

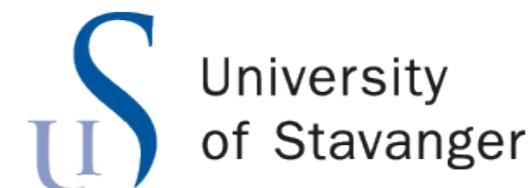
February 24-28

Entity Linking and Retrieval for Semantic Search

Edgar Meij – @edgarmeij
Yahoo Labs



Krisztian Balog – @krisztianbalog
University of Stavanger



Daan Odijk – @dodijk
University of Amsterdam





Minnesota Children's Museum

Deloitte.

GUIDANT

NORTHWEST AIRLINES

ECOLAB

Yalspar



Donaldson
Filtration Solutions



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

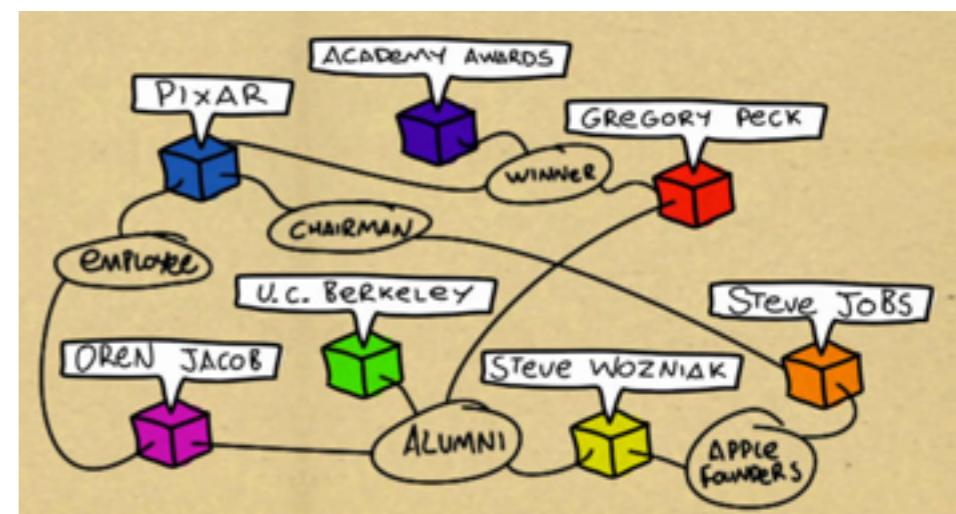


Entity?

- 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



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?

nyc - Google Search

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

Google nyc

Web Maps Images News Videos More Search tools Sign In

About 104,000,000 results (0.32 seconds)

[Welcome to NYC.gov | City of New York](#)
www1.nyc.gov/ ▾ New York City ▾
The official website of the City of New York. Find information about important alerts, 311 services, news, programs, events, government employment, the office of ...
[Jobs - General Information - 311 - Office of the Mayor](#)

[New York City - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/New_York_City ▾ Wikipedia ▾
New York is the most populous city in the United States and the center of the New York metropolitan area, one of the most populous urban agglomerations in the ...

[The Official New York City Guide to NYC Attractions, Dining, Hotels ...](#)
www.nycgo.com/ ▾ New York City ▾
Visit NYCgo for official NYC information on travel, hotels, deals and offers like Restaurant Week, and the best restaurants, shops, clubs and cultural events.

[News for nyc](#)

 Tax break helps NYC lure 'The Tonight Show' home
Boston.com - 5 hours ago
NEW YORK (AP) — 'The Tonight Show' made its return to New York City with a splashy opening sequence showcasing Grand Central ...
San Francisco C...

Man did 23 years for NYC murder he didn't commit, gets \$6.4M
CBS News - 1 day ago

NYC attorneys go after Florida pizza chain for using NYPD logo without visible...
New York Daily News - 19 hours ago

[NYC.com | New York Hotels - Broadway Tickets - New York ...](#)
www.nyc.com/ ▾
Save more with NYC.com, the trusted New York City experts. Book New York hotels and lodging, buy New York Broadway tickets and event tickets, reserve New ...

 Map data ©2014 Google

New York City

New York is the most populous city in the United States and the center of the New York metropolitan area, one of the most populous urban agglomerations in the world. [Wikipedia](#)

Weather: 42°F (6°C), Wind E at 2 mph (3 km/h), 57% Humidity
Local time: Sunday 8:06 AM
Population: 8.337 million (2012)
Colleges and Universities: [New York University](#), More

Upcoming events

New York Rangers vs. Boston Mar 2 The New York Times Travel Sh... Feb 28 Flogging Molly Feb 27 Book of Mormon Feb 23 STRIVE PREPARED TO WORK DETERMINED TO SUCCEED Healthcare Careers Advisory... Feb 27

Points of interest

Entity Retrieval?

spanish restaurants in nyc - Google Search

<https://www.google.com/#q=spanish+restaurants+in+nyc>

spanish restaurants in nyc

Google Sign In

Web Maps Images Shopping News More Search tools

Spanish restaurants near New York, NY

Price, Rating... ▾

Image	Rating	Reviews	Name	Cuisine	Price Range
	3.8 ★★★★☆	40 reviews	Spain Restaurant	Spanish	\$ - \$• Spanish
	4.1 ★★★★☆	317 reviews	Txikito	Spanish	z - \$\$\$
	3.9 ★★★★☆	630 reviews	Sevilla Restaurant & Bar	Spanish	z - \$\$ - Spanish
	3.8 ★★★★☆	99 reviews	Meson Sevilla	Spanish	\$ - Spanish
	4.1 ★★★★☆	1,251 reviews	Casa Mono	Spanish	z - \$\$\$ - Spanish
	3.9 ★★★★☆	751 reviews	Tia Pol	Spanish	z - \$\$\$ - Spanish
	4.0 ★★★★☆	1,114 reviews	Boqueria	Spanish	z - \$\$ - Spanish
	4.1 ★★★★☆	138 reviews	Tertulia	Spanish	z - \$\$ - Spanish
	4.4 ★★★★☆	466 reviews	Degustation	Spanish	z - \$\$\$ - Spanish
	4.0 ★★★★☆	196 reviews	El Charro Espanol	Spanish	\$\$\$ - Spanish
	4.1 ★★★★☆	492 reviews	La Fonda	Spanish	z - \$\$\$ - S

Spanish Restaurants in New York City | Find the Best Spanish Fo...

www.zagat.com/c/new-york-city-ny/spanish-restaurants Zagat Survey

30+ items - Discover the best Spanish restaurants in New York City with ...

Degustation. Spanish • East Village. 20 \$84. 24

Solera Restaurant. Spanish • East 50s. 20 \$55. 23

[100 best New York restaurants: Spanish food and tapas - Time Out](http://www.timeout.com/newyork/restaurants/100-best-new-york-restaurants-s...)

www.timeout.com/newyork/restaurants/100-best-new-york-restaurants-s...

Funky Basque tapas, boutique jamon and modern Iberian cuisine are the draws at these top spots, the best New York restaurants for a lusty Spanish spread.

Tertulia - Boqueria - Txikito - Salinas

[Spanish Restaurants - New York City, Menus, Ratings, Reviews ...](http://www.menupages.com/restaurants/all-areas/all-neighborhoods/spanish/)

www.menupages.com/restaurants/all-areas/all-neighborhoods/spanish/

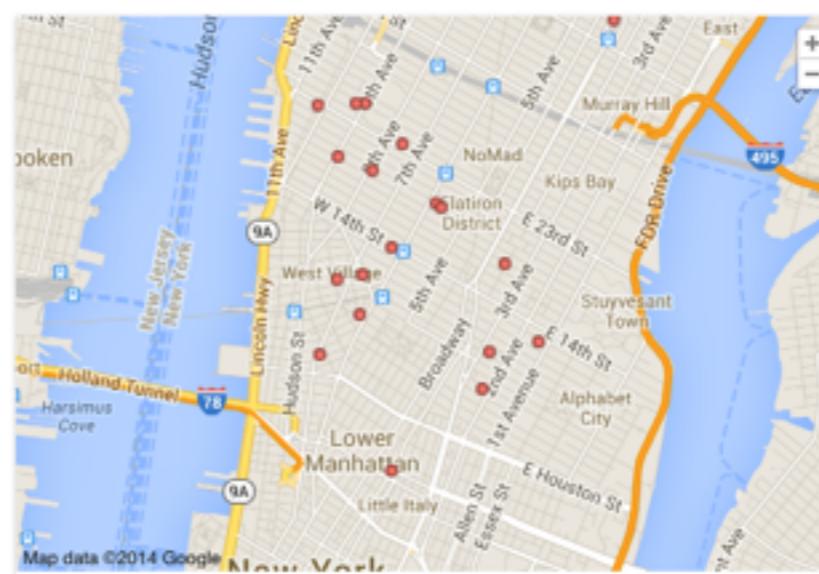
Results 1 - 100 of 129 - New York City Restaurants: Spanish. Filter by Meal:

1 100 Montaditos 176 Bleecker St | Btwn Sullivan & Mac Dougal St 2.
2 Alcala 246 E 44th St | Btwn 2nd & 3rd Ave 15.

[10 Classic Spanish Restaurants to Try Before You Die - Eater Maps ...](http://ny.eater.com/archives/2013/04/spanish_restaurants_map.php)

ny.eater.com/archives/2013/04/spanish_restaurants_map.php

Apr 12, 2013 - Here's a guide to 10 classic Spanish restaurants in New York City. These aren't the modern imports from Andalucía, the sleek tapas bars, or the



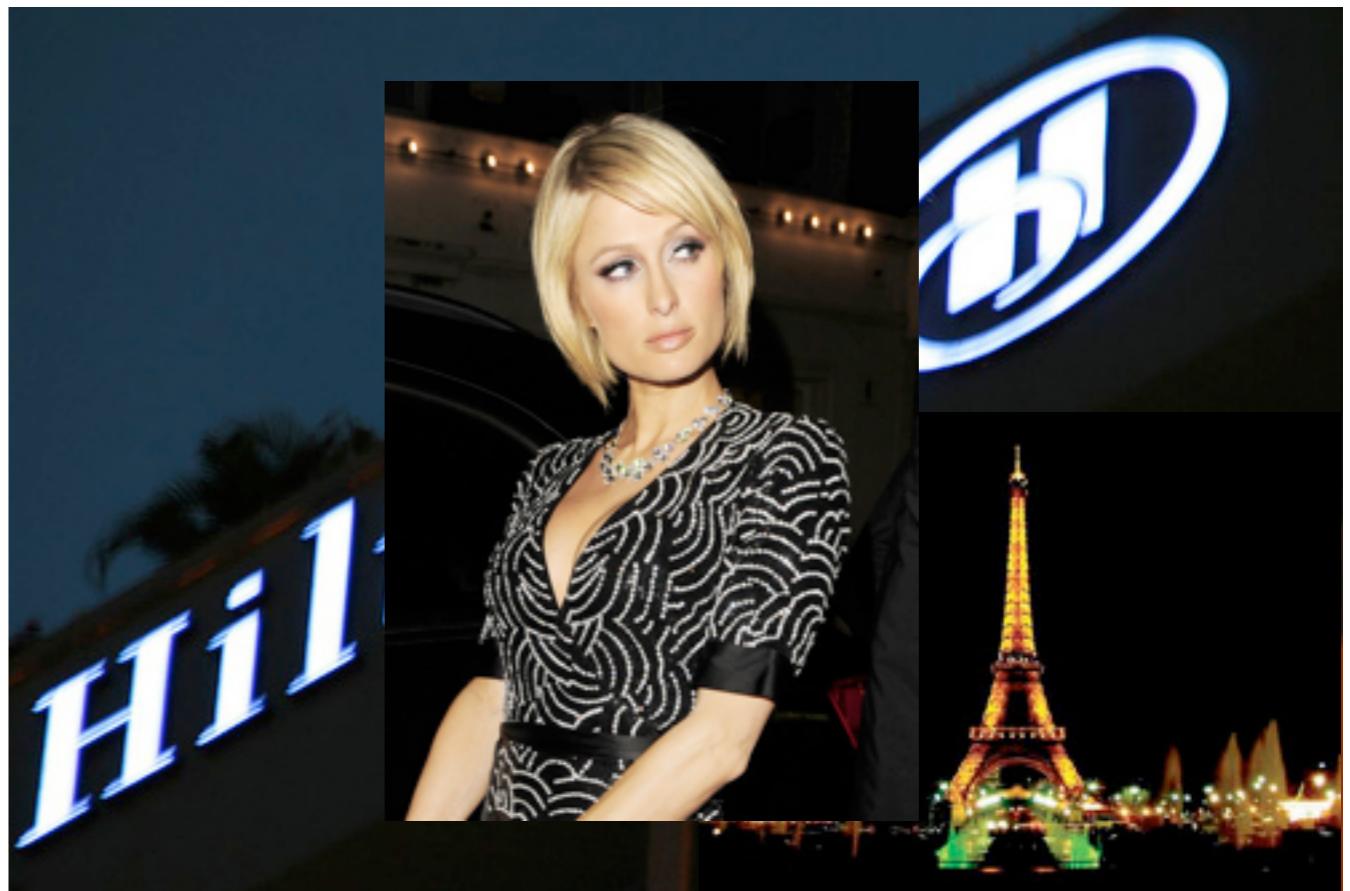
Map for spanish restaurants in nyc

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’



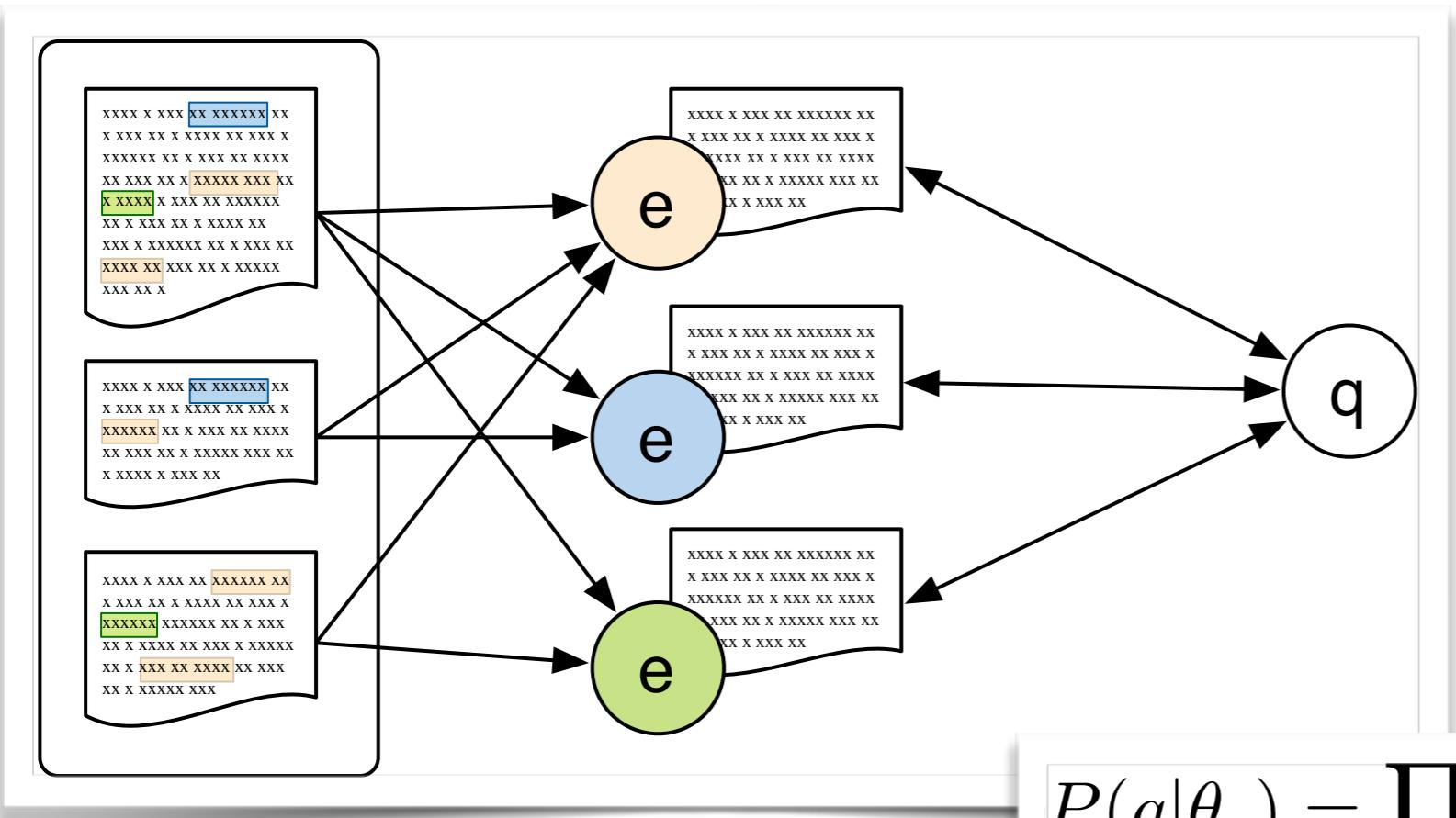
Semantic search

- 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 NYC?”
- Entity/Attribute/Relationship retrieval
 - + social, + personal
 - + (hyper)local

Semantic search

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

IR



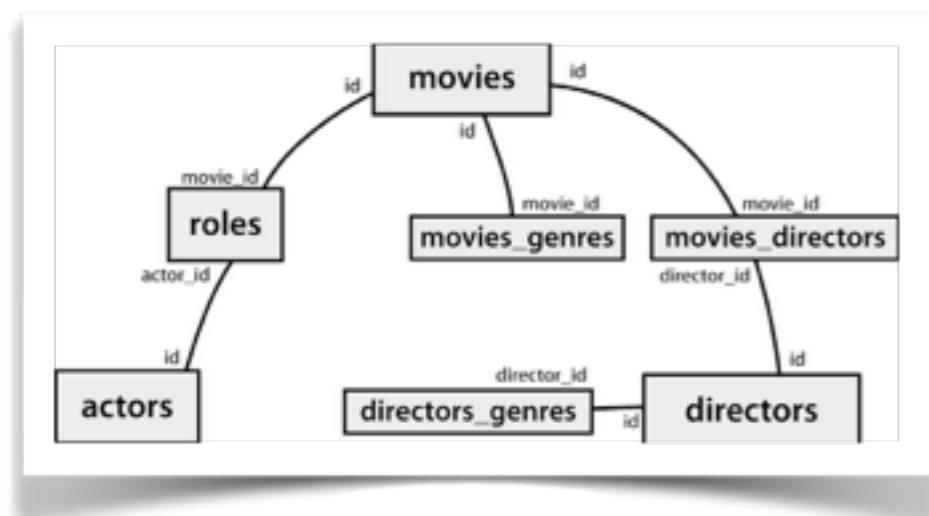
$$\begin{aligned}
 P(q|\theta_e) &= \prod_{t \in q} \underbrace{P(t|\theta_e)}_{(1 - \lambda)P(t|e) + \lambda P(t)}^{n(t,q)} \\
 &\quad \downarrow \\
 &\quad \sum_d P(t|d, e) P(d|e)
 \end{aligned}$$

