USING OF PROCESS ONTOLOGIES FOR DECISION SUPPORT IN INFORMATION MANAGEMENT SYSTEMS

T. Atanasova

Abstract: *The use of ontologies for data mining and knowledge discovery in databases is proposed by many researchers. In this paper it is proposed to use a lot of data collected in the information systems, to develop methods for decision support, based on previous experience in implementing similar tasks. It is shown that the investigations are directed towards to reduce the time and the resources, necessary to achieve efficient solutions in the work of enterprises and organizations. In this paper emphasis is placed on the use of problem-oriented thesauri to create ontology of processes. The development will be useful in analyzing the results and the formation of future strategies of management.*

Keywords*: process ontologies, semantics, data sources, information management*

ACM Classification Keywords: *H. Information Systems, H.4 Information Systems Applications*

Introduction

Large databases within information management systems often contain interesting hidden knowledge model that could be extracted using data mining [Mansingh, 2011]. Review of the literature shows that both objective and subjective methods for the extraction of the interesting links have to be applied. Ontologies (as an objective method) allow to reveal hidden implicit knowledge and to make them explicit by means of conceptualization, which can be used in different applications and organizations.

The approach proposed in this paper involves the use of process ontologies for the implementation of decision support, based on previous experience in implementing similar problems in the information management systems.

Ontologies

The term "ontology" in the context of knowledge management is considered as a formal explicit specification of shared conceptualization, which covers the consistent knowledge. Well-formed knowledge representation language can be adopted as ontology.

The main components of the ontology are concepts, relationships, instances, rules and axioms that restrict the interpretation of the components of the ontology. The concept is a class of objects (entities) in the area. Relationships describe the interactions between concepts or properties. The relationships are of two types: taxonomy and the associative relationship. Taxonomies systematize concepts as a hierarchical tree, and the associative relationship disposals the concept on the tree. Instances are specifications of concepts, together with the taxonomy and the relationships they form the knowledge domain.

Axioms are used to restrict the values of classes or objects (examples). Ontology may have logic inference, and then it is so-called formal ontology. Formal ontology must have axioms that restrict the possible interpretations of logical expressions.

Ontologies are created in various forms - from lexicon to dictionary terms, or as first-order logic. In a broad sense, they can be distributed over three categories: general, domain or applied ontology.

General ontologies represent knowledge, applicable in various fields of knowledge. The domain ontology focuses on the refinement of a more narrow meaning of the terms used in a certain area, and may represent a basic reality, in this specific area, but independent of a specific task. Domain ontology has four levels: domain, category, class and instance.

Applied ontology is a specific sub-ontology that contains concepts and relationships which are relevant only to the definite task, such as thesauri, which are semantic relations between lexical units. Usually they contain a small number of concepts with relationships and inference rules, which are defined in detail for solution of particular independent task.

Ontology construction

Ontologies can be expressed in different modelling techniques. The choice of an appropriate semantic model to represent ontology depends on the purpose for which the ontology is build and the underlying assumptions for achieving these goals. The mechanism of the formal description of ontology for information management is proposed in [Загорулько, 2009].

Lot of researches makes extensive engineering efforts in ontologies construction using a relational database as a data source. Construction rules of ontology elements based on relational database, which are used to generate ontology concepts, properties, axioms, instances are put forward. Most approaches are based on converting a database schema (DB) to the ontology as, for example, it is shown in [Zhang and Li, 2011].

Most domain knowledge about domain entities with their properties and relationships is embodied in document collections. On this base in the [Hou et al, 2011] a graph-based formalism to represent ontologies has been proposed. The procedure for ontology construction based on the Concept Map is considered in [Huang and Diao, 2008].

To construct a hierarchy of classes it is necessary not only to use the analysis of the database schema but to make identification of hierarchies in the data itself, that's why machine learning methods have to be involved.

The elicitation of ontology goes through few steps [Santoso et al, 2011]:

1. As the basic knowledge, a hierarchy of concepts gives information about the domain, a hierarchy of concepts is represented as a graph $G = (N, E)$; N - the set of nodes, E - the set of branches, where $N=\{n_1, n_2,\ldots n_n\}$, $E=\{e_1,e_2,\ldots e_n\}$. The graph can be described using XML Schema Datatype (XSD).

