

# **Utilizing Knowledge Graphs in Text-centric Information Retrieval**

**Laura Dietz (@lauradietz99)**

University of New Hampshire

**Alexander Kotov (@rusillini)**

Wayne State University

**Edgar Meij (@edgarmeij)**

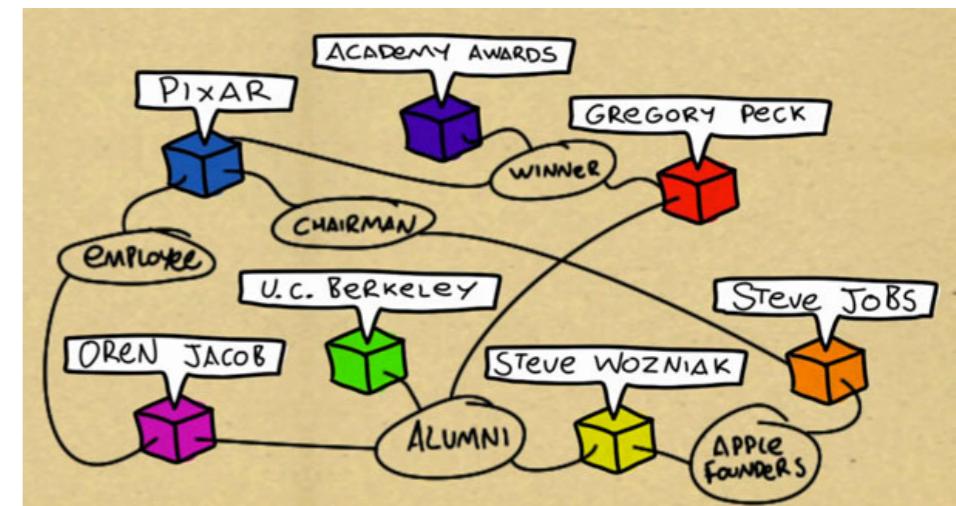
Bloomberg L.P.

# Entity?

- Uniquely identifiable *thing* or *object*
  - “A thing with a distinct and independent existence”
  - people, places, products, companies, etc. etc.

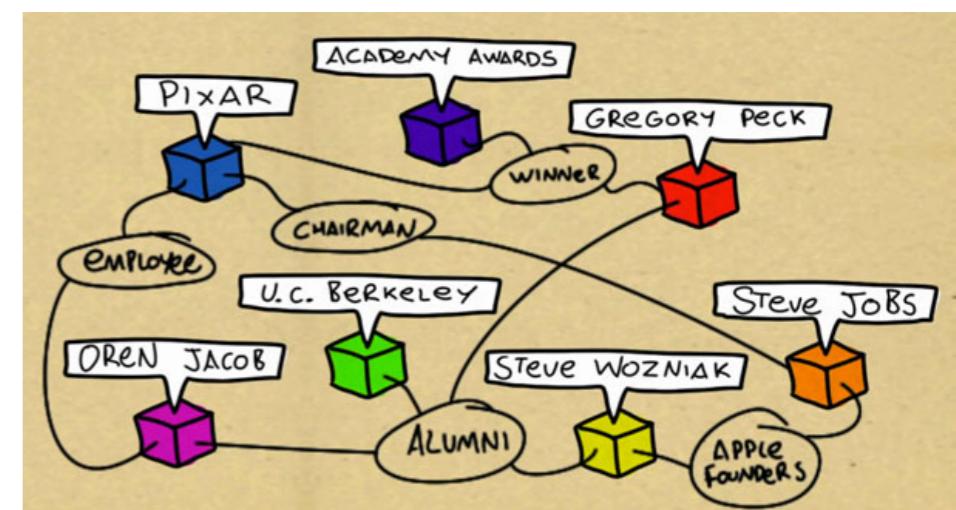
# What's so special about entities?

- ID
- Name(s)
- Type(s)
- Attributes (/Descriptions)
- Relationships to other entities



# Knowledge graphs

- The “backbone” of semantic search
- They define
  - entities
  - attributes
  - types
  - relations
  - (provenance, sometimes)
  - and more
    - external links, homepages, features, ...



# Knowledge graphs

**dbpedia:Cambridge**

<b>foaf:name</b>	Cambridge
<b>rdfs:label</b>	Cambridge
<b>rdfs:comment</b>	Cambridge (/ 'keɪmbrɪdʒ/ KAYM-brij) is a university city and the county town of Cambridgeshire, England, on the River Cam about 50 miles (80 km) north of London. At the United Kingdom Census 2011, its population was 123,867, including 24,488 students [...]
<b>dbpedia-prop:areaCode</b>	1223
<b>dbpedia-prop:areaTotalSqMi</b>	15.71
<b>dbpedia-prop:timezone</b>	<a href="#">dbpedia:Greenwich Mean Time</a>
<b>owl:sameAs</b>	<a href="#">nyt:Cambridge</a> <a href="#">freebase:Cambridge</a> ...
<b>rdf:type</b>	<a href="#">dbpedia-owl:PopulatedPlace</a> <a href="#">dbpedia-owl:City</a> ...
is <b>dbpedia-owl:birthPlace</b> of	<a href="#">dbpedia:Olivia Newton-John</a> <a href="#">dbpedia:Richard Attenborough</a> ...
...	

