

Análisis e implementación de una interfaz para visualización de un repositorio RDF



Grado en Ingeniería Multimedia

Trabajo Fin de Grado

Autor: Julián Sánchez García

Tutores: María Pilar Escobar Esteban
Gustavo Candela Romero

Febrero 2019



Universitat d'Alacant
Universidad de Alicante

Análisis e implementación de una interfaz para visualización de un repositorio RDF

Visor de datos RDF

Autor

Julián Sánchez García

Directores

María Pilar Escobar Esteban
Lenguajes y Sistemas Informáticos

Gustavo Candela Romero
Lenguajes y Sistemas Informáticos



GRADO EN INGENIERÍA MULTIMEDIA



Escuela
Politécnica
Superior



Universitat d'Alacant
Universidad de Alicante

ALICANTE, 1 de febrero de 2019

Resumen

El presente Trabajo Fin de Grado se desarrolla en el ámbito de la Web Semántica, repositorios que ofrecen Datos abiertos enlazados expresados en RDF, del inglés *Resource Description Framework* y el impacto que estos pueden tener en las bibliotecas digitales.

La aparición y desarrollo de nuevas tecnologías y los estándares de publicación para Datos abiertos enlazados o *Linked Open Data (LOD)* ofrecen grandes oportunidades para la reutilización de datos, en concreto, en el sector de las bibliotecas digitales. El método de publicación RDF se presenta como un gran aliado para convertir la información de la Web en recursos accesibles para otros usuarios. RDF ofrece soluciones para la distribución de conocimiento pero no está exento de problemas. En la actualidad, para recuperar datos en RDF los repositorios ofrecen puntos de consulta SPARQL, un lenguaje similar a SQL, del inglés *Structured Query Language* y difícil para usuarios sin los conocimientos necesarios. Visualizar este tipo de datos de una forma clara y legible también entraña dificultades por diferentes motivos. Los formatos de serialización de datos, la inconsistencia en las bases de datos de los repositorios LOD y los diferentes vocabularios empleados para etiquetar entidades, hacen que los datos recuperados sean impredecibles y por consiguiente difíciles de manipular.

Los objetivos principales de este trabajo consisten en el análisis del uso de Datos abiertos enlazados en la actualidad y en la creación de un formato de acceso gráfico para recuperar información semántica almacenada en repositorios que ofrecen sus datos en RDF. Para ello, se ha implementado una interfaz HTML que permite explotar la información almacenada en el repositorio Wikidata¹ y la Biblioteca Virtual Miguel de Cervantes (BVMC)².

Se ha realizado un análisis de las principales bases de conocimiento LOD así como de importantes web's que hacen uso de Datos abiertos enlazados. A partir del análisis de las soluciones que ofrecen para la visualización de datos expresados en RDF, se desarrolla un prototipo de interfaz HTML para recuperar y visualizar datos RDF procedentes de Wikidata en el dominio del Siglo de Oro de la Literatura Española. Para tratar de recuperar la mayor información posible de este periodo literario el visor RDF se nutre de diferentes fuentes de datos o repositorios.

¹https://www.wikidata.org/wiki/Wikidata:Main_Page

²<http://www.cervantesvirtual.com>

Preámbulo

La principal motivación para la realización de este proyecto ha sido el deseo de aprender y profundizar conocimientos sobre los métodos de publicación de Datos abiertos y el uso de datos estructurados. Ambos campos han sido tratados de forma superficial a lo largo del Grado en Ingeniería Multimedia y este Trabajo Fin de Grado se presenta como una oportunidad para ampliar conocimientos.

El tema principal sobre el que se centra este proyecto son los datos estructurados, la publicación de Datos abiertos enlazados y el diseño de interfaces usables, con la finalidad de entender mejor el método estándar de publicación RDF y de cómo trabajar con este formato para la recuperación y la correcta visualización de datos.

Quiero mostrar agradecimiento principalmente a mis dos tutores, que me han apoyado y orientado en todo momento en un ámbito complejo como es el del presente trabajo. En segundo lugar, me gustaría agradecer el esfuerzo y dedicación a los usuarios de Internet que participan de forma altruista en el desarrollo de plataformas que proporcionan datos libres para su reutilización. Estos repositorios ofrecen la oportunidad de hacer uso de los Datos abiertos enlazados que pueden aportar gran valor a los sistemas de información.

*A Macarena y a mi hija Cloe,
por su paciencia y amor.
A mi madre, por todo.*

*Se cometen muchos menos errores
usando datos incorrectos que no
empleando dato alguno*

Charles Babbage.

Índice general

Resumen	v
1. Introducción	1
2. Modelo de negocio	5
2.1. Lean Canvas	5
2.1.1. Segmento de clientes/usuarios	5
2.1.2. Problemas	6
2.1.3. Proposición única de valor	7
2.1.4. Soluciones	7
2.1.5. Canales de difusión	7
2.1.6. Flujo de ingresos	8
2.1.7. Estructura de costes	8
2.1.8. Métricas clave	8
2.1.9. Ventaja especial	8
3. Objetivos	9
3.1. Generales	9
3.2. Específicos	9
4. Estudio del problema	11
4.1. Modelo de datos RDF	11
4.2. Web Semántica, Datos abiertos y Datos enlazados	14
4.3. Bases de conocimiento	16
4.3.1. Wikidata	16
4.4. Bibliotecas Digitales y Datos enlazados	18
4.4.1. Repositorios que ofrecen Datos abiertos	20
5. Metodología	23
5.1. Iteración 1	24
5.2. Iteración 2	25
5.3. Iteración 3	25
6. Análisis y especificación	27
6.1. Ámbito del Sistema	27
6.2. Funciones del producto	27
6.3. Restricciones	27
6.4. Suposiciones y dependencias	28

6.5. Requisitos Funcionales	28
6.6. Requisitos No Funcionales	36
6.7. Casos de Uso	38
7. Diseño	41
7.1. Arquitectura seleccionada	41
7.2. Tecnologías	42
7.3. Diagrama de clases	43
7.3.1. Diagrama de Bloques	44
7.4. Mockups	44
8. Implementación	47
8.1. Iteración 1: Estudio y planificación	47
8.2. Iteración 2: Implementación	48
8.3. Iteración 3: Mejoras, pruebas y conclusiones	48
9. Pruebas y validación	49
10. Resultados	53
11. Conclusiones	63
11.1. Conclusiones	63
11.2. Líneas de trabajo futuras	64
12. Lista de acrónimos	67
Bibliografía	70
A. Anexo I. Guía de estilo	71
B. Anexo II. Estudio de Interfaces	77
C. Anexo III. SPARQL y recuperación de datos del Siglo de Oro	87
D. Anexo IV. MockUp's	111
E. Anexo V. XML vs JSON	117

Índice de figuras

2.1. Lienzo de negocio realizado con Canvanizer.	5
2.2. Interfaz para visualizar datos expresados en RDF.	7
4.1. Ejemplo de una solicitud mediante una tripleta RDF en DBpedia.	12
4.2. Representación de una tripleta RDF.	13
4.3. Punto de acceso a Datos abiertos SPARQL.	15
4.4. Timeline con autores del Siglo de Oro generado con Histropedia.	18
4.5. Consulta SPARQL para generar el timeline.	18
4.6. Volumen de Datos abiertos en diferentes repositorios.	21
4.7. Gráfico con los formatos RDF más usados.	21
5.1. Distribución de las tareas en tres iteraciones. Realizado con MS Project. .	23
5.2. Gestor de tareas Trello.	24
6.1. Diagrama de casos de uso.	39
7.1. Representación de la arquitectura del visor RDF.	41
7.2. Diagrama de clases.	43
7.3. Diagrama de Bloques.	44
7.4. Boceto a mano alzada de la interfaz.	45
7.5. Mapa web.	46
8.1. Primera interfaz para pruebas de lectura de ficheros.	47
9.1. Imagen del visor RDF en Opera.	49
9.2. Imagen del visor RDF en Firefox.	50
9.3. Imagen del visor RDF en Microsoft Edge.	50
9.4. Imagen del visor RDF en Google Chrome.	51
10.1. Interfaz principal	53
10.2. Página listado de autores.	54
10.3. Ventana emergente con información ampliada	55
10.4. Información ampliada con las consultas federadas	56
10.5. Buscador por título y autor	57
10.6. Interfaz para pruebas de lectura de ficheros RDF	58
10.7. Buscador Wikipedia	58
10.8. Interfaz acerca del TFG	59
10.9. Tiempos de trabajo por meses	60

E.1. Ejemplo de nodos en una estructura XML.	118
E.2. Representación de la estructura del archivo JSON devuelto por Wikidata.	119
E.3. Datos del archivo JSON devuelto por Wikidata para libros.	119

Índice de tablas

4.1. Algunos repositorios que ofrecen Datos abiertos.	20
6.1. RF 1 - Visualizar datos RDF de la categoría del Siglo de Oro	28
6.2. RF 2 - Visualizar autores RDF	28
6.3. RF 3 - Visualizar libros RDF	29
6.4. RF 4 - Visualizar cronología autores Siglo de Oro	29
6.5. RF 5 - Visualizar cronología de libros del Siglo de Oro	30
6.6. RF 6 - Ver ficha autor	30
6.7. RF 7 - Ver ficha libro	31
6.8. RF 8 - Paginación de resultados	31
6.9. RF 9 - Buscador de Wikipedia	32
6.10. RF 10 - Contador de resultados	32
6.11. RF 11 - Reutilización de herramientas RDF de terceros	33
6.12. RF 12 - Buscar autor	33
6.13. RF 13 - Buscar libro	34
6.14. RF 14 - Cargar archivo XML, JSON o RDF	34
6.15. RF 15 - Visualizar archivo XML, JSON y RDF	35
6.16. RF 16 - Acceder a datos BVMC	35
6.17. RNF 1 - Enlazar a fuentes	36
6.18. RNF 2 - Mostrar texto alternativo en imágenes	36
6.19. RNF 3 - Carga asíncrona	36
6.20. RNF 4 - Reutilizar información de BVMC	37
6.21. RNF 5 - Limitar consultas a Wikipedia	37
6.22. RNF 6 - Abrir enlaces en nueva ventana	37
6.23. RNF 7 - Realizar peticiones paginadas	38
6.24. RNF 8 - Usar BBDD de Wikidata	38

Índice de Listados

1.1. Consulta SPARQL que devuelve datos de libros a partir de un título dado.	2
1.2. Ejemplo RDF en formato JSON.	2
1.3. Ejemplo RDF en formato XML.	3
1.4. Ejemplo de W3C de una tripleta RDF en formato N-Triples.	3
1.5. Ejemplo de W3C de una tripleta RDF en formato XML.	3
E.1. Ejemplo de la estructura del formato JSON.	117
E.2. Ejemplo de la estructura del formato XML.	117

1. Introducción

Los últimos avances y desarrollos de la sociedad de la información han permitido la publicación de una gran cantidad de datos que se ofrece por medio de la Web. Sin embargo, mucha de esta información carece de estructura, dificultando el procesamiento por parte de las máquinas. Los diferentes modelos de datos estructurados junto al uso del método de publicación RDF, del inglés *Resource Description Framework* [Group, 2014], tratan de poner algo de orden en este sentido.

En esencia, la Web Semántica se desarrolla mediante el uso de un conjunto de prácticas que tiene como principal objetivo ofrecer todos los recursos de la Web de forma legible tanto para humanos como para máquinas. Los datos estructurados es la forma en la que esto es posible. Mediante la organización de la información en modelos de datos se codifican los datos como recursos para que sean accesibles. El uso de datos estructurados para organizar la información y los métodos de publicación de Datos abiertos enlazados o *Linked Open Data (LOD)* [W3C, 2009a] ofrecen un sinfín de posibilidades para la recuperación y reutilización de información.

Hoy en día la Web ofrece gran cantidad de Datos abiertos accesibles para la libre reutilización. Muchos de estos datos se encuentran publicados en portales y repositorios presentados mediante el método de publicación Linked Data (LD) o Datos enlazados [Lee, 2009], es decir, ofrecer recursos que ya existen en otro repositorio. Todas las posibilidades que ofrece este método de publicación se escapan al objetivo de este trabajo, si bien, se ahondará acerca del problema en la visualización de Datos abiertos expresados en RDF a partir del estudio de diferentes repositorios LOD así como las opciones que ofrecen para gestionar y visualizar sus datos. En concreto, el análisis se centra en las bibliotecas digitales y en cómo ofrecen y visualizan sus datos expresados en RDF.

Dada la actual relevancia de las tecnologías junto con los nuevos métodos de publicación de datos, es necesario preguntarse el valor que pueden aportar a las colecciones digitales atendiendo a la preservación y difusión del patrimonio cultural y cómo las bibliotecas digitales que ofrecen Datos abiertos enlazados podrían enriquecerse. El movimiento LOD plantea un escenario ideal para el desarrollo de las plataformas digitales de forma que estas puedan incorporar herramientas para la gestión, difusión y reutilización de la información.

A partir del estudio de diferentes plataformas que hacen uso de datos RDF y que siguen el manual de buenas prácticas para la publicación de Datos abiertos enlazados [W3C, 2009a], se han detectado una serie de problemas comunes que afectan principalmente a la recuperación y visualización de la información.

El primer problema se encuentra en el acceso a la información. Para la recuperación de información en los repositorios que almacenan sus datos en formatos basado en tripletas (RDF/XML, Turtle, N-Triple, N3, JSON-LD o JSON), ofrecen puntos de

acceso SPARQL, del inglés SPARQL Protocol and RDF Query Language [W3C, 2008], un lenguaje complejo para cualquier persona no especializada.

SPARQL es un lenguaje estandarizado para realizar consultas sobre repositorios que publican sus datos en RDF y que permite consultar diversas fuentes de datos en la misma consulta. Es una tecnología muy importante en la Web Semántica y surge como recomendación del consorcio W3C¹. Es un lenguaje similar a SQL que nos permite recuperar sentencias RDF. Por ejemplo, si queremos recuperar un listado de libros a partir del nombre podemos emplear la consulta del ejemplo 1.1 extraída de Wikipedia.

Listado 1.1: Consulta SPARQL que devuelve datos de libros a partir de un título dado.

```

1 {
2 PREFIX dc: <http://purl.org/dc/elements/1.1/>
3 SELECT ?title
4 WHERE { <http://ejemplo.org/libros> dc:title ?title }
5 }
```

Su estructura básica consiste en *Prefix* para determinar el espacio de nombres o *namespace*, *Select* para indicar el dato que se quiere recuperar y *Where* para indicar la condición para el conjunto de datos. Los resultados de las consultas SPARQL pueden ser conjuntos de resultados o grafos RDF.

El otro gran problema reside en el formato en el que se ofrecen los datos, ya que cuando se realiza una consulta SPARQL podemos recuperar la información en diferentes formatos. RDF es un modelo abstracto con varios formatos de serialización, es decir, pueden presentarse con distintos lenguajes de marcado por lo que la forma particular en la que se codifica un recurso o tripleta varía de un formato a otro. JSON, XML, N-Turtle o N-triples son algunos de los formatos más empleados.

Como vemos en el listado 1.2, el documento JSON organiza las URL's de los recursos en diferentes líneas, de modo que la lectura del documento es sencilla, por ejemplo, si se almacena en un *array*. De esta forma, cada objeto recuperado a su vez puede organizarse en un nuevo *array* en donde se almacenan sus propiedades y valores.

Listado 1.2: Ejemplo RDF en formato JSON.

```

1 {
2   "http://dbpedia.org/resource/Traveling_Wilburys" : { "http://dbpedia.org/
3     ontology/associatedBand" : [ { "type" : "uri", "value" : "http://dbpedia.org
4     /resource/The_Beatles" } ] } ,
5     .....
6   "http://dbpedia.org/resource/The_Beatles:_The_Biography" : { "http://dbpedia.org
7     /ontology/nonFictionSubject" : [ { "type" : "uri", "value" : "http://dbpedia
8     .org/resource/The_Beatles" } ] }
9 }
```

En el ejemplo del listado 1.3, el documento XML está estructurado de manera diferente. La información se organiza en nodos mediante etiquetas. Cada uno de estos nodos indica el recurso recuperado y sus propiedades, pero también la ontología empleada y

¹<https://www.w3.org/>

en ocasiones sus relaciones. El principal inconveniente de este formato reside en la imposibilidad de conocer de antemano el número de nodos y etiquetas de cada documento. Dependiendo del conjunto de datos que se pretenda recuperar así como del repositorio del que se obtenga, la estructura del documento puede variar. Existen librerías externas que facilitan la lectura de documentos XML pero pueden ralentizar la velocidad de respuesta del sistema.

Listado 1.3: Ejemplo RDF en formato XML.

```

1 <?xml version="1.0" encoding="utf-8" ?>
2 <rdf:RDF
3     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4     xmlns:dc="http://purl.org/dc/elements/1.1/"
5 <rdf:Description rdf:about="http://dbpedia.org/resource/The_Beatles">
6   <rdf:type rdf:resource="http://dbpedia.org/class/yago/WikicatParlophoneArtists
7     " />
8   <rdf:type rdf:resource="http://dbpedia.org/ontology/Agent" />
9   ...
10  <dbo:wikiPageExternalLink rdf:resource="http://www.guardian.co.uk/media/pda
11    /2008/nov/25/thebeatles-apple" />
12  ...
13  <rdf:Description rdf:about="http://dbpedia.org/resource/The_Beatles">
14    <dbo:wikiPageRedirects rdf:resource="http://dbpedia.org/resource/The_Beatles"
15    />
16  </rdf:Description>
17 </rdf:RDF>

```

En el ejemplo N-triples 1.4, el documento es un texto plano organizado en triplas separadas por un punto. N-triples es simplemente una forma de simplificar un documento XML como el del listado 1.5, de mayor dificultad para interpretar.

Listado 1.4: Ejemplo de W3C de una tripleta RDF en formato N-Triples.

```

1 <http://www.w3.org/2001/sw/RDFCore/ntriples/> <http://purl.org/dc/elements/1.1/
2   creator> "Dave Beckett" .
3 <http://www.w3.org/2001/sw/RDFCore/ntriples/> <http://purl.org/dc/elements/1.1/
4   creator> "Art Barstow" .
5 <http://www.w3.org/2001/sw/RDFCore/ntriples/> <http://purl.org/dc/elements/1.1/
6   publisher> <http://www.w3.org/> .

```

Listado 1.5: Ejemplo de W3C de una tripleta RDF en formato XML.

```

1 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
2     xmlns:dc="http://purl.org/dc/elements/1.1/">
3   <rdf:Description rdf:about="http://www.w3.org/2001/sw/RDFCore/ntriples/">
4     <dc:creator>Art Barstow</dc:creator>
5     <dc:creator>Dave Beckett</dc:creator>
6     <dc:publisher rdf:resource="http://www.w3.org/">
7   </rdf:Description>
8 </rdf:RDF>

```

Como se aprecia en estos ejemplos, existen diferentes formas de expresar un recurso RDF y en diferentes notaciones así como diferentes estructuras para organizar la información. Esto plantea un problema a la hora de interpretar el archivo para la recuperación

de la información y visualización en HTML. Este proceso se realiza mediante la lectura de los datos y la localización de nodos y etiquetas en el documento. Mediante el uso de lenguajes como Javascript podemos extraer la información de cada nodo para mostrarla en HTML. Al disponer de diferentes formas de notación resulta complicado aplicar generalizaciones a la hora de interpretar que información contiene cada línea del documento. En este sentido, JSON se presenta como un formato más amigable como se explica en el Anexo E con más detalle.

Por último, existe un problema inherente a las bases de conocimiento colaborativas que al ser de uso libre y de participación altruista no ofrecen la integridad deseada en sus datos. Esto produce un problema de inconsistencia en los datos y dificultad para extraer información completa y correcta. Las entidades, aun siendo del mismo dominio de datos, pueden no contener las mismas propiedades. Por ejemplo, si se analizan las entidades Miguel de Cervantes² y Francisco de Quevedo³ en Wikidata⁴, podemos ver que la propiedad P97 *Título Nobiliario* no aparece en la entidad Cervantes pero si en Quevedo. Es común en las bases de datos colaborativas encontrar entidades de un mismo dominio que no ofrecen las mismas propiedades o propiedades sin valor. La duplicidad en los datos es otro de los problemas debido a que muchos usuarios crean las mismas entidades pero con valores diferentes en sus propiedades, lo que da lugar a la misma entidad pero con algún dato diferente.

Las *consultas federadas* surgen como una oportunidad para abordar parte del problema en la inconsistencia de datos. Los sistemas de bases de datos federadas son sistemas de bases de datos cooperativos y autónomos. Los usuarios pueden acceder a datos de distintos sistemas de información en una misma consulta. De esta forma es posible buscar información sobre una misma entidad en diferentes repositorios y bases de datos.

Wikidata es una base de conocimiento colaborativa y de uso libre. Permite las consultas federadas a partir de acuerdos de cooperación con otros sistemas de información, como es el caso de la Biblioteca Virtual Miguel de Cervantes (BVMC) [Wikidata, 2017]. Wikidata ha federado su punto de servicio SPARQL lo que en la práctica quiere decir que es posible recuperar información de la BVMC mediante sentencias SPARQL en Wikidata sin pasar por la web BVMC.

El presente TFG describe una solución aplicada para el acceso a plataformas de Datos abiertos así como el diseño de una interfaz que permita recuperar y visualizar datos en formato RDF. Esta, deberá gestionar un conjunto de datos concretos, recuperar la información solicitada y visualizarla.

²<https://www.wikidata.org/wiki/Q5682>

³<https://www.wikidata.org/wiki/Q201315>

⁴https://www.wikidata.org/wiki/Wikidata:Main_Page

2. Modelo de negocio

Para el análisis de la viabilidad del proyecto se ha realizado un estudio a partir de la creación de un lienzo de negocio siguiendo el modelo Canvas Business Model [Alexander Osterwalder, 2010] para obtener una idea general del proyecto y su viabilidad.

2.1. Lean Canvas

Mediante la herramienta Canvanizer¹ se ha diseñado el lienzo de negocio.

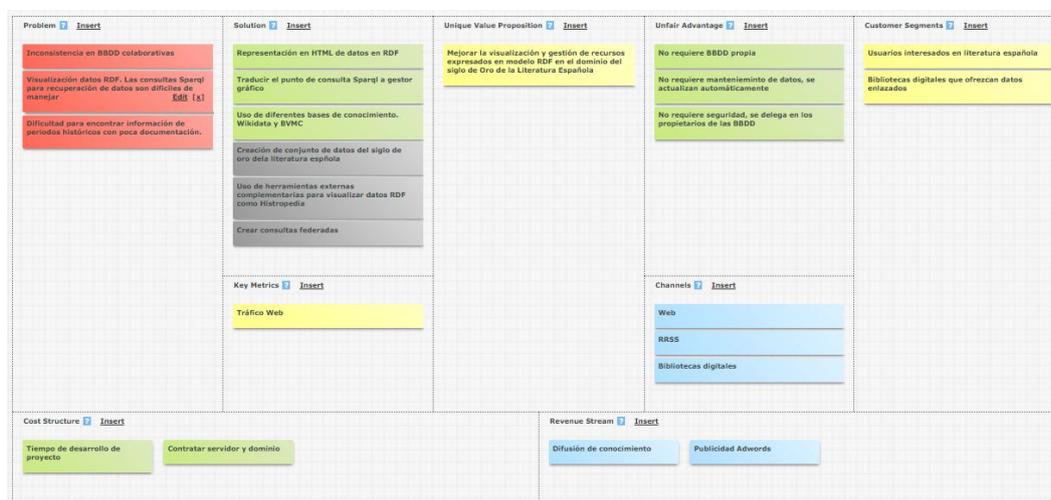


Figura 2.1.: Lienzo de negocio realizado con Canvanizer.

2.1.1. Segmento de clientes/usuarios

El grupo de clientes al que va destinado este visor RDF lo forman principalmente usuarios interesados en la literatura. A grandes rasgos, se dividen en dos segmentos: usuarios interesados en el Siglo de Oro de la Literatura Española y bibliotecas digitales que ofrecen Datos abiertos enlazados.

¹<https://canvanizer.com/canvas/wl790A5x0043J>

El primer segmento de clientes son los usuarios interesados en esta época de la literatura española y que mediante el visor de datos RDF puedan recuperar información de diferentes bases de conocimiento. Cabe destacar que la documentación sobre este periodo literario se ubica en la edad media, lo que añade una dificultad a la hora de recuperar datos exactos. Recuperar la información de diferentes bases de datos puede aportar un gran valor para actualizar y depurar información.

El segundo grupo correspondería a los *Early Adopters* o usuarios tempranos, ya que, como se estudia en el capítulo 4 son muchas las plataformas y bibliotecas digitales que ofrecen Datos abiertos enlazados pero carecen de herramientas para su correcta visualización. Para este segmento, este visor puede resultar de ayuda para mostrar de mejor manera los datos que incorporan de otras plataformas LOD.

2.1.2. Problemas

Los principales problemas detectados para los que este proyecto trata de dar soluciones son los siguientes:

- Dificultad para visualizar datos expresados en RDF. Actualmente, los puntos de consulta sobre estos datos se realizan principalmente mediante el lenguaje SPARQL.
- Inconsistencia de las bases de conocimiento colaborativas.
- Dificultad para encontrar información de periodos históricos con poca documentación.

Como se explica en el capítulo 4 y en el Anexo B, actualmente existen alternativas para visualizar datos RDF. Wikimedia tiene diferentes proyectos² que usan datos de esta base de conocimiento para mostrar información de forma particular. Un ejemplo es Histropedia³, una aplicación web que permite generar timelines o cronologías a partir de consultas SPARQL, por ejemplo, para mostrar libros por fechas de publicación.

Otro ejemplo de visor RDF es Monumental⁴, que podemos ver en la figura 2.2 y que permite buscar y visualizar información sobre lugares y ciudades a partir de datos expresados en RDF procedentes de Wikidata.

²<https://tools.wmflabs.org/admin/tools>

³<http://histropedia.com/>

⁴<https://tools.wmflabs.org/monumental/>



Figura 2.2.: Interfaz para visualizar datos expresados en RDF.

2.1.3. Proposición única de valor

Mejorar la visualización y gestión de recursos expresados en RDF en el dominio del Siglo de Oro de la Literatura Española. Ofrecer estos recursos como Datos abiertos enlazados desde diferentes fuentes poniendo a disposición de usuarios el mayor volumen de información posible.

2.1.4. Soluciones

Las soluciones planteadas pasan por la creación de un visor web que nos permita recopilar datos RDF de repositorios que ofrezcan Datos abiertos y poder visualizarlos de una forma clara y entendible. Para ello se emplearán diferentes acciones.

- Acceso a diferentes bases de conocimiento para completar y ampliar información.
- Crear una interfaz que nos permita representar en HTML datos expresados en RDF. Traducir un punto de consulta SPARQL a una interfaz más amigable, ágil e intuitiva.
- Creación de diferentes conjuntos de datos con SPARQL del dominio del Siglo de Oro. Autores, libros y datos básicos de esta época.
- Uso de herramientas externas complementarias para visualizar datos RDF.

