DEPLOYMENT OF ONTOLOGIES IN BUSINESS INTELLIGENCE SYSTEMS

Carsten Felden
Technical University Bergakademie Freiberg, Lessingstraße 45,09596 Freiberg,Germany

Daniel Kilimann
Mercator School of Management, University Duisburg-Essen,Campus Duisburg,47057 Duisburg,Germany

Keywords: Management Information Systems, Ontologies, Ontological Engineering.

Abstract: The consideration of integrated structured and unstructured data in management information systems requires a new kind of metadata management. Ontologies constitute a possibility to solve the resulting problems. Process models describe the development of ontologies which can be utilised in the context of management information systems, are discussed.

1 INTRODUCTION

The penetration of companies with information systems constitutes the basis of information supply for decision makers by means of structured and unstructured data integrated from internal and external sources. More and more heterogeneous data are available and this leads to the often quoted information flooding. Furthermore, heterogeneous data formats lead to the problem that related data cannot be found and shown to the user. Semantic annotation of natural language documents and the integration of domain ontologies can enable semantic inquiries.

This paper presents an idea of melting data dictionary data with ontologies in favour of queries in an integrated database. First of all, we introduce business intelligence (BI) systems which are using such an integrated database in order to gain an impression of available data for the reader. Chapter three addresses the issue of ontologies and their possible application fields. A restraint on the ontology usage is their development as well as their maintenance. Therefore chapter four discusses current approaches of ontology development. Chapter five recapitulates the results and gives an outlook.

2 BI SYSTEMS USING STRUCTURED AND UNSTRUCTURED DATA

Up to 90 percent of the information in a company are not available in a machine processable format, e.g. structured data, but as unstructured, non-machine processable, data. These kinds of data are generally natural language documents (Kantardzic, 2003). Due to this reason, there is a considerable potential which can be managed by an adequate handling of information flooding. Data in BI-systems derive from different heterogeneous sources. They have to be differentiated between intra-corporate and external sources. The first ones are operational application systems. The operational system environment is heterogeneous, because these systems were normally developed isolated from each other. Operational applications are using different data structures and formats. External data can be for instance purchased data streams from news services like *Reuters* or the result of queries sent to search engines like *Google*. Appropriate systems are required in order to retrieve all stored information of the central database.

The integration process of internal and external data is comparable. During the transformation, data are syntactically and semantically adjusted. The

Felden C. and Kilimann D. (2006).
DEPLOYMENT OF ONTOLOGIES IN BUSINESS INTELLIGENCE SYSTEMS.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - DISI*, pages 306-309
DOI: 10.5220/0002453603060309

syntactic data transformation is the necessary conversion of formats into a uniform standard. Semantic transformation deals with clearing up textual senseless field contents, the decomposition of semantic overburdened fields and the elimination of synonyms and homonyms. The results of the transformation process are data structures which correspond to the design of the database included in the Business Intelligence system.

3 ONTOLOGIES IN BI SYSTEMS

In the scope of computer science an *ontology* is formally a defined system of concepts. This paper is based on the definition of Studer: "An ontology is a formal, explicit specification of a shared conceptualisation." (Studer et al., 1998)

Conceptualisation corresponds to an abstract model of a domain which identifies the relevant concepts and their relationships. *Explicit* means that the used concepts are unique and their usage is formally confined. *Formal* refers to the fact that ontologies should be machine-readable. *Shared* indicates that an ontology is accepted by a group of people and used corporately.

Important components of a data warehouse are metadata. They link the operational information systems and the data warehouse of a BI system. Metadata are located in a directory which enables analysts to discover data in a meta database system. This directory is called data dictionary or repository (Froeschl, 1997). Metadata consist of all information which simplify development, maintenance and administration of a data warehouse system as well as enable the acquisition of information for the data warehouse (Bauer and Günzel, 2001). They explain the transformations during the data integration process. Furthermore, they characterize the algorithms operating in the data warehouse so that the result is the linkage between the aggregation processes and the subject orientation of the entire database (Inmon and Hackathorn, 1994).

More precisely, a central requirement of a data dictionary is the documentation of the data fields and database structures including data origin, data validation, data definition, possible influencing variables, details about the acquisition of data, and links to other information (Wertz, 1986). It can be stated that a data dictionary concentrates on structured data. But, as mentioned above, a BI system also covers unstructured data. Natural language documents are characterized by a confusing variety of terms. According to this situation, metadata have to consider synonyms and multilingual terms. Ontologies offer corresponding assistance in this context. Their major requirement is to make such information machine-processable and to simplify accessing data.

