

# Realizing the Hidden – Interactive Visualization and Analysis of Large Volumes of Structured Data

Olaf Noppens  
Institute of Artificial Intelligence  
Ulm University, Germany  
olaf.noppens@uni-ulm.de

Thorsten Liebig  
Institute of Artificial Intelligence  
Ulm University, Germany  
thorsten.liebig@uni-ulm.de

## ABSTRACT

An emerging trend in Web computing aims at collecting and integrating distributed data. For instance, community driven efforts recently have build ontological repositories made of large volumes of structured and interlinked data from various Web sources. Those repositories are extreme in the sense that they are extraordinary in size and dominated by assertional data incorporating only a small and typically lightweight schema. So far, users can find tools for building and browsing through large schemas but there is virtually no support in navigating, visualizing or even analyzing the data part of such a structured repository appropriately. This paper describes how to combine techniques from visual analytics and logical reasoning for interactive exploration of large volumes of interrelated data. Our approach utilizes visual abstraction techniques, semantical filters, as well as various level of detail information. The underlying visualization rationale is driven by the principle of providing detail information with respect to qualitative as well as quantitative aspects on user demand while offering an overview at any time. By means of our prototypical implementation and two real-world data sets we show how to answer several data specific tasks by interactive visual exploration.

## 1. MOTIVATION

The two building blocks of an ontology are its terminology or schema and its assertional data. The terminology introduces concepts and properties and gives structure to them in terms of defining axioms using the language constructs of the underlying ontology language. The data part originates concrete individuals and relationships between those individuals utilizing the definitions of the terminology.

As an example, consider a social-network ontology describing members of research communities defined by concepts such as `Project`, `Person`, `Publication`, `Institution` as well as properties such as `has-Author`, `has-Project-Member`, `has-Research-Interest` etc. A corresponding real-world ontology will define a general vocabulary for refining these terms.

A researcher, for instance, can be a working person, an affiliated person, a student, a PhD student, an employee can be characterized by his title, his affiliations, his degrees, his project memberships in the past and present etc. Keeping this in mind, a corresponding data part spans a network of people associated with their current and past affiliations, project memberships, authorships, research interests etc. Besides rather simple questions such as about the project membership of some people one can also use the information repository to answer more complex questions such as “Does a specific author (or a group of authors) tend to publish with certain co-authors or with people from certain affiliations?”, “Are there isolated affiliations within a specific research area?” or “Are there research topics which were only studied by PhD students?”.

The ratio of terminological to assertional data of an ontology depends on its intended purpose. So called reference ontologies, such as the SNOMED CT or NCI medical terminologies, try to serve as an unifying vocabulary for various other schemas. Therefore, they typically are made up of terminological axioms solely. On the other hand, real-world domain ontologies mostly contain a much larger amount of individuals as opposed to concept or property definitions. For instance, the data volume of the Gene Ontology (GO) [4] exceeds the number of concepts about seven times. More often, assertional data of those ontologies outnumbers the terminological part by more than thousand times (i. e. SEMINTEC). Artificial data suites such as the Lehigh University Benchmark (LUBM) [5] try to reflect this in that they increase the number of individuals rather than the number of concepts or properties for benchmarking. Not surprisingly, scalability of reasoning with assertional data is still an actual research issue albeit recent optimizations have shown significant increase in speed for answering queries with respect to large volumes of individual data under specific conditions [8].

At the same time the trend of collecting and integrating distributed data into one large repository is gaining more and more momentum. As an example, community efforts such as DBpedia [2] or ReSIST [1] have recently extracted large volumes of structured data from the Web (Wikipedia, US Census Data, DBLP, Citeseer, ACM, etc.). Those repositories are extreme in the sense that they are extraordinary in size and dominated by assertional data incorporating only a small and typically lightweight schema. So far, users can find tools for building and browsing through large terminolo-

gies (i. e. Protégé, SWOOP, TopBraidComposer) but there is virtually no support in navigating, visualizing or even analyzing the assertional part of an ontology appropriately.

In this paper we will present our approach of combining techniques from visual analytics and reasoning for interactive exploration of large volumes of interrelated instance data. The following will describe the various selection, exploration, and filtering techniques which have been implemented and integrated into our ONTOTRACK [11] ontology authoring environment. This is done on the basis of two data sets introduced in the next section.

## 2. DATA SETS

For the rest of the paper, we have chosen two real-world data sets from different domains in order to show how to make implicitly encoded knowledge explicitly visible with help of our approach. The first one has been extracted from the MONDIAL Database and the second from the ReSIST Network of Excellence. Both assertional data sets consist of more than hundred thousands of individuals and were mapped to the Web Ontology Language OWL 1.1 [14].

