Supporting Public Browsing of an Art Gallery Collections Database

Sean Bechhofer, Nick Drummond, and Carole Goble

Department of Computer Science,
University of Manchester,
Oxford Road, Manchester M13 9PL, UK
seanb@cs.man.ac.uk,
WWW home page: http://img.cs.man.ac.uk

Abstract. Increased public awareness and usage of the Web suggests that a commensurate Web presence is required from providers of cultural heritage information such as Galleries and Museums. While galleries have traditionally supplied goal-driven search facilities to specialists, they must now provide browsing and query facilities to casual users, with less precise information seeking requirements. Hypertext systems provide an appropriate technology to support the networks of associations required in order to provide path-based browsing. Requirements are twofold: browsing support for the users; and authoring support in the creation of pathways. In our prototype, we combine techniques from Hypertext and Information Retrieval to provide access to artifacts drawn from the costume collection of the Manchester City Art Gallery. We provide similarity based browsing, using terms from the artifacts' metadata to calculate similarities. The approach is simple, yet effective as the results of a user evaluation show.

1 Introduction

In the past it has been difficult for members of the public to access information due to a lack of available resources, libraries or expert help. However, galleries and museums are under pressure to increase public access to collections, and technologies such as the WWW have made information less restricted by medium and more widely available. Increased public awareness and usage of the Web suggests that a commensurate Web presence is required from providers of cultural heritage information.

The Manchester City Art Gallery of Costume at Platt Hall, was the first institution in this country to be solely concerned with the history of clothes. The collections held by the Gallery are extensive, covering all aspects of dress, dress care and dressmaking, as well as aids to the appearance. There are associated collections of textiles, embroideries, lace and dolls. The total number of items in these collections is approximately 19,400, including 12,850 items of costume, 3,800 OPUA (objects of personal use and adornment), 2,400 textile items and 376 dolls or items of dolls' clothing. There are in addition approximately 1,600 unaccessioned items used for study and as handling collections.

Along with many other galleries and museums, the Gallery is investigating the use of IT as a way to improve public appeal and effectiveness as an information provider.

The Gallery's visitors fall into three main categories:

1. Goal-driven specialists who are seeking specific information and have a precise information need;
2. Casual but goal-driven seekers, for example students given a specific assignment, that have an imprecise information need;
3. Casual visitors who are not goal-directed and have an imprecise information need.

While galleries have traditionally supplied search facilities to specialists, they must now provide browsing and query facilities to casual users of category 3, whose information seeking requirements are imprecise and less goal-driven. The truly casual user will have no idea of the content of the collections, nor of how they might be categorised, classified or indexed. Not only is their information need vague, but they would not be able to articulate it with any degree of confidence of acquiring a result if they could [5]. Access to collections must be tailored towards identified audience needs.

A considerable body of research exists on information seeking [4, 14, 22], predicated on the analysis of the interaction of an expert mediator, e.g. a gallery specialist or librarian, with a searcher. A common scenario is that the visitor has some idea of the kind of artifact they would like to look at, and they describe this to the gallery front-desk advisor. These descriptions vary widely in form and clarity, including concrete requests: *the Sunflowers by Van Gogh*; simple "kind of" categories: *a wedding dress*; vague requests, using abstract categories: *something glamorous*; and complex, narrative descriptions where the customer "paints a picture" of what they would like: *19th century men's evening suit*.

Analytical search, as defined by Pejterson [19] is supported on the WWW by search engines, ideal for goal-driven precise retrievals, such as concrete requests or possibly narrative description. They are less ideal for supporting search by analogy or browsing, though they can be used to identify a sub-collection that can be browsed.

Our focus here is primarily on browsing [17] – an undirected navigation around the collection of objects, taking short steps from point to point. Browsing provides an overview of a space, requires a smaller cognitive load than an analytical search strategy and perhaps most importantly in this context, can aid in discovery and learning. In this case the overview provides an idea of the content of the collection. *Across-document browsing* allows the user to gain a sense of the form of the collection – a casual user in the gallery is unlikely to invest time and effort in a specific search.

In order to support such browsing, the crucial question that the system must address is "*where can I go from here?*" or "*what other things like this are there?*". To support navigation, documents should be dynamically linked together in a range of diverse ways; pathways, such as Walden Paths [10] should be formed that

act as trails through the network. Users have diverse perspectives [6]. It must be possible to locate information from a range of entry points, and associate information that is similar in a variety of ways. Information can be richly linked and *classified* into multiple clusters simultaneously, based on some shared notion of similarity. Hypertext systems would appear to be an appropriate technology to support the complex networks of associations required to support path-based browsing. Thus requirements are twofold: better **browsing** support for the users; and better **authoring** support for the dynamic creation of browsing pathways.

