Relevance Criteria Dynamics: A Study of Online News Selection on SERPs

Yung-Sheng Chang
School of Information, University of Texas at Austin, USA.
yscchang@utexas.edu

Jacek Gwizdka
School of Information, University of Texas at Austin, USA.
asist2018@gwizdka.com

ABSTRACT

A lab-based experiment was conducted to understand the multidimensional and dynamic changes in the use of relevance criteria on online news on search engine result pages (SERPs). Data from retrospective think-aloud (RTA) was first analyzed to investigate the criteria used when participants selected a search result on SERPs. The number and frequency of criteria used for relevance judgments were calculated. The criteria used on subsequent SERPs visits and on results at different ranks on SERPs were then analyzed to understand the differences in relevance criteria that were applied. The results showed that among the top six identified criteria, *specific source* and *topicality* were the two most often used relevance criteria. 59.6% of the times participants used a single criterion to make relevance judgment. Considering how criteria changed on subsequent visits to SERPs and on search results at different ranks, *specific source* was the most often used criterion initially but was then surpassed by *topicality* on subsequent visits to SERPs and on results at lower ranks. This could be explained by users using *specific source* as the primary cue in information evaluation and by users' expectation that Google's search engine algorithm shows more relevant information on the first SERP and at the top result ranks.

KEYWORDS

Relevance, Relevance criteria, Web search, SERP.

INTRODUCTION

With the current ubiquitous access to internet, users primarily access news online. The internet changed the ways in which people seek news. People used to receive news passively from radio, television, or newspapers. Access to news online lets news readers have the autonomy to actively select news. When news readers search news online, SERPs are the first stage of online news selection and evaluation. Hence it is vital to understand how people judge the relevance of online news on SERPs.

Relevance is a vital but complex concept that aids an information retrieval system to retrieve relevant information. Given its complexity, previous studies (Borlund, 2003; Saracevic, 1996; Schamber, Eisenberg, & Nilan, 1990) have defined relevance as having two characteristics: multidimensionality and dynamics. Relevance multidimensionality refers to the many criteria used in relevance assessment by different users, while relevance dynamics refers to the perception of relevance changes over time for the same users (Borlund, 2003). In past studies, Zhang, Zhang, Lease, and Gwizdka (2014) applied crowdsourcing and verified five relevance factors: reliability, topicality, scope, novelty, and understandability. Sundar (1999) focused on online news information evaluation and identified four criteria: credibility, liking, quality, and representativeness. Taylor, Zhang, and Amadio (2009) examined dynamic changes in relevance judgment within four information search processes (learning, focusing, browsing, and extracting) and found a number of common criteria used. Although there have been studies focusing on the multidimensional or dynamic aspects of relevance, there is lack of studies focusing on how news readers apply relevance criteria in their assessment of online news and how it changes with subsequent visits to SERPs. To bridge this literature gap, we conducted a lab-based experiment in which RTA was employed to elicit the relevance criteria used by participants and how the use of relevance criteria changed on subsequent SERPs and on results displayed at different ranks on SERPs.

METHOD

40 native English speakers with no vision correction participated in the study (26 females). Only participants who rated topic familiarity below medium on a five-point Likert scale were recruited. Since news information is time sensitive, we created naturalistic search scenarios related to recent news. When participants arrived in the lab, they read the experiment instructions, signed the consent form, and filled in demographic information. The experiment presentation was controlled by iMotions software (https://imotions.com/). Tobii TX-300 eye-tracker captured participants' eye-movement. Participants first performed a training task that guided them through the procedure. Then they conducted four assigned search tasks (order was randomized) on following topics: Zika virus, Brexit, 2016 Summer Olympics, and Bob Dylan Nobel Prize award. Each search task started with a display of task scenario and two questions to be answered by searching for information. Next, participants started searches by using Google search engine. While performing search tasks, participants were asked to save and annotate webpages they thought were relevant. The annotations were answers responding to the questions. Lastly, a gaze-cued RTA was conducted by replaying search sessions recorded from iMotions and participants were asked to explain their relevance judgments and search behaviors. Replay of participants' own eyes gaze helped them recall where they looked at on SERPs. The RTA sessions were open coded. Camtasia software was used to record the RTA sessions. The experiment took 1.5 to 2 hours to complete.

