



MediaEval Benchmark 2015

MediaEval Benchmarking Initiative for Multimedia Evaluation

The "multi" in multimedia: speech, audio, visual content, tags, users, context

Retrieving Diverse Social Images Task

- task overview -

Bogdan Ionescu (UPB, Romania)

Adrian Popescu (CEA LIST, France)

Mihai Lupu (TUW, Austria)

Henning Müller (HES-SO in Sierre, Switzerland)



University Politehnica
of Bucharest



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology



September 14-15, Wurzen, Germany

Outline

- The Retrieving Diverse Social Images Task
- Dataset and Evaluation
- Participants
- Results
- Discussion and Perspectives

Diversity Task: Objective & Motivation

Objective: image search **result diversification** in the context of *social photo retrieval*.



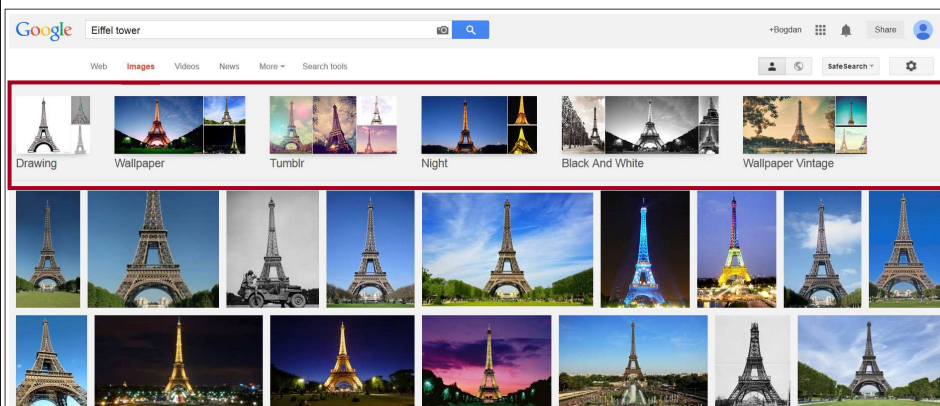
Why diversifying search results?

- to respond to the needs of different users;
- as a method of tackling queries with unclear information needs;
- to widen the pool of possible results (increase performance);
- to reduce the number/redundancy of the returned items;
- ...

3

Diversity Task: Objective & Motivation #2

The concept appeared initially for text retrieval but regains its popularity in the context of multimedia retrieval:



[Google Image Search ("Eiffel tower"), >2014]

4

Diversity Task: Use Case

To disambiguate the diversification need, we introduced a very focused use case scenario ...

Use case: we consider a **tourist use case** where a person tries to find more information about a place she is potentially visiting. The person has only a vague idea about the location, knowing the name of the place.

... e.g., looking for **Rialto Bridge** in Italy

Diversity Task: Use Case #2

... learn more information from Wikipedia

Rialto Bridge

From Wikipedia, the free encyclopedia

The **Rialto Bridge** (Italian: **Ponte di Rialto**) is one of the four bridges spanning the Grand Canal in Venice, Italy. It is the oldest bridge across the canal, and was the dividing line for the districts of San Marco and San Polo.

Contents	[hide]
1	History
2	See also
3	References
4	External links

History



Two gondoliers pull out with clients on board from a row of gondolas on the Grand Canal near Rialto Bridge.



Detail of the bridge

The first dry crossing of the Grand Canal was a pontoon bridge built in 1181 by Nicolò Barattieri. It was called the Ponte della Moneta, presumably because of the mint that stood near its eastern entrance.^[2]

The development and importance of the Rialto market on the eastern bank increased traffic on the floating bridge, so it was replaced in 1255 by a wooden bridge.^[2] This structure had two inclined ramps meeting at a movable central section, that could be raised to allow the passage of tall ships. The connection with the market eventually led to a change of name for the bridge. During the first half of the 15th century, two rows of shops were built along the sides of the bridge. The rents brought an income to the State Treasury, which helped maintain the bridge.

Maintenance was vital for the timber bridge; it was partly burnt in the revolt led by Bajamonte Tiepolo in 1310. In 1444, it collapsed under the weight of a crowd watching a boat parade and it collapsed again in 1524.

