[Markov et al, 2004] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. *Formal Definition of the Concept "INFOS"*. Proceedings of the Second International Conference "Information Research, Applications and Education" i.TECH 2004, Varna, Bulgaria. Sofia, FOI-Commerce, 2004, pp. 71-74. Int. Journal "Information Theories and Applications", 2004, Vol.11, No.1, pp.16-19

[Markov P., 2002] P.Markov. Think with Your Mind. Markov College. Sofia, 2002. (in Bulgarian)

[Pavlov, 1987] T.Pavlov. Collection of Selected Works, Vol. II. Theory of Reverberation. Science and Art, Sofia, 1987. (in Bulgarian).

[Rosenblueth, Wiener, 1945] A.Rosenblueth, N.Wiener. Role of Models in Science. Philosophy of Science, Vol.12, No.4, 1945.

[Shannon, 1949] C.E.Shannon. The Mathematical Theory of Communication. In: The Mathematical Theory of Communication. Ed. C.E.Shannon and W.Weaver. University of Illinois Press, Urbana, 1949.

[Tomov, 1991] K.Tomov. The Resomal-Isomorphic Principle. Arges, Sofia, 1991. (in Bulgarian)

[Wartofsky, 1979] M.W.Wartofsky. Models. Representation and the Scientific Understanding. D.Reidel Publishing Company, Dordrecht: Holland /Boston: USA, London: England/, 1979.

[Popper, 1968] K.R.Popper. Conjectures and Refutations: The Growth of Scientific Knowledge. Harper & Row, Publishers, New York and Evanston. 1968.

Authors' Information

Krassimir Markov - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; Institute of Information Theories and Applications FOI ITHEA, P.O.Box: 775, Sofia-1090, Bulgaria; e-mail: foi@nlcv.net

Krassimira Ivanova - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria

Ilia Mitov - Institute of Information Theories and Applications FOI ITHEA, P.O.Box: 775, Sofia-1090, Bulgaria; e-mail: foi@nlcv.net

TECHNOLOGY FOR ONTOLOGICAL ENGINEERING LIFECYCLE SUPPORT [1]

Vladimir Gorovoy, Tatiana Gavrilova

***Annotation**: Presented paper describes software system project ONTOLINGE-KAON that provides technological support for the whole lifecycle of ontological engineering. The main stress is put on the evaluation of maturity and quality of ontologies and on the usage of ontologies with the help of automated generation of knowledge portals, based on ontologies. Possibility of creation of knowledge portals built on top of ontologies can become a big step forward in the field of e-learning. The paper presents advantages provided by knowledge portals based on top on ontologies.*

***Keywords**: ontological engineering, knowledge engineering.*

***ACM Classification Keywords**: H.0 Information systems – General, I.2.6 Artificial intelligence - Learning*

Introduction

ONTOLONGE-KAON is aiming at providing technological support for the full lifecycle of ontological engineering. At present a great number of components and software products are implementing various tasks in work with

[1] The research is partially supported by Russian Foundation for Basic Research (grants **06-01-81005 and 07-01- 00053)**

ontologies. According to this fact an urgent problem of integration of these components in a common system supporting all stages from creation of an ontology, evaluation of its maturity and quality up to using of it by target users emerges.

Evaluation of quality of developed ontologies and support of the phase of ontology usage is a bottleneck of present instruments providing technological support for working with ontologies. And this is regardless of the fact, that today we can state a broad interest to ontological engineering. Thus there are more than 50 ontological editors:

- Protégé,
- GALEN Case Environment (GCE),
- ICOM
- Integrated Ontology Development Environment
- IsaViz
- JOE
- KAON (including OIModeller)
- KBE -- Knowledge Base Editor (for Zeus AgentBuilding Toolkit)
- LegendBurster Ontology Editor
- LinKFactory Workbench
- Medius Visual Ontology Modeler
- NeoClassic
- OilEd
- OLR3 Schema Editor
- OntoBuilder
- And others.

But all of them don't support development of corporate knowledge portals. The first step in this direction was done in the KAON project (http://kaon.semanticweb.org) [Motik et al., 2002]. One of the developed subsystems (KAON Portal) allowed generating portal providing web-interface for browsing of ontologies created by OI-Modeler. One of the KAON Portal's shortcomings is absence of any information bound to ontology (moreover, binding of information is not supported). Besides KAON Portal generates a portal on the basis of ontologies, kept in an internal format (private expansion of RDFS), not being a conventional standard. Thus, ontologies used for generation of a portal should be created only in KAON OI-Modeler what is certainly an essential restriction.

