

# Entity Linking and Retrieval for Semantic Search

**Edgar Meij** – @edgarmeij  
Yahoo Labs



**Krisztian Balog** – @krisztianbalog  
University of Stavanger



**Daan Odijk** – @dodijk  
University of Amsterdam



# WiFi

- Network: Delta-Meeting
- Password: not needed(?)



Minnesota Children's Museum Deloitte.

GUIDANT

NORTHWEST AIRLINES. ECOLAB

Valspar

Donaldson  
Filtration Solutions



Carlson Companies

AMS  
Solutions for Life



Medtronic  
Alleviating Pain - Restoring Health - Extending Life

Thrivent Financial for Lutherans

MARVIN  
Windows and Doors

3M Worldwide

POLARIS  
The Way Out.

GENERAL MILLS

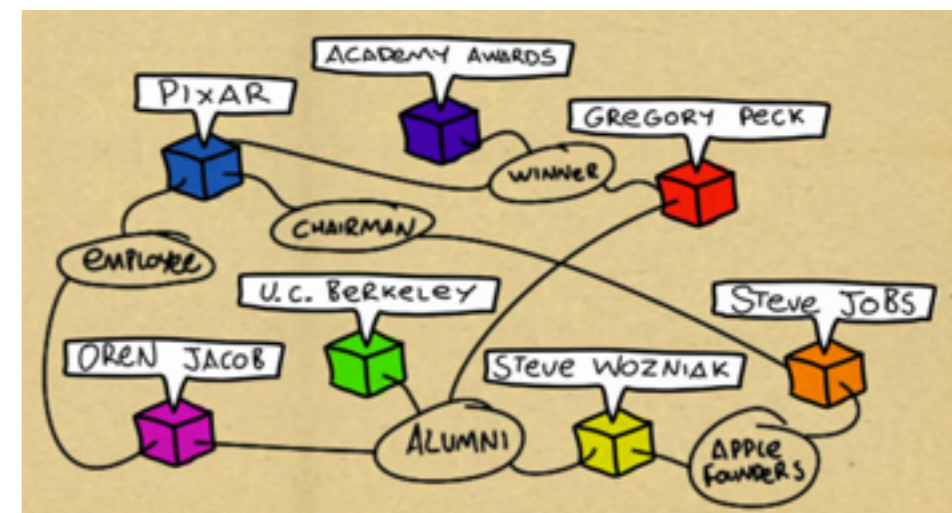


# Entity?

- Uniquely identifiable *thing* or *object*
  - “A thing with a distinct and independent existence”

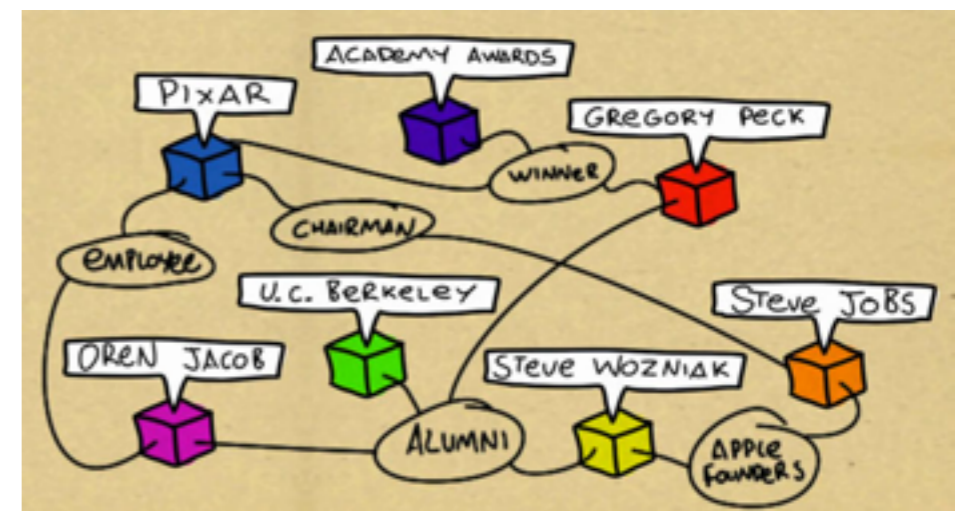
# What's so special about entities?

- 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, ...



# Entity Linking?

## Iranian POW negotiator holds talks with Iraqi ministers

The head of [Iran's prisoner of war](#) commission met with two [Iraqi](#) Cabinet ministers Saturday in a bid to glean information about thousands of Iranian POWs allegedly in Iraq, the official Iraqi News Agency reported.

Iraqi Foreign Minister [Mohammed Saeed al-Sahhaf](#) told Abdullah al-Najafi that the two states needed to "speed up the closure of what remains from the POW and Missing-In-Action file," INA said.

The issue of POWs and missing persons remains a stumbling block to normalizing relations between the two neighbors.

Iraq has long maintained that it has released all Iranian prisoners captured in the [1980-88 Iran-Iraq War](#). The countries accuse each other of hiding POWs and preventing visits by the [International Committee of the Red Cross](#) to prisoner camps.

The ICRC representative in [Baghdad](#), Manuel Bessler, told The [Associated Press](#) that his organization has had difficulty visiting POWs on both sides on a regular basis.

In April, Iran released 5,584 since [1990](#).

More than 1 million people w

### Baghdad

Baghdad is the capital of Iraq and of Baghdad Governorate. With a metropolitan area estimated at a population of 7,000,000, it is the largest city in Iraq. It is the second-largest city in the Arab world (after Cairo) and the second-largest city in southwest Asia (after Tehran).

[open in wikipedia](#)

fied as civil law detainees in the largest exchange

# Entity Linking/Retrieval?

montreal - Google Search

https://www.google.com/#q=montreal

Google  +Edgar Share

Web Maps News Images Videos More Search tools

About 75,900,000 results (0.52 seconds)

**Montreal - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Montreal](http://en.wikipedia.org/wiki/Montreal) - Wikipedia  
In 2009, **Montreal** was named North America's leading host city for ... In 2012, QS World University Rankings ranked **Montreal** the 10th-best place in the world to ...  
[Demographics of Montreal - Second largest city - Greater Montreal - Mount Royal](#)

**Tourism Montreal: Traveller**  
[www.tourisme-montreal.com](http://www.tourisme-montreal.com) - Montreal Official Tourist Information Web Site  
**Montréal** in a Nutshell. Info. **Montréal** is as romantically traditional as it is cutting-edge innovative, as cosmopolitan dynamic as it is small-town friendly.  
[What to do](#) | [Major](#) | [Festival](#) - [Traveller](#) - [Discover](#) - [Hotels](#)



**bienvenue - welcome to montreal.com**  
[www.montreal.com/](http://www.montreal.com/)  
montreal.com. Montreal City Weblog. Find **Montreal** Vacation Rentals on FlipKey. Old **Montreal** by Patrick Martineau. See your horizontal **Montreal** photo in this ...

**Montréal - Lonely Planet**  
[www.lonelyplanet.com/canada/montreal](http://www.lonelyplanet.com/canada/montreal) - Lonely Planet  
Few cities can compete with **Montréal's** mouthwatering mix of food, festivals and fun-centric living. Cuisine Capital Blessed with one of the most...

**News for montreal**

**RAW: Sinkhole on De La Montagne**  
[CBC.ca](#) - 1 hour ago  
A sinkhole big enough to swallow a person opened up right in the middle of downtown **Montreal** on De La Montagne Street on Saturday ...

