



UNIVERSIDAD PONTIFICIA DE SALAMANCA
FACULTAD DE INFORMÁTICA
Grado en Informática

Trabajo Fin de Grado

Sistema automatizado para la publicación de autoridades y artículos científicos en Wikidata

Brais Rodríguez Calvo

Dr./Dra. Ana M^a Feroso García
Director

M^a Isabel Manzano García
Codirector

Salamanca, Julio de 2021



UNIVERSIDAD PONTIFICIA DE SALAMANCA
FACULTAD DE INFORMÁTICA
Grado en Informática

Trabajo Fin de Grado

Sistema automatizado para la publicación de autoridades y artículos científicos en Wikidata

Brais Rodríguez Calvo

Dr./Dra. Ana Mª Feroso García
Director

Firmado por MANZANO
GARCIA, MARIA ISABEL
(FIRMA) el día 15/07/2021
con un certificado
emitido por AC DNIE 004

Mª Isabel Manzano García
Codirector

Salamanca, Julio de 2021

AGRADECIMIENTOS

Quiero antes de nada transmitir mis más sinceros agradecimientos a todas esas personas que me han ayudado durante toda esta etapa de mi vida.

En primer lugar, a mis directoras del proyecto Ana M^a Feroso García catedrática de la Universidad Pontificia de Salamanca y M^a Isabel Manzano García Directora de la biblioteca de la Universidad Pontificia de Salamanca por toda la información y apoyo recibido para la culminación de este Trabajo de Fin de Grado

En segundo lugar, a mi familia, pero sobre todo a mi madre María y a mi padre Carlos que han estado apoyándome y animándome durante toda mi vida y a mis abuelos que siempre me ayudaron todo lo posible.

Al experto en Wikidata Stefano Bargioni, subdirector de la Biblioteca de la Universidad Pontificia Santa Croce (Roma) y a Camillo Carlo Pellizzari di San Girolamo administrador de Wikidata por todo el apoyo y consejos recibidos en la implementación del proyecto.

Finalmente, a mi universidad la Universidad Pontificia de Salamanca por cambiarme la vida desde el día en que a ella entré y por hacerme sentir como uno más de su familia.

Llevar a cabo este proyecto me ha hecho madurar como estudiante, como futuro trabajador y como persona.

Resumen

El propósito de este proyecto es obtener un método automatizado para la publicación en la base de datos de Wikidata, de autoridades y registros bibliográficos correspondientes a PDI de la Universidad Pontificia de Salamanca (UPSA). Wikidata constituye una de las bases de conocimiento abierto y global más importantes y es la fuente de conocimiento de la famosa Wikipedia. Concretamente la publicación se concretará en autoridades (PDI) de la Universidad Pontificia de Salamanca y además en la incorporación de los artículos que ese PDI de la UPSA ha escrito para la revista “Cuadernos Salmantinos de Filosofía”. Esta revista se trata de la de mayor calidad de la UPSA, estando en diversos rankings internacionales y en múltiples bases de datos especializadas. De este modo tanto las personalidades y/o autores de los artículos de la revista, y esos artículos, adquieran una mayor transcendencia, visibilidad y apertura al entrar en una fuente de conocimiento tan globalizada, abierta, globalizada, democrática y sostenible como es la familia Wikimedia. Para llevarlo a cabo, se extraerán los autores de nuestra Universidad a través de los ficheros aportados por la biblioteca y de los artículos de la revista publicada por nuestra Universidad a través de una serie de peticiones a servicios web.

Seguidamente, se combinarán los resultados y mediante un algoritmo se obtendrán las autoridades que se deben subir a Wikidata. Para el registro en Wikidata se utilizarán tres herramientas externas. La primera será Mix'n'match, la cual nos servirá para identificar, obtener o crear los elementos que debemos añadir a la fuente de conocimiento. La segunda es OpenRefine, una aplicación que nos servirá para realizar operaciones (actualizaciones) sobre los datos que finalmente se subirán. También se podrán subir directamente los datos desde esta aplicación, sin embargo, usaremos otro sistema o herramienta más eficiente a la hora de realizar esta tarea, sobre todo, para el caso de que se obtenga un gran volumen de entradas llamado QuickStatements.

Abstract

The purpose of this project is to obtain an automated method for publication in the Wikidata database of authorities and bibliographic records corresponding to PDIs of the Pontifical University of Salamanca (UPSA). Wikidata constitutes one of the most important open and global knowledge bases and is the source of knowledge for the famous Wikipedia. Specifically, the publication will take place in authorities (PDI) of the Pontifical University of Salamanca and also in the incorporation of the articles that this PDI of the UPSA has written for the magazine “Cuadernos Salmantinos de Filosofía”. This magazine is the highest quality of the UPSA, being in various international rankings and in multiple specialized databases. In this way, both the personalities and / or authors of the journal articles, and those articles, acquire greater significance, visibility and openness by entering a source of knowledge as globalized, open, globalized, democratic and sustainable as is the family. Wikimedia. To carry it out, the authors of our University will be extracted through the files provided by the library and from the articles in the journal published by our University through a series of requests to web services.

Then, the results will be combined and through an algorithm the authorities that must be uploaded to Wikidata will be obtained. For this, three external tools will be used. The first one will be Mix'n'match which will help us to identify, obtain or create the elements that we must

add to the source of knowledge. The second one is OpenRefine, an application that will help us to perform operations on the data that will finally be uploaded. It can also be uploaded the data directly from this application, however it will be used another more efficient system when carrying out this task, especially for a large volume of entries called QuickStatements..

Descriptores

Mix'n'match, OpenRefine, QuickStatement, Autoridades, Registro Bibliografico, Wikidata, Java, Autoridades UPSA, Autores Revista Cuadernos Salmantinos.

Keywords

Mix'n'match, OpenRefine, QuickStatement, Authorities, Bibliographic, Wikidata, Java, UPSA Authorities, Authors of the Magazine "Cuadernos Salmantino

Contenido

1.	INTRODUCCIÓN	1
1.1	MOTIVACIÓN.....	1
1.2	OBJETIVOS	1
1.3	ESTRUCTURA DOCUMENTO	1
2.	BASES TEÓRICAS.....	3
2.1	WIKIDATA	3
2.2	HERRAMIENTAS DE WIKIDATA.....	6
2.3	HERRAMIENTA MIX'N'MATCH	8
2.3.1	FUNCIONAMIENTO.....	8
2.3.2	PÁGINA PRINCIPAL	10
2.3.3	PÁGINA IMPORTACIÓN O ACTUALIZACIÓN CATÁLOGOS	11
2.3.4	FORMATO DEL ARCHIVO VÁLIDO PARA IMPORTAR	13
2.3.5	VINCULACIÓN DE LAS ENTRADAS	18
2.3.6	MÉTODOS DE TRABAJO PARA VINCULAR ENTRADAS A WIKIDATA	21
2.3.7	TRUCOS PARA HACER MÁS RÁPIDA LA VINCULACIÓN	24
2.3.8	RETOMAR EL TRABAJO	25
2.3.9	EXTRACCIÓN DE LOS DATOS NO EXISTENTES DE WIKIDATA	26
2.3.10	VALORACIÓN DE LA HERRAMIENTA.....	28
2.4	HERRAMIENTA OPENREFINE	29
2.4.1	PRINCIPALES FUNCIONALIDADES.....	30
2.4.2	CONCILIACIÓN DE LOS DATOS	31
2.4.3	TRABAJAR CON FACETAS.....	38
2.4.4	EDITAR COLUMNAS.....	42
2.4.5	TRANSFORMACIÓN DE STATEMENTS Y SUBIDA A WIKIDATA	47
2.4.6	VALORACIÓN DE LA HERRAMIENTA.....	52
2.5	HERRAMIENTA OPENREFINE	53
2.5.1	MÉTODO DE TRABAJO.....	54
3.	DESARROLLO Y APLICACIÓN PRÁCTICA.....	59
3.1	DESCRIPCIÓN DEL CASO DE USO PARA NUESTRA BIBLIOTECA.....	60
3.2	DESCRIPCIÓN DEL CASO DE USO PARA NUESTRA BIBLIOTECA.....	61
3.3	APLICACIÓN DEL PROCESO DE MIGRACION A WIKIDATA	62
3.3.1	FASE 1: PREPARACIÓN INICIAL DE LOS DATOS.....	63
3.3.2	FASE 2: FASE DE GENERACIÓN DE ENRIQUECIMIENTO Y CREACIÓN DE LAS AUTORIDADES DE LA UNIVERSIDAD PONTIFICIA DE SALAMANCA	68

3.3.3	FASE 3: FASE DE CREACIÓN DE LOS ARTICULOS PUBLICADOS POR LAS AUTORIDADES DE LA UPSA	83
3.4	DESCRIPCIÓN DE LA APLICACIÓN JAVA.....	89
4.	RESULTADOS.....	93
4.1	APLICACIÓN DEL PROCESO DE MIGRACION A WIKIDATA	94
5.	CONCLUSIONES	97
6.	LÍNEAS FUTURAS	98
7.	BIBLIOGRAFÍA	99

Ilustración 2-1 Esquema del mecanismo de Wikidata	4
Ilustración 2-2 Instancias de tipo para la ciudad de San Francisco.....	4
Ilustración 2-3 Instancias de la propiedad "Head of Government" de la ciudad de San Francisco	5
Ilustración 2-4 Guía de herramientas de Wikidata (https://www.wikidata.org/wiki/Wikidata:Tools)	6
Ilustración 2-5 Categoría Editar Elementos (https://www.wikidata.org/wiki/Wikidata:Tools/Edit_items)	6
Ilustración 2-6 Ventana creación nuevo usuario	7
Ilustración 2-7 Catalogo Netflix FR (https://mix-n-match.toolforge.org/#/catalog/2179)	9
Ilustración 2-8 Mix'n'match (https://mix-n-match.toolforge.org/#/).....	10
Ilustración 2-9 Pantalla de importación parte 1 (https://mix-n-match.toolforge.org/#/import/)	11
Ilustración 2-10 Pantalla de importación parte 2 (https://mix-n-match.toolforge.org/#/import/)	12
Ilustración 2-11 Ejemplo de las columnas asimiladas por mix'n'match en la importación...	14
Ilustración 2-12 Ejemplo datos para ser importados a los catálogos	15
Ilustración 2-13 Unos catálogos de prueba creados en la herramienta (Aparecen marcados como [TEST])	16
Ilustración 2-14 Catálogo de Prueba autoridades UPSA.....	17
Ilustración 2-15 Catálogo recién creado	18
Ilustración 2-16 Ejemplo de elementos de la lista de matches automáticos	19
Ilustración 2-17 Filosofía de trabajo Mix'n'match.....	21
Ilustración 2-18 Pantalla modo juego parte 1	22
Ilustración 2-19 Pantalla modo juego parte 2	22
Ilustración 2-20 Pantalla modo lista	23
Ilustración 2-21 Menú de descarga de las entradas de Mix'n'match	27
Ilustración 2-22 Muestras del archivo json descargado	27
Ilustración 2-23 Pantalla principal de la aplicación de navegador de OpenRefine	31
Ilustración 2-24 Ventana de inserción de datos de OpenRefine	32
Ilustración 2-25 Opciones de cotejamiento de datos OpenRefine	33
Ilustración 2-26 Elección de la base para reconciliar los datos	33
Ilustración 2-27 Opciones para cotejar una columna OpenRefine	34
Ilustración 2-28 Entradas después de ser cotejadas.....	35
Ilustración 2-29 Resultado tras la reconciliación	35
Ilustración 2-30 Tabla principal de datos sobre el objeto.....	36
Ilustración 2-31 Búsqueda en Wikidata del elemento a identificar	36
Ilustración 2-32 Entradas después de reconciliar los datos.....	37
Ilustración 2-33 Facetas sobre columnas reconciliadas.....	38
Ilustración 2-34 Faceta por parámetro	38
Ilustración 2-35 Faceta por puntuación	39
Ilustración 2-36 Todas las facetas sobre datos conciliados	39
Ilustración 2-37 Opciones de facetas generales con respecto a una columna cualquiera	40
Ilustración 2-38 Faceta de texto estándar	40
Ilustración 2-39 Faceta por duplicados.....	41
Ilustración 2-40 Faceta por longitud de texto.....	41
Ilustración 2-42 Menú de división de columnas	42
Ilustración 2-43 Utilización de GREL para editar una columna.....	43
Ilustración 2-44 Ventana de cambio de nombre de columna	43
Ilustración 2-45 Opciones de edición de columnas	44

Ilustración 2-46 Ventana emergente con las columnas de datos extraídos de Wikidata.....	45
Ilustración 2-47 Acceso al menú de funciones de Wikidata	47
Ilustración 2-48 Pantalla de inicio de Schema	48
Ilustración 2-49 Schema enriquecimiento de identificador UPSA para autoridades.....	48
Ilustración 2-50 Ejemplo de creación de nuevos elementos en OpenRefine	49
Ilustración 2-51 Lista de errores elementos nuevos	50
Ilustración 2-52 Ejemplo de previsualización sobre las ediciones	50
Ilustración 2-53 Menú desplegable de opciones de Wikidata	51
Ilustración 2-54 Ejemplo de los comandos de las ediciones sobre un elemento en el archivo Quickstatements que obtenemos al exportar de OpenRefine.....	55
Ilustración 2-55 Página principal QuickStatements	55
Ilustración 2-56 Ejemplo de formato de comandos importados en Quickstatements	56
Ilustración 3-2 Estructura archivo json de publicaciones.....	64
Ilustración 3-3 Muestra del archivo json con las publicaciones de la revista	65
Ilustración 3-4 Archivo original de autoridades de la UPSA.....	66
Ilustración 3-5 Datos de las autoridades en formato csv totalmente limpios	67
Ilustración 3-6 Menú de operaciones (PROVISIONAL)	68
Ilustración 3-7 Vista de operaciones con las autoridades de la aplicación(PROVISIONAL)	69
Ilustración 3-8 Muestra del formato del archivo tsv para importar los datos a mix'n'match	69
Ilustración 3-9 Ventana de importación de Autoridades Mix'n'match parte 1	70
Ilustración 3-10 Ventana de importación de Autoridades Mix'n'match parte 2	71
Ilustración 3-11 Testeo de columnas de datos	71
Ilustración 3-12 Testeo de datos importados	72
Ilustración 3-13 Visión del catálogo recién creado	72
Ilustración 3-14 Lista de entidades ligadas automáticamente	73
Ilustración 3-15 Ejemplo de entrada de autoridad	73
Ilustración 3-16 Ejemplo de entidad ligada automáticamente errónea	74
Ilustración 3-17 Resultado después de gestionar los elementos ligados automáticamente	74
Ilustración 3-18 Menú de descarga de información catálogo	75
Ilustración 3-19 Vista de operaciones con las autoridades de la aplicación(PROVISIONAL)	76
Ilustración 3-20 Muestra de los datos del archivo de enriquecimiento para OpenRefine ..	76
Ilustración 3-21 Proceso de creación de nuevos elementos	77
Ilustración 3-22 Muestra del dato creado en Wikidata	77
Ilustración 3-23 Estado final del catálogo	78
Ilustración 3-24 Formato de cotejamiento	78
Ilustración 3-25 Resultado tras reconciliar Autoridades a enriquecer	79
Ilustración 3-26 Elementos no conciliados	79
Ilustración 3-27 Schema de enriquecimiento de autoridades	82
Ilustración 3-28 Resultado de ejecutar comandos en Quickstatements	82
Ilustración 3-29 Ventana operaciones con publicaciones de la aplicación (PROVISIONAL) ..	84
Ilustración 3-30 Formato de salida del archivo tsv de autoridades y artículos	84
Ilustración 3-31 Cotejado de las publicaciones	85
Ilustración 3-32 Entradas seleccionadas para su creación	85
Ilustración 3-33 Schema de creación de artículos	88
Ilustración 3-34 importación de los artículos mediante Quickstatements	89
Ilustración 3-35 Esquema de las partes de la aplicación	90
Ilustración 3-36 Paquete funcionalidad csv	90

Ilustración 3-37 Esquema funcionalidad json.....	91
Ilustración 3-38 Esquema de funcionalidad TSV	91
Ilustración 3-39 Esquema interfaz de usuario	91
Ilustración 3-40 Conjunto de clases usadas.....	92
Ilustración 3-41 Esquema operaciones generales	92
Ilustración 4-1 Ventana de inicio aplicación	94
Ilustración 4-2 Ventana de Operaciones con las Autoridades.....	95
Ilustración 4-3 Ventana de operaciones con publicaciones de la revista	96

1. INTRODUCCIÓN

1.1 MOTIVACIÓN

El agente motivador de este proyecto no es sino otro que fomentar la interoperabilidad de la información propia con otras instituciones, abriendo nuestras puertas al mundo a través de Wikidata para fomentar a los autores y personalidades de nuestra Universidad y mostrar sus logros y labores

1.2 OBJETIVOS

El objetivo de este proyecto es la publicación referente a la Biblioteca de la Universidad Pontificia de Salamanca, de autores propios y ajenos a la misma, en la base de datos de Wikidata para fomentar el conocimiento interconectado con otras organizaciones académicas similares o instituciones que necesiten información en ese dominio.

Este objetivo principal a su vez incluye dos principales objetivos derivados:

- Estudiar, analizar, valorar y aprender a utilizar las principales herramientas existentes para la publicación efectiva de datos en Wikidata para su difusión y enriquecimiento.

Diseñar el proceso que permita optimizar y conseguir la publicación de la forma más automatizada posible, las autoridades de nuestra universidad y de cualquier otra referenciada en la revista que se toma como caso de uso o aplicación práctica, “Cuadernos Salmantinos de Filosofía”, aunque igualmente extrapolable a otras publicaciones.

1.3 ESTRUCTURA DOCUMENTO

Este documento se distribuye en seis extractos bien diferenciados. Comienza con la exposición de las bases teóricas en las que se basan las herramientas Wikidata necesarias para realizar las ediciones de los datos en su base de datos. Se hablará de la base de datos de Wikipedia (Wikidata), de las funcionalidades de Mix'n'match y OpenRefine, así como el uso que le vamos a dar a la herramienta QuickStatements.

Le sigue un desarrollo de nuestro proyecto más en detalle con todas las herramientas nombradas el cual dividiremos en tres fases, así como la implementación de una aplicación en lenguaje java que permite la transformación al formato correcto de todos los datos usados por las aplicaciones antes nombradas.

También se hará una exposición de los resultados del proyecto donde se mostrará lo que se ha conseguido, las conclusiones a las que llegamos después de realizar todo este trabajo y sus líneas futuras y mejoras posibles que pudiesen incluirse en un futuro.

Finalmente, toda la bibliografía de referencia del proyecto que garantiza el acceso a una mayor información sobre el tema.

2. BASES TEÓRICAS

Para conseguir los objetivos propuestos lo primero es adquirir conocimiento tanto de la base de datos Wikidata, como de las herramientas a usar para publicar en ella, así como de la propia estructura y metodología seguida para publicar los elementos.

En este apartado trataremos de todo ello y así poder adquirir los conocimientos que necesitamos para su posterior puesta en práctica.

2.1 WIKIDATA

Pertenece a la familia de Wikimedia y es hermana de la más conocida Wikipedia que junto con otras como Wikinoticias, Wikidiccionarios, Wikilibros... conforman el grupo de Wikimedia. Wikidata es por tanto una base de conocimiento *“editada en colaboración”*, eso quiere decir que son los usuarios los que mediante el navegador crean, modifican y eliminan el contenido. Como norma general de todas las wikis la información está plagada de referencias a otros elementos que tienen sus propias páginas y continuando así sucesivamente, creando una red interconectada de conocimiento.

Wikidata en concreto es la fuente de conocimiento para otras wikis como Wikipedia y opera con el sistema Wikibase que es un conjunto de extensiones de Mediawiki que permiten trabajar con un repositorio central. Estos vínculos que se generan entre los distintos bloques de datos permiten que la información solo tenga que ser añadida una vez y será disponible para todos. Cada elemento posee un identificador único *“Q Number”* (con formato Qxxx donde las xxx representan un número entero que identifica el tipo de objeto en cuestión) al que se le atribuyen otra serie de propiedades que caracterizan y permiten crear una gran red interconectada. Estos vínculos son muy útiles porque nos permiten realizar operaciones como las consultas con SPARQL, este es un lenguaje de consultas para bases de datos de conocimiento aplicable a Wikidata, para una mayor comprensión sobre el tema ver referencia de la bibliografía (1).

Un ejemplo del mecanismo funcional de Wikidata puede ser el que aparece en la siguiente ilustración Ilustración 2-1 Esquema del mecanismo de , en la que se puede apreciar distintos elementos Wikidata, cada uno representado por instancias de tipo. En el centro de la imagen se puede ver el elemento de Wikidata San Francisco, que representa a la ciudad americana de identificador Q62. Presenta a su vez una serie de propiedades como su geolocalización y el número de habitantes o la cabeza de gobierno, por otro lado, también está vinculada a por ejemplo, su estado California y este a su vez a su nación EEUU.

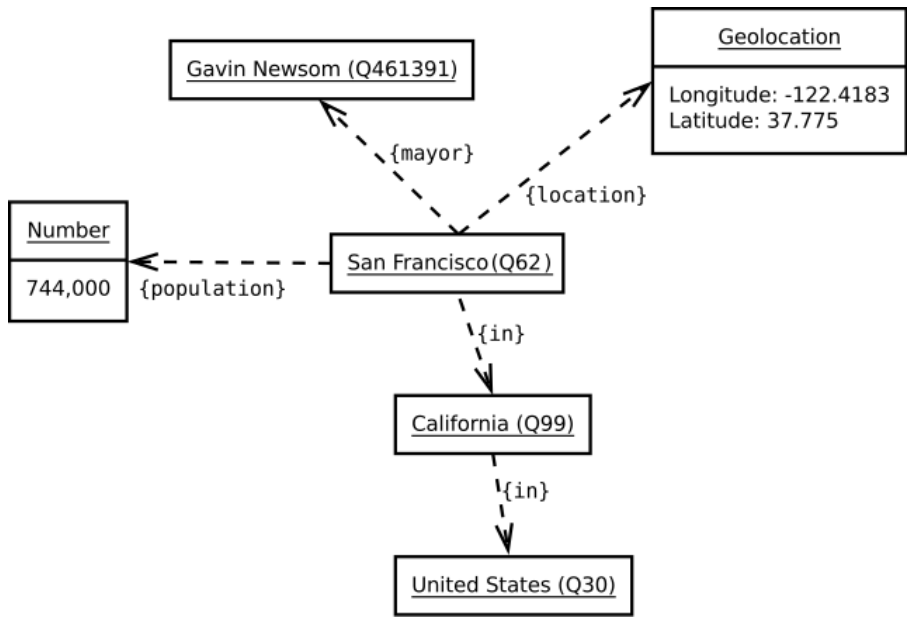


Ilustración 2-1 Esquema del mecanismo de Wikidata

De esta forma es como se va hilando la red que conforma esta base de conocimiento. En la siguiente imagen Ilustración 2-2 Instancias de tipo para la ciudad de San Francisco, se puede ver un ejemplo de las instancias de tipo que tiene la ciudad de San Francisco (Q62)

Statements	
instance of	<div><div><div><div></div><div>consolidated city-county</div></div><div>2 references</div></div></div>
	<div><div><div><div></div><div>city of the United States</div></div><div>4 references</div></div></div>
	<div><div><div><div></div><div>sanctuary city</div></div><div>2 references</div></div></div>
	<div><div><div><div></div><div>city</div></div><div>1 reference</div></div></div>
	<div><div><div><div></div><div>big city</div></div><div>1 reference</div></div></div>
	<div><div><div><div></div><div>charter city</div></div><div>1 reference</div></div></div>

Ilustración 2-2 Instancias de tipo para la ciudad de San Francisco

Un ejemplo de las instancias de las propiedades de la ciudad se puede ver en la siguiente imagen Ilustración 2-3 Instancias de la propiedad "Head of Government" de la ciudad de San Francisco





head of government	 Ed Lee
	start time 11 January 2011
	end time 12 December 2017
	▼ 0 references
	 Gavin Newsom
	start time 8 January 2004
	end time 10 January 2011
	▼ 0 references
	 Willie Brown, Jr.
	start time 8 January 1996
	end time 7 January 2004
	▼ 0 references
	 Frank Jordan
	start time 8 January 1992
	end time 7 January 1996

Ilustración 2-3 Instancias de la propiedad "Head of Government" de la ciudad de San Francisco

Por un lado, tenemos el país que contiene al estado, que a su vez contiene a la ciudad y esta tiene propiedades como cualquier otro elemento ciudad puede tener.

Todos estos datos pueden desplegarse en sus 300 lenguajes, por esta razón puede ser muy útil utilizar esta información para lanzar nuevas aplicaciones y páginas webs y enlazar y enriquecer la información. A partir de este punto es donde empieza lo más importante de este proyecto, las herramientas que nos van a permitir editar todos estos datos.

2.2 HERRAMIENTAS DE WIKIDATA

Wikidata posee una gran lista de herramientas que facilitan su utilización haciendo que sea más rápido y sencillo su uso como podemos ver en la siguiente imagen, Ilustración 2-4 Guía de herramientas de Wikidata (<https://www.wikidata.org/wiki/Wikidata:Tools>).



Ilustración 2-4 Guía de herramientas de Wikidata (<https://www.wikidata.org/wiki/Wikidata:Tools>)

Como podemos ver en su página oficial (2) tenemos acceso a ellas a través de una serie de categorías que nos permiten diferenciarlas según su funcionalidad, o también podemos acceder a la documentación existente para ellas a través del buscador con el nombre en cuestión.

Por ejemplo, si lo que queremos es acceder a la documentación OpenRefine podemos acceder por la categoría de “*edición de elementos*” y dentro de esa lista podremos ver una pequeña descripción sobre la herramienta de edición y fotografías del entorno. Si hacemos click veremos la documentación como se puede ver en la Ilustración 2-5 Categoría Editar Elementos (https://www.wikidata.org/wiki/Wikidata:Tools/Edit_items).

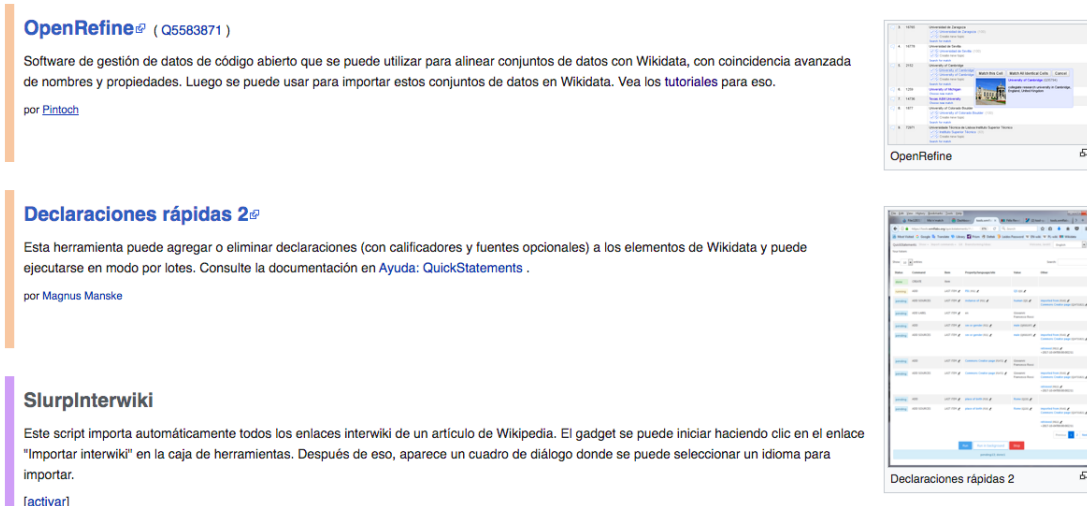


Ilustración 2-5 Categoría Editar Elementos (https://www.wikidata.org/wiki/Wikidata:Tools/Edit_items)

Después de realizar una búsqueda para encontrar el software que resuelva nuestras necesidades, se llega a la conclusión de que la mejor opción para la búsqueda e identificación de objetos Wikidata existentes resulta ser Mix'n'match (esta herramienta da la opción de crear nuevos elementos, pero solo es recomendable en el caso de que ya existan la mayoría de elementos y solo queramos añadir unos pocos, ya que deberá hacerse manualmente como se explicará a continuación); para la limpieza, importación y manipulación de los datos además de su importación a la base de datos de conocimiento consideramos que la mejor opción es OpenRefine (como se explicará con un mayor detenimiento más adelante OpenRefine se encarga de transformar de forma interactiva nuestras órdenes a comandos Quickstatements, que después importaremos por esta herramienta)- Por último en nuestro caso trabajaremos con un gran volumen de datos por lo que la opción más recomendable para hacer la importación de datos a Wikidata es QuickStatements

Para tener acceso a las herramientas de Wikimedia es necesario que nos registremos como usuarios del sistema. Una vez hagamos eso podremos realizar ediciones sobre la información. Si necesitamos generar un nuevo usuario en el sistema podemos hacerlo en (3) de la forma que se puede ver en la Ilustración 2-6 Ventana creación nuevo usuario

Create account

Language: [Deutsch](#) | [English](#) | [Esperanto](#) | [français](#) | [español](#) | [italiano](#) | [Nederlands](#)

Username [\(help me choose\)](#)

Password

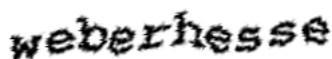
It is recommended to use a unique password that you are not using on any other website.

Confirm password

Email address (optional)

To protect the wiki against automated account creation, we kindly ask you to enter the words that appear below in the box [\(more info\)](#):

CAPTCHA Security check



[Refresh](#)

[Can't see the image?](#)

Wikidata is made by people like you.



1,458,702,264

edits



94,180,721

items



23,064

recent contributors

Ilustración 2-6 Ventana creación nuevo usuario

2.3 HERRAMIENTA MIX'N'MATCH

Es una herramienta de importación de conjuntos de datos, creada por Magnus Manske un científico y desarrollador de software de una de las primeras versiones del software de Media Wiki.

La aplicación está plagada de catálogos (posee una cantidad de casi 5000 conjuntos de datos, un número que varía constantemente). Estos catálogos son amplias listas de datos vinculados con entradas de Wikidata. Su función principal es la de identificar cuáles existen ya en la base de datos de Wikipedia y cuales aún no tienen ninguna entrada. El proceso de coincidencia se puede llevar a cabo por varios métodos distintos pero el objetivo final siempre es “*hacer matches*”. Entendemos por hacer un match el hecho de identificar un elemento existente sin la mayor probabilidad posible de coincidencia con los posibles registros identificados por similitud, en Mistmatch

2.3.1 FUNCIONAMIENTO

La herramienta divide los elementos importados a partir del raspado o extracción automática desde una página web, Google sheets o de un archivo en disco local, en cinco categorías de trabajo

- **Ligado Automáticamente:** (EN: **Automatically Matched**) este es el caso en el que el sistema es capaz de asociar esa entrada a una o varias coincidencias de Wikidata, estas selecciones aparecerán por orden de similitud.
- **Ligado Manualmente:** (EN: **Manually Matched or Fully Matched**) ocurre cuando un usuario crea un match manualmente, vincula el elemento con uno de Wikidata (bien sea porque existe y lo ha identificado correctamente o porque ha decidido crear uno nuevo).
- **No aplicable a Wikidata:** (EN: **Not Applicable to Wikidata**) son entradas no útiles para el sistema, como duplicados, redirecciones o temas inapropiados.
- **Sin match:** (EN: **Unmatched**) on las entradas más interesantes y que llevan una mayor labor de investigación, dado que automáticamente no se le ha vinculado a ningún dato Wikidata, pero cabe la posibilidad de que existan y que el sistema no haya podido identificar correctamente, en cuyo caso si se requiere hacer una búsqueda más detallada. Sin embargo, no suele ser el caso ya que los elementos a parte de su identificador Wikidata cuentan por lo general con propiedades que son identificadores únicos en sí mismos como por ejemplo el VIAF para los autores, que nos sirve para aliviar cualquier duda acerca de su identidad.

Toda la información sobre la funcionalidad de la herramienta se puede encontrar en Wikidata en el apartado siguiente (4)

El catálogo que se muestra en la Ilustración 2-7 Catalogo Netflix FR (<https://mix-n-match.toolforge.org/#/catalog/2179>) es un ejemplo de un catálogo que se encuentra

visible públicamente en la herramienta, que ya ha sido creado y trabajado por distintos usuarios de la comunidad, porque como se puede ver la lista de elementos ligados manualmente está muy avanzada. Éste es el formato de los catálogos de Mix`n`match

Netflix FR

[Action ▾](#)

Netflix movies and TV shows available in France in January 2019

Importado por Tpt | [Actualizar](#)

Entradas

Ligado manualmente	3541	94.7%
Ligado automáticamente	24	0%
No Wikidata	0	0%
Not applicable to Wikidata	0	0%
Unmatched	173	4.6%
Total	3738	

Ilustración 2-7 Catalogo Netflix FR (<https://mix-n-match.toolforge.org/#/catalog/2179>)

2.3.2 PÁGINA PRINCIPAL

La herramienta está disponible en la dirección <https://mix-n-match.toolforge.org/#/>. A primera vista en su página principal podemos ver de forma recurrente la palabra “Catalogo” como se puede observar en la *Ilustración 2-8 Mix’n’match* (<https://mix-n-match.toolforge.org/#/>)

Los catálogos no son más que bloques de entradas de un mismo tipo. Podemos acceder a nuestros catálogos de forma específica usando la barra de búsqueda escribiendo el título del catálogo por un identificador que lo identifica directamente y que se le confiere en el momento de su creación. Para una visión más general es muy cómodo navegar por las distintas categorías de catálogos, que podemos ver en la parte izquierda de la imagen. También podemos observar las últimas actualizaciones en la lista de “Últimos catálogos” donde aparecen los que han sido creados recientemente.

Nuestro proyecto se centrará en la importación y extracción de los datos y que después de procesados procederemos con ellos de una forma o de otra.

The screenshot shows the Mix'n'match website interface. At the top, there is a navigation bar with 'Mix'n'match', a language dropdown set to 'español', a globe icon, a 'Log in for actions' link, and a search bar with 'Buscar' buttons. Below the navigation bar, there is a descriptive paragraph about the tool's functionality. The main content area is divided into two sections: 'Catalog groups' on the left and 'Search catalogs' on the right. The 'Catalog groups' section contains a table with two columns: 'Agrupar' and 'Catálogos'. The 'Search catalogs' section has a search input field and a list of 'Latest catalogs' with links to various catalog entries.

Agrupar	Catálogos
	8
Cinema	49
Music	11
Software	14
Wolfram Language entity	85
Archives	27
Art	126

Search catalogs

Start typing here

Latest catalogs

- [Women in Gaming: 100 Professionals of Play](#) people profiled in the "Women in Gaming: 100 Professionals of Play" book
- [Rekhta author](#) identifier for a author in the Rekhta database of literature of the Urdu language of India and Pakistan
- [Russians of Latvia](#) identifier for a person in the website Russians of Latvia
- [Biographia Benedictina](#) identifier for a Benedictine monk in the Biographia Benedictina
- [Brapci author](#) identifier for an author in the Brazilian database Brapci
- [Discogs music styles](#) List of music styles on Discogs
- [Visuotinė lietuvių enciklopedija](#) identifier for an item in the online edition of the Visuotinė lietuvių enciklopedija
- [AllMusic music genres](#) List of music genres and styles on AllMusic

Ilustración 2-8 Mix’n’match (<https://mix-n-match.toolforge.org/#/>)

2.3.3 PÁGINA IMPORTACIÓN O ACTUALIZACIÓN CATÁLOGOS

Podemos ver el panel de importación de datos de la herramienta en la Ilustración 2-9 Pantalla de importación parte 1 (<https://mix-n-match.toolforge.org/#/import/>) y también en la Ilustración 2-10 Pantalla de importación parte 2 (<https://mix-n-match.toolforge.org/#/import/>)

Desde el principio en la pantalla de importación nos da la opción de utilizar un catálogo existente añadiendo su identificador correspondiente, lo cual nos valdría para actualizar el catálogo con los nuevos datos que estamos transfiriendo), o crear directamente un nuevo catálogo, para lo cual dejaremos ese apartado en blanco.

Los siguientes apartados serán el nombre del catálogo que vamos a crear la descripción del mismo, esta descripción se mostrara en la herramienta junto al título, lo cual permitirá a los usuarios que estén haciendo matches saber un poco más sobre el contenido del catálogo.

Les sigue la URL en el caso de que queramos vincular los datos importados con referencias de alguna dirección web, en este caso biblioteca.upsa.es.

La WD property o propiedad de Wikidata que va ser usada como identificador de las entradas del catálogo.

El type o tipo que nos va a permitir incluir el catálogo en una de las categorías que nombramos antes.

El primary language o lenguaje primario que usaran las entradas del catálogo (suele dejarse como inglés por defecto) y el tipo de entrada por defecto

Estos últimos elementos del formulario de importación son opcionales.

Import or update catalog

Catálogo

Update an existing catalog (give ID), or create a new catalog (leave ID empty).
Note: if you enter the property first and then click another field, some information will be filled in automatically.

Identificador del catálogo Leave empty to create a new catalog. If given, other data in this section will be ignored.

Nombre del catálogo

Descripción

URL optional

WD property optional Type

Primary language Default entry type optional

Ilustración 2-9 Pantalla de importación parte 1 (<https://mix-n-match.toolforge.org/#/import/>)

The screenshot shows a web form titled 'Data' with a note about using Google Sheets. It contains several sections: 'Data format' with radio buttons for 'Comma-separated values' (selected) and 'Tab-separated values'; 'Source' with radio buttons for 'URL' (selected) and 'File', and a 'Log in for actions' link; 'Updating' with checkboxes for 'Add new entries' (checked), 'Update existing descriptions', and 'Update all descriptions'; and 'Looks good' with an unchecked checkbox. There are also radio buttons for update frequency: 'No regular update' (selected), 'Update weekly', 'Update monthly', and 'Update once a year'.

Ilustración 2-10 Pantalla de importación parte 2 (<https://mix-n-match.toolforge.org/#/import/>)

Seguimos con el apartado del formato de los datos, tanto si creamos desde cero el catálogo como si lo actualizamos deberemos añadirle una serie de datos. Los datos se pueden importar con un archivo de formato .csv (separado por comas) o .tsv. En estos archivos hay siempre una única fila de encabezado. Para evitar problemas en los datos importados es recomendable utilizar el formato .tsv en lugar del .csv ya que si entre los datos existen códigos, identificadores o palabra que lleven “,” entre sus caracteres, producirán errores en las entradas. También se da la opción de usar la URL de su página, donde se producirá periódicamente el rascado de datos de forma automatizada o las hojas de cálculo de Google.

En nuestro caso que usaríamos los datos locales de nuestro ordenador, la fuente que debemos escoger es file, y deberemos importar el fichero mediante el explorador de archivos.

Deberemos indicar también si lo que queremos es crear nuevas entradas, actualizar descripciones existentes o actualizar todas las descripciones.

Para finalizar si creemos que todo lo que hemos hecho deberemos pulsar el checkbox looks Good, para poder continuar con el proceso.

En nuestro proyecto usaremos los datos que nos aporta la biblioteca sobre las Autoridades de nuestra universidad por medio de un fichero independiente, por lo que la opción recomendable será el archivo tsv o csv.

2.3.4 FORMATO DEL ARCHIVO VÁLIDO PARA IMPORTAR

Cuando se importan los datos a Wikidata desde un .csv o .tsv (la diferencia entre uno y otro es simplemente el separador que se usa a la hora de delimitar las columnas de datos en el primero es una "," o ";" y en el segundo caso es un tabulador).

En nuestro caso será el formato tsv porque garantiza una mayor corrección a la hora de que la herramienta analice la información. Todo formato csv o tsv debe de tener una línea inicial que defina los nombres de las columnas, para poder diferenciar unos datos de otros como si de una tabla se tratara. Por lo tanto, el archivo que vamos a usar deberá de tener un formato específico.

Será del todo obligatorio que aportemos una columna de identificador (con nombre *id*) para la entrada y un nombre para la misma (nombrado como *name*).

De forma optativa podremos añadir, tantas propiedades de Wikidata como deseemos. xxx es el formato de cualquier propiedad Wikidata, donde xxxx es un número entero- Las propiedades deberían de ponerse en la primera fila de título, con su formato Pxxx, salvo las propiedades expuestas a continuación que deben de escribirse con el nombre que se especifica a continuación:

- descripción: (nombrado como *desc*) para mejor identificación
- url: (nombrado como *url*)
- Tipo: (nombrado como *name*) evitando así que se le asigne uno automático que puede que no sea el adecuado
- Fechas (nacimiento, muerte): (nombradas como *born* y *died*) con un formato determinado que puede ser tanto "AAAA", "AAAA-MM" o "AAAA-MM-DD"
- Autoq: (nombrado como *Qxxx*) se establecerá como un match automático
- Qnumber: (nombrado como *Qxxx*) se establecerá como un match manual
- Propiedad: (nombrado como *Pxxx*) se pueden nombrar por cadenas de texto, pero resulta más óptimo hacerlo directamente por su identificador. Por ejemplo, P1607 representa al DIALNET de autor y podríamos ponerlo como nombre de columna de las dos maneras.

Si queremos más información sobre el tema debemos ir a (5)

Los atributos que se han presentado en los puntos anteriores no representan los datos que hay que añadir, sino la variedad de datos que se pueden introducir en la herramienta de una forma genérica. Más adelante en la aplicación práctica, se mostrará los datos que se deben añadir para las autoridades en esta herramienta. Para los artículos científicos no se utilizará esta herramienta por las razones que se explicaran en sucesivos apartados

Antes de que se lleve a cabo la importación se mostrará las columnas que el sistema detecta antes de confirmar la importación (justo cuando pulsamos el anteriormente

explicado botón *looks good*). Una vez vistas se podrá retroceder y hacer las modificaciones necesarias en caso de que algo no vaya bien.

Un ejemplo de ello lo podemos ver en la imagen Ilustración 2-11 Ejemplo de las columnas asimiladas por mix'n'match en la importación donde podemos ver una serie de columnas en las que aparece la columna introducida y el modo en que lo ha entendido la herramienta.

Columnas de datos
Reload
Test

#	Column label	Use as
0	id	id
1	name	name
2	born	born
3	died	died
4	P950	P950
5	P1607	P1607
6	P213	P213
7	P214	P214
8	P496	P496
9	P1960	P1960
10	desc	desc

Ilustración 2-11 Ejemplo de las columnas asimiladas por mix'n'match en la importación

También podremos ver una lista de los elementos que se han capturado correctamente.

Columnas de datos
Reload
Test

Test results (≤1000)

Type	Rows
all	421
added	420
updated	0

Looks good, queue import/update

Finalmente, una vez creado el catálogo se facilitará el Identificador de este dónde podremos consultar sus entradas. Para acceder al catálogo desde el navegador se podrá hacer añadiendo a la siguiente extensión el código que identifica al catálogo de la siguiente forma <https://mix-n-match.toolforge.org/#/catalog/4580> (catálogo final de las autoridades de la Universidad Pontificia de Salamanca)

Después de un breve intervalo de tiempo podremos actualizar el catálogo (No actualizar el navegador, sino pulsar en la propia herramienta el botón que pone actualizar). Después de todo esto la herramienta ya habrá trabajado sobre las entradas que le hemos dado y las añadirá a las listas de datos que hemos explicado en el apartado 2.3.4 Funcionamiento. Más adelante se explicará más concretamente en el proceso que le sigue, hacer matches.

El siguiente es un ejemplo Ilustración 2-12 Ejemplo datos para ser importados a los catálogos se podría decir que es el archivo más simple admitido por la herramienta, en él se incluyen solo dos columnas. Son las dos columnas obligatorias que citamos en apartados anteriores (el identificador local y el nombre de ciertas autoridades de la UPSA). En este caso la herramienta sería capaz de asociarles elementos a las autoridades, pero no en muchas ocasiones esos elementos no se corresponderían con los que nosotros queremos, ya que con el nombre Fernando Galindo Rubio, por ejemplo, podrían existir distintas personas y el programa no tendría forma de distinguirlas.

Cuando se crea un catálogo de prueba es necesario indicarlo en la descripción para que los identificadores de Wikidata puedan identificarlo como tal y no sea eliminado.

```
0,Konstan", David
1,Tirado San Juan", Víctor Manuel
2,Galindo Rubio", Fernando
3,Rensoli Laliga", Lourdes
4,Fernández Vallina", Emiliano
5,Palacios Vicario", Beatriz
6,Berjón Gallinas", Roberto
7,Fernández del Campo Carranza", Lucas
8,Santamaría Domínguez", Mónica
9,Pérez Herrero", Pilar
10,Marín Serrano", Benito
11,Rivera Picón", Cristina
12,Álvarez García", Francisco José
13,Villa García", Julia
14,Tamarit Rodríguez", Ana
15,Fernández Blanco", Elena
16,Bécares Botas", Vicente
17,Merayo Pérez", Arturo
18,Suárez Carballo", Fernando
19,Beato Gutiérrez", María Encarnación
20,González Yagüe", Alfonso
21,Castro Cardoso", Dionisio de
22,Rodríguez Molinero", José Luis
23,Flores García", María Ángeles
24,Sánchez Vázquez Juan Francisco
```

Ilustración 2-12 Ejemplo datos para ser importados a los catálogos

La siguiente imagen Ilustración 2-13 muestra una lista de los últimos catálogos creados en la herramienta, presente en su pantalla principal.

Latest catalogs

[TEST] languages data	Learning test
[TEST] Prueba Autoridades universidad	Este catalogo es una prueba (missing human type)
[TEST] Actores de prueba	Este es un catalogo de prueba (incorrect type)
Women in Gaming: 100 Professionals of Play	people profiled in the "Women in Gaming: 100 Professionals of Play" book
Rekhta author	identifier for a author in the Rekhta database of literature of the Urdu language of India and Pakistan
Russians of Latvia	identifier for a person in the website Russians of Latvia
Biographia Benedictina	identifier for a Benedictine monk in the Biographia Benedictina
Brapci author	identifier for an author in the Brazilian database Brapci

Ilustración 2-13 Unos catálogos de prueba creados en la herramienta (Aparecen marcados como [TEST])

A continuación, se ve la imagen correspondiente al catálogo de prueba creado a partir de cinco de los datos que se muestran en Ilustración 2-12 Ejemplo datos para ser importados a los catálogos. Como podemos ver es capaz de ligar automáticamente algunas autoridades, pero lo más probable es que no sean las que andábamos buscando, esto se debe a que hemos dado como dato únicamente el nombre del autor, para el cual se pueden encontrar múltiples entradas.

[TEST] Prueba Autoridades universidad

Action ▾

Este catalogo es una prueba (missing human type)

Importado por Braisrodriguezcalvo | Actualizar

Entradas

Ligado manualmente	0	0%
Ligado automáticamente	4	80%
No Wikidata	0	0%
Not applicable to Wikidata	0	0%
Unmatched	1	20%
Total	5	

Tipos de entrada

Ilustración 2-14 Catálogo de Prueba autoridades UPSA

2.3.5 VINCULACIÓN DE LAS ENTRADAS

Una vez se ha creado el catálogo y tenemos asignadas de forma automática o sin asignar cada entrada deberemos realizar una serie de operaciones aún con ellas para finalizar el proceso. El esquema de trabajo es el siguiente.

Autoridades UPSA

Action ▾

es una lista combinada de Autoridades de la Universidad Pontificia de Salamanca y autores de la revista Cuadernos Salmantinos de Filosofía

Importado por Braisrodriguezcalvo | [Actualizar](#)

Entradas		
Ligado manualmente	0	0%
Ligado automáticamente	376	62%
No Wikidata	0	0%
Not applicable to Wikidata	0	0%
Unmatched	224	37.3%
Total	600	

Ilustración 2-15 Catálogo recién creado

Después de crear el catálogo comenzamos con matches automáticos y no matches como podemos ver en la Ilustración 2-15 Catálogo recién creado. Para los “matches automáticos” deberemos confirmar o no que el match es correcto- En caso de que escojamos que es correcto quedará completamente fijado (Ligado manualmente). En caso contrario, eliminaremos el match y pasará a estado de “Unmatched”. Un ejemplo de la forma de presentar la información de las entradas de Wikidata se muestra en la Ilustración 2-16 Ejemplo de elementos de la lista de matches automáticos.

Autoridades de la Universidad Pontificia de Salamanca

Action ▾

Registro de autoridades de la Universidad Pontificia de Salamanca

1 2

José Manuel de Mena Ramos	José Manuel de Mena Ramos es una Autoridad de la Universidad Pontificia de Salamanca José Manuel *de *Mena *Ramos *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
José Manuel de Mena Ramos [Q99838791]	España profesor universitario, entrenador deportivo, profesor ♂	Confirmar Borrar [all]
# Marcos Cabezas González	Marcos Cabezas González es una Autoridad de la Universidad Pontificia de Salamanca Marcos *Cabezas González *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
Marcos Cabezas-González [Q58681061]	Investigador ♂	Confirmar Borrar [all]
M. Luísa Díez Platas	M. Luísa Díez Platas es una Autoridad de la Universidad Pontificia de Salamanca M. *Luísa Díez *Platas *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
M. Luisa Díez Platas [Q99197449]	España profesor universitario ♀	Confirmar Borrar [all]
María Paz de Blas Fernández	María Paz de Blas Fernández es una Autoridad de la Universidad Pontificia de Salamanca María Paz *de Blas Fernández *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
María Paz de Blas Fernández [Q100715817]	España profesor universitario ♀	Confirmar Borrar [all]
Rubén González Crespo	Rubén González Crespo es una Autoridad de la Universidad Pontificia de Salamanca Rubén González *Crespo *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
Rubén González Crespo [Q60976901]	Investigador ♂	Confirmar Borrar [all]

Ilustración 2-16 Ejemplo de elementos de la lista de matches automáticos

El elemento puede ser “No match” desde un principio o bien porque hemos eliminado el “matches automático” debido a que después de investigar su descripción y buscarlo en Wikidata o Wikipedia concluimos que no es lo que andamos buscando. La forma de hacer estas comprobaciones es acceder a la página principal de Wikidata y escribir en la barra de búsqueda el nombre del elemento que queremos comprobar. Seguidamente nos saldrán una serie de resultados y podremos entonces consultar si realmente alguno de ellos es el que buscamos echando un vistazo en los a las propiedades de cada elemento (en base a los datos que tenemos y a los que aparecen en Wikidata podemos sacar conclusiones).

En el caso de que sea el que andamos buscando deberemos obtener el “Qxxx” de dicho elemento y asignárselo, en caso contrario debemos plantearnos si es un ítem apropiado para Wikidata y debemos crearlo como un elemento nuevo (Crear nuevo elemento) o simplemente desecharlo (No aplicable)

Antes de crear un elemento se deben de sopesar los requisitos de importancia propuestos por Wikidata, que son:

- Debe contener al menos un sitio de referencia válido a una página de Wikipedia, Wikivoyage, Wikisource ..., cualquiera de las wikis de Wikimedia.
- Debe de hacer referencia a una entidad material o a un concepto claramente identificable.
- Debe de ser de relevancia, en el sentido de que puede ser descrito mediante referencias serias y de disponibilidad pública.

- Satisface una necesidad estructural, como por ejemplo hacer más útiles y complementar a otros ítems existentes.

Si cumple dichos requisitos es seguro que se debe incluir en su base de datos.

Para una información más detallada se puede consultar en (6)

Por otro lado, Wikidata nos insta a que añadamos la mayor información posible para enriquecer sus fuentes.

Si no se está del todo seguro de que hacer con uno de los elementos lo que se recomienda es dejarlo para que otro usuario que pueda tener un mayor conocimiento, lo trámite.

2.3.6 MÉTODOS DE TRABAJO PARA VINCULAR ENTRADAS A WIKIDATA

Encontramos dos métodos de trabajo de para trabajar con la herramienta, de todas formas, para una mayor comprensión se recomienda (7), pero la filosofía que se sigue en todo momento es la de la imagen siguiente Ilustración 2-17 Filosofía de trabajo Mix'n'match

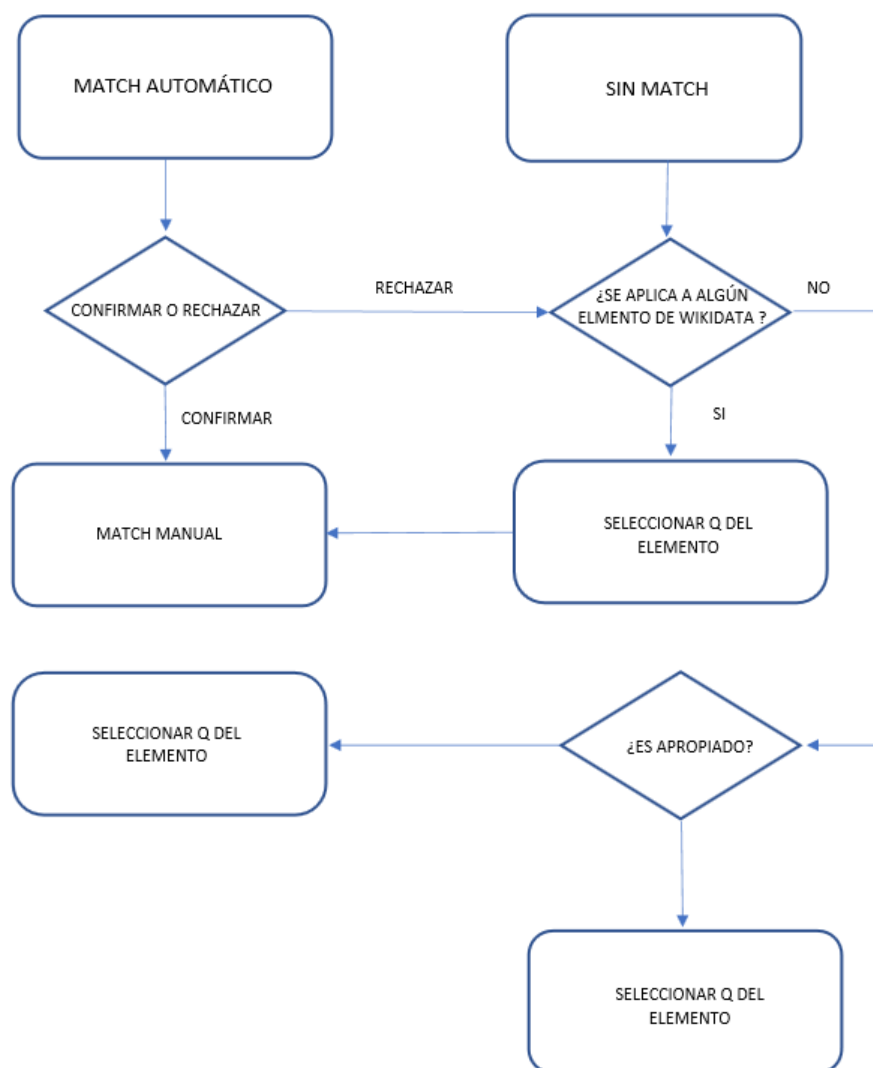


Ilustración 2-17 Filosofía de trabajo Mix'n'match

- **Modo de juego:** en él se presentan sólo elementos que no tienen match encontrado todavía y en un orden aleatorio, para que poco a poco vayamos vinculando las entradas del catálogo creado previamente. Para ello nos aparece la descripción completa del mismo y únicamente vemos uno a la vez, como podemos apreciar en la siguiente imagen Ilustración 2-18 Pantalla modo juego parte 1 y en Ilustración 2-19 Pantalla modo juego parte 2

Netflix FR

Action ▾

Netflix movies and TV shows available in France in January 2019

Derren Brown: The Push 🔍

Nueva entrada

Entrada	67228112
Identificador del catálogo	80220000
Descripción del catálogo	reality, variety and talk show with Derren Brown Realidad, variedad y espectáculo de charla con "Derren Brown"
Type	film [Q11424]

Introduce el número Q de un elemento coincidente

Fijar Q

Nuevo elemento

N/D

Buscar

[Search Wikidata](#) | [Search en.wikipedia](#) | [Google-search Wikipedias](#) | [Google-search Wikisource](#) | [Google-search Wikidata](#)

Wikidata search results

No matches found

Ilustración 2-18 Pantalla modo juego parte 1

en.wikipedia search results

programming		
Deb Webber	show uses cold reading and other techniques used by entertainers like Derren Brown. On 13 August 2004 Australian Channel 7 broadcast a show called "Caught"	Q5247796 [↑]
Eugene Washington	Eugene the first black actor to star in a professional Agatha Christie production in the UK). He also played Dog Fox in The Lion, the Witch and the Wardrobe	Q5407840 [↑]
Netflix	CNET. Archived from the original on March 14, 2017. Retrieved March 13, 2017. "Netflix Switch Lets You Netflix And Chill with the Push of a Button". Tech	Q907311 [↑]

Results from other catalogs [\[Candidatos para creación\]](#)

No matches found

Ilustración 2-19 Pantalla modo juego parte 2

Para entrar en este modo de juego debemos entrar en la etiqueta “Action” y escoger modo de juego o “Match Mode”.

Para cada elemento podremos ver sus características, un menú de opciones en el que vemos que es lo que podemos hacer con él, y listas de resultados de Wikidata, Wikipedia y otras categorías para que nos ayude a elegir qué hacer con él.

- **Modo lista:** nos separa las entradas por categoría, es decir podemos encontrar todos los tipos anteriores y aparecen ordenadas en una lista en la que podremos hacer Scroll y movernos por ella sin problemas

En este modo no encontraremos sugerencias como en el anterior, seremos nosotros los que tenemos que investigar. Para entrar en este modo lo único que deberemos hacer es acceder a cualquiera de las listas de los tipos de datos que tenemos en el catálogo, un ejemplo de la visión de las entradas es el siguiente Ilustración 2-20 Pantalla modo lista.

Netflix FR

Action ▾

Netflix movies and TV shows available in France in January 2019

Basilisk	anime with Kohsuke Toriumi, Kiyoshi Kobayashi, Ryusaku Chijiwa, and Risa Hayamizu <i>*anime Con *Kohsuke *Toriumi, *Kiyoshi *Kobayashi, *Ryusaku *Chijiwa, y *Risa *Hayamizu</i>	Ligado automáticamente 3▼
Basilisk [Q282464]	Serie de televisión de anime, serie de manga de Japón	Confirmar Borrar [all]
Crusoe	period piece created by Justin Bodle and Stephen Gallagher <i>Pieza de periodo creada por Justin *Bodle y Stephen Gallagher</i>	Ligado automáticamente
Crusoe [Q1388988]	Serie de televisión por James Seymour Brett, Stephen Gallagher de Estados Unidos	Confirmar Borrar [all]
Deadman Wonderland	anime with Romi Park, Kana Hanazawa, Masayuki Kato, and Yuki Kaji <i>*anime Con *Romi Parque, *Kana *Hanazawa, *Masayuki *Kato, y *Yuki *Kaji</i>	Ligado automáticamente
Deadman Wonderland [Q482790]	2007 serie de manga por Jinsei Kataoka publicado en Gekkan Shōnen Ace de Japón	Confirmar Borrar [all]
Hellsing	anime created by Kouta Hirano <i>*anime Creado por *Kouta *Hirano</i>	Ligado automáticamente 2▼
Hellsing [Q951078]	Serie de televisión, serie de televisión de anime por Umanosuke Iida, parte de Hellsing de Japón	Confirmar Borrar [all]
Big	asian tv show with Gong Yoo, Rhee Min-jung, Bae Suzy, and Jang Hee-jin <i>*asian *tv Espectáculo con Gong *Yoo, *Rhee *Min-*jung, *Bae *Suzy, y *Jang *Hee-*jin</i>	Ligado automáticamente 3▼
Big [Q488633]	Serie de televisión de Corea del Sur	Confirmar Borrar [all]
Iris	asian tv show with Lee Byung-hun, Kim Tae-hee, Jeong Jun-ho, and Kim Young-chul <i>*asian *tv Espectáculo con Lee *Byung-*hun, Kim *Tae-*hee, *Jeong *Jun-*ho, y Kim Young-*chul</i>	Ligado automáticamente
Iris [Q496296]	Serie de televisión de Corea del Sur	Confirmar Borrar [all]

Ilustración 2-20 Pantalla modo lista

2.3.7 TRUCOS PARA HACER MÁS RÁPIDA LA VINCULACIÓN

- Si son sujetos humanos, es probable que tengan como datos propios nacimiento y defunción, esto es muy útil para verificar que es quien creemos que es, o desambiguar la entrada.
- Comprobar si su descripción se relaciona con la de las sugerencias
- Cuando se busca en Wikidata y no encontramos ningún resultado porque la búsqueda es demasiado específica (por ejemplo, incluimos su fecha de nacimiento), es recomendable borrar datos como los años en el buscador, para obtener resultados más generales (puede que la información de la fecha no exista todavía y si otros tipos de informaciones)

2.3.8 RETOMAR EL TRABAJO

Si por alguna razón tenemos que volver a retomar el trabajo en el catálogo para añadirle nuevos datos o para buscar matches podremos hacerlos buscando el catálogo por el nombre de la categoría (su nombre) en el buscador y una vez dentro de él pulsar en el botón de acción importación o actualización para actualizar o en el editor de catálogo para cambiar sus características o de forma más directa como se comentó en apartados anteriores en lo que se especifica como debe de construirse la URL de acceso al catálogo

2.3.9 EXTRACCIÓN DE LOS DATOS NO EXISTENTES DE WIKIDATA

Si se necesita extraer los datos de las listas del catálogo, tenemos la posibilidad de hacerlo. Para ello debemos acceder a la pestaña de “Acción” y a continuación escoger la opción “Descargar”, pasaremos a otra ventana en la que deberemos especificar las columnas de datos que queremos obtener para cada elemento, o lo que es lo mismo y que es lo que se nos obliga a hacer en la herramienta, las cualidades de los elementos que queremos que no estén en el archivo que descargamos. También deberemos de establecer el formato de salida, que puede ser Json o .tsv (se recomienda usar el formato json dado que es más fácil de trabajar con él cuando implementamos código).

ACLARACIÓN: aquí tenemos un apartado dedicado a los datos auxiliares, estos datos auxiliares son datos que le aportamos en nuestro archivo de importación y que no muestra en las listas de esta aplicación pero que sí que tiene en cuenta a la hora de hacer el match como por ejemplo el VIAF.

En nuestro proyecto deberemos crear finalmente las entradas para las que no se ha encontrado un elemento equivalente en Wikidata, y como veremos más adelante vamos a necesitar separar los datos que la herramienta ha podido encontrar de los que no, o simplemente descargar dos listas distintas. Otra opción es realizar las operaciones con la herramienta, dejando completamente aislados los elementos que tengan match da igual automático que manual y por otro lado los elementos que no tengan match, una vez se ha conseguido esto descargar directamente todas las entradas con el mínimo número de atributos posibles. Realmente solo nos hace falta el atributo que nos dice a que Qxxx se le ha atribuido a cada entrada, si no se ha encontrado aparecerá como null y será fácilmente detectable esta diferenciación mediante código, creando así dos listas. Una contendrá los elementos que ya existen y para los cuales querremos actualizar o añadir alguna propiedad nueva y por otro lado los que vamos a crear desde cero, ambas operaciones realizadas por la siguiente herramienta que vamos a mencionar a continuación, Openrefine.

A continuación, podemos ver unas imágenes Ilustración 2-21 Menú de descarga de las entradas de Mix’n’match y Ilustración 2-22 Muestras del archivo json descargado del menú de descarga y la forma en la que se recomienda hacerlo, ya que el hecho de esconder lo que no quieres resulta muy confuso.

Descargar entradas

Catálogos	4580
Columnas de datos	<input checked="" type="checkbox"/> URL externa <input checked="" type="checkbox"/> Ligado por nombre de usuario <input type="checkbox"/> Datos auxiliares <input type="checkbox"/> Fecha de nacimiento/fallecimiento <input type="checkbox"/> Latitud/Longitud <input checked="" type="checkbox"/> Multiple matches
Hide matches	<input type="checkbox"/> Hide any matched entries, even automatic ones <input type="checkbox"/> Hide entries matched by users, via auxiliary data, or name&dates <input type="checkbox"/> Hide entries matched by users <input type="checkbox"/> Hide unmatched entries <input type="checkbox"/> Hide automatically (transiently, "automatched") matched entries <input type="checkbox"/> Hide entries automatically matched via name&dates <input type="checkbox"/> Hide entries matched via auxiliary data (e.g. via shared VIAF ID) <input type="checkbox"/> Hide entries without multi-matches
Formato	<input type="radio"/> Texto tabulado <input checked="" type="radio"/> JSON <input type="checkbox"/> Descargar como archivo (del otro modo, se abre en una nueva pestaña)
<div>Descargar</div>	

Ilustración 2-21 Menú de descarga de las entradas de Mix'n'match

```
{
  "entry_id": "114985244",
  "catalog": "4580",
  "external_id": "46936",
  "external_url": "https://koha.upsa.es/cgi-bin/koha/opac-authoritiesdetail.pl?marc=1&authid=46936",
  "name": "Antonio Pintor Ramos",
  "description": "Antonio Pintor Ramos es una Autoridad de la Universidad Pontificia de Salamanca",
  "entry_type": "",
  "mnm_user_id": "4",
  "q": "Q988645962",
  "matched_on": "20210625200115",
  "matched_by_username": "Auxiliary data matcher",
  "multi_match_candidates": null
},
{
  "entry_id": "114985245",
  "catalog": "4580",
  "external_id": "1 Follow link (cmd + click)",
  "external_url": "https://koha.upsa.es/cgi-bin/koha/opac-authoritiesdetail.pl?marc=1&authid=168669",
  "name": "José Manuel Alfonso Sánchez",
  "description": "José Manuel Alfonso Sánchez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964",
  "entry_type": "",
  "mnm_user_id": "4",
  "q": "Q98995910",
  "matched_on": "20210625200115",
  "matched_by_username": "Auxiliary data matcher",
  "multi_match_candidates": null
},
}
```

Ilustración 2-22 Muestras del archivo json descargado

2.3.10 VALORACIÓN DE LA HERRAMIENTA

La herramienta Mix'n'match es útil para saber qué elementos de entre una lista dada existen o no en la BD de Wikidata. En casos como el de las Autoridades de la Universidad Pontificia de Salamanca en donde muchas entidades ya existen, aunque sin mucha información y el resto aún no existe, pero no es creador de información propiamente dicho, es verdad que entre sus funciones está la de crear elementos Wikidata, pero el proceso tiene que ser manual lo cual resulta muy tedioso para grandes volúmenes de datos. Si lo que queremos es añadir muchos nuevos elementos deberemos usarlo otra herramienta como por ejemplo OpenRefine. A pesar de esto nos es muy útil la herramienta que acabamos de describir, para obtener listas en formato Json o tsv de los elementos que no existen y a los que queremos añadir más datos como se ha explicado en el apartado anterior

2.4 HERRAMIENTA OPENREFINE

Es otra de las herramientas que usaremos para desarrollar nuestro proyecto. Es una herramienta java Open Source con un interfaz web similar a una hoja de cálculo con muchas aplicaciones, principalmente para limpieza de tablas de datos y vinculación con bases de conocimiento como puede ser Wikidata, o también enriquecimiento de fuentes y la creación de elementos Wikidata en masa usando Quickstatements.

A diferencia de otros softwares del estilo todo el trabajo se lleva a cabo desde el navegador, lo que para gustos puede ser más o menos cómodo.

El nombre original de este sistema fue Google Refine y fue creada por el grupo Google. Actualmente es sostenido por la comunidad y le ha llevado unos 10 años prepararse y convertirse en un sistema de confianza renombrado como OpenRefine.

En nuestro proyecto lo usaremos tanto para enriquecer como para crear nuevos elementos wikidata

2.4.1 PRINCIPALES FUNCIONALIDADES

Entre sus funciones principales encontramos:

- **COTEJAMIENTO DE DATOS:** se entiende por ello la vinculación de datos externos en formato bruto con elementos propios de bases de conocimiento. En la herramienta esto se conoce como reconciliación de los datos. Estos datos pueden proceder de museos, universidades bibliotecas, de usuarios o de nuestro repertorio local de archivos. Estos datos simplemente se comparan con los que ya existen para obtener una coincidencia lo más exacta posible, en eso consiste la reconciliación. La reconciliación funciona de una forma similar a la de Mix'n'match ya que también está automatizado y también requiere de la validación manual. También se recomienda que los datos estén lo más limpios posibles sin caracteres extraños para facilitar su identificación.
- **ENRIQUECIMIENTO DE DATOS:** está disponible en la versión 2.8 y superiores. Permite crear columnas adicionales a partir de los datos de Wikidata (debe de haber una columna cotejada ya que en base a los elementos que se sabe completamente que existen, se obtienen las propiedades que puedan tener estos datos). De esta forma podremos enriquecer nuestros propios datos y por ejemplo descargarlos para otros fines.
- **TRANSFORMACIÓN AL FORMATO WIKIDATA:** se rige por un "Schema", este *Schema* es el método que se usa para especificar la estructura que queremos darles a los objetos que vamos a crear o a las propiedades que queremos enriquecer, un ejemplo sería decirle que propiedades y con qué valores van a aparecer, así como una referencia que sirva de justificante para ese dato.
- **LIMPIEZA DE DATOS:** no es necesario que los datos se usen para incrementar fuentes, pueden simplemente limpiar el formato y hacer cambios sobre lo que ya tenemos, y devolverlos en un formato determinado.

2.4.2 CONCILIACIÓN DE LOS DATOS

En cuanto a la conciliación de los datos la herramienta cuenta con una lista de fuentes de reconciliación bastante amplia, llamando la atención en especial la que es contra el VIAF (el VIAF es un identificador que sirve para unir ficheros de autoridad de distintas bibliotecas). Una opción de hecho para el enriquecimiento de nuestro proyecto gracias a esta funcionalidad de OpenRefine, sería que en los casos de las líneas, es decir, de autores de nuestro fichero de autoridades en que no exista dicho atributo, reconciliar contra esta base de conocimiento para intentar rellenar los huecos vacíos.

Wikidata está incluida en el paquete de instalación, al que se pueden añadir extensiones que pueden mejorar su funcionalidad como es el caso del VIAF. Conciliar puede usarse también para:

- Corrección ortográfica y cambios en los nombres.
- Limpieza de encabezados con respecto a los proporcionados por autoridades de confianza
- Agregar datos a servicios como Wikidata
- Concretar en qué otras listas aparecen los elementos conciliados

Para obtener más información sobre este tema se recomienda consultar el manual de OpenRefine en (8)

La ventana principal de OpenRefine nos proporciona una serie de opciones sobre como iniciar nuestro trabajo, ofreciéndonos la oportunidad de crear un nuevo proyecto abrir uno existente o importarlo, además de una serie de opciones sobre el lenguaje de la aplicación. Por lo tanto, podremos cambiar el idioma de la aplicación fácilmente desde el principio (mi recomendación es usarlo siempre en inglés). Desde la Ilustración 2-23 Pantalla principal de la aplicación de navegador de OpenRefine se puede ver lo que acabamos de mencionar

En esta demostración vamos a abrir un nuevo proyecto desde un archivo .csv separado por comas que se ha creado anteriormente con los datos de alguna de las autoridades de la UPSA (en esta ocasión se ha usado el csv con comas, pero se recomienda siempre hacerlo con un formato tsv).

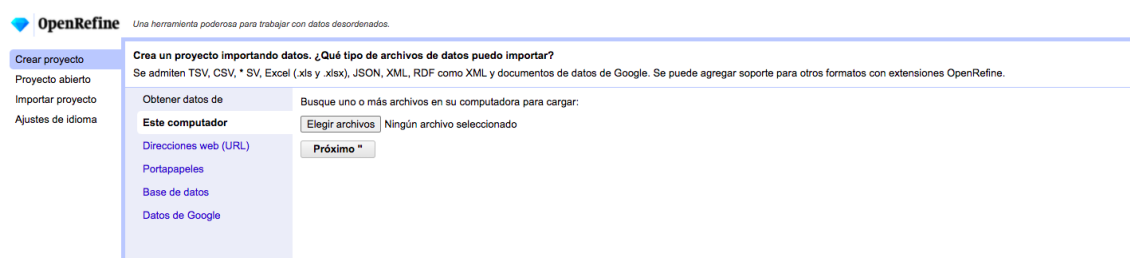


Ilustración 2-23 Pantalla principal de la aplicación de navegador de OpenRefine

Tras añadir el archivo nos da una vista preliminar de él antes de crear el proyecto, donde podremos asignarle un nombre.

En la parte inferior de la pantalla se puede ver un menú en el que podemos especificar distintas configuraciones a la hora de abrir el archivo. Todo lo mencionado se puede apreciar en Ilustración 2-24 Ventana de inserción de datos de OpenRefine

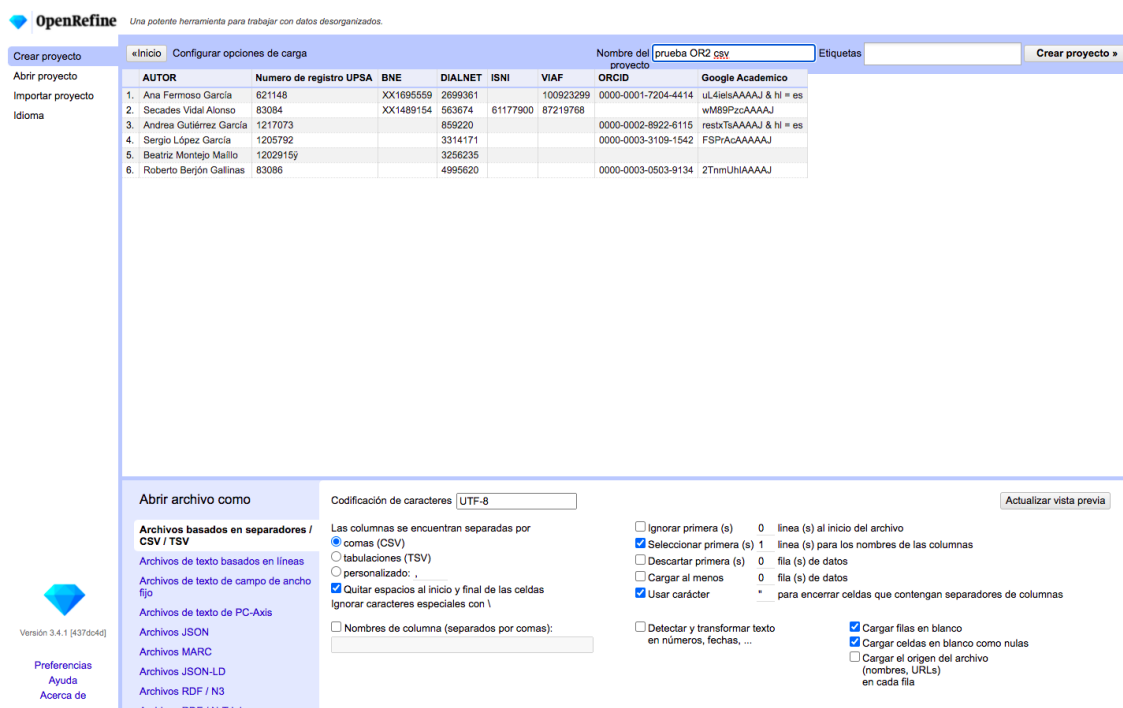


Ilustración 2-24 Ventana de inserción de datos de OpenRefine

Lo que vamos a hacer ahora es reconciliar con respecto al dato o columna que creamos que es más identificativa posible, que en este caso será la columna "Autor". Para ello se debe clicar en el menú desplegable de la columna en cuestión y entrar en la pestaña de "Cotejar" → "Iniciar" como se puede ver en la Ilustración 2-25 Opciones de cotejamiento de datos OpenRefine

6 filas Extensiones: Wikidata ▼

Mostrar como: **filas** registros Mostrar: 5 10 25 50 filas «Primera < anterior 1 - 6 siguiente > última»

Que hace	AUTOR	Numero de regis	BNE	DIALNET	ISNI	VIAF	ORCID	Google Academi
1. Facetas	621148	XX1695559	2699361			100923299	0000-0001-7204-4414	uL4ielsAAAAJ & hl = es
2. Filtro de texto	83084	XX1489154	563674		61177900	87219768		wM89PzcAAAAJ
3. Editar celdas	1217073		859220				0000-0002-8922-6115	restxTsAAAAJ & hl = es
4. Editar columnas	1205792		3314171				0000-0003-3109-1542	FSPrAcAAAAAJ
5. Transponer	1202915y		3256235					
6. Ordenar ...	83086		4995620				0000-0003-0503-9134	2TnmUhlAAAAJ

Ver
 Cotejar ▶ Iniciar ...
 Facetas ▶
 Acciones ▶
 Copiar información de cotejo ...
 Usa valores como identificadores
 Agregar columna de identificadores de entidad

Ilustración 2-25 Opciones de cotejamiento de datos OpenRefine

A continuación, escogeremos el servicio sobre el que queremos cotejar los datos, por defecto viene Wikidata. Aunque tenemos opción a agregar más, para nuestro caso no será necesario, salvo que intentemos recopilar más información como podría ser el ya mencionado caso del VIAF. La Ilustración 2-26 Elección de la base para reconciliar los datos es un ejemplo de esto.

Cotejar columna "AUTOR"

Servicios ✕

Wikidata (en) ✕

Agregue un servicio o extensión a la izquierda

Agregar servicio estándar ...
 Cotejar Cancelar

Ilustración 2-26 Elección de la base para reconciliar los datos

A continuación, el sistema intentará reconocer el tipo de dato de la columna que hemos usado y nos da unas cuantas sugerencias en Ilustración 2-27 Opciones para cotejar una columna OpenRefine, también podremos escoger nosotros uno que tal vez se amolde mejor para nuestro caso. Para este ejemplo nos vale con el sugerido.

Algo muy recomendable para aumentar la probabilidad de reconciliar el mayor número de entradas posible, es el tomar como apoyo alguna de las otras columnas de nuestra tabla, por ejemplo, el identificador DIALNET. Como este valor es único, el resultado de la reconciliación es Óptimo.

Lo que estamos haciendo en la Ilustración 2-27 Opciones para cotejar una columna OpenRefine, es decirle que nos busque elementos que tengan ese nombre, que sean del tipo humano (Q5) y que además tienen la propiedad DIALNET.

Cotejar columna "AUTOR"

» Ir a API del servicio

Cotejar cada celda con los valores de una de estas clases:

☒ humano
Q5

☐ artículo académico
Q13442814

☐ artículo de revisión
Q7318358

Usar también detalles relevantes de otras columnas:

columna	¿Incluir?	Como propiedad
Numero de registro	<input type="checkbox"/>	
UPSA	<input type="checkbox"/>	
BNE	<input type="checkbox"/>	
DIALNET	<input checked="" type="checkbox"/>	Dialnet author ID
ISNI	<input type="checkbox"/>	
VIAF	<input type="checkbox"/>	
ORCID	<input type="checkbox"/>	
Google Academico	<input type="checkbox"/>	

☐ Cotejar contra la clase:

☐ Reconciliar contra ninguna clase en particular

☒ Cotejar automáticamente candidatos con alta confianza

Máximo número de candidatos a devolver

Agregar servicio estándar ...

Cotejar

Cancelar

Ilustración 2-27 Opciones para cotejar una columna OpenRefine

El proceso lleva un tiempo razonable. Cuando termina nuestra columna estará reconciliada. Justo debajo de su nombre podremos apreciar una barra de progreso de color verde que nos indica el porcentaje de reconciliación que hemos conseguido. Podemos verlo en la Ilustración 2-28 Entradas después de ser cotejadas

OpenRefine prueba OR2 csv Enlace permanente

Facetas / Filtros Deshacer / Rehacer 1/1

Actualizar Restablecer todos Remover todos

Mostrar como: **filas** registros Mostrar: 5 10 25 50 filas

Extensiones: Wikidata

«Primera < anterior 1 - 6 siguiente > última»

Que hace	AUTOR	Numero de regis	BNE	DIALNET	ISNI	VIAF	ORCID	Google Academi
1.	Ana Fermoso García Escojer nueva coincidencia	621148	XX1695559	2699361		100923299	0000-0001-7204-4414	uL4ieIAAAAAJ & hI = es
2.	Secades Vidal Alonso Escojer nueva coincidencia	83084	XX1489154	563674	61177900	87219768		wM89PzcAAAAJ
3.	Andrea Gutiérrez García ✓ Crear nuevo elemento Buscar una coincidencia	1217073		859220			0000-0002-8922-6115	restxTsAAAAJ & hI = es
4.	Sergio López García ✓ Sergio López-García (100) ✓ Sergio López García (100) ✓ Sergio C López García (95) ✓ Crear nuevo elemento Buscar una coincidencia	1205792		3314171			0000-0003-3109-1542	FSPiAcAAAAJ
5.	Beatriz Montejó Mallo Escojer nueva coincidencia	1202915y		3256235				
6.	Roberto Berjón Gallinas Escojer nueva coincidencia	83086		4995620			0000-0003-0503-9134	2TrmUHIAAAAJ

Facetas / Filtros Deshacer / Rehacer 1/1

Actualizar Restablecer todos Remover todos

2 opciones Ordenar por: **AZ** conteo

emparejado 4

ninguno 2

Facetas por conteo de opciones

AUTOR: juicio

AUTOR: puntuación del mejor candidato

100,00 - 101,00

Número ☒ No número ☐ Blanco ☒

Error ☐

Ilustración 2-28 Entradas después de ser cotejadas

En la columna de la izquierda de Ilustración 2-28 Entradas después de ser cotejadas podemos ver una serie de cuadros, estos cuadros se llaman “FACETAS”, y son una forma de analizar los datos de la columna reconciliada o de cualquier otra. En la primera sobre el Autor podemos ver el número de elementos emparejados y los que no se han conseguido emparejar.

Esta parte, Ilustración 2-28 Entradas después de ser cotejadas, se asemeja mucho a lo que ya hemos visto de Mix’n’match porque en las entradas de la tabla podemos ver que se nos da la opción de reconciliar manualmente y que se nos ofrecen una serie de opciones de elementos existentes según su porcentaje de similitud. Para cerciorarnos de que es el mismo podemos buscarlo a mano en Wikidata para ello basta con ponernos encima del posible valor cotejado y nos saldrá su información y cómo podemos acceder a él en Wikidata como podemos ver en el cuadro de la Ilustración 2-29 Resultado tras la reconciliación. También se puede optar por la creación de un nuevo ítem directamente si lo consideramos apropiado para Wikidata.

OpenRefine prueba OR2 csv Enlace permanente

Facetas / Filtros Deshacer / Rehacer 1/1

Actualizar Restablecer todos Remover todos

Mostrar como: **filas** registros Mostrar: 5 10 25 50 filas

Extensiones: Wikidata

«Primera < anterior 1 - 6 siguiente > última»

Que hace	AUTOR	Numero de regis	BNE	DIALNET	ISNI	VIAF	ORCID	Google Academi
1.	Ana Fermoso García Escojer nueva coincidencia	621148	XX1695559	2699361		100923299	0000-0001-7204-4414	uL4ieIAAAAAJ & hI = es
2.	Secades Vidal Alonso Escojer nueva coincidencia	83084	XX1489154	563674	61177900	87219768		wM89PzcAAAAJ
3.	Andrea Gutiérrez García ✓ Crear nuevo elemento Buscar una coincidencia	1217073		859220			0000-0002-8922-6115	restxTsAAAAJ & hI = es
4.	Sergio López García ✓ Sergio López-García (100) ✓ Sergio López García (100) ✓ Sergio C López García (95) ✓ Crear nuevo elemento Buscar una coincidencia	1205792		3314171			0000-0003-3109-1542	FSPiAcAAAAJ
5.	Beatriz Montejó Mallo Escojer nueva coincidencia	1202915y		3256235				
6.	Roberto Berjón Gallinas Escojer nueva coincidencia	83086		4995620			0000-0003-0503-9134	2TrmUHIAAAAJ

Facetas / Filtros Deshacer / Rehacer 1/1

Actualizar Restablecer todos Remover todos

2 opciones Ordenar por: **AZ** conteo

emparejado 4

ninguno 2

Facetas por conteo de opciones

AUTOR: juicio

AUTOR: puntuación del mejor candidato

100,00 - 101,00

Número ☒ No número ☐ Blanco ☒

Error ☐

Wikidata

Sergio López-García (Q287109291)

researcher

WIKIDATA

Coincidir con esta celda

Coincidir con todas las celdas idénticas

Cancelar

Ilustración 2-29 Resultado tras la reconciliación

Como podemos ver en Ilustración 2-30 Tabla principal de datos sobre el objeto y Ilustración 2-31 Búsqueda en Wikidata del elemento a identificar, nos lleva a su objeto Wikidata.

Sergio López-García (Q87109291)

investigador

 [editar](#)

[En más idiomas](#)

[Configurar](#)

Idioma	Etiqueta	Descripción	También conocido como
inglés	Sergio López-García	investigador	
Español	Sin etiqueta definida	investigador	
catalán	Sin etiqueta definida	Sin descripción definida	
gallego	Sin etiqueta definida	Sin descripción definida	

[Todos los idiomas ingresados](#)

Ilustración 2-30 Tabla principal de datos sobre el objeto

empleador



Universidad pontificia de salamanca

hora de inicio

5 de octubre de 2011

[1 referencia](#)

[+ añadir valor](#)

[+ agregar declaración](#)

Identificadores

ORCID ID



0000-0003-3109-1542

[0 referencias](#)

[+ añadir referencia](#)

[+ añadir valor](#)

[+ agregar declaración](#)

Ilustración 2-31 Búsqueda en Wikidata del elemento a identificar

En las capturas anteriores podemos ver su ORCID y que pertenece al PDI de la UPSA por lo tanto coincide con la primera sugerencia.

En este caso que no tenemos muchos datos podemos hacer la comprobación de forma manual, pero si lo que queremos es hacerlo más rápidamente, podemos bajar el porcentaje de conciliación aceptado en la parte izquierda de la pantalla. Si hacemos esto lo que pasará es que se asimilaran como válidas coincidencias que tengan un menor porcentaje de similitud.

OpenRefine prueba OR2 csv [Enlace permanente](#) Abrir ... Exportar Ayuda

Facetas / Filtros Deshacer / Rehacer 3/3 **6 filas** Extensiones: Wikidata

Actualizar Restablecer todos Remover todos Mostrar como: filas registros Mostrar: 5 10 25 50 filas «Primera < anterior 1 - 6 siguiente > última»

AUTOR: juicio cambiar

2 opciones Ordenar por: **AZ** conteo

emparejado 4

ninguno 2

Facetas por conteo de opciones

AUTOR: puntuación del mejor candidato cambiar restaurar

100,00 - 101,00

☒ Numérico ☐ No numérico ☒ Blanco

☐ Error

Que hace	AUTOR	Numero de regis	BNE	DIALNET	ISNI	VIAF	ORCID	Google Academico
1.	Ana Fermoso García	621148	XX1695559	2699361		100923299	0000-0001-7204-4414	uL4ieIsAAAAJ & hI = es
2.	Secades Vidal Alonso	83084	XX1489154	563674	61177900	87219768		wM89PzAAAAAJ
3.	Andrea Gutiérrez García	1217073		859220			0000-0002-8922-6115	restxTsAAAAAJ & hI = es
4.	Sergio López-García	1205792		3314171			0000-0003-3109-1542	FSPfAcAAAAAJ
5.	Beatriz Montejó Maíllo	1202915y		3256235				
6.	Roberto Berjón Gallinas	83086		4995620			0000-0003-0503-9134	2TnmUHIAAAAAJ

Ilustración 2-32 Entradas después de reconciliar los datos

El resultado final después de reconciliar los datos sería Ilustración 2-32 Entradas después de reconciliar los datos

2.4.3 TRABAJAR CON FACETAS

Las facetas son unos mecanismos que nos permiten agrupar las entradas de datos en función de una serie de criterios los cuales podemos escoger en nuestros menús de facetas

Tenemos varios tipos de facetas que nos ayudan a trabajar sobre los datos, y estás se encuentran en dos apartados distintos que nombramos a continuación.

- **Facetas con respecto a la columna reconciliada:** son facetas que nos permiten actuar sobre los datos que ya han sido reconciliados, y agrupar estos por distintos criterios pero que tienen que ver directamente con la conciliación de elementos. Podemos ver todas las posibles funcionalidades en la Ilustración 2-33 Facetas sobre columnas reconciliadas

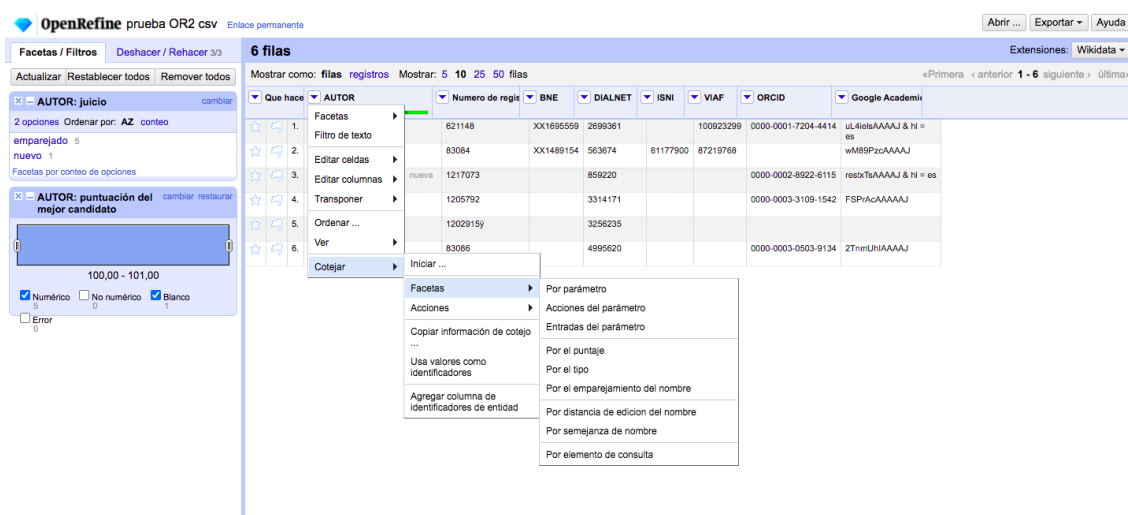


Ilustración 2-33 Facetas sobre columnas reconciliadas

El más usado y que aparece por defecto cuando se reconcilian datos, es la faceta “Por parámetro” que nos muestra agrupadas las entradas que han conseguido conciliarse o no Ilustración 2-34 Faceta por parámetro

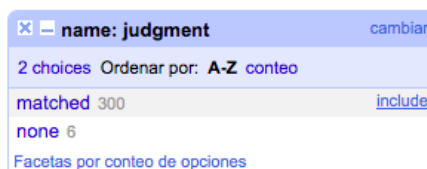


Ilustración 2-34 Faceta por parámetro

Y la otra faceta que aparece por defecto es la que mide la puntuación de la conciliación en la Ilustración 2-35 Faceta por puntuación

name: por el puntaje del mejor candidato

81.00 — 101.00

☒ Numérico ☐ Non-numeric ☒ Blank ☐ Error

0 0 6 0

Ilustración 2-35 Faceta por puntuación

Hay muchas más facetas que se pueden usar, pero las dos más útiles son las dos anteriores. Aunque hay más solo para los elementos cotejados, como se puede ver en Ilustración 2-36 Todas las facetas sobre datos conciliados y dado que el objetivo del proyecto no es la investigación en concreto de cada funcionalidad, no se explicará el funcionamiento del resto, aunque si se necesitara para otro proyecto se puede encontrar más información en (9)

AUTOR: ¿El tipo del mejor candidato encaja?

2 opciones Ordenar por: AZ conteo

(sin tipo) 1

verdadero 5

Facetas por conteo de opciones

AUTOR: ¿El emparejamiento del nombre del mejor candidato encaja?

3 opciones Ordenar por: AZ conteo

(no reconciliado) 1

falso 1

verdadero 4

Facetas por conteo de opciones

AUTOR: por distancia de edición del nombre del mejor candidato

0,00 - 1,01

☒ Numérico ☐ No numérico ☒ Blanco

5 0 1

☐ Error 0

AUTOR: Por semejanza del nombre del mejor candidato

0,33 - 1,00

☒ Numérico ☐ No numérico ☒ Blanco

5 0 1

☐ Error 0

AUTOR: juicio

2 opciones Ordenar por: AZ conteo

emparejado 5

nuevo 1

Facetas por conteo de opciones

AUTOR: puntuación del mejor candidato

100,00 - 101,00

☒ Numérico ☐ No numérico ☒ Blanco

5 0 1

☐ Error 0

AUTOR historial de entradas

3 opciones Ordenar por: AZ conteo

1619543375674 4

1619545678091 1

1619546148915 1

Facetas por conteo de opciones

AUTOR: por el puntaje del mejor candidato

100,00 - 101,00

☒ Numérico ☐ No numérico ☒ Blanco

5 0 1

☐ Error 0

AUTOR: tipos de mejores candidatos

2 opciones Ordenar por: AZ conteo

(no reconciliado) 1

Q5 5

Facetas por conteo de opciones

Ilustración 2-36 Todas las facetas sobre datos conciliados

- **Facetas Generales:** son operaciones que se pueden hacer sobre cualquier columna y que definen a sus elementos de una forma en concreto agrupándolos.

Son más numerosas aun que las facetas de conciliación, como podemos ver en el menú desplegable presente en la Ilustración 2-37 Opciones de facetas generales con respecto a una columna cualquiera

Nombraremos unas pocas que parecen más relevantes por su utilidad

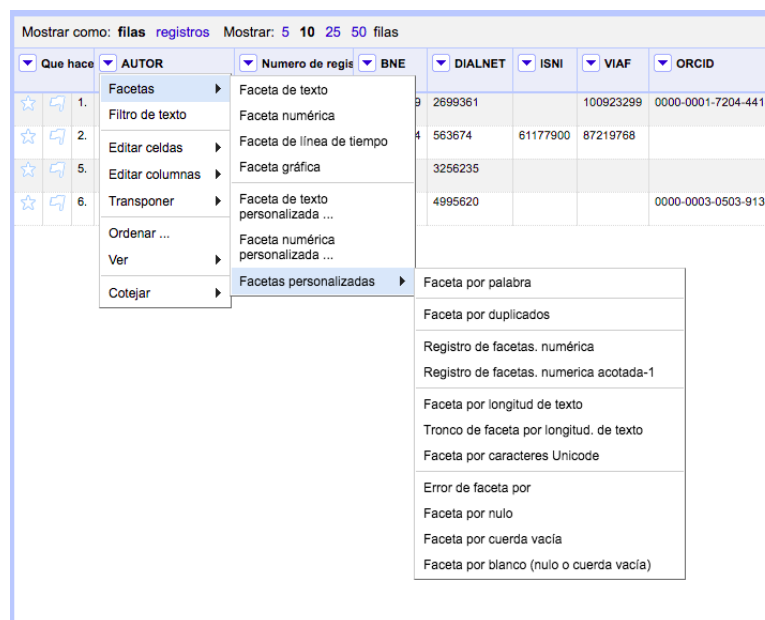


Ilustración 2-37 Opciones de facetas generales con respecto a una columna cualquiera

La faceta de texto: es una faceta que nos agrupa las entradas por su valor textual, como se puede ver en el ejemplo de la en la Ilustración 2-38 Faceta de texto estándar que agrupa las entradas por nombre de autor. Para este caso ocurre que cada autor es único, por esa razón solo aparecen agrupaciones de un elemento.



Ilustración 2-38 Faceta de texto estándar

La faceta por duplicados: es una faceta que indica los duplicados que puede haber por columna dentro de nuestras entradas. Se puede apreciar un ejemplo en Ilustración 2-39 Faceta por duplicados, donde nos da que todos los elementos son “false” dado que no encontró ningún duplicado.

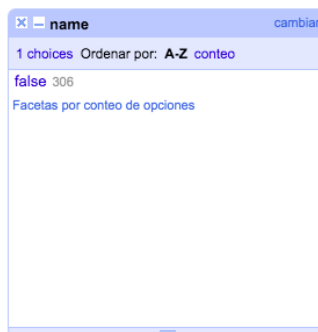


Ilustración 2-39 Faceta por duplicados

La faceta por longitud de texto: donde podemos ver una gráfica agrupando los tamaños de los datos de la columna en cuestión según el número de caracteres. Está presente en el ejemplo de la Ilustración 2-40 Faceta por longitud de texto

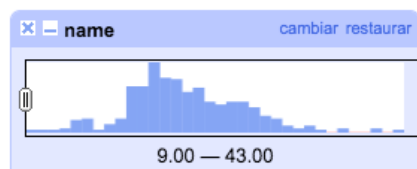
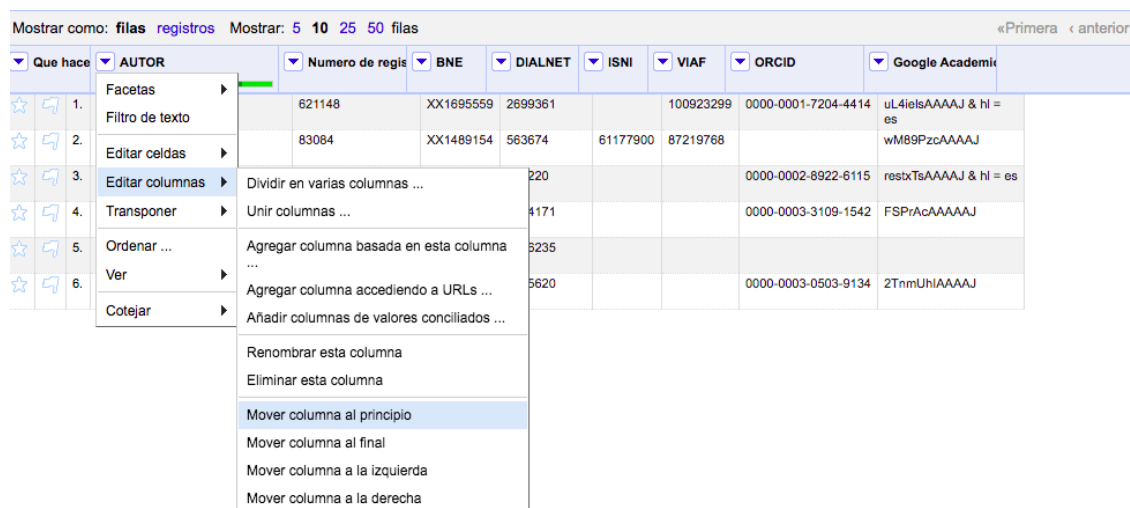


Ilustración 2-40 Faceta por longitud de texto

2.4.4 EDITAR COLUMNAS

La edición de columnas es tremendamente importante ya que podremos hacer todo tipo de operaciones sobre la columna objetivo para obtener el resultado esperado. Podremos hacer operaciones como crear nuevas columnas, moverlas, renombrarlas, dividir, unir... Entre ellas destaca la de enriquecimiento de datos que explicaremos más adelante. Podemos observar el menú de acciones en la Ilustración 2-41 Menú desplegable de edición de columnas



Su funcionamiento es exactamente el definido en el nombre de la funcionalidad, pero se explicará el funcionamiento de algunas de estas opciones para ver su dinámica.

- **Dividir una columna en varias:** esta funcionalidad nos permite crear más columnas a partir de una existente que presente múltiples informaciones que puedan ser separadas. Un ejemplo de esto podría ser Ilustración 2-42 Menú de división de columnas en la que podemos ver como por medio de un separador se pueden separar las informaciones.

Dividir columna qNumber en varias columnas

Dividir por

☒ Separador

separador

☐ regex

Dividir en

columnas máximo (deje en blanco para no limitar)

☐ Longitud del campo

Separe listas de números con comas, p. ej., 5, 7, 15

Aceptar

Cancelar

Luego de la división

☒ Interpretar el tipo de celda

☒ Eliminar esta columna

Ilustración 2-42 Menú de división de columnas

- **Agregar columna basada en una columna:** esta funcionalidad permite crear nuevas columnas en base a otra ya existente, la forma de hacer operaciones específicas sobre las columnas es usar el lenguaje de expresión GREL, un lenguaje sencillo y fácil de usar, pero que requiere un proyecto a parte para su explicación. Se puede encontrar toda la información sobre las funcionalidades que ofrece este lenguaje en (10)

Agregar columna basada en la columna AUTOR

Nuevo nombre de la columna

En error ☒ cambiar a en blanco ☐ guardar error ☐ copiar valor de la columna original

Expresión Lenguaje Lenguaje de expresión general refinado (GREL) ▼

No hay error de sintaxis.

Vista previa Historial Con estrella Ayuda

fila	valor	valor
1.	Ana Feroso García	Ana Feroso García
2.	Secades Vidal Alonso	Secades Vidal Alonso
3.	Andrea Gutiérrez García	Andrea Gutiérrez García
4.	Sergio López García	Sergio López García
5.	Beatriz Montejo Maíllo	Beatriz Montejo Maíllo
6.	Roberto Berjón Gallinas	Roberto Berjón Gallinas

Aceptar Cancelar

Ilustración 2-43 Utilización de GREL para editar una columna

En la vista de la Ilustración 2-43 Utilización de GREL para editar una columna, podemos ver la interfaz de trabajo para trabajar con GREL sobre las columnas. Dentro de esta ventana podríamos hacer todo lo que necesitamos, por ejemplo, extraer el valor del “Q Number” de cada elemento.

→ `cell.recon.match.id`

- **Cambio de nombre de columna:** Esta funcionalidad es útil cuando por ejemplo queremos que los nombres de las columnas tengan el nombre de su propiedad Wikidata con el formato Pxxx, que nos resultara más útil en algunas ocasiones. Podemos ver su ventana en la Ilustración 2-44 Ventana de cambio de nombre de columna

127.0.0.1:3333 dice

Nuevo nombre de la columna

Cancelar Aceptar

Ilustración 2-44 Ventana de cambio de nombre de columna

- **Enriquecer con los datos cotejados:** Entre todas las opciones dichas puede que la más interesante sea la de añadir columnas en base a los datos reconciliados. Esto nos permite extraer los datos que Wikidata posee para dichas entradas en

su base de datos y enriquecer nuestro propio archivo con más información o hacer operaciones de comprobación. En este caso vamos a obtener los datos que tiene acerca de la ocupación de cada uno de los autores como podemos ver en la Ilustración 2-45 Opciones de edición de columnas y en Ilustración 2-46 Ventana emergente con las columnas de datos extraídos de . En esta última podemos ver que nos permite agregar una propiedad cualquiera y a la derecha nos deja ver los valores que se encuentran en Wikidata para esa propiedad y para el correspondiente elemento cotejado.

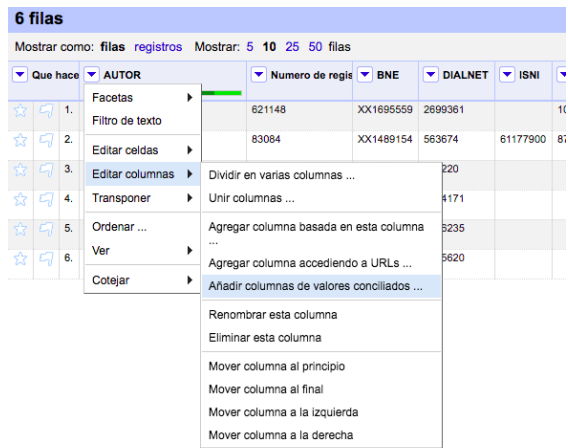


Ilustración 2-45 Opciones de edición de columnas

Añadir columnas de la columna conciliada AUTOR

Avance	
AUTOR	ocupación retirar configurar
Ana Feroso García	profesor universitario
Secades Vidal Alonso	prorector
	profesor universitario
<no reconciliado>	
Sergio López-García	investigador
Beatriz Montejo Maíllo	profesor universitario
Roberto Berjón Gallinas	investigador
	profesor universitario

Agregar propiedad	
<input type="text"/>	
Propiedades sugeridas	
<ul style="list-style-type: none"> País de ciudadanía empleador grupo étnico padre campo de trabajo imagen miembro de miembro del partido político mamá lengua materna ocupación Lugar de nacimiento lugar de entierro lugar de la muerte cargo que ocupa 	

OK Cancelar Reiniciar

Ilustración 2-46 Ventana emergente con las columnas de datos extraídos de Wikidata

- **Copiar parámetros de cotejo:** copiar los parámetros de cotejo que obtuvimos anteriormente, a la columna deseada para enriquecer lo que ya tenemos

	AUTOR	Numero de registro	BNE	DIALN
1.	Facetas	621148	XX1695559	2699361
2.	Filtro de texto	83084	XX1489154	563674
3.	Editar celdas	nueva 1217073		859220
4.	Editar columnas	1205792		3314171
5.	Transponer	1202915y		3256235
6.	Ordenar ...	83086		4995620
7.	Ver			
8.	Cotejar	Iniciar ...		
	Facetas			
	Acciones			
	Copiar información de cotejo ...			
	Usa valores como identificadores			
	Agregar columna de identificadores de entidad			

Copiar a columnas

Numero de registro UPSA

BNE

DIALNET

ISNI

VIAF

ORCID

Google Academico

Opciones para copiar

☒ Aplicar a celdas con parámetros

☒ Parámetros de cotejo nuevos

☒ Parámetros de cotejo coincidentes

Copiar

Cancelar

Todo el enriquecimiento que podamos sacar de Wikidata lo podremos usar para lo que necesitemos, usando las opciones de esta herramienta para descarga de archivos de nuestros proyectos en distintos formatos. Esa es la verdadera utilidad del uso de los datos de la BD de Wikidata.

2.4.5 TRANSFORMACIÓN DE STATEMENTS Y SUBIDA A WIKIDATA

Para poder importar nuestros datos de enriquecimiento o creación de elementos a Wikidata tenemos dos métodos distintos. Utilizar OpenRefine directamente para subir las ediciones a Wikidata o Exportarlo a formato QuickStatements.

Realmente la única diferencia que hay entre un método u otro es la herramienta que lo llevará a cabo. En el caso de OpenRefine se transforma las ediciones a comandos de una forma transparente, invisible para el usuario y se suben directamente (resulta muy cómodo para pequeños volúmenes de datos). El otro método, la exportación de un archivo .txt con los comandos se trasladará a la herramienta QuickStatements para su subida en masa, este proceso se explicará más adelante con mayor detenimiento.

En la **Ilustración 2-47 Acceso al menú de funciones de Wikidata** se puede ver el menú de funcionalidades de Wikidata (desplegable en la parte superior derecha de la pantalla). Pero antes de aplicar ninguna edición debemos de generar el *schema* de las ediciones.



Ilustración 2-47 Acceso al menú de funciones de Wikidata

Esquema es una plantilla que muestra nuestros datos tabulares de una forma accesible para ser transformados en ediciones de Wikidata. Inicialmente la interfaz nos muestra las columnas de nuestras tablas, diferenciando las conciliadas.

En la parte superior de la pantalla **Ilustración 2-48 Pantalla de inicio de Schema**, podemos ver un índice de pestañas donde tendremos las filas de los elementos del *Schema* que se ve en la imagen, los mensajes de los errores que vayan surgiendo y una vista previa de las ediciones que cuando todo es correcto nos muestra el resultado que se verá en Wikidata. Las filas de los apartados anteriores, la de *Schema* en la que estamos, una de *Asuntos* y otra que se llama *Avance*. Posteriormente podemos ver los nombres de las columnas de datos que en este caso han sido nombrados con el formato de propiedades de Wikidata (Pxxx) por simple comodidad a la hora de trabajar

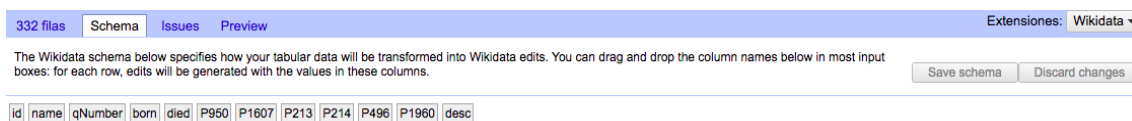


Ilustración 2-48 Pantalla de inicio de Schema

Lo primero que debemos hacer es añadir el objeto conciliado del cual queremos hacer modificaciones, en este caso el *name* (representa al nombre de una autoridad), que hemos conciliado con el objeto Humano (Q5). La modificación que le vamos a hacer es añadir la propiedad del Identificador de la Universidad Pontificia de Salamanca porque muchos de nuestros autores no lo tienen. El procedimiento puede verse en la **Ilustración 2-49 Schema enriquecimiento de identificador UPSA para autoridades**. Lo primero que debemos hacer es añadir el elemento a modificar como un nuevo ítem del *schema*.

Después de fijar el nuevo ítem como el dato de la columna *name* lo que debemos hacer es añadir un nuevo *statement*, esto se traduce como identificar la propiedad Wikidata que queremos editar, en este caso identificador de la Universidad Pontificia de Salamanca (P9247) y asignarle la correspondiente columna de nuestros datos, tal y como aparece en la imagen.

Siempre es necesario añadir referencias de la localización donde se puede verificar que efectivamente se encuentra el dato en cuestión, asegurándonos de que si se necesita se puede llegar hasta él. En este caso usamos la propiedad de afirmado en (P248) que nos indica que se encuentra en el Catálogo de la UPSA. También el identificador de la UPSA correspondiente al autor con la propiedad Identificador de la UPSA (P9247) y el valor del elemento de nuestra columna de datos. Por último, también es conveniente añadir la fecha en la que se insertaron los datos en el formato siguiente (yyyy-mm-dd)

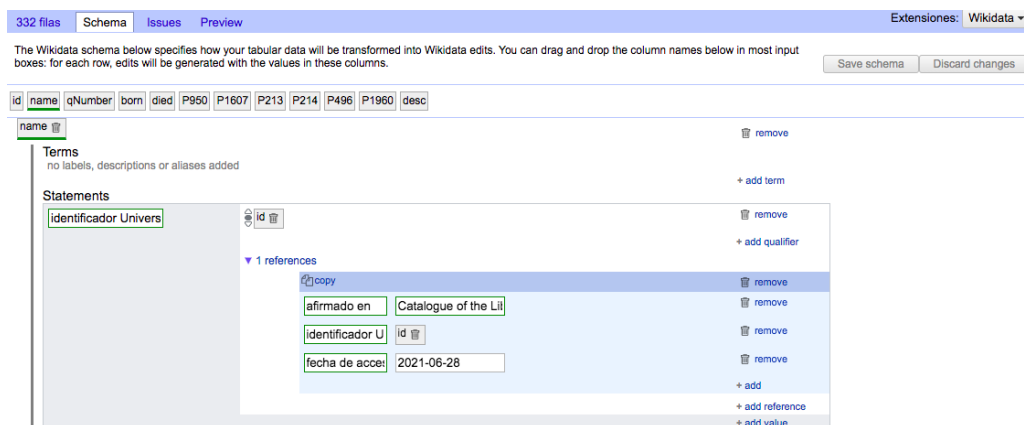


Ilustración 2-49 Schema enriquecimiento de identificador UPSA para autoridades

Hasta ahora sólo se ha explicado cómo podemos enriquecer. En caso de que lo que queramos es crear nuevos elementos de Wikidata, el schema cambiará ligeramente.

En la **Ilustración 2-50 Ejemplo de creación de nuevos elementos en OpenRefine**, podemos ver un ejemplo en el que se crean Publicaciones de la revista Cuadernos Salmantinos. Para poder crear ítems desde cero será totalmente obligatorio la creación de *términos* en el *schema*. Siendo imprescindible la creación de un *label* y opcional la inclusión de una descripción. También se

deberá hacer uso de la propiedad *instancia de* para especificar el tipo de elemento Wikidata al que hace referencia, en este caso la publicación de la revista es un *artículo científico* y se establecerá como tal.

El resto de los elementos que se usan en el esquema, serán los mismos que cuando enriquecemos datos ya existentes.

Ilustración 2-50 Ejemplo de creación de nuevos elementos en OpenRefine

En la pestaña de *Issues* o *Asuntos* es donde se nos presentan los posibles mensajes de errores o advertencias sobre lo que estamos haciendo en el *schema*. En nuestro caso simplemente estamos enriqueciendo lo que ya existe y no nos aparecerán más mensajes de error que la advertencia de la falta referencias en nuestros datos en el caso de que no se hayan referenciado los datos.

Por otro lado, en el caso de que lo que hagamos sea crear nuevos elementos desde cero nos aparecerán una serie de requisitos para que sea posible su importación a Wikidata, estos serán los comentados en párrafos anteriores sobre la creación de nuevos elementos. Los posibles errores que nos saldrán en ese caso serán los que aparecen en la **Ilustración 2-51 Lista de errores elementos nuevos**, todo esto en el caso de que se haga sin conocimiento de el procedimiento a seguir, si se sigue lo comentado en párrafos anteriores no aparecerá mención alguna.

Por lo tanto como se ve en la imagen, para elementos nuevos deberemos proporcionarles un *label* o *etiqueta* y una *descripcion*, aunque esta segunda no es algo puramente obligatorio ya que aparece como una advertencia y no como un error fatal. También nos dice que debemos *referenciar* los datos que añadamos y que debemos asociarles algún tipo mediante la propiedad que se comentó anteriormente de *instancia de*.

Finalmente también sale un mensaje informativo diciendo que si importas los datos en cuestión se estará creando nuevos elementos Wikidata.

Para un mayor conocimiento de los elementos usables en schema de OpenRefine se recomienda consultar (11)

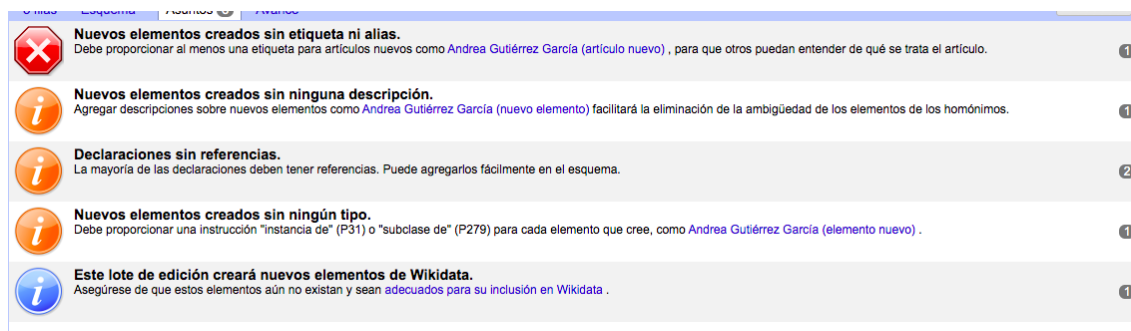


Ilustración 2-51 Lista de errores elementos nuevos

Para finalizar sólo nos queda hablar de la ventana que aparece en la anterior imagen, la *preview* o *previsualización*, en ella como se comento se puede ver una vista previa de las ediciones. Un ejemplo de esto seria la **Ilustración 2-52 Ejemplo de previsualización sobre las ediciones**, en donde se puede ver que cambios se harán sobre una autoridad

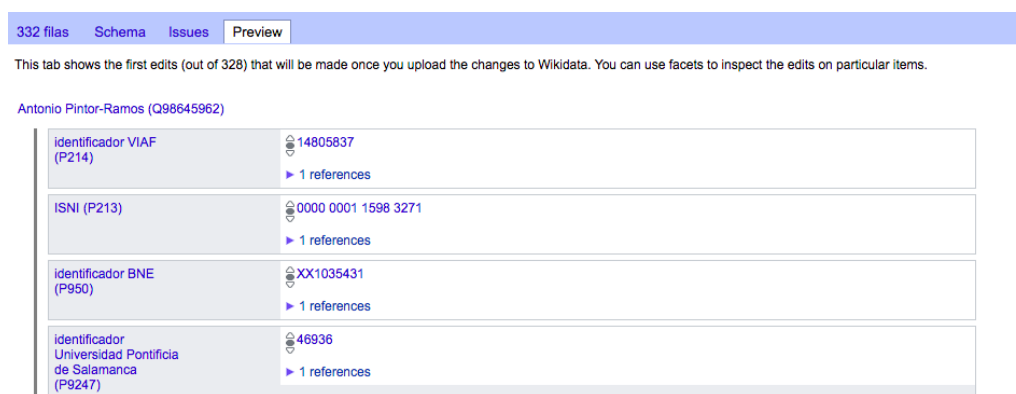


Ilustración 2-52 Ejemplo de previsualización sobre las ediciones

Se recomienda separar los datos de nueva creación de los que ya hemos conseguido conciliar y añadirlos a Wikidata en dos operaciones distintas. También se debe señalar que las entradas que no se reconcilien con nada o que no se las establezca como nuevos elementos de creación, no se tendrán en cuenta a la hora de exportar los resultados.

Resulta muy interesante la propiedad de exportar el *Schema* como un fichero Json para poder reutilizarlo en futuras ocasiones, se puede acceder a esta funcionalidad en el menú desplegable de Wikidata que está en la parte superior derecha de la ventana visible en la **Ilustración 2-53 Menú desplegable de opciones de** .

Si los datos que añadimos son los mismos que ya existen simplemente serán descartados por Wikidata, por lo que no tendremos que preocuparnos de que se sobrescriban. Por otro lado, si ya existe un dato y se añade otro distinto simplemente se añadirá al lado del que ya existe y serán los administradores de Wikidata lo que decidan qué hacer.

Finalmente, una vez hayamos terminado de preparar todo el proyecto accederemos de nuevo al menú desplegable de Wikidata como podemos ver en la imagen de la **Ilustración**

2-53 Menú desplegable de opciones de . En donde podremos exportar los datos a un fichero de comandos Quickstatements o subirlo directamente desde la herramienta.

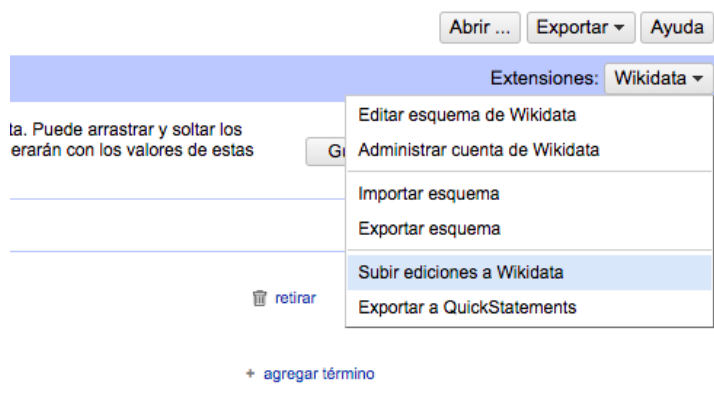


Ilustración 2-53 Menú desplegable de opciones de Wikidata

2.4.6 VALORACIÓN DE LA HERRAMIENTA

La herramienta OpenRefine es una buena interfaz para trabajar con ediciones de Wikidata, tanto nuevas creaciones como enriquecimiento de datos o su limpieza, realmente es muy completa. Dicho esto recalcar que aunque pueda hacer las operaciones de las demás herramientas, no siempre es la adecuada para hacerlo, como por ejemplo el caso de la autoridades es como se explico en apartados anteriores ideal para el uso de Mix'n'Match por la simple razón que algunos de ellos existen y otros no, siendo pocos los que no existen y Quickstartements más necesaria para importaciones de gran tamaño

2.5 HERRAMIENTA OPENREFINE

Es una herramienta diseñada por Magnus Manske al igual que lo fue Mix'n'Match, a diferencia de que la principal función de QuickStatements es usar una serie de comandos para hacer ediciones en Wikidata. Puede añadir todo tipo de declaraciones, como las que se han usado en el *Schema* de OpenRefine para editar los objetos Wikidata.

Los comandos para trabajar con esta herramienta se pueden escribir directamente en la ventana de importación o copiar y pegar de un documento de texto externo. Una vez que ya hemos escrito los comandos se ejecutará la herramienta y terminará la importación de los datos a Wikidata. En nuestro caso los comandos los copiaremos del fichero que exportamos desde OpenRefine.

Existen dos versiones de la aplicación y aunque las dos funcionan la opción válida y que se recomienda es la versión 2 (QuickStatements V2)

Esta aplicación está preparada para la subida de grandes bloques de datos a Wikidata, evitando las entradas duplicadas y permitiendo una inserción limpia de los mismos.

2.5.1 MÉTODO DE TRABAJO

En nuestro caso usaremos la versión 2 de la aplicación y usaremos el archivo que extraemos de OpenRefine como explicamos anteriormente. Podemos ver una muestra de los comandos en la **Ilustración 2-54 Ejemplo de los comandos de las ediciones sobre un elemento en el archivo Quickstatements que obtenemos al exportar de OpenRefine.**

```
Q98645962      P9247      "46936" S248      Q105667100      S9247      "46936" S813      +2021-06-28T00:00:00Z/11
Q98645962      P950       "XX1035431" S248      Q105667100      S9247      "46936" S813      +2021-06-28T00:00:00Z/11
Q98645962      P213       "0000 0001 1598 3271" S248      Q105667100      S9247      "46936" S813      +2021-06-28T00:00:00Z/11
Q98645962      P214       "14805837" S248      Q105667100      S9247      "46936" S813      +2021-06-28T00:00:00Z/11
```

Ilustración 2-54 Ejemplo de los comandos de las ediciones sobre un elemento en el archivo Quickstatements que obtenemos al exportar de OpenRefine

Deberemos de copiar y pegar los comandos que hay en este archivo .txt y acceder a la herramienta disponible en la dirección (12).

La página principal de la herramienta podemos verla en la **Ilustración 2-55 Página principal QuickStatements**

Ilustración 2-55 Página principal QuickStatements

La metodología de trabajo es la siguiente. Pulsaremos en el botón “Lote nuevo”, nos aparecerá un cuadro donde podremos pegar los comandos copiados anteriormente y de nuevo pulsamos en el botón “importar comandos V2”. Al hacer esto los interpretará como sentencias no como simplemente código y el formato del texto cambiará. También podemos pulsar el botón de importar los datos mediante archivo .csv, pero para nuestro caso usaremos el anterior. Todo lo comentado anteriormente se puede ver en el ejemplo de la **Ilustración 2-56 Ejemplo de formato de comandos importados en Quickstatements.**

QuickStatements español Lote nuevo Últimos lotes Charla Git Ayuda Braisrodriguezalvo Tus últimos lotes

Lote #58169 "Some artices magazine Cuadernos Salmantinos de Filosofia" on Wikidata por Braisrodriguezalvo [Lotes] [Discutir/revertir lote](#)

Estado: DONE 100% (5) de 5 hecho

1	init	CREATE Item	<p>es:El espíritu en C.G. Jung</p> <p>es:Artículo de la revista Cuadernos Salmantinos</p> <p>instance of [P31]:scholarly article [Q13442814] retrieved [P813]:2021-07-01</p> <p>title [P1476]:es:"El espíritu en C.G. Jung" retrieved [P813]:2021-07-01</p> <p>author [P50]:Antonio García Vázquez [Q16874986] retrieved [P813]:2021-07-01</p> <p>stated in [P248]:Cuadernos Salmantinos de Filosofia [Q65929181]</p>
2	init	CREATE Item	<p>es:Ser y conocer: Peculiaridades informáticas de la especie humana</p> <p>es:Artículo de la revista Cuadernos Salmantinos</p> <p>instance of [P31]:scholarly article [Q13442814] retrieved [P813]:2021-07-01</p> <p>title [P1476]:es:"Ser y conocer: Peculiaridades informáticas de la especie humana" retrieved [P813]:2021-07-01</p> <p>author [P50]:Marceliano Arranz Rodrigo [Q99197429] retrieved [P813]:2021-07-01</p> <p>stated in [P248]:Cuadernos Salmantinos de Filosofia [Q65929181]</p>
3	init	CREATE Item	<p>es:Biobibliografía</p> <p>es:Artículo de la revista Cuadernos Salmantinos</p> <p>instance of [P31]:scholarly article [Q13442814] retrieved [P813]:2021-07-01</p> <p>title [P1476]:es:"Biobibliografía" retrieved [P813]:2021-07-01</p> <p>author [P50]:Vicente Muñoz Delgado [Q55229413] retrieved [P813]:2021-07-01</p> <p>stated in [P248]:Cuadernos Salmantinos de Filosofia [Q65929181]</p>
4	init	CREATE Item	<p>es:Abrahan H. Maslow y la teoría holístico/dinámica de las necesidades</p> <p>es:Artículo de la revista Cuadernos Salmantinos</p> <p>instance of [P31]:scholarly article [Q13442814] retrieved [P813]:2021-07-01</p> <p>title [P1476]:es:"Abrahan H. Maslow y la teoría holístico/dinámica de las necesidades" retrieved [P813]:2021-07-01</p> <p>author [P50]:Abrahan H. Maslow [Q13442814] retrieved [P813]:2021-07-01</p> <p>stated in [P248]:Cuadernos Salmantinos de Filosofia [Q65929181]</p>

Ilustración 2-56 Ejemplo de formato de comandos importados en Quickstatements

Para continuar deberemos escoger el modo de ejecución que puede ser “Ejecución” o “Ejecución por lotes”. Esta última cuenta con una serie de ventajas, dado que no se ejecuta directamente en el navegador del PC sino que se ejecuta en un servidor Wikimedia, con lo que si por casualidad nuestro ordenador falla o se cierra el navegador, continuará la ejecución sin problemas y podremos ejecutar varios lotes paralelamente sin preocupación de que se ralentice el procesamiento de nuestra computadora. Para conocer más de este modo de funcionamiento se puede consultar (13).

También nos preguntará sobre el nombre que le queremos dar a nuestro lote, algo importante para identificar las ediciones que realizan los distintos usuarios en Wikidata y empezará la ejecución. En la parte superior de la ventana podremos ver una barra de progreso que irá marcando las entradas creadas correctamente y las que tienen errores. Los errores en ciertos casos son muy comunes debido a que puede que no termine de ejecutar la parte del elemento Wikidata y ya empiece con sus propiedades provocando que no se envíen los datos correctamente, pero esto se puede corregir rápidamente dándole al botón “Try to reset errors” y veremos cómo se van corrigiendo.

Si el error sea de otro tipo no se conseguirá subir correctamente y deberemos hacer cambios en el elemento para que todo vaya bien.

Comentar también que en la ventana Y rodeando a las entradas, hay una serie de herramientas que nos permiten visualizar tanto las mismas entradas como los errores de subida, así como el número de entradas o los lotes con los que hemos trabajado antes.

Para poder ver las contribuciones que hemos hecho a Wikidata deberemos de iniciar sesión con nuestro usuario y contraseña en nuestra cuenta Wikimedia y aparecerán los distintos lotes que hemos subido. Si por casualidad algo en el proceso sale mal y queremos que se eliminen los cambios deberemos de acudir a los administradores de Wikidata mediante la propia herramienta. Puedes acceder a tus lotes en el botón de la parte de arriba a la derecha de la ventana donde obtendremos todos los lotes como aparece en la

QuickStatements español Lote nuevo Últimos lotes Charla Git Ayuda Braisrodriguezcalvo Tus últimos lotes

Últimos lotes de la cuenta Braisrodriguezcalvo

#	Usuario	Nombre	Estado	Última modificación	Acciones
#58600	Braisrodriguezcalvo	create other publications of the journal Cuadernos Salmantinos de Filosofia	DONE DONE:138 ERROR:1	2021-07-06 10:31:08	Discutir/revertir lote
#58506	Braisrodriguezcalvo	Muestra creacion publicaciones revista Cuadernos Salmantinos	DONE DONE:3	2021-07-05 11:25:14	Discutir/revertir lote
#58169	Braisrodriguezcalvo	Some articles magazine Cuadernos Salmantinos de Filosofia	DONE DONE:3	2021-07-01 08:40:14	Discutir/revertir lote
#57801	Braisrodriguezcalvo	Data enrichment for UPSA authorities	DONE DONE:3251 ERROR:1	2021-06-28 21:29:46	Discutir/revertir lote
#57684	Braisrodriguezcalvo	test enriquecer autoridades de la UPSA	DONE DONE:167	2021-06-27 19:33:52	Discutir/revertir lote
#57683	Braisrodriguezcalvo		STOP DONE:1 INE:169	2021-06-27 19:27:02	Discutir/revertir lote
#57682	Braisrodriguezcalvo		STOP INE:167	2021-06-27 19:26:27	Discutir/revertir lote
#57134	Braisrodriguezcalvo	dialnet for authorities	DONE DONE:1065	2021-06-22 16:21:28	Discutir/revertir lote

En la parte de la derecha podemos ver unos botones que ponen *discutir/revertir* donde podremos instar a los administradores a que hagan cambios sobre nuestras ediciones.

Para esta herramienta no se diseñará una valoración ya que únicamente la utilizamos para la importación final de los datos, pero si que es necesario recalcar que inicialmente para poder realizar cualquier importación de datos sobre esta herramienta será necesario por supuesto contar con una cuenta de usuario Wikimedia, como el caso de las otras herramientas, y además como requisito extra tener una antigüedad en la cuenta de unos días y haber realizado unas 50 ediciones con esa cuenta para poder trabajar con la herramienta

Para poder hacer las ediciones previas lo que se debe hacer es trabajar con un pequeño conjunto de datos (≥ 50) con OpenRefin por ejemplo y hacer una importación directa sin necesidad de exportar el fichero txt de comandos QuickStatement

3. DESARROLLO Y APLICACIÓN PRÁCTICA

En este apartado se explicará los procedimientos que se deben de seguir para conseguir cumplir el caso de uso de publicación de la información sobre sobre autoridades de la UPSA y publicaciones de los artículos escritos por éstas. Las ediciones que se han llevado a cabo en este proyecto han sido refrentadas por un experto y un administrador de Wikidata para garantizar su corrección.

3.1 DESCRIPCIÓN DEL CASO DE USO PARA NUESTRA BIBLIOTECA

El caso de uso que debemos llevar a cabo se puede definir como “Migración de autoridades de la Universidad Pontificia de Salamanca y las publicaciones que a éstas correspondan de la revista Cuadernos Salmantinos de Filosofía”. Aplicando lo presentado en apartados anteriores, se deberá de llevar a cabo la importación de las autoridades de la UPSA y las publicaciones que hayan realizado en la revista antes citada.

Atendiendo al caso de uso descrito, para llevar a cabo la aplicación a nuestro proyecto deberemos obtener los datos pertenecientes a las autoridades que realicen publicaciones en la revista “*Cuadernos Salmantinos de Filosofía*”. Esta información se extraerá directamente del repositorio institucional de la UPSA, Summa. Este servicio es ofrecido por la Biblioteca de la Universidad Pontificia de Salamanca para ofrecer acceso a una extensa cantidad de contenido digital con relación a toda la información tanto pasada como futura de la universidad, entre los documentos podemos encontrar libros, revistas, audiolibros, conferencias... En nuestro caso nos servirá para obtener la información sobre los autores y sus artículos de la revista que necesitamos, desde 1974 - 2020.



Para hacer este proyecto es necesario tener conocimiento básico de la filosofía y funcionamiento de la base de conocimiento que utilizaremos para exportar nuestras autoridades y darles acceso público, explicado en los apartados anteriores.

Se debe crear un algoritmo en un lenguaje de programación que sea capaz de extraer de los archivos de datos y seleccionar la información que debemos usar de forma que se pueda usar toda la información en conjunto como fuente de datos para las herramientas citadas. Con este procedimiento se intenta reducir el número de duplicados y datos inútiles, usando unos datos más limpios desde el principio.

Se debe de crear una cuenta de Wikimedia para poder trabajar con las herramientas necesarios sobre Wikidata. De todos los cambios que hagamos sobre Wikidata quedará constancia en nuestro perfil.

Deberemos de fijar los conocimientos básicos sobre funcionalidad y filosofía de uso de las herramientas Mix'n'match, OpenRefine y QuickStatement que se han adquirido en el apartado anterior.

3.2 DESCRIPCIÓN DEL CASO DE USO PARA NUESTRA BIBLIOTECA

FASES DEL PROYECTO

- Fase 1: Fase de preparación inicial de los datos
- fase 2: Fase de creación de enriquecimiento y creación de las autoridades de la Universidad Pontificia de Salamanca.
- Fase 3: Fase de creación de los artículos publicados por las autoridades de la UPSA

3.3 APLICACIÓN DEL PROCESO DE MIGRACION A WIKIDATA

La aplicación del proceso estaría formada por 4 partes. La aplicación escrita en lenguaje java que permite interactuar y hacer operaciones con los distintos archivos, la herramienta Mix'n'match, Openrefine y Quickstatements. Los cuatro programas funcionando en conjunto.

3.3.1 FASE 1: PREPARACIÓN INICIAL DE LOS DATOS

Extracción de los datos de las autoridades de la UPSA se hará a través de un archivo .csv proporcionado por la biblioteca y también la obtención de la información de las distintas publicaciones de la revista “Cuadernos Salmantinos”, a través una petición a SUMMA, de donde extraeremos un archivo json con la información sobre las distintas publicaciones.

CONSULTA A SUMMA:

[https://summa.upsa.es/json/select.vm?query=id:0000000488+OR+id%3A%280000001007+OR+0000001032+OR+0000001051+OR+0000001068+OR+0000108406+OR+0000001085+OR+0000001103+OR+0000001123+OR+0000001142+OR+0000001142+OR+0000001162+OR+0000001162+OR+0000001222+OR+00000029046+OR+00000029065+OR+00000029846+OR+00000030293+OR+00000030355+OR+00000030371+OR+00000030662+OR+00000031400+OR+00000032425+OR+00000033498+OR+00000039699+OR+00000039733+OR+00000044926+OR+00000048613+OR+0000000489+OR+0000000509+OR+0000000533+OR+0000000559+OR+0000000580+OR+0000000609+OR+0000000642+OR+0000000671+OR+0000000694+OR+0000000717+OR+0000000744+OR+0000000783+OR+0000000808+OR+0000000827+OR+0000000849+OR+0000000872+OR+0000000891+OR+00000092686+OR+0000000932+OR+0000000949+OR+0000000968+OR+0000000988+OR+0000132178%29&length=1000000](https://summa.upsa.es/json/select.vm?query=id:0000000488+OR+id%3A%280000001007+OR+0000001032+OR+0000001051+OR+0000001068+OR+0000108406+OR+0000001085+OR+0000001103+OR+0000001123+OR+0000001142+OR+0000001142+OR+0000001162+OR+0000001162+OR+0000001222+OR+00000029046+OR+00000029065+OR+00000029846+OR+00000030293+OR+00000030355+OR+00000030371+OR+00000030662+OR+00000031400+OR+00000032425+OR+00000033498+OR+00000039699+OR+00000039733+OR+00000044926+OR+00000048613+OR+000000489+OR+0000000509+OR+0000000533+OR+0000000559+OR+0000000580+OR+0000000609+OR+0000000642+OR+0000000671+OR+0000000694+OR+0000000717+OR+0000000744+OR+0000000783+OR+0000000808+OR+0000000827+OR+0000000849+OR+0000000872+OR+0000000891+OR+00000092686+OR+0000000932+OR+0000000949+OR+0000000968+OR+0000000988+OR+0000132178%29+OR+parent%3A%280000001007+OR+0000001032+OR+0000001051+OR+0000001068+OR+0000108406+OR+0000001085+OR+0000001103+OR+0000001123+OR+0000001142+OR+0000001142+OR+0000001162+OR+0000001162+OR+0000001222+OR+00000029046+OR+00000029065+OR+00000029846+OR+00000030293+OR+00000030355+OR+00000030371+OR+00000030662+OR+00000031400+OR+00000032425+OR+00000033498+OR+00000039699+OR+00000039733+OR+00000044926+OR+00000048613+OR+0000000489+OR+0000000509+OR+0000000533+OR+0000000559+OR+0000000580+OR+0000000609+OR+0000000642+OR+0000000671+OR+0000000694+OR+0000000717+OR+0000000744+OR+0000000783+OR+0000000808+OR+0000000827+OR+0000000849+OR+0000000872+OR+0000000891+OR+00000092686+OR+0000000932+OR+0000000949+OR+0000000968+OR+0000000988+OR+0000132178%29&length=1000000)

El resultado obtenido por pantalla debe guardarse como un nuevo fichero en formato json,. A continuación, se puede ver en la Ilustración 3-1 Estructura archivo json de publicaciones y también un extracto del mismo en la Ilustración 3-2 Muestra del archivo json con las publicaciones de la revist. Este archivo contará con tres tipos de elementos (article, magazine, magazinePublication) que se explicarán más adelante.

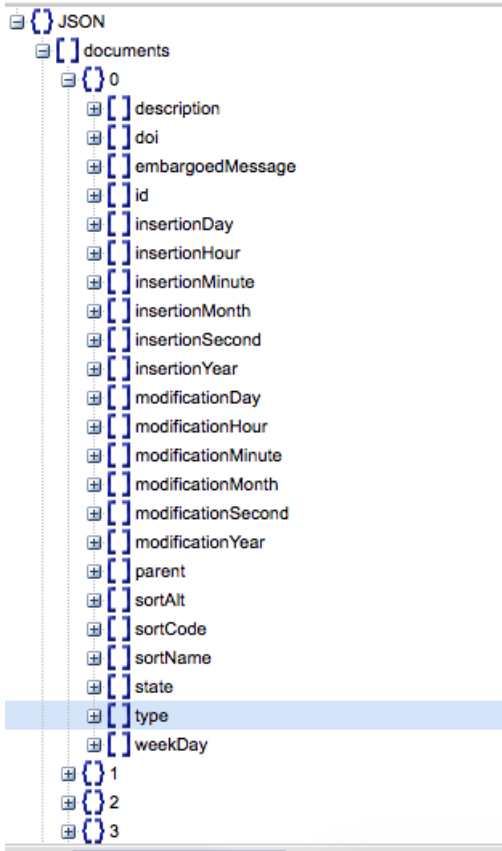


Ilustración 3-1 Estructura archivo json de publicaciones

```

"documents": [
  {
    "description": [
      ""
    ],
    "doi": [
      "10.36576/summa.488"
    ],
    "embargoedMessage": [
      ""
    ],
    "id": [
      "0000000488"
    ],
    "insertionDay": [
      "10"
    ],
    "insertionHour": [
      "11"
    ],
    "insertionMinute": [
      "40"
    ],
    "insertionMonth": [
      "04"
    ],
    "insertionSecond": [
      "31"
    ],
    "insertionYear": [
      "2012"
    ],
    "modificationDay": [
      "06"
    ],
    "modificationHour": [
      "02"
    ],
    "modificationMinute": [
      "00"
    ],
    "modificationMonth": [
      "12"
    ],
    "modificationSecond": [
      "22"
    ],
    "modificationYear": [
      "2020"
    ],
    "parent": [
      "0000033056"
    ],
    "sortAlt": [
      "Cuadernos Salmantinos de Filosofía"
    ],
    "sortCode": [
      "CSF"
    ],
    "sortName": [
      "Cuadernos Salmantinos de Filosofía"
    ],
    "state": [
      "00"
    ],
    "type": [
      "magazinePublication"
    ],
    "weekDay": [
      ""
    ]
  },
  {
    "doi": [
      "10.36576/summa.1142"
    ],
    "endDay": [
      "31"
    ],
    "endIssueNumber": [
      ""
    ]
  }
]

```

Ilustración 3-2 Muestra del archivo json con las publicaciones de la revista

El archivo .csv de partida que se nos ha proporcionado desde la biblioteca requiere una limpieza de caracteres extraños antes de ser utilizado ya que su formato no es compatible con el entorno de NetBeans, la causa que posiblemente haya originado este problema es probablemente el copiado y pegado en el formato de origen el texto en cuestión, lo cual provoca las incompatibilidades en la lectura de los caracteres como se puede ver en la Ilustración 3-3 Archivo original de autoridades de la UPSA.

```
AUTOR;N. Reg. UPSA;BNE (el 001);DIALNET;ISNI;VIAF;ORCID;Google AcadMico;
;;;;;;
Abad Buil, Jos  Mar a, 1928?-2018;474959;XX1719730;1157097;0000000066363657;283
Abelairas G mez, Cristian;1227805;;3226578;;39149233632676500000;0000-0002-1056
Acevedo Gonz lez, Horacio;755773;;3575569;;;;;
Alameda Garc a, David;206790;XX1501135;607256;;87226734;0000-0001-7016-5896;ysu
Alfonso S nchez, Jos  Manuel, 1964-;168669;XX1506436;1228379;0000000059372985;4
Alonso Gonz lez, Carmen Mar a;320005;XX1687312;2122396;0000000081617969;8742707
Alonso Secades, Vidal;83084;XX1489154;563674;0000000061177900;87219768;;wM89Pzc
 lvarez Alonso, Jos  F lix, 1984-;1199842;XX5587581;3839855;;514602936983582000
 lvarez  lvarez, Jorge Luis;1235398;;;;;;
 lvarez Garc a, Francisco Jos , 1974-;786272;XX1582987;3071602;0000000072838196
 lvarez G mez, Mariano, 1935-2017;75768;XX1724462;15067;0000000066387181;173363
 lvarez Rico, Manuel;211340;XX1092260;2484833;0000000080160274;56671976;;
 lvarez Turienzo, Saturnino, 1920-;82361;XX856814;155522;0000000117370137;25123
Alves Vas, Francisco Javier, 1980-;1193890;;3356719;;0000-0003-0853-1765;;
Am ria Jim nez, Mario;1225826;;2277231;;0000-0001-7826-256X;;
```

Ilustraci n 3-3 Archivo original de autoridades de la UPSA

Los problemas surg an por los caracteres que llevaban tildes, di resis u otros s mbolos que se desv an de la implementaci n m s simple de la escritura. Era en esos casos en los que estos s mbolos aparec an substituidos por uno o m s caracteres extra os que el propio programa interpret  y dado que pr cticamente todos los nombres contaban con errores y tras buscar una soluci n, se intent  que por c digo se especificara el modo de lectura de las l neas de fichero como UTF-8 para solucionar las incompatibilidades.

Este m todo no result , pasando a la siguiente idea que ser a implementar un peque o programa en java que solucionara las incompatibilidades l neas a l nea usando patrones.

Esto no se pudo hacer ya que no se sigue un patr n fijo a simple vista, y finalmente se opt  por la soluci n m s simple pero correcta, que fue usar la herramienta de reemplazo de caracteres deduciendo cuales eran los cambios para obtener el nombre correcto.

Pero no solo fueron los nombres los que se vieron afectados por estos problemas, ya que los c digos de los identificadores tambi n sufrieron variaciones, las cuales no se han podido identificar hasta el momento de la inserci n de los datos.

El trabajo de limpieza de los datos es la parte del proyecto m s tediosa y que m s quebraderos de cabeza han dado, ya que es del todo imposible saber si los datos est n bien o mal sin tener una referencia y conlleva momentos de frustraci n y vueltas a empezar a programar con cosas que considerabas que eran correctas un n mero elevado de veces.

Mi recomendaci n es cerciorarse de que todos los datos son correctos desde el principio

Así es como quedaría el formato del fichero tras la limpieza Ilustración 3-4 Datos de las autoridades en formato csv totalmente limpios.

```

1  AUTOR;nacimiento;defuncion;N_Reg_UPSA;BNE;DIALNET;ISNI;VIAF;ORCID;Google Académico
2  ;;;;;;
3  Abad Buil, José María;1928;2018;474959;XX1719730;1157097;0000 0000 6636 3657;28337773;;
4  Abelairas Gómez, Cristian;;;1227805;;3226578;;39149233632676512753;0000-0002-1056-7778;5zrm09wAAAAJ
5  Acevedo González, Horacio;;;755773;;3575569;;;
6  Alameda García, David;;206790;XX1501135;607256;;87226734;0000-0001-7016-5896;ysurk-cAAAAJ
7  Alfonso Sánchez, José Manuel;1964;;168669;XX1506436;1228379;0000 0000 5937 2985;46174455;0000-0002-2674-1021;3W7AmzoAAAAJ
8  Alonso González, Carmen María;;320005;XX1687312;2122396;0000 0000 8161 7969;87427079;0000-0001-5528-7705;dWAS4DAAAAAJ
9  Alonso Secades, Vidal;;83084;XX1489154;563674;0000 0000 6117 7900;87219768;;wM89PzcAAAAJ
10 Álvarez Alonso, José Félix;1984;;1199842;XX5587581;3839855;;5146029369835820278;0000-0001-9380-3010;bm8vQXkAAAAJ
11 Álvarez Álvarez, Jorge Luís;;;1235398;;;
12 Álvarez García, Francisco José;1974;;786272;XX1582987;3071602;0000 0000 7283 8196;102423191;0000-0001-8102-5418;WaqSJREAAAJ
13 Álvarez Gómez, Mariano;1935;2017;75768;XX1724462;15067;0000 0000 6638 7181;17336382;;
14 Álvarez Rico, Manuel;;;211340;XX1092260;2484833;0000 0000 8016 0274;56671976;;
15 Álvarez Turienzo, Saturnino;1920;;82361;XX856814;155522;0000 0001 1737 0137;2512361;;
16 Alves Vas, Francisco Javier;1980;;1193890;;3356719;;0000-0003-0853-1765;
17 Amatria Jiménez, Mario;;1225826;;2277231;;0000-0001-7826-256X;
18 Andaluz Romanillos, Ana María;1957;;82482;XX1085421;51622;0000 0000 5925 0267;44984848;0000-0002-9033-0710;TUPwN7EAAAAJ
19 Andrés Marcos, Luís;;572573;;2022618;;;
20 Arias García, Elisa;;;1187815;;2291521;;0000-0001-6208-3460;
21 Arranz García, Olga;;794503;;3003106;;0000-0003-2145-6173;AraENNEAAAAJ
22 Arranz Rodrigo, Marceliano;1943;;359216;;348948;0000 0000 6232 4220;313230408;;
23 Arriba de la Fuente, José Ángel de;;171981;;1967437;;;
24 Arroyo del Bosque, Rubén;;1226595;;4440738;;0000-0002-2460-2140;51E9PlcAAAAJ
25 Aubach, María Teresa;1930;2002;1981;XX963381;532531;0000 0000 7978 2172;77128019;;
26 Aznar Gil, Federico;1953;;1207110;XX824449;266256;0000 0001 1839 0519;89037400;;
27 Baigorri, Artemio J.;1956;;1204924;XX1036284;172326;0000 0000 5922 2338;7714813;0000-0003-4427-900X;AoZXT24AAAAJ

```

Ilustración 3-4 Datos de las autoridades en formato csv totalmente limpios

3.3.2 FASE 2: FASE DE GENERACIÓN DE ENRIQUECIMIENTO Y CREACIÓN DE LAS AUTORIDADES DE LA UNIVERSIDAD PONTIFICIA DE SALAMANCA

Una vez hayamos conseguido limpiar los datos y tengamos los archivos base necesarios para realizar las importaciones haremos uso de una aplicación que transformará esos datos en formato adecuado para que los procesen las herramientas de importación de datos como Mix'n'match o OpenRefine.

Este proceso se llevará a cabo mediante un algoritmo programado en Java Application with Maven con IDE Apache NetBeans 12.2 y su interfaz de usuario se ha implementado por medio java swing.

El código programado está acompañado de los comentarios pertinentes que facilitan su entendimiento

La aplicación contará con un menú con dos elementos, uno que hace referencia a las operaciones que podremos hacer "*Operaciones*" y otro que nos permite finalizar la ejecución de la aplicación "*Salir*". El apartado de *Operaciones* cuenta con más funcionalidades, las "*Operaciones con Autoridades*" y las "*Operaciones con Publicaciones*" como se puede ver en la siguiente imagen **Ilustración 3-5 Menú de operaciones (PROVISIONAL)**.

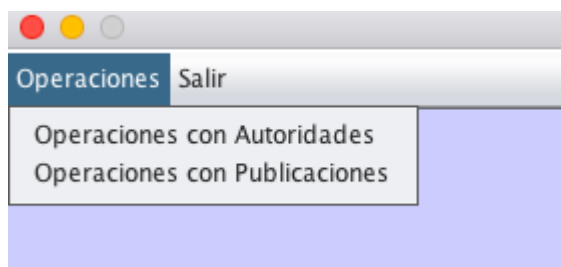


Ilustración 3-5 Menú de operaciones (PROVISIONAL)

La primera funcionalidad que se va a explicar corresponde a las autoridades que se presenta en la aplicación de la forma en la que se ve en la **Ilustración 3-6 Vista de operaciones con las autoridades de la aplicación**.

Tenemos cuatro cuadros con los que deberemos de trabajar. Se deberá empezar por el que se encuentra en la parte superior izquierda de la pantalla. En él se escogerá el archivo csv de autoridades del que se ha hecho la limpieza anteriormente. Usando el explorador de archivos se seleccionará en la localización en que esté y se cargará el archivo. Una vez hecho esto deberemos pasar al cuadro que se encuentra a su derecha para escoger la carpeta en la que queremos que se generen los resultados de la ejecución del programa.



Ilustración 3-6 Vista de operaciones con las autoridades de la aplicación (PROVISIONAL)

Lo que se generará será un archivo tsv que deberemos usar para subir a Mix'n'match los datos de las autoridades de la UPSA y crear el catálogo siguiendo los pasos que explicamos en apartados anteriores. En la siguiente imagen podemos verlo, **Ilustración 3-7 Muestra del formato del archivo tsv para importar los datos a mix'n'match**

Id	name	born	died	P950	P1607	P213	P214	P496	P1960	desc	
46936	Antonio Pintor Ramos				P1607	P213	P214	P496	P1960	desc	14805837
168669	José Manuel Alfonso Sánchez	1964			XX1506436	1228379	0000	0000	5937	2985	46174455
591776	Covadonga Monte Río				XX1149708	535897	0000	0000	6096	4339	86885309
743978	Mark Beyebach	1963			XX1616709		0000	0001	1079	3508	90673403
539959	Francisco García Martínez	1967					0000	0001	0775	1825	121131391
1190764	Lorenzo de Santos Martín	1961					0000	0003	2798	353X	
82208	Luis. Jiménez Díaz				XX957501	1238485	0000	0000	6080	3586	87945626
224243	Xosé Manuel Domínguez Prieto	1962			XX968289	923819	0000	0000	5954	4222	42894848
1199842	José Félix Álvarez Alonso	1984			XX5587581	3839855	5146029369835828278	0000	0001	9380	3010
779429	José Magaña Romera	1955			592411	305375240					
758245	Oscar Prieto Domínguez	1981			2389763	0000	0001	2273	6877	172649336	0000-0002-4143-1734
634198	Manuel Ángel Martínez Juan	1964			XX4581605	2033240	169424708				
628686	Emilio J. Martínez López				XX1589785	4592847	0000	0000	5135	458X	39171633
218925	María Carmen Santos Asensi				XX918733	35811	192111442	0000	0001	6911	6704

Ilustración 3-7 Muestra del formato del archivo tsv para importar los datos a mix'n'match

Las columnas del archivo serán las necesarias para la herramienta:

- Id: Identificador de la UPSA
- name: Nombre completo del autor en el orden corriente (nombre-apellidos)
- born: fecha de nacimiento en formato (yyyy)
- died: fecha de defunción en formato (yyyy)
- P950: Identificador de la Biblioteca nacional Española (BNE)
- P1607: DIALNET de autor
- P213: ISNI
- P214: VIAF
- P496: ORCID

- P1960: Google Académico

Estos datos proceden de la recolocación y readaptación de los datos procedentes de las columnas del archivo original que conservan su nombre y no usan como en este caso el formato de propiedad Wikidata (Pxxx). Esto es así como ya comentamos en apartados anteriores porque la Herramienta Mix'n'match identifica mejor los elementos con este formato.

Pasamos ahora a Mix'n'match para realizar la importación de los datos de las autoridades a un nuevo catálogo como se muestra en las imágenes **Ilustración 3-8 Ventana de importación de Autoridades Mix'n'match parte 1** y **Ilustración 3-9 Ventana de importación de Autoridades Mix'n'match parte 2**.

Para ello dejaremos en blanco el identificador del catálogo, le pondremos un nombre y una descripción al mismo, también se le pondrá una URL de referencia, la cual corresponde a la localización de las autoridades en la biblioteca. Se escribirá la WP que representan los elementos del identificador, en este caso el identificador de la UPSA. También se le dará el tipo de "Publishers" y como lenguaje primario se dejará por defecto el inglés (EN)

En la segunda parte escogeremos el archivo separado por tabulaciones y la opción de fichero local. Seleccionaremos el fichero en cuestión y la opción de añadir nuevas entradas ya que todo es completamente nuevo.

Import or update catalog



Catálogo

Update an existing catalog (give ID), or create a new catalog (leave ID empty).
Note: if you enter the property first and then click another field, some information will be filled in automatically.

Identificador del catálogo	<input type="text"/>	Leave empty to create a new catalog. If given, other data in this section will be ignored.	
Nombre del catálogo	Autoridades de la Universidad Pontificia de Salamanca		
Descripción	Registro de autoridades de la Universidad Pontificia de Salamanca		
URL	https://koha.upsa.es/cgi-bin/koha/opac-authorities-home.pl		optional
WD property	P9247	optional	Type Publishers ↕
Primary language	en	Default entry type	Qxxx optional

Ilustración 3-8 Ventana de importación de Autoridades Mix'n'match parte 1

Data

Note: To use Google Sheets as a source, use menu "File/Publish to the web", select the correct sheet, check "Automatically republish when changes are made", and copy the URL from that dialog (not the main sheet URL!) here.

Data format ☐ Comma-separated values ☒ Tab-separated values

Source ☐ URL

☒ No regular update ☐ Update weekly ☐ Update monthly ☐ Update once a year

☒ File Seleccionar archivo salidaParaMi...454a50e-516e-4d00-9b7d-430c0ca5c4aa

Updating ☒ Add new entries ☐ Update existing descriptions ☐ Update all descriptions

Looks good ☐

Ilustración 3-9 Ventana de importación de Autoridades Mix'n'match parte 2

Cuando finalicemos estos pasos pulsaremos el Check Button de “looks good” y pasaremos a testear si la importación es correcta, en la **Ilustración 3-10 Testeo de columnas de datos** y **Ilustración 3-11 Testeo de datos importados**. Tras verificar que todo es en efecto correcto podemos mandar a la cola de importación. También obtendremos como dato el identificador del catálogo justo después de importar los datos para que podamos acceder a él.

El catálogo está disponible en la dirección <https://mix-n-match.toolforge.org/#/catalog/4580>

#	Column label	Use as
0	id	id
1	name	name
2	born	born
3	died	died
4	P950	P950
5	P1607	P1607
6	P213	P213
7	P214	P214
8	P496	P496
9	P1960	P1960
10	desc	desc

Ilustración 3-10 Testeo de columnas de datos

Columnas de datos

Reload

Test

Test results (≤1000)

Type	Rows
all	421
added	420
updated	0

Looks good, queue import/update

Ilustración 3-11 Testeo de datos importados

A continuación, en la Ilustración 3-12 Visión del catálogo recién creado se muestra la vista del catálogo tras la creación

Autoridades de la Universidad Pontificia de Salamanca

Action

Registro de autoridades de la Universidad Pontificia de Salamanca

Importado por Braisrodriguezcalvo | Actualizar

Entradas

Ligado manualmente	0	0%
Ligado automáticamente	304	72%
No Wikidata	0	0%
Not applicable to Wikidata	0	0%
Unmatched	115	27.4%
Total	419	

Tipos de entrada

419

Ilustración 3-12 Visión del catálogo recién creado

Lo siguiente que debemos buscar será obtener la mayor cantidad posible de “matches” manuales. Para ello empezaremos con la lista de entradas ligadas automáticamente. La forma más cómoda de abordarlos es usando el modo lista (clicando sobre el botón de ligado automáticamente) como se ve en la Ilustración 3-13 Lista de entidades ligadas automáticamente

Autoridades de la Universidad Pontificia de Salamanca

[Action ▼](#)

Registro de autoridades de la Universidad Pontificia de Salamanca

1 2

José Manuel de Mena Ramos	José Manuel de Mena Ramos es una Autoridad de la Universidad Pontificia de Salamanca José Manuel *de *Mena *Ramos *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
José Manuel de Mena Ramos [Q99838791]	España profesor universitario, entrenador deportivo, profesor ♂	Confirmar Borrar [all]
# Marcos Cabezas González	Marcos Cabezas González es una Autoridad de la Universidad Pontificia de Salamanca Marcos *Cabezas González *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
Marcos Cabezas-González [Q58681061]	Investigador ♂	Confirmar Borrar [all]
M. Luisa Díez Platas	M. Luisa Díez Platas es una Autoridad de la Universidad Pontificia de Salamanca M. *Luisa Díez *Platas *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
M. Luisa Díez Platas [Q99197449]	España profesor universitario ♀	Confirmar Borrar [all]
María Paz de Blas Fernández	María Paz de Blas Fernández es una Autoridad de la Universidad Pontificia de Salamanca María Paz *de Blas Fernández *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
María Paz de Blas Fernández [Q100715817]	España profesor universitario ♀	Confirmar Borrar [all]
Rubén González Crespo	Rubén González Crespo es una Autoridad de la Universidad Pontificia de Salamanca Rubén González *Crespo *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
Rubén González Crespo [Q60976901]	Investigador ♂	Confirmar Borrar [all]

Ilustración 3-13 Lista de entidades ligadas automáticamente

Deberemos de ir comprobando cada entrada de forma que nos cerciorem de que coincide con la persona que estamos buscando. Por lo general la correspondencia es correcta ya que como hemos dicho en apartados anteriores, se usan los identificadores de autores para asegurarse de que la autoridad es quien debe de ser, pero además de eso si la descripción perteneciente a nuestro elemento y la del elemento existente muestra a simple vista una relación consistente, no será necesario una mayor comprobación. Podemos ver la relación de las descripciones en el siguiente ejemplo **Ilustración 3-14 Ejemplo de entrada de autoridad**.

M. Luisa Díez Platas	M. Luisa Díez Platas es una Autoridad de la Universidad Pontificia de Salamanca M. *Luisa Díez *Platas *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Ligado automáticamente
M. Luisa Díez Platas [Q99197449]	España profesor universitario ♀	Confirmar Borrar [all]

Ilustración 3-14 Ejemplo de entrada de autoridad

Tendremos que dedicarle una mayor investigación en caso de que la descripción de la persona sea muy genérica por ejemplo que la descripción del elemento sea "Humano" o en el caso de que la descripción no nos hable de un ser humano como el caso de la **Ilustración 3-15 Ejemplo de entidad ligada automáticamente**

<div>Francisco Marín Serrano</div> <div>[Q99447080]</div>	<div>Francisco Marín Serrano es una Autoridad de la Universidad Pontificia de Salamanca</div> <div>Francisco Marín *Serrano *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca</div> <div>40 EASD Annual Meeting of the European Association for the Study of Diabetes : Munich, Germany, 5-9 September 2004</div> <div>[Q56836084]</div>	<div>Ligado automáticamente</div> <div>Confirmar Borrar [all]</div>
------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------

Ilustración 3-15 Ejemplo de entidad ligada automáticamente errónea

Tras convertir todas las entradas ligadas manualmente por el método anterior, obtenemos lo que aparece en la Ilustración 3-16 Resultado después de gestionar los elementos ligados automáticamente

Autoridades Universidad Pontificia de Salamanca			Action ▾
Todas las Autoridades Universidad Pontificia de Salamanca			
Importado por Braisrodriguezcalvo Actualizar			
Entradas			
Ligado manualmente	332	<div><div></div></div>	72%
Ligado automáticamente	0	<div><div></div></div>	0%
No Wikidata	0	<div><div></div></div>	0%
Not applicable to Wikidata	0	<div><div></div></div>	0%
Unmatched	129	<div><div></div></div>	27.9%
Total	461		

Ilustración 3-16 Resultado después de gestionar los elementos ligados automáticamente

Llegados a este punto deberemos de descargar los datos del catálogo para que nuestra aplicación obtenga la lista de autoridades que debemos enriquecer posteriormente mediante OpenRefine.

Para ello usaremos el botón desplegable de “Action” donde accederemos al menú de descarga como se ha explicado en apartados anteriores, marcando las opciones de la forma en la que se muestra en la Ilustración 3-17 Menú de descarga de información catálogo

Descargar entradas

Catálogos	4580
Columnas de datos	<input checked="" type="checkbox"/> URL externa <input checked="" type="checkbox"/> Ligado por nombre de usuario <input type="checkbox"/> Datos auxiliares <input type="checkbox"/> Fecha de nacimiento/fallecimiento <input type="checkbox"/> Latitud/Longitud <input checked="" type="checkbox"/> Multiple matches
Hide matches	<input type="checkbox"/> Hide any matched entries, even automatic ones <input type="checkbox"/> Hide entries matched by users, via auxiliary data, or name&dates <input type="checkbox"/> Hide entries matched by users <input type="checkbox"/> Hide unmatched entries <input type="checkbox"/> Hide automatically (transiently, "automatched") matched entries <input type="checkbox"/> Hide entries automatically matched via name&dates <input type="checkbox"/> Hide entries matched via auxiliary data (e.g. via shared VIAF ID) <input type="checkbox"/> Hide entries without multi-matches
Formato	<input type="radio"/> Texto tabulado <input checked="" type="radio"/> JSON <input type="checkbox"/> Descargar como archivo (del otro modo, se abre en una nueva pestaña)
<div>Descargar</div>	

Ilustración 3-17 Menú de descarga de información catálogo

Una vez descargado el archivo json correspondiente se deberá de usar de nuevo la aplicación en el cuadro de la parte inferior a la izquierda usando el explorador de archivos acceder al elemento que acabamos de descargar. Una vez hecho esto nos vamos al cuadro a su derecha y escogemos la carpeta donde queremos que la aplicación nos dé el resultado. El resultado consistirá en un archivo tsv que contiene las autoridades que ya existen en Wikidata con sus datos correspondientes para hacer el enriquecimiento posterior en OpenRefine que explicaremos más adelante. La interfaz de usuario se puede ver claramente en la **Ilustración 3-18 Vista de operaciones con las autoridades de la aplicación(PROVISIONAL)**.

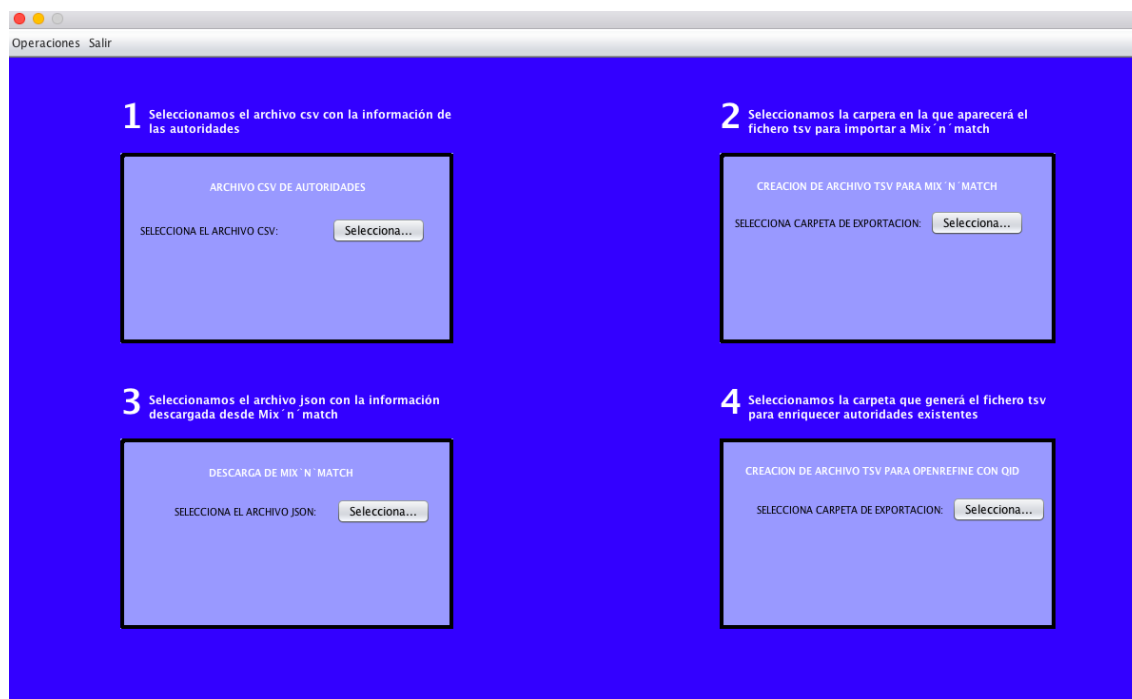


Ilustración 3-18 Vista de operaciones con las autoridades de la aplicación (PROVISIONAL)

La salida del programa nos da un fichero tsv con el formato que se ve a continuación, se diferencia del anterior en que le hemos añadido el Qnumber correspondiente al elemento en cuestión que se ha identificado en Mix'n'match como se puede ver en la **Ilustración 3-19 Muestra de los datos del archivo de enriquecimiento para OpenRefine**

id	name	QNumber	born	died	P958	P1607	P213	P214	P496	P1960	desc	0000	0001	1598	3271	14805837	
46936	Antonio Pintor Ramos	090645962															Antonio Pintor Ramos es una Autoridad de la Universidad Pontificia de Salamanca
168669	José Manuel Alfonso Sánchez	080995018	1964														José Manuel Alfonso Sánchez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
591776	Covadonga Monte Río	099156608															Covadonga Monte Río es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1963
743978	Mark Beyebach	089543851	1963														Mark Beyebach es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1963
1198764	Lorenzo de Santos Martín	099343854	1961														Lorenzo de Santos Martín es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1961
82280	Luis. Jiménez Díaz	099528768															Luis. Jiménez Díaz es una Autoridad de la Universidad Pontificia de Salamanca
224243	Xosé Manuel Domínguez Prieto	012402829	1962														Xosé Manuel Domínguez Prieto es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1962
1199842	José Félix Álvarez Alonso	099804232	1984														José Félix Álvarez Alonso es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1984
779429	José Magaña Romero	0107313528	1955														José Magaña Romero es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1955
758245	Oscar Prieto Domínguez	098828969	1981														Oscar Prieto Domínguez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1981
634198	Manuel Ángel Martínez Juan	0107313854	1964														Manuel Ángel Martínez Juan es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
628686	Emilio J. Martínez López	086931381															Emilio J. Martínez López es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
218925	María Carmen Santos Asensi	099343226															María Carmen Santos Asensi es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
1205792	Sergio López García	087109291															Sergio López García es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
1205631	Raúl Ramírez Rodríguez	098831607															Raúl Ramírez Rodríguez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
286790	David Alameda García	090805155															David Alameda García es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964

Ilustración 3-19 Muestra de los datos del archivo de enriquecimiento para OpenRefine

Para terminar con la parte de Mix'n'match, vamos a crear las Autoridades que no se han encontrado y que se encuentran en la lista de *“unmatched”*. Para ello abriremos dicha lista y empezaremos a crear elementos uno a uno pulsando en *nuevo elemento* como se ve en la **Ilustración 3-20 Proceso de creación de nuevos elementos**

Juan Antonio González Iglesias [Q4893838]	España poeta, escritor, crítico literario (*1964) ♂	Borrar
Manuel José Moro Rodríguez	Manuel José Moro Rodríguez es una Autoridad de la Universidad Pontificia de Salamanca Manuel José *Moro Rodríguez *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	By Braisrodriguezcalvo
Manuel José Moro Rodríguez [Q107345676]	No se puede auto-describir	Borrar
Jesús Rojano Martínez	Jesús Rojano Martínez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1962 Jesús *Rojano Martínez *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca *con *fecha *de *nacimiento *en *el 1962	By Braisrodriguezcalvo
Jesús Rojano Martínez [Q107345677]	No se puede auto-describir	Borrar
Noa María Carballa Rivas	Noa María Carballa Rivas es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1987 Noa María *Carballa *Rivas *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca *con *fecha *de *nacimiento *en *el 1987	By Braisrodriguezcalvo
Noa María Carballa Rivas [Q107345678]	No se puede auto-describir	Borrar
Alfonso José López Rivero	Alfonso José López Rivero es una Autoridad de la Universidad Pontificia de Salamanca Alfonso José López *Rivero *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Not matched
Search Wikidata Search en.wikipedia Google-search Wikipedias Google-search Wikidata		Fijar Q Nuevo elemento N/D
Carmen Yebra Rovira	Carmen Yebra Rovira es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1972 Carmen *Yebra *Rovira *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca *con *fecha *de *nacimiento *en *el 1972	Not matched
Search Wikidata Search en.wikipedia Google-search Wikipedias Google-search Wikidata		Fijar Q Nuevo elemento N/D
Francisco Javier Martín Jiménez	Francisco Javier Martín Jiménez es una Autoridad de la Universidad Pontificia de Salamanca Francisco Javier Martín Jiménez *es *una *Autoridad *de *la *Universidad *Pontificia *de Salamanca	Not matched

Ilustración 3-20 Proceso de creación de nuevos elementos

Cuando creamos un nuevo elemento automáticamente se nos genera un Qnumber para ese nuevo elemento, esto se puede ver en el cuadro rojo de la fotografía anterior. Si se clica en el enlace del identificador nos lleva al elemento dentro de Wikidata que se ha creado.

Noa María Carballa Rivas (Q107345678)

No description defined
edit

In more languages
Configure

Language	Label	Description	Also known as
English	Noa María Carballa Rivas	No description defined	
Spanish	No label defined	No description defined	
Catalan	No label defined	No description defined	
Galician	No label defined	No description defined	

Statements
+ add statement

Identifiers

VIAF ID	315089407 edit <div> 0 references + add reference </div> + add value
Biblioteca Nacional de España	XX5470393 edit

Ilustración 3-21 Muestra del dato creado en Wikidata

Como se puede apreciar al crear un nuevo elemento con esta herramienta no se genera una descripción para el nuevo elemento, aún a pesar de introducir los datos de ese modo, pero sí se puede asignar una posteriormente con OpenRefine.

Finalmente, el catálogo de autoridades debe de quedar de esta forma **Ilustración 3-22 Estado final del catálogo**



Ilustración 3-22 Estado final del catálogo

Para finalizar con esta fase del proyecto nos trasladaremos a la otra herramienta, OpenRefine. En ella enriqueceremos las autoridades que ya existían como comentamos anteriormente. Crearemos un nuevo proyecto en OpenRefine con el archivo que acabamos de crear con la aplicación y reconciliaremos el nombre del autor de la siguiente forma **Ilustración 3-23 Formato de cotejamiento**

Cotejar columna "nombre"

Cotejar cada celda con los valores de una de estas clases:

- ☒ humano (Q5)
- ☐ artículo académico (Q13442814)

Usar también detalles relevantes de otras columnas:

columna	¿Incluir?	Como propiedad
identificación	<input checked="" type="checkbox"/>	Pontifical University of Salama
qNumber	<input type="checkbox"/>	
Nació	<input checked="" type="checkbox"/>	date of birth
fallecido	<input checked="" type="checkbox"/>	date of death
P950	<input checked="" type="checkbox"/>	Biblioteca Nacional de España
P1607	<input checked="" type="checkbox"/>	Dialnet author ID
P213	<input checked="" type="checkbox"/>	ISNI
P214	<input checked="" type="checkbox"/>	VIAF ID
P496	<input checked="" type="checkbox"/>	ORCID iD
P1960	<input checked="" type="checkbox"/>	Google Scholar author ID
desc	<input type="checkbox"/>	

☐ Cotejar contra la clase:

☐ Reconciliar contra ninguna clase en particular

☒ Cotejar automáticamente candidatos con alta confianza

Máximo número de candidatos a devolver

Agregar servicio estándar ...

Ilustración 3-23 Formato de cotejamiento

Este es el resultado que obtendremos después de conciliar **Ilustración 3-24 Resultado tras reconciliar Autoridades a enriquecer**. Como se puede ver el resultado es idéntico al de OpenRefine Salvo por 4 elementos no conciliados que podemos solucionar bajando la puntuación de conciliación (los elementos que se ven en Ilustración 2-12 Ejemplo datos para ser importados a los catálogos)

328 matching filas (332 total)												Extensiones: Wikidata
Mostrar como: filas registros												« primera » anterior 1 - 50 siguiente » última »
▼ Todo	▼ Id	▼ name	▼ qNumber	▼ born	▼ died	▼ P950	▼ P1607	▼ P213	▼ P214	▼ P496	▼ P1960	▼ desc
★	1.	46936	Antonio Pintor-Ramos Escojer nueva coincidencia	Q98645962		XX1035431	0000 0001 1598 3271	14805837				Antonio Pintor Ramos es una Autoridad de la Universidad Pontificia de Salamanca
★	2.	168669	José Manuel Alfonso Sánchez Escojer nueva coincidencia	Q9895910	1964	XX1506436	1228379	0000 0000 5937 2985	46174455	0000- 0002- 2674- 1021	3W7AmzoAAAAJ	José Manuel Alfonso Sánchez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964
★	3.	591776	Covadonga Monte Río Escojer nueva coincidencia	Q99156600								Covadonga Monte Río es una Autoridad de la Universidad Pontificia de Salamanca
★	4.	743978	Mark Beyebach Escojer nueva coincidencia	Q89543851	1963	XX1149708	535897	0000 0000 6096 4339	86885309	0000- 0002- 4490- 5915		Mark Beyebach es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1963
★	5.	1190764	Lorenzo de Santos Martín Escojer nueva coincidencia	Q99343854	1961			0000 0001 0775 1825	121131391	0000- 0003- 2798- 353X		Lorenzo de Santos Martín es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1961
★	6.	82200	Luis Jiménez Díaz Escojer nueva coincidencia	Q99528760		XX957501	1238485	0000 0000 6080 3586	87945626			Luis. Jiménez Díaz es una Autoridad de la Universidad Pontificia de Salamanca
★	7.	224243	Xosé Manuel Domínguez Prieto Escojer nueva coincidencia	Q12402829	1962	XX960289	923019	0000 0000 5954 4222	42094048			Xosé Manuel Domínguez Prieto es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1962
★	8.	1199842	José Félix Álvarez Alonso Escojer nueva coincidencia	Q99004232	1984	XX5587581	3839855		5146029369835820278	0000- 0001- 9380- 3010	bm8vQXxAAAAJ	José Félix Álvarez Alonso es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1984
★	9.	779429	José Magaña Romero Escojer nueva coincidencia	Q107313528	1955		592411		305375240			José Magaña Romero es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1955
★	10.	758245	Óscar Prieto Domínguez Escojer nueva coincidencia	Q98828969	1981		2309763	0000 0001 2273	172649336	0000- 0002- 4143-	CWqg-ZN8AAAAJ	Óscar Prieto Domínguez es una Autoridad de la Universidad Pontificia de Salamanca con fecha

Ilustración 3-24 Resultado tras reconciliar Autoridades a enriquecer

4 matching filas (332 total)												Extensiones: Wikidata
Mostrar como: filas registros												« primera » anterior 1 - 4 siguiente » última »
▼ Todo	▼ Id	▼ name	▼ qNumber	▼ born	▼ died	▼ P950	▼ P1607	▼ P213	▼ P214	▼ P496	▼ P1960	▼ desc
★	13.	218925	Maria Carmen Santos Asensi ✓ Oscar Sanjuán Martínez (100) ✓ Maria Carmen Santos Asensi (100) ✓ Crear nuevo elemento Buscar una coincidencia	Q99343226		XX918733	35811		192111442	0000- 0001- 6911- 6704		Maria Carmen Santos Asensi es una Autoridad de la Universidad Pontificia de Salamanca
★	71.	1188876	Alberto Royo Mejía ✓ Alberto Valverde Pérez (100) ✓ Alberto Pérez (100) ✓ Alberto Royo Mejía (100) ✓ Crear nuevo elemento Buscar una coincidencia	Q98954533		XX5389594	22627					Alberto Royo Mejía es una Autoridad de la Universidad Pontificia de Salamanca
★	168.	110834	José L. Segovia Bernabé ✓ Belén Sanz Barbero (100) ✓ José L. Segovia Bernabé (100) ✓ Crear nuevo elemento Buscar una coincidencia	Q93427456	1960	XX1159950	1885Q6dri	0000 0000 5927 7152	338063	0000- 0001- 6030- 061X		José L. Segovia Bernabé es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1960
★	216.	183994	Manuel Rodríguez Becerra ✓ Manuel Rodríguez Becerra (100) ✓ Manuel Rodríguez Becerra (100) ✓ Crear nuevo elemento Buscar una coincidencia	Q5994227					117117618	0000- 0002- 8763- 4369	e2kCoVMAAAAJ	Manuel Rodríguez Becerra es una Autoridad de la Universidad Pontificia de Salamanca

Ilustración 3-25 Elementos no conciliados

Una vez reconciliados los elementos en su totalidad, debemos crear el esquema correspondiente para poder generar los comandos que después de insertaran en

QuickStatements. El formato que debe de tener el schema es el siguiente **Ilustración 3-26 Schema de enriquecimiento de autoridades**

idnameqNumberborndiedP950P1607P213P214P496P1960desc

name

remove

Terms

no labels, descriptions or aliases added

+ add term

Statements

identificador Univers

id

remove

+ add qualifier

▼ 1 references

copy

remove

afirmado enCatalogue of the Lii

remove

identificador Uid

remove

fecha de acce:2021-06-28

remove

+ add

+ add reference

+ add value

fecha de nacimiento

born

remove

+ add qualifier

▼ 1 references

copy

remove

afirmado enCatalogue of the Lii

remove

identificador Uid

remove

fecha de acce:2021-06-28

remove

+ add

+ add reference

+ add value

fecha de fallecimiento

died

remove

+ add qualifier

▼ 1 references

copy

remove

afirmado enCatalogue of the Lii

remove

identificador Uid

remove

fecha de acce:2021-06-28

remove

+ add

+ add reference

+ add value

80

identificador BNE	P950	remove
		+ add qualifier
▼ 1 references		
copy		remove
afirmado en	Catalogue of the Lli	remove
identificador U	id	remove
fecha de acces	2021-06-28	remove
		+ add
		+ add reference
		+ add value

identificador Dialnet	P1607	remove
		+ add qualifier
▼ 1 references		
copy		remove
afirmado en	Catalogue of the Lli	remove
identificador U	id	remove
fecha de acces	2021-06-28	remove
		+ add
		+ add reference
		+ add value

ISNI	P213	remove
		+ add qualifier
▼ 1 references		
copy		remove
afirmado en	Catalogue of the Lli	remove
identificador U	id	remove
fecha de acces	2021-06-28	remove
		+ add
		+ add reference
		+ add value

identificador VIAF	P214	remove
		+ add qualifier
▼ 1 references		
copy		remove
afirmado en	Catalogue of the Lli	remove
identificador U	id	remove
fecha de acces	2021-06-28	remove
		+ add
		+ add reference
		+ add value

ORCID

P496

remove

+ add qualifier

▼ 1 references

copy

remove

afirmado en

Catalogue of the Lii

remove

identificador U

id

remove

fecha de acce:

2021-06-28

remove

+ add

+ add reference

+ add value

identificador Google

P1960

remove

+ add qualifier

▼ 1 references

copy

remove

afirmado en

Catalogue of the Lii

remove

identificador U

id

remove

fecha de acce:

2021-06-28

remove

+ add

+ add reference

+ add value

Ilustración 3-26 Schema de enriquecimiento de autoridades

Para finalizar el proceso exportaremos al fichero txt de comandos QuickStatements que después pegaremos en el cuadro de importación de la aplicación QuickStatements y ejecutaremos obteniendo el siguiente resultado **Ilustración 3-27 Resultado de ejecutar comandos en Quickstatements.**

Lote #57684 "test enriquecer autoridades de la UPSA" on Wikidata por Braisrodriguezcalvo [Lotes] [Discutir/revertir lote](#)

Estado: DONE

99.4% (167) de 167 hecho

1	done	Antonio Pintor-Ramos (Q98645962)	ADD	Description	es:Antonio Pintor Ramos es una Autoridad de la Universidad Pontificia de Salamanca
2	done	Antonio Pintor-Ramos (Q98645962)	ADD	Statement	Pontifical University of Salamanca ID (P9247): "46936"
3	done	Antonio Pintor-Ramos (Q98645962)	ADD	Sources to	Pontifical University of Salamanca ID (P9247): "46936" stated in (P248):Catalogue of the Library of the Pontifical University of Salamanca (Q105667100) Pontifical University of Salamanca ID (P9247): "46936" retrieved (P813):2021-06-27
4	done	Antonio Pintor-Ramos (Q98645962)	ADD	Statement	Biblioteca Nacional de España ID (P980): "XX1035431"
5	done	Antonio Pintor-Ramos (Q98645962)	ADD	Sources to	Biblioteca Nacional de España ID (P980): "XX1035431" stated in (P248):Catalogue of the Library of the Pontifical University of Salamanca (Q105667100) Pontifical University of Salamanca ID (P9247): "46936" retrieved (P813):2021-06-27
6	done	Antonio Pintor-Ramos (Q98645962)	ADD	Statement	ISNI (P213): "0000 0001 1598 3271"
7	done	Antonio Pintor-Ramos (Q98645962)	ADD	Sources to	ISNI (P213): "0000 0001 1598 3271" stated in (P248):Catalogue of the Library of the Pontifical University of Salamanca (Q105667100) Pontifical University of Salamanca ID (P9247): "46936" retrieved (P813):2021-06-27
8	done	Antonio Pintor-Ramos (Q98645962)	ADD	Statement	VIAF ID (P214): "14805837"
9	done	Antonio Pintor-Ramos (Q98645962)	ADD	Sources to	VIAF ID (P214): "14805837" stated in (P248):Catalogue of the Library of the Pontifical University of Salamanca (Q105667100) Pontifical University of Salamanca ID (P9247): "46936" retrieved (P813):2021-06-27
10	done	José Manuel Alfonso Sánchez (Q98995910)	ADD	Description	es:José Manuel Alfonso Sánchez es una Autoridad de la Universidad Pontificia de Salamanca con fecha de nacimiento en el 1964

Primero **Página 1** Siguiente Último

All ☐ errores ☐ Init 10

Ilustración 3-27 Resultado de ejecutar comandos en Quickstatements

3.3.3 FASE 3: FASE DE CREACIÓN DE LOS ARTICULOS PUBLICADOS POR LAS AUTORIDADES DE LA UPSA

En esta fase crearemos las entradas de los artículos de la revista Cuadernos Salmantinos que hayan sido elaborados por las autoridades de la UPSA con las cuales hemos trabajado en el apartado anterior.

Para ello se cargará los datos de las Autoridades que tenemos de la misma forma que en el apartado anterior. La interfaz de usuario está presente en la **Ilustración 3-28 Ventana operaciones con publicaciones de la (PROVISIONAL)**. Debemos seleccionar el fichero json de publicaciones de la revista con el cuadro superior izquierdo y el fichero de autoridades de autoridades csv en el cuadro superior derecho. Estos dos pasos cargaran los datos necesarios para los cálculos posteriores.

El formato del fichero de autoridades tsv que se genera tiene el mismo formato que el del apartado anterior, de hecho, es la misma funcionalidad. La razón de que no se guarden los datos de la anterior operación en una variable para su uso posterior en esta ventana, es la posibilidad de que en un futuro se pudiera hacer lo mismo para distintas revistas y por lo tanto no sería necesario añadir los datos de las autoridades.

El fichero tiene el formato que se muestra en la Ilustración 3-1 Estructura archivo json de publicaciones de apartados anteriores.

El siguiente paso es calcular que artículos corresponden a dichas autoridades y generar un fichero tsv en el que estén presentes los datos necesarios para identificar a cada una de las autoridades y al mismo tiempo permitan crear los nuevos artículos con OpenRefine.

Para ello usaremos el cuadro inferior de mayor tamaño en el que seleccionaremos la carpeta en la que queremos exportar el archivo creado y para que se termine de generar pulsaremos el botón *“Calcular resultado”*. De esta forma ya tendremos el archivo para su uso.

Algo importante para la correcta implementación del proceso es tener en cuenta que para una publicación puede haber más de un autor. Para este caso en concreto lo que la aplicación hace es generar otra segunda lista con estos autores. No podemos ponerlos todos en la misma lista debido a que nos aparecería un error en OpenRefine diciendo que no se pueden crear elementos idénticos, con los mismos datos. Por esa razón es necesario hacerlo en un segundo paso



Ilustración 3-28 Ventana operaciones con publicaciones de la aplicación (PROVISIONAL)

Los datos correspondientes a las columnas de este archivo son:

- `articleName`: nombre del artículo en cuestión
- `pageNumber`: número de páginas del contenido del artículo
- `pages`: indica la página de inicio y fin del artículo en el volumen
- `volumen`: volumen de la revista
- `fechaPublicacion`: fecha de publicación del artículo en formato (yyyy)
- `doi`: identificador propio del artículo
- `separador`: un guion que separa la info. Del autor de la del artículo
- Resto de información del autor que ya se ha explicado en apartados anteriores.

Se puede apreciar una muestra del formato usado en la **Ilustración 3-29 Formato de salida del archivo tsv de autoridades y artículos**

articleName	pageNumber	pages	volumen	fechaPublicacion	doi	separator	autorName	born	died	num_reg_upsa	BNE	DIALNET	ISNI	VIAF	ORCID	Google_academico
El espíritu en c.G. Jung	11	21-31	15	1988	10.36576/summa.854	-	Antonio Vázquez Fernández	1926	2020	76991	XX1868098	1370280				
Ser y conocer: Peculiaridades informáticas de la especie humana	15	5-19	15	1988	10.36576/summa.853	-	Marcellano Arranz Rodrigo	1943							359216	348948

Ilustración 3-29 Formato de salida del archivo tsv de autoridades y artículos

Usaremos en este caso OpenRefine porque como explicamos en apartados anteriores la herramienta de Mix'n'match no es útil para la generación de una gran cantidad de elementos nuevos, ya que tendríamos que hacerlo de forma manual, de uno en uno.

Creamos un nuevo proyecto en OpenRefine y reconciliamos la columna del nombre del artículo contra la clase “*articulo científico*” o “*sholarly article*”. Para obtener mejores resultados utilizamos las columnas del número de páginas, la fecha de publicación, el DOI del artículo y el nombre de su autor. Esto lo podemos ver en la **Ilustración 3-30 Cotejado de las publicaciones**

Cotejar columna "articleName"

» Ir a [API del servicio](#)

Cotejar cada celda con los valores de una de estas clases:

☐ Q35120

Usar también detalles relevantes de otras columnas:

columna	¿Incluir?	Como propiedad
pageNumber	<input checked="" type="checkbox"/>	number of pages
pages	<input type="checkbox"/>	
volumen	<input type="checkbox"/>	
fechaPublicacion	<input checked="" type="checkbox"/>	publication date
doi	<input checked="" type="checkbox"/>	DOI
separator	<input type="checkbox"/>	
autorName	<input checked="" type="checkbox"/>	author
born	<input type="checkbox"/>	
died	<input type="checkbox"/>	
num_reg_upsa	<input type="checkbox"/>	
BNE	<input type="checkbox"/>	

☒ Cotejar contra la clase:

☐ Reconciliar contra ninguna clase en particular

☒ Cotejar automáticamente candidatos con alta confianza

Máximo número de candidatos a devolver

Ilustración 3-30 Cotejado de las publicaciones

El resultado obtenido tiene que ser de 0% de cotejamiento ya que no se ha registrado ninguno de estos artículos hasta el momento. Por lo tanto, deberemos indicarle que lo que queremos hacer con todos los elementos es crear nuevos ítems. Para ello lo que debemos hacer es usar la acción sobre los datos cotejados que nos permite marcar todas las filas como elementos de nueva creación y nos dará como resultado el que podemos ver en la imagen **Ilustración 3-31 Entradas seleccionadas para su creación**.

5 filas Extensiones: W

Mostrar como: **filas registros** Mostrar: **5 10 25 50** filas « primera < anterior **1** - **5** siguiente »

	Todo	articleName	pageNumber	pages	volumen	fechaPublicacio	doi	separator	autorName	born	died	r
☆	1.	Explicación científica y biología: insuficiencias de la metodología hartmanniana ante el problema de la vida <small>NEW</small> Esoger nueva coincidencia	18	93-111	6	1979-01-01	10.36576/summa.618	-	Marceliano Arranz Rodrigo	1943		3592
☆	2.	El dualismo antropológico en Rousseau <small>NEW</small> Esoger nueva coincidencia	41	113-154	34	2007-01-01	10.36576/summa.29047	-	Antonio Pintor Ramos			4693
☆	3.	Juicio teleológico y sistema de razón en Kant <small>NEW</small> Esoger nueva coincidencia	57	163-220	34	2007-01-01	10.36576/summa.29049	-	Ana María Andaluz Romanillos	1957		8248
☆	4.	Sobre el estatuto epistemológico de la sociología de la comunicación humana <small>NEW</small> Esoger nueva coincidencia	45	201-246	14	1987-01-01	10.36576/summa.840	-	Octavio Uña Juárez	1945		1123
☆	5.	Una antropología cosmológica y psicósomática en el siglo XVI: Nuevo intento de comprensión de la obra del bachiller M. Sabuco y Álvarez <small>NEW</small> Esoger nueva coincidencia	19	407-426	5	1978-01-01	10.36576/summa.601	-	Francisco Rodríguez Pascual	1927	2007	2631

Ilustración 3-31 Entradas seleccionadas para su creación

Será necesario que se reconcilien también los datos del autor (que estarán entre las autoridades que creamos antes) de la misma forma que lo hicimos en apartados anteriores usando los datos de distintas columnas como el VIAF, ORCID de autor y demás informaciones para garantizar que todos son vinculados con sus elementos Wikidata.

Tras hacer este paso deberemos de pasar a la creación del schema necesario para la creación de los nuevos artículos.

El esquema que se debe de hacer es el que aparece en la **¡Error! No se encuentra el origen de la referencia.**

Este esquema es el más extenso de los que hemos usado en el proyecto debido en parte a que son elementos nuevos.

Comenzaremos con la parte de los términos del esquema. En ella por el hecho de ser nuevos elementos debemos darles un *“label”* de forma obligatoria, este se corresponde con el título del artículo en cuestión.

Como el elemento *“label”* es un texto monolingüe (tiene que especificarse el idioma) debemos decirle que el idioma es español (es) porque el artículo está originalmente en español. También es muy recomendable hacer su versión en inglés (en) aunque como el título original está en español no deberá de traducirse.

Las descripciones son opcionales, pero también es recomendable ponerlas, éstas deben ser muy generales. Siguiendo las recomendaciones de los expertos en Wikidata con los que he contactado, quedando con el texto que aparece en la imagen.

En cuanto a la parte de los statements debemos empezar hablando de la propiedad Wikidata *“instancia de”* (P31) que nos permite decir el tipo al que pertenece el artículo, éste será *“artículo científico”* (Q13442814).

Le sigue la propiedad de *“autor”* (P1476) que se representa por los autores que hemos reconciliado (por eso es necesario que estén reconciliados)

La siguiente propiedad será la de *“titulo”* (P1476) que se representa por el título del artículo.

La *“idioma de la obra”* (p407) que será el español (Q1321)

“Número de páginas” (P1104) que vendrá representado por *“pages”* al que necesariamente habrá que añadirle el valor de las unidades que en este caso serán *“pages”* (Q1069725).

“publicado en” (P1433) que vendrá representado por Revista Cuadernos Salmantinos(Q65929181)

El *“tomo”* de la publicación (P478) que se representará por el número del volumen

La *“fecha de publicación”* (P577) con el formato (yyyy)

Por último, el *“DOI”* (P356) que vendrá representado por el identificador doi.

Todo lo dicho se puede ver de forma visual en la **Ilustración 3-32 Schema de creación de artículos**

articleName	pageNumber	pages	volumen	fechaPublicacion	doi	separator	autorName	born	died	num_reg_upsa	BNE	DIALNET	ISNI	VIAF	ORCID	Google_academico	Column
<div> <div>articleName</div> <div>remove</div> </div>																	
<div> <div>Terms</div> <div> <div> <div>Description</div> <div>es</div> <div>artículo de la revista Cuadernos Sal</div> <div>override if present</div> <div>remove</div> </div> <div> <div>Label</div> <div>es</div> <div>articleName</div> <div>override if present</div> <div>remove</div> </div> <div> <div>Description</div> <div>en</div> <div>article of the journal Cuadernos Sal</div> <div>override if present</div> <div>remove</div> </div> <div> <div>Label</div> <div>en</div> <div>articleName</div> <div>override if present</div> <div>remove</div> </div> </div> <div>+ add term</div> </div>																	
<div> <div>Statements</div> <div> <div> <div>instancia de</div> <div>artículo científico</div> <div>remove</div> </div> <div>+ add qualifier</div> <div> <div>▼ 1 references</div> <div> <div>copy</div> <div>remove</div> </div> <div> <div>fecha de acceso</div> <div>2021-07-06</div> <div>remove</div> </div> <div> <div>afirmado en</div> <div>Cuadernos Salman</div> <div>remove</div> </div> <div>+ add</div> <div>+ add reference</div> <div>+ add value</div> </div> </div> </div>																	
<div> <div>autor</div> <div> <div> <div>autorNa...</div> <div>remove</div> </div> <div>+ add qualifier</div> <div> <div>▼ 1 references</div> <div> <div>copy</div> <div>remove</div> </div> <div> <div>fecha de acceso</div> <div>2021-07-06</div> <div>remove</div> </div> <div> <div>afirmado en</div> <div>Cuadernos Salman</div> <div>remove</div> </div> <div>+ add</div> <div>+ add reference</div> <div>+ add value</div> </div> </div> </div>																	
<div> <div>título</div> <div> <div> <div>es</div> <div>articleName</div> <div>remove</div> </div> <div>+ add qualifier</div> <div> <div>▼ 1 references</div> <div> <div>copy</div> <div>remove</div> </div> <div> <div>fecha de acceso</div> <div>2021-07-06</div> <div>remove</div> </div> <div> <div>afirmado en</div> <div>Cuadernos Salman</div> <div>remove</div> </div> <div>+ add</div> <div>+ add reference</div> <div>+ add value</div> </div> </div> </div>																	
<div> <div>idioma de la obra o c</div> <div> <div> <div>español</div> <div>remove</div> </div> <div>+ add qualifier</div> <div> <div>▼ 1 references</div> <div> <div>copy</div> <div>remove</div> </div> <div> <div>fecha de acceso</div> <div>2021-07-06</div> <div>remove</div> </div> <div> <div>afirmado en</div> <div>Cuadernos Salman</div> <div>remove</div> </div> <div>+ add</div> <div>+ add reference</div> <div>+ add value</div> </div> </div> </div>																	
<div> <div>número de páginas</div> <div> <div> <div>pageNumber</div> <div>página</div> <div>remove</div> </div> <div>+ add qualifier</div> <div> <div>▼ 1 references</div> <div> <div>copy</div> <div>remove</div> </div> <div> <div>fecha de acceso</div> <div>2021-07-06</div> <div>remove</div> </div> <div> <div>afirmado en</div> <div>Cuadernos Salman</div> <div>remove</div> </div> <div>+ add</div> <div>+ add reference</div> <div>+ add value</div> </div> </div> </div>																	

The screenshot displays the Wikidata article creation interface for four properties. Each property has a value field and a list of references.

- publicado en**: Value is 'Cuadernos Salmantinos de Filosofía'. It has one reference with 'fecha de acceso' (2021-07-06) and 'afirmado en' (Cuadernos Salma).
- tomo**: Value is 'volumen'. It has one reference with 'fecha de acceso' (2021-07-06) and 'afirmado en' (Cuadernos Salma).
- fecha de publicación**: Value is 'fechaPublicac...'. It has one reference with 'fecha de acceso' (2021-07-06) and 'afirmado en' (Cuadernos Salma).
- DOI**: Value is 'doi'. It has one reference with 'fecha de acceso' (2021-07-06) and 'afirmado en' (Cuadernos Salma).

Each reference row includes a 'copy' icon, a 'remove' icon, and buttons for '+ add', '+ add reference', and '+ add value'.

Ilustración 3-32 Schema de creación de artículos

Para finalizar la tercera fase deberemos extraer los comandos Quickstatements para importarlos mediante la herramienta con el mismo nombre como se ve en la **Ilustración 3-33 importación de los artículos mediante Quickstatements**

Como se comentó antes en el caso de los artículos que tienen varios autores, será necesario abrir un nuevo proyecto con la lista de autores y artículos correspondientes donde a diferencia de lo que acabamos de hacer, sólo es un enriquecimiento de datos del autor del artículo.

En nuestro proyecto solo se da un caso en el que en un artículo exista más de un autor y que ese autor coincida con una autoridad de la Universidad Pontificia de Salamanca.

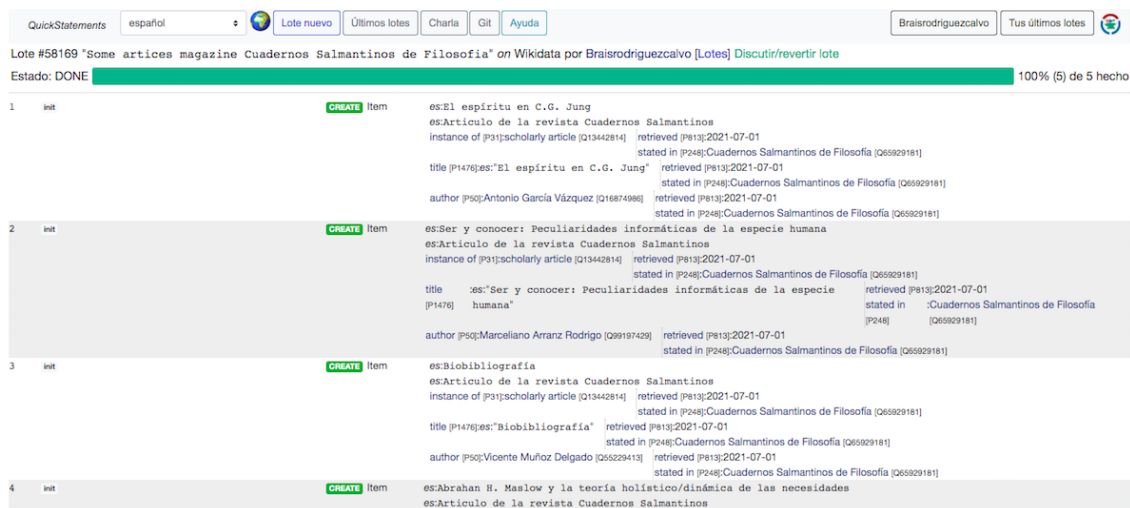


Ilustración 3-33 importación de los artículos mediante Quickstatements

3.4 DESCRIPCIÓN DE LA APLICACIÓN JAVA

Referente a la aplicación java, como se ha comentado en apartados anteriores se realizará en un entorno de desarrollo de NetBeans y para la elaboración de la interfaz se usará JavaSwing

En la **Ilustración 3-34 Esquema de las partes de la aplicación**, podemos ver COMO las funcionalidades principales de la aplicación están ordenadas por paquetes. Estos paquetes en su interior tienen clases y interfaces.

Es una aplicación para la biblioteca de la Universidad Pontificia de Salamanca con el objetivo de facilitar la transformación automatizada de los datos que poseen, al formato adecuado para su utilización en las herramientas que se han nombrado en este proyecto.

Goza de una interfaz simple y transparente que facilita el procesamiento por cualquier persona con tan sólo una lectura rápida al manual y guía práctica que se expone más adelante en el documento.

Su valor radica sobre todo en que su uso es aplicable no sólo a esta revista en concreto sino también a cualquier otra revista y autoridades, siempre que cumplan el formato de datos de entrada. Con formato de datos de entrada me refiero a que por ejemplo los ficheros de los datos de autoridades deberán de tener formato csv y que sus nombres de columnas coincidan con los de los archivos usados en este proyecto, algo que no será un problema dado que los archivos se extraen directamente del sistema en ese formato.

Lo mismo ocurre con el fichero json de publicaciones, al descargarse directamente del sistema siempre vendrá con el formato correcto sea cual sea la revista

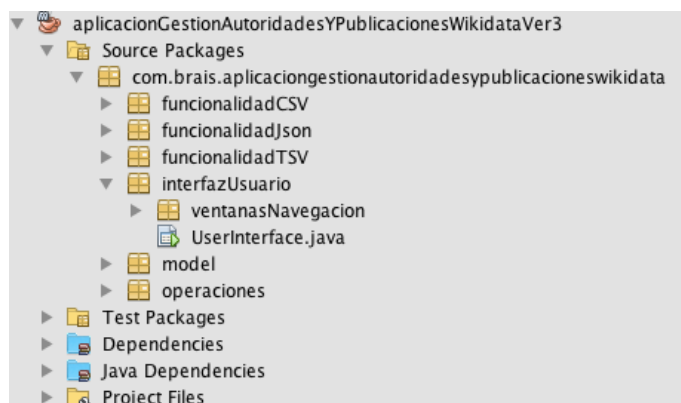


Ilustración 3-34 Esquema de las partes de la aplicación

A continuación, se irán describiendo las funcionalidades, implementadas en cada paquete de una forma detallada.

El paquete “funcionalidadCSV” incluye una interfaz con un único método que se desarrolla en la implementación y que permite la lectura de los datos del csv usando jerárquicamente las tres clases que se encuentran en el subpaquete de “Reader”. Estas clases con el “AutoridadCsvReader” que es la clase más general donde se produce la lectura de los datos, la siguiente “AutoridadCsvParser”, que se encarga de la lectura línea a línea de los datos y por último “AutoridadBuilder”, que nos permite que nos permite crear las instancias de la clase “Autoridad”.

Con este apartado conseguimos leer las autoridades del csv original y obtener un diccionario de autoridades con un identificador que es el nombre de la autoridad en cuestión, para que en procesos posteriores resulte mucho más fácil identificar la autoridad.

Todo lo que se ha dicho antes sobre este paquete se puede ver esquematizado en la **Ilustración 3-35 Paquete funcionalidad csv**

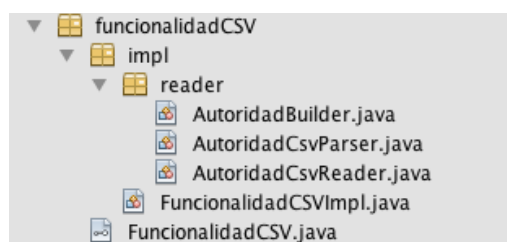


Ilustración 3-35 Paquete funcionalidad csv

El siguiente paquete es el de “funcionalidadJson” cuenta también en su interior con una interfaz y una implementación por medio de las cuales podemos leer los datos del json de publicaciones usando JsonBinding. El esquema se puede apreciar en la **Ilustración 3-36 Esquema funcionalidad json**. La aplicación que tiene es la de poder analizar los datos del archivo json de publicaciones de la revista que poseemos como base de información y generar una lista de elementos que se utilizaran para crear los datos necesarios para importar los artículos con posterioridad.

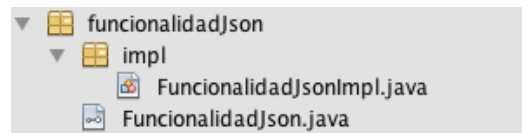


Ilustración 3-36 Esquema funcionalidad json

El siguiente paquete es el correspondiente a la “funcionalidadTSV” que como el resto en su interior tiene una interfaz y una implementación. En la implementación se desarrollarán las distintas funcionalidades que permiten la creación de todos los ficheros que hemos visto hasta ahora, los cuales hemos usado para importar la información a las herramientas de edición de Wikidata. En el paquete “writerClasses” tenemos tres clases. La primera “CreatorCsvFileArticulosAutoridad” permite que creamos el fichero combinado de artículos y autoridades, el “CreatorCsvFileAutoridades” permite que creamos el fichero de autoridades que utilizamos en mix’n’match y por ultimo “CreatorCsvFileAutoridadesConQ”, genera el archivo que usaremos para enriquecer los autoridades existentes. El esquema de lo que se ha hablado esta visible en la **Ilustración 3-37 Esquema de funcionalidad TSV**

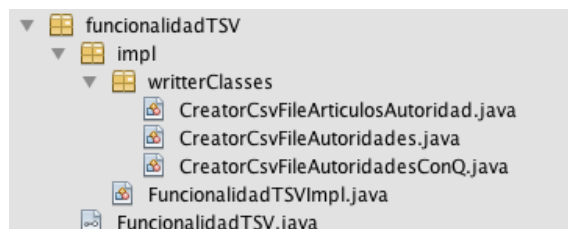


Ilustración 3-37 Esquema de funcionalidad TSV

La interfaz de usuario representa la parte visible del programa. Está formada por un JFrame que se llama “UserInterface” donde se ejecuta el main de la aplicación y por dos JPanel que se llaman “OperacionesAutoridades” y “OperacionesPublicaciones”. Ambas realizan las funcionalidades que su propio nombre indica. Todo lo dicho se puede apreciar en la **Ilustración 3-38 Esquema interfaz de usuario**. Se explicará cómo se debe interactuar con la interfaz con más detalle en el apartado de guía práctica para la migración de datos a Wikidata.

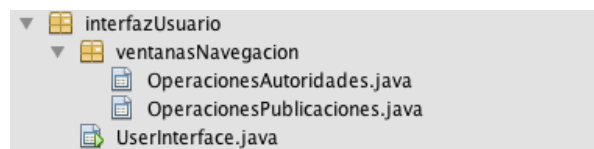


Ilustración 3-38 Esquema interfaz de usuario

El siguiente apartado representa a todas las clases usadas a lo largo de la aplicación para hacer las sucesivas lecturas y generar los correspondientes archivos de trabajo. Todas se pueden ver en la **Ilustración 3-39 Conjunto de clases usadas**. Con estas clases se puede leer y generar cualquier tipo de dato que necesitamos en el proyecto.



Ilustración 3-39 Conjunto de clases usadas

Para terminar, se encuentra el paquete en el que se desarrollan algunas operaciones generales de la aplicación visible en

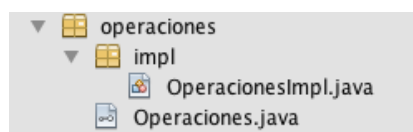


Ilustración 3-40 Esquema operaciones generales

4. RESULTADOS

Sin duda el resultado principal de este proyecto es la migración de los datos, y la descripción de las herramientas y proceso seguido para ello; de las autoridades de la Universidad Pontificia de Salamanca y los artículos científicos que estas autoridades (PDI) hayan publicado en la revista Cuadernos Salmantinos de Filosofía. La incorporación a Wikidata de estos contenidos supone visibilizar y enriquecer los contenidos científicos propios de la UPSA en una base de conocimiento global, abierta, sostenible y reutilizable.

También la documentación necesaria para la realización del trabajo ofrece una base a proyectos similares o para versiones sucesivas en el futuro que incluyan más funcionalidades.

El proporcionar un método automatizado para la transformación de los datos base a datos usados para la elaboración de las ediciones de elementos Wikidata, es otra aportación e importante resultado del proyecto.

Proporcionar un medio usable, simple y transparente para que el personal que en un futuro pueda llevar a cabo estos procesos sin necesitar un excesivo esfuerzo en el proceso, proporcionando un manual de usuario como fuente de entendimiento para su uso, también es otra aportación relevante.

Por último, la extensibilidad que esta aplicación en particular y todo el proyecto en general, permitirá en relación con cualquier otra revista de la biblioteca de nuestra universidad y/o en cualquier otro corpus similar existente en los portales de datos de la UPSA, de su Biblioteca o de cualquier otra institución similar.

4.1 APLICACIÓN DEL PROCESO DE MIGRACION A WIKIDATA

La guía práctica y manual de usuario del proyecto es bastante simple. Una vez que se obtengan los dos ficheros necesarios, csv de autoridades de la UPSA y fichero json de las publicaciones de la revista haremos uso de nuestra aplicación de escritorio.

Al abrir dicha aplicación nos encontramos con una ventana de inicio como la que podemos ver en la Ilustración 4-1 Ventana de inicio aplicación

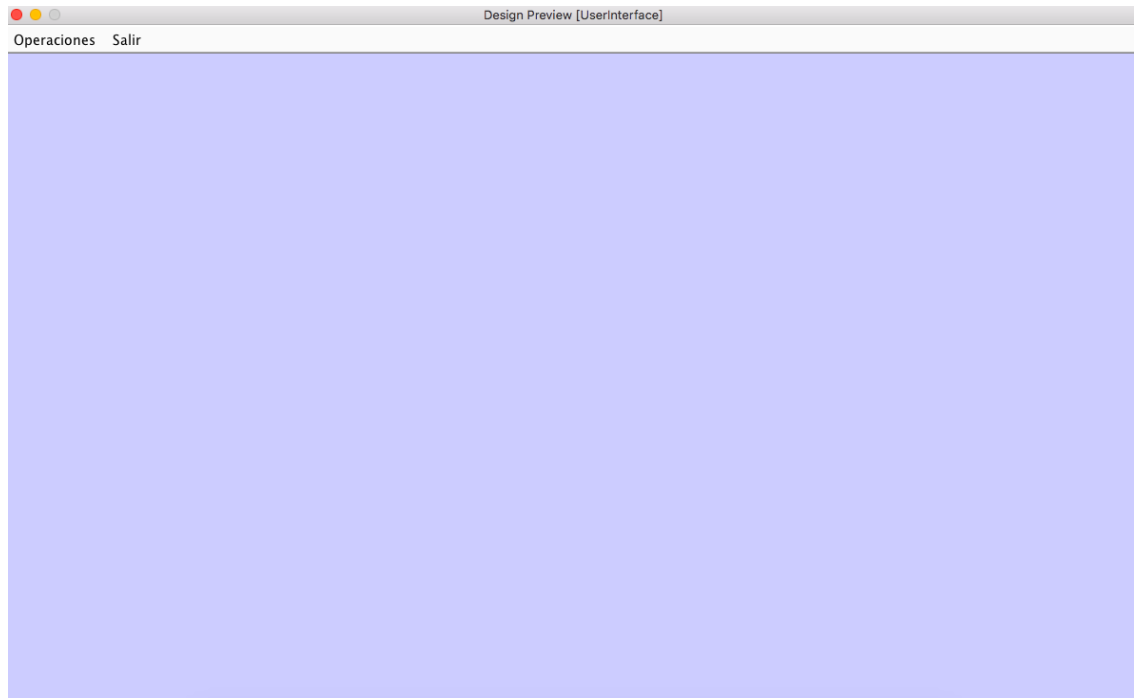


Ilustración 4-1 Ventana de inicio aplicación

Si queremos empezar a hacer operaciones con los datos que tenemos en los ficheros deberemos ir al menú de la aplicación, al elemento que pone operaciones. Dentro del desplegable que aparecerá tendremos la posibilidad de hacer operaciones con las autoridades o con las publicaciones.

Cada uno de los ítems contiene su propia ventana de operaciones. Si en algún momento deseamos finalizar la ejecución del programa lo único que debemos hacer es pulsar en el segundo ítem del menú que pone “Salir”

Comencemos con las operaciones sobre las Autoridades de la UPSA. La ventana de trabajo es visible en la **Ilustración 4-2 Ventana de Operaciones con las Autoridades**

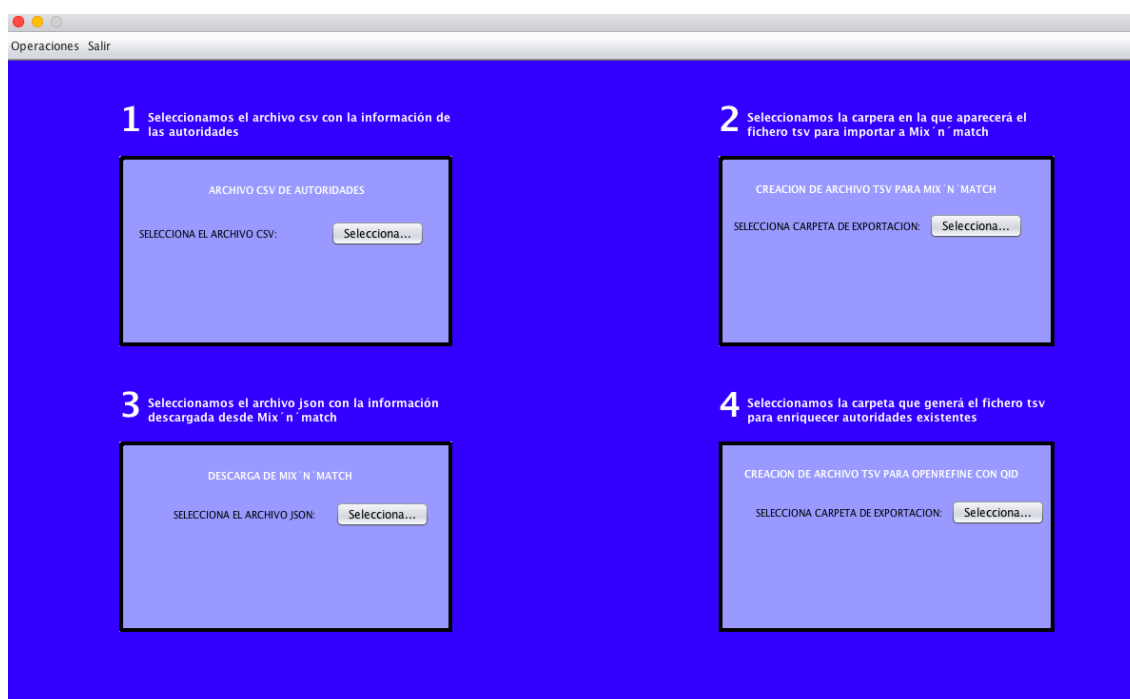


Ilustración 4-2 Ventana de Operaciones con las Autoridades

Los cuadros que se encuentran en la parte superior de la ventana serán los que nos permitirán obtener la información del fichero .csv de autoridades de la Universidad Pontificia de Salamanca y obtener el fichero .tsv que usaremos para crear el catálogo. Para ello deberemos empezar usando el cuadro de la izquierda en el cual usando el administrador de archivos podremos seleccionar el fichero csv. Obtendremos un mensaje con el nombre del fichero en cuestión justo de debajo para cerciorarnos de que en efecto es el archivo deseado. A continuación, pasaremos al cuadro de la derecha donde podemos escoger la localización del dónde queremos que se genere el fichero tsv y crear con el catálogo de Autoridades de la UPSA en la herramienta Mix'n'match.

Deberemos de acceder al navegador para usar Mix'n'match y crear el catálogo en cuestión como se comentó en los apartados anteriores del proyecto. Una vez logrado esto deberemos descargar los datos correspondientes al mismo en el formato json desde su ventana de importación.

Finalmente usaremos los cuadros que se encuentran en la parte inferior de la ventana para hacer uso de los datos descargados. Empezaremos de nuevo por el cuadro de la izquierda, en él podremos seleccionar el fichero json con la información del catálogo mediante el administrador de archivos. También veremos el nombre del archivo seleccionado lo que nos permite cerciorarnos de que en efecto es el correcto.

Finalmente, para terminar con las autoridades pasaremos al cuadro de su derecha, donde escogeremos el lugar donde se descargará el archivo tsv que importaremos en OpenRefine posteriormente para enriquecer las autoridades que de forma previa ya existan en Wikidata.

Una vez hayamos terminado con la parte correspondiente a las autoridades de la UPSA, pasaremos a la ventana de las operaciones con las publicaciones. Lo único que debemos hacer es seleccionar en el menú de operaciones la entrada con el mismo nombre, y ya podremos ver

el menú que aparece en la imagen **Ilustración 4-3 Ventana de operaciones con publicaciones de la revista**



Ilustración 4-3 Ventana de operaciones con publicaciones de la revista

Como en la ventana de las operaciones con autoridades empezaremos con los cuadros que se encuentran en la parte superior de la ventana, concretamente con el elemento de la izquierda. Lo que haremos con él es seleccionar también mediante el administrador de archivos el archivo json donde tenemos almacenados los datos sobre las publicaciones de la revista Cuadernos Salmantinos de Filosofía. Justo después podremos ver el nombre del fichero que acabamos de seleccionar para verificar que es el correcto como en los casos anteriores.

Lo mismo haremos con el cuadro de su derecha, pero en este caso el fichero será el perteneciente a la información de las autoridades que utilizamos al principio.

Cuando hayamos hecho todo esto ya dispondremos de la información necesaria para generar el fichero que contenga las publicaciones de la revista Cuadernos Salmantinos correspondientes a las autoridades de la UPSA.

Para obtener los datos que usaremos para crear un nuevo proyecto en OpenRefine para la creación de los artículos de la revista usaremos el cuadro inferior de la ventana.

En este nuevo cuadro deberemos indicarle la carpeta a la que queremos exportar los datos sobre los artículos en formato tsv y a continuación pulsar en el botón “Calcular resultado” con lo que obtendremos el ultimo archivo necesario para la migración

5. CONCLUSIONES

EL objetivo último de este trabajo es la creación de un sistema capaz de gestionar la migración tanto de las autoridades de la Universidad Pontificia de Salamanca, como los artículos que éstas hayan publicado en la revista Cuadernos Salmantinos de Filosofía, también la posibilidad de extensión posterior a otras revistas, no únicamente a la revista nombrada. Todo esto se lleva a cabo en la consecución del presente proyecto y también se concluye el trabajo con las siguientes aportaciones de interés.

La búsqueda de una método sencillo y accesible para el personal que posteriormente se encargará de aplicar esta metodología al resto de las revistas, lo cual se consigue mediante la interfaz gráfica de la aplicación java que permite la transformación de los datos para su posterior edición con las herramientas de Wikidata.

Generar la documentación necesaria para generar una base de conocimiento accesible tanto para el personal como para los desarrolladores que puedan en un futuro desarrollar aplicaciones similares o versiones sucesivas de la aplicación

Con las herramientas, la aplicación y la documentación que se exponen en este documento podemos afirmar que tenemos un buen mecanismo para afrontar la propuesta de forma rápida y sencilla. Para propuestas similares se requerirán una serie de modificaciones o actualizaciones que seguirían el mismo patrón que los procedimientos llevados a cabo en el presente trabajo.

Personalmente la realización de este trabajo, aunque no es comparable a lo que tendré que hacer en el futuro, en un trabajo, ha conseguido que empiece a desarrollar las habilidades y aptitudes necesarias para cumplir con los objetivos que se me encomienden. Por otro lado, me ha dado seguridad en mí mismo para lograr mis objetivos personales. Cuando empecé a elaborarlo lo veía como un trabajo más para terminar la carrera, pero una vez que me acercaba a su conclusión, he visto que es más que eso y alcance a comprender porque es tan importante el TFG en los grados

6. LÍNEAS FUTURAS

Como mejoras futuras de la implementación del presente proyecto podría plantearse la inclusión de la API de Wikidata para realizar todos o algunos de los procesos, en lugar de las herramientas mencionadas. Sin embargo, no se considera del todo útil esta propuesta ya que nunca podrá ser superior a las herramientas que la propia Wikimedia ha implementado para trabajar con su base de datos Wikidata y que se encuentran a nuestra absoluta disposición. Lo único que debemos hacer para acceder a su contenido es registrarnos como usuarios de la plataforma.

La interfaz de usuario que presenta este proyecto es muy simple, es algo que podría mejorarse en futuras versiones.

El código podría ser más eficiente a la hora de hacer los cálculos, tal vez con el uso de contenedores más especializados.

Finalmente, o más importante a tener en cuenta es la posibilidad de expandirse incluyendo en el código fuente nuevas funcionalidades para trabajar con otros elementos de la biblioteca

7. BIBLIOGRAFÍA

1. wikidata tutorial SPARQL. [En línea] [Citado el: 05 de mayo de 2021.] https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial.
2. tools Wikidata. [En línea] [Citado el: 05 de mayo de 2021.] <https://www.wikidata.org/wiki/Wikidata:Tools>.
3. Wikimedia. Crear nuevo usuario. [En línea] [Citado el: 09 de 07 de 2021.] <https://www.wikidata.org/w/index.php?title=Special:CreateAccount&returnto=Wikidata%3AMain+Page>.
4. Wikidata. Wikidata. [En línea] [Citado el: 9 de 7 de 2021.] https://meta.wikimedia.org/wiki/Mix%27n%27match/Manual#How_does_it_work?.
5. Wikidata Mix'n'match características importación. [En línea] [Citado el: 05 de mayo de 2021.] <https://meta.wikimedia.org/wiki/Mix'n'match/Import>.
6. Mix'n'match notability. [En línea] [Citado el: 06 de mayo de 2021.] <https://www.wikidata.org/wiki/Wikidata:Notability>.
7. Manual de uso de Mix'n'match. [En línea] [Citado el: 06 de mayo de 2021.] https://meta.wikimedia.org/wiki/Mix%27n%27match/Manual#Match_mode.
8. Manual Openrefine . [En línea] [Citado el: 09 de 07 de 2021.] <https://docs.openrefine.org/>.
9. Manual Openrefine. [En línea] [Citado el: 10 de 7 de 2021.] <https://docs.openrefine.org/manual/reconciling#reconciliation-facets>.
10. Manual OpenRefine GREL. [En línea] [Citado el: 06 de mayo de 2021.] <https://docs.openrefine.org/manual/grelfunctions>.
11. Wikidata OpenRefine. [En línea] [Citado el: 10 de 07 de 2021.] https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine/Editing/Schema_alignment.
12. Herramienta QuickStatements. [En línea] [Citado el: 11 de mayo de 2021.] <https://quickstatements.toolforge.org/#/>.
13. Ayuda de QuickStatements. [En línea] [Citado el: 12 de mayo de 2021.] <https://www.wikidata.org/wiki/Help:QuickStatements>.
14. Mix'n'match. Mix'n'match-principal. [En línea] [Citado el: 01 de mayo de 2021.] <https://mix-n-match.toolforge.org/#/>.