Multimedia Content Analysis and Search: New Perspectives and Approaches

HongJiang Zhang

CTO, Microsoft China R&D

Oct 20, 2009

How Has Web Search Changed CBIR after 20 Years

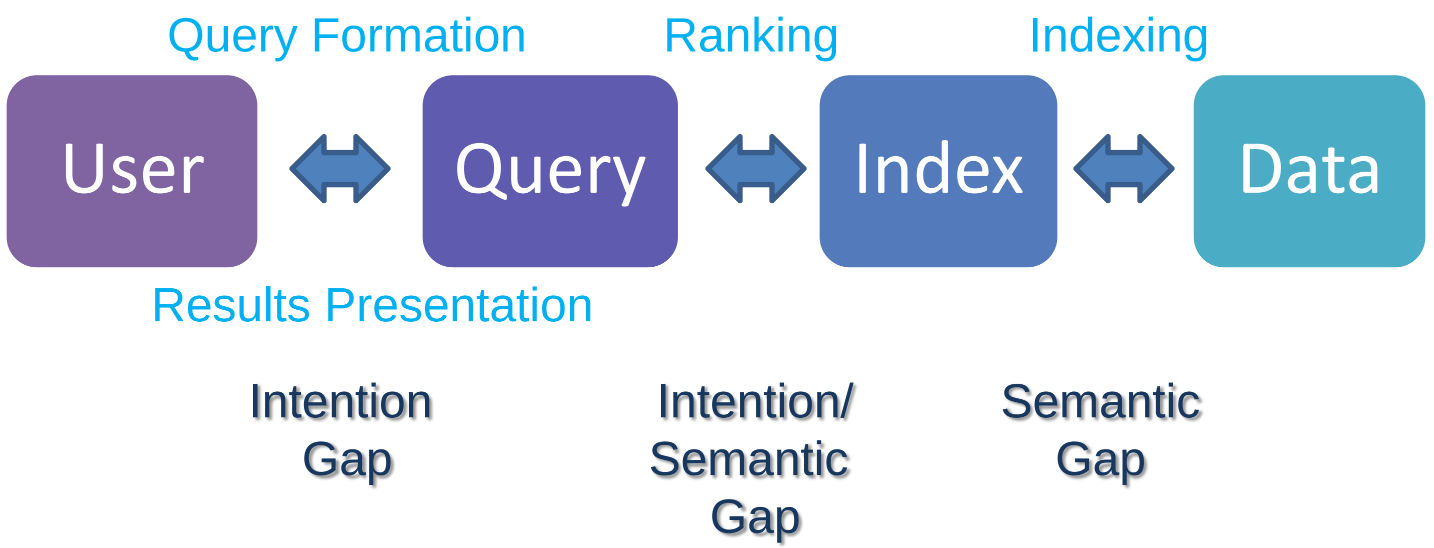
Outlines

- CBIR: a glance back
  - Issues and recent work
- Web Image Search
  - Label collection
  - What are useful and feasible categories ?
  - How to build automatic concept classifiers and annotation models ?
    - Data driven, model less
    - Inter play between text data and visual features
  - How to capture user intention?
    - UI and query formation
    - Search result organization
- Outlook

Acknowledgments

- Many colleagues at Microsoft
  - Xian-Sheng Hua, Lei Zhang, Wei-Ying Ma, Yong Rui
- My Ph.D students
  - Changhu Wang, Dong Liu, Guo-Jun Qi, ...
- Shih-Fu Chang

A Typical Image Search System

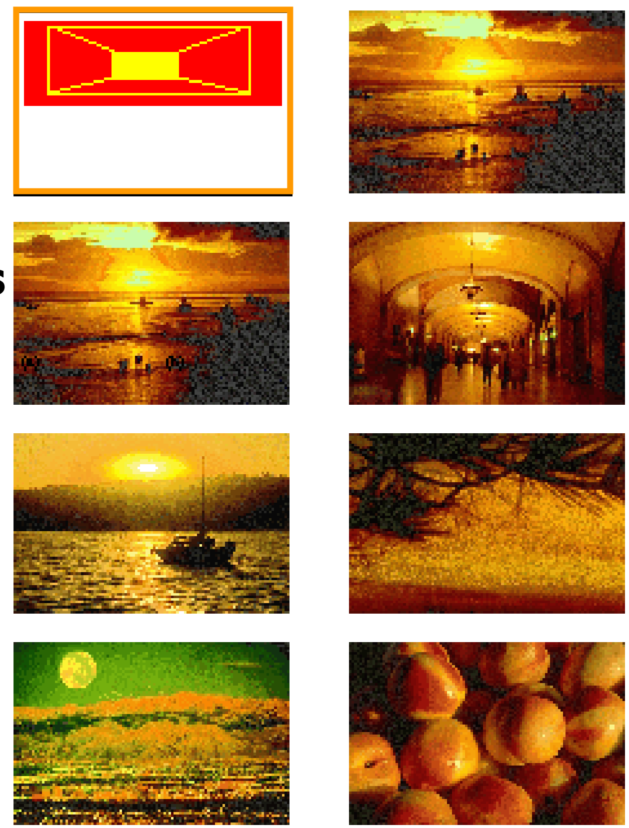


Bridging Semantic Gaps

CBIR

Query

results

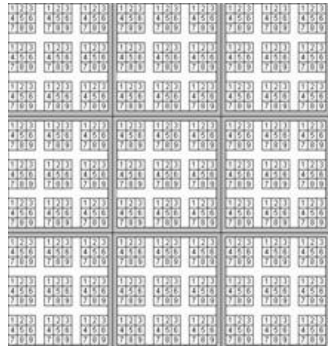


IBM QBIC (Flickner *et a*l ’95)

Issues in Image Search

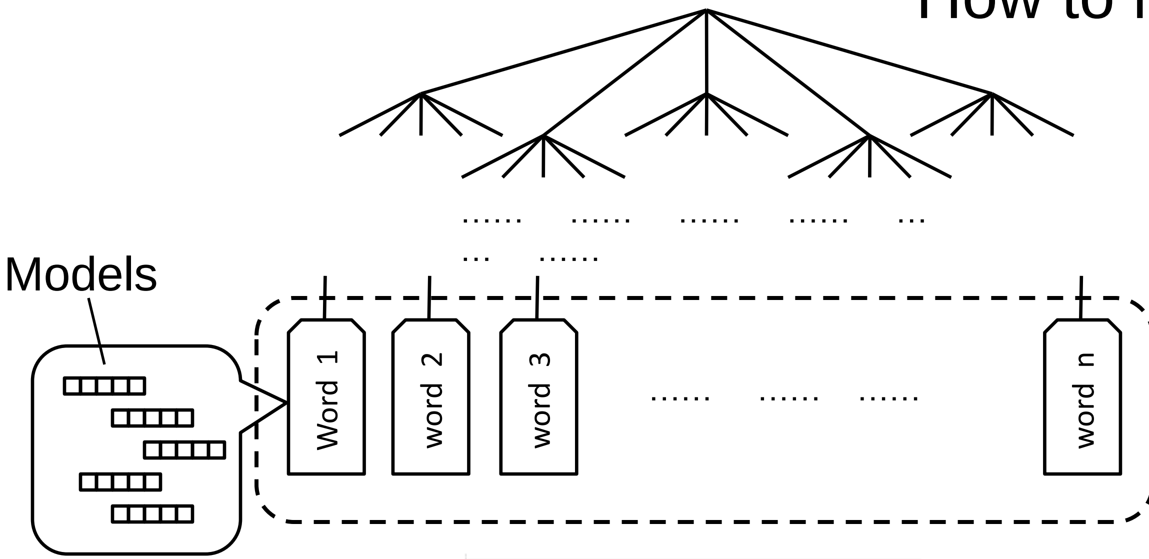
- Label collection
  - Scale and relevance
- What are useful and feasible categories ?
  - A picture is worth a thousand words …
  - Which 1000 words?
- Semantic gap: How to build automatic concept classifiers and image annotation models?
- Intention gap: How to capture user intention?

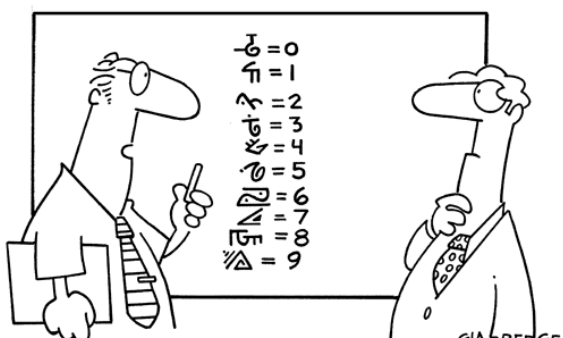
Feature Vectors



What kinds of image features should be used?

How to map them to words?





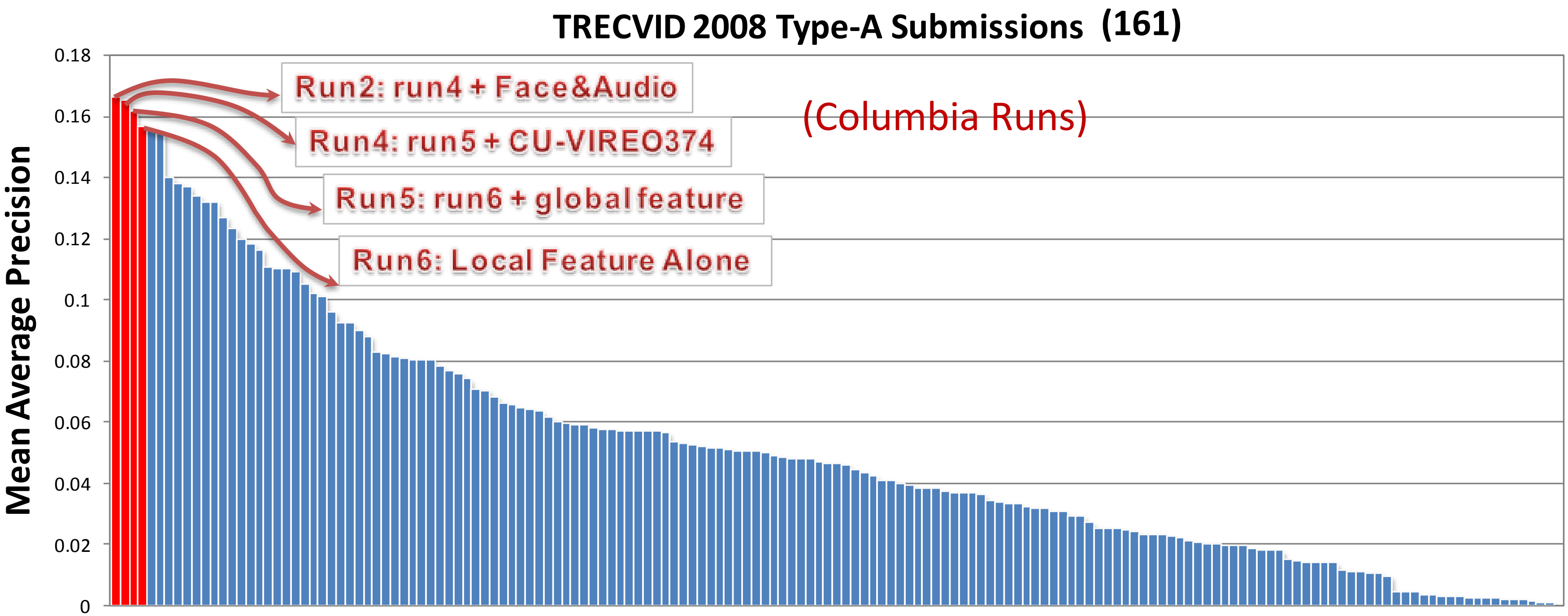
Let’s view the feature groups as words!

Concept Classifiers and Image Annotation Models

- Area of extensive studies
- Many models developed
  - Machine learning is the core
- Recent works
  - Content-based soft annotation, E. Chang, *et al*
  - Real-time annotation of pictures. J. Li & J. Wang
  - Simple Classifiers using Global Features
  - Counting Local Key Points: Bags of Words (BoW)
  - Local Features Multi-BoW Spatial Pyramid Kernel
  - Multi-label annotation framework
  - …

Local Features Prevail in Recent TRECVID

(Chang, *et al*, TRECVID 2008)



– Spatial local features achieve near top performance
– Other features (global, context, face, audio) help

(Slide courtesy of S.F. Chang)

Sample Detection Results (TRECVID2008)

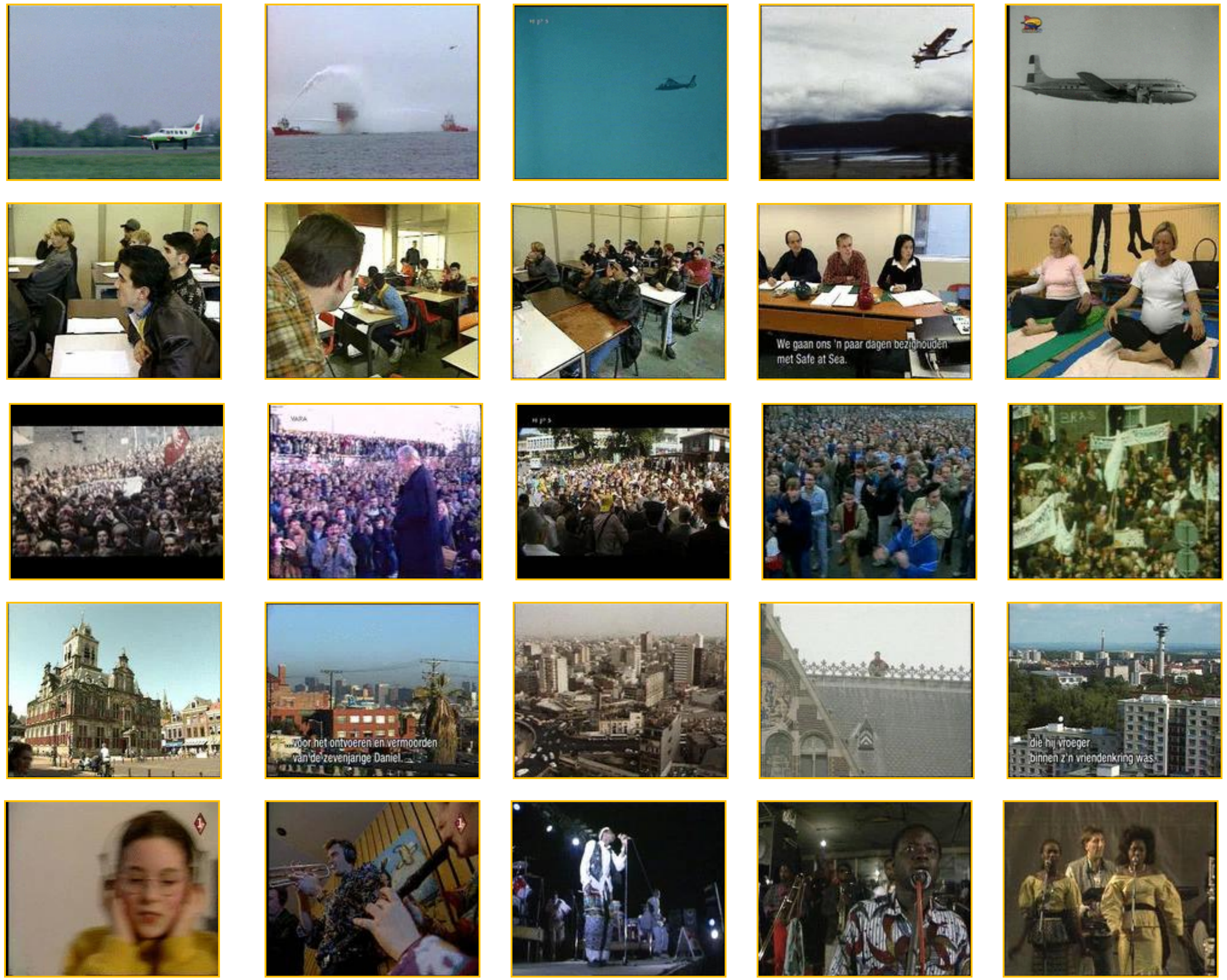
Airplane flying

Classroom

Demonstration Or Protest

Cityscape

Singing



(Slide courtesy of S.F. Chang)

Summary: Concept Classifiers and Annotation Models

- Try to answer two questions:
  - What kinds of image features should be used?
  - How to map them to words?
- Many models developed
  - Machine learning is the core
  - Success in relative small-scale image databases
- Key issue: Scalability
  - Low in precision and non-generalizable, due to scarcity of training samples , visual diversity
  - Training large amount of classifiers computationally prohibitive

How Has WWW Changed CBIR ?

