processes to the rest of the innovation management study and enterprise activities such as product implementation and successful selling of ideas (Bakker et al., 2006).

In the thesis the relationship between Idea Management Systems and Innovation Management is investigated in more detail from the point of view of data creation and management. The thesis contributes a notion for reusing elements of the theoretical models of Innovation Management for describing ideas in an Idea Management System. Therefore, a more detailed state of the art review of innovation models in presented in the next section.

3.2 Organizational Innovation Models

The Innovation Management domain introduced in the previous section has been built and is evolving based on various theories and models on how innovation processes work, i.e. how new ideas are created, what are the parties involved, what are the types of innovation, what are the effects of innovation etc. Throughout the existence of Innovation Management those questions have been risen in order to improve understanding of innovation in an organization and help enterprises to manage it in a better way with respect to changing situation in the organization ecosystem.

One of the first scientists who shed light on this topic was Schumpeter (Schumpeter, 1934) who's contributions on destructive nature of innovation and its various types led other researchers to develop e.g. the split between radical and incremental innovation. In the state of the art of Innovation Management there have been a number of models created based on different studies and with different goals in mind (Garcia and Calantone, 2002; Popadiuka and Choo, 2006; Damanpour and Gopalakrishnan, 1998; Chuang et al., 2010). Often the creation of those models has been stimulated by the changes in economies, e.g. globalization (Ortt and van der Duin, 2008) or a widespread shift from manufacturing to service oriented economies (Susman et al., 2006).

In context of those theoretical studies on the nature of innovation, one of the research problems of this thesis is to understand how ideas differ from each other beyond the domain context (i.e. topic of idea). To accomplish this goal the thesis refers to innovation models of Innovation Management domain and compares those models to the reality of community created innovation in Idea Management Systems 6. Therefore, in the following section, an overview of most important organizational innovation models is presented. The selection of models presented is based on the state of the art reviews by other authors and presents the contributions considered as most important for the field (Garcia and Calantone, 2002; Afuah, 1998). However, since the thesis research focuses on innovation taxonomies, the models of Innovation Management are only overviewed from the point of view of new innovation types that they introduce. The goal of the section is to give the reader an idea how vthe interpretation of innovation types progressed and what were the particular goals and reasons behind the new categorizations. The discussion of usage of those models as well as various definitions and interpretations of innovation for Idea Management Systems is presented later in the chapter referring to contribution of model for idea characteristics 6. The economical aspects of the presented models and impact of the models on organization management are omitted since those topics are not related to the thesis and do not present any major important for understanding the presented contributions.

3.2.1 Overview of Innovation Models

Schumpeter model

Joseph A. Schumpeter is considered as one of the pioneers of Innovation Management who proposed to look at different types of innovation in economy studies. In his work he popularized the concept of "creative destruction" which refers to the notion that every new innovation inevitably causes destruction of value of the prior state (Schumpeter, 1934) . Based on this theory, Drejer (Drejer, 2004) interprets the Schumpeterian perspective as a model covering five types of innovation (see Fig. 3.2). In this categorization the distinction between different types of innovation is in terms of object of innovation, i.e. item or concept that is the subject of change though introducing an innovation (e.g. change in a product or change an organizational process).

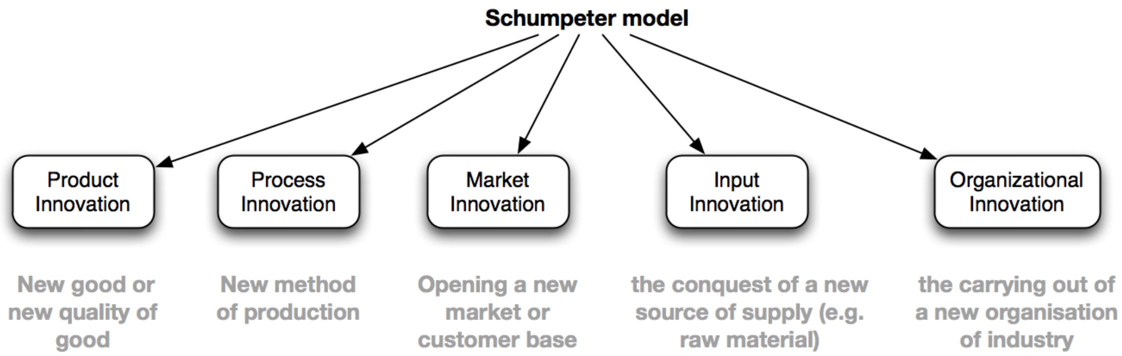


Figure 3.2: Schumpeterian types of innovation as interpreted by Drejer (Drejer, 2004)

Incremental and Radical Innovation

Following the initial contributions of Schumpeter, one of the popularly used dimensions of innovation is the split between incremental and radical. Such interpretation is difficult to attribute to a single researcher as it has been noticed in a number of theories under various names e.g. by Abernathy (Abernathy, 1978) as incremental and radical or by Porter as continuous and discontinuous innovation (Porter, 1986). In principle, the distinction between those two terms can be interpreted as follows: the incremental innovation refers to gradual progress made based on observations of previous changes in a product and existing knowledge of the organization, while the radical innovation in perceived as appliance of a completely new mindset and resulting in major advancements that make previous state irrelevant (thus connecting to the Schumpeterian theories of destructive innovation). Nevertheless, as noticed by Garcia (Garcia and Calantone, 2002), the understanding of innovation, even with regard to those two terms can be quite different depending on scope of analysis of the organization and the market. This can be observed in the more recent models proposed as an extension to this distinction between radical and incremental innovation.

Henderson Clark model

The model proposed by Henderson and Clark (Henderson and Clark, 1990) notices that the initial split between radical and incremental is insufficient to explain effects and origins of innovation in an organization. As a result, architectural and modular innovations are added that differentiate innovation according to the scope of technological change and the amount of knowledge required to make it possible (see Fig. 3.3). The architectural innovation can be defined as change in assembly rules of the same components as before, while modular innovation is a technological change in internal design some particular element.

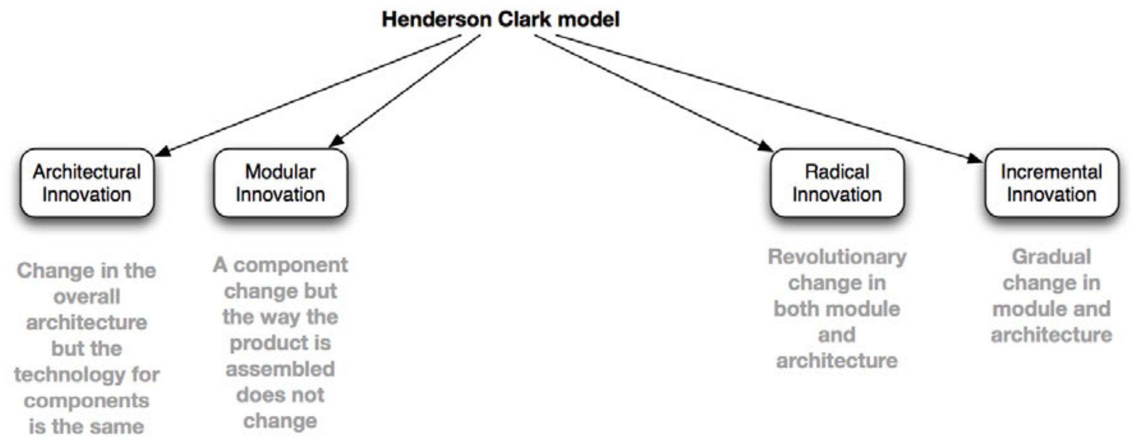


Figure 3.3: Types of innovation as perceived by Henderson and Clark

Abernathy-Clark Model

The another model proposed by Clark together with Abernathy (Abernathy and Clark, 1985) discusses similar problems to the previous one but explores a different aspect of big and small changes in product. In particular, the model discusses the context of expanding the market along with introduction of product changes (see Fig. 3.4). Although the naming of innovation types is similar in this model as in the previous one, the novelty is in the new interpretation and positioning of the established types in relation to each other along two dimensions: innovation that causes product change vs. lack of change; and secondly new markets and customers vs. old markets.

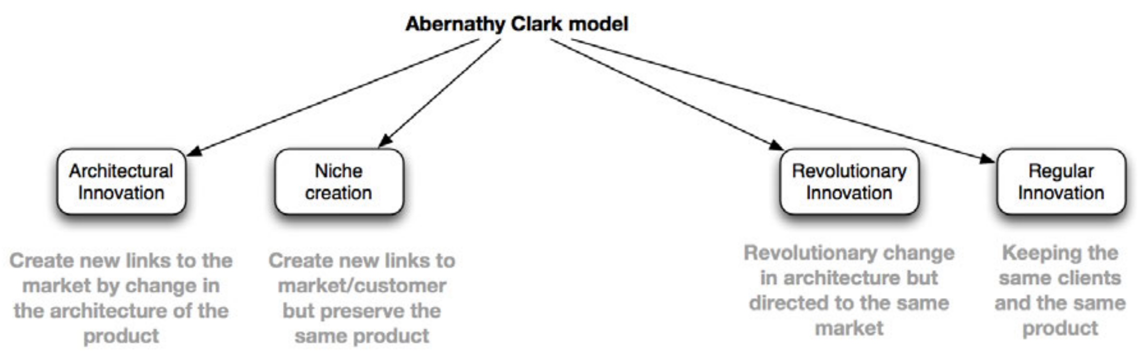


Figure 3.4: Interpretation of types of innovation by Abernathy and Clark

Furthermore, Abernathy and Clark introduced a competence model that lists the changes in organizational knowledge depending on the innovations made along the aforementioned innovation dimensions: technological product change impacting production knowledge (e.g. product design or manufacturing competency change); and market change impacting knowledge about customer relationships and models of customer communication (e.g. necessity for new distribution methods).

Teece Model

The Teece model (Teece, 1998) expands on the previous theories and introduces the relation between imitability of innovation and complementary assets (see Fig. 3.5). The imitability of innovation is described as capability of competitors to copy a product or deliver a similar one that would resemble the characteristics of original. On the other hand, the complementary assets are the non-technological elements that enable to deliver the innovation to the market (e.g. distribution channels, brand name etc.).

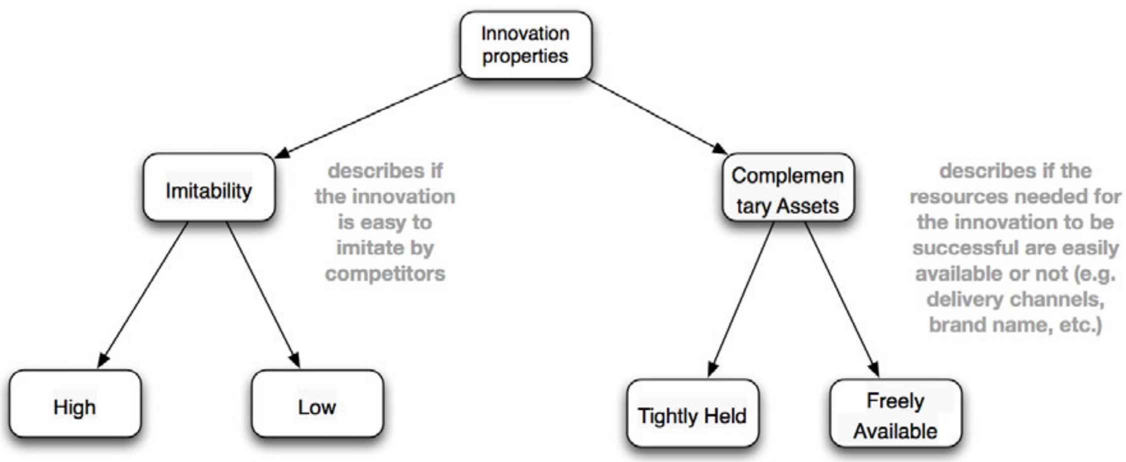


Figure 3.5: Innovation types as proposed by Teece

Based on this model, Teece points out who holds the upper hand after the innovation is introduced and who will get most profit out of introducing the innovation to the market (e.g. innovations that have freely available complementary assets and high imitability are unlikely to give profit to the original innovator, as they will easily get copied by competitors who have better or similar distribution channels).

Chandy Tellis

The model proposed by Chandy and Tellis (Chandy and Tellis, 1998) refers to a number of earlier mentioned types of innovation (i.e. radical, incremental and market/ technological) but defines a new interpretation of them using two new characteristics: newness of technology and fulfilment of customer needs (see Fig. 3.6). In the Chandy-Tellis model combinations of those characteristics lead to certain types of innovation, e.g. if the innovation is proposed in an old technology and the customer need fulfilment is low than the innovation can be characterised as incremental, while innovations in new technologies that fulfil a lot of needs are radical. Similarly, technology / market dimension: innovation in new technologies that lead to satisfy little needs of the customer are characterised as technological breakthroughs, while the market breakthroughs are associated with innovations in old technologies that bring high customer fulfilment.

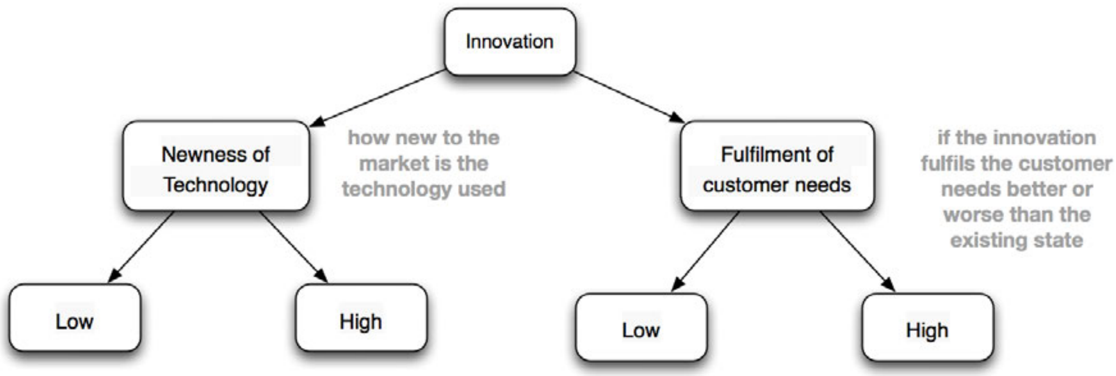


Figure 3.6: Interpretation of innovation types by Chandy and Tellis using two new categories

S-Curve model

According to Afuah (Afuah, 1998) innovation models can be split into static and dynamic. Contrary to all the previously presented, the S-Curve model is a dynamic model. That implies the model analyses dynamics of market or organizational processes and describes types of innovation with regard to the changing conditions (Rogers, 1962).

In particular, the S-Curve model relates to the adoption of innovation, i.e. phases such as introduction, growth and maturity with respect to the amount of adopters on the market (see Fig. 3.7). Based on various stages of adoption, the model describes the profits and difficulties of introducing an innovation (e.g. early adopters get a head start over the competition but technologically need to overcome many obstacles, their risk of failure is high as are the costs of innovating).

Abernathy Utterback

Similarly as the S-Curve, the model of Abernathy and Utterback (Abernathy and Utterback, 1978) can be classifeid as dynamic. It does not apply a categorization of innovation byt defines life cycle phases of innovation. On each stage those life cycles the model defines certain characteristics that influence adoption of the innovation and profitability. Therefore, the model presents the dynamic of innovation as it progresses though the organization and

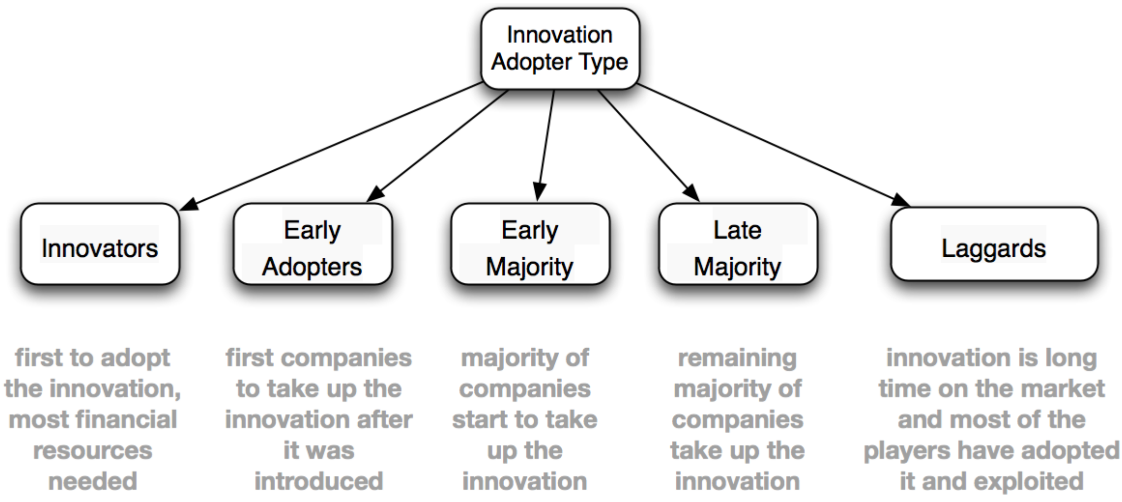


Figure 3.7: Categorization of innovation based on the S-Curve model of innovation adoption

becomes mature. With this respect the model relevant for the later proposed Idea Life Cycle in chapter 4 of the thesis.

Tushman

The model of Tushman and Anderson (Tushman et al., 1997) is another example of dynamic approach to the theory of organizational innovation. Similarly as the Abernathy-Clark model it looks into concepts of existing market vs. new market and incremental vs. radical, however describes the dynamics of the process. In particular, the new contribution is about understanding of the intersections between those categories and crossing from one to another. Additionally, Tushman and Anderson put in the middle "generational innovation" which refers to releasing new product iterations (see Fig. 3.8). The dynamic aspects of the model are its alignment to perception of time and how the innovation progresses over time.

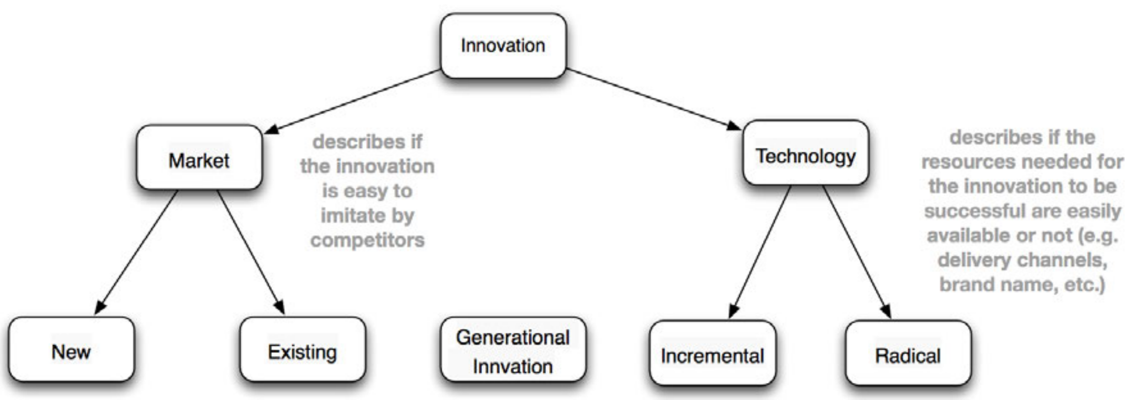


Figure 3.8: Types of innovation derived from the dynamic model of Tushman and Anderson

3.3 Opinion Mining

Recent years have brought the burst of popularity of community portals across the Internet. Alongside of the content created by the portal editors the so called user-generated content is an important part of many websites. Users provide their input not only through discussions and personal notes in various social web spaces (boards, blogs etc.) but also on mass scale leave comments and reviews of products and services on numerous commercial websites. The fast growth of such content has not been fully harnessed yet. Information contributed by the users is often very disorganized and many portals that enable user input leave the user added information unmoderated.

Opinion mining (often referred as sentiment analysis) is an attempt to take advantage of the vast amounts of user generated content. It employs computer processing power to formalize the knowledge taken from user opinions and analyze it for further reuse. Although there are some early works about recognition of subjective texts from early 80s and 90s, the real progress in the area started with the rise of Web 2.0. The new types of Internet content enforced new ways of data management which ,as a consequence, caused new problems and opportunities to arise. Over the last decade a huge increase of interest in the sentiment analysis research is clearly visible (Esuli, 2007a; Wiebe, 2011).

In this thesis the opinion mining technologies are used as a tool to analyse community feedback submitted in form of idea comments. Therefore, in order to provide an introduction to the topic, the following section summarizes the key directions in opinion mining research. Furthermore, the novelty that this thesis proposes is a new usage context for opinion mining: metric generation in Idea Management Systems. Therefore, we also present the previous use cases for the opinion mining algorithms and provide a short discussion of their importance mostly with connection to Internet technologies (see Sec. 3.3.5).

3.3.1 Domain overview

The opinion mining is often associated with another research topic information retrieval (IR). Nevertheless, opinion mining proves to be a lot difficult task. The primary reason is characteristics of the data sources. In IR, the algorithms operate on factual data, while in opinion mining input data is only subjective information. In practice, this means that opinion mining is needed to go a step further than information retrieval and analyse sentences and phrases deeper with respect to their semantics. During the facts analysis one is interested in simple characteristics and extracting it. In opinion mining the additional task is to determine the nature of opinion: whether it is positive or negative in general; what features does it describe; what features are valued, which are not etc.

As mentioned before, the rise of interest in the area has been caused by the growth of user-generated content on the Web. One of the primary characteristics of such content is its textual disorder and high diversity. The style of writing opinions varies a lot within a single portal but even more if one was to analyse a given topic in the Internet wide scale. Opinions are expressed with informal language. Therefore sentence construction can very a lot depending on the community (which can go even as far as altering grammar within a single language). For instance the product reviews contributed on Amazon about a movie and a computer game based on it can be written totally different and even sometimes not understandable for people outside of the community. Typically, information retrieval techniques achieve best results when applied to highly structured, formalized text, in most cases opinion mining does not have this comfort. In order to give more insight into the problem, in the subsequent subsections we describe various attempts to classify and formalize different opinion types.

Types of evaluation

In general, there are two main ways to express sentiments: direct opinions and comparisons. Direct opinions usually describe one object and contain some adjectives that refer to it (i.e. the image quality of this camera is good). In contrast, the comparative statements mention more then one object and describe some sort of relation (i.e. the image quality of camera X is much better then camera Y).

Types of context

To extract the opinion one has to to know what the opinion is about. Depending on the location/portal the descriptive information can be stated in many different ways. On review portals it is often relatively easy to extract sentiment information but for instance on a forum it is considerably harder to identify the subject of discussion or subject of a single post.

As it is shown further (see Sec. 3.3.4), the software that extracts opinions and performs any kind of automatic sentiment recognition is often built for specific contexts. The generic engines and algorithms perform much worse then applications meant only to analyze particular types of text (i.e. movie reviews). Although this is a huge limitation , in practice such techniques still hold a great value even with scope narrowed down to opinion mining posts from a single portal.

Level of interest

People can express their opinions with different detail. Some will give general information while others will provide more in depth review. Additionally some skip from one product feature to another with only a brief description while others elaborate on certain features a lot more.

This factor has a particular importance during the overall classification of the text orientation (positive/negative). One has to judge if separate sentences refer to the same attribute/object or different. Similarly depending on the user interest one sentence can express many opinions within.

Querying formula

Depending on the person and the place where people share their opinions, statements and queries can be expressed in a different way. Some users tend to use keywords or short sentences while others provide full text. For example: iPhone advantages or what are the advantages of the iPhone?

Type of vocabulary used

Opinions can be expressed in many different ways depending on the manner the vocabulary is used. One can use words that directly refer to the sentence subject (i.e. I think this product is bad!) or use affect vocabulary that contains more emotions and can be much harder to recognize (i.e. I love the way this switch works! or I was stunned to see all those special effects).

Additionally, with regard to vocabulary and grammar, we can variate opinions between stated explicitly (as in using simple language constructs and clear statements) or implicitly (i.e. This phone fits right into my pocket).

3.3.2 Document level sentiment analysis

Document opinion analysis is about classifying the overall sentiments expressed by the authors in the entire document text. The task is determine whether the document is positive, negative or neutral about a certain object. When applied to a single type of text those techniques typically have a range of accuracy from 70% to 80% depending on amount of human input and type of text (Liu, 2008b). In the rest of the following section we present a number of most representative solutions in the area, some present an novel algorithms (Turney, 2002), while other try to implement approaches proven in other domains.

The work done by Turney (Turney, 2002) on review classification presents an approach based on distance measure of adjectives found in text from preselected words with known polarity (excellent and poor). The author presents a three step algorithm which processes documents without human supervision. First, the adjectives are extracted along with a word that provides contextual information. Words to extract are identified by applying predefined patterns (for instance: acjective-noun or adverb-noun etc.).Next, the semantic orientation is measured. This is done by measuring the distance from words of known polarity. The mutual dependence between two words is found by analysis of hit count with AltaVista search engine for documents that contain two words in a certain proximity of each other. At the end the algorithm counts the average semantic orientation for all word pairs and classifies a review as recommended or not.

In contrast, Pang et al. (Pang et al., 2002) present an work based on classic topic classification techniques. The proposed approach aims to test whether a selected group of machine learning algorithms can produce good result when sentiment analysis is perceived as document topic analysis with two topics: positive and negative. Authors present results for experiments with: Nave Bayes (Lewis, 1998), Maximum Entropy (Berger et al., 1996) and Support Vector Machine algorithms (Joachims, 1998). Interestingly the performed tests have shown results comparable to other solutions ranging from 71% to 85% depending on the method and test data sets.

3.3.3 Sentence level sentiment analysis

The sentence level opinion mining is an action that can be associated with two tasks. Initial work is to identify whether the sentence is subjective (opinionated) or objective. The second task is to classify a subjective sentence and determine if it is positive, negative or neutral. Similarly as with document level most techniques use forms of machine learning.

Riloff and Wiebe (Riloff and Wiebe, 2003) put most of impact in their work on the task of subjective sentences identification. They propose a method that at bootstrap uses a high precision (and low recall) classifiers to extract a number of subjective sentences. During this phase sentences are labelled by two classifiers: first for high confidence subjective sentences, second for high confidence objective sentences. The sentences that are not clearly classified into any category are left unlabelled and omitted at this stage. Both of the classifiers are based on preset list of words that indicate sentence subjectivity. The subjective classifier looks for the presence of words from the list, while the objective classifier tries to locate sentences without those words. According to the results presented by authors their classifiers achieve around 90

In the second step, the gathered data is used a for training an extraction algorithm that generates patterns for subjective sentences. The patterns are used to extract more sentences in the same text. The presented method has such split in order to increase recall after the initial bootstrap phase (however, as expected, author report the precision to fall between 70-80%).

During the learning phase the algorithm uses a predefined set of syntactic templates that are matched against the subjective sentences (see Fig. 3.9). After the entire training set is processed the extracted patterns are ranked based on their occurrence frequency and according to some preset conditions only the best patterns are selected for next iteration of base text analysis.

SYNTACTIC FORM	EXAMPLE PATTERN
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
<subj> active-verb dobj	<subj> dealt blow
<subj> verb infinitive	<subj> appear to be
<subj> aux noun	<subj> has position
active-verb <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
verb infinitive <dobj>	get to know <dobj>
noun aux <dobj>	fact is <dobj>
noun prep <np>	opinion on <np>
active-verb prep <np>	agrees with <np>
passive-verb prep <np>	was worried about <np>
infinitive prep <np>	to resort to <np>

Figure 3.9: Extraction patterns learning phase syntactic templates and corresponding extraction patterns (Riloff and Wiebe, 2003)

Although the presented work does achieve quite good results it only concerns one task put ahead for sentence sentiment analysis. In opposition to it, work done by Yu and Hatzivassiloglou (Yu and Hatzivassiloglou, 2003) discusses both sentence classification (subjective/objective) and orientation (positive/negative/neutral). For the the first step of sentence classification, authors present test results for three different algorithms: sentence similarity detection, nave Bayens classification and Multiple nave Bayens classification. In the second step of sentence orientation recognition authors use a technique similar to the one used by Turney (Turney, 2002) for document level sentiment analysis (see Sec. 3.3.2). The main different is that the algorithm is extended to use more then two (excellent/poor) base words to which all others are compared.

3.3.4 Feature based sentiment analysis

The feature level of sentiment analysis is the most detailed study of the text. Being most useful it is also the hardest to perform. The goal is to not only determine text subjectivity and polarity but also what in particular the text author liked or disliked about the object. Typical this objective is split into the following tasks:

- extract object features that are commented
- determine orientation of opinions (positive/negative/neutral)
- group feature synonyms and produce a summary (see Fig. 3.10)

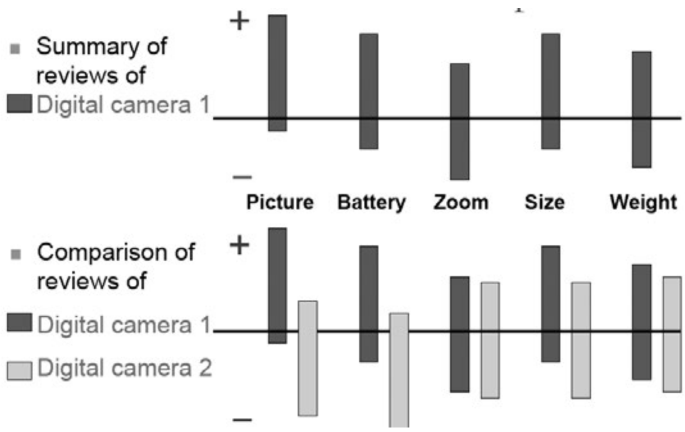


Figure 3.10: Sample output of the feature based sentiment analysis [3]

Similarly as with both previously described level (see Sec. 3.3.2 and Sec. 3.3.3) often the feature sentiment analysis experiments are conducted only for a single selected text type. Sometimes authors go even further and present methods for specific text format, for instance reviews where positive and negative features are explicitly separated is different areas. Such approach is presented by Hu and Liu in their work about customer reviews analysis (Hu and Liu, 2004). In their research authors present opinion mining based on feature frequency. Only the most frequent features recognized by precessing many review are taken into consideration during summary generation.

3.3.5 Opinion mining appliances

Aside of research on improving algorithms researchers gave a lot of attention to appliance of their opinion mining technologies in practice. Below are presented selected and most prominent use cases for opinion mining techniques. Some have been already related in various research experiments previously mentioned while others still remain a goal to achieve in the future.

Additionally, in the context of those uses cases, their potential business benefits did not go unnoticed for the industry. Therefore, not surprisingly, the interest in the topic in recent year is growing very rapidly among both: small specialized companies (e.g. Nielsen (Nielsen, 2012), Biz360 (Biz360, 2012), Cymfony (Cymfony, 2012); and large corporations (e.g. IBM TAKMI system (IBM, 2012)).

Product benchmarking and market intelligence

The key to selling a product is responding to customers demands in proper time and in the right location. Many companies spend huge money on market analysis and hire external specialized consulting companies. The opinion mining techniques could aid this effort and potentially minimize costs. Market analysis done by specialized companies is needed to take certain amounts of time and effort, while in many cases getting fast access to accurate market data can be a key factor. The right opinion mining tools could create a business advantage for a company to get ahead of its competitors and swiftly react to customer needs.

Additionally opinion mining opens new frontiers. With the immense amounts of community created data on the Internet its analysis becomes impossible or at least very difficult and expensive without some automatic methods. This domain is huge and the amount of appliances possible is vast.

In response to those challenges and needs researchers in the area of opinion mining have investigated product review analysis in a number of contexts. Liu et al. in a number of works (Ding et al., 2008; Liu et al., 2005) investigated recognition of features discussed in a product review followed by sentiment detection for reviews originating from Amazon.com. Pak and Paroubek (Pak and Paroubek, 2010) discuss similar problems and present promising results but with relation to product benchmarking based on smaller excerpts of text originating from Twitter. Based on those works, both small and long text sentiment analysis as well as formal and informal text sentiment analysis experiments have shown that contemporary algorithms can successfully summarize product reviews if properly adjusted to the analysed domain. Furthermore, moving from products to assessment of market situation and trends, Das and Chen (Das and Chen, 2007) proposed a solution for analysing sentiments from stock message boards of Yahoo!. The results of their evaluation go in line with contemporary work (e.g. of Antweiler and Frank (Antweiler and Frank, 2004)) and show that sentiment analysis of such resources can help to determine the market activity for prediction of increases in volatility rather than particular stock movements.

Advertisement placement

Advertisements over the Internet are best to be placed in places where they can reach the biggest group of potential customers. For instance it is best to advertise selected specialized computer equipment on tech forums, while entire desktop computer sets will find better audience among more common Internet users. Therefore, often topic-based mining techniques are used. Nevertheless sometimes this can be insufficient. For instance one could imagine a situation where a tech review website releases a negative review of a product and the topic mining techniques select to display the advertisement of this product next to the review (because topic matches). In such case opinion mining would help to analyze the polarity of the article and not display the ad.

Additionally detection of text polarity and semantics with relation to advertisement topic can help to detect whether content of the website and commercial message contextually fit to each other in other not to bring harm to company reputation or brand popularity. For instance it would be very bad to display a commercial of airlines next to a news post about an airplane crash.

In their analysis of such opinion mining solutions for ad placement in blogs, Fan and Chang (Fan and Chang, 2010) show an improvement can be reached in comparison to typically used content based solutions such as Google AdSense (Google, 2012).

Individual needs

The opinion mining system could be potentially used by casual Internet users. The aforementioned feature level analysis (see Sec. 3.3.4) can be a very good way (if accurate) to provide a summarized view of posts for community review sites. Research on inclusion of opinion mining into end user experience has been evaluated for movie reviews or product reviews on Amazon book store by Hu and Liu (Hu and Liu, 2004) as well as number of other websites by (Hariharan et al., 2010). Those works show that such enhancements can greatly improve user experience thus being beneficial for both content consumers and producers.

Furthermore, in relation to this user centric opinion mining use cases, Qiu el at. (Qiu et al., 2010) present an alternative approach with regard to the aforementioned ad placement use case. In particular, they propose to mine sentiments of the content consumer rather than the content provider for selecting ads. In the proposed solution, user activity is analysed and based on consumer sentiments the advertisements of products that address consumer dissatisfaction are proposed, while advertisements of products disliked are hidden.

Opinion spam detection

The opinion spam is a direct result of the user generated content popularity. The opinions given by the users about various products and services have gained huge commercial value over the recent years. The modern Internet is being ruthlessly used just like other media as a battle front for clients in between companies and corporations. Therefore, not surprisingly, the systems that enable to post opinions are often abused. Fake or misinforming comments are posted to mislead the potential client into buying or not buying a product. This can be done both automatically but also by humans. In small scale opinion content is easy to moderate, however on big and popular forums, message boards or even internet shops this can be a very hard task. Systems that detect bogus product comments could improve the credibility of any community portal thus increasing the potential revenue. Nevertheless this domain is still not discussed that much in the open.

One reason is that misleading clients and subtly deceiving them to buy a particular product instead of others is a target of many profit oriented company on the market. Therefore, it is in direct interest of such companies to use any legal means to criticize all competitive products, and make this criticism look as credible as possible (even if it is not). On the other hand, it would be extremely useful to be able to eliminate all false comment about own products (and have this ability exclusive).

Secondly, it has to be noted that the problem of detecting opinion spam is even harder task than opinion mining itself. One has not only to detect sentences with opinions or types of opinions but also judge which opinion is correct and which is deliberately formed false. In some situations this can be impossible to determine for human not to even mention any machine AI like techniques.

One of the first solutions in this area has been presented by Jintal and Liu (Jindal and Liu, 2008). In their work, Jintal and Liu propose classification of opinion spam types and evaluate a number of machine based learning techniques to detect those different spam types. In relation to opinion mining algorithms as described before, the motivation of this work and similar attempts of other researchers is to improve the credibility of former solutions by eliminating opinions that might bias the metrics for opinion summarization (Jindal and Liu, 2008; Li et al., 2011).

3.4 Semantic Web for knowledge formalization and interoperability

The Semantic Web is a vision of next generation World Wide Web where all information is machine processable and thus computer agents may interpret available content and its relationships to aid humans in accessing, browsing and searching information. The term has been coined by Tim Berners-Lee (Berners-Lee et al., 2001) who also outlined the elements of Semantic Web architecture, enumerated the core technologies that would enable it (Berners-Lee, 2000) and drew the roadmap on how the Semantic Web research should progress (Berners-Lee, 1998).

Back then, this vision of Semantic Web has been formulated as a response to an observation on state of the World Wide Web where most information is stored only in a human readable form. Such data, encoded with HTML, does not provide much information about the meaning or relationships (i.e. semantics) of the web resources. Additionally, as the World Wide Web is an international and distributed information silo created by huge amount of people, the languages and ways in which content is described are very diverse and hard to analyse by machines. According to Semantic Web this brings a need for agreement on properly representing and characterizing those descriptions.

Therefore, the core research effort in the area is related to modelling descriptive information about web resources (i.e. web ontologies) as well as ways of publishing, consuming and storing this data. According to the W3C definition by Herman (Herman, 2004), the Semantic Web is essentially a "[..] metadata based infrastructure for reasoning on the Web [...]". The improvement of the current Web is achieved via computer agents reasoning about the web resources though the analysis of their metadata and more importantly relationships between the metadata concepts. However, as Herman states (Herman, 2004), the Semantic Web is not about giving birth to artificial intelligence for the Web which may be in a distant future a layer much above it.

With relation to this thesis, the Semantic Web technologies are being used as a tool to introduce interoperability for Idea Management Systems and model as well as formalize the highly interconnected information that is being stored. For this reason, to introduce the reader to the concepts used further in the thesis, an overview of Semantic Web technologies is presented in the following sections.

Firstly, the most important technologies of the Semantic Web from the point of view of this thesis are discussed in section 3.4.1. Later, the main research areas that try to cope with problems of implementing Semantic Web in practice are outlined (see Sec. 3.4.2). Furthermore, since this thesis is an attempt to utilize Semantic Web methodologies in practice, we provide an overview of similar activities for a number of domains and point out how and why Semantic Web was proposed for them (see Sec. 3.4.3).

3.4.1 Overview of Core Semantic Web Technologies

The originally proposed stack of technologies by Tim Berenes-Lee (Berners-Lee, 2000) as well as its later refinements (see Fig. 3.11) contain a number layers that make the final vision possible. However, this section outlines only the key elements of the stack that are utilized in the thesis for formalization of Idea Management metadata and further use of this metadata for idea analysis and assessment beyond the borders of a single system.

RDF - Model for metadata descriptions

The Resource Description Framework (RDF) (Manola and Miller, 2004) provides a data representation model and syntax for describing resources on the web. RDF works as a level above XML or some other data serialization (e.g. JSON). The mark-up language used for serialization provides a universal format of encoding data, while within the boundaries of this language RDF imposes use of certain structures and constructs for modelling information and expressing its relationships.

The main principal of RDF model is the use of triples - statements consisting of three parts: *subject*, *object* and *predicate* (sometimes referred to as *property*). The *subject* defines resource that a given statement is about. In RDF all resources are identified by URI address - unique identifiers similar to URLs used on the web. Every resource can be described by

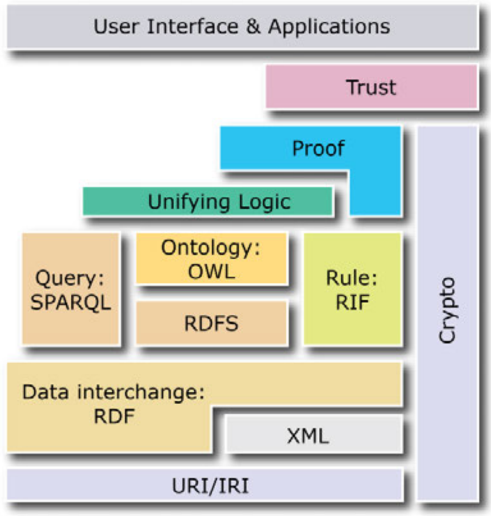


Figure 3.11: Semantic Web layer cake (Berners-Lee, 2009)

a number of *predicates* which state some kind of characteristics of the resource. Finally, the *object* part of an RDF sentence expresses the value of the property identified with the predicate. This value can be either a literal (e.g. a number or a string) or a pointer to another resource. As a result of using this model, RDF information is perceived as a graph, not a tree like in XML. The *subjects* and *objects* of RDF statements are graph nodes while predicates are edges (see Fig. 3.12). One of the benefits of such approach is the ease to extend the graph and integrate data across systems by referring to common URIs of resources.

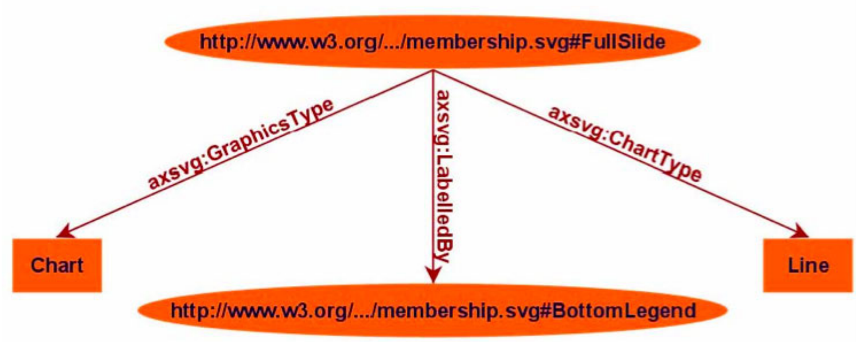


Figure 3.12: Example of a RDF statement modelled as a graph (Herman, 2004)

RDFS and OWL - languages for expressing ontologies

The RDF Schema (RDFS) (Brickley and Guha, 2004) is used for defining ontologies, i.e. ”specifications of a conceptualization” (Gruber, 1993). In practice of Semantic Web, the ontologies defined with RDFS describe particular domains or selected topics via a set of classes and their properties as well as relationships. The previously introduced RDF takes advantage of those RDFS definitions to describe particular data of a certain domain.

In this context, the Ontology Web Language (OWL) (Manola and Miller, 2004) extends the notion of RDFS and provides new constructs that allow to express the particularities of class relationships in more detail. Furthermore, OWL enables to define more complex relationships between concepts to restrict the interpretation of the knowledge base. In comparison to RDFS, this is achieved by a number of new features such as logical expressions, local properties, possibility to define certain values as required or optional and limit their range. Those additional description capabilities are defined in OWL as constructs (union, intersection etc.) and axioms (subclass, equivalent class, definition of symmetry etc.).

OWL consists of three sub-languages: OWL Lite, OWL DL and OWL Full. Each of the following encapsulates the previous one and the most important differences that the next level brings are based on level of restrictions in expressing an ontology. The OWL Lite provides classification hierarchies such as capability to define sub-classes or sub-properties and has a number of basic constraints. Next, the OWL DL adds more constructs and axioms allowing to express more (e.g. enumerations) but still giving computational completeness (i.e. all conclusions are guaranteed to be computable by a reasoner). Furthermore, OWL DL has some restrictions about the use of earlier introduced constructs (e.g. classes cannot be used as instances). Finally, the OWL Full language gives the maximum level of expressiveness by no restrictions on use of vocabulary as long as it is legal according to RDF rules. As a result, OWL Full at the cost of expressiveness does not deliver computational guarantees as the two previous versions.

SPARQL and others - querying the semantic web

Apart of defining metadata modelling recommendations the vision of Semantic Web also delivers a method for retrieval of data. The SPARQL (Prud'hommeaux and Seaborne, 2008) is a W3C recommendation for the ontology query language for the Semantic Web. SPARQL is built similarly as the query language for relational databases - Structured Query Language (SQL). For instance, in order to retrieve any RDF data SELECT query is used, to add new information INSERT query can is defined etc. (see Fig. 3.13). However, while query construction follows similar methodologies as in regular SQL, there are also some inevitable differences resulting from the underlying data model. For example, the WHERE keyword in the SELECT sentence requires stating the patterns of matching RDF triples aside of the conditions for variables like in SQL. Another notable addition is the definition of PREFIX section at the start of the query, where all referenced ontologies are declared in order to abbreviate URIs of properties and classes used in the query.

3.4.2 Semantic Web research areas

After years of research, the aforementioned core technologies are fairly well established and provided as W3C recommendations for the World Wide Web use. However, as researchers and practitioners of Semantic Web started to built solutions on top of those technologies it became apparent that many new problems for web data processing emerge. For this reason,

```
PREFIX  bibtexrdf: <http://www.marcont.org/bibtexrdf >
SELECT  ?title
WHERE   {
             ?resource bibtexrdf: hasTitle ?title .
             ?resource bibtexrdf: hasAuthor "Adam Westerski"
}
ORDER BY ?title
```

Figure 3.13: An example of a SPARQL query

a number of research areas related to Semantic Web focuses on solving those issues related to deployment of the vision outlined by Tim Berners-Lee or even adjusting the outdated details of original concepts to the reality of the evolving Web. The following are some of those key problem areas accompanied with references to examples of solutions and more detailed state of the art studies:

- semantic data management - the application of RDF model for the Web brings new challenges in terms of storage of this information in databases, therefore a number of solutions are the topic of investigation (Erling and Mikhailov, 2009; Broekstra et al., 2002). Furthermore, the bigger complexity of metadata in comparison to the current Web, brings forth new problems of scalability for efficient data processing, indexing and querying (Cai and Frank, 2004; Wilkinson et al., 2003; Delbru et al., 2012; Bishop et al., 2011).

- ontology creation and deployment - ontologies are the core of Semantic Web, yet there are many problems related to defining: how ontologies should be modelled and engineered (Gómez-Pérez, 1999), how ontologies should be maintained and updated (Plessers et al., 2007), how to mediate between different ontologies (Tang et al., 2006) and detect similar ones (Jean-Mary et al., 2009).

- annotation - to make the Semantic Web real aside of ontology creation there has to be a process for applying the metadata to web content. Therefore, there has been a number of works that deal with the problem of migrating from the current Web to the Semantic Web via content annotation (Uren et al., 2006). Researchers try to find solutions to encourage regular web users to start using the proposed ontologies for describing their own data (Siorpaes and Hepp, 2008) or develop solutions to obtain metadata in an automatic (Dill et al., 2003) or semi-automatic (Handschuh et al., 2003) manner from unstructured or poorly structure content (e.g. via various web mining methods (Stumme et al., 2006) like HTML scraping (Fernández-Villamor et al., 2011)). Furthermore, there has been significant effort on developing ways to technically incorporate the annotations into the contemporary web content (e.g. RDFa (Adida et al., 2004)).

- reasoning - as mentioned before Semantic Web not only delivers constructs to describe resources but also relationships between the resources and even relationships between the descriptions. Therefore, there has been a lot of research (Baader et al., 2003;

Horrocks et al., 2003; Horrocks, 2008; Papataxiarhis et al., 2009; Keller and Feier, 2005; Lisi and Esposito, 2009) on use of description logic and rule languages to extract information more precisely and obtain information that is not explicitly put in metadata but can be inferred using ontology definition (e.g. via inheritance or transitivity of properties and classes).

- human interaction - on top of providing the viable infrastructure to make Semantic Web technically functional, researcher have analysed method of interaction with the Semantic Web data (Hachey and Gasevic, 2012) e.g. data browsing (Berners-lee et al., 2006; Tummarello et al., 2010; Bizer et al., 2005) and information visualisation (Geroimenko and Chen, 2002).

- linking data - the vision of Semantic Web is a one where web resources connect to each other and interlink on various levels thus weaving the web (Berners-Lee and Fischetti, 1999). Therefore, the Semantic Web research also extends on methods to build this web of linked data, i.e. is best practices for publishing and connecting structured data (Bizer et al., 2009). Among others, this effort is achieved by initiatives like the Linking Open Data community project (Linking Open Data, 2011), the creation of DBpedia (Auer et al., 2007a) and linking information it and similar large datasets. The mentioned Linked Open Data project describes this as bootstrapping the Semantic Web by providing the so-called Linked Data cloud - a set of dataset that mutually reference each other. Apart of methods for producing and consuming linked data, the research is this area also refers to issues such as data quality (Hogan et al., 2010) or specific solutions for all aforementioned Semantic Web areas but with particular interest on problems of data relationships (e.g. user interfaces for linked data (Dadzie and Rowe, 2011) or linked data indexing and search (Oren et al., 2008)).

The enumerated research areas show how big and complex as well as difficult to implement in practice the Semantic Web vision is. However, at the same time, those research problems have emerged as a response to the opportunities for progress envisioned by use cases of the Semantic Web. The proposed solutions in those research areas deliver proof that starting from the simple model of RDF and a conceptualization of a domain (e.g. Idea Management Systems), one can achieve a lot in terms of system interoperability and expand on data analysis capabilities with unified methodologies for information querying.

3.4.3 Semantic Web appliance examples

This thesis describes a particular use of Semantic Web rather than progress in research on any of the aforementioned technologies or problem areas. Therefore, below as a supplement to the description of the Semantic Web state of the art some examples of applications in various domains are given. More specifically three cases of different scope of applications are presented: 1) modelling ontologies for communities that use a number of tools but within a single interaction interaction paradigm or environment, 2) a narrow case for application in a single type of system or technology, and finally 3) the uses related to particular organizations or societies.

Ontologies for community modelling

One of the breakthroughs in recent history of the Web was the introduction of Web 2.0 technologies and the surge of participatory role of the on-line web communities in web content creation. The novelty of Web 2.0 was not only technologies but design paradigms that put impact on social collaboration and social networking (O'Reilly, 2007). This shift did not go unnoticed for the Semantic Web research community and therefore there has been some engagement in modelling new information sources and taking advantage of the Web 2.0 phenomena for the adoption of Semantic Web. Among others, there have been initiatives like Friend-of-a-Friend (FOAF) (Brickley and Miller, 2010) to model social network connections or Semantically Interlinked On-line Communities (SIOC) (Breslin et al., 2005) to model the metadata of web resources created by the communities and links between identities as well as content of various social spaces (Bojars et al., 2008). Furthermore, various researchers have approached modelling user profile (Bojars and Breslin, 2007) and the presence on the web in general (Stankovic, 2007) to take advantage of this data in a variety of ways, e.g. to recommend and personalize community created information (Kapanipathi et al., 2011; Abel, 2011). Finally, aside of opening new possibilities, the Semantic Web researchers have proposed to improve the existing Web 2.0 technologies and interaction rules e.g. through semantic wikis (Krotzsch et al., 2007; Auer et al., 2006) or improving tagging activities in general via semantics to close the gaps between folksonomies (disorganized but easily created hierarchies of content) and ontologies that do better in content organization but are more difficult to create and maintain (Kim et al., 2008).

Ontologies for specific systems and technologies

Aside of general purpose applications that can work for a variety of systems and environments, ontologies and Semantic Web concepts have been proposed for a more narrow scope of application like specific IT systems (e.g. Digital Libraries (Kruk and McDaniel, 2009)) or particular technologies (e.g. Semantic Web Services (Nixon et al., 2001)). In contrast to the previous application example, in case of specific system applications like digital libraries in many cases the modelling focus is not on the person or the content creator but more on the constraints that a given system imposes and well as opportunities that it delivers (Kruk et al., 2006). In case of digital libraries the core of the problem has been improving search and browsing (Kruk et al., 2005) as well as seeking convergence with other disciplines to make libraries more attractive for regular people (e.g. though interoperability with aforementioned Social Web (Kruk et al., 2008) or by making digital libraries part of larger software ecosystem like e-Learning (Westerski et al., 2006)). In comparison, Semantic Web appliances targeted for specific technologies are often focused on very specific and usually long standing complex problems of those technologies, for example automatic Web Service composition, matching or discovery (Haller et al., 2005; Li and Horrocks, 2003).

Enterprise and organization applications

The applications of Semantic Web have also been deliberated in the context of entire organizations rather than systems or communities. The typical scenarios for those cases often start with the desire to fulfil particular business or social needs of the organisations (Baker et al., 2012). Among the discussed Semantic Web appliances the major activity areas are related to e-Government (Klischewski, 2003) and various areas of industry (Cardoso et al., 2007). In the industrial solutions, depending on the particular organization type and area of activity the goals and benefits of Semantic Web can be very different. Some of the solutions focus on private metadata repositories and aim to aid functioning of an organization based on the resources that it produces within its own ecosystem, e.g. to improve manufacturing processes (Macrini et al., 2005; Anastasiou et al., 2012) or management of the enterprise via applying Semantic Web for typical enterprise systems like ERP (Anjomshoaa et al., 2006) or CRM (Liwen and Min, 2004). However, other Semantic Web researchers approach the enterprise from a different angle, they propose to connect to the earlier mentioned Linked Data initiative and then reuse the publicly available data to improve information search inside the enterprise via linking to common concepts (Wood, 2010). With regard to appliances in public institutions and government similar data linking attempts can be observed. In particular, there is a lot of impact on producing open data (Shadbolt et al., 2012; Hendler et al., 2012) that could be later reused e.g. for improvement of information services for citizens (Gómez-Pérez et al., 2005; Sacco, 2006) or better transparency of public institutions (Jarvenpaa et al., 2006).

Following those and the earlier presented case studies, it can be seen that Semantic Web has found a wide area of appliances and can deliver a variety of benefits. As it is shown later throughout this thesis, the use of Semantic Web for Idea Management Systems faces problems and has the opportunities common with a number of the presented uses cases. Like the SIOC Project the proposal of the thesis relates to on-line communities and content generated by them, yet like the enterprise ontologies the presented work faces a number of limitations and turns to a narrow and specialized group to deliver the main benefits. Additionally, like the digital library ontologies the proposal of the thesis is aimed at a particular system type therefore the design process is subject of a number of constrains. Aside of those similarities the deployment of Semantic Web for Idea Management Systems brings new experiences and observations. All those are detailed in the next chapter as the Generic Model for Idea Management Systems is introduced and modelled using the described principles and technologies of Semantic Web.

3.5 Summary and Relation to Thesis Contributions

In the Foundations chapter we have discussed state of the art in domains related to development of the thesis. The goal of this description has been to make the reader familiar with the methodologies and technologies of selected domains and present the understanding of those areas as they are used by the thesis to leverage capabilities of the contemporary Idea Management Systems. The critical review of specific elements of the state of the art related to thesis contributions is done later, individually per each contribution chapter.

In this chapter, firstly, the contribution area of the thesis has been described though analysis of **the state of the art in Idea Management Systems**. We have shown how Idea Management Systems evolved from suggestion boxes embedded in other systems into separate class of systems that support various activities related to organization of ideas. Additionally, the state of the art in Idea Management Systems research has been presented giving insight into attempts of scientists from different areas to solve problems of the systems as their use evolved in the industry.

Following the description of Idea Management Systems state of the art, this chapter has presented an overview of the state of the art in areas that are utilized by the thesis to deliver its contributions, specifically: Innovation Management, Opinion Mining and Semantic Web.

The **Innovation Management models for innovation modelling** have been described as they are the foundations for the thesis contributions in the area of modelling *idea concept* and extending the metadata of an idea beyond the plain text that is available in the contemporary Idea Management Systems. The presented overview of models developed across past years has shown a large variety of approaches and presence of many different perspectives on: how innovation can be classified, where does innovation originate from, what are the results of innovation. In Chapter 6 this knowledge is utilized to construct a model for idea characteristics and assess the value of past innovation models in terms of usage for Idea Management Systems data modelling and comparison of on-line web communities that create the content of those systems.

Next, the Foundations chapter provided an **overview of Option Mining domain** that focuses on algorithms for natural language analysis of subjective texts in order to identify the emotions expressed by the author of the text and quantify it. The achievements of this area have been applied in the thesis for the proposal of extending Idea Management model towards community opinions analysis. The overview of state of the art in Opinion Mining research has shown the presence of three main areas of activity: sentence based analysis, document base analysis and feature based analysis. Across those areas a variety of solutions have been identified ranging from keyword based analysis to pattern based algorithms, either with various domain specific enhancements possible to increase the quality of the solution. In the thesis the knowledge from this domain has been used in Chapter 5 to: (1) construct a proposal for formalization of opinions based on analysis of metadata generated by available Opinion Mining algorithms, (2) construct a keyword based opinion mining prototype to quantify opinions in Idea Management Systems and evaluate the use of obtained metadata for generation of metrics for idea assessment.

Finally, the Foundations chapter presented the **domain of Semantic Web** which focuses on solving problems of systems integration though introducing extensive use of rich metadata on the Web. The overview of this domain has shown the presence of many technologies that address publishing, storing, indexing, querying and processing metadata delivered by distributed systems. Furthermore, the analysis of Semantic Web deployments for different domains delivered insight into uses of discussed technologies in enterprise including Idea Management Systems. The thesis builds on those achievements and proposes the use of Semantic Web technologies in Idea Management Systems in order to support integration of distributed Idea Management Systems and facilitate comparison and analysis of Idea Management System data regardless of the domain or system architecture. For this reason, Semantic Web is not the contribution area of the thesis but the technological background that is used as an enabler to construct the evaluation experiments described throughout the contribution chapters. Firstly, this notion is introduced and explained in detail in Chapter 4 with Gi2MO Ontology being formalized using Semantic Web methodologies. Further this proposal is consequently supported across all of thesis contribution chapters as new idea assessment techniques are introduced and evaluated with prototypes constructed based on Semantic Web technologies.

Chapter 4

Generic Knowledge Model for Idea Management System

This chapter describes the research on a generic model to describe data of Idea Management Systems. The goal of the presented model is to deliver a single point of reference for Idea Management Systems and provide a formalization that can be applied for any platform to achieve data portability and interoperability with other enterprise and web systems.

Based on the analysis of the state of the art in Idea Management Systems, we propose a theoretical framework for the Idea Life Cycle - stages through which an idea progresses as it changes over time in response to various actions of actors involved in the idea management processes. We point out the dependencies between the defined life cycle stages and detail the impact that those stages have on each other.

We leverage the Life Cycle model and introduce a Generic Idea and Innovation Management Ontology (Gi2MO) - a formal definition of the concepts described in the theoretical framework. In terms of model implementation, to facilitate idea dataset comparison, we propose to use of Semantic Web technologies as a gap closer between heterogeneous Idea Management software and achieving interoperability on metadata level.

The contributions presented in the chapter are:

- Idea Life Cycle proposal
- Gi2MO Ontology for Idea Management Systems

4.1 Introduction

The concept of innovation in organizations has become an important issue along with the increasing competitiveness of markets. Many companies realised that it is crucial to constantly develop their value proposition and innovate not only to attract new clients but also to avoid loosing current ones. On the other hand, reports indicate that the global financial crisis in 2007 had an impact on the innovation process within enterprises. The harsh economic conditions indeed sometimes lead to reducing investments (Kanerva and Hollanders, 2009) but more interestingly change the motivation for innovation. Apart from increasing competitiveness or customer satisfaction, companies seek to use innovation as a tool to reduce production costs (Andrew et al., 2009a). This can lead to a conclusion that implementation of innovation programmes can diversify greatly depending on the particular motivations that an organization has.

However, regardless of those motivations, similarly as with all other activities in the modern organizations, innovation is aimed to be a repeatable and formalized process that can be measured (Andrew et al., 2009b). To achieve best results, innovation should be managed and developed as part of an organizations culture with appropriate procedures thus creating: an innovation management process (Brem and Voigt, 2009; Cormican and O'Sullivan, 2004). Ever since the brink of the information technology era it has been common to use software tools to aid the selected aspects of the management processes in organizations. In this matter, innovation management is no different (Adamides and Karacapilidis, 2006; Conn et al., 2009). However, the systems that are used diversify greatly depending on the organization size, scope, area of activity, enterprise procedures and utilized support systems.