**Towns in Montreal celebrate 10 demerged years**  
[CBC.ca](#) - 1 hour ago  
**Montreal** the go-to place for raunchy bachelor parties




**Montreal**  
City in Canada


Montreal is a city in the Canadian province of Quebec. It is the largest city in the province, the second-largest in Canada and the 8th-largest in North America. [Wikipedia](#)

**Area:** 141 sq miles (365.1 km<sup>2</sup>)  
**Weather:** 85°F (29°C), Wind 0 mph (0 km/h), 11% Humidity  
**Local time:** Sunday 5:02 PM  
**Population:** 1.621 million (2006) UNdata  
**Province:** Québec


**Points of interest** [View 40+ more](#)




[Mount Royal](#)




[Notre-Dame Basilica of Montreal](#)



[Montreal Museum of Fine Arts](#)



[Montreal Biodome](#)



[Montreal Botanical Garden](#)



# Entity Retrieval?

restaurants in montreal - Google Search

https://www.google.com/#q=restaurants+in+montreal

Google

restaurants in montreal

+Edgar

Share

Web Maps Shopping Images News More Search tools

About 66,300,000 results (0.29 seconds)

**The Keg Steakhouse & Bar - Old Port**  
www.kegsteakhouse.com  
4.3 ★★★★★ 60 Google reviews · Google+ page

**Restaurant Kashima**  
www.kashima.ca  
4.9 ★★★★★ 9 Google reviews

**O.Noir**  
www.onoir.com  
3.9 ★★★★★ 31 Google reviews

**Restaurant Trang An**  
plus.google.com  
4.3 ★★★★★ 7 Google reviews

**L'Auberge Saint-Gabriel**  
www.lesaint-gabriel.com  
4.0 ★★★★★ 12 Google reviews · Google+ page

**Vago**  
www.restaurantvago.com  
3 Google reviews

**Bistro On The Avenue**  
www.bistroontheavenue.com  
3.7 ★★★★★ 5 Google reviews

**Map results for restaurants in montreal**

**A** 25 Rue Saint Paul Est  
Montréal, QC, Canada  
+1 514-871-9093

**B** 1232 Avenue Greene  
Westmount, QC, Canada  
+1 514-934-0962


**C** 1631 Rue Sainte-Catherine Ouest  
Montréal, QC, Canada  
+1 514-937-9727

**D** 7259 Rue Saint-Denis  
Montréal, QC, Canada  
+1 514-272-9992

**E** 426 St Gabriel St  
Montreal, QC, Canada  
+1 514-878-3561

**F** 1336 Avenue Greene  
Westmount, QC, Canada  
+1 514-846-1414

**G** 1362 Avenue Greene  
Westmount, QC, Canada  
+1 514-939-6451



Map for restaurants in montreal

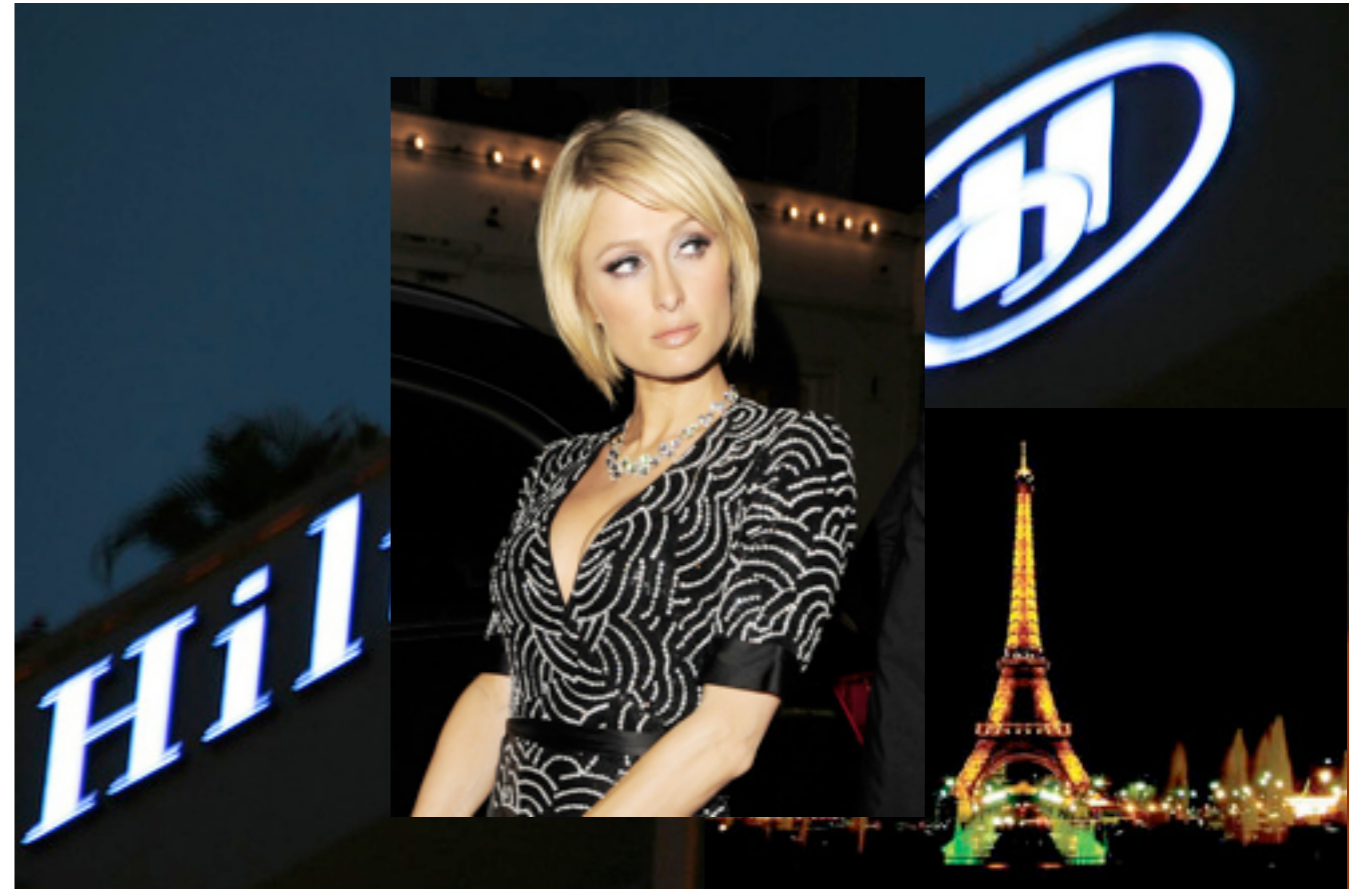
The map shows a portion of Montreal, Quebec, Canada, with several red location pins labeled A through G. The pins are placed in various neighborhoods including Villeray, Saint-Michel-Parc Dension, Rosemont-La Petite-Patrie, Le Plateau-Mont-Royal, Côte-Des-Neiges, and Westmount. The map also shows major roads like Avenue Greene, Rue Saint-Denis, and Rue Saint-Paul, and the St. Lawrence River.

