The Web as a **Semantic Source**

Ricardo Baeza-Yates

VP, Yahoo! Research

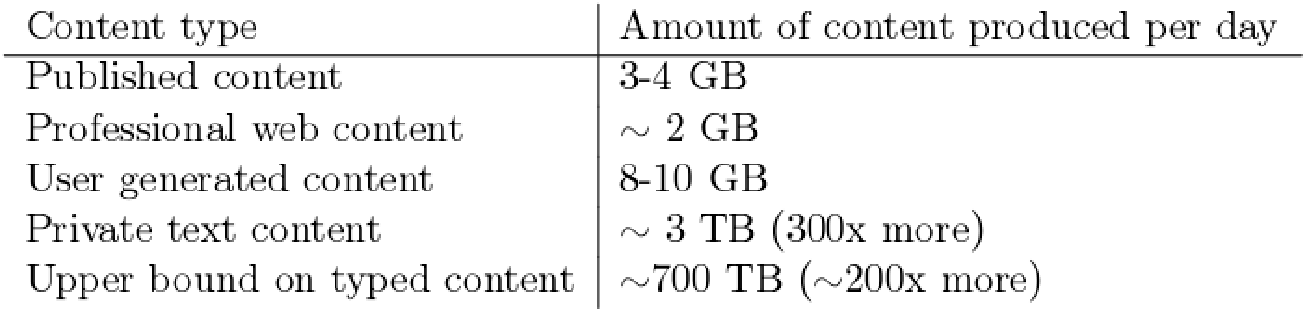
and Applications

Outline

- Internet moved from a curiosity to a substrate for life activity
- Content growing, changing, diversifying, fragmenting
- Semantics of content unlock the value of the data ecosystem
- Explicit and implicit semantic sources
- Applications
- A virtuous feedback cycle for enhancing semantics

Content Growth and Trends

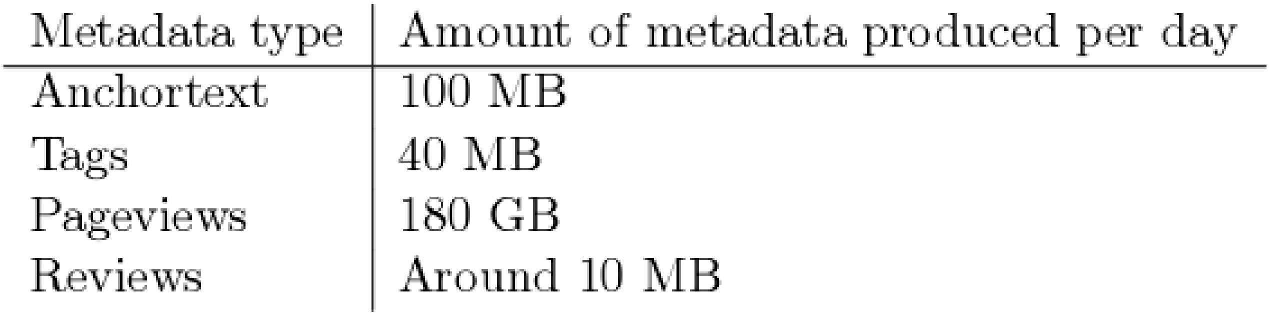
Content trends



Content type	Amount of content produced per day
Published content	3-4 GB
Professional web content	∼ 2 GB
User generated content	8-10 GB
Private text content	∼ 3 TB (300x more)
Upper bound on typed content	∼700 TB (∼200x more)

[Ramakrishnan and Tomkins 2007]

Metadata trends



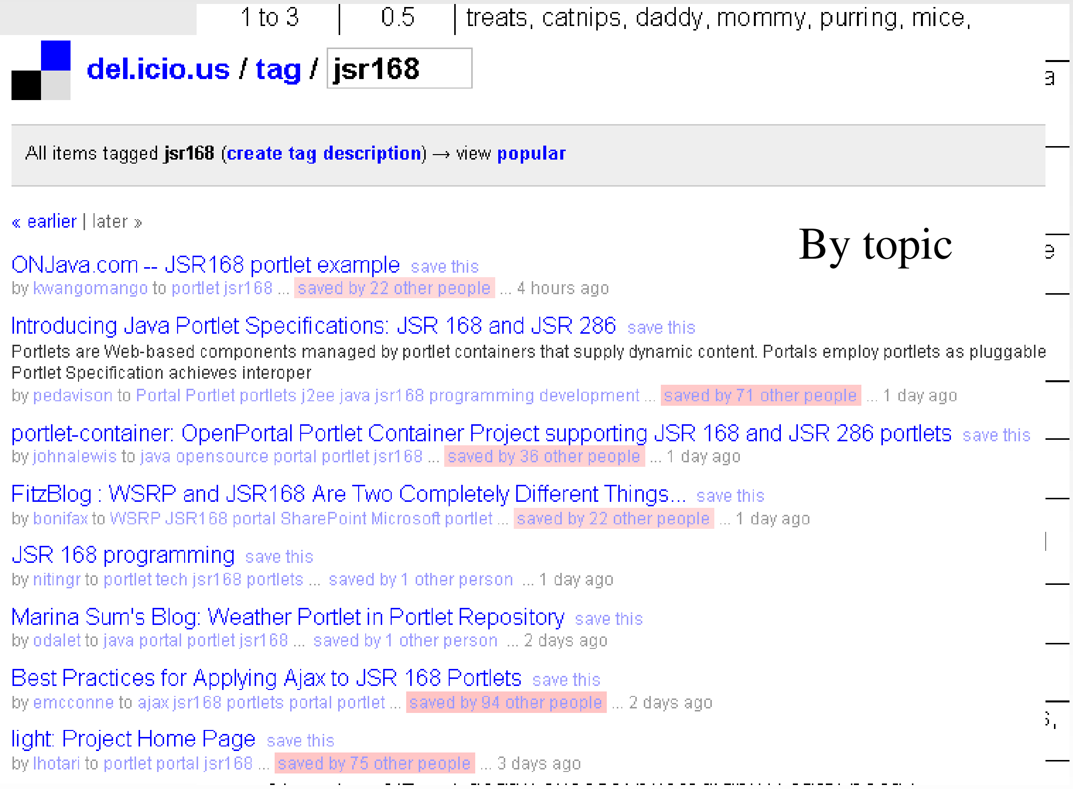
Metadata type	Amount of metadata produced per day
Anchortext	100 MB
Tags	40 MB
Pageviews	180 GB
Reviews	Around 10 MB

[Ramakrishnan and Tomkins 2007]

Content ownership

- Content consumption is fragmenting – nobody owns more than 10% of the Web page views
- No single place will own all the content
- Best of breed processing will operate on the web version (?)
- Value transitions to ecosystem

Content Consumption is fragmenting



1 to 3 | 0.5 | treats, catnips, daddy, mommy, purring, mice,

All items tagged **jsr168** (**create tag description**) → view **popular**

« earlier | later »

By topic

ONJava.com -- JSR168 portlet example save this
by kwangomango to portlet jsr168 ... saved by 22 other people ... 4 hours ago

Introducing Java Portlet Specifications: JSR 168 and JSR 286 save this
Portlets are Web-based components managed by portlet containers that supply dynamic content. Portals employ portlets as pluggable
Portlet Specification achieves interoper
by pedavison to Portal Portlet portlets j2ee java jsr168 programming development ... saved by 71 other people ... 1 day ago

portlet-container: OpenPortal Portlet Container Project supporting JSR 168 and JSR 286 portlets save this
by johnalewis to java opensource portal portlet jsr168 ... saved by 36 other people ... 1 day ago

FitzBlog : WSRP and JSR168 Are Two Completely Different Things... save this
by bonifax to WSRP JSR168 portal SharePoint Microsoft portlet ... saved by 22 other people ... 1 day ago

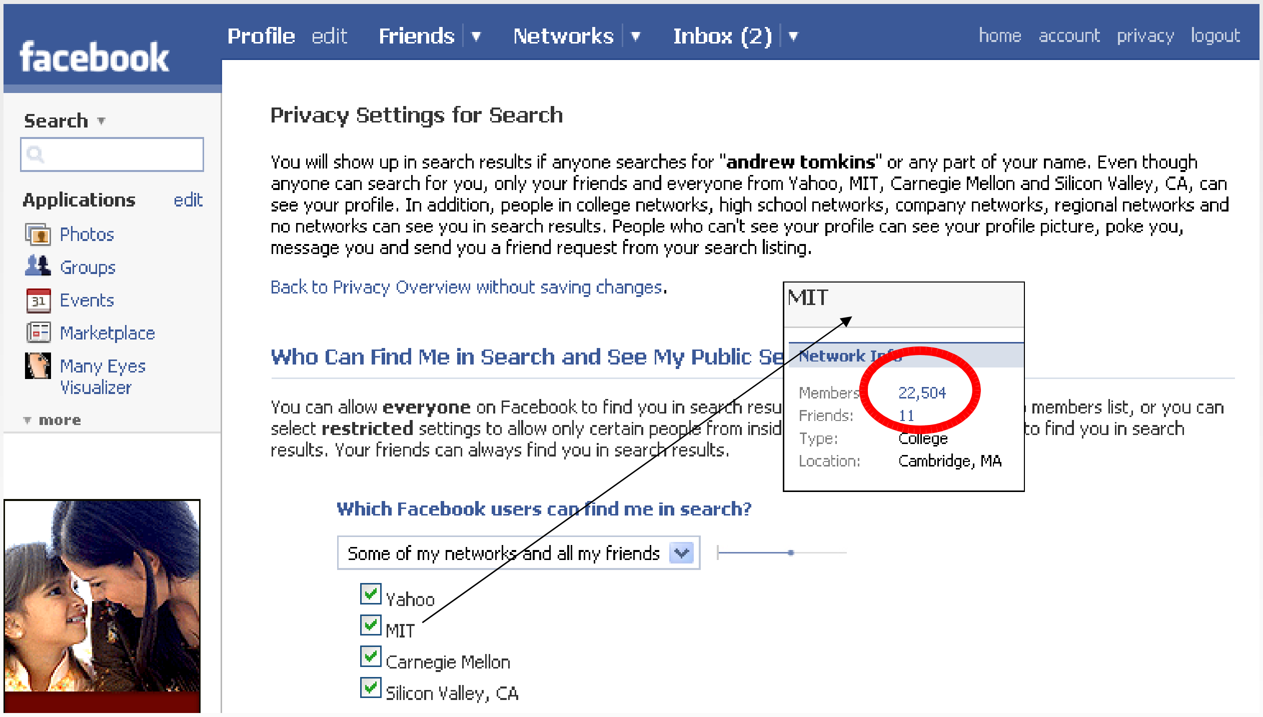
JSR 168 programming save this
by nitingr to portlet tech jsr168 portlets ... saved by 1 other person ... 1 day ago

Marina Sum's Blog: Weather Portlet in Portlet Repository save this
by odalet to java portal portlet jsr168 ... saved by 1 other person ... 2 days ago

Best Practices for Applying Ajax to JSR 168 Portlets save this
by emcconne to ajax jsr168 portlets portal portlet ... saved by 94 other people ... 2 days ago

light: Project Home Page save this
by lhotari to portlet portal jsr168 ... saved by 75 other people ... 3 days ago

Content access is fragmenting



Privacy Settings for Search

You will show up in search results if anyone searches for "**andrew tomkins**" or any part of your name. Even though anyone can search for you, only your friends and everyone from Yahoo, MIT, Carnegie Mellon and Silicon Valley, CA, can see your profile. In addition, people in college networks, high school networks, company networks, regional networks and no networks can see you in search results. People who can't see your profile can see your profile picture, poke you, message you and send you a friend request from your search listing.

Back to Privacy Overview without saving changes.

Who Can Find Me in Search and See My Public Se

You can allow **everyone** on Facebook to find you in search resu … members list, or you can select **restricted** settings to allow only certain people from insid … to find you in search results. Your friends can always find you in search results.

MIT

Network Info

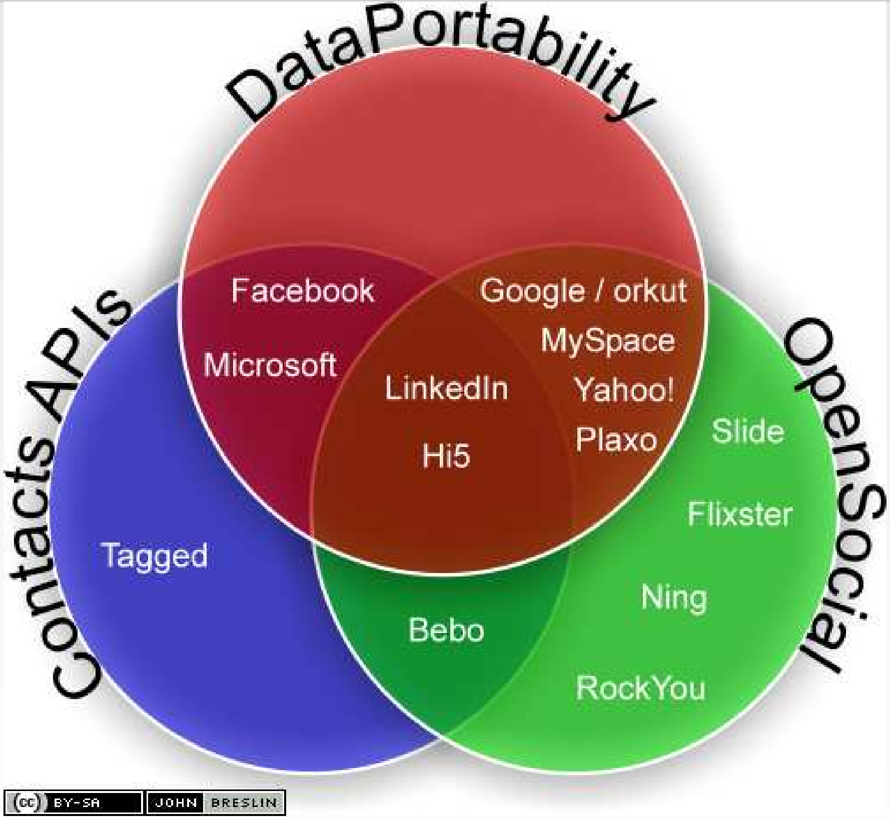
Members:	22,504
Friends:	11
Type:	College
Location:	Cambridge, MA

Which Facebook users can find me in search?

Some of my networks and all my friends ▾

☑ Yahoo
☑ MIT
☑ Carnegie Mellon
☑ Silicon Valley, CA

Time of Openness

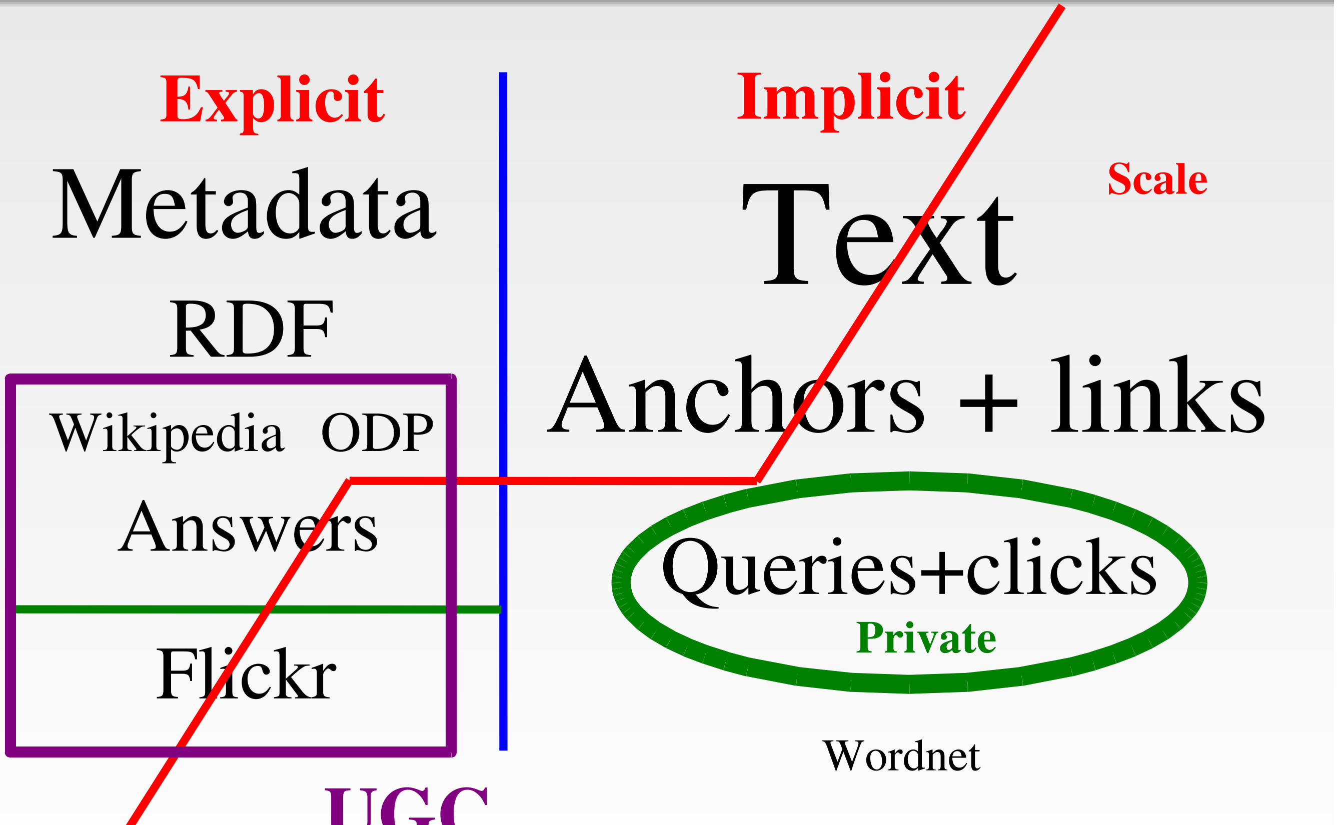


Semantic Sources

Semantic Sources in the Web

- Intrinsically Collaborative
- Explicit or Implicit
- Taxonomies vs. Folksonomies
- Different Size and Growth
- Diversity of Quality
- Public or Private

Examples



Explicit

Metadata

RDF

Wikipedia ODP

Answers

Flickr

UGC

Implicit

Scale

Text

Anchors + links

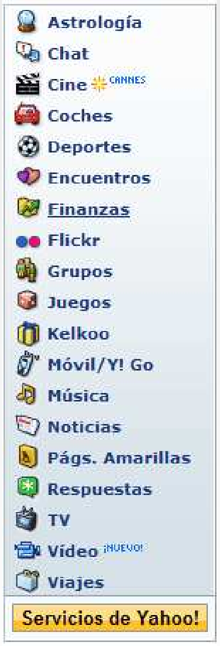
Queries+clicks

Private

Wordnet

User Generated Content at Yahoo!

