

# Using Ephemeral Clustering and Query Logs to Organize Web Image Search Results on Mobile Devices

Jose G. Moreno

HULTIG - University of Beira Interior, Portugal  
DLU/GREYC - University of Caen, France  
LISI - National University of Colombia, Colombia  
jose@hultig.di.ubi.pt

Gaël Dias

HULTIG - University of Beira Interior, Portugal  
DLU/GREYC - University of Caen, France  
ddg@di.ubi.pt

## ABSTRACT

The recent shift in human-computer interaction from desktop to mobile computing fosters the needs of new interfaces for web image search results exploration. In this paper, we present two different strategies to cluster results gathered from an image search engine and propose an adapted interface for handled devices. For that purpose, we suggest to expand the original query based on labels of Ephemeral Clusters and compare it to a Query Log based approach. Consistent results were obtained for both strategies from manual and automatic evaluations, confirming that organizing image search results into clusters can improve mobile image information retrieval.

**Categories and Subject Descriptors:** H.3.3 [Information Storage and Retrieval]: Information search and retrieval [clustering]; H.5.3 [Information interface and presentation]: Group and Organization Interfaces[mobile base interaction]

**General Terms:** Algorithms, Experimentation

**Keywords:** Search results organization, Search results clustering, Mobile image search

## 1. INTRODUCTION

In recent years, the growing number of mobile devices with Internet access has changed the way to access the Internet content as well as user interaction [11]. In particular, new specifically designed applications are developed, which take into account mobile peculiarities to enhance user interaction. However, performing web search in a commercial search engine is still made in a similar way as in desktop computers, i.e. a simple list of ranked results is shown to the user. Furthermore, ranked lists are not suitable for exploration and selection of relevant results, especially on mobile devices. In the context of mobile web image search, enhanced user interaction is crucial as mobile devices have small screens that restrict the number, quality, size of images displayed. Moreover, in mobile devices browsing must be as simple as possible. For example, by showing a sim-

ple ranked list of image results, the user needs to browse in the whole list to look for the relevant image using networking transfer. To address these problems, image collection visualization techniques have been used within novel image retrieval systems to allow new user interactions to explore web image results. However, most of these systems have been made under desktop conditions and do not take into account restrictions present when the user uses a mobile device.

Recently, the image search engine service provided by Google, Google Images, updated the way that mobile users<sup>1</sup> may interact with the application [3] but the new interaction still has problems such as screen space waste and page jumping difficulties to mention just but a few. In the context of desktop computers, researchers have proposed new user interaction solutions using artificial intelligence techniques to enhance web image retrieval such as the “*Sort by subject*” facility proposed in [2]. However, previous works on human-computer interaction have shown that mobile user needs are different than desktop computers user needs [18]. There are even differences between the user needs depending of the mobile devices that are used [14]. Although the desktop solutions could easily be implemented for mobile devices the proposed methodologies so far show different drawbacks such as language dependency [2], use of fixed image collections [7], or predefined categorization that restricts the use of queries with unusual categorizations.

Performing a successful text query is still a research and commercial challenge. To enhance the search process, search engines provide a common tool in web search retrieval called query suggestion. Query suggestions have been used in a variety of ways to help the user query definition process. There are two types of query suggestions facilities: (1) pre-submitted suggestions that appear as an auto-complete box and (2) post-submitted suggestions, which are usually links used to redirect to new queries. Usually, post-submitted suggestions are shown by search engines after a text like “*Including results for ...*”, “*Showing results for ...*” or “*We have included ...*” and some others offer more than one result shown as “*Searches related to ...*”. In particular, typography errors can be solved with post-submitted suggestions and more complex problems related to (1) the query such as semantic disambiguation, (2) the device such as typography difficulties or (3) the user such as users’ time consumption are addressed by search engines using pre-submitted suggestions. In particular, typography difficulties are common in mobile devices because the keys in these keyboards are

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IMMPD’11, November 29, 2011, Scottsdale, Arizona, USA.  
Copyright 2011 ACM 978-1-4503-0995-0/11/11 ...\$10.00.