# **Semantic search**

# Semantic search

- Improve search accuracy by **understanding** searcher intent and the **contextual meaning** of terms/documents/...
- Move beyond “ten blue links” (towards actually answering information needs) using rich context

‘hilton paris’



‘countries in africa’



‘good camera  
under \$300’



# Semantic search

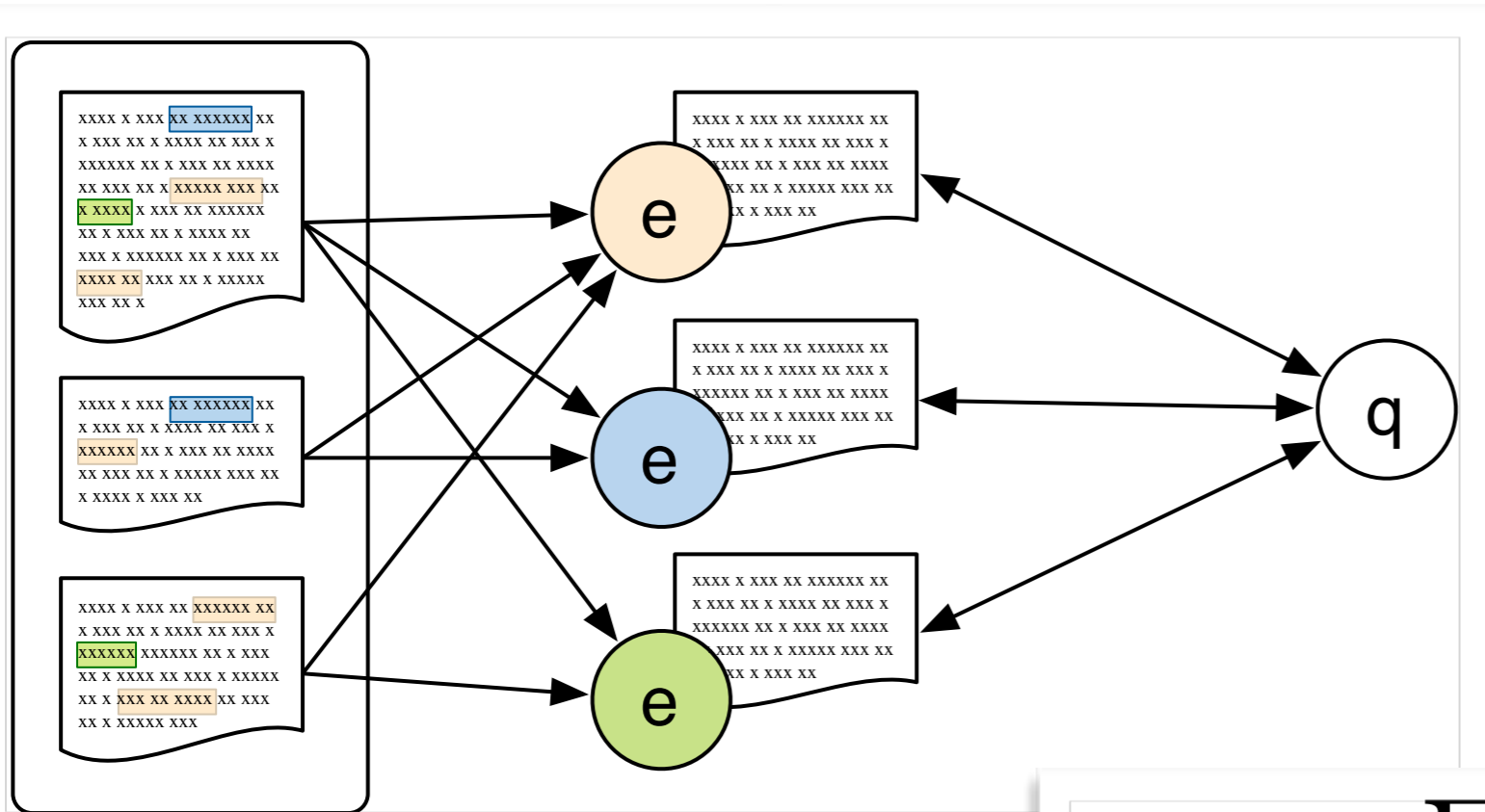
- Centers around entities
  - “Who was the first human in outer space?”
  - “How tall is the Eiffel tower?”
  - “Who is Brad Pitt married to?”
  - “Where is the closest Starbucks?”
  - “Which airlines fly the Airbus A380?”
  - “What is the best Chinese restaurant in Montreal?”
- Entity/Attribute/Relationship retrieval
  - + social, + personal
  - + (hyper)local

# Semantic search

- Combination of entity-related techniques, from various fields
  - IR
  - NLP
  - DB
  - Semantic Web



# IR



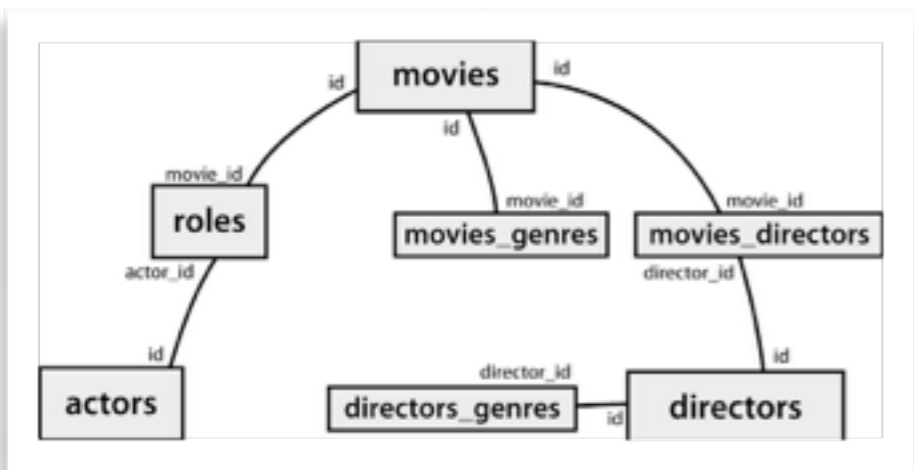
$$P(q|\theta_e) = \prod_{t \in q} \underbrace{P(t|\theta_e)}_{(1-\lambda)P(t|e) + \lambda P(t)}^{n(t,q)}$$
$$\downarrow$$
$$(1-\lambda)P(t|e) + \lambda P(t)$$
$$\downarrow$$
$$\sum_d P(t|d, e)P(d|e)$$

# NLP

- Question answering, relationship extraction



# Databases



**SPARK** searching, probing & ranking

DBLP IMDb

tom cruise nicole kidman

[Preferences](#)

IMDB Results 1 - 10 for "tom cruise nicole kidman"

- actresses** : Name: **kidman, nicole** Id: 206700 54.08
- actressplay** : Charactor: Herself Id: 1053008
- movies** : Name: "Love Chain" (2003) {**tom cruise** (#1.4)} Id: 663077

---

- actors** : Name: **cruise, tom** Id: 142703 31.02
- actorplay** : Charactor: Himself Id: 841769
- movies** : Name: 101 Most Shocking Moments in Entertainment (2003) (TV) Id: 1318
- actressplay** : Charactor: Herself Id: 1052857
- actresses** : Name: **kidman, nicole** Id: 206700

---

- actors** : Name: **cruise, tom** Id: 142703 31.02
- actorplay** : Charactor: Himself Id: 841840
- movies** : Name: Road to the Red Carpet (2002) (TV) Id: 344540
- actressplay** : Charactor: Herself Id: 1052938
- actresses** : Name: **kidman, nicole** Id: 206700

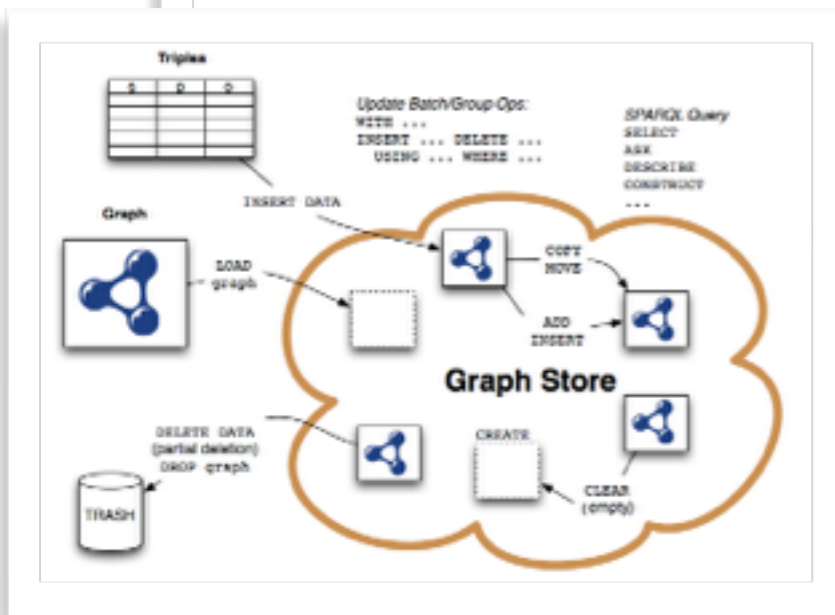
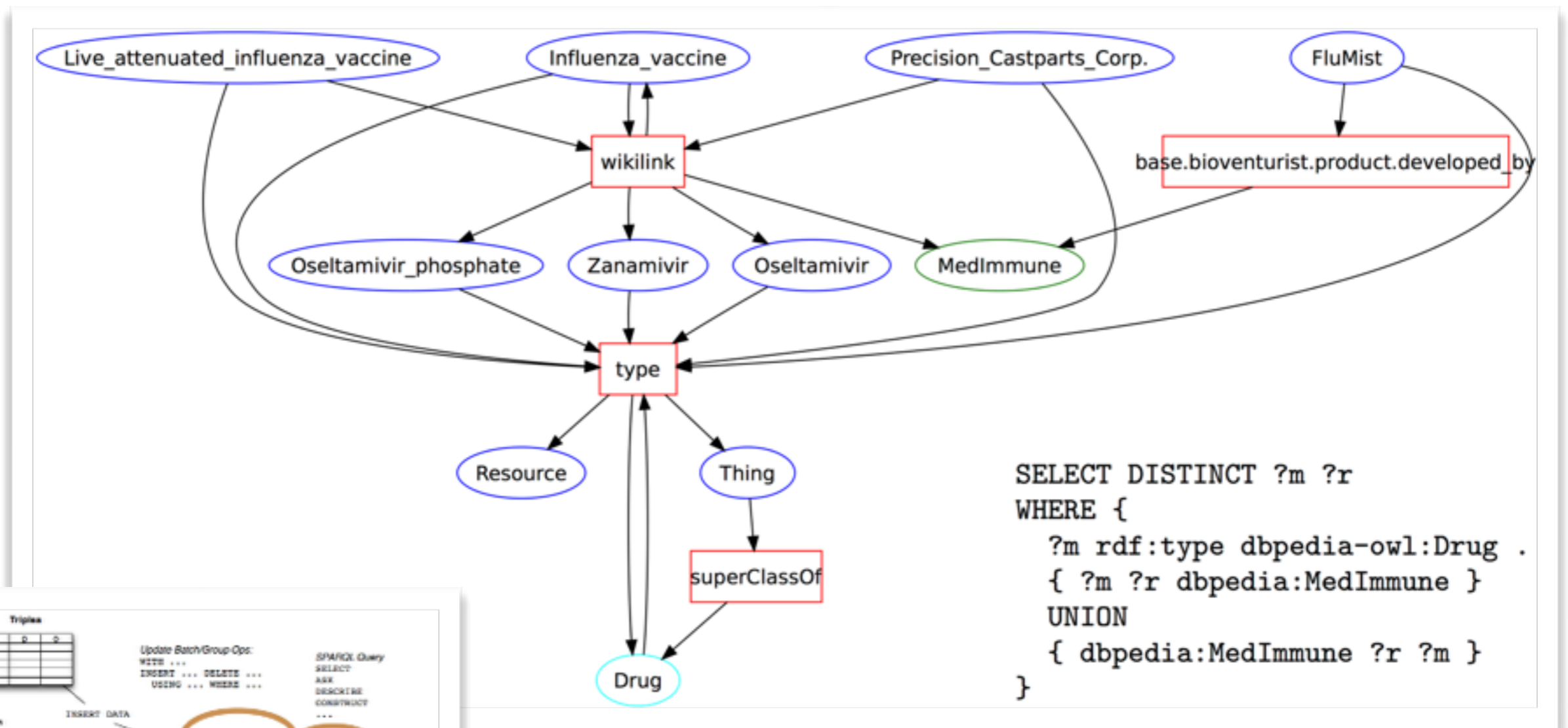
---

- actors** : Name: **cruise, tom** Id: 142703 31.02
- actorplay** : Charactor: Himself Id: 841842
- movies** : Name: Shirtless: Hollywood's Sexiest Men (2002) (TV) Id: 369194
- actressplay** : Charactor: Herself Id: 1052942
- actresses** : Name: **kidman, nicole** Id: 206700

