
Explicaciones para un recomendador de música basadas
en datos enlazados

Title in English (defined in Cascaras\cover.tex)



Trabajo de Fin de Grado
Curso 2019–2020

Autor

Adrián Garrido Sierra
Diego Sánchez Muniesa

Director

Guillermo Jiménez Díaz
Marta Caro Martínez

Grado en **Ingeniería Informática**

Facultad de Informática

Universidad Complutense de Madrid

Explicaciones para un recomendador de
música basadas en datos enlazados
Title in English (defined in
Cascaras\cover.tex)

Trabajo de Fin de Grado en **Ingeniería Informática**
Departamento de **XXXXXXXXXXXXXX**

Autor
Adrián Garrido Sierra
Diego Sánchez Muniesa

Director
Guillermo Jiménez Díaz
Marta Caro Martínez

Convocatoria: *Febrero/Junio/Septiembre 2020*
Calificación: *Nota*

Grado en **Ingeniería Informática**
Facultad de Informática
Universidad Complutense de Madrid

DIA de MES de AÑO

Dedicatoria

*A Pedro Pablo y Marco Antonio, por crear TeXiS
e iluminar nuestro camino*

Agradecimientos

A Guillermo, por el tiempo empleado en hacer estas plantillas. A Adrián, Enrique y Nacho, por sus comentarios para mejorar lo que hicimos. Y a Narciso, a quien no le ha hecho falta el Anillo Único para coordinarnos a todos.

Resumen

Explicaciones para un recomendador de música basadas en datos enlazados

Un resumen en castellano de media página, incluyendo el título en castellano. A continuación, se escribirá una lista de no más de 10 palabras clave.

Palabras clave

Máximo 10 palabras clave separadas por comas

Abstract

Title in English (defined in Cascaras\cover.tex)

An abstract in English, half a page long, including the title in English. Below, a list with no more than 10 keywords.

Keywords

10 keywords max., separated by commas.

Índice

1. Introducción	1
1.1. Motivación	2
1.2. Objetivos	2
1.3. Plan de trabajo	2
1.4. Explicaciones adicionales sobre el uso de esta plantilla	2
1.4.1. Texto de prueba	2
2. Estado de la Cuestión	7
2.1. Conceptos sobre el ámbito	7
2.2. WikiData	8
2.3. SPARQL	8
3. Explicaciones	11
3.1. Explicaciones directas	11
3.2. Explicaciones indirectas	15
3.2.1. Explicaciones basadas en el artista	15
3.2.2. Explicaciones basadas en el género	18
4. Diseño de la interfaz de explicaciones	21
4.1. Diseño del grafo	21
4.2. Diseño de la interfaz	23
5. Implementación	25
5.1. Procesamiento y manejo de la muestra	25
5.2. Arquitectura de la aplicación	26
6. Conclusiones y Trabajo Futuro	27
6.1. Trabajo futuro	27
7. Introduction	29
8. Conclusions and Future Work	31
Bibliografía	33
A. Título del Apéndice A	35

Índice de figuras

3.1. Ejemplo de explicación por artista	12
3.2. Ejemplo de explicación por álbum	12
3.3. Ejemplo de explicación por década	13
3.4. Ejemplo de explicación por premio	13
3.5. Ejemplo de explicación por género	14
3.6. Ejemplo de explicación por compañía discográfica	14
3.7. Ejemplo de explicación por single	14
3.8. Ejemplo de explicación por idioma	15
3.9. Ejemplo de explicación por país de origen	16
3.10. Ejemplo de explicación por influencia	16
3.11. Ejemplo de explicación por integrantes	17
3.12. Ejemplo de explicación por tipo de voz	18
3.13. Ejemplo Relaciones que pueden tener artistas por premios	18
4.1. Primer diseño de la interfaz web	22
4.2. Primer diseño de la interfaz web	23

Índice de tablas

Capítulo 1

Introducción

“Frase célebre dicha por alguien inteligente”

— Autor

El estudiante elaborará una memoria descriptiva del trabajo realizado, con una **extensión mínima recomendada de 50 páginas** incluyendo al menos una introducción, objetivos y plan de trabajo, resultados con una discusión crítica y razonada de los mismos, conclusiones y bibliografía empleada en la elaboración de la memoria.

La memoria se puede redactar en castellano o en inglés, pero en el primer caso la introducción y las conclusiones de la memoria tienen que traducirse también al inglés y aparecerán como capítulos **al final de la memoria**. En ambos casos, el título de la memoria aparecerá en castellano y en inglés.

Además del cuerpo principal describiendo el trabajo realizado, la memoria contendrá los siguientes elementos, que no computarán para el cálculo de la extensión mínima del trabajo:

- un resumen en inglés de media página, incluyendo el título en inglés,
- ese mismo resumen en castellano, incluyendo el título en castellano,
- una lista de no más de 10 palabras clave en inglés,
- esa misma lista en castellano,
- un índice de contenidos, y
- una bibliografía.

La portada de la memoria deberá contener la siguiente información:

- "Máster en NOMBRE DEL MÁSTER, Facultad de Informática, Universidad Complutense de Madrid"
- Título
- Autor
- Director(es)
- Colaborador externo de dirección, si lo hay

- Curso académico
- Solo en la versión final: convocatoria y calificación obtenida

Para facilitar la escritura de la memoria siguiendo esta estructura, el estudiante podrá usar las plantillas en LaTeX o Word preparadas al efecto y publicadas en la página web del máster correspondiente.

Todo el material no original, ya sea texto o figuras, deberá ser convenientemente citado y referenciado. En el caso de material complementario se deben respetar las licencias y copyrights asociados al software y hardware que se emplee. En caso contrario no se autorizará la defensa, sin menoscabo de otras acciones que correspondan.

1.1. Motivación

Introducción al tema del TFM.

1.2. Objetivos

Descripción de los objetivos del trabajo.

1.3. Plan de trabajo

Aquí se describe el plan de trabajo a seguir para la consecución de los objetivos descritos en el apartado anterior.

1.4. Explicaciones adicionales sobre el uso de esta plantilla

Si quieres cambiar el **estilo del título** de los capítulos, edita `TeXiS\TeXiS_pream.tex` y comenta la línea `\usepackage[Lenny]{fncychap}` para dejar el estilo básico de L^AT_EX.

Si no te gusta que no haya **espacios entre párrafos** y quieres dejar un pequeño espacio en blanco, no metas saltos de línea (`\\`) al final de los párrafos. En su lugar, busca el comando `\setlength{\parskip}{0.2ex}` en `TeXiS\TeXiS_pream.tex` y aumenta el valor de `0,2ex` a, por ejemplo, `1ex`.

TFMTeXiS se ha elaborado a partir de la plantilla de TeXiS¹, creada por Marco Antonio y Pedro Pablo Gómez Martín para escribir su tesis doctoral. Para explicaciones más extensas y detalladas sobre cómo usar esta plantilla, recomendamos la lectura del documento `TeXiS-Manual-1.0.pdf` que acompaña a esta plantilla.

