Using Collaboratively Constructed Document Collections to Simulate Real-World Object Comparisons

Karl Grieser,[♡] *Timothy Baldwin,*[♠] *Fabian Bohnert*[♣] and *Liz Sonenberg*[♡]

♡ DIS
University of Melbourne
VIC 3010 Australia
kgrieser@csse.unimelb.edu.au
l.sonenberg@unimelb.edu.au

♠ CSSE
University of Melbourne
VIC 3010 Australia
tim@csse.unimelb.edu.au

♣ Faculty of IT
Monash University
VIC 3800 Australia
fabianb@csse.monash.edu.au

Abstract *While the layout of a museum exhibition is largely prescribed by the curator, visitors to museums view connections between exhibits in ways unique to themselves. With the assistance of a large-scale survey of museum visitors we identify that the view taken by museum visitors of a collection of exhibits can be represented by similarity over documents associated with each exhibit. We show that even when using a basic document similarity measure there is a correlation between document similarity and visitors' judgements of relatedness of exhibits aligned to these documents.*

Keywords User Studies Involving Documents, Web Documents, Cognitive Aspects of Documents.

1 Introduction

Recently there has been a move towards providing visitors to museums and Cultural Heritage (CH) spaces with personalised tours. These tours can be created explicitly by a visitor prior to entering the collection, or tailored to a visitor while browsing a collection. In order to create a dynamic tour for a given visitor, there is the need to (1) model a visitor's preferences (Zukerman and Albrecht [14]), and (2) have knowledge about the content of individual exhibits and connections between pairs of exhibits (Aroyo et al. [1], Bohnert et al. [3], Cox et al. [4], and Grieser et al. [7]). The focus of this paper is on the second of these requirements, using web documents to represent museum exhibits, and document similarity to model similarities between them.

Museum exhibitions are generally designed around a common theme (e.g., *Melbourne* or *marine life*), and professionally curated so that exhibits are organised in a coherent fashion relevant to that theme with closely-related objects in close physical proximity of each other (e.g., artefacts from the same era or of the same function are often presented together). The task of tour personalisation can be seen as one of matching the interests of a visitor to the themes represented in the museum.

Proceedings of the 13th Australasian Document Computing Symposium, Hobart, Australia, 8 December 2008.

However, visitors to a museum or cultural heritage site can categorise the museum space in a way particular to the context of their visit (e.g., preferring to visit more tactile exhibits to entertain small children, or choosing to visit all exhibits from a particular location or era, irrespective of theme). That is, they often have their own opinions on the degree of relatedness of exhibits, independent of the themed design of a gallery or exhibition.

While various computational methods have been developed to identify relationships between documents or words (e.g., Rubenstein and Goodenough [13] and Ponzetto and Strübe [12]), there is currently no standard method of identifying the manner in which people view the relationships between real-world objects. The objective of this research is to test the portability of document-based similarity methods to the task of estimating museum exhibit relatedness. In particular, we map each of 41 exhibits from Melbourne Museum to the most closely associated Wikipedia article, and perform a simple pairwise cosine similarity calculation over the weighted document vector for each document.

We compare our calculated document similarities against two data sets: (1) real-world relatedness estimates of pairs of exhibits in Melbourne Museum provided by over 500 museum visitors, and (2) a calculation of the physical distance between each pairing of exhibits. In the first instance, we calculate how well document similarity models the museum visitors' notions of exhibit relatedness. In the second instance, we determine how closely our similarity estimates mimics the physical layout of the museum. We also compare physical distance with the visitors' ratings to gauge how faithful the ratings are to the prescriptive theming of the museum space.

2 Related Research

Cultural Heritage spaces such as historical sites and museums are providing greater access to their collections through mobile computing (e.g., Benelli et al. [2] and Oppermann et al. [10]) and the web (Aroyo et al. [1]). This has enabled museums to reach wider audiences and to better communicate the importance of their collections. The personalisation of content through digital collections has been a major focus of many CH projects, and multiple methods have been used to tailor content or tours to the visitor (Aroyo et at. [1], Benelli et al. [2], and Cox et al. [4]). Previous approaches such as the ones in Aroyo et al. [1] and Cox et al. [4] used the attributes of previously rated exhibits to identify other exhibits that the visitor may find interesting. Grieser et al. [7], on the other hand, used common attributes of exhibits in the current visit to predict future exhibits the visitor may visit, while Bohnert et al. [3] used the amount of time a visitor spent viewing exhibits to infer a visitor's interests and pathways. The content-based models explored in Grieser et al. [7] and the collaborative models proposed by Bohnert et al. [3] have the advantage of being non-intrusive, as they do not require explicit exhibit ratings. However, these techniques suffer from the so-called cold-start problem in the initial stages of a visit.

