

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Generating Paths through Cultural Heritage Collections

### Conference or Workshop Item

How to cite:

Fernando, Samuel; Goodale, Paula; Clough, Paul; Stevenson, Mark; Hall, Mark and Agirre, Eneko (2013). Generating Paths through Cultural Heritage Collections. In: Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Association for Computational Linguistics, Sofia, Bulgaria pp. 1–10.

For guidance on citations see [FAQs](#).

© 2013 Association for Computational Linguistics



<https://creativecommons.org/licenses/by-nc-sa/>

Version: Version of Record

Link(s) to article on publisher's website:  
<http://www.aclweb.org/anthology/W13-2701>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Generating Paths through Cultural Heritage Collections

Samuel Fernando<sup>1</sup>, Paula Goodale<sup>2</sup>, Paul Clough<sup>2</sup>,  
Mark Stevenson<sup>1</sup>, Mark Hall<sup>2</sup>, Eneko Agirre<sup>3</sup>

<sup>1</sup>Department of Computer Science, University of Sheffield

<sup>2</sup>Information School, University of Sheffield

<sup>3</sup>Computer Science Department, University of the Basque Country

{s.fernando, p.goodale, p.d.clough,  
r.m.stevenson, m.mhall}@sheffield.ac.uk  
e.agirre@ehu.es

## Abstract

Cultural heritage collections usually organise sets of items into exhibitions or guided tours. These items are often accompanied by text that describes the theme and topic of the exhibition and provides background context and details of connections with other items. The PATHS project brings the idea of guided tours to digital library collections where a tool to create virtual paths are used to assist with navigation and provide guides on particular subjects and topics. In this paper we characterise and analyse paths of items created by users of our online system. The analysis highlights that most users spend time selecting items relevant to their chosen topic, but few users took time to add background information to the paths. In order to address this, we conducted preliminary investigations to test whether Wikipedia can be used to automatically add background text for sequences of items. In the future we would like to explore the automatic creation of full paths.

## 1 Introduction

Paths (or trails) have been studied as a means of assisting users with the navigation of digital collections as an alternative to standard keyword-based search (Furuta et al., 1997; Reich et al., 1999; Shipman et al., 2000; White and Huang, 2010). Paths can be particularly useful to users who are unfamiliar with the content of digital collections (e.g. historical documents) and may find it difficult to formulate appropriate queries (Wilson et al., 2010). Paths can be used to assist users with the navigation of collections through the provision of narratives and subject guides. From an

educational perspective paths can provide tangible learning objects, created by teachers and followed by students. Alternatively from a cultural heritage perspective paths can be used to create activity trails and guided tours support exploration by visitors through collections of cultural artefacts. This echoes the organised galleries and guided tours found in physical museums. The existence of tools, such as Walden's paths<sup>1</sup>, Trailmeme<sup>2</sup> and Storify<sup>3</sup>, provide functionalities for users to record and share paths through web resources and digital libraries. From this perspective everyone can take on role of curator and provide access to their own personal collections.

We have developed an online system called PATHS that allows curators and end-users to create and view paths to navigate through the Europeana<sup>4</sup> cultural heritage collection. As part of evaluations of the prototype PATHS system participants have created paths on various topics. In this paper we describe a number of these paths and their characteristics. Analysing paths that are created manually and characterising them can be seen as a first step towards developing methods to support the creation of paths automatically and semi-automatically. Within the context of the PATHS project this is being considered to deal with the following limitations of manual creation of paths. Firstly, the effort required in generating them often means that a sufficient number of paths on a variety of topics are not available. Secondly, the manual creation of paths is a very time-consuming process that would benefit from computational support in whatever form this might take. This paper presents initial work in automatically creating paths and provides the following novel con-

<sup>1</sup><http://www.csd1.tamu.edu/walden/>

<sup>2</sup><http://open.xerox.com/Services/xerox-trails>

<sup>3</sup><http://storify.com/>

<sup>4</sup><http://www.europeana.eu/>

tributions: (1) we present results of user studies describing what people want from paths and how they use them to navigate digital collections; (2) we analyse a set of manually-created paths to identify their properties and be able to characterise them; and (3) we present work on automatically generating background text for sequences of items, thus providing an efficient way to enrich paths with additional information with little manual input required.

