

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРИ РАЗРАБОТКЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ДОСТУПА К УЧЕБНО-МЕТОДИЧЕСКОЙ ИНФОРМАЦИИ

Ворона С.П., Савкова Е.О.

Донецкий национальный технический университет
кафедра автоматизированных систем управления
E-mail: vorona.sonya@gmail.com

Аннотация:

Ворона С.П., Савкова Е.О. Использование онтологического моделирования при разработке интеллектуальной системы доступа к учебно-методической информации. В данной статье представлены основные подходы, принципы и методологии разработки онтологии. Рассмотрена возможность повторного использования существующих онтологий и предложен подход к построению онтологической модели базы знаний.

Abstract:

Vorona S., Savkova E. Use of ontological modeling in the development of an intellectual system of access to educational and methodical information. This article presents the main approaches, principles and methodologies of ontology development. The possibility of reusing existing ontologies is considered and an approach to building an ontological model of the knowledge base is proposed.

Введение

Существенное противоречие между огромным объемом информации, с одной стороны, и дефицитом хорошо структурированной образовательной, научно-методической, учебной, справочной и других видов информации, на использование которой направлена профессиональная деятельность педагогов, с другой – возникло на протяжении давнего времени. Стремительное развитие информационных технологий привело к лавинообразному росту и образовательной информации, окружающей современного человека. Это, в свою очередь, приводит к необходимости повышения эффективности использования образовательных ресурсов, поиска новых способов систематизации, хранения, представления и структуризации, а также автоматической обработки знаний. Следовательно, возникает потребность в разработке системы, позволяющей осуществлять интенсивное и динамическое наращивание научно-образовательного контента, повышать качество учебно-методической работы и научных исследований и создавать условия для развития виртуальной академической мобильности преподавателей и научных работников [1].

Лучшая организация данных и информации может быть получена с помощью информационных технологий управления знаниями на основе моделей представления данных в специальной семантической форме. Целью семантических моделей представления данных является предоставление возможности записи информации в форме, которая может быть обработана компьютером, с учетом ее значения. Неотъемлемым компонентом семантических технологий является онтология, представляющая формализованное описание основных понятий предметной области и отношений между ними. Стоит отметить, что спрос на семантические технологии растет с каждым годом. Разработка современных методов искусственного интеллекта и анализа данных привела к активному применению во всем мире интеллектуальных (семантических) методов поиска и создания баз знаний и онтологий, которые могли бы значительно улучшить качество управления в этой области. Таким

образом, актуальна задача создания модели информационного пространства в области научно-образовательного контента, которая позволит эффективно использовать данные, собранные из имеющихся в ВУЗе ресурсов, для чего и требуется разработка онтологии.

Современным интеллектуальным системам для работы необходима некоторая формализация реального мира. Одним из видов такой формализации является построение онтологической модели. Прежде чем перейти к онтологической модели следует акцентировать внимание на понятии онтологии. В инженерии знаний под онтологией понимают общее описание основных понятий предметной области и того, как они между собой связаны. Для более конкретного описания предметной области используется онтологическая модель. Она содержит информацию об отдельных объектах, процессах и событиях предметной области. Вся информация записывается в терминах, заданных онтологией [2].

В качестве ядра модели выбрана онтология, которая является удобным средством описания областей знаний и содержит наряду с традиционным описанием предметной области соотнесенное с ним метаописание структуры и типологии соответствующих хранилищ данных и сетевых ресурсов знаний. Реализация стратегий управления знаниями основывается на использовании разработанных методов построения и работы с онтологической моделью знаний и множеством моделей объектов знаний [3].

Разработка онтологической модели

Концептуальное представление информационно-образовательного пространства в виде онтологий обеспечивает семантическую интерпретацию запросов к образовательному контенту, реализацию унифицированного интеллектуального доступа к множеству источников знаний [4]. В основе концептуального уровня структуры знаний лежит таксономия используемых понятий или онтология, предназначенная для идентификации различных компонентов знания. Онтологию можно рассматривать, как систему рубрикации предметной области, благодаря которой интегрируются разнородные источники знаний. С другой стороны, онтология часто рассматривается, как словарь-тезаурус, совместно используемый для упрощения коммуникации пользователей, интерпретации и формулирования их запросов.

При использовании программ оптимизации, конвертирования онтологических описаний и структуризации (Protégé, GrOWL бChimaera, ODE) применяют оригинальные подходы к унификации и структуризации онтологических описаний [5].

