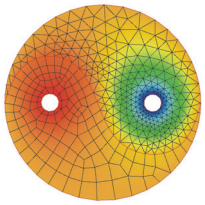
Publishing House
AKAPIT

COMPUTER METHODS IN MATERIALS SCIENCE

Informatyka w Technologii Materiałów

Vol. 7, 2007, No. 2



ONTOGRATOR – AN INTELLIGENT ACCESS TO HETEROGENEOUS KNOWLEDGE SOURCES ABOUT CASTING TECHNOLOGY

STANISŁAWA KLUSKA-NAWARECKA[1], GRZEGORZ DOBROWOLSKI[2], ROBERT MARCJAN[2], EDWARD NAWARECKI[2]

[1] Instytut Odlewnictwa, ul. Zakopiańska 73, 30-418 Kraków
[2] Katedra Informatyki, Akademia Górniczo Hutnicza, al. Mickiewicza 30, 30-059 Kraków

Abstract

OntoGRator is a system for integration and organization of access to various information and knowledge stored in independent computerized sources that stay external to it. The sources can be of various structuralization: data in relational data bases, documents, XML files, web pages etc. It is assumed that the sources gathered knowledge of some domain like casting technology or metallurgy. Integration of such sources is realized by defining two kinds of ontologies: "domain otologies" representing concepts of the domain together with dependences between these concepts, and "data source ontologies" describing each data source. Use of the knowledge gathered in the system consists in navigating through notions of the domain ontology and - after identifying the interesting concept – in retrieval of the information about chosen this way topic (coming from different data sources).

The paper discusses some architectural and implementation issues of OntoGrator as well as its distinctive application in the field of casting technology.

Key words: information systems, ontologies, data integration, casting technology

1. INTRODUCTION

Fast and effective access to information is a condition of success for almost all activities and even can be taken as a measure of modernity. Of course, the matter is when the information is crucial to some project and can influence on what-ever decisions that must be made during its realization. The above statement is of rather general nature but it acquires quite concrete sense in industrial "spheres" and particularly in the contemporary casting industry (Dobrowolski et al., 2003; Dobrowolski et al., 2004; Lepatti & Berger, 2004). As justification one can enumerate the following features of the industry:

- great variety of technologies and materials,
- high quality requirements that must be met by majority of products to be competitive at the global market,
- existence of the wide and strong scientific support that open possibility of new technologies implementation based on advanced theoretical methods, modern measurement techniques and computer tools application,
- possibility of the production intensification by introducing new ideas in the fields of organization and management.

All these above indicate that innovation activities in the industry involve usually several scientific institutes and firms and their gathered knowledge as well as information must be interchange in course. A part of this information which does not constitute know-how of its owners can be (or even must be) publicized. It deals with information that can be called as encyclopedic one. Moreover the informa-

 ISSN 1641-8581

tion is often computerized already.

The paper discusses some architectural and implementation issues of OntoGrator (Mackowiak, 2006) – a system of a large variety of heterogeneous knowledge resources. The issues deal with both their content as well as modes of their utilization. Such a system can be characterized briefly as an easily extensible (modified) platform that organizes co-operation of components implementing these resources. It is assumed that the components have been built at different moments of time, under different conditions, and using computer technologies (tools) that did not keep exactly the same standards and, in consequence, are hardly integrated. In order to make the co-operation of the components possible as well as to facilitate use of the system by users of different profiles, providing possibly full knowledge about the components and system in a symbolic and ready to process form is assumed. In this way a formal basis for the architecture becomes the idea of *systems with explicit knowledge* proposed in (Dobrowolski et al., 2004).