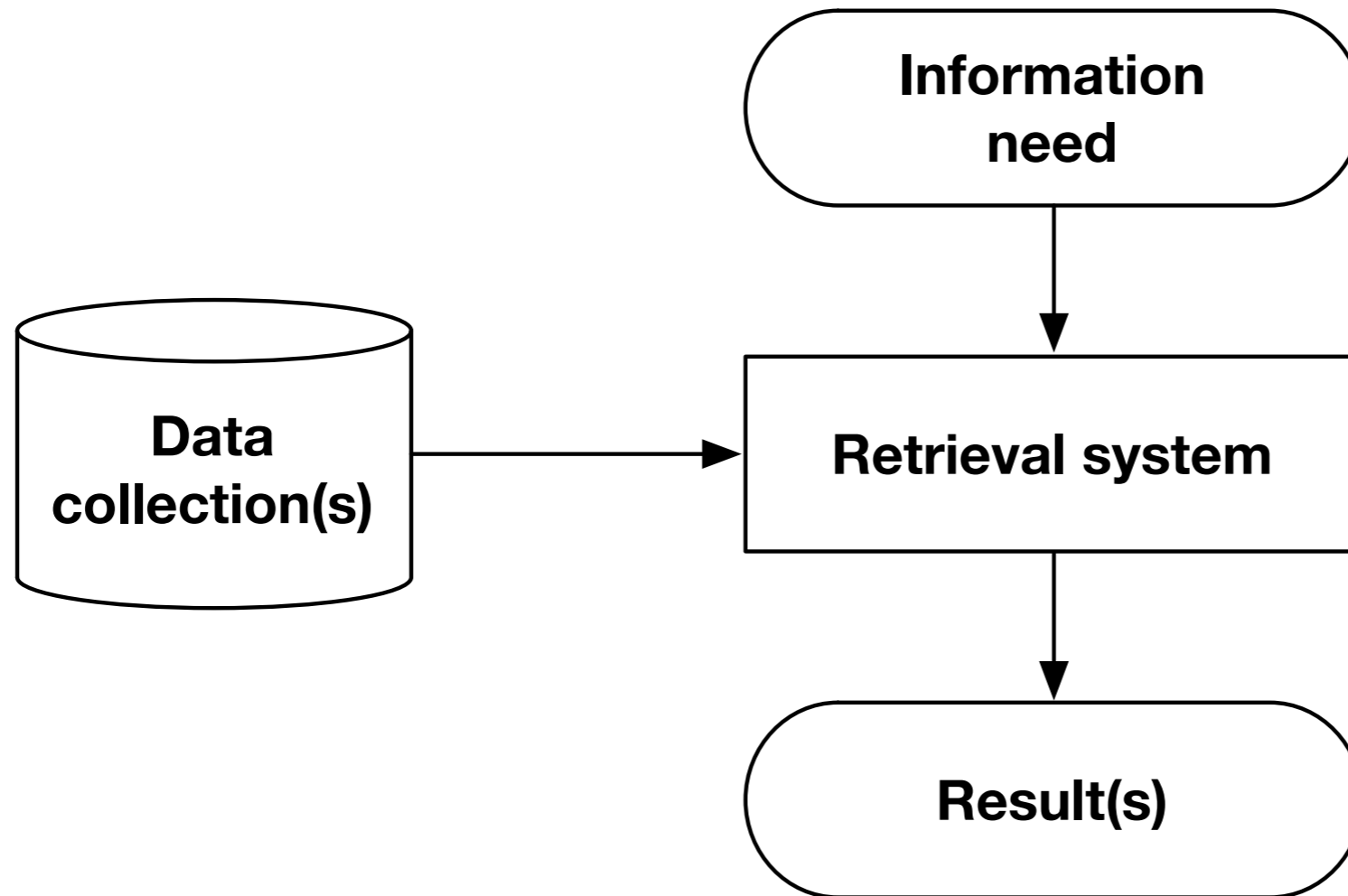
---

- actors** : Name: **cruise, tom** Id: 142703 31.02
- actorplay** : Charactor: Himself (Narrator) Id: 841844
- movies** : Name: Stanley Kubrick: A Life in Pictures (2001) Id: 386247
- actressplay** : Charactor: Herself Id: 1052946
- actresses** : Name: **kidman, nicole** Id: 206700

# Semantic Web

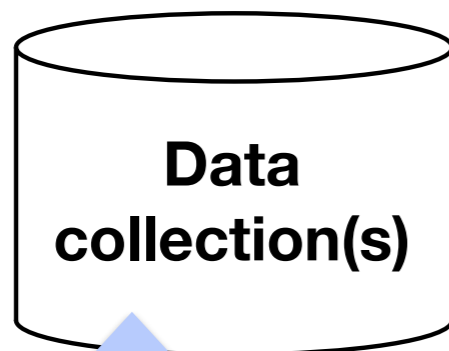


# Birds-eye view



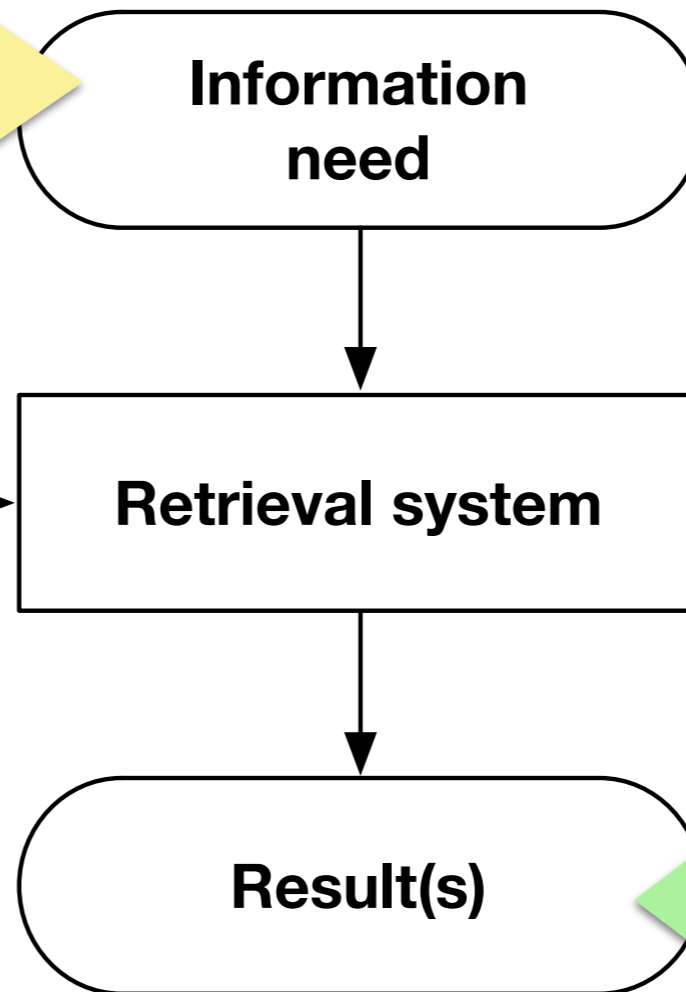
## Many ways to express

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



## Different types of data

- Unstructured
- Semistructured
- Structured



## Result format

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

# The possibilities are endless

IR

DB

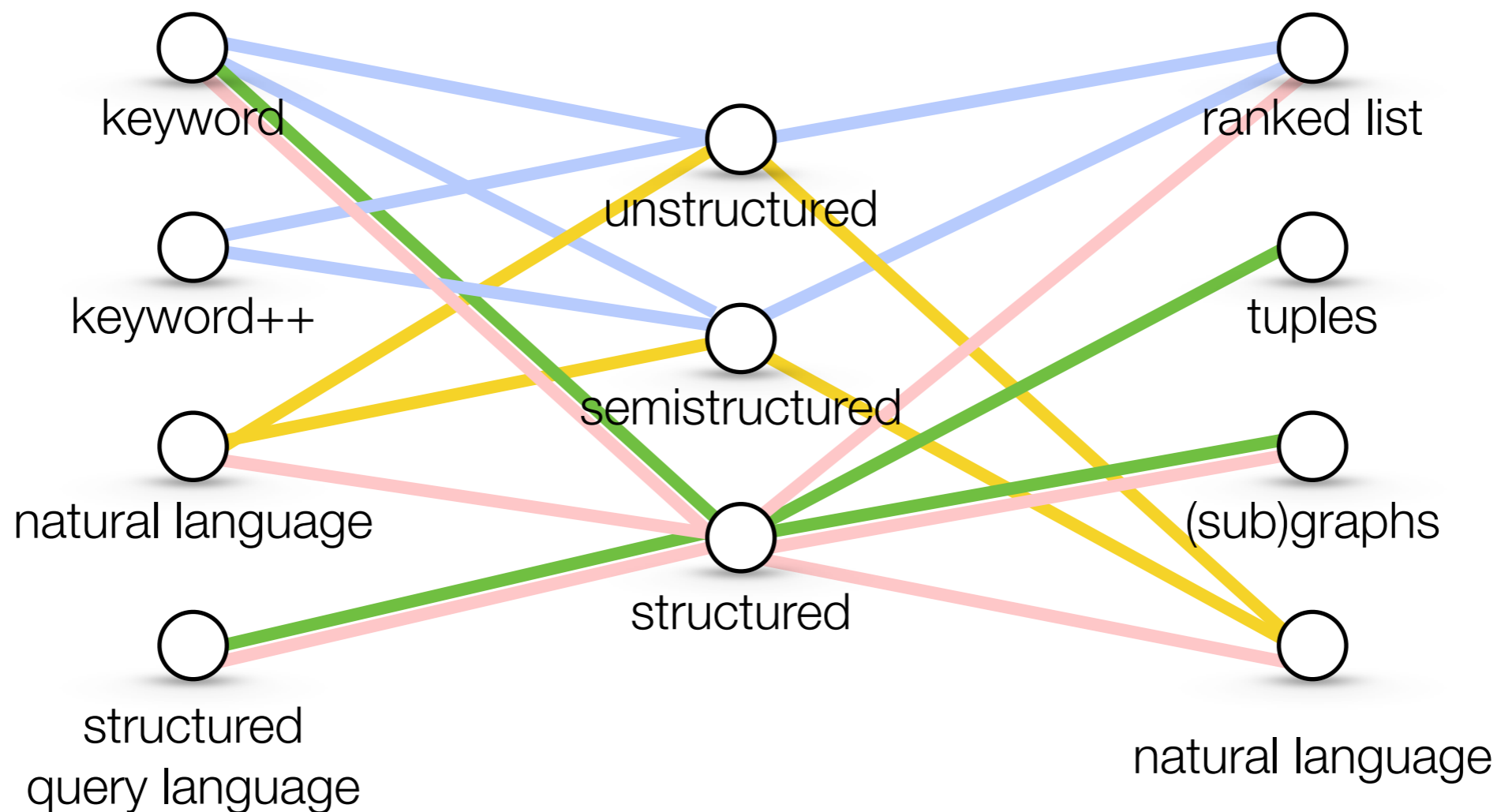
NLP

SW

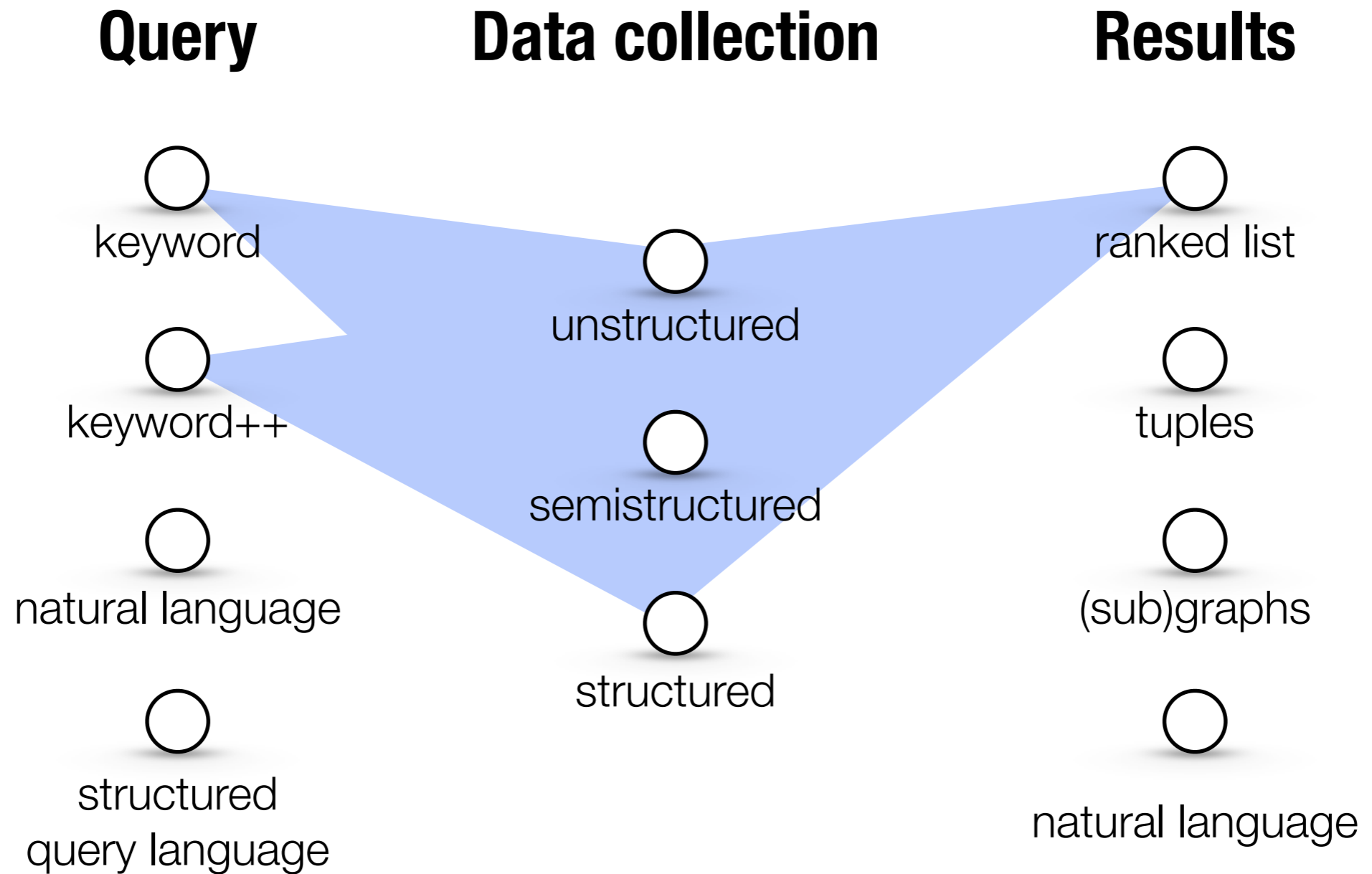
Query

Data collection

Results



# Our focus

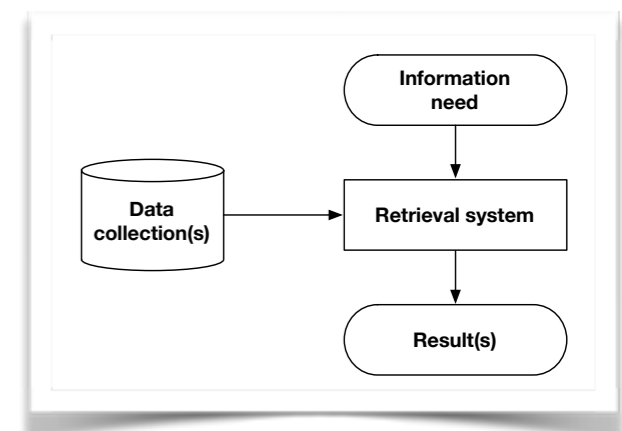




# Data collection

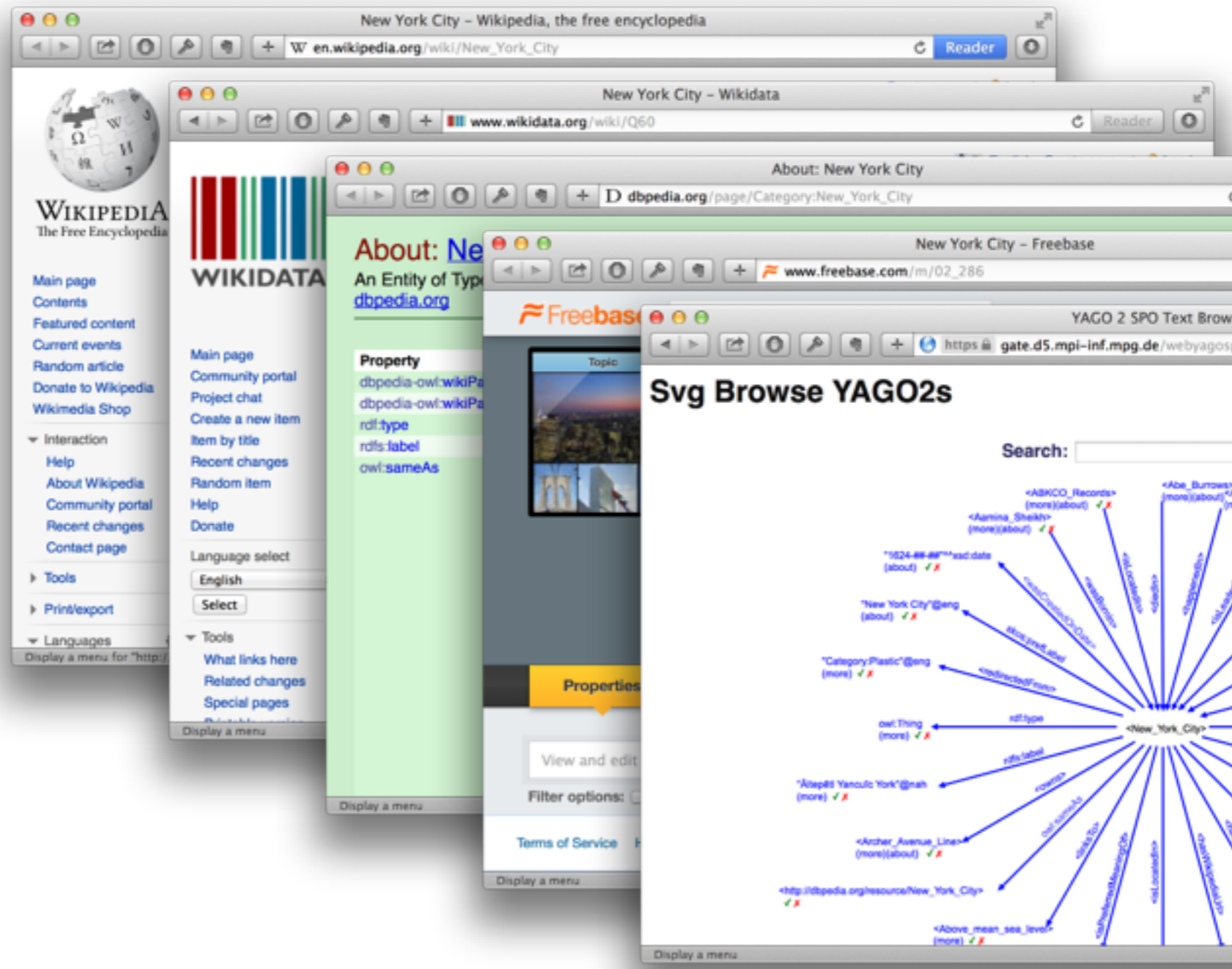
- Unstructured
  - Documents, web pages, snippets, ...
- Semistructured
  - XML, RDF, ...
- Structured
  - Relational DBs, RDF, ...

} Often organized around entities

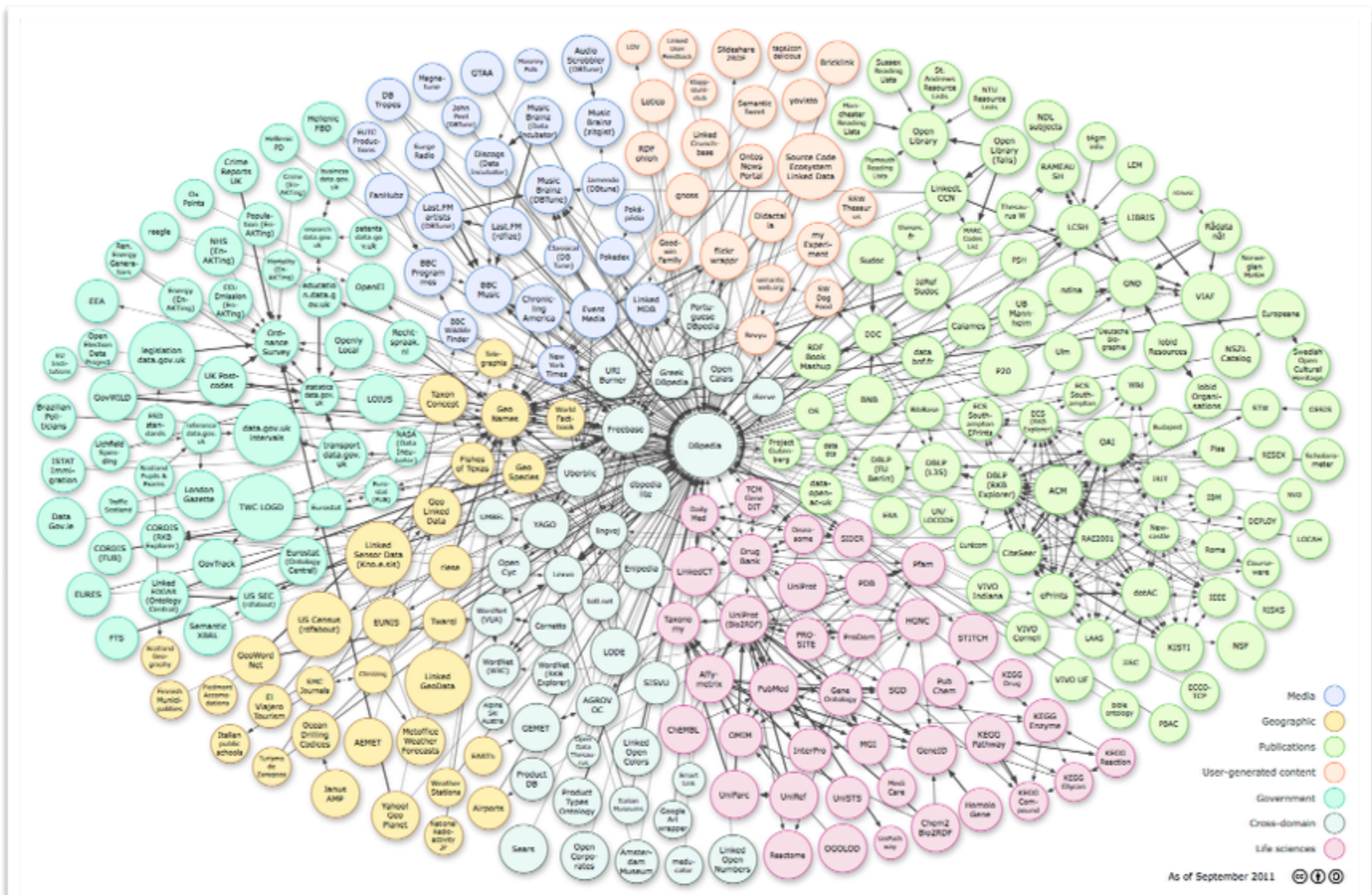


# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# Linking Open Data (LOD)?



# RDFa


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




[The Wire \(TV Series 2002–2008\) - IMDb](#)  
[www.imdb.com/title/tt0306414/](http://www.imdb.com/title/tt0306414/)   
★★★★★ 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](#).



[Logitech Revue 097855070906 | eBay](#)  
[www.ebay.com/ctg/Logitech-Revue-/97019743](http://www.ebay.com/ctg/Logitech-Revue-/97019743)   
★★★★★ 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**...



[Washington Huskies Mens Basketball tickets, and ... - Ticketmaster](#)  
[www.ticketmaster.ca/Washington-Huskies-Mens-Basketball-tickets/art...](http://www.ticketmaster.ca/Washington-Huskies-Mens-Basketball-tickets/art...)   
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](#)



[Recette Tarte Tatin - testée et approuvée - calorie et ... - iSaveurs](#)  
[www.isaveurs.com/recette/recette\\_tart...](http://www.isaveurs.com/recette/recette_tart...)   
45 mins  
il y a 1 jour – Recette de cuisine gourmande - Tarte Tatin - pour virginie63 :  
J'adooore j'adore, ...  
[http://www.isaveurs.com/recette/recette\\_tarte\\_tatin.php](http://www.isaveurs.com/recette/recette_tarte_tatin.php).

# Queries

- **Keyword queries**
  - Single-search-box paradigm
  - Typical web search queries
  - “Telegraphic”, i.e., neither well-formed nor grammatically correct
- **Keyword++ queries**
  - Augmented with context
    - form/facet-based input
    - location/date/TOD/...

# Example keyword+++ queries

The image shows a screenshot of the Amazon website's search interface. The search bar contains the text "kawasaki ac". Below the search bar, a dropdown menu displays search suggestions. The suggestions are:

- kawasaki accessories**
  - in Arts, Crafts & Sewing
  - in Automotive
  - in Tools & Home Improvement
  - in Electronics
- kawasaki accessories**
- kawasaki accessories motorcycle**
- kawasaki acoustic guitar**
- kawasaki actuator**

On the left side of the page, there are navigation menus for "Departments", "Motorcycle & Powersports", "Sports & Outdoors", and "Clothing & Accessories". In the center, there is a section for "kawasaki" with a "Related Searches" link and a "Showing 1 - 16 of 21" indicator. Below this, a green t-shirt with the Kawasaki logo is displayed.



john smith



Advanced

## SEARCH

Advanced >

All

**People**

More...

