Using Text-based Web Image Search Results Clustering to Minimize Mobile Devices Wasted Space-Interface

Jose G. Moreno and Gaël Dias

Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France

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 order to leverage users' efforts, we present a set of state-of-the-art ephemeral clustering algorithms, which allow to summarize web image search results into meaningful clusters. This way of presenting visual information on mobile devices is exhaustively evaluated based on two main criteria: clustering accuracy, which must be maximized, and wasted space-interface, which must be minimized. For the first case, we use a broad set of metrics to evaluate ephemeral clustering over a public golden standard data set of web images. For the second case, we propose a new metric to evaluate the mismatch of the used space-interface between the ground truth and the cluster distribution obtained by ephemeral clustering. The results evidence that there exist high divergences between clustering accuracy and used space maximization. As a consequence, the trade-off of cluster-based exploration of web image search results on mobile devices is difficult to define, although our study evidences some clear positive results.

1 Introduction

In recent years, the growing number of mobile devices with internet access has changed the way to access web contents as well as user interaction [1]. However, performing mobile web image search is still made in a similar way as in desktop computers, i.e. a simple list or grid of ranked image results is returned to the user. Previous works on human-computer interaction have also shown that mobile user needs are different than the ones for desktop computers [2]. In particular, ranked lists are not suitable for exploration and selection of relevant results on mobile devices as they involve repeated scrolling, sliding and zooming.

In the specific context of Information Retrieval, [3] propose a study of novel interfaces for web image search. In particular, they conducted a large scale analysis of search logs based on a set of 55 million queries. Their findings suggest several interesting implications. Web image searchers view more pages of search results, they spend more time looking at those pages and they click on more results than web page searchers. According to the authors, one of the main reasons for this situation is the fact that there is often no definitive answer to a query, which means that the sought after image could be one of many.

To remedy this situation, one common way to present web search results on mobile devices is to build meaningful ephemeral clusters [4]. Ephemeral clustering, also called Search Results Clustering (SRC), is a methodology that discovers clusters from a set of web search results retrieved for a given query. Once clusters have been presented to the users, they are discarded and do not provide information for future results. As a consequence, SRC can be seen as a clustering methodology, which deals with small data sets and focuses on the dynamics and volatility of the web. Outside the mobile computing context, this methodology has mainly been used for web page search results organisation [5] [4] [6] [7] and web image search results exploration [8] [9] [10].

In this paper, we are particularly interested in studying web image SRC algorithms to improve search engine interfaces for mobile devices. Although different approaches have been proposed for web image SRC [8] [9] [10], few studies have specifically been dealing with mobile devices [11] [12]. Moreover, both studies draw approximative conclusions as their evaluation frameworks are incomplete. As a consequence, we propose to evaluate the trade-off between clustering accuracy and used space-interface over a public data set of 71478 web images [13] and 353 text queries for three common state-of-the-art SRC algorithms: Suffix Tree Clustering (STC) [14], LINGO [15] and HISGK-means [16].[1] In the context of mobile devices, we hypothesize that SRC systems should successfully combine two main criteria: maximum cluster accuracy and minimum wasted space-interface. The underlying idea is simple. In order to leverage the users' efforts, a given interface should clearly list all of the many sought of images as well as present them in a compact representation. For the first case, ephemeral clustering should be as accurate as possible. For this purpose, we propose a broad set of clustering evaluation metrics [17] in the specific context of text-based web image SRC systems. For the second case, the diversity of the web image search results should be presented in an effortless interface, which limits repetitive scrolling, sliding or zooming. In order to quantitatively measure the compactness of a given interface, we propose a new metric, which evaluates the mismatch of the used space-interface between the ground truth and the cluster distribution obtained by ephemeral clustering. The results evidence that there exist high divergences between clustering accuracy and used space maximization. As a consequence, the trade-off of cluster-based exploration of web image search results on mobile devices is difficult to define, although our study evidences some clear positive results.

In the next section, we present the related work about web image ephemeral clustering. In the third section, we propose a quick overview of text-based SRC systems and present our methodology to improve mobile web image search results exploration. Then, we define the experimental set-ups and list the different evaluation metrics for clustering accuracy. Subsequently, we theoretically define a new metric to measure the wasted space-interface. Finally, we discuss the obtained results and draw conclusions about the use of text-based web image SRC to minimize mobile devices wasted space-interface.

