UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

DESIGN AND DEVELOPMENT OF GAMIFIED SMART OBJECTS FOR MUSEUMS BASED ON AUTOMATICALLY GENERATED QUIZZES EXPLOITING LINKED DATA.

APPLICATION: TELECOMMUNICATIONS MUSEUM AT ETSIT.

ALEJANDRO LÓPEZ MARTÍNEZ

JULIO 2019

TRABAJO DE FIN DE GRADO

Título:	Diseño y desarrollo de objetos inteligentes gamificados basados en la generación automática de preguntas explotando Linked Data. Aplicación: Museo de Telecomunicación de la ETSIT.
Título (inglés):	Design and development of gamified smart objects for museums based on automatically generated quizzes exploiting Linked Data. Application: Telecommunications Museum at ETSIT.
Autor:	Alejandro López Martínez
Tutor:	Carlos Á. Iglesias Fernández
Departamento:	Departamento de Ingeniería de Sistemas Telemáticos

MIEMBROS DEL TRIBUNAL CALIFICADOR

Presidente:	—
Vocal:	—
Secretario:	—
Suplente:	—

FECHA DE LECTURA:

CALIFICACIÓN:

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE
INGENIEROS DE TELECOMUNICACIÓN

Departamento de Ingeniería de Sistemas Telemáticos
Grupo de Sistemas Inteligentes

TRABAJO FIN DE GRADO

DESIGN AND DEVELOPMENT OF GAMIFIED SMART OBJECTS FOR MUSEUMS BASED ON AUTOMATICALLY GENERATED QUIZZES EXPLOITING LINKED DATA.

APPLICATION: TELECOMMUNICATIONS MUSEUM AT ETSIT.

Alejandro López Martínez

Julio 2019

Resumen

El proyecto pretende realizar una plataforma de gamificación para revalorizar obras culturales, mejorando la experiencia de los usuarios al visitar un museo. En concreto, el proyecto tiene como objetivo realizar mediante minería de textos una identificación de las entidades de las descripciones de los objetos culturales, y su posterior enlace mediante técnicas semánticas con las entidades de la base de datos enlazados DBPedia

Un módulo de generación automática de preguntas permitirá a los visitantes conocer más de las obras expuestas, empleando una técnica de gamificación para capturar su atención. Este módulo será realizado mediante un módulo de inferencia semántica, mediante la generación de plantillas SPARQL. Con el fin de determinar el objeto, se emplearán códigos QR que serán leídos por una aplicación móvil, que mostrará el juego.

El sistema se validará en el museo de la ETSIT de Telecomunicación, y podrá ser utilizado tanto por los visitantes como para nuevos alumnos en las jornadas de puertas abiertas, pudiendo ser extendido posteriormente a otros elementos de la escuela (p.ej. para las visitas al departamento DIT, cámara anecoica, casa solar, etc.), y pudiéndose también adaptar a otros museos.

Palabras clave: Web Semántica, Linked Data, gamificación, Entity Linking

Abstract

This project aims to deploy a gamification platform to enrich and enhance cultural works, improving users' experience during visits to museums by providing them with more knowledge about the collection of exhibits through an enjoyable learning mechanism.

In order to achieve this objective, a text mining procedure, done on the description of the cultural objects , will yield relevant pieces of information which will subsequently be linked to DBPedia's linked data knowledge base using semantic technologies.

An automated question generation module will allow museum visitors to deepen their comprehension of the exhibits by means of gamification techniques to draw their attention. The aforementioned module will be created based on a semantic inference module, using SPARQL templates.

With the purpose of determining the smart object, QR codes will be printed which will be read by a mobile phone application. The designed app will not only display the game but will also show more details and complementary information about the smart objects, therefore creating smart visit experiences through which visitors can interact with the environment and the smart objects in it to acquire knowledge about the exhibits.

The system will be validated at the Telecommunications Museum at ETSIT and it can be used by both visitors and future students in open house days. It can also be extended to be adapted for any other museum.

Keywords: Semantic Web, Linked Data, Gamification, Entity Linking

Agradecimientos

Quisiera agradecer a mis padres, Alejandro y Raquel, todo lo que han hecho por mí, porque sin ellos no hubiera sido posible llegar hasta aquí. Gracias a mi hermano, Mario, por su positivismo constante que me ha alegrado y motivado en numerosas ocasiones.

También me gustaría dar las gracias a mi tutor, Carlos Ángel, y a los compañeros del Grupo de Sistemas Inteligentes por su ayuda y consejos para la realización del proyecto.

Muchas gracias.

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Context

Problem-based learning (PBL) is defined as an approach to a learning methodology based on the use of problems as the basis to acquire and consolidate knowledge on a particular subject [1].

Ever since its beginning over three decades ago [2], PBL has had a significant impact on education because people are given the chance to learn by solving problems. The aim of this approach is to help individuals primarily acquire flexible knowledge and intrinsic motivation [3].

Nowadays there is a general interest in the concept of gamification, which is defined as *"a process of enhancing services with (motivational) affordances in order to invoke gameful experiences and further behavioural outcomes"*[4].

Therefore, gamification and problem-based learning can turn out to be closely related if we desire to provide a flexible way of learning about a particular subject by developing a game containing questions about the topic of interest. This is the concept upon which our project is based.

Additionally, our focus is on designing and developing a multiple choice based game based on Semantic Web [5] technologies, because of the capability to potentially produce large, varied and entertaining question sets which can be automatically generated.

In order to achieve this goal, data from extensive Linked Open Data knowledge graphs such as DBpedia [6] has to be extracted. There are several projects that have successfully accomplished to produce question sets based on these structured data stores, culminating in a smartphone application to play quiz-like games[7][8].

Some museums around the world, like Prado Museum [9] have realised the significance and benefits of generating high quality linked data to publish their holding online. Having linked data allows museums to closely relate and categorise artworks with one another, their epoch, authors and other intricate facts. This results in a very accessible and flexible structure to allow users to browse through the online exhibits easily and intuitively. This is clearly an attempt to modernise themselves to enhance visitors' experiences.

Moreover, there have also been projects to promote and enrich museum visits by digitalising works and providing a tangible and embodied interaction with these. [10] This concept is known as *Smart Objects*.

Our idea was born from taking into account each and every consideration previously commented, because we are going to offer museum visitors a smartphone application where they can view the description of every object of interest and additionally they will be able to play a multiple choice game asking questions about them.

This means that our proposal offers a succulent value proposition to promote and enhance access to cultural heritage in museums, combining a set of techniques to produce Smart Objects and smart visiting experiences [11] in museums in a unique and original manner that could potentially revolutionise the way cultural heritage in museums is appreciated.

1.2 Project goals

The main goal of the project is to deploy a gamification platform to enrich and enhance the Telecommunications Museum "Joaquín Serna" at ETSIT UPM, improving user's experience during visits to the Museum by providing them with more knowledge about the collection of exhibits through and enjoyable learning mechanism.

In order to accomplish this, we need to set goals that must be achieved by the end of

the project. These are:

- Formation in semantic technologies to be able to process Linked Data and query Linked Open Data graphs such as DBPedia.
- Creation of ETSIT Museum's knowledge base, in order to obtain the objects' description from the Museum's online collection by means of web scraping technologies.
- Design and implementation of the module that can process descriptions, covering the detection and linking of entities.
- Design and implementation of the smart object question generation semantic system, based on SPARQL templates.
- Design and implementation of a REST API that hosts the information on Museum's objects and the generated questions, so that the mobile application can request this data.
- Design QR codes to access the smart objects.
- Design and implementation of the smartphone application of gamification, integrating the reading of QR codes.
- Evaluation of the system by deploying it at the Museum.

1.3 Structure of this document

In this section we provide a brief overview of the chapters included in this document. The structure is as follows:

Chapter 1 describes the state of the art of the context in which this project is developed. Additionaly, it provides a description of the goals set to accomplish during this project.

Chapter 2 outlines the principal technologies used to carry out the project.

Chapter 3 describes the architecture of the gamification system

Chapter 4 expands on the extraction of the objects' descriptions, creation of a knowledge graph and generation of the question templates based on SPARQL queries.

Chapter 5 describes the REST API implemented to retrieve information on the scanned smart objects by the mobile application.

Chapter 6 describes the design process of QR codes together with the design and implementation of the smartphone app of gamification.

CHAPTER 2

Enabling Technologies

In this chapter we will describe the technologies used to bring about the implementation of this project.

2.1 Python libraries

- **Srapy**

  Scrapy [12] is a collaborative open-source framework that extracts data from websites. It scrapes web data easily and efficiently. It has the following features:

  - **Quick and powerful:** the user has to write the set of rules to extract the desired data and Scrapy does the rest.
  - **Easily extensible:** given its configuration, new functionalities can be added without having to modify its source code.
  - **Portable and pythonic**: it is written in Python and it can run on Linux, Windows, Mac and BSD.

  Furthermore, with Scrapy offers several tools. These comprise:

  - Suport to extract and selec data from HTML/XML using CSS selectors by means of Xpath, with help functions to extract information using regular expressions.
  - An interactive iPython console to test CSS and Xpath expressions to extract data.
  - Possibility to export data registries into multiple formats, such as JSON, CSV or XML, amongst others.
  - Strong extensibility due to its ability to accept user-defined extensions and pipelines.

