Visualizing Temporal-Semantic Relations in Dynamic Information Landscapes

Vedran Sabol
vsabol@know-center.at
Know-Center Graz
Graz, Austria

Arno Scharl
scharl@modul.ac.at
MODUL University Vienna
Vienna, Austria

Abstract

Topical relations, trends and geographic context are three dimensions of heterogeneous data sets that play an important role in many semantic applications. Several visual representations have been developed which separately address each of these three dimensions. However, in application domains with data sets characterized by patterns not only within each separate dimension, but also by patterns between the dimensions, these patterns may resist analysis unless all dimensions can be analyzed simultaneously. This paper aims to provide a framework for the simultaneous analysis and visualization of multiple dimensions by means of tightly coupled views [20]. The resulting user interface allows investigating the evolution of knowledge by simultaneously showing topical, temporal and geospatial patterns in heterogeneous data sets. The synchronization of the tightly coupled views will build upon *Multiple Coordinated Views Technology* [17] and on lessons learned from the FIT-IT Semantic Systems Project IDIOM [26], which has introduced the interface metaphor of *Knowledge Planets* [23]. The *Media Watch on Climate Change* [27] implements this metaphor and demonstrates the potential of geospatial Web technology for visualizing semantic associations within virtual spaces.

This prototype is just a first step towards building a portfolio of modular visualization services that use geobrowsers such as NASA World Wind and Google Earth as *generic image rendering engines* for the following types of data: (i) geo-referenced information (news articles, statistical data, user profiles, etc.), (ii) semantic associations, (iii) ontological knowledge, (iv) organizational workflows, and (v) social networks.

Knowledge Planets

Diverting geobrowsers from their traditional purpose and associating them with *semantically* referenced information, they can be used to visualize *knowledge planets* based on layered thematic maps (see Figure 2). Such maps are visual representations of virtual spaces based on a landscape metaphor [4]. Much has been written about the geography of virtual spaces [3; 8; 9; 24]. In contrast to knowledge planets, however, many of these spaces bear little resemblance to the Earth's physical features.

Generally, two sets of information need to be integrated and mapped to latitude and longitude – image tiles and terrain information. Knowledge planets are generated by orthographically projecting and tiling thematic maps. The planet metaphor allows visualizing massive amounts of textual data. At the time of map generation, the knowledge planet's topography is determined by the content of the knowledge base. In this process, spatial proximity in the layout is a measure for the relatedness of documents. The peaks of the virtual landscape thus indicate abundant coverage on a particular topic (their height being related to the number of documents within the cluster), whereas valleys represent sparsely populated parts of the information space.

Extending the planet metaphor, search results can be visualized as cities, landmarks or other objects of the manmade environment. Zooming provides an intuitive way of selecting the desired level of aggregation. Unique resource identifiers link concepts embedded in the thematic maps to related news articles, encyclopedia entries or papers in scientific journals. With such a query interface that hides the underlying complexity, exploring complex data along multiple dimensions is as intuitive as using a geobrowser to get a glimpse of the next holiday destination.

The knowledge planet prototype of Figure 2 is based on *VisIslands*, a thematic mapping algorithm similar to SPIRE's Themescape [25] and its commercial successor Cartia/Aureka, which supports dynamic document clustering [1; 21]. Initially, the document set is pre-clustered using hierarchical agglomerative clustering [12], randomly distributing cluster centroids in the viewing rectangle. The documents belonging of a cluster, as determined by the pre-clustering, are then placed in circles around each centroid. A linear iteration force-directed placement algorithm adapted from Chalmers [5] optimizes the arrangement. The result resembles a contour map of islands. Fortunately, algorithms based on force models easily generalize to the knowledge planets' spherical geometries.

