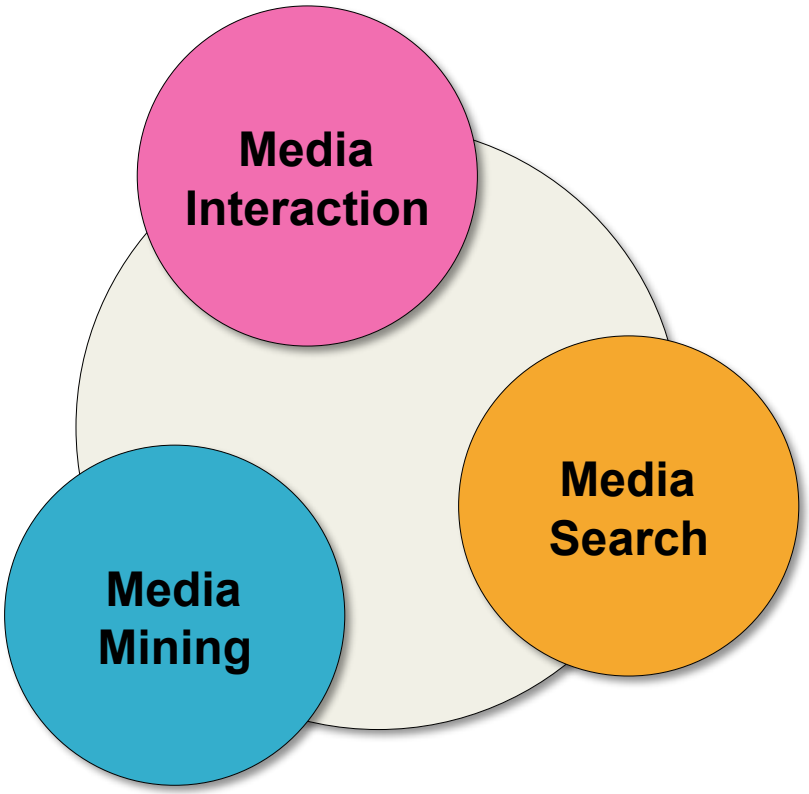
Powering NG Media Search on the Web

Roelof van Zwol

roelof@yahoo-inc.com

YAHOO!

Media Interaction



What makes Flickr successful?

Flickr: Who is looking?

Video Tag Game

What makes Flickr special?
Media Interaction

Flickr

Tags

- bird
- fly
- 50200mm
- south
- iceland
- ZD

Go South
From: Jón Ragnarsson

Yellow
From: Pelican Eyes

Tags

- green
- evening
- orange
- shadow
- nature
- yellow
- sunlight
- colors

What makes Flickr special?

1. User Generated Content

- Content not licensed from providers such as Corbis or Getty, but rather contributed by users.

sometimes it snows in april
From Joui

Mum and...
From Chrissie64

Africa Masai boy
From housden photos

slippers
From benjaminhamilton

What makes Flickr special?

2. User Organized Content

- Content is tagged, described, organized, discovered, etc. not by “editors” but by the users themselves.

What makes Flickr special?

3. User Distributed Content

- Flickr achieved distribution across the internet, not through “business deals” per se, but rather through the Flickr community which distributed Flickr content on 3rd-party blogs.

What makes Flickr special?

4. User Developed Functionality

- Flickr exposed APIs (PHP, Perl, etc.) that allowed the community of developers to build against the Flickr platform.

What makes Flickr special?

1. User Generated Content

- Content not licensed from providers such as Corbis or Getty, but rather contributed by users.

2. User Organized Content

- Content is tagged, described, organized, discovered, etc. not by “editors” but by the users themselves.

3. User Distributed Content

- Flickr achieved distribution across the internet, not through “business deals” per se, but rather through the Flickr community which distributed Flickr content on 3rd-party blogs.

4. User Developed Functionality

- Flickr exposed APIs (PHP, Perl, etc.) that allowed the community of developers to build against the Flickr platform.

Entire ecosystem created by less than ten employees…
aided by millions in the Flickr community.

Flickr – tag cloud

Time

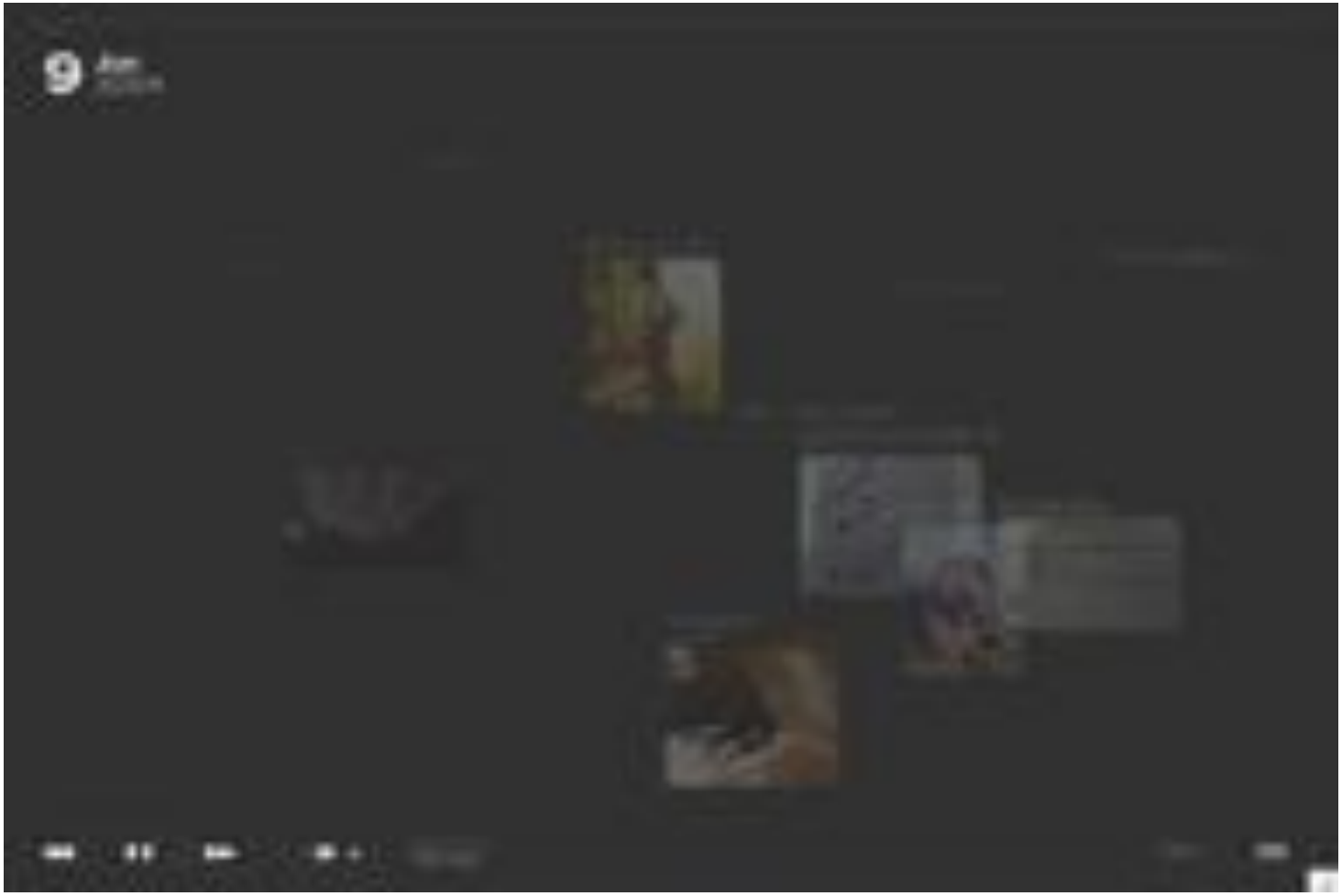
and

Space

All time most popular tags

amsterdam animal animals april architecture art australia baby barcelona beach berlin bird birthday black blackandwhite blue boston bridge building bw california cameraphone camping canada car cat cats chicago china christmas church city clouds color colorado concert day dc dog dogs england europe family festival fireworks florida flower flowers food france friends fun garden geotagged germany girl graduation graffiti green hawaii holiday home honeymoon house india ireland italy japan july june kids lake landscape light london losangeles macro march may me mexico moblog mountains museum music nature new newyork newyorkcity newzealand night nyc ocean orange oregon paris park party people phone photo pink portrait red reflection river roadtrip rock rome sanfrancisco school scotland sea seattle sign sky snow spain spring street summer sun sunset taiwan texas thailand tokyo toronto travel tree trees trip uk unfound urban usa vacation vancouver washington water wedding white winter yellow zoo

Flickr – tag lines



Flickr – tag lines

What makes an object “Interesting” for an interval?

- Occurs more frequently in the interval, and less frequently outside
- Don’t be fooled by sparse occurrences (1 versus 0)
- Proposed model: simple approach based on tf-idf:

$$\frac{\textbf{Occurrences (per second) during the interval}}{\textbf{epsilon + Occurrences (per second) overall}}$$

Goal:

- Return top K most representative objects (for a specific time frame)

Flickr – zone tag

Flickr: Who is looking?
Media Interaction

Roelof van Zwol,
WI'07

About Flickr

On-line photo sharing service

> 3.5 Billion photos uploaded

> 12 Million Web-users registered

> 4000 photos uploaded per minute

> 12.000 photos served per second, at peak times (August 2007)

Who is looking?

A characterization of usage behavior on Flickr, with focus on:

- When? -- *Temporal characteristics*
- Who? -- *Social*
- Where? -- *Spatial*

Not about:

- Why? or What?
  - Social incentives
    - G.W. Furnas et al. “Why do tagging systems work?”
    - C. Marlow et al. “Ht06, tagging paper, taxonomy, Flickr academic article, to read”
    - M. Ames and M. Naaman. “Why we tag: motivations for annotation in mobile and online media”

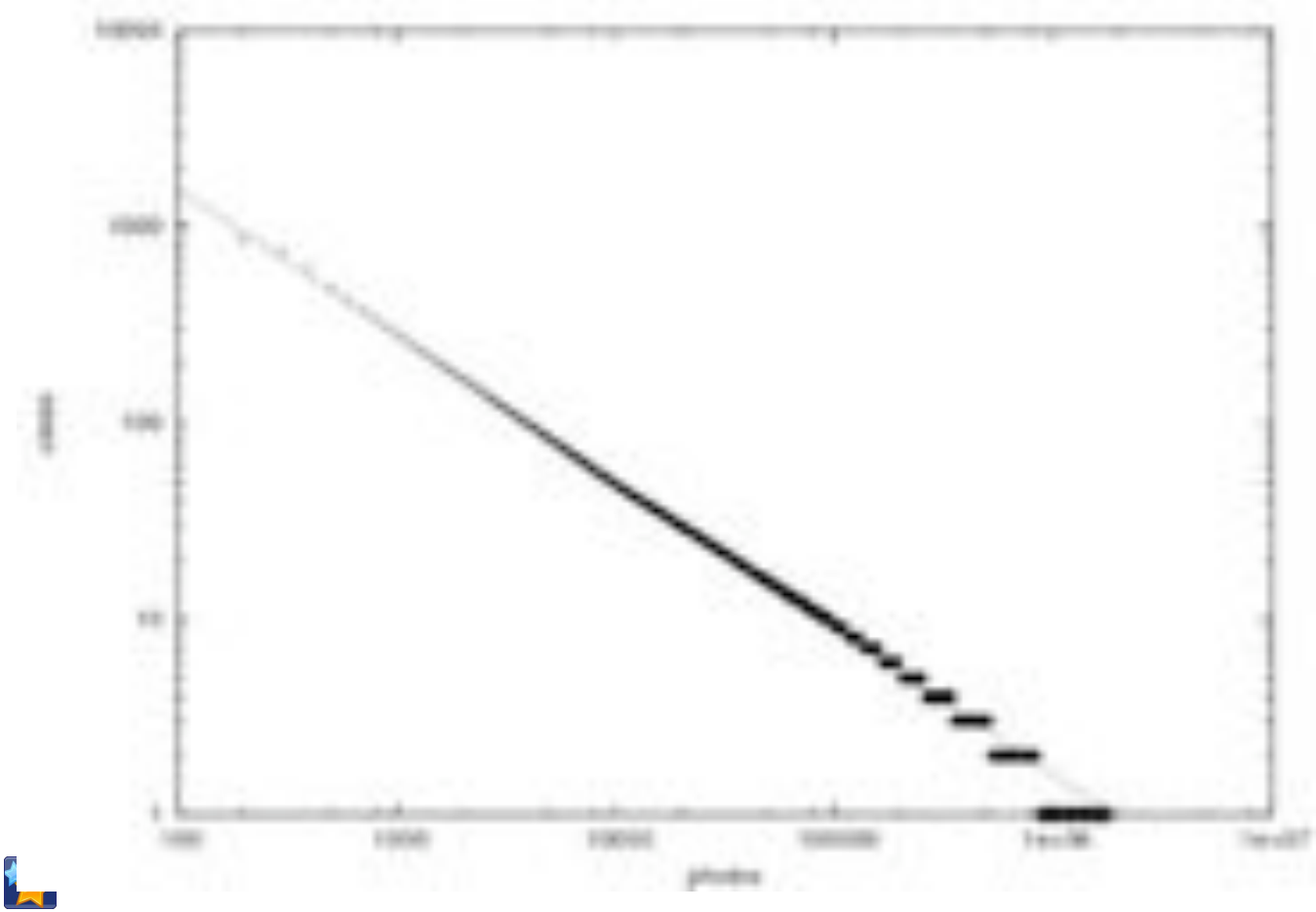
Data Collection

Analysis is based on:

- HTTP access logs of Flickr, spanning a 60 day period
  - 1.83 Million public photos
    - uploaded in the first 10 days
    - and their views in the consecutive 50 days
  - limited to the detailed photo views on Flickr:
    - .*flickr.com/photos/<owner id>/<photo id>/?.*
- Data collected through public Flickr API:
  - flickr.photos.getInfo
  - flickr.photos.getAllContexts
  - flickr.contacts.getPublicList
- Mapping service from IP to long/lat coordinates

Characterisation of Photo Views

- 1.83 million photos; 6.72 million views
- Power law - the probability of having x visits is proportional to $x^{-0.7}$



Characterisation of Photo Views

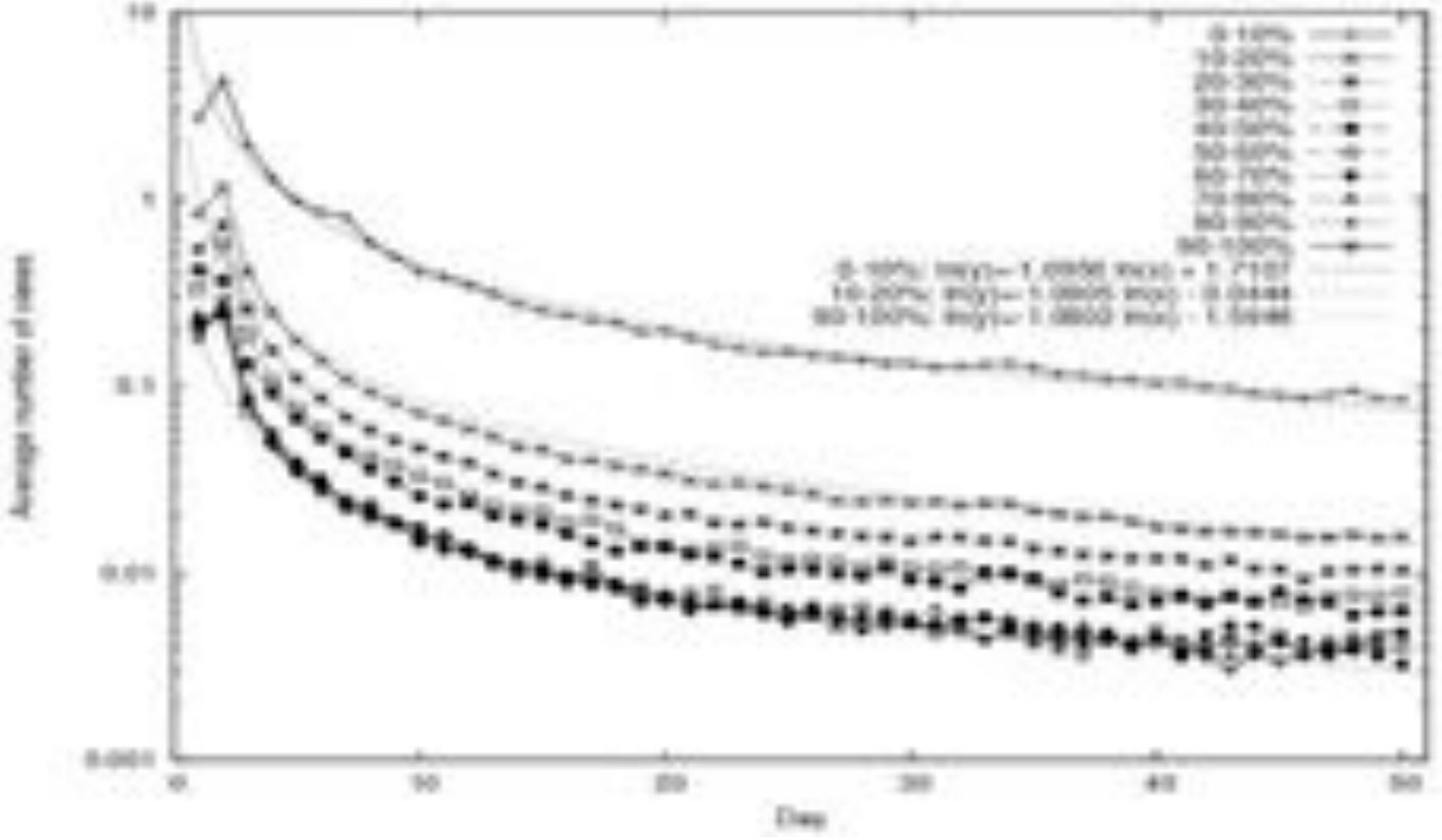
- Dividing the collection into equal slices, based on the number of photos
- Where slice 0-10% contains the top 10% most frequently viewed photos
- Emphasize on the skewedness of the distribution of photo views: 0-10% slice already covers >50% of all views

slice	views
0-10%	3,802,875
10-20%	812,131
20-30%	515,532
30-40%	365,712
40-50%	312,270
50-60%	182,856
60-70%	182,857
70-80%	182,856
80-90%	182,857
90-100%	182,857

Characterization of Photo Views

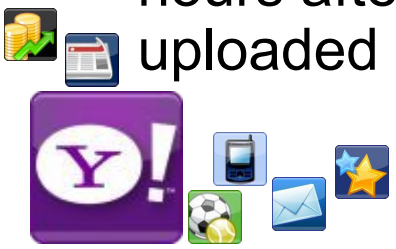
The average number of photo views per day for the slices over a 50 day period.

