Personalizing Access to Cultural Heritage Collections using Pathways

Paul Clough  
University of Sheffield  
Sheffield, S1 4DP, UK  
p.d.clough@sheffield.ac.uk

Nigel Ford  
University of Sheffield  
Sheffield, S1 4DP, UK  
n.ford@sheffield.ac.uk

Mark Stevenson  
University of Sheffield  
Sheffield, S1 4DP, UK  
marks@dcs.shef.ac.uk

ABSTRACT
This paper discusses mechanisms for personalizing access to cultural heritage collections and suggests that paths or trails are a flexible and powerful model for this and could link with existing models of cognitive information behaviour. We also describe a European project called PATHS (Personalized Access To cultural Heritage Spaces) that aims to support information exploration and discovery through digital cultural heritage collections. This project aims to implement the user models discussed in this paper and provide a mechanism for users to create and share pathways through information spaces for learning and knowledge discovery. Personalized access to digital cultural heritage resources will be provided by adapting suggested routes to the requirements of individual users and groups, such as students/teachers, professional archivists and historians and scholars.

Author Keywords
Adaptive systems, personalization, paths, user models, cultural heritage resources, social navigation.

INTRODUCTION
Significant amounts of cultural heritage material are now available through online digital library portals. However, this vast amount of cultural heritage material can also be overwhelming for many users who are provided with little or no guidance on how to find and interpret this information. Potentially useful and relevant content is hidden from the users who are typically offered simple keyword-based searching functionality as the entry point into a cultural heritage collection. The situation is very different within traditional mechanisms for viewing cultural heritage (e.g. museums) where items are organized thematically and users guided through the collection. Users of cultural heritage portals have diverse information needs and exhibit highly individualistic information seeking behaviours (e.g. information encountering and foraging) which are not well supported in standard search interfaces. Recent trends in information access services have recognized the necessity of providing support for more exploratory and serendipitous search behaviours if services are to be effective in helping users with discovering and assimilating knowledge [11, 12, 13].

This paper suggests the metaphor of “paths” through a collection as a powerful and flexible model for navigation that can enhance the user’s experience of cultural heritage collections and support them in their learning and information seeking activities. It also describes a new project funded by the European Union, PATHS (Personalized Access To cultural Heritage Spaces), that uses this model to provide users with innovative ways to access and utilize the contents of digital libraries that enrich their experiences of these resources. The PATHS project aims to create a system that acts as an interactive personalized tour guide through existing digital library collections, offering suggestions about items to look at and assist in their interpretation by providing relevant contextual information from related items within specific collections and items from external sources.

The remainder of this paper is organized as follows. We begin with a description of existing approaches that have been used to personalize navigation systems, including details of ways in which cognitive styles of users can be modeled. We then introduce the notion of a path through a collection which we suggest as a useful model of navigation. Finally, we describe the PATHS project which will make use of this model and apply it to a large on-line cultural heritage collection.

BACKGROUND
Cultural Heritage and the User Experience
Cultural heritage institutions hold an enormous and rich variety of digital content covering a broad range of subjects: natural history, ethnography, archaeology, historic monuments, fine and applied arts which often cross national and linguistic boundaries. There is strong motivation to bring together content from different cultural institutions into aggregated portals, which have typically offered access services based on traditional catalogues used in libraries,
museums and archives. Search services have been geared towards subject specialists and experienced users. Yet the environment in which users and digital library services are operating has changed. People come to digital libraries with experience of using the web and with new expectations [8].

Cultural institutions wish to be able to offer users of their portals with an experience that is continuous with the way people experience the web. They are seeking to enable richer user experiences that support connectivity between people, content and applications, to support writers as well and readers, and to enable collaborations with and between users. A new generation of cultural portals is encouraging user participation by offering people with opportunities to interact with content (for example encouraging them to tag resources) and to make recommendations to other users [6, 19].