The idea of rebuilding the bridge in stone was first proposed in 1503. Several projects were considered over the following decades. In 1501, the authorities requested proposals for the renewal of the Rialto Bridge, among other things. Plans were offered by famous architects, such as Jacopo Sansovino, Palladio and Vignola, but all involved a Classical approach with several arches, which was judged inappropriate to the situation. Michelangelo also was considered as designer of the bridge.

The present stone bridge, a single span designed by Antonio da Ponte, was finally completed in 1591. It is similar to the wooden bridge it succeeded. Two inclined ramps lead up to a central portico. On either side of the portico, the covered ramps carry rows of shops. The engineering of the bridge was considered so audacious that architect Vincenzo Scamozzi predicted future ruin. The bridge has defied its critics to become one of the architectural icons of Venice.

See also

Miracle of the Relic of the Cross at the Ponte di Rialto (depiction of wooden bridge)

References

- ^[1] Fulton, Charles Carroll (1874). *Europe Viewed Through American Spectacles* (Google books). Philadelphia: J.P. Lippincott & Co. p. 242. Retrieved 2008-09-05. "There being no vehicles or horses in Venice, it is simply for pedestrians."
- ^[2] ^a ^b Momenti, Pompeo; Horatio Forbes Brown (1906-10-13). *Venice: Its Individual Growth from the Earliest Beginnings to the Fall of the Republic* (Google books). Chicago: A.C. McClurg & Co. p. 29. Retrieved 2008-09-05.



WIKIPEDIA
The Free Encyclopedia

Coordinates: 45°43′03″N 12°33′58″E﻿ / ﻿45.71750°N 12.56611°E﻿ / 45.71750; 12.56611



The Rialto Bridge

Carries	pedestrian bridge ^[1]
Crosses	Grand Canal
Locale	Venice, Italy
Design	stone arch bridge
Width	22.80 metres (75 ft)
Height	7.32 metres (24 ft)
Longest span	28.80 metres (94 ft)
Construction	1588
Begin	
Construction end	1591
Coordinates	45°43′03″N 12°33′58″E﻿ / ﻿45.71750°N 12.56611°E﻿ / 45.71750; 12.56611



Diversity Task: Use Case #3

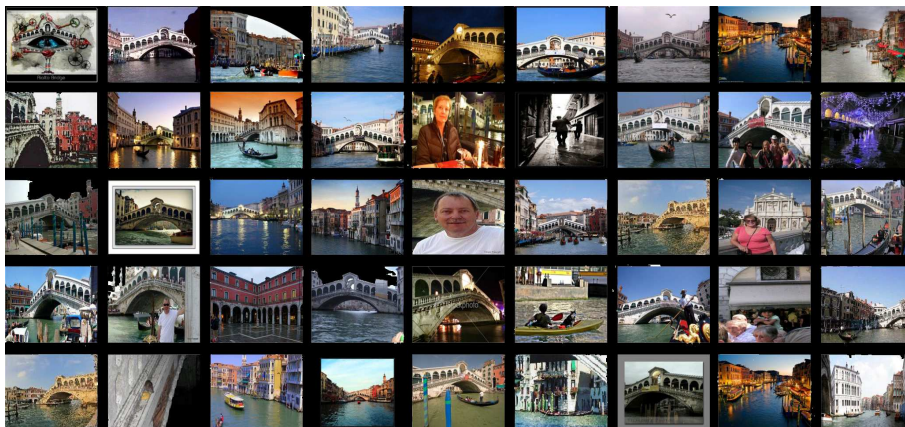
... how to get some more accurate photos ?



query using text "Rialto Bridge" ...

... browse the results

Diversity Task: Use Case #4



Diversity Task: Use Case #5



page n

Diversity Task: Use Case #6

... too many results to process,

inaccurate, e.g., people in focus, other views or places



meaningless objects



redundant results, e.g., duplicates, similar views ...



Diversity Task: Use Case #7



page 1

Diversity Task: Use Case #8



page n

Diversity Task: Definition

Participants receive a *ranked list* of photos with locations retrieved from Flickr using its default “relevance” algorithm.

Goal of the task: *refine* the results by providing a *ranked* list of *up to 50 photos (summary)* that are considered to be both *relevant* and *diverse* representations of the query.

relevant*: a common photo representation of the query concepts: sub-locations, temporal information, typical actors/objects, genesis information, and image style information;

diverse*: depicting different visual characteristics of the target concepts, e.g., sub-locations, temporal information, etc with a certain degree of complementarity, i.e., most of the perceived visual information is different from one photo to another.