The paper describes an experiment into how such an information provider can present information in an effective, efficient and satisfactory manner, focusing in particular on the task of public access browsing. In our experimental prototype, we draw on and combine techniques from Hypertext and Information Retrieval to provide scalable access to artifacts drawn from the costume collection of the Manchester City Art Gallery. The specific browsing paradigm we follow is one of analogy based browsing, using terms in the artifacts' metadata to calculate similarities.

2   Hypertext

The word "hypertext" was first introduced by Nelson [18] although its origins can be traced back to the Memex of Bush [7]. A hypertext consists of a collection of linked nodes or documents. The user navigates from node to node, traversing or following the links. The gallery objects and their descriptions in the collection form a hypertext − browsing the objects in the collection is a form of link-following.

2.1   First Generation Hypertext

A first generation hypertext system consists of a collection of nodes or documents along with manually authored, static, untyped point-to-point links. Although such systems (for example the WWW) are extremely powerful and offer flexibility, they suffer from several problems. Maintenance is difficult, particularly if the number of nodes is high, or the information represented by the nodes changes frequently. If a new node is to be added, we may have to add many new links to pre-existing nodes. Similarly, deletion of nodes can lead to problems of "dangling links". This is a crucial issue as we expect the system to be extensible as further objects are added to the collection. If we are to use a hypertext model as a delivery mechanism for our gallery browsing, broken links are highly undesirable.

The link structure in a first generation system is prescriptive − the only links available are those provided by the system designer, resulting in systems which are less able to adapt to the user. The lack of any semantics or types on the links can also lead to difficulties in interpreting the structure or defining how the hypertext should behave. Although a hard-wired static predetermined links structure has its advantages (particularly if the intention is to educate the user

or provide a guided trail through the collection), if it is the only navigational facility offered, it constrains the reader [12]. The designer has to "second-guess" the requirements of the user and consider all eventualities. The issue of link types is particularly important as objects may have different kinds of associations between them.

The major issue is scale. The difficulty of creating large numbers of manually-authored links limits the size of the hypertext. A concrete example is found in typical WWW public access systems that do not reflect the diversity of user perceptions, and where there is limited scope for locating information from a range of viewpoints. Categories are poorly interconnected – frequently, beneath a couple of categorisation layers, the information structure breaks down to simple lists, offering limited scope for following up an interest. Where keyword searching exists, the onus is on the user to determine suitable terms. If the request gives no answer, the user must determine which keywords to try next.

2.2   Second Generation Hypertext

Second generation hypertexts have attempted to address these shortcomings through the use of more principled authoring. In particular, there has been a move to separate the links from the documents, and specify the structure and behaviour in terms of a well-defined conceptual schema, typing documents and links. Information about the hypertext is now represented explicitly from the information in it. Links between documents can be derived by querying the schema, allowing richer and more elaborate associations.

A considerable amount of research has been applied to the automatic construction of hypertext structure. However, the time to supervise link creation for large and growing collections is prohibitive [21]. Moreover, too many links can overwhelm the reader.

Such a view can be taken to its limit, with all access to objects being provided through query rather than linking. This is more the view of the Information Retrieval (IR) community, where the system maintains a descriptor for each document or object that characterises its content. Queries consist of a composition of descriptors and the system returns those documents matching the query. Navigation from document to document is conducted entirely through queries. Link creation using IR methods has been applied to hypertexts to suggest links to authors, to inform retrieval algorithms to retrieve hypertext nodes, and to infer links among documents (e.g. [15, 16]). SuperBook [20] used keyword queries instead of static hypertext links as a navigation mechanism, although this relied on a table of contents interface metaphor that we consider inappropriate. See [11] for a list of references.

