

The impact of task phrasing on the choice of search keywords and on the search success of inexperienced internet searchers / Elena Barsky

Abstract

As an integral part of our daily lives, information retrieval remains a challenging activity to many searchers. The reasons behind search success and failure are not always transparent. The factors that are believed to contribute to search success are diverse and include the characteristics of the searcher, the system and the search task. Among the task properties that are typically believed to shape the search process are the quantity of information required, predictability of the source of information, urgency, novelty, and other characteristics. This study focused on one particular task property – the way the task was phrased – and aimed to demonstrate that task phrasing shapes the search process and outcome.

At the center of the study was the concept of task difficulty. There were two definitions of difficulty: first, the difficulty of task phrasing was defined as the existence or absence of a direct match between the phrasing of the task and the text of an existing Web page – the target page. Two types of task phrasings (i.e., questions) were created. The *easy* task phrasing consisted of keywords extracted from an existing Web page. Queries that relied on the *easy* task phrasing retrieved the target Web page instantly. The *difficult* task phrasing was a rewording (i.e., rephrasing) of the *easy* task. Dealing with the same topic as the *easy* task phrasing, the *difficult* phrasing had no direct match with the target page and required a more thorough search. In addition, an effort was made to ensure that the *difficult* task phrasing retrieved no alternative pages that would contain the same exact keywords and the answer to the question. Second, the difficulty of the task phrasing was defined as the existence or absence of clear search instructions in the body of the task. The *easy* task contained two related requests that were presented to the searcher as a clear set of smaller questions. The *diffi-*

cult task also contained two related requests, but was presented to the searcher as one sentence with no clear request boundaries.

A group of 88 searchers received three pairs of tasks – two English pairs and one Hebrew – where each pair contained one *difficult* and one *easy* task phrasing on different search topics. All the searchers' search activities, including the queries submitted, result pages received and Web pages visited, were logged by a software application. The findings confirmed that the task phrasing shapes the search process and outcome. Overall, in most *easy* tasks, searchers were more likely to retrieve the target page among the top 10 results, visit it and find the target answer. The task assessment depended on the search outcome: in most *easy* tasks searchers were likely to be more satisfied with the search results and perceive the task as easy. Several phenomena associated with different search steps were observed. User-originated keywords, shorter queries with missing key concepts, persistent use of phrases and spelling mistakes usually negatively affected target page retrieval. Low target page rank, usually caused by shorter queries, irregular occurrence of the target page among the search results and availability of alternative pages with title tags matching the task phrasing, negatively affected target page visit rates. Finally, topic difficulty and complex text and page structure negatively affected target answer retrieval from the target page. Some search topics seemed to have broader coverage than others, which was reflected in the number of non-target and zero answers.

The experiment uncovered several Hebrew language-related search patterns: searchers tended to omit the definite article and prefixes, and also tended to prefer nouns over infinitive forms of verbs. Another finding was that some keywords were more likely to be selected for queries than other keywords.

The research method suggested in this study allowed identification of individual combinations of factors that affect each of the major stages of information retrieval.

System No.

1209866