

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЙ В СИСТЕМЕ СЕМАНТИЧЕСКОГО АНАЛИЗА КЕЙСОВ

А.В. Шишин,

аспирант факультета бизнес-информатики и прикладной математики Нижегородского филиала Государственного университета – Высшей школы экономики

О.Р. Козырев, профессор факультета бизнес-информатики Государственного университета – Высшей школы экономики
okozyrev@hse.ru

Доказавший свою эффективность на практике кейс-метод сегодня активно применяется многими учебными заведениями. Нелинейность кейсов обуславливает необходимость наличия развитых средств компьютерной поддержки процесса обучения, обладающих функциями семантической обработки. В работе формулируются концепции применения онтологий в информационной системе, призванной повысить эффективность решения преподавателями, практикующими кейс-метод, задачи отбора релевантного материала, формирования последовательности кейсов и ее адаптации к обучаемым.

Применение информационных технологий в сфере образования

Исследование вопроса разработки и применения различного рода компьютерных систем в образовательных целях насчитывает длительную историю. Сегодня в фокусе внимания преподавателей и исследователей находятся не проблемы создания отдельных пакетов прикладных программ (основная задача сводилась к обеспечению практических занятий и лабораторного практикума), а вопросы проработки концепций и создания масштабных информационно-образовательных сред, понимаемых как совокупность знаний, средств их представления и обеспечение доступа к ним различных категорий пользователей [1].

Существующая методологическая база в этой области позволяет создавать образовательные комплексы, включающие в себя: учебные пособия; автоматизированные учебные курсы; различные программные тренажеры. Последние служат основным средством приобретения и освоения обучаемыми так называемых неартикулируемых знаний, таких как профессиональная интуиция и опыт, которые невозможно передать через текст [2].

Развитие информационных технологий, тотальное распространение Интернета, обеспечившего широкую доступность информационных ресурсов,

привело к увеличению спроса на образовательные продукты. Одно из направлений в области компьютерной поддержки процесса обучения – разработка и развитие сетевых образовательных средств и продуктов, ориентированных на различные категории обучаемых, например, поддержка программ дистанционного обучения. В спектр таких проектов попадают и различные информационно-образовательные порталы [3] – огромные массивы данных и знаний. Наиболее эффективные из них построены на базе онтологий, описывающих семантику представленных ресурсов [4].

Одна из самых сложных проблем в работе над автоматизированными средствами обучения – проблема адаптации учебного материала к группам обучаемых с разным уровнем подготовки.

Большинством учёных констатируется факт отставания методических аспектов информационных технологий обучения от развития технических средств. К существующим проблемам добавляются и такие аспекты, как распространение сравнительно новых методов обучения, несущих в себе новые трудности существующим системам [1].

Один из таких методов, зарекомендовавший себя у преподавателей разных учебных заведений и получивший широкое распространение сравнительно недавно, – это кейс-метод, или метод анализа

ситуаций — интерактивный метод обучения, призванный развивать в обучаемых личностный компонент знания, все то, что не может быть передано непосредственно от преподавателя к обучаемому. Такие аспекты профессионального знания могут быть получены учеником только самостоятельно в ходе анализа представленных ситуаций, заведомо не имеющих единственно верного решения. Наиболее активно кейс-метод применяется в рамках курсов подготовки предпринимателей и менеджеров в бизнес-школах во многих странах. В ведущих учебных заведениях мира накоплен обширный материал, активно используемый преподавателями в рамках кейс-метода. Простое аккумулятивное хранение документов в слабоструктурированных хранилищах делает эти материалы практически недоступными ввиду их количества и неэффективности поиска по ключевым словам, когда не учитывается семантика слов и значение контекста, в котором они употребляются.

Цель данной работы — представление концепции информационной системы, призванной обеспечить помощь преподавателям, практикующим кейс-метод в процессе обучения, в задаче поиска и отбора релевантных материалов, а также организации кейсов в последовательность, составляющую адаптированный к обучаемым курс.

