

Visualization of Heritage Content in the Singapore Memory Portal to Support User Learning (Paper ID: 111)

Christopher S.G. Khoo, Myo Thu Ta, Kaung Pyie Win, & Chit Su San Thi

Wee Kim Wee School of Communication & Information

Nanyang Technological University

chriskhoo@pmail.ntu.edu.sg; {MYOTHUTA001; KAUNGPYI001; CHITSU002}@ntu.edu.sg

ABSTRACT

Background. This paper describes ongoing work to develop a Web application to perform summarization and visualization of memory postings in the Singapore Memory Portal, a crowdsourced online heritage portal. The motivation is to organize the information into knowledge structures based on information categories that users would use in writing essays and creating mindmaps on heritage topics.

Objective. A sentence categorization approach to text summarization was adopted in the study. The paper describes the initial sentence categorization method implemented, that makes use of cue words/phrases associated with information categories.

Contribution. A prototype Web application has been implemented that retrieves memory posts via the Web service API of the Singapore Memory Portal, and displays a mindmap-like graphical presentation of sentences organized by the information categories.

INTRODUCTION

This paper describes ongoing work to develop a Web application to perform summarization and visualization of memory postings in the Singapore Memory Portal, a crowdsourced online heritage portal set up and maintained by the Singapore National Library Board. The Singapore Memory Project is “a national initiative started in 2011 to collect, preserve and provide access to Singapore's knowledge materials, so as to tell the Singapore Story” (<http://www.singaporememory.sg/Help-Info#faqs>). It aims to “capture and document precious moments and memories related to Singapore” from individual Singaporeans, residents as well as organizations (<http://www.singaporememory.sg/Help-Info#about-us>). The portal supports posting and sharing of recollections in the form of text and digital media. A typical memory post consists of a photograph with a few lines of text describing it. There are nearly a million

posts, mainly on Singapore’s history and significant events, culture and customs, life and society, places and architecture, famous people, and national issues and government policies.

Current online heritage portals, including the Singapore Memory Portal, are organized based on records, collections and in-house knowledge organization schemes. In our opinion, the knowledge organization schemes used to organize heritage content do not support user learning and open-ended exploration. This project attempts to develop a knowledge organization scheme and a Web application that performs summarization and visualization of social media content to support user learning of Singapore cultural heritage topics.

PROBLEM STATEMENT

Figure 1 shows the main screen of the Singapore Memory Portal, which indicates that the content is organized by collection, year and location. A search using the keyword “Chingay” displays the summary search result screen shown in Figure 2. Figure 3 gives an example of a detailed memory post. Chingay is a street performance and float parade, held annually during the first weekend of the Lunar New Year period (<https://chingay.org.sg/about-chingay>).

To learn about Chingay in Singapore, the user has to read many memory posts in sequence. The memory posts are not organized in any particular way, and can be on different aspects of the topic. The information in a set of memory posts is thus disjointed and not coherently organized. To learn about Chingay, the user has to synthesize the information into a coherent understanding of the topic. We assume that this involves finding relationships among the pieces of information, and organizing the information into a knowledge structure based on the relationships. In a sense, it involves linking information together to tell a “story”.

Heritage professionals are realizing that for heritage portals to attract and engage public users, the heritage resources 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) examined the “organizing metaphors and storytelling strategies that support narrative coherence” (p. 277) in previous cultural heritage digital library projects. He explained “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). An example of an attempt to support narrative coherence in online heritage is the PATHS (Personalised Access to Cultural Heritage Space) Project funded by the European Commission to develop an interface that acts as a tour guide through the Europeana collections by using pathways—sembled sequences of heritage records with alternative routes (Hall et al., 2013; About PATHS, n.d.; About the PATHS prototype, n.d.).

In this project, we had earlier attempted to identify the knowledge structures that users synthesize to achieve a coherent understanding, or in Dalbello’s words “narrative coherence”, by asking three graduate students to read memory posts on selected heritage

topics, and to outline an essay on each of the topic. They were also asked to draw a mindmap for each topic to reflect their understanding of the topic. An example mindmap on Chingay is shown in Figure 4. The outline essays and mindmaps were analyzed to identify the knowledge structures and conceptual relations used to organize the information taken from the memory posts. The results have been reported in Khoo, Teng, Ng & Wong (2014).

We noticed that most of the essays started with one to three sentences summarizing the basic facts about the topic. This suggests that people have some idea what constitute basic facts about a particular type of event or entity. Another common knowledge structure is the timeline—a list of dates or years, and an associated characteristic for each year. Some timelines list only particularly significant years associated with notable events, disasters or developments. The writer may also summarize the development or evolution of an event or entity over time, or compare a past situation with the present situation.

The main types of information for cultural, religious and national festivals that we identified in the student essays and mindmaps are listed in Table 1. The types of information can be represented as conceptual relations that link pieces of information to the topic. These conceptual relations can thus be represented in an ontology or graphically in a mindmap. Wikipedia defines a mind map as “a diagram used to visually organize information. A mind map is often created around a single concept, drawn as an image in the center of a blank landscape page, to which associated representations of ideas such as images, words and parts of words are added.” (Mind map, 2015)

In developing a Web application to summarize and visualize the content of memory posts on a particular topic, we assume that if the information is organized according to the knowledge structures typically used by people in essays and mindmaps, it will help users to synthesize these knowledge structures in their minds and achieve a coherent understanding of a topic more quickly. We decided to display the organized information in two ways: graphically in the form of a mindmap, and textually in the form of table.

In this project, we adopt the sentence categorization approach to text summarization. The set of memory posts on a topic are segmented into sentences, and automatic sentence categorization is performed to assign sentences to the top-level information categories in Table 1. The subcategories are ignored for the moment, and used only to clarify the scope of the top-level categories. This paper reports our initial attempt to categorize the sentences into the top-level information categories, focusing on the topic of Singapore festivals, including religious and cultural festivals (e.g., new year celebrations of various ethnic groups) and national celebrations (e.g., national day parade). Other topics such as places and buildings, famous persons, events (disasters and crises), and life activities (e.g., memories of school days, family outings) are left for future studies. We collected a comprehensive list of Singapore festivals and alternative names for them from various online sources, and used the

terms to filter out memory posts on these festivals (over 7000 posts) from a corpus derived from a 2013 memory dump from the Singapore Memory Portal database. After data cleaning, we ended with 5315 posts in English language on various Singapore festivals.

SENTENCE CATEGORIZATION METHOD

A simple-minded method of sentence categorization was used that looked for cue words/phrases associated with the different information categories. To identify potential cue words/phrases, we analyzed frequently occurring words and phrases in the sample memory posts. We generated n-grams from the texts starting with unigrams (i.e. single words), 2-grams (2 adjacent words), 3-grams (contiguous sequence of 3 words) and 4-grams. N-grams with frequency lower than 5 were dropped from the analysis. The rest were manually screened for cue words/phrases that suggest a particular category of information. This was done by retrieving sentences containing the n-grams, and manually assigning the sentences to one of the information categories. If the majority of the sentences containing a specific n-gram were assigned to a particular category X, then the n-gram was accepted as a cue phrase for category X. As an example, if a sentence contains the words “to go to”, then the sentence may be categorized as *location*. “I participated in” and “to celebrate” are associated with the *name* category. Table 2 gives example cue words, a sample sentence containing each, and the manually assigned information category.

As these frequently occurring cue words/phrases are mainly functional words that can be used in many contexts, the sentence categorization accuracy is not high. The current focus of the project is to improve on the sentence categorization.

PROTOTYPE WEB APPLICATION

A prototype Web application has been implemented to submit query keywords to the Web service API (application programming interface) of the Singapore Memory Portal to retrieve memory postings. This was implemented using the Microsoft .Net framework and the MVC (Model View Controller) framework. The user interface was implemented using the JQuery JavaScript library. An example summary search result screen is shown in Figure 5. The user can select an information category on the right panel to filter out posts containing a particular category of information, with the cue words highlighted.

On clicking on the mindmap icon on the left column, a mindmap-like graphical presentation of the information is displayed (Figure 6). The sentences extracted from the retrieved memory posts are categorized into the different information categories, and linked to the topic. The graphical presentation was implemented using a data visualization JavaScript library, *D3.js*, that can run on a Web browser to display graphics using HTML, SVG and CSS (D3, 2016).

CONCLUSION

We have implemented a prototype Web application to retrieve memory postings from the Singapore Memory Portal, extract and categorize sentences into different information categories, and display the categorized sentences in a mindmap-like graphical representation. The information categories are modelled on knowledge structures and conceptual relations found in student essays and mindmaps on heritage topics. A simple-minded sentence categorization method using cue words/phrases was implemented.

Current work in the project is focused on:

1. improving the automatic sentence categorization
2. developing a clustering program to cluster sentences with similar content, to reduce repetitive information
3. investigating different ways of presenting the summarized information graphically as well as in a text summary.

Future evaluation of the Web application will include experiments to find out to what extent it supports user learning and student essay writing on heritage topics.

Table 1. Categories of information (or conceptual relations) related to *festivals* (the top 2 levels)

Name	- Personal significance
- Alternative name [including nickname]	- Making or strengthening friendship
- Current name	- Experience with family/relatives
Function	Emotion/sentiment [including fond memory]
- Definition [what it is]	- Current sentiment [including nostalgic sentiment, fond memory]
Significance	Associated personality
- Historical significance	- Person officiating the opening/closing
- Cultural significance	- Participant
- Social significance [in people's lives]	- Role of a personality
- Religious significance	- Activity of a personality
Typical date [when it is celebrated, e.g. month]	Past situation [compared to the present]
Location [geographic area where it is celebrated]	- Past activity [related to <i>Associated activity/event</i>]
Held at [location/building/area]	- Past performance item
Story	- Past rule/policy
- Origin story [reason for holding it; how it began]	Timeline [dates/years of significant or memorable events; related to <i>Has event</i>]
Has scenery/sight [visual impact]	- Date of origin [date first held/celebrated]
Has atmosphere [including sound]	- Date of termination
Cultural attribute	- Date-significant feature
- Associated food [traditional food]	- Development over time
- Associated attire [dress, costume]	- Date-particular celebration
- Associated object	Related organization
- Nationalistic/multicultural element	- Organized by
- Associated belief	Associated people group

<ul style="list-style-type: none"> - Spirit/attitude/cultural trait embodied - National/cultural achievement 	<ul style="list-style-type: none"> - Associated ethnic group - Associated age group - Associated religious group
Associated activity/event	Programme item
<ul style="list-style-type: none"> - Has activity [that people do regularly at the place; personal or family activity] - Has event [specific public/historic event, or annual event] 	<ul style="list-style-type: none"> - Performance item Related festival
Experience/memory [of an experience or activity; related to <i>Associated activity/event</i>]	Publication
<ul style="list-style-type: none"> - Visual experience [related to <i>Has scenery/sight</i>] - Participant's experience - Audience's experience 	<ul style="list-style-type: none"> - Book - News report - Movie Interesting fact

Table 2. Sample cue words and matching sentence, and the manually assigned information category

Cue words	Sentence context	Information category
during the	During the final day of Chingay, everyone was a bit sad because it was the last day that Chingay 2012 is, and after the performance during the last day.	Experience/memory
looking forward to	Looking forward to Chingay 2013!	Name
the first day	I usually visit my relatives' home on the first day of Chinese New Year.	Name
be part of	My memories of Chingay was when I get to be part of Chingay'12 and also meet up with all the performers from all the community clubs in Singapore.	Associated people group

REFERENCES

- About PATHS. (n.d.). Retrieved from <http://www.paths-project.eu/eng/About>
- About the PATHS prototype. (n.d.). Retrieved from <http://www.paths-project.eu/eng/Prototype>
- Are museums about stories or objects? (2009). *Museum Identity*, 2, 26-27.
- D3. (2016). D3 Data-Driven Documents. Retrieved from <https://d3js.org/>
- Dalbello, M. (2004). Institutional shaping of cultural memory: Digital library as environment for textual transmission. *Library Quarterly*, 74(3), 265–298.
- Hall, M. M., Clough, P. D., Fernando, S., Goodale, P., Stevenson, M., Agirre, E., ... & Bergheim, R. (2013). Information seeking in digital cultural heritage with PATHS. In *Proceedings of the 36th international ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1105-1106). New York: ACM.

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.

Mind map. (2015) In Wikipedia. Retrieved April 17, 2015, from http://en.wikipedia.org/wiki/Mind_map