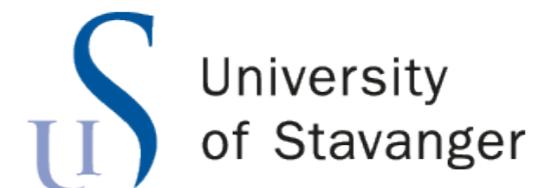


# Entity Linking and Retrieval for Semantic Search

**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**

**Yalspar**



**Donaldson**  
Filtration Solutions



**BEST BUY**

**Carlson Companies**

**AMS**  
Solutions for Life®



**Medtronic**  
Alleviating Pain - Restoring Health - Extending Life

**Thrivent Financial for Lutherans**

**MARVIN**  
Windows and Doors

**POLARIS**  
The Way Out.

**3M** Worldwide

**GENERAL MILLS**

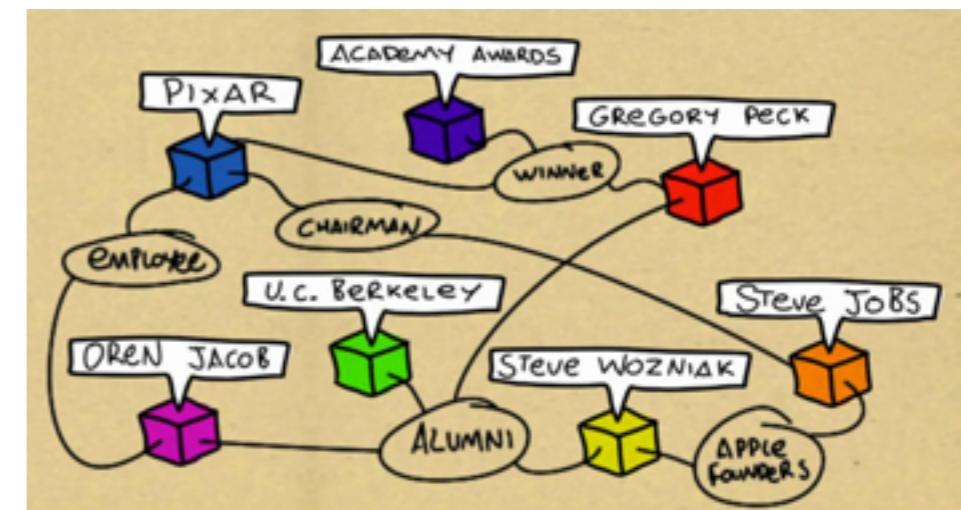


# Entity?

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

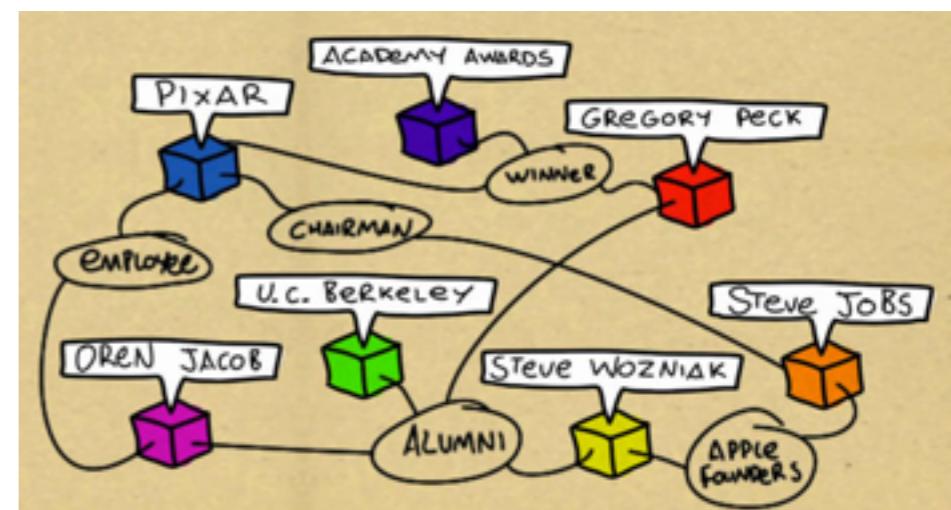
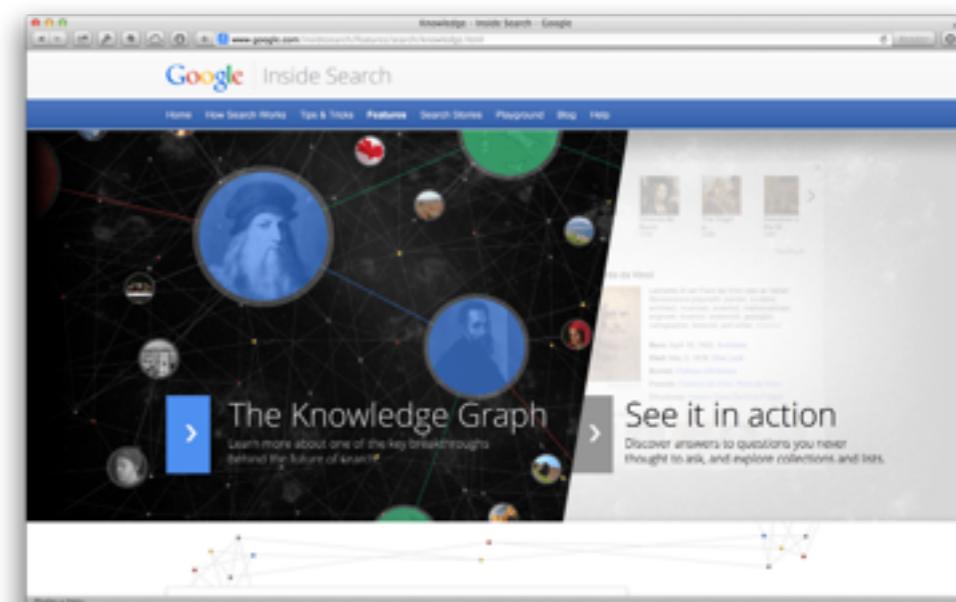
# What's so special about entities?

- 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-Sahaf](#) 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 montreal

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](https://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.ca/](http://www.tourisme-montreal.ca/) ▾ 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

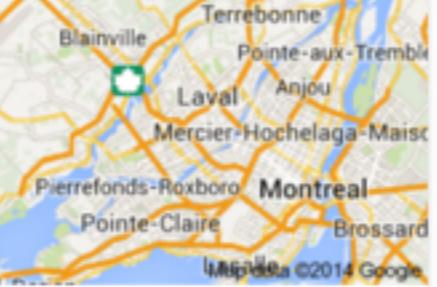
**bienvenue - welcome to montreal.com**  
[www.montreal.com/](http://www.montreal.com/) ▾  
montreal.com. Montreal City Weblog. Find Montreal Vacation Rentals on FlipKey. Old Montreal by Patrick Martineau. See your horizontal Montreal photo in this ...