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

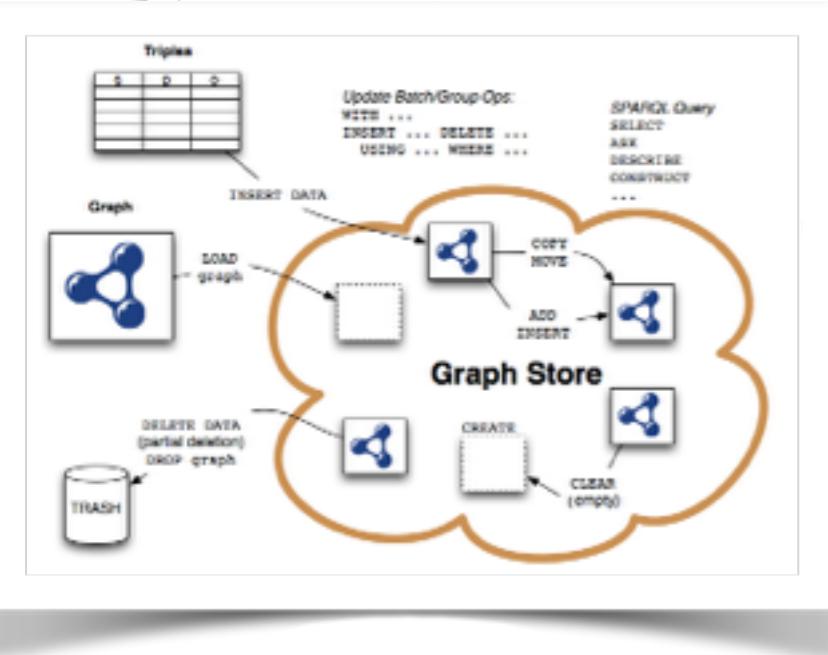
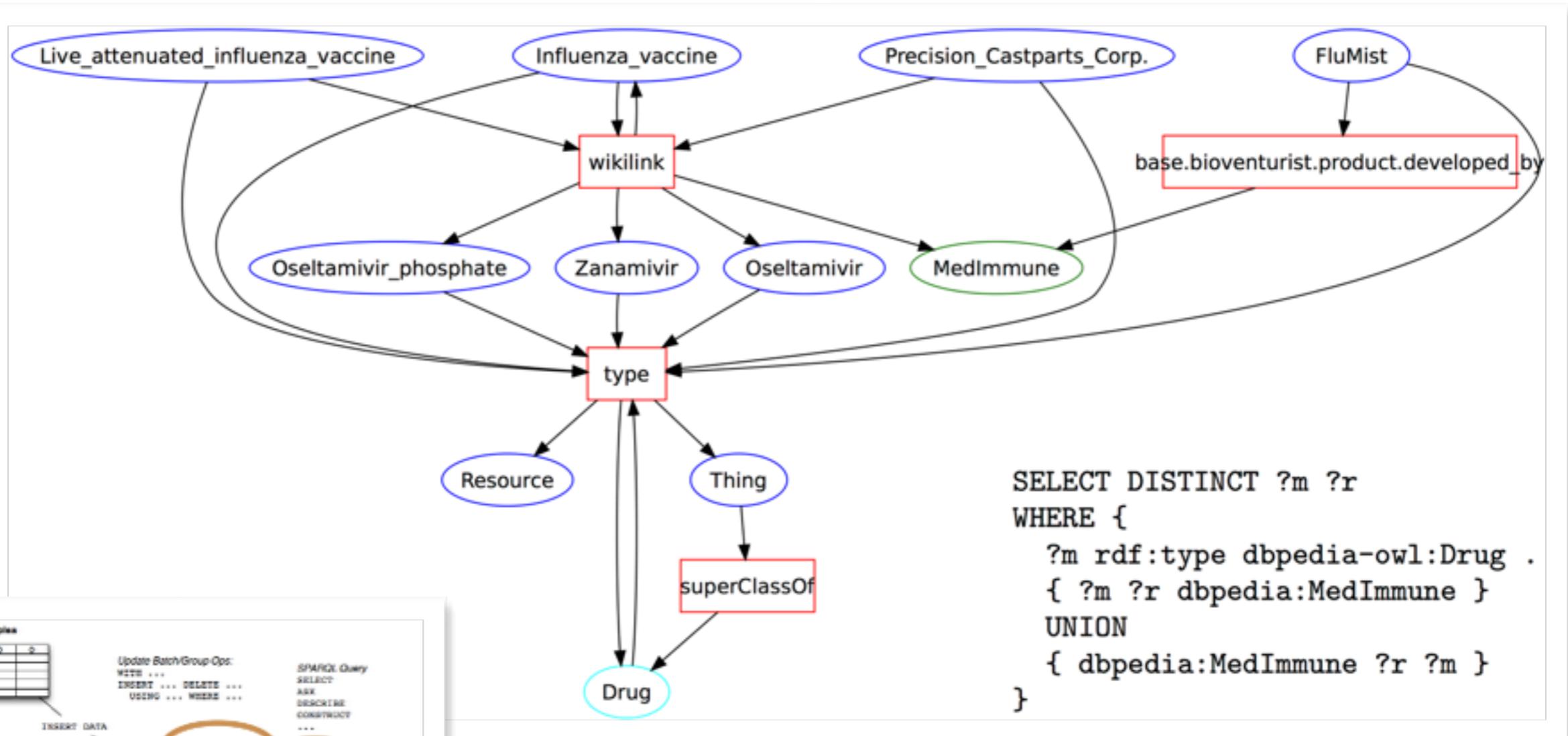
actorplay : Charactor: Himself (Narrator) Id: 841844