# **Popular (semi)structured data sources**

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO

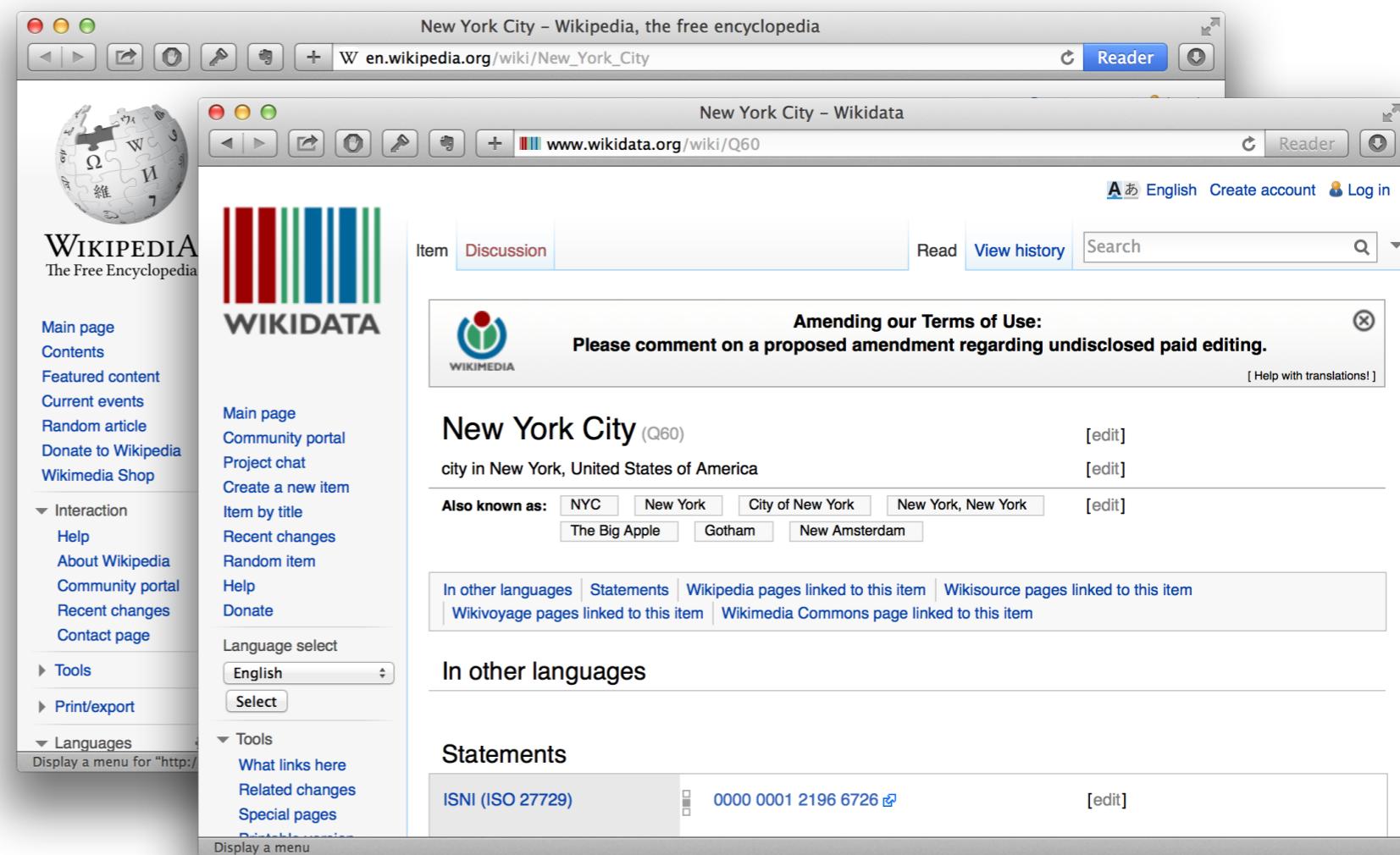
# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# Popular (semi)structured data sources

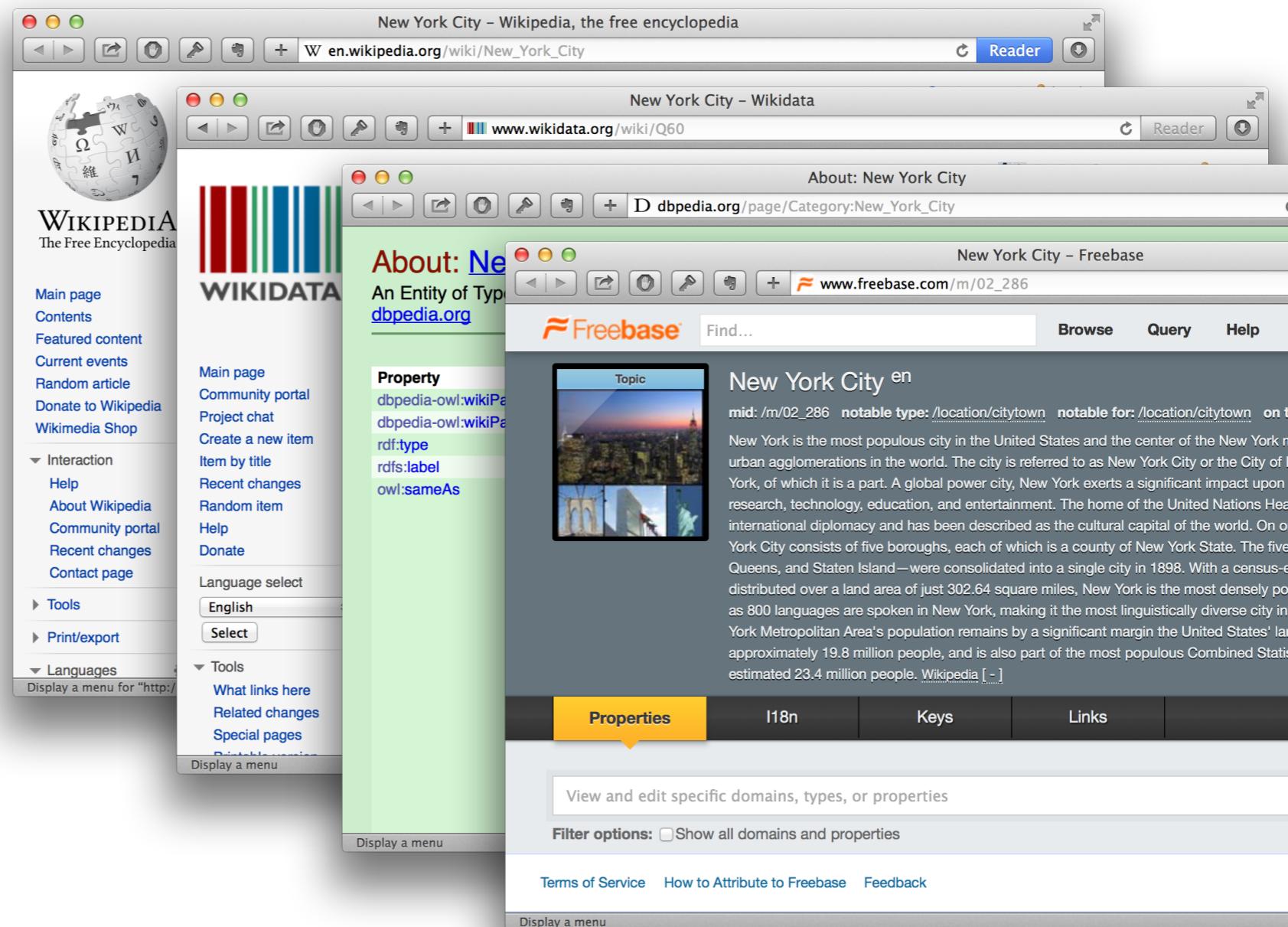
- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO

The image displays three separate browser windows side-by-side, each showing a different structured data source for "New York City".

- Left Window (Wikipedia):** Shows the main Wikipedia page for New York City.
- Middle Window (Wikidata):** Shows the Wikidata item page for New York City (Q60). It includes a sidebar with navigation links like Main page, Community portal, and Project chat. The main content area shows the entity's type as Concept and lists properties and values such as dbpedia-owl:wikiPageID (700743), dbpedia-owl:wikiPageRevisionID (543794976), rdf:type (skos:Concept), rdfs:label (New York City), and owl:sameAs (multiple links to various DBpedia pages).
- Right Window (DBpedia):** Shows the DBpedia category page for New York City. It lists the entity as an Entity of Type Concept from the dbpedia.org Named Graph. A table below shows properties and their corresponding values, such as dbpedia-owl:wikiPageID (700743), dbpedia-owl:wikiPageRevisionID (543794976), rdf:type (skos:Concept), rdfs:label (New York City), and owl:sameAs (multiple links to various DBpedia pages).

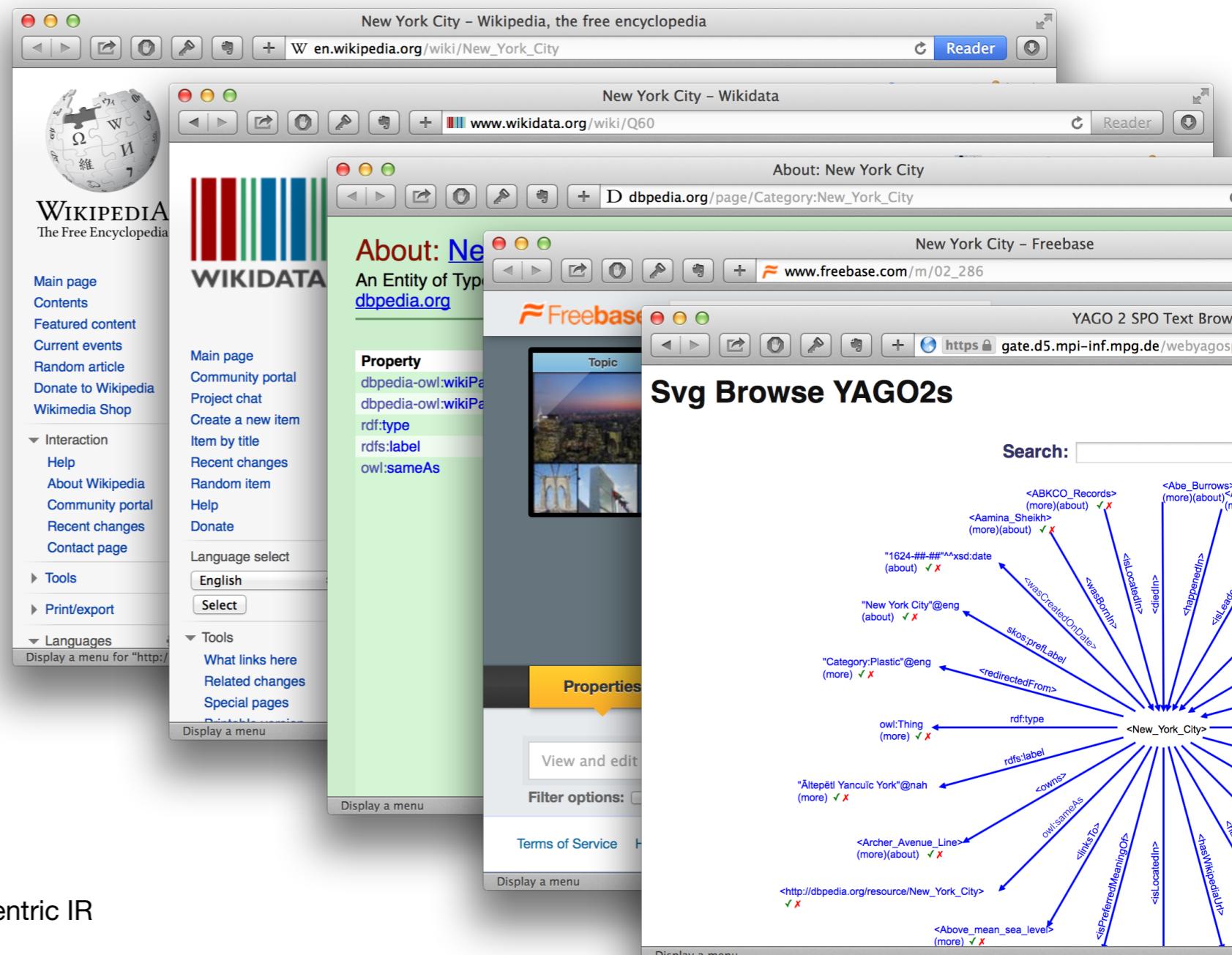
# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# DBpedia

