National Taiwan University

MSR-Bing Image Retrieval Challenge 2013
Bellevue, USA

Search-Based Relevance Association with Auxiliary Contextual Cues

Chun-Che Wu, Kuan-Yu Chu, Yin-Hsi Kuo, Yan-Ying Chen, Wen-Yu Lee, Winston H. Hsu
NTU MiRA, National Taiwan University, Taiwan

2013.10.07

CMLab
Communications & Multimedia Lab.

Challenge

Constructing image retrieval models from **huge collections**, 23M **history queries, images** and **clicks**, to measure relevance of any **new coming** image-query pairs in an **online** system (< 12 seconds).

Input: image & query

Output: relevance score

suri and
katie cruise

?

How relevant

0.68

Database

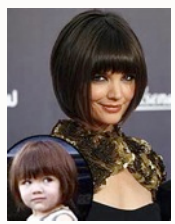
		
fall :113;fall pictures :85;fall leaves :48;fall backgrounds :33;fall images :28;fall foliage :21;fall colors :18;fall pics :16;fall trees :14;autumn images :13	barack obama :414;barak obama :60;barack obama pictures :44;barrack obama :21;presidents :12;pictures of barack obama :3;pictures of barak obama :2;images of barak obama :2;barrak obama :1;barack obama image :1	food :513;food pictures :13;pictures of food :11;food pics :5;picture of a food :4;fast food :3;food images :3;resturant food :2;foood :2;food picture :2

Strategies and Framework

Input:
image-query pair

?