Специфика кейс-метода

Особенность любого кейса — его сюжетность. Нелинейность кейса, его насыщенность ролями делает разработку хорошего кейса нетривиальным длительным процессом, требующим серьёзных интеллектуальных затрат. С другой стороны количество удачных кейсов все время растёт, и этот факт обуславливает необходимость наличия удобных, интуитивно понятных средств поддержки работы с хранилищами кейсов.

Особенности, присущие кейс-методу, с которыми может столкнуться преподаватель, рассмотрены ниже.

1. Проблема подбора материала или организации кейсов в последовательности, где каждый элемент несёт новизну, потенциально новые знания и опыт, но система в целом сохраняет тематическую канву. Требования, предъявляемые к такой последовательности, самые разные, например, непротиворечивость, уникальность, нарастающая сложность материала. С одной стороны, последовательное освоение кейсов с опорой на уже полученный в предыдущих обсуждениях опыт сделает анализ текущего материала эффективнее и продуктивнее; с другой

стороны, органично дополняющие друг друга тексты способствуют приобретению чётко сформулированных профессиональных навыков. Таким образом, это проблема тематической группировки текстов, а её решение — ручная проработка всех доступных материалов. В рамках этой же проблемы решается вопрос адаптации материала к обучаемым с разными уровнями подготовки.

2. Любая информация имеет свойство устаревания. Работа с неактуальными данными, снижает вовлечённость обучаемых в процесс. Эффективность такого кейса может сильно уступать запланированной. Здесь представляется проблема подбора альтернативы с небольшими девиациями устаревшему кейсу с сохранением общей направленности всей последовательности или всего курса.

Периодическое обновление кейсов по одной и той же теме позволяет проводить ретроспективный анализ способов решения сходных проблем и проследить эволюцию решений, что представляет определённый профессиональный интерес.

Наличие системы семантического анализа текстов с функциями поиска и группировки семантически близких текстов позволит повысить эффективность и скорость решения описанных выше проблем.

От системы требуется работа с текстовыми материалами на уровне семантики слов и контекста, в котором они употребляются. Решение задачи семантической обработки текстов тесно связано с вопросами представления знаний предметной области и способами аннотирования текстов. Знания о предметной области позволяют оперировать текстом как совокупностью объектов и фактов, идентифицируемых и интерпретируемых в рамках имеющихся знаний.

Онтологии

Онтология — воплощение концептуального знания о предметной области — состоит из следующих структурных компонентов:

- ✧ таксономии;
- ✧ описания отношений, в которых находятся объекты предметной области.

Создание онтологий — сложный и интерактивный процесс. В нём участвуют эксперты конкретных предметных областей и специалисты инженерии знаний. На сегодняшний день разработаны подходы, позволяющие до определенной степени автоматизировать этот процесс. Однако подавляющее

большинство существующих онтологий разработаны «вручную» с применением специальных технических средств – редакторов онтологий (например, Protégé, OntoEdit) [10, 11].

Количество качественных доступных онтологий в мире сегодня не так велико, чтобы удовлетворить растущий спрос. Поэтому коллективы разработчиков вынуждены адаптировать наиболее подходящие версии, либо разрабатывать собственные онтологии для решения индивидуальных практических задач. Часто онтологии разрабатываются в рамках решения одной конкретной задачи, и требования, предъявляемые к онтологии, диктуются спецификой выбранного подхода и поставленными целями. Онтологии не создаются один раз «на века». В процессе жизненного цикла они могут претерпевать значительные изменения, поскольку:

- ✧ онтологии могут содержать ошибки, заложенные на этапе проектирования;
- ✧ представления о предметной области могут меняться со временем, что делает ряд предположений неактуальными или противоречащими действительности;
- ✧ сами требования к онтологии могут также изменяться со временем [5].