- **RDFLib**

  Python's RDFlib allows to create, manipulate, query and store RDF graphs. Its main structure is the *Graph* object. Definitions and facts representing a domain can be stored. These graphs are made up of triple patterns (subject/predicat/object) and these elements can either be nodes (an URI) or literals. A node can be created in several ways: by using a specific URI using **URIRef** class; by using blank nodes with the **BNode** class and by using literals (represented by the **Literal**) class.

- **SPARQLWrapper**

  We will use the *SPARQLWrapper* Python library that is a wrapper around SPARQL services. Therefore, we can produce queries to extract data from SPARQL endpoints - in our case, DBPedia- using Python scripts.

  SPARQL query templates will be used to generate our set of questions by extracting relevant additional entities related to ETSIT's Museum objects.

- **Pandas**

  During the data analisys stage we made use of the Pandas library. It is is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for Python. The main benefits of using Pandas is that provides ease for grouping, merging and querying Pandas data structures.

2.2 SPARQL

SPARQL Protocol and RDF Query Language is a specification proposed by W3C and provides a syntax to handle RDF graphs on the web and data endpoints.

RDF is based on triple patterns that define a universe that contain entities and properties

2.3 DBpedia

Dbpedia [13] is an open web repository containing structured data coming from Wikipedia and unlike Wikipedia, which consists of articles and documents, Dbpedia is a collection of structured data.

In its english database version there are more than 3,77 million entities described, amongst them are 764 thousand persons, 563 thousand places, 112 thosand music albums, 72 thousand movies and 18 thousand videogames.

Figure 2.1 represents the sequence of functions that take place when a query to DBPedia is made.

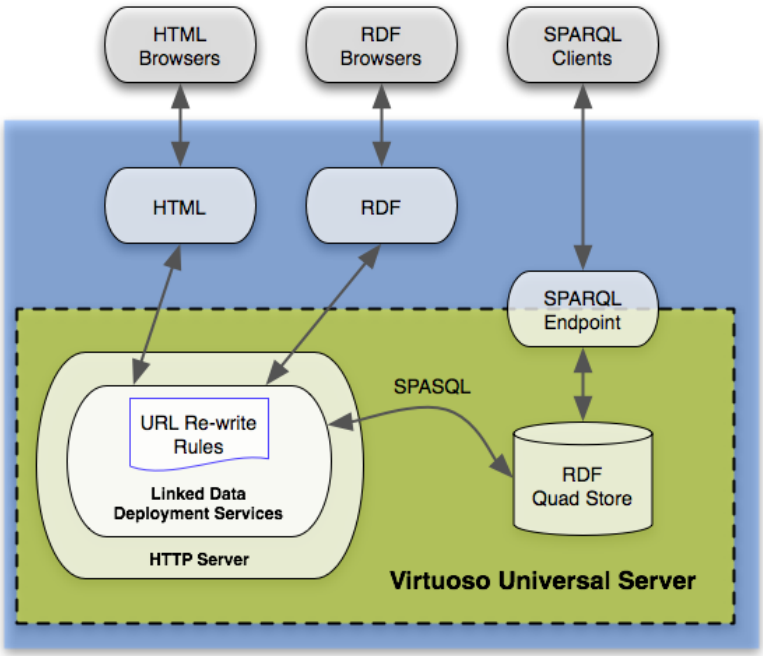

Figure 2.1: DBPedia Data Provision Architecture

We will use the entities gathered in DBPedia to annotate our objects with relevant data

that will allow us to generate automated questions.

2.4 Fuseki

Apache Jena Fuseki is a Semantic Web framework for Java that provides a store service for RDF data. It is a SPARQL server that we will use to store the obtained dataset of the scraped museum objects.

2.5 Babelfy

Babelfy is a software algorithm used for the disambiguation and recognition of text [14]. It performs two tasks:

1. Named Entity Recognition : seeks to locate and named entity mentions in unstructured text

2. Named Entity Disambiguation (Entity Linking) : clasifies the recognised entities into pre-defined categories such as the person names, organizations, locations, amongst others. This means that a decision must be made to classify this recognised entity. i.e. in the sentence "England is the best soccer team" the tool would need to identify that "England" refers to the English national soccer team but not specifically to the country. This therefore requires a knowledge base containing the entities to which entity mentions can be linked.

   Babelfy will be used to process the text description and to retrieve the URL linking to the corresponding matches entities so that we can extract more data about them.

2.6 Heroku

Heroku is a platform as a cloud computing service that supports many programming languages [15].

Dynos are the name given to Heroku's main artifacts of its architecture model. They are based on Linux Containers and they are units that provide computing capacity within this platform.

Heroku's main features are:

1. **Flexibility and growth:** the number of Dynos assigned to an application can be changed at any moment from Heroku's command line or the dashboard

2. **Size:** Heroku offers different types of dynos, each with its own processing and memory capacity.

3. **Routing:** Internally the routers track the location of the Dynos that are running and redirect traffic accordingly.

4. **Tracking:** There is Dyno orchestrator that continously monitors all running dynos.

5. **Redundancy:**Dynos are completely isolated from one another. This implies that in case there are faults in the internal infrastructure of a given dyno, the rest of them are not affected and thus the application is not affected.

   All of this features combined with the fact that there are free dynos makes Heroku an ideal platform of choice to deploy our web server.

CHAPTER 3

Smart Objects Gamification Platform

3.1 Introduction

In this chapter we will present the global architecture of the system developed. We will provide an outline of the project's architecture in order to offer the reader a general view of the purpose of each module and how they are interlinked with each other to provide our smart object gamification platform.

In the following chapters we will expand on these modules, thoroughly describing the artefacts comprising them.

3.2 Architecture

We will proceed to describe the architecture of the project, enumerating and overviewing the main modules that it consists of.

The system has been divided into three main components. These are:

1. **Data extraction and annotation system:** it is in charge of extracting the museum

objects' data from their webpage by means of web scraping technologies, then storing each item as RDF triples and afterwards complementing every object with additional information (i.e. annotate) exploiting the DBpedia Linked Data knowledge base . Finally, the annotated dataset will be used to feed a set of question templates to yield a batch of semi-automated generated questions stored ad JSON files.

2. **Web service system:** the server's main purpose is to store the questions and the museum objects' information that were generated in the data extraction and annotation stage in order to make them available to the gaming system. These will be saved into a NoSQL database, being MongoDB our choice. This data will be requested by the gaming system through a REST API in the Flask web application server.

3. **Gaming system:** it comprises two main artefacts, QR codes and a smartphone application. The user will scan the desired QR code about a particular object. The QR codes contain URIs to use the REST API and retrieve the information and questions on the object scanned. The user will therefore be presented with additional facts about the object of interest and there will be an option to play a random set of questions concerning it or its theme.

The combination of these subsystems is shown in Fig. 3.1, which portrays the global architecture of the platform.

In the following chapters we will proceed to provide a comprehensive description from the design phase to the implementation of the artefacts involved in each and every of the aforementioned subsystems.

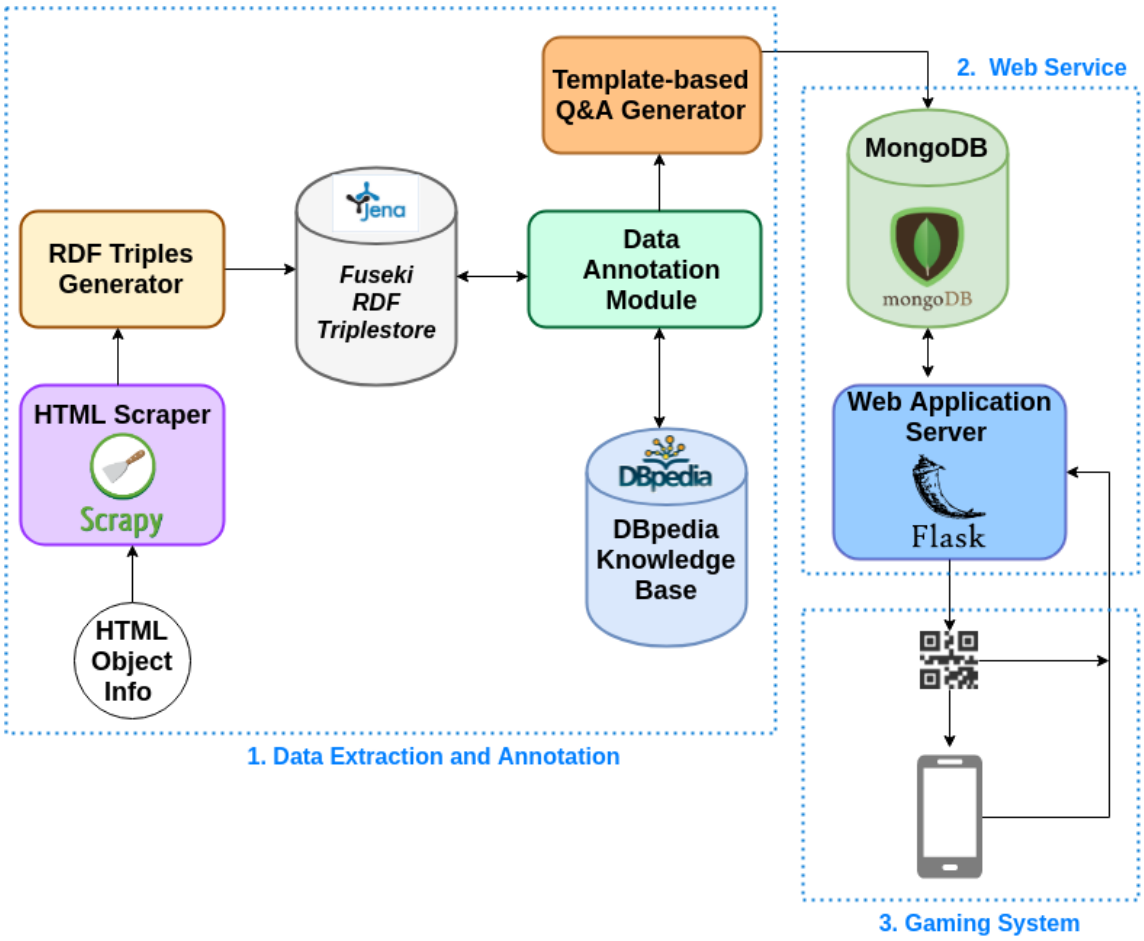

Figure 3.1: Architecture of the Smart Objects gamification platform

3.3 Web Service

We have used Heroku PaaS in order to make publicly available the REST API we have implemented.

