VIScover: Visualizing, Exploring, and Analysing Structured Data

VAST 2009 Flitter Mini Challenge Award: *Novel visualization of effect of rule application*

Thorsten Liebig*
Ulm University
derivo GmbH
Germany

Olaf Noppens†
Ulm University
derivo GmbH
Germany

Friedrich von Henke‡
Ulm University
Germany

Abstract

Today's challenging task in intelligent data processing is not to store large volumes of interlinked data but to visualize, explore, and understand its explicit or implicit relationships. Our solution to this is the VIScover system. VIScover combines semantic technologies with interactive exploration and visualization techniques able to analyze large volumes of structured data. We briefly describe our VIScover system and show its potential using the example of the VAST 2009 social network and geospatial data set.

Index Terms: I.2.4 [Computing Methodologies]: Artificial Intelligence—Knowledge Representation Formalisms and Methods; H.3.3 [Information Systems]: Information Storage And Retrieval—Information Search and Retrieval H.5.2 [Information Interfaces And Presentation]: User Interfaces—

1 Introduction

Access to knowledge within documents, databases, etc. has become a critical factor at a time where electronically stored data doubles at decreasing time frames. Many relevant pieces of information or correlations are stored electronically but archived in a way which prevents their access. Furthermore, common retrieval systems only offer insufficient search, analysis and visualization options.

The underlying problem is twofold. First, the currently used data representation mechanisms are extremely weak in preserving the meaning of the stored information. Data typically is stored in documents or tables without any meta-data which states their relationship between each other in an explicit and machine processable way. Second, traditional interfaces and presentation techniques do not provide adequate support for browsing through large volumes of interlinked data repositories or to uncover important coherences.

Our VIScover system solves this problem by combining semantic technologies with an innovative exploration and visualization approach enabling users to reveal and understand encoded knowledge. Pure data is enriched by a knowledge net whose hidden relationships and facts are unlocked by an inference process. Its interactive visualization interface supports even non-expert users in analyzing and exploring significant qualitative correlations. Novel content-aware abstraction, aggregation and filtering techniques allow to grasp complicated chains of dependencies effectively.

2 Treating the Social Network Semantically

The provided social network and geospatial data has been enriched to a knowledge net (aka. ontology) which reflects implicitly given characteristics of the domain in terms of ontological axioms. For instance, some of the relationships of the provided data set are symmetric, such as the `flitters` relationship amongst persons or the `nearby` relationship amongst cities. We have encoded these kind of textually expressed (and therefore only implicitly given) data constrains as well as the data itself as ontology axioms using W3C's Web Ontology Language (OWL) [2] as representation language.

This is the source of an inference process which automatically makes hidden findings explicitly available. We used the RacerPro [1] inference engine as our underlying reasoning system because it showed the best performance and service features for this social network ontology.

3 Meaningful Social Network Analysis

On top of this infrastructure there is our exploration and visualization system VIScover [3] that displays significant correlations within an ontology in a clear and concise manner using modern visualization and analysis techniques. It visualizes hidden information dependencies, abstracts from detail information when needed, and automatically groups data by its meaning. Even long chains of correlations and causalities can easily be tracked. With respect to the social network data, VIScover allows to incrementally explore branches of the flitter connectivity network in parallel. Relationships are graphically depicted as clubs originating from the set of source entities to the relationship's fillers. The "club visualization" metaphor of VIScover also displays quantities of relatedness and allows to qualify clubs with help of logical as well as numerical filters in an on-demand fashion.

4 Gaining Insight into the Social Structure

The main task of the VAST 2009 social network and geospatial challenge was to identify the most likely social structure from some given hypothesis. In order to identify such a hypotheses we defined the supposed social sub-structures as formal concepts of our ontological model. For instance, the suspected employee is assumed to have about 40 Flitter contacts where three of these contacts are his "handlers" which probably itself have between 30 to 40 Flitter contacts. Figure 1 shows this definition of `Suspected-Employee`

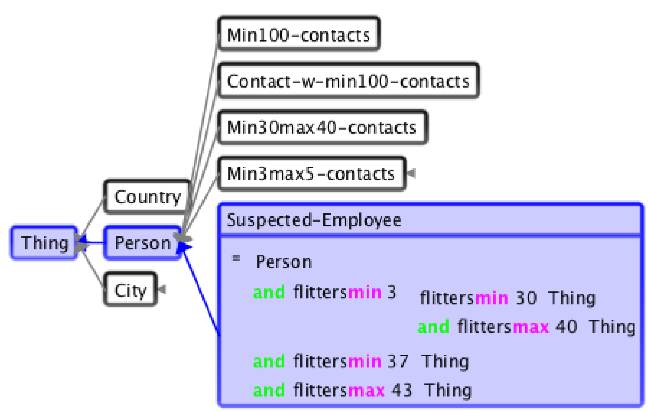

Figure 1: Ontology concepts for solving this mini challenge

*e-mail:thorsten.liebig@uni-ulm.de
†e-mail:olaf.noppens@uni-ulm.de
‡e-mail:friedrich.von-henke@uni-ulm.de

