Factors Affecting Click-Through Behavior in Aggregated Search Interfaces

Shanu Sushmita
University of Glasgow
shanu@dcs.gla.ac.uk

Hideo Joho
University of Tsukuba
hideo@slis.tsukuba.ac.jp

Mounia Lalmas
University of Glasgow
mounia@acm.org

Robert Villa
University of Glasgow
villar@dcs.gla.ac.uk

ABSTRACT

An aggregated search interface is designed to integrate search results from different sources (web, image, video, blog, etc) into a single result page. This paper presents two user studies investigating factors affecting users click-through behavior on aggregated search interfaces. We tested two aggregated search interfaces: one where results from the different sources are blended into a single list (called *blended*), and another, where results from each source are presented in a separate panel (called *non-blended*). A total of 1,296 search sessions performed by 48 participants were analysed in our study. Our results suggest that 1) the position of search results is significant only in the blended and not in the non-blended design; 2) participants' click-through behavior on videos is different from other sources; and finally 3) capturing a task's orientation towards particular sources is an important factor for further investigation and research.

Categories and Subject Descriptors

H.5 [**INFORMATION INTERFACES AND PRESENTATION**]

General Terms

Human Factors, Measurement

Keywords

Aggregated search, result presentation, task-based user study, click-through behavior, blended and non-blended designs

1. INTRODUCTION

Aggregated search interfaces are now a common paradigm for search result presentation. An aggregated search interface is designed to integrate search results from different sources (web, image, video, news, blog, tweet, etc) into a single result page. An objective of aggregated search is to facilitate the access to the increasingly diverse content available on the web. There appears to be at least two main types of integration; **blended** and **non-blended**. A blended integration (as initiated by Google Universal Search[1], and now used by many other search engines) presents results from different sources within a single ranked list, whereas a non-blended integration (e.g., Alpha Yahoo![2] or Naver[3]) presents results from each source in a separate panel.

Although a large number of studies devoted to the design and evaluation of conventional web search interfaces have been reported in the literature (e.g. [8, 22, 23]), less is known about aggregated search interfaces. The main aspect distinguishing an aggregated search interface from a conventional one is that results from different sources are integrated within the default web search results. This increases the complexity in the design of search interfaces and result presentation. This distinction raises two main issues:

1. How should we determine the relevance of a source to a search task when given a query?
2. How should we organise the search results with multiple sources?

There have been several studies looking at the first issue e.g. [17, 20, 19]. On the other hand, research on the second issue in the context of aggregated search interfaces is still limited. To examine the second issue, we carried out two user studies: one using a blended design and another using a non-blended design. The objective of our studies was to investigate the impact of factors on people's click-through behavior on the aggregated search result presentation. The factors we studied included position of search results, source types, and strength of search task's orientation towards a particular source type.

We elaborate on the last factor. An important task of an aggregated search system is to determine, for a given query, not only whether a document is relevant, but from which sources relevant documents should be retrieved. To capture this aspect, we introduce the notion of the "source orientation" of an information need to refer to the degree to which documents *from* a specific source would be relevant to complete the corresponding search task. It should

[1]http://googleblog.blogspot.com/2007/05/universal-search-best-answer-is-still.html
[2]au.alpha.yahoo.com/
[3]www.naver.com/

be emphasised that it was not the aim of this work to study how to select the sources from which documents should be retrieved. Instead, given that we know the source orientation of an information need, we aim to investigate the effect of this on users click-through behavior. Source-orientation can be seen as a dimension of relevance [3], one necessary to investigate aggregated search.

To summarise, the work reported in this paper is concerned with the effect of the source orientation of an information need, the positioning of search results, and source types on users click-through behavior. We study these effects for both bended and non-blended aggregated search interfaces.

1.1 Research Questions

This subsection defines the research questions investigated in our study. In this paper, we refer to conventional web search results as the *base elements* of aggregated result presentation, and those from all other sources such as images, videos, and news, as *augmented elements*.

R1 *How does the position of augmented elements affect users' click-through behavior in aggregated search interfaces?*

It has been shown in previous studies that click-through behavior is strongly affected by the position of results in the ranked list [9, 10, 13, 15]. However, less is known for aggregated result presentation. Therefore, we aim to investigate how the position of augmented elements affect user behavior, in both blended and non-blended designs, to gain a further insight into this problem.

R2 *How does the source orientation of an information need affect users' click-through behavior in aggregated search interfaces?*

While the thematic (or topical) relevance of documents is important, we argue that the source type may also be an important dimension in the relevance space for investigating aggregated search, in a similar manner to how structural relevance is considered as another dimension of relevance in XML retrieval research [14]. Therefore, we decided to incorporate the source orientation factor into our experimental design.

R3 *How does the effect of position and source orientation differ over the source types of augmented elements?*

The complexity of aggregated search interfaces partly stems from the range of potential sources that can be aggregated, and because of this, it is important to understand the influence different types of source can have on the use of the augmented elements. Therefore, the last research question is concerned with the interaction effect between the augmented source types and the other two factors described in R1 and R2 (position and source orientation).

1.2 Structure

The rest of the paper is structured as follows. The experimental design to address our research questions is described in Section 2. Section 3 reports the results from our studies and their analyses, whereas Section 4 contains a discussion of our findings. Related work is presented in Section 5. Finally, we finish with conclusions and future work in Section 6.