suri and
katie cruise



- Search-Based Approach
- Relevance Propagation
- Name-to-Face Estimation

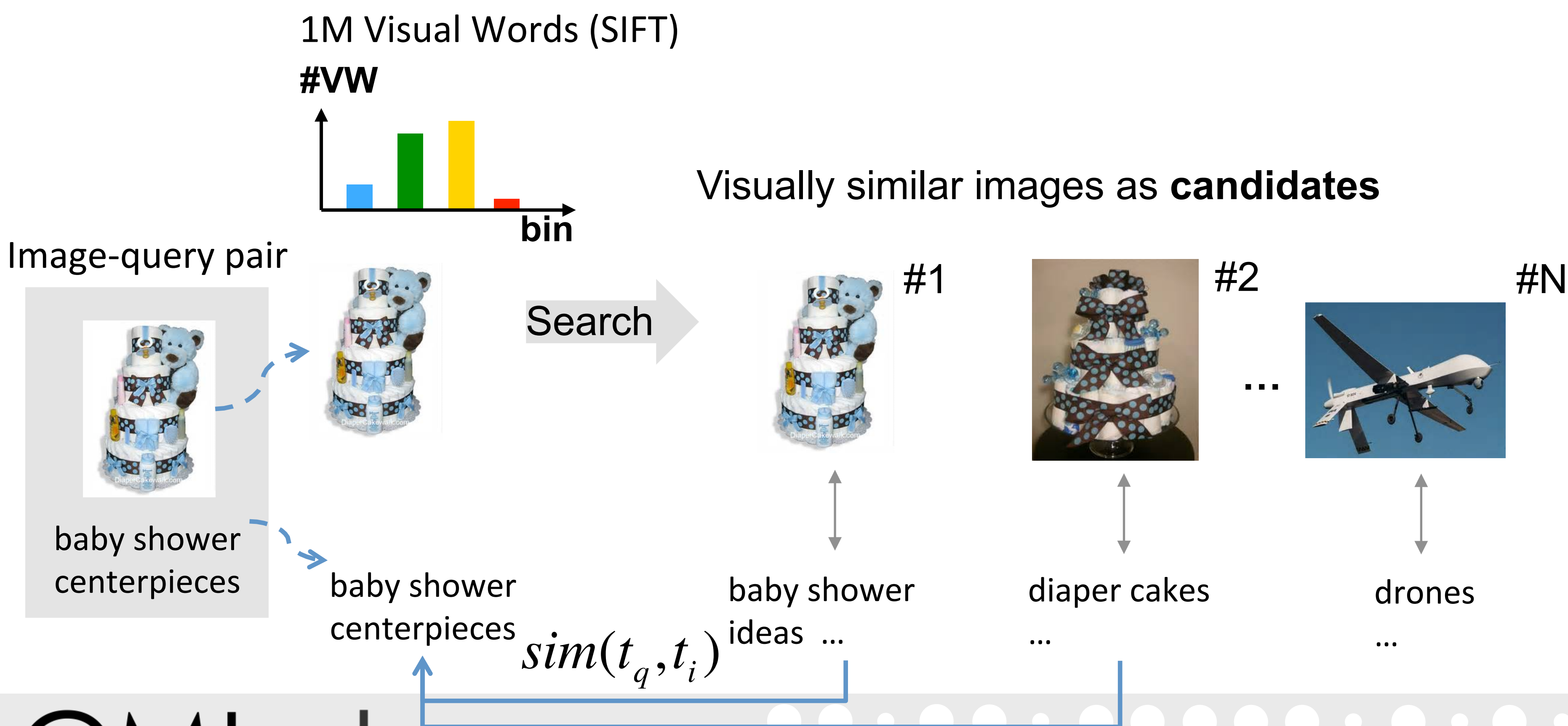
Output:
relevance score 0.68

- Ensuring scalability
- Leveraging prior knowledge
- Strengthening popular queries
- Dealing with unseen data

Evaluated by ranking results of a given query: $DCG_{25} = 0.01757 \sum_{i=1}^{25} \frac{2^{rel_i} - 1}{\log_2(i+1)}$

Search for Reference Candidates

- The query is relevant to the image, if the other visually similar images are associated with similar queries.



Search-Based Approach

- Retrieving candidates by visual (CBIR) & text (TBIR)
- Similarity are weighted by reliability of candidates (#clicks)

$$relevance = \sum_{i \in C} sim(v_q, v_i) \cdot sim(t_q, t_i) \cdot click,$$

q: query
C: top ranked images

Query

baby shower
centerpieces

search

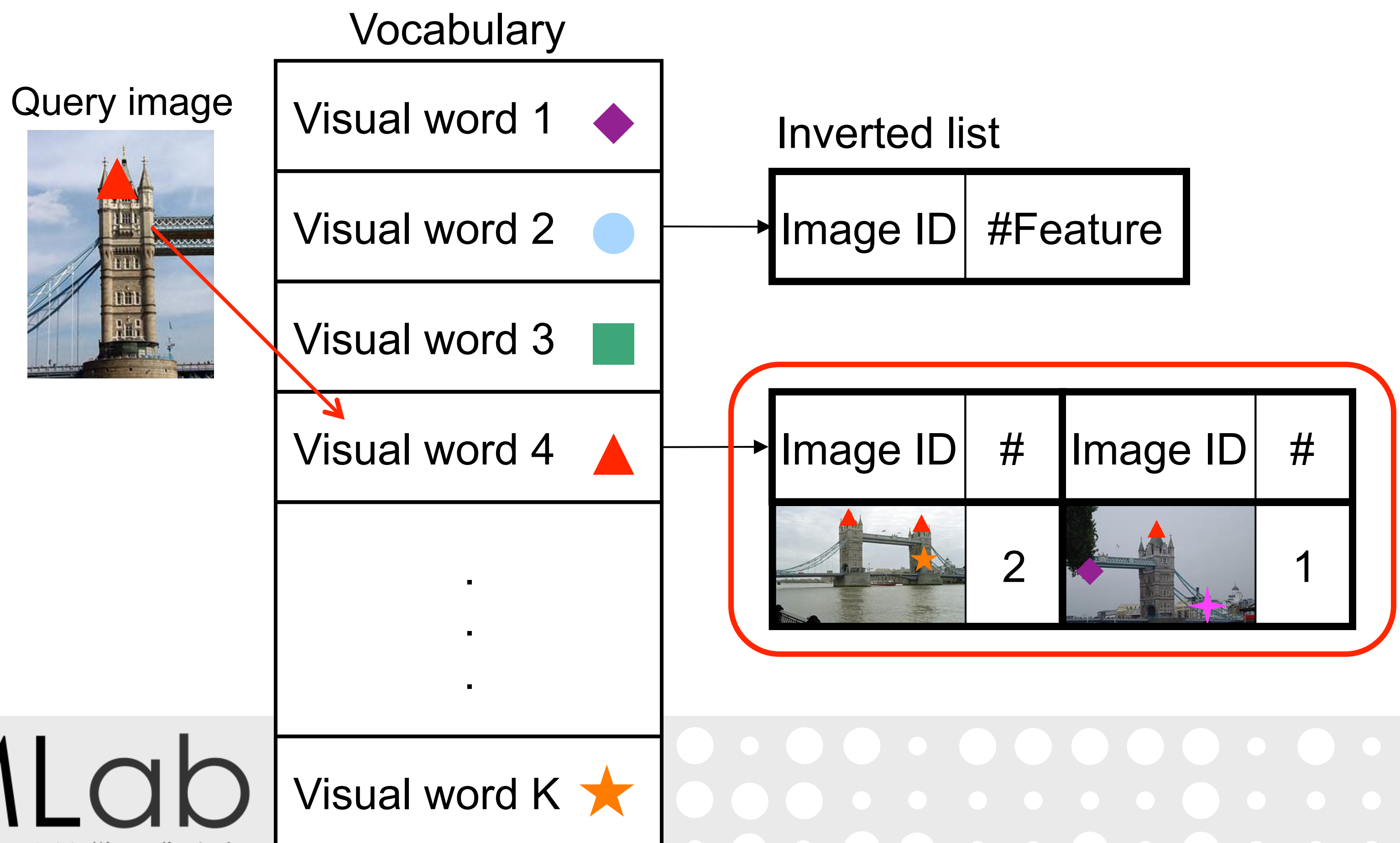
diaper cakes (21)
baby shower ideas (14)