<sup>1</sup>Iphone and Android users in particular.

used for different letters or because of their smaller size [12]. These pre- and post-submitted suggestions are already implemented in mobile devices, however picking the right pre-submitted suggestion is not a easy task [17].

Google suggestions are defined by the query logs that they capture from users' search activities [1]. In the same way, Yahoo! suggestions come from users' query definition behaviors [4]. The use of query logs as suggestions can reduce keystrokes in common search and help to reduce user time consumption in the query determination process [17] or sometimes can increase the number of keystrokes to improve the precision of the search [13]. As this kind of approaches use query logs information, they are highly dependent on previous users' queries and are in fact unusable for uncommon queries.

To address the problems related to the organization of web image search results for mobile devices and improve the results obtained with image search systems, we explore two customized mobile image visualization techniques based on text information: (1) through an Ephemeral Clustering technique as presented in [9] and (2) through the exploration of Query Logs to generate query suggestions as cluster names. An evaluation with frequently used queries is performed to evaluate the results under users perception. We also propose an automatic evaluation so that it presents good characteristics such as user independence and reproducibility.

This papers is organized as follows. Section 2 describes previous works on web cluster image retrieval. Section 3 explains our proposals. The experimental set-up is presented in the Section 4 and the results are discussed in Section 5. Finally, the concluding remarks are presented in Section 6.

## 2. RELATED WORK

The most popular way to access to the image content in the web is using text based queries [8]. Normally, popular search images engines as Yahoo! and Google use the text image information as the name, caption or surrounding text to allow text based searches. Even when services like TinEye, GazoPa or Google Images offer content-based visual searches, the query text based or keywords search are still the preferred way to access web images. This fact is supported by the high accuracy that text query-based systems can get compared to visual query-based systems [16].

Web cluster retrieval has been an interesting way to present search results and has been investigated by researchers with different approaches from text retrieval techniques [9] [22] to image retrieval methodologies [5] [21] [19] [10], but without notable impact on most used search engines such as Yahoo!, Google or Bing in their text retrieval interface. It is however included in the Google's image retrieval interface [2]. Nevertheless, large number of clusters and imprecise names are the main reasons to avoid the use of web cluster retrieval on main search engines [9]. Ephemeral Clustering has been used for web search results classification and has shown good results in previous works for mobile applications, in which the number of obtained clusters and the cluster label are important factors to expedite the user interaction time expense in the retrieval process [15].

Text-based and mixed approaches have been used to address the image web cluster retrieval problem. Mixed approaches include visual feature extraction and analysis as well as text features. One of the most interesting works tackling the mixed approach is proposed by [10]. They first

reformulate the text query with frequent co-occurring key phrases. Then, they organize the images inside text clusters obtained through ephemeral clustering by visual contents. Although, the visual similarity between the images can be guaranteed, semantically related images may not belong to the same visual cluster and many clusters may be generated to include uncategorised images. Moreover, [10] use the text ephemeral clustering algorithm proposed by [22], which over-generates clusters and builds an *Other Topics* cluster for unsolved web pages. As such, many possible results are lost as the cluster can not be used reliably to show semantically related images. Comparatively, [21] address the visual semantic gap including visual phrasal analysis but do not provide names to the clusters as they do not use ephemeral clustering but just combine text and visual information in terms of vectors for classical centroid clustering. In [5], the authors propose a novel technique to combine web page structure, text information and low level image features. The user time consumption problem is addressed using a cluster presentation. In particular, they produce three different image representations: a textual feature, a visual feature and a link graph feature. Their algorithm consists in an initial text cluster where a spectral technique is applied to define the appropriate number of clusters. Then, the web structure and visual information are combined, which show interesting results for one query. The major drawback of this work is the fact that they do not propose a general evaluation with more than one query. As such, the importance of the approach can not be accessed.