Visualizing Temporal Data

Various approaches have been developed for visualizing temporal data [18]. In the information visualization literature, well-known metaphors for presenting data with a pronounced temporal component are ThemeRiver [10] or the Perspective Wall [16]. In the context of this paper, temporal activity and intensity views are of particular interest (see Figure 1). *Temporal activity views* visualize presence and duration of events, which are related to a set of color-coded entities, such as different persons or topical clusters present in a document repository. Events are visualized as rectangles with the duration of events being proportional to the width of the corresponding rectangles. Activity views and similar visualizations provide a clear representation of event boundaries, but are not suitable for overlapping events. *Temporal intensity views* visualize the intensity of those events, whereby overlapping events are stacked over each other to produce "hills". The height of these "hills" indicates the cumulative intensity of events at the given point in time. Intensity view and similar visualizations are suitable for representing overlapping events, but cannot provide a clear representation of event boundaries.

Despite the visual appeal of ThemeRiver visualizations, there are conceptual limits with regards to how many dimensions developers can integrate into a single display. Inexperienced, non-technical users in particular can get overwhelmed when too many dimensions are presented within the same diagram. Coordinated multiple views provide effective navigation mechanisms for complex data sets, whereby each visualization reveals a different facet of the data (such as temporal, topical or spatial dimensions). Truly scalable visual applications – in terms of always having the option to add additional dimensions whenever the need arises –require a reliable high-performance architecture based on a unified data model for synchronizing tightly coupled views in response to queries and user actions such as zooming, selection and filtering.

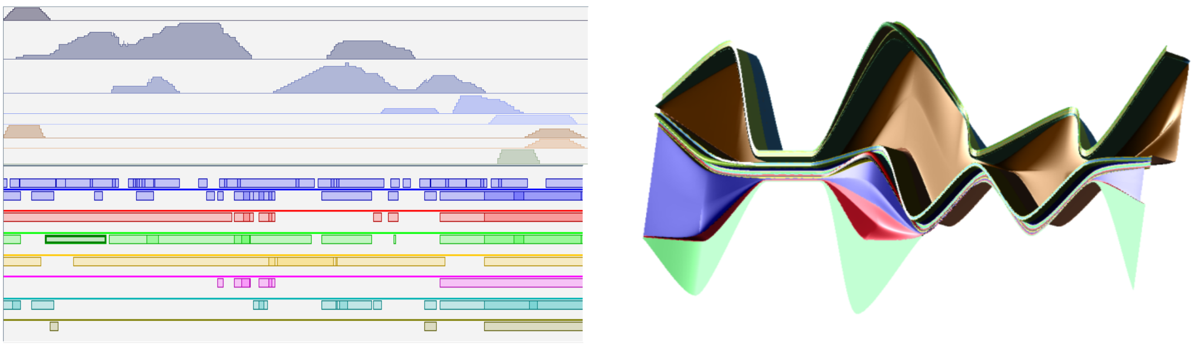

Figure 1. Temporal activity and intensity diagram [20]; three-dimensional ThemeRiver [11]

Snap Together [19] is an early example in which users dynamically combine and bind different visualizations on-the-fly to produce a customized user interface. Oculus GeoTime [14] is an example of a representation that merges geospatial and temporal developments within a single 3D view. Brodbeck and Girardin [2] employ multiple coordinated views for analysis of geo-referenced, high-dimensional data sets. The IDIOM Project [22] also falls into the group of geospatial-topical systems as it employs a geospatial view and a semantic landscape for browsing large document sets within a single, coordinated interface.

Dynamic Topographies for Semantic Maps

While temporal activity and temporal intensity views are suitable for discovering temporal patterns and temporal behavior of entities, they generally cannot express complex, manifold relations and patterns in heterogeneous datasets. Semantic maps, by contrast, are capable of presenting complex relations in the data set due to the fact that spatial proximity in the layout is a measure for relatedness between objects. But temporal behavior cannot be visualized in a static landscape. The next generation of semantic map algorithms proposed in this paper will show how virtual globe architectures are ideally suited to visualize the evolution of information spaces. These algorithms will specifically address *temporal changes in the underlying topography*.

