Curran Kelleher
Proposed Research

I propose to design, implement and deploy a collaborative Internet-based multi-scale visual analytics platform. The system will have the multiple coordinated view functionality of Weave, but will also be augmented by two main differences: 1.) it is designed to have as its data representation foundation the Semantic Web and Universal Data Cube 2.) it will address the visual representation and manipulation of ontologies and knowledge bases using domain specific visual encodings.

The proposed platform will be built on top of many existing technologies, filling in the usability gaps which prevent mainstream usage of powerful data management and visual analytics tools. The three primary technologies which will be tightly integrated to fulfill this goal are Weave, the Semantic Web, and the Universal Data Cube. The resulting integrated platform will ultimately provide an interactive visual environment for dealing with the following three realms of interrelated tasks: ontology (conceptual model) design, management and presentation; knowledge base management and presentation; and analytical exploration of multi-scale, multidimensional knowledge base summaries. In addition, the results at any stage can be published in Web documents for public consumption.

The amount of publicly available data published in the Semantic Web is astounding and growing every day. Currently, Weave can only visualize data which has been published using the Weave server. The first step of my plan of research is to adapt Weave to navigate and visualize Semantic Web content. This step alone would enhance Weave tremendously, as it would instantly give Weave users (both data providers and end users) the ability to put their data in the context of all data available in the Semantic Web and perform comparative analyses. Also, when browsing the Semantic Web in Weave, all published metadata about entity classes and properties will be made readily accessible from within the tool, enhancing usability and ensuring correct interpretation of the data.

Generally speaking, when performing analysis tasks on large amounts of data, it is tremendously useful to perform aggregation. In business intelligence communities, a commonly accepted method of aggregation is the (hierarchical) data cube, which allows aggregation of relational databases in a multi-scale, multidimensional fashion. In my work on the Universal Data Cube (UDC) [1], I developed a technology allowing for publication of federated hierarchical data cubes built on the Semantic Web. The second step of my plan of research is to modify the UDC model such that arbitrary Semantic Web content can be automatically aggregated into hierarchical data cubes and published for consumption.

Weave currently supports visual representation of data tables. The third part of my research plan is to develop a general theoretical model of how semantic graphs can be transformed into interactive visualizations. This visualization model will then be implemented in Weave. In this activity, I plan to build upon known works such as Lee Wilkinson's "Grammar of Graphics", and extend the visual mappings with interaction specifications such as selection, probing, layout refinement and data manipulation. One specific case of this visualization model I plan to address is the visual presentation and interactive navigation of hierarchical data cubes from the UDC. Another important specific case I plan to address is that of ontology design, which can be framed as a special case of knowledge base management.

In addition to domain-independent visualizations, I plan to develop the interactive visual encoding model to support domain-specific diagrammatic languages as well. The idea is that for any ontology, a set of interactive visual encodings can be developed for each class and property in that ontology. This feature would enable end users to create and disseminate their own ad-hoc visual languages. The combination of this feature with the powerful asynchronous collaboration feature of Weave will allow data-driven communities to continually generate novel visualization content for the Web in an organic way.

The overall purpose of this system is to extended the cognitive capabilities of individuals and organizations using a unique synthesis of interactive visual analytics, data management and Internet technologies. The proposed research plan would in the end enable user activity to span three distinct task categories: ontology management, knowledge base management, and knowledge base summarization. I believe that the real power in the system will be in the dynamic interplay between these three realms, and the worldwide collaboration on visualization content that follows.

The power of the envisioned platform is not in that it can do certain predefined things, but in that it will allow users to define their own visual encodings in novel ways that suit their tasks, then publish their results and methods on the Internet for others to use. The net effect will be the emergence of an enabling technology for production, consumption, and sharing of visualization content. The asynchronous collaboration feature of Weave will enable the evolution of global visual analysis content in much the same way as open source software - once a work is published, others can derive new works from it and publish that, and so on. The synchronous collaboration feature of Weave will enable people to closely collaborate on visual analysis tasks in real time. The combination of the two modes would be useful in scenarios such as the classroom, in which an instructor provides a “live feed” of his actions to students, but the students are free to break from the instructor session and independently explore, then merge the new findings back into the instructor’s session when, for example, asking a question.

I hope to disseminate the resulting software and specifications as openly as possible within the constraints of the University, and publish academic research papers on the work as it progresses. The metamodel mapping between semantic graphs and hierarchical data cubes may be a candidate for consideration as an open standard. I hope to be able to pursue this direction eventually with a standards body such as the Object Management Group or World Wide Web Consortium.

The resulting set of tools will allow the average computer user to visually navigate, explore and create public data, as well as previously published analyses of that data. The primary desired outcome is that this platform will be used by ordinary people in daily life to enhance their education and quality of life. Since the domains of application are so diverse, these tools could also be fully integrated into teaching of courses at the middle school, high school and university levels. The tool is also intended to aid researchers in many domains in their data analysis, collaboration and knowledge dissemination processes.

[1] Additional material for consideration is available at www.curransoft.com/grfp .