In parallel, [19] propose a web image search strategy using text clustering of web search results to reformulate the queries. This approach is based on text analysis only as image textual indexing is supposed to be processed in a first step. A set of new related queries are then defined based on the original query combined with the cluster names. New expanded queries are used to search for new results. Each independent expanded query is then treated as a semantically well-defined cluster of images. A user case study has been realized with 24 volunteers and the results show a preference over two different clustered versions i.e. image retrieval and theme retrieval. Unfortunately, as they use the ephemeral clustering algorithm of [22], they show similar drawbacks as in [10].

Even with these efforts, the most common problems of text cluster retrieval are also present in image cluster retrieval. Recently, [9] described a novel way to obtain hierarchical clusters including semantic knowledge through the use of the InfoSimba similarity measure over web snippets. The results show that this technique outperforms other state of the art techniques in ephemeral clustering as they obtain few but well-semantically separated clusters and propose a way to get well named clusters. In this paper, only text-based approaches are explored avoiding visual information. In particular, we propose two different approaches: (1) a similar framework as proposed in [19] but using the HISGK-means algorithm presented in [9] and (2) a Query Log-based query expansion technique. These approaches are particularly relevant to mobile web image retrieval as only a few clusters of semantically related images can be retrieved and as a consequence provide an effortless interface for mobile devices as browsing will be intrinsically limited. In particular, the HISGK-means algorithm builds few clusters comparatively

to state of the art algorithms and the expansion by query logs is also limited in size.

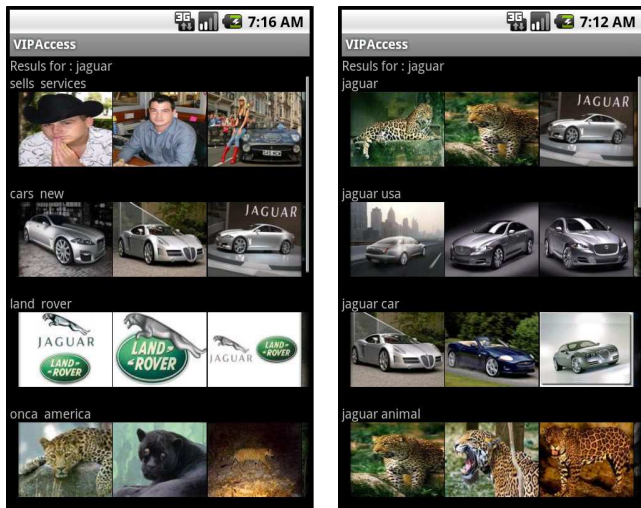
### 3. TWO TEXT-BASED APPROACHES

Our approaches consist in two different methodologies to address the organization of web image search results. First, an Ephemeral Clustering approach is presented and a second one is based on Query Logs. Independently of the approach, the proposed algorithms must achieve the following goal: building a web image search results taxonomy for mobile devices in response to a text query in order to facilitate user exploration. Moreover, the methodologies will have to be based on commercial search engines services to retrieve web images, web snippets and query suggestions, so that we provide a meta image search engine for mobile devices capable to deal with real-world queries.

As previous works in text document retrieval [22] [15] [6], we propose to use a new algorithm [9] based on an informative similarity measure, called the HISGK-means, which allows to fit the image search results on a small screen of a mobile device. In these works, a clustering process is included to discover groups of semantically related documents and presented in a taxonomy to the user. In [9], the authors address the additional challenges of finding the right label for each cluster, which can be different from words contained in a cluster and proposing a method that finds only a few number of clusters. In contradistinction with these works, our system has to group a collection of images that comes from the web in response to a text query. For that purpose, we propose to expand any text query following two different approaches. With the Ephemeral Clustering based approach, the query is expanded with each cluster name to form a cluster of related images. With the Query Log based approach, we use the users previous queries to suggest new query expansion to form different groups of similar pictures.

