

Curso 2019/2020

Grupo 18

Semantic Web, Linked Data and Knowledge Graphs

Javier del Río
Claudio Baldman
Daniel de la Vega
Montserrat Silva
Pablo Beltrán

Elección de Dataset

- Eventos en los próximos 100 días en Madrid
 - Eventos en Bibliotecas en Madrid
-



<https://datos.madrid.es>

Requisitos del Dataset

Eventos en los próximos 100 días en Madrid

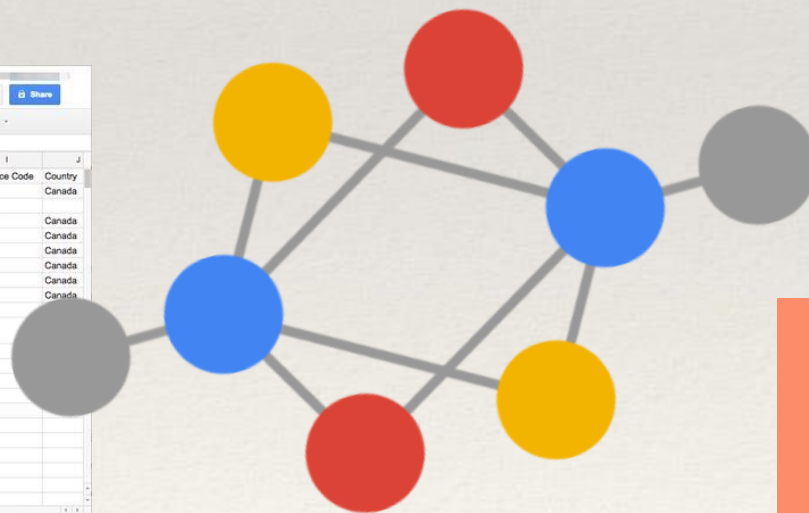
- ✓ Los datos pertenecen a una ciudad
 - ✓ Están disponibles en CSV
-
- ✓ Tienen licencia abierta
 - ✓ Pueden enlazarse con otras entidades

Objetivos

Proyecto

- Crear un grafo de conocimiento a partir de datos proporcionados en formato csv.
- Crear una aplicación que permita hacer uso de dicho grafo.

The image shows a web-based spreadsheet application. At the top, the title bar reads "New John's Apparel CSV". Below it is a menu bar with options: File, Edit, View, Insert, Format, Data, Tools, Add-ons, and Help. To the right of the menu bar, it says "Last edit was 2 minutes ago". On the far right, there are buttons for "Comments" and "Share". The spreadsheet itself has columns labeled A through J. The first row (row 1) contains headers: "First Name", "Last Name", "Email", "City", "Province", "Province Code", and "Country". Rows 2 through 13 contain data for various individuals. A large green circular overlay with a white document icon and the text "CSV" is centered over the spreadsheet data. At the bottom of the screen, there is a status bar with a plus icon, a minus icon, and the text "customers_export -". On the far right of the status bar, there is a green button with a white plus icon and the text "Export".