2. EXPERIMENTAL DESIGN

To investigate the factors affecting click-through behavior in aggregated search interfaces, we carried out two separate within subject studies with 48 participants (24 using a blended interface design and 24 using a non-blended one). It should be emphasised that it was not the aim of this work to compare the performance of blended and non-blended interfaces, but instead to make separate observations for each. Since these two main approaches to aggregating results exist (blended and non-blended), we tested the factors for both approaches.

Research reported in [19] and [20] showed that the typical pattern with multiple-sources (in terms of click-through and relevance) was that at most two sources are involved, with the conventional web source being highly predominant. As a consequence, in this paper only the aggregation of two sources (conventional web and one other source) is considered at any one time. In addition, we chose image, news and video as the non-web sources, as these were some of the most frequently used sources in existing aggregated search interfaces.

The factors examined in our study were represented by three independent variables: position, source of the augmented elements, and the source orientation of the search task. Each variable had three levels, as listed below:

1. **Position (P):**
   Blended: Top, Middle and Bottom (Figure 1).
   Non-Blended: Left, Top-Right and Bottom-Right (Figure 2).
2. **Augmented Element (AE):** Image, News and Video.
3. **Source-Orientation (SO):** High, Medium and Low.

To measure the effect of the independent variables on users click-through behavior, we proposed two dependent variables: the frequency of clicks on augmented elements, and the ratio of those clicks that are bookmarked by participants. Our participants were asked to bookmark the search results perceived to be needed to complete the search tasks.

Deciding which interface positions augmented elements could be placed and thus examined in this study was not trivial. There are many possible combinations in both the blended and non-blended interface designs. Based on the observations of existing aggregated search interfaces, we decided to focus on the positions listed above, with further examination of other positions left for future work.

For both (blended and non-blended) studies, we used the same experimental design, unless otherwise stated. That is, the same number of participants, the same search tasks and topics, the same duration of experiments, and the same questionnaires. The same search results were shown on both interfaces.

2.1 Interfaces

Two separate aggregated search interfaces were designed, one for each type of aggregation. A blended interface (Figure 1) was used to study the factors affecting blended aggregation, and a non-blended interface (Figure 2) was used to investigate the factors affecting non-blended integration.

Each interface contained a fixed set of search results, one set for each of the twenty seven topics used in the experimented (we describe in the next section the search tasks

and topics). Participants were shown one result page at a time, one for each topic. All the search results were fetched prior to the experiment, so that all participants were presented with the same set of results for a given topic. We used the Yahoo! search API to fetch the results using the topic texts as queries. For each query, the results fetched corresponded to the top 10 results from the conventional web, and the top 5 from each of the other three sources. We kept the same ranking of results within each source (no re-ranking was performed). The only difference between the blended and non-blended interfaces was the organisation of the results (i.e. their positions on the interface).

For a given topic, the interface was set to display the corresponding result page showing 15 items from two selected sources. These items were fixed and were the same for all participants. The result page showed 10 results from the conventional web (base elements) and five from the non-web source (augmented elements). These sizes are similar to typical aggregated interfaces. The positions of the augmented elements (image, news and video) were varied; at the top, the middle and the bottom of the web results in the blended interface, and left, top-right and bottom-right in the non-blended interface. The base and the augmented elements were not merged[4] in our blended design. This is the strategy adopted by most current search engines (with the sources used in our study), and avoids the need to devise a merging algorithm, and consider its effect on users click-through behavior.

Since three positions were to be tested in both interfaces, this resulted in three panels for the non-blended design. To use the same number (2) of sources in both studies, two panels were used to display, respectively, the base elements and the augmented elements; the third panel was kept empty. To distinguish between the three panels, the empty panel had a light background colour (see Figure 2). Questions in the exist questionnaire checked whether participants were distracted or disturbed by the presence of the coloured panel. It was found that 98% of the participants did not notice the coloured panel. Those who did (2%) confirmed that the colour did not distract them. Therefore, we can assume that the coloured panel did not lead to any biased observations for the non-blended design.

2.2 Tasks

Since our experimental design had three independent variables (position, source of the augmented elements, and source orientation of the search task) and each variable had three levels, we formulated one search task (a topic and a task description) for each combination of these variables ($3 \times 3 \times 3 = 27$) as shown in Table 1. This led to a total of twenty-seven search tasks.

The degree of source orientation of a task was determined according to our intuition. For instance, we speculated that a highly video-oriented search task was one where some visual learning would be highly desirable (e.g. dance steps, aerobic movements). A mediumly video-oriented search task was one where we assumed that video results may or may not be required; e.g. for the topic "baking pancakes", some participants might prefer reading a recipe rather than watching a video.

For each topic, participants were shown the search results

[4]In search engine terminology, we slotted them at three different positions on the result page (top, middle and bottom).

Table 1: Position and source orientation combinations for each source. For blended P_1=top, P_2=middle and P_3=bottom. For non-blended P_1=left, P_2=right-top and P_3=right-bottom. Here, each cell (combination c_n) is tested for three sources, image, video and news. Thus leading to twenty-seven combinations.

Position	High	Medium	Low
P_1	c_1	c_2	c_3
P_2	c_4	c_5	c_6
P_3	c_7	c_8	c_9

Figure 3: Example of a bookmarked page, showing web page with an extra feedback bar at the top of the page, which allowed the users to mark the result as relevant or not relevant.

associated with that topic, displayed in the blended or the non-blended design. Participants were asked to bookmark those results they judged useful in completing the task (see Figure 3). A brief task description for each topic was provided, which did not refer to the source orientation of the corresponding task (e.g. words such as photo, image, video, were not used). For instance, for the topic *salsa dance*, the task description provided was *"provide examples of dance steps"*.

