Effect of Snippets on User Experience in Web Search

Mari-Carmen Marcos
Universitat Pompeu Fabra
Barcelona, Spain
mcarmen.marcos@upf.edu

Ferran Gavin
Softonic
Barcelona, Spain
ferran.gavin@softonic.com

Ioannis Arapakis
Yahoo Labs
Barcelona, Spain
arapakis@yahoo-inc.com

ABSTRACT

In recent years, the search engine results pages (SERP's) have been augmented with new markup elements that introduce seamlessly additional semantic information. Examples of such elements are the aggregated results disseminated by vertical portals, and the enriched snippets that display meta-information from the landing pages. In this paper, we investigate the gaze behaviour of web users who interact with SERP's that contain plain and rich snippets, and observe the impact of both types of snippets on the web search experience. For our study, we consider a wide range of snippet types, such as multimedia elements (Google Images, Google Videos), recommendation snippets (Author, Google Plus, Reviews, Google Shopping Product), and geo-location snippets (Google Places). We conduct two controlled user studies that employ eye tracking and mouse tracking, and analyse the search interactions of 213 participants, focusing on three factors: noticeability, interest, and conversion. Our findings indicate that ranking remains the most critical factor in relevance perception, although in certain cases the richness of snippets can capture user attention.

Categories and Subject Descriptors

H.5.m [**Information interfaces and presentation (e.g., HCI)**]: Miscellaneous

General Terms

Measurement, Design, Experimentation, Human Factors

Keywords

Search engines, rich snippets, eye tracking, user experience

1. INTRODUCTION

Information access is an area of research that, among other aspects, deals with the improvement of SERP's, both in terms of the presentation as well as the relevance of the retrieved results. In the last decades, companies with online presence developed a strong dependence on tracking metrics provided by search engines. As a result, many companies turn to specialists for optimising their web content for search engines and improving their ranking, a process known as Search Engine Optimisation (SEO). At the same time, search engines constantly upgrade their algorithms and interfaces, seeking to facilitate better services in terms of performance, content relevance, and user experience. One such example are the semantic markup elements (multimedia items) introduced in the SERP's, which show aggregated results from vertical portals. If we consider the case of Google, we can identify several other types of similar elements sharing the space with the typical results, composed by title, snippet (defined in [1] as "query-biased search result summaries"), and URL.

This work aims to understand better how t he enrichment of snippets with multimedia elements, recommendations, and geo-location information affects user behaviour during interactions with SERP's. More specifically, we investigate the following three dimensions: (i) noticeability, (ii) interest, and (iii) conversion. As noticeability we regard the ability of rich snippets to compete for user attention against other types of elements in a SERP. Interest is the overall engagement and involvement with a given resource, whereas conversion is how this attention translates to clickthroughs and, eventually, monetisation. Our experimental approach involves the use of eye tracking technology and click logging, for capturing the effects of rich snippets shown on SERP's. This is a high-value research area, yet poorly understood. We further believe that this work can impact future web search interfaces and interaction techniques for studying UX, as well as complement the common batch evaluations in information retrieval.

2. RELATED WORK

Until recently, the results of a SERP were presented using the same layout, and the relative ranking position of a result was the most determining factor that users considered [14]. This has been demonstrated by previous studies [8] that have altered the original rankings of the results in SERP's and observed that users would still assign more importance to top-ranked results. In the same work [8], the authors introduced additional information to the snippets and found that the performance for informational tasks improved significantly. The reason for that was that users clicked on results while being certain about the relevance of the corresponding landing pages, hence the importance of snippets.

Interacción 2015, September 07-09, Vilanova i la Geltrú, Spain.

A critical step in designing and building efficient search services is understanding how people interact with SERP's. In promoting such knowledge, past work has proposed techniques ranging from direct observation and user surveys to log analysis, and most recently eye-tracking studies. Eye tracking is a promising technique for the study of user behaviour in web search, since it can register accurately short-term changes in gaze activity that are not measurable by other means. Furthermore, It provides an account of users' unconscious behaviours and cognitive processing that are needed for interpreting their actions, as well as useful for mapping the user experience [6]. Some of the first works that studied user interactions with SERP's with the help of eye tracking hardware, were published in the last decade [12, 13]. Since then, there has been a surge of research [2, 4, 3, 16, 26, 9, 5, 27, 19, 24, 7, 18, 10] that has employed eye tracking and mouse tracking to analyse different aspects of the SERP's.