*we thank the task survey respondents for their precious feedback on these definitions.

13

Dataset: General Information & Resources

~300 location queries – **single-topic** (e.g., "Aachen Cathedral") + 70 queries related to events and states associated with locations – **multi-topic** (e.g., "Oktoberfest in Munich");

Provided information:

- query text formulation & GPS coordinates;
- links to Wikipedia web pages;
- up to 5 representative photos from Wikipedia;
- ranked set of Creative Commons photos from Flickr* (up to 300 photos per query);
- metadata from Flickr (e.g., tags, description, views, #comments, date-time photo was taken, username, userid, etc);
- visual, text & user annotation credibility descriptors;
- relevance and diversity ground truth.

14

Dataset: Provided Descriptors

General purpose visual descriptors (color/texture/feature):

- e.g., color histograms, Histogram of Oriented Gradients, Locally Binary Patterns, Color Moments, etc;

Convolutional Neural Network based descriptors:

- Caffe framework based;

General purpose text descriptors:

- e.g., term frequency information, document frequency information and their ratio, i.e., TF-IDF;

User annotation credibility descriptors (give an automatic estimation of the quality of users' tag-image content relationships):

- e.g., measure of user image relevance, the proportion of bulk taggings in a user's stream, the percentage of images with faces.

15

Dataset: Basic Statistics

- **devset** (designing and validating the methods)

<i>#single-topic</i>	<i>#images</i>	<i>min-average-max #img. per query</i>
153	45,375	281 - 297 - 300

+ 12M images* via credibility information.

- **credibilityset** (training/designing credibility desc.)

<i>#single-topic</i>	<i>#images*</i>	<i>#users</i>	<i>average #img. per user</i>
300	3,651,303	685	5,330

- **testset** (final benchmarking)

<i>#single-topic</i>	<i>#images</i>	<i>min-average-max #img. per query</i>
69	20,700	300 - 300 - 300

<i>#multi-topic</i>	<i>#images</i>	<i>min-average-max #img. per query</i>
70	20,694	176 - 296 - 300

+ 15M images* via credibility information.

* images are provided via Flickr URLs.

16

Dataset: Ground Truth - annotations

Relevance and diversity annotations were carried out by **expert annotators***:

- **devset**: relevance (3 annotations from 11 experts), diversity (1 annotation from 3 experts + 1 final master revision);
- **testset**: relevance single-topic (3 annotations from 7 experts), relevance multi-topic (3 annotations from 5 experts), diversity (1 annotation from 3 experts + 1 final master revision);
- **credibilityset**: only relevance for 50,157 photos (3 annotations from 9 experts);
- lenient majority voting for relevance.

* advanced knowledge of location characteristics mainly learned from Internet sources.

17

Dataset: Ground Truth #2 - basic statistics

- **devset**:

relevance	% relevant img. 68.5	
diversity	avg. #clusters per location 23	avg. #img. per cluster 8.9

- **credibilityset**:

relevance	% relevant img. 69
-----------	-----------------------

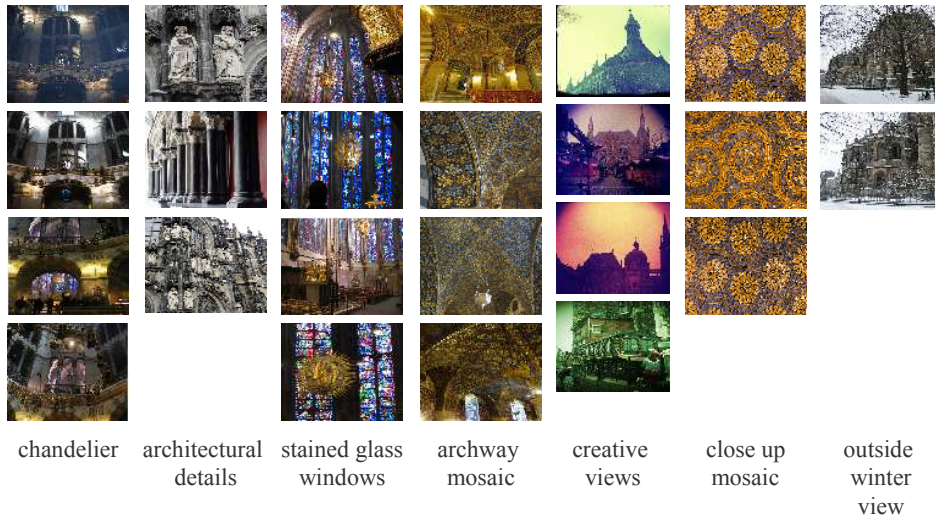
- **testset**:

relevance	% relevant img. 66	
diversity	avg. #clusters per location 19	avg. #img. per cluster 10.8

