Enabling Semantic Access to Cultural Heritage: A Case Study of Tate Online

Paul Clough[1], Jennifer Marlow[1] and Neil Ireson[2]

[1]Department of Information Studies, [2]Department of Computer Science
University of Sheffield, UK
{p.d.clough; j.marlow}@shef.ac.uk; n.ireson@dcs.shef.ac.uk

Abstract. This paper describes a study in which we enable semantic access to cultural heritage information from Tate Online, a large online art collection. An online survey was conducted with 635 visitors of Tate Online to assess the utility of semantic access in this domain. Results from this survey were used to develop a prototype system to demonstrate advanced search and browse functionalities (including faceted browsing, timelines and maps) based on semantic enrichment of data from Tate Online. A task-based user experiment involving 14 participants was conducted to evaluate the prototype system. Results showed the benefits of semantic enrichment for certain tasks, such as open-ended browsing.

Keywords: Semantic organization, faceted browsing, cultural heritage

1 Introduction

Cultural heritage institutions around the world often rely on the Internet to enable access to digitised versions of their collections. Similarly to digital library access, many institutions provide users with both free-text searching of collection content, together with browsable categories, such as object type or subject matter, also useful in organizing items. However, these broad groupings may not always allow the individual to locate desired material quickly and easily. Advances in technology have made it increasingly possible to search and browse for items using richer sets of parameters based on semantic information. For example, faceted browsing enables a user to broaden or narrow a search based on multiple criteria at once [1].

The study described here used Tate Online[1] (the web presence of Britain's Tate art galleries) as a case study to examine how Semantic Web technologies could be applied to enhance access to a large online art collection. An online questionnaire was used to elicit feedback from 635 site visitors in order to gauge their needs and preferences with regards to browsing and exploring material related to artists and artworks. This feedback helped to guide the design of a prototype faceted browsing system based upon material from Tate Online. An initial task-based user evaluation has also been carried out with 14 people to obtain qualitative feedback regarding the prototype. Although this study focused on the Tate, the findings and technical principles behind the design could likely be generalized to other collections in the art and cultural heritage domains as well.

[1] http://www.tate.org.uk/

2 Background

According to Fluit et al. [2], "the Semantic Web is an extension of the current World Wide Web, based on the idea of exchanging information with explicit, formal and machine-accessible descriptions of meaning." As Maedche & Staab [3] explain, these descriptions can be utilised to facilitate finding, integrating and connecting information in a way above and beyond that which can be done with a simple keyword search. Hildebrand et al. [4] outline the various elements of the semantic search process, which include construction of the query, execution of the search algorithm, and presentation of results. With regards to interface design matters, they mention both typical and more experimental visualization techniques ranging from ranked lists, clustered result displays, tag clouds, cluster maps, and data-specific designs such as timelines.

Benjamins et al. [5] highlight the value of semantics in the humanities domain, stating that most information-seeking in this area involves "events, persons, and movements in a historical or cultural context." Similarly, Hyvönen [6] asserts that the cultural heritage domain is well suited to the creation of semantic portals. These can, among other things, (1) give an aggregated, global overview of heterogeneous content and (2) provide a more "intelligent" way of examining content through semantic linkages. There are several ways in which said intelligent services can utilize semantic information. These include semantic search, semantic auto-completion, faceted semantic search, semantic browsing and recommendation links, relational search, and visualizations on maps and timelines. Projects such as MultimediaN[2] and eChase[3] have also explored the use of semantic enrichment in the cultural heritage domain.

However, despite the intuitiveness of providing semantic access to cultural heritage material, the types of browsing systems described here have primarily been adopted on experimental sites or for small amounts of data, rather than being deployed by large cultural heritage organizations. Faceted browsing has been previously evaluated, but few results are reported in the cultural heritage domain. This study attempted to explore the feasibility (and potential uses) of semantic organization in the context of a larger online art collection, with an emphasis on different types of task and user.

3 Methodology

As an initial means of gathering background information on users' typical tasks and needs when using a site like Tate Online, a questionnaire was published targeting visitors of the site. This survey was offered as a pop-up to individuals visiting the Tate Collection website. It was conducted to get an idea of what people use the Collection site for and to use this input to help guide the design of a system for browsing and exploring material related to artists and artworks.