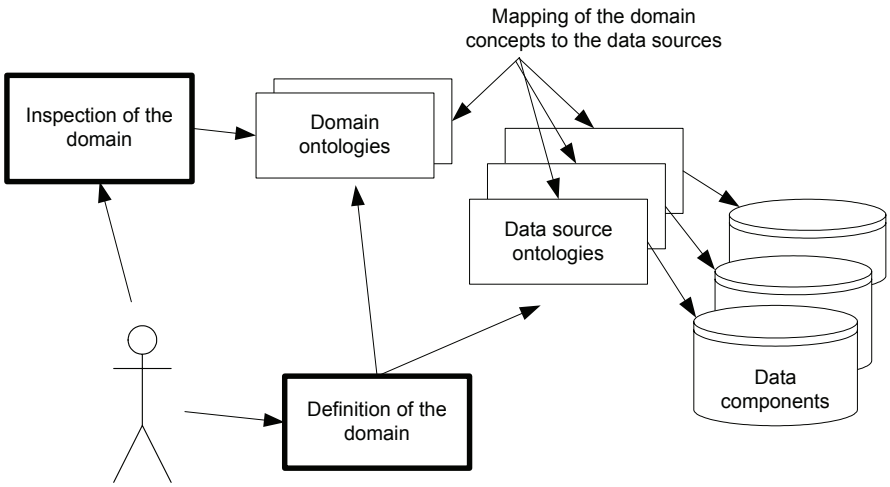


Fig. 1. *Ontology driven access to heterogeneous data sources – general concepts.*

2. ONTOGRATOR - AN INTELLIGENT ACCESS TO HETEROGENEOUS KNOWLEDGE SOURCES

OntoGRator is a system that makes available information stored in independent computerized sources that stay external to it. In order to allow coordinated presentation of these data and, in particular, to gain a synergy effect that can arise from some differences (with respect to the origin and scope), their sources must be subjects of prior integration that is supported by the system also. Both the integration process and mechanism necessary to aid an effective queries in the, building this way, conglomerate of data intensively use the applied to sources shared interpretation given in the form of an ontology (Guarino, 1998; Staab & Studer, 2004). The integrated sources become components of the discussed information environment that here is dedicated to casting technology but can be prepared also for whatever engineering domain.

It must be mentioned that presented implementation of OntoGRator assumes that a component does not need any input data except those necessary to carry out a query and produce information that are directly presented to a user. Such assumption does not exclude "classic" data sources (e.g. data bases) but cooperation (data exchange) among components is not supported. Such a bit simplified solution is taken because majority of candidates for components are legacy systems operating in the above sketched mode. Moreover integration of those systems seems to be of far more importance than design of an environment fully flexible and complete.

3. INTEGRATION OF SOURCES

OntoGRator is ready for modification of the subject. It is armed with mechanisms aiding successive integration of new domains as well as sources about domains already included. Integration of a new domain needs design and introduction into the system an ontology for such domain. Building-in of a new source (creation of a component) – working-out of it in the aspect of compatibility with a binding in the system domain ontology (it is possible to change it also) and constructing of semantic links between the ontology and structures (objects) of the source. An object means for instance a row of some view of the data base.

In fact, a question of the links is equivalent to implementation of an addressing system for the objects. It ought to be stressed that identification (addressing) of the objects in not only a software engineering problem but can depend on characteristics of sources of the given domain.

Possibility of ontology modeling in the system is almost unlimited. Description logics (Baader et al., 2003) is taken as a basis without any constraints with respect to defined operators or features of relations. It stems from the fact that queries can be real-

ized via simple look-up in the ontologies (not unconstrained inference). Objects of the sources are treated as extensions of the description logics in the presented approach.

Besides the domain ontologies OntoGRator maintains also ontologies of the sources (components) themselves. Ones more the extensions are here objects of the sources what ideally completes knowledge-based description of the system. The components ontologies are specializations of the OntoGRator ontology. It embraces definitions of concepts and their attributes describing the proper way of processing of the particular source type (e.g. data base – identification and authorization data, forms of queries declaration etc.). The idea of ontology driven integration and access to heterogeneous data sources is shown in figure 1.