diaper cakes (18)
diaper cake (6)

Method	Initial	Search
DCG@25	0.469	0.484

Speed Up by Indexing

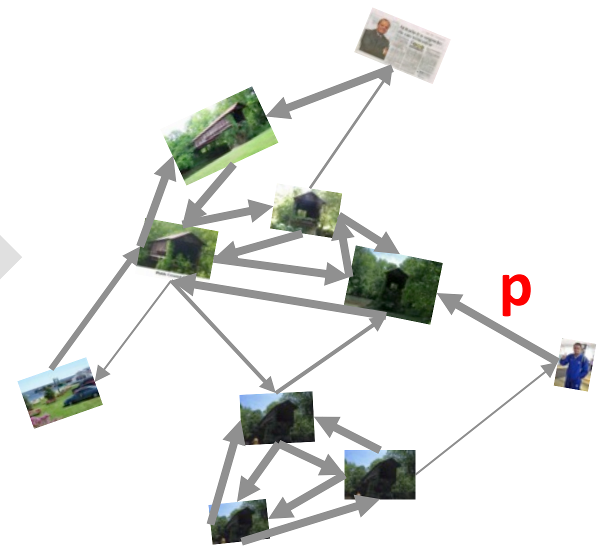
- Why search? Flexible for indexing: inverted-index, KD-tree, …
- Computation & memory cost will not surge with incremental data size -> Scalable (< 1 sec/query)



Majority in Visual Consistency

Images of the query “Waldo Alabama”

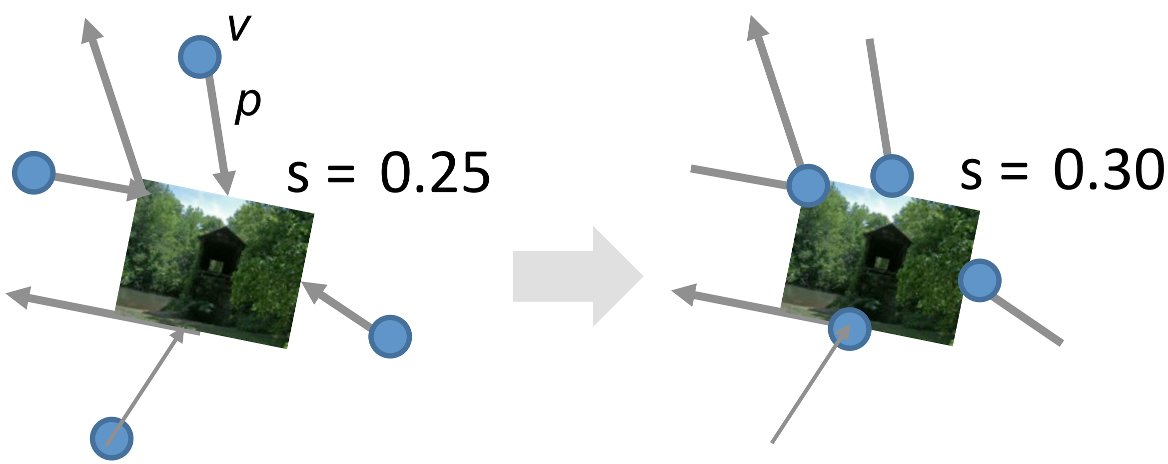
Visual Similarity as transition probability (p)