#### 3.1 Mobile Interface

An interface for mobile devices has been developed using the Android platform for touch screen devices. An initial start page is displayed on which the user can enter the text query and the image search results are displayed in groups of images as shown in Figure 1 for the query “jaguar” based on the ephemeral clustering strategy and the Query Log approach. This is a typical query used to check the ability of the system to find semantically separated results. In particular, both approaches retrieve semantically separated results. The Ephemeral Clustering based approach finds the cluster names “sells services”, “cars new”, “land rover” and “onca america” (the animal) for the first four image groups and the Query Log based approach suggests “jaguar”, “jaguar usa”, “jaguar car” and “jaguar animal”. To facilitate the exploration phase, each group of image results can be explored with left and right movements as a typical gallery exploration. Moreover, the different groups can be explored using up and down movements allowing a quick exploration of different meanings involved in the original query. Finally, when the user picks an image, he gets the basic information of the image and with a longer pick, the user is redirected to the image website.



(a) HISGK

(b) Query Log

Figure 1: Results for “jaguar” using Ephemeral Clustering (HISGK-means) and Query Logs.

#### 3.2 Ephemeral Clustering based Approach

The Ephemeral Clustering algorithm called Hierarchical InfoSimba-based Global K-means (HISGK-means) [9] has been used for this first approach. The algorithm uses web snippet results to generate an automatic hierarchical representation. The clusters are groups of web page results and are labelled within the clustering process. Additionally, the HISGK-means offers three important characteristics for this work: (1) optimum clustering is guaranteed, (2) the labelling step is included in the clustering process to avoid unlabelled clusters and (3) it is language independent. The procedure to get web image search clusters is defined in algorithm 1.

---

**Algorithm 1** Image clusters using the HISGK-means.

---

**Input:** *TextQuery* **Output:** *ImagesCluster*

1. *RankedList* = *getWebResults(TextQuery)*
2. *ClusterSet* = *HISGK(RankedList)*
3. For each element *cluster<sub>i</sub>* in *ClusterSet*
4. *ClusterName* = *getClusterName(cluster<sub>i</sub>)*
5. *ExpandQuery* = *concat(TextQuery, ClusterName)*
6. *ImagesCluster<sub>i</sub>* = *getImageResults(ExpandQuery)*
7. *ImageClusterName<sub>i</sub>* = *ClusterName*
8. return *ImagesCluster*

---

The results are displayed in the mobile device as sorted groups as shown in Figure 1. In fact, each group is a set of images obtained by the Image Search API from Google and labeled by the cluster names obtained from the HISGK-means algorithm concatenated to the original query.

#### 3.3 Query Log based Approach

This approach is based on Query Logs obtained through the Google Suggestions API [1]. As far as we know, there are no previous similar works to organize image web search topics, even when previous works have been developed for web organizing search results [20]. Google Suggestions API service allows get frequently used queries related to a given query. Word completion is a particular procedure, which al-

lows the user to be aware of users queries constructed using the original query as prefix. For example, when the query “app” is used in the Google Suggestions API, the most important suggestions are “applebees”, “apple store”, “apple trailers” and so on and so forth. In fact, these queries do not represent adequately the original query intentions. However, if we include the blank space, we get suggestions like “app store”, “app world”, “app planet” that are more related to the original query. As the previous approach, an image web search is addressed after the query expansion and the results are displayed in the same way as in the first approach. Algorithm 2 describes the Query Log based approach.