На проблему эволюции онтологий и на развитие средств технического сопровождения онтологий на всём жизненном цикле направлено сегодня внимание учёных.

Системы семантического анализа неструктурированной информации

Одной из масштабных разработок в области семантического анализа текстов является продукт компании IBM – UIMA (Unstructured Information Management Architecture) [8]. Данная платформа предназначена для использования её в качестве основы для построения сложных систем семантического анализа неструктурированной информации. Архитектура UIMA унифицирует процесс обработки не только текстов на естественных языках, но и позволяет анализировать мультимедиа-файлы. Например, рассматриваемый в нашем случае кейс может сопровождаться звуковыми файлами, несущими дополнительную, относительно текста, информацию. UIMA позволяет создавать сложные приложения, где каждый компонент выполняет определённые функции: идентификация языка; синтаксический анализ или непосредственное аннотирование текста. Для каждого элемента системы чётко определены входные параметры и формат результирующих данных на выходе, что позволяет компонентам

с лёгкостью использовать для своего алгоритма данные-результаты других блоков. Чёткая спецификация взаимодействия составляющих системы предоставляет разработчику возможность повторного использования компонентов других авторов [8].

Синтаксис системы типов позволяет описывать иерархии одиночного наследования. Каждый пользовательский тип расширяется набором свойств (features). Таким образом, система типов может интерпретироваться как таксономия понятий предметной области. Система типов отображается утилитой автоматической генерации кода в набор Java-классов. Каждый аннотатор имплементирует специальную логику идентификации набора символов в тексте как экземпляр того или иного класса. Результаты аннотирования индексируются и могут быть доступны третьим компонентам, отвечающим непосредственно за предоставление определённой функциональности пользователю. Функциональная специализация отдельных блоков обуславливает их относительную самостоятельность и возможность повторного использования, а система, построенная из отдельных компонент, является масштабируемой.

Типичная схема построения приложений на базе UIMA представлена на *рис. 1*.

Представленная на рисунке подсистема обработки коллекций (CPE – Collection Processing Engine) состоит из набора компонент-аннотаторов. Обмен данными между различными компонентами системы осуществляется посредством CAS (Common Analysis Structure), представляющую результаты аннотирования текста. Применение технологии IBM для построения приложений семантического анализа данных не ограничивается научно-исследовательскими проектами. Удачная концепция нашла свое применение в ряде коммерческих продуктов, перечень которых варьируется от интеллектуального поиска (OmniFind от IBM) до систем автоматизированной обработки заявок в службе поддержки пользователей [9].

Концептуальное описание системы

Объединение преимуществ онтологического подхода к спецификации концептуальных знаний о предметной области с потенциалом архитектуры UIMA представляет определённый интерес. В рамках проводимых исследований и экспериментов по созданию приложений семантической обработки текста на базе названных технологий системы типов и онтологии рассматриваются независимо друг от друга. Их назначение различно, последние служат для представления концептуальных знаний о предметной области и логического вывода на них,