One more example of the system focused on creation of a portal is PORTO [Gavrilova et al., 2003]. Work in PORTO system consists of the following stages:

1. Creation of a portal ontology by an analyst by means of the visual editor.
2. Creation of design of a portal and binding ontology concepts to representation by the web-designer.
3. Online generation of pages of a portal by server PORTO in reply to the user queries.

One of the disadvantages of PORTO is that it lacks integration with other systems for ontology development. Thus ontology for portal generation can be created only in the offered visual editor that kept its data in own internal format.

Automated generation of knowledge portals on the top of ontologies would become a great step forward in the field of e-learning. For example, automatic creation of a portal the course would be invaluable to students and would release teachers from a part of work for creation of the portal during development of an ontology for a training course or a topic of one lecture.

The fact of portal being built on the top of ontology can help users during its use. In this case it is possible to enhance standard text search on a portal with the reasoning system being able to run queries on ontology. As the interface for the advanced users it is possible to use the form for input of RQL-inquiries [Karvounarakis et al., 2003] in this case. Executing of such queries will release users from a filtration of superfluous information that he often receives as a result of standard text queries.

System components and their interaction

Components of the system are focused on solving of the following problems:

1. Ontologies producing
2. Ontologies evaluation
3. Ontologies usage

For ontologies producing it is possible to use any ontology editor or some other instrument for ontologies creation allowing saving ontologies in OWL format (http://www.w3.org/TR/2003/CR-owl-features-20030818). Among ontological editors it is worth mentioning Protege [Noy et al., 2001] or SWOOP [Kalyanpur et al., 2004]. For creation of OWL ontologies programmatically Jena (HP labs - http://www.hpl.hp.com/semweb/downloads.htm), KAON2 (http://kaon2.semanticweb.org), IODT (IBM Integrated Ontology Development Toolkit - http://www.alphaworks.ibm.com/tech/semanticstk) or OWL-API (http://owl.man.ac.uk/api.shtml) can be used. These APIs considerably facilitate life for software developers who implement import of their internal ontologies to OWL format. Thus we get an ontology in OWL format as an output of the first component.

Important part of ONTLINGE-KAON system is an evaluation module for created ontology and providing recommendations for its improvement. One can use existing tools for ontologies evaluation (OntoAnalyser, KAON2) as such a component. Unfortunately these instruments for evaluation haven't reached serious level of maturity and are not widely used. This is caused by the fact that recommendations of knowledge engineers in the domain of ontologies forming can hardly be formalized and implemented in software. As a result of this, problem of building new instruments for ontologies meeting needs of the majority of users is topical. Some ideas regarding building of such a module are presented below in the section devoted to evaluation component.

There are the following requirements for evaluation component in ONTOLINGE-KAON system:

1. Ability to work with ontologies in OWL format.

2. Inference of remarks concerning ontology quality and providing suggestions for quality improvement.

Corporate and educational portals, generated on the basis of ontologies, formed by other system components, are targeted at implementation of ontology using. We formulated the following requirements for a portal:

- Generation on top of ontology in OWL format
- Access to created ontology
- Adding relevant information to concept instances for displaying on the page generated by the portal
- Ordinary portal search
- Portal search using that portal is built on the top of ontology