RESULTS

The frequency and percentage of relevance criteria used were first calculated. Low frequency criteria (<0.05%) were excluded. This left six criteria for further analysis: *specific source* (28%), *topicality* (27.8%), *result rank* (9.4%), *credibility* (6%), *familiarity* (4.9%), and *recency* (4.3%). For clarification, *specific source* means the origin of the news information, for example, BBC. *Specificity*, which is a low frequency code, means how focused was the information in relation to the search topic. We analyzed participants' use of single and multiple criteria when they selected a result judged by them as relevant on SERP (Table 1). The changes in the use of relevance criteria on subsequent SERP visits and per result rank on SERPs are shown in Figure 1.

Table 1. Frequency of single and multiple-criteria used in assessment of relevant results displayed on SERPs

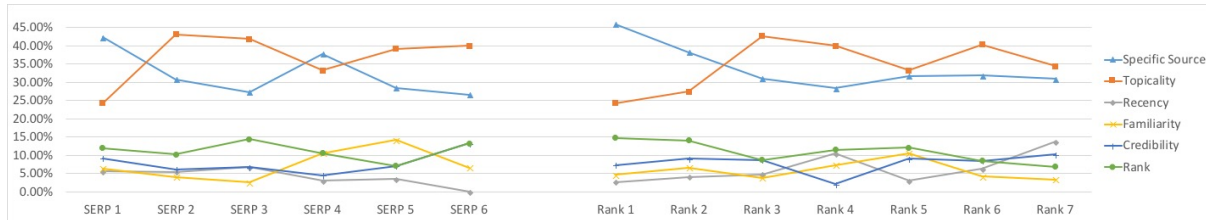| | Single Criterion | | Multiple Criteria | | | |
			Two		Three	
	#	%	#	%	#	%
Source	122	31.3	87	41.4	18	33.3
Topicality	175	44.9	42	20.0	9	16.7
Recency	23	5.9	9	4.3	3	5.6
Familiarity	10	2.6	22	10.5	8	14.8
Credibility	10	2.6	26	12.4	13	24.1
Rank	50	12.8	24	11.4	3	5.6
Total	390	-	210	-	54	-

Figure 1. Percentage of criteria used on subsequent SERP visits (left) and per result rank on SERPs (right).

DISCUSSION

We identified six relevance criteria, with *specific source* and *topicality* being the two most used criteria. Participants often used a single criterion to make relevance judgments (59.6%). The use of *topicality* decreased as the number of criteria used increased. *Specific source* remained above 30% regardless of the number of criteria used. Interestingly, the emphasis on *credibility* increased as more criteria were used. These results differ from Sundar's study (1999) due to the differences in information items on which relevance is assessed. Sundar (1999) investigated online news story itself while we focused on online news information on SERPs. The use of relevance criteria in relation to the order of SERP visits and result rank on SERPs showed the highest use of *specific source* followed by *topicality* on the first visited SERPs in search session and on the top ranked results on SERPs. However, after the first visit to SERPs and on the lower result ranks, the use of *topicality* increased and surpassed the use of *specific source*. This could be explained by the website source being the primary cue in user's online information evaluation (Greer, 2003), leading participants to focus on *specific sources* in the early stages of the search process. The relationship between relevance criteria use and the result rank may be explained by searchers' expectations of how well Google's search engine works – it is expected to display results which are more relevant on the first SERP page and at higher ranks. Our results are similar to White's study (2013), showing that users tend to select the top ranked results. In the future, we plan to gain more insights into the relevance judgment process by investigating how many relevant pages were found and how they are distributed on different SERPs and ranks. Furthermore, we will combine and analyze the use of criteria with eye fixation data.

REFERENCES

Borlund, P. (2003). The concept of relevance in IR. JASIST, 54(10), 913-925.

Greer, J. D. (2003). Evaluating the credibility of online information: A test of source and advertising influence. Mass Communication and Society, 6(1), 11-28.

Saracevic, T. (1996). Relevance reconsidered. Proceedings of CoLIS'1996.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition∗. Information processing & management, 26(6), 755-776.

Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. Journalism & Mass Communication Quarterly, 76(2), 373-386.

Taylor, A., Zhang, X., & Amadio, W. J. (2009). Examination of relevance criteria choices and the information search process. Journal of Documentation, 65(5), 719-744.

White, R. (2013). Beliefs and biases in web search. Paper presented at the Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval.

Zhang, Y., Zhang, J., Lease, M., & Gwizdka, J. (2014). Multidimensional relevance modeling via psychometrics and crowdsourcing. Paper presented at the Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval.