

The importance of sources in the selection of online newspaper articles: a study of Google Noticias using eye-tracking

Cristòfol Rovira
Jofre Capdevila
Mari-Carmen Marcos *

*Article received on:
March 19, 2013.*

*Article accepted on:
November 19, 2013.*

ABSTRACT

This paper presents the results of a study of the news search engine *Google Noticias* (Spanish news aggregator) to determine the degree of importance users lend to titles, sources and summaries as these appear in the lists of search results. By use of eye-tracking, researchers tallied the duration of the gaze of fifty users viewing search results displays to determine which of these elements captured their attention the longest. The results indicate that more attention is paid to sources than to the summaries; while there are no significant differences between time spent viewing sources versus the title, indicating that online media users also consider the sources when deciding which articles to pursue.

* The three authors belong to the Universidad Pompeu Fabra, Barcelona, Spain. (cristofol.rovira@upf.edu); (jofre.capdevila@gmail.com); (mccarmen.marcos@upf.edu)

Keywords: Search engines; News search engines; Search engines result lists; SERP, news services; User behavior; Eye-tracking, information sources.

RESUMEN

La importancia de las fuentes en la selección de artículos de prensa en línea: un estudio de Google Noticias mediante seguimiento ocular (eye tracking)

Cristòfol Rovira, Jofre Capdevila and Mari Carmen Marcos

Se presentan los resultados de un estudio realizado en el buscador Google Noticias para determinar la importancia que le dan los usuarios a los elementos principales de una noticia que aparece en un listado de resultados: titular, fuente y resumen. Mediante la técnica de seguimiento ocular (*eye-tracking*) se analiza la duración de la mirada de 50 usuarios frente a las páginas de resultados y se determina qué elementos captan más tiempo su atención. Los resultados del estudio indican que se presta más atención a las fuentes que al resumen pero que no hay diferencias significativas entre el tiempo dedicado a la fuente con respecto al título, lo cual indica que en los medios en línea los usuarios también consideran a la fuente como elemento para decidir qué artículos leer.

Palabras clave: Buscadores; Buscadores de noticias; Listados de resultados; Servicios de noticias; Comportamiento de los usuarios; Seguimiento ocular (*eye-tracking*); Fuentes de información.

INTRODUCTION

Online newspapers have been the object of many studies often focused on comparative analysis (Díaz-Noci & Meso Ayerdi, 1998; Schultz, 1999) or the assessment of a concrete aspect, such as multimedia contents (Díaz-Noci, 2009; Guallar, Rovira and Ruiz, 2010), social networks (Rodríguez-Martínez, Codina and Pedraza-Jiménez, 2010), information retrieval (Martínez Rubio, 2008; Guallar & Abadal, 2009; Rubio Lacoba, 2010), or blog contents (Sánchez Vigil et al., 2010).

Studies about newspaper reading using specialized search engines like the one evaluated here, *Google Noticias* (Spanish news aggregator), are less common. This is a news search service that collects articles from numerous newspaper sources, so the user may find information of the same subject from different media.

Newspaper search services, like *Google Noticias*, have three blocks of information for each item listed as a result: news title, news source and a summary (*Illustration 1*). Based on these three blocks of information, the user has to decide if he/she will read the full news. The fact that we read diagonally adds to this relatively little information. Nicholas Carr (2011), a Pulitzer Prize finalist, indicates that the Internet promotes a quick and distracted pecking of small fragments of information from several sources at a great speed, leaving no room for reflection or critical thinking.

In such context, deciding what news to read ends up being a quick process guided by a personal interest in the subject and the information provided by the search engine. In this study we question if readers consider the news source as important data to decide if they continue reading. In the paper format, the accuracy or point of view of each source is clearer, however, in the Web, especially in a news search engine, what role does the source play in the decisions of the readers?