Tones of data
Billions of users
Billions of interactions
Thousands of results

What's Happening

4 billion (June 2009)

- **~4000** uploads/minute
- 128 years to view all of them (1s per image)
- 2% Internet users visit
- Daily time on site: 4.7 minutes

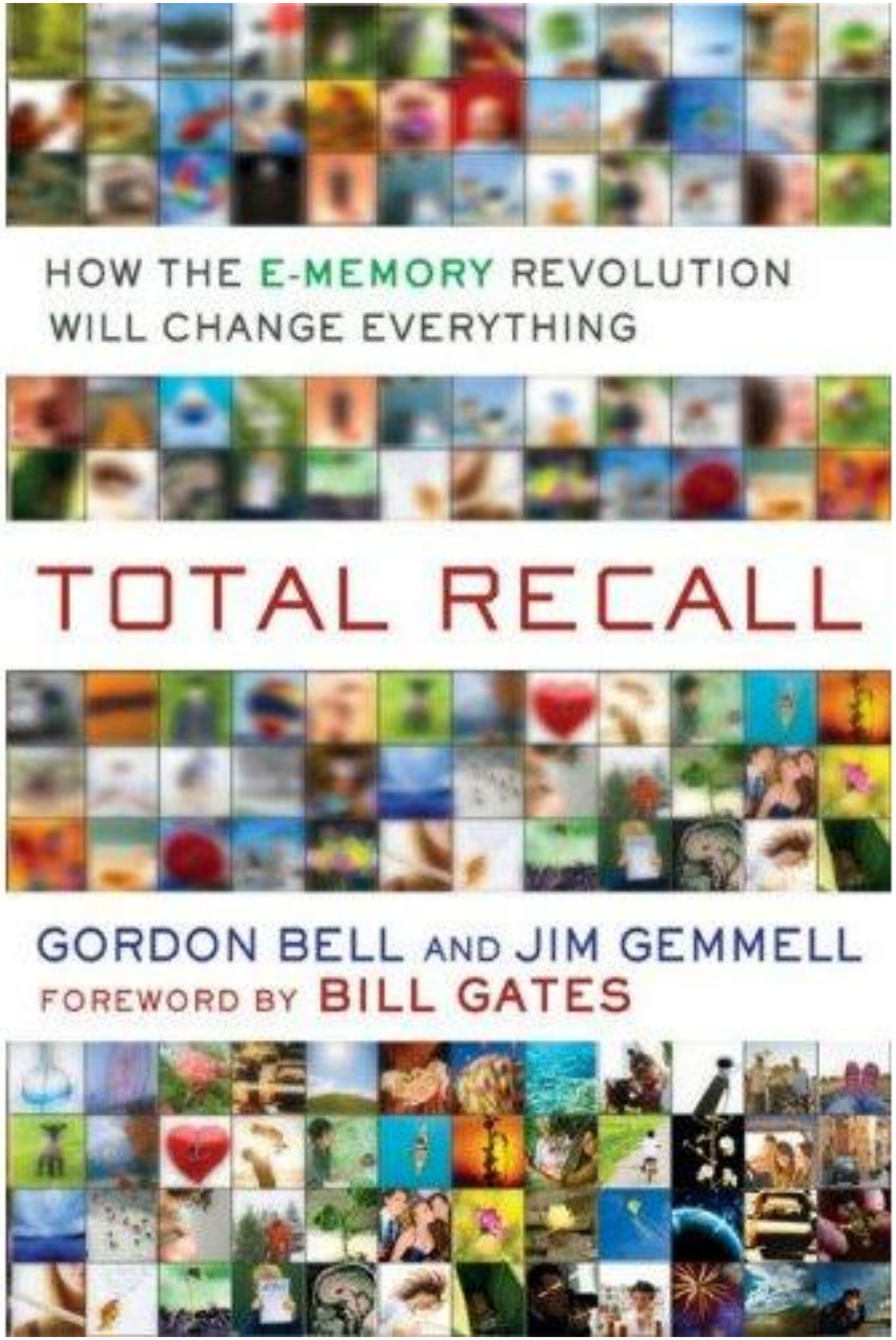
120 million (July 2009)

- **~20 hours** uploaded/minute
- 600 years to see all of them
- 20% Internet users visit
- Daily time on site: 23 minutes

- 2007 bandwidth = entire Internet in 2000
- March 2008: bandwidth cost US$1M a day

15 billion (April 2009)

- **~22000** uploads/minute
- 480 years to view all of them (1s per image)
- 24% Internet users visit
- Daily time on site: 30 minutes


HOW THE E-MEMORY REVOLUTION
WILL CHANGE EVERYTHING
TOTAL RECALL
GORDON BELL AND JIM GEMMELL
FOREWORD BY BILL GATES

Inventors of the CCD Camera Chip Share Nobel in Physics

IEEE Fellows Willard Boyle and George Smith started the digital-image revolution

7 October 2009—Willard Boyle and George Smith, formerly of Bell Laboratories, in Murray Hill, N.J., will share half of this year's Nobel Prize in Physics "for the invention of an imaging semiconductor circuit-the CCD," the basis for digital imagery in everything from pocket cameras to the Hubble Space Telescope. (The "imaging" part of the citation is in dispute, as the first imaging CCD was developed by IEEE Fellow Michael F. Tompsett, a colleague of Boyle and Smith.) In announcing the awards, the Royal Swedish Academy of Sciences called Boyle and Smith "masters of light" and said that, with fellow winner and optical-fiber pioneer Charles Kuen Kao, they "helped to shape the foundations of today's networked societies."

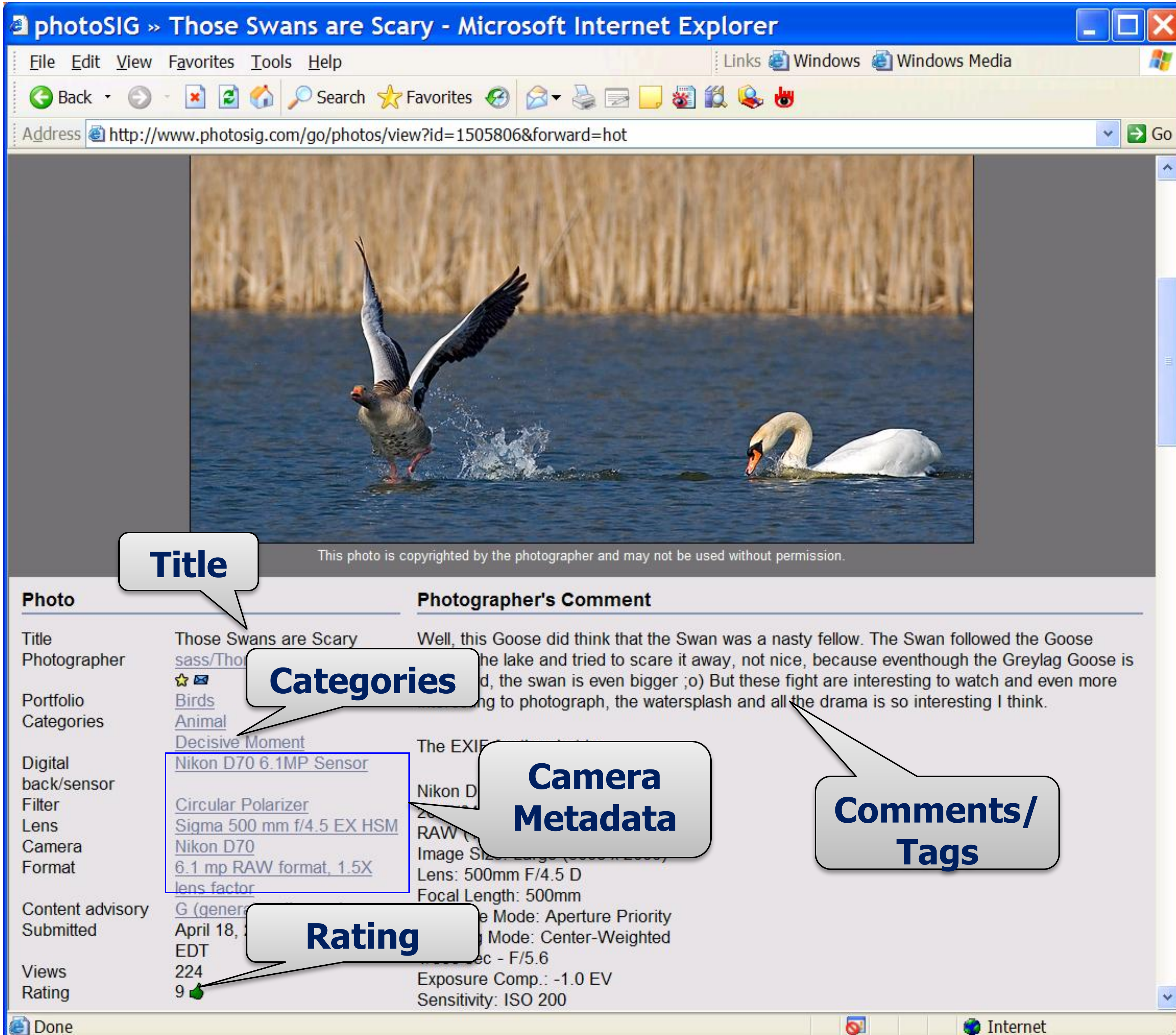
- IEEE Spectrum

How Has WWW Changed CBIR ?

Tones of data
Billions of users
Billions of interactions
Thousands of results

Issues in Web Image Search

- Label collection
  - Billions of user tags
- What are useful and feasible categories ?
- Semantic gap: How to build automatic concept classifiers and text annotation models ?
  - Data driven, model less
  - Inter play between text data and visual features
- Intention gap: How to capture user intention?
  - UI and query formation
  - Search result organization


photoSIG » Those Swans are Scary - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Links Windows Windows Media
Back Search Favorites
Address http://www.photosig.com/go/photos/view?id=1505806&forward=hot
Go
This photo is copyrighted by the photographer and may not be used without permission.
Title
Photo
Photographer's Comment
Title
Those Swans are Scary
Photographer
Well, this Goose did think that the Swan was a nasty fellow. The Swan followed the Goose
he lake and tried to scare it away, not nice, because eventhough the Greylag Goose is
d, the swan is even bigger ;o) But these fight are interesting to watch and even more
ng to photograph, the watersplash and all the drama is so interesting I think.
Categories
Portfolio
Birds
Categories
Animal
Decisive Moment
The EXIF
Digital back/sensor
Nikon D70 6.1MP Sensor
Camera Metadata
Comments/ Tags
Filter
Circular Polarizer
Lens
Sigma 500 mm f/4.5 EX HSM
Camera
Nikon D70
Format
6.1 mp RAW format, 1.5X lens factor
RAW
Image Si
Lens: 500mm F/4.5 D
Focal Length: 500mm
Content advisory
e Mode: Aperture Priority
Submitted
April 18,
EDT
Rating
Mode: Center-Weighted
ec - F/5.6
Views
224
Exposure Comp.: -1.0 EV
Rating
9
Sensitivity: ISO 200
Done
Internet

Texts Associated with Web Images:
Surrounding Texts and Tags

- Large but not systematic vocabularies
- Often low relevance to visual content
- No keyword annotation or ranking
- Ambiguous, subjective
- Incomplete, noisy

Texts need to be *extracted*, *processed* and *ranked*

Data Driven Approaches to Web Image Search

- Image annotation by search and mining
- Tag ranking and refinement
- Model selection
  - Finding high-level concepts with small semantic gaps
  - Learning a new similarity measure to reduce semantic gaps

Data Driven Approaches to Web Image Search

- Image annotation by search and mining
  - X. Wang, L. Zhang*, et al, AnnoSearch: Image Auto-Annotation by Search*, CVPR 2006
  - X. Li, L. Chen, L. Zhang, *et al, Image Annotation by Large-Scale Content-based Image Retrieval,* ACM MM'06
  - X. Wang, L. Zhang*, et al, Annotating Images by Mining Image Search Results,* PAMI'08
  - Many other related and continued works…

Image Annotation by Search + Mining:

Data Driven, Model Free

- Two basic stages:
  - **Searching similar images:** For an uncaptioned image *Iq, we first find a set of visually similar images Φs from a large-scale image database.*
  - **Mining representative keywords:** Given the image set *Φs, we further cluster the descriptive texts of Φs (i.e., image title, surrounding text, etc.) to find the most representative keywords as the annotations to Iq***.**

$$
\begin{aligned}
\mathbf{w}^* &= \arg\max_{\mathbf{w}} p(\mathbf{w} \mid I_q) \\
&= \arg\max_{\mathbf{w}} \sum_i p(\mathbf{w} \mid I_i) \cdot \underbrace{p(I_i \mid I_q)}_{\text{Search}} \qquad \cdots (*) \\
&= \arg\max_{\mathbf{w}} \sum_i \Big(\underbrace{\sum_j p(\mathbf{w} \mid t_j) \cdot p(t_j \mid I_i)}_{\text{Mining}} \Big) \cdot p(I_i \mid I_q) \qquad \cdots (**)
\end{aligned}
$$