Furthermore, when an organization decides to invest in open innovation with Idea Management Systems than another factor is involved: communities gathered around and inside the enterprise. The systems that gather innovation from distributed online communities are being adjusted and tailored to stimulate the characteristics of a particular community in the best possible way. Together with the other aforementioned causes, this leads to a fragmentation in the Idea Management Systems area. Therefore, in this chapter, we propose a single model that would generalize the entire concept of such systems and allow to reference it during further work.

In the following sections, firstly we recap the earlier introduced state of the art and point out how contributions of this chapter build on top of the related work (see Sec. 4.2). Further, in Section 4.3, we present the 'Idea Life Cycle' - a set of consequent stages in the idea management process driven by interactions of different actors and communities with the system and the changes in data. We describe the actions and classify methodologies per each of the stages in reference to current state of the industry as well as research. Building on on top of that framework, we propose how the quality of the entire process can be improved through gathering feedback on each stage of the life cycle (see Sec. 4.3.6). Finally, we describe the process of formalizing of the Idea Life Cycle with an ontology (see Sec. 4.4). To assess the quality and value of our research, we present an evaluation of the ontology though a coverage study (see Sec. 4.5) and summarize the results obtained (see Sec. 4.6).

4.2 Related Work

The topic of domain driven ontologies and their design has been investigated in numerous works for different areas and with different scope in mind e.g. BBC Music (Raimond et al., 2007) Ontology, GoodRelations ontology (Hepp, 2008) for e-commerce, SIOC ontology (Breslin et al., 2005) and many others. In our work we tried to learn from the best practices from those attempts and employ them in our research. Furthermore, based on observations from a number of case studies (Bontas et al., 2005) and ontology engineering research (Pinto and Martins, 2000; Uschold et al., 1998; Gruninger, 1996), we reference or reuse concepts from a number of previously developed ontologies that intersect with our domain.

However, since Idea Management Systems are a rising technology there has not been much research done in terms of application of metadata and assets interlinking. To our knowledge Riedl et al. (Riedl et al., 2009b) are the only ones who present a similar attempt to ours and aim to provide technical means for describing and integrating data of IMS. Their Idea Ontology applies a different approach where less impact is put on interlinking (i.e. relationships and dependencies between concepts) and more on the sole goal of integration of idea repositories. Furthermore, we present a different methodology for constructing our ontology and put more impact on theoretical definition of Idea Life Cycle as the foundation for our proposal. As a result, concepts such as the idea metadata changes in time and the role of various actors in the Idea Management process influence our knowledge modelling decisions in a significantly bigger degree than in case of Idea Ontology.

Aside of ontologies aimed strictly for Idea Management Systems, there have been attempts to construct models for concepts related to innovation processes in general. Bullinger (Bullinger, 2008) proposes the concept of OntoGate for idea assessment though usage of ontologies that model domain specific knowledge (e.g. product structure, market description, organization strategy etc.). In comparison to our contribution described in this chapter, the proposal of Bullinger compliments Gi2MO Ontology as a tool that can be connected with existing Idea Management System metadata to provide a new solution for idea assessment. More similary to our Gi2MO proposal, Stankovic et al. (Stankovic, 2010) as well as Lorenzo et al. (Lorenzo et al., 2011), propose ontologies related to innovation modelling that cover serialization of information system metadata for integration. However, Stankovic mainly targets Idea Marketplaces and as a result only focuses on modelling aspects related to challenges and competitions that are central for this group of systems. On the hand, Lorenzo proposes an ontology for brainstorming systems that covers a large number of concepts related to idea modelling and communities. Nevertheless, unlike our contribution, Lorenzo focuses on specific of modelling community collaborative processes and pays less attention to the management, assessment and metric measurement aspects that are central for Idea Management Systems.

4.3 Idea Life Cycle

Based on the observations of the state of the art presented in the previous chapter as well as highlighted case studies we propose to interpret Idea Management System as a software aided approach to manage innovation on its **stages of evolution**:

- Idea Generation
- Idea Improvement
- Idea Selection
- Idea Implementation
- Idea Deployment

Ideally, input and output of all of those stages should be closed in a cycle to reuse the data for improving the quality of future ideas and idea management procedures (see Sec. 4.3.6). Each of the stages can involve participation of many actors coming often from different communities, either inside the organization or from an external environment (see Fig. 4.1).

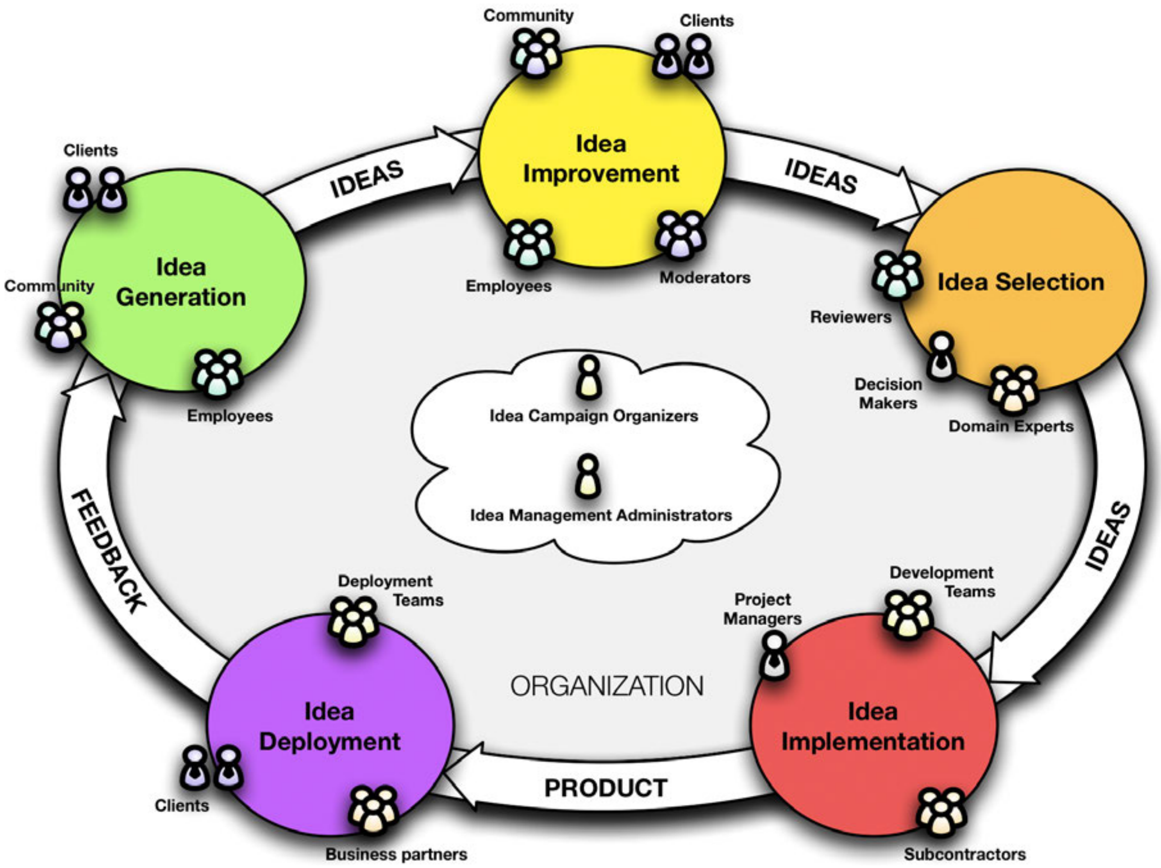


Figure 4.1: Idea Life Cycle and Communities

Idea Generation is about reaching out to the community or a particular group of people and ***extracting the ideas*** from them.

Idea Improvement is about enabling people to ***collaborate with each other to improve the ideas gathered***.

Idea Selection aims to harness the high volume of data submitted by the crowds and ***choose the best ideas***.

Idea Implementation starts at a point when an idea gets a positive review and is accepted to be put into production. The goal of this stage is to ***transform ideas into products or services***.

Idea Deployment is the process that ***tracks the successfulness of ideas*** after they have been delivered to the target audience as products.

In the next subsections, we shall highlight the practices and activities characteristic for each of the phases. Furthermore, while doing so, we also detail the techniques in each stage that push the data changes in ideas across the life cycle (see Fig. 4.2).

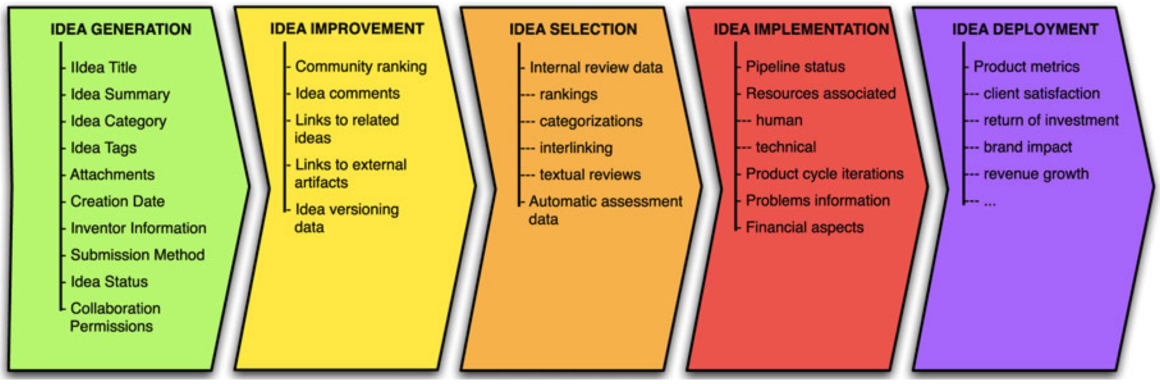


Figure 4.2: Idea Life Cycle Data Evolution

4.3.1 Idea Generation

The input for this phase is gathered from the people that interact with a computer system or telecommunication infrastructure. The end product of this phase is a *semi-formalized idea*. This goal can be achieved in a number of ways depending on the idea capture method:

- push methods (user is explicitly asked for ideas on a given topic)
- pull methods (user ideas are extracted or inferred from some content)

Among the **push methods** the most popular solution is simple *web input form* where user fills out the data corresponding to the idea formalization such as: title, summary etc. (e.g. used in products of Salesforce (SalesForce, 2009), BrightIdea (WebStorm, 2009) and most of other ones on the market). However, some of the other possibilities for push idea input are: *a guided process* (e.g. indirect questions that lead to formalization of idea in Ingenuity Bank (IngenuityBank, 2009)) or *web services* that allow connecting with various input devices (e.g. mobile phone (IngenuityBank, 2009)). Additionally, systems based on the push methods can be constructed to support either a single user idea generation process or a collaborative idea generation process (e.g. through brainstorming (IngenuityBank, 2009; Idearium, 2009)).

On the other hand, the **pull methods** are about extracting ideas either from textual content (e.g. social media) or based on verbal contacts with the client. The key element of this method is that information analysed is not submitted by the user with the intention of idea generation. The techniques used, aim to separate ideas from unrelated opinions and unwanted content. Among those techniques, we can distinguish: *data mining* (Cabena et al., 1997) in conjunction *opinion mining* (Liu, 2008c) for textual content located outside organization systems or integration with other systems and implementing data flows for content within the organizations systems e.g. Customer Relationship Management integration (e.g. implemented by Salesforce (SalesForce, 2009)).

Apart of deciding upon the usage of either push and pull input techniques the items that especially matter at the idea generation stage are: 1) encouraging the inventors to actually approach the system and contribute their ideas or opinions; 2) to ensure the good quality of the submitted content. The support for such activities is being quite often built into Idea Management Systems as part of the preparation process for idea generation competitions (e.g. as a reward system for best innovators).

Based on the analysis of the aforementioned literature and systems that employ either the push or pull methods, we propose to define the outcome of the idea generation phase as an interlinked set of data that can be broken down into following:

- Idea Title (brief summary of idea)
- Idea Description (detailed textual description of an idea)
- Idea Category (assignment to some fixed predefined categories)
- Idea Tags (categorization with custom keywords)
- Attachments (rich media like pictures, videos etc.)
- Creation Date (the date when idea was submitted to the system)
- Inventor Information (idea is interlinked with user profile)
- Submission Method (optional depending on system capabilities)
- Idea Status (indicates the position of idea in the company internal process pipeline)

- Collaboration Permissions (some default preset depending on the system and selected scenario for idea collaboration)

4.3.2 Idea Improvement

Once the ideas are submitted it is a good practice to immediately share them with public and see what other participants of the idea competition think. This way, before ideas are assessed by dedicated staff from the organization, data is incubated in the community for a period of time, improved and confronted with mass opinion. Idea Improvement is about community interaction and collaboration. Therefore, this stage includes:

- all the post processing of ideas done by the community after the original content is submitted
- the moderation practices needed to organize that content and support the community

The post processing techniques can be directed towards modification of an existing idea content or extending it. In case of modifications the same input techniques as used during the idea generation are valid, however in addition it is needed to set the rules for modifications and track changes. The modification policies require inclusion of profiling, authentication and privilege lists inside the Idea Management System. Once this is available a direct extension is traceability of changes which can be resolved though *idea versioning* (e.g. in Accept Ideas (AcceptIdeas, 2009)) handled similar to Source Code Management (SCM) such as SVN (Subversion, 2009) or CVS (CVS, 2009). Sometimes both profiling and versioning challenges are resolved with existing technologies e.g. through implementing wiki-like input (AcceptIdeas, 2009).

The support for modifying ideas by community members is useful, however it requires a lot of dedication and effort from an individual. Therefore, the techniques that allow users to make small additions to extend ideas are equally important: discussion support, community ranking methods, and idea interlinking.

The discussions between idea competition participants are most often facilitated with the model taken directly from Web 2.0 social spaces such as forums, blogs etc. In practice, this is implemented as comments for ideas (e.g. in IdeaScale (IdeaScale, 2009)) but also sometimes extends to additional forums, dedicated blogs or even external popular community sites (such as Facebook or Twitter) integrated with the Idea Management System e.g. in Salesforce Ideas deployments from Dell (Dell, 2009) or Starbucks (Starbucks, 2009)).

Idea comments and discussions are a natural way to improve ideas and express opinions, however this type of user input is not quantified and hard to analyse when it grows in size. Therefore, Idea Management Systems often introduce additional tools for quantified community based idea ranking:

- *simple up/down ranking* (often similar to Digg e.g. Salesforce Ideas (SalesForce, 2009))
- *buying and selling idea shares* (in systems that implement prediction market mechanisms (Spann and Skiera, 2003) e.g. Nesco Idea Exchange (Nesco, 2009) or IDEM (Bothos et al., 2008))

- *idea games* (idea competition participants compete according to a set of rules e.g. ref-Quest (Baalsrud Hauge et al., 2008) or Idealyst (Toubia, 2006))
- *hybrid ranking systems* (e.g. up/down ranking combined with a limited pool of votes that is refiled based on some rules, e.g. Newsfutures Idea Pageant (Newsfutures, 2009))

The above ranking methods are one of the attempts to move some of the problems of the idea assessment phase (see Sec. 4.3.3) into the community improvement stage. However, it is not the only technique practised for community supported assessment. In addition, quite often Idea Management Systems deliver simple support for idea interlinking. In most systems this is implemented as duplicate detection that results in a decrease of information volume during assessment phase. However, it could also be possible to extend this concept up to similarity comparison (e.g. feature similarity based on research done in opinion mining (Hu and Liu, 2004)), time-line dependencies (partially implemented in reference to idea requirements in Accept Ideas (AcceptIdeas, 2009)), or idea evolution dependencies (done in many systems in a simple form of idea status tracking).

Similarly as in the idea generation phase, all types of activities performed during the idea improvement phase result in additional data added to the idea description:

- Community ranking data
- Idea comments
- Links to related ideas
- Links to artifacts outside the Idea Management Systems (e.g. social collaborative portals, external media linked by users etc.)
- Idea versioning data (full versioning information or partial e.g. modification date)

4.3.3 Idea Selection

The goal of this stage is to select the best ideas and propose them for implementation. This can be achieved with data browsing and search techniques. However, the task is not straightforward and gets complex due to the characteristics of data from previous stages (Jouret, 2009; Turrell, 2008): high volume, big redundancy of data, large amount of trivial ideas. The three most important techniques to cope with those problems are:

- idea assessment (reviews run periodically and in parallel to the selection process)
- machine aided data preprocessing (computational heavy tasks such as statistics, pattern detection etc.)
- filtering and clustering (textual and graphical methods applied during selection to enhance idea browsing and search)

The idea assessment done by internal organization reviewers is supposed to enrich the community created idea description with alignment to organization strategy, goals and current needs. To gather the input from reviewers similar tools as during the Idea Improvement stage can be used:

- ranking tools
- categorization
- interlinking
- textual reviews

In contrast to community assessment the reviews done internally can be much more complex and demanding, e.g. ranking can be split into many themed categories (e.g. in Accept Ideas (AcceptIdeas, 2009)). Furthermore, the assessment can be potentially customized through profiling of reviewers who can provide better assessment if it is aligned to their area of expertise e.g. market analysis, strategic planning, product cycle placement,financial analysis (e.g. cost vs. return of investment) etc.

The input given by reviewers during this stage and by community earlier can be processed with machine algorithms to extract additional value and calculate metrics. The algorithms can be oriented towards mining connections in structured data (Cabena et al., 1997) (e.g. measure average similarity ratio based on different categorizations or review metrics) or to extract valuable information from textual comments and reviews with natural language processing technologies (NLP) (e.g. measure opinion polarity for ideas with opinion mining technologies (Pang et al., 2002)). Furthermore, if the Idea Management System has a well developed personalization module then connections between users and submitted content can be tracked and reasoned upon (e.g. detecting patterns in community behaviour to measure individual users reputation and expertise).

In the end, both algorithms aided assessment and human assessment ultimately produce a number of characteristics of an idea. In the selection process all this data is utilized to deliver different view points for the person responsible to choose the final ideas (or best candidates) for implementation. The idea database is explored by defining criteria aligned with idea characteristics for idea filtering, ordering and search. The techniques can be either textual (tables and lists) or graphical (diagrams, charts, other innovative graphical presentation or navigation techniques).

On the idea selection stage ideas are enriched with the following data:

- internal review data
- automatic assessment data
- idea ranking and selection data

4.3.4 Idea Implementation

The idea implementation phase starts when selected ideas are approved for implementation. The goal is to transform ideas into products, services or perhaps just actions. At this stage, Idea Management Systems come very close to project management tools, product life cycle management etc. In those areas, quite often organizations already have dedicated and specialized systems that support management and development activities. Therefore, Idea Management Systems take a number of approaches ranging from complex to very limited:

- full embedded support for project management (allocation of resources, definition of tasks and requirements, reporting support etc.)
- integration with popular project management/ product life cycle tools (e.g. through open APIs)
- no development management aside of status reporting

Each of these approaches has been implemented in practice by companies that successfully deliver commercial idea management platforms. When implemented by the same vendor, the support for project management is either a module in the idea management platform (e.g. in Salesforce Ideas (SalesForce, 2009)) or a separate product with very tight integration (e.g. BrightIdea Pipeline (BrightIdea, 2009b)). On the other hand, the interfaces to popular project management software or open APIs limit the scope and complexity of idea management software to a more consistent range of tools (e.g. Accept (AcceptIdeas, 2009) or Imaginatik (IdeaCentral, 2009) solutions). This way it is easier to harness the new software by using it only for the first three stages of the idea life-cycle.

The necessity to integrate idea implementation stage to up most with the previous idea management phases is often advocated by using statistics about high research activity that resulted in low innovation adoption ratio (e.g. such statistics on innovation performance are delivered annually in European Innovation Scoreboard (Scoreboard, 2009) or as innovation reports by BCG (Andrew et al., 2009c)). Vendors that deliver fully integrated solutions tend claim that some of the innovations are lost or not properly implemented because idea information is not properly communicated between different organization environments. Nevertheless, it should be noted that the aforementioned statistics most often only stress the lack of proper innovation management processes in organizations and do not reject or favour any methods or tools to fix this process continuity. With the following section we do not take a side is this discussion, we only wish to indicate the necessity to take account of the idea implementation phase and raise full awareness of it. From the point of view of idea life cycle and idea management, the biggest value of this phase is located in the metrics and the feedback that can be taken from the implementation stage and used to improve the entire innovation process (see Sec. 4.3.6).

During the idea implementation stage ideas are enriched with the following data:

- status and progress update on idea in the implementation pipeline

- resources associated with idea implementation (technical, human etc.)
- information about iterations of the product cycle (how much effort did production take)
- information about problems encountered (e.g. what was the idea lacking)
- financial data (cost of implementing idea, cost of resources etc.)

4.3.5 Idea Deployment

After ideas are successfully implemented as products they need to be delivered to the customers. Similarly as with idea implementation we wish to stress that the biggest value of this phase for idea management is gathering data about the deployment process rather than actual management of activities that need to be done to deliver a product.

The data added to idea description is fully related to the reception of the implemented idea by clients. Later this can be translated into various innovation metrics (Andrew et al., 2009b), e.g. :

- client satisfaction
- return of investment
- brand impact
- revenue growth

4.3.6 Dependencies between idea life cycle stages

Earlier (see Sec. 4.3) we have presented an order of continuous stages in the idea life cycle process. However, it has to be noted that, in practice, the cycle for each idea should not end with the last phase described. For the best results the output of each stage should be used to improve the predecessors and the entire quality idea management methodology in the organization (see Fig. 4.3).

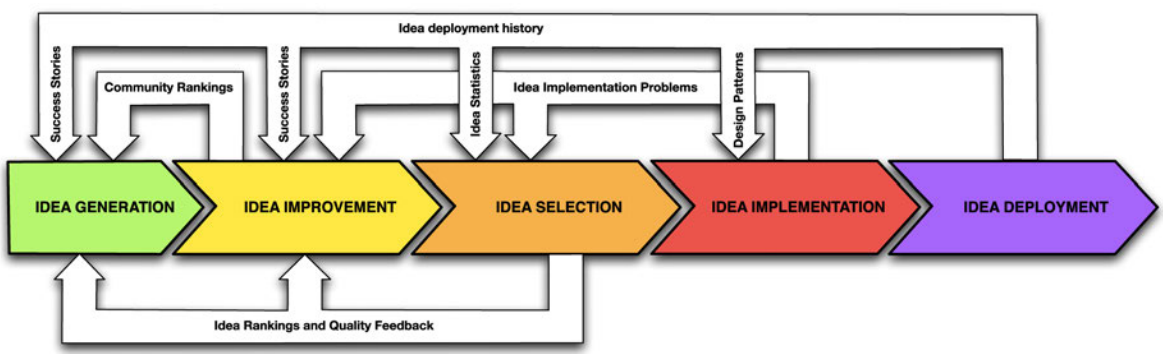


Figure 4.3: Idea Life Cycle Dependencies

Idea Improvement

The community rankings described earlier can be used for idea self-organization. This way the community's top rated ideas can be promoted and exposed stimulating creativity during the *Idea Generation* phase. In addition in Idea Management Systems based on game research (Baalsrud Hauge et al., 2008; Toubia, 2006) the community rankings can be shown to create game winning ideas.

Idea Selection

The idea ranking and assessment metadata can be easily reused in *Idea Generation* and *Idea Improvement* phases. The data can be passed to community moderators and users can be notified at generation time about some additional criteria for the ideas that the organization currently seeks. Also the metrics defined during assessment phase can be employed to provide hints in real time for idea usefulness (e.g. tag analysis- comparing user input with keywords for current idea campaign). Furthermore, the defined metrics and internal idea rankings can be used to order ideas so that the most valued ones are additionally promoted among users during the idea competition event in the *Idea Improvement* phase. Such practices help to show what is valuable for the company and give a better idea for the users on how to improve their own ideas.

Idea Implementation

During idea implementation the development team is given information provided by the inventor and has to relate it to the reality of the organization (e.g. technology process, organizational capabilities, available resources etc.). This way some potentially valuable and promising ideas are intersected with typical product or service development problems. This information can be also valuable to transfer to the *Idea Selection* phase for improving selection of ideas in the future (for instance as encountered problems and issues that reviewers should pay extra attention to).

In addition, the information can be used by community moderators during the *Idea Improvement* phase. The more the moderators are aware of desirable idea descriptions the better they can steer and direct the community to improve the current ideas. Moderators can point out and stimulate contributions from the crowd based on the feedback from implementation teams that were missing particular information or in large part found some data useless (or even making their work harder).

Idea Deployment

The idea deployment phase can potentially bring a lot of valuable data as feedback for every stage in the Idea Management cycle. In addition, it is not only important to reuse the data in real time as they come but also run statistics and detect patterns of successful and unsuccessful ideas.

For the *Idea Implementation* phase the outcomes of ideas such as product opinions or financial statistics like sales data or return of investment can help to identify problems in the implementation phase (for instance two equally promising ideas selected for implementation but due to different development team composition one got less successful; potential reason could be e.g. too big time to market, choosing bad technical solution, or even skipping some of the original idea information). In practice, this information can aid process improvement and making some strategic decisions for future improvement. However, it has to be noted that to apply such analysis the Idea Implementation process needs to be very well defined.

In the case of the *Idea Selection* phase, similar statistics as for Idea Implementation can aid greatly to choose the correct ideas and in identifying patterns for ideas that turn out to be bad in practice. Similarly, to analyse faults of the idea assessment process and improve it, it has to be very well defined and documented (e.g. the reason why a particular idea was chosen has to be clear and document).

In addition, properly prepared idea outcome data can be used as a motivator both in *Idea Generation* and *Idea Improvement* phases. The ideas that got implemented and furthermore had very good reception as services or products can be exposed as success stories. Such practices shall both encourage potential contributors to share their ideas and in addition deliver patterns that show how to describe ideas so that they become successful.

4.4 Generic Idea and Innovation Management Ontology

With the definition of the Idea Life Cycle we have classified the activities and data flows present in a variety of Idea Management Systems. However, our goal is to apply this theoretical model in practice of deployed systems to serialize their data and enable idea comparison regardless of the underlying IT system layer. In order to fulfil this goal there is a need to deliver a formal specification of the presented model. In terms of this formalization we propose defining an ontology - an explicit specification of a conceptualization (Gruber, 1993). In particular, we take the approach presented by the research in the Semantic Web area, a domain that focuses on data portability and interoperability between Web systems through the use of web ontologies.

Amongst other, one of the important reasons to develop Semantic Web (Berners-Lee, 1998) and rich metadata is to bring order to the current Web and harness the ever growing informational chaos. There have been a number of visions to achieve this at full in the Internet wide scale. Some initiatives took a global approach (e.g. Cyc (Lenat, 1995)), while others claimed that building the new Web is through starting in small domains and interconnecting those islands slowly approaching the desired state (e.g. Linking Open Data community project (Linking Open Data, 2011)). However, in addition to those movements, as Semantic Web research grew in popularity, the industrial sector started to experiment with applying the technologies in closed environments, not to solve the inconveniences of the global Web but their own local problems with information overflow. In the following section, we present how the same could be done for the domain on Idea Management Systems to achieve our final step of Idea Life Cycle formalization.

4.4.1 Use Case Study

We present two use cases that can give an image how Semantic Web annotations can aid Idea Management in practice and improve the current systems. Firstly, we show how very basic Semantic Web technologies deliver interoperability between different systems (Scenario 1). Next, we present a more sophisticated example where usage of common ontologies leads to discovering new useful data (Scenario 2). The primary goal for both examples is to expose the benefits of interlinking Idea Management Systems with other systems.

Scenario 1: Extracting idea metrics via direct links to other systems

John is a working in a medium but rapidly growing enterprise. To wisely allocate the sudden influx of money his company invests in innovation. The enterprise has a large number of products and a huge client base, so John sets up Idea Management facilities that will help to gather the feedback from the clients. Nevertheless, as John discovers, when new products are released, clients suddenly get very active and the feedback grows to incredible amounts nobody is capable to asses within reasonable time and effort costs. Furthermore, the metric generation capabilities embedded in the software are either insufficient or require a lot of effort to manually input business data for every idea to fully compare and judge client submissions. Therefore, John turns for help to emerging technologies and convinces his company management to invest in integration of systems with Semantic Web technologies. As time passes newly adapted technologies start to pay off. When a game changing product is released clients turn again to company website to submit their ideas with volume never seen before. However, this time John is prepared! The Idea Management platform is tightly interconnected with other development and corporate management systems that deliver a huge number of metrics and new capabilities, e.g.:

- Based on connections between Idea Management platform and project management system John can see which similar past ideas became successful and which failed in development. Therefore, he can asses the probability of success for new ideas.

- John discovers the true power of Semantic Web based integration as he can see how past ideas have been causing problems during and after implementation. Although the Idea Management System has never been integrated with company bug tracking environment, it has been with the project management suit, which data in turn is semantically interlinked with ideas and via simple reasoning delivers desired metrics (see Fig. 4.4).

- John is not an engineer, neither in charge of the product production cycle - it is hard for him to judge accurately the production difficulties that might emerge, as well as time and cost implications. However, thanks to the integration with the PLM system, metrics for past similar ideas are automatically extracted and John can see how much time and effort it took to develop them.

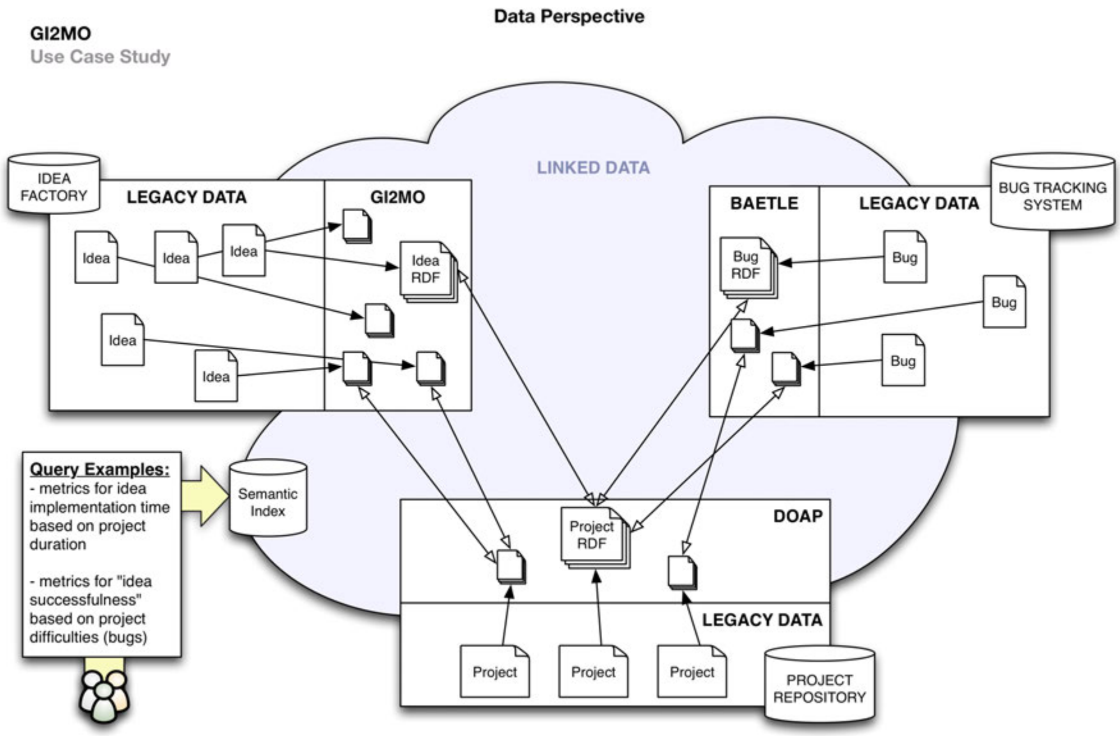


Figure 4.4: Scenario 1: Mining new metrics and interlinked data from internal company systems.

- John prises the flexibility of the Semantic Web technologies. When integrating the systems he did not have an idea what kind of metrics or connections between the systems he would need. However, as it became clear over time, the tight interlinking between ontologies for different systems and reasoning capabilities allow to quickly add new metrics without much effort.

Scenario 2: Discovering new ideas and assets through usage of common ontologies

The Idea Management Systems maintained by John turn out to work very good and provide a valuable supplement to the company innovation management process. However, John notices that he is missing a big amount of potentially good ideas that are submitted via other systems on the Web rather than his Idea Management facility (e.g. social portals, boards or blogs).

Fortunately, the Semantic Web technologies embedded in company facilities allow to discover and easily pull this new data inside the Idea Management System. The connections between the assets in the Idea Management System and in other systems are discovered via user profiling described with common ontologies on all portals. In case of John's Idea Management System, the ontology favoured to describe users is FOAF (Brickley and Miller, 2010). Fortunately, it happens to be linked with a popular solution used to describe users across social spaces (SIOC ontology (Breslin et al., 2005)). With little effort new idea mining system is deployed and starts to track connections via user profiles. In his growing happiness, while assessing his top contributors, John sees in an aggregated view that Mary also often publishes ideas using her blog. Thanks to using common technologies and ontologies that are interlinked, those ideas are pulled inside the Idea Management facility and can be assessed an analysed just like they would be posted normally via Idea Management System front-end (see Fig. 4.5).

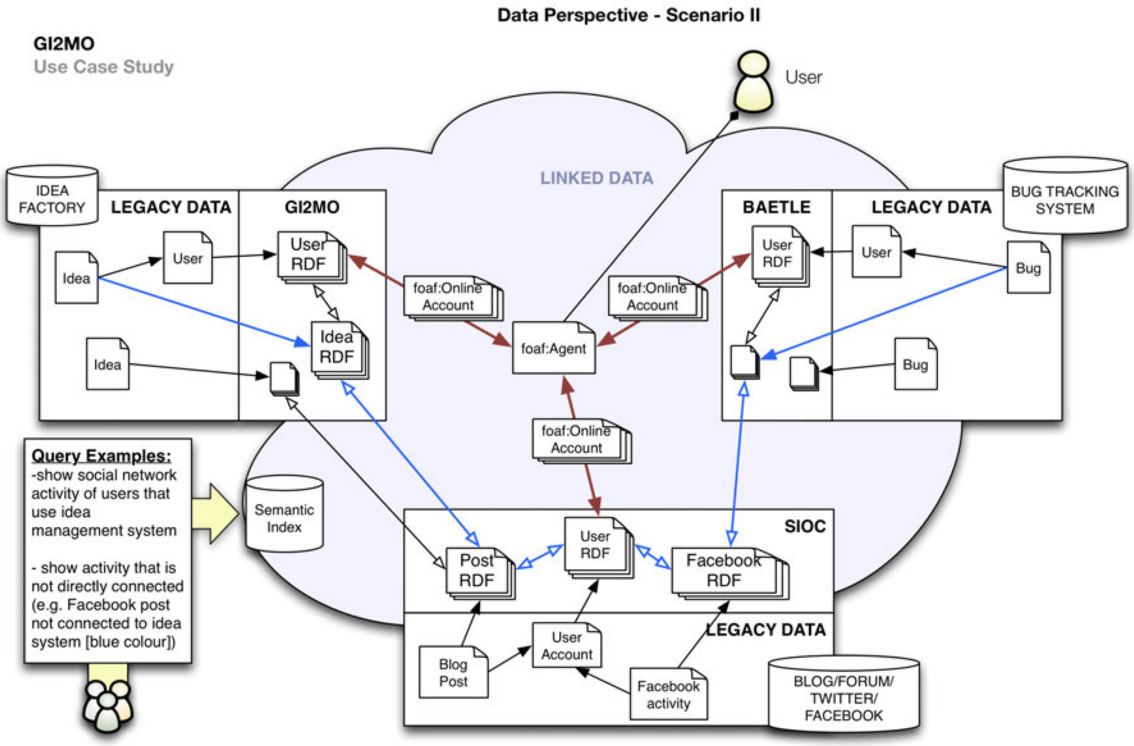


Figure 4.5: Scenario 2: Using common user profiling to discover new ideas.

4.4.2 Ontology Design and Implementation

To enable the presented use cases we have created an ontology to cover all the concepts described in the Idea Management Systems. As a preparation for that task we used a certain number of sources as a guide for modelling the data structure of this particular domain:

- analysis of publicly published data from operational Idea Management Systems (e.g. Dell IdeaStorm (Dell, 2009), myStarbucks (Starbucks, 2009) and othes)
- work with a sample commercial system (Atos PGI 2.0 (Atos, 2010))
- analysis of cases studies from the industry (presentations, publications, conference publications etc.)
- analysis of data based on a research done on definition of the Idea Life Cycle

As a result, we have defined a data model that serves as a base for designing the ontology and later applying it to a number of heterogeneous systems (see evaluation in Section 4.5).

One of the goals for our ontology was to make it available for others to apply to more than just the handful of systems that we experimented with. Therefore, one of the biggest challenges was to maintain the integrity with Semantic Web trends and standards yet keep the ontology simple and put impact on its usability and ease to appliance to encourage other developers. This resulted in a number of problems that can be generalized for every ontology design task but had to be resolved with our specific domain context in mind:

a. **Modelling open data vs. closed data.**
A large number of data stored in Idea Management Systems is not published for users that generate ideas (internal metrics, assessments, internal reviews, business analyses etc.). Furthermore, often the main means of idea assessment are statistics that differ very much depending on Idea Management System implementation (e.g. number of posts by given user or complex business metrics). Such richness and diversity of information results in a situation, were a big number of solutions for Idea Management are very simple ones, collecting little data in comparison to the sophisticated platforms. Therefore, a question rises whether the ontology should be extremely generic and simple or cover in detail the most sophisticated types of systems. Secondly, whether the ontology should be aimed for the sole goal of data publishing and search (e.g. like SIOC ontology) or further data analysis and reasoning (e.g. for multimedia operations (Simou et al., 2005)).

b. **Modelling for distributed publishing of Idea Management concepts vs. centralized model.**
In case of the distributed model (e.g. embedding RDFa on each page that represents different Idea Management concepts like Idea, Idea Comment etc.) the ontology grows double in size because of the necessity to implement inverse properties. Excluding such a possibility makes the ontology much easier to comprehend yet limits its use[1].

c. **Usage of existing ontologies for modelling Idea Management concepts.**
Describing common concepts over many different systems on the entire Internet with same vocabularies brings many benefits and simplifies overall perspective of the Semantic Web (Antoniou and van Harmelen, 2004; Fensel et al., 2005). However, when narrowing down to a single domain, the necessity to comprehend all those vocabularies to model a single system becomes a problem for a potential developer. Therefore, a question that we had to face in our research was whether or not to model Idea Management Systems with the use of external ontologies and to what extent.

We addressed all of the above questions during the ontology design phase and when applying the schema to operational Idea Management Systems as detailed below.

[1]The evolution of SIOC ontology specification (Bojars and Breslin, 2010) is a very good example of problems that come with the distributed model and the needs to preserve data schema simplicity.

Idea Management System Data Model

As an outcome of domain analysis and work with a number of different Idea Management Systems, we have listed all the data that is created or modified during each of the Idea Life Cycle phases. Based on that, we created a model for the Idea Management Systems that could be used as a reference to design the ontology. Following the earlier presented Idea Life Cycle analysis (see Sec. 4.3), the basic concepts for each of the respectable phases are:

- Idea Generation data (idea title, summary, creation/modification dates, attachments, categorizations etc.)
- Idea Improvement data (comments, user ratings, idea versions etc.)
- Idea Selection data (internal reviews, metrics, analysis and assessments)
- Idea Implementation data (information related to development process of a product/service based on the selected idea)
- Idea Deployment metrics (most often business metrics such as Return Of Investment, total cost etc.)

Those data assets created across life cycle stages have been associated with a number of concepts that are presented on Figure 4.6 and relate to the central concept of an idea in various ways.

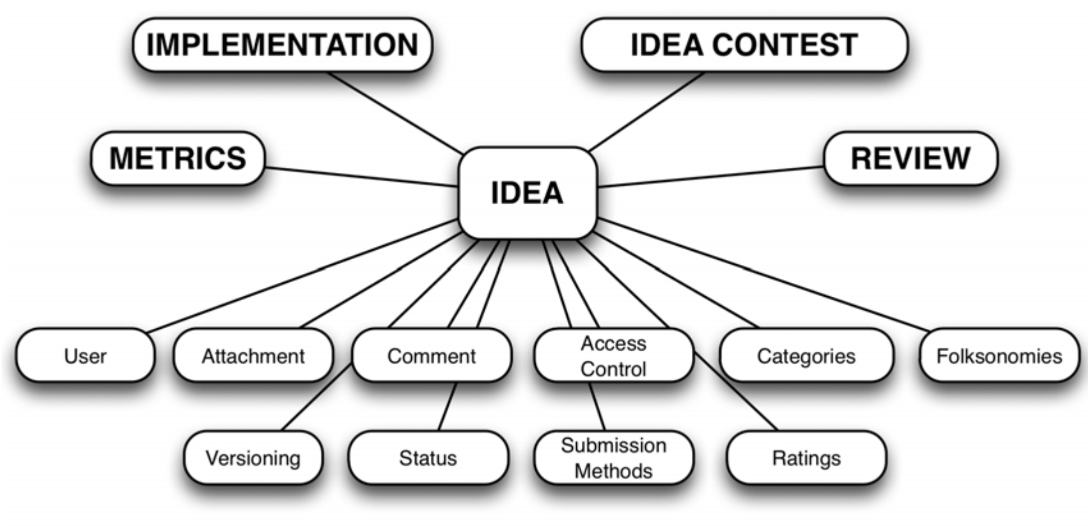


Figure 4.6: Idea Management System - a simplified idea centric data model

Apart of the information that is subject of change or is added during different life cycle states, there is a number of stable concepts that are present on each of the stages and can deliver useful information for idea selection and assessment, most interesting being:

- user data (Idea Management Systems can involve people working in a variety of roles that impact the innovation process in different ways (Westerski, 2012b))
- idea contest (a particular asset for Idea Management Systems - a themed event in time that initiates idea collection, e.g. "collecting ideas for the next product version release")
- idea status in the pipeline (can refer to the general Idea Life Cycle but also often has additional internal stages).

All together, the key concepts that define Idea Management Systems and differentiate its data from other social platforms are: Idea, Idea Contest, Idea Review, Idea Implementation and content Metrics that may refer to any of those concepts (see Table 4.1).

Table 4.1: Idea Management Data Model - core concepts overview

IMS Concept	**Description**
Idea	Idea is a central concept to the IMS, it models the content submitted by community inventors with respect to calls announced by organizers of IMS instances. Most important elements of idea concept are: idea title, idea textual description, idea submission date. Ideas can be subject of reviews; user ratings, comments and may evolve in time thus changing idea status.
Idea Contest	Idea Contests are events organized by owners of the IMS in order to stimulate the community to submit ideas on particular topic. Those events may have strict time boundaries, access rules and incentives such as money rewards attached. Contests are characterized by title and contest textual description.
Review	Idea reviews are assessments of idea value for the organization done internally and most often invisible for regular community members. Reviews may aggregate a number of rating mechanisms and criteria (e.g. financial reviews, technical reviews etc.).
Implementation	IMSes often report and point to implementation of ideas in order to stimulate the communities and motivate innovators via success stories. Furthermore, some IMSes offer agile project management facilities that make the transition from idea management ot implementation seamless. The Implementation concept refers to description of those implementations.
Metrics	Metrics are statistics associated with ideas, contests, users or entire IMS instances. The difference between metrics and reviews is that metrics are automatically generated based on analysis of the content (e.g. idea count for user, comment count for idea), while reviews are ratings and textual assessments performed by humans in order to judge idea value.
Status	One of the key elements of IMS that distinguishes its content from forums or blogs is the status of an idea and Idea Life Cycle. Ideas are organized based on their maturity and results assessment, which impact the status of an idea.

The remaining concepts recognized for the Idea Management Systems are a consequence of design based on social interaction and networking. Therefore, in majority those concepts are typical for many social platforms. However, depending on particular case and connection to core concepts, those social networking concepts may gain some special meaning (see Table 4.2).

Table 4.2: Idea Management Data Model - social networking concepts overview

IMS Concept	**Description**
User	Users are actors that create content of an Idea Management System. They may have a variety of access permissions to different content and can be organized in user groups.
Attachment	Attachments are non-textual extensions supplemented to ideas or idea contests, e.g. images, videos etc.
Comment	Comments are the main feedback instrument in IMSes. They can be attached to concepts such as idea or idea contests and are used to interact with the authors of content and discuss on idea improvement and status change.
Access Control	Idea Management Systems have a variety of access levels for most of their content. Some elements such as reviews are usually only visible for internal staff while ideas and idea contests can be restricted to specific user groups (e.g. employees of a certain department).
Categories	Categories in IMSes are used to organize content. Categories can be visible for regular users to enable self-organization of content as ideas are being created (e.g. product categories) or can serve as internal taxonomies (e.g. for organization of review types).
Folksonomies	Folksonomies were adopted in IMSes similar as in other social platforms with the advent of Web 2.0. Tags and tagging activities are used as an additional tool for organization of ideas.
Versioning	In Idea Management Systems ideas are not set in stone, they evolve, are being updated by the original innovators or can be merged with other ideas by moderation activities. The concept of versioning provides references to different version of an idea to preserve information on how an idea has evolved over time.
Submission Method	While majority of the IMSes are Web oriented platforms the continuous progress of IT brings new interaction methods with the social platforms. The concept of submission method refers to describing the input method with which an idea has been generated (e.g. via web input form, via mobile device but also via web scraping or brainstorming software).
Ratings	Ratings are second to comments in tools for crowdsourcing community assessments of ideas. Depending on the system, ratings can be organized in different ways (percent or point ratings). Furthermore, rating can be included in internal reviews as a supplement to textual reviews.

Ontology Schema

Based on the presented earlier data schema we constructed an ontology that aims to clip all the phases of idea management process together and allow to analyse the connections between (e.g. how idea input phase influences idea implementation etc.). Since we put most impact on connections between assets the natural choice was to base on research done in Semantic Web and support our work with achievements of technologies such as OWL, RDFS and research on other domain ontologies (e.g. GoodRelations (Hepp, 2008)).

The most important concepts of Idea Management Systems that we wanted to interconnect and that have driven the design of the Gi2MO ontology are:

- idea version control (track history of changes to e.g. see if ideas improved a lot provide better results in terms of different metrics such as revenue, cost etc.)
- idea pipeline modelling (building dependencies between phases and infrastructure o establish links)
- modelling dependencies with internal assets (other ideas) and external assets (ideas from other systems or other media resources)

The overview of classes included in the ontology is shown on Figure 4.7. For a detailed technical information with listing of all properties please refer to the Appendix B as well as the full ontology specification on Gi2MO Project website (Westerski, 2012b).

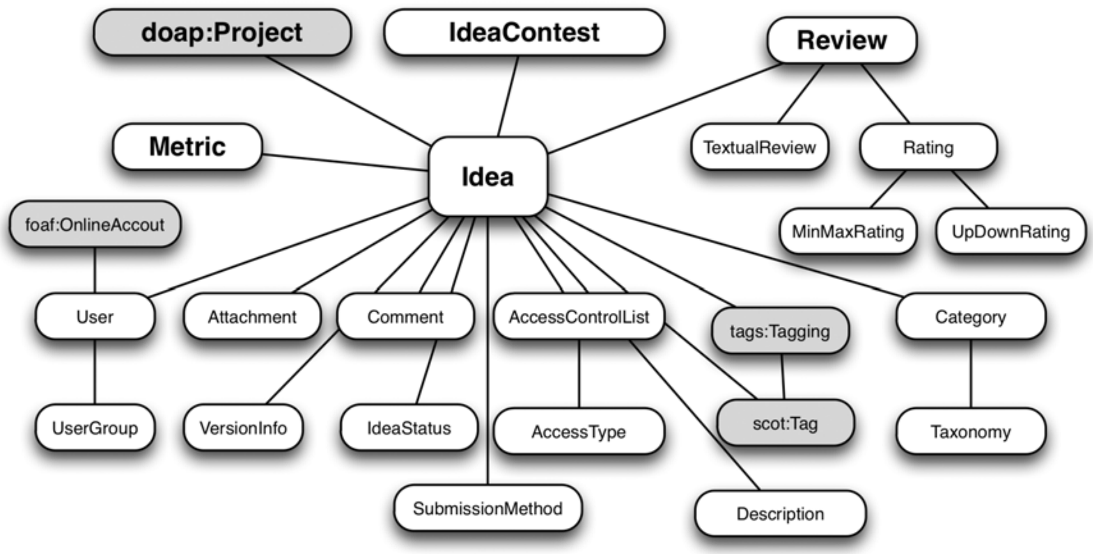


Figure 4.7: Gi2MO ontology class design overview

Connections with other ontologies

During the creation of the Gi2MO ontology we decided to model a part of the Idea Management System using existing ontologies. To simplify the task we took a two step approach: first we modelled the entire data model with a single name space (Gi2MO v0.1) and next we gradually started introducing other ontologies in certain areas (Gi2MO v0.2). The list of all imports can be seen in Table 4.3.

Table 4.3: Gi2MO ontology imports

Ontology	Description of concepts modelled
DCMI Terms (DCMI Usage Board, 2012)	Generic properties for many assets, e.g.: 'title', 'description' etc.
DOAP (Dumbill, 2012)	Idea implementation information
FOAF (Brickley and Miller, 2010)	Relation between User Account and personal data
SKOS (Miles and Bechhofer, 2009)	Idea Categorization and taxonomy hierarchy modelling
SCOT (Kim and Breslin, 2012)	Tags and tagging activities

4.5 Evaluation

We have identified that the two most critical actions during applying an ontology to a domain specific system are: **data migration** and **connecting the new data layer with the application logic** to take advantage of the new capabilities. Therefore, we divided the evaluation activities into two separate phases respectable for those problems. Firstly, we wanted to test the coverage of the ontology on different data sets available and recognize any potential problems. Secondly, we aimed to evaluate the ontology in a development environment where data encoded with Gi2MO vocabulary would be put into a particular use.

4.5.1 Ontology Coverage Study

For the first evaluation task we tested the coverage of ontology defined as: *the amount of properties that could be modelled using the Gi2MO Ontology in relation to all the data properties discovered for a particular dataset.* During our experiments we used the data available on-line from the Idea Management facilities open to public use. The datasets were as follows:

- Dell IdeaStorm - based on SalesForce Ideas platform
- myStarBucks Ideas - based on SalesForce Ideas platform
- Adobe Acrobat Idea - based on BrightIdea platform
- Cisco i-Prize - based on Spigit platform

Data coming from the two instances of SalesForce Ideas platform were used to see the differences that can occur within a deployment of the same system but profiled for different companies. Next, we pursued to mine data from two other systems and see further the variations that occur. In order to obtain data from the Idea Management Systems, we

Table 4.4: Quantitative results in RDF triples for ontology coverage experiment

Portal Name	# of Ideas	# of Triples	# of Gi2MO Triples[1]		
			v0.1	v0.2	v0.3
Dell IdeaStorm	9851	520330	427248	250869	250869[2]
myStarBucks	10949	194086	153040	89638	89638[2]
Adobe Ideas	579	17859	13292	7499	8798
Cisco i-Prize	826	133413	94262	69628	81950

developed custom HTML scraping tools, while to encode the information in RDF we used D2RQ tool (Bizer, 2004; D2RQ, 2012) with specific mappings to Gi2MO for each test case. To make the experiment more reliable we asked university students, being independent to the ontology creation process, to perform the mappings without our supervision. Furthermore, we repeated the experiment three times for different versions of the ontology:

- Gi2MO v0.1 - entire Idea Management System data model covered by Gi2MO
- Gi2MO v0.2 - the same data model based on Idea Life Cycle as in v0.1 but the formalization introduces imported name spaces of other ontologies to describe a number of concepts (as earlier proposed in table 4.3)
- Gi2MO v0.3 - final version constructed after the main evaluation tasks to adjust the ontology to the preliminary results and include new concepts omitted by our methdology based on the Idea Life Cycle data modelling

The quantitative results of RDFization with those three different iterations of the Gi2MO ontology are presented in Table 4.4.

Analysing the results we can make two interesting observations. First, the amount of ideas does not always have a direct impact on how data size and complexity scales. Although IdeaStorm and myStarBucks are based on the same system with almost identical capabilities, a smaller amount of ideas in IdeaStorm produced a much larger number of connections between assets than in myStarBucks. As we found out, in this particular case the reason was huge user activity in IdeaStorm in terms of idea reviews. Furthermore, as we experimented with systems from other vendors, we noticed that the amount of descriptive data and interconnections produced can rise into very high numbers just because of the amount of metrics published. In those terms Cisco i-Prize was the most rich, while the SalesForce systems had least of such data. This state translated in a great way into relation between amount of ideas and triples: respectably 161 triples/idea in i-Prize and 17 triples/idea in myStarbucks.

Secondly, apart of the quantitative analysis more important for us was the coverage of the ontology versus the data mined from different Idea Management facilities. By performing this experiment we wanted to verify if the methodology to build an ontology based on our Idea Life Cycle model was correct. Taken into account all the four IMS instances from the

[1] Triples where the predicate is expressed with Gi2MO ontology

[2] The experiment with Gi2MO Ontology in version v0.2 already covered the entire dataset so no changes were introduced by modifications in version v0.3

Table 4.5: Property mapping results for ontology coverage experiment

Portal Name	**Ontology**	**Properties Covered**					
		v0.1		**v0.2**		**v0.3**	
		property count	property percent	property count	property percent	property count	property percent
Dell IdeaStorm	Gi2MO	21	100%	11	52%	11	52%
	imported	0	0%	10	48%	10	48%
	uncovered	0	0%	0	0%	0	0%
	total	21	100%	21	100%	21	100%
myStarBucks	Gi2MO	14	100%	8	57%	8	57%
	imported	0	0%	6	43%	6	43%
	uncovered	0	0%	0	0%	0	0%
	total	14	100%	14	100%	14	100%
Adobe Ideas	Gi2MO	24	75%	13	40%	18	56%
	imported	0	0%	11	35%	14	44%
	uncovered	8	25%	8	25%	0	0%
	total	32	100%	32	100%	32	100%
Cisco i-Prize	Gi2MO	43	72%	29	49%	36	60%
	imported	0	0%	14	23%	24	40%
	uncovered	17	28%	17	28%	0	0%
	total	60	100%	60	100%	60	100%
Average	Gi2MO	25	**87%**	15	**49%**	18	**56%**

experiment, our ontology together with suggested ontology imports (v0.2) covered on average 87% of the metadata. In particular, as we noticed, for the two first front-ends of the same vendor (IdeaStorm and myStarBucks), the data structure was quite similar. Therefore, as a result we got 100% coverage for Gi2MO ontology for every iteration of the experiment with quite similar mappings reused in both cases. However, the last system (Cisco i-Prize) proved to be quite different and served us as a valuable lesson on how Idea Management Systems can variate depending on the vendor. In case of this platform, the coverage was around 72% mostly due to rich metadata assigned to user profiles that we did not take into consideration before. Such evaluation made us rethink some of the elements of the ontology and suggestions for ontology reuse that were included v0.3 iteration of Gi2MO specification. The full results in terms of coverage of Gi2MO properties are shown in Table 4.5.

The presented data additionally shows use of ontology imports and Gi2MO ontology properties. On average Gi2MO ontology was used to serialize 49% of the properties in version v0.2, while 56% in the final v0.3 version. This can lead to an interesting insight that our choice of referring to Semantic Web research was quite successful. Although a significant part of metadata has been covered due to the contribution of our proposal, the existent Semantic Web ontologies provided a formalization for nearly half of the Idea Management metadata.

4.5.2 Ontology Utilization Study

To test how the ontology would work in practice for an end user driven use case we turned for help to university students again and asked to implement a visualisation mechanism that would allow to categorize and view ideas from many heterogeneous sources. As a result, we got a web application (see Fig. 4.8) with the data back-end of the system entirely RDF driven and capable to work with the data mined earlier. During the experiment we did not observe any major issues related to comprehending the Gi2MO ontology or the documentation delivered. However, the biggest standing problem that emerged was scalability - the application could not handle RDF dumps from the previous experiment at their full size. The specifics of the implementation challenges and detailed discussion of the scalability problems are presented by Rico (Rico, 2010) who developed the Idea Browser application as part of his graduate thesis under the supervision of Gi2MO Project.

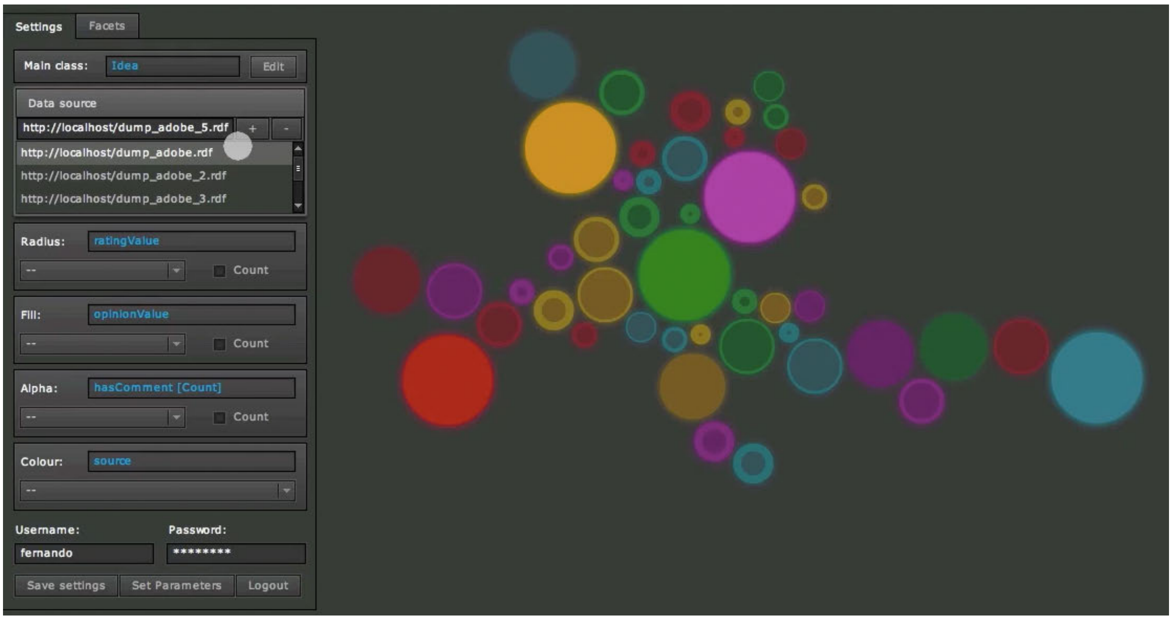


Figure 4.8: Web application for faceted browsing of Ideas

Additionally, apart of data visualisation, the discussed Idea Browser application was an experiment to provide idea assessment functionality based only on the data extracted from the existing systems and serialized according to Gi2MO Ontology. The tool allowed to select, browse and filter ideas based on idea properties. In particular, we experimented with analysis of: 1) RDF object properties that formalize the relationships between various concepts describing an idea or relationships between ideas; 2) RDF data properties that explicitly express literals and values submitted by the community or generated by the particular IMS instance (e.g. comment count). The conclusion from prototyping phase and usage experiments of the ontology with Idea Browser was that community generated information is very sparse in terms of amount of relationships between concepts as well as amount of metadata attached to individual idea in general. Therefore, in order to fulfil to objective of the thesis to improve the contemporary idea assessment methodologies, more idea metadata would be required. This conclusion set the direction for the rest of the contributions of the thesis, as we decided to investigate new ways of obtaining idea metadata and increasing the granularity of data submitted by community members in Idea Management Systems. The further research that was conducted to achieve those goals is presented in the next chapters.

