

Semantic Browsing of Digital Collections

Trevor Collins, Paul Mulholland, and Zdenek Zdrahal

Knowledge Media Institute, The Open University, UK
{t.d.collins, p.mulholland, z.zdrahal}@open.ac.uk

Abstract. Visiting museums is an increasingly popular pastime. Studies have shown that visitors can draw on their museum experience, long after their visit, to learn new things in practical situations. Rather than viewing a visit as a single learning event, we are interested in ways of extending the experience to allow visitors to access online resources tailored to their interests. Museums typically have extensive archives that can be made available online, the challenge is to match these resources to the visitor's interests and present them in a manner that facilitates exploration and engages the visitor. We propose the use of knowledge level resource descriptions to identify relevant resources and create structured presentations. A system that embodies this approach, which is in use in a UK museum, is presented and the applicability of the approach to the broader semantic web is discussed.

1 Introduction

This paper presents an approach that exploits the use of semantics to create and present online digital collections for museum visitors to explore after their visit. Learning from museums is an example of what has been referred to as “free-choice learning” [1]. Other examples include learning “when watching television, reading a newspaper, talking with friends, attending a play, or surfing the internet” [1]. Characteristically, the learning that occurs is a product of how we choose to spend our playtime rather than the product of a formal education or job activity. As noted in [2] “play is not just mindless entertainment, but an essential way of engaging with and learning about our world and ourselves - for adults as well as children.”

1.1 Background

In a survey of visitors to a museum web site Kravchyna and Hastings [3] found that 57% of respondents visited a museum's web site both before and after visiting the museum. Yet, other than the hours of business, admission price and travel advice, little information is tailored to the pre- or post-visit reader. For someone who has never visited the museum, this essential information is very valuable, but for those who have already been, additional information can be made available that enables the visitor to build on their museum experience and in doing so encourage return visits. Falk and Dierking [1] reported finding extensive evidence that showed how visitors are able to draw on their museum experience, long after their visit, to construct new knowledge when they come across applicable situations in their everyday lives.

Rather than considering the museum visit as a one off event, the web offers an opportunity for visitors to continue learning when they get home [4].

1.2 Search

Rose and Levinson [5] proposed a hierarchical framework for classifying the goals of search engine users. They suggested that the purpose of an internet search can be broadly categorized as either 'navigational,' 'resource' or 'informational'. A navigational search is to locate to a known web site. Resource searches are used to obtain a resource available on the web. The purpose of informational searches is to learn something new by reading or viewing web pages. Across three samples of approximately 500 Alta Vista queries an average of 13.50% were categorized as navigational, 24.57% as resource, and 61.93% as informational. Although the navigational goal is least common, it is the best supported by traditional search engines [6].

To support exploratory informational searches several researchers have investigated the use of categories for presenting search results [7], [8]. Dumais, Cutrell and Chen [7] found in a series of four experimental studies, each involving between 18 and 20 people, that category interfaces were more effective than list interfaces. From a two month longitudinal study of sixteen users' search behaviors, Kaki [8] found categories were more beneficial than lists when more than one result was sought.

Guha, McCool and Miller [9] make the case that informational searches can significantly exploit the semantic web. Specifically, they showed how semantics can be used to augment the results of traditional information retrieval search techniques and to improve the text retrieval part of a search engine by identifying the context of the concepts denoted in the search phrase. We are particularly interested in identifying information related to a museum visitor's interests that can be drawn together to form a collection of resources for them to explore when they return home.

1.3 Collections and Narratives

An organized collection of objects forms a narrative that expresses relationships across the included items [10]. For example, a curated exhibition of paintings constitutes a narrative that expresses a story across the selected works. Here we are interested in forming a digital collection from a set of resources (according to the knowledge level description of those resources) and creating a series of hypertext presentations to support the reader's exploration of the resources in the collection. Exploring a collection involves the active interpretation of the included objects and the discovery of relationships between them.

Several systems for generating narratives have been developed recently in the cultural heritage domain. For example, Geurts, Bocconi, van Ossenbruggen and Hardman [11] present an ontology-driven approach for producing multimedia presentations. For a given query (such as 'life and work of Rembrandt'), presentation

genre (such as ‘biographies’ or ‘curriculum vitae’) and presentation medium (such as ‘screen’ or ‘paper’), a multimedia presentation is created through a two step process. A semantic graph, produced in response to the user query, is first of all transformed into a structured progression according to the chosen genre. The media items that represent the concepts identified in the structured progression are then retrieved to produce a multimedia presentation in the chosen medium. Within the Artequakt system [12] information extraction procedures are used to populate a knowledge base. Facts in the knowledge base are then used to fill in a predefined presentation template to produce a narrative. For example, information regarding an artist’s place and date of birth and date of death can be used to complete the opening sections of a biographical template.