2.3   Implementation Strategy

Our approach employs a second generation architecture, with links derived from the content of the object descriptions, forming a document space and a link space of the kind described in [1, 6]. We create links automatically based on
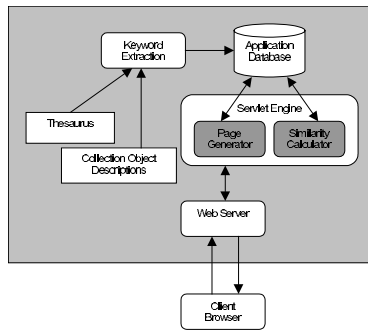
Fig. 1. System Architecture

the content of documents describing the artifacts in the costume collection and take the simple approach of extracting keywords (taken from a thesaurus) from those descriptions. The question "*what is there like this?*", is then answered through an examination of the extracted keywords and a comparison with other objects in the collection. Our experimental prototype is intended to test the hypothesis that such a straightforward approach will produce an adequate and usable browsing environment for casual exploration of the collections.

Manually authored links still play a part in providing browsing facilities. The system designers may wish to provide particular trails through the collection. Such a mechanism should be seen, however, as one of a number of complementary navigational devices, which can be employed and should be incorporated in the system.

3   System Architecture and Implementation

The system maintains a database of objects along with their descriptions and other catalogue data. In addition to the textual descriptions, we generate keyword annotations based on those descriptions. The annotations are then used to determine the "similarity" of objects in the collection.

Figure 1 shows the basic architecture of the system. The use of an HTML based front end will allow us to develop the application for WWW usage, providing remote access to the collection. In addition, with the growing public usage of the Web, familiarity with navigational mechanisms such as clicking on links or images to see further information is widespread. The database (stored as an XML document) contains basic information about each object in the collection as shown in Figure 2. Not all fields (such as donor or date) are present for each object.

Keywords are drawn from a thesaurus of terms related to costume. The thesaurus has a simple structure [2], including broader, narrower and related terms, and is based loosely on the ICOM Costume Classification [13]. A Keyword Analyser and Extractor tokenizes the descriptions applied to the objects,

Field	Purpose
id	Unique identifier
accession number	Gallery information
name	The object's name
date	Date information
donor	Where the item came from
picture	A link to an image of the object
links	Objects in the database considered to be closely related
description	A textual description of the object

Fig. 2. Database Fields

Fig. 3. Application Screenshots

applies stemming to extract and match terms against those in the thesaurus and annotates each entry with the terms that it finds.

Servlets generate HTML descriptions for particular objects, which are then delivered to the client browser. The results of a page generation are shown on the left of Figure 3. The page has a frame at the left providing access to the front pages and a tool bar along the bottom, which provides access to linked or related objects. For example, selecting the magnifying glass yields the page shown on the right of Figure 3. The system has found 10 objects matching the focus, and displays them in rank order (those with most matching keywords at the top). A list of the keywords applying to the object is shown, and users can use these to refine the list, in a sense combining the browsing and searching notion. However, this is not a straightforward search, but is a form of search by analogy, along the lines of Query-By-Example [23], where the focus object provides an initial query, which can then be refined.

4   User Evaluation

Evaluation of browsing is a difficult task. Standard quantitative measures such as recall and precision can be used to evaluate performance of search systems, but in our particular application, they are less appropriate. Instead, we conducted a qualitative evaluation, and assess the satisfaction that users felt when using the system through a questionnaire.
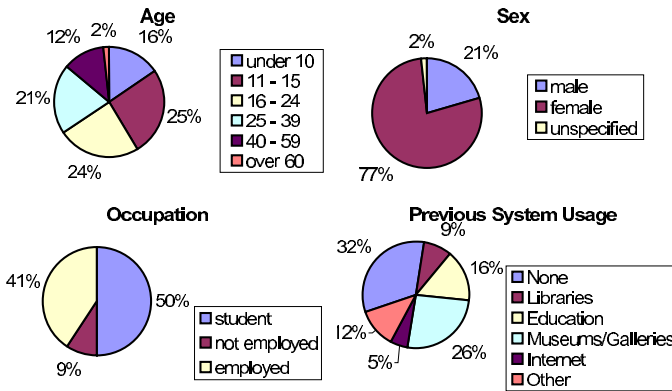
Fig. 4. User Demographics

4.1 Questionnaire

Users of the system were asked to fill out a questionnaire requesting their name, age, sex, occupation status (employed/unemployed/student), whether they had used any similar information systems in the past and if so, where. They were then asked to rate how the system compared to others they had used (if appropriate) and whether they found the system easy to use and useful. Responses to these latter questions were based on a seven-point scale. Additional information about the users was gained through interview and observation.