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

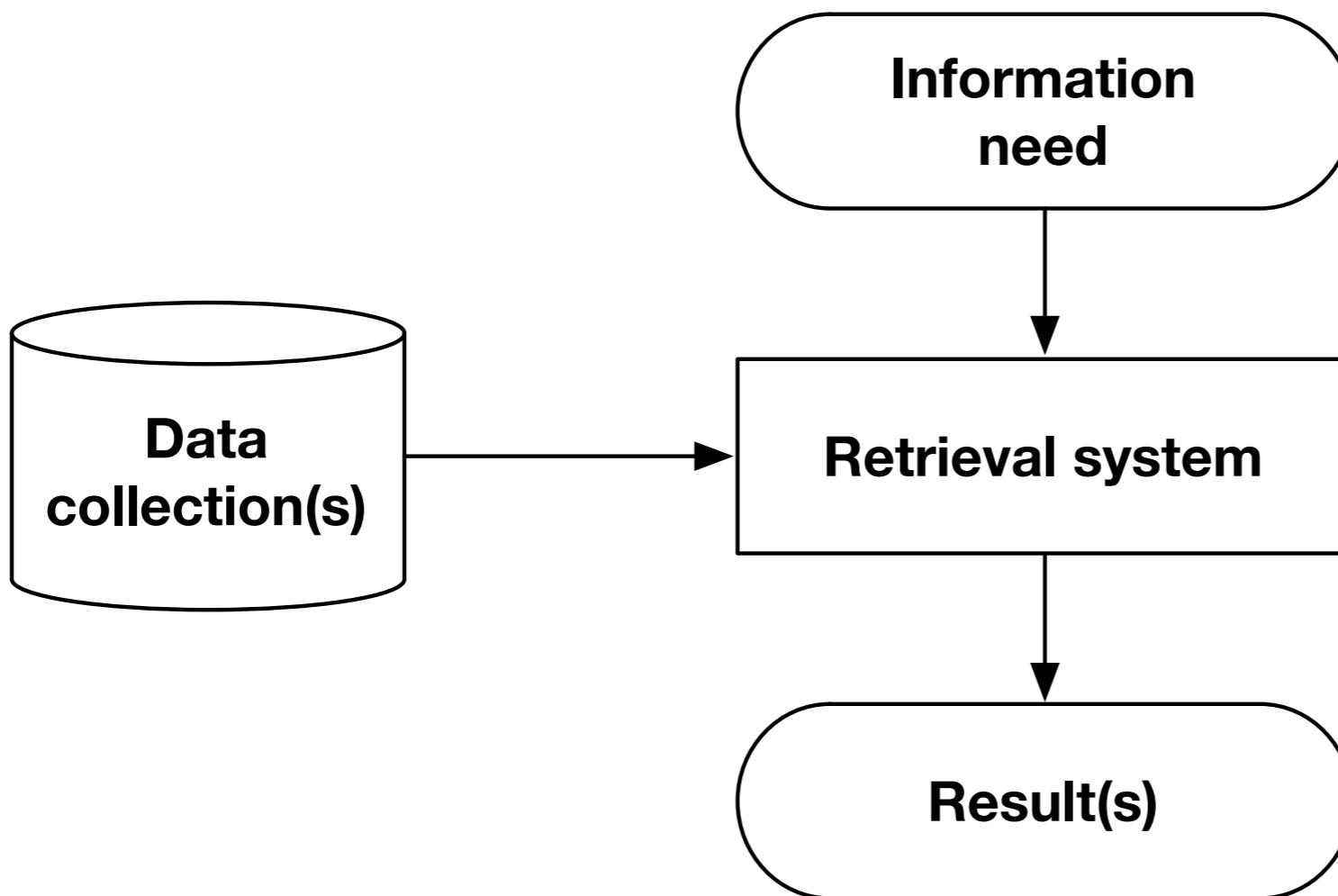
actressplay : Charactor: Herself Id: 1052946