---

[1] More SRC algorithms can be used, but we motivate our decision in section 3.

2   Related Work

In previous web image SRC research works, two different methodologies can be distinguished: single-step and multiple-steps result processing. Following the first approach, clustering is directly performed on image information, which can be visual or/and textual. Within this context, [9] extract codewords based on image analysis, which are then ranked based on a regression model. The salient visual phrases are then used to gather images and form clusters in a monothetic like-wise strategy. The proposed strategy is evaluated on a set of crawled web images for 15 different text queries. For each query, the images are manually clustered and 5 evaluators are asked to name each cluster. Finally, the results evidence smaller entropy compared to the baseline. One major drawback of this work is the fact that no cluster labels are provided as only visual features are used, which may unease the process of information search. In [8], the authors propose a technique to combine web page structure, text information and low level image features. In a first step, the appropriate number of clusters is defined using a spectral analysis technique over the textual features. Then, web structure and visual information are combined to re-organise images in clusters. Unfortunately, the authors only evaluate their contribution against the ambiguous query "Pluto" based on 496 images and 6 clusters.[2] As such, the importance of their approach cannot be generalized.

Within the second approach, ephemeral clustering is first performed on text information to retrieve salient text query meanings/facets and then query reformulation is performed for each cluster label over an image search engine. Within this context, [18] propose a text-based strategy called IGroup. First, the STC algorithm, proposed earlier by [14], clusters (text) web image snippets into semantic groups and associates a label to each one. The clusters are then visualized through a navigational panel, together with their labels and representative thumbnails. For evaluation, they propose a user study realized with 24 volunteers and results show a preference of cluster-based interfaces for image retrieval and multi-faceted retrieval. Unfortunately, as they use STC, which tends to over-generate clusters [16] and builds an *Other Topics* cluster for unsolved web pages, many unrelated results are gathered in this "unlabeled cluster". Moreover, the evaluation is not reproducible as the list of queries and web image search results are not known. Following the same ideas, [10] propose to reformulate a text query with frequent co-occurring key phrases within previously encountered ephemeral clusters. Then, they organize the images inside text clusters by visual contents. To evaluate the performance of their HiCluster algorithm, 12 participants were involved in a user study to compare HiCluster, IGroup and Google Image search engine over a set of 14 keywords. 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. Also, they show the same drawbacks as in [18]: they use the STC al-

---

[2] The authors acknowledge that the maximization of the clustering algorithm also gives best results for 21 partitions.

gorithm, the evaluation data set is not provided and a user study is preferred over a qualitative evaluation. More recently, [12] propose to use the HISGK-means SRC algorithm [16] and compare it to a query log strategy. The results are evaluated based on a crowdsourcing study with 45 users. Additionally, a new evaluation metric is proposed and computed over a set of 97 queries. However, the results are not conclusive, as stated by the authors. Moreover, similar to previous studies, their evaluation is not reproducible. Finally, they state that their SRC algorithm proposes a compact representation of web image search results. But, as they do not compare it to well-known SRC strategies, their "real" contribution cannot be assessed.

While initial works have been focusing on the single-step strategy, recent works prefer to focus on the multiple-steps approach. This situation can easily be understood as correct text labeling of meaningful web image clusters is the essential key to facilitate user exploration of web search results. As a consequence, the text feature plays a crucial role in web image ephemeral clustering, both for single-step and multiple-steps approaches. In fact, the main meanings/facets of queries are discovered through text SRC algorithms. Then, in a second step, web image clustering may be improved by the introduction of visual or web structure features. Nevertheless, although promising results are presented, most studies propose superficial evaluation results, which do not allow to draw definitive conclusions. As a consequence, we propose to exhaustively evaluate a new hypothesis in the context of mobile web image search against a standard data set [13] of 353 text queries and 71478 web images crawled from the web for three common state-of-the-art SRC systems: Suffix Tree Clustering (STC) [14], LINGO [15] and HISGK-means [16].[3] Our main hypothesis is that mobile interfaces for web image search should satisfy two main constraints. The first one aims to provide relevant cluster results to the user. As a consequence, (text) web image snippet clustering must reveal high accuracy. The second one aims to leverage the user's effort in exploring web search results. As such, interface operations such as scrolling, sliding or zooming should be minimized. To quantitatively measure such possible interactions, we introduce a new evaluation metric based on the comparison between the ground truth layout and the cluster distribution obtained by ephemeral clustering in terms of wasted space-interface.