- The declining trend followed by each of the slices can be modeled by an exponential decay
- The number of views on day x after being uploaded is proportional to e−1.1x



Characterisation of Photo Views

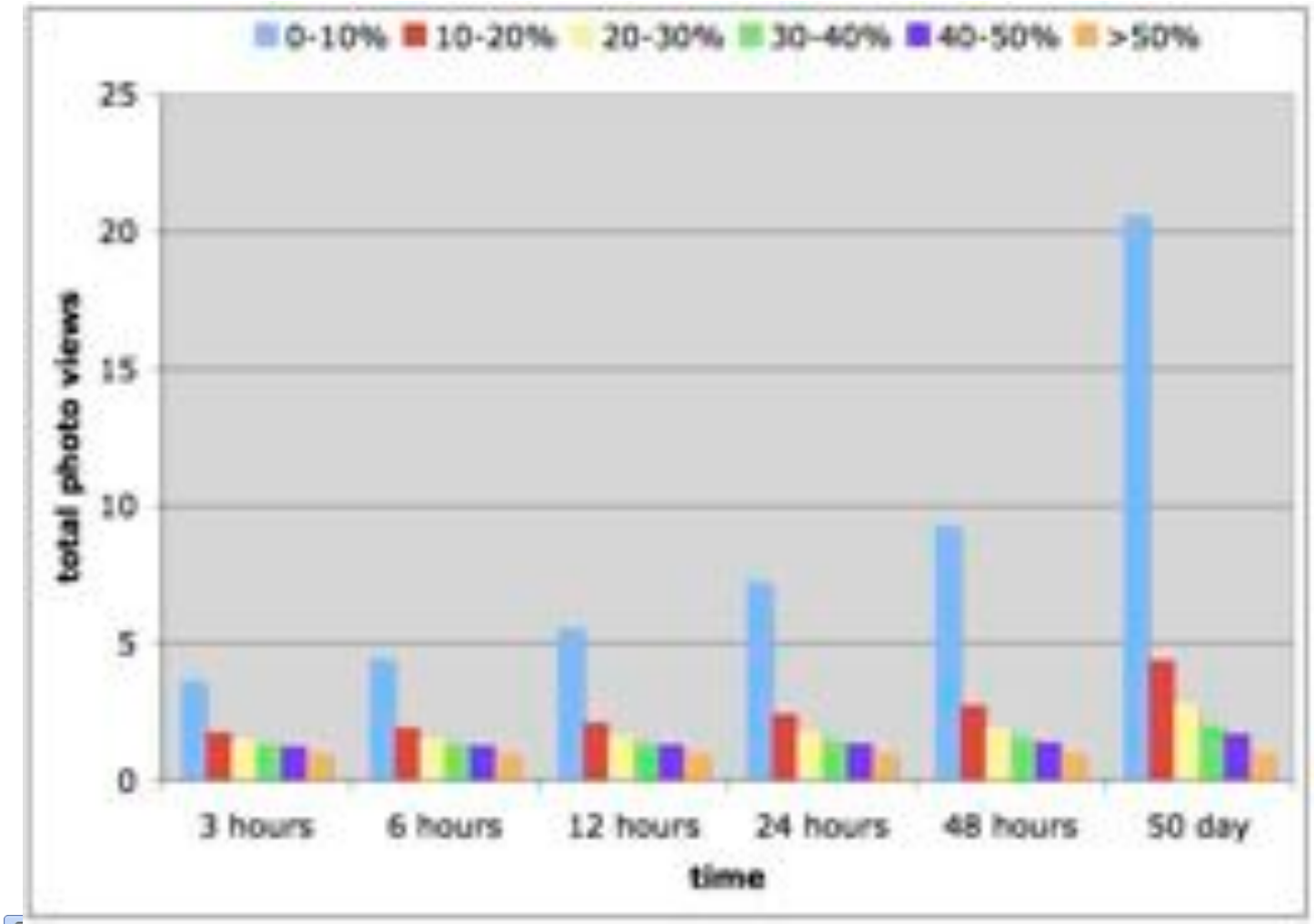
- Focus on first 48 hours
- Shows similar behaviour for different trends (slices)
- After 48 hours, a photo already received ~50% of the total number of views it will receive after 50 days
- Moreover, popular photos are already discovered within 3 hours after being uploaded



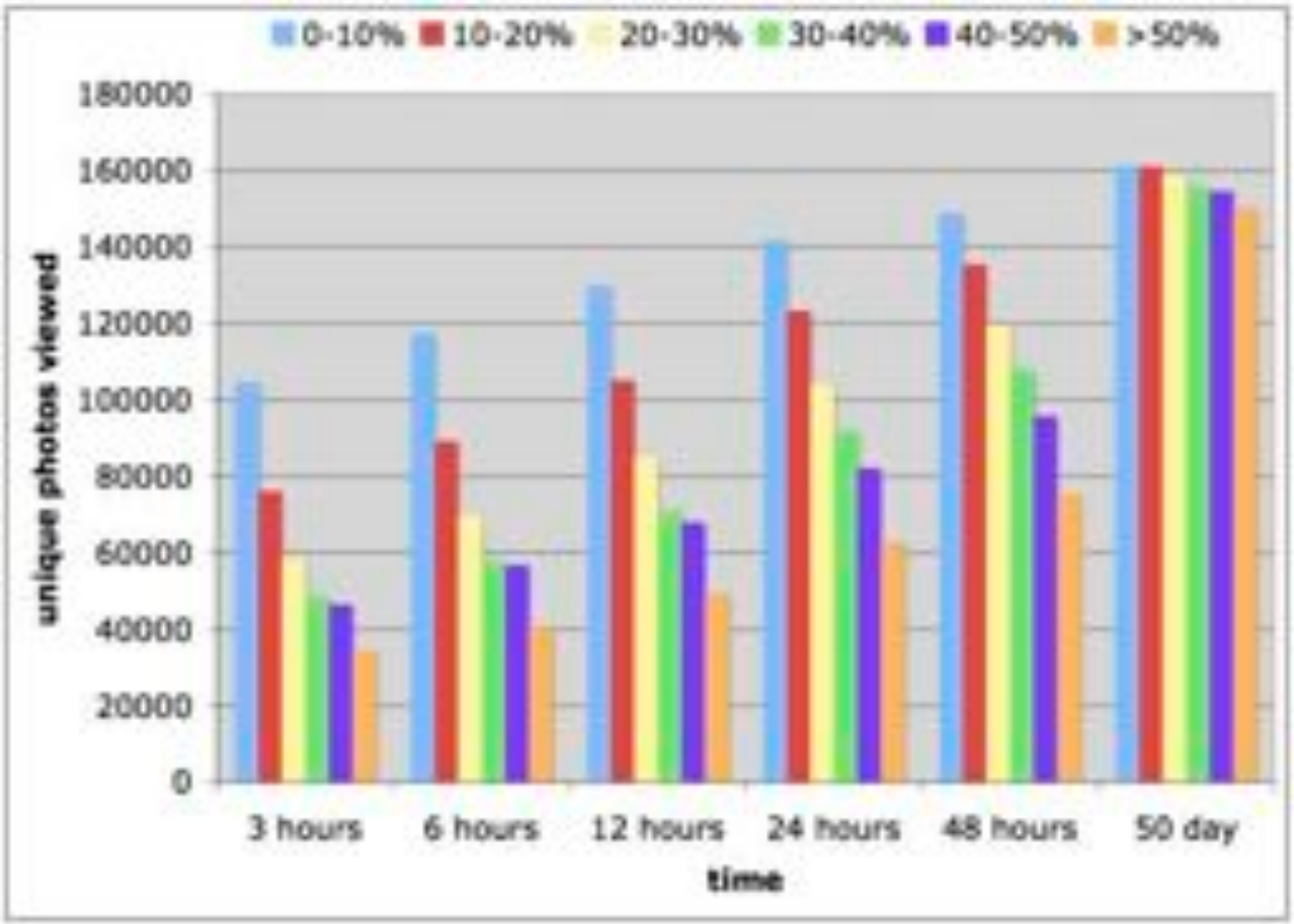
	3 hours		6 hours		12 hours	
slice	avg	std	avg	std	avg	std
0-10%	3.63	8.25	4.44	12.51	5.55	18.66
10-20%	1.77	0.97	1.92	1.05	2.11	1.12
20-30%	1.47	0.67	1.54	0.7	1.62	0.73
30-40%	1.3	0.46	1.33	0.47	1.36	0.48
40-50%	1.25	0.43	1.27	0.44	1.3	0.46
>50%	1	0	1	0	1	0

	24 hours		48 hours		50 day	
slice	avg	std	avg	std	avg	std
0-10%	7.24	26.5	9.28	37.6	20.6	87.7
10-20%	2.43	1.22	2.75	1.28	4.4	0.7
20-30%	1.77	0.77	1.92	0.79	2.8	0.4
30-40%	1.44	0.5	1.52	0.5	2	0
40-50%	1.35	0.48	1.39	0.49	1.7	0.45
>50%	1	0	1	0	1	0

Characterisation of Photo Views



Characterisation of Photo Views



Applications

What can you do with this knowledge?

- **Predict the popularity** of a photo (using temporal, and social indicators)
- **Develop caching strategies** for frequently viewed media content
- Develop a hybrid model for serving multimedia content that implements a **P2P** storage strategy for in-frequently viewed content, in combination with a **content distribution network** for serving popular media content

by: thepres6

Video Tag Game
Media Interaction

Roelof van Zwol,
Lluis Garcia,
Borkur Sigurbjornsson,
Georgina Ramirez
WWW'08

About & Motivation

About

- Time-based annotation of streaming video, in a multi-player game

Motivation

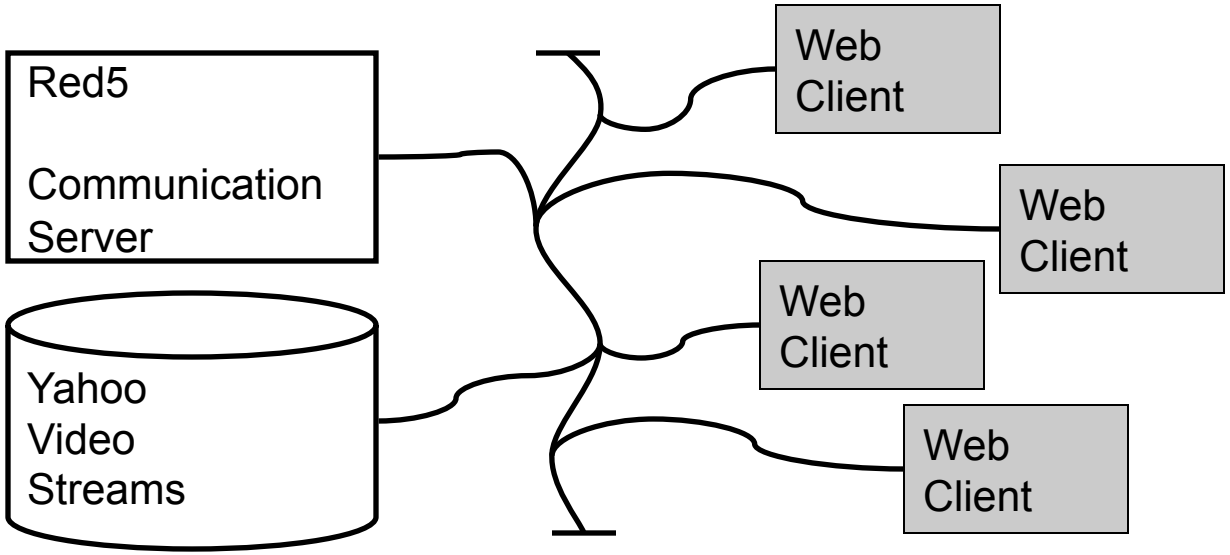
- To collect dense, time-based annotations of video
- Investigate users accuracy when tagging streamed video
- Enable retrieval of video-fragments

How?

Set-up

- In a multi-player game setting
- Tagging of streaming video
- Temporal scoring mechanism, that rewards tag-agreement between users

Architecture



Video Tag Game

Temporal Scoring Mechanism

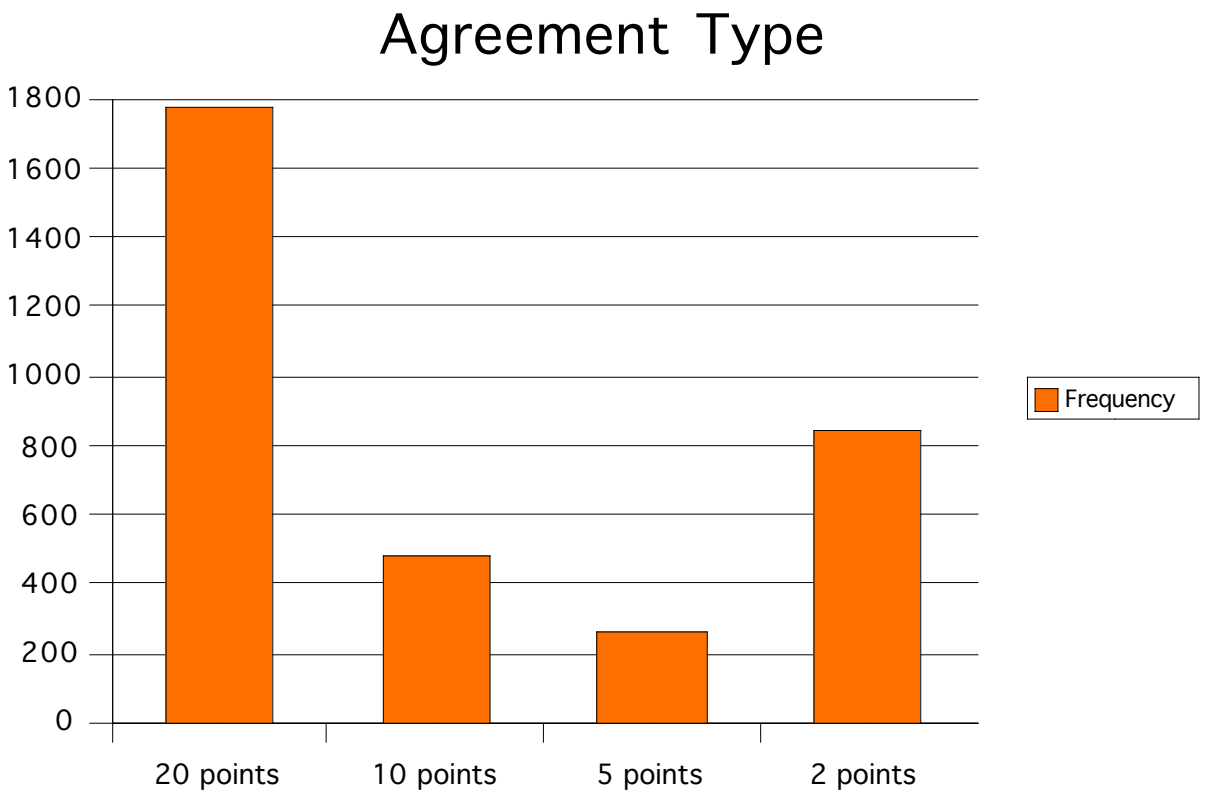
- If two players agree on a tag, the players get points
- More points should be rewarded for a tag if the difference in time between two players, submitting that tag, is smaller
- Entering the same tag twice within a short period of time should not be rewarded (for that user, others can however benefit)

Video Tag Game

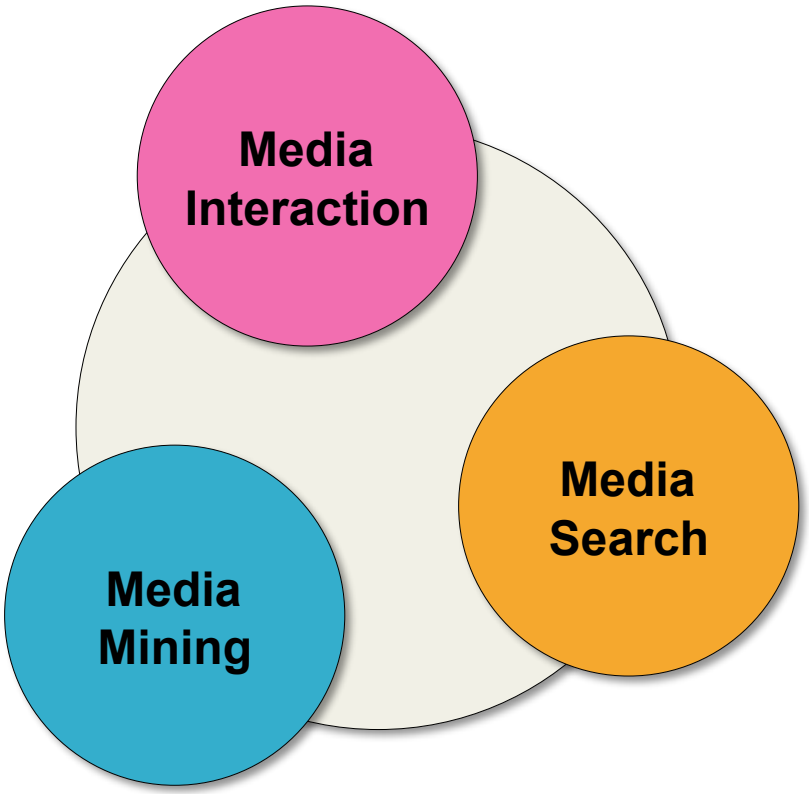
Video Tag Game

SIGIR demo:

- 27 games / 59 players / 5890 tags
- 0.57 agreement (3360 scoring tags), on avg. 12.88 points per agreed tag:



Media Mining



Classifying tags

Collective and personalized tag recommendation

Tag explorer

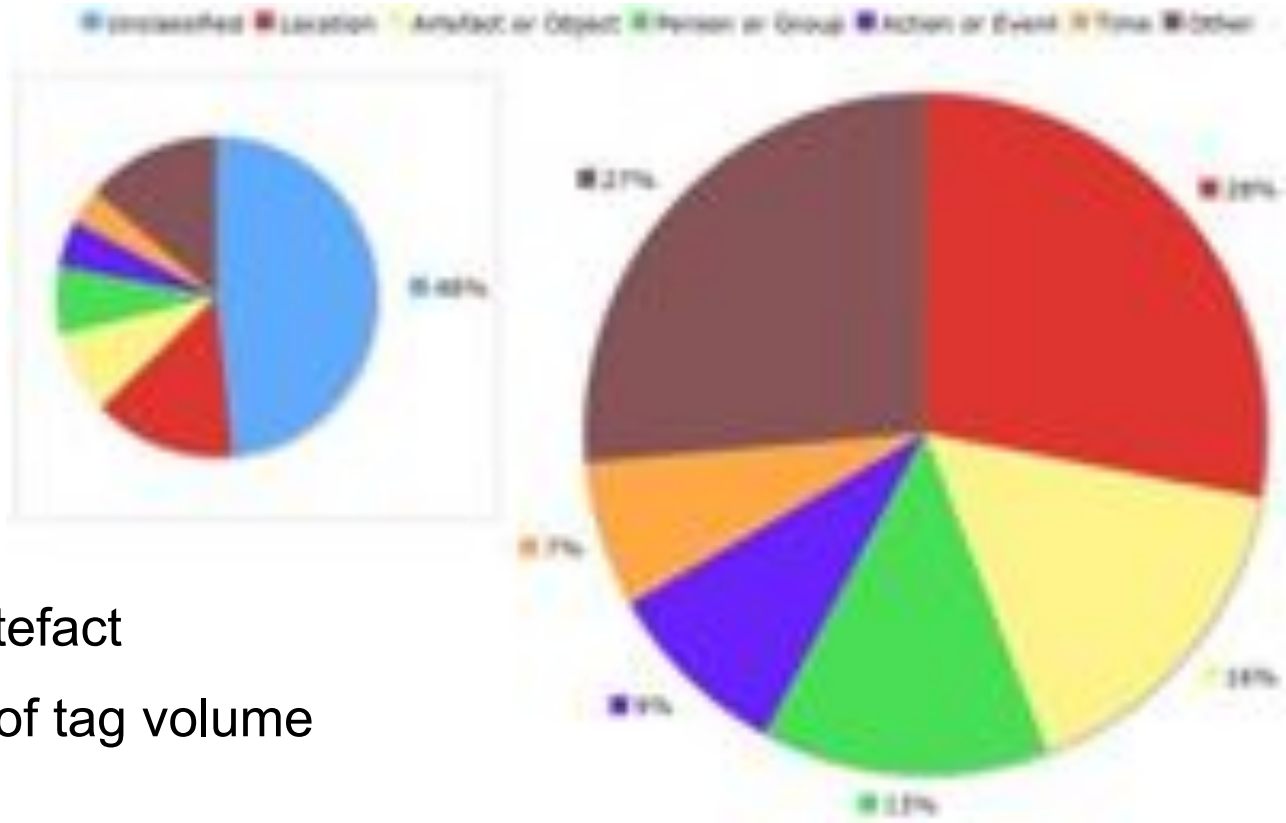
Placing Flickr Images on a Map

Classifying Tags
Media Mining

Simon Overell
Borkur Sigurbjornsson
Roelof van Zwol
WSDM'09

Syntactic Classification

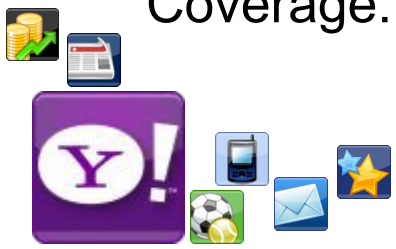
- Objective: syntactic classification of tags using open source content (WordNet, Wikipedia, ODP, etc.)
- Assign tag semantics using WordNet broad categories



Paris :: location

Eiffel Tower :: artefact

Coverage: 52% of tag volume



How…

To extend coverage of syntactic classification?

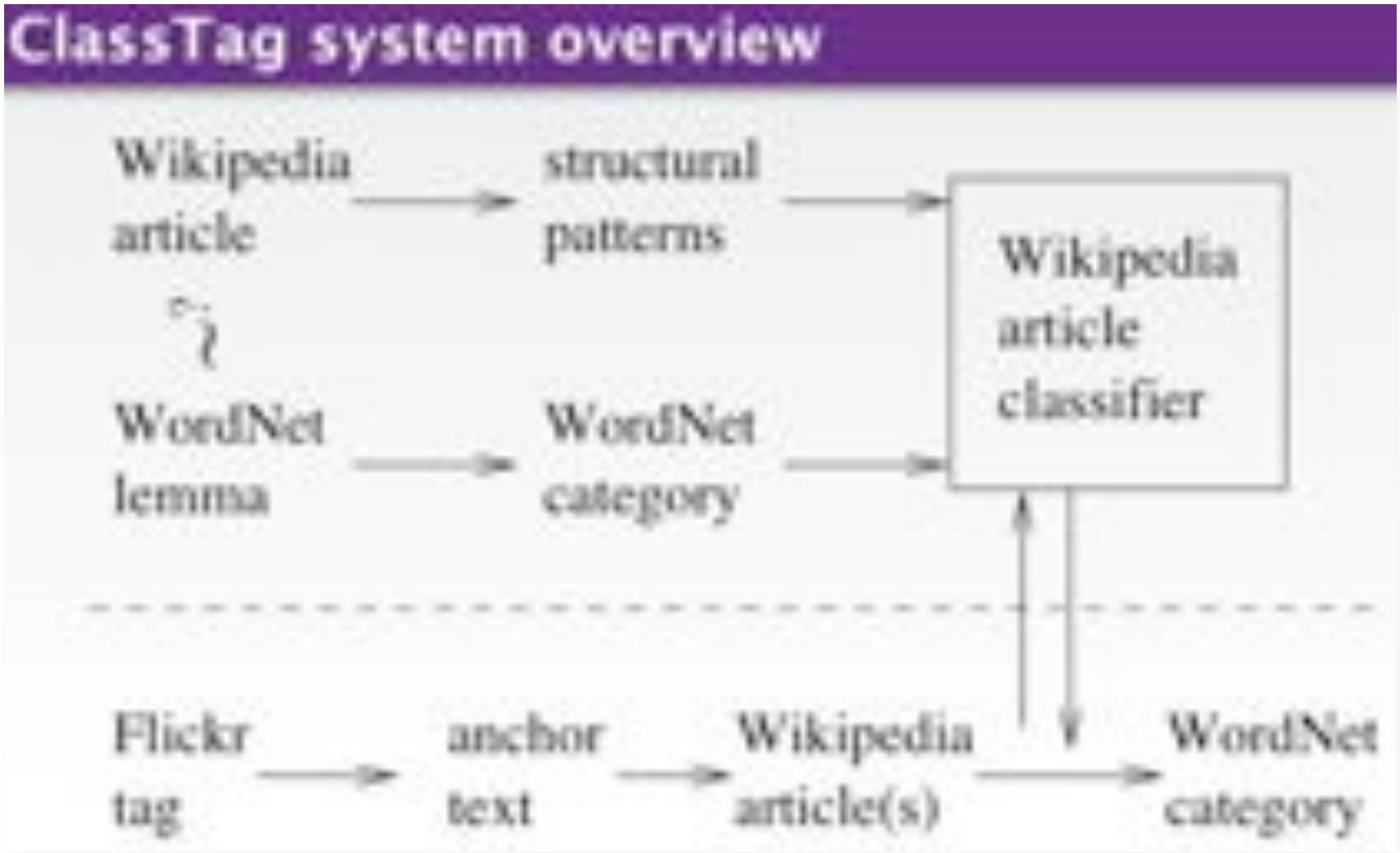
- Based on classification of Wikipedia pages
- Mapping from tags to classified Wikipedia pages
- Upperbound for coverage: 78.6% of the tag volume

To classify Wikipedia pages?