The paper is structured as follows: Section 2 describes related work on the use of narratives in cultural heritage and previous approaches to automatically generate paths; Section 3 defines the problem of generating paths and describes the datasets used in the experiments; Section 4 presents analysis of manually-created paths; Section 5 shows results of using automatic methods to generate background text; and finally Section 6 concludes the paper and provides avenues for further work.

## 2 Related Work

### 2.1 Narratives and Cultural Heritage

The potential of narrative in digital CH to support learning, creativity and exploration is clear, providing opportunities for supporting a more active user interaction, including deeper engagement with context, representation of the collecting process, and facilitation of a more entertaining experience of learning (Mulholland and Collins, 2002). Walker et al. (2013) also propose narrative as a major element of interaction and informal learning, suggesting that meaning is made when the links between people and artefacts, and interpretation and ideas are surfaced, especially within social groups. Their experiments involve the use of mobile and handheld technologies in a physical museum environment, capturing audio annotations, but have much in common with experimental systems designed for path creation online. In a similar vein the StoryBank project utilises collections of photographs and audio narratives to create and share stories as information in the developing world (Frohlich and Rachovides, 2008).

Whilst technologies have aided the creation and sharing of narratives in physical cultural encounters, Manovich (1999) critiques the lack of narrative in digital cultural environments, offering that online collections and many CH web sites are databases with constantly changing content that inevitably lack a cohesive and persistent story.

However, since “narrative is constructed by linking elements of this database in a particular order” (Manovich, 1999), it is possible to offer users any number of explicit ‘trajectories’ (narratives) through a digital information space, and by merging database and narrative in this way, creating a more dynamic, discovery-led experience. This view might be interpreted at its simplest level as a virtual representation of the guided tours routinely offered in physical CH spaces, and indeed there is a small strand of research into the creation of systems for generating and exploring online exhibitions and tours from items held within digital collections. A scenario of users creating and editing trails in a CH context is described by Walker (2006), including functionality for collecting, ordering and annotating museum objects.

### 2.2 Automatically Creating Paths

Generation of implicit trails through physical and virtual museum spaces has been related to the learning process (Peterson and Levene, 2003). In this example, trails are automatically created by users as they navigate their way through an information space, and may be used for individual or collaborative purposes. Research on the application of curated pathways in web environments has often focused on providing trails pre-prepared by experts (e.g. curators, educationalists) as a means of assisting novice users to navigate information online (Shipman et al., 2000). Indeed, it has been found that domain knowledge or expertise can considerably enhance the quality of trails created (Yuan and White, 2012). Automatic extraction and generation of trails in information spaces has been explored as a means of harnessing the wisdom of crowds, using the mass actions of earlier user behaviour to establish relevance, and recommend content or navigation routes to later users. Such trails can be readily mined from search engine transaction logs and have been shown to provide added value (White and Huang, 2010; Hassan and White, 2012; Liao et al., 2012). West and Leskovec (2012) take this notion a stage further and attempt to identify wayfinding strategies employed by browsers in Wikipedia, with the goal of assisting future users in their navigation by surfacing potentially useful hyperlinks.