Aplicación

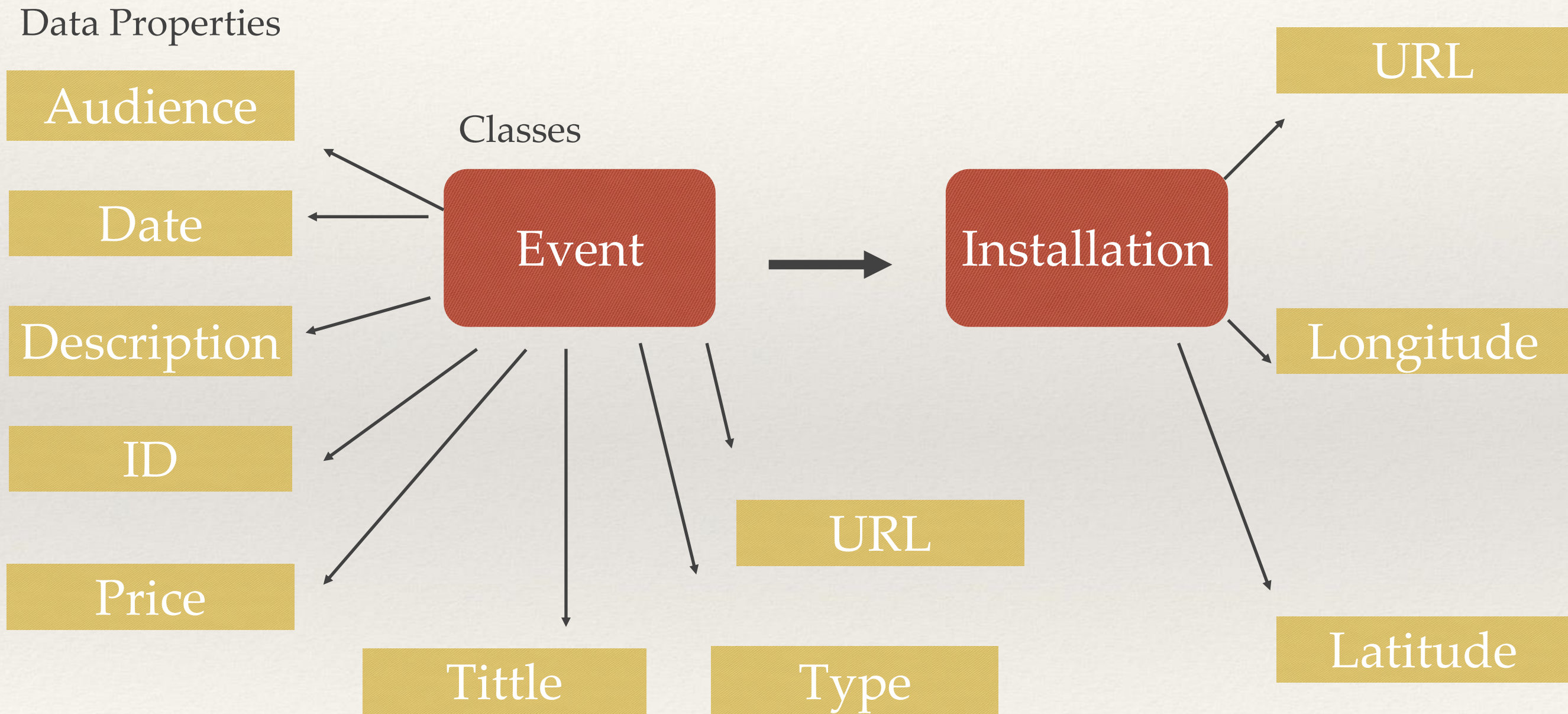
- Presentar los eventos realizados en Madrid
- Permitir el filtrado de información
- Consultar mediante SPARQL a nuestro grafo de conocimiento.



Análisis de Data Set

Data	Type	Range	Description	Problems
ID-EVENTO	Integer	[0-9]	Event ID	No data for many columns. Not a uniform format.
TITULO	String	[a-z][A-Z][0-9]	Event name	
PRECIO	String	[a-z][A-Z][0-9]	Price of admittance	
GRATUITO	Boolean	[0-1]	Indicates whether the event is free to attend	
LARGA-DURACION	Boolean	[0-1]	Indicates whether the event lasts more than a day	
DIAS-SEMANA	String	[L,M,X,J,V,S,D]	Indicates which days of the week the event is attendable	Format is different to other columns
DIAS-EXCLUIDOS	Date	[dd-MM-YYYY]	Indicates which days of the week the event is closed	
FECHA	DateTime	[YYYY-MM-dd HH:mm:ss.s]	Start date of the event	
FECHA-FIN	DateTime	[YYYY-MM-dd HH:mm:ss.s]	End date of the event	
HORA	Time	[HH:mm]	Start time of the event	
DESCRIPCION	String	[a-z][A-Z][0-9]	Summary of the event	
CONTENT-URL	String	[a-z][A-Z][0-9][/:.-?=%@#]	Web page for the contents of the activity	
TITULO-ACTIVIDAD	String	[a-z][A-Z][0-9]	Title for the activity	
URL-ACTIVIDAD	String	[a-z][A-Z][0-9][/:.-?=%@#]	Web page of the activity	
URL-INSTALACION	String	[a-z][A-Z][0-9][/:.-?=%@#]	Web page of the venue	
NOMBRE-INSTALACION	String	[a-z][A-Z][0-9]	Name of the venue	
COORDENADA-X	Integer	[0-9]	ED50 X-coordinate on the city map	
COORDENADA-Y	Integer	[0-9]	ED50 Y-coordinate on the city map	
LATITUD	Double	[0-9]	Latitude on WGS84 system	
LONGITUD	Double	[0-9]	Longitude on WGS84 system	
DISTRITO	String	[a-z][A-Z][-]	District name	
TIPO	String	[a-z][A-Z][/]	Event type	
AUDIENCIA	String	[a-z][A-Z][/,]	Target audience	

