THE PLATFORM FOR SEMANTIC INTEGRATION AND SHARING TECHNOLOGICAL KNOWLEDGE ON METAL PROCESSING AND CASTING

STANISŁAWA KLUSKA-NAWARECKA¹, EDWARD NAWARECKI², GRZEGORZ DOBROWOLSKI², ARKADIUSZ HARATYM², KRZYSZTOF REGULSKI²*

¹ Foundry Research Institute, Zakopianska 73; Cracow, Poland
² AGH University of Science and Technology, Mickiewicza 30, 30-059 Cracow, Poland
*Corresponding author: regulski@tempus.metal.agh.edu.pl

Abstract

This paper presents the concept of knowledge sharing platform, which uses an ontological model for integration purposes. The platform is expected to serve the needs of the metals processing industry, and its immediate purpose is to build an integrated knowledge base, which will allow the semantic search supported by domain ontology. Semantic search will resolve the difficulties encountered in the class of Information Retrieval Systems associated with polysemy and synonyms, and will make the search for properties (relations), not just the keywords, possible. An open platform model using Semantic Media Wiki in conjunction with the author’s script parsing the domain ontology will be presented.

Key words: knowledge integration, knowledge base, decision support, semantic search, metal processing, casting

1. CONTEXT OF THE RESEARCH WORKS

For many years, at the Foundry Research Institute in Cracow, construction tools for integrated knowledge bases have been developed (Dobrowolski et al., 2007; Górny et al., 2009; Kluska-Nawarecka et al., 2002; Nawarecki et al., 2012; Kluska-Nawarecka et al., 2005). The studies carried out at present are aimed at improving the information retrieval systems (IR systems) in such a way as to make the collection and sharing of documents easier and more functional from the user’s point of view. In 1997, the Institute launched a SINTE database, which is a bibliographic casting database containing abstracts of over 38,000 articles published in various casting journals (American, English, French, German, Czech, Slovenian, Russian, Ukrainian), proceedings of conferences, and R&D works written by the staff of the Foundry Research Institute. Together with NORCAST, CASTSTOP and a CASTEXPERT diagnostic system, the SINTE database forms a part of the INFOCAST system, a decentralised decision-making information system which is intended to support the casting technology both in industry and in scientific and research work (Marcjan et al., 2002). In thus extended knowledge base it becomes increasingly difficult to reach to the information searched. This is particularly true when the user does not know in advance what kind of resources will be of interest to him, that is, whether he is looking for publications on the subject indicated, or for the information on standards and certificates, or for knowledge in the form of rules or guidance on the characteristics of materials and physico-chemical properties.
The knowledge sharing platform should act as an intermediary between the user and heterogeneous sources of knowledge. Using appropriate knowledge formalisms, such as description logic-based ontologies, it aims at integration of resources in such a way that going through a variety of sources is done with a significant benefit to the user, and also in a way transparent to him. The platform aims at knowledge sharing for non-routine tasks that are difficult to predict in the normal course of production. The solutions and implementations presented in the previous section comprise the knowledge modules that can operate independently (as evidenced by the experience), but using them to create an integrated information and decision-making system enriched ontologically would provide additional functionalities and make it easier for the user to use the functionalities already existing.

The platform is to be conceived as an ontological tool for integration of various subsystems in a semantic network, where individual packets of information and data, as well as components of the knowledge transferred in the system are described using metadata in accordance with the shared ontological model, so that they can be explicitly used (shared) by the individual modules and also remain ready for reuse by other computerised systems, as the Platform is an open system.

Recent studies done at the Foundry Research Institute among experts from the world of industry led to the conclusion that the goal should be to create a platform for knowledge sharing that by giving the user an easy-to-use interface would provide him with a steady supply of current information and knowledge in the field of the metalcasting practice, coming not only from the literature, but also from all other sources, such as databases, or knowledge obtained algorithmically from the process data. To achieve this, these sources will need to be properly integrated and ready for processing (at least indexed for easier search). The system should allow the design of a knowledge base in such a way that it is interactive and makes the codification of expert technological knowledge possible. The task of the proposed system will be the semantic integration of the collected data, information and knowledge. The aim will be to provide the end user with a transparent access to the integrated knowledge base based on

2. THE PLATFORM FOR SEMANTIC INTEGRATION AND SHARING KNOWLEDGE ON METAL CASTING – CASTWIKI

An attempt to create an algorithm which will enable searching the material by its physico-chemical or mechanical parameters, taking also into account the possible upgrading process, requires the use of a knowledge model of the treatment processes. The answer and the way to solve this problem, also in this situation, can be the semantic search based on ontological model.
heterogeneous resources. The integration tool will be ontologies.