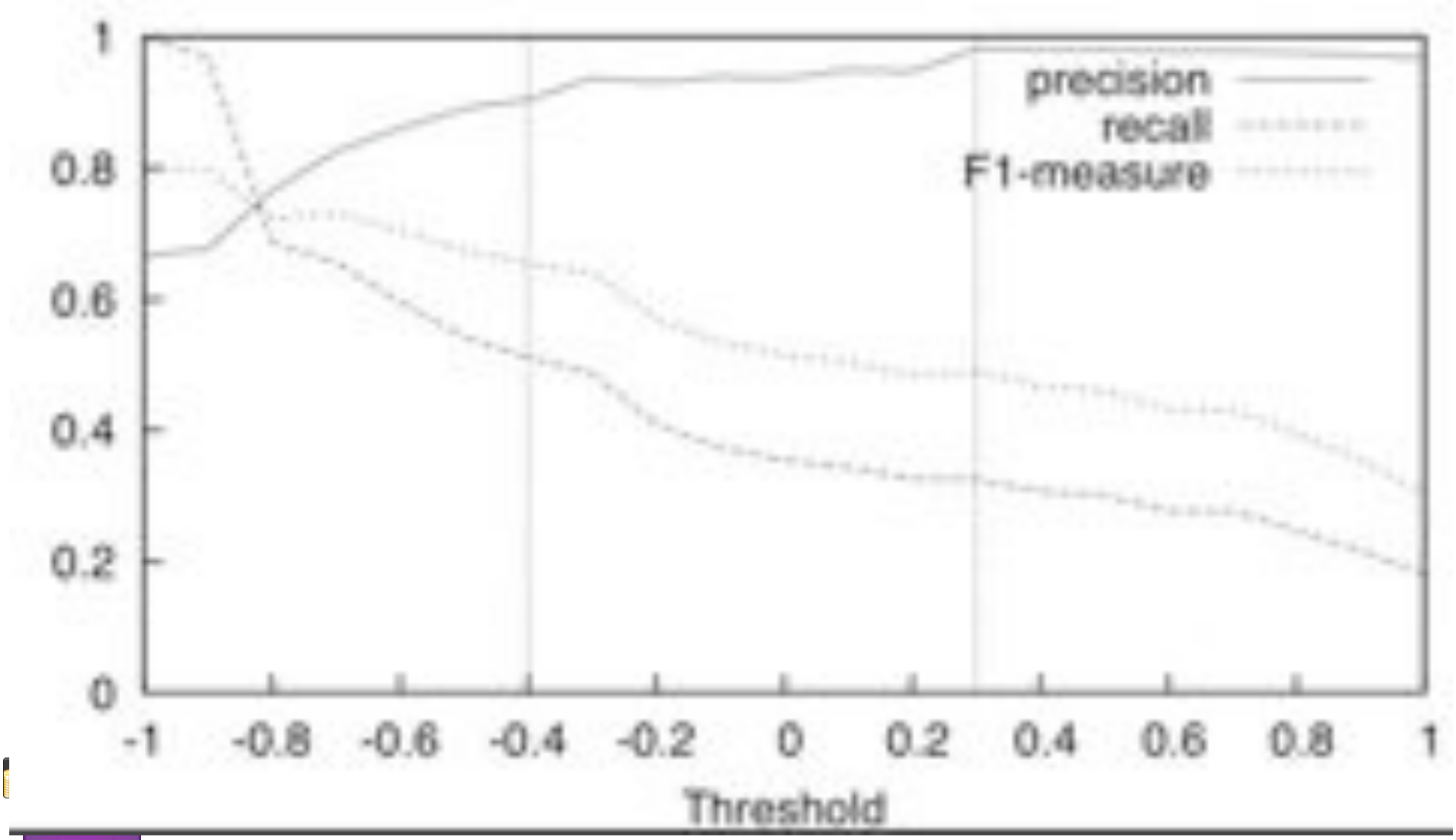
- Use structural patterns found in Wikipedia pages
  - templates and categories
- Achieved extended coverage: 68% of the tag volume

Example

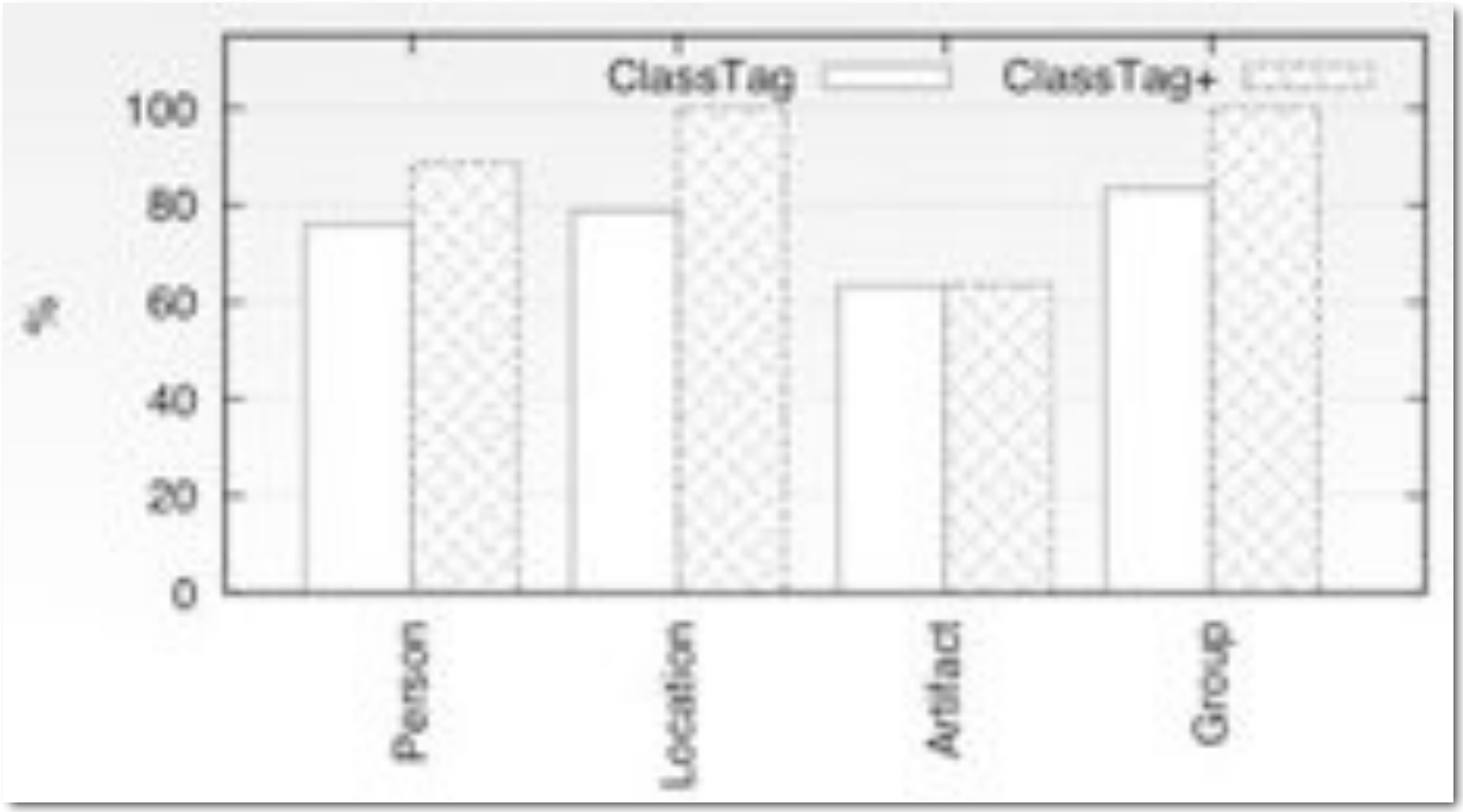
System


ClassTag system overview
Wikipedia article
structural patterns
Wikipedia article classifier
WordNet lemma
WordNet category
Flickr tag
anchor text
Wikipedia article(s)
WordNet category

Performance



TagClass


ClassTag
ClassTag+
100
80
60
40
20
0
Person
Location
Artifact
Group

REST-API

<tagclass tag="iwo jima”>

<classification source="wordnet" class="location” instanceof="island” rank="1" />

<classification source="wordnet" class="act" instanceof="amphibious assault” rank=”2"/>

<classification source="wikipedia" class="location" rank="1" support="0.80"/>

<classification source="wikipedia" class="act" rank="2" support="0.10"/>

<classification source="wikipedia" class="artifact" rank="3" support="0.07"/>

</tagclass>

<tagclass tag="bigapple" >

<classification source="wikipedia" class="location" rank="1” support="0.79"/>

<classification source="wikipedia" class="act" rank="2” support="0.20"/>

</tagclass>

Collective Tag Recommendation
Media Mining

Borkur Sigurbjornsson
Roelof van Zwol
WWW'08

Motivation

I went to Barcelona, took a photo, tagged it:

- “Sagrada Familia”

2 years later I want to find the photo:

- query: *church Barcelona Gaudí*
- no pictures found

Task:

- Help users to provide rich annotations

Flickr Annotations

Characteristics:

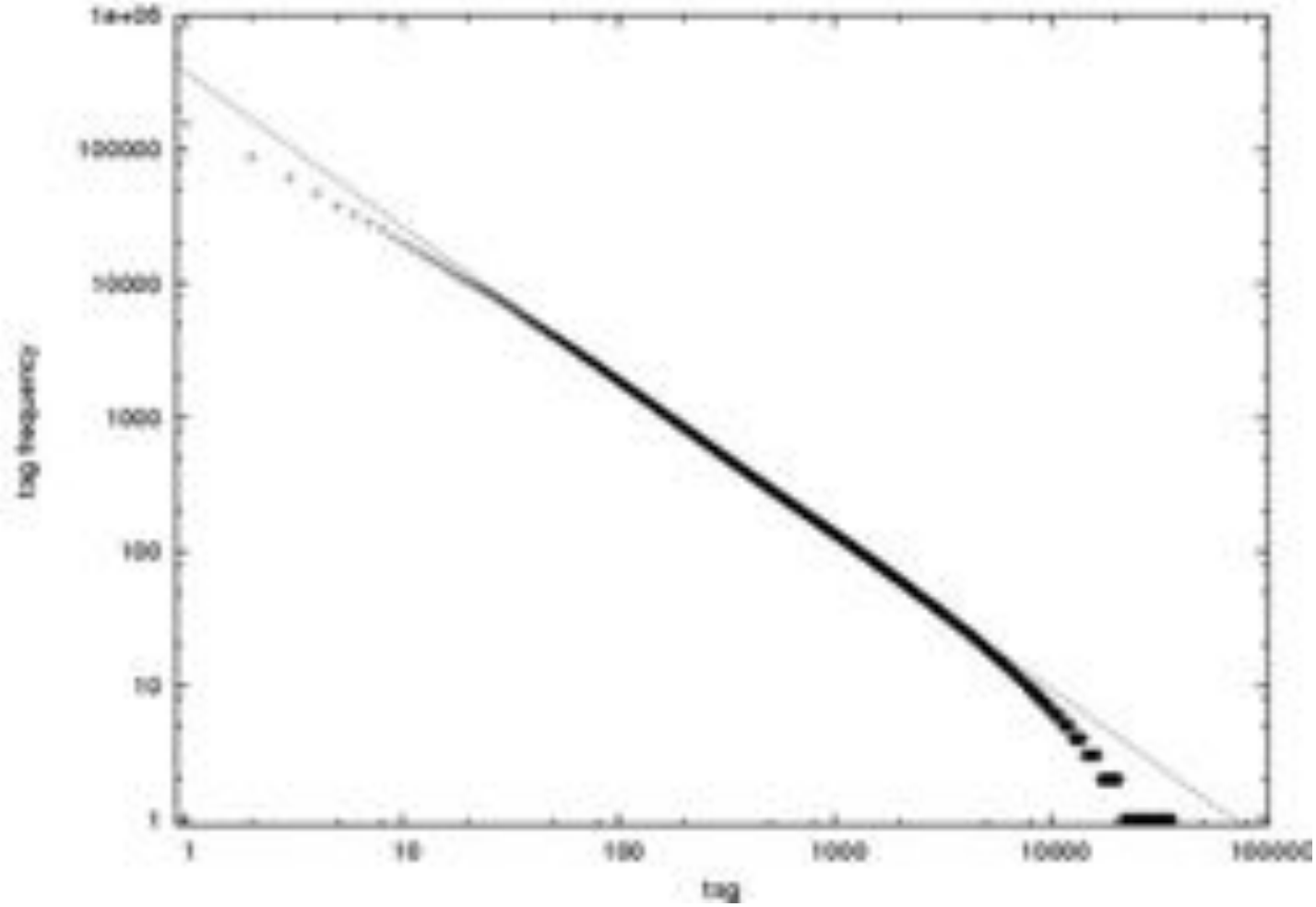
- Most photos have few tags
- Few photos have many tags

Tags per photo	Percentage of photos[1]
1	30%
2-3	34%
4-6	23%
> 6	13%

[1] based on a random sample of 100 million tagged Flickr photos

Flickr Tag Frequency

- Few tags are used to describe many photos
- Most tags are used to describe few photos



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 Co-occurrence Statistics

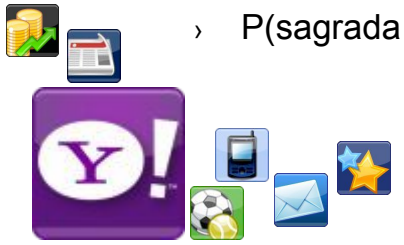
Input: A snapshot of 500M public photos on Flickr, with annotations

Approach is based on probabilistic framework

- Assume an photo is labelled with a set of tags T = {ta, tb , ...}
- Define I (T) as the number of photos that contain the tag set T
- For any pair of tags ti , tj , we denote the number of image co-occurrences by I (ti∩tj)
- Estimate the probability that a tag, ti , appears in presence of tag tj , by calculating:

$$p(t_i|t_j) = \frac{I(t_i \cap t_j)}{\sum_k I(t_k \cap t_j)}$$

- **Examples:**
  - P(barcelona|sagradafamilia) = 0.46
  - P(sagrada familia|gaudi) = 0.14



Tag Co-occurrence Statistics

- Probabilistic framework cont'd:
  - Estimate the probability that any one tag is used on an image by:

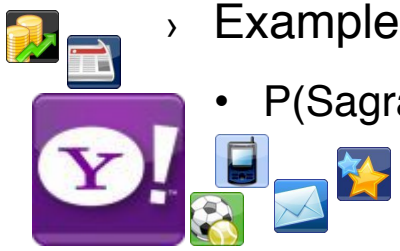
$$p(t_i) = \frac{\sum_j I(t_i \cap t_j)}{\sum_{j,k} I(t_k \cap t_j)}$$

  - Objective is to calculate the probability of a tag in any context, e.g. a set of tags T:

$$p(T|t_i) = \prod_{t \in T} p(t|t_i)$$

$$p(t_i|T) = \frac{p(T|t_i)p(t_i)}{p(T)} = \frac{p(t_i)\prod_{t \in T} p(t|t_i)}{\sum_j p(t_j)\prod_{t \in T} p(t|t_j)}$$

  - Example:
    - P(Sagrada Familia | {church, Barcelona})=0.67



Tag Recommendation System

- Task: Given a partially annotated photo, recommend additional annotations
- Approach: Use the aggregated annotation term co-occurrence

Summary

Tagging is sparse but diverse

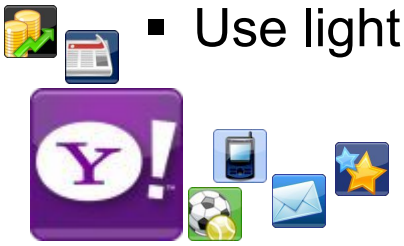
- Few tags per photo
- Tag frequency distribution follows a power law

Use the collective knowledge to recommend tags

- For 68% of photos our first suggestion is good
- For 94% of photos we provide a good suggestion among top 5
- For top 5 suggestions, 54% are good

Future work

- Use additional data sources (User profile, social contacts)
  - TagSuggest 2.0P
- Use light weight image features



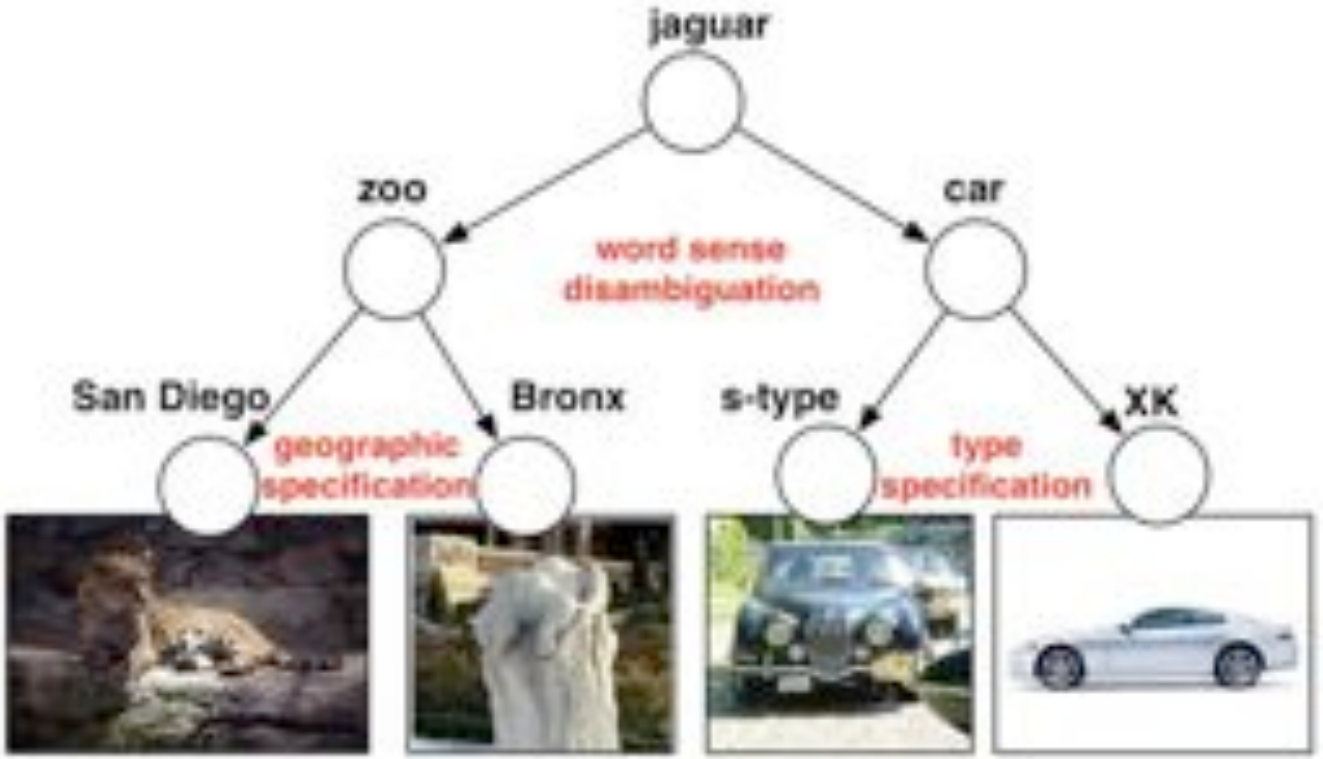
Resolving Tag Ambiguity
Media Mining

Killian Weinberger,
Malcolm Slaney,
Roelof van Zwol
ACM MM'08

Resolving Tag Ambiguity

The objective of this research is to determine when additional tags are needed. Two scenarios:

- A tag set has an ambiguous meaning
- The tag set is not sufficiently specific

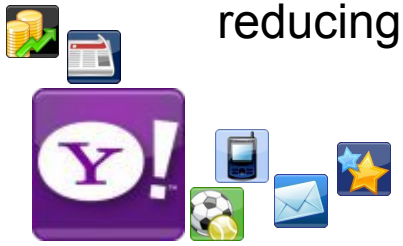


Resolving Tag Ambiguity

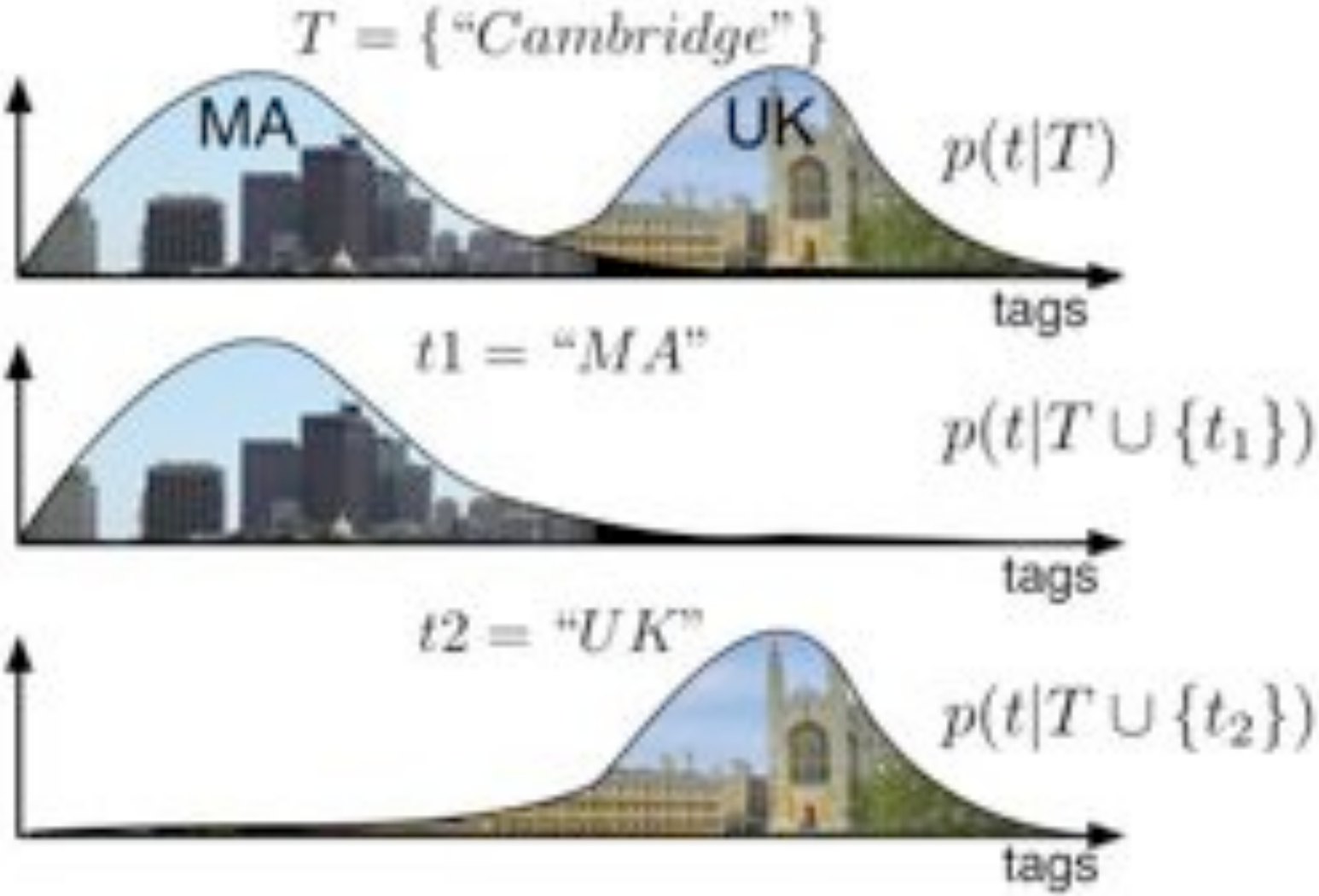
- Two contributions:
  1. A statistical approach is proposed to measure the ambiguity of a tag set, and the user is only interrupted, when the ambiguity score is above a certain threshold
  2. The method introduces pair wise disambiguation to recommends two tags that would reduce the ambiguity of the existing tag set the most

Resolving Tag Ambiguity

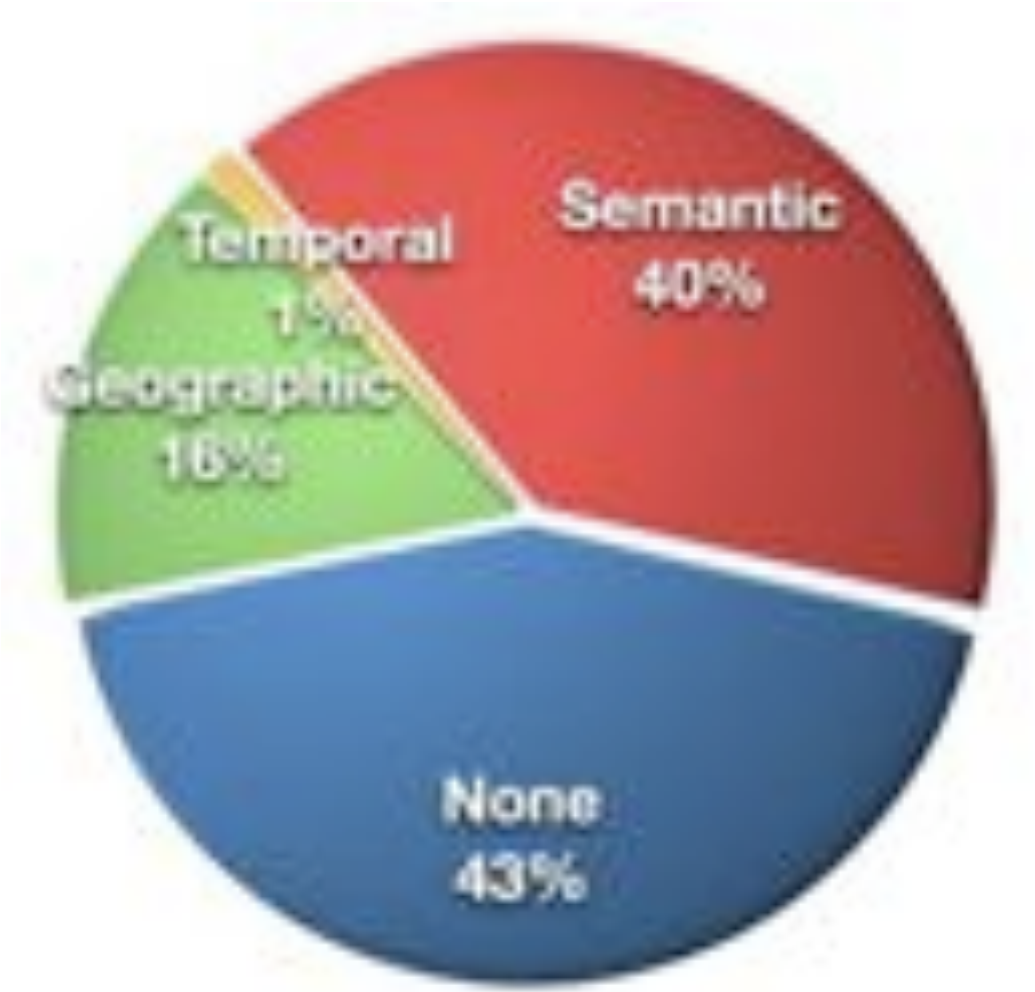
- Intuition:
  - A tag set is ambiguous if it can appear in two different tag contexts
    - Geographic locations, time-based events, languages, topical, social, or any combination of the mentioned contexts ("Java": location, programming language, coffee, etc.)
  - Example: "Cambridge"
    - Considered ambiguous, based on spatial context
    - Tag suggestions: "Massachusetts" or "United Kingdom"
    - Alternative tag suggestion "university" is highly relevant, but will not resolve the ambiguity.
- Approach:
  - Extends the probabilistic framework of TagSuggest, and uses a *weighted KL divergence* for detecting pairs of tags that have the largest impact on reducing the ambiguity



Resolving Tag Ambiguity



Results



Tag Explorer
Media Mining

Borkur Sigurbjornsson
Roelof van Zwol

TagExplorer

A prototype for browsing Flickr photos

Provides query refinement for …

- … drilling in on 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

http://sandbox.yahoo.com/TagExplorer

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


TagExplorer
YAHOO!
RESEARCH

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

Interaction Options

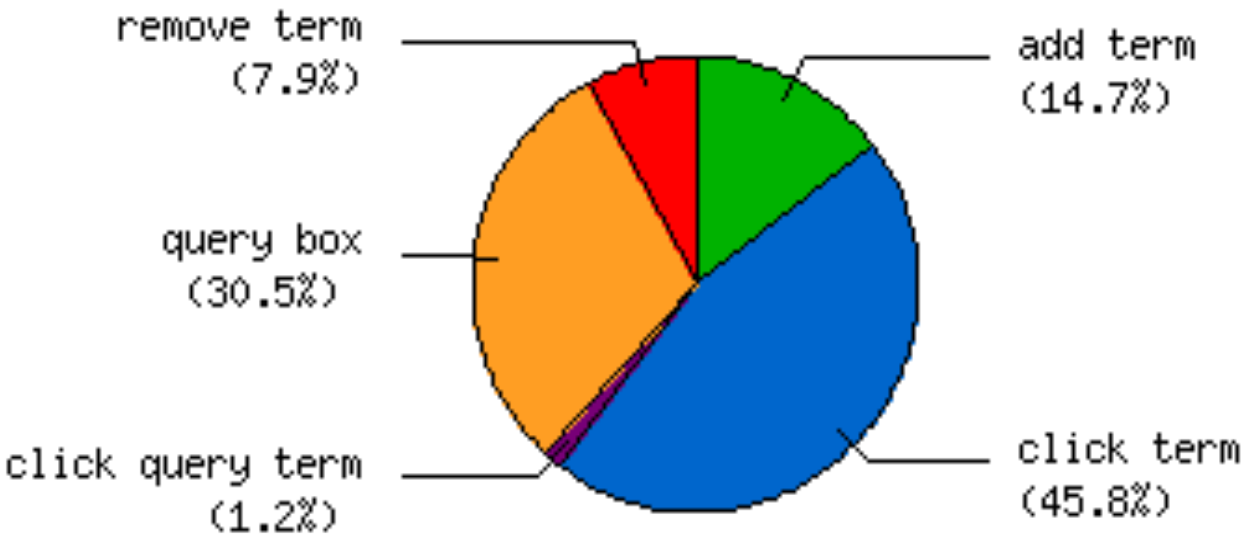
- Given the query paris
  - Clicking tag: museum+
    - New query: museum
  - Add tag: museum+
    - New query: paris museum
- Given the query museum paris
  - Remove tag: museumX parisX
    - New query: museum

User Interaction

- Based on 1500+ interaction sessions

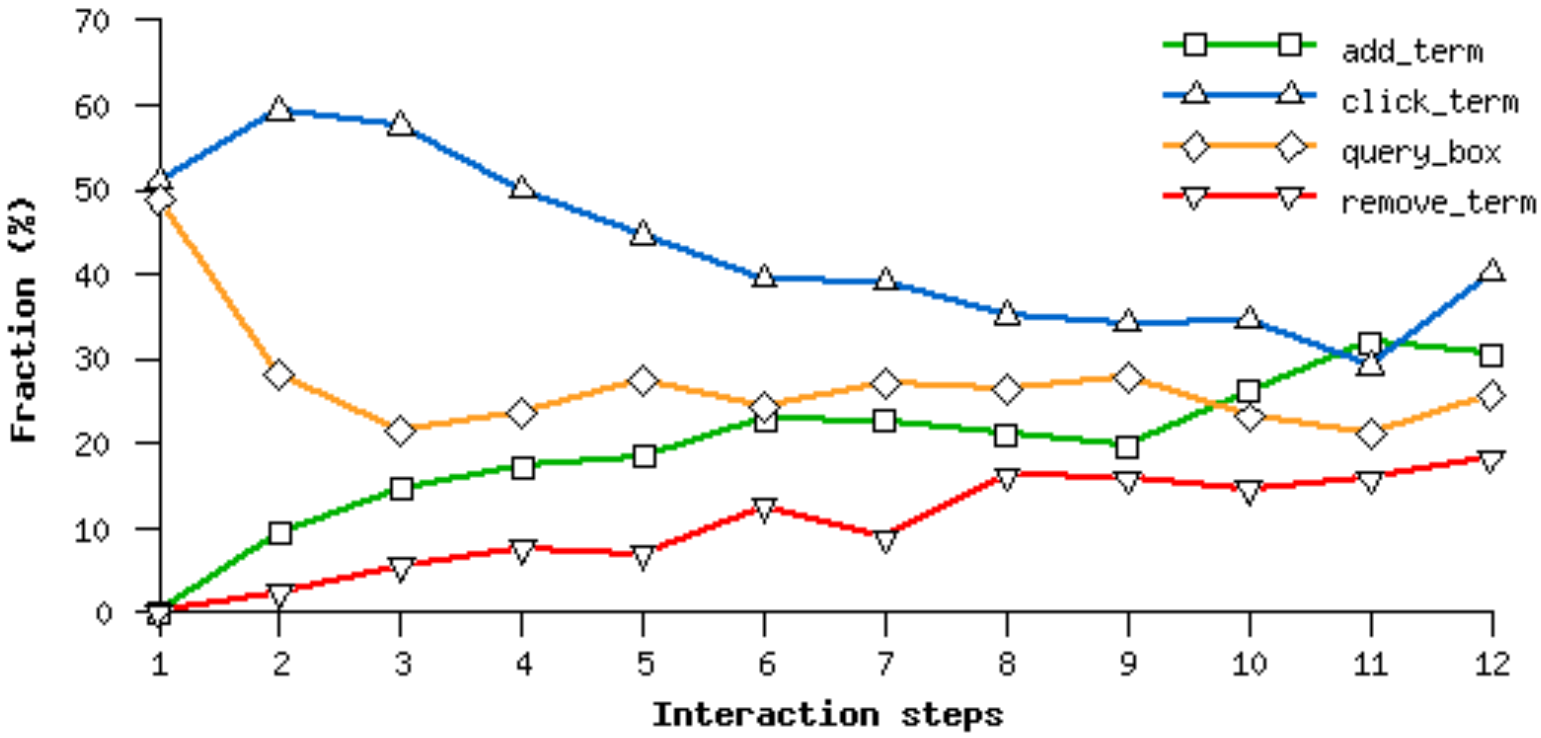
- Tag-based refinement is used more than query box

- Clicking tags is more common than adding tags



User Interaction

- Tag-clicks and query-box are dominant for beginning of interaction trails



Placing Flickr Images on a Map
Media Mining

Pavel Serdyukov
Vanessa Murdock
Roelof van Zwol
SIGIR'09

Personal content gets “geo-tagged”

geo-tags

Photo sharing sites become “Geo”

- Yahoo! (Flickr) and Google (Panoramio)

How it works… with GPS?

- Photos with “tenerife” tag near Tenerife
- In many cases works perfect!
- In some cases works awful!

University of Twente
Enschede - The Netherlands

How it works… without GPS?

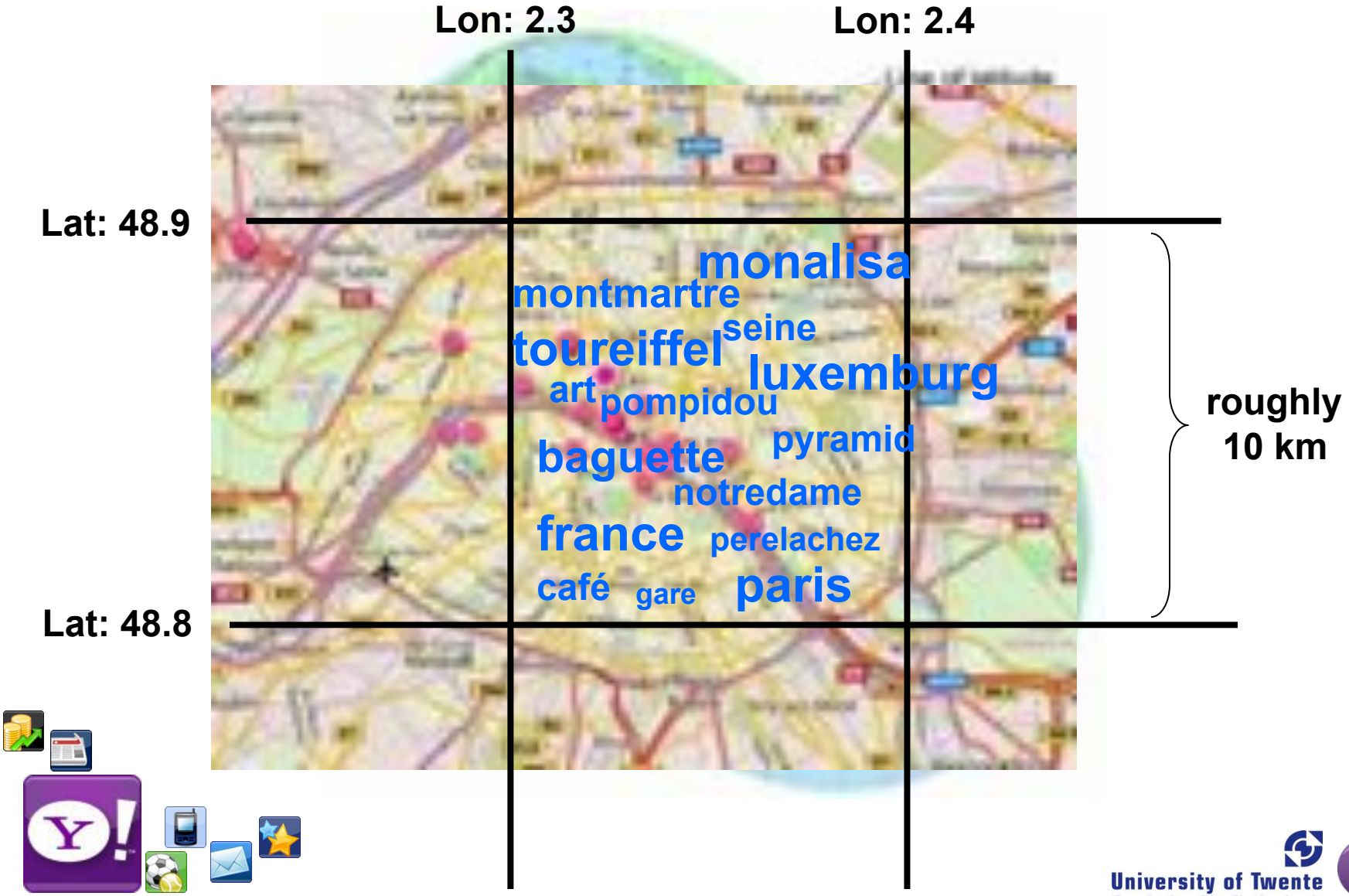
- Suppose you have a photo
- And you upload it to Flickr
- Flickr suggests you to put it on a map
- So, please, find where polar bears live

So, let's help users…

- **Map a photo using user tags only**
  - Around 96% of photos are not geo-tagged!
- **Use tags of many geo-tagged photos**
  - There are more than 100 millions of them
- **Do not rely on gazetteers entirely**
  - They are never complete

University of Twente
Enschede - The Netherlands

How to model locations?



Why all tags matter?

St. Petersburg

Popular tags

russia, **church**, bridge, cathedral, florida, **pier**, **sunrise**, tampa, st, light, **neva**, petersburg, water, tampabay, vinoypark, **pelican**, water, **hermitage**, russian, **winter**, baltic, **warpedtour**, bird, petersburg, bay

Finding "relevant" locations

- Locations are documents (L)
- Tagsets are queries (T)
- Tags of photos are query terms (t_i)
- Let's apply LM-based IR techniques
- How likely that location L produced the image with a tagset T :

$$P(T \mid L) = \prod_{i=1}^{|T|} P(t_i \mid L)$$

$$P(t \mid L) = \frac{|L|}{|L| + \lambda} P(t \mid L)_{ML} + \frac{\lambda}{|L| + \lambda} P(t \mid G)_{ML}$$

University of Twente
Enschede - The Netherlands

Smoothing with neighbors

- Locations are cells of the Earth grid
  - So, they all have close and distant **neighbors**
- Good locations are from good neighborhoods?
  - Support locations with evidence from surroundings
- Term-based smoothing:

$$P(t \mid L) = \mu \frac{|L|}{|L| + \lambda} P(t \mid L)_{ML} + (1 - \mu) P(t \mid NB(L)) + \frac{\lambda}{|L| + \lambda} P(t \mid G)_{ML}$$

- Cell-based smoothing:

$$P(T \mid L) = \alpha P(T \mid L) + (1 - \alpha) P(T \mid NB(L))$$

average over neighbors of L

University of Twente
Enschede - The Netherlands

Considering spatial ambiguity

- Not all toponyms are equally location-specific!
  - sanfrancisco (28 places) vs. tokyo (1 place)
  - bath – not only the city in UK

- And even not all tags are equally useful
  - chihuahua (world popular dog) vs. dingo (australia)

- Let's increase the influence of unique, less ambiguous tags
  - Let's use **spatial** features of tags for that

University of Twente
Enschede - The Netherlands

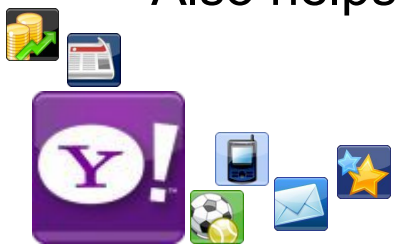
Tag-specific smoothing

- Let's make individual probabilities of ambiguous tags less decisive
- How to characterize **spatial ambiguity**?
- **Standard deviation** of coordinates of images using the tag:

$$\lambda(t) = \lambda + \gamma(\sigma_{lat}(t) + \sigma_{lon}(t))$$

$$P(t \mid L) = \frac{|L|}{|L| + \lambda(t)} P(t \mid L)_{ML} + \frac{\lambda(t)}{|L| + \lambda(t)} P(t \mid G)_{ML}$$

- Also helps not to over-boost ambiguous place names!

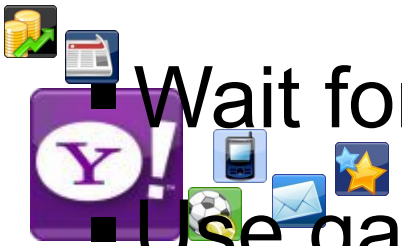


University of Twente
Enschede - The Netherlands

Hard cases (I)

- No models for rarely visited locations

tasiilaq greenland



- Wait for more data?
- Use gazetteers?