The twenty-seven topics were distributed into three sessions. Each session was composed of all search tasks with the same level of source orientation: high, medium or low. What then varied in each session were the source of the augmented elements (3) and the position of the augmented elements (3), leading to (3×3) 9 topics per session (as seen in Table 1).

Every result page was timed for two minutes and participants were asked to spend it in their usual search manner. That is, participants could view all 'fifteen' results if they wished or just 'one' result within the two minute time period. On average each session lasted for 18-20 minutes (2 minutes each for 9 topics). The sessions were rotated to minimise learning effects, and ordering effects were reduced by allowing participants to select the topics in random order within each session.

2.3 Participants

The experiment for the blended integration was carried

(a) Image on top (b) Image in the middle (c) Image at the bottom

Figure 1: Examples of result pages in the blended interface.

(a) Image on right-top (b) Image at the right-bottom (c) Image in the left

Figure 2: Examples of result pages in the non-blended interface.

out by 18 males and 6 females from the University of Glasgow, of which 16 were graduates, 7 postgraduates and 1 an undergraduate. The experiment for the non-blended integration was carried out by 18 males and 6 females, of which 1 was a graduate, 21 were postgraduates and 2 were undergraduates. To prevent any learning effect or biases, participants from one study were not involved in the second study. Participants in each study did all twenty-seven topics, thus leading to a total of $(27 \times 24 \times 2)$ 1296 sessions.

The participants were from various educational fields, e.g. computing, law, life sciences, real estate, aerospace, business management, arts and commerce. Participants were recruited through an email distributed to several mailing lists. An entry questionnaire was used to capture participants' profile and search background. As the experiment was divided into three sessions, post-session questionnaires were used to capture subjective assessments on the sessions and tasks. Exit-questionnaires provided participants' perceptions of sessions and tasks as a whole. Participants were rewarded fifteen pounds (GBP) for their participation.

3. RESULTS AND ANALYSIS

This section presents the results of our experiments. In Section 3.1, we present the overall results of the main and interaction effects of position, task source orientation, and augmented element type on users click-through behavior. We then look into the details of each of the independent variables in Sections 3.2, 3.3, and 3.4, respectively. Finally we present the result of participants' perceptions of the augmented elements in Section 3.5.

3.1 Overall Result

3.1.1 Click-through frequency

We first looked at the overall frequency with which augmented elements are clicked for the two designs of aggregation. The results are shown in Table 2, which shows the total number of clicks on each augmented element, the percentage with which augmented elements were clicked compared to the web results, the mean number of augmented click-throughs per session, and the percentage of clicked augmented elements which were also bookmarked. The bottom row of the table shows that the total number of clicks on the augmented elements was comparable between the blended and non-blended designs (978 and 1039). The ratio of the clicks was also found to be comparable (denoted as % Click in Table 2). This is perhaps unsurprising since the set of presented results was the same, the only difference being their organisation. When we look at the click frequency across the augmented element types, again no large difference was observed within each type of aggregation, as well as across the two types of aggregation. These results suggest that the studies of the blended and non-blended designs are comparable in our analysis of click-through frequency.

A noticeable difference, however, was the ratio of bookmarked results over those clicked (denoted as % Bookmark in Table 2). Participants were asked to bookmark the results when they were perceived to be relevant to the search tasks. The bottom row of the table suggests that participants bookmarked the clicked results more frequently in the blended design than the non-blended design. This suggests that, overall, participants performed 10% better in the blended design than non-blended in finding perceived relevant documents from the augmented elements. While the performance is not the focus on this paper, this could be considered as an advantage of the blended design. For the rest of this paper, however, we mainly report the click-through frequency results, due to space limitations.

3.1.2 ANOVA tests

The next analysis performed looked at the significance of the effect of the three independent variables such as; Posi-

Table 2: Distribution of clicks. Total = total number of clicks on augmented elements, % click = percentage of clicks on augmented elements (augmented element/(web+augmented element)), Mean (SD) = mean and standard deviation of click-through frequency on augmented elements, and % Bookmark = percentage of bookmarked results (number of bookmarked augmented elements/number of clicks on augmented elements).

	Blended				Non-Blended			
Source	**Total**	**Click %**	**Mean (SD)**	**% Bookmark**	**Total**	**Click %**	**Mean (SD)**	**% Bookmark**
Image	360	33.7	1.7 (1.72)	53.6	410	37.3	1.9 (1.28)	41.5
Video	324	34.4	1.6 (1.48)	63.6	339	33.0	1.6 (1.21)	58.4
News	294	30.1	1.4 (1.59)	46.9	290	28.6	1.3 (1.39)	33.4
All	978	32.8	1.6 (1.60)	54.9	1039	33.1	1.6 (1.31)	44.8

Table 3: Statistical significance (p-values) of the main and interaction effects of the position (P), Source-Orientation (SO), and Augmented Element type (AE) on the click-through ratio.

Factors	Blended	Non-Blended
P	**0.0000**	0.5437
SO	**0.0000**	**0.0000**
AE	**0.0171**	**0.0005**
PxSO	0.1084	0.0632
PxAE	**0.0000**	0.6641
SOxAE	**0.0000**	**0.0000**
PxSOxAE	**0.0000**	**0.0110**

tion (P), Source-Orientation (SO), and Augmented Element type (AE), on participants' click-through behavior. ANOVA tests were used to examine the significance of three main effects (P, SO, AE), three interaction effects of two factors (PxSO, PxAE and SOxAE), and one interaction effect of three factors (PxSOxAE).