18

Dataset: Ground Truth #3 - example

Aachen Cathedral, Germany:



19

Evaluation: Required Runs

Participants are required to submit up to 5 runs:

- **required runs:**

- run 1: automated using *visual information only*;
- run 2: automated using *textual information only*;
- run 3: automated using *textual-visual* fused without other resources than provided by the organizers;

- **general runs:**

- run 4: automated using *credibility information*;
- run 5: *everything allowed*, e.g., human-based or hybrid human-machine approaches, including using data from external sources (e.g., Internet).

20

Evaluation: Official Metrics

- **Cluster Recall* @ X = N_c/N (CR@X)**

where X is the cutoff point, N is the total number of clusters for the current query (from ground truth, $N \leq 25$) and N_c is the number of different clusters represented in the X ranked images;

- **Precision @ X = R/X (P@X)**

where R is the number of relevant images;

- **F1-measure @ X = harmonic mean of CR and P (F1@X)**

official ranking F1@20

Metrics are reported for different values of X (5, 10, 20, 30, 40 & 50) on per location basis as well as overall (average).

*cluster recall is computed only for the relevant images.

21

Participants: Basic Statistics

- **Survey (February 2015):**

- 83 (66/55) respondents were interested in the task, 33 (26/23) very interested;

- **Registration (May 2015):**

- 24 (20/24) teams registered from 18 (15/18) different countries (3 teams are organizer related);

- **Crossing the finish line (August 2015):**

- 14 (14/11) teams finished the task, 11 (12/8) countries, including 3 organizer related teams (1 late submission);
- 59 (54/38) runs were submitted from which 1 (1/2) **brave human-machine!**

- **Workshop participation (September 2015):**

- 10 (10/8) teams are represented at the workshop.

* the numbers in the brackets are from 2014/2013.

22

Participants: Submitted Runs (59)

team	country	1-visual	2-text	3-text-visual	4-cred.	5-free
BME-DCLAB	Hungary	✓	✓	✓	✓	x
CSLU*	United States	✓	✓	x	✓	visual-text-cred.
ETH-CVL	India-Switzerland	✓	✓	✓	x	x
Imcube	Germany	✓	✓	✓	x	visual-cred
LAPI**	Romania	✓	✓	✓	✓	CNN
MIS	Austria	✓	✓	✓	x	visual-text
PRa-MM	Italy	✓	✓	✓	✓	visual-text
RECOD	Brazil	✓	✓	✓	✓	visual-text-cred.
SAILUSC	United States	✓	x	✓	x	visual-text
sigma2320	China	✓	✓	✓	✓	CNN-text-cred.
TUW**	Austria	✓	✓	✓	✓	text-visual
UNED	Spain	✓	✓	✓	x	text-visual-human
UPC-UB-STP	Spain-Austria	✓	✓	✓	x	visual-text
USEMP**	Greece-France	✓	✓	✓	x	visual-text-CNN

*late submission; **organizer related team.

23

Results: Official Ranking (F1@20; all)

> team best runs only (full ranking will be sent via email);

team/run	P@10	P@20	CR@10	CR@20	F1@10	F1@20
TUW_run3	0.7633	0.7309	0.3163	0.4963	0.4309	0.5727
USEMP_run2sMMRtextual	0.7583	0.7788	0.2913	0.4483	0.4089	0.5494
PRaMM_run5	0.759	0.7486	0.2934	0.4522	0.4118	0.5481
MIS_run3	0.8022	0.7784	0.2784	0.445	0.3985	0.5473
ETH-CVL_run1_Visual	0.7201	0.6853	0.3037	0.4724	0.4185	0.5453
UNED_run5RelFeedback	0.7856	0.7766	0.2909	0.4344	0.4103	0.538
LAPI_HC-RF_run4_credibility	0.7043	0.7126	0.2824	0.449	0.3891	0.5336
sigma2320_run5	0.718	0.6957	0.287	0.4435	0.3988	0.5241
RECOD_run4	0.7058	0.7198	0.2744	0.4309	0.3829	0.5219
IMCUBE_run2_text	0.7647	0.7478	0.2813	0.4182	0.3967	0.5195
SAILUSC_VisualRun5	0.7158	0.705	0.2715	0.4158	0.3824	0.5093
UPC-STP_run3_A_20	0.6633	0.6878	0.2766	0.422	0.3776	0.5076
bmedclab_run1	0.7058	0.7094	0.2202	0.378	0.3246	0.4782
Flickr initial results	0.7086	0.7	0.2402	0.3684	0.3466	0.4672
cslu_f1a_run4	0.595	0.6029	0.2444	0.3933	0.3353	0.46

