



Mining the Web 2.0 to improve Image Search

Ricardo Baeza-Yates
VP, Yahoo! Research

SEMEDIA

- Search Environments for Media
 - Focused on image & video search



Agenda

- Motivation: Search
- The Power of Data
- Examples
 - Exploiting Flickr Tags
 - Using Visual Annotations
 - Faceted Image Retrieval
- Impacts not only relevance but also the UI
- Concluding Remarks

-3-

Motivation

- Web search is no longer about document retrieval
 - Means for web-mediated goals
- New breed of search experiences
 - Demands search ecosystem combining content with intent
 - Exploiting the Wisdom of Crowds behind the Web 2.0

-4-

Trends

- User Generated Content
 - Massive (quality vs. quantity)
 - Social Networks
 - Real time (people + sensors)
- Impact
 - Fragmentation of ownership
 - Fragmentation of access (longer tail)
 - Fragmentation of right to access
- Viability
 - Business model based in advertising

-5-

Search is Evolving

- Already, more than a list of docs
- Moving towards identifying a user's task
- Enabling means for task completion
- New experiences based on the Web 2.0
- Challenges: on-line, scalability

-7-

More complete information

Shortcuts

Deep Links

Enhanced Results

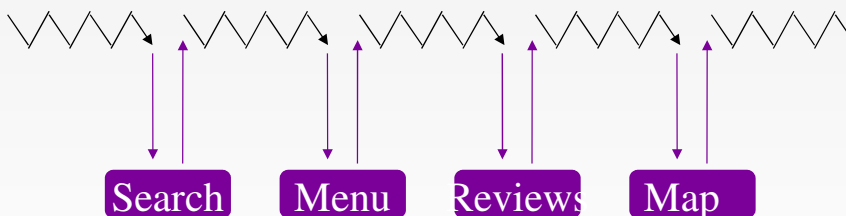
- 8 -

Search: Content vs. Intent

Premise:

- People don't want to search
- People want to get tasks done and get straight to their answers

Start *I am craving for good coffee and baklava in Thira* → Finish

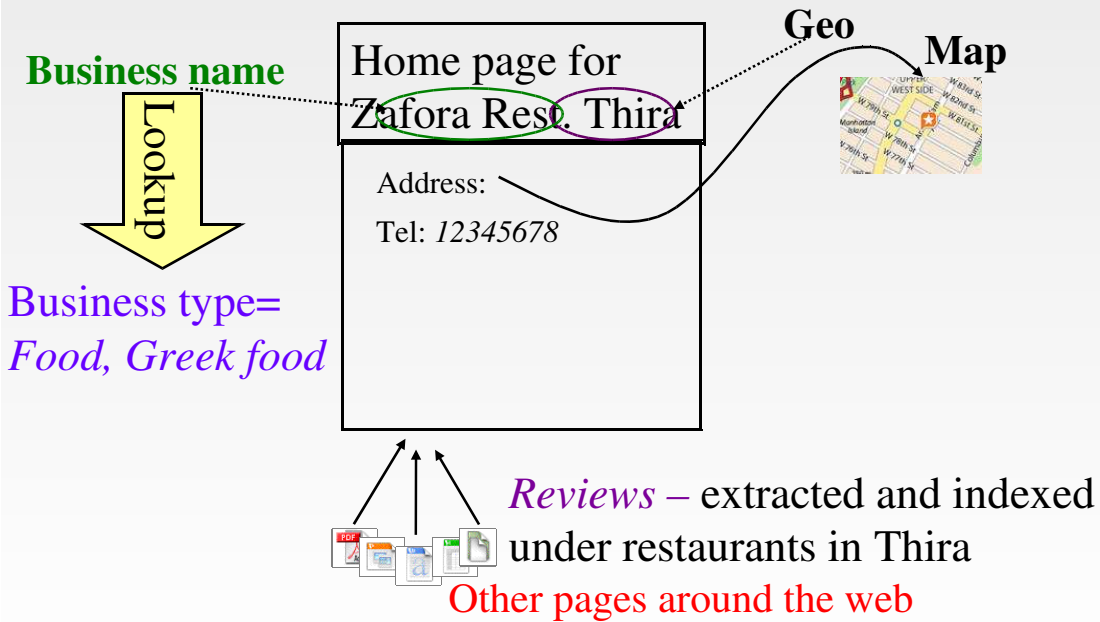


9

- 9 -

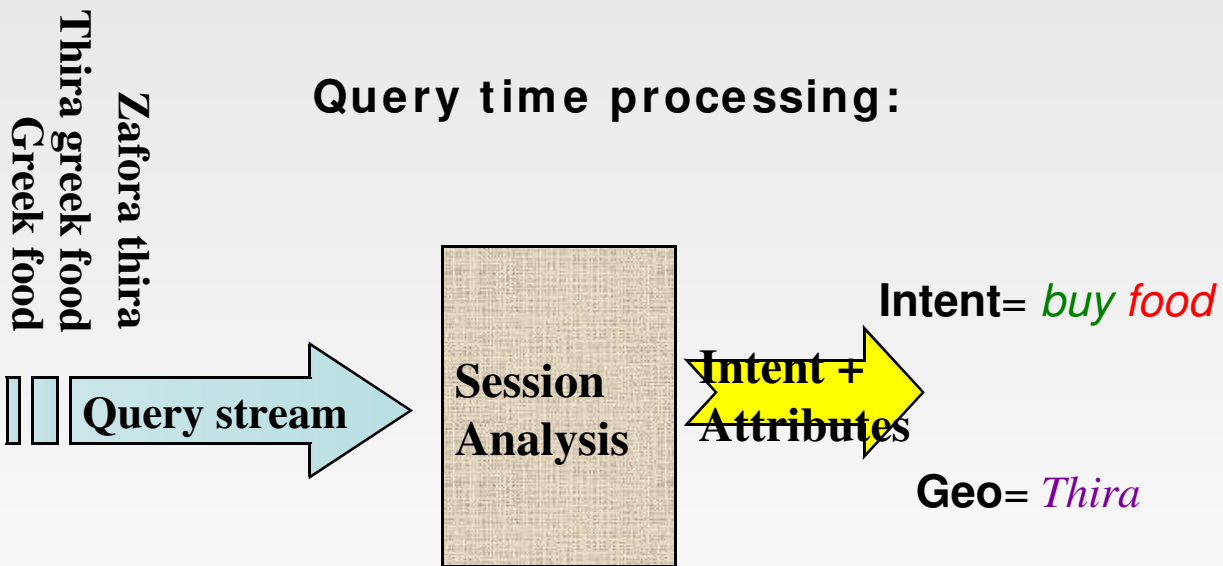
How this might work – I

Index time processing:



How this might work – II

Query time processing:



Net

- We move from a web of pages to a **web of objects**
- Objects are **people, places, businesses, restaurants ...**
- Objects have attributes
 - Missing, noisy, etc.
- Intents are satisfied by presenting **objects and attributes**
- Attributes define faceted search