The raw click-through frequencies were first converted into ratios (click-through frequency on augmented element/total click-through frequency) before performing the ANOVA test. When the difference was found to be significant among the levels of independent variables, we further performed multiple post-hoc tests to find significant pairs among the factor levels. We ran multiple t-tests in the post-hoc analysis while adjusting the p-values using the Bonferroni method, where the critical value was set to 0.05/number of multiple comparisons.

The overall ANOVA results are given in Table 3, which shows the p-value of the tests. The significant effects are highlighted with bold. It can be seen that the position effect was found to be significant in the blended design but not in the non-blended design. The effect of the source orientation and augmented element types was found to be significant in both designs of aggregation. This suggests that some factors affected click-through behavior in only one of the aggregation designs, while other factors affected both designs.

Furthermore, an interaction between the position and augmented element type was found to be significant only in the blended design, while an interaction effect between the source orientation and augmented element type was found to be significant in both designs. This indicates that participants' click-through behavior was not affected by a single factor but by a combination of multiple factors. Finally, the bottom row of Table 3 shows that there was a significant interaction of the three factors on participants' click-through behavior.

Table 4: Blended: Click-through frequency across different positions.

	Top	Middle	Bottom
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.2 (1.90)	1.5 (1.43)	1.5 (1.73)
Video	1.8 (1.53)	1.6 (1.41)	1.3 (1.46)
News	1.8 (1.71)	1.1 (1.25)	1.4 (1.71)
All	1.9 (1.72)	1.4 (1.38)	1.4 (1.63)

Table 5: Multiple t-tests (p-value) on position combinations in blended design

Position	Blended
Top-Middle	**0.0010**
Top-Bottom	**0.0000**
Middle-Bottom	0.1059

These results are helpful in presenting a high-level overview of the ANOVA tests. The following subsections will discuss these results in greater detail.

3.2 Position Effect

Our first research question [**R1**] considered the effect of the augmented elements' position on users' click-through behavior. We will show the result of the blended design first (Section 3.2.1), followed by the non-blended design (Section 3.2.2), and finally, a comparison between the two designs is discussed as summary (Section 3.2.3).

3.2.1 Blended

As we saw in the previous section, the position had a significant effect on participants' click-through behavior in the blended design. Table 4 shows the breakdown of the click-through frequency based on the three levels of positions (i.e., top, middle, and bottom). The overall mean click frequency (row "All" in Table 4) shows that there is a position effect between the top and middle/bottom positions, but less so between the middle and bottom positions. We performed multiple comparisons to test that the differences observed were significant. These post-hoc tests revealed that differences between the top and middle, and top and bottom positions were significant (Table 5). This suggests that participants clicked the augmented elements more frequently when they were located at the top of the results than the other positions.

Table 6: Non-Blended: Click-through frequency across different positions.

	Left	Top-Right	Bottom-Right
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.0 (1.36)	1.8 (1.21)	1.9 (1.27)
Video	1.6 (1.17)	1.5 (1.21)	1.6 (1.26)
News	1.4 (1.48)	1.3 (1.47)	1.3 (1.21)
All	1.7 (1.37)	1.6 (1.32)	1.6 (1.24)

Table 7: Blended: Click-through frequency across different source orientation levels.

	High	Medium	Low
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.5 (1.92)	1.4 (1.47)	1.3 (1.46)
Video	1.7 (1.66)	1.9 (1.43)	1.2 (1.26)
News	2.2 (1.89)	1.0 (1.20)	1.0 (1.25)
All	2.1 (1.85)	1.4 (1.41)	1.2 (1.33)

3.2.2 Non-Blended

Unlike the blended design, the position effect was found to be insignificant in the non-blended design. The breakdown of the click-through frequency based on the position of the non-blended design can be found in Table 6. As can be seen from the bottom row of the table, the overall frequency was similar across the position levels. On the other hand, there appears to be some difference among the augmented source types, which will be discussed in Section 3.4.2.

3.2.3 Summary

To summarise, our results for research question one [**R1**] suggest that there is a significant position effect on users' click frequency in the blended interface, whereas in the non-blended interface, the position of augmented elements does not seem to affect users' click frequency.

3.3 Source Orientation

The second research question [**R2**] looks at the effect of the source orientation of an information need on users' click-through behavior. The order of the experimental results is the same as Section 3.2.

3.3.1 Blended

As we saw in Table 3, the source orientation was found to be significant in the blended design. The breakdown of the results based on the three levels of the orientation is shown in Table 7. As can be seen from the bottom of the table, the click-through frequency decreases as the level of source orientation weakens. The post-hoc tests show that the differences between the high level and middle level, and between the high level and low level, are significant (second column of Table 8). This suggests that participants clicked the augmented results more frequently when an information need has a strong orientation towards a particular information source type, in the blended design.

3.3.2 Non-Blended

Like the blended design, participants' click-through frequency was also significantly affected by the source orientation in the non-blended design. The breakdown of the results based on the three levels of orientation in the non-blended design is shown in Table 9. Again, the frequency appears to decrease as the level of orientation weakens. The post-hoc tests show that the differences between the high level and medium level, and between the high level and low level, are significant (third column of Table 8). This suggests that participants clicked the augmented results more frequently when the task's source orientation was high in the non-blended design.