This study begins with the hypothesis that users who employ new technologies to be informed about current topics tend to be less rigorous in the evaluation of information. We believe that the overload of information conditions users to select the news by an attractive title. So, our first consideration is that when a user reviews a list of results of a news search engine, most of his/her gaze will concentrate in the titles, therefore, much more time will be devoted to looking at the titles than the sources.

For this study, users were tested in the laboratory using an eye-tracker. This technology has been used in previous studies to establish the visual behavior of people who review web pages of search engine results (Granka, Joachims and Gay, 2004; Rele & Duchowski, 2005; González-Caro & Marcos, 2011; Nielsen & Pernice, 2009). However, no research has been published yet about web pages of news results.

The article structure is as follows: the section “Previous Works” includes studies which have been useful as a starting point; the section “*Google Noticias*” details the experimental design carried out; next, results of the test analysis and finally, the conclusions of this study.

PREVIOUS WORKS

Eye-tracking is a technique that establishes the position of the eyes and their movements while a specific action is taking place. It uses an infrared wave technology capable of detecting the pupils. This technique has been mainly employed for research on the visual system, cognitive psychology and product design. The marketing field has used it to improve product placing in grocery store shelves, choose advertising spaces in web pages, increase conversion ratios of web sales, or for benchmarking (to compare users' actions in their website with those of the competition). In recent years, this technique has been quite common in the evaluation of websites especially in usability studies of digital interfaces, where most studies have been motivated by commercial interests. The book of Jakob Nielsen & Kara Pernice (2009) gives numerous examples of eye-tracking in usability studies. Chapters include studies that laid the foundations to evaluate navigation menus, links, images and advertising in web design.

Search interfaces and information retrieval interfaces are fields of study that have also received a lot of attention, mainly interfaces for Internet Search Engine Results, also known as Search Engine Results Pages (SERPs). In this area, eye-tracking technology allowed significant findings, such as evidence from heat maps of the golden triangle. The golden triangle (Enquiro, 2003) is the zone which attracts more gazes from most users and is situated in the top-left corner of a list of results.

Granka, Joachims and Gay (2004) followed these studies and their tests established that although the number of clicks is very high for the first result, users spend the same time analyzing the first and second results, and they look up to position 5 or 6 if they face more complicated questions or tasks. Later, Guan & Cutrell (2007) observed that users click much more frequently the two first results, even though the answer to their search is not among them. A study of Cornell University (Pan et al., 2007) changed the order of results and concluded that users have "blind faith" in search engine classification, regardless of the relevance of these results.

Other studies consider different zones of results pages (organic search results and advertising, or even in more detail: title, snippet and URL of each result). Rele and Duchowski (2005) reveal that the snippet (summary) receives a higher percentage of fixations than the titles, while the URL receives the least. These conclusions are reinforced by more recent studies, like those of Marcos & González-Caro (2010) and González-Caro & Marcos (2011), who demonstrate that the element most looked in the results is the summary (44% of the time), followed by the title (39%) and the URL (17%).

In the specific area of digital newspapers, the eye-tracking studies with the most citations are those of the Poynter Institute (*Eyetrack I* in 1990-91, *Eyetrack II* in 1999-2000, *Eyetrack III* in 2003-2004 and *Eyetrack IV* in 2006-2007). Along with the results, heat maps of newspaper web pages provide interesting suggestions about when and how to incorporate advertising, multi-media elements and short texts in the covers of digital newspapers.

The study proposed here does not analyze digital newspaper web pages or general search engines. It questions the influence of the information source at the time of choosing in the news aggregator *Google Noticias*. There have been no published works yet using eye-tracking technology to study this platform.

GOOGLE NOTICIAS

Google Noticias is the Spanish market edition of Google News, which is a news aggregator and search engine that constantly and automatically tracks information from the main online communication media, considering the user's geographical location. This service was launched by Google, Inc. in 2002, however, it remained as a test version until 2006. In June 2012, 72 regional editions for concrete geographical sectors were accessible, including Spain. The aggregator is available in 28 languages, works by an automated aggregation algorithm that tracks more than 25,000 digital media every 15 minutes to feed the database, and releases the news of the last 30 days in its server. This study will evaluate the results page interface, which we describe below.