3. ONTOLOGIES – MODELS IN DESCRIPTION LOGIC – APPLICATION

The Descriptive or Description Logic (DL) is a subset of First Order Logic (FOL), which can be used to represent a domain in a formalised and structured and, at the same time, computer-processable manner. The basic element of representation are unary predicates corresponding to a set of objects and binary predicates mapping relationships between objects. The descriptive logic allows creating definite descriptions that depict the domain using concepts (unary predicates) and roles (binary predicates).

For example, a steel casting with hot cracks in two places, in the DL record may look like this:

\[ \text{Casting} \sqsubseteq \exists \text{made}\, \text{CastSteel} \sqcap (\geq 2 \text{ defective}) \sqcap \forall \text{defective}\, \text{HotCrack} \]

Hence it follows that concepts can be built from atomic ideas (casting, cast steel, HotCrack), atomic roles (made = made of, defective = having a defect), and constructors. Using logic one can create definitions of concepts:

\[ \text{Casting} \equiv \text{Product} \sqcap \exists \text{cast}\, \text{Mould} \]

or axioms:

\[ \text{CastSteel} \sqsubseteq \text{Alloy} ("\text{every CastSteel is Alloy}"") \]

Description logic was created for the ontology, and therefore a very important issue was to create simultaneously such a language that would allow the symbolic language of logic to be written in the form of computer code. Such language for the description logic has proved to be OWL (Web Ontology Language). Creating a kind of shared formal language, ontologies allow integrating a wide variety of distributed sources of knowledge in a given field, overcoming the problem of differences in the systemic, syntactic, and semantic areas which, so far, has been the biggest challenge for computer tools, often preventing a clear identification of the concept, and therefore, finding information related to this concept.

Ontology is not a database schema, but a simplification can be used: for knowledge repositories, ontology is that what the entity diagram is for a database; it is a diagram, a model of a field of knowledge, understandable by both computers and humans.

The Department of Applied Informatics and Modelling at the AGH Department of Metals Engineering and Industrial Computer Science has developed a domain ontology for different cast iron grades and changes in their properties under the influence of treatment presented in figure 1. Directly under the parent class there are 5 main categories: Treatment, Alloying_elements, Carbon_form, Properties, and Cast Iron. All together, accumulate more than 90 general terms, on which the reasoning in CastWiki Platform will be based.

Fig. 1. Symbolic representation with directed graph of a fragment of domain ontology in the field of cast iron.
3.1. OntoGRator System

At the Foundry Research Institute in Cracow, attempts to use ontologies for integration of knowledge in the field of metal processing have already been going on for several years in cooperation with the AGH Department of Computer Science (Dobrowolski et al., 2007; Regulski et al., 2008). The OntoGRator system developed at that time allowed describing in a strict and formal manner the area that the integrated data were related with and also specifying the semantics of integrated resources. Although the OntoGRator system was solving problems of a semantic description of the area which the processing of metals is, it did not enjoy the sympathy of its potential users. The lack of success was due to, among others, a very complex structure of the system, which consisted of the two main subsystems:

- **OntoGRator Engine** - engine that integrates data from multiple heterogeneous sources, including information on the problem area contained in domain ontology. This subsystem provides the data, which it has integrated, in the form of new, expanded by the data from external sources, ontology, operating through the Jena API programming interface,

- **OntoGRator Web** - application in J2EE technology presenting in the form of web pages the integrated ontologies available through the Jena API programming interface.

A user who was not a knowledge engineer might have great difficulties in understanding the operating principle of the system. Without basic knowledge of the ontology, the system was becoming incomprehensible. Additionally, the mere idea of the system assumed its ability to integrate structured knowledge resources, such as databases and documents from servers located in the Internet, which the user had access to (Adrian et al., 2007). However, this assumption turned out to be ahead of its time - the actual databases often did not provide any API, and user had no access to them via the network interface.

Much more functional has proved to be the system that provided the ability to insert the content directly, instead of placing the URL / URI for each resource.

3.2. CastWiki Platform

The example of Wikipedia (http://en.wikipedia.org/) shows that it is possible to create a system in which each user has the ability to edit and add content, and at the same time a high level of quality of the accumulated knowledge is maintained through supervision and control, and the ability of other users to introduce their own amendments. Participation in editing Wikipedia is voluntary and unpaid, and millions of users around the world every day add new definitions and edit the existing ones. Wikipedia's success has inspired software developers to implement industrial systems operating on the same principle, but being the sole repository of a company. Systems operating in this way, known as content management systems (CMS) or idea management systems, inherit from Wikipedia several advantages:

- Wiki tools are a popular source of information and knowledge, which most of internet users have already encountered and become familiar with,

- Wiki keeps the knowledge resources constantly updated through editing, but discussion leading to the development of a final version of the problem is an integral part of the entry in Wiki,

- Wiki technology is as simple as possible, it requires minimal skills to edit and add new content, and is available to everyone,

- Wiki structure provides a description of the concepts in natural language, and at the same time contains a unique URI which is an effective way to identify concepts in the knowledge model.