Table 8: Multiple comparison for source orientation combinations on click-through frequency.

Orientation	Blended	Non-Blended
High-Medium	**0.0000**	**0.0000**
High-Low	**0.0000**	**0.0000**
Medium-Low	0.4268	0.0619

Table 9: Non-Blended: Click-through frequency across different source orientation levels.

	High	Medium	Low
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.5 (1.31)	1.6 (1.17)	1.6 (1.16)
Video	1.6 (1.14)	1.9 (1.30)	1.2 (1.11)
News	2.3 (1.44)	0.9 (1.17)	0.9 (1.03)
All	2.1 (1.35)	1.5 (1.28)	1.2 (1.14)

3.3.3 Summary

As we have seen, the effect of the source orientation was found to be equally significant in both the blended and non-blended design. The standard deviation in the bottom row of Tables 7 and 9 suggests that this trend is slightly more consistent in the non-blended design than the blended design. However, in terms of the frequency, the results are comparable. This suggests that the effect of the source orientation is an important factor to be investigated in aggregated search research.

3.4 Interaction across Augmented Elements

Our last research question [**R3**] aims to investigate the interaction effect between the augmented source types and the position or source orientation. We were interested to see if the interaction effect of factors such as position and source orientation vary, or were similar across augmented elements. To answer this question we compared the click-through data from image, video and news results across different levels of position and source orientation. For each interface, we first report results for the interaction effect of position and augmented elements (PxAE), and then we present results for the interaction of source orientation and augmented element (SOxAE). Finally, we compare the interaction effect between the two aggregation designs.

3.4.1 Blended

Returning to Table 3 again, we can see that the interaction effect of position and augmented element (PxAE) is significant in the blended interface (fifth row of Table 3). This suggests that participants' click-through behavior was affected by a combination of the two factors. To understand this effect better, we plotted an interaction plot using the position and augmented elements. The result is shown in Figure (a) in Table 10.

The interaction plot shows that the pattern of videos is

different from images and news. The post-hoc tests show that, at the middle position, participants clicked video results more frequently than images or news results (See Table (a) in Table 10).

Next, results from the ANOVA tests suggest that the interaction effect of source orientation and augmented element (SOxAE) is another significant factor affecting user's click-through behavior in the blended design (sixth row of Table 3). The interaction plot of the two factors is shown in Figure (a) of Table 11. Again, the pattern of video results appears to be different from images and news results. The post-hoc tests show that, at the middle level of source orientation, participants clicked the video results more frequently than the other sources in the blended design (See Table (a) of Table 11).

3.4.2 Non-Blended

For the non-blended design, an interaction effect of position and augmented element type (PxAE) was not found to be significant. As can be seen from the interaction plot shown in Figure (b) of Table 10, all sources show a similar pattern.

On the other hand, participants' click-through frequency was affected significantly by the interaction effect of the source orientation and augmented element types (SOxAE). The interaction plot of the two variables is shown in Figure (b) of Table 11. As can be seen from the plot, the video results again have a different pattern from the image and news results. While the significant effect was found only in the middle level of source orientation in the blended design, the non-blended design have significant effects in the high and low levels of orientation, too (See Table (b) of Table 11).

3.4.3 Summary

A highlight of the above analysis is the difference in the click behaviour for video results, compared to news and image. We observed multiple cases where the frequency pattern of videos differed from that of images and news. Videos tended to have a low frequency even when the position was high or source orientation was high. On the other hand, participants tended to click video results more frequently than the other sources when the position was middle or the orientation was middle.

We will discuss the implications of this finding in Section 4.

3.5 User Perception

Recall that our experimental design had three sessions (Section 2) with each session corresponding to one level of source orientation (high, medium and low). After each session, participants were asked to provide their preferred position for the augmented elements on the result page. More specifically, participants were asked if they prefer to view non-web results (image, news, etc.) at the top, middle or bottom of the result page for the blended interface, and left, top-right or bottom-right of the result page for the non-blended interface. We now present the results from the questionnaires for both the blended and non-blended studies.

The questionnaire results are shown in Figure 4. Results from the blended interface study suggest that for highly oriented sessions users prefer to have augmented results on the top. Users tended to get less specific and their preferences

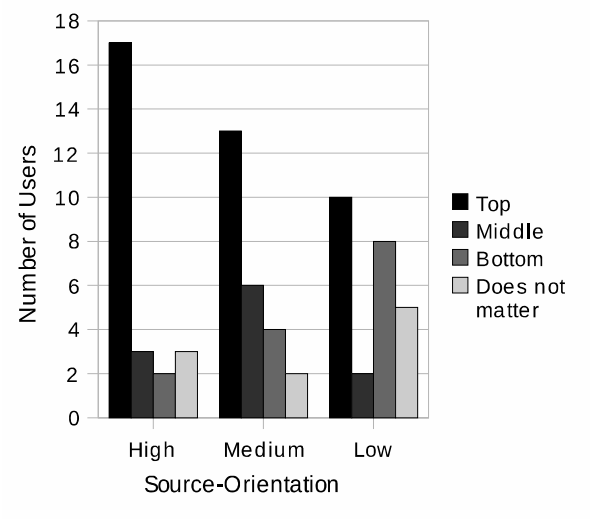


Figure 4: Blended: User preference on position of augmented elements across source orientations.

becomes mixed when the source orientation decreases. This is consistent with our findings from the click data where a clear position effect was observed on users' click-through behavior (Section 3.2.1)

For the non-blended interface, results show that user preferences are mixed when considering the position of augmented results for different orientations (Figure 5). Also, many users suggested that since they were able to see all the panels (base and augment elements) on the screen without having to scroll, the position of the panels did not matter. Again, this is consistent with our findings from the click data analysis, where position was not significant for the non-blended interface (Section 3.2.2), although there is a trend for users to prefer the top-right position for the display of non-web results.

