



Making things findable

Peter Mika

Researcher and Data Architect

Yahoo! Inc.

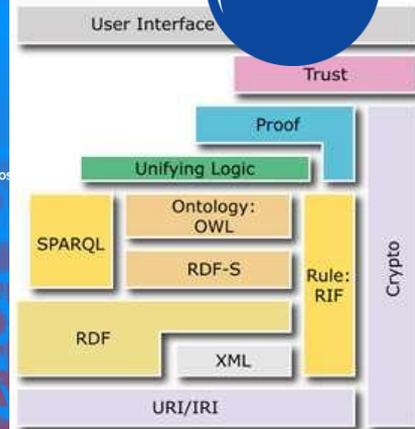
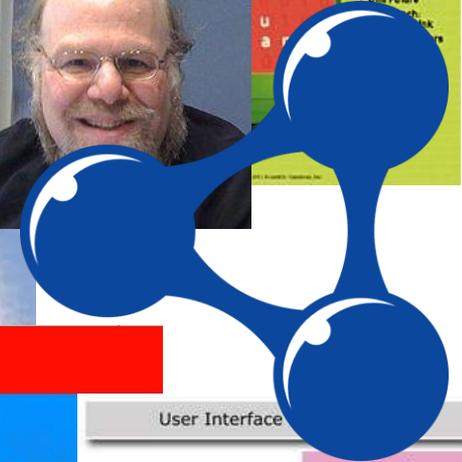
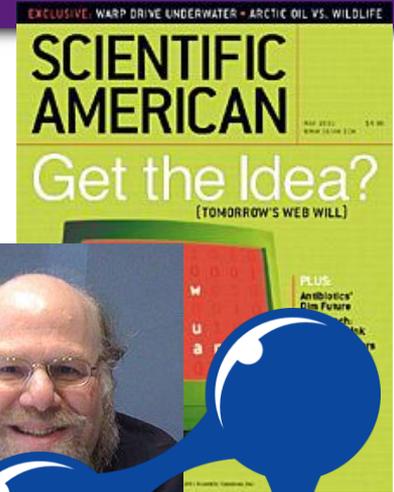
This was planned to be a presentation (only)
about Semantic Search...

But we are celebrating!



Anno 2001: The birth

- Scientific American article
- Semantic Web Working Symposium at Stanford
- Semantic Web standardization begins
- EU funding! DAML funding!
- The Semantic Web starts a career... and so *am I*...



ON
knowledge
TO



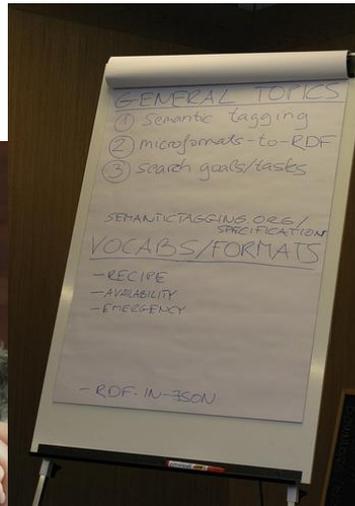
Anno 2004-2006: Reality sets in

- Agents? Semantic Web Services? Common sense?
- Engineers are not logicians
- Humans will have to do (most of) the work
- Two separate visions
- No one actually invests... other than the EU
- Where are the ontologies?
- Bad reputation for SW research
- The Semantic Web is looking for a job.
and so am I ☹



Anno 2007: A second chance

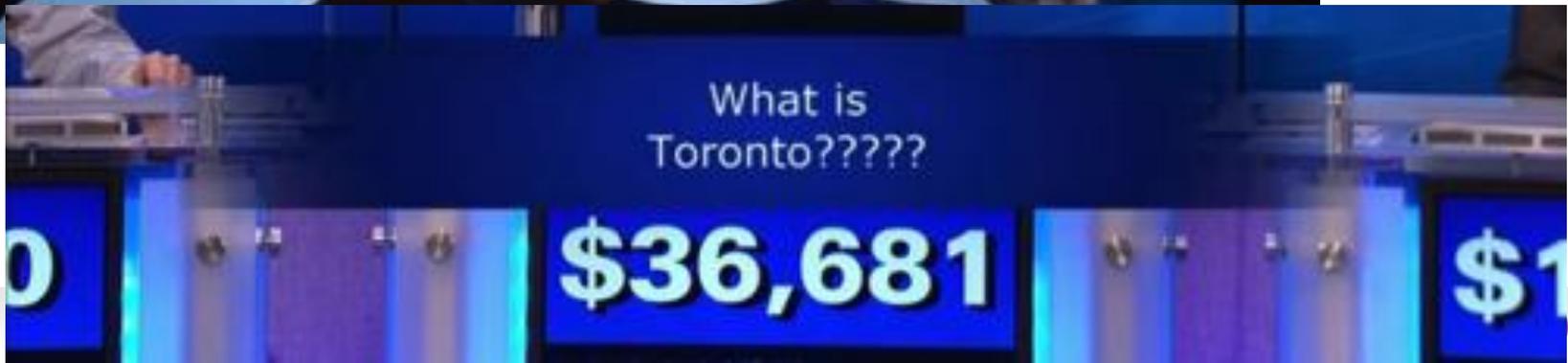
- Data first, schema second...logic third
- Linked Data.... billions of triples
- A sense of community
- Some ontologies
- We get the standards we need
- Startups, tool vendors appear
- SemTech
- The Semantic Web is in business...
and so am I!





Meanwhile, in Search...

Search is really fast, without necessarily being intelligent



Why Semantic Search? Part I

- Improvements in IR are harder and harder to come by
 - Machine learning using hundreds of features
 - Text-based features for matching
 - Graph-based features provide authority
 - Heavy investment in computational power, e.g. real-time indexing and instant search
- Remaining challenges are not computational, but in modeling user cognition
 - Need a deeper understanding of the query, the content and/or the world at large
 - *Could Watson explain why the answer is Toronto?*

Poorly solved information needs

- Multiple interpretations
 - paris hilton
- Long tail queries
 - george bush (and I mean the beer brewer in Arizona)
- Multimedia search
 - paris hilton sexy
- Imprecise or overly precise searches
 - jim hendler
 - pictures of strong adventures people
- Searches for descriptions
 - countries in africa
 - 32 year old computer scientist living in barcelona
 - reliable digital camera under 300 dollars

Poorly solved information needs

- Multiple interpretations
 - paris hilton
- Long tail queries
 - george bush (and I mean the beer brewer in
- Multimedia search
 - paris hilton sexy
- Imprecise or overly precise searches
 - jim hendler
 - pictures of strong adventures people
- Searches for descriptions
 - countries in africa
 - 32 year old computer scientist living in barcelona
 - reliable digital camera under 300 dollars

Many of these queries would not be asked by users, who learned over time what search technology can and can not do.

Dealing with sparse collections

Hi, **Guest** | [Sign In](#) | [Help](#) Bucket: VIP004 ▾ highconf model ... bad results / ads or bugs? tell us! [hide] [Make Yahoo! your homepage](#) | [Mail](#)

YAHOO! [Web](#) [Images](#) [Video](#) [Local](#) [Shopping](#) [News](#) [More ▾](#)

[Search](#) [Options ▾](#)

Also try: [rio blanco county](#), [rio blanco herald times](#), [more...](#)

[Search Pad](#)

[SearchScan - On](#)

303,000 results for **roi blanco**:

[Show All](#)

[Facebook](#)

[LinkedIn](#)

[Amazon.com](#)

[Sponsor Results](#)

Rio Blanco Hotel Sponsored Results

Visiting Ocho **Rios**? Read Reviews and Book **Rio Blanco** Hotel.
[TripAdvisor.co.uk/rioblancohotel](#)

Cheap Blanco Sinks Sponsored Results

Great Discounts on **Blanco** Sinks. Up To 50% Off **Blanco** Sinks.
[Sinks-Taps.clickcompare.info](#)

Roi Blanco - HomePage

Phone contact: Voice: +34 981 167 000 ext. 1276 Fax: +34 981 167 160
[www.dc.fi.udc.es/~roi](#) - [Cached](#)

Roi Blanco - Spain | LinkedIn

Experience: Researcher, Yahoo! Research; Visitor, University of Glasgow
[www.linkedin.com/pub/roi-blanco/8/58/2b0](#) - [Cached](#)