Необходимость в разработке онтологий объясняется рядом следующих причин [6]:

- 1) разработкой и управлением терминологией;
- 2) совместным использованием людьми или программными агентами общего понимания структуры информации;
- 3) возможностью повторного использования знаний в предметной области;
- 4) отделением знаний в предметной области от оперативных знаний;
- 5) получением надежного семантического базиса в определении содержания;
- 6) возможностью использования онтологий для поддержки функционирования и роста нового вида цифровых библиотек, реализованных как распределенные интеллектуальные системы.
- 7) получением логической теории, которая состоит из словаря и набора утверждений на некотором языке логики, что позволяет на основе этой теории получать вывод новых знаний, явно не заложенных в онтологии;

При разработке онтологии учитываются принципы, приведенные в [7]:

- Не существует одного единственного правильного способа моделирования предметной области, зачастую существуют жизнеспособные альтернативы. Наилучшее решение почти всегда зависит от его применения и возможного ожидаемого расширения.
- Разработка онтологий — это неизбежно итеративный процесс.

- Концепты в онтологии должны быть как можно более близки к объектам (физическим или логическим) и связям в рассматриваемой предметной области. Они наиболее вероятно являются существительными (объекты) и глаголами (связи) в предложениях, описывающих рассматриваемую предметную область.

К наиболее распространённым методологиям разработки онтологий относятся On-To-Knowledge, METHONTOLOGY, DILIGENT и NeOn. Рассмотрим подробнее последнюю методологию:

- учитывает наличие имеющихся онтологий и предусматривает повторное использование уже существующих онтологий;
- предлагает несколько сценариев разработки, которые выбираются в зависимости от входных требований.

Учитывая методологию NeOn, разработка онтологической модели включает в себя следующие шаги:

- 1) идентификация цели, языка описания и области действия онтологии;
- 2) идентификация предполагаемых конечных пользователей;
- 3) идентификация предполагаемых сценариев применения;
- 4) идентификация функциональных и нефункциональных требований;
- 5) группирование функциональных требований (компетентностных вопросов);
- 6) валидация требований;
- 7) определение приоритетов требований;
- 8) извлечение терминологии и оценка её частоты;
- 9) обзор существующих онтологических ресурсов (онтологий, словарей, тезаурусов, и т.д.) как в формате RDFS или OWL, так любых других для повторного использования в разрабатываемой модели;
- 10) непосредственное структурирование или кодирование модели на языке RDFS или OWL;
- 11) оценка соответствия разработанной модели требованиям, которые были сформулированы на первом шаге.

Разрабатываемая онтология должна соответствовать следующим нефункциональным требованиям:

- наименование концептов онтологии должно производиться на английском языке; наименование локальных имен концептов должно соответствовать UpperCamelCase-стилю для классов и индивидов, и lowerCamelCase стилю для связей;
- онтология не должна напрямую импортировать ни одну из существующих онтологий, чтобы не вносить прямые зависимости от внешних онтологий;
- онтология должна использовать только конструкции подмножества языка RDF Schema..

Определение возможности повторного использования существующих онтологий

При разработке новых онтологий одной из рекомендаций является повторное использование концептов из существующих онтологий. На текущий момент разработано много онтологий, с помощью которых можно описать проблемно ориентированную модель. В связи с этим необходимо провести анализ и определить, какие из существующих онтологий могут быть использованы [8]. Проведение анализа онтологии следует проводить по следующим параметрам:

- предназначение (с какой целью была разработана онтология?);
- пространство имен онтологии;
- концепты для повторного использования (какие из концептов, определённых в данной онтологии, целесообразно использовать повторно?).

В данный момент в мировой практике уже сложилось некоторое количество стандартизованных верхнеуровневых онтологических моделей, которые соответствуют в предметной области:

DC (Dublin Core) - метаонтология, представляющая собой словарь, используемый для описания предметных областей и ресурсов в целом. Этот словарь содержит общие свойства, обычно используемые для элементов метаданных на различных типах ресурсов, включая названия, авторов, даты, издателей, события и т.д.

BIBO (The Bibliographic Ontology) - используется в качестве онтологии цитирования, онтологии классификации документов или просто как способ для описания любого вида документов в RDF. При разработке онтологии были использованы многие существующие форматы описания метаданных документа, поэтому BIBO может быть использована в качестве общей основы для преобразования произвольных источников библиографических данных к общему формату.