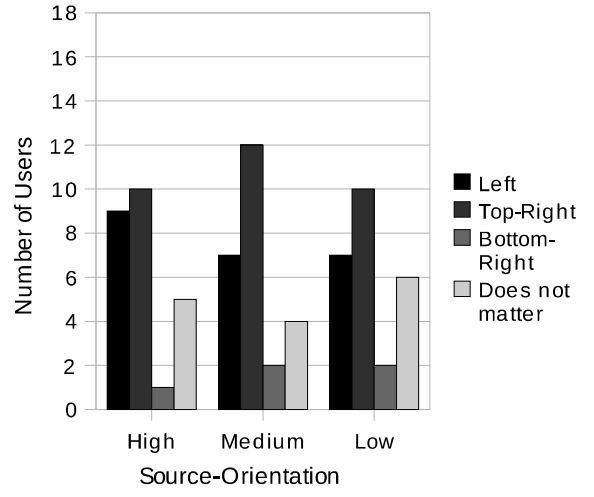
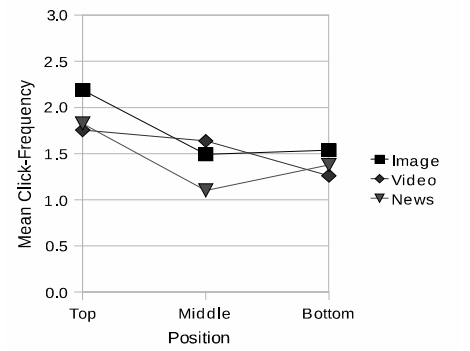


Figure 5: Non-Blended: User preference on position of augmented elements across source orientations.

Table 10: Interaction effect of position and augmented elements (PxAE) in the blended design (Figure (a)) and non-blended design (Figure (b)). Tables (a) and (b) show the results of post-hoc pair wise comparisons.

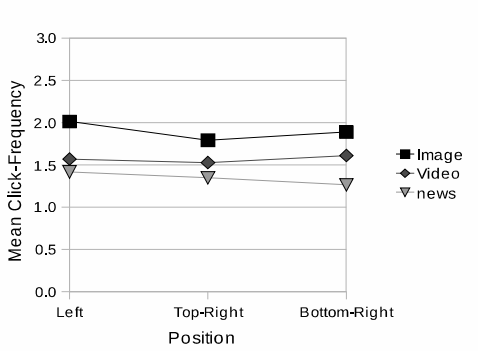
Figure (a)



	Top	Middle	Bottom
Image-Video	0.6759	**0.0023**	0.9853
Image-News	0.5643	**0.0159**	0.4724
News-Video	0.8619	**0.0000**	0.4786

Table (a)

Figure (b)



	Left	Right-Top	Right-Bottom
Image-Video	0.5597	1.0000	1.0000
Image-News	0.1507	0.1490	**0.0147**
News-Video	0.3677	0.5451	0.1238

Table (b)

4. DISCUSSION

Our work was motivated by the prominent complexity of aggregated search interfaces compared to the conventional single-source design. Despite the complexity, there has been a limited number of studies looking at the effects of factors such as augmented elements' positions, source types, and the search tasks source orientation, on people's search behavior.

As we mentioned earlier, it was not our intention to decide which type of aggregation is more effective than the other. Instead, we were interested in understanding the characteristics of both designs so that we can leverage their advantages depending on the context of system use. The following discusses the main findings of our study and their implications on the design of aggregated search interfaces.

Our first finding is that the factors that affect participants' click-through behavior differ between the blended and non-blended designs. This may sound obvious, however this should not be underestimated, since it suggests that the way in which we present results from different sources indeed matters. For example, participants' click-through behavior was significantly affected by the position of augmented elements in the blended design, echoing the findings of previous studies [9, 10, 13, 15], yet not in the non-blended design. This suggests that we need a careful estimation of the relevance of augmented elements with respect to the base elements when the blended design is employed. When we cannot afford to measuring the relevance of augmented elements, the non-blended design is more appropriate since participants' click-through behavior was not affected by the position in this type of aggregation. Such a situation may arise in digital libraries and elsewhere. These results address our first research question [**R1**]

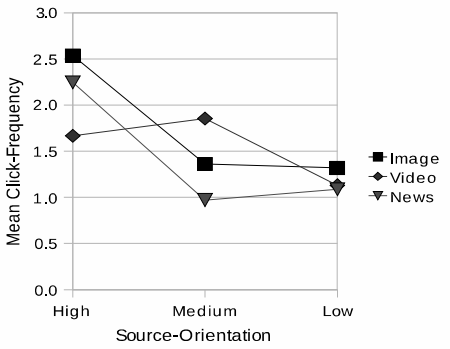
The second finding is that videos resulted in a different click-through pattern from news and images. This trend was common in both the blended and non-blended designs. This suggests that, when deciding to retrieve videos, different behavior from other sources may be observed. Or more generally, our study suggests that participants' click-through behavior can be different across source types. These observations address our third research question [**R3**]

