

Manuscript of conference paper: Khoo, C.S.G., Teng, T.B.R., Ng, H.C., & Wong, K.P. (2014). Developing a taxonomy to support user browsing and learning in a digital heritage portal with crowd-sourced content. In W. Babik (Ed.), Proceedings of the 13th International ISKO Conference, 19-22 May 2014, Krakow, Poland (pp. 266-273). Wurzburg: Ergon Verlag.

Christopher S.G. Khoo, Toby Bing-Ren Teng, Hua-Chen Ng & Kae-Perng Wong

Developing A Taxonomy to Support User Browsing and Learning in a Digital Heritage Portal With Crowd-Sourced Content

Abstract

A taxonomy is being developed to organize the content of a cultural heritage portal called Singapore Memory Portal, that provides access to a collection of memory postings about Singapore's history, culture, society, life/lifestyle and landscape/architecture. The taxonomy is divided into an upper-level taxonomy to support user browsing of topics, and a lower-level taxonomy to represent the types of information available on specific topics, to support user learning and information synthesis. The initial version of the upper-level taxonomy was developed based on potential users' expectations of the content coverage of the portal. The categories are centered on the themes of daily life/lifestyle, historically significant events, disasters and crises, festivals, a variety of cultural elements and national issues. The lower-level taxonomy was derived from attributes and relations extracted from essays and mindmaps produced by coders after reading memory postings for a sample of topics.

Introduction

Many galleries, libraries, archives and museums (GLAMs) have set up online portals to provide public access to their heritage collections in digitized form. However, little is known about how the content in these portals should be organized to support user searching, browsing and learning. Many online cultural heritage portals adopt an organization scheme that is content-oriented or institution-oriented that makes little sense to public users, does not inform users what they can learn from the portal content, and does not support user searching for specific information or open-ended exploration.

This study is part of a project to develop a taxonomy to organize the content of a cultural heritage portal called Singapore Memory Portal (<https://www.singaporememory.sg/>). The portal was set up by the National Library Board, Singapore, to collect, organize and provide access to a digital collection of "memories" about Singapore's history, culture, society, life/lifestyle and landscape/architecture. The content of the portal is contributed by users and professional collectors, though some are contributed by the National Library from its collections. A typical memory contributed by users consists of a photograph with a few lines of text describing it; while memories from professional collectors are more elaborate and may contain video footage and an interview transcript.

The purpose of the taxonomy is to support user browsing and learning, and to be a component of the information architecture for the portal. The taxonomy is envisaged to have two parts:

- An upper-level taxonomy to provide a topical overview of the content of the portal and support user browsing;
- A lower-level taxonomy to show for a particular topic what aspects of the topic and what associated concepts are covered in the content, to supporting user learning and information synthesis.

A small user study has earlier been carried out with twelve Singapore citizens to find out what topics and subtopics they expected to be in the Singapore Memory Portal, and

what they thought were the themes related to national identity (Sriedar & Khoo, 2013). A taxonomy with eleven top-level categories was derived from the users' responses. The taxonomy can serve as the top-level browse structure for the Portal.

In our opinion, a taxonomy should also help users to synthesize an understanding of a heritage topic by indicating the types of information or concepts that are covered in the memory postings. The upper-level taxonomy is probably not sufficiently fine-grained to represent related aspects and topics. To identify more fine-grained categories, participants were asked what they had learnt from the memory postings and what interesting insights they had obtained. From the responses, we identified eleven aspects and related concepts, which can be represented as attributes and relations in an ontology for representing the types of information available on a heritage topic.

We report the results of a follow-up study to validate and enrich the upper-level taxonomy, as well as the effort to flesh out the lower-level taxonomy with topic-specific attributes and relations.

Literature Review

It is difficult for public users to search or browse cultural heritage portals because the portals are not designed to meet their needs and requirements. The knowledge organization in heritage portals is usually based on an institutional or resource perspective, for example based on the institution's collections or an in-house specialist classification system. Little is known about the users of heritage portals—their needs and behaviour (MacArthur, 2013; Xie, 2006). They are not well studied and are seldom involved in the design of heritage systems (Dobrevá et al., 2010). The subjective nature of user-generated content in portals with Web 2.0 functionalities presents additional challenges (Loverance, 2006).