3   Text-Based Web Image Search Results Clustering

Web image SRC systems consist in different processing steps to address the organization of web image search results following two different approaches: single-step and multiple-steps. The common processing in both approaches is text-based web image search results clustering. Indeed, both methodologies strongly depend on (text) web image snippet ephemeral clustering, which may eventually be combined with other features. As a consequence, we propose to study different text-based SRC algorithms. Related studies in web image SRC have privileged two different clustering algorithms: STC [14] and HISGK-means [16]. However,

---

[3] The application of OPTIMSRC [6] and TOPICAL [7] is discussed in the next section.

many successful works have been proposed for ephemeral text clustering such as LINGO [15], OPTIMSRC [6] and TOPICAL [7]. As a consequence, we propose a comparative study to acknowledge the behavior of each one of these five algorithms in terms of text-based clustering accuracy over the ODP-239 [6] data set. Our main idea is to motivate the use of a given text-based SRC algorithm in the context of web image search results clustering for mobile devices.

3.1 A Review of SRC Algorithms

Different SRC algorithms have been proposed. In this section, we review the most relevant and recent ones in the context of (text) web snippet ephemeral clustering.

STC: The Suffix Tree Clustering algorithm is proposed by [14] to efficiently identify sets of text documents that share common phrases. In particular, they propose a monothetic clustering technique, which merges base clusters with high overlap. The main distinguishing characteristics of STC are its linear time complexity and a string text representation instead of the common bag−of−words.

LINGO: [15] propose the LINGO monothetic clustering algorithm. They first extract frequent phrases based on suffix-arrays. Then, they reduce the term-document matrix to reduce space dimensionality using Single Value Decomposition to discover latent structures of diverse topics. Finally, they match group descriptions with the extracted topics and assign relevant documents to them.

OPTIMSRC: [6] show that the characteristics of the outputs returned by SRC algorithms suggest the adoption of a meta clustering approach. As such, they introduce a novel criterion to measure the concordance of two partitions of objects into different clusters based on the information content associated to the series of decisions made by the partitions on single pairs of objects. Then, the meta clustering phase is viewed as an optimization problem of the concordance between the clustering combination and the given set of clusters. The results demonstrate that meta SRC is superior over individual clustering techniques.

HISGK-means: [16] propose a strategy based on the Global K−means algorithm and an informative similarity measure, which captures the similarity between web snippets by mixing first-order and second-order similarity measures. In particular, the HISGK-means includes the labeling step to avoid unlabeled clusters and builds a compact hierarchical representation suitable for mobile devices according to [12].

TOPICAL: [7] recently propose to move away from the bag−of−words representation towards a collaborative knowledge-based representation in the form of graphs using a new wikipedia-based annotator [19]. Based on this representation, they develop a monothetic clustering algorithm through the spectral properties of the graph. The final result is a topical decomposition, which achieves relative improvements of up to 20% with respect to current state-of-the-art algorithms.

It is interesting to note that, compared to STC, LINGO and OPTIMSRC, HISGK-means and TOPICAL algorithms access external knowledge i.e. endoge-

nous corpus statistics for the former and collaborative knowledge-base for the latter. So, in order to confirm the first point of our initial hypothesis i.e. clustering accuracy, we propose a comparative experiment over the well-known ODP-239 [6] data set for each one of the five algorithms. In particular, we use STC and LINGO APIs[4] and the java code provided by [16]. For the OPTIMSRC and the TOPICAL algorithms, we replicate the results evidenced by the authors in their respective papers. The results are illustrated in Table 1.

	$F_\beta-$measure	STC	LINGO	OPTIMSRC	HISGK-means	TOPICAL
	F_1-measure	0.324	0.273[6]	0.313[6]	0.375	0.413[7]
ODP-239	F_2-measure	0.319	0.283[6]	0.341[6]	0.426	N/A
	F_5-measure	0.322	0.294[6]	0.380[6]	0.474	N/A

Table 1: Ephemeral Clustering $F_\beta-$measure Results.

