

САССА АНТОН ОЛЕГОВИЧ

**МОДЕЛИ И МЕТОДЫ ИЗВЛЕЧЕНИЯ И АНАЛИЗА ЗНАНИЙ ИЗ
ИНФОРМАЦИОННЫХ РЕСУРСОВ ОБРАЗОВАТЕЛЬНОГО ПОРТАЛА
ВУЗА**

Автореферат

диссертации на соискание академической степени магистра
техники и технологий по специальности
6N0704 – Вычислительная техника и программное обеспечение

Республика Казахстан
г. Усть-Каменогорск, 2011

Работа выполнена в Восточно-Казахстанском государственном техническом университете им. Д. Серикбаева

Научные руководители: кандидат технических наук,
Балова Т.Г.

Официальные оппоненты: к.ф.-м.н., доцент кафедры ММ и
КТ ВКГУ им. С. Аманжолова
Попова Г.В.

Защита состоится 27 января 2011 г. в 09-30 на заседании государственной аттестационной комиссии при Восточно-Казахстанском государственном техническом университете им. Д. Серикбаева по адресу: 070004, г. Усть-Каменогорск, ул. Серикбаева 19, аудитория ГЗ-322

С диссертацией можно ознакомиться в библиотеке Восточно-Казахстанского государственного технического университета им. Д. Серикбаева.

Автореферат разослан « 27 » декабря 2010 г.

Секретарь
ГАК

Денисова Н.Ф.

Общая характеристика работы

Актуальность темы. Для высшего учебного заведения знания являются одним из важнейших ресурсов наравне с финансовыми и материальными. Вуз приобретает несомненные конкурентные преимущества, если быстрее других создает и находит новые знания, обеспечивает их развитие и практическое внедрение, формируя научные и профессиональные компетенции как сотрудников так и выпускников вузов. Знания, интеллектуальный капитал, интеллектуальная собственность получают растущее признание в качестве нового источника развития. В связи с этим инновационные вузы стремятся эффективно распоряжаться, управлять имеющимися у них знаниями.

Актуальными являются извлечение и анализ знаний интернет-ресурсов образовательного портала вуза, осуществление которых возможно с помощью системы онтологий.

Онтологии играют решающую роль в модели описания знания, без которой, как утверждают специалисты, вход в любую предметную область запрещен. Онтологии разрабатываются и могут быть использованы при решении различных задач, в том числе для совместного применения людьми или программными агентами, для возможности накопления и повторного использования знаний в предметной области, для создания моделей и программ, оперирующих онтологиями.

Онтологии могут быть использованы везде, где требуется обработка данных, учитывающая их семантику. В силу изначальной ориентированности языка OWL на машинную обработку, правильное применение онтологий может, с одной стороны, существенно упростить и, с другой стороны, открыть новые возможности в разработке приложений, решающих задачи автоматизированной обработки и доступа к данным. Проектирование онтологий – это творческий процесс, и поэтому потенциальные приложения онтологий, а также понимание разработчиком предметной области и его точка зрения на нее будут, несомненно, влиять на принятие решений.

До настоящего времени в достаточной степени не разработан комплексный набор методов и алгоритмов описания и использования содержащих знания объектов образования с применением онтологических моделей и семантических методов и методик построения на их основе систем управления знаниями.

Одним из современных направлений развития электронного портала вуза является преобразование его информационного пространства в пространство знаний и электронных образовательных ресурсов. В связи с этим, задача извлечения и анализа знаний интернет-ресурсов образовательного портала вуза является актуальной как с научной точки зрения, так и практически значимой, поскольку позволяет применить возможности технологии Semantic Web для создания семантического web-портала системы управления знаниями современного вуза.

Целью исследования является извлечение и анализ электронных образовательных ресурсов из электронного портала вуза. Для достижения поставленной цели необходимо решить следующие **задачи**:

- исследовать возможности онтологического подхода для развития информационной системы образовательного портала вуза и поддержки принятия решений по управлению системой знаний и компетенций вуза;
- создать онтологию электронного образовательного ресурса с целью описания разнородных типов объектов знаний;
- создать онтологическую базу знаний электронного образовательного ресурса, содержащую единую онтологическую модель знаний и множество семантических метаописаний объектов знаний;
- разработать набор методов для реализации основных функций работы со знаниями, таких как объединение, пересечение, классификация.

Предметом исследования являются модели знаний образовательных ресурсов вуза с использованием системы онтологий.

Методы исследования. В ходе диссертационного исследования были использованы онтологический подход и таксономии, методы теории множеств и системный анализ.

