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

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
  - Exploting the Wisdom of Crowds behind the Web 2.0

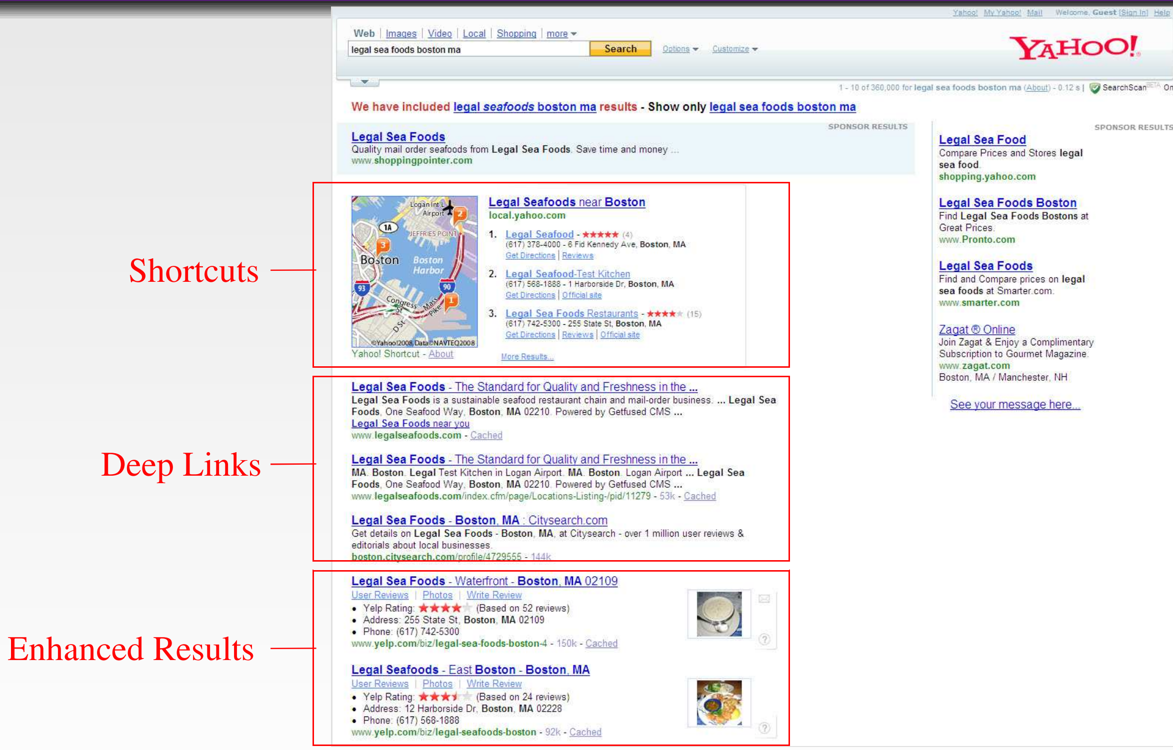
Trends

- 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

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

More complete information

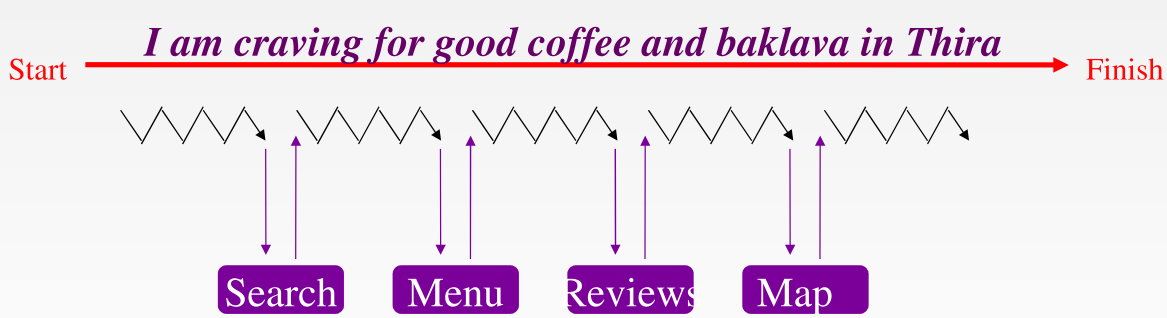


Search: Content vs. Intent

Premise:

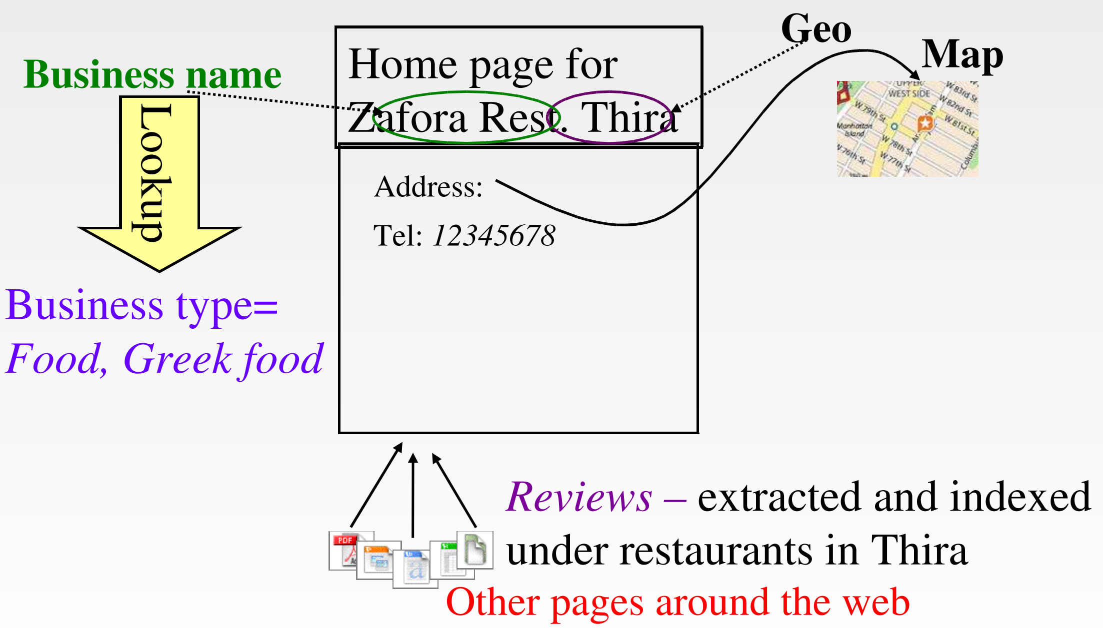
– People don’t want to search

– People want to get tasks done and get straight to their answers



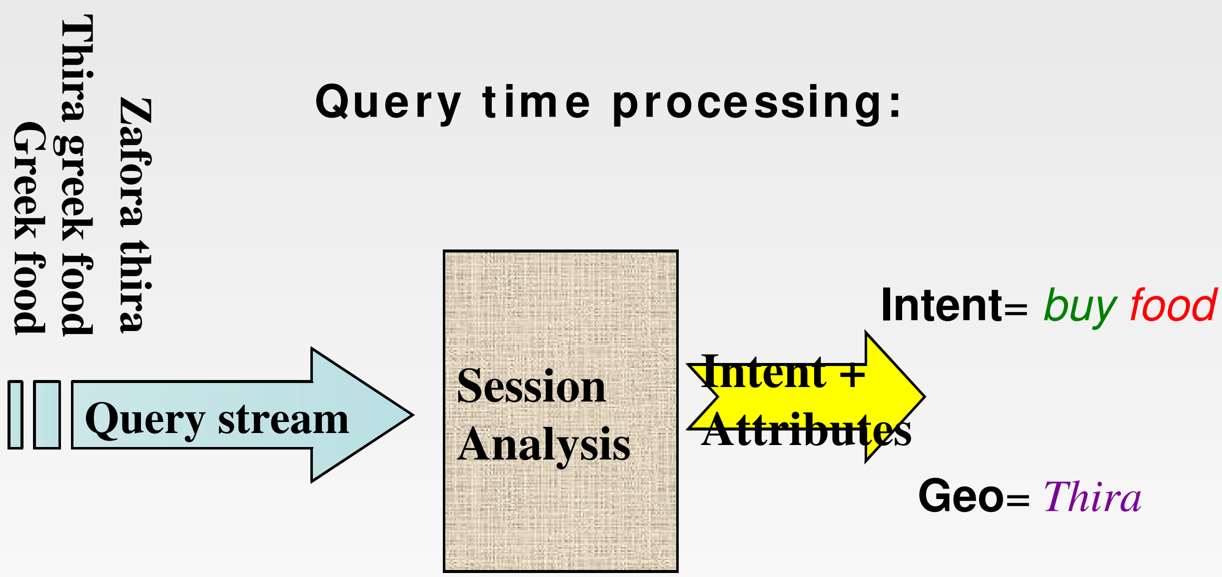
How this might work – I

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

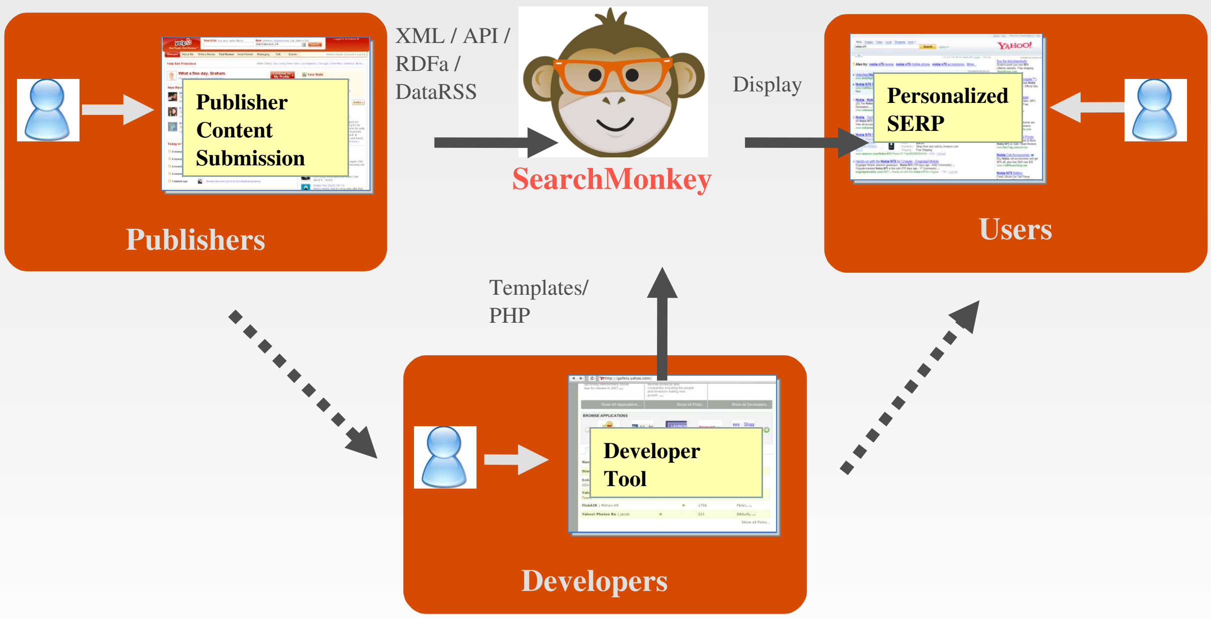
12

How do we get structured objects/attributes?

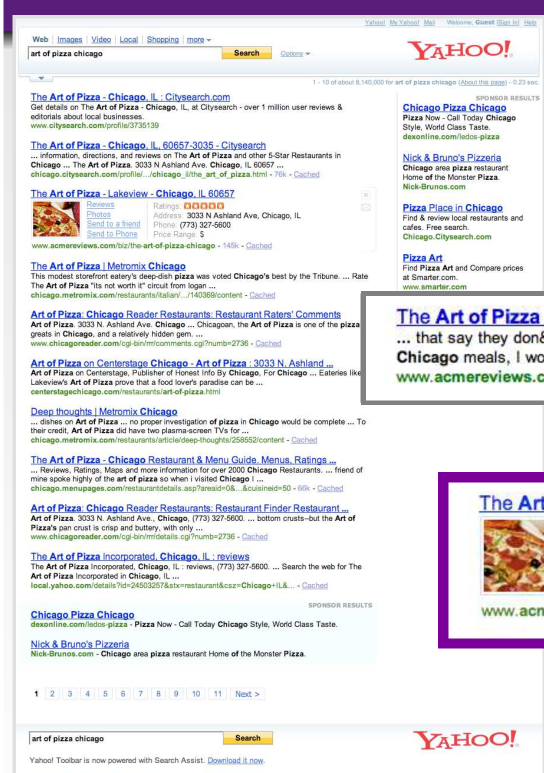
- Web Content
  - Metadata/Taxonomies/Folksonomies
  - Classification/ML/Extraction/Semantic Web
- Web 2.0 & Web Usage
  - Explicit & Implicit relations
- Building out an open ecosystem
  - Publishers have incentives to contribute

13

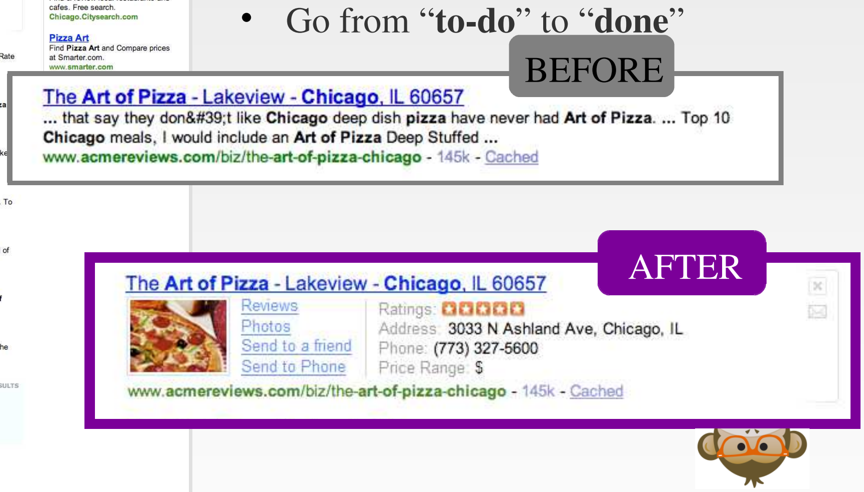
The SearchMonkey Ecosystem



Opening search - what does it mean?