This API grants access to the database where the data concerning our Museum objects and questions is stored. A CRUD pattern has been implemented in order to allow the smartphone application perform all necessary operations on the items of information stored in MongoDB.

The use of the API is completely transparent to the user, who will only interact with the smartphone application.

We have to distinguish between three different types of data found in our database, because every type has distinct uses. These are:

1. Data of Museum objects.
2. Questions classified into the 4 main categories.
3. User data.

Since the questions are only presented when a particular Museum object's QR code is scanned by the app(outlined in Section 3.2), this means that it is reasonable to deliver the information of the exhibit of interest and a selection of its associated questions in the same HTTP transaction.

Therefore, and because the idea of this gamification platform is to make users interact with the Museum's objects and learn more about them whilst enjoying themselves, the only way to be presented the information of an object of the Collection and thus questions, is to scan a QR code. The intricacies of the QR code schemes are explained later.

When a user scans a QR about a particular object, the following sequence of actions take place between the server and the app:

- The app requests the specified resource by communicating with the server through the URI **/objects/:PID** where **PID** is the unique identification number assigned by the Library's Management to each object belonging to the Telecommunications Museum.
- The server receives a GET request asking for a resource with a given *PID* (identifier), and executes a function to read data from the database (MongoDB) whose signature is *read(id)*.
- The database returns the JSON object whose identifier (PID) coincides with the one requested.
- Once the requested Museum object's data is obtained by the REST API, another function takes as input argument the category of the retrieved object. Its signature is *read_random_question(category)*. This function queries the database and returns three randomly chosen questions.
- The server prepares and return to the app a GET Response in JSON format by combining altogether the resources obtained in a JSON object.

Figure 3.2 depicts this sequence of actions clearly so that the reader may understand the communication between the app and the server (API plus database).

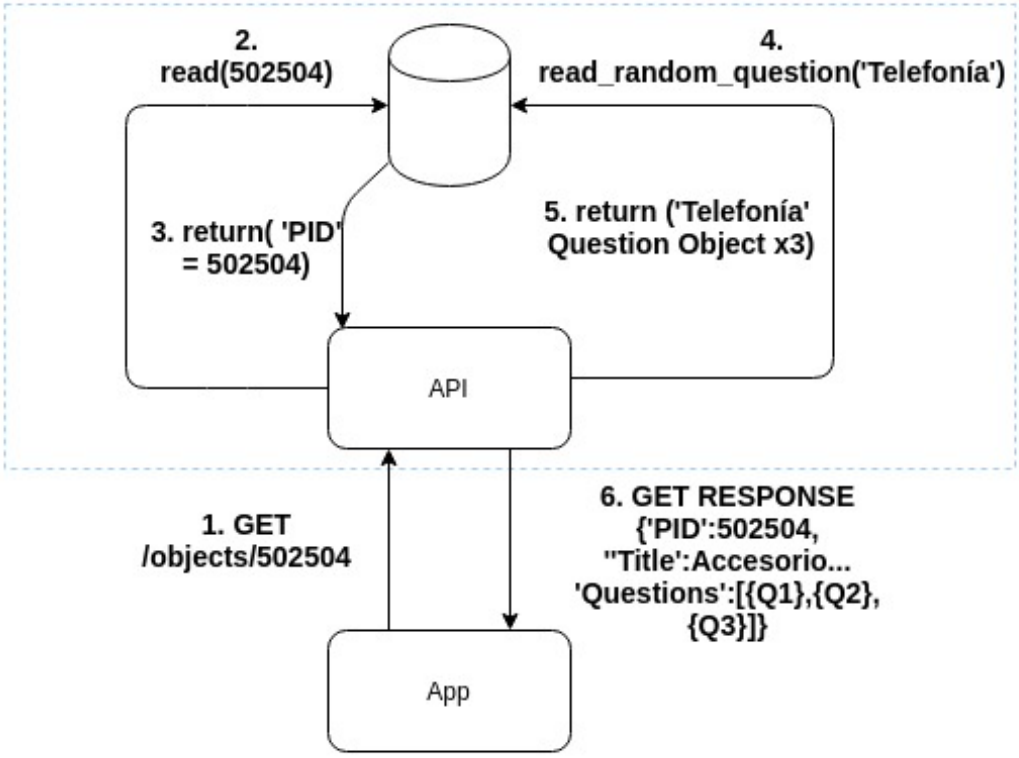

Figure 3.2: Diagram representing a typical sequence of communication between the app, API and database.

CHAPTER 4

Data Extraction and Annotation

4.1 Introduction

In this chapter we will elaborate on the data extraction process and object annotation stages that result in the conception of the automated questions. First, we will describe the data obtention from the online digital collection of the telecommunications museum's objects. Afterwards, we will comment on the process of storing these objects as RDF triples using well known ontologies. Then, we will depict the approach to the gathering of relevant items of information about the museums' objects from DBPedia and telecommunication glossaries in order to produce a varied dataset that will be used by the question generator. Therefore, and finally, the template-based question and answer generator scheme will be described.

4.2 Telecommunication Objects Data Obtention

The items belonging to the collection of the ETSIT Museum are catalogued in an open platform of digital documents named UPM Digital Collection ("Colección Digital Politécnica)[1]. The data regarding these objects has been extracted from this online repository. Figure **??**

[1] http://cdp.upm.es

depicts an entry of one of the Museum's objects and shows its properties and the description property we will process later on.

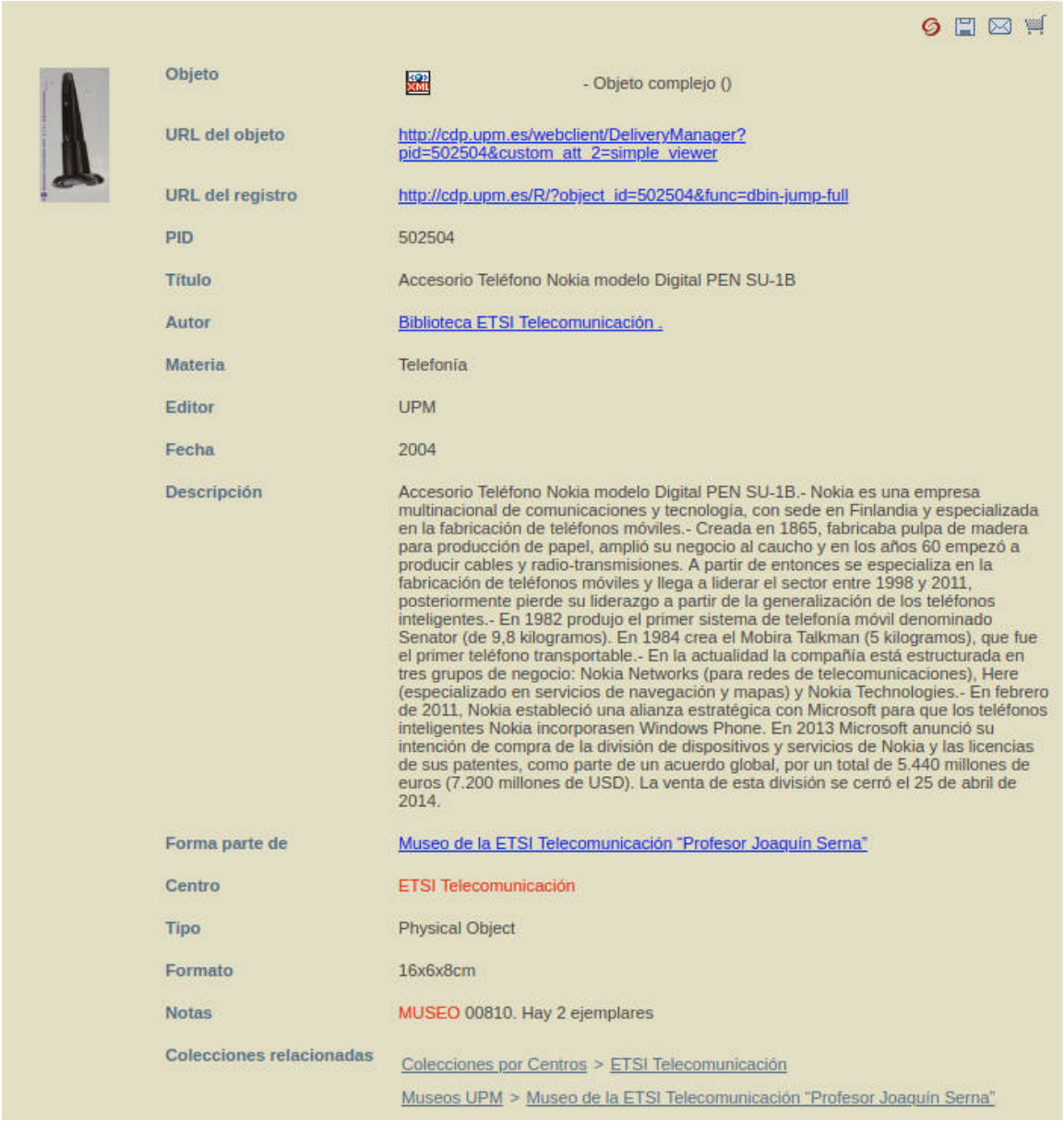

Figure 4.1: Example of an object as displayed in ETSIT's Museum website

Every object's information entry follows the same structural pattern with the same properties to classify them. Hence, this allows us to easily extract the desired pieces of information in an automated manner, by preparing a Python script using the Scrapy web crawler framework. The data is afterwards stored in a JSON file. Amongst these properties we only extracted the most relevant ones for our purpose.

