

# Position Paper: Promoting User Engagement and Learning in Search Tasks By Effective Document Representation

Piyush Arora      Gareth J. F. Jones  
ADAPT Centre, School of Computing,  
Dublin City University, Dublin 9, Ireland  
{parora,gfjones}@computing.dcu.ie

## ABSTRACT

Much research in information retrieval (IR) focuses on optimization of the rank of relevant retrieval results for single shot ad hoc IR tasks with straightforward information needs. Relatively little research has been carried out to study and support user learning and engagement for more complex search tasks. We introduce an approach intended to improve topical knowledge of a user while undertaking IR tasks. Specifically, we propose to explore methods of finding useful and informative textual units (*semantic concepts*) within retrieved documents, with the objective of creating improved *document surrogates* for presentation within the search process. We hypothesize that this strategy will promote improved implicit learning within search activities. We believe that the richer document representations proposed in the paper would help to promote engagement, understanding and learning as compared to more traditional search engine document snippets. We propose a framework for holistic evaluation of our proposed document representations and their use in search.

## 1. INTRODUCTION

Search engines are used to support a wide range of different task types ranging from simple fact finding to complex topic exploration. At present, search engines are optimized for look-up tasks where the searcher is generally looking for a specific piece of information, and not for tasks that require multiple interactions with information to examine a topic in a less focused way.

Jiang et al. report an experimented study for known-items, known-subject, interpretive and exploratory tasks [1]. This study showed that users often become less interested in the results after a few search iterations, since the results overlap significantly or include very similar information. Further, they demonstrated that presenting information using current document surrogate methods does not work well for different types of search tasks and information needs.

*Document surrogates* are the primary way in which potentially interesting documents are revealed to searchers and are expected to reflect the information contained in the documents which is significant to the current query. However, in practice it is often hard to infer the contents of individual ranked results just by reading the surrogates, as used by current search engines, e.g. Google<sup>1</sup>, Bing<sup>2</sup> etc. as shown in Figure 2.

<sup>1</sup><https://www.google.com>

<sup>2</sup><https://www.bing.com/>

Further, information in current surrogates is only intended to indicate the relevance of a document rather than to provide important and useful information about the document to address the information need or to improve user topical knowledge. We believe it is important to address this limitation of current search systems, and study how to *select* and *present information* for different information needs to users to measure and promote learning and engagement effectively by development of enhanced document surrogates. Commercial search system such as Google, have already started using different methods for information representation for many popular known item queries, rather than displaying only surrogates to view the information. This motivates us further to search for alternate representations for complex search tasks for more general information needs regardless of their popularity. A sample collection of queries and alternative document representation gathered during our initial investigation using Google search system is available at the following link <https://goo.gl/Tz2w1y>.

**Our Position:** We propose that providing *informative surrogates* would lead to an overall improved user search experience in terms of satisfaction and increase in topical knowledge, which will promote user engagement and implicit learning within a search process. We hypothesize that developing technologies to select and present information from documents which capture relevant topical information will promote topical knowledge of a user in a search session for more complex search tasks. In this paper, we discuss a mechanism of user information presentation by developing alternative representations of document surrogates using more *meaningful summaries* and *concepts*, one of such alternative representation is shown in Figure 3.

**Motivation:** The motivation for our study is based on the observation that as we read documents, we consume and interact with information, and we increase our knowledge about a particular topic and theme. Thus the principal question that needs to be investigated is: *How can retrieved information be selected and arranged for search tasks to encourage and enhance user learning and engagement?*

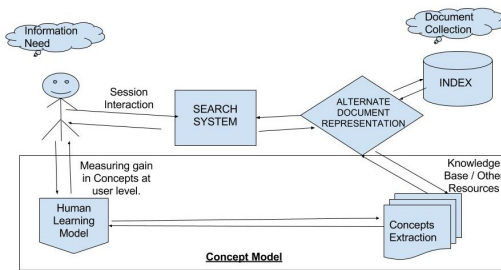
## 2. PROPOSITION

We believe it is important to study the improvement in the topical knowledge of a user with respect to the query, and develop methods for analyzing documents by measuring concepts within them to promote user learning and engagement. We hypothesize that understanding document similarities and differences at the concept level can help us to identify cues to better formulate document surrogates (in terms of the document summaries, topics, visualisations etc.) to improve the learning experience.

There has been quite some work in developing and analyzing alternative surrogates by varying: the summaries length and rep-

resentation by adding thumbnails, importance sentences etc. from the document [2, 4]. We propose to use the motivation and findings from this previous work with the aim of capturing human topical knowledge and learning. Our work differs from this earlier work since our main focus is on selection and representation of information from a document as informative summaries and meaningful concepts inclined towards improving learning and increasing topical knowledge, rather than only indicating the relevance of a document.