Institutions are looking for ways to recreate in the digital information space the opportunities that visitors to libraries, museums and archives have of sharing books, objects and ideas with each other. They would like to be able to personalize the experience for their digital library users, for example suggesting content that is more likely to be of interest based on their profile and highlighting associations between related items [1]. But cultural heritage portals lack the massive volumes of users and interactions that are required to enable the effective analysis of user behaviours, usage patterns and interests. Issues related to user-adaptivity such as controllability, trust, privacy, obtrusiveness, predictability, transparency and breath of user’s experience are also important [14].

Adaptive Personalized Systems

A number of adaptive personalized systems have been developed, particularly to support learning. The Adaptive Personalized eLearning Service (APeLS) reported by Conlan et al. [9] can take metadata describing an individual’s learning needs, prior competences and personal characteristics, and construct a personalized course by discovering and sequencing appropriate learning objects discovered within one or more distributed specialist repositories. Input to a rule-based adaptive engine includes metadata relating to the learner, content in the form of learning resources, and narratives. A narrative is a description of the required conceptual pathway for a particular learner, and is built prior to searching for particular learning resources, taking account for example of prerequisite concepts and their sequencing. Thus, the system’s narrative model refers to concepts rather than specific resources and can be made to model the learner’s prior knowledge and preferences.

O’Keeffe et al. [16] note that although early adaptive hypermedia systems personalized learning according to learners’ prior knowledge, goals and preferences, they did not explicitly address pedagogy. They report the incorporation and further development of the APeLS system in iClass, which is a framework of services designed to support teachers and learners in providing personalized learning experiences. In terms of standards, OWL is used for domain ontologies; IMS Learning Design (IMS LD) for structuring learning activities; IMS LIP for learner profiles; and SCORM for learning object manifests. Like APeLS, iClass separates the generation of personalized conceptual learning paths from specific resources that can instantiate such paths. However, an important way in which iClass differs from APeLS is that pedagogy is explicitly represented in a separate model distinct from domain knowledge. The representation of pedagogies in a separate model means that they can be reused.

Castillo, Gama and Breda [7] report GIAS, a system also designed to select learning resources from a repository which are appropriate to individual learners’ current states of knowledge and their preferred learning styles. However, the system differs from others in that from a baseline initial psychometric assessment of learning style, it fine-tunes its model of each learner’s style in response to feedback from interactions between the learner and learning resources. This on-going adaptation is necessary because of inherent uncertainty in any psychometric assessment of learning style, and also possible drift in learners’ preferences over time and as a result of interactions with resources.

A number of systems are designed to enable resource discovery not only of materials stored in specialised repositories, but also open corpus material available on the Web. Dolog et al. [10], for example, report development of the “Personal Reader”, which is also based on metadata-based reasoning mechanisms. From the starting point of a learning resource being studied by a learner, the “Personal Reader” discovers resources that are related to the currently viewed resource. It can recommend resources which provide a different perspective on the topic – for example, a summary, more general or more specific material, or examples illustrating the concepts involved. The authors’ aim is to work towards the vision of an adaptive web capable of leveraging open corpus material [3, 4].

Social Navigation Systems

A number of social navigation systems have been developed that attempt to harness the collective knowledge of their user communities. As noted by Brusilovsky et al. [5], such systems use a variety of techniques, which they categorize in terms of:

- Types of past user behaviour recorded;
- How these data are used to generate collective knowledge;
- How this collective knowledge is used in order to benefit users in terms of their information accessing.

Brusilovsky et al. describe the use of social navigation in digital libraries as implemented in the Ensemble system. Users can follow “trails” of library items which are created on specific topics in the manner being adopted in our work.
An early system for the development of educational systems based on an “ecological” paradigm is reported by McCalla [2]. This is an approach to providing adaptive personalized learning which not only makes use of standard pedagogical metadata, but also dynamically collects and cumulates other metadata relating to resources, learners, and the interactions between the two, at the time of use. The approach also entails the mining of this data in order to generate new pedagogic knowledge as required for different purposes. Metadata is gathered at the time of use of the resource as opposed to being pre-assigned by human experts. As a learner accesses a particular resource, metadata is attached to that resource, which relates to the learner, the resource and the interaction between the two (e.g. dwell time on a resource). Over time, each learning resource will accumulate many models. This data can then be subjected to data mining in order to discover patterns that are useful in achieving particular tasks.