Guided tours or pathways are essentially more structured, purposeful forms of trails, taking the user through a specific sequence of information

nodes and may also be automatically generated, rather than manually curated as in the examples above. Wheeldon and Levene (2003) offer an algorithm for generating trails from site-search, enabling elements of structure and context to be incorporated into the trails created in this way, but noting potential scalability issues for web scale search tasks. In the CH domain, a small number of projects have attempted to automatically generate digital content in the form of exhibitions, tours and trails. Mäkelä et al. (2007) describe a system which utilises semantically annotated content to generate personalised ‘exhibitions’ from a structured narrative-based search query. Similarly, Zdrahal et al. (2008) demonstrate how pathways can be generated through a collection of semantically related documents to provide a means of exploration, using non-NLP clustering and path creation techniques. Sophisticated approaches such as linear programming and evolutionary algorithms have also been proposed for generating summaries and stories (McIntyre and Lapata, 2010; Woodsend and Lapata, 2010). In contrast, Wang et al. (2007) use a recommender system approach to generate museum tours on the basis of ratings stored within a dynamic user model, and Pechenizkiy and Calders (2007) propose the additional use of data mining techniques on log data to improve this type of tour personalisation.

In summary, online tours and trails are made possible either through manually curated content generated through the efforts of experts or other end users, or have been automatically generated from the mining of large scale search logs, or from collections benefitting from semantically-linked content and/or detailed user models.

### 3 Methodology

This study brings together work from several areas of the PATHS project. An analysis of what paths might be used for and what form they are expected to take, has had implications for the system design and functionality and evaluation measures. A user study focused upon evaluation of the first prototype has provided manually-created paths as a basis for analysing path content and attributes, which in turn informs the desired characteristics of automated paths and the algorithm designed for generating paths automatically.

#### 3.1 Utilisation of Paths

Initial user requirements interviews with 22 expert users in the heritage, education and professional domains found a strong affinity with the path metaphor, revealing a range of different interpretations of what it means in the CH context and how they could be employed in an online environment to engage with key audiences. Eight interpretations of the path metaphor emerged:

1. Path as search history
2. Path as information seeking journey
3. Path as linked metadata
4. Path as a starting point or way in
5. Path as a route through
6. Path as augmented reality
7. Path as information literacy journey / learning process
8. Path as transaction process

The first three of these are closest to the idea of hypertext trails, with trails defined by user interaction in 1 and 2, and trails defined automatically, by the system in 3. Variations 4-6 are more creative interpretations, all suggesting opportunities for guiding the user into and through collections, encouraging exploration and/or offering an immersive experience, conducive with our initial vision for the PATHS system.

In addition to expert-defined routes, 5 also incorporates the idea of users being able to see and follow “well-trodden path” defined by the cumulative interactions of other users, thus extending the opportunities for utilizing search histories. Conversely, 7 and 8 are both process oriented, although 7 is experiential, user-defined, learning-oriented, typified by trial and error and unique to the individual, whilst 8 is a rigid process designed to escort all users consistently through a standard process of pre-defined steps.

A strong emphasis was placed on path content being carefully selected or ‘curated’ by the path-creator, with the addition of context and interpretation so that the objects within the path convey a narrative or meaning. Content may be derived from one collection, but there were seen to be significant benefits from including objects from diverse collections, along with other materials from external web sites.

Paths facilitate topic-based information retrieval typified by the berry-picking mode of interaction (Bates, 1989), rather than known item searching. Furthermore, paths may be a useful tool

for personal information management in both formal and informal research scenarios, enabling the user to record, reuse and share their research activity, or helping them to organize their ideas. Creativity is also encouraged, as user-generated paths provide the means to repurpose CH objects into users' own narratives for private or public consumption.

A summary of specific user scenarios highlighted by participants is given below:

- Teachers/lecturers presentations and classroom activities
- Museum personnel curating collections, giving an overview, or covering a topic in depth
- Leisure users browsing, collecting interesting and/or visually appealing content
- Researchers to aid image-based research, sharing and discussing findings with fellow researchers and supervisors
- Non-academic specialists (e.g. local historians) collecting and sharing items of interest with other enthusiasts

### 3.2 Defining the Problem

To create a path or narrative that guides a user through a set of items from a collection, whether as a manual process or automatically, there are three main activities: (1) the selection of items to include in the path; (2) the arrangement of items to form a path or narrative and (3) the annotation of the path to with descriptive text and background information. We envision techniques to automate the entire process; however, a first step is to analyse existing manually-created paths to identify their characteristics and inform the automatic creation of similar structures.