2.1 Aggregated Search

In addition to web search, the major commercial search portals offer access to specialised search services (e.g., news, local business, online products) or verticals of different information sources (e.g., images, videos, books). Aggregated search attempts to achieve diversity by fusing results from different verticals into one SERP, and complement the standard web results. Arguello and Capra [3] employed several aggregated verticals like images, news, shopping and video, and evaluated whether these verticals can influence user interactions with other components in the SERP. The results from a large-scale crowdsourcing study revealed that the level of influence may depend on the vertical, its surrogate representation, where it is displayed, and how it is distinguished from other components in the SERP. In [2, 4], the authors examined the aggregated search coherence, i.e., the extent to which results from different sources focus on similar senses of an ambiguous or underspecified query. Both studies provided evidence that users are more likely to interact with the web results when the vertical results are more consistent with the users' intended query-sense, e.g., a user searching for information about the planet "Saturn" is more likely to interact with the web results if the blended images contain pictures of the planet versus the car.

2.2 Gaze and Cursor Behaviour in Web Search

Several studies have investigated the gaze and click behaviour in aggregated search. Liu et al. [16] collected eye tracking data, click-through data, and users' feedback on their examinations of SERP's. The findings of this study reveal that a large proportion (45.8%) of the results fixated by users were not recognised as being "read". In addition, the authors of [16] found that before users actually read the results, there is often a skimming step during which they briefly scan the results without reading them, and perform judgments according to different signals.

Wang et al. [26] and Diaz et al. [9] found that different result appearances can create different biases on gaze behaviour for both vertical and other results on SERP's. A number of studies [5, 27] also showed that multimedia components in SERP's and result attractiveness may influence users' gaze and click-through behaviour. Furthermore, Navalpakkam et al. [19] conducted a controlled study where they varied the presence and relevance of a rich informational panel placed to the right of organic search results. The authors discovered that the information panels containing information relevant to the user's task can attract more attention and facilitate longer mouse cursor hovers.

Sushmita et al. [24] investigated the factors affecting users' click-through behaviour on aggregated search interfaces. Their study led to several findings. Foremost, it revealed that the position of search results was only significant for the blended interface. Secondly, participants' click-through behaviour on videos was different compared to other sources. Finally, capturing a task's orientation towards particular sources was identified as an important factor that warrants further investigation.

Chen et al. [7] studied click behaviour in aggregated SERP's and proposed a novel federated click model, which accounts for the fact that user attention is attracted by vertical results. The study provided evidence which indicates that the visual attention received by the vertical results can increase the chances that other nearby web results are also examined, and that click behaviour on vertical results can lead to more clues of search relevance due to their presentation style.

2.3 Rich Snippets & Social Annotation

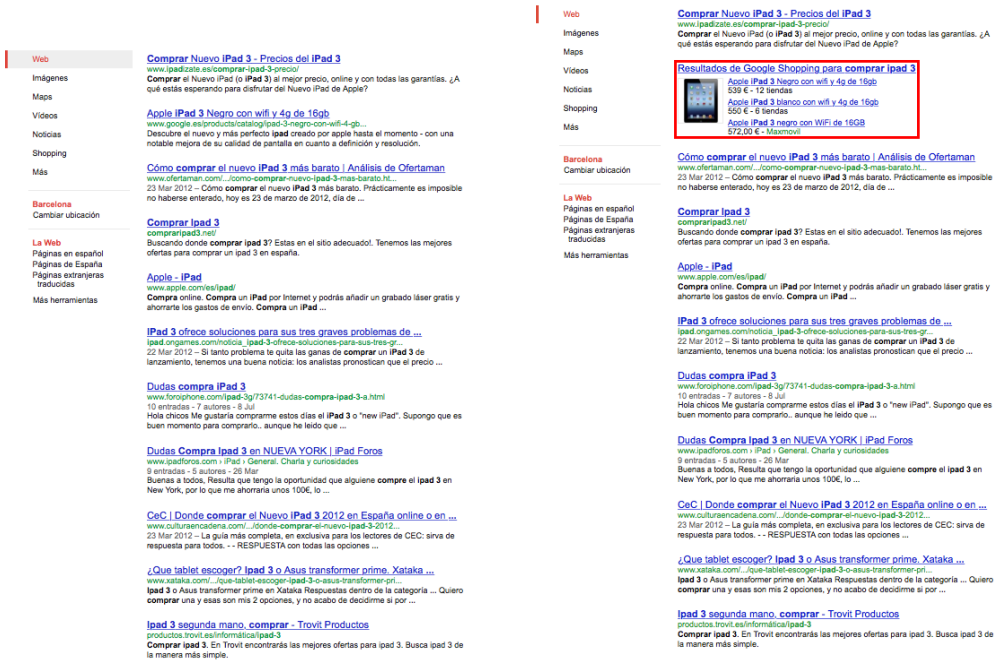
Several works have also examined the interaction between rich snippets and social annotations. A study by Muralidharan, Gyongyi, and Chi [18] demonstrated that placing a social annotation at the bottom of the snippet block reduces the probability of users fixating on it. The same study also revealed that positioning the social annotation at the top of the snippet block can mitigate this issue. The authors attribute this behaviour to the "inattentional blindness effect" [17], which leads people to notice mainly what they are actively looking for. In this study, the tasks were conducted using mock-ups and expert searchers, which places the reported effects in an artificial environment rather than a real-life setting.