In terms of clustering accuracy, the HISGK-means demonstrates a competitive behavior compared to state-of-the-art SRC algorithms i.e. OPTIMSRC and TOPICAL. As a consequence, in the next experiments, we propose to compare STC and LINGO (through available carrot2 APIs) against an available implementation of a representant of state-of-the-art SRC algorithms, the HISGK-means. We also produce the results for K−means using the carrot2 implementation, which automatically decides the best number of clusters.

3.2   A Gallery-Based Interface

Within the scope of our study, we also propose a gallery-based interface to present the results obtained by web image SRC algorithms. Under this interface, each cluster is displayed with its label and a gallery of web images that belongs to it. So, each cluster of image results can be explored with left-right movements as in a typical gallery exploration (see Figure 1). Moreover, shift between clusters is performed using up-down movements allowing a quick exploration of different meanings/facets discovered by the SRC algorithm. Note that this interface takes advantage of touch-screen capabilities allowing the smooth integration of clusters in mobile interfaces.[5] With this interface, the users can explore the top image results in the first clusters like in a common grid interface. However, as additional feature, they can explore the next image results of potentially interesting clusters depending on their query intentions. This interface offers two main advantages for cluster-based interfaces: user adaptability (because of the use of widely-known interactions through the use of a gallery interface) and navigation (because when the shift between clusters is performed, the state of the gallery is kept and the users can restart the exploration of previously explored clusters in the deserted point). Although the interface implementation is thought to minimize users' efforts in both scrolling and sliding,

---

[4] http://carrot2.org [Last Accessed: October, 2012]
[5] Note that the same interface can be used through multidirectional button or keyboard-based navigation.
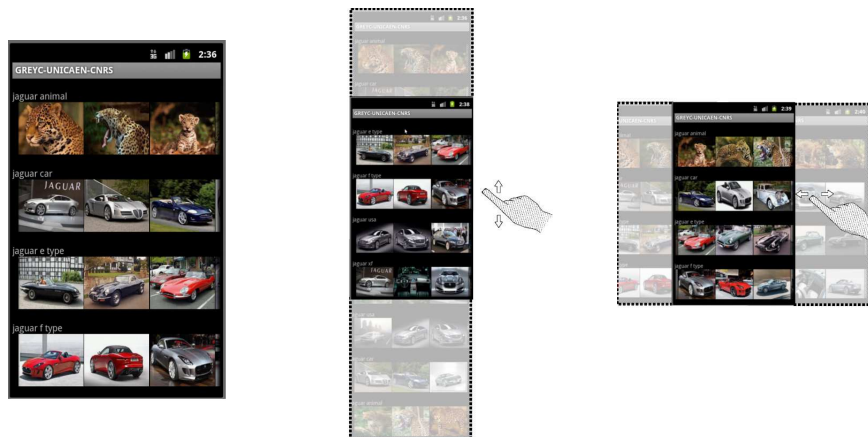
Fig. 1: Mobile Interface. (Left) Mobile Interface, (Center) Scroll Interface Operation and (Right) Slide Interface Operation.

it is important to note that what determines the "real" user energy to use this kind of interfaces is the quantification of the average number of images in each cluster and the average number of clusters to display. We will discuss this issue in the next section.

4   Evaluation and Results

4.1   Data Sets

To evaluate SRC algorithms in the context of mobile web image search, we use a public data set of 71478 images proposed by [13]. The data set is built with top web image results for 353 text queries. For each web image result, its associated metadata are available. Each (text) web image snippet is constructed from the text surrounding the image (10 words before and after), the title of the web page and the alternative image description, available through the metadata. For each web image, a binary label is included to assess if it is relevant or irrelevant to the query. As a consequence, the golden standard is composed of only two groups of images for each query (relevant or irrelevant).

However, for the sake of our evaluation, we need to extend the queries to ambiguous/multi-faceted cases. As such, we propose to merge related queries in terms of string matching and generate a list of new queries. The construction of this new data set is described by the following steps: (1) each query is tokenized and the frequency of each unique token is calculated, (2) unique tokens with frequency higher than two are selected as new queries, (3) the image list of each new query is determined by the union of all the image lists of the queries in which the new query is present and (4) the new query membership is defined in a similar way as in the original data set. Following this procedure, a total of 61

new ambiguous/multi-faceted queries are obtained and the number of clusters for each new query varies between 4 (when 2 queries are merged) and 50 (when 25 queries are merged). Note that the obtained data set has been formatted to the standard proposed by [6] and is available for public access. Some examples of the new data set are illustrated in Table 2.

