Utilizing Knowledge Bases in Text-centric Information Retrieval

Laura Dietz (@lauradietz99)
University of New Hampshire

Alexander Kotov (@rusillini)
Wayne State University

Edgar Meij (@edgarmeij)
Bloomberg

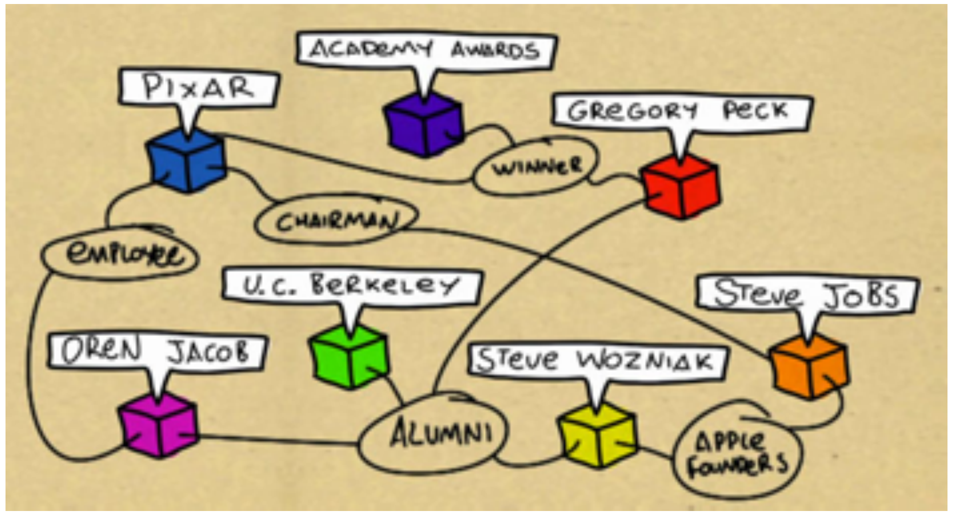
Slides at https://github.com/laura-dietz/tutorial-kb4ir

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

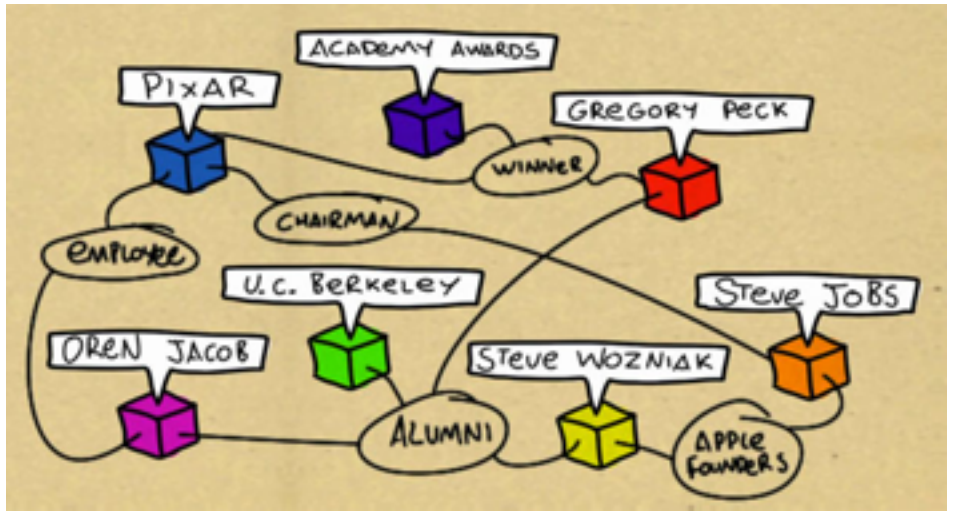
- The “backbone” of semantic search
- They define
  - entities
  - attributes
  - types
  - relations
  - (provenance, sometimes)
  - and more
    - external links, homepages, features, …

The Knowledge Graph

Learn more about one of the key breakthroughs behind the future of search

See it in action

Discover answers to questions you never thought to ask, and explore collections and lists.

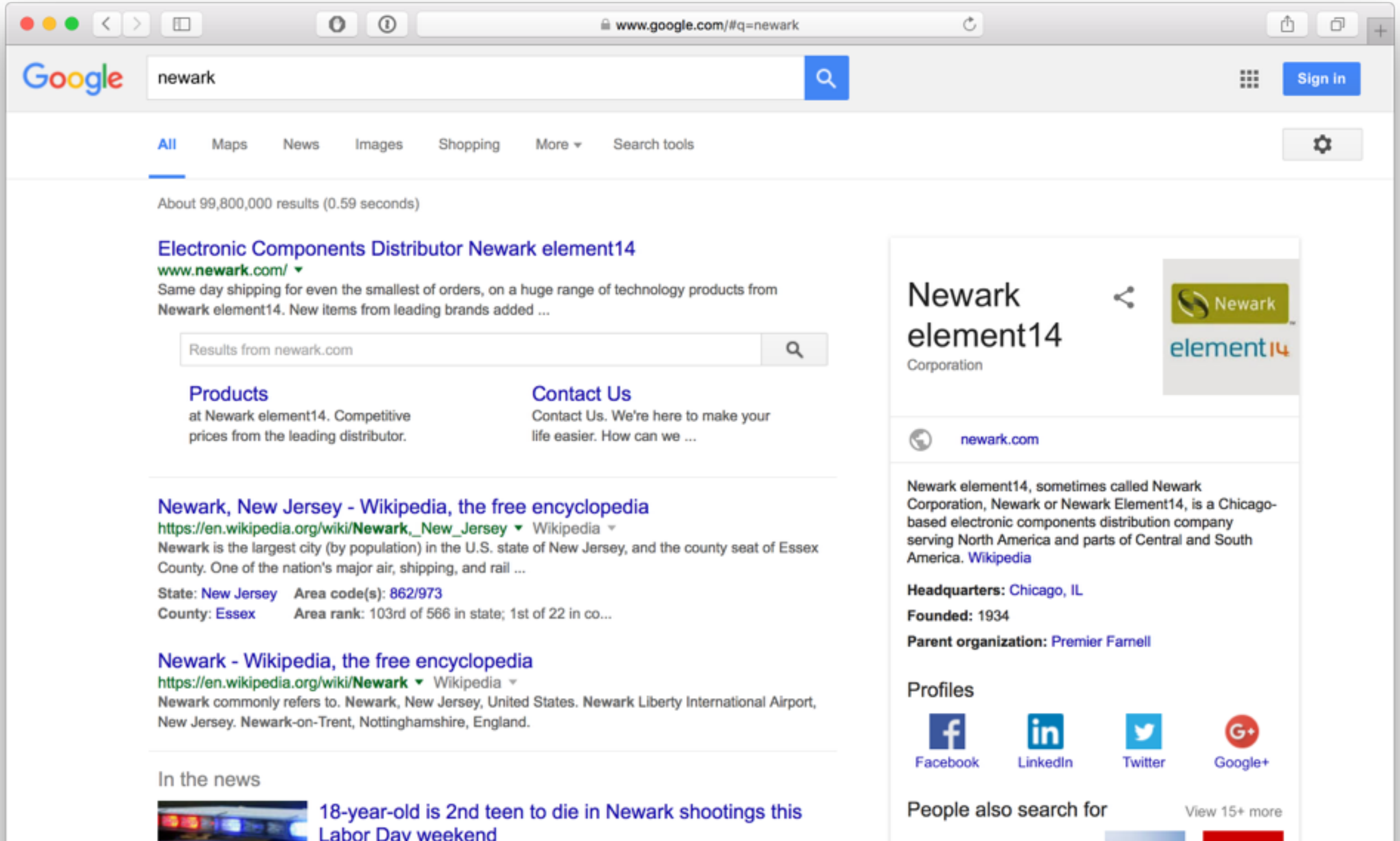


Knowledge graphs

dbpedia:Audi_A4

foaf:name	Audi A4
rdfs:label	Audi A4
rdfs:comment	The Audi A4 is a compact executive car produced since late 1994 by the German car manufacturer Audi, a subsidiary of the Volkswagen Group. The A4 has been built [...]
dbpprop:production	1994 2001 2005 2008
rdf:type	**dbpedia-owl:MeanOfTransportation** **dbpedia-owl:Automobile**
dbpedia-owl:manufacturer	**dbpedia:Audi**
dbpedia-owl:class	**dbpedia:Compact_executive_car**
owl:sameAs	**freebase:Audi A4**
is **dbpedia-owl:predecessor** of	**dbpedia:Audi_A5**
is **dbpprop:similar** of	**dbpedia:Cadillac_BLS**

Entity Linking/Retrieval



About 99,800,000 results (0.59 seconds)

Electronic Components Distributor Newark element14
www.newark.com/ ▾
Same day shipping for even the smallest of orders, on a huge range of technology products from Newark element14. New items from leading brands added ...

Products
at Newark element14. Competitive prices from the leading distributor.

Contact Us
Contact Us. We're here to make your life easier. How can we ...

Newark, New Jersey - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Newark,_New_Jersey ▾ Wikipedia ▾
Newark is the largest city (by population) in the U.S. state of New Jersey, and the county seat of Essex County. One of the nation's major air, shipping, and rail ...

State: New Jersey   Area code(s): 862/973
County: Essex   Area rank: 103rd of 566 in state; 1st of 22 in co...

Newark - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Newark ▾ Wikipedia ▾
Newark commonly refers to. Newark, New Jersey, United States. Newark Liberty International Airport, New Jersey. Newark-on-Trent, Nottinghamshire, England.