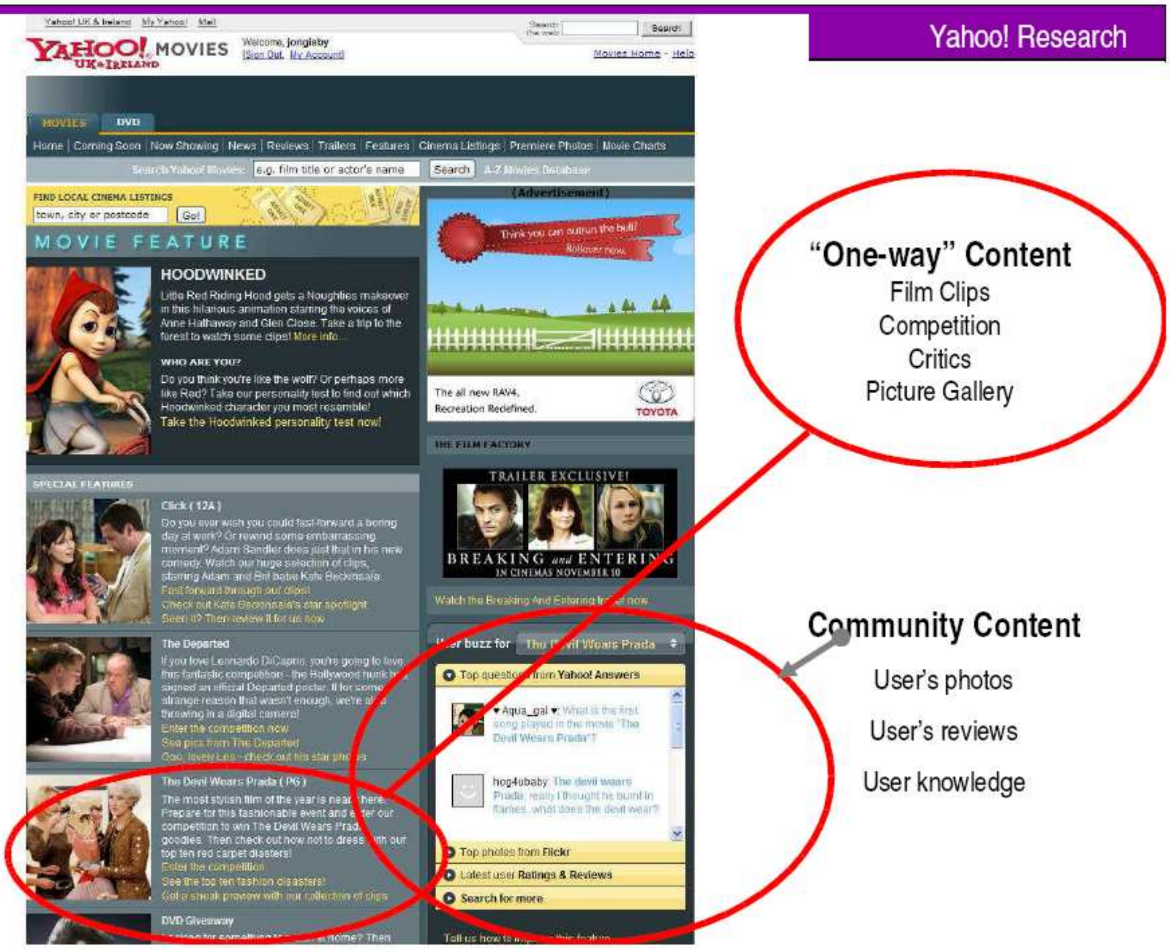
Social properties had 115M unique visitors worldwide, 56M "under 35".



Astrología
Chat
Cine CANNES
Coches
Deportes
Encuentros
Finanzas
Flickr
Grupos
Juegos
Kelkoo
Móvil/Y! Go
Música
Noticias
Págs. Amarillas
Respuestas
TV
Vídeo ¡NUEVO!
Viajes
Servicios de Yahoo!

• **Yahoo! Grups**	8 million groups, 1 in 10 Internet users
• **Del.icio.us**	2 million users
• **Flickr**	1 million photos uploaded daily
• **Yahoo! Answers**	90M unique users, 250M answers
• **Messenger**	85M unique users

(2007 data)



Yahoo! UK & Ireland My Yahoo! Mail
Search the web Search
YAHOO! UK & IRELAND MOVIES Welcome, jongiaby [Sign Out, My Account]
Movies Home - Help

MOVIES DVD
Home | Coming Soon | Now Showing | News | Reviews | Trailers | Features | Cinema Listings | Premiere Photos | Movie Charts
Search Yahoo! Movies: e.g. film title or actor's name Search A-Z Movies Database

FIND LOCAL CINEMA LISTINGS town, city or postcode Go!

MOVIE FEATURE

HOODWINKED

Little Red Riding Hood gets a Noughties makeover in this hilarious animation starring the voices of Anne Hathaway and Glen Close. Take a trip to the forest to watch some clips! More info...

WHO ARE YOU?

Do you think you're like the wolf? Or perhaps more like Red? Take our personality test to find out which Hoodwinked character you most resemble!
Take the Hoodwinked personality test now!

SPECIAL FEATURES

Click (12A)

Do you ever wish you could fast-forward a boring day at work? Or rewind some embarrassing moment? Adam Sandler does just that in his new comedy. Watch our huge selection of clips, starring Adam and Brit babe Kate Beckinsale.
Fast forward through our clips!
Check out Kate Beckinsale's star spotlight
Seen it? Then review it for us now

The Departed

If you love Leonardo DiCaprio, you're going to love this fantastic competition - the Hollywood hunk has signed an official Departed poster. If for some strange reason that wasn't enough, we're also throwing in a digital camera!
Enter the competition now
See pics from The Departed
Ooo, lovely Leo - check out his star photos

The Devil Wears Prada (PG)

The most stylish film of the year is nearly here. Prepare for this fashionable event and enter our competition to win The Devil Wears Prada goodies. Then check out how not to dress with our top ten red carpet disasters!
Enter the competition
See the top ten fashion disasters!
Get a sneak preview with our collection of clips

DVD Giveaway

(Advertisement)
Think you can outrun the bull? Rollover now.
The all new RAV4. Recreation Redefined. TOYOTA

THE FILM FACTORY
TRAILER EXCLUSIVE! BREAKING and ENTERING IN CINEMAS NOVEMBER 10
Watch the Breaking And Entering trailer now

User buzz for The Devil Wears Prada
Top questions from Yahoo! Answers
♥ Aqua_gal ♥: What is the first song played in the movie "The Devil Wears Prada"?
hog4ubaby: The devil wears Prada, really I thought he burnt in flames, what does the devil wear?
Top photos from Flickr
Latest user Ratings & Reviews
Search for more
Tell us how to improve this feature

"One-way" Content
Film Clips
Competition
Critics
Picture Gallery

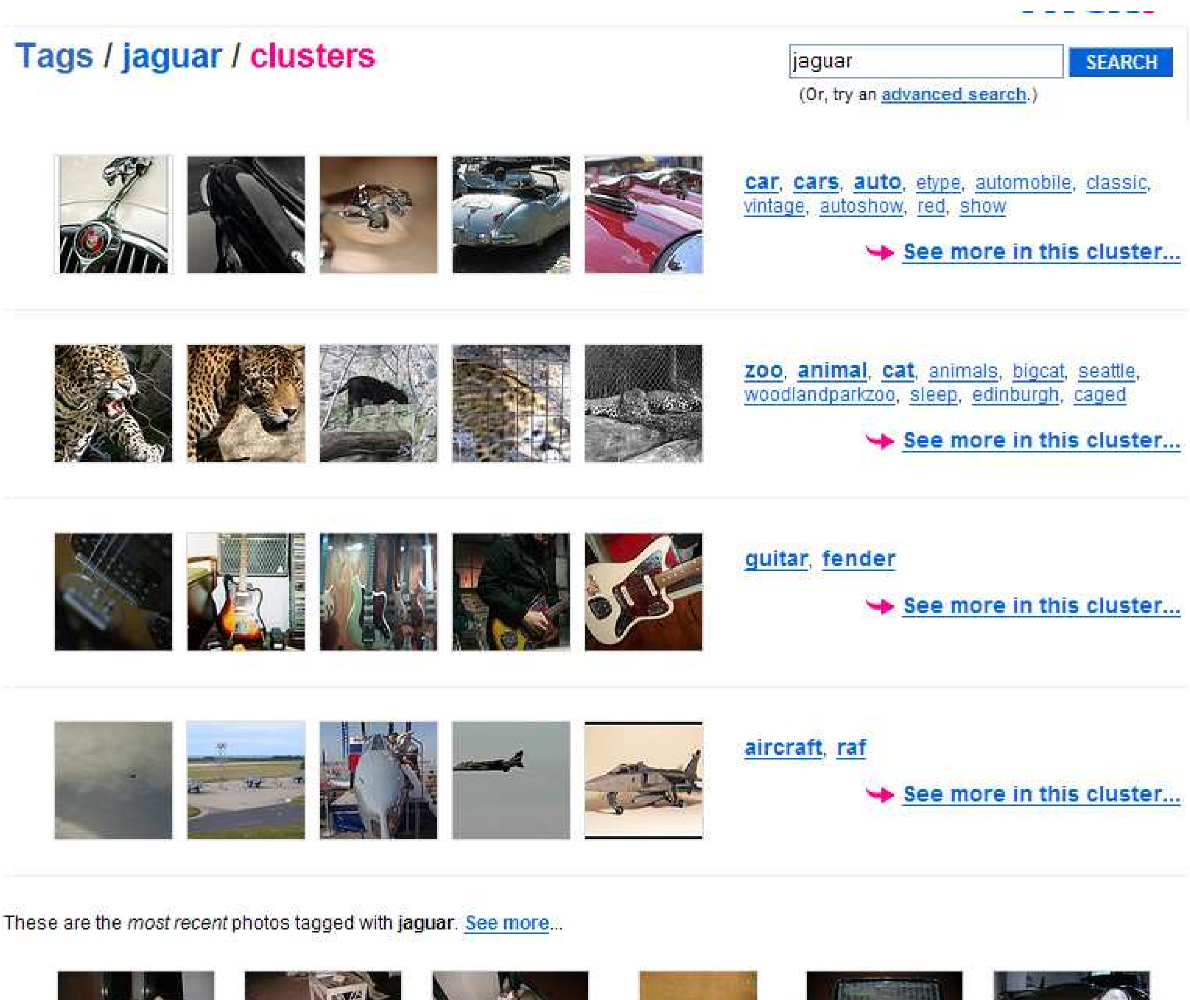
Community Content
User's photos
User's reviews
User knowledge

The Wisdom of Crowds

- James Surowiecki, a ***New Yorker*** columnist, published this book in 2004
  - **“Under the right circumstances, groups are remarkably intelligent”**
- Importance of diversity, independence and decentralization

Aggregating data

> *“large groups of people are smarter than an elite few, no matter how brilliant—they are better at solving problems, fostering innovation, coming to wise decisions, even predicting the future”.*



Tags / jaguar / clusters

(Or, try an advanced search.)

car, cars, auto, etype, automobile, classic, vintage, autoshow, red, show

See more in this cluster...

zoo, animal, cat, animals, bigcat, seattle, woodlandparkzoo, sleep, edinburgh, caged

See more in this cluster...

guitar, fender

See more in this cluster...

aircraft, raf

See more in this cluster...

These are the *most recent* photos tagged with jaguar. See more...

The Wisdom of Crowds

–Crucial for Search Ranking

–Text: Web Writers & Editors

- not only for the Web!

–Links: Web Publishers

–Tags: Web Taggers

–Queries: All Web Users!

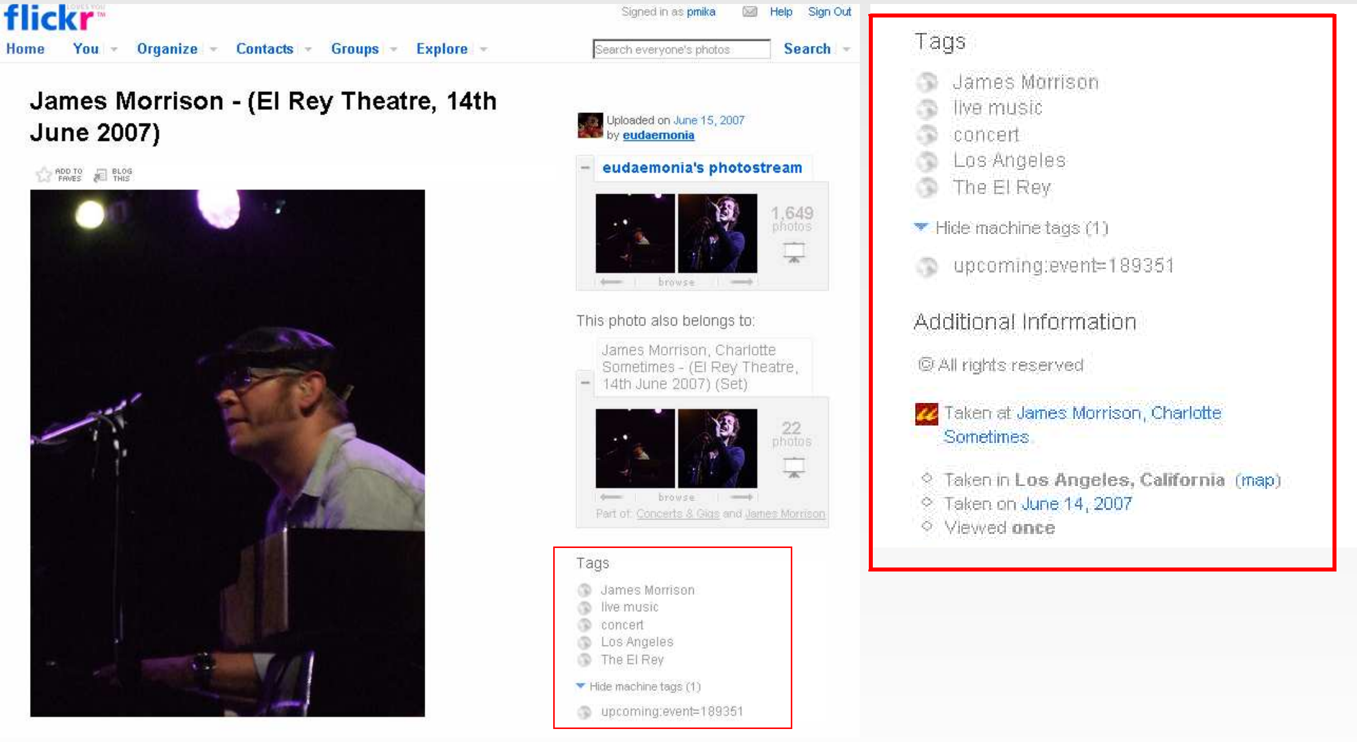
- Queries and actions (or no action!)