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

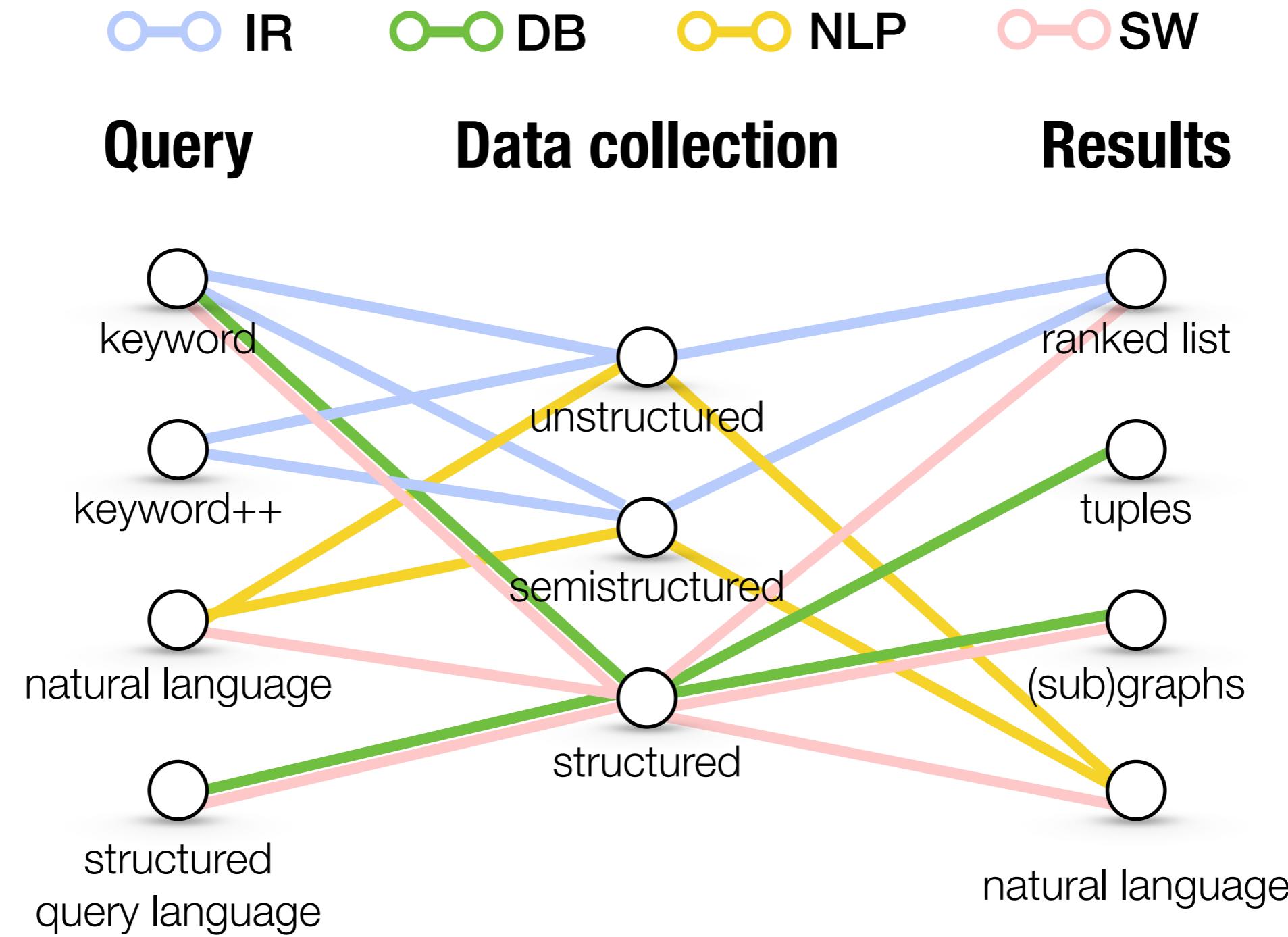
Retrieval system

Result(s)