### 3.3 User Study

The manually generated paths used for this study were created as part of a more detailed user study to evaluate the first prototype, conducted using a protocol informed by the Interactive IR evaluation framework (Borlund, 2003). Twenty-two users, including subject experts, students and general users (subject novices), each completed a 2-hour session, during which they participated in the following activities:

- Profile questionnaire and cognitive style test
- Familiarisation with the system
- 4x short information seeking tasks (5 minutes each)

- 1x long simulated work task - path creation (30 minutes)
- Task feedback questionnaire
- Session/system feedback questionnaire
- Think-after interview based upon the complex task

Of most interest here is the simulated work task, with associated observations, feedback and reflections. This task focused on the creation of a path, using a scenario adapted to the type of user. Freedom was given in choosing a subject for the path, and limited instructions were provided in what might be needed to complete the task, for example:

“Imagine you are a student who has been asked to create a path as part of a university assignment. You have been asked to use primary source materials to create a mini online exhibition suitable for a target group within the general public and/or school visitor categories. Your goal is to introduce a historical or art-focussed topic in a popular, accessible way, and to encourage further use and exploration of cultural heritage resources.”

Data on the tasks was captured via log files, as well as screen recording and observations using the Morae usability software. Detailed analysis was undertaken of user behaviour in the process of completing the task, and of the paths created, from both quantitative and qualitative perspectives.

## 4 Analysing Manually-created Paths

In this section we describe the results of analysing the 22 paths created manually in the PATHS prototype system.

### 4.1 User behaviour

On average users spend 25.3 mins on creating a path (min=11.7; max=33.6) with an average of 201 mouse clicks (min=53; max=380). From the observations, it was noted that some participants spent quite a lot of time thinking about the task and pondering their next move, whilst others engaged in more rapid fire activity in the face of uncertainty. Analysis of the screen recordings showed a variety of primary interaction styles for this task, with a fairly even split between serial searching (33%) and serial browsing (39%), as the two most popular strategies. Serial searching involves repetitive search and reformulation, with only a page or two of search results viewed before searching again, and serial browsing involves very

few searches, with large numbers of search results pages viewed (over 50 pages in some cases). These are then in effect, polar opposites of interaction. Only 6% engaged primarily in exploring behaviour (using the explore and similar items content), and 22% of participants occupied the middle ground, utilising a mix of search, browse and explore, with no strong preference for any one style.

## 4.2 Properties of paths

The mean number of items in a path was 10.7 (std dev=6.7 items) with a minimum of 5 items and maximum of 29 items. Most popular bin is 6-10 items in a path (59%). We found 85% of the items included in the paths included an image with the metadata. The paths created were manually categorised by theme to ascertain whether there are any distinct preferences for the subject matter of content included. The most popular categories were paths about places (23%), art subjects (23%) and history subjects (32%). These themes are likely to have been influenced at least partly by what content is currently available in our collection, although the amount of art-related content is much less than for history, and also appear to have been influenced by the topics covered in existing paths in the system (e.g. places, topics related to the world wars). There were, however a significant number of expert users who attempted to build paths related to their own research interests, with varying degrees of success.

## 4.3 Descriptions and ordering

Once items have been selected and they have been transferred in the path creation workspace, users have the opportunity to modify and enhance their path with a number of tools for adding content and metadata, and for re-ordering the content. On creating the path, most users immediately went to the metadata fields and added information for the path description and duration fields, as well as a number of tags (or keywords). A short 1-2 line description of the path appears to be the norm and was added in 91% of cases. Tags were added by 82% of users and a duration by only 46% of users. It is clear from further investigation that the tags were added incorrectly (without commas between them) by a significant number of users and a tip for successful use is required.

The items within a path can be annotated with the user's own contextual information, and can be re-ordered into a more meaningful sequence, such

as a chronological or narrative sequence. These more advanced features were used by significantly fewer users, which could indicate a learning issue, a lack of need, or a time constraint. On reviewing the paths created by our evaluation participants it is found that in 41% of cases, contextual information was not added to any items in the path. There are however 32% in which annotations were added to all items (generally these were shorter paths with fewer items), and a further 27% where annotations were added to some or most of the items.