Metadata: Microformats

Microformats

- microformats.org
- Originated by Tantek Celik and others
- Agreements on the way to encode certain kinds metadata in HTML
  - Reuse of semantic-bearing HTML elements
  - Based on existing standards
  - Community process
  - Persons, events, listings etc. but also syntactic metadata: licenses, tags
- Microformats have no shared syntax
  - Each microformat has a separate syntax tailored to the vocabulary
- Microformats are not ontologies
  - No formal descriptions of schema, only text
  - Limited reuse, extensibility of schemas
  - No datatypes
- No namespaces, unique identifiers (URIs)
  - no interlinking
  - mapping between instances is required
- Relationship to page context is unclear
- Widely used in millions of documents
  - User-generated as well as automatically generated

Example: tags and machine tags



James Morrison - (El Rey Theatre, 14th June 2007)

Uploaded on June 15, 2007
by eudaemonia

eudaemonia's photostream

This photo also belongs to:

James Morrison, Charlotte Sometimes - (El Rey Theatre, 14th June 2007) (Set)

Part of: Concerts & Gigs and James Morrison

Tags

- James Morrison
- live music
- concert
- Los Angeles
- The El Rey

▼ Hide machine tags (1)

- upcoming:event=189351

Additional Information

© All rights reserved

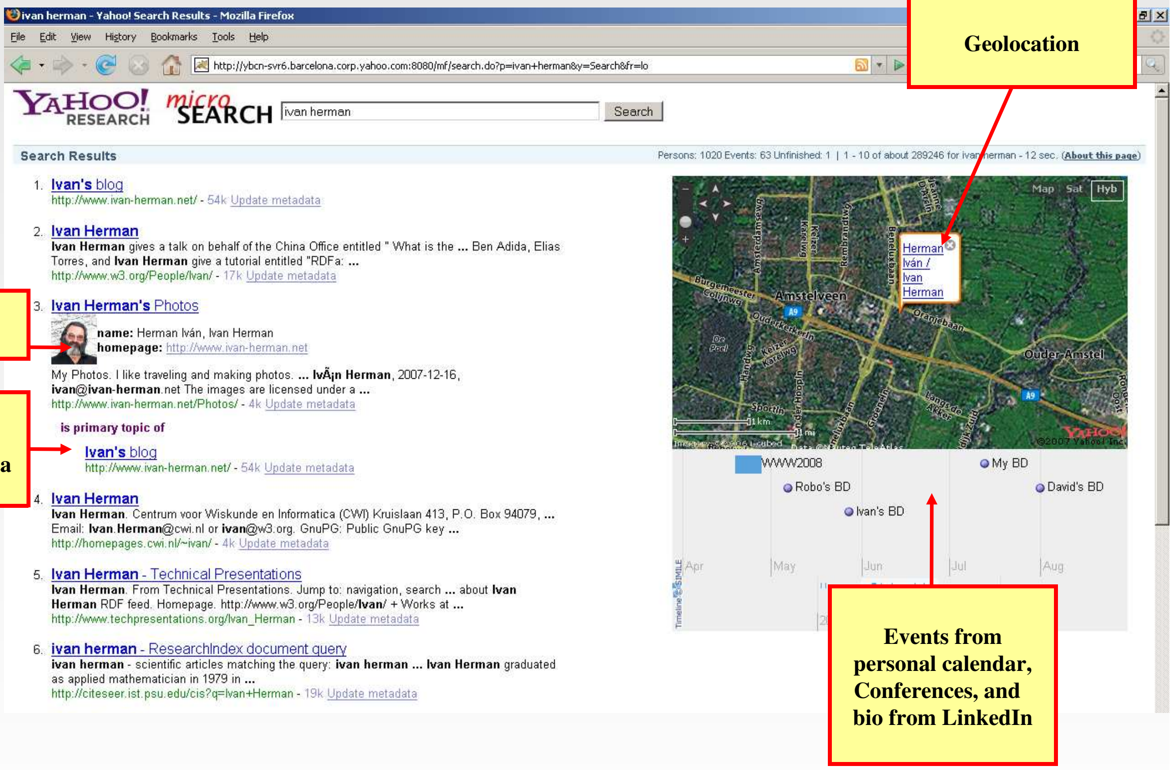
Taken at James Morrison, Charlotte Sometimes

- Taken in Los Angeles, California (map)
- Taken on June 14, 2007
- Viewed once

Metadata is out there

- Question:
  - Just how much data is out there?
  - What is the quality?
- Idea: bring metadata to the surface of search
- How does it work?
  - User enters query
  - Metadata is extracted dynamically
  - Entity reconciliation
  - Metadata is used to display
    - rich abstracts,
    - related pages
    - spatial, temporal visualization
- Microsearch prototype
    - Play at http://www.yr-bcn.es/demos/microsearch/

Example: ivan herman



http://ybcn-svr6.barcelona.corp.yahoo.com:8080/mf/search.do?p=ivan+herman&y=Search&fr=lo

ivan herman

Search Results

Persons: 1020 Events: 63 Unfinished: 1 | 1 - 10 of about 289246 for ivan herman - 12 sec. (**About this page**)

1. **Ivan's** blog
   http://www.ivan-herman.net/ - 54k Update metadata
2. **Ivan Herman**
   Ivan Herman gives a talk on behalf of the China Office entitled " What is the **...** Ben Adida, Elias Torres, and **Ivan Herman** give a tutorial entitled "RDFa: **...**
   http://www.w3.org/People/Ivan/ - 17k Update metadata
3. **Ivan Herman's** Photos
   name: Herman Iván, Ivan Herman
   homepage: http://www.ivan-herman.net
   My Photos. I like traveling and making photos. **... IvÃ¡n Herman**, 2007-12-16, **ivan@ivan-herman**.net The images are licensed under a **...**
   http://www.ivan-herman.net/Photos/ - 4k Update metadata
   is primary topic of
   Ivan's blog
   http://www.ivan-herman.net/ - 54k Update metadata
4. **Ivan Herman**
   Ivan Herman. Centrum voor Wiskunde en Informatica (CWI) Kruislaan 413, P.O. Box 94079, **...** Email: **Ivan**.**Herman**@cwi.nl or **ivan**@w3.org. GnuPG: Public GnuPG key **...**
   http://homepages.cwi.nl/~ivan/ - 4k Update metadata
5. **Ivan Herman** - Technical Presentations
   Ivan Herman. From Technical Presentations. Jump to: navigation, search **...** about **Ivan Herman** RDF feed. Homepage. http://www.w3.org/People/**Ivan**/ + Works at **...**
   http://www.techpresentations.org/Ivan_Herman - 13k Update metadata
6. **ivan herman** - ResearchIndex document query
   ivan herman - scientific articles matching the query: **ivan herman ... Ivan Herman** graduated as applied mathematician in 1979 in **...**
   http://citeseer.ist.psu.edu/cis?q=Ivan+Herman - 19k Update metadata

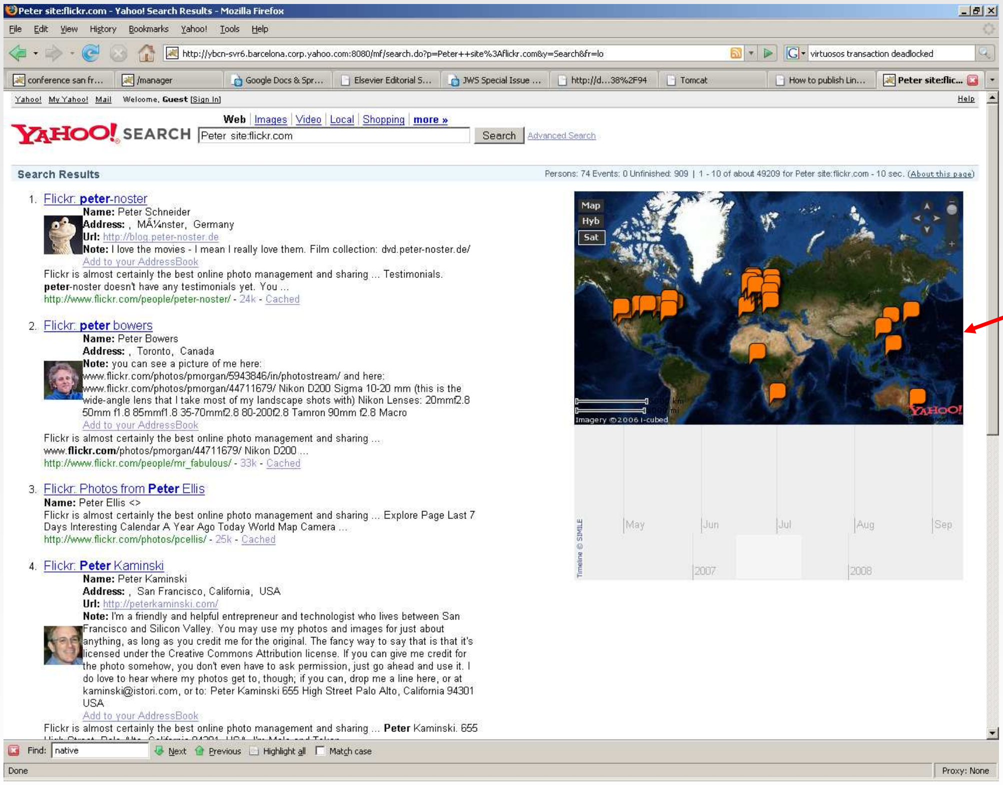
Geolocation

Rich abstract

Related pages based on metadata

Events from personal calendar, Conferences, and bio from LinkedIn

Example: peter site:flickr.com


Peter site:flickr.com - Yahoo! Search Results - Mozilla Firefox
File Edit View History Bookmarks Yahoo! Tools Help
http://ybcn-svr6.barcelona.corp.yahoo.com:8080/mf/search.do?p=Peter++site%3Aflickr.com&y=Search&fr=lo
virtuosos transaction deadlocked
conference san fr… /manager Google Docs & Spr… Elsevier Editorial S… JWS Special Issue … http://d…38%2F94 Tomcat How to publish Lin… Peter site:flic…
Yahoo! My Yahoo! Mail Welcome, Guest [Sign In] Help
Web Images Video Local Shopping more »
YAHOO! SEARCH
Search Advanced Search
Map
Hyb
Sat
Imagery ©2006 i-cubed
YAHOO!
Timeline © SIMILE
May
Jun
Jul
Aug
Sep
2007
2008
Find: native Next Previous Highlight all Match case
Done
Proxy: None

Peter site:flickr.com

Search Results

Persons: 74 Events: 0 Unfinished: 909 | 1 - 10 of about 49209 for Peter site:flickr.com - 10 sec. (About this page)

1. Flickr: **peter**-noster
   Name: Peter Schneider
   Address: , MÃ¼nster, Germany
   Url: http://blog.peter-noster.de
   Note: I love the movies - I mean I really love them. Film collection: dvd.peter-noster.de/
   Add to your AddressBook
   Flickr is almost certainly the best online photo management and sharing ... Testimonials. **peter**-noster doesn't have any testimonials yet. You ...
   http://www.flickr.com/people/peter-noster/ - 24k - Cached

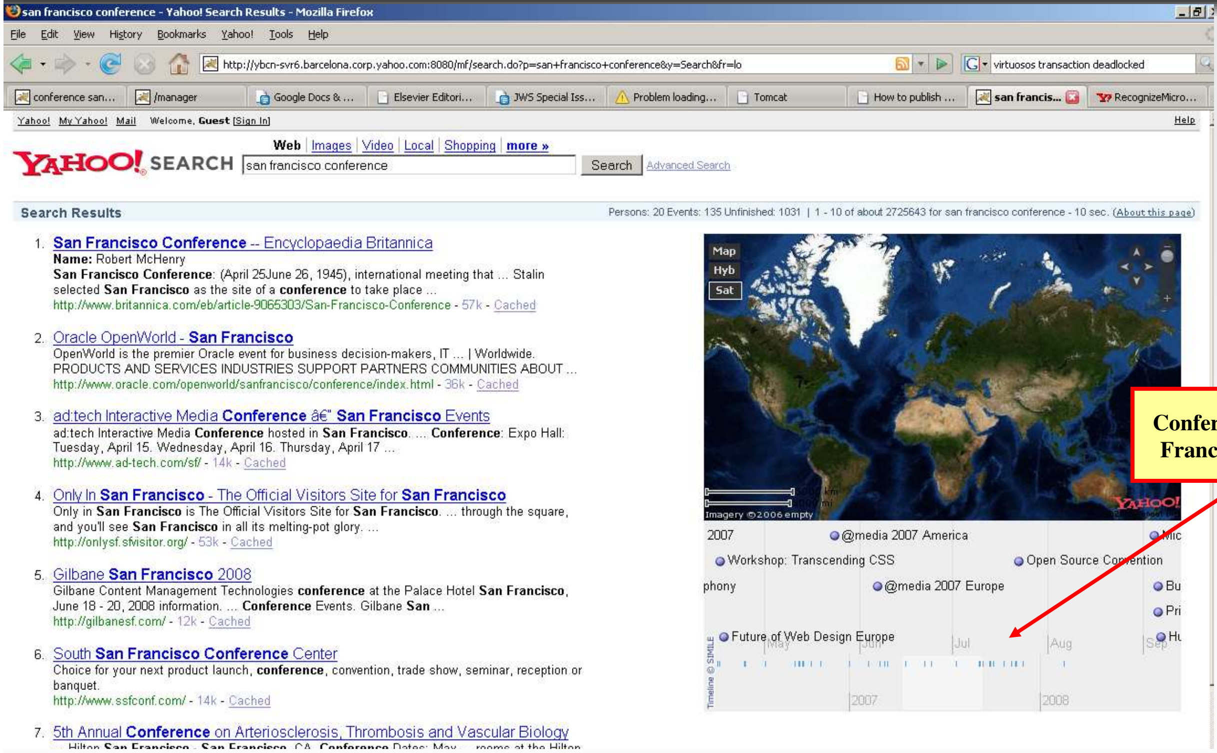
2. Flickr: **peter** bowers
   Name: Peter Bowers
   Address: , Toronto, Canada
   Note: you can see a picture of me here: www.flickr.com/photos/pmorgan/5943846/in/photostream/ and here: www.flickr.com/photos/pmorgan/44711679/ Nikon D200 Sigma 10-20 mm (this is the wide-angle lens that I take most of my landscape shots with) Nikon Lenses: 20mmf2.8 50mm f1.8 85mmf1.8 35-70mmf2.8 80-200f2.8 Tamron 90mm f2.8 Macro
   Add to your AddressBook
   Flickr is almost certainly the best online photo management and sharing ... www.**flickr.com**/photos/pmorgan/44711679/ Nikon D200 ...
   http://www.flickr.com/people/mr_fabulous/ - 33k - Cached

3. Flickr: Photos from **Peter** Ellis
   Name: Peter Ellis <>
   Flickr is almost certainly the best online photo management and sharing ... Explore Page Last 7 Days Interesting Calendar A Year Ago Today World Map Camera ...
   http://www.flickr.com/photos/pcellis/ - 25k - Cached

4. Flickr: **Peter** Kaminski
   Name: Peter Kaminski
   Address: , San Francisco, California, USA
   Url: http://peterkaminski.com/
   Note: I'm a friendly and helpful entrepreneur and technologist who lives between San Francisco and Silicon Valley. You may use my photos and images for just about anything, as long as you credit me for the original. The fancy way to say that is that it's licensed under the Creative Commons Attribution license. If you can give me credit for the photo somehow, you don't even have to ask permission, just go ahead and use it. I do love to hear where my photos get to, though; if you can, drop me a line here, or at kaminski@istori.com, or to: Peter Kaminski 655 High Street Palo Alto, California 94301 USA
   Add to your AddressBook
   Flickr is almost certainly the best online photo management and sharing ... **Peter** Kaminski. 655

Flickr users named "Peter" by geography

Example: san francisco conference


san francisco conference - Yahoo! Search Results - Mozilla Firefox
File Edit View History Bookmarks Yahoo! Tools Help
http://ybcn-svr6.barcelona.corp.yahoo.com:8080/mf/search.do?p=san+francisco+conference&y=Search&fr=lo
virtuosos transaction deadlocked
conference san… /manager Google Docs & … Elsevier Editori… JWS Special Iss… Problem loading… Tomcat How to publish … san francis… RecognizeMicro…
Yahoo! My Yahoo! Mail Welcome, Guest [Sign In] Help
Web Images Video Local Shopping more »
YAHOO! SEARCH
Search Advanced Search
Map
Hyb
Sat
Imagery ©2006 empty
YAHOO!
2007
@media 2007 America
Workshop: Transcending CSS
Open Source Convention
phony
@media 2007 Europe
Future of Web Design Europe
May
Jun
Jul
Aug
Sep
Timeline © SIMILE
2007
2008

san francisco conference

Search Results

Persons: 20 Events: 135 Unfinished: 1031 | 1 - 10 of about 2725643 for san francisco conference - 10 sec. (About this page)

1. **San Francisco Conference** -- Encyclopaedia Britannica
   Name: Robert McHenry
   San Francisco Conference: (April 25June 26, 1945), international meeting that ... Stalin selected **San Francisco** as the site of a **conference** to take place ...
   http://www.britannica.com/eb/article-9065303/San-Francisco-Conference - 57k - Cached

2. Oracle OpenWorld - **San Francisco**
   OpenWorld is the premier Oracle event for business decision-makers, IT ... | Worldwide. PRODUCTS AND SERVICES INDUSTRIES SUPPORT PARTNERS COMMUNITIES ABOUT ...
   http://www.oracle.com/openworld/sanfrancisco/conference/index.html - 36k - Cached

3. ad:tech Interactive Media **Conference** â€" **San Francisco** Events
   ad:tech Interactive Media **Conference** hosted in **San Francisco**. ... **Conference**: Expo Hall: Tuesday, April 15. Wednesday, April 16. Thursday, April 17 ...
   http://www.ad-tech.com/sf/ - 14k - Cached

4. Only In **San Francisco** - The Official Visitors Site for **San Francisco**
   Only in **San Francisco** is The Official Visitors Site for **San Francisco**. ... through the square, and you'll see **San Francisco** in all its melting-pot glory. ...
   http://onlysf.sfvisitor.org/ - 53k - Cached