Learning has been measured in terms of query reformulation strategies [3], we propose to measure learning in terms of how people interact and consume different parts of the presented information based on their prior topical knowledge. Analyzing user interaction with the enhanced representation and measuring the changes in the concepts, in terms of *clicks*, *views* etc. in a session can indicate the improvement in user learning. Describing document content at the concept level would also enable the presentation of relevant information, as well as to support presentation of interesting and intriguing facts related to the user query.



**Figure 1: Architecture for measuring engagement and learning in a search task**

### 3. PROPOSED METHODOLOGY

In this research, we propose a new method of result representation by analyzing initial retrieved results to extract meaningful units from the documents, once we identify the relevant and important information from the document, we can represent it effectively as meaningful surrogates to be presented to the user. An example of our proposed alternative representation is shown in Figure 3, where the relevant concepts for a document on: “Depression Symptoms and Treatments” are indicated. Figure 2 represents the traditional surrogate representation by the current Google search system.

Figure 1 shows the proposed system architecture for supporting enhanced representation. There are two main processes added to the traditional search model: *Concept Extraction*, this deals with forming models and a framework to extract relevant and important information from a document to be represented effectively e.g. using a knowledge base or resources such as wikipedia, DMOZ etc.; *Human learning modeling*, this aims at forming models to measure the increase in topical knowledge of a user by analyzing their interactions with the enhanced representation.

The main challenges to be investigated and addressed within this proposed framework are contained in the following Research Questions (RQ’s):

- **RQ1:** *How to make efficient models to find and compare semantic concepts from documents ?*  
– In RQ1, the task is to identify effective methods for extracting semantic concepts to represent documents.
- **RQ2:** *How can we provide representations of documents using semantic concepts ?*

– In RQ2, the task is to explore different methods of presenting meaningful summaries and interpretive concepts to users in the form of enriched surrogates which promote engagement and learning during search.

- **RQ3:** *Does user learning and engagement change when results are presented using more descriptive and interpretive concepts?*

– In RQ3, the task is to measure learning and engagement by conducting task based experimental user studies. The main idea is to measure the relation between two important aspects: *increase in topical knowledge* and *enhanced results representation*.

Depression Symptoms & Warning Signs: How to Recognize ...  
www.helpguide.org/.../depression/depression-signs-and-symptoms.htm  
Learn about depression symptoms, warning signs, and causes, plus what you can do to ... Either insomnia, especially waking in the early hours of the morning, ...

**Figure 2: Surrogate from Google search engine**

Depression Symptoms & Warning Signs: How to Recognize ...  
www.helpguide.org/.../depression/depression-signs-and-symptoms.htm  
Learn about depression symptoms, warning signs, and causes, plus what you can do to feel better.  
Depression, Signs & Symptoms, Faces of Depression, Types of Depression, Causes and Risk Factors, Depression Recovery

**Figure 3: Enhanced surrogate with marked semantic concepts**

Evaluation of alternative representations should be carried out by studying search effectiveness in a task-based study in terms of different facets such as user search experience and satisfaction, and the increase in their topical knowledge to measure engagement and learning. We believe that investigating these facets will improve understanding of information representation in search engines to promote more active, interactive and engaged search processes.

### 4. CONCLUSION

In this paper, we propose to study the use of richer document representations to understand and promote user learning and engagement, as compared to traditional search engine document surrogates for the same user queries. We highlight the advantages of richer alternate representations as compared to state-of-the-art representations. We also present the main challenges which need to be addressed to support improved document representation. We believe that this research will help to measure and promote user engagement and learning effectively, and further also lead to effective representation of summaries for web documents for mobile and emerging search environments in future.

**Acknowledgment:** This research is supported by Science Foundation Ireland (SFI) as a part of the ADAPT Centre at Dublin City University (Grant No: 12/CE/I2267).

### 5. REFERENCES

- [1] J. Jiang, D. He, and J. Allan. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of ACM SIGIR 2014*, pages 607–616.
- [2] H. Joho and J. M. Jose. Effectiveness of additional representations for the search result presentation on the web. *Information processing & management*, 44(1):226–241, 2008.
- [3] P. Vakkari. Exploratory searching as conceptual exploration. *Proceedings of HCIR*, pages 24–27, 2010.
- [4] R. W. White, J. M. Jose, and I. Ruthven. Using top-ranking sentences to facilitate effective information access. *JASIST*, 2005, 56(10):1113–1125.