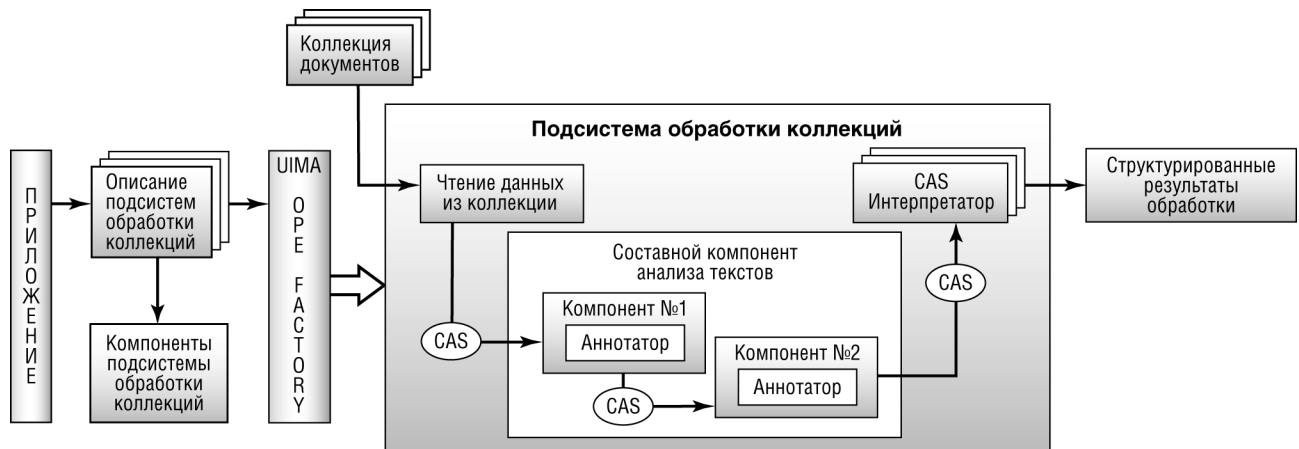


Рис. 1. Архитектура системы семантической обработки текстов

первые – разрабатываются для анализа конкретного текста. В общем случае разработчики – разные люди. Отметим, что проектирование масштабной системы типов, также как и создание онтологий требует больших временных затрат. Нетрудно проследить и сходства в жизненном цикле представителей обеих форм концептуализации предметной области. Учитывая сложность процесса разработки онтологий, таксономий, а также их эволюцию, целесообразнее получать системы типов непосредственно из описания онтологий.

На рис. 2 приведена архитектура, на основе которой разрабатывается система семантического анализа кейсов. Приложение, основанное на предлагаемой архитектуре, предоставит пользователям возможность работать с кейсами на уровне семантики, заложенной в онтологии.

Система, построенная на представленной архитектуре, позволяет оптимизировать решение некоторых инженерных задач.

1. Значительно облегчается процесс создания системы типов, используемой компонентами UIMA на этапе анализа документов. Весь процесс сводится к разработке онтологии интересующей предметной области. На этом этапе могут быть использованы специализированные визуальные средства (например, Protégé).
2. Актуализация автоматической синхронизации системы типов UIMA с эволюционирующей онтологией. Наличие такого механизма гарантирует, что компоненты системы при анализе текста всегда используют актуальную систему типов.
3. Обновление базы данных аннотированных документов с учётом изменений в системе типов.