### The MONDIAL Database

The Mondial Database<sup>1</sup> (MONDIAL) is a collection of geographic information compiled from different Web data sources such as the CIA World Factbook, Global Statistics and the Terra database [12]. The core of a MONDIAL record consists of data about countries, cities, as well as deserts, rivers, or ethnic groups mainly collected from the CIA World Factbook. In addition the collection includes statistical data about populations, area, or length. There are several formats and extracts of MONDIAL freely available. First, we mapped this data set into OWL format via a XSL transformation. We then enriched the assertional data with a manually created terminology describing concepts such as **Country**, **City**, **Water-Area**, **Continent** and properties such as **has-City**, **encompassed**, **has-province-capital**. For instance, the latter property relates a city to a province and is defined as a subproperty of **has-City**. Furthermore, for most properties we have added an inverse counterpart (e. g. **access-to-river** is the inverse of **flows-through-country**). Such a basic vocabulary at hand we defined more complex concepts. For example, a **EuropeanCountry** has been defined as a **Country** which is encompassed by the continent **Europe**. In addition, we have enhanced the resulting ontology with additional information such as a reference to a country flag gathered from the web.

### ReSIST Network – Resilience Knowledge Base

The Resilience Knowledge Base (RKB) has been created during the first year of the European Network of Excellence in Resilience Computing. The RKB aims at supporting researchers in accessing knowledge on resilience concepts, methods, tools, and the community itself. For that purpose, resilience data has been captured from each partner’s information resources such as research interest details and courseware. This data has been complemented by external sources captured from research information services CORDIS and NSF. Moreover, meta-data about publications

<sup>1</sup><http://www.dbis.informatik.uni-goettingen.de/Mondial/>

and the RISKS index of “Computer-related Risks to the Public” has been gathered from the Citeseer and ACM repositories forming a social-network of researchers and publications. The modeling also includes special ontologies for Dependability and Security and uses vocabulary of the portal and support ontology of the AKT project<sup>2</sup>. Both vocabularies are general purpose ontologies describing, among others, people, projects, and publications and their relationships in a very fine-grained way introducing detailed taxonomies. For instance, an **Affiliated Person** is a person which has an affiliation with some organization (multiple affiliations are allowed), and a **Student** is defined as an affiliated person which studies at an **Educational Organization** and is affiliated within that organization to an **Organization Unit**. Moreover, such taxonomies can also be found for relationships, e. g. **works-for** is a subproperty of **has-affiliation** because if a person works for an organization the organization is also its affiliation. Figure 1 shows part of the taxonomy of definitions in the upper part, e. g. the concept definitions of person and student in a logical notation. The data is held in a RDFS triple store and accessible via a SPARQL interface.<sup>3</sup> The system incorporates a consistent reference service which maps different URIs from various sources into one reference [9]. We extracted a significant amount of data from the RKB containing more than hundred thousand individuals and their relationships to a local repository in order to avoid web-based latency delays. Thereby redundant URIs have been cleared as the whole data set was translated into the OWL 1.1 format.

## 3. VISUAL ANALYSIS THROUGH INTERACTIVE EXPLORATION

Analysing real-world domain ontologies such as those described above is a complex task when it comes to repositories containing a large amount of heavily interlinked individuals. We claim that a combination of semantic technologies with methods from Visual Analytics, two currently very agile and vivid research areas, will generate new and strongly needed data discovery and illustration options able to uncover hidden and entailed information. Visual Analytics is the formation of abstract visual metaphors in combination with a human information interaction. To a certain extent it also allows for visually reasoning about data from a human perspective. However, it is an outgrowth of many scientific fields and includes technologies from a wide range of areas and unfold its power in combination with them [17].

One lesson learnt from the visual analysis of large data sets in general is that it is not advisable to arbitrarily visualize both all dependencies and all particulars at any time [10]. Therefore, our approach follows the *Visual Information-Seeking Mantra* of “Overview first, zoom and filter, then details-on-demand” [16] by providing detail information on user demand while offering an overview at the same time.

Figure 1 shows a part of the hierarchy of concepts defined in the schema data of the RKB on the upper part and the data analysis pane on the lower part. The interaction, clustering and exploring techniques of the concept hierarchy pane is out of scope of this paper and can be found in our description

<sup>2</sup><http://www.aktors.org/>

<sup>3</sup><http://www.rkbexplorer.com/>

of the ontology environment ONTOTRACK [11]. However, as shown in the following, there will be a tight interaction with the concept hierarchy pane while interactively exploring and analysing the data.

### 3.1 Reasoning

Analysing ontological data without taking its semantics into account is not of much use. Just considering the given triples of a RDF(S) model or the told axioms of an OWL ontology does not reflect the intended meaning of the knowledge base. A serious data analysis need to uncover entailed information by making implicit knowledge explicitly available. This has to be done by a reliable reasoning system. For instance, in OWL the subconcept relationship is transitive. Without reasoning, this information would be hidden such that in the RKB domain, a student would not be classified as a person. As mentioned before, a student is a subconcept of person which is also reflected in the concept hierarchy shown in the upper part of Figure 1d. Note that classification of data also means that it is not necessary to explicitly state that an individual “StudentA” is an instance of the concept **Student**. Instead this information can also be automatically derived from property definitions and property instantiations: For instance, if the domain of `studies-at` is declared to be a student and it is known that “StudentA” studies at “OrganizationA” (property instantiation), “StudentA” will be derived to be an instance of **Student**. Another example deals with the fact that in OWL individuals may have different identifiers and are not disjoint by default. Consequently, with help of a reasoning system, two fillers of a functional property (i.e. a property with at most one filler per source) will be merged to one single individual. In addition, sophisticated data exploration or filtering tasks correspond to queries which can only be answered with help of a reasoning engine. Our approach therefore inherently requires to have access to a high-performance reasoning system. Besides RacerPro [6] our tool particularly relies on the relational reasoning engine U2R2 [18] when dealing with assertion data at very large scale.

### 3.2 Abstraction and Clustering

Following the Information-Seeking Mantra and similar studies, all the connections and relationships between individuals within an ontology can not be visualized and understood at once. From the modeling point of view, concept and property filler extensions seem to build obvious or “natural” clusters with concise semantics. In both cases individuals with similar characteristics are pooled. For instance, in the MONDIAL domain, all capitals of European countries, and all headquarters which are located at a specific country can be pooled within a so-called *property filler cluster* as shown in Figure 2. Here, individuals are visualized as small filled circles and clusters are drawn as clubs originating from the individual which is considered as the origin of the filler set (e.g. capital **Paris** in Figure 2), i.e. the subject of the inverse property `headquater-at`.

From an abstract point of view, there is a direct mapping between visual representation and semantics: in other words, each cluster can be exactly described with a logical expression. Following this idea, the origin of a cluster club need not to be a single individual but a set of individuals which are considered as a logical union. For instance, in Figure 3

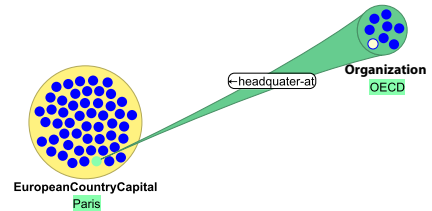


Figure 2: Clustering with clubs and circles.

the cluster on the right hand side contains the union of all fillers of the predecessor cluster, i.e. countries to which the European capitals belong to. This general representation principle allows to map the instances of any legal Description Logics expression to be clustered and abstracted with help of a club. As one can see in the previous figures, each club is labeled with the corresponding expression, resp. in the case of named concepts with the label of those concepts (e.g. `EuropeanCountryCapital` in Figure).

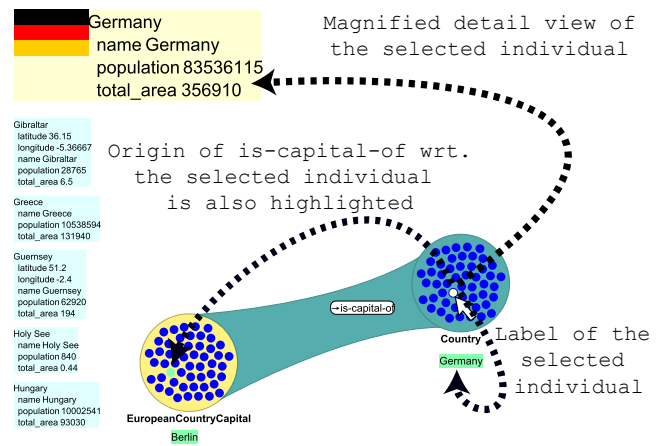


Figure 3: Group expanding club.

Abstraction also means that only individuals in a cluster are drawn if their number is below a user-definable limit. Moreover, the diameter of clusters showing no individuals explicitly approximates the number of individuals and allows to easily compare the number of individuals by the clusters rendering size. In addition, the number of fillers are drawn within a cluster as can be seen in the upper cluster of Figure 1. To downsize the number of individuals, other methods are available as shown later in the section 3.5.