4.6 Summary

In this chapter we have described the research done on building a consistent and theoretically grounded model for Idea Management Systems. The goal of the proposed model was to deliver a single point of reference that could be extended in various ways to fulfil the objectives of the thesis regarding improvement of idea assessment process.

In particular, we have contributed a life cycle proposal for the data stored in the Idea Management Systems as well as a list of dependencies between information created on the various stages of the proposed life cycle.

Furthermore, based on the proposed life cycle, we have contributed an ontology for Idea Management Systems - a formalization of metadata that can be used to describe ideas and associated information. The ontology was proposed as a universal metadata schema to be applied in any sort of Idea Management System regardless of vendor or deployment use case.

According to presented evaluation results, we have fulfilled this goal though achieving 87% ontology coverage on average for a variety of completely different IMS deployments. Furthermore, we described the use of datasets formalized with the ontology during prototyping experiments with the students of UPM university who developed idea assessment applications. As a result of this stage, we observed that although our model was complete, there is very little amount of idea metadata that could lead to development of techniques for idea assessment. In conclusion, to fulfil the remaining goals of the thesis, we decided to research on extending the amount of information available though various post processing activities as well as introducing new activities into the Idea Life Cycle workflow.

The remaining elements of the solution architecture that implement this research are presented in the following chapters (Chapter 5, 6, and 7) building on top of the so far proposed Gi2MO ontological model.

Chapter 5

Community Opinions Model for Idea Management System

In this chapter we focus on community activity that is manifested in Idea Management Systems by comments attached to ideas. In particular, we propose the use of opinion mining techniques to analyse the comments of ideas and construct a new metric (opinion rating) that aggregates the sentiment value of all comments of a particular idea.

Firstly, following the analysis of the opinion mining state of the art, we propose an ontological model that extends the previously introduced Gi2MO ontology with formalization of opinion concepts. On the road to achieving this, we answer the research question to what extent opinion information can be formalized in a unified way. We analyse various domains and scenarios for usage of such formalization to deliver a final solution that would integrate opinion data of Idea Management Systems as well as other web systems. The goal of opinion ontology is to enable data portability and comparison of Idea Management community activity beyond the boundaries of a single system instance.

Secondly, we propose how to utilize the results of the opinion mining process for assessment of ideas. In particular, we present a method for aggregating opinions and calculating a metric that expresses the sentiment of comments attached to an idea. We compare the use of this metric in relation to previously available statistics for assessment of idea successfulness.

The contributions presented in the chapter are:

- Marl Opinion Ontology
- A metric for ranking ideas based on opinions expressed in idea comments

5.1 Introduction

One of the core characteristics of the Idea Management Systems is the participatory role of the community. The notion of crowd-sourcing is employed by inviting customers or employees to comment and collaboratively improve the submitted ideas. While this solution enables to extend ideas in a unique way and delivers new valuable knowledge about the enterprise environment, it also introduces a number of problems related to the amount of data that needs to be processed during idea evaluation.

As reported in case studies (Jouret, 2009), observing the reactions of clients on ideas is a very time consuming yet important element for the competition organizers in Idea Management Systems. In the contemporary solutions, the organization of large amounts of data is harnessed usually in two ways (Hrastinski et al., 2010): 1) by idea rankings based on automatically generated community statistics (comment count, vote count, view count etc.); 2) expert reviews (that is: designated reviewers filling out metrics based on idea text and submitted comments). The community generated statistics require little effort to generate but do not give a full view of the community opinions expressed in the comments. This is partially resolved by up/down voting mechanism, however the users voting activity does not necessarily reflect the emotions of a discussion that the user initiates in repose to an idea. On the other hand, thorough moderation and expert reviews solve the problem of omitting the significance of community discussions. Unfortunately, the biggest problem of this method is that it is inefficient in terms of time and human resources that need to be delegated to analyse thousands of ideas and hundreds of thousands of comments. Furthermore, studies have shown (Gangi and Wasko, 2009) that either of the two idea assessment methods have a small impact on the final choice of ideas that are implemented.

Taking into account the aforementioned achievements of research on Idea Management metrics and leveraging our own research presented in the previous chapter, we sought to answer the question: how to transform or extend the available data of an Idea Management System to improve the idea review capabilities with respect to understanding the semantics of community opinions. For achieving that goal, we propose the analysis of the comments that users create when discussing the value of ideas.

To find a particular method for efficient analysis of comments, we looked into previous research done in the area of on-line web communities and natural language processing. The rise of the Social Web has stimulated progress in those disciplines and gave birth to new trends. One of the research domains that noted especially big progress within recent years is opinion mining. From the information systems point of view, opinion mining aims to harness the flows of unstructured (or poorly structured) subjective user generated textual content that otherwise is hard to analyse, accurately categorise and reason upon. In web systems such as Idea Management Systems, opinion mining could be incorporated to improve browsing and assessment of data. However, as Softic et al. (Softic and Hausenblas, 2008) point out, in order to use opinion mining to improve browsing the Social Web there is a need for a common metadata proposal that would enable interoperability between heterogeneous systems and allow comparison of their data.

In the research presented in this chapter, we establish this common web metadata schema that would enable to publish in a formalized manner the results of the opinion mining process. We motivate our work by presenting a global scenario for all web applications as well as one specifically aligned for Idea Management Systems (see Sec. 5.2). As we report on the research done, firstly we introduce an abstract data model - an ontology that formalizes all concepts derived from the opinion mining process (see Sec. 5.4). Further, in connection to the Gi2MO ontology presented in the previous chapter, we propose the use of Semantic Web technologies to adapt that opinion ontology for web use and show exactly what profits can that bring (see Sec. 5.5). Finally, we propose the use of formalized opinion mining data in Idea Management to generate a metric that together with the ontology can be employed to compare distributed communities associated with different Idea Management instances (see Sec. 5.6.2).

5.2 Motivation

5.2.1 Publishing and consuming opinion data on the web

Embedding opinion mining functionality for websites that are rich in user comments can aid to automatically rank comments and let users faster reach the types of opinions that they seek (Pang and Lee, 2008). Furthermore, given the same data, opinion mining algorithms can be used to supply additional metrics to rate products and content (Tian et al., 2009). However, all of this value is often limited only to the single site of origin that performed the opinion mining algorithm.

Based on the achievements and research done in the area of Semantic Web (Berners-Lee et al., May 2001) and more specifically its evolution into proposal of Linked Data (Berners-Lee, 2006), we motivate our work with publishing opinion information using a universal metadata format that would extend the usability of such data. First and foremost, when having opinions described across the Internet in a unified way it is possible to compare them and perform an Internet wide search and statistics. At the moment it is possible to find opinions of desired polarity about selected product using contemporary Internet search engines, however the simple text based indexing is far less accurate and less flexible than what could be achieved with metadata indexing (Hemminger et al., 2007). Furthermore, if the opinion mining data would be accompanied and linked with other metadata that describes the context of the subjective content, then the capabilities of search and browsing would rise even more (e.g. with regard to aggregation and data mashups (Catasta et al., 2009)).

Finally, above of the aforementioned motivations, we would like to point to what currently seems to be the principal argument for content providers to publish metadata: improve visibility on the Web and in the search engines. Metadata can help to increase the precision and recall of search (Beall, 2010) but also the value of metadata becomes more visible as the search results in the leading Internet search engines start to contain data extracted form the metadata published along with HTML (e.g. Google Rich Snippets [1]), thus making particular search results more attractive in comparison to competitive links. Through annotation of

[1] http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html

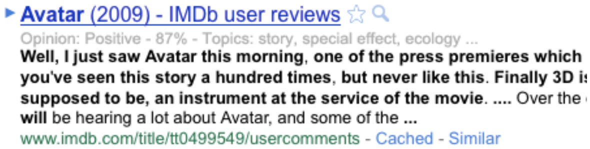


Figure 5.1: A Google snippet modified with Greasemonkey script and enriched with data extracted from RDF

opinions, exactly the same benefit could be delivered for the websites that provide opinion mining results over subjective content posted on them or remote sources (see Fig. 5.1).

5.2.2 Using opinion metadata in Idea Management Systems

Idea Management Systems are often deployed as web systems that connect to the customer base of the organization via Internet. Therefore, the previously described scenarios for publishing and consuming metadata from other web systems can be applied as well (see Fig. 5.2). Additionally, the introduction of interoperability of community generated data can help also within a more narrow scope of IMS to IMS interoperability: 1) compare communities of different Idea Management deployments; or 2) integrate multiple instances that collect similar ideas but from different language groups and aggregate opinions on the same topics.

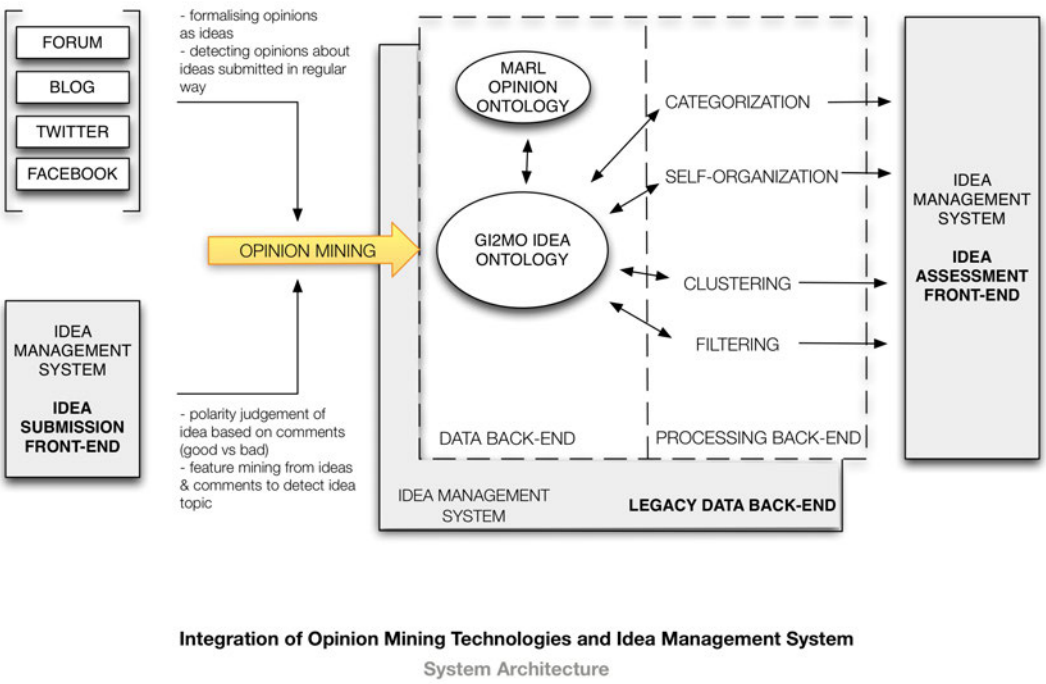


Figure 5.2: Use of opinion mining in Idea Management Systems - motivation overview.

5.3 Related Work

5.3.1 Formalizing Community Activity

The research presented in this section is primary focused on developing a universal model for describing and comparing opinions on the World Wide Web and in ecosystems of various computer systems. As such, it is tied to efforts of the Semantic Web research community, which goals have been outlined by Sir Tim Berners-Lee (Berners-Lee, 1998). Furthermore, as much as we are interested in reasoning and giving birth to the intelligent web, our research is focused to a much more extent on the sole goal of publishing and consuming data. Therefore, we have aligned our investigation with the efforts undertaken by the Linking Open Data project[2] - an attempt to build an interlinked Web of Data using Semantic Web technologies.

In terms of related research conducted in those areas, to our knowledge, there has been only one attempt to achieve a similar goal as our. Softic et al. (Softic and Hausenblas, 2008) has proposed an opinion ontology and performed a number of experiments to show its use. However, as authors claim themselves the ontology is unfinished and missing the key element of opinion formalization leaving it for later research which has not done yet. In our work we aimed to use the opinion mining as a tool in our main research area of Innovation Management, therefore we needed a full solution for metadata publishing that could be applied in practice. In comparison to Softic et al. we propose a different conceptual model for the opinion ontology, deliver new properties that describe not only a generic concept but enable to publish the numerical values from the opinion mining process (which is impossible using Softic et al. opinion ontology). Furthermore, with our research we propose a different evaluation framework and test our solution in different cases, which in the end delivers new conclusions and opens new possibilities.

Within commercial services related to the area of opinion mining there are different data serialization methods used for APIs but all use own vocabularies. In relation to our work, a standing out service by Opendover (Opendover, 2012) moves towards the Semantic Web technologies but the vocabularies used refer only to individual sentiments (thus being more similar to a dictionary) rather than full opinions like in case of Marl ontology.

On the other hand, not related to opinion mining, we recognize that for a practical solution, opinions could be conceptually modelled as reviews. Therefore, in terms of related work we also considered vocabularies created for describing online reviews. Among those, the most popular are: hReviews (Allsopp, 2007), the RDF mapping of hReview (Ayers and Heath, 2007), Google's RDF vocabulary for reviews[3] and Schema.org Review vocabulary[4]. In comparison to our work the existing review formalization vocabularies are much more generic and conceptually describe less referring to the entire review body, whereas we see that the opinion ontology needs to describe particular elements of the review and features discussed in the review (e.g. one might imagine a query using both concepts "show all sci-fi movie

[2]http://esw.w3.org/SweoIG/TaskForces/CommunityProjects /LinkingOpenData
[3]http://www.google.com/support/webmasters/bin/answer.py?answer=146645
[4]http://schema.org/Review

reviews that contain positive **opinions** about director"). Furthermore, we see reviews as judgement based on factual information and comprehensive knowledge whereas opinions are less formal, smaller pieces of information. For those reasons we believe there is a need for making a distinction between the two concepts in terms of metadata and web search.

5.3.2 Community Activity Measures

Aside of contributing a formalization for community opinions, we propose a specific use of our work through obtaining and post-processing opinion data with opinion mining algorithms. In the past years opinion mining has been a very active domain that vastly increased its research activity (Esuli, 2007b; Wiebe, 2012) along with the evolution of the Social Web and the growing popularity of Web 2.0 technologies (Lytras and de Pablos, 2009). The variety of approaches can be split into (Liu, 2008a): document subjectivity judgement, sentence analysis, or feature analysis. Depending on the taken approach, the contemporary solutions deliver accuracy ranging from 60% for simple keyword methods though 80% for various pattern matching or machine learning solutions (Pang et al., 2002), up to 90% and above for domain optimized algorithms (Yu and Hatzivassiloglou, 2003). In comparison to this classification, for the needs of our research, we constructed a tool that treats idea comments as single documents and employs a keyword based approach.

In addition to development of different opinion mining approaches that improve the entire process accuracy, researchers have also proposed the inclusion of opinion mining into a pipeline of a larger scope. The usage of sentiment analysis has been evaluated in a number of domains such as: product review mining and summarization (Zhuang et al., 2006), business and government intelligence (e.g. trend prediction in sales (Mishne and Glance, 2006)), analysis of public opinions before political elections (Mullen and Malouf, 2008). Our work fits into this landscape as we do not contribute to opinion mining algorithms but innovate though proposing a novel application of opinion mining and based on it advance in the area of Idea Management though adding a new metric and method for idea comparison.

Within the domain of Idea Management Systems in specific, there have been a few other attempts to employ opinion mining to improve idea review practices. In particular, Bothos et al. (Bothos et al., 2008) proposed using opinion mining to improve prediction markets technique for rating ideas. In contrast our research in Gi2MO project focuses on comparison of distributed Idea Management Systems via sentiments of associated communities (Westerski et al., 2011b). Furthermore, we go a step further, and rather than focusing on details of application we pursue the evaluation of usefulness of textual opinions in Idea Management Systems in general. More precisely, we verify if inspecting community generated comments does actually influence the idea review workflow in a different way than other contemporary metrics. As such, our study aims to supplement previous work done in the area.

5.4 Marl: An Ontology for Describing Community Opinions

When designing the ontology our aim was to analyse the properties that characterize opinions expressed on the web or inside various IT systems. The final set of concepts that we propose (see Fig. 5.3) consists of constructs that describe 3 main areas of the opinion metadata: *opinion context*, *described object information* and *opinion mining metadata*. The first group (*opinon context*) contains information about the location where the opinion has been expressed and opinion text. The second part (*described object*) points to information that would allow to identify the item or object that is being reviewed or analysed in the opinion. The final part of our ontology (*opinion mining data*) groups information that can be obtained with opinion mining algorithms.

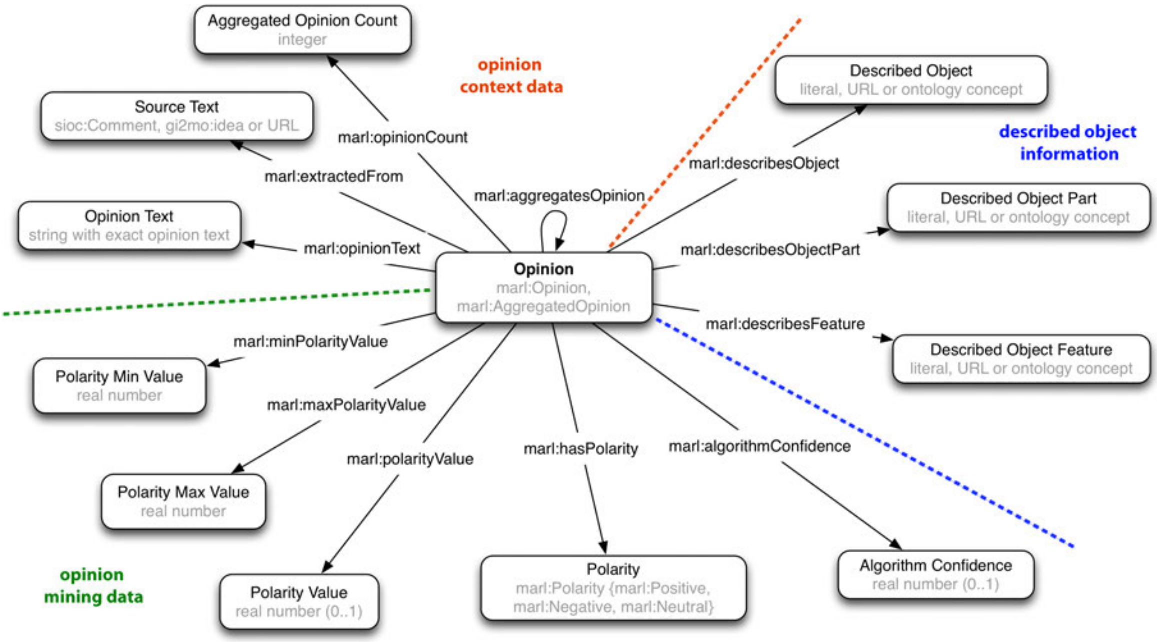


Figure 5.3: Conceptual model for opinion and the proposed **Marl ontology**

Such organization of the ontology is a result of a two step research process. First, we analysed different kinds of subjective data sources and produced a common model that was formalized as Marl Ontology v0.1. For this part we started with three common cases of opinions expressed on the Web: movie opinions, movie review opinions and products opinions. Later, in addition, we also analysed characteristics of opinions in enclosed communities and used an enterprise open innovation system as a case study. In the second phase, we evaluated the proposed ontology against live data and corrected the discovered drawbacks in version 0.2 of the ontology. The description of particular properties and explanation of their meaning can be found in Table 5.1.

In the particular model that we created we attempted to center all the data properties

[1] Properties added in Marl v0.2

Table 5.1: Marl ontology: classes and properties breakdown

Name	**Description**
Opinion	Class that represents the opinion concept
extractedFrom	Indicates the source text from which the opinion has been extracted.
opinionText[1]	The exact string that contains the calculated sentiment.
hasPolarity	Points to either entity or literal that indicates if the opinion is positive/negative or neutral
polarityValue	A numerical representation of the polarity value.
maxPolarityValue	Maximal possible numerical value for the opinion
minPolarityValue	Lowest possible numerical value of the opinion
describesObject	Indicates the object that the opinion refers to
describesObjectPart	Indicates a particular element of the object that the opinion refers to (eg. laptop *battery*)
describesFeature	Points to a feature of an object that the opinion refers to (eg. laptop battery *life*)
algorithmConfidence	A number that describes how much the algorithm was confident with its assessment
AggregatedOpinion	Subclass of Opinion class that aggregates a number of opinions.
aggregatesOpinion	Points to Opinion instances that are aggregated
opinionCount[1]	Amount of opinions aggregated.
Polarity	Instances of this class represent the positive, neutral or negative polarity

around a single opinion class. This and other ontology design choices that we made with Marl relate to one of the common problems of modelling ontologies for web use: the choice between modelling certain concepts fully as classes of domain ontologies, literals or simply URLs. While for using the full potential of Semantic Web it is best to model metadata concepts as entities described by particular ontologies the reality proves that this is far from being a practical solution. Therefore, we propose a model that accommodates both (see Fig. 5.4) assuring future extendibility yet facilitating more simple and practical use. In the next section we describe the benefits and applications of either of the cases.

```
(1) Example A: Entity referencing for describing contextual information
marl:describesObject <http://dbpedia.org/
                      resource/Avatar_%282009_film%29>
marl:describesObjectPart dbpedia-owl:director
marl:extractedFrom <http://www.gi2mo.org/
                    index.php?sioc_type=comment&sioc_id=157>
marl:polarity marl:Positive
marl:polarityValue "0.6"

(2) Example B: Using literal values to describe contextual information
marl:describesObject "Avatar"
marl:describesObjectPart "director"
marl:extractedFrom <http://www.gi2mo.org/
                    2010/09/introducing-marl/#comment-157>
marl:polarity marl:Positive
marl:polarityValue "0.6"
```

Figure 5.4: Referencing entities (1) and using literals (2) with Marl ontology

5.5 Publishing and consuming opinion metadata on the web with Marl

Following the description of the opinion ontology we show its possible uses and the differences that various closed and open systems impose. Furthermore, to support the ontology design decisions described earlier, we expose the benefits and drawbacks of publishing opinion data in different forms and with a different level of detail using the Marl ontology.

5.5.1 Internet wide keyword search and comparison of opinion values

In the simplest case where opinion ontology would be used only with properties expressed with literals, the structure information (connection between opinion text, opinion value and the full body of text) can still be very useful. Even with the contemporary keyword search engines publishing opinion metadata could make a lot of sense. While the discovery of information remains impaired and inaccurate, once actually having found the desired textually expressed opinions, thanks to the metadata it is possible to compare them or transform in different ways. Furthermore, as the research on semantic metadata indexing (Oren et al., 2008) progresses it is already possible to utilize these simple relationships to make useful search queries on large data sets (see Fig. 5.5).

```
* <http://purl.org/marl/ns#extractedFrom> * AND
* <http://purl.org/marl/ns#hasPolarity> <http://purl.org/
                                        marl/ns#Positive> AND
* <http://purl.org/marl/ns#describesObject> "Avatar"
```

Figure 5.5: Sindice Semantic Index (Oren et al., 2008) sample query for: "Search positive opinions about Avatar"

5.5.2 Internet wide entity based search and/or improved data discovery

One of the envisioned bold goals of Semantic Web is to provide entity based search. This would allow to point exact concepts that the user is referring to and eliminate ambiguity of user query present in the keyword search. Slowly this is becoming achievable much due to popularization of big linked data silos (e.g. DBpedia (Auer et al., 2007b)) and wide adoption of certain ontologies (e.g. GoodRelations (Hepp, 2008)). In our research, we also considered using opinion metadata in such scenario.

```
* <http://purl.org/marl/ns#extractedFrom> * AND
* <http://purl.org/marl/ns#hasPolarity> <http://purl.org/
                                        marl/ns#Positive> AND
* <http://purl.org/marl/ns#describesObject> <http://dbpedia.org/
                                        resource/Avatar_%282009_film%29>
```

Figure 5.6: Sindice Semantic Index (Oren et al., 2008) sample query for: "Search positive opinions about Avatar" using DBpedia Avatar entity for disambiguation

In comparison to the previous case, instead of using literals to describe opinion context, Marl ontology properties point to the exact concepts defined in one of the commonly refereed datasets. This, for example, allows to formulate queries that distinguish opinions about "Avatar" movie by James Cameron from other meanings of this word (see Fig. 5.6).

From a technical point of view, the establishment of such metadata infrastructure would physically link the opinions together with the Linked Data cloud and therefore each other as well via reference to similar topics. In turn, this would allow to traverse the distributed graph in many different ways for numerous use cases, such as aggregation of opinions (see Fig. 5.7).

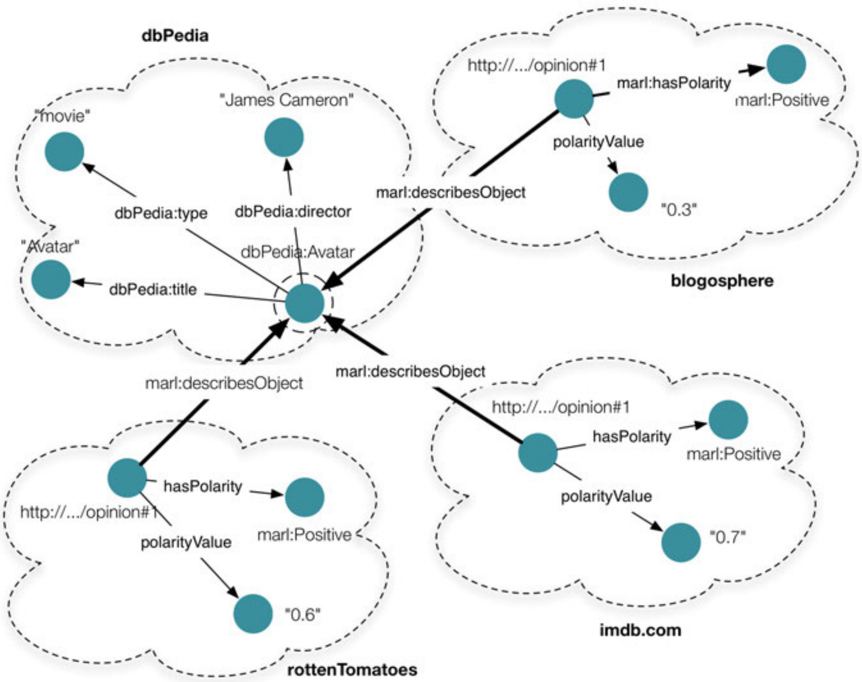


Figure 5.7: Sample RDF graph with opinions linked indirectly via metadata references to common entities.

5.5.3 Semantic search engines for dedicated systems

The large scale entity search engines still cope with a number of problems such as insufficient data, efficiency problems etc. even in aforementioned cases of vertical search (e.g. single topic or content type, like the movies). Nevertheless, we also would like to show that similar techniques, that expose the benefits of Marl ontology, can be very useful even if limited to very narrow systems or groups of heterogeneous systems where most of the problems of Internet wide search are eliminated (e.g. in an enterprise).

Following the example of movies that we used in previous cases, the local search could limit to a single website but thanks to the rich data descriptions with the ontological structure it would enable more precise queries than in text search. In this case Marl fills the gap for describing opinions in conjunction with complex taxonomy trees that enable to query for opinions related to particular elements in the hierarchy that characterizes a given domain.

Finally, one can move away from the World Wide Web context to the enterprise environments or other closed systems. In such case the difference is the full control over created data and very strictly defined vocabularies that do not need to be aligned with Web publishing standards. In that case, Marl can be used to together with the variety of enterprise ontologies in the enterprise collaborative systems (e.g. Idea Management Systems or collaborative knowledge management systems). The opinions can be linked via products that they refer to, innovation proposals that are commented by employees, projects in which context the opinions are expressed etc.

5.6 Evaluation

5.6.1 Data model coverage and annotation experiments

In order to evaluate our proposal for annotation of opinions we did two experiments. In the first, the goal was to analyse the coverage of the proposed schema against different datasets. In the second experiment we wanted to test in practice how the linked opinion metadata would work with the capabilities of the contemporary search engines and semantic web query endpoints.

During the coverage experiments we analysed two kinds of data: (a) datasets created by other researchers and annotated with opinion mining data; (b) services available on-line that use opinion mining for various goals. The final list consisted of 5 research datasets and 8 services, for each we analysed the data that is exposed and provided Marl mappings. Next, we calculated the coverage as *an amount of properties that were possible to describe with Marl over the total amount of data properties used in a dataset.* In the first experiment we considered all the dataset fields and the average coverage we got was 64%. However, it has to be noted that the individual characteristics of the data sources varied a lot. As a result very often the metadata uncovered by the ontology was proprietary information specific only for a single deployment (see Fig. 5.8).

According to ontology design goals presented by Noy et al. Noy and McGuinness (2001) one of the characteristics of good design is not to cover the very individual elements of datasets. Therefore, after removing the dataset fields that did not repeat at least once, we ran the experiment again and got the average coverage of 76%. The results of the experiments per dataset have been summarized in Table 5.2.

In the second part of our experiments we tested the capabilities of Marl to be used in context of Semantic Web queries. We started with creating a list of competency questions and tested them against the ontology (a total of 20 query templates where created). Later, for a more practical approach, we extracted small parts of datasets mapped in the previous experiment and used them to check with software prototypes if the queries involving Marl deliver anticipated results with different kinds of search. On this stage the problem that we encountered in most cases was insufficient data to create rich links to expose true power of Marl. Ultimately, for Internet wide data, we did our tests in the context of movie reviews and filtering opinions by polarity from different sites such as Tweetsentiments, IMDB (via Cornell

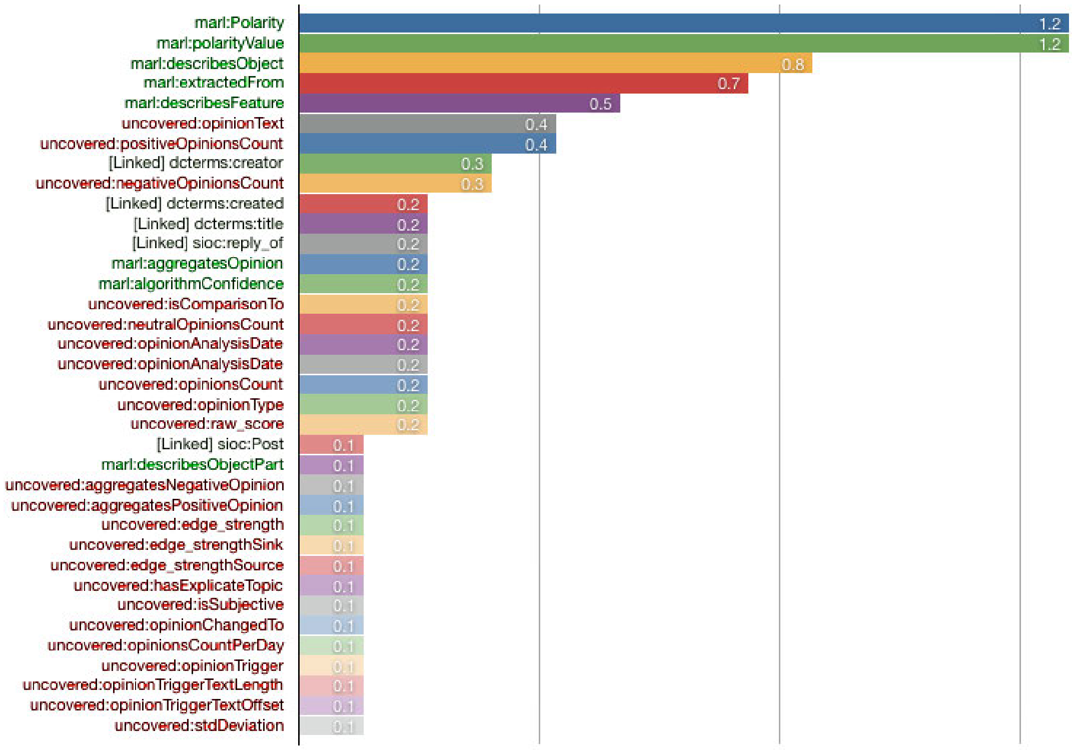


Figure 5.8: Property usage in mappings for datasets and services (average amount a property was used in a data source)

dataset Pang and Lee (2004)) and Swotti in a single query. We repeated this both for references to movies expressed as literals and for the entity search (with DBpedia entity references). In both cases we used Sindice search engine as back-end for the demonstration. Finally, for tests of metadata search in closed private environments we have setup a local SPARQL endpoint and used the OPAL opinion mining module in conjunction with technologies from

Table 5.2: Marl ontology coverage experiment results, considering all dataset fields (exp1) and after removing fields that did not repeat at least once (exp2).

Dataset/service name	**#covered/#total**		**coverage**	
	exp1	**exp2**	**exp1**	**exp2**
Congressional speech data (Thomas et al., 2006)	7 / 12	7 / 7	58%	100%
Movie Review Data (Pang and Lee, 2004)	3 / 4	3 / 3	75%	100%
Customer Review Data (Hu and Liu, 2004)	5 / 9	5 / 6	56%	83%
French Newspaper Articles (Bestgen et al., 2004)	1 / 3	1 / 2	33%	50%
Multi-Domain Sentiment Dataset (Blitzer et al., 2007)	4 / 4	4 / 4	100%	100%
Swotti (www.swotti.com)	9 / 13	9 / 13	69%	69%
Tweetsentiments (www.tweetsentiments.com)	6 / 11	6 / 11	55%	55%
Mombo (www.mombo.com)	10 / 16	10 / 12	63%	83%
Opinion Crawl (www.opinioncrawl.com)	4 / 9	5 / 9	44%	44%
OPAL (www.gi2mo.org/apps/opal/)	8 / 11	8 / 11	73%	73%
OPfine (www.jane16.com)	6 / 6	6 / 6	100%	100%
Evri (www.evri.com)	3 / 5	3 / 5	60%	60%
Opendover (www.opendover.nl)	4 / 9	4 / 6	44%	67%
Average	**5 / 8**	**5 / 7**	**63%**	**76%**

Gi2MO project Westerski et al. (2010) to extract opinions from independent Idea Management Systems and visualise them together. The additional challenge was that the two systems had data in different languages: one Spanish and the other English. As an outcome, the opinion mining algorithm enabled us to leverage the multilingual instances to the same level but ultimately the Marl ontology in conjunction with other Semantic Web vocabularies worked as an enabler to integrate the systems and run queries over the data to aggregate all information in a single view (e.g. show all ideas with community opinions and compare aggregated opinion scores, or compare the amount of positively received ideas by idea categories etc.). All together the query experiments proved that the ontology is capable in answering all the formulated questions in test scenarios of: movie opinions, product opinions, Idea Management Systems. A common problem, that confirmed the test results of coverage experiment, was that many queries expected the direct link to text fragment of the opinion - which is not facilitated by Marl. An example of a query constructed for data serialized with Marl v0.1 during our experiments can be seen at figure 5.9.

```
PREFIX gi2mo: <http://purl.org/gi2mo/ns#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX marl: <http://purl.org/marl/ns#>
SELECT ?idea_uri
       COUNT(?negative_opinion_uri) AS ?negative_opinions
       COUNT(?positive_opinion_uri) AS ?positive_opinions
FROM <http://etsit.gi2mo.org/etsit_ideas_en.rdf>
WHERE {
{
    ?idea_uri a gi2mo:Idea .
    ?idea_uri gi2mo:hasComment ?comment_uri .
            ?positive_opinion_uri marl:extractedFrom ?comment_uri .
            ?positive_opinion_uri marl:hasPolarity marl:Positive .
}
UNION {
    ?idea_uri a gi2mo:Idea .
    ?idea_uri gi2mo:hasComment ?comment_uri .
            ?negative_opinion_uri marl:extractedFrom ?comment_uri .
            ?negative_opinion_uri marl:hasPolarity marl:Negative .
} } GROUP BY ?idea_uri
```

Figure 5.9: A sample SPARQL query for "Show amount of positive and negative opinions for all ideas submitted into the Idea Management System". The source data was serialized using Marl v0.1 therefore aggregation operator was used to go around the lack of opinion count information.

Concluding both of the experiments, we used the acquired knowledge to produce a second iteration of the ontology (Marl 0.2) and included the new properties that according to our tests were uncovered and often used by other datasets; or were expected as output for search queries. After repeating the coverage experiments with the new version of the ontology we got 79% coverage for experiment 1 (all dataset fields considered) and 94% coverage for experiment 2 (dataset fields that did not repeat at least one time across different sources ignored).

5.6.2 Leveraging Opinion Metadata to Idea Metrics

Having evaluated the ontological model against a number of different systems on the Web, we move to more focused tests and evaluate our proposal against the particular use case of idea assessment in Idea Management Systems.

We propose the use opinion mining to calculate a new metric that aggregates sentiment of comments attached to an idea. Later, apply Marl ontology in conjugation with the Gi2MO Ontology concepts and utilize both to describe the community generated content. In the following section, we describe methodology for generating the said metric and proceed with the evaluation tests to verify the usefulness of opinion mining in Idea Management in general.

Hypothesis

As shown in the previous sections, the value of mining opinions from comments can be studied from many different angles and its impact can differ depending on how the mined information is applied in practice. In our evaluation, we focus on two main hypotheses that relate opinion mining to Idea Management Systems:

H1. Organizations choose to implement ideas based on opinions of the community.

H2. Community opinions are not fully reflected by the currently used community activity metrics.

With **H1** we put forward a hypothesis that idea reviewers and managers of the idea competitions investigate not only the summary statistics like idea ratings but read the comments and those comments influence the final decisions that managers make in regard which ideas are implemented and which not.

With **H2** we suggest that the commonly used metrics in Idea Management Systems are not fully accurate about the opinions of the community regarding a certain idea. We hypothesise that evaluating opinions submitted in comments can deliver new knowledge that could potentially have additional impact on the final idea selections.

Evaluation setting and measures

Taken into account both of the stated hypotheses, we propose to evaluate if they are indeed supported by evidence through calculating a single metric for every individual idea based on the following algorithm:

- calculate the opinion rating separately of every comment attached to the idea
- calculate the idea rating as a sum of ratings of its comments

We applied the above methodology in practice using the dataset of Ubuntu Brainstorm (Ubuntu Brainstorm, 2012) - an Idea Management System instance run by Canonical to collect ideas for improving their Ubuntu Linux distribution (see details in Table 5.3).

Table 5.3: Ubuntu Brainstorm dataset statistics

Metric	**Metric Value**
Idea number	21690
Comments number	133090
Users number	10062
Implemented Ideas number	576
Amount of Votes cast	2608917

The distinctive feature of Ubuntu Brainstorm dataset, in comparison to other Idea Management data, is the possibility to submit new solutions for already existing ideas (see Fig. 5.10). The first solution is provided by the author of the idea, while the following solutions can be submitted by any member of the community. Each solution can be individually voted on, however the comments for all solutions are submitted in the same space, only referencing the root idea.

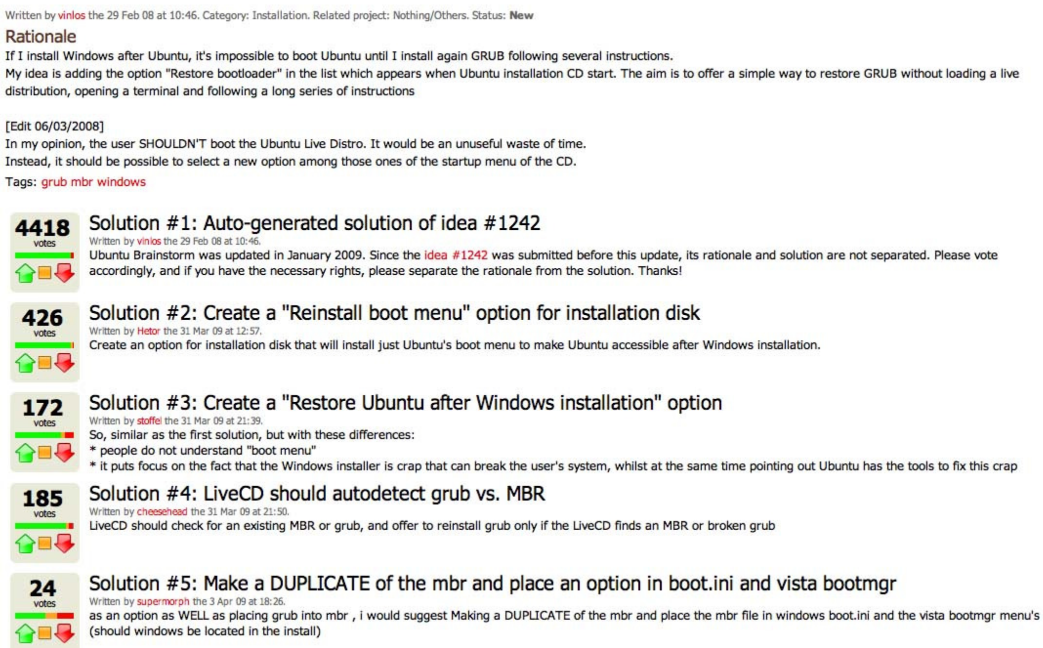


Figure 5.10: A sample **single** idea with solutions (Ubuntu Brainstorm Ubuntu Brainstorm (2012)).

In our evaluation, in order to calculate the opinion rating per each comment, we constructed a simple prototype (OPAL (OPAL, 2012; Rico, 2010)) that sums the word ratings of all words in the comment text (the word ratings were obtained using SentiWordNet library (Baccianella et al., 2010)). We measured the performance of such solution by manually annotating 50 idea comments (with positive, negative or neutral ratings) and compared the results with the automatic annotation done with OPAL. For this task, we used measures typically applied in opinion mining research: precision (Eq. 5.1), recall (Eq. 5.2) and their harmonic mean, i.e. F-measure (Eq. 5.3).

$$precision = \frac{|\, correct\ automatic\ annotations\, |}{|\, all\ automatic\ annotations\, |} \tag{5.1}$$

$$recall = \frac{|\, correct\ automatic\ annotations\, |}{|\, all\ evaluation\ set\ idea\ annotations\, |} \tag{5.2}$$

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5.3}$$

The proposed solution achieved 66% precision, 67% recall and 67% F-measure. In relation to the earlier mentioned related work, this places our prototype between keyword and pattern based solutions. Based on the state of the art analysis in opinion mining, this result could be further improved e.g. by applying enhancements for expressions typical to Ubuntu community. However, since providing the best possible algorithm is not the goal of this evaluation, we considered the obtained result sufficient and proceeded with the main experiments on relating opinion rating to other Idea Management System metrics.

Using the OPAL prototype, we automatically annotated comments for 50 ideas: 10 implemented ideas, 10 highest rated (with up/down rating), 10 lowest rated, 10 top commented, and 10 least commented (but having at least 1 comment). All together, we obtained opinion ratings for 1796 comments which were used to calculate the opinion ratings for the aforementioned 50 ideas.

Including the legacy metrics, we used the dataset to calculate the following information:

- comment count - amount of comments attached to an idea
- solution count - amount of solutions submitted for an idea
- maximal solution up/down rating - the highest rating of a solution attached to an idea
- minimal solution up/down rating - the lowest rating of a solution attached to an idea
- average solution up/down rating - average of ratings of all solutions attached to an idea
- idea age - time (in days) since idea was submitted until the day experiment was conducted
- opinion rating - rating based on opinion mining algorithm run over comments attached to an idea
- idea adoption - indicates if an idea was implemented (equals 1) or not (equals 0).

To verify hypothesis **H1**, we analysed the impact of all legacy metrics on idea adoption (if an idea was implemented or not) and compared with the results for our opinion metric. To address hypothesis **H2** we analysed various correlations of our opinion metric with a number of currently utilised metrics in Idea Management Systems: community rating, comment count. etc. The results of those experiments are presented in the next section.

Results

In case of hypothesis **H1**, for each of the aforementioned metrics we measured and compared the bivariate correlation with idea adoption to check if any of the metrics has a determining impact on whether ideas have been ultimately selected for implementation or rejected (see Table 5.4).

Table 5.4: Bivariate correlations of metrics with Idea Adoption

Metric name	**Correlation**
Comment count	0.03
Solution count	0.04
Max. solution rating	0.3
Min. solution rating	0.24
Avg. solution rating	0.37
Idea age	0.12
Opinion rating	**0.04**

The results show that correlation of opinion metric is one of the lowest. This suggests that reviewers and decision makers of the Ubuntu Brainstorm system did not pay attention to user opinions expressed in the comments. Such results indicate that hypothesis **H1** is not supported.

In the second activity to verify hypothesis **H2** we took the same metrics but measured the correlations between each other to see if opinion metric delivers new information or has the same behaviour as some other metric (see Table 5.5).

Table 5.5: Bivariate correlations of metrics with each other (including opinion rating).

		1	**2**	**3**	**4**	**5**	**6**	**7**
1	Comment count	1	0.37	0.68	0.11	0.26	0.02	**0.28**
2	Solution count	x	1	0.28	-0.32	-0.21	-0.65	**-0.08**
3	Max. solution rating	x	x	1	0.32	0.51	0.32	**0.25**
4	Min. solution rating	x	x	x	1	0.95	0.26	**0.38**
5	Avg. solution rating	x	x	x	x	1	0.26	**0.41**
6	Idea age	x	x	x	x	x	1	**0.19**
7	Opinion rating	x	x	x	x	x	x	1

The obtained results show that opinion metric has a medium positive correlation with average rating, however weak correlation with max. rating and medium correlation with min. rating. Taking into account this result, we can make a statement that in the particular settings of Ubuntu Brainstorm good ratings of idea solutions do not reflect the community opinions, while poor ratings usually go in line with bad comments. To confirm this observation, we also investigated the raw data of the max. and min. solution rating metrics. Figure 5.11 shows that the behaviour of min. rating is similar to opinion rating in the area of solutions with lowest rating (2), while in other areas the similarities are much harder to observe (especially in top voted area (1) where some of the top ideas have lowest opinion rating of the entire sample).

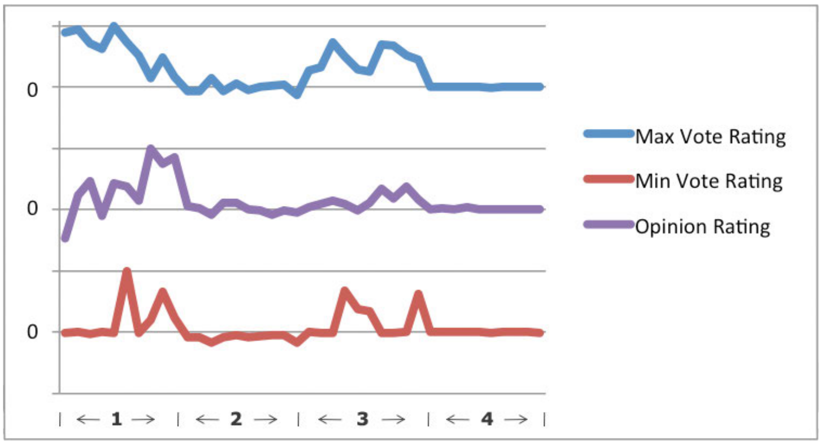


Figure 5.11: Comparison of normalized rating values for ideas from the experiment. (1) - top voted solutions, (2) - lowest voted, (3) - implemented, (4) - least commented

Moreover, this criticism of most down ranked ideas or lack of support for top ranked ideas should not be understood in terms of quantity of opinions (due to weak correlation between opinion rating and comment count) but strength and verboseness of sentiment expressions in the comments. Taking into account those results and observing the correlation of opinion rating metric with the remaining legacy metrics we can conclude that the new metric does not duplicate the behaviour of other Idea Management indicators. Therefore, hypothesis **H2** can be considered as supported.

5.7 Summary

In this chapter we have proposed an ontology for describing opinions and related assessment data obtained via opinion mining algorithms. The formalization has been discussed in the context of Web systems in general but more importantly in terms of appliance in Idea Management Systems.

Based on the presented opinion ontology as well as Gi2MO ontology from the previous chapter, we have created a new metric for idea assessment called *idea opinion rating*. The metric based on data from opinion mining process aggregates the sentiments of comments attached to ideas and enables to analyse them independently of IMS domain or language.

The goal of our contribution is to achieve comparison of idea opinions across heterogeneous systems or independent IMS deployments. This chapter has presented the required theoretical background and proposed a solution to achieve the aforementioned goal: (1) though opinion ontology as enabler for data interoperability and (2) by using opinion mining as a tool that delivers necessary data for analysing community feedback and serializing it with the ontology.

Furthermore, in this chapter, we evaluated a number of aspects related to the deployment of the ontology in real case studies. Those evaluation activities have shown that **79% of metadata has been covered** by our ontology when applying it for data generated by opinion analysis solutions from either commerce or research areas. Furthermore, we discovered that the coverage increased up to **94%** if we consider only information that repeats across different solutions at least once.

Moving towards Idea Management Systems appliance, we asked and answered crucial questions regarding the adoption of the proposed solution. More precisely, we hypothesised that decision makers in Idea Management Systems pay special attention to the community opinions expressed in comments and that therefore opinion mining techniques could deliver a new valuable tool to summarize this information and be used for rating ideas.

Concluding our investigation we can state that measuring community sentiments related to ideas through comments does indeed deliver a new supplementary tool for judgement of ideas performance, however not fully distinct from the already present idea rating metric. The strongest detected correlation with average solution rating suggests that the up/down rating does partially reflect what users write in the comments. The interesting observation of our analysis is that very well rated solutions of ideas often attract high amount of criticism as well as positive feedback (low correlations of the top and lowest ratings with opinion rating). This leads to a conclusion that the up/down ratings does not fully reflect the intensity of emotions in user opinions in idea comments and it is worthwhile to use opinion mining to get the full picture of community sentiments.

In addition, our study has shown that the impact of user opinions is very low on final idea selections in the systems used so far. This finding proves that idea reviewers and contest managers responsible for selecting ideas to implement do not make their choice based on opinion of the community. According to case studies this fact has been attributed so far to overwhelming amount of comments difficult to analyse in reasonable time - an issue that is resolved by our solution through use of opinion mining algorithms.

Chapter 6

Idea Characteristics Model for Idea Management System

One of the problems of Idea Management Systems is the difficulty to accurately depict the distinctive features of ideas in a rapid manner and use them for judgement of proposed innovations. The research described in this chapter aims to solve this problem by introducing annotation of ideas with a domain independent taxonomy that describes various characteristics of ideas. Furthermore, we propose to transform such annotations into new metrics that allow the comparison of ideas or entire idea datasets.

In the chapter, firstly, we introduce a taxonomy based on the various concepts of innovation research that we align with the reality of Idea Management Systems. We describe, how the well known concepts of past models apply to Idea Management and propose a categroization of those concepts in a hierarchy that would be applicable for annotation of ideas.

Secondly, we propose to leverage the annotations made with this taxonomy and automatically generate metrics that would enable the idea reviewer to assess various elements of an idea or entire idea datasets.

The contributions presented in the chapter are:

- Gi2MO Types taxonomy for describing non-domain characteristics of ideas
- Gi2MO Types metrics generated from taxonomy annotations to summerize the characteristics of idea datasets

6.1 Introduction

Much before the age of information scientists have attempted to understand how innovation works and developed numerous models for describing those mechanics. As it can be seen throughout the progress of research on Innovation Management, the way creative processes in organizations were perceived has changed along with evolution in production methods (Utterback, 1987), new market types (Ortt and van der Duin, 2008), switch from manufacturing to service based economies (Susman et al., 2006) etc. Numerous case studies have shown that the established theoretical models help practitioners to understand organizational innovation and build strategies for management of innovation (Warren and Susman, 2002; Hobday, 2005).

On the other hand, Idea Management Systems in majority have been developed and moved forward by the industry rather than the academia. As a result, most of the studies available focus on appliance of metrics existing in other domains that would give immediate results for IMSes rather than contributing to long term goals of improving the understanding of mechanics that rule Idea Management in particular. In this chapter, we refer to the approach already successfully utilized in innovation management and study the specifics of computer supported community innovation in contrast to previous theoretical models. Our goal is to understand what are the elements that could be perceived as characteristics of ideas in Idea Management Systems and how could those aid in assessment of innovation leading to selection of best ideas for implementation.

In particular, based on the analysis of existing innovation management models and experiments with data from Idea Management Systems, we propose a taxonomy that characterises an idea in an Idea Management System. Furthermore, we continue the line of inquiry from previous chapters and attempt to apply the discovered taxonomy for idea assessment and idea comparison.

We focus on challenges that arise when trying to quantify the value of information contained in ideas and its impact on innovation in the enterprise. As mentioned in the previous chapters and noted by the literature (Hrastinski et al., 2010), in the contemporary solutions those problems of idea assessment are approached by: 1) the use of a handful of automatically generated yet very simple community statistics; 2) expert reviews that require a considerable amount of knowledge and impose serious time constraints thus increase the costs of the entire idea management process.

In this chapter, we present a solution for idea assessment that combines the advantages of those two cases mentioned by Harstinski: rapid generation of metrics that require little expert knowledge yet offer more diversity and versatility than the current community metrics. In particular, we deliver a methodology for obtaining the metrics via analysis of idea annotations made with a domain independent taxonomy that expresses idea characteristics. The focus is to show that the proposed set of metrics can be applied to Idea Management Systems in a meaningful way that would allow to capture the distinctive features of ideas and compare entire idea datasets.

The chapter is structured as follows: firstly we summarize the past research achievements in terms of metric generation for Idea Management Systems as well as other kinds of computer-supported cooperative work systems (see Sec. 6.2). Additionally, in the same section, we discuss research on capturing the meaning of innovation in general and show how it influenced our work. Afterwards, we introduce our contribution in a form of a taxonomy for describing idea characteristics and present in more detail the theoretical grounding by referring to particular innovation models (see Sec. 6.3). Finally, we show how to utilize the proposed taxonomy in practice of Idea Management Systems by transforming the idea annotations into metrics that characterise entire systems (see Sec. 6.4). At the end, we present the results of our experiments that test the usage of the taxonomy for annotation (see Sec. 6.5.1) as well as verify the performance of metrics generated from those annotations in relation to the contemporary parameters of Idea Management Systems (see Sec. 6.5.2).

6.2 Related Work

Having a significant presence in the industry, Idea Management Systems have also been investigated by the academia in search of problems and patterns that emerge when using this class of systems in an organization (e.g. (Bailey and Horvitz, 2010)). In our case, the investigative work on idea assessment is of special interest. Within this area, Hrastinski et al. (Hrastinski et al., 2010) surveyed a number of selected products and pointed out that the current commercial systems employ rather simple idea evaluation methods most often being analysis of community statistics (number of ideas per user, community voting results, number of idea comments etc.) or internal business metrics that are delivered by designated experts (e.g. return of investment, market value etc.). Both of those approaches have been evaluated by Gangi et al. (Gangi and Wasko, 2009) and compared to conclude that in practice none of current methods have a significant impact on which ideas are being implemented by the organizations. Following those conclusions, there have been various approaches that attempted to find a solution to time efficient and effective automatic idea assessment problem e.g. with prediction markets (Bothos et al., 2008), by applying problem solving algorithms (Adamides and Karacapilidis, 2006), calculating metrics for the quality of management (Conn et al., 2009) or using data from other enterprise systems to automatically assess ideas (Ning et al., 2006; Westerski and Iglesias, 2011). All those solutions are based on the notion of reusing existing data whereas the approach proposed by us claims that there is a necessity to attach some additional data to ideas in order to improve assessment and selection phases of the idea life cycle.

Apart of Idea Management Systems domain research, there has been a huge number of works that attempt to analyse characteristics of discussions or content created by communities in a collaborative way e.g. (Stromer-Galley, 2007; Alexandru Spatariu and Bendixen, 2004; Nisbet, 2004). Among those, Perey (Perey, 2008) describes a necessity to go beyond simple metrics that count number of interactions with the system in time. However, in contrast to us, in his work Perey focused only on measuring characteristic features of users and their interactions with each other rather than metrics on content that those users create. Klein (Klein, 2012) notices similar problems with regard to difficulties in assessment and browsing community submissions but he attempts to find a remedy through experimenting with novel system interaction methods, in particular argumentation tools (Kirschner et al., 2002). While this approach is different to ours it shows an interesting alternative not only for generation of new metrics but altering the entire philosophy of Idea Management System front-end which in turn can create new opportunities in the back-end.

Outside of the aforementioned areas of computer science, there has been a large number of works that investigate ways of categorising innovation and attempt to quantify it. While preparing for the creation of the taxonomy and validating it afterwards we analysed those models as a reference. We started from the very origins of Schumpeter's innovation theories (Schumpeter, 1934) and finished with the contemporary work on the topic. The selection of models that we have analysed as related was based on studies from a number of works that attempt to revise the state of the art on innovation models (Eris and Saatcioglu, 2006; Popadiuka and Choo, 2006; Damanpour and Gopalakrishnan, 1998; Chuang et al., 2010; Garcia and Calantone, 2002). During our work, we prepared a taxonomy model that included the various perceptions of innovation from those models. The preliminary experiments with this taxonomy version have shown that most reviewers (i.e. annotators asked to use the taxonomy) did not know how to apply the terms. Consequently we made a choice to propose the taxonomy, as described in the next section, only based on the analysis of idea content from Idea Management instances that we gathered.

6.3 A Domain Independent Taxonomy for Idea Annotation

In this Section, we introduce a taxonomy that captures the characteristics of ideas published in an Idea Management System. In our methodology, the taxonomy is used to annotate ideas with terms that later serve as a base for calculating metrics. The choice of terms that establish the taxonomy is based on our experience with different kinds of Idea Management datasets gathered during the course of Gi2MO project (Westerski et al., 2010). This initiative aimed to enrich contemporary Idea Management Systems with an extensive use of metadata according to the Semantic Web principles. During the project we gathered various datasets ranging from ideas for technology to products for mass consumer (see Sec. 6.5.2 for detailed dataset description). Based on the analysis of those datasets, we enumerated the key characteristics of ideas that could be inferred from the idea text and organized them into a hierarchy. The taxonomy model that we propose can be summarized by the following hypothesis:

"Every idea that was **proposed** has been **triggered** by a particular experience and describes a certain **innovation** put in context of application in a given **object**."

"Proposed", "triggered", "innovation" and "object" represent the four main characteristics of an idea that we established as the root for further taxonomy terms which detail a particular aspect of the idea characteristics (see Fig. 6.1).

The **trigger** branch details aspects related to experiences that influenced creation of the idea. While analysing the ideas gathered in different Idea Management Systems, we noticed that users often tend to mention how they came up with a particular innovation in order to

Figure 6.1: Gi2MO Types taxonomy as proposed in Section 6.3.

justify their claims. Similarly, innovation models of Kelly and Kranzberg (Kelly et al., 1975), Usher (Usher, 1954), Myers (Myers and Marquis, 1969), Hughes (Hughes, 1975) as well as contemporary research (Narasimhalu, 2005) notice the existence of various causes that lead to idea generation. In particular, innovation is described as being a result of recognition of a problem, need for changes or recognition of technical feasibility or demand. Those different types of triggering experiences are referred by us in the trigger branch as **Observation Types**. Additionally, Usher (Usher, 1954) has shown that innovation is not only triggered by experiences related to a personal observation but also events that influence the innovator and lead to an act of insight. We relate to this by characterising the type of event that led to the idea with **Creativity Origin** classification and by identifying the connection between the triggering experience and the object that is innovated (**Associated Object**).

The **innovation** branch relates the idea proposal to the reality of the enterprise and the state of the Idea Management facility. As such, the assessment made by annotators that use this taxonomy branch goes to the origins of the very understanding of innovation in enterprise discussed for tens of years since the original contributions by Schumpeter (Schumpeter, 1934). The variety of models proposed since then show that interpretation of innovation can be extended in many different directions depending on the context and goals. In our work, we took into account the previous models (see Sec. 6.2), however we narrowed down the amount of terms based on experiences with idea datasets and inferences that could be made based on idea text. As a result, we noticed three key aspects that were mentioned by innovators and reviewers in Idea Management Systems: relations to other ideas previously posted in the system or innovations introduced by the organisation (**Dependence**), descriptions of usefulness of the idea for a particular group (**Target Audience**), references to idea originality with respect to current state of organisation or entire market (**Originality**).

