Can a Clipboard Improve User Interaction and Experience in Web-based Image Search?

Leon Kastler and Ansgar Scherp

Institute for Web Science and Technologies
Universität Koblenz-Landau, Germany

Abstract. We investigate if a clipboard as an extension to standard image search improves user interaction and experience. In a task-based summative evaluation with 32 participants, we compare plain Google Image Search against two extensions using a clipboard. One clipboard variant is filled with images based on DCG ranking. In the other variant, the clipboard is filled based on gaze information provided by an eyetracker. We assumed that the eyetracking-based clipboard will significantly outperform the other conditions due to its human-centered filtering of the images. To our surprise, the results show that eyetracking-based clipboard was in almost all tasks worse with respect to user satisfaction. In addition, no significant differences regarding effectiveness and efficiency between the three conditions could be observed.

1 Introduction

A study on web-based multimedia search [10] revealed that 56% of all image searches have a follow up. On average, 2.8 queries are executed during an image search session [10]. Thus, multiple pages of query results are inspected and different images are reviewed and compared until the final decision is made. On the other hand, 90% of image search sessions last less than 5 minutes [10]. This raises the question for appropriate user interface designs to conduct image search tasks effectively and efficiently. Search engine providers constantly improve their information retrieval methods and add techniques like content-based filters to enhance the image search experience. However, selecting and using content-based features can be challenging to users. Our idea is to improve the user interaction and experience by extending standard web-based image search engines with the concept of a *clipboard*. A clipboard can be considered as overlay that is automatically filled with images during the users' search session. The users can open the clipboard at any time and view the selected images and review their details.

In this paper, we compare the unmodified Google Image Search with two variants using a clipboard. In the first variant, the clipboard is filled with the query search results of the users by applying an image ranking based on the discounted cumulative gain (DCG) algorithm. In the second variant, the user's gaze information provided by an eyetracking device is leveraged to identify relevant images. The eyetracking device is used without any specific interaction by the user. We have formulated the null hypothesis saying that there is no difference in the three

Fig. 1: Screenshot of clipboard within Google Image Search

conditions (plain Google Image Search, DCG-based clipboard, eyetracking-based clipboard) with respect to participants' efficiency, effectiveness, and satisfaction. We have conducted a sumamtive evaluation with 32 participants comparing the three different conditions. Each participant has conducted six different image search tasks of different complexity. The tasks have been taken from the top 20 queries worldwide of Google Insight in 2011. A natural assumption would be that the results show that the eyetracking-based clipboard with the human-centred filtering of the images will significantly outperform the other conditions. To our surprise, however, the results show that the eyetracking-based clipboard performed worse. The users' satisfaction with respect to the clipboard's support for conducting the image search tasks was in almost all tasks lower. In addition, there are no significant differences between the three conditions regarding efficiency and effectiveness in conducting the tasks.

2 Web-based Image Search by a Clipboard

Using state of the art image search engines, we often encounter the problem that these systems are not suitable for more complex tasks. In situations where users are looking for multiple images to a specific topic with probably several queries, image search engines do not support them properly. Users have to keep a list of possibly interesting images. These are realized through saving images on the local machine or through opening images in different browser tabs. Both solutions need additional effort by the user. To handle this scenario, we introduce the concept of a clipboard to enhance traditional web-based image search.

Our approach of a clipboard is a session-wide, separated, automatically generated collection of interesting result objects as shown in Figure 1. It was designed in order to meet the needs of a multiple image search situation. Throughout a single search session, a user can formulate queries, e.g., different keywords or search engine specific filters, and look through returned result sets. During the search session, the user can open the clipboard. Here, the user receives a presentation of all images that were regarded as interesting by the system. Images are selected based on two different algorithms.