It is important to comment the utility of the PID property shown in Table 4.1, since it

is a unique six digit number which will allow us to univocally identify the object and build a URI to request the desired resource through our REST API.

Property	**Comment**
PID (PID)	Identification number used by UPM Digital Collection.
Title (Título)	Name of the object.
Date (Fecha)	Date of manufacture/creation.
Type (Tipo)	Type of object, they have been classified as a "Physical Object".
Subject (Materia)	Category to which the object belongs.
Centre (Centro)	Location of the object (ETSIT's Museum of Telecommunication "Professor Joaquín Serna")
Description (Descripción)	Information describing the history of the object, its manufacture, uses and/or its constituent parts.

Table 4.1: Summary of the main properties that were retrieved for each object and a comment explaining the meaning of each one.

4.2.1 Item Categorisation

Initially, this stage was not meant to be necessary because we assumed that the *Subject* property correctly classified the museum objects into the most suitable category(subject) of telecommunications. However, during the first meeting with the directress of ETSIT's library, we discovered that the object database was being dumped into the School's Library blog[2] in order to improve the readability of the information , since the Digital Collection's website aforementioned proved to be more difficult to navigate through and with a more complex user interface.

Consequently, the Library is actively updating the Museum's objects in the blog whilst the Museum's Digital Collection is lately being left slightly unattended. This has led us to identify a number of inconsistencies in the Digital Catalogue with respect to the updated blog:

- The Digital Collection's subjects of classification tend to differ from the subject(s) assigned to the same object in the blog. For instance, there is a device named *Direct Current Analog Voltmeter* which is sorted into two subjects according to the Digital

[2] `http://biblog.etsit.upm.es/`

Collection[3]: *Voltmeters* and *Wattmeter*. However, the blog[4] classifies this same object into the *Instrumentation* category.

- The Digital Collection also has **11** objects which are mistakenly duplicated. Additionally, there are **5** objects in the blog whose URLs to their pictures in the image gallery of the Digital Collection are not correct.

Regarding the first problem identified, whereas the Collection has a grand total of **75** subjects (the majority tend to simply state again the type of telecommunications object as we have already illustrated with an example), the blog sorts objects into **6** possible main categories. This is the preferred way by the Library's management of classifying these objects and it brings us simplicity towards creating the Quiz part of the platform, where players will need to master six broad categories instead of seventy-five.

In relation to the repetition and broken URL problem, we provided the Library's directress a list of these conflictive objects so that they could be corrected.

As a consequence of this, the original file with the information on the Museum's objects extracted from the Collection was modified accordingly. This resulted in **873** objects processed and sorted just like they are classified on the blog. Note that an object can be sorted into more than one subject. The categories are illustrated in Table 4.2.

	Sound	**Image**	**Telephony**	**Telegraphy**	**Radio**	**Instrumentation**
# of matches	62	10	596	66	78	76

Table 4.2: Breakdown of objects into their categories.

The final result of the object extraction and categorisation procedures can be appreciated by the example provided in Listing 4.1.

[3] `http://cdp.upm.es/R/?object_id=486143&func=dbin-jump-full`
[4] `http://biblog.etsit.upm.es/?p=11020`

Listing 4.1: JSON object after its extraction and categorisation


```
"PID": "502504",
"Titulo": "Accesorio Teléfono Nokia modelo Digital PEN SU-1B",
"Materia": [
"Telefonía"
],
"Fecha": "2004",
"Descripcion": "Accesorio Teléfono Nokia modelo Digital PEN SU-1B.- Nokia es
    una empresa multinacional de comunicaciones y tecnología, con sede en
   Finlandia y especializada en la fabricación de teléfonos móviles.-
   Creada en 1865, fabricaba pulpa de madera para producción de papel,
   amplió su negocio al caucho y en los años 60 empezó a producir cables y
   radio-transmisiones. A partir de entonces se especializa en la fabricaci
   ón de teléfonos móviles y llega a liderar el sector entre 1998 y 2011,
   posteriormente pierde su liderazgo a partir de la generalización de los
   teléfonos inteligentes.- En 1982 produjo el primer sistema de telefonía
   móvil denominado Senator (de 9,8 kilogramos). En 1984 crea el Mobira
   Talkman (5 kilogramos), que fue el primer teléfono transportable.- En la
    actualidad la compañía está estructurada en tres grupos de negocio
   Nokia Networks (para redes de telecomunicaciones), Here (especializado
   en servicios de navegación y mapas) y Nokia Technologies.- En febrero de
    2011, Nokia estableció una alianza estratégica con Microsoft para que
   los teléfonos inteligentes Nokia incorporasen Windows Phone. En 2013
   Microsoft anunció su intención de compra de la división de dispositivos
   y servicios de Nokia y las licencias de sus patentes, como parte de un
   acuerdo global, por un total de 5.440 millones de euros (7.200 millones
   de USD). La venta de esta división se cerró el 25 de abril de 2014."
"Centro": "ETSI Telecomunicación",
"Tipo": "Physical Object",
"Formato": "16x6x8cm",
"Notas": "00810. Hay 2 ejemplares",
"UrlObjeto": "http://cdp.upm.es/R/?object_id=502504&func=dbin-jump-full",
"UrlGaleria": "http://cdp.upm.es:80/webclient/DeliveryManager?pid=502504&
   custom_att_2=direct",
"file_urls": [
"http://cdp.upm.es/webclient/StreamGate?folder_id=200&dvs=1540034752354~565"
],
"files": [
"MUSEO_502504.jpeg"
```

4.3 Resource Description Framework Graph Generation

After the data extraction and categorisation process, the resulting JSON file mentioned before is used as input for a script that takes such file and generates a Resource Description Framework (RDF) graph using Turtle syntax.

The purpose of generating such graph is primarily that it will be useful for us when preparing the dataset to feed the template-based question generator since we can rapidly query properties of these telecommunications objects as well finding relationships between them (i.e. obtain the objects that were created before/after or even on a specific date).

Another notable aspect to take into account is that, although the generated graph is simple in terms of the lack of extensive and deep relationships between objects, it can serve as the first step towards converting ETSIT's Telecommunications Museum into an *Open Museum*[16], offering a much more user-oriented approach towards data presentation of the collections available, whereby visitors can benefit from more structured and well-connected information of the museum's exhibits in digital format through Linked Data.

However, the creation of such comprehensive graph is beyond the scope of this project since it would require a considerable amount of time and effort to yield a decent result. Nevertheless, we will offer our help to the Library's management in case this task was to be undertaken in the future.

Concerning the graph that we were able to generate, each object is identified by 6 predicates, enough for our purpose. We used CIDOC-CRM and Dublin Core ontologies to structure our objects. Listing 4.2 depicts an object from the Museum Objects' RDF graph in order to show how some of the properties named in Section 4.2 have been used to produce a graph of exhibits. Note that, as we defined in the forenamed section, all the objects are of type *"Physical Object"*, which happens to be an entity class of CIDOC-CRM ontology.

The resulting graph has been uploaded to an endpoint provided at GSI servers and it is at the reader's disposal[5].

[5] `http://sefarad-fuseki.cluster.gsi.dit.upm.es/museoETSIT/query`

Listing 4.2: Excerpt from the Museum Objects' RDF Graph

```
<http://cdp.upm.es/R/?object_id=510058&func=dbin-jump-full> a crm:E19
    _Physical_Object ;
    crm:P1_is_identified_by "Magnetófono de bobinas Tesla" ;
    crm:P54_has_current_permanent_location "Museo de la ETSI
    Telecomunicación 'Profesor Joaquín Serna'" ;
    dc:date 1963 ;
    dc:description "Magnetófono de bobinas Tesla.- Marca TESLA.- Modelo
        SONET B3.- Grabadora de cinta.Mono, 4 pistas, 2 velocidades de
        grabación.- Alimentación 220V c.a.- Potencia de salida 1,5 W (2.5 W
        max.).-Fabricado en Checoslovaquia" ;
    dc:subject "Sonido" .
```

4.4 Object Annotation

We will now proceed to describe the object annotation stage.