In 72% of cases the items in the paths created were re-ordered to some degree, with 17% spending a considerable amount of time on this activity. This finding is encouraging, as the default is for items to be included in the path in the order they were saved to the workspace, and re-ordering indicates that users are thinking about their path as a whole and trying to make sense of the information it is intended to convey. Typical types of ordering included chronology (32%), narrative (23%), geography (for example, a walking tour - 9%), theme (9%) and 'interestingness' (5%).

## 5 Enriching paths with background information

This section describes preliminary work on the task of semi-automated path creation. In particular we describe efforts to enrich paths with background contextual information using relevant Wikipedia articles. The related work described in Section 2.2 shows that there have been previous efforts to automatically select cultural heritage items to form paths, trails and exhibitions. However to our knowledge no significant effort has been made to automatically annotate such paths with descriptive or contextual information. The interviews described in Section 3.1 highlighted the importance CH experts placed on having additional information to give context for the items in the path. It was also noted during the manual path-creation exercise (Section 4.3) that a significant number of the users did not add any such information to the path. The reasons for this are unclear, but nevertheless there seems to be sufficient motivation to devise automatic methods for this task. Although the methods have previously been well established in other tasks<sup>5</sup>, we believe

<sup>5</sup>INEX Tweet Contextualization Track (<https://inex.mmci.uni-saarland.de/tracks/qa/>) and Link-the-wiki Track (<http://www.inex.otago.ac.nz/tracks/wiki-link/wiki-link.asp>)

this is the first time they have been applied for the task of annotating sequences of items in this way.

## 5.1 Method

Manually generated paths contain sequences of items selected from Europeana on some topic or theme. Creators provide their own title, subject keywords and description for the path. To aid creation of paths we explore whether background information could be generated automatically for such paths. An approach is presented here which shows promise as a potential way to achieve this task. The input for this approach is a sequence of items and a key Wikipedia article which describes the overall topic of the path. The output comprises sentences taken from a relevant Wikipedia article. The aim is for this output to provide useful and interesting additional background information related to the items and theme of the path. In this paper experiments are focussed on how to select good quality text to present as additional information for the path. For this reason the key Wikipedia article is manually chosen, and the task is to find a good approach for selecting the most relevant sentences from this key article for the text.

Two methods are tested in this paper. The first method simply takes the first  $n$  sentences of the article and outputs this. Since Wikipedia articles are always structured to have a summary of the article in the first paragraph we can expect this text to perform well as a summary of the path topic.

The second method is more advanced and attempts to find text in the article that is relevant to the actual items that have been chosen for the path. This approach uses the Wikipedia Miner software (Milne and Witten, 2008) to add inline links to the text in the items for this approach. This software disambiguates terms in the text and then detects links using various features such as the commonness of the term, the overall relatedness of the terms in the text and so on. The result is text enriched with inline links to relevant Wikipedia articles. Each link also has an associated confidence value which indicates how sure the software is that the link is correctly disambiguated and relevant to the text.

The approach works as follows for a sequence of items  $S$  and a key article  $K$ . First Wikipedia Miner is run over the items in  $S$ . The text input to Wikipedia Miner comprises the title, subject and description fields of each item. The output is a set

of article titles  $W$  comprising the titles of all the linked articles which were found in the text fields of  $S$ . For each title in  $W$  we also have the associated confidence value for the link as calculated by Wikipedia Miner. The next step is to select from  $K$  the most relevant sentences to output as the generated text. For each sentence in  $K$  a score is assigned if any of the words in the sentence match one of the titles in  $W$ . The score is then simply the sum of the confidence values associated with these titles. The top scoring sentences are then output as the background text. This method can be considered to be a kind of query based summarisation (Jurafsky and Martin, 2008).