El siguiente texto se genera con el comando `\lipsum[2-20]` que viene a continuación en el fichero `.tex`. El único propósito es mostrar el aspecto de las páginas usando esta plantilla. Quita este comando y, si quieres, comenta o elimina el paquete `lipsum` al final de `TeXiS\TeXiS_pream.tex`

1.4.1. Texto de prueba

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet,

¹<http://gaia.fdi.ucm.es/research/texis/>

tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a,

dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus

orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congue neque id dolor.

Donec et nisl at wisi luctus bibendum. Nam interdum tellus ac libero. Sed sem justo, laoreet vitae, fringilla at, adipiscing ut, nibh. Maecenas non sem quis tortor eleifend fermentum. Etiam id tortor ac mauris porta vulputate. Integer porta neque vitae massa. Maecenas tempus libero a libero posuere dictum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aenean quis mauris sed elit commodo placerat. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Vivamus rhoncus tincidunt libero. Etiam elementum pretium justo. Vivamus est. Morbi a tellus eget pede tristique commodo. Nulla nisl. Vestibulum sed nisl eu sapien cursus rutrum.

Nulla non mauris vitae wisi posuere convallis. Sed eu nulla nec eros scelerisque pharetra. Nullam varius. Etiam dignissim elementum metus. Vestibulum faucibus, metus sit amet mattis rhoncus, sapien dui laoreet odio, nec ultricies nibh augue a enim. Fusce in ligula. Quisque at magna et nulla commodo consequat. Proin accumsan imperdiet sem. Nunc porta. Donec feugiat mi at justo. Phasellus facilisis ipsum quis ante. In ac elit eget ipsum pharetra faucibus. Maecenas viverra nulla in massa.

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Estado de la Cuestión

Actualmente Internet aloja una cantidad ingente de información de todo tipo. A pesar de los beneficios evidentes que supone tener tanta información subida a la web, este hecho también trae consigo problemas debido a dificultades semánticas, un más que elevado número de fuentes de información y el propio tamaño de los datos. Como consecuencia, la información accesible de forma efectiva es muy inferior a lo que cabría esperar.

A continuación estudiaremos tecnologías cuyo objetivo es facilitar contexto a los datos y facilitar el proceso de búsqueda al proveer información verdaderamente relevante para los usuarios.

2.1. Conceptos sobre el ámbito

La llamada **Web Semántica** es una extensión de la World Wide Web propuesta por Tim Berners-Lee [2], cuyo objetivo es proveer a los datos estructura y significado. Con esto se consigue que la información de la web sea más fácilmente comprensible para las máquinas (o agentes de software). De esta forma se desea lograr que los agentes no solo analicen gramaticalmente, sino que comprendan la gran cantidad de información contenida en la web. Gracias a este entendimiento, los agentes podrán integrar datos de diversas fuentes, inferir hechos ocultos y responder a consultas complejas fácilmente [8].

Para poder hacer realidad esta Web Semántica es necesario que la información de la web sea accesible en un formato estandarizado, accesible y manejable por las herramientas de la Web Semántica. Además, no basta con tener acceso a los datos sino también a la relación entre la información, lo cual marca la diferencia entre la Web Semántica y una simple colección de datasets. Esta colección de datos interrelacionados es lo que llamamos **Linked Data** (o datos enlazados) y será la base de nuestro trabajo en el dominio de los temas musicales.

El sistema estructural más importante que nos interesa para este proyecto es el **Esquema RDF** (Resource Description Framework), que es un modelo de datos basado en declaraciones sobre recursos web mediante expresiones sujeto-predicado-objeto. Dichas expresiones se llaman “tripletas”, donde el sujeto representa al recurso, el objeto representa el valor del recurso y el predicado supone los rasgos o aspectos del recurso que relacionan sujeto y objeto [8].

Estos tres argumentos se representan mediante un Identificador de Recursos: una

URI (Uniform Resource Identifier) [8, 1].

A su vez, esta URI puede ser una URL (Uniform Resource Locator), que es un identificador de recursos cuyos recursos referidos pueden variar en el tiempo, o una URN (Uniform Resource Name), que es un identificador de recursos independiente de la localización de estos recursos [3, 7, 8].

Esencialmente, estas tres tuplas están relacionadas entre ellas de una forma que podemos representar mediante un grafo. Cada una de estas tuplas puede estar relacionada también con otras, creando así un grafo más grande y completo que enlaza una gran cantidad de información.

Como hemos comentado, si empleamos este lenguaje seremos capaces de relacionar un sujeto y un objeto mediante un predicado de forma que obtendremos un grafo similar al siguiente:

Podríamos decir que estamos estableciendo una relación entre dos objetos unidos mediante una propiedad. Esta propiedad es una “explicación” que nos sirve para relacionar objetos. En nuestro proyecto, estos objetos son canciones.

2.2. WikiData

Wikidata es una base de datos libre, colaborativa, multilingüe y secundaria, que recopila datos estructurados para dar soporte a Wikipedia, Wikimedia Commons, así como a otras wikis del movimiento Wikimedia y a cualquier persona en el mundo [4].

Gracias a Wikidata, tenemos una gran cantidad de información relevante acerca de toda nuestra cultura reunida en un mismo sitio. La usaremos para consultar e investigar la información relacionada con nuestro dataset (o conjunto de datos), esencialmente datos acerca de artistas y sus temas musicales.

Necesitamos un método para que nuestra aplicación pueda navegar por esta información, por lo que usaremos SPARQL.

2.3. SPARQL

El lenguaje principal que vamos a utilizar para consultar los datos enlazados de la web es SPARQL. Este es un lenguaje especializado para buscar y consultar datos RDF. Tiene una sintaxis sencilla e incluye uso de prefijos para simplificar las URL.

Wikidata está estructurada de tal manera que resulta sencillo acceder a su información utilizando el lenguaje de consultas SPARQL. Cada objeto tiene ciertas propiedades a las que podemos acceder seleccionando el identificador correcto de la propiedad.

Un ejemplo de una consulta sencilla podría ser:

-Obtén todos los Artistas cuyo género musical sea Rock:

```
SELECT ?singer ?singerLabel ?genre ?genreLabel
WHERE
{
  ?singer wdt:P31 wd:Q215380;
           wdt:P136 wd:Q7749;
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" . }
}
```

En esta consulta estamos accediendo a todos los objetos que contengan las propiedades(wdt) P31(Tipo Instancia) y P136(Género) con el valor(wd) que nosotros estamos seleccionando, Q215380 y Q7749, forzando así a que las dos propiedades sean Grupo Musical y Rock and Roll respectivamente. SPARQL es un lenguaje muy potente que puede abarcar gran cantidad de datos en función de la web (se pueden crear consultas mucho más complejas para obtener datos más específicos).

Capítulo 3

Explicaciones

El objetivo de nuestro trabajo consiste en explicar las relaciones entre dos canciones dadas por un recomendador de música que trabaja con un dataset concreto. De esta manera, llamaremos **explicación** a cada factor común que compartan ambas canciones, siendo principalmente propiedades que poseen las propias canciones o los valores de otras propiedades previas. A lo largo de este capítulo concretaremos todas las explicaciones que vamos a estudiar.

