Just looking around: Supporting casual users initial encounters with Digital Cultural Heritage

David Walsh
Department of Computing
Edge Hill University
St Helens Road
Ormskirk, L39 4QP
United Kingdom
walshd@edgehill.ac.uk

Mark M. Hall
Department of Computing
Edge Hill University
St Helens Road
Ormskirk, L39 4QP
United Kingdom
mark.hall@edgehill.ac.uk

ABSTRACT

Cultural Heritage institutions have developed numerous ways of supporting visitors who have simply wandered in through the front door. However, for their digital collections, the CH institutions mostly provide a simple search-box, which supports the expert user, but which does not support the casual user who has just stumbled across the collection. These casual users frequently have no goal or topic in mind, but just want to have a look around what is available in the collection. For these users the blank search-box presents a significant obstacle, as without a goal or topic it is very difficult to formulate an appropriate query. In this paper we propose extending current exploratory search and information seeking models to support the initial interaction between the casual user and the collection.

Keywords

information seeking, information retrieval, casual user, exploration

1. INTRODUCTION

Cultural Heritage (CH) institutions such as museums and galleries are accustomed to dealing with visitors who arrive at their doors with no knowledge of what they could see or even what they would like to see. At the moment of arrival the only goal these visitors have is spending some time in the museum or gallery. CH institutions have developed a number of successful strategies for supporting these visitors by providing them with floor plans, fliers, guide-books, audio-guides, and guided tours.

The major limitation that CH institutions face is that the physical space available for displaying the institution's artefacts and the time required to curate the various introductory guides severely limit the number of artefacts that can be shown to visitors, with the majority of artefacts held in storage rather than on display. To address this limitation CH institutions have been engaged in a massive digitisation process, making large parts of their holdings available to the public on their websites.

Access to these collections is primarily provided through the search box. This works very well for expert users who know exactly what they want and which keywords they need to use to find what they are looking for [25]. However, CH institutions also have to support the casual user who has just stumbled across their collection in the same way that they would wander into the CH institution's physical space. For these casual users, who have no specific goal in mind when they come to the collection, the blank search-box presents an almost insurmountable problem [30]. To access anything in the collection, they are required to know at least one keyword and these users are not provided with any alternative way to find what might be available and develop an overview of the collection [15, 23, 14]. This is illustrated by the following quote from a casual user of such a collection:

> So what use are the digital libraries, if all they do is put digitally unusable information on the web. [5]

There is thus a clear need to support these casual users in their use of the collection. However, in the information seeking (IS) and information retrieval (IR) domains, the focus has always been on users with a more or less clearly defined goal, as the lack of goal breaks the fundamental assumptions in all current IS and IR models [12, 31]. Much work has been done on supporting exploration and discovery when the user has at least a very vague goal [13, 17, 7], but there remains a gap in our understanding around that first, casual interaction between the user and the collection.

2. BACKGROUND

Opening the CH institutions' archives to the world through digitisation makes them available to a much wider audience than just the museum curators and CH researchers [15]. With this expanded audience comes the requirement of supporting users outside the core group of experts, who have previously explored the CH institutions archives, to include the casual user [9, 27]. This presents a problem to current information retrieval models and techniques, as these are generally built around the concept that the user has some kind of goal, however vague, in mind.

ECIR Supporting Complex Search Task Workshop '15 Vienna, Austria
Published on CEUR-WS: `http://ceur-ws.org/Vol-1338/`.

The reasons users turn to digital information systems cover the whole spectrum from (re-)locating a piece of information they know exists to exploring an unknown topic to develop an understanding [6]. These interactions with digital information systems can roughly be classified as known-item or known-topic searches, where the user knows what they are looking for and what they expect to see, and exploratory search interactions, where the goal is to explore, learn, interpret, synthesize, and understand [18, 28, 30].

The traditional IR model describes a simple loop consisting of problem identification, query formulation, and result evaluation [25], which successfully supports the known-item and known-topic search tasks [29]. To support the more open-ended exploration interactions, this basic model has been expanded to create exploratory search models [18, 20], which have much wider scope, complexity, and duration [19, 3, 28].

