Learning when searching for web data

Laura Koesten
The Open Data Institute
London EC2A 4JE, UK
University of Southampton
Southampton SO17 1BJ, UK
laura.koesten@theodi.org

Emilia Kacprzak
The Open Data Institute
London EC2A 4JE, UK
University of Southampton
Southampton SO17 1BJ, UK
e.kacprzak@theodi.org

Jenifer Tennison
The Open Data Institute
London EC2A 4JE, UK
jeni.tennison@theodi.org

ABSTRACT

Searching on the web increasingly involves searching for data as well as searching for traditional web pages. To support learning from data on the web and facilitating learning through searching for data, the different characteristics of these sources need to be considered. Data usually needs additional context in order to be transformed into information and subsequently knowledge of an individual. Searching for data on the web requires a means to understand, analyse and interpret the data found. This can either be provided by the system; by the way context is presented; or by the user's prior knowledge of the topic and general data literacy skills. Therefore searching for data on the web should be considered an area in its own right for future research in the context of search as a learning activity.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.1.2 [**User/Machine Systems**]: Human information processing

Keywords

Web search, Data discovery, Context awareness, Sensemaking, Human information interaction

1. INTRODUCTION

Searching the web is a daily activity for people from a variety of backgrounds and skill sets [8] and is used for learning and discovery. Within this text learning is conceived as processing of information and as construction of knowledge, which adheres to cognitive and constructivist approaches [12]. Learning is always based on prior knowledge and is therefore different for people depending on their experience, context and abilities [7, 12, 14]. This is summarised by [7], as a personal information infrastructure, which provides the basis for processing information in order to construct new knowledge.

Search as Learning (SAL), July 21, 2016, Pisa, Italy

Searching and learning are intrinsically linked. Learning has been conceptualized as the interactive intention of searching by [10]. Learning can be the explicit aim of a search - often involving several search sessions and results that need to be interpreted and evaluated; or a byproduct of search rather than a specified goal [8]. This is especially typical for exploratory search tasks where sensemaking and learning are inherent to the task [8]. This paper focuses on how exploratory search for data can present different issues than exploratory search for web pages. Data in this case refers to structured, mostly numerical or factual data, available on the web to download.

The remainder of this paper is structured as follows. Section 2 introduces searching for data on the web as a distinct activity opposed to searching for web pages. The importance of context in enabling understanding when searching for data on the web is discussed in section 3.

2. DATA SEARCH

The majority of research about search as learning focuses on traditional web pages, which can also contain data, opposed to looking at data on the web as an independent source [6, 12]. Data published on the web is used alongside the content on traditional web pages to enable decision making about complex situations [1].

Techniques to support the task of searching for data are less advanced than those for searching for web pages. Web search engines are based on algorithms which are designed to rank web pages and do not equally support the indexing of structured content [9]. Additionally, users are likely to be not as familiar with the process of searching for data, as different skills might be needed for a successful search activity. [1] provide a hierarchy of levels of information that has data at the bottom - which is defined as raw facts; when context is added to the data it is defined as being information; and when this information is integrated it is considered to be knowledge, which means an understanding of the situation. In that sense data can be seen as the raw source, but the construction of knowledge requires an additional process.

As stated in section 1, a person's personal information infrastructure at the time of the search task determines the ability of building relationships between information sources [1, 8]. The ability to transform data to information is dependent on the context provided by the system as well as on data literacy skills of the individual. A learning process might be harder to predict or evaluate for data search opposed to search for web pages.

3. CONTEXT

Web pages often offer textual information and provide therefore curated and processed data, or information, that comes with context. Furthermore search engines are very advanced in providing additional context - they can provide contextual and personalised results by combining explicit queries with implicit feedback, such as e.g. integrating the user's browsing behaviour into a ranking system [11, 13].

Context is a necessary source of meaning [4], and there is added complexity of context within data search due to the additional information required to create meaning from data opposed to from text documents. This additional information can partly be provided through information about the data - metadata. Learning can be enhanced by providing reference points with the data or in the presentation of data - to enable the user to build a web of relationships between the different bits of information, which is needed to understand complex information [1]. For example, sensemaking of geographical data is easier when displayed in a map, and meaning can be attached to numbers if a range or a graph is presented that supports relating those numbers to reference points.

[10] describe the sensemaking process as creating knowledge structures between the data or information that has been acquired through the information seeking task. Decisions about the amount of context provided with the data are made by data publishers or by those designing the system; interface design plays a key role in representing the context [5].

The presentation of data influences sensemaking [14]. Interfaces should enable discovery of connections between different data points, that represent data in a network to make a user understand its meaning within the context of other data. An overview of search results can enhance orientation and understanding of the information provided, which can enable learning activities [10]. For data search, learning can be supported by allowing to zoom in and out of levels of data, allowing filtering and cross filtering [10], rather than displaying one piece of content at a time, such as is done with a list of documents. Navigational structures can support the cognitive representation of information [10] and this is even more important when searching for data on the web, to facilitate the transition of data to information and subsequently to knowledge. Publishing structured data as Linked Data can be seen as a partial realisation of this idea, as it provides a basis for interlinking data by providing context [2], however the majority of data on the web is not published as Linked Data.

4. CONCLUSIONS

Current search engines are optimised for searching and learning factual knowledge from web pages [10], but are not yet fully facilitating searching for data on the web or providing the means to understand, analyse and synthesise this data. Document search differs to data search as finding, accessing, understanding and using data requires additional skills. The user's prior knowledge and experience with the domain or topic determine the ability to understand the data. Skills such as accessing, interpreting and critically assessing data are part of a user's data literacy [3]. Data usually requires additional context to be interpreted, as discussed in section 3. Hence potentially more complex search interfaces are required, that offer different viewpoints to facilitate learning during the search process.

Further research is needed to understand how people make use of data resources and progress from finding to understanding [8]; which can be defined as learning. The challenges of searching for data should be an area of attention in its own right, rather than extrapolating results from traditional document search. The better the sensemaking process from data to knowledge is understood, the better systems we can create to facilitate learning from and by data search.

5. REFERENCES

[1] M. J. Albers. Human–information interaction with complex information for decision-making. In *Informatics*, volume 2, pages 4–19. Multidisciplinary Digital Publishing Institute, 2015.

[2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227, 2009.

[3] J. Calzada Prado and M. Á. Marzal. Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2):123–134, 2013.

[4] B. Dervin. Given a context by any other name: Methodological tools for taming the unruly beast. *Information seeking in context*, 13:38, 1997.

[5] S. Greenberg. Context as a dynamic construct. *Human-Computer Interaction*, 16(2):257–268, 2001.

[6] S. Halford, C. Pope, and M. Weal. Digital futures? sociological challenges and opportunities in the emergent semantic web. *Sociology*, 47(1):173–189, 2013.

[7] G. Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, New York, NY, USA, 1995.

[8] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1998.

[10] S. Y. Rieh, K. Collins-Thompson, P. Hansen, and H.-J. Lee. Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, 42(1):19–34, 2016.

[11] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 525–534. ACM, 2007.

[12] P. Vakkari. Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1):7–18, 2016.

[13] R. W. White, P. Bailey, and L. Chen. Predicting user interests from contextual information. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 363–370. ACM, 2009.

[14] M. L. Wilson, B. Kules, B. Shneiderman, et al. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1):1–97, 2010.