Научная новизна исследования и положения, выносимые на защиту:

- онтология электронного образовательного ресурса;
- методы создания и поддержки в актуальном состоянии единой онтологической модели знаний;
- комплексная онтологическая база знаний электронного образовательного ресурса;
- методы реализации оценки семантической близости классов и индивидов, позволяющие классифицировать и идентифицировать классы онтологий.

Практическая значимость исследования – разработанный электронный образовательный ресурс, созданный с использованием системы онтологии позволяет управлять качеством образовательного процесса на основе имеющихся в университете знаний.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на научных конференциях различного уровня:

- X Республиканская научно-техническая конференция аспирантов и молодых преподавателей: «Творчество молодых - инновационному развитию Казахстана», Казахстан, Усть-Каменогорск, ВКГУ им.Д.Серикбаева, 2010 г.

По основным результатам исследований опубликовано 1 работа.

Объем диссертации составляет 104 страниц машинописного текста. В тексте диссертационного исследования содержится 42 рисунка, 4 таблицы, 3 приложения. Список использованной литературы включает в себя 57 наименований источников.

Основное содержание работы

В первом разделе «Анализ использования онтологического подхода к извлечению знаний из электронных образовательных ресурсов» определяется понятие онтологии, её классификация и основные задачи онтологического подхода.

В диссертационном исследовании под онтологией понимается спецификация концептуализации предметной области, зависящая от целей ее создания, или задач, при решении которых планируется ее применять.

Согласно классификации по цели создания в основе онтологической базы знаний диссертационного исследования лежит онтология верхнего уровня. Ее основным назначением является создание единой «правильной» онтологии, фиксирующей знания, общие для предметных областей электронного университета, и в многократном использовании данной онтологии. В соответствии с классификацией по содержанию, онтология электронного портала вуза является предметной, описывающей реальные предметы (ГОСО, дисциплины, учебные планы), участвующие в какой-либо деятельности (учебном процессе).

Использование онтологического подхода в диссертационной работе позволит решить следующие задачи:

- создание и использование баз знаний;
- организация эффективного поиска в информационных и образовательных ресурсах и базах знаний;
- каталогизация и классификация информационно-образовательного ресурса (в том числе неструктурированной информации) путем создания аналитических метаданных;
- извлечение знаний и доступ к любому уровню каталогизации;
- обеспечение общей терминологии для множества специалистов и совместно используемых приложений;
- многократное применение БЗ и информационных массивов.

Применение онтологического подхода существенно повышает ценность корпоративного портала вуза, как и большинства систем управления знаниями, за счет организации и структурирования информационных образовательных ресурсов, развития его поисковых возможностей и интеграции в международную образовательную систему.

Во втором разделе «Онтологическая модель базы знаний электронного образовательного ресурса» в качестве языка описания онтологий выбран язык OWL, который основан на DAML/OIL, имеет широкую теоретическую и практическую поддержку, является стандартом Семантического Веб (Semantic Web). Используется диалект языка OWL DL, который поддерживает дескриптивную логику, предоставляет максимальные возможности для логического вывода новых и заложенных в онтологии знаний.

Формальная модель онтологии является разновидностью сетевой модели знаний и представляется в виде тройки конечных множеств - классов, отношений и свойств.

Для разработки таксономии классов онтологии электронного образовательного ресурса выбран комбинированный метод, который сочетает нисходящий и восходящий подход: сначала мы определяем более заметные понятия, а затем соответствующим образом обобщаем и ограничиваем их.

В работе рассмотрены основные операции над онтологиями:

- оценка семантической близости двух классов и двух индивидов, которые позволяет сравнить соответствие структуры (сравнение классов) и состава (сравнение индивидов) электронных образовательных ресурсов портала вуза;

- интеграция онтологий, которая обеспечивает накопление информации и облегчает достижение интероперабельности взаимодействующих информационных ресурсов, основанных на разных исходных онтологиях.

Совокупность описаний онтологической модели и метаданных всех ресурсов электронного портала вуза составляют онтологическую базу знаний электронного образовательного ресурса, использование которой позволяет реализовать семантическую поддержку процессов работы со знаниями электронного университета.

В **третьем** разделе «Проектирование и анализ базы знаний электронного образовательного ресурса с использованием системы проектирования онтологий Protégé» показана реализация всех этапов жизненного цикла онтологии, а именно построение, заполнение, проверка, использование, поддержка и развитие.