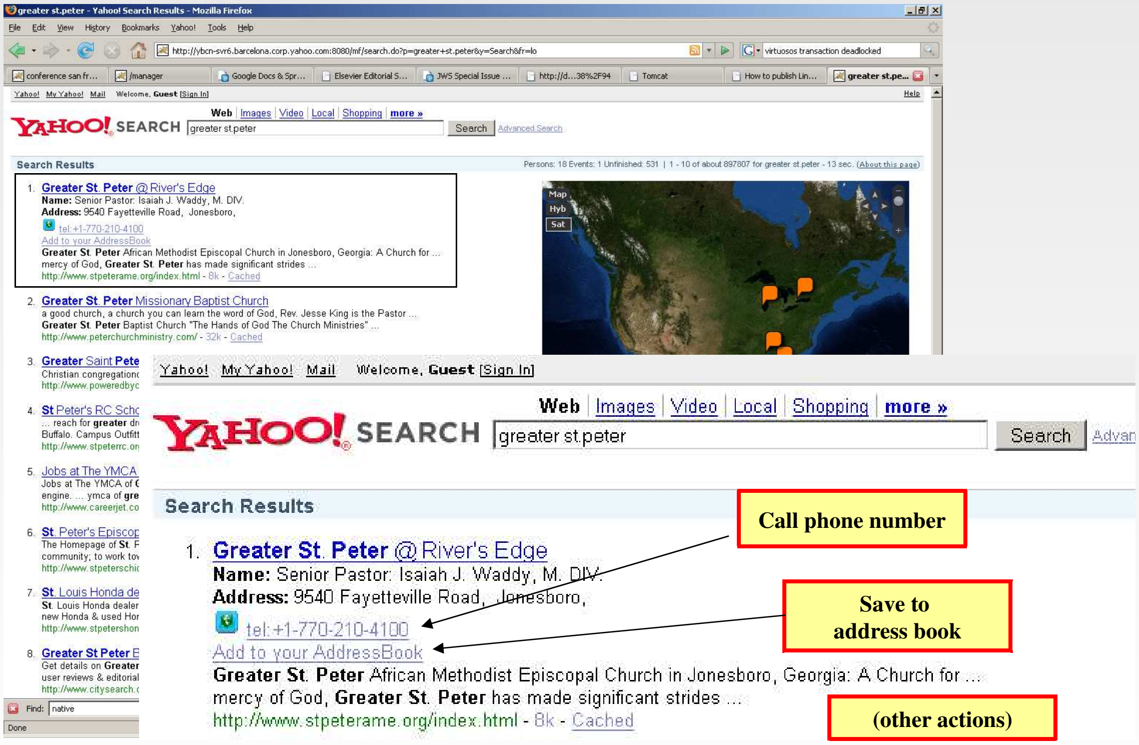
5. Gilbane **San Francisco** 2008
   Gilbane Content Management Technologies **conference** at the Palace Hotel **San Francisco**, June 18 - 20, 2008 information. ... **Conference** Events. Gilbane **San** ...
   http://gilbanesf.com/ - 12k - Cached

6. South **San Francisco Conference** Center
   Choice for your next product launch, **conference**, convention, trade show, seminar, reception or banquet.
   http://www.ssfconf.com/ - 14k - Cached

7. 5th Annual **Conference** on Arteriosclerosis, Thrombosis and Vascular Biology

Conferences in San Francisco by date

Example: greater st. peter



greater st.peter - Yahoo! Search Results - Mozilla Firefox

http://ybcn-svr6.barcelona.corp.yahoo.com:8080/mf/search.do?p=greater+st.peter&y=Search&fr=lo

File Edit View History Bookmarks Yahoo! Tools Help
virtuosos transaction deadlocked
conference san fr... /manager Google Docs & Spr... Elsevier Editorial S... JWS Special Issue ... http://d...38%2F94 Tomcat How to publish Lin... greater st.pe...
Yahoo! My Yahoo! Mail Welcome, Guest [Sign In] Help
Web | Images | Video | Local | Shopping | more »
YAHOO! SEARCH
Search Advanced Search
Map
Hyb
Sat

greater st.peter

Search Results

Persons: 18 Events: 1 Unfinished: 531 | 1 - 10 of about 897807 for greater st.peter - 13 sec. (About this page)

1. Greater St. Peter @ River's Edge
   Name: Senior Pastor: Isaiah J. Waddy, M. DIV.
   Address: 9540 Fayetteville Road, Jonesboro,
   tel:+1-770-210-4100
   Add to your AddressBook
   Greater St. Peter African Methodist Episcopal Church in Jonesboro, Georgia: A Church for ... mercy of God, **Greater St. Peter** has made significant strides ...
   http://www.stpeterame.org/index.html - 8k - Cached
2. Greater St. Peter Missionary Baptist Church
   a good church, a church you can learn the word of God, Rev. Jesse King is the Pastor ...
   Greater St. Peter Baptist Church "The Hands of God The Church Ministries" ...
   http://www.peterchurchministry.com/ - 32k - Cached
3. Greater Saint Pete
   Christian congregationc
   http://www.poweredbyc
4. St Peter's RC Scho
   ... reach for **greater** dr
   Buffalo. Campus Outfitt
   http://www.stpeterrc.or
5. Jobs at The YMCA
   Jobs at The YMCA of G
   engine. ... ymca of **gre**
   http://www.careerjet.co
6. St. Peter's Episcop
   The Homepage of **St**. F
   community; to work tov
   http://www.stpeterschic
7. St. Louis Honda de
   St. Louis Honda dealer
   new Honda & used Hor
   http://www.stpetershon
8. Greater St Peter B
   Get details on **Greater**
   user reviews & editorial
   http://www.citysearch.c

Find: native

Done

Yahoo! My Yahoo! Mail Welcome, Guest [Sign In]

Web | Images | Video | Local | Shopping | more »

YAHOO! SEARCH greater st.peter Search Advan

Search Results

1. Greater St. Peter @ River's Edge
   Name: Senior Pastor: Isaiah J. Waddy, M. DIV.
   Address: 9540 Fayetteville Road, Jonesboro,
   tel:+1-770-210-4100
   Add to your AddressBook
   Greater St. Peter African Methodist Episcopal Church in Jonesboro, Georgia: A Church for ... mercy of God, **Greater St. Peter** has made significant strides ...
   http://www.stpeterame.org/index.html - 8k - Cached

Call phone number

Save to address book

(other actions)

Lessons

- More metadata than we expected
  - 53% of unique queries have at least one metadata-enabled page in top 10 (n=7848)
- Performance is poor
  - Metadata needs to come from the index for performance
- Metacrap does exist
  - Users have to see metadata to spot mistakes in their markup, warn others
- RDF templating is hard
  - Adds extra complexity
- Scalability

Exploting Metadata: SearchMonkey

SearchMonkey

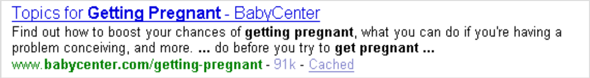
- Creating an ecosystem of publishers, developers and end-users
  - Motivating and helping publishers to implement semantic annotation
  - Providing tools for developers to create compelling applications
  - Focusing on end-user experience
- Rich abstracts as a first application
- Addressing the long tail of query and content production
- Standard Semantic Web technology
  - dataRSS = Atom + RDFa
  - Industry standard vocabularies

What is SearchMonkey?

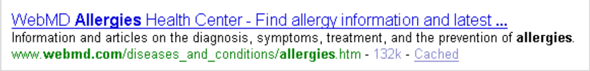
SearchMonkey

an open platform for using structured data to build more useful and relevant search results

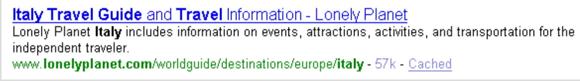
Before



[Topics for **Getting Pregnant** - BabyCenter]
Find out how to boost your chances of **getting pregnant**, what you can do if you're having a problem conceiving, and more. **...** do before you try to **get pregnant ...**
www.**babycenter.com/getting-pregnant** - 91k - Cached

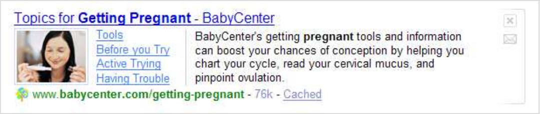


WebMD **Allergies** Health Center - Find allergy information and latest **...**
Information and articles on the diagnosis, symptoms, treatment, and the prevention of **allergies**.
www.**webmd.com**/diseases_and_conditions/**allergies**.htm - 132k - Cached

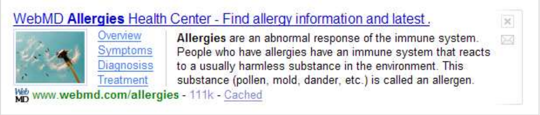


Italy Travel Guide and **Travel** Information - Lonely Planet
Lonely Planet **Italy** includes information on events, attractions, activities, and transportation for the independent traveler.
www.**lonelyplanet.com**/worldguide/destinations/europe/**italy** - 57k - Cached

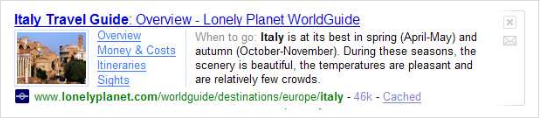
After



Topics for **Getting Pregnant** - BabyCenter
Tools
Before you Try
Active Trying
Having Trouble
BabyCenter's getting **pregnant** tools and information can boost your chances of conception by helping you chart your cycle, read your cervical mucus, and pinpoint ovulation.
www.**babycenter.com/getting-pregnant** - 76k - Cached

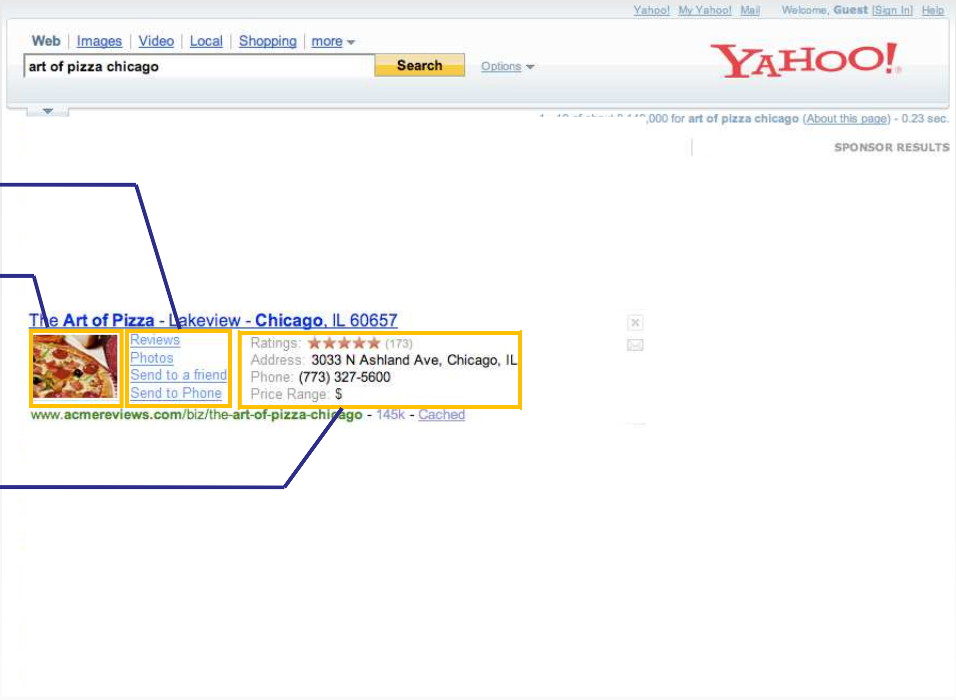


WebMD **Allergies** Health Center - Find allergy information and latest .
Overview
Symptoms
Diagnosiss
Treatment
Allergies are an abnormal response of the immune system. People who have allergies have an immune system that reacts to a usually harmless substance in the environment. This substance (pollen, mold, dander, etc.) is called an allergen.
www.**webmd.com/allergies** - 111k - Cached



Italy Travel Guide: Overview - Lonely Planet WorldGuide
Overview
Money & Costs
Itineraries
Sights
When to go: **Italy** is at its best in spring (April-May) and autumn (October-November). During these seasons, the scenery is beautiful, the temperatures are pleasant and are relatively few crowds.
www.**lonelyplanet.com**/worldguide/destinations/europe/**italy** - 46k - Cached

Enhanced Result



Yahoo! My Yahoo! Mail Welcome, **Guest** [Sign In] Help
Web | Images | Video | Local | Shopping | more
Search
Options
YAHOO!
SPONSOR RESULTS

art of pizza chicago

1 - 10 of about [illegible],000 for **art of pizza chicago** (About this page) - 0.23 sec.

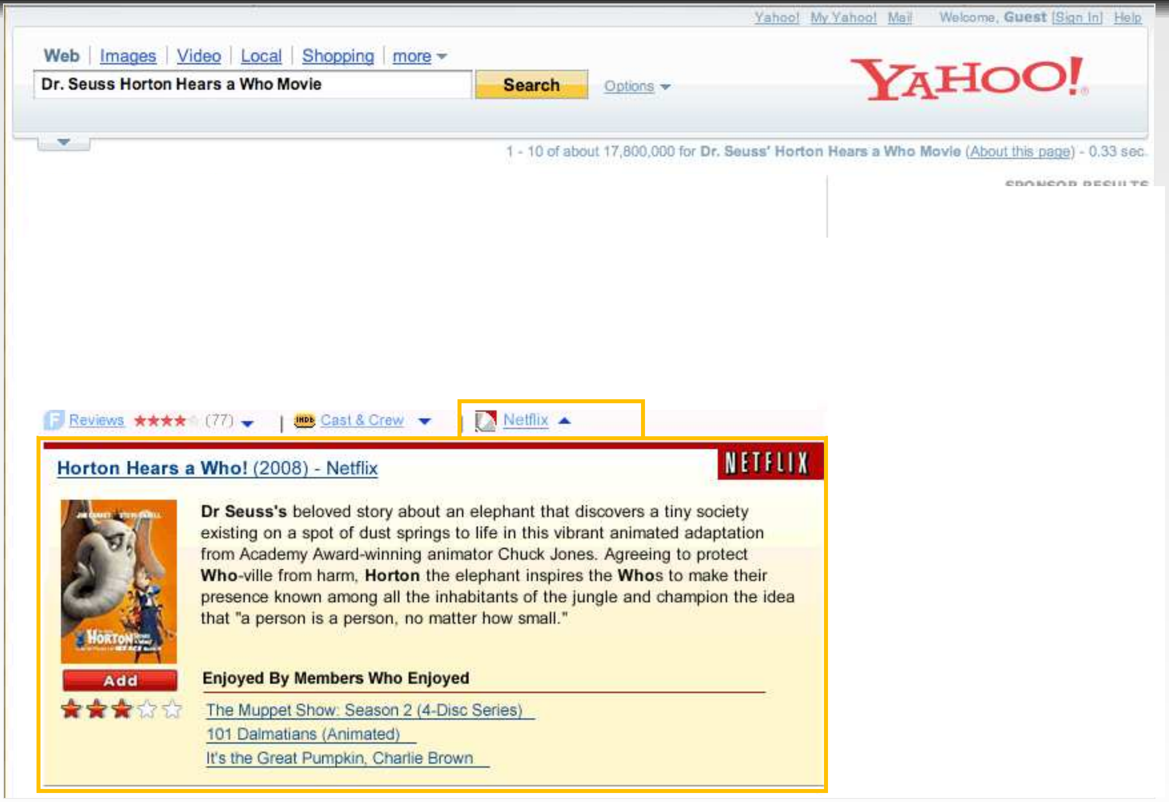
The **Art of Pizza** - Lakeview - **Chicago**, IL 60657
Reviews
Photos
Send to a friend
Send to Phone
Ratings: ★★★★★ (173)
Address: 3033 N Ashland Ave, Chicago, IL
Phone: (773) 327-5600
Price Range: $
www.**acmereviews.com**/biz/the-**art-of-pizza-chicago** - 145k - Cached

deep links

image

name/value pairs or abstract

Infobar



Yahoo! My Yahoo! Mail Welcome, **Guest** [Sign In] Help

Web | Images | Video | Local | Shopping | more

Dr. Seuss Horton Hears a Who Movie

Search

Options

1 - 10 of about 17,800,000 for **Dr. Seuss' Horton Hears a Who Movie** (About this page) - 0.33 sec.

SPONSOR RESULTS

Reviews ★★★★ (77) | Cast & Crew | Netflix

Horton Hears a Who! (2008) - Netflix

Dr Seuss's beloved story about an elephant that discovers a tiny society existing on a spot of dust springs to life in this vibrant animated adaptation from Academy Award-winning animator Chuck Jones. Agreeing to protect **Who**-ville from harm, **Horton** the elephant inspires the **Who**s to make their presence known among all the inhabitants of the jungle and champion the idea that "a person is a person, no matter how small."