Image Annotation by Search + Mining

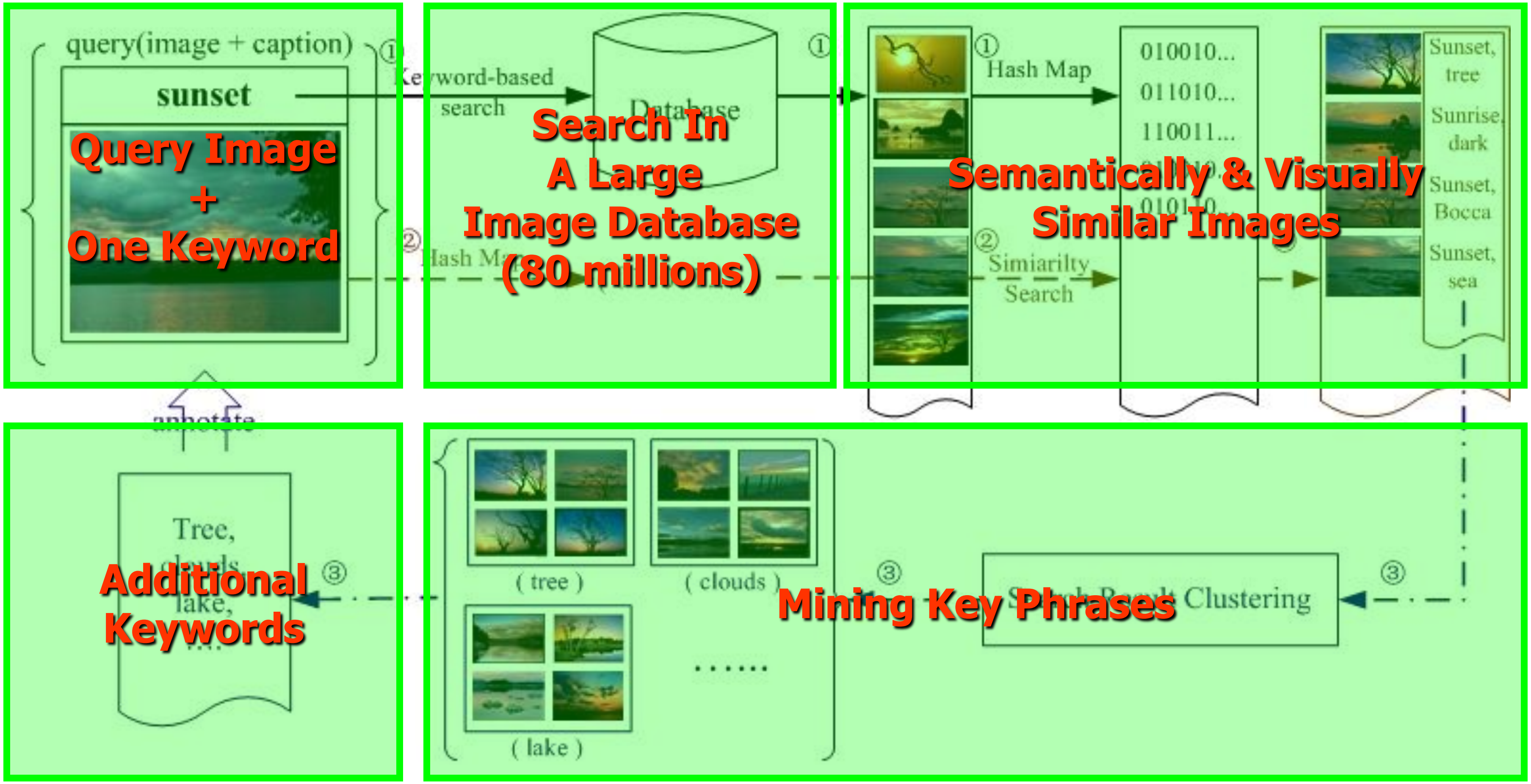
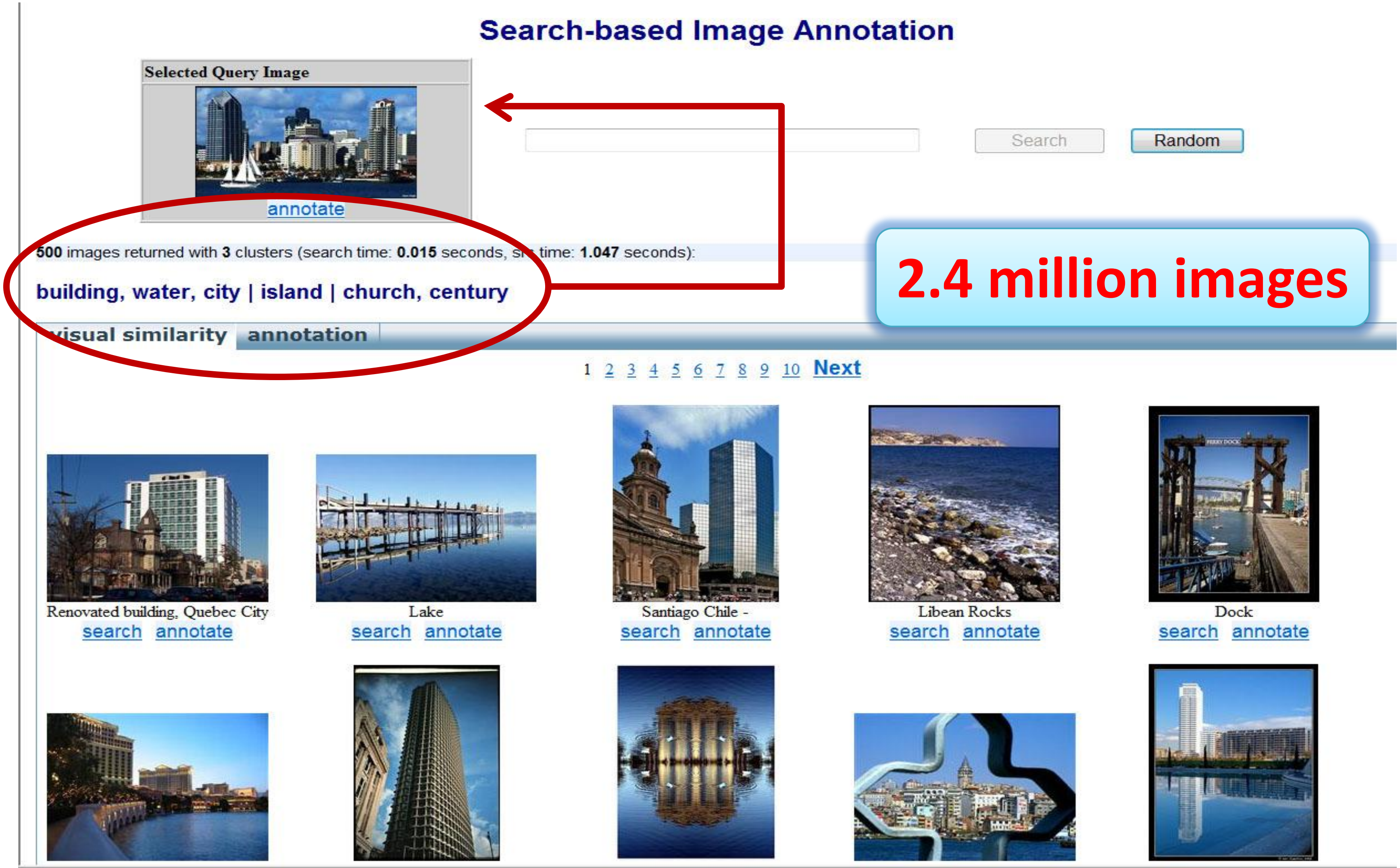
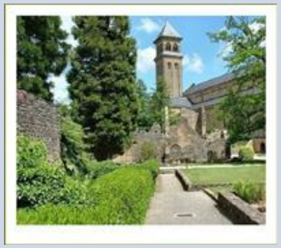
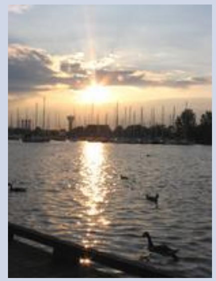
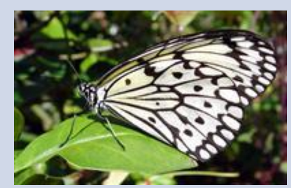


Image Annotation by Search + Mining



Annotation Examples (2.4M Images)

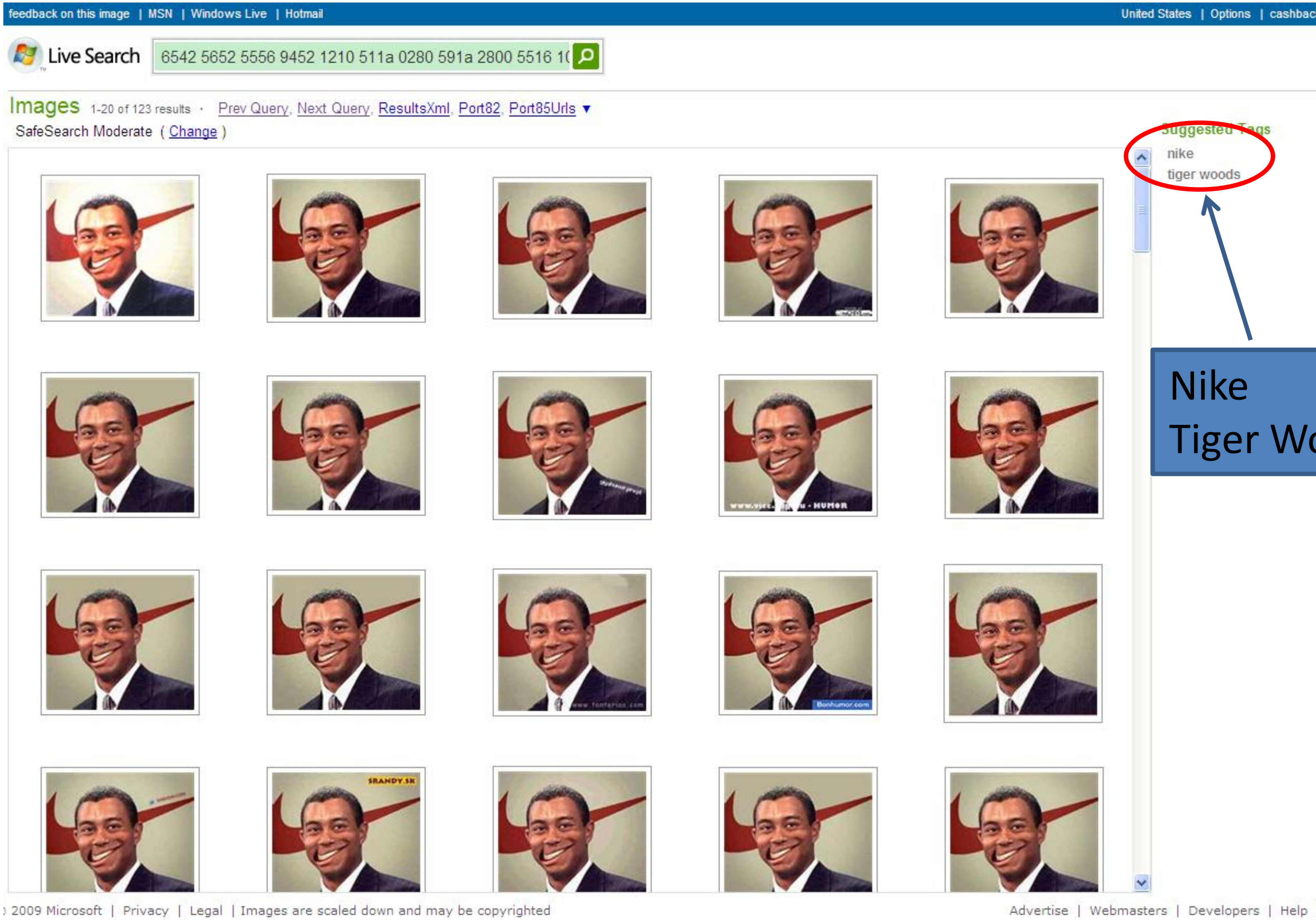
	house, castle, church, summer ; garden ; trees, water, sky ; ruins		sunset, water ; beach ; zoo ; lake		sky ; lake, water, river ; clouds ; trees, mountains, snow ; building
	summer, mountains		snow ; city, sky		model ; girl ; studio
	mountain, lake, water, tree ; hills, valley ; sky		house, town, window, village		butterfly ; flower ; fly ; frog, water ; tree, ground

Database – The Larger The Better

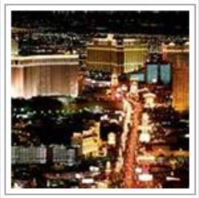
- Increase the size of image database

- Annotation based searching duplicate images in the web (2B)
  - Bin Wang, et.al. *Large-scale duplicate detection for web image search*, ICME 2006.

Duplicate Search from 2 Billion Images



Nike
Tiger Woods

	mercedes benz; swarovski crystal		Logo; mercedes benz; mercedes van; mercedes logo		chocolate, Red, Favorites
	Las vegas		Vegas; las vegas		sacre coeur; Paris; location vacances
	paris hilton; hollywood gossip;		barack obama; presidential candidate		bill gates
	frida kahlo; hope,tree,art; masters painter		van gogh; oil painting; drinkers, vangogh		van gogh; night café; oil paintings
	Happy birthday dog balloons; Glitter		Simpsons movie		travel inn; premier inn; Accommodation; city centre; basildon hotel
	pearl harbor josh hartnett		timber wolf		Monkey

Annotation Based On Duplicate Search from 2B Images

- Perfect for popular images
  - Celebrity, Product, Landmark, Cartoon, Paintings …
- However, not well for personal images
  - When there is no duplicate, the system will fail
  - Tag quality need to be improved

Data Driven Approaches to Web Image Search

- Image annotation by search and mining
- Tag ranking and refinement
  - D. Liu, X. Hua, H. Zhang, *Tag Ranking*, WWW 09
  - D. Liu, *et al*,Tag quality improvement for social images, ICME09
  - X. Li, C. Snoek, M. Worring, *Learning Tag Relevance by Neighbor Voting for Social Image Retrieval*, MIR08

Issues with User Tags of Images

- The most relevant tag often NOT ranked at the top in a tag list

1. alex
2. meditating
3. love
4. winter
5. dog
6. golden retriever
7. zen
8. calm
9. perfect
10. 1-5-fav
11. 5-10-fav
12. top_v111
13. 100v 10f
14. top20dogpix
15. top_f25
16. 25-0-fav

Issues with User Tags of Images

- The most relevant tag often NOT ranked at the top in a tag list
  - Only <10% images with the most relevant tag at the top of their tag list

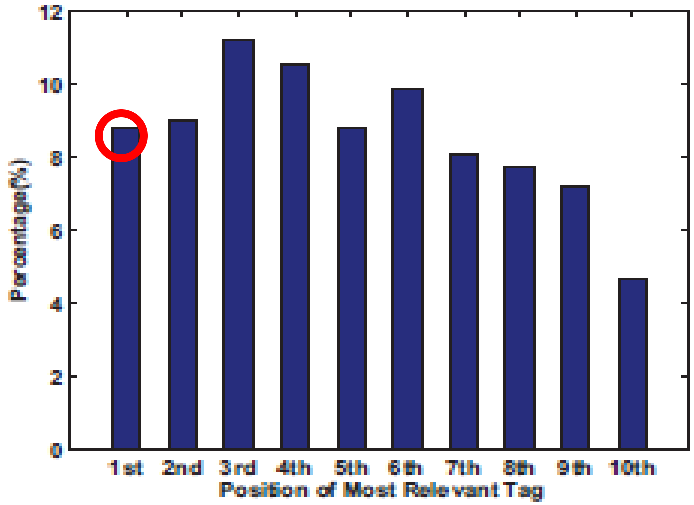
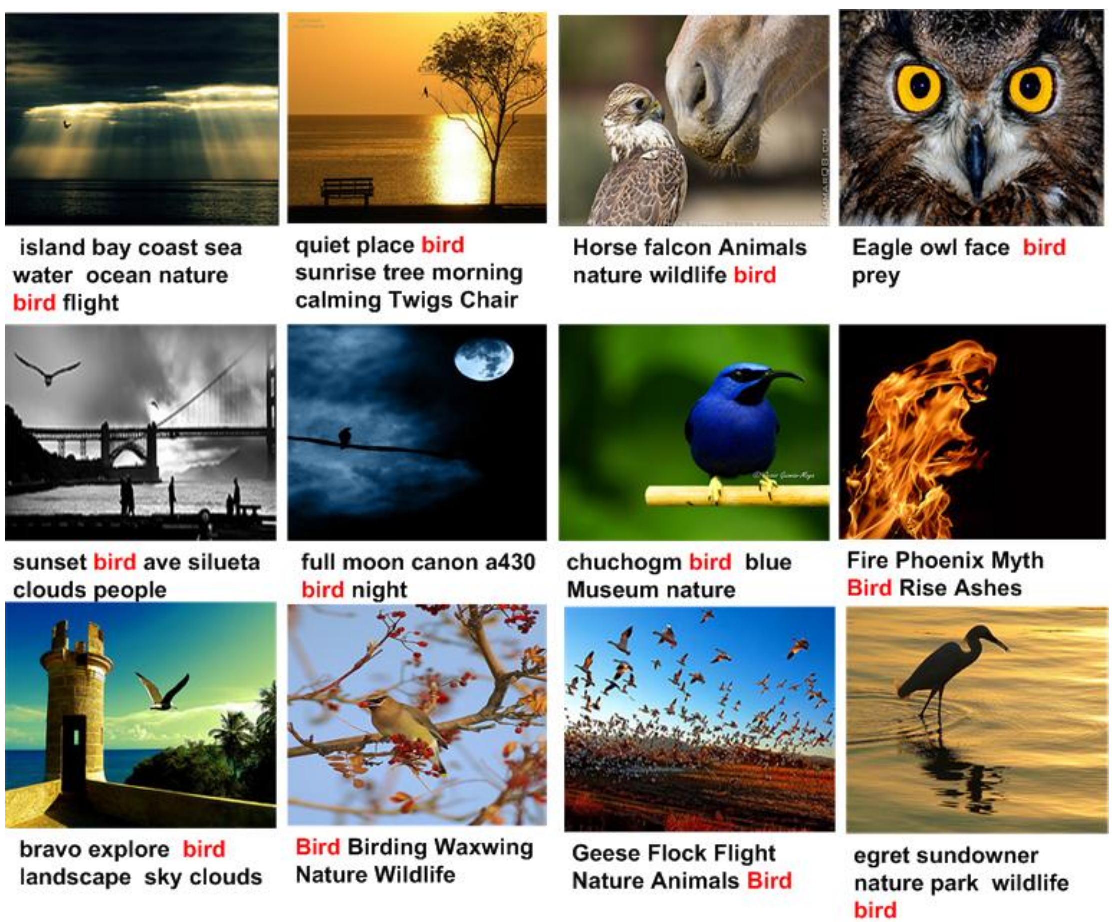


Figure 2: Percentage of images that have their most relevant tag at the n-th position in the associated tag list, where $n = 1, 2, 3 \ldots, 10$.