While it is not entirely clear why videos are different, it is possible to suggest some possible reasons. First, videos are multimodal media [16] combining text, images, and audio in a dynamic way. The dynamism and multimodality of the information source might cause a user to give a different priority to videos during a search task. The different priority may cause the different click-through pattern of videos when compared to news and images. Secondly, it may be due to the type of surrogate used to represent videos results, being less informative of the contents of the video than the equivalent image or text representations. The title of news articles and the thumbnail of an image can provide a good indication of the respective content of the documents. On the other hand, although basic metadata of the videos were presented, this may not have been as informative of the contents. This difficulty in getting the preview of videos might cause the different click-through pattern [11]. It should be noted that, the task time limit of 2 minutes did not seem to discourage users from viewing video results.

The third finding was that a search task's orientation towards a particular source can affect participants' click-through behavior. This trend was common to both the blended and non-blended designs. Traditional information retrieval research has been focused on the modelling of thematic (or topical) relevance of documents. However, research on XML document retrieval [14] and geographic information retrieval (GIR) [6] has demonstrated that relevance can be multidimensional, there being a structural relevance in XML retrieval, and geographic relevance in GIR, which can be considered apart from the thematic relevance. In a similar way, our experimental design controlled the level of orientation towards a particular source (i.e., news, images, and videos). The significant effect of the source orientation observed in our experiments suggests that the task's source

Table 11: Interaction effect of source orientation and augmented elements (SOxAE) in the blended design (Figure (a) and Table (a)) and non-blended design (Figure (b) and Table (b)).

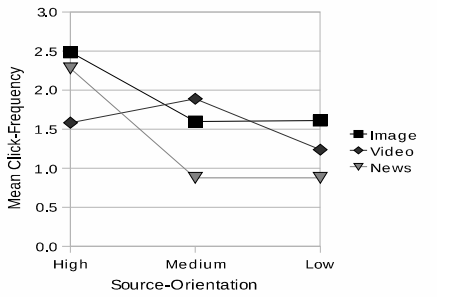
Figure (a)



	High	Medium	Low
Image-Video	0.0975	**0.0010**	0.1611
Image-News	0.8643	**0.0175**	0.6302
News-Video	0.1099	**0.0000**	0.0708

Table (a)

Figure (b)



	High	Medium	Low
Image-Video	**0.0001**	**0.0028**	0.2394
Image-News	0.8904	**0.0006**	**0.0000**
News-Video	**0.0011**	**0.0000**	**0.0045**

Table (b)

orientation is an important factor to investigate in research on aggregated search. On the design level, it suggests that devising a means of capturing a searcher's intent about the source is an important problem to tackle. These observations address our second research question [**R2**].

Finally, we should clarify the limitations of this study. First, we generated a priori the results to be displayed on the interfaces. Although this made the investigation fair, the implications of our results is limited to this particular set of search results. Second, to reduce the complexity and duration of the experiment, we tested only three sources; image, video and news. Finally, we used only a single search engine throughout our experiments. Therefore, our findings may not apply to different environments. Further studies should be carried out to deepen our understanding of aggregated search interfaces. Also, we did not make any attempt to ensure that the documents retrieved by the search engine were relevant to the tasks, to ensure that the presented results were representative of real world search engine performance. However, as the ratio of bookmarked documents over all clicked documents suggests, participants did perceive some of the retrieved documents as relevant in order to complete the search tasks. Furthermore, one of the authors of this paper was there to deal with any technical problems during the experiment, and we did not observe cases where participants were deliberately bookmarking clearly irrelevant documents. Therefore, we assert that the backend engine did retrieve relevant documents and participants were able to find some of them although we did not perform any quality control of retrieval.

5. RELATED WORK

The usefulness of aggregated search systems has already been validated in a number of studies. For example, [17] showed the positive effect (in relevance ranking) of properly integrating news within web results using a blended design, [19] showed through a log analysis that users access documents from various sources within the same search session, and [18] demonstrated through a user study the effectiveness of aggregated search for non-navigational queries.

Aggregated search can be compared to federated search [12], metasearch [7] and distributed information retrieval [5], which are all concerned with the retrieval of documents from several sources (e.g. digital libraries, search engines, specialist collections, etc.). Aggregated search can be seen as an alternative, where results from the conventional web, verticals of the *same* search engine, and other sources (such as Tweeter and Facebook) covered by that same search engine are retrieved. In aggregated search, the sources have a more distinctive nature (e.g. image vs. text, compared to distributed text collections more common in federated search). In addition, for a large percentage of queries, relevant results mostly come from the conventional web. However, for an increasing number of queries (e.g. person, red car), results from other sources (the augmented elements in aggregated search) are relevant, and thus could be *added* to the standard web results (base elements). This raises two main issues, as discussed in the introduction, the relevance of the sources, and the position of the augmented elements.

Recent work has looked at identifying and/or predicting the relevance of a source to a given query [17, 20, 18]. Our paper is concerned with the second issue, more precisely, with the impact of the relevance of the sources on user behavior. Nonetheless, this previous work has influenced some of the design of our study, e.g. the number of sources to be considered at any one time.

