Visualize History

Computer Science 490

Sam Strasser

September 28, 2007

1 Introduction

Visualize History aims to provide a new way of presenting and understanding history. The application will provide an intuitive, interactive way to navigate the complex relationships within history.

Historical events are often restricted to the pages of books, newspapers and magazines, and from a very young age, people learn to approach history exclusively through words. When images are used, they often supplement or summarize the prose; only rarely does visual data stand on its own.

Yet no intrinsic property of the historical data enforces the convention that words should trump images. In fact, much of history is naturally presented visually. Any interesting history will consist of various parts, interlinked and roughly ordered. Students must reconstruct the multi-dimensional story from the linear prose they read in order to understand the complete picture.

Visualize History will attack the general problem by presenting that multi-dimensional story in multiple dimensions. For example, many key events are directly linked to a specific geographic location and all are linked to a specific time or time period. The application will therefore provide a way to navigate through both time and space. However, logical relationships between events are just as important as the temporal or spatial ones, and the application will also provide a way to navigate from an event to other related events.

I expect to lay out the semester approximately as follows (for more details, see below):

- Start up: meeting with professors; shaping idea; getting sign off (September)
- Research (Oct. 1 – 14)
- Design (Oct. 15 – 31)
- Implementation (Nov. 1 – 30)
- Polish and finalize (Dec. 1 – 8)

2 Research

The problem of visualizing historical data is in no way new. David Staley predicts that computers will change the way history is taught (Staley, 2003). In his book, he gives philosophical arguments and predictions for the evolution of historians and historiography, as well as examples of new ways to present data. Edward Tufte, perhaps the most famous person to visualize data, gives example after example of historical data that was properly presented, and of some that was not (Tufte, 1997). He tells of the cholera outbreak in London in 1854 near the Broad Street pump, and the role of the diagram in evaluating the cause of the outbreak. The list of examples, even as given only by these two authors, goes on and on.

The problem has not, however, been addressed in general. Each of the example diagrams had its own special considerations that factored into the final product. Each case was to be considered separately, and often yielded vastly different diagrams. My research will focus on the characteristics that can be abstracted out of this long list of examples to characterize the more general problem, and those characteristics that are truly specific to the individual case. Specifically, the research will aim to define

history in the abstract case, and especially try to narrowly define the relationships between and among historical events. For example, to visualize wars in which the United States has participated, the Civil war might be represented as one blob that spans 1860-1864. To visualize the end of slavery, though, the civil war would need to be broken into smaller pieces, such as the emancipation proclamation or Southern secession. By examining many such examples, I hope to be able to abstract a loose but formal structure for historical data in the general case.

While looking at these specific solutions, one quickly sees the problem become unmanageable. That is, more events are related than one might want to show at any time. Another main goal of my research therefore will be to define the application's domain in such a way that the data remains manageable and comprehensible. That domain must allow at least for the most important types of history, and should allow for as much flexibility as possible.

3 Design

The design process will be to formalize that loose structure into a formal structure that can be efficiently implemented. The design will consist of three pieces: data storage, abstract historical objects and the user interface. I expect to produce a diagram documenting the flow of data within the application, both between abstract objects and also between the user interface and those objects. It will label all inputs and all outputs, and map those inputs can follow through the application. I will write the design in loosely object-oriented terms, demonstrating the objects and the relationships between them.

Part of that design will include large portions that will not be implemented due to the large scope of the project. Where possible, I will note what I intend to implement. The implemented pieces will assume that the designed (but not implemented) features will eventually be implemented.

4 Implementation

The implementation will depend heavily on the design that I come up with. I will have to make several high level implementation decisions, along with the more mundane ones.

I expect to use one of three general platforms. The Google Maps API provides a lot of core functionality, such as converting natural language locations to latitude and longitude pairs or doing the actual display of the graphical data. The second potential platform would be Microsoft Silverlight and the third would be a Java applet. In either case, Visualize History will be easily embedded in a web site. The decision between the three depends almost entirely on what tools they provide in this specific area, namely for geographic translation and display.

The language will be implied by the platform. For example, Google Maps require JavaScript, for which I know syntax but am relatively inexperienced. Silverlight would allow for any .NET language, and I am very familiar with C#. A Java applet would imply Java, which I am also fairly familiar with.

Because so much of the design will not be implemented this semester, the implementation will allow for the addition of the documented features, even if they will not yet be added.

5 Tentative Design

While doing some preliminary research, I began to guess about the details of the design. These are only my current thoughts, and some are more concrete than others. For now, it is the design plan.