Fernquist and Chi [10] investigated the impact of enriched elements with social annotation in SERP's. More specifically, the authors conducted an eye-tracking study using a retrospective think-aloud protocol, based on the design guidelines discussed in [18]. The authors blended the users' personal organic search results with regular, live search results, using a personalised ranking function. Their findings reveal that users employ annotations for local and shopping query types more often than fact-finding and news-related query types. Additionally, the authors discovered that users make use of annotations only when the task context and situation fit the need to look for social resources, such as subjective queries like shopping, products, and restaurant reviews.

Despite the above research efforts, there is still little understanding on what consitutes a well-designed, attractive, and engaging aggregated SERP. This paper provides some initial insights by examining the design of aggregated results and their effect on user gaze and click-through behaviour, in the context of web search. In what follows, we provide a description of our experimental method and discuss the main findings.

3. METHODOLOGY

To demonstrate the impact of enriched snippets on user search experience we carried out two controlled experiments, thereof referred to as Study 1 and Study 2. Both stud-



(a) SERP with plain snippets.

(b) SERP with rich snippet (#2).

Figure 1: SERP example showing the original plain results (a) and the results with the enriched snippet (b).

ies asked a number of participants to complete a series of search tasks using a commercial search engine, and shared a similar objective: investigate to what extent the richness of the snippets can affect user behaviour in web search. Study 2 followed Study 1, and aimed at validating the preliminary findings discovered in Study 1.

3.1 Study 1

Study 1 consisted of an offline and an online part. The offline part was conducted in a laboratory setting, where the experimental procedure was followed and eye tracking data were collected. The online part repeated the same experimental design in an online setting, and collected mouse tracking data.

3.1.1 Design

The study used a mixed design with three independent variables: snippet richness (with two levels: "plain", "rich"), snippet position (with two levels: "top-ranked", "bottom-ranked"), and snippet type (with five levels: "Author", "Google Plus", "Google Places", "Multimedia", "Review"). To control for snippet richness, we prepared two versions of every SERP shown in the study: one version containing a rich snippet among the organic results (at top-ranked position #2 or #3, or bottom-ranked position #6 or #7) and another version without the rich snippet, showing only plain web results. More specifically, for each task introduced in the study (Section 3.1.3), predefined search queries were submitted to Google Search until obtaining SERP's that contained at least one of the types of rich snippets shown in Table 1, in the 2nd, 3rd position, 6th, or 7th position. While retaining the original look and feel of the retrieved SERP's, we edited the HTML code of the pages so that they either contained only one rich snippet or no rich snippets; any additional, unwanted rich snippets were modified to appear as plain snippets instead.

Given that the retrieved SERP's showed the original top ten results obtained from Google search, we considered them as topically relevant to the associated search query. This assumption was further validated by a manual examination. We also controlled that none of the snippet text (plain or rich) contained directly the answer to the query. The reason for that is because we wanted to study click behaviour (conversion), therefore displaying the answer in a snippet would bias to click on it, or avoid it if its not the correct one

Our dependent variables were noticeability, interestingness, and conversion, all aspects of the search experience that can be potentially affected by the richness of the snippets. To study the effects of our experimental manipulation on our dependent variables, we obtained several metrics of gaze behaviour and a metric of search task performance [6], as shown bellow:

- Noticeability: Time to First Fixation, Fixations Before
- Interest: Total Fixation Duration, Fixation Count, Visit Duration
- Conversion: Click count

3.1.2 Apparatus

The relationship between attention and eye movements has been investigated extensively in the past [11, 20, 21, 22]. When we read, examine a scene, or search for an object, we continuously make eye movements called saccades. Saccades are rapid movements that occur when we change focus, and can reach velocities as high as 500° per second. When the visual gaze is maintained on a single location for several milliseconds we have a fixation. The importance of gaze in the assessment of attention focus lies in the fact

Table 1: Types of rich snippets and search tasks used in Study 1

Rich snippet	Search task
Google Places	1. Find a hotel near Sants Station 2. Find a destination management company in Barcelona
Google Plus	3. Find the Fitur events program 4. Find the definition of "revenue management"
Author	5. Find statistics of the use of Facebook in companies of the touristic sector 6. Find what actions is taking the Ushuaia Hotel in Facebook
Multimedia	7. Find in what Facebook Places does consist on 8. Find a webpage with reviews of the Ushuaia Hotel
Review	9. Find a webpage with reviews of the Pachá Hotel 10. Find a good blogs directory about tourism and travel

Table 2: Eye metrics used to analyse gaze behaviour

1. **Time to First Fixation:** Time taken (in seconds) before a participant fixates on an AOI for the first time.

2. **Fixations Before:** Number of times a participant fixates on the media before fixating on an AOI for the first time.