The **object** branch focuses on describing the entity that is being innovated and the changes proposed in relation to the original. Apart of the classical distinction between products and processes (Abernathy and Utterback, 1978; Utterback and Abernathy, 1975; Adner and Levinthal, 2001), we also recognize service innovation as it has been advocated by a number of researchers that studied innovation past the time when manufacturing was the dominant element of economies (Susman et al., 2006; Drejer, 2004; Rubalcaba et al., 2010). With regard to classifying how those entities are transformed by ideas, Gilfillan (Gilfillan, 1935) noticed that innovation is often a chain of small improvements, modifications and additions rather than a single act of brilliance of one innovator. We relate to this observation by classifying the type of changes proposed for an object (**Structure**), as well as recognizing if the introduced change is a reoccurring innovation from some past iteration or a completely new proposal (**History Relationship**). Additionally, following the observations of Abernathy and Clark on firm competence (Abernathy and Clark, 1985), we noticed that proposed changes in the analysed ideas may have a different impact on the current design of the product as well as associated product knowledge. Some ideas propose adding or removing elements in an existing design while others introduce a totally new product. Those kind of differences are classified in the **Offering Placement** sub-tree. Finally, following research in the engineering design (Jarratt et al., 2010) we notice that the proposed changes in the object and their implementation may affect existing related products in a different way. A report by AberdeenGroup (Brown, 2006) shows that analysis of those kind of changes is of crucial importance for organizations when making decisions on adopting certain innovations or not.

The final **proposal type** branch is connected to the way the text of an idea has been written. The analysis of idea datasets has shown that not all users express their requests for innovation in the same way. Some of the ideas differ on the level of completeness of the description, while others vary in the way the entire idea has been formulated. We perceive those differences as lack or presence in description of selected innovation process stages such as problem definition or solution (as defined in many innovation management models, e.g. (Kelly et al., 1975; Baker and Freeland, 1972; Usher, 1954)). The goal of the Proposal Type taxonomy branch is to capture those differences and later allow the idea reviewers or moderators of idea contests to filter out certain proposal types that are not wanted at all.

6.4 Calculation of Metrics Based on Idea Annotations

The taxonomy presented in the previous section enables to identify the characteristics for individual ideas. Nevertheless, in big datasets the amount of idea annotations made using the taxonomy terms can be overwhelming and therefore difficult to analyse and interpret for a practitioner. Since our goal is to facilitate idea dataset comparison, we propose to summarize the annotations and describe their meaning for the entire dataset.

In particular, in the next step of our methodology, we propose to utilize the taxonomy described in the previous section to annotate ideas and afterwards produce metrics based on the quantitative analysis of the annotations.

The said methodology for generating metrics includes:

- assuming a certain interpretation of terms in the taxonomy and assigning a metric to each taxonomy sub-tree
- calculating the metric value for every idea individually based on idea annotations with the taxonomy terms
- calculating the metric value for entire dataset as a median of metric value of all ideas from the dataset
- supplementing every calculated dataset metric with the diversity measure of annotations per each taxonomy sub-tree (using information entropy)

In the first step, we define 14 metrics (see Table 6.1) corresponding to different branches of taxonomy tree and an additional single metric (Idea Completeness) that measures how many branches of the taxonomy are used for describing an idea. The metrics that relate to particular sub-trees have an ordinal scale based on the particular interpretation of term order in the respective taxonomy sub-tree. The explanation of the approach taken for each sub-tree can be observed in table 6.1, while the results of applying the median for calculations of metrics for entire datasets can be seen in the next section when we report on evaluation results (see Sec. 6.5.2).

The aforementioned metrics summarise the information expressed by the annotations and transform it by providing a certain interpretation. However, the problem that arises is that some of the information is lost in comparison to term frequency analysis. In particular, one cannot say what is the diversity of terms just looking at the metric (e.g. Idea Originality is 0.5 if half of the ideas are *New* and half have *no innovation* but also when all ideas are tagged as *Incremental*). Therefore, apart of the metrics based on interpretation of the taxonomy terms we also propose to measure the diversity of terms in the annotated datasets. Whereas

Table 6.1: List of metrics based on Gi2MO Types taxonomy.

Taxonomy Branch	**Metric**	**Explanation**	**Scale**
Trigger/ Observation Type	Trigger Experience Completeness	How complete was the experience with the triggering object	1 = Potential Opportunity 0.5 = Faulty Experience 0 = Lack Of Feature
Trigger/ Creativity Origin	Trigger Situational Dependence	How much is the trigger dependence on occurrence of some particular event	1 = Event 0 = Object Interaction
Trigger/ Associated Object	Trigger Relatedness	How closely related are the object of innovation and the object that triggered the idea	1 = Object Of Innovation 0 = Other Object
Innovation/ Dependence	Idea Dependability	How much is the idea connected to other ideas (how much does it change influence other)	1 = (Proceeds, Follows, Duplicates, Is Part Of) 0.5 = (Encapsulates, Excludes) 0 = None
Innovation/ Target Audience	Idea Adaptiveness	How much are the ideas meant for new markets and how much for existing ones	1 = Existing Audience 0 = New Audience
Innovation/ Originality	Idea Originality	How original is the idea	1 = New 0.5 = Incremental 0 = None
Innovation/ Related to	Idea Originality Scope	How far does the originiality of the idea reach. Is it only the particular element of the organisation, entire organisation or the entire market ?	1 = Market 0.5 = Organization 0 = Innovation Proposals
Proposal Type	Community Cooperativeness	How well do users formulate and communicate their ideas and the implementation of ideas.	1 = Solution 0.75 = Suggestion 0.25 = Request 0 = Issue Report
Object/ History Relationship	Implementation Freshness	How much are the new ideas related to former products	1 = Evolutionary 0 = Regressive
Object/ Type	Implementation Integrity	How tangible is the object of innovation	1 = Product 0.5 = Service 0 = Process
Object/ Type/ Product	Implementation Applicability Scope	How broad is the application of the idea (measured by product offering scope)	1 = Product Type 0.5 = Product Line 0 = Specific Product
Object/ Offering Placement	Implementation Constructiveness	How much ideas are creating new products or just improving the old products	1 = New 0 = Existing
Object/ Structure	Implementation Scope	How much the ideas want to modify of the current state	1 = Complete 0.5 = Element 0 = Characteristic
Object/ Relationships	Implementation Dependability	How much does the introduction of innovation impact other products	1 = (Part Of, Composed Of) 0 = Complementary With

the first set of metrics flattens the perception of terms to a common level, in the second case we sought for a solution that will enable the reviewer of the idea contests to assess the diversity of terms and judge how much ideas are similar to each other under certain criteria. As a result we have chosen information entropy as a statistic that would fulfil this need (see Eq. 7.1).

$$E(tb_x) = -\sum_{i=1}^{n} p(i) \log_2 p(i) \tag{6.1}$$

When applied to our case, $E(tb_x)$ is the diversity for the tb_x taxonomy branch; $p(i)$ is the frequency of annotations with the certain (i) term combination; while n is the number of all such combinations in a given taxonomy branch (we assume that all combinations are possible, e.g. an idea can describe product and service innovations). Based on the above, we propose to calculate term diversity understood in such way for every taxonomy branch and for each dataset. As an outcome, our hypothesis is that the entropy should allow to decide how similar to each other are ideas of different datasets. The results of experiments that evaluate this hypothesis in practice and calculate entropy for particular datasets are presented in next section.

6.5 Evaluation

In order to test our hypothesis about the taxonomy and the formulated metrics, we performed a series of experiments to cover the entire presented methodology. Firstly, we studied how does the taxonomy perform when annotations are applied by groups of people of different sizes and different expertise levels, as well as how does manual annotation compare to the automated approach (see Sec. 6.5.1). Further, having obtained satisfactory results with the annotation experiments, we evaluated the second step of the methodology that delivers the actual metrics. The performed experiments aimed to evaluate the feasibility to use the metrics for comparison of datasets (see Sec. 6.5.2) as well as usage of metrics for selection process of best ideas by measuring their correlation with some of the currently used statistics (see Sec. 6.5.2).

6.5.1 Annotation of Data in Idea Management Systems

Our ultimate desire was to construct a taxonomy that could be complex enough to cover all the idea characteristics but at the same time suitable for usage by non-experts or with automatic annotation algorithms. During the experiments we realised that this might be a difficult task to achieve due to some characteristics being very detached from the sole idea text. Therefore. we downsized the taxonomy in different ways to find the set of its elements that would fit the desired goals best. We present the results of a number of experiments that compare performance of annotators when using the full taxonomy as well as certain parts of taxonomy for: manual annotation of ideas (see Sec. 6.5.1) and automatic annotation (see Sec. 6.5.1).

Manual Annotation

In case of manual annotation, we measured the differences in annotations proposed by different people as well as the differences in annotations of the same annotator repeated in certain time intervals. For the first experiment, we arranged for 10 people to individually annotate the same set of 10 ideas. All participants of the experiment were computer scientists, aged 25-30 and working in the academia; none of them have had any previous contact with innovation theory or our taxonomy in specific. We did not limit the annotators in any way with regard to annotation rules (e.g. annotators could apply many different terms of the same branch to a single idea). Following the experiment, we measured the agreement of annotators as a percent of cases in which they either agreed to put the same annotation or agreed on not putting a certain annotation at all. As a result, we discovered the differences in decisions were quite considerable with only 34% of cases where annotators fully agreed and 5% of cases where no agreement could be reached at all (half of the annotators put an annotation and the other half did not).

Pursuing the same line of inquiry, we repeated the experiment inviting 5 innovation theory experts to provide the annotations for the same 10 ideas. In comparison with the first experiment, the innovation experts reached a consensus in 2% more cases than non-experts. For full results and comparison please see figure 6.2.

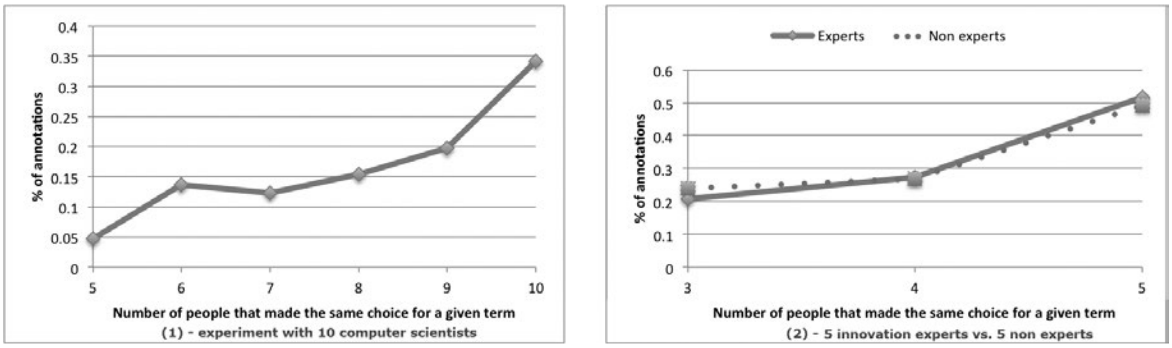


Figure 6.2: Evaluation results for manual annotation of 10 random ideas taken from IdeaStorm dataset.

In the second experiment referring to manual annotation, we asked the same person to annotate the same set of 100 ideas twice but in time distance of 3 months. The differences in that case were smaller than in the first experiment with 10 different annotators - 70% of annotations turned out the same in second annotation round as in the first. The worst result was noted for the Trigger/CreativityOrigin branch (only 48% of the same annotations and the only branch below 50%) due to the annotator categorising specifically types of triggering events very differently in consequent iterations of the experiment. On the other hand, the best results were achieved for Object type and Originality branches (82 % and 79 % respectably). Additionally, if we include in our calculations the cases of agreement on not marking a certain annotation, the final result for single annotator agreement rises to 90 %.

To relate those results to the theoretically grounded studies on measuring agreement we refer to the literature on metric reliability (Neuendorf, 2001; Krippendorff, 2004). In particular, we verified our results using two common indicators: Cohen kappa (Cohen, 1960) and Krippendorff alpha (Krippendorff, 1970) (preferred by some researchers as a better measure for metrics based on ordinal values (Hayes and Krippendorff, 2007)). We calculated those indicators for the final case of single annotator agreement that so far seemed to be our recommendation based on the best percentage agreement results. The obtained Cohen kappa was 0.613 as well as the Krippendorf alpha. As suggested by Landis and Koch (Landis and Koch, 1977) our results for those indicators can be considered as 'substantial agreement' in case of Cohen kappa and 'fair agreement' in case of Krippendorff alpha according to Taylor and Watkinson (Taylor and Watkinson, 2007).

Taken into account the above analysis, the presented manual annotation experiments show that the characterisation of innovation can be very subjective and relies in a great manner on the understanding of the topic by the annotator. This is in line with statement made by Garcia et al. (Garcia and Calantone, 2002) who presented a number of examples where the same innovation was labelled as radical or incremental depending on very small differences in understanding of those terms. In the case of annotation of ideas in Idea Management System the experiments showed that this is quite a valid problem if the annotations are made by a collective of reviewers (regardless if they are innovation experts or not). However, in the case of a single person doing all the work, the annotations are quite coherent, especially if limited to certain taxonomy branches.

Automatic Annotation

Regarding the automatic annotation, we tested whether it is possible to automatically extract features of the ideas that would suggest certain annotations without the need of pointing to keywords or using any additional knowledge base. Therefore, we experimented with a machine learning approach that was based on comparing similarity of idea texts. In particular, we used a a supervised machine learning approach and the weighted k-nearest neighbour (kNN) algorithm (Dasarathy, 1990). Our evaluation was done using a tool called GoNTogle that was previously proven to successfully work for automatic annotation of documents (Bikakis et al., 2010). In the implementation of GoNTogle the nearest neighbours are selected based on text similarity calculated by the document similarity algorithm of Lucene library (Lucene, 2012).

During our experiments, we used the annotated data corpus of 400 ideas from the previous manual annotation tests: 200 ideas were used as a training set and 200 for evaluation of the accuracy of the automatic annotation proposals. Taking into account the results of the manual annotation experiment, both of the datasets used during the automatic annotation were prepared by the same single annotator. In our first approach, we merged idea description with idea title into a single block of text and treated it as a document. For the analysis of results we used the typical measures for judgement of information retrieval effectiveness: precision (Eq. 6.2), recall (Eq. 6.3) and their harmonic mean, i.e. F-measure (Eq. 6.4).

$$precision = \frac{|\, correct\ automatic\ annotations \,|}{|\, all\ automatic\ annotations \,|} \tag{6.2}$$

$$recall = \frac{|\, correct\ automatic\ annotations \,|}{|\, all\ evaluation\ set\ idea\ annotations \,|} \tag{6.3}$$

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{6.4}$$

In the described case, we allowed annotation with all taxonomy terms and as a result the average F-measure was 0.46. To investigate further reasons for such performance, we analysed the results for particular elements of the taxonomy to discover which branches of the taxonomy could be fit for use with automatic annotation algorithms (see Fig. 6.3). In particular, we found that most promising elements of the taxonomy are located in the Trigger sub-tree.

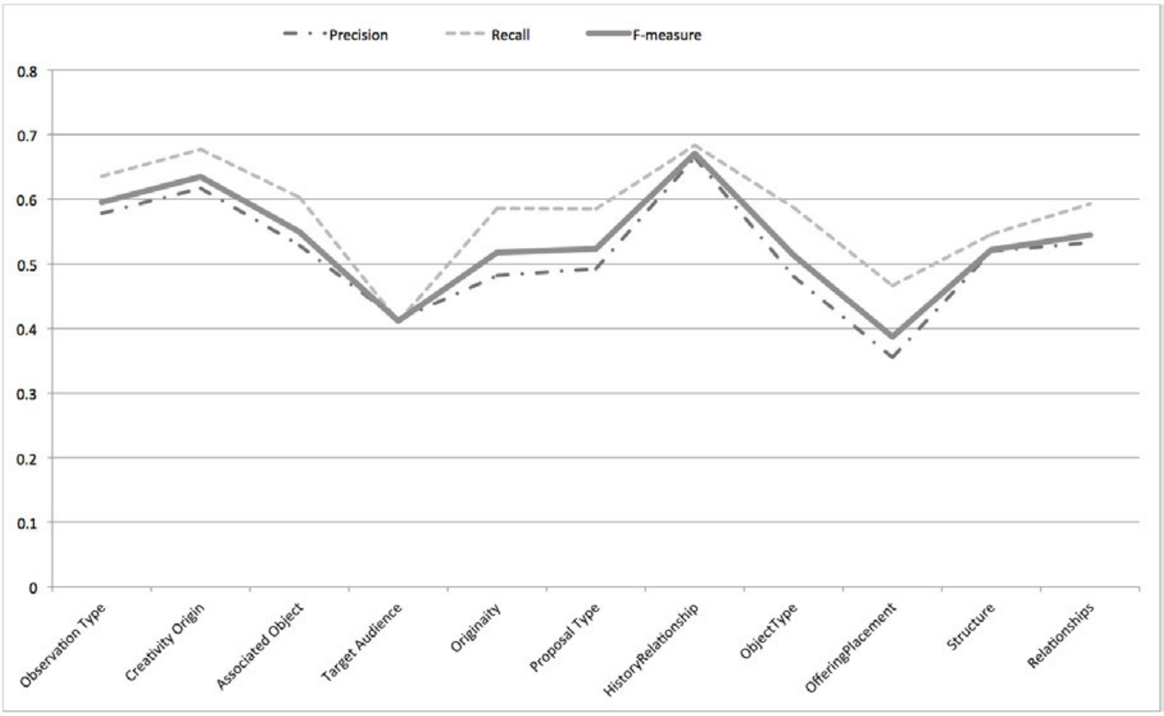


Figure 6.3: Evaluation results of automatic annotation split per taxonomy sub-tree (IdeaStorm dataset).

In addition, some elements of the Object branch also gave interesting results but we detected that in some cases the high variance of learning set had a big impact on those results (which was not the case for the Trigger branch as shown on Figure 6.4).

In an attempt to search for different options and improve the automatic annotation results, we took a few paths to change our process, most notably: split the ideas into paragraphs and treat them as separate documents during annotation time, increase the learning set size,

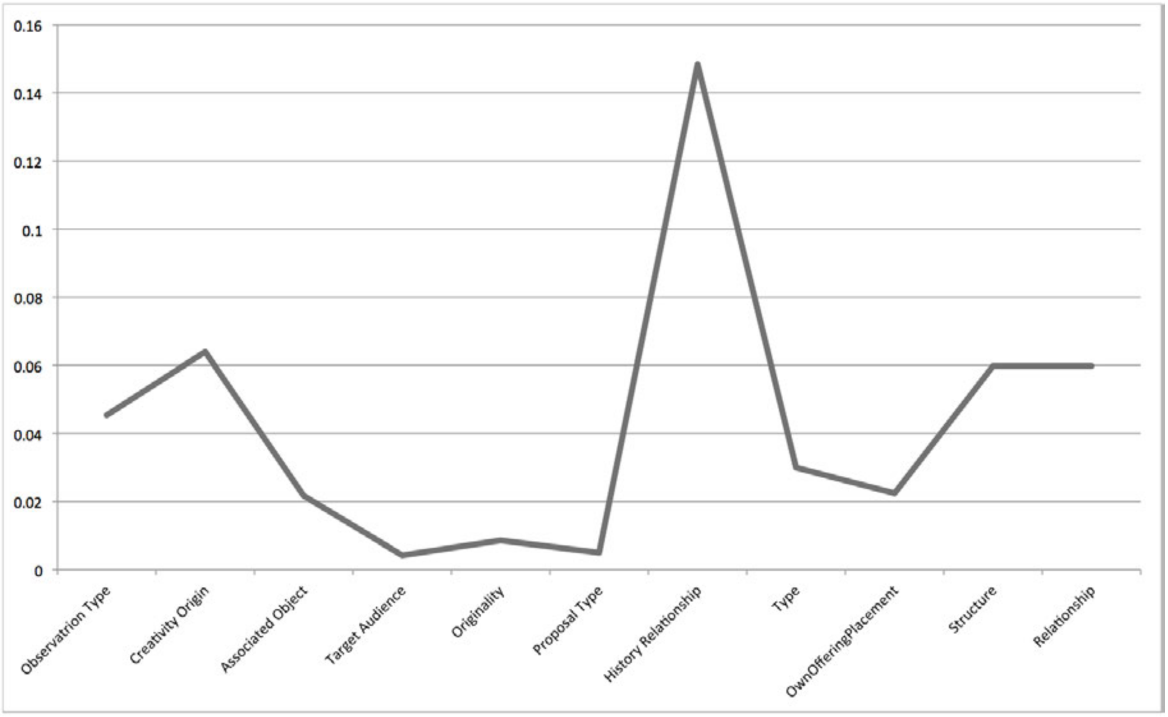


Figure 6.4: Variance of the dataset manual annotations split per taxonomy sub-tree (IdeaStorm dataset).

add additional rules for annotation (e.g. define terms that exclude each other etc.). In the first case, splitting of ideas into paragraphs brought a quite substantial improvement and interesting observations. Taking into account different taxonomy branches, on average the F-measure increased by 24% (with best case of 83% F-measure for Creativity Origin branch). In addition, we noticed that the amount of annotations per idea shrank because individual paragraphs did not hold enough information to assign terms from certain taxonomy branches (see Fig. 6.5).

In case of increasing the learning set size by 50% (up to 300 ideas) we got a 2% F-measure improvement in case of taking the full taxonomy into account and 5% F-measure improvement if analysing only the term branches filtered out earlier during the paragraph experiment. Finally, by adding some additional rules on top of the regular algorithm we did not get any improvement at all.

Concluding the experiment, we noticed that utilising the full taxonomy as originally proposed is very challenging if we desire to obtain all annotations in an automatic manner. Nevertheless, by measuring the performance of particular branches of the taxonomy we got interesting insight into the elements of the taxonomy that are already fit to be used with automatic annotation and which should be left for manual process. In addition, our results have shown that splitting idea text into paragraphs proved to work best for the type of textual submissions provided by innovators in Idea Management Systems.

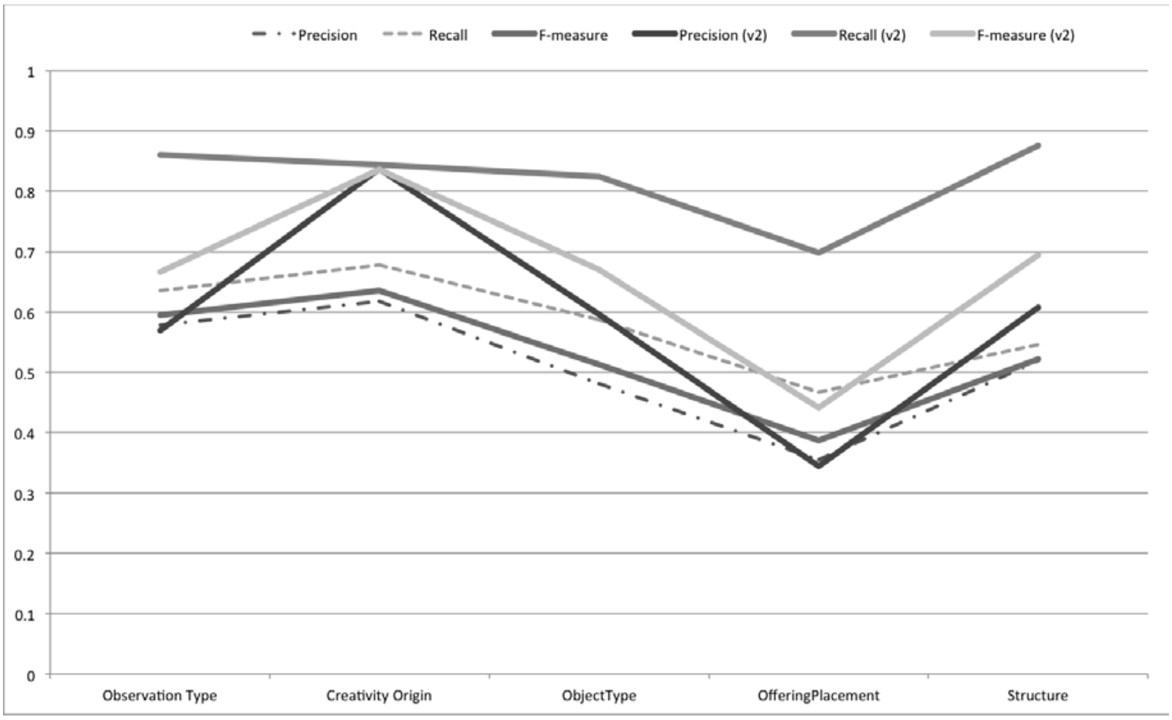


Figure 6.5: Differences in precision, recall and F-measure between annotation of IdeaStorm dataset when using full idea text and text split into paragraphs (v2).

6.5.2 Testing metrics with datasets

After achieving satisfactory results with taxonomy annotation tests, we proceeded with experiments to evaluate the metrics that can be generated after the annotations are delivered. The goal of the following tests was to verify if the metrics diversify enough between different datasets to be able to observe distinctive features of selected Idea Management instances and make assumptions about the types of communities engaged in the innovation process. Furthermore, we compared the proposed metrics with the currently available and measured if they have any correlation with the successfulness of ideas or each other.

We analysed a total of 4 datasets (see Table 6.2), from each we extracted and manually annotated 200 ideas: 120 random selected ideas, 40 ideas that have been implemented, 10 top rated ideas, 10 lowest rated ideas, 10 top commented ideas, 10 least commented ideas. The ideas were selected based on the analysis of the entire lifetime of the respectable instances since their start until the time our experiment was conducted (February 2011).

Two of the chosen instances are based on the same SaleForce Idea Management System. Both are administered in a similar manner as indefinite idea competitions: Dell IdeaStorm exists since February 2007, while the myStarbucks system is running since March 2008. In both cases, the organizations that own the systems are large multinational corporations with huge user base (e.g. Dell sold 44 million PC units just in 2009 (Dell, 2011), while Starbucks claimed to serve 60 million customers weekly in 2011 (Starbucks, 2011)). Up until the time of our experiment both instances presented similar user interface and workflow for the innovators as well as participants of the community. We have chosen those two instances to see if systems deployed in the same way from the perspective of infrastructure as well as idea management practices would diversify due to the fact that ideas are collected for different kinds of products (see Table 6.2).

Table 6.2: List of datasets used for the experiments.

System name	**#Ideas/ #Comments/ #Users**	**Area**	**Case Characteristic**
Dell IdeaStorm	15.000/ 90.000/ 2.000	Computers, telecommunication devices and related services.	Focused on collecting ideas for existing products over indefinite amount of time with periodically organized focus sessions
myStarbucks Ideas	8.000/ 80.000/ 3.000	Coffee and related products sold in a coffeehouse chain.	Focused on collecting ideas for existing products and changes in services over indefinite amount of time
Cisco i-Prize	1.000/ 4.000/ 1.000	Computer, networking and communications equipment.	Viewable only after registration and available only during a set amount of time. Focused on collecting very abstract ideas for new area of activity. Introduces considerable money incentives for best inventors.
Ubuntu Brainstorm	27.000/ 90.000/ 2.000	Open-source operating system and related software.	Very collaborative, computer literate community gathered around open-source software products. Apart of ideas system enables submission of proposed implementation methods for ideas.

The third instance included in our tests was Canonical's Ubuntu Brainstorm that was opened in February 2008 and is based on an open-source IdeaTorrent platform. In comparison to the previous instances, the idea submission rules are different and force innovators to deliver solutions for their ideas. Another major difference is that Canonical user base is smaller in comparison to Dell or Starbucks (20 million users total as estimated by Canonical (Ubuntu, 2011)) but also very collaborative (Lakhani and von Hippel, 2003; Feldstein, 2007) and only focused on a single type of an open-source product. The implementation process of ideas is significantly more transparent due to the fact that Ubuntu is an open-source project and all its production infrastructures are available to public and linked to Brainstorm. We have chosen to analyse this instance to see if the computer technology literate audience of Canonical that is used to giving contributions for free would propose ideas that differ in comparison to mass consumer customer base of Dell and Starbucks.

The final dataset that we analysed came from an instance called i-Prize, operated by multinational corporation called Cisco. The instance started running in February 2010 and was only open for three months. Apart of setting a limited time frame for the collection of ideas, Cisco also offered considerable money incentives for the winners that proposed the best ideas. In contrast other instances do not have any incentives apart of public mentions of the winning ideas. Additionally, the goal of i-Prize contest was to collect ideas for a new major future Cisco business while in all three other instances there were no precise goals other than gathering feedback from clients on current products and services.

Taking into account the described differences between the datasets we applied the previously introduced metrics to see if those four different datasets would indeed differ as expected when measured with our metrics. The process of applying the metrics to a dataset included: calculation of metric value per every idea individually, calculating the median value out of all 200 ideas annotated. We followed this methodology with all metrics and for all data samples from every dataset. When visualised on a chart (see Fig. 6.6) we were able to observe the differences between the datasets and interpret them. As hypothesised before, the biggest similarities can be observed with Starbucks and IdeaStorm instances which gather ideas in competently different areas but are run by the same operator (11 out of 14 metrics had the same values). The most standing out difference between these two datasets can be observed with regard to Innovation Freshness: IdeaStorm ideas in majority were never implemented before while most of Starbucks ideas are reoccurring requests to bring back old innovations.

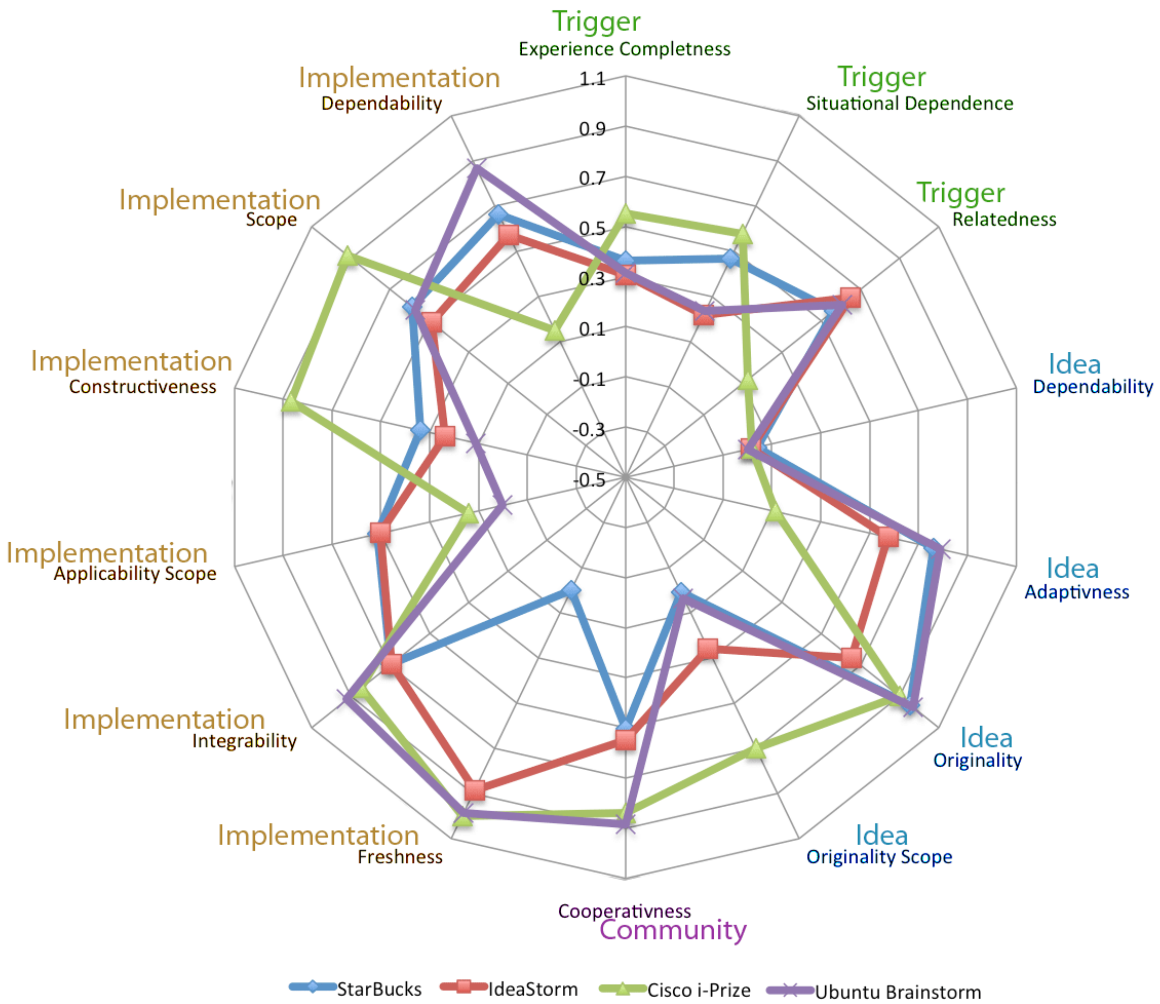


Figure 6.6: Comparison of metrics based on interpretation of taxonomy terms.

Based on the metrics calculations, the Cisco i-Prize is the instance that exceeds others by a large margin in many areas but also has most contrasts (5 top scores and 4 lowest scores). Most notably this instance is characteristic for remarkably high Implementation Constructiveness and Scope, which could be attributed to i-Prize contest explicitly asking for ideas in new areas covering very board scope. This assumption is also confirmed by high Idea Originality Scope that shows that most proposed ideas are original with regard to a very broad scope of markets.

Lastly, the ideas originating from Ubuntu open-source community stand out most in two areas: Implementation Dependability and Applicability Scope. The first metrics shows that Ubuntu users most often propose changes in key elements of offering that have impact on many software modules. The second metric shows that the proposed ideas are very specific and aimed only for particular products from Canonical offering.

Concluding the above analysis we observed that the metrics enabled to verify judgement about certain instances and deliver proof to how certain communities exceed others. In addition to such interpretation we evaluated the diversity of datasets measured with entropy. When visualised on a radar chart (see Fig. 6.7) the area taken by dataset determined how similar to each other are ideas of different datasets. In this particular case our experiment has shown that overall ideas posted in IdeaStorm and myStarbucks instances were most diversified while Cisco and Ubuntu Brainstorm least.

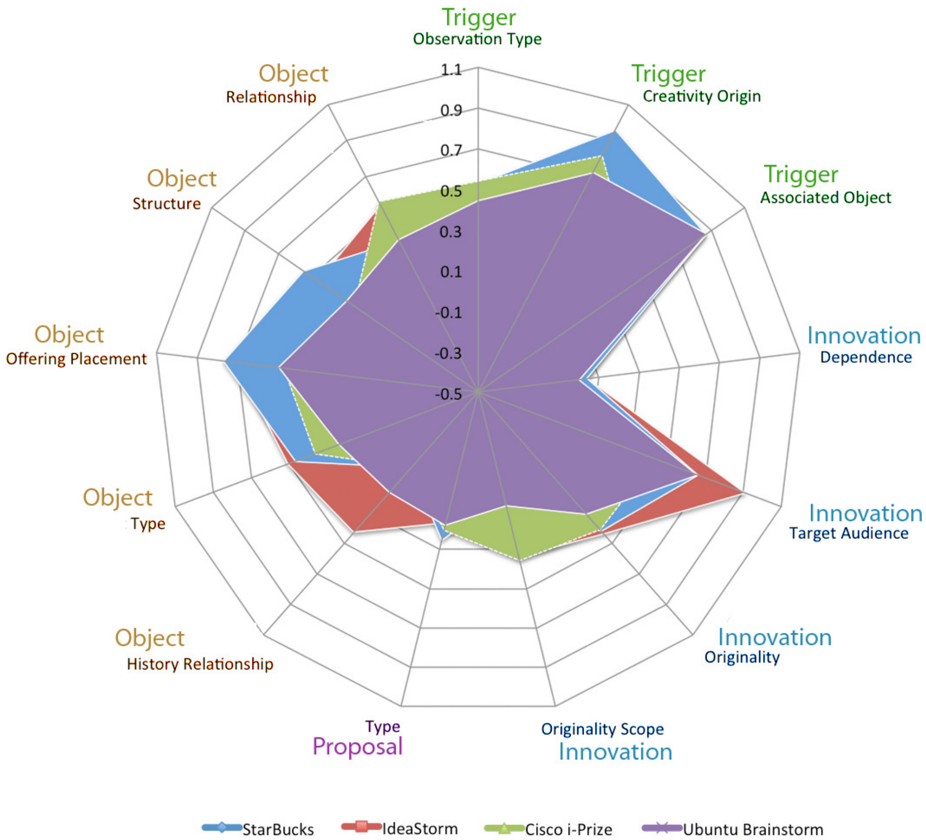


Figure 6.7: Comparison of diversity of datasets with respect to different taxonomy branches based on entropy measure.

Metrics Correlation Analysis

The analysis presented in the previous section has evaluated the metrics when used for judgement of entire datasets or groups of ideas gathered in idea contests. However, Idea Management Systems suffer not only from lack of tools for assessment of entire instances but assessment of individual ideas in particular. Therefore, in the final experiment, we compared the currently used metrics for idea assessment and idea selection processes (Jouret, 2009) with the metrics proposed by us. Our goal was to check: 1) if there would be a meaningful change in the proposed metrics values or correlations when calculated for particular idea subsets (e.g. top commented ideas or top rated ideas); and 2) if the relationships between our metrics and idea adoption would be similar to the impact of legacy metrics on idea adoption.

To achieve the stated goals, we related our metrics with the following legacy metrics used in contemporary Idea Management Systems: idea rating value, number of comments for idea and idea age (amount of days until idea gets implemented; for not implemented ideas days until the date of newest idea in the test dataset). In particular, we measured the bivariate correlations between our metrics and the legacy metrics (see example of Dell IdeaStorm in Table 6.3). The correlation between all variables turned out small (according to Cohen scale (Cohen, 1988)) which suggests that there is little point for analysis of our metrics in border line conditions of community metrics typically used in Idea Management Systems.

Table 6.3: Bivariate correlations between the proposed metrics and legacy metrics (Dell IdeaStorm dataset)

Metric	# Comments	Rating	Idea Age
Completeness	0.04	0.11	0.01
Experience Completeness	-0.15	0.03	-0.15
Situational Dependence	0.17	0.28	0.13
Relatedness	0.04	-0.04	-0.13
Dependability	-0.04	-0.03	-0.1
Adaptiveness	-0.04	-0.19	-0.16
Originality	-0.17	-0.13	-0.11
Originality Scope	-0.01	0.03	-0.01
Cooperativeness	-0.08	-0.14	-0.07
Freshness	-0.03	0.08	0.07
Integrability	-0.17	-0.22	-0.18
Applicability Scope	0.11	0.07	0.09
Constructiveness	-0.09	-0.04	-0.049
Scope	-0.10	-0.15	-0.22
Dependability	0.22	0.15	0.46

To assure that those results were not only the case of a single dataset we measured the aforementioned correlations for all other test datasets and observed the differences between the correlations of the same metrics. While in most cases the correlations remained small as in IdeaStorm, the standard deviations were quite significant in comparison to the mean correlation value of all datasets (see Table 6.4). This could lead to a conclusion that the behaviour of idea characteristics (expressed with our metrics) in relation to community activity (measured with legacy metrics) is individual for every particular environment and setting of idea campaign.

[1] correlation undefined for one of the datasets

Table 6.4: Mean and standard deviation of bivariate correlations between all datasets for Gi2MO Types metrics and legacy metrics

Metric	# Comments		Rating		Idea Age		Idea Adoption	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Completeness	0.06	0.02	0.08	0.04	-0.02	0.04	0.12	0.05
Experience Completeness	0.02	0.13	-0.01	0.05	-0.01	0.13	0.00	0.14
Situational Dependence	0.08	0.07	0.11	0.11	0.10	0.10	0.10	0.13
Relatedness	-0.01	0.10	-0.01	0.08	-0.10	0.12	0.01	0.12
Dependability	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]
Adaptiveness	0.01	0.06	-0.05	0.13	-0.07	0.06	-0.01	0.10
Originality	0.00	0.12	0.01	0.10	-0.08	0.16	-0.01	0.18
Originality Scope	0.02	0.05	-0.01	0.03	0.03	0.14	-0.01	0.16
Cooperativeness	0.07	0.13	0.02	0.14	-0.07	0.16	0.06	0.10
Freshness	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]	n/a[1]
Integrability	-0.10	0.09	-0.11	0.11	-0.08	0.08	-0.06	0.06
Applicability Scope	-0.04	0.12	0.01	0.06	0.02	0.08	-0.09	0.06
Constructiveness	0.04	0.16	0.04	0.10	-0.02	0.03	-0.03	0.08
Scope	0.04	0.13	-0.02	0.13	-0.06	0.12	-0.03	0.08
Dependability	0.08	0.11	0.06	0.12	0.14	0.22	0.11	0.13

As a follow up, we also measured correlations exclusively between the metrics proposed in this article. Yet again, the results were very different depending on the dataset, however some metrics within the scope of a single dataset were strongly correlated allowing to make interesting observations about the communities:

- **Starbucks**: ideas for products impose more modifications in existing offering than ideas for services or processes (strong correlation of Implementation Dependability and Integrity)
- **Cisco**: the only instance where inventors connect new products with gain of new type of customers (strong negative correlation between Constructiveness and Adaptiveness); generic ideas for product types are related to modifications in existing offering, while specific and detailed ideas are more typical for completely new items (strong negative correlation between Constructiveness and Applicability Scope)
- **Ubuntu**: very original ideas are also the ones that deliver most complete description (strong correlation of Originality and Idea Completeness); similar as in StarBucks product ideas impose more modifications in related items of offering (strong correlation of Implementation Dependability and Integrity)
- **IdeaStream**: none of the metrics had a strong correlation

Except for the IdeaStorm instance, all other datasets had one single standing out similarity: ideas that proposed a complete structure change of products often referred to creating new products rather than redesigning old ones (very strong correlation between Implementation Scope and Constructiveness).

Finally, in addition to correlations between metrics, we measured and compared the correlations of all metrics to idea adoption (determines if an idea was implemented or not) to see if our metrics do better or worse as a tool for detecting good ideas. In case of IdeaStorm (see Table 6.5), in majority of cases, our metrics had a better correlation with idea adoption than the legacy metrics. The most standing out results were achieved by Innovation and

Object metrics. Nevertheless, according to the Cohen scale the impact in best cases can be described as medium. After repeating the experiment for the 3 other datasets (myStarBucks, Cisco and Ubuntu), the final conclusions were similar.

Table 6.5: Bivariate correlations between metrics and idea adoption (Dell IdeaStorm dataset)

Metric	**Idea Adoption**		**Metric**	**Idea Adoption**
Completeness	0.16		Freshness	0.05
Experience Completeness	-0.03		Integrability	-0.1
Situational Dependence	-0.07		Applicability Scope	-0.09
Relatedness	0.02		Constructiveness	-0.15
Dependability	-0.07		Scope	0.14
Adaptiveness	0.05		Dependability	0.23
Originality	-0.27		# Comments	-0.04
Originality Scope	-0.2		Rating	-0.04
Cooperativeness	0.01		Idea Age	-0.06

Concluding all experiments with the correlation measure, the proposed metrics provide a small improvement over the legacy metrics in terms of picking the winning ideas. Our results show that Idea Originality as well as Object Dependability are better indicators than any other. Additionally, the correlation analysis delivered another proof that our metrics can be used for comparison of different environments and discovering characteristics of the communities.

6.6 Summary

In this chapter we have proposed a set of new automatically generated metrics to aid the decision making process during the assessment of ideas in Idea Management Systems. Our hypothesis was that these metrics could be derived from annotations made with a specially crafted taxonomy, and used to characterise community generated innovation in a sufficient way to compare the gathered data. This hypothesis has been confirmed with a number of experiments that used the taxonomy as a tool to discover differences and similarities of various case studies.

Furthermore, we presented an evaluation of all the steps underlying the generation of metrics and obtained valuable insight into conditions under which our methodology performs best. We determined that out of four proposed taxonomy sub-trees substantial parts of two (Trigger and Object) can be applied automatically with satisfactory results, while the characteristics represented in the two remaining sub-tress (Innovation and Proposal Type) should be analysed and applied by a human. Furthermore, we have shown that the manual annotation delivers significantly better results when done by a single annotator rather than a group (regardless of the level of expertise with innovation theory).

Finally, we evaluated the use of metrics not only for comparison of entire datasets but also for decision making process of selecting the individual ideas for implementation. We determined that the borderline cases of community activity that are currently used for filtering ideas (vote count, comment count etc.) do not influence the values of metrics proposed by us (e.g. more original ideas are not more commented or voted on). In addition, the obtained results have shown that our metrics deliver slightly better results to predict winning ideas in comparison with the contemporary used community metrics. Most notably, our results show best performance for Idea Originality and Object Dependability as best measures of idea adoption, standing out in comparison to any other metric.

Chapter 7

Idea Relationships Model for Idea Management System

In the practice of case studies, ideas submitted in the Idea Management Systems are rarely isolated from each other. It is often the case that community members duplicate each other or connect to parts of ideas made by other people. Furthermore, as we have shown in the previous chapter, innovations rely of past activities and events from the enterprise environment. All those relationships are the topic of research presented in the following chapter.

Firstly, we analyse the relationships between ideas created inside a single Idea Management System. We note that in the current systems the only type of relationship utilized is 'duplicate'. Therefore, we propose an extended hierarchy of relationships between ideas.

Secondly, we extend the scope of our research and look into the relationships that go beyond the borders of an Idea Management System. We propose a classification of enterprise and public systems related to Idea Management and define a methodology for extending the earlier presented Gi2MO Ontology with idea metrics based on relations of an idea with data concepts from other systems.

The contributions presented in the chapter are:

- Gi2MO Links: hierarchy of relationships between ideas in an Idea Management System
- Methodology for extending the Gi2MO ontology for interlinking ideas with other enterprise data to generate idea metrics
- Idea aggregation metric to compare the capabilities to summerize idea datasets based on relationships

7.1 Introduction

A number of definitions of an *idea concept* point to the fact that a new innovation is very often a combination of existing ones (Foster, 2007; Young, 2003; Koestler, 1990). This implies that ideas reuse each others elements and duplicate parts of their descriptions. Moreover, according to a number of studies, in Idea Management Systems the intersections between ideas are not only the result of incremental innovation but also occur due to the fact that: 1) innovators do not find it worthwhile to review the huge database of existing ideas in search of duplicates before posting their own idea (Ford and Mohapatra, 2011); 2) idea authors are not presented with proper tool support for duplicate search (Ford and Mohapatra, 2011); or 3) innovators submit willingly duplicates because of the effort that the author has put in creating his idea or to show that the author of duplicate was thinking in the same direction before as well (Bailey and Horvitz, 2010). As a result, duplicates are a significant part of the gathered ideas and make idea assessment a time consuming, tedious process (Geffen and Judd, 2004).

For this reason, the research and industrial efforts in the area of Idea Management have focused on methods that would improve automatic duplicate detection (Ford and Mohapatra, 2011). However, aside of duplicate relationship, the innovation management research points out a number of other types of relations between incremental ideas (e.g. based on product dimensions that idea discusses (Normann, 1971) or based on type of changes in product structure subsystems (Hubert Gatignon and Anderson, 2002)). Similarly, in other enterprise activity areas, relationships have been the topic of many studies (Maier et al., 2005; Wood, 2010; Rebstock et al., 2008; Decker, 2002; Zhou et al., 2010; wen Ma, 2009) and it has been shown that the understanding of semantics of those relationships can contribute to data summarization (O'Leary, 1998; Rao et al., 2012), information search (Yoo and Kim, 2002; Ho et al., 2004) and aid assessment of gathered information (Chen et al., 2010; Han and Park, 2009).

Furthermore, in case of Idea Management, as we have shown in the previous chapter, ideas are not only related with each other but also connect to the environment, needs and events that influence the innovator as well as the organization that seeks the new ideas. In fact, according to a survey of Boston Consulting Group (Andrew et al., 2009b), the most important criteria for innovation adoption in enterprises is derived from knowledge that is not located in the Idea Management System (e.g. financial impact, development effort required, production cost, relation to own products and competitors etc.).

For the reasons described above, in this chapter, we propose to investigate the complexity of relationships that govern Idea Management Systems. In particular, we propose a hierarchy of relationships between ideas inside of a single Idea Management System as well as investigate what kind of enterprise and public systems are related to information contained in ideas of an Idea Management System.

With regard to relationships limited to a scope of a single system we investigate how recognition of those relationships could aid to summarize the data of Idea Management System and minimize the effort that an idea reviewer or idea contest manager has to spend in order to asses ideas.

With respect to relationships between Idea Management System and other external systems we recognize that there is a big variety of systems and information that can be connected to an idea. Therefore, we do not attempt to list all possible related systems but categorize the nature of relations and propose a methodology for extending the Gi2MO Ontology to embrace the new data with the use of the Linked Data paradigm (Berners-Lee, 2006). We also describe how to utilize the formalized links for generation of metrics that would allow comparing and ranking of ideas.

In both cases, we relate our study of relationships to the main research questions of the thesis, i.e. idea comparison during the idea assessment phase.

7.2 Related Work

7.2.1 Relationship modelling

Semantics of relationships is a highly investigated topic in the area of linguistics and psychology. Its application in the computer science has been reviewed in a number of works for domains such as information retrieval or information extraction (Khoo and Na, 2006; Green et al., 2002; Myaeng and McHal, 1991). Myaeng et al. (Myaeng and McHal, 1991) reviewed the created classifications of relationships in those areas and split them into pragmatic and linguistic. During our research on idea relationships we used those relationship hierarchies as a reference. In particular, we analysed a taxonomy of relationships proposed by Bejar et al. (Bejar et al., 1990) and attempted to transform the language relationships into idea relationships. In many cases our conclusion was that relationships applicable for language constructs either did not make sense when applied for innovation or intersected with each other making classification of ideas a difficult task.

Such a debate about relevance of linguistic relationships in other areas has been the topic of interest of knowledge management research (Bean and Green, 2001; Stephens et al., 1995). In this area, the concept of relationships has been often investigated and modelled on the level of entire knowledge objects rather than just language constructs. A number of works in the ontology research (e.g. Cyc (Lenat, 1995)) and Semantic Web in particular (e.g. OWL (Patel-Schneider et al., 2004)) attempt to define such knowledge relationships on a generic level. Additionally, in many cases researchers have analysed semantics of relationships for specific narrow domains. For instance, the Learning Object Model (LOM) specification (IEEE, 2002) defines a simplified model that has been argued and extended in a number of works (Rodriguez et al., 2009; Fischer, 2001). We did not encounter similar studies of relationship types for Idea Management Systems in particular, however we used the achievements from other domains such as aforementioned e-learning or multimedia (Marsh and White, 2003) to recognize how information objects can be linked for delivering a more complete overview of the entire knowledge repository. Additionally, in comparison to related work on relationship hierarchies in both knowledge management and earlier described linguistics, our research does not attempt to find a most complete or suitable relationship taxonomy for Idea Management but to determine if there is any point of introducing such.

With reference to the domain of Idea Management in general, Kornish et al. (Kornish and Ulrich, 2011) has studied idea similarity and its appliance for clustering. Nevertheless, Kornish did not point explicitly to naming or identifying any relationships, just the summerization capabilities of identifying similar ideas in general. The results of his study show that clustering indeed can aid idea assessment phase and achieve good results. Following those conclusions, our work compliments the study of Kornish and proposes in the next step to identify the relationships between ideas and perform further clustering based on that knowledge.

7.2.2 Relationship formalization and metric building

Aside of naming idea relationships and using them for clustering, in this chapter, we also propose the formalization of discovered relationships and establish Linked Enterprise Data within the context of Idea Management ecosystem to generate idea metrics. While to our knowledge this particular solution has not been tested in context of idea management, there is a number of different approaches that relate to our work in both the research fields of innovation management and the Semantic Web.

In relation to exclusively Idea Management Systems, Hrastinski et al. (Hrastinski et al., 2010) surveyed a number of selected products and pointed out that the current commercial systems employ rather simple idea evaluation methods most often being analysis of community statistics (number of ideas per user, community voting results, number of idea comments etc.) or internal business metrics that are delivered by designated experts. On the other hand, shifting towards the scientific research in the area, there have been various approaches that attempted to find a solution to time efficient and effective automatic idea assessment problem e.g. with prediction markets (Bothos et al., 2008), by applying problem solving algorithms (Adamides and Karacapilidis, 2006) or calculating metrics for the quality of management (Conn et al., 2009). However, non of those did direct towards the use of metadata to integrate idea management with other business systems as we propose. The previous research that does take into consideration use of ontologies most often is discussed in context of innovation management which is a more broad yet also more generic point of view on innovation than Idea Management Systems. For instance, Ning el at. (Ning et al., 2006) introduces a vision of the semantic extended enterprise where Semantic Web technologies are used to collect similar data from different innovation oriented systems yet omits the particularities of using different ontologies in systems distributed across the enterprise. To our knowledge, specifically in the area of Idea Management Systems and Semantic Web, only Rield et al. (Riedl et al., 2009a) proposed an ontology for describing the Idea Management System data structure but did not discuss the ontology in the context of interoperability with other enterprise systems and their dedicated ontologies.

In relation to Semantic Web research carried out for other domains and Linked Data approaches to the enterprise environment modelling there have been numerous solutions proposed. In many cases, the research carried out so far focuses on very specific systems - the most relevant ones from the point of view of Idea Management are presented later as we discuss the contribution details of our methodology (see Sec. 7.4.2). In those solutions, when approaching knowledge management problems, in most cases the focus is put on getting deep into details of representing domain specific knowledge or system structure and taking advantage of this with various reasoning scenarios (Mika and Akkermans, 2003). Contrary to such methods in our work we simplify the technical approach and attempt to direct the research effort towards investigating benefits that come from particular links between the data of very different systems. As such, from the technical and conceptual point of view, we align our vision of Semantic Web in the enterprise more to the principles presented by the Linking Open Data project (Linking Open Data, 2011), however with the obvious distinction of not publishing the data in the open and just using the same lightweight data linking approach. There have been some initial initiatives for establishing Enterprise Linked Data but so far the focus has been put on pulling the information from the Linked Open Data cloud into the enterprise and reusing it (Wood, 2010). In our work, we also notice the huge benefit of open data for the enterprise but at the same time we dedicate to the concept of creating an Enterprise Linked Data cloud that would be private and individual for a given corporation.

Finally, in relation to using the Linked Data in practice, as part of our evaluation we presented the notion of generating charts over the interlinked datasets. Similar work on calculating metrics over the open datasets has been presented by Zembowicz et al. (Zembowicz et al., 2010). In comparison to our implementation that evaluates charting in a particular domain, Zembowicz focuses more on the user interface side and translating between complex SPARQL queries to enable a simple human-computer interaction method.

7.3 Relationships between Ideas of an Idea Management System

Idea Management Systems are an implementation of open innovation paradigm (Chesbrough, 2003) with the use of social collaborative Web platforms. Traditionally, open innovation concept involved asking parties not directly involved in product development for ideas in a suggestion box-like fashion (e.g. as practised by Toyota for over 50 years, much before open innovation term or Internet were born (Yasuda, 1990)). Nowadays, the huge popularity of social networking platforms and increasing literacy of consumers with Web tools allows to extend those practices with crowdsourcing activities (Howe, 2004) that not only gather ideas from consumers on mass scale but also to make the innovators more aware of each others innovations and encourage them to collaborate on idea improvement and reuse of each others ideas. As an outcome, the new ideas submitted to the Idea Management Systems are often interconnected in a variety of ways.

In our research, we refer to the large data volume problem by discussing the concept of idea relationships and clustering in idea datasets based on types of relationships that connect ideas (see Fig. 7.1). In the contemporary systems this problem is typically handled by duplicate detection in conjunction with crowdsourcing methodologies that employ users in submitting duplicate reports rather than utilizing fully automatic approaches. With reference to this state, we investigate if there is room to introduce new types of relationships between ideas and if this change would allow to make a meaningful impact on downsizing the idea dataset in instances of Idea Management Systems that contain tens of thousands of ideas.

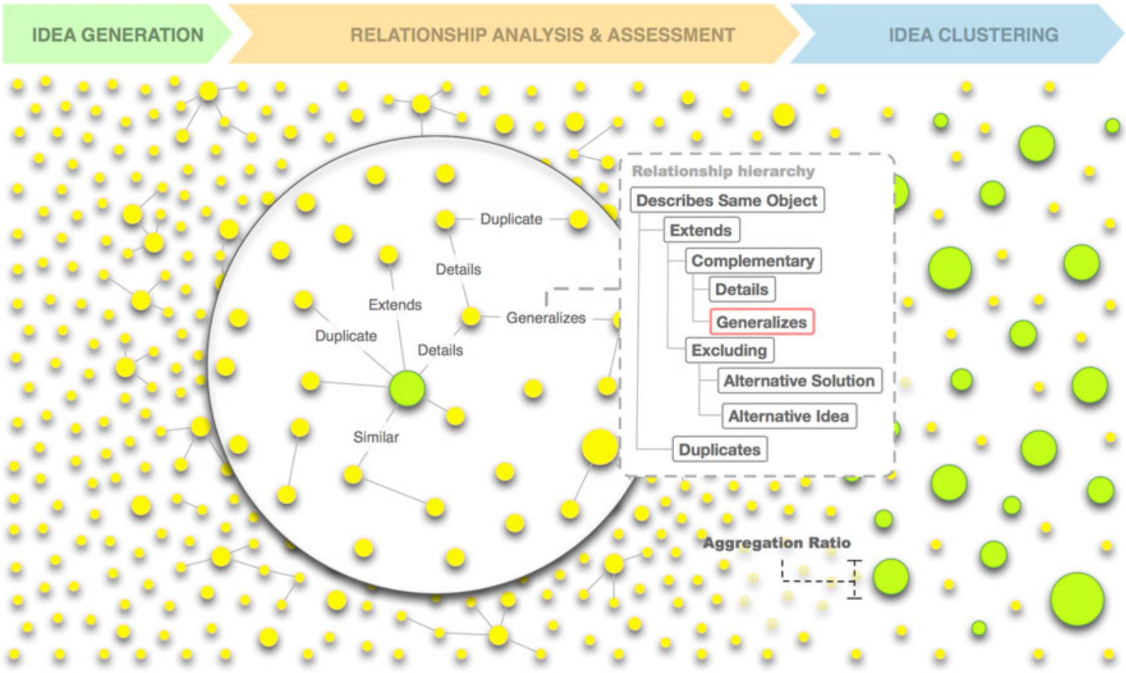


Figure 7.1: Approach taken for investigating idea relationships in Idea Management System

To achieve our goal, we propose an extension of idea relationships (see Sec. 7.3.1). Afterwards, we evaluate the proposed hierarchy by formulating a number of hypotheses that verify if our proposal indeed aids idea assessment and solves specific problems of introducing new idea relationships in open innovation (see Sec. 7.5).

7.3.1 Idea Relationship Hierarchy

To facilitate achieving the aforementioned goals, we contribute a hierarchy of relationships between ideas in an Idea Management System. The preliminary version of the hierarchy has been created based on our past experiences in the Gi2MO project with various idea datasets (Westerski et al., 2010; Westerski and Iglesias, 2012). Later, we refined the hierarchy by running a number of test annotation experiments with various datasets and by referring to the earlier described related work. The final version of hierarchy proposal used during our experiments is presented in Table 7.1.

The relationships that can be established between ideas have been separated into two categories: those that can be identified by analysis of the text of two ideas (A - Based on knowledge) and relationships that are created based on user interactions with the system (B - Based on Action). This state is a result of experiments with applying the presented relationships to various idea datasets. Although relationship models referenced before (e.g.

Table 7.1: Proposed idea relationship hierarchy for Idea Management Systems.

A Based on knowledge	Relationship existing between knowledge content of ideas created independently
A1 Similar	Ideas similar to each other
A1.1 Describes Same Object	Ideas that propose a similar innovation for the same item
A1.1.1 Extends	One idea extends other
A1.1.1.1.1 Complementary	Ideas that can work together
A1.1.1.1.1.1 Details	One idea focuses on part that other neglects
A1.1.1.1.1.2 Generalizes	One idea describes a more broad vision of other
A1.1.1.1.2 Excluding	Implementations of ideas exclude each other
A1.1.1.1.2.1 Alternative Solution	Ideas refer to the same object and problem but solved in different ways
A1.1.1.1.2.1 Alternative Idea	Two completely distinct ideas that in effect are impossible to implement together
A1.1.2 Duplicates	Ideas describe exactly the same innovation
A1.2 Describes Related Object	Ideas that propose innovation for different objects that are somehow related to each other
A2 Disjoint	Ideas not having any meaningful similarities
B Based on Action	The relationship is created by an action operating on both ideas by a user of the system
B1 Based on Moderator Action	Action taken by moderator of the system in reaction to submitted ideas and relationship annotations
B1.1 Follows	Implementation of an idea should follow some other idea
B1.2 Proceeds	Implementation of an idea should proceed some other idea
B1.3 Merged	Two ideas merged into a single one
B2 Based on Innovator Action	Relationships created based on user interaction with ideas
B2.1 Originates	Ideas created by extending some other idea
B2.2 Is version	Created by updating an idea (e.g. in reaction to community feedback)
B2.3 References	One idea referencing other idea (or resource from outside the system)

LOM (IEEE, 2002)) do not apply such distinction, we identified that annotators were unable to put any of the relationships from group (B) just based on the idea text and without the contextual knowledge of the entire idea repository, including history of the examined ideas.

Additionally, we propose a certain interpretation of the dependencies between the relationships in the proposed hierarchy, namely:

- *similar*, *disjoint*, *describesRelatedObject*, and all *excluding* relationships are symmetric
- *details* relationship is inverse of *generealizes* relationship
- *extends* and *duplicate* relationships are not symmetric and during annotation we provided additional *is extended* and *is duplicated* relationships being inverse to the aforementioned

7.4 Relationships between Ideas and Other Enterprise Data

Aside of the proposed hierarchy that models mutual relationships between ideas we also notice that Idea Management Systems are part of the enterprise ecosystem and idea descriptions may relate to information stored in external systems. In search of a methodology to model those relationships, we refer to the earlier mentioned research on Semantic Web and integration between heterogeneous systems via modelling the semantics of relations. The Semantic Web in its origins was supposed to be a remedy to information overflow of the ever-growing Internet where machines through analysis of content relationships could help human to reach the desired data in a fast manner (Berners-Lee, 1998). As the topic gained interest it became obvious that the same technologies aimed for organising the global Internet network can deliver value to internal, closed environments of large enterprises that suffer similar problems of information overflow and disorganization (Feigenbaum et al., 2009).

The first attempts in both areas have put a lot of effort in development of reasoning techniques and algorithms related to the artificial intelligence. However, as this approach did not succeed to bring the desired solutions to mainstream development, more lightweight approaches were born to introduce metadata annotations to the Web and their simplistic use. Among them is Linking Open Data (Linking Open Data, 2011) initiative and research gathered around it that tries to draw simple patterns for usage and publication of online metadata linked across independent systems.

In the following section, we relate our observations about relationships between Idea Management Systems and enterprise data together with the notion of linking data for improving knowledge management. In particular, we conform to the trend of transforming the Web of Data into Web of Linked Data and focus specifically on the benefits that this might bring to the enterprise, i.e. data analysis for innovation management and interlinking various enterprise systems to support innovation processes in the organization. The principal research questions that we attempt to answer are: what enterprise systems and which of their data can be useful for innovation management, how to extend the earlier proposed Gi2MO ontology towards linking data and finally how to utilize the connections to calculate innovation metrics.

In that context, we follow a research methodology (see Figure 7.2) that leads to extending the Generic Idea and Innovation Management Ontology (Gi2MO) (Westerski, 2012a) towards establishing links with enterprise systems and exploiting their data. In particular, we motivate our work with the desire to extract innovation metrics though analysis of linked data (see Sec. 7.4.1). On the road to achieving this goal, we establish a classification of systems present in the idea management ecosystem and proceed with the analysis of their current status in terms of ontologies and interlinking efforts (see Sec. 7.4.2). Further, we show how the data can be exploited to create new capabilities for Idea Management Systems and propose particular interlinking methods (see Sec. 7.4.3).

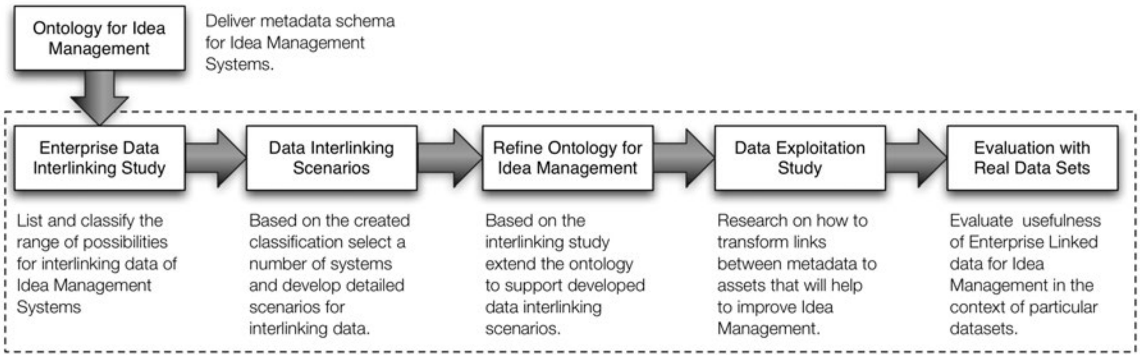


Figure 7.2: Research approach taken for investigating Enterprise Linked Data for Idea Management

7.4.1 Motivation

One of the most important and troublesome stages Idea Management is data assessment. Separating good ideas from bad is the core reason for existence of idea management. Currently, to perform this task, human reviewers fill out forms and deliver assessments which are the means for standardized comparison of ideas, their filtering and finally selecting the best candidates for implementation. On the other hand, the automatically generated metrics in most cases are limited to simple statistics derived from community activity (e.g. average number of comments in time per idea, per user etc.).

In relation to those activities the key problems of idea management are: information overflow (e.g. when a new product is announced by a company, the idea management facilities are flooded with new ideas), information redundancy (often ideas duplicate each other) or information triviality (simple and obvious ideas do not provide genuine value). Each of those issues impact in a negative way the idea assessment process and moderation activities which in turn discourages people from submitting new ideas because of slow feedback and little impact.

As an improvement over this state, in our research we propose to use datasets of other enterprise and public systems to supply additional data for idea management to generate new metrics and aid idea reviewers (see Fig. 7.3). For example, assessment of idea value based on links to supporting material, or judgement of how bad idea implementation went based on links to system that gathers implementation problems (e.g. bug-tracking). In the next sections of this chapter we describe how we cope with this problem through use of Semantic Web technologies and specifically extending the Idea Management Ontology to facilitate various interlinking scenarios.

7.4.2 Enterprise Data Interlinking Study

In order to fulfil the goals stated in our motivation, firstly we needed to identify the possible systems involved in the IMS ecosystem and face a challenge described by the question: *"What*

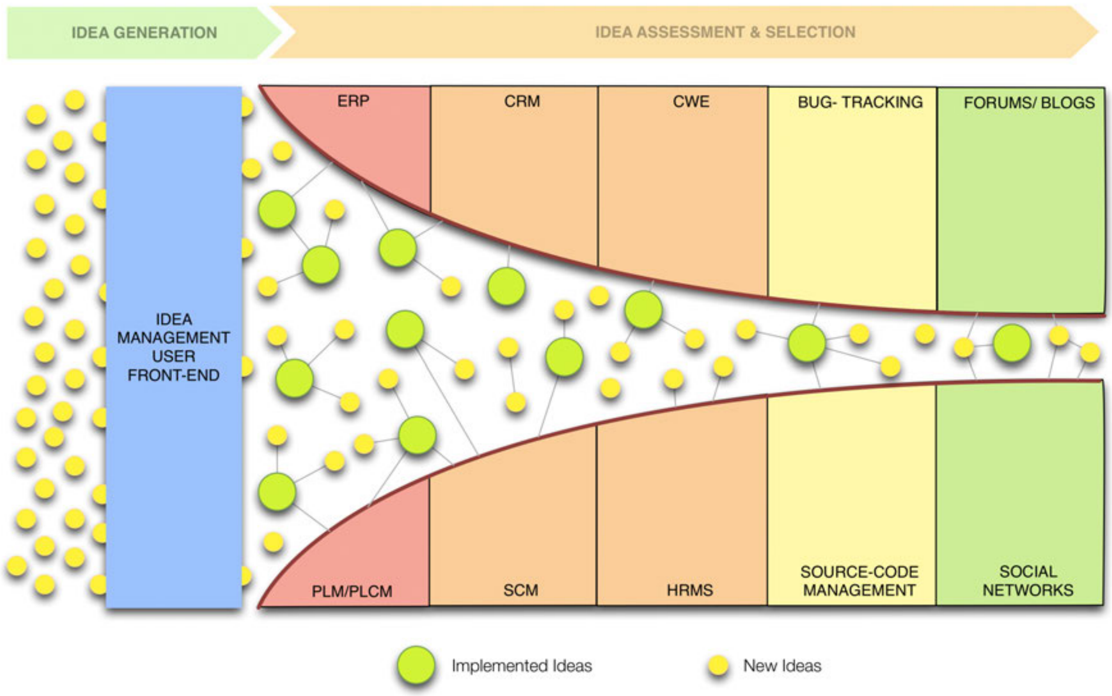


Figure 7.3: The concept of funnelling ideas based on their metrics derived from connections to data in other systems

data can be used to interlink with Idea Management Systems?". Based on the origin of data valuable for idea management we propose to classify it into three categories (see Fig. 7.4).

Starting from least complex linking scenario, we interpret those categories as:

- **interlining Idea Management System internal assets.** The simplest case where we interlink only internal data of Idea Management System to deliver better tractability and allow analysis of how different phases of idea life cycle impact each other. In the previous section we have detailed possible kinds of those relationships, while in Chapter 4 we shown what kind of impact relating ideas of different maturity can have. Here we will view those relationships in terms of general similarity and show how they may be interpreted in a generic way just like all other relationships of an idea to any sort of information, inside or outside the IMS.

- **interlinking internal data across the enterprise.** This is a case of enterprise systems integration that are not shared with the public and transferring the benefits of that information onto Idea Management Platform. The difference in comparison to the first case is that data spans over multiple systems of different types. Therefore, we are presented with the systems integration and data mediation problems.

- **interlining Idea Management data with public data.** This is a case where assets from Idea Management Systems are linked to data published in other independent systems that are available for public use (e.g. social networking portals). The evolution

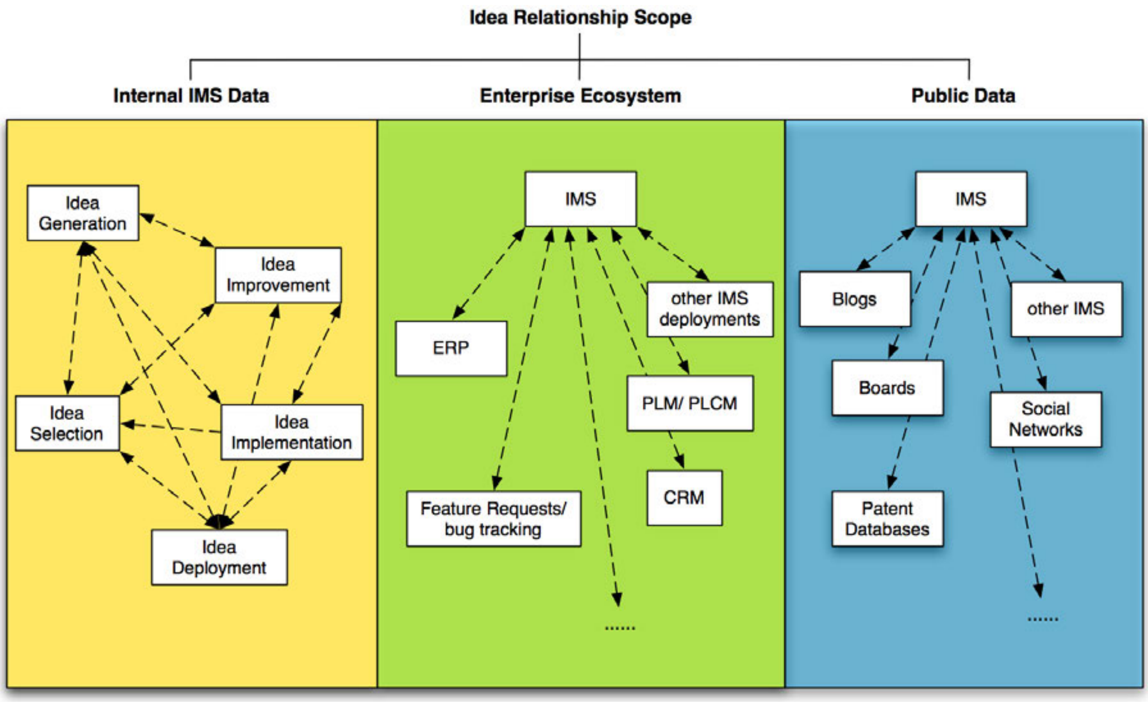


Figure 7.4: Dependencies between idea information and other data, categorized in terms of IT system scope.

of the problem in this case, in comparison to the previous, is that there is no control over systems maintained by other companies, and possibly the data as well because it is created by large communities moderated by independent parties.

Each of the mentioned categories can be further detailed, however is has to be noted that at some point the type of the systems that can be interlinked start to be very dependent on the enterprise profile, size and a number of other aspects that determine what kind of IT support systems are used (e.g. a software development company will use different systems to support their management process than a hardware design company). Moreover, while implementing the use cases in practice (see Sec. 7.4.3), we noticed that the amount of data and its growth rate in correlation to amount of information submitted to the Idea Management System plays an important role for effectiveness of integration in terms of benefits delivered (e.g. it makes little sense to integrate a bug tracking system that produces a significantly smaller rate of bugs in time than the efficiency of IMS in terms of the implemented ideas). We list the most important systems for idea management per each category, describe their current status with respect to ontologies (see Table 7.2) and later, on top of the presented classification, we chose a particular scenario and detail it on data level so that it can be an inspiration for other cases as well.

Table 7.2: Ontologies for Enterprise Systems

Scope	System Name	Acronym	Goal	Ontology
Internal	Idea Management System	IMS	Collect and manage ideas	Gi2MO (Westerski, 2012a), IO (Riedl, 2009)
	Global Approaches to Enterprise Management			
Enterprise	Enterprise Resource Planing	ERP	Manage business execution	REA (Gailly and Poels, 2005; Bialecki, 2001; O'Leary, 2004), TOVE (Fox and Gruninger, 1998), EO (Uschold et al., 1995), E3 (Gordijn and Akkermans, 2002), BMO (Osterwalder and Pigneur, 2002)
Enterprise	Product Life Cycle Management	PLM/PLCM	Manage product development and engineering	SOM based (Khanna, 2003)
	Specialised (Dedicated) Systems for Enterprise Management			
Enterprise	Client Relationship Management	CRM	Manage input from customers	CMMI based (Lee et al., 2007), O-CREAM (Magro and Goy, 2008), Customer Ont. (Liwen and Min, 2004)
Enterprise	Supply Chain Management	SCM	Manage the flow of products, services	SCOR based (Yiqing et al., 2009), SCM Ontologies (Chandra and Tumanyan, 2004; Khanna, 2003)
Enterprise	Project Management	PMS	Plan the project, assign tasks and set deadlines	PROMONT (Abels et al., 2006), PMO (PMO, 2010), IT-CODE (IT-Code, 2010), DOAP (Dumbill, 2012)
Enterprise	Human Resources Management System	HRMS	Gather information about employees	Organization ontology (Reynolds, 2012), Reference Ontology (Gómez-Pérez et al., 2007), ResumeRDF (Bojars and Breslin, 2007)
Enterprise	Collaborative Working Environment	CWE	Share documents and information	SIOC (Lee et al., 2008)
	Product Development Support Systems (Examples for Software Development)			
Enterprise	Bug-tracking System	-	Collect and organize issues	BAETLE (BEATLE, 2010)
Enterprise	Software Configuration Management	SCM	Manage configuration aspects	SCM ontologies (de Oliveira Arantes et al., 2007; Shahri et al., 2007)
Public	Blog/Forum/Lists	-	Publish information and engage into discussions	SIOC (Breslin et al., 2005; Fernandez et al., 2008)
Public	Idea Management System	-	Collect and manage ideas	Gi2MO (Westerski, 2012a), IO (Riedl, 2009)
Public	Social Networks	-	Connect with other people and publish/access personal data	SIOC (Breslin et al., 2005), FOAF (Brickley and Miller, 2010)
Public	Wikis	-	Publish information and collaborate on improving it	SWIVT (Krotzsch and Vrandecic, 2012; Krötzsch et al., 2005)
Public	Online Patent Databases	-	Collect and publish patent information	PSO (Giereth et al., 2007)
Public	Mindmapping	-	Create and publish mind maps	Mindraider ontology (Dvorak, 2010)

Scenario case study: Interlinking Innovation Data with Human Resources Management System

In the following scenario we aim to extract employee characteristics from the Human Resources Management System (HRMS) and try to connect it to the data produced in Idea Management System (IMS) so that we can deliver some additional benefits. The common denominator of both systems is the concept of the person therefore we can attempt to draw scenarios based on integration on the level of personal profile. In terms of Semantic Web this is most often achieved by using the FOAF ontology for both systems and interlinking the common profile with data in each system (see Fig. 7.5). The technical particularities of establishing links in each of the systems can be solved by using certain dedicated ontologies (e.g. Gi2MO (Westerski, 2012a) for Idea Management System and Organization Ontology (Reynolds, 2012) together with ResumeRDF (Bojars and Breslin, 2007) for HRMS).

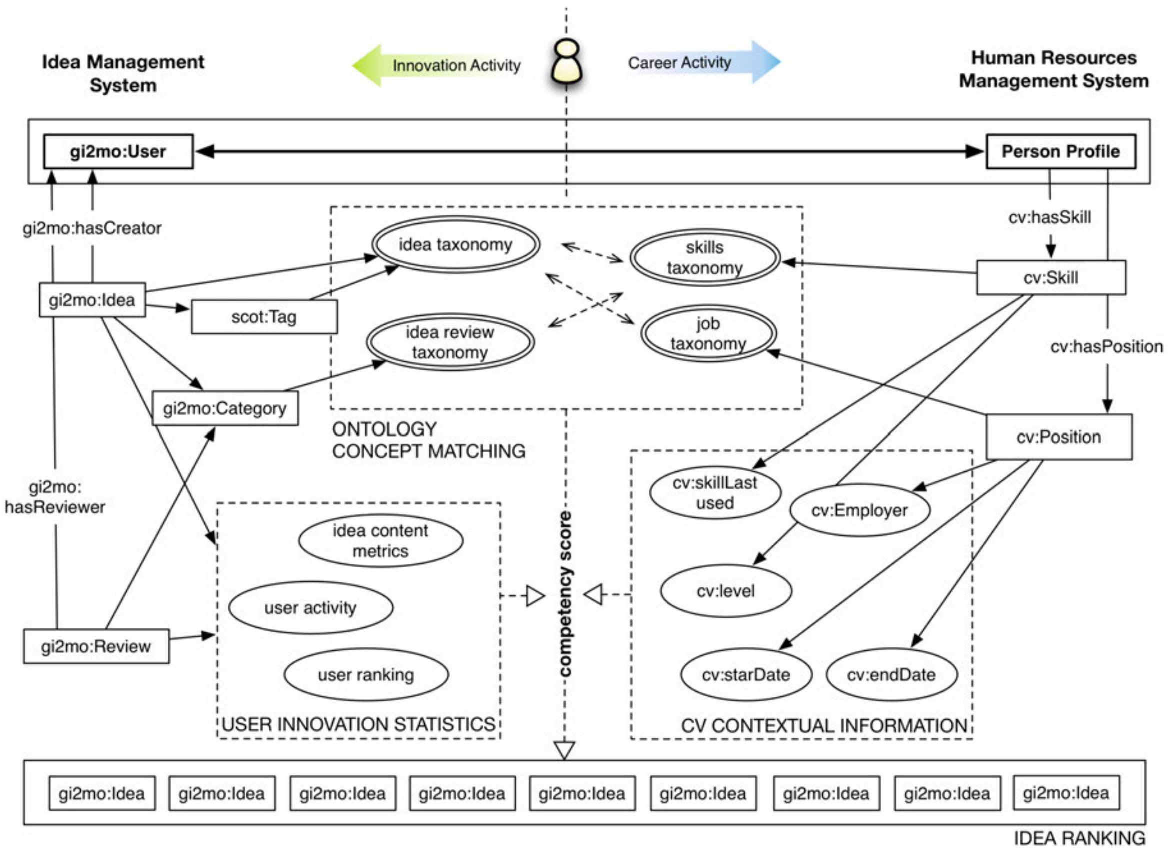


Figure 7.5: Using links between HRMS and IMS to get deeper understanding of the innovator profile and asses ideas by competencies

Using the links established in such a manner it is possible to relate ideas of given characteristics with personal skills, competencies etc. to achieve a number of goals, for example:

- assess ideas based on the competencies, experience and skills of the person that submitted the idea

- recommend idea reviewers based on their skills and relation to idea topic
- judge the efficiency of idea reviewers or idea submitters based on their activity in the IMS and regular projects of the enterprise (e.g. to promote people who are clearly more active than others in many areas or to see if employees from certain departments are better for participation in the innovation efforts).

7.4.3 Extending the Ontology to facilitate interlinking scenarios

Following the analysis of enterprise systems and their dedicated ontologies, we continued by relating those systems to idea management through enumerating metrics that could be extracted from each and developing the necessary Idea Management ontology extensions that would facilitate the data integration (see Table 7.3).

During our research we encountered a number of problems related to activities of interlinking independent systems using Semantic Web technologies:

- in a number of cases data can be linked indirectly (e.g. bugs linked to ideas via project management system). However this creates a problem when a certain system is not present in particular company environment.
- should the links be established via a single property (could result in big number of properties) or via classes that describe type of relation and additional characteristics
- should there be individual properties for links with every kind of system or generic ones (eg. gi2mo:hasRelated). In case of generic ones the ontology is more simple but processing data becomes more complex (e.g. type of relation can be identified in SPARQL query by checking rdf:type).
- the ontologies established over the past years for enterprise systems were not created with the intent to expose structured data but to perform very specialised tasks related to knowledge management within the scope of those systems. Therefore, not only those ontologies do not facilitate interlinking but in addition often are insufficient for publishing even the most basic data of the systems.
- adding new properties and extending the ontology makes it more powerful and useful but at the same time more complex and harder to comprehand by non Semantic Web experts, while the core design assumption for Gi2MO is to maintain a simple and easy to implement data schema (Westerski et al., 2010).

The choices that we have made in terms of the above problems are reflected in particular decisions for Gi2MO ontology enhancements presented in Table 7.3. Following the original design assumptions of the Gi2MO ontology in most cases we opt for making the data schema as simple as possible even at the cost of increasing the complexity of SPARQL queries required to extract the data. The ontology extensions presented in Table 7.3 lay the foundations

Table 7.3: Linking Idea Management with Other Systems

System	Link Example	Metric Example	Gi2MO Properties
Internal IMS Assets	Link ideas based on similarty (eg. duplicates, similar topic, one idea part of another etc.)	Amount of similar ideas (e.g. with a certain degree of similarity)	gi2mo:hasRelated, gi2mo:hasSimilar, gi2mo:describesPartOf, gi2mo:hasDuplicate
	Global Approaches to Enterprise Management		
ERP	Link idea to financial data of processes that implements it	Return of Investment for particular implemented ideas	gi2mo:hasRelated
PLM	Link ideas to products that implement them	Amount of resources involved in product engineering	gi2mo:hasImplementation, gi2mo:hasRelated
	Specialised (Dedicated) Systems for Enterprise Management		
CRM	Link ideas to client complaint/ suggestion logs	Amount of complaints filed for a product that evolved from Idea Management	gi2mo:hasRelated
SCM	Link ideas to supply chain activities that occurred during sales of products based on ideas	Average delay in product deliveries based on certain idea	gi2mo:hasRelated, gi2mo:hasImplementation
PMS	Link ideas to projects	Time beyond set deadline that it took to develop certain product	gi2mo:hasRelated, gi2mo:hasImplementation
HRMS	Link ideas to people in the company that are responsible for different aspects	Employment duration in the company for idea reviewers	via persons's foaf:Agent having OnlineAccount in both systems
CWE	Link ideas to documents and discussions that occur in the company	Amount of discussions regarding product based on idea	gi2mo:hasOrigin, gi2mo:hasRelated
	Product Development Support Systems (Examples for Software Development)		
Bug-tracking	Link ideas to bugs that were submitted in relation to their implementation	Amount of bugs submitted to a product that implements certain idea	gi2mo:hasImplementation to project instance or gi2mo:hasRelated directly to bug
SCM	Link ideas to software projects that implement them	Amount of commits in time for changes based on idea category	gi2mo:hasRelated, gi2mo:hasImplementation
Blog/ Forum/ Lists	Link ideas to posts that discuss them	Amount of comments for post related to idea	gi2mo:hasRelated, gi2mo:hasOrigin
Separate IMS Instances	Link the same ideas across different language versions of the IMS deployed by a single company	Amount of ideas in external systems related to certain idea	gi2mo:hasRelated, gi2mo:hasOrigin
Social Networks	Link ideas to posts that describe their topic	Amount of comments on the topic related to idea	gi2mo:hasRelated, gi2mo:hasOrigin
Wiki	Link ideas with wiki pages on which the ideas are further developed	Number of revisions of a wiki page that describes an idea	gi2mo:hasRelated, gi2mo:hasOrigin
Patent Databases	Link ideas to patents that describe similar topics	Amount of patents that cover the idea	gi2mo:hasRelated
Mindmaps	Link ideas to particular mindmaps that describe them	Amount of concepts that create the idea	gi2mo:hasRelated, gi2mo:hasOrigin

for experimenting with different integration scenarios and utilizing extensive links spanning across a number of systems to evaluate the benefits gained from particular datasets. As an example we detail one of such evaluation activities in the next section.

7.5 Evaluation

7.5.1 Usage of relationships between ideas for idea clustering

In the previous sections we have shown that relationships between knowledge concepts have been investigated in many different domains and with different results. Furthermore, with regard to our own contributions, we have shown that relationships in IMS can be modelled with different granularity and scope depending on the goal. In the following subsection, we focus on evaluation of our first contribution - the hierarchy of relationships between ideas inside the Idea Management System. Moreover, we focus only on certain aspects of introducing the relationship hierarchy and therefore to make our goals more clear we have defined the following hypotheses:

H1. The semantics of idea relationships are more complex than duplicate relationship.

H2. Wider range of relationships can be used to summarize datasets better than with just duplicate relationship.

H3. Apart of idea topic there are idea characteristics (e.g. innovation type) and idea text features that have an impact on how idea annotators perceive the type of relationship between ideas.

With **H1** we put forward a hypothesis that duplicate relationship is insufficient to describe all relationships between ideas stored in the Idea Management Systems. To verify this hypothesis we propose to annotate a subset of ideas using a broad set of relationships identified during our research and compare the results to annotations done only when a duplicate relationship was available.

With **H2** we suggest that the newly proposed relationships, once applied as annotations to ideas, can help in information summerization and would allow to aggregate more ideas than it was possible before when just using the duplicate relationship. To evaluate that hypothesis we propose to reuse the annotations from the previous experiment and aggregate ideas based on their similarities, inheritance of relationship types and transitivity of relationships.

Finally, with **H3** we suggest that annotators pick relationships for ideas not only as a function of similarity on the topic level (e.g. discussing the same product) but also based on the scale of innovation that an idea proposes, how detailed the description is etc. To verify this hypothesis, we refer to our previous research on innovation taxonomies (see Chap. 6) and compare the idea relationship annotations from previous hypotheses experiments with idea similarity expressed with metrics derived from annotations with taxonomy terms for describing idea characteristics.

Experiments setting and data preparation

During the evaluation stage we conducted three experiments, one per each hypothesis. The content used for all experiments was taken from Ubuntu Brainstorm Idea Management System (see Table 7.4). The distinctive characteristics of this system are:

- the topic of all ideas is improvement or introduction of innovations into an open-source linux operating system distribution, related products and services.
- users submit not only ideas but also solutions. The original creator of an idea posts the

first solution and afterwards any member of the community is allowed to add his own solution for implementing the same innovation.

Table 7.4: Ubuntu Brainstorm dataset statistics

Metric	**Metric Value**
Idea number	21690
Comments number	133090
Users number	10062
Implemented Ideas number	576
Amount of Votes cast	2608917

During the experiment we collected all data of the Ubuntu BrainStrom instance and imported into our own system (IdeaStream, 2012). Next, a single annotator was asked to provide relationships for 200 ideas that included: 120 random selected ideas, 40 ideas that have been implemented, 10 top rated ideas, 10 lowest rated ideas, 10 top commented ideas, 10 least commented ideas. The annotator was only presented the idea text (without the complementary solutions). Per each idea the annotator was presented with 5 similar ideas for which he had to specify the relationships (see Fig. 7.6). The similar ideas were selected by the system based on Lucene keyword similarity algorithm (McCandless et al., 2010) run over the index of all 21690 Ubuntu ideas. As a result, we obtained annotations for 1000 idea relationships. This data was used in each of the following hypotheses evaluations.

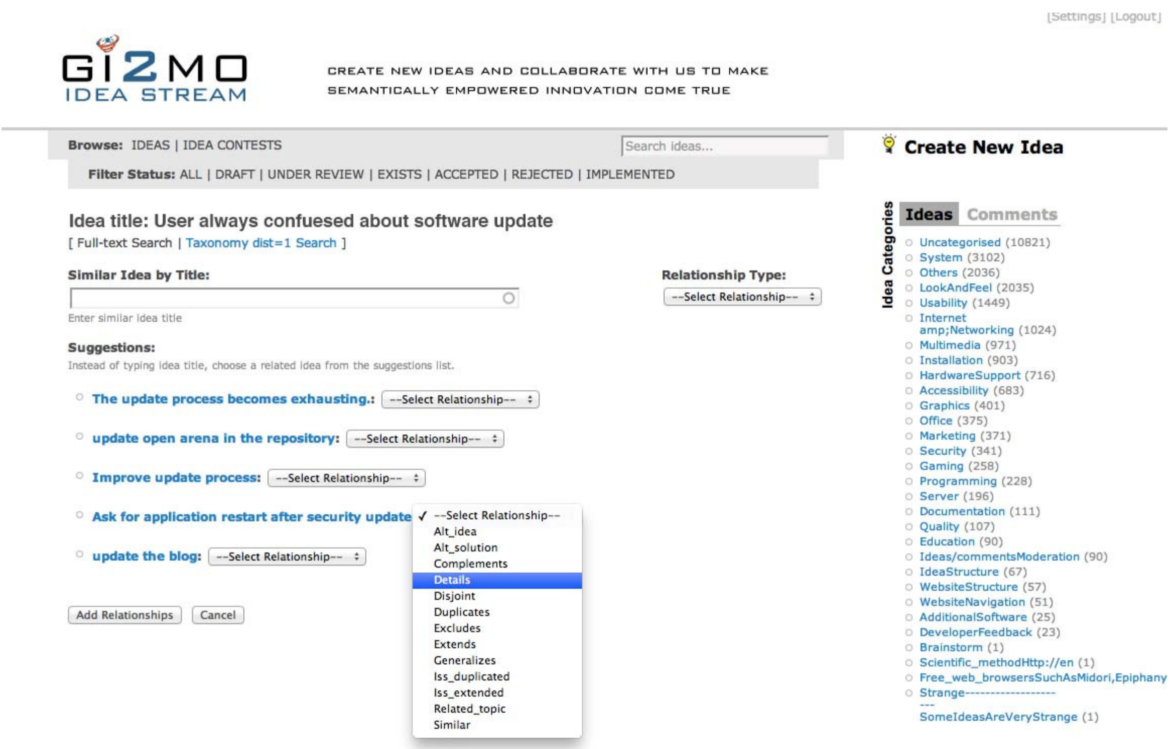


Figure 7.6: Annotator using a tool for similar idea detection.

Experiment I: Relationship amount comparison

In order to determine if any other relationship apart of duplicate is valid we checked if among the obtained relationships were any other than duplicates as well as compared the amount of particular relationships. Apart of data obtained during our own annotation experiment, we also compared our results with the duplicate annotations already present in the online version of Ubuntu Brainstorm (limiting the maximal amount of duplicates to 5 per idea just like we did in our own experiment). As a result, within our own annotations the duplicate relationship accounted for only 25% of all annotations. In comparison to the Ubuntu community annotations we got an increase of 76.7% in relationship count in favour of our solution with regard to what was available before. The detailed results can be observed in Table 7.5.

Table 7.5: Comparison of relationship count across different experiments (200 ideas annotated, 5 relationships max. per each, no inheritance or transitive relationships reasoning)

Ubuntu Brainstorm			
Duplicate	249		

Gi2MO Relationships			
total 440 (328 without duplicates)			
similar	4	disjoint	558
related object	111	extends	2
is extended	3	complements	0
details	136	generalizes	27
excludes	0	alternative solution	19
alternative idea	26	duplicates	112
is duplicated	0		

As an outcome of this experiment we can confirm that introducing new relationships resulted in more metadata and annotators taking advantage of the new power they were given. Therefore, based on the presented results, hypothesis **H1** is supported.

Experiment II: Idea Relationships and Clustering

Idea aggregation for information summarization

In the previous experiment we have shown that by introducing a more broad set of relationships we were able to obtain a much bigger amount of annotations. Nevertheless, this does not imply that the amount of unique similar ideas would rise in the same degree (different relationships can point to the same ideas).

Therefore, to answer a question if annotations made with the new set of relationships would allow to summarize the data more than just the previously present duplicate relationship, we processed the annotated Ubuntu dataset by aggregating all similar ideas into a single one (just as it is done in the contemporary systems when duplicate ideas are detected). In contrast to the previous experiment the main difference is that we count the amount of unique ideas that can be aggregated rather than total number of relationships obtained. In particular, we analysed the amount of unique ideas that could be aggregated in relation to entire dataset size. The results were: 1.13 % of dataset were duplicates that could be aggregated based on Ubuntu community annotations, 0.5 % of dataset aggregated based on duplicate relationships from our experiment and finally 1.95 % of dataset aggregated while using the full relationship hierarchy and aggregating all similar ideas. This indicates that the summerization of our solution with respect to Ubuntu community annotations gave a 95% increase.

Additionally, to see if the summierization ratio of our solution could be further improved, we analysed two ways of extending the knowledge base using inference of:

- inheritances between relationship types, e.g. when aggregating *complementary* ideas also *details* and *generalizes* annotations are taken into account
- transitive relationships, e.g. if A *extends* B and C *extends* B, than both B and C are aggregated into A

To compare all three options (user made annotations, inherited relationships, transitivity of relationships), we defined **idea aggregation rating** metric that states how many unique ideas have been aggregated per a single idea in the system (see Table 7.6). Observing the results, it can be seen that using transitivity gives a significant increase of ideas aggregated which can be particularly seen for top relationships in the hierarchy when relationship inheritance is applied (e.g. amount for similar ideas aggregated change from 0.02 per idea to 3.37 after applying inheritance and inferring related ideas via transitive relationships).

Table 7.6: Idea aggregation ratio in different inferencing scenarios

Relationship	**No Inheritance**		**Inheritance**	
	No Transitivity	Transitivity	No Transitivity	Transitivity
Similar	0.02	0.02	2.85	**3.37**
Related Object	0.72	0.75	0.72	**0.75**
Extends	0.01	0.01	1.39	**1.52**
Complements	0	0	1.06	**1.18**
Details	0.87	0.90	0.87	**0.90**
Generalizes	0.18	0.21	0.20	**0.21**
Excludes	0	0	0.29	**0.30**
Alternative Solution	0.12	0.12	0.12	**0.12**
Alternative Idea	0.17	0.18	0.17	**0.18**
Duplicates	0.71	0.71	0.71	**0.73**

Taking into account quite a considerable difference in dataset summerization that is mainly the result introducing new relationships but also the application of logic operators for relationships, we can state hypothesis **H2** as supported in the conditions of our experiment.

Experiment III: Idea metric similarity and its impact on idea relationships

In the final evaluation, we aimed to verify if the similarity as perceived by idea annotators is related to characteristics of ideas other than idea topic. To define those characteristics we referred to our previous research on idea classification and reused the Gi2MO Types taxonomy (Westerski, 2012c) that advocates the use of 4 main idea characteristic areas:

- trigger type (type of event that caused the creation of idea, e.g. faulty experience or improvement request)
- innovation type (how much innovative is the idea, e.g. radical innovation vs. incremental innovation)
- proposal type (how broad is the description, e.g. full solution vs. bug report)
- object type (is innovation proposed for object or service, is it a complete change or element change etc.)

All together Gi2MO Types delivers 74 terms grouped into the above 4 categories. Based on presence or absence of particular terms in idea annotations Gi2MO Types defines 14 metrics that characterise an idea. For our experiment, we annotated ideas with Gi2MO Types taxonomy terms and calculated the aforementioned metrics. The metric similarity S between two related ideas i_A and i_B was calculated individually per each metric M_x as an absolute difference of a given metric value for two related ideas:

$$S(i_A, i_B, M_x) = 1 - |M_x(i_A) - M_x(i_B)| \tag{7.1}$$

The calculations, as described above, were made for a subset of 50 ideas taken from the previous evaluations, in particular: 10 implemented, 10 top rated, 10 most down-ranked, 10 top commented, 10 least commented (but having at least 1 comment). Using the taxonomy we annotated the those 50 ideas as well as all their related ideas with Gi2MO Types terms that identified their characteristics. As a result we got 250 annotated ideas (each of the 50 ideas had 5 related ideas).

Having such dataset we analysed the correlations between presence or absence of particular idea relationships and the idea characteristic similarity expressed with Gi2MO Types metrics. The results are presented in Table 7.7.

Table 7.7: Bivariate correlations between the Gi2MO Types metric value similarities and top level relationships from the proposed hierarchy

Metric/Relationship	Similar	Disjoint	Duplicate	Related Object	Extends	Complements	Excludes
Completeness	**0.18**	**-0.19**	0.08	0.03	**0.11**	0.08	0.07
Experience Completeness	-0.04	0.06	0.07	0.01	-0.09	-0.07	-0.04
Situational Dependence	0.01	-0.03	0.02	0.05	-0.04	-0.05	0.01
Relatedness	0.01	-0.02	0.11	-0.04	-0.07	-0.02	-0.08
Adaptiveness	**0.18**	**-0.20**	0.01	0.01	**0.18**	**0.12**	**0.11**
Originality	**0.15**	**-0.17**	0.06	0.05	0.09	0.05	0.07
Originality Scope	0.06	-0.01	0.08	**-0.14**	**0.13**	0.06	-0.05
Cooperativeness	**0.14**	**-0.14**	**0.12**	0.03	0.04	0.06	-0.02
Freshness	0.09	-0.09	0.01	**-0.11**	**0.15**	**0.10**	0.09
Integrability	**0.24**	**-0.25**	0.08	0.02	**0.19**	**0.16**	0.07
Applicability Scope	**0.14**	**-0.14**	-0.01	0.01	**0.10**	-0.06	0.08
Constructiveness	-0.02	-0.01	0.01	-0.05	0.00	**0.14**	0.08
Scope	**0.11**	**0.13**	-0.01	0.09	0.07	0.09	-0.02
Object Dependability	**0.24**	**-0.21**	-0.01	0.01	**0.25**	**0.21**	**0.11**

To analyse the results we used Cohen correlation scale for social sciences (Cohen, 1988). According to that scale the majority of correlations between relationships and idea characteristics are irrelevant. More precisely, only 48 correlations out of 180 can be described as small (between 0.1 and 0.3). Most of those small correlations for a single relationship have been observed in case of disjoint and similar relationships (8 metrics out of 14 possible).

Taking into account the presented results the final hypothesis **H3** is proven as unsupported in the conditions of the conducted experiment.

7.5.2 Usage of enterprise data interlinking for generating idea metrics

In the following section we compliment our study on linking ideas with enterprise data by presenting the results of our work implemented in practice. As mentioned earlier (see Sec. 7.4.1) our primary motivation with regard to enterprise linked data is extracting innovation metrics. A popular way the metrics are utilized in the contemporary Idea Management Systems is in data visualisations. Therefore, to prove that the metrics that we have pointed can be extracted in practice, we followed this notion of data visualisation and constructed an application called Idea Analyst (Westerski, 2012d) that would map data extracted with SPARQL queries from distributed datasets to bubble charts (see Fig. 7.7).

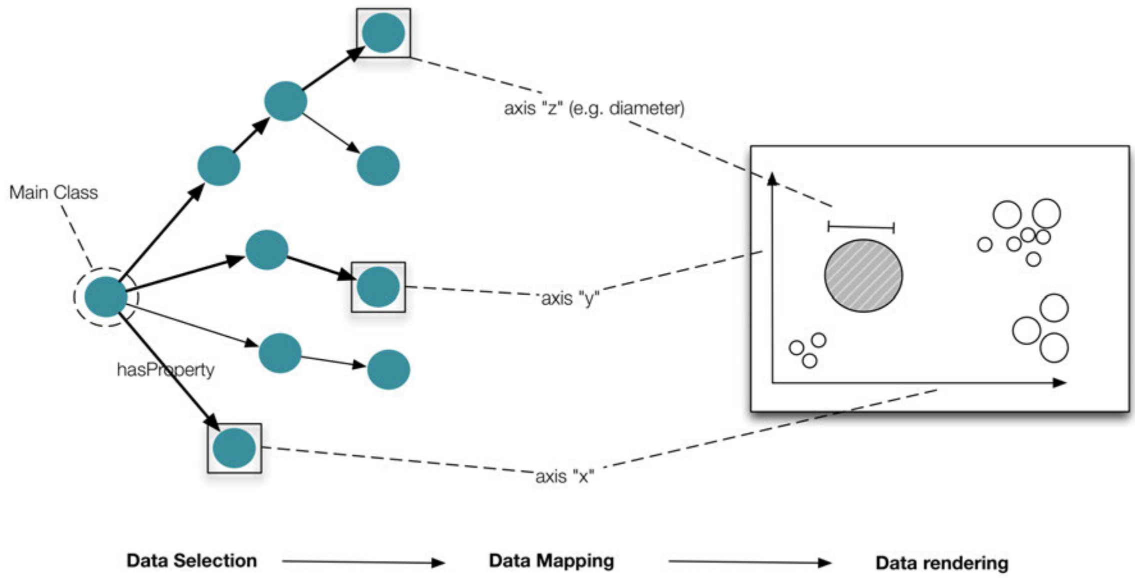


Figure 7.7: Mapping of RDF encoded data into a bubble chart.

In our implementation data series for a multidimensional diagram are first extracted independently from each of the datasets and then bound together by a common concept that must be present in each result set. For example values used to visualise the radius of spheres plotted onto the chart have to refer to the same root property in the Idea Management System (e.g. idea URI) as values extracted by another query that delivers sphere fill color values. Furthermore, as we noticed when working with particular datasets, most of the data that is published in the linked data cloud as well as web systems related to idea management is not numerical. Therefore, one important observation is the necessity of using the SPARQL endpoint implementation that supports aggregate functions (COUNT, SUM, MAX etc.).

Having met those requirements, we used Idea Analyst to experiment with the interlinking scenarios proposed in the previous sections. We present one of them - extracting metrics derived from the integration between Idea Management System and HRMS.

In this case study rather than assessing ideas we use the data to recognize the **effectiveness of employees as ideas authors**. This is visualised by comparing the amount of skills that employees have to the amount of ideas that they created and amount of those ideas that have proven successful enough to get implemented.

The main ontologies used are: Gi2MO for Idea Management System and ResumeRDF (Bojars, 2007) for HRMS. The idea management dataset comes from one of the publicly available instances (Adobe, 2012a), whereas the HRMS dataset was prepared manually without relation to any particular system.

To visualise the data we created a 2 dimensional bubble chart with two data series mapped to x and y axis while the third data series is mapped as the sphere diameter.

```
I. Y axis: amount of implemented ideas per author (IMS)

prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix gi2mo: <http://purl.org/gi2mo/ns#>
SELECT ?foaf_person_uri, COUNT(?idea_uri)
FROM <http://lab.gsi.dit.upm.es/~adam/ld_test/test_adobe_uc3.rdf>
WHERE {
?idea_uri rdf:type gi2mo:Idea .
?idea_uri gi2mo:hasStatus gi2mo:Implemented .
?idea_uri gi2mo:hasCreator ?idea_author_uri .
?idea_author_uri gi2mo:isAccountOf ?foaf_person_uri
} GROUP BY ?foaf_person_uri

II. X axis: amount of skills per author (HRMS)

prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix cv: <http://kaste.lv/~captsolo/semweb/resume/cv.rdfs#>
prefix cvx: <http://gi2mo.org/hrms_cvx#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?foaf_person_uri, COUNT(?skill_uri)
FROM <http://lab.gsi.dit.upm.es/~adam/ld_test/test_skills.rdf>
WHERE {
?skill_uri rdf:type cv:Skill .
```

```
?hrms_cv_uri cv:hasSkill ?skill_uri .
?hrms_user_uri cvx:hasCV ?hrms_cv_uri .
?hrms_user_uri cvx:hasAgent ?foaf_person_uri
} GROUP BY ?foaf_person_uri

III. Diameter: total amount of submitted ideas per author (IMS)

prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix gi2mo: <http://purl.org/gi2mo/ns#>
SELECT ?foaf_person_uri, COUNT(?idea_uri)
FROM <http://lab.gsi.dit.upm.es/~adam/ld_test/test_adobe_uc3.rdf>
WHERE {
?idea_uri rdf:type gi2mo:Idea .
?idea_uri gi2mo:hasCreator ?idea_author_uri .
?idea_author_uri gi2mo:isAccountOf ?foaf_person_uri
} GROUP BY ?foaf_person_uri
```

The end result is a bubble chart where it can be observed that on top of the huge number of ideas that never get implemented the two most valuable groups of employees for the companies innovation policy are: people with very little technical knowledge but huge motivation (a large number of submitted ideas) and very skilled people that share just a few ideas but almost always are successful (see Fig. 7.8).

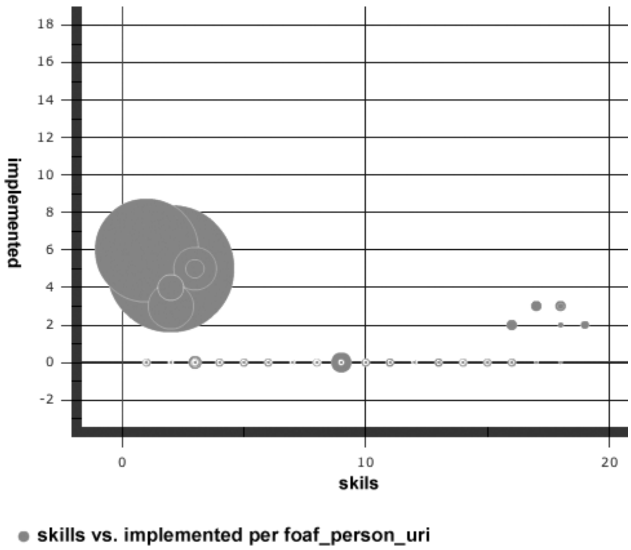


Figure 7.8: Bubble chart generated by the Idea Analyst for the presented SPARQL queries.

7.6 Summary

In this chapter we have presented an analysis of relationships of the information stored in an Idea Management System. Firstly, we investigated the context of a single Idea Management System deployment and proposed a classification of relationships that emerge as a result of the collaborative open innovation processes. The contribution of this part is a hierarchy of relationships between ideas and a number of observations towards idea clustering that can be evaluated with the proposed hierarchy.

Secondly, we looked at a more broad scope of the information relationships in the entire organization ecosystem. We investigated the connections that emerge between the ideas of Idea Management Systems and concepts of other data silos in the enterprise or Internet space. In particular, we have contributed: 1) a classification of such related information systems; 2) proposed a methodology for extending the Gi2MO ontology to facilitate establishment of relationships in a form of metadata links; 3) proposed transformation of the metadata links between ideas and enterprise data into idea metrics.

The aforementioned contributions propose certain solutions to the problems put ahead of this thesis, i.e. information summarization and data assessment. Therefore, each requires an evaluation in order to be validated and judged in terms of contribution significance. The evaluation activities in case of mutual idea relationships in a single IMS have shown that the current state of the art system loose a huge amount of information by neglecting or simplifying the semantics of relationships. Namely, by introducing our hierarchy of relationships the amount of new relationships detected rose by 76% in comparison to traditional solutions for duplicate detection. Furthermore, using the logical dependencies between proposed relationships (e.g. relationship inheritance or transitivity) we proven that the contemporary clustering methods can be improved by 95% in terms of dataset summarization. Finally, in terms of Enterprise Linked Data metrics, we have presented a prototype and a number of data processing queries that delivered proof that our solution is not only theoretically feasible but also technically.

Chapter 8

Conclusions

This thesis has presented a number of solutions that contribute to the area of knowledge management in Idea Management Systems. In the following chapter those contributions are summarized and final conclusions are presented.

Furthermore, having taken account of the achievements of the thesis, this chapter presents the possible future extensions of the research done.

The main topics discussed in this chapter are:

- conclusions regarding four key areas of thesis contritions: generic model for Idea Management Systems, community model, idea characteristics and idea relationship models

- possible future lines of research and new possibilities that the thesis contributions open

8.1 Conclusions

The research for the goals of this thesis has begun in late 2009. Up until that time Gartner has noted in a number of annual reports that the Idea Management Systems are maintaining a stable position on the market of rising technologies, yet struggle to move into the so called 'Slope of Enlightenment' where the technologies gain traction and adoption in a global sense. Similarly as many other promising new technologies, Idea Management Systems in their state back then faced many technical problems and very few academic researchers tried to address those issues. After approximately three and a half years in development, this thesis has managed to establish certain contributions to the state of the art of Idea Management Systems but it also has to be noted that the area into which the thesis contributes has changed in many ways. Concluding the thesis, we summarize the contributions and reflect on the significance of each with respect to the point of departure as well as current state of the art and concurrent developments of other researchers.

The research presented in this thesis set out with a goal to identify the most critical problems of Idea Management Systems, select specific issues to resolve and contribute solutions that would allow to progress the state of the art in the area. Based on the analysis of the domain and available literature, idea assessment was selected as the key problem area for the thesis. Throughout the course of the thesis, a number of contributions have been delivered that can be gathered under **four main contribution areas**:

- **Generic Model of Idea Management Systems** - the thesis proposed a framework that binds all data creation processes of Idea Management System in a cycle. Based on this proposal the thesis analysed the metadata of the contemporary Idea Management Systems and verified the usefulness of this information for idea assessment. Based on that research, the final contributions in the problem area of describing the contemporary state of Idea Management Systems are: 1) a theoretical Idea Life Cycle proposal; and 2) its formalization in a form of ontology. The final conclusions obtained via evaluation of those contributions are: 1) data of IMS is highly interconnected and mutually dependent; 2) formalization of this data using proposed ontology and Semantic Web modelling techniques allows to cover majority of information from the contemporary IMSes and benefit in the area of system integration to assess ideas in a single environment and compare different instances based on contemporary metrics; 3) yet at the same time there is very little machine processable metadata available in the contemporary systems that would enable to identify idea relationships in a systematic manner. As a result of those conclusions the thesis contributed in three more areas to obtain more metadata that would allow better comparison and assessment of ideas.

- **Community Opinions Model for Idea Management Systems** - the research on community generated content of the thesis points to little use so far of feedback of users in form of idea comments. The thesis has contributed a proposal to use opinion mining technique in idea management to analyse this community created data and generate additional metrics for idea assessment. The results of evaluation of thesis solutions have shown that the utilized technology and solutions build on top of it deliver new information for decision makers that have impact on their judgement with regard to acceptance or rejection of ideas. Furthermore, the thesis has contributed a model for modelling opinions by the means of an ontology and connected this model to the previously introduced ontology for IMS. On top of the evaluation for the use in Idea Management Systems, the results of experiments show that the proposed model can be used in other web systems that invite community into a deliberation process and cover majority of data generated by available opinion mining solutions.

- **Idea Characteristics Model for Idea Management Systems** - the thesis has proposed usage of innovation models and theories on characteristics of innovation. Based on the previous models the thesis has identified the characteristics specific for ideas in the Idea Management Systems, however independent of the domain or market segment in which the system in deployed. Furthermore, the thesis has contributed a number of experiments which have shown that such characteristics can be applied both manually and in an automatic manner using a machine learning approach. Finally, based on the proposed idea annotations, the thesis has delivered a study of transforming annotations intro metrics that identify information stored in the IMS. The conclusions of this part are that metrics derived from innovation models are equally relevant to identification of winning ideas in Idea Management Systems as any other currently used community metrics (up/down ratings, comment count etc.) but do not show any significant improvement. Nevertheless, results of evaluation have shown the added value of presented model in terms of analysis and comparison of various IMS deployments. The proposed metrics allow to identify community behaviour to verify if submitted ideas follow the initially set goals of organizers. Furthermore, since the metrics scales are not related to any particular domain or organization, the proposal has proven successful for comparison of different IMS instances and their communities.

- **Idea Relationships Model for Idea Management Systems** - the two previous contributions areas related to obtaining new information about ideas that would enable their assessment and comparison. The final contribution goes back to the original statement of the thesis regarding idea dependencies and explores further the relationship types as well as identifies if the new metadata obtained via previous contributions can be used to facilitate relationship identification. The contributions in this area are: 1) classification of systems related to Idea Management Systems; 2) a methodology for interlinking those systems by means of extending the previously proposed ontology; 3) hierarchy of relationships between ideas and its use for clustering of ideas. The conclusions obtained based on evaluation of the contributions are: 1) contemporary systems omit the topic of relationships in a huge degree that otherwise, using the proposal of the thesis, enables to improve the current clustering capabilities over 90%; 2) the non-domain idea characteristics derived from innovation models have a rather small impact on idea similarity or dissimilarity and almost no impact whatsoever on how people perceive types of similarity

The order in which the contributions of the thesis have been presented was not accidental. Each of the consecutive contributions builds on the experiences and solutions delivered by the previous. The firstly proposed generic model has created a base for all remaining research of the thesis. The thorough analysis of dependencies between Idea Management processes by the thesis and establishment of Idea Life Cycle has provided the thesis author and other researchers (as proven in the Impact Chapter) with knowledge on how to approach the problem area and classify solutions proposed by either academic researchers or the industry. The generic IMS ontology build on top of that life cycle definition has shown the lacks in metadata of the current systems which led the thesis to three major novel proposals on how to obtain such metadata in order to facilitate better idea assessment.

8.2 Future Work

The development of this thesis and its contributions to the state of the art in Idea Management Systems have opened new possibilities for future research. The experiments conducted have delivered proof for usefulness of certain solutions or excluded particular approaches. Additionally, the related software research prototypes has stimulated development of new ideas for improving Idea Management Systems. In terms of conclusions for the thesis research the following lines of future research can be pointed out:

- **Extended evaluation of opinion mining.** Thesis area: community opinions. The metrics generated based on use of opinion mining were tested only with a single IMS instance. It would be desirable to perform similar experiments but with a wider range of systems. The experiments with idea characteristics have shown the communities gathered around IMSes and their contributions may differ to a large extent, therefore it is possible that usefulness of contributed opinion metrics might be different depending on the system.

- **Further research on automatic idea annotation.** Thesis area: idea characteristics. The thesis has experimented with automatic idea annotation and obtained satisfactory results for selected parts of the proposed taxonomy. However, the presented evaluation took into account only a single method for automatic annotation: supervised machine learning approach based on k-NN algorithm with nearest neighbours detected using keyword similarity. It remains to be seen whether other methods would enable to achieve better results and recommend automatic annotation for the full taxonomy. In light of evaluations that have proven usefulness of the idea characteristics for dataset comparison, investing more effort in research on automatic appliance of those characteristic annotations would improve the contribution of the thesis in a significantly.

- **Clustering based on idea characteristics.** Thesis area: idea characteristics. The thesis has delivered a number of studies on idea characteristics and quantitative analysis of the characteristics. However, the remaining future work is to see how similar are ideas with respect to more than a single characteristic. A potential direction for this kind research could be the use of clustering algorithms and treatment of idea characteristics as feature vectors. It remains to be seen if such methods could deliver distinctive clusters of ideas that could have a meaningful impact on analysis of idea datasets.

- **Idea annotation with domain ontologies.** Thesis area: idea characteristics, idea relationships. The thesis has studied automatic annotation in the context of domain independent taxonomy in order to deliver a tool for comparison of different IMS deployment. As an extension of this work, the thesis also investigated impact of those annotations on idea similarity. In terms of future work, a possible line of research would be to evaluate the use domain ontologies in the same way, i.e. automatic annotation of ideas with concepts related to domain, development of metrics based on those annotations, and computation of idea similarity based on domain related annotations.

- **Improved Enterprise Linked Data evaluation.** Thesis area: idea relationships. The major problem of evaluation efforts in the area of idea relationships was lack of sufficient data and large enterprise partners that would share their information for the needs of experiments. In terms of future work, it would be a valuable contribution to evaluate the proposed solution in the environment of a large enterprise that would indeed posses the discussed enterprise systems and be willing to give access to them for experiments.

- **Research on automatic idea relationship discovery and manual annotation.** Thesis area: idea relationships. The results of thesis experiments have shown that there is huge potential beyond the current use of idea relationships. However, the introduction of new relationship types also makes the annotation process more difficult. The thesis delivered experiments with semi-automatic annotation by a single annotator as well as ruled out the use of idea characteristics for automatic discovery of idea relationship type. Potential future lines of research in this area could include: a) research on how to obtain the idea relationships in a fully automatic way (e.g. based on keyword similarity or domain ontologies) b) evaluation of manual annotation with multiple annotators to verify ambiguity of terms in the proposed relationship taxonomy.

- **Usage of the newly discovered idea relationships.** Thesis area: idea relationships. Aside of experiments with idea relationship annotation, the thesis presented results of clustering based on the new relationships. However, it remains a topic of future research if the new relationships could be utilized in other ways like idea recommendation or ranking (e.g. similar as used by web search engines (Brin and Page, 1998) or for social network analysis (Wasserman and Faust, 1994)). The future work in this area should verify if such use of idea relationships would aid idea assessment and if ranking generated based on relationships would have an impact on ideas implemented.