12

- 12 -

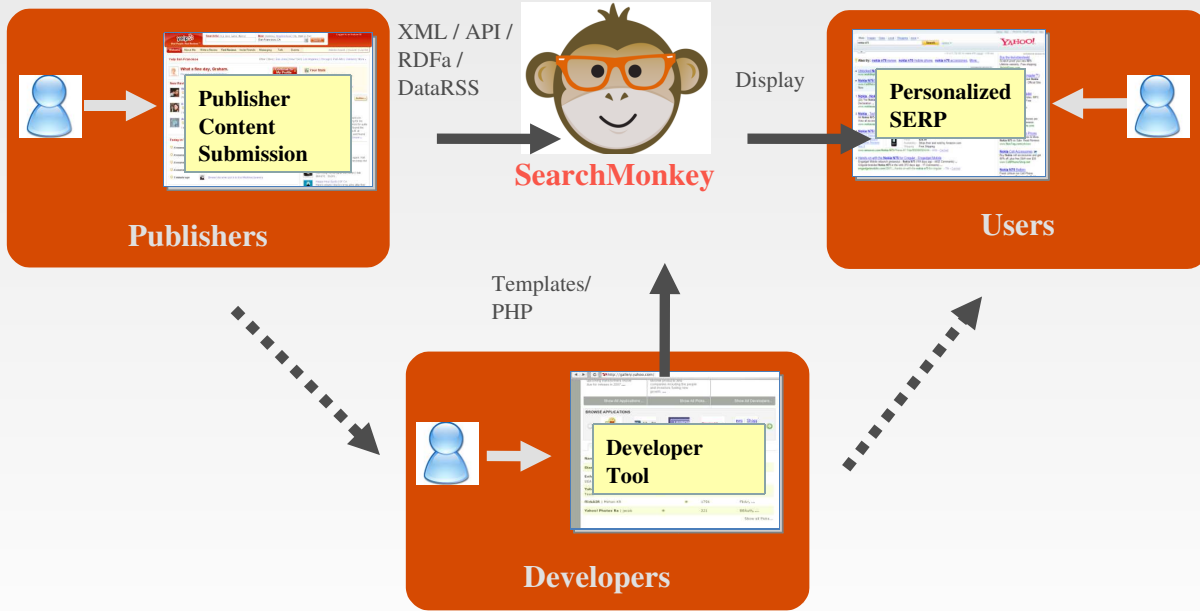
How do we get structured objects/attributes?

- Web Content
 - Metadata/Taxonomies/Folksonomies
 - Classification/ML/Extraction/Semantic Web
- Web 2.0 & Web Usage
 - Explicit & Implicit relations
- Building out an open ecosystem
 - Publishers have incentives to contribute

13

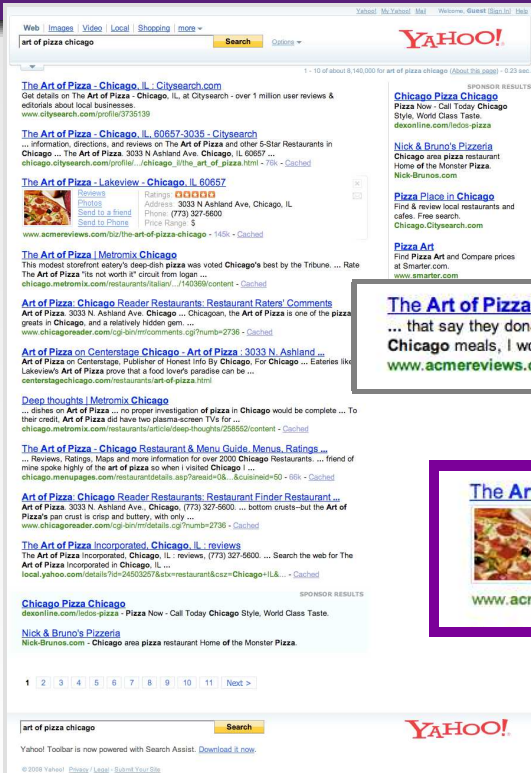
- 13 -

The SearchMonkey Ecosystem



- 14 -

Opening search - what does it mean?



- Clear win for: developers, site owners, users and Yahoo!
- Go from “to-do” to “done”

BEFORE

The Art of Pizza - Lakeview - Chicago, IL 60657
 ... that say they don't like Chicago deep dish pizza have never had Art of Pizza. ... Top 10 Chicago meals, I would include an Art of Pizza Deep Stuffed ...
www.acmerreviews.com/biz/the-art-of-pizza-chicago - 145k - Cached

AFTER



- 15 -

Content and Metadata 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)

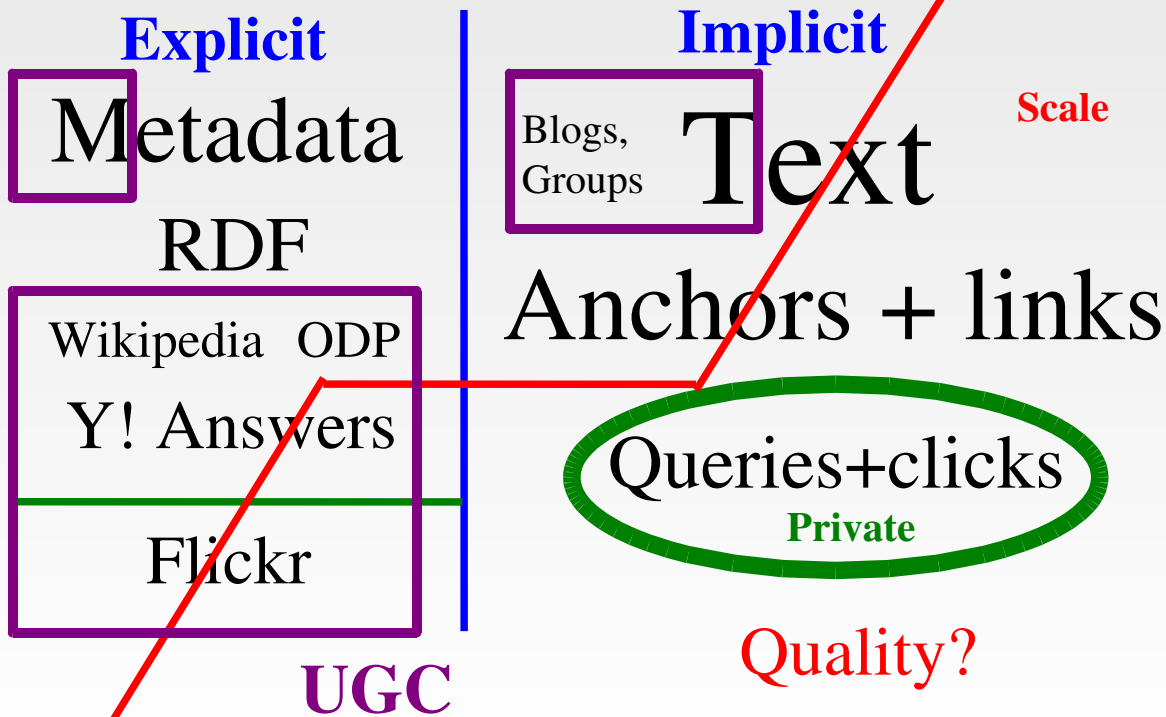
Metadata type	Amount of metadata produced per day
Anchortext	100 MB
Tags	40 MB
Pageviews	180 GB
Reviews	Around 10 MB