4.4.1 Annotation Problems

In order to extract additional information related to every museum's item or the category they are sorted into, we rely on the use of DBPedia as our Linked Data knowledge graph. This statement is of great significance because the quality, abundance and detail of the information existent on DBPedia about a particular entity or subject will directly impact on the quality and relevance of the questions that can be generated.

For instance, let us consider for a moment that the museum we were working with is the Prado Museum. An artwork such as *The Clothed Maja (La Maja Vestida)*, one of Goya's most popular pieces, has its own entry in DBPedia[6]. This means that if we were to produce questions related to this painting, it would be done with ease and it would most presumably yield potential data for questions with a reasonable level of intricacy.

However, we can draw the conclusion that the Telecommunications Museum at ETSIT is not regarded as a renowned museum based on the fact that it has no Wikipedia dedicated article and the main websites that refer to it are from the university's web domain. Consequently, the Museum's Collection data is also not available in Wikipedia nor DBPedia.

[6] http://dbpedia.org/page/La_maja_vestida

Therefore, finding additional information on any of these objects other than the provided by UPM's Digital Collection is very difficult. Taking up on the previous example, the *Direct Current Analog Voltmeter*, there is no specific DBPedia entry for it. This at first may not seem to be a problem as one would think that what is more important is to know about voltmeters in general. Indeed, we could extract data about voltmeters from the Voltmeter entity [7] in DBPedia.

Unfortunately, this entry is not very descriptive, because, for example, one could think of asking about the difference between the types of available voltmeters as a possible legitimate question. However, they are referenced only as a redirect link to the voltmeter entity, so it would not be possible to generate this question. Similarly, there is for instance no information about the inventor of this device.

Overall, we are left with scarce information that we can automatically extract about details on any of the objects in general, as we wanted to illustrate this problem with an example, but the reality is that this lack of information to work with is recurrent for almost every item.

Nevertheless, we must mention that the "rdfs:comment" DBPedia property of items of interest tend to be detailed and if we could automatically extract data from those texts we could yield relevant data to generate good questions. What is to be highlighted is that although these descriptions are usually detailed, there are little and not very useful properties of these entities (which is what we can query and automatically process).

Altogether, from this analysis performed we can conclude that the information on the subject of telecommunications available on DBPedia is generally not well described. Although a query of entities directly related to the subject "dbc:Telecommunications" yields 2644 results and 10000+ in the case of a search with graph depth[8] of 2, these entities are mainly not described in the level of detail we expect in order to produce a batch of pertinent questions.

This led us to consider training a machine learning model to recognise and extract relevant information from text descriptions, not only the ones provided by DBPedia, but also from the ones extracted from the Museum's Digital Collection. We will comment more on the latter later. Building such model would greatly overextend the project and would not allow us to finish it on time, although this idea is very appealing and it could seem to be worth implementing it if the project were to be continued in the future.

[7] `http://dbpedia.org/page/Voltmeter`

[8] `https://www.w3.org/TR/sparql11-property-paths/`

Additionally, we had the intention of using Named Entity Recognition and Named Entity Disambiguation Tools (Refered in section XX on the description of the objects provided by the Digital Collection. After manually analising various sets of 50 of these objects and identifying which of the recognised entities could be of the greatest relevance towards the generation of object-specific automated questions, we realised that overall the descriptions provided are also not very detailed, hence seriously limiting the entities identified by the Entity Linking tools that could be regarded as potentially valid to generate questions.

The descriptions could all preferably mention for instance the components or materials used to manufacture the invention, its uses, or the manufacturers amongst others. Very few objects satisfy all of these desired qualities and therefore it was also difficult to extract additional information in the form of entities linking to their DBPedia entry.

Of course, the aforementioned fact that DBPedia's information on the subject of telecommunications lack the level of desired detail, together with the absence of dedicated DBPedia entries for exactly our objects, makes us resort to other approaches in order to obtain data to annotate our objects and produce interesting questions.

Our first possibility was to expand on the RDF graph we obtained (See Section 4.3). However, this implied having to include rather meticulous details in the graph, establishing relationships between entities that depend on each other, adding many more properties to objects,etc. The cost of this solution, time wise, would exceed the foreseen deadline of the project. This basically implies that we were going to transform the Telecommunications Museum's data into Linked Open Data(not a simple graph like we have generated), which by itself may acquire the dimensions of a whole new project.

4.4.2 Strategies Devised to Annotate Objects

Based on the analysis made, we introduce the different strategies that were ultimately devised to retrieve additional data related to the objects. We will afterwards expand on them :

- **NER and NED tools:** they are used to mine from the description of each object, if available, entities that represent **companies** and **countries/cities of manufacture**. Mentions to people(inventors of the object primarily) have been disregarded because there is only an 8% of the objects that mention them in their descriptions (7 out of 873), thus considered insignificant.
- **SPARQL queries to DBPedia based on the category of the object:** taking

into consideration the category sorting process explained in Section 4.4.1, we have decided to generate queries based on each of these 6 subjects. This process is further detailed in Section 4.4.2.2.

- **Glossary of telecommunications:** a comprehensive glossary of telecommunications has been used to obtain definitions of terms on this area. It contains many entries that do not appear in DBPedia, or if they do, their DBPedia definition is not as good as the one provided by this glossary.

Retrieving information about companies and locations using NER and NED tools will yield a very small dataset compared to those generated through SPARQL queries and the telecommunications' glossary. Nevertheless, it is worth mentioning this strategy because it was initially the principal strategy devised to annotate data, but we have shown that given the lack of information on DBPedia about our Museum's Objects, this task turns out to be rather unfruitful. However, if this were to be applied to any renowned museum with DBPedia articles, the results obtained could be quite interesting.

4.4.2.1 NER and NED. Babelfy.

As mentioned in Section in the introduction, we have used Babelfy's Named Entity Recognition and Disambiguation API because it can process, unlike DBPedia Spotlight, text in Spanish. Table 4.3 shows the evaluation metrics for the companies recognised amongst the text descriptions.

	expected matches	**matches**	**total objects**	**match percent**
Babelfy	620	549	873	62,8%

	F-Score	**Precision**	**Recall**
Babelfy	0.88	0.87	0.88

Table 4.3: Evaluation Metrics for "Organisation" type entity recognition and links

4.4.2.2 Pre-confectioned SPARQL queries

We manually analysed the entities returned under queries about our 6 categories of interest. Let us remind the reader that these were : **Sound, Image, Telegraphy, Telephony,**

Radio and Instrumentation.

It is important to highlight that, under the circumstances in which we have presented our problems, generating intricate questions about a particular object is not possible. Let us illustrate this better: if for instance, our object is a voltmeter, we have seen that we cannot ask about neither its inventor nor types of voltmeters nor anything else specific to it. However, we can query about the different measuring instruments used in telecommunications, and we could therefore ask for example questions about which of the given instruments best matches the definition provided.

Listing 4.3: Example of SPARQL query to obtain data on the category of "Instrumentation"

```
%INSTRUMENTS RELATED TO TELECOMMUNICATIONS%
SELECT distinct ?uri ?label ?comment ?img WHERE {
  {?uri dct:subject ?s .
  ?s skos:broader{,1} dbc:Measuring_instruments.
  ?uri rdfs:label ?label.
  ?uri rdfs:comment ?comment.
  ?uri dbo:thumbnail ?img.
  ?uri dct:subject/skos:broader{,4} dbc:Telecommunications.
  filter( regex(str(?uri),"meter|analyzer|scope|radar","i" ) && lang(?label)
       = 'es' && lang(?comment) = 'es')}
   UNION
  {?uri dct:subject dbc:Electrical_meters.
  ?uri rdfs:label ?label.
  ?uri rdfs:comment ?comment.
  ?uri dbo:thumbnail ?img.
  filter(lang(?label) = 'es' && lang(?comment) = 'es')}
}
------------------------------------------------------------------------------
%INSTRUMENTS NOT/LESS RELATED TO TELECOMMUNICATIONS%
SELECT distinct ?uri ?label ?comment ?img WHERE {
  {?uri dct:subject ?s .
  ?s skos:broader{,1} dbc:Measuring_instruments.
  ?uri rdfs:label ?label.
  ?uri rdfs:comment ?comment.
  ?uri dbo:thumbnail ?img.
  FILTER NOT EXISTS {?uri dct:subject/skos:broader{,4} dbc:
      Telecommunications}
  FILTER NOT EXISTS {?uri dct:subject/skos:broader{,4} dbc:Electricity}
  FILTER NOT EXISTS {?uri dct:subject/skos:broader{,4} dbc:Chemistry}
  filter( regex(str(?uri),"meter","i" ) && lang(?label) = 'es' && lang(?
```

```
        comment) = 'es')}
}
```

This means that our question generation is focused on the category the object belongs to, as we have commented. Under these main 6 categories we can specify other subcategories where we can generate interesting and varied questions.

Refering back to Section , we stated that the percentage of Museum object's descriptions mentioning inventors was negligible. We also commented on the fact that the entities related to telecommunications in DBPedia are not well described and that they lack information on, for example, their inventors or scientists/engineers related to the item. However, we thought that including data on these people could add flair to the generated questions, so instead we searched in DBPedia for telecommunication and IT related personalities and we were able to extract an image of them and the list of awards they have been given (if any).

With these pieces of information we can perform questions such as *"Who is the person in the image"?("¿Quién es la persona de la foto?")*, or *"Which of the following awards has been given to Alan Turing?"("¿Cuál de los siguientes galardonas ha sido obtenido por Alan Turing?")*.