Visually consistent images have higher relevance to the query

Random Walk

Relevance Propagation by Visual Similarity



$$s = (\alpha P + (1-\alpha) v \mathbf{1}^T) s,$$

$$\text{where } P(i,j) = sim(i,j) / \sum_i sim(i,j)$$

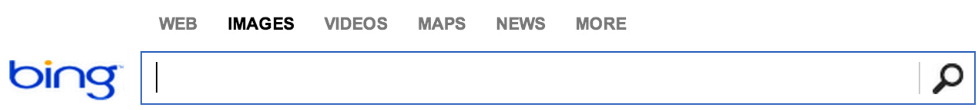
s: Score,
v: Prior knowledge,
P: Transition probability

15.78% relative improvement

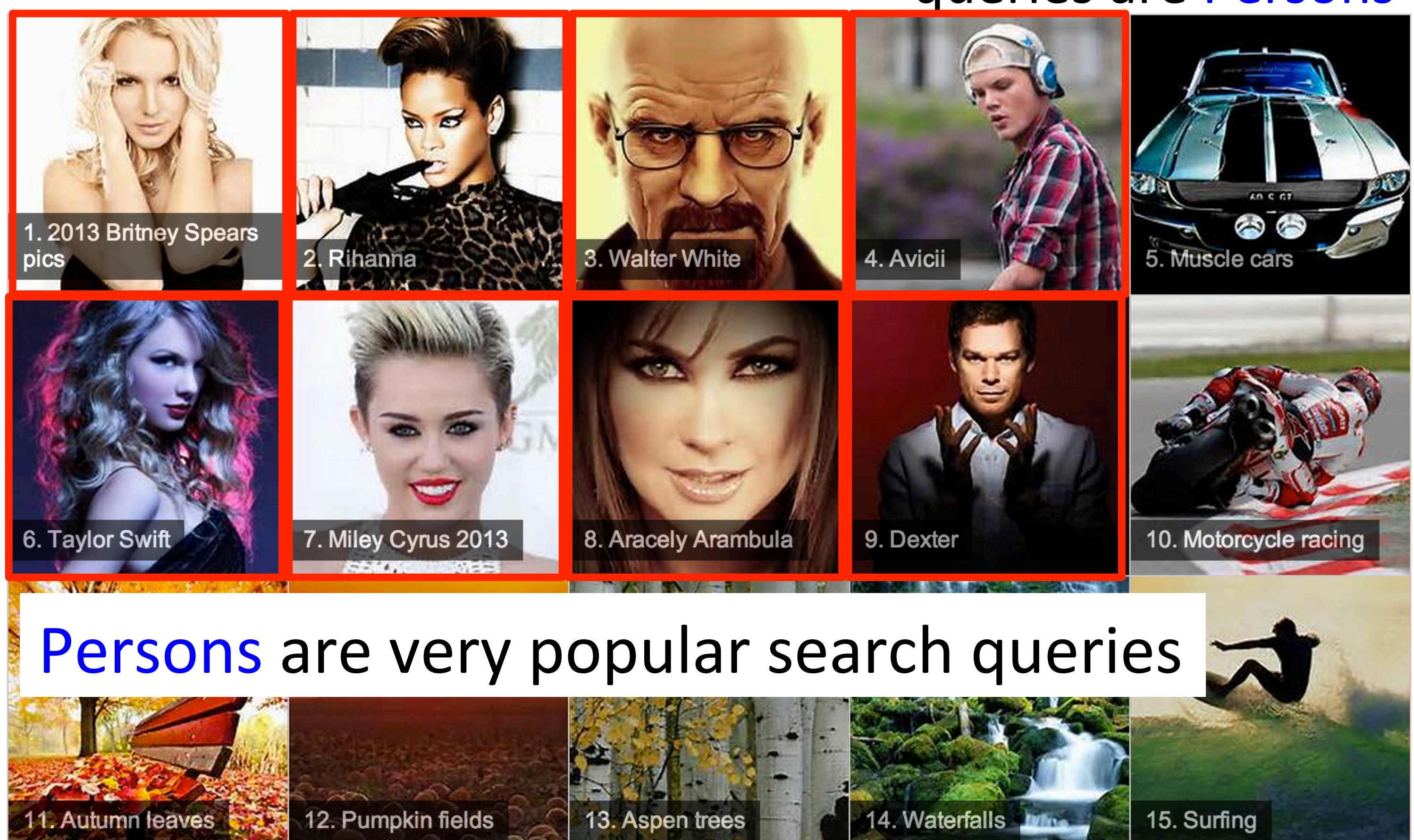
Method	Initial	Search	Propagation
DCG@25	0.469	0.484	0.543

