Bringing Lives to Light: Lives and Event Representation in Temporal and Geographic Context

Ryan Shaw
School of Information
University of California, Berkeley
Berkeley, California, USA, 94720-4600
ryanshaw@ISchool.Berkeley.EDU

Ray R. Larson
School of Information
University of California, Berkeley
Berkeley, California, USA, 94720-4600
ray@ISchool.Berkeley.EDU

ABSTRACT

Our demonstration system consists of a set of tools for identifying life events in biographical texts and linking them to relevant contextual resources.

Categories and Subject Descriptors: H.3.7 [Digital Libraries]: standards, systems issues, user issues; H.3.5 [Online Information Services]: web-based services

General Terms: Design

BIOGRAPHY AND LIFE EVENTS

Cultural heritage, history, and the social sciences are fundamentally about human activity. Everyone is interested in what other people do and have done. Chronological, geographical and biographical data lend themselves naturally to being connected: an event is associated with a place, a time and potentially with particular people; places are associated with different events and people; and individual people are also associated (in a variety of ways) with different places and events.

Life-events in sequence constitute a narrative that can engage interest and spark inquiry. History, geography, and most other subjects can come alive in the travelogues, journeys of discovery, and the life-stories of those involved. Science can be explained through the work of scientists. Engineering is routinely explained through the heroic struggles of inventors.

But mere narrative is not enough. Understanding the context of these life events differentiates education from memorizing. It is understanding the circumstances of people's actions that illuminates their lives, but there is a significant gap in the infrastructure developed by libraries, museums, and publishers in this area. Our objective in this project is to design, demonstrate, and evaluate techniques that would bring lives to light by revealing them in their contexts.

BIOGRAPHICAL MARKUP AND TOOLS

Our system consists of a set of tools for enhancing digital documents or document surrogates by identifying references to or representations of people, places, or events, disambiguating these references by linking them to identifiers from naming authorities, and using the disambiguated references to provide useful information in context as well as to link to related resources available on the web.

JCDL'08, June 16–20, 2008, Pittsburgh, Pennsylvania, USA.
ACM 978-1-59593-998-2/08/06.

For textual documents, we use standard natural language processing techniques to identify named entities: the names of people, organizations, places, or events. Named entities can also be identified manually in both texts and images using a web browser-based annotation tool. We will demonstrate how these tools can be used to enhance a variety of different types of document, including articles from the Citizendium open encyclopedia project and the Biographical Directory of the U.S. Congress, scanned page images from a digital library of Irish Studies resources, and various documents from the Emma Goldman Papers Project.

Once named entities have been identified, our tools provide a simple interface for resolving ambiguities by linking names to the appropriate authoritative identifiers. For person and organization names, these include Library of Congress and Deutsche Nationalbibliothek authority files and WorldCat Identities URLs. For place name we use identifiers from the GeoNames geographical database. The system is extensible so that new sources of identifiers can be added as they appear; for example we are experimenting with using Freebase URLs to identify historical events, for which no established naming authority currently exists.

After entities have been unambiguously linked to autoritative identifiers, we can use these identifiers to find additional information about those entities. For example, a linked map and timeline displayed alongside a document may indicate the locations of places and the locations and times of events referenced in the document. Access to unstructured information and related documents can be provided by using identifiers or aliases to construct dynamic links representing queries on appropriate reference sources or special collections. Finally, URLs of single static resources can also be added to named entities so that links to them will appear whenever those entities are references in browsed documents.

We are experimenting with a number of ways of making our linked data available, including RDF/XML, enhanced Atom feeds, and metadata embedded in HTML pages. In all cases we are reusing standard metadata vocabularies wherever possible, so that other systems may easily integrate, and ideally contribute to, our contextual information.

ACKNOWLEDGMENTS

The work presented draws on two projects partially supported by Institute of Museum and Library Services and the National Endowment for the Humanities: *Bringing Lives to Light: Biography in Context* (http://ecai.org/imls2006/) and *Context and Relationships: Ireland and Irish Studies.*