Listing 4.3 provides an example of the generation of two datasets related to the **Instrumentation** category by querying DBPedia for measuring instruments related to telecommunications and for those less or not related to telecommunications. Let us draw attention to the properties extracted from these queries, Table 4.4

VARIABLE NAME	**?label**	**?comment**	**?img**
EXPLANATION	Name of the instrument entity	Definition provided by DBPedia	Image of the instrument

Table 4.4: Example of properties extracted from DBPedia about entities belonging to the "Instrumentation" category

As a result, it springs us to mind a set of legitimate questions that could be asked(we will explain this in more detail in the question template section), such as:

- "Which of the following instruments is related to telecommunications? (¿"Cuál de los siguientes instrumentos está relacionado con las telecomunications"?)
- "Which of the following instruments is NOT related to telecommunications"?" ("¿Cuál de los siguientes instrumentos NO está relacionado con las telecomunicaciones?")

- "Which of the following instruments is related to telecommunications?(¿"Cuál de los siguientes instrumentos está relacionado con las telecomunications"?)
- "Which instrument suits this definition?" ("¿A qué instrumento corresponde esta definición?")
- "Which instrument is portrayed in the image"? ("¿A qué instrumento corresponde esta imagen?")

4.4.2.3 Telecommunications Glossary

After an exhaustive search for online glossaries of telecommunications, we found a very detailed and well structured online telecommunications dictionary [17] created by Universitat Politècnica de Catalunya -UPC (Polytechnic University of Catalonia) - in conjunction with Enciclopèdia Catalana [18] (Catalan Encyclopedia).

This dictionary [9] has about 4200 telecommunications-specific terms and their corresponding definitions. One drawback of using this glossary is that, although there is a Spanish version, only the name of the terms appear in Spanish, whilst the definitions are typed in Catalan.

Since it was by far the most extensive and detailed online dictionary we could find, it was worth translating these definitions into Spanish. This was not a difficult task, however it was time-consuming since we had to manually revise the translations made by Google from Catalan into Spanish and therefore it is worth mentioning the stages involved in the retrieval and processing of this potent glossary, as depicted in Figure 4.2.

It is to be noted that there is a better dictionary of telecommunications provided by the International Telecommunication Union [19], unfortunately there is no online version and it is only presented in document format.

The preference of searching for an online telecommunications glossary relies on the fact that it is notably easier to extract them through the use of web crawlers, as HTML is a strongly typed language and therefore it is easy to automate extraction of data, unlike an electronic document where the task can be considerably more arduous. This is the reason why ITU's glossary was diregarded.

[9] https://www.termcat.cat/es/diccionaris-en-linia/235/es

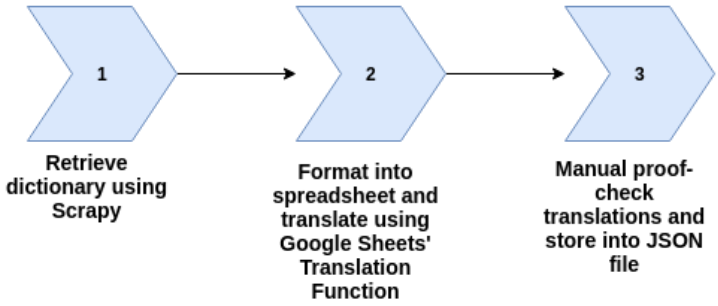

Figure 4.2: Stages of extraction, translation and refination of the dictionary of telecommunications

4.4.2.4 (Annotation Data)

As a result of the methods, techniques and strategies employed to obtain relevant data to annotate our Museum's objects, we will summarise here the data obtained, classified by category.

During the process we also came to the realisation that data to annotate our Museum's objects was for some categories scarce or it was difficult to draw a line between entities of some categories. Therefore, the course of action we took was to encapsulate some categories together for the data used to feed the question templates.

Consequently, the question template has been designed to process 4 main areas:

1. **Sound and Image**
2. **Radio and Telepraphy**
3. **Instrumentation**
4. **Telephony**

For example, concerning *Telephony* category, we extracted, amongst others, additional data about international country codes; definitions about telephone and telephone networks and specifications of smartphones.

Figure 4.3 represents the amount of data items we could extract using the glossary of telecommunications and the prepared SPARQL queries based on our initial 6 main subjects. The information of inventors is not included in the figure because it can be considered a wildcard in the sense that a question about inventors can fit in any of the chosen categories. Similarly, data about companies is not presented here because it has been appended as a

property of Museum items, i.e. a key/value pair in the corresponding JSON object has been added, with the key named *"Company"*.

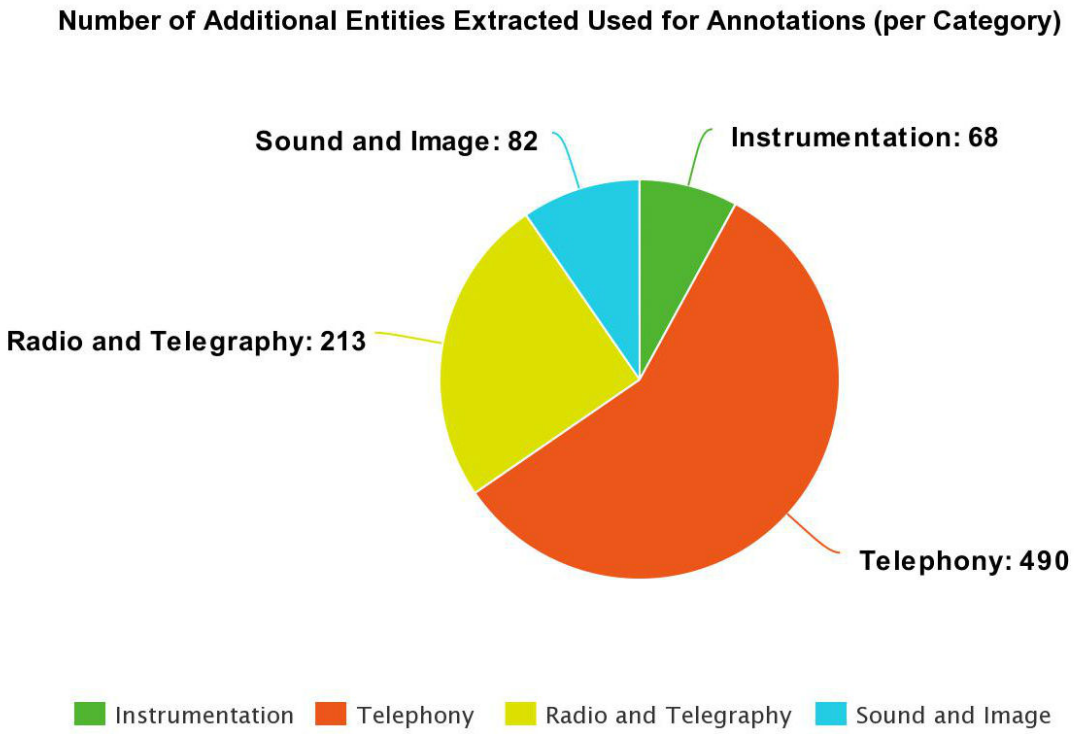

Figure 4.3: Number of additional entities extracted used for annotations (per category)

4.5 Template-based Question Generator

Once we solved the problems about the obtention of relevant additional data to complement the existing information on the Museum's objects and thus generate questions based on them, we could proceed to finally generate batches of questions for each of the four core categories.

Our main purpose, as we commented in the introduction of the document is to achieve the gamification of the Telecommunications Museum's objects by providing an enjoyable learning mechanism in the form of a mobile application with multiple choice questions.

The structure of a multiple choice question consists in the following three items[20]:

- The question to formulate, called the **stem**.
- The correct answer: the **key**.
- A set of incorrect but still plausible answers : the **distractors**.

4.5.0.1 Question Templates

The question generator works by taking as input lists of the data we have extracted in the data annotation stage and the set of templates we have devised.

A question template is just a JSON object with a list of key and value pairs.

In order to achieve the semi-automated generation of questions, we will summarise the tasks this module has to perform in order to successfully produce decent questions:

1. Feed the annotated data to the question generator.

2. Produce with this data and the prepared templates questions based on the 4 main categories : **"Sound and Image", "Telephony","Instrumentation" and "Radio and Telegraphy")** . Table 4.5 shows a sample of *Instrumentation* templates and a sample of *Telephony* templates.

3. Additionally generate questions with templates that are generic, i.e. questions that can be sorted into any of the categories. These include questions about scientists/engineers related to the field telecommunications as well as questions about the company that manufactured our object from the Telecommunications Museum.

4. Finally, after the questions have been generated, they are sorted according to their difficulty. In order to do so, we make use of DBpedia's popularity score, DBPedia Pagerank [21], together with Sematch [22] (Refer to Section **SECTION!**), to obtain an estimate of the popularity and semantic similarity between the key and the distractors. The score as a percentage, indicates the difficulty of the question if the score is low, and the higher the score, the easier the question, in theory.

We have to note that the questions are randomly generated. The question templates randomly choose 4 entities : 1 key and 3 distractors. The reasonable amount of extracted entities to feed the template allows us to generate many variations of the same question template.

This is, although there are not many question types (primarily because we could not think of other templates where questions could be interesting and make sense at the same time, due to the lack of data as we have commented in earlier sections.), the amount of entities for each question type is large enough in order to produce varied and entertaining questions.