[2] http://e-culture.multimedian.nl/
[3] http://www.echase.org/

Table 1. Example task types given to participants

Task type	Example
Specific fact-finding	How many works by Henry Bishop are in the collection?
Extended fact-finding	Which of the following artists lived during the same time period?
Open-ended browsing	Find as many artists as you can who lived between 1800 and 1900.
Exploration	Find an abstract painting that you like in the collection

The feedback from the questionnaire helped to guide and justify the design of the prototype faceted browsing system, discussed in Section 5 below. Once a working prototype system was developed, this was then evaluated by a set of users with characteristics similar to Tate Online visitors (both with and without specialized artistic knowledge). Users were asked to perform four search tasks with the faceted browsing system and an equivalent four tasks with the current Tate Collection website (Table 1). These tasks addressed the various information seeking behaviours as proposed by Shneiderman [7], and are also applicable in the cultural heritage domain: specific fact-finding, extended fact-finding, open-ended browsing, and exploration of availability. Performance on the tasks was observed, yielding both quantitative data in the form of time taken and tactics used, and qualitative data in the form of self-reported system satisfaction and task difficulty levels.

4 Results of Tate Online Questionnaire

A total of 635 individuals world-wide answered the online questionnaire. Of these responses, 42% stated their primary reasons for visiting the Tate site were related to academic/research objectives, and 34% were using it out of personal interest. Roughly 2/3 of people visited the site looking for something specific, such as a particular artist (45%) or artwork (19%), or both (12%). Alternatively, 14% looked for types of artworks, and 10% were just browsing the collection.

Table 2. Most important criterion for search, by user type

Topic	Total %	% of general users	% of expert users
Artwork subjects	44.7	34.6	51.5
Relationships	31.1	28.8	23.3
Dates	13.1	13.5	15.2
Gender of artist	3.5	13.5	4.0
Nationality of artist	3.6	7.7	3.0
Locations	4.0	1.9	3.0

When asked which criteria would be the most important when searching for an artwork or artist (from the list in Table 2), overall 45% of respondents mentioned subjects of artworks, 31% voted for relationships between artists, and 13% selected dates (such as artists' birth dates or artwork creation dates). Table 2 shows similarities between user groups (general and expert user): although absolute percentages vary, the ranking of functionalities remains the same.

Based on these responses, two of the features are already partially offered by the Tate Collection search functionality: artwork subject search is offered based on a hierarchy of related subjects (this is useful in the case where the artists' names or

artwork title are unknown). Dates are also incorporated into the Collection site's advanced search, however in this case the user is required to directly input a specific year or range of years. Enhanced presentation of date-based information could include an interactive timeline view. However, also of interest (and currently not possible) would be the ability to explore relationships between artists (e.g. finding people who were inspired by, worked with, or were related to a given artist). Presumably this is important information for students of art (and also of potential interest to casual browsers).

Respondents were asked which of a range of possible features would be most useful in enhancing access to material in the Tate Collection. The "most useful" feature as chosen by the greatest number of respondents (26.2%) was faceted browsing (the ability to search for information based on several criteria at once, e.g. "find female French artists from the 19th century"). Also of high interest was the ability to explore relationships between artists. Finally, the possibility of accessing other (related) links in English via the Collection pages was deemed to be useful. This also emerged as a theme in a previous 2004 internal Tate Online survey, in which people expressed a wish to be able to access links to other sites (artists' official pages, other high quality art/museum web pages) provided on the Tate pages, in order to further their information seeking and exploring process. Upon investigation, the answers were similar between expert and general users; the most notable difference being the percentage of people who would find it useful to explore relationships between creators and creations: this was highly ranked by the expert users, but of low importance to the general users (for whom links in English were more important).

5 Prototype Design

The design of the prototype was influenced by the desires of the Tate users (expressed in the online questionnaire and discussed above), the availability of (semantic) data and the toolkit user to develop the interface.

With the permission of the Tate we scraped the Tate Online website and created a number of "wrappers" to extract the structured data used to generate the pages. This provided information on the (approximately 3,000) artists on the site (i.e. their names, birth/death dates), and information and images on their (approximately 30,000) related artworks (i.e. titles, subject). This information was augmented by linking the Tate artists to the information provided by Getty Union List of Artist Names[4] (ULAN). ULAN contains information on over 100,000 artists, including name variations, nationality, birth and death dates/places, role, gender, relationships; thus these form the facets for exploration[5].