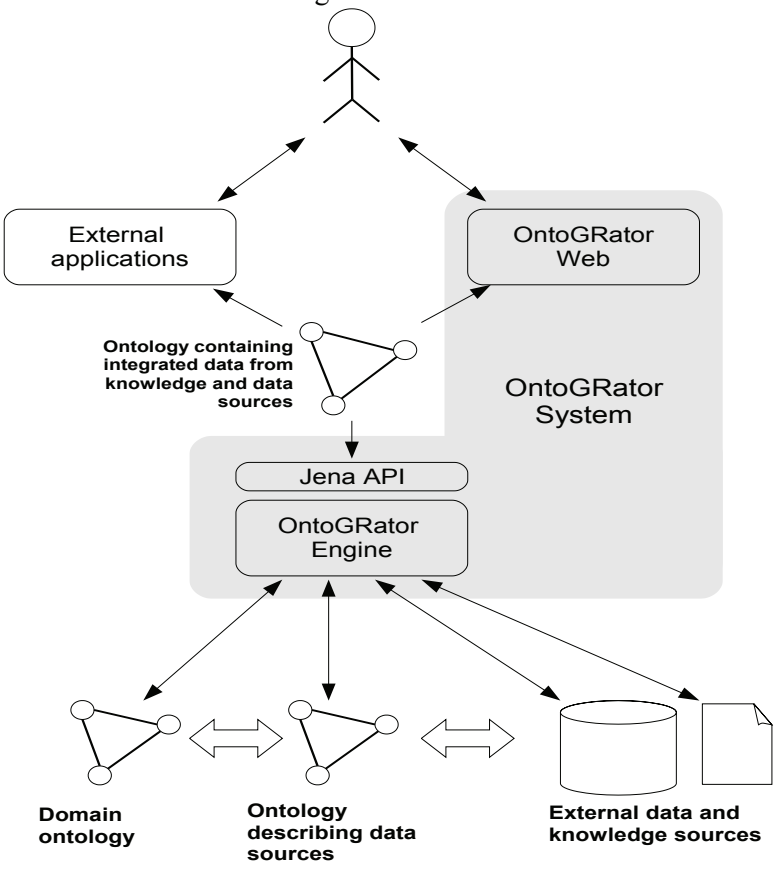


Fig. 2. OntoGRator – architecture.

The problem of identification of the objects is solved basing on adopted thesauri, one for each supported domain. The basic requirement is the ability to describe each data object and ontology notion with the appropriately composed set of keywords. The solution is justified by the fact that many legacy sources (e.g. bibliographical data bases) are already structuralized in this manner (e.g. Sinte and Norcast data bases (Dobrowolski et al., 2003) integrated in the presented implementation of OntoGRator with the thesaurus by Foundry Research Institute in Krakow). Moreover working with sets of keywords is easy and natural for people preparing such sources and their experience can be used also for new ones.

The presentation layer of OntoGRator is oriented towards not only showing chosen objects of the appropriate components but also visualizing the ontology as a means of supporting data retrieval as well as informing a user about notion structure and "knowledge" contents of the system. The first aspect is realized by system tools according to features of the objects (e.g. reproducing a picture of the given format). The second – by the standard interface.

4. ARCHITECTURE AND FUNCTIONALITY OF THE SYSTEM

The OntoGRator (figure 2) system consists of two main subsystems:

- OntoGRator Engine – reasoning engine integrating data from external heterogenous sources together with information from the problem domain described by the domain ontology. The module is based on Jena software (McBride, 2002). Integrated data are available in the form of extended ontology (concepts from domain ontology supplemented with object instances from data sources.) and are provided by Jena API to the upper layers of he system.
- OntoGRator Web – web based application presenting delivered by the Jena API (that is integrated by the OntoGRator engine) ontologies in a form of web pages. OntoGRator Web allows for searching, querying and navigation through the ontology and access to the individual instances of data objects coming from external data sources.

Moreover there is a possibility to connect to the system from external applications. Data served by the OntoGrator Engine are delivered in a standard form of the ontology (e.g. can be easily transformed to the RDF or OWL format) and using the Jena API could be accessed by external tools such as ontology browsers (e.g. Jena, Protage or KAON tools) or even more complicated software like reasoning or expert systems.

The principles of the system operation can be summarized as follows:

- There is (defined earlier) base domain ontology defining the concepts and relations appearing in the field (domain).
- Additionally, there are a couple of external data or knowledge sources containing information about that domain – various documents, data in relational data bases, XML files, web pages etc. Information is distributed and difficult to access as a whole.
- Integration of such knowledge and data sources (including of a new data source to the system) depends on description of the particular data or knowledge source in terms of the ontology, mapping of its terms to the domain concepts, and definition of the access methods for that data source.
- OntoGRator Engine takes such ontologies (domain ontology and ontologies describing data sources) and produces the final ontology which is a sum of the input ontologies extended with object instances that come from the data sources. The process is carried out in online mode. Queries to the final ontology are transformed by the OntoGRator Engine to the queries to the knowledge sources.

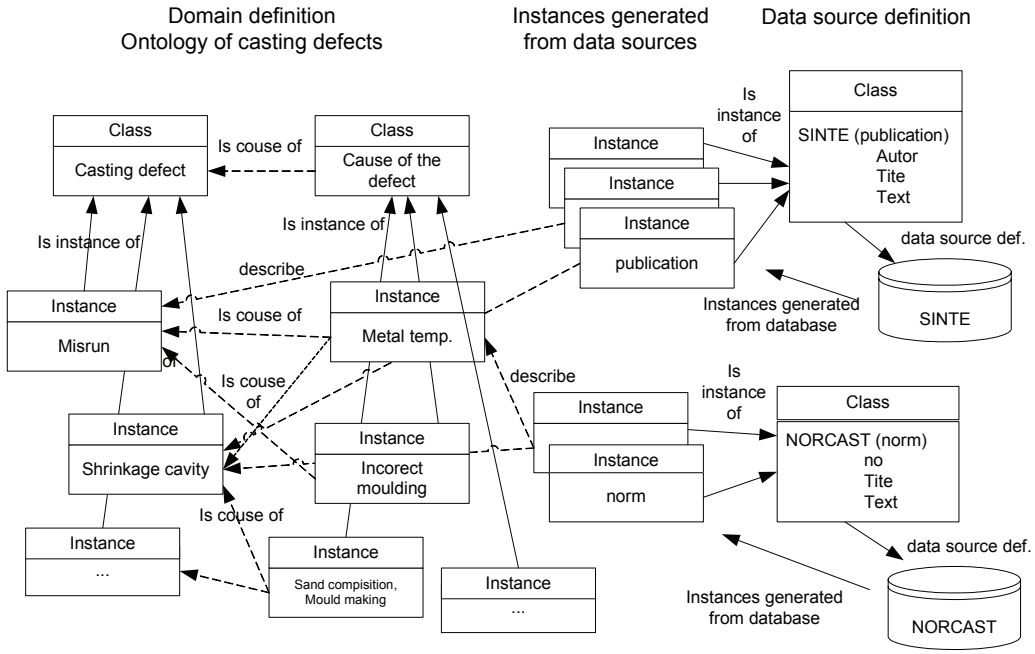


Fig. 3. Ontology representing knowledge about casting defects (a fragment).

As a result, OntoGrator Engine delivers integrated data in a form of the graph with nodes representing concepts and their instances, and arcs - relations between these concepts and instances. The graph can be inspected using the OntoGrator, Web applications as well as various tools designed for ontology manipulation.

The system was implemented using the Java technology. The modules for ontology management and reasoning of the first version was based on KAON tools (Voltz et al. 2003) (http://kaon.semanticweb.org/), later some of these modules was reimplemented using Jena software (McBride, 2002) (http://jena.sourceforge.net/).

5. APPLICATION TO THE CASTING TECHNOLOGY

The OntoGRator system was applied to integration of knowledge sources about casting technology, particularly to the knowledge about casing defects. The application consists in the definition of the domain ontology of casting defects (fragment of such ontology is schematically shown on figure 3). The ontology defines concepts of casting defect, cause of the defect. It introduces taxonomy of the defects and causes (according to the Polish and French norms), particular instances of the defects and causes and relationships between them. There are also ontologies representing various technological processes of the casting technology. A technological process is described in a form of the graph with nodes representing technological operations, its description, parameters, input and output materials, sub-products etc. Both ontologies are merged to show the relationships between technological operations and possible casting defects.

There are several independently developed and managed data sources containing information on the subject. Actually system integrates the following ones:

- databases managed by Foundry Research Institute in Krakow:
  - SINTE – database of publications on casting technology,
  - NORCAST – database of norms,
- sets of various documents containing description and analysis of casting defects,
- database of pictures of casting defects.

Each data source is integrated into the system by its data source ontology which is merged with the concepts of the domain ontology. This allows generation of the instances representing objects coming from the data source and extension of the graphs representing the domain ontology. A user can explore the domain navigating through the graph defined by the ontology (following links representing relations between particular domain concepts and their instances) and access the documents coming from integrated data sources that describe particular concepts finally.