- Extract structured information from Wikipedia
  - infoboxes, categories, and more
  - crowd-sourced community effort
- Open source
  - written in Scala, Java and VSP
  - Virtuoso Universal Server Operating system
- See <http://dbpedia.org/About>

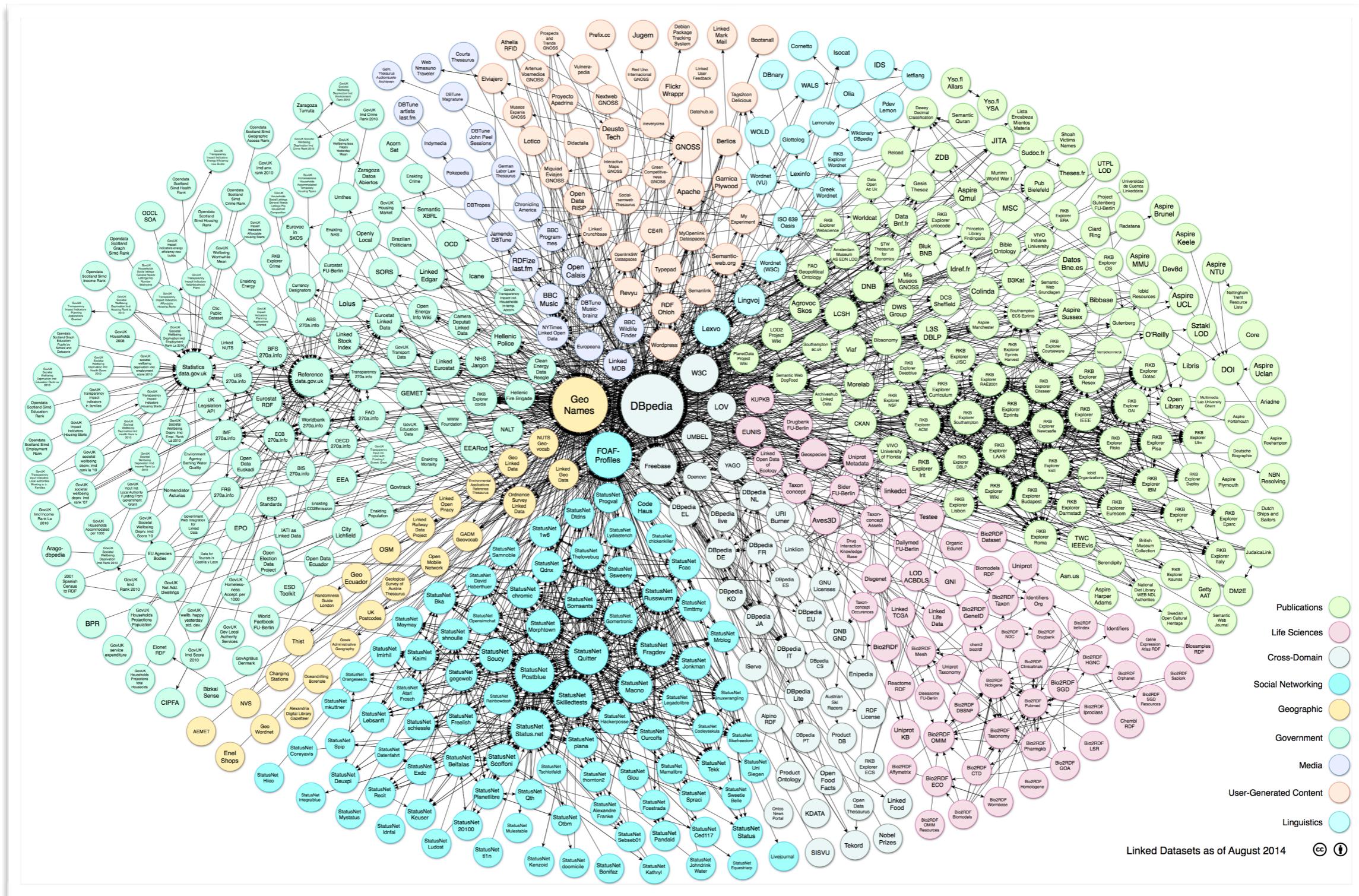
# Freebase

- Initially seeded from high-quality open data
  - then maintained mainly by community
- Harvested from many sources
  - Wikipedia, MusicBrainz, and others.
- Acquired by Google in 2010 (GKG)
  - now in read-only mode
- See <http://www.freebase.com/>

# YAGO

- Accuracy manually evaluated
  - confirmed accuracy of 95%
  - relations annotated with confidence values
- Anchored in Time and Space
  - Thematic domains (e.g. "music" or "science")
- Based on wikipedia, includes WordNet
- See <http://www.mpi-inf.mpg.de/yago-naga/yago/>

# Linking Open Data (LOD)?



# RDFa, microdata

- schema.org, sitemaps.org
  - used by Google, Bing, Yandex, Yahoo!, IPTC, etc.

# RDFa, microdata

- schema.org, sitemaps.org
  - used by Google, Bing, Yandex, Yahoo!, IPTC, etc.

The Wire (TV Series 2002–2008) - IMDb  
www.imdb.com/title/tt0306414/ +1  
★★★★★ Rating: 9.6/10 - 44,375 votes  
Baltimore drug scene, seen through the eyes of drug dealers, and law enforcement.  
Starring Dominic West, John Doman, Deirdre Lovejoy.