Only relationships between concepts (i.e. object properties) are visually represented via clubs, values of datatype properties such as the title of a publication or additional country information such as the total size of its area and number of the population are available in a detail view when hovering over a cluster as shown on the left hand side of Figure 3. When hovering over an individual with the mouse pointer the list of detail view entries will be scrolled to the corresponding entry and it will be magnified. It is possible to determine which datatype properties are shown in the detail entry and to specify which datatype properties provide additional graphical information such as an image.

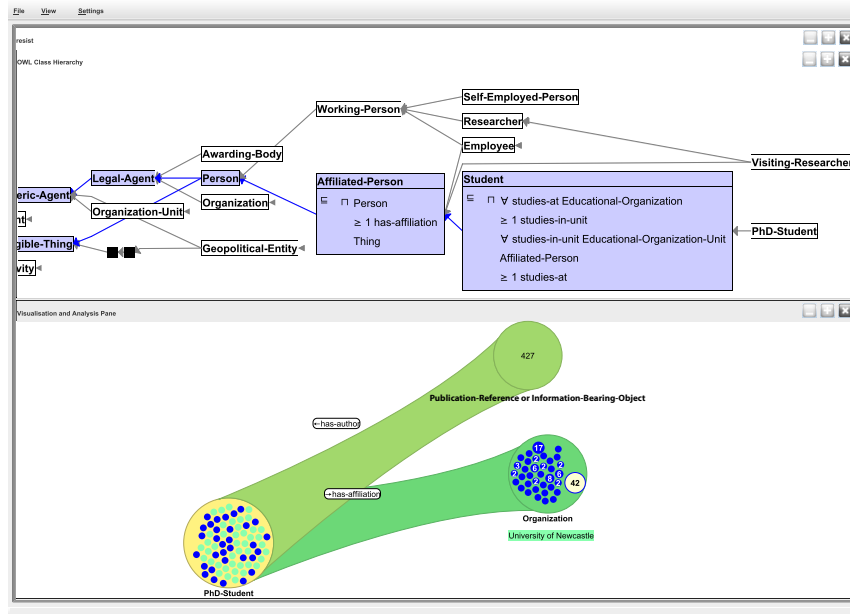


Figure 1: Application screen shot showing schema pane on the upper part and data analysis pane on the lower part.

To easily grasp the individual(s) a specific individual within a cluster is related to with respect to the given properties all individuals of the origin of the property and, if further cluster expansion is available, of the fillers of the corresponding property with respect to the highlighted individual are recursively highlighted, too. In addition, the labels of these individuals are rendered at the bottom of the cluster. For instance, in Figure 3 the mouse pointer is hovering over the individual “Germany”. As a result its label is rendered and because “Berlin” is the only origin of the *is-capital-of* property the corresponding graphical representation is also highlighted and its label also rendered at the bottom of the *EuropeanCountryCapital* cluster.

### 3.3 Interactive Exploration

As clustering helps to abstract from specific data in an overview manner, a user-directed interactive exploration strategy allows for concentrating on interesting parts of an ontology, or fractions which promise to unveil deeper insights. Initially, one can either start with a user selected individual or with the instances of a concept via ‘n drop operations. This will result either in showing the graphical representation of that individual or a circle containing all individuals which are instances of the given concept. For instance, one can drag a concept from the concept view as shown on the upper part of Figure 1 on the data analysis pane on the lower part. As mentioned in the previous subsection, only individuals in a cluster are drawn if their number is below a user-definable limit. Otherwise only the number will be shown within the graphical cluster representation.

After clicking on the graphical representation of an individual or a cluster a graphical preview menu of all object properties an individual or, in the case of a cluster, the union of individuals is related to. To allow a more natural exploration of an ontology, the exploration direction is

not limited to the direction of the property but also allows the inverse direction (even if no inverse property has been explicitly defined). For instance, Figure 4 shows that the individual labeled “Thorsten Liebig” is related via the property *has-author* with exactly 48 publications. The properties are arranged in a radial tree layout reflecting the underlying property hierarchy as shown in Figure 4. Here, one can easily grasp that the property *has-affiliation* is a superproperty of the *works-for* affiliations. All properties are encoded with the same color schema with a decreasing brightness according to the hierarchy level. Due to the tree-like layout, for a property with several superproperties its preview slice will be duplicated and drawn below the previews of each of its superproperties. However, we do not follow the color encoding for properties with multiple superproperties. Instead of choosing different colors because of the different color schemas of the superproperties, one color schema is used for all these representations. Note also that the color of a property and its inverse is always the same.