Ontología del Dataset



LODREFINE

- Realizamos la importación de nuestro CSV a LODRefine
- Se realizó un tratamiento de los datos
 - Limpieza
 - Homogeneidad de formatos
 - Agrupación de similares

LODREFINE

- Tratamiento de los datos (Limpieza)

Antes

▼ TIPO
/contenido/actividades/ExcursionesItinerariosVisitas
/contenido/actividades/CineActividadesAudiovisuales
/contenido/actividades/DanzaBaile
/contenido/actividades/ProgramacionDestacadaAgendaCu edit

Después

▼ TIPO
ExcursionesItinerariosVisitas
CineActividadesAudiovisuales
DanzaBaile
ProgramacionDestacadaAgendaCu edit

LODREFINE

- Tratamiento de los datos (Limpieza)

Antes

▼ AUDIENCIA
/usuario/Niños
/usuario/Niños
/usuario/Niños,/usuario/Fam edit

Después

Niños
Niños
Niños,Familias

LODREFINE

- Tratamiento de los datos (Homogeneidad de formatos)

Antes

▼ FECHA	▼ FECHA-FIN
2019-10-25 00:00:00.0	2019-10-25 23:59:00.0
2019-10-24 00:00:00.0	2019-10-26 23:59:00.0
2019-10-11 00:00:00.0	2019-10-11 23:59:00.0
2019-12-11 00:00:00.0	2019-12-15 23:59:00.0
2019-10-13 00:00:00.0	2019-10-13 edit 23:59:00.0

Después

▼ FECHA	▼ FECHA-FIN
25/10/2019	25/10/2019
24/10/2019	26/10/2019
11/10/2019	11/10/2019
11/12/2019	15/12/2019
13/10/2019	13/10/2019

Enlazado de datos

- Mediante DBpedia y WikiData
- Logrado en :
 - Distrito
 - Nombre / Instalación

The screenshot displays a data linking interface. At the top right, a dropdown menu labeled 'DISTRITO' is open, showing a progress bar and a list of options: 'Ciudad Lineal' (with a 'Choose new match' link), 'CENTRO', 'Mancha Júcar-Centro (0.316)', 'Centro Cultural Islámico Rey Fahd (0.182)', and 'Centro Histórico de la Ciudad de' (with a value of 0.154). Below this, a 'Search for Match' panel is visible. It contains a search input field with the text 'RETIRO', two radio buttons for 'Match other cells with same content' (selected) and 'Match this cell only', and three buttons: 'Match', 'New Topic', and 'Don't Reconcile Ce'. Below the search panel, a table shows three columns of data: '4475508', '40.42802889001877', and '-3.71018235358'. To the right of the table, a list of search results is shown, each with a description and a URL: 'Jardines del Retiro de Madrid', 'Rascacielos de Retiro', 'Retiro (Buenos Aires)', 'Retiro (Buenos Aires)', 'Retiro (Madrid)', 'Retiro (Madrid)' (highlighted in orange), and 'Calles del barrio de Retiro (Buer'. An 'edit' button is visible on the right side of the interface.

DISTRITO

Ciudad Lineal
Choose new match

CENTRO

Mancha Júcar-Centro (0.316)

Centro Cultural Islámico Rey Fahd (0.182)

Centro Histórico de la Ciudad de (0.154)

Search for Match

Search for "RETIRO"

☒ Match other cells with same content
☐ Match this cell only

Match New Topic Don't Reconcile Ce

4475508 40.42802889001877 -3.71018235358

RETIRO

Select an item from the list:

Jardines del Retiro de Madrid <http://es.dbpedia.org/resour>

Jardines del Retiro de Madrid <http://es.dbpedia.org/resour>

Rascacielos de Retiro <http://es.dbpedia.org/resour>

Retiro (Buenos Aires) <http://es.dbpedia.org/resour>

Retiro (Buenos Aires) <http://es.dbpedia.org/resour>

Retiro (Madrid) <http://es.dbpedia.org/resour>

Retiro (Madrid) <http://es.dbpedia.org/resour>

Calles del barrio de Retiro (Buer <http://es.dbpedia.org/resour>

edit

Aplicación