En una primera fase, generaremos explicaciones básicas para relacionar diferentes tipos de canciones, por ejemplo: género, artista, álbum, etc. Asignaremos una complejidad de $k=1$ a las explicaciones directas. Estas explicaciones son una relación directa entre dos canciones relacionadas por una propiedad.

Después generaremos explicaciones más complejas que normalmente un humano pasaría por alto. Estas explicaciones complejas pueden estar formadas por relaciones indirectas entre los datos de una forma que se puede representar con un grafo. Estas explicaciones pueden tener un nivel de complejidad diferente ($k=2, 3, 4...$) dependiendo de cuántas aristas del grafo separen ambos elementos.

3.1. Explicaciones directas

En esta sección vamos a recopilar todas las propiedades propias de la canción en sí. Esto significa que son propiedades directamente relacionadas con la canción siguiendo el modelo RDF que podemos obtener con una consulta básica. Aquí encontraremos propiedades como el artista que interpreta la canción, la compañía discográfica que la ha publicado o los premios que ha recibido, entre otras.

Artista

El artista es una de las explicaciones más obvias pero también es una de las más importantes. Esta explicación hace referencia a cuando dos temas son interpretados por el mismo artista, por lo que tienen una clara relación directa ya que suele existir similitud entre las canciones de un artista. Además, a partir del artista podemos obtener otras relaciones más complejas en función de los datos de este mismo.

Como ejemplo podemos tomar los temas *One More Time* y *Get Lucky*. Ambos son

interpretados por el dúo musical **Daft Punk**, así que podemos explicar su relación gracias a este dato.

Para esta explicación utilizamos la propiedad “performer” (intérprete) de Wikidata. Siguiendo el modelo RDF, el sujeto es la canción (*One More Time*, por ejemplo), el predicado es la propiedad intérprete y el objeto es el artista (**Daft Punk**).



Figura 3.1: Ejemplo de explicación por artista

Álbum

Una explicación muy potente será que ambas canciones pertenezcan al mismo álbum. A menudo esta explicación aparecerá acompañada de la explicación del artista y, en cualquier caso, la relación que existe entre dos temas del mismo álbum suele ser más estrecha debido a que poseen más puntos en común, como puede ser el género, la fecha o la temática.

Tomemos como ejemplo *Smells Like Teen Spirit* y *Come as You Are*. Estas dos canciones son muy cercanas porque tienen varios puntos en común, pero una de las explicaciones que podemos ofrecer es que ambas pertenecen al álbum *Nevermind*, de **Nirvana**. *About a Girl* es otro tema que se podría recomendar a raíz de *Smells Like Teen Spirit*, pero es una peor elección que *Come as You Are* porque pertenece a un álbum diferente.

Volvemos a utilizar Wikidata para obtener esta explicación, concretamente la propiedad “part of” (parte de). Esta propiedad se utiliza para varias cosas, entre ellas indicar los álbumes a los que pertenecen las canciones. Así obtenemos que *Smells Like Teen Spirit* es “parte de” *Nevermind*.



Figura 3.2: Ejemplo de explicación por álbum

Década

Creemos que hay una mayor probabilidad de que exista una relación entre dos canciones que pertenezcan a la misma década. Esto se debe a que a lo largo del tiempo ha habido periodos marcados por uno o varios géneros musicales. Esto también ayuda a estudiar cómo estos distintos géneros están relacionados entre sí, lo cual es otro punto importante a tener en cuenta ya que hay géneros íntimamente relacionados entre sí: techno y house, heavy metal y thrash metal, etc.

Gracias a esta explicación, podemos relacionar dos canciones que a priori son muy distintas, como es el caso de *September* de **Earth, Wind & Fire** y *Bohemian Rhapsody* de **Queen**. Estos temas no comparten algo tan básico como el artista o el género, pero ambos fueron publicados en los años 70 y por ello pueden resultar interesantes para un usuario que busque escuchar los ritmos de esa época.

Para obtener esta explicación, emplearemos la propiedad “publication date” (fecha de publicación) de las canciones en Wikidata y comprobaremos si las dos fechas se sitúan dentro de la misma década.



Figura 3.3: Ejemplo de explicación por década

Premios

La siguiente explicación son los premios recibidos. Existe una variedad de premios compartidos por diferentes canciones. Algunos de estos premios son más específicos que otros, pero en cualquier caso pueden resultar una relación útil para nuestras explicaciones ya que son un reflejo de la repercusión de las canciones.

Así explicaríamos la relación entre *Single Ladies (Put a Ring on It)* de **Beyoncé** y *Rolling in the Deep* de **Adele**, ya que ambos temas han recibido el Premio Grammy a la canción del año.

Usaremos la propiedad “award received” (premio recibido) encontrada en Wikidata. De esta forma, *Rolling in the Deep* sería el sujeto, “award received” sería el predicado y *Grammy Award for Song of the Year* sería el objeto.

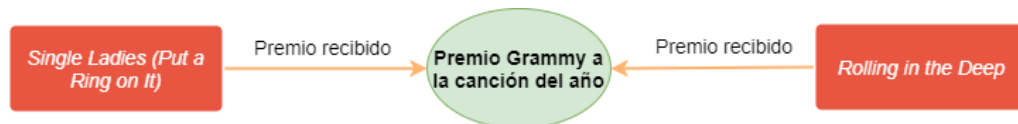


Figura 3.4: Ejemplo de explicación por premio

Género

La explicación Género es una de las más representativas ya que las personas se guían por el género que prefieren para escuchar canciones similares, aunque en ocasiones no deba ser así. Un usuario que sea fan del jazz querrá escuchar canciones de ese mismo género o géneros similares porque todas las temas de un mismo género comparten una serie de puntos en común como la elección de los instrumentos, temáticas similares en sus letras o patrones de composición.

Gracias al género, podemos explicar la relación entre canciones como *Master of Puppets* de **Metallica** e *Indians* de **Anthrax**, que pertenecen al **thrash metal**.

Para esta relación utilizamos la propiedad “genre” (género) de Wikidata. Así tenemos que la canción *Master of Puppets* pertenece al género **thrash metal**.

Compañía discográfica

Compañía discográfica hace referencia a la compañía por la que ha firmado el artista para publicar un tema concreto. Hemos podido observar que una misma compañía puede



Figura 3.5: Ejemplo de explicación por género

firmar a artistas que a veces no tienen relación ninguna en cuanto al estilo musical, pero hay otras causas que sí los pueden relacionar indirectamente como la tendencia del momento, el target del público que generan, etc.

Para ilustrarlo con un ejemplo, tanto la canción *Born in the U.S.A.* de **Bruce Springsteen** como la canción *Ashes* de **Céline Dion** fueron publicadas por el sello discográfico **Columbia Records**.

Podemos alcanzar esta explicación gracias a la propiedad “record label” (sello discográfico) de Wikidata. *Born in the U.S.A.* tiene un “record label” cuyo nombre (o valor) es **Columbia Records**.