3. **Total Fixation Duration:** Sum of the duration for all fixations within an AOI.

4. **Fixation Count:** Number of times a participant fixates on an AOI.

5. **Visit Duration:** Duration of each individual visit within an AOI.

that, although looking might appear to be a process that is under voluntary control, conscious and deliberate control of fixation happens infrequently. As with other components of voluntary performance, looking is controlled by a general intention, and consciousness plays a minor role in the execution of the intended sequence of fixations [15].

To analyse gaze behaviour, we used a Tobii 1750 eye tracker, integrated into a 17" TFT monitor with a 1280×1024 resolution. When activated, the eye tracker illuminates the user with two infrared projections that generate reflection patterns on the corneas of the eyes. A video camera gathers these reflection patterns along with the position of the user and, through digital image processing, the pupil locations are extracted at a rate of 50 Hz. The pupil positions are then mapped to gaze locations on the screen. For the gaze behaviour analysis we used the eye metrics listed in Table 2, which were extracted automatically using the Tobii Studio Statistics application. The metrics were calculated based on defined Areas of Interest (AOIs) and data selection time intervals. We defined as our AOI(s) the results shown in the SERP. In addition to using eye tracking in the offline part of Study 1, we also used the software CrazyEgg[1] to log participants' mouse tracking data for the online part of the study.

3.1.3 Search Tasks

A total of ten tasks (shown in Table 1) were performed by every participant, two for each type of rich snippet. In the context of these search tasks, the participants assumed the role of an expert in social media for the tourism sector and were invited to participate in a panel for a particular city. Examples of the tasks involved finding a hotel near the venue, retrieving the conference program, and other. For each search task, two SERP's were available: one that contained a rich snippet at a top-ranked (#2 or #3) or a low-ranked position (#6 or #7), and one showing only plain snippets.

3.1.4 Participants

A total of 60 participants (female=38, male=22), between the ages of 18 to 58 and and free from any obvious physical or sensory impairment, were recruited through a campus-wide ad. All participants were frequent users of web search engines. Besides the 60 participants who were involved in the offline study that collected eye tracking data, another 110 participants were involved in the online study that collected mouse tracking data. The participants were mainly locals (i.e., Catalan, Spanish) and had graduated from, were currently studying, or working at the at Pompeu Fabra University, Barcelona. Finally, participants were all proficient with the English and Spanish languages.

3.1.5 Procedure

For every search task, participants were presented with an initial search query and the retrieved SERP for this search query. They were instructed to examine the SERP as they would normally do, although they were not able to issue additional search queries. As long as the answer did not appear in the snippet description of the retrieved results, participants were encouraged to click on any of the result links and examine the landing pages, prior to continuing to the next task. Two tracks were designed in order to show the SERP's in an alternating manner (Figure 2). Therefore, for every type or rich snippet participants were shown a SERP with a rich snippet (at a top- or low-ranked position) and a SERP containing only plain snippets. Selecting a particular result as the one the participants felt would answer the search task question, would conclude the task. To enrich the sample for the conversion measure, 110 participants performed the same search tasks online, while we were logging their mouse tracking data.

3.1.6 Results

To choose an appropriate statistical test, we first examine the distribution of our data using the Anderson-Darling and Cramer-von Mises tests. These tests are known to perform better compared to the Kolomorov-Smirnov test [23, 25], although in large samples they tend to be significant even for scores that are marginally different from a normal distribution; we thus interpret them in conjunction Q-Q plots, while also accounting for the skew and kurtosis values. Since in all cases we observe a non-normal distribution in the abso-

[1] `http://www.crazyegg.com/`

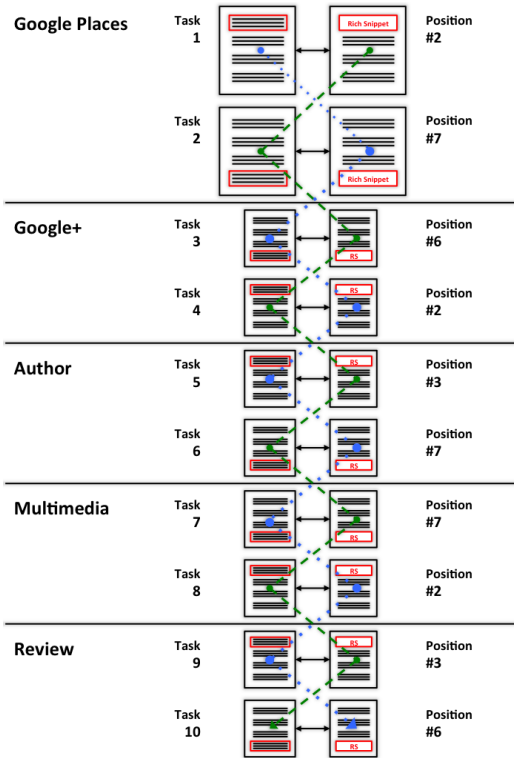
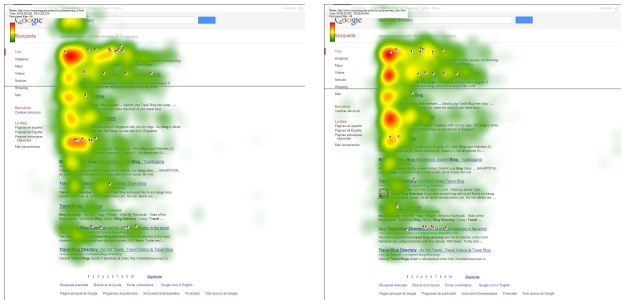