Roi Blanco | Facebook

Roi Blanco is on Facebook. Join Facebook to connect with **Roi Blanco** and others you may know. Facebook gives people the power to share and makes the world more open and ...
[www.facebook.com/roicho](#) - [Cached](#)

Roi Blanco - ACM author profile page

[www.acm.org](#) - The premier society in computing brings you the Computer Portal.
[portal.acm.org/author_page.cfm?id=81100367648](#) - [Cached](#)

DBLP: Roi Blanco

2010; 21 : **Roi Blanco**, Hugo Zaragoza: Finding support sentences for entities. SIGIR 2010: 339-346: 20 : Gianluca Demartini, Malik Muhammad Saad Missen, **Roi Blanco**, Hugo Zaragoza ...
[www.informatik.uni-trier.de/.../a-tree/b/Blanco:Roi.html](#) - [Cached](#)

Rio de Janeiro Holidays from £1374 Sponsored Results

Rio, Iguazu Falls, Buenos Aires. 11 Days, all flights & hotels.
[www.LlamaTravel.com/rio-de-janeiro](#)

Rio Blanco Hotel

Great **Rio blanco** hotel Deals
[Expedia.co.uk](#)

Rio

Great Deals on Used Cars. Many Models & Competitive Offers.
[www.ciao.co.uk/kiario](#)

All Mayfair Taps & Showers

Low prices and fast delivery, UK & Roi, 0044 (0) 20 8506 9725
[Showers-Direct2u.co.uk/mayfair](#)

Blanco at Amazon.co.uk

Low prices on **Blanco**. Free UK Delivery on Amazon Orders
[Amazon.co.uk](#)

Blanco - UK Sale

Save on all top **Blanco** products & more. Compare prices & order now

Dealing with sparse collections

Hi, **Guest** | Sign In | Help

Bucket: VIP004 ▾ highconf model ... bad res

Web Images Video Local Shopping

YAHOO!

roi blanco

Search

Options ▾

Also try: [rio blanco county](#), [rio blanco her](#)

Rio Blanco Hotel
Visiting Ocho Rios? Read Reviews and Book Rio Bla
[TripAdvisor.co.uk/riobl](#)

Cheap Blanco Sink
Great Discounts on Blai
[Sinks-Taps.clickcomp](#)



Roi Blanco - Home
Phone contact: Voice: +
[www.dc.fi.udc.es/~roi](#)

Roi Blanco - Spain
Experience: Researcher
[www.linkedin.com/pub/roi-blanco/8/58/2b0](#) - [Cached](#)

Roi Blanco | Facebook
Roi Blanco is on Facebook. Join Facebook to connect with Roi Blanco and others you may know. Facebook gives people the power to share and makes the world more open and ...
[www.facebook.com/roicho](#) - [Cached](#)

Roi Blanco - ACM author profile page
[www.acm.org](#) - The premier society in computing brings you the Co
[portal.acm.org/author_page.cfm?id=81100367648](#) - [Cached](#)

DBLP: Roi Blanco
2010; 21 : Roi Blanco, Hugo Zaragoza: Finding support sentences
2010; 339-346: 20 : Gianluca Demartini, Malik Muhammad Saad Mi
Hugo Zaragoza ...
[www.informatik.uni-trier.de/.../a-tree/b/Blanco:Roi.html](#) - [Cached](#)



Sorted Results



Rio Blanco Hotel
Great Rio blanco hotel Deals
[Expedia.co.uk](#)



Rio
Gre
Y
...

All Mayfair Taps & Showers
Low prices and fast delivery, UK &
Roi, 0044 (0) 20 8506 9725
[Showers.Direct2u.co.uk/mayfair](#)

Note: don't solve the sparsity problem where it doesn't exist

Contextual Search: content-based recommendations

Hi, Peter | Sign Out | Help

Yahoo! Mail

YAHOO! NEWS

Search

Web Search

HOME U.S. BUSINESS WORLD ENTERTAINMENT SPORTS TECH POLITICS SCIENCE HEALTH OPINION MOST POPULAR

World Video Middle East Europe Latin America Africa Asia Canada Australia/Antarctica Kevin Sites

Search All News News Search

TRENDING NOW: wikileaks college football manny pacquiao black friday lebron james

US, SKorea launch war games in tense Yellow Sea

Associated Press

Buzz up! 1178 votes | Share 1030 | retweet 169 | Email | Print

By DAVID GUTTENFELDER and JEAN H. LEE, Associated Press – 1 hr 39 mins ago



AFP – US and S.Korea plan naval show of force off tense peninsula

YEONPYEONG ISLAND, South Korea – A U.S. supercarrier and South Korean destroyer took up position in the tense Yellow Sea on Sunday for joint military exercises that were a united show of force just days after a deadly North Korean artillery attack.

As tensions escalated across the region, with North Korea threatening another "merciless" attack, China belatedly jumped into the fray. Beijing's top nuclear envoy, Wu Dawei, called for an emergency meeting in early December among regional powers involved in nuclear disarmament talks, including North Korea.



Slideshow: North Korea fires on South Korean island



North Korea Video: U.S.-Korea military drills start Reuters



North Korea Video: Tension in Korea peninsula Australia 7 News



nuclear disarmament talks

Search

[China calls for emergency talks on N.Korea](#)
China called Sunday for an emergency meeting in early December of envoys to the six-nation talks on North Korean nuclear...
<http://news.yahoo.com/>

[China calls for urgent talks on North Korea](#)
China quickened its diplomatic efforts to ease tensions between North and South Korea, calling Sunday for an emergency meeting...
<http://news.yahoo.com/>

More >>

YAHOO! NEWS

1 stock to buy tomorrow!
We have tomorrow's big winner - Join our Free New sletter today!
www.AwesomePennyStocks.com

VISIT SITE

Top Stories

- US, SKorea launch war games in tense Yellow Sea
- Ireland says bailout talks complete, EU to OK deal
- FBI: OR bomb suspect spoke of best site for blast
- US asks WikiLeaks to halt document release
- Egypt picks parliament after roundup of opposition
- Rio police claim victory after invading gang haven
- Election in Haiti beset by cholera, confusion
- 3 missing Mich. boys feared to be in danger
- No. 5 Wisconsin routs NU 70-23 for B10 title share
- 5 rare Sumatran elephants found dead in Indonesia

More Top Stories >

AdChoices

> Learn more

* Roll over for important details.



Contextual Search: content-based recommendations

Hi, Peter | Sign Out | Help

Yahoo! Mail

YAHOO! NEWS

Search [] Web Search

HOME U.S. BUSINESS WORLD ENTERTAINMENT SPORTS TECH POLITICS SCIENCE HEALTH OPINION MOST POPULAR

World Video Middle East Europe Latin America Africa Asia Canada Australia/Antarctica Kevin Sites

Search All News News Search TRENDING NOW: wikileaks college football manny pacquiao black friday lebron james

US, SKorea launch war games in tense Yellow Sea

Associated Press Buzz up! 1178 votes Share 1030 retweet 169 Email Print



AFP - US and S.Korea plan naval show of force off tense peninsula

By DAVID GUTTENFELDER and JEAN H. LEE, Associated Press - 1 hr 39 mins ago

YEONPYEONG ISLAND, South Korea - A U.S. supercarrier and South Korean destroyer took up position in the tense Yellow Sea on Sunday for joint military exercises that were a united show of force just days after a deadly North Korean artillery attack.

As tensions escalated across the region, with North Korea threatening another "merciless" attack, China belatedly jumped into the fray. Beijing's top nuclear envoy, Wu Dawei, called for an emergency meeting in early December among regional powers involved in nuclear disarmament talks, including North Korea.

- Slideshow: North Korea fires on South Korean island
- North Korea Video: U.S.-Korea military drills start Reuters
- North Korea Video: Tension in Korea peninsula Australia 7 News



Top Stories

US, SKorea launch war games in tense Yellow Sea

Hovering over an underlined phrase triggers a search for related news items.

nuclear disarmament talks Search

[China calls for emergency talks on N.Korea](#)
China called Sunday for an emergency meeting in early December of envoys to the six-nation talks on North Korean nuclear...
<http://news.yahoo.com/>

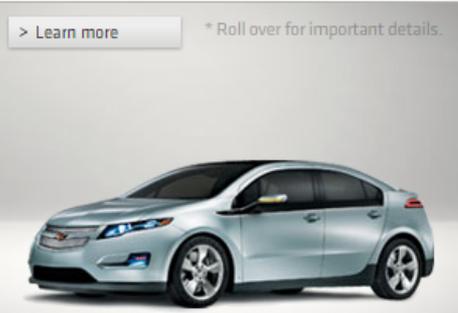
[China calls for urgent talks on North Korea](#)
China quickened its diplomatic efforts to ease tensions between North and South Korea, calling Sunday for an emergency meeting...
<http://news.yahoo.com/>

More >> YAHOO! NEWS

1 stock to buy tomorrow! We have tomorrow's big winner - Join our Free New sletter today!
www.AwesomePennyStocks.com VISIT SITE

AdChoices

Learn more Roll over for important details.