Therefore, the amount of questions randomly produced for each template can be arbitrarily high, up to a given point where too many yields would result in repeated questions (i.e. the exact same 4 possible answers with the exact same key) and also questions with very similar possible answers.

Through testing and experimentation we have decided, as a compromise, to generate a number of questions per category proportional to the amount of data used.

Template	Details
Instrumentation	Match item with the definition shown *Stem: Which instrument matches the definition shown?*
	Match image to instrument *Stem: Which instrument is depicted in the image?*
	Identify the instrument related to telecommunications *Stem: Which of the following instruments is the most related to telecommunications?*
	Identify the instrument unrelated to telecommunications *Stem: Which of the following instruments is the least related to telecommunications?*
Telephony	Ask about the HW specifications of a smartphone (RAM, battery or storage) *Stem: Which of the following cellphones has 4GB of RAM ?*
	Ask about a comparison of min./max. value of HW specs of smartphones *Stem: Which of the following smartphones has the lowest battery capacity?*
	Match definition with the correct answer *Stem: which telephony term best suits the definition given?*
	Match a international country code to its country *Stem: which country has +34 as international country code?*
	Match a given country to its international country code? *Stem: what is the international country code of Mexico?*

Table 4.5: Sample of supported question templates for Instrumentation and Telephony categories with stem examples.

CHAPTER 5

Gaming System

5.1 Introduction

In this chapter we are going to describe the gaming subsystem of the platform. This is the system the user directly interacts with. As we have mentioned throughout the document, the user will use the smartphone app to scan QR codes which grant access to the information of a Museum object plus the right to play some questions concerning the object.

The process of the extraction of Museum object's information and the generation of questions has been addressed in previous sections (See Section 4).

Firstly, we will expand on the QR codes because it is a component of the uttermost importance towards achieving the goal of creating the desired gamification platform.

Secondly, and finally, we will describe how does the application work. This application is the culmination of our project and throughout this point we will appreciate how all artefacts involved in the whole system work together to produce a gamified environment of the Telecommunications Museum's object, therefore achieving to turn these ordinary exhibits into Smart Objects.

5.2 QR Codes

Our initial idea was to generate a batch of simple QR codes, one for each Museum object, which would link to the REST API to request the information on that specific exhibit and play random questions as it has already been thoroughly explained throughout Section 4.

The QR codes initially pointed to the HTTP GET primitive , /objects/:PID , uniquely identifying each Museum object.

A script was programmed to generate the URI to the resource identifying the object, and then converted into a .csv file. This allowed to use Mail Merge functions from Microsoft Word to easily generate in **batch 873 QR codes**. The first version can be seen in Figure 5.1.

Figure 5.1: Example of the initial aspect of the QR codes

These QR codes would be cut out and pasted on the showcases of the Museum's object. This posed a problem because there are, more often than not , many objects per showcase, which would mean that the objects would be barely visible due to the QR blocking the view.

Moreover, in order to be granted permission to paste these QR codes, we had to explain to the Library's directress, Reyes Albo, the project we were undertaking and its implications for the Museum.

We were granted permission to generate QR codes, however, not the way we had anticipated, and this meant a rather important amount of work for us. There were several considerations that needed to be taken into account:

- They already had in mind generating QR codes for every exhibit, but these QR codes

would neatly be formatted and printed on hard cardboards.

- The QR codes were going to be linked to the corresponding URL of the Museum object located at the Library's Blog (we mentioned in Section 4)

This meant that we had to modify the project in order to meet the directress' requirements. As a result:

- **Ambivalent QR codes have been generated:** this means that if a QR code is scanned with our application, we will be able to view its information in the app and play quizzes. If they are scanned with any simple barcode reader, the QR code will take the user to the entry of that object on the blog.
- **QR codes have been formatted, coloured and printed into hard carboards:** we designed the appearance of the printed hard cardboard, satisfying the directress' needs. The aim is to have a space where to place these hard cardboards, and users interested in learning more about any Museum object can take these cardboards around the School, look for the objects printed on the cardboards and scan QR codes to view more details on the object and play quizzes.

We shall note that the generation of these QR codes in this format was deceiving because there was no rapid way of producing a batch of these cardboards to generate all required items at once. Microsoft Word did not respond well to this extra formatting and the Mailing options did not reproduce qr codes for every entry in our source file.

Therefore, we had to implement a solution based on a Microsoft Word macro using VBA scripts. Figure 5.2 portrays a prototype of the aforementioned hard cardboards, which has been used to test the app and the interaction with the Smart Object environment.

Figure 5.2: Prototype of a hard cardboard with objects belonging to the category of Telegraphy.

5.3 The Application. SmartExhibits ETSIT-UPM

As we commented in Section **ENABLINGTECHH**, the Application has been developed in JavaScript using React Native framework.

The GSI is owner of a Google Play account where they upload their own apps and thus *SmartExhibits ETSIT-UPM* has been uploaded to Play Store. Therefore, although the app could potentially work for iOS there is no guarantee it will, since the design, implementation and test stages have been oriented towards Android devices.

We will explain now the typical user interaction sequence with the application:

Firstly, when the user initialises the app, they are presented with an introductory screen where there is a camera icon. If this icon is pressed, it will pop up the next screen, which opens our QR scanner.

The scanner will not react to any code other than our QR format. Once a valid QR code is scanned, a screen with three tabs pops up.

From this moment, the user has access to an elegant presentation of the scanned object's information and to three randomly generated questions regarding the category the object belongs to.

After scanning the object, the user will initially be presented with the first of the three tabs, showing the name of the scanned object and its image.

The middle tab provides a list summarising the most interesting properties provided by the Digital Collection's descriptions, where the user can learn facts about this particular item.

The third tab grants access to the quiz game, the user must press the play button placed in the centre of this tab and three random questions will appear, one after another.

After the three questions have been answered, regardless of whether they were correct or incorrect, the user is brought again to the three-tabbed screen.

5.3.1 Gamification Rules

The game allows registered users to be able to ascend in a ranking system we created. At first, questions are easier, and as users obtain more points when answering questions correctly, they tend to become harder.

A user will work their way up the ranks of a "Telecommunications Engineer", this is, there will be 4 stages of expertise in the game. They will earn a badge when they have acquired enough points to ascend to the next level of difficulty. The badges/ranks are:

1. **Beginner**
2. **Intermediate**
3. **Advanced**
4. **Expert**

Additionally, there is a time limit to answer every question. This time is reduced when users ascend up the ranks of expertise.

CHAPTER 6

Conclusions and future work

In this chapter we will describe the conclusions drawn from this project.

6.1 Conclusions and achieved goals

Throughout the completion of the project we have achieved the goals that we initially set for the project:

The main objective of this project was to create a gamification platform in order to enrich and enhance the Telecommunications Museum "Joaquín Serna" at ETSIT-UPM, improving user's experience during visits to the Museum by providing them with more knowledge about the exhibits through an enjoyable learning mechanism.

We have successfully managed to create the gamification platform with over 1000 possible questions and ranging levels of difficulty.

The first step to approach the implementation of the platform was to extract the description of the museum's objects from ETSIT's Digital Colletion using Scrapy, a web crawler.

Once we had the extracted descriptions we proceeded to sort objects into categories based on six different subjects: *Telephony, Radio, Telegraphy, Instrumentation, Image* and

Sound, a task that was initially not foreseen but had to be done because ETSIT's Digital Collection was not sorting objects correctly into sensible categories.

Afterwards, we made use of the six categories aforementioned to query DBPedia using SPARQL based on them and extract entities related to these subjects. This was one of the two main approaches to obtaining additional data to feed the question templates, since processing text descriptions of the objects was very difficult using semantic web technologies due to the lack of information on DBPedia about them. The other approach was to use a glossary of telecommunications where we could obtain a wide range of telecommunication terms that could also be included in questions.

Then, we designed a REST API web service that would allow us to store the gathered data (object descriptions together with the automatically generated sets of questions) and be able to access it by means of API requests made by the smartphone application to view the smart object.

In the final stages we generated hard cardboard prints showing ,per object, its image, name and the corresponding QR that makes the smart object available on the smartphone application.

Finally, the smartphone gamification application was designed. In order to pique the interest of visitors as much as possible we devised several gamification techniques, such as rewarding milestone badges to those users who answer questions correctly.

We can conclude that, as we have described, every stage of the project has been successfully completed and this is reflected by the culmination of the smartphone application. We consider this project a very interesting value proposition for ETSIT's Museum, because of its innovative nature. We are offering a different way of experiencing museum visits by bringing about the chance to interact and learn more about exhibits than just by merely reading descriptions.

6.2 Problems Encountered

Throughout the development of our project there were several problems found. We will list them and explain how they have been approached to solve them:

- **Use of new technologies:**

This was the starting problem of our project since most of the technologies needed to

carry out the implementation of the gamification platform were unknown to us. Linked Data technologies such as the formation of RDF graphs, query of SPARQL endpoints and comprehending how ontologies work were the main challenge at the start of the project. Therefore, the first stage of our project was solely dedicated to studying and understanding these, and all of the technologies we were going to make use of.

- **Generation of questions:**

This was undoubtedly the main problem faced during the project. Initially we thought we could apply entity recognition schemes to the descriptions of the Museum's objects and based on that we could potentially find related and relevant additional information on DBPedia.