Visually resembling tectonic processes in the natural world, dynamic rendering will reflect both long-term trends and short-term fluctuations in the knowledge space (in geology, these processes are referred to as epeirogeny and orogeny, respectively). A major international event that causes intensive media coverage can be visualized as a volcano-like structure rapidly rising from the seabed, for example, forming a new island that might remain as a permanent addition to the knowledge space (compare Figure 2; the photo serves to illustrate the underlying idea; to visualize the rise/decay of a topic, the actual mapping algorithm will elevate/lower a set of concentric contour lines).

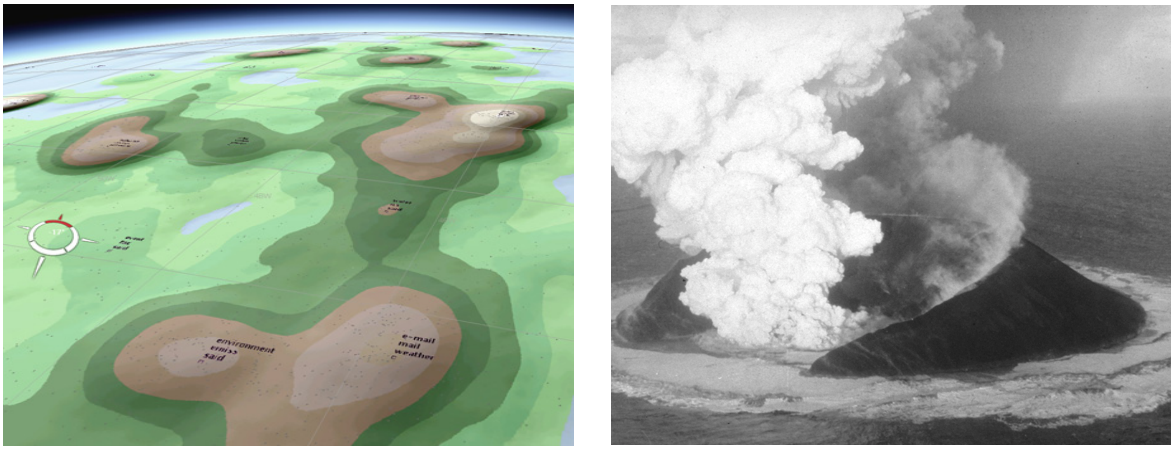

Figure 2. Knowledge planet prototype [23]; volcanic cone of an island off Iceland's coast [28]

NASA World Wind's *Time Control Widget*, a new feature of release 1.4, provides an effective way to control such animations and explore the evolution of information spaces. Even more powerful as a control element are temporal activity and temporal intensity views. To understand how the dynamic topography of the knowledge planet is created, it is best to imagine the surface of the planet being placed so that the temporal axis of the temporal view is orthogonal to the landscape (Figure 3; right); the surface becomes a slice of the temporal view for a chosen interval in time. The landscape view thus shows relations between entities, but only depending on the data from the selected time interval. As the user slides the selected time interval along the time axis (panning along the time axis), or modifies the width of the interval (temporal zooming), the landscape view updates itself dynamically to reflect the changes in the selected subset of the data.

Note that when the visualized data subset is modified (some documents are removed, new documents are added) the semantic map is not just filtered – it is its topography that is altered. Old island and hills may disappear, change their shape or even new ones may arise from the sea. Other modifications of the topography, such as drifting of hills towards each other (correspond to merging of previously separate clusters) or splitting of an island (cluster breakup) may also occur. Transitions of the landscape topography from an old to a new temporal configuration must be incremental and adaptive in the sense that only those changes should be introduced in the topography which are really necessary. The configuration of the parts of the topography, which are little or not at all affected by the modification of the selected time interval, should remain as stable as possible with respect to their previous position and shape. In this way the user will be able to understand the modified topography immediately through the recognition and orientation provided by the already known, preserved (or scarcely modified) elements of the topography. These adaptive, incremental transitions shall be smoothly animated so that the user can follow and understand the changes.