Contextual Search: personalization

Web Images Video Local Shopping More the web UK only

YAHOO!

Y! UK & IE | My Yahoo! | Get Yahoo! on your phone

Hi, Peter

YAHOO! SITES

- Mail
- Finance (Dow Jones)
- Flickr
- Games
- Messenger
- Movies
- Music
- TV
- Weather (4°C)

More Yahoo! Sites

MY FAVORITES

- Facebook
- All Things Digital
-

TODAY - November 28, 2010



Acrobatic leap leads to stunning play

Oklahoma State's Broderick Brown vaults far out of bounds to reach a ball, and then surprises everyone. [Watch >>](#)

- Hogs make BCS case
- Conference races
- College scoreboard



Catch may be play of the year



Who's that with Kim Kardashian?



Moves that make you look cheap



Rare bears brought to light

1 - 4 of 24

NEWS WORLD LOCAL FINANCE

- U.S., S. Korea take up position in tense Yellow Sea
- Haiti presidential election begins amid cholera, confusion
- U.S. asks whistleblower WikiLeaks to halt document release
- Rio de Janeiro police claim victory after invading gang haven

TRENDING NOW

- | | |
|------------------------|--------------------------|
| 1. Jessica Simpson | 6. Sarah Palin |
| 2. Anne Hathaway | 7. Christina Aguilera... |
| 3. Dwayne Johnson | 8. Yao Ming |
| 4. Online photo alb... | 9. Debt management |
| 5. Flu symptoms | 10. Medicare |

As a reward, we're giving them a gift of even more minutes.



£10

For £10, get...

Latest video picks



Contextual Search: personalization

The image shows a screenshot of the Yahoo! homepage from November 28, 2010. The page features a search bar at the top, a navigation menu on the left, and a main content area with a featured story and a list of related items. Two callout boxes provide context:

- Machine Learning based 'search' algorithm selects the main story and the three alternate stories based on the users demographics (age, gender etc.) and previous behavior.** This callout points to the featured story about Oklahoma State's Broderick Brown.
- Display advertising is a similar top-1 search problem on the collection of advertisements.** This callout points to a Giffgaff advertisement for a £10 gift.

The main content area includes a featured story titled "Acrobatic leap leads to stunning play" about Oklahoma State's Broderick Brown, and a list of related items such as "Catch may be play of the year", "Who's that with Kim Kardashian?", "Moves that make you look cheap", and "Rare bears brought to light". The page also features a "Web Search" button and a "Web Search" button.

Contextual Search: new devices

Show related content

YAHOO! CONNECTED TV

Poll: "Who Will Win Tonight?"	
SHOWTIME BOXING	
Who will win the WBO Featherweight Title tonight?	
"JuanMa" Lopez	59%
Rafael Marquez	39%
Draw	2%

Who will win the fight tonight? Vote now! Press [button] on your TV remote

SHOWTIME



Contextual Search: new devices

Show related content

YAHOO! CONNECTED TV

Who will win the fight tonight? Vote now! Press [button] on your TV remote

WBO Championship BOXING

Who will win the WBO Featherweight Title tonight?

"JuanMa" Lopez	59%
Rafael Marquez	39%
Draw	

ROUND 3 OF 12

O.N

Connect to friends watching the same



Aggregation across different dimensions

Web Images Video Local Shopping More ▾

YAHOO!

Web Search

Yahoo.com Mail My Yahoo!

Hi, serendipity588 | Sign Out

BROOKLYN HEIGHTS Including Downtown change ▾

Top Picks News Things to Do Deals

The best news, deals and things to do around you

Brooklyn Heights Blog

News »

Love Lane Mews: Gordon Gekko Meets Mary Poppins

4 hrs ago

The Wall Street Journal recently caught up with Love Lane Mews architect Stephen B. Jacobs. The takeaway — the new condos there will be “supercalafragalisticexpialadoshus” meets “greed is good”: WSJ: But when it comes to preservation, the...

Share Comment Like

Dumbo NYC

News »

Dumbo Links Week of 21Nov2010



18 hrs ago

The following are selected links from this past week on blogs and websites with discussion about Dumbo (and its neighboring areas): More Apple Talk in Dumbo?, 24Nov10, Brownstoner 30 Washington St Two FI Extension?, 22Nov10, Brownstoner Toll Bros...

Share Comment Like

Brooklyn Heights Blog

News »



Contribute to Brooklyn Heights ▾

ADD AN EVENT GET INVOLVED
SUBMIT A LOCAL SOURCE WRITE

Posts From Community Writers



Charles Manley wrote an article:
Aeropostale Opens on Newly Paved
Fulton Street in Downtown Brooklyn



Lucrezia Wise wrote an article:
Sing Along at the Jingle Bell
Jamboree!

About this experience >

Send us feedback >

Aggregation across different dimensions

Web Images Video Local Shopping More ▾

YAHOO![®]

Web Search

Yahoo.com Mail My Yahoo!

Hi, serendipity588 | Sign Out

BROOKLYN HEIGHTS Including Downtown change ▾

Top Picks News Things to Do Deals

The best news, deals and things to do around you

Brooklyn Heights Blog

Love Lane Mews: Gordon Gekko Meets Mary Poppins

4 hrs ago

The Wall Street Journal recently caught up with Love Lane Mews architect Stephen... new condos there will be "supercalafragalisticexpialadoshus" meets "greed is go... preservation, the...

Share Comment Like

Dumbo NYC

Dumbo Links Week of 21Nov2010



18 hrs ago

The following are selected links from this past week on blogs and websites with discussion about Dumbo (and its neighboring areas): More Apple Talk in Dumbo?, 24Nov10, Brownstoner 30 Washington St Two FI Extension?, 22Nov10, Brownstoner Toll Bros...

Share Comment Like

Brooklyn Heights Blog

Hyperlocal: showing content from across Yahoo that is relevant to a particular neighbourhood.

38°F
Fair

Route to Brooklyn Heights ▾

EVENT GET INVOLVED
SUBMIT A LOCAL SOURCE WRITE

Posts From Community Writers

Charles Manley wrote an article: **Aeropostale Opens on Newly Paved Fulton Street in Downtown Brooklyn**

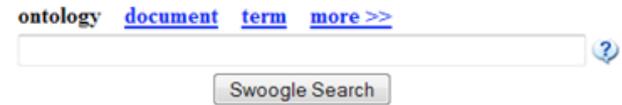
Lucrezia Wise wrote an article: **Sing Along at the Jingle Bell Jamboree!**

About this experience >

Send us feedback >

Why Semantic Search? Part II

- The Semantic Web is now a reality
 - Billions of triples
 - Thousands of schemas
 - Varying quality
- End users
 - Keyword queries, not SPARQL
- Searching data instead or in addition to searching documents
 - Providing direct answers (possibly with explanations)
 - Support for novel search tasks



Searching over 10,000 ontologies

Direct answers in search

Hi, serendipity588 | Sign Out | Help | Bucket: VIP038 | Quickkapps YAP -... bad results / ads or bugs? tell us! [hide] | Make Yahoo! your homepage | Mail

Web Images Video Local Shopping News More ▾

YAHOO!

Search Pad | SearchScan - On | 17,400,000 results for **vienna, austria**

Points of interest in Vienna, Austria

- Maria am Ges...
- Kunsthistorisch...
- Hundertwasserha...
- Hawelka
- Spanish Riding ...
- Alsergrund
- Kahlenberg
- Schatzkammer

Weather vienna austria, map of vienna austria, more...