We examined the search and browse interfaces of a few well-known heritage portals, including American Memory (<http://memory.loc.gov/ammem/index.html>), Europeana (<http://www.europeana.eu/>), and Australia's Trove's Pictures, Photos, Objects (<http://trove.nla.gov.au/picture>). Most portals expect the user to initiate a search by entering a search query and then filtering the results by specifying values for the facets of topic/subject, time period, place, format and source of content. These facets may also be used as top-level categories for browsing. Few portals have a well-developed browse structure.

Chaudhry and Tan (2005) described a project to develop a taxonomy for an integrated museum and archives portal. The taxonomy was developed based on terms/concepts harvested from existing taxonomies, classification schemes and finding aids within the institutions as well as obtained from external sources. The taxonomy was developed for use by non-specialist users within the heritage institutions, and not for public users. Our taxonomy is developed based on end-user expectations.

There is increasing realization among museum and heritage professionals that for heritage resources to engage the general public, they need to be organized to tell a story, embedded in a narrative context, or stimulate storytelling ("Are museums about stories or objects?", 2009). Dalbello (2004) surveyed cultural heritage digital library projects in the U.S. from 1997 to 2002, and examined the "organizing metaphors and storytelling strategies that support narrative coherence" (p. 277). He characterized narrative coherence as "the presence of a storytelling process in which order is imposed

on disjointed pieces of information and fragments of information become meaningful” (p. 277). He found that digital library projects of the time made use of topic (category), biographical approach, critical incident and temporal organization as organizing principles to improve narrative coherence. However, the narrative coherence achieved was limited and episodic. In this study, we sought to identify the conceptual relations and information organization structures that end-users might use to achieve narrative coherence with information extracted from a set of memory postings on a topic.

Development of the Upper-Level Taxonomy

In a small user study reported in Sriedar and Khoo (2013), twelve Singapore citizens were asked to indicate what topics and subtopics they expected to be in the Singapore Memory Portal, and what they thought were the themes related to national identity. An upper-level taxonomy was derived from the users’ responses, with eleven top-level categories: Place, Activity, Event, Personality, Organization/Institution, Emotion, Format, Culture/Society, Daily life/Lifestyle, National issue/Government policy, and Others.

To validate the taxonomy and enrich it further, students in the Knowledge Organization class in the Fall Semester of 2013 were set an assignment to develop a taxonomy for the Singapore Memory Portal. Each team of two or three students were asked to interview five users. Each user interview comprised the following steps:

1. Ask user to browse the portal and become familiar with the content.
2. Ask the user to respond to the following questions:
 - What topics and sub-topics do you expect to find in the Singapore Memory Portal?
 - What are the themes that you feel make up the Singaporean national identity. Examples of themes include ‘Historic Places’, ‘Food’ and ‘National Events’.
 - For each of the following topics, list a few sub-topics: Place, Activity, Event, Personality, Organization/Institution, Emotion, Format, Culture/Society, Daily life/Lifestyle, and National issue/Government policy.

Thirteen student teams submitted reports of their data collection and the taxonomies they had constructed. The information and taxonomies in the thirteen student reports were analyzed to identify alternative perspectives and alternative top-level categories, and to flesh out levels two to four of the taxonomy. The revised taxonomy has approximately twice as many categories as the earlier version. The top two levels of the taxonomy are listed in Appendix 1 (the taxonomy has up to four levels).

The categories are centered on the following six themes:

1. *Daily life/lifestyle* of the past. Relevant categories include places that are significant in people’s lives and that evoke vivid or fond memories, including places of recreation (such as parks, the zoo, aquarium, public library and tourist attractions), food establishments (such as hawker centers and food stalls), neighborhoods where people grew up in or frequented, and transportation hubs such as train stations and bus stations that were exciting places in the past. Activities associated with these places are also frequently cited categories. Related to these are memories of childhood and school days.