The Tate Online Collection does provide an "Advanced Search" facility, which allows the user to search on artist names and dates and artwork titles, date and subjects. However this facility still provides limited exploration of the collection as the search results only returns information on matching artworks and refocusing the query requires submitting a new search. The prototype system aims to provide access to the same collection via a faceted browsing interface. There are a number of faceted

[4] http://www.getty.edu/research/conducting_research/vocabularies/ulan/

[5] Further details of this system will be provided in subsequent publications.

browsing development tools available (e.g. SlashFacet[6], mSpace[7], Flamenco[8]). The prototype system was implemented using the Simile Exhibit toolkit[9] from MIT, which offers a lightweight (it is implemented entirely in JavaScript and the interface is configurable via the webpage HTML) and comprehensive system (including a variety of types of facet (i.e. numerical, hierarchical) and views (i.e. timeline, maps)), which allows for flexible and fast prototyping.

The prototype allows users to explore the collection using one of four views:

- Artist view (providing information on artists and the titles of their artworks).
- Artwork view (providing more information on the artwork and featuring thumbnails of all artworks).
- Timeline view of artists' birth/death dates.
- Map view of artists' birth places.

The results displayed in all the views are constrained by the values selected in both the artist and artwork facets. Figures 1-4 illustrate these various views. Unfortunately only a basic facility for exploring relationships was implemented (i.e. the user can see the information on immediately related artists). This was mainly due to the fact ULAN provided relatively sparse network of relationships between the Tate artists, but also because there was no readily available means to provide a relationship exploration view (e.g. a hyperbolic graph) in the Simile Exhibit Tool.

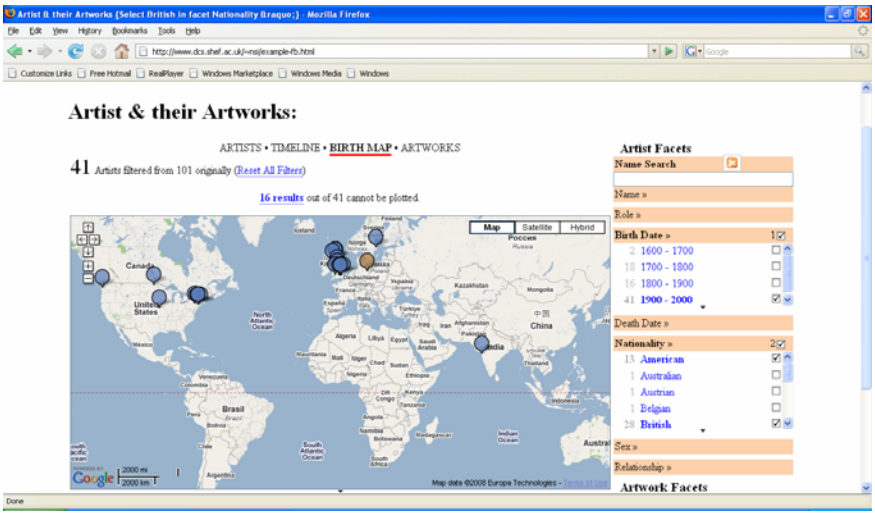
Figure 1. Artists view using facets to narrow down results to British artists born between 1900 and 2000. Basic information shown includes biographical information and list of artworks (along with thumbnail image and creation date)

[6] http://slashfacet.semanticweb.org/
[7] http://mspace.fm/
[8] http://flamenco.berkeley.edu/
[9] http://simile.mit.edu/

Figure 2. Artworks view displaying larger thumbnails of all artworks relating to the artists whose facets are selected (again, the same birth date and nationality facets as in Figure 1 are chosen)

Figure 3. Timeline view (once again, depicting all British artists born between 1900 and 2000)

Figure 4. Map view (plotting artists on a map based upon their place of birth)



6 Results of the User Evaluation

After developing a prototype system implementing some of the functionality in Table 2 (predominantly faceted browsing), a user experiment was carried out to evaluate the prototype. In total, 14 individuals participated in the user evaluation as described in Section 3. As the vast majority of the participants were not art experts (therefore representing the segment of Tate Online visitors, who come to the site for leisure purposes or out of personal interest), it is perhaps to be expected that they would be most likely to demonstrate an open-ended browsing type of behavior. However, although this task was most realistic, it was also deemed to be the most difficult kind to complete with both systems (Table 3).