Add

★★★☆☆

Enjoyed By Members Who Enjoyed

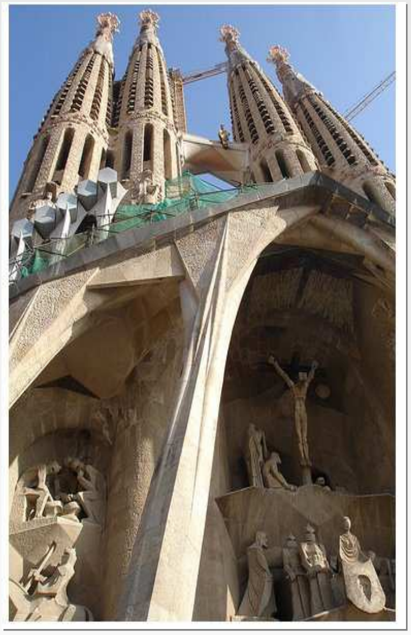
- The Muppet Show: Season 2 (4-Disc Series)
- 101 Dalmatians (Animated)
- It's the Great Pumpkin, Charlie Brown

UGC: Exploiting Flickr Tags

Tag Mining

- Objective:
  - Deploy collective knowledge that exists within Social Media services (Flickr and Delicious)
- Approach:
  1. Use tag co-occurrence statistics for media annotation and retrieval
  2. Semantic analysis of large tag-spaces

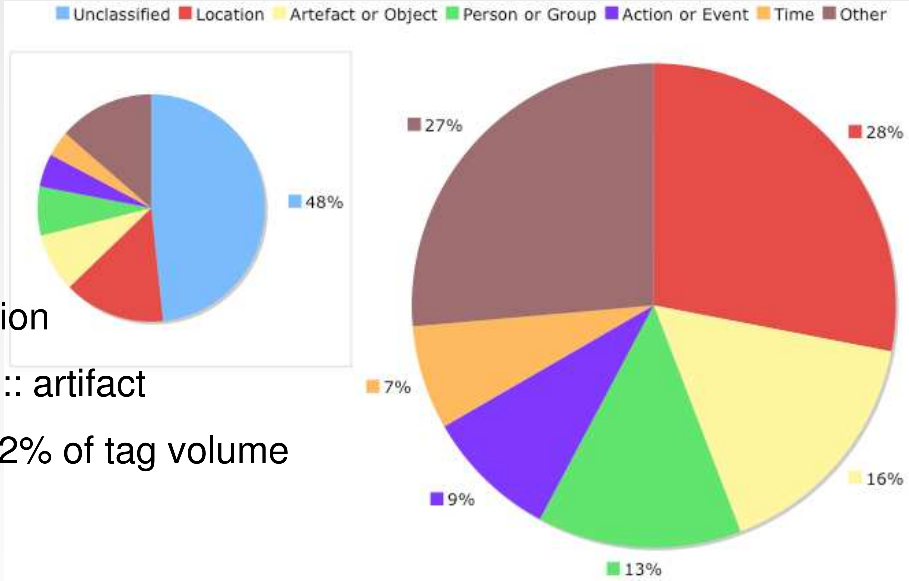
Tag Mining - Collective Knowledge



- Many users annotate photos of “La Sagrada Familia”:
  - Sagrada Familia, Barcelona
  - Sagrada Familia, Gaudi, architecture, church
  - church, Sagrada Familia
  - Sagrada Familia, Barcelona, Spain
- Derived collective knowledge:
  - Barcelona, Gaudi, church, architecture

Tag Mining - Semantics

- Assign tag semantics using WordNet broad categories

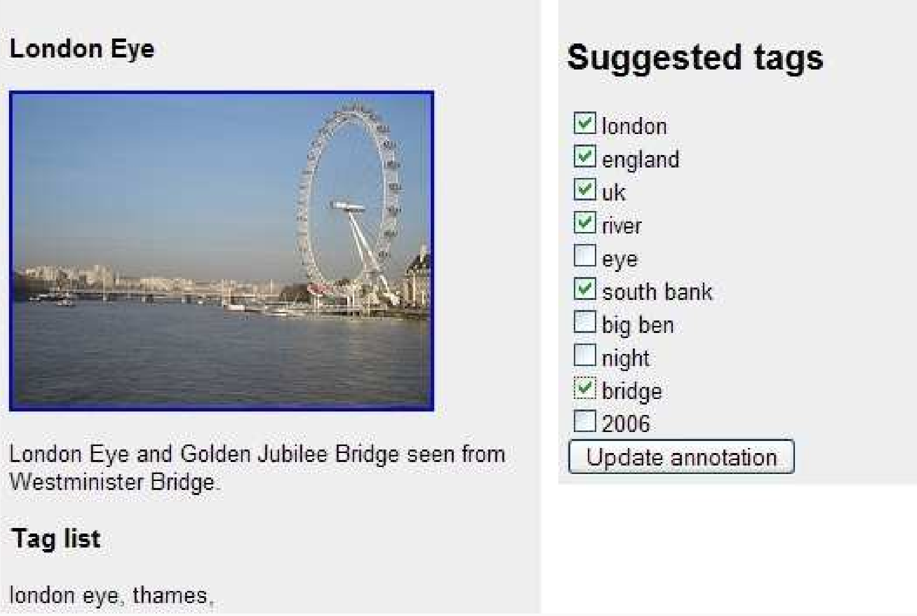


  - Paris :: location
  - Eiffel Tower :: artifact
  - Coverage: 52% of tag volume

Tag Mining - Semantics

- Extend this mapping using patterns found in Wikipedia
  - Upperbound for coverage: 78.6% of the tag volume
  - Based on SVM approach
    - Features: Wikipedia templates and categories
    - Training data: Wikipedia entries found in WordNet
  - Extended coverage: 68% of the tag volume
  - Mapping from Wikipedia pages to tags
    - Reduces ambiguity in the classification

Understanding tags



London Eye

London Eye and Golden Jubilee Bridge seen from Westminister Bridge.

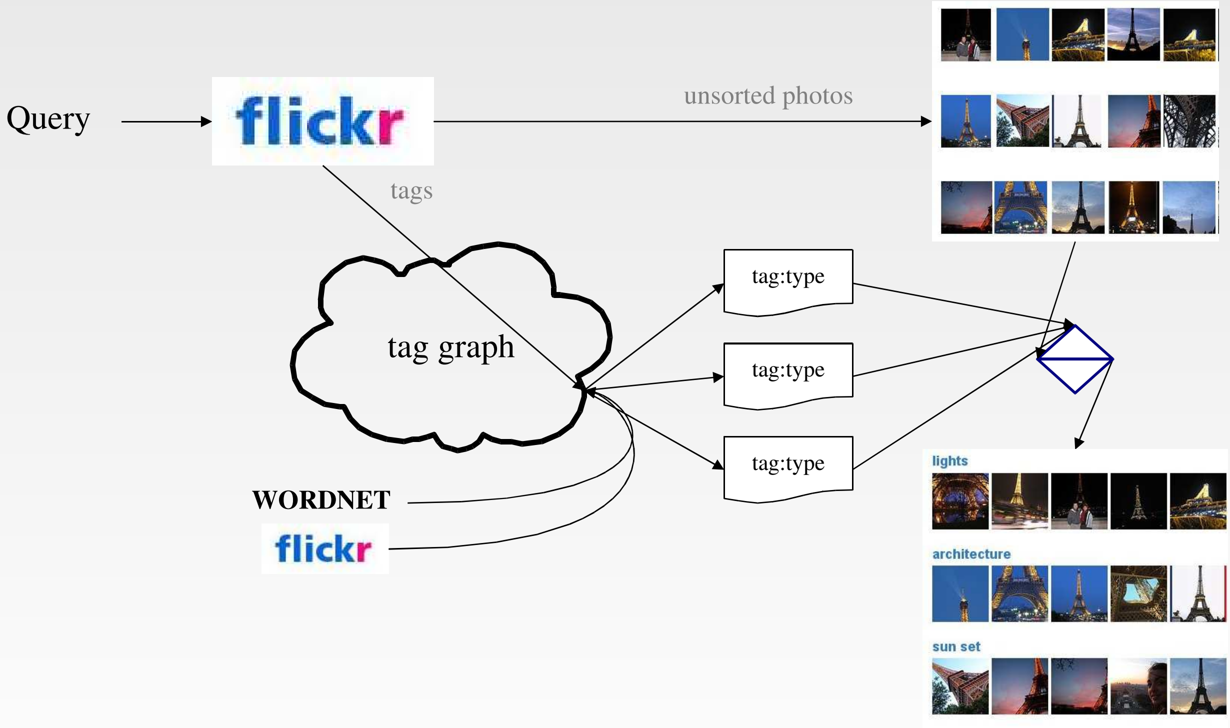
Tag list

london eye, thames,

Suggested tags

☑ london
☑ england
☑ uk
☑ river
□ eye
☑ south bank
□ big ben
□ night
☑ bridge
□ 2006

Semantic Photo Search



Query

flickr

unsorted photos

tags

tag graph

tag:type

tag:type

tag:type

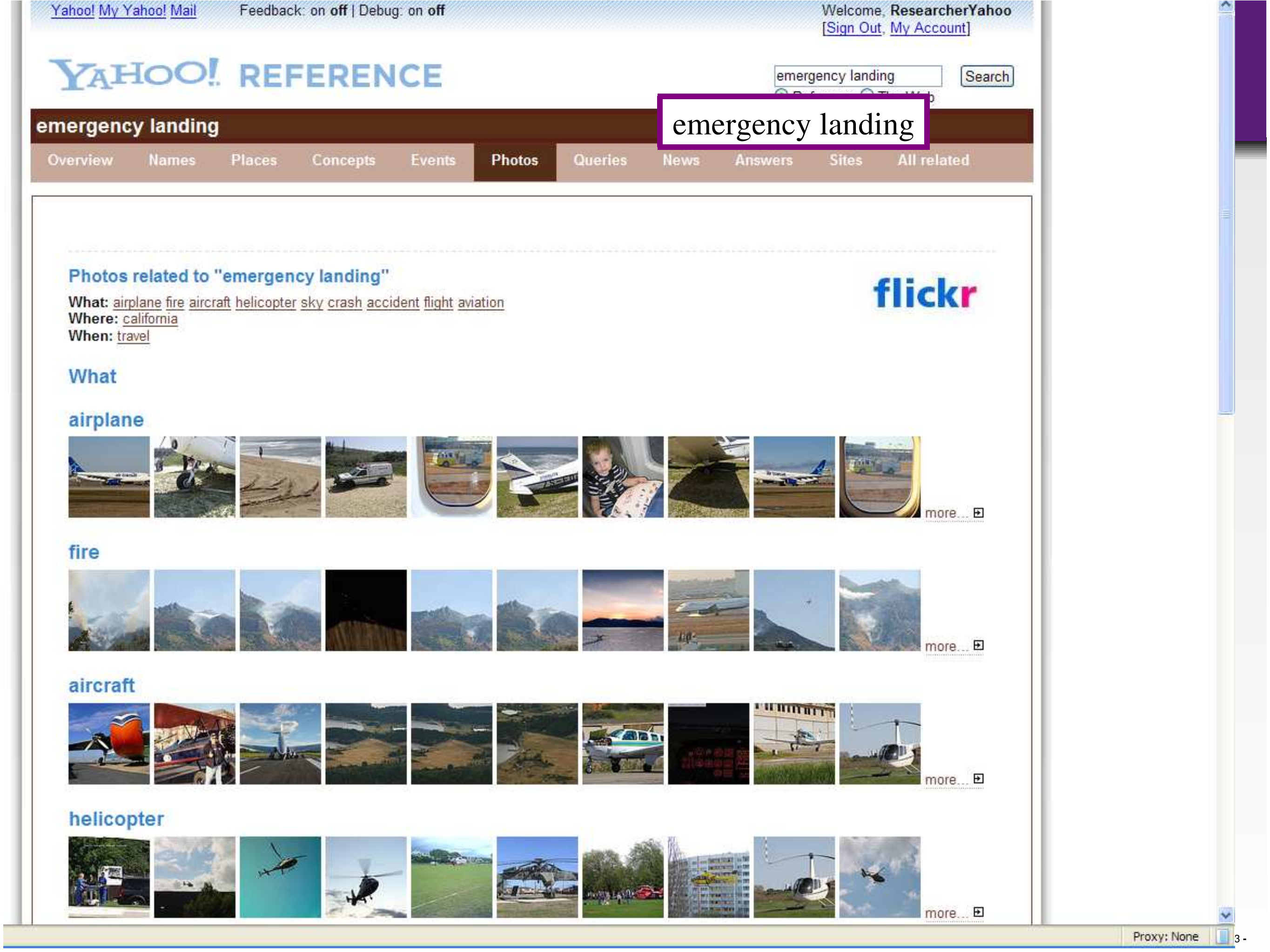
WORDNET

flickr

lights

architecture

sun set



emergency landing

emergency landing

Photos related to "emergency landing"

What: airplane fire aircraft helicopter sky crash accident flight aviation
Where: california
When: travel

What

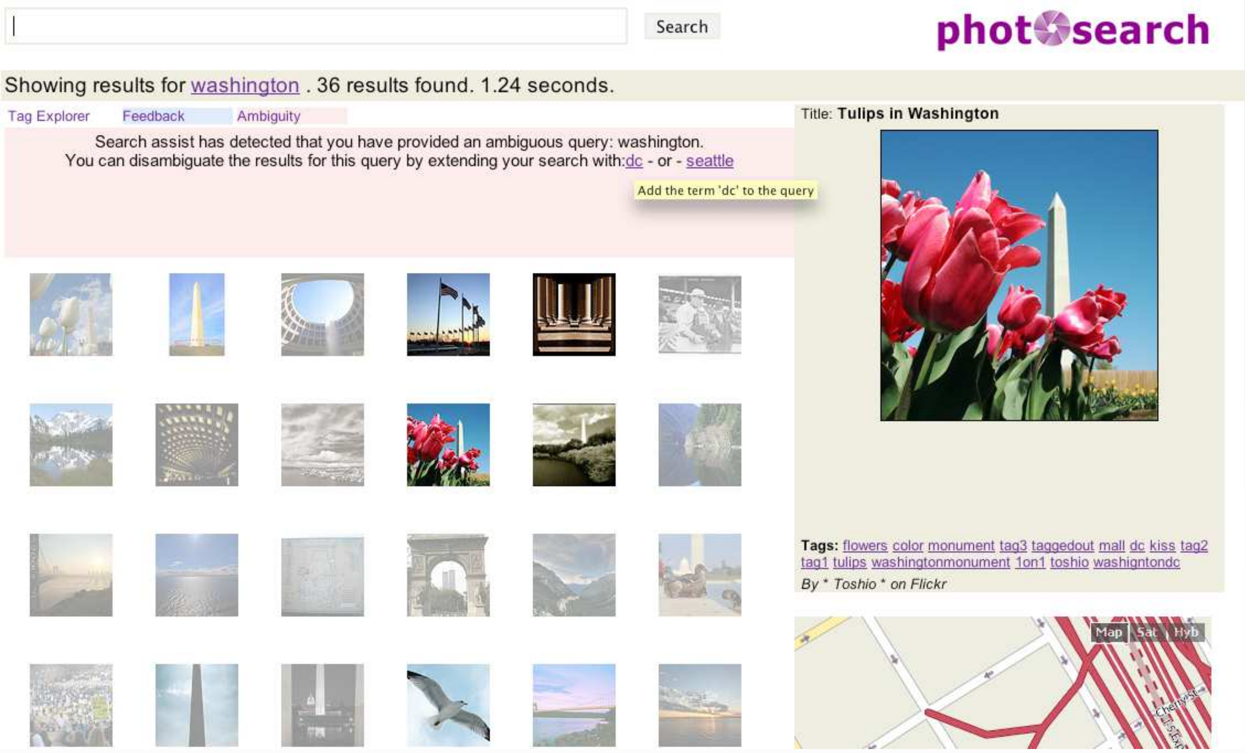
airplane

fire

aircraft

helicopter

Media Search - Demo



Showing results for washington . 36 results found. 1.24 seconds.

Search assist has detected that you have provided an ambiguous query: washington.
You can disambiguate the results for this query by extending your search with:dc - or - seattle

Add the term 'dc' to the query

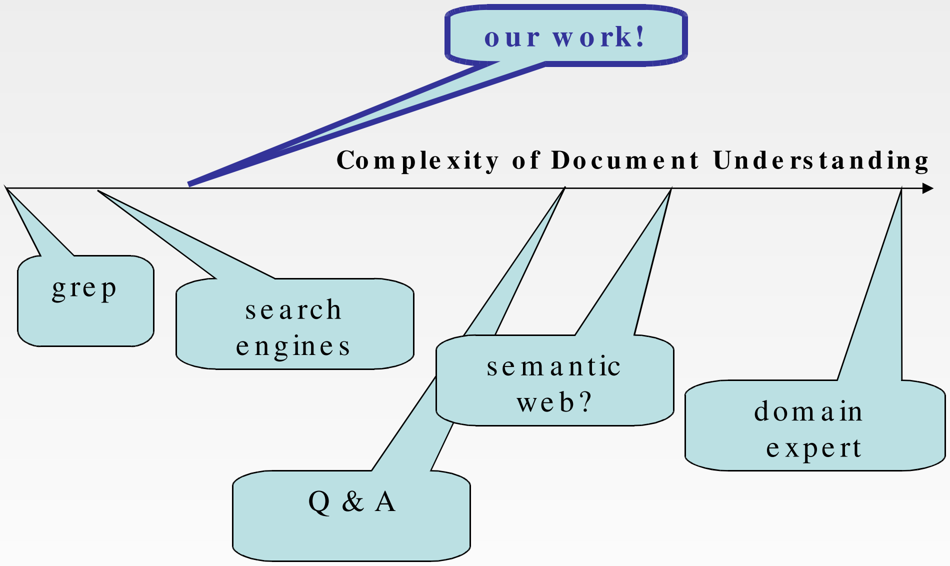
Title: **Tulips in Washington**

Tags: flowers color monument tag3 taggedout mall dc kiss tag2 tag1 tulips washingtonmonument 1on1 toshio washigntondc

*By * Toshio * on Flickr*

Understanding Text

Document Understanding Cartoon



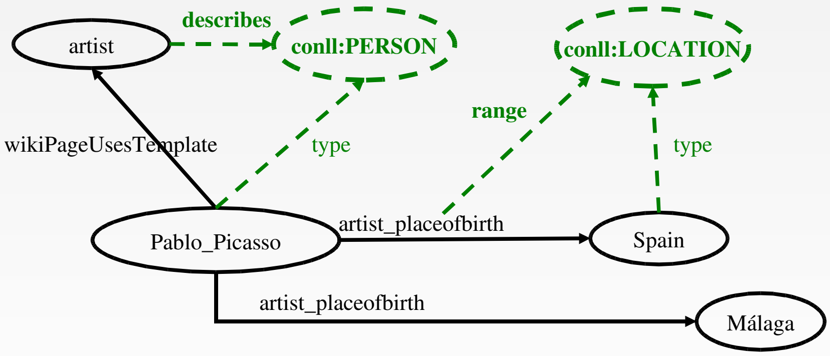
Extending metadata

```
Pablo Picasso was born in Málaga, Spain.
     PER                   LOC     LOC
```