Rather than building a coherent narrative from relevant facts (taken from a semantic graph or information extraction procedures) we are interested in building collections out of units that are meaningful in their own right, that is, the resources being used are lexia [13]. Examples include textual story passages, meaningful video clips and paintings. This avoids the problem of creating low-level narrative coherence, but raises the challenge of ensuring coherence across a collection of resources. The approach taken in response is to ensure the relationships between the included resources are clear and transparent to the user. The Topia [14] and Noadster [15] systems apply a concept lattice clustering approach to hierarchically group components that share characteristics. In the Topia system this is applied to group media components in order to generate hypermedia presentations. Similarly, we are interested in creating structured presentations that identify groups of related resources and show the relationships between resources in an online collection.

1.4 Summary and Overview

In summary, we are interested in applying semantic web technologies to support free-choice learning. We believe the semantic web offers an opportunity for museums to extend their visitors’ museum experience in a rewarding manner that can be used to encourage future museum visits.

The following section explains the ontologies we use for describing heritage resources and our approach for retrieving and presenting resources related to a given set of concepts. Section 3 describes how these approaches were put together to form the Bletchley Park Text system, which is now in daily use at Bletchley Park, a museum in the UK. Several examples of how the system is being used to explore a collection of resources are given in Section 4. The limitations of our approach and its potential application across the semantic web are discussed in Section 5. Section 6 summarizes what we consider to be the primary costs and benefits of adopting a semantic approach to support the exploration of digital collections.

2 Approach

We can explain the approach we took by first describing the context of the work, the ontologies used to represent the museum resources, and the method used to represent

the key information taken from the tour guides' presentation. We then describe how a set of resources are retrieved to produce a collection of related resources, and how these are organized into structured presentations to support the exploration of the relationships, which are used to connect concepts across the resource archive and form groups within the visitor's collected set of resources.

2.1 Bletchley Park Museum

This work has been carried out in collaboration with Bletchley Park Museum. The Park was the headquarters of the British Government Code and Cipher School during the Second World War. The Park was closed in 1945 and did not become a museum until the early 1990s after decades of secrecy. Since then the museum has been piecing together much of its hidden history. Several of the original buildings are still there and a number of them have been restored. The museum includes a range of exhibits that seek to explain the life of the people that worked there, the significance of the work they carried out, and how this pioneering work shaped modern computing and communication technologies.

2.2 Representing Resources

Bletchley Park Trust provided us with a set of transcripts of interviews with people that worked in the Park during the Second World War and a set of historical accounts of the activities related to the work of Bletchley Park for each month during the war. The knowledge level description of these resources was created using three distinct ontologies: the CIDOC Conceptual Reference Model (CRM), a Story and Narrative ontology (created as part of the Story Fountain system [16]), and a Bletchley Park domain ontology. CIDOC is the International Committee for Documentation of the International Council of Museums. Their CRM is a high-level ontology for describing cultural heritage objects and events, and is currently being considered as a potential ISO standard [17]. The story and narrative ontology was used to represent the historical accounts and first person interviews that make up a significant portion of the archive. The story and narrative ontology follows structuralist theories of narrative in distinguishing between a story (i.e. the conceptualization of what is told) and a narrative (i.e. how that story is told and what media is used) [18].

An example of the type of metadata used to describe a story is given in Figure 1. A story is represented as having any number of central actors (i.e. the main people or groups in the story), existents (i.e. the main physical objects), themes and events. Each event is described as having actors, existents, locations and a time specification. Depending on the type of event, existing properties were specialized or additional properties added. For example, an interview event had an interviewer and interviewee, and a creation event had a creator and an object of creation. Currently, the archive contains over 400 stories, which refer to over 1,700 distinct concepts, these include approximately: 450 people, 250 groups, 500 places, 200 physical objects, 300 conceptual objects and 50 work roles.

Sample resource:

Margot McNeely and Diana Lauder interview

My name is Margot McNeely; I was 17 and a half a schoolgirl in Burnham when I decided to volunteer for the WRNS that was in 1944. I took my school certificate and done some basic training then had lots of interviews. I don't know how they chose us, we were sent down here not knowing what we were going to be doing. ...