Popular Queries in Image Search

- 8/15 trending image searches* are Persons
- 31.8% Bing dev image queries are Persons



Trending image searches



Persons are very popular search queries

* Bing trending image searches on Sep. 24

Learning Identity Classifiers by Names and Faces

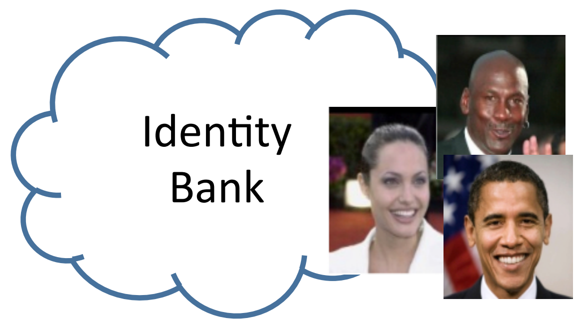
- Name Detection:
  - *Collections of celebrity names (2,221)
  - Mutual combinations of First-Name (1,164) and Last-Name (1,681)
- Identity Bank: 6,762 identity models
  - Training by 35,092 face-name pairs in Bing training image set
- Challenges
  - Accuracy
  - Persons out of Identity Bank

Queries

Tyler Swift album

highlights **Michael Jordan**

Barack Obama election

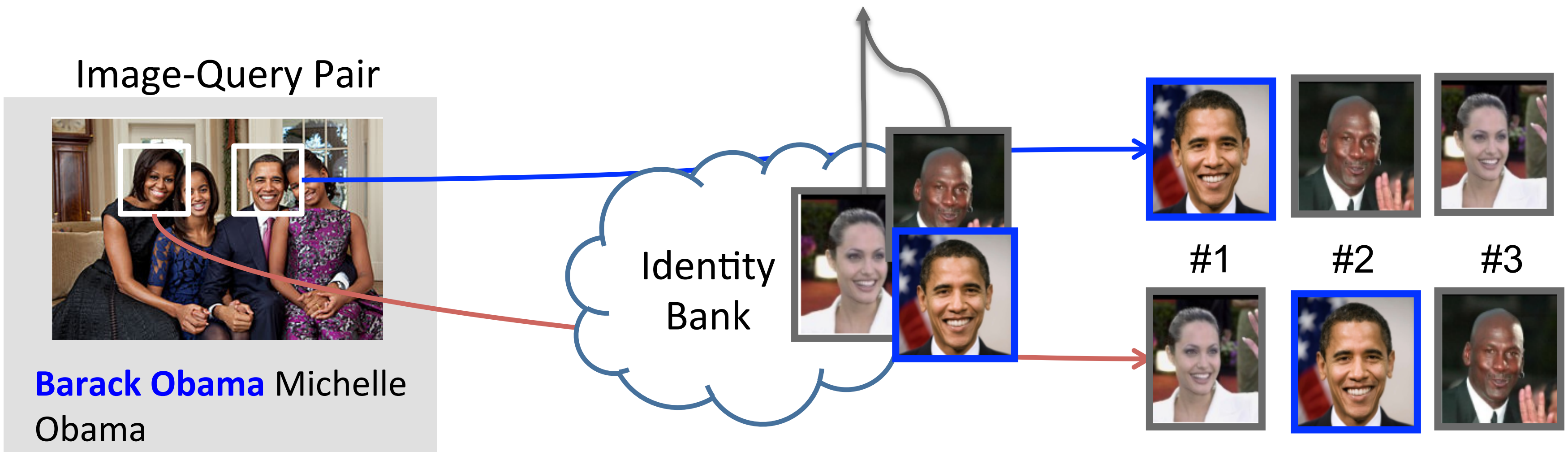


*http://www.posh24.com/celebrities/

Improve Classification Accuracy

- Identity Bank with background models (IB)

Randomly selected background models → Hypothesis testing statistically significant?



s: classification score by the model of detected name → Adjust **s** by rank

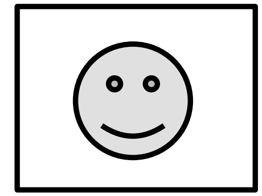
$$s' = \frac{s}{r}$$

	Initial	IB
DCG@25 (I)	0.352	0.496

(I): names included in Identity Bank

Persons out of Identity Bank

- Face number as reference (FN) for smoothing
  - Once a name is detected in a tag, the associated image should comprise at least one face.
  - More names are detected, more faces are expected to appear in image content.