---

**Algorithm 2** Image clusters with Query Logs.

---

**Input:** *TextQuery* **Output:** *ImagesCluster*

1.  $TextQuery = \text{concat}(TextQuery, \text{blankSpace})$
  2.  $QueriesSet = \text{getQueryLogSuggestions}(TextQuery)$
  3. For each element  $QueriesSet_i$  in  $QueriesSet$
  4.  $ExpandQuery = \text{concat}(TextQuery, QueriesSet_i)$
  5.  $ImagesCluster_i = \text{getImageResults}(ExpandQuery)$
  6.  $ImageClusterName_i = ExpandQuery$
  7. return *ImagesCluster*
- 

## 4. EXPERIMENTAL SET-UP

To check the performance of both approaches i.e. Ephemeral Clustering and Query Logs, a user evaluation is proposed. In this initial evaluation, the entire web image content is used as dataset through the Google Search API. Moreover, to avoid subjectivity of biased user evaluations, we proposed to evaluate the results using the Amazon crowdsourcing platform called Mechanical Turk<sup>2</sup> (AMTurk). With respect to the query set to evaluate the methodologies, we chose the top fastest rising and top falling queries from 2010 provided by Google Zeitgeist 2010<sup>3</sup>. Some of these queries were removed as they contained straight characters. Finally, 97 queries were used to retrieve and evaluate the approaches. Table 1 shows the 11 categories, the number from each category and respective queries used. In the last column, the original query (in bold font) and examples of query expansions found by both approaches are included. These queries were used both in the users and automatic evaluations.

### 4.1 Users Evaluation

With the previous queries, the approaches were applied to retrieve the top four relevant images of the most relevant three clusters. The relevance order was defined using the estimated number of results for each query provided by the Google API. Both results were presented to three different users using AMTurk, who had to choose between the following options: (1) left side is better, (2) both are bad, (3) both are good and (4) right side is better. To avoid any position bias, the presentation position for each query was randomly assigned to right or left depending on the approach. Three different situations had to be considered by the evaluators to assess the results: (1) taking into account just the text

<sup>2</sup><https://www.mturk.com/mturk/welcome> [Online; accessed June-2011].

<sup>3</sup><http://www.google.com/intl/en/press/zeitgeist2010/> [Online; accessed June-2011].

label, (2) taking into account just the image results and finally, (3) taking into account both text label and image results. An agreement phase was then introduced to avoid the unusual judgements in the post-processing task of the recollected data. Each query was evaluated by three judges and when two out of three judges agreed, the answer was taken into account, otherwise the judgement was avoided. The user agreement obtained was 78%, i.e., at least two users assigned the same answer to 75 out of the 97 original queries.

### 4.2 Automatic Evaluation

To evaluate the performance of both methodologies in an automatic way, two different factors were analysed. First, we propose a new evaluation measure, the average accumulative estimated results (AccER). The AccER is calculated using the estimated size of retrieved results for each query, eventually expanded. The objective of the AccER measure is to define how many images are retrieved accumulatively with  $r$  query expansions. The AccER is defined in equation 1 where  $S_{1j}, S_{2j}, \dots, S_{rj}, \forall j = 1, \dots, n$  are the estimated sizes of image results for the different expanded queries 1, 2, ...,  $r$  of the original query  $j$ , knowing that there exist  $n$  original queries.

$$AccER_r = \sum_{j=0}^n S_{rj} + \sum_{k=1}^{r-1} AccER_k \quad (1)$$

Second, to evaluate if the expanded queries adequately overlap, a percentage of this overlapping is calculated from the original query. The overlapping percentage is defined in equation 2 where  $R_{q_0}$  is the result set of retrieved images for the query  $q_0$  and  $R_{q_i} \forall i = 1, \dots, r$  are the result sets for each expansion of the query  $q_0$ . In particular,  $r$  is the number of expansions obtained and  $m$  is the maximum size of the real retrieved images (in our experiments  $m = 64$ ).

$$\%Overlapping = \sum_{j=0}^n \frac{\sum_{i=0}^r \text{size}(R_{q_j} \cap R_{q_i})}{r * m} * 100 \quad (2)$$

Higher values of this metric indicate a better response of overlapping between the original query and the expanded queries and can be interpreted as the average percentage of images in one expanded query, which were retrieved by the original query.