2.1.5. Canales de difusión

Los principales canales de difusión serán la propia web y las redes sociales. En caso de poder introducir este visor en otras plataformas como bibliotecas digitales, estas podrían

ayudar también para atraer tráfico a la plataforma mediante el envío de enlaces.

2.1.6. Flujo de ingresos

Aunque no se estima una fuente de ingresos mas allá de la propia difusión de conocimiento, el uso de publicidad Adwords⁵ podría resultar interesante en caso de que la web tuviese un tráfico considerable.

2.1.7. Estructura de costes

Los costes de este proyecto se basan principalmente en el gasto de un dominio y alojamiento así como en el tiempo de desarrollo, que en su versión beta está en torno a las 300 horas de trabajo.

2.1.8. Métricas clave

Al ser un servicio con un fin más altruista que comercial no se consideran muchas de las claves de negocio habituales, como tasas de conversión, usuarios segmentados, etc. Se establece la métrica visitas de usuario para evaluar el impacto del proyecto. Si esta web es visitada o no, si el usuario pasa un tiempo determinado en el sitio web o el número de referencias consultadas por sesión, son sólo algunas métricas que pueden dar una idea aproximada del éxito o fracaso de la plataforma.

Estas métricas son configurables mediante el uso de aplicaciones como Google Analytics⁶.

2.1.9. Ventaja especial

Mediante la reutilización de Datos abiertos de otras plataforma se consiguen las siguientes ventajas especiales:

- El sistema no requiere base de datos ya que emplea las de otras plataformas.
- El sistema no requiere de seguridad en BBDD ya que la delega en los proveedores de datos.
- El sistema no requiere de mantenimiento de datos ya que estos se actualizan automáticamente de las bases de conocimiento que los proveen.

⁵<https://ads.google.com>

⁶<https://analytics.google.com>

3. Objetivos

3.1. Generales

Análisis de repositorios LOD actuales y bibliotecas digitales que ofrecen sus datos expresados en RDF. Desarrollo de una interfaz web que permita la recuperación y visualización de datos expresados en RDF así como la reutilización de datos de Wikidata y la Biblioteca Virtual Miguel de Cervantes (BVMC) para recuperar toda la información posible sobre el Siglo de Oro de la Literatura Española.

3.2. Específicos

- Estudio de herramientas actuales para la visualización de datos RDF así como las principales iniciativas que ofrecen LOD siguiendo el modelo RDF.
- Análisis del grado de aplicación del modelo RDF en las principales bibliotecas digitales.
- Analizar el dominio de datos y la recuperación de datos a través de consultas SPARQL dentro del dominio del Siglo de Oro de la Literatura Española.
- Diseño e implementación de una interfaz adaptativa en HTML. Mapear la información recibida para la correcta visualización asignando etiquetas HTML y CSS para generar el contenido web de forma dinámica.
- Organizar la información para una ágil y cómoda gestión, ofreciendo opciones de navegación al usuario.
- Añadir opciones a la interfaz para que permita consultar y mostrar datos de diferentes dominios (consultas federadas).
- Facilitar al usuario la búsqueda, visualización y reutilización de datos procedentes de Wikidata y la Biblioteca Virtual Miguel de Cervantes (BVMC) en el dominio del Siglo de Oro de la Literatura Española. Crear un buscador que permita localizar libros y autores del Siglo de Oro.
- Transformar la información visualizada en formato JSON para su reutilización y ofrecer al usuario la opción de descargar del fichero.

4. Estudio del problema

Como se ha visto en el capítulo 1, gran parte del problema de la Web Semántica se encuentra en cómo se organiza la información y el uso de diferentes formatos para representar los datos. La correcta organización permite además que los datos sean interpretables por máquinas, el motor principal de la Web Semántica. Es por ello que se requieren de los nuevos métodos de publicación y uso de tecnologías para que esta información sea más fácil de entender e interpretar.

Actualmente, son muchas las iniciativas que hacen uso de Datos abiertos enlazados, desde grandes bases de conocimiento como Wikidata hasta ayuntamientos locales, bibliotecas digitales o universidades pero, en gran parte, no ofrecen herramientas adecuadas para la gestión y visualización de datos RDF. Para obtener más información ver el Anexo B donde se estudian las interfaces de diferentes repositorios y bibliotecas digitales que hacen uso de datos RDF y las soluciones que ofrecen para visualizarlos.

- Modelo de datos RDF.
- Web Semántica, Datos abiertos y Datos enlazados.
- Bases de conocimiento.
- Bibliotecas Digitales y Datos enlazados.

4.1. Modelo de datos RDF

RDF es un marco común para la descripción de recursos en la Web, que ofrece un modelo de datos para metadatos y que a su vez nos permite clasificar la información de una manera más eficiente y mejor interpretable. Es el estándar de modelos de datos para el intercambio de información en la Web ya que proporciona información descriptiva sobre los recursos que se encuentran en la Web así como sus relaciones con otros objetos [Group, 2014].

Mediante el uso del modelo RDF se pueden alojar los recursos en la Web de forma distribuida. De esta forma, la Web Semántica se comporta como un grafo dirigido en donde las diferentes bases de datos ofrecen sus recursos codificados de forma que estos pueden ser accesibles desde diferentes puntos de consulta.

- RDF es el formato de datos idóneo para Datos abiertos enlazados.
- Describe información del recurso y sus relaciones.

- RDF usa URIs¹ para nombrar elementos unívocamente.

RDF se fundamenta en el uso de *tripletas* formadas por *sujeto-verbo-predicado* para determinar el recurso que se está enlazando, en donde el sujeto corresponde con la entidad *persona*, el predicado con sus propiedad *nacido en* y el objeto el valor de la propiedad *España*, tal como se indica en el ejemplo de la figura 4.1.

Recurso -> Propiedad -> Valor

DBpedia -> DBprop (nacimiento) -> DBpedia (España)

DBprop (residencia) -> DBpedia (Suiza) -> DBprop (provincia) -> DBpedia (Jura)

Sujeto (recurso) -> Predicado (propiedad) -> Objeto (valor)

Figura 4.1.: Ejemplo de una solicitud mediante una tripleta RDF en DBpedia.

Este modelo se basa en recuperar un recurso y describir su contenido y relación. Es un lenguaje que representa información así como un modelo de datos para objetos. La finalidad de utilizar RDF es simplificar la recuperación de datos que están alojados en diferentes repositorios y dominios. En el modelo RDF, las propiedades describen el recurso, de modo que el metadato nos da la información del recurso mediante el binomio propiedad y valor [Group, 2014].

Tal como se muestra en la figura 4.2, las entidades *libro* y *autor* existen de forma independiente. Con RDF podemos indicar no sólo su contenido, también sus relaciones y tipos que pueden haber entre las diferentes entidades, en este ejemplo, la relación *write by*.

¹<https://www.w3.org/TR/uri-clarification/>

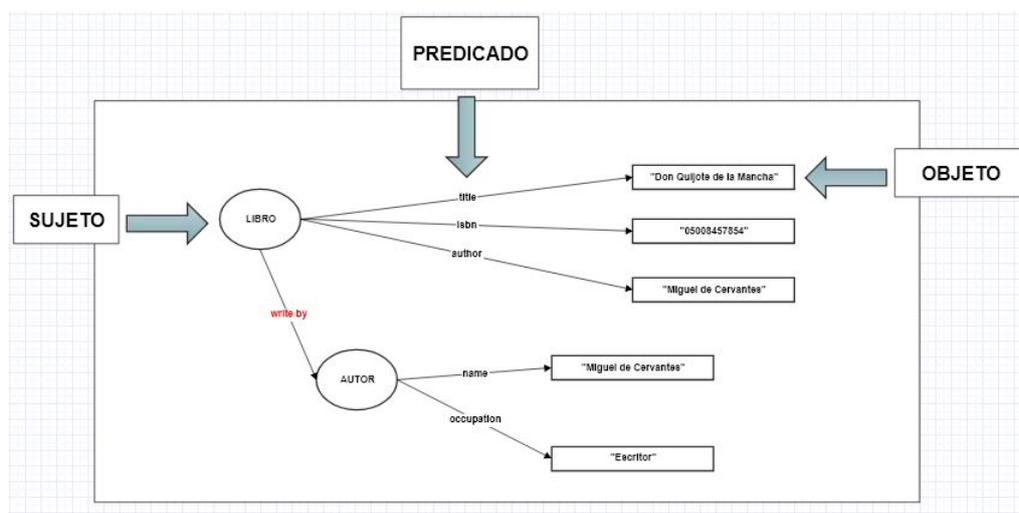


Figura 4.2.: Representación de una tripleta RDF.

Para ello, es necesario el uso de modelos estándares para definir los vocabularios de los metadatos mediante el uso de las ontologías.

Una ontología es una definición formal de tipos de propiedades y relaciones entre entidades [W3C, 2009b]. Es similar a crear un vocabulario que identifique el nombre de una propiedad y sus relaciones. Las ontologías son importantes en la Web Semántica ya que permiten definir vocabularios en un marco común. Esto también forma parte del problema que se aborda en este TFG, ya que existen diferentes vocabularios y estándares en función de los objetos o datos que queremos modelar. Para ello, OWL y Dublin Core tratan de aportar soluciones.

OWL (Lenguaje de Ontologías Web OWL)² es el lenguaje que permite representar explícitamente el significado de términos en vocabularios y las relaciones entre esos términos [W3C, 2009b]. OWL plantea un modelo mediante etiquetas que definen qué datos se almacena en cada tipo de entidad así como de que forma puede relacionarse con otras entidades.

Actualmente existen iniciativas que proponen un marco común para la creación de ontologías y vocabularios, como Schema³ y Dublin Core⁴. Ambas organizaciones están dedicadas a fomentar la adopción de los estándares interoperables de los metadatos y a promover el desarrollo de los vocabularios especializados para describir recursos en forma de datos estructurados. Actualmente, la propuesta OWL es más potente que Schema ya que permite más formas de clasificación y relación.

²<https://www.w3.org/2007/09/OWL-Overview-es.html>

³<https://schema.org/>

⁴<http://dublincore.org>

En resumen, las ontologías indican la relación y su tipo entre entidades. Mediante el lenguaje OWL, podemos crear vocabularios para definir las ontologías mientras que Dublin Core y Schema proponen un marco común para la definición de los vocabularios de metadatos.

4.2. Web Semántica, Datos abiertos y Datos enlazados

Convertir la Web en una gran fuente de información interconectada está más próximo gracias a la aparición y desarrollo de nuevas tecnologías y modelado de datos. Los estándares de publicación LOD permiten acceso a grandes volúmenes de información y esto plantea nuevos retos y oportunidades que permitan establecer sistemas de información combinando fuentes de diferentes repositorios. El creciente número de plataformas y repositorios que ofrecen Datos abiertos abre un mundo de posibilidades para la reutilización de información. La construcción de una Web Semántica permite interconectar de forma automática muchos de los recursos de la Web como los repositorios Open Data (OD) o *Datos abiertos*, que ofrecen sus recursos en diferentes formatos y de los cuales hacen uso otras muchas web's mediante los Datos enlazados.

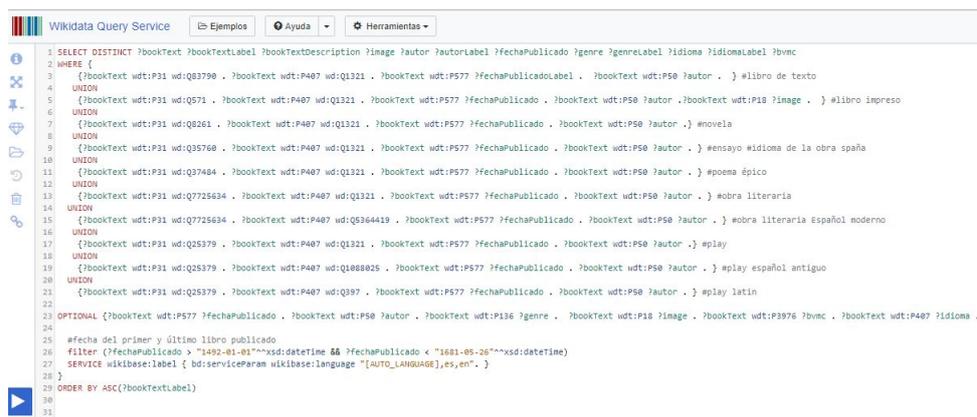
El concepto de Web Semántica nace casi con la propia Web. Tim Bender Lee⁵, considerado el padre de la Web, ha trabajado durante su carrera para la creación de una Web más interconectada, con contenido enlazado de forma dinámica que ayude así a completar la información en nuestras búsquedas [Web, 2012]. Podríamos considerar que el hipertexto es la primera piedra para la creación de la Web Semántica pero en la actualidad este concepto de Web posee mayor significado y tiene como finalidad ofrecer al usuario respuestas a sus preguntas de forma más rápida y sencilla. Esto es posible gracias a una información mejor definida y a la utilización de una infraestructura común (RDF) mediante la cual es posible compartir, procesar y transferir información. En términos generales, la Web Semántica consiste en trabajar con datos estructurados en lugar de texto y enlaces.

RDF, SPARQL y OWL⁶ son los mecanismos actuales mediante los cuales podemos compartir y reutilizar información.

No obstante, la Web Semántica no está exenta de obstáculos que todavía hoy ralentizan su desarrollo. Actualmente faltan puntos de acceso a datos más amigables a diferencia de SPARQL 4.3 y no se aprecia una clara voluntad de invertir en datos para compartir al no tener una clara proyección comercial.

⁵<https://www.w3.org/People/Berners-Lee/>

⁶<https://es.wikipedia.org/wiki/OWL>



```

1 SELECT DISTINCT ?bookText ?bookTextLabel ?bookTextDescription ?image ?autor ?autorLabel ?fechaPublicado ?genre ?genreLabel ?idioma ?idiomaLabel ?bvmc
2 WHERE {
3   {?bookText wdt:P31 wd:Q83790 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicadoLabel . ?bookText wdt:P50 ?autor . } #libro de texto
4   UNION
5   {?bookText wdt:P31 wd:Q571 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . ?bookText wdt:P18 ?image . } #libro impreso
6   UNION
7   {?bookText wdt:P31 wd:Q8261 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor .} #novela
8   UNION
9   {?bookText wdt:P31 wd:Q35760 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #ensayo #idioma de la obra spaña
10  UNION
11  {?bookText wdt:P31 wd:Q37484 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #poema épico
12  UNION
13  {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #obra literaria
14  UNION
15  {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q5364419 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #obra literaria Español moderno
16  UNION
17  {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor .} #play
18  UNION
19  {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1888825 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #play español antiguo
20  UNION
21  {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q397 . ?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . } #play latín
22  }
23 OPTIONAL {?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . ?bookText wdt:P136 ?genre . ?bookText wdt:P18 ?image . ?bookText wdt:P3976 ?bvmc . ?bookText wdt:P407 ?idioma .}
24
25 #fecha del primer y último libro publicado
26 filter (?fechaPublicado > "1492-01-01"^^xsd:date && ?fechaPublicado < "1681-05-26"^^xsd:date)
27 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es,en" . }
28 }
29 ORDER BY ASC(?bookTextLabel)
30
31

```

Figura 4.3.: Punto de acceso a Datos abiertos SPARQL.

Según la Carta Internacional de Datos abiertos (Open Data Charter)⁷, Datos abiertos son todos aquellos datos digitales que se ofrecen a disposición del usuario con ciertas características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar. Los principios básicos sobre los que se basa *Open Data Charter* son los siguientes:

- Datos abiertos por defecto.
- Oportunos y exhaustivos.
- Accesibles y utilizables.
- Comparables e interoperables.
- Para mejorar la gobernanza y la participación ciudadana.
- Para el desarrollo incluyente y la innovación.

El uso de estas prácticas ofrecen un gran número de beneficios como mejorar el flujo de información entre gobiernos y ciudadanos o poner a disposición información que puedan tener un tratamiento de investigación o difusión de conocimiento. Gracias a la disposición de estos datos de forma libre, muchas plataformas pueden ampliar y completar sus información mediante el uso de Datos enlazados.

El objetivo principal de los Datos enlazados [Lee, 2009] es el de completar la información con recursos de otras fuentes. Para ello, muchos repositorios ofrecen sus recursos como datos estructurados para que otras plataformas los reutilicen como Datos enlazados.

⁷<http://opendatacharter.net/principles-es>

Estos datos pueden ser interconectados entre sí a través de tecnologías estándar como HTTP, RDF y URI, y que en lugar de ofrecer páginas web para lectores humanos al mismo tiempo ofrece esa información con un formato concreto interpretable por ordenadores. Esto permite enlazar información de distintas fuentes ofreciendo un contenido mucho más rico.

En el sector de la bibliotecas digitales se pueden obtener muchos beneficios, por ejemplo, enriqueciendo sus plataformas con datos de otros repositorios y a su vez ofrecer los propios datos en abierto para que otras bibliotecas puedan hacer lo mismo. El fin del uso de esta filosofía de publicación es crear una gran fuente de conocimiento, global, colaborativa, dinámica y que permita el acceso a datos que están en constante actualización. En este sentido, las bases de conocimiento colaborativas se presentan como una gran oportunidad dado el gran número de personas que colaboran en ellas y la gran fuente de información que albergan.

4.3. Bases de conocimiento

Las bases de conocimiento nacen como una evolución necesaria a las bases de datos. La automatización de procesos da lugar al requerimiento de consultas constantes sobre datos por lo que las bases de conocimiento se comportan como un tipo especial de bases de datos [Wikipedia, 2018]. Deben cumplir una serie de requisitos como proveer de herramientas para la recolección, organización y recuperación (visualización) de datos. Generalmente, se clasifican en dos grupos, las bases de conocimiento legibles por máquinas y las legibles por humanos. Las primeras se emplean para aplicar razonamientos deductivos automatizados. Las bases de conocimiento legibles por humanos están diseñadas para procesos de aprendizaje, difusión de conocimiento y colaboración con el objetivo final de encontrar soluciones a problemas ya resueltos. El presente trabajo se centra en este segundo grupo.

Algunas plataformas como DBpedia⁸, Freebase⁹ o Wikidata¹⁰ ofrecen una base de conocimiento colaborativa compuestas principalmente por datos estructurados que permiten una organización de la información. Estas plataformas permiten la creación, manipulación y recuperación de datos bajo licencias libres.

En general, todas cumplen la función de proveer libre acceso a datos pero es Wikidata sobre la que se apoyan el mayor número de proyectos LOD. Por este motivo, se ha elegido como base de conocimiento para este proyecto.

4.3.1. Wikidata

Uno de los proyectos más importantes en LOD es el llevado a cabo por la Fundación Wikimedia¹¹. Esta organización sin ánimo de lucro y conocida fundamentalmente por su proyecto Wikipedia, ofrece diferentes plataformas para la divulgación de conocimiento.

⁸<http://es.dbpedia.org/>

⁹<https://developers.google.com/freebase/>

¹⁰<https://www.wikidata.org/>

¹¹<http://www.wikimedia.es/wiki/Portada>

Wikibooks, Wikcionario o Wikidiversidad son sólo algunos ejemplos de proyectos Wikimedia que trabajan con Datos abiertos. Estas herramientas utilizan las bases de datos de Wikidata para diversos fines [Tomás Saorín, 2018].

El presente trabajo se sustenta sobre su posiblemente más ambicioso proyecto, Wikidata¹², un gran almacén de Datos abiertos que Wikimedia pone a disposición de los usuarios.

Wikidata es una base de conocimiento estructurada y mantenida de forma colaborativa. El objetivo de esta plataforma es crear una fuente común de conocimiento bajo licencia Creative Commons¹³. De esta manera sus datos están disponibles para ser reutilizados en otros proyectos Wikimedia o por terceros [Pellissier Tanon et al., 2016].

La manera en que ofrece sus datos de forma estructurada es mediante entidades, representadas por su URI. Por ejemplo, la entidad *Q5682* representa a *Miguel de Cervantes* y sus propiedades.

En Wikidata además de entidades se definen también sus propiedades. Hay más de 4.700 propiedades [Tomás Saorín, 2018] agrupadas en 8 ámbitos: Generic, Person, Organization, Events, Works, Terms, Geographical feature y Others.

Estas propiedades se representan mediante espacio de nombres¹⁴ o *namespaces*. Las propiedades se comportan como el vocabulario para añadir metadatos a una entidad.

Por ejemplo: la propiedad *P31* corresponde con el namespace *instancia de*. Esta propiedad puede asumir diferentes valores, por ejemplo un ser *humano* (Q5) o una *forma literaria* (Q25379).

El principal problema que genera esta forma de organizar la información se encuentra en que no todas las entidades de un mismo ámbito requieren de las mismas propiedades. Una instancia de *humano* (Q5) puede ser un pintor o un político y cada uno de ellos se puede representar por diferentes propiedades. Por lo tanto, no se puede encontrar la homogeneidad deseada en este sentido y se deben de estudiar las propiedades de los objetos que se pretenden recuperar. De esta manera podemos establecer las propiedades mínimas que deberían de tener las entidades de una misma clase. Se amplía información en el Anexo C.

Gracias a la infraestructura proporcionada por Wikidata así como su forma de publicación LOD, esta base de conocimiento permite la reutilización de sus datos por terceros y cada día crece el número de iniciativas para la reutilización y visualización de Datos abiertos. Histropedia, que podemos ver en la figura 4.4, es un aplicación web que permite crear líneas de tiempo o cronologías a partir de datos proporcionados por Wikidata. Mediante consultas SPARQL como la de la figura 4.5 podemos recuperar información y asignar etiquetas a la información recuperada para su representación.

¹²<http://www.wikidata.org>

¹³<http://es.creativecommons.org>

¹⁴<https://www.wikidata.org/wiki/Help:Namespaces>

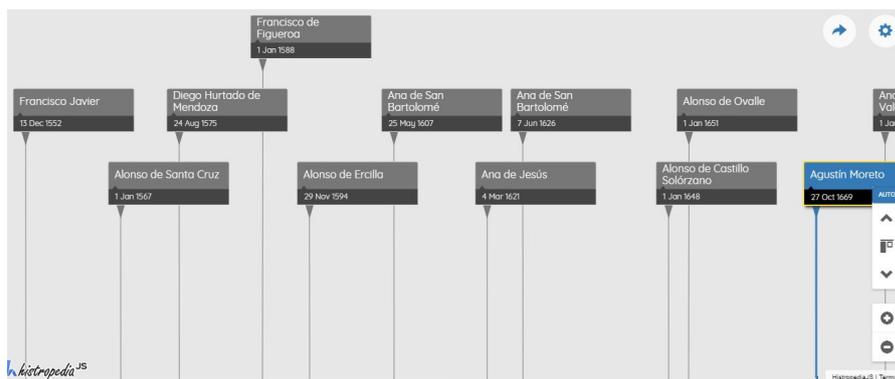


Figura 4.4.: Timeline con autores del Siglo de Oro generado con Histropedia.

Enter a SPARQL query:

```

SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel ?image ?firma
WHERE{
  {
    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
    ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en espanyol
    ?autor wdt:P27 wd:Q29 . #pais de procedencia
    ?autor wdt:P106 wd:Q36180 . #Escritor
    ?autor wdt:P569 ?birth.
    ?autor wdt:P570 ?birthDeath.
    ?autor wdt:P373 ?category .
    ?autor wdt:P106 ?ocupacion.
  }
  UNION
  {
    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
    ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en espa
    ?autor wdt:P27 wd:Q29 . #pais de procedencia
    ?autor wdt:P106 wd:Q49757 . #Poeta
    ?autor wdt:P569 ?birth.
    ?autor wdt:P570 ?birthDeath.
  }
}

```

Figura 4.5.: Consulta SPARQL para generar el timeline.

4.4. Bibliotecas Digitales y Datos enlazados

Gracias a las diferentes iniciativas LOD, la Web de los datos se presenta como un gran repositorio de información que ofrece nuevas posibilidades a las bibliotecas digitales, facilitando la ampliación de la información desde otras fuentes de datos así como la gestión de sus datos mediante tecnologías semánticas.

El movimiento LOD y el método de publicación de datos RDF pueden convertirse en poderosos aliados de las bibliotecas digitales y archivos de información. Estos repositorios alojan documentos que requieren altos niveles de estructuración por lo que se

hace necesario el uso de nuevas formas de clasificación. La organización de los recursos mediante RDF añade nuevas posibilidades también para la reutilización de datos.

Actualmente, la tendencia en publicación LOD va en aumento [Sulé, 2015] y diferentes webs, fundaciones o instituciones públicas ofrecen gran cantidad de datos de acceso libre y a su vez estos datos son reutilizados por terceros para enriquecer sus sitios web. El desarrollo de las bases de datos colaborativas y el método de publicación RDF, junto con la forma en la que muchas instituciones sirven sus datos, da pie a que estos puedan incorporarse a otros sistemas de información. Para conocer en qué grado se aplica el modelo RDF así como las herramientas que se facilitan para la recuperación de información, se han estudiado diferentes portales e instituciones, el número y tipos de documentos que ofrecen y las soluciones para la visualización y gestión. En primer lugar, se han analizado algunas bibliotecas digitales relevantes.

La **Biblioteca Virtual Miguel de Cervantes (BVMC)**¹⁵ colabora con el proyecto Wikidata ofreciendo su repositorio semántico con miles de autores y obras. Esta biblioteca está federada por Wikidata y permite realizar consultas federadas desde su punto de acceso SPARQL hacia su repositorio. De esta forma, es posible recuperar información de BVMC a través de consultas a Wikidata. También ofrece información enlazada desde Wikipedia completando así información de libros y autores. Dispone de unos 200.000 registros recientemente migrados a tripletas RDF [Candela et al., 2017]. Ofrece una interfaz con un buscador básico para consulta. Ver más detalles en el Anexo B.

La **Biblioteca Nacional Francesa (BNF)**¹⁶ ofrece un amplio catálogo de libros, películas e imágenes principalmente. Cuenta con más de 2 millones de autores y 260.000 obras [BNF, 2014]. La BNF permite consultas federadas en Wikidata con el *namespace* P268. Muestra la información de forma legible pero no permite la navegación con información relacionada.

Por último, la **Biblioteca Nacional Española (BNE)**¹⁷ a través de su portal, ofrece Datos abiertos enlazados disponibles en formato RDF-Turtle principalmente y RDF-XML. Cuenta con más de 2 millones de referencias y también permite consultas federadas en Wikidata con el identificador *P950*. Ofrece 4 clases de objetos, entidades, obras, autores y personas [Muñoz, 2014]. Incluye un buscador avanzado que permite filtros y buscar por dominios (autor, obra...). Muestra la información de forma legible y ofrece información relacionada.