New Query	Merged Queries (Amb./Facets)	Web Image Results Examples
logo	logo psg, logo fc barcelona, logo apple, logo windows, logo renault, logo ferrari, logo adidas, logo nike, etc.	
france	stade de france, france flag, map france, france team jersey	
simpson	bart simpson, homer simpson	

Table 2: Examples of the Built Data Set with Ambiguous/Multi-faceted Queries.

F_β−measure	F&M	F−measure
$\frac{(\beta^2+1)*P_\beta*R_\beta}{\beta^2*P_\beta+R_\beta}$	$\sqrt{(\frac{SS}{SS+SD})(\frac{SS}{SS+DS})}$	$\frac{2*P*R}{P+R}$
P b-cubed (P_{b^3})	R b-cubed (R_{b^3})	F b-cubed (F_{b^3})
$Avg_e[Avg_{e'.C(e)=C(e')}[Cness(e,e')]]$	$Avg_e[Avg_{e'.L(e)=L(e')}[Cness(e,e')]]$	$\frac{2*R_{b^3}*P_{b^3}}{R_{b^3}+P_{b^3}}$

Table 3: Metrics to Evaluate SRC Algorithms.

4.2 Cluster Accuracy

Recently, [17] showed the inconvenience of typical metrics used in clustering evaluation. Their results indicate that common metrics such as the F−measure are good to assign higher scores to clusters with high homogeneity, but fail to evaluate cluster completeness. They also show a similar behavior for metrics such as Mutual Information and F&M. To remedy this situation, [17] propose new metrics called b-cubed Precision, Recall and F−measure. In particular, they evidence that only the b-cubed F−measure is robust enough to assign adequate scores to different situations for the clustering task. In order to have an exhaustive evaluation, "traditional" and more recent metrics are evaluated over the previously described data set. Note that if any SRC algorithm generates overlapped results, the duplicated images are removed as proposed in [6]. The used metrics are presented in Table 3 following the definitions and notations in [6] (F_β−measure) and [17] (F&M, F−measure, P_{b^3}, R_{b^3} and F_{b^3}).

The clustering results are presented in Table 4. Note that the HISGK-means is a hierarchical algorithm. As such, we provide the results, which correspond to the first level of the tree-structure i.e. $HISGK_1$. The results clearly demonstrate that the HISGK-means algorithm outperforms the other SRC algorithms for text-based web image search results flat clustering.

Data Set	Metric	STC	LINGO	K-means	HISGK$_1$
Original 353 Web Queries	F$_1$−measure	0.369	0.257	0.315	0.611
	F$_2$−measure	0.308	0.197	0.261	0.657
	F$_5$−measure	0.283	0.175	0.241	0.692
	F&M	0.399	0.310	0.355	0.625
	F−measure	0.504	0.438	0.401	0.649
	P_{b3}	0.624	0.679	0.599	0.591
	R_{b3}	0.288	0.182	0.249	0.706
	$\mathbf{F_{b3}}$	**0.386**	**0.280**	**0.331**	**0.623**
61 Ambiguous or Multi-faceted Web Queries	F$_1$−measure	0.469	0.247	0.414	0.553
	F$_2$−measure	0.483	0.207	0.430	0.623
	F$_5$−measure	0.500	0.194	0.461	0.676
	F&M	0.486	0.287	0.441	0.571
	F−measure	0.534	0.390	0.472	0.574
	P_{b3}	0.523	0.569	0.473	0.490
	R_{b3}	0.513	0.218	0.482	0.690
	$\mathbf{F_{b3}}$	**0.475**	**0.267**	**0.414**	**0.554**

Table 4: Ephemeral Clustering Results over the Standard Data Sets.

4.3   Wasted Space-Interface

In this work, we strongly believe that the diversity of the web image search results should be presented in an effortless interface, which can limit repetitive scrolling, sliding and zooming. But, objectively measuring this issue in mobile devices is still an open problem. In order to address this issue, a new metric called Wasted Space-Interface (WSI) is proposed. The idea behind is simple: the less (interface) space is used to present a given quantity of information, the more the users' efforts will be leveraged. So, the WSI should evaluate the mismatch of the used space-interface between the ground truth and the cluster distribution obtained by ephemeral clustering. One direct implication is that the measure does not depend on the characteristics of the mobile device, but is correlated to the used gold standard.