Vienna, Austria - Visitor Guide
travel.yahoo.com
Hotels | Restaurant Guide | Flights | Map

Top Rated Things To Do (176)

- Stephansdom
- Schönbrunner Schloßpark
- Church of St. Mary of the Angels

[More Things To Do...](#)

Vienna - Wikipedia, the free encyclopedia
Name | History | Geography and... | Districts and...
Vienna lies in the east of Austria and is close to the borders of the Czech Republic, Slovakia, and Hungary. These regions work together in a European Centrope border region. Along with...
en.wikipedia.org/wiki/Vienna - [Cached](#)

Vienna Austria - Welcome
Vienna is the capital of Austria. Informations about hotel, restaurant, bed and breakfast, weather, map, travel, tourism and much more ...
www.viennaaustria.co.uk - [Cached](#)

Vienna Austria Tourist Information and Vacation Guide
Tourist attractions, museums, art galleries, public transit, airport information, train information, weather, maps, hotels and other accommodation in Vienna Austria
viennaaustria.ca - [Cached](#)

APO Europe holds regional conference in Vienna ...
VIENNA, Austria - The Alpha Phi Omega (APO) in Europe held its 2nd regional conference in Vienna from October 30 to November 1, 2010. APO members from all over Europe ...
www.abs-cbnnews.com/global-filipino/11/03/10/apo-europe... - [Cached](#)

Vienna, Austria - Image Results

Information box with content from and links to Yahoo! Travel

Vienna
Fantas... on art and prints. Feed your... on eBay.co.uk.
www.ebay.co.uk/art

Vienna, Austria
Check Rates at Over 360 Hotels in Vienna. Compare Top Sites & Save
BookingBuddy.co.uk

200 Hotels in Vienna
Book your hotel in Vienna online. Find your hotel on a city map.
www.booking.com/vienna-hotels

Fly to Austria
At Opodo™ we search 500 airlines to find cheap flights to Austria.
Opodo.co.uk/austria

Vienna Austria What to Do
Find and book 55+ things to do in Vienna on Viator.
www.viator.com/vienna

Hotel Deals in Vienna, Austria
Discounts up to 70% on Vienna hotels in Austria.
travel.hotels-and-discounts.com

Hotel Austria
Visiting Vienna? Read Reviews and



Direct answers in search

blu ray player

Search

Options

Also try: [blu ray player reviews](#), [best blu ray player](#), [more...](#)

Samsung blu-ray player

A Samsung Blu-Ray Player Is A Great Gift To Give. Shop Best Buy® Now!
www.BestBuy.com

blu ray player

Exclusive Deals On 3D-Ready Blu-ray Players. Official Sony Site
www.SonyStyle.com/Blu-Ray

More Sponsors: [blu ray player](#), [blu ray dvd](#), [blu ray](#), [blu ray disc player](#)

Top Blu Ray Players

 Samsung BD-C6900 Blu-ray ... ★★★★★ \$197.77 - \$349.99	 LG BX580 Blu-ray Disc ... ★★★★★ \$169 - \$287	 Panasonic DMP-BD65 ... ★★★★★ \$99 - \$179.95	 Panasonic DMP-BD65 ... ★★★★★ \$179.99	
 Samsung BD-P3600 Blu-ray ... ★★★★★ (5) \$169.96 - \$299.99	 Samsung BD-C7500 Blu-ray ... ★★★★★ \$179.99 - \$349.98	 Sony BDP-S500 Blu-ray ... ★★★★★ (8) \$605.09	 Sharp BD-HP20U Blu-ray ... ★★★★★ (12) \$89.99	 LG BD530 Blu-ray Player ★★★★★ \$79.99

More Blu Ray Players from Yahoo! Shopping »

Products from Yahoo! Shopping

Since Aug, 2010, 'regular' search results are 'Powered by Bing'

Search Pad

SearchScan - On

17,400,000 results for vienna, austria

Related Points of Interest



Maria am Ge



Kunsthistoris



Hundertwass



Hawelka



Spanish Ridi



Alsergrund



Kahlenberg



Schatzkamr

Search Pad

SearchScan - On

839,000 results for blu ray player:

Show All

Best Buy

Wikipedia

CNET Reviews

Shopping Sites

Sponsored Results

Related Searches

blu ray software

best buy

playstation 3

Blu-Ray Player at Amazon

Big Savings on Blu-Ray Player. Free Shipping on Qualified Orders
www.Amazon.com/electronics

Shop Target.com

Find Great Deals Online. Free Shipping on 100,000 Items!
Target.com/FreeShipping

Buy.com® Blu Ray Players

This Week Only - Holiday Savings on Blu Ray Players! Most Ship Free.
www.Buy.com/Electronics

Sears® Electronics Sale

Get More this Holiday Season with Your Way Rewards. Save Now!
www.Sears.com/Shop-Your-Way

Netflix Official Site

Blu-ray Movies \$0.32 Each

Novel search tasks

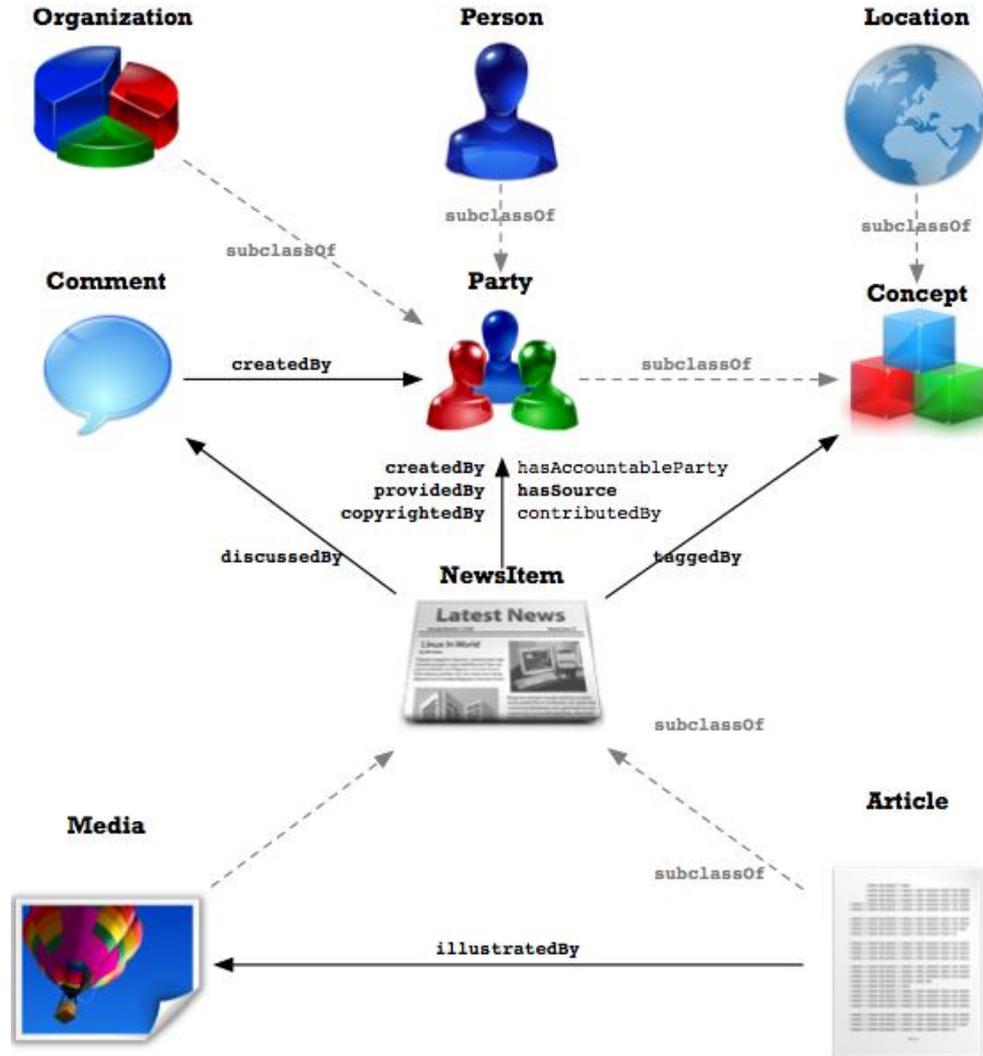
- Aggregation of search results
 - e.g. price comparison across websites
- Analysis and prediction
 - e.g. world temperature by 2020
- Semantic profiling
 - recommendations based on particular interests
- Semantic log analysis
 - understanding user behavior in terms of objects
- Support for complex tasks (search apps)
 - e.g. booking a vacation using a combination of services

Why Semantic Search? Part III

- There is a model
 - Publishers are (increasingly) interested in making their content searchable, linkable and easier to aggregate and reuse
 - So that social media sites and search engines can expose their content to the right users, in a rich and attractive form
- This is a about creating an ecosystem...
 - More advanced in some domains
 - In others, we still live in the tyranny of rate-limited APIs, licenses etc.

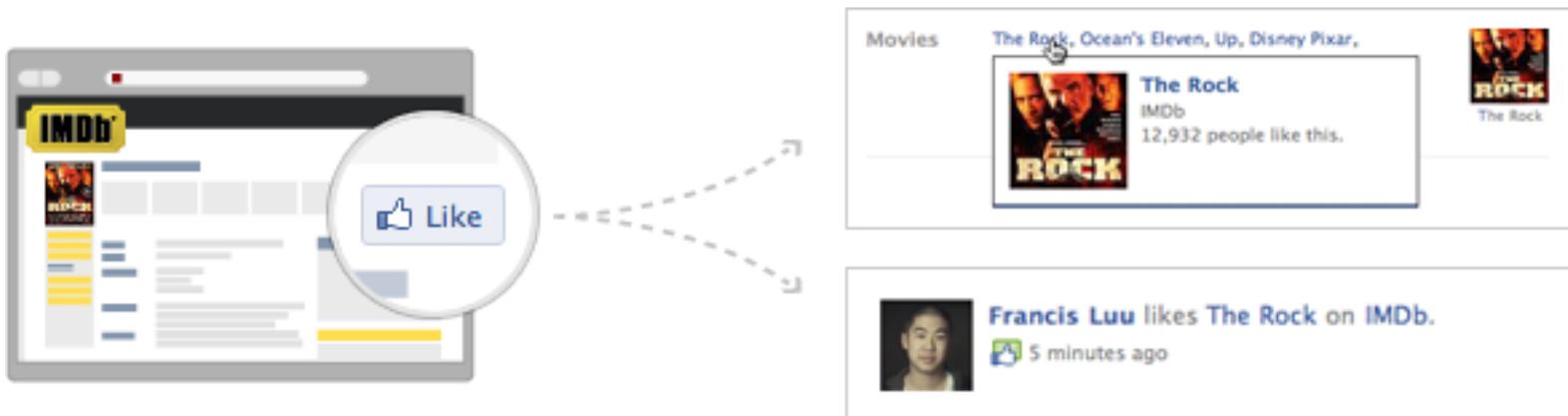
Example: rNews

- RDFa vocabulary for news articles
 - Easier to implement than NewsML
 - Easier to consume for news search and other readers, aggregators
- Under development at the IPTC
 - March: v0.1 approved
 - Final version by Sept



Example: Facebook's Like and the Open Graph Protocol

- The 'Like' button provides publishers with a way to promote their content on Facebook and build communities
 - Shows up in profiles and news feed
 - Site owners can later reach users who have liked an object
 - Facebook Graph API allows 3rd party developers to access the data
- Open Graph Protocol is an RDFa-based format that allows to describe the object that the user 'Likes'



Example: Facebook's Open Graph Protocol

- RDF vocabulary to be used in conjunction with RDFa
 - Simplify the work of developers by restricting the freedom in RDFa
- Activities, Businesses, Groups, Organizations, People, Places, Products and Entertainment
- Only HTML <head> accepted
- <http://opengraphprotocol.org/>

```
<html xmlns:og="http://opengraphprotocol.org/schema/">
<head>
  <title>The Rock (1996)</title>
  <meta property="og:title" content="The Rock" />
  <meta property="og:type" content="movie" />
  <meta property="og:url"
content="http://www.imdb.com/title/tt0117500/" />
  <meta property="og:image" content="http://ia.media-
imdb.com/images/rock.jpg" /> ...
</head> ...
```



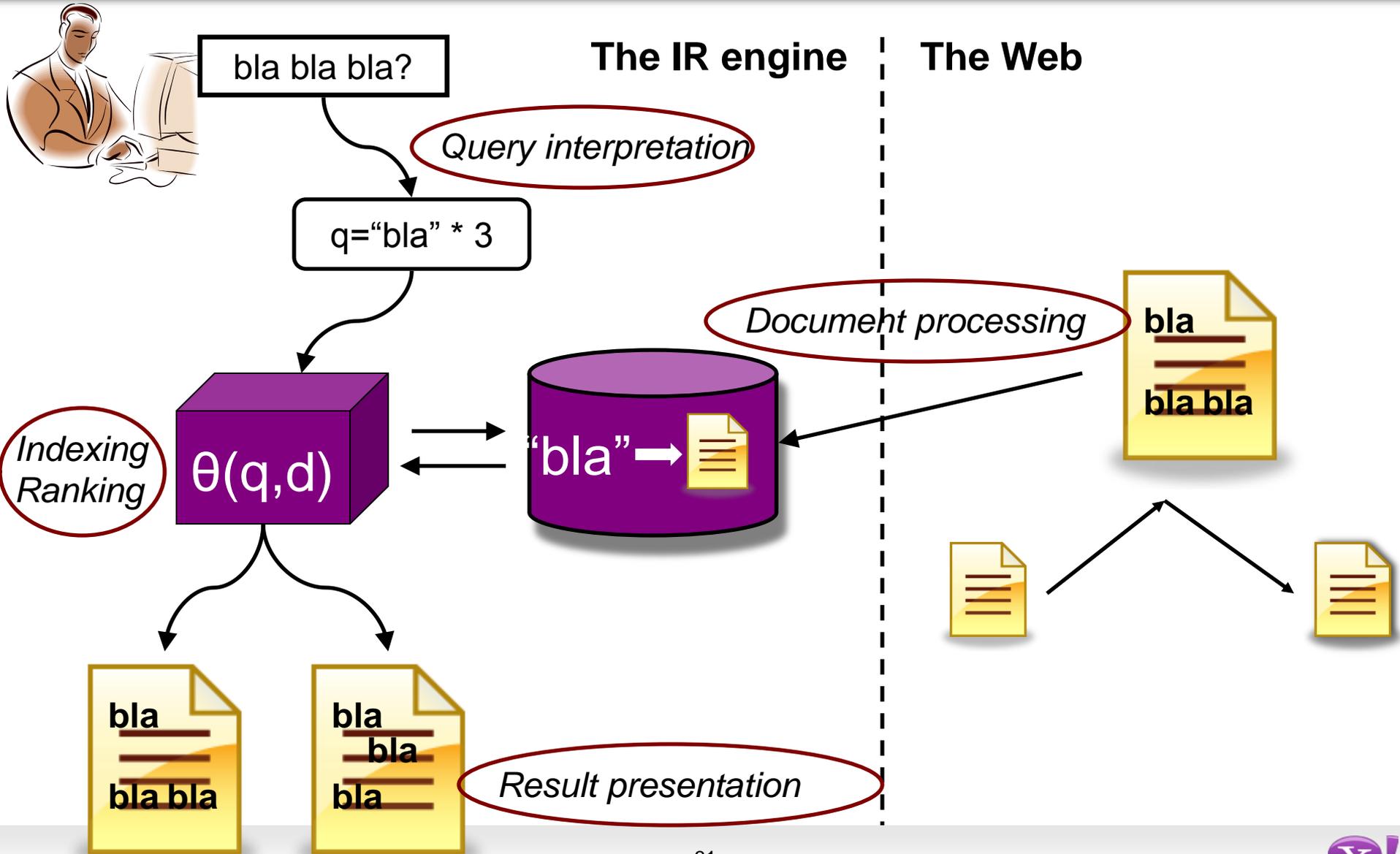


Semantic Search

Semantic Search: a definition

- Semantic search is a retrieval paradigm that
 - **Makes use of the structure of the data or explicit schemas to understand user intent and the meaning of content**
 - **Exploits this understanding at some part of the search process**
- Combination of document and data retrieval
 - Documents with metadata
 - Metadata may be embedded inside the document
 - *I'm looking for **documents** that mention countries in Africa.*
 - Data retrieval
 - Structured data, but searchable text fields
 - *I'm looking for **directors**, who have directed movies where the synopsis mentions dinosaurs.*

Semantics at every step of the IR process





Data on the Web

Data on the Web

- Data on the Web as a complement to professional data providers
- Most web pages are generated from databases, but the data is not always directly accessible
 - APIs offer limited views over data
- The structure and semantics (meaning) of the data is not directly accessible to search engines
- Two solutions
 - Extraction using **Information Extraction (IE)** techniques (implicit metadata)
 - Relying on publishers to expose structured data using standard **Semantic Web** formats (explicit metadata)