Figura 3.6: Ejemplo de explicación por compañía discográfica

Singles

Existe también una cierta relación entre aquellos temas que sean singles o sencillos, así que consideramos esto como una explicación más. El razonamiento para esta decisión es que los singles son canciones que se publican de forma independiente por razones promocionales, por lo que suelen convertirse en los temas más populares y representativos del trabajo del artista. Por ello, pueden poseer más valor para el recomendador que otras canciones porque los singles tienen más probabilidades de coincidir con los gustos del usuario.

El tema *Bad Guy* de **Billie Eilish** es un single, así que podemos decir que tiene cierta relación con *Shallow* de **Lady Gaga** y **Bradley Cooper**, que también es un single.

Para esta explicación emplearemos la propiedad “instance of” (instancia de) en Wikidata y su valor “single”. De esta forma comprobaremos que tanto *Bad Guy* como *Shallow* son “instancia de” “single”.



Figura 3.7: Ejemplo de explicación por single

Idioma

La explicación Idioma denota, como su nombre sugiere, que ambas canciones están en el mismo idioma o lenguaje. Existen numerosos ejemplos de temas musicales que están escritos

en un idioma que no se corresponde con la lengua materna del artista que los publican, o sencillamente están escritas en un idioma que se habla en varias partes del mundo distintas, como el inglés o el español. En ciertos casos el compartir idioma es un indicativo de otras relaciones adicionales, como puede ser el caso de canciones que pertenecen a un folclore regional determinado. En cualquier caso el compartir lenguaje ya es una relación directa en sí misma.

Como ejemplo podemos tomar el tema *Like a Virgin* de **Madonna** y *Rocket Man* de **Elton John**, los cuales están escritos en inglés.

Para esta explicación emplearemos la propiedad “language of work or name” (idioma de la obra o del nombre) de Wikidata. Así tenemos que la canción *Like a Virgin* tiene idioma “inglés”.



Figura 3.8: Ejemplo de explicación por idioma

3.2. Explicaciones indirectas

Las explicaciones indirectas nos van a ofrecer relaciones más complejas que nos proporcionen información que a simple vista no podríamos relacionar, ya que están relacionadas con sujetos externos como son el artista que interpreta la canción y el género de la canción.

El punto de partida principal serán las explicaciones directas. De cada una de las explicaciones directas podemos obtener más propiedades que indirectamente estarán relacionadas con la canción. Así pues, desde una canción podemos obtener su artista, a partir del artista los géneros musicales con los que está relacionado y pasamos a estudiar subgéneros de esos géneros obteniendo así un grafo en forma de árbol que puede dividirse en varios niveles. Podemos profundizar en estos niveles tanto como nosotros creamos conveniente.

A continuación enumeramos las explicaciones indirectas adicionales que buscamos para establecer la relación entre dos canciones. Cabe destacar que las dos principales ramas o propiedades directas que vamos a explorar son el género y el artista/grupo de la canción.

3.2.1. Explicaciones basadas en el artista

La primera propiedad de las canciones que hemos estudiado en profundidad es el artista. Se trata de uno de los aspectos que más influye en los temas musicales y, lo que es más importante, nos permite acceder a mucha información útil que no se relaciona de forma directa con la propia canción.

Las siguientes explicaciones se basan en propiedades de los artistas que interpretan las canciones analizadas:

País de origen (Artista)

El primer estudio de relaciones indirectas viene dado por investigar y obtener propiedades del artista que interpreta la canción.

Para empezar podemos fijarnos en el país de origen del artista. Esta explicación puede ser especialmente útil para países relativamente pequeños donde la música nacional presenta unos patrones claros. Además, a lo largo de la historia en un país se genera una tendencia o nuevo género musical el cual es propio de ese país.

La canción *Fire* de la banda **BTS** y *Kill This Love* de **Blackpink** guardan cierta similitud porque ambos grupos son originarios de **Corea del Sur**.

Para comprobarlo, obtendremos la propiedad “country of origin” (país de origen) del artista que previamente ya hemos almacenado al estudiar las relaciones directas en el caso de las bandas o la propiedad “country of citizenship” (país de nacionalidad) en el caso de los artistas en solitario. *Fire* tiene el “performer” **BTS**, cuyo “country of origin” es Corea del Sur.

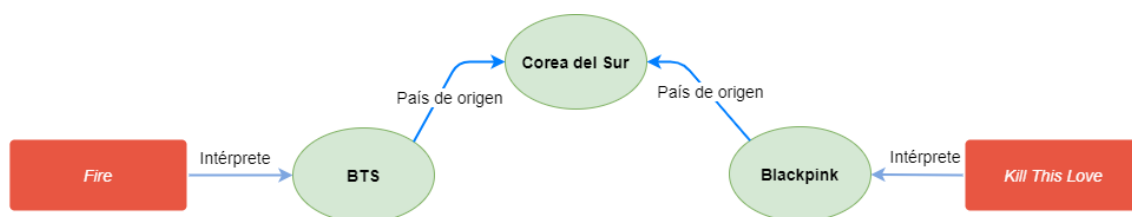


Figura 3.9: Ejemplo de explicación por país de origen

Cabe destacar que también hemos incluido una propiedad “location of formation” (lugar de formación) que hace referencia a la localidad de formación del grupo o artista, ejerciendo una selección más precisa pues nos da un núcleo de población más pequeño. Un ejemplo de ello podría ser dos míticas bandas de la ciudad de Boston: Aerosmith y Boston. Creemos que la unión de la ciudad natal con la música puede hacer a un usuario el escuchar grupos de su misma ciudad o estado.

Influencia

La **influencia de los artistas** es otra explicación importante. A menudo el trabajo de un artista se ve influenciado por otros artistas de formas que no siempre están ligadas a un género musical concreto, así que podemos encontrar una relación entre dos canciones examinando estas influencias.

Tomando como ejemplo las canciones *Billie Jean* de **Michael Jackson** y *Paint It Black* de **The Rolling Stones**, podemos establecer una relación entre ellas al ver que ambos artistas fueron influenciados por la banda **The Beatles**.

Obtenemos esta explicación gracias a la propiedad “influenced by” (influenciado por) de Wikidata, además de la propiedad “performer” para relacionar el tema con su artista. Así averiguamos que *Billie Jean* tiene de “intérprete” a **Michael Jackson**, quien fue “influenciado por” **The Beatles**.



Figura 3.10: Ejemplo de explicación por influencia

Integrantes

Los **integrantes** son cada uno de los miembros que conforman un grupo. En caso de ser un único artista, esta propiedad es irrelevante ya que haríamos referencia a la propiedad directa de artista. Esta propiedad resulta útil en casos en los que miembros de la banda han cambiado de grupo a lo largo de su carrera musical. Esto nos da acceso a otros grupos que por valores o gustos musicales del artista nos hacen pensar que pueden ser muy parecidos.

Si tomamos los temas *Lithium* de **Nirvana** y *The Pretender* de **Foo Fighters** podemos observar que existe una relación entre ellos porque el músico **Dave Grohl** ha formado parte de ambas bandas.

