

An Interface to Search Human Movements Based on Geographic and Chronological Metadata

Wilma Bainbridge
Montgomery Blair High School
Silver Spring, MD 20901, USA
noin_of_oz@yahoo.com

Ryen W. White[†] and Douglas W. Oard^{†,‡}
[†]Institute for Advanced Computer Studies, [‡]College of Information Studies
University of Maryland
College Park, MD 20742, USA
{ryen,oard}@umd.edu

ABSTRACT

Historians and scholars can better understand historic events by studying the geographic and chronological activity of individuals who witnessed them. A lack of adequate tools to help users study these activities can hinder the process of learning and discovery. In this paper we present an interface to address this problem that contains three components: a map, a timeline, and a text representation of a survivor's movements. These components simultaneously provide query input (where users can specify their needs) and dynamic results display (where users can immediately see the effect of their decisions). The results of a pilot study show that users reacted positively to the interface.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]

General Terms

Human Factors, Experimentation.

1. INTRODUCTION

The Survivors of the Shoah Visual History Foundation has assembled an extensive collection of interviews with survivors of the Holocaust. More than 50,000 survivors provided video testimonies in 32 languages, chronicling their experiences before, during and after the Second World War. This large collection of first-hand experiences has been extensively indexed by subject matter experts, greatly increasing its potential value to historians and other scholars. Scholars often use oral histories to complement written records, tell of undocumented events and provide explanations, interpretations and perspectives [1]. One aspect of that broad effort is characterization of human migration, both voluntary and forced. One of the goals of the MALACH (Multilingual Access to Large Spoken Archives) project is to develop interface tools that will help scholars understand and analyze the human movements that are described in the Shoah Foundation's collection.

The geographic and temporal nature of the data clearly calls for presenting more than a simple ranked list of retrieval results. The interface design presented here therefore uses graphical representations of the time and location to facilitate interaction, serendipity and learning. The results update dynamically in response to inputs from the user without requiring an explicit "search" action. Our interface uses a zoomable map display as in [2], but it has been designed to depict individual experiences, rather than describing historical events. To this end, the interface only shows data for one survivor at a time.

Copyright is held by the author/owner(s).
SIGIR '05, August 15-19, 2005, Salvador, Brazil.
ACM 1-59593-034-5/05/0008.

2. INTERFACE DESIGN

A search on the system is a two-stage process. Stage One begins with query input (as free text) and ends with identification of testimonies that match the query. The manually assigned metadata for each segment includes segment boundaries, an average of five concept keywords drawn from a custom-built thesaurus, at least one pre-coordinated location-time pair, a carefully edited three-sentence summary, and names of all mentioned individuals (in normalized form). Testimony-scale metadata (e.g., the survivor's name and birthplace) is also searched. The searcher can re-sort the initial results by speaker name, speaker date of birth, the names of mentioned individuals, locations described in the segment, or the time frame described by the segment (typically with one-year granularity) and select a speaker whose movements are of interest. The result of that selection (Stage Two of the process, depicted in Figure 1) is the focus of this paper. The interface design is based on the principle of intentional redundancy, depicting search results in three overlapping ways: a zoomable *map*, a *timeline*, and a *results list*.

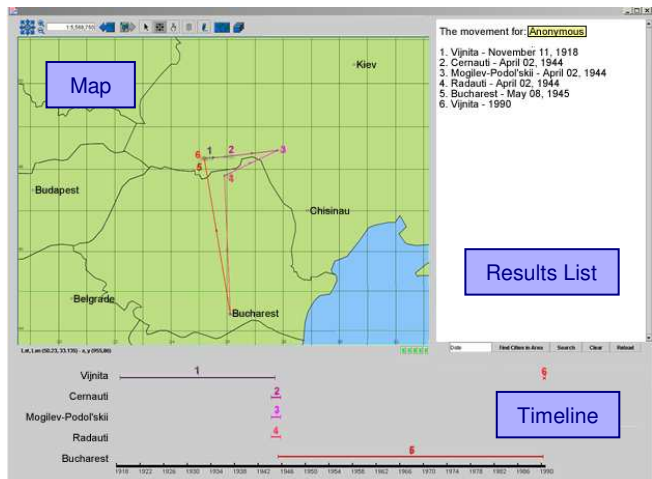


Figure 1. Search Interface.

Map: The main purpose of the map is to show the movement of a selected survivor between geographic locations, represented by connected line segments. Each location is numbered indicating the order in which the locations were visited (based on automatic decomposition of the pre-coordinated location-time pairs). Arrowheads indicating the direction of movement are included. Color ("red-shift") is also used to visually convey the passage of time, the initial line segment is rendered in a (short wavelength) violet hue, proceeding through the spectrum to a (long

wavelength) red hue at the end. In addition to use as a display, the map can serve as a source of spatial queries; users can mark a region and then locate all survivors that passed through a location in that region. The timeline allows further refinement, limiting a query to those who were there during a specific period.