## 5. RESULTS AND DISCUSSION

In the users evaluation through AMTurk, a total of 45 workers were involved. They solved 291 tasks that include three times each one of the 97 queries in table 1. The obtained distribution of the results is shown in Figure 2. It shows the percentage results obtained under a user agreement rule as was described in the previous section.

Figure 2 includes the distribution results for Image, Text and Combined (Image and Text) judgements. The results show a higher percentage of agreement values for the answer “both are good”. The results also show higher values for the result obtained by the Query Log based approach, which means that users tend to better understand the results shown for the Query Log approach rather than the Ephemeral Clustering strategy. Nevertheless, it is important to note that the queries used were the most rising ones in 2010, which can explain the good performance obtained for

Category	#	Queries	Expanded Queries Examples
Fastest Rising	10	chatroulette, ipad, justin bieber, nicki minaj, friv, myxer, katy perry, twitter, gamezer, facebook	<b>chatroulette</b> : 2010 new, alternative, alternative for adults, ban, chat video, chat world, clones, code script, gifs, ip blocked, like sites, not working, random website, screenshots, service know, service new, sites, source code
Fastest Falling	10	swine flu, wamu, new moon, mininova, susan boyle, slumdog millionaire, circuit city, myspace layouts, michael jackson, national city bank	
Entertainment	8	shakira, eminem, netflix, youtube videos, lady gaga, kesha, grooveshark, transformers 3	<b>shakira</b> : 1977 born, albums biography, biography, downloads songs, facebook official, loca, loca lyrics, lyrics, news lyrics, pictures gallery, rabiosa, rabiosa lyrics, waka waka
Sports	10	mundial 2010, olympics, espn3, fifa 11, randy moss, miami heat, mourinho, wayne rooney, cricket live score, david villa	
Consumer Electronics	8	iphone 4, nokia 5530, htc evo 4g, nokia n900, blackberry apps, duracell mygrid, otterbox, pданet	<b>iphone 4</b> : 2011 phone, 3gs apple, accessories, apple 2010, apps, cases, covers, jailbreak, recording calling, reviews, specs, unlock, verizon
Food & Drink	8	masterchef, cupcakes, jimmy johns, dominos pizza menu, tudo gostoso receitas, guacamole recipe, applebees menu, rachel ray	
Maps Searches	10	anhembi parque, wm gucken, world cup, bundeskanzleramt, rio branco, mt everest, kew gardens, tour eiffel, oxford street, nür-burging	<b>anhembi parque</b> : america latin, brazil pavilion, complex shows, inn holiday, paulo brazil, paulo turismo, sao paulo, travel youtube
People	6	selena gomez, kim kardashian, miley cyrus, taylor lautner, megan fox, robert pattinson	
News	7	haiti, besiktas, chile, earthquake, jörg kachelmann, mobile technology, oil spill	<b>haiti</b> : earthquake, earthquake facts, earthquake struck, economy finance, election, flag, history, libre, map, news, news information, president, weather, western country, world development
Health Queries	10	hcg diet, dr oz, aspergers, mcdonalds nutrition, vitamin d deficiency, appendicitis symptoms, cholera, nfp, vacina h1n1, whooping cough	
Humanitarian Aid	10	donate to haiti, donate to pakistan, text to donate, doctors without borders, download to donate, red cross canada, blood donation restrictions, donate blood australia, donate now button, csl plasma	<b>donate to haiti</b> : 2010 relief, donations earthquake, earthquake, help earthquake, hope now, red cross, red cross, relief, text

Table 1: Categorization, number and queries used for the evaluations.