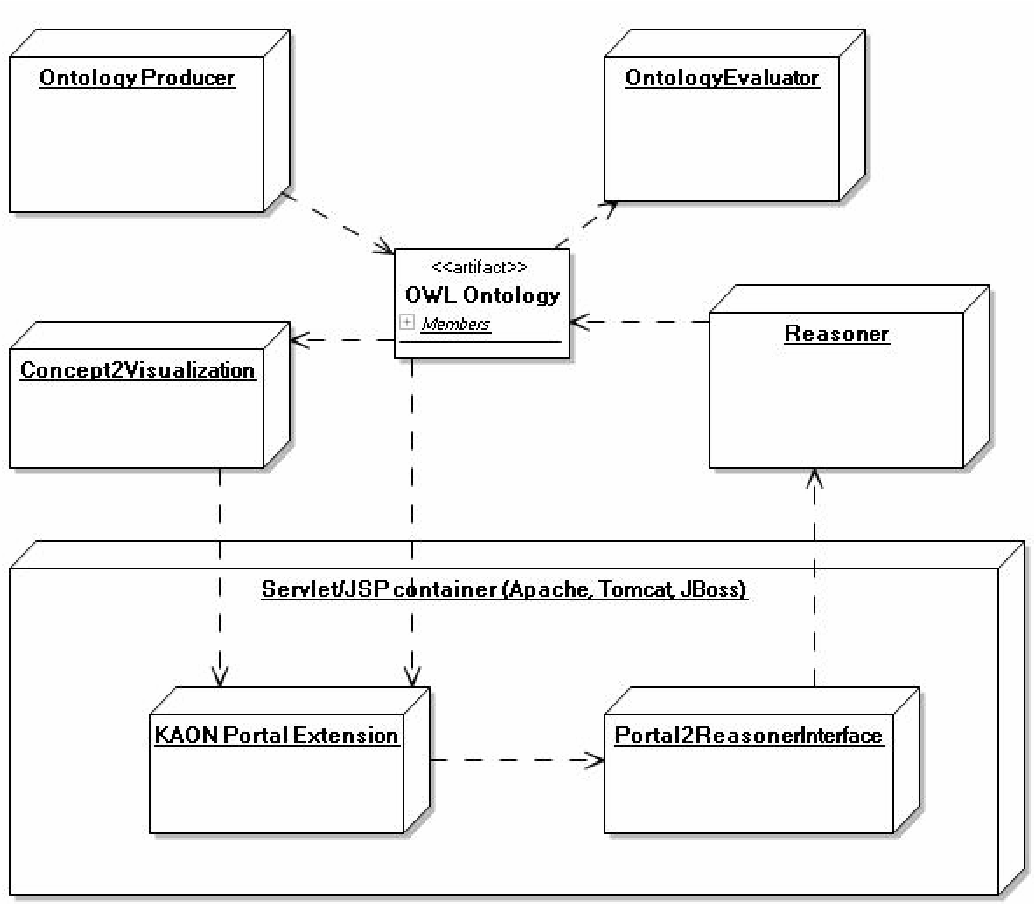
Architecture of the system solving all stated problems is presented on picture 1. Its main components are:

- OntologyProducer – component of ontology forming
- OntologyEvaluator – component for evaluation of quality and maturity of ontologies
- KAON Portal Extension – extension of KAON Portal'a. Mission of this component includes generation of a portal built on ontologies
- Concept2Visualization – module providing possibility of matching presentation with concepts of the ontology (for exampled concept "project OntoWeb" is matched to some visualization on the generated page which contains all necessary information concerning the project)
- Portal2ReasonerInterface – module providing interface for search queries bound to portal ontology

- Reasoner - module for reasoning (for example Pellet OWL Reasoner).
- Servlet/JSP container – server supporting Servlets' and JSPs'. On such a server KAON Portal Extension module can work (for example Apache Tomcat or JBoss).

Thus, in created architecture we develop the following components and integrate them with each other and other components of ONTOLINGE-KAON system:

- KAON Portal Extension
- Concept2Visualization
- Portal2ReasonerInterface
- OntologyEvaluator



Pic. 1. ONTOLINGE-KAON architecture

Component of ontologies evaluation – OntologyEvaluator

Existing instruments for ontologies evaluation (OntoAnalyser, KAON2) can be used as a component of ontologies evaluation. Unfortunately existing evaluation tools haven't reached serious level of maturity and are not widely used. Suggested solution is supposed to eliminate shortcomings peculiar to existing evaluation tools.

Analysis of student works on creating ontologies in simple and well-known domains revealed several primary factors that distinguished good ontologies from bad ones. These laws can be reformulated and made applicable for a practicing knowledge engineer. The main hypothesis can be stated as: "Harmony = conceptual balance + clarity".

At that conceptual balance means that:

- Concepts of the same level are connected with parent concept with the same type of relations (e.g. “class-subclass”, “whole-part”).
- Depth of ontological tree branches should be about the same (±2).
- The whole picture should be pretty symmetric.
- Cross links should be avoided as far as possible

Clarity includes:

- Minimization: Thus maximum number of concepts of the same level or branch depth shouldn’t exceed famous Ingve-Miller number (7±2) [Miller, 1956].
- Clarity for reading. Relations’ type should be as obvious as possible, not to overload ontology scheme with unnecessary information and skip names of relations.

All these laws correspond with some results of Gestalt psychology formulated by Maks Wertheimer as early as 1944 [Wertheimer, 1944]. Thus the main principle of good gestalt (good shape) or Pragnanz law was formulated in such a way:

“Organization of any structure in nature or cognition will be as good (regular, complete, balanced, or symmetrical) as the prevailing conditions allow “

The major part of enumerated factors may be formalized and their verification may be implemented in OntologyEvaluator. Thus using of this component can contribute to create harmonic ontologies.

Ontology using

On this phase our solution suggests generation of knowledge portals based on built ontologies and further using of a portal as an ontology navigation tool, for necessary information search and for querying ontology.

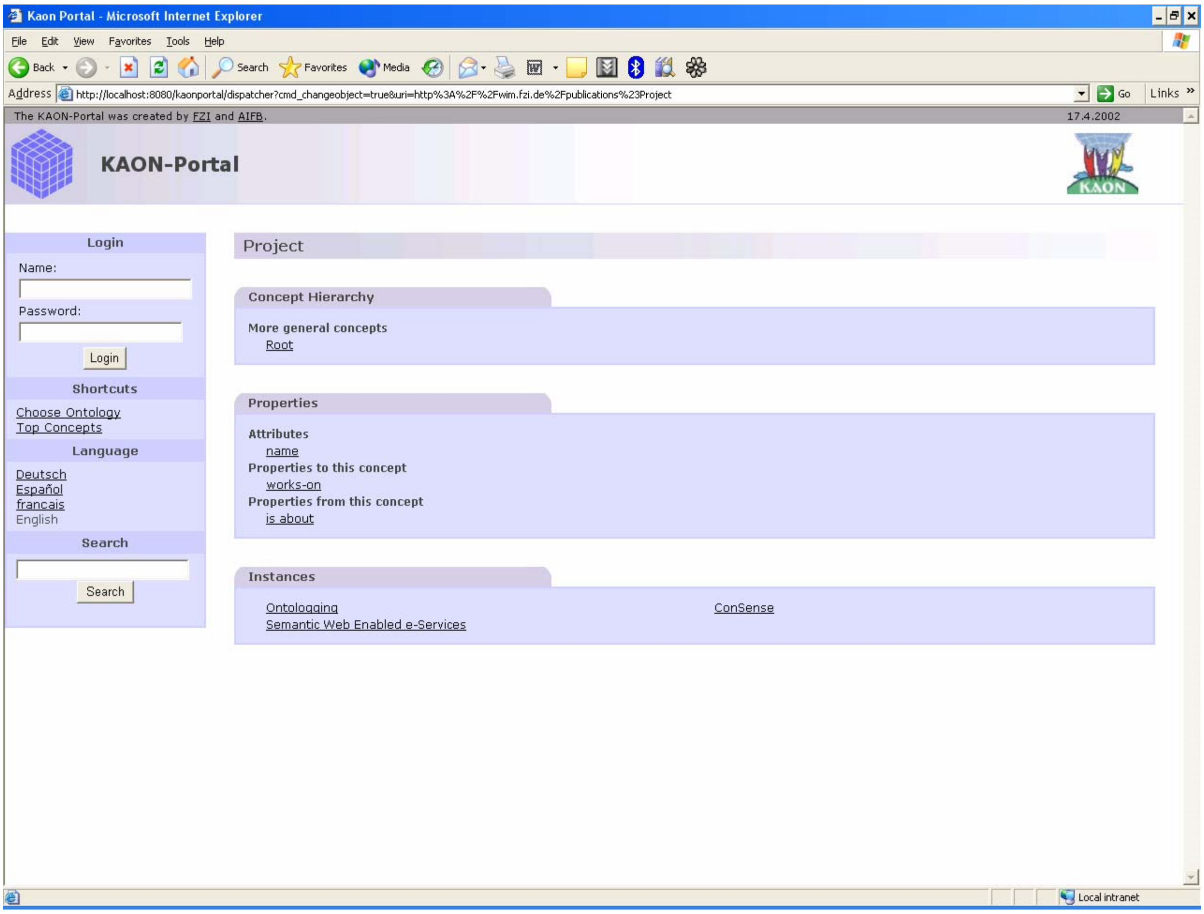
The components below are related to the ontology using:

- KAON Portal Extension
- Concept2Visualization
- Portal2ReasonerInterface
- OntologyEvaluator
- Reasoner
- Servlet/JSP container.

We can use KAON Portal module developed in KAON project as a base prototype for KAON Portal Extension. KAON Portal allows generating portal for navigation of ontology, described on proprietary extension of RDFS (ontology in this format can be produced by OI-Modeler). Screenshot devoted to project concept is presented on pic. 2.

For practicable using of the generated portal as an educational knowledge portal it is necessary to implement in KAON Portal Extension the following features:

- OWL support, because ability to work with conventional standard in the field of ontologies description would contribute better interoperability and integration with other instruments of ontological engineering.
- Ability to bind presentations with ontology concepts. Without this functionality it is impossible to create a usable knowledge portal. This feature is implemented by Concept2Visualization module.
- Interface for search queries of a portal ontology. In the simplest case it can be ability to input RQL-queries providing results on them. Portal2ReasonerInterface implements this functionality.