Timeline: The timeline organizes events chronologically (i.e., as-lived, rather than as-told). Places visited by the person are listed on the left side and ranges corresponding to time intervals are drawn to the right of the location names, vertically aligned with the timeline below. The map and timeline are coupled; the numbered timeline segments correspond to the numbers on the map, and the timeline colors correspond to the colors on the map. Using a combination of the map and timeline, users can get a feeling for the distances traveled and the time spent doing so.

Results List: When the second stage search is initially completed, the survivor's name and movements are shown in the list in the format *place, time*. The results list may be searched by *place* (i.e., find all survivors who went to a particular place) or by *time and place* (i.e., find all survivors in the same place at the same time).

The metadata characteristics pose some design challenges. This is evident in Figure 1 where the survivor is listed as being at three locations on April 2, 1944. In actuality, this is the start date of a period that is correctly depicted in the timeline (our present text display was optimized for single dates (typically full years), which are more common in the collection). Even the timeline contains ambiguity however; the indexed granularity cannot determine the order in which those three locations were visited (or, indeed, whether repeated visits occurred during that period).

3. PILOT STUDY

We recruited 4 subjects for a pilot evaluation to test the usability of the interface. All were graduate students studying either History (2), Information Science (1) or Computer Science (1). The pilot study was designed to gather qualitative data on subject perceptions of the interface and of the search process. Two of the four subjects (one historian and the information scientist) had some knowledge of the collection's metadata structure before participating. Short task scenarios were devised that required subjects to use the interface to: (i) find many people who were in certain locations at certain times (e.g., *Find survivors who were in Bucharest on May 8 1945*), (ii) track the location of one survivor over time (e.g., *Select a survivor and describe their movements*) and (iii) comment on differences between two given survivors that relate to their movements (e.g., *Select two survivors and describe the differences in their movements*). Sessions lasted about 30 minutes, and the subjects were asked to "think aloud" as they searched. We observed them and made written notes on their interaction and comments that may influence design decisions.

4. FINDINGS AND DISCUSSION

Overall, subjects responded positively to the interface, although there was a noticeable divide between the historians and others. Generally, the subjects found the different forms of result presentation useful, especially when there was uncertainty over the location. The map was the most commonly used form of query input by all subjects as they felt it encouraged exploration, was not confusing and was a relatively robust way to refine the result set. All subjects felt that the timeline should be used only in combination with the map and/or the results list because the information it displayed was occasionally ambiguous.

The two historians generally wanted to compare a number of individuals over a period of time, arrange people by status (such as in a camp, in hiding, or in transport) and wanted to be able to compare an individual to others in the same situation. In essence, the historians wanted to construct a story. In contrast, our computer and information scientists wanted to narrow their search using query terms rather than using the other interface components to directly manipulate the results display. The computer and information scientists generally wanted to find small pieces of relevant information. The historians reacted more positively to the interface, probably because they were more likely to conduct exploratory searches, something they felt was well supported in this interface.

The interface described in this paper includes an unusually close coupling between output and input, reflecting the way in which answers to initial questions naturally facilitate the generation of new questions. At present, each user interaction is treated in isolation and leads directly to output at the interface. Subjects liked this as it allowed them to immediately see the outcome of their actions and follow new search directions where appropriate. However, the computer and information scientists both felt that a useful addition would be the ability to perform a series of related interactions, *then* see the output. They felt this would allow them to construct complex queries and refine their search more rapidly.

Our study also found that different interface components (or combinations of components) were useful for different searches. When subjects searched for survivors in a given place on a given date, they used only the results list. When asked to track the location of an individual over time, the map, the timeline and the results list were all used. Finally, when asked to compare the movements of two survivors, subjects used only the map and the timeline. Since the interface should reasonably be expected to handle all such searches we felt it would be wise to show the map, timeline and results list as the default display. However, two subjects suggested that it may also be helpful to give users control over which components are displayed at the interface and the proportion of the interface devoted to each component to reduce potential distraction and interface clutter.

Work is ongoing to simultaneously display the movements of multiple survivors and to visually represent uncertainty that results from metadata characteristics, including future sources of metadata (e.g., extraction of location names from automatically produced transcripts). Our initial pilot study suggests that our basic design is suitable for use by well trained scholars, and with further enhancement we hope to develop these ideas into a set of effective and intuitive interfaces for exploration of a broad array of oral history collections by scholars, teachers, and others.

5. ACKNOWLEDGMENTS

This work has been supported in part by NSF IIS Award 0122466.

6. REFERENCES

- [1] Gustman, S., Soergel, D., Oard, D., Byrne, W., Picheny, M., Ramabhadran, B. and Greenberg, D. (2002). Supporting access to large digital oral history archives. *Proceedings of the Joint Conference on Digital Libraries*, 18-27.
- [2] Smith, D. (2002). Detecting and Browsing Events in Structured Text. *Proceedings of the 25th Annual International ACM SIGIR Conference*, 73 – 80.