Tang and McCalla, [18] report a system based on this approach, designed to search Citeseer and recommend relevant papers to research students. Each time a paper is read by a learner, an instance of the learner’s model is attached to the paper’s metadata. The system also enables the learner to annotate the resource, this information feeding in to model (recording the learner’s interactions with resources). Work is reported in further developing approaches to the learner modelling that is a key feature of ecological systems [15, 20].

USER MODELS
Some existing systems employ explicit user models to drive adaptivity. Such models may entail, for example, levels of expertise or cognitive styles. Others make use of recorded user behaviour without any explicit underlying cognitive models.

User models may be utilized in both inductive and deductive approaches to system design. An inductive system is defined here as one in which the model driving adaptivity does not derive from any pre-existing user model, and is created by analyzing user behaviours and inputs. A deductive system is one in which an existing user model is used to generate adaptive behaviour. Deduction entails using pre-existing user models to driver adaptive behaviour (e.g. recommended paths and next moves). Induction entails gathering and using data generated by users (the navigational behaviour they exhibit, and the comments/reactions/annotations/tags they provide relating to nodes and paths).

These categories are not mutually exclusive: the two may be used interactively – to begin with parameters derived from known user models are likely to be effective, and allow the refinement and modification of these models via data from user behavior and inputs during system use, for example user-system interactions recorded in transaction logs.

Insofar as decisions are made as to which user behaviours and inputs are recorded and used as input to an adaptivity engine, such selection is based on at least an implicit model or theory of which factors are likely to be most influential in the calculation of appropriate adaptive system behaviour. A criticism of a number of the explicit or implicit user models employed by a number of existing adaptive systems is that the models selected are not those which are arguably best suited to drive adaptivity relating to the navigation of information spaces.

Although there are many models of both information seeking and learning via pedagogical mediation, few map directly onto the type of choices most relevant to an adaptive system designed to advise on paths through a complex set of information sources. Of particular relevance here is the work of Pask and Witkin [21, 22], the models of whom have been extensively studied for almost 40 years. This (and related) work suggests that cognitive processing takes place across two major orthogonal dimensions (Figure 1).

![Figure 1. Key cognitive dimensions.](image)

The work suggests that:

- Different individuals may have different predominant navigational styles. These appear to be linked to more fundamental cognitive styles. High academic achievers as well as less academically achieving people may still have a predominant style. These navigation styles translate into different paths.
- Adopting a navigation path that matches one’s predominant style can influence the effectiveness of the resultant learning. Where an individual navigates using a path that mismatches their predominant style their learning may be disrupted; matching a navigational path may enhance learning (for a given learning task). However, these results have been found in experimental rather than more natural learning conditions.
- Navigation paths adopted by individuals may also vary according to their level of subject expertise in the area being navigated.
- Different paths may also be more, and less appropriate for achieving different types of goal/task (e.g. relatively convergent fact-finding versus more divergent creative exploration).
• Individuals also vary in the extent to which they thrive in navigational conditions characterised by external mediation (guidance) versus autonomy. This difference appears to be linked to fundamental cognitive style.

Table 1 shows characteristics, as identified in this work, of the horizontal (local/global) dimension in terms of a number of aspects of cognitive processing relevant to users’ navigation of a semantic space.