IMDb

Logitech Revue 097855070906 | eBay  
www.ebay.com/ctg/Logitech-Revue-/97019743 +1  
★★★★★ from 51 users - \$99.99 to \$190.00  
eBay: The Logitech Revue is a media streamer with Google TV, which serves as a complete entertainment system. With its Ethernet Interface, this Logitech

eBay

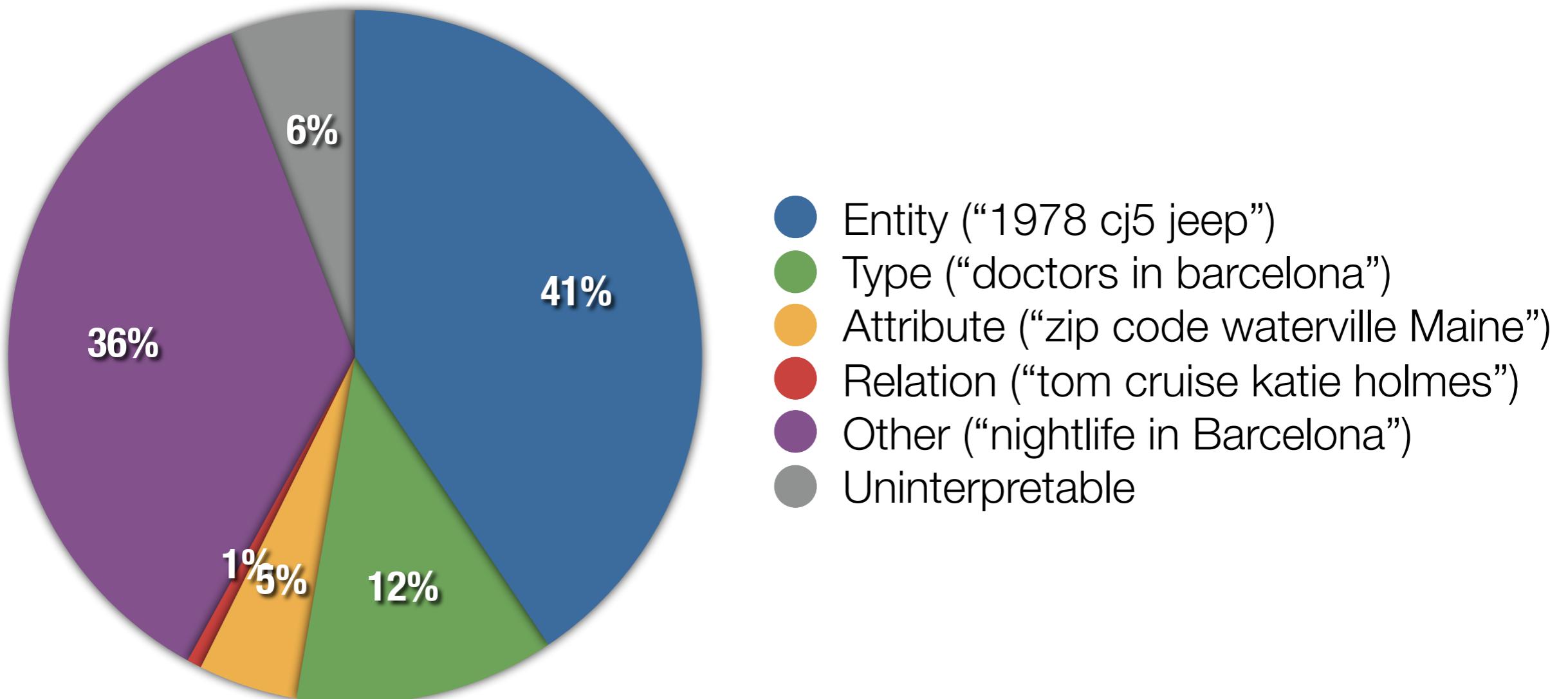
Washington Huskies Mens Basketball tickets, and ... - Ticketmaster  
www.ticketmaster.ca/Washington-Huskies-Mens-Basketball-tickets/art... +1  
Results 1 - 10 of 19 – Find and buy Washington Huskies Mens Basketball ...  
Sat 12 Nov Washington Huskies Mens ... - Alaska Airlines Arena at ...  
Sun 13 Nov Washington Huskies Mens ... - Alaska Airlines Arena at ...  
Mon 14 Nov Washi

ticketmaster®

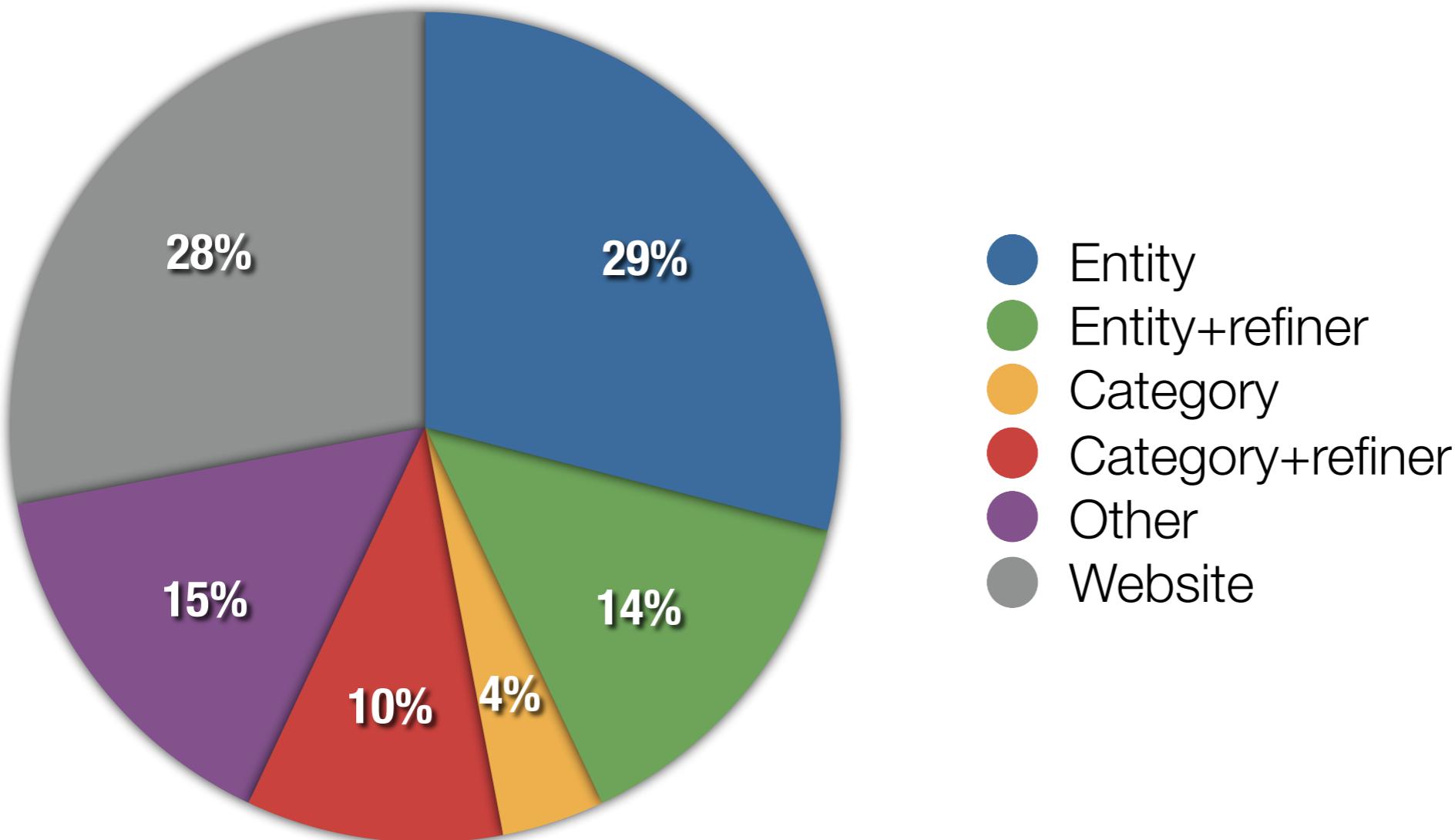
Recette Tarte Tatin - testée et approuvée - calorie et ... - iSaveurs  
www.isaveurs.com/recette/recette\_tart... - Translate this page +1  
45 mins  
il y a 1 jour – Recette de cuisine gourmande - Tarte Tatin - pour virginie63 : J'adooooore j'adore, ...  
http://www.isaveurs.com/recette/recette\_tarte\_tatin.php.

