

УДК 37

Разработка онтологии курса «Экспертные системы и базы знаний»**Мьинт У**

Аспирант,
кафедра программного обеспечения и администрирования
информационных систем,
Курский государственный университет,
305000, Российская Федерация, Курск, ул. Радищева, 33;
e-mail:oomyint261@gmail.com

Аннотация

В статье анализируются особенности разработки онтологии курса «Экспертные системы и базы знаний», приводятся принципы отбора содержания, обосновывается необходимость использования онтологической модели представления информации. Автор рассматривает этапы создания онтологии курса «Экспертные системы и базы знаний» на основе текстов, созданных на естественном языке. Этапы разработки онтологии обозначаются как группа действий, отвечающих за процессы разработки онтологий и их жизненных циклов, за составление методологии их построения, наборов языков и инструментов.

Для цитирования в научных исследованиях

Мьинт У. Разработка онтологии курса «Экспертные системы и базы знаний» // Педагогический журнал. 2017. Т. 7. № 4А. С. 209-216.

Ключевые слова

Электронный образовательный ресурс, онтология, онтологическая модель.

Введение

Процесс построения онтологий является сложным и занимающим много времени. С целью ускорить данный процесс в середине 90-х годов исследователи впервые стали создавать специальную среду для разработки онтологий. В последние годы было создано большое количество инструментов для разработки онтологий. К примеру, сайт W3C предлагает около 50 различных инструментов для редактирования [Абруков и др., 2016]. Этапы разработки онтологии представляется возможным обозначить как группу действий, отвечающих за процессы разработки онтологий и их жизненных циклов, за составление методологии их построения, наборов языков и инструментов.

Особенности разработки онтологии курса «Экспертные системы и базы знаний»

Рассмотрим этапы создания онтологии курса «Экспертные системы и базы знаний» на основе текстов, созданных на естественном языке.

На данный момент тексты на естественном языке являются основным способом хранения и дальнейшей передачи знаний. В связи с все более и более усиливающейся тенденцией хранения текстов в цифровом виде, а также значительным увеличением доступных информационных ресурсов и объемов передаваемой информации становится актуальной проблема автоматизированной обработки информации, в частности задача информационного поиска и машинного анализа текста. К данному типу проблем можно отнести разработку эффективных подходов к систематизации контента (информационного содержимого текста, отражающего предметную область), а также задачу обработки текста с целью выявления его смысла и структурирования информационного содержимого [Грищенко, 2016].

Задача систематизации контента остро встает и в сфере образования в процессе структурирования учебной информации при ее компьютерном представлении. Связано это в первую очередь с тем, что характерной особенностью информационной культуры педагога в настоящий момент является необходимость осмысления большого объема информации, выраженной в различных видах, например, в виде временных диаграмм, логических функций, таблиц, электрических схем, алгоритмов функционирования и микропрограмм. Кроме того, при работе со средствами вычислительной техники имеет место множество разнообразных проявлений человеческих знаний о реальных объектах. Все это предполагает применение новых обучающих технологий, современных методов представления учебной информации, ее использование при логическом выводе.

Введем понятие лингвистической онтологии предметной области «Экспертные системы и базы знаний». Различные подходы к определению онтологии рассматриваются в трудах некоторых исследователей [Загоруйко, 2014], далее онтологией предметной области будем считать:

$$O = \langle X, R, F \rangle_*$$

где X – конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология O ;

R – конечное множество отношений между концептами (понятиями, терминами) заданной предметной области;

F – конечное множество функций интерпретации (аксиоматизации).

Накладывая ограничения $R = \emptyset$ и $F = \emptyset$ онтология O трансформируется в простой словарь:

$$O = V = \langle X, \{ \}, \{ \} \rangle, \text{ или вырожденную онтологию [Сазыкин, Кудряков, 2015].}$$

Видоизменим выражение $*$ следующим образом

$$O = O' = \langle X, Z, R, \{ \}, \{ \} \rangle_{**},$$

где Z не пустое, конечное множество определений терминов из множества X .

Таким образом, лингвистическая онтология – это иерархическая сеть терминов. Каждое понятие связывается отношениями с другими понятиями онтологии.

Определим отношения между понятиями, характерными для лингвистической онтологии.

Первый тип отношений – родовидовое отношение «ниже- выше», обладает свойством транзитивности и наследования. Второй тип отношений – отношение «часть-целое». Используется не только для описания физических частей, но и для других внутренних сущностей понятия, таких как свойства или роли для ситуаций. Еще один тип отношения, называемого несимметричной ассоциацией $асц2$ - $асц1$, связывает два понятия, которые не могут быть связаны выше рассмотренными отношениями, но одно из понятий не существовало бы без другого. Последний тип отношений – симметричная ассоциация, он связывает, например, понятия, очень близкие по смыслу, но которые нельзя склеить в одно понятие.

Отношения «ниже-выше», «часть-целое» и несимметричная ассоциация являются иерархическими отношениями. Таким образом, на основе свойств иерархичности, транзитивности и наследования для каждого понятия может быть определена совокупность понятий, которые являются для него нижестоящими понятиями по иерархии. Построение данного вида онтологии (лингвистической онтологии) подразумевает определение множества

терминов из обрабатываемого текста и сопоставление им соответствующего определения.

Создание онтологии лингвистического типа предполагает необходимость выполнения следующих действий:

- 1) сформировать терминологическую базу для какой-либо предметной области, опираясь на массив текстовой информации;
- 2) проанализировать полученную информацию, чтобы отфильтровать необходимые отношения и указать их определения;
- 3) установить отношения между наборами терминов данной области.

Данные этапы создания онтологии, кроме первого, предполагают работу эксперта (человека).

Итак, в рамках первого этапа необходимо из текста выбрать «кандидатов» для включения во множество X . Заведомо отметим одно важное допущение: при отборе терминов кандидатов мы основываемся на той гипотезе, что словарный запас и частоты использования слов зависят от темы текста. Другими словами, можно предположить, что ключевые слова текста (то есть слова, частота повторения которых в тексте выше других) являются основными и предположительно терминами. Решение использовать статистический метод при поиске ключевых слов в структуре текста обусловлено их простотой и ресурснезависимостью.

Обрабатываются все строки входного массива. В результате имеем таблицу лексем, содержащую все слова из входного потока (за исключением стоп-слов). С точки зрения структур данных наша таблица лексем представляет собой хеш-таблицу. Получая в качестве аргумента некоторое слово, функция выдает в результате некоторое целое число – индекс в таблице лексем, под которым следует хранить это слово. Реализуется это при помощи функции, выдающей по заданной букве ее номер в русском алфавите, и функции, суммирующей коды букв слова [Рыбина, Блохин, 2015].

Чтобы разрешить коллизии, можно использовать линейный список. Для каждого элемента в списке имеется два поля: в строковом содержится сама лексема, а в целочисленном указывается, сколько раз лексема повторялась в тексте входного потока. На следующем этапе выделяются возможные термины за счет поиска родственных слов на основе морфемного подхода в таблице и заменяются на так называемое главное слово. Для этих целей используется еще один проход по хеш-таблицам и использование дополнительной структуры данных (морфологического словаря).

Морфословарь является древовидной сильноветвящейся структурой, все вершины которой являются массивом, содержащим элементы, которые имеют три поля: уникальный номер, поле флагов, а также указатель к вершине следующего уровня.

Дерево строится таким образом, что все элементы пронумерованы, им соответствуют буквы русского алфавита. В таком случае слово является путем в дереве, начинающимся от

корня и идущим к вершине. Структуры полей флагов имеет такие признаки, как окончание слова, признаки основ, признаки главных слов в группах родственных слов.

Процесс поиска морфологически родственных слов сводится к рекурсивному спуску по дереву с целью нахождения элементов вершины, который имеет признак основы. Далее производится обход дерева, а все слова, которые были найдены, добавляют к списку родственных слов.

После разбора текста необходим дополнительный проход по таблице лексем для поиска и замены групп родственных слов соответствующим главным словом, складывая частоту их появления в тексте.

По завершении описанного этапа таблица лексем имеет ряд главных слов для всех групп словродственных, встретившихся в структуре входного потока. Если в тексте имеются слова, которые не присутствуют в тексте морфологического словаря, тогда формируются выходной список слов.

При выходе из системы получим множество X' слов, которые являются «кандидатами», вида:

$$X' = \{x, V\},$$

где: x – слово-«кандидат»;

V – вес «кандидата».

Заключение

Таким образом, этапы разработки онтологии представляют собой группу действий, отвечающих за процессы разработки онтологий и их жизненных циклов, за составление методологии их построения, наборов языков и инструментов. Разработанная выше модель лингвистической онтологии, а также метод и алгоритмы ее создания реализованы в программном виде объектно-ориентированной библиотеке классов на Java. Объектно-ориентированная реализация библиотеки на основе языкового стандарта Java является востребованной и приемлемой в современных условиях разработки программного обеспечения.

Библиография

1. Абдулгалимов Г.Л., Гулюта А.А., Казагачев В.Н. Профессиональная подготовка будущего учителя информатики к преподаванию робототехники // Успехи современной науки и образования. 2016. Т. 2. № 12. С. 29-31.
2. Аbruков В.С. и др. База знаний: эксперимент, интеллектуальный анализ данных,

-
- искусственные нейронные сети // Сборник трудов II Всероссийской научной конференции «Наноструктурированные материалы и преобразовательные устройства для солнечной энергетики». Чебоксары, 2014. С. 15.
3. Грищенко М.А. Применение модельно-управляемого подхода для создания продукционных экспертных систем и баз знаний // Искусственный интеллект и принятие решений. 2016. № 2. С. 16-29.
 4. Данилин С.Н., Макаров М.В., Щаников С.А. Комплексный показатель качества работы нейронных сетей // Информационные технологии. 2013. № 5. С. 57-59.
 5. Загорюлько Ю.А. О формализации семантики областей знаний в информационных и интеллектуальных системах на основе онтологий // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS). Минск, 2014. С. 117.
 6. Кадырова Г.Р., Кирасирова Ю.С. Использование интеллектуальной системы при выборе учебного заведения // Современная наука: актуальные проблемы и пути их решения. 2015. № 4. С. 190-192.
 7. Рыбина Г.В. Экспертные системы и инструментальные средства для их разработки: некоторые итоги // Информационно-измерительные и управляющие системы. 2013. Т. 11. № 5. С. 35-48.
 8. Рыбина Г.В., Блохин Ю.М. Методы и средства интеллектуального планирования: применение для управления процессами построения интегрированных экспертных систем // Искусственный интеллект и принятие решений. 2015. № 1. С. 75-93.
 9. Сазыкин В. Г., Кудряков А.Г. Особенности поддержки решения технических задач с помощью экспертных систем // Путь науки. 2015. № 8. С. 21-23.
 10. Самойлова Е.М., Игнатъев А.А. Моделирование динамической системы автоматизированного токарного модуля при разработке экспертной системы // Модели, системы, сети в экономике, технике, природе и обществе. 2016. № 2. С. 268-277.

Development of the ontology of the course "Expert systems and knowledge bases"

Myint Oo

Postgraduate,

Department of software and administration of information systems,

Kursk State University,

305000, 33 Radishcheva st., Kursk, Russian Federation;

e-mail: oomyint261@gmail.com

Myint Oo

Abstract

The article analyzes the features of development of the ontology of the course "Expert systems and knowledge bases", principles of content selection, the necessity of using an ontological model of information presentation. The process of constructing ontology is difficult. To expedite this process, in the mid 90-ies researchers began to create a special environment for ontology design. In recent years, a large number of tools for ontology development were created. The author considers the stages of creation of an ontology of the course "Expert systems and knowledge base" based on the texts created in a natural language. The stages of ontology development are identified as a group of activities in the field of the processes of ontology development and their life cycles, the methodology of their construction, the sets of languages and tools. Today the natural language texts are the primary means of storage and further transfer of knowledge. In connection with the increasingly growing trend of storing texts in digital form, as well as a significant increase in available information resources and the volume of the transmitted information, the problem of automated information processing becomes actual, in particular the task of information retrieval and machine analysis of the text. The developed model of linguistic ontology, and method and algorithms of its creation is implemented in software object-oriented Java Library. Object-oriented implementation of a library based on Java is a popular and acceptable in modern software development.