In the news

18-year-old is 2nd teen to die in Newark shootings this Labor Day weekend

Newark element14
Corporation

newark.com

Newark element14, sometimes called Newark Corporation, Newark or Newark Element14, is a Chicago-based electronic components distribution company serving North America and parts of Central and South America. Wikipedia

Headquarters: Chicago, IL

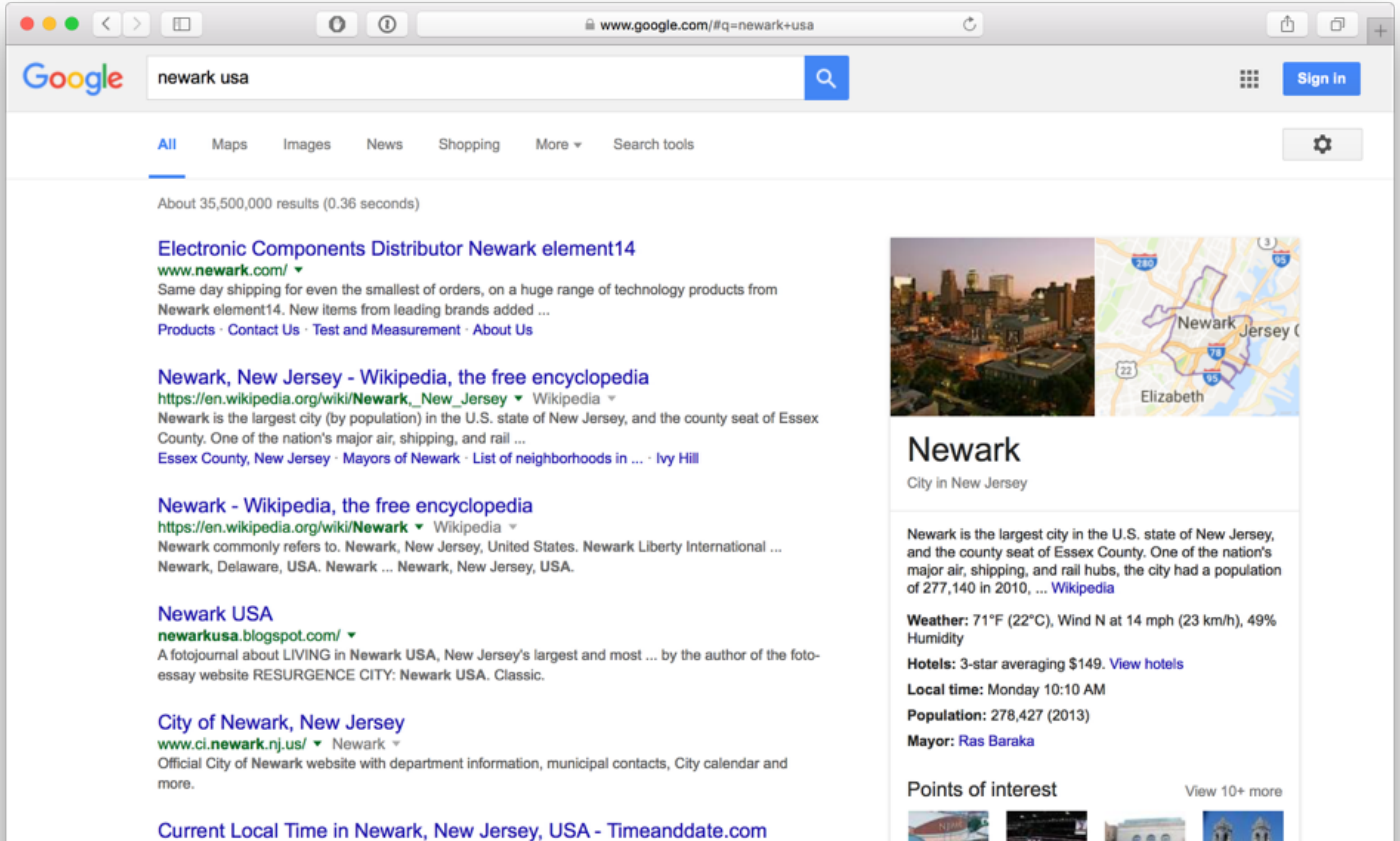
Founded: 1934

Parent organization: Premier Farnell

Profiles

People also search for

Entity Linking/Retrieval



newark usa

About 35,500,000 results (0.36 seconds)

Electronic Components Distributor Newark element14
www.newark.com/ ▾
Same day shipping for even the smallest of orders, on a huge range of technology products from Newark element14. New items from leading brands added ...
Products · Contact Us · Test and Measurement · About Us

Newark, New Jersey - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Newark,_New_Jersey ▾ Wikipedia ▾
Newark is the largest city (by population) in the U.S. state of New Jersey, and the county seat of Essex County. One of the nation's major air, shipping, and rail ...
Essex County, New Jersey · Mayors of Newark · List of neighborhoods in ... · Ivy Hill

Newark - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Newark ▾ Wikipedia ▾
Newark commonly refers to. Newark, New Jersey, United States. Newark Liberty International ... Newark, Delaware, USA. Newark ... Newark, New Jersey, USA.

Newark USA
newarkusa.blogspot.com/ ▾
A fotojournal about LIVING in Newark USA, New Jersey's largest and most ... by the author of the foto-essay website RESURGENCE CITY: Newark USA. Classic.

City of Newark, New Jersey
www.ci.newark.nj.us/ ▾ Newark ▾
Official City of Newark website with department information, municipal contacts, City calendar and more.

Current Local Time in Newark, New Jersey, USA - Timeanddate.com

Newark
City in New Jersey

Newark is the largest city in the U.S. state of New Jersey, and the county seat of Essex County. One of the nation's major air, shipping, and rail hubs, the city had a population of 277,140 in 2010, ... Wikipedia

Weather: 71°F (22°C), Wind N at 14 mph (23 km/h), 49% Humidity

Hotels: 3-star averaging $149. View hotels

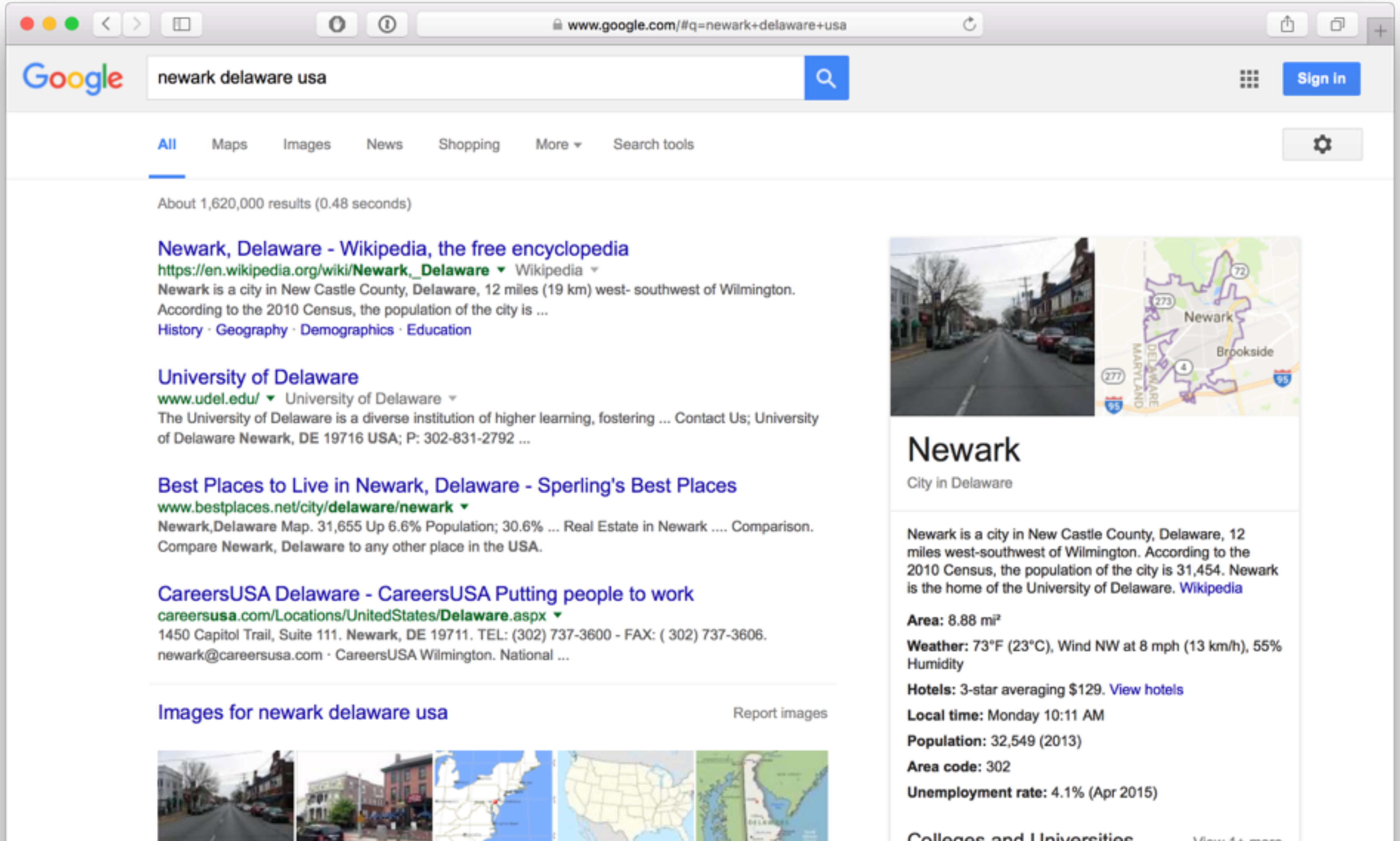
Local time: Monday 10:10 AM

Population: 278,427 (2013)

Mayor: Ras Baraka

Points of interest
View 10+ more

Entity Linking/Retrieval



About 1,620,000 results (0.48 seconds)

Newark, Delaware - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Newark,_Delaware ▾ Wikipedia ▾
Newark is a city in New Castle County, Delaware, 12 miles (19 km) west- southwest of Wilmington. According to the 2010 Census, the population of the city is ...
History · Geography · Demographics · Education

University of Delaware
www.udel.edu/ ▾ University of Delaware ▾
The University of Delaware is a diverse institution of higher learning, fostering ... Contact Us; University of Delaware Newark, DE 19716 USA; P: 302-831-2792 ...

Best Places to Live in Newark, Delaware - Sperling's Best Places
www.bestplaces.net/city/delaware/newark ▾
Newark,Delaware Map. 31,655 Up 6.6% Population; 30.6% ... Real Estate in Newark .... Comparison. Compare Newark, Delaware to any other place in the USA.

CareersUSA Delaware - CareersUSA Putting people to work
careersusa.com/Locations/UnitedStates/Delaware.aspx ▾
1450 Capitol Trail, Suite 111. Newark, DE 19711. TEL: (302) 737-3600 - FAX: (302) 737-3606. newark@careersusa.com · CareersUSA Wilmington. National ...

Images for newark delaware usa

Newark

City in Delaware

Newark is a city in New Castle County, Delaware, 12 miles west-southwest of Wilmington. According to the 2010 Census, the population of the city is 31,454. Newark is the home of the University of Delaware. Wikipedia

Area: 8.88 mi²

Weather: 73°F (23°C), Wind NW at 8 mph (13 km/h), 55% Humidity

Hotels: 3-star averaging $129. View hotels

Local time: Monday 10:11 AM

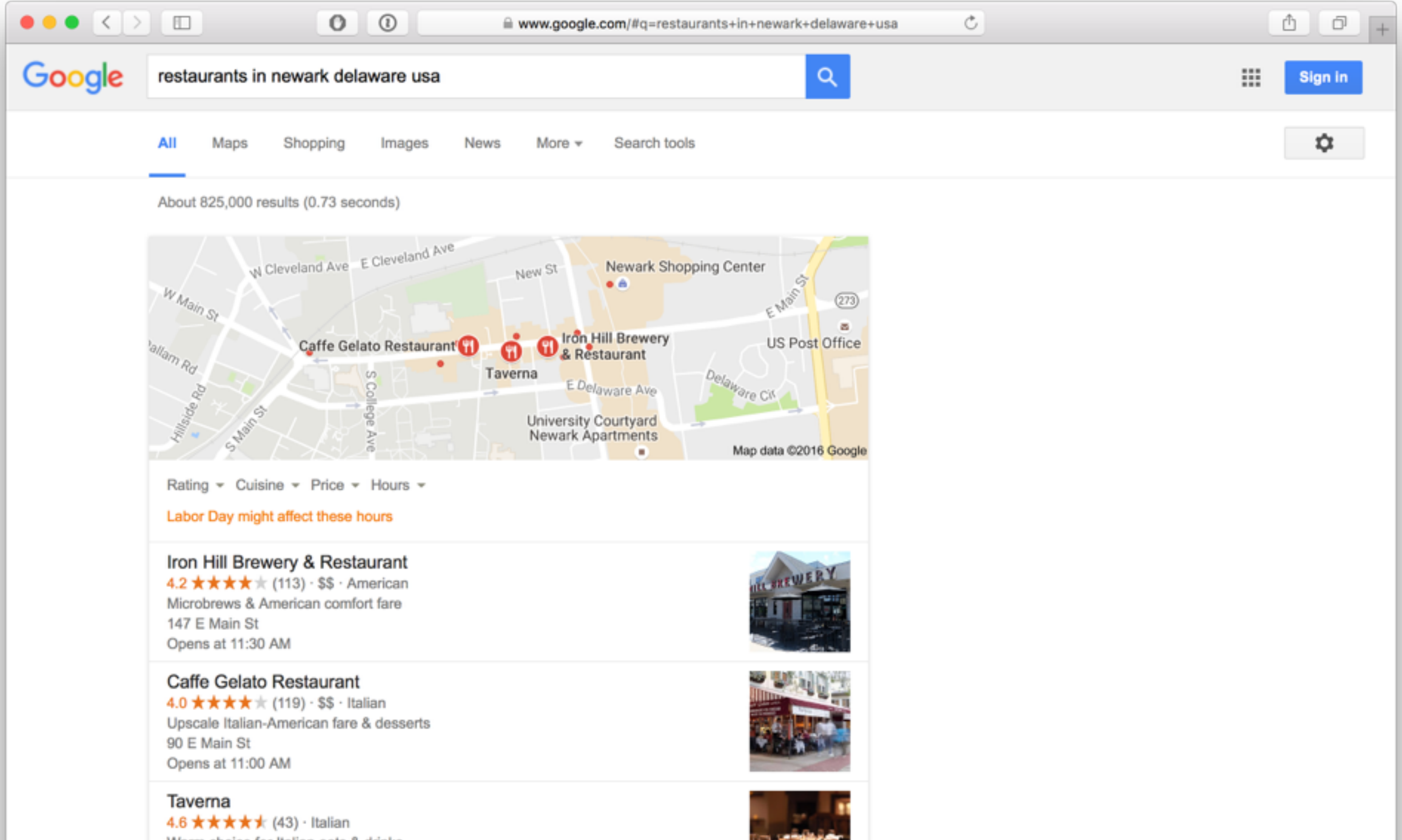
Population: 32,549 (2013)

Area code: 302

Unemployment rate: 4.1% (Apr 2015)

Colleges and Universities

Entity Retrieval



restaurants in newark delaware usa

About 825,000 results (0.73 seconds)

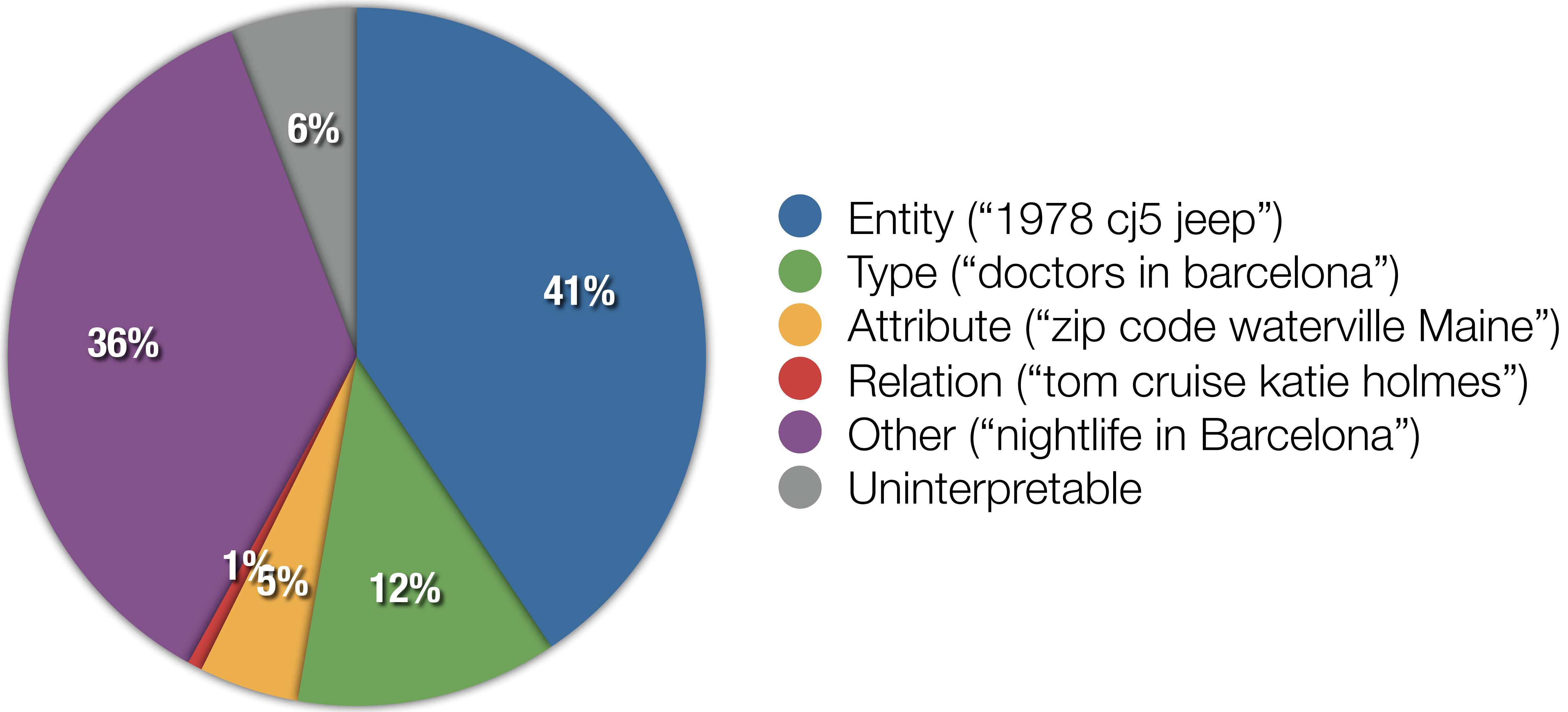
Labor Day might affect these hours

Iron Hill Brewery & Restaurant
4.2 ★★★★☆ (113) · $$ · American
Microbrews & American comfort fare
147 E Main St
Opens at 11:30 AM

