A Query Interface for an Event Gazetteer

Robert B. Allen
College of Information Studies
University of Maryland
College Park, MD 20742
allen@acm.org

ABSTRACT

We introduce an event gazetteer which stores and presents "locations in time". Each event is coded with attributes of event type, location, actor, and beginning and ending times. Events can also contain sets of other events. This paper reports the development of an interface for generating searches to these "part-of" relationships. For instance, we can search for all named Battles in the event database which occurred during the Civil War. Ultimately, we envision a flexible, broad-based service that is a resource for users ranging from students to genealogists and researchers interested in historical events.

Categories & Subject Descriptors: H.1.2 User/Machine Systems - Human factors, K.3.1. Computer Uses in Education
General Terms: Design, Human Factors
Keywords: Events, Causation, Metadata, Timelines.

1. THE VISION

Electronic spatial gazetteers provide a comprehensive resource of place names and descriptions [5]. While there is wide recognition that time and space are interrelated, there is no comprehensive resource for describing events analogous to the online spatial gazetteer. There have been previous prototypes of timeline interfaces (e.g., [1, 8]) and several temporal schema have been proposed. One of these, TimeML [4], emphasizes the richness of temporal descriptions in natural language documents. The History Events Mark Up Language [9] provides an XML Schema with links to historical documents and a non-interactive timeline interface.

We report the development of a robust yet extensible event gazetteer. The creation of, search for, and presentation of event schemas are integral to this system. It also allows tagging with flexible metadata lists, hierarchical grouping of events into components, the extension of timelines to "actors", and presentation of alternative views of events to represent the opinions of different users.

There are many possible applications for a generalized framework for describing events. The most obvious involves history. A student might browse event descriptions about the Civil War. With an interactive timeline interface that the student might compare different streams of events. For instance, Acts of Congress during the Civil War could be compared to military events in the War. An event gazetteer could be the basis of for browsing newspapers stories (e.g., [3, 7]). It could also manage events presented as part of scientific explanations.

JCDL'04, June 7–11, 2004, Tucson, Arizona, USA.

2. EVENTS AND EVENT SETS

2.1 Events

Naturally, an Event Gazetteer should be focused on presenting events, but "events" are not easy to define. We might agree that the inauguration of a President is an event. But what about the speech that is given as part of the inauguration? What about speaking the first word of that speech? In this project, we focus on named events and subevents.

Events have attributes including names, owners, descriptions, location, time (beginning and ending times), type, and actors. In the current system, each of these is a text string with the exceptions of time, which is an Oracle Date Type, and Actor, which is a separate high-level entity.

2.2 Events Sets

Each event can contain a set of other events. These eventsets may indicate that one event is "part-of" another event. Thus, Pickett's Charge is a part of the Battle of Gettysburg and the Battle of Gettysburg is a part of the Civil War. This part-of relationship is structured with predefined lists of Event Type metadata. In those a War can be composed of Battles and Battles can be composed of Attacks. To prevent cycles, the events included in these sub-sets must be shorter than and occur during the parent event. These sets of events can also be useful for creating timelines and our system allows events in an event set to be interconnected by user-defined Annotations.

2.3 Searches and the Query Interface

An event server has been implemented with JSP and Oracle. The implementation currently offers basic functionality for creating, storing, searching, and deleting Events. Simple queries such as "Find all Battles in Virginia" are completed by requesting Event Type="Battle" and Location="Virginia". With the complex part-of Event-Type relationships we might consider a search such as "Find all Civil War Battles in Virginia". That is: Parent Event Name="Civil

War", Child Event Type="Battle", and Child Event Location="Virginia". The query interface produces SQL strings which are, in turn, submitted to the database. In the current system the part-of relationships are implemented with Event Types.

Figure 1: Query interface for searches on events and event sets.

3. FUTURE WORK

Our vision is for a flexible architecture with a broad range of event-authors and searchers. The initial system reported here needs to be extended to include defining user groups and permissions. Similarly, we will also introduce graphical hypertextual interaction with applets. We are considering implementing a true object-oriented model (e.g., [6]) and temporal data model (e.g., [10]). In the current version, temporal order is enforced with post-processing the result sets. In the following description, we elaborate on two topics for future work: metadata and relationships among events.