One way to fill up the clipboard is an algorithm that only uses the result set of the image search engine. We use an algorithm based on the discounted cumulative gain algorithm (DCG). The DCG algorithm is normally used to evaluate information retrieval systems [6]. The function is shown in Equation 1. Here, $R_{B,i}$ is our selection function for image i, q_j is the j-th query of the complete query set Q, and $pos_q(i)$ is the position of image i in query q. If images are contained in several result sets, the system adds up the calculated values. The order of images is determined by the decreasing order of $R_{B,i}$ of all i. If the system calculates the same value for two images, it takes the first added image as higher ranked.

$$R_{B,i} = \sum_{q_j \in Q} \frac{1}{pos_{q_j}(i)} \tag{1}$$

The eyetracking-based clipboard is identical to the DCG-based clipboard in terms of user interface design. The difference lies in selecting interesting images for the user. The eyetracking-based clipboard decides based on the sum of gaze events and the total duration of fixations of an image, if it can be considered as interesting (see Equation 2). If two images have exactly the same total fixation time and numbers of events, the image that was fixated first will be ranked higher. $R_{C,i}$ is our selection algorithm for image i and $gazeEvents_q(i)$ are all gaze events in the result set of q for image i. $fixationDuration_q(i)$ is the total duration of all fixations of image i in query q.

$$R_{C,i} = \sum_{q_j \in Q} gazeEvents_{q_j}(i) + fixationDuration_{q_j}(i) \tag{2}$$

Since this algorithm is highly influenced by the user interaction, we face the following problem: it is necessary for our evaluation that the user does not know how images are added to the clipboard. Knowing how the selection algorithm works may lead to an explicit usage of the eyetracker by the participants, i.e., the eyetracker is applied as explicit input device. To prevent that the participants compare images in the clipboard with images outside the clipboard, we span the clipboard over the result set. This is done in both clipboard variants to keep them comparable.

3 Evaluation Design

To investigate the usefulness of a clipboard for image search, we designed a task-based, summative evaluation with three variants of a web-based image search engine. Our first variant, called variant A, uses the unmodified Google Image Search and acts as baseline. Variant B uses the DCG-based algorithm and variant C the eyetracking-based approach as introduced in Section 2.

Procedure: The evaluation consists of three phases. The introduction begins with a short welcome of the participant, followed by an explanation of the set-up, and the calibration of the eyetracker. The participant begins now with the

Task	Description	Complexity
1	Good portrait of Michael Jackson	Simple
2	Beautiful, large image (min. 1200x900px) of the romantic rhine"	Filter
3	Five images of places of interest in Berlin	Composition
4	Good portrait of Lady Gaga	Simple
5	Clipart presentation of "Ampelmaennchen" (Icon of Berlin)	Filter
6	Three images about "Christmas" and "Loriot" (German Comedian)	Composition

Table 1: Evaluation tasks of different complexity

machine-based evaluation process, starting with an overview of the system, the specific evaluation object and the task execution process. To ensure that the participant understands the task evaluation process, we give him an example task at the beginning, that is ignored in later data analysis. This helps to reduce curiosity in using the eyetracker. The task evaluation phase consists of a number of tasks that are executed by the participant. At the beginning of every task, the participant gets a short motivational description with a list of actions. When the participant clicks on the "next" button, a new window opens with the specific variant where the task is executed. Closing this window leads to a task-related questionnaire. When all tasks are processed, the participant is asked to fill in a final questionnaire with general questions about the evaluation. In addition, we encourage the participant to give a short free feedback, either written or oral.

Participants: For the evaluation, 32 participants (9 female) were randomly selected and assigned to the three different system variants. The average age was 25.7 (SD: 3.0). The majority, 28 participants, were students from our university. 20 of the students studied computer science, seven studied education and one participant studied in the field of humanities. The other four participants had a finished, non-academic training in one of the fields of economy, computer science, or physical therapy. All participants were asked to state their experience with Google Image Search on a five-point Likert-Scale. The average experience level was 3.78 (SD: 0.67). Concerning the question of their usage frequency of Google Image Search, eight participants reported that they use it on a monthly basis, 14 on a weekly basis, five on a daily basis and five participants use Google Image Search several times a day. Overall, we can say that all participants use Google Image search on a regular basis and are experienced in its usage.

Tasks: We developed three common scenarios with an increasing level of complexity, based on information from Google Insight[1]. Each scenario was translated into two tasks, rendering a total number of six tasks (as shown in Table 1) to be completed by participant. The first scenario was based on a simple image search task. The participant had to formulate a good query for the task, search through the result set, and select an image as his/her solution. The second scenario includes the usage of filters like specific image types or colouring. Here, the process of the first task was extended by either using a more complex query or filters of the search engine directly. The last scenario was designed to create a set of images given a specific topic where several queries were necessary.