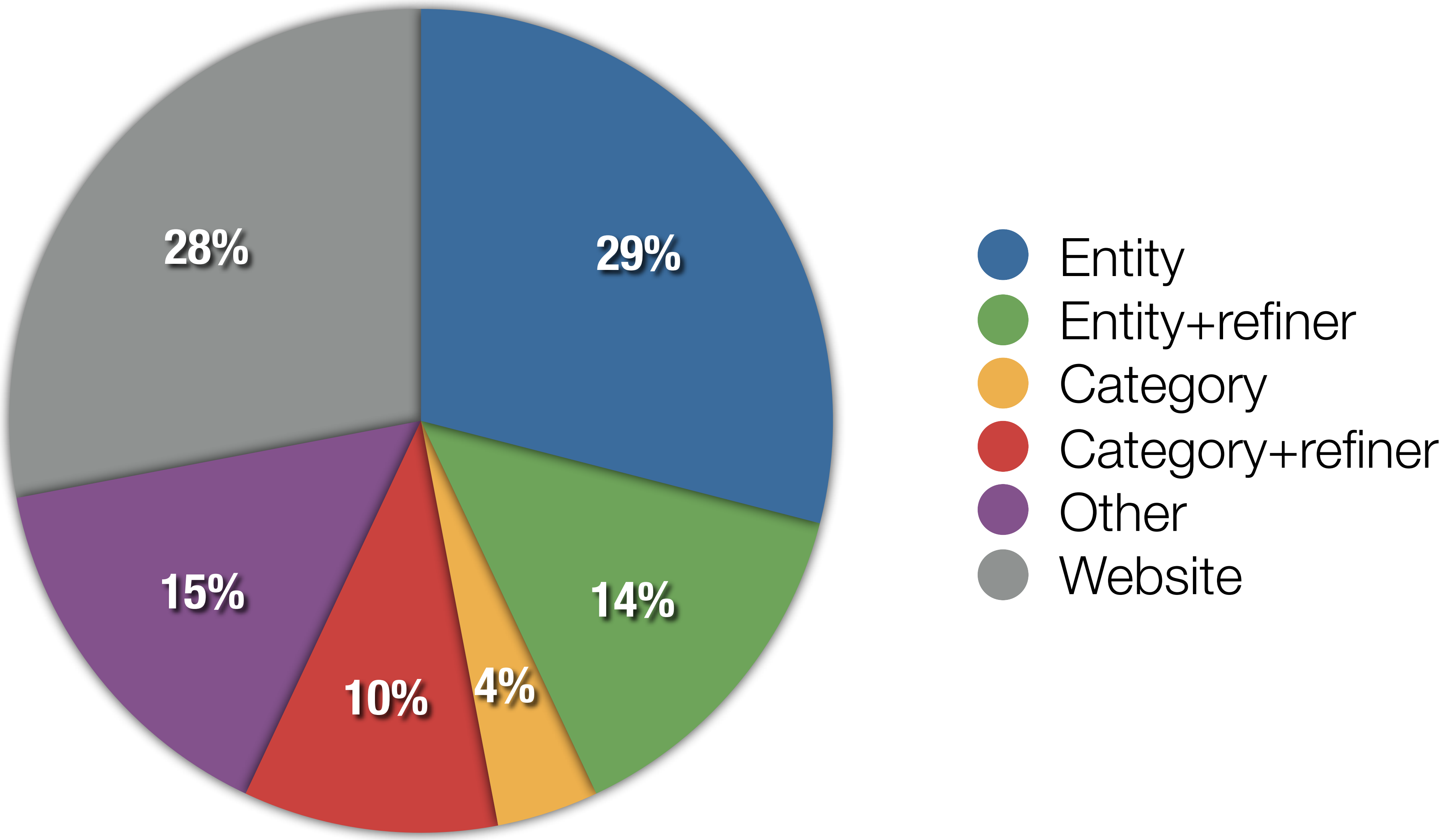
Caffe Gelato Restaurant
4.0 ★★★★☆ (119) · $$ · Italian
Upscale Italian-American fare & desserts
90 E Main St
Opens at 11:00 AM

Taverna
4.6 ★★★★★ (43) · Italian

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



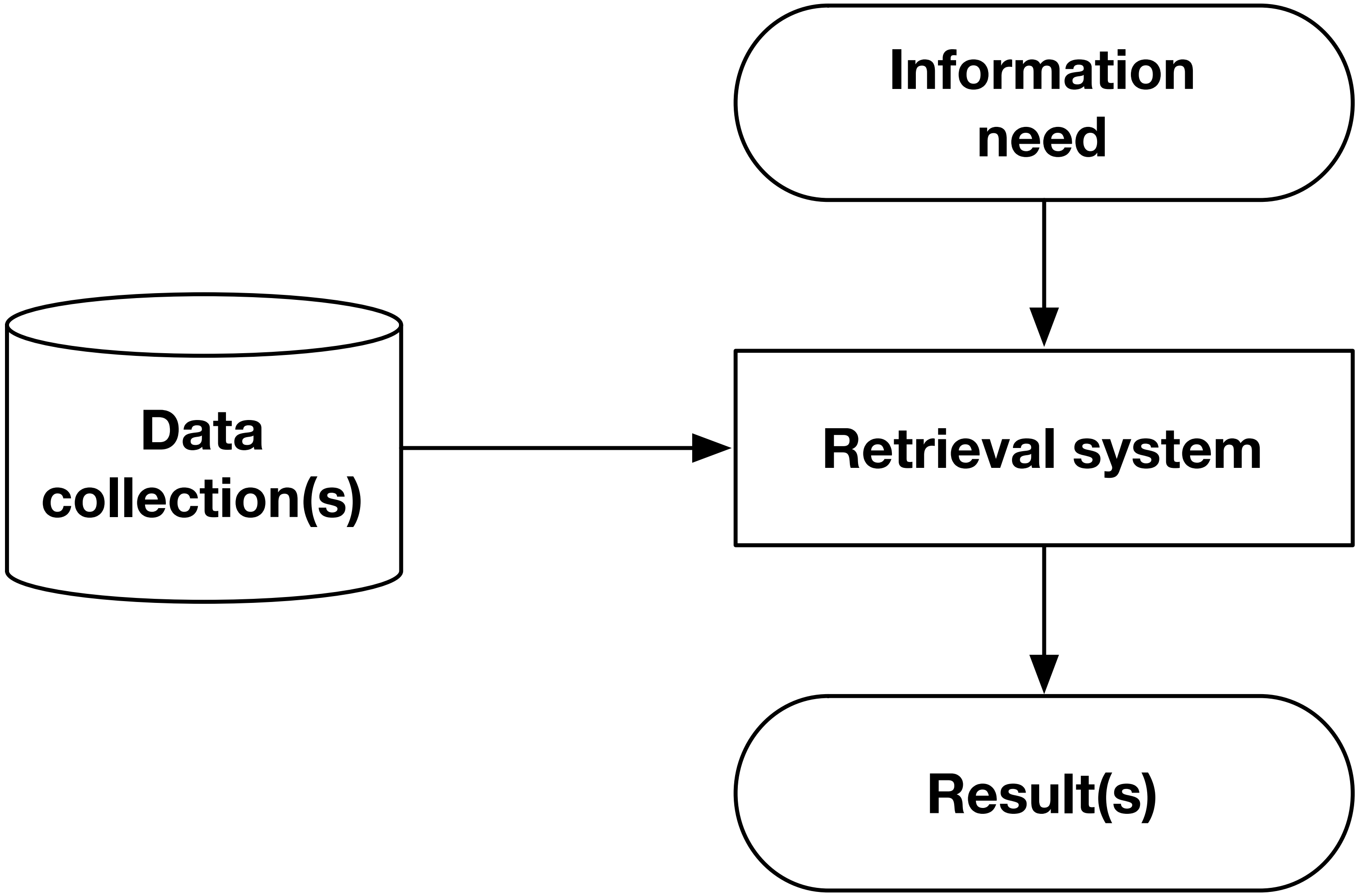
Distribution of web search queries [Lin et al. 2011]



Today’s focus

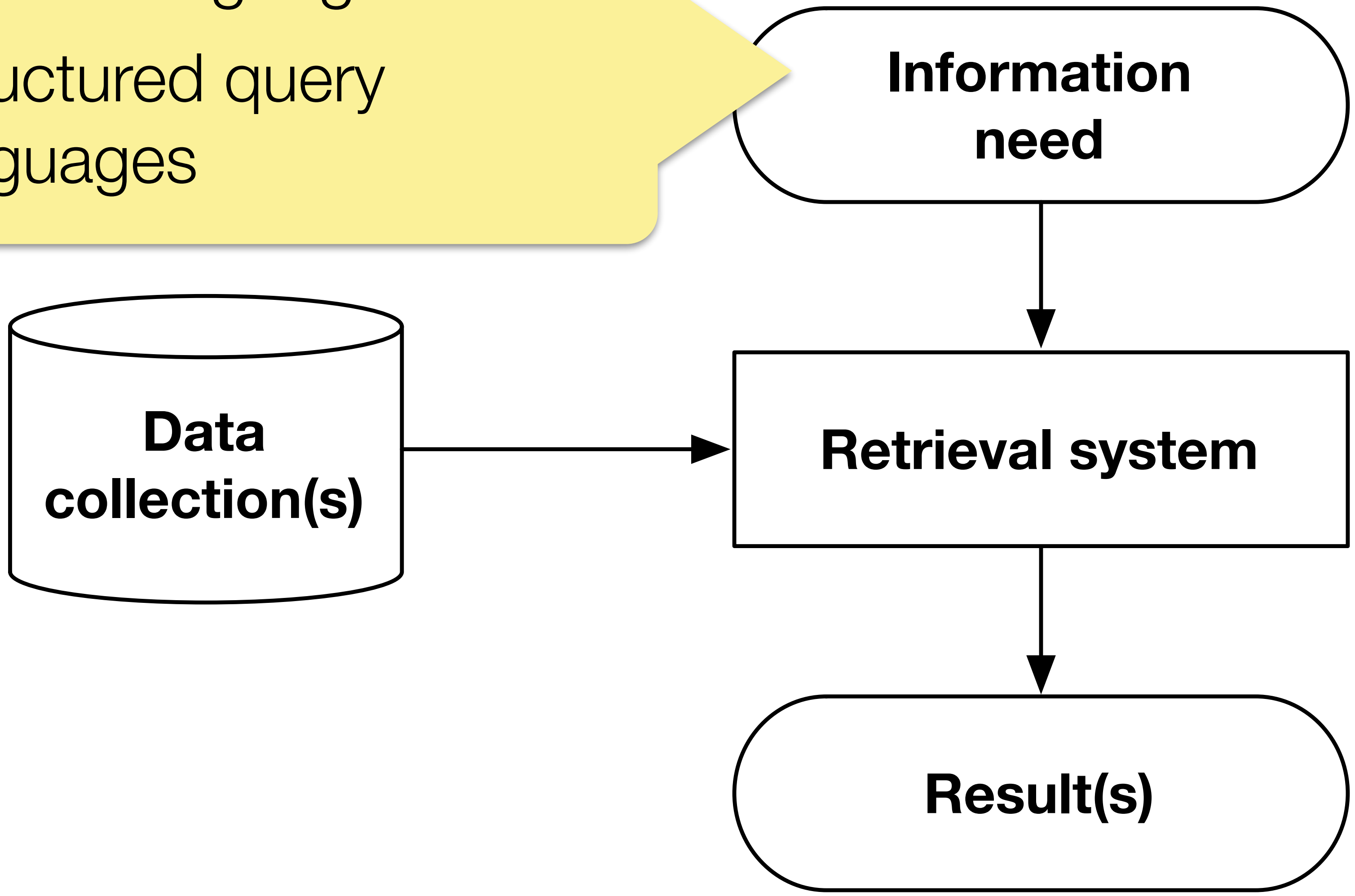
How to use KGs to improve information access.