The arrow left to the property label indicates the direction of the expansion. A left directed arrow denotes that the actual root of the cluster is the filler of the property. Moreover, the size of each property slice can be weighted with respect to the number of fillers while preserving the constraint that slices of subproperties must not overhand the slices of the corresponding superproperties. Though this representation gives a good understanding of the distribution of property fillers, property labels can not always be statically provided, instead when hovering over a property slice with the mouse pointer the corresponding name is shown. Preliminary experiments have shown us that this is not a drawback as the user is used to the color-property encoding rapidly while using the system.

Each cluster as well as each (visible) individual can serve

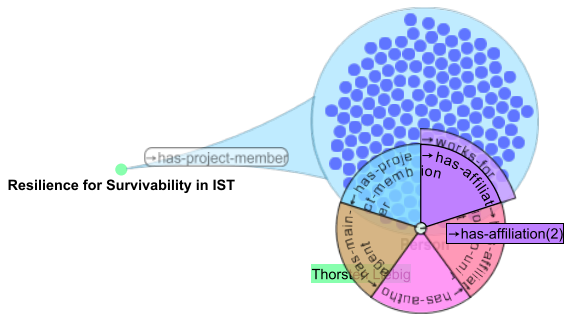


Figure 4: Previews for expanding property fillers.

as a follow-up point for further expansions. This also allows to branch the expansion by selecting other properties or de-expand clusters. More information about the various expansion options can be found in [13].

### 3.4 Analysing Quantities

Though it depends on the kind of questions to answer and on the ontology domain there are lot of questions that inherently require to take quantities into account. Even if one can easily grasp how many individuals are related to a given one with respect to a given cluster club source by counting the individuals within the club, it is more complex to visually answer how many individuals in a cluster are related to one specific individual within the successor cluster. Note that this is completely independent from the fact whether it is an inverse property or not. For instance, to answer the question within the MONDIAL domain which is the country in which the highest number of deserts can be found, one would start with the cluster representing all deserts as one can see on the left hand side of Figure 5. After expanding this cluster with respect to the `located-in-country` property one would get a cluster consisting of all corresponding countries. Even if one could manually analyse and count for each country how many deserts it has it is more appropriate to visually show quantities which a very important issue to better understand information. Derived from well-known visual analytics methods we propose a simple but powerful solution: the diameters of each genre circle scales proportionally with the number of related deserts in the predecessor club. In case that there are more than one source individual their number is also drawn within the individual circle as shown in Figure 5. One can also see that the most deserts can be found on the African continent (when hovering with the mouse pointer over the circle labeled “14” in the `encompassed` cluster the individual’s name is shown).

Even if the original question does not refer to quantities, an additional visual receiving information about quantities is a good benefit. Consider a follow-up expansion of Figure 4 (b) to get all co-authors of all publication of “Thorsten Liebig”. At a glance one gets the information which person is the co-author of most of the publications as shown in Figure 6 (a). The circle labeled “48” is Thorsten himself. Note that the same approach could answer from which affiliation tend to come his most co-authors (Figure 6 (b)).

### 3.5 Dimensions of Filtering

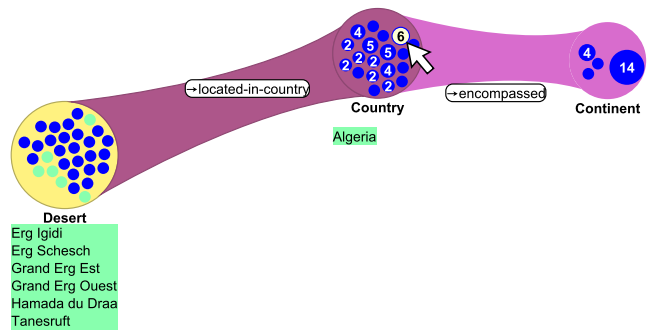


Figure 5: Utilizing quantities to answer questions such as “In which countries can be found the highest number of deserts?”.

As shown before, abstraction based on clustering allows for a bird-eye view of large volumes of interrelated individuals, and close-by detail information about individuals is provided on user demand. However, this technique which is widely used in information visualization only considers two points of a fictitious scale of granularity, namely the lower and upper point. Filtering is another important step towards a goal-directed visual analysis. In our approach one can distinguish two kinds of filter dimensions: local filtering based on object properties and global filtering on datatype properties. Remember that values of datatype properties are not visible as first-class elements but only in the detail view such as the length of rivers, the population of countries, the depth of seas in the MONDIAL domain.

*Local filtering* allows to refine a cluster to those individuals which are instances of an arbitrary concept definition in the schema: To establish such filter, one can drag an arbitrary concept from the concept hierarchy view of ONTOTRACK onto an existing cluster. For instance, within the MONDIAL domain, any set of countries can be restricted to European countries simply by dragging the `EuropeanCountry` concept onto a cluster with country instances as shown in Figure 7.