<table>
<thead>
<tr>
<th>Local (analytic)</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning/problem-solving goals</strong></td>
<td></td>
</tr>
<tr>
<td>Convergent goals.</td>
<td>Divergent goals.</td>
</tr>
<tr>
<td>“Find an answer”.</td>
<td>Creatively explore.</td>
</tr>
<tr>
<td>Learn pre-defined content.</td>
<td>Come up with new ideas.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process goals</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concerned with procedures and vertical deep detail (procedure building).</td>
<td>Concerned with conceptual overview and horizontal broad inter-relationships (description building).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Navigation styles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Serialist navigation style</td>
<td>Holist navigation style</td>
</tr>
<tr>
<td>Narrow focus.</td>
<td>Broad global focus.</td>
</tr>
<tr>
<td>One thing at a time.</td>
<td>Many things on the go at the same time.</td>
</tr>
<tr>
<td>Short logical links between nodes.</td>
<td>Rich links between nodes.</td>
</tr>
<tr>
<td>Intolerance of strictly irrelevant material.</td>
<td>Welcoming of enrichment (but strictly irrelevant) material.</td>
</tr>
<tr>
<td>Finish with one topic before going on to the next.</td>
<td>Layered approach returning to nodes at different level of detail.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive learning outcomes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Good grasp of detailed evidence.</td>
<td>Well developed conceptual overview.</td>
</tr>
<tr>
<td>Deep understanding of individual topics.</td>
<td>Broad inter-relationship of ideas.</td>
</tr>
<tr>
<td>In-depth understanding of the parts.</td>
<td>Good grasp of the “big picture”.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic learning pathologies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor appreciation of topic inter-relationships.</td>
<td>Poor grasp of detail.</td>
</tr>
<tr>
<td>Failing to see the “big picture”.</td>
<td>Over-generalisation.</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of local and global information processing styles.

**Scenarios of use**

**Global**
The stereotypical global user will want to explore a new topic broadly at first, first establishing a relatively tentative/provisional conceptual overview before going back to nodes and filling in details. S/he will take a broad focus, sampling nodes at different levels taking a “layered” approach. S/he will have a high tolerance of uncertainty and ambiguity, and will – relative to his or her local counterparts – be happy to spend quite a long time in tentative exploratory uncertain mode. The more extreme global person is likely to be a divergent (creative) thinker and will want to be stimulated by new connections. S/he will want to explore widely spaced nodes – not keep to close and strictly logical connections. S/he may be looking for new ideas/directions “off the beaten track”.

**Local (analytic)**
The stereotypical local (analytic) user will be concerned to learn in a step by step way, mastering one aspect of the topic (node) in depth before going on to the next. The “big picture” will tend to emerge relatively late in the learning process. S/he will find distracting information which the global information processor might consider valuable enrichment material that would enhance learning. S/he is likely to be a relatively convergent thinker who knows what s/he wants and does not want to be distracted by anything “irrelevant”. S/he will want to explore step by step in a “one thing at a time” mode, with a relatively narrow focus. At any given point in time, the optimal next node to be explored will be local (the next logical step). S/he will have a low tolerance of uncertainty and ambiguity, and will not like to spend long in tentative exploratory uncertain mode.

**Dependent**
The dependent learner may be new to the subject area, and may want to follow a path (whether global or local) that has been approved and has been pre-trodden by experts and other authoritative figures.

**Autonomous**
The more autonomous learner may want to beat his/her own exploration path, and may want to (make and possibly be stimulated to make) more original less conventional connections between nodes.

**PATHS AS A METAPHOR FOR NAVIGATION**
The notion of a “path” or “trail” through digital library collections provides a flexible model of navigation that can provide a powerful model onto which various levels of personalisation can be added. In hypertext systems, the notion of guided tours and trails refer to scenarios when users are introduced to an unfamiliar subject and activities carried out with individual steps [17, 22].

A path is defined here as a route through a semantic space. In this case, the semantic spaces are defined digital collections of heritage resources. A “route through” a semantic space only makes sense in terms of a sequence of processing of connected components of that semantic space. The components of semantic space that enable sequential processing are defined here as nodes. The semantic space is navigable by the sequenced selection of connected nodes.
Figure 2. Stereotypical “local/analytic” path. Characteristics: small steps between nodes in the sequence, mainly depth-first; infrequent re-visiting of nodes. Logical “one step at a time” brick-by-brick approach.

Figure 3. Stereotypical “global” path. Characteristics: large steps between nodes; significantly breadth-first; frequent re-visiting of nodes attending to different aspects in a layered “parallel processing” approach. Frequent accessing of “enrichment” material if offered at any node.