Figure 2: Experimental protocol.

lute differences of the estimate errors, we opt for the Mann-Whitney test and report our results at an α level of .05. Finally, to take an appropriate control of Type I errors in multiple pair-wise comparisons we apply the Bonferroni correction.

When comparing top-ranked vs. bottom-ranked snippets, we identify several statistically significant differences across our eye metrics, and for both rich and plain snippets, and different snippet types (Author, Google Plus, Google Places, Multimedia, Review). More specifically, for the metric Time to First Fixation, the Mann-Whitney test reveals that top-ranked plain snippets ($Mdn = 2.88$) received their first fixation significantly faster than bottom-ranked plain snippets ($Mdn = 15.71$), $U = 1289.50, z = -10.03, p = .000, r = -.48$. Similarly, the top-ranked rich snippets ($Mdn = 3.25$) received their first fixation significantly faster than bottom-ranked plain snippets ($Mdn = 15.30$), $U = 1135.50, z = -9.71, p = .000, r = -.46$. For the metric Fixations Before, the Mann-Whitney test reveals that top-ranked plain snippets ($Mdn = 9.00$) had significantly fewer fixations on other items before being noticed than bottom-ranked plain snippets ($Mdn = 51.50$), $U = 1405.50, z = -9.79, p = .000, r = -.46$. The same highly significant effect is observed when comparing top-ranked rich snippets ($Mdn = 9.00$) against bottom-ranked rich snippets ($Mdn = 46.00$), $U = 1112.50, z = -9.77, p = .000, r = -.46$. Additionally, when examining the metric Total Fixation Duration, the Mann-Whitney test reveals that the top-ranked plain snippets ($Mdn = 2.08$) received longer fixations than bottom-ranked plain snippets ($Mdn = 0.81$), $U = 3644.00, z = -5.08, p = .000, r = -.24$. Likewise, participants fixated for longer times on the top-ranked rich snippets ($Mdn = 2.08$) than bottom-ranked rich snippets ($Mdn = .90$), $U = 3364.50, z = -4.65, p = .000, r = -.22$. For the Fixation Count metric, the Mann-Whitney test reveals that the top-ranked plain snippets ($Mdn = 10.00$) received significantly more fixations than bottom-ranked plain snippets ($Mdn = 4.00$), $U = 3422.50, z = -5.56, p = .000, r = -.26$. The same effect, is observed when comparing top-ranked rich snippets ($Mdn = 10.00$) against bottom-ranked rich snippets ($Mdn = 5.00$), although it has a smaller effect size, $U = 3550.00, z = -4.23, p = .000, r = -.20$. Finally, for the metric Visit Duration, participants spent significantly more time examining the top-ranked plain snippets ($Mdn = .81$) than the bottom-ranked plain snippets ($Mdn = .59$), $U = 4975.50, z = -2.28, p = .022, r = -.10$. Likewise, participants spent significantly more time examining the top-ranked rich snippets ($Mdn = .74$) against bottom-ranked rich snippets ($Mdn = .53$), $U = 4035.00, z = -3.12, p = .002, r = -.14$. These findings provide further evidence that the results shown in top-ranked positions attract significantly more attention, and for longer time, compared to the results shown in bottom-ranked positions. But does this also hold for plain vs. rich snippets that are displayed in the same position in a SERP? Does the richness or type of snippet introduce any effect?



(a) Only plain snippets. (b) With rich snippet (#7).

Figure 3: Heatmaps showing allocation of user attention in web search, on two SERP's.

To answer the above questions, we perform a comparison between rich and plain snippets that have the same relative position in the layout of the SERP, across all eye metrics. We observe that rich snippets generally tend to attract more attention. For example, rich snippets receive their first fixation faster, have fewer fixations on other elements prior to being noticed, and receive more and longer fixations than plain snippets. However, none of these differences appear to be statistical significant, and even less when comparing snippets in top-ranked positions (Figure 3). Most likely, when a snippet appears in a top-ranked position, this by itself introduces a bias which makes users consider it as topically relevant [1]; thus the absence of significant effects in top-ranked snippets.