OCML knowledge model:

```
(def-instance margot-mcneely-interview story
  ((describes-event mmc-event)
   (describes-existent bombe)
   (has-associated-narrative margot-mcneely-narrative)
   (has-central-actor margot-mcneely)
   (has-theme life-and-work-in-bletchley-park)))

(def-instance mmc-event bletchley-park-life-experience
  ((has-actor margot-mcneely wrns)
   (has-billeting-location woburn-abbey crawley-grange)
   (has-work-location hut-11)
   (has-working-object bombe)
   (has-working-role bombe-operator p5)
   (is-described-in-story margot-mcneely-interview)))

(def-instance margot-mcneely-narrative cipher-digital-narrative
  ((has-associated-story-object margot-mcneely-interview)
   (has-uri "http:// ... ")))
```

Fig. 1. An extract from a sample resource and an example of the type of metadata used to describe it

2.3 Representing the Museum Tour Experience

While visiting the Park visitors are given a guided tour of the grounds and about told the history of the Park. A knowledge level description of the key facts given in the tour was stored as a set of fact triples in the knowledge base. These facts are not necessarily mentioned in the interviews or historical accounts, but provide useful background knowledge that is helpful when reading the resources.

Within educational hypertext, the concepts of vertical and horizontal navigation are used to describe the types of links within a hypertext [19]. Vertical navigation refers to traversing the hierarchical structure of a hypertext from parent to child and child to parent, whereas horizontal navigation links associated pages across the content hierarchy. Horizontal navigation is particularly beneficial within educational hypertext for referring the reader to related examples, counter examples, and sources of related information.

In our approach the fact triples are used to produce horizontal links that highlight relationships between groups of resources. For example, during the tour of Bletchley Park visitors are told that Alan Turing was the head of Hut 8. This is represented by a fact triple where ‘Alan Turing’ is an instance of a person, ‘was head of’ is a relation, and ‘Hut 8’ is an instance of a place within Bletchley Park. This additional fact is then used to provide a horizontal link between groups of resources relating to Alan Turing and groups relating to Hut 8. These horizontal links provide a navigational aid and serve to reinforce the story of Bletchley Park told by the tour guides.

2.4 Retrieving Related Resources

The visitor to the museum can identify any class, slot or instance as a concept of interest. For example, they may be interested in Alan Turing (an instance of a person), the places where people lived referred to as billeting locations, (an example of a slot), or a broader interest in decryption machines (an example of a class of machine). For a given query such as ‘Alan Turing, billeting locations and decryption machines’ the relevant stories are those where the knowledge level description refers to the instance Alan Turing, the slot billeting location, or any instance of the class ‘decryption machine’ (such as the Bombe or Colossus). Logical OR is used to identify all of the related stories. The combined set is referred to as the visitor’s collection.

2.5 Identifying Connections Between Concepts

Connections can be found when they exist between any pair of concepts (i.e. instances, slots or classes) by applying a path finding algorithm. The algorithm connects concepts in the archive using the slot values in each story event. For example, if one story explains that Alan Turing invented the Bombe, represented by an event of type ‘birth’ with a value of Alan Turing in the ‘has-actor’ slot and a value of Bombe in the ‘has-recipient’ slot, then this would constitute a connection between Alan Turing and the Bombe decryption machine as explained in that story.

For connections involving more than one story, common slot values are used to connect them. For example, to connect Alan Turing and Block G (one of the locations within Bletchley Park) two stories are required, one story describes that Alan Turing invented the Bombe and another explains that the Bombe machines were used in Block G. In terms of the knowledge level description, the first story includes an event of type ‘birth’ that identifies Alan Turing as the value of the ‘has-actor’ slot and the Bombe machine as the value of the ‘has-recipient’ slot, the second story includes an event of type ‘bletchley park life experience’ with a value of Bombe in the ‘has working object’ slot and a value of ‘Block G’ in the ‘has work location’ slot.

Within the archive there are over 1,700 concepts, and all concept pairs can be connected within a maximum of seven stories. The most common path length involves just three stories.

2.6 Identifying Categories Within a Collection

To group stories in a collection, categories are formed according to the most frequently used slot-value pairs (see Table 1). Two forms of categories are produced: a flat list of categories and a hierarchical list of categories. A flat list is simply a set of categories ordered by frequency, such as ‘has actor Alan Turing (three stories), has actor Winston Churchill (two stories), has actor John Tiltman (two stories), has actor Frank Birch (one story), and billeting location Woburn Abbey (one story)’. The same story can be included in multiple categories. To form a hierarchy of categories the same approach is used recursively within each category, down to a minimum category size. In the example given in Table 1, the category formed for ‘has actor Alan Turing’ contains three stories (i.e. story 1, 4 and 5), and within this category a sub-category can be formed containing the two stories that include ‘has actor Winston Churchill’ (i.e. story 1 and 4). The nested category includes stories that contain all of the parent slot-value pairs (i.e. ‘has actor Alan Turing’ AND ‘has actor Winston Churchill’).

Table 1. An example of the type of frequency data used to form a set of categories within a collection of stories. In this case the most common category would contain the stories that describe the activities of Alan Turing.

