

# ACM Multimedia 2014 Grand Challenge MSR-Bing Image Retrieval Challenge

<http://research.microsoft.com/irc/>

(Latest Update: June 14, 2014)

## Challenge Overview

With the success of the 1st MSR-Bing Image Retrieval Challenge (MSR-Bing IRC) at ACM Multimedia 2013, Microsoft Research in partnership with Bing is happy to propose MSR-Bing IRC at ACM Multimedia 2014.

Do you have what it takes to build the best image retrieval system? Enter this Challenge to develop an image scoring system for a search query.

In doing so, you can:

- Try out your image retrieval system using real world data;
- See how it compares to the rest of the community's entries;
- Get to be a contender for ACM Multimedia 2014 Grand Challenge;

## Task

The topic of the Challenge is web image retrieval. The contestants are asked to develop systems to assess the effectiveness of query terms in describing the images crawled from the web for image search purposes. A contesting system is asked to produce a floating-point score on each image-query pair that reflects how relevant the query could be used to describe the given image, with higher numbers indicating higher relevance. The dynamic range of the scores does not play a significant role so long as, for any query, sorting by its corresponding scores for all its associated images gives the best retrieval ranking for these images.

## Datasets

The data is based on queries received at Bing Image Search in the EN-US market and comprises two parts: (1) the Training Dataset which is a sample of Bing user click log, and (2), the Dev Dataset which, though may differ in size, is created to have consistent query distribution, judgment guidelines and quality as the Test Dataset. The two datasets are intended for contestants' local debugging and evaluation. Below table shows the dataset statistics.

	#Distinct Queries	#Distinct unigrams
Training Dataset	11,701,890	7,174,869
Dev Dataset	1,000	4,144

$\frac{\text{\#Distinct unigrams in both Training and Dev Datasets}}{\text{\#Distinct unigrams in Dev Dataset}} = 92.3\%$

This dataset is called Clickture-Lite. More details about the dataset please see the [dataset document](#), and the dataset can be downloaded at the [MSR-Bing Image Retrieval Challenge 2013 website](#). A paper introducing this dataset can be found [here](#).

This year, we also provide a much big dataset (Clickture-Full) with 40M images, which is a superset of Clickture-Lite. But this dataset is optional to be used. That is, systems that based on Clickture-Lite will be used for final award evaluation, but systems based on Clickture-Full can be a different run to submit and get evaluated.

## Evaluation Metric

Each entry to the Challenge is ranked by its respective Discounted Cumulated Gain (DCG) measure against the test set. To compute DCG, we first sort for each query the images based on the floating point scores returned by the contesting entry. DCG for each query is calculated as

$$DCG_{25} = 0.01757 \sum_{i=1}^{25} \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

where  $rel_i = \{Excellent = 3, Good = 2, Bad = 0\}$  is the manually judged relevance for each image with respect to the query, and 0.01757 is a normalizer to make the score for 25 Excellent results 1. The final metric is the average of for all queries in the test set.

In addition to DCG, the average latency in processing each image-query pair will be used as a tie-breaker. For this Challenge, each entry is given at maximum 12 seconds to assess each image-query pair. Tied or empty (time-out) results are assigned the least favorable scores to produce the lowest DCG.

## Process (New!)

In the evaluation stage, you will be asked to download one compressed file (evaluation set) which contains two files in text formats. One is a list of key-query pairs, and the other is a list of key-image pairs. You will be running your algorithm to give a relevance score for each pairs in the first file, and the image content, which will be Base64 encoded JPEG files, can be found in the second file through the key.

The evaluation set, which is encrypted, will be available for downloading 2 to 3 days before the challenge starts. A password will be delivered to all participants to decrypt the evaluation set when the challenge starts.

One full day (24 hours) will be given to the participants to do predictions on all the query-image pairs in the evaluation set. Before the end of the challenge, participants need to submit the evaluation results (which is a TSV file containing a list of triads: key, query, score) to a CMT system (will be announced at: <http://research.microsoft.com/irc/>). The order of the triads in the text file is not important. Running prediction in parallel is encouraged.

The evaluation set will be different from what we used last year. The number of query-image pairs will be increased significantly this time. A trial set will be available around one week before the start of the challenge.

One team can submit up to 3 runs in 3 zipped text files, and each file corresponds to the results of one run. The team need to clearly specify one of the run as the "master" run, which will be used for final ranking. The results for other runs will be also sent back to the teams for their reference.

## Participation and Prizes

The Challenge is a team-based contest. Each team can have one or more members, and an individual can be a member of multiple teams. No two teams, however, can have more than 1/2 shared members. The team membership must be finalized and submitted to the organizer prior to the Final Challenge starting date.

At the end of the Final Challenge, all entries will be ranked based on the metrics described above. The top three teams will receive award certificates and/or cash prizes (prize amounts TBD).

## Paper Submission

Please follow the guideline of ACM Multimedia 2014 Grand Challenge for the corresponding paper submission.

## Detailed Timeline

- Feb 15, 2014: Dataset available for download (Clickture-Lite) and hard-disk delivery (Clickture-Full).
- June 18: Trail set available for download and test.
- June 24: Final evaluation set available for download (encrypted)
- June 26: Evaluation starts (password for decrypt the evaluation set delivers at 0:00am PDT)
- June 28: Evaluation results announce.
- June 29, 2014: Paper submission (please follow the guideline of the main conference)

## More information

- Information about MSR-Bing IRC @ ICME 2014 (containing important message about evaluation data).
- [Dataset Download](#)
- [MSR-Bing IRC 2013](#)
- [MSR-Bing IRC 2013 Workshop](#)
- Research paper about the dataset: "[Clickage: Towards Bridging Semantic and Intent Gaps via Mining Click Logs of Search Engines](#)"
- [Looking into "MSR-Bing Challenge on Image Retrieval"](#)
- [Can I Use Additional Data?](#)

Please note this time we don't separate the Challenge to two tracks as we did in MSR-Bing IRC 2013. Instead, we only have one track this time. The evaluation will be based on both the final challenge results and the paper submissions. And please also note that though we use the same training data as MSR-Bing IRC 2013, the test data in final challenge will be different.

## Challenge Contacts

Questions related to this challenge should be directed to:

Xian-Sheng Hua ([xshua@microsoft.com](mailto:xshua@microsoft.com)), Microsoft Research

Ming Ye ([mingye@microsoft.com](mailto:mingye@microsoft.com)), Microsoft Bing