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



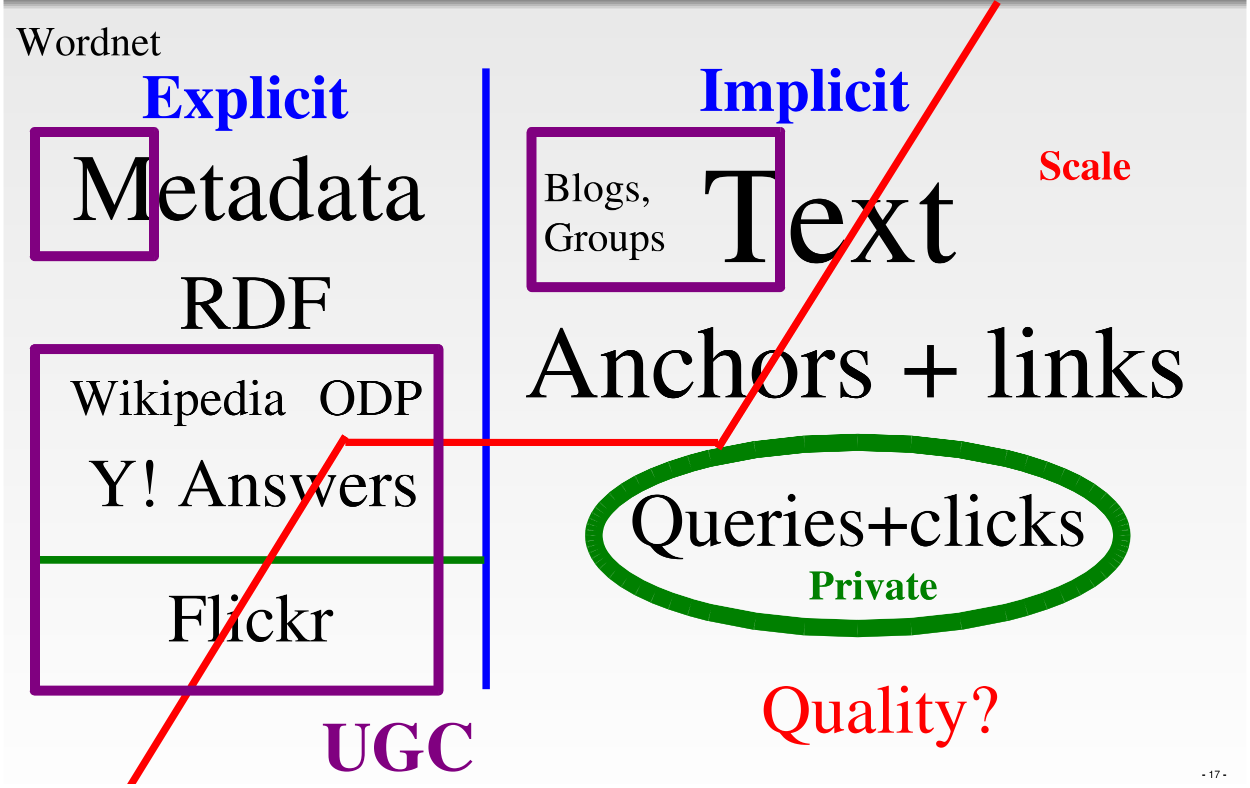
Content and Metadata trends

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

Examples

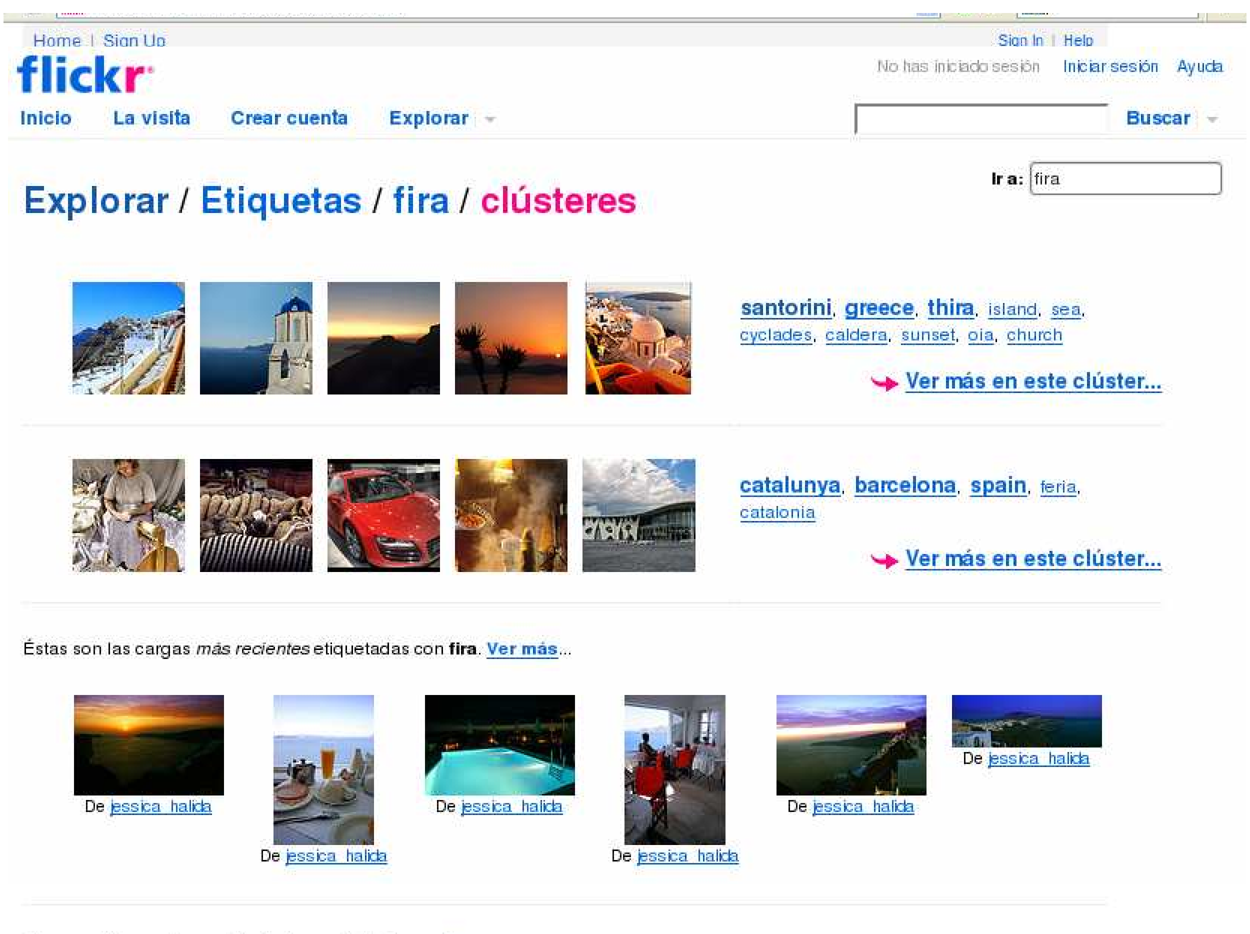


The Wisdom of Crowds

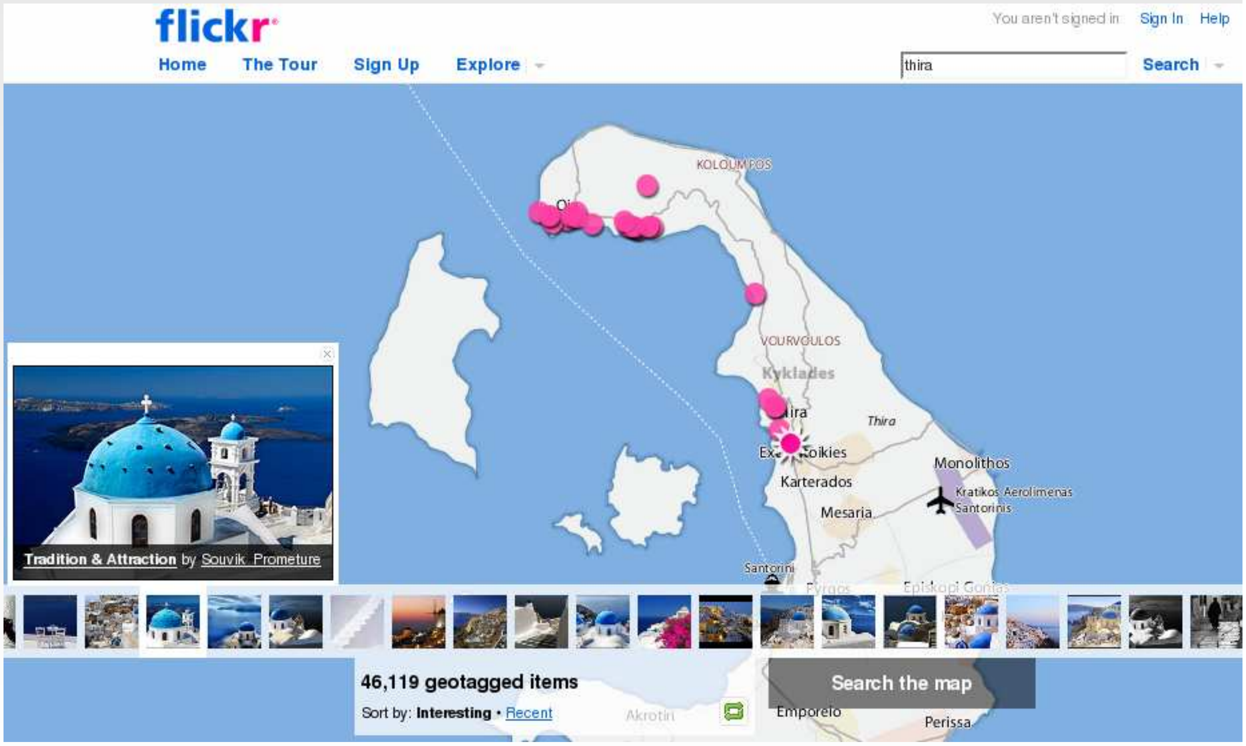
- James Surowiecki, a ***New Yorker*** columnist, published this book in 2004
  - **“Under the right circumstances, groups are remarkably intelligent”**
- Importance of diversity, independence and decentralization

