

РЕДАКТОР ОНТОЛОГИИ ДЛЯ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ СЕМАНТИЧЕСКОЙ ЗНАЧИМОСТИ ПОНЯТИЙ

Овчаров А. Э., Мамонова Т. Е.

Научный руководитель: Мамонова Т. Е.

e-mail: fakiri41@sibmail.com

Введение

В последние годы развитие современного общества характеризуется ростом информационных технологий в науке и производстве [1]. Вследствие этого накапливаются большие разнокачественные объемы информации, которые нуждаются не только в хранении и извлечении этой информации, но и в создании способов представления, формализации и обработке посредством компьютерных систем. В настоящее время наиболее развиты такие средства поиска и представления информации как поисковые системы Google, Yandex, тематические каталоги, различные информационные порталы [2].

Преимущества онтологий

Во всех перечисленных средствах задача навигации не решается на должном уровне, так как нет структурированного списка ссылок или рубрикаций, а только лишь огромные базы данных, состоящие из различных веб-страниц, которые индексируются при помощи компьютеров (программ-роботов «пауков»).

Поиск информации осуществляется следующим образом. Пользователь вводит запрос – слово или фразу, по которой поисковая система будет искать необходимые веб-сайты. Содержимое запроса сопоставляется с веб-страницами, которые находятся в базе данных поисковой системы. При наличии совпадений эти страницы добавляются в список результатов поиска.

Целесообразность и результативность такого поиска применима в случае когда: пользователь может точно сформулировать свой запрос, необходима свежая и актуальная информация, касающаяся какого-либо события, имеется время для отбора нужной информации из огромного массива найденных материалов, пользователь готов сам оценить качество найденной информации.

Отрицательными сторонами данных поисковых систем является то, что поисковые системы не занимаются поиском следующих сайтов и веб-ресурсов: с узкой сферой знаний, представляющие из себя базы данных, быстро стареющие и теряющие актуальность, с непрямыми адресами.

По своей сути базы данных описанных систем не имеют структурированных баз знаний, а лишь огромный объем хаотично расположенных материалов, у которых отсутствует контроль качества информации.

В последнее десятилетие все чаще используются понятия «онтология» и «онтологические технологии». Данный термин впервые появился в

работе Томаса Грубера [3], в которой рассматривались различные аспекты взаимодействия интеллектуальных систем между собой и с человеком. В его формулировке (1991) онтология определена как «формальная спецификация концептуализации, которая имеет место в некотором контексте предметной области». Под концептуализацией понимается представление предметной области через описание множества понятий (концептов) и связей (отношений между ними). Преимуществом онтологий в качестве способа представления знаний является их формальная структура, которая делает возможной автоматическую обработку онтологической информации, что может иметь применение в: искусственный интеллект, информационный поиск, вопросно-ответные системы, машинный перевод, создание электронных обучающих систем, системный анализ предметной области, интегрирование данных и знаний, системы понимания языка (аннотирование текста, реферирование и т.д.).

Онтология позволяет более эффективно обрабатывать сложную и разнообразную информацию. Этот способ представления знаний позволяет приложениям распознавать те семантические отличия, которые являются само собой разумеющимися для людей, но не известны компьютеру [4].

Редакторы онтологий на примере Protégé

Основная функция любого редактора онтологий состоит в процессе формализации знаний и представлении онтологии как спецификации (точное и полное описание).

Важной характеристикой является функциональность редактора, т.е. множество сценариев его использования. Базовый набор функций обеспечивает [5]: работу с одним или более проектами, сохранение проекта в нужном формате (экспорт), открытие проекта и представление его структуры, импорт из внешнего формата, редактирование метаданных проекта (от настройки форм редактирования и представления данных, до поддержки версий проекта), редактирование онтологии (создание, редактирование, удаление понятий, отношений, аксиом и прочих структурных элементов онтологии).

С момента его создания Protégé многие годы использовался экспертами в основном для концептуального моделирования в области медицины. В последнее время его стали использовать в других предметных областях [6].

Protégé основан на модели представления знаний ОКВС (Open Knowledge Base Connectivity).

