

Baseline approach to MSR-Bing Challenge on Image Retrieval

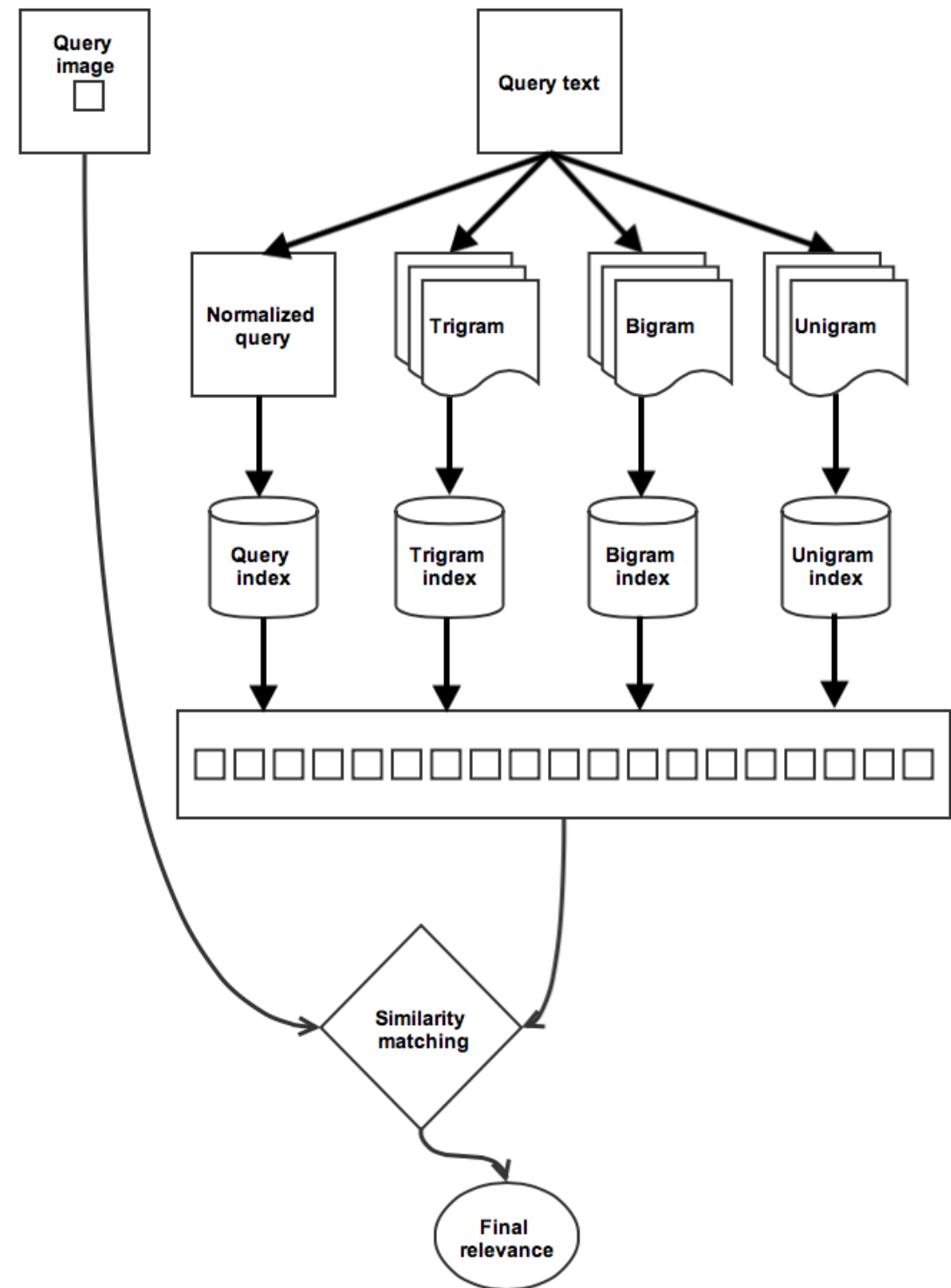
Aleksandr Sayko, Anton Slesarev
7 October 2013

Plan

- Approach overview
- Click-relevant images index
- Query expansion and examples fetching
- Similarity matching
- Conclusion

System overview

- Index on click log
- Text query processing
- Similarity matching using top 100 pics



Clicked images indices

highchair birthday party	img42	3	
highchair birthday party	img7	4	
arseny's birthday party	img7	1	

Normalized query index	Unigram index	Bigram index	Trigram index
birthday_highchair_party (img42; 3; 1) (img7; 1; 1)	highchair (img42; 3; 1) (img7; 1; 1)	birthday_highchair (img42; 3; 1) (img7; 4; 1)	birthday_highchair_party (img42; 3; 1) (img7; 1; 1)
arseny_birthday_party (img7; 1; 1)	birthday (img42; 3; 1) (img7; 5; 2)	highchair_party (img42; 3; 1) (img7; 4; 1)	arseny_birthday_party (img7; 1; 1)
	party (img42; 3; 1) (img7; 5; 2)	birthday_party (img42; 3; 1) (img7; 5; 2)	
	arseny (img42; 3; 1) (img7; 1; 1)	arseny_birthday (img7; 1; 1)	
		arseny_party (img7; 1; 1)	

Clicked images issues

Retrieval artifacts



Distracting images



Query expansion

Text query: “highchair party”.