## 5.2 Results

The automatic approaches for generating background text were run over the items in the 22 manually created paths described in the previous section. To ensure a fair test the user-added text was discarded and only the original Europeana metadata for the items was used as source for the methods.

For each path a single key Wikipedia article was manually found which best captured the overall theme or topic of the path. For all paths at least one reasonably good article was found, although for some paths the articles were better matches for the path topic than for others. Some examples are shown in Table 1.

With the input ready, the approaches described were run over the items and key articles to generate the additional text. This was limited to 3 sentences to provide a short but useful piece of text for viewers of the paths. For example one path was entitled “Bronze Age metallurgy”. The user provided description was as follows:

*Bronze Age metalsmiths produce a variety of objects, however, we learn as much from the tools, leftover debris and the way in which objects were deposited as we do from the finished objects themselves.*

The key article chosen for the path was “Bronze Age”. The items in the path included various artifacts and working debris from the Bronze Age. For the first approach the output is the first 3 sentences from this article:

*The Bronze Age is a period characterized by the use of copper and its alloy bronze as the chief hard materials in the manufacture of some implements and weapons. Chronologically, it stands be-*

Path title	Key Wikipedia article
Canals in England	Canals of the United Kingdom
A brief history of Sheffield	History of Sheffield
Art and Criticism of John Ruskin	John Ruskin
Early medieval features in English Parish Churches	Architecture of the medieval cathedrals of England
Brooches through the ages	Brooch

Table 1: Key articles for paths

tween the Stone Age and Iron Age. The term Stone Age implies the inability to smelt any ore, the term Bronze Age implies the inability to smelt iron ore and the term Iron Age implies the ability to manufacture artifacts in any of the three types of hard material.

For the second approach the top 3 highest scoring sentences were output:

*A region could be in the Bronze Age either by smelting its own copper and alloying with tin or by trading for bronze from production areas elsewhere. The Bronze Age in Ireland commenced around 2000 BC, when copper was alloyed with tin and used to manufacture Ballybeg type flat axes and associated metalwork. The Bronze Age is a period characterized by the use of copper and its alloy bronze as the chief hard materials in the manufacture of some implements and weapons.*

These sentences scored most highly since they contained the most high-confidence terms from the items, for example terms such as ‘copper’, ‘alloy’ and ‘Bronze Age’.

### 5.3 Evaluation

To evaluate the two approaches, 5 annotators were presented with the paths and the text and asked to rate each path on 3 dimensions:

- The *relevance* of the text to the theme and items of the path. Text which relates strongly to the path is scored highly while off-topic or irrelevant text is given a low score.
- The *coherence* or quality of the text itself. Text which appears well-written and well-structured is scored highly, while poorly written or incoherent text is given a low score.
- The *contextualisation* of the text in relation to the path. To achieve a high score the text should offer useful or interesting additional information which is not found elsewhere within the content, i.e. the text helps to provide a context for items in the path.

Annotators were asked to grade from A (very good) to E (very poor) on each dimension. The results are shown in Figure 1. The results for the first 3 sentences are shown as **First3** and for the weighted approach as **Weighted**. For each dimension, the distribution of judgements across the paths is shown. The **First3** approach was found to be superior in every dimension. For relevance scores 90% of the scores were either A or B compared to 63% for the **Weighted** approach. Similarly for the coherence judgements 97% were A or B compared to 62% for the weighted approach. The reason for this superior performance seems to be that the first few sentences of Wikipedia articles are deliberately created to give a short summary introduction of the topic of the article. This explains the high scores for relevance and coherence.

Both approaches scored lower on the contextualisation dimension, with **First3** getting 67% A or B grades and the **Weighted** approach getting 43%. There may be several reasons for this. Firstly one problem is that the auto-generated text sometimes repeats information that is already in the path and item descriptions; thus the text fails to meet the requirement of ‘useful additional information’. Secondly the text is sometimes quite general and vague, rather than focussing on specific details which might be most relevant to the items chosen for the path.