[Ramakrishnan and Tomkins 2007]

- 16 -

Examples

Wordnet



- 17 -

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

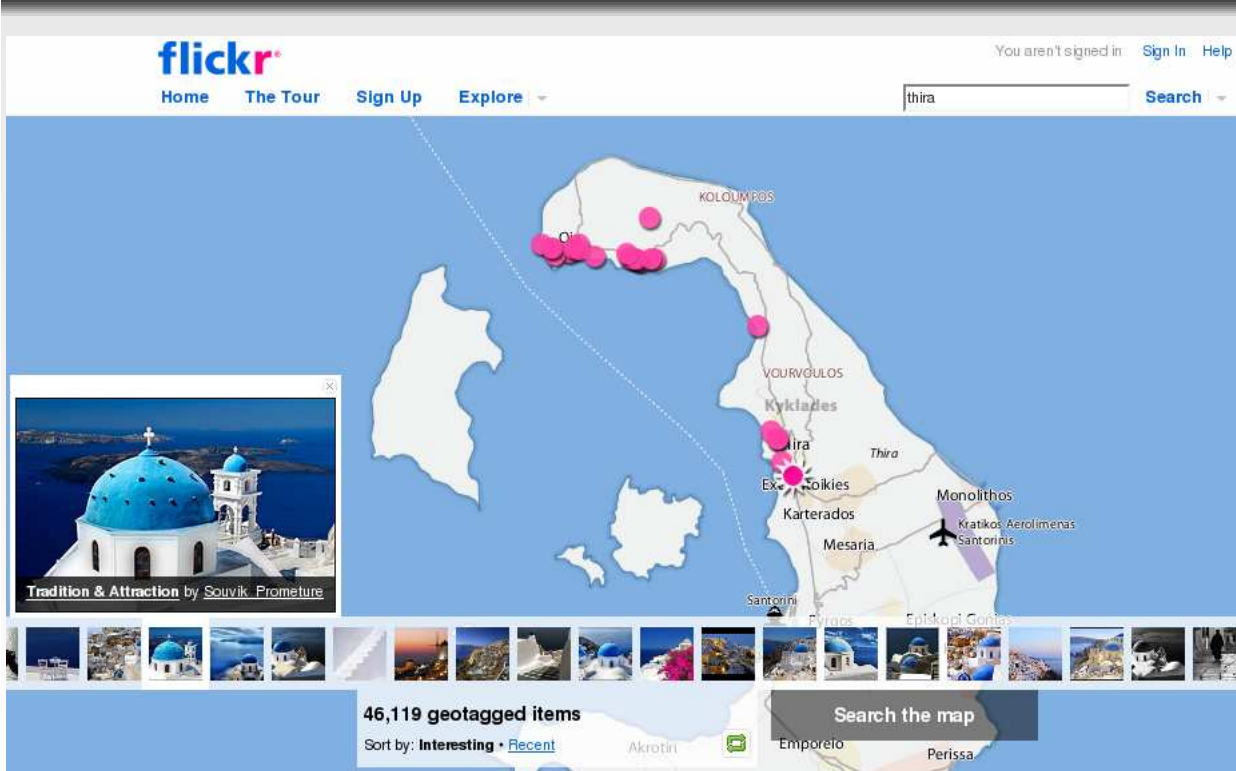
Aggregating data

“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”.

- 18 -

The screenshot shows the Flickr website interface. At the top, there are navigation links for 'Home', 'Sign Up', 'Sign In', 'Help', 'Ayuda', and 'Iniciar sesión'. Below the Flickr logo, there are links for 'Inicio', 'La vista', 'Crear cuenta', and 'Explorar'. A search bar is visible with the text 'Ir a: fira' and a 'Buscar' button. The main content area displays two clusters of photos. The first cluster is for the tag 'fira' and includes photos of Santorini, Greece, with a caption: 'santorini, greece, thira, island, sea, cyclades, caldera, sunset, oia, church' and a link 'Ver más en este clúster...'. The second cluster is for the tag 'catalunya' and includes photos of Barcelona, Spain, with a caption: 'catalunya, barcelona, spain, feria, catalonia' and a link 'Ver más en este clúster...'. Below these clusters, a message states: 'Éstas son las cargas más recientes etiquetadas con fira. Ver más...'. At the bottom, there are six individual photo thumbnails, each with a caption 'De jessica halida'.

The Wisdom of Crowds



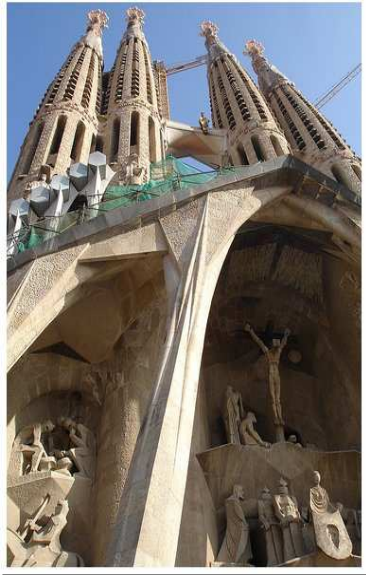
- 20 -

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!)