Three interactive paths are possible: first one can drop an existing concept onto a non restricted cluster to restrict the cluster accordingly. Second, one can drop it onto an already restricted cluster. This will be interpreted as a logical conjunction of the existing filters and the given one. Third, a new (temporary or permanent) concept can be defined within the concept hierarchy view by using all graphical features and guidance of ONTOTRACK to allow any arbitrary concept expression such as nominals, complements, unions. This is a powerful selection mechanism since concepts can be defined on demand during exploration. The underlying reasoning engine supports the user by instant classification of individuals as well as in providing feedback about nonsensical or even unsatisfiable restriction. For instance, if fresh water areas and salt water areas are modeled as disjoint concepts, their conjunction is unsatisfiable. Therefore, the unsatisfiability of the combination of a fresh water area and a salt water area (or even subconcepts of such as lake resp. sea) filter is recognized and the user is warned to avoid such unsatisfiable filter combinations. Moreover, a local filter may also affect adjunctive or prospective clusters. Consequently, a filter will propagate its restrictions. As a

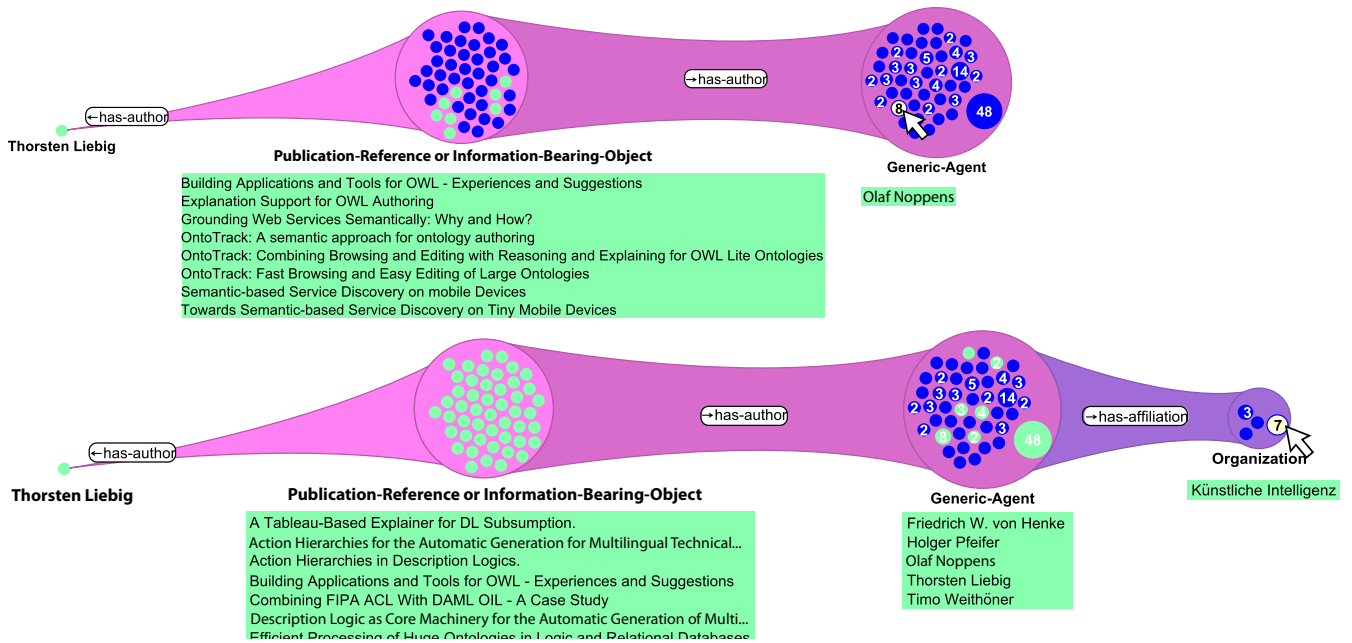


Figure 6: (a) Who is the most co-author of papers involving Thorsten. (b) From which affiliation the most co-authors come from?

result, each prospective cluster will adapt its set of individuals accordingly. This might also mean that whole clusters will disappear in an animated fashion if they do not longer possess any expansion source.