OntoGRator working in the field of the casting technology is available in the internet (http://lab.iisg.agh.edu.pl/~ontogrator/).

6. SUMMARY

The paper concerns the problem of integration and then access to various data sources containing knowledge about a given domain. The presented solution is based on a formal definition of the domain in a form of ontology. The ontology specifies the concepts and relations that appear in the domain and creates a general schema for the integrated data. As a result a user gets possibility of the "ontology driven" access to the knowledge about the inspected domain. The system allows navigation, searching, querying and reasoning about the concepts and relations from the domain and finally access the integrated data sources.

Experiments with the OntoGrator about casting technology show that the system is characterized by the great potential an expressive power. It is also worthwhile to mention that tools of the OntoGrator kind can provide information services of the quite new quality.

REFERENCES

Baader, F., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., editors, 2003, *THE DESCRIPTION LOGIC HANDBOOK: Theory, implementation, and applications*. Cambridge University Press.

Dobrowolski, G., Marcjan, R., Nawarecki, E., Kluska-Nawarecka, S., Dziaduś, J., 2003, Development of INFOCAST: Information system for foundry industry. *TASK Quarterly*, 7 (2), 283-289.

Dobrowolski, G., Kisiel-Dorohinicki, M., Nawarecki, E., 2004, Agent technology in information systems with explicit knowledge. In Bubnicki, Z., Grzech, A., editors, *Proc. 15th Int. Conf. on Systems Science*, volume III, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 105-112.

Guarino, N., 1998, Formal ontology and information systems. In Guarino, N., editor, *Formal Ontology in Information Systems. Proceedings of FOIS'98*, Amsterdam. IOS Press, 3-15.

Lepratti, L., Berger, U., 2004, Ontology-based approach for the future production of European Industry, in *Management and Control of Production and Logistics 2004*, ed., Lefranc, G., Santiago de Chile. IFAC/IEEE/ACCA, Universidad de Las Americas, 123-128.

Mackowiak K., 2006, *Ontologies in the integration of heterogeneous data sources*, Master Thesis, KI AGH, Kraków.

McBride B., 2002, Jena: A semantic Web toolkit. *IEEE Internet Computing*, 6(6), 55-59.

Staab S., Studer, R., editors, 2004, *Handbook on Ontologies*, Springer.

Volz R., Oberle D., Steffen S., Motik B., 2003, KAON SERVER - A Semantic Web Management System, *Proc. 12th Int. World Wide Web Conference*, Budapest, ACM, 20-24.

ONTOGRATOR – INTELIGENTNY SYSTEM UDOSTĘPNIANIA WIEDZY W ZAKRESIE TECHNOLOGII ODLEWNICZYCH

Streszczenie

OntoGRator jest systemem do integracji i udostępniania wiedzy oraz danych przechowywanych w niezależnych, zewnętrznych, skomputeryzowanych źródłach. Źródłami danych mogą być zasoby zgromadzone w relacyjnych bazach danych, zbiory dokumentów, materiały o charakterze multimedialnym itp. Źródłami wiedzy mogą być też ontologie opisujące fragmenty wiedzy dziedzinowej z danego obszaru (np. odlewnictwa, metalurgii itp.). Integracja źródeł wiedzy realizowana jest przez zdefiniowanie dwóch grup ontologii: „ontologii dziedzinowych" opisujących pojęcia z danej dziedziny oraz zależności między tymi pojęciami, oraz „ontologii źródeł danych" opisujących poszczególne źródła danych udostępnianych przez system. Korzystanie z wiedzy zgromadzonej w systemie polega na nawigowaniu po pojęciach ontologii dziedzinowej oraz po - zidentyfikowaniu interesującego pojęcia - na uzyskaniu informacji na tak wybrany temat (pochodzącej z różnych źródeł danych).

Artykuł prezentuje kwestie związane z architekturą i implementacją systemu oraz zastosowanie systemu do integracji i udostępniania wiedzy w zakresie technologii odlewniczych.

Submitted: October 25, 2006
Submitted in a revised form: December 13, 2006
Accepted: December 20, 2006