Visual Analyses on Linked Data – An Opportunity for both Fields

Michael Granitzer, Vedran Sabol, Wolfgang Kienreich, Dickson Lukose, Kow Weng Onn

The Social Semantic Web (Breslin et. al. 2009) constitutes the next evolutionary step of the World Wide Web by combining semantic technologies with social computing paradigms. A core concept of the Social Semantic Web is Linked Data (LD), which allows for publishing and interlinking data in a machine understandable way on the World Wide Web (WWW). Linked Data uses established W3C techniques, such as Unique Resource Identifiers (URI) for uniquely identifying data sources, the Resource Description Framework (RDF) for semantic representation of those data sources and SPARQL as query language for data access. One recent result is the Linked Open Data (LOD) cloud, which aims in massively connecting available information in all domains, e.g., Friend of a Friend (FOAF), DBPedia, GeoData as well as Academic Data. The LOD cloud currently includes over 100 different datasets and has been growing from one billion triples and 250k links in mid-2007 to 25 billion triples and 395 million RDF links in September 2010. With the recent drive towards publication of both high quality metadata together with datasets supporting the publications, exponential growth is foreseen in infrastructure for data sharing, tools for documenting and moving data and semantic approaches to metadata management (Steinhart, 2011).

However, utilization of the wealth of data contained in the Linked Open Data cloud is mostly restricted to technical experts due to the complex nature of the available access mechanisms. In the past, large text repositories have been made accessible to the general public through a narrow set of document retrieval techniques, cumulating in simple search boxes supplemented with suggestion and refinement facilities. We expect that the highly structured nature of semantic entities, and the fact that semantic relations form a typed graph, will require techniques – and produce results – which go far beyond this state of the art. Visual interfaces will play an important role in both everyday search and in-depth analysis of linked data.

Visual Analytics, the science of analytical reasoning facilitated by interactive visual interfaces, combines automatic analysis, visual representation and user interaction in a closed loop intended to provide users with new insights [Thomas and Cook 2005]. This approach emphasizes the use of visual abstractions to represent aggregated information and facilitate the formulation and validation of hypothesis by expert analysts. Linked Data would benefit from the application of Visual Analytics methods, which could foster identification of hidden but valuable patterns and trends. Visual Analytics, on the other hand, would benefit from scenarios involving Linked Data which could foster development and evaluation of scalable Web-based visual analysis techniques for huge, distributed networks.

Knowledge Visualization utilizes visual representations to foster the communication of knowledge between persons (Eppler & Burkhard 2004). This approach emphasizes the use of visual metaphors to represent relevant information and facilitates collaborative dissemination and decision making by domain experts. Linked Data would benefit from utilizing Knowledge Visualization tools for crowdsourcing the creation of visual metaphors for communication of discovered facts and fostering collaborative creation

and gardening of knowledge. Knowledge Visualization, on the other hand, would benefit from scenarios based on Linked Data which could enable large-scale evaluation of educational, cultural and other effects on knowledge communication.

We suggest that application scenarios in the context of Linked Data should be addressed by a combination of Knowledge Visualization and Visual Analytics. Visualizing the complex semantic structures underlying Linked Data is hard. Making exciting new discoveries in Linked Data, and presenting them to a public audience in a comprehensible way, is even harder. By joining forces, Visual Analytics and Knowledge Visualization could tackle this challenge. The application to such a large and diverse scenario would boost research in both disciplines and potentially change the way in which knowledge is consumed on a global level.

In order to exploit these synergies, we envision a system for accessing and exploring Linked Data to be composed of two types of visual components, the actual selection and integration of which could be adjusted to a given scenario and user group:

Discovery Components focus on the flow of information from the system to the user. They facilitate the interpretation of data and the generation of new insights, for example through explorative navigation or analytical reasoning. Typical visual means include abstractions like charts and graphs as well as geospatial, temporal and network visualizations. A major challenge arises from the necessity to query various linked data sets, transform and normalize the retrieved data, and map the information onto appropriate visual channels. Manual configuration of such a visualization pipeline is clearly an expert task. A high degree of automation would be required to create accessible components, and could be achieved by (semi-automatically) aligning (Granitzer et al. 2010) relevant parts of the underlying ontologies to each other and to the semantically described visual components.

Description components focus on the flow of information from the user to the system. They facilitate visual metaphors empowering the users to efficiently and intuitively express newly learned facts and knowledge, with the purpose of transferring and communicating this knowledge to a broader audience. The system would offer several types of visualization metaphors suitable to address the user's goal. Starting from an empty skeleton of the chosen metaphor the user would apply a Semantic Visualization Builder Tool to construct a visual representation expressing the newly acquired knowledge both visually and semantically. When published, such a visual representation would provide a social platform for other users to learn from and to extend with further knowledge. As the knowledge expressed in the visualization is effectively incorporated into LOD issues like quality and trustworthiness of the information pose a major challenge.

Visual interfaces for discovery and communication of knowledge have the potential to significantly increase the accessibility and applicability of Linked Data: the cloud will become an enabling platform for everyone to engage in collaborative knowledge creation and collective thinking. In analogy to the rise of Web 2.0, large social networks building around Linked Data will fuel the growth of Web 3.0. Usage will be increased and with usage comes sustainability. So let us start exploiting the synergies by bringing intuitive visual interfaces and Linked Data closer together.

Breslin, J. G., Passant, A., & Decker, S. (2009). *The Social Semantic Web*. *Media* (p. 300). Springer. doi: 10.1007/978-3-642-01172-6.

Eppler, M.J., Burkhard, R.A., Knowledge Visualization – Towards a New Discipline and its Fields of Application, ICA Working Paper #2/2004, University of Lugano, Switzerland, 2004.

Granitzer, M., Sabol, V., Onn, K.W., Lukose, D. and Tochtermann, K., Ontology Alignment - A Survey with Focus on Visually Supported Semi-Automatic Techniques, Future Internet, Volume 2, Issue 3, pages 238-258, MDPI AG, 2010.

Keim, D., Mansmann, F., & Thomas, J. (2010). Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations Newsletter*, *11*(2), 5–8. ACM. Retrieved February 20, 2011, from http://portal.acm.org/citation.cfm?id=1809403.

Steinhart, G., (2011). DataStaR: A Data Sharing and Publication Infrastructure to Support Research, Agriculture Information Worldwide, Vol. 4, No. 1, pp. 16-20.

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics* (p. 186). IEEE Computer Society. Retrieved from http://nvac.pnl.gov/agenda.stm.