Each item of the result list of *Google Noticias* has up to nine elements of information (data from January 2012). The four basic elements that always appear in every result providing the main information of the news (*Illustration 1*) are:

1. Title. The main headline of the news, an active link to get to the original article.
2. Main source. The name of the source or communication media that published the news.
3. Date. The publication date of the news.
4. Summary. It captures the first 200 characters of the original article.

Three other elements appear only when Google finds news very similar to the main one in another media.

5. Secondary titles. Clickable titles of a news selection similar to the main headline.
6. Secondary sources. Sources or information media where the news listed as secondary were published
7. More information. Access to the complete list of news similar to the main one.

Finally, an image related with the news often appears, although not in every case.

8. Image linked with the news.
9. The source from which the image was extracted.

So, elements of the three groups is the most commonly observed.

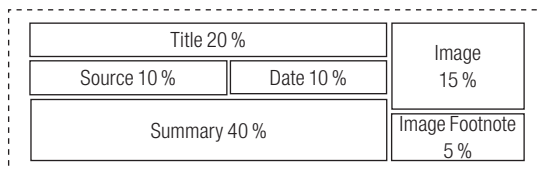


Illustration 1. Areas of an item of the result list of Google News.

METHODOLOGY

Participants

User selection for the test is a complex process when trying to fulfill the requirements to generalize the results of the sample to a particular population. Normally, usability studies using eye-tracking are qualitative and do not look for a generalization, but a detection of anomalies to improve in a web page design. Thus, user samples are usually small, between 10 and 15 participants. On the contrary, this investigation has a quantitative orientation, since we wish to obtain solid evidence on the degree of influence of the sources at the moment of choosing what item to click, on the result page of *Google Noticias*. Therefore, sample selection has been more demanding, despite the limitations inherent to every recruitment process, especially with tight budgets.

Given test characteristics, we were interested in users or potential users of *Google Noticias*, so the profile for the test was regular Internet users and legal adults.

Generally, the number of users for a quantitative study should be higher than 30 participants (Marcos, Nettleton and Sáez, 2012). According to the timing, means and resources to develop the project, 60 users would participate in the study. Afterwards, users in whom the eye-tracker recorded less than 80% of their fixations were eliminated, as well as two users with incoherent data or data completely away from the mean. The definite sample included 50 subjects distributed as follows:

Table 1. Distribution of users for the test.

Age	Men	%	Women	%	Total	%
18 - 20 years	5	10 %	14	28 %	19	38 %
21 - 25 years	0	0 %	7	7 %	7	14 %
26 - 30 years	4	8 %	2	4 %	6	12 %
31 - 35 years	3	6 %	3	6 %	6	12 %
36 - 40 years	3	6 %	2	4 %	5	10 %
41 - 50 years	3	6 %	2	4 %	5	10 %
+50 years	1	2 %	1	2 %	2	4 %
Total	19	38 %	31	62 %	50	100 %

Eye-tracking device

Tests were performed with an eye-tracker device model Tobii 1750. This is a monitor that includes an infrared light system, which captures eye movement. This technology detects with great precision the gaze position on the screen, as well as the length of time the gaze remained on a specific point of the screen. These objective data require a cognitive interpretation (Hassan & Herrero, 2007). Traditionally, this interpretation has assumed that when a person fixes his/her gaze on an element, he/she is thinking of such element. Despite the fact that this relationship is not always true, it is sufficiently consistent to obtain solid results, especially if the number of users who participate in the study is relatively high.

Experimental design

The aim of this study is to understand the user's visual behavior facing an information search. Therefore, the test will focus on the visualization of the result lists of this service.

To achieve a systematic and comparable analysis among users, the test was carried out under the same conditions, trying to standardize the initial search as much as possible to get similar results. This is the reason why users were

presented with previously elaborated result lists and asked to choose the most interesting news to solve a need of information stated earlier.