5.1 Representation of Historical Data

Historians usually present history as part of a topic; you might learn about Israeli independence or about the American Civil War, and then each event implicitly belongs to that topic. Many events usually make up these topics, and, as taught today, history suggests that each topic is separate. The topics are not at all separate, which anyone realizes as they study more and more history: everything seems to relate. The representation that I predict I will use is best explained with an example.

A class on the Civil War rarely deviates toward a discourse on the American Presidency, other than outside the time period of about 1856 to 1868, and if it did, it would probably be considered a bad course. A student exploring the Civil War might like to understand the evolution of the party system then to comprehend fully what a split in the Democratic Party might suggest. That student should then be able to pause her exploration of the Civil War to look at past Presidents. So in this case, Abraham Lincoln would belong to the topic "American Civil War" and would also belong to the topic "American Presidents". I will call these topics "datasets" although that is fairly likely to change. I will also interchange "nodes" and "events" depending on the context.

Another level of complication arises in that datasets can also be events. In looking at cathedrals of Western Europe, you might expect to find Notre Dame as a single event, as it should be. If you wanted to explore Paris from 1150 to 1350, though, you would almost certainly want to see big events in the buildings history, such as the beginning of construction (1163), the completion of the Nave (1196) or the final touches (1345). Notre Dame the dataset and Notre Dame the node both need to be represented.

Not only should datasets be able to be nodes as well, but the two should be able to be related. To continue the above example, while looking at architecture of Western Europe, our user might decide that she wants to learn all the details of Notre Dame. By focusing on Notre Dame (the node), she should be able to interact with the details of Notre Dame (the dataset). In fact, for the user, there should probably be no distinction between datasets and nodes.

To summarize, I will treat history as graph whose nodes are events, people, datasets and possibly more. The edges will be relationships among those nodes. Specifically, some set of edges (perhaps implied) will denote that two nodes are related by time, and another set will denote that two nodes are related in space. The final type of edge is what I have been calling "logical relationships". My research will be focused on the attributes that should be allowed. These attributes will make up the rest of the edges. For example, the node "Abraham Lincoln" will be connected to the node "Civil War" and "American Presidents", and both of those edges will be "logical" edges.

5.2 User Interface

The user interface, the key component of the application, will be broken down into two parts. The display area will take up about 80% of the width of the application, and the navigation panel will take up the rest. The display area will be responsible for showing the content of the application. It will use several different views that will correspond to different ways to visualize the data.

One view will be based on top of a map. (This is the only view that I intend to implement as part of this project). It will correlate data that is related geographically. For example, if focusing on architecture from 1000 to 1500, it might focus on France and England to demonstrate the similarities and differences as the two developed and exchanged their techniques for designing and constructing architecture.

The map view will include a slider to represent time, with some notion of "current time". Then, events will be less and less opaque as they are farther from "current time". Even more specifically, I can imagine a general frame of time that the user is viewing. Events from within that time period would be shown and others would not. Then the frame (and not just "current time") could be dragged, stretched

or shrunk as the user wanted to scroll through time. By extending the frame, the user would be broadening the scope of their focus, and by shifting current time, would be changing the specific focus (and not its breadth). Along with this ability to zoom in and out of time, the user would also be able to pan and zoom along the geographic axes (latitude and longitude). Note here the benefit to the user's experience from extending an already familiar concept such as Google Maps or map interfaces more generally. Being able to pan and zoom is so natural that by now it is expected in such an application. If the time axis is implemented with a similar UI, the user is already equipped to use the application.

The nodes will then be overlaid on the map. If a node is naturally related with a place then it will be placed at that location. Abraham Lincoln of 1858 might be shown in Illinois (during the Lincoln-Douglas debates) but Abraham Lincoln of 1861 might be shown in Washington D.C. Clicking on a node will increase zoom in on that node. For example, clicking on Notre Dame might then show the construction and completion of key pieces of the cathedral. Connections will be shown through cartoon lines or arrows. Napoleon's march of 1812 might for example be shown as an arrow to Moscow, and another arrow for the return trip.

5.3 Generation of Historical Data

For Visualize History, the data will be uploaded by a human. There will be a small initial dataset that I upload. That will aim to demonstrate the breadth of the application. On top of that dataset, the application will include an easy way to upload additional information. That upload mechanism will also be aimed at a user familiar with technology only from the end-user perspective.

As a side note, I plan to devote next semester to this problem. A lot of data exists in various forms, and after this semester's proof-of-concept, I hope to make the application more practical by accumulating data from the vast, public, free locations where it lives right now.

6 Deliverables

A web-based application that allows for manual data entry of generic historical data, and displays that data in an interactive way that is intuitive to navigate and makes exploring history easy.