

СИСТЕМА КОЛЛЕКТИВНОГО ПОСТРОЕНИЯ ОНТОЛОГИЙ С ОТКРЫТЫМ ЯЗЫКОМ ПРЕДСТАВЛЕНИЯ

Бениаминов Е. М., Лапшин В. А., Перов Д. В.
Российский государственный гуманитарный университет
г. Москва, Миусская площадь д. 6, ГСП-3, 125993, Россия
e-mail: mefrill@yandex.ru

Аннотация — В работе описывается концепция реализации программной системы ЭЗОП, предназначенной для коллективного построения онтологий. Одной из наиболее значимых особенностей данной программной системы является возможность изменения языка представления онтологий пользователями системы и приведение его в соответствие с традициями предметной области. Одна из проблем формализации знаний специалистами предметных областей знаний - это невозможность представления специфических знаний без использования сложившегося в этой области языка. Описываемый в докладе подход призван решить эту проблему, если не в полной мере, то частично, предоставив специалистам средства настройки языка по мере пополнения библиотеки онтологий онтологиями из их области знаний.

I. Введение

В первом десятилетии нового тысячелетия человечество столкнулось с вопросами, во многом схожими с теми, ответы на которые приходилось искать в конце 60-х, начале 70-х годов прошлого века. С экспоненциальным ростом объемов информации, которые пришлось обрабатывать в эпоху научно-технической революции, остро встала проблема автоматизации этой обработки. Эта проблема была решена построением специализированных инструментов, позволяющих применить машины для обработки огромных потоков данных, с которыми люди не в состоянии были справиться.

Сейчас перед человечеством встала подобная проблема, но уже в области обработки знаний. С ростом всемирной сети Интернет и количества знаний, которые стали доступны исследователям во многих областях науки, исследователи уже физически не в состоянии даже прочитать весь объем публикаций, выходящий в многочисленных периодических изданиях по всему миру, не говоря уже об их творческом осмыслении. Таким образом, назрела острая необходимость в инструментах автоматической систематизации знаний и вычленения их наиболее важных элементов. Исследования в направлении обработки знаний машинами и, особенно, в направлении разработки инструментов формального описания знаний, сейчас объединены одним общим термином – *разработка онтологий* [5]. В данной работе авторы постараются привлечь внимание к одному аспекту этой обширной тематики, а именно, к разработке тематических онтологий научными коллективами.

В работе описывается программная система ЭЗОП [6] – инструмент, предоставляющий возможности коллективного создания онтологий, функционирующий в стиле системы Википедия [3]. В отличие от Википедии, статьи которой представляют собой размеченные тексты на естественном языке, основными элементами системы ЭЗОП являются онтологии.

Одной из основных особенностей системы является возможность изменения формального языка, на котором описываются онтологии так, чтобы язык онтологии был приближен языку предметной области

онтологии. В результате тексты онтологии становятся понятны специалистам в данной научной области, которые получают также возможность записывать факты из этой области на привычном языке. При этом внутреннее представление этих онтологий описывается на стандартном языке едином для всех областей знаний и доступно для компьютерного анализа.

II. Описание принципов реализации

В этом разделе описываются принципы, лежащие в основе реализации программной системы ЭЗОП. В соответствии с принципами Веб 2.0, система ЭЗОП обеспечивает удобную среду в Интернет для объединения пользователей системы по интересам, распределению ролей между пользователями и формированию библиотек онтологий самими пользователями на формальном языке, который они формируют в процессе построения библиотеки.

Система ЭЗОП предоставляет возможности коллективной разработки онтологий. Онтология представляет собой фиксацию на формальном открытом языке договоренностей пользователей о том, что как называется в их области, какими свойствами обладает, и каким ограничением удовлетворяет.

Онтология научной группы может уточняться и, как правило, уточняется в процессе формирования. Более того, по мере определения онтологий предметной области научной группы, в этой группе формируется свой язык, на котором выражается данная онтология, и выражаются способы ее использования. Степень формальности такого языка зависит от традиций и особенностей предметной области, лежащей в основе интересов данной группы, а также от желания и возможности использования в данной области компьютерных моделей.

Первоначально пользователям системы предоставляется некоторая базовая онтология ядра системы и язык ядра, обеспечивающие пользователям возможность формировать произвольные онтологии и вводить новые конструкции языка, переопределять старые конструкции – подстраивать язык под конкретную проблемную область.

Далее, пользователи могут объявить любую построенную ими онтологию средой для построения новых онтологий. В этом случае онтология среды и конструкции языка, доступные из среды, становятся доступными для вновь создаваемых онтологий.

В ядре системы имеются возможности для модульного объектного построения новых онтологий. В результате деятельности пользователей строятся библиотеки взаимосвязанных онтологий и словарей языков системы. В системе имеются средства, поддерживающие процессы обсуждения вариантов онтологий, построения и хранения черновиков онтологий, версий онтологий, тестирования и отладки онтологий, а также средства управления публикацией онтологий (доступностью онтологий для различных групп пользователей).

SYSTEM OF COLLECTIVE CONSTRUCTION OF ONTOLOGY WITH EXTENDED LANGUAGE OF REPRESENTATION

Beniaminov E. M., Lapshin V. A., Perov D. V.
Russian State University for Humanities
GSP-3, Moscow, Miuskaya square 6, 125993, Russia
e-mail: mefrill@yandex.ru

Abstract — The paper describes the concept of implementing of a system EZOP for the collaborative construction of ontologies. The main problem in the formalization of knowledge of groups of researchers in the form of ontologies is the need to study the language of representation of ontology. One of the most significant features of the system described in this report is the ability to extend the language by which users interact with the system. The approach described in the report allows experts to create own knowledge of language of representation that reflects their knowledge.

I. Introduction

The paper describes a software system EZOP [6] - a tool that provides the possibility of collective creation of the ontology that operates in the style of Wikipedia [3]. Unlike Wikipedia, where articles are tagged as texts in natural language, the basic elements of the system are EZOP ontology:

One of the main features of the system is the ability to change the formal language for describing ontologies so that the language of the ontology was approached language of a domain ontology. As a result, the texts of the ontology become clear to those skilled in the scientific field, which also receive the opportunity to record the facts of this area at usual language.

II. Description of the Principles of Implementation

EZOP system provides possibilities of collective development of ontologies. Ontology represents a fixation on the formal language of the agreements of users about what and how everything is named, what properties and what constraints they satisfy. It is assumed that formal language for description of ontologies is collectively formed by users as the replenishment system of ontologies and patterns of language, introduced in ontologies:

Ontology of a research group might be specified and is usually specified in the process of formation. Moreover, so far as ontology defined of a specific area of a research group, in this group a specific language is formed, which expresses this ontology and methods of its application. Degree of formality of such language depends on traditions and peculiarities of the domain, the underlying interests of the group, as well as the desire and the possibility of using of computer models in the field.

Initially, users of the system provide some basic ontology kernel and language kernel, providing users the ability to generate arbitrary ontology and to introduce the new language constructs to override the old design - adjust the language for a specific problem area.

Further, users might declare any of constructed ontologies as the environment to construct new ontologies. In this case, the ontology of the medium and language construct, available from the environment, become available to the newly created ontologies.

IV. Conclusion

EZOP system provides opportunities for the collective formation of ontologies in the style of Wikipedia, i.e. with the support of versioning, the module-in particular, providing of testing tools and debugging of ontologies in the process of writing.

Ontologies are formed in the open formal language. The openness of ontology language means that users of the system might modify language of the ontology, specifying the new syntactical and lexical constructions, as well as changing the semantics of existing structures.

Онтологии в системе ЭЗОП представляют собой не просто описание знаний, с каждым текстом онтологии система содержит внутреннее машинное представление этой онтологии. Система предоставляет возможности задания вопросов к онтологии, на которые система пытается найти ответы, используя для этого внутреннее представление онтологии. Язык запросов к онтологии также может расширяться пользователями системы, т.е. в процессе создания онтологии пользователи системы могут формировать также язык диалога с системой в контексте данной онтологии.

Система ЭЗОП основана на алгебраическом подходе к представлению знаний [1]. При этом подходе онтология представляется как алгебраическая теория, в которой знания представляются в виде термов, составленных из операций, имеющим в качестве параметров другие операции, с параметрами или без них (константы).

Система предоставляет возможности по расширению языка онтологии с помощью языковых шаблонов [2]. Каждый языковой шаблон может быть автоматически трансформирован в правило контекстно-свободной порождающей грамматики Хомского [4]. Таким образом, язык онтологии синтаксически представляет собой контекстно-свободный язык. Однако, возможность задания вместе с каждым шаблоном условий на его применение, позволяет уточнить синтаксис языка так, чтобы система могла различать более тонкие синтаксические конструкции.

IV. Заключение

В данной работе описаны принципы, используемые при разработке системы ЭЗОП, предназначенной для коллективной разработки онтологий с открытым языком представления.

Система ЭЗОП предоставляет возможности по коллективному формированию онтологий в стиле Википедия, т.е. с поддержкой версионности, модульности, предоставлением инструментов тестирования и отладки онтологий в процессе их написания.

Тексты онтологий формируются на открытом формальном языке. Открытость языка онтологии означает, что пользователи системы могут модифицировать язык онтологии, задавая новые синтаксические и лексические конструкции, а также изменяя семантику уже имеющихся конструкций.

В данный момент реализован рабочий прототип системы, который доступен на сайте проекта [6].

Работа над проектом ЭЗОП ведется по гранту РФФИ 09-07-00079-а.

V. Список литературы

- [1] *Бениаминов Е. М.* Алгебраические методы в теории баз данных и представлении знаний. М.: Научный мир, 2003.
- [2] *Бениаминов Е. М., Манушина М. Ю.* Принципы построения открытого языка шаблонных выражений в системе представления знаний // НТИ Сер. 2, Информ. процессы и системы. 7. 2000.
- [3] Сайт системы Википедия. URL <http://www.wikipedia.org> (дата обращения 12.05.2011).
- [4] *Лапшин В. А.* Лекции по математической лингвистике. М.: Научный мир, 2010.
- [5] *Лапшин В. А.* Онтологии в компьютерных системах. М.: Научный мир, 2010.
- [6] Сайт проекта ЭЗОП. URL <http://www.ezop-project.ru> (дата обращения 12.05.2011).