Slot-value pairs		Stories (an illustrative sample of five)					Frequency
Slot	Value	Story 1	Story 2	Story 3	Story 4	Story 5	
has actor	Alan Turing	Y	N	N	Y	Y	3
has actor	Winston Churchill	Y	N	N	Y	N	2
has actor	John Tiltman	N	N	Y	N	Y	2
has actor	Frank Birch	N	Y	N	N	N	1
billeting location	Woburn Abbey	N	Y	N	N	N	1

3 Bletchley Park Text

The knowledge level description of the resources and tour experience were developed using the Apollo knowledge modeling application [20]. The resulting knowledge model was exported as OCML [21] and placed on a Lisp server. A reasoning engine written in Lisp was used to produce story collections for any given set of concepts, generate pathways connecting concepts, and identify categories within a story collection.

Museum visitors express their interests using SMS text messages. Around the museum additional labels have been posted in locations and on exhibits to identify concepts of interest. One or more messages can be sent by a visitor, which are automatically downloaded from a mobile phone and stored in a database using commercially available text messaging software (i.e. SMS Demon available from www.dload.com.au). When a visitor enters their mobile phone number to log onto the

web site, the concepts identified in their messages are used to query the knowledge model and identify a collection of related stories. Alternative mobile technologies for recording the visitor’s interests, such as Radio Frequency Identification (‘RFID’) tags and readers, and location tracking Personal Digital Assistants (‘PDAs’) were also considered. However, the cost to the museum of providing and maintaining these technologies and the effort involved for the visitor to learn how to use them was prohibitive for our purposes.

Bletchley Park Text was built using the Story Fountain system [16]. Story Fountain was designed to support the investigation of questions and topics that require the accumulation, association or triangulation of information across a story archive. The underlying architecture of the Story Fountain is shown in Figure 2 along with the addition of the mobile phone service used to create the Bletchley Park Text application. An Apache web server coordinates the delivery of the site and uses the ModPython and ModLisp modules to access the presentation module (i.e. Python) and the knowledge module (i.e. the OCML knowledge model).

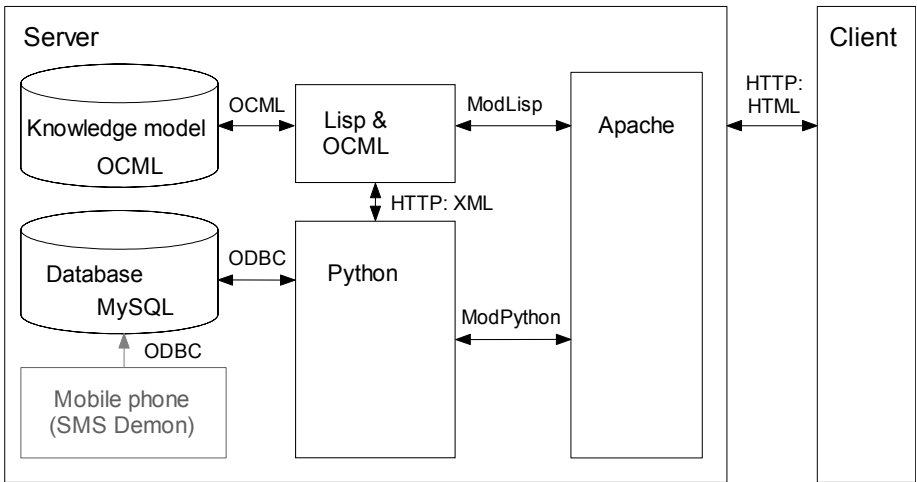


Fig. 2. The system architecture of the Bletchley Park Text application

Following a series of pilot trials the text application has been in daily use by visitors to Bletchley Park since May 2005.

4 Examples

After sending a text message the visitor can access the web site by entering their mobile phone number (see Figure 3, left). Six areas are available for the visitor to explore: stories, connections, categories, hierarchy, spotlight and modify (Figure 3, right). The first five present the collection of stories in different formats, and the sixth enables the visitor to change their selected set of concepts and thereby modify their story collection.

The ‘stories’ area presents all of the related stories from the archive (see Figure 4, left). In the five presentation areas each story is represented as a preview containing the title, an image (if available) and the first few lines of text (if available). The visitor can view any single story by clicking on its title (see Figure 4, right). The ‘stories’ area initially presents the collection as a single set, which the user can reduce by choosing one or more of their concepts as a filter. For example, by selecting Alan Turing as a filter concept only the stories relating to Alan Turing will be displayed (see Figure 5, left).

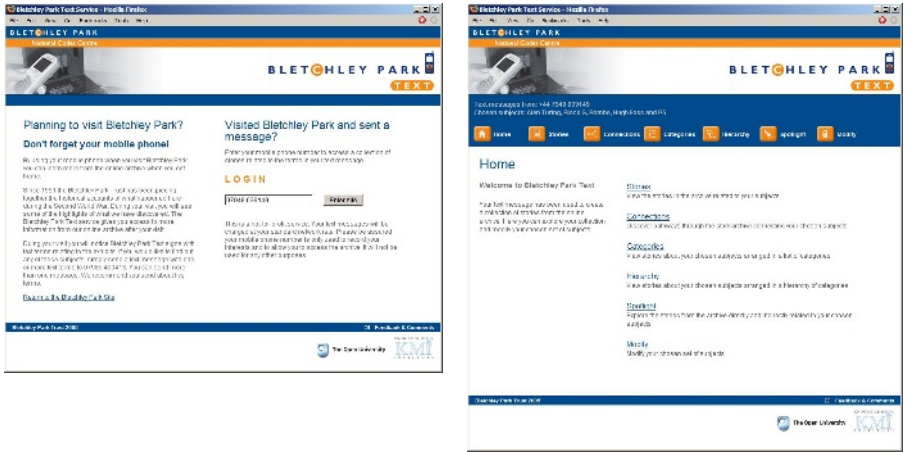


Fig. 3. The Bletchley Park Text login page (left), and the home page (right)

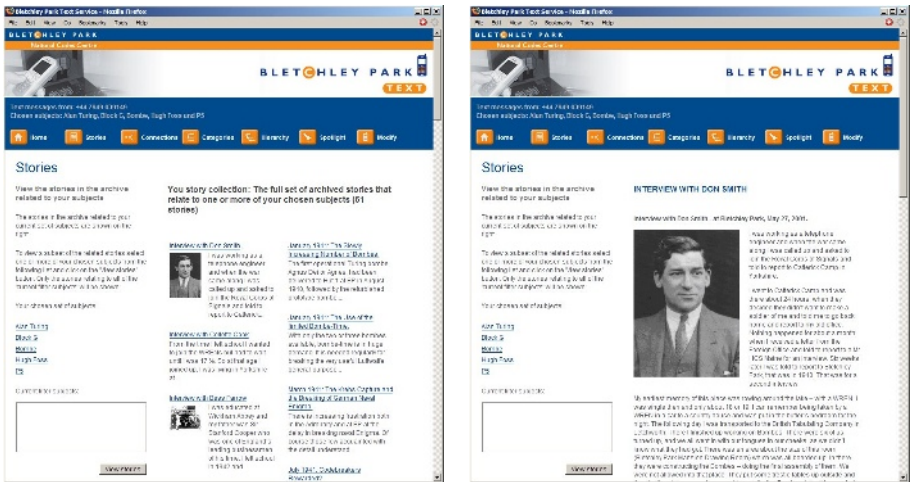


Fig. 4. The ‘stories’ area showing all of the stories in the archive related to the visitor’s chosen set of concepts (left), and a subset of the stories relating to a specific concept, Alan Turing (right)

The ‘connections’ area allows the visitor to chose two of their concepts as start and end nodes, and presents a connection extracted from the knowledge model (see Figure 5, right). The first two of the visitor’s concepts are used initially by default.

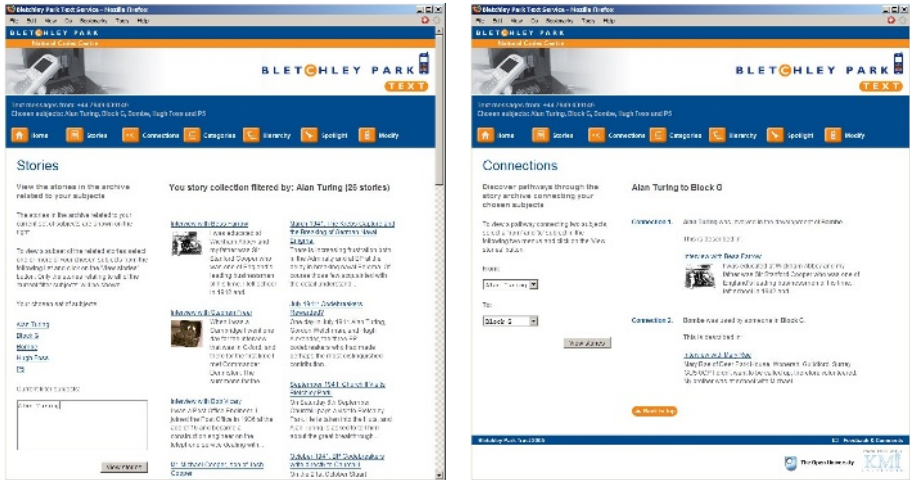


Fig. 5. An example of an interview story being displayed in the stories area (left), and an example of a pathway using two stories to link Alan Turing to Block G shown in the ‘connections’ area (right)