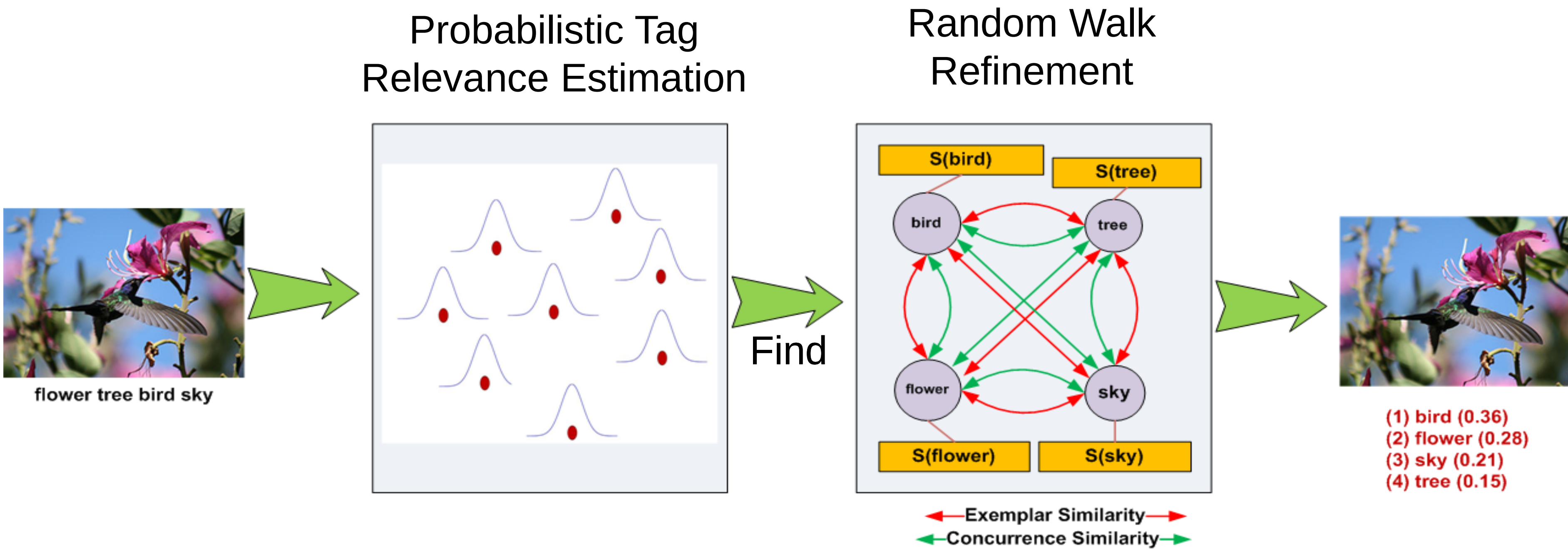
Issues with User Tags of Images

- The most relevant tag often NOT ranked at the top in a tag list
  - Only <10% images with the most relevant tag at the top of their tag list
  - Significantly limit the performance of tag-based image search



Automatic Tag Ranking

Liu, Hua, Zhang, *Tag Ranking*, WWW 09



- Basic ideas:
  - Large *tag clusters* should be promoted
  - *Semantically close tags* should be ranked closely
  - Initial tag relevance estimatic $s(t,x) \doteq p(x|t)$

    Density of image *x* in the image space with tag *t*

Tag Distance

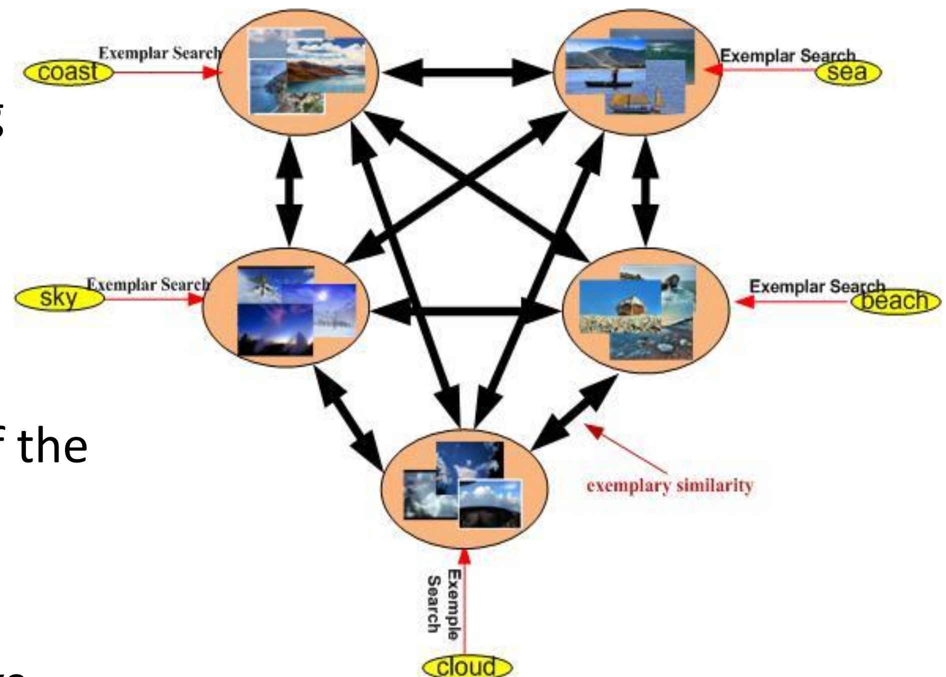
- WordNet distance
- Google distance

text-based, image irrelevant

- Tag Concurrence Distance

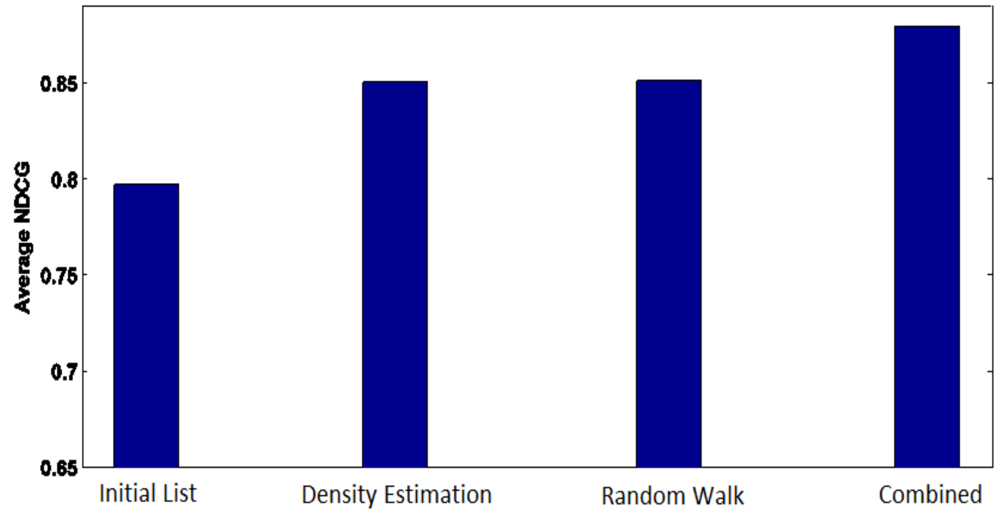
  - Tags are not complete
  - Image independent

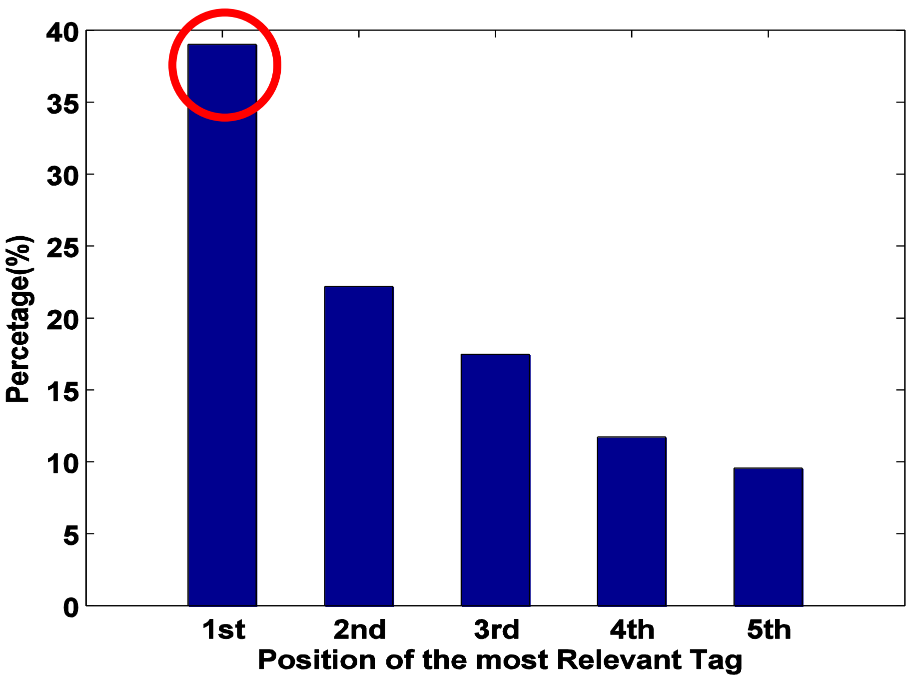
- Tag2Image Distance
  - Find images with a particular tag
  - Keep those close to the target image (finding N neighborhood)
  - Named as “Tag2Image Set”
  - Tag2Image: Distance between corresponding tag2image sets of the two tags
  - Pros and Cons
    - Image dependent
    - Finding neighbors may be expensive



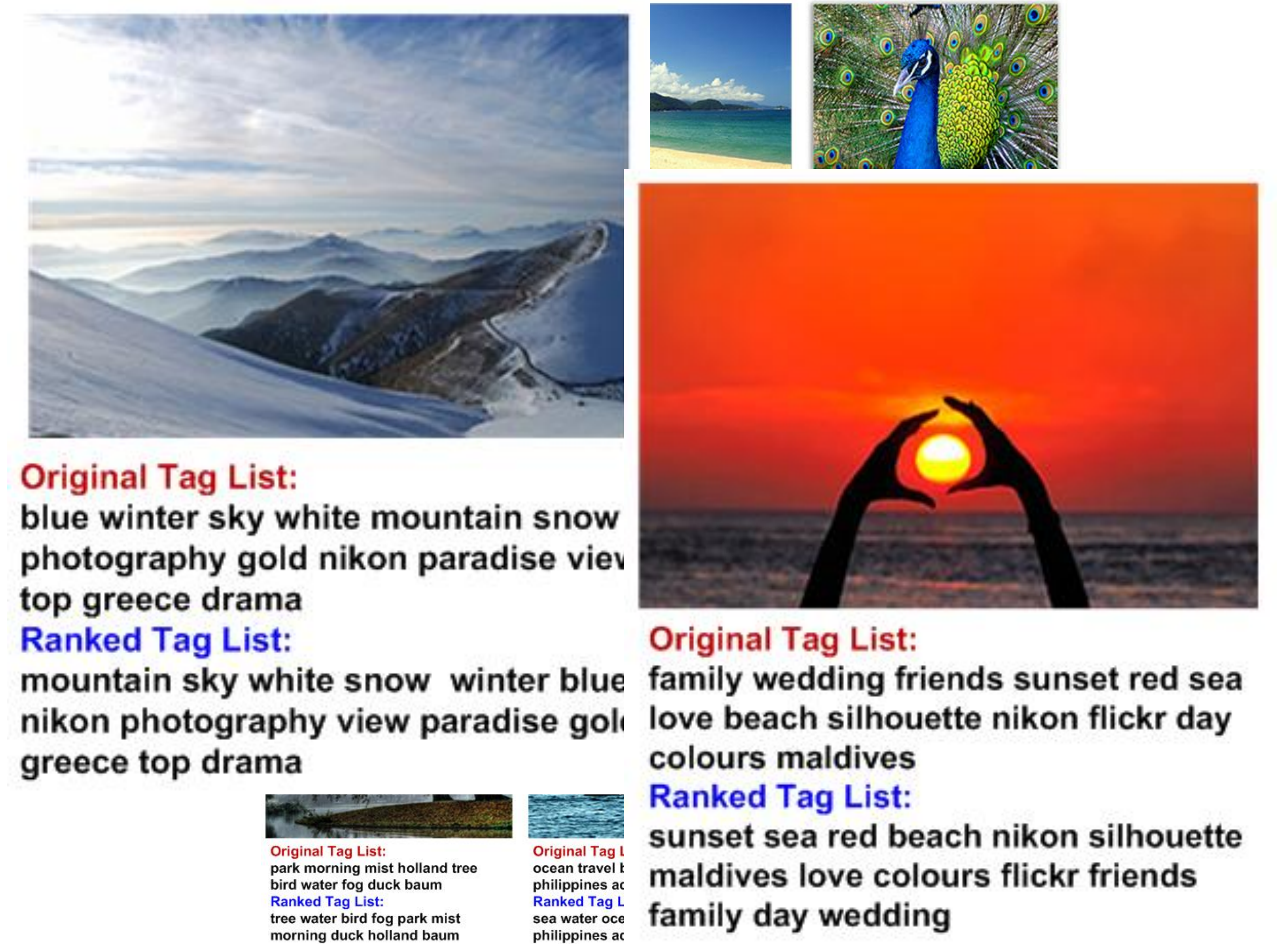
Results

- In term of average NDCG
  - 50,000 Flickr images (to mine distance and estimate density)
  - 13,330 unique tags
  - 10,000 test images (each labeled by 5 persons with five levels of relevance)



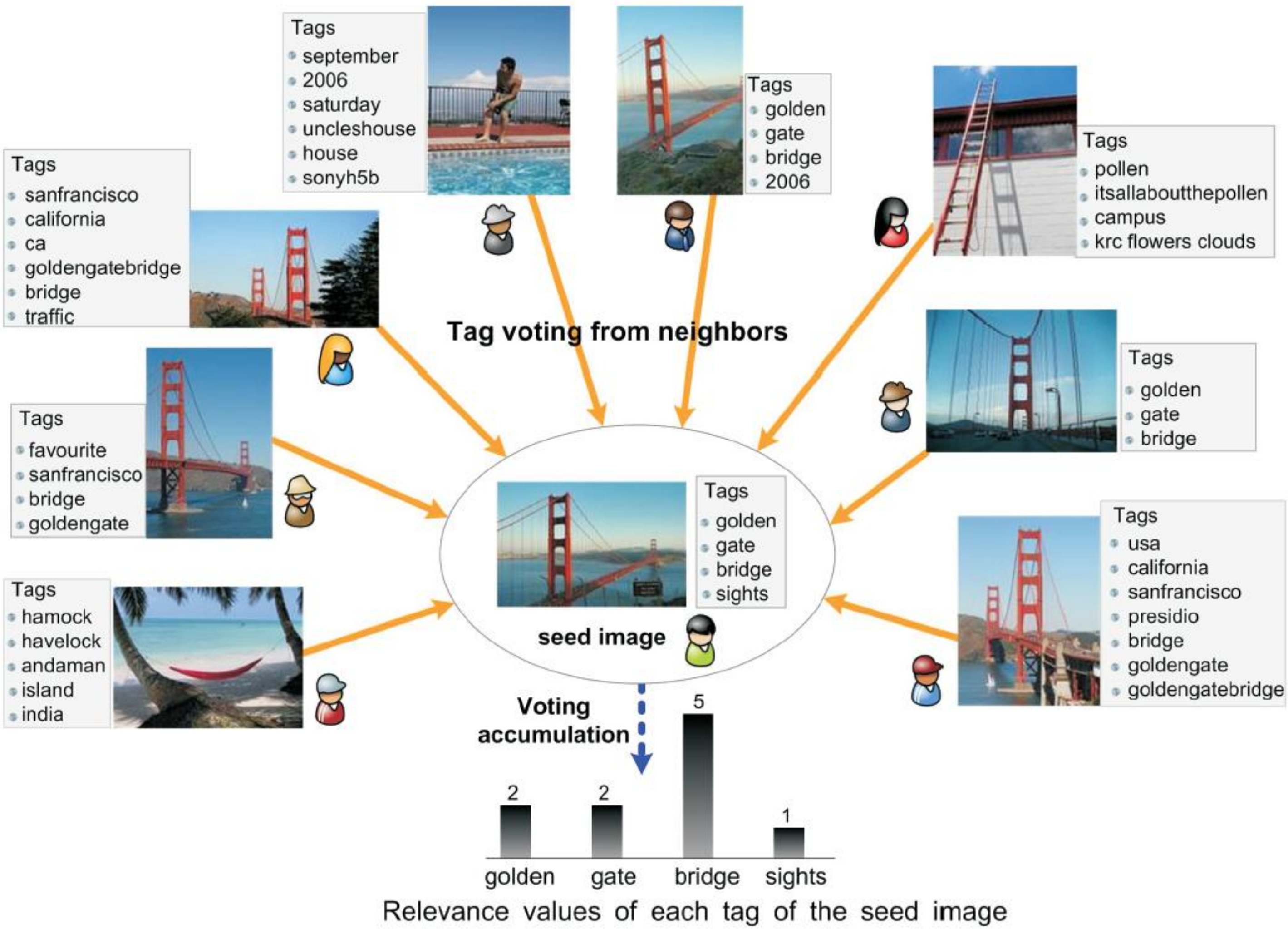


- After tag ranking, ~40% images have their most relevant tag appear at the top position in their tag list.


Original Tag List:
blue winter sky white mountain snow
photography gold nikon paradise
top greece drama
Ranked Tag List:
mountain sky white snow winter blue
nikon photography view paradise
greece top drama
Original Tag List:
family wedding friends sunset red sea
love beach silhouette nikon flickr day
colours maldives
Ranked Tag List:
sunset sea red beach nikon silhouette
maldives love colours flickr friends
family day wedding
Original Tag List:
park morning mist holland tree
bird water fog duck baum
Ranked Tag List:
tree water bird fog park mist
morning duck holland baum
ocean travel
Ranked Tag
sea water

Learning Tag Relevance by Neighbor Voting for Social Image Retrieval

Li, Snoek, & Worring, MIR08



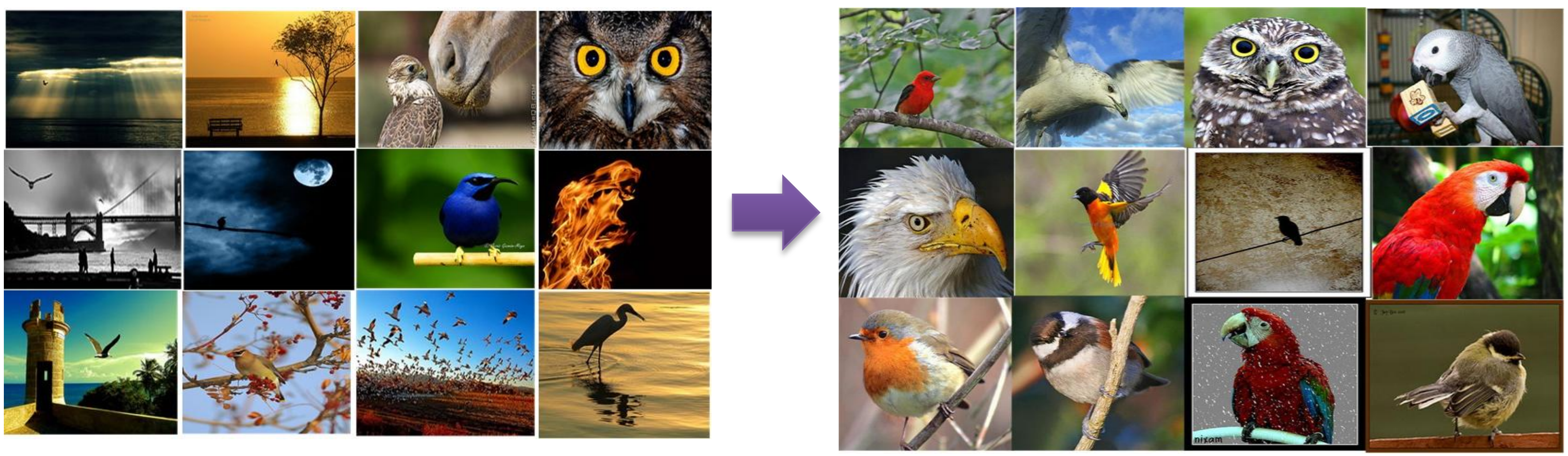
The relevance value of each tag is estimated by accumulating neighbor votes it receives from visually similar images of the seed image.

Application 1: Tag-based search

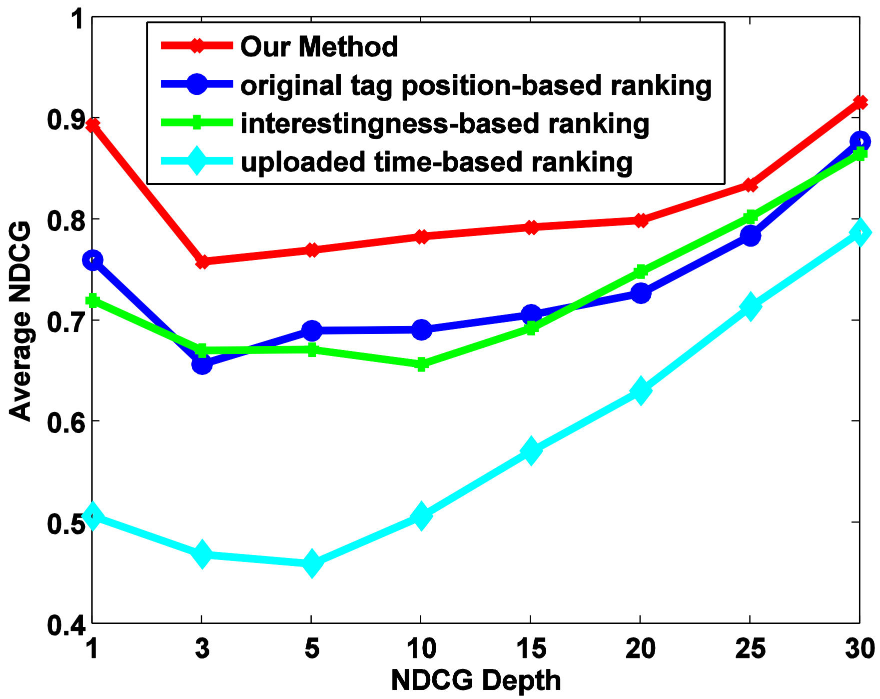
- Use tag position as relevance measure $r(x_i) = -\tau_i + 1/n_i$
- Ranking result for query “water”

Application 1: Tag-based search

- Use tag position as relevance measure $r(x_i) = -\tau_i + 1/n_i$
- Ranking result for query "bird"



Performance of Tag-Based Search



Our tag position-based ranking strategy outperforms all other image ranking strategies on Flickr

Application 2: Auto Tagging

- Use top tags of similar images as tags for a new uploaded image
- Performance

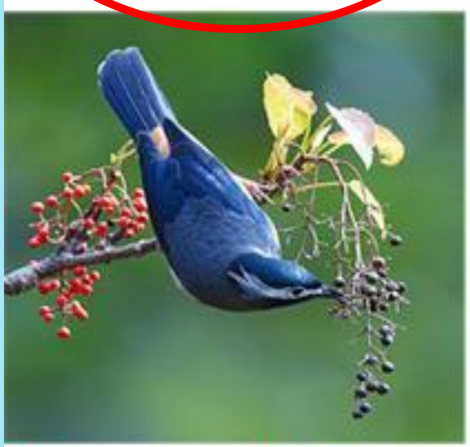
	Prec@1	Prec@5	Prec@all
Original(Baseline)	0.5858	0.4980	0.4980
Recommendation	0.7255	0.5799	0.5772
Improvement(%)	23.9	16.5	15.9

Recommended Tags:
water sky blue snow beauty landscape nature sea earth storm mountain cloud sunset light river

Recommended Tags:
flower plant flor red rose tree color

Recommended Tags:
sunset yellow red tree texture sunrise hill

Recommended Tags:
cat architecture tiger wildlife white sunlight mountain animal sunset bird eye yellow



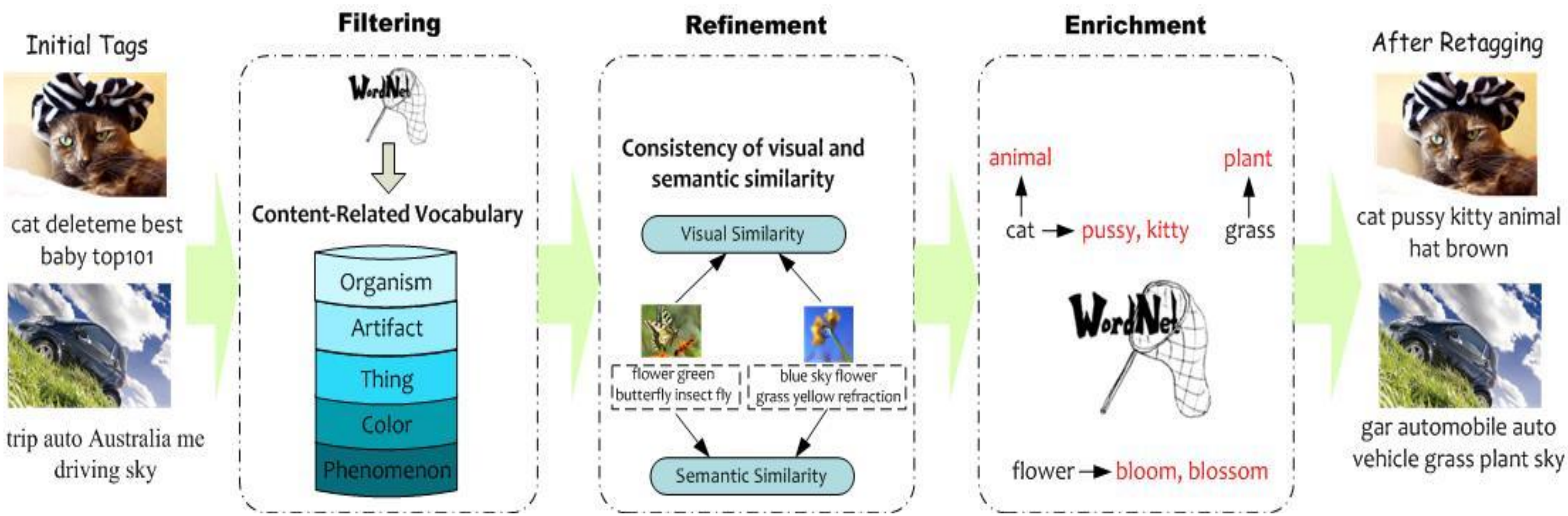
Recommended Tags:
bird flower water green

Recommended Tags:
sea mountain sky water blue beach landscape

Recommended Tags:
mountain sky landscape nature tree

Recommended Tags:
nature green forest tree water mountain

Automatic Tag Ranking: Extension



Discussion: Tag Ranking and Refinement

- Basic assumption of tag processing
  - Similar images <--> similar tags
  - Tags correlate with image content
- Applications
  - Tag based image search
  - Automated tagging of image by visual similarity
- A great example of data, user and feature inter play in image indexing

Data Driven Approaches to Web Image Search

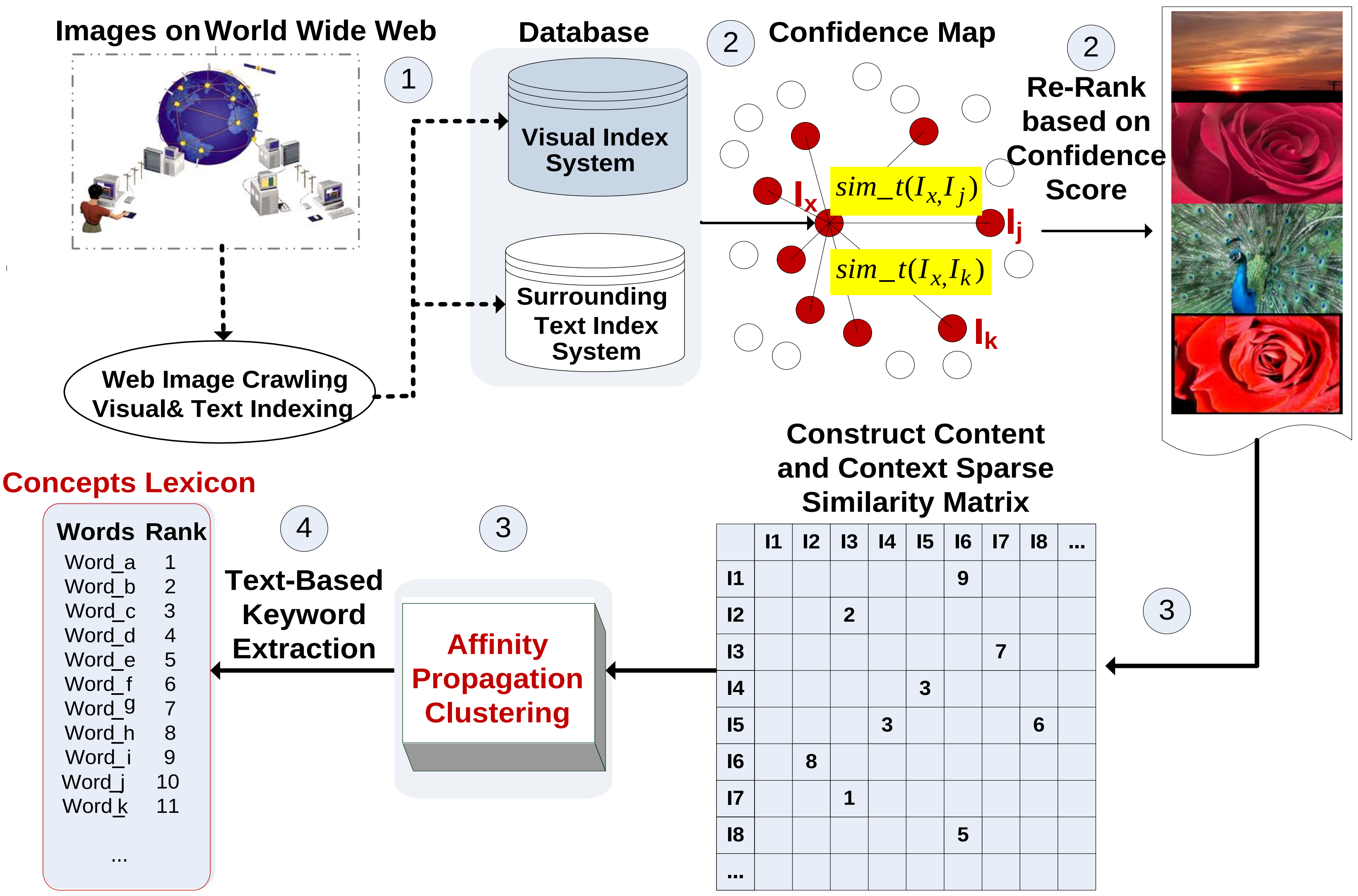
- Image annotation by search and mining
- Tag ranking and refinement
- Model selection: utilize rich textual data of Web images to reduce semantic gaps
  - Finding high-level concepts with small semantic gaps
  - Learning new similarity measures to reduce semantic gaps

Finding High-Level Concepts with Small Semantic Gaps

Y. Lu, *et al*, CVPR 2008

- Problem: How to find meaningful indexing terms from noisy surrounding texts of images?
- Basic idea
  - Different features for different concept modeling
    - Color feature → sunset, landscape, …
- Given a feature space, can we identify *high-level concepts with small semantic gaps*?
  - Images with small semantic gaps are selected and clustered by a confidence map and content-context similarity matrix
  - Mine a concept lexicon with small semantic gaps and high co-occurrences from the surrounding text of images
  - The mined lexicon builds an index of images

Framework

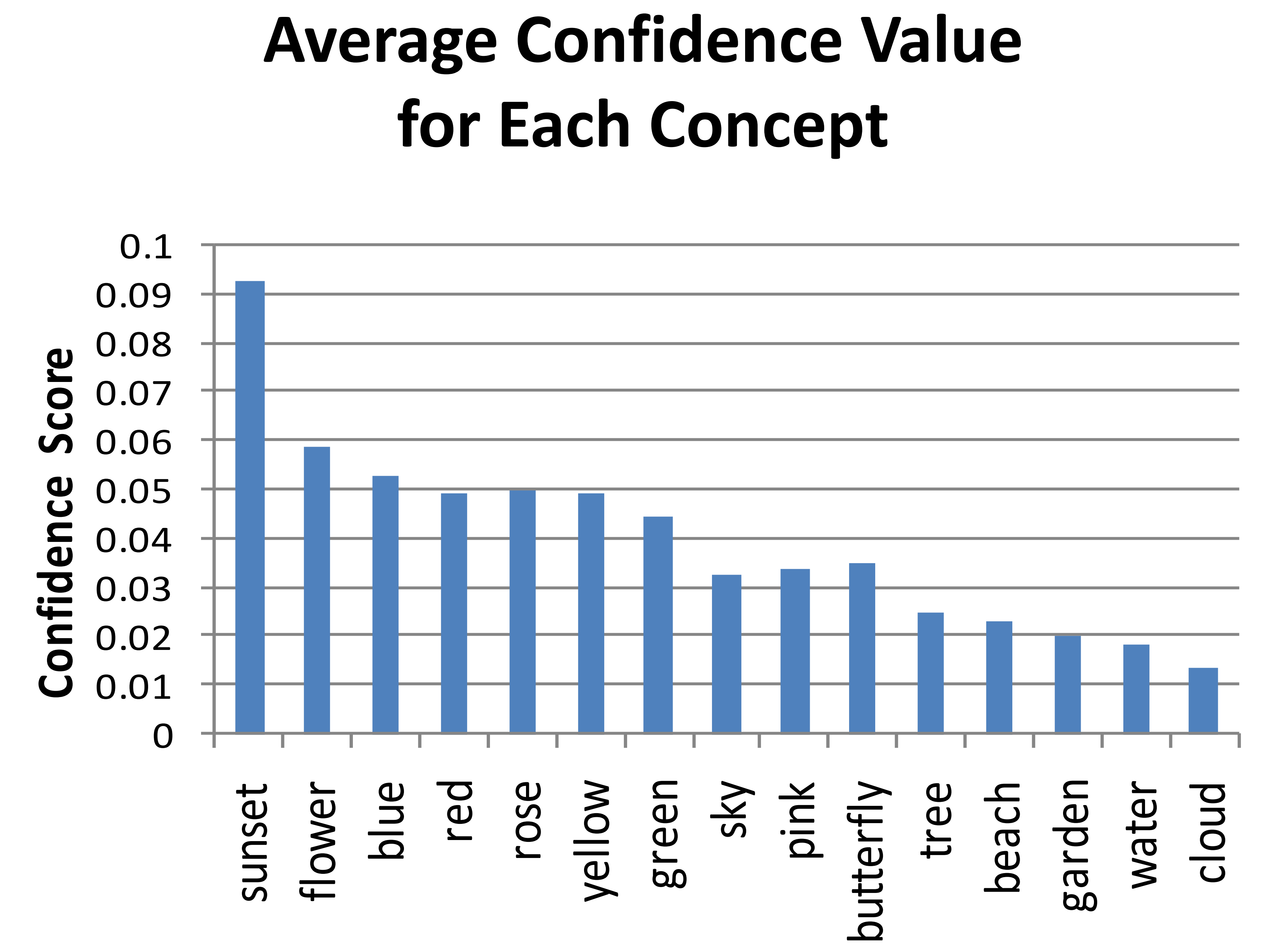


	I1	I2	I3	I4	I5	I6	I7	I8	...
I1						9			
I2			2						
I3							7		
I4					3				
I5				3				6	
I6		8							
I7			1						
I8						5			
...									

