Eurographics Conference on Visualization (EuroVis) 2017
B. Kozlíková, T. Schreck, and T. Wischgoll (Editors)

Short Paper

The Situation Universe: Visualizing the Semantics of Integrated Data Structures

B. Karer, D. Fernández-Prieto, and H. Hagen

Computer Graphics and HCI Group, TU Kaiserslautern, Germany

Abstract

The efficient extraction and communication of information in heterogeneous data domains is a major challenge in the context of the ongoing digitalization efforts in industry and in the public sector. The heterogeneity of the data itself and the diverse interests of the users addressing it demand the integration of structural and semantic information about data aggregated from multiple sources into a single model and unified visualization. In this paper, we present an approach to visualize the possible interpretations of data integrated from heterogeneous environments, including the sequences of operations applied to filter, transform, and reinterpret the data, such that the result supports these interpretations. Users can thereby access and explore integrated data from the perspective specific to their respective fields of experience.

Categories and Subject Descriptors (according to ACM CCS): I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods—Semantic Networks; I.3.m [Computer Graphics]: Miscellaneous—

1. Introduction

Recent developments in the field of the Internet of Things and the ongoing digitalization efforts in industry nourish a strong and still increasing interest in data analytics. Decision support for strategic planning and process optimization are major drivers for data analytics in industry as well as in the public sector. Complex applications require the aggregation and integration of data from heterogeneous environments, where context- or vendor-specific naming conventions limit the compatibility of data from different sources or domains. For data bases, this problem is solved by semantic integration, matching data items by their meaning rather than their labels. However, data integration alone is insufficient for proper analytics, since domain-specific interpretations cannot simply be overridden by a superordinate naming convention. Different users will still interpret the same data differently. As an example, consider an industry where a sales manager and a production manager monitor the percentage of defective parts being produced by a milling machine. The sales manager applies it as a correction factor for the time to produce a given number of parts, whereas the production manager uses the exact same fraction as a performance indicator for the scheduling of maintenance intervals. While, formally, the percentage of defective parts is an error rate and therefore likely to be labeled as such after (semantic) integration, the managers' different points of view require different interpretations of this data to infer the desired information.

Current models for data analytics commonly focus on either the structural composition or the semantics of data. Consequently, visualization techniques also focus on either the structure or the semantics. However, the demand for context-specific data representations in integrated environments motivates an integral visualization of the data's structure and the different applicable interpretations.

For the visualization of the semantics of integrated data structures, we identify the following requirements:

[R1] Explicit encoding of the domain-specific semantics.
[R2] Support for context-sensitive data interpretation and dynamic binding of actual data sources.
[R3] Concise and clear presentation of structure and semantics.
[R4] Concise depiction of data provenance revealing the transformation paths from raw data to interpreted information.

Towards satisfying these requirements we propose to combine context-sensitive semantics with a concise yet expressive graph-based visualization. Our contribution is twofold:

1. We introduce situation semantics and situation theory to the field of visualization as a model to define semantics and context.
2. We propose a visualization for integrated data in heterogeneous environments that – enriched with elements for operations performed on the data – concisely conveys the applicable semantics.

The remainder of this paper is structured as follows: In the next section, we introduce relevant aspects of situation theory and explain how we model possible interpretations that may apply after binding the abstract structure to actual data. We then turn to the description of our visualization technique and demonstrate its application in an example. From the discussion of our method's performance in satisfying the above requirements, we derive directions

DOI: 10.2312/eurovisshort.20171153

for the next steps we are planning to take in the course of this research. We conclude the paper with a summary of the key features of the proposed visualization approach.

2. Model

2.1. Context-Sensitive Semantics

Semantic data models often organize data in a labeled graph which can be interpreted in a way that allows the inference of certain information about the data. A popular example for such a model is the semantic web [BLHL01]. Yet, the inference rules are usually not explicitly encoded in the network. The inherent ambiguity resulting from this fact can be alleviated by applying an ontology providing the correct interpretation of data items based on syntactic rules. This property enables the integration of data from heterogeneous sources with potentially differing naming conventions [Gar05]. However, such a procedure requires interpretation and transformation to map the data to a common ontology. In setups where data instances and semantics can change independently, this introduces synchronization problems. For example, it requires an updated ontology to contain legacy elements to remain consistent with the data. It would thus be more convenient if the data itself could directly provide the correct interpretation. We model these interpretation rules explicitly, using situation semantics as introduced by Barwise and Perry in 1983 [BP83]. In particular, our technique employs the following features of this theory:

1. *inherent context-sensitivity*: the **situation** is a collection of facts, the **infons**, explicitly stating that within the situation an expression does or does not hold. The notion of holding with respect to a situation is called the **polarity**. If an expression e has polarity 1 with respect to situation s, we say s **supports** e.
2. *partiality*: the situation is never completely known. Therefore, its description is always considered incomplete but extendable.
3. *dynamic interpretation*: so-called **types** model objects and situations by describing their properties. We identify types describing objects directly with data entities. Since this renders objects into collections of properties, in our work, every type is also a situation. Types are linked by **constraints** that model general relations between them. Constraints are satisfied if the situation supports the incident type. Intuitively, if the situation is the same on both sides of a constraint, it is an inference of polarities. In our application, this is generally the case. **Conditional constraints** may only take effect if the situation supports the additional information defined in the condition.
4. Additional information introduced into the situation may change the set of applicable constraints and therefore change the possible semantics. The same holds for the introduction or removal of new data instances determining infon polarities.

For a more detailed discussion on situation semantics, the reader is kindly referred to Keith Devlin's review [Dev06] and for a more elaborated discussion on situation theory, the reader may be interested in John Barwise's original introduction of the idea [Bar86]. Rather than modeling the situation directly, we consider the types as free variables and only model the constraints between them. The resulting network of constraints is applicable to arbitrary situations by binding (subsets of) the available free variables allocating the infon polarities. By the conditions, the infon polarities determine the applicability of interpretations. To also consider derived data that is not directly available but has to be computed from other data, we extend the graph by the syntax of the data's structural composition and the operations that can be performed on it. Like the applicability of constraints, the results of operations performed on data also depend on individual data instances. Thereby, we model a universe of possible situations according to the data's structural composition and the known semantics. The actual situation is then captured by evaluating the network of transitions (syntax) and constraints (semantics) based on the actual data instances. It is determined by exactly the subset of semantic concepts that is reachable from the data concepts for which at least a single data item exists or can be computed and the constraint and interpretation paths leading there.

2.2. Information Provenance

To allow a user to make informed decisions based on interpreted and transformed information rather than raw data, it is important to convey the history of operations performed on the data. Incorporating information provenance into the model and visualization reveals the data sources involved in the process of deriving new information. After all, the derived information can only be as trustworthy as the data it is obtained from. One approach of providing this information provenance is to specify proof traces in a formal proof modeling language [dSDMM03], [dSMF06]. While this allows the computer to estimate the trustworthiness of derived information based on a set of inference rules, it is comparably hard to read for a human user. Visual representations of data transformation paths, on the other hand, typically focus on certain aspects of the transformation pipeline rather than on the change of the data or its meaning. This different focus often necessitates the inclusion of additional detail information. While such decisions are perfectly sound for their respective applications, our focus on the possible changes of data syntax and semantics usually does not require this level of detail which is why we favor a more simple representation. We abstract all possible operations into three types:

1. *Filters* are set operations that neither change the data nor the interpretation.
2. *Transformations* directly change the data. The output data might have different semantics than the input.
3. *Interpretations* assign new semantics to the data. These new semantics can be associated with a different concept not reachable under the prior interpretation. An interpretation's applicability can depend on additional conditions.

This high level abstraction provides an overview over the sequences of operations applied to data until it supports certain information. Information provenance can therefore be assessed by following these paths backwards until raw data is reached.

3. Visualization

To visualize the possible interpretations of integrated data, we propose a graphical annotation for the graph structure we developed in Section 2.1. Note that without the operations, the graph reduces to an interlinked web of data, similar to the internet. This resemblance renders semantic web technologies natural candidates for our visualization. Recent studies have shown that VOWL, a graphical annotation for the Web Ontology Language [NL13, NHL13, LNHE16], can considerably improve the communication between domain and ontology experts if it is integrated into an interactive framework [NHL13, LNHE14]. To benefit from these findings, we visualize the graph as a node-link diagram based on VOWL2 [NLH14] and the WebVOWL platform [LLMN15].

Although this notation is directly applicable to the interlinked concepts, we need to extend it to include the operations and the distinction between data transformation syntax and the applicable semantics. Extensions of VOWL notation have been proposed to cover dynamics or to be applied to text visualization [DLCW15, BL15]. In contrast to these extensions, our focus is the integration of different semantics from multiple domains and their applicability with respect to the context defined by the data.