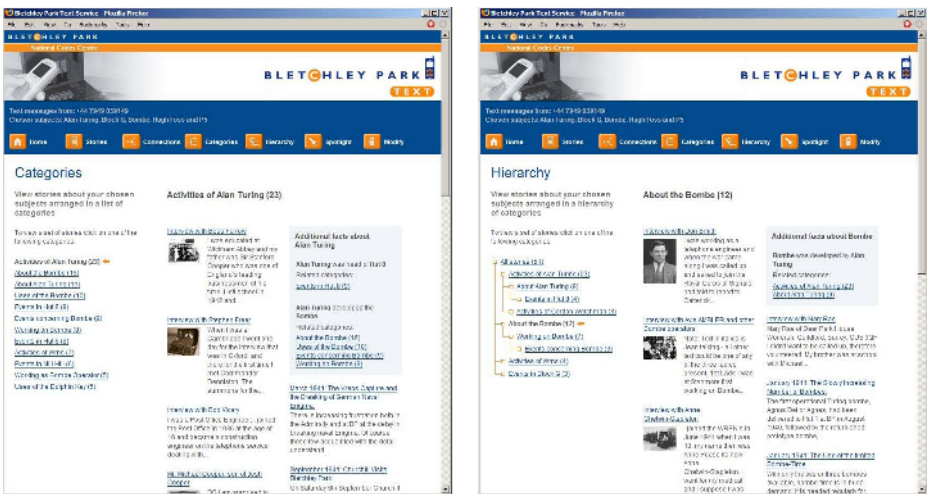


Fig. 6. Examples of the ‘categories’ (left) and ‘hierarchy’ (right) areas showing the use of the ‘Additional facts’ box to provide horizontal links between categories of stories

Along with a preview of the stories used in the connection the semantics from the knowledge model are transformed to produce a natural language description of each part of the connection.

The additional facts, used to represent the key information from the visitor’s tour experience, are applied to produce horizontal links in the ‘categories’, ‘hierarchy’ and ‘spotlight’ presentation areas. The ‘category’ area shows the visitor’s story collection organized in a list of up to 12 categories (see Figure 6, left). The ‘hierarchy’ shows the same collection in a hierarchically organized list of categories (see Figure 6, right). These categories are formed in a bottom-up fashion from the resources retrieved by the visitor’s chosen concepts. Forming groups within a collection of resources based on the underlying semantics enables the visitor to see how the resources can be organized and highlights distinctions between the resources in separate groups.

The ‘spotlight’ area separates out the list of categories into two sets (see Figure 7). One set contains the five most frequent categories that explicitly mention the visitor’s chosen concepts. The other contains the five most frequent categories that do not mention any of the visitor’s concepts. Although these categories may overlap those shown in the ‘categories’ area, the distinction between the directly and indirectly related categories emphasizes the relationship between the concepts used to identify the collection and those closely related according to the knowledge level descriptions. The spotlight metaphor is used here to convey a sense that the visitor’s set of concepts illuminate a section of the online archive. Choosing concepts in the indirectly related set of categories will slowly move the spotlight to a nearby section. Conversely, choosing a set of completely unrelated concepts will make the spotlight jump to an entirely new section of the archive.

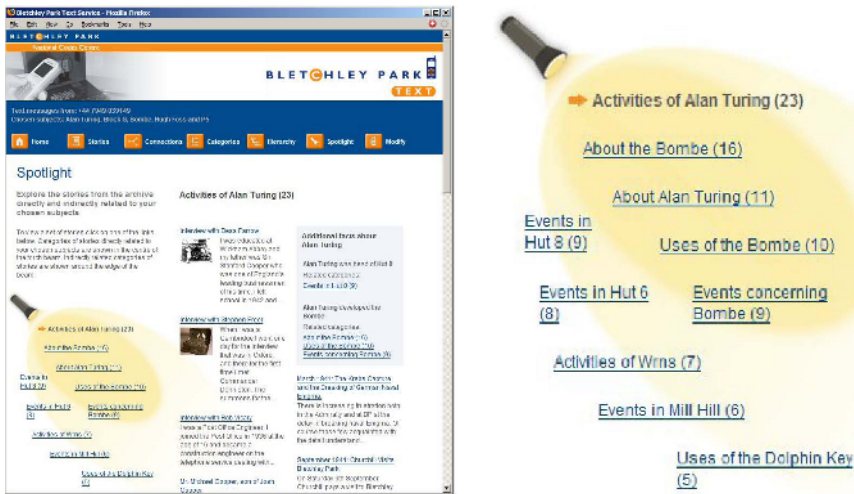


Fig. 7. The ‘spotlight’ area showing sets of categories directly and indirectly related to the visitor’s chosen concepts (left). Directly related categories are shown in the focus of the torch beam, indirectly related categories are shown to the side of the torch beam (right detail).