AISO (the Academic Institution Internal Structure Ontology) - онтология внутренней структуры академических институтов, используется для представления структуры научных организаций в терминах подразделений и учебных программ. Она содержит классы для представления таких объектов, как факультеты, кафедры, учебные модули. AISO используется в качестве основной схемы для описания организационных диаграмм образовательных учреждений и курсов, которые преподаются в конкретном учреждении.

SKOS (Simple Knowledge Organization System) - модель связывания научных данных, которая адаптирована для компьютерной обработки. В частности, SKOS включает контролируемые структурные словари семантических значений для связывания научных данных.

MLO (Metadata for Learning Opportunities) - метаданные для образовательных возможностей. Данная онтология является стандартной европейской моделью метаданных для рекламы и продвижения возможностей обучения (модули и программы курсов). Она разработана в целях единообразного представления информации об образовании в различных учреждениях и странах.

W3C - онтология для медиаресурсов, представляет расширенные метаданные для медиаресурсов в Интернет.

Рассмотренные онтологии характеризуются концептуальной локальностью и локальностью в использовании. Отсутствие базисной онтологии затрудняет процесс объединения существующих онтологий, зачастую несовместимых между собой. На основе описанных функциональных требований можно выделить следующее множество основных терминов: пользователь, источник данных, тип доступа к данным, набор данных, предметная область, язык публикации данных, термин, перевод термина, значение термина, курс, факультет, модуль, программа, предмет, описание, название, тип, тема, формат, дата, отношения, и т.д.

Описание онтологической базы знаний

Модель онтологической базы знаний приведена в общем виде таким образом (см. формула 1):

$$Z = \langle O_m, O_p, M \rangle \quad (1)$$

Где O_m онтология верхнего уровня (метаонтология), которая включает в себя наиболее абстрактные понятия и отношения между ними; O_p набор доменных онтологий; M это механизм познания умозаключения. Таким образом, представляется наиболее удобным создание онтологической базы знаний для интеллектуальной системы доступа к информационно-интеллектуальному контенту на основе метаонтологии (онтологии верхнего уровня), например, как представлено на рисунке 1.



Рис. 1. Мета модель Dublin Core.

Каждый узел в онтологии верхнего уровня является вершиной иерархии классов для онтологий низшего уровня, который можно представить следующим образом [9]:

$$O_p = \langle C, R, S, G, T, D_s, D_g, E \rangle \quad (2)$$

где $C = \{c_i | i = 1, \dots, n\}$ - конечное непустое множество классов, описывающих понятия области;

$R = \{r_i | i = 1, \dots, m\}$ - конечное множество бинарных отношений между классами, $R \subseteq C \times C, R = \{R_{ISA}\} \cup R_{ASS}$, где R_{ISA} — это антисимметричное, транзитивное, нереклексивное иерархическое отношение "класс-подкласс", определяющее частичный порядок на множестве классов; R_{ASS} - конечный набор ассоциативных отношений;

$S = \{s_i | i = 1, \dots, k\}$ - конечный набор слотов (атрибуты класса);

$G = \{gs_i | i = 1, \dots, l\}$ - конечный набор концептов (атрибутов слота);

T является конечным непустым множеством, которое определяет управляемый словарь терминов домена, построенный на множестве базовых терминов $B = \{b_i | i = 1, \dots, n\}$ являющихся множеством имен классов онтологии:

$$T = \bigcup_{i=1}^n T_i \quad (3)$$

$$T_i = \{b_i\} \cup E_q(b_i), \bigcap_{i=1}^n T_i = \emptyset \quad (4)$$

$E_q(b_i)$ - набор синонимичных терминов, каждый термин связан с базовым термином $b_i \in B$;

D_s - конечный набор типов слотов;

D_g - является конечным набором типов фасетов;

$E = \{e | i = 1, \dots, u\}$ - конечное множество экземпляров классов.

Структура класса определяется следующим образом:

$$c = \langle Name_c, (isa\ c_{parent}), (s_1, s_2, \dots, s_{n(c)}) \rangle \quad (5)$$

где $c, c_{parent} \in C$ классы онтологии, связанные иерархическими отношениями R_{ISA} , $s_j \in S$ – слоты, $Name_c \in B$ — это имя класса, являющееся базовым термином контролируемого словаря T . Иерархия классов формируется с помощью отношения «is-a».

Множество классов делится на два непересекающихся множества $C = C_{abstract} \cup C_{concrete}$. Для класса из подмножества $C_{concrete}$ можно определить экземпляры $e \in E$.