To measure the agreement among the annotators the following approach was used. First the scores were converted to numeric values; A to 1, B to 2 and so on. Then the scores for each annotator were compared to the average of the scores of all the other annotators. The correlation was computed using Spearman’s correlation coefficient. These scores were then averaged amongst all annotators to give a final agreement value. The results are shown in Table 2.



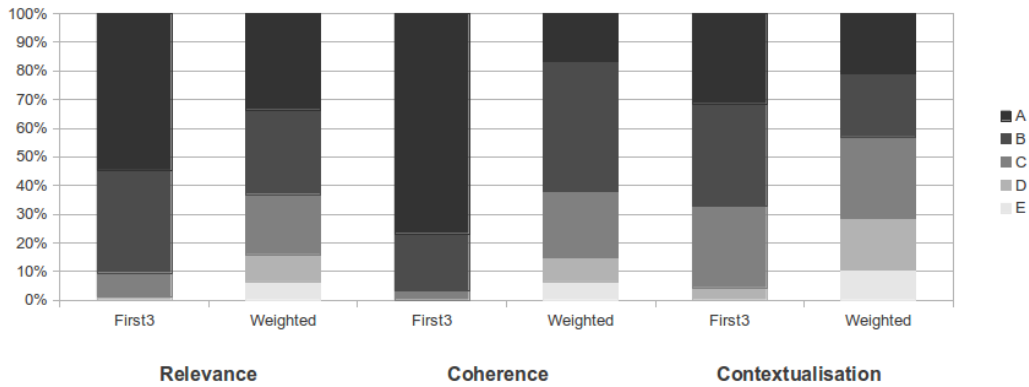


Figure 1: Comparing the results of the two methods.

	<b>First3</b>	<b>Weighted</b>
Relevance	0.57	0.57
Coherence	0.28	0.56
Contextualisation	0.56	0.78

Table 2: Agreement amongst annotators.

For both approaches there was good agreement on the Relevance dimension. For the Coherence dimension the **First3** approach got quite a low score. This may be because one annotator gave lower scores for all paths, while the others all gave consistently high scores, which seems to have skewed the correlation co-efficient. For the contextualisation dimension the correlation scores for high for both approaches, and the **Weighted** approach in particular achieved a very high agreement value.

## 6 Conclusions

This paper presented results of interviews about creating paths through cultural heritage collections. These results inform us on how people want to navigate through cultural heritage collections using the path metaphor, how they wish to make use of paths for their work and education, and what information and qualities they consider it important for a path to contain. The paper also presents results from studies using the PATHS prototype software where users were able to search and explore a large digital library collection and create their own paths of items from the collection on topics of their interest.

From the interviews it was clear that the experts considered it important that the paths contain additional information to convey contextual informa-

tion to understand the meaning of the items in the path. The results from the user studies showed that this need was not being met in a significant number of cases; users were putting items together on a topic but adding little or no descriptive text about the topic and the items in the path. Therefore we identified this as a key task which might benefit from automatic methods. The simpler approach which output the first  $n$  sentences from the key Wikipedia article was found to generate the best results. The resulting generated text was found to be relevant and coherent. In most cases the text was also found to add useful context about the topic.

Future work will further refine the text generation approach. The approach depends on successfully identifying a good key article for each path. In these experiments the key article was manually chosen, however we are devising methods to select this article automatically. To correct the problem with repeated information a filtering approach could eliminate information that is already contained within the paths.

## Acknowledgments

The research leading to these results was carried out as part of the PATHS project (<http://paths-project.eu>) funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270082.