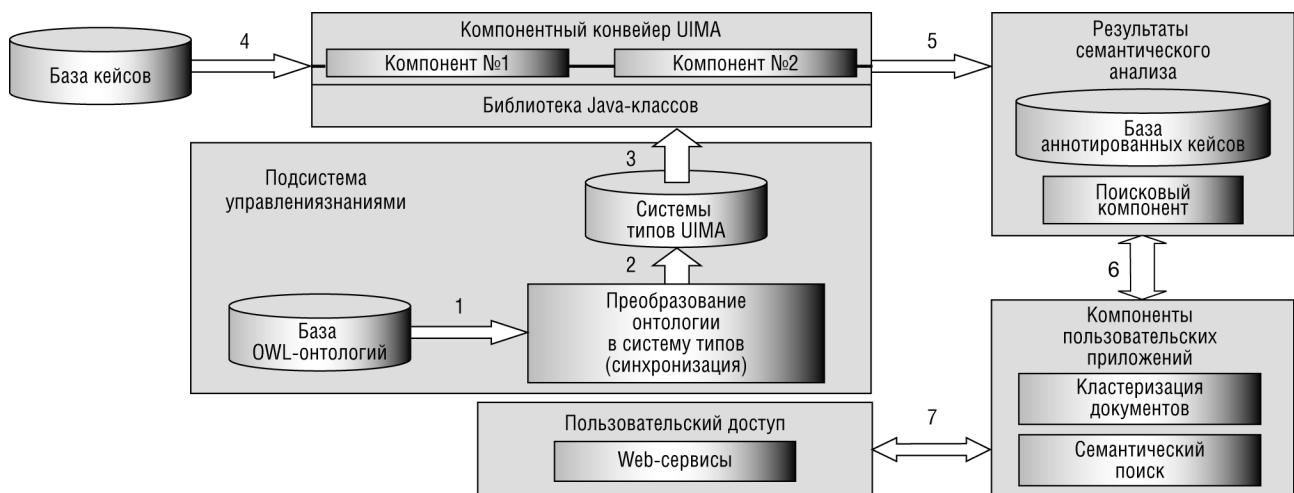


Рис. 2. Структурная схема системы семантического анализа текстовой информации

Такой подход позволит избежать проведения повторного анализа всего массива документов, и незначительные изменения системы типов не приведут к неоправданным временным затратам со стороны системы.

Предлагаемая архитектура гарантирует:

- ✧ при аннотировании документов всегда используется актуальная версия системы типов, отражающая последние тенденции представлений о предметной области;
- ✧ база данных аннотированных документов, всегда находится в актуальном состоянии и не содержит результатов анализа документов, базирующихся на неактуальных знаниях.

В рамках разработки прототипа системы делается ряд ограничений, которые могут быть устранены в будущем. Описание онтологии предметной области должно быть представлено на языке OWL Lite. Будем считать, что сама онтология не содержит противоречий. Выразительные возможности языка OWL [7] не позволяют трансформировать все конструкции языка в описание типов UIMA непосредственно, особенно отображения ряда ограничений на свойства. Решение этой технической проблемы – модификация исходного кода Java-классов, для генерации которых и служит система типов. Опыт отображения специфичных для OWL языковых конструкций в Java-классы подробно рассмотрен в [6].

На базе данной архитектуры предлагается построить систему семантической обработки кейсов. С учётом описанных в предыдущих разделах особенностей систем поддержки образования и ряда специфических особенностей кейс-метода, активно используемого сегодня в различных учебных заведениях, исследование вариантов построения таких систем актуально. Главное назначение системы – оказание помощи преподавателям в навигации в больших массивах кейсов, тематической группировки материала с целью создания адаптированной последовательности кейсов. Решение, базирующееся на применении методов семантического анализа текстов, даст ощутимое преимущество перед классическими подходами в решении описанных выше проблем обработки неструктурированной инфор-

мации. Применение онтологий в качестве точки отсчёта позволяет использовать богатый опыт ведущих инженеров в области представления знаний.

Основные функции, выполняемые такой системой:

- ✧ семантическое индексирование больших объёмов неструктурированной информации;
- ✧ отражение специфицированных знаний о предметной области в информационных портретах документов (эволюция информационных портретов вместе с развитием онтологий).

Определим целевую категорию пользователей и задачи, которые могут быть решены с помощью системы.

Таблица 1

Категории пользователей системы

Категории пользователей системы	Выполняемые функции/решаемые задачи
Разработчики онтологий предметных областей	<ul style="list-style-type: none"> ✧ Добавление новых онтологий ✧ Модификация существующих онтологий
Преподаватели	<ul style="list-style-type: none"> ✧ Поиск релевантных кейсов ✧ Поиск тематически близких кейсов ✧ Составление адаптированной последовательности кейсов с учетом подготовленности обучаемых
Администраторы образовательного процесса	<ul style="list-style-type: none"> ✧ Добавление/удаление кейсов

Заключение

Рассмотрены вопросы целесообразности применения системы семантического анализа слабо структурированной информации в сфере образования, а именно в рамках кейс-метода, технические аспекты проектирования и реализации такой системы. Концепция системы предусматривает серьёзную помощь в навигации и структуризации обширных объёмов педагогических материалов с целью последующего их объединения в полноценный курс практических занятий, тематически ориентированный и адаптированный к определенной категории обучаемых. ■