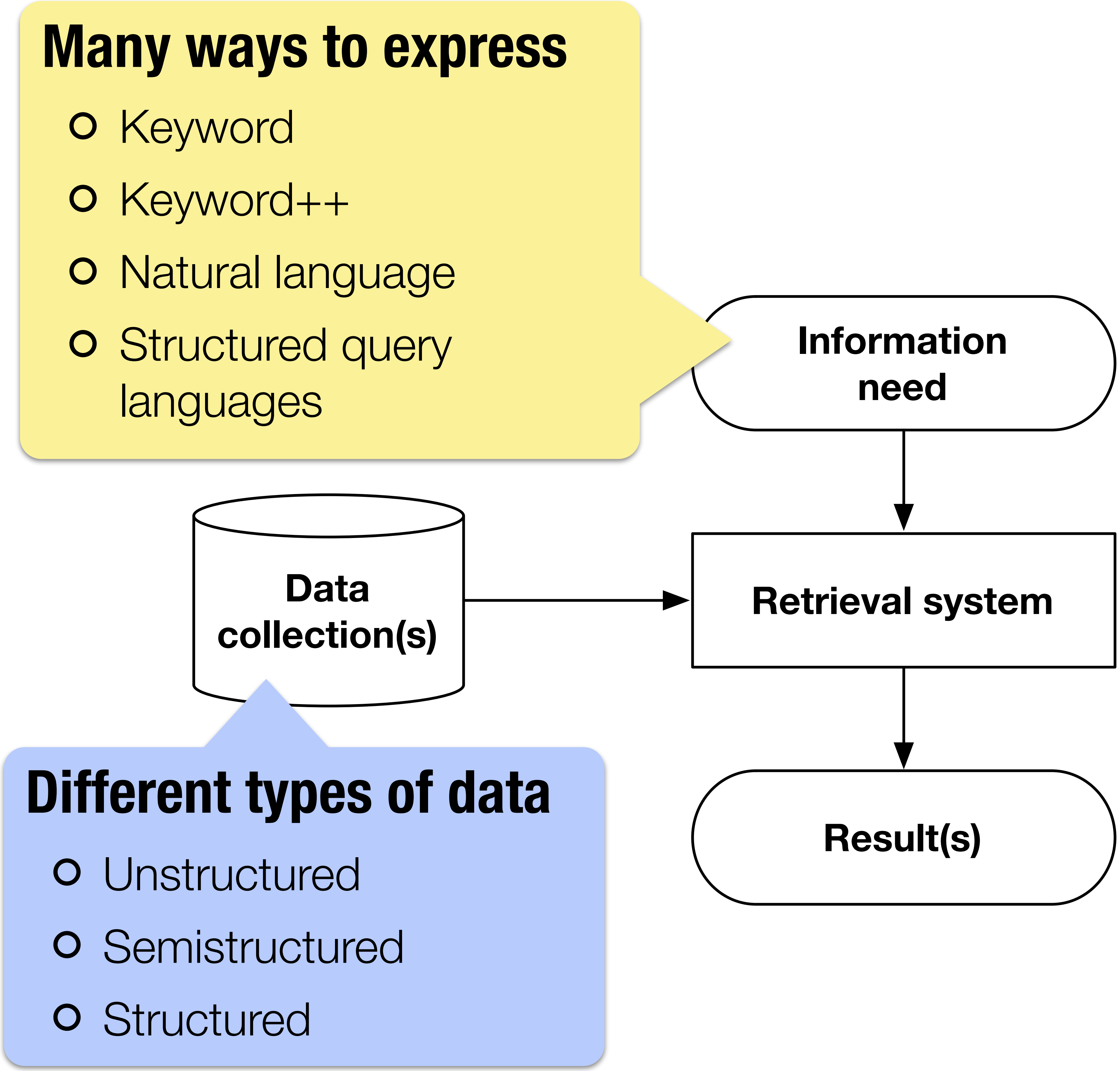
Birds-eye view



Many ways to express

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





Many ways to express

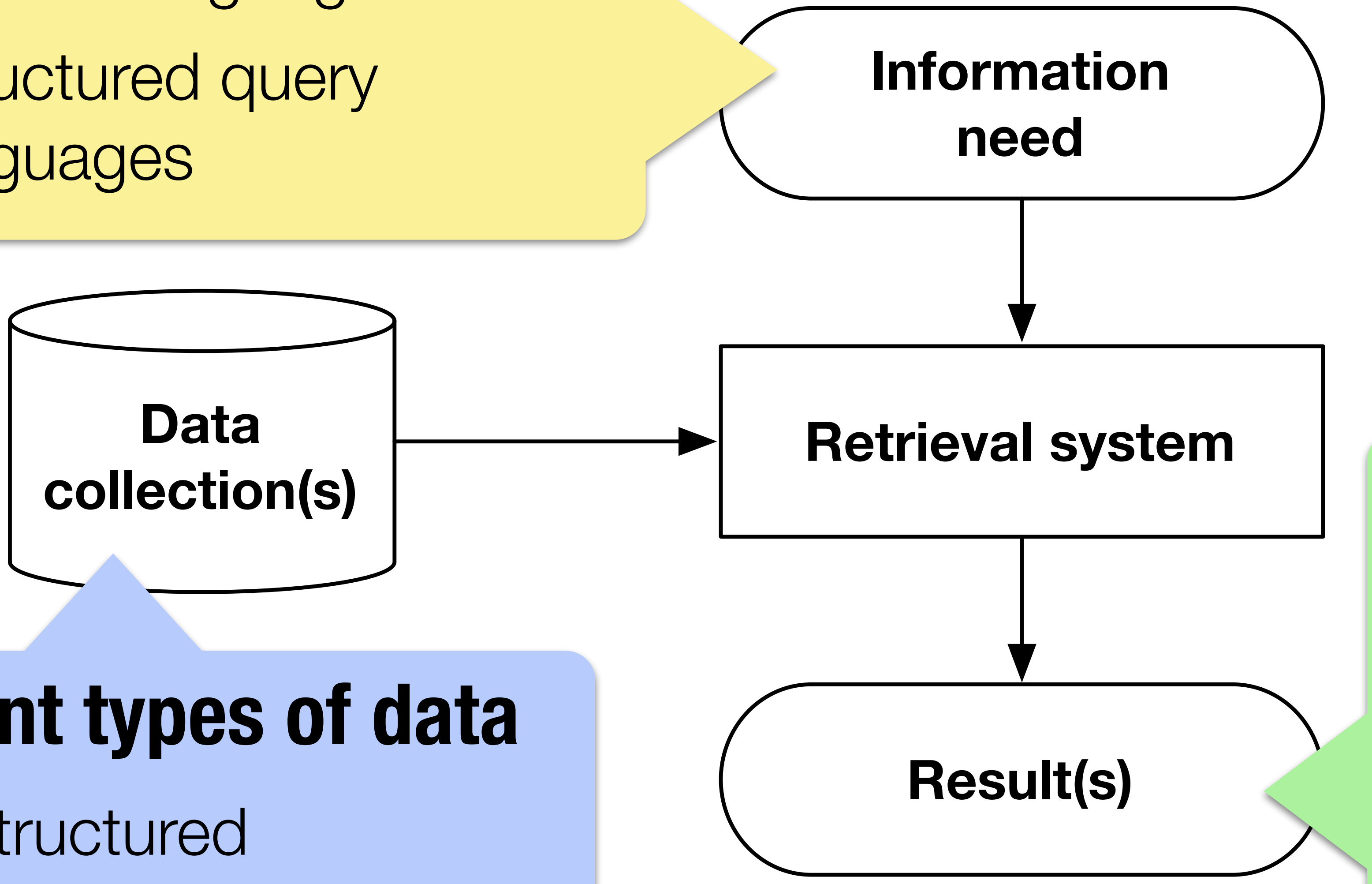
- Keyword
- Keyword++
- Natural language
- Structured query languages