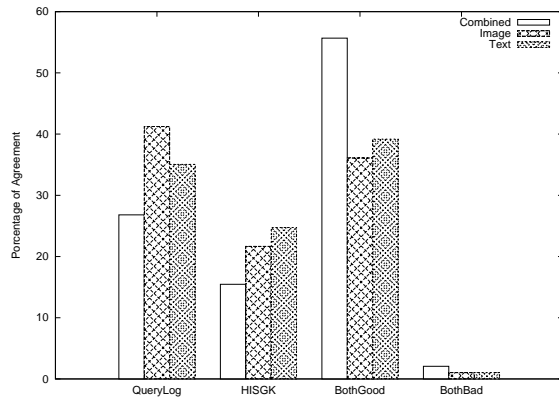


Figure 2: Percentage of agreement for Text, Images and Combined questions from 45 AMTurk workers.

the Query Log approach. However, the Ephemeral Clustering approach also gets an important approval from the users as both methodologies are accepted as good by the users. Moreover, it tends to show better cluster names than the Query Log methodology comparatively to the correctness of image clusters. These results show a promising field to explore when both approaches can be combined in a mixed strategy that may explore the best of each one. On the one hand, the Ephemeral Clustering approach is a better option when the query is rarely used and it is not possible to find good Query Logs. On the other hand, the Query Log approach is a good option when the queries are frequent ones.

In the automatic evaluation, the Figure 3 allows to check the average accumulative estimated results (AccER). The original query average was obtained using the original query without modifications.

As the curve of the HISGK-means approach has a better approximation to the original query curve, this can be interpreted as a better generation of new Query names with-

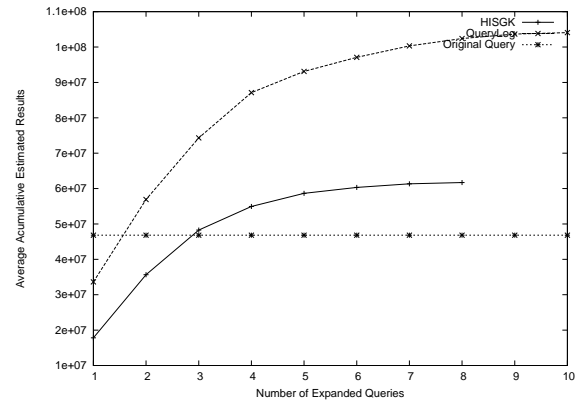
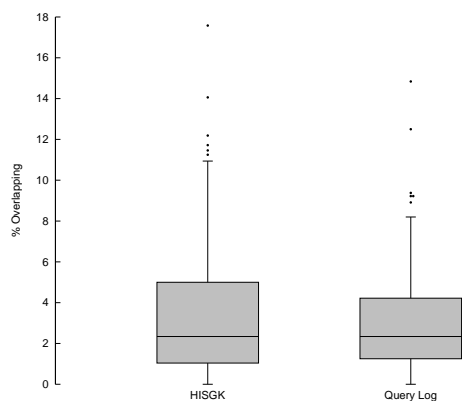


Figure 3: Estimated Average Accumulative results over the 97 queries for the first 10 images groups.

out overlapping, while Query Log based estimated results show that this approach includes an important number of overlapped results. This factor is important in the moment of determining the extra redundant results included in each cluster. This issue is particularly important for image exploration, as more overlapping images immediately decreases the user interface successfulness. Additionally, in order to check the overlapping between the original query and the new expanded queries, the percentage of overlapping was calculated and the results for the 97 queries are shown in Figure 4. The results show similar behaviours in the both approaches and the paired t-student test indicates that there is not a significant difference between both approaches.

With these evaluations, user-based and automatically-based, the obtained results confirm in a consistent way that the proposed approaches both perform well for the organization of web image results. Moreover, they both seem to be complementary.



**Figure 4: Boxplot for percentage of overlapping between the first 64 results of the original query and the first 64 results of the expanded queries.**

## 6. CONCLUSIONS

In this paper, two novel approaches to organize image search results for mobile devices have been presented and evaluated. A novel Ephemeral Clustering technique and Query Log approach have been used to get cluster names and organize image web results into clusters. Both approaches were evaluated with top rising and falling queries. A User-based and an automatically-based evaluations have been proposed, which show that the proposed approaches perform well for real web queries in a real-world environment. Ephemeral clustering showed more compact results compared to Query Logs, but Query Logs evidenced better user acceptance rates. Both solutions have already been implemented with a novel combination of common galleries interface in which the results are organized by semantic groups. As future work, we aim at evaluating the results of the approaches on the mobile device itself to access definitive conclusions.