Next, we repeat our analysis for each snippet type separately. As before, we observe differences between the rich and plain snippets that span across the different types of snippets, like Google Places, Author, Multimedia, etc., although most of them appear not to be statistically significant. As side-findings, we report that participants fixated faster on the top-ranked rich Multimedia snippets ($Mdn = 1.87$) than the top-ranked plain Multimedia snippets ($Mdn = 2.49$), $U = 218.00, z = -2.36, p = .018, r = -.24$. Moreover, the bottom-ranked rich Multimedia snippets ($Mdn =$

Figure 4: Visual representation of the four rich snippet types used in Study 2.

.45) were observed for less time than the bottom-ranked plain Multimedia snippets ($Mdn = .71$), $U = 104.50, z = -1.99, p = .045, r = -.20$.

When analysing the click-through data of the 110 participants who performed the online part of Study 1, we observe that the top-ranked snippets of type Review, Author, and Google Plus (all recommendation snippets), receive the highest percentage of clicks: 32%, 21%, and 18% respectively. In 4 out of 5 rich snippet typologies, the number of clicks is similar; a statistically significant difference is only found for the Google Plus recommendation snippet ($p = .04$). For this type of rich snippet, the recommended result received eight clicks, while the plain result received only two. However, we cannot conclude that Google Plus recommendations per se had a strong influence on the participants' click behaviour. We need to account for the context as well as the person that recommends the result, as discussed in [18]. In this case, the picture we used as a recommender is a public and acknowledged person. We speculate that the results could have been different if the recommender was someone less popular or unknown to the wide public. Finally, the rich snippet with the least clicks is the Multimedia snippet, most likely because people seeking information do not take so much into consideration multimedia results. Another possible interpretation is that any deviation from the traditional, textual presentation of he results may lead to "banner blindness", and consequently result in less attention to the enriched snippet. No significant differences in performance are found when we compare enriched results with plain snippets.

3.2 Study 2

The second study is a follow-up to Study 1. It was performed to investigate users' behaviour while interacting with a different set of snippets (Figure 4) and also allowed us to replicate the preliminary findings of the first study. More specifically, we introduce the Google Shopping snippet and remove the Google Plus (given recent findings reported in [18]) and Multimedia snippet types. The reason for that was the increasing popularity and visibility of Google's Shopping Product snippet. In addition, we now compare only top-ranked rich vs. plain snippets, and exclude from our analysis bottom-ranked results, in the light of the findings provided by Study 1.

Table 3: Types of rich snippets and search tasks used in Study 2

Rich snippet	Search task
Author	1. iPad 3 Vs. iPad 2 comparison
Google Shopping	2. Buy iPad 3
Google Places	3. Apple store Madrid
Review	4. Best moment to sell an iPhone

3.2.1 Design

The study used an independent measures design with two independent variables: snippet richness (with two levels: "plain", "rich"), and snippet type (with five levels: "Author", "Google Places", "Google Shopping", "Review"). To control for snippet richness, we prepared two versions of every SERP shown in the study: one version containing a rich snippet among the organic results (at position #2) and another version without the rich snippet, showing only plain web results. More specifically, for each task introduced in the study (see Section 3.2.3), predefined search queries were submitted to Google Search until obtaining SERP's that contained at least one of the types of rich snippets shown in Table 3 in the 2nd or 3rd position. While retaining the original look and feel of the retrieved SERP's, we edited the HTML code of the pages so that they either contained only one rich snippet or no rich snippets; any additional, unwanted rich snippets were modified to appear as plain snippets instead.

Given that the retrieved SERP's showed the original top ten results obtained from Google search, we considered them as topically relevant to the associated search query. This assumption was further validated by a manual examination. Similarly to the previous study, we controlled that none of the snippet text (plain or rich) contained directly the answer to the query. Our dependent variables were again noticeability, interestingness, and conversion, all aspects of the search experience that can be potentially affected by the richness of the snippets. To study the effect of our experimental manipulation on the dependent variable, we obtained the same five metrics of gaze behaviour and one metric of search task performance discussed in Section 3.1.1.

3.2.2 Apparatus

The study used the setup discussed in Section 3.1.2.

3.2.3 Search Tasks

Four search tasks were used in the study, as shown in Table 3. Each search task asked the participants to retrieve relevant information to answer a question, e.g., find information on how to arrive to the AppStore located in Madrid. For each search task, participants were presented with an initial search query and the retrieved SERP for this search query. Participants were instructed to examine the SERP as they would normally do, although they were not able to issue additional search queries. As long as the answer did not appear in the snippet description, the participants were encouraged to click on any of the results and examine the web pages prior continuing to the next task. Selecting a particular result as the one that answered the search task

question would conclude the task. Half of the participants were shown SERP's that contained only plain snippets and the other half SERP's that included a rich snippet at a top-ranked position.