Significant efforts have been devoted to the design and evaluation of conventional web search interfaces. "Advanced" factors such as interface attractiveness [8] and result trustworthiness [22] have been evaluated. A taxonomy of result presentation techniques has been proposed as a reference for designers of web search systems [23]. However, there is still little understanding of what makes a well-designed, attractive, trustworthy and engaging aggregated search interface. This paper provides some initial insights, by looking at basic factors, such as position, source and source relevance, and

their effect on users' click-through behavior.

Studies, e.g. log analysis and eye-tracking experiments, which look at the effect of result position in the context of conventional web search, are not new. For instance, [13] showed that when results were placed relatively low in the result page, people spent more time searching and were less successful in their search task. Similar behavior is likely to be observed with aggregated search interfaces; however, it remains interesting to see the effect of source orientation on this. This has motivated us to investigate the position effect across results from different sources.

6. CONCLUSION AND FUTURE WORK

Aggregation is an emerging paradigm in search result presentation, which has many unexplored questions. We tested two aggregated search interfaces: one where results from the different sources are blended into a single list (*blended*), and the other, where results from each source are presented in a separate panel (*non-blended*). A total of 1,296 search sessions performed by 48 participants were analysed. Our studies led to three main findings. First, the position of search results was only significant for the blended interface. Second, participants' click-through behavior on videos was different compared to other sources. Finally capturing a task's orientation towards particular sources is an important factor to consider when considering the use of an aggregated search design.

These findings led us to formulate new research questions for future work: "How can we accurately incorporate the user behavior into retrieval models when the behavioral patterns differ across the source types?", "How will the click-through pattern on augmented elements change as the number of aggregated sources increase?", and "What kind of interaction design will help us capture a user's source preference of an information need?".

Acknowledgements: This work was carried out in the context of research partly funded by a Yahoo! Research Alliance Gift. Mounia Lalmas is currently funded by Microsoft Research/Royal Academy of Engineering.

7. REFERENCES

[1] O. Zamir and O. Etzioni. Grouper: a dynamic clustering interface to Web search results. *WWW*, pp 1361–1374, 1999.

[2] J. Goldstein, V. Mittal, J.Carbonell and M. Kantrowitz. Multi-document summarization by sentence extraction. *NAACL-ANLP Workshop on Automatic summarization*, pp 40–48, 2000.

[3] E. Cosijn and P. Ingwersen Dimensions of relevance. *IP&M*, 36(4):533–550, 2000.

[4] J. Reid. A Task-Oriented Non-Interactive Evaluation Methodologyfor Information Retrieval Systems. *Information Retrieval*, 2(1):115–129, 2000.

[5] Y. Rasolofo, F. Abbaci and J. Savoy. Approaches to collection selection and results merging for distributed information retrieval. *ACM CIKM*, pp 191–198, 2001.

[6] R. Purves and C. Jones, Chris. SIGIR Workshop on geographic information retrieval. *SIGIR Forum*, 38(2):53–56, 2004.

[7] A. Gulli and A. Signorini. Building an open source meta-search engine. *WWW*, pp 1004–1005, 2005.

[8] I. Nakarada-Kordic and B. Lobb. Effect of perceived attractiveness of web interface design on visual search of web sites. *ACM SIGCHI* , pp 25–27, 2005.

[9] T. Joachims, L. Granka, B. Pan, H. Hembrooke and G. Gay. Accurately interpreting clickthrough data as implicit feedback. *ACM SIGIR*, pp 154–161, 2005.

[10] E. Agichtein and Z. Zheng. Identifying "best bet" web search results by mining past user behavior. ACM SIGKDD, pp 902–908, 2006.

[11] Y. Song and G. Marchionini. Effects of audio and visual surrogates for making sense of digital video. *ACM SIGCHI*, pp 867–876, 2007.

[12] M. Shokouhi, M. Baillie and L. Azzopardi. Updating collection representations for federated search. *ACM SIGIR*, pp 511–518, 2007.

[13] Z. Guan and E. Cutrell An eye tracking study of the effect of target rank on web search. *ACM SIGCHI*, pp 417–420, 2007.

[14] M. Lalmas and A. Tombros. INEX 2002-2006: Understanding XML retrieval evaluation. *ICDL*, pp 187–196, 2007.

[15] M.T. Keane, M. O'Brien and B. Smyth Are people biased in their use of search engines? *CACM*, 51(2):49–52, 2008.

[16] M. Halvey, D. Vallet, D. Hannah and J.M. Jose. ViGOR: a grouping oriented interface for search and retrieval in video libraries. *ACM/IEEE-CS JCDL*, 87–96, 2009.

[17] F. Diaz. Integration of news content into web results. *ACM WSDM*, pp 182–191, 2009.

[18] S. Sushmita, H. Joho and M. Lalmas. A Task-Based Evaluation of an Aggregated Search Interface. *SPIRE*, pp 322–333, 2009.

[19] S. Sushmita, B. Piwowarski and M. Lalmas. Dynamics of Genre and Domain Intents. *AIRS*, 2010 (In Press).

[20] J. Arguello, F. Diaz, J. Callan and J. Crespo. Sources of evidence for vertical selection. *ACM SIGIR*, pp 315–322, 2009.

[21] J. Nielsen and K. Pernice. Eyetracking Web Usability. *New Riders Press*, 2009.

[22] Y. Kammerer and P. Gerjets. How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface. *Symposium on Eye-Tracking Research (Applications)*, pp 299–306, 2010.

[23] M.L Wilson, B. Kules, m.c. schraefel and B. Shneiderman. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science.*, 2(1):1–97, 2010.