**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. Blessed with one of the most...

**News for montreal**

**RAW: Sinkhole on De La Montagne**  
[CBC.ca](http://www.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](http://www.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

Reader Share

Google restaurants in montreal

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](http://www.kegsteakhouse.com)  
4.3 ★★★★★ 60 Google reviews · Google+ page

**Restaurant Kashima**  
[www.kashima.ca](http://www.kashima.ca)  
4.9 ★★★★★ 9 Google reviews

**O.Noir**  
[www.onoir.com](http://www.onoir.com)  
3.9 ★★★★★ 31 Google reviews

**Restaurant Trang An**  
[plus.google.com](http://plus.google.com)  
4.3 ★★★★★ 7 Google reviews

**L'Auberge Saint-Gabriel**  
[www.lesaint-gabriel.com](http://www.lesaint-gabriel.com)  
4.0 ★★★★★ 12 Google reviews · Google+ page

**Vago**  
[www.restaurantvago.com](http://www.restaurantvago.com)  
3 Google reviews

**Bistro On The Avenue**  
[www.bistroontheavenue.com](http://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

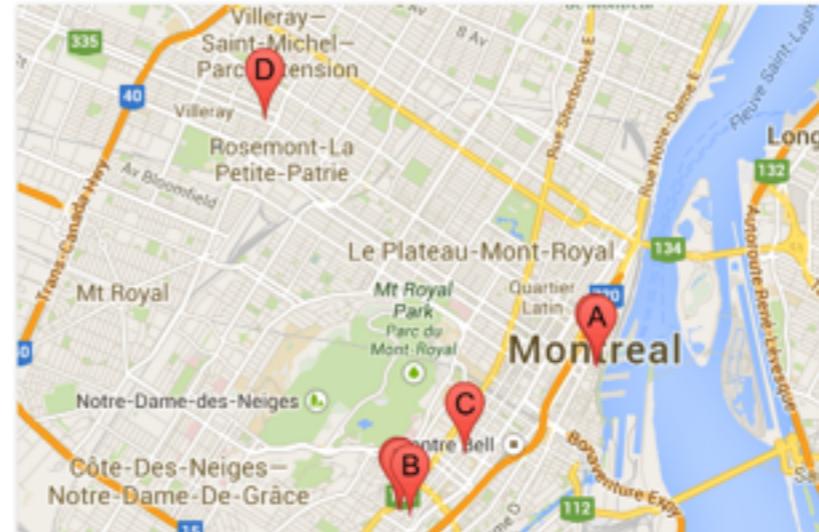
C 1631 Rue Sainte-Catherine Ouest  
Montréal, QC, Canada  
+1 514-937-9727

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 data ©2014 Google

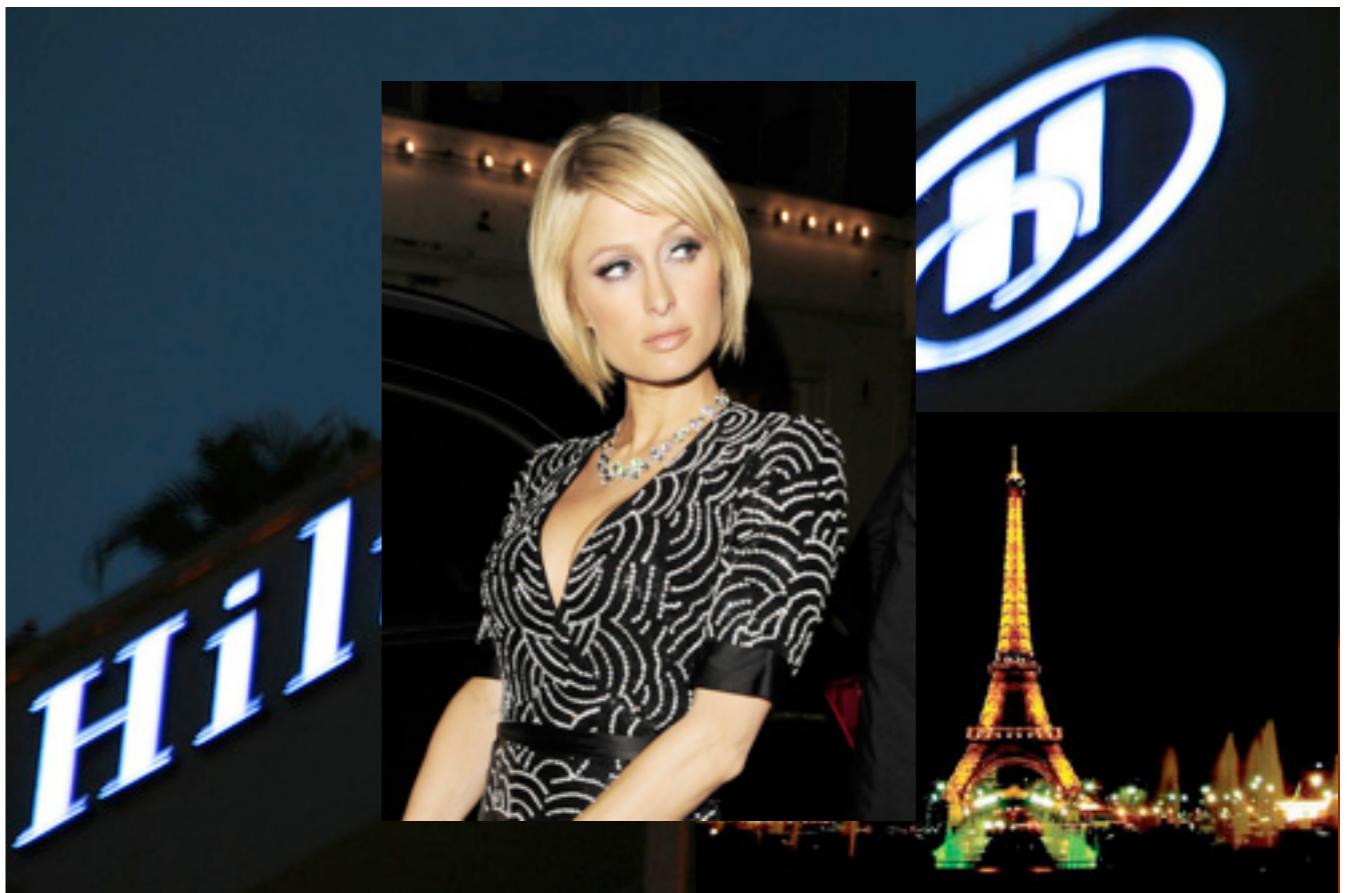
**restaurants in Montreal - TripAdvisor**  
[www.tripadvisor.com](http://www.tripadvisor.com) › Canada › Quebec › Montreal › TripAdvisor