2. *Historically significant events*, such as Singapore's independence and Japanese Occupation during World War II. Places, buildings, personalities and activities related to these events are also noted categories.
3. *Disasters and crises* events, especially those that involve loss of lives such as war, epidemics and building collapse.
4. *Festivals* and national celebrations, including the new year celebrations of the various ethnic groups and the national day celebration. These celebrations involve many people, much effort, a variety of activities, colorful sights and an atmosphere that evoke vivid and pleasant memories, that are reinforced annually.
5. *Cultural elements*, such as food, behavior/attitude, language and humor that distinguish a group of people from others, and therefore define ethnic or national identity.
6. *National issues and government policy*-related categories. Users appear to be sensitive to national issues and government policies/regulations that affect their lives.

The taxonomy is based on user warrant—potential users' perspective of the topics that are expected to be covered in the Singapore Memory Portal and the expected terms used. We shall next check content coverage to ensure that the taxonomy covers most of the topics addressed in the memory postings. User studies can then be carried out to evaluate whether the browse structure based on the taxonomy actually helps to give the user an overview of the portal content and guide the user to the desired topic.

Development of the Lower-Level Taxonomy

The lower-level taxonomy will indicate, for a particular topic, what aspects and concepts are associated with the topic—to supporting user learning and information synthesis. Our assumption is that learning about a topic involves synthesizing information into a coherent understanding and linking together related aspects and attributes of the topic, as well as identifying relations to other topics. The lower-level taxonomy should guide the user to link related information together so as to synthesize a coherent understanding. Whereas the upper-level taxonomy can be displayed as a hierarchical menu structure for user navigation, we propose to display the lower-level taxonomy as a mindmap showing different types of information available on a topic.

In the earlier user study with twelve Singapore citizens, the participants had been asked what they had learnt from the memory postings and what interesting insights they had obtained. From the responses, we had identified the following eleven attributes or relations: place, role/participant, crisis/adverse situation, history, program item, time, personal experience, activity, subtype, event, and sentiment.

In the follow-up study reported here, we sought to determine how users might synthesize an understanding of a topic from the memory postings, what kinds of information they might select from the postings and how the selected information might be organized. As we cannot directly examine a person's understanding and information organization in his or her brain, we made use of a written essay as an explicit representation of a person's implicit understanding of a topic. We selected twelve topics in the areas of festivals, iconic/heritage buildings and personalities, and three of the authors of this paper read memory postings on the topics and wrote outline

essays (i.e. essays written in point form) on the topics. In addition, they drew mindmaps to represent their standing of the topics. The procedure is thus as follows:

- Read the memory postings on the selected topic to identify important or recurrent information and themes;
- Write an outline essay on the topic to summarize what has been learnt from the memory postings;
- Draw a mindmap to indicate the important concepts and relationships between concepts.

Each of the twelve topics was assigned to two persons. We then analyzed the outline essays and mindmaps to identify the types of information (i.e. attributes and relations) that were used in the essays and mindmaps, and organization structures used to achieve “narrative coherence.”

As expected, the attributes and relations associated with Festivals and Iconic/Heritage Buildings are very similar, whereas Personalities are associated with rather different types of information. The list of attributes and relations derived for Festivals is given in Appendix 2.

We highlight some interesting characteristics of the attributes and relations related to Festivals. The four festivals analyzed were:

- Chingay—a street parade held during Chinese New Year but has become a national event
- Hari Raya—Malay new year and celebration of the end of Ramadan
- Mid-autumn (lantern/mooncake) festival—a Chinese festival
- Seventh (Hungry Ghost) Month—a Chinese Taoist festival.

We noticed that most of the essays started with one to three sentences summarizing the basic facts about the festival, building or personality. This suggests that people have some idea what constitutes basic facts about a particular type of event or entity. From the essays, we identified the following as basic facts about festivals: name and alternative names, definition (what it is), date of origin (when first held) and origin story (how it came about), typical date (e.g., month it is typically held) and frequency (e.g., annual), associated people group (who celebrates it) and the main activities and performance items of the celebration. The basic facts may include related festivals (especially the parent festival that it is a part of) and a short interesting or amusing fact about the festival.