The identification of reasons for visitors finding commonality between exhibits is a key step in personalising a tour. Previous studies have used common attributes to align exhibits and identify similarities (e.g., same artist, same style of jewelery, as used in Cox et al. [4]). This has lead to the use of ontological frameworks as a basis for these comparisons (e.g., The Getty AAT, Iconclass, and the CIDOC Conceptual Reference Model). For CH sites such as art galleries where all exhibits have the same attributes, this method is appropriate. However, for CH sites that have exhibits of differing backgrounds (e.g., natural history museums or national parks) this method does not adequately account for the diversity of the exhibit structure. We aim to address this gap by using an alternative semi-structured data source that is able to identify relationships between concepts within its hierarchy.

Estes [5] showed that for concepts that do not have a common conceptual frame (or physical structure), people relate concepts using a process of integration. An integrative relationship is the interaction that occurs between two concepts. This is different from attributive comparison, where concept attributes are compared in order to determine similarity. This indicates that for CH sites with diverse collections, highly structured data sources and ontologies are unable to sufficiently identify the interactions between exhibits that visitors will make when considering them. Their key failure is that they do not simulate the thought process that the average museum visitor will go through (often an integrative relationship), but rather focus on the organisational hierarchy designed by the collection's curator.

For this study, we will use a non-expert data source that provides relationships between highly different entities, and that is able to represent the information at a common visitor level: Wikipedia. In recent years, Wikipedia has been used increasingly in document processing tasks, due to its sheer size, multilinguality, and domain diversity. Extensive conceptual similarity experiments have been performed in other studies, such as the ones discussed in Gabrilovich and Markovitch [6] and Milne et al. [9]). Particularly interesting is the category hierarchy, as each article must be a member of at least one category. This hierarchy has been investigated by Ponzetto and Strübe [11, 12] as a parallel to other existing hierarchies such as WordNet. In Wikipedia, the articles are created with the intention of being understandable to all users, and even its place in the category hierarchy is reached through discussion and consensus, meaning that an article's content and its organisation is designed to make sense to the majority of people viewing it. This collaboratively constructed social nature of Wikipedia (Mathes [8]) is the reason for choosing it as a data source for this research.

3 Museum Visitor Survey

For the purposes of this research, we identified 41 exhibits from Melbourne Museum which could readily be aligned with Wikipedia documents as per Wikipedia guidelines[1] – some trivially as named entities (e.g., *Phar Lap*), others less convincingly via more general articles (e.g., *gold mining* for a diorama of a Ballarat gold mine). We then designed a web survey drawing heavily on the psycholinguistic research of Rubenstein and Goodenough [13] on lexical similarity. In their research, subjects were presented with a standardised set of word pairs (presented in random order), and asked to rate their relatedness on a discrete scale of 0 to 4. In our case, rather than words, we present the subject with images of two exhibits from Melbourne Museum, and ask them to rate their relatedness on a scale of 0 to 4, keeping with the standardised scale defined by Rubenstein and Goodenough [13]. We also asked for a justification of the rating.

Subjects were presented with 15 exhibit pairs in random order, 3 of which were common to all respondents and the remaining 12 of which were chosen randomly. The images were presented adjacent to each other in a web browser, again in randomised order.

In order to ensure that the ratings were relative to actual visits to Melbourne Museum, we targeted Museum Victoria members exclusively, and asked respondents to indicate how frequently they had visited the museum in the preceding 12 months (as well as other demographic and profiling data which is irrelevant to this current research). We received over 500 responses over a three-week period, and recorded at least one rating for every exhibit pair. Of these, we filtered out a small number where the same relatedness value was given for all 15 exhibit pairs. We then calculated the mean of the relatedness values for a given exhibit pair, and use this as our gold-standard relatedness data.