2. The inclusion of basic knowledge in the process of building ontology on OWL (Web Ontology Language) allows to speed up the process of ontologies construction. Metadata from the relational database are displayed in OWL with group of rules for the tables, attributes, and rows: R - set of tables r_i, $R=\{r_1,r_2,\ldots, r_n\}$, r_i consists of a set of tuples t, $t_i=\{t_{i1}, t_{i2}, \ldots, t_{in}\}$; $p\_a()$ - is predicate to obtain a set of attributes a from r, $p\_a(r_i)=\{a_{i1}, a_{i2}, \ldots, a_{in}\}$; $p\_k(r_i)$ - predicate for getting the primary key from the r_i; $f\_k(r_i)$ - a predicate of obtaining a secondary key from the r_i. Then the rules for mapping tables, attributes, and instances into ontology can be design [Santoso et al, 2011]. Only database scheme is used in [Santoso et al, 2011], without considering the actual data and processes that accompany them. While a large amount of data stored in relational databases, represent a valuable resource for building ontologies.

The use of ontologies for data mining and knowledge discovery in databases is proposed by many researchers [Jung, 2009; Savvas and Bassiliades, 2009; Зуева et al, 2008; Ferilli, 2011; Pomares et al, 2011]. In this paper, attention is focused on the process ontologies and its use for information management.

Process ontology

One of the most famous books on the philosophy of the processes is Process and Reality, (1929), written by Alfred North Whitehead. The author argues that all events are connected with each other and with the environment in which they arise. Thus, the world is represented as an interconnected system of large and small events; some of them are relatively stable. The events are always changing. The changes represent the actualization of certain features and disappearance of others. The world is not just exists, it always becomes. As in [Palomäki and Keto, 2006] here the idea is discussed that everything is a process and consists of the processes. These processes are divided into constant processes that are interpreted as *concepts*, and the processes which are interpreted as *events*, represent a finite set of four-dimensional space-time. Thus, the world is built from events, i.e., ontologically, all consists of processes.

The consideration of processes includes [Jussupova-Mariethoz and Probst, 2007]:

- *when* a process should be initiated and finished;
- *who* participates in this process;
- *how* this process should be performed; which results must be examined, analyzed and taken into account.

Spatial and temporal order is constructed from the relation between events $e_1, e_2, ..., e_n$. The relationship is a causal connection between events, where the cause of the event is happening before its effect. The process in its entirety is interpreted as a topological space ***T***. The space ***T*** has starting point ***a*** and the end point ***b*** of the process, which are events. The chain from ***a*** to ***b*** contains a sequence of events. The formal description of the process ontology is proposed in [Palomäki and Keto, 2011].

In context of information management systems different processes can be divided into core processes (associated with the creation of a product or service); providing processes (varying depending on changes in the composition of the basic processes and technology) and external processes of interaction. Additionally, processes can be identified that determine the trends and directions of development of basic processes, depending on the analysis and estimated trends aiming in improving management procedures in the organization [Черний and Тузовский, 2009].

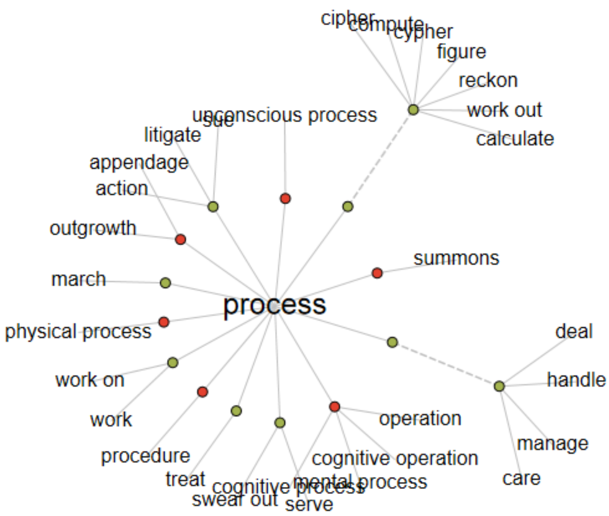

Figure 1. A part of thesaurus for the word “process” (designed using resources from http://www.visualthesaurus.com)

Using process ontology to support the decision-making in information management

Observations show that in the systems of organizational management a mixture of workflow and dataflow technology can be seen. Thus this work focuses on how to identify implicit patterns (as processes) in the streams of data available in the documents held in the systems of organizational management.