Data Collection & Feature Space

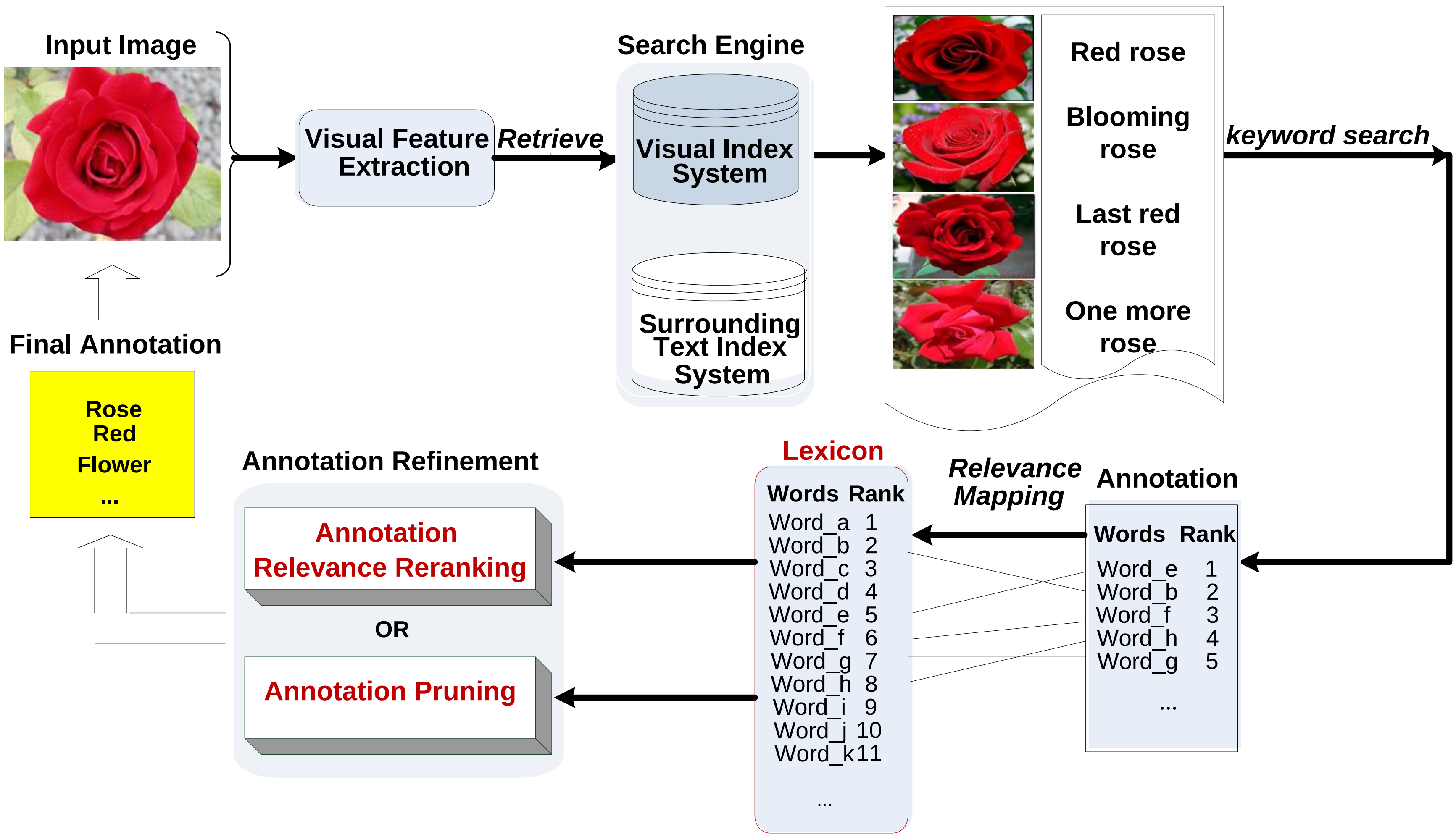
Image	Title	Descriptions
	Sea sunset	Sunset at the sea
	Red Rose	A rose in my garden taken June 8th 2002 (My other hobby is rose gardening)…
	The Falls	This is a waterfall that is about 3 miles from my house. It's called The Falls…

- 2.4 million web images from photo forums
- 64 dimensional global visual feature
  - color moments, color correlogram and color-texture moments



Average Confidence Value
for Each Concept
Confidence Score
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
sunset
flower
blue
red
rose
yellow
green
sky
pink
butterfly
tree
beach
garden
water
cloud

Applications: Annotation Refinement

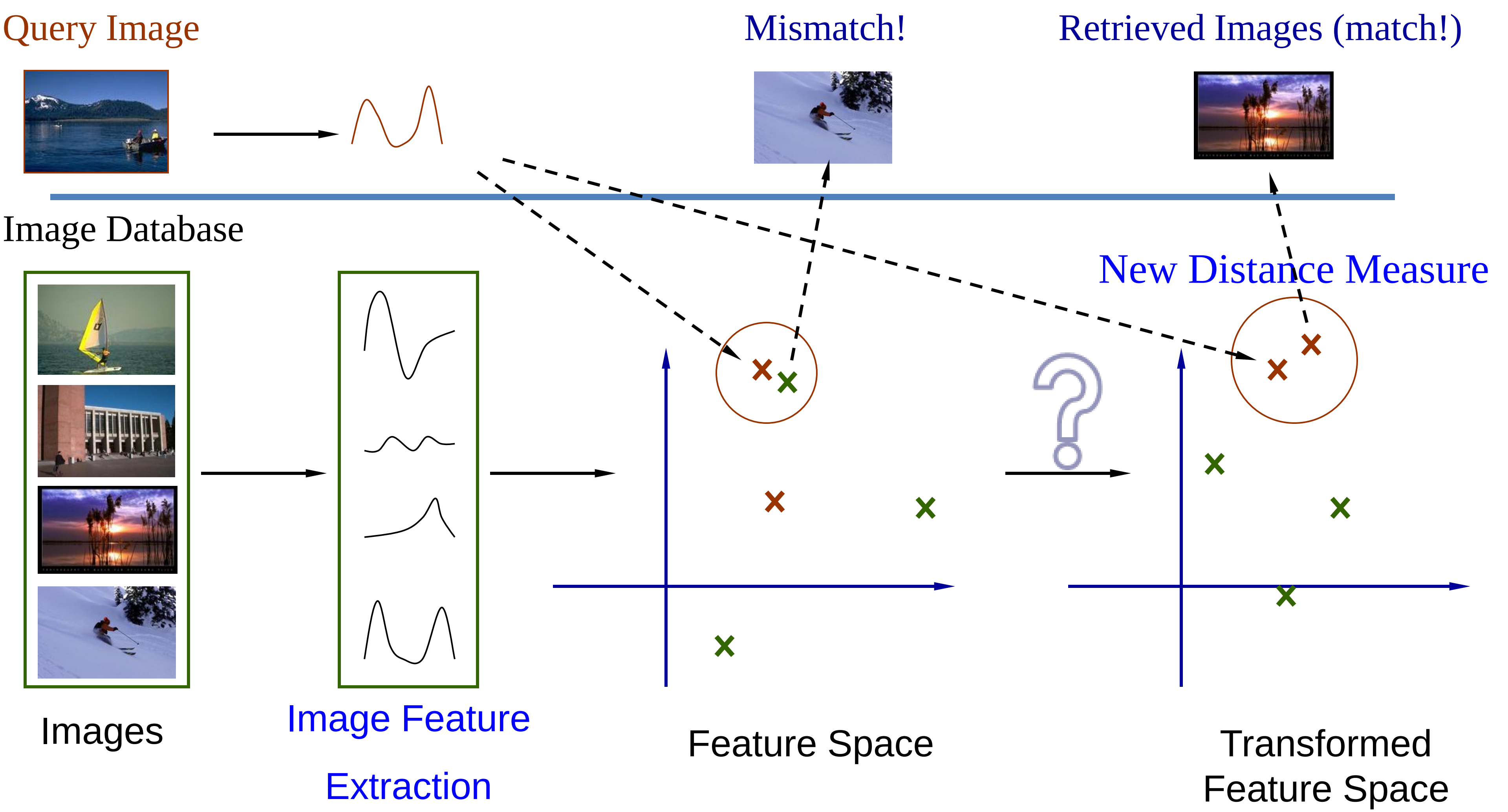


Learning New Similarity Measures to Reduce Semantic Gap

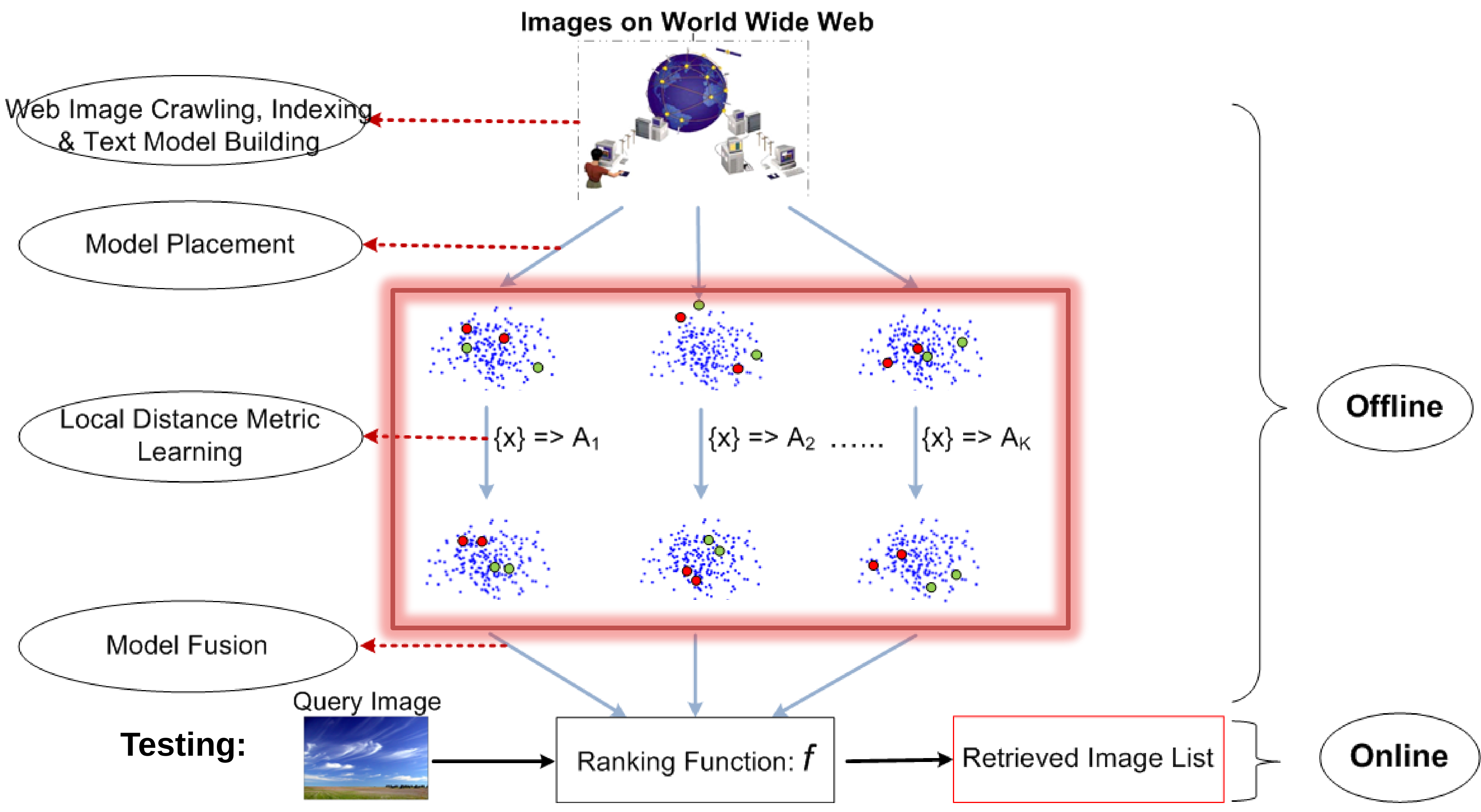
C. Wang, L. Zhang, H.J. Zhang, SIGIR 2008

- Basic idea
  - Input
    - A large scale Web image database with rich textual data
    - A query image with no textual descriptions
  - Output
    - Learn a new distance measure in the visual space to approximate the distance in the textual space
- Challenges
  - Scalability problem:
    - Local models learning, followed by a fusion stage
  - Noisy textual labels: document similarity
    - Term level similarity → cosine similarity
    - Topic level similarity → **LDA** similarity

Learning A New Similarity Measure

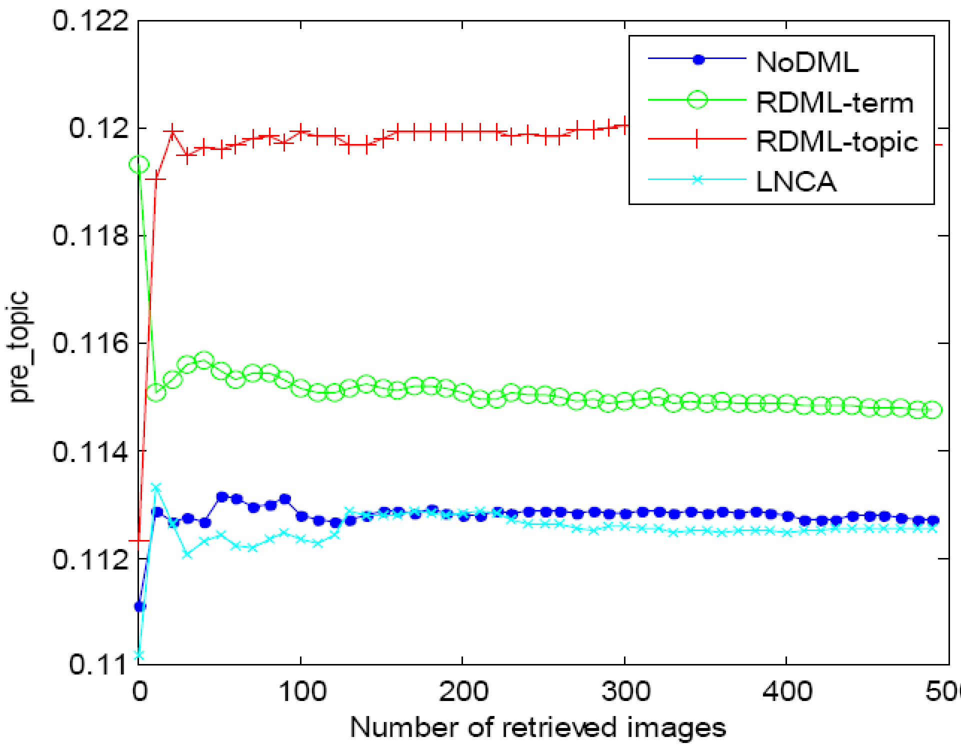


Framework – Local Distance Learning

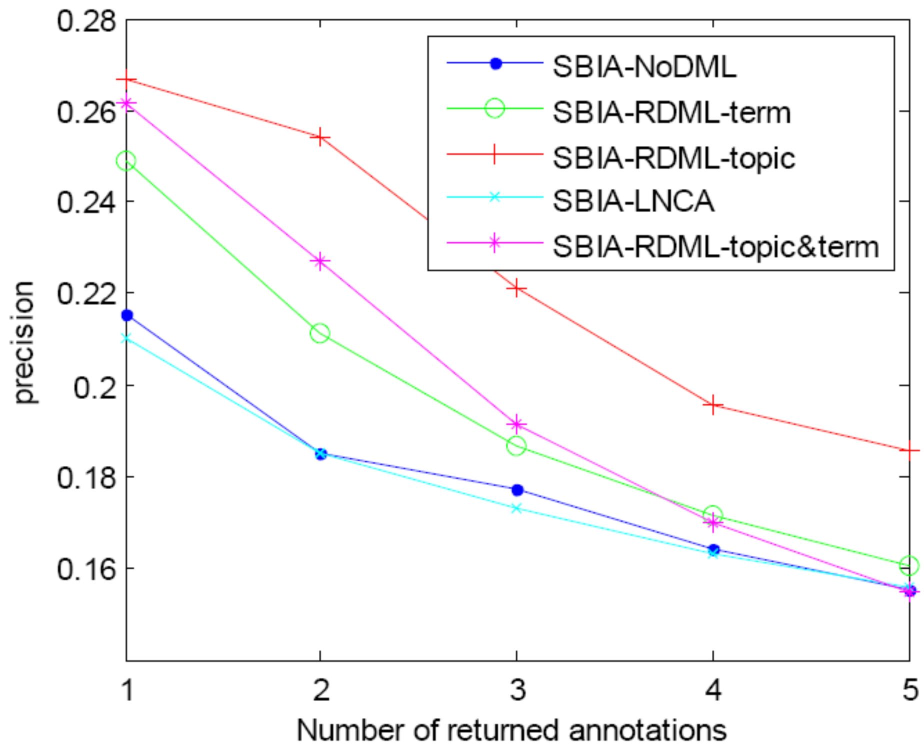


Experiments

- Training set: 2.4M web images
- Testing set: UW image dataset (1109 images)



Experiment 1: Content-based image retrieval: retrieval precisions by different models

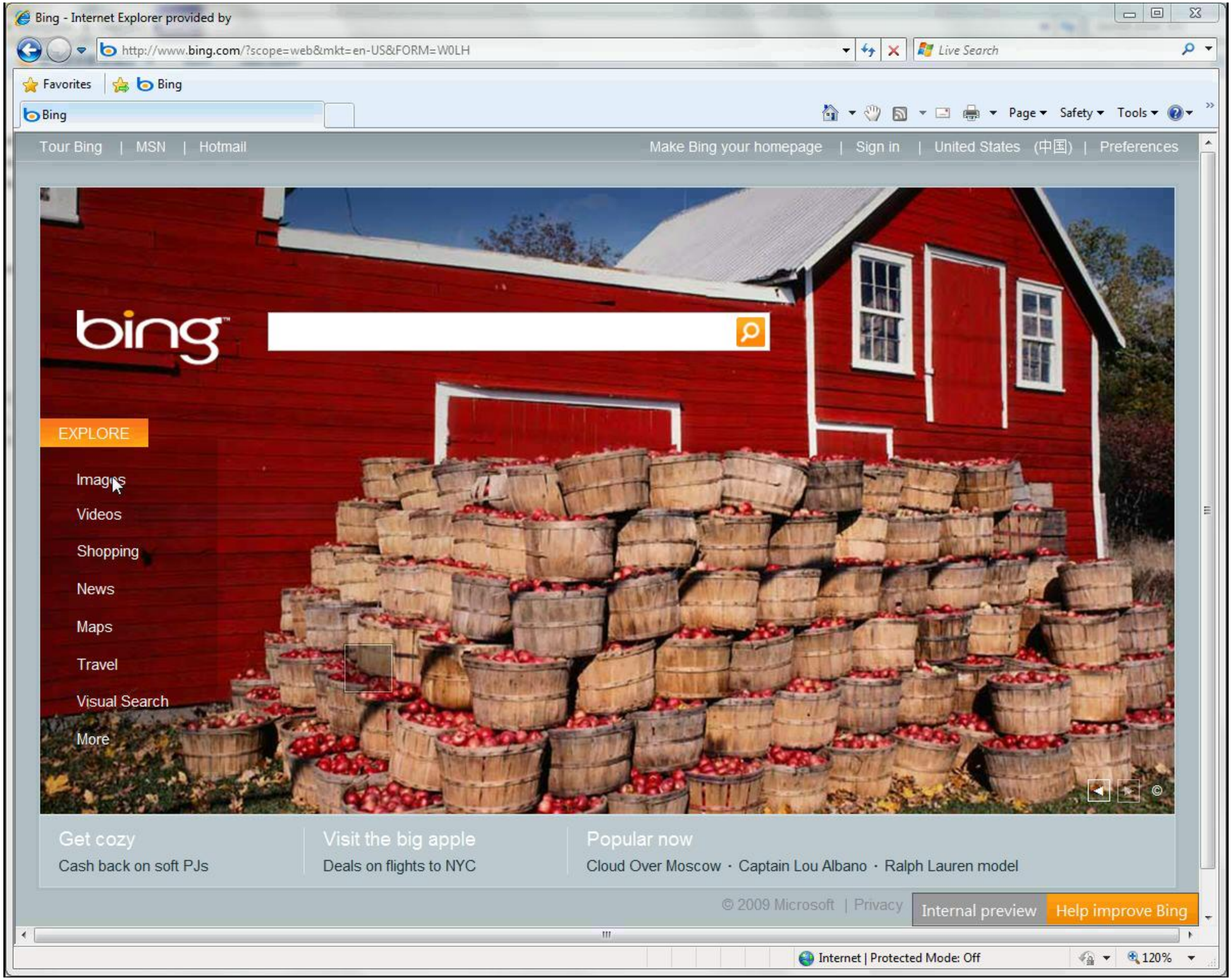


Experiment 2: Search-based image annotation: annotation precisions by different algorithms

Discussions

- Finding high-level concepts with small semantic gaps
  - High-level concepts with small semantic gap lexicon can be automatically selected
  - Useful for many CBIR applications
    - Auto annotation
    - Annotation refinement and rejection
- Learning new similarity measures to reduce semantic gaps
  - Text info is valuable for learning a better visual distance
  - Promising results in both image retrieval and annotation
  - Similar methods can be used in model selection

bing Image Search



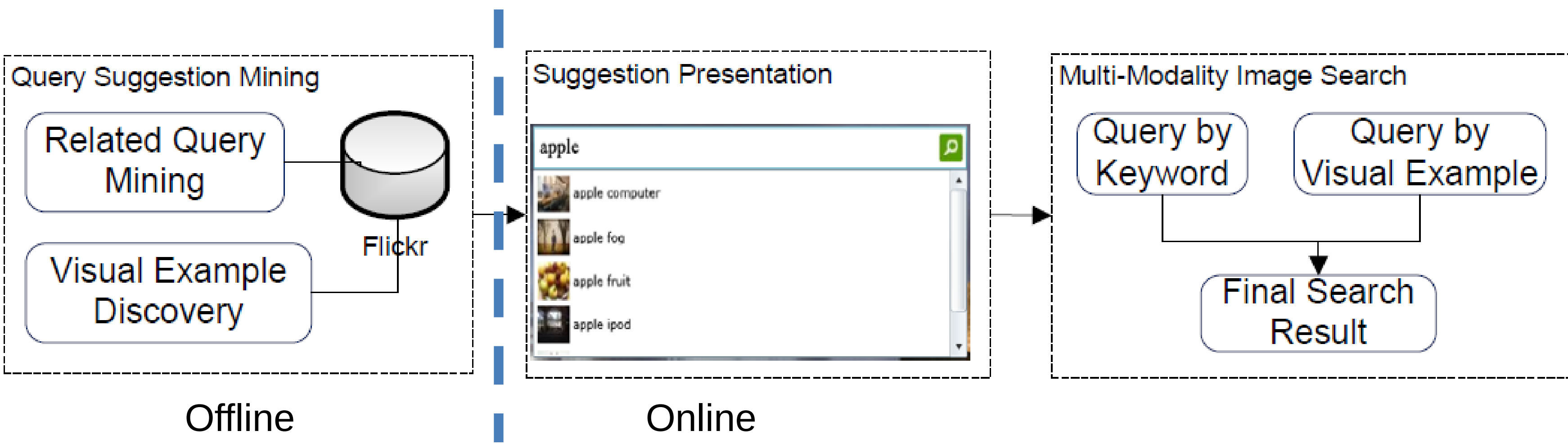
Issues in Web Image Search

- Label collection
  - Billions of user tags
- What are useful and feasible categories ?
- Semantic gap: How to build automatic concept classifiers and text annotation models ?
  - Data driven, model less
  - Inter play between text data and visual features
- Intention gap: How to capture user intention?
  - UI and query formation
  - Search result organization

Visual Query Suggestion

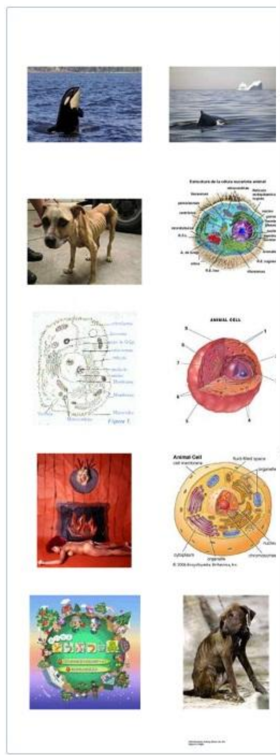
Z. Zha, et al, *Visual Query Suggestion*, ACM09

- Problem
  - *Intention gap*: the incapability of key word query to express the search intention in image search.
- Solution
  - Suggest visual queries in addition to keywords



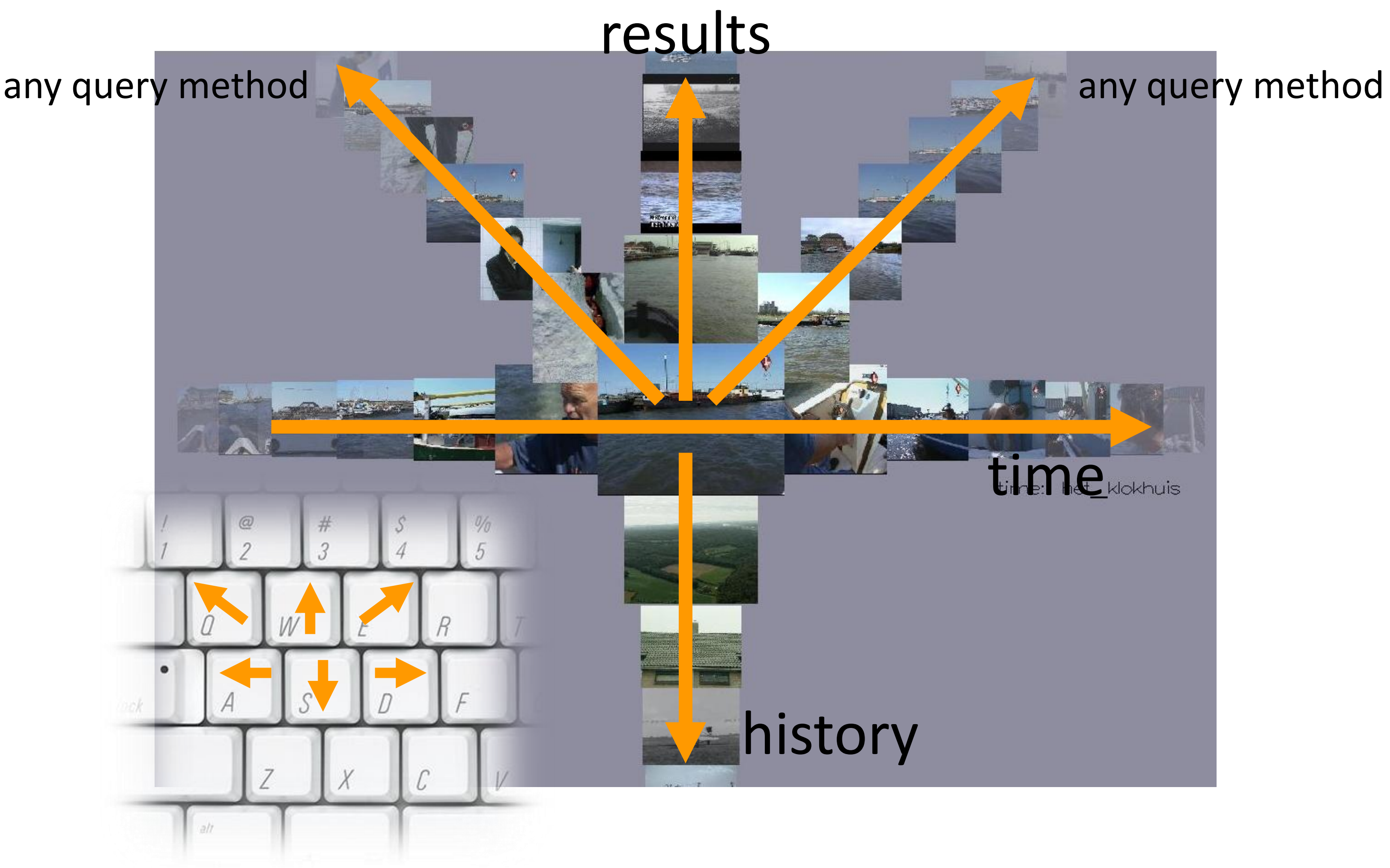
Visual Query Suggestion

- Sample keyword-image suggestions for three initial queries.
- Search results



The MediaMill Multi-Dimensional ForkBrowser

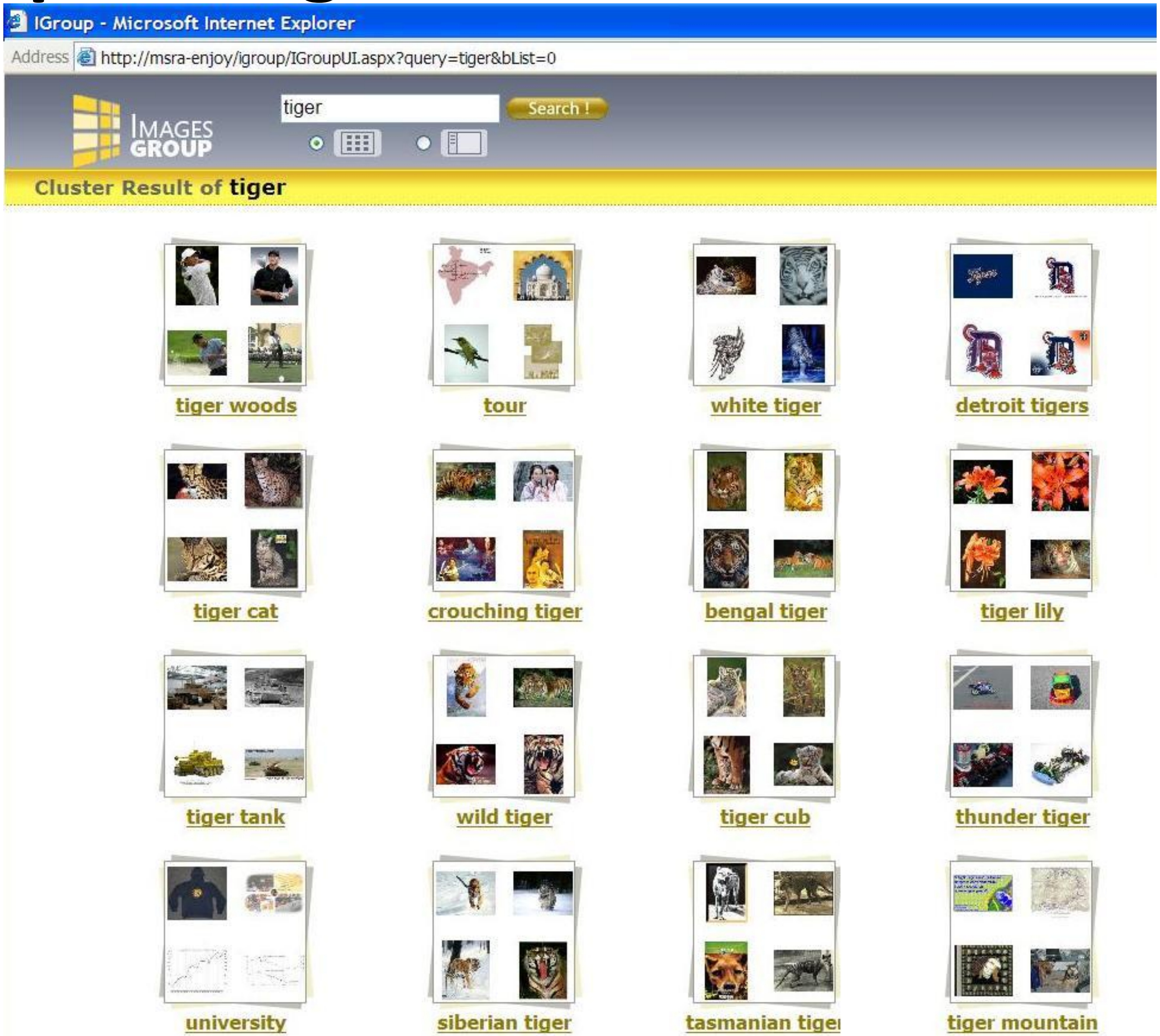
O. de Rooij, C G M Snoek, and M Worring, CIVR'08