Synsets of “highchair”: “feeding chair”...

Synsets of “party”: “political company” ...

Expanded query: “highchair party” OR

“feeding chair party” OR “highchair

political company” OR “feeding chair

political company”

Ranking ideas

birthday_party

(img42; 3; 1) (img7; 5; 2)

- Query match is better than trigram match, trigram is better than bigram...
- Ngram match is considered more relevant when there are more clicks and other queries with the ngram for which the image was clicked.

Examples fetching

dunk competition img_id: default, img_file:ZGVmYXVsdA==.jpeg Relevance: default

/9j/4AAQSkZJRgABAQEAAAAAAAAAD/2wBDAAoHBwgHBgoICAgLCgoLDhgQDg0NDh



Rzn2Wi26WuUSaA, 41.472517



RxvfG++xVpxGCw, 39.090944



h0kuQfqoK527lA, 33.133415

grand challenge on image retrieval img_id: default, img_file:ZGVmYXVsdA==.jpeg Relevance: default

/9j/4AAQSkZJRgABAQEAAAAAAAAAD/2wBDAAoHBwgHBgoICAgLCgoLDhgQDg0NDh0VFhEYIx8lJC



6yMZsK/i3MW+iw, 117.865128



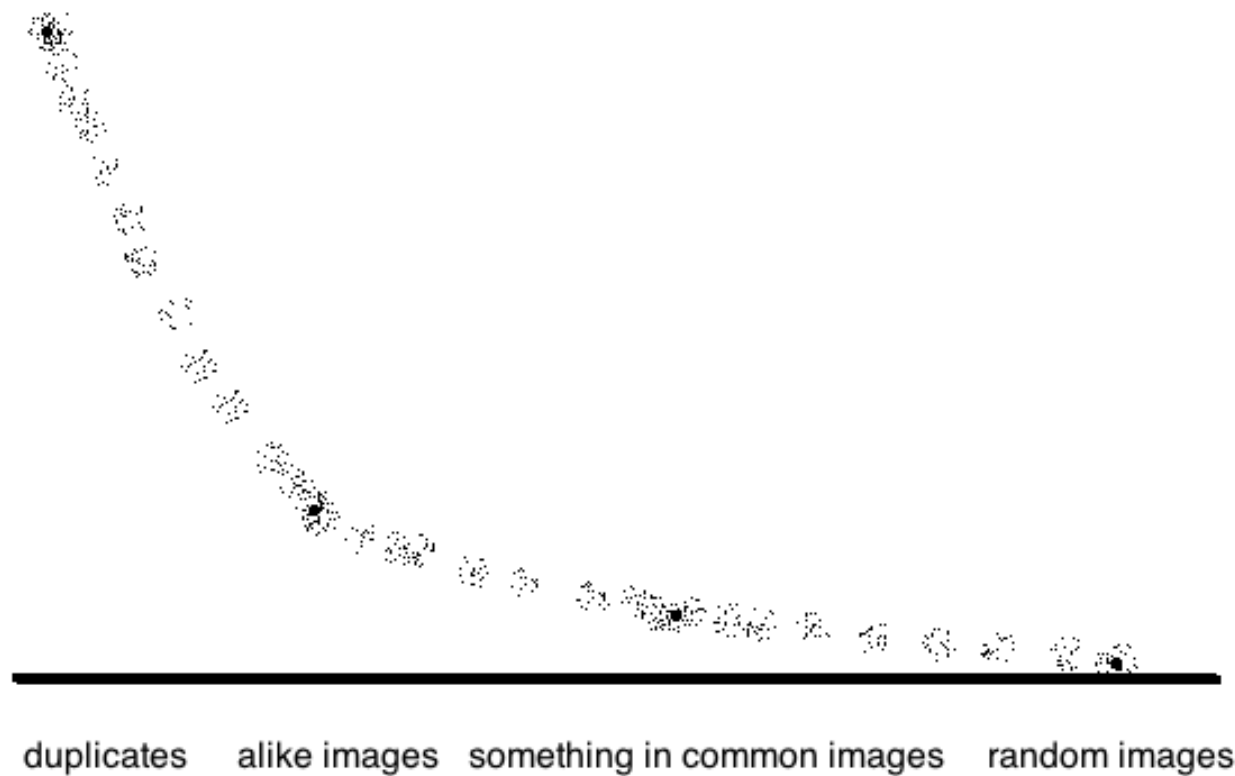
B97VFbaYXvgQRw, 81.910159



UKI1iGOISHZk4w, 68.039194

Final relevance estimation

$$\sum_{Img \in ExtractedImages} SimScore(Img, QueryImg)$$



Similarity matching

- Tentative matches using Lowe's algorithm
- Spatial verification exploiting RANSAC

Visual similarity score

$$\frac{I}{\sqrt{N_1 N_2}}$$

I - the number of inliers

N_1, N_2 - number of descriptors on the first and the second images

Performance issues on calculating SimScore

- Filter out non-similar candidates
- L2 distance on histograms of visual words + threshold
- Visual words vocabulary size is 32k

Conclusion

- 0.486 target metric, still baseline
- Clicks contain information (still noisy) on relevance
- Visual similarity gives the opportunity to propagate the information on new images