En conclusión, las tres bibliotecas ofrecen Datos abiertos enlazados y están federadas en Wikidata, lo que permite extraer información de libros y autores a través del punto de consulta Wikidata¹⁸. Las tres plataformas muestran los datos de forma legible pero, salvo la Biblioteca Nacional Española (BNE), las opciones de navegación que ofrecen son deficientes ya que no permiten la navegación por épocas o fechas. En el Anexo B se incluyen imágenes y características sobre cada una de las plataformas.

¹⁵<http://www.cervantesvirtual.com/>

¹⁶<https://data.bnf.fr>

¹⁷<http://datos.bne.es>

¹⁸<https://query.wikidata.org>

4.4.1. Repositorios que ofrecen Datos abiertos

Se analiza en qué grado y forma se aplica el modelo RDF en importantes colecciones digitales y portales que ofrecen Datos abiertos enlazados. Se muestra un resumen sobre algunos repositorios de interés y en qué formato sirven sus datos. De esta forma se decide sobre qué base de conocimiento y conjunto de datos se desarrollará el visor RDF. Se han recogido datos al respecto en fecha 09/12/2018 y pueden consultarse en la tabla 4.4.1.

Portal	Formato	Archivos	URL
Datos abiertos UE	RDF	12692	data.europa.eu
	JSON	178	
	XML	130	
		1368	
Gobierno de España	JSON	12692	datos.gob.es
	XML	7680	
	RDF-TRUTLE	2028	
	RDF-N3	1298	
	RDF-XML	257	
		719	
BVMC	JSON	233691	data.cervantesvirtual.com
	RDF	812	
		812	
Wikidata	JSON	?	wikidata.org
	JSON	128237	Libros
		269360	Autores
BNE	RDF-TURTLE	2014824	datos.bne.es
		?	Datos sobre obras
BNF	?	140772	catalogue.bnf.fr/index.do

Tabla 4.1.: Algunos repositorios que ofrecen Datos abiertos.

Como podemos ver en la figura 4.6, la Biblioteca Nacional Española (BNE) ofrece más recursos que el resto. Se puede concluir que la tendencia a la publicación de Datos abiertos con el modelo RDF es creciente a tenor de la importancia de los sitios analizados. Cabe indicar que sobre Wikidata solo se han considerado datos de autores y obras. Los datos de esta tabla 4.4.1 son aproximados y sólo sirven para analizar la tendencia en cuanto a publicación en RDF.

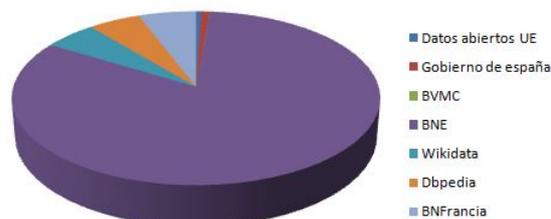


Figura 4.6.: Volumen de Datos abiertos en diferentes repositorios.

Se ha estudiado el formato más habitual en el que estos repositorios ofrecen sus datos y se concluye que RDF y JSON son los formatos más usados en los repositorios analizados, tal como indica la figura 4.7. Como RDF se estructura de manera similar a XML entraña una dificultad similar a la hora de leer su contenido, como explica el capítulo 1. Por lo tanto, se elige JSON como formato candidato para la recuperación de información como se indica en el Anexo E. En la figura 4.7 se agrupan en RDF los formatos RDF-Turtle, RDF-N3 y RDF-XML.

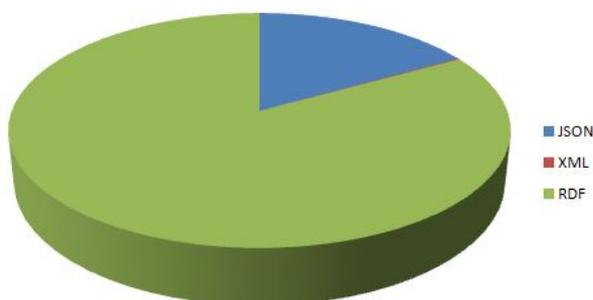


Figura 4.7.: Gráfico con los formatos RDF más usados.

Otras instituciones como ayuntamientos y administraciones públicas, ofrecen gran cantidad de Datos abiertos para su reutilización. El caso del Ayuntamiento de Valencia¹⁹ o la Diputación de Alicante²⁰ son sólo algunos ejemplos de la tendencia hacia esta filosofía de publicación. Estas y otras instituciones públicas ponen a disposición de ciudadanos

¹⁹<http://gobiernoabierto.valencia.es/va/data>

²⁰<http://datos.diputacionalicante.es>

Datos abiertos para su reutilización. En su mayor parte, son datos estadísticos como niveles de contaminación, pluviosidad, censos, etc... y aunque este tipo de datos no tengan un interés particular para este proyecto aporta una idea general sobre la tendencia actual en publicación de Datos abiertos.

Sin embargo, las herramientas para la visualización de grandes volúmenes de datos o de fuentes poco homogéneas es un área poco explorada en la actualidad. Se analizan las interfaces de importantes repositorios que publican sus datos en RDF y que se puede leer de forma ampliada en el Anexo B. También se incluyen en el estudio otras herramientas de terceros para la visualización de datos RDF.

Mediante el análisis de diferentes visores de datos, en su mayoría RDF, se extraen las conclusiones que sirven como punto de partida para el diseño de la interfaz de este proyecto. Se analizan sus ventajas y desventajas para entender parte del problema al que se enfrenta este TFG. Los problemas principales se podrían resumir en dos bloques:

- Cómo de organizados están los datos a los que accedemos.
- Dificultad para visualizar datos expresados en formatos como RDF, JSON o XML.

Tras el estudio, se observa que las interfaces que mejor muestran la información son herramientas externas como por ejemplo, Monumental²¹ o Reasonator²², pero requieren del uso de datos externos y no ofrecen opciones para navegar ni buscador avanzado. Por el contrario, los repositorios suelen ofrecer mejores buscadores, opciones de navegación y filtros pero pierden en cuanto a visualización de datos.

En líneas generales, la aplicación del modelo RDF es desigual en los portales analizados y los formatos de representación son desiguales. En la mayoría de los casos se requieren herramientas externas, bien para transformar los datos a tripletas RDF bien para visualizar esos datos. Además, muchos de estos portales no ofrecen punto de consulta SPARQL, fundamental para la consulta de datos RDF.

²¹<https://tools.wmflabs.org/monumental/>

²²<https://tools.wmflabs.org/reasonator/>

5. Metodología

Este capítulo se centra en el proceso que se ha seguido para el desarrollo del proyecto y marca las pautas de trabajo. En líneas generales, la metodología empleada ha sido la siguiente.

- Análisis de sistemas similares.
- Descripción en UML de casos de uso y diagramas de clase.
- Implementación del sistema en HTML, CSS y Javascript.
- Implantación del sistema de búsqueda.
- Pruebas de funcionamiento y carga.

El desarrollo de este proyecto se ha realizado siguiendo la metodología *Scrum*¹. Esta ha permitido dividir el trabajo en 3 iteraciones para poder organizar las tareas y estimar los plazos de trabajo como se aprecia en la figura 5.1.

tarea	Nombre de tarea	Duración	Comienzo	Fin	Previ	los
★	▪ Hito 1: Estudio del arte y dominio de datos	134 hrs	sáb 01/09/18	lun 17/09/18		
✎	Planificación de proyecto	4 hrs				
✎	Análisis de visores RDF	16 días				
✎	DATOS Wikidata - BVMC (siglo de oro)	8 hrs				
✎	SPARQL y consultas de información	12 hrs				
✎	Hoja de Estilos	3 hrs				
✎	MOCKUPS	5 hrs				
✎	Lectura de fichero XML RDF JSON	12 hrs				
✎	▫ Documentación	24 hrs				
✎	▫ Memoria - 1	32 hrs				9
★	▪ Hito 2 Implementar HTML y Buscador	100 hrs	lun 15/10/18	sáb 27/10/18		
✎	▪ Desarrollo Visor					
✎	IEEE	8 hrs				
✎	Implementación Mockup's HTML	12 hrs				
✎	▪ Programación JS					
✎	Conexión AJAX	1 hr				
✎	Implementar Consultas Sparql	6 hrs				
✎	Mostrar datos recuperados	24 hrs				
✎	▫ Memoria - 2					
★	▪ Hito 3 Diseño Mejoras	100 hrs	mar 20/11/18	dom 02/12/18		21
✎	Maquetación web. CSS	12 hrs				
✎	Mejoras: Descargar fichero JSON	12 hrs				
✎	Mejoras: filtros en el buscador	20 hrs				
✎	Mejoras: consultas federadas parametrizadas	6 hrs				
✎	Mejoras: parser JSON XML RDF	48 hrs				
✎	Mejoras: buscador wikipedia	6 hrs				
✎	Pruebas de validacion	10 hrs				
✎	Consultas Federadas	8 hrs				
✎	▫ Memoria - 3	60 hrs				

Figura 5.1.: Distribución de las tareas en tres iteraciones. Realizado con MS Project.

¹[https://es.wikipedia.org/wiki/Scrum_\(desarrollo_de_software\)](https://es.wikipedia.org/wiki/Scrum_(desarrollo_de_software))

Para la organización y planificación del trabajo, se ha utilizado Microsoft Project (MS Project)². El motivo de la elección de este software se debe a que permite, entre otras cosas, organizar las tareas por tiempos de forma sencilla y ofrece gráficos automatizados como los diagrama de Gantt, que ayudan a estudiar el desarrollo del trabajo y prever retrasos. También se ha empleado el tablero Trello³ de la figura 5.2 para tener una visión general del trabajo mediante el uso de tarjetas y técnicas Kahnban⁴.

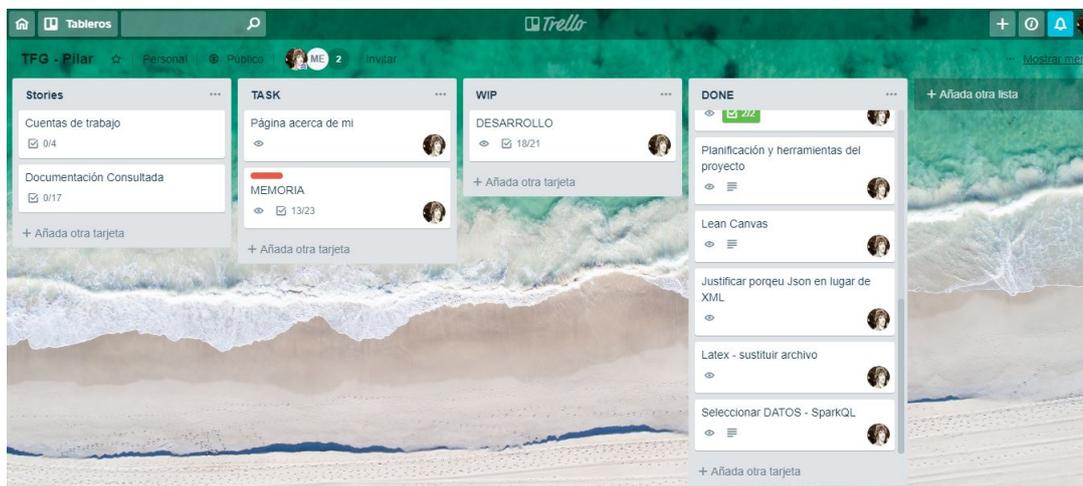


Figura 5.2.: Gestor de tareas Trello.

5.1. Iteración 1

Estudio del problema y análisis de diferentes visores RDF estudiando sus ventajas y desventajas que se puede consultar en el Anexo B. Análisis de web's que usan LOD y las soluciones que ofrecen destacables bibliotecas digitales para mostrar datos en RDF. Estudio acerca de los principales repositorios [Pellissier Tanon et al., 2016] LOD y RDF. Planificación del proyecto con las siguientes herramientas:

- MS Project.
- Toggle para medir tiempos por tarea.
- Trello.
- GitHub⁵.

²<https://products.office.com/es-es/project/project-and-portfolio-management-software>

³<https://trello.com/b/1iT4rK7o/tfg-pilar>

⁴<https://es.wikipedia.org/wiki/Kanban>

⁵<https://github.com/juliansanchez/rdfvisor>

Diseño de mockup's que se puede consultar en el Anexo D. Creación de consultas SPARQL para la recuperación de datos RDF que se puede ver de forma ampliada en el Anexo C. Estudio de los formatos de serialización en RDF ampliado en el Anexo E.

5.2. Iteración 2

Creación del lienzo de negocio o Lean Canvas [Alexander Osterwalder, 2010] y a partir del estudio del problema, redacción del documento IEEE con las especificaciones descritas en el capítulo 6. Implementación de mockup's en HTML y principales requisitos funcionales.

5.3. Iteración 3

Implementación del resto de requisitos y aplicar estilo CSS. Implementación de consultas federadas Wikidata y BVMC. Depuración de código y pruebas de validación W3C. A lo largo de las iteraciones se han ido redactando los diferentes puntos de la memoria del proyecto.

6. Análisis y especificación

En este capítulo se presentan los principales requerimientos funcionales y no funcionales según la norma IEEE 830.

6.1. Ámbito del Sistema

El visor de datos RDF realiza consultas en la Web sobre diferentes repositorios que ofrecen Datos abiertos en RDF para extraer información de un periodo de la literatura española. El ámbito donde se ubica el proyecto es el de la Web Semántica, el movimiento LOD y las bibliotecas digitales.

6.2. Funciones del producto

- Conectar con repositorios de Datos abiertos para recuperar información.
- Visualizar de forma legible datos expresados en RDF.
- Listar autores y obras literarias.
- Realizar búsquedas sobre la información recuperada.
- Realizar búsquedas en Wikipedia.
- Recuperar información de otras bibliotecas mediante consultas federadas.
- Cargar archivos en RDF, JSON o XML para su visualización.

6.3. Restricciones

- Aplicación para ser utilizada con Internet.
- El servidor debe ser capaz de atender consultas concurrentemente.
- La interfaz se diseñará siguiendo un modelo adaptativo.
- Los lenguajes de programación empleados serán SPARQL, HTML5, CSS3 y JS.
- Habilidad: usuario básico de Internet
- Criticabilidad: el sistema no puede funcionar de forma autónoma ya que depende de otros servicios externos de terceros, como los proveedores de datos.

6.4. Suposiciones y dependencias

El sistema no requiere de almacenamiento ya que la información mostrada proviene de bases de datos externas.

6.5. Requisitos Funcionales

Se describen las funciones principales del sistema.

RF 1: Visualizar datos RDF categoría del Siglo de Oro

Requisito	RF 1
Nombre requisito	Visualizar datos RDF del Siglo de Oro
Descripción	Visualizar datos de la categoría Wikidata
Prioridad	Alta

Tabla 6.1.: RF 1 - Visualizar datos RDF de la categoría del Siglo de Oro

- Justificación: Ubicar al usuario en la época literaria.
- Precondición: Acceder a la página principal.
- Descripción: Mostrar datos básicos de la categoría Wikidata Siglo de Oro de la Literatura Española.

RF 2: Visualizar autores RDF

Requisito	RF 2
Nombre requisito	Visualizar autores RDF
Descripción	Visualizar datos RDF de autores del Siglo de Oro
Prioridad	Alta

Tabla 6.2.: RF 2 - Visualizar autores RDF

- Justificación: Visualizar datos RDF de autores de Wikidata.
- Precondición: Acceder a la página autores.
- Descripción: Listar todos los autores del Siglo de Oro de la Literatura Española.

RF 3: Visualizar libros RDF

Requisito	RF 3
Nombre requisito	Visualizar libros RDF
Descripción	Visualizar libros del Siglo de Oro en RDF
Prioridad	Alta

Tabla 6.3.: RF 3 - Visualizar libros RDF

- Justificación: Visualizar libros de Wikidata en formato RDF.
- Precondición: Acceder a la página libros.
- Descripción: Visualizar todos los libros del Siglo de Oro de la Literatura Española en RDF.

RF 4: Visualizar cronología autores

Requisito	RF 4
Nombre requisito	Visualizar cronología autores Siglo de Oro
Descripción	Usar Histropedia para crear cronología de autores
Prioridad	Baja

Tabla 6.4.: RF 4 - Visualizar cronología autores Siglo de Oro

- Justificación: Reutilización de otros visores RDF y ofrecer una vista más atractiva.
- Precondición: RF 11.
- Descripción: Mediante la reutilización de las consultas SPARQL ofrecer al usuario otro tipo de vista de los datos RDF recolectados.

RF 5: Visualizar cronología libros

Requisito	RF 5
Nombre requisito	Visualizar cronología libros Siglo de Oro
Descripción	Usar Histropedia para crear cronología de libros
Prioridad	Baja

Tabla 6.5.: RF 5 - Visualizar cronología de libros del Siglo de Oro

- Justificación: Reutilización de otros visores RDF y ofrecer una vista más atractiva.
- Precondición: RF 11.
- Descripción: Mediante la reutilización de las consultas SPARQL ofrecer al usuario otro tipo de vista de los datos RDF recolectados.

RF 6: Ver ficha autor

Requisito	RF 6
Nombre requisito	Ver ficha autor
Descripción	Visualizar información ampliada del autor
Prioridad	Media

Tabla 6.6.: RF 6 - Ver ficha autor

- Justificación: Ofrecer una información ampliada del autor
- Precondición: Clicar sobre uno de los autores.
- Descripción: Mediante un elemento emergente (PopUp) se se muestra más información del autor.

RF 7: Ver ficha libro

Requisito	RF 7
Nombre requisito	Ver ficha libro
Descripción	Visualizar información ampliada del libro
Prioridad	Media

Tabla 6.7.: RF 7 - Ver ficha libro

- Justificación: Ofrecer información ampliada sobre un libro.
- Precondición: Clicar sobre uno de los libros.
- Descripción: Mediante un elemento emergente se muestra información ampliada del libro.

RF 8: Paginación de resultados

Requisito	RF 8
Nombre requisito	Paginación de resultados
Descripción	Listar todos los datos recuperados de forma paginada
Prioridad	Alta

Tabla 6.8.: RF 8 - Paginación de resultados

- Justificación: Facilitar la navegación del usuario.
- Precondición: Mostrar listas de autores o libros.
- Descripción: Ofrecer al usuario 2 botones para mostrar página anterior y siguiente en la lista de resultados de autores y libros.

RF 9: Buscador de Wikipedia

Requisito	RF 9
Nombre requisito	Buscador de Wikipedia
Descripción	Ofrecer un buscador de Wikipedia
Prioridad	Alta

Tabla 6.9.: RF 9 - Buscador de Wikipedia

- Justificación: Ofrecer al usuario poder contrastar información de Wikidta en Wikipedia sin salir de la interfaz.
- Precondición: Uso de la API de Wikipedia.
- Descripción: Crear una página con un buscador de información en Wikipedia que liste los resultados.

RF 10: Contador de resultados

Requisito	RF 10
Nombre requisito	Contador de resultados
Descripción	Mostrar número de items de resultados
Prioridad	Media

Tabla 6.10.: RF 10 - Contador de resultados

- Justificación: Ofrecer información adicional al usuario en sus búsquedas.
- Precondición: Realizar una búsqueda.
- Descripción: Añadir el número de resultados de una consulta en el buscador y mostrar por pantalla.

RF 11: Reutilización de herramientas RDF

Requisito	RF 11
Nombre requisito	Reutilización de herramientas RDF LOD
Descripción	Histropedia para crear cronologías
Prioridad	Media

Tabla 6.11.: RF 11 - Reutilización de herramientas RDF de terceros

- Justificación: Reutilizar herramientas libres para enriquecer el visor RDF.
- Precondición: Crear consultas SPARQL.
- Descripción: Enriquecer el visor con Histropedia para generar cronologías a partir de consultas SPARQL.

RF 12: Buscar autor

Requisito	RF 12
Nombre requisito	Buscar autor
Descripción	Buscador de autores
Prioridad	Alta

Tabla 6.12.: RF 12 - Buscar autor

- Justificación: Localizar un autor determinado.
- Precondición: Acceder a página de autores.
- Descripción: Ofrecer un campo de texto para poder localizar un autor por nombre o apellido.

RF 13: Buscar libro

Requisito	RF 13
Nombre requisito	Buscar libro
Descripción	Buscador de libros por título
Prioridad	Alta

Tabla 6.13.: RF 13 - Buscar libro

- Justificación: Localizar un libro determinado.
- Precondición: Acceder a página libros.
- Descripción: Ofrecer un campo de texto para poder localizar un libro por título o autor.

RF 14: Cargar archivo XML, JSON o RDF

Requisito	RF 14
Nombre requisito	Cargar archivo XML, JSON o RDF
Descripción	Carga de ficheros RDF para visualizar
Prioridad	Baja

Tabla 6.14.: RF 14 - Cargar archivo XML, JSON o RDF

- Justificación: Cargar archivos para visualizarlos.
- Precondición: Disponer de un archivo XML, RDF o JSON.
- Descripción: Ofrecer un campo para subir un archivo XML, RDF o JSON. Debe mostrar los datos básicos del archivo seleccionado como nombre, tamaño y fecha de actualización.

RF 15: Visualizar archivo XML, JSON y RDF

Requisito	RF 15
Nombre requisito	Visualizar archivo XML, JSON y RDF
Descripción	Leer y visualizar ficheros RDF, XML y JSON
Prioridad	Baja

Tabla 6.15.: RF 15 - Visualizar archivo XML, JSON y RDF

- Justificación: Visualizar archivos RDF.
- Precondición: RF14.
- Descripción: Visualizar de forma legible información de archivos XML, RDF o JSON.

RF 16: Acceder a datos BVMC

Requisito	RF 16
Nombre requisito	Acceso datos BVMC
Descripción	Opción de recuperar datos de la BVMC
Prioridad	Alta

Tabla 6.16.: RF 16 - Acceder a datos BVMC

- Justificación: Mostrar datos de diferentes fuentes. Completar la información de Wikidata con datos de la BVMC.
- Precondición: Indicar en el *checkbox* que se desea recuperar datos de la BVMC.
- Descripción: De forma opcional, el usuario puede recuperar todos los libros de la BVMC para un autor de Wikidata dado.

6.6. Requisitos No Funcionales

RNF 1: Enlazar a fuentes

Requisito	RNF 1
Nombre requisito	Enlazar a fuentes
Descripción	Añadir enlaces a las fuentes de datos
Prioridad	Alta

Tabla 6.17.: RNF 1 - Enlazar a fuentes

RNF 2: Mostrar texto alternativo en imágenes

Requisito	RNF 2
Nombre requisito	Mostrar texto alternativo en imágenes
Descripción	Añadir etiqueta HTML <i>alt</i> a las imágenes
Prioridad	Media

Tabla 6.18.: RNF 2 - Mostrar texto alternativo en imágenes

RNF 3: Carga asíncrona

Requisito	RNF 3
Nombre requisito	Carga asíncrona
Descripción	Uso de <i>AJAX</i> para cargar datos de forma asíncrona
Prioridad	Alta

Tabla 6.19.: RNF 3 - Carga asíncrona

RNF 4: Reutilizar información de BVMC

Requisito	RNF 4
Nombre requisito	Reutilizar información de BVMC
Descripción	Mostrar datos RDF de la BVMC
Prioridad	Alta

Tabla 6.20.: RNF 4 - Reutilizar información de BVMC

RNF 5: Limitar consultas a Wikipedia

Requisito	RNF 5
Nombre requisito	Limitar consultas a Wikipedia
Descripción	Máximo 200 items por resultado de búsqueda
Prioridad	Alta

Tabla 6.21.: RNF 5 - Limitar consultas a Wikipedia

RNF 6: Abrir enlaces en nueva ventana

Requisito	RNF 6
Nombre requisito	Abrir enlaces en nueva ventana
Descripción	Uso de la etiqueta HTML <i>target</i>
Prioridad	Baja

Tabla 6.22.: RNF 6 - Abrir enlaces en nueva ventana

RNF 7: Realizar peticiones paginadas

Requisito	RNF 7
Nombre requisito	Realizar peticiones paginadas
Descripción	Realizar las consultas SPARQL paginadas en 20 items
Prioridad	Alta

Tabla 6.23.: RNF 7 - Realizar peticiones paginadas

RNF 8: Usar BBDD de Wikidata

Requisito	RNF 8
Nombre requisito	Usar BBDD de Wikidata
Descripción	Todos los datos se extraen de esta base de conocimiento
Prioridad	Alta

Tabla 6.24.: RNF 8 - Usar BBDD de Wikidata

6.7. Casos de Uso

En la figura 6.1 se muestra el diagrama de casos de uso en el que se indican las acciones que puede realizar el usuario. La etiqueta *include* indica que la acción es obligatoria mientras que *extend* significa que es opcional.

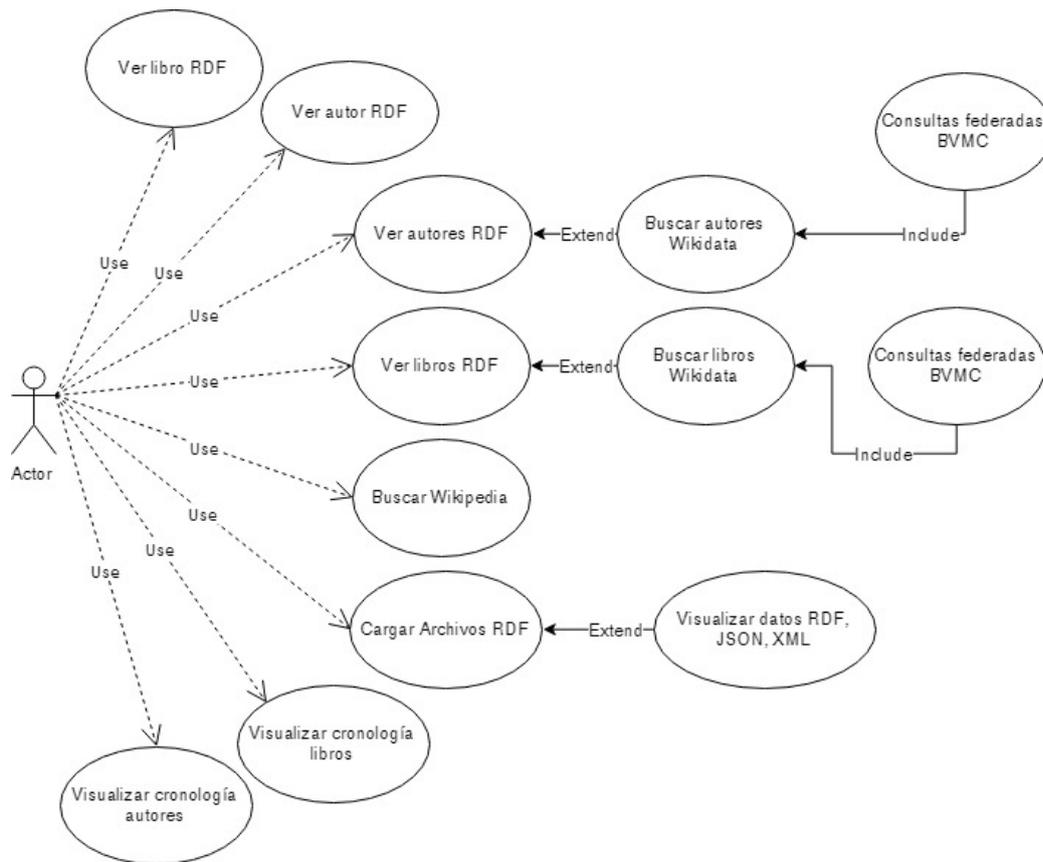


Figura 6.1.: Diagrama de casos de uso.