Литература

1. Соловов А. В. Проектирование компьютерных систем учебного назначения, Самара, СГАУ, 1995 (<http://www.informika.ru/text/inftech/edu/design>).
2. Норенков И.П., Уваров М.Ю. Информационно-образовательные среды на базе онтологического подхода. // В сборнике научных

- статей «Интернет-порталы: содержание и технологии». Выпуск 3. / Редкол.: А.Н. Тихонов (пред.) и др.; ФГУ ГНИИ ИТТ «Информика». – М.: Просвещение, 2005. – С. 367–378.
3. Радаев В.В. Идеология создания специализированного образовательного портала на примере портала по экономике, менеджменту и социологии / Сб. научн. ст. «Интернет-порталы: содержание и технологии». Вып. 1. ГНИИ ИТТ «Информика». – М.: Просвещение, 2003. – С. 91–109.
 4. Боровикова О.И., Загорюлько Ю.А. Организация порталов знаний на основе онтологий. // Труды международного семинара Диалог'2002 «Компьютерная лингвистика и интеллектуальные технологии». Протвино, 2002. Т.2, С.76–82.
 5. L. Stojanovic and B. Motik. Ontology evolution within ontology editors. In Proceedings of the OntoWeb-SIG3 Workshop at the 13th International Conference on Knowledge Engineering and Knowledge Management EKAW 2002, volume 62, September 2002.
 6. Kalyanpur, A., Pastor, D., Battle, S. and Padget, J., Automatic mapping of owl ontologies into java. In Proceedings of Software Engg. – Knowledge Engg. (SEKE) 2004, Banff, Canada, June 2004.
 7. World Wide Web Consortium (W3C). OWL – Web Ontology Language. <http://www.w3.org/TR/owl-ref>.
 8. UIMA Tutorial and Developers' Guides. <http://www.research.ibm.com/UIMA>.
 9. Marc Andrews Observations On Demand. http://marcandrews.typepad.com/marc_andrews/2006/07/real_world_uima.html.
 10. The prot?g? ontology editor and knowledge acquisition system. <http://protege.stanford.edu>.
 11. OntoEdit Ontology Engineering Environment <http://www.ontoknowledge.org/tools/ontoedit.shtml>.
 12. The Case Method of Teaching Science. <http://ublib.buffalo.edu/libraries/projects/cases/teaching/teaching.html>.



ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ – ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

представляет свои периодические издания

**ЭКОНОМИЧЕСКИЙ ЖУРНАЛ
ВЫСШЕЙ ШКОЛЫ ЭКОНОМИКИ
ЕЖЕКВАРТАЛЬНЫЙ НАУЧНО-ИНФОРМАЦИОННЫЙ
ЖУРНАЛ**

Издается с 1997 г.

Главный редактор –
Евгений Евгеньевич Гавриленков

Журнал освещает теоретические и прикладные проблемы экономической науки. В каждом* номере – статьи ведущих российских экономистов. Рецензии, методологические и лекционные материалы. Эксклюзивные статьи зарубежных экономистов. Данные официальной статистики по широкому кругу вопросов.

Каталог Агентства «Роспечать» – индекс 79264
Объединенный каталог «Пресса России» – индекс 29233

Координаты редакции:
101990 Москва, ул. Мясницкая, 20, офис 235
e-mail: redact@hse.ru, тел./факс: (495) 628-0442

**РОССИЙСКАЯ ЭКОНОМИКА:
ПРОГНОЗЫ И ТЕНДЕНЦИИ
ЕЖЕМЕСЯЧНЫЙ СПРАВОЧНО-АНАЛИТИЧЕСКИЙ
ЖУРНАЛ**

Издается с 1993 г.

Главный редактор –
Елена Анатольевна Иванова

Журнал освещает состояