

A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from the bar, containing the date.

16-12-2022

# Alexandria: Sistema de recomendación basado en reglas

Coñecemento e Razoamento Automático

Several thin, curved lines in shades of blue and grey sweep upwards from the bottom left corner of the page.

Blanco Filgueira, Guillermo

Freiría Alonso, Nerea

## Contenido

1.	Introducción .....	<b>¡Error! Marcador no definido.</b>
2.	Base de hechos inicial .....	<b>¡Error! Marcador no definido.</b>
i.	Limpieza de los datos resultantes .....	<b>¡Error! Marcador no definido.</b>
ii.	Creación de hechos .....	<b>¡Error! Marcador no definido.</b>
3.	Plantillas .....	<b>¡Error! Marcador no definido.</b>
4.	Reglas .....	<b>¡Error! Marcador no definido.</b>
5.	Casos de prueba .....	<b>¡Error! Marcador no definido.</b>
6.	Integración con Java.....	<b>¡Error! Marcador no definido.</b>

## 1. Introducción

En esta práctica se creará un sistema de recomendación de libros en base a los gustos del usuario. Los datos utilizados se han extraído de la base de datos pública conocida como *Wikidata*, y posteriormente se han tratado esos datos para incluirlos en nuestro programa. Con el objetivo de integrar estos datos en una aplicación usable se ha integrado el sistema en CLIPS a través de su interfaz de Java con la librería gráfica Java Swing.

En este documento se detallarán la obtención de los datos y su tratamiento, así como el desarrollo del sistema inteligente y su base de conocimiento asociada. Finalmente, se comentarán detalles de implementación de la aplicación y decisiones de diseño de su interfaz.

## 2. Base de hechos inicial

Para poder extraer los datos que se van a utilizar se ha empleado el grafo de conocimiento *Wikidata* para lo que se ha utilizado la consulta en SPARQL de la Ilustración 1.

```
select ?libro ?titulo ?generoLabel ?idiomaLabel ?autorLabel
where
{
    ?libro wdt:P31 wd:Q571 .
    ?libro wdt:P1476 ?titulo .
    ?libro wdt:P136 ?genero .
    ?libro wdt:P407 ?idioma .
    ?libro wdt:P50 ?autor .
    SERVICE wikibase:label {bd:serviceParam wikibase:language "es".}
}
```

Ilustración 1 - Base de hechos inicial- Consulta Wikidata

Como se puede observar en la consulta se han seleccionado los datos del nombre del título, el género, el idioma y, por último, el autor. En cuanto a la consulta propiamente dicha, primero se ha seleccionado las instancias del libro y luego se han extraído los datos descritos con anterioridad dado el resultado de la primera consulta. Los campos utilizados son los siguientes:

- **wdt:P31**: se corresponde con *instancia de* (este elemento es un ejemplar de otro elemento).
- **wdt:Q571**: libro (*medio para registrar información en forma de escritura o imágenes impresas.*)
- **wdt:P1476**: se corresponderá con el título (*título de una obra incluyendo el idioma*)
- **wdt:P136**: género (*género de una obra creativa o campo de trabajo de un artista*).
- **wdt:407**: idioma de la obra o del nombre (*idioma de la obra, edición o traducción*)
- **wdt:50**: autor (*creador de un texto*)

Para finalizar, también se incluye la última línea de la consulta para indicar que se utilice el idioma español para la realización de la búsqueda. Los resultados obtenidos son más de 2.200 libros, que fueron tratados mediante la herramienta Excel según se detalla en el siguiente apartado.

## i. Limpieza de los datos resultantes

Tras descargar los resultados de la consulta en formato XLSX (opción que ofrece la plataforma) se procedió a su limpieza en Excel, que constó de los siguientes pasos:

1. Eliminación de los ejemplares que no tenían alguno de los campos de datos.
2. Eliminación de los libros en idiomas que podrían presentar problemas debido a caracteres no ASCII. Se decidió escoger 8 idiomas: alemán, catalán, español, euskera, francés, italiano y portugués; excluyendo el resto.
3. Eliminación de los libros que, tras la anterior limpieza, aún presentaban en su título caracteres no UTF-8.
4. Agrupación por título de los libros que presentaban varios idiomas, géneros y/o autores, asumiendo el nombre del libro como clave primaria y agrupando el resto de los valores separados por comas en sus respectivos campos.
5. Por último, estandarización de formato capitalizando todos los valores (mayúscula primera letra de cada palabra, excepto en el título) y reemplazando los espacios por guion bajo, a modo de carácter separador que no tenga valor semántico para el intérprete de CLIPS.

Para la realización del paso 4 se ha utilizado un plugin para Excel conocido como *Ablebits*. Tras el análisis, se exportaron los datos a CSV con separador de punto y coma, obteniendo una base de datos de aproximadamente 500 libros.

## ii. Creación de hechos

Para compilar los datos en el formato declarativo de hecho en CLIPS se ha empleado un script en Python que toma como entrada el anterior CSV y devuelve la base de hechos volcada en un archivo CLP de CLIPS. En esta sección se procederá a explicar su funcionamiento:

```
entrada = open("baseHechos.csv")
output = open("baseHechos.clp", "w")
output.write("(deftemplate Usuario\n"
             + "    (multislot generosFavoritos)\n"
             + "    (multislot autoresFavoritos)\n"
             + "    (multislot generosVetados)\n"
             + "    (multislot autoresVetados)\n"
             + ") \n"
             + "(deftemplate Libro\n"
             + "    (slot titulo)\n"
             + "    (multislot generos)\n"
             + "    (multislot idiomas)\n"
             + "    (multislot autores)\n"
             + "    (slot favorito (default no))\n"
             + "    (slot recomendado (default no)) \n"
             + ") \n")
```

```

        + ")\n"
        + "(deffacts libros\n")
line = entrada.readline()
while line is not None and line != '':
    atribs = line.split(";")
    titulo = atribs[0]
    generos = atribs[1].replace(",", " ")
    idiomas = atribs[2].replace(",", " ")
    autores = atribs[3].replace(",", " ")
    output.write(f"\t(Libro (titulo {titulo}) (generos
{generos}) (idiomas {idiomas}) (autores {autores[: -1]}))\n")
    line = entrada.readline()
output.write("\n")

```

### Ilustración 2- Creación de Hechos: script Python

La primera parte crea las plantillas correspondientes a los objetos de libro y usuario con los que trabajará el sistema de recomendación. Nótese que en ambos se emplean *multislots* para campos que serán listas de valores. La segunda parte lee línea a línea en archivo CSV separando los campos de atributos en base al separador de punto y coma. Además, en los campos multivaluados (idiomas, autores y géneros) se sustituye la coma por un espacio, que el separador semántico en CLIPS (recordemos que los espacios dentro de los nombres de autores o géneros, así como en el título, ya fueron reemplazados en Excel por guiones bajos). Y tras este tratamiento, genera una nueva línea de hecho dentro del *deffacts* de libros, con el formato que espera el intérprete de CLIPS.

El resultado se puede observar en el archivo *baseHechos.clp* dentro de los archivos del programa que acompaña a este informe, con la definición de las dos plantillas y la definición de los aproximadamente 500 hechos de tipo Libro en la base a las plantillas.

## 3. Plantillas

El anterior resultado compone el conocimiento bruto de nuestro sistema, consistente de datos acerca de los productos (libros) que conoce nuestro sistema de recomendación. Antes de explicar el otro fragmento que completa la base de conocimiento de nuestro sistema (añadiendo las reglas y variables de control necesarias para su ejecución) profundizaremos un poco más en los campos de estas plantillas para entender su interés en el funcionamiento del sistema.

La plantilla de **Usuario** almacena 4 campos multivaluados (listas), de sus géneros y autores favoritos y vetados. Dentro del sistema el usuario podrá indicar que le gusta o disgusta un género o autor. En el primer caso le recomendará más libros que contengan ese género o autor (campos *generosFavoritos* y *autoresFavoritos*), pero en el segundo quedarán vetados, impidiendo al sistema recomendarle libros que contengan ese género o autor (campos *generosVetados* y *autoresVetados*). Este mecanismo es bastante directo de cara al usuario, permitiéndole afinar los resultados en base a

los gustos que ya conoce de si mismo. En la siguiente explicación de la plantilla de libro se menciona otro mecanismo que es más transparente de cara al usuario.

La segunda plantilla de **Libro** almacena los campos de valor único título (que identifica al libro), favorito y recomendado (dos campos lógicos de *si/no* que por defecto son *no*); además de 3 listas de autores, géneros e idiomas. En el sistema de recomendación se podrán marcar como favoritos los libros que le gusten al usuario, activando este campo y haciendo que se recomienden libros similares en género y autor a este. Esto proporciona un mecanismo de recomendación más transparente en función a los ejemplares que le gustan en lugar de preguntar directamente por géneros y autores.

#### 4. Reglas

El archivo de reglas *reglas.clp* en los archivos del programa que acompaña a este documento completa el sistema de recomendación con reglas de conocimiento acerca de qué libros son similares, qué libros recomendar en base a esto y los gustos del usuario (géneros y autores favoritos o vetados, y sus libros favoritos) y una serie de reglas de control para regular la generación de sucesivos ciclos de recomendaciones.

Al inicio del archivo se definen 3 hechos ordenados: el primero la instanciación de una plantilla de usuario que

#### 5. Integración con Java

- Cambiar \_ a ‘ ‘
- Como se usa a interfaz

Creemos reglas para clips que chamme función `regex_max` tamen customizasa.

#### 6. Casos de prueba

En esta sección se detallarán los casos de prueba realizados además de una pequeña explicación sobre el razonamiento seguido para determinar los outputs correctos. Con el objetivo de que las pruebas no fuesen de una complejidad demasiado alta se optó por seleccionar tan solo una parte de los datos disponibles, concretamente 20 ejemplares de libros. Esta batería de datos se puede consultar en el archivo adjunto a este documento (*'baseHechosReducidos.txt'*).

El flujo habitual será que el usuario seleccione uno o varios libros favoritos en la sección de búsqueda y, después, el sistema mostrará libros recomendados en base a los autores y géneros de esos libros seleccionados. Posteriormente, en la pestaña de *Recomendaciones* el usuario podrá refinar la búsqueda en base a los géneros o autores que sean de su gusto, y vetar aquellos que no le agraden.

El output esperado serán aquellos libros que tengan comparten género o autor con al menos uno de los libros favoritos del usuario (estos serán también incluidos en las recomendaciones). Además, se añadirán libros cuyo género y autor el usuario haya indicado explícitamente que le gustan, y no se recomendarán aquellos que tengan un género o autor vetado por el usuario.

PERFIL 1			
INPUT	Búsqueda		-
	Recomendaciones	Gusta	Ensayo, No Ficción
		Favoritos	-
		No gusta	Poesía
OUTPUT			<ul style="list-style-type: none"><li>- <i>A Brief History of Time: From the Big Bang to Black Holes</i></li><li>- <i>The Universe in a Nutshell</i></li><li>- <i>Animal Liberation</i></li><li>- <i>Scouting for boys</i></li><li>- <i>The Open Society and Its Enemies</i></li><li>- <i>Bacchus en Canada</i></li><li>- <i>La Corriveau</i></li></ul>

PERFIL 2			
INPUT	Búsqueda		A Brief History of Time: From the Big Bang to Black Holes
	Recomendaciones	Gusta	-
		Favoritos	-
		No gusta	Poesía y No Ficción
OUTPUT			<ul style="list-style-type: none"><li>- <i>A Brief History of Time: From the Big Bang to Black Holes</i></li><li>- <i>The Universe in a Nutshell</i></li><li>- <i>The Open Society and Its Enemies</i></li></ul>

PERFIL 3			
----------	--	--	--

INPUT	Búsqueda		Effluves
	Recomendaciones	Gusta	Poesía y Peter Singer
		Favoritos	El cavi
		No gusta	Divulgación y no Ficción
OUTPUT			<ul style="list-style-type: none"><li>- Effluves</li><li>- Divagations surrealistes</li><li>- Daughter for Sale</li><li>- El canvi</li><li>- The Open Society and its Enemies</li></ul>

PERFIL 4			
INPUT	Búsqueda		La corriveau
	Recomendaciones	Gusta	Ficción
		Favoritos	Le sommeil des loutres
		No gusta	novela, poesía, Stephen Hawking, no ficción
OUTPUT			<div><div>- All my fallen angelas</div><div>- The Open Society and its Enemies</div></div>

PERFIL 5			
INPUT	Búsqueda		La balada del herido pájaro y otros cuentos
	Recomendaciones	Gusta	-
		Favoritos	-
		No gusta	Julio Escoto
OUTPUT			<div><div>-</div><div>Everything’s Eventual</div><div>-</div><div>All my fallen angelas</div></div>

PERFIL 6			
INPUT	Búsqueda		-
	Recomendaciones	Gusta	Karl Popper
		Favoritos	-
		No gusta	Terror



OUTPUT	- The Open Society and Its Enemies
--------	------------------------------------

PERFIL 7			
INPUT	Búsqueda		-
	Recomendaciones	Gusta	Karl Popper
		Favoritos	The Open Society and Its Enemies
		No gusta	Ensayo
OUTPUT			Sin recomendaciones

PERFIL 8			
INPUT	Búsqueda		La Floraison des nenuphars
	Recomendaciones	Gusta	Merie-Christine Chartier
		Favoritos	Les sommeil des loutres
		No gusta	-
OUTPUT			<div>- La Floraison des nenuphars</div> <div>- Le sommeil des loutres</div> <div>- All My Fallen Angelas</div>

PERFIL 9			
INPUT	Búsqueda		La Floraison des nenuphars
	Recomendaciones	Gusta	Marie-Christine Chartier
		Favoritos	Les sommeil des loutres, All My Fallen Angelas
		No gusta	Julio Escoto
OUTPUT			<ul style="list-style-type: none"><li>- Everything’s Eventual</li><li>- La Floraison des nenuphars</li><li>- Le sommeil des loutres</li><li>- Daughters for Scale</li><li>- All My Fallen Angelas</li></ul>

PERFIL 10			
INPUT	Búsqueda		<ul style="list-style-type: none"><li>- IT</li><li>- Animal Liberation</li></ul>
	Recomendaciones	Gusta	Terror y Cuento
		Favoritos	Everything's Eventual
		No gusta	Ensayo
OUTPUT			<ul style="list-style-type: none"><li>- La balada del herido pájaro y otros cuentos</li><li>- It</li><li>- Everything's Eventual</li><li>- All My Fallen Angelas</li></ul>

PERFIL 11			
INPUT	Búsqueda		La corriveau
	Recomendaciones	Gusta	Karl Popper, Kristin Hannah
		Favoritos	-
		No gusta	Ensayo
OUTPUT			- The Nigtingale

PERFIL 12			
INPUT	Búsqueda		<ul style="list-style-type: none"><li>- Bajo el almendro...junto al volcán</li><li>- A brief history of Time: From the Big Band to Black Holes<ul style="list-style-type: none"><li>- La Corriveau</li></ul></li></ul>
			Gianna Patriarca, Julio Escoto, No Ficción
	Recomendaciones	Gusta	-
		Favoritos	Novela Histórica, Ensayo
OUTPUT			<ul style="list-style-type: none"><li>- La balada del herido pájaro y otros cuentos</li><li>- Daughters for Scale</li></ul>

	- All My Fallen Angelas
--	-------------------------

PERFIL 13			
INPUT	Búsqueda		Universe in nutshell
	Recomendaciones	Gusta	Kristin Hannah
		Favoritos	-
		No gusta	No Ficción
OUTPUT			<ul style="list-style-type: none"><li>- The nightingale</li><li>- El cavi</li><li>- A brief History of Time: From the Big Bang to Black holes</li><li>- The Universe in a Nutshell</li><li>- The Open Society and its Enemies</li></ul>

PERFIL 14			
INPUT	Búsqueda		Scouting for boys
	Recomendaciones	Gusta	No ficción
		Favoritos	El Cavi
		No gusta	Ensayo, Ficción
OUTPUT			<div>- Efflueves</div> <div>- Divagations surrealistes</div> <div>- Daughters for Scale</div>

PERFIL 15			
INPUT	Búsqueda		It
	Recomendaciones	Gusta	Novela Histórica
		Favoritos	Bajo el Almendro
		No gusta	Cuento, Krisitin Hannah
OUTPUT			<div>- Bajo el almendro...junto al volcán</div> <div>- Hsiang Fei</div> <div>- It</div>

PERFIL 16			
INPUT	Búsqueda		Effluves
	Recomendaciones	Gusta	Ensayo
		Favoritos	Daughters for sale
		No gusta	Filosofía, Ficción, Catherine Fernland, Stephen Hawkings
OUTPUT			<ul style="list-style-type: none"><li>- Effluves</li><li>- Divagations surrealistes</li><li>- Daughters for Sale</li><li>- El cavi</li><li>- Animal Liberation</li><li>- Scouting for boys</li></ul>

PERFIL 17			
INPUT	Búsqueda		The universe in a nutshell
	Recomendaciones	Gusta	Stephen King
		Favoritos	-
		No gusta	Filosofía, poesía y cuentos
OUTPUT			<ul style="list-style-type: none"><li>- It</li><li>- A brief History of Time: From the Big Bang to Black Holes</li><li>- The Universe in a Nutshell</li><li>- Animal Liberation</li><li>- Scouting for boys</li><li>- Bacchus en Canada</li><li>- La Corriveau</li></ul>

Todos los casos de prueba han expuesto los resultados esperados por lo que se puede afirmar que se trata de una aplicación robusta y el correcto funcionamiento del sistema experto.