7. Diseño

En esta sección se detalla la arquitectura del sistema, tecnologías empleadas y diseños de interfaz.

7.1. Arquitectura seleccionada

Como muestra la figura 7.1, la arquitectura se describe principalmente por una comunicación HTTP (*HyperText Transfer Protocol*) mediante AJAX que envía consultas SPARQL sobre los repositorios Wikidata para recuperar datos expresado en RDF. Opcionalmente, el sistema permite recuperar datos de la BVMC a través de Wikidata mediante las consultas federadas.

Los repositorios responden devolviendo un objeto con la información devuelta de la consulta en formato JSON.

Finalmente se leen los datos recuperados y en función de la información leída se aplica una determinada regla de estilo CSS y se inserta en el documento HTML accediendo al DOM.

El sistema no almacena información por lo que no requiere del uso de base de datos propia.

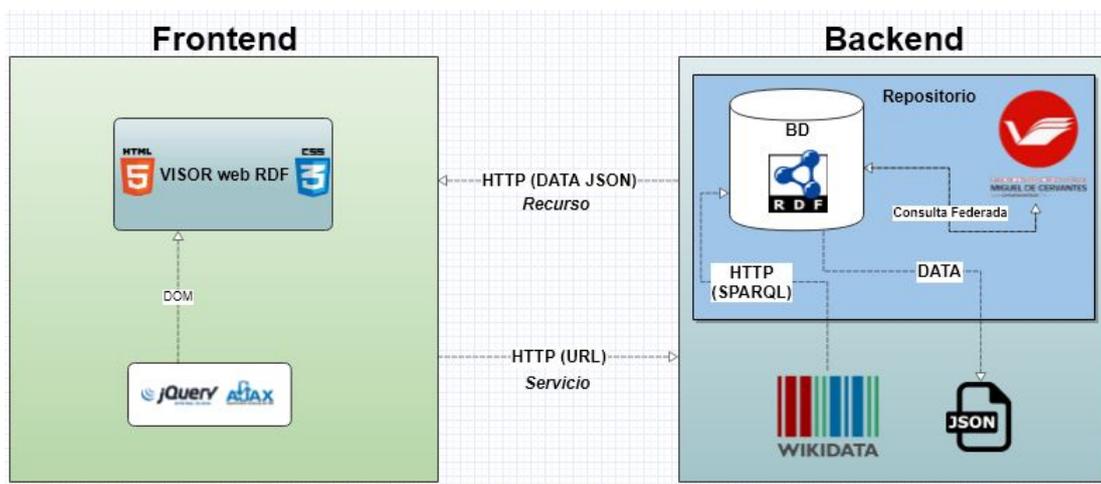


Figura 7.1.: Representación de la arquitectura del visor RDF.

7.2. Tecnologías

Para el desarrollo del proyecto se han utilizado diferentes herramientas, lenguajes de programación y tecnologías. A continuación se describen las más relevantes.

Lenguajes y tecnologías

Los lenguajes principales para el desarrollo del visor RDF han sido HTML, CSS, Javascript y jQuery. La memoria del proyecto se ha realizado en LaTeX con la aplicación web www.overleaf.com.

Se han utilizado las tecnologías HTTP, AJAX y SPARQL para realizar las consultas sobre los datos y CORS¹ para poder habilitar en el navegador la recepción de datos de diferentes dominios.

IDE

Para la creación del código se ha usado el editor de texto ATOM² que permite plugins para facilitar la escritura y depuración de código. Para el control de versiones se ha usado el cliente Git, Gitkraken³.

Control del proyecto

Cronograma y planificación de trabajo con MS Project y Toggle para el control de tiempos de trabajo. Para la organización de la tareas de forma visual se ha usado Trello.

Servidor

El servidor empleado pertenece a Wikidata. Mediante llamadas AJAX se realizan las peticiones desde el visor (cliente) para recuperar la información de las bases de datos de Wikidata.

Herramientas externas

Histropedia es una aplicación Web que funciona como visor de datos RDF. Permite crear timelines o cronogramas a partir de consultas SPARQL. Ofrece la posibilidad de exportar e incrustar en HTML mediante la etiqueta *iframe*.

Para la creación de las consultas de datos SPARQL se ha usado Query Wikidata⁴. Esta aplicación online actúa como punto de acceso a las bases de datos de Wikidata a través de consultas SPARQL para crear conjuntos de datos. Permite descargar la información consultada en diferentes formatos y códigos para usar su API.

Para el diseño de los prototipos de la interfaz se ha usado la aplicación Justinmind⁵.

¹<https://tinyurl.com/omd5o73>

²<https://atom.io/>

³<https://www.gitkraken.com/>

⁴<https://query.wikidata.org/>

⁵<https://www.justinmind.com/>

7.3. Diagrama de clases

Aunque el proyecto no requiere de bases de datos ya que el sistema se nutre de datos de terceros, se ha diseñado el diagrama de clases que muestra como se organiza la información y las relaciones de las entidades que intervienen.

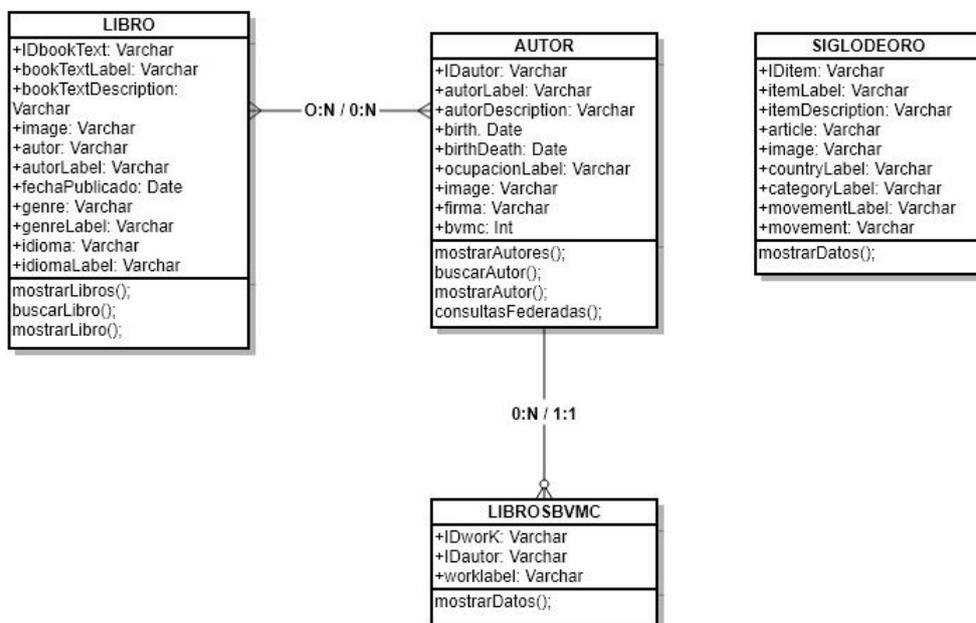


Figura 7.2.: Diagrama de clases.

Como vemos en la figura 7.2 se han creado 4 tablas con sus atributos. Las tablas *AUTOR* y *LIBRO* se identifican inequívocamente por su URI como cadena de caracteres a través de sus claves primarias *IDauthor* y *IDbookText* respectivamente. Los atributos de cada tabla corresponden con la información que se desean mostrar en el visor (propiedades de la entidad). Se genera una tercera tabla de la relación *LIBRO-AUTOR* muchos a muchos, ya que existen libros que pueden no tener autor (libros con autores anónimos o sin datos en la base de conocimiento) o libros escritos por varios autores. Al mismo tiempo, pueden haber autores que tengan muchos libros o ninguno (faltan datos en el repositorio).

La tercera tabla *LIBROSBVMC* se emplea para almacenar los datos de los libros en la BVMC para un autor dado. Los atributos seleccionados son: *idAuthor* como clave primaria y clave ajena a la tabla *AUTOR* (dependencia de identificador). Almacenamos la URI del libro en el atributo *work* y el título en *worklabel*.

La cuarta tabla *SIGLO DE ORO* almacena los datos básicos de la categoría Siglo de

Oro para mostrar en la página principal y no se relaciona con otras tablas.

Las decisiones tomadas para este diseño de diagrama de clases se fundamentan en la inconsistencia de las bases de datos colaborativas y se eligen estos atributos y claves primarias teniendo en cuenta que muchos valores de las propiedades pueden no existir en una entidad.

7.3.1. Diagrama de Bloques

El diagrama de bloques en la figura 7.3 describe el funcionamiento principal del visor RDF. La primera acción se realiza de forma automática y es la carga de datos RDF del Siglo de Oro procedentes de Wikidata. El usuario visualiza datos RFD de libros y autores y dispone de un buscador que realiza peticiones paginadas al servidor de Wikidata. El sistema ofrece la posibilidad de recuperar datos de la BVMC en la misma petición. Finalmente se visualizan datos RDF de libros y autores del Siglo de Oro de la Literatura Española.

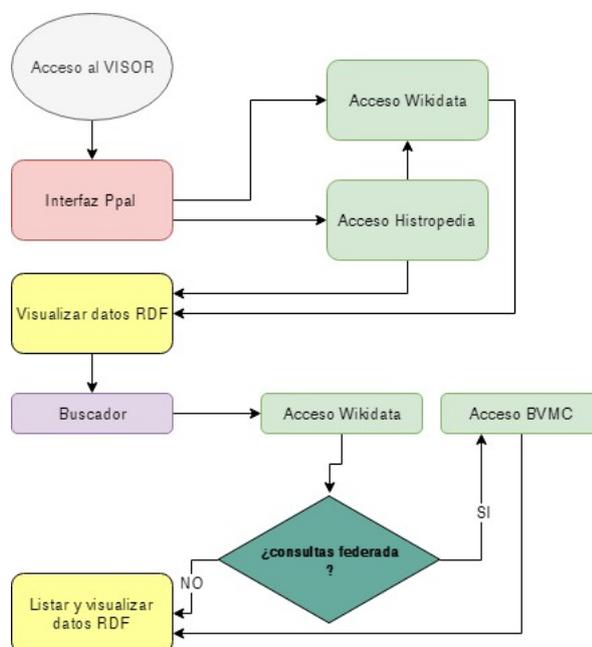


Figura 7.3.: Diagrama de Bloques.

7.4. Mockups

El objetivo principal de este apartado es diseñar una interfaz que permita la búsqueda de autores y libros RDF así como su correcta visualización. Para ello, se presentarán

listas de datos de forma organizada de manera que el usuario pueda tener accesos a toda la información y pueda visualizarla correctamente.

El proceso de diseño comenzó con el estudio de diferentes web's que ofrecen Datos abiertos enlazados así como sus interfaces detalladas en el Anexo B. En particular, el caso de la BVMC ha servido como punto de partida para este diseño. Tras este estudio se elabora el boceto a mano alzada que podemos ver en la figura 7.4. En este primer diseño se esboza el número de interfaces necesarias así como los datos que deben mostrar.

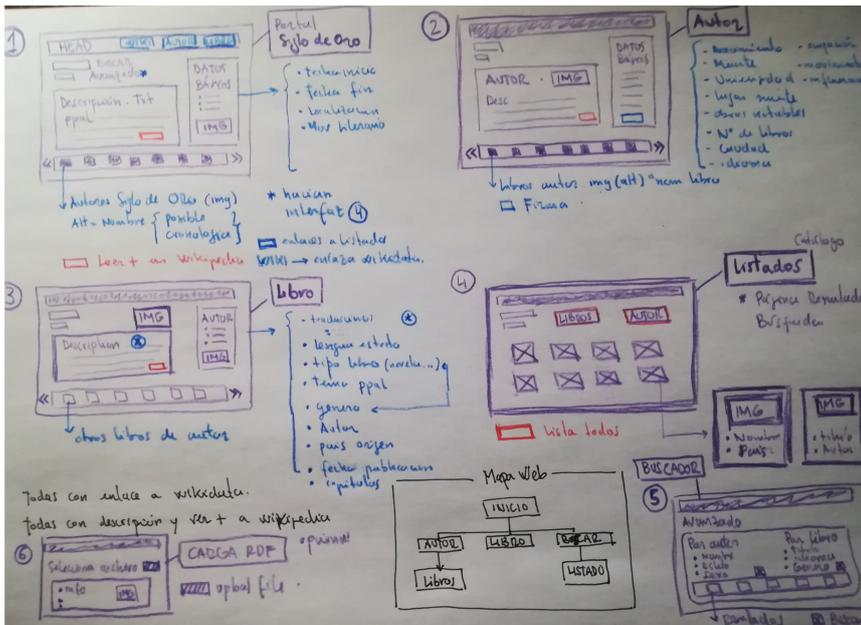


Figura 7.4.: Boceto a mano alzada de la interfaz.

Posteriormente se elaboró el mapa web de la figura 7.5 para diseñar un flujo de navegación sencillo e intuitivo.

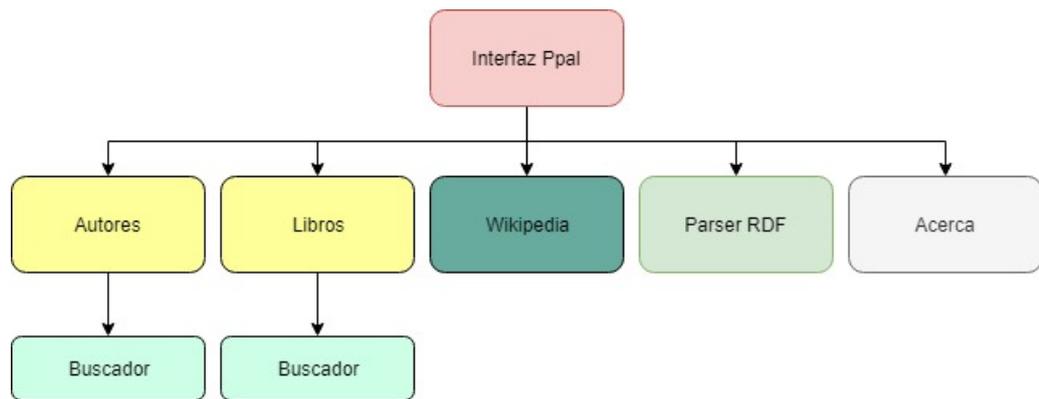


Figura 7.5.: Mapa web.

La elección de este mapa web se justifica en simplificar al máximo los niveles de profundidad web (clicks que hace el usuario hasta llegar al recurso).

Posteriormente estos prototipos se han mejorado con la herramienta Justinmind. Estos se han realizado en escala de grises para fijar toda la atención en la estructura. Para conocer más detalles de los diseños del prototipo ver el Anexo D.

El último paso del diseño de mockups ha sido la creación de la guía de estilos que figura en el Anexo A.

8. Implementación

8.1. Iteración 1: Estudio y planificación

Durante la primera iteración se analizaron las herramientas y tecnologías disponibles para llevar a cabo el proyecto y se realizaron las primeras pruebas para la lectura y extracción de la información de archivos *RDF*, *XML* y *JSON* que se refleja de forma más extendida en el Anexo E.

Para la realización de pruebas se creó la primera interfaz que podemos ver en la figura 8.1.



Figura 8.1.: Primera interfaz para pruebas de lectura de ficheros.

Esta interfaz permite cargar ficheros *XML*, *RDF* o *JSON* y en función de si localiza alguna etiqueta (en el ejemplo, el título de un libro) muestra por pantalla el valor de esa etiqueta y toda la información del archivo en formato texto.

Al mismo tiempo se llevó a cabo un estudio sobre el periodo histórico del Siglo de Oro de la Literatura Española así como sobre el lenguaje de consulta SPARQL para extraer la información de Wikidata. Tanto el estudio del periodo literario como las consultas SPARQL empleadas en el proyecto se pueden ver de forma ampliada en el Anexo C. Posteriormente se diseñaron los bocetos para la interfaz del proyecto y la hoja de estilo

a seguir del Anexo D y en el Anexo A respectivamente.

Los principales problemas abordados en esta iteración han sido la inconsistencia de los datos de Wikidata y la lectura y visualización de formatos RDF.

8.2. Iteración 2: Implementación

En la primera fase de esta iteración se implementaron los mockup's en HTML y CSS así como la mayor parte de los requisitos funcionales y no funcionales. Estos fueron principalmente el buscador, paginación y visualización de datos.

El principal problema en esta iteración ha consistido en gestionar la ausencia de información en algunos campos de los datos recuperados.

8.3. Iteración 3: Mejoras, pruebas y conclusiones

En esta iteración se ha refinado el estilo de las interfaces con CSS. Se ha incorporado a la interfaz la herramienta *Histropedia* para reutilizar las consultas SPARQL y enriquecer así el visor con herramientas ya creadas. Se han implementado las consultas federadas para recuperar datos RDF de diferentes fuentes. Se han realizado las pruebas de validación y redactado las conclusiones en la memoria del proyecto.

9. Pruebas y validación

Dado que el visor de RDF diseñado en este trabajo se plantea como un prototipo, no se han realizado pruebas de rendimiento en el servidor, como por ejemplo, las pruebas de carga. Sin embargo, en su implementación las solicitudes a Wikidata se realizan de forma paginada para evitar la sobrecarga.

Se ha depurado el código a partir de los validadores recomendados por W3C tanto para HTML¹, CSS² y JS³.

Mediante el uso de CSS y Bootstrap⁴ se han diseñado una interfaz *responsive* o adaptable a los principales navegadores y dispositivos móviles.

Responsive en Opera



Figura 9.1.: Imagen del visor RDF en Opera.

¹<https://validator.w3.org/>

²<https://jigsaw.w3.org/css-validator/>

³<https://www.jshint.com/>

⁴<https://getbootstrap.com/>

Responsive en Firefox

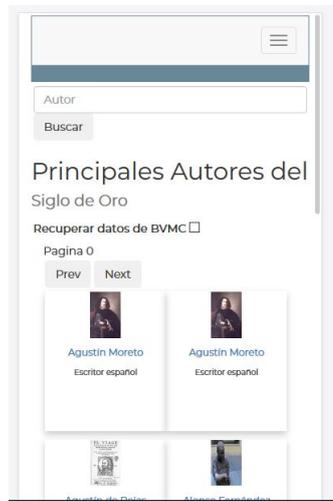


Figura 9.2.: Imagen del visor RDF en Firefox.

Responsive en Edge



Figura 9.3.: Imagen del visor RDF en Microsoft Edge.

Responsive en Chrome



Figura 9.4.: Imagen del visor RDF en Google Chrome.

En las imágenes superiores se puede observar que el aspecto de la interfaz es muy similar para diferentes tamaños de dispositivos móviles y en diferentes navegadores web.

10. Resultados

Los resultados de este trabajo se dividen en el análisis abordado en el capítulo 4 y la implementación del visor RDF. A continuación, se muestra el resultado de las interfaces creadas.

Interfaz principal

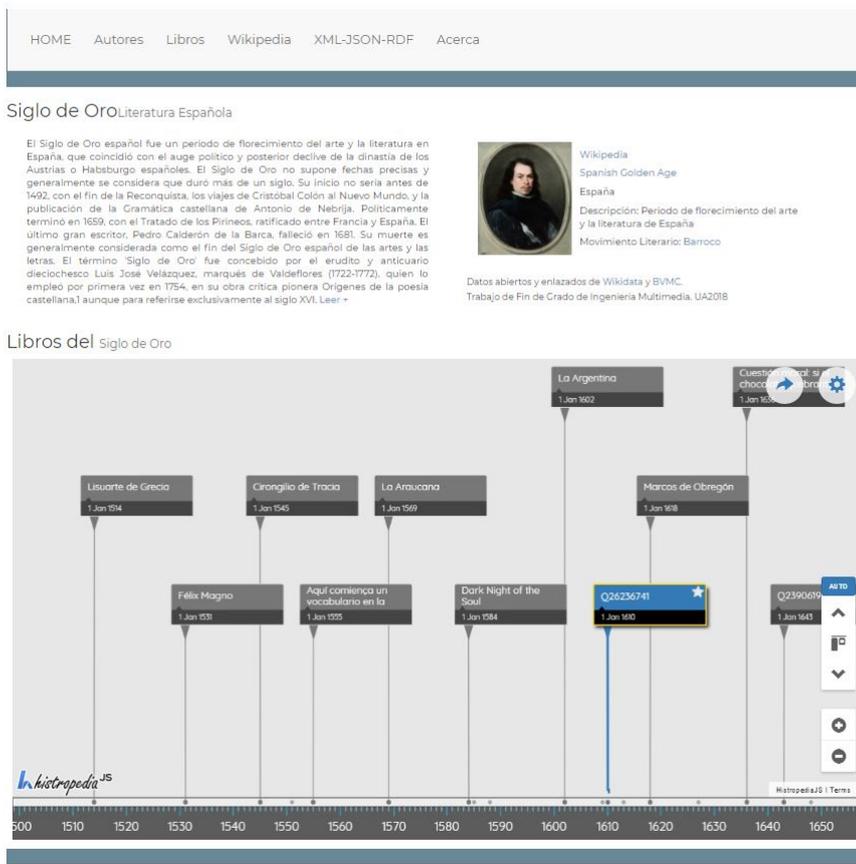


Figura 10.1.: Interfaz principal

La figura 10.1 muestra la interfaz principal. Esta permite visualizar datos RDF de la categoría Siglo de Oro en Wikidata. Incorpora la herramienta Histropedia en donde se muestra la cronología de libros publicados en esta época literaria. En esta interfaz se reflejan los requisitos: RF 1, RF 5, RF 11, RNF 1, RNF 2, RNF 3, RNF 6 y RNF 8.

Interfaz listado autores/libros

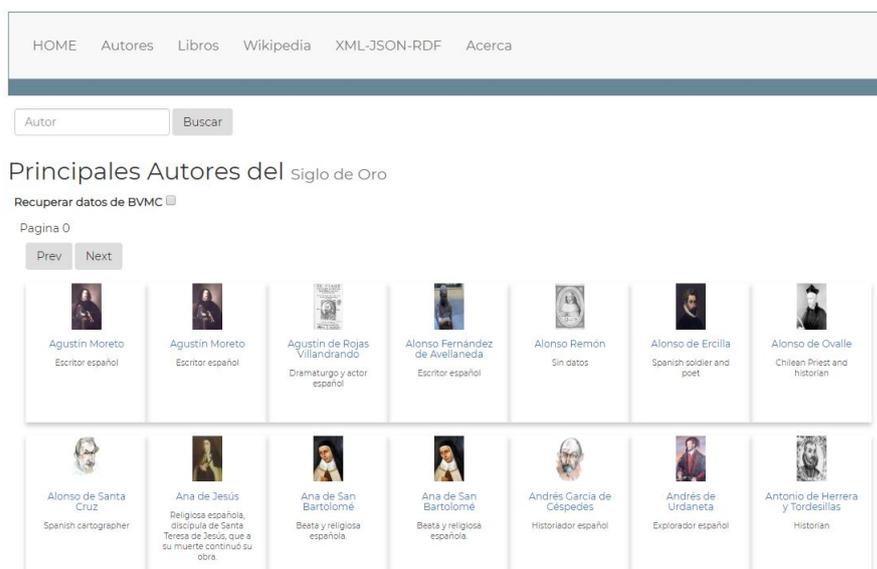


Figura 10.2.: Página listado de autores.

La interfaz de la figura 10.2 permite visualizar listados de autores en RDF. Incorpora un buscador y paginación de resultados. La interfaz para libros comparte el mismo diseño pero con diferentes datos. Se incluye un selector tipo *checkbox* para recuperar datos de la BVMC. Se implementan los requisitos RF 2, RF 3, RF 4, RF 8, RF 12, RF 13, RF 16, RNF 4 y RNF 7.

Interfaz ficha autor/libro

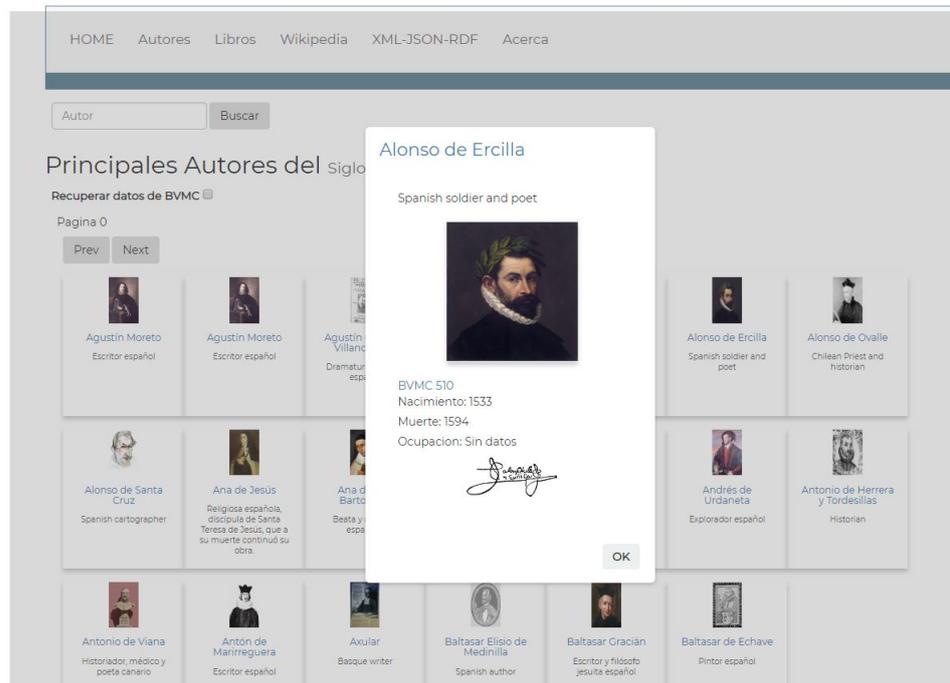


Figura 10.3.: Ventana emergente con información ampliada

En la imagen 10.3 vemos información ampliada del autor o libros. Con respecto al diseño de mockup's se ha decidido implementar esta interfaz con una ventana emergente en lugar de una página. De esta manera se facilita la navegación del usuario. En esta interfaz se reflejan los requisitos RF 6 y RF 7.

