Biography as Events in Time and Space

Fredric Gey
University of California, Berkeley
UC Data Archive & Technical Assistance
Berkeley, CA 94720-5100
510-643-1298

gey@berkeley.edu

Ryan Shaw, Ray Larson, Barry Pateman
University of California, Berkeley
School of Information
Berkeley, CA 94720-4700

510-642-6046

{ryanshaw, ray@ ischool.berkeley.edu}

ABSTRACT
In digital humanities projects, particularly for historical research and cultural heritage, GIS has played an increasingly important role. However, most implementations have concentrated on displays which ignore the temporal dimension or express it as multiple snapshots for fixed or periodic points in time. Our project concentrates on historical biography and expresses a biography as a sequence of life events with in time and space. We utilize named entity recognition and extraction to automatically mark up biographies so that they can be displayed as dynamic maps. In so doing, contextual features and related happenings and people can be overlaid to facilitate serendipitous discovery of unanticipated and seemingly unrelated connections.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *abstracting methods, linguistic processing*

General Terms
Experimentation

Keywords
Geographic search, Biographical text processing, Digital humanities.

1. INTRODUCTION
Cultural heritage, history, and social sciences are fundamentally about human activity. Everyone is interested in what other people do and have done. Biographies are regularly among the best-selling books. Not only History, but also Geography and most other subjects can come alive in the travelogues, journeys of discovery, and the life-stories of those involved. But mere narrative is not enough. Understanding the context differentiates education from memorizing. Building and supporting a community of learners needs more than facts. Understanding the circumstances of people's actions illuminates their lives. However, there is a significant gap in the infrastructure developed by libraries, museums, and publishers in this area. We have

standards for the computerized handling people's names, but not for their lives. Take an example from American history, the US frontier hero Davy Crockett – how many people know that after serving as a Colonel under General Andrew Jackson, he was twice elected a congressman from Tennessee? Knowing these facts leads to other contextual questions: What was the population of Tennessee when Crockett was a congressman? Who were his contemporaries in Congress (e.g. Daniel Webster)? Connecting contexts and people in biographical texts is the subject of this demonstration for ACM GIS 2008.

2. BIOGRAPHY AS EVENTS
We define a biography B as an ordered sequence of events (E_0 … E_i, … E_L). An *Event* E_i is a 4-tuple (*Activity, Date-range, Place, Other people*). A *Date-range* consists of a start-date and an end-date where *Date* can be either year or year, month, day combination. It is the time during which the event happened. *Place* is the location where the event happened and is usually represented by latitude and longitude coordinates. Activity is a description of what happened. Two important Activities are fixed: for E_0 the activity is "Born", for E_L, the activity is "Died" (the last activity in a biography).

3. DEMONSTRATION CONTENT
We will show mapping of biographies and historical events and their demographic contexts being developed in two closely-related projects: *Bringing Lives to Light: Biography in Context* (http://ecai.org/imls2006) and *Contexts and Relationships: Ireland and Irish Studies* (http://ecai.org/neh2007). The conceptual framework for this exploration is found in [1, 2, 4].

The presentation will include prototype interfaces and an "under-the-hood" look at pioneering XML event markup standard proposals which facilitate the demonstration. We have developed markup grammars using the Gate technology [2] and, in particular, the Golden Gate XML editor from Karlsruhe [4] to automatically markup biographies for time and place.

In addition to working with USA congressional biographies, we have followed the timeline of the great anarchist thinker and lecturer Emma Goldman as she toured the USA from 1910 to 1916, lecturing about drama, status of women, and anarchists such as Francisco Ferrer y Guardia, the Spanish-Catalan anarchist. An example screenshot is shown below in Figure 1.

We developed a timeline marked up with geography using the XML Atom format (http://en.wikipedia.org/wiki/Atom) which allows us to directly input the data file into Google Maps and provide a geographic user interface to the Goldman travels and lectures:
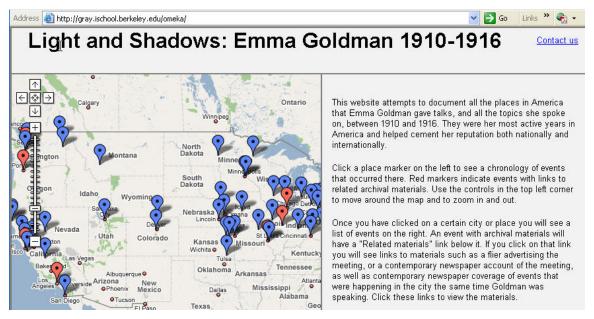
Figure 1: Travels of Emma Goldman Screenshot

4. FUTURE DIRECTIONS

Our technology for marking up and extracting events from biographies will be extended to tools for editorial markup of digitized cultural materials and tested upon a body of digitized resources from Irish historical and cultural journals and documents being prepared by collaborators at the Centre for Data Digitisation and Analysis of Queens University, Belfast, Northern Ireland. We expect to be challenged by the complexities of spelling variants for Celtic and middle English person and place names, as well as inexact time periods. The project will work on the development of a gazetteer of historical Irish place names and their latitude and longitude coordinates. We also hope to connect the textual information to statistical data from historical Irish censuses.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Buckland, M., Chen, A., Gey, F., Larson, R, Mostern, R., and Petras, V. 2007. Geographic search: catalogs, gazetteers, and maps. College & Research Libraries (Sept 2007): 376-387.
http://www.ala.org/ala/acrl/acrlpubs/crljournal/collegeresearch.cfm

[2] Buckland, M., Gey, F., Larson, R., Access o Heritage Resources using What, When, and Who. Museums and the Web 2007: The International Conference for Culture and Heritage Online, April 11-14, 2007, San Francisco.
http://www.archimuse.com/mw2007/papers/buckland/buckland.html

[3] Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, USA.

[4] Larson, R Geographic information retrieval and spatial browsing. In GIS and Libraries: Patrons, Maps and Spatial Information, pages 81–124. UIUC - GSLIS, Urbana-Champaign, IL, 1996.

[5] Sautter, G., Padberg, K and Tichy, W. 2007. Empirical Evaluation of Semi-Automated XML Annotation of Text Documents with the GoldenGATE Editor, in Proceedings of ECDL 2007, Budapest, Hungary, 2007.