Структура экземпляра класса аналогична структуре класса c , которая формирует экземпляр:

$$e(c) = \langle Name_e, (s_1^e, s_2^e, \dots, s_{n(c)}^e) \rangle \quad (6)$$

Где $s_1^e, s_2^e, \dots, s_{n(c)}^e$ Образцы слотов заполнены указанными значениями атрибутов.

Определение ассоциативных отношений, составляющих множество R_{ASS} осуществляется посредством явного указания в качестве значения слота имени ассоциированного класса, а также типа ассоциативных отношений, существующих между этими классами. Для реализации ассоциативных отношений используются типы D_{class} (тип «Класс») и $D_{instance}$ (тип "Экземпляр") вместе со стандартными типами D_{SS} (символ, строка, число с плавающей точкой, т.д.):

$$D_S = D_{SS} \cup \{D_{class}\} \cup \{D_{instance}\} \quad (7)$$

Определение типов D_{class} и $D_{instance}$ включает в себя указание дополнительного аргумента – ассоциированного учебного класса. Если один из слотов класса c_1 имеет тип $D_{instance}$ с ассоциированным классом c_2 , то в слоте Значения при создании экземпляров класса c_1 могут быть использованы экземпляры набора $Tr(c_2)$, который является транзитивным замыканием c_2 , относящегося к R_{ISA} , включая класс c_2 и все его подклассы ниже иерархии:

$$Tr(c_2) = \{c_2\} \cup \{c_i \in C | \exists R_{ISA}(c_i, c_2)\} \quad (8)$$

В этом случае классы c_1 и c_2 связаны с ассоциативным отношением, т.е. $\exists R_{ASS}(c_1, c_2)$. Если один из слотов класса c_1 имеет тип D_{class} с ассоциированным классом c_2 , тогда в качестве значения слота при создании экземпляры класса c_1 могут быть использованы классы множества $Tr(c_2)$. В этом случае классы c_1 и c_2 также связанных с ассоциативным отношением, т.е. $\exists R_{ASS}(c_1, c_2)$. Таким образом, в качестве значения слота может быть не только выборка ассоциативных классов, но и базовый термин. Последний может быть использован для описания сложных объектов, таких как виды требований с условиями контролируемой лексики. Семантические отношения между отдельными областями домена могут быть описаны с использованием стандартизированной модели RDF для описания

объектов данных для семантической сети. Это позволит подключить данные, представленные в машиночитаемом формате.

Заключение

Для успешной разработки и эффективного функционирования интеллектуальной системы доступа к учебно-методической информации в рамках ВУЗа необходимо грамотно организовать и описать пространство знаний. Как было отмечено, наиболее удобным и очевидным способом организации такого пространства знаний является онтология. В связи с этим были рассмотрены основные принципы и методологии создания онтологической модели, а также возможность повторного использования некоторых существующих онтологий для упрощения разработки. В дальнейшем планируется создание детального механизма связывания доменных онтологий через развитую онтологию верхнего уровня в интегрированной базе онтологических знаний.

Литература

1. Андриевская Н.К. Основные принципы и подходы при разработке системы управления профессиональными знаниями ВУЗа. Информатика и кибернетика. – 2019. – С. 49-56.
2. Пальчунов Д. Е. Моделирование мышления и формализация рефлексии. I: Теоретико-модельная формализация онтологии и рефлексии // Философия науки. –2006. – № 4 (31). – С.62-99.
3. Кравченко Ю.А., Бова В.В. Нечеткое моделирование разнородных знаний в интеллектуальных обучающих системах // Открытое образование. – 2013. – № 4 (99). – С. 70-74.
4. Тельнов Ю.Ф. Принципы и методы семантического структурирования информационно-образовательного пространства на основе реализации онтологического подхода // Вестник УМО. Экономика, статистика, информатика. – 2014. – № 1. – С. 187–191.
5. Навроцкий М.А., Жукова Н.А., Муромцев Д.И. Онтология проектирования, применения и сопровождения порталов научно-технической информации // Онтология проектирования. – 2018. – №1 (27).
6. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер. – 2000. – С. 384.
7. Suárez-Figueroa, M. The NeOn methodology for ontology engineering / M. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López // Ontology engineering in a networked world. – Springer Berlin Heidelberg. – 2012. – P. 9-34.
8. Janowicz K. et al. Five stars of linked data vocabulary use. Editorial // Semantic Web 0. – 2014. – IOS Press. 4 p.
9. Avdeenko T V, Bakaev M A. Scientific Bulletin of NSTU. – 2016 – 3 84.