Dave xxx

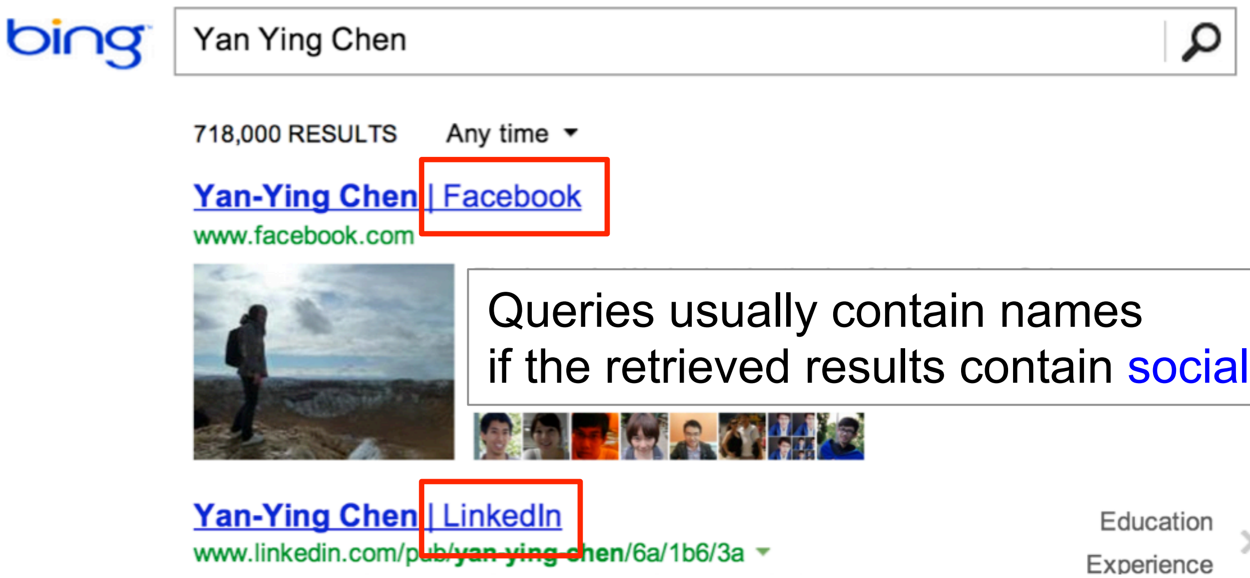
45.71% relative improvement

	Initial	IB	FN	IB+FN
DCG@25 (N)	0.481	0.496	0.508	0.516
DCG@25 (I)	0.352	0.496	0.500	0.510

(N): partial names appear in name list
(I): names included in Identity Bank

Unseen Names: More Auxiliary Cues

- Detect names out of name list (non-celebrities)



Queries usually contain names
if the retrieved results contain social websites

	Name List	Facebook	LinkedIn	Twitter
Precision	0.967	0.73	0.959	0.795
Recall	0.279	0.585	0.438	0.509

> 20% improvement in recall of name detection

Ongoing and Future Work

- Seeing the unseen data
  - Tag expansion: Snippet, Hash tags
  - Topic modeling
- Describing the unseen data
  - Embedded contexts: Angelina (female) vs. Michael (male)
  - Attributes
- Search or Classification? Query-dependent strategy
  - General queries: search-based
    - Few training data; minor improvement & less scalability by learning
  - Trending queries: classification-based
    - Rich training data; significant improvement & better user experiences by learning

Summary

- We propose an image-query relevance measurement approach considering four major strategies,
  - Ensuring scalability
  - Leveraging prior knowledge
  - Strengthening popular queries
  - Dealing with unseen data
- We demonstrate the efficiency (< 2 seconds/query) and 16% relative improvement in DCG compared to the original ranking results

Acknowledgements

- Thanks for Microsoft Research and Bing for organizing the challenging and great event!

- With the research supports in part by
  - Quanta Research Institute (**cloud cluster**)
  - Chunghwa Telecom Laboratories
  - Microsoft Research Asia

Thank you for your attention

NTU MiRA, National Taiwan University, Taiwan