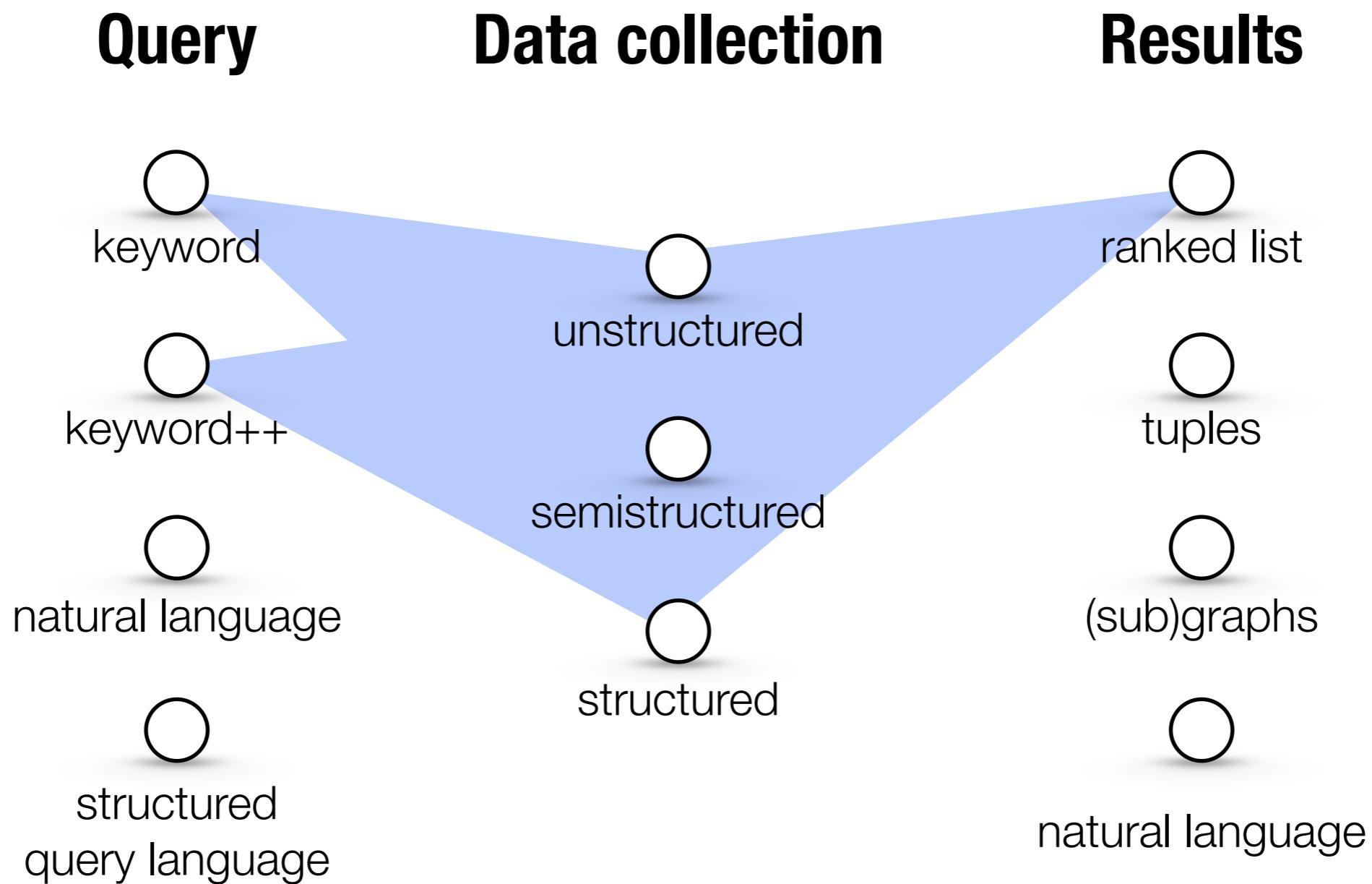
Result format

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

The possibilities are endless

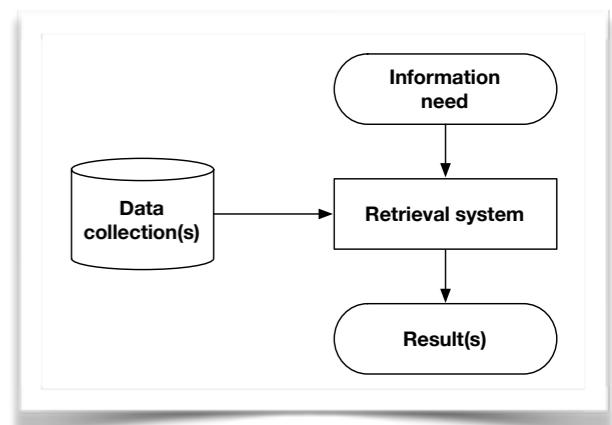


Our focus



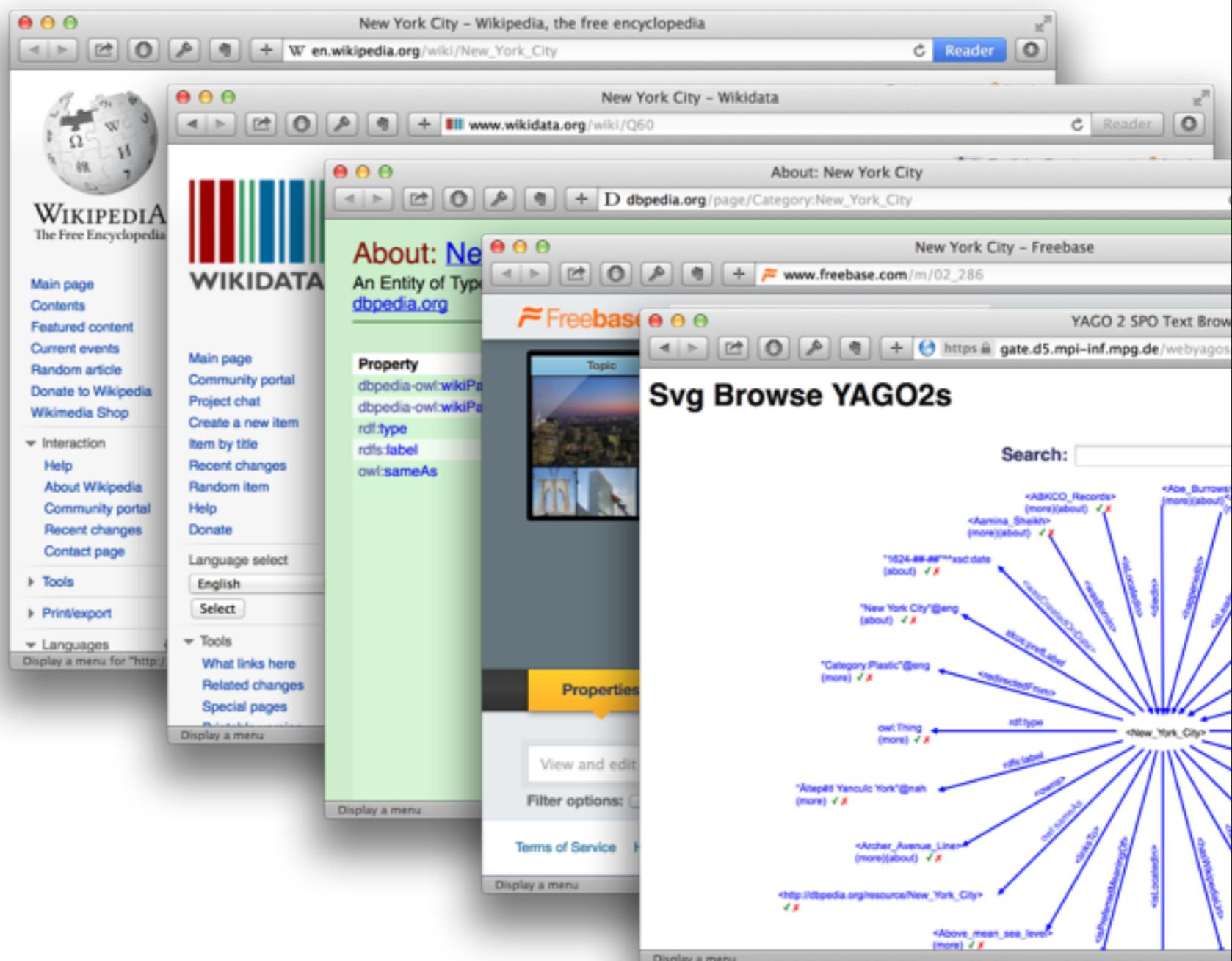
Data collection

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

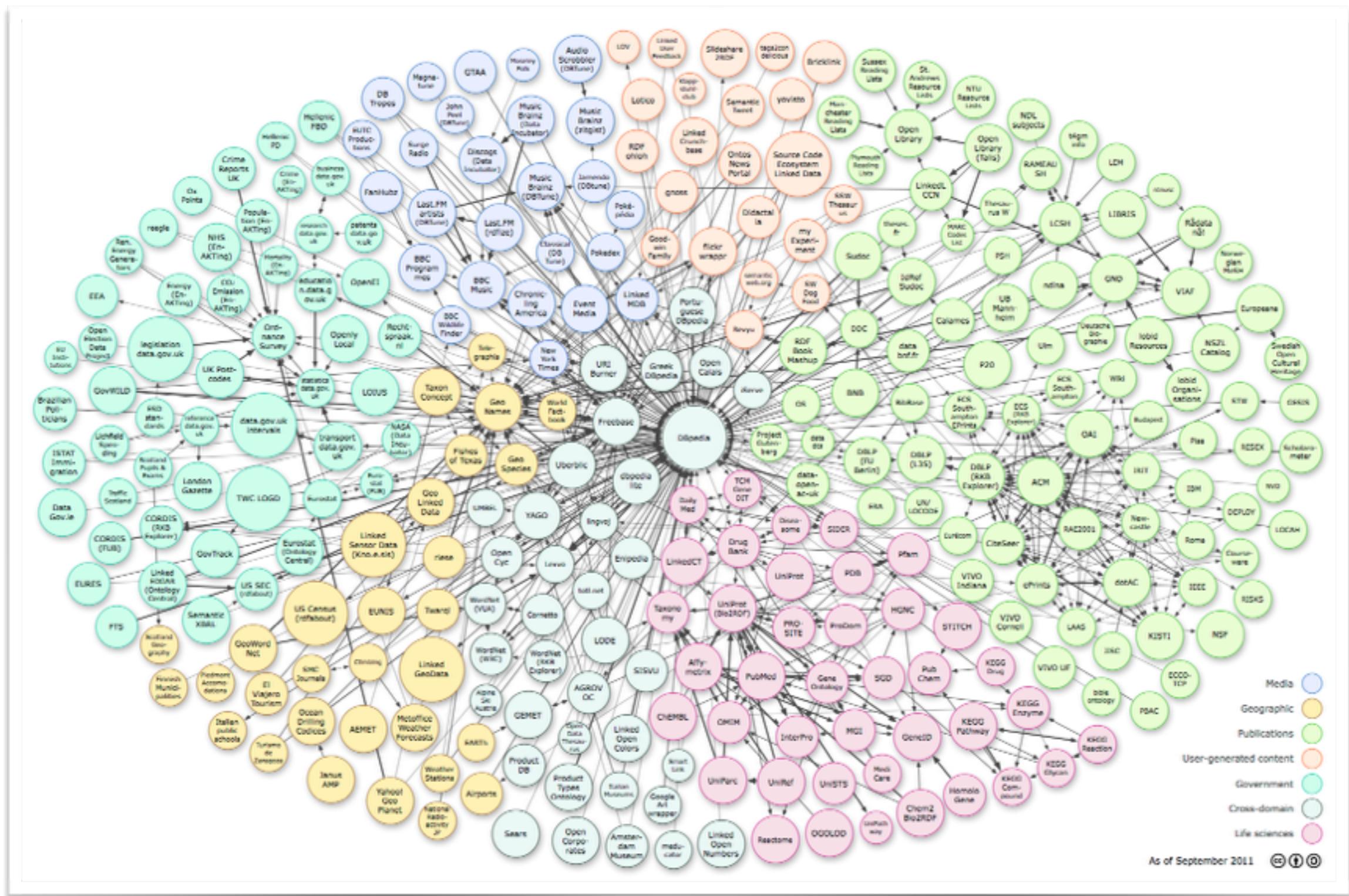


Popular (semi)structured data sources

- 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/ +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

Queries

- **Keyword queries**
 - Single-search-box paradigm
 - Typical web search queries
 - “Telegraphic” — neither well-formed nor grammatically correct
- **Keyword++ queries**
 - Augmented with form/facet-based input

Example keyword++ queries

amazon Try Prime

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

Shop by Department

Search All Departments Go

Search suggestions

Departments

Motorcycle & Powersports

- Motorcycle Protective Coats & Vests
- Powersports Body Parts

Sports & Outdoors

- Men's Athletic Clothing
- Men's Athletic Hoodies

