Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-349

Visualization of Large Ontologies in University Education from a Tool Point of View

Michael SCHAAF[a,1], Franziska JAHN[a], Kais TAHAR[a], Christian KÜCHERER[b], Alfred WINTER[a] and Barbara PAECH[b]
[a] *Institute of Medical Informatics, Statistics and Epidemiology, Medical Faculty, University of Leipzig*
[b] *Software Engineering Heidelberg, Institute for Computer Science, University of Heidelberg*

Abstract. The realization of ontology visualization requirements in university education is a challenging task and should be supported by appropriate tools. This applies in particular, if the ontology to be visualized is based on a large text corpus that comprises a huge number of concepts, relations and annotations. In *SNIK*, we developed such an ontology of information management in hospitals in order to support the transfer of knowledge in the context of the university education. The challenge is to identify tools and methods, which are capable to support ontology visualization and usage as efficiently as possible. Related research fields (e.g. bioinformatics) are confronted with similar visualization problems. These tools and methods used could provide a suitable solution in our research field. In total, we assessed eight tools concerning the visualization of large ontologies to evaluate their suitability representing knowledge in the field of medical informatics.

Keywords. ontology, ontology visualization, ontology visualization tools, knowledge representation

1. Introduction

In medical informatics, students need a deep understanding about the management of health information systems (HIS) and the interrelations between the different concepts of strategic, tactical and operational information management as well as the concepts of IT governance and IT service management. A variety of approaches of information management, especially in textbooks, address the diversity of task and information exchange from their point of view without providing formal models [1][2][3]. The absence of formal models, hampers the access, retrieval and usage of information in a certain area of knowledge [4]. For this reason, ontologies were introduced in order to facilitate information access and information retrieval by explicitly defining the specific concepts, relations and instances. In *SNIK* (semantic network of information management in hospitals), we developed a theoretically and empirically founded and tested ontology of information management in hospitals [5]. The main focus now is on the visualization, usage and query of this large ontology in order to use its formalized

[1] Corresponding Author: mschaaf@imise.uni-leipzig.de

knowledge in the context of the university education. In the last two decades, ontology development and, subsequently, the development of ontology development methodologies have gained considerable attention [2],[6]. Thus, quite a number of ontology development and visualization tools have been proposed [7]. In related research fields (e.g. bioinformatics), the visualization and usage of large ontologies (e.g. *Gene Ontology - http://geneontology.org) is a known* problem and the tools and methods used could provide a suitable solution in our research field [8]. *SNIK* is intended as the basis for two different application scenarios. First, several visualization and query tasks are important within the university education and for the international Frank-vanSwieten-Lectures (FvSL) on strategic information management in hospitals. Based on an interactive blended learning scenario, we want to teach students of medical informatics the information management concepts and their interrelations. Second, the ontology will be used as the underlying data for a decision-supporting tool for CIOs. Therefore, the focus of this paper is on introducing and evaluating eight tools supporting the visualization and usage of large ontologies exemplified by *SNIK*.

2. Materials and Methods

A performant visualization and the query of large ontologies without noticeable skips or dropped frames is a problem for many visualization tools. The *SNIK* ontology comprises more than 2000 concepts represented by nodes, 4500 relations represented by edges and approximately 15000 annotations containing additional information about the domain (e.g. definition, type, synonym, reference). The original format of our ontology is a spreadsheet file in csv format. When using an ontology within the university education, students need a deep understanding of its structure, content and the general rules of application. This is quite easy with regard to small ontologies. However, as the size of the ontology grows, the task of understanding will become considerably harder. Therefore, the visualization of large ontologies necessarily requires an appropriate tool environment, which enables an easy access to the modelled knowledge and limits the amount of information that students receive. This means, for example, that tools need to provide filter and query functions in order to focus on specific parts of the ontology. Typical teaching scenarios require to identify the interrelation between two given concepts. If we interpret the ontology as a directed graph, we will need basic graph algorithms to identify a path (e.g. the shortest path) between those two concepts. In another case, the immediate neighborhood of a concept and its interrelations might be important. In SNIK, one could identify all interrelations between the two concepts “Chief Information Officer” and “Chief Executive Officer”. For the implementation of our ontology-based teaching scenarios we determined some important key criterions for the visualization and features for the usage of the ontology:

- **Performance:** performant ontology visualization and navigation with no noticeable skips or dropped frames
- **Maintenance:** import, export functions (csv, owl), ontology manipulation
- **Usability:** general tool handling, documentation, plugin-management
- **Functions:** concept search, concept property overview, creation and query of subgraphs, graph algorithms (e.g. shortest path, concept neighborhood), static views, hierarchy overview, display filters
- **Topicality:** up-to-dateness of the tool**,** further development

The tools presented in [7] provide an excellent overview of the latest tools concerning the visualization of large ontologies. We tested more than 20 tools for their fundamental suitability concerning the visualization of large ontologies by importing *SNIK*. A few tools were not capable to import the ontology and were not assessed.