### Relationship

- All
- 1st Connections (0)
- 2nd Connections (0)
- Group Members (0)
- 3rd + Everyone Else (9)

### Location

- All
- United States (71)
- United Kingdom (9)
- Greater Los Angeles Area (7)
- Greater New York City A... (7)
- Australia (6)

+ Add

### Current Company

- All
- NHS (29)
- Tesco PLC (24)
- None (20)
- Information Limited (13)
- IBM (12)
- McDonald's Corporation (9)

+ Add

9 results for **john smith**

United Kingdom x

Current Company: McDonald's Corporation x

Reset



**john smith**

nothing  
Reading, United Kingdom · Accounting  
[Similar](#)

Connect



**John Smith**

Operations Supervisor at Rookery Foods  
Warrington, United Kingdom · Restaurants  
[Similar](#)

Connect

Current: Operations Supervisor at Rookery Foods T/A McDonald's



**John Smith**

Cashier at McDonald's  
Newcastle upon Tyne, United Kingdom · Restaurants  
[Similar](#)

Connect



**john smith**

chef at McDonalds  
Coventry, United Kingdom · Restaurants  
[Similar](#)

Connect



**John Smith**

Burger flipper at McDonald's  
London, United Kingdom · Restaurants  
[Similar](#)

Connect



**john smith**

Accountant at Mcdonalds  
Peterborough, United Kingdom · Food & Beverages  
[Similar](#)

Connect



**John Smith**

Manager at McDonald's  
London, United Kingdom · Restaurants  
[Similar](#)

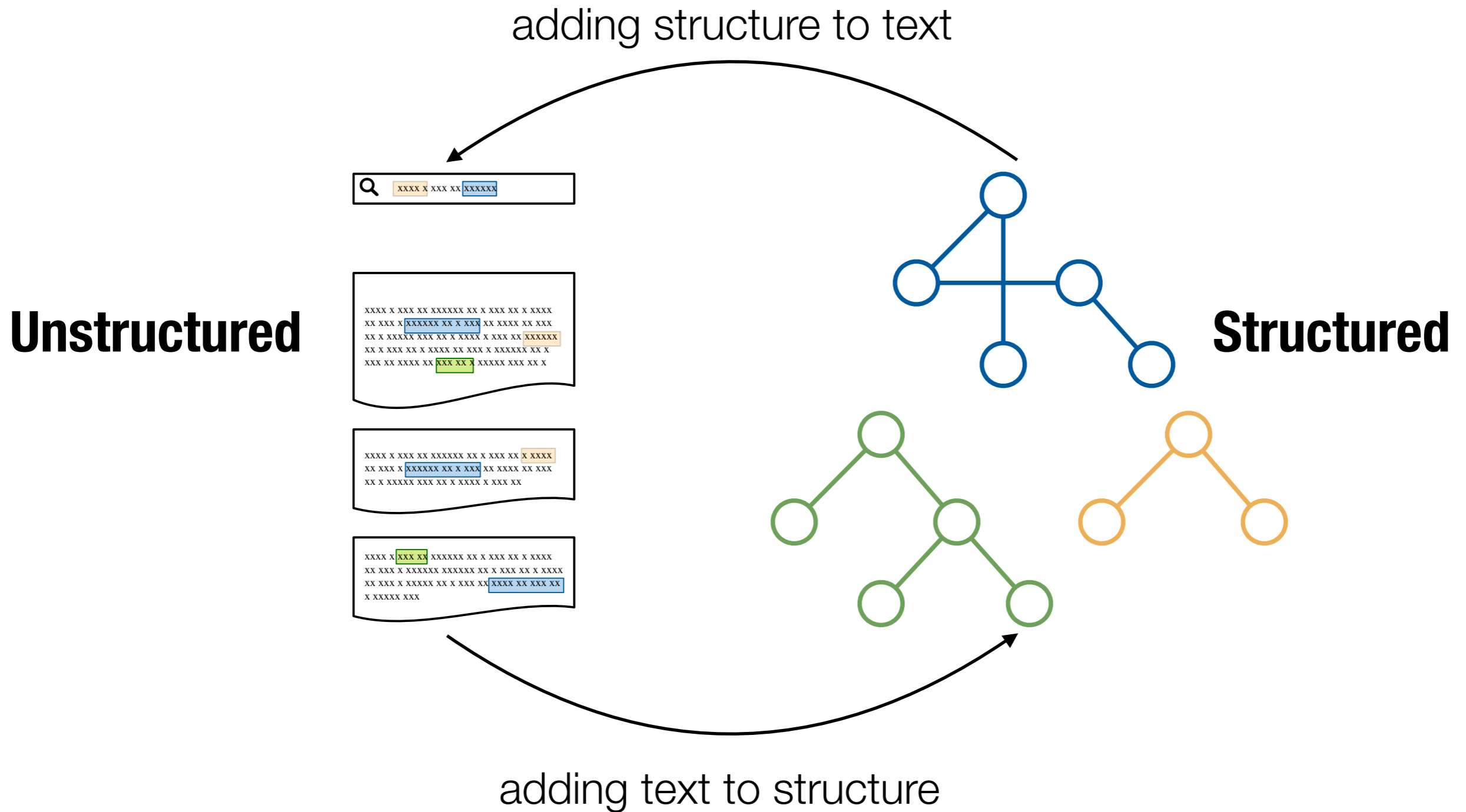
Connect

# Example keyword++ queries





# Interplay: (un)structured data



# Menu

- Introduction
- Part 1 – Entity Linking
- Part 2 – Entity Retrieval
- Part 3 – Semantic Search
- Wrap up

# Program

<b>10:00 - 10:30</b>	Welcome, introduction
<b>10:30 - 11:30</b>	Entity linking
<b>11:30 - 12:00</b>	Entity linking practical
<b>12:00 - 13:00</b>	Lunch
<b>13:00 - 14:00</b>	Entity retrieval
<b>14:00 - 14:30</b>	Entity retrieval practical
<b>14:30 - 15:00</b>	Coffee break
<b>15:00 - 15:30</b>	Semantic search
<b>15:30 - 16:00</b>	Wrap up, Q&A

# References

The screenshot shows the Mendeley Group page for 'Entity Linking and Retrieval'. The browser address bar shows the URL: [www.mendeley.com/groups/3339761/entity-linking-and-retrieval/](http://www.mendeley.com/groups/3339761/entity-linking-and-retrieval/). The page features the Mendeley logo and navigation tabs for 'Get Mendeley', 'What is Mendeley?', 'Papers', and 'Groups'. The group title is 'Entity Linking and Retrieval' with 113 papers and 33 members. A sidebar on the left includes 'Overview', 'Papers', and 'Members' tabs, along with a 'Feedback' button. The main content area is divided into 'Group activity' and 'About this group'. The 'Group activity' section lists recent additions by Daan Odijk and Edgar Meij, including papers like 'Active objects: Actions for entity-centric search' and 'Natural language questions for the web of data'. The 'About this group' section identifies the owner as Edgar Meij, a researcher at a non-academic institution, and describes the group as a public library of tutorials. It also shows 48 members and tags for 'entity linking' and 'entity retrieval'. A 'Sign up today - FREE' banner is visible at the bottom right.

<http://www.mendeley.com/groups/3339761/entity-linking-and-retrieval/> (or <http://bit.ly/ELR-bib>)