Clothing & Accessories

- Men's Activewear
- Men's Shirts
- Men's Fashion Hoodies & Sweatshirts
- Novelty T-Shirts

"**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



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 ×

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 & Beverages

[Similar](#)

Connect

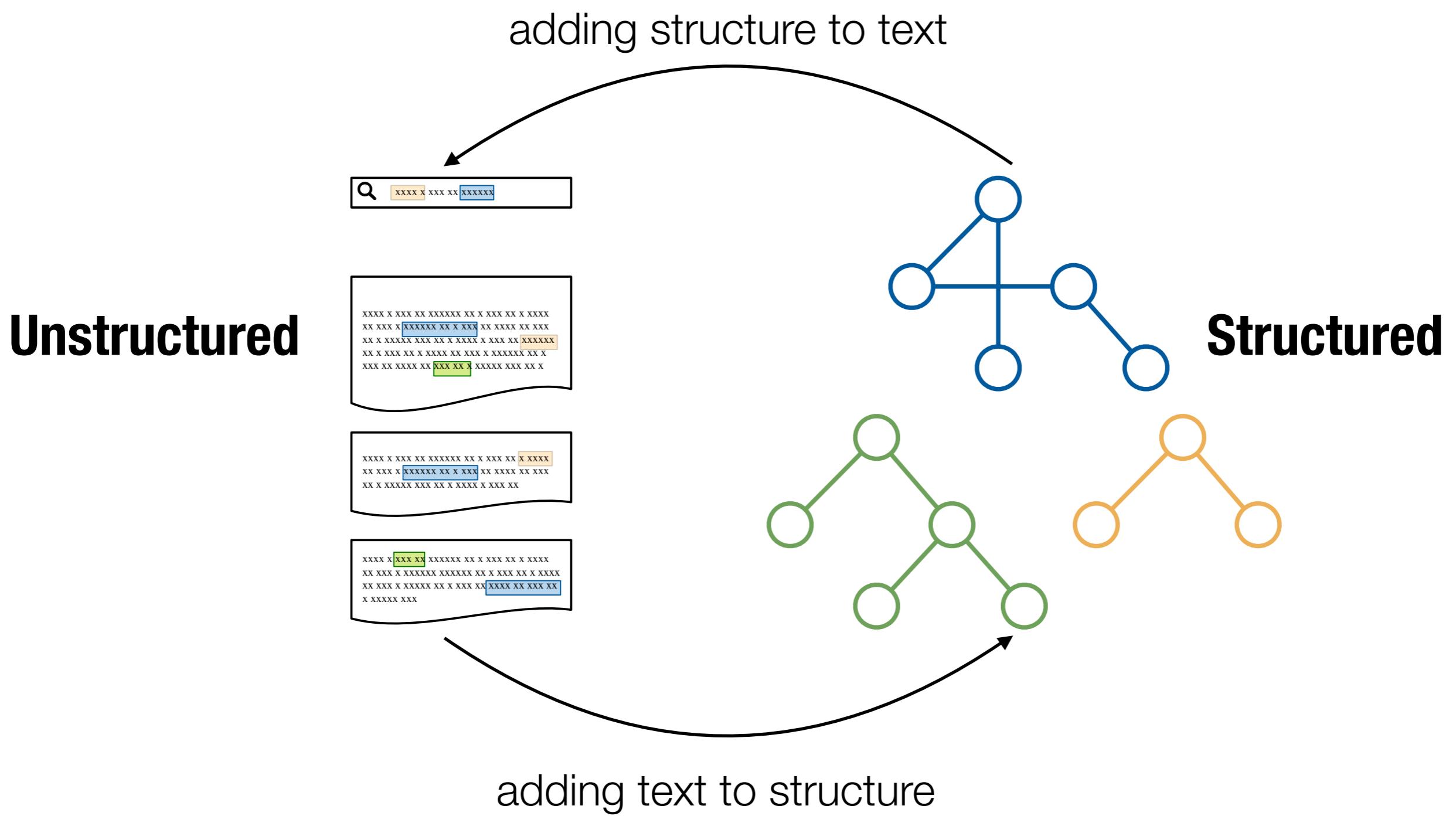
**John Smith**

Manager at McDonald's

London, United Kingdom · Restaurants

Connect

Interplay: (un)structured data



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

09:00 - 09:30 Welcome, introduction

09:30 - 10:30 Entity linking

10:30 - 11:00 Coffee break

11:00 - 12:00 Entity retrieval

12:00 - 12:20 Semantic search

12:20 - 12:30 Wrap up

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/. 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 and has 48 members. The 'About this group' section describes it as a public library for Entity Linking and Retrieval tutorials. A 'Sign up today - FREE' banner at the bottom encourages users to join.

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