However, it was completely contrary to what we had foreseen, primarily because DBPedia's entries on telecommunication-related categories were not very descriptive, i.e. there is usually no information, amongst others, about the inventors of the instruments or classification into the different types of instrument.

For example, when we performed Entity Linking procedures to the description of a voltmeter belonging to the Museum's Collection, the corresponding DBpedia entry about voltmeters did not describe types of voltmeters or any additional relevant information that could make sense to use in an automated question generation process.

Moreover, the collection of these objects had no entry in DBPedia. This meant that it was increasingly difficult to find additional information to annotate our objects in order to generate sensible questions automatically through templates.

Usually museum exhibits tend to be described in DBPedia and in an initial stage of design and viability of the project we tried to think of possible questions that could be automatically generated about renowned paintings. Since these were fully described in DBPedia, it was easy to extract a lot of relevant and complimentary information through some SPARQL template prototypes and Entity Linking procedures applied to the description of the painting.

The solution to this great inconvenience, although not being ideal, was satisfactory enough given the time constraints to develop the project. Based on the categories the objects were classified into, we decided to produce questions by querying for entities related to that category. Therefore, quizzes generated have not been as object-specific as we desired them to be, nevertheless we could produce a varied dataset that produced interesting and mostly sensible questions about concepts related to the category of the object.

6.3 Future work

In this section we will outline possible implementations to enrich our project if it were to be continued in the future:

- **Build and train a natural language processing Machine Learning model:** this would be used to process the descriptions of the Museum's objects. If we were to train such model, we could potentially extract very useful pieces of information specific on the concerning object, not about the category it belongs to as we have done in our project. For instance, we could automatically identify uses of the invention or its constituent parts (as long as the description contains such information, which it usually does as we have analysed but we could not extract with our approach).

- **Produce high quality linked data from ETSIT's museum objects.**

- **Adapt the platform to be able to extend it to other renowned museums:** this gamification system could be extended beyond ETSIT's Museum to popular crowded museums where this gamification platform that turns ordinary exhibits into smart objects could contribute to enhancing the value of these cultural artworks.

APPENDIX A

Impact of this project

This appendix reflects, quantitatively or qualitatively, on the possible socio-economic, environmental and ethical implications of the project.

A.1 Socio-economic impact

Museums hold the history of humanity expressed in the form of art. A museum may gather thousands of exhibits and some of them are not as popular as others. An example of this statement is the vast amount of people that stop and stand a long time admiring Picasso's *Guernica* at Reina Sofía, whilst other Picasso's paintings on the next room attract many less visitors. People tend to pay more attention to the popular exhibits than to the rest, probably because they have heard from those before and they would like to know more about it.

Another comment is that children or adolescents may not be mature enough to fathom the significance of the exhibits shown and they may tend to find museums boring. Usually adults are found in these type of museums, contrarily to interactive science museums which are *fun* to these younger persons primarily because they can experiment with their senses and *feel* the museum.

This gamification platform could turn out to be very appealing to address these two cases. Turning ordinary exhibits into Smart Objects by using this application could really contribute towards piquing adolescents' and children's interest in cultural museums. They will be able to interact with the objects and become more interested in the works due to the quizzes, because these questions make the processing of information easier and more importantly, enjoyable. This could also mean that not only popular works are the most visited exhibits in museums, because the gamification app can be adjusted and redesigned to only granting badges and rewards to the user if they have answered correctly questions on many different exhibits, which means that the less popular ones can be promoted with this initiative.

Overall, a platform like the one designed could potentially redefine the way museum visits take place. A museum implementing this system could attract more visitors because it is a new and entertaining way of learning about the exhibits. Therefore this increase in visitors implies more revenue for the museum.

A.2 Environmental Impact

This section will define the main environmental effects of this project.

Computers and other information technology infrastructures consume significant amounts of electricity.If museums use efficient servers to host the platform's data, the energy consumption can be significantly reduced. Also, since the server will be accessed by the application when scanning QR codes, this means that it makes sense to use the application during a museum visit, which means that the server can be turned off overnight when it will not be used.

A.3 Ethical Implications

The first ethical problem we may face is that if users are given all information and entertaining questions in their smartphones about exhibits they may be less keen on hiring a museum guide, so it may pose a threat to these jobs at museums.

Another consideration to take into account is that users have to register on the platform in order to be able to advance through the game and be rewarded with badges, which means that we would have at our disposal data such as email addresses. We will not disclose any information about our users as stated by the Spanish Law.

APPENDIX B

Economic budget

This appendix details an adequate budget to bring about the project...

B.1 Physical resources

The budget for the physical resources necessary to carry out this project is limited to the cost of the computer used to develop it.

The technical specification of the computer where the platform has been developed are:

- **CPU:** Intel Core i7 7700
- **RAM:** 16 GB
- **Graphics card:** GTX 1060
- **Storage** 256GB SSD

This computer had a cost of 1100€ although the project could have been done in a cheaper computer with less resources.

B.2 Human Resources

The annual salary of a junior telecommunications engineer in Spain ranges from 20.000€ to 25.000€ euros.

B.3 Licenses

The cost of the software licences necessary to do the project is zero because all of the software packages and tools used are open-source and therefore free.

Bibliography

[1] LuAnn Wilkerson and Wim Gijselaers. Bringing problem-based learning to higher education: Theory and practice. *New directions in teaching and learning*, 68, 1996.

[2] David Boud and Grahame Feletti. *The challenge of problem-based learning*. Routledge, 2013.

[3] Cindy E Hmelo-Silver. Problem-based learning: What and how do students learn? *Educational psychology review*, 16(3):235–266, 2004.

[4] Juho Hamari, Jonna Koivisto, Harri Sarsa, et al. Does gamification work?-a literature review of empirical studies on gamification. In *HICSS*, volume 14, pages 3025–3034, 2014.

[5] Michael Meder, Till Plumbaum, Ernesto William De Luca, Sahin Albayrak, and TU DAI-Labor. Gamification: A semantic approach for user driven knowledge conservation. In *LWA*, pages 265–268, 2011.

[6] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

[7] Guillermo Vega-Gorgojo. Clover quiz: a trivia game powered by dbpedia. *Semantic Web*, (Preprint):1–15, 2018.

[8] Chenghua Lin, Dong Liu, Wei Pang, and Zhe Wang. Sherlock: A semi-automatic framework for quiz generation using a hybrid semantic similarity measure. *Cognitive computation*, 7(6):667–679, 2015.

[9] Ana María Bravo, Javier Docampo Capilla, Javier Pantoja, Carlos Navalón Casado, and Ricardo Maturana. The new prado museum website: A (semantic) challenge . the annual conference of museums and the web — 2016. 01 2016.

[10] Martin Risseeuw, Dario Cavada, Elena Not, Massimo Zancanaro, Mark Marshall, Daniela Petrelli, and Thomas Kubitza. An authoring environment for smart objects in museums: the mesch approach. 2016.

[11] Carmelo Ardito, Paolo Buono, Giuseppe Desolda, and Maristella Matera. From smart objects to smart experiences: An end-user development approach. *International Journal of Human-Computer Studies*, 114:51–68, 2018.

[12] Daniel Myers and James W McGuffee. Choosing scrapy. *Journal of Computing Sciences in Colleges*, 31(1):83–89, 2015.

[13] Eva Blomqvist, Diana Maynard, Aldo Gangemi, Rinke Hoekstra, Pascal Hitzler, and Olaf Hartig. *The Semantic Web: 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28–June 1, 2017, Proceedings*, volume 10249. Springer, 2017.

[14] Tiziano Flati and Roberto Navigli. Three birds (in the llod cloud) with one stone: Babelnet, babelfy and the wikipedia bitaxonomy. *Proceedings of SEMANTiCS2014*, 2014.

[15] Neil Middleton and Richard Schneeman. *Heroku: Up and Running: Effortless Application Deployment and Scaling.* " O'Reilly Media, Inc.", 2013.

[16] Alain Dupuy, Brigitte Juanals, and Jean-Luc Minel. Towards open museums: The interconnection of digital and physical spaces in open environments, 2015.

[17] Jordi Berenguer Sau. Diccionari de telecomunicacions. *Llengua i ús: revista tècnica de política lingüística*, (39):72–73, 2007.

[18] Albert Balcells, Enric Pujol, and S Izquierdo. Història de l'institut d'estudis catalans. vol. i: 1907-1942. *Barcelona: Institut d'Estudis Catalans. Balcells, Albert, Izquierdo, Santiago & Pujol, Enric (2007): Història de l'Institut d'Estudis Catalans*, 2:843–846, 2002.

[19] International Telecommunication Union. Vocabulario de términos de las telecomunicaciones móviles internacionales. recomendación uit-r m.1224-1 (03/2012). page 139, 2012.

[20] Tahani Alsubait, Bijan Parsia, and Uli Sattler. Generating multiple choice questions from ontologies: Lessons learnt. In *OWLED*, pages 73–84. Citeseer, 2014.

[21] Andreas Thalhammer and Achim Rettinger. Pagerank on wikipedia: towards general importance scores for entities. In *European Semantic Web Conference*, pages 227–240. Springer, 2016.

[22] Ganggao Zhu and Carlos A Iglesias. Sematch: Semantic similarity framework for knowledge graphs. *Knowledge-Based Systems*, 130:30–32, 2017.

[23] Renzo Angles and Claudio Gutierrez. The expressive power of sparql. In *International Semantic Web Conference*, pages 114–129. Springer, 2008.