Best improvements compared to Flickr (percentage points): P@20 - 3, CR@20 - 13.

24

Results: Official Ranking (F1@20; single-topic)

> team best runs only (full ranking will be sent via email);

team/run	P@10	P@20	CR@10	CR@20	F1@10	F1@20
USEMP_run3sMMRvisual+textual	0.8435	0.8333	0.3177	0.5044	0.4507	0.6177
TUW_run3	0.7928	0.7326	0.3237	0.5037	0.4443	0.5802
ETH-CVL_run3_VisualText	0.871	0.842	0.291	0.442	0.4249	0.5674
MIS_run3	0.8696	0.7935	0.2908	0.4546	0.418	0.5595
IMCUBE	0.8043	0.7819	0.2874	0.438	0.41	0.5478
LAPI_HC-RF_run2_text_TF	0.7681	0.7391	0.2778	0.4392	0.3978	0.5402
RECOD_run4	0.7551	0.763	0.2659	0.4301	0.3879	0.539
UPC-STP_run3_A_20	0.729	0.7493	0.2878	0.4306	0.402	0.5334
sigma2320_run5	0.7159	0.6804	0.2864	0.4521	0.3985	0.5298
PRaMM_run5	0.7638	0.7362	0.2754	0.4288	0.3949	0.529
UNED_run5RelFeedback	0.7841	0.7645	0.2774	0.4194	0.3928	0.524
SAILUSC_VisualRun5	0.7087	0.7022	0.2602	0.4211	0.3694	0.5116
bmedclab_run1	0.7203	0.7022	0.2129	0.3702	0.3204	0.4751
Flickr initial results	0.6986	0.6877	0.2367	0.3681	0.3427	0.4676
cslu_f1a_run4	0.6029	0.5986	0.2372	0.3695	0.3303	0.4433

Best improvements compared to Flickr (percentage points): P@20 - 15, CR@20 - 15.

25

Results: Official Ranking (F1@20; multi-topic)

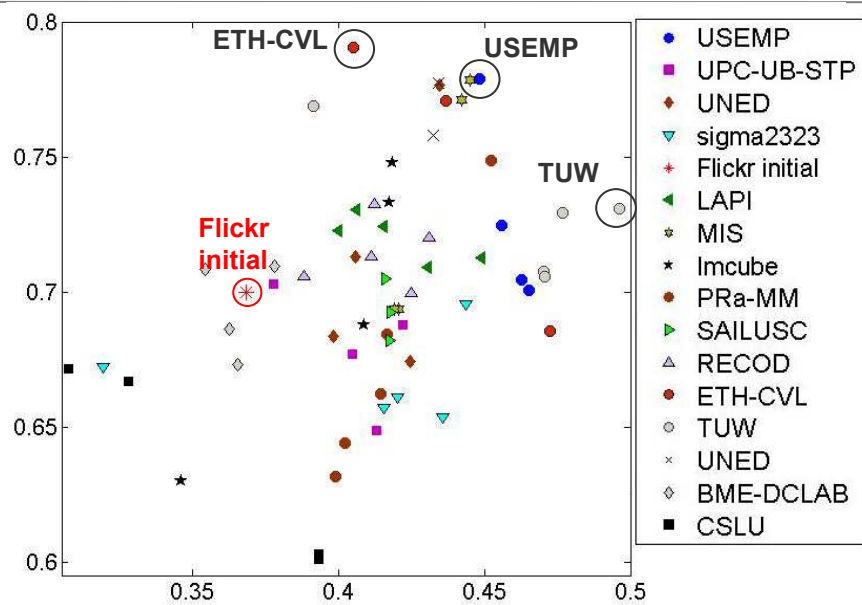
> team best runs only (full ranking will be sent via email);