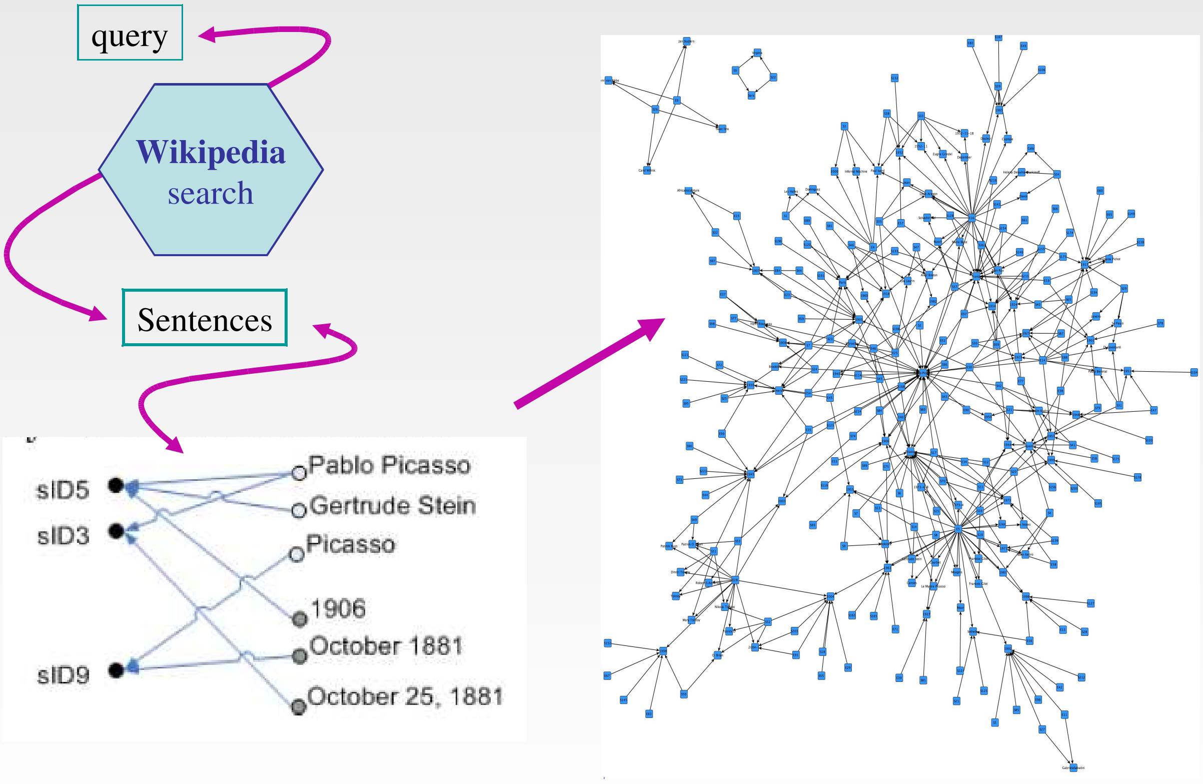
E:PERSON — GPE:CITY — GPE:COUNTRY

artist:name — artist:placeofbirth — artist:placeofbirth

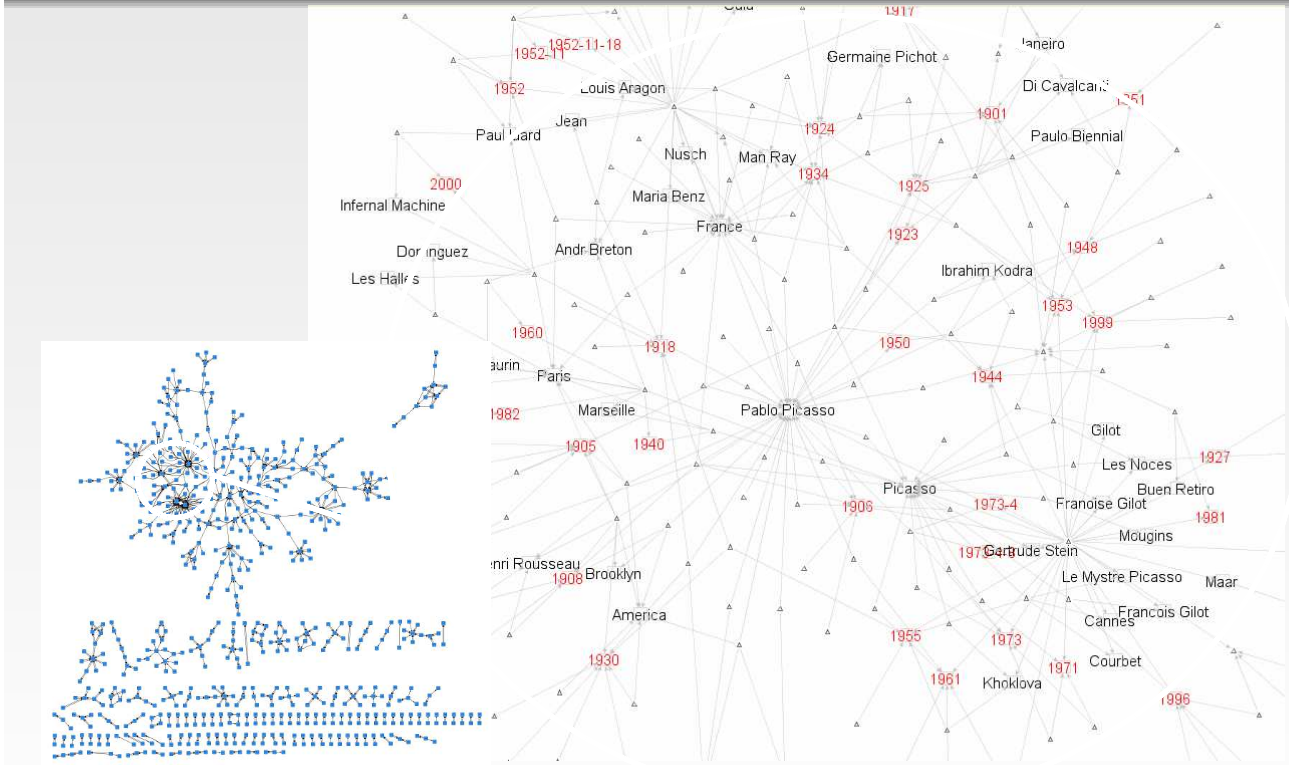
If most artists are persons, than let’s assume all artists are persons.
If most places of birth are locations, then let’s assume all are.

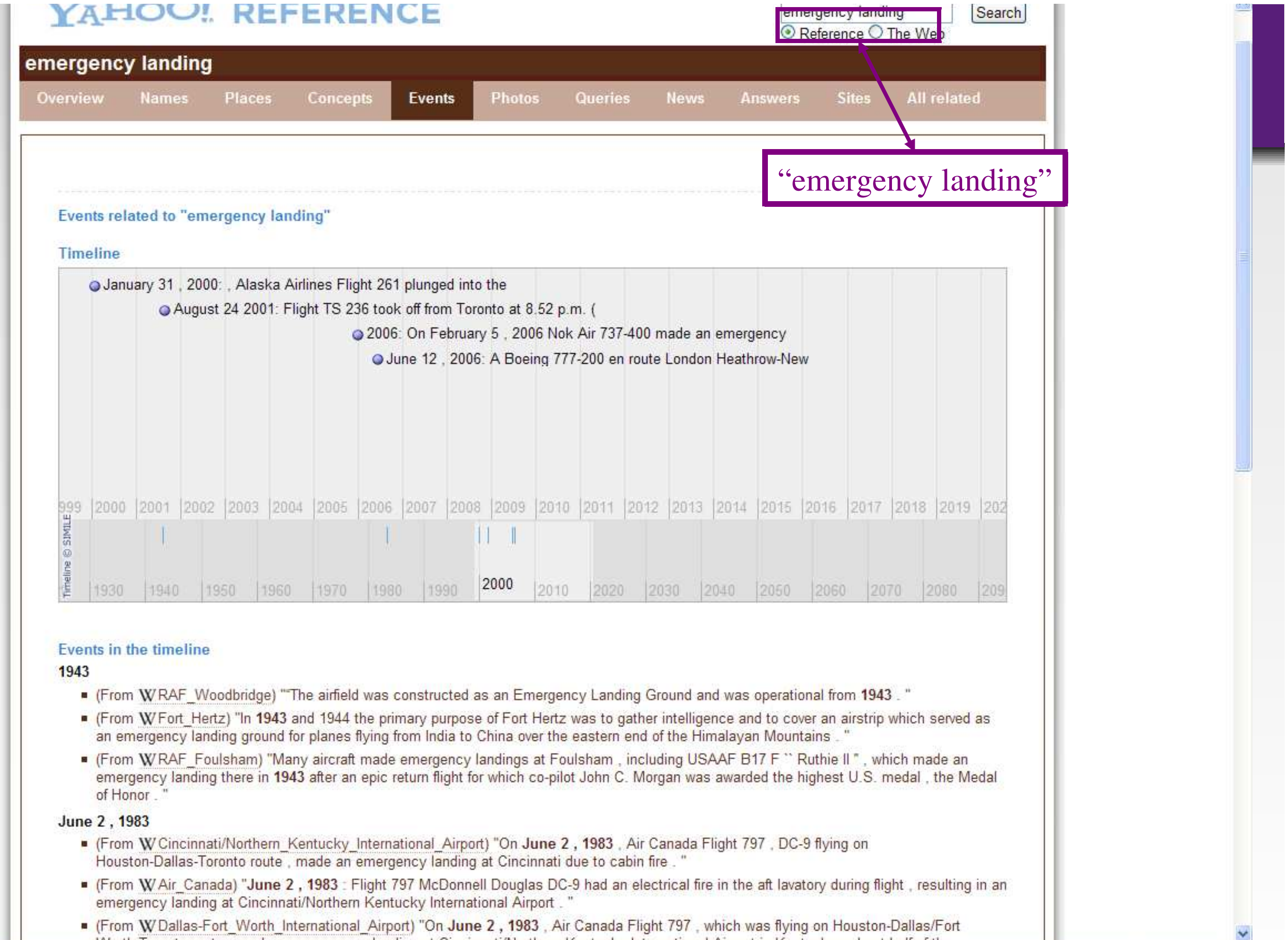


Entity Containment Graph



Example: Picasso





emergency landing

Events related to "emergency landing"

Timeline

- January 31 , 2000: , Alaska Airlines Flight 261 plunged into the
- August 24 2001: Flight TS 236 took off from Toronto at 8.52 p.m. (
- 2006: On February 5 , 2006 Nok Air 737-400 made an emergency
- June 12 , 2006: A Boeing 777-200 en route London Heathrow-New

Events in the timeline

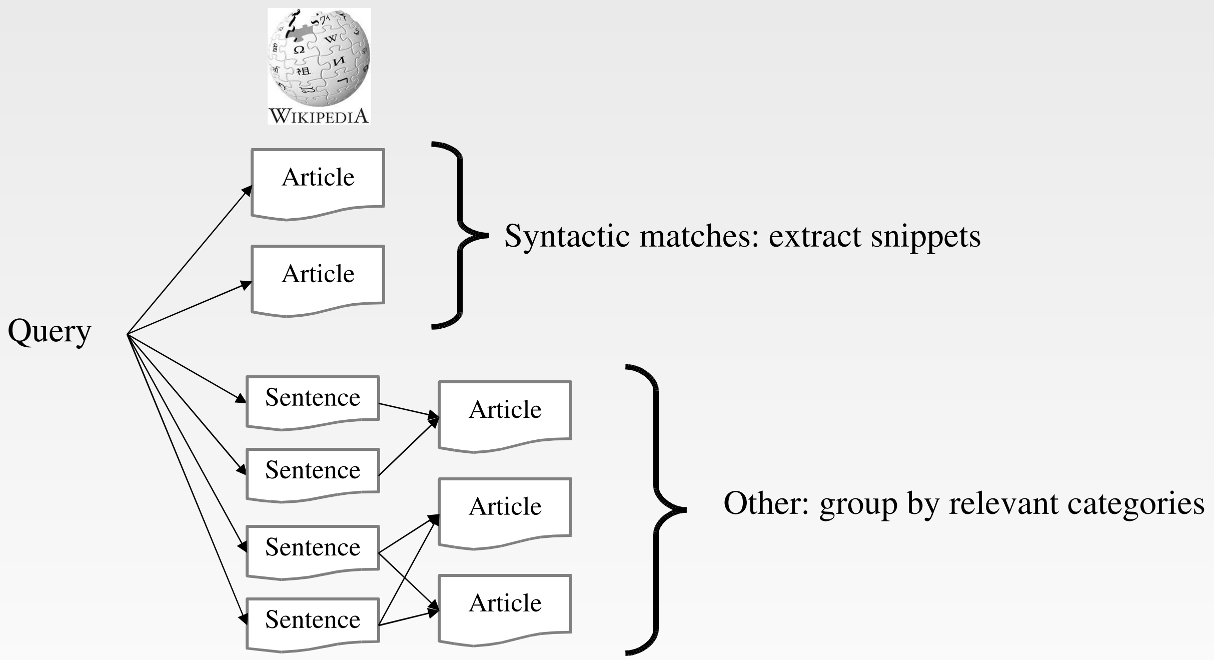
1943

- (From W RAF_Woodbridge) ""The airfield was constructed as an Emergency Landing Ground and was operational from **1943** . "
- (From W Fort_Hertz) "In **1943** and 1944 the primary purpose of Fort Hertz was to gather intelligence and to cover an airstrip which served as an emergency landing ground for planes flying from India to China over the eastern end of the Himalayan Mountains . "
- (From W RAF_Foulsham) "Many aircraft made emergency landings at Foulsham , including USAAF B17 F `` Ruthie II " , which made an emergency landing there in **1943** after an epic return flight for which co-pilot John C. Morgan was awarded the highest U.S. medal , the Medal of Honor . "

June 2 , 1983

- (From W Cincinnati/Northern_Kentucky_International_Airport) "On **June 2 , 1983** , Air Canada Flight 797 , DC-9 flying on Houston-Dallas-Toronto route , made an emergency landing at Cincinnati due to cabin fire . "
- (From W Air_Canada) "**June 2 , 1983** : Flight 797 McDonnell Douglas DC-9 had an electrical fire in the aft lavatory during flight , resulting in an emergency landing at Cincinnati/Northern Kentucky International Airport . "
- (From W Dallas-Fort_Worth_International_Airport) "On **June 2 , 1983** , Air Canada Flight 797 , which was flying on Houston-Dallas/Fort

Synthetic Document



Syntactic matches: extract snippets

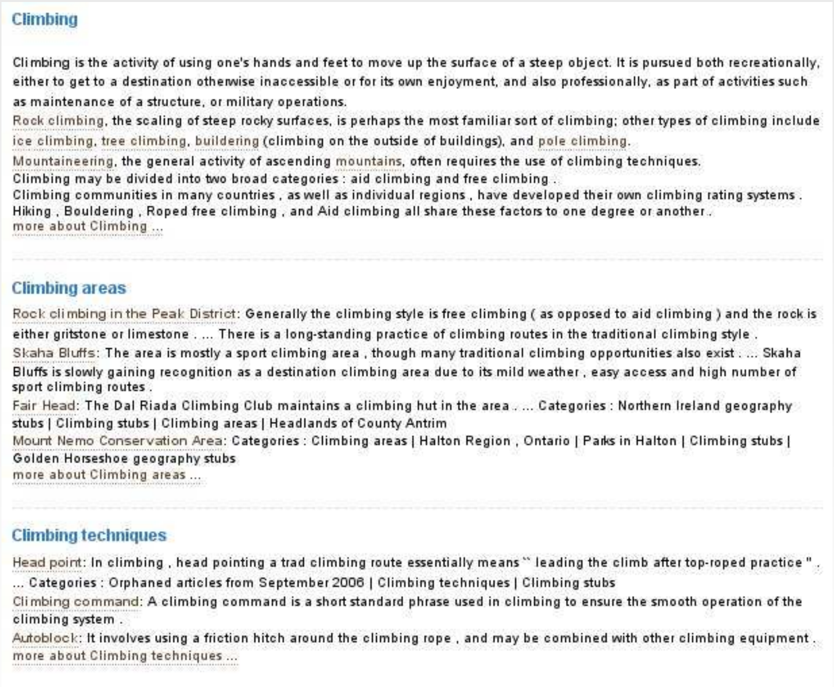
Other: group by relevant categories

Synthetic Document

Article: Article

Category: Category, Article, Article, Article, Article

Category: Category, Article, Article, Article



Climbing

Climbing is the activity of using one's hands and feet to move up the surface of a steep object. It is pursued both recreationally, either to get to a destination otherwise inaccessible or for its own enjoyment, and also professionally, as part of activities such as maintenance of a structure, or military operations.
Rock climbing, the scaling of steep rocky surfaces, is perhaps the most familiar sort of climbing; other types of climbing include ice climbing, tree climbing, buildering (climbing on the outside of buildings), and pole climbing.
Mountaineering, the general activity of ascending mountains, often requires the use of climbing techniques.
Climbing may be divided into two broad categories : aid climbing and free climbing .
Climbing communities in many countries , as well as individual regions , have developed their own climbing rating systems .
Hiking , Bouldering , Roped free climbing , and Aid climbing all share these factors to one degree or another ..
more about Climbing ...