We adopt the basic graphical representation of data attributes and relations from VOWL2. The nodes symbolizing the classes in VOWL2 denote the concepts in our model. To encode the operations, we introduce special nodes symbolizing *filters*, *transformations*, and *interpretations*. We apply colors to clearly distinguish between the nodes and edges that are associated with semantic information and those that represent or act on data. The annotation is explained in Figures 1 and 2. In the resulting graph of concepts and operations, the subgraph induced by the blue nodes and edges denotes the syntax of possible data transformations, its orange counterpart defines a conceptual model on the semantic level. Wherever the same concept is connected to multiple interpretation nodes, different domains assign different semantics to the same data. The user decides which of the possible interpretations is to be applied.

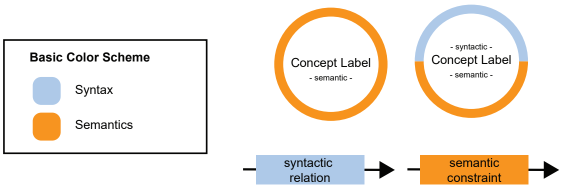

Figure 1: *Visual representation of concepts and relations. A basic color scheme distinguishes syntax and semantics. Circles with an orange border indicate concepts that only contain semantic information and are not directly identified with data. Circles with half-blue and half-orange border indicate concepts that directly relate data to semantics. The same color scheme applies to edges.*

3.1. Application Example

We demonstrate the usage of our model and graphical annotation by the example introduced in Section 1. The resulting graph is shown in Figure 3. A milling machine processes raw material (e.g. stainless steel) to produce metal parts. We consider the scenario of a sales manager and a production manager, two experts from different domains, both of which are interested in the amount of defective parts being produced.

The sales manager needs to determine how many parts can be offered to customers based on the actual production rate. Relying on the syntactic information this expert can only learn that a production rate can be computed by using data associated to the milling machine. In the sales management domain, the relevant factor is named the "reject rate". Even though what is called the "error rate" in the integrated data is directly linked to the milling machine, associating it with the reject rate requires extensive detail knowledge of the data structure. In the graph obtained using our model, this information is carried by the interpretation operation linking the error rate to the reject rate. Following the constraints, a less experienced user will be able to infer that the production rate of a milling machine is determined by the reject rate, which in the data structure is referred to as the error rate, the percentage of defective parts.

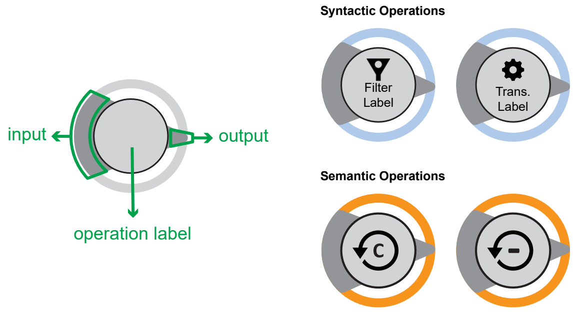

Figure 2: *Visual representation of operations. A shaded circle segment indicates the direction of the data flow from input to output. Since they operate on the syntax, filters (top-left) and transformations (top-right) are depicted with a blue border. The orange border of conditional and unconditional interpretations (bottom-left and bottom right) indicates their influence on the semantics.*

Similarly, the production manager is interested in scheduling maintenance intervals to prevent the risk of critical failures of the milling machine. From the semantic information, the expert learns that the input for this estimation is the same error rate that the sales manager applies to compute the production rate.

4. Discussion and Future Work

The proposed extension of VOWL2 and its integration with situation theory satisfies the requirements defined in Section 1 well:

[R1] Domain-specific semantics are encoded by explicitly associating data with its several possible interpretations (visualized as orange concepts and relations).

[R2] With situation semantics and situation theory, we apply an inherently context-sensitive model for data semantics. Dynamic binding of data sources is possible, but requires a reevaluation of the applicable semantics.

[R3] Structural and semantic information are clearly distinguishable by applying the proposed color scheme.

[R4] Information provenance is easily traceable following the paths marked as syntactic transitions, i.e. the sequences of applied data transformations and relations between data elements. Detail information about operations can be accessed and edited on demand (see Figure 3).

The ongoing integration of our approach into an interactive visualization system based on WebVOWL will further enhance compliance with the requirements. For example, automatic extraction and

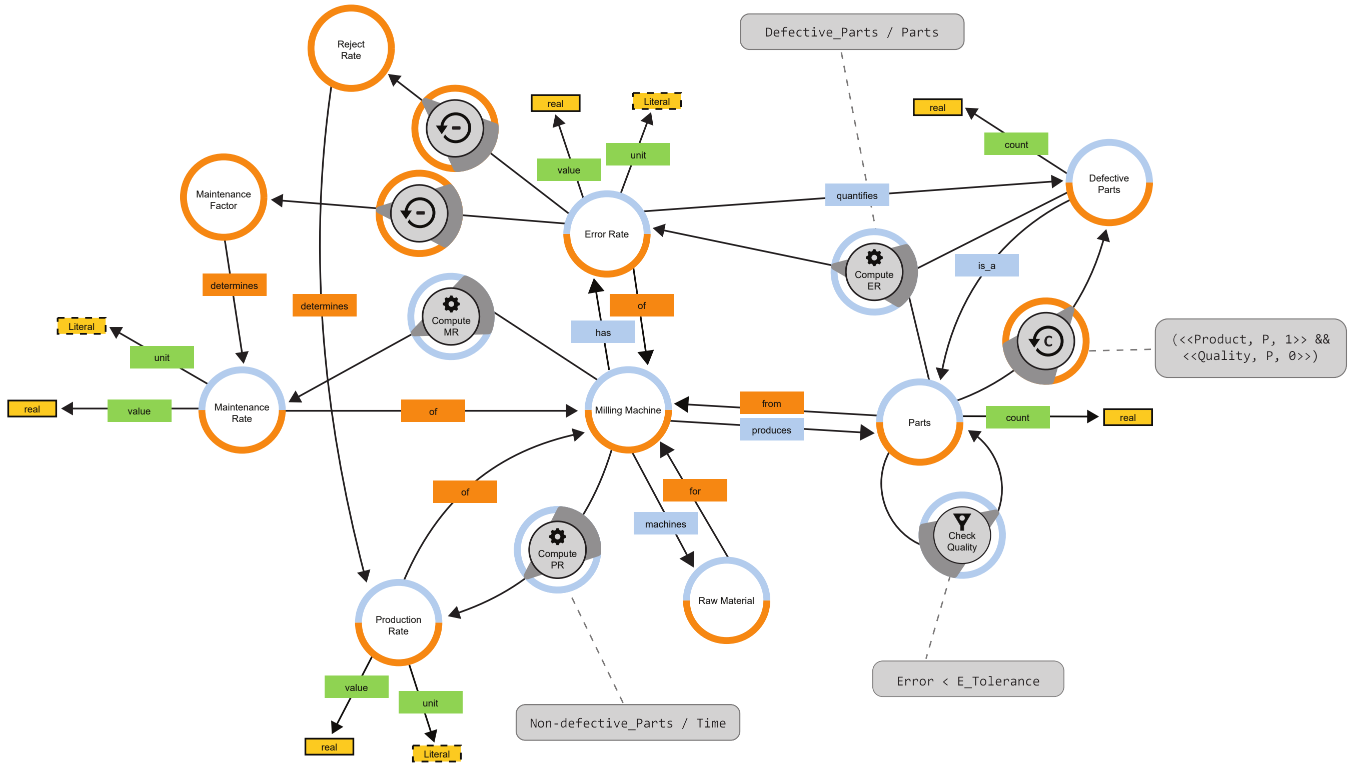

Figure 3: *Graph of the milling machine example. Transformation operations are applied to infer the production rate and maintenance rate of the milling machine. A filter identifies the produced parts that do not comply to an error tolerance. A conditional interpretation encodes the recognition of these parts as defective. Two unconditional interpretations link the error rate to the domain-specific semantics. Following the graph's edges, the user can infer how the concepts are related on the semantic level and how operations translate between data with different associated semantics.*

highlighting of applicable interpretations can reduce the graph's complexity. This could be combined with user and task profiles to highlight concepts and interpretations by domain-specific relevance. Likewise, the interactive extension and modification of structure, semantics, and operations can enrich the analysis experience by enabling the direct incorporation of newly obtained insight. Studies with experts from different application domains will evaluate the usefulness and applicability of our approach and identify directions for further improvement.