В качестве инструмента для построения онтологии и базы знаний выбран редактор онтологии Protégé 4.1, разработанный Knowledge Modeling Group Стэнфордского университета как часть проекта Protégé, выполняемого в течение последних 15 лет и включающего разработку набора инструментов моделирования знаний. Этот инструмент может быть легко расширен с помощью пользовательских плагинов для работы с другими, основанными на знаниях, приложениями. С помощью плагинов OWLViz и OntoGraf мы можем графически представить структуру нашей онтологии.

Для реализации этапов построения онтологической базы знаний электронного образовательного ресурса необходимыми и достаточными являются знания структуры и задач онтологического подхода, а также использование таких методов исследования, как системный анализ и теория множеств.

В данной диссертационной работе используется Reasoner Pellet, который является открытым исходным кодом Java основанным на OWL DL рассуждениях. Pellet осуществляет классификацию классов, проверку согласованности и правильности иерархии классов, а также дает возможность исправления и изменения онтологии.

Для выполнения операций над онтологиями, таких как сравнение и объединение, используется OWLDiff Protege Plugin и язык запросов DL Query. В результате осуществления данных действий, разработанная онтологическая база знаний позволяет контролировать качество образовательного процесса.

Заключение

В ходе диссертационного исследования были выполнены извлечение и анализ электронных образовательных ресурсов из электронного портала вуза.

Решение поставленных в ходе исследования задач позволило достичь следующего результата: были исследованы возможности онтологического подхода для развития информационной системы образовательного портала вуза и поддержки принятия решений по управлению системой знаний и компетенций вуза; создана онтология электронного образовательного ресурса с целью описания разнородных типов объектов знаний; создана онтологическая база знаний электронного образовательного ресурса, содержащую единую онтологическую модель знаний и множество семантических метаописаний объектов знаний; разработаны методы создания и поддержки в актуальном состоянии единой онтологической модели знаний, а также набор методов для реализации основных функций работы со знаниями, таких, как объединение, пересечение, классификация, идентификация.

На основе онтологического подхода были разработаны методы реализации оценки семантической близости классов и индивидов, позволяющие классифицировать и идентифицировать классы онтологий.

Экспериментальная проверка разработанной в ходе исследования онтологической базы знаний электронного образовательного ресурса показала, что данный программный продукт позволяет управлять качеством образовательного процесса на основе имеющихся в университете знаний. Разработанная онтологическая база знаний была апробирована и успешно зарекомендована, тем самым, подтвердив свою работоспособность и эффективность структуры приложения.

Достоверность результатов и выводов исследования обеспечена корректным применением комплекса методов решения поставленных задач, адекватных предмету автоматизации.

Резюме

Предметом исследования данной диссертационной работы являются модели знаний образовательных ресурсов вуза с использованием системы онтологий. Цель исследования - извлечение и анализ электронных образовательных ресурсов из электронного портала вуза. Для достижения поставленной цели необходимо решить следующие задачи:

- исследовать возможности онтологического подхода для развития информационной системы образовательного портала вуза и поддержки принятия решений по управлению системой знаний и компетенций вуза;
- создать онтологию электронного образовательного ресурса с целью описания разнородных типов объектов знаний;
- создать онтологическую базу знаний электронного образовательного ресурса, содержащую единую онтологическую модель знаний и множество семантических метаописаний объектов знаний;
- разработать набор методов для реализации основных функций работы со знаниями, таких как объединение, пересечение, классификация, идентификация.

Для разработки онтологической базы знаний электронного образовательного ресурса был использован онтологический подход с учетом методов системного анализа и теории множеств.

Научная новизна исследования заключается в следующем:

- создана онтология электронного образовательного ресурса;
- разработаны методы создания и поддержки в актуальном состоянии единой онтологической модели знаний;
- разработана комплексная онтологическая база знаний электронного образовательного ресурса;
- разработаны методы реализации оценки семантической близости классов и индивидов, позволяющие классифицировать и идентифицировать классы онтологий.

Предлагаемый в данной работе электронный образовательный ресурс, созданный с использованием системы онтологий позволяет управлять качеством образовательного процесса на основе имеющихся в университете знаний и активизировать их для создания новых знаний.

ТҮЙІНДЕМЕ

Диссертация көлемі 104 бет машинамен жазылған мәтіннен тұрады. Диссертациялық зерттеу мәтінде 42 сурет, 4 кесте, 3 қосымша бар. Пайдаланылған әдебиеттер тізіміне 57 деректер атауы енеді.