Pic. 2: Concept of project

Conclusion

Having all the features described above ONTOLINGE-KAON can become a big step forward on the way to using technologies and methodologies of ontological engineering for creating educational knowledge portals. Suggested approach for evaluation of quality and maturity of ontologies seems to be interesting. It can assist in creating high quality harmonic ontologies. Compared to existing systems of ontological engineering new is ability for dynamic generation of portal based on created ontology and containing possibilities for ontology navigation and information constituent related to ontology instances. Proposed solution is flexible and allows to automate reflection of changes made in ontology on generated pages. New is also a feature of portal search leveraging that portal is based on ontology. This functionality radically differs from ordinary search with the help of search systems because it provides only semantically correct results and saves user from the necessity of choosing from many variants much of which are very far related to expected results.

Bibliography

[Gavrilova et al., 2003] T.A. Gavrilova , V. A. Gorovoy. Ontological Engineering for Corporate Knowledge Portal Design // In "Processes and Foundations for Virtual Organizations", Eds. L. Camarinha -Matos and H. Afsarmanesh, Kluwer Academic Publishers, 2003. - p.289-296.

[Kalyanpur et al., 2004] A. Kalyanpur, E. Sirin, B. Parsia, J. Hendler. Hypermedia inspired ontology engineering environment: Swoop. // In Proceedings of 3rd International Semantic Web Conference (ISWC-2004), Japan (Poster).

[Karvounarakis et al., 2003] G. Karvounarakis, A. Magkanaraki, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, K. Tolle. Querying the Semantic Web with RQL. // In Computer Networks and ISDN Systems Journal, Vol. 42(5), August 2003, pp. 617-640.

[Miller, 1956] G.A. Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. // In The Psychological Review, 1956, vol. 63, pp. 81-97

[Motik et al., 2002] Boris Motik, Alexander Maedche, Raphael Volz. A Conceptual Modeling Approach for Semantics-Driven Enterprise Applications. // In Proceedings of the First International Conference on Ontologies, Databases and Application of Semantics (ODBASE-2002). Springer, 2002.

[Noy et al., 2001] N.F. Noy, M. Sintek, S. Decker, M. Crubezy, R.W. Fergerson, M.A. Musen. Creating Semantic Web Contents with Protege-2000. // IEEE Intelligent Systems 16(2), pp. 60-71, 2001

[Wertheimer, 1944] M. Wertheimer. Gestalt theory. // In Social Research, 11, 78-99.

Authors' Information

Vladimir Gorovoy – PhD student, Saint-Petersburg State Polytechnic University, Intelligent Computer Technologies Dpt. 195251, Politechnicheskaya 29/9, St. Petersburg, Russia; e-mail: vgorovoy@mail.ru

Tatiana Gavrilova – Professor, Saint-Petersburg State Polytechnic University, Intelligent Computer Technologies Dpt. 195251, Politechnicheskaya 29/9, St. Petersburg, Russia; e-mail: gavr_csa@rambler.ru

INTELLIGENT SEARCH AND AUTOMATIC DOCUMENT CLASSIFICATION AND CATALOGING BASED ON ONTOLOGY APPROACH

Vyacheslav Lanin, Lyudmila Lyadova

Abstract: This paper presents an approach to development of intelligent search system and automatic document classification and cataloging tools for CASE-system based on metadata. The described method uses advantages of ontology approach and traditional approach based on keywords. The method has powerful intelligent means and it can be integrated with existing document search systems.

Keywords: electronic document, automatic document classification and cataloging, ontology approach, information system development.

ACM Classification Keywords: I.2.7 Artificial Intelligence: Natural Language Processing – Text analysis; D.2.2 Software Engineering: Design Tools and Techniques – Computer-aided software engineering (CASE).

Introduction

Development tools used for implementing large distributed information systems, which consist of separated subsystems and should be installed in territorially remote organizations, should meet the requirements, providing possibility of its customization on various maintenance conditions and user's requirement during installation and dynamically during maintenance. Organizations has various technical possibilities, organizational and business forms, it makes information system development difficult. Implementation of these requirements provides efficiency of expenses for system creation, a high degree of its adaptability and scalability, robustness of the system.

The CASE-system METAS bases on interpretation of the multilevel metadata. The metadata describes an information system from different points of view and with a various grain size. Opportunities of dynamic system customization are provided by re-structuring of a database, generation and customization of user interface, generation of queries and creation of reports [Lyadova, 2003].