3. Results

3.1. Knowledge Visualization with Protégé

Protégé [9] is one of the most frequently used tools for developing ontologies and provides, beside a simple tree view, numerous visualization plugins in order to navigate on the hierarchy and relationships of ontologies. The import of *SNIK* (csv) was realized using the self-developed tool *Excel2OWL* [10]. The integrated *OntoGraf*-plugin of *Protégé* offers several automatic layouts and visualizes classes as nodes of graph and the relations between them, unfortunately, as unlabeled edges. With regard to *SNIK*, the visualization of a large number of nodes quickly reached its limits. It is impossible to "freeze" the arrangement of a modelled subgraph and add some further concepts or save the current view. Thus, the user easy loose track and a traceable construction of relevant parts of the ontology is almost impossible. Furthermore, the range of functions is limited to arranging nodes or filtering nodes and edges according to their types. There are several more plugins available for Protégé (e.g. OWLViz, GraphViz, IsaViz), that are incapable to solve the mentioned problems or are just out-of-date.

3.2. Knowledge Visualization with Tulip

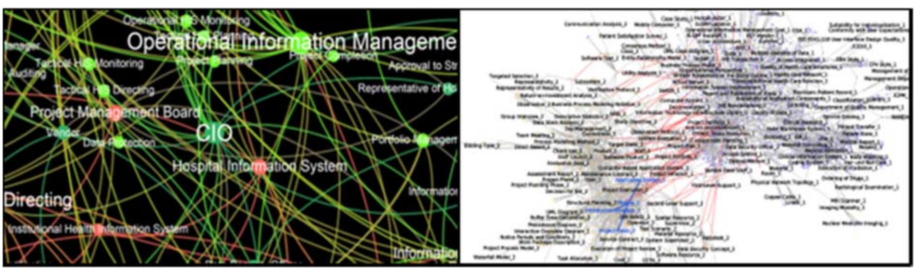

Figure 1. Fraction of the SNIK-Ontology visualized with Gephi (left) and Tulip (right).

The information visualization framework *Tulip* is another promising tool for the analysis and visualization of ontologies [11]. As shown in Figure 1, the visual complexity of large ontologies can be reduced by providing a wide range of node- and edge-metrics and several clustering algorithms. Tulip offers the opportunity to define and navigate large ontology hierarchies or nested subgraphs. The import of *SNIK* was easily accomplished by gradual adding the specified concepts and relations as nodes and edges. Each node in *SNIK* was supplemented by adding our specified annotations (definitions, synonyms, references). An important function of Tulip is the creation of subgraphs by choosing certain concepts or visualization shortest paths between concepts. This enables the visualization of important connected concepts without burdening the user with irrelevant concepts. In this way, required areas of *SNIK* could be analyzed and assessed separately. Tulip is easily extensible by providing a wide range of plugins, thus additional functions required for the visualization can be added.

3.3. Knowledge Visualization with Gephi

Gephi [13] is an open source tool for exploring large network graphs that focus on dynamic network visualization. The integrated visualization module allows the representation of large ontologies with up to 100.000 nodes and 1.000.000 edges in real-time. Gephi distinguishes between the three data types node, edge and attribute. In order to import *SNIK*, the tool provides an easy data import function for the integration of several data formats (e.g. csv, gexf, gdf, graphml). In a similar way to Tulip, we gradually added the specified concepts and relations as nodes and edges. Using the text module allows to show labels for each node from associated data attributes. For the visualization of *SNIK*, *Gephi* provides several highly parameterizable layout algorithms (e.g. Force Atlas, Fruchterman, Reingold) to increase clarity and readability.
In order to handle the complexity of *SNIK*, we created complex dynamic filters to query the graph and build subgraphs from the filtering results. The functions range from simple metrics for network analysis (betweenness centrality, closeness diameter, shortest path) to computational intensive clustering algorithms. The range of functions is comparable with Tulip and can be extended by plugins.

3.4. Knowledge Visualization with Cytoscape

The open source platform *Cytoscape* [12] was originally developed for visualizing, modeling and analyzing large molecular and genetic interaction networks. However, due to its flexibility by the provision of several plugins, the tool has become established in many fields of research. The data import plugin allows the import of *SNIK* by defining nodes, edges and attributes. Both, nodes and edges can have their own attributes. The tool provides, similar to *Tulip* and *Gephi*, several layout force-directed algorithms to layout the nodes and the edges of the network. As one of the few tools, it provides a plugin for importing the owl data format – unfortunately is not yet supported in the latest version version of *Cytoscape*. The creation of subgraphs of specific edges and nodes works smoothly and was much easier than with *Tulip* or *Gephi*. The basic function range of the tool is not that impressive but it can be easily extended by numerous plugins supporting clustering and network analysis.