On the other hand, users may be conditioned by an individual interest (or lack of interest) on a particular theme. To avoid this type of bias as far as possible, two result lists with current subjects were elaborated, one for each task (*Illustration 2*):

Task 1: sports news

Task 2: politics or society news

Results pages shown were static screen captures taken from *Google Noticias*. These captures were slightly manipulated to establish a small space between the zones of each result: its title, summary, source and URL. This way the eye-tracker would have a better certainty of the exact position of the user's gaze.

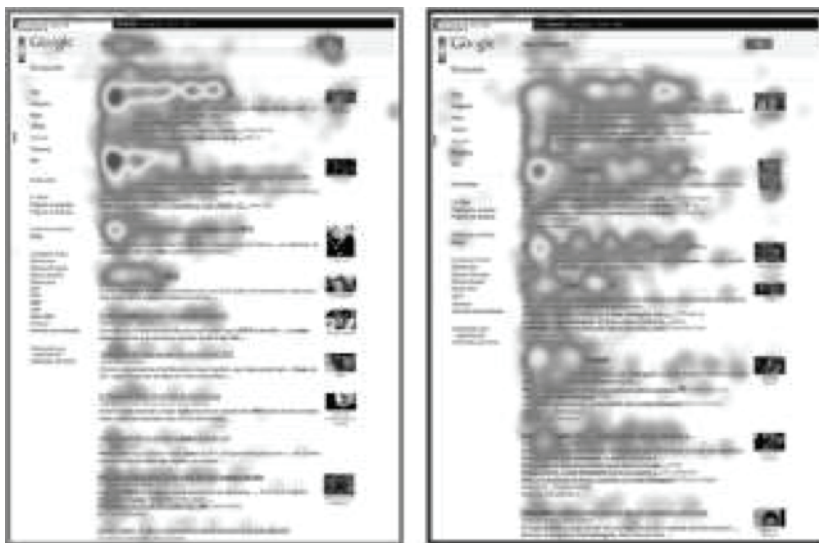


Illustration 2. Heat maps of the analyzed result lists: in the left, sports news and in the right, politics or society news.

Metrics

As mentioned before, each result of *Google Noticias* has a similar structure, title, source, date, summary and image. Each Area of Interest (AOI) was defined with the Tobii Studio software to obtain the metrics of all.

An important decision in any study using an eye-tracker is choosing which metrics will be employed to reach the aims of the study. In usability studies (Marcos, 2011), metrics related with the level of effectiveness and efficiency of the tasks at hand are usually used. For example:

- Percentage of users who fixed their gaze in an AOI.
- Time elapsed from the beginning of the task until the end.
- Number of user fixations in each AOI.
- Fixation duration in each AOI.

In this study, fixation duration was used as the main metric to determine the importance users give to each zone of the results.

RESULTS

Due to the objectives of the study, data analysis will be focused on the element “main source”, which will be compared with the principal areas of all items (title and summary). Data has been normalized since every AOI has a different size. Normalization was carried out dividing the fixation duration in each AOI by the surface of these AOIs, obtaining a normalized data of time by surface unit.

Table 2. Normalized time of fixation duration in the main areas.

	Title	Summary	Source	Differences Title and Source	Differences Summary and Source
Users	50	50	50	50	50
Mean	787,660	232,080	983,520	-195,860	-751,440
Shapiro-Wilk	0,884	0,731	0,809	0,909	0,774
Shapiro-Wilk Sig. P-value	0,000	0,000	0,000	0,001	0,000

Table 3. Wilcoxon Signed-Rank Test.