- 21 -

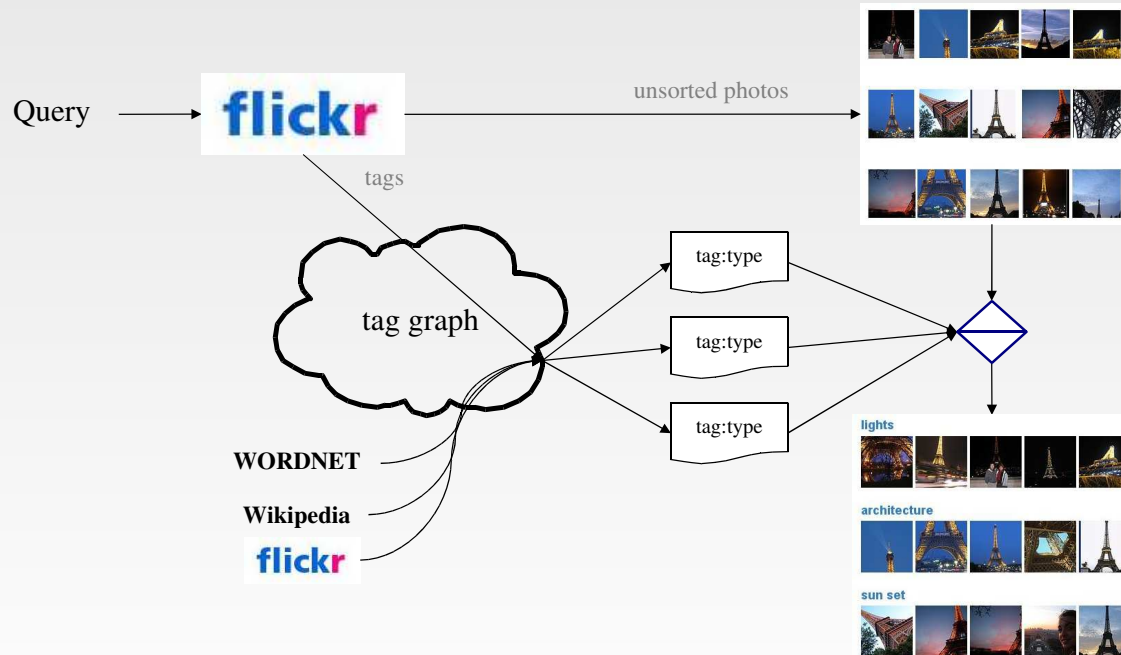
Tag Mining - Collective Knowledge



- 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

- 22 -

Improving Image Search



- 23 -

TagExplorer

- <http://sandbox.yahoo.com/TagExplorer>
- A prototype for browsing Flickr photos
- Provides query refinement for ...
 - ... drilling in to more **specific topics**
 - ... zooming out to more **general topics**
 - ... side-track to a **related topic**
- Organizes refinement terms ...
 - ... in a **tag-cloud**
 - ... groups together **semantically similar** terms

- 24 -

Dynamic Tag Clouds

- For the user query a list of related terms is presented and can be used to refine the query (visualized as a tag-cloud)
- The related terms are derived using tag co-occurrence among 250 million Flickr photos
- The related terms are calculated using a probabilistic framework using different conditional probabilities to get a mixture of general and specific terms

- 25 -

Semantic Breakup of Tag Clouds

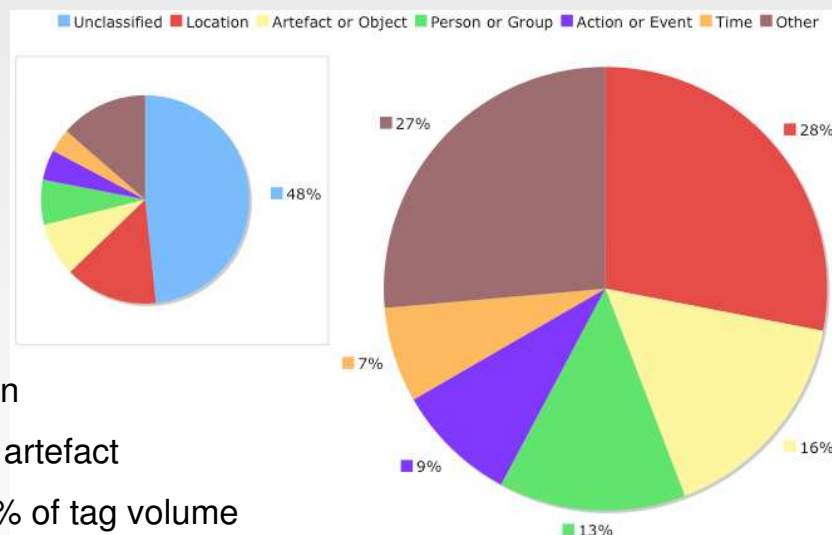
- Tag-cloud is organized by grouping together tags that have similar meaning
- The grouping is a two levels
 - Where? What? When?
 - Locations, subjects, names, activities, time
- The classification of tags is derived using a machine learned classification of Wikipedia pages

Overell, Sigurbjornsson and van Zwol, WSDM 2009

- 26 -

Tag Mining - Classification

- Assign tag semantics using WordNet broad categories



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

- 27 -

Tag Mining – Classification

- Extend this mapping using patterns found in Wikipedia
 - Upper bound 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

- 28 -

TagExplorer - Example

TagExplorer
Powered by [Flickr](#)

santorini

YAHOO! RESEARCH

Query: [santorini](#) ✕

locations
[akrotiri](#) [athens](#) [cyclades](#) [europe](#) [greece](#) [greek islands](#) [mykonos](#) [oia](#) [thera](#)

subjects
[aegean](#) [island](#) [islands](#) [sea](#) [sun set](#) [names](#) [greek](#)

activities
[travel](#) [time](#) [honeymoon](#)

Help
You can refine your query using the tag-cloud on the left.

- Use [tag](#) to post new query using tag
- Use [+](#) to add terms to query
- Use [✕](#) to remove terms from query

Photo Results

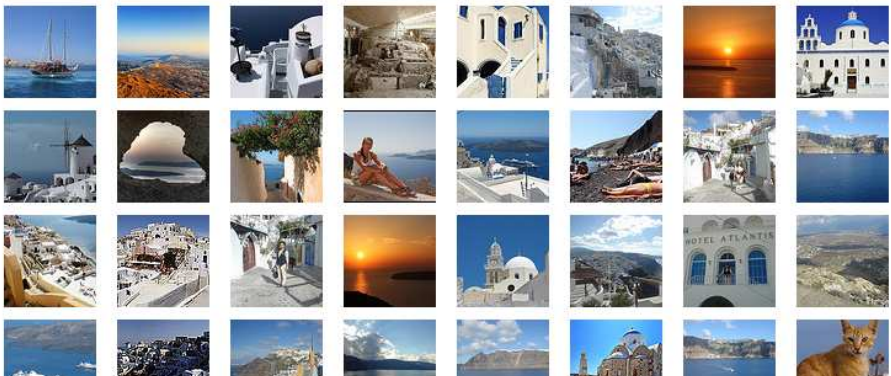



Photo Details



The Heart of Santorini
Taken by: [Klearchos Kapoutsis](#)
[View photo on Flickr](#)
Tags: heart santorini caldera καρδιά megalochori Ελλάδα σαντορινη κυκλάδες καλντέρα μεγαλοχωρι καλδέρρα

Could suggest tags: nice but

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

Update annotation

- 33 -

Faceted Image Search

Web Images Video Local Shopping more ▾

brussels

Search

Options ▾



Show only: Wallpaper size Black & White [More filters](#)

1 - 20 of 2,315,140 for brussels - 0.08 s. | 🔍

Explore [Bruxelles, Belgium](#) ▾

 **Statue of the Pissing Boy**
62 images

 **Central Square**
71 images

 **Royal Palace**
403 images

 **Place du Jeu de Balle**
456 images

 **Autoworld**
41,247 images

 **Europalia**
210 images

 **Belgian Royal Museums of Fine Arts**



brussels.jpg
700 x 497 | 175k
[trekearth.com](#)



brussels grandplace.jpg
800 x 600 | 117k
[trekearth.com](#)



brussels grand place
[Jason's Trave... flickr.com](#)



Brussels.jpg
1536 x 1024 | 424k
[isavetravel.com](#)



couple of days in brussels
[Rui M Leal flickr.com](#)



xmas in brussels ii.jpg
785 x 572 | 199k
[trekearth.com](#)



great market brussels belgium
[gbatistini flickr.com](#)



man piss in brussels belgium
[mambol935 flickr.com](#)