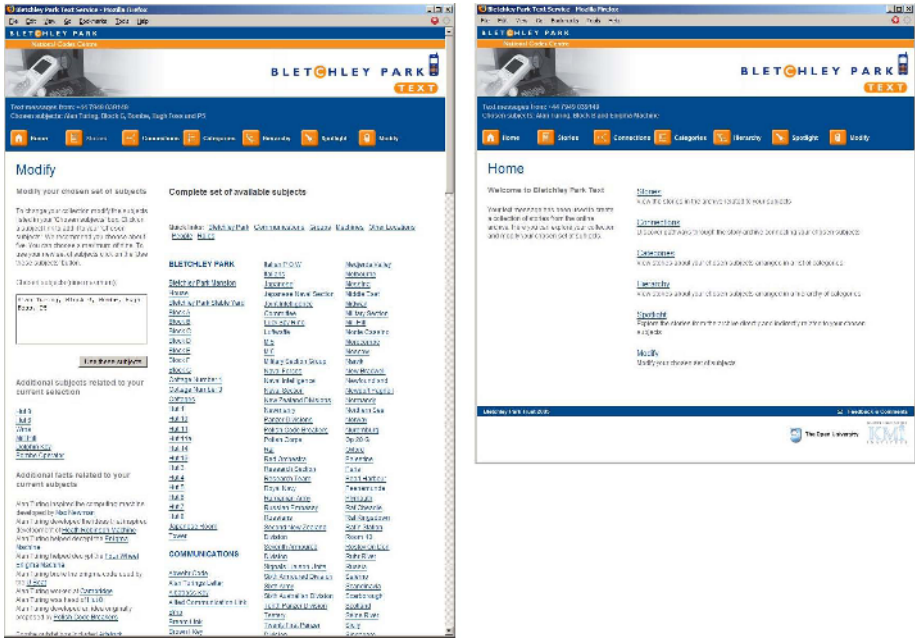


Fig. 8. The ‘modify’ area where the visitor can edit their chosen set of concepts (left). When the visitor finishes editing their concepts and selects the ‘Use these concepts’ button, they are taken back to the home page with a new story collection (right).

Finally, the ‘modify’ area allows the visitor to edit their set of concepts (Figure 8, left). Five components make up the modify page. The current set of concepts is listed in an editable text box at the top left hand corner of the page. The related concepts (i.e. those shown on the edge of the spotlight) are listed below the current concepts. These are followed by a list of additional facts related to the current set of concepts (i.e. those displayed in the categories, hierarchy and spotlight areas). At the bottom left hand corner of the page is a menu listing all of the visitor’s previous sets of concepts. The right hand side of the page shows an ordered list of over 500 concepts. All concept labels are displayed as hyperlinks. Clicking on a label (in the related concept, additional fact or the larger list of concepts) will add it to the set displayed in the chosen set of concepts text box. Once a new set of concepts has been chosen, clicking on the ‘Use these subjects’ button will use these concepts to create a new collection and the visitor will be returned to the home page to begin the exploration process again (Figure 8, right).

5 Discussion

The content of the Bletchley Park Text system is primarily text with illustrative pictures, but the approach is applicable to any media type. Within the current system a single content source and knowledge model is used. However, providing the

ontologies are used consistently to describe the resources there is no reason why the content and knowledge model could not be distributed. The Apollo tool used to develop the knowledge model can also export models using the standard RDF and XML syntaxes. The OCML syntax was used in our case because it was directly compatible with a Lisp-based reasoning engine, which offered us a fast and efficient means for retrieving and organizing the resources.

Currently, the Bletchley Park Text system presents the shortest path connecting any pair of concepts. However, we have also been exploring ways of presenting alternative pathways. The choice of connections in a pathway could be biased towards or away from the visitor's current set of concepts. Including the visitor's concepts may highlight the connections between their chosen set of concepts, whereas excluding them would bring the visitor in contact with related but yet unconsidered concepts. Another possibility is to use the pathways as a means to construct time-ordered sequences of events. For example, introducing a particular person the work they carried out at Bletchley Park and where they went afterwards.

The category based presentations currently show the most frequently occurring categories within the visitor's collection. Revealing the complete set of categories within any given collection would provide a more flexible means for exploring the resources. However, this may also provide the visitor with too much information. One solution being considered is to enable the user to select which slots and values are used, or not used, to form categories.

A pilot study involving a group of 35 high school pupils and their teachers was carried out in September 2004 [22]. All of the pupils used their mobile phones to send messages, 20 of them chose to follow up their visit by signing onto the Bletchley Park Text web site. Six of the students were asked to write an essay after their visit and a follow-up interview was carried out with this group. In both the essays and interview the pupils clearly demonstrated that they had drawn on the resources available from the web site, and several of the pupils indicated that they wanted to continue to use the site. As noted earlier, the Bletchley Park Text system is now available to all visitors, and we are currently monitoring the messages received and the changes visitors make to their concepts. Further evaluation studies are being planned to explore visitors' use of the web site and alternate forms of presentation.

