Processing of Wiki Resource Semantics on Base of Ontological Analysis

Julia Rogushina
Institute of Software Systems of National Academy of Sciences of Ukraine
Kyiv, Ukraine
ladamandraka2010@gmail.com

***Abstract*—The work objects the development of ontology-based means for processing of semantically matched Wiki resources. Such means allows more relevant representation of domain-specific information objects corresponded with Wiki pages. Proposed approach to semantic processing of Wiki resources is based on methods of ontological analyses of complex information objects. Ontological analyses provides matching of Wiki resource with ontology that allows to use various powerful mechanisms of knowledge management. Main results: methodological recommendations for development of hierarchy of categories and semantic properties of Wiki resource; their use for efficient search and navigation; formal representation of semantic elements of Wiki and their matching with relevant domain ontology. The results of proposed research work are used for design of the electronic version of the Big Ukrainian Encyclopedia (e-BUE) on technological platform of the Semantic MediaWiki. Processing of the semantic properties and categories of Wiki provides the e-BUE transformation into the distributed knowledge base with extended functionality**

***Keywords*—semantic Wiki, ontology, semantic properties**

I. Introduction

Now the use of the Web distributed knowledge determs the effectiveness of information systems. Therefore, the semantization of the Web information resources (IRs) that provides the possibility of their automated processing at the knowledge level is a topical issue of modern information technology. This is caused by great amount of such information as well as the complexity of its structure.

The creation of modern encyclopedic information sources, which includes the electronic version of the Big Ukrainian Encyclopedia (e-BUE), should be focused on intelligent technologies of information processing: the possibility of automated acquisition of relevant knowledge from the IR by other Web applications will make this resource really useful.

II. Formulation of the problem

The transformation of the Wiki resource into the distributed knowledge base requires the definition of the basic principles for page classification and the development of the system of categories and semantic properties that allow formalizing and integrating the knowledge contained on individual Wiki pages. Due to the fact that built-in semantic tools of Semantic MediaWiki do not support the analysis and visualization of this knowledge system, it is advisable to use external knowledge management tools based on ontological analysis. Therefore we propose to develop the rules for comparing the semantic markup of Wiki pages with elements of ontologies and the methods of their use.

III. Semantic MediaWiki

The Wiki technologies [1] provide distributed information processing on the Web and allows users to freely edit the content of the pages. Today, there is a large number of semantic extensions of this technology. One of them is Semantic MediaWiki (SMW) which turns MediaWiki into semantic resource by allowing automatic integration of information from various Wiki pages and supporting of complex semantic queries that process semantic properties and categories of information objects [2].

SMW has a fairly high expressive power, reliable implementation and user-friendly interface, and IRs on it‘s base are dynamically updated and rapidly expanding by the entire community of users, have a well-defined and easy to understand structure and provide information processing at the semantic level. Therefore SMW can be used as a technological platform for group knowledge management. It should be noted that now Semantic MediaWiki is a base for increasing number of sites, portals and encyclopedias. Such interest to semantization of Wiki technology causes the development of tools for intelligent processing of Wiki IRs.

Built-in tools of SMW allows to build ontologies represented in the OWL language but these tools are not flexible enough to generate the ontology relevant to needs of external application. Therefore we propose to develop advanced methods for creating and replenishing of ontologies based on the Wiki-resources, as well as comparing of ontologies created and improved on base of different variants of the Wiki-resources.

A. Categories in SMW

Categorization of Wiki pages is a convenient tool for classification of knowledge represented by Wiki resource. Wiki allows to use an arbitrary number of classifiers for every page. The main purposes of classifiers for Wiki resource: 1. to combine a group of pages with common properties or semantics;2. to retrieve the pages by classifiers.

Each page of the Wiki resource can be assigned to one or more categories of any classifier (depending on the classifier specifics). There is very important for realization of various encyclopedic IR on base of Wiki because the encyclopedia

implies the co-operation and coexistence of different points of view without competitive extrusion where this is not required.

So the system of categories in any area is not obliged to be an attribute of only one classification tree, but may be a system of tree-shaped structures that are superimposed one on another.

That is, the same category or Wiki page can be classified into categories from different classifications, and categorizing of the page within some classification system is not important for it‘s categorized within another classification systems. For example, a Wiki page that describes organization can be classified by the type of institution (category "the National Academy of Sciences"), by its location (category "organization of Kiev"), by the number of employees (category "big organization").

Wiki-based encyclopedias can use either standard classifiers, such as UDC, or systems of categories proposed by IR developers. For example, the e-BUE uses the classification of pages by scientific directions and the set of categories-tags (common and specific for sciences).

B. Semantic properties in SMW

Semantic properties provide data binding to Wiki pages. Properties, in contrast to the Wiki categories, have not only a name but also a value. Each property has a type, name, and value; in addition, it has separate Wiki page in a special namespace that allows to specify a property type, define it’s position in the property hierarchy, and also documents its usage.

Semantic properties in SMW have values of various types – either common datatypes (Boolean, Number, Text etc.) or specific for Wiki (Page, Geographic coordinate, URL etc.). Properties and types are the basic way of entering semantic data.

IV. Types of information objects as a search component in Wiki resources

Information object (IO) is an information model of an object of a certain subject domain that defines the structure, attributes, limitation of integrity and, possibly, the behavior of this object [3]. In case of Wiki IRs IO is the content of some Wiki page together with it’s semantic matching. For example, every entry of e-BUE is a description of some IO.

Categories and semantic properties can be used as an instrument of representation of information objects: categories of encyclopaedias and guides based on Wiki group pages that have similar structure and the same semantic properties. Encyclopaedia developers can create a prototype pages with a unified content placement which simplify the information perception by users. The group of Wiki pages based on the same prototype page is named the type of IO.

For e-BUE, the type of IO is a group of e-BUE entries that has common semantic properties.

The presence of various IO types greatly improves the quality of the search and navigation into e-BUE, but developers of this IO system have to do it systematically. They have to define correctly (according to domain specifics) name spaces, hierarchy and other relations between categories, range of values and range of definition of semantic properties etc. It is advisable to use two approaches for creation of IO types – top-down and bottom-up.

Top-down approach – from top-level categories of IO to their subcategories that more clearly characterize the content. For example, at first the category “Person” is defined with the set of it‘s semantic properties, subsections and categories. On the next step the category “Scientist” is defined as a subcategory of “Person”. This subcategory inherits all properties, subsections and categories of the supercategory but can obtain additional –properties (such as “academic degree” or “publications”).

The bottom-up approach – from Wiki pages to IO types. In process of creating of new articles or reworking of existing ones expert tries to choose some similar pages and then defines the type of IO for them. For example, under the e-BUE article "Austro-Prussian War of 1866" expert can create IO type "War", and under the article "Abstract" – IO type "Document". Then these types can also be used for classification of other articles.

V. Search and navigation in e-BUE

One of the most important factors in the use of modern distributed knowledge bases is the organization of search and navigation, which allows the user to receive the relevant information [4]. In this case, both the access time and the clarity and convenience of the user interface are important.

Search tools can be subdivided into the following main groups: 1. by keywords (for e-BUE – by the article’s title or by the initial letters of the title); 2. by domain topics (for e-BUE – by knowledge directions); 3. by type of information object (for e-BUE – the search for personalities and concepts and by other IO types); 4. by IO semantics (for e-BUE – by semantic queries with categories and values of the semantic properties of Wiki-pages). The first one provides the fastest access to information, the second is based on commonly used classifiers (they allow to formalize the classification, but complicate the search for ordinary users), and the third one takes into account the semantics of the common user’s information needs and allows to obtain admittance to simular articles.

The fourth type of search is a semantic search. It results the information object with a complex structure that is used for search procedures. It can be considered as a special case of object recognition problem. It‘s implementationin e-BUE requires the installation of semantic plug-ins designed to handle semantic properties of articles.

VI. Components of semantic Wiki-resource

The formal model of any Wiki-resource consists of the following elements:

- the set of Wiki-pages P that includes:
  - the set of pages created by users;
  - the set of pages describing categories;
  - the set of pages describing templates;
  - the set of other special pages;

- the one-element set L that describes the links from one Wiki page of this resource to another Wiki page of this resource (although Wiki-resources provide links to other types of pages, they are not included in this model).

The formal model of semantically enriched Wiki-resources is more complex and includes a number of elements related to semantic properties: the set P is supplemented by the set of the Wiki pages that describe semantic properties, and the set L is supplemented by the set of links deal with semantic properties with type “Page”.

VII. Ontologies as a means of distributed knowledge representation

Ontologies are widely used to represent knowledge in the web-oriented information systems because they provide reuse of knowledge in different applications. It causes the great interest to methods and tools of ontological analysis [5].
One of the important areas of such research is the integration of ontology with other information resources of the Web, namely, with semantic Wiki resources. Such properties of the Web as heterogeneity and dynamism create a number of problems associated with the replenishment, use and evaluation of ontologies, and cause the need for more dynamic semantic Wiki resources that can be used to update these ontologies.

Processing of large amounts of knowledge (for example, into e-BUE with tens of thousands of pages) requires to automate this work on base of the approaches, methods and tools currently used for knowledge management. Creating of consistent, complete and understandable system of the IO types is extremely difficult even for domain experts that haven’t the skills and experience of knowledge engineer. Development of system with a large number and complex structure of IO types is implemented much more efficient with the help of methods and tools of ontological analysis: main elements of semantic markup are considered as classes, instances, and attitudes of the ontology of the corresponding domain.

In order to use such ontological tools, it is necessary to develop a method for transforming ontology elements into Semantic MediaWiki (categories and semantic properties). The next stage deals with refining the initial domain ontology by analysis of semantically-tagged Wiki-resources that requires a method for transforming Semantic MediaWiki constructs into ontological representation in OWL. Tools of ontological analysis enable to evaluate properties of this ontology and it’s pertinentness to user’s conceptions about domain. Reiteration of these actions should ensure the formation of an adequate domain ontology that can be used in different applications as a knowledge base.

VIII. Formal model of ontology

In the general case, the formal model of domain ontology is an ordered triple O = <T, R, F>, where T - the set of domain concepts; R – the set of relations between them; F – the set of interpretation functions for of concepts and relations.

This formal model can be specified in different ways depending on the purpose and scope of the ontology purpose and domain [6].

Now the Web-oriented information systems use most often various dialects of the OWL (Web Ontology Language) language for ontology representation [7]. OWL is one of components of Semantic Web [8]. Ontology in OWL is a sequence of axioms and facts, as well as references to other ontologies. OWL enables to obtain by logical inference of facts that are not represented directly in ontology but specified by its formal semantics. There are a lot of approaches to ontology formalization [9]. For processing semantics of the Wiki resources we should use formal model of ontology that consists of the following elements:

- X – the set of ontology concepts that joinsthe set of classes and the set of class instances;

- R - the set of relations between elements of ontology that joins: - hierarchical relation between ontology classes (with such properties as antisymmetry and transitivity);- the set of object properties that define relationships between class instances; the set of data properties that define relationships between class instances and values of these properties;

- F – the set of properties of ontology classes, class instances and their properties that can be used for logical output (for example, equivalence, difference);

- T – the set of data types (for example, string, integer).

The choice of such ontology model causes by the following reasons. Firstly, it quite enough corresponds to the intuitive representation of the ontology laid in the user interface of the broadly used tool for ontological analysis Protege [10]. Secondly, this model integrates quite easily with various ontology-based applications. Thirdly, this model allows to compare the ontological representation of domain knowledge with semantic constructions of Semantic MediaWiki.

Now ontologies are widely used for semantic markup of various information resources [11]. The problem of matching ontologies and semantic Wiki resources arises in several cases. First, development of semantic Wiki resources requires to create the set of categories and semantic properties. But Semantic MediaWiki’s built-in tools do not allow either to visualize this information or to assess its integrity and consistency. Therefore, there is advisable to construct the ontology for domain of Wiki resource and then use this ontology as the basis for semantic markup. Secondly, Wiki resources are more dynamic in comparison with ontologies because many users can participate in their development and improvement, and therefore they can be useful for improving the corresponding domain ontologies by actual information [12].

IX. Matching of ontologies and semantic Wikis

Some correspondences between the elements of domain ontology and the Semantic MediaWiki pages are one-to-one and can be detected automatically, some of which require additional refinements by the user but can also be partially automated. Matching is one-to-one to such elements:
- Link to another Wiki page with The "Page" object property;

- Semantic property of type "page" with Object property;
- Semantic property of any other type with Value of data. Some Wiki elements can be tansformed into ontological elements:
- Category into Class;
- Category Hierarchy into Class Hierarchy;
- Wiki page into Individual of class;
- Template into Class. But domain expert has to define what element of Wiki resource would be built for class of ontology – Category or Template: either Template or Category would be developed only for frequently used classes with unified structure, and Category only – for all other ones. For example, e-BUE contains templates and categories for Personality, Author BUE and Organization, and categories for Instrument, Group of persons and Region.

If OWL ontology is already built then it's easy to use it in Semantic MediaWiki. But the reverse process can not be completely automated. Moreover, automatically generating ontology for Semantic MediaWiki will lose the information from OWL ontology that deals with characteristics of classes and properties that have no analogues in the Wiki (in particular, about the equivalence of classes and properties, their non-crossing, their region of value, and definition). At the same time, some part of the Semantic MediaWiki content can not be directly transformed into ontology. For example, the fact that the pages use the same template suggests that these pages describe IOs of the same type. But representation of this fact in ontology requires to create a specific class and link it to the page element. In addition, it is not possible to associate with the ontology class the specific fragment of Wiki page.

X. Conclusions

The proposed approach to the use of semantic properties of Wiki pages to expand the functionality of the information resource can enhance the search and navigation efficiency in such resources by more complete and relevant satisfaction of information needs of users. Matching the elements of Wiki resource with elements of ontology allows, on the one hand, to take advantage of the ontological presentation for knowledge management of information from Wiki pages (for example, to check their consistency, perform logical output or perform complex queries) of various encyclopedias and Wikis. On the other hand, it allows to transform the encyclopedia into a distributed knowledge base that can be used not only by people but also by other intelligent applications. The availability of semantic markup on the basis of the domain ontology greatly extends both the scope of the use of such information resource and the effectiveness of working with them. Use of modern standards of knowledge representation ensures the interoperability of Wiki content.

References

[1] Wagner C. Wiki: A technology for conversational knowledge management and group collaboration *The Communications of the Association for Information Systems*, 2004, V. 13(1), pp. 264-289.

[2] Krotzsch M., Vrandecic D., Volkel M. Semantic MediaWiki. Available at: http://c.unik.no/images/6/6d/SemanticMW.pdf (accessed 2017, Nov).

[3] Rogushina J.V. The Use of Ontological Knowledge for Semantic Search of Complex Information Objects *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, 2017, pp.127-132.

[4] Rogushina J. Use of the Ontological Model for Personification of the Semantic Search *International Journal of Mathematical Sciences and Computing(IJMSC)*, V.2, N.1, 2016, pp.1-15.

[5] Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases. Available at: http://www.cin.ufpe.br/ mtcfa/files/ 10.1.1.35.1743.pdf (accessed 2017, Oct).

[6] Guarino N. Formal Ontology in Information Systems. Formal Ontology in Information Systems. *Proceedings of FOIS'98*, 1998, pp.3-15.

[7] OWL 2 Web Ontology Language Document Overview. W3C. 2009 Available at: http://www.w3.org/TR/owl2-overview/ (accessed 2017, Oct).

[8] Davies J., Fensel D., van Harmelen F. *Towards the Semantic Web: Ontology-driven knowledge management*, John Wiley & Sons Ltd, England, 2002.

[9] Uschold M., Grüninger M. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, 11(2), 1996, pp.93–155.

[10] Tudorache T., Noy N.F., Musen M.A. Collaborative Protege: Enabling Community-based Authoring of Ontologies. *International Semantic Web Conference*, 2008.

[11] Vargas-Vera M., Motta E., Domingue J., Lanzoni M., Stutt A., Ciravegna F. MnM: Ontology driven tool for semantic markup. *Proc. Workshop Semantic Authoring, Annotation & Knowledge Markup (SAAKM 2002)*, 2002, pp.43-47.

[12] Rogushina, J.V. Development of distributed intelligent systems on base of ontological analysis and semantic wiki technologies. *Ontology of designing. 2017; 7(4): 453-472.*, 2017, pp.453-472.

ОБРАБОТКА СЕМАНТИКИ WIKI-РЕСУРСОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО АНАЛИЗА

Рогушина Ю.В.
Институт программных систем НАН Украины,
Киев, Украина

Работа посвящена разработке онтологических средств обработки семантически размеченных Wiki-ресурсов, которые позволяют более релевантно отобразить специфичные для предметной области информационные объекты, представленные на отдельных Wiki-страницам.

Предлагаемый подход к семантической обработке ресурсов Wiki базируется на методах онтологического анализа сложных информационных объектов. Онтологический анализ обеспечивает сопоставление Wiki-ресурса с онтологией, что позволяет применять к ним различные мощные механизмы менеджмента знаниями.

Основные результаты: методологические рекомендации для разработки иерархии категорий и семантических свойств Wiki-ресурса; использование семантической разметки для эффективного поиска и навигации; формальное представление семантических элементов Wiki и их сопоставление с релевантной онтологией предметной области.

Результаты предлагаемых исследований используются для разработки электронной версии Большой украинской энциклопедии (е-БУЕ) на технологической платформе Semantic MediaWiki. Обработка семантических свойств и категорий Wiki обеспечивает преобразование е-БУЕ в распределенную базу знаний с расширенными функциональными возможностями.