# **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’



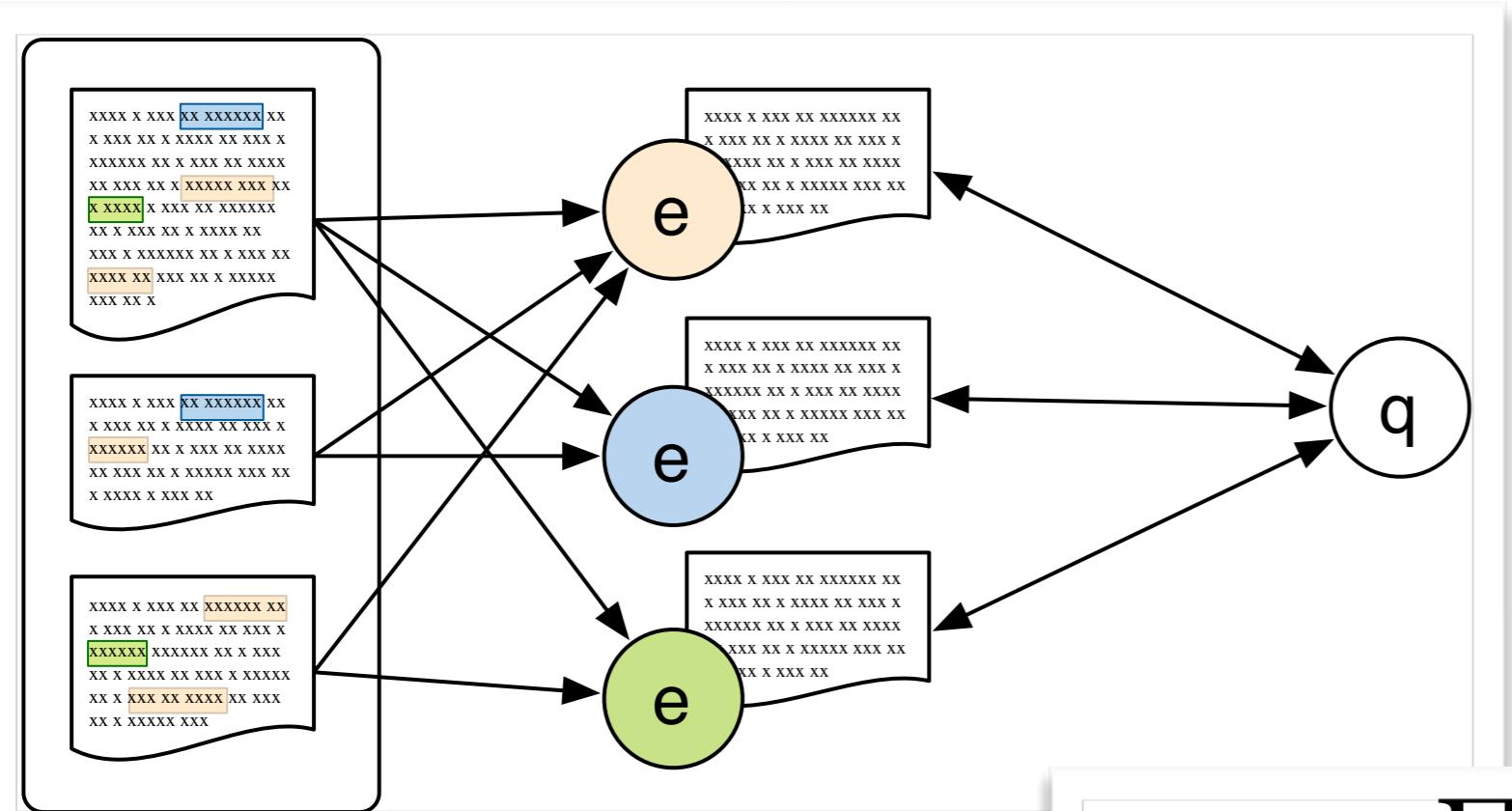
# Semantic search

- 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



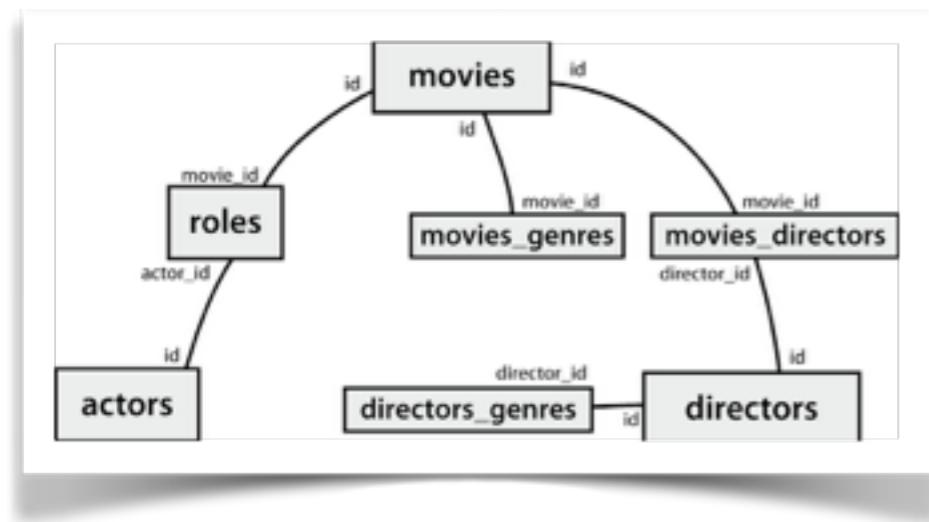
$$\begin{aligned}
 P(q|\theta_e) &= \prod_{t \in q} \underbrace{P(t|\theta_e)}_{(1 - \lambda)P(t|e) + \lambda P(t)}^{n(t,q)} \\
 &\quad \downarrow \\
 &\quad \sum_d P(t|d, e) P(d|e)
 \end{aligned}$$