Table 1. Assessment of selected tools for the visualization of large ontologies.

Tool	Performance	Usability	Maintenance	Functionality	Topicality
Ontograf	O	+	-	-	-
OWLViz	-	+	-	-	-
GraphViz	-	O	-	O	+
IsaViz	--	+	-	O	--
Tulip	++	+	++	+	++
Gephi	++	+	++	+	++
Cytoscape	+	++	++	+	++
Neo4j 2	+	O	O	+	++

4. Discussion and Outlook

The visualization and the usage of large ontologies in the context of the university education is a challenging task. The main reason is the lack of specialized and matured tools supporting the visualization and usage of large ontologies. Although there are several promising tools, many of them lack of basic important functions required for our ontology-based scenario for teaching the management of health information systems.

A major problem for many tools is a performant visualization of the huge number of nodes and edges of the ontology. A few promising tools (e.g. *Tulip*, *Gephi* or *Cytoscape*) may be able to visualize our large ontology, more specialized functions (e.g. edge detection and filtering) are not yet available or need to be programmed itself. In summary, none of the tools has convinced us completely concerning the criteria defined. On the other hand, the extensibility of *Gephi*, *Tulip* and *Cytoyscape* by plugins is promising for future developments.

Acknowledgement

This work is supported by the DFG (German Research Foundation) under the Project SNIK, Grant no. 1605/7-1 and 1387/8-1.

References

[1] A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung, and F. Jahn, Health Information Systems: Architectures and Strategies, 2nd ed. 20., vol. 64, no. 6. London, New York: Springer, 2011.
[2] L. J. Heinrich, R. Riedl, and D. Stelzer, Informationsmanagement: Grundlagen, Aufgaben, Methoden, 11th ed. Oldenbourg: De Gruyter Oldenbourg, 2014.
[3] E. Ammenwerth, R. Haux, and P. Knaup-Gregori, IT-Projektmanagement im Gesundheitswesen: Lehrbuch und Projektleitfaden - Taktisches Management von Informationssystemen, 2nd editio. Stuttgart: Schattauer GmbH, Verlag für Medizin und Naturwissenschaften; 2 edition, 2014.
[4] A. Heitmueller, S. Henderson, W. Warburton, A. Elmagarmid, A. S. Pentland, and A. Darzi, "Developing public policy to advance the use of big data in health care.," *Health Aff. (Millwood).*, vol. **33**, no. 9, pp. 1523–30, Sep. 2014.
[5] M. Schaaf, F. Jahn, K. Tahar, C. Kücherer, A. Winter, and B. Paech, "Entwicklung und Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus," in INFORMATIK 2015. GI-Jahrestagung, 2015, pp. 783–796.
[6] M. K. Mike Uschold, "Towards a Methodology for Building Ontologies."
[7] M. Bergman, "Listing of 185 Ontology Building Tools | AI3:::Adaptive Information," Listing of 185 Ontology Building Tools, 2016. [Online]. Available: http://www.mkbergman.com/904/listing-of-185-ontology-building-tools/. [Accessed: 11-Feb-2016].
[8] S. Lohmann, S. Negru, F. Haag, and T. Ertl, "Knowledge Engineering and Knowledge Management," in *19th International Conference, EKAW 2014*, Linköping, Sweden, November 24-28, 2014. Proceedings, 2014, p. pp 266–281.
[9] N. F. Noy, M. Sintek, S. Decker, M. Crubézy, F. W., and M. A. Musen, "Creating Semantic Web Contents with Protégé-2000," *IEEE Intell. Syst.*, vol. **16**, no. 2, pp. 60–71, 2001.
[10] K. Tahar, F. Jahn, M. Schaaf, C. Kücherer, B. Paech, and A. Winter, "Eine Ontologie für die Unterstützung der Lehre und des Informationsmanagements im Gesundheitswesen," in *13th Leipzig Research Festival for Life Sciences*, 2014, p. 45.
[11] D. Auber, D. Archambault, R. Bourqui, A. Lambert, M. Mathiaut, P. Mary, M. Delest, J. Dubois, and G. Melançon, "The Tulip 3 Framework: A Scalable Software Library for Information Visualization Applications Based on Relational Data," *INRIA*, Jan. 2012.
[12] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks.," *Genome Res.*, vol. **13**, no. 11, pp. 2498–504, Nov. 2003.
[13] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks.," in *International AAAI Conference on Weblogs and Social Media.*, 2009.