Information Extraction methods

- Natural Language Processing
- Extraction of triples
 - *Suchanek et al. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia, WWW, 2007.*
 - *Wu and Weld. Autonomously Semantifying Wikipedia, CIKM 2007.*
- Filling web forms automatically (form-filling)
 - *Madhavan et al. Google's Deep-Web Crawl. VLDB 2008*
- Extraction from HTML tables
 - *Cafarella et al. WebTables: Exploring the Power of Tables on the Web. VLDB 2008*
- Wrapper induction
 - *Kushmerick et al. Wrapper Induction for Information Extraction Text extraction. IJCAI 2007*

Semantic Web

- Sharing data across the Web
 - Publish information in standard formats (RDF, RDFa)
 - Share the meaning using powerful, logic-based languages (OWL, RIF)
 - Query using standard languages and protocols (HTTP, SPARQL)
- Two main forms of publishing
 - Linked Data
 - Data published as RDF documents linked to other RDF documents and/or using SPARQL end-points
 - Community effort to re-publish large public datasets (e.g. Dbpedia, open government data)
 - RDFa
 - Data embedded inside HTML pages
 - Recommended for site owners by Yahoo, Google, Facebook

RDFa: metadata embedded in HTML

Roi's homepage

...

```
<p typeof="foaf:Person" about="http://example.org/roi">
```

```
<span property="foaf:name">Roi Blanco</span>.
```

```
<a rel="owl:sameAs" href="http://research.yahoo.com/roi"> Roi Blanco </a>.
```

```
You can contact him at <a rel="foaf:mbox" href="mailto:roi@yahoo-inc.com">
```

via email .

```
</p>
```

...

Crawling the Semantic Web

- Linked Data
 - Similar to HTML crawling, but the the crawler needs to parse RDF/XML (and others) to extract URIs to be crawled
 - [Semantic Sitemap/VOID](#) descriptions
- RDFa
 - Same as HTML crawling, but data is extracted after crawling
 - *Mika et al. Investigating the Semantic Gap through Query Log Analysis, ISWC 2010.*
- SPARQL endpoints
 - Endpoints are not linked, need to be discovered by other means
 - [Semantic Sitemap/VOID](#) descriptions

Data fusion

- Ontology matching
 - Widely studied in Semantic Web research, see e.g. list of publications at ontologymatching.org
 - Unfortunately, not much of it is applicable in a Web context due to the quality of ontologies
- Entity resolution
 - Logic-based approaches in the Semantic Web
 - Studied as record linkage in the database literature
 - Machine learning based approaches, focusing on attributes
 - Graph-based approaches, see e.g. the work of [Lisa Getoor](#) are applicable to RDF data
 - Improvements over only attribute based matching
- Blending
 - Merging objects that represent the same real world entity and reconciling information from multiple sources

Data quality assessment and curation

- Heterogeneity, quality of data is an even larger issue
 - Quality ranges from well-curated data sets (e.g. Freebase) to microformats
 - In the worst of cases, the data becomes a graph of words
 - Short amounts of text: prone to mistakes in data entry or extraction
 - Example: mistake in a phone number or state code
- Quality assessment and data curation
 - Quality varies from data created by experts to user-generated content
 - Automated data validation
 - Against known-good data or using triangulation
 - Validation against the ontology or using probabilistic models
 - Data validation by trained professionals or crowdsourcing
 - Sampling data for evaluation
 - Curation based on user feedback



Query Interpretation

Query interpretation in search

- Provide a higher level representation of queries in some conceptual space
 - Ideally, the same space in which documents are represented
 - Limited user involvement in the case document retrieval
 - Examples: search assist, facets
- Interpretation happens before the query is executed
 - Federation: determine where to send the query
 - Example: show business listings from Yahoo! Local for local queries
 - Ranking feature
 - Blend multiple possible interpretations of the same query
 - Deals with the sparseness of query streams
 - 88% of unique queries are singleton queries (Baeza et al)
 - Spell correction (“we have also included results for...”), stemming

Query interpretation in Semantic Search

- “Snap to grid”
 - Using the ontologies, a summary of the data or the whole data we find the most likely structured query matching the user input
 - Example: “starbucks nyc” -> company name:”Starbucks” in location:”New York City”
- Larger user involvement
 - Guiding the user in constructing the query
 - Example: Freebase Suggest
 - Displaying back the interpretation of the query
 - Example: TrueKnowledge



Indexing and Ranking

Indexing

- Search requires matching and ranking
 - Matching selects a subset of the elements to be scored
- The goal of indexing is to speed up matching
 - Retrieval needs to be performed in milliseconds
 - Without an index, retrieval would require streaming through the collection
- The type of index depends on the query language to support
 - SPARQL is a highly-expressive SQL-like query language for experts
 - DB-style indexing
 - End-users are accustomed to keyword queries with very limited structure (see Pound et al. WWW2010)
 - IR-style indexing

IR-style indexing

- Index data as text
 - Create virtual documents from data
 - One virtual document per subgraph, resource or triple
 - typically: resource
- Key differences to Text Retrieval
 - RDF data is structured
 - Minimally, queries on property values are required
- MapReduce is an ideal model for building inverted indices
 - Map creates (term, {doc1}) pairs
 - Reduce collects all docs for the same term: (term, {doc1, doc2...})
 - Sub-indices are merged separately
 - Term-partitioned indices
- *Peter Mika. Distributed Indexing for Semantic Search, SemSearch 2010.*

Ranking

- Previously, expert applications using specialized datasets
 - Queries given as logical formulas or highly selective DB-style queries
 - Expert users
 - Small, high quality datasets
 - Possible to give a precise answer (question-answering)
- Increasingly, end-user applications using Web data
 - Using keyword queries at least as a starting point
 - Non-expert users
 - Large datasets with potential mistakes
 - Not possible to give precise answers, only to provide relevant answers

Ranking methods

- The unit of retrieval is either an object or a sub-graph
 - The sub-graph induced by matching keywords in the query
- Ranking methods from Information Retrieval
 - TF-IDF, BM25, probabilistic methods (language models)
 - Methods such as BM25F allow weighting by predicate
- Machine learning is used to tune parameters and incorporate query-independent (static) features
 - Example: authority scores of datasets computed using PageRank (Harth et al. ISWC 2009)
- Additional topics
 - Relevance feedback and personalization
 - Click-based ranking
 - De-duplication
 - Diversification



Evaluation

Harry Halpin, Daniel Herzig, Peter Mika, Jeff Pound,
Henry Thompson, Roi Blanco, Thanh Tran Duc

Semantic Search evaluation at SemSearch 2010 and 2011

- Evaluation is a critical component in developing IR systems
- Keyword search over RDF data
 - Entity Search
 - Queries where the user is looking for a single real world object
 - *Pound et al. Ad-hoc Object Retrieval in the Web of Data, WWW 2010.*
 - List search (new in 2011)
 - Queries where the user is looking for a class of objects
- Focus on relevance, not efficiency
- Real queries and real data
 - Yahoo! and Microsoft query logs
 - Billion Triples Challenge 2009 dataset
- TREC style evaluation
 - Focusing on ranking, not question answering
 - Evaluated using Amazon's Mechanical Turk

Hosted by Yahoo! Labs at semsearch.yahoo.com

SEMSEARCH
CHALLENGE from **YAHOO!**
LABS

[Home](#) [Datasets](#) [Instructions](#) [Registration](#) [Submission](#) [FAQs](#)



SEMSEARCH CHALLENGE

IMPORTANT DATES

-  **February 15, 2011**
Registration Opens
-  **March 1, 2011**
Submission Opens
-  **March 21, 2011**
Submission Close
-  **March 29, 2011**
Workshop & Results

Be a part of the next wave of Web Search

The Web is rapidly evolving into a vast knowledge base of structured information, thanks to advances in technology and standards for data sharing, and the willingness of businesses and public organizations to publish information in an open, interoperable manner.

This emerging Web of Objects promises to herald a new area of Web Search, where search engines index specific items of information about objects, instead of matching keywords in text. These new types of search engines enable Semantic Search, i.e. giving precise answers to complex queries. In the context of the Web, such search engines help users with the additional challenge of consolidating information from multiple sources.