# NLP

- Question answering, relationship extraction



# Databases



SPARK  
searching, probing & ranking

IMDB

Results 1 - 10 for "tom cruise nicole kidman"

tom cruise nicole kidman  Preferences

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

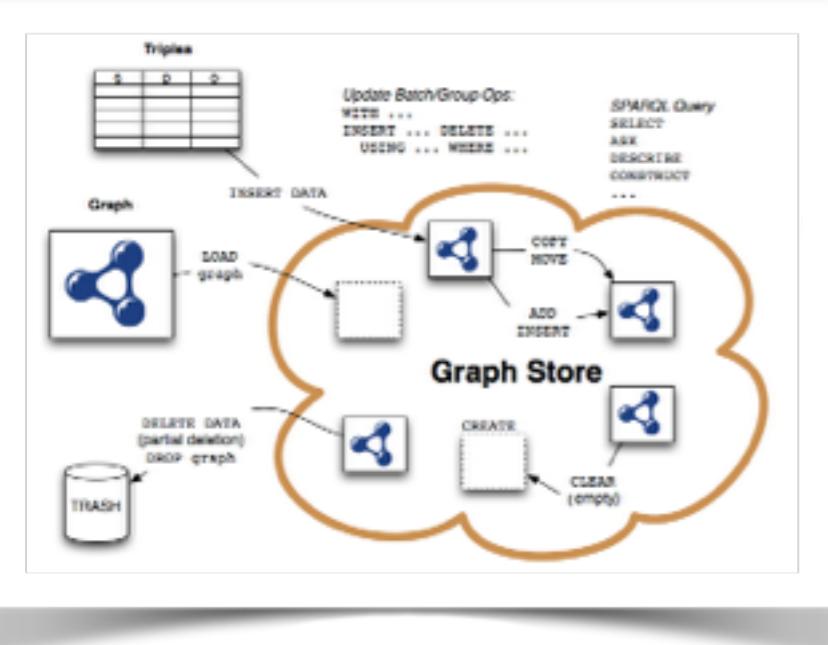
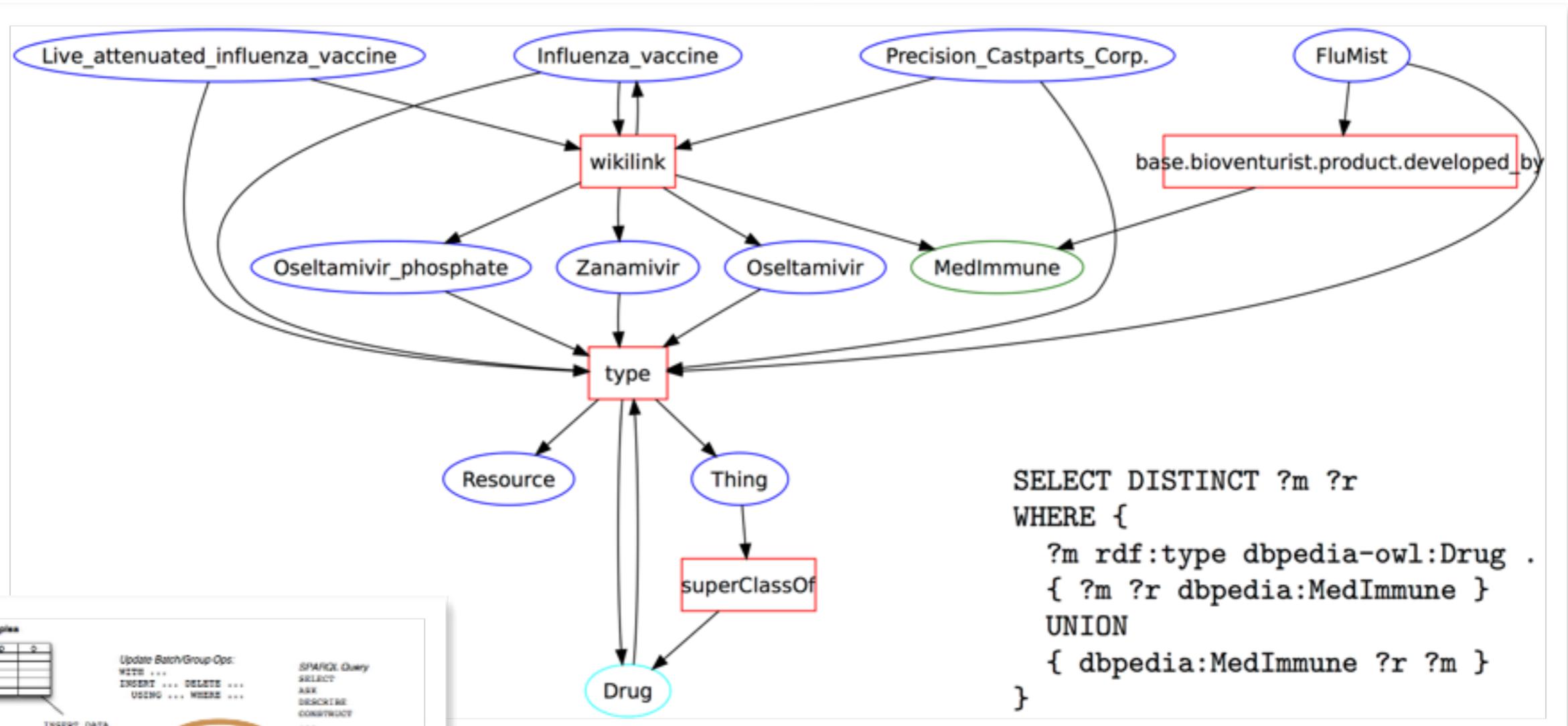
actorplay : Charactor: Himself (Narrator) Id: 841844