Done

Diversification of image search results

Topical and Visual diversification

ACM MIR 2008, Vancouver, Canada

WWW2009, Madrid, Spain

- 35 -

Dimensions of Diversity

- **Topical diversity**

Query: "Jaguar"



- **Visual diversity**

Query: "Jaguar X-type"



- Other dimensions: spatial, temporal, social

- 36 -

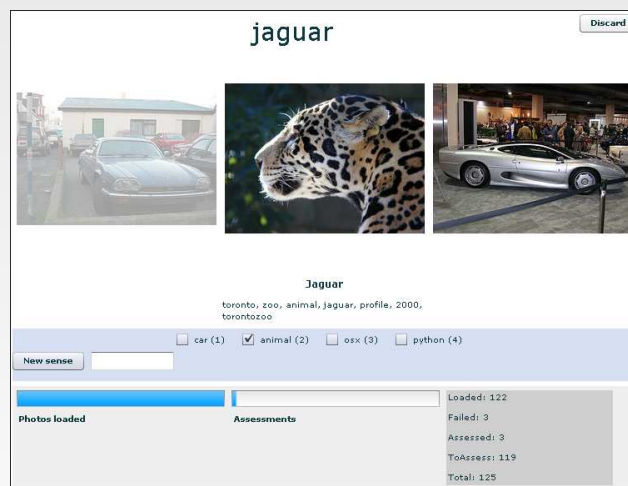
Topical Diversity

- **Propose: diversification as part of the retrieval model**
 - Query Likelihood (full index, tags only)
 - Relevance model (full index, tags only, dual index)
- Topics
 - 95 topics extracted from Flickr search logs
 - 25 ambiguous topics
- Collection:
 - 6M public photos from Flickr (Title, description and tags)

- 37 -

Topical Diversity

- Blind pooling, 51.000 images judged for relevance.
- Two step assessment:
 - Binary relevance judgement
 - Sense classification
- Measured inter-assessor agreement for 20% of topics
 - >85% for all topics
 - most topics >90%



- 38 -

Topical Diversity – Retrieval Performance

- Unambiguous topics

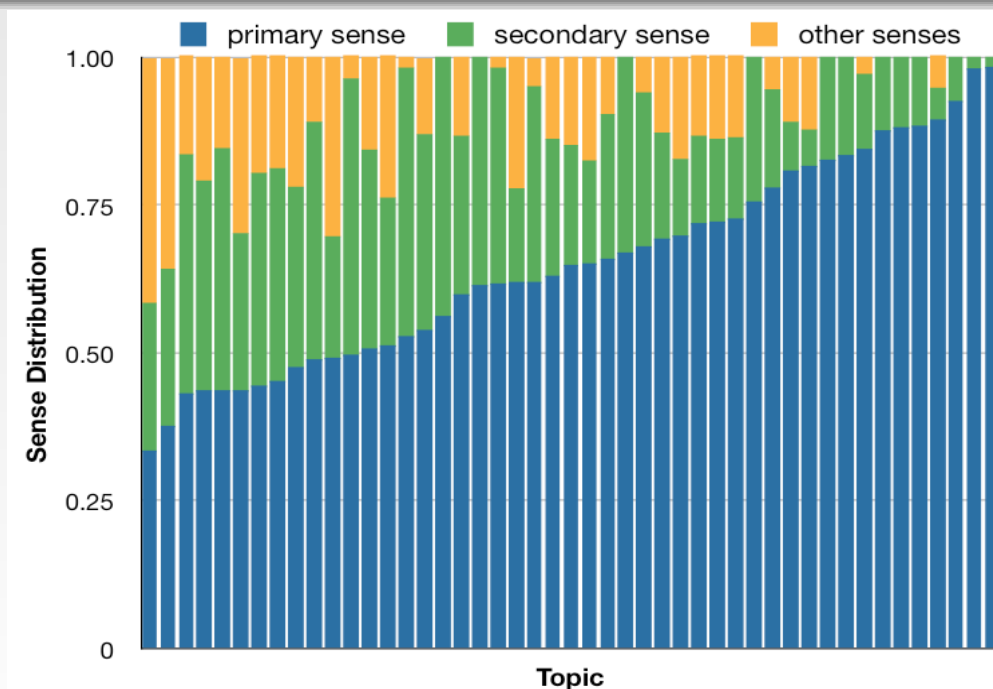
Model	P@1	P@5	P@10	P@15	P@20	P@25	P@50
Query Likelihood	0.747	0.733	0.733	0.719	0.709	0.701	0.667
Query Likelihood (Tags Only)	0.779	0.749	0.720	0.712	0.703	0.700	0.673
Relevance Model	0.758	0.743	0.720	0.708	0.706	0.699	0.677
Relevance Model (Tags Only)	0.779	0.726	0.717	0.719	0.714	0.710	0.683
Relevance Model (Dual Index)	0.768	0.754	0.739	0.726	0.719	0.716	0.680

- Ambiguous topics

Model	P@1	P@5	P@10	P@15	P@20	P@25	P@50
Query Likelihood	0.680	0.760	0.720	0.725	0.734	0.744	0.734
Query Likelihood (Tags Only)	0.800	0.736	0.732	0.720	0.736	0.736	0.734
Relevance Model	0.720	0.760	0.768	0.784	0.788	0.792	0.778
Relevance Model (Tags Only)	0.840	0.728	0.744	0.741	0.756	0.752	0.735
Relevance Model (Dual Index)	0.720	0.776	0.768	0.755	0.754	0.760	0.763