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”.



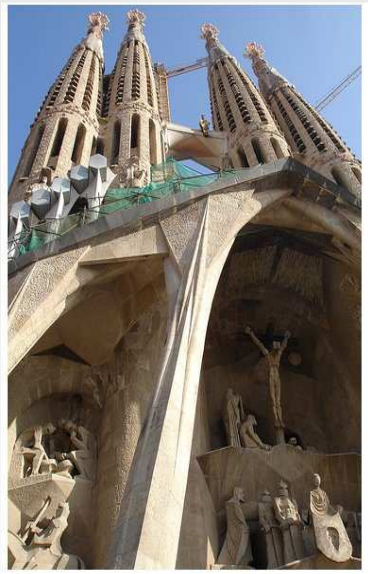
The Wisdom of Crowds



The Wisdom of Crowds

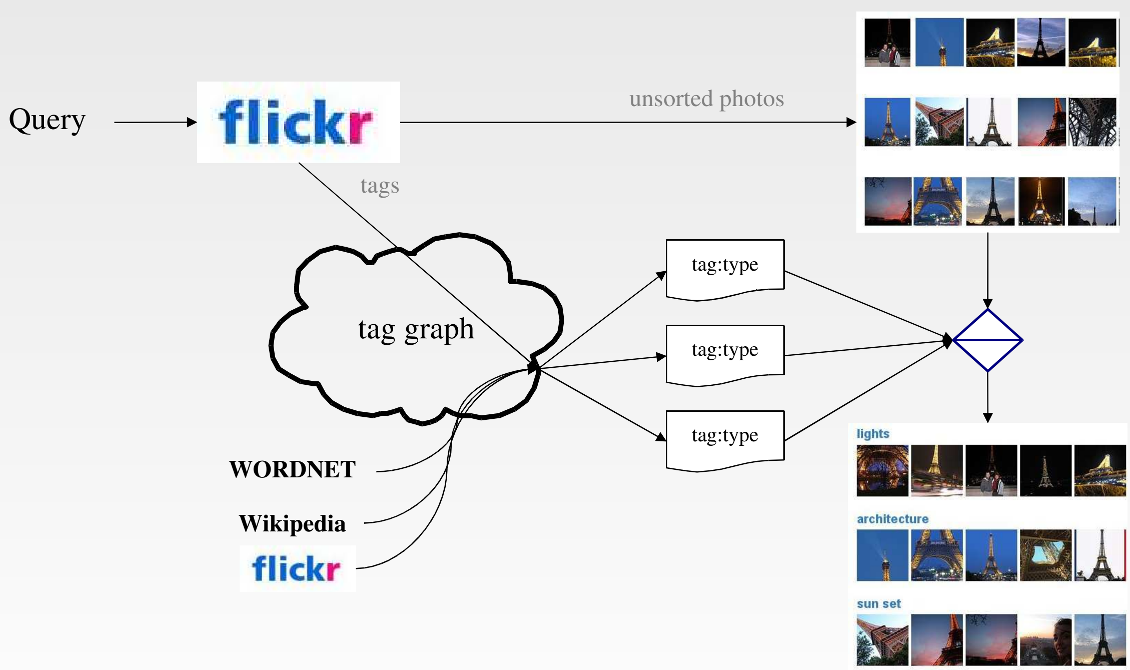
- –Crucial for Search Ranking
- –Text: Web Writers & Editors
  - not only for the Web!
- –Links: Web Publishers
- –Tags: Web Taggers
- –Queries: All Web Users!
  - Queries and actions (or no action!)