The data dictionary provides an appropriate basis to construct an ontology, because it ensures the unambiguousness of the used terms within the database and contains the necessary metadata. Because of this it is a suitable foundation to identify appropriate concepts und their relationships. But, the modelling process has to be executed manually. Models are especially important in order to recognise and eliminate possible restraints during the development of the ontology and to simplify the maintenance procedure.

4 COMMON ONTOLOGY DEVELOPMENT PROCEDURE MODELS

Although ontology development is comparable with software development life cycles, special requirements of ontologies have to be kept in mind. In the recent years, numerous suggestions were made how to develop an ontology (Staab et al., 2001).

4.1 Ontology Development Approaches

The METHONTOLOGY approach was published by Fernandez-Lopez, Gomez-Perez and Juristo in 1997 (Fernandez-Lopez et al., 1999). METHONTOLOGY is a comprehensive ontology development methodology according to the IEEE-norm in the fields of software development and knowledge management. The activities of the ontology development process are divided into three categories: *project management activities, development activities* and *supporting activities.* D*evelopment activities* describe the procedure of ontology construction in detail. *Project management activities* include planning, control, and quality assurance. Both have to be distinguished from the accompanying *supporting activities*. These activities are divided into knowledge acquisition, integration, evaluation, documentation, and configuration management (Corcho et al., 2003).

The *On-To-Knowledge* project is concentrated on the design of an ontology based knowledge management system. The On-To-Knowledge procedure model consists of the following phases: feasibility study, kickoff-phase, refinement, evaluation, and maintenance (Sure, 2002). The feasibility study identifies chances and risks and analyses the primary application areas. The results of this phase are the basis for the kickoff-phase. The created application specification contains the domain, the objective, design directives, available resources, and potential users. Subsequent competence questions are formulated in order to collect domain specific terms in an informal manner.

The main focus of the *TOVE (Toronto Virtual Enterprise)* methodology, created by Grüninger and Fox, is to provide a series of competence questions (Grüninger and Fox, 1995). Questions on the problems that have to be solved are formulated and should be answered afterwards by the ontology. They are used in order to build the concept hierarchy and to evaluate the ontology.

The *SENSUS* approach was introduced by Swartout (Swartout et al., 1996). The initial point is the SENSUS ontology itself. It represents an extensive ontology including 70,000 domain independent concepts. Representative concepts of this domain are selected and manually linked with the SENSUS ontology. Afterwards, all concepts are inserted which are located directly at the path from the specific terms to the root. Further concepts are included manually. The remaining SENSUS concepts are discarded as irrelevant.

The *KACTUS* approach was developed with the scope of the Esprit-project. It postulates already existing ontologies which are reused or customized in order to create a new one (HCS, 1996). First of all the applications, thus the relevant concepts and objectives, are specified. A new ontology is developed by adjusting and refining the already existing top-level or reusable ontologies.

4.2 Critical Review

Following the framework of Gómez-Pérez, Fernández-López, and Corcho, two different kinds of criteria are used to evaluate the shown approaches (Gómez-Pérez et al., 2004). The first criteria type follows the IEEE-standard using the same criteria as in the field of software development (Fernández-López et al., 1999). According to the IEEE-norm 1074-1997, criteria can be classified into three categories: ontology management activities, development oriented activities, and accompanying activities (IEEE, 1998). The ontology management activities constitute the tasks and functions of the project management within the development process. The development oriented activities splits into predevelopment, development, and postdevelopment. The accompanying activities as mentioned above support the development process and are executed parallel to it.

The activities described in the Grüninger and Fox methodology concern pre- and postdevelopment as well as ontology and configuration management. There are fewer activities described in the KACTUS approach. Furthermore the accompanying activities are missing. METHONTOLOGY refers to almost every activity, but the descriptions differ in their level of detail. These which are part of the predevelopment and the implementation are missing. The On-To-Knowledge methodology covers the entire spectrum of suggested activities including the ones for the predevelopment process. Finally, SENSUS does not include the phases conceptualisation and formalisation.

Further criteria are life cycle, application dependence, and the usage of a core ontology. There are two options an incremental life cycle or evolutionary prototyping. There are three different specifications concerning the next criterion: application dependent, application independent and semi-application dependent. Reusability of existing ontologies enables an efficient handling of available knowledge. The usage of a data dictionary as a basis of ontology development fits in this context.

The approach of Grüninger and Fox as well as On-To-Knowledge supports both possible life cycles. In contrast, concerning KACTUS and METHONTOLOGY evolutionary prototypes are recommended. The SENSUS approach provides no life cycle at all.

The Grüninger and Fox methodology just as the SENSUS approach are characterized as semi-application dependent. The KACTUS and On-to-Knowledge approaches are application dependent by definition. In contrast to this, METHONTOLOGY is application independent.

