Concept Maps to Support Oral History Search and Use

Ryen W. White[1], Hyunyoung Song[2], and Jay Liu[2]
[1]Institute for Advanced Computer Studies, [2]Department of Computer Science
University of Maryland
College Park, MD 20742, USA
ryen@umd.edu, hsong@cs.umd.edu, jmliu@wam.umd.edu

ABSTRACT
In this paper we describe a novel technique to support information seeking in oral history archives using concept maps. We conducted a pilot study with teachers engaged in work tasks using a prototype concept mapping tool. Results suggest that concept maps can help searchers, especially when tasks are complex.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process*

General Terms
Experimentation, Human Factors

Keywords
Concept mapping, oral histories

1. INTRODUCTION
Oral histories are a potentially powerful and inspiring source of information about past events. Video- and audio-taped interviews can yield expressive insight into how events were perceived by those who participated in or witnessed them. However, to access useful information in these archives searchers must watch or listen to media sources sequentially, a time-consuming activity. To address this problem advances in Automatic Speech Recognition (ASR) [8] technology allow it to be used to transcribe the spoken word that can later be searched. Despite these advances, oral history search systems generally offer limited support in helping searchers decide which terms or concepts to choose when composing their queries. As a result, searchers may represent their needs in such a way as to lead to ineffective retrieval.

In this paper we describe the initial steps toward providing interface support for oral history search using searcher-constructed concept maps. These maps allow searchers to build a representation of their interests that may be helpful in their search. Oral history archives are rich in named entities and inter-entity relationships that can be tagged and made accessible to a search system. Concepts maps containing these entities and relationships may therefore be a reasonable way to facilitate search and use in these archives. We have developed a prototype concept mapping tool that adopts a "berrypicking" search model [1], allowing searchers to store information fragments on a canvas, and link these fragments interactively to create a concept map. We present the tool, a pilot evaluation of its utility, and conclude with a brief discussion of evaluation findings.

JCDL'06, June 11–15, 2006, Chapel Hill, North Carolina, USA.

2. CONCEPT MAPPING TOOL
Concept mapping is a technique for visualizing the relationship between concepts by labeled nodes and links widely used to support the attainment of pedagogical goals, foster creativity, and represent the structure of knowledge [5,7]. They have already been used to support search and use, but generally to represent relevant information in a collection [6], Anomalous State of Knowledge (ASK) structures derived from problem statements elicited from searchers [2], or the construction of knowledge models [4] *The contribution of this paper is the support for search and use in oral histories using searcher-constructed concept maps.*

The Survivors of the Shoah Visual History Foundation (VHF) has created an archive of video interviews from over 52,000 Holocaust survivors, liberators and rescuers. Entities present in the archive (e.g., people, places) have been tagged as part of manual indexing by trained VHF indexers. During this process, interviews were divided into topically coherent *segments* (approximately 3 minutes long), and assigned keywords from a thesaurus developed by the VHF for this purpose that contains concepts, people, and place names. This indexing is used by the tool we have developed as the basis for concept map construction. In a similar way to systems such as DLITE [3] this tool takes the form of a two-dimensional workspace onto which snippets of information can be dragged, dropped, directly manipulated, linked, and used.

The concept mapping tool is part of an oral history search interface divided into two components: "retrieve" and "store/use." The "retrieve" component allows searchers to submit textual queries and retrieve lists of segments for examination. Segments are ranked based on the number of query terms that appear in their portion of the ASR transcript or keywords assigned to them during indexing. Searchers can drag entities such as people, places, and segments from anywhere on the "retrieve" portion of the interface (e.g., any result list entry), and drop them directly onto the concept map canvas in the "store/use" part of the interface, where they appear as *nodes*. Figure 1 shows an example of a concept map. It has been constructed interactively, and maintains its state for an entire search session, or longer if explicitly saved by the searcher.

Searchers can annotate nodes by attaching notes to them. Named directed links (or relationships) can be formed between two nodes (shown in Figure 1).[1] Link names vary depending on node attributes. For example, "Lived in" may be appropriate for person-place relationships, but not place-segment relationships. The next section describes a pilot evaluation of the tool.

[1] Additional functionality visible in Figure 1, such as "Find more like…" and "Relate…" was disabled for the pilot evaluation, to allow us to focus on concept map construction.

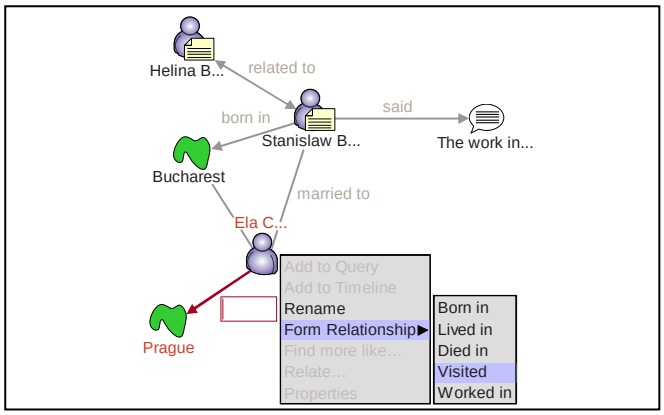

Figure 1. Concept map created using tool (names anonymized).

3. PILOT EVALUTION

We conducted a pilot study with six middle and high school teachers. We asked them to produce two stories for their students using the VHF archive: a "simple" story and a "complex" story. The simple story (i.e., *Build a story about how wartime affects peoples' lives*) was intentionally general and focused on gathering information about people; the complex story (i.e., *Build a story about what happened between people in Warsaw ghettos*), was intentionally specific and focused on gathering information about relationships between people. Story construction was felt to be a reasonable approximation of a task teachers would engage in with the archive, and afforded us an opportunity to assess the utility of the concept mapping tool.

The study used a mixed-design: three teachers used the oral history search system with concept mapping tool for their two tasks and three used the system without the tool. This study design was chosen because a significant learning effect was observed between the two conditions during a preliminary test, and we could not guarantee its removal through counterbalancing. We used a subset of the archive comprising approximately 400 English interviews that had been manually indexed as described in the previous section. Resources made available included the ASR transcripts for each interview (35% word error rate), interviewee audio, and a tool to browse the VHF thesaurus. The search system (i.e., the "retrieve" component) provided capabilities to search transcripts and assigned keywords. Participants were asked to write one or two paragraphs of a story for each task on a sheet of paper, using five sources of evidence taken from search results. Participants were asked to enumerate the sources in the paragraph as they wrote. A score (between 1 and 5) was computed for the stories based on a count of the number of sources used. Participants' satisfaction with their stories was measured in questionnaires with Likert scales and semantic differentials.

The experiment ran for up to one-and-a-half hours. The experiment began with 15 minutes training on the VHF archive and ASR technology. Four participants were aware of VHF and its mission before the experiment, but only one had seen ASR transcript before. The next 60 minutes were spent on two tasks with or without the concept mapping tool (depending on the experimental design). Participants completed a post-experiment questionnaire to elicit their opinion on the tool and task outcomes.

4. FINDINGS AND DISCUSSION

Initial analysis indicates that in constructing complex stories compared with simple stories, participants generally:

- Built richer concept maps, with more nodes and links;
- Explored the VHF thesaurus in greater detail, and used some of the thesaurus entries in their search statements, and;
- Took longer to reach task completion. All participants used the entire 30 minutes for the complex task but used on average 19 minutes for the simple task.

Although participants complained about the accuracy of the ASR transcripts in all cases, the search, audio playback, and thesaurus browser used in the study appeared sufficient for both tasks. Participants who used the concept maps for complex tasks built stories containing significantly more sources of evidence ($\underline{M}$ = 3.67) than those without concept maps ($\underline{M}$ = 3.33),[2] the same was not true for the simple tasks.

Participants commented that building concept maps enlarged their working memory, supported their highly iterative searching style (for complex stories), and led to the creation of good stories. This is promising, especially when we consider that research on automatically detecting entities in archives such as this makes offering concept maps possible without incurring the costs of a large manual indexing effort. In future work we will test higher-fidelity prototypes, investigate the use of automated entity detection, and test concept map effectiveness in other domains. There may also be value in using concept maps for applications other than search and use, such as story construction.

5. REFERENCES

[1] Bates, M. J. (1989). Design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-424.

[2] Belkin, N.J. and Kwasnik, B.H. (1986). Using structural representation of anomalous states of knowledge for choosing document retrieval strategies. In *Proc. of SIGIR*, 11-22.

[3] Cousins, S., Paepcke, A., Winograd, T., Bier, E.A., and Pier, K. (1997). The digital library integrated task environment (DLITE). In *Proc. of ICDL*, 142-151.

[4] Leake, D.B., Maguitman, A., Reichherzer, T., Cañas, A.J., Carvalho, M., Arguedas, M, Brenes, S., and Eskridge, T. (2003). Aiding knowledge capture by searching for extensions of knowledge models. In *Proc. of K-CAP*, 44-53.

[5] Novak, J. D. and Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press: Cambridge, U.K.

[6] Richardson, R. and Fox, E.A. (2005). Using concept maps in digital libraries as a cross-language resource discovery tool. In *Proc. of JCDL*, 256-257.

[7] Shneiderman, B. (1999). User interfaces for creativity support tools. In *Proc. of Conference on Creativity and Cognition*.

[8] Whitaker, S., Choi, J., Hirschberg, J., and Nakatani, C.H. (1998). What you see is (almost) what you hear: Design principles for user interfaces for accessing speech archives. In *Proc. of ICSLP*.

[2] Two-way mixed-design ANOVA $\underline{F}(1,4) = 10.05$, $\underline{p} = .03$ ($\alpha = .05$)