ЭЛЕКТРОНДЫҚ БІЛІМ БЕРЕТІН ҚОР, ЖОО-ның порталы, ОНТОЛОГИЯ, ТАКСОНОМИЯ, КЛАСС, ҚАСИЕТТЕРІ, ИНДИВИД, СЕМАНТИКАЛЫҚ ЖАҚЫНДЫҚ, БІЛІМ БЕРУ БАЗАСЫ.

Магистрлық диссертация тақырыбы «ЖОО-ның білім беретін порталының интернет-қорларынан білімдерді шығару және талдауы».

Зерттеу пәні жоғары оқу орындарының білім беретін қорлардың онтология жүйесін қолданумен жасалатын білім моделі болып табылады. Зерттеу мақсаты – ЖОО-ның ақпараттық білім беретін қорларына рұқсаттың қамтамасыз етуі үшін ЖОО-ның электрондық порталынан білімдердің шығаруы және талдануы. Қойылған мақсатқа жету үшін келесі тапсырмаларды шешу қажет:

- ЖОО-ның құзырларын және білімдер жүйесін басқарудың шешімдерін қабылдауды қолдау және ЖОО-ның білім беретін порталының ақпараттық жүйесінің дамытуы үшін онтологиялық тәсілдің мүмкіндіктерін зерттеу;

- білім объектілерінің әртүрлі түрлерінің сипаттау мақсатымен электрондық білім беретін қордың онтологиясын жасау;

- білім объектілерінің семантикалық метасипаттамаларының жиыны және білімдердің біртұтас онтологиялық үлгісі бар электрондық білім беретін қордың онтологиялық білім базасын жасау;

- біріктірулер, қиылысу, классификация сияқты біліммен жұмыс жасау негізгі функцияларын іске асыруы үшін әдістерді жетілдіру.

Электрондық білім беретін қордың онтологиялық білім базасын өңдеу үшін жүйелік талдау және жиындар теориясының әдістері есепке алына отырып онтологиялық тәсіл қолданылған болатын.

Зерттеудің ғылыми жаңашылдығы:

- электрондық білім беретін қордың онтологиясы өңделді;

- білімдердің біртұтас онтологиялық моделін өзекті күйінде қолдау және жасау әдістері жасалған;

- электрондық білім беретін қордың кешенді онтологиялық білім базасы жасалған;

- онтология класстарын классификациялауға және белгіленуге мүмкіндік беретін класстар және индивидтердің семантикалық жақындығын бағалауды іске асыру үшін әдістер жасалған.

Осы жұмыста ұсынылатын онтология жүйесін қолданып жасалған электрондық білім беретін қор университетте бар білімдердің негізінде білім беру үдерісінің сапасын басқаруға және жаңа білімдерді жасауы үшін оларды іске қосуға мүмкіндік береді.

Abstract

The volume of thesis is 104 typewritten pages. There are 44 pictures, 3 tables, 3 supplements in the text of investigation of thesis. There are 57 types of sources.

ONLINE EDUCATIONAL SOURCE, UNIVERSITY ONLINE PORTAL, ONTOLOGY, TAXONOMY, CLASS, PROPERTY, INDIVIDUAL, SEMANTIC PROXIMITY, KNOWLEDGE BASE.

The theme of master's thesis «The retrieving and analysis of knowledge from sources of online educational portal ».

The subject of research is the model of knowledge of university's educational sources using ontology. The purpose of research is the retrieving and analysis of knowledge from sources of online educational portal for access to information resources of the university. To achieve this goal it is necessary to solve the following tasks:

- to investigate the possibility of an ontological approach for the development of online educational portal and for the support of decisions on the management of the system of knowledge and competencies of high school;

- to create the ontology of online educational source for describing the diverse types of objects (resources) of knowledge;

- to create an ontological base of knowledge of online educational source, which contains the united ontological model of knowledge and a great number of semantic meta objects of knowledge;

- to develop a set of methods to realize the main functions of work with knowledge, such as the union, intersection, classification.

For the development of ontological base of knowledge of online educational source was used ontological approach, with the account the methods of system analysis and set theory.

Scientific novelty of the study is the following:

- the ontology of online educational source was created;

- the methods of creating and maintaining in an actual state of a single ontological model of knowledge was developed;

- a comprehensive ontological base of knowledge of online educational source was developed;

- the methods of assessing of implementation of the semantic proximity of classes and individuals, were developed which allows them to classify and identify classes of ontologies.

The proposed online educational source founded on ontological base of knowledge in this work allows to manage the quality of the educational process on the base of knowledge in the university and strengthen them to create new knowledge.