[1] http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

Pairing of methods	ρ-value	p-value
Human & Physical	$+0.196$	1.5×10^{-8}
Human & Document	$+0.157$	6.6×10^{-6}
Physical & Document	$+0.038$	2.3×10^{-1}

Table 1: Two-tailed Pearson correlation and p-value between the different methods (Human Judgements, Physical Distance and Document Similarity)

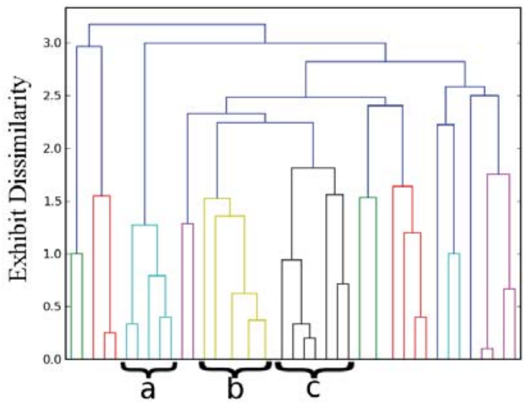

Figure 1: Clustering of exhibits based on the pairwise relatedness ratings

In order to carry out preliminary analysis of the ratings from the museum members, we performed agglomerative hierarchical clustering over the survey results, and identified distinct groupings of exhibits. The hierarchy created from the relatedness scores (translated into *dis*similarity ratings by subtracting from the maximum relatedness value, i.e., 4) is shown in Figure 1.

Encouragingly, we found that the results followed broad thematic boundaries, with cluster (a) revolving around geology, cluster (b) revolving around birds and trees, and cluster (c) revolving around prehistoric animals and fossils, for example. However, in many cases, these clusters do not correspond to the thematic/physical layout present in the museum. We will discuss this further in Section 5.

4 Exhibit Comparison

The arrangement of exhibits within a museum exhibition is often planned around a central theme. This can be a rather broad theme such as *science*, or a more specific one such as documenting the growth of a city over time. In the more specific case, the arrangement of the exhibits is key to an exhibition's interpretation. However, this interpretation may not be the interpretation that a visitor considers when identifying relationships between exhibits.

At this early stage of the research, we calculate simple cosine similarity between the term vectors of the Wikipedia documents to estimate the relatedness of a given pairing of exhibits, weighting terms with a basic tf·idf model. In addition to the document similarity model, we explore the hypothesis that physical walking distance between exhibits is inversely proportional to their degree of relatedness, i.e., closely-related exhibits should be in close physical proximity, and less-related exhibits should be further apart from each other. This derives simply from the careful theming of the museum space by curatorial staff. We calculate the physical distance between exhibits via an SVG image of the museum space, mapped onto a graph structure which preserves the physical layout of the museum (i.e., preventing paths from passing through walls or ceilings).

Identifying which of these measures most closely aligns with the ratings provided in the survey may provide an indication of which viewpoint the average visitor takes: the expert view of the curator, or the more common interpretation supported by the socially-constructed documents.

5 Results and Discussion

We evaluated the relative "fit" between each pairing of human judgements, walking distance and document similarity via the ρ-value of a two-tailed Pearson correlation test over the corresponding lists of exhibit pairings.[2] The results are presented in Table 1.

The highest correlation (at level of statistical significance, $p \ll 0.01$) was obtained for the pairing of physical distance between exhibits with the human judgements. The most obvious explanation for this result is that the visitors' view of exhibits mirrors that intended by the curators to a certain degree. Preliminary analysis of the justifications for relatedness from the web survey supports this observation, with a number of respondents citing physical proximity as the reason for a higher relatedness value.

The second highest correlation was achieved for the pairing of human judgements and document similarities (again at a level of statistical significance), indicating that our document similarity model was moderately successful at capturing exhibit relatedness, despite our relatively simple approach. Note that as the documents were sourced from Wikipedia, there is nothing specific to Melbourne Museum in them, and no indication of how the exhibits are interpreted in the museum space. In this sense, the results are highly encouraging.

We get very low correlation between the physical distance and document similarity (not at a level of statistical significance). When combined with the above two results, this indicates that the document similarity model is modelling something removed from the physical layout of the museum, and yet agrees with our human subjects. Hence, it appears to be picking up on cross-gallery relatedness, and complementing the physical distance model.

[2]Note that we reverse the sign of the ρ-value in the case that we are comparing a similarity with a distance (i.e., similarity vs. dissimilarity).

It is unclear at this point exactly what the degree of influence of the museum layout was on the survey responses. We intend to carry out further analysis of the survey data to clarify this point.