movies : Name: Stanley Kubrick: A Life in Pictures (2001) Id: 386247

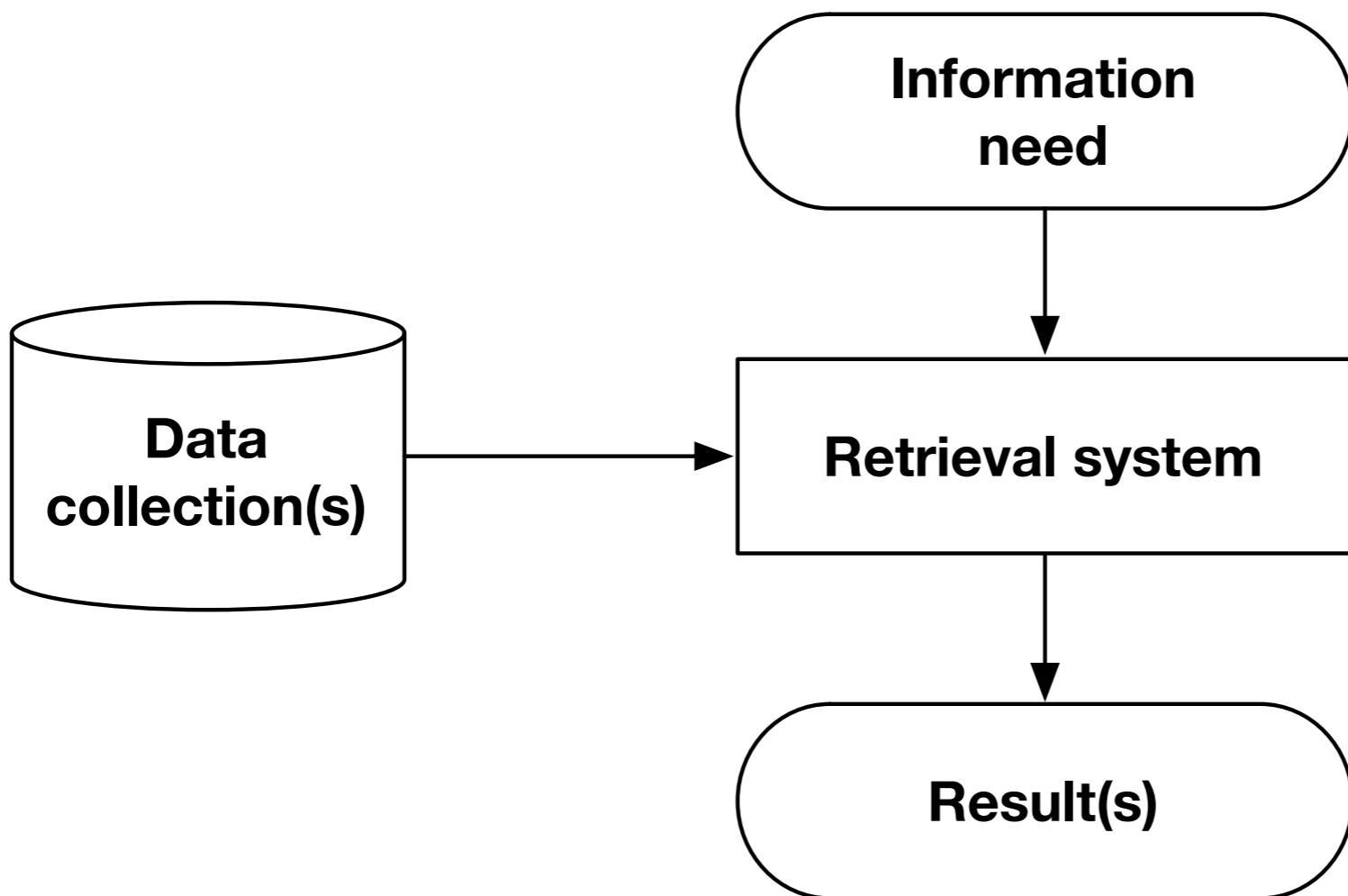
actressplay : Charactor: Herself Id: 1052946

actresses : Name: **kidman, nicole** Id: 206700

# Semantic Web

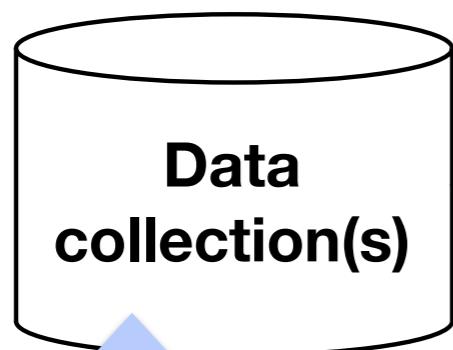


# Birds-eye view



# Many ways to express

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



# Different types of data

- Unstructured
- Semistructured
- Structured

Information  
need

Retrieval system

Result(s)