Nodes for a resource or information item might contain the following information:

- System-provided subject metadata (subject keywords) about this resource;
- User-provided subject metadata (tags) for this resource;
- Broader, narrower and related terms according to the default (“authority”) subject ontology for the topics that this resource is about;
- Recommended “next” nodes(s) according to the pedagogical ontology for the topic(s) the resource is about;
- Information about the resource at abstract level (e.g. overview, summary, image thumbnail etc.);
- Inter-relationships showing how the information stored at this node inter-relates with other nodes to form the bigger picture;
- Link(s) to the detailed contents of this resource (e.g. full text, detailed images etc.);
- Information on which ontologies are available for the topic(s) this resource is “about” – both subject and pedagogical;
- User annotations about this resource (this user and others);
- Information on paths that are available relating to the topic(s) this resources is “about” – both subject and pedagogical;
- Information on available paths in which the current resource forms a node.

Nodes may be connected in a number of different ways. They may be pre-defined or computed on the fly. They may be defined by the system/designers and/or by users – individual or aggregated. A node consists of items of information aggregated and represented by means of one or shared tags (metadata). A route is a sequence in which one or more nodes are attended to. Attending to a node may entail intellectual processing of different types and at different levels. Thus a user may return to previously “attended to” nodes and attend to it in a different way and/or at a different level in order to increase (clarify/modify/deepen/modify/refute/etc.) his or her understanding.

Nodes for a resource may link to a variety of items:

- Subject ontologies, e.g. a generally agreed default “authority” ontology;
- Pedagogical ontologies that contain information on, e.g. which topics are pre-requisite for the understanding of others;
- User metadata relating to goals, preferred navigation styles, levels of topic knowledge etc.;
- Information on interactions in which this node has been involved including user goals in using it, its part in any paths, user annotations (including user-perceived usefulness of this node and ultimate success of the path of which is formed a part);
- Information on the user’s current path;
• Information on other available paths, with their own metadata for describing for whom/which learning goals/at which level they may be appropriate.

There are potentially alternative routes to the same destinations. A destination is defined here in terms of a user’s desired state of knowledge or understanding. This may be defined by the user and/or someone else (for example, a teacher and/or subject expert). Different users may have different starting points for similar destinations. By way of example, Figures 2 and 3 illustrate stereotypical local and global pathways through a set of nodes that are arranged in a hierarchical order, which represent a typical hierarchical subject representation, with top level topics subdivided into lower order subordinate topics. The numbers represent the order in which the imagined “local” and “global” users might navigate the nodes. It is proposed that:

• Users can construct their own paths (“independent paths”) which can be saved for future reference, edited or shared with other users. These paths will be more than a simple list of items in a collection that the user has visited; they will also contain information about the links between the items, details of others items connected to them and connections to information, both within and outside the collection that provides context.

• Groups of users can work collectively to create paths (“collaborative paths”), adding new routes of discovery and annotations that can build upon the contributions made by others.

• User-generated paths will exist as information objects in their own right: they can be indexed, organised and shared with others, and will be potential learning objects that can be offered to users alongside the cultural heritage content.

• Users can also follow pre-defined paths (“guided paths”) created by domain experts, such as scholars or teachers. Guided paths provide an easily accessible entry point to the collection that can be either followed in their entirety or left at any point to create an independent path. Guided paths can be based around any theme, for example artist and media (“paintings by Picasso”), historic periods (“the Cold War”), places (“Venice”), famous people (“Muhammed Ali”), emotions (“happiness”), events (“the World Cup”) or any other topic (e.g. “Europe”, “food”).

• Users’ paths can be private and only accessible to them or made public and stored in a common (searchable) repository of organised paths on different themes and subjects.

Different users will have differing needs from pathways and an important aspect of the research will be to build up knowledge and understanding of user profiles, their needs, interests and preferences and feed this into the system development. It is expected that both explicit information input from the user will be utilized (e.g. in the form of a profile to be completed) and implicit information from users’ interaction (e.g. from transaction log files) will also shape their information seeking experience.

THE PATHS PROJECT

The PATHS project aims to make use of current knowledge of personalisation to develop a system for navigating cultural heritage collections that is based around the metaphor of paths and trails through them. The PATHS system will provide an adaptive and rich information-seeking encounter for the user. It system will make user-specific recommendations about items of potential interest as individuals navigate through the collection. The user will be offered links to information both within and outside the collection that provide contextual and background information, individually tailored to each user and their context.