- **Integration of IMS with Social Web.** Thesis area: idea relationships. In scale of the entire Internet, the information submitted to IMS is but a fraction of the feedback published on-line. As part of research on Enterprise Linked data, the thesis contributed a solution for linking IMS with various data silos of an enterprise and pointed to use of public information sources like blogs and forums in terms of future work. A potential line of future work could be related to appliance of business intelligence and NLP techniques to import information from the Social Web into IMS and analyse it together with well formatted ideas submitted by the users. Furthermore, Social Web data could be used for ranking already submitted ideas (e.g. rating ideas based on popularity of idea topic on the Social Web).

- **Automatic idea mashups.** Thesis area: entire framework. The IMSes have a huge base of ideas and our research has shown that those ideas are not only duplicates but are connected to each other in a variety of ways. So far we have experimented with downsizing the idea dataset via clustering but perhaps a viable future lines of research could be allowing users to mashup ideas together from the existing idea database. The room for novelties is quite broad there and could include research on: idea mashup operators, idea similarity for automatic mashup suggestions, metrics for ranking the mashed ideas vs. regular ones, and finally research on incentives and community takeup with relation to reusing ideas of other people.

Concluding the presented lines of future work: the thesis has investigated and proposed solutions for idea assessment in Idea Management Systems but as pointed out in the contribution of Idea Life Cycle, all processes of Idea Management Systems are interconnected and dependent on each other. Therefore, aside of answering new questions that the thesis rose, future work should investigate further the impact of thesis contributions on other Idea Life Cycle phases, with special interest in idea selection and research on how the proposed idea assessment solutions could be implemented as assets for decision makers.

Bibliography

Abel, F., 2011. Contextualization, user modeling and personalization in the social web. Ph.D. thesis, L3S Research Center, Gottfried Wilhelm Leibniz Universitaet Hannover.

Abels, S., Ahlemann, F., Hahn, A., Hausmann, K., Strickmann, J., 2006. Promont - a project management ontology as a reference for virtual project organizations. In: On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. pp. 813–823.

Abernathy, W., Utterback, J., 1978. Patterns of innovation in technology. Technology Review 80 (7), 40– 47.

Abernathy, W. J. (Ed.), 1978. The productivity dilemma: roadblock to innovation in the automobile industry. Johns Hopkins University Press.

Abernathy, W. J., Clark, K. B., February 1985. Innovation: Mapping the winds of creative destruction. Research Policy 14 (1), 3– 22.

AcceptIdeas, 2009. Accept Ideas homepage.
URL http://www.acceptsoftware.com/

Adamides, E. D., Karacapilidis, N., 2006. Information technology support for the knowledge and social processes of innovation management. Technovation 26 (1), 50–59.

Adida, B., Birbeck, M., McCarron, S., Herman, I., 2004. RDFa Core 1.1.
URL http://www.w3.org/TR/rdfa-syntax/

Adner, R., Levinthal, D., 2001. Demand heterogeneity and technology evolution:implications for product and process innovation. Management Science 47, 611–628.

Adobe, 2012a. Adobe acrobat ideas.
URL http://na5.brightidea.com/ct/ctlist.bix?c=8FBBEA8F-D8E6-4E34-A7C1-7C74FB3B4EFA

Adobe, 2012b. Adobe systems homepage.
URL http://www.adobe.com/

Afuah, A., 1998. Innovation management: Strategies, implementation, and profits. Oxford University Press., New York.

Aiken, M., Krosp, J., Shirani, A., Martin, J., 1994. Electronic brainstorming in small and large groups. Information & Management 27 (3), 141 – 149.

Alexandru Spatariu, K. H., Bendixen, L. D., 2004. "defining and measuring quality in online discussions. Journal of Interactive Online Learning 2 (4).

Allsopp, J., 2007. Review and resume microformats: hreview and hresume. In: Microformats: Empowering Your Markup for Web 2.0.

Anadiotis, G., Kafentzis, K., Pavlopoulos, J., Westerski, A., April 2012. Building consensus via a semantic web collaborative space. In: WWW 2012 Companion Proceedings. Semantic Web Collaborative Spaces Workshop (SWCS2012). Lyon, France.

Anastasiou, D., Ramos, L., Krima, S., Chen, Y.-J. (Eds.), 2012. Proceedings of the Second International Workshop on Ontology and Semantic Web for Manufacturing (OSEMA 2012).

Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., Werthimer, D., November 2002. Seti@home: An experiment in public-resource computing. Communications of the ACM 45 (11).

Andrew, J. P., Haanæs, K., Michael, D. C., Sirkin, H. L., Taylor, A., 2009a. Innovation 2009. making hard decision in the downturn. Tech. rep., Boston Consulting Group.

Andrew, J. P., Haanæs, K., Michael, D. C., Sirkin, H. L., Taylor, A., 2009b. Measuring innovation 2009. the need for action. a BCG senior management survey. Tech. rep., The Boston Consulting Group.

Andrew, J. P., Stover DeRocco, E., Taylor, A., 2009c. The innovation imperative in manufacturing. how the united states can restore its edge. Tech. rep., Boston Consulting Group.

Anjomshoaa, A., Karim, S., Shayeganfar, F., Tjoa, A., 2006. Exploitation of semantic web technology in erp systems. In: Tjoa, A., Xu, L., Chaudhry, S. (Eds.), Research and Practical Issues of Enterprise Information Systems. Vol. 205 of IFIP International Federation for Information Processing. Springer US, pp. 417–427.
URL http://dx.doi.org/10.1007/0-387-34456-X_42

Antoniou, G., van Harmelen, F., 2004. A Semantic Web Primer. MIT Press.

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise ? the information content of internet stock message boards. Journal of Finance, 1259–1294.

Applegate, L. M., 1986. Idea management in organization planning. Ph.D. thesis, University of Arizona.

Applegate, L. M., Konsynski, B. R., Nunamaker, J. F., 1986. A group decision support system for idea generation and issue analysis in organization planning. In: Proceedings of the 1986 ACM conference on Computer-supported cooperative work. New York, NY, USA.

Atos, 2010. Atos PGI 2.0 Idea Management Platform homepage.
URL `http://pgi2-en.atosorigin.es/`

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Ives, Z., 2007a. Dbpedia: A nucleus for a web of open data. In: International Semantic Web Conference. Springer, Busan, Korea, pp. 11–15.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Ives, Z., 2007b. Dbpedia: A nucleus for a web of open data. In: In 6th Int'l Semantic Web Conference, Busan, Korea.

Auer, S., Dietzold, S., , Riechert, T., Riechert, T., 2006. Ontowiki - a tool for social, semantic collaboration. In: The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006. Springer, pp. 736–749.

Ayers, D., Heath, T., 2007. RDF Review Vocabulary Specification. http://hyperdata.org/xmlns/rev/.

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., Patel-Schneider, P. F. (Eds.), 2003. The description logic handbook: theory, implementation, and applications. Cambridge University Press, New York, NY, USA.

Baalsrud Hauge, J., Duin, H., Thoben, K. D., 2008. Applying serious games for supporting idea generation in collborative innovation processes. In: Proceedings of the 14th International Conference on Concurrent Enterprising, ICE2008. Lisbon, Portugal.

Baccianella, S., Esuli, A., Sebastiani, F., May 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta.

Baez., M., Convertino, G., 2012. Designing a facilitator's cockpit for an idea management system. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW.

Bailey, B. P., Horvitz, E., 2010. Whats your idea? a case study of a grassroots innovation pipeline within a large software company. In: CHI '10 Proceedings of the 28th international conference on Human factors in computing systems.

Baker, N. R., Freeland, J. R., 1972. Structuring information flow to enhance innovation. Management Sceince 19, 105–116.

Baker, T., Noy, N., Swick, R., Herman, I., 2012. Semantic web case studies and use cases.
URL `http://www.w3.org/2001/sw/sweo/public/UseCases/`

Bakker, H., Boersma, K., Oreel, S., 2006. Creativity (ideas) management in industrial r&d organizations: A crea-political process model and an empirical illustration of corus rd&t. Creativity and Innovation Management 15 (3), 296–309.

Barki, H., Pinsonneault, A., 2001. Small Group Brainstorming and Idea Quality: Is Electronic Brainstorming the Most Effective Approach? Small Group Research 32 (2), 158.

Baumgartne, J., 2008. An introduction to idea management. Tech. rep., Bwiti bvba.

Baumgartner, J., 2004. Big and little innovation, report 103.
URL http://www.jpb.com/report103/archive.php?issue_no=20040427

Beall, J., 2010. How google uses metadata to improve search results. The Serials Librarian 59 (1).

Bean, C. A., Green, R., 2001. Relationships in the Organization of Knowledge. Information Science and Knowledge Management. Kluwer Academic Publishers.

BEATLE, 2010. Bug and enhancement tracking language.
URL http://code.google.com/p/baetle/

Beaudouin-Lafon, M. (Ed.), 1999. Computer Supported Co-operative Work. John Wiley & Sons.

Bejar, I. I., Chaffin, R., Embretson, S., 1990. Cognitive and Psychometric Analysis of Analogical Problem Solving. Springer.

Belecheanu, R., 2009. D6.3.1 SAP Living Lab Environment. Tech. rep., Laboranova Collaboration Environment for Strategic Innovation.

Berger, A. L., Pietra, V. J. D., Pietra, S. A. D., 1996. A maximum entropy approach to natural language processing. Computational Linguistics 22 (1), 39–71.

Berners-Lee, T., 1998. Semantic web road map.

Berners-Lee, T., 2000. Semantic web. In: XML 2000. Washingon, USA.
URL http://www.w3.org/2000/Talks/1206-xml2k-tbl/

Berners-Lee, T., 2006. Linked data. http://www.w3.org/DesignIssues/LinkedData.

Berners-Lee, T., 2009. Semantic web and linked data.
URL http://www.w3.org/2009/Talks/0120-campus-party-tbl/

Berners-lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D., 2006. Tabulator: Exploring and analyzing linked data on the semantic web. In: In Proceedings of the 3rd International Semantic Web User Interaction Workshop.

Berners-Lee, T., Fischetti, M., 1999. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor, 1st Edition. Harper San Francisco.

Berners-Lee, T., Hendler, J., Lassila, O., May 2001. The Semantic Web. Scientific American. URL http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&pageNumber=1&catID=2

Berners-Lee, T., Hendler, J., Lassila, O., May 2001. The semantic web. Scientific American, 29–37.

Bestgen, Y., Fairon, C., Kerves, L., 2004. Un baromètre affectif effectif: Corpus de référence et méthode pour déterminer la valence affective de phrases. In: Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles.

Bialecki, A., 2001. REA Ontology Formalization. URL http://www.getopt.org/ecimf/contrib/onto/REA/

Bikakis, N., Giannopoulos, G., Dalamagas, T., Sellis, T., 2010. Integrating keywords and semantics on document annotation and search. In: The 9th International Conference onOntologies, DataBases, and Applications of Semantics (ODBASE 2010).

Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., Velkov, R., 2011. OWLIM: A family of scalable semantic repositories. Semantic Web Journal 2 (1).

Biz360, 2012. Biz360 homepage. URL http://www.biz360.com/

Bizer, C., 2004. D2RQ Treating Non-RDF Databases as Virtual RDF Graphs. In: In Proceedings of the 3rd International Semantic Web Conference (ISWC2004).

Bizer, C., Cyganiak, R., Gauss, T., Maresch, O., 2005. The triql.p browser: Filtering information using context-, content- and rating-based trust policies. In: Proceedings of the Semantic Web and Policy Workshop at the 4th International Semantic Web Conference. Galway, Ireland, pp. 12–20.

Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked data - the story so far. International Journal on Semantic Web and Information Systems 5 (3), 1–22.

Blitzer, J., Dredze, M., Pereira, F., 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proceedings of the Association for Computational Linguistics (ACL).

Bojars, U., 2007. ResumeRDF Ontology Specification. URL http://rdfs.org/resume-rdf/

Bojars, U., Breslin, J. G., 2007. ResumeRDF: Expressing skill information on the Semantic Web. In: First International ExpertFinder Workshop. Berlin, Germany.

Bojars, U., Breslin, J. G., 2010. SIOC Core Ontology Specification. URL http://rdfs.org/sioc/spec/

Bojars, U., Breslin, J. G., Finn, A., Decker, S., 2008. Using the semantic web for linking and reusing data across web 2.0 communities. Web Semantics 6 (1), 21–28.

Bontas, E. P., Mochol, M., Tolksdorf, R., 2005. Case studies on ontology reuse. In: In Proceedings of the 5th International Conference on Knowledge Management (IKNOW05).

Bothos, E., Apostolou, D., Mentzas, G., 2008. IDEM: A prediction market for idea management. designing e-business systems. markets, services, and networks. In: 7th Workshop on E-Business, WeB 2008. Paris, France.

Bower, M., October 1930. The merchandising of ideas. Harvard Business Review, 26–34.

BrainBank, 2009. International olympic committee case study.
URL `http://www.brainbankinc.com/Casestudies1.aspx`

Brem, A., Voigt, K.-I., 2007. Innovation management in emerging technology ventures - concept of an integrated idea management. International Journal of Technology, Policy and Management 7 (3).

Brem, A., Voigt, K.-I., 2009. Integration of market pull and technology push in the corporate front end and innovation management—insights from the german software industry. Technovation 29 (5), 351–367.

Breslin, J. G., Harth, A., Bojars, U., Decker, S., 2005. Towards semantically-interlinked online communities. In: In Proceedings of the 2nd European Semantic Web Conference (ESWC '05). Heraklion, Greece.

Brickley, D., Guha, R., 2004. RDF Vocabulary Description Language.
URL `http://www.w3.org/TR/rdf-schema/`

Brickley, D., Miller, L., 2010. FOAF vocabulary specification.
URL `http://xmlns.com/foaf/spec/`

BrightIdea, 2009a. Bright Side Foundation.
URL `http://www.brightidea.com/about-us-foundation.bix`

BrightIdea, 2009b. BrightIdea Pipline product page.
URL `http://www.brightidea.com/pipeline.bix`

BrightIdea, 2012. BrightIdea homepage.
URL `http://www.brightidea.com/`

Brin, S., Page, L., April 1998. The anatomy of a large-scale hypertextual web search engine. In: Proceedings of the 7th International Conference on World Wide Web. Brisbane, Australia.

Broekstra, J., Kampman, A., van Harmelen, F., 2002. Sesame: A generic architecture for storing and querying RDF and RDF Schema. In: International Semantic Web Conference. pp. 54–68.

Brown, J., 2006. Managing product relationships: Enabling iteration and innovation in design. Tech. rep., AberdeenGroup.

Brown, M., Powers, S., Nicolson, N., March 2009. Vendor landscape: Innovation management software. Tech. rep., Forrester.

Bullinger, A. C., 2008. Innovation and Ontologies. Gabler.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A., 1997. Discovering Data Mining: From Concept to Implementation. Prentice Hall.

Cai, M., Frank, M., 2004. RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network. In: Proceedings of the 13th international conference on World Wide Web. WWW '04. New York, NY, USA, pp. 650–657.

Cardona, G. P., Westerski, A., Garijo, M., December 2012. Application of semantic search in idea management systems. In: The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012). London, UK.

Cardoso, J., Hepp, M., Lytras, M. D. (Eds.), 2007. The Semantic Web: Real-World Applications from Industry. Vol. 6 of Semantic Web And Beyond Computing for Human Experience. Springer.

Catasta, M., Cyganiak, R., Tummarello, G., 2009. Towards ecsse: live web of data search and integration. In: Semantic Search 2009 Workshop (SemSearch2009).

Chandra, C., Tumanyan, A., 2004. Ontology driven knowledge design and development for supply chain management. In: Proceedings of IERC'2004,.

Chandy, R. K., Tellis, G. J., 1998. Organizing for radical product innovation: The overlooked role of willingness to cannibalize. Journal of Marketing Research 35 (4), 474– 478.

Chen, T.-Y., Chen, Y.-M., Chia-Jou Lin and, P.-Y. C., 2010. A fuzzy trust evaluation method for knowledge sharing in virtual enterprises. Computers & Industrial Engineering 59 (4), 853–864.

Chesbrough, H. W., March 2003. Open Innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

Chuang, L.-M., Liu, C.-C., Tsai, W.-C., Huang, C.-M., May 2010. Towards an analytical framework of organizational innovation in the service industry. African Journal of Business Management 4 (5), 790– 799.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum.

Conn, S. J., Torkkeli, M. T., Bitran, I., 2009. Assessing the management of innovation with software tools: an application of InnovationEnterprizer. International Journal of Technology Management 45 (3/4), 323–336.

Cormican, K., O'Sullivan, D., 2004. Auditing best practice for effective product innovation management. Technovation 24 (10), 819–829.

CVS, 2009. Concurrent Versions System (CVS) project homepage.
URL http://www.nongnu.org/cvs/

Cymfony, 2012. Cymfony homepage.
URL http://www.cymfony.com/

D2RQ, 2012. D2RQ Project Page. http://d2rq.org/.

Dadzie, A.-S., Rowe, M., 2011. Approaches to visualising linked data: A survey. Semantic Web 2 (2), 89–124.

Dahlander, L., Gann, D. M., 2010. How open is innovation? Research Policy 39 (6), 699 – 709.

Damanpour, F., Gopalakrishnan, S., March 1998. Theories of organizational structure and innovation adoption: the role of environmental change. Journal of Engineering and Technology Management 15 (1), 1– 24.

Das, M., Puri, R., September 2003. Using innovation management for cost reduction in the auto industry. Tech. rep., Infosys.
URL http://www.infosys.com/research/publications/Documents/cutting-edge-CE-07-03.pdf

Das, S. R., Chen, M. Y., September 2007. Yahoo! for amazon: Sentiment extraction from small talk on the web. Management Science 53 (9), 1375–1388.

Dasarathy, B. V., 1990. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. IEEE Computer Society.

DCMI Usage Board, 2012. Dcmi metadata terms specification. http://dublincore.org/documents/dcmi-terms/.

de Oliveira Arantes, L., de Almeida Falbo, R., Guizzardi, G., 2007. Evolving a software configuration management ontology. In: II Workshop on Ontologies and Metamodeling in Software and Data Engineering.

Decker, S., 2002. Semantic web methods for knowledge management. Ph.D. thesis, University of Karlsruhe.

Delbru, R., Campinas, S., Tummarello, G., 2012. Searching web data: An entity retrieval and high-performance indexing model. Web Semantics 10, 33–58.

Dell, 2009. Dell IdeaStorm homepage.
URL http://www.ideastorm.com/

Dell, 2011. Form 10-k for fiscal 2009. Tech. rep., Dell Inc.

Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R. V., Jhingran, A., Kanungo, T., McCurley, K. S., Rajagopalan, S., Tomkins, A., Tomlin, J. A., Zien, J. Y., 2003. A case for automated large-scale semantic annotation. Web Semantics 1 (1), 115–132.

Ding, X., Liu, B., Yu, P. S., 2008. A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. Heraklion, Greece.

Drejer, I., April 2004. Identifying innovation in surveys of services: a schumpeterian perspective. Research Policy 33 (3), 551– 562.

Driver, E., 2003. Company size more than vertical industry defines idea management software customer. Tech. rep., Forrester.

Dumbill, E., 2012. Description of a project ontology.
URL http://trac.usefulinc.com/doap

Dvorak, M., 2010. Mindraider homepage.
URL http://mindraider.sourceforge.net/semantic-web.html

Eris, E. D., Saatcioglu, O. Y., July 2006. A system look for technological innovation: Firm based perspective. In: European and Mediterranean Conference on Information Systems (EMCIS). Costa Blance, Alicante, Spain.

Erling, O., Mikhailov, I., 2009. Virtuoso: RDF Support in a Native RDBMS. Semantic Web Information Management, 501–519.

Esuli, A., 2007a. Sentiment classification bibliography.
URL http://liinwww.ira.uka.de/bibliography/Misc/Sentiment.html

Esuli, A., 2007b. Sentiment classification bibliography. http://liinwww.ira.uka.de/bibliography/Misc/Sentiment.html.

Fan, T.-K., Chang, C.-H., 2010. Sentiment-oriented contextual advertising. Knowledge and Information Systems 23, 321–344.

Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., Stephens, S., January 2009. The semantic web in action. Scientific American.

Feldstein, A. P., 2007. Brand communities in a world of knowledge-basedproducts and common property. In: MIT5: creativity, ownership and collaboration in digital age.

Fenn, J., LeHong, H., July 2011. Hype cycle for emerging technologies, 2011. Tech. rep., Gartner.

Fensel, D., Hendler, J. A., Lieberman, H., Wahlster, W., 2005. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press.

Fernandez, S., Berrueta, D., Labra, J. E., 2008. A semantic web approach to publish and consume mailing lists. International Journal on WWW/Internet 6 (1).

Fernández-Villamor, J. I., Blasco, J., Iglesias, C. A., Garijo, M., January 2011. A semantic scraping model for web resources - applying linked data to web page screen scraping. In: 3rd International Conference on Agents and Artificial Intelligence (ICAART 2011). Rome, Italy.

Fischer, S., 2001. Course and exercise sequencing using metadata in adaptive hypermedia learning systems. ACM Journal of Educational Resources in Computing 1 (1).

Ford, E., Mohapatra, S., 2011. Idea de-duplication in an innovation community.

Foster, J., 2007. How to Get Ideas. Berrett-Koehler Publishers.

Fox, M. S., Gruninger, M., 1998. Enterprise modeling. AI Magazine 19 (3).

Gailly, F., Poels, G., 2005. Towards an OWL formalization of the resource event agent business domain ontology. In: Proc. 10th Int'l Conf. Business Information Systems, LNCS.

Gamble, P., Blackwell, J., 2002. Knowledge Management - A State Of The Art Guide. Kogan Page Business Books.

Gangi, P. M. D., Wasko, M., 2009. Steal my idea! organizational adoption of user innovations from a user innovation community: A case study of dell ideastorm. Decision Support Systems 48, 303– 312.

Garcia, A. C. B., Vivacqua, A. S., Tavares, T. C., 2010. mesa: a model for collective decision making. In: CSCW10 Workshop on Collective Intelligence In Organizations: Toward a Research Agenda.

Garcia, R., Calantone, R., 2002. A critical look at technological innovation typology and innovativeness terminology: a literature review. The Journal of Product Innovation Management 19, 110—132.

Geffen, C., Judd, K., 2004. Innovation through initiativesa framework for building new capabilities in public sector research organizations. Journal of Engineering and Technology Management 21 (4), 281–306.

Geroimenko, V., Chen, C. (Eds.), 2002. Visualizing the Semantic Web. Springer.

Giereth, M., Koch, S., Kompatsiaris, Y., Papadopoulos, S., Pianta, E., Serafini, L., Wanner, L., 2007. A modular framework for ontology-based representation of patent information. In: Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference. Amsterdam, Netherlands.

Gilfillan, S. C., 1935. The Sociology of Invention. Follett Publishing Company.

Glassman, B. S., 2009. Improving idea generation and idea management in order to better manage the fuzzy front end of innovation. Ph.D. thesis, Purdue University.

Gómez-Pérez, A., 1999. Ontological engineering: A state of the art. Expert Update: Knowledge Based Systems and Applied Artificial Intelligence 2 (3), 3343.

Gómez-Pérez, A., Ortiz-Rodríguez, F., Villazón-Terrazas, B., 2005. Legal ontologies for the spanish e-government. In: Current Topics in Artificial Intelligence (CAEPIA). pp. 301–310.

Gómez-Pérez, A., Ramírez, J., Villazón-Terrazas, B., 2007. An ontology for modelling human resources management based on standards. In: Proceedings of the 11th international conference, KES 2007 and XVII Italian workshop on neural networks conference on Knowledge-based intelligent information and engineering systems: Part I. Berlin, Heidelberg.

Google, 2012. Google AdSense.
URL www.google.com/adsense

Gordijn, J., Akkermans, H., 2002. Value based requirements engineering: Exploring innovative e-commerce ideas. Requirements Engineering Journal 8.

Gorski, C., Heinekamp, E. J., 2002. Capturing employee ideas for new products. In: The PDMA ToolBook 1 for New Product Development. John Wiley & Sons. Inc.

Greeley, M., 2009. BrightIdea, in partnership with adobe acrobat ideas, competes for 2009 forrester groundswell awards. http://blog.brightidea.com/innovation_work/2009/09/brightidea-in-partnership-with-adobe-acrobat-ideas-competes-for-2009.html.

Green, R., Bean, C. A., Myaeng, S. H., 2002. The Semantics of Relationships: An Interdisciplinary Perspective. Kluwer Academic Publishers.

Gruber, T. R., 1993. A translation approach to portable ontology specifications. Knowledge Acquisition 5, 199–220.

Grudin, J., May 1994. Computer-supported cooperative work: history and focus. Computer 27 (5), 19–26.

Gruninger, M., 1996. Designing and evaluating generic ontologies. In: Proceedings of the Workshop on Ontological Engineering, European Conference on Artificial Intelligence.

Hachey, G., Gasevic, D., 2012. Semantic web user interfaces: A systematic mapping study and review.
URL http://www.semantic-web-journal.net/sites/default/files/swj316.pdf

Haller, A., Cimpian, E., Mocan, A., Oren, E., Bussler, C., July 2005. Wsmx - a semantic service-oriented architecture. In: Proceedings. 2005 IEEE International Conference on Web Services (ICWS).

Han, K. H., Park, J. W., 2009. Process-centered knowledge model and enterprise ontology for the development of knowledge management system. Expert Systems with Applications 36 (4), 7441–7447.

Handschuh, S., Staab, S., Studer, R., September 2003. Leveraging metadata creation for the semantic web with cream. In: Proceedings of KI 2003: Advances in Artificial Intelligence. Springer, Berlin, Germany, pp. 19–33.

Hariharan, S., Srimathi, R., Sivasubramanian, M., Pavithra, S., 2010. Opinion mining and summarization of reviews in web forums. In: Proceedings of the Third Annual ACM Bangalore Conference.

Hayes, A. F., Krippendorff, K., 2007. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures 1 (1), 77–89.

Hemminger, B. M., Saelim, B., Sullivan, P. F., Vision, T. J., 2007. Comparison of full-text searching to metadata searching for genes in two biomedical literature cohorts. J. Am. Soc. Inf. Sci. Technol. 58 (14).

Henderson, R. M., Clark, K. B., March 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative Science Quarterly 35 (1), 9– 30.

Hendler, J. A., Holm, J., Musialek, C., Thomas, G., 2012. Us government linked open data: Semantic.data.gov. IEEE Intelligent Systems 27 (3), 25–31.

Hepp, M., 2008. GoodRelations: An ontology for describing products and services offers on the web. In: Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management (EKAW2008). Vol. 5268. Springer LNCS, Acitrezza, Italy, pp. 332–347.

Herman, I., 2004. Tutorial on basic sw technologies. Sankt Augustin, Germany.
URL `http://www.w3.org/2004/Talks/0608-StAugustin-IH/`

Ho, C.-T., Chen, Y.-M., Chen, Y.-J., Wang, C.-B., 2004. Developing a distributed knowledge model for knowledge management in collaborative development and implementation of an enterprise system. Robotics and Computer-Integrated Manufacturing 20 (5), 439–456.

Hobday, M., 2005. Firm-level innovation models: Perspectives on research in developed and developing countries. Technology Analysis & Strategic Management 17 (2), 121–146.

Hogan, A., Harth, A., Passant, A., Decker, S., Polleres, A., 2010. Weaving the pedantic web. In: Proceedings of the Linked Data on the Web WWW2010 Workshop. Raleigh, USA.

Horrocks, I., December 2008. Ontologies and the semantic web. Communications of the ACM 51 (12), 58–67.

Horrocks, I., Patel-Schneider, P. F., van Harmelen, F., 2003. From SHIQ and RDF to OWL: The Making of a Web Ontology Language. Journal of Web Semantics 1 (1), 7–26.

Howe, J., 2004. The rise of crowdsourcing. Wired 14 (6).

Hrastinski, S., Kviselius, N. Z., Edenius, M., 2010. A review of technologies for open innovaton: Characteristics and future trends. In: Proceedings of the 43rd Hawaii International Conference on System Sceiences. Koloa, USA.

Hu, M., Liu, B., 2004. Mining and summerizing customer reviews. In: Proceeding of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

Hubert Gatignon, Michael L. Tushman, W. S., Anderson, P., 2002. A structural approach to assessing innovation: Construct development of innovation locus, type, and characteristics. Management Science 48 (9), 1103–1122.

Hughes, T. P., 1975. Innovators: the problems they choose, the ideas they have, and the innovations they make. In: Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley, F. A., Mitzner, M. (Eds.), Technological innovation: A critical review of current knowledge. Advanced Technology and Science Studies Group, Georgia Tech.

Huizingh, E. K., 2011. Open innovation: State of the art and future perspectives. Technovation 31 (1), 2 – 9.

Husig, S., Kohn, S., June 2003. Factors influencing the front end of the innovation process: A comprehensive review of selected empirical npd and explorative ffe studies. In: Proceedings of the 10th. International Product Development Management Conference. Brussels, Belgium.

i-Prize, 2010. Cisco i-Prize idea collection page.
URL http://i-prize.spigit.com/

IBM, 2012. IBM TAKMI Text mining system homepage.
URL http://www.trl.ibm.com/projects/textmining/takmi/takmi_e.htm

IdeaCentral, 2009. Imaginatik IdeaCentral product page.
URL http://www.imaginatik.com/webdoc_prod_idc_description

Idearium, 2009. Idearium project page.
URL http://www.laboranova.com/tools/idearium

IdeaScale, 2009. IdeaScale product page.
URL http://ideascale.com/

IdeaStream, 2012. Gi2MO IdeaStream idea management system project page. http://ideastream.gi2mo.org/.

IdeaTorrent, 2012. Ideatorrent homepage. http://www.ideatorrent.org/.

IEEE, 2002. Standard for learning object metadata. technical report 1484.12.1, ieee inc. Tech. rep., LTCS WG12: IEEE Learning Technology Standards Committee.

Imaginatik, 2009. Best practice: The idea management process.
URL `http://www.imaginatik.com/webdoc_idea_best_practice`

Imaginatik, 2012. Imaginatik homepage.
URL `http://www.imaginatik.com/`

IngenuityBank, 2009. Ingenuity Bank homepage.
URL `http://www.ingenuitybank.com/`

IT-Code, 2010. It-code homepage.
URL `http://www.civil.auc.dk/~i6ycl/itcode`

Jarratt, T. A. W., Eckert, C. M., Caldwell, N. H. M., Clarkson, P. J., 2010. Engineering change: An overview and perspective on the literature. Research in Engineering Design 22 (2), 103–124.

Jarvenpaa, M., Virtanen, M., Salminen, A., 2006. Semantic portal for legislative information. In: Electronic Government. Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Jean-Mary, Y. R., Shironoshita, E. P., Kabuka, M. R., 2009. Ontology matching with semantic verification. Web Semantics 7 (3), 235–251.

Jindal, N., Liu, B., 2008. Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. pp. 219–230.

Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of the European Conference on Machine Learning (ECML). Chemnitz, Germany.

Jouret, G., September 2009. Inside cisco's search for the next big idea. Harvard Business Review.

Kanerva, M., Hollanders, H., April 2009. The impact of the economic crisis on innovation. analysis based on the innobarometer 2009 survey. Tech. rep., PRO INNO Europe.

Kapanipathi, P., Orlandi, F., Sheth, A., Passant, A., October 2011. Personalized filtering of the twitter stream. In: Semantic Personalized Information Management Worshop at the International Semantic Web Conference. Koblenz, Germany.

Keller, U., Feier, C., 2005. State of the art and requirements on reasoning with semantic web services. Tech. rep., Digital Enterprise Research Istitute (DERI).

Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley, F. A., Mitzner, M., 1975. Ecology of innovation. In: Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley,

F. A., Mitzner, M. (Eds.), Technological innovation: A critical review of current knowledge. Advanced Technology and Science Studies Group, Georgia Tech.

Khanna, N., 2003. An ontology for a lean supply chain. Master's thesis, Wichita State University.

Khoo, C. S. G., Na, J.-C., 2006. Semantic relations in information science. Annual Review of Information Science and Technology 40 (1).

Kim, H. L., Breslin, J. G., 2012. SCOT Ontology Specification. http://rdfs.org/scot/spec/.

Kim, H. L., Scerri, S., Breslin, J. G., Decker, S., Kim, H. G., 2008. The state of the art in tag ontologies: a semantic model for tagging and folksonomies. In: Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications. Dublin Core Metadata Initiative, pp. 128–137.

Kirschner, P. A., Buckingham-Shum, S. J., Carr, C. S., 2002. Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer.

Klein, M., 2012. Enabling large-scale deliberation using attention-mediation metrics. Computer Supported Cooperative Work 21 (4-5), 449–473.

Klischewski, R., 2003. Semantic web for e-government. In: Proceedings of the Second International Conference in E-Government (EGOV'03). Springer, pp. 288–295.

Koen, P., Ajamian, G., Burkart, R., Clamen, A., Davidson, J., DAmore, R., Elkins, C., Herald, K., Incorvia, M., Johnson, A., Karol, R., Seibert, R., Slavejkov, A., , Wagner, K., 2001. Providing clarity and a common language to the "fuzzy front end". Research Technology Management 44 (2), 46–55.

Koestler, A., 1990. The Act of Creation. Penguin Books.

Kornish, L. J., Ulrich, K. T., 2011. Opportunity spaces in innovation: Empirical analysis of large samples of ideas. Management Science 57 (1), 107–128.

Kraut, R. E., 1996. Network-based human-computer interaction. In: Human-Computer Interaction Technologies in Japan. International Technology Research Institute.

Krippendorff, K., 1970. Estimating the reliability, systematic error and random error of interval data. Educational and Psychological Measurement 30, 61–70.

Krippendorff, K., 2004. Reliability in content analysis: Some common misconceptions and recommendations. Human Communication Research 30 (3), 411–433.

Krotzsch, M., Vrandecic, D., 2012. SWIVT Ontology Specification.
URL http://semantic-mediawiki.org/swivt/

Krötzsch, M., Vrandecic, D., Vr, D., Völkel, M., 2005. Wikipedia and the semantic web - the missing links. In: Proceedings of Wikimania 2005.

Krotzsch, M., Vrandecic, D., Vlkel, M., Haller, H., Studer, R., 2007. Semantic wikipedia. Journal of Web Semantics 5, 251–261.

Kruk, S. R., Decker, S., Zieborak, L., 2005. Jeromedl - adding semantic web technologies to digital libraries. In: In Proceedings of DEXA2005 Conference. Copenhagen, Denmark.

Kruk, S. R., Haslhofer, B., Piotrowski, P., Westerski, A., Woroniecki, T., September 2006. The role of ontologies in semantic digital libraries. In: 10th European Conference on Research and Advanced Technology for Digital Libraries. Alicante, Spain.

Kruk, S. R., Kruk, E., Stankiewicz, K., 2008. Evaluation of semantic and social technologies for digital libraries. In: European Conference on Research and Advanced Technology for Digital Libraries (ECDL). Aarhus, Denmark.

Kruk, S. R., McDaniel, B. (Eds.), 2009. Semantic Digital Libraries. Springer.

Lakhani, K. R., von Hippel, E., 2003. How opensource software works: free user-to-user assistance. Research Policy 32 (6), 923943.

Landis, J. R., Koch, G. G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Larrinaga, F., Lizarralde, O., Perez, A., 2012. Gi2MO Wave ontology specification. URL `http://purl.org/gi2mo/wave/ns`

Larrinaga, F., Santos, I., Lizarralde, O., Perez, A., 2011. A case study on the use of community platforms for inter-enterprise innovation. In: 17th International Conference on Concurrent Enterprising (ICE), 2011. Aachen, Germany.

Lee, C.-S., Wang, Y.-C., Liu, W.-M., Lin, Y.-C., 2007. Crm ontology based on cmmi project planning for business application. In: 2007 International Conference on Machine Learning and Cybernetics.

Lee, D., Peristeras, V., Loutas, N., 2008. Reusing the SIOC Ontology to Facilitate Semantic CWE Interoperability. In: On the Move to Meaningful Internet Systems 2008.

Leitzelman, M., Trousse, B., 2011. Supporting the selection of open innovation software tools. In: 17th International Conference on Concurrent Enterprising (ICE), 2011. Aachen, Germany.

Lenat, D., 1995. CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM 38, 33–38.

Lewis, D. D., 1998. Naive (bayes) at forty: The independence assumption in information retrieval. In: Proceedings of the European Conference on Machine Learning (ECML). Chemnitz, Germany.

Li, F., Huang, M., Yang, Y., Zhu, X., 2011. Learning to identify review spam. In: Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three.

Li, L., Horrocks, I., 2003. A software framework for matchmaking based on semantic web technology. In: Proceedings of the 12th international conference on World Wide Web. Budapest, Hungary.

Linking Open Data, 2011. Linking Open Data Project website.
URL http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

Lisi, F. A., Esposito, F., 2009. Learning semantic web rules: State of the art and directions of research. In: IRMLeS 2009: 1st ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web. Heraklion, Crete.

Liu, B., April 2008a. Opinion mining & summarization - sentiment analysis. Tutorial at International World Wide Web Conference 2008 (WWW2008), Beijing, China.

Liu, B., 2008b. Opinion mining and summarization. In: Tutorial given during World Wide Web conference (WWW-2008). Beijing, China.

Liu, B., 2008c. Opinion mining and summarization. In: World Wide Web Conference 2008 (WWW2008). Beijing, China.

Liu, B., Hu, M., Cheng, J., 2005. Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web. Heraklion, Greece.

Liwen, H., Min, X., 2004. Research on modelling customer ontology under crm framework. In: PACIS 2004 Proceedings.

Lorenzo, L., Lizarralde, O., Santos, I., Passant, A., July 2011. Structuring e-brainstorming to better support innovation processes. In: Fifth International AAAI Conference on Weblogs and Social Media (ICWSM11). Barcelona, Spain.

Lu, I. M., Mantei, M. M., 1991. Idea management in a shared drawing tool. In: Proceedings of 2nd European Conference on Computer Supported Cooperative Work.

Lucene, 2012. Lucene implementation of knn classifier.
URL http://lucene.apache.org/core/old_versioned_docs/versions/3_0_1/api/core/index.html

Lytras, M. D., de Pablos, P. O., 2009. Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies. Information Science Reference.

Macrini, D., Maggiore, A., Santagata, A., 2005. Semantic web in the automotive industry: a case study. In: SWAP 2005 - Semantic Web Applications and Perspectives. Trento, Italy.

Magro, D., Goy, A., 2008. Towards a first ontology for customer relationship management. In: Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology. Cergy-Pontoise, France.

Maier, R., Hadrich, T., Peinl, R., 2005. Enterprise Knowledge Infrastructures. Springer.

Manola, F., Miller, E., 2004. RDF Primer.
URL http://www.w3.org/TR/rdf-primer/

Marsh, E. E., White, M. D., 2003. A taxonomy of relationships between images and text. Journal of Documentation 56 (6).

McCandless, M., Hatcher, E., Gospodnetic, O., 2010. Lucene in Action,. Manning Publications.

Mika, P., Akkermans, H., 2003. Analysis of the state-of-the-art in ontology-based knowledge management. Tech. rep., Vrije Universiteit, Amsterdam.

Miles, A., Bechhofer, S., 2009. Skos simple knowledge organization system. reference. http://www.w3.org/TR/2009/REC-skos-reference-20090818/.

Mishne, G., Glance, N., 2006. Predicting movie sales from blogger sentiment. In: AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW). New York, NY, USA.

Montoya-Weiss, M. M., O'Driscoll, T. M., 2000. From experience: applying performance support technology in the fuzzy front end. Journal of Product Innovation Management 17 (2), 143 – 161.

Mullen, T., Malouf, R., 2008. Taking sides: User classification for informal online political discourse. Internet Research 18, 177– 190.

Myaeng, S. H., McHal, M. L., 1991. Toward a relation hierarchy for information retrieval. In: 2nd ASIS SIG/CR Classification Research Workshop.

Myers, S., Marquis, D. G., 1969. Successful Commercial Innovation,. National Science Foundation.

Narasimhalu, A., 2005. Innovation cube: Triggers, drivers and enablers for successful innovations. In: ISPIM.

Nesco, 2009. Nesco Idea Exchange product page.
URL http://www.nosco.dk/product.html

Neuendorf, K. A. (Ed.), 2001. The Content Analysis Guidebook. Sage Publications.

Newsfutures, 2009. Newsfutures Idea Pageant product page.
URL http://us.newsfutures.com/home/ideaPageant.html

Nielsen, 2012. Nielsen online homepage.
URL http://www.nielsen-online.com/

Nielsen, J., 2006. Managing innovation. Tech Monitor Jan-Feb 2006, 62–63.

Ning, K., O'Sullivan, D., Zhu, Q., Decker, S., 2006. Semantic innovation management across the extended enterprise. International Journal on Industrial and Systems Engineering 1.

Nisbet, D., 2004. Measuring the quantity and quality of online discussion group interaction. Journal of eLiteracy 1.

Nixon, L., Paslaru, E., Zaremba, M., Dente, E., Lara, R., Binder, W., Constantinescu, I., Jurga, R., Schickel-Zuber, V., Tanasescu, V., Carman, M., Penserini, L., Pistore, M., July 2001. State of the art of current semantic web services initiatives. Tech. rep., Knowledge Web.

Normann, R., 1971. Organizational innovativeness: Product variation and reorientation. Administrative Science Quarterly 16, 203–215.

Noy, N. F., McGuinness, D. L., 2001. Ontology development 101: A guide to creating your first ontology. Tech. rep., Stanford Medical Informatics, Stanford.

O'Leary, D. E., 1998. Enterprise knowledge management. IEEE Computer 31 (3), 54–61.

O'Leary, D. E., 2004. On the relationship between rea and sap. International Journal of Accounting Information Systems 5 (1), 65 – 81.

OPAL, 2012. OPAL (opinion analyser) application homepage. `http://www.gi2mo.org/apps/opal-opinion-analyser/`.

Opendover, 2012. Opendover service homepage.
URL `http://www.opendover.nl/`

OpenInnovators, 2009. Open innovation and crowdsourcing examples. open innovators blog.
URL `http://www.openinnovators.net/list-open-innovation-crowdsourcing-examples/`

O'Reilly, T., 2007. What is web 2.0: Design patterns and business models for the next generation of software. Communications and Strategies (65), 17–37.

Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Tummarello, G., 2008. Sindice.com: A document-oriented lookup index for open linked data. International Journal of Metadata, Semantics and Ontologies 3 (1).

Ortt, R. J., van der Duin, P. A., 2008. The evolution of innovation management towards contextual innovation. European Journal of Innovation Management 11 (4), 522–538.

Osborn, A. F., 1957. Applied imagination: Principles and procedures of creative problem-solving. Charles Scribner's Sons, New York.

Osterwalder, A., Pigneur, Y., 2002. An e-business model ontology for modeling e-business. In: 15th Bled Electronic Commerce Conference.

Pak, A., Paroubek, P., May 2010. Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10).

Pang, B., Lee, L., 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the ACL.

Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2 (1-2).

Pang, B., Lee, L., Vaithyanathan, S., May 2002. Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP). Philadelphia, PA, USA, pp. 79–86.

Papataxiarhis, V., Tsetsos, V., Karali, I., Stamatopoulos, P., 2009. Developing rule-based applications for the web: Methodologies and tools. In: Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches.

Patel-Schneider, P. F., Hayes, P., Horrocks, I., 2004. Ontology language semantics and abstract syntax, W3C recommendation. Tech. rep., OWL Web.

Perey, C., 2008. Beyond eyeballs: Improving social networking metrics. In: W3C Workshop on the Future of Social Networking.

Pinto, H. S., Martins, J., 2000. Reusing ontologies. In: In AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes.

Plessers, P., Troyer, O. D., Casteleyn, S., March 2007. Understanding ontology evolution: A change detection approach. Web Semantics 5 (1), 39–49.

PMO, 2010. Project metrics ontology.
URL `http://www.daml.org/ontologies/349`

Poltrock, S., Grudin, J., 1999. Cscw, groupware and workflow: experiences, state of art, and future trends. In: CHI '99: CHI '99 extended abstracts on Human factors in computing systems. ACM, New York, NY, USA, pp. 120–121.

Popadiuka, S., Choo, C. W., 2006. Innovation and knowledge creation: How are these concepts related? International Journal of Information Management 26 (4), 302– 312.

Porter, M. E. (Ed.), 1986. Competition in Global Industries. Harvard Business Review Press.

Prud'hommeaux, E., Seaborne, A., 2008. SPARQL Query Language for RDF.
URL `http://www.w3.org/TR/rdf-sparql-query/`

Qiu, G., He, X., Zhang, F., Shi, Y., Bu, J., Chen, C., September 2010. Dasa: Dissatisfaction-oriented advertising based on sentiment analysis. Expert Systems With Applications 37 (9), 6182–6191.

Raimond, Y., Abdallah, S., Sandler, M., Giasson, F., 2007. The music ontology. In: Proceedings of the International Conference on Music Information Retrieval.

Rao, L., Mansingh, G., Osei-Bryson, K.-M., 2012. Building ontology based knowledge maps to assist business process re-engineering. Decision Support Systems 52 (3), 577–589.

Rebstock, M., Janina, F., Paulheim, H., 2008. Ontologies-Based Business Integration. Springer.

Resulta, 2011. RESULTA project page.
URL http://resulta.grupogesfor.com/

Reynolds, D., 2012. Organization ontology.
URL http://www.epimorphics.com/public/vocabulary/org.html

Rico, F. T., 2010. Aplicacion de tecnicas de analisis de sentimientos a la gestion de innovacion con fabrica de ideas. Master's thesis, Universidad Politecnica de Madrid.

Riedl, C., 2009. Idea ontology vocabulary specification.
URL http://www.ideaontology.org/spec/

Riedl, C., May, N., Finzen, J., Stathel, S., Kaufman, V., Krcmar, H., 2009a. An idea ontology for innovation management. International Journal on Semantic Web and Information Systems.

Riedl, C., May, N., Finzen, J., Stathel, S., Leidig, T., Kaufman, V., Belecheanu, R., Krcmar, H., 2009b. Managing service innovations with an idea ontology. In: Proceedings of XIX International RESER Conference, September 24th-26th 2009. Budapest, Hungary.

Riloff, E., Wiebe, J., 2003. Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP). Sapporo, Japan.

Rodriguez, M. E., Conesa, J., Sicilia, M. A., 2009. Clarifying the semantics of relationships between learning objects. In: Third International Conference on Metadata and Semantics Research (MTSR'09).

Rogers, E. M., 1962. Diffusion of innovations. New York: Free Press.

Rozwell, C., Harris, K., Caldwell, F., 2002. Survey of innovation management technology. Tech. rep., Gartner.

Rozwell, C., Harris, K., Mesaglio, M., August 2010. Who's who in innovation management technology. Tech. rep., Gartner.

Rubalcaba, L., Gallego, J., Gago, D., 2010. On the differences between goods and services innovation. Journal of Innovation Economics 30 (4), 611–628.

Sacco, G. M., 2006. User-centric access to e-government information: e-citizen discovery of e-services. In: AAAI Spring Symposium: Semantic Web Meets eGovernment. pp. 114–116.

SalesForce, 2009. Salesforce Ideas product page.
URL `http://www.salesforce.com/salesforceideas/`

Schumpeter, J., 1934. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle. Harvard University Press, Cambridge, MA.

Scoreboard, 2009. European innovation scoreboard 2008. comparative analysis of innovation performance. Tech. rep., PRO INNO Europe.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., Schraefel, M. C., 2012. Linked open government data: Lessons from data.gov.uk. IEEE Intelligent Systems 27 (3), 16–24.

Shahri, H. H., Hendler, J. A., Porter, A. A., 2007. Software configuration management using ontologies. In: Proceedings of the 3rd International Workshop on Semantic Web Enabled Software Engineering at the 4th European Semantic Web Conference (ESWC'07). Innsbruck, Austria.

Simou, N., Saathoff, C., Dasiopoulou, S., Spyrou, V., Voisine, N., Tzouvaras, V., Kompatsiaris, Y., Avrithis, Y., Staab, S., 2005. An ontology infrastructure for multimedia reasoning. In: Proceedings of the International Workshop VLBV05. Sardinia, Italy.

Siorpaes, K., Hepp, M., May 2008. Games with a purpose for the semantic web. IEEE Intelligent Systems 23 (3), 50–60.

Softic, S., Hausenblas, M., 2008. Towards opinion mining through tracing discussions on the web. In: Social Data on the Web (SDoW 2008) Workshop at the 7th International Semantic Web Conference.

Spann, M., Skiera, B., 2003. Internet-based virtual stock markets for business forecasting. Management Sceince 49, 1310–1326.

Spigit, 2012. Spigit homepage.
URL `http://www.spigit.com/`

Stankovic, M., October 2007. Modeling online presence. In: Proceedings of the First Social Data on the Web Workshop. Karlsruhe, Germany.

Stankovic, M., 2010. Open innovation and semantic web : Problem solver search on linked data. In: In Proceedings of International Semantic Web Conference (ISWC) 2010.

Starbucks, 2009. My Starbucks Idea homepage.
URL `http://mystarbucksidea.force.com/home/home.jsp`

Starbucks, 2011. Starbucks corporation fiscal 2011 annual report. Tech. rep., Starbucks Corporation.

Stephens, L. M., Stephens, L. M., Chen, Y. F., Chen, Y. F., 1995. Principles for organizing semantic relations in large knowledge bases. IEEE Transactions on Knowledge and Data Engineering 8, 492–496.

Stromer-Galley, J., 2007. Measuring deliberation's content: a coding scheme. Journal of Public Deliberation 3 (1).

Stumme, G., Hotho, A., Berendt, B., 2006. Semantic web mining: State of the art and future directions. Web Semantics 4 (2), 124–143.

Subversion, 2009. Subversion (SVN) project homepage.
URL `http://subversion.tigris.org/`

Susman, G., Warren, A., Ding, M., 2006. Product and service innovationinsmall and medium-sized enterprises. Tech. rep., Smeal College of Business. The Pennsylvania State University.

Tang, J., Li, J., Liang, B., Huang, X., Li, Y., Wang, K., December 2006. Using bayesian decision for ontology mapping. Web Semantics 4 (4), 243–262.

Taylor, J., Watkinson, D., 2007. Indexing reliability for condition survey data. The Conservator 30, 49–62.

Teece, D. J., 1998. Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. California Management Review 40 (3), 55– 79.

Theodoros Bozios, Fedon komisopoulos, N. F., May 2009. Analyzing the relationships among innovation activities and organization innovation. In: Proceedings of the 2nd International Conference: Quantitative and Qualitative Methodologies in the Economic and Administrative Sciences. Athens, Greece.

THOFU, 2012. THOFU project page.
URL `http://www.thofu.es/`

Thomas, M., Pang, B., Lee, L., 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In: Proceedings of EMNLP.

Tian, P., Liu, Y., Liu, M., Zhu, S., 2009. Research of product ranking technology based on opinion mining. In: Second International Conference on Intelligent Computation Technology and Automation.

Toubia, O., 2006. Idea generation, creativity, and incentives. Marketing Science 25, 411– 425.

Trott, P. (Ed.), 2008. Innovation Management and New Product Development. Prentice Hall.

Tummarello, G., Cyganiak, R., Catasta, M., Danielczyk, S., Delbru, R., Decker, S., 2010. Sig.ma: Live views on the web of data. Web Semantics 8 (4), 355–364.

Turney, P. D., 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA.

Turrell, M., August 2002. Idea management and the suggestion box. Tech. rep., Imaginatik.

Turrell, M., 2008. Collecting ideas from consumers & customers.
URL http://www.youtube.com/watch?v=QVZLrcOJR6Y

Tushman, M. L., Anderson, P. C., O'Reilly, C., 1997. Technology cycles, innovation streams and ambidextrous organizations. Managing Strategic Innovation and Change 34 (3), 3– 23.

Ubuntu, 2011. Ubuntu homepage.
URL http://www.ubuntu.com/

Ubuntu Brainstorm, 2012. Ubuntu brainstorm idea management instance. http://brainstorm.ubuntu.com/.

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F., January 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. Web Semantics 4 (1), 14–28.

Uschold, M., Healy, M., Williamson, K., Clark, P., Woods, S., 1998. Ontology reuse and application. In: Proceedings of the 1st International Conference on Formal Ontology in Information Systems (FOIS98. IOS Press, pp. 179–192.

Uschold, M., Uschold, M., King, M., King, M., House, S. B. R., Moralee, S., Moralee, S., Zorgios, Y., Zorgios, Y., 1995. The enterprise ontology. The Knowledge Engineering Review 13.

Usher, A. P., 1954. A History of Mechanical Invention. Harvard University Press.

Utterback, J. M., 1987. Innovation and industrial evolution in manufacturing industries. In: Guile, B. R., Brooks, H. (Eds.), Technology and Global Industry: Companies and Nations in the World Economy. National Academy Press, pp. 16–48.

Utterback, J. M., Abernathy, W. J., 1975. A dynamic model of product and process innovation. Omega 3 (6), 639–656.

van Dijk, C., van de Ende, J., 2002. Suggestion systems: Transferring employee creativity into practicable ideas. R&D Management 32 (5), 387–395.

Vivacqua, A. S., Expedito, C., Galuzzo, F., Borges, M. R. S., da Silva, S. T., 2010. Moving from ideas to proposals. In: CSCW10 Workshop on Collective Intelligence In Organizations: Toward a Research Agenda.

Voigt, K.-I., Brem, A., June 2006. Integrated idea management in emerging technology ventures. In: IEEE International Conference on Management of Innovation and Technology.

Warner, F., July 2002. In a word – Toyota drives for innovation. Fast Company 61, 36. URL http://www.fastcompany.com/magazine/61/toyota.html

Warren, A., Susman, G., 2002. Review of innovation practices in small manufacturing companies. Tech. rep., National Institute of Standards and Technology, United States Department of Commerce.

Wasserman, S., Faust, K. (Eds.), 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

WebStorm, 2009. BrightIdea WebStorm product page. URL http://www.brightidea.com/webstorm.bix

wen Ma, S., 2009. Research on the relationships between enterprise knowledge structure, knowledge creation and knowledge innovation. In: Information Science and Engineering (ICISE), 2009 1st International Conference on. pp. 5086–5090.

Westerski, A., May 2011. Gi2MO: Interoperability, linking and filtering in idea management systems. In: Extended Semantic Web Conference 2011. PhD Symposium Poster. Heraklion, Greece.

Westerski, A., 2012a. Gi2MO ontology specifrication. URL http://purl.org/gi2mo/ns

Westerski, A., 2012b. Gi2MO Project homepage. URL http://www.gi2mo.org/

Westerski, A., 2012c. Gi2MO Types specification. http://www.gi2mo.org/taxonomy/.

Westerski, A., 2012d. IdeaAnalyst webpage. URL http://gi2mo.org/apps/idea-analyst

Westerski, A., Anadiotis, G., 2012. eDialogos Consensus ontology specification. URL http://www.imc.com.gr/ontologies/eDialogos/consensus/

Westerski, A., Dalamagas, T., Iglesias, C. A., 2012a. Classifying and comparing community innovation in idea management systems. Decision Support Systems.Accepted for Publication.

Westerski, A., Iglesias, C. A., 2011. Exploiting structured linked data in enterprise knowledge management systems: An idea management case study. In: Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International.

Westerski, A., Iglesias, C. A., 2012. Mining sentiments in idea management systems as a tool for rating ideas. In: Large-Scale Idea Management and Deliberation workshop. 10th International Conference on the Design of Cooperative Systems (COOP2012). Marseille, France.

Westerski, A., Iglesias, C. A., Garcia, J. E., October 2012b. Idea relationship analysis in open innovation crowdsourcing systems. In: 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing. Pittsburgh, United States.

Westerski, A., Iglesias, C. A., Nagle, T., 2011a. The road from community ideas to organisational innovation: A life cycle survey of idea management systems. A Special Issue of the Journal Web-Based Communities. Community-based Innovation: Designing Shared Spaces for Collaborative Creativity.

Westerski, A., Iglesias, C. A., Rico, F. T., 2010. A model for integration and interlinking of idea management. In: Metadata and Semantic Research: 4th International Conference, MTSR 2010. Alcalá de Henares, Spain.

Westerski, A., Iglesias, C. A., Rico, F. T., October 2011b. Linked opinions: Describing sentiments on the structured web of data. In: 4th international workshop Social Data on the Web (SDoW2011). Bonn, Germany.

Westerski, A., Kruk, S. R., Samp, K., Woroniecki, T., Czaja, F., O'Nualláin, C., 2006. E-learning based on the social semantic information sources. In: 1st Latin American conference on Learning Objects. Guayaquil.

Wiebe, J., 2011. Sentiment classification bibliography.
URL `http://www.cs.pitt.edu/~wiebe/subjectivityBib.html`

Wiebe, J., 2012. Opinion mining bibliography by jan wiebe. `www.cs.pitt.edu/~wiebe/subjectivityBib.html`.

Wilkinson, K., Sayers, C., Kuno, H. A., Reynolds, D., 2003. Efficient RDF storage and retrieval in Jena2. In: SWDB. pp. 131–150.

Witt, M., Scheiner, C., Robra-Bissantz, S., 2011. Gamification of online idea competitions: Insights from an explorative case. In: INFORMATIK 2011 - Informatik schafft Communities. Berlin, Germany.

Wood, D., 2010. Linking Enterprise Data. Springer.

Yasuda, Y., 1990. 40 Years, 20 Million Ideas: The Toyota Suggestion System. Productivity Pr.

Yiqing, L., Lu, L., Chen, L., April 2009. Decision-making for supplier selection based on ontology and rules. In: Second International Conference on Intelligent Computation Technology and Automation, 2009. ICICTA '09.

Yoo, S. B., Kim, Y., 2002. Web-based knowledge management for sharing product data in virtual enterprises. International Journal of Production Economics 75 (1-2), 173–183.

Young, J. W., 2003. A Technique for Producing Ideas. McGraw-Hill.

Yu, H., Hatzivassiloglou, V., 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP). Sapporo, Japan.

Zainol, A. S., Yusof, W. Z. M., Mastor, K. A., Sanusi, Z. M., Ramli, N. M., 2012. Using group brainstorming in industrial design context: Factors inhibit and exhibit. Procedia - Social and Behavioral Sciences 49, 106 – 119.

Zembowicz, F., Opolon, D., Miles, S., 2010. openchart: Charting quantitive properties in lod. In: LDOW2010.

Zhou, J., Yang, H., Wang, M., Zhang, R., Yue, T., Zhang, S., Mo, R., 2010. A survey of semantic enterprise information integration. In: Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on.

Zhuang, L., Jing, F., Zhu, X.-Y., 2006. Movie review mining and summarization. In: Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM '06). New York, NY, USA.

List of Figures

List of Tables

Appendix A

Thesis Contributions List

The research of this thesis has stimulated creation of a project for knowledge management in Idea Management Systems called Gi2MO. The project is an embodiment of the achievements of the thesis. It has been established for dissemination of the work of the thesis as well as an incubator for research prototypes which matured during the course of the thesis to become software components used by a number of institutions.

Furthermore, the research presented in the thesis and disseminated through scientific publications, has been utilized in a number of funded projects, has stimulated development for a range of open-source software and has initiated various research collaborations.

This appendix lists the achievements of the thesis in terms on impact that it put for both software development and research.

The main aspects of impact discussed are:

- Usage of thesis research outcomes in other projects by Mondragon univeristy, DERI institute, INRIA research centre, and IMC Systems
- Research collaborations established between Gi2MO Project and other research centres
- Software development of Gi2MO Project, its usage, dissemination and impact so far
- List of thesis contributions in terms of scientific publications and participation in research events

A.1 Thesis research impact on work of other scientists

The research conducted as part of this thesis has been acknowledged by a number of researchers independent to the thesis author. As a result selected elements of the contributions the we presented have been used as a point of reference or component to conduct further investigation in Idea Management Systems or related areas. We list the areas, research centres and their related publications that refer to the work of this thesis.