3.2.4 Participants

A total of 43 participants (female=26, male=17), between the ages of 18 to 45 and and free from any obvious physical or sensory impairment, were recruited through a campus-wide ad. All participants were frequent users of web search engines. The participants were mainly locals (i.e., Catalan, Spanish) and had graduated from, were currently studying, or working at the at Pompeu Fabra University, Barcelona. Finally, participants were all proficient with the English and Spanish languages.

3.2.5 Procedure

Two groups were created using an independent measures design. Each group performed four search tasks. The search tasks were part of a larger scenario that instructed the participants as follows: *"You want to buy the new iPad 3, but before you do so, you need to know how much it costs, what are the technical pros and cons in comparison with iPad 2, how to arrive to the AppStore in a particular city, and the best moment to sell your iPad 2"*. Four SERP's were selected and were enriched with four types of rich snippets: Product, Author, Places, and Reviews. For this study, the rich snippets and its respective plain snippet were always placed at a top position (#2).

For each group, half of users would see the SERP's with the plain results and the other half would see the SERP's that contained the rich snippet. As in Study 1, the participants were asked to click on the result that they felt would answer the question presented in the search task. After performing the four tasks, the users in the group that saw the enriched SERP's were shown the rich snippet highlighted and were asked to answer the following questions: (i) Did you notice the highlighted result?, (ii) If you noticed it, did you click on it?, (iii) What was the reason behind (not) clicking on it?, and (iv) Did the image (pictures, maps, stars) influence your click decision?

3.2.6 Results

To analyse the user behaviour (noticeability, interestingness, and conversion,) we use the Mann-Whitney test on all eye metrics. Our analysis does not reveal any statistically significant differences between plain and rich snippets, for any of the SERP's and snippet types (Author, Google Places, Google Shopping, Review). In addition, we apply the Chi-Square test of Association on the click data obtained by the eye tracking software. A significant association between the snippet richness (plain, rich) and the number of participants who clicked on the enriched results is observed for the task that involved the Google Places snippet, ($\chi^2(1) = 4.00, p = .045$). The reason could be that the map and address shown in the rich snippet were perceived as helpful and topically relevant, and the participants felt inclined to select it.

Furthermore, we apply the Chi-Square Goodness-Of-Fit test on the questionnaire data, and more specifically to questions (i), (ii), (iii), and (iv) presented in Section 3.2.5. The test reveals a consistent effect across all snippet types. More specifically, the Chi-Square Goodness-Of-Fit test indicates that significantly more participants reported having noticed the rich snippet of type Author ($\chi^2(1) = 6.00, p = .014$), Google Shopping ($\chi^2(1) = 8.16, p = .004$), Google Places ($\chi^2(1) = 8.16, p = .004$), and Review ($\chi^2(1) = 6.00, p = .014$). Finally, significantly more participants reported that the presence of the pictures, images, or stars in rich snippets did not influence their decision of clicking the result, that those who reported that it did. Again, this effect is consistent across all snippet types, i.e., Author ($\chi^2(2) = 15.75, p = .000$), Google Shopping ($\chi^2(2) = 19.00, p = .000$), Google Places ($\chi^2(2) = 13.00, p = .002$), and Review ($\chi^2(2) = 7.00, p = .030$).

4. DISCUSSION & CONCLUSIONS

In this paper we presented a two-part study of the gaze and mouse behaviour of web users, while interacting with SERP's. Considering our User Experience (UX) context, our goal was to investigate whether the richness of snippets can affect user behaviour and introduce a bias to the subjective perception of relevance, i.e., which results people perceive as more useful or relevant in a SERP.

Foremost, the findings of our first study indicate that the relative position of a result in a SERP remains the most influential factor of click behaviour, although snippet richness appears to become a more important variable, especially when examining bottom-ranked results. Our analysis did not indicate an effect of snippet richness on the gaze behaviour for top-ranked positions: none of the eye metrics we examined was found to be statistically significantly different for rich snippets, followed by similar findings in regards to the recorded clicks. Nevertheless, snippet richness is a factor that must be taken into account when considering bottom-ranked results, since they are noticed much earlier than the plain snippets and for longer periods of time. Furthermore, Multimedia snippets were the most noticeable element across the evaluated types of snippets in this study. Also, we demonstrated that the social content in a snippet has a growing importance, as indicated by the higher click count it achieved, compared to the equivalent plain results of the same ranking.

The second study compared rich and plain snippets displayed at top positions. Our analysis did not provide any evidence that rich snippets attract attention faster or for longer periods of time than their corresponding plain snippets. Additionally, the number of clicks did not differ between the rich and plain snippets, apart from the Google Places snippet. The reported answers in the questionnaire suggested that the participants were aware of the presence of these enriched results, but their decision of clicking was not conditioned by their presence in the SERP.

In conclusion, our study provides further evidence which confirms the importance of ranking in relevance judgments, and also indicates that snippet richness is not as influential as one would originally anticipate. However, this may change in the future, but more work is needed to understand how rich snippets should be presented, and in which position in a SERP they are most effective. We consider that this work can impact future web search interfaces and interaction techniques for studying UX, as a complement to the common batch evaluation that dominates the IR field. Eye-tracking is a suitable method for such research, as it accurately captures a user's initial (low-level) attention, much more accurately than other methods, such as clicks logs for

example.

Finally, we acknowledge that the results discussed in this paper are preliminary, and further testing is warranted with additional snippet types and search tasks. We leave for future work search tasks where the scenario is ambiguous: will rich snippets play a more important role when the user cannot provide a definitive search phrase? Nevertheless, we feel that we have raised a main issue for a topic that is largely unexplored and it is of importance to research conducted in human-computer interaction (HCI), information retrieval (IR), as well as SEO practitioners.

5. REFERENCES

[1] M. Ageev, D. Lagun, and E. Agichtein. Towards task-based snippet evaluation: preliminary results and challenges. In C. L. A. Clarke, L. Freund, M. D. Smucker, and E. Yilmaz, editors, *Proceedings of the SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation (MUBE 2013)*, pages 1–2, 2013.

[2] J. Arguello and R. Capra. The effect of aggregated search coherence on search behavior. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1293–1302, New York, NY, USA, 2012. ACM.

[3] J. Arguello and R. Capra. The effects of vertical rank and border on aggregated search coherence and search behavior. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 539–548, New York, NY, USA, 2014. ACM.

[4] J. Arguello, R. Capra, and W.-C. Wu. Factors affecting aggregated search coherence and search behavior. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, CIKM '13, pages 1989–1998, New York, NY, USA, 2013. ACM.

[5] J. Bar-Ilan, K. Keenoy, M. Levene, and E. Yaari. Presentation bias is significant in determining user preference for search results—a user study. *Journal of the American Society for Information Science and Technology*, 60(1):135–149, 2009.

[6] A. Bojko. *Eye Tracking the User Experience: A Practical Guide to Research*. Rosenfeld Media, 2013.

[7] D. Chen, W. Chen, H. Wang, Z. Chen, and Q. Yang. Beyond ten blue links: Enabling user click modeling in federated web search. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 463–472, New York, NY, USA, 2012. ACM.

[8] E. Cutrell and Z. Guan. What are you looking for?: An eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 407–416, New York, NY, USA, 2007. ACM.

[9] F. Diaz, R. White, G. Buscher, and D. Liebling. Robust models of mouse movement on dynamic web search results pages. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13, pages 1451–1460, New York, NY, USA, 2013. ACM.

[10] J. Fernquist and E. H. Chi. Perception and understanding of social annotations in web search. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 403–412, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[11] M. H. Fischer. An investigation of attention allocation during sequential eye movement tasks. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 52(3), 1999.

[12] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Trans. Inf. Syst.*, 23(2):147–168, Apr. 2005.

[13] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 478–479, New York, NY, USA, 2004. ACM.

[14] G. Hotchkiss. Eye tracking on universal and personalized search, September 2007.

[15] D. Kahneman. *Attention and Effort (Experimental Psychology)*. Prentice Hall, 1973.

[16] Y. Liu, C. Wang, K. Zhou, J. Nie, M. Zhang, and S. Ma. From skimming to reading: A two-stage examination model for web search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 849–858, New York, NY, USA, 2014. ACM.

[17] A. Mack and I. Rock. *Inattentional blindness*. The MIT Press, Cambridge, MA., 1998.

[18] A. Muralidharan, Z. Gyongyi, and E. Chi. Social annotations in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1085–1094, New York, NY, USA, 2012. ACM.

[19] V. Navalpakkam, L. Jentzsch, R. Sayres, S. Ravi, A. Ahmed, and A. Smola. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 953–964, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[20] K. Rayner. Eye movements and visual cognition: Scene perception and reading. In K. Rayner, editor, *Springer series in neuropsychology*, pages 46–65. Springer-Verlag, 1992.

[21] R. W. Remington. Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 6(4):726–744, 1980.

[22] M. Shepherd, J. M. Findlay, and R. J. Hockey. The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, 38(3):475–491, August 1986.

[23] M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

[24] S. Sushmita, H. Joho, M. Lalmas, and R. Villa. Factors affecting click-through behavior in aggregated search interfaces. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 519–528, New York, NY, USA, 2010. ACM.

[25] H. C. Thode. *Testing for normality*. CRC Press, 2002.

[26] C. Wang, Y. Liu, M. Zhang, S. Ma, M. Zheng, J. Qian, and K. Zhang. Incorporating vertical results into search click models. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 503–512, New York, NY, USA, 2013. ACM.

[27] Y. Yue, R. Patel, and H. Roehrig. Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 1011–1018, New York, NY, USA, 2010. ACM.