iSaveurs!  
n°1 des gourmands

# Distribution of web search queries [Pound et al. 2010]



# Distribution of web search queries [Lin et al. 2011]



# Entity Linking/Retrieval

Display a menu

All Maps News Images Books More Settings Tools

About 351,000,000 results (1.16 seconds)

**University of Cambridge**  
<https://www.cam.ac.uk/> ▾  
The mission of the University of Cambridge is to contribute to society through the pursuit of education, learning and research at the highest international levels of ...

**Cambridge - Wikipedia**  
<https://en.wikipedia.org/wiki/Cambridge> ▾  
Cambridge is a university city and the county town of Cambridgeshire, England, on the River Cam about 50 miles (80 km) north of London. At the United ...  
University of Cambridge · Cambridgeshire · Cambridge, Massachusetts · St Ives

**Cambridge, Massachusetts - Wikipedia**  
[https://en.wikipedia.org/wiki/Cambridge,\\_Massachusetts](https://en.wikipedia.org/wiki/Cambridge,_Massachusetts) ▾  
Cambridge is a city in Middlesex County, Massachusetts, United States, in the Boston metropolitan area. Situated directly north of the city of Boston, across the ...  
City: 1846 State: Massachusetts  
County: Middlesex Area code(s): 617 / 857

**Cambridge Dictionary | Free English Dictionary, Translations and ...**  
[dictionary.cambridge.org/us/](https://dictionary.cambridge.org/us/) ▾  
Free online dictionary from Cambridge University Press.

**Welcome to the City of Cambridge**  
<https://www.cambridgema.gov/> ▾  
Find all you need to know about living in, working in, and visiting the city of Cambridge.


**Cambridge**  
City in England

Cambridge is a city on the River Cam in eastern England, home to the prestigious University of Cambridge, dating to 1209. University colleges include King's, famed for its choir and towering Gothic chapel, as well as Trinity, founded by Henry VIII, and St John's, with its 16th-century Great Gate. University museums have exhibits on archaeology and anthropology, polar exploration, the history of science and zoology.

**Weather:** 52°F (11°C), Wind S at 19 mph (31 km/h), 85% Humidity  
**Local time:** Thursday 4:19 PM  
**Population:** 123,900 (2011)

**Plan a trip**

 Cambridge travel guide

 3-star hotel averaging \$107

# Entity Linking/Retrieval

www.google.com/?gfe\_rd=cr&ei=O1yTWMbuMbHA8gf15LygDw&gws\_rd=cr&fg=1

Google restaurants in cambridge

All Maps Shopping News Images More Settings Tools

About 44,300,000 results (1.07 seconds)



Map data ©2017 Google

Rating ▾ Price ▾ Hours ▾

**Border Cafe**  
4.2 ★★★★☆ (602) · \$\$ · Restaurant  
Casual spot for Tex-Mex & Cajun eats  
32 Church St  
Usually a wait · Great cocktails · Casual



**Oleana**  
4.6 ★★★★★ (263) · \$\$\$ · Mediterranean  
Memorable Eastern Mediterranean menus  
134 Hampshire St  
Opens at 5:30 PM



**Giulia**  
4.5 ★★★★★ (60) · \$ · Italian  
Intimate spot for classic Italian fare  
1682 Massachusetts Ave  
Opens at 5:30 PM



# Entity Linking/Retrieval

A screenshot of a Google search results page. The search bar at the top contains the query "UK EU relationships". Below the search bar, the "All" tab is selected, along with "News", "Images", "Videos", "Shopping", and "More". To the right of the search bar are "Settings" and "Tools" links, and a "Sign in" button. The page indicates "About 55,300,000 results (0.80 seconds)".

The first result is a link to "Foreign relations of the United Kingdom - Wikipedia" with the URL [https://en.wikipedia.org/wiki/Foreign\\_relations\\_of\\_the\\_United\\_Kingdom](https://en.wikipedia.org/wiki/Foreign_relations_of_the_United_Kingdom). The snippet describes the diplomatic foreign relations of the United Kingdom since 1973, mentioning its membership in the European Union and a 2016 referendum.