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

Improving Image Search



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

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

Sigurbjornsson and van Zwol, WWW 2008

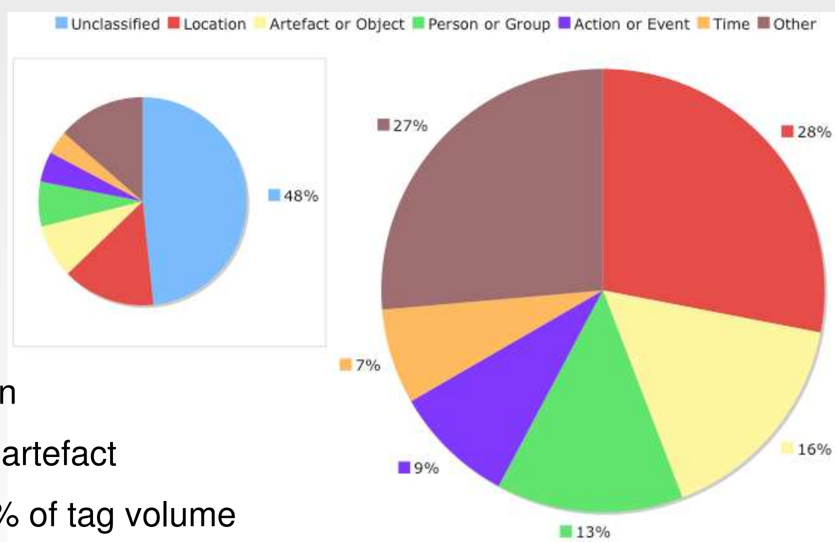
Semantic Breakup of Tag Clouds

- 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

Tag Mining - Classification

- Assign tag semantics using WordNet broad categories

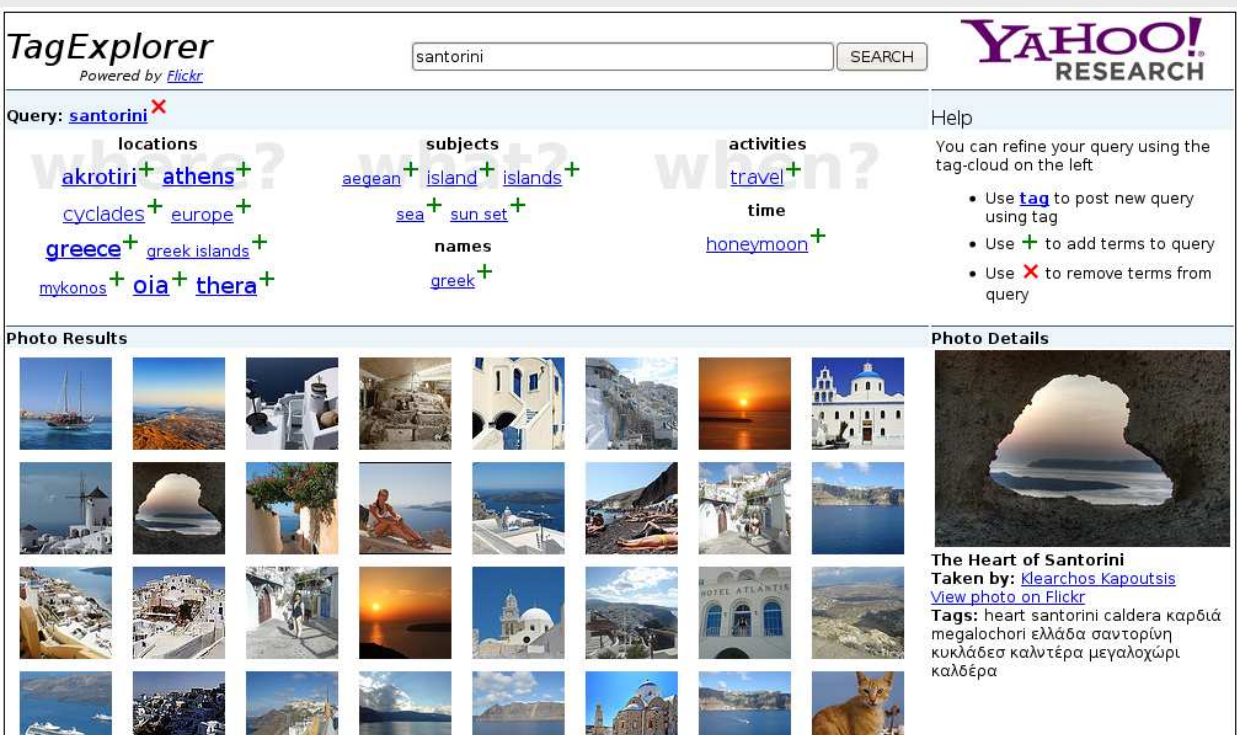


  - Paris :: location
  - Eiffel Tower :: artefact
  - Coverage: 52% of tag volume

Tag Mining – Classification

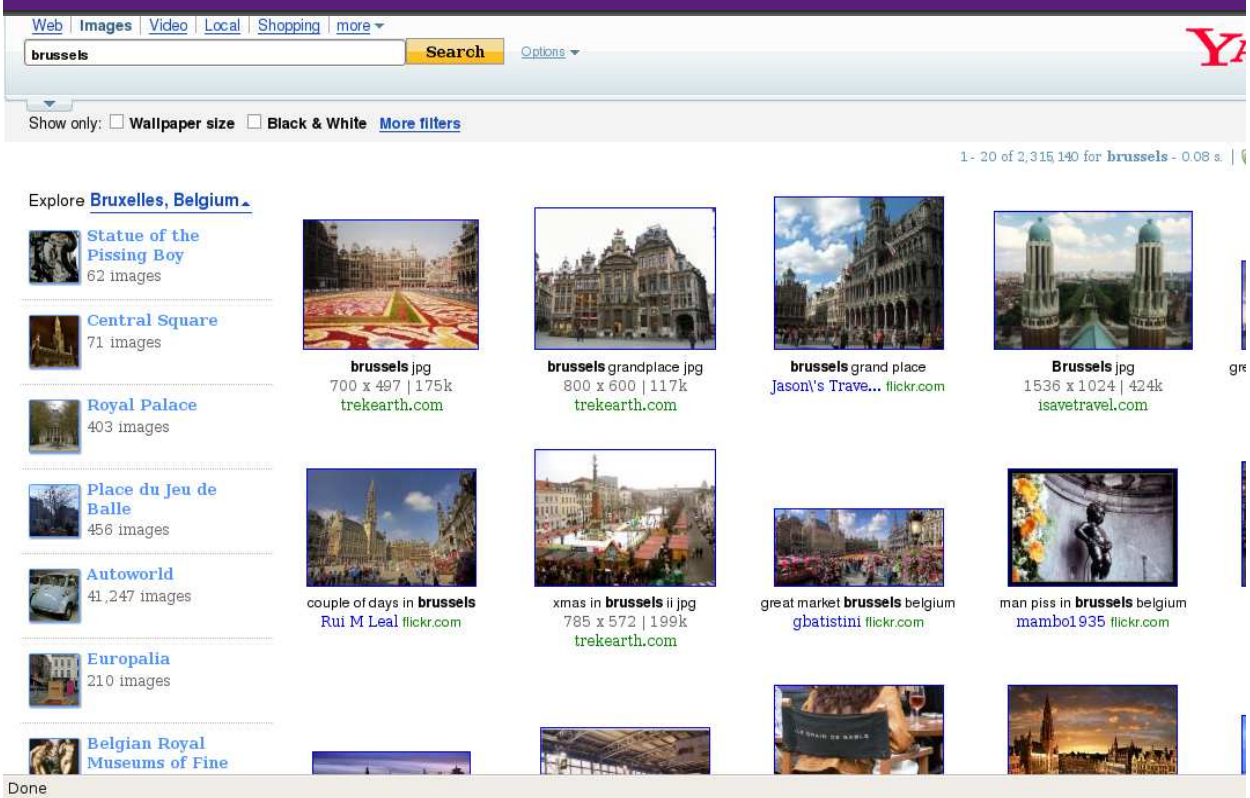
- Extend this mapping using patterns found in Wikipedia
  - Upper bound for coverage: 78.6% of the tag volume
  - Based on SVM approach
    - Features: Wikipedia templates and categories
    - Training data: Wikipedia entries found in WordNet
  - Extended coverage: 68% of the tag volume
  - Mapping from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

TagExplorer - Example



Could suggest tags: nice but ....

Faceted Image Search



Diversification of image search results

Topical and Visual diversification

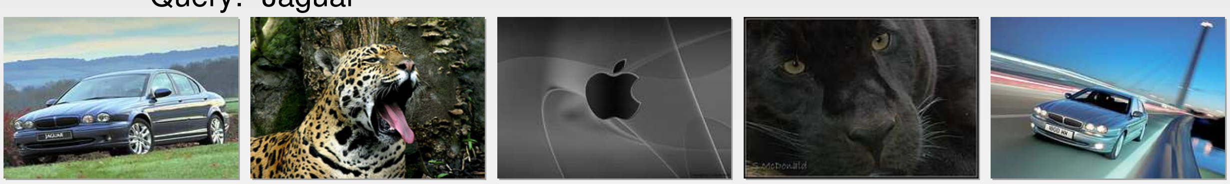
ACM MIR 2008, Vancouver, Canada

WWW2009, Madrid, Spain

Dimensions of Diversity

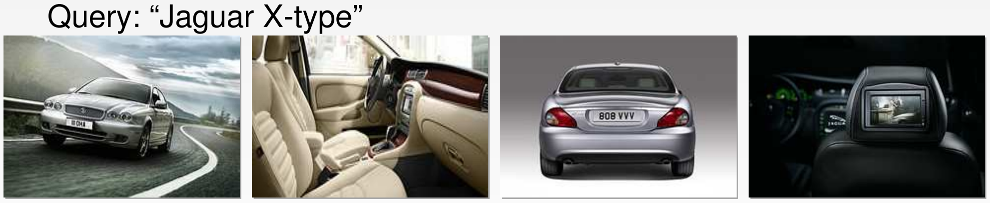
- **Topical diversity**

  Query: “Jaguar”



- **Visual diversity**

  Query: “Jaguar X-type”