All information and data in the organization is contained in a multitude of sources ***S*** (documents, files, databases, external resources), ***S*** = $\{S_1, ..., S_n\}$. Such information is structured and includes numerical data and text fragments. Extracting knowledge from a specific domain can be considered as the construction of ontology. Using a hierarchy of concepts as the main (basic) knowledge allows the expression of the identified knowledge at a higher level of abstraction.

As an ontology a symbolic system $\{C, T, P, F, A\}$ will be considered, where *C* is the set of concepts, *T* - a thesaurus, or partial order on the set *C*, the hierarchy of relationships, "subclass" and "superclass"; *P* - the set of predicates (properties); *F* - a function that assigns to each element of *P* an element from the set of C (considering them in *T*); *A* -is a set of axioms of the ontology.

Thus, the information resources can be distributed among a set of hierarchically organized categories of thesaurus. Each category is described semantically. Software module can be developed that allocates annotation of documents concepts, instances of concepts and relationships between them.

When in a large organization a document is receiving (this is a starting point of the process) there is a problem of its proper handling and direction to the appropriate department. Resolution of the document and its classification is sometimes difficult and can bring to the loss of time and possible non-efficiency in its work. Each document represents a specific event; the relationship between documents defines the process. Each action is associated with the document; it is an event in the chain process of passage of the document. The process, which has been formed during solution of certain problems, is implicitly stored in the database. Then the objective is to identify processes that have brought to the event, as it reflected in the set of documents, and to predict future events based on the past experience.

Description of the incoming document is summarized in terms of problem-oriented thesaurus with text processing technology. Based on this description the graph of the document is constructed. According to the obtained descriptions semantic search can be organized later. Then the concept - synonymous are normalized. The conceptual closeness is derived from existing concepts in the domain.

To determine the relationship of units with one or more specified activities in the information system a method may be used for calculation of the semantic proximity, which also assesses the relationship of these units to the selected areas. Semantic interoperability can be calculated using the various existing methods. For example, an algorithm for computing the conceptual similarity with the use of problem-oriented thesauri can be based on fuzzy logic that is used to ease the difficulties in developing and analyzing complex systems.

Concepts in the ontology have many attributes, operations and associations. The mechanism of fuzzy inference is used to calculate the conceptual closeness as the similarity of attribute x_A, the operation of x_O, by association in this area x_D and the range of the similarity x_R - as the degree of affiliation between the concepts. That is, the concept can probably be qualified to a given node in the ontology if the conceptual closeness is high, and so a group of entities with identical attributes, operations, and associations with other entities is formed.

Discovering process models from information system event logs is definitely non-trivial. Within the analysis of event logs, process can be defined as the automated construction of structured process models from event logs [Goedertiera et al, 2011].

By the classification of records and the use of the process ontology the system evaluates the documents. The results are provided to the experts in the form of a semantic network of documents, which have a conceptual affinity with concepts and relations in the ontology. In addition, the ontology of processes will determine what information to extract and how to accelerate the semantic search. Semantic search is looking for documents in the information resources of the organization, whose semantic descriptions are similar to the semantic query. As a result of the semantic search, a list of found by the inquiry documents is ordered according to the semantic proximity value and further grouped according to the process. This allows extracting and using of the information and knowledge stored during operation of information systems.

Conclusion

The main problem for the information management systems is to manage the information and knowledge. The ontology systems specify semantic information about documents, services, and workflows used in various information systems. Ontology is also used to model the relations between documents and services and to manage process configurations.

The primary aim to use of ontologies is to integrate different applications. The diverse of service models and tools in information systems is constantly increasing, the necessity arises to have uniform base for their using [Daskalova and Kolchakov, 2000]. Exchanging of various data needs to take into consideration not only syntaxes but semantics of the data and processes.

The problem of inclusion of intelligent information systems constitute a scientific problem, ensuring that the theoretical development of decision support (DSS) in real-world application areas is based on their specificity. The knowledge-intensive approaches such as ontology are suitable for modeling complex systems.

To provide support decision making in information systems that is focused on specific goals and objectives of management process the problem-oriented thesauri are used for the formation process ontology.