Finally reuse of existing ontologies is discussed. Reusability is not part of the Grüninger and Fox methodology. According to the KACTUS approach new ontologies are developed by reusing or adjusting existing ones. Analogue to this, the SENSUS ontology is used as a basis for constructing the designated ontology. A reusable ontology is not mentioned explicitly in METHONTOLOGY as well as in On-To-Knowledge, but the idea is taken into account in both approaches.

4.3 Application of Ontologies in BI Systems

The following figure shows the assignment between a data-dictionary entry and an ontology entry.

```
Data-Dictionary-Entry:
Dimension Product   =    {product entry}
product entry =    Product_ID + Product_Name +
Type + (Hour) + (Day) + (Week) + (Month) +
(Quarter) + Start_Year + End_Year
-----------------------------------------------
Ontology-Entry:
< daml: ObjectProperty rdf: ID = “traded at” >
   < rdfs: domain rdf: resource = “{product entry}” >
   < rdfs: range rdf: resource = ”APX”/ >
< /daml: ObjectProperty >
< daml: Class rdf: ID = “Amsterdam Power Exchange” >
   < daml: sameClassAs rdf: resource = “’#APX”/ >
</ daml: Class >
```

Figure 1: Data dictionary with connected ontology.

The data dictionary entry in the superior part of the figure describes the product dimension of a database as part of a BI system. The lower part shows an ontology entry. The line < rdfs: domain rdf: resource = “{product entry}” > references to the respective data dictionary entry. This means an enhancement of the ontology model and creates a connection between the technical and semantic product description. Users can benefit from an integrated view on structured and unstructured data based on the above described connection.

5 CONCLUSION

BI systems provide their users access to structured and unstructured data. The problem of an integrated metadata management is not solved, yet. Ontologies are a presently discussed proposal. An existing data dictionary administrating structured data should be enriched with functionalities of an ontology in order to be able to handle unstructured data as well. The development of an ontology based on an existing data dictionary requires a large manual effort. Due to this reason, common ontology development procedure models are discussed in this paper. A terminal decision is not yet possible, because models are based on different assumptions and aims. In addition, new approaches have to be recommended. Furthermore, it has to be clarified in the future, if semi-automatic methods can be integrated into a standardized ontology development process.

REFERENCES

Bauer, A. and Günzel, H., 2001. *Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung*, dpunkt.verlag. Heidelberg.

Corcho, Ó.; Fernández-López, M. and Gómez-Pérez, A., 2003. Methodologies, Tools and Languages for Building Ontologies. Where is their meeting point? In *IEEE Transactions on Data and Knowledge Engineering 46 (2003) 1*, p. 41-64. http://portal.acm.org/citation.cfm? id=864179.2003, accessed 2005-08-20.

Fernández-López, M., Gómez-Pérez, A., Pazos-Sierra, A. and Pazos-Sierra, J., 1999. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. In: *IEEE Intelligent System 14 (1999) 1*, p. 37-46.

Froeschl, K. A., 1997. *Metadata Management in Statistical Information Processing - A Unified Framework for Metadata-Based Processing of Statistical Data Aggregates*, Springer. Wien, New York.

Gómez-Pérez, A., Fernández-López, M., Corcho, O., 2004. *Ontological Engineering*, Springer, London et al.

Grüninger, M. and Fox, M. S., 1995. Methodology for the Design and Evaluation of Ontologies. In *Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Canada.

Human-Computer Studies Lab, 1996. *The KACTUS Booklet version 1.0. Esprit Project 8145 KACTUS*, http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html, accessed 2005-08-20.

Institute of Electrical and Electronics Engineers, 1998. *IEEE Standard for Developing Software Life Cycle Processes.* IEEE Std 1074-1997. http://ieeexplore.ieee.org/xpl/standardstoc.jsp?isnumber=16018. 1998, accessed 2005-08-21.

Inmon, W. H. and Hackathorn R. D., 1994. *Using the data warehouse*. Wiley, New York et al..

Kantardzic, M., 2003. *Data Mining. Concepts, Models, Methods, and Algorithms*. IEEE Press, Piscataway.

Staab, S., Schnurr H. P., Studer, Rudi and Sure, York, 2001. Knowledge Processes and Ontologies. In *IEEE Intelligent Systems 16 (2001) 1*, p. 26-34.

Studer, R., Benjamins, R. V., Fensel, Dieter, 1998. Knowledge Engineering. Principles and Methods. In *IEEE Transactions on Data and Knowledge Engineering 25 (1998) 1-2*, p. 161-197.

Sure, Y.: On-To-Knowledge. Ontology-based Knowledge Management Tools and their Application. In *KI 14 (2002) 1*, p. 35-37.

Wertz, C. J., 1986. *The Data Dictionary - Concepts and Uses.* North-Holland, Amsterdam.