Appliance of Idea Management in technological clusters of enterprises

The researchers of Mondragon University have studdied the appliance of Idea Management Systems in the environment of multiple companies united under the barner of the Mondragon Corporation. With relation to our work, Larrinaga et al. (Larrinaga et al., 2011) reference the notion of Idea Life Cycle presented by this thesis as well as refer to our Gi2MO Ontology is terms of building a Semantic Web oriented solution for data integration of the cluster companies they worked with. In their article, they present the preliminary evaluation of an Idea Management solution built with a number of enhancements with regard to the related work.

Related publication:

Larrinaga, F., Santos, I., Lizarralde, O., Perez, A., 2011. A case study on the use of community platforms for inter-enterprise innovation. In: 17th International Conference on Concurrent Enterprising (ICE), 2011. Aachen, Germany

Modelling brainstorming use for support of innovation processes

In a joint effort between DERI, Mondragon Univeristy and ISEA institute researchers proposed to model brainstorming processes based on SIOC ontology and by referring to previous achievements in the Idea Management area, including Gi2MO ontology. In their work, the aforementioned scientists refer to Gi2MO and use it to contrast with their different modelling solution that proposes stronger re-use of the existing ontologies instead of naming and knowledge organization specific for Idea Management Systems.

Related publication:

Lorenzo, L., Lizarralde, O., Santos, I., Passant, A., July 2011. Structuring e-brainstorming to better support innovation processes. In: Fifth International AAAI Conference on Weblogs and Social Media (ICWSM11). Barcelona, Spain

Use of Idea Management for Green Services and citizen participation

The researchers of INRIA institute propose to pursue the usage of Idea Management platforms to connect with the citizens and encourage them in collaboration on creation of new services focused on green technologies and addressing environmental problems. In their case study, Leitzelman et al. (Leitzelman and Trousse, 2011) relate to the notion of Idea

Life Cycle as proposed by this thesis, as well as our proposal of semantic idea management. Including those concepts as well as analysing other Idea Management platforms capabilities, Leitzelman proposes a methodology to select an idea Management tool best fitting the specific needs of an organization and use case.

Related publication:

Leitzelman, M., Trousse, B., 2011. Supporting the selection of open innovation software tools. In: 17th International Conference on Concurrent Enterprising (ICE), 2011. Aachen, Germany

Use of Idea Management for question answering with large scale deliberation

IMC Technologies is a company focused on deploying Semantic Web based solutions in various enterprises and public organizations. As part of their activity the company maintains a research department that contributes to large scale deliberation. During their research, the scientists of IMC proposed the use of Gi2MO Ontology to model the data of deliberation systems that were applied in medicine for a problem solving system. In this solution, Gi2MO has been tested from the point of view of integration with other vocabularies due to being incorporated in a bigger model that among others included deliberation and problem modelling ontologies.

Related publication:

Anadiotis, G., Kafentzis, K., Pavlopoulos, J., Westerski, A., April 2012. Building consensus via a semantic web collaborative space. In: WWW 2012 Companion Proceedings. Semantic Web Collaborative Spaces Workshop (SWCS2012). Lyon, France

A.2 Gi2MO Project for thesis dissemination and software incubation

The Gi2MO Project (see Fig. A.1) is an initiative started by the author of the thesis and it has been inspired by the research done for the needs of this thesis. The project goal is to improve knowledge management and interaction in Idea Management Systems. Throughout its existence, the project has served for dissemination purposes of thesis research results, a testing ground for the solutions proposed and incubator for software prototypes that could be used as components in Idea Management Systems.

The achievements of the project include use of thesis results in form of either research or software by a number of companies or research centres. Furthermore, through the Gi2MO Project the author of the thesis has initiated research collaboration with a number of institutions that contributed to research and development of either Gi2MO Project or deliberation systems in general.

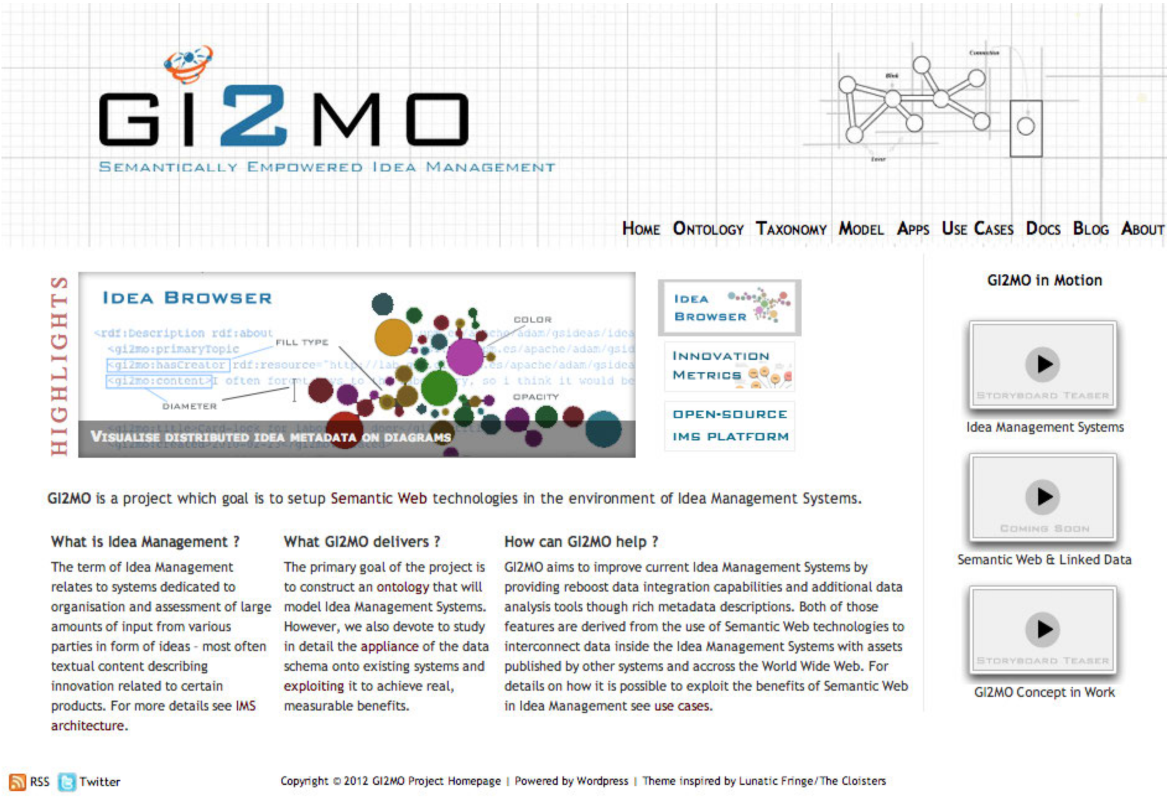


Figure A.1: Internet website of the Gi2MO Project

List of software prototypes developed as part of Gi2MO Project

The Gi2MO Project started based on a proposal of this thesis for an ontology for Idea Management Systems. In terms of software implementations the project envisioned three main application areas (see Fig. A.2): data publishing, data portability and analysis; and data integration.

The developments throughout the project reference those goals and attempt to fulfil some of their aspects:

Data publishing:

1 Drupal RDFme Plugin - a Drupal CMS extension that allows to publish RDF data and attach it to regular HTML pages. For Gi2MO it was used in conjungtion with IdeaStream IMS based on Drupal to export and import data mapped accoridng to the Gi2MO ontology.

2 WordPress RDFa Plugin - a plugin for WordPress blogging engine. It embeds a new button in the post editor that allows to insert RDFa markup for ideas expressed using the Gi2MO vocabulary.

Data portability and analysis:

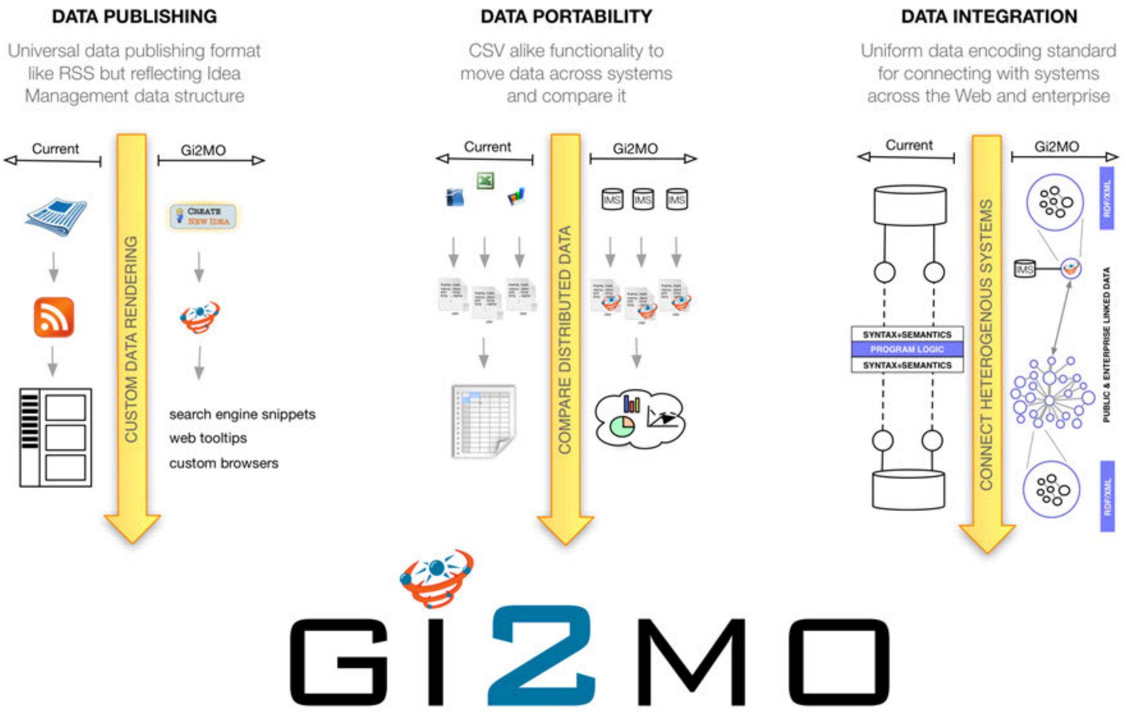


Figure A.2: Three application branches under which Gi2MO prototypes were developed

1 Gi2MO RDF2HTML - a script that take RDF as input and converts it into HTML based on a XSL template. The script wasproposed for Gi2MO in order to render metadata embedded in HTML and show additional information about ideas.

2 Gi2MO Stats - an iPhone application that allows to monitor and compare the statistics of Idea Management Systems. Uses Gi2MO ontology for data portability across systems and services.

3 Idea Analyst - a data visualisation web application that takes RDF as input and turns it into various charts. Was proposed for Gi2MO as an analytical tool that uses SPARQL queries to transform raw data into data series.

4 Idea Browser - a data visualisation web application, that takes RDF input and allows to plot the data it into a bubble diagram. In comparison to the Idea Analyst, the tool focused more on interface problems and usability rather than technical aspects of mapping RDF onto graphs.

5 IdeaStream Analytics - an extension to IdeaStream IMS (based on Drupal) that delivers charts, statistics and summerised table views for Idea Management System data.

6 IdeaStream Similarity - an extension to IdeaStream IMS (based on Drupal) that enables creation of idea relationships, their management and moderation over time.

7 OPAL (Opinion Analyser) - an opinion mining module for Drupal CMS that analyses comments and determines if they are positive or negative. Was proposed for Gi2MO as a tool to experiment with building metrics based on community opinions.

Data integration:

1 Google Wave Plugin - a plugin for Google Wave platform that lets to create and send ideas to the Idea Management System. Build for Gi2MO Project as part of experiments with social networking platforms used in enterprises.

2 IdeaStream Recommender - an extension to IdeaStream IMS (based on Drupal) that shows integration with Human Resources Management System (HRMS) and allows to generate idea rankings based on user profiles. Inspired by the research of Gi2MO on Enterprise Linked Data and its usage for Idea Management Systems.

Use of matured Gi2MO Project software by other institutions

Aside of a rich offering of research prototypes Gi2MO has stimulated development of open-source tools for Idea Management that could be used in practice of organizations. A number of those developments have matured enough to make an impact on the Idea Management System market and be used in real case studies of organizations independent to the author of the thesis.

1. **An Open-Source Idea Management System: Gi2MO IdeaStream.**
   Gi2MO IdeaStream is a module for Drupal Content Management System that initially was created as a base environment to present all developments and research prototypes of Gi2MO Project. However, its further progress as a mature software application got stimulated and encouraged by the interest of other organizations that sought an open-source alternative for the commercial tools of the contemporary Idea Management market. Up until the time of the publication of this thesis IdeaStream has been evaluated and/or used in practice by a number of organizations, including:

   - large enterprises (e.g. subsidy of Saab group utilizing IdeaStream for connecting to the employees and gathering ideas on products)
   - small-medium companies (e.g. department of Ericpol consulting using IdeaStream for collecting ideas internally about company activities and current projects)
   - research laboratories (e.g. INRIA institute during their evaluation of IMS technologies for citizen deliberation on green services)
   - university associations (e.g. ETSIT UPM fendetel using IdeaStream for organizing design competitions for students)

2. **Idea visualisation tools: Idea browser and Idea Analyst.**
   The research on knowledge management in Idea Management Systems presented in this thesis has stimulated building a number of data visualisation tools that would take

advantage of the metadata generated for the goals of information assessment and idea selection. The aim of those tools was to demo the portability of IMS information that can be achieved with the Gi2MO ontology.

However, tools such as Idea Browser and Idea Analyst also aimed to propose a new concepts for graphical visualisation of the idea database that could be used to browse and find ideas easier than with the textual form normally used in Idea Management Systems. As a result, aside of our own experiments, those prototypes have been evaluated and used by the researchers of the INRIA institute during their research on evaluation of Idea Management frameworks (Leitzelman and Trousse, 2011).

3. **Metadata sharing and consuming: RDFme module for Drupal.**
The RDFme module for Drupal CMS was developed in order to provide a demo of Gi2MO ontology being used in Idea Management Systems based on Drupal platform. Aside of our own projects that utilized the very module in conjunction with research that followed (e.g. on idea taxonomies and idea relationships), the RDFme module was also utilized outside of our own context.

In connection to the aforementioned tool for data visualisation the researchers of INRIA institute have used RDFme to export that data of their Idea Management System and analyse it with Idea Browser as well as Idea Analyst.

On the other hand, the researchers of the Mondragon University, took advantage of our RDFme solution together with the Gi2MO ontology to experiment with their own vision of enterprise supported by Semantic Web technologies (Larrinaga et al., 2011).

A.3 Cross university and enterprise collaboration

1. **Results achieved for RESULTA Project**
The RESULTA project funded by the Spanish Ministry of Industry, Tourism and Trade was envisioned to investigate novel techniques for organizing communication and cooperation in the environment of consulting companies. In its initial period, the project focused on social networking platforms and evaluation of their use for establishing contacts between employees as well as clients. The contributions of this thesis for the project are located in the area of solutions for integration of social networking infrastructures with Idea Management. The thesis proposed solutions such as Gi2MO ontology and a number of implementations that allowed to seamlessly transport information from the poorly organized conversations in social network into a structured form in an Idea Management System where the contributions could be assessed and analysed. Furthermore, during this period, we collaborated with Universidad de Alcalá (UAH) on applying their research on ontology matching for integration of different innovation environments in the enterprise with the use of our Gi2MO ontology.

Having established IMS as part of the project infrastructure, in the second stage of the project, we focused more on expanding on the use of other enterprise systems. The thesis has contributed its theories for the use of Enterprise Linked Data to rate and assess ideas in the Idea Management Systems. In the context of the project those contributions have been evaluated through integration of Human Resources Management System and Idea Management. In particular, we organized a series of experiments where participants provided their CVs and filled out details about their skills. This information was later used to rank the ideas those participants posted on the topics related to the project dissemination and future use.

2. **Results achieved for THOFU Project**
   The THOFU project goal was to experiment with and establish new technologies for the use in the future of the hotel business. As part of this bold vision, the thesis contributed to the use of Idea Management Systems for the hotels and tourism. Most notably, the thesis proposed the use of the Gi2MO ontology as well as Gi2MO Types taxonomy to establish metadata that could provide a base for analysis of data submitted by the hotel customers as well as hotel staff.

   Further, the thesis recognized the key differences in the THOFU scenarios to the previously investigated enterprise environment of RESULTA project. in contrast to limited and controlled environment of consulting companies, the hotel open innovation system would gather more ideas and much less organized due to its openness to various customers from many different locations. Therefore, we contributed a study of idea relationships and methods for clustering of similar ideas. The results enabled to significantly downsize the amount of data to be analysed in the case studies investigated as part of the project.

3. **Collaboration with IMIS intitute on automatic annotation (Greece)**
   During our study on idea descriptions and ways of categorizing open innovation data, we proposed a taxonomy for idea characteristics. As part of this research work, we evaluated the use of our contribution working together with automatic annotation solutions. The motivation was to deliver a solution that would decrease the amount of effort and time needed for describing newly submitted ideas into the system.

   On the road to achieving our goal, we established a cooperation with researchers of IMIS institute that specialized in automatic annotation algorithms. The IMIS institute provided us with the necessary implementation and guided us with choosing and adjusting the automatic annotation algorithms specifically for the types of texts present in the Idea Management Systems. Using the established cooperation, together we managed to evaluate the proposal of the thesis for idea characteristics and contributed interesting results about the use of automatic text annotation in the environment of Idea Management.

4. **Collaboration with IMC on consensus ontology and deliberation modelling**
   IMC technologies is an Small/Medium Enterprise (SME) mainly operating in the area of Greece. The company focuses on delivering semantic technologies for other organizations to enhance information search. As part of their research on large scale deliberation systems in medicine, they invited the author of this thesis to collaborate together on researching a solution for automatic analysis of question-answering data. In particular,

the author of the thesis, together with the employees of IMC, contributed the eDialogos Consensus Ontology (Westerski and Anadiotis, 2012). The ontology relies heavily on existing vocabularies, including the Gi2MO ontology.

5. **General collaboration with a number of institutes and companies for improving both research and software prototypes**
   The impact that this thesis has made in terms of research and software, has resulted in cooperation with organizations interested in the contributions of the thesis. While some the established partnerships focused on help with deployment of the thesis solutions, others transformed into more serious collaborations that resulted in improvements of thesis contributions and more mature releases of the software delivered by the Gi2MO Project.

   Among others, the collaboration with other institutions has stimulated the development of Gi2MO ontology in a form of ontology extensions. Those additions to the original schema, maintained now as the core ontology, were created to satisfy various needs of particular implementations where Gi2MO was used as a metadata schema for a number of different needs. For instance, the researchers of Mondragon University have proposed an extension called Gi2MO Wave Ontology (Larrinaga et al., 2012). Their proposal originates from the requirements of the InnoWEB project run together with the Mondragon Corporation and testing take-up of open innovation in cluster companies of a large enterprise. The particular novelties included more detailed metadata about idea contests and a more strict view on idea descriptions by defining a particular set of allowed ones like Outcome, Target, Technology, Market etc.

   Similarly, as Mondragon University, other institutions that we described earlier as using some of the contributions of this thesis, had an impact on the development of the thesis results by delivering requirements and feedback that stimulated certain design and research decisions made in the Gi2MO Project.

A.4 List of scientific publications

During the work of this thesis the author has contributed to a number of research areas. This work has been acknowledged and accepted for presentation during a number of research events. The following scientific publications refer to those contributions:

- Westerski, A., Iglesias, C. A., Rico, F. T., 2010. A model for integration and interlinking of idea management. In: Metadata and Semantic Research: 4th International Conference, MTSR 2010. Alcalá de Henares, Spain

- Westerski, A., Iglesias, C. A., Nagle, T., 2011a. The road from community ideas to organisational innovation: A life cycle survey of idea management systems. A Special Issue of the Journal Web-Based Communities. Community-based Innovation: Designing Shared Spaces for Collaborative Creativity

- Westerski, A., May 2011. Gi2MO: Interoperability, linking and filtering in idea management systems. In: Extended Semantic Web Conference 2011. PhD Symposium Poster. Heraklion, Greece
- Westerski, A., Iglesias, C. A., 2011. Exploiting structured linked data in enterprise knowledge management systems: An idea management case study. In: Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International
- Westerski, A., Iglesias, C. A., Rico, F. T., October 2011b. Linked opinions: Describing sentiments on the structured web of data. In: 4th international workshop Social Data on the Web (SDoW2011). Bonn, Germany
- Anadiotis, G., Kafentzis, K., Pavlopoulos, J., Westerski, A., April 2012. Building consensus via a semantic web collaborative space. In: WWW 2012 Companion Proceedings. Semantic Web Collaborative Spaces Workshop (SWCS2012). Lyon, France
- Westerski, A., Iglesias, C. A., 2012. Mining sentiments in idea management systems as a tool for rating ideas. In: Large-Scale Idea Management and Deliberation workshop. 10th International Conference on the Design of Cooperative Systems (COOP2012). Marseille, France
- Westerski, A., Iglesias, C. A., Garcia, J. E., October 2012b. Idea relationship analysis in open innovation crowdsourcing systems. In: 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing. Pittsburgh, United States
- Cardona, G. P., Westerski, A., Garijo, M., December 2012. Application of semantic search in idea management systems. In: The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012). London, UK
- Westerski, A., Dalamagas, T., Iglesias, C. A., 2012a. Classifying and comparing community innovation in idea management systems. Decision Support Systems.Accepted for Publication

Appendix B

Gi2MO Ontology Specification

One of the contributions of the thesis is the Gi2MO ontology - a formal specification of a conceptualization of the information stored in an Idea Management System. The following appendix contains a summarized version of the specification. In comparison, the full specification available on-line is a more structured document with a grater number links and back-references that facilitate improved specification browsing in the web environment. The summarized version presented in the appendix contains all the information required for the proposed modelling process as does its web counterpart.

The following appendix is primary a supplement for chapter 4, which describes the research done on modelling of Idea Management System information as well as results of experiments with the Gi2MO ontology, its performance in terms of system coverage and practical utilization for data integration.

Gi2MO Ontology Specification

V0.5 - 29 March 2012

This version: http://purl.org/gi2mo/0.5/ns (RDF/XML, HTML)
Latest version: http://purl.org/gi2mo/ns
Editors: Adam Westerski
Authors: Adam Westerski
Contributors: See acknowledgements

Abstract

Generic **I**dea and **I**nnovation **M**anagement **O**ntology (**Gi2MO**) is a standardised data schema (also referred as "ontology" or "vocabulary") designed to annotate and describe resources gathered inside Idea Management facilities. The following document contains the description of ontology and instructions how to connect it with descriptions of other resources.

Table of Contents

Appendixes

1 Introduction

The following specification is a formal description of metadata schema proposal that can be applied to data gathered in the so-called Idea Management Systems. The goal of the following section is to introduce both Semantic Web and Idea Management experts to the topic and goals of the ontology and provide the basic knowledge to comprehend the technical part of the specification.

An important note is that Gi2MO ontology is not a complete model of the Idea Management System. It marly defines concepts that are not described yet by the means of other ontologies. For detailed instructions how to completely model Idea Management Systems with Gi2MO and other ontologies see homepage of the Gi2MO project.

1.1 Idea Management Systems and Innovation Management Process

The Idea Management Systems are referred most often as an application used by organisations to collect input about various ideas regarding their products and services; and manage them afterwards providing certain assessment and screening facilities.

Although, the concept of such systems and vendors that deliver them are available on the market for quite a while, the scope and technologies involved are constantly evolving. The most most basic idea management process includes:

- idea generation (facilities for submitting ideas)
- idea improvement (usually Web 2.0 facilities to modify ideas in collaborative environment)
- idea assessment (idea analysis, assessment and selection tools, e.g. diagrams and reports)

However, in the contemporary systems the tendency is to extend this process towards:

- idea implementation (lightweight project management tools or integration with more complex suits)

- idea deployment (analysis of business metrics related to products based on ideas from IM process, e.g. ROI)

In relation to other IT systems used in modern organisations, at the time of writing this specification the Idea Management Systems have not been yet introduced as a standard component by the key players as part of their ERP, PLM/PLCM or CRM suites. Idea Management, in relation to those systems, is still being considered as a rising market and a complementary solution that can be supplied depending on organisation's interest in investment in innovation.

In relation to other processes run in the modern enterprises, Idea Management can be defined as part of Innovation Management Process run as part of innovation strategy in many contemporary enterprises. Idea Management is perceived as complementary solution to this process that delivers tools that can enhance the successfulness of the innovation management process and open it to new possibilities.

1.2 The Semantic Web

The Semantic Web is a W3C initiative that aims to introduce rich metadata to the current Web and provide machine readable and processable data as a supplement to human-readable Web.

Semantic Web is a mature domain that has been in research phase for many years and with the increasing amount of commercial interest and emerging products is starting to gain appreciation and popularity as one of the rising trends for the future Internet.

One of the corner stores of the Semantic Web is research on inter-linkable and interoperable data schemas for information published online. Those schemas are often referred to as ontologies or vocabularies. In order to facilitate the concept of ontologies that lead to a truly interoperable Web of Data, W3C has proposed a series of technologies such as RDF and OWL. **Gi2MO** uses those technologies and the research that comes within to propose an ontology set in the domain of Idea Management.

1.3 What is Gi2MO for?

The goals of the Gi2MO ontology to achieve as a data schema are:

- enable to publish raw data from Idea Management Systems online and in compliance with current and future Internet trends
- interconnect Idea Management assets with other resources (projects, product metrics, basic knowledge management concepts)
- utilise all above to enhance current idea assessment capabilities with algorithms based on analysis of rich semantic descriptions of idea management assets

For more information please refer to Gi2MO project website.

2. Gi2MO ontology at a glance

An alphabetical index of Gi2MO terms, by class (concepts) and by property (relationships, attributes), are given below. All the terms are hyperlinked to their detailed description for quick reference.

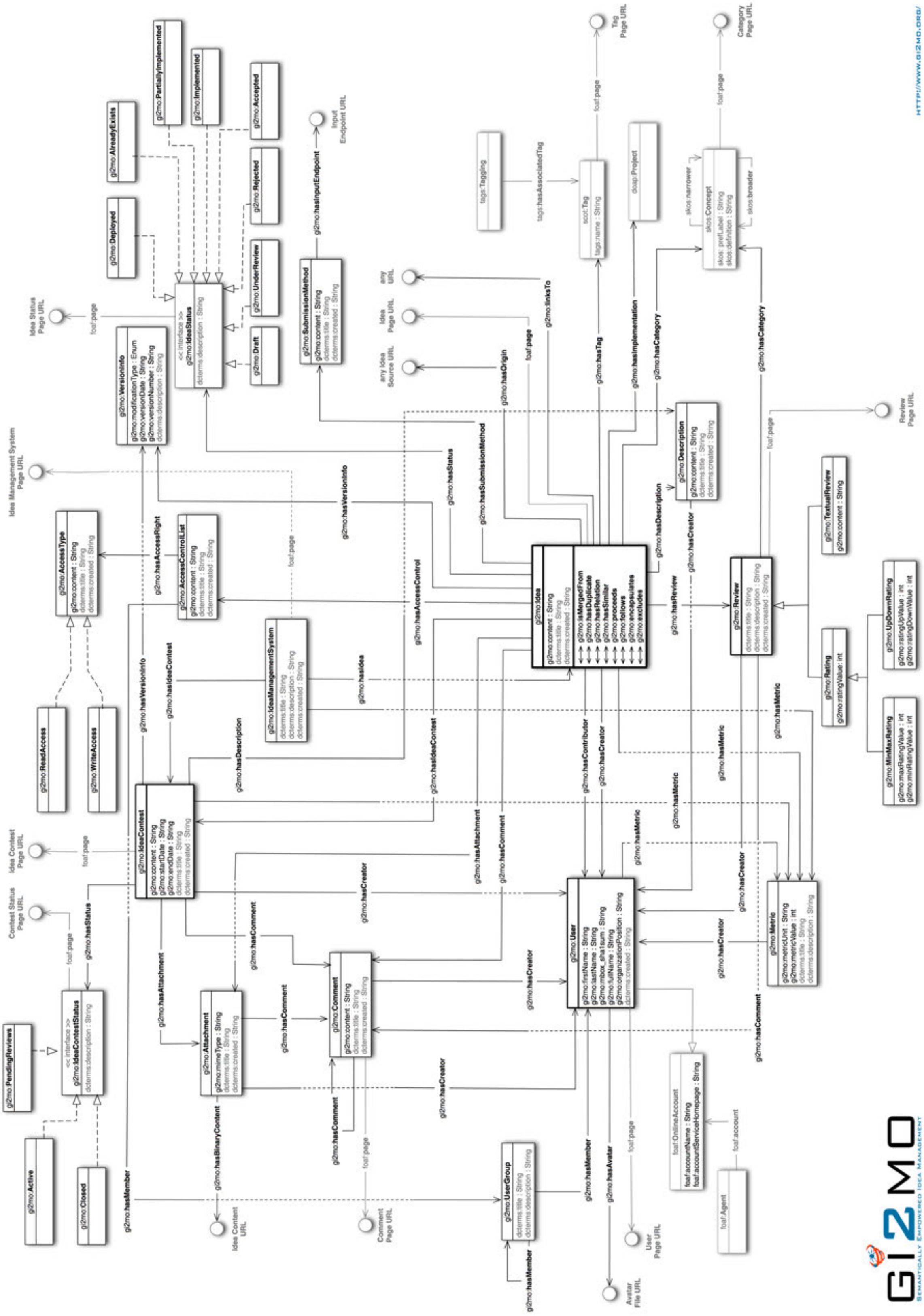
Classes: AccessContolList, AccessType, Attachment, Category, Comment, Description, Idea, IdeaCategory, IdeaContest, IdeaContestStatus, IdeaManagementSystem, IdeaStatus, IdeaTag, Metric, MinMaxRating, Project, Rating, Review, SubmissionMethod, TextualReview, UpDownRating, User, UserGroup, VersionInfo,

Properties: accessRight, content, created, describesPartOf, description, endDate, firstName, fullname, hasAccessContol, hasAccessRight, hasAttachment, hasAvatar, hasBinaryContent, hasCategory, hasComment, hasContributor, hasCreator, hasDescription, hasDuplicate, hasEditor, hasIdea, hasIdeaContest, hasImplementation, hasInputEndpoint, hasMember, hasMetric, hasOrigin, hasRelated, hasReview, hasSimilar, hasStatus, hasSubCategory, hasSubmissionMethod, hasSupplement, hasTag, hasTagging, hasTopCategory, hasVersionInfo, ideasNumber, isAccessControlOf, isAccountOf, isAttachmentOf, isCategoryOf, isCommentOf, isContributorOf, isCreatorOf, isDescribedIn, isDescriptionOf, isEditorOf, isIdeaContestOf, isIdeaOf, isImplementationOf, isLinkedBy, isMemberOf, isMergedFrom, isMergedInto, isMetricOf, isOriginOf, isReviewOf, isStatusOf, isSubmissionMethodOf, isSupplementOf, isTagOf, isTopicOf, isVersionInfoOf, lastName, linksTo, maxRatingValue, mboxSha1sum, metricUnit, metricValue, mimeType, minRatingValue, modificationType, organizationPosition, postsNumber, primaryTopic, ratingDownValue, ratingUpValue, ratingValue, startDate, title, username, versionDate, versionNumber,

Instances: Accepted, Active, AlreadyExists, Closed, Deployed, Draft, Implemented, PartialyImplemented, PendingReviews, ReadAccess, Rejected, UnderReview, WriteAccess,

3. Gi2MO ontology overview

The Gi2MO UML diagram presented below shows connections between classes that implement the data model of Idea Management Systems.



UML Class Diagram for the Gi2MO Ontology

3.1. Example

A very basic example below shows a single idea annotated with Gi2MO metadata:

```
<rdf:Description rdf:about="http://ideas.gi2mo.org/idea/012345">
  <foaf:page rdf:resource="http://ideas.gi2mo.org/ideaView?id=012345"/>
  <gi2mo:hasCreator rdf:resource="http://ideas.gi2mo.org/user/Pablo"/>
  <gi2mo:content>I often forget keys to the laboratory, so i think it would be great to install
                 a card access lock with a reader for chips in spanish ID
  </gi2mo:content>
  <dcterms:title>Card-lock for laboratory door</gi2mo:title>
  <dcterms:created>2010-02-23</gi2mo:created>
  <gi2mo:hasStatus rdf:resource="http://www.upm.es/gi2mo#Implemented"/>
  <gi2mo:hasComment rdf:resource="http://ideas.gi2mo.org/comment/054321"/>
  <gi2mo:hasComment rdf:resource="http://ideas.gi2mo.org/comment/054322"/>
  <gi2mo:hasCategory rdf:resource="http://ideas.gi2mo.org/category/General"/>
  <gi2mo:hasDuplicate rdf:resource="http://ideas.gi2mo.org/idea/02345"/>
</rdf:Description>
```

For more examples please see a Gi2MO RDF export for a simple Idea Management System instance installed on lab GSI severs.

4. Cross-reference for Gi2MO classes and properties

Below see a comprehensive list of all Gi2MO classes, properties and their descriptions.

Class: gi2mo:AccessContolList

URI: http://purl.org/gi2mo/ns#AccessContolList

AccessControlList - An instance of this class indicates the access rights for particular users to a resource that it is connected to (e.g. write privileges for a given idea or idea contest).

Class: gi2mo:AccessType

URI: http://purl.org/gi2mo/ns#AccessType

AccessType - Defines access type to a resource. See class instances for possible options.

Class: gi2mo:Attachment

URI: http://purl.org/gi2mo/ns#Attachment

Attachment - An instance of this class represents attachment of any entity in idea management system (e.g. image attachment to idea).

Class: gi2mo:Category

URI: http://purl.org/gi2mo/ns#Category

Category - An instance of this class represents a category defined to classify ideas or other assets in the Idea Management system. For ideas predefined categories are commonly used in many systems to group ideas by topic (e.g. in IM system for computer equipment vendor categories could be Desktops; Laptops etc.). For reviews this property can be used to groups reviews by type (e.g. "Return of Investment" , "Customer Benefit" , "Cost analysis" etc.) (Deprecated. Use skos:Concept instead to describe categories for ideas and other assets of the Idea Management System.)

Class: gi2mo:Comment

URI: http://purl.org/gi2mo/ns#Comment

Comment - An instance of this class represents comments attached to other entities throughout the idea management system (e.g. comments for idea, idea contests, reviews etc.)

Class: gi2mo:Description

URI: http://purl.org/gi2mo/ns#Description

Description - An instance of this class models any textual description that is added as a supplement to the basic textual content of an entity of an Idea Management system. For example this can be a special description of an idea (e.g. Beneficiaries described separately from the idea summary).

Class: gi2mo:Idea

URI: http://purl.org/gi2mo/ns#Idea

Idea - An instance of this class represents a single idea stored in the Idea Management system.

Class: gi2mo:IdeaCategory

URI: http://purl.org/gi2mo/ns#IdeaCategory

IdeaCategory - An instance of this class represents a category defined to classify ideas in the Idea Management system. (Deprecated. Use skos:Concept instead to describe categories for ideas and other assets of the Idea Management System.)

Class: gi2mo:IdeaContest

URI: http://purl.org/gi2mo/ns#IdeaContest

IdeaContest - An instance of this class represents the so-called Idea Contests (also referred as "idea campaign", "idea event" etc.). Idea contests are events that stimulate idea submission, bind ideas thematically and attach their collection to a particular time period.

Class: gi2mo:IdeaContestStatus

URI: http://purl.org/gi2mo/ns#IdeaContestStatus

IdeaContestStatus - An instance of this class describes status of idea contest. For a list of recommended instances of this class see the individuals list associated to this class in the ontology definition.

Class: gi2mo:IdeaManagementSystem

URI: http://purl.org/gi2mo/ns#IdeaManagementSystem

IdeaManagementSystem - An instance of this class represents an Idea Management Systems as a container for knowledge associated to the Idea Management process. The class models the relationships that the systems has with it's data.

Class: gi2mo:IdeaStatus

URI: http://purl.org/gi2mo/ns#IdeaStatus

IdeaStatus - An instance of this class describes status that idea and indicates the position of the idea in the Idea Management life cycle (e.g. Draft, Implemented, Deployed etc.) For a list of recommended instances of this class see the individuals list associated to this class in the ontology definition.

Class: gi2mo:IdeaTag

URI: http://purl.org/gi2mo/ns#IdeaTag

IdeaTag - An instance of this class models tags associates to idea in the Idea Management system. (Deprecated. Use scot:Tag and tags:Tagging instead to describe tags in the Idea Management system.)

Class: gi2mo:Metric

URI: http://purl.org/gi2mo/ns#Metric

Metric - This class models any metrics related to Ideas. For example Idea Deployment metrics (Revenue, ROI etc.) or any other metrics associated to the idea as it progresses across the idea pipeline (idea implementation cost, idea assessment costs etc.) The metric can be associated with an individual Idea or an Idea contest.

Class: gi2mo:MinMaxRating

URI: http://purl.org/gi2mo/ns#MinMaxRating

MinMaxRating - An instance of this class describes a rating where users giving votes have to select a value from a predefined scale (for example 1..10).

Class: gi2mo:Project

URI: http://purl.org/gi2mo/ns#Project

Project - Represents a project that implements an idea. (Deprecated. Use doap:project instead.)

[back to top]

Class: gi2mo:Rating

URI: http://purl.org/gi2mo/ns#Rating

Rating - An instance of this class describes a numerical rating attached to the idea. E.g. star review or a quantified review.

Class: gi2mo:Review

URI: http://purl.org/gi2mo/ns#Review

Review - An instance of this class identifies review of a idea. This can be, a textual review, a rating or any kind of assessment made by a number people (e.g. star reviews done by users but also idea assessment made by internal company reviewers). The instance of this class can be used both to describe a single review action (connected to a user thus having hasCreator property) or a global review summary (e.g. description summarization all votes and combined rating value). If not sufficient is recommended to extend this class via inheritance mechanism and add additional type of idea assessment that is particular for a given implementation of Idea Management System.

Class: gi2mo:SubmissionMethod

URI: http://purl.org/gi2mo/ns#SubmissionMethod

SubmissionMethod - An instance of this describes an input interface (e.g. device or software) that was used to submit an idea. This class can be used for Idea Management systems that have many data entry points, for example web forms, collaborative tools, mobile device access etc. Also this class can be used to interlink detailed input information encoded via other ontologies through rdfs:seeAlso (e.g. operating system, browser, screen size etc. can be expressed in DCCI Ontology).

Class: gi2mo:TextualReview

URI: http://purl.org/gi2mo/ns#TextualReview

TextualReview - An instance of this class models a textual review of any kind attached to the idea.

Class: gi2mo:UpDownRating

URI: http://purl.org/gi2mo/ns#UpDownRating

UpDownRating - An instance of this class describes a rating type where users only vote up or down, without any particular grading scale. In case of modelling a single vote ratingValue can be set to 1 or -1.

Class: gi2mo:User

URI: http://purl.org/gi2mo/ns#User

User - An instance of this class models a user account in the Idea Management system. User types can be defined by creating UserGroups.

Class: gi2mo:UserGroup

URI: http://purl.org/gi2mo/ns#UserGroup

UserGroup - An instance of this class allows to group users to reflect the user management model popular in many IT systems.

Class: gi2mo:VersionInfo

URI: http://purl.org/gi2mo/ns#VersionInfo

VersionInfo - An instance of this class allows to attach versioning information to ideas.

Property: gi2mo:accessRight

URI: http://purl.org/gi2mo/ns#accessRight

accessRight - Describes type of access privileges (READ, WRITE) (Deprecated. Use gi2mo:hasAccessRight object property instead.)

Property: gi2mo:content

URI: http://purl.org/gi2mo/ns#content

content - Textual content associated to an entity (e.g. user entered idea summary, comment text, review text etc.)

Property: gi2mo:created

URI: http://purl.org/gi2mo/ns#created

created - Creation date. (Deprecated. Use dcterms:created instead.)

Property: gi2mo:describesPartOf

URI: http://purl.org/gi2mo/ns#describesPartOf

describesPartOf - Indicates that an idea includes a description of what has been proposed in another idea but they might not have been merged. For example: single complex idea describes improving a laptop model and can include ideas about new keyboard, new screen that are submitted also as separate ideas by other users.

Property: gi2mo:description

URI: http://purl.org/gi2mo/ns#description

description - Text providing a description of an entity (e.g. idea category or idea contest description) (Deprecated. Use dcterms:description instead.)

Property: gi2mo:endDate

URI: http://purl.org/gi2mo/ns#endDate

endDate - Indicates end date of an event.

Property: gi2mo:firstName

URI: http://purl.org/gi2mo/ns#firstName

firstName - First name of a user.

Property: gi2mo:fullname

URI: http://purl.org/gi2mo/ns#fullname

fullname - Users full name. Depending on the culture/nationality can be more then just a joint of gi2mo:firstName+gi2mo:lastName.

Property: gi2mo:hasAccessContol

URI: http://purl.org/gi2mo/ns#hasAccessContol

hasAccessControl - Property indicating an Idea/idea contest having certain collaborations permissions (e.g. public, private, editable for all etc). In case of Idea Contests this property can be used to indicate who can take part in the competition.

Property: gi2mo:hasAccessRight

URI: http://purl.org/gi2mo/ns#hasAccessRight

hasAccessRight - Indicates that access control list describes privileges for a certain type of access right.

Property: gi2mo:hasAttachment

URI: http://purl.org/gi2mo/ns#hasAttachment

hasAttachment - Property indicating an Idea having an Attachment.

Property: gi2mo:hasAvatar

URI: http://purl.org/gi2mo/ns#hasAvatar

hasAvatar - Point to an avatar (picture) of a user

Property: gi2mo:hasBinaryContent

URI: http://purl.org/gi2mo/ns#hasBinaryContent

hasBinaryContent - Link (URL or URI) to the attachment (file)

Property: gi2mo:hasCategory

URI: http://purl.org/gi2mo/ns#hasCategory

hasCategory - Property indicating the category of idea or a review.

Property: gi2mo:hasComment

URI: http://purl.org/gi2mo/ns#hasComment

hasComment - Property indicating an idea or other entity having a comment.

Property: gi2mo:hasContributor

URI: http://purl.org/gi2mo/ns#hasContributor

hasContributor - Indicates that an idea has a particular contributor (a person that did not create the idea but now actively participates in its improvement/evolution)

Property: gi2mo:hasCreator

URI: http://purl.org/gi2mo/ns#hasCreator

hasCreator - Property indicating an idea being created by a certain User.

Property: gi2mo:hasDescription

URI: http://purl.org/gi2mo/ns#hasDescription

hasDescription - This property indicates that an idea or another entity of the Idea Management systems (e.g. idea contest) has a more broad description then just the standard summary (e.g. idea benefits or idea contest requirements).

Property: gi2mo:hasDuplicate

URI: http://purl.org/gi2mo/ns#hasDuplicate

hasDuplicate - Indicates identical ideas.

Property: gi2mo:hasEditor

URI: http://purl.org/gi2mo/ns#hasEditor

hasEditor - Indicates the user responsible for change

Property: gi2mo:hasIdea

URI: http://purl.org/gi2mo/ns#hasIdea

hasIdea - Indicates an idea that was posted inside a particular Idea Management System instance.

Property: gi2mo:hasIdeaContest

URI: http://purl.org/gi2mo/ns#hasIdeaContest

hasIdeaContest - Property indicating an idea being part of a Idea Contest or an Idea Contest being posted within an IdeaManagementSystem instance .

Property: gi2mo:hasImplementation

URI: http://purl.org/gi2mo/ns#hasImplementation

hasImplementation - Indicates idea having an implementation as a product or a feature

Property: gi2mo:hasInputEndpoint

URI: http://purl.org/gi2mo/ns#hasInputEndpoint

hasEndpoint - Indicates the data entry endpoint of the submission method. This can be a web form URL, a web service address, a device URI etc.

Property: gi2mo:hasMember

URI: http://purl.org/gi2mo/ns#hasMember

hasMember - Describes Access Control List/ UserGroup having a certain member(s).

Property: gi2mo:hasMetric

URI: http://purl.org/gi2mo/ns#hasMetric

hasMetric - Associates an idea or idea contest with a metric.

Property: gi2mo:hasOrigin

URI: http://purl.org/gi2mo/ns#hasOrigin

hasOrigin - Indicates that an idea origins (was created based on other) from other idea.

Property: gi2mo:hasRelated

URI: http://purl.org/gi2mo/ns#hasRelated

hasRelated - Describes relations between ideas and other resources, without specifying exact link type.

Property: gi2mo:hasReview

URI: http://purl.org/gi2mo/ns#hasReview

hasReview - Indicates an idea having a review

Property: gi2mo:hasSimilar

URI: http://purl.org/gi2mo/ns#hasSimilar

hasSimilar - Indicates similar idea but not duplicates (e.g. idea about improving car appearance, and idea about car speed; both can mention new type of tires but also a number of different elements).

Property: gi2mo:hasStatus

URI: http://purl.org/gi2mo/ns#hasStatus

hasStatus - Property indicating an Idea or IdeaContest have a given status. In case of ideas, this gives information about the position of the idea in the idea life cycle (e.g. newly submitted idea, under review, implemented etc.).

Property: gi2mo:hasSubCategory

URI: http://purl.org/gi2mo/ns#hasSubCategory

hasSubCategory - Describes relationship between categories where one is classified as a part of another. (Deprecated. Use skos:narrower instead.)

Property: gi2mo:hasSubmissionMethod

URI: http://purl.org/gi2mo/ns#hasSubmissionMethod

hasSubmissionMethod - Property indicating an Idea being created with a particular submission method or device (e.g. thought web CMS, via call phone etc).

Property: gi2mo:hasSupplement

URI: http://purl.org/gi2mo/ns#hasSupplement

- Indicates review being a supplement to another review (e.g. this can be used to model textual information added to Rating Review or to create compound reviews made of many ratings etc.)

Property: gi2mo:hasTag

URI: http://purl.org/gi2mo/ns#hasTag

hasTag - Property indicating tag associated to the idea - a user created categorization of an idea.

Property: gi2mo:hasTagging

URI: http://purl.org/gi2mo/ns#hasTagging

hasTag - Indicates that an idea has been tagged by a certain person at a certain point of time. The following property points to a class that describes an activity of tagging while gi2mo:hasTag points to a Tag itself.

Property: gi2mo:hasTopCategory

URI: http://purl.org/gi2mo/ns#hasTopCategory

hasTopCategory - Describes relationship between categories where one is classified as a part of another. (Deprecated. Use skos:broader instead.)

Property: gi2mo:hasVersionInfo

URI: http://purl.org/gi2mo/ns#hasVersionInfo

hasVersionInfo - Indicates that Idea has been modified over time and has more then 1 version

---

Property: gi2mo:ideasNumber

URI: http://purl.org/gi2mo/ns#ideasNumber

ideasNumber - Number of ideas posted by a user. (Deprecated. Use gi2mo:Metric instead to describe metrics for gi2mo:User.)

Property: gi2mo:isAccessControlOf

URI: http://purl.org/gi2mo/ns#isAccessControlOf

isAccessControlOf - Property indicating that access control list is attached to a certain resource (e.g. idea). In case of Idea Contests this property can be used to indicate who can take part in the competition.

Property: gi2mo:isAccountOf

URI: http://purl.org/gi2mo/ns#isAccountOf

isAccountOf - Indicates that account belongs to a particular person that has a foaf description.

Property: gi2mo:isAttachmentOf

URI: http://purl.org/gi2mo/ns#isAttachmentOf *isAttachementOf* - Property indicating an attachment being associated to an idea/idea contest.

Property: gi2mo:isCategoryOf

URI: http://purl.org/gi2mo/ns#isCategoryOf

isCategoryOf - Property indicating the idea that a category characterises a particular entity in the Idea Management System.

Property: gi2mo:isCommentOf

URI: http://purl.org/gi2mo/ns#isCommentOf

isCommentOf - Property indicating an comment being attached to a certain resource.

Property: gi2mo:isContributorOf

URI: http://purl.org/gi2mo/ns#isContributorOf

isContributorOf - Indicates a person being a contributor in a certain idea (not the original creator but a participant in idea creation or improvement)

Property: gi2mo:isCreatorOf

URI: http://purl.org/gi2mo/ns#isCreatorOf

isCreatorOf - Property indicating that User is an author of a certain resource.

Property: gi2mo:isDescribedIn

URI: http://purl.org/gi2mo/ns#isDescribedIn

isDescribedPartiallyBy - Indicates that an idea (as a whole) is described in the text of another idea. For example: single complex idea describes improving a laptop model and can include ideas about new keyboard, new screen that are also submitted as separate ideas by other users.

Property: gi2mo:isDescriptionOf

URI: http://purl.org/gi2mo/ns#isDescriptionOf

isDescriptionOf - This property indicates that the description is attached to a certain idea or another resource in the Idea Management systems (e.g. idea contest).

Property: gi2mo:isEditorOf

URI: http://purl.org/gi2mo/ns#isEditorOf

isEditorOf - Indicates that user has created (edited) a certain version of an idea.

Property: gi2mo:isIdeaContestOf

URI: http://purl.org/gi2mo/ns#isIdeaContestOf

isIdeaContestOf - Indicates an idea that belongs to an idea contest; optionally a Contests being posted in a particular Idea Management System.

Property: gi2mo:isIdeaOf

URI: http://purl.org/gi2mo/ns#isIdeaOf

isIdeaOf - Indicates an Idea Management System in which the idea has been posted.

Property: gi2mo:isImplementationOf

URI: http://purl.org/gi2mo/ns#isImplementationOf

isImplementationOf - Describes that project is a result of previously submitted idea.

Property: gi2mo:isLinkedBy

URI: http://purl.org/gi2mo/ns#isLinkedBy

isMentionedBy - This property can be used to link idea with any other entities that mention it (e.g. via links or citations). This can include both entities inside the Idea Management system, however also objects outside the system (e.g. Twitter comments, forums posts, blogs posts etc.).

Property: gi2mo:isMemberOf

URI: http://purl.org/gi2mo/ns#isMemberOf

isMemberOf - Describes that a User or a UserGroup are a member of other entity (Access Control List, UserGroup)

Property: gi2mo:isMergedFrom

URI: http://purl.org/gi2mo/ns#isMergedFrom

isMergedFrom - Indicates that an Idea was created out of merging other ideas or resources

---

Property: gi2mo:isMergedInto

URI: http://purl.org/gi2mo/ns#isMergedInto

isMergedInto - Indicates an idea has been merged into a "master" idea.

Property: gi2mo:isMetricOf

URI: http://purl.org/gi2mo/ns#isMetricOf

isMetricOf - Indicates that a metric is associated to and idea or idea contest

---

Property: gi2mo:isOriginOf

URI: http://purl.org/gi2mo/ns#isOriginOf

isOriginOf - Indicates that an idea was the base for creation of some other idea.

Property: gi2mo:isReviewOf

URI: http://purl.org/gi2mo/ns#isReviewOf

isReviewOf - Indicates a review being assigned to an idea.

Property: gi2mo:isStatusOf

URI: http://purl.org/gi2mo/ns#isStatusOf

isStatusOf - Indicates Status being assigned to a particular idea or idea contest.

Property: gi2mo:isSubmissionMethodOf

URI: http://purl.org/gi2mo/ns#isSubmissionMethodOf

isSubmissionMethodOf - Indicates that a submission method has been used to create certain idea.

Property: gi2mo:isSupplementOf

URI: http://purl.org/gi2mo/ns#isSupplementOf

isSupplementOf - Indicates review being a supplement to another review (e.g. this can be used to model textual information added to Rating Review or to create compound reviews made of many ratings etc.)

Property: gi2mo:isTagOf

URI: http://purl.org/gi2mo/ns#isTagOf

isTagOf - Indicates that a tag is assigned to a certain Idea or other resource

---

Property: gi2mo:isTopicOf

URI: http://purl.org/gi2mo/ns#isTopicOf

isTopicOf - Indicates that an idea is a topic of some other resource. (e.g. forum, blog post, twitter post, news article etc.)

Property: gi2mo:isVersionInfoOf

URI: http://purl.org/gi2mo/ns#isVersionInfoOf

isVersionOf - Indicates that version info is attached to a certain idea or other resource under version control.

Property: gi2mo:lastName

URI: http://purl.org/gi2mo/ns#lastName

surname - Surname of a user.

Property: gi2mo:linksTo

URI: http://purl.org/gi2mo/ns#linksTo

linksTo - Describes idea linking to other assets (e.g. can be used to list all hyperlinks associated in the idea summary)

Property: gi2mo:maxRatingValue

URI: http://purl.org/gi2mo/ns#maxRatingValue

maxRatingValue - Defines maximal possible value for the rating.

Property: gi2mo:mboxSha1sum

URI: http://purl.org/gi2mo/ns#mboxSha1sum

mboxSha1sum - user mailbox sha1 sum

Property: gi2mo:metricUnit

URI: http://purl.org/gi2mo/ns#metricUnit

metricUnit - Identified in what kind of unit metric is defined (e.g. for money related metrics it can be currency: USD, EUR, for effort related metrics it can be KLOC etc.).

Property: gi2mo:metricValue

URI: http://purl.org/gi2mo/ns#metricValue

metricValue - Identified the value of a metric

Property: gi2mo:mimeType

URI: http://purl.org/gi2mo/ns#mimeType

mimeType - Indicates the MIME type of the attachment (e.g. image/jpg, audio/mp3)

Property: gi2mo:minRatingValue

URI: http://purl.org/gi2mo/ns#minRatingValue

minRatingValue - Defines minimal possible value for the rating.

Property: gi2mo:modificationType

URI: http://purl.org/gi2mo/ns#modificationType

modificationType - Indicates what type of modification has been performed (edit, delete etc)

Property: gi2mo:organizationPosition

URI: http://purl.org/gi2mo/ns#organizationPosition

organizationPosition - Describes the occupation of the employee in the organisation.

Property: gi2mo:postsNumber

URI: http://purl.org/gi2mo/ns#postsNumber

postsNumber - Number of posts done by a user (Deprecated. Use gi2mo:Metric instead to describe metrics for gi2mo:User.)

Property: gi2mo:primaryTopic

URI: http://purl.org/gi2mo/ns#primaryTopic

hasPrimaryTopic - Describes the original web resources that the class models. (Deprecated. Use foaf:page instead)

Property: gi2mo:ratingDownValue

URI: http://purl.org/gi2mo/ns#ratingDownValue

ratingDownValue - Amount of times people voted down.

Property: gi2mo:ratingUpValue

URI: http://purl.org/gi2mo/ns#ratingUpValue

ratingUpValue - Amount of times people voted up.

Property: gi2mo:ratingValue

URI: http://purl.org/gi2mo/ns#ratingValue

ratingValue - Numerical value of the rating

Property: gi2mo:startDate

URI: http://purl.org/gi2mo/ns#startDate

startDate - Indicates starting date of an event.

Property: gi2mo:title

URI: http://purl.org/gi2mo/ns#title

title - Describes the title of an entity (e.g. idea title, idea contest tile, review title etc.). (Deprecated. Use dcterms:title instead.)

Property: gi2mo:username

URI: http://purl.org/gi2mo/ns#username

username - Username of a user. (Deprecated. Use foaf:accountName instead.)

Property: gi2mo:versionDate

URI: http://purl.org/gi2mo/ns#versionDate

versionDate - Version creation or modification date.

Property: gi2mo:versionNumber

URI: http://purl.org/gi2mo/ns#versionNumber

versionNumber - Indicates version number

Instance: gi2mo:Accepted

URI: http://purl.org/gi2mo/ns#Accepted

Accepted - Indicates an idea that has been accepted for implementation.

Instance: gi2mo:Active

URI: http://purl.org/gi2mo/ns#Active

Active - Indicates an idea contest is active and accepting idea submissions.

Instance: gi2mo:AlreadyExists

URI: http://purl.org/gi2mo/ns#AlreadyExists

AlreadyExists - Indicates that the following idea has been already implemented the moment it was submitted.

Instance: gi2mo:Closed

URI: http://purl.org/gi2mo/ns#Closed

Closed - Indicates an idea contest is closed, no longer accepting idea submissions and has been kept in the system for archival purposes and possible future reference.

Instance: gi2mo:Deployed

URI: http://purl.org/gi2mo/ns#Deployed

- Indicates an idea that has been implemented as a product and deployed on the market.

Instance: gi2mo:Draft

URI: http://purl.org/gi2mo/ns#Draft

Draft - Indicates an new idea submitted to the system.

Instance: gi2mo:Implemented

URI: http://purl.org/gi2mo/ns#Implemented

Implemented - Indicates an idea that was accepted and later implemented.

Instance: gi2mo:PartialyImplemented

URI: http://purl.org/gi2mo/ns#PartialyImplemented

PartialyImplemented - Indicates that an ideas has been accepted for implementation but only partially implemented.

Instance: gi2mo:PendingReviews

URI: http://purl.org/gi2mo/ns#PendingReviews

PendingReviews - Indicates that an idea contest is currently closed for submissions but the submitted ideas have not received their reviews and their statuses may be subject of change. This status should be used when the idea contest organizers decide to close submissions and start review process to choose the winners of the competition.

Instance: gi2mo:ReadAccess

URI: http://purl.org/gi2mo/ns#ReadAccess

- Read access rights

Instance: gi2mo:Rejected

URI: http://purl.org/gi2mo/ns#Rejected

Rejected - Indicates an idea that has been rejected and shall not be implemented.

Instance: gi2mo:UnderReview

URI: http://purl.org/gi2mo/ns#UnderReview

UnderReview - Indicates an idea being under review.

Instance: gi2mo:WriteAccess

URI: http://purl.org/gi2mo/ns#WriteAccess

WriteAccess - Write access right

A Acknowledgements

This documentation has been generated automatically from the most recent ontology specification in OWL using a python script called SpecGen. The style formatting has been inspired on FOAF specification.

Special thanks for support with ontology creation and research to: Prof. Carlos A. Iglesias and members of the GSI Group of DIT department of Universidad Politécnica de Madrid.

Appendix C

Gi2MO Types Specification

One of the contributions of the thesis is the Gi2MO Types - a taxonomy for describing non-domain characteristics of idea stored in an Idea Management System. The following appendix contains a summarized version of the taxonomy specification. In comparison, the full specification available on-line is a more structured document with a grater number links and back-references that facilitate improved specification browsing in the web environment. The summarized version presented in the appendix contains all the information necessary to model non-domain characteristics of ideas as proposed by the thesis.

The following appendix is primary a supplement for chapter 6, which describes the research done on modelling of characteristics for ideas Idea Management Systems based on contemporary and past innovation management models. For details of evaluation of this taxonomy in terms of automatic/ manual annotation, metric generation and performance in terms of use for data comparison see chapter 6.

Gi2MO Types Taxonomy Specification

V0.2 - July 2011

This version: http://purl.org/gi2mo/types/0.2/ns (RDF/XML, HTML)
Latest version: http://purl.org/gi2mo/types/ns
Editors: Adam Westerski
Authors: Adam Westerski
Contributors: See acknowledgements

Abstract

Gi2MO Types is a taxonomy designed to be used for annotation of ideas in Idea Management Systems. The following document contains the description of taxonomy and instructions how to connect it with other metadata of Idea Management Systems and their other resources.

Table of Contents

Appendixes

1 Introduction

The following specification is a formal description of taxonomy proposal that can be applied to annotate ideas stored in Idea Management Systems. The goal of the following section is to provide the basic knowledge to comprehend the technical part of the specification. As such it shall introduce the topic of Idea Management Systems and areas related to the Gi2MO project - Semantic Web and publication of open data on the web.

