Open Identification and Linking of What, Where, When & Who

Ryan Shaw & Michael Buckland
School of Information
University of California, Berkeley
Understanding an item in a digital library or museum requires context: information about the topics, places, events, and people to which the item is related.
Students, journalists and investigators traditionally discover contextual background by asking "the four Ws": what, where, when and who. The DCMI Kernel Metadata Community has recognized the four Ws as fundamental elements of descriptions.* Making better use of metadata to answer such questions via links to appropriate resources has been our focus in a series of research projects over the past few years:

**Support for the Learner: What, Where, When, and Who**
http://ecai.org/imls2004/

**Bringing Lives to Light: Biography in Context**
http://ecai.org/imls2006/

**Context and Relationships: Ireland and Irish Studies**
http://ecai.org/neh2007/

Currently we are building a system to enable readers to relate any topic, place, event or person mentioned in a text to explanatory resources. It is being developed with two corpora: a variety of biographical texts with dense mentions of people, events, and places, and a large collection of books, journals and manuscripts relating to Irish culture and history.
Our system consists of tools for referring to topics, places, events or people, disambiguating these references by linking them to unique identifiers, and using the disambiguated references to provide useful information in context and to link to related resources.

We use identifiers from both established and emerging naming authorities. Central to our strategy is to move from appearances in a text to naming authorities to the construction of links for searching or querying trusted resources.
Ordinarily work on mapping between vocabularies focuses on similar vocabularies, e.g. two different place name gazetteers. We are interested in also mapping among dissimilar vocabularies, such as those found in different genres of reference works (now made available as web services rather than paper books). Semantic associations can be established by vertical vocabulary mapping and contextual associations can be established through horizontal associations.
Names of historical events and time periods have not been accorded the same kind of treatment as person names and place names. Thus we have identified a need for a genre of terminology service that we call an event directory. Events are interesting not only as another kind of named entity that can be used to index documents, but also as rich sources of syntagmatic relations among other kinds of entities.

Identifying kinds of events
battle, treaty, election, earthquake

Relations to places: where did it happen?
Spain, Texas, Eastern Europe, Constantinople

Relations to time: when did it happen?
May 1976, sometime before 1811, Stone Age

Participation: who and what were involved?
Emma Goldman, Standard Oil, the Mona Lisa
We are taking two complementary approaches to building an event directory for use with our collection of Irish texts:

- A top-down approach, using FAST subject headings, Wikipedia, and chronologies found in the appendices of general Irish history books to identify major events in Irish history and their names.

- A bottom-up approach, mining our collection of scanned texts for common co-occurrences of place names and dates that may identify events of interest.


We gratefully acknowledge the Institute of Museum and Library Services and the National Endowment for the Humanities for funding and supporting our work.