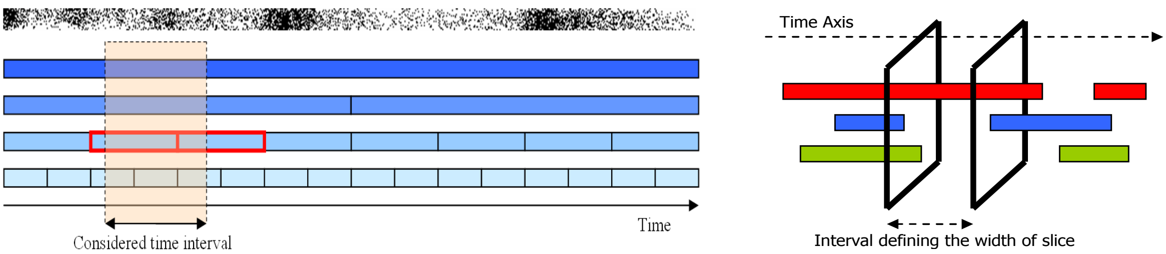

Figure 3. Document distribution over time on top and temporal subdivision into segments (left); knowledge planet surface as a temporal slice across the time axis (right).

The above described behavior requires fast, incremental, scalable layout algorithms: when a data set is modified a full re-computation of the whole data set shall not be performed – the changes are merely incorporated into the existing configuration which saves processing time and ensures that unmodified structures can be recognized by the user. A class of layout algorithms based on the force-directed placement approach [7] fulfils the requirements of incrementality and scales very well in their optimized versions. *O(n*log n)* time complexity is possible, for example with the algorithm [13] used to generate the prototype in Figure 2.

Dynamic landscape topographies have only been realized for very small data sets of up to few hundred objects [15]. To address the increased computational and storage demands resulting from dynamic topographies and very large data sets, a *'data squashing'* approach [6] will be applied. Data squashing can be defined as the construction of a summarized, "compressed" dataset which is significantly smaller than the original one, but leads to approximately the same analysis results as the original. Construction of the summarized, "squashed" data set will be realized by employing automatically generated ontologies (see previous section) to guide the subdivision of the original information space. The advantages of this approach are a potentially large decrease of both memory and CPU-time consumption at the possible cost of some reduction in accuracy: Instead of calculating a similarity matrix between one million documents, for example, processing ten sets of 100,000 documents each is computationally less expensive. The subdivision will also help improve the automated generation of captions for topological features; an essential component for visualizing dynamic topographies through interfaces that support multiple layers of abstraction.

To address requirements arising from the adaptive dynamic topographies and the capability to choose the time interval which shall be visualized, the data set will be subdivided into temporal segments. This is represented in Figure 3 (left) with the document distribution over time shown on top (a dot represents an event or a document at the given time point) and a hierarchical temporal subdivision of the data set (shown in blue). Temporal segments are produced starting with the top-level segment (dark blue, contains the whole data set), which is recursively subdivided to produce a hierarchy of ever shorter segments. A separate layout for each segment is pre-computed and stored, so when the user changes the selected time interval (shown as a transparent box) suitable pre-computed layouts are chosen (segments with red borders in the figure bellow), filtered and used as a starting position for the on-the-fly computation.

Figure 4 shows a mock-up of the RAVEN interface, composed out of nine different views. The interface employs three monitors with the dynamic landscape shown on the left, the geospatial browser on the right, and a central component offering several views: document view (top-left), search results view (top-right), temporal activity and intensity views (bottom-left), and an ontology browser (bottom-right).

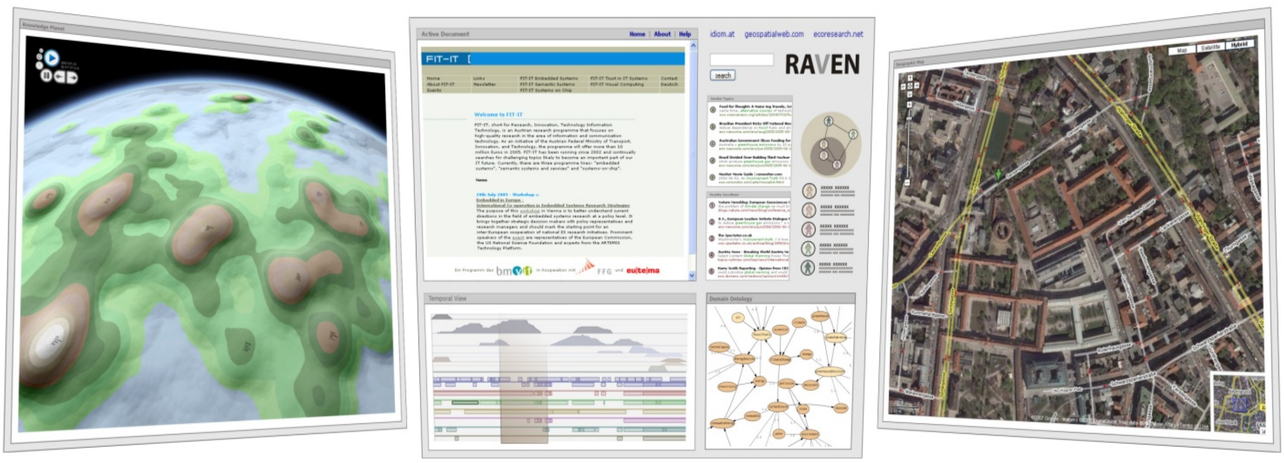

Figure 4. Interface mockup with temporal controls (left: *Time Control Widget;* middle: activity and intensity diagrams) to control an ensemble of nine tightly coupled views on three different screens.