Thus, a wiki-type tool successfully meets the most important demands of the knowledge management: it allows the codification of knowledge, recording of experience and results of the creation of new (experimental) knowledge by free editing of entries, supports discussion on specific concepts, giving the opportunity to generate the phenomenon of externalisation of knowledge, and by maintaining the history of discussion on a given topic can also be personalised. The ability to create a "stub article", which is only a draft definition allowing for the extension, too short to serve as a definition of the encyclopaedic nature, but still giving some information about the topic, is a key aspect here. It is precisely in this way, by creating first a short description, incomplete and uncertain, that we allow the discussion to be started on a given topic. Other
users can participate in determining the definition of the concept, adding fragments of the description. Such a scheme of action allows the collective creation of knowledge resources, so-called, shared conceptualisation. Creating the "stub articles" is a voluntary activity, the aim of which is to liberate the externalisation of knowledge by encouraging discussion on a given topic.

However, Wikipedia as a public system is not an acceptable solution for companies that need to restrict access to their knowledge. Knowledge in industrial plants is valuable, but also highly specialised. That is why it was necessary to create a separate platform, using standard Wiki specifically for the needs of foundry plants. MediaWiki software was applied as a platform used by Wikipedia and made available under the GPL licence.

The system called CastWiki schematically presented in figure 2 is designed to provide a platform for the exchange of knowledge and saving the casting knowledge by specialists in the field of metal processing. Wiki mechanism allows the inclusion of such types of content as:
- descriptions in natural language,
- graphic files,
- hypertext links to other concepts in CastWiki,
- links to all the resources available in the network (documents, catalogues, images, photographs, animations, databases) and having its own URL.

In this way, it allows the integration of knowledge already stored in digital form, being a component of other knowledge systems (e.g. INFOCAST, CastExpert + etc).

The integration of these data and knowledge resources (as well as those added during the use of sources) consists in describing the resources by terms (concepts) of ontology, then mapping their structure to the underlying ontology components. For each class in the ontology, a description in natural language can be added. Ontology editing tools such as Protégé permit the placement of descriptions in text form directly in the description of the OWL ontology. They also allow user to place references in the form of a URL. This gives the possibility to transfer the unstructured knowledge, which the definitions in natural language are, and photographs to CastWiki knowledge base. Each concept (article) in CastWiki acquires its unique URI / IRI, which can also serve as a reference to the class description in ontology.

The problem in Wiki-class systems is page duplication and redundancy of knowledge. Co-creation by multiple users makes the situation when for the same substantive term there are several articles under various entries (e.g., L200HNM cast steel, which is also G200CrMoNi4-3-3 cast iron). In this situation, it is necessary to integrate duplicate articles and create redirection of individual entries to the integrated article. Ontology facilitates the analysis of overlapping terms through the rdf:SeeAlso property, which allows placing adequate terms directly in ontological description, thus greatly facilitating the work of CastWiki editors.

Another problem that is solved due to this structure of the knowledge model is the problem of homonyms. Concepts with the same name require the creation of an additional article in CastWiki, which will be a list of words that share the same spelling and pronunciation with a short note about the context of each word. Ontology also solves the problem of homonyms: the model itself cannot have two classes with the same name, which forces the
ontology engineer to extend the class name in such a way as to give the context in accordance with the namespace.

Ontology also provides the ability to create a more structured knowledge base than the traditional wiki approach. Namely, Wikipedia does not allow the definition of relationships. The concept, which is a non-autonomous object, cannot have a representation in the form of a Wikipedia entry. Creating one’s own CastWiki system is a way to avoid this limitation. Users gain the ability to create descriptions not only for classes and instances in the ontology, but also for relationships (object properties).

For example, in an ontology there is a relationship preventive_means, for which the basic Wiki version has no place in a separate article. Therefore, knowledge about how to prevent the casting defect would have to be included in the description of a particular type of defect. CastWiki allows us to create an article integrated with the preventive_means relationship so that the user can easily create a document that contains basic information about how to prevent defects, while collecting all the known resources of knowledge on this subject (including specification of defects which are related in the article, or links to specific procedures to prevent defects). This form of knowledge collection provides hypertext structure of the system and ease of navigation across the resources offered to the user who needs no preliminary knowledge of the conceptual model of the system. At the same time, the user navigating across the ontology can easily find a definite description of relationship, not just its member classes.