That a basic document similarity measure over a single set of documents could achieve these results is highly encouraging. Clearly there is much more that can be done. Areas of future research we are interested in are analysing cross-article links and the category hierarchy in Wikipedia, and combining these with the document similarity model (inspired in part by the work of Ponzetto and Strübe [11]). We are also interested in exploring a broader range of term weighting, feature selection, and similarity metrics in the document similarity model, as well as different document sets (including documents from the Melbourne Museum website). We anticipate that this will provide a more thorough picture of the way in which museum visitors conceptualise relationships between exhibits.

6 Conclusions

When identifying relationships between exhibits within a museum, previous methods have used highly structured methods often created by a curator highly familiar with the collection. We have proposed a document similarity model which makes use of content authored by non-experts to overcome the curator-centric design as well as to identify associations between exhibits.

Through comparison of the conceptual design of the museum in its physical layout (based on exhibit locality within the museum), the content of exhibits (represented by collaboratively-constructed documents relating to the exhibit content), and the ratings of museum visitors (obtained through a web survey), we have shown the following: (1) visitors' impressions of exhibit relatedness is affected by the physical layout of the museum, although less than might have been expected; (2) a basic document similarity model is surprisingly effective at capturing visitor ratings of exhibit relatedness, in a manner largely orthogonal to the relatedness derived from the museum layout.

Acknowledgements Thanks to the staff at Melbourne Museum for assistance and access to their members, and in particular Carolyn Meehan for helping with the survey design and distribution.

References

[1] L Aroyo, R Brussee, L Rutledge, P Gorgels, N Stash and Y Wang. Personalized museum experience: The Rijksmuseum use case. In *Proceedings of Museums and the Web*, page Online proceedings, San Francisco, United States, 2007.

[2] G Benelli, A Bianchi, P Marti, D Sennati and E Not. HIPS: Hyper-interaction within physical space. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, pages 1077–1078, Florence, Italy, 1999.

[3] F Bohnert, I Zukerman, S Berkovsky, T Baldwin and L Sonenberg. Using collaborative models to adaptively predict visitor locations in museums. In *Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 42–51, Hanover, Germany, 2008.

[4] R Cox, M O'Donnell and J Oberlander. Dynamic versus static hypermedia in museum education: an evaluation of ILEX, the intelligent labelling explorer. In *Proceedings of the Artificial Intelligence in Education Conference*, pages 181–188, Le Mans, France, 1999.

[5] Z Estes. A tale of two similarities: comparison and integration of conceptual combination. *Cognitive Science*, Volume 27, Number 6, pages 911–921, 2003.

[6] E Gabrilovich and S Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 1606–1611, Hyderabad, India, 2007.

[7] K Grieser, T Baldwin and S Bird. Dynamic path prediction and recommendation in a museum environment. In *Proceedings of the Workshop on Language for Cultural Hetitage Data*, pages 49–56, Prague, Czech Republic, 2007.

[8] A Mathes. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. University of Illinois Urbana-Champaign, `http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf`, 2004. Unpublished Paper, Retrieved on 11 Aug, 2008.

[9] D N Milne, I H Witten and D M Nichols. A knowledge-based search engine powered by Wikipedia. In *Proceedings of the 16th ACM conference on Conference on Information and Knowledge Management*, pages 445–454, Lisbon, Portugal, 2007.

[10] R Oppermann and M Specht. A context-sensitive nomadic information system as an exhibition guide. In *Proceedings of the Handheld and Ubiquitous Computing Second International Symposium*, pages 127–142, Bristol, United Kingdom, 2000.

[11] S P Ponzetto and M Strübe. An API for measuring the relatedness of words in Wikipedia. In *Proceedings of the ACL Demo and Poster Sessions*, pages 49–52, Prague, Czech Republic, 2007.

[12] S P Ponzetto and M Strübe. Deriving a large scale taxonomy from Wikipedia. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, pages 1440–1445, Vancouver, Canada, 2007.

[13] H Rubenstein and J B Goodenough. Contextual correlates of synonymy. *Communications of the ACM*, Volume 8, Number 10, pages 627–633, 1965.

[14] I Zukerman and D.W. Albrecht. Predictive statistical models for user modeling. *User Modeling and User-Adapted Interaction*, Volume 11, Number 1-2, pages 5–18, 2001.