Interfaz página datos federados

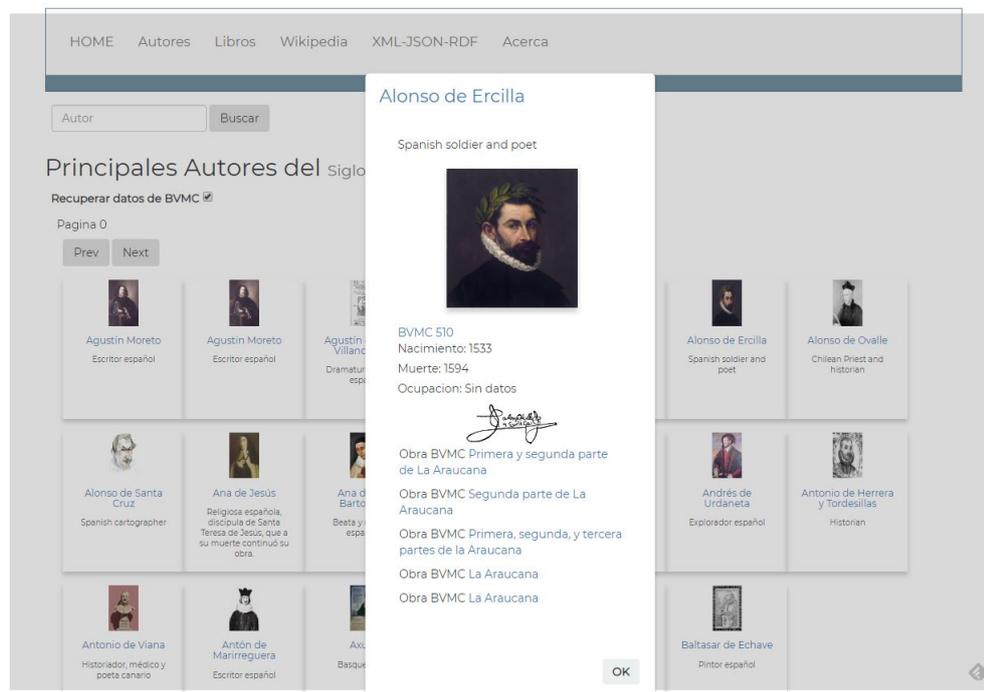


Figura 10.4.: Información ampliada con las consultas federadas

La interfaz de la imagen 10.4 comparte el mismo diseño que la de la figura 10.3 pero incorpora los datos de la BVMC. Para un autor se muestran todos los libros disponibles en la BVMC. Dado que se desconocen el número de items devueltos en las consultas federadas, el diseño de la ventana emergente se adapta al número de resultados recuperados.

Interfaz buscador



Figura 10.5.: Buscador por título y autor

Para el buscador de la figura 10.5 se ha decidido implementar uno para cada ámbito, autores y libros. Esto se debe a que facilita la creación de consultas SPARQL. Permite localizar libros por título y autor y localizar autores por nombre o apellidos. Incorpora paginación y un contador de resultados. Inicialmente se plantearon más filtros para esta interfaz que no se han conseguido implementar y se plantean como mejoras en el capítulo 11.2. Los requisitos implementados en esta interfaz son RF 10 y RNF 9.

Interfaz lector de archivos RDF



Figura 10.6.: Interfaz para pruebas de lectura de ficheros RDF

La interfaz de la figura 10.6 se diseñó para realizar las primeras pruebas de lectura de ficheros como se explica en el apartado 8. Permite cargar archivos RDF, leer su contenido y mostrar por pantalla en formato texto plano. No aplica estilos y se plantea como una tarea futura. Se implementan parcialmente los requisitos RF 14 y RF 15

Interfaz buscador en Wikipedia

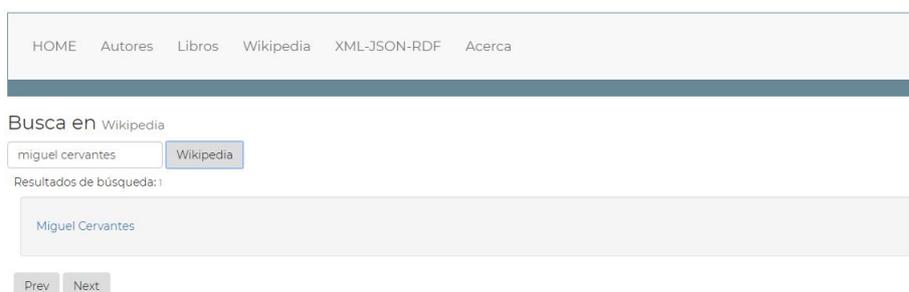


Figura 10.7.: Buscador Wikipedia

Esta interfaz se implementó para pruebas. El buscador de la imagen 10.7 permite localizar recursos en Wikipedia para contrastar la información RDF con la que se está trabajando. No incorpora paginación. Se postpone para tareas futuras. Los requisitos implementados parcialmente son RF 9 y RNF 5.

Interfaz acerca del TFG

HOME Autores Libros Wikipedia XML-JSON-RDF Acerca

Trabajo Fin de Grado Ingeniería Multimedia 2018



Julián Sánchez García
Ingeniero Multimedia
Universidad de Alicante 2013-2018

Relación de asignaturas

- Programación Hipermedia
- Fundamentos de Bases de Datos
- Análisis y Especificación de Sistemas Multimedia
- Programación I y II
- Sistemas de Difusión Multimedia
- Usabilidad y Accesibilidad
- Proyectos Multimedia
- Sistemas Multimedia Avanzados
- Gestión de Contenidos Multimedia
- Diseño de Sistemas Multimedia

Análisis e implementación de una interfaz para visualización de un repositorio RDF

El presente TFG consiste en el análisis de las tecnologías actuales y el estado de los principales repositorios que ofrecen datos abiertos enlazados en formato RDF para la implementación de un prototipo de visor de datos RDF que permita recolectar y visualizar datos abiertos de la base de conocimiento colaborativa Wikidata.

Los datos pertenecen al dominio de Wikidata, en concreto al Siglo de Oro de la Literatura Española. Se pretende recuperar toda la información de este periodo literario. Para ello se recurre a otras fuentes, como la BVMC, que ofrece su catálogo también en RDF y además están federados en Wikidata. Esto quiere decir que Wikidata incluye en su catálogo todos los datos expresados en RDF que ofrece la BVMC, de forma que accediendo a Wikidata, podemos recuperar también información de la BVMC.

El principal problema se encuentra en la dificultad para leer y visualizar datos expresados en RDF así como la inconsistencia de las bases de conocimiento colaborativas.

Para tratar de dar solución a este problema, se ha implementado un prototipo para visualizar datos RDF. Este se ha realizado en HTML, con CSS y jQuery. El proceso consta de las siguientes fases:

- Recolectión de datos de Wikidata mediante Consultas SPARQL
- Conexión AJAX para recuperar la información de los repositorios en JSON
- Implementación de consultas federadas
- Recuperación, lectura y visualización de los datos

El diseño es responsive y se adapta a la mayoría de navegadores y dispositivos móviles

Este proyecto es un prototipo y continúa en fase de desarrollo. Se ha diseñado para explicar las posibilidades en la actualidad para la recuperación y visualización de Datos abiertos y enlazados.

© 2018 UA Ingeniería Multimedia.
TFG Julián Sánchez García

Figura 10.8.: Interfaz acerca del TFG

La interfaz de la figura 10.8 se ha creado para incluir una breve descripción del proyecto así como datos del autor.

En líneas generales el resultado de la implementación del visor RDF ha sido satisfactorio. En cuanto al diseño han habido ligeras modificaciones sobre el prototipo inicial en aras de mejorar la navegabilidad. Aunque los datos recuperados podrían ser más exactos y precisos se ha conseguido el objetivo de facilitar al usuario la recuperación y visualización de datos RDF para el dominio de datos del Siglo de Oro.

El principal problema encontrado ha sido la gestión de datos recuperados en donde las propiedades de una entidad no poseen valor. Para ello se han creado condiciones para mostrar en pantalla diferentes estilos, etiquetas HTML y CSS en función del valor de cada propiedad de la entidad.

Otro de los problemas encontrados son las diferencias entre ontologías y vocabularios en diferentes repositorios. Muchas plataformas u organismos plantean diferentes formas de nombrar a las propiedades de determinados objetos, por ejemplo: un libro puede tener como nombre de propiedad *Book* pero también puede llamarse *Text* según el modelo seguido. Aunque Dublin Core plantea diferentes vocabularios según temática, el uso

de estos para elaborar un modelo de metadatos difiere, por ejemplo, del elegido por Wikidata. Es probable que con el tiempo se usen formatos de metadatos y vocabularios más homogéneos, del mismo modo que ya está ocurriendo con los microformatos RDFa¹. Un ejemplo de esto serían los formatos enriquecidos que Google muestra en sus resultados de búsqueda (recipe, article, item review...), que cada vez más, se ofrecen en diferentes plataformas pero con la misma notación o tipo de datos estructurado, y por consiguiente, se visualizan de la misma manera en todos los navegadores.

Por último, se detallan los tiempos de trabajo registrados en Toggle que se pueden ver en la figura 10.9.

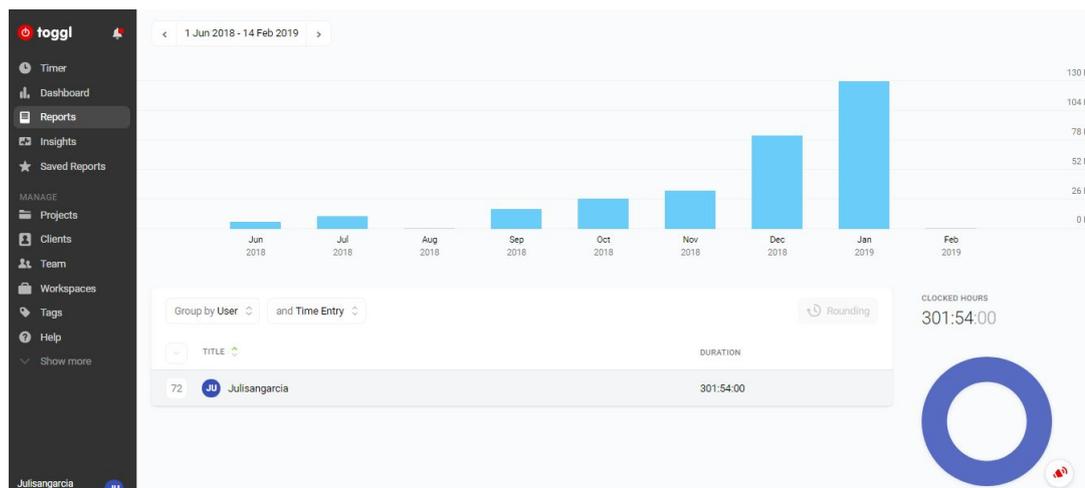


Figura 10.9.: Tiempos de trabajo por meses

Las asignaturas aprendidas durante el Grado en Ingeniería Multimedia que han permitido desarrollar este proyecto han sido:

- Programación Hipermedia I y II
- Fundamentos de Bases de Datos
- Análisis y Especificación de Sistemas Multimedia
- Programación I y II
- Sistemas de Difusión Multimedia
- Usabilidad y Accesibilidad
- Proyectos Multimedia

¹<https://es.wikipedia.org/wiki/RDFa>

- Sistemas Multimedia Avanzados
- Gestión de Contenidos Multimedia
- Diseño de Sistemas Multimedia

11. Conclusiones

11.1. Conclusiones

Se han conseguido alcanzar los principales objetivos planteados en el capítulo 3, analizar el estado actual de repositorios LOD y el uso que se hace de ellos así como el desarrollo de un visor HTML que permita buscar y visualizar datos expresados en RDF de repositorios que ofrecen Datos abiertos.

El resultado final es un prototipo de visor de datos RDF en HTML, adaptable a diferentes dispositivos para un modelo de datos concreto, el Siglo de Oro de la Literatura Española y reutilizando datos de la base de conocimiento Wikidata. Se ha conseguido desarrollar una interfaz que permite al usuario navegar por esta categoría de la literatura española, listando autores y libros del Siglo de Oro.

El prototipo de visor RDF se ha desarrollado con el objetivo de servir de ejemplo para explicar el análisis realizado. Es por ello que en su implementación no se ha buscado el desarrollo de un producto final, más bien se ha pretendido mostrar como integrando diferentes tecnologías y protocolos actuales es posible obtener grandes beneficios con la reutilización de datos.

Se han realizado diferentes estudios acerca de las tecnologías actuales empleadas en la Web Semántica, formatos de publicación de Datos abiertos, repositorios LOD, bibliotecas digitales y cómo muestran estas sus Datos abiertos enlazados. Aunque este análisis se ha realizado de manera general, ha permitido extraer conclusiones previas a la implementación del visor, como la elección del formato de archivo a utilizar o la base de conocimiento de donde se extrae la información entre otras.

Se concluye que aumenta la tendencia en la publicación de datos expresados en RDF y un claro interés en fomentar el método de publicación LOD. Las tecnologías actuales permiten la creación de herramientas para la gestión y visualización de este tipo de datos y, aunque todavía existen problemas inherentes a la Web Semántica, cada día son mayores las posibilidades para la reutilización de información.

En el ámbito de las bibliotecas digitales, el uso de los métodos de publicación RDF y herramientas para la gestión y visualización de este tipo de datos aportaría un gran valor en cuanto a patrimonio cultural, ofreciendo una gran cantidad de información de muchas fuentes de datos sobre un tema concreto e interconectando sus repositorios con otras bases de conocimiento.

Las conclusiones personales acerca del trabajo son satisfactorias. A partir del trabajo inicial de investigación se ha conseguido entender la problemática acerca del método de publicación RDF y LOD. Se ha realizado un estudio del estado actual de las tecnologías, repositorios LOD y bibliotecas digitales que han permitido establecer el punto de partida para la implementación de visor RDF. Para ello, se ha creado una metodología que ha

servido para abordar la implementación del prototipo que finalmente se ha terminado con la mayoría de requisitos funcionales planteados inicialmente.

Algunos objetivos específicos se han abordado de forma parcial como el buscador RDF y Wikipedia, el lector de archivos RDF y la descarga de los datos en JSON que se abordarán en líneas futuras de trabajo.

11.2. Líneas de trabajo futuras

Dada la constante actualización y crecimiento de las bases de conocimiento así como tendencia creciente al uso de RDF para compartir datos, cabe esperar mejoras en cuanto a la integridad de las bases de conocimiento y al aumento Datos abiertos expresados en RDF. Por ello es recomendable la actualización del visor para la depuración de la información, mejorar la exactitud de los datos mostrados o adaptarlo para poder visualizar datos RDF de diferentes dominios. A continuación se plantean futuras líneas de trabajo.

Depurar las consultas SPARQL

Dada la complejidad del lenguaje SPARQL y la inconsistencia de los datos recuperados, las consultas creadas han sido sobre autores y libros únicamente, sobre entidades muy concretas. Con ellas no se ha conseguido extraer con exactitud todas las referencias por lo que este apartado es mejorable. Una posible estrategia consistiría en extraer datos de diferentes categorías y usarlos junto a los actuales. Algunas de estas categorías en Wikidata son:

- Literatura del Siglo de Oro (Q9596300)
- Escritores de España del Silo de Oro (Q9631693)
- Dramaturgos de España del siglo XVI (Q6223260)
- Escritores en español del siglo XVI (Q6232242)
- Escritores en latín del siglo XVI (Q8100449)
- Escritores en español del siglo XV (Q6232241)
- Novelistas de España del siglo XVI (Q6285989)

Por otro lado se pueden completar las consultas con información de otras bibliotecas federadas en Wikidata (BNE y BNF).

Consultas Prototipo

Analizando el número de propiedades de cada entidad recuperada en un archivo RDF, podemos establecer pautas para mostrar la información. Según el número de propiedades que posean valor o no, se pueden establecer diferentes reglas CSS para mostrar los datos. Para ello, debe crearse un buscador avanzado con todas las opciones disponibles

(propiedades para entidades en Wikidata) de tal modo que según los parámetros introducidos se tradujese en una consulta SPARQL. De esta forma, el usuario podría listar información de diferentes dominios de datos. Por ejemplo, un primer campo podría ser *tipo de instancia*, como un libro, un ser humano, un edificio, etc.

Analíticas de uso y visitas

Vincular el visor RDF a plataformas de analítica web. Este elemento permitiría monitorizar el uso del visor RDF y analizando comportamientos como el número de consultas y el término elegido que han arrojado 0 resultados, permitiría conocer que datos se podrían incorporar.

Consultas parametrizadas

Consiste en ofrecer al usuario la opción de recuperar la información en función de una condición. Por ejemplo, recuperar todos los libros de Wikidata que estén en la BVMC para un autor dado y para un *género literario* determinado en BVMC. En este ejemplo, *género literario* sería el parámetro, que podría ser sustituido por rango de fechas, editor o publicador entre otros.

Preprocesado de imágenes

En ocasiones, las imágenes como la firma de autor o portada del libro en Wikidata están disponibles en formatos *svg* o *tiff* que no son manipulables en HTML mediante la etiqueta *img*. Por lo tanto necesitan de un cambio de formato para poder incrustar estas imágenes en HTML. Además, aplicando técnicas de preprocesado las imágenes podrían comprimirse antes de mostrarse mejorando así el rendimiento de carga y velocidad web.

Mejoras en el buscador RDF

Añadir nuevos parámetros de búsqueda ofreciendo al usuario filtros que permitan seleccionar datos más concretos como rango de fechas, género literario, etc.

Mejoras en el buscador Wikipedia

Actualmente el buscador de Wikipedia está limitado a 200 artículos. Añadir nuevos parámetros como el número de items de respuesta o implementar la paginación de resultados mejorarían esta funcionalidad.

Despliegue en producción

Este visor se ha diseñado como prototipo para respaldar el análisis del modelo RDF y su utilidad en la bibliotecas digitales. En el caso de poner el visor en producción se debe contratar un servidor y dominio o bien optar por alojar el visor en Toolforge ¹, un

¹<https://tools.wmflabs.org/admin/tools>

entorno para el desarrollo de proyectos dentro del movimiento Wikimedia. Para habilitar las peticiones de datos de diferentes dominios (los navegadores bloquean estas peticiones por seguridad) es necesario instalar y habilitar CORS² en el servidor.

Multidioma

Dado que Wikidata ofrece sus recursos en diferentes idiomas a partir de la etiqueta *language*, resultaría sencillo implementar la visualización de la información en diferentes idiomas.

²https://developer.mozilla.org/es/docs/Web/HTTP/Access_control_CORS

12. Lista de acrónimos

- TFG: Trabajo Fin de Grado
- RDF: Resource Description Framework
- LOD: Linked Open Data
- LD: Linked Data
- SPARQL: SPARQL Protocol and RDF Query Language
- SQL: Structured Query Language
- OD: Open Data
- XML: Extensible Markup Language
- Turtle: Tarse RDF Triple Language
- N-Triples: Texto plano formateado para la representación RDF
- N3: Forma abreviada de serialización no-XML de modelos RDF
- JSON-LD: Javascript Object Notation for Linked Data
- JSON: Javascript Object Notation
- URI: Uniform Resource Identifier
- DOM: Document Object Model
- AJAX: Asynchronous JavaScript And XML
- RN: Requisito Funcional
- RNF: Requisito No Funcional
- UML: Unified Modeling Language
- HTTP: Hypertext Transfer Protocol
- API: Application Programming Interface

Bibliografía

- [Alexander Osterwalder, 2010] Alexander Osterwalder, Y. P. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley and Sons Inc. Hoboken, New Jersey. ISBN: 978-0470-87641-1.
- [BNF, 2014] BNF, editor (2014). *We grew up together: data.bnf.fr from the BnF and Logilab perspectives*, Paris, Bibliothèque nationale de France, Petit auditorium. IFLA Information Technology Section ; IFLA Semantic Web Special Interest Group ; Bibliothèque nationale de France, IFLA Information Technology Section ; IFLA Semantic Web Special Interest Group ; Bibliothèque nationale de France.
- [Candela et al., 2017] Candela, G., Escobar, P., Carrasco, R., and Marco, M. (2017). Migration of a library catalogue into RDA linked open data. *Semantic Web Journal*, Preprint(Preprint):1–11.
- [Group, 2014] Group, R. W. (2014). Rdf.
- [Lee, 2009] Lee, T. B. (2009). Linked data.
- [Muñoz, 2014] Muñoz, R. S. (2014). Launching of beta version of datos.bne.es, a LOD service and a FRBR-based catalogue view. *SCATNews*, 1(42):13–21.
- [Pellissier Tanon et al., 2016] Pellissier Tanon, T., Vrandečić, D., Schaffert, S., Steiner, T., and Pintscher, L. (2016). From freebase to wikidata: The great migration. In *Proceedings of the 25th International Conference on World Wide Web, WWW '16*, pages 1419–1428, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- [Sulé, 2015] Sulé, A.; Centelles, M. F. J. G. J. (2015). Aplicación del modelo de datos rdf en las colecciones digitales de bibliotecas, archivos y museos de españa. *Revista Española de Documentación Científica*, 39(1):121. <http://redc.revistas.csic.es/index.php/redc/article/view/924/1340>.
- [Tomás Saorín, 2018] Tomás Saorín, J. A. P. S. (2018). Wikidata y dbpedia: viaje al centro de la web de datos. *IweTel*, 12:207–214.
- [W3C, 2008] W3C (2008). Sparql query language for rdf. Web. <https://www.w3.org/TR/rdf-sparql-query/>.
- [W3C, 2009a] W3C (2009a). Best practices for publishing linked data. Web.
- [W3C, 2009b] W3C (2009b). Owl.

[W3schools, 2019] W3schools (2019). Json vs xml.

[Web, 2012] Web, S. (2012). Semantic web.

[Wikidata, 2017] Wikidata (2017). SPARQL federation input/Archive. https://www.wikidata.org/wiki/Wikidata:SPARQL_federation_input/Archive. [Online; accessed 10-July-2018].

[Wikipedia, 2018] Wikipedia (2018). Base de conocimiento. Wikipedia, 27 junio de 2007.

A. Anexo I. Guía de estilo

Guía de estilos Visor RDF

Trabajo Fin de Grado

Ingeniería **Multimedia**

Universidad de Alicante 2018

Autor Julián Sánchez García

Introducción

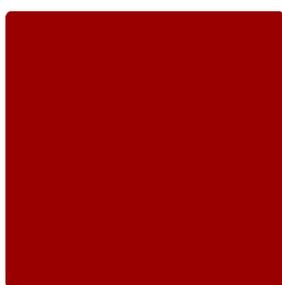
Este documento refleja la guía de estilo del proyecto. Tiene como objetivo principal establecer un marco de actuación para el diseño de una interfaz para la visualización de datos RDF.

Para ello se establecen una guía para el desarrollo coherente y comprensible que permita una ejecución correcta del estilo. De esta forma, la presente guía brinda un conjunto de pautas que pueden ir siendo modificadas en el tiempo hasta obtener el estilo deseado en las distintas versiones de este documento.

Estas pautas permitirán asegurar la coherencia del estilo del producto final.

Paleta **cromática**

Colores primarios



#990000

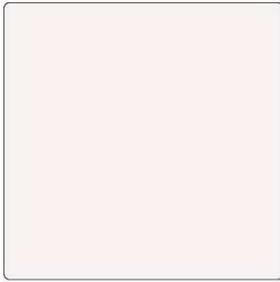


#339966



#006699

Fondos y Bordes



#f6f0f0



#7b988a



#688797

Texto

\$normal-text
#333333

\$mid-text #666666

\$subtle-text #999999

\$ghost-text #999999

\$error-text #d32323

\$elite-text #f15c00

\$bold-text #333333

\$link-color #0073bb

\$green #41a700

Tipografías

La tipografía seleccionada proviene de **Google Fonts** que tiene como principal ventaja que la fuente siempre se muestra.

Thin
Thin Italic
Extra-Light
Extra-Light Italic
Light
Light Italic
Regular
Regular Italic
Medium
Medium Italic
Semi-Bold
Semi-Bold Italic
Bold
Bold Italic
Extra-Bold
Extra-Bold Italic
Black
Black Italic

Sin embargo, **@font-face** es posible que en algunas versiones de ciertos navegadores no esté disponible.

La principal desventaja es que añades más peticiones HTTP a la carga de la página lo que se traduce en una carga más lenta.

Otra desventaja de Google Fonts es que puede haber una demora en mostrarse la fuente, lo que implica que primero se verá la fuente por defecto.

Se ha optado por usar la fuente Montserrat:

<https://fonts.google.com/specimen/Montserrat>

- Raleway +

- Montserrat Regular ▼
- Raleway Regular ▼

- Open Sans +

- Roboto +

- Oswald +

- Lato +

The spectacle before us was indeed sublime.

Apparently we had reached a great height in the atmosphere, for the sky was a dead black, and the stars had ceased to twinkle. By the same illusion which lifts the horizon of the sea to the level of the spectator on a hillside, the sable cloud beneath was dished out, and the car seemed to float in the middle of an immense dark sphere, whose upper half was strewn with silver. Looking down into the dark gulf below, I could see a ruddy light streaming through a rift in the clouds.

H1	30/36px
H2	21/27px
H3	16/21px
H4	14/18px
Texto	14/18px
link	14/18px
Small	12/16px

Botones y elementos de navegación

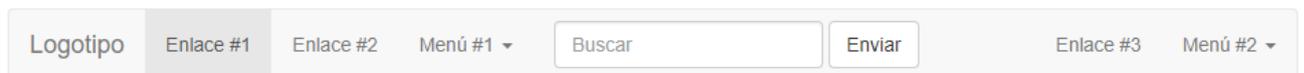
Forma



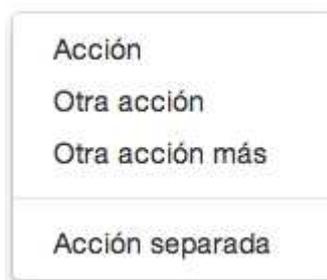
Tamaño



Navegación



Desplegable

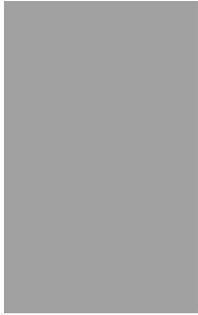


Imágenes

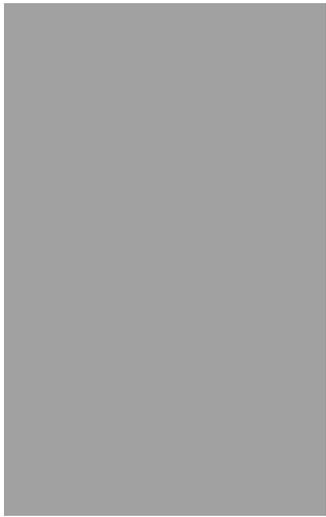




255 x 255



300 x 480



500 x 800



800 x 1280

B. Anexo II. Estudio de Interfaces

Estudio de visores RDF

En este anexo se muestran las principales ventajas y desventajas en los diseños de interfaces estudiados. Muchos de los repositorios analizados muestran sus datos en RDF pero existen muchas diferencias en las herramientas que ofrecen para visualizar la información. Inicialmente se estudian las destacables bibliotecas digitales.