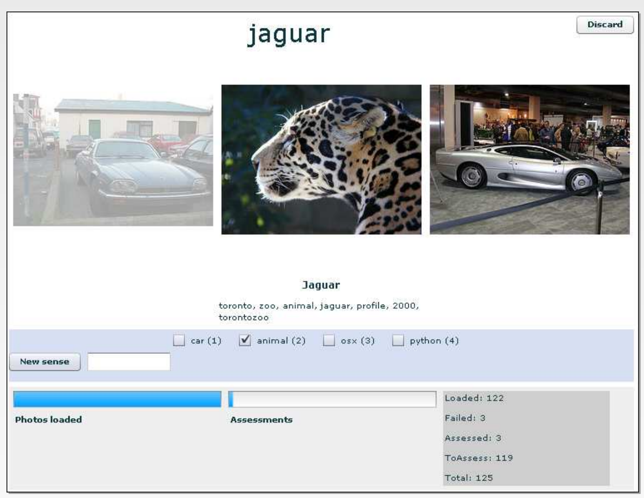
- Other dimensions: spatial, temporal, social

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)

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%



Topical Diversity – Retrieval Performance

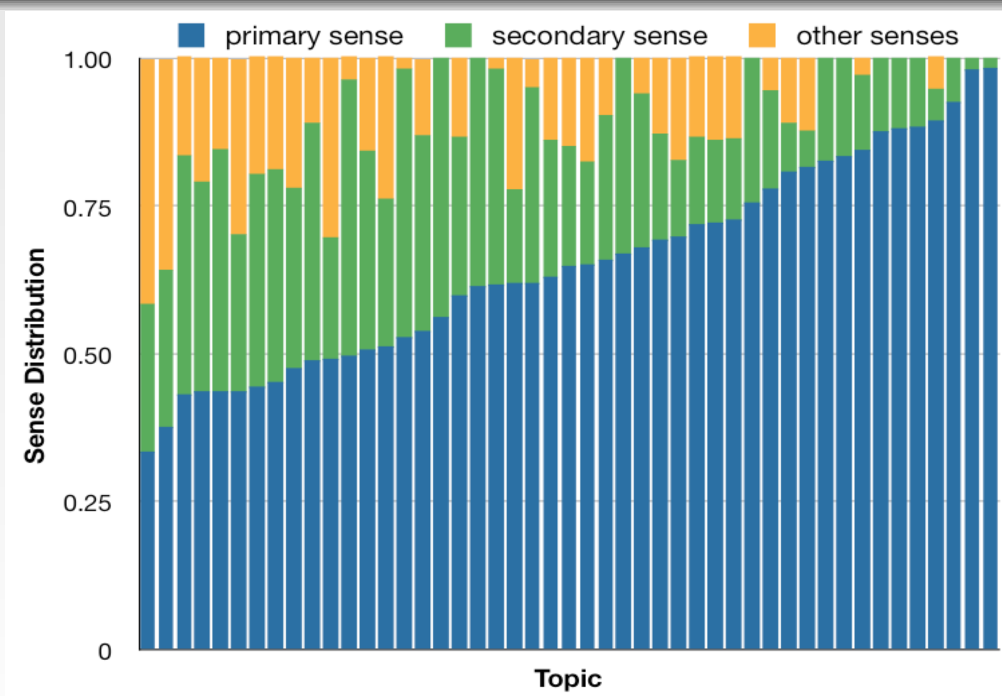
- Unambiguous topics

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

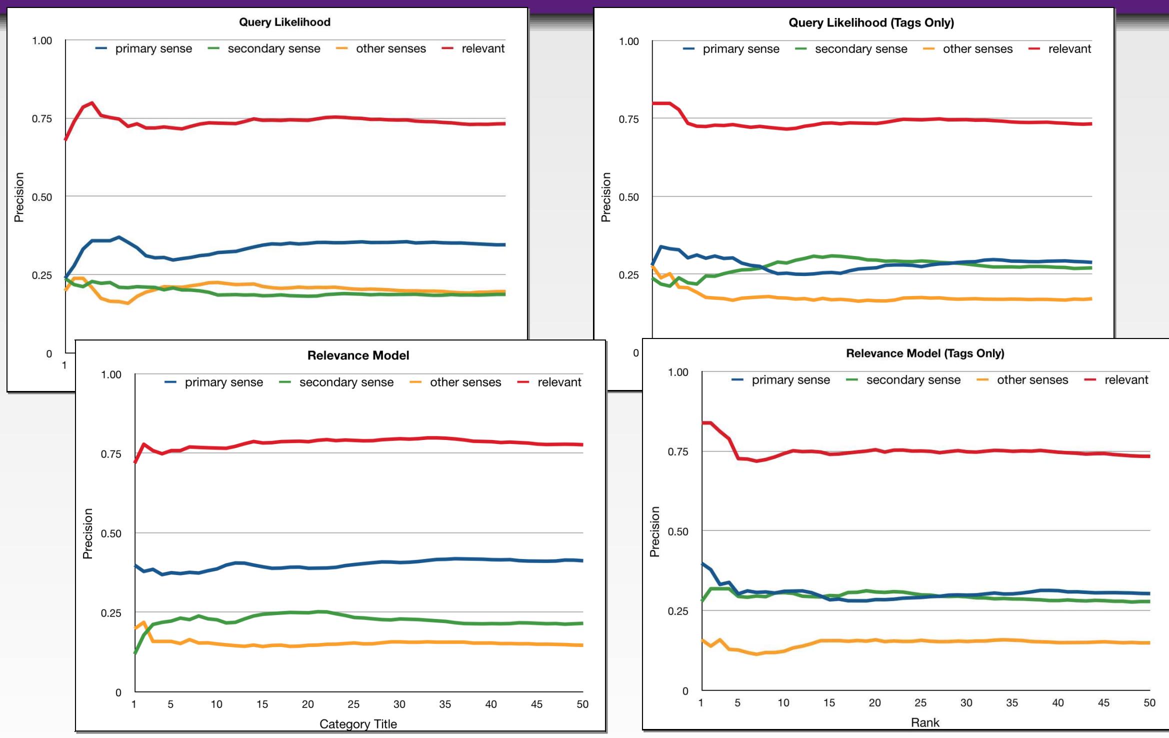
- Ambiguous topics

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

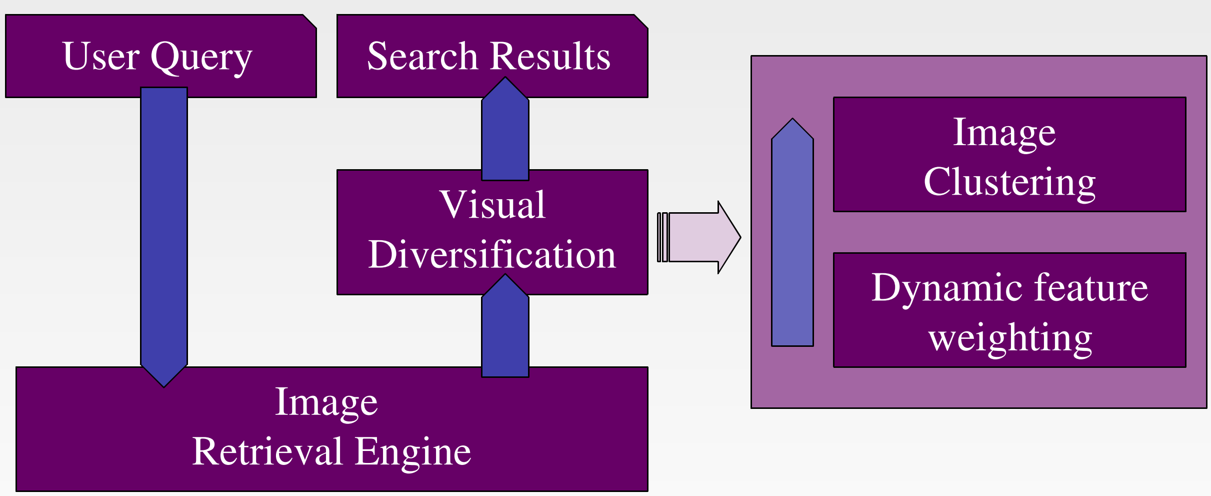
Topical Diversity – Distribution of word senses