team/run	P@10	P@20	CR@10	CR@20	F1@10	F1@20
PRaMM_run5PRaMM_run5	0.7543	0.7607	0.3111	0.4753	0.4285	0.567
TUW_run3	0.7343	0.7293	0.3091	0.489	0.4177	0.5654
UNED_run5RelFeedback	0.7871	0.7886	0.3041	0.4491	0.4275	0.5519
sigma2320_run3	0.7043	0.685	0.2978	0.4626	0.4118	0.5393
LAPI_HC-RF_run4_credibility	0.67	0.6814	0.2936	0.4684	0.3932	0.5364
MIS_run3	0.7357	0.7636	0.2662	0.4354	0.3793	0.5353
ETH-CVL_run1	0.7214	0.6829	0.3096	0.4622	0.4204	0.5333
USEMP_run2sMMRtextual	0.72	0.7343	0.301	0.4417	0.4068	0.5299
SAILUSC_VisualRun3	0.7143	0.7079	0.2756	0.4259	0.3849	0.5174
RECOD_run1	0.7457	0.735	0.2758	0.4221	0.3932	0.5133
IMCUBE_run1_vis	0.6671	0.6743	0.2673	0.4209	0.3702	0.5027
UPC-STP_run3_A_20	0.5986	0.6271	0.2656	0.4136	0.3536	0.4822
bmedclab_run1	0.6914	0.7164	0.2273	0.3857	0.3288	0.4813
Flickr initial results	0.7186	0.7121	0.2436	0.3687	0.3504	0.4667
cslu_f1a_run4	0.5871	0.6071	0.2515	0.4167	0.3402	0.4764

Best improvements compared to Flickr (percentage points): P@20 - 5, CR@20 - 10.

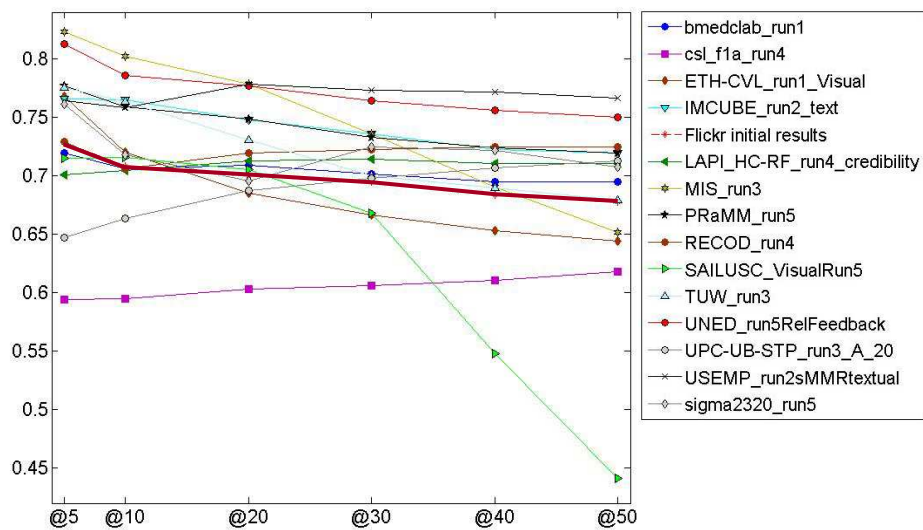
26

Results: P vs. CR @20 (all runs - all testset)



27

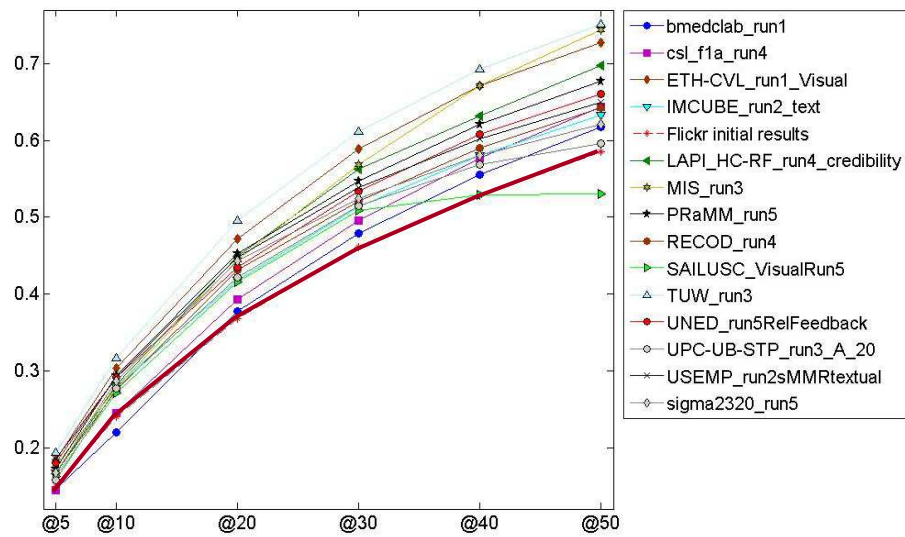
Results: Best Team Runs (Precision @)