Основными элементами являются классы, экземпляры, слоты (представляющие свойства классов и экземпляров) и фасеты, задающие дополнительную информацию о слотах. Окно редактора онтологии Protégé приведено на рисунке 1.

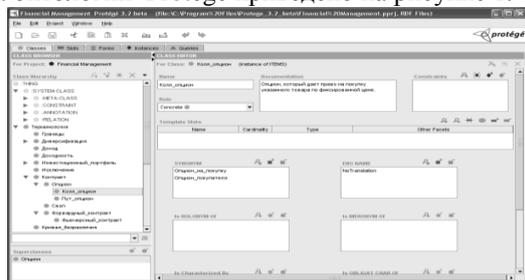


Рис. 1. Окно редактора Protégé

Недостатками Protégé и других существующих редакторов онтологии, например, таких как Fluent Editor и On-To-Knowledge является то, что эксперт либо пользователь не имеет возможности редактировать созданную онтологию. Ни один из редакторов не способен обеспечивать автоматическое определение семантической значимости понятий или определений.

Анализ актуальности разработки программного продукта

Анализируя выше изложенные существующие редакторы онтологии на примере Protégé можно сделать вывод, что задача по разработке метода построения онтологии является актуальной и разработанный продукт должен соответствовать следующим требованиям.

- Может быть использован, как для начального формирования онтологии, так и для дополнения существующей онтологии.
- Должен обеспечивать оценку семантической значимости сформированных множеств понятий (концептов).

При этом в случае возникновения семантической неопределенности решение может быть принято экспертом. Также существует возможность автоматизации данного вопроса без участия эксперта. Для правильного выбора значения программой, модель онтологии должна быть дополнена концептами.

В соответствии с задачами и требованиями, которые должен решать программный продукт наиболее подходящей средой проектирования является Visual Studio и язык программирования C#. На рисунке 2 представлена структурная схема разрабатываемого редактора онтологии.



Рис. 2. Структурная схема

Редактор онтологии включает в себя следующие модули.

1. Интерфейс пользователя – реализация возможности управления программой пользователем посредством кнопок, выпадающих меню, вспомогательных окон.

2. Управление – реализация алгоритма управления программным продуктом в соответствии с идеей разработки метода построения онтологии.

3. Модуль онтологии – реализация операций, которые могут быть применимы к концептам.

4. Модуль ввода-вывода – позволяет загружать модели онтологии из базы данных, работать и править их для дальнейшего сохранения.

5. Подключаемые компоненты – обеспечивают загрузку данных об экземплярах и концептах онтологии.

Использование онтологии изучаемой области делает возможным структурировать данные, т.е. формируется модель области знания, что можно сравнить с освоением определенной области изучения человеком.

В дальнейшем планируется использовать разработанный редактор онтологии для создания электронных обучающих систем, которые могут быть применимы преподавателями любых специальностей и направлений высших учебных заведений. Также редактор может быть использован для создания онтологии как образовательной программы по изучению русского языка иностранными студентами.

Список использованных источников

1. Роль информации в современном обществе [Электронный ресурс]. – Режим доступа: http://www.rusnauka.com/12_KPSN_2010/Informatic_a/60717.doc.htm, свободный. – Загл. с экрана.
2. Классификация поисковых систем [Электронный ресурс]. – Режим доступа: http://wiki.iteach.ru/index.php/klassifikaciya_poiskovich_system, свободный. – Загл. с экрана.
3. Gruber T. R. The role of common ontology in achieving sharable, reusable knowledge bases // Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference. J.A. Allen, R. Fikes, E. Sandewell – eds. Morgan Kaufmann, 1991. – 601 p.
4. Константинова Н. С., Онтологии как системы хранения знаний. – Санкт-Петербург, 2012. – 54 с.
5. Соловьев В. Д., Добров Б. В., Иванов В. В., Лукашевич Н. В. Онтологии и тезаурусы. – Казань, 2006. – 157 с.
6. Применение онтологии в задачах эксплуатации кораблей [Электронный ресурс]. – Режим доступа: <http://www.swsys.ru/index.php?page=article&id=3453>. – Загл. с экрана.