BVMC

Biblioteca Virtual Miguel de Cervantes.

<http://www.cervantesvirtual.com/>

Este portal hace uso de datos abiertos de Wikidata y los ofrece como datos enlazados.

The screenshot shows the BVMC website interface for Miguel de Cervantes Saavedra (1547-1616). It features a search bar at the top, a profile section with a portrait and a table of personal data (Nombre, Fecha de nacimiento, etc.), and a section for his works. A sidebar on the right lists external links such as Wikidata, Wikipedia, and VIAF.

En este ejemplo, la BVMC ofrece datos enlazados desde Wikidata además de información propia.

Miguel de Cervantes / dirección Florencio Sevilla Arroyo

Más información



Título: Miguel de Cervantes / dirección Florencio Sevilla Arroyo
Fecha de distribución: 2002
Idioma: español
Director: Sevilla Arroyo, Florencio
Materia: Cervantes Saavedra, Miguel de, 1547-1616
Enlace permanente: <http://www.cervantesvirtual.com/nd/ark:/59851/bmcbg2n3>

[Ver catálogo](#)

[Ver en origen](#)

Obras relacionadas

[Ver las 1141 publicaciones >](#)



IV Centenario de Don Quijote de la Mancha. Contenidos. Colección Cervantina



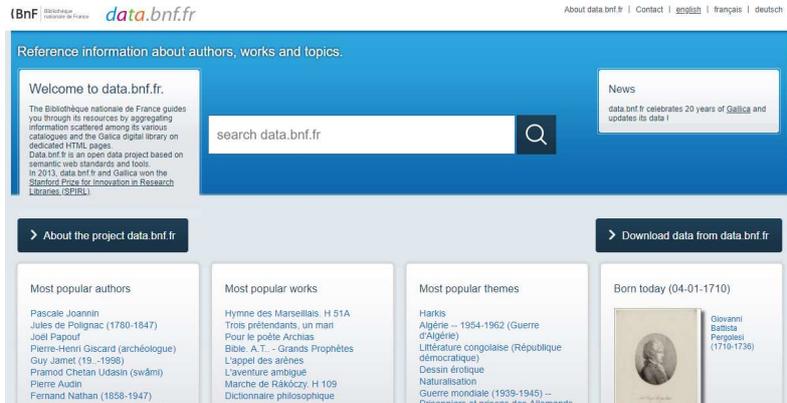
El ingenioso hidalgo Don Quijote de la Mancha. Primera parte. Capítulo I / Miguel de Cervantes Saavedra
Edición digital basada en la edición de Madrid, Ediciones de La Lectura, 1911-1913
Grabaciones sonoras

En líneas generales, muestra información completa. No permite más parámetros de búsqueda que texto. No permite navegación fluida, de autores a categorías y libros en la misma página. Requiere mucho del uso de volver atrás en el navegador. No permite ver el número total de documentos. No ofrece punto de acceso a datos, se hace a través de query.wikidata y su buscador.

BNF Biblioteca Nacional Francesa

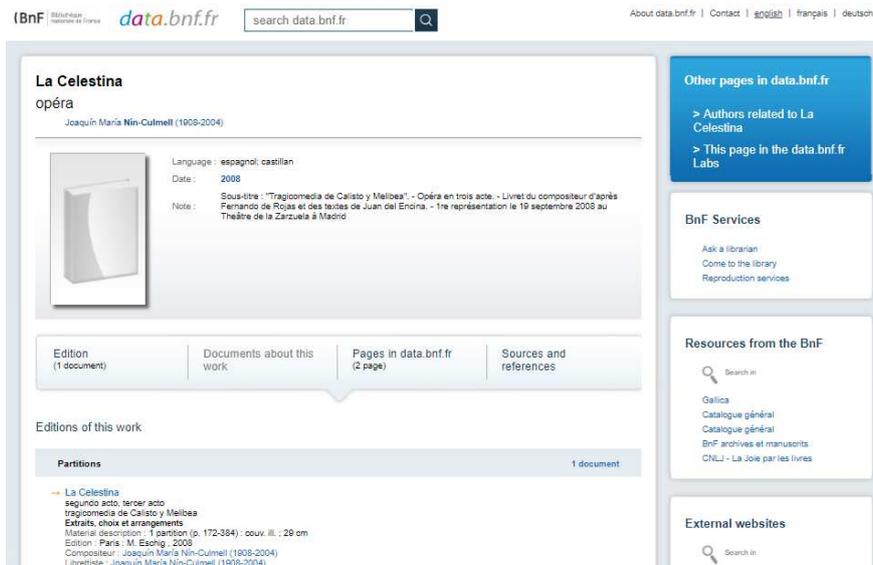
https://catalogue.bnf.fr/index.do

Catálogo general de la Biblioteca Nacional Francesa. Su principal problema es que no permite la descarga de RDF.



The screenshot shows the homepage of data.bnf.fr. At the top, there is a navigation bar with the BnF logo and the text "data.bnf.fr". Below this, a blue banner contains the text "Reference information about authors, works and topics." and a search bar with the placeholder "search data.bnf.fr". To the right of the search bar, there is a "News" section with a small article snippet. Below the banner, there are two buttons: "About the project data.bnf.fr" and "Download data from data.bnf.fr". The main content area is divided into four columns: "Most popular authors" (listing Pascale Joannin, Jules de Polignac, etc.), "Most popular works" (listing Hymne des Marseillais, etc.), "Most popular themes" (listing Harkis, Algérie, etc.), and "Born today (04-01-1710)" (featuring a portrait of Giovanni Battista Pergolesi).

No ofrece un buscador avanzado ni filtros. No permite la navegación con información relacionada.



The screenshot shows a detailed record for "La Celestina" on data.bnf.fr. The record includes the title "La Celestina", the genre "opéra", and the composer "Joaquín María Nin-Culmell (1908-2004)". There is a small image of a book cover. The record also lists the language as "español, castillan", the date as "2008", and a note about the work. Below the main information, there are four tabs: "Edition (1 document)", "Documents about this work", "Pages in data.bnf.fr (2 page)", and "Sources and references". The "Edition" tab is active, showing a list of partitions with one document listed. To the right of the main record, there are several sidebars: "Other pages in data.bnf.fr" with links to authors and this page in the data.bnf.fr Labs; "BnF Services" with links to ask a librarian, come to the library, and reproduction services; "Resources from the BnF" with a search bar and links to Gallica, Catalogue général, and BnF archives; and "External websites" with another search bar.

Biblioteca Nacional Española

http://datos.bne.es/inicio.html

The screenshot shows the 'Búsqueda avanzada' (Advanced Search) page of the DATOS-BNE-ES database. The page features a navigation menu with 'Inicio', 'Personas', 'Entidades', 'Obras', 'Temas', 'Búsquedas avanzadas', and 'Ayuda'. Below the menu is a search bar with the text '1878. Buscar en palabras sobre el conjunto (estado de acceso)'. The main search area is titled 'Búsqueda avanzada' and includes several input fields: 'Título', 'Autor(es)', 'Materia(s)', 'Colección/Serie', 'Editorial', 'Identificador (ISBN, ISSN, etc.)', 'Intérpretes (cine o música)', 'Lugar de edición', and 'Año de publicación'. There are also dropdown menus for 'Idioma' (set to 'Cualquiera') and 'Tipo de documento' (set to 'Todos'). At the bottom of the form are two buttons: 'Buscar ediciones' and 'Limpiar formulario'. The footer contains contact information for 'DATOS-BNE-ES' and the logo of the Biblioteca Nacional Española.

Ofrece buscador avanzado. Permite listar por categorías. Ofrece información relacionada. Duplicidad de contenidos. Permite descarga en RDF-Turtle. Buen diseño de interfaz y navegación.

The screenshot shows the search results page for 'Obras' (Works). The search bar contains the text 'Buscar el título de una obra o el nombre de su autor. Ej: Cien años de soledad, La Iliada, Ring des Nibelungen, De bello Gallico, Wagner'. Below the search bar is a pagination control showing page 1 of 12. The results are displayed in a grid of three items:

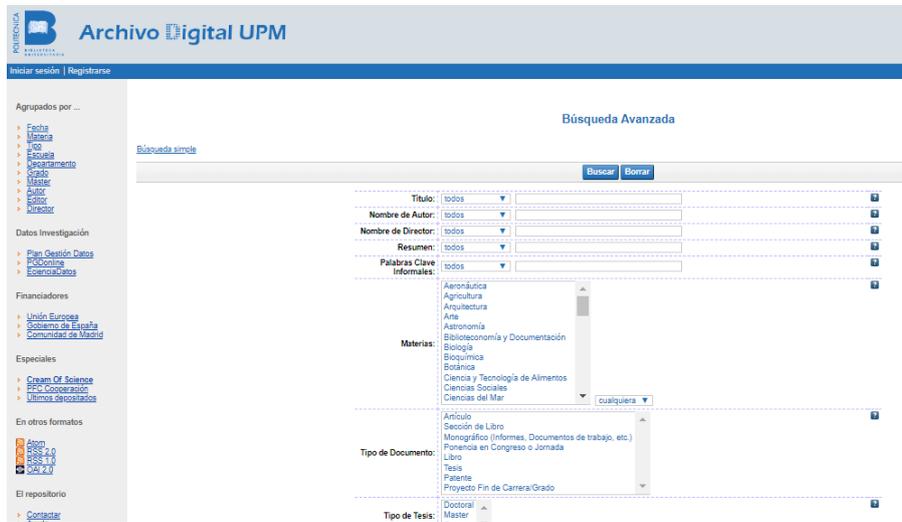
- La petit prince** by Saint-Exupéry, Antoine de (119 ediciones)
- La vida es sueño** by Calderón de la Barca, Pedro (280 ediciones). A 'Disponible en digital' button is visible.
- La Iliada homero** (376 ediciones). A 'Disponible en digital' button is visible.

The screenshot shows the search results page for a specific work. The search bar contains the text 'Buscar una persona, grupo, entidad, el título de una obra o un tema. Ej: Leonardo da Vinci, ONU, Grapes of Wrath, Aztecas'. Below the search bar is a breadcrumb trail: 'Calderón de la Barca, Pedro (1600-1681) / La vida es sueño / Español / La vida es sueño auto sacramental, historial y alegórico. Emp.: Yo soy la negra sombra (h. 1)... Fin.: y grave verdad soñada (h. 23v)'. The main title is 'La vida es sueño auto sacramental, historial y alegórico. Emp.: Yo soy la negra sombra (h. 1)... Fin.: y grave verdad soñada (h. 23v)'. Below the title is a table with the following data:

Título	La vida es sueño auto sacramental, historial y alegórico. Emp.: Yo soy la negra sombra (h. 1)... Fin.: y grave verdad soñada (h. 23v); de Pedro Calderón de la Barca;	Tipo	Manuscrito
Lugar de publicación	S.XVII	Temas	Teatro
Descripción	l. 23 h.		

Biblioteca Digital Politécnica Madrid

<http://oa.upm.es/>



Permite múltiples filtros. Permite descargar en múltiples formatos. Mala visualización de la información. Permite la navegación gracias a su panel lateral.

Portal de datos abiertos Europeo

<http://data.europa.eu/euodp/en/home>

Ofrece diferentes herramientas de visualización gráficas pero no para manejar colecciones. La consulta de datos es mediante Sparql.



Ofrece un buscador y opciones de descarga de ficheros.

Monumental

<https://tools.wmflabs.org/monumental/>

The screenshot shows the Monumental tool interface for the Eiffel Tower. The header includes the tool name 'MONUMENTAL', the item name 'Torre Eiffel', and the user 'JULISANCHEZ1980'. A large image of the Eiffel Tower is displayed on the left. To the right, the title 'Torre Eiffel' is followed by 'tour Eiffel'. Below this, there are several data points: 'VII Distrito de París · París · Isla de Francia · Métropole du Grand Paris · Francia', '1190 images and 20 subcategories', '146 languages', 'registered historic monument', and 'toureffel.paris'. A row of eight small thumbnail images shows various views of the tower. At the bottom, there is a 'Wikipedia' section with language options 'ES', 'EN', and 'FR'.

Herramienta para visualizar datos de Wikidata. No posee buscador avanzado. Buen diseño y cantidad de información. No permite navegación a información similar. No permite descarga de fichero.

Wikidata / Wikipedia

<https://www.wikidata.org/wiki/Q243>

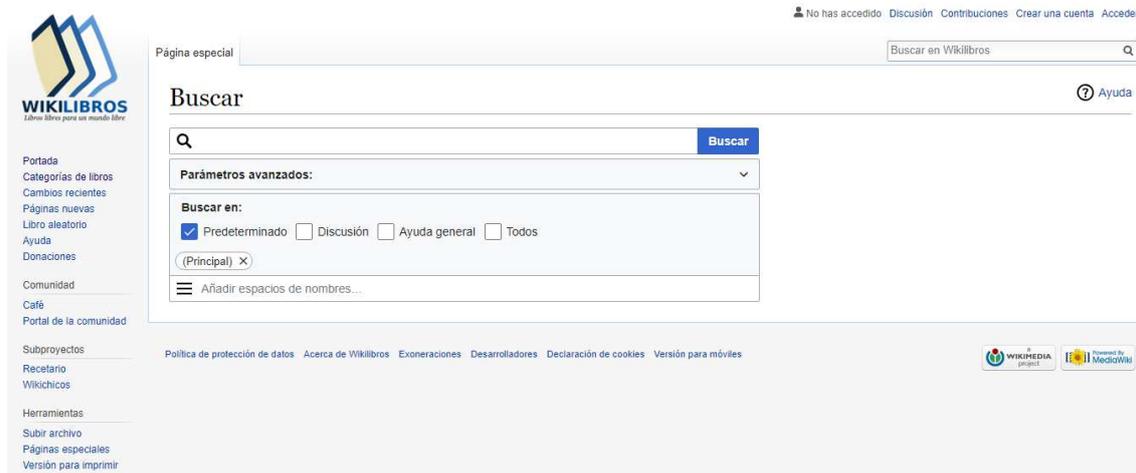
The screenshot shows the Wikidata page for the Eiffel Tower (Q243). The header includes the Wikidata logo, the item name 'Eiffel tower (Q243)', and navigation options like 'Read', 'View history', and 'Search Wikidata'. The main content area displays the item's description: 'tower located on the Champ de Mars in Paris, France' and 'Tour Eiffel | The Eiffel Tower'. Below this, there is a table with columns for 'Language', 'Label', 'Description', and 'Also known as'. The table lists the tower's name and descriptions in English, Spanish, Catalan, and Galician. A 'Statements' section is visible at the bottom, showing 'instance of' with a value of 'lattice tower' and '0 references'.

Language	Label	Description	Also known as
English	Eiffel tower	tower located on the Champ de Mars in Paris, France	Tour Eiffel The Eiffel Tower
Spanish	Torre Eiffel	monumento en París, Francia	Tour Eiffel
Catalan	Torre Eiffel	No description defined	
Galician	Torre Eiffel	torre situada no Campo de Marte en París	Tour Eiffel

Gran cantidad de información difícil de leer. No ofrece buscador avanzado. No permite descarga de fichero. No permite navegación.

Wikilibros

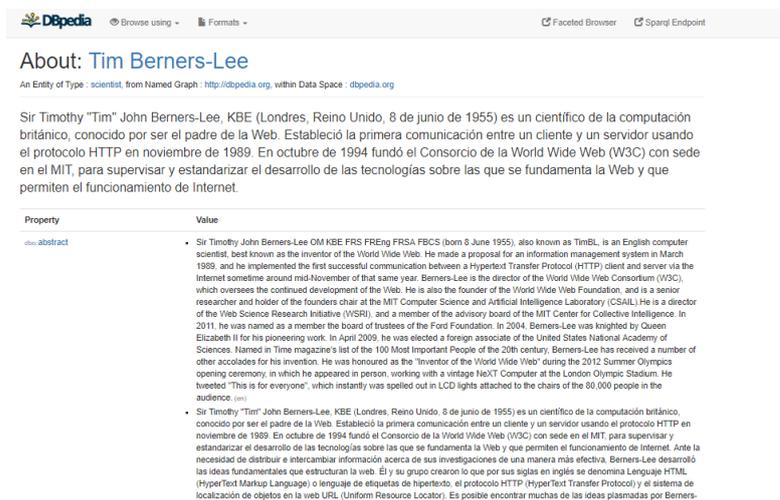
https://es.wikibooks.org/wiki/Portada



No ofrece buscador avanzado. No permite acceso a información relacionada, conjunto de datos de un mismo dominio. Ofrece menos información que Wikidata a pesar de ser del mismo organismo, Wikimedia.

DBpedia

http://dbpedia.org/page/Tim_Berners-Lee



Property	Value
abstract	<ul style="list-style-type: none">Sir Timothy John Berners-Lee OM KBE FRSE FRSE FRCGS (born 8 June 1955), also known as TimBL, is an English computer scientist, best known as the inventor of the World Wide Web. He made a proposal for an information management system in March 1989, and he implemented the first successful communication between a Hypertext Transfer Protocol (HTTP) client and server via the Internet sometime around mid-November of that same year. Berners-Lee is the director of the World Wide Web Consortium (W3C), which oversees the continued development of the Web. He is also the founder of the World Wide Web Foundation, and is a senior researcher and holder of the founders chair at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). He is a director of the Web Science Research Initiative (WSRI), and a member of the advisory board of the MIT Center for Collective Intelligence. In 2011, he was named as a member of the board of trustees of the Ford Foundation. In 2004, Berners-Lee was knighted by Queen Elizabeth II for his pioneering work. In April 2009, he was elected a foreign associate of the United States National Academy of Sciences. Named in Time magazine's list of the 100 Most Important People of the 20th century, Berners-Lee has received a number of other accolades for his invention. He was honoured as the "inventor of the World Wide Web" during the 2012 Summer Olympics opening ceremony, in which he appeared in person, working with a vintage NeXT Computer at the London Olympic Stadium. He tweeted: "This is for everyone", which instantly was spelled out in LCD lights attached to the chairs of the 80,000 people in the audience. ^(en)Sir Timothy "Tim" John Berners-Lee, KBE (Londres, Reino Unido, 8 de junio de 1955) es un científico de la computación británico, conocido por ser el padre de la Web. Estableció la primera comunicación entre un cliente y un servidor usando el protocolo HTTP en noviembre de 1989. En octubre de 1994 fundó el Consorcio de la World Wide Web (W3C) con sede en el MIT, para supervisar y estandarizar el desarrollo de las tecnologías sobre las que se fundamenta la Web y que permiten el funcionamiento de Internet. Ante la necesidad de distribuir e intercambiar información acerca de sus investigaciones de una manera más efectiva, Berners-Lee desarrolló las ideas fundamentales que estructuran la web. Él y su grupo crearon lo que por sus siglas en inglés se denomina Language HTML, (HyperText Markup Language) o lenguaje de etiquetas de hipertexto, el protocolo HTTP (HyperText Transfer Protocol) y el sistema de localización de objetos en la web URL (Uniform Resource Locator). Es posible encontrar muchas de las ideas plasmadas por Berners-

Permite visualizar los datos en diferentes formatos. Gran cantidad de información. Dispone de punto de acceso Sparql. No ofrece buscador avanzado. No permite navegación, no ofrece información relacionada.

<http://dbpedia.org/sparql/>

Buscador limitado. Lista de resultados poco atractiva visualmente. Demasiada información mal organizada que no permite acceso desde resultados de búsqueda.

Conclusiones

En líneas generales, las interfaces que mejor muestran la información son herramientas externas como por ejemplo, Monumental, pero requieren del uso de datos externos y no ofrecen opciones para navegar en la búsqueda de información del mismo dominio. Al mismo tiempo, se observa que los repositorios que alojan información propia disponen de peores interfaces para la visualización que las herramientas externas. También se observa que en las herramientas externas para visualización, los buscadores ofrecen menos posibilidades que los propios repositorios.

En general, los repositorios suelen ofrecer mejores buscadores, opciones de navegación y filtros pero pierden en cuanto a visualización de datos. Ocurre todo lo contrario con las herramientas externas, como Raesonator o Monumental.

C. Anexo III. SPARQL y recuperación de datos del Siglo de Oro

Recuperación de datos Siglo del Oro de la Literatura Española

Consultas SPARQL

Introducción

Para la realización de este trabajo, se han seleccionado datos de Wikidata, una base de conocimiento colaborativa por lo que es muy habitual encontrarnos con una falta de homogeneidad de la información consultada. De esta forma, podemos encontrar autores con diferentes tipos y propiedades. En particular, algunos de ellos aparecen con la propiedad “writer” y otros como “autor”, no siguiendo en muchos casos un mismo patrón.

Consultas SPARQL

Para acceder a los datos del repositorio LOD Wikidata y para la creación de las consultas de los datos que va a mostrar el visor RDF, se ha usado la herramienta <https://query.wikidata.org>. Mediante ella podemos realizar las consultas en el lenguaje *SPARQL* y obtener información para posteriormente mostrarla desde nuestra interfaz.

El funcionamiento de *SPARQL* se asemeja a SQL. Se basa en el uso de etiquetas *namespace* (*espacio de nombres*) y un identificador único.

La estructura de una sentencia Sparql es la siguiente:

```
SELECT *
WHERE { ?s ?p ?o }
LIMIT 10
```

En el **select** podemos emplear variables (elementos que comienzan por **?**) o bien recuperar todo el contenido.

En el **where** introducimos las tripletas RDF para ir filtrando la información. El espacio de nombres **wdt:** se refiere a las propiedades, mientras que en el espacio de nombres **wd:** se refiere a entidades. Los diferentes filtros pueden concatenarse mediante puntos (.).

Ejemplo: lista de todas las mujeres que estudiaron en la universidad de Alicante.

```
SELECT ?person ?personLabel ?LugarNacimientoLabel ?coordenadas
?FechaNacimiento ?image
WHERE {
  ?person wdt:P69 wd:Q2037040 . #han estudiado en la universidad de alicante
  ?person wdt:P21 wd:Q6581072 . #del genero femenino
  ?person wdt:P19 ?LugarNacimiento .
  ?LugarNacimiento wdt:P625 ?coordenadas .
  ?person wdt:P569 ?FechaNacimiento .
```

```

?person wdt:P18 ?image

SERVICE wikibase:label { bd:serviceParam wikibase:language
"[AUTO_LANGUAGE],en,es" }
#Muestra en pantalla las etiquetas según idioma especificado
}

```

Vemos que tiene una estructura similar a las consultas SQL pero en esta ocasión, usamos las etiquetas *wdt* y *wd*, que hacen referencia al predicado WDT “Educado en”, y al objeto WD “Universidad de Alicante”.

Por lo tanto, si analizamos esta consulta como una tripleta de datos RDF, tendríamos:

Sujeto = **?person** Persona (identificador del recurso)
 Predicado = **wdt:P69** “Estudia en” (nombre de la propiedad)
 Objeto = **wd:Q2037040** Universidad de Alicante (valor de la propiedad)

Es muy importante conocer los valores de los *namespace* para cada atributo que queramos recuperar.

Para ello debemos de ver en el recurso de [Wikidata](https://www.wikidata.org), el valor único (Ej: property: P69)



IMG: CAPTURA DE HTTPS://WWW.WIKIDATA.ORG

Debido a que las entidades pueden contener algunas propiedades y otras no, es recomendable utilizar la instrucción **Optional**.

```

select *
where {
  ?x wdt:P31 wd:Q5.
  ?x wdt:P135 wd:Q530936.
  OPTIONAL {?x wdt:P18 ?imagen }
  OPTIONAL {?x wdt:P735 ?givenname }
}

```

Esta sentencia solicita todas las instancias (P31) de seres humano (Q5) que pertenecen al movimiento (P135) del Siglo de Oro (Q530936).

Los datos recuperados muestran sólo 7 autores lo cual es incorrecto, por lo que debemos de estudiar otra manera de recuperar el mayor número de autores de esta época, por ejemplo,

recuperando otras categorías que devuelvan más resultados, autores españoles, autores nacidos en, autores muertos en, autores con descendencia.... Ver capítulo “Líneas de trabajo futuro”.

Mediante la instrucción *describe wd:Q5682*, podemos obtener todas las propiedades de una entidad (ej: Miguel de Cervantes):

```
describe wd: Q5682
```

La sentencia que describe una entidad es la siguiente: <http://tinyurl.com/ycqu5rov>

Estudio de la información con la que se va a trabajar

En este apartado se detalla la información que queremos obtener para mostrar en el visor RDF.

Justificación

La selección de los datos se fundamenta en el interés que se prevé que pueda tener para el usuario. En esta ocasión se trabajará en el dominio del Siglo de Oro de la literatura española. Esta época se ubica en la edad media, que por la distancia y por falta de medios de impresión de la época, los datos pueden ser incompletos. El uso de las consultas federadas puede ayudar a clarificar y verificar la falta de información. Por ejemplo, puede que en Wikidata no encontremos información sobre un autor que si se encuentra en la Biblioteca Virtual Miguel de Cervantes.

La referencia inicial para la búsqueda de la información ha sido el portal del Siglo de Oro de Wikipedia: https://es.wikipedia.org/wiki/Portal:Siglo_de_Oro

Previo selección de datos. Hitos importantes de la época.

El periodo sobre del que se pretenden recuperar datos está comprendido entre 1492 y 1659. Los primeros libros considerados dentro de esta época literaria son “Gramática Castellana” de 1492 y “La Celestina” de 1499. El fin de este movimiento literario se fija en la fecha de la muerte del autor Calderón de la Barca, concretamente, en la fecha de la publicación de su último libro, “Autos Sacramentales” en 1481 (inacabado). Otras referencias de sus últimos libros son “Hado y Divisa de Leonido y Marfisa” o “Andrómeda y Perseo”.

Selección de datos

Para la creación de la consulta de datos, se han estudiado los recursos sobre autores del Siglo de Oro de la Literatura española a partir de los datos de Wikipedia / Wikidata.