An important note is that Gi2MO Types taxonomy presented here is not a complete model to address the problem of describing and linking ideas inside and between distributed Idea Management Systems. It marly defines concepts that are not described yet by the means of other taxonomies or standards and provides the data attributes that create the basis for establishing clustering and classification facilities for Idea Management Systems. For detailed instructions and recommendations how to use the taxonomy to achieve the aforementioned goals please refer to the guidelines of the Gi2MO project.

1.1 Idea Management Systems

Idea Management Systems, can be considered as an evolution of suggestions boxes used for collecting input from customers. The said systems have went above and beyond mere collection of ideas and aspire to be tools for idea assessment, analysis and sometimes even deployment of innovation. Idea Management Systems take advantage of computer networks and web technologies to reach a large number of customers or enterprise employees at the same time, gather their ideas and invite to collaboration for improving those ideas.

The use of IT infrastructure enables to collect big amounts of ideas in a fairly small amount of time but also brings the problem how to respond to the ideas from the community in a timely manner and how to continuously stimulate this crowd to keep supplying new ideas. Furthermore, in systems running over large periods of time the amount of ideas counted in tens of thousands becomes almost impossible to browse and manage either by moderators or users. All the said problems lower the efficiency of the systems or even put their usage into question. Gi2MO Types taxonomy is one of the building blocks for improving data organisation in Idea

Management Systems and thus finding a solution to the aforementioned problems.

In relation to other IT systems used in modern organisations, at the time of writing this specification the Idea Management Systems have not been yet introduced as a standard component as part of the most popular ERP, PLM or CRM suites. Idea Management, in relation to those systems, is still considered as a rising market and a complementary solution that can be supplied depending on organisation's interest in investment in innovation.

In relation to other processes run in the modern enterprises, Idea Management can be defined as part of Innovation Management Process run as part of innovation strategy in many contemporary enterprises. Idea Management is perceived as complementary solution to this process that delivers tools that can enhance the successfulness of the innovation management process and open it to new possibilities.

1.2 The Semantic Web

The Semantic Web is a W3C initiative that aims to introduce rich metadata to the current Web and provide machine readable and processable data as a supplement to human-readable Web.

Semantic Web is a mature domain that has been in research phase for many years and with the increasing amount of commercial interest and emerging products is starting to gain appreciation and popularity as one of the rising trends for the future Internet.

One of the corner stones of the Semantic Web is research on inter-linkable and interoperable data schemas for information published online. Those schemas are often referred to as ontologies or vocabularies. In order to facilitate the concept of ontologies that lead to a truly interoperable Web of Data, W3C has proposed a series of technologies such as RDF and OWL. **Gi2MO Types** uses those technologies and the research that comes within to publish the characteristics of ideas together with other metadata and enable the easy portability and integration of this information across different Idea Management Systems.

1.3 What is Gi2MO Types for?

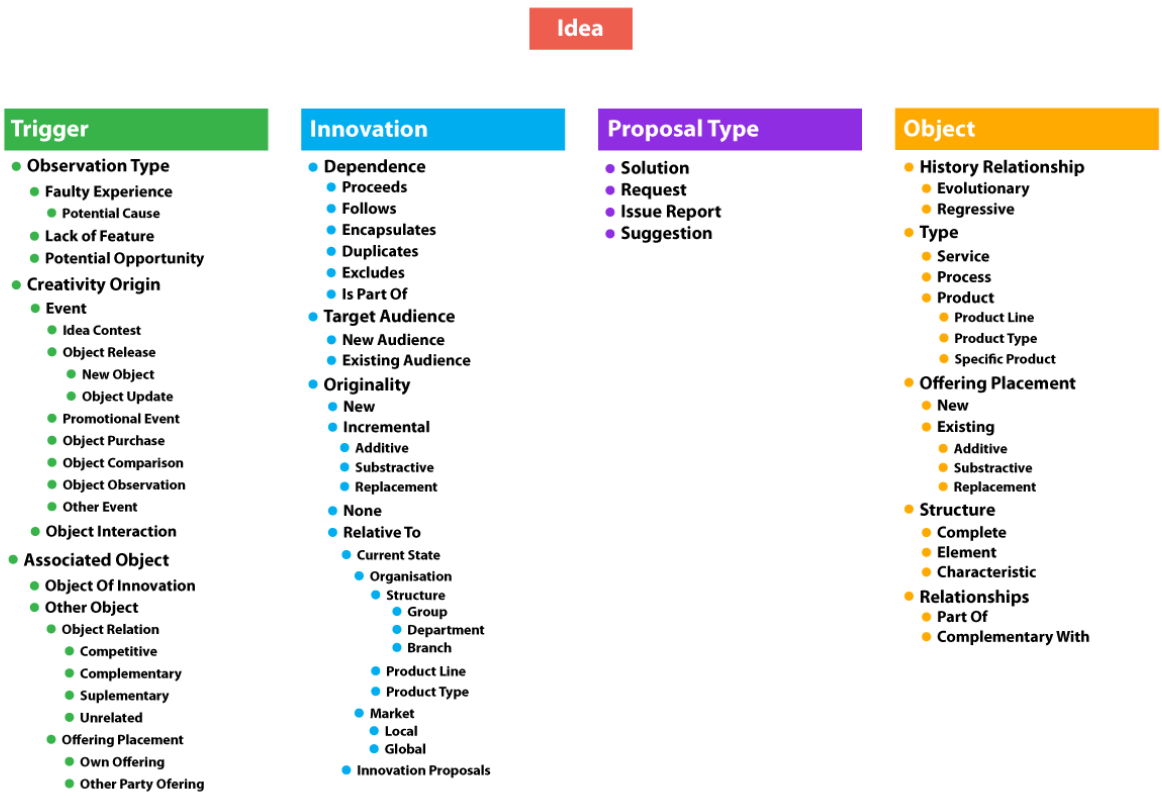
The goals of the Gi2MO Types taxonomy to achieve as a knowledge modelling framework are:

- enable characterisation of ideas beyond the regular tools available in contemporary Idea Management Systems
- aid idea assessment by providing summarised idea information
- provide a corner store to build classification and clustering facilities based on the annotated data corpus
- facilitate idea ranking and similarity detection based on comparison of ideas and their annotations

2. Gi2MO Types taxonomy at a glance

The Gi2MO Types diagram presented below shows the hierarchy of terms organised following the main assumption of the taxonomy that:

*Every idea that was **proposed** has been **triggered** by a particular experience and describes a certain **innovation** put in context of application in a given **object**.*



3. Summary sheet of Gi2MO Types terms

Level	Term Name	Description	Example
1	TRIGGER	What has caused creation of the idea / what is the inspiration	-
2	Observation Type	Type of Trigger. Defines what was the observation of the inventor due to the event that he experienced.	-
3	Faulty Experience	A problem/ failure of a product that caused a the creative process.	-
4	Potential Cause	If a cause of the problem was identified.	-
3	Lack of Feature	The innovation was triggered by the fact that object lacked some kind of feature.	-
3	Potential Opportunity	The ideas was triggered by a person that was not displeased with the way it worked by has foreseen that an addition to the product could improve it.	-
2	Creativity Origin	Points to what kind of event triggered the innovation.	-
3	Event	The innovation was not triggered by direct innovation with some object but by some event.	-
4	Idea Contest	Idea was specifically triggered by idea contests and incentives that come with	-
4	Object Release	Premiere of a new object to the market or audience.	-
5	New Object	Premiere of a fully new object.	-
5	Object Update	Premiere of a new refreshed version of an already existing object.	-
4	Promotional Event	Event related to the object of innovation.	-
4	Related Event	Undefined event related in some way to the object.	-
4	Object Purchase	Idea was created as a result of purchasing the object by the client.	-
4	Object Observation	Idea triggered by plain object observation or interacting with object related information without any particular intent.	-
4	Object Comparison	Idea created during comparison with some other object.	-
3	Object Interaction	A person interacting with the object and discovering some new need or problem.	-
2	Associated Object	Object connected to the trigger of innovation.	-
3	Object of Innovation	Innovation was triggered by interaction with the same object that is the topic of idea.	-
3	Other Object	Type of Trigger. The inventor has came up with an idea by comparison of the object with some other product/service/process.	-
4	Object Relation	Defines the relation of the object that triggered the idea and the object that is the topic of the idea.	-
5	Competitive	If the object that triggered innovation was a competitive solution.	-
5	Complementary	If the object that triggered innovation was a complementary solution that works together with the object to fulfil some OTHER function/goal. (removing one of the objects makes it impossible to fulfil this function).	Idea: I use a scanner a lot so I would like it to be embedded in the laptop.
5	Supplementary	If the object that triggered innovation was a solution that works on top of the existing and delivers some new value but does redefine the end functionality.	Idea: Working with a mouse made me furious when I could not put the pad next to laptop because of the bad USB ports positioning.
5	Unrelated	The innovation was triggered by neither complementary or competitive solution.	-
4	Offering Placement	Determines if the other object that is the means of comparison is part of own offering or some other party.	-
5	Own	Object that triggered the innovative ideas come from own offering.	-
5	Other Party	Object that triggered the innovative ideas come from offering of some other party.	-
1	INNOVATION	Characteristics of the proposal and its impact	-
2	Dependence	Innovation is dependent in some way on some other idea.	-
3	Proceeds	After the described innovation is implemented it is going to be possible to unlock other idea.	ideaA: Keyboard for iPhone. ideaB: Backlit keyboard for iPhone.
3	Follows	The innovation realization is dependent on implementing some other idea first.	ideaA: Keyboard for iPhone. ideaB: Backlit keyboard for iPhone.
3	Encapsulates	Innovation that reproduces other idea(s) but also adds new value.	-
3	Duplicaes	Innovation that exactly reproduces already existing idea.	-
3	Excludes	Introducing this innovation causes the inability to implement some other innovation.	-
3	Is Part Of	Innovation that has been also described in another idea that also shows a bigger scope and discusses other items.	-
2	Target Audience	Defines if the innovation imposes a change in the audience or not (relative to current state of the object of innovation).	-
3	New Audience	Innovation relates to shifting the object to fit new market/customer.	-
3	Existing Audience	Updates of an existing object to fit the changing needs of current customers.	-
2	Orginality	Defines the size/scope of innovation in relation to existing	-

		ideas, state or offering.	
3	New	Completely new idea/design not based on anything created before.	-
3	Incremental	Improvement for an existing object.	-
4	Additive	Improvement made by adding a new element to a previous proposal.	ideaA - new keyboard, ideaB - new backlit keyboard
4	Substractive	Improvement made by simplifying other proposal and removing an element from the previous proposal.	ideaA - serve breakfast with bread, tea, cornflakes, ideaB - serve breakfast only with cornflakes.
4	Replacement	Improvement made just be replacing some element of the previous proposal.	ideaA - colorA, ideaB - colorB OR putting glossy screen to laptop offerring (incremental modification because there are already glossy screens just not for this particular line).
3	None	This is a special instance for idea mgt systems that indicates that idea hold no real innovation at all.	-
3	Relative To	Describes innovation characteristics abstraction level (e.g. if something is innovation with respect to local market or global market).	idea: Laptop keyboard dock for iPhone. Organization level: Breakthrough (product does not exist yet). Global Market: None (Motorola Atrix already has this)
4	Current State	Innovation described in relation to the current state of some entity or environment	-
5	Organisation	Innovation described with respect to organization processes/product/services	-
6	StructureUnit	Innovation describe with respect to some particular element of the administrative structure of the organization.	-
7	Group	Innovation described with respect to group level of the company.	-
7	Department	Innovation described with respect to department level of the company.	-
7	Branch	Innovation described with respect to branch level of the company.	-
6	Product Line	Innovation describe with respect to a product line.	E.g. Vosotro Laptops or iPods.
6	Product Type	Innovation described with respect to a particular product type.	E.g. Laptops.
5	Market	Innovation described with respect to the entire market.	-
6	Local	Innovation described with respect to the local market.	-
6	Global	Innovation described with respect to the global market.	-
1	PROPOSAL	Describes the way a person has formalized his ideas (what parts are there, what are missing etc.) What is the completeness of the proposal.	-
2	Request	A person has supplied a request for a certain feature.	-
2	Solution	The person has supplemented the request with a technical or theoretical solution proposal.	-
2	IssueReport	Some idea are not innovative at all but nevertheless they are submitted to the idea management systems. This caregory indicates idea is not innovative at all just reports on a bug/issue/error/failure of an object.	-
2	Suggestion	Contrary to request does not state the necessity or a dire need but rather says what would be good in the opinion of a person for the company or organisation to do something.	-
1	OBJECT	Characteristics of the item that is the object of the innovation.	-
2	History Relationship	Relationship of the innovation with the previous iterations of the objects (or previous innovation proposals for this object). The difference with "Offering placement" is that it looks to compare the change with the past states of the object, not the current).	-
3	Evolutionary	Indicates that the innovation for the given product puts it in a state that it has never been before.	-
3	Regressive	Indicates that the innovation for the given product was already implemented at some point but has been removed.	-
2	Type	Type of object (is it product, service etc.).	-
3	Service	Innovation relates to a service. Opposing products, services are: intangible (cannot be physically possessed), inseparable (production cannot be separated from consumption), perishable (cannot store for future use), variable (the quality and consistency differs depending on the person that uses it).	-
3	Process	Innovation relates to a process (newelements into an organization's operations such as input materials, task specifications, work and informationflow mechanisms, and equipment used to produce a product or render a service).	-
3	Product	Innovation relates to a product.	-
4	Product Line	An innovation is proposed for entire product line.	Inspiron line
4	Product Type	Innovation proposed for a specific product type.	Laptops
4	Specific Product	Innovation proposed for an individual product.	Dell M1330
2	Offering Placement	Object placement in the current offering / state.	-
3	New	The innovation proposes a new product.	New laptop.
3	Existing	The innovation proposes innovating an existing product.	New laptop screen.
4	Additive	The innovation proposes adding some new element to the object.	Add delete key on the laptop keyboard.

4	Substractive	The innovation proposes removing some element from the object.	Remove dvd drive from a laptop.
4	Replacement	The innovation proposes replacing or modifying an existing element of the object.	Replace the current screen with a better quality one.
2	Structure	Defines the element in the object structure that the innovation refers to. This should be connected with a domain ontology that identifies specific elements of products etc.	-
3	Complete	Innovation related to the entire product.	Dell M1330
3	Element	Innovation relates to a particular element of an object.	Dell M1330 Keyboard
3	Characteristic	Innovation relates to particular feature of an object.	Dell M1130 weight
2	Relationship	Defines if an object has relationship with other items of offering (reason: adding an innovation into a certain object might impact its relationships with other products and increase/decrease the value of innovation).	-
3	Part Of	Object is part of some other object.	-
3	Complementary With	Object is complementary with some other object.	-

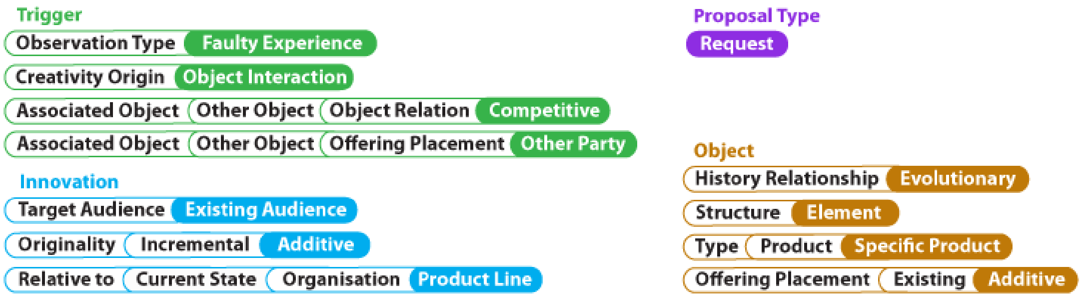
2.1. Example

A basic example of annotations created for a sample idea with Gi2MO Types taxonomy:

Idea Title: More buttons on the tablet.

Idea Summary:
I would like to propose adding more physical buttons on the 9 inch tablet that is currently available in your offer. When using the tablet I feel that the single 'home' button is not enough for many activities that the tablet is advertised for making the experience bad.

For example, for reading ebooks, it would be very useful to have "back" and "forward" buttons for scrolling pages of the book. I own a e-paper reader and I think those buttons could be also used for different activities (for example web browsing or games).



The listing below shows the metadata of the same idea annotated with the use of Gi2MO Idea Ontology and Gi2MO Types taxonomy:

```
<rdf:Description rdf:about="http://ideas.gi2mo.org/idea/012345/rdf">
  <foaf:page ref:resource="http://ideas.gi2mo.org/idea/012345"/>
  <gi2mo:hasCreator rdf:resource="http://ideas.gi2mo.org/user/Pablo/rdf"/>
  <gi2mo:content>I would like to propose adding more physical buttons on the 9 inch tablet that
  is currently available in your offer. When using the tablet I feel that the single 'home'
   button is not enough for many activities that the tablet is advertised for making the
   experience bad.

  For example, for reading ebooks, it would be very useful to have "back" and "forward" buttons
  for scrolling pages of the book. I own a e-paper reader and I think those buttons could be
  also used for different activities (for example web browsing or games).
  </gi2mo:content>
  <dcterms:title>More buttons on the tablet.</gi2mo:title>
  <dcterms:created>2011-07-23</gi2mo:created>
  <gi2mo:hasStatus rdf:resource="http://purl.org/gi2mo/ns#Implemented"/>
  <gi2mo:hasComment rdf:resource="http://ideas.gi2mo.org/comment/054321/rdf"/>
  <gi2mo:hasComment rdf:resource="http://ideas.gi2mo.org/comment/054322/rdf"/>
  <gi2mo:hasCategory ref:resource="http://ideas.gi2mo.org/category/General/rdf"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#FaultyExperience"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#ObjectInteraction"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#CompetitiveObject"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#OtherPartyOffering"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#ExistingAudience"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#AdditiveInnovation"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#ProductLine"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#Request"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#Evolutionary"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#SpecificProduct"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#AdditiveChange"/>
  <gi2mo:hasCharacteristics rdf:resource="http://purl.org/gi2mo/types/ns#ElementChange"/>
</rdf:Description>
```

4. Cross-reference for linking Gi2MO Types terms with web metadata

An alphabetical index of Gi2MO Types terms are given below. All the terms are hyperlinked to their detailed description for quick reference.

Below see a comprehensive list of all Gi2MO Types terms and their descriptions.

Class: gi2mo_types:AdditiveChange

URI: `http://purl.org/gi2mo/types/ns#AdditiveChange`

AdditiveChange - The innovation proposes adding some new element to the object.

sub-class-of:
`gi2mo_types:ExistingObject`

[back to top]

Class: gi2mo_types:AdditiveInnovation

URI: `http://purl.org/gi2mo/types/ns#AdditiveInnovation`

AdditiveInnovation - Improvement made by adding a new element to a previous proposal.

sub-class-of:
`gi2mo_types:Incremental`

[back to top]

Class: gi2mo_types:AssociatedObject

URI: `http://purl.org/gi2mo/types/ns#AssociatedObject`

AssociatedObject - Object connected to the trigger of innovation

sub-class-of:
`gi2mo_types:Trigger`

[back to top]

Class: gi2mo_types:Branch

URI: `http://purl.org/gi2mo/types/ns#Branch`

Branch - Innovation relative to branch level of the organization (e.g. to use when the organization has already introduced the innovation in a certain branch but in antoher branch its completly unknown).

sub-class-of:
`gi2mo_types:StructureUnit`

[back to top]

Class: gi2mo_types:CharacteristicChange

URI: `http://purl.org/gi2mo/types/ns#CharacteristicChange`

CharacteristicChange - Innovation relates to particular feature of an object.

sub-class-of:
`gi2mo_types:Structure`

[back to top]

Class: gi2mo_types:CompetitiveObject

URI: http://purl.org/gi2mo/types/ns#CompetitiveObject

CompetitiveObject - If the object that triggered innovation was a competitive solution

sub-class-of:
gi2mo_types:ObjectRelation

[back to top]

Class: gi2mo_types:ComplementaryObject

URI: http://purl.org/gi2mo/types/ns#ComplementaryObject

ComplementaryObject - If the object that triggered innovation was a complementary solution that works together with the object to fulfil some OTHER function/goal. Removing one of the objects will make it impossible to fulfil this new function.

sub-class-of:
gi2mo_types:ObjectRelation

[back to top]

Class: gi2mo_types:ComplementaryWith

URI: http://purl.org/gi2mo/types/ns#ComplementaryWith

ComplementaryWith - Object is complementary with some other object.

sub-class-of:
gi2mo_types:Relationships

[back to top]

Class: gi2mo_types:CompleteChange

URI: http://purl.org/gi2mo/types/ns#CompleteChange

CompleteChange - Innovation related to the entire product.

sub-class-of:
gi2mo_types:Structure

[back to top]

Class: gi2mo_types:ComposedOf

URI: http://purl.org/gi2mo/types/ns#ComposedOf

ComposedOf - Is built of some other objects.

sub-class-of:
gi2mo_types:Relationships

[back to top]

Class: gi2mo_types:CreativityOrigin

URI: http://purl.org/gi2mo/types/ns#CreativityOrigin

CreativityOrigin - Points to what kind of event triggered the innovation.

sub-class-of:
gi2mo_types:Trigger

[back to top]

Class: gi2mo_types:CurrentState

URI: http://purl.org/gi2mo/types/ns#CurrentState

CurrentState - Innovation described in relation to the current state of some entity or environment.

sub-class-of:
gi2mo_types:RelativeTo

[back to top]

Class: gi2mo_types:Department

URI: http://purl.org/gi2mo/types/ns#Department

Department - Innovation described with respect to department level of the company.

sub-class-of:

gi2mo_types:StructureUnit

[back to top]

Class: gi2mo_types:Dependence

URI: http://purl.org/gi2mo/types/ns#Dependence

Dependence - Innovation is dependent in some way on some other idea.

sub-class-of:
gi2mo_types:Innovation

[back to top]

Class: gi2mo_types:Duplicates

URI: http://purl.org/gi2mo/types/ns#Duplicates

Duplicates - Innovation that exactly reproduces already existing idea.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:ElementChange

URI: http://purl.org/gi2mo/types/ns#ElementChange

ElementChange - Innovation relates to an element of a product (a subelement or a characteristic of an element).

sub-class-of:
gi2mo_types:Structure

[back to top]

Class: gi2mo_types:Encapsulates

URI: http://purl.org/gi2mo/types/ns#Encapsulates

Ecapsulates - Innovation that reproduces other idea(s) but also adds new value.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:Event

URI: http://purl.org/gi2mo/types/ns#Event

Event - The innovation was not triggered by direct innovation with some object but by some event.

sub-class-of:
gi2mo_types:CreativityOrigin

[back to top]

Class: gi2mo_types:Evolutionary

URI: http://purl.org/gi2mo/types/ns#Evolutionary

Evolutionary - Indicates that the innovation for the given product puts it in a state that it has never been before.

sub-class-of:
gi2mo_types:HistoryRelationship

[back to top]

Class: gi2mo_types:Excludes

URI: http://purl.org/gi2mo/types/ns#Excludes

Excludes - Introducing this innovation causes the inability to implement some other innovation.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:ExistingAudience

URI: http://purl.org/gi2mo/types/ns#ExistingAudience

ExistingAudience - Adaptive Innovation. Updates of an existing object to fit the changing needs of current customers.

sub-class-of:
[gi2mo_types:TargetAudience]

[[back to top]]

Class: gi2mo_types:ExistingObject

URI: http://purl.org/gi2mo/types/ns#ExistingObject

ExistingObject - The innovation proposes innovating an existing product.

sub-class-of:
gi2mo_types:OwnOfferingPlacement

[back to top]

Class: gi2mo_types:FaultyExperience

URI: http://purl.org/gi2mo/types/ns#FaultyExperience

FaultyExperience - A problem/ failure of a product that caused a the creative process.

sub-class-of:
gi2mo_types:ObservationType

[back to top]

Class: gi2mo_types:Follows

URI: http://purl.org/gi2mo/types/ns#Follows

Follows - The innovation realization is dependent on implementing some other idea first.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:GlobalMarket

URI: http://purl.org/gi2mo/types/ns#GlobalMarket

Global - Innovation described with respect to the global market.

sub-class-of:
gi2mo_types:Market

[back to top]

Class: gi2mo_types:Group

URI: http://purl.org/gi2mo/types/ns#Group

Group - Innovation described with respect to group level of the company.

sub-class-of:
gi2mo_types:StructureUnit

[back to top]

Class: gi2mo_types:HistoryRelationship

URI: http://purl.org/gi2mo/types/ns#HistoryRelationship

HistoryRelationship - Relationship of the innovation with the previous iterations of the objects (or previous innovation proposals for this object). The difference with "Offering placement" is that it looks to compare the change with the past states of the object, not the current).

sub-class-of:
gi2mo_types:Object

[back to top]

Class: gi2mo_types:IdeaContest

URI: http://purl.org/gi2mo/types/ns#IdeaContest

IdeaContest - Idea was specifically triggered by idea contests and incentives that come with.

sub-class-of:
gi2mo_types:Event

[back to top]

Class: gi2mo_types:Incremental

URI: http://purl.org/gi2mo/types/ns#Incremental

Incremental - Improvement for an existing object.

sub-class-of:
gi2mo_types:Originality

[back to top]

Class: gi2mo_types:Innovation

URI: http://purl.org/gi2mo/types/ns#Innovation

Innovation - Characteristics of the proposal and its impact.

[back to top]

Class: gi2mo_types:InnovationProposal

URI: http://purl.org/gi2mo/types/ns#InnovationProposal

InnovationProposal - Innovation described in relation to other proposals in the idea management system.

sub-class-of:
gi2mo_types:RelativeTo

[back to top]

Class: gi2mo_types:IsPartOf

URI: http://purl.org/gi2mo/types/ns#IsPartOf

IsPartOf - Innovation that has been also described in another idea that also shows a bigger scope and discusses other items.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:IssueReport

URI: http://purl.org/gi2mo/types/ns#IssueReport

IssueReport - Some idea are not innovative at all but nevertheless they are submitted to the idea management systems. This caregory indicates idea is not innovative at all just reports on a bug/issue/error/failure of an object.

sub-class-of:
gi2mo_types:ProposalType

[back to top]

Class: gi2mo_types:LackOfFeature

URI: http://purl.org/gi2mo/types/ns#LackOfFeature

LackOfFeature - The innovation was triggered by the fact that object lacked some kind of feature.

sub-class-of:
gi2mo_types:ObservationType

[back to top]

Class: gi2mo_types:LocalMarket

URI: http://purl.org/gi2mo/types/ns#LocalMarket

Local - Innovation described with respect to the local market.

sub-class-of:
gi2mo_types:Market

[back to top]

Class: gi2mo_types:Market

URI: http://purl.org/gi2mo/types/ns#Market

Market - Innovation described with respect to the entire market

sub-class-of:
gi2mo_types:CurrentState

[back to top]

Class: gi2mo_types:New

URI: http://purl.org/gi2mo/types/ns#New

New - Completely new idea/design not based on anything created before.

sub-class-of:
gi2mo_types:Originality

[back to top]

Class: gi2mo_types:NewAudience

URI: http://purl.org/gi2mo/types/ns#NewAudience

NewAudience - Shifted Innovation. It relates to shifting the object to fit new market/customer.

sub-class-of:
gi2mo_types:TargetAudience

[back to top]

Class: gi2mo_types:NewObject

URI: http://purl.org/gi2mo/types/ns#NewObject

NewObject - The innovation proposes a new product.

sub-class-of:
gi2mo_types:OwnOfferingPlacement

[back to top]

Class: gi2mo_types:NewObjectRelease

URI: http://purl.org/gi2mo/types/ns#NewObjectRelease

NewObjectRelease - Premiere of a fully new object.

sub-class-of:
gi2mo_types:ObjectRelease

[back to top]

Class: gi2mo_types:None

URI: http://purl.org/gi2mo/types/ns#None

None - This is a special instance for idea mgt systems that indicates that idea hold no real innovation at all

sub-class-of:
gi2mo_types:Originality

[back to top]

Class: gi2mo_types:Object

URI: http://purl.org/gi2mo/types/ns#Object

Object - Characteristics of the item that is the object of the innovation

[back to top]

Class: gi2mo_types:ObjectComparison

URI: http://purl.org/gi2mo/types/ns#ObjectComparison

ObjectComparison - Idea was triggered by comparison of the object of innovation with the some other product/object.

sub-class-of:
gi2mo_types:Event

[back to top]

Class: gi2mo_types:ObjectInteraction

URI: http://purl.org/gi2mo/types/ns#ObjectInteraction

ObjectInteraction - A person interacting with the object and discovering some new need or problem.

sub-class-of:
gi2mo_types:CreativityOrigin

[back to top]

Class: gi2mo_types:ObjectObservation

URI: http://purl.org/gi2mo/types/ns#ObjectObservation

ObjectObservation - The idea was initiated by observing an object (e.g. due to the desired to purchase it or general interest in the object).

sub-class-of:
[gi2mo_types:Event]

[back to top]

Class: gi2mo_types:ObjectOfInnovation

URI: http://purl.org/gi2mo/types/ns#ObjectOfInnovation

ObjectOfInnovation - Innovation was triggered by interaction with the same object that is the topic of idea.

sub-class-of:
[gi2mo_types:AssociatedObject]

[back to top]

Class: gi2mo_types:ObjectPurchase

URI: http://purl.org/gi2mo/types/ns#ObjectPurchase

ObjectPurchase - Purchase of the object by the client.

sub-class-of:
[gi2mo_types:Event]

[back to top]

Class: gi2mo_types:ObjectRelation

URI: http://purl.org/gi2mo/types/ns#ObjectRelation

ObjectRelation - Defines the relation of the object that triggered the idea and the object that is the topic of the idea.

sub-class-of:
[gi2mo_types:OtherObject]

[back to top]

Class: gi2mo_types:ObjectRelease

URI: http://purl.org/gi2mo/types/ns#ObjectRelease

ObjectRelease - Premiere of an object to the market or audience.

sub-class-of:
[gi2mo_types:Event]

[back to top]

Class: gi2mo_types:ObjectUpdateRelease

URI: http://purl.org/gi2mo/types/ns#ObjectUpdateRelease

ObjectUpdateRelease - Premiere of a new refreshed version of an already existing object.

sub-class-of:
[gi2mo_types:ObjectRelease]

[back to top]

Class: gi2mo_types:ObservationType

URI: http://purl.org/gi2mo/types/ns#ObservationType

ObservationType - Type of Trigger. Defines what was the observation of the inventor due to the event that he experienced.

sub-class-of:
[gi2mo_types:Trigger]

[back to top]

Class: gi2mo_types:OfferingPlacement

URI: http://purl.org/gi2mo/types/ns#OfferingPlacement

OfferingPlacement - Determines if the other object that is the means of comparison is part of own offering or some other party.

sub-class-of:
[gi2mo_types:OtherObject]

[back to top]

Class: gi2mo_types:Organization

URI: http://purl.org/gi2mo/types/ns#Organization

Organization - Innovation described with respect to organization processes/product/services

sub-class-of:

gi2mo_types:CurrentState

[back to top]

Class: gi2mo_types:OrganizationsProductLine

URI: http://purl.org/gi2mo/types/ns#OrganizationsProductLine

ProductLine - Innovation originality is described with relation to a particular product line (e.g. to be used when innovation is based on what has been already achieved in some other product line).

sub-class-of:

gi2mo_types:Organization

[back to top]

Class: gi2mo_types:OrganizationsProductType

URI: http://purl.org/gi2mo/types/ns#OrganizationsProductType

ProductType - Innovation originality described in relation to product type (e.g. to be used when certain products in the organisation already have the innovation but a similar innovation is proposed for a totally new product type).

sub-class-of:

gi2mo_types:Organization

[back to top]

Class: gi2mo_types:Originality

URI: http://purl.org/gi2mo/types/ns#Originality

Originality - Defines the size/scope of innovation in relation to existing ideas, state or offering.

sub-class-of:

gi2mo_types:Innovation

[back to top]

Class: gi2mo_types:OtherEvent

URI: http://purl.org/gi2mo/types/ns#OtherEvent

OtherEvent - Undefined event related in some way to the object.

sub-class-of:

gi2mo_types:Event

[back to top]

Class: gi2mo_types:OtherObject

URI: http://purl.org/gi2mo/types/ns#OtherObject

OtherObject - The inventor has came up with an idea by comparison of the object with some other product/service/process.

sub-class-of:

gi2mo_types:AssociatedObject

[back to top]

Class: gi2mo_types:OtherPartyOffering

URI: http://purl.org/gi2mo/types/ns#OtherPartyOffering

OtherPartyOffering - Object that triggered the innovative ideas come from offering of some other party.

sub-class-of:

gi2mo_types:OfferingPlacement

[back to top]

Class: gi2mo_types:OwnOffering

URI: http://purl.org/gi2mo/types/ns#OwnOffering

OwnOffering - Object that triggered the innovative ideas come from own offering.

sub-class-of:
[gi2mo_types:OfferingPlacement]

[back to top]

Class: gi2mo_types:OwnOfferingPlacement

URI: http://purl.org/gi2mo/types/ns#OwnOfferingPlacement

OwnOfferingPlacement - Object placement in the current offering / state.

sub-class-of:
gi2mo_types:Object

[back to top]

Class: gi2mo_types:PartOf

URI: http://purl.org/gi2mo/types/ns#PartOf

PartOf - Object is part of some other object.

sub-class-of:
gi2mo_types:Relationships

[back to top]

Class: gi2mo_types:PotentialCause

URI: http://purl.org/gi2mo/types/ns#PotentialCause

PotentialCause - If a cause of the problem was identified.

sub-class-of:
gi2mo_types:FaultyExperience

[back to top]

Class: gi2mo_types:PotentialOportunity

URI: http://purl.org/gi2mo/types/ns#PotentialOportunity

PotentialOportunity - The observation was not due to an negative experience but assumption that the object could be better if extended in some way.

sub-class-of:
gi2mo_types:ObservationType

[back to top]

Class: gi2mo_types:Proceeds

URI: http://purl.org/gi2mo/types/ns#Proceeds

Proceeds - After the described innovation is implemented it is going to be possible to unlock other idea.

sub-class-of:
gi2mo_types:Dependence

[back to top]

Class: gi2mo_types:Process

URI: http://purl.org/gi2mo/types/ns#Process

Process - Innovation relates to a process. Eg. new elements into an organization’s operations such as input materials, task specifications, work and information flow mechanisms, and equipment used to produce a product or render a service.

sub-class-of:
gi2mo_types:Type

[back to top]

Class: gi2mo_types:Product

URI: http://purl.org/gi2mo/types/ns#Product

Product - Innovation relates to a product.

sub-class-of:
gi2mo_types:Type

[back to top]

Class: gi2mo_types:ProductLine

URI: http://purl.org/gi2mo/types/ns#ProductLine

ProductLine - Indicates more specifically that the object of innovation is the entire product line.

sub-class-of:
gi2mo_types:Product

[back to top]

Class: gi2mo_types:ProductType

URI: http://purl.org/gi2mo/types/ns#ProductType

ProductType - Innovation is targeted for a specific product type.

sub-class-of:
gi2mo_types:Product

[back to top]

Class: gi2mo_types:PromotionalEvent

URI: http://purl.org/gi2mo/types/ns#PromotionalEvent

PromotionalEvent - Event related to the object of innovation.

sub-class-of:
gi2mo_types:Event

[back to top]

Class: gi2mo_types:ProposalType

URI: http://purl.org/gi2mo/types/ns#ProposalType

ProposalType - Describes the way a person has formalized his ideas (what parts are there, what are missing etc.) What the the completeness of the proposal.

[back to top]

Class: gi2mo_types:Regressive

URI: http://purl.org/gi2mo/types/ns#Regressive

Regressive - Indicates that the innovation for the given product was already implemented at some point but has been removed.

sub-class-of:
gi2mo_types:HistoryRelationship

[back to top]

Class: gi2mo_types:Relationships

URI: http://purl.org/gi2mo/types/ns#Relationships

Relationships - Defines if an object has relationship with other items of offering (reason: adding an innovation into a certain object might impact its relationships with other products and increase/decrease the value of innovation).

sub-class-of:
gi2mo_types:Object

[back to top]

Class: gi2mo_types:RelativeTo

URI: http://purl.org/gi2mo/types/ns#RelativeTo

RelativeTo - Describes innovation characteristics abstraction level (e.g. if something is innovation with respect to local market or global market)

sub-class-of:
gi2mo_types:Originality

[back to top]

Class: gi2mo_types:ReplacementOfElement

URI: http://purl.org/gi2mo/types/ns#ReplacementOfElement

ReplacementOfElement - The innovation proposes replacing or modifying an existing element of the object.

sub-class-of:
gi2mo_types:ExistingObject

[back to top]

Class: gi2mo_types:ReplacementOfInnovationElement

URI: http://purl.org/gi2mo/types/ns#ReplacementOfInnovationElement

ReplacementOfInnovationElement - Improvement made just be replacing some element of the previous proposal.

sub-class-of:
gi2mo_types:Incremental

[back to top]

Class: gi2mo_types:Request

URI: http://purl.org/gi2mo/types/ns#Request

Request - A person has supplied a request for a certain feature.

sub-class-of:
gi2mo_types:ProposalType

[back to top]

Class: gi2mo_types:Service

URI: http://purl.org/gi2mo/types/ns#Service

Service - Innovation relates to a service. Opposing products, services are: intangible (cannot be physically possessed), inseparable (production cannot be separated from consumption), perishable (cannot store for future use), variable (the quality and consistency differs depending on the person that uses it).

sub-class-of:
gi2mo_types:Type

[back to top]

Class: gi2mo_types:Solution

URI: http://purl.org/gi2mo/types/ns#Solution

Solution - The person has supplemented the request with a technical or theoretical solution proposal.

sub-class-of:
gi2mo_types:ProposalType

[back to top]

Class: gi2mo_types:SpecificProduct

URI: http://purl.org/gi2mo/types/ns#SpecificProduct

SpecificProduct - Innovation targeted for a particular product in the offering.

sub-class-of:
gi2mo_types:Product

[back to top]

Class: gi2mo_types:Structure

URI: http://purl.org/gi2mo/types/ns#Structure

Structure - Defines the element in the object structure that the innovation refers to. This should be connected with a domain ontology that identifies specific elements of products etc.

sub-class-of:
gi2mo_types:Object

[back to top]

Class: gi2mo_types:StructureUnit

URI: http://purl.org/gi2mo/types/ns#StructureUnit

StructureUnit - Innovation relative to a certain organization structure element (e.g. when innovation has been already introduced on some level of the organization administrative unit but not in another where it could be a brand new idea).

sub-class-of:
gi2mo_types:Organization

[back to top]

Class: gi2mo_types:SubstractiveChange

URI: http://purl.org/gi2mo/types/ns#SubstractiveChange

SubstractiveChange - The innovation proposes removing some element from the object.

sub-class-of:
gi2mo_types:ExistingObject

[back to top]

Class: gi2mo_types:SubstractiveInnovation

URI: http://purl.org/gi2mo/types/ns#SubstractiveInnovation

SubstractiveInnovation - Improvement made by simplifying other proposal and removing an element from the previous proposal.

sub-class-of:
gi2mo_types:Incremental

[back to top]

Class: gi2mo_types:Suggestion

URI: http://purl.org/gi2mo/types/ns#Suggestion

Suggestion - Contrary to request does not state the necessity or need observed by the innovator but rather says what would be good in the opinion of a person for the particular company or organization.

sub-class-of:
gi2mo_types:ProposalType

[back to top]

Class: gi2mo_types:SuplementaryObject

URI: http://purl.org/gi2mo/types/ns#SuplementaryObject

SuplementaryObject - Innovation triggered by an object that can be used on top of the existing item and delivers some new value by improving the characteristics of the item however does not redefine the end functionality (removing this object will still make it possible to perform the activity).

sub-class-of:
gi2mo_types:ObjectRelation

[back to top]

Class: gi2mo_types:TargetAudience

URI: http://purl.org/gi2mo/types/ns#TargetAudience

TargetAudience - Defines if the innovation imposes a change in the audience or not (relative to current state of the object of innovation).

sub-class-of:
gi2mo_types:Innovation

[back to top]

Class: gi2mo_types:Trigger

URI: http://purl.org/gi2mo/types/ns#Trigger

Trigger - What has caused creation of the idea / what is the inspiration

[back to top]

Class: gi2mo_types:Type

URI: http://purl.org/gi2mo/types/ns#Type

Type - Type of object (is it product, service etc.).

sub-class-of:
gi2mo_types:Object

[back to top]

Class: gi2mo_types:UnrelatedObject

URI: http://purl.org/gi2mo/types/ns#UnrelatedObject

UnrelatedObject - The innovation was triggered by neither complementary or competitive solution.

sub-class-of:
gi2mo_types:ObjectRelation

[back to top]

A Changelog

2011-06-10

- Added a number of terms based on ontology evaluation with live datasets
  - Replaced "ModificationOfInnovtion" and "Modification" with "ReplacementOfInnovationElement" and "ReplacementOfElement"
  - "Originality/Breakthrough" changed to "Originality/New"
  - Under "Originality/RelatedTo" adde references to "OrganisationsProductLine", "OrganisationsProductType", also put organisation administrative units under "StructureUnut" and added "Branch"
  - Added "Trigger/CreativityOrigin/ObjectPurchase"
  - Added "Trigger/CreativityOrigin/ObjectObservation"
  - Added "Trigger/TriggerType/PotentialOpportunity"
  - Added "Trigger/AssociatedObject/OtherObject/ObjectType/Supplementary"
  - Added "Trigger/CreativityOrigin/ObjectComparison"
  - Uner "Object/Type/Product" added "ProductLine", "ProductType" and "SpecificProduct"
  - Added "ProposalType/IssueReport"
  - Added "ProposalType/Suggestion"
  - Under "Innovation/Dependence" added "Excludes" and "isPartOf"

2011-05-30

- First version of the document

B Acknowledgements

This documentation has been generated automatically from the most recent ontology specification in OWL using a python script called SpecGen. The style formatting has been inspired on FOAF specification.

Special thanks for support with ontology creation and research to: Prof. Carlos A. Iglesias and members of the GSI Group of DIT department of Universidad Politécnica de Madrid and Dr. Theodore Dalamagas, Nikos Bikakis and Giorgos Giannopoulos from IMIS institute.

Appendix D

Marl Ontology Specification

One of the contributions of the thesis is the Marl Ontology - a formal specification of a conceptualization of opinions as expressed in the collaborative systems on the World Wide Web. The following appendix contains a summarized version of the ontology specification. In comparison, the full specification available on-line is a more structured document with a grater number links and back-references that facilitate improved specification browsing in the web environment. The summarized version presented in the appendix contains all the information necessary to model nuser opinions as proposed by the thesis.

The following appendix is primary a supplement for chapter 5, which describes the research done on modelling of community opinions on the Web and the use of this model for Idea Management Systems. For details of evaluation of this taxonomy in terms of coverage of data model from generic Web systems as well as performance of application of the ontology for Idea Management Systems see chapter 6.

gsi UPM Grupo de Sistemas Inteligentes

Marl

Marl Ontology Specification

V0.2 - 13 March 2011

This version:	http://purl.org/marl/0.2/ns (RDF/XML, HTML)
Latest version:	http://purl.org/marl/ns
Previous version:	http://purl.org/marl/0.1/ns
Editors:	Adam Westerski
Authors:	Adam Westerski
Contributors:	See acknowledgements

Abstract

Marl is a standardised data schema (also referred as "ontology" or "vocabulary") designed to annotate and describe subjective opinions expressed on the web or in particular Information Systems. The following document contains the description of ontology and instructions how to connect it with descriptions of other resources.

Table of Contents

Appendixes

1 Introduction

The following specification is a formal description of metadata schema proposal that can be applied to data representing subjective opinions published on the Web. The goal of the following section is to provide the basic knowledge to comprehend the technical part of the specification. As such it shall introduce both Semantic Web and general topic of opinion representation and sentiment analysis.

An important note is that Marl ontology presented here is not a complete model to address the problem of describing and linking opinions online and inside information systems. It marly defines concepts that are not described yet by the means of other ontologies and provides the data attributes that enable to connect opinions with contextual information already defined in metadata created with other ontologies. For detailed instructions and recommendations how to fully model opinions and the results of opinion mining process refer to analysis done by Gi2MO project.

1.1 Opinions on the Web and the opinion mining process

With the birth of Web 2.0 users started to provide their input and create content on mass scape about their subjective opinions related to various topics (e.g. opinions about movies). While this kind of content can be very beneficial for many different uses (e.g. market analysis or predictions) it's accurate analysis and interpretation has not been fully harnessed yet. Information left by the users is often very disorganised and many portals that enable user input leave the user added information unmoderated.

Opinion mining (often referred as sentiment analysis) is one of the attempts bring order to those vast amounts of user generated content. The domain focuses to analyse textual content using special language processing tools and as output provides a quantified judgement of the sentiments contained in the text (e.g. if the text expresses a positive or negative opinion).

Due to the complexity of the problem and attempts to provide efficient and fast tools the area can be divided into three main research directions:

- document wide sentiment analysis
- sentence sentiment analysis
- feature-based sentiment analysis

In relation to the World Wide Web, there is a number of common uses of opinion formalisation and analysis. Firstly, it can be applied on top of search engines to find the desired content and next run it through opinion analysis software to obtain desired statistics (e.g. Swotti). Secondly, such algorithms can used within dedicated systems that use the Web to connect to particular communities and gather their opinions on very specific topics (e.g. Internet shops or review websites).

In relation to the dedicated systems (e.g. Enterprise Systems), there the community collaborative models that have proven successful in the open web are often transferred to large enterprise to enhance knowledge exchange and bring the employees together. The same opinion mining techniques can be applied in such cases to extract particular information and use it for internal statistics and to improve knowledge search across the enterprise (e.g. see use of opinion mining in Idea Management [link]).

1.2 The Semantic Web

The Semantic Web is a W3C initiative that aims to introduce rich metadata to the current Web and provide machine readable and processable data as a supplement to human-readable Web.

Semantic Web is a mature domain that has been in research phase for many years and with the increasing amount of commercial interest and emerging products is starting to gain appreciation and popularity as one of the rising trends for the future Internet.

One of the corner stones of the Semantic Web is research on inter-linkable and interoperable data schemas for information published online. Those schemas are often referred to as ontologies or vocabularies. In order to facilitate the concept of ontologies that lead to a truly interoperable Web of Data, W3C has proposed a series of technologies such as RDF and OWL. **Marl** uses those technologies and the research that comes within to propose an ontology for the particular goal of describing opinions and linking them with contextual information (such as opinion topic, features described in the opinion etc.).

1.3 What is Marl for?

The goals of the Marl ontology to achieve as a data schema are:

- enable to publish raw data about opinions and the sentiments expressed in them
- deliver schema that will allow to compare opinions coming from different systems (polarity, topics, features)
- interconnect opinions by linking them to contextual information expressed with concepts from other popular ontologies or specialised domain ontologies

For more information please refer to Marl usage study done as part of the research in the Gi2MO project.

2. Marl ontology at a glance

An alphabetical index of Marl terms, by class (concepts) and by property (relationships, attributes), are given below. All the terms are hyperlinked to their detailed description for quick reference.

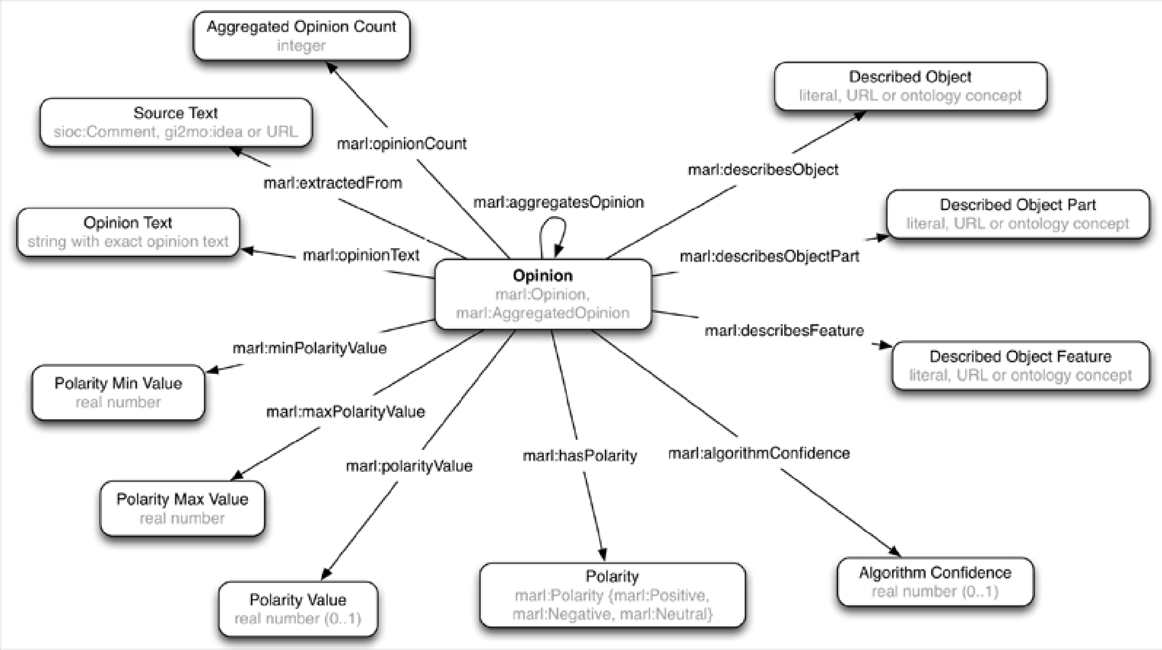
Classes: AggregatedOpinion, Opinion, Polarity,

Properties: aggregatesOpinion, algorithmConfidence, describesFeature, describesObject, describesObjectPart, extractedFrom, hasOpinion, hasPolarity, maxPolarityValue, minPolarityValue, negativeOpinionCount, neutralOpinionCount, opinionCount, opinionText, polarityValue, positiveOpinionsCount,

Instances: Negative, Neutral, Positive,

3. Marl ontology overview

The Marl class diagram presented below shows connections between classes and properties used for describing opinions.



Class and Properties Diagram for the Marl Ontology

3.1. Example

A very basic example below shows a single opinion annotated with Marl metadata (the second class maps the opinion structure and is shown as reference):

```
<rdf:Description rdf:about="http://gi2mo.org/marl/avatar/opinion/012345/rdf">
  <marl:extractedFrom rdf:resource="http://gi2mo.org/marl/blog/avatar-review/comment/321/rdf"/>
  <marl:describesObject rdf:resource="http://dbpedia.org/resource/Avatar_(2009_film)"/>
  <marl:describesFeature rdf:resource="http://dbpedia.org/property/runtime"/>
  <marl:polarityValue>-0.2</marl:polarityValue>
  <marl:minPolarityValue>-1</marl:minPolarityValue>
  <marl:maxPolarityValue>1</marl:maxPolarityValue>
  <marl:hasPolarity rdf:resource="http://purl.org/marl/ns#Negative"/>
  <rdf:type rdf:resource="http://purl.org/marl/ns#Opinion"/>
</rdf:Description>

<rdf:Description rdf:about="http://gi2mo.org/marl/blog/avatar-review/comment/321/rdf">
  <dcterms:title>Re: Avatar Review</dcterms:title>;
  <sioc:has_creator rdf:resource="http://gi2mo.org/marl/blog/author/user345/"/>
  <dcterms:created>Fri, 3 Jun 2010 13:53:54 +0200</dcterms:created>;
  <sioc:reply_of rdf:resource="http://gi2mo.org/marl/blog/avatar-review/"/>
  <sioc:content>Awful movie, way to long!</sioc:content>
  <foaf:primaryTopic rdf:resource="http://gi2mo.org/marl/blog/avatar-review/comment/321"/>
  <rdf:type rdf:resource="http://rdfs.org/sioc/ns#Post"/>
</rdf:Description>
```

For more examples please see a Marl RDF export for a opinions taken from a simple idea management system instance installed for ETSIT school of Universidad Politécnica de Madrid. Furthermore, we recommend reading Marl Use Cases document for more examples and hints how to properly describe opinions with the ontology.

4. Cross-reference for Marl classes and properties

Below see a comprehensive list of all Marl classes, properties and their descriptions.

Class: marl:AggregatedOpinion

URI: http://purl.org/marl/ns#AggregatedOpinion

AggregatedOpinion - The same as Opinion class but indicates that the properties of this class aggregate all the opinions specified in the "extractedFrom" source. Optionally, if the aggregatesOpinion property is used this class could be created to aggregate only certain opinions (e.g. in a text about political scene it there could be many AggregatedOpinion classes each with opinions per different politician).

sub-class-of:
- marl:Opinion

in-domain-of:
- marl:aggregatesOpinion
- marl:opinionCount
- marl:positiveOpinionsCount
- marl:negativeOpinionCount
- marl:neutralOpinionCount

Class: marl:Opinion

[back to top]

URI: http://purl.org/marl/ns#Opinion

Opinion - Describes the concept of opinion expressed in a certain text.

in-domain-of:

- marl:describesObjectPart
- marl:extractedFrom
- marl:hasPolarity
- marl:describesFeature
- marl:describesObject
- marl:polarityValue
- marl:minPolarityValue
- marl:opinionText
- marl:algorithmConfidence
- marl:maxPolarityValue

in-range-of:

- marl:aggregatesOpinion
- marl:hasOpinion

[back to top]

Class: marl:Polarity

URI: http://purl.org/marl/ns#Polarity

Polarity - Class that represents the opinion polarity. Use instances to express if the polarity is positive, neutral or negative.

in-range-of:

- marl:hasPolarity

[back to top]

Property: marl:aggregatesOpinion

URI: http://purl.org/marl/ns#aggregatesOpinion

aggregatesOpinion - Indicates that the polarity described with the class is a calculation (eg. sum) of other opinions polarity (eg. aggregated opinion about the movie derived from many sentiments expressed in one text).

OWL Type:
ObjectProperty
Domain:
marl:AggregatedOpinion
Range:
marl:Opinion

[back to top]

Property: marl:algorithmConfidence

URI: http://purl.org/marl/ns#algorithmConfidence

algorithmConfidence - A numerical value that describe how much the algorithm was confident of the assessment of the opinion (eg. how much the opinion matches a gives object/product).

OWL Type:
DatatypeProperty
Domain:
marl:Opinion

[back to top]

Property: marl:describesFeature

URI: http://purl.org/marl/ns#describesFeature

describesFeature - Indicates a feature of an object or object part that the opinion refers to (eg. laptop battery life or laptop battery size etc.).

OWL Type:
ObjectProperty
Domain:
marl:Opinion

[back to top]

Property: marl:describesObject

URI: http://purl.org/marl/ns#describesObject

describesObject - Indicates the object that the opinion refers to.

OWL Type:
ObjectProperty
Domain:
`marl:Opinion`

[back to top]

Property: marl:describesObjectPart

URI: `http://purl.org/marl/ns#describesObjectPart`

describesObjectPart - Indicates a particular element or part of the object that the opinion refers to (eg. laptop screen or camera battery).

OWL Type:
ObjectProperty
Domain:
`marl:Opinion`

[back to top]

Property: marl:extractedFrom

URI: `http://purl.org/marl/ns#extractedFrom`

extractedFrom - Indicates the text from which the opinion has been extracted.

Inverse:
`marl:hasOpinion`
OWL Type:
ObjectProperty
Domain:
`marl:Opinion`

[back to top]

Property: marl:hasOpinion

URI: `http://purl.org/marl/ns#hasOpinion`

hasOpinion - Indicates that a certain text has a subjective opinion expressed in it.

Inverse:
`marl:extractedFrom`
OWL Type:
ObjectProperty
Range:
`marl:Opinion`

[back to top]

Property: marl:hasPolarity

URI: `http://purl.org/marl/ns#hasPolarity`

hasPolarity - Indicates if the opinion is positive/negative or neutral. Use instances of class marl:Polarity.

OWL Type:
ObjectProperty
Domain:
`marl:Opinion`
Range:
`marl:Polarity`

[back to top]

Property: marl:maxPolarityValue

URI: `http://purl.org/marl/ns#maxPolarityValue`

algorithmConfidence - Maximal possible numerical value for the opinion.

OWL Type:
DatatypeProperty
Domain:
`marl:Opinion`

[back to top]

Property: marl:minPolarityValue

URI: `http://purl.org/marl/ns#minPolarityValue`

minPolarityValue - Lowest possible numerical value of the opinion.

OWL Type:
DatatypeProperty
Domain:
marl:Opinion

[back to top]

Property: marl:negativeOpinionCount

URI: http://purl.org/marl/ns#negativeOpinionCount

negativeOpinionCount - Amount of negative opinions aggregated.

OWL Type:
DatatypeProperty
Domain:
marl:AggregatedOpinion

[back to top]

Property: marl:neutralOpinionCount

URI: http://purl.org/marl/ns#neutralOpinionCount

neutralOpinionCount - Amount of neutral opinions aggregated.

OWL Type:
DatatypeProperty
Domain:
marl:AggregatedOpinion

[back to top]

Property: marl:opinionCount

URI: http://purl.org/marl/ns#opinionCount

opinionCount - Amount of all aggregated opinions.

OWL Type:
DatatypeProperty
Domain:
marl:AggregatedOpinion

[back to top]

Property: marl:opinionText

URI: http://purl.org/marl/ns#opinionText

opinionText - The exact text that expresses the opinion. This can be used when entity/text pointed by extractedFrom contains many opinions. For example extractedFrom can point to a comment that contains many opinions about a movie, each opinion should have a separate marl:Opinion and optionally an opinionText property to indicate the specific text fragment of the comment.

OWL Type:
DatatypeProperty
Domain:
marl:Opinion

[back to top]

Property: marl:polarityValue

URI: http://purl.org/marl/ns#polarityValue

polarityValue - A numerical representation of the polarity value. The recommended use is by specifying % by using a real number from 0..1. In case this is not feasible in a given solution use minOpinionValue and maxOpinionValue to provide additional information.

OWL Type:
DatatypeProperty
Domain:
marl:Opinion

[back to top]

Property: marl:positiveOpinionsCount

URI: http://purl.org/marl/ns#positiveOpinionsCount

positiveOpinionCount - Amount of positive opinions aggregated.

OWL Type:
DatatypeProperty

Domain:
`marl:AggregatedOpinion`

[back to top]

Instance: marl:Negative

URI: `http://purl.org/marl/ns#Negative`

Negative - Negative polarity.

RDF Type:
`marl:Polarity`

[back to top]

Instance: marl:Neutral

URI: `http://purl.org/marl/ns#Neutral`

Neutral - Neutral polarity

RDF Type:
`marl:Polarity`

[back to top]

Instance: marl:Positive

URI: `http://purl.org/marl/ns#Positive`

Positive - Positive polarity

RDF Type:
`marl:Polarity`

[back to top]

A Changelog

2011-03-13

- Added a number of properties based on ontology evaluation with live datasets
  - marl:opinionCount, marl:negativeOpinionCount, marl:positiveOpinionCount, marl:neutralOpinionCount can be used with AggregatedOpinion class to express amount of aggregated opinions
  - marl:opinionText can be used with Opinion class to describe the exact text extract that the Opinion class describes (in case extractedFrom has many opinions)

2011-01-27

- First version of the document

B Acknowledgements

This documentation has been generated automatically from the most recent ontology specification in OWL using a python script called SpecGen. The style formatting has been inspired on FOAF specification.

Special thanks for support with ontology creation and research to: Prof. Carlos A. Iglesias and members of the GSI Group of DIT department of Universidad Politécnica de Madrid.