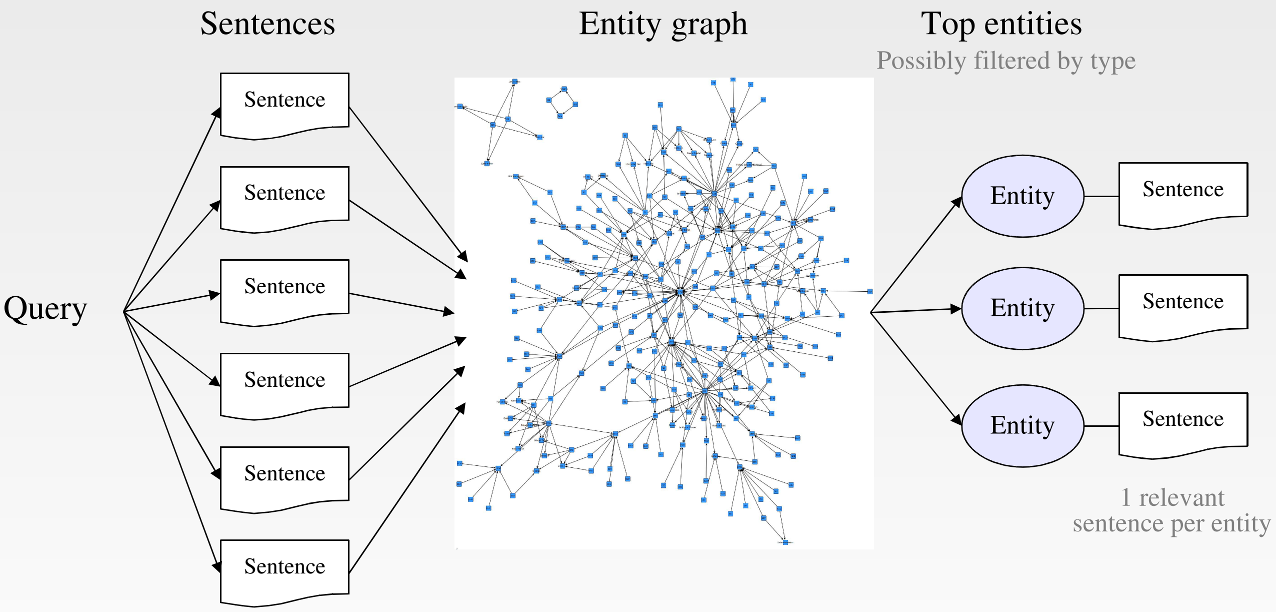
Climbing areas

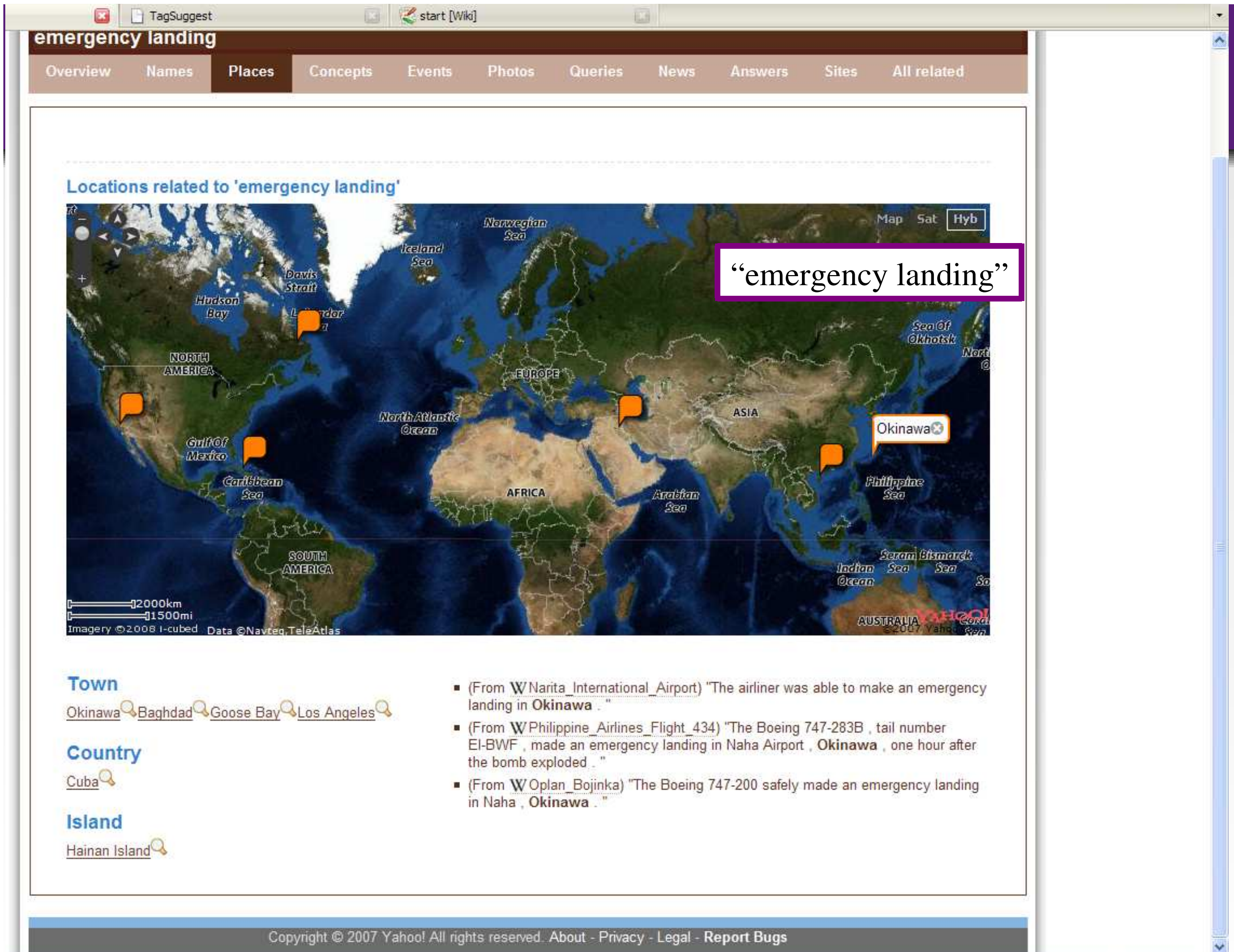
Rock climbing in the Peak District: Generally the climbing style is free climbing (as opposed to aid climbing) and the rock is either gritstone or limestone . ... There is a long-standing practice of climbing routes in the traditional climbing style .
Skaha Bluffs: The area is mostly a sport climbing area , though many traditional climbing opportunities also exist . ... Skaha Bluffs is slowly gaining recognition as a destination climbing area due to its mild weather , easy access and high number of sport climbing routes .
Fair Head: The Dal Riada Climbing Club maintains a climbing hut in the area .. ... Categories : Northern Ireland geography stubs | Climbing stubs | Climbing areas | Headlands of County Antrim
Mount Nemo Conservation Area: Categories : Climbing areas | Halton Region , Ontario | Parks in Halton | Climbing stubs | Golden Horseshoe geography stubs
more about Climbing areas ...

Climbing techniques

Head point: In climbing , head pointing a trad climbing route essentially means `` leading the climb after top-roped practice " . ... Categories : Orphaned articles from September 2006 | Climbing techniques | Climbing stubs
Climbing command: A climbing command is a short standard phrase used in climbing to ensure the smooth operation of the climbing system .
Autoblock: It involves using a friction hitch around the climbing rope , and may be combined with other climbing equipment .
more about Climbing techniques ...

Related Entities Search





“emergency landing”

emergency landing

Locations related to 'emergency landing'

Town

Okinawa Baghdad Goose Bay Los Angeles

Country

Cuba

Island

Hainan Island

- (From W Narita_International_Airport) "The airliner was able to make an emergency landing in **Okinawa** . "
- (From W Philippine_Airlines_Flight_434) "The Boeing 747-283B , tail number EI-BWF , made an emergency landing in Naha Airport , **Okinawa** , one hour after the bomb exploded . "
- (From W Oplan_Bojinka) "The Boeing 747-200 safely made an emergency landing in Naha , **Okinawa** . "

Syntactic/Semantic Tagging

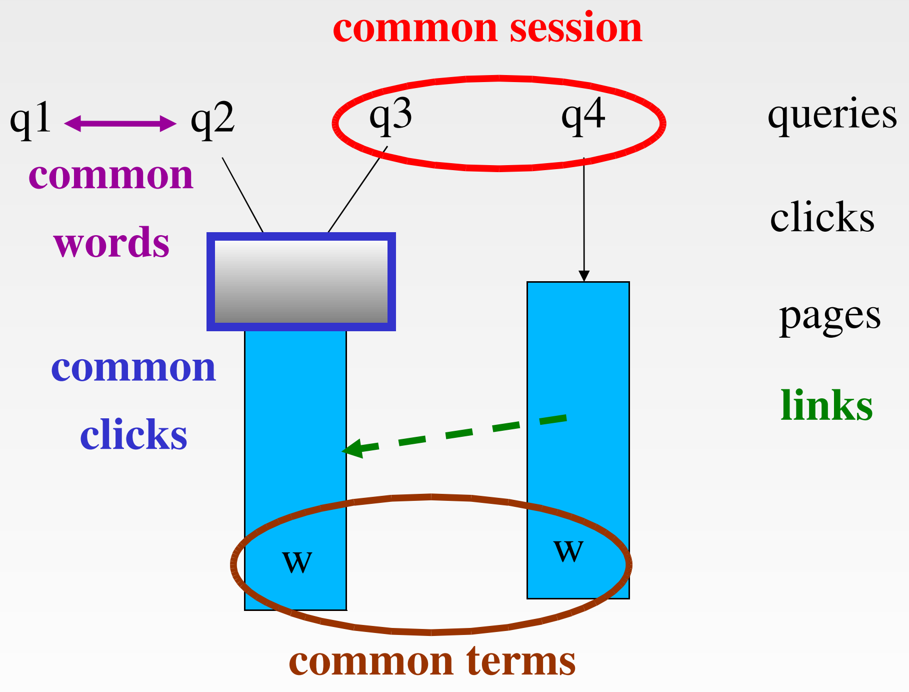
- Goals: Identify multiword expressions and entities, support generalization, coarse disambiguation
- Tagger: Average perceptron HMM (Collins, 2002) general purpose tagger: efficient (millions of features, hundreds of classes), fast (thousands of sentences/sec)
- State of the art: 3rd NIST Automatic Content Extraction Evaluation (ACE) 2007 (Surdeanu & Ciaramita, 2007)
- Tasks: PoS Tagging, supersense tagging (Ciaramita & Altun, 2006), named entity detection (CoNLL, BBN-WSJ, ACE, etc.)
- Research problems: robustness on Web data (domain adaptation), learning/evaluating from user-generated data (Mika et al., forthcoming)

Parsing/SRL

- Goal: extract structured information at sentence level (beyond the bag of words/document-centric models)
- Dependency parsing:
  - Parser: Fast discriminative multilingual Shift/Reduce parsing (hundreds of sentences/sec) (Attardi, 2006); 2nd in Adaptation Task of CoNLL 2007 (Attardi et al., 2007)
- Semantic role labeling:
  - Joint parsing and SRL: 3rd best system at CoNLL 2008 (Forthcoming)
- Research problems: how can structured linguistic representations be used to improve search/ranking problems? (Surdeanu, Ciaramita & Zaragoza, ACL 2008)

Web Usage:
Extracting Semantics from Queries

Relating Queries (Baeza-Yates, 2007)



common session

q1 q2 q3 q4

queries

common

words

clicks

pages

common

clicks

links

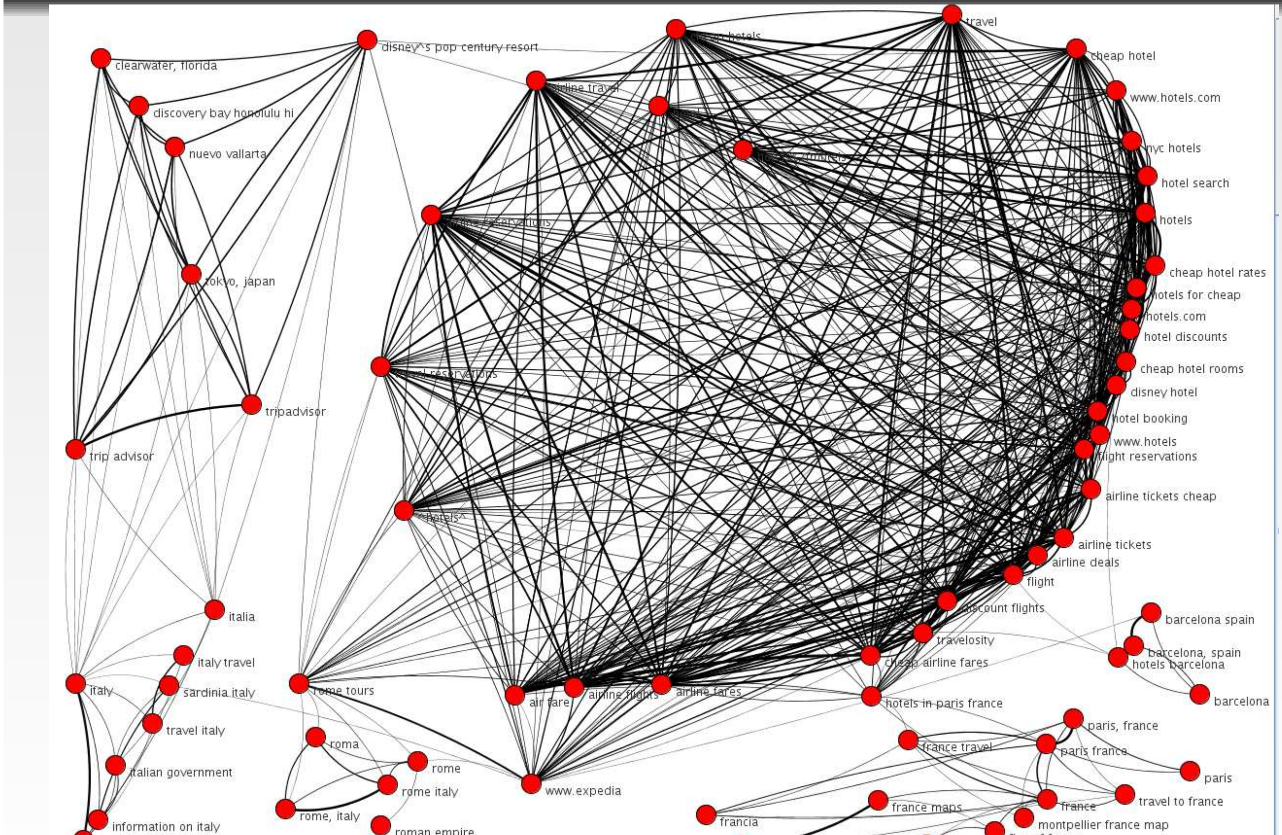
w w

common terms

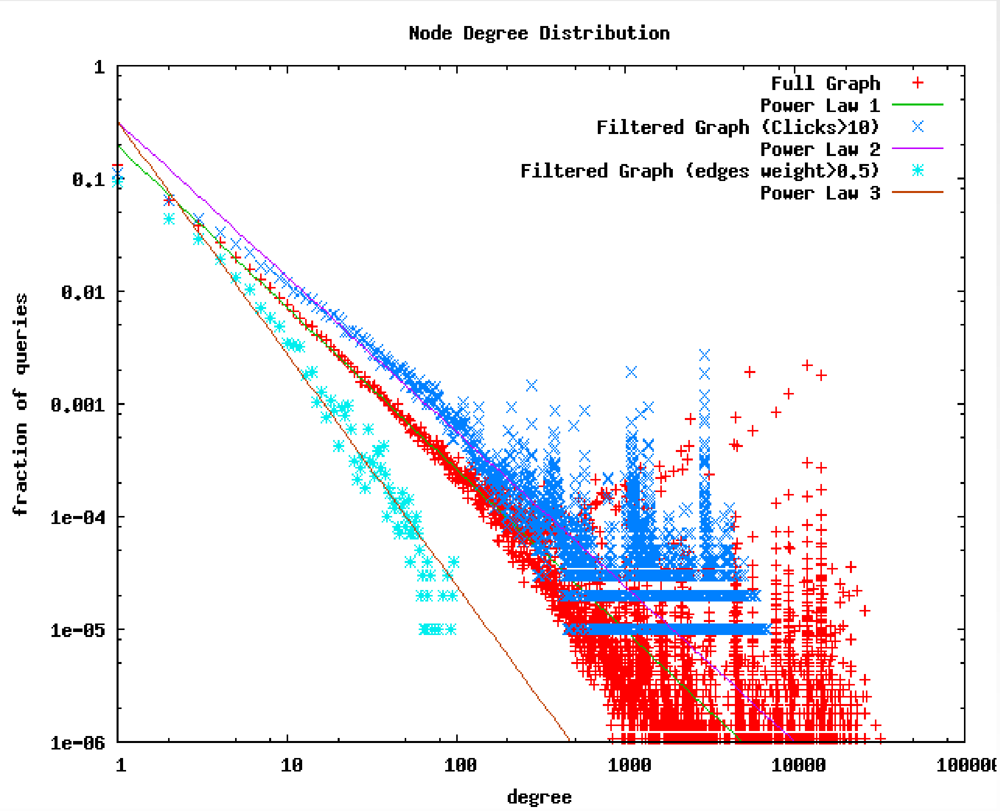
Qualitative Analysis

Graph	Strength	Sparsity	Noise
Word	Medium	High	Polysemy
Session	Medium	High	Physical sessions
Click	High	Medium	**Multitopic pages** **Click spam**
Link	Weak	Medium	Link spam
Term	Medium	Low	Term spam

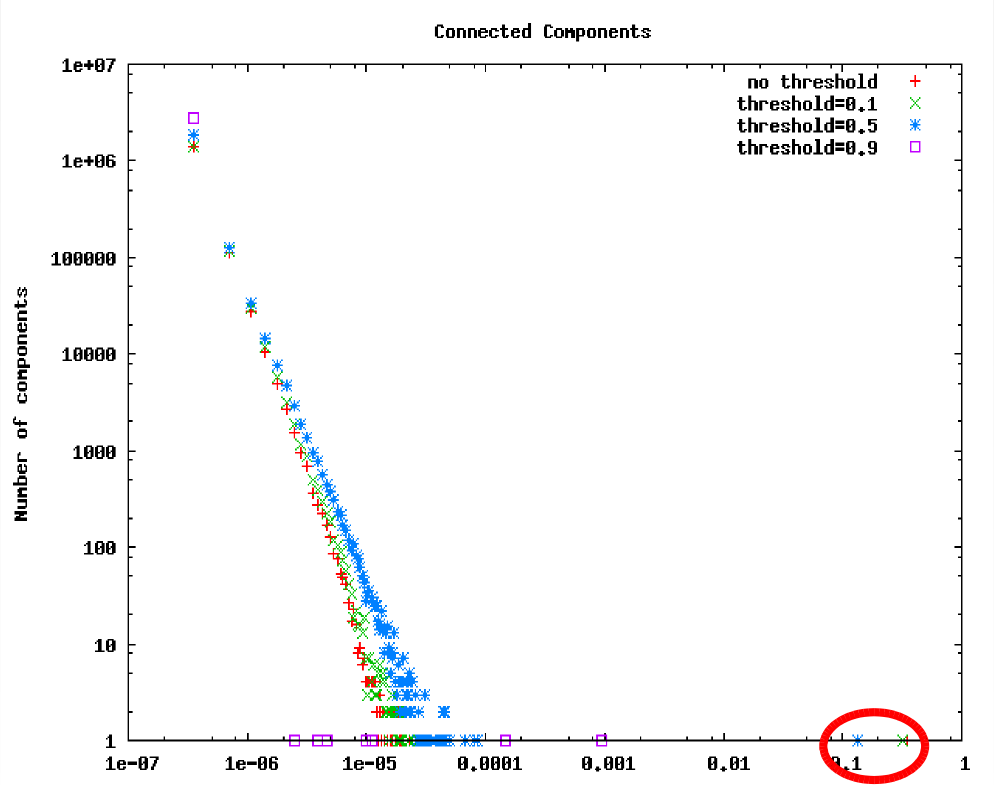
Click Graph



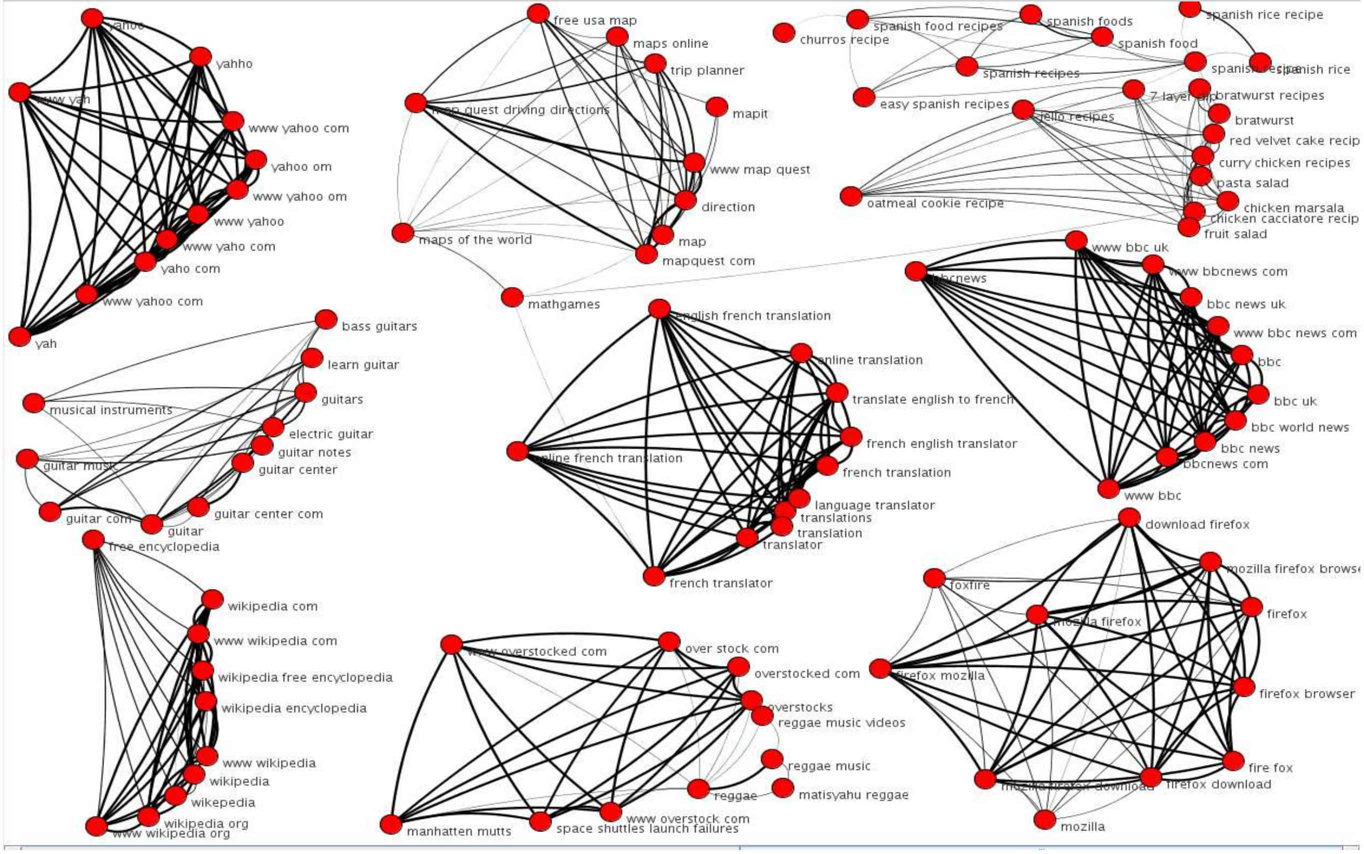
Node Degree Distribution



Connected Components



Implicit Folksonomy?



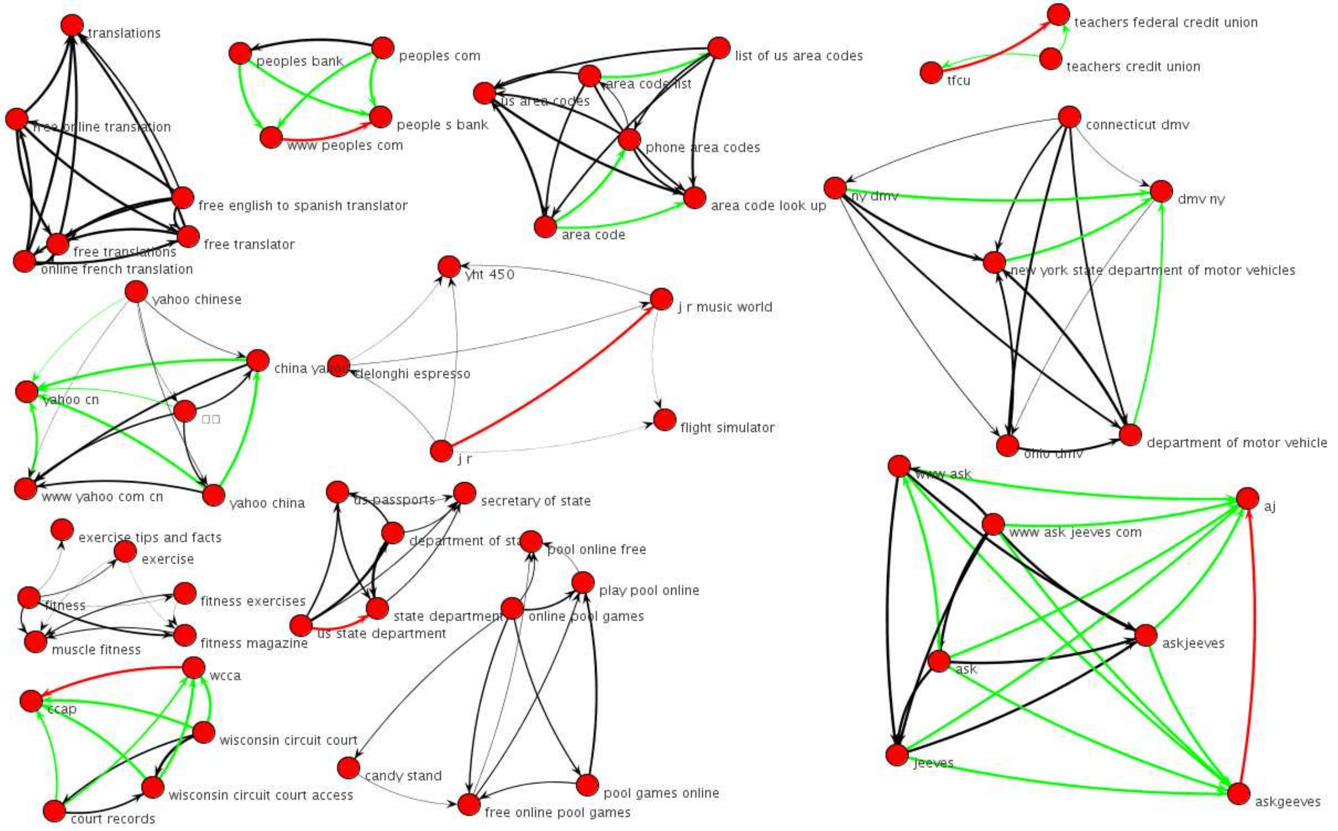
Set Relations and Graph Mining

- Identical sets: **synonyms**
- Subsets: **specificity**
  - directed edges
- Non empty intersections (with threshold)
  - degree of relation
- Dual graph: URLs related by queries
  - High degree: multi-topical URLs

Baeza-Yates & Tiberi

ACM KDD 2007

Implicit Knowledge? Webslang!



Evaluation: ODP Similarity

- A simple measure of similarity among queries using ODP categories
  - Define the similarity between two categories as the length of the longest shared path over the length of the longest path
  - Let *c_1,.., c_k* and *c'_1,.., c'_k* be the top *k* categories for two queries. Define the similarity (@*k*) between the two queries as *max{ sim(c_i,c'_j) | i,j=1,..,K }*

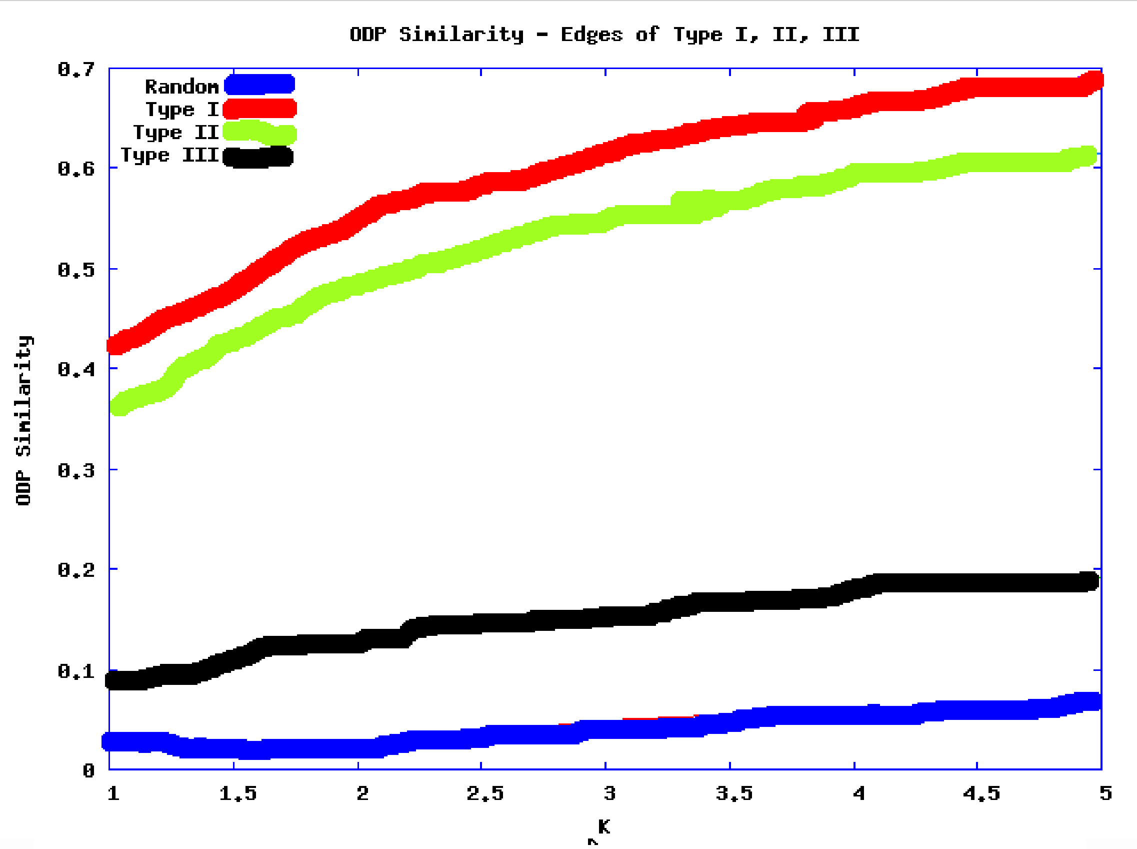
ODP Similarity

- Suppose you submit the queries “*Spain”* and “*Barcelona*” to ODP.
- The first category matches you get are:
  - Regional/ Europe/ Spain
  - Regional/ Europe/ Spain/ Autonomous Communities/ Catalonia/ Barcelona
- Similarity @1 is 1/2 because the longest shared path is “Regional/ Europe/ Spain” and the length of the longest is 6

Experimental Evaluation

- We evaluated a sample of 1,000 thousand edges for each kind of relation
- We also evaluated a sample of 1,000 random pairs of not adjacent queries (baseline)
- We studied the similarity as a function of *k* (the number of categories used)

Experimental Evaluation

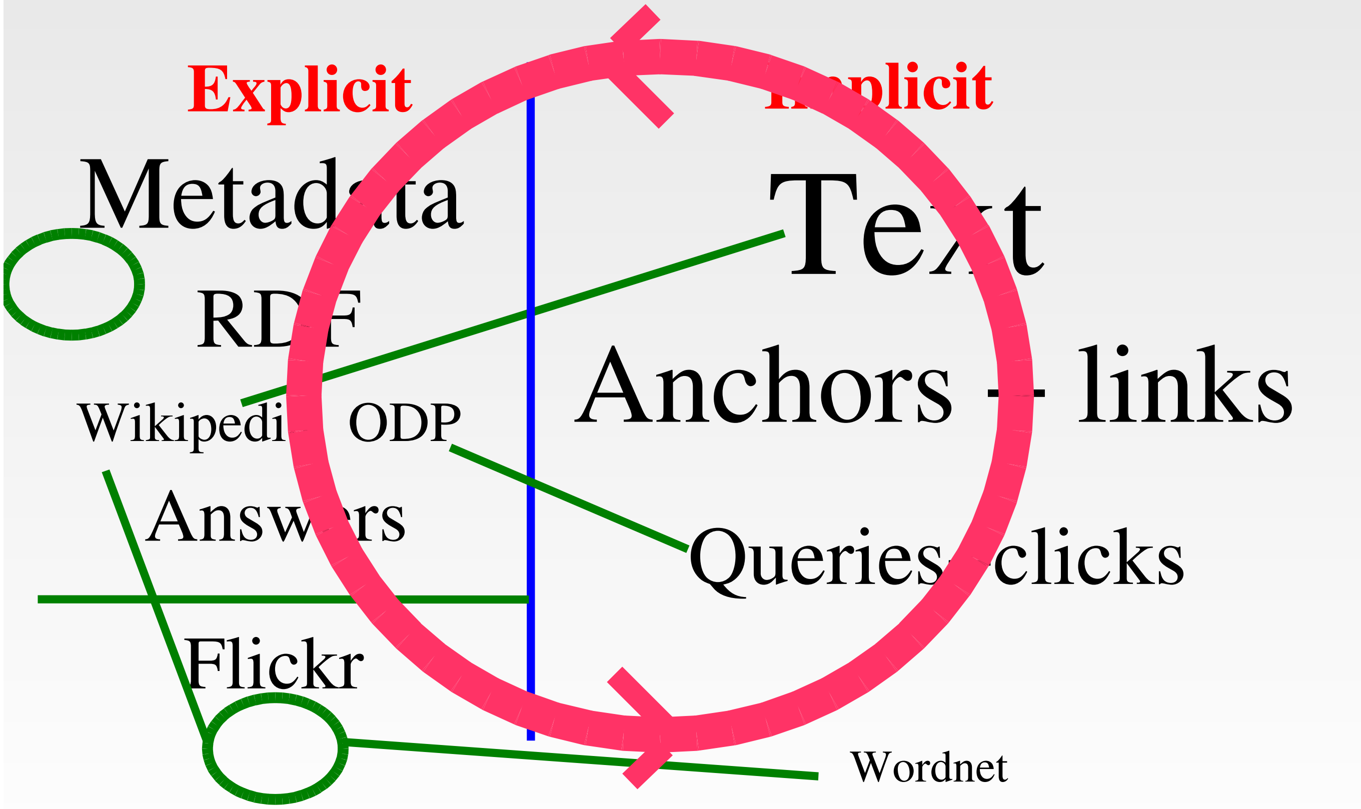


Final Remarks

Open Issues

- Explicit vs. implicit social networks
  - Any fundamental similarities?
- How to evaluate with (small) partial knowledge?
  - Data volume amplifies the problem
- User aggregation vs. personalization
  - Optimize common tasks
  - Move away from privacy issues

The Virtuous Cycle



Explicit

Metadata

RDF

Wikipedia ODP

Answers

Flickr

Implicit

Text

Anchors -- links

Queries, clicks

Wordnet

The Future: Web 3.0?

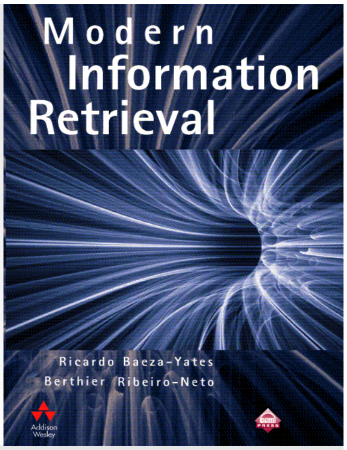
- We are at Web 2.0 beta
- People wants to get tasks done
  - Where I do go for a original holiday with 1,000 US$?
- Take in account the context of the task

Start

Yahoo! Experience

Finish

Second edition coming soon



No Questions?

Contact: rbaeza@acm.org

Thanks to Massi Ciaramita, Peter Mika, Borkur Sigurbjornsson, Mihai Surdeanu, Andrew Tomkins, Roelof van Zwol, Hugo Zaragoza