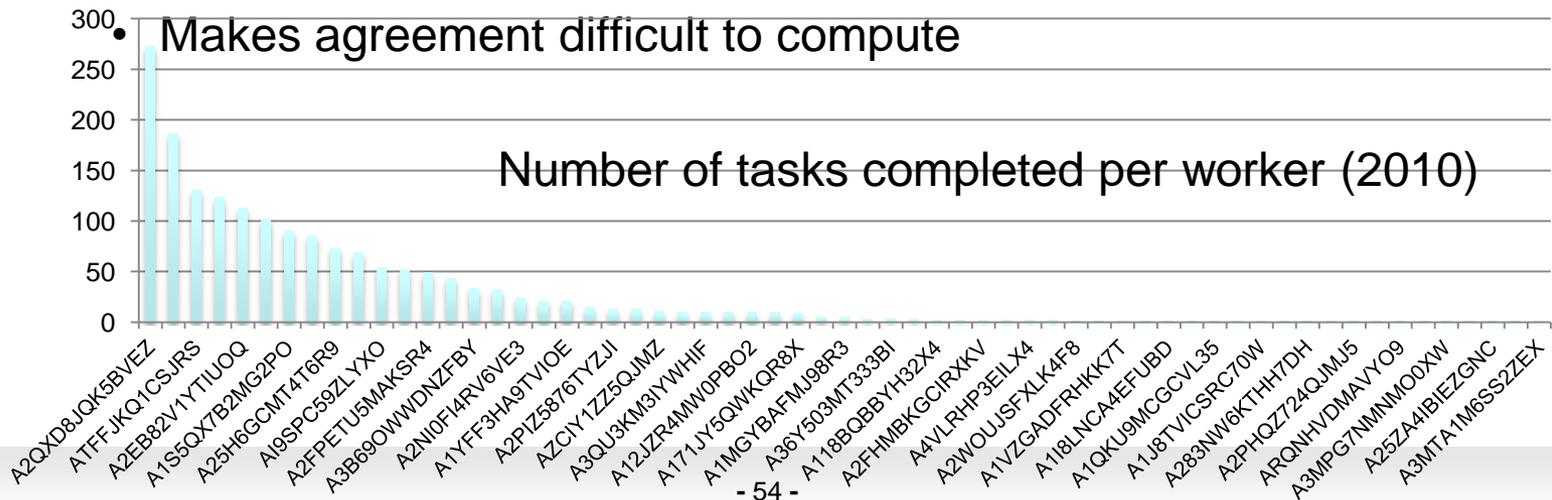
Organizing committee

- Roi Blanco (*Yahoo! Research*)
- Harry Halpin (*University of Edinburgh*)
- Daniel M. Herzig (*Karlsruhe Institute of Technology*)
- Peter Mika (*Yahoo! Research*)
- Henry S. Thompson (*University of Edinburgh*)
- Thanh Tran Duc (*Karlsruhe Institute of Technology*)



Assessment with Amazon Mechanical Turk

- Evaluation using non-expert judges
 - Paid \$0.2 per 12 results
 - Typically done in 1-2 minutes
 - \$6-\$12 an hour
 - Sponsored by the European SEALS project
 - Each result is evaluated by 5 workers
- Workers are free to choose how many tasks they do



Evaluation form

william penn university

Assess this search result for the above query:

property	value
name	William_Penn_University
label	William Penn University
type	University
type	University108286163
type	Thing
subject	Category:Council_of_Independent_Colleges
subject	Category:Mahaska_County%2C_Iowa
subject	Category:Universities_and_colleges_affiliated_with_the_Religious_Society_of_Friends
comment	William Penn University is a private, liberal arts university in Oskaloosa, Iowa, United States. It was founded by members of the Religious Society of Friends (Quakers) in 1873 as Penn College. In 1933, the name was changed to William Penn College, and finally to William Penn University in 2000. Ath. . .
sameAs	Mx4rv8bUMpwpEbGdrcN5Y29ycA
reference	http://www.wmpenn.edu/
page	William_Penn_University
wordnet_type	synset university noun 2

- Excellent - describes the query target specifically and exclusively
- Not bad - mostly about the target
- Poor - not about the target, or mentions it only in passing

Results

- 5-6 teams participated in both 2010 and 2011
 - Web queries and web data... difficult task!
- The Semantic Web is not (necessarily) all DBpedia
 - Though a search engine on Dbpedia only would have done well
- Computing agreement is difficult
 - Each judge evaluates a different number of items
 - Follow up experiments validated the Mechanical Turk approach
 - *Blanco et al. Repeatable and Reliable Search System Evaluation using Crowd-Sourcing, SIGIR2011*
- Result rendering is important
 - Influences the perception of relevance
 - No 'canonical' result rendering in Semantic Web search
- Assessments are made public
 - Measure your own system against the best of the field

Next steps

- New tasks
 - More complex queries (but finding them is difficult)
 - Retrieval on RDFa data
 - Aggregated search
 - Ranking properties of objects
- Achieve repeatability
 - Simplify our process and publish our tools
 - Automate as much as possible... except the Turks ;)
- Positioning compared to other evaluation campaigns
 - TREC Entity Track
 - Question Answering over Linked Data
 - SEALS campaigns
- Join the discussion at semsearcheval@yahoogroups.com



Search interface

Search Interface

- Semantic Search brings improvements in
 - Snippet generation
 - Adaptive and interactive presentation
 - Presentation adapts to the kind of query and results presented
 - Object results can be actionable, e.g. buy this product
 - User can provide feedback on particular objects or data sources
 - Aggregated search
 - Grouping similar items, summarizing results in various ways
 - Filtering (facets), possibly across different dimensions
 - Query and task templates
 - Help the user to fulfill the task by placing the query in a task context

Snippet generation using metadata

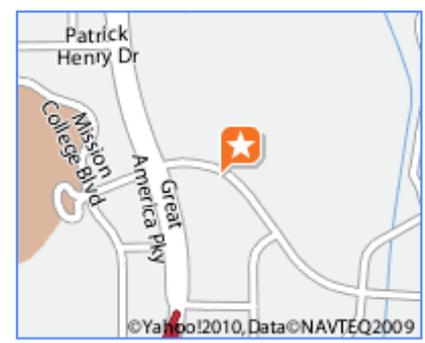
- Yahoo displays enriched search results for pages that contain microformat or RDFa markup using recognized ontologies
 - Displaying data, images, video
 - Example: GoodRelations for products
 - Enhanced results also appear for sites from which we extract information ourselves
- Also used for generating facets that can be used to restrict search results by object type
 - Example: “Shopping sites” facet for products
- Documentation and validator for developers
 - <http://developer.search.yahoo.com>
- Formerly: SearchMonkey allowed developers to customize the result presentation and create new ones for any object type
- *Haas et al. Enhanced Results in Web Search. SIGIR 2011*

Example: Yahoo! Enhanced Results

9,550 results for **parcel 104 sunnyvale**:

- [Show All](#)
- [Yelp](#)
- [TripAdvisor](#)
- [Yahoo! Travel](#)
- [Yahoo! Local](#)

Parcel 104
www.parcel104.com
(408) 970-6104
2700 Mission College Blvd, Santa Clara, CA ▾
[Make Reservations](#) | [Hours & More Info »](#)



★★★★★ (7)
7 Reviews - Parcel 104 is a real treat! The food is fantastic and it is nice not to be buried in tons of food that is too much for one meal. The [more](#)

Parcel 104 - Santa Clara, CA
[User Reviews](#) | [Photos](#) | [Write Review](#)

- Yelp Rating: ★★★★★ 246
- Address: 2700 Mission College Blvd, Santa Clara, CA 95054, USA
- Phone: (408) 970-6104

www.yelp.com/biz/parcel-104-santa-clara - [Cached](#)

Enhanced result with deep links, rating, address.

Parcel 104, Santa Clara - Restaurant Reviews - TripAdvisor
★★★★★ - (408) 970-6104 - Santa Clara, CA 95054, 95054
#11 of 36 hotels in **Sunnyvale**, CA; Holiday Inn Express Mountain View (Town Center) ... provides unbiased reviews, articles, recommendations and opinions on **Parcel 104** ...
www.tripadvisor.com/Restaurant_Review-g33046-d342112... - [Cached](#)

Sunnyvale American (New) Restaurants
Parcel 104 is a diamond in the ruff for sure!! This place is hidden at the Marriott ... Located in the heart of Downtown **Sunnyvale**, this place is really cute with super ...
www.yelp.com/c/sunnyvale-ca/newamerican - [Cached](#)

Example: Yahoo! Vertical Intent Search

Hi, **Guest** | Sign In | Help

Bucket: VIP004 ▾ highconf model ... bad results / ads or bugs? tell us! [hide]

Make Yahoo! your homepage | Ma

YAHOO!

Web Images Video Local Shopping News More ▾

harrison ford Search Options ▾

Search Pad

SearchScan - On

25,200,000 results for **harrison ford**

Related People

-  Mark Hamill
-  Calista Flockhart
-  Carrie Fisher
-  Sean Connery
-  Shia LaBeouf
-  George Lucas
-  Ridley Scott

Also try: [harrison ford movies](#), [harrison ford biography](#), [more...](#)

[Harrison Ford - Image Results](#)

News & Photos Videos Twitter

[more Harrison Ford photos...](#)

Latest News:

- [Colorado small forward commits to UB](#)
- The Buffalo News Blogs - Se
- [Harrison Ford talks abou](#)
- St. Petersburg Times - Sep 6
- [Harrison Ford's Harrison](#)
- Best Week Ever - Sep 08 08
- [more Harrison Ford new](#)

[Harrison Ford \(I\)](#)
His father was Irish, his mother Russian-Jewish. He was a lackluster student... Date of Birth: 13 July 1942 , Chicago, Illinois, USA.
www.imdb.com/name/nm0000148 - [Cached](#)

[Harrison Ford - Wikipedia, the free encyclopedia](#)
[Early life](#) | [Early career](#) | [Milestone...](#) | [Other film...](#)
Harrison Ford is an American film actor and producer. Ford is

Sponsored Results

[Harrison Ford the Films Sale](#)
Save on **Harrison Ford** the Films.
Shop Smart, Save Big
www.shop.com

[Harrison Ford Bluray at Amazon.co.uk](#)
Over 10,000 DVDs under £10. Free UK Delivery on Amazon Orders
Amazon.co.uk/dvd

[See your message here...](#)

Related actors and movies

Example: Sig.ma Semantic Information Mashup (DERI)



Help About Forum

Version: 1.1.33

giovanni tummarello [Add More Info](#) [Start New](#) [Order](#) [Options](#) [Use it](#)

Giovanni Tummarello

picture:  [15]

title: Dr. [15]

given name: Giovanni [3,4,15]

family name: Tummarello [3,4,15]

is creator of: [A Node Indexing Scheme for Web Entity Retrieval](#) [3,4]
[Hierarchical Link Analysis for Ranking Web Data](#) [3,4]
[ESWC 2006 Demo: DBin - enabling SW P2P communities](#) [3,4]
[Rapid Prototyping of Semantic Mash-Ups through Semantic Web Pipes](#) [3,4]
[Context Dependent Reasoning for Semantic Documents in Sindice](#) [3,4]
[An Entity Name System for Linking Semantic Web Data](#) [3,4]
[Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web](#) [3,4,19,20]
[Sindice.com: Weaving the Open Linked Data](#) [3,4,5,6,7,8,9,10,11]
[RDFSyc: efficient remote synchronization of RDF models](#) [3,4,5,6,7,8,9,10,11]
[Exposing Large Datasets with Semantic Sitemaps](#) [5,6,7,8,9,10,11]
[Enabling Semantic Web communities with DBin: an overview](#) [5,6,7,8,9,10,11]
[show 100 more values](#) [+]

is alternate of: [Giovanni Tummarello - semanticweb.org](#) [16,17,18]

alternate: http://www.semanlink.net/tag/giovanni_tummarello.rss [14]

Sources (20) Approved (0) Rejected (0)

- [Giovanni Tummarello](#) 80 facts | 2010-12-28
http://dblp.l3s.de/d2r/resource/authors/Giovanni_Tu...
- [RDF Description of Giova...](#) 80 facts | 2010-12-28
http://dblp.l3s.de/d2r/data/authors/Giovanni_Tummar...
- [Untitled document](#) 44 facts | 2010-12-28
<http://data.semanticweb.org/person/giovanni-tummare...>
- [Giovanni Tummarello](#) 44 facts | 2010-12-23
<http://data.semanticweb.org/person/giovanni-tummare...>
- [Giovanni Tummarello](#) 38 facts | 2010-12-28
http://semanticweb.org/id/Giovanni_Tummarello
- [Untitled document](#) 38 facts | 2010-12-23
<http://ontoworld.org/wiki/Special:ExportRDF/Giovann...>
- [Untitled document](#) 38 facts | 2010-12-23
<http://ontoworld.org/wiki/Special:URIResolver/Giova...>
- [Untitled document](#) 38 facts | 2010-12-23
<http://semanticweb.org/wiki/Special:ExportRDF/Giova...>
- [Untitled document](#) 38 facts | 2010-12-28
<http://semanticweb.org/index.php?title=Special:Expo...>
- [Untitled document](#) 42 facts | 2010-02-19
http://semanticweb.org/id/Giovanni_Tummarello&fetch...
- [Untitled document](#) 42 facts | 2010-03-03
http://semanticweb.org/id/Giovanni_Tummarello&fetch...
- [Untitled document](#) 2 facts | 2010-07-30
<http://www.bibsonomy.org/swrc/author/Giovanni+Tumma...>
- [Giovanni Tummarello](#) 17 facts | 2010-11-23
http://www.semanlink.net/tag/giovanni_tummarello

[reject all](#) [approve all](#)

[add source url](#)

[hide](#)

Example: Time Explorer

Demonstration Application

home | help

Living Knowledge
Time Explorer

iraq

Future

Search

YAHOO!
RESEARCH



Search in Time

Results 1 - 10 of 40 for Iraq

Congressional Office Forecasts Drop in Deficit, With Possibility of a Later Surplus



If the United States reduced the number of troops in Iraq to 30,000 in 2010, from more than 130,000 today, the agency estimated, the spending on Iraq would be about \$300 billion less than it assumes
query.nytimes.com/gst/fullpage.html?res=9...

2007-01-25

Keywords: John M. Spratt Jr., Iraq, Congressional Budget Office, White House

source

Dates: 2007-01-25, 2010

The Cost of Invading Iraq: Imponderables Meet Uncertainties



Another study of Iraq war costs, by Linda J. Bilmes of Harvard and Joseph E. Stiglitz of Columbia, comes up with an eye-catching estimate of \$2.2 trillion, assuming the United States is no longer... in Iraq in 2015.
query.nytimes.com/gst/fullpage.html?res=9...

2006-03-30

Keywords: Alan B. Krueger, Robert H. Topel, Steven J. Davis, Saddam Hussein

source

Dates: 2015

Iraq Is Strategic Issue for Oil Giants, Too



Mr. Chalabi, in his paper, wrote, "Iraq has the potential to produce 4.7 million barrels a day more oil from discovered fields that are ready to be developed." With the right investments, Iraq
query.nytimes.com/gst/fullpage.html?res=9...

2003-02-22

Keywords: Saddam Hussein, Raad Alkadiri, Saudi Arabia, Persian Gulf

Person Location NYT Per NYT Loc

- As McBride (1)
- David Brooks (1)
- David Rieff (1)
- Jack Bauer (1)
- Jonathan Raban (1)
- Joseph E. Stiglitz (1)
- Kathleen Hastings (1)
- Linda J. Bilmes (1)
- More Humble Hawk (1)
- Paul D. Wolfowitz (1)
- Right Sp... Hannity (1)



Opinion Panel

Overall view Trend Adjustments



Future work in Semantic Web Search

- (Semi-)automated ways of metadata creation
 - How do we go from 5% to 95%?
- Data quality
 - How do we assess the quality of data?
- Reasoning
 - Automated ontology mapping, instance mapping and blending
 - To what extent is reasoning useful in general?
- Scalability
 - What is between databases and IR engines?
- Ontology reuse
 - How do we get people to reuse ontologies?
- Display
 - How do we automatically generate effective displays for data we don't understand or only partially understand?

In 2011

- The Semantic Web is still evolving
 - Finally, a JSON syntax for RDF!
- Leaner and meaner
 - Bottom-up approaches: graphs, statistics, mining, retrieval, visualization...
- We get some credit in other fields
 - RDF data management is now a topic at VLDB, others
 - CIKM 2011 has 67 submitted abstracts on Semantic Search!
- The Semantic Web is 10 years wiser... *and so am I*

The End