Para comprobarlo usaremos la propiedad de Wikidata “performer” y después comprobaremos la propiedad “has part” (compuesto de) para ver todos los miembros de la banda y buscar coincidencias. De este modo, podemos ver que *The Pretender* tiene un “performer” que es **Foo Fighters**, y a su vez **Foo Fighters** “has part” **Dave Grohl**.

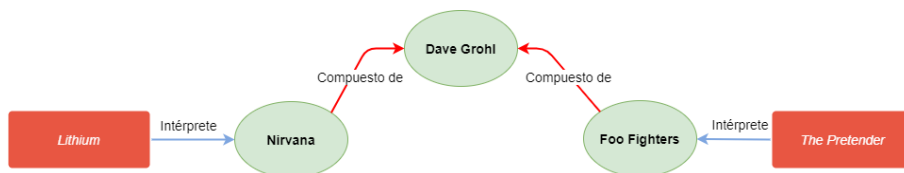


Figura 3.11: Ejemplo de explicación por integrantes

Tipo de voz

Siguiendo con el estudio de los artistas, el **tipo de voz** de los vocalistas es una buena explicación para relacionar dos canciones. El tipo de voz influye en el sonido general del tema y puede ser especialmente determinante en ciertos géneros musicales. Por limitaciones técnicas, solo aplicaremos esta explicación con artistas en solitario.

La voz de un artista puede encajar en más de un tipo, como veremos en el siguiente ejemplo, aunque en esta explicación buscamos que coincidan en al menos un tipo. La cantante **Lady Gaga** se asemeja a **Amy Winehouse** por su tipo de voz, ya que ambas son *mezzo-soprano* y *contralto*. Así podemos explicar la relación entre dos canciones de estas artistas como *Poker Face* y *Rehab*.

Para esta explicación resulta útil la propiedad “voice type” (tipo de voz) en Wikidata. Partiendo de *Poker Face*, utilizamos la propiedad “performer” para llegar a **Lady Gaga** y finalmente usamos su “voice type” para obtener *mezzo-soprano*.

Premios (Artista)

Por último nos queda comentar la propiedad de premios. En esta sección hemos recogido dos propiedades: “award received” (premio recibido) y “nominated for” (nominado a). Estas categorías hacen referencia a los premios que ha recibido los artistas a lo largo de su carrera y la segunda a los premios a los que han sido nominados. Normalmente los premios se reparten según distintas categorías en función al estilo de música, pero también hay otros que nos aportan otro tipo de información al estudio como Grammy al Mejor Videoclip, Grammy Award a la Mejor Canción Escrita, Grammy Award a la Mejor Interpretación

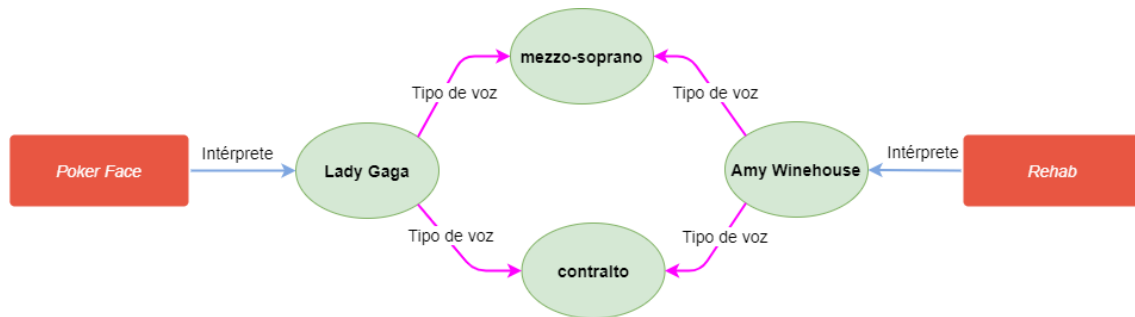


Figura 3.12: Ejemplo de explicación por tipo de voz

Rap/Sung etc...

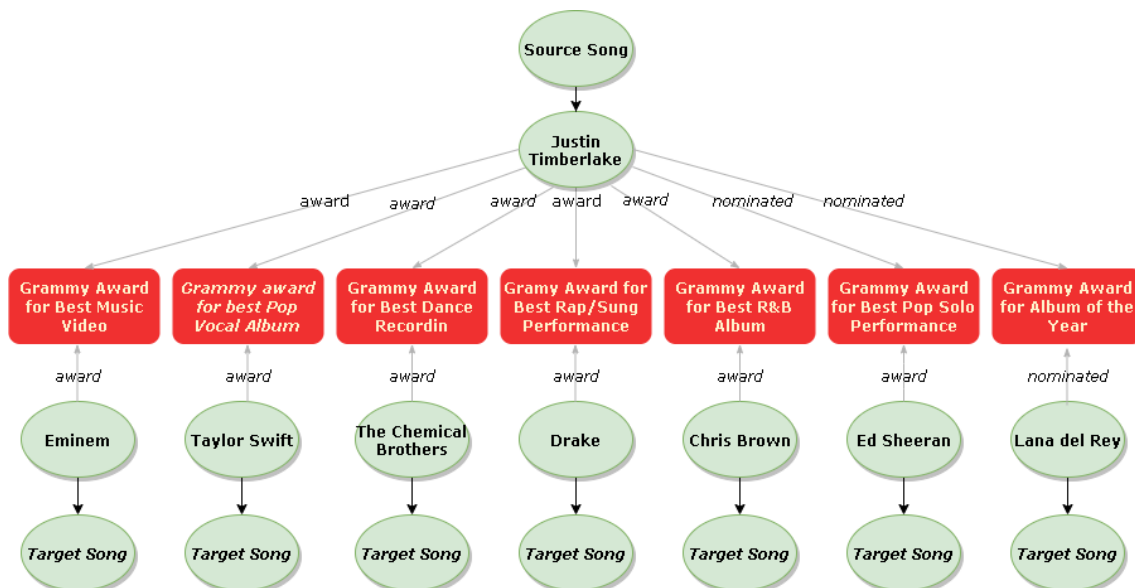


Figura 3.13: Ejemplo Relaciones que pueden tener artistas por premios

3.2.2. Explicaciones basadas en el género

El segundo estudio que hemos hecho ha sido el que viene ligado al género de la canción. Es una de las propiedades más detalladas en Wikidata y por lo tanto una de de las que más información nos pueden proporcionar. Hay muchos géneros y subgéneros documentados, lo que nos facilita la obtención y clasificación de estos.

En cuanto a esta propiedad, hemos sacado las siguientes relaciones:

Influenciado por y Subgénero

Al igual que pasaba con los artistas, los géneros también son influenciados por otros los cuales ha servido para el desarrollo de nuevos estilos a lo largo de la historia. Gracias a la evolución del Rock hay otros géneros que han marcado tendencia como el Rock and

Roll, el Heavy Metal, Hard Rock, Garage Rock etc... Cabe destacar la diferencia entre las propiedades Influenced by y Subgenre, mientras que la primera solo indica que ese estilo musical ha marcado una cierta influencia tomando algunos aspectos, la segunda señala que es un Subgénero, una tendencia muy parecida y que toma como base ese estilo musical.

Fecha de fundación

Con esta propiedad intentamos tener en cuenta la historia de la música o su cronograma. Hay géneros que fueron desarrollados conjuntamente a lo largo del tiempo. Un ejemplo de este caso se dio en curiosamente el tecnopop que tanto caracterizó a los años 80, pues convivió a veces con el soul y el estilo afroamericano doo-wop, de Billy Joel.

País de origen (Género)

Por último, recogemos la información del país donde se formó ese género. Con esta propiedad queremos poder recoger movimientos artísticos que surgieron en el mismo país. Los géneros que fueron formados en un determinado país están mucho más arraigados a este a pesar de que se hayan extendido por todo el mundo. Es por esto que si queremos relacionar una artista que cualquier parte del mundo, que tiene un estilo Techno, será más propenso a estar relacionado con artistas o canciones Holandesas, cuna del este estilo musical.

Capítulo 4

Diseño de la interfaz de explicaciones

El objetivo de nuestro proyecto, como ya hemos expresado anteriormente en este documento, consiste en proporcionar explicaciones que justifiquen la relación existente entre dos canciones proporcionadas por un recomendador de música. Una vez realizado el estudio de las canciones y las distintas explicaciones posibles, es necesario mostrar el resultado de nuestro estudio de una forma gráfica, comprensible para un usuario humano.

Los datos que manejamos en nuestro estudio consisten en objetos y las relaciones que existen entre ellos, pues siguen el modelo de datos RDF. Necesitamos representar todos los caminos que se forman entre las dos canciones iniciales, mostrando las conexiones existentes entre los objetos intermedios. Por estos motivos hemos decidido llevar a cabo la visualización mediante un grafo porque consideramos que se adecuaba mejor a nuestras necesidades.

Es importante que la interfaz sea fácil de entender y usar, así que siempre trataremos de mantenerla lo más sencilla posible. La base de nuestro diseño es el grafo de explicaciones ya mencionado, pues es el elemento más importante debido a la gran cantidad de información que aporta y, por lo tanto, es también la parte central de la interfaz alrededor de la cual se posicionarán el resto de elementos.

4.1. Diseño del grafo

En la Figura 4.1 se puede observar la primera iteración de nuestro diseño. En ella se ve un ejemplo de grafo donde los nodos son los distintos sujetos y objetos (según el modelo RDF), mientras que las aristas representan los predicados que relacionan a esos nodos.

Los nodos posicionados a ambos extremos son las canciones sobre las que estamos realizando el estudio, diferenciados por el color rojo. A partir de ellas parten todas las explicaciones. En el centro se puede ver también un nodo coloreado de un rojo más suave, este representa al objeto de una explicación directa (en este caso la explicación “Idioma”).

Los nodos coloreados de color naranja representan los artistas de ambas canciones y las relaciones que existen entre ellos, de la misma forma que en el anterior caso con las canciones y las explicaciones directas. Estas explicaciones obtenidas por el estudio de los artistas son indirectas, ya que requieren un estudio más profundo para relacionar las can-

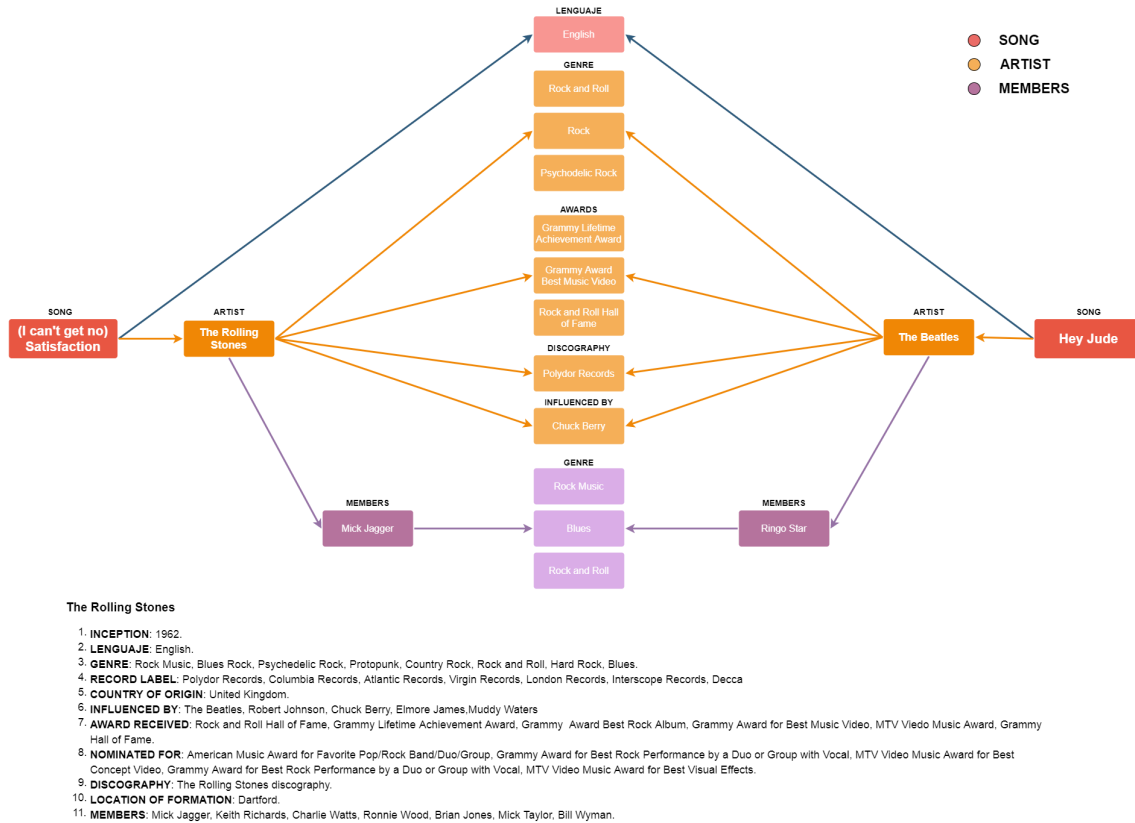


Figura 4.1: Primer diseño de la interfaz web

ciones originales.

Por último tenemos nodos morados que representan a los miembros o integrantes de los artistas, pues en este caso dichos artistas son bandas formadas por varias personas. Su funcionamiento es el mismo que el descrito en el párrafo anterior, con la salvedad de que estos miembros se obtienen del artista y, por lo tanto, sus explicaciones presentan un nivel extra de profundidad.

En esta versión del diseño también se propone una funcionalidad que permite visualizar datos adicionales. Haciendo click en los nodos sobre los que se hacen los estudios (canciones, artistas y miembros), se despliega una sección inferior en la que se muestran todos los datos obtenidos en el estudio del elemento concreto, aparezcan en el grafo o no. Así podemos ver todos los sellos discográficos con los que trabajaron **The Rolling Stones** aunque solo **Polydor Records** los relacione con **The Beatles**.