En concreto, el punto de partida de investigación son los siguientes enlaces:

- https://es.wikipedia.org/wiki/Portal:Siglo_de_Oro
- https://es.wikipedia.org/wiki/Categor%C3%ADa:Literatura_del_Siglo_de_Oro
- https://es.wikipedia.org/wiki/Categor%C3%ADa:Escritores_de_Espa%C3%B1a_del_Siglo_de_Oro
- https://es.wikipedia.org/wiki/Categor%C3%ADa:Obras_literarias_del_Siglo_de_Oro

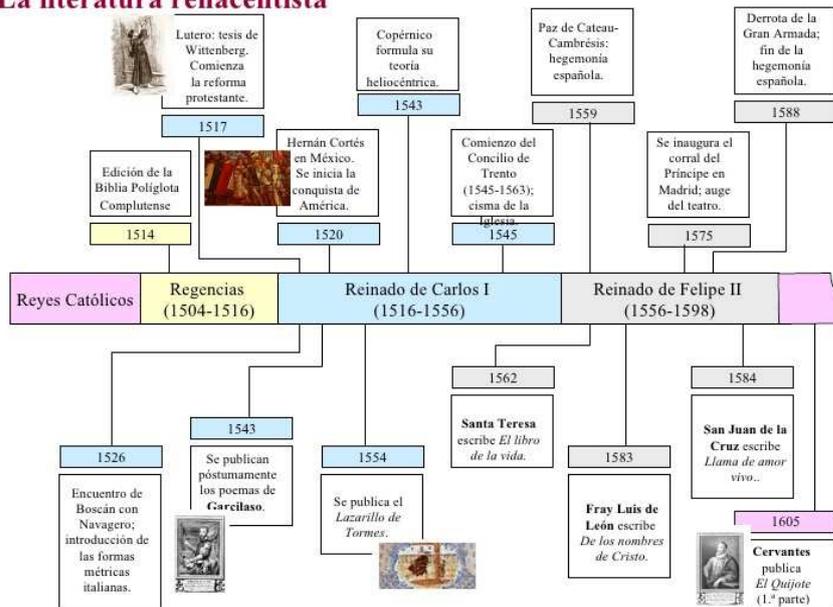
Tras estudiar el modelo de los autores, se han elegido variables para proceder a las consultas. El objetivo es seleccionar el menor número posible de PROPIEDADES de los recursos para recuperar el mayor número de información acerca de los autores del Siglo de Oro. Esto tiene el riesgo de que los filtros no sean los correctos y no aparezcan, por ejemplo, autores de esa época que escribían en otra lengua distinta a la española.

Información de interés

Se han establecido una serie de parámetros para ubicar los datos en el contexto y tiempo. Para ello se tienen en cuenta los siguientes parámetros:

- La celestina, Fernando de Rojas. Considerada la **primera obra** después de la edad media, 1499. **Nota:** muchos de los libros creados no fueron publicados hasta mucho después, por lo tanto, la propiedad “año de publicación” no nos serviría del todo. Deberíamos estudiar otras propiedades.
- Lazarillo de Tormes, Anónimo, 1554. Algunos libros no tienen autor o es anónimo.
- **Última obra** publicada “El cordero de Isaias”, Calderon de la Barca 1681, fecha de muerte del autor y publicación del libro.
- **Boscán.** Aunque no se reconoce como autor, fue el creador teórico de este movimiento literario, adaptando las métricas italianas a la literatura española. Junto a Garcilaso de la Vega son considerados los precursores de este movimiento literario.
- Garcilaso de la Vega, considerado el **primer autor** de esta época.
- El **periodo** comprendido corresponde a casi dos siglos, XV y XVI, durante los Austrias Mayores y Menores, entre **1492 y 1659** aproximadamente.
- **Estilos literarios** de la época: Barroco (culteranismo y conceptismo), Novela (picaresca, religiosa, pastoril y caballeresca), Teatro (religioso y profano) y Poesía (religiosa y profana o popular).
- **Escuelas** representativas del movimiento: La escuela Salmantina y la escuela Sevillana.
- El número de **autores estimados** son 92.
- La primera novela caballeresca inicia este periodo. Las fechas de publicaciones con la temática picaresca española son propias de esta a época.

La literatura renacentista



IMG: deleitedeletras.jimdo.co

Lista de autores estimados: 92

Agustín de Rojas Villandrando
Agustín de Salazar
Alonso de Acevedo (escritor)
Alonso de Bonilla
Alonso de Castillo Solórzano
Alonso de Ledesma
Alonso Fernández de Avellaneda
Alonso Núñez de Reinoso
Amarilis
Ana Caro de Mallén
Ana de Castro Egas
Ana Francisca Abarca de Bolea
Andrés Rey de Artieda
Antonio de Eslava
Antonio de Herrera y Tordesillas
Antonio de Lofraso
Antonio Hurtado de Mendoza
Baltasar del Alcázar
Baltasar Gracián
Bartolomé Jiménez Patón
Bartolomé Leonardo de Argensola
Bartolomé y Lupercio Leonardo de Argensola
Bernardino de Rebolledo
Bernardo de Balbuena
Conde de Villamediana
Constanza Ossorio
Cristóbal de Castillejo
Cristóbal de Mesa
Cristobalina Fernández de Alarcón
Diego de Saavedra Fajardo
Diego de Silva y Mendoza
Diego Hurtado de Mendoza (poeta y diplomático)
Diego Jiménez de Enciso
Diego Ortiz de Zúñiga
Diego Pérez de Valdivia
Feliciano de Silva
Francisco de Aldana
Francisco de Avellaneda
Francisco de Figueroa
Francisco de la Cueva
Francisco de Medrano
Francisco de Quevedo
Francisco de Rojas Zorrilla
Francisco Delicado
Francisco Fernández de Córdoba, abad de Rute
Gabriel Bocángel
Gabriel López Maldonado
Garcilaso de la Vega
Gaspar de Baeza
Gonzalo Fernández de Oviedo
Guillén de Castro
Gutierre de Cetina
Hernando de Acuña
Jacinto Abad de Ayala
Jerónimo de Pasamonte
Juan Bautista Diamante
Juan de la Cruz
Juan de Moncayo
Juan de Salinas y Castro
Juan de Zabaleta
Juan Francisco Andrés de Uztarroz
Juan Luis Vives

Juan Ruiz de Alarcón
Leonor de la Cueva y Silva
Lope de Rueda
Lope de Vega
Lorenzo van der Hamen
Luis Barahona de Soto
Luis de Góngora
Luis de León
Luis de Ulloa
Luis Díez de Aux
Luis Ferrer de Cardona
Luis Quiñones de Benavente
Luis Vélez de Guevara
Lupercio Leonardo de Argensola
María de Guevara
María de Rojas y Garay
María de Zayas
Martín Miguel Navarro
Mateo Alemán
Miguel de Cervantes
Pedro Calderón de la Barca
Pedro Laínez
Pedro Núñez Vela
Pedro Simón Abril
Pedro Valero Díaz
Rodrigo Caro
Salvador Jacinto Polo de Medina
Segimon Arquer
Tirso de Molina
Vicente Espinel

Lista de libros estimados: 114

A Dios por razón de estado
A secreto agravio, secreta venganza
Agudeza y arte de ingenio
Amor, honor y poder
Andrómeda y Perseo
Arauco Domado (Pedro de Oña)
Arte de ingenio, tratado de la agudeza
Arte nuevo de hacer comedias en este tiempo
Atheneo de grandesa
Bernardo (obra perdida de Cervantes)
Brevisima relación de la destrucción de las Indias
Cancionero de Upsala
Casa con dos puertas, mala es de guardar
Cautiverio feliz
Clásicos Castellanos
Códice Boxer
Códice Daza
Crónica de la Nueva España
Descripción de la muy noble y más antigua ciudad de Gibraltar y de su celebrado monte llamado Calpe
Don Gil de las calzas verdes
Don Quijote de la Mancha
Égloga de Juan de Tovar

El alcalde de Zalamea
El alguacil endemoniado
El amante liberal
El burlador de Sevilla y convidado de piedra
El casamiento engañoso
El celoso extremeño
El cerco de Numancia
El chitón de las tarabillas
El coloquio de los perros
El Comulgatorio
El condenado por desconfiado
El Criticón
El curioso impertinente
El diablo cojuelo
El Discreto
El entremés de los romances
El esclavo del demonio
El galán fantasma
El gallardo español
El gran mercado del mundo
El gran teatro del mundo
El Héroe
El José de las mujeres
El juez de los divorcios
El licenciado Vidriera
El mágico prodigioso
El médico de su honra
El mundo por de dentro
El Político
El príncipe constante
El retablo de las maravillas
El rufián dichoso
El siglo pitagórico
El trato de Argel
El vergonzoso en palacio
El viejo celoso
Fábula de Píramo y Tisbe
Fábula de Polifemo y Galatea
Guzmán de Alfarache
Historia general de las Indias
Historia verdadera de la conquista de la Nueva España
Instrucción de música sobre la guitarra española
Itinerario de la armada
La Araucana
La aurora en Copacabana
La cena del rey Baltasar

La cueva de Salamanca
La cuna y la sepultura
La dama duende
La Dorotea
La española inglesa
La Fastiginia
La fiera, el rayo y la piedra
La fuerza de la sangre
La Galatea
La Gatomaquia
La gitanilla
La gran sultana
La ilustre fregona
La Isla de los Monopantos
La Lozana andaluza
La pícara Justina
La puente de Mantible
La señora Cornelia
La vida del Buscón
La vida es sueño
Las mocedades del Cid
Las semanas del jardín
León prodigioso
Los cabellos de Absalón
Los encantos de la culpa
Los trabajos de Persiles y Sigismunda
Marcos de Obregón
Nafragios y comentarios
No hay burlas con el amor
Oráculo manual y arte de prudencia
Pedro de Urdemalas (Cervantes)
Peregrinación de las tres casas sanctas de Jherusalem, Roma y Santiago
Relación breve de la conquista de la Nueva España
Relación de las cosas de Yucatán
Relación de méritos y servicios del conquistador Bernardino Vázquez de Tapia
Rimas humanas y divinas del licenciado Tomé de Burguillos
Rinconete y Cortadillo
Soledades
Soneto a Cristo crucificado
Sueño de la muerte
Sueño del Infierno
Sueño del Juicio Final
Sueños y discursos
Tesoro de la lengua castellana o española
Trilogía de los Pizarros
Viaje del Parnaso

A partir de estos datos se establecen las PROPIEDADES de las entidades de Wikidata que debemos considerar. Estas son las Siguietes:

Propiedades de Autores

PROPIEDADES	CODE	Valor	Code
instance of	p31	human	Q5
sex or gender	p21	male	Q6581097
		female	Q6581072
country of citizen	p27	spain	Q29
given name	p735		
date of birth	p569		
date of death	p570		
languages spoken, written or signed	p1412	spanish	Q1321
		latin	Q397
		valencian	Q32641
ocupation	p106	playwright	Q214917
		dramaturge	Q487596
		writer	Q36180
		poet	Q49757
		novelist	Q6625963
		poetess	Q49757
		lyricist	Q822146
		renaissance humanist	Q1636680
		author	Q482980
bith name	p1477		
given name	p735		
place of birth	p19		
place of death	p20		
cause of death	p509		
floruit	p1317	17. century	
family name	p734		
genre	p106	novel	Q8261
		poetry	Q482
		tragedy	Q80930
		theater	Q11635
		drama	Q25372

		comedy	Q408321
		Eclogue (poema)	Q11072974
		dramaturgy	Q492537
		picareque novel	Q44563
educated at	p69	Salamanca	Q308963
		complutense	Q495225
		granada	Q1232180
		Univ osuna	Q7896031
		Zaragoza	Q633561
		instituto san isidro	Q5917707
		universidad de valencia	Q383568
notable wok	p800		
commons category	p373		
member of	p463	spanish golden age	Q530936
movement	p135	mannerism	Q131808
		spanish golden age	Q530936
		barroquismo	Q1807353
		baroque literature	Q1431376
native language	p103	early modern spanish	Q5364419
influenced by	p737		
work period start	p2031		
work period end	p2032		
time period	p2348	spanish golden age	Q530936
bvmc person ID	p2799		

Propiedades Libros

PROPIEDADES	CODE	Valor	CODE
Instancia	P31	Libro de texto	Q83790
		Libro	Q571
		Novela	Q8261
		Ensayo	Q35760
		Epopeya	Q37484

		Obra Literaria	Q7725634
		Obra de teatro	Q25379
Idioma de la obra	P407	Español	Q1321
		español antiguo	Q1088025
fecha de publicación	P577		
BVMC id work	P3976		

A partir de estas propiedades de Wikidata, se debe de estudiar la mejor estrategia para seleccionar las propiedades que vamos a usar en el Query, con el objetivo de recuperar todos los autores y libros de esta época con el menor número de propiedades usadas en la consulta.

Consultas SPARQL

HOME - DATOS BASICOS EDAD DE ORO

Devuelve información básica de ese periodo literario.

```
SELECT distinct ?item ?itemLabel ?itemDescription ?article ?image ?countryLabel
?categoryLabel ?movementLabel ?movement WHERE{

  ?item ?label "Spanish Golden Age"@en.

  ?article schema:about ?item .

  ?article schema:inLanguage "en" .

  ?item wdt:P18 ?image .

  ?item wdt:P17 ?country .

  ?item wdt:P373 ?category .

  ?item wdt:P31 ?Movement .

  ?article schema:isPartOf <https://en.wikipedia.org/>.

  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }

}
```

TODOS LOS LIBROS

```
SELECT DISTINCT ?bookText ?bookTextLabel ?bookTextDescription ?image ?autor ?autorLabel
?fechaPublicado ?genre ?genreLabel ?idioma ?idiomaLabel

WHERE {

  {?bookText wdt:P31 wd:Q83790 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicadoLabel . } #libro de texto

  UNION
```

```

    {?bookText wdt:P31 wd:Q571 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #libro impreso

UNION

    {?bookText wdt:P31 wd:Q8261 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #novela

UNION

    {?bookText wdt:P31 wd:Q35760 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #ensayo #idioma de la obra spaña

UNION

    {?bookText wdt:P31 wd:Q37484 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #poema épico

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #obra literaria Q5364419

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q5364419 . ?bookText wdt:P577
?fechaPublicado .} #obra literaria Español moderno

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #play

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1088025 . ?bookText wdt:P577
?fechaPublicado .} #play español antiguo

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q397 . ?bookText wdt:P577
?fechaPublicado .} #play latin

OPTIONAL {?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P18 ?image . ?bookText
wdt:P50 ?autor . ?bookText wdt:P136 ?genre . ?bookText wdt:P407 ?idioma . }

#fecha del primer y último libro publicado

filter (?fechaPublicado > "1492-01-01"^^xsd:dateTime && ?fechaPublicado < "1681-05-
26"^^xsd:dateTime)

SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es,en". }
}

ORDER BY ASC(?bookTextLabel)

```

Versión 2

```

SELECT DISTINCT ?bookText ?bookTextLabel ?bookTextDescription ?image ?autor ?autorLabel
?fechaPublicado ?genre ?genreLabel ?idioma ?idiomaLabel ?bvmc

WHERE {

    {?bookText wdt:P31 wd:Q83790 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicadoLabel . } #libro de texto

UNION

```

```

    {?bookText wdt:P31 wd:Q571 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #libro impreso

UNION

    {?bookText wdt:P31 wd:Q8261 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #novela

UNION

    {?bookText wdt:P31 wd:Q35760 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #ensayo #idioma de la obra spaña

UNION

    {?bookText wdt:P31 wd:Q37484 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #poema épico

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #obra literaria Q5364419

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q5364419 . ?bookText wdt:P577
?fechaPublicado .} #obra literaria Español moderno

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #play

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1088025 . ?bookText wdt:P577
?fechaPublicado .} #play español antiguo

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q397 . ?bookText wdt:P577
?fechaPublicado .} #play latin

OPTIONAL {?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P18 ?image . ?bookText
wdt:P50 ?autor . ?bookText wdt:P136 ?genre . ?bookText wdt:P407 ?idioma . ?bookText
wdt:P3976 ?bvmc }

#fecha del primer y último libro publicado

filter (?fechaPublicado > "1492-01-01"^^xsd:dateTime && ?fechaPublicado < "1681-05-
26"^^xsd:dateTime)

SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es,en". }

}

ORDER BY ASC(?bookTextLabel)

```

Versión 3 y final

```

SELECT DISTINCT ?bookText ?bookTextLabel ?bookTextDescription ?image ?autor ?autorLabel
?fechaPublicado ?genre ?genreLabel ?idioma ?idiomaLabel ?bvmc

WHERE {

    {?bookText wdt:P31 wd:Q83790 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicadoLabel . ?bookText wdt:P50 ?autor . } #libro de texto

```

```

UNION

    {?bookText wdt:P31 wd:Q571 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor .?bookText wdt:P18 ?image . } #libro impreso

UNION

    {?bookText wdt:P31 wd:Q8261 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor .} #novela

UNION

    {?bookText wdt:P31 wd:Q35760 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #ensayo #idioma de la obra spaña

UNION

    {?bookText wdt:P31 wd:Q37484 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #poema épico

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #obra literaria

UNION

    {?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q5364419 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #obra literaria Español moderno

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor .} #play

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q1088025 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #play español antiguo

UNION

    {?bookText wdt:P31 wd:Q25379 . ?bookText wdt:P407 wd:Q397 . ?bookText wdt:P577
?fechaPublicado . ?bookText wdt:P50 ?autor . } #play latin

OPTIONAL {?bookText wdt:P577 ?fechaPublicado . ?bookText wdt:P50 ?autor . ?bookText
wdt:P136 ?genre . ?bookText wdt:P18 ?image . ?bookText wdt:P3976 ?bvmc . ?bookText
wdt:P407 ?idioma .}

#fecha del primer y último libro publicado

filter (?fechaPublicado > "1492-01-01"^^xsd:dateTime && ?fechaPublicado < "1681-05-
26"^^xsd:dateTime)

SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es,en". }
}

```

```
ORDER BY ASC(?bookTextLabel)
```

TODOS LOS AUTORES

```
#todos los autores del Siglo de Oro de la literatura española
```

```
SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel
?categoryLabel
```

```
WHERE{
```

```
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q36180 . #Escritor
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
}
```

UNION

```
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q49757 . #Poeta
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
}
```

UNION

```
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q6625963 . #Novelista
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
}
```

UNION

```
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q487596 . #Dramaturgo
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
}
```

```

}

#nacimiento de garcilaso y calderon de la barca y muerte de calderon

filter (?birth > "1492-01-01"^^xsd:dateTime && ?birth < "1659-01-01"^^xsd:dateTime)

filter (?birthDeath > "1540-01-01"^^xsd:dateTime && ?birthDeath < "1681-05-
26"^^xsd:dateTime)

OPTIONAL{?autor wdt:P106 ?ocupacion . ?autor wdt:P373 ?category.}

SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }

}

ORDER BY ASC(?ocupacionLabel)

```

Versión 2

```

SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel
?image

```

```

WHERE{

```

```

    {

?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

?autor wdt:P27 wd:Q29 . #pais de procedencia

?autor wdt:P106 wd:Q36180 . #Escritor

?autor wdt:P569 ?birth.

?autor wdt:P570 ?birthDeath.

?autor wdt:P373 ?category .

?autor wdt:P106 ?ocupacion.

    } UNION {

?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

?autor wdt:P27 wd:Q29 . #pais de procedencia

?autor wdt:P106 wd:Q49757 . #Poeta

?autor wdt:P569 ?birth.

?autor wdt:P570 ?birthDeath.

?autor wdt:P373 ?category .

?autor wdt:P106 ?ocupacion .

    } UNION {

?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

```

```

?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q6625963 . #Novelista
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
} UNION {
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q487596 . #Dramaturgo
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
}
#nacimiento de garcilaso y calderon de la barca y muerte de calderon
filter (?birth > "1492-01-01"^^xsd:dateTime && ?birth < "1659-01-01"^^xsd:dateTime)
filter (?birthDeath > "1540-01-01"^^xsd:dateTime && ?birthDeath < "1681-05-26"^^xsd:dateTime)
#filter (?birthDeath < "1681-05-26"^^xsd:dateTime) #muerte de calderon de la barca
OPTIONAL{?autor wdt:P106 ?ocupacion . ?autor wdt:P373 ?category. ?autor wdt:P18 ?image}
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }
}
ORDER BY ASC(?ocupacionLabel)

```

Version 3

```

SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel
?image ?firma ?bvmc
WHERE{
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q36180 . #Escritor

```

```

?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion.
} UNION {
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q49757 . #Poeta
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
} UNION {
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q6625963 . #Novelista
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
} UNION {
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q487596 . #Dramaturgo
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
}

#nacimiento de garcilaso y calderon de la barca y muerte de calderon
filter (?birth > "1492-01-01"^^xsd:dateTime && ?birth < "1659-01-01"^^xsd:dateTime)

```

```

filter (?birthDeath > "1540-01-01"^^xsd:dateTime && ?birthDeath < "1681-05-
26"^^xsd:dateTime)

#filter (?birthDeath < "1681-05-26"^^xsd:dateTime) #muerte de calderon de la
barca

OPTIONAL{?autor wdt:P106 ?ocupacion . ?autor wdt:P373 ?category. ?autor wdt:P18
?image . ?autor wdt:P109 ?firma . ?autor wdt:P2799 ?bvmc}

SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }

}

ORDER BY ASC(?ocupacionLabel)

```

Versión 4 y final

```

SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel
?image ?firma ?bvmc

```

```

WHERE{

  {

    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

    ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

    ?autor wdt:P27 wd:Q29 . #pais de procedencia

    ?autor wdt:P106 wd:Q36180 . #Escritor

    ?autor wdt:P569 ?birth.

    ?autor wdt:P570 ?birthDeath.

    ?autor wdt:P18 ?image

  } UNION {

    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

    ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

    ?autor wdt:P27 wd:Q29 . #pais de procedencia

    ?autor wdt:P106 wd:Q49757 . #Poeta

    ?autor wdt:P569 ?birth.

    ?autor wdt:P570 ?birthDeath.

    ?autor wdt:P18 ?image

  } UNION {

    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

    ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español

    ?autor wdt:P27 wd:Q29 . #pais de procedencia

    ?autor wdt:P106 wd:Q6625963 . #Novelista

    ?autor wdt:P569 ?birth.

    ?autor wdt:P570 ?birthDeath.
  }
}

```

```

        ?autor wdt:P18 ?image
    } UNION {
        ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
        ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
        ?autor wdt:P27 wd:Q29 . #pais de procedencia
        ?autor wdt:P106 wd:Q487596 . #Dramaturgo
        ?autor wdt:P569 ?birth.
        ?autor wdt:P570 ?birthDeath.
        ?autor wdt:P18 ?image
    } UNION {
        ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
        ?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en español
        ?autor wdt:P27 wd:Q29 . #pais de procedencia
        ?autor wdt:P106 wd:Q482980 . #Autor
        ?autor wdt:P569 ?birth.
        ?autor wdt:P570 ?birthDeath.
        ?autor wdt:P18 ?image .
    }
    #nacimiento de garcilaso y calderon de la barca y muerte de calderon
    filter (?birth > "1492-01-01"^^xsd:dateTime && ?birth < "1659-01-01"^^xsd:dateTime)
    filter (?birthDeath > "1540-01-01"^^xsd:dateTime && ?birthDeath < "1681-05-26"^^xsd:dateTime)
    #filter (?birthDeath < "1681-05-26"^^xsd:dateTime) #muerte de calderon de la barca

    OPTIONAL{?autor wdt:P18 ?image . ?autor wdt:P109 ?firma . ?autor wdt:P2799 ?bvmc .}
    SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }
}

ORDER BY ASC(?autorLabel)

```

CONSULTAS FEDERADAS

Dado que la BVMC está federada en Wikidata podemos recuperar información de ambas fuentes. En este caso, se recuperan todas las obras de la BVMC por autor.

```

SELECT distinct * WHERE {
    wd:Q5682 wdt:P2799 ?id
    BIND(uri(concat("http://data.cervantesvirtual.com/person/", ?id)) as ?bvmcID)

```

```

SERVICE <http://data.cervantesvirtual.com/openrdf-sesame/repositories/data> {
  ?bvmcID <http://rdaregistry.info/Elements/a/authorOf> ?work .
  ?work rdfs:label ?workLabel
}

```

NOTA: Q5682 se reemplaza por un dato pasado por parámetro.

Adicionalmente se ha usado una herramienta externa para enriquecer la plataforma.

Herramientas RDF Adicionales

<http://histropedia.com/showcase/wikidata-viewer.html>

Se la utilizado el visor de datos RDF Histropedia. Esta herramienta permite visualizar cronogramas con datos extraídos de Wikidata. Para ello, se han empleado las consultas SPARQL creadas y ligeramente adaptadas.

De esta manera enriquecemos la plataforma, no sólo con datos abiertos, también con herramientas libres diseñadas para este propósito.

Ejemplo Mujeres que estudiaron en la UA

```

SELECT ?person ?personLabel ?LugarNacimientoLabel ?coordenadas
?FechaNacimiento ?image

WHERE {

  ?person wdt:P69 wd:Q2037040 . #han estudiado en la universidad de alicante

  ?person wdt:P21 wd:Q6581072 . #del genero femenino

  ?person wdt:P19 ?LugarNacimiento .

  ?LugarNacimiento wdt:P625 ?coordenadas .

  ?person wdt:P569 ?FechaNacimiento .

  ?person wdt:P18 ?image

  SERVICE wikibase:label { bd:serviceParam wikibase:language
"[AUTO_LANGUAGE],en" }
}

```

URL: <https://tinyurl.com/y9mkrz2v>

Autores del Siglo de Oro

```

SELECT DISTINCT ?autor ?autorLabel ?autorDescription ?birth ?birthDeath ?ocupacionLabel
?image ?firma

WHERE{

  {

    ?autor wdt:P31 wd:Q5 . # todas las instancias de humanos

```

```

?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en espanyol
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q36180 . #Escritor
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion.
}
UNION
{
?autor wdt:P31 wd:Q5 . # todas las instancias de humanos
?autor wdt:P1412 wd:Q1321 . #lenguas escritas o habladas en espanyol
?autor wdt:P27 wd:Q29 . #pais de procedencia
?autor wdt:P106 wd:Q49757 . #Poeta
?autor wdt:P569 ?birth.
?autor wdt:P570 ?birthDeath.
?autor wdt:P373 ?category .
?autor wdt:P106 ?ocupacion .
}
#nacimiento de garcilaso y calderon de la barca y muerte de calderon
filter (?birth > "1492-01-01"^^xsd:dateTime && ?birth < "1659-01-01"^^xsd:dateTime)
filter (?birthDeath > "1540-01-01"^^xsd:dateTime && ?birthDeath < "1681-05-26"^^xsd:dateTime)
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }
}
ORDER BY ASC(?autorLabel)

```

URL: <https://tinyurl.com/y7tq598d>

Libros del Siglo de Oro

```

SELECT DISTINCT ?bookText ?bookTextLabel ?bookTextDescription ?image ?autor ?autorLabel
?fechaPublicado ?genre ?genreLabel ?idioma ?idiomaLabel
WHERE {
{?bookText wdt:P31 wd:Q83790 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicadoLabel . } #libro de texto
UNION
{?bookText wdt:P31 wd:Q571 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577
?fechaPublicado .} #libro impreso

```

UNION

```
{?bookText wdt:P31 wd:Q8261 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577  
?fechaPublicado .} #novela
```

UNION

```
{?bookText wdt:P31 wd:Q35760 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577  
?fechaPublicado .} #ensayo #idioma de la obra
```

UNION

```
{?bookText wdt:P31 wd:Q37484 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577  
?fechaPublicado .} #poema epico
```

UNION

```
{?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q1321 . ?bookText wdt:P577  
?fechaPublicado .} #obra literaria Q5364419
```

UNION

```
{?bookText wdt:P31 wd:Q7725634 . ?bookText wdt:P407 wd:Q5364419 . ?bookText wdt:P577  
?fechaPublicado .} #obra literaria Espanyol moderno
```

#fecha del primer y año ultimo libro publicado

```
filter (?fechaPublicado > "1492-01-01"^^xsd:dateTime && ?fechaPublicado < "1681-05-  
26"^^xsd:dateTime)
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es,en". }
```

}

URL: <https://tinyurl.com/y7awxk8k>

Conjuntos de datos objetivo

Estos y muchos otros conjuntos de datos del Siglo de la edad de oro son candidatos para traducirse a una consulta Sparql e introducirlo en nuestro sistema, por ejemplo, como filtros avanzados para el buscador.