	Title and Source	Summary and Source
Z	-1,057	-5,913
Sig. P-value	0,290	0,000

In the statistical study we are working with numerical data with related samples, the same subject observes every AOI analyzed. We are especially in-

terested in identifying if significant differences exist between the variable “Title” in relation to the “Source” and the variable “Summary” also in relation to the “Source”. According to the Shapiro-Wilk test, none of the variables follows normal distribution, as it is shown in the two columns of differences (Table 2), where the p-value is lower than 0.05. Therefore, parametric tests, like the Student’s t test, cannot be applied to determine significant differences. Since these are dependent data, the non-parametric Wilcoxon Signed-Rank Test is used, proving that statistically significant differences only exist between the variables “Summary” and “Source” where a p-value lower than 0.05 is obtained. On the contrary, no significant differences exist between the variables “Title” and “Source” where a p-value of 0.290 is obtained (Table 3). Although the center of attention of this investigation lies in the behavior of users facing the zones where information sources appear, it must also be pointed out that significant differences exist between the title and the summary, favoring the title.

Differences in normalized time data between source and summary are important. A difference of 751 in favor of the source is observed, meaning there is an attention increase of 323% gazing from the summary to the sources. In other words, the source receives 4.24 times more attention than the summary (Table 2).

Therefore, according to obtained data and the statistical significance, we can state that the source area receives more attention than the summary.

CONCLUSIONS

This study analyzed the behavior of a group of users facing the result lists of *Google Noticias* at the moment of deciding which item to click. Our purpose was to understand the importance users give to the news sources that appear in the description of each item of the result list. An eye-tracking system was used to achieve this. Such system recorded the fixation duration of users in different zones of the result lists. To avoid the bias that could be generated in the areas of greater surface, relative data of fixation duration by surface unit was used.

According to obtained data, users pay more attention to the sources than to the summary. On the contrary, no significant differences exist between the time spent looking at the titles in relation with the time spent looking at the sources, as stated in the hypothesis. This indicates that users consider the source where information comes from as much as the title. A statement which does not corroborate the initial hypothesis stating that digital media make people less rigorous while choosing an article, at least this is the conclusion from this study of *Google Noticias*.

Even though these findings are solid, to generalize conclusions for the general population complementary studies are required. On the other hand, new investigations should also be carried out to confirm that the high value of the relative fixation duration in the source area in relation with the summary is due to the interest on this content, and not a consequence of this area being located in the upper section of each item, just below the title.

Speaking in terms of journalism, the study contributes with a new vision about how users face news services, realizing the importance media headings have in the moment of choosing which item to click from the result list of a news search engine.

Note:

This work is part of the investigation project CSO2011-22691-New strategies of advertising and promotion of Spanish tourist brands in the Web. Ministry of Science and Innovation, 2012-2014.

REFERENCES

- Carr, Nicholas (2011), *Superficiales. ¿Qué está haciendo internet con nuestras mentes?*, Madrid: Taurus.
- Díaz-Noci, Javier (2009), "Multimedia y modalidades de lectura: una aproximación al estado de la cuestión", in *Comunicar*, núm. 33, pp. 213-219.
- ; Meso Ayerdi, Koldo (1998), "Desarrollo del periodismo electrónico", in *El Profesional de la Información*, vol. 7, No. 12, pp. 4-11.
- Enquiro/Eyetoools Google Report* (2003), article available in: http://eyetools.com/research_google_eyetracking_heatmap.html
- González-Caro, Cristina; Marcos, Mari-Carmen (2011), "Different Users and Intents: An Eye-tracking Analysis of Web Search", in Irwin King, Wolfgang Nejdl, Hang Li (eds.), *wsdm'11 Fourth acm International Conference on Web Search and Data Mining* (Kowloon, Hong Kong, February 09-12, 2011), New York: ACM.
- Granka, Laura; Joachims, Thorsten; Gay, Geri (2004), "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* (Sheffield, United Kingdom, July 25-29, 2004), New York: ACM Press, pp. 478-479, available in en: http://www.cs.cornell.edu/People/tj/publications/granka_etal_04a.pdf
- Guallar, Javier; Abadal, Ernest (2009), "Evaluación de hemerotecas de prensa digital: indicadores y ejemplos de buenas prácticas", in