For that purpose, we first define two different spatial quantities: A_{gsr} and A_{src}. The area used by each gold standard query result ($A_{gsr} = GS_{max} \times n$) is defined by the product between the number of elements of the cluster with more images ($GS_{max} = max(|L_1|, |L_2|, .., |L_n|)$), where $|L_i|$ is the number of images in each cluster L_i) and the number of clusters (n). Note that this area is related to the number of repetitive scrolls and slides to explore the overall results. Indeed, the exploration of the images in a given cluster is related to sliding and the shift between clusters is related to scrolling (or vice and versa depending of the screen position). Correspondingly, the area used by any SRC algorithm for each query result ($A_{src} = CS_{max} \times m$) is the product between the size of the biggest discovered cluster ($CS_{max} = max(|C_1|, |C_2|, .., |C_m|)$) and the number of produced clusters (m). So, the WSI is defined as the difference between the

used area-interface of any SRC algorithm minus the used area-interface by the gold standard i.e. $WSI = max\{(A_{src} - A_{gsr}), 0\}$.

Note that, if the golden standard clustered results were presented for a given query on an imaginary screen, the WSI metric would evaluate the need for an extra screen size to display all the information present in the results.[6] Therefore, less compact SRC algorithms would include additional users' effort as repetitive interface interactions would be necessary to visualize all information. As a consequence, more compact SRC algorithms are more likely to afford less repetitive scrolling, sliding and zooming and accordingly better user explorations of the clustered web image results can be achieved. To illustrate this situation, we present in Figure 2 the discovered clusters distributions for an example query over the standard data set for each one of the four SRC tested algorithms. In particular, each circle represents a web image result, the cluster membership is determined by the horizontal organization (each line is a different cluster) and the membership to the gold standard clusters is represented by different colors.

Fig. 2: Distribution shapes for SRC algorithms. From Left to Rigth: STC, LINGO, K-means and HISGK$_1$.

To allow direct comparison of the WSI metrics between different queries, we define the Normalized WSI (NWSI). In particular, we take into account the fact that the worst SRC algorithm would maximize the A_{src} value and therefore NWSI is obtained by dividing WSI with A_{src}^{max}. It is easy to prove that $A_{src}^{max} = ((n_{q_j} + 1)/2)^2$, where n_{q_j} is the number of images of the gold standard for any query (q_j). As a consequence, the NWSI metric is defined $NWSI = WSI/A_{src}^{max}$ and the normalized results of the wasted space-interface are shown in Table 5.

STC	LINGO	K-means	HISGK$_1$
0.082	0.305	0.031	0.027

Table 5: Ephemeral Clustering NWSI Results.

Note that the smaller the NWSI is, the less space is needed to present the clustered information and therefore users' efforts are minimized. The results clearly demonstrate that the NWSI is smaller on average for the HISGK-means than the other SRC algorithms, thus providing a more compact representation of the clustered information and confirming our second hypothesis.

Overall results evidence that there exist high divergences between clustering accuracy and used space-interface maximization for classical SRC algorithms.

---

[6] The negative extra space is avoided by the zero value in the definition of WSI.

For cluster accuracy, we base our conclusions on the F_{b^3} results following the recommendations of [17]. Within this context, the HISGK-means algorithm outperforms by 17.7% the second best result achieved by the STC algorithm. The same situation can be observed for the NWSI evaluation as best results are obtained for the HISGK-means, although comparative results are obtained for K-means. In fact, these results indicate that even though K-means does not achieve as good performances as STC in terms of cluster accuracy, it is more suitable for interface definition. Comparatively, the results for LINGO show bad performances for both experiments. As such, our study clearly and exhaustively demonstrates that the state-of-the-art SRC algorithm, HISGK-means, achieves good performances both in terms of clustering accuracy and used space-interface. These results indicate that current SRC algorithms are appropriate to deal with web image SRC exploration for mobile devices. As far as we know, we are the first to propose a reproducible quantitative evaluation based on a standard data set, which theoretically and practically assesses the contribution of SRC algorithms for mobile web image search results exploration.