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

Result format

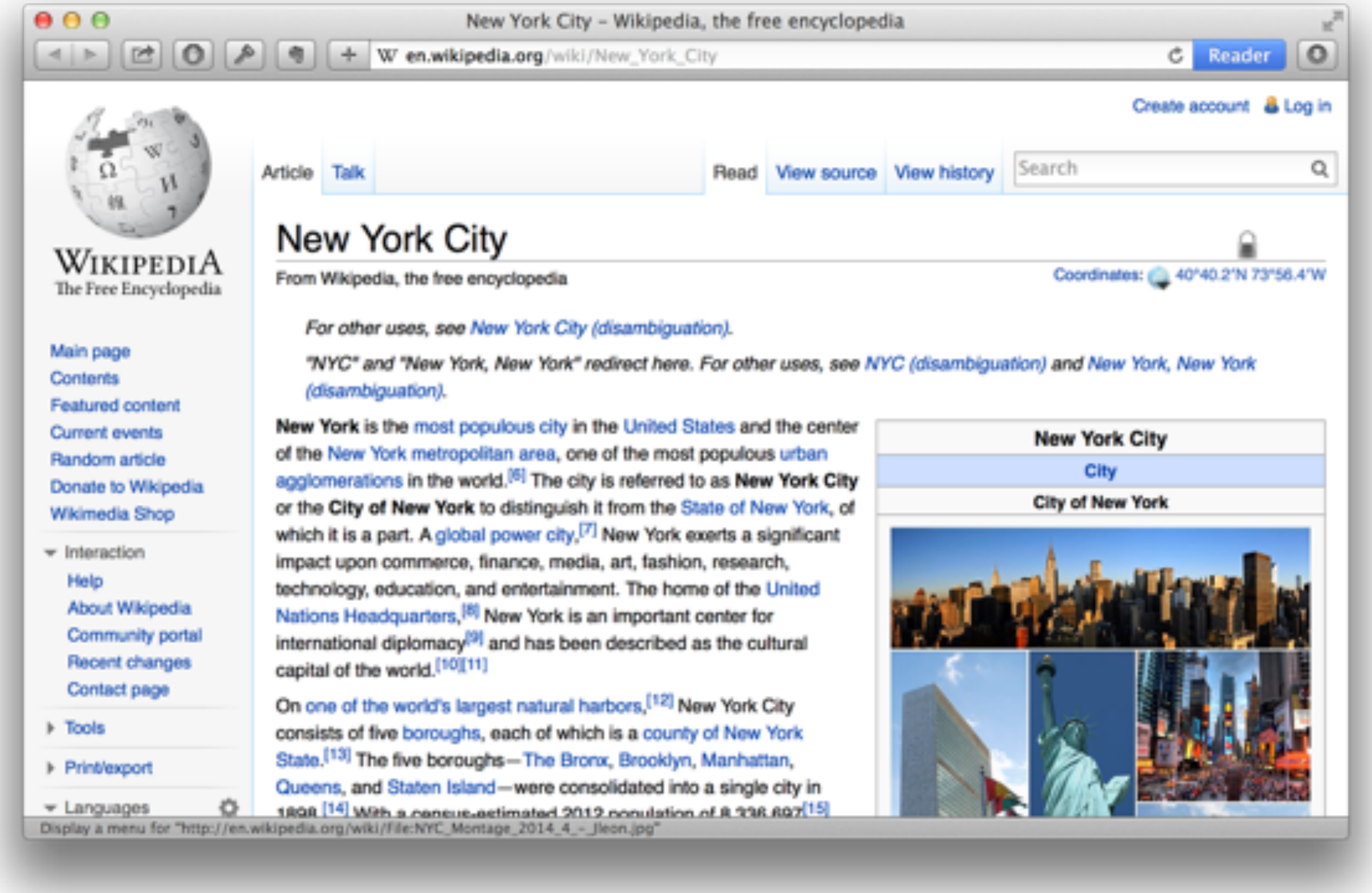
- Ranked list
- Tuples
- (Sub)graphs
- Natural language

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



New York City

From Wikipedia, the free encyclopedia

Coordinates: 40°40.2′N 73°56.4′W

For other uses, see New York City (disambiguation).

"NYC" and "New York, New York" redirect here. For other uses, see NYC (disambiguation) and New York, New York (disambiguation).

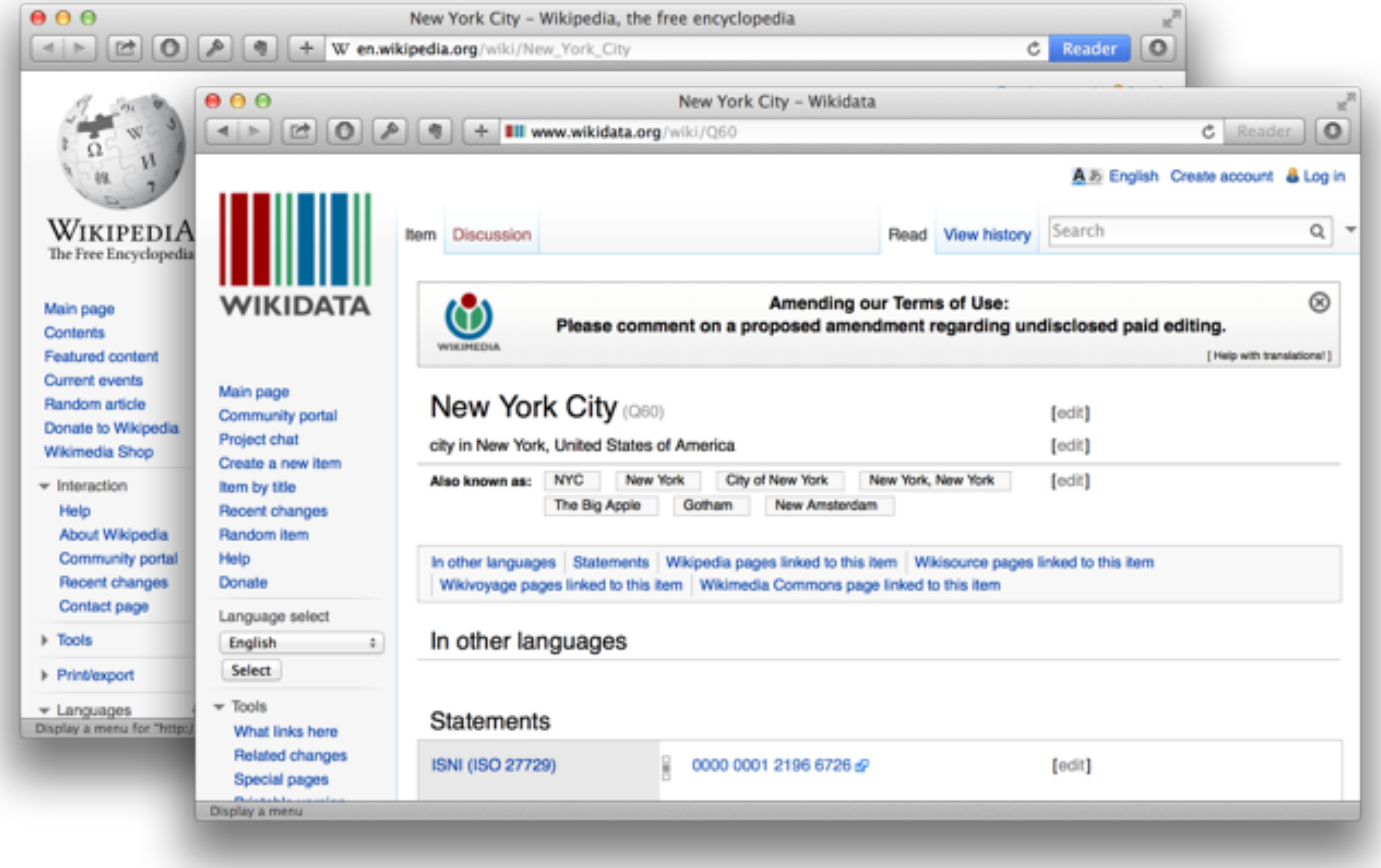
New York City
City
City of New York

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.[6] The city is referred to as **New York City** or the **City of New York** to distinguish it from the State of New York, of which it is a part. A global power city,[7] New York exerts a significant impact upon commerce, finance, media, art, fashion, research, technology, education, and entertainment. The home of the United Nations Headquarters,[8] New York is an important center for international diplomacy[9] and has been described as the cultural capital of the world.[10][11]

On one of the world's largest natural harbors,[12] New York City consists of five boroughs, each of which is a county of New York State.[13] The five boroughs—The Bronx, Brooklyn, Manhattan, Queens, and Staten Island—were consolidated into a single city in 1898.[14] With a census-estimated 2012 population of 8,336,697[15]

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



Amending our Terms of Use:
Please comment on a proposed amendment regarding undisclosed paid editing.
[Help with translations!]

New York City (Q60) [edit]

city in New York, United States of America [edit]

Also known as: NYC | New York | City of New York | New York, New York | The Big Apple | Gotham | New Amsterdam [edit]

In other languages | Statements | Wikipedia pages linked to this item | Wikisource pages linked to this item | Wikivoyage pages linked to this item | Wikimedia Commons page linked to this item