- El Profesional de la Información*, noviembre-diciembre, Vol. 18, No. 3, pp. 255-269.
- ; Rovira, Cristòfol; Ruiz, Sara (2010), “La multimedialidad en la prensa digital. Presencia de elementos multimedia y sistemas de recuperación en los principales diarios digitales españoles”, in *El Profesional de la Información*, November-December, Vol. 19, No. 6, pp. 620-629.
- Guan, Zhiwei; Cutrell, Edward (2007), “An eye tracking study of the effect of target rank on Web search”, in *Procs of the Sigchi conf on human factors in computing systems*, New York: ACM Press, available in: <http://dl.acm.org/citation.cfm?id=1240691>
- Hassan Montero, Yusef; Herrero Solana, Víctor (2007), “Eye-Tracing in Interacción Persona-Ordenador”, in *No sólo usabilidad*, October, available in: <http://www.nosolousabilidad.com/articulos/eye-tracking.htm>
- Marcos, Mari-Carmen (2011), *Disseny centrat en l'usuari: com millorar l'experiència d'ús en llocs web*, Barcelona: COBDC.
- Marcos, Mari-Carmen; González-Caro, Cristina (2010), “Comportamiento de los usuarios en la página de resultados de los buscadores. Un estudio basado en *eye tracking*”, in *El Profesional de la Información*, July-August, Vol. 19, No. 4, pp. 348-358.
- ; Nettleton, David; Sáez, Diego (2012), “Evaluation of User Search Behavior using a Web Search Log and Eye Tracking data”, in *Proceedings of the BCS HCI. People & Computers XXVI* (Birmingham, UK), pp. 262-267.
- Martínez Rubio, Rosa (2008), “La recuperación de la información en los periódicos digitales valencianos”, in Guillermo López García, *Comunicación local y nuevos formatos periodísticos en internet: cibermedios, confidenciales y weblogs*, Valencia: Servei de Publicacions de la Universitat de València.
- Nielsen, Jakob; Pernice, Kara (2009), *Eyetracking web usability*, Berkeley: New Riders Press.
- Pan, B. *et al.* (2007), “In Google we trust: Users’ decisions on rank, position, and relevance”, in *Journal of Computer-Mediated Communication*, 12 (3), article 3, available in: <http://jcmc.indiana.edu/vol12/issue3/pan.html>
- Poynter Institute (2004), *Eyetrack III: online news consumer behavior in age of multimedia*, available in: <http://www.poynter.org/extra/eyetrack2004/index.htm>
- Rele, Rachana S.; Duchowski, Andrew T. (2005), “Using eye tracking to evaluate alternative search results interfaces”, in *Proceedings of the Human Factors and Ergonomics Society* (Sept 26-30, Orlando, FL, HFES), available in: http://andrewd.ces.clemson.edu/research/vislab/docs/Final_HFES_Search.pdf
- Rodríguez-Martínez, Ruth; Codina, Lluís; Pedraza-Jiménez, Rafael (2010), “Cibermedios y web 2.0: modelo de análisis y resultados

- de aplicación”, in *El Profesional de la Información*, January-February, Vol. 9, No. 1, pp. 35-44.
- Rubio Lacoba, María (2010), “Mejor que un buscador, un ‘encontrador’”, in *Documentación de las ciencias de la información*, No. 33, pp. 273-287.
- Sánchez Vigil, Juan Miguel; Marcos Recio, Juan Carlos; Olivera Zaldua, María (2010), “Presencia y contenidos de los blogs en los principales diarios españoles”, in *El Profesional de la Información*, noviembre-diciembre, Vol. 19, No. 6, pp. 602-609.
- Schultz, Tanjev (1999), “Interactive Options in Online Journalism: A Content Analysis of 100 U.S. Newspapers”, in *Journal of Computer-Mediated Communication*, vol. 5, issue 1, September.