[1] `http://www.google.com/insights/search/`

Name	Group A	Group B/C
Q1	I can conduct the complete task with Google Image Search.	I can conduct the complete task with GoogleBoard.
Q2	Google Image Search offers me all things I need to conduct the task.	GoogleBoard offers me all things I need to conduct the task.
Q3	The results of Google Image Search meet my expectations I had during the formulation of my queries.	The results of GoogleBook meet my expectations I had during the formulation of my queries.
Q4	Too much input steps were needed to conduct the task.	Too much input steps were needed to conduct the task.
Q5	The presentation of results was suitable for the conduct of the task.	The presentation of results was suitable for the conduct of the task.
Q6	N/A	The GoogleBoard supported me in conducting the task.
Q7	N/A	The presentation of results in the Google Image Search was lucid.
Q8	N/A	The presentation of results in the GoogleBoard was lucid.
Q9	N/A	The images selected for the GoogleBoard were suitable for the task.
Q10	N/A	The ordering of images in the GoogleBoard was suitable for the task.

Table 2: Task-related questionnaire

Name	Group A	Group B/C
F1	The usage of Google Image Search is intuitive for me	The usage of Google Image Search with GoogleBoard is intuitive for me
F2	The presence of the eyetracker confused me during the task execution	The presence of the eyetracker confused me during the task execution
F3	I knew in every moment that Google Image Search was working	I knew in every moment that Google Image Search and GoogleBoard was working
F4	N/A	It is easy to switch between Google Image Search and clipboard
F5	N/A	The clipboard is a useful extension for the Google Image Search
F6	N/A	The usage of the clipboard was easy to learn
F7	N/A	There were system errors during my work with the clipboard

Table 3: Final questionnaire

Measurements: We recorded the same data for all participants: the duration of task execution, eyetracking events like gaze-in and gaze-out, mouse events like mouse-over and mouse-out, the number of queries, the number of filters used, questionnaire answers, and free feedback. Taking this measurement in all conditions A, B, and C ensures comparability of the groups and reduces the rule of error by bias.

Questionnaire: We have created two different questionnaires, a task-related (see Table 2) and a final one (see Table 3). The task-related questions aimed on the usability of the variant for the given task. The final questionnaire included questions about the evaluation as a whole. All questions were based on the IsoMetrics Questionnaire Inventory for summative evaluations [3]. Group A was asked less questions in both questionnaires than for group B and C because several questions were related to the clipboard, which group A using the standard Google Image Search did not have.

Fig. 2: slider element, initial state

Fig. 3: slider element, after clicked on so-so

4 Evaluation Wizard and System Architecture

To guide the participants through the evaluation, we have designed a browser-based evaluation wizard. With this wizard, we are able to create a seamless transition between the description, execution, and evaluation of the tasks. All data created during this session is recorded by the system. The answers in the questionnaires are given through a slider element. The limiting values on this scale are defined as "predominantly disagree" and "predominantly agree", taken from the standard Lickert scale. An important part for the evaluation is the absence of the slider handle at the beginning (Figure 2). The handle is only visible after the participant has decided for value on the scale (Figure 3). By this, no priming by a pre-selected value, e.g., a position of the handle on the scale is conducted.

5 Evaluation Results

We measure the effectiveness by comparing the amount of images saved on the computer to the amount of images requested in the task description. To measure the efficiency, we mainly analyse the duration for every task per participant. We also analyse the number of gaze-in events on the images in the result set, number of mouse events, number of queries, and number of filter usages. We measure the user satisfaction by evaluating both questionnaires, the task-related and the final one. We had to remove two participants from the evaluation results. The first tried actively to manipulate the eyetracking-based selection algorithm. The second participant was removed because we lost the session data by a critical error during the evaluation process.

Effectiveness: For the other participants, we checked the images of all participants and verified, if they stored the amount of images as requested in the task description. We found out that one participant of group A and two participants of group C missed to store the correct amount. One participant of group A stored no image in task 1. In group C, one participant stored only one image in task 6, where three were requested and another participant in group C stored for task 3 one image less than requested.

Efficiency: For efficiency, we compared the durations of all three groups per task. Figure 4 shows an average-high-low diagram of these values. We analysed all three groups pairwise with the Mann-Whitney U-test at a significance level of $\alpha = 0.05$. These analyses show a significant difference between group A and C in task 1 ($p = .029$). In all other tasks, we had no significant differences.

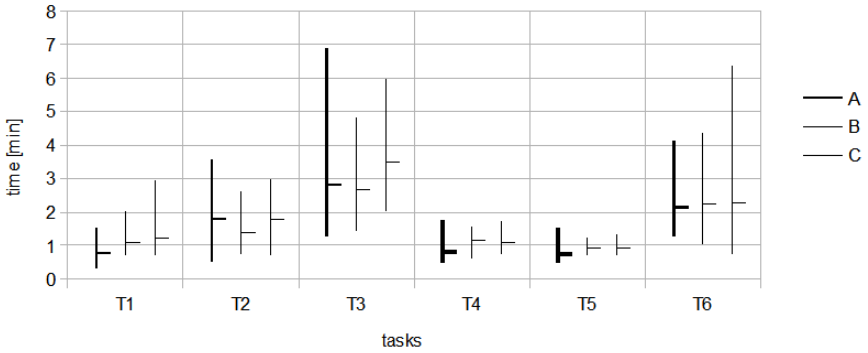

Fig. 4: Min-Max-Average diagram of processing duration per task and group

		Task 1		Task 2		Task 3		Task 4		Task 5		Task 6	
		avg	SD	avg	SD	avg	SD	avg	SD	avg	SD	avg	SD
Duration [s]	A	45.15	21.08	94.35	65.34	153.91	103.7	49.36	26.84	44.57	21.5	134.51	65.12
	B	66.06	26.41	84.73	43.13	160.60	75.94	68.73	20.94	105.43	122.27	135.49	69.47
	C	74.69	39.56	108.06	52.69	207.91	75.64	66.10	17.77	55.16	13.73	138.81	101.55
Number of Gaze-In Events	A	34.5	58.44	73.4	31.49	143.8	155.41	25.5	58.44	20.67	31.49	109.71	155.41
	B	37.3	54.15	42.3	73.12	107.5	140.61	41	42.03	50.83	168.35	74	159.28
	C	50.4	68.7	58.1	105.12	113.8	140.76	40.88	50.01	25.71	28.93	84.5	179.36
Number of Mouse Events	A	20.8	15.19	74	50.62	112.2	68.21	20.38	10.39	17.88	9.25	97.86	79.86
	B	34	17.01	46	29.8	102.8	44.3	39.33	27.13	66.67	108.31	62.17	26.45
	C	36.9	23.69	59.5	54.11	148.3	88.84	36.38	22.36	36.63	30.3	86.38	86.38
Number of Queries and Filters	A	1	0	4	1.63	2.8	3.12	1.38	0.75	1.88	1.13	2.14	1.93
	B	1.3	0.68	2.1	0.32	1.7	1.06	1.17	0.4	2.5	1.76	1.17	0.4
	C	1.5	0.97	2.4	0.97	3.1	2.56	1.25	0.46	2.13	2.42	2.25	2.82
Images saved via GoogleBoard	B	0.6	0.52	0.4	0.52	1.6	1.58	0.17	0.41	0.67	0.52	.067	0.82
	C	0.1	0.32	0.3	0.48	1.2	1.69	0.38	0.52	0.25	0.46	1	1.41

Table 4: Events sorted by task and group

However, we could observe that variant C was on average slower than both other variants. Besides duration, we also investigated all other measured data as shown in Table 4, but no significant differences was inferable from it. To analyse the user satisfaction, we looked for statistically significant differences between the groups by using Mann-Whitney U-Test. For the task-related questionnaire, we tested every question for every task. The results are presented in Table 5. We found three significant differences between group B and C. Here, question Q6 ("The Clipboard supported me in the conducting the task.") and question Q9 ("The images selected for Clipboard were suitable for the task.") show advantages for variant B over variant C in task 3 ($p_{Q6}^{T3} = .021$) and task 5 ($p_{Q6}^{T5} = .027$, $p_{Q9}^{T5} = .045$). In regard of the final questionnaire, we compared the groups pairwise with each other, obtaining results that are presented in Table 6. Here, no significant differences could be found.