The trials were conducted at the Platt Hall Gallery over a three-week period. Platt Hall is located in South Manchester, close to the Universities and many local primary and secondary schools. The high student population in the area, along with the style of gallery, strongly influences the type of visitor. During the period of the trials it was the end of the school and University summer break.

The logs kept of each visitor's activities indicate that there were 155 users of the system during the 16 days of research over the 3 weeks. Of these, 7 can be deducted from a particular pair of users who exited and re-entered the system several time as a method of getting their bearings. Even if approximately 10% of sessions were due to gallery assistants re-use of the system, during consultation sessions, there were still around 133 valid users, of which 58 subjects returned completed questionnaires – around a 44% feedback rate.

Figure 4 shows the basic demographics of the sample population. The sample was predominantly female, and broad ranging in age from four years old to over 60. 50% of the subjects who answered the question about their occupation identified themselves as students with half (25% of total subjects) of those identifying their education as pre-A-Level. From the ages of the subjects, the total number of users under 16 is actually even higher (41%), suggesting that even more of the users were in full-time school education. From the other 50%
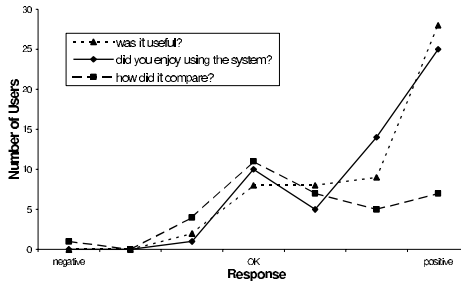
Fig. 5. User Satisfaction

of subjects, 9% of the total subjects were unemployed and 17% worked for City Art Gallery or other galleries.

The users of the system were reasonably typical of the normal visitors to the gallery, with a few exceptions. The gallery has a relatively high number of third age (over 60) visitors, few of whom were happy to use the system. In addition, the number of degree level students was lower than usual, due to the timing of the trials. Research by the Arts About Manchester project [3] into visitors to the City Art Gallery city site can be used to demonstrate strong similarities between our sample population and the population visiting City Art Gallery. Comparing the two populations they agree on the dominance of female visitors, a broad spread of age groups, predominantly young (often students), and far less third-agers. One noticeable difference is the larger number of family groups in our population, perhaps explained by the time of year.

Figure 5 shows the levels of user satisfaction recorded. Users gave a positive response to the system in terms of its usefulness, ease of use and comparison with other systems (where appropriate).

5   Discussion

The application described here makes use of some simple, well-understood techniques in order to deliver the required functionality. Our original requirements were better browsing support for the users; and better authoring support for the dynamic creation of browsing pathways.

The evaluation results suggest that browsing the hypertext generated by keyword linking was satisfactory. As opposed to the VOIR (Visualisation of Information Retrieval) system [11], we do not turn words above a frequency threshold into a hypertext anchor. Instead two anchors are associated with every artifact returned: the first returns a list of related artifacts; the second shows the most relevant unread artifact. Thus, we avoid the "over linking" problem whereby readers are overwhelmed by too many link choices. An alternative mechanism would be to filter anchors by some context-specific condition, but this adds to the complexity of the implementation. The returning of clusters of artifacts that

share common keywords, linked through a thesaurus, is a lightweight mechanism to support the search by analogy information seeking mechanism advocated by [19].

Better authoring support has been achieved by the creation of links purely through keyword associations. Dynamic link creation effectively removes the authoring task.

There are areas for improvement and further investigation.

5.1   Use of Thesaurus Structure

The similarity matching between object descriptions is achieved through the use of simple keywords. The structure in the thesaurus (broader, narrower and related terms), however, was not used for the purposes of this experiment, but could be used during the similarity calculations to improve the retrieval of "similar" objects. For instance, the fact that a tie and a scarf are both worn around the neck should lead to scarves and ties being more closely associated than, say a shoe and a scarf. Investigations are needed into how we might use the thesaurus relationships to calculate similarities, and whether such an approach provides significant improvement in the performance, or at least in the reported user satisfaction. The Semantic Hypermedia Architecture of Glamorgan [9] uses a thesaurus in this way to calculate similarities.

5.2   An Open Hypermedia Architecture

An Open Hypermedia architecture, such as the Distributed Links Server (DLS) [8] allows the integration of third party applications and documents into a system. The use of the DLS eases the process of constructing links through the use of link bases, which explicitly separate links from documents. Through the COHSE project, we are investigating the combination of an open hypermedia system with conceptual models (such as the thesaurus used in the annotation of collection objects).