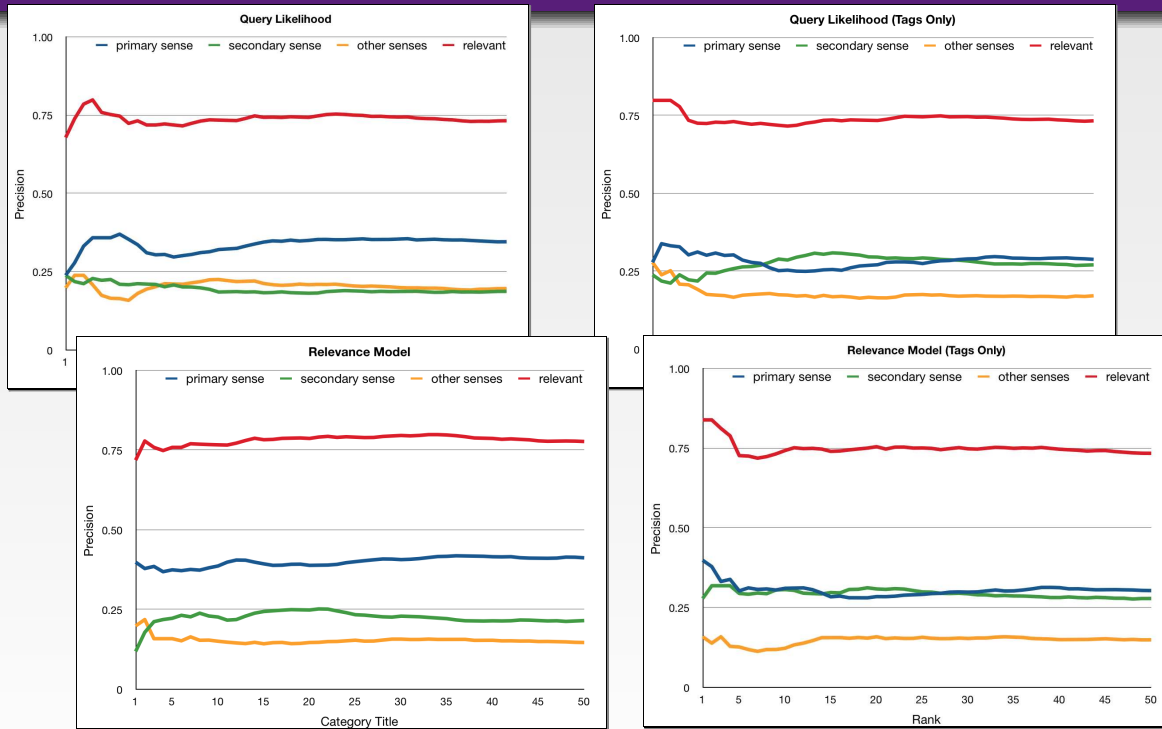
- 39 -

Topical Diversity – Distribution of word senses



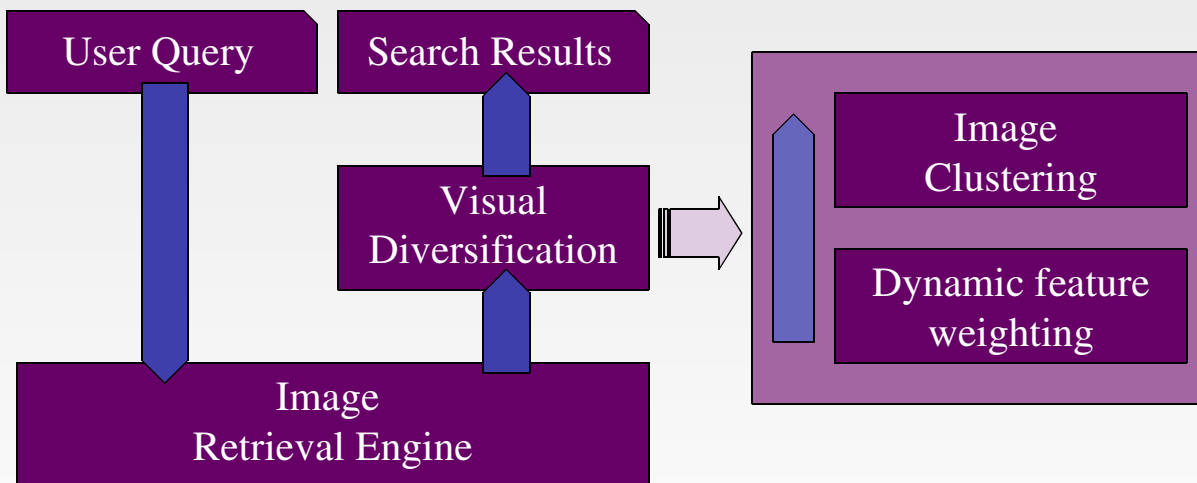
- 40 -

Topical Diversity – Distribution of word senses



- 41 -

Visual Diversity



- 42 -

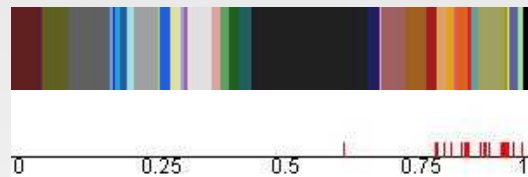
Visual Diversity

- Visual diversity is needed when topical ambiguity is low
- Depends of visual characteristics of images returned
 - Dynamic feature weighting
 - Capture visual characteristics of the list of images
 - Light-weight clustering methods
 - (Top-K) Folding, Reciprocal Election

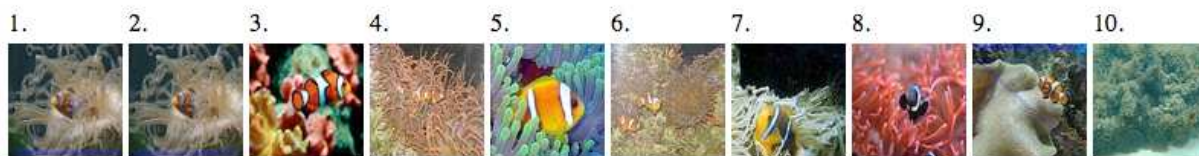
-43-

Visual Diversity – Dynamic feature weighting

- Query: “clownfish”
 - Avg. color histogram:
 - Distance to average:
 - Color variance: 0.46



Textual ranking (using dual index relevance model)



-44-

Visual Diversity – Dynamic feature weighting

- Distance between a and b :

$$d(a, b) = \frac{1}{f} \sum_{i=0}^f \frac{1}{\sigma_i} d_i(a, b)$$

Distance between a and b according to i^{th} feature

Total number of features

Variance of distances according to i^{th} feature

- 45 -

Visual Diversity – clustering of results

Textual ranking (using dual index relevance model)



Folding selection (features: color histogram, edge corellogram)



Top-k folding selection (features: color histogram, edge corellogram)



- 46 -

Boosting Image Retrieval through Aggregating Search Results based on Visual Annotations

ACM Multimedia 2008, Vancouver, Canada

- 47 -

Use Visual Annotations

Flickr allows another kind of annotations (notes)

- Associate **text** with **visual area**
- Highly relevant to content
→ **Visual Annotation**
- Valuable to learn different the visual representations of an object



Olivares, Ciaramita, van Zwol. ACM Multimedia 2008

- 48 -

Motivation

Main objective:

Improve retrieval performance by **combining visual annotations**, with **textual** and **visual information**

Use visual annotations (text & image) and rank aggregation to improve retrieval

- 49 -

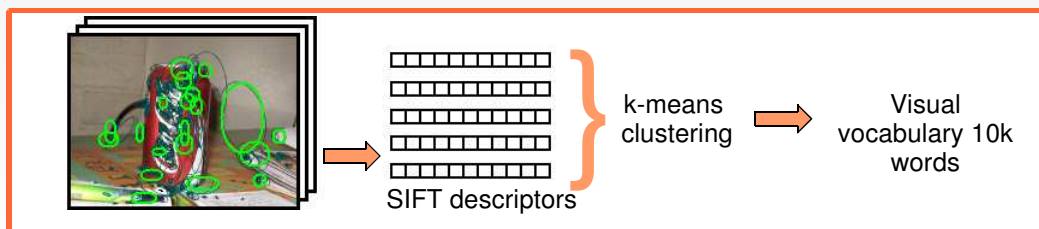
Content-based Image Retrieval

1. Extract visual features and describe them

- Processed 12,000 images.
- Computed Harris and Hessian features
- Described using SIFT