5   Conclusions

In this paper, we present the first exhaustive evaluation for the design of enhanced web image SRC systems on mobile devices. We base our study on two different hypotheses. First, cluster accuracy must be maximized and second, used space-interface is ought to be maximized. For that purpose, we develop a complete evaluation framework based on (1) a selection process of recent state-of-the-art SRC algorithms over the ODP-239 data set [6], (2) the definition of a new data set for multi-faceted web image search from the original data set proposed in [13], (3) the use of newly introduced clustering evaluation metrics presented in [17] and (4) the definition of the Normalized Wasted Space-Interface, which evaluates the users' effort to explore web search results based on a spatial definition that does not rely on hardware specificities. The results evidence that there exist divergences between cluster accuracy and NWSI for "classical" SRC algorithms. However, state-of-the-art algorithms for text-based ephemeral clustering, such as the HISGK-means, propose the best trade-off between accuracy and usage. As a consequence, new SRC algorithms are well-suited to generate compact interfaces for mobile devices, thus inherently reducing scrolling, sliding and zooming.

References

1. Kamvar, M., Baluja, S.: A large scale study of wireless search behavior : Google mobile search. In: 24th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI). (2006)
2. Kamvar, M., Kellar, M., Patel, R., Xu, Y.: 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) 801–810

3. André, P., Cutrell, E., Tan, D.S., Smith, G.: Designing novel image search interfaces by understanding unique characteristics and usage. In: 12th International Conference on Human-Computer Interaction (INTERACT). (2009) 340–353

4. Carpineto, C., Romano, G.: Mobile information retrieval with search results clustering : Prototypes and evaluations. Journal of the American Society for Information Science **60** (2009) 877–895

5. Ferragina, P., Gulli, A.: A personalized search engine based on web-snippet hierarchical clustering. Software: Practice and Experience **38**(2) (2008) 189–225

6. Carpineto, C., Romano, G.: Optimal meta search results clustering. In: 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). (2010) 170–177

7. Scaiella, U., Ferragina, P., Marino, A., Ciaramita, M.: Topical clustering of search results. In: 5th ACM International Conference on Web Search and Data Mining (WSDM). (2012) 223–232

8. Cai, D., He, X., Li, Z., Ma, W.Y., Wen, J.R.: Hierarchical clustering of www image search results using visual, textual and link information. In: 12th Annual ACM International Conference on Multimedia (MM). (2004) 952–959

9. Wang, X.J., He, Q.C., Li, X.: Grouping web image search result. In: 12th Annual ACM International Conference on Multimedia (MM). (2004)

10. Ding, H., Liu, J., Lu, H.: Hierarchical clustering-based navigation of image search results. In: 16th Annual ACM International Conference on Multimedia (MM). (2008) 741–744

11. Liu, H., Xie, X., Tang, X., Ma, W.Y.: Clustering-based navigation of image search results on mobile devices. In: 2004 international conference on Asian Information Retrieval Technology (AIRS). (2004) 325–336

12. Moreno, J.G., Dias, G.: Using ephemeral clustering and query logs to organize web image search results on mobile devices. In: 2011 International ACM workshop on Interactive Multimedia on Mobile and Portable Devices (IMMPD). (2011) 33–38

13. Krapac, J., Moray, A., Verbeek, J., Jurie, F.: Improving web-image search results using query-relative classifiers. In: IEEE Conference on Computer Vision & Pattern Recognition (CVPR). (2010) 1094–1101

14. Zamir, O., Etzioni, O.: Web document clustering: A feasibility demonstration. In: 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). (1998) 46–54

15. Osinski, S., Stefanowski, J., Weiss, D.: Lingo: Search results clustering algorithm based on singular value decomposition. In: Intelligent Information Systems Conference (IIPWM). (2004) 369–378

16. Dias, G., Cleuziou, G., Machado, D.: Informative polythetic hierarchical ephemeral clustering. In: 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT). (2011) 104–111

17. Amigó, E., Gonzalo, J., Artiles, J., Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information Retrieval **12**(4) (2009) 461–486

18. Wang, S., Jing, F., He, J., Du, Q., Zhang, L.: Igroup: Presenting web image search results in semantic clusters. In: 25th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI). (2007) 587–596

19. Vitale, D., Ferragina, P., Scaiella, U.: Classification of short texts by deploying topical annotations. In: 34th European Conference on Advances in Information Retrieval (ECIR). (2012) 376–387