# Result format

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

# The possibilities are endless

IR

DB

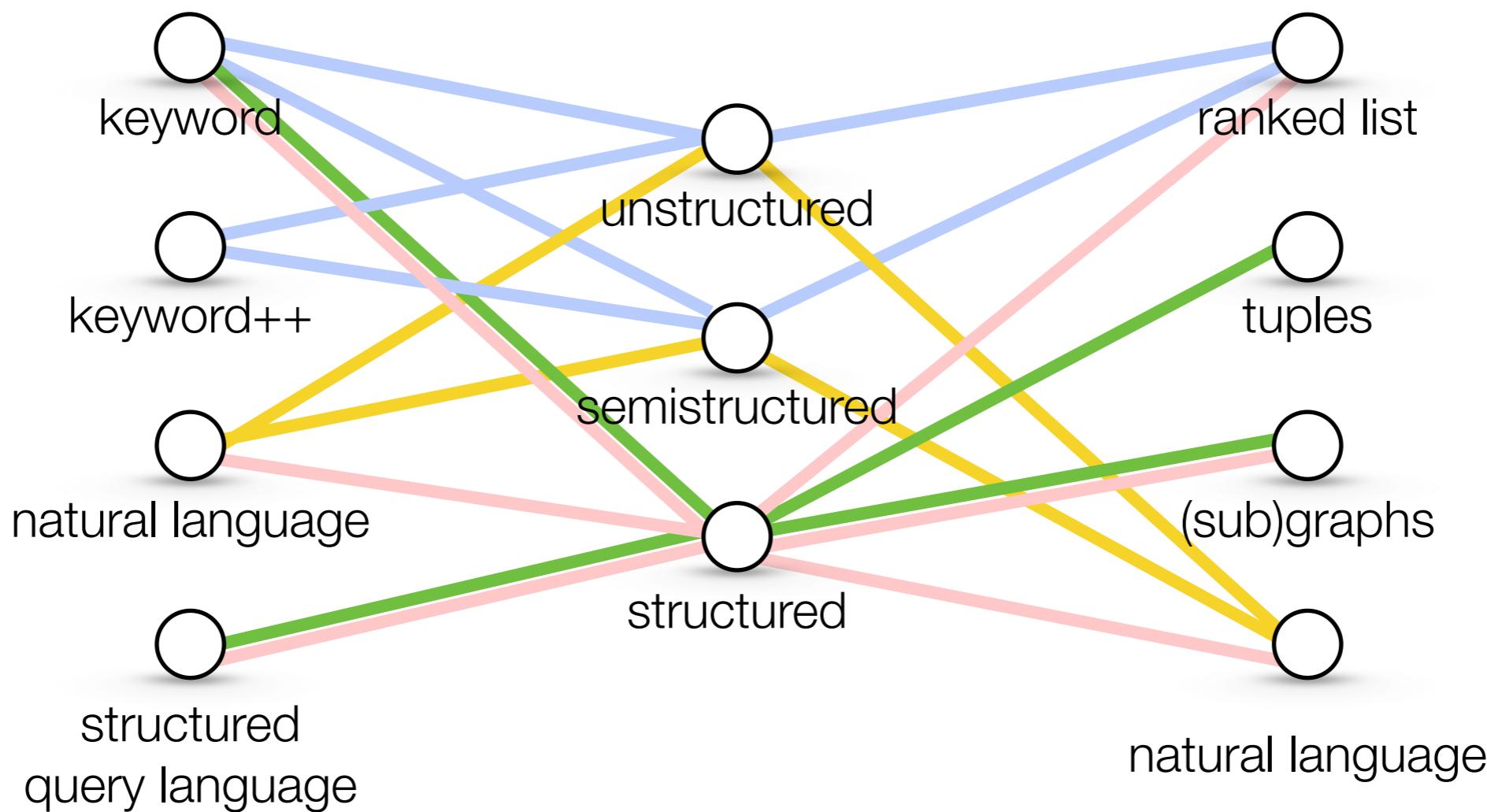
NLP

SW

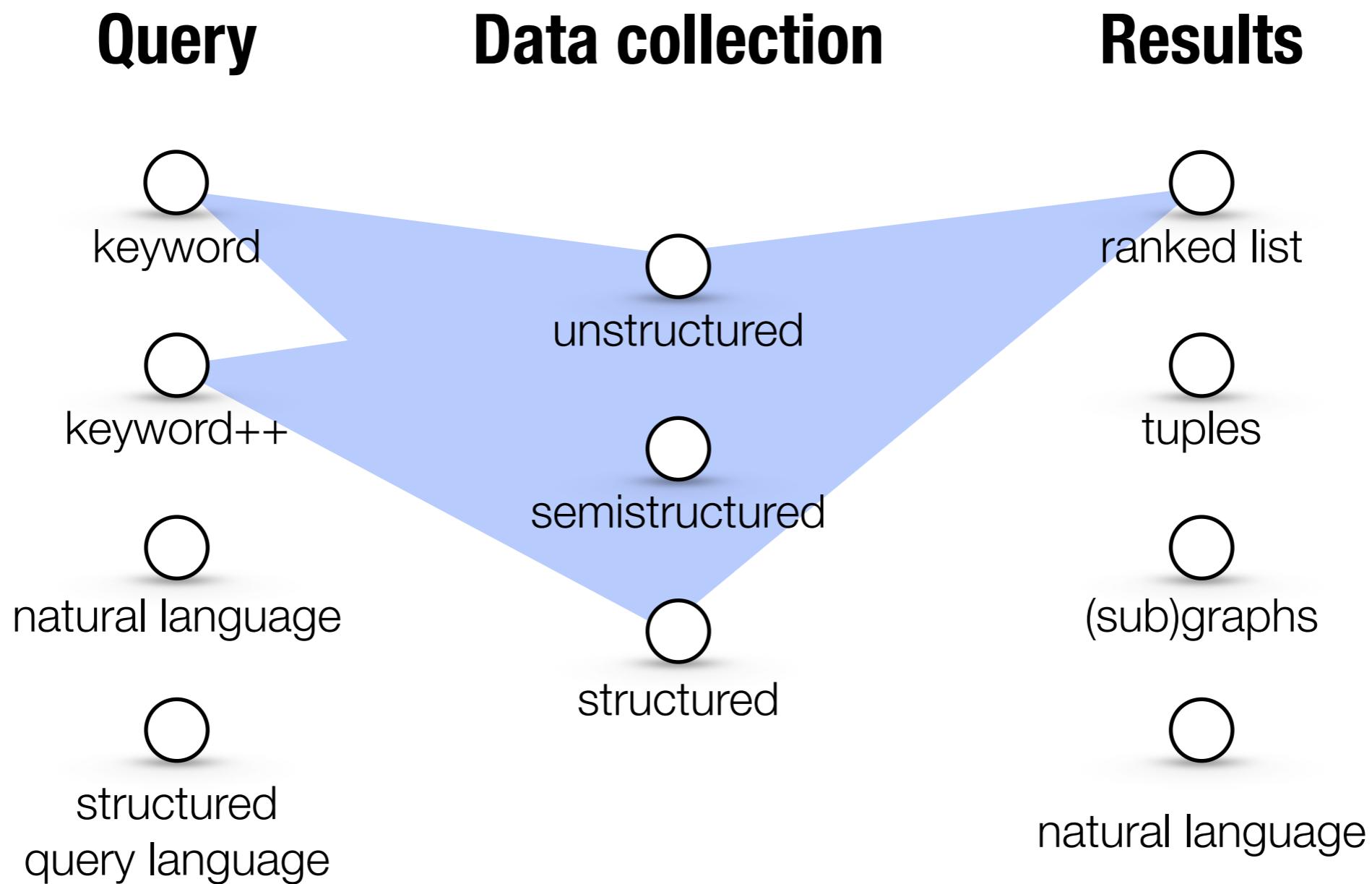
**Query**

**Data collection**

**Results**

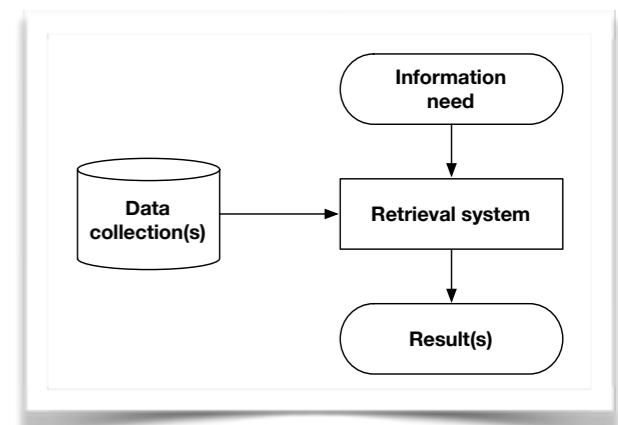


# Our focus



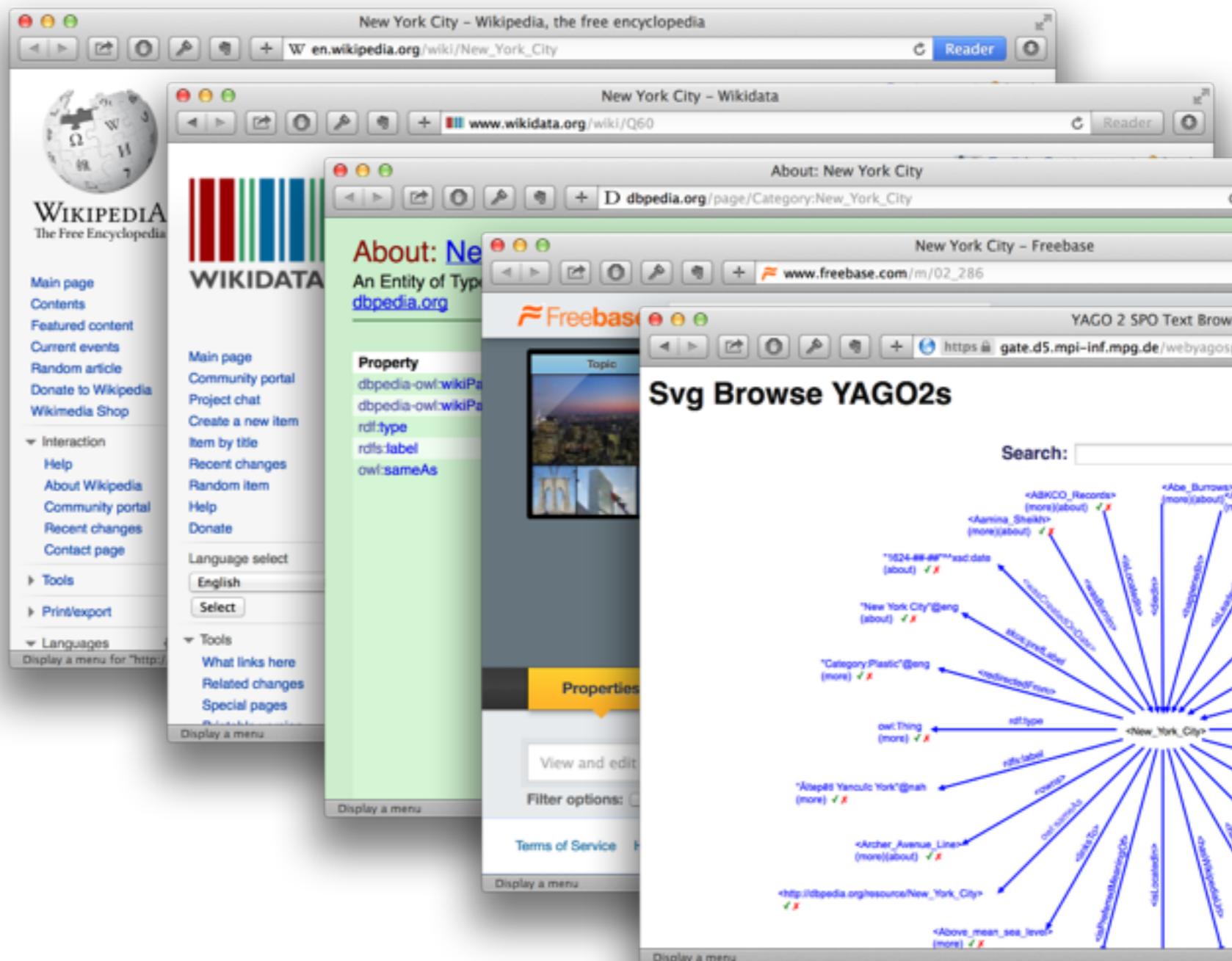
# Data collection

- Unstructured
    - Documents, web pages, snippets, ...
  - Semistructured
    - XML, RDF, ...
  - Structured
    - Relational DBs, RDF, ...
- Often organized around entities