For citation

Myint Oo. (2017) Razrabotka ontologii kursa "Ekspertnye sistemy i bazy znaniy" [Development of the ontology of the course "Expert systems and knowledge bases"]. *Pedagogicheskii zhurnal* [Pedagogical Journal], 7 (4A), pp. 209-216.

Keywords

Electronic educational resource, ontology, ontological model.

References

1. Abdulgalimov G.L., Gulyuta A.A., Kazagachev V.N. (2016) Professional'naya podgotovka budushchego uchitelya informatiki k prepodavaniyu robototekhniki [Professional training of future teachers of informatics to teaching robotics]. *Uspekhi sovremennoi nauki i obrazovaniya* [Successes of modern science and education], 12(2), pp. 29-31.
2. Abrukov V.S. et al. (2014) Baza znaniy: eksperiment, intellektual'nyi analiz dannykh, iskusstvennye neironnye seti [Knowledge base: experiment, data mining, artificial neural network]. In: *Sbornik trudov II Vserossiiskoi nauchnoi konferentsii "Nanostrukturirovannye*

-
- materialy i preobrazovatel'nye ustroystva dlya solnechnoi energetiki*" [Proc. All-Russia Conf. "Nanostructured materials and devices for cone energy]. Cheboksary, p. 15.
3. Danilin S.N., Makarov M.V., Shchanikov S.A. (2013) Kompleksnyi pokazatel' kachestva raboty neironnykh setei [Complex indicator of quality of work of neuron networks]. *Informatsionnye tekhnologii* [Information technology], 5, pp. 57-59.
 4. Grishchenko M.A. (2016) Primenenie model'no-upravlyaemogo podkhoda dlya sozdaniya produktsionnykh ekspertnykh sistem i baz znaniy [Application of a model-driven approach for creating rule-based expert systems and knowledge bases]. *Iskusstvennyi intellekt i prinyatie reshenii* [Artificial intelligence and decision-making], 2, pp. 16-29.
 5. Kadyrova G.R., Kirasirova Yu.S. (2015) Ispol'zovanie intellektual'noi sistemy pri vybore uchebnogo zavedeniya [The use of intelligent systems in the choice of educational institution]. *Sovremennaya nauka: aktual'nye problemy i puti ikh resheniya* [Modern science: actual problems and ways of their solution], 4, pp. 190-192.
 6. Rybina G.V. (2013) Ekspertnye sistemy i instrumental'nye sredstva dlya ikh razrabotki: nekotorye itogi [Expert systems and tools for their development: some results]. *Informatsionno-izmeritel'nye i upravlyayushchie sistemy* [Information-measuring and control system], 5(11), pp. 35-48.
 7. Rybina G.V., Blokhin Yu.M. (2015) Metody i sredstva intellektual'nogo planirovaniya: primeneniye dlya upravleniya protsessami postroeniya integrirovannykh ekspertnykh sistem [Methods and tools for the intelligent planning: application to manage processes of construction the integrated expert systems]. *Iskusstvennyi intellekt i prinyatie reshenii* [Artificial intelligence and decision-making], 1, pp. 75-93.
 8. SamoiloVA E.M., Ignat'ev A.A. (2016) Modelirovaniye dinamicheskoi sistemy avtomatizirovannogo tokarnogo modulya pri razrabotke ekspertnoi sistemy [Modeling of dynamic systems of computer-aided lathe module in the development of expert system]. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* [Models, systems, networks in economics, technology, nature and society], 2, pp. 268-277.
 9. Sazykin V G., Kudryakov A.G. (2015) Osobennosti podderzhki resheniya tekhnicheskikh zadach s pomoshch'yu ekspertnykh sistem [The features of support of solving technical problems with expert systems]. *Put' nauki* [Path of science], 8, pp. 21-23.
 10. Zagorul'ko Yu.A. (2014) O formalizatsii semantiki oblastei znaniy v informatsionnykh i intellektual'nykh sistemakh na osnove ontologii [On the formalization of the semantics of the fields of knowledge in the information and intellectual systems on the bases of on ontology]. In: *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem (OSTIS)* [Open semantic technology for intelligent systems (OSTIS)]. Minsk, p. 117.
-