3.1 Metadata

Each of the metadata elements we have introduced can be elaborated. Events have *owners*, the user who created them. *Locations* are text descriptions but they could also be coordinates or links to identification numbers in spatial gazetteers such as the ADL-ID numbers. Locations may also include non-contiguous locations, extra-terrestrial locations, and even fictitious locations. *Times* could periodic or they could be in different time scales such as geologic time. Structurally, *Actors* are similar to events, they can have attributes and be associated with other sets of events. Actors should also have Roles; but the value of adding roles needs to be weighed against the complexity.

We have emphasized *Event Types* here because they show the part-of relationship. In the current system Event Types are predefined. In future systems, the event type lists should be created and managed by users and sets of Event Types could be bundled, perhaps by cultural context. Indeed, other gazetteer attributes such as hierarchical spatial political structures could take advantage of a similar search techniques. Moreover, we need to consider developing scripts which are collections of interrelated objects and events.

We have employed an ad hoc, simplified metadata system but other much more complete metadata systems have, of course, be developed. For instance, the LC Classification Schedules include Chronology Tables and the Name Authorities describe attributes of individuals. We are exploring the use of those records.

3.2 Causation

Beyond the part-of relationships among events, more needs to be done for other types of associations among events. The most challenging of these is probably causation. We model causation as an event that causes a transition between two other events. This would be implemented as triples of events.

A further goal for this system is to facilitate clear presentation to users of causal threads to support the development of stories about events (e.g., the evolution of the dinosaurs into birds). Some of our earlier work described stories as chains of events [2] and we expect to re-implement that in this more general framework.

4. ACKNOWLEDGEMENTS

This work is funded in part by NSF award #0329111 for "Interacting with Threaded Event Scenarios". Several students have participated in this project including Neetu Ahuja, Kapil Kher, Manav Kher, Jun Luo, Craig Murray, and Yejun Wu.

5. REFERENCES

[1] Allen, R. B. Timelines as information system interfaces. In *Proceedings International Symposium on Digital Libraries* (1995), pp. 175–180. http://raven.umd.edu/ rba/PAPERS/TL/isdl.pdf.

[2] Allen, R. B., and Acheson, J. A. Browsing structured multimedia stories. In *Proceedings ACM Digital Libraries Conference* (2000), pp. 11–18.

[3] Allen, R. B., and Schalow, J. Metadata and data structures for the historical newspaper digital library project. In *Proceedings ACM Conference on Information and Knowledge Management* (Nov. 1999), pp. 147–153.

[4] Ferro, L., Gerber, L., Mani, I., Sundheim, B., and Wilson, G. Instruction manual for the annotation of temporal expressions. Tech. rep., MITRE Washington C3 Center, McLean, Virginia, 2002.

[5] Hill, L. Core elements of digital gazetteers: Placenames, categories, and footprints. In *Proceedings European Digital Library Conference* (2000).

[6] Murray, N., Paton, N., and Goble, C. Kaleidoquery: A visual query language for object databases. In *Proceedings of the Conference on Advanced Visual Interfaces (AVI)* (1998), pp. 247–257.

[7] Papka, R., and Allan, J. Topic Detection and Tracking: Event clustering as a basis for first story detection. In *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*, W. B. Croft, Ed. Kluwer, 2000, pp. 97–126.

[8] Plaisant, C., Milash, B., Rose, A., Widoff, S., and Shneiderman, B. LifeLines: Visualizing personal histories. In *Proceedings ACM SIGCHI: Human Factors in Computing Systems* (1996), pp. 221–227.

[9] Robertson, B. A developer's guide to the resources of the historical event markup and linking project, 2001-2002. http://www.heml.org/.

[10] Tsotras, V., Jensen, C., and Snodgrass, R. An extensible notation for spatiotemporal index queries. *ACM SIGMOD Record 27* (1998), 47–53.