5. Conclusions

In this paper, we propose an approach to the visualization of the semantics of integrated data structures. We apply situation semantics to model all possible interpretations of the data as a universe of situations. The actual data instances determine the subset of applicable interpretations and thereby the concrete situation observed. Being inherently context-sensitive, the semantics and therefore the visualization can be applied to dynamically changing data by rebinding the abstract model to different data sources. A graph-based visualization captures the applicable semantics and reveals the sequences of operations applied to filter, transform, and interpret the available raw data into a form supporting these semantics. Encoding data semantics explicitly rather than deriving it from external knowledge, our model is essentially domain independent. Consequently, the proposed model and visualization have a diverse range of applications, including highly complex fields such as business analytics, factory planning, predictive maintenance, and generally every domain where an integrated view on a heterogeneous data environment meets diverse interests of users addressing the data. All these areas benefit significantly from the integrated visualizations of semantic and structural information. Therefore, we consider this work as a first step towards a visualization of context-dependent meaning in heterogeneous data environments.

Acknowledgment

This work was supported by the DFG (German Research Foundation), as a project of the International Research Training Group (IRTG) 2057.

References

[Bar86] BARWISE J.: The situation in logic. In *Logic, Methodology and Philosophy of Science VIIProceedings of the Seventh International*

Congress of Logic, Methodology and Philosophy of Science, Ruth Barcan Marcus G. J. D., Weingartner P., (Eds.), vol. 114 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 1986, pp. 183 – 203. 2

[BL15] BURCH M., LOHMANN S.: Visualizing the evolution of ontologies: A dynamic graph perspective. *CEUR Workshop Proceedings 1456* (2015), 69–76. 3

[BLHL01] BERNERS-LEE T., HENDLER J., LASSILA O.: The semantic web. *Scientific American 284*, 5 (May 2001), 34–43. 2

[BP83] BARWISE J., PERRY J.: *Situations and Attitudes*. MIT Press, Cambridge, MA, 1983. 2

[Dev06] DEVLIN K.: Situation theory and situation semantics. In *Handbook of the History of Logic*, Woods J., Gabbay D. M., (Eds.), vol. 7. 2006, pp. 601–664. 2

[DLCW15] DASIOPOULOU S., LOHMANN S., CODINA J., WANNER L.: Representing and Visualizing Text as Ontologies: A Case from the Patent Domain. In *Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data* (2015), pp. 83–90. 3

[dSDMM03] DA SILVA P. P., DEBORAH S., MCGUINNESS D. L., MCCOOL R.: Knowledge provenance infrastructure, 2003. 2

[dSMF06] DA SILVA P. P., MCGUINNESS D. L., FIKES R.: A proof markup language for semantic web services. *Inf. Syst. 31*, 4-5 (June 2006), 381–395. 2

[Gar05] GARDNER S. P.: Ontologies and semantic data integration. *Drug Discovery Today 10*, 14 (2005), 1001 – 1007. 2

[LLMN15] LOHMANN S., LINK V., MARBACH E., NEGRU S.: WebVOWL: Web-Based Visualization of Ontologies. *Lecture Notes in Computer Science 8982* (2015), 225–232. 3

[LNHE14] LOHMANN S., NEGRU S., HAAG F., ERTL T.: VOWL 2: User-Oriented Visualization of Ontologies. *Knowledge Engineering and Knowledge Management - 19th International Conference, {EKAW} 2014, Linköping, Sweden, November 24-28, 2014. Proceedings 8876* (2014), 266–281. 3

[LNHE16] LOHMANN S., NEGRU S., HAAG F., ERTL T.: Visualizing ontologies with VOWL. *Semantic Web 7*, 4 (2016), 399–419. 2

[NHL13] NEGRU S., HAAG F., LOHMANN S.: Towards a unified visual notation for OWL ontologies. *Proceedings of the 9th International Conference on Semantic Systems - I-SEMANTICS '13* (2013), 73. 2, 3

[NL13] NEGRU S., LOHMANN S.: A Visual Notation for the Integrated Representation of OWL Ontologies. *Vis.Uni-Stuttgart.De* (2013), 308–315. 2

[NLH14] NEGRU S., LOHMANN S., HAAG F.: VOWL: Visual Notation for Ontologies (Specification of Version 2.0), 2014. 3