Hard cases (II)

- No location-specific tags
- No disambiguating tags

beach coast rocks lovers

michigan cats dogs

- Additional evidence needed
  - › IP location? Image analysis?

University of Twente
Enschede - The Netherlands

Hard cases (III)

- Conflicting location specific tags
- Metonymy:
  - “Italy scored on the last minute”

italy australia

- Evidence from photo groups?

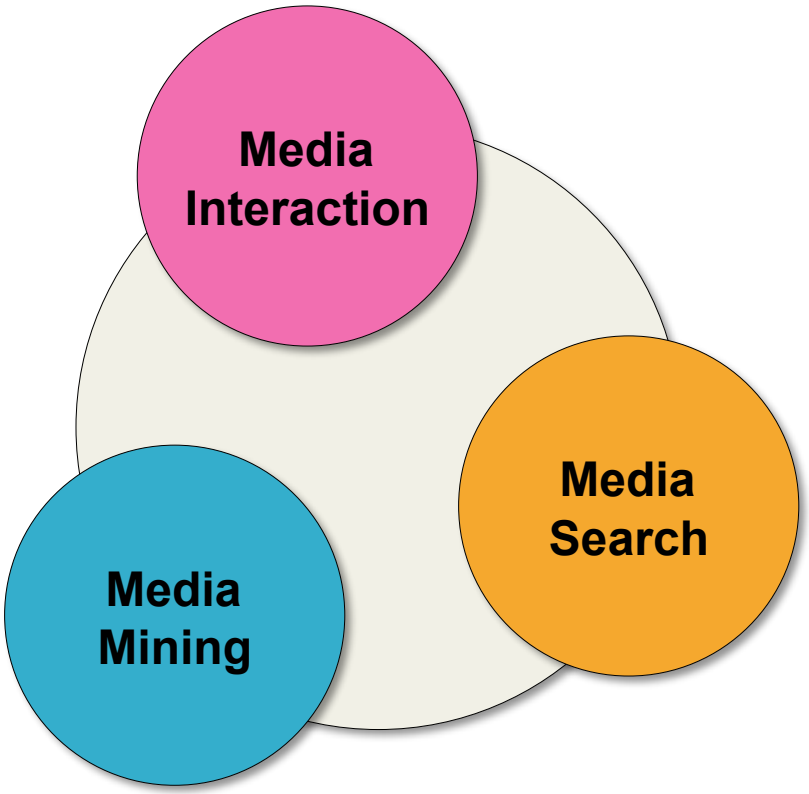
Hard cases (IV)

- Tags specific to very large regions

russia river bear

- Guess the zoom level?
- Diversify the result?

Media Search



Search is changing…

Faceted Image Search

Ranking images with clicks, textual, and visual features

Search result diversification

“Web of Objects” paradigm
Search is Changing…

Ricardo Baeza-Yates
Roelof van Zwol

Next Generation Web Search

Web search is no longer about document retrieval

- Means for web-mediated goals

Witness a new breed of search experiences

- Demands search ecosystem that combines content with intent
- Exploiting the “Wisdom of Crowds” behind the Web 2.0

We are going to:

“the Web-of-Objects”

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:**

"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".

Trends

User Generated Content

- Massive (quality vs. quantity)
- Social Networks
- Real time (people + sensors + mobile)

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

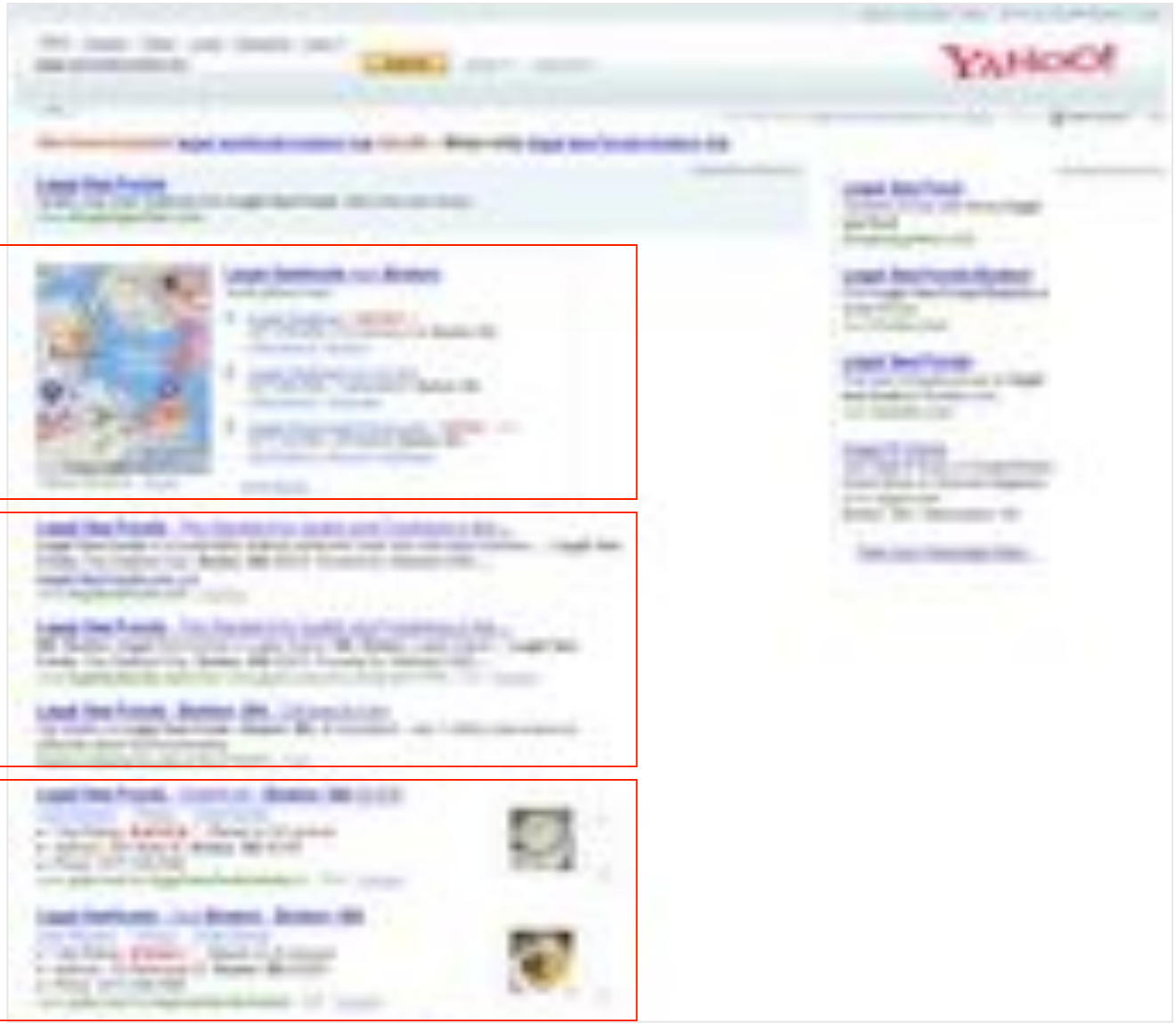
- Moving towards identifying a user's task
- Enabling means for task completion
- ***Media integration***

New user experiences based on the Web 2.0

Challenges:

- on-line
- scalability

More complete information



Shortcuts

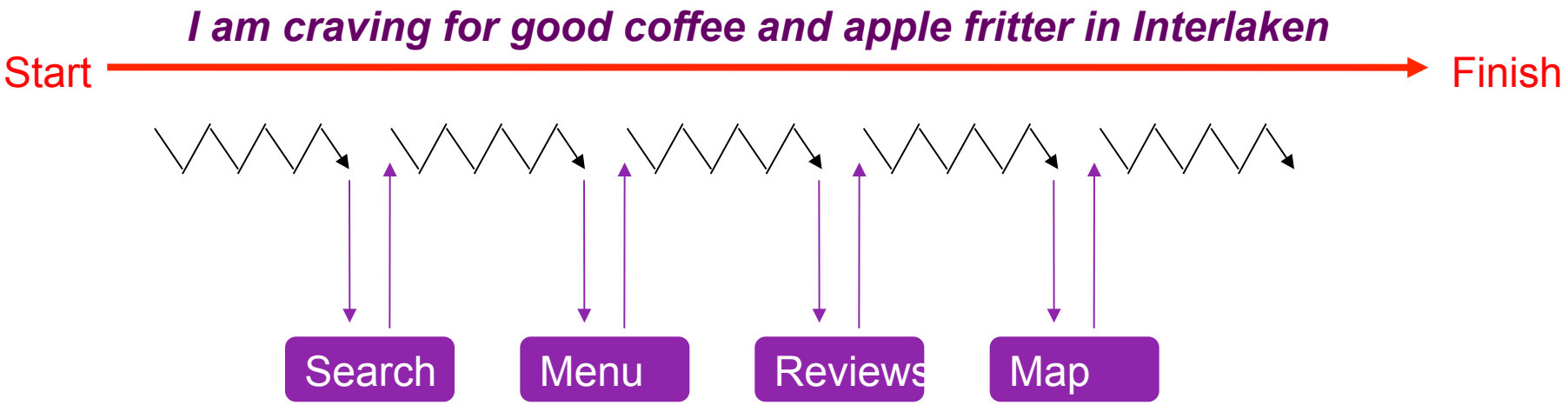
Deep Links

Enhanced Results

Search: Content vs. Intent

Premise:

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



Next gen?

We move from a web of pages to a web of objects

- Objects are people, places, businesses, restaurants …
- Objects have attributes
- Missing, noisy, etc.

Discover and satisfy intent by presenting objects and attributes

- Objects define faceted search

How to obtain structured objects?

Web Content

Metadata/Taxonomies/Folksonomies

Classification/ML/Extraction/Semantic Web

Media analysis

- What images to show, what objects depicted, salient objects, visually similar, etc. … But ALL needs to be done on a very large scale!

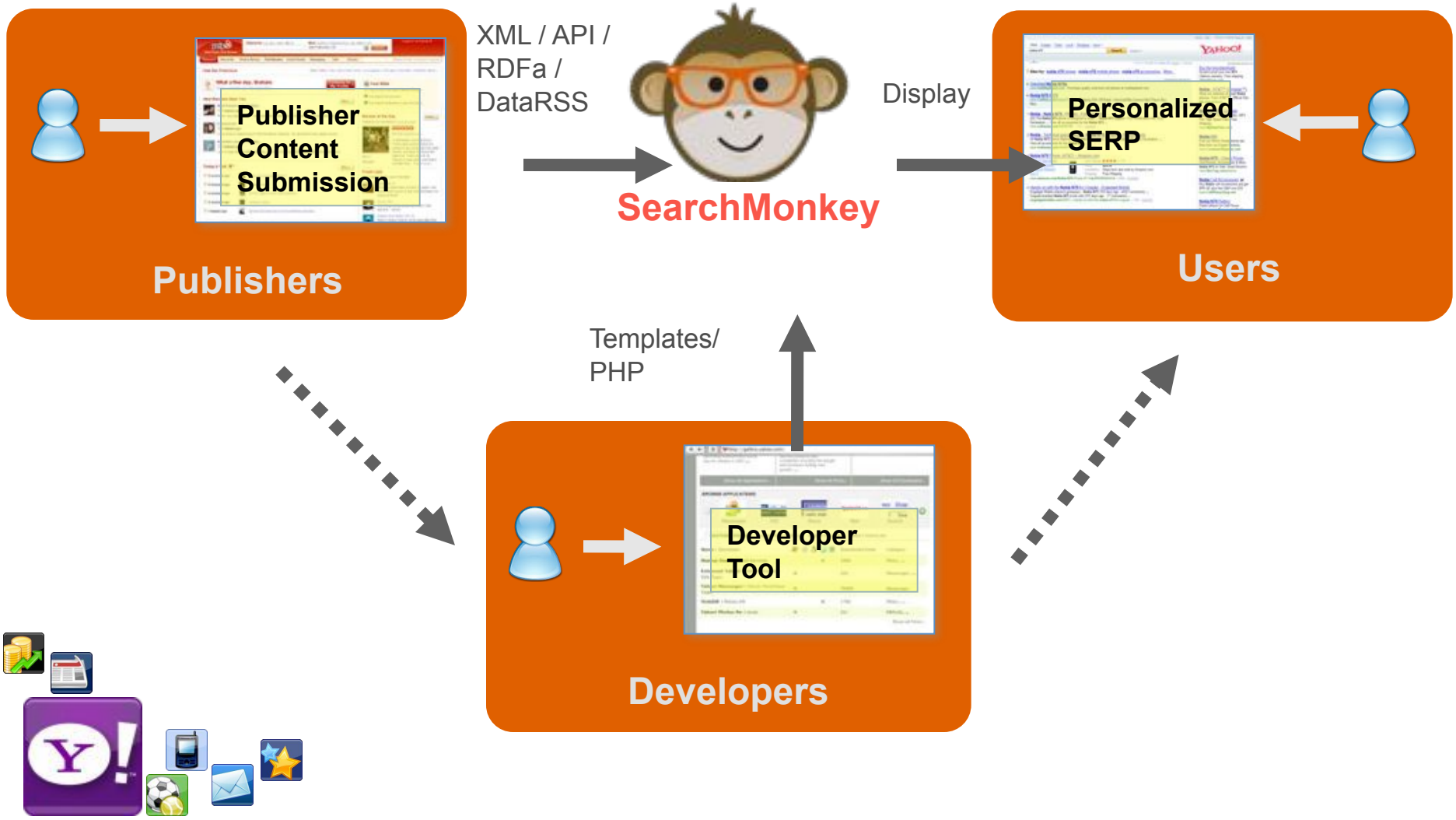
Web 2.0 & Web Usage

Explicit & Implicit relations

Building out an open ecosystem

Publishers have incentives to contribute

The SearchMonkey Ecosystem



Opening search - what does it mean?

Clear win for: **developers**, **site owners**, **users** and **Yahoo!**

Go from “**to-do**” to “**done**”

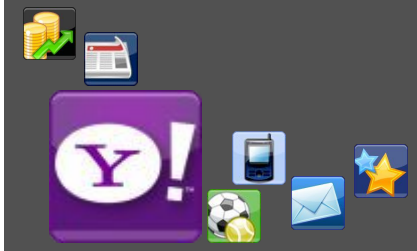
BEFORE

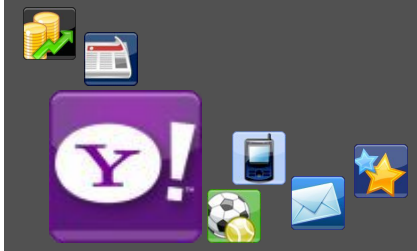
AFTER

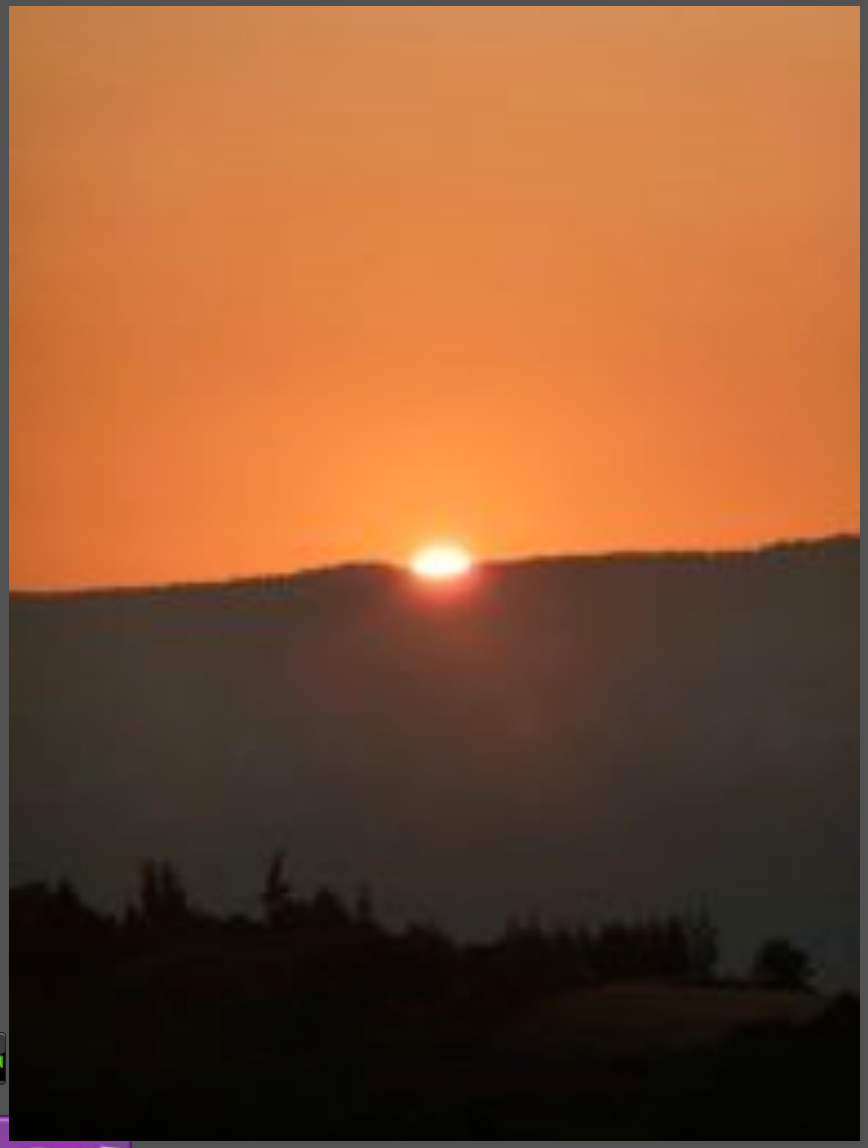
How does that affect (social) media
Search is Changing…

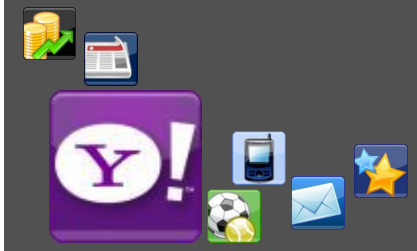
Roelof van Zwol

A small experiment…





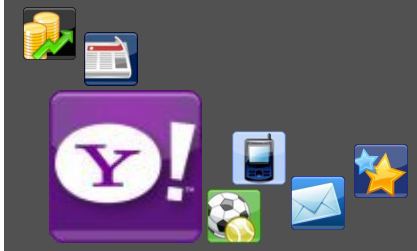




What did you see?

Where did your attention go?

How long to interpret the picture?

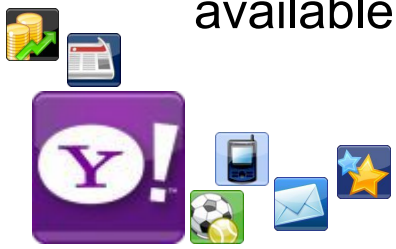


Next Generation … Media Search

Searching for images or video on the Web,

more than a matter of ranking by relevancy!

- **Entertain**: Not necessarily task oriented: ~ 40% of page views are related to celebrities and entertainment
- **Curiosity**: People tend to click on seemingly unrelated images out of curiosity
- **Diversity**: offer topical and visual diversity to satisfy the needs of many, and to compensate for "lack of power in query formulation"
- **Visual quality**: First notion of quality can already be obtained from thumbnails and image dimension, people search for people.
- **Exploratory in nature**: More than one media object needed to satisfy a user's need
- **Novelty**: Being able to serve new content as soon as it becomes available.