5.3   Extending Coverage

The prototype is based on a small subset of items in the Platt collection. Currently, most electronic catalogue information in the gallery tends to be administrative rather than descriptive information. Extending the prototype will require the addition of more descriptive information in electronic form, which can be a labour intensive task. The successful reception of the prototype, however, suggests that this may be worthwhile.

References

[1] Agosti, M., Gradenigo, G., and Marchetti, P. G., *A hypertext environment for interacting with large textual databases*. Information Processing & Management, 1992. **28**(3): pp. 371-387.

[2] Aitchison, J. and Gilchrist, A., *Thesaurus construction - A practical manual*. 2nd ed. 1987, London: Aslib.

[3] Arts about Manchester, `http://www.aam.org.uk/`.

[4] Bates, M., *How to Use Controlled Vocabularies More Effectively in Online Searching*. Online, 1988. **12**(November): pp. 45–56.

[5] Bates, M., *Indexing and Access for Digital Libraries and the Internet: Human Database and Domain Factors*. Journal of the American Society for Information Science, 1998. **49**(13): pp. 1185–1205.

[6] Bullock, J. and Goble, C., *TourisT: The Application of a Description Logic based Semantic Hypermedia System for Tourism*, in Hypertext '98, Pittsburgh. 1998.

[7] Bush, V., *As We May Think*. The Atlantic Monthly, July 1945.

[8] Carr, L., et al. *The Distributed Link Service: A Tool for Publishers, Authors and Readers*. World Wide Web Journal, 1995. **1**(1): pp. 647- 656.

[9] Cunliffe, D., Taylor, C., and Tudhope, D., *Query-based Navigation in Semantically Indexed Hypermedia*, in Hypertext '97, Southampton, UK. 1997.

[10] Furuta, R., et. al, *Hypertext Paths and the World-Wide Web: Experiences with Walden's Paths*, in Hypertext '97, Southampton, UK. 1997, pp. 167-176.

[11] Golovchinsky, G., *What the Query Told the Link: the integration of hypertext and information retrieval*, in Hypertext '97, Southampton, UK. 1997, pp. 67–74.

[12] Halasz, F. G., Moran, T. P., and Trigg, R. H., *Notecards in a Nutshell*, in Proc. of ACM CHI 87, Toronto, Canada,. 1987, ACM Press. pp. 45-52.

[13] ICOM Classification, `http://www.open.gov.uk/mdocassn/costume/vbt00e.htm`.

[14] Ingwersen, P., *Cognitive perspectives of information-retrieval interaction - elements of a cognitive IR theory*. Journal Of Documentation, 1996. **52**(1): pp. 3–50.

[15] Kheirbek, A. and Chiaramella, Y., *Integrating hypermedia and information retrieval with conceptual graphs*, in Proc. HIM '95 (Konstanz). 1995, Universittsverlag Konstanz. pp. 47–60.

[16] Lucarella, D. and Zanzi, A., *Information-retrieval from hypertext - an approach using plausible inference*. Information Processing & Management, 1993. **29**(3): pp. 299– 312.

[17] Marchionini, G., *Information Seeking in Electronic Environments. Cambridge Series on Human-Computer Interaction*. Vol. 9. 1995: Cambridge University Press.

[18] Nelson, T. H., *Literary Machines*. 1981.

[19] Pejtersen, A. M., *A Library System for Information Retrieval based on a Cognitive Task Analysis and Supported by an Icon-based Interface*, in Proc. of SIGIR '89, Cambridge, MA, USA,. 1989, pp. 40–47.

[20] Remde, J. R., Gomez, L. M., and Lanauer, T. K., *SuperBook: an automatic tool for information exploration-hypertext?*, in Hypertext '87, San Antonio, Texas, USA. 1987, pp. 175–188.

[21] Robertson, J., Merkus, E., and Ginige, A., *The Hypermedia Authoring Research Toolkit (HART)*, in ECHT'94, Edinburgh, UK. 1994, ACM Press. pp. 177–185.

[22] Saracevic, T., Spink, A., and M-M. Wu In , *Users and Intermediaries on Information Retrieval: what are they talking about?*, in User Modelling: Proc. of the 6th International Conference UM97. 1997, Springer: New York. pp. 43–54.

[23] Zloof, M. M., *Query-by-Example*, in Proc. of the National Computer Conference. 1975, AFIPS Press: Montvale, NJ. p. 431–438.