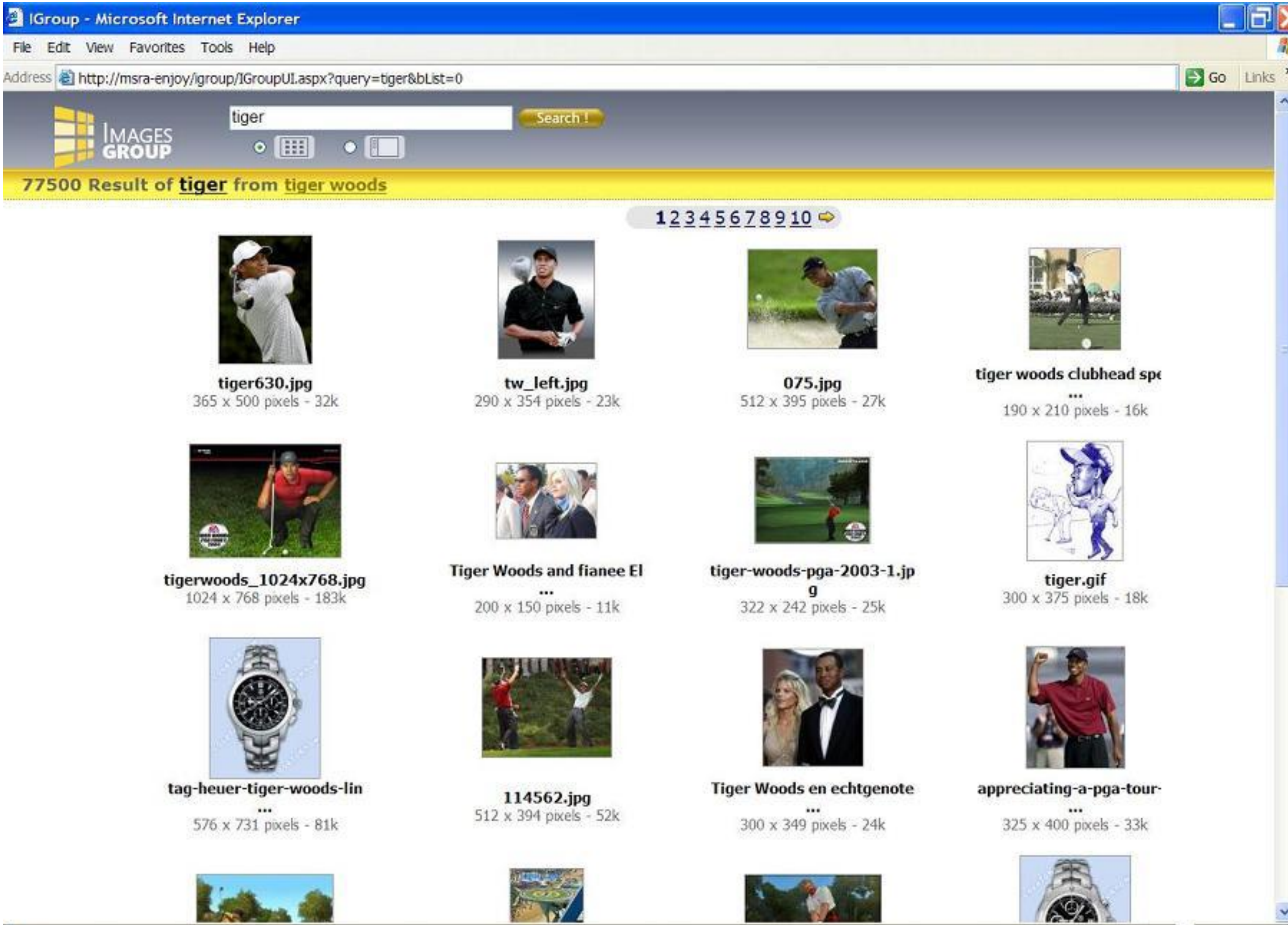
IGroup – Image Search Result Clustering

- A powerful navigation tool
  - Generate cluster names from general web search results and query log analysis
  - Tune cluster names specifically for image searches
  - Re-search for images by cluster names
  - Merge clusters from different sources
  - Group clusters according to image analysis

IGroup – Image Search Result Clustering



IGroup – Image Search Result Clustering

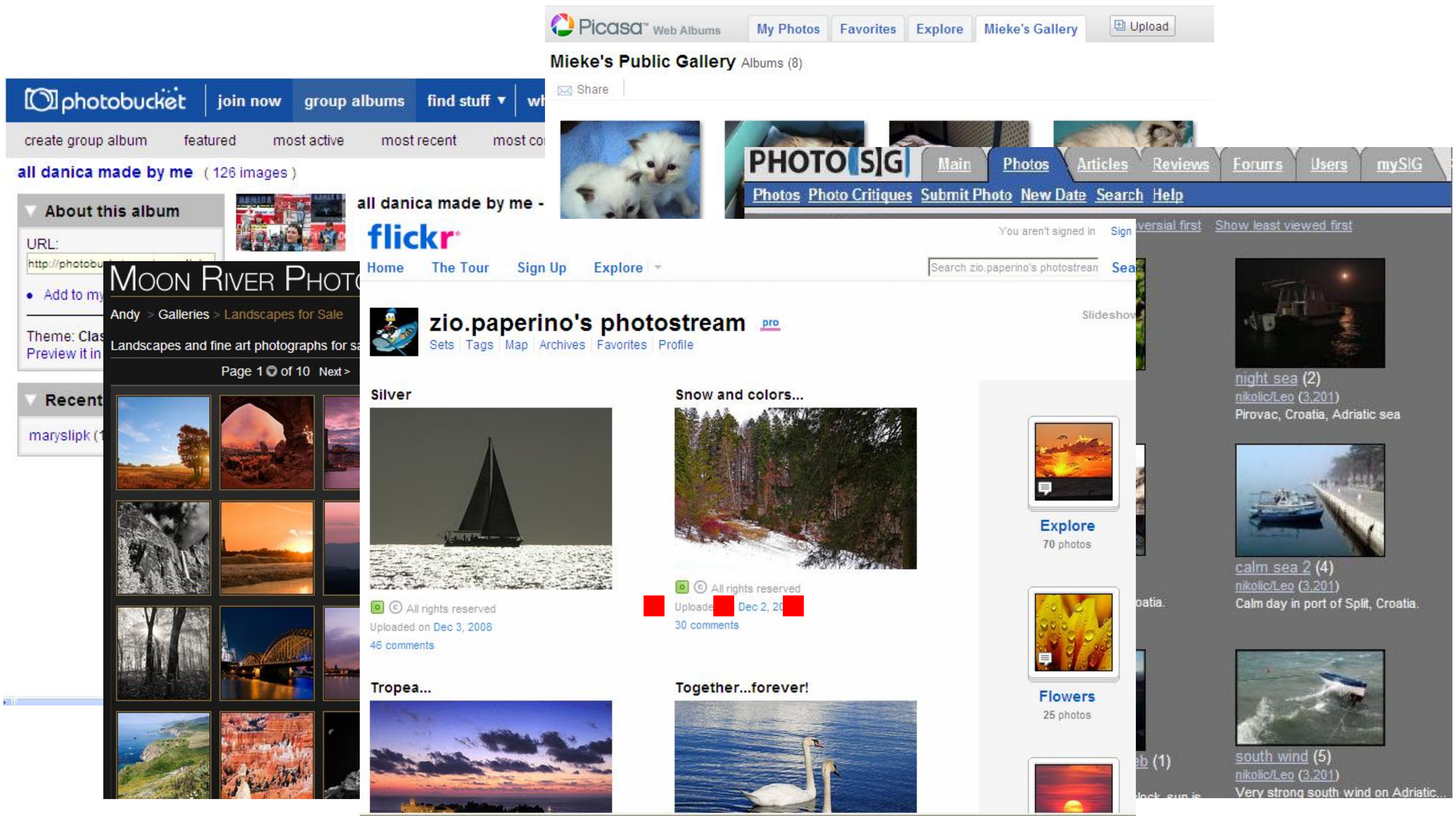


Issues in Web Image Search

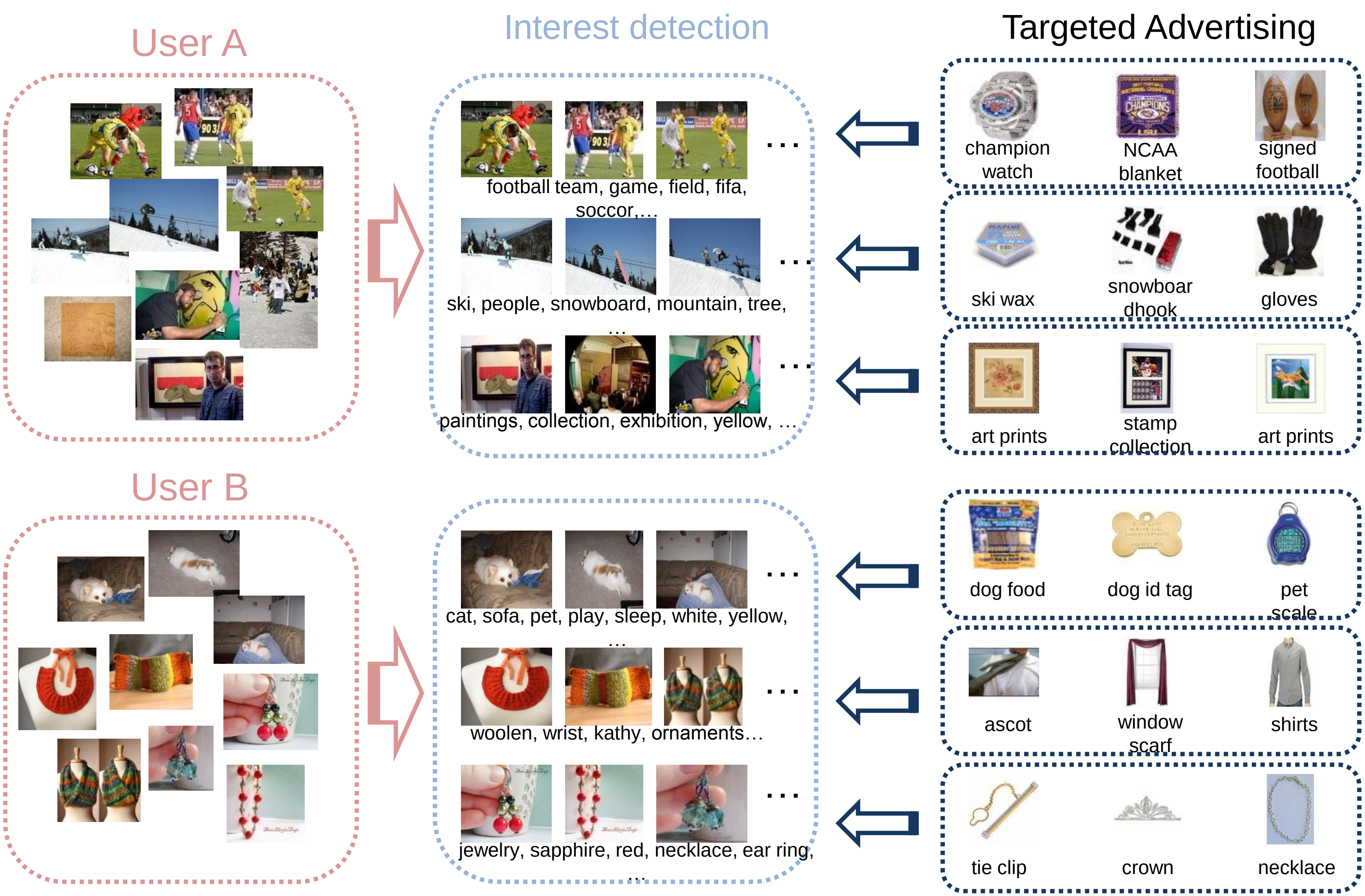
- Label collection
  - Billions of user tags
- What are useful and feasible categories ?
- Semantic gap: How to build automatic concept classifiers and text annotation models ?
  - Data driven, model less
  - Inter play between text data and visual features
- Intention gap: How to capture user intention?
- Other applications

Social Media Is Booming

- Online photo sharing attracts everyday users

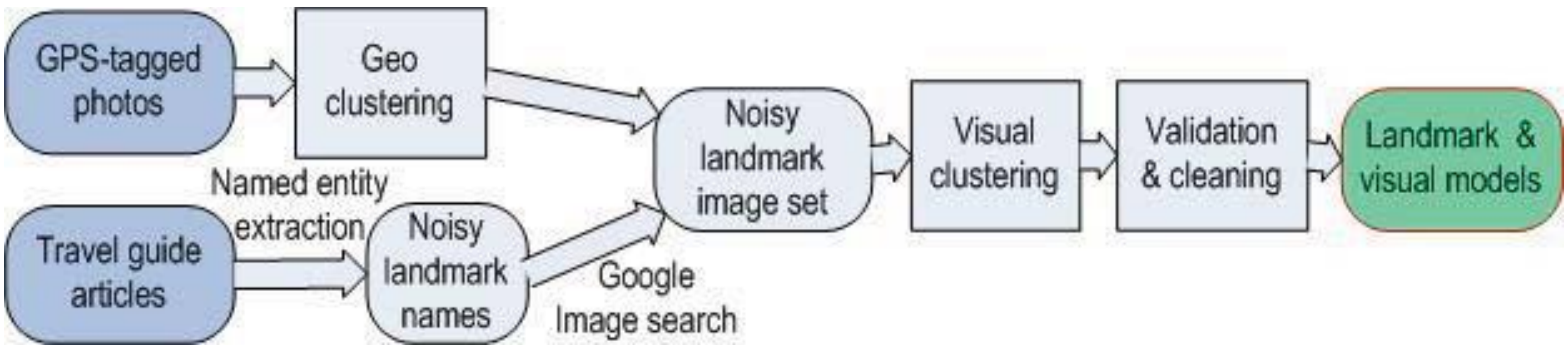


Detect User's Interests Mining based on Photo Collections



Tour the World: A web-scale landmark recognition engine

Y-T Zheng, … T Chua, CVPR09



Summary

Is Text Search Much Better Off Today?

25% of search results pointing to totally irrelevant websites；

30% of searches were given up at the end, due to non satisfying results；

35% of the users are not happy with search results;

40% of the users need to modify keywords to restart/refine a search;

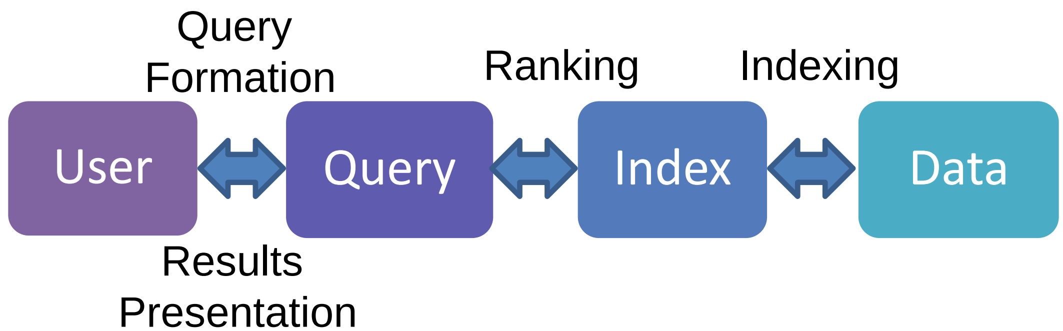
46% of the users spend more than 30 minutes in one search;

50% of the time is wasteful during search sessions;

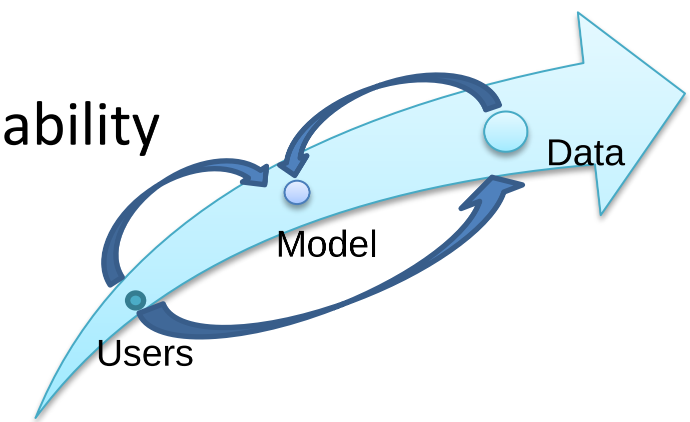
72% of users feel the result page too messy.

CBIR's golden time is just beginning!

Summary



- Web brings tons of data(metadata), billions of users, interactions and accesses
- Inter play between text data and visual features
- **Three keys:**
  - Data driven, machine learning, scalability
- New interaction model/UI
- New applications



Thanks!