3.3. Semantic CastWiki

Semantic MediaWiki (SMW) is a complex semantic extension of the MediaWiki platform (a free Wiki-type solution licensed by open source, developed by the Wikimedia Foundation, which is a basis for most of the projects such as Wikipedia, Wiktionary and Wikinews.) by the mechanisms to improve extraction, search, valuation, marking and publishing of Wiki content. It also provides a platform for software development, which makes it the fastest growing project of this type in the network. In the present work, this application provides the basic mechanisms for the performance of semantic issues.

Semantic annotations that have been developed for SMW are designed in such a way as to enable a faithful export of ontologies in OWL DL format. It is worth mentioning that the SMW user interface does not require a formal interpretation of OWL DL, nor does it impose restrictions on expressiveness. OWL DL ontology structures can be divided into instances that represent individual elements of a particular domain, classes which are aggregates of instances with the same characteristics, and attributes describing logical relationships between instances. The way in which SMW represents knowledge was partly inspired by solutions such as Web Ontology Language, which allows performing an easy transcription from one format to another. From a technical point of view, the MediaWiki platform uses the namespaces to group the pages by content. This mechanism was also used in the clustering of ontology elements:

– OWL individuals - these elements are represented as regular articles. Pages of this type account for a significant portion of the data contained in Wiki. Usually they are grouped in the main namespace, but can be stored in some other spaces, too (People space, Image space).

– OWL classes - they have a counterpart in the basic mechanisms of MediaWiki as a category. The category system, which has been an integral part of the MediaWiki platform since 2004, quickly became the main tool for the classification of documents in Wikipedia and other Wikimedia Foundation projects. Category pages are grouped in the namespace of the same name. They can be organised hierarchically in a similar way as it is done in OWL ontologies.

– OWL properties - the relationships between ontology elements have no counterpart in the MediaWiki engine, and are supplied with the extension of SMW. OWL distinguishes the relationships between data (assigning numerical value to ontological element) and between objects (the relation of two ontological elements). Semantic MediaWiki simplifies this division by aggregating all types of relationships in the namespace called Property.

In order to easily browse the semantic annotations found on the Wiki page, a factbox is used. It allows users to view the most important facts about the subject. For those who are supporting and complementing Wiki, it is also a tool to validate the correctness of the Wiki engine "reasoning" in respect to the annotations introduced earlier. The in-
formation is displayed in two columns: the first, starting from the left, contains attributes used on the page (e.g. the population), while the second one stores the values assigned to them (e.g. 340,000). Each attribute name is also a link to its site, where one can usually find basic information about this attribute (meaning, use). Depending on the design of the attribute, the fact table may include, for example, its value in different units of measurement.

Next to each attribute there is a special magnifying glass icon which is a link to a simple semantic search engine. For example, if a web page is labelled [[is an alloy::Iron]], the user will receive a list of all the sites that meet this requirement. Below the list there is a form in which one can specify any desired attribute-value pair. If the attribute value is numeric, the search engine can also provide pages with an approximation of this value. In the header of the fact table there is an eye icon, which allows quick browsing of all the semantic annotations as presented in figure 3.

3.4. Parsing of domain ontologies – implementation of the script

The recommended method of entry of the ontology into the Wiki structure is by creating one’s own Wiki parsing script. SMW initially circulated such a possibility, but a multitude of ontological formats and change in the approach of the authors of the application to the construction of semantic structures in Wiki caused giving up the idea of a development of this tool (in most of the recently launched scenes it has been removed completely). Author’s script, in turn, allows a relatively easy optimisation, depending on the user needs. Taking into account the available libraries working with MediaWiki it also gives the possibility of parsing most of the popular ontology formats.

The script applied in SemaWiki to load the ontology was implemented using the Python Wikipedia (pywikipedia) programming platform. It is a set of tools to automate the work on the pages of MediaWiki and other popular Wiki engines using web crawler. From the point of view of the Wiki platform, robot is a normal user with specific access rights.
The first step is to import the necessary libraries that allow us to edit the ontology while preserving the logic graph. For this task, a rdflib package is used. Then are defined namespaces used in parses ontology. After verifying that the robot correctly loaded the ontology and logged on to the designated Wiki, actual parsing begins. First in line is the class structure. An algorithm in each iteration of the loop finds a subject-object pair joined with verb such as rdfs: subClassOf. In this way, no class shall be neglected. It is worth noting that during transcription of ontology, no reasoning is carried out, i.e. if a relationship is not defined directly, it shall not be reflected in the ontology.