These models all treat the search process as if it is completed in a single session. However, the process of satisfying an information need will often extend over multiple search sessions as the users slowly develop and refine their precise understanding of what they are looking for. A number of models of this extended process have been created to describe this information seeking journey [16, 26]. These models all describe a process in which the user starts with a very vague notion of what they are looking for and what the journey's end-point will be. Then, as the user interacts with the search system, they develop a clearer understanding of their information need and their searches become evermore focused until they develop the final queries that satisfy their information need.

The final phases of the information seeking journey are generally well supported by the traditional search model and interfaces. For the earlier, more open-ended stages, a number of exploratory interfaces have been developed. Hierarchical systems [10] were intended to help organize large sets of documents into groups or categories [8] enabling searchers to perform more sophisticated browsing tactics such as traversing and exploring nearest neighbour categories [2, 30]. Clustering approaches [7] group together related documents to give the user an overview over the "topics" in their search results. Faceted classification [24, 18, 13, 21] generates a list of the most frequent keywords for the collection (or search result) and shows these to the user. The user can then explore the collection by clicking on the keywords, rather than having to type them into the search box. Tag-clouds provide a similar visualization of the most frequent keywords. Socially curated systems [22, 11, 1] allow users to curate their own mini-exhibitions and then share these with other users, providing the new and casual user with a starting point for exploring the collection.

These approaches all suffer from a number of technical limitations, primarily around the difficulty of scaling to the massive amount of information that is available in modern Digital Cultural Heritage collections. The manual processes that create hierarchical systems cannot deal with the millions of items that exist in modern big-data DCH collections and that need to be classified. Socially curated systems suffer from the same lack-of-manpower issue and additionally to provide a comprehensive overview over a collection, they would require so many mini-exhibitions that they simply replace the problem of finding an item that the user is interested in with the problem of finding a mini-exhibition that the user is interested in. Clustering and visualization approaches can deal with the amount of data, but the resulting visualizations tend to suffer from information overload and do not provide a usable overview over the large collections. Similarly, faceted interfaces can process the amount of data available, but DCH collections are very heterogeneous [15] and showing the most frequent 20 or 30 keywords does not give the user an overview and access to more than a very small fraction of the total content.

More importantly, however, is that all these theories and interfaces start with the assumption that the user has at least a very vague goal in mind. They do not model or support the completely undirected casual user.

3. SUPPORTING THE CASUAL USER

To support the casual user in their initial interaction with the collection, the major change we propose is to let go of the concept of the "information need" as the reason for interacting with an information system. The casual user has a motivation for coming to the CH institution's site, but this it not necessarily a need for information, they might just want to procrastinate . The focus for supporting the casual user has to shift from supporting them in exploring and finding what they are looking for to supporting them in understanding what is available in the collection and where they might start browsing.

We envision a number of different interfaces that could enable such access. One approach would be to generate textual summaries that describe the type of content available from the collection. Such an approach would need to analyse the individual items meta-data using an algorithm such as LDA [4], then combine that with a textual resource such as Wikipedia to generalise the topics, and finally generate textual descriptions such as "The collection contains historical artefacts from ancient Egypt, space exploration, horology, and a modern collection of oceanographic specimens." The user could then click on any of the topics to get a summary of the content in the selected area of the collection, enabling them to freely explore.

An alternative would be to use the topic structure to generate an exploratory semantic map, that the users can interact with and explore like they would a physical map. Another approach could be to look at developing an automatic measure for the "interestingness" of items in the collection. This could then be used to sample items from the collection to show the casual user the "highlights" of the collection.

The investigation of potential interfaces will have to be accompanied by a series of user studies that investigate how casual users develop a topic they are interested in, when confronted with a new collection. This will enable us to extend the existing models for exploratory search and information seeking by providing a more detailed understanding of the initial phase in which the user develops their information need. This extension will enable information systems to support the complete information journey, from the development of the information need to its final fulfilment.

Finally, while in DCH this issue is particularly prevalent, understanding the casual user who has no immediate need could also have significant impact in the area of E-commerce. E-commerce is a major growth area, but currently does not support browsing the available things in the same way that you can browse through a shop.