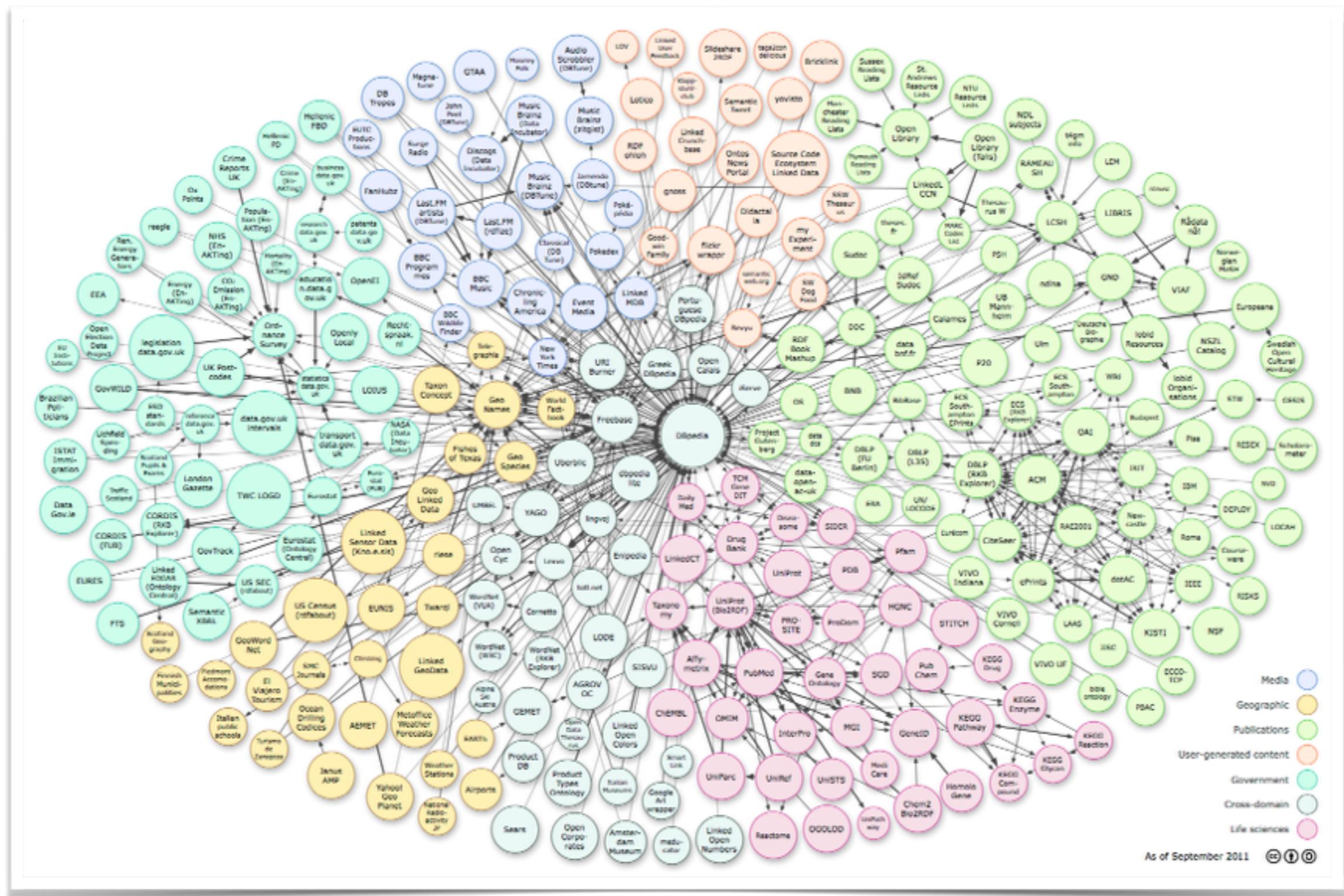


# Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



# Linking Open Data (LOD)?



# RDFa

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

The image displays four separate web snippets illustrating the use of RDFa:

- IMDb Snippet:** A snippet from the IMDb page for "The Wire (TV Series 2002–2008)". It shows the title, URL (www.imdb.com/title/tt0306414/), rating (9.6/10 - 44,375 votes), and a brief description of the show's plot.
- eBay Snippet:** A snippet from an eBay listing for a Logitech Revue media streamer. It shows the item title, URL (www.ebay.com/ctg/Logitech-Revue-/97019743), rating (★★★★★ from 51 users - \$99.99 to \$190.00), and a brief description of the product.
- Ticketmaster Snippet:** A snippet from a Ticketmaster page for Washington Huskies Mens Basketball tickets. It shows the event title, URL (www.ticketmaster.ca/Washington-Huskies-Mens-Basketball-tickets/art...), results count (Results 1 - 10 of 19), and a list of dates and locations.
- iSaveurs Snippet:** A snippet from an iSaveurs recipe page for "Recette Tarte Tatin". It shows the title, URL (www.isaveurs.com/recette/recette\_tart...), preparation time (45 mins), a thumbnail image of the dish, and a short description.

# 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

amazon Try Prime Krisztian's Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search All Departments kawasaki ac Go

Search suggestions

**"kawasaki"**  
Related Searches: [ka](#)

Showing 1 - 16 of 21



**kawasaki accessories**  
in Arts, Crafts & Sewing  
in Automotive  
in Tools & Home Improvement  
in Electronics

**kawasaki accessories**  
**kawasaki accessories motorcycle**  
**kawasaki acoustic guitar**  
**kawasaki actuator**

Cotton. Accessories: See



john smith



Advanced

## SEARCH

Advanced &gt;

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 ×

Current Company: McDonald's Corporation ×

Reset

**john smith**

nothing

Reading, United Kingdom · Accounting

[Similar](#)

Connect

**John Smith**

Operations Supervisor at Rookery Foods

Warrington, United Kingdom · Restaurants

[Similar](#)

Connect

**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 McDonald's

Peterborough, United Kingdom · Food &amp; Beverages

[Similar](#)

Connect

**John Smith**

Manager at McDonald's

London, United Kingdom · Restaurants

Connect

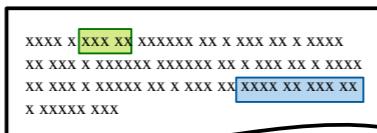
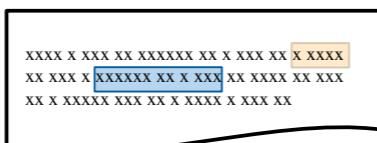
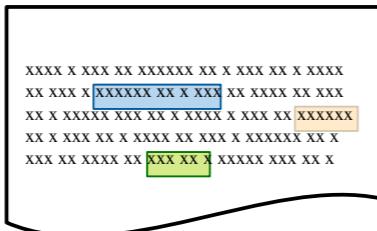
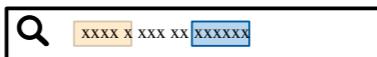
# Example keyword++ queries



# Interplay: (un)structured data

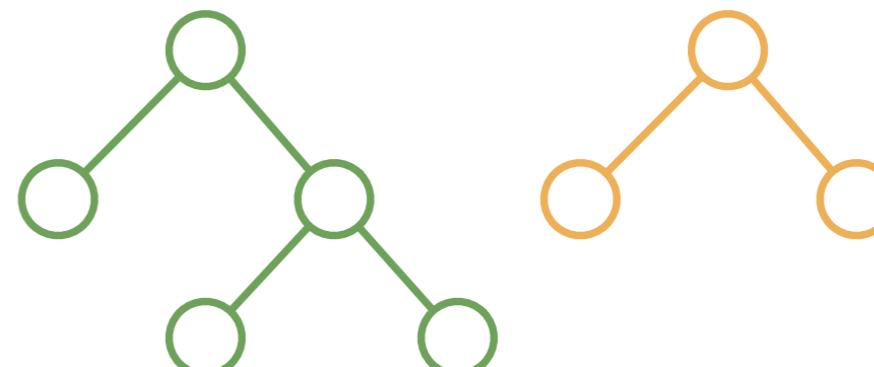
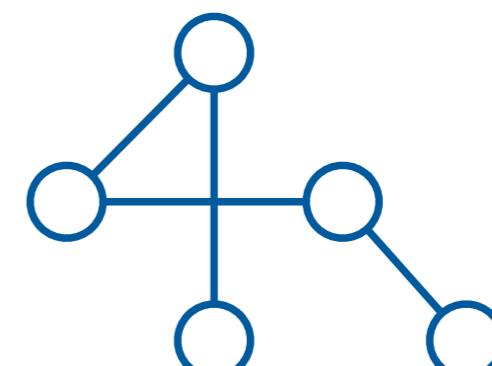
**Unstructured**

adding structure to text



**Structured**

adding text to structure



# Menu

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

---

See <http://ejmeij.github.io/entity-linking-and-retrieval-tutorial/> (or <http://bit.ly/ELR-slides>) for the slides.

# 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 a web browser window displaying the Mendeley Group page for 'Entity Linking and Retrieval'. The URL in the address bar is [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 links for 'Get Mendeley', 'What is Mendeley?', 'Papers', and 'Groups'. A search bar and a 'Sign up & Download' button are also present. The main content area displays the group's activity feed, which includes posts from Daan Odijk, Edgar Meij, and others. The group is owned by Edgar Meij, a researcher at a non-academic institution, and is associated with Computer and Information Science. The group has 48 members and focuses on entity linking and retrieval. A sign-up banner at the bottom encourages users to join for free.

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