Topical Diversity – Distribution of word senses



Visual Diversity

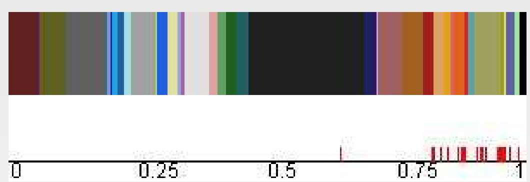


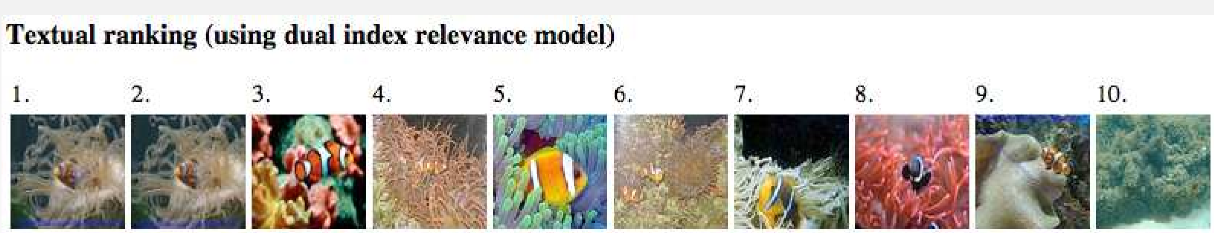
Visual Diversity

- 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

Visual Diversity – Dynamic feature weighting

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





Visual Diversity – Dynamic feature weighting

- Distance between *a* and *b*:

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

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

Total number of features

Variance of distances according to i^{th} feature

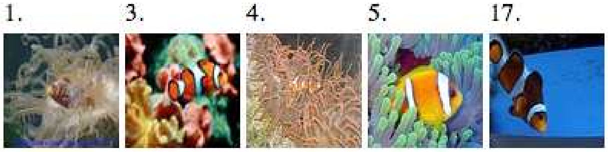
Visual Diversity – clustering of results

Textual ranking (using dual index relevance model)

1. 2. 3. 4. 5. 6. 7. 8. 9. 10.

Folding selection (features: color histogram, edge corellogram)

1. 3. 4. 5. 17.



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

1. (15) 3. (12) 23. (6) 18. (5) 5. (4)

Boosting Image Retrieval through Aggregating Search Results based on Visual Annotations

ACM Multimedia 2008, Vancouver, Canada

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

Motivation

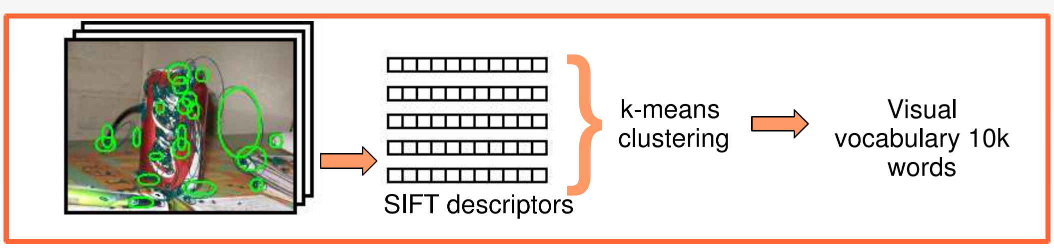
Main objective:

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

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

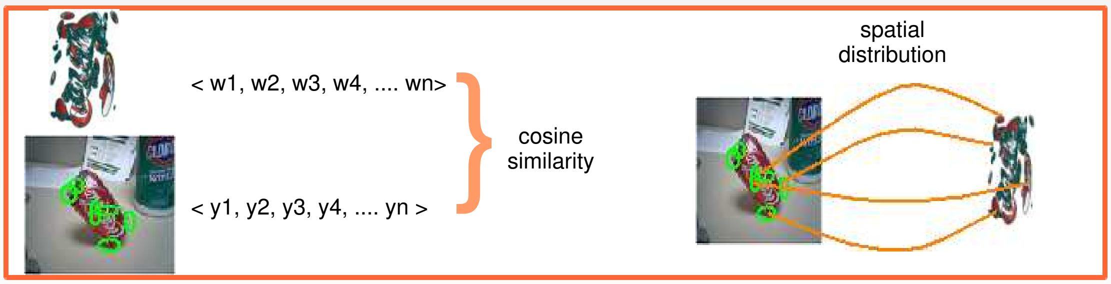
Content-based Image Retrieval

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.



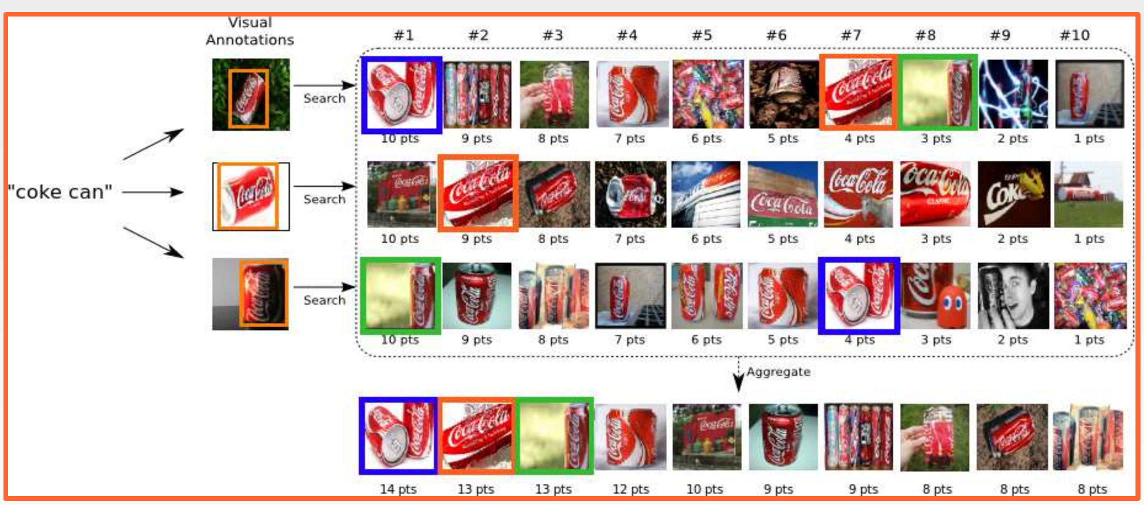
Content-based Image Retrieval

- 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

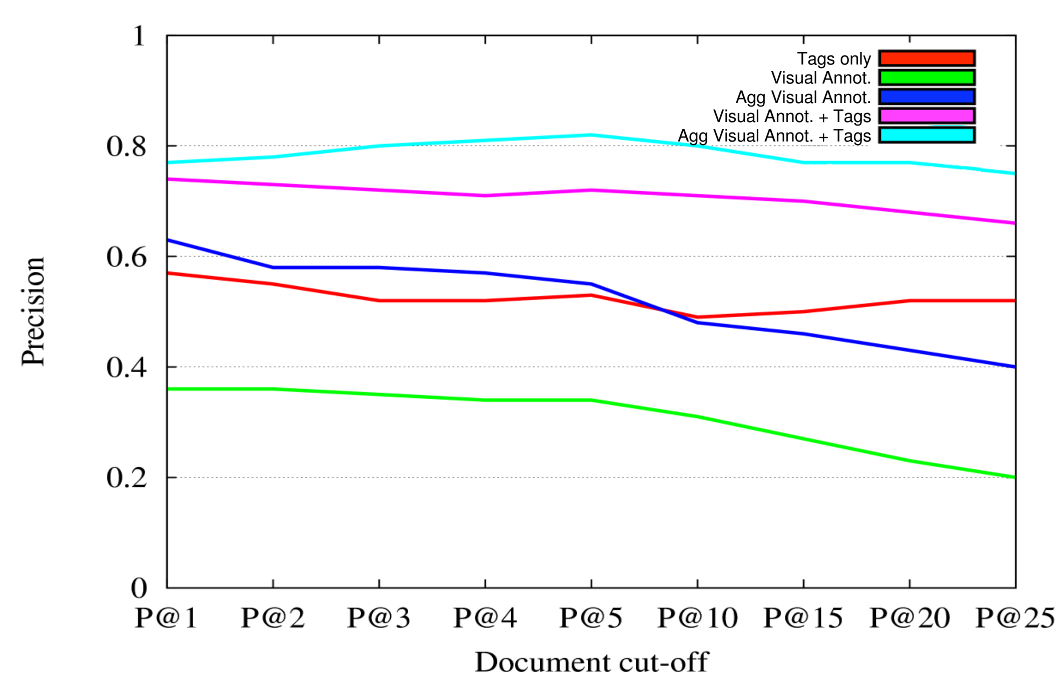


Aggregating visual annotations

Borda count [Aslam & Montague]