*ranking based on official metrics (F1@20).

28

Results: Best Team Runs (Cluster Recall @)



*ranking based on official metrics (F1@20).

29

Results: Visual Results - Flickr Initial Results



e.g., Pingxi Sky Lantern Festival

CR@20=0.28 (18 clusters), P@20=0.95, F1@20=0.43.

30

Results: Visual Results #2 - Best CR@20



e.g., Pingxi Sky Lantern Festival

CR@20=0.72 (18 clusters), P@20=0.85, F1@20=0.78.

31

Results: Visual Results #3 - Lowest CR@20



e.g., Pingxi Sky
Lantern Festival

CR@20=0.44 (18 clusters),
P@20=0.95,
F1@20=0.61.

32

Brief Discussion

Methods:

- this year mainly **classification/clustering** (& fusion), **re-ranking**, **optimization-based & relevance feedback** (incl. machine-human);
- best run F1@20: improving relevancy (text & Greedy) + diversification via clustering (learning on devset the best clustering-feature-distance); use of visual-text information (team TUW);

Dataset:

- getting very complex (read diverse);
- still low resources for Creative Commons on Flickr;
- multi-topic diversity annotations slightly easier to perform;
- descriptors were very well received (employed by most of the participants as provided).

33

Present & Perspectives

For 2015:

- new multi-topic queries related to location events,
- the entire dataset is to be publicly released (soon).

For 2016:

- general purpose queries?



**EUROPEAN
SCIENCE
FOUNDATION**

Acknowledgements

MUCKE project: *Mihai Lupu*, TUW, Austria & *Adrian Popescu*, CEA LIST, France (for funding part of the annotation process);

European Science Foundation, activity on "Evaluating Information Access Systems" (for funding the attendance to the workshop);

Task auxiliaries: *Alexandru Gînsacă*, CEA LIST, France & *Bogdan Boteanu*, UPB, Romania.

Task supporters: *Ioan Chera*, *Ionuț Duță*, *Andrei Filip*, *Florin Guga*, *Tiberiu Loncea*, *Corina Macovei*, *Cătălin Mitrea*, *Ionuț Mironică*, *Irina Emilia Nicolae*, *Ivan Eggel*, *Andrei Purică*, *Mihai Pușcaș*, *Oana Pleș*, *Gabriel Petrescu*, *Anca Livia Radu*, *Vlad Ruxandu*, *Gabriel Vasile*.

34

Questions & Answers

Thank you!

... and please contribute to the task by
uploading free Creative Commons
photos on social networks!
(you are doing a great work so far ;-))

see you at the **poster session** and for the
technical retreat ...

35

CBMI 2016 in Bucharest



CBMI aims at bringing together the various communities involved in all aspects of content-based multimedia indexing for retrieval, browsing, visualization and analytics.

The CBMI proceedings are traditionally indexed and distributed by IEEE Xplore. In addition, authors of the best papers of the conference will be invited to submit extended versions of their contributions to a special issue of *Multimedia Tools and Applications* journal (MTAP).

For more information see <http://cbmi2016.upb.ro> or follow us on Twitter <https://twitter.com/cbmi16> and Facebook <https://www.facebook.com/CBMI2016>.

CBMI 2016
14th International
Workshop on
Content-Based
Multimedia
Indexing

15-17 June, 2016
Bucharest, Romania

Important dates

-Full/short paper
submission deadline:
February 1, 2016;

-Notification of
acceptance:
March 31, 2016;

- Camera-ready papers
due: *April 14, 2016.*

36