Another common information organization structure in the essays is the timeline—a list of dates or years, and an associated characteristic for each year (e.g., location of a parade). Some timelines list only particularly significant years associated with significant events, disasters or developments. Sometimes exact dates are not provided: events are mentioned in roughly chronological order. The writer may also summarize the development or evolution of the festival over time, or compare a past situation (activity, performance item, or policy/rule) with the present situation.

The essays often include a list of activities associated with the festival (especially performance items), as well as customs/practices, food and attire. Religious festivals will include associated beliefs.

Many memory postings include a personal element in the description: the poster’s personal experience of an activity in the festival, the scenery/sight and the atmosphere, which the poster found to be memorable. This personal element is often reflected in the

essays. The posters also often express their emotional reaction to the festival activity/scenery/atmosphere (e.g., excitement) as well as the sentiment of the festival participants and audience.

The posters as well as the essay writers are also mindful of the multicultural/multi-ethnic nature of the society in Singapore and nationalistic ideals. Many postings point out multicultural, multi-ethnic and even multi-national elements in the festival celebrations. They also highlight the spirit, attitude or cultural trait embodied in the festival celebration.

The attributes and relations derived for each type of topic can be used in various ways in a digital heritage portal. They can be used as ontological relations to link concepts in the memory postings to create a semantic network. We propose to use them instead as taxonomy categories to tag and categorize sentences in a set of memory postings on a topic. Categories with one or more sentences can be displayed in a hierarchically organized mindmap to help the user to make sense of the types of information available on the topic. If there are too many categories to display, categories can be selected based on some importance criteria, for example whether they are a subcategory of “basic facts” or have many sentences tagged with the category.

Conclusion

We have described a project to develop a taxonomy for a digital heritage portal with crowd-sourced content. The taxonomy is divided into an upper-level taxonomy to support user browsing, and a lower-level taxonomy to represent the types of information available on a specific topic. The categories in the lower-level taxonomy are semantically attributes and relations linking topics to related aspects and concepts.

The next steps in the project are:

1. To check the content coverage of the taxonomy and to extend the taxonomy to cover any gaps in the topics and types of information found in the memory postings;
2. To develop a method for automatic tagging of memory postings and sentences in the postings with taxonomy categories;
3. To develop a browse interface based on the upper-level taxonomy, and a visualization interface based on the lower-level taxonomy to display a mindmap of available types of information on a particular topic;
4. To carry out user studies to evaluate the effectiveness of the interface to support browsing, learning and essay writing.

References

- Are museums about stories or objects? (2009). *Museum Identity*, 2, 26-27.
- Chaudhry, A.S., & Tan, P.J. (2005). Enhancing access to digital information resources on heritage: A case of development of a taxonomy at the Integrated Museum and Archives System in Singapore. *Journal of Documentation*, 61(6), 751-776.
- Dalbello, M. (2004). Institutional shaping of cultural memory: Digital library as environment for textual transmission. *Library Quarterly*, 74(3), 265–298.
- Dobrev, M., McCulloch, E., Birrell, D., Ünal, Y., & Feliciati, P. (2010). Digital natives and specialised digital libraries: A study of Europeana users. In *2nd International Symposium on Information Management in a Changing World, IMCW 2010, Ankara, Turkey, September 22-24, 2010* (pp. 45-60). Berlin: Springer, 2010.
- Loverance, R. (2006). Digital preservation of user-generated content: The next challenge for museums, libraries and archives. In *Cultural Heritage On Line: International Conference, Firenze, 14-16 December*

- 2006 (pp. 232-239). Retrieved Sep 17, 2013, from <http://www.bibsonomy.org/bibtex/29a564159c38521f972044d7f315bbf66/gonaut>
- MacArthur, M. (2013). Can museums allow online users to become participants? Blog post on Museums Remixed. Retrieved Sep 17, 2013, from <http://museumsremixed.blogspot.sg/2007/04/can-museums-allow-users-to-become.html>
- Persson, Per. (n.d.) AGNETA and FRIDA: A Narrative Experience of the Web?
- Sriedar, J., & Khoo, C.S.G. (2013). A user study of the Singapore Memory Portal to derive a taxonomy for content organization. In *Proceedings of the 2013 Digital Heritage International Congress, 28 Oct-1 Nov 2013, Marseille, France* (IEEE catalog no. CFP1308W-USB, pp. 297-305).
- Xie, H. (2006). Evaluation of digital libraries: Criteria and problems from users' perspectives. *Library and Information Science Research*, 28(3), 433-452.