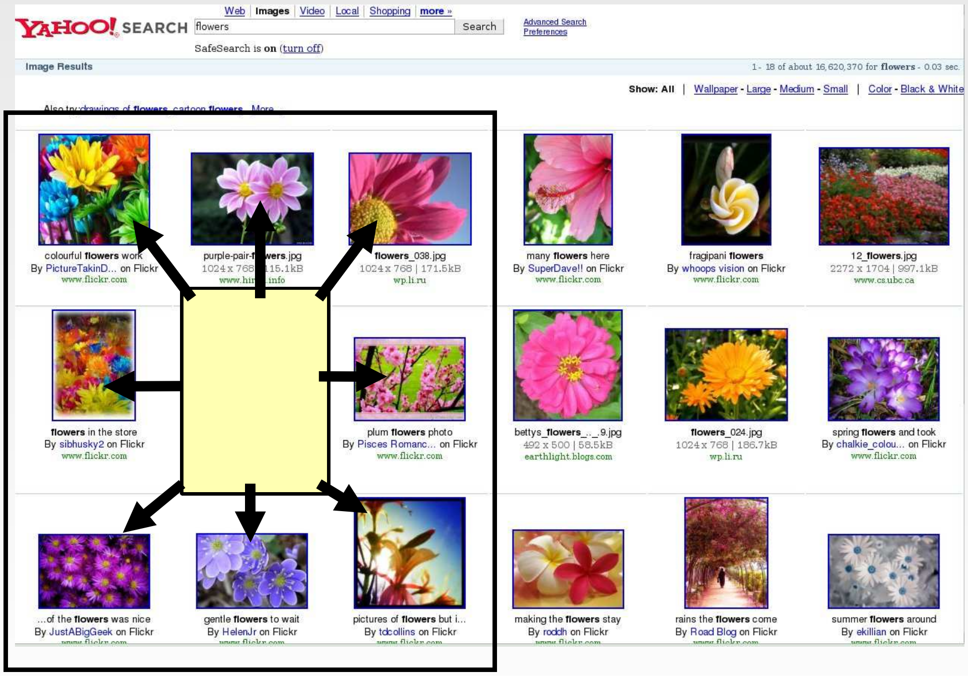
Results: Systems comparison



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

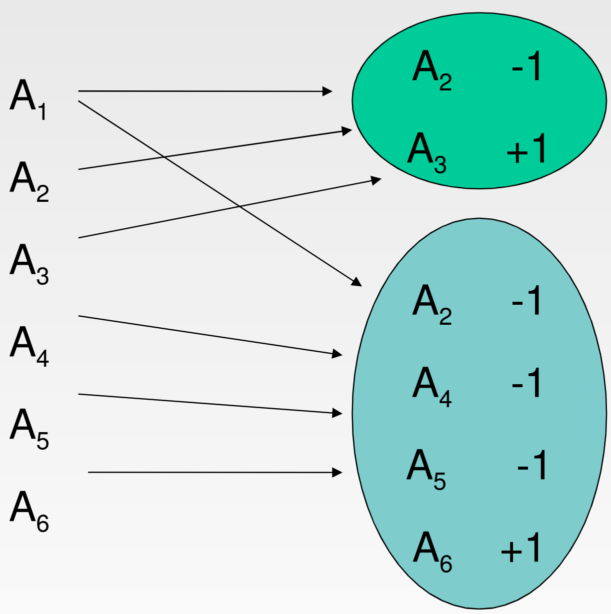
Bridging the semantic gap in image retrieval with user-clicks on images…

Ranking Images with Click Data



Learning from Clicks

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



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

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

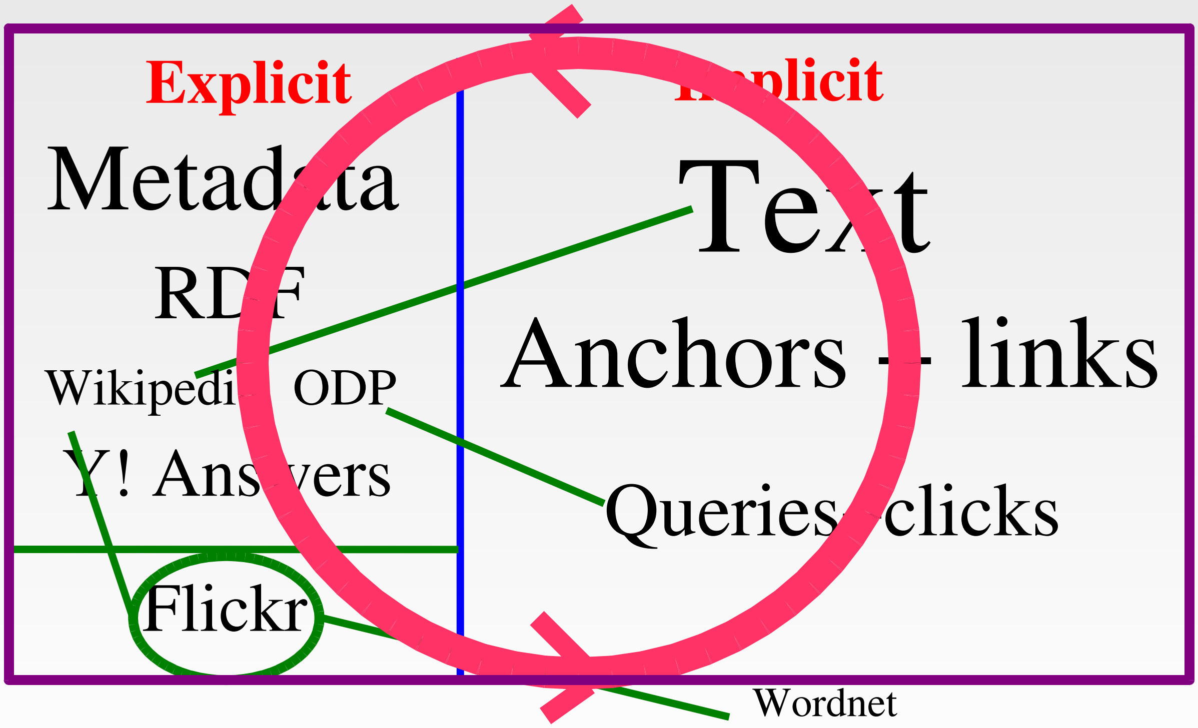
Results

	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***

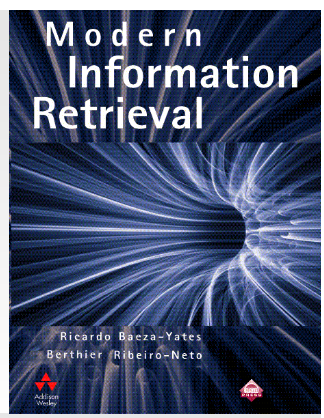
Open Issues

- 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

The Virtuous Cycle



Second edition coming soon



Questions?

Contact: rbaeza@acm.org

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