Semantic technologies and metadata systematisation for evaluating time series in the context of driving experiments

Ulf Noyer, Dirk Beckmann, and Frank Köster
German Aerospace Center, Institute of Transportation Systems
{ulf.noyer|dirk.beckmann|frank.koester}@dlr.de

1 Motivation and background

The design of assistance and automation systems in the automotive domain often has a strong focus on human machine interaction. Therefore test vehicles and simulators play a major role within state-of-the-art development processes. In particular, during assessment/evaluation of developed assistance and automation systems these facilities are used to record extensive data sets (e.g. multivariate time series as well as video and audio data), which describe driver behavior, man-machine interactions, dynamics of real or simulated vehicles, etc. These data has to be explored and analyzed to support further design activities. [1]

At the Institute of Transportation Systems of the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt – DLR) e.g. multivariate time series are stored in a relational database. Metadata is used to link additional information to recorded data (e.g. experimental setup, sensor configuration and types of sensors). During the analysis of this data, significant conclusions can arise, that should be associated with measured data or already associated metadata. This newly created information is also expressed as metadata. The main intention is to provide all stakeholders in a development project with a comprehensive view on measured, transformed, and analyzed data as well as to support the flexible integration of further insights.

2 Architecture and approach

Our approach is founded on techniques from scientific workflow systems and semantic web [2]. All metadata is modeled using RDF-triples and stored by using the Jena Semantic Framework. To use time series elements from the relational database in RDF-triples and annotate them with metadata, a mapping from database elements into URIs is proposed (cf. [3]). The mapping is a bijection, which transforms database table, tuples, attributes and attribute instances into unique URIs. It is not generally necessary to transform time series data itself into RDF. Even though a realization in RDF would be feasible in a similar process [4].

Figure 1 shows an example, where database elements are annotated with metadata[1]. The attributes are associated with sensors (*direction indicator* etc.). Tuples are annotated, if certain actions (*indicator left* or *turn left*) or environmental events (*vehicle ahead lost*) were detected. Annotations can be generated manually (e.g. by video rating) or automatically (e.g. by algorithms or data mining). They can serve as input in advanced analysis steps to derive an even higher abstraction level. On the basis of a comprehensive ontology a reasoner can be also used for this purpose. In Fig. 1 the maneuver *overtaking* is estimated based on *indicator left* and *turn left*.

Data analysis activities are supported by using ontologies through a strict systematization of concepts. That helps a team to establish a common understanding of concepts and relations

[1] there are no predicates of annotations illustrated for a better overview

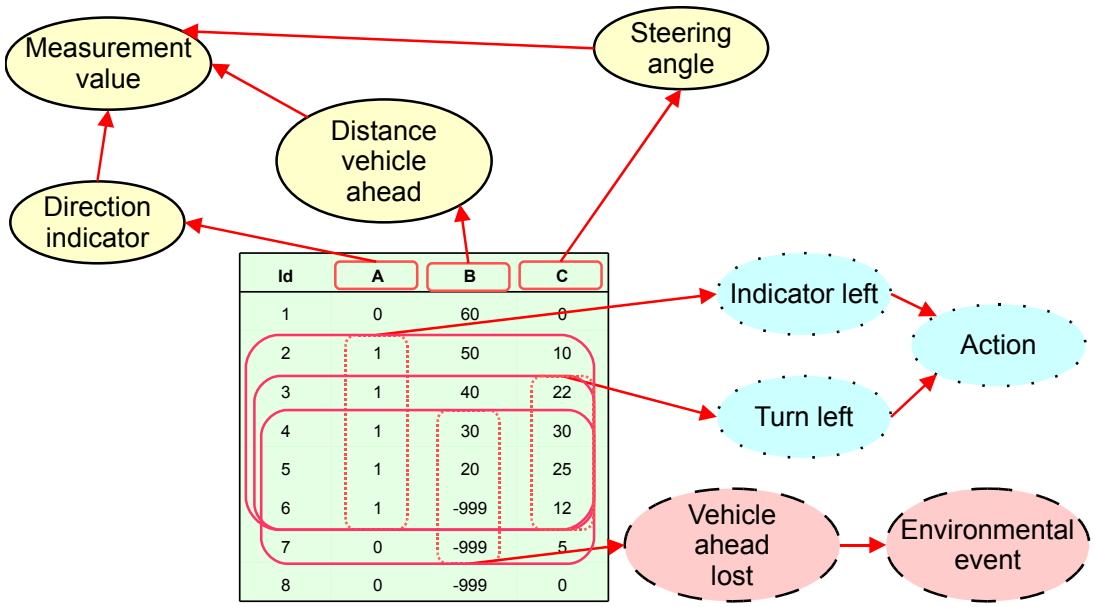

Figure 1: Annotations on database elements

within the domain. Furthermore, ontologies are reliable in aspects of flexibility and extensibility — this supports dynamic processes during the analysis of data. This is an important foundation for the application of automated analysis techniques in terms of data mining.

Another well suited task for metadata annotations is the support of data quality management. Annotations can be used to mark table attributes with sensor configurations to provide information about the expected properties of recorded data. Interesting configuration properties for quality management can be valid co-domain, sample rate, monotony, permitted discrepancies between values, et cetera. If quality checks now identify suspicious tuples, they can be connected with the violated configuration annotations.

3 Summary

In contrast to traditional relational data modeling, the presented method is characterized by its flexibility and extensibility of the metadata for new domain specific concepts. The design exploits the power and flexibility of ontologies and maintains good manageability of relational databases. So the team can profit from a common and gradually evolving ontology. Our research activities and developed tools are currently evaluated in a research project, which focuses on methodologies for the development of safety-critical embedded systems (Integrated Modeling of Safe Transportation – IMoST).

References

[1] F. Köster and U. Noyer, "Leistungsfähiges Datenmanagement als Rückgrat einer menschzentrierten Entwicklung von Automation / Assistenz," in *24. VDI/VW-Gemeinschaftstagung — Integrierte Sicherheit und Fahrerassistenzsysteme*, 2008.

[2] M. Gertz and K.-U. Sattler, "Integrating scientific data through external, concept-based annotations," in *In Second International Workshop Data Integration over the Web*, 2002.

[3] T. Berners-Lee, "Relational Databases on the Semantic Web," 1998.

[4] C. Bizer, "D2R MAP — A Database to RDF Mapping Language," in *WWW2003*, 2003.