Esta función no es integral para el objetivo de nuestro proyecto, tan solo es un complemento que ofrece opciones al usuario. Por ello se le ha dado una prioridad baja durante el desarrollo de este trabajo y acabó siendo retirado de diseños posteriores por carecer del interés suficiente para la materia que nos ocupa.

4.2. Diseño de la interfaz

Pasaremos ahora a explicar el resto de la interfaz con la ayuda de la Figura 4.2. En esta iteración, el diseño del grafo se mantiene intacto en la mayor parte. Las novedades a destacar es que se pueden ver las explicaciones indirectas obtenidas por el estudio de los géneros de la canción, representados por nodos de color verde, además de que el nombre de las explicaciones (el predicado según RDF) se traslada a las aristas. (WIP. SEGURAMENTE HAYA QUE CAMBIAR ESTE PÁRRAFO Y/O LA IMAGEN)

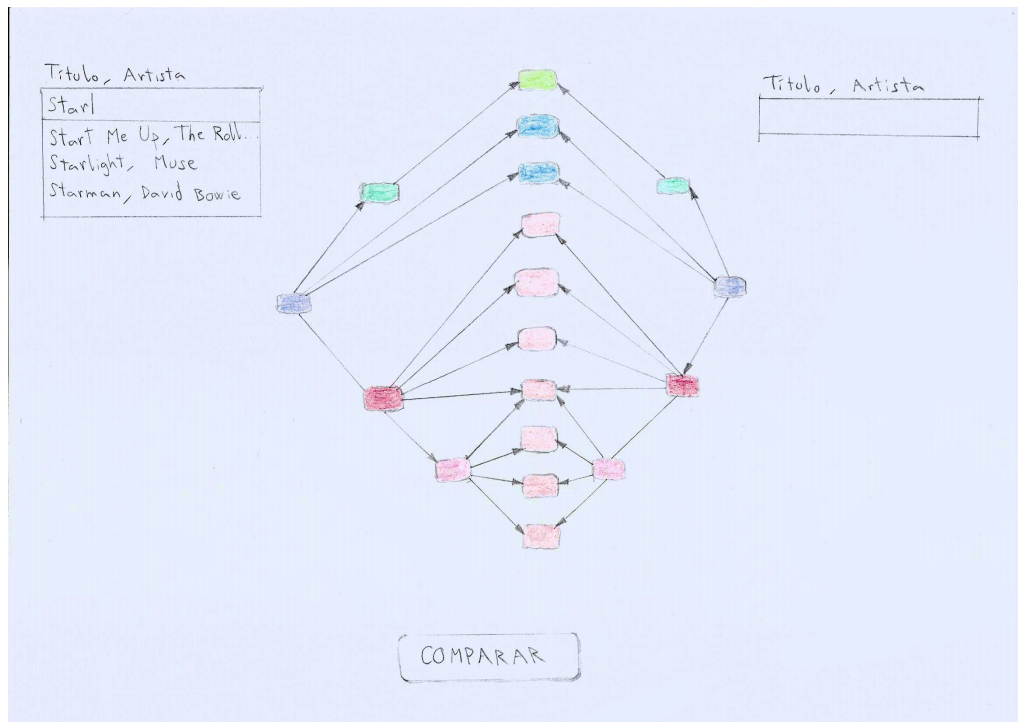


Figura 4.2: Primer diseño de la interfaz web

El resto de la interfaz consiste en los elementos con los que interactuará el usuario antes de generar el grafo de explicaciones. A ambos lados de la pantalla se sitúan los campos a rellenar con los datos básicos de las canciones. Para identificar la canción deseada es necesario escribir en ellos el título de la canción y el nombre de su artista, separados por una coma. En realidad este campo de entrada es un buscador que mostrará en un desplegable la lista de canciones que coinciden con el texto introducido y el usuario deberá seleccionar la que está buscando. Si intenta comparar una canción no contemplada por el sistema, aparecerá un mensaje de error.

La decisión de limitar las posibles entradas ha sido tomada por motivos prácticos. En Wikidata hay una cantidad ingente de temas musicales registrados, sin embargo los datos de cada canción concreta no siempre son tan completos como cabría esperar. A menudo falta información o la que hay no aparece siguiendo los mismos estándares que el resto, especialmente cuando se trata de canciones menos populares. Por ello hemos elaborado una lista de canciones que poseen información útil en Wikidata y que podemos tratar en nuestra aplicación sin problemas.

Por último, en la parte inferior de la pantalla hay un botón para “COMPARAR”. Una vez rellenos los campos de las dos canciones con opciones válidas, el usuario deberá pulsar este botón para que se dibuje el grafo de explicaciones. La primera vez que se use la aplicación no aparecerá ningún grafo en el centro de la interfaz, pero en las comparaciones siguientes el grafo previo no se borrará hasta pulsar de nuevo este botón para dibujar un grafo nuevo.

Capítulo 5

Implementación

5.1. Procesamiento y manejo de la muestra

Partimos de una muestra (o **dataset**) de alrededor de 19 millones de líneas, obtenido a partir de la API de **Last.fm** [5], una plataforma que almacena y proporciona mucho contenido musical. Constaba de los siguientes campos: `<user, timestamp, artist, song>`, los cuales hacen referencia al usuario que ha escuchado la canción, la fecha en la que fue escuchada, el nombre del artista y el título de la canción respectivamente.

El primer paso fue hacer un pequeño estudio del dataset para hacernos una idea de cómo era nuestra muestra y qué podríamos sacar de ella. Se trataba de un dataset con muy pocos campos que no daba información ninguna sobre la canción o el artista, sino que solo proporcionaba un medio para poder obtenerla de una fuente externa. Hicimos una limpieza del dataset, ya que había numerosos valores que eran nulos en determinados campos o no tenían una codificación correcta.

Es en este punto donde comenzamos a pensar cómo queríamos mostrar la información que podíamos ofrecer previa al funcionamiento de la aplicación. Dudábamos entre permitir utilizar cualquier objeto del dataset o restringirlo solo a algunos.

En la librería de Wikidata, por motivos obvios, se debe poner como entrada de cualquier función de búsqueda el código identificador del objeto sobre el que se quiere obtener información. Estos identificadores son fijos y únicos para cada uno de los objetos que están registrados en la página. Con nuestra librería somos capaces de, a partir de un string que represente un título de canción o un artista, género, etc., obtener su respectivo identificador para posteriormente procesar las consultas.

El problema aquí es que para obtener el identificador del objeto, su nombre o título debía ser exacto al que aparecía en Wikidata, pues de cualquier otra forma se lanzaría una excepción. Por ejemplo, intentar obtener el identificador de la canción “Don’t Stop me Now” sería incorrecto, porque en Wikidata figura como “Don’t Stop Me Now”. Debido a esto, es necesario hacer un parseo previo de cualquier String (o cadena de caracteres) que se vaya a utilizar como entrada.

Finalmente decidimos crear una lista preseleccionada y parseada de las canciones más

populares de todo el dataset. Para ello ordenamos las canciones del dataset por popularidad descendente, entendiendo como popularidad la cantidad de veces que aparecían en la muestra. Después elaboramos un script que recorría todas ellas, ejecutaba el parseo y finalmente comprobaba si era posible obtener los datos de Wikidata.