Appendix 1. Top two levels of the upper-level taxonomy

<p>Place</p> <ul style="list-style-type: none"> General [Singapore overall, or in general] Recreation place Tourist attraction [see also <i>Recreation place</i>] Food-related place Historic place Neighborhood Iconic/heritage building [or structure] School Shopping place Medical/health-related place Sports/exercise place Place of worship Transportation place Cemetery 	<ul style="list-style-type: none"> Non-profit organization Political party
<p>Personality</p> <ul style="list-style-type: none"> Politician Athlete Artist Entertainer Scientist Historical figure [includes pioneer] Celebrity 	<p>Emotion</p> <ul style="list-style-type: none"> Success story Sad story Overcoming adversity
<p>Activity/event</p> <ul style="list-style-type: none"> Leisure activity National event [including national celebration] Historic event Disaster/crisis Cultural/religious celebration Sports event 	<p>Culture/Society [related to Activity/event, especially Cultural/religious celebration]</p> <ul style="list-style-type: none"> General Arts Food [cuisine] Language Humor Attitude & behavior Custom & religious practice [see also <i>Cultural/religious celebration</i>] Traditional game [also under <i>Leisure activity</i>]
<p>Organization/institution [overlaps with Place, focusing on specific named organizations]</p> <ul style="list-style-type: none"> Government agency School Corporate organization Shop [see also <i>Shopping place</i>] Healthcare organization [see also <i>Medical/health-related place</i>] 	<p>Daily life/lifestyle</p> <ul style="list-style-type: none"> General [Life in Singapore in the past] Time period Life stage Entertainment Occupation Transportation Nightlife
	<p>National issue/government policy</p> <ul style="list-style-type: none"> Housing Land planning Education Healthcare Transport Natural resource Social/family Nation building Politics [including political changes] National campaign

Appendix 2. List of attributes/relations and sub-relations for Festivals

Name	Experience/memory [strong or fond memory of an experience or activity; related to <i>Associated activity/event</i>]
- Alternative name [including nickname]	
- Current name	
Function	- Visual experience [related to <i>Has scenery/sight</i>]
- Definition [what it is]	- Participant's experience
Significance	- Audience's experience
- Historical significance	- Personal significance
- Cultural significance	- Making or strengthening friendship
- Social significance [in people's lives]	- Experience with family/relatives
- Religious significance	Emotion/sentiment [including fond memory]
Typical date [when it is celebrated, e.g. month]	- Current sentiment [including nostalgic sentiment, fond memory]
Location [geographic area where it is celebrated]	Associated personality
Held at [location/building/area]	- Person officiating the opening/closing
Story	- Participant
- Origin story [Reason for holding it; how it began]	- Role of a personality
Has scenery/sight [visual impact]	- Activity of a personality
Has atmosphere [including sound]	Past situation [compared to the present]
Cultural attribute	- Past activity [related to <i>Associated activity/event</i>]
- Associated food [traditional food]	- Past performance item
- Associated attire [dress, costume]	- Past rule/policy
- Associated object	Timeline [dates/years of significant or memorable events; related to <i>Has event</i>]
- Nationalistic/multicultural element	- Date of origin [date first held/celebrated]
- Associated belief	- Date of termination
o Origin legend	- Date-significant feature
o Religious belief	o Date-location
- Spirit/attitude/cultural trait embodied	- Development over time
- National/cultural achievement	- Date-particular celebration
Associated activity/event	Related organization
- Has activity [that people do regularly at the place; personal or family activity]	- Organized by
o Children's activity	Associated people group
o Participant's activity	- Associated ethnic group
o Audience's activity	- Associated age group
o Game/play	- Associated religious group
o Religious practice	Programme item
o Custom [practice]	- Performance item
o Preparation and cleaning up	Related festival
o Interesting public behavior	Publication
- Has event [specific public/historic event, or annual event]	- Book
o Historic event	- News report
o Disastrous event [adverse situations, crisis]	- Movie
	Interesting fact