## 7. REFERENCES

- [1] Google Suggestions. [Online; accessed June-2011]. <http://labs.google.com/intl/en/suggestfaq.html>.
- [2] Sort by subject in Google Images. [Online; accessed June-2011]. <http://googleblog.blogspot.com/2011/05/sort-by-subject-in-google-images.html>.
- [3] The New Image Search for Android and iPhone. [Online; accessed June-2011]. <http://googlemobile.blogspot.com/2010/04/new-image-search-for-android-and-iphone.html>.
- [4] Yahoo Suggestions. [Online; accessed June-2011]. <http://developer.yahoo.com/search/web/V1/relatedSuggestion.html>.
- [5] D. Cai, X. He, Z. Li, W.-Y. Ma, and J.-R. Wen. Hierarchical clustering of www image search results using visual, textual and link information. In *12th Annual ACM International Conference on Multimedia (MM 2004)*, 2004.
- [6] C. Carpineto and G. Romano. Mobile information retrieval with search results clustering : Prototypes and evaluations. *Journal of the American Society for Information Science*, 60:877–895, 2009.
- [7] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, and N. Papamarkos. Accurate image retrieval based on compact composite descriptors and relevance feedback information. *International Journal of Pattern Recognition and Artificial Intelligence*, 24(2), 2010.
- [8] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60, 2008.
- [9] G. Dias, G. Cleuziou, and D. Machado. Informative polythetic hierarchical ephemeral clustering. In *IEEE/WIC/ACM International Conference on Web Intelligence (WIC 2011)*, 2011.
- [10] H. Ding, J. Liu, and H. Lu. Hierarchical clustering-based navigation of image search results. In *16th Annual ACM International Conference on Multimedia (MM 2008)*, pages 741–744, 2008.
- [11] M. Kamvar and S. Baluja. A large scale study of wireless search behavior : Google mobile search. In *24th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, 2006.
- [12] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8):58–62, 2007.
- [13] M. Kamvar and S. Baluja. Query suggestions for mobile search: Understanding usage patterns. In *26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, 2008.
- [14] M. Kamvar, M. Kellar, R. Patel, and Y. Xu. Computers and iphones and mobile phones, oh my!: a logs-based comparison of search users on different devices. In *18th International World Wide Web Conference (WWW 2009)*, pages 801–810, 2009.
- [15] D. Machado, T. Barbosa, S. a. Pais, B. Martins, and G. Dias. Universal mobile information retrieval. In *13th International Conference on Human Computer Interaction (HCI 2009)*, pages 345–354, 2009.
- [16] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. Kahn, and W. Hersh. Overview of the clef 2009 medical image retrieval track. In *10th Workshop of the Cross-Language Evaluation Forum (CLEF 2009)*, 2009.
- [17] T. Paek, B. Lee, and B. Thiesson. Designing phrase builder : A mobile real-time query expansion interface. *Human Factors*, 2009.
- [18] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, 2008.
- [19] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang. Igroup: Presenting web image search results in semantic clusters. In *25th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2007)*, pages 587–596, 2007.
- [20] X. Wang and C. Zhai. Learn from web search logs to organize search results. In *30th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2007)*, 2007.
- [21] X.-J. Wang, Q.-C. He, and X. Li. Grouping web image search result. In *12th Annual ACM International Conference on Multimedia (MM 2004)*, 2004.
- [22] H.-J. Zeng, Q.-C. He, Z. Chen, W.-Y. Ma, and J. Ma. Learning to cluster web search results. In *27th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2004)*.