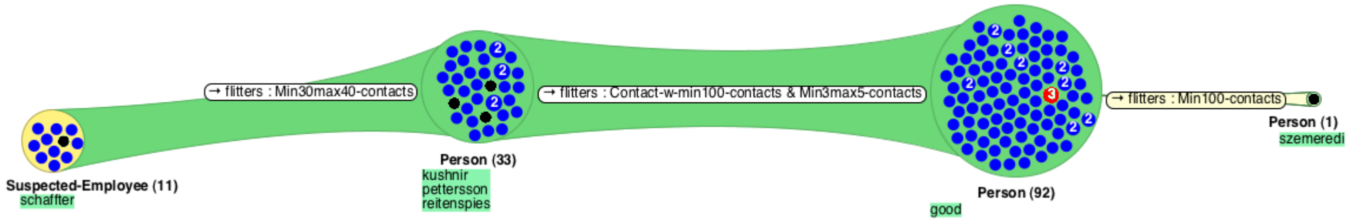

Figure 2: Identified social structure within the challenge data

(light-blue background), which formally characterizes this pattern using one of the OWL syntaxes as rendered with our tool. Figure 1 also depicts the remaining four other concepts (hiding their detailed definitions) solely needed to solve this challenge.

The set of individuals which comply to the definition of a suspected employee is computed (inferred) by the underlying reasoner as a logical consequence of the formal ontological axioms. When dropping the concept onto the explorer pane this set is visualized as a disk covering all the individuals as blue circles. This set of eleven persons, out of overall 6000, is shown in Figure 3.

Figure 3: The set of suspected employees

To validate one of the hypothesis it is necessary to expand the flitter contacts of those employees, so that we can identify the related handlers. This can be done by clicking on the set which then shows a preview of potentially expandable relations (those which actually have fillers wrt. the originating set). Figure 4 displays the relationship preview for the set of eleven suspected employees.

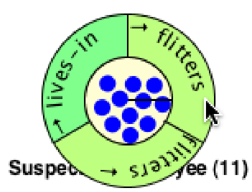

Figure 4: Preview of potentially expandable relations

As a result of expanding the flitters relationship, a club with all related Flitter contacts appears as shown in Figure 5. This allows users to interactively browse even through large volumes of interlinked data. At any time of interaction, a user can branch his exploration by expanding a second relationship or dig into the network by expanding any of the succeeding clubs.

Some of these 401 persons do have Flitter contact with up to four suspected employees. This sort of quantity (amount of related individuals of preceeding club) is shown by a number within the individuals circle in VIScover as can be seen in Figure 5. In addition the connectivity chain of related individuals in preceeding clubs are highlighted and their label are listed as an mouse-over effect (see green circles and lists of labels in Figure 5).

Now, hypotheses A assumes that the three handlers of the employee have 30 to 40 own Flitter contacts. We can easily narrow down the initial expansion by using the concept description `Min30max40-contacts` as a restricting filter by dropping this concept on the corresponding club expansion. A consequence to this is a cut down from 401 to 33 potential handlers (cf. leftmost club expansion in Figure 2).

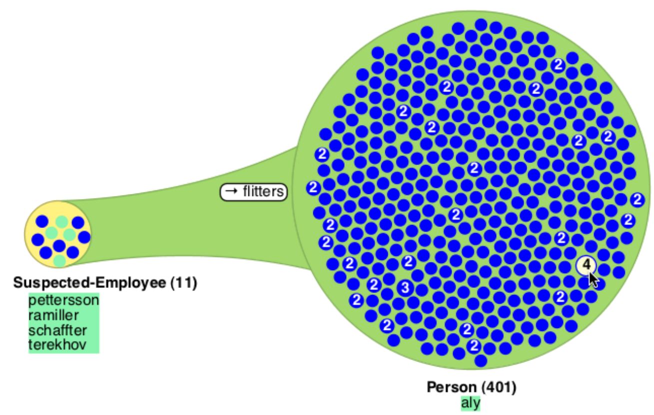

Figure 5: Graph after expansion of contacts of suspected employees

We finally can approve one of the social structures of the challenge (namely A) with just two more expansions and three additional filters. The resulting social network graph is depicted in Figure 2. It shows a structure consisting of an suspected employee (schaffter), three handlers (reitenspies, pettersson, kushnir), a middleman (good), and Fearless Leader (szemeredi).

5 Conclusion

Our exploration and visualization system VIScover displays significant correlations within a knowledge net in a clear and concise manner using modern visualization and analysis techniques. In combination with Semantic Technologies it visualizes hidden information dependencies and automatically groups data by its meaning. Even long chains of correlations and causalities can easily be tracked. Using semantic filters information can be refined from the perspective of its content. Furthermore, VIScover has more analysis features not mentioned here. For instance, its sliders allow to filter individuals with respect to associated numerical data ranges, such as time, age, etc.

References

[1] V. Haarslev and R. Möller. Racer: A Core Inference Engine for the Semantic Web. In *Proc. of the 2nd Int. Workshop on Evaluation of Ontology-based Tools (EON 2003)*, pages 27–36, 2003.

[2] B. Motik, P. F. Patel-Schneider, and I. Horrocks. OWL 2 Web Ontology Language: Structural specification and functional-style syntax. W3C Working Draft, 11 April 2008, World Wide Web Consortium, April 2008.

[3] O. Noppens and T. Liebig. Understanding Large Volumes of Interconnected Individuals by Visual Exploration (System Description). In *Proc. of the 4th European Semantic Web Conference (ESWC 2007)*, pages 799–808, Innsbruck, Austria, June 2007.