Information Seeking in Digital Cultural Heritage with PATHS

Mark M. Hall,
Paul D. Clough,
Samuel Fernando,
Paula Goodale,
Mark Stevenson
Sheffield University
S1 4DP, Sheffield, UK
{m.mhall,p.d.clough,
s.fernando,p.goodale,
r.m.stevenson}@sheffield.ac.uk

Eneko Agirre,
Arantxa Otegi,
Aitor Soroa
University of the Basque
Country
Donostia, Spain
{e.agirre,arantza.otegi,
a.soroa}@ehu.es

Kate Fernie,
Jillian Griffiths
MDR Partners
United Kingdom
kate.fernie@mdrpartners.com
jillian.griffiths@mdrprojects.com

ABSTRACT

Current Information Retrieval systems for digital cultural heritage support only the actual search aspect of the information seeking process. This demonstration presents the second PATHS system which provides the exploration, analysis, and sense-making features to support the full information seeking process.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces

Keywords

exploration; information seeking; interactive information retrieval; demonstration

1. INTRODUCTION

Current Information Retrieval (IR) systems in the domain of digital cultural heritage provide good access for experts through the standard search-box paradigm [8], as these users are familiar with the collections, have specific information needs, and know which keywords to use to satisfy these information needs. However, for users who are new to the collections or do not have a specific information need the search box presents an almost insurmountable obstacle [9]. Additionally these IR systems only support a small fraction of the information seeking process [3, 5], forcing users to augment the IR systems with other tools.

To support the whole information seeking process for both experts and novices, IR systems are required that provide an

initial overview over the collection [2], functions for exploring collections [4] such as thesauri [7] and faceted browsing [1], somewhere to collect potentially relevant items [5], and finally the ability to organise the items into a sense-making structure.

The PATHS (Personalised Access To cultural Heritage Spaces) project aims to address this issue through a combination of content pre-processing and front-end application functionality. The demo presented in this paper showcases the second PATHS prototype that has been developed based on the experience of building, using, and evaluating the first prototype. While the prototype uses cultural heritage data, it can be applied to data-sets from other domains as well.

2. THE PATHS SYSTEM

When a new user comes to the PATHS system, they are offered two main avenues for overviewing and exploring the collection. One is a set of guided tours, called *paths*, that have been curated both by experts and other users, and represent a more focused version of the paths concept introduced by [6]. These *paths* take the user on a trip through an aspect of the collection, providing a first introduction to the kind of information available. The second avenue is through the use of an automatically generated thesaurus that provides topic-based exploration support. Figure 1 shows the textual visualisation that enables the user to explore and drill down into the thesaurus. Tag-cloud and map-based visualisations are also provided to support users with different cognitive styles. The inset in figure 1 shows the map-based visualisation, which creates a novel, two-dimensional semantic map of the thesaurus that supports a visual exploration of the collection.

As the user explores the collection they can collect individual items into their workspace. This allows them to keep track of what they have explored and of items that might help them to fulfil their information needs. The items in the workspace can be annotated and arranged to support the analysis of what the user has found.

When the user has found a sufficient number of items to fulfil their information need they can arrange the items into a *path* structure to support the final sense-making step in the information seeking journey (fig. 2). The *path* is an ordered structure of items that supports branching and
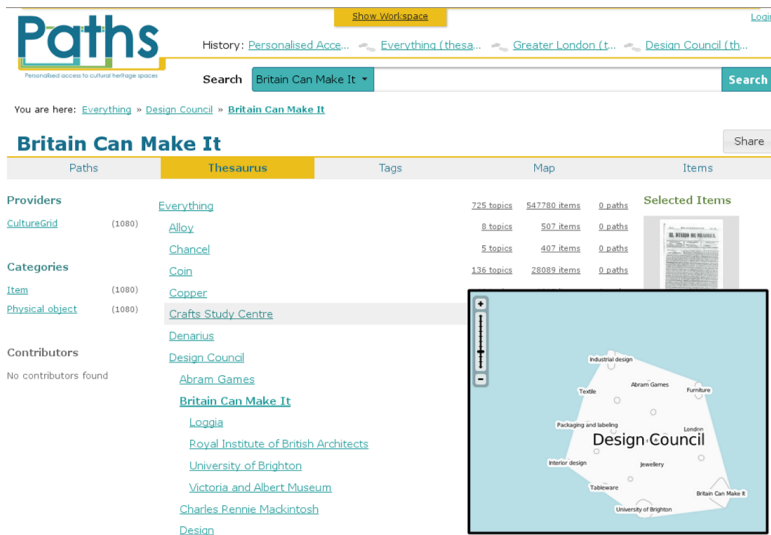
Figure 1: Screenshot showing some of the exploration features provided by the PATHS system. The main area shows the thesaurus for exploring the collection. The inset in the bottom-right corner shows a semantic map that has been developed to enable a more visual exploration of the collection.
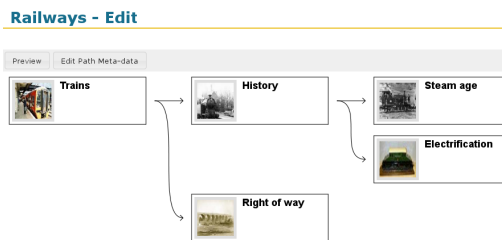

Figure 2: Overview over an example *path* on the topic of "Railways", showing the ability to structure the information by creating multiple branches.

user-annotation of each step in the path, creating the user's narrative through parts of the collection. Potential uses for the *path*s are as a personal overview over a topic that can evolve as the user further explores the collection, as a virtual exhibition or story created by a collection curator, or as a learning object to support teaching activities.

3. DEMO DESCRIPTION

The PATHS demo will provide a guided tour to the PATHS system and demonstrate how a user can explore the collection, collect items and then form these into the sense-making *path* structure. Participants can then explore the prototype system on their own, which will be provided at http://www.paths-project.eu. The demo will provide access to a sub-set of the data available in Europeana[1], covering approximately 1.8 million items from the English and Spanish collections.

---

[1]http://www.europeana.eu

4. ACKNOWLEDGEMENTS

5. REFERENCES

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

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

[3] C. Kulthau. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science and Technology*, 42(5):361–271, 1991.

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

[5] T. Russell-Rose and T. Tate. *Designing the search experience: The information architecture of discovery.* Morgan Kaufmann, 2013.

[6] F. Shipman, C. Marshall, R. Furuta, D. Brenner, H.-w. Hsieh, and V. Kumar. Creating educational guided paths over the world-wide web. In *Proceedings of Ed-Telecom*, volume 96, pages 326–331. Citeseer, 1996.

[7] A. Shiri, C. Revie, and G. Chowdhury. Thesaurus-enhanced search interfaces. *Journal of information science*, 28(2):111–122, 2002.

[8] A. Sutcliffe and M. Ennis. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10:321–351, 1998.

[9] M. Wilson, K. B, S. MC, and S. B. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1):1–97, 2010.