Each resulting subject-object pair will be represented in CastWiki as a category page, so its name must begin with the keyword "Category". Additionally, to obtain the required transparency, from each URI, the natural class name is extracted. The resulting pages are tested under four conditions. The first checks if parent category is not a class Thing (highest class in OWL, all classes declared in any ontology are subordinate to the Class Thing). If the condition is not met, the child class becomes the parent class.

Entering attributes to Wiki structure is done in the same way. The algorithm takes subject-object pairs combined by a relationship: rdfs: subProper-
ytyOf. Such a condition, however, does not guarantee the extraction of all classes from the ontology. This is due to the fact that, in contrast to classes, attributes are not grouped under one and the same parent attribute.

3.5. Semantic search

With ready-to-action ontology, the user can start adding pages. This process is not much different from the completion of a database content in simple Wiki. The easiest way is to enter into a standard MediaWiki search engine the name of a specific term. If it has not been already included in the base, the application will ask whether to create a page to an earlier question.

Simple search by attributes, with the continuously expanding knowledge bases of a Wiki type, would be insufficient in the long run. Therefore CastWiki also provides the ability of search based on formal questions. For this purpose, a special syntax has been designed, similar to the solutions used in the same tags. For example, the query [[has alloying element :: Nickel]] will generate all pages, for which the attribute "has alloying element" assumes the value "Nickel". Of course, the introduced phrases can be much more advanced. The syntax allows creating queries based on the logic of sets such as [[Category: Cast Iron]] [[has alloying element ::! Nickel]], which will generate a list of all pages in the category "Cast Iron" with value "nickel" for the attribute "has alloying element". In the case of attributes taking numerical value, there is the additional possibility of declaring the search ranges ([[has content of C :: > 0.1%]] [[has content of C :: <2%]]).

4. SUMMARY

The designed platform is a complete, functional tool that allows for the creation in an enterprise of new channels of communication and knowledge transfer. Such a system can be built at minimum cost - the cost is actually just the time. The platform provides employees with complete information about all the resources of knowledge that are available in the organisation, can also easily share new resources and integrate the ones already existing, but still not catalogued. The implementation of such a system can prevent employees from repeating the same job many times, give easy access to proven best practices that exist in the company, and facilitate the development and transfer of knowledge.

Thus shaped system has a huge advantage over the dedicated systems with a ready knowledge base. It is above all much cheaper. CastWiki must be extended by the staff, which takes the time, but it is cheap and easy to use. Every employee can participate in the development of a knowledge base, which allows taking full advantage of the knowledge accumulated in the company. At the same time it is possible in the process of implementing CastWiki to fill a basic knowledge base with the resources accumulated previously, or with information from other purchased systems.

The proposed tools - ontologies - can significantly affect the competitiveness of the casting plants, support knowledge management and reduce the barriers of entry for companies wishing to expand the range of products. Implementation of integrated knowledge management systems, as well as decision support systems requires long-term investments, especially experts’ time, but in the long run it could decide about the survival and competitiveness of an industrial plant.
solutions proposed in this article greatly improve the search process and data distribution. They are justified by the relatively simple solutions that do not require a lot of time to assimilate, mainly owing to the fact that they are based on commonly used technologies. A few years ago, the main problem associated with the implementation of similar techniques was little interest from serious investors. Today, many multinational companies driving development of information technology (Google, Apple) and most popular social networking sites (Facebook, last.fm) successfully use their own semantic solutions. Within the last few years, several ontology-based Wiki platforms have been created, which were used as knowledge bases not only in IT-related companies. Among them, a foundry plant could find its place without any major obstacles. As regards conversion into semantic knowledge, information used in foundry practice is in no way different from other data.

Acknowledgement. Scientific work financed from funds for the scientific research as an international project. Decision No. 820/N-Czechy/2010/0 and 0R0B0008 01

REFERENCES


Ciszewski, S., Kluska-Nawarecka, S., Document driven ontological engineering with applications in casting defects diagnostic, Computer Methods in Materials Science: quarterly / Akademia Górniczo-Hutnicza. 2004 t. 4 nr 1-2 s. 56-64.

Dobrowolski, G., Kluska-Nawarecka, S., Marcjan, R., Nawarecki, E., OntoGRator – an intelligent access to heterogeneous knowledge sources about casting technology, Computer Methods in Materials Science: quarterly / Akademia Górniczo-Hutnicza; ISSN 1641-8581 - 2007 vol. 7 no. 2 s. 324-328.