In other languages

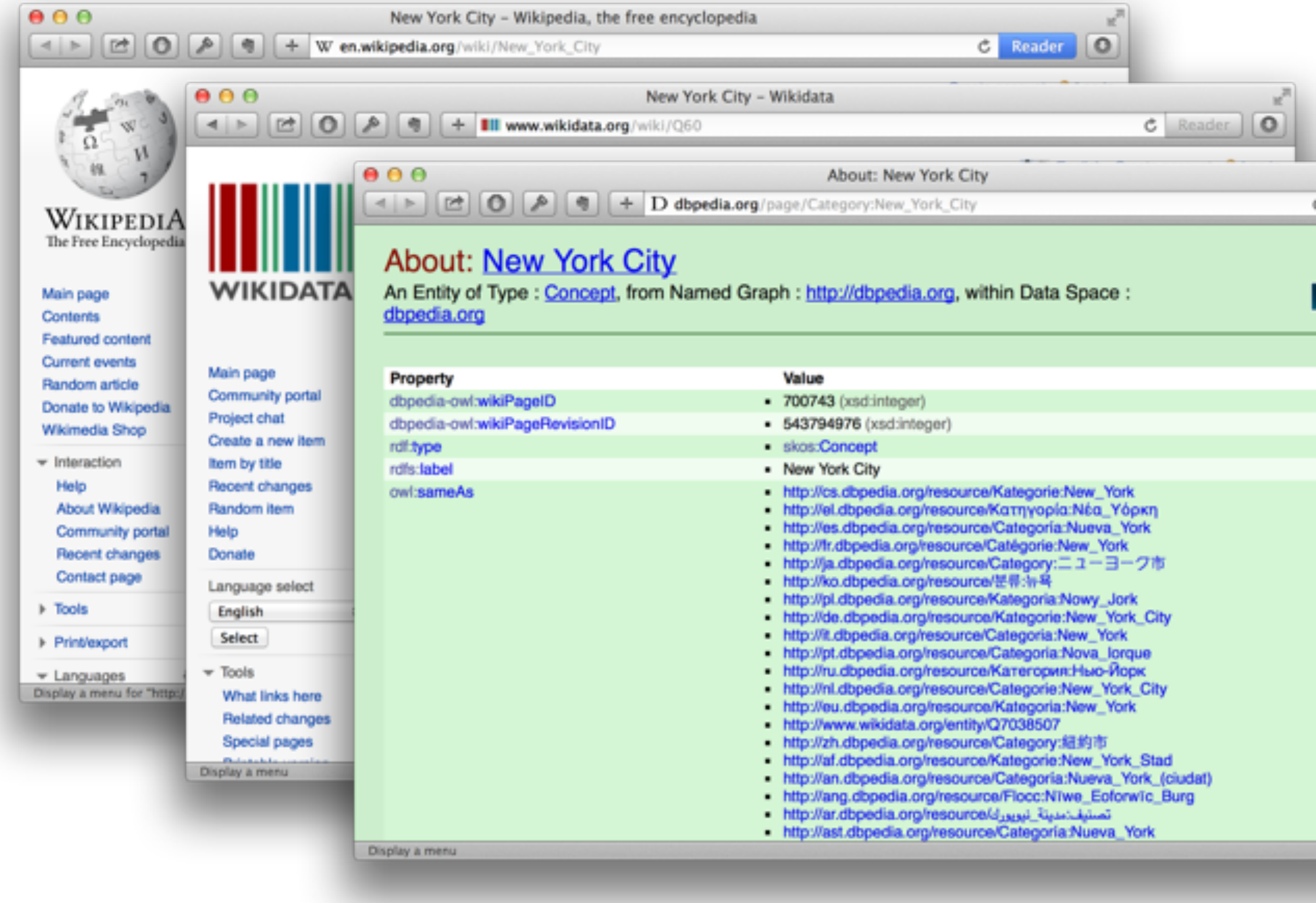
Statements

ISNI (ISO 27729) | 0000 0001 2196 6726 [edit]

ICTIR 2016 Tutorial on Utilizing KGs in Text-centric IR

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



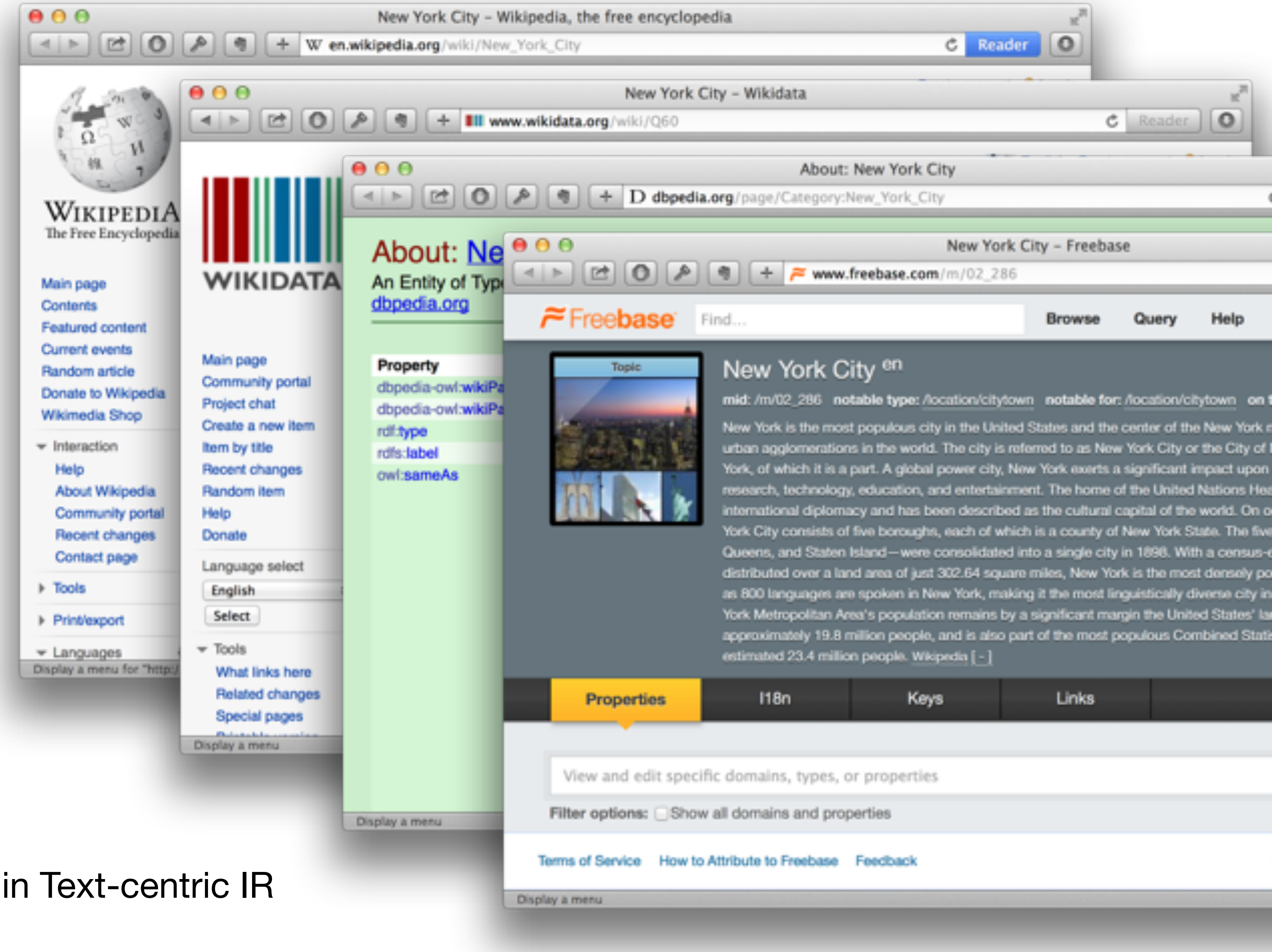
About: New York City

An Entity of Type : Concept, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Property	Value
dbpedia-owl:wikiPageID	700743 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	543794976 (xsd:integer)
rdf:type	skos:Concept
rdfs:label	New York City
owl:sameAs	http://cs.dbpedia.org/resource/Kategorie:New_York http://el.dbpedia.org/resource/Κατηγορία:Νέα_Υόρκη http://es.dbpedia.org/resource/Categoría:Nueva_York http://fr.dbpedia.org/resource/Catégorie:New_York http://ja.dbpedia.org/resource/Category:ニューヨーク市 http://ko.dbpedia.org/resource/분류:뉴욕 http://pl.dbpedia.org/resource/Kategoria:Nowy_Jork http://de.dbpedia.org/resource/Kategorie:New_York_City http://it.dbpedia.org/resource/Categoria:New_York http://pt.dbpedia.org/resource/Categoria:Nova_Iorque http://ru.dbpedia.org/resource/Категория:Нью-Йорк http://nl.dbpedia.org/resource/Categorie:New_York_City http://eu.dbpedia.org/resource/Kategoria:New_York http://www.wikidata.org/entity/Q7038507 http://zh.dbpedia.org/resource/Category:紐約市 http://af.dbpedia.org/resource/Kategorie:New_York_Stad http://an.dbpedia.org/resource/Categoría:Nueva_York_(ciudat) http://ang.dbpedia.org/resource/Flocc:Nīwe_Eoforwīc_Burg http://ar.dbpedia.org/resource/تصنيف:مدينة_نيويورك http://ast.dbpedia.org/resource/Categoría:Nueva_York

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



New York City en

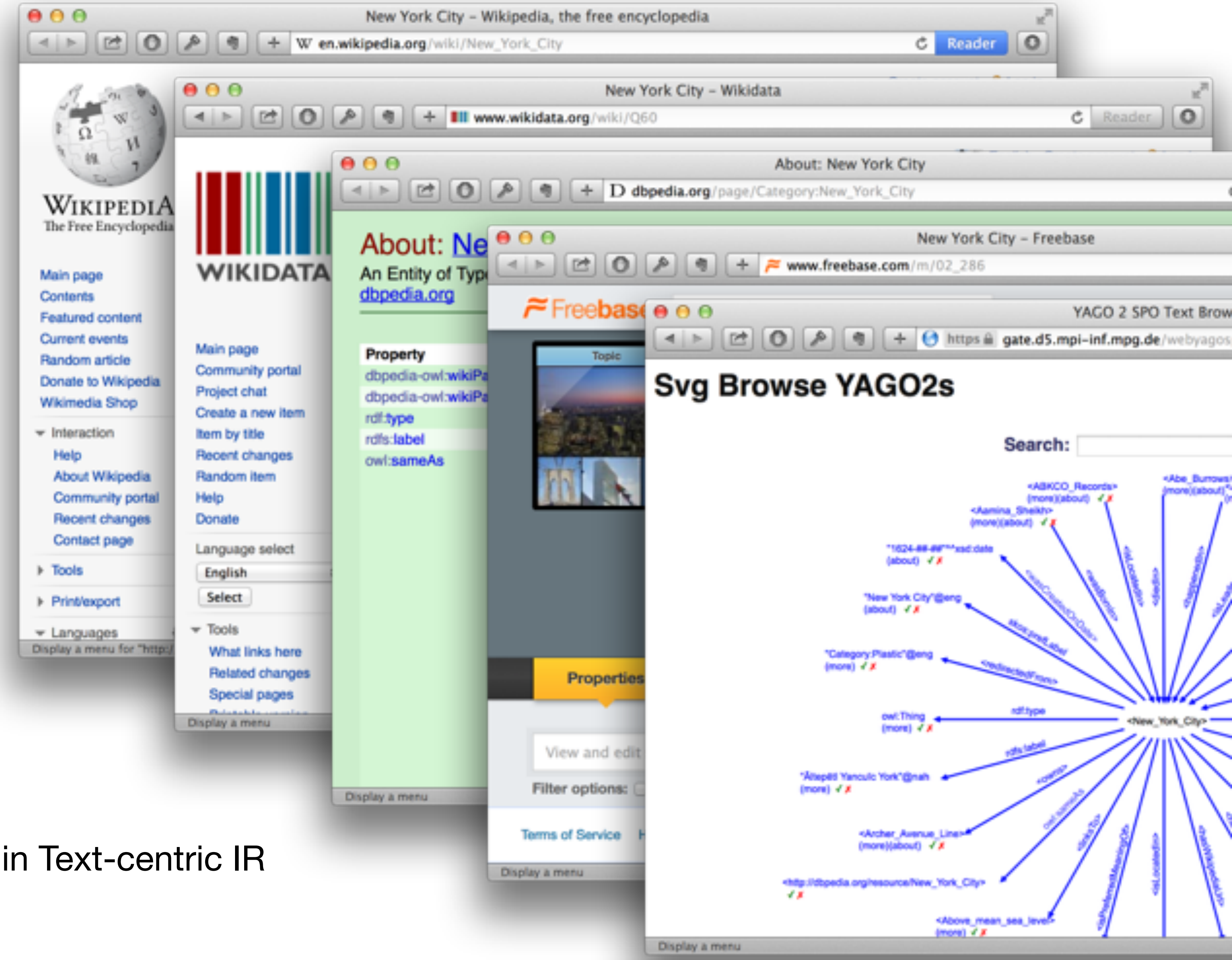
mid: /m/02_286 notable type: /location/citytown notable for: /location/citytown on