4. REFERENCES

[1] E. Agirre, N. Aletras, P. Clough, S. Fernando, P. Goodale, M. Hall, A. Soroa, and M. Stevenson. Paths: A system for accessing cultural heritage collections. In *ACL (Conference System Demonstrations)*, pages 151–156. Citeseer, 2013.

[2] M. J. Bates. Information search tactics. *Journal of the American Society for information Science*, 30(4):205–214, 1979.

[3] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[5] C. L. Borgman. The digital future is now: A call to action for the humanities. *Digital humanities quarterly*, 3(4), 2009.

[6] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Information processing & management*, 31(2):191–213, 1995.

[7] C. Carpineto, S. Osiński, G. Romano, and D. Weiss. A survey of web clustering engines. *ACM Computing Surveys (CSUR)*, 41(3):17, 2009.

[8] M. Chen, M. Hearst, J. Hong, and J. Lin. Cha-cha: A system for organising intranet search results. 1999.

[9] A. S. Cifter and H. Dong. User characteristics: Professional vs. lay users, 2009.

[10] E. Clarkson, K. Desai, and J. D. Foley. Resultmaps: Visualization for search interfaces. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1057–1064, 2009.

[11] M. Hall, P. Goodale, P. Clough, and M. Stevenson. The paths system for exploring digital cultural heritage'. In *Clare Mills, Michael Pidd and Esther Ward. Proceedings of the Digital Humanities Congress*, 2012.

[12] M. Harvey, M. Wilson, and K. Church. Workshop on searching for fun 2014. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, pages 6–6, New York, NY, USA, 2014. ACM.

[13] M. A. Hearst. Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4):59–61, 2006.

[14] K. Hornbæk and M. Hertzum. The notion of overview in information visualization. *International Journal of Human-Computer Studies*, 69(7-8):509 – 525, 2011.

[15] A. Johnson. Users, use and context: supporting interaction between users and digital archives. *What Are Archives?: Cultural and Theoretical Perspectives: A Reader*, pages 145–64, 2008.

[16] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *JASIS*, 42(5):361–371, 1991.

[17] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

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

[19] D. A. Norman. *The psychology of everyday things*. Basic books, 1988.

[20] P. Pirolli. Powers of 10: Modeling complex information-seeking systems at multiple scales. *Computer*, 42(3):33–40, 2009.

[21] P. L. Schmitz and M. T. Black. The delphi toolkit: Enabling semantic search for museum collections. In *Museums and the Web 2008: the international conference for culture and heritage on-line*, 2008.

[22] F. M. Shipman, R. Furuta, D. Brenner, C.-C. Chung, and H.-w. Hsieh. Guided paths through web-based collections: Design, experiences, and adaptations. *Journal of the American Society for Information Science*, 51(3):260–272, 2000.

[23] M. Skov and P. Ingwersen. Exploring information seeking behaviour in a digital museum context. In *Proceedings of the second international symposium on Information interaction in context*, pages 110–115. ACM, 2008.

[24] D. A. Smith, A. Owens, A. Russell, C. Harris, M. Wilson, et al. The evolving mspace platform: leveraging the semantic web on the trail of the memex. In *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, pages 174–183. ACM, 2005.

[25] A. Sutcliffe and M. Ennis. Towards a cognitive theory of information retrieval. *Interacting with computers*, 10(3):321–351, 1998.

[26] P. Vakkari, M. Pennanen, and S. Serola. Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information processing & management*, 39(3):445–463, 2003.

[27] P. Vilar and A. Šauperl. Archival literacy: Different users, different information needs, behaviour and skills. In *Information Literacy. Lifelong Learning and Digital Citizenship in the 21st Century*, pages 149–159. Springer, 2014.

[28] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proceedings of the 16th international conference on World Wide Web*, pages 21–30. ACM, 2007.

[29] R. W. White and R. A. Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009.

[30] M. L. Wilson and D. Elsweiler. Casual-leisure searching: the exploratory search scenarios that break our current models. In *Proceedings of HCIR*, pages 28–31, 2010.

[31] C. Ye and M. Wilson. The characteristics of casual sessions in search behaviour logs, 2014.