Faceted Media Search
Media Search

Roelof van Zwol,
Lluis Garcia,
Mridul Muralidharan,
Borkur Sigurbjornsson
and many others!
WWW'10

Overview

- Serving facets for image search

- Extracting entity facets
- Extracting ranking features
- Ranking candidate entity facets
- Evaluation
- Conclusions

People

Product & design

- Kaushal Kurapati
- Anuj Sahai
- Polly Ng

Engineering

- Anand Ramani
- Sriram ‘Thiru’ Sathish
- Ramu Adapala
- Abhinav Katiyar
- Murali Krishna
- Balaji Kanan

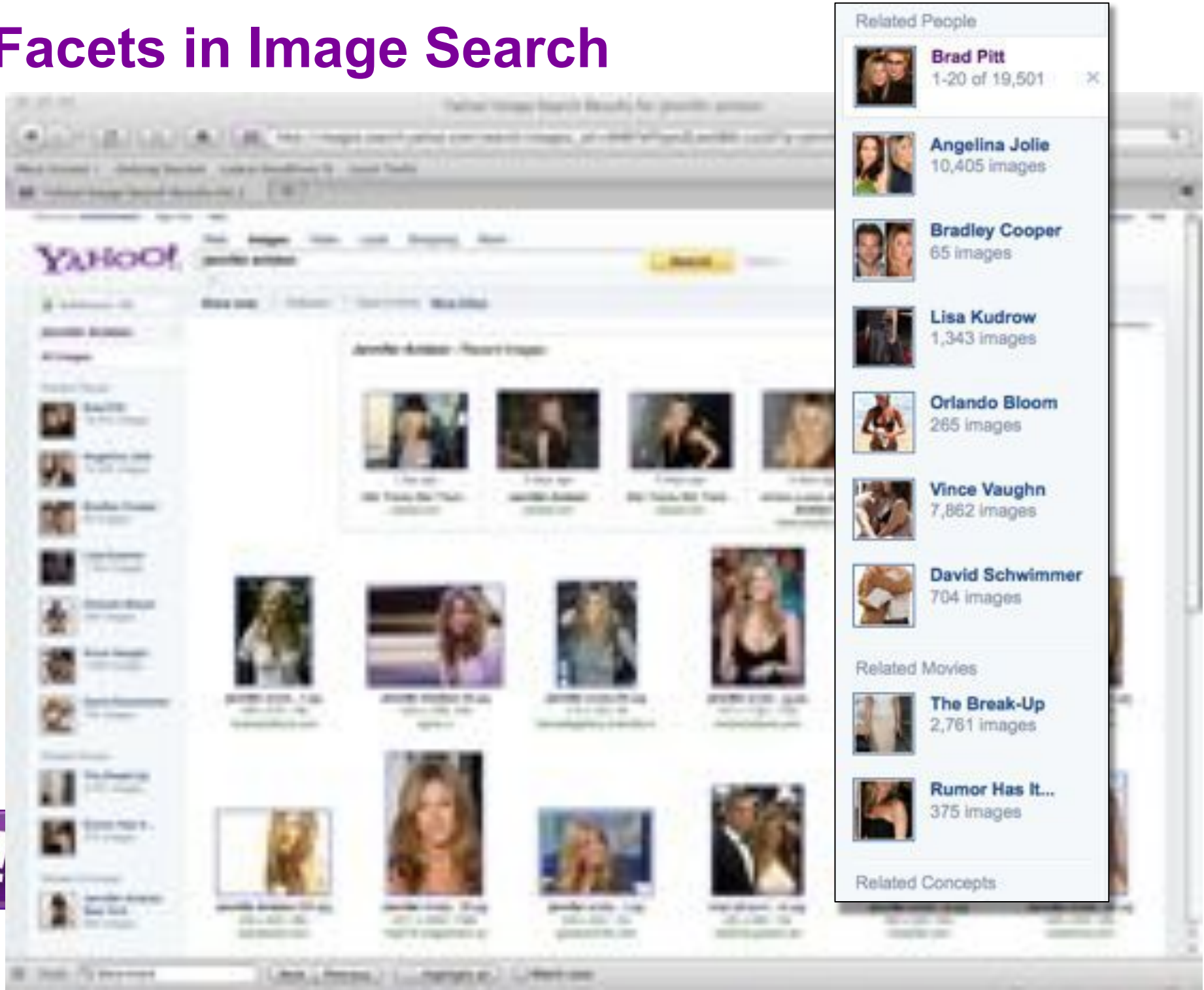
Research

- Roelof van Zwol
- Borkur Sigurbjornsson
- Lluis Garcia
- Mridul Muralidharan

Sciences

- Nicolas Torzec

Facets in Image Search



Facets in Image Search (cont'd)

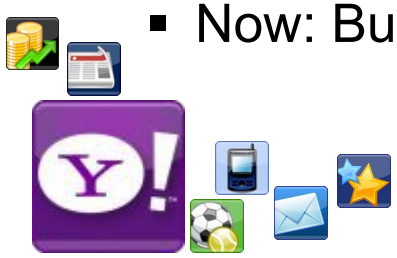
GridFaces

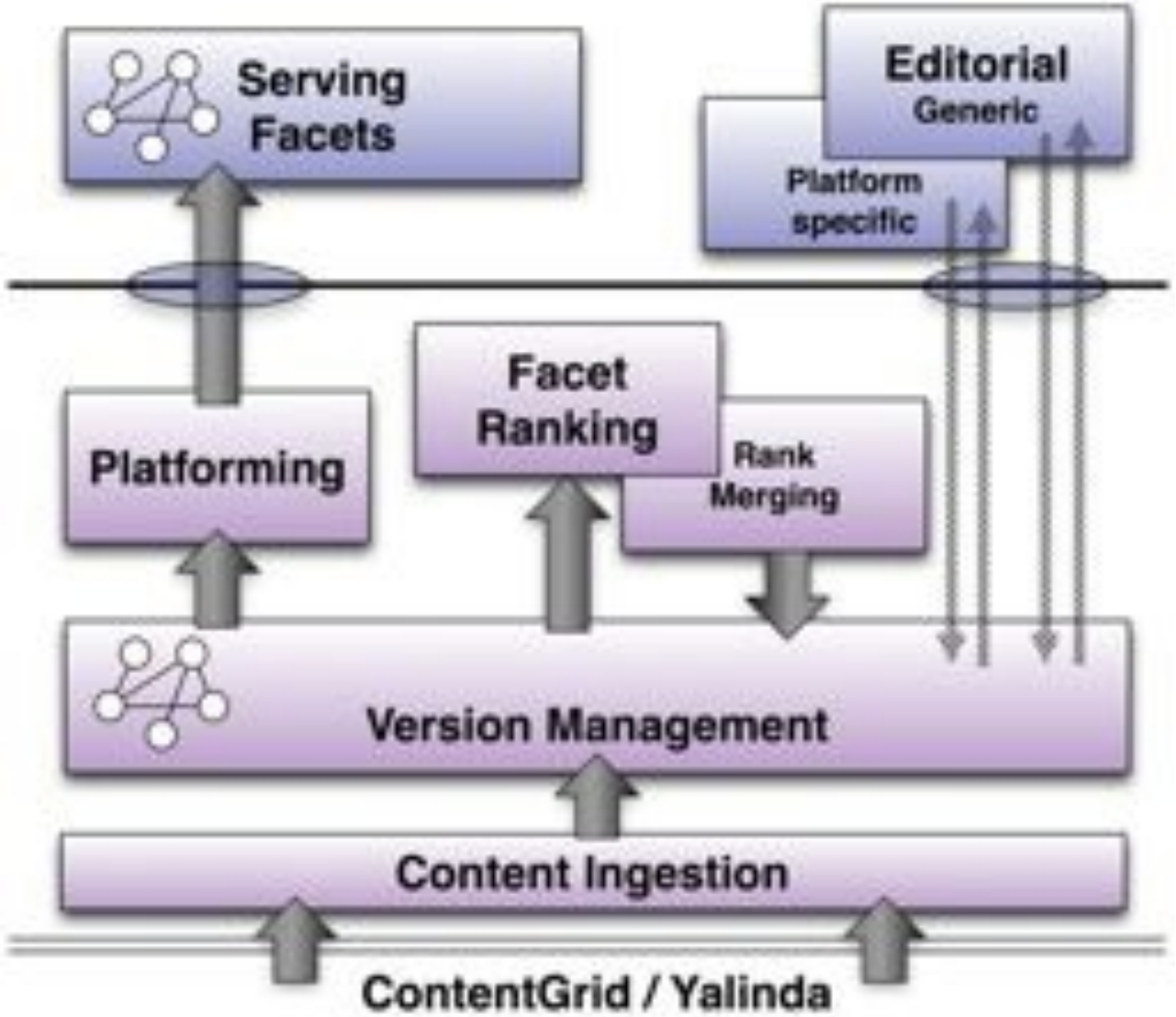
Goals:

- Power the faceted search experience of image search
- Promote the "Web-of-Objects" paradigm through the introduction of facets in the SERP.

Main milestones

- May 15th: Travel facets in bucket test
- July 23th: Travel facets launched
- August 6th: Celebrity facets in bucket test
- September 21st: Celebrity facets launched
- December: video search, mobile adopts facets
- Now: Bucket tests in web search




Serving
Facets
Editorial
Generic
Platform
specific
Facet
Ranking
Rank
Merging
Platforming
Version Management
Content Ingestion
ContentGrid / Yalinda

Extracting Facets

- A facet is defined as the directed relationship between two entities (e,f).
- For a given entity e, a set of candidate facets F is collected. We refer to entity f ∈ F as the facet of entity e.
- Entities and facets are extracted from structured sources, such as: Y! Movies, GeoPlanet, Y! Travel, Wikipedia, etc.
- Pool of 6M+ entities and 80M+ candidate facets

name	**Justin Timberlake**
aliases	JT, Justin Randall Timberlake, J. Timberlake
type	Person (musician, actor)

entity *e*	Justin Timberlake
entity *f*	Jessica Biel
type	Romantic relationship

Extracting Features

Extract a set of features from different ranking sources:

- Image search query terms
- Image search user sessions
- Annotated photos in Flickr
- Favorites in Y!music

Pre-process sources into a common format

Extract statistical features:

- Atomic features
- Symmetric features
- A-symmetric features
- Combined features

Extracting Features

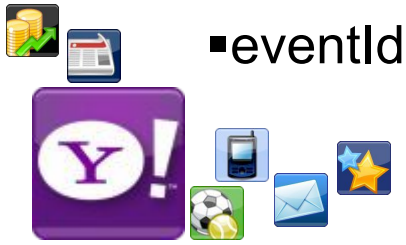
Query term analysis:

- For every query entered by a user, we extract co-occurring entity pairs:

User query:	**Cubbon park in Bangalore, India**
Tokenization:	Cubbon+park+in+Bangalore+India
Normalization:	cubbon+park+in+bangalore+india
Segmentation:	cubbon+park+in+bangalore+india
Entity detection:	cubbon park; bangalore; india; bangalore india
Cooc pairs:	(cubbon park, bangalore india), (cubbon park, india), (cubbon park, bangalore), (bangalore,india)

Per event collect (common format):

- eventId, userId, timestamp, (e1,e2)+



Extracting Features

Independent from the source, the following set of features is extracted:

Atomic features
P(e), P(f)
E(e), E(f)
Symmetric features
P(e,f), P_u(e,f), SI(e,f), CS(e,f)
A-symmetric features
P(e|f), P(f|e), P_u(f|e), KL(e||f), …
Combined features
P_u(e|f)xP(f), P_u(f|e)xP(f) …

* *P_u(f|e)* is a variant of P(f|e), where each entity *e* and entity pair *(e,f)* is counted once per user. To make the feature less prune to the impact of a single user.

User Click Feedback

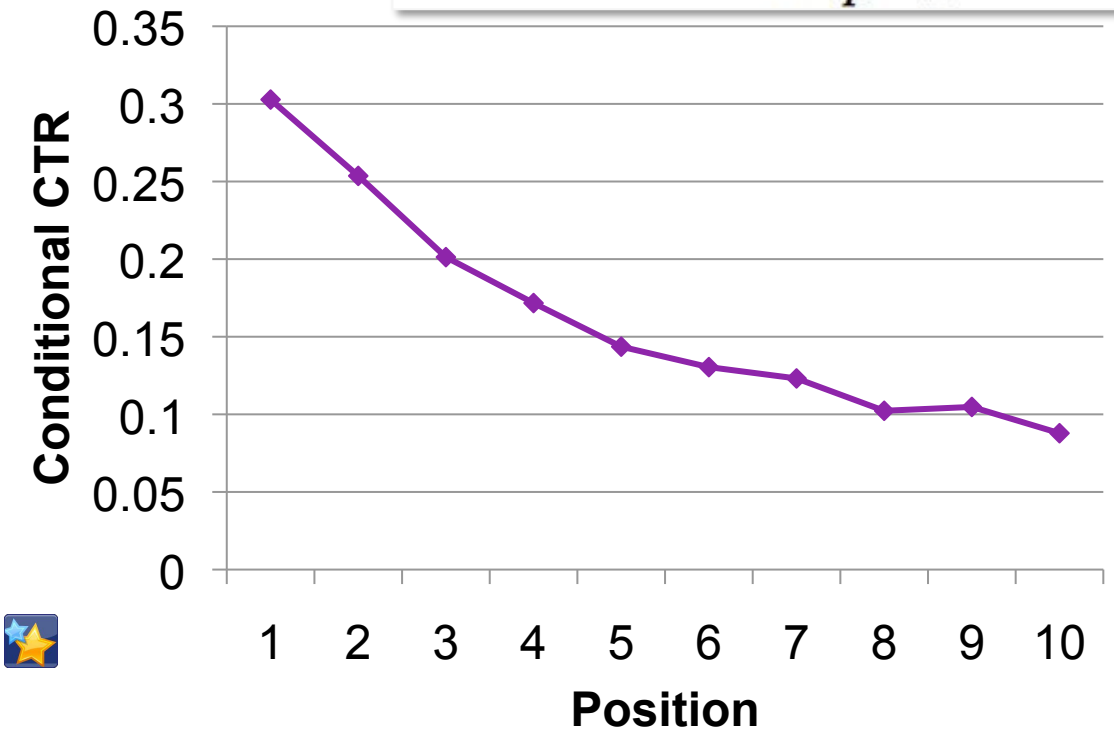
Adopted two click-feedback models:

- $CTR_{e,f}$

$$ctr_{e,f} = \frac{clicks_{e,f}}{views_{e,f}}$$

- $COEC_{e,f}$

$$coec_{e,f} = \frac{clicks_{e,f}}{\sum_{p=1}^{P} views_{e,f_p} \times ctr_p}$$



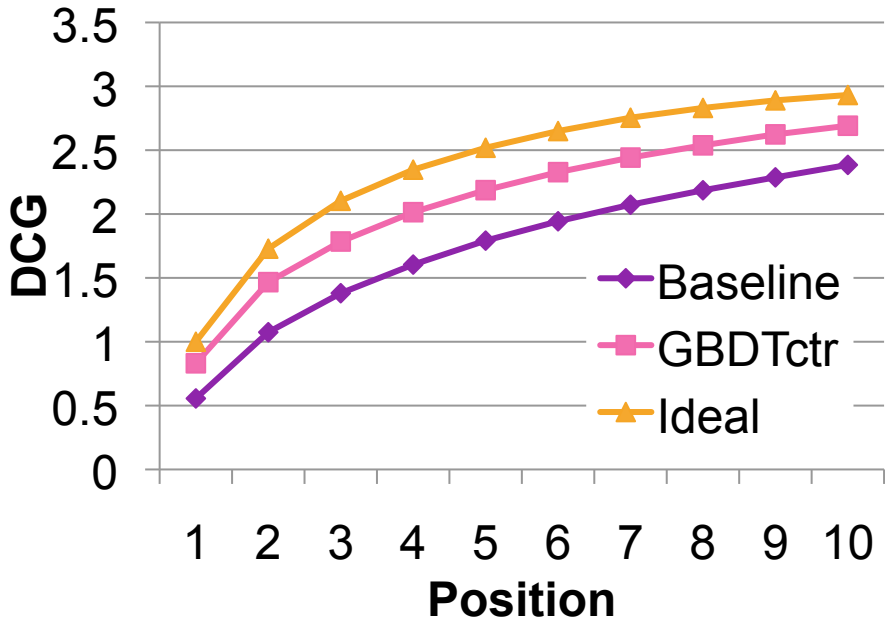
Evaluation – Overall Performance

	CTR		COEC	
Run	**mDCG**	**mnDCG**	**mDCG**	**mnDCG**
Ideal	2.375	--	2.594	--
Baseline	1.728	0.709	1.812	0.677
GBDT$_\star$	2.090	**0.874**	2.436	**0.930**

- Based on the mnDCG computed over the first 10 results.

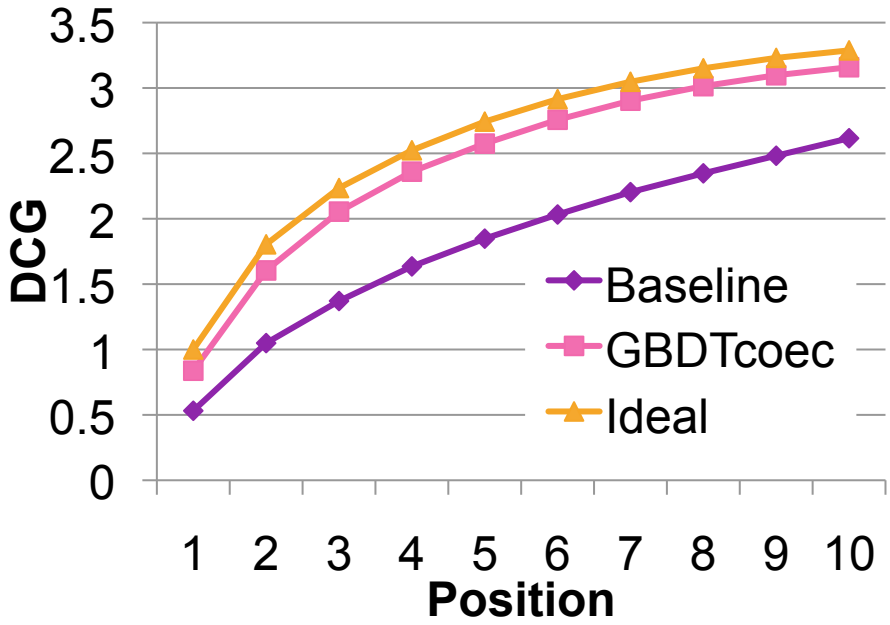
- Comparing baseline performance against the CTR/COEC is *unfair*, due to the grouping of facets by category!

Evaluation – DCG@p

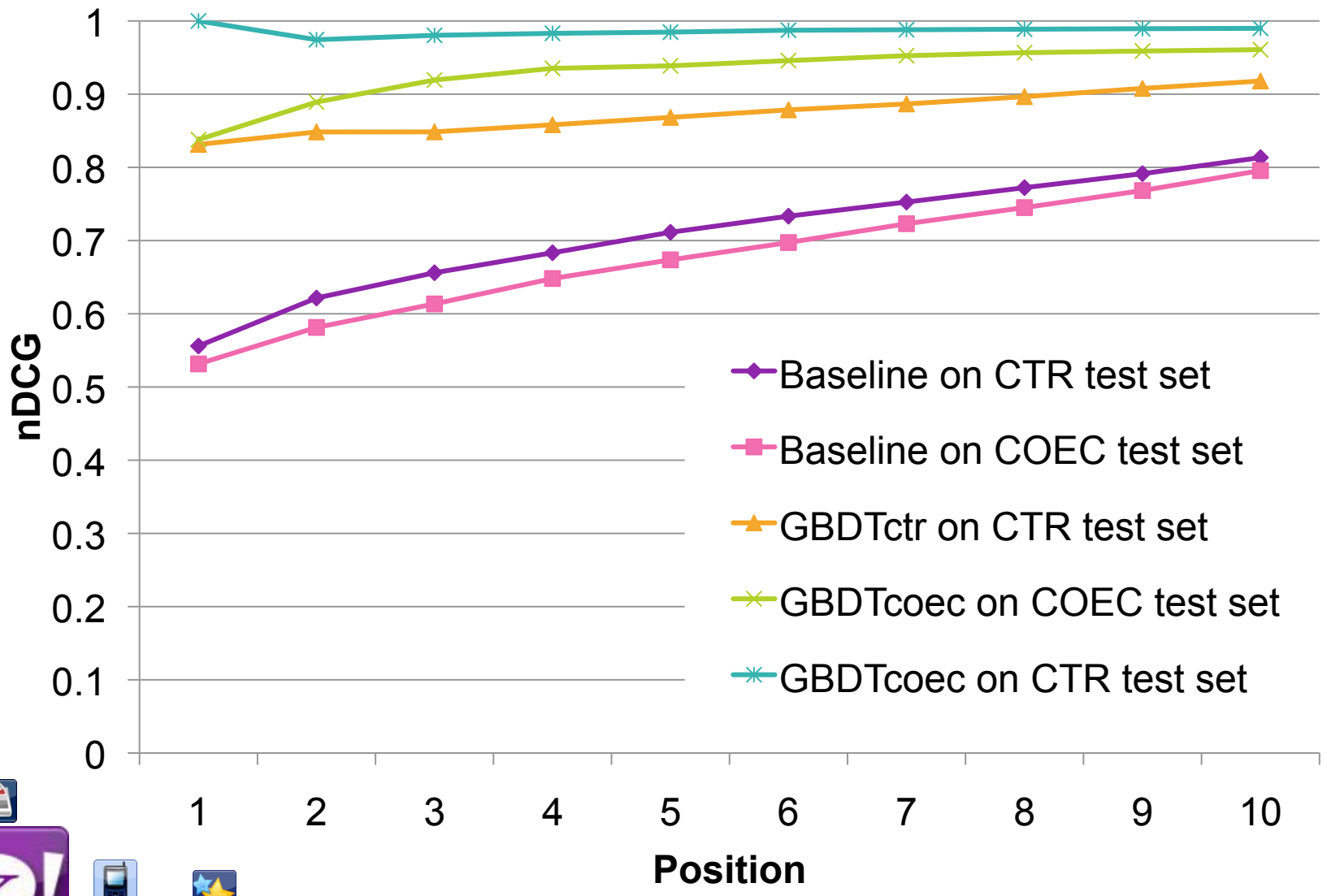


CTR test set

COEC test set



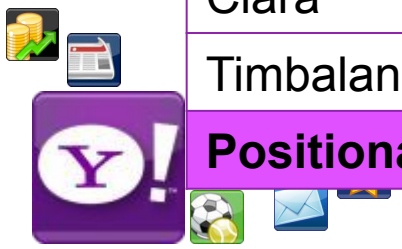
Evaluation – nDCG@p



Evaluation – per query

- All results reported are statistically significant ($p<0.001$)
- "Justin Timberlake" example:

Facet	CTR	COEC	Basel.	G_{coec}	G_{ctr}
Jessica Biel	1	7	5	1	1
Jesse Mc Cartney	2	1	12	2	4
Britney Spears	3	5	10	5	3
NSync	4	2	7	3	9
Alpha Dog	5	3	1	6	10
Cameron Dias	6	10	4	4	6
JC Chasez	7	9	2	8	8
T.I.	8	4	13	9	11
Ciara	9	8	9	11	5
Timbaland	10	6	11	10	13
Positional error	**--**	**28**	**41**	**10**	**23**



Evaluation – Feature Importance

$GBDT_{ctr}$		$GBDT_{coec}$	
Feature	**Weight**	**Feature**	**Weight**
QS $P_u(e\|f) * P(f)$	100	QT $P_u(e,f)/P(f)$	100
QT $P(e)$	85.11	FT $P(e)$	11.56
QS $P(e)$	76.88	QT $P(e)$	9.57
QT $P_u(e,f)$	69.32	QS $P(e)$	9.22
QT $P_u(e\|f) * P(f)$	69.21	FT $E(e)$	9.22
QT $P_u(f\|e) * P(f)$	64.38	FT $KL(e)$	8.84
QS $P(e,f)$	59.78	QT $P_u(f\|e) * P(f)$	8.19
QT $P(e,f)$	52.98	QT $P_u(e\|f) * P(f)$	8.14
QS $P_u(e,f)$	48.26	QS $P(f)$	7.53
FT $P(e)$	43.71	QT $P_u(e,f)$	7.25

QT: Query term; QS: Query session; FT: Flickr tag.