- Build visual vocabulary

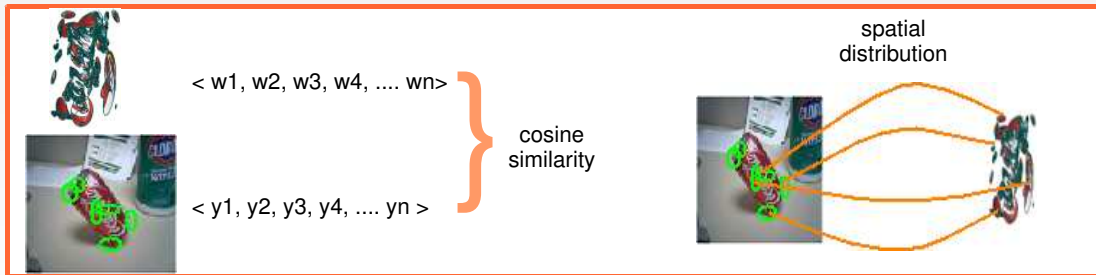
- Clustered SIFT descriptors to create vocabulary of 10,000 words
- Implemented an approximate K-means algorithm
- 3 resulting vocabularies: based on Harris, Hessian and a combination of those 2 features.



- 51 -

Content-based Image Retrieval

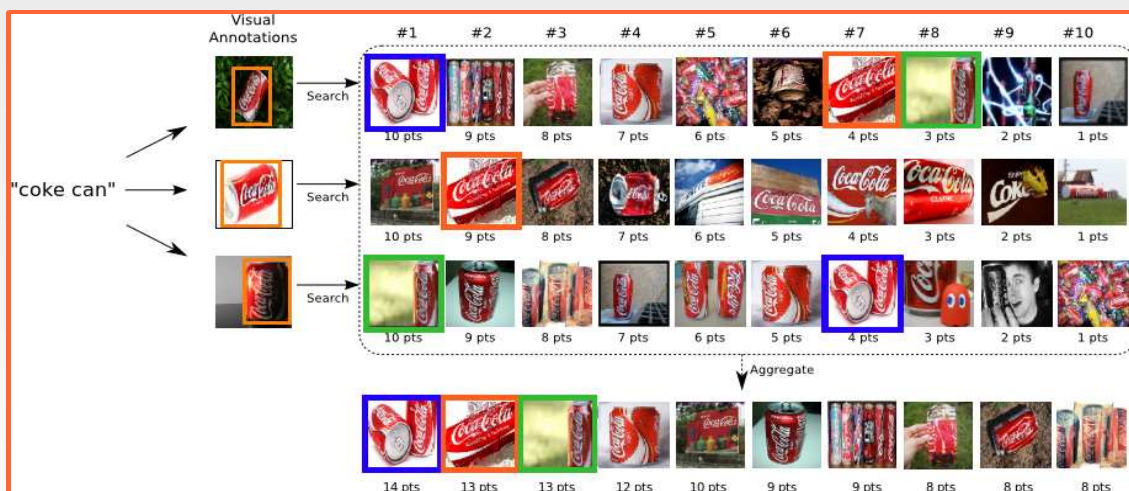
- Image represented as a set of visual words
 - Represented in the vector space model
 - Similarity measured by cosine similarity
- Spatial distribution more important in images than in text



- 52 -

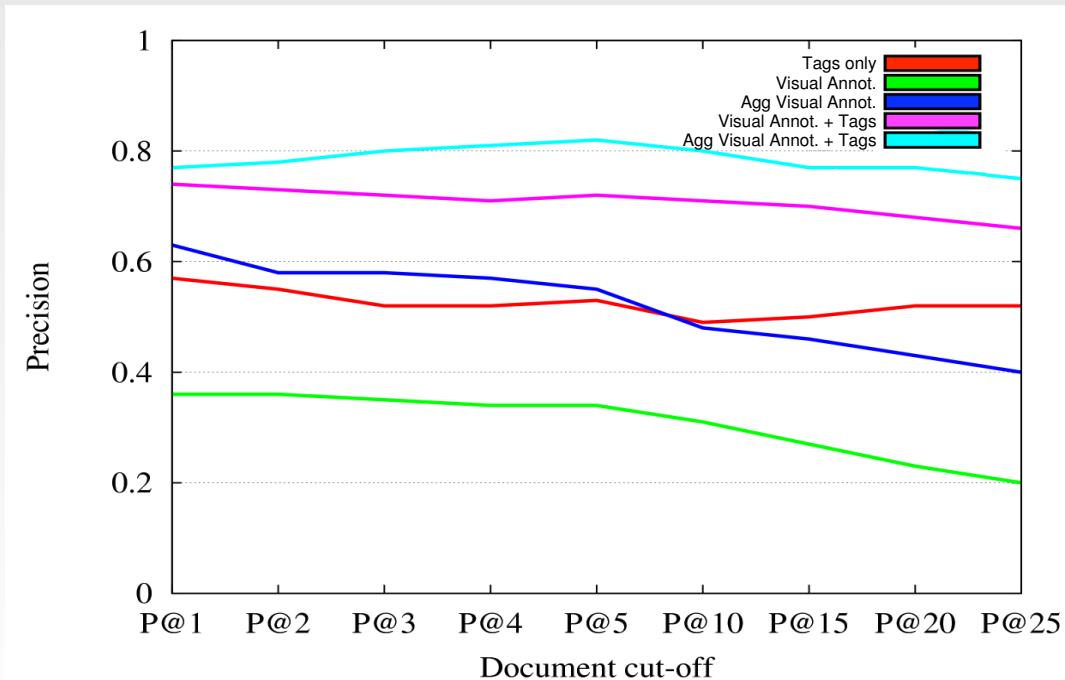
Aggregating visual annotations

Borda count [Aslam & Montague]