		T1	T2	T3	T4	T5	T6
Q1	A-B	.7	.82	.88	.9	.95	.8
	A-C	.85	.55	.57	.75	.53	.6
	B-C	.76	.82	.65	.8	.8	.75
Q2	A-B	.55	.88	.82	.9	.3	.85
	A-C	.34	.6	.26	.6	.6	.14
	B-C	.94	.79	.14	.52	.8	.44
Q3	A-B	.08	.41	.73	.12	.7	.52
	A-C	.21	.14	.55	.6	.12	.29
	B-C	.6	.68	.68	.07	.3	.52
Q4	A-B	.23	.5	.73	.44	.2	.18
	A-C	.65	.97	.45	.4	.6	.29
	B-C	.29	.7	.97	.95	.65	.9
Q5	A-B	.36	.91	.91	.52	.25	.52
	A-C	.27	.4	.5	.67	.34	.46
	B-C	.7	.15	.27	.8	.24	.9
Q6	B-C	.07	.13	**.021**	.12	**.027**	.56
Q7	B-C	.26	.55	.6	.65	.9	.95
Q8	B-C	.57	.36	.43	.16	.8	.8
Q9	B-C	.5	.31	.07	.4	**.045**	.52
Q10	B-C	.31	.29	.15	.09	.7	1

Table 5: p-values of Mann-Whitney U tests of task-related questionnaire (significances marked)

	A-B	A-C	B-C
F1	.706	.880	.791
F2	.569	.620	.819
F3	.070	.791	.325
F4	N/A		.447
F5	N/A		.162
F6	N/A		.594
F7	N/A		.137

Table 6: p-values of Mann-Whitney U tests of final questionnaire

6 Discussion

In terms of effectiveness, all three variants can be considered similar. Although variant C has two participants with missing images, the interface for saving images is identical to variant B, where no image was missing. We have not observed any further differences with respect to the effectiveness between the three variants.

Regarding efficiency, all three variants are not significantly different from each other, besides for task 1 where group A was significantly faster than group C. In this task, the means of both groups were 15 seconds apart but due the fact that the task only took 30 seconds for A and 45 seconds for C at average makes it hard to derive further conclusions. As there are no further significant differences, we derive that the two different clipboard variants are equal to the Google Image Search and to each other. This shows on the one hand that the clipboard does not hinder users in executing tasks but also brings no advantages, what possibly indicates that the Google Image Search might be sufficient for most participants. Considering the average usage of Google Image Search of our participants, we can see that the group is familiar with this site.

The fact that the DCG-based clipboard (group B) got better results than the eyetracking-based (group C) is a surprise for us. For this, we investigated how often the clipboard was used in each task. For group B the average usage per task was 1.01 (SD: 0.36) and group C with 0.88 (SD: 0.42). Thus, the clipboard was only opened once or not at all in each task, although the participants were instructed to do so. This shows that the participants did only accept both clipboard variants to a minimum extent. We further investigated the amount of images shown in the clipboard when first opening it. We found out that group B had an average of 48.52 images in the clipboard (SD: 7.2) where group C had

only 15.4 images (SD: 5.57). We also checked how many times the clipboard contained less than 20 images, which is the predefined maximum number of images stored in the clipboard. For group B, this occurred three times, for group C 19 times. Given this difference, the lower number of selected images for variant C might lead to a less satisfying selection as in variant B.

7 Related Work

This work is related to approaches that use eyetracking devices as input devices and approaches that receive implicit information from collected eyetracking data to improve ranking (relevance feedback). Cosato et al. [2] used eyetrackers as input devices for their so called Rapid Serial Visualization Presentation methods for large data sets of images. They compared their methods with a grid-based presentation of images. There are also many other applications that make use of an eyetracker as input device like for example drawing shapes [5] or composing texts [11, 12]. A general problem of using eyetrackers as direct input device is that it was considered by participants as "unnatural" use of one's eyes in a focused way to control an application. By this, users get quickly tired using such a system. In contrast, in our approach we only collect normal gaze movements of the users while using an application they are familiar with. The users do not control the image search nor the clipboard by using the eyetracker as direct input device.

Another way to use eyetracking data is for implicit relevance feedback in information retrieval tasks. For instance, in textual information retrieval, Hardoon et al. [4] extracted implicit relevance feedback from eyetracking information. Another work is the Text 2.0 project from Buscher et al. [1]. They used eyetracking information to recognize text areas of interest and to provide additional content depending on these areas. Besides these text-based approaches, there are works like Pasupa et al. [9] that used eyetracking information to rank images. A work concerning the retrieval of relevance feedback from eyetracking data was published by Jaimes et al. [7]. They recognized different patterns of observation for images of different semantic categories. The system GaZIR, created by Kozma et al. [8], is an approach for visual image search engines using an eyetracker. In contrast to these systems, our approach aimed to use the feedback for filling the clipboard and so improving the usability of results, rather than improving the search itself. Also, prior work used predefined, closed sets of images, we use live data received from a popular web-based image search engine.

8 Conclusion

We have presented an approach to use eyetracking information in web-based image search engines. We have evaluated a baseline variant with two clipboard versions using different selection algorithms. To our surprise, all three system variants showed almost no significant differences ($\alpha < 0.05$). Comparing both clipboard variants, the selection algorithm using eyetracking information was

overall less favoured by the participants than the one using the DCG-based algorithm applied on the Google image results. We identified and discussed several properties of both clipboard variants that explain why this unexpected result may have happened.

Acknowledgement

The research reported here was partially supported by the SocialSensor FP7 project (EC under contract number 287975).

References

1. R. Biedert, G. Buscher, S. Schwarz, M. Möller, A. Dengel, and T. Lottermann. The Text 2.0 Framework – Writing Web-Based Gaze-Controlled Realtime Applications Quickly and Easily. In *Proc. International Workshop on Eye Gaze in Intelligent Human Machine Interaction (EGIHMI)*, 2010.
2. S. Corsato, M. Mosconi, and M. Porta. An eye tracking approach to image search activities using rsvp display techniques. In *Proc. Working Conference on Advanced Visual Interfaces*, AVI '08, pages 416–420, New York, NY, USA, 2008. ACM.
3. I. D. Günther Gedigaa, Kai-Christoph Hamborg. The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour and Information Technology*, 18:151–164, 1999.
4. D. R. Hardoon and K. Pasupa. Image ranking with implicit feedback from eye movements. In *Proc. Symposium on Eye-Tracking Research and Applications*, ETRA '10, pages 291–298, New York, NY, USA, 2010. ACM.
5. A. J. Hornof and A. Cavender. Eyedraw: enabling children with severe motor impairments to draw with their eyes. In *Proc. Conference on Human Factors in Computing Systems*, pages 161–170. ACM, 2005.
6. K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, Oct. 2002.
7. A. J. Jeff, R. Jaimes, J. Pelz, T. Grabowski, J. Babcock, and S. fu Chang. Using human observers' eye movements in automatic image classifiers. In *Proc. of SPIE Human Vision and Electronic Imaging VI*, pages 373–384, 2001.
8. L. Kozma, A. Klami, and S. Kaski. GaZIR: gaze-based zooming interface for image retrieval. In *Proc. international conference on Multimodal interfaces*, ICMI-MLMI '09, pages 305–312, New York, NY, USA, 2009. ACM.
9. K. Pasupa, A. Klami, C. J. Saunders, S. Kaski, S. Szedmak, and S. R. Gunn. Learning to rank images from eye movements. In *Proc. 12th International Conference on Computer Vision*, pages 2009–2016, 2009.
10. D. Tjondronegoro and A. Spink. Web search engine multimedia functionality. *Inf. Process. Manage.*, 44(1):340–357, 2008.
11. D. J. Ward and D. J. C. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418(6900):838, 2002.
12. X. A. Zhao, E. D. Guestrin, D. Sayenko, T. Simpson, M. Gauthier, and M. R. Popovic. Typing with eye-gaze and tooth-clicks. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 341–344, New York, NY, USA, 2012. ACM.