Topotime: Representing historical temporality

Karl Grossner and Elijah Meeks, *Stanford University Libraries*

Historical analysis within any discipline depends in part upon establishing chronologies (Sewell 2005), but historical data are problematic. The spreadsheets and database systems used for representing and computing over digital chronologies do not handle vague or otherwise uncertain data well. How does one encode "for 6 months before the war," "around 1832," or "during harvest seasons in her youth?" And when dates for events lasting days or months are given only in years, how can we calculate contemporaneity? What if our data include both precise dates and vague date ranges with varying granularity?

To date, historical researchers and digital humanities application developers have managed temporal uncertainty in ad hoc fashion, normally with one or two date fields using the ISO-8061 standard (e.g. YYYY-MM-DD) for the Gregorian calendar— most would say with less than optimal results. Meanwhile, researchers in computer science, geographical information science, and other fields have done considerable work on some challenges in temporal representation, including uncertainty and qualitative temporal reasoning. (cf. Kauppinen et al 2010; Holmen & Ore 2009; Crescioli, D'Andrea Niccolucci 2000; Plewe 2002). Some of those results, which often demonstrated only in small exemplars, can be brought to bear on humanists' requirements. However, we must make explicit our desiderata for temporal representation and computation in order to make headway towards fulfilling them.

In light of this, we have initiated the *Topotime* project[1], with these initial goals: 1) a specification for computable digital representations of the kinds of temporal entities typically found in historical texts, and some relationships between them; 2) one or more graphical timeline layout programs to parse and render data written in that form; 3) software tools to facilitate the encoding process, and to transform data in two stages—converting spreadsheet exports to the flexible and human-readable data format, JSON-LD[2], then parsing and transforming those JSON data files into "temporal geometries" for calculations of distance, similarity, and topological relations. In this paper, we briefly outline the draft Topotime specification as it stands, and the software development progress we have made so far.

The goals as set out above are admittedly ambitious. Temporal entities found in historical texts and records include a wide range of scales and imprecision, and refer to many calendars and modes of temporal reasoning. We have begun what will be a long-term iterative process of enumerating examples and fine-tuning a data model to handle them, written in JSON.

---

[1] http://dh.stanford.edu/topotime; https://github.com/ComputingPlace/Topotime

[2] JavaScript Object Notation for Linked Data (http://json-ld.org/).

Two perspectives

We are approaching this work from two directions in parallel to meet requirements for both drawing timelines and for calculating temporal relations. These are not mutually exclusive and parsers for both cases have considerable overlap of functionality and comparable complexity; generalities in formal representation that are useful in both cases are emerging. For example, both parsers convert various date expressions to Julian dates for calculations. There are also distinct differences, for example between the data objects best suited for efficiently drawing time bands, dots, and arrows on a timeline, and the temporal geometries referred to earlier.

The basic elements of Topotime

A Topotime data file describes a *PeriodCollection*. Each *Period* is of a *class* (either *Event*, *HistoricalPeriod*, or *Lifespan*) and has temporal extents described by one or more typed timespans (*tSpan*). PeriodCollections have Projection definitions which include *atom* (granularity, such as day or year), *origin* (day zero on the reference calendar, in Gregorian date terms), and *scale* (used for timeline rendering). Periods must have a unique id, a source attribution, and a label for graphical display. They can also have any number of optional properties (attributes), although Topotime software does not handle these directly. A PeriodCollection can also include a set of asserted *relations*, both between periods and between periods and places. These are distinguished from those purely temporal relations between Period timespans, which can be calculated and may be incidental.
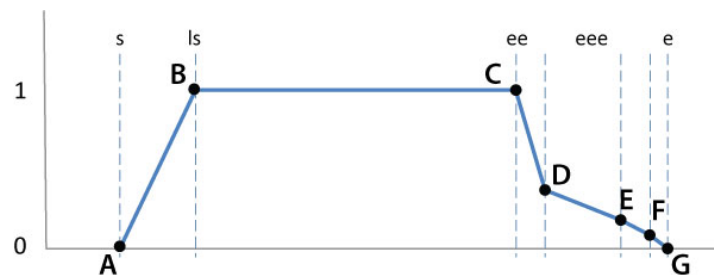
Timespans

When describing the "when" of an occurrence we ordinarily mean that it took place either *throughout* some timespan, or for *some time during* it. Someone born in 1723 was not born for the entire year! In Topotime, a *tSpan* describes temporal extents and *throughout* is the default; *some time during* is noted by adding a ("during": True) statement, and a duration, e.g. ("d": "1d") for a duration of a birth day. In both cases (throughout and during), date ranges describe bounds with a required start ("s") and optional latest-start ("ls"), earliest-end ("ee"), and end ("e"). This conforms to a pattern commonly seen in graphical representations, from Joseph Priestley's 18th century timelines to the popular Simile Timeline software[3], and the recent

---

[3] http://www.simile-widgets.org/timeline/

formalizations cited earlier. In this way, timespans can be represented as having either fixed or "fuzzy" bounds[4]. The result is a "temporal geometry" such as pictured in Figure 1. The shapes of the curves between s-ls and ee-e can be articulated more completely by adding "sls" and/or "eee" arrays, as shown, reflecting an author's understanding of the probabilities over the course of those sub-spans.

Time values for s, ls, ee, and e can be either a day, a month, or a year, and can be qualified by operators for "before" (<), "after" (>), and "about" (~). They can also be pointers to other Period timespans or parts thereof. For example, ">38.s" refers to "after the start of Period 38 in this collection." Omission of a referent part (e.g. >38 or <38) is taken to mean either its end ("e") in the first case or its "s" in the latter.

Topotime recognizes not only *date ranges* with fixed or fuzzy bounds, but *durations*, *cyclical timespans* (regularly recurring ranges), and *multi-spans* (arbitrary discontinuous spans) as well. Examples of notation for these are listed in Table 1.

Table 1 - Timespan notation in Topotime, partial listing

fixed range (throughout)	{"s": "1901-04-01", "e": "1963-01-12"} *A lifespan: born April 1, 1901; died Jan 12, 1963*
fuzzy range (throughout)	{"s": "1923-03-21", "ls": "1923-06-20", "e": "1930-10-01", "ee": "1930-12-31"} *Employed from spring of 1923 to late 1930.*
fixed range (during)	{"s":"1934", "during": True, "d": "4m"} *Traveled in Spain for 4 months in 1934*
fuzzy range (during)	{"s": "1923-03-21", "ls": "1923-06-20", "e": "1930-10-01", "ee": "1930-12-31", "during": True, "d": "~6m"} *Hospital stay for about 6 months during studies*
cyclical	{"s": "1951-05-01", "e": "1999-05-01", "cduration": "18m", "cstep": "4y"} *US Presidential campaign seasons in late 20th century*
multi-part	[{"s": "1901-01-01", "ls": "1901-02-02", "ee": "1919-01-01", "e": "1920-05-05"}, {"s": "1931-01-01", "ls": "1935-02-02", "ee": "1961-01-01", "e": ">12"}] *intermittently as specified, until after Period #12*
duration	{"s": ">1", "duration": "2m"} *tSpan for Period beginning after Period 1, lasting 2 months*

[4] Note that the term "fuzzy" means indeterminate and probabilistic here; this does not correspond with its meaning in fuzzy set theory, as percent membership in a set.

Period relations

Meronomic (parts)
Purely temporal relationships between Periods (overlap, adjacency, containment) can be calculated from tSpans geometrically. But there are more relationships we routinely assert and represent, for example *part-of*. We might say, "these 18 events occurring at these times, or in this order, were *part-of* that larger event"—e.g. a lifespan, war, or political campaign. Another scholar's chronology for the same composite event might include an entirely different set of sub-events. We may wish to model The Bronze Age as an historical period having spatial-temporal parts such as "Late Bronze Age Southern Levant" and "Bronze Age – Malta." A Topotime *relation* consists of a subject, predicate and object (at minimum) in the following form:

> { "subj": 23, "pred": "has-part", "obj": 14}

Among other things, Topotime part-of relations enable rendering sub-events within parent containers on timelines.

Time and place
Events and other occurrences are wholly bound to places. Parenthetically, we would argue that places may be best characterized by what has occurred in them. Certainly historical periods are often defined geographically, or are relevant only in particular regions. Some are equally geographic and temporal constructs, e.g. "Pre-dynastic Egypt" (4500-2950 BC), or "The Neolithic Levant." "Song Dynasty in the Third Imperial Period" is relevant in China and neighboring places, but not elsewhere.

Furthermore, simple and complex events all have spatial extension we often want to display on a map alongside a timeline. Period locations in Topotime can be specified with a single spatial location expressed in a standardized format (GeoJSON or WKT), and with an optional name in this form:

> { "subj": 23, "pred": "has-location", "obj": {"name": "Venice",
>   "geom": "POINT (45.4375, 12.3358)", "geomType": "WKT" }

Standards for specifying Places in data objects like these for gazetteers are now emerging, thanks to the coordinating efforts of projects like Pelagios and national historical GIS projects Great Britain Historical GIS and the China Historical GIS.

Participation
Periods having "class": "Lifespan" can be asserted to participate-in other kinds of periods, such as *Event* and *HistoricalPeriod*.

Looking ahead

Topotime is an open source software development project. We know that many further challenges exist for representing not simply time, but temporality in digital humanities works. Our goal has been to help initiate what will hopefully be an ongoing collaborative process with some concrete steps and functioning software. We are hopeful this work will contribute to the development of interoperable gazetteers of place and period, temporal extensions of the popular GeoJSON format, and improved capabilities for timeline visualizations.

Works cited

Crescioli M., D'Andrea A., Niccolucci F. (2000). A GIS-based analysis of the Etruscan cemetery of Pontecagnano using fuzzy logic, in G.R. Lock (ed.), *Beyond the Map: Archaeology and Spatial Technologies*, Amsterdam, lOS Press, 157-179.

Holmen, J., and Ore, C. (2009). Deducing event chronology in a cultural heritage documentation system. In *CAA 2009 Proceedings* retrieved 29 Jul 2013 from http://www.edd.uio.no/artiklar/arkeologi/holmen_ore_caa2009.pdf

Kauppinen,T. Mantegari,G., Paakkarinen, P., Kuittinen, H., Hyvonen, E., Bandini, S. (2010). Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International Journal of Human-Computer Studies 68* (2010) 549–560

Plewe, B. (2002). The Nature of Uncertainty in Historical Geographic Information. *Transactions in GIS*, 6(4): 431-456.

Sewell, W. (2005). *Logics of History*. Chicago: University of Chicago Press