Conclusions – Ongoing work

Image search first to introduce the WOO in the SERP

Introduce a machine learned approach for ranking facets, based on user-click feedback

- Extract features in generic manner from various sources (query term-, query session-, and Flickr tag analysis)
- Enriched feature space (user prone, and combined)
- Adopt/evaluated two click-feedback models
- GBDT models outperform baseline

Experiment with models for different categories.

Ranking images with clicks, textual, and visual features
Media Search

Roelof van Zwol
Vanessa Murdock
Lluis Garcia
Ximena Olivares
TechPulse'09

Image MLR

Observations:

- “Few” clicks on images (but still >> M clicks / day)
- SERP based on grid display

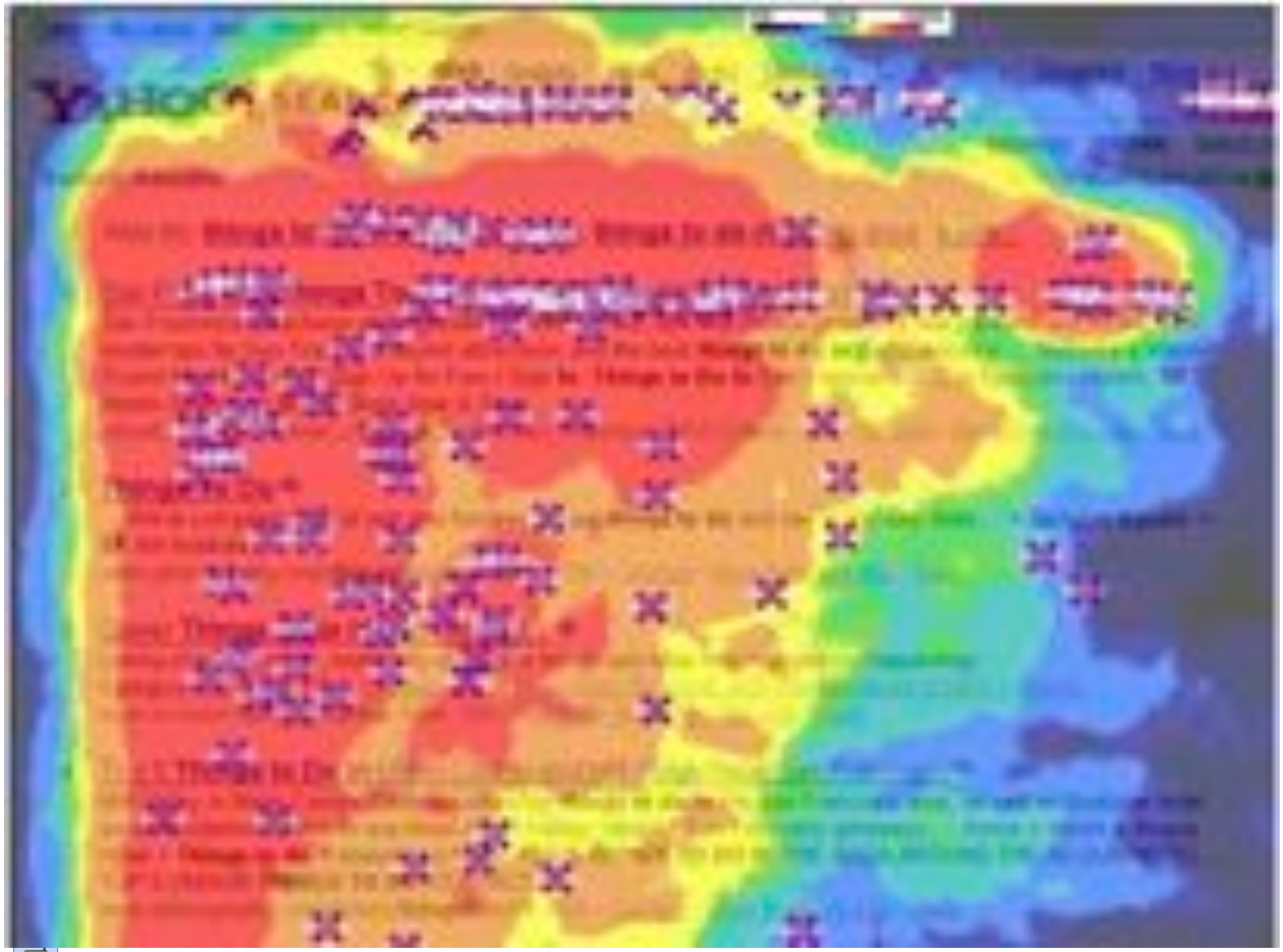
Hypotheses:

- Image thumbnails convey more information than document snippets
- Relevance judgments made on visual relevance, not textual relevance

Focus on:

- Block construction and models for deploying query-clicks in image search
- Incorporating visual features in MLR

The Golden Triangle of Web Search



How do users scan the SERP?

High resolution image seen by the **fovea**

2° = Diameter 0.8” (60 pixels)

Reduced visual acuity experienced by the **parafovea**

5° = Diameter 2.1” (150 pixels)

Progressively reducing visual acuity from the periphery of the retina

How customers look at the image SRP

In Image Search, it is easier for customers to use peripheral view to determine relevance and efficiently scan to see more images on the page.

The content in images are easier to see in the peripheral view when compared to text, where the center of the eye must focus to obtain information.

In Web Search, customers use peripheral view to identify the parts most likely to have relevant information based on the location of boldfaced terms.*

Text areas must be directly looked at in order to obtain the information.

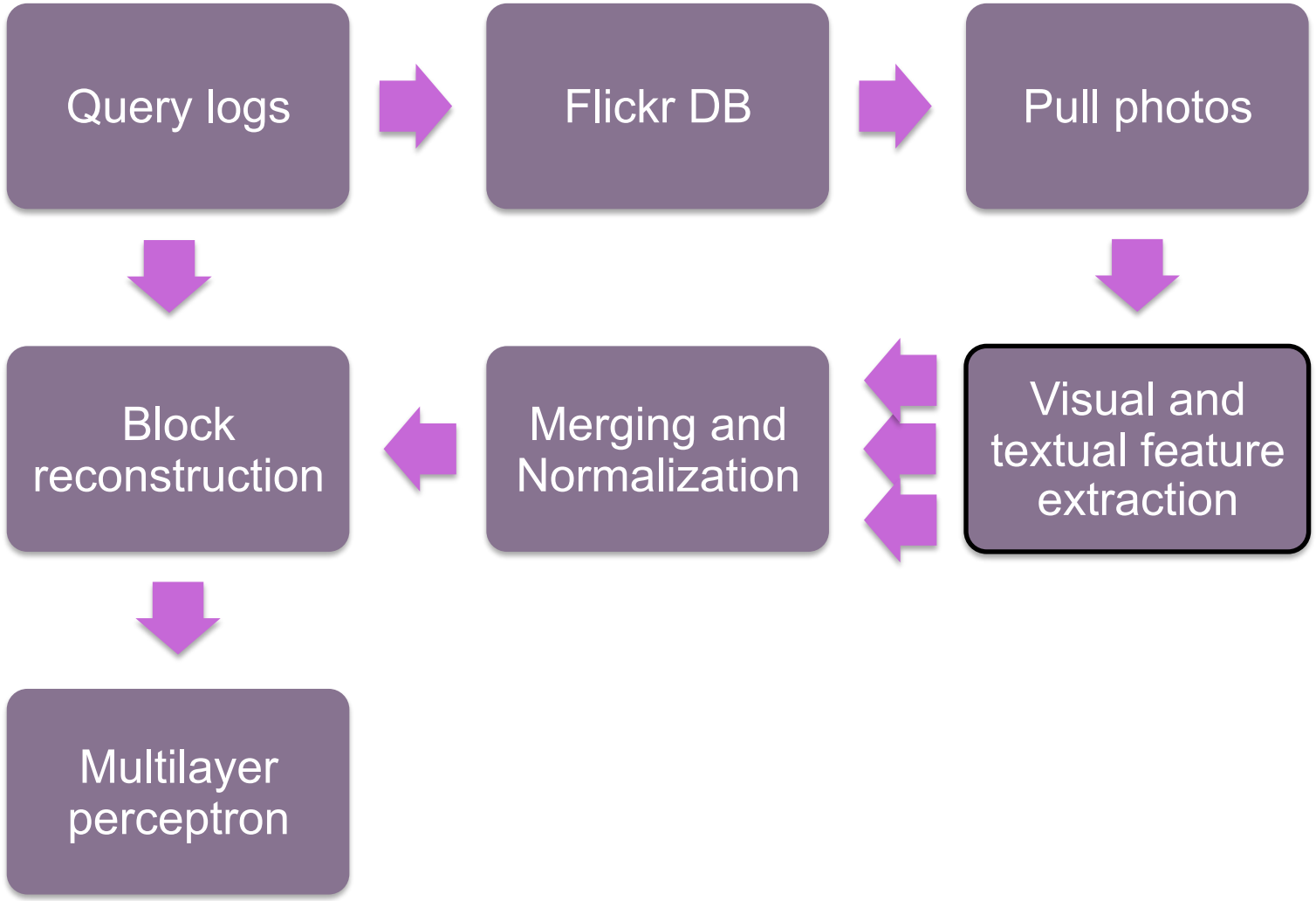
Ranking Images with Click Data

Ranking Images with Click Data

Ranking Images with Click Data

Ranking Images with Click Data

Image feature extraction pipeline – Hadoop Map/Reduce



Data collection

3.5 million distinct photos from Flickr:

- Meta-data: tags, titles, descriptions
- Only public photos

Randomly-sample 600,000 unique queries from image search logs

- Include the search results: clicks and views

Filter out non-flickr results

Filter out non-public flickr photos

Paris – Gare du Pont Cardinet – 28-07-2007 – 9h03

Jolie gare des années 1920. Influence angkorienne et ressemblance frappante avec le palais du roi Narai, á Lopburi… Nice, small Paris railway station from the 1920's. Through the Angkorian influence, strong similarities with the King Narai's Palace, in Lopburi…

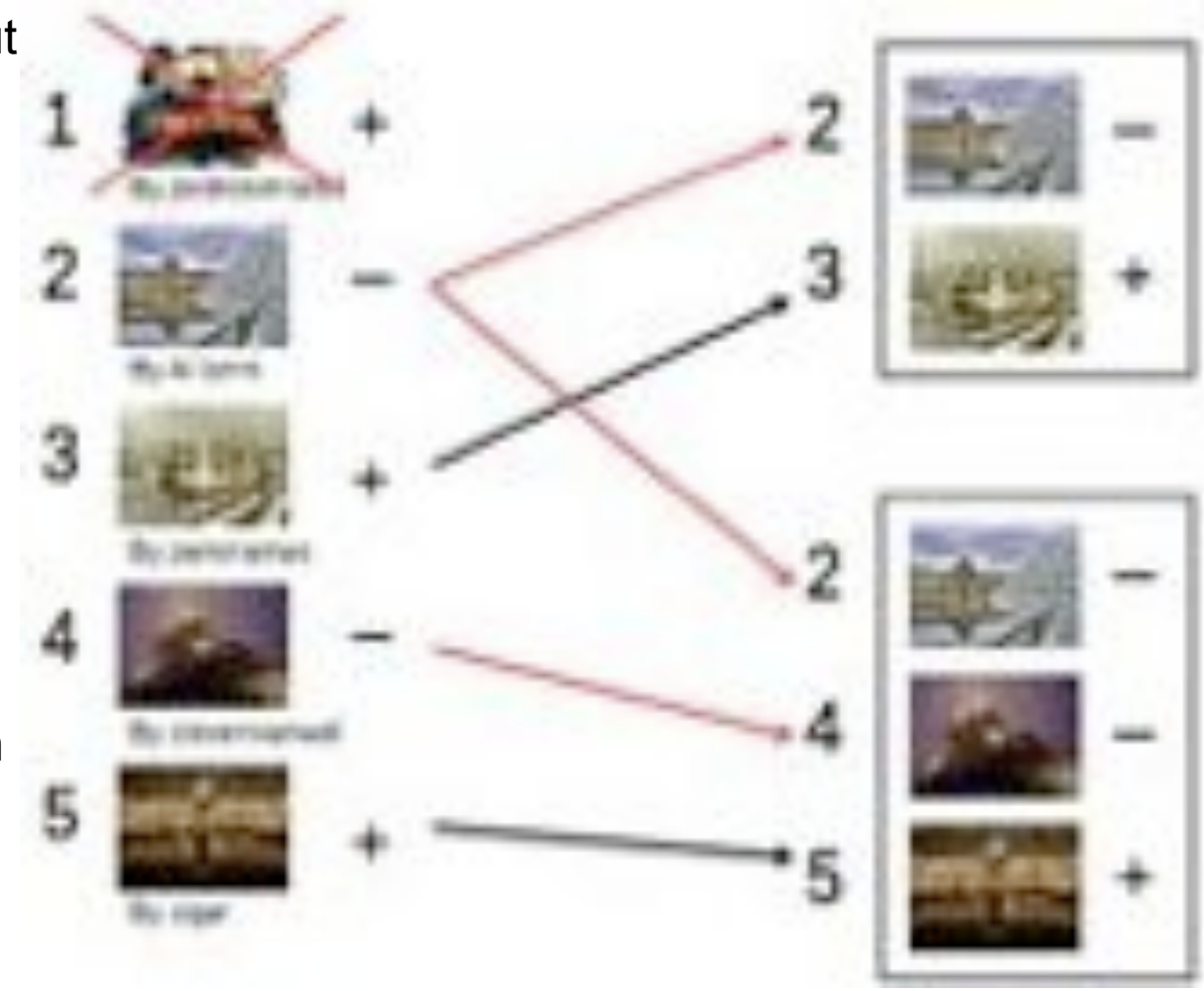
Image by Panoramas on Flickr

Tags:

gare
pont
cardinet
paris
RER
1920
1925
Angkor
ankorien
Angkorian
Lopburi
Narai
pluie
rain
reflet
reflection
zebra

Learning from Clicks

- Replicate block construction from literature [Joachims, Ciaramita]
- Discard blocks without negative examples
- User clicks give relative preference
- Clicks at rank 1 ignored
- Train and evaluate in “blocks”
- Multi-layer perceptron



Training details

Training: 1,167,000 blocks

Testing: 250,000 blocks

Parameters tuned only on the textual features

Multi-layer Regression

- One hidden layer
- Ten training iterations

Textual Features

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

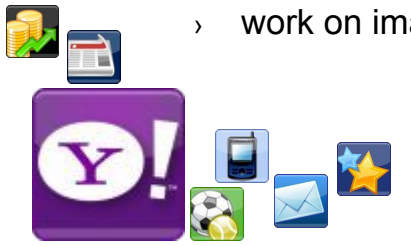
- Scores normalized by column and by row

Visual Features

- Extract low-level global (and local) features

Color
Color histogram
Color layout
Scalable color
Edge
CEDD
Edge histogram
Texture
Tamura

- Focus on light-weight features
  - work on image thumbnails (120*160 pixels)



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 confidence score

- Use the score to rank images in each block

Results

System	Accuracy	MRR
Retrieval baseline	0.4198	0.6286
Learned baseline	0.4073	0.6104
Textual features	0.5484	0.7034*
Visual features	0.5805	**0.7233**
Combined features	0.7512	**0.8365**

Conclusions

- Textual features improve results over the baseline retrieval
- Visual features improve results over the text-based features
- Combination of Text and Visual most powerful
- Block construction effective even though doesn't mirror human gaze

Diversifying Image Search Results
Media Search

Dimensions of Diversity

- **Topical diversity**

  Query: “Jaguar”





- **Visual diversity**

  Query: “Jaguar X-type”



- Other dimensions: spatial, temporal, social

Topical Diversity
Retain relevancy, improve diversity

Roelof van Zwol
Vanessa Murdock
Lluis Garcia
Georgina Ramirez
ACM MIR'08

Topical Diversity

Diversification as part of the retrieval model through variation of content types

- 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%

Retrieval Performance

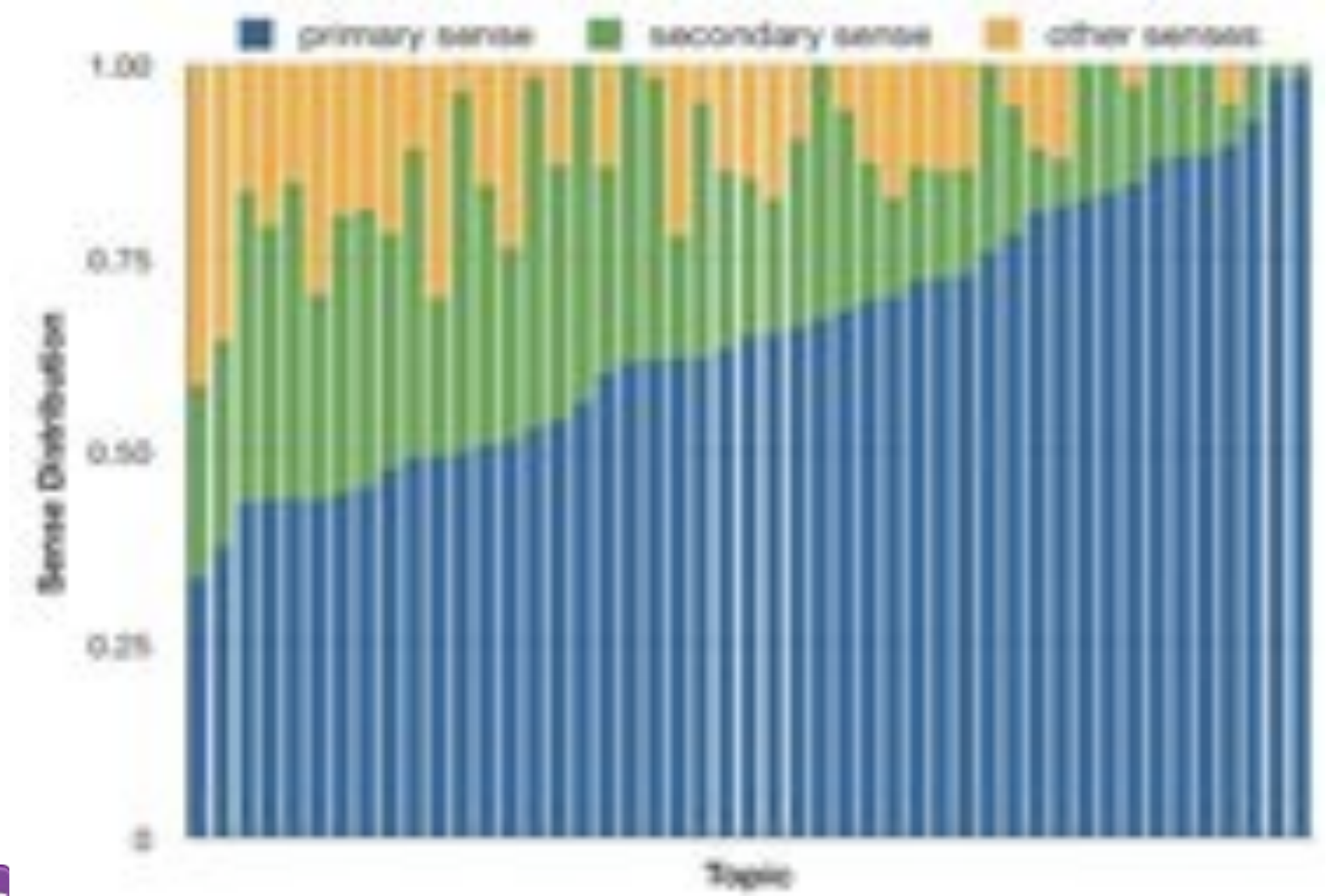
Unambiguous topics

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

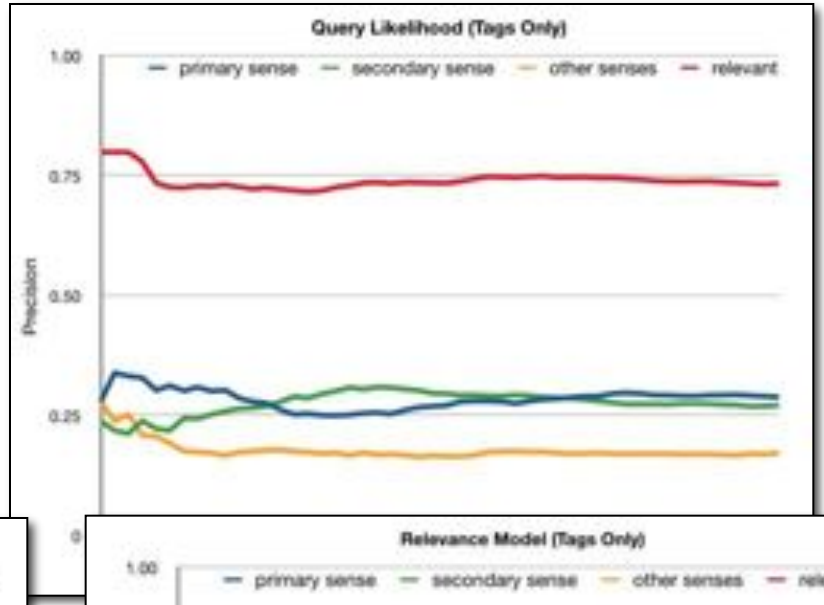
Model	P@1	P@5	P@10	P@15	P@20	P@25	P@50
Query Likelihood	0.680	0.760	0.720	0.725	0.734	0.744	0.734
Query Likelihood (Tags Only)	0.800	0.736	0.732	0.720	0.736	0.736	0.734
Relevance Model	0.720	0.760	**0.768**	**0.784**	**0.788**	**0.792**	**0.778**
Relevance Model (Tags Only)	**0.840**	0.728	0.744	0.741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

