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

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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Distinct Queries</th>
<th>#Distinct unigrams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>11,701,890</td>
<td>7,174,869</td>
</tr>
<tr>
<td>Dev</td>
<td>1,000</td>
<td>4,144</td>
</tr>
</tbody>
</table>

#Distinct unigrams in both Training and Dev Datasets / #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
As mentioned above, a dataset based on Bing Image search index is available for offline training purposes. Detailed descriptions of the dataset can be found at the “Datasets” section of this page. In addition, the organizer will also make available a web service accessible from the “Team” section of this site for online test runs three months before the final submission deadline. Each contestant can enter the URI of the web service implementing a contending entry at the website. Upon receiving an entry, the Challenge web site will schedule a job to call the web service, evaluate the responses of the web service and post the results on the “Team” and the “Leaderboard” sections of the web site if the entry is designated to show its result in public.

Web Service Development Phase
Initially, the web site will evaluate each entry by computing the DCG on a trial data set with 10 queries and 50 images each. Contestants can therefore submit as many test runs as necessary.

Final Challenge
At the final challenge starts (time TBD), the web site will switch the data set to the Challenge test set. All contestants must ensure their entries are properly registered with the website prior to this time and their web services are up and running for at least 1 week. No further revisions to the entry are allowed at this point.

Once the winners of the Challenge are determined, the website will resume accepting submissions and evaluating results from the general public. The web site with the Challenge test set will be maintained indefinitely after the Challenge for future researchers to include in their studies as a baseline.

Web Service Interface
Each entry is a URI to REST-based web service hosted by each team. The web service must be publicly accessible with HTTP POST through the internet with the following parameters:

<table>
<thead>
<tr>
<th>name</th>
<th>Type</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>runID</td>
<td>UTF-8 string</td>
<td>A unique identifier to name a particular run when a system is submitted for evaluation. White spaces are not allowed in the string</td>
</tr>
<tr>
<td>query</td>
<td>UTF-8 string</td>
<td>A text query in its raw form of user input (all capitalization, punctuations etc retained)</td>
</tr>
<tr>
<td>image</td>
<td>Base 64 string</td>
<td>A base64 encoded JPEG image thumbnail, processed so that the larger dimension between width and height is at most 300 pixels</td>
</tr>
</tbody>
</table>
A contesting system should process the image and respond as soon as possible with a HTTP 200 OK. The response body should be encoded in UTF-8 with the MIME type ‘text/plain’ and contain a floating-point score with higher scores indicating more relevant result for the query. The organizer can call each contesting web service multiple times during the Final Challenge week to obtain statistically significant results to determine winners.

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).
- TBD: Evaluation system available for testing web service
- TBD: Challenge starts
- TBD: Challenge ends
- TBD: Paper submission

More information
- 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 (xhua@microsoft.com), Microsoft Research
Ming Ye (mingye@microsoft.com), Microsoft Bing