The use of ontologies allows processes to implement algorithms and decision support functions. The article focuses on the use of problem-oriented thesauri to create an ontology of processes. Rules of generalization and analysis of data and methods for finding solutions based on problem-oriented thesauri and ontologies processes still have to be developed.

Bibliography

Daskalova H., K. Kolchakov, Some investigations for optimizations of information interchange. International Scientific Conference "Communication, Electronic and Computer Systems 2000", May 2000, Sofia, Bulgaria. Reports, Vol.1, pp.206-210.

Ferilli S., Automatic Digital Document Processing and Management, Advances in Pattern Recognition, Springer-Verlag London (2011)

Goedertiera Stijn, Jochen De Weerdta, David Martensa,b, Jan Vanthienena, Bart Baesensa. Process discovery in event logs: An application in the telecom industry, Applied Soft Computing 11 (2011) 1697–1710

Hou Xin, S.K. Ong, A.Y.C. Nee, X.T. Zhang, W.J. Liu, GRAONTO: A graph-based approach for automatic construction of domain ontology, Expert Systems with Applications 38 (2011) 11958–11975

Huang Ning, ShiHan Diao, Ontology-based enterprise knowledge integration, Robotics and Computer-Integrated Manufacturing 24 (2008) 562–571

Jung Jason J., Semantic business process integration based on ontology alignment, „Expert Systems with Applications" 36 (2009) 11013–11020

Jussupova-Mariethoz Yelena, Andre-Rene Probst, Business concepts ontology for an enterprise performance and competences monitoring, Computers in Industry 58 (2007) 118–129

Lee C.-S. et al. Automated ontology construction for unstructured text documents, „Data & Knowledge Engineering” 60 (2007) 547–566

Mansingh G. et al. Using ontologies to facilitate post-processing of association rules by domain experts, „Information Sciences” 181 (2011) 419–434

Palomäki J. and H. Keto / A Formal Presentation of the Process-Ontological Model, Information Modelling and Knowledge Bases XXII A. Heimburger et al. (Eds.) IOS Press, (2011) pp.194-205

Palomäki Jari and Harri Keto, A Process-Ontological Model for Software Engineering, 2nd Int. Workshop on Philosophical Foundations of Information Systems Engineering PHISE'06, (2006), Luxembourg, pp.720-726

Pomares Alexandra, Claudia Roncancio, Van-Dat Cung, and Maгна-del-Pilar Villamil, Improving Source Selection in Large Scale Mediation Systems through Combinatorial Optimization Techniques, A. Hameurlain, J. Küng, and R. Wagner (Eds.): TLDKS III, LNCS 6790, pp. 138–166, 2011.

Santoso H.A. et al., Ontology extraction from relational database: Concept hierarchy as background knowledge, „Knowledge-Based Systems” 24 (2011) 457–464

Savvas, N. Bassiliades, A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration, „Expert Systems with Applications” 36 (2009) 4467–4478

Zhang Lei, Jing LI, Automatic Generation of Ontology Based on Database, Journal of Computational Information Systems 7:4 (2011) 1148-1154

Загорулько Ю.А., Е.А. Сидорова, О.И. Боровикова, Онтологический подход к построению систем информационной поддержки научной и производственной деятельности, II Всероссийская конф. «ЗНАНИЯ – ОНТОЛОГИИ – ТЕОРИИ» с межд. уч. 20–22 октября 2009 г. Новосибирск (2009) http://math.nsc.ru/conference/zont09/reports/91Zagorulko-Sidorova-Borovikova.pdf

Зуева А. Г., Б. В. Носков, Е. В. Сидоренко, Е. И. Всяких, С. П. Киселев „Практика и проблематика моделирования бизнес-процессов” „ИТ-Экономика” ДМК Пресс, (2008)

Черний А.В., Тузовский А.Ф. Развитие информационной системы организации с использованием семантических технологий, „Знания-Онтологии-Теории” (ЗОНТ-09), 20-22 октября 2009 г. Новосибирск (2009)

Acknowledgment

The paper is published with financial support of the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA (www.ithea.org) and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine (www.aduis.com.ua).

Authors' information

Tatiana Atanasova – Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria, 1113, Sofia, Acad. G. Bonchev 2, e-mail: atanasova@iit.bas.bg

Major Fields of Scientific Research: Knowledge-Based Systems, Learning Structures, Semantic Technologies