The second result is a link to "Britain and the EU: A long and rocky relationship - BBC News" with the URL [www.bbc.com/news/uk-politics-26515129](http://www.bbc.com/news/uk-politics-26515129). The snippet discusses the United Kingdom's relationship with the EU, noting it as one of the most divisive issues in British politics.

A "Top stories" section features a box for "UK Unveils Brexit Plan Calling for Close New EU Relationship" from ABC News, posted 3 hours ago. It includes a link to "More for UK EU relationships".

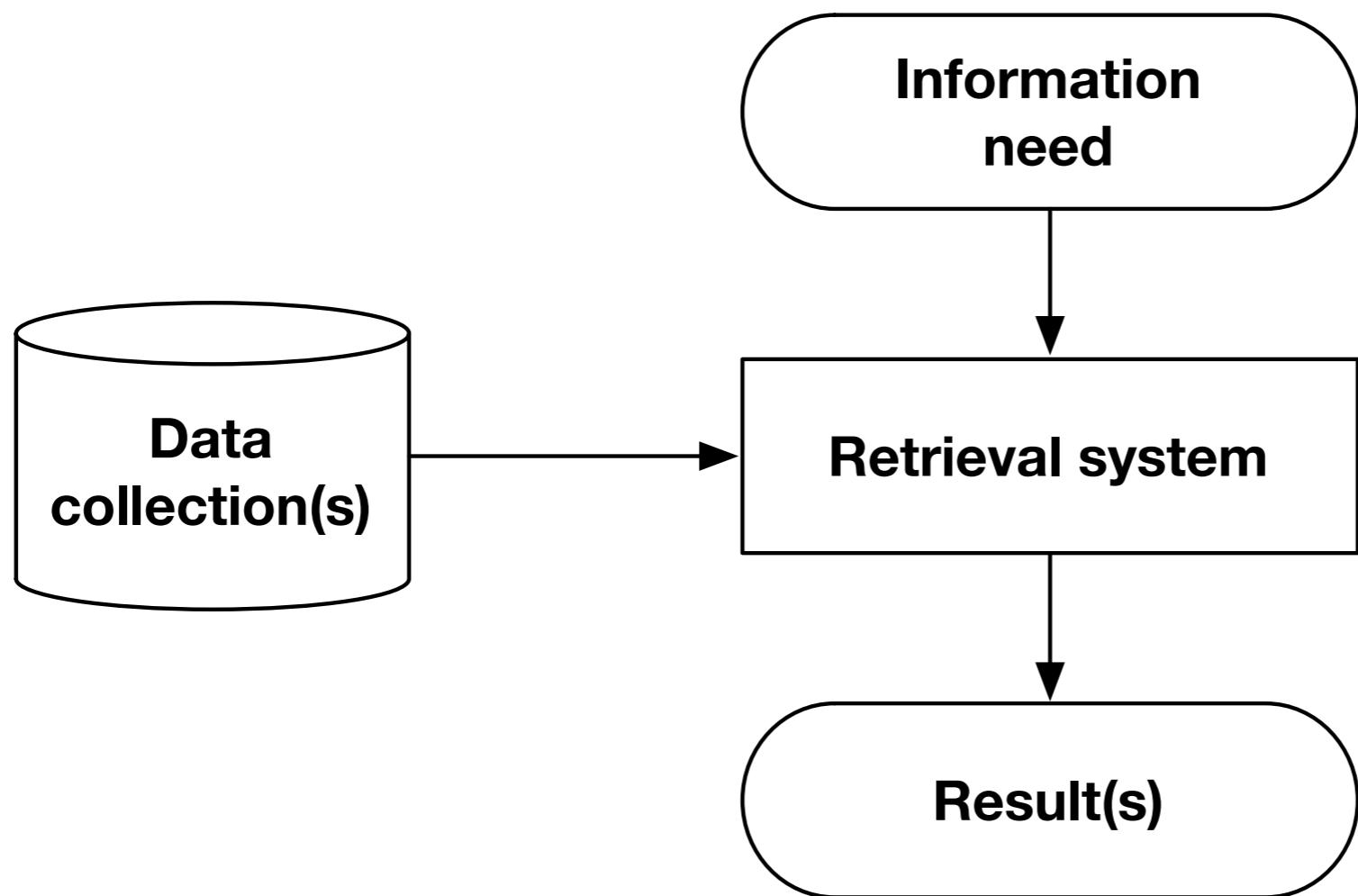
Another result is a link to "European Union - United Kingdom relations" with the URL [www.robert-schuman.eu/en/european-union-united-kingdom-relations](http://www.robert-schuman.eu/en/european-union-united-kingdom-relations). The snippet mentions the UK general election on 7th May 2015 and its complicated relations with European neighbours.

At the bottom left, there is a "Display a menu" link, and at the bottom center, a link to "Great Britain's relationship with Europe - BRIT POLITICS".

# **Today's focus**

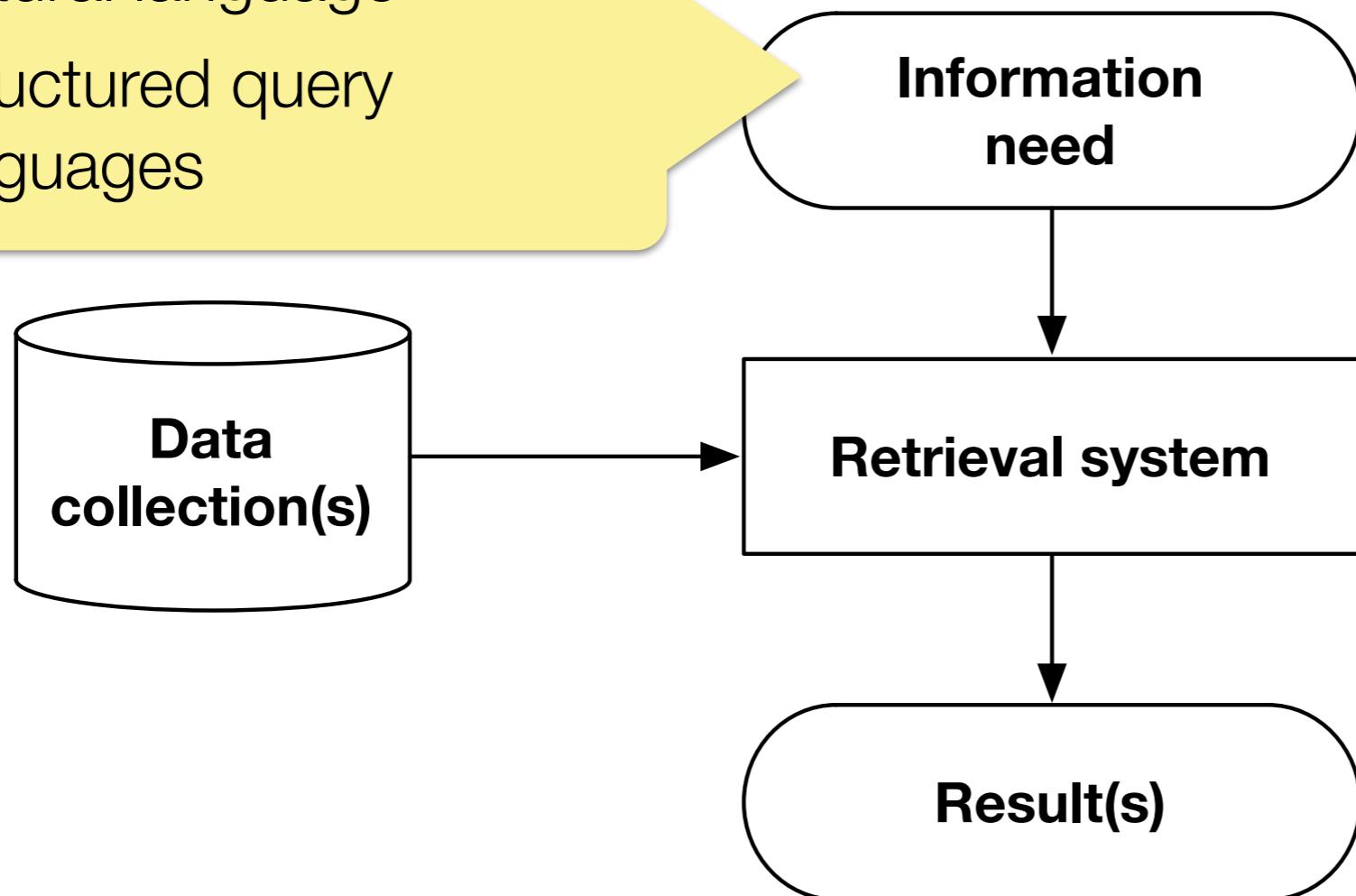
How to use KGs to improve information access.

# Birds-eye view



# Many ways to express

- Keyword
- Keyword++
- Natural language
- Structured query languages



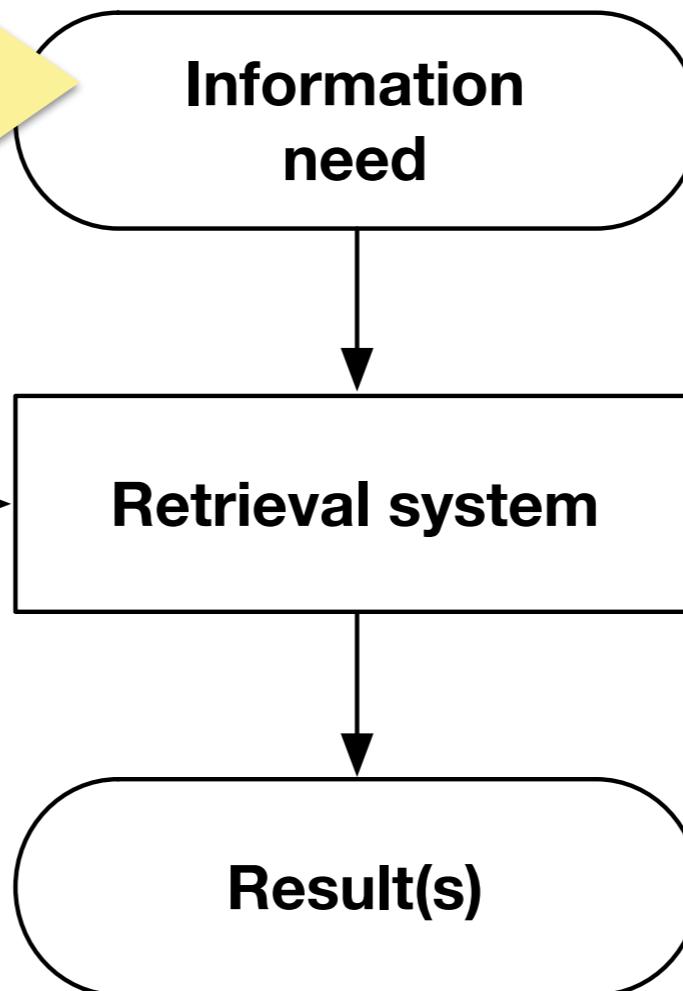
# Many ways to express

- Keyword
- Keyword++
- Natural language
- Structured query languages



# Different types of data

- Unstructured
- Semistructured
- Structured



# Many ways to express

- Keyword
- Keyword++
- Natural language
- Structured query languages



# Different types of data

- Unstructured
- Semistructured
- Structured

Information  
need

Retrieval system

Result(s)

# Result format

- Ranked list
- Tuples
- (Sub)graphs
- Natural language

# Menu

- |                      |   |
|----------------------|---|
| <b>14:00 - 14:15</b> | Introduction                                    |
| <b>14:15 - 14:45</b> | Part 1 – Entity linking                         |
| <b>14:45 - 15:30</b> | Part 2 – Entity Representation<br>and Retrieval |
| <b>15:30 - 16:00</b> | Coffee break                                    |
| <b>16:00 - 17:00</b> | Part 3 – Utilizing KGs in<br>Text-centric IR    |
| <b>17:00 - 17:30</b> | Discussion and wrap-up                          |