Empezamos con un dataset de las 2500 canciones más populares y fuimos capaces de obtener la información de 1408 canciones con su determinado artista. Cabe señalar que en cada búsqueda de una canción se debe añadir su artista, pues hay varias canciones con el mismo título que no podrían diferenciarse de otro modo.

Nuestra aplicación final funciona con este dataset limitado pero que cuenta con la seguridad de que se pueden obtener datos fiables sin importar la canción que se elija. Además, al haber escogido las canciones con mayor popularidad nos vamos a encontrar con mayor cantidad de datos ya que estas eran las que más documentadas estaban. a diferencia de las que estaban en la parte inferior del dataset, que eran muy poco conocidas y apenas se podían sacar datos valiosos sobre ellas.

5.2. Arquitectura de la aplicación

Como inicio tenemos varios scripts que nos ayudan a organizar y manejar mucho mejor los datos, como por ejemplo el script de limpieza y organización del dataset que, como ya hemos explicado antes, elimina cualquier tipo de valor nulo y lo ordena por popularidad, o el script de parseo, que nos permite acercarnos más a la sintaxis gramatical de Wikidata.

Como punto troncal, hemos desarrollado una librería que nos permite trabajar con la API de Wikidata. Hace uso de SparQL Library, la cual nos permite hacer uso del lenguaje SPARQL y hacer todo tipo de consultas a un punto. Esta librería se encarga de crear un punto de conexión a la API de Wikidata y de crear y ejecutar queries de consulta u obtención de datos.

Como siguiente punto tendríamos el modelo de la aplicación, que es una serie de clases que se encargan de ejecutar todo el proceso de obtención de datos, el cual es genérico para todas las canciones.

Cuenta con una lista de propiedades que son el objeto de estudio de cada canción y métodos para añadir o eliminar alguna de estas. Actualmente cuenta con un número de 21 propiedades totales, las cuales pueden tener su vez aún más de un modo recursivo.

Por último tenemos la clase principal, que es la que se encarga de hacer todas las llamadas al modelo de cada una de las dos canciones y artistas que son introducidos. También se encarga del manejo de excepciones y creación del CSV con las relaciones totales.

Capítulo 6

Conclusiones y Trabajo Futuro

6.1. Trabajo futuro

El esquema RDF y el uso de SPARQL para su uso y análisis puede ser muy potente y abarcar grandes cantidades de datos. En nuestro caso nos hemos centrado en trabajar con la herramienta y los datos de Wikidata por una cuestión de cantidad de información, fiabilidad y resultados obtenidos.

Con Wikidata hemos sido capaces de crear conexiones entre canciones con toda la información posible en cada una de sus respectivas páginas descriptivas. Sin embargo hay más opciones de donde recopilar más información sobre canciones y artistas.

En un principio intentamos obtener información que se alejaba un poco más de lo que hemos creado, que es un análisis musical. Hay otras opciones y datos que se pueden sacar sobre canciones, como por ejemplo las películas o series en las que aparecen como banda sonora, o también canciones que han sido emitidas durante grandes eventos como la Superbowl. El problema que tuvimos es que el proyecto en ese punto se desviaba de su rumbo original, ya que en Wikidata no había datos tan específicos para la gran mayoría de canciones y debíamos recurrir a otras fuentes.

El problema de otras bases de datos alternativas es que tampoco poseen ese tipo de información tan detallada, no siguen el modelo de datos RDF o tienen la información tan solo en casos muy concretos y sin una forma efectiva de acceder a ella a partir de nuestro dataset, como comprobamos al estudiar la posibilidad de utilizar **MusicBrainz** [6] para obtener las bandas sonoras en las que aparece una canción.

Otro aspecto en el que se puede trabajar para llegar a un trabajo mucho más completo y poder hacer una explicación de la relación de dos canciones, sería un estudio a nivel de usuario. En nuestro proyecto hemos hecho un análisis avanzado de las relaciones que tienen las entidades en un ámbito musical. Sin embargo, hay otro estudio posible que es el análisis de los usuarios que han escuchado esas canciones. Partimos de la base de que dos personas que escuchan una misma canción, pueden compartir parte de sus gustos musicales, y se pueden recomendar y crear relaciones conforme a su historial.

Cada usuario tiene un historial de temas escuchados, de hecho uno de los datasets que

obtuvimos al principio del proyecto era de este carácter. Si se hiciese un estudio individual de un usuario, se podrían analizar sus gustos musicales, canciones más escuchadas, últimos géneros más populares, etc. . .

Como ejemplo podemos seleccionar un usuario cuya mayoría de temas escuchados pertenecen al género pop, pero también cuenta con algunos pertenecientes al género indie durante los últimos meses. Si se encuentra una tendencia similar entre una cantidad significativa de usuarios, podríamos sacar una explicación que establezca una relación entre un tema de género pop y otro de género indie por la frecuencia con que aparecen estos géneros juntos entre las listas de temas escuchados de estos usuarios mencionados.

Chapter 7

Introduction

Introduction to the subject area. This chapter contains the translation of Chapter 1.

Chapter 8

Conclusions and Future Work

Conclusions and future lines of work. This chapter contains the translation of Chapter 6.

Bibliografía

*Y así, del mucho leer y del poco dormir, se le
secó el cerebro de manera que vino a perder el
juicio.*

(modificar en Cascaras\bibliografia.tex)

Miguel de Cervantes Saavedra

- [1] T. Berners-Lee, R. Fielding, L. Masinter, et al. Uniform resource identifiers (uri): Generic syntax, 1998.
- [2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001.
- [3] T. Berners-Lee, L. Masinter, M. McCahill, et al. Uniform resource locators (url). 1994.
- [4] W. contributors. Wikidata: Introduction, 2020. Accessed 13-April-2020.
- [5] M. B. T. W. R. J. Felix Miller, Martin Stiksel. Last.fm. Accessed 02-June-2020.
- [6] R. Kaye. Musicbrainz. Accessed 02-June-2020.
- [7] P. Saint-Andre and J. Klensin. Uniform resource names (urns). *Internet Engineering Task Force (IETF), RFC*, 8141, 2017.
- [8] S. Sakr, M. Wylot, R. Mutharaju, D. Le Phuoc, and I. Fundulaki. *Linked Data: Storing, Querying, and Reasoning*. Springer, 2018.

Apéndice **A**

Título del Apéndice A

Contenido del apéndice

Apéndice	B
----------	----------

Título del Apéndice B

Este texto se puede encontrar en el fichero Cascaras/fin.tex. Si deseas eliminarlo, basta con comentar la línea correspondiente al final del fichero TFMTeXiS.tex.

*–¿Qué te parece desto, Sancho? – Dijo Don Quijote –
Bien podrán los encantadores quitarme la ventura,
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

*–Buena está – dijo Sancho –; fírmela vuestra merced.
–No es menester firmarla – dijo Don Quijote–,
sino solamente poner mi rúbrica.*

*Primera parte del Ingenioso Caballero
Don Quijote de la Mancha
Miguel de Cervantes*