6 Conclusion

This paper has explained our approach for using semantics to create structured presentations of personalized collections. The Bletchley Park Text system uses this approach to produce a web site where visitors can read historical accounts and interviews with the people that worked at Bletchley Park.

The cost in effort of providing this (or similar) sites is primarily the annotation of the resources. Identifying and digitizing appropriate content is a relatively small task compared to the development of the knowledge level description. Although considerable advances are being made in the automatic annotation of text resources (as illustrated by the Artequakt system [12]), developing an accurate knowledge model is a difficult and critical aspect of this approach. However, the benefits are clear. The automated identification of pathways connecting concepts and the

formation of meaningful categories, as used in the Bletchley Park Text system, are currently not possible without the use of knowledge level descriptions. Furthermore, once produced, the knowledge model can be used to provide a range of services, not just for museum visitors but also museum staff.

The majority of the existing approaches for searching the web have focused on the retrieval of single pages. This form of presentation fails to support the exploration and analysis of web resources. By exploiting the semantics of online resources (such as museum archives) the semantic web is creating an opportunity for us to support people's playful explorations in ways that highlight the connections across web resources and categories within collections of resources.

References

1. Falk, J.H. and L.D. Dierking, Learning from museums: Visitor experiences and the making of meaning. American Association for State and Local History, ed. R. Rodgers. 2000, Walnut Creek, CA: AltaMira Press. 272.
2. Gaver, B., Designing for Homo Ludens, in *i3* magazine. 2002. p. 2-6.
3. Kravchyna, V. and S.K. Hastings, Information Value of Museum Web Sites. *Fisrt Monday*, 2002. 7(2).
4. Anani, N. Enhancing the Heritage Experience. in *Museums and the Web*. 2005. Toronto, Canada: Archives & Museum Informatics.
5. Rose, D.E. and B. Levinson. Understanding User Goals in Web Search. in *WWW*. 2004. New York, USA: ACM Press.
6. Yee, K.-P., et al. Faceted Metadata for Image Search Browsing. in *CHI*. 2003. Fort Lauderdale, Florida, USA: ACM Press.
7. Dumais, S., E. Cutrell, and H. Chen. Optimizing Search by Showing Results in Context. in *SIGCHI*. 2001. Seattle, WA, USA: ACM Press.
8. Kaki, M. Findex: Search Result Categories Help Users when Document Ranking Fails. in *CHI*. 2005. Portland, Oregon, USA: ACM Press.
9. Guha, R., R. McCool, and E. Miller. Semantic Search. in *WWW*. 2003. Budapest, Hungary: ACM Press.
10. Pearce, S.M., *On collecting: An investigation into collecting in the European tradition*. 1995, London: Routledge.
11. Geurts, J., et al. Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations. in *International Semantic Web Conference*. 2003. Florida, USA: Springer Verlag.
12. Alani, H., et al., Automatic Ontology-Based Knowledge Extraction from Web Documents. *IEEE Intelligent Systems*, 2003. 18(1): p. 14-21.
13. Landow, G., *Hypertext: The Convergence of Contemporary Critical theory and Technology*. 1992: The John Hopkins University Press, Baltimore, USA.
14. Rutledge, L., et al. Finding the Story - Broader Applicability of Semantics and Discourse for Hypermedia Generation. in *Hypertext*. 2003. Nottingham, UK.: ACM Press.
15. Rutledge, L., J.v. Ossenbruggen, and L. Hardman. Making PDF Presentable. in *WWW*. 2005. Chiba, Japan: ACM Press.
16. Mulholland, P., T. Collins, and Z. Zdrahal. Story Fountain: Intelligent support for story research and exploration. in *International Conference on Intelligent User Interfaces*. 2004. Madeira, Portugal.: ACM Press.
17. CIDOC Conceptual Reference Model. Available from: <http://zeus.ics.forth.gr/cidoc/>

18. Chatman, S., *Story and Discourse: Narrative structure in fiction and film*. 1980, New York: Cornell University Press.
19. Brusilovsky, P. and R. Rizzo. *Map-Based Horizontal Navigation in Educational Hypertext. in Hypertext*. 2002. College Park, Maryland, USA.: ACM Press.
20. Apollo Knowledge Modelling Application. Available from: <http://apollo.open.ac.uk>
21. OCML: Ontological Conceptual Modelling Language. Available from: <http://kmi.open.ac.uk/projects/ocml/>
22. Mulholland, P., T. Collins, and Z. Zdrahal. *Spotlight Browsing of Resource Archives. in Hypertext*. 2005. Salzburg, Austria: ACM Press.