The PATHS project will take a user-centred approach to development by bringing users into the research cycle from the beginning of the project, gathering their input at all stages in the development on how it can help to meet their needs and feedback on the functionality as prototypes are field-tested. The PATHS project consists of several separate, but connected, packages of work, including the following:

• Gathering user requirements and creating functional specifications from a broad range of users including those belonging to different groups, e.g. students, family historians and photographers and of different types, e.g. learning styles and needs from Cultural Heritage collections. These requirements will be used to develop a functional specification for the systems developed during the project. These requirements will build upon those identified in previous work for cultural heritage information access systems [24, 25].

• Processing cultural heritage content and enriching it through identifying connections between items within a collection and complementing connections with existing relations and providing links to material both within and external to the collection that provides background information (e.g. to Wikipedia).

• Designing effective user interfaces through which users will interact with the PATHS system. These interfaces will provide users with personalised navigation through Cultural Heritage collections that is enriched with the additional information added through processing the digital content. The user interface will allow users to follow pathways created by other users and to share their own. This will build on previous work on personalisation in museums and digital libraries [26, 27].
• Designing evaluation methodologies and conducting of field trials to assess the performance (effectiveness, efficiency and satisfaction) of the systems implemented in PATHS in realistic scenarios. Evaluation will culminate in field trials in end-user scenarios. Particular focus will be on evaluating users’ search sessions1 and the value of paths generated by the user. Also, focusing on the evaluation of browsing techniques will form part of this research.

The vision is to build a system that:

• Exploits existing knowledge of users to optimise the effectiveness of interacting with digital heritage resources.
• Enables the testing and refinement of such knowledge.
• Enables new knowledge to be discovered.

This system will provide personalized access to resources by adapting suggested routes to the personalized requirements of individual users and groups. It will seek to:

• Respond to users in a cognitively ergonomic way – i.e. by matching navigation to a learner’s preferred style and minimising any mismatch and consequent additional cognitive processing load. In this way, the learner will find exactly what s/he wants with the least effort. Navigation entails travelling the shortest path between starting point and desired end point.
• Challenge and “stretch” the user by via controlled and constructive mismatching. In this way, learners may develop increasing autonomy and versatility – i.e. the ability autonomously to thrive in information environments not necessarily matched to their own preferred style. PATHS will also explore the extent to which users may be encouraged and helped to engage in cognitive processing in which they are less strong. For example, the extreme divergent thinker may usefully be encouraged (in certain learning circumstances) not to underplay complementary convergent processing. Cognitive research suggests that s/he may, without such complementary processing, exhibit over-generalisation and lack of grasp of detail. Conversely, the strong convergent thinker may be encouraged to explore and think more divergently (creatively) to avoid fragmented learning and failing to see the wood for the trees. PATHS will explore the potential of suggesting sub-optimal, but constructive paths to users.

CONCLUSIONS
With growing amounts of digital cultural heritage information becoming available online users will require assistance, particularly for more exploratory information seeking tasks. We suggest that the notion of paths through collections forms a powerful metaphor onto which personalisation can be added that will map onto existing models of user’s information behaviour. Pathways can be used to support various styles of cognitive information processing which will surface as different routes users may take through an information space. Offering users suggested routes or paths through an information space will locate and associate information that will ultimately help them to fulfill broader activities, such as constructing knowledge around a given subject or theme.

The PATHS project aims to investigate and implement pathways in a naturalistic setting for a range of users and groups that regularly make use of cultural heritage information. A large-scale operational system will developed for navigating on-line cultural heritage collections in a more effective manner than current searching functionalities. Pathways will be used to guide and assist individuals and user communities with information discovery and exploration within cultural heritage information spaces for learning and information seeking activities. This will support multiple information seeking behaviours and enhance the user’s information access experience of digital library resources.

ACKNOWLEDGMENTS
The work in this project is supported by the European Commission, grant number 270082.

REFERENCES

---

1 We began work on the evaluation of search sessions through organising a session at the US Text RETrieval Conference (TREC) in 2010 called the Session Track: http://ir.cis.udel.edu/sessions/