- 53 -

Results: Systems comparison



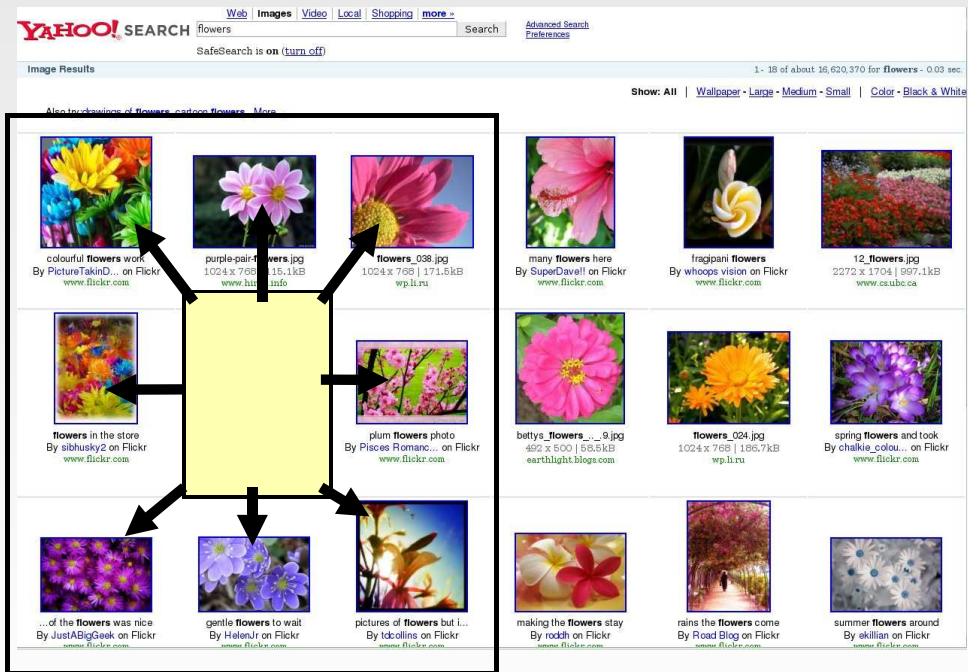
- 54 -

Ranking Images on the Web with Clicks, Tags and Visual Features

Bridging the semantic gap in image retrieval with user-clicks on images...

- 55 -

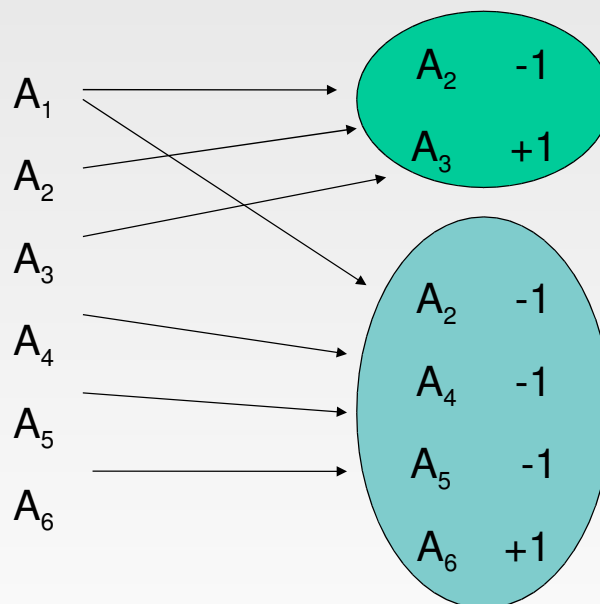
Ranking Images with Click Data



- 56 -

Learning from Clicks

- User clicks give relative preference
- Clicks at rank 1 ignored
- Train and evaluate in “blocks”



- 57 -

Classification

- Two classes: clicked and nonclicked
 - Assume they are separable by a hyperplane
- Train on patterns independently
- Binary perceptron
 - Averaging: Average weight vector of all models posited during training
 - Uneven margin:
 - Clicked class outnumbered by nonclicked class
- Perceptron produces a score
 - Use the score to rank images in each block

- 58 -

Features: Visual + Textual

- Visual features described earlier
- Tf.idf term weights (query, image) pairs
 - Tags
 - Title
 - Description
 - All as one “document”
- Cosine similarity
- Maximum tf.idf score
- Average tf.idf score
- Bias feature is 1.0 for every example
- Scores normalized by column and by row

- 59 -

Results

	Accuracy	MRR
Retrieval Baseline	0.4198	0.6186
Learned Baseline	0.4073	0.6104
Textual Features	0.5484	0.7034*
Visual Features	0.5805	0.7233*
Text + Visual	0.7512	0.8365*

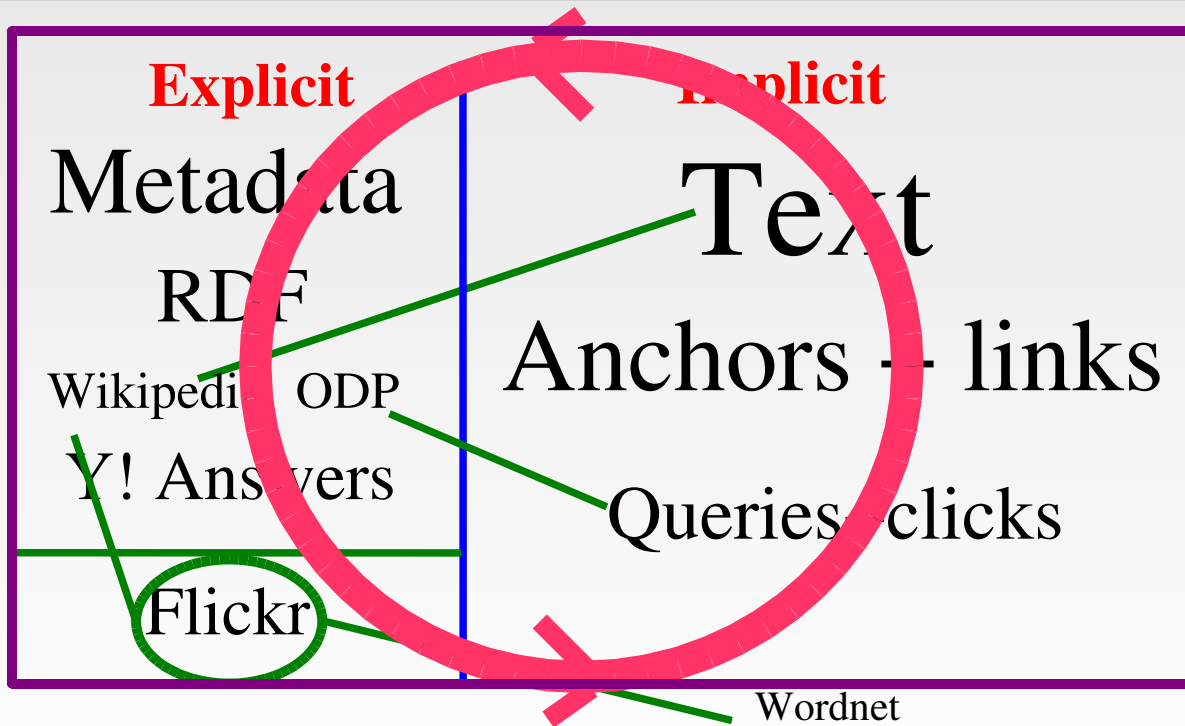
- 60 -

Open Issues

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

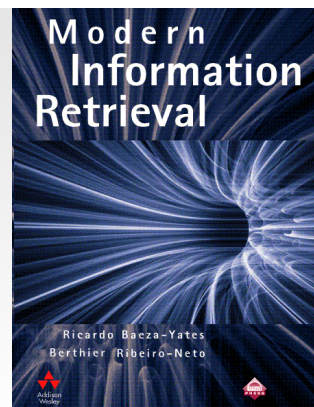
- 64 -

The Virtuous Cycle



- 65 -

**Second edition
coming soon**



Questions?

Contact: rbaeza@acm.org

Thanks to Borkur Sigurbjornsson, Vanessa Murdock,
Lluís Garcia, and Roelof van Zwol