## References

Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search inter-

- face. *Online Information Review*.
- Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information research*, 8(3).
- David M Frohlich and Dorothy Rachovides. 2008. Using digital stories for local and global information sharing. In *Community and International Development, CHI 2008 Workshop*.
- R. Furuta, F. Shipman, C. Marshall, D. Brenner, and H. Hsieh. 1997. Hypertext paths and the World-Wide Web: experiences with Walden's Paths. In *Proceedings of the eighth ACM conference on Hypertext*, pages 167–176, New York, NY.
- Ahmed Hassan and Ryen W White. 2012. Task tours: helping users tackle complex search tasks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1885–1889. ACM.
- Daniel Jurafsky and James H. Martin. 2008. *Speech and Language Processing (2nd Edition) (Prentice Hall Series in Artificial Intelligence)*. Prentice Hall.
- Zhen Liao, Yang Song, Li-wei He, and Yalou Huang. 2012. Evaluating the effectiveness of search task trails. In *Proceedings of the 21st international conference on World Wide Web*, pages 489–498. ACM.
- Eetu Mäkelä, Osmo Suominen, and Eero Hyvönen. 2007. Automatic exhibition generation based on semantic cultural content. In *Proc. of the Cultural Heritage on the Semantic Web Workshop at ISWC+ASWC*, volume 2007.
- Lev Manovich. 1999. Database as symbolic form. *Convergence: The International Journal of Research into New Media Technologies*, 5(2):80–99.
- Neil McIntyre and Mirella Lapata. 2010. Plot induction and evolutionary search for story generation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1562–1572. Association for Computational Linguistics.
- D. Milne and I.H. Witten. 2008. Learning to link with wikipedia. In *Proceeding of the 17th ACM conference on Information and knowledge management*, pages 509–518. ACM.
- Paul Mulholland and Trevor Collins. 2002. Using digital narratives to support the collaborative learning and exploration of cultural heritage. In *Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on*, pages 527–531. IEEE.
- Mykola Pechenizkiy and Toon Calders. 2007. A framework for guiding the museum tours personalization. In *Proceedings of the Workshop on Personalised Access to Cultural Heritage (PATCH07)*, pages 11–28.
- Don Peterson and Mark Levene. 2003. Trail records and navigational learning. *London review of Education*, 1(3):207–216.
- S. Reich, L. Carr, D. De Roure, and W. Hall. 1999. Where have you been from here? Trails in hypertext systems. *ACM Computing Surveys*, 31.
- Frank M Shipman, Richard Furuta, Donald Brenner, Chung-Chi Chung, and Hao-wei Hsieh. 2000. Guided paths through web-based collections: Design, experiences, and adaptations. *Journal of the American Society for Information Science*, 51(3):260–272.
- K. Walker, A. Main, and Fass. J. 2013. User-Generated Trails in Third Places. In *HCI-3P Workshop on Human Computer Interaction for Third Places at Computer Human Interaction 2013*.
- Kevin Walker. 2006. Story structures. building narrative trails in museums. In *Technology-Mediated Narrative Environments for Learning*, pages 103–114. Sense Publishers.
- Yiwen Wang, Lora M Aroyo, Natalia Stash, and Lloyd Rutledge. 2007. Interactive user modeling for personalized access to museum collections: The rijksmuseum case study. In *User Modeling 2007*, pages 385–389. Springer.
- Robert West and Jure Leskovec. 2012. Human wayfinding in information networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 619–628. ACM.
- Richard Wheeldon and Mark Levene. 2003. The best trail algorithm for assisted navigation of web sites. In *Web Congress, 2003. Proceedings. First Latin American*, pages 166–178. IEEE.
- Ryen W White and Jeff Huang. 2010. Assessing the scenic route: measuring the value of search trails in web logs. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 587–594. ACM.
- M. Wilson, Kulesm B., M. Schraefel, and B. Schneiderman. 2010. From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1):1–97.
- Kristian Woodsend and Mirella Lapata. 2010. Automatic generation of story highlights. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 565–574. Association for Computational Linguistics.
- Xiaojun Yuan and Ryen White. 2012. Building the trail best traveled: effects of domain knowledge on web search trailblazing. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 1795–1804. ACM.

Zdenek Zdrahal, Paul Mulholland, and Trevor Collins.  
2008. Exploring pathways across stories. In *Proc.  
of International Conference on Distributed Human-  
Machine Systems*.