New York is the most populous city in the United States and the center of the New York m
urban agglomerations in the world. The city is referred to as New York City or the City of I
York, of which it is a part. A global power city, New York exerts a significant impact upon
research, technology, education, and entertainment. The home of the United Nations He
international diplomacy and has been described as the cultural capital of the world. On o
York City consists of five boroughs, each of which is a county of New York State. The five
Queens, and Staten Island—were consolidated into a single city in 1898. With a census-e
distributed over a land area of just 302.64 square miles, New York is the most densely po
as 800 languages are spoken in New York, making it the most linguistically diverse city in
York Metropolitan Area's population remains by a significant margin the United States' la
approximately 19.8 million people, and is also part of the most populous Combined State
estimated 23.4 million people. Wikipedia [-]

Filter options: □ Show all domains and properties

Popular (semi)structured data sources

- Wikipedia
- Wikidata
- DBpedia
- Freebase
- YAGO



ICTIR 2016 Tutorial on Utilizing KGs in Text-centric IR

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

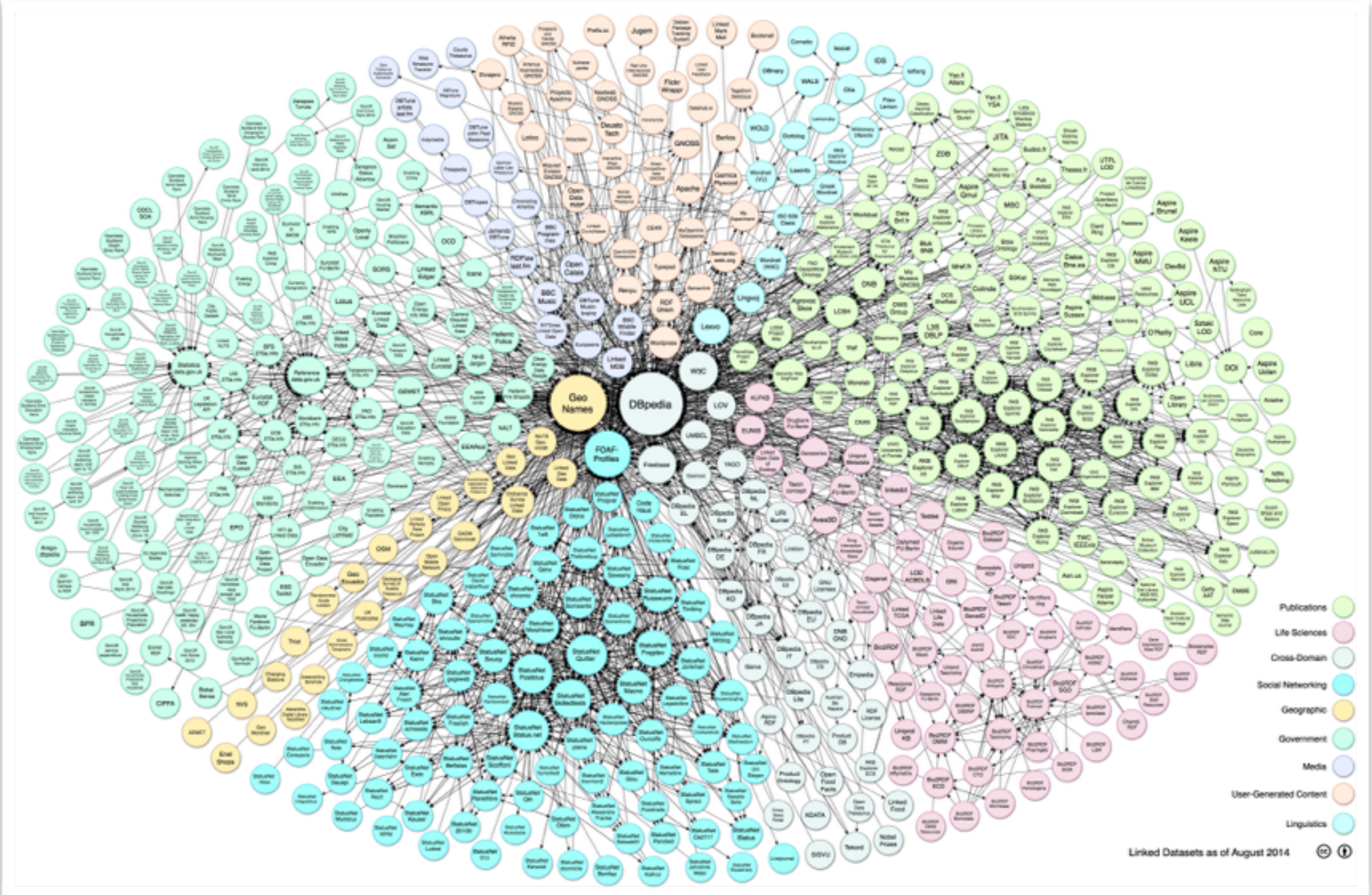
Freebase

- 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")
- Includes WordNet
- See http://www.mpi-inf.mpg.de/yago-naga/yago/

Linking Open Data (LOD)?

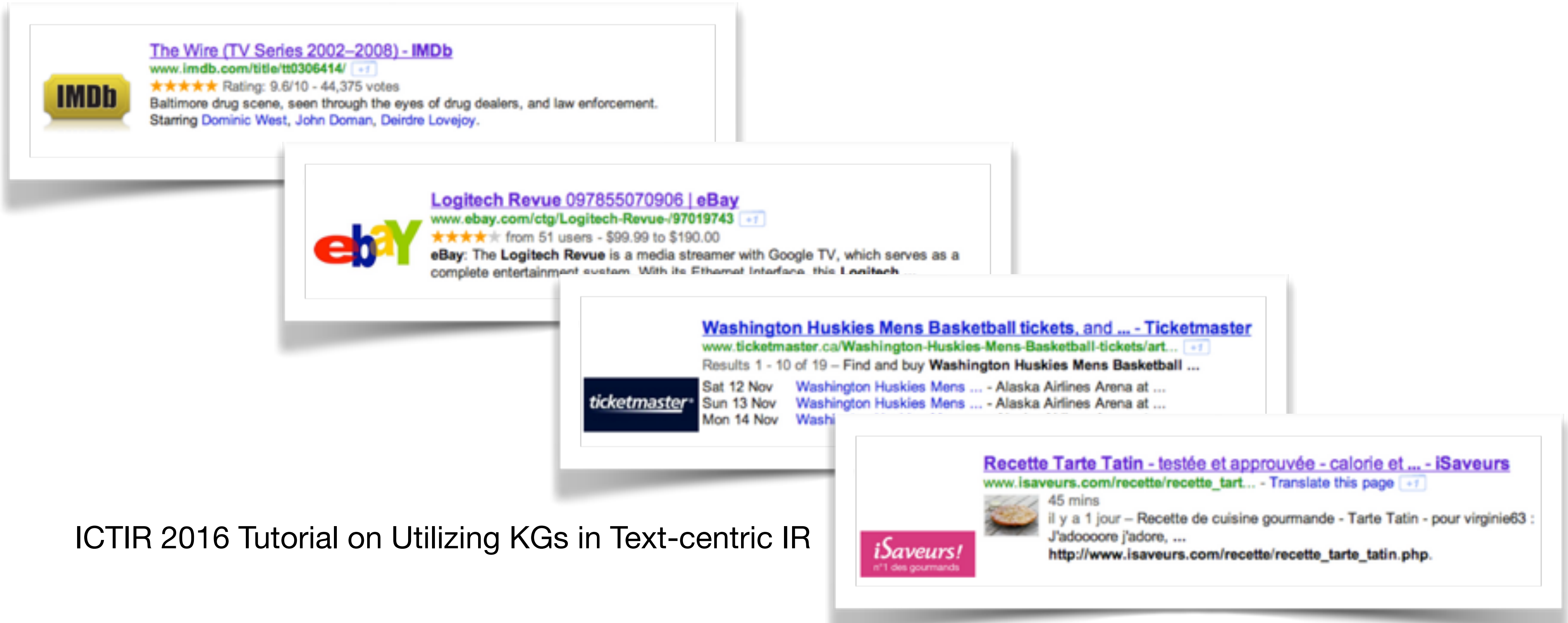


RDFa

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

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/
★★★★★ 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
★★★★☆ 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...
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... - Translate this page
45 mins
il y a 1 jour – Recette de cuisine gourmande - Tarte Tatin - pour virginie63 :
J'adoooore j'adore, ...
http://www.isaveurs.com/recette/recette_tarte_tatin.php.

Menu

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