*Global filters* are based on values of datatype properties and are applied to all visible individuals. In contrast to local filters, filtering out an individual does not imply its removal from clusters by default. This is motivated by typical queries and the typical usage of datatype properties: on the one hand their values are only strings such as names, labels, comments, addresses. On the other hand they are based on enumeration types (even user-defined types) as well as numerical data. Even if string matching is a powerful filter mechanism, from a visual analytics perspective it is more or less a utility method and we do not consider it, for string matching please refer to a description of our preliminary system [13]. Numerical datatypes such as integers possess a minimum and maximum value. However, for a given knowledge base, the real used minimum resp. maximum value may be different from the declared ones. Similar to enumeration types (e.g. months) they span a finite range. For instance, in the MONDIAL domain rivers are correlated to specific length. As one can see in Figure 8 concrete domains of datatype properties can be adjusted with sliders. Here, all those rivers which are briefer than 5066 kilometers will be instantly colored grey and one can easily grasp those which fall into the interval. As one can see, the quantity measurement of individuals in the cluster containing the seas are also proportionally greyed out, e.g. there are 9 of original 17 rivers which flow into the Atlantic Ocean.

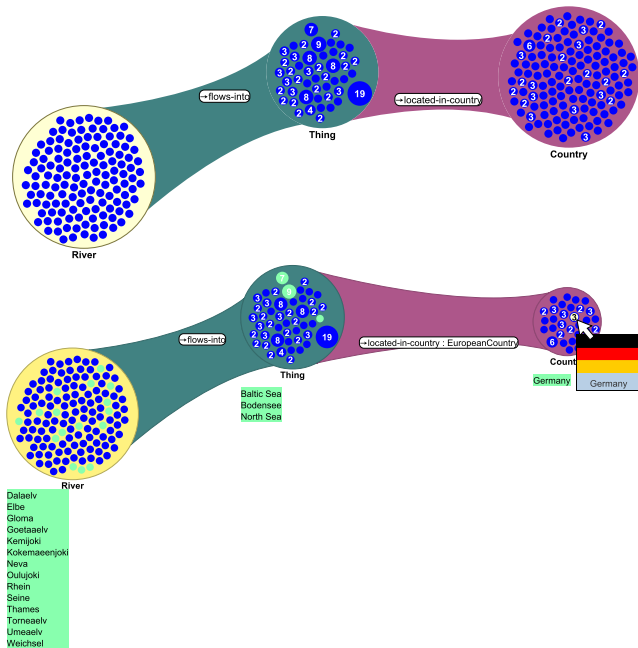
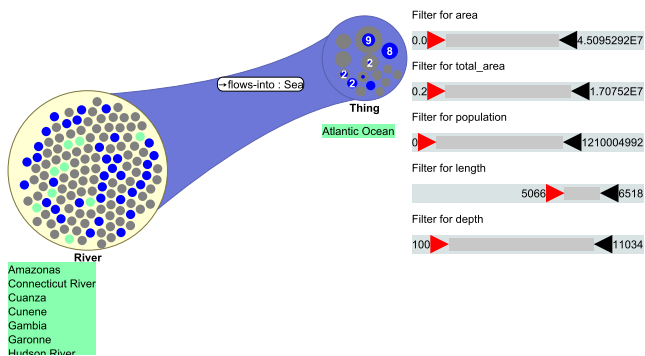


Figure 7: Local restriction of countries to European countries.

Applying several slider filters to the data set, the conjunction of the single filters is considered. However, in less structured data sets such as the RKB, it has shown that concepts



**Figure 8: Global restriction via sliders (e. g. wrt. the length of rivers).**

are not strictly defined with respect to datatype properties. For example, none, one or more publication years can be assigned to a publication. This is because the data has been automatically extracted from different sources such as publication lists of homepages which does not guarantee that publication years are available and could be extracted. This directly leads to the question how to deal with individuals which could have (a) specific data value(s) according to the concept description but does not have any, individuals which have different data values for a specific property. In both cases, an individual which have at least one value for a specific property which should be filtered out, will be drawn in grey.

#### 4. IMPLEMENTATION

In contrast to widely-used small or middle-sized ontologies, real-world domain ontologies consist of thousands of thousands of individuals and relationships between them. This makes great demands on the scalability and performance of knowledge-based systems. In our case, the challenge refers to the visualization, analysing as well as the reasoning component. Note that the instant reasoning process is a key feature within our approach and therefore the visualization and analysing methods are closely coupled to it. As our approach is highly interactive, exploration and analysis steps should not be performed with long delays.

From a visualization perspective, we address this requirement with our decision for the Piccolo framework which can manage huge numbers of graphical objects and allows for arbitrary animations, zooming and scaling [3]. The presented visualization and analysing component has been developed as a plugin-in for our ONTOTRACK ontology environment which itself utilizes Piccolo for its interactive visualization of both the concept and property hierarchy. The visualization of these hierarchies is an extension to the elaborated SpaceTree paradigm offering features such as animation of changes and thumbnails [15]. Moreover, our component benefits from ONTOTRACK's elaborated authoring and interaction features, as well as its ontology management capabilities which are based on the wide-spread Java OWL 1.1 API [7].