Table 3. Mean ratings of difficulty and time taken (mins) to complete task, by task type

Task type	Mean rating (1=not difficult, 7=very difficult)	Time taken (Faceted browser)	Time taken (Tate)	Significance (2-tailed t test)
Specific fact-finding	1.82	0.27	0.22	.416
Extended fact-finding	2.43	1.32	1.72	.338
Open-ended browsing	3.10	3.70	4.90	.036* ($p<0.05$)
Exploration	3.50	1.95	2.15	.838

Table 4 shows the percentage of participants using (at least once) the available functionality, by task type (e.g. 71.4% of users performed an artist/artwork search for specific fact-finding). As Table 4 demonstrates, the task type affected the strategies employed and features used by the participants. Whilst search was more popular for simple fact-finding, facets were employed by at least some people for all tasks. The

features used corresponded with the search criteria highlighted in Table 2; however, what emerges is that different features are useful for different information needs.

Table 4. Percent of participants using functionality, by task type

Task type	Artist/Artwork Search	Facets	Timeline View	Artwork View	Artwork Subjects/ Keywords
Specific fact-finding	71.4	28.5	0.0	35.7	0.0
Extended fact-finding	64.3	50.0	42.8	0.0	0.0
Open-ended browsing*	0.0	85.7	28.5	7.1	0.0
Exploration	42.8	42.8	0.0	71.4	51.7

*Users were given a maximum of 5 minutes to complete this task

When individuals were asked which system they preferred to use for each of the various tasks, the faceted browser overall was viewed as equally or more preferable than the current Tate site, although again, the degree to which this was felt depended on the task (Table 5). In most cases one site did not have any major advantage over another, with the exception of the open-ended browsing task (in which the scenario was slightly biased towards the faceted browser's functionalities).

Table 5. System preference, by task type

Task type	Percent preferring Faceted Browser	Percent preferring Tate Online	Percent with no preference
Specific fact-finding	30.8	30.8	38.4
Extended fact-finding	64.3	21.4	14.3
Open-ended browsing	84.6	7.7	7.7
Exploration	50.0	35.7	14.3

7 Conclusion

The study described here used both an online questionnaire and task-based evaluation as means of providing input into the design of a faceted browsing system, specifically designed to enable access to cultural heritage information. The focus of the content related specifically to artists and artworks. The initial results indicate that users can have different information seeking needs, depending on their level of expertise in the art domain, and their reasons for visiting an art-related site.

The preliminary results of the evaluation, which was based on tasks encompassing a range of search behaviors, support findings by Capra et al. [8] that faceted systems may be most useful to find a specific type of information or to help narrow down a search. This was confirmed by users' comments and observed patterns of use. However, Tvarožek & Bielikova [9] bemoan the fact that many existing systems do not support browsing and exploratory tasks sufficiently. Further research may consider the role of a user's expertise more closely, involve a larger number of expert users, and focus on the creation of tasks (and other means of information visualization) that are most likely to be of use in various situations.

Acknowledgments

Work partially supported by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project MultiMATCH contract IST-033104. The authors are solely responsible for the content of this paper.

References

[1] Hearst, M. A. 2006. Clustering versus faceted categories for information exploration. Commun. ACM 49, 4 (Apr. 2006), 59-61.

[2] Fluit, C., Sabou, M., & Van Harmelen, F. (2005). Ontology-based information visualization: Towards Semantic Web applications. In V. Geroimenko (Ed.), Visualizing the Semantic Web (2nd ed.): Springer Verlag.

[3] Maedche, A., & Staab, S. (2002). Applying semantic web technologies for tourism information systems. 9th International Conference for information and communication technologies in tourism (Enter-2002), 311-319.

[4] Hildebrand, M., van Ossenbruggen, J., & Hardman, L. (2007). An analysis of search-based user interaction on the semantic web: Centrum voor Wiskunde en Informatica.

[5] Benjamins, V.R., Contreras, J., Blazquez, M., Dodero, J.M., Garcia, A., Navas, E., Hernandez, F., & Wert, C. (2004). Cultural heritage and the semantic web. Proceedings of 1st European Semantic Web Symposium, 433-444.

[6] Hyvönen, E. (2007). Semantic portals for cultural heritage, http://www.seco.tkk.fi/publications/2007/hyvonen-portals-2007.pdf (Accessed 2 June 2008.).

[7] Shneiderman, B. (1998). Designing information-abundant web sites: issues and recommendations. International Journal of Human-Computer Studies, 47(1), 5-29.

[8] Capra, R., Marchionini, G., Oh, J., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. Proceedings of 7th ACM/IEEE Joint Conference on Digital Libraries, 442-451.

[9] Tvarožek, M., & Bielikova, M. (2007). Personalized faceted navigation for multimedia collections. Proceedings of 2nd International Workshop on Semantic Media Adaptation and Personalization, 104-109.