References

1. Andrews, K., Guetl, C., et al. (2001). "Search Result Visualisation with xFIND", *Second International Workshop on User Interfaces to Data Intensive Systems (UIDIS-2001)*. Zurich, Switzerland: IEEE Press. 50-58.
2. Brodbeck, D. and Girardin, L. (2003). "Visualization of Large-Scale Customer Satisfaction Surveys Using a Parallel Coordinate Tree", *9th IEEE Symposium on Information Visualization (InfoVis 2003)*. Seattle, USA: IEEE Computer Society.
3. Castells, M. (2001). *The Internet-Galaxy: Reflections on the Internet, Business, and Society*. Oxford: Oxford University Press.
4. Chalmers, M. (1993). "Using a Landscape Metaphor to Represent a Corpus of Documents", *Spatial Information Theory: A Theoretical Basis for GIS (Lecture Notes in Computer Science, Vol 716)*. Eds. A.U. Frank and I. Campari. Berlin: Springer. 377-390.
5. Chalmers, M. (1996). "A Linear Iteration Time Layout Algorithm for Visualising High-Dimensional Data", *7th Conference on Visualization*. San Francisco, USA: IEEE Computer Society. 127-132.
6. DuMouchel, W. (2002). "Data Squashing: Constructing Summary Data Sets", *Handbook of Massive Data Sets*Norwell, USA: Kluwer Academic Publishers. 579-591.
7. Fruchterman, T.M.J. and Reingold, E.M. (1991). “Graph Drawing by Force-directed Placement”, *Software - Practice and Experience,* 21(11): 1129-1164.
8. Gibson, W. (1984). *Neuromancer*. New York: Ace Books.
9. Giese, M. (1998). “Constructing a Virtual Geography: Narratives of Space in a Text-Based Environment”, *Journal of Communication Enquiry,* 22(2): 152-176.
10. Havre, S., Hetzler, E., Whitney, P. and Nowell, L. (2002). “ThemeRiver: Visualizing Thematic Changes in Large Document Collections”, *IEEE Transactions on Visualization and Computer Graphics,* 8(1): 9-20.
11. Imrich, P., Mueller, K., et al. (2003). "Interactive Poster: 3D ThemeRiver", *IEEE Information Visualization Symposium*. Seattle, USA.
12. Jain, A., K., Murty, M.N. and Flynn, P.J. (1999). “Data Clustering: A Review”, *ACM Computing Surveys,* 31(3): 264-323.
13. Jourdan, F. and Melancon, G. (2004). "Multiscale Hybrid MDS", *8th IEEE International Conference on Information Visualisation*. London, UK. 388-393.
14. Kapler, T. and Wright, W. (2005). “GeoTime Information Visualization”, *Information Visualization,* 4(2): 136-146.
15. M., G., W., K., V., S. and G., D. (2003). "WebRat: Supporting Agile Knowledge Retrieval through Dynamic, Incremental Clustering and Automatic Labeling of Web Search Result Sets", *1st IEEE Workshop on Knowledge Management for Distributed, Agile Processes*. Linz, Austria.
16. Mackinlay, J.D., Robertson, G.G. and Card, S.K. (1991). "The Perspective Wall: Detail and Context Smoothly Integrated", *SIGCHI Conference on Human Factors in Computing Systems*. New Orleans, USA. 173-176.
17. Müller, F. (2005). *Granularity Based Multiple Coordinated Views to Improve the Information Seeking Process*, Konstanz, Germany: University of Konstanz.
18. Müller, W. and Schumann, H. (2003). "Visualization Methods for Time-dependent Data - An Overview", *Winter Simulation Conference (WSC-2003)*. New Orleans, USA: IEEE Press. 737- 745
19. North, C. and Shneiderman, B. (1999). Snap-Together Visualization: Coordinating Multiple Views to Explore Information. College Park: University of Maryland, Department of Computer Science. http://www.cs.umd.edu/hcil/pubs/.
20. Sabol, V., Granitzer, M. and Kienreich, W. (2007). "Fused Exploration of Temporal Developments and Topical Relationships in Heterogeneous Data Sets", *3rd International Symposium of Knowledge and Argument Visualization, 11th International Conference Information Visualisation*. London, UK: IEEE Computer Society.
21. Sabol, V., Kienreich, W., et al. (2002). "Applications of a Lightweight, Web-Based Retrieval, Clustering, and Visualisation Framework", *4th International Conference on Practical Aspects of Knowledge Management (= Lecture Notes in Computer Science, Vol 2569)*. Eds. D. Karagiannis and U. Reimer. Berlin: Springer. 359-368.

22. Scharl, A. (2006). “Tightly Coupled Geospatial Interfaces for Collaborative Systems and Just-in-Time Information Retrieval Agents”, *Research in Computing Science (Special Issue: Semantics in Geoinformatics Applications),* 25: 3-18.
23. Scharl, A. (2007). "Towards the Geospatial Web: Media Platforms for Managing Geotagged Knowledge Repositories", *The Geospatial Web - How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. Eds. A. Scharl and K. Tochtermann. London: Springer. 3-14.
24. Schroeder, R., Huxor, A. and Smith, A. (2001). “Activeworlds: Geography and Social Interaction in Virtual Reality”, *Futures,* 33(7): 569-587.
25. Wise, J.A. (1999). “The Ecological Approach to Text Visualization”, *Journal of the American Society for Science,* 50(9): 814-835.

Online Resources

26. IDIOM Research Project. http://www.idiom.at/.
27. Media Watch on Climate Change. http://www.ecoresearch.net/climate.
28. National Oceanic & Atmospheric Administration (NOAA). http://www.noaa.gov/.