As already mentioned our component closely relies on a reasoning engine. Because it is absolutely not feasible to compute any analysis results in advance, analysis is performed

on user interaction. One of the biggest problem of standard reasoning engine is their scalability and performance with respect to large data sets. Therefore, to allow instant reasoning as well as real-time queries during exploration and interaction we utilize besides RacerPro the very fast U2R2 reasoning engine which performs inference within its secondary storage and allows for advanced features such as retraction of concept and property definitions and incremental reasoning [18].

#### 5. CONCLUSION

In this paper we presented a gainful combination of established methods from visual analytics and semantic technologies. The exploration examples throughout the paper should give an idea how this will help to gain further insights into large and heavily interlinked assertional data sets. Exploration direction, level of detail as well as global filter conditions are determined by the user. At the same time the underlying reasoning engine ensures the semantically correct data by representing told but also entailed facts. Furthermore, the interlocking of semantic and visual technologies adds new exploration options such as filler set abstractions, local filtering with help of schema axioms, and the clarification of quantities not found in current tools. All these features have been implemented and integrated into our ONTOTRACK ontology authoring environment.

#### Acknowledgments.

This work has been partly supported under the ReSIST Network of Excellence, which is sponsored by the Information Society Technology (IST) priority in the EU Sixth Framework Programme (FP6) under contract number IST 4 026764 NOE.

#### 6. REFERENCES

- [1] T. Anderson, Z. Andrews, J. Fitzgerald, B. Randell, H. Glaser, and I. Millard. The ReSIST Resilience Knowledge Base. In *Proc. of the 37th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2007)*, June 2007.
- [2] S. Auer, C. Bizer, J. Lehmann, G. Kobilarov, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In *Proc. of the 6th International Semantic Web Conference (ISWC 2007)*, volume 4805 of *LNCS*, pages 722–735. Springer, 2007.
- [3] Ben Bederson, Jesse Grosjean, and Jon Meyer. Toolkit Design for Interactive Structured Graphics. Technical Report CS-TR-4432, University of Maryland, January 2002.
- [4] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, May 2000.
- [5] Yuanbo Guo, Zhengxiang Pan, and Jeff Heflin. LUBM: a benchmark for OWL knowledge base systems. *Journal of Web Semantics*, 3(2):158–128, July 2005.
- [6] Volker Haarslev and Ralf 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.
- [7] Matthew Horridge, Sean Bechhofer, and Olaf Noppens. Igniting the OWL 1.1 Touch Paper: The OWL API. In *Proc. of the 3rd OWL Experiences and*

*Directions Workshop (OWLED'07) at the ESWC'07*, Innsbruck, Austria, 2007.

- [8] Ian Horrocks. Applications of Description Logics: State of the Art and Research Challenges. In *Proc. of the 13th Int. Conf. on Conceptual Structures (ICCS'05)*, 2005.
- [9] Afraz Jaffri, Hugh Glaser, and Ian Millard. URI Identity Management for Semantic Web Data Integration and Linkage. In *Proc. of the 3rd International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2007)*, Vilamoura, Algarve, Portugal, 2007. Springer.
- [10] Daniel A. Keim. Visual exploration of large data sets. *Communications of the ACM*, 44(8):38–44, 2001.
- [11] Thorsten Liebig and Olaf Noppens. ONTOTRACK: A semantic approach for ontology authoring. *Journal of Web Semantics*, 3(2):116 – 131, 2005.
- [12] Wolfgang May. Information extraction and integration with FLORID: The MONDIAL case study. Technical Report 131, Universität Freiburg, Institut für Informatik, 1999.
- [13] Olaf Noppens and Thorsten 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.
- [14] Peter F. Patel-Schneider and Ian Horrocks. OWL 1.1 Web Ontology Language Overview, December 2006.
- [15] Catherine Plaisant, Jesse Grosjean, and Benjamin B. Bederson. SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. In *Proc. of the IEEE Symposium on Information Visualization (INFOVIS 2002)*, pages 57 – 64, Boston, USA, October 2002.
- [16] Ben Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. of the IEEE Symposium on Visual Languages*, pages 336–343, Washington, USA, 1996.
- [17] James J. Thomas and Kristin A. Cook. *Illuminating the Path*. IEEE Computer Society, 2005.
- [18] Timo Weithöner. U2R2 – The Ulm University Relational Reasoner: System Description. In *Proc. of the 14th Int. Conf. on Logic Programming Artificial Intelligence and Reasoning (LPAR'07)*, Yerevan, Armenia, October 2007.