- Autores de la edad de oro de la literatura española
- Autores de la edad de oro de la literatura española que han muerto fuera de España
- Lugar de la muerte de los autores de la edad de oro
- Familiares también autores, de los autores de la edad de oro
- Fecha de la muerte de los autores
- Año de publicación de los libros de los autores de la edad de oro
- Autoras de la edad de oro de la literatura española
- Autores de la edad de oro de la literatura española
- Mapa lugares de nacimiento de autores del siglo de oro de la literatura española
- Cronograma con las fechas de nacimiento de los autores del siglo de oro
- Cronograma con las fechas de publicaciones de los autores del siglo de oro

Fuentes

- <https://query.wikidata.org/>
- <http://data.cervantesvirtual.com/help>
- <http://histropedia.com/>
- <https://www.w3.org/TR/rdf-sparql-query/>
- https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries

D. Anexo IV. MockUp's

Diseño gráfico de las interfaces para la visualización de datos RDF.

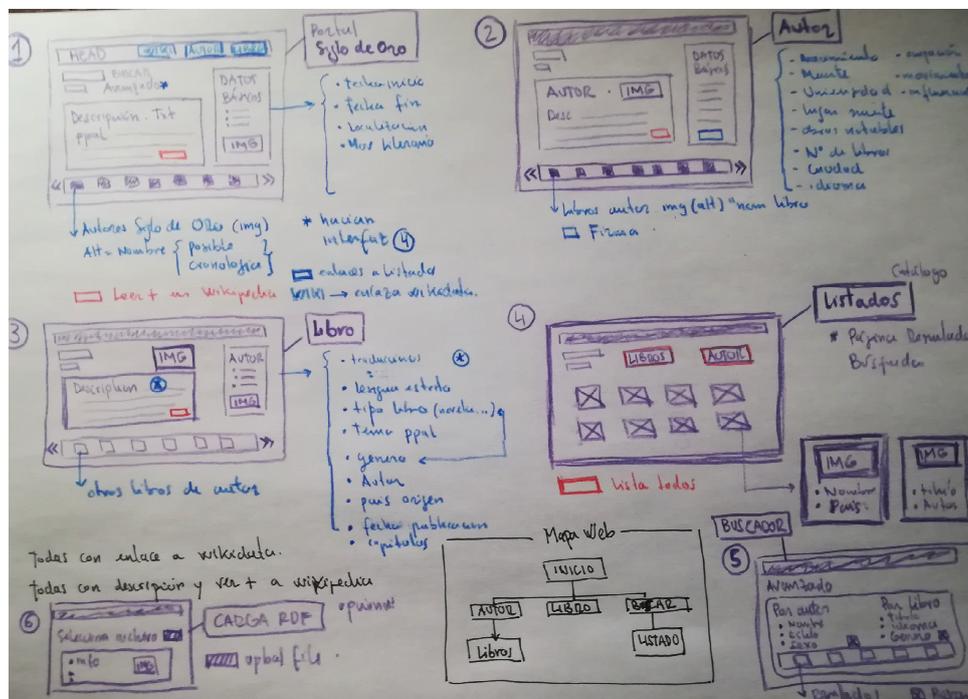
Boceto inicial

En un primer momento se han esbozado los dibujos de las 6 interfaces principales del visor de datos. Estas irán evolucionando en el tiempo. Para ello, he trabajado con la herramienta Justinmind para el prototipado. Es una aplicación de escritorio que permite crear prototipos dinámicos dando una visión general tanto del diseño y estructura como del flujo de navegación. Los prototipos se han creado en escala de grises para enfatizar la estructura y la organización de los datos RDF que se quieren visualizar.

El menú principal debe ser visible en todas las páginas así como el pie de página.

La herramienta utilizada para el prototipado ha sido <https://www.justinmind.com/>.

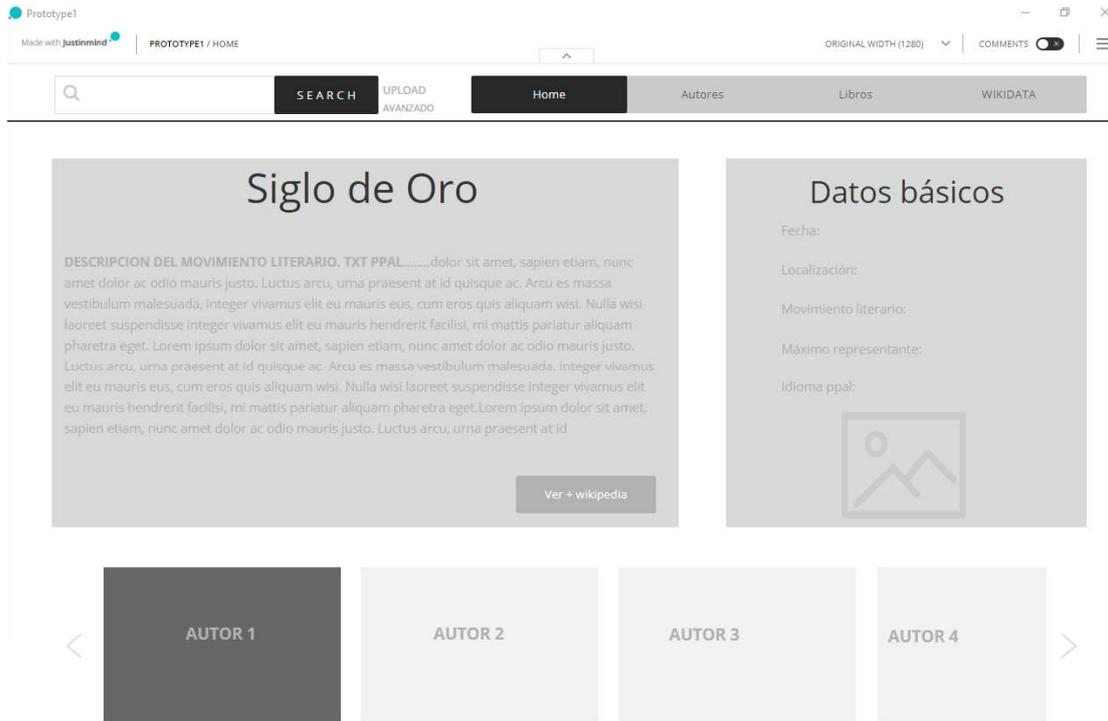
Visión General



En primer lugar se diseñaron las diferentes interfaces a mano alzada. De esta forma se tiene más libertad creativa y permite realizar correcciones rápidamente. Un vez terminado el boceto, este rediseña con la herramienta Justinmind que permite dar más detalle al diseño.

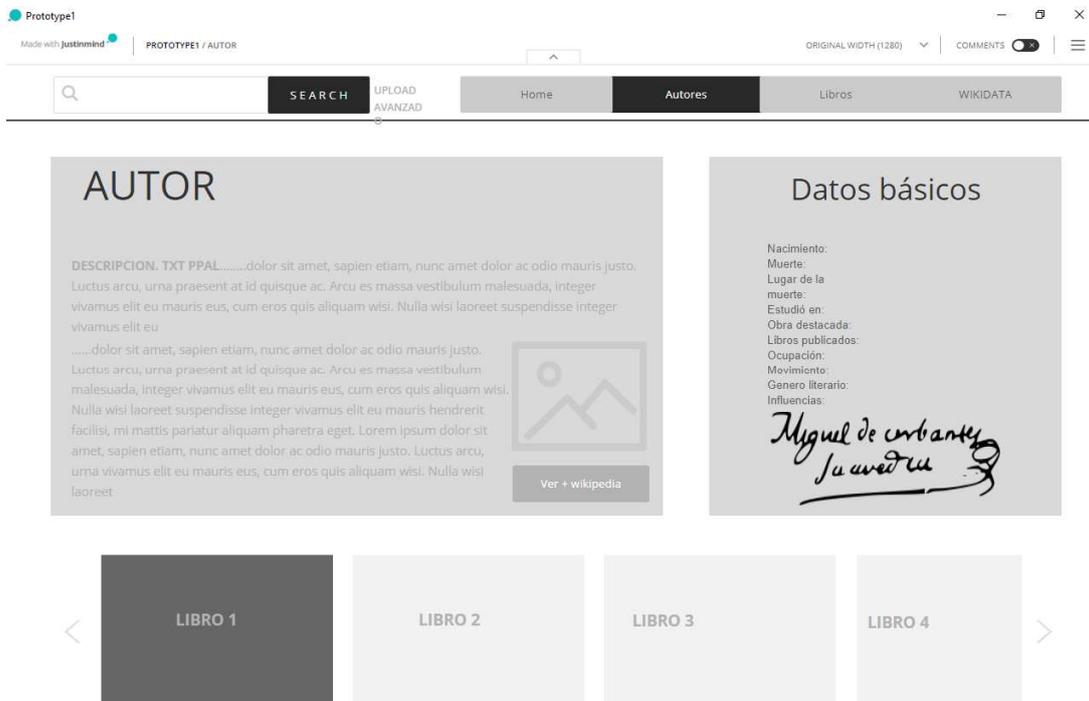
A continuación se muestran los mockup's de cada una de las interfaces.

Interfaz 1: Home page



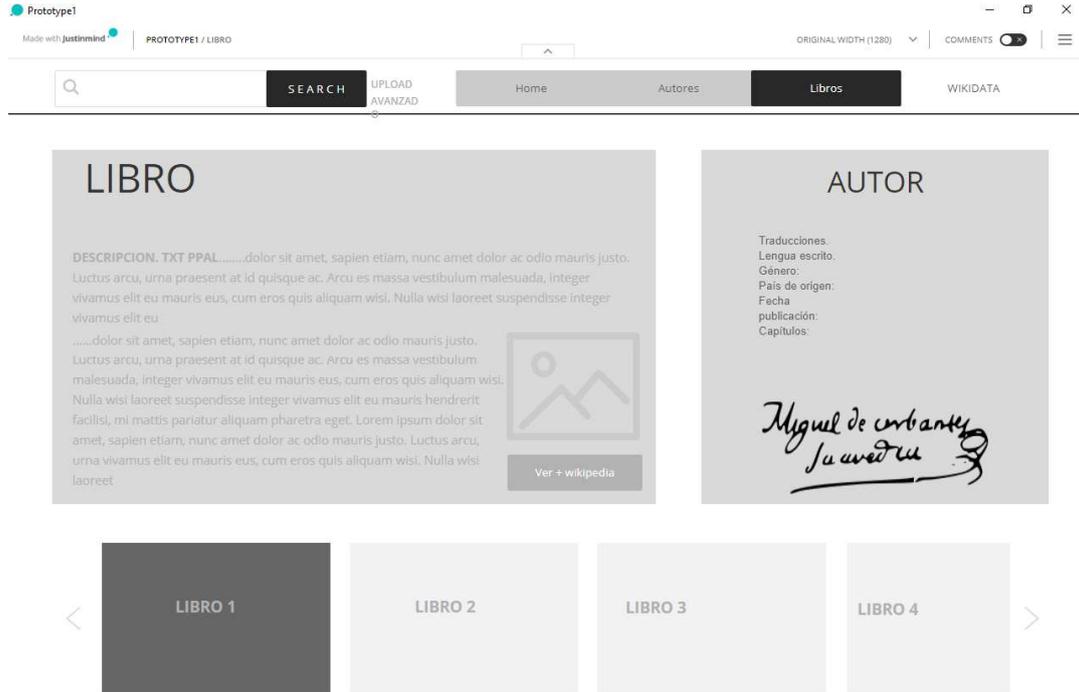
Página principal del visor RDF. Debe mostrar datos básicos de la categoría del Siglo de Oro de la Literatura Española y una lista de autores y/o libros de la época.

Interfaz 2: autor



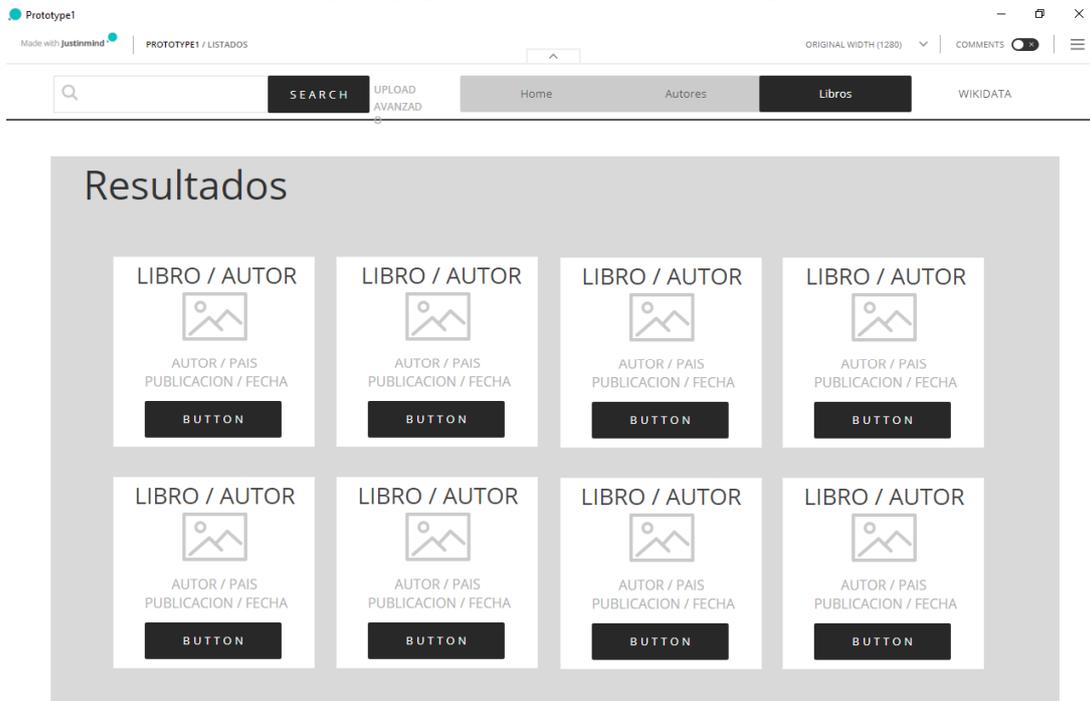
Interfaz que muestra datos de autores RDF. La interfaz debe mostrar datos relativos al autor así como su bibliografía. Debe incorporar imagen de autor y de la firma si la hubiera.

Interfaz 3: libro



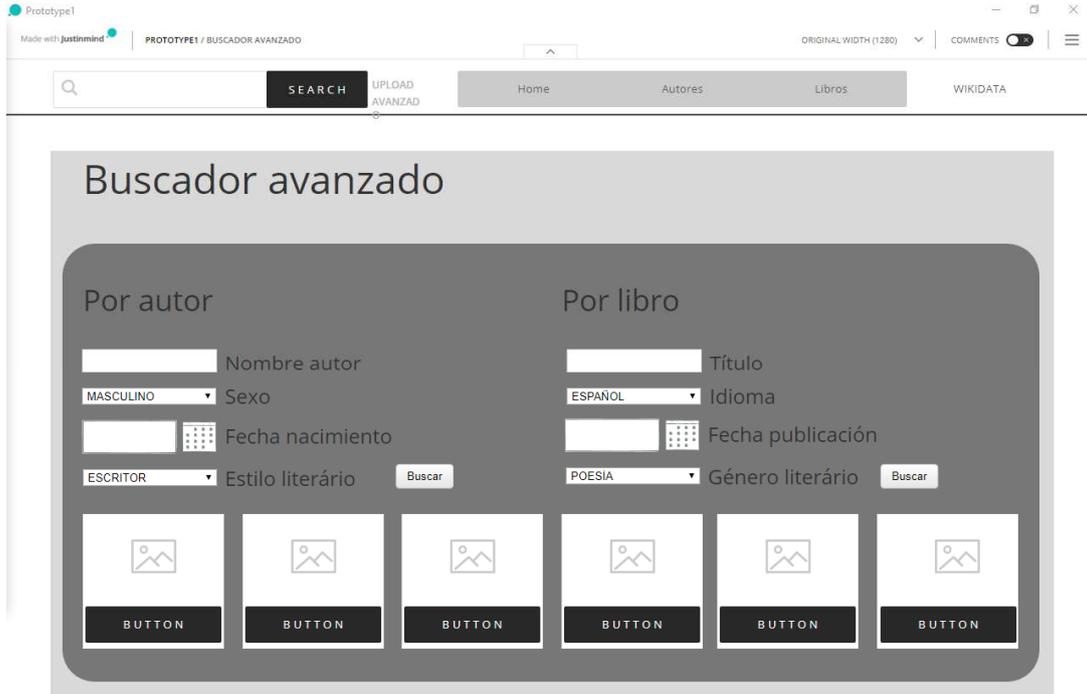
Interfaz para visualizar datos RDF sobre libros del Siglo de Oro. Debe mostrar datos relativos al libro así como el autor y enlaces al mismo.

Interfaz 4: listados (catálogo / página de resultados de búsqueda)



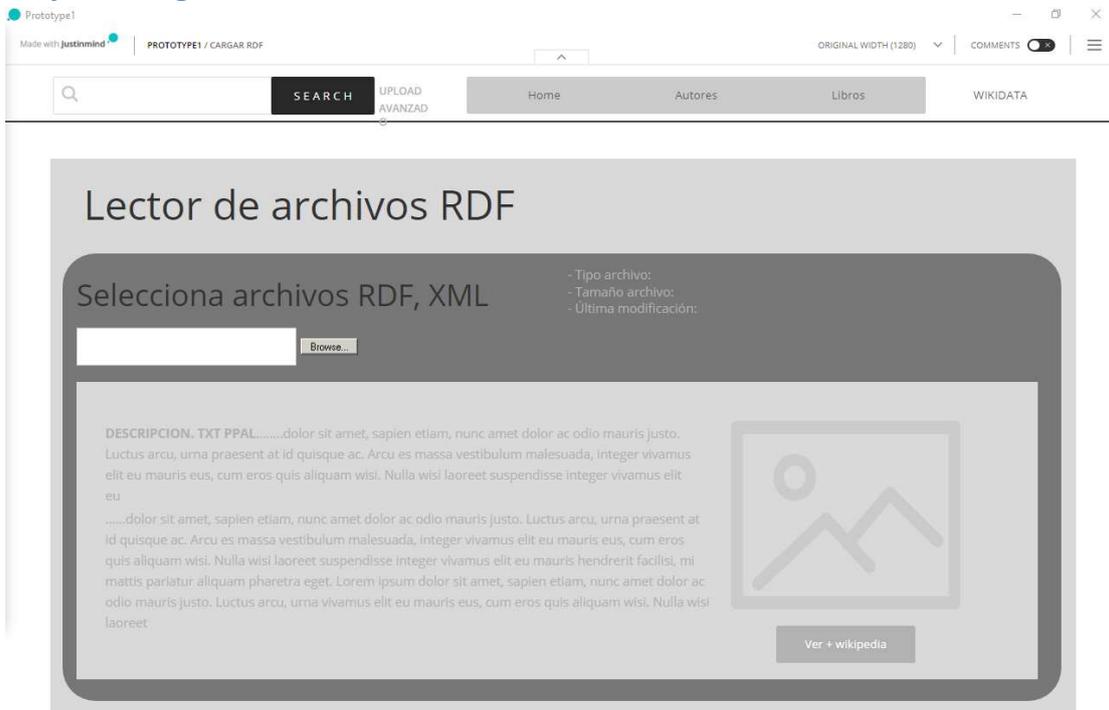
Interfaz para mostrar listados de autores y libros del Siglo de Oro. Debe mostrar información básica así como botones de paginación. Permitir el acceso a información ampliada clicando sobre el libro o autor.

Interfaz 5: Buscador avanzado



Interfaz para localizar libros y autores del Siglo de Oro. Debe permitir introducir, al menos, un campo de búsqueda. Ofrecer filtros para fecha, idioma o género literario entre otros. Los resultados se mostrarán en la misma página.

Interfaz 6: carga manual de archivos



Interfaz creada para pruebas de lectura de ficheros RDF. Debe permitir la carga de archivos RDF, JSON y XML para visualizar su contenido por pantalla. Debe mostrar detalles del archivo como el tipo, nombre o tamaño.

E. Anexo V. XML vs JSON

Se han estudiado las ventajas y desventajas de estos formatos para determinar con qué tipo de archivo se va a trabajar en el proyecto. Como se ha visto en el capítulo 1, tanto JSON como XML pueden utilizarse para recibir datos de un servidor web.

JSON es un formato de texto ligero para trabajar con el intercambio de datos mientras que XML es un meta-lenguaje que permite definir lenguajes de marcas para almacenar datos en forma legible. Ambos son formatos útiles para recibir datos de un servidor y tienen como finalidad organizar y estructurar la información.

Tanto JSON como XML son *autodescriptivos* y legibles por humanos. Ambos son jerárquicos ya que permiten almacenar valores dentro de otros valores y pueden ser analizados y utilizados por muchos lenguajes de programación. Tanto JSON como XML se pueden recuperar con una petición XMLHttpRequest ¹.

Vemos en un ejemplo de W3C [W3schools, 2019] como JSON en la figura E.1 es un texto más sencillo que el ejemplo de XML de la figura E.2.

Listado E.1: Ejemplo de la estructura del formato JSON.

```
1 {"employees":[
2   { "firstName":"John", "lastName":"Doe" },
3   { "firstName":"Anna", "lastName":"Smith" },
4   { "firstName":"Peter", "lastName":"Jones" }
5 ]}
```

Listado E.2: Ejemplo de la estructura del formato XML.

```
1 <employees>
2   <employee>
3     <firstName>John</firstName> <lastName>Doe</lastName>
4   </employee>
5   <employee>
6     <firstName>Anna</firstName> <lastName>Smith</lastName>
7   </employee>
8   <employee>
9     <firstName>Peter</firstName> <lastName>Jones</lastName>
10  </employee>
11 </employees>
```

- JSON no usa etiqueta final
- JSON es más corto
- JSON es más rápido de leer y escribir

¹<https://developer.mozilla.org/es/docs/Web/API/XMLHttpRequest>

- JSON puede usar matrices

Por el contrario, XML tiene una estructura más compleja como vemos en la figura E.1.

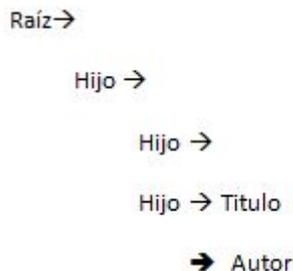


Figura E.1.: Ejemplo de nodos en una estructura XML.

Pasos para extraer la información de un documento XML.

- Obtener un documento XML
- Utilizar el DOM XML para recorrer el documento
- Extraer árbol de nodos.
- Extraer valores y almacenar en variables.

El principal problema es que XML permite muchos lenguajes. Mediante el uso de la función *nodeName* en JS, podemos identificar cada nodo y su información pero en ocasiones, en los documentos RDF todos sus nodos se etiquetan (en el primer nivel) con la misma etiqueta *rdf* y es necesario mirar en el siguiente nivel o nodo para comprobar que información contiene, lo que dificulta la manipulación del texto de respuesta. Además, si el documento es muy grande ralentiza la carga y visualización de los datos. Sin embargo, el proceso de manipulación de JSON es más simple.

- Obtener una cadena JSON
- Con la función *JSON.Parse* formatea la cadena JSON

Aunque los principales repositorios ofrecen sus datos en ambos formatos, para ese proyecto se ha decidido trabajar con JSON porque se presenta como un formato más sencillo y ligero de XML además de que se adapta mejor al uso de tecnología AJAX. También es uno de los formatos más ofrecidos en los repositorios LOD como se explica en el capítulo 4.

En la figura E.2, podemos ver la estructura del document JSON que devuelve Wikidata con nuestra consulta.

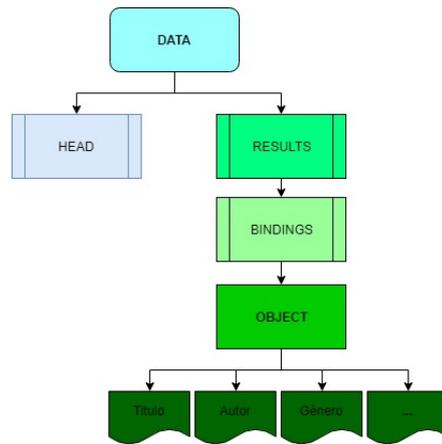


Figura E.2.: Representación de la estructura del archivo JSON devuelto por Wikidata.

En la figura E.3 vemos el ejemplo real que se ha usado en este proyecto para la entidad *libro*.

```

{
  "head": {
    "vars": [
      "bookText",
      "bookTextLabel",
      "bookTextDescription",
      "image",
      "autor",
      "autorLabel",
      "fechaPublicado",
      "genre",
      "genreLabel",
      "idioma",
      "idiomaLabel",
      "bvnc"
    ],
    "length": 12,
    "bindings": {}
  },
  "results": {
    "bindings": [
      {
        "autor": {
          "type": "uri",
          "value": "http://www.wikidata.org/entity/Q5699869"
        },
        "autorLabel": {
          "xml:lang": "es",
          "type": "literal",
          "value": "Antonio Colmenero de Ledesma"
        },
        "bookText": {
          "type": "uri",
          "value": "http://www.wikidata.org/entity/Q23986197"
        },
        "bookTextDescription": {
          "xml:lang": "es",
          "type": "literal",
          "value": "libro de Antonio Colmenero de Ledesma"
        },
        "bookTextLabel": {
          "type": "literal",
          "value": "Q23986197"
        },
        "fechaPublicado": {
          "datatype": "http://www.w3.org/2001/XMLSchema#dateTime",
          "type": "literal",
          "value": "1643-01-01T00:00:00Z"
        },
        "image": {
          "type": "uri",
          "value": "http://commons.wikimedia.org/wiki/Special:FilePath:divis3C3A9%20em328quatre%20partes%2C%201643.png"
        },
        "bvnc": {}
      }
    ]
  }
}
  
```

Figura E.3.: Datos del archivo JSON devuelto por Wikidata para libros.