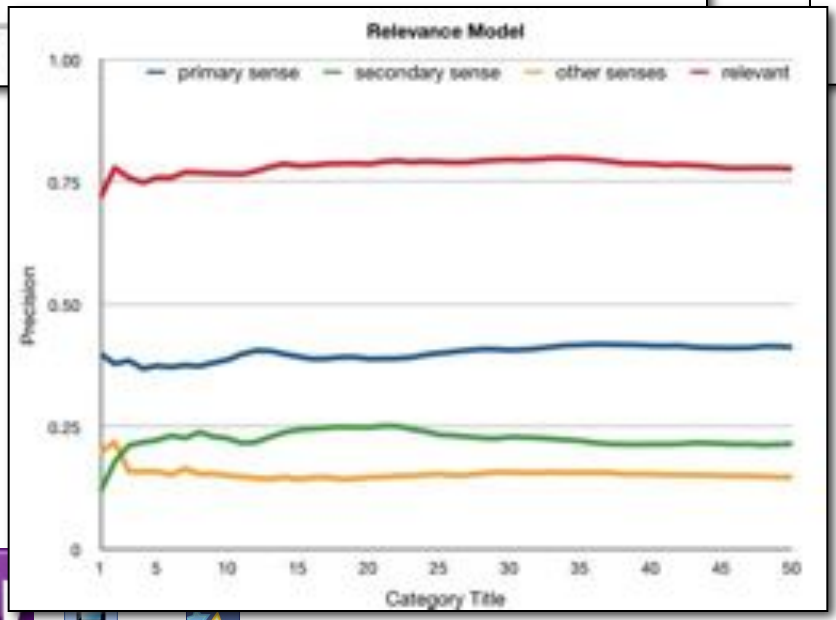
Sense Distribution

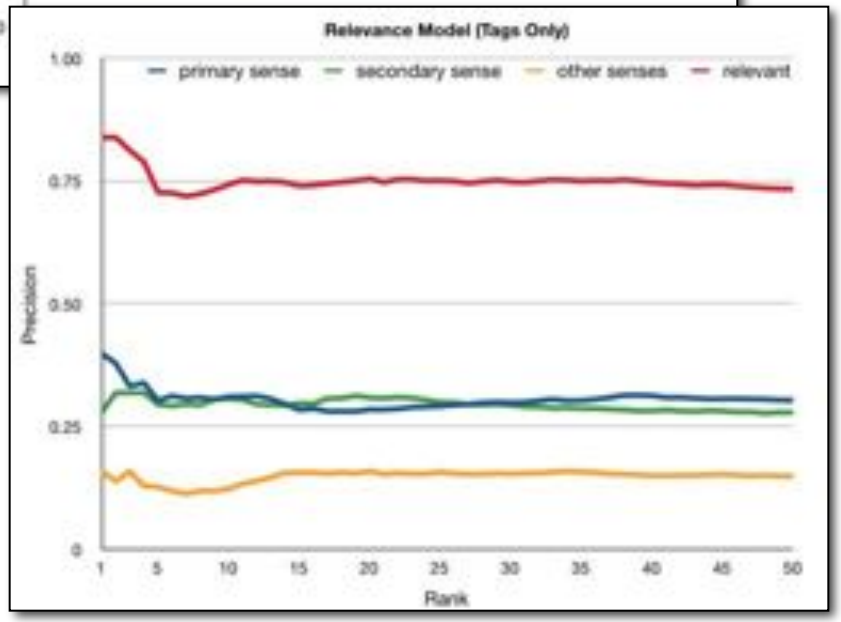


Results


Query Likelihood
primary sense
secondary sense
other senses
relevant
Precision
1.00
0.75
0.50
0.25
0


Query Likelihood (Tags Only)
primary sense
secondary sense
other senses
relevant
Precision
1.00
0.75
0.50
0.25
0


Relevance Model
primary sense
secondary sense
other senses
relevant
Precision
1.00
0.75
0.50
0.25
0
1
5
10
15
20
25
30
35
40
45
50
Category Title


Relevance Model (Tags Only)
primary sense
secondary sense
other senses
relevant
Precision
1.00
0.75
0.50
0.25
0
1
5
10
15
20
25
30
35
40
45
50
Rank

For focussed queries
Visual Diversity

Reinier van Leuken
Lluis Garcia
Ximena Olivares
Roelof van Zwol
WWW'09

Need for Diversification of Results in Image Search

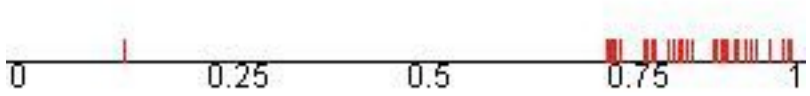
- Image Search on the Web relies on textual information associated with an image
- Textual information is key to retrieve relevant results
- Textual information lacks discriminative power to deliver visually diverse search results
- “Limited” query formulation power

Contributions

- **Dynamic weighting of visual features**
  - **To capture the discriminative aspects of a set of images**

Average color histogram:

Distances to average:



- **Methods for visual diversification of image search results**
  - **Post-retrieval step**
    - We assume relevance of images retrieved is good.
  - **Deploy lightweight clustering methods**
    - Folding -- obey original clustering
    - Maxmin -- maximize the visual diversity, irrespective of ranking
    - Reciprocal election -- images cast votes for other images to be its representative.

Visual Features

- Selection of 6 global features, based on MPEG-7 recommendation:

Color
Color histogram -- *Bhatta Charrya distance*
Color layout – *Angular distance*
Scalable color -- *L1 norm*
Edge
CEDD -- *Tanimoto coefficient*
Edge histogram -- *L1 norm*
Texture
Tamura -- *L2 norm*

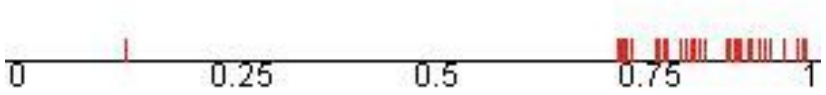
Dynamic Feature Weighting

- In context of a set of images, the relative importance of the different visual features is a-priori undefined
- Depends on the characteristics of the images in the set

Jaguar:

Average color histogram:

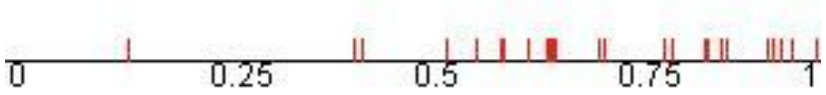
Distances to average:



Fireworks:

Average color histogram:

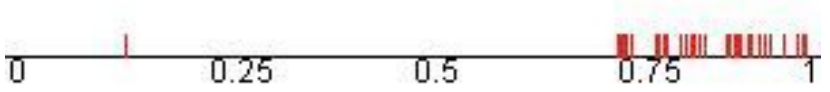
Distances to average:



Dynamic Feature Weighting

Average color histogram:

Distances to average:

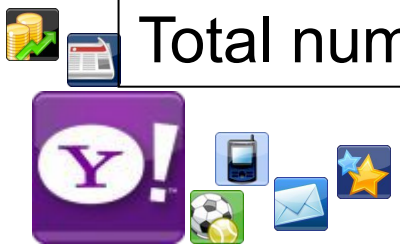


Distance between a and b according to i^{th} feature

$$d(a,b) = \frac{1}{f} \sum_{i=0}^{f} \frac{1}{\sigma_i^2} d_i(a,b)$$

Total number of features

Variance of distances according to i^{th} feature



MaxMin - Folding - Reciprocal Rank

Methods for Diversification

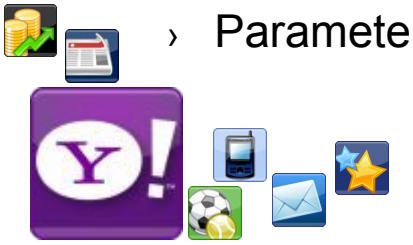
Notation

- A set of images search results *I* contains *n* images
- *I* can be stored in:
  - A ranked list $L=L_1,L_2,\ldots,L_n$, *with decreasing relevance*
  - An unorder set $S=S_1,S_2,\ldots,S_n$
- Methods
  - Input: *L* or *S*
  - Output: a clustering *C* (partitioning of *I*)
  - Images divided over *K* clusters: $C_1, C_2,\ldots,C_k$, with:

$$C_l \mid C_m \neq \varnothing$$

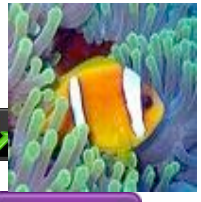
$$\bigcup_{k=1}^{K} C_k = I$$

  - One image is declared cluster representative R_k
  - All representatives together form the set *R*



  - Parameter free -- threshold is set dynamically

Folding

1. 
1. 
2. 
5. 
3. 
7. 
4. 
14. 
5. 
17. 
6. 
20. 

Algorithm 1 Folding

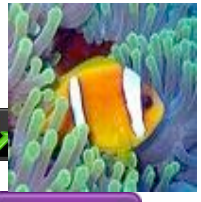
Input: Ranked list L of I

Output: Clustering C

1: Let the image L_1 be the first representative R_1
2: **for** Each image L_i **do**
3: **if** $d(L_i, R_j) > \epsilon(^*)$ for all representatives R_j **then**
4: add L_i to the set of representatives R
5: **for** Each image $L_i \notin R$ **do**
6: Find representative R_j that is closest to L_i
7: Assign L_i to the cluster of R_j

$(^*)\epsilon$ is defined as the mean distance all images have to the *average image* in I

Maxmin

1. 
32. 
2. 
50. 
3. 
16. 
4. 
12. 
5. 
39. 
6. 
9. 

Algorithm 2 Maxmin

Input: Set S containing I

Output: Clustering C

1: Select the first representative R_1 randomly
2: **while** All pairwise distances in $R > \epsilon$ **do**
3: **for** Each image $L_i \notin R$ **do**
4: Let d_i be $\arg\min_{d(L_i,R_j),R_j \in R}$
5: Add to R the image with $\arg\max_{d_i}$
6: **for** Each image $S_i \notin R$ **do**
7: Find representative R_j that is closest to S_i
8: Assign S_i to the cluster of R_j

Reciprocal election

Algorithm 3 Reciprocal election

Input: Set S containing I, parameter m

Output: Clustering C

```
Initialize Votes map V[0, ..., k] = 0, ..., 0
for Each image i in S do
    Rank S into L_i based on visual similarity to i
    for Each image j in L_i do
        V[j] += 1/r, where r is the rank of j in L_i
while V is not empty do
    Let R_i be the item with the highest score in V
    Remove R_i from V
    Initialize new cluster C with representative R_i
    for All items s in V do
        if R_i is in top-m of L_s then
            add s to cluster C
            remove s from V
```

Evaluation

Experimental Setup

Collection: 8.5 Million public photos from Flickr

Topics: 75 queries, top 50 results -- 25 ambiguous queries and 50 unambiguous queries

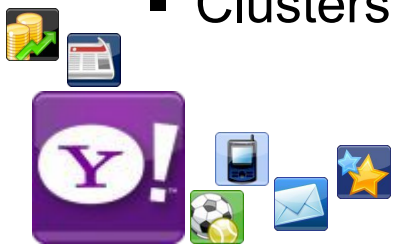
- apple, jaguar, clownfish, tattoo, butterfly, …
- See also: “Diversifying Image Search with User Generated Content” @ ACM MIR 2009

For ambiguous queries:

- Balanced topical diversity
- Clusters to represent word-senses and visual representation

For the un-ambiguous queries:

- Results focused on one topic
- Clusters driven by visual representation only



Experimental Setup

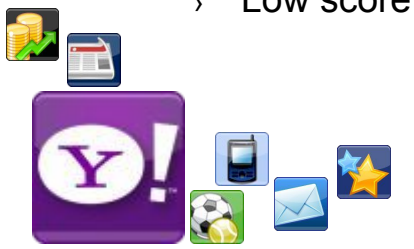
Human assessments:

- 8 independent, unbiased assessors
- Task: “cluster images for a given topic into clusters, based on visual characteristics”.
- Assessment tool:
  - Select topic, and inspect the top 50 results during >1 minute
  - Assign each image to a cluster (max. 20 clusters, undo last action)
  - Label each cluster, and select cluster representative
- 200 human clusterings collected.
- Inter-assessor variability provides baseline for algos

Experimental Setup

Evaluation Criteria

- Objective: Compare the quality of (two) clusterings
- Given a set of images I and two clusterings C and C':
  - N11: image pairs in same cluster under both C and C'
  - N00: image pairs in different cluster under both under C and C'
  - N10: image pairs in same cluster in C but not in C'
  - N01: image pairs in same cluster in C' but not in C
- Fowlkes-Mallows Index:
  - Clustering equivalent of precision/recall
  - High score on FM index indicates cluster similarity
- Variation of Information:
  - Measures the difference in the relationship between a point and a cluster over two clusterings
  - Low score on VI indicates cluster similarity

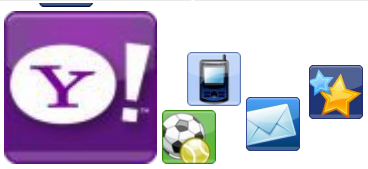


Results – Over All Topics

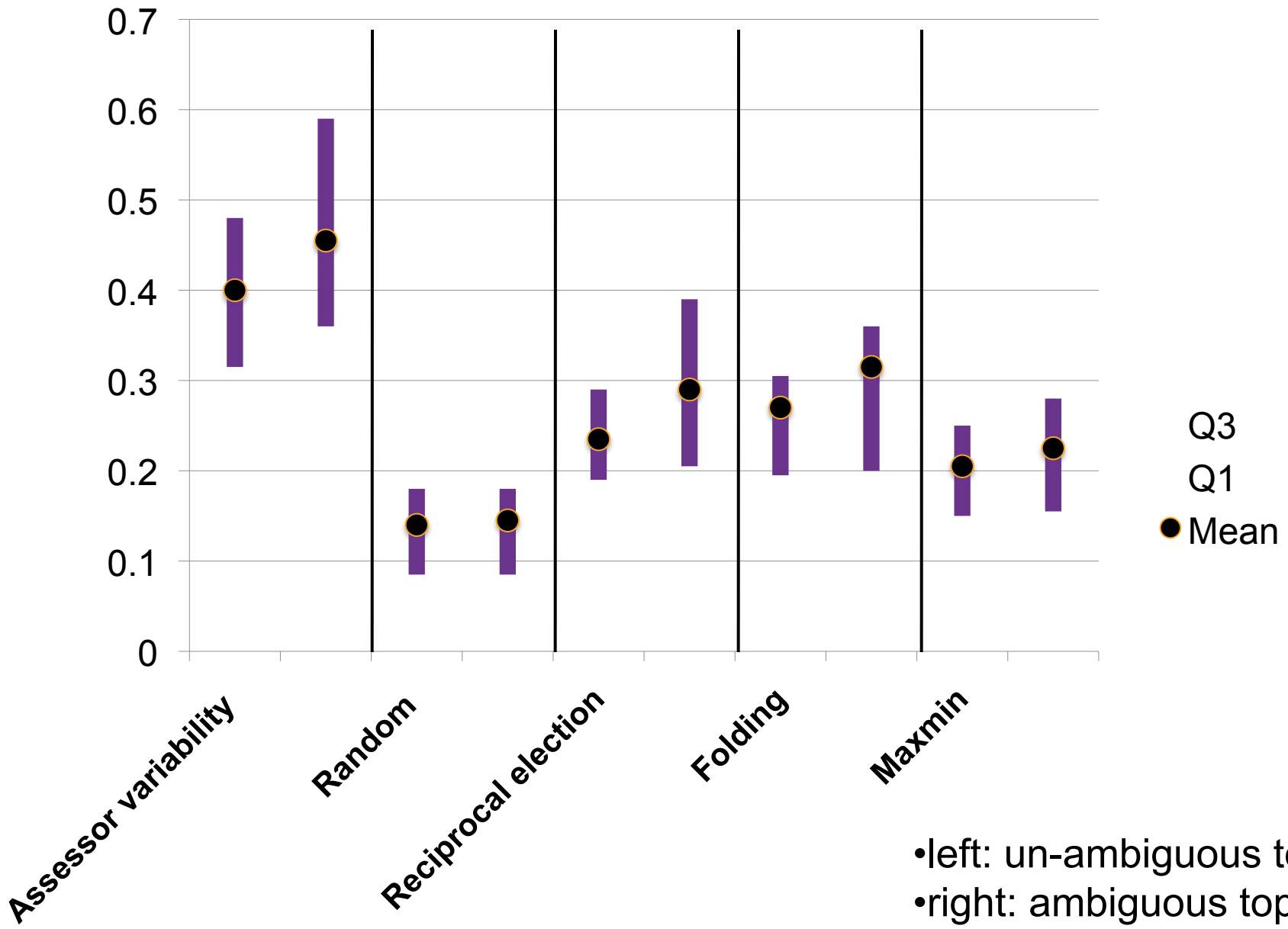
- Performance Bounds:
- Upper-bound: Inter assessor agreement
- Lower-bound: Random clustering

- Overall performance (over all topics):

	Inter assessor variability	Random	Folding	Maxmin	Reciprocal election
FM index	*0.419*	*0.139*	**0.282**	0.214	0.250
VI	*1.463*	*2.513*	2.081	2.129	**1.975**

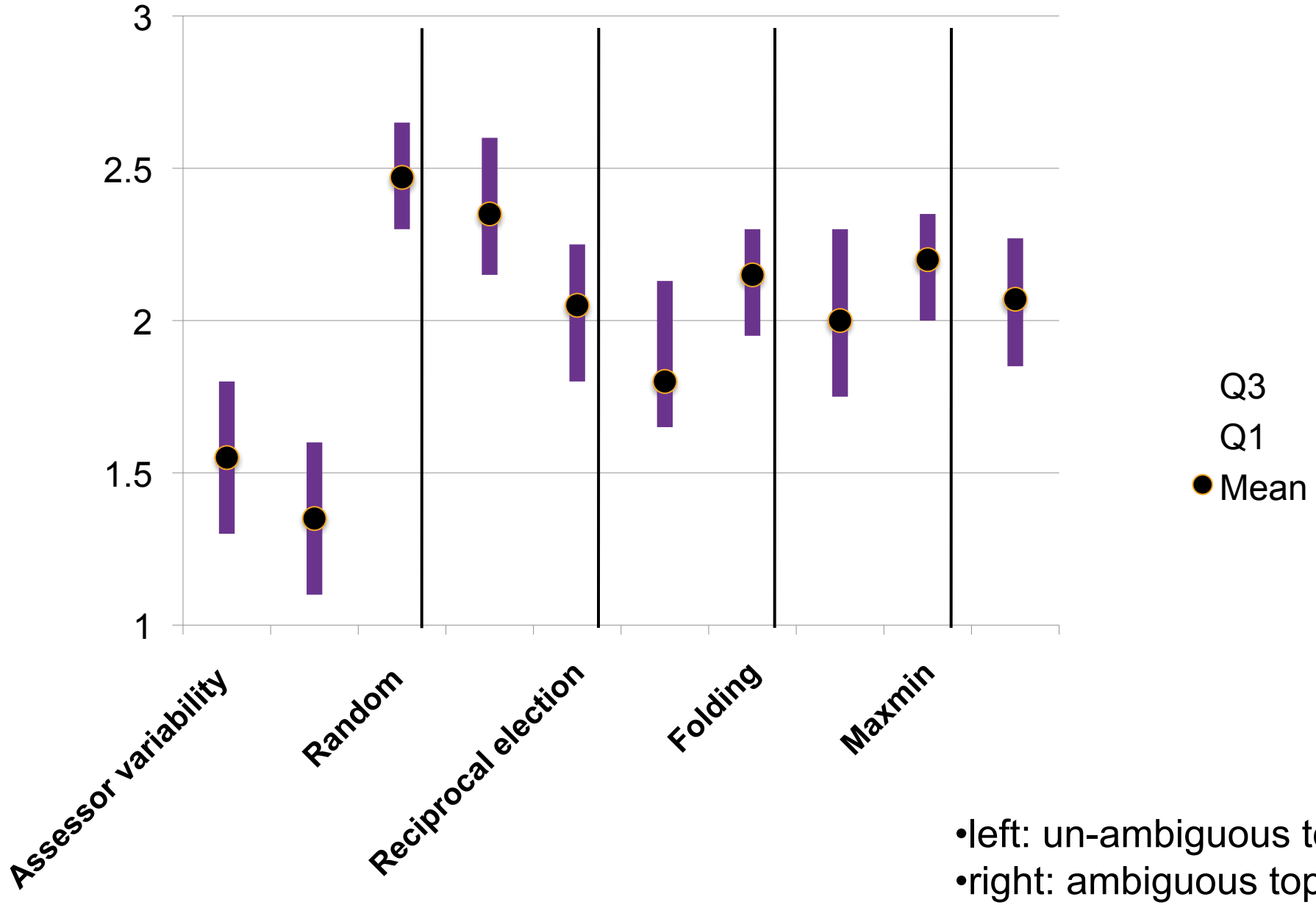


Results – Fowlkes-Mallows Index



- left: un-ambiguous topics
- right: ambiguous topics

Results – Variation of Information



- left: un-ambiguous topics
- right: ambiguous topics

Conclusions

Need for diversification of image search results – topical, visual, etc…

Method for dynamic weighting of (visual) features for a given set of images

Methods for clustering and visual diversification of image search results

- Effective, efficient, no parameters*, no training
- Automatically adopt to characteristics of a set of images

Folding respects ordering of initial ranking

Reciprocal election focuses more on cluster quality

Questions?

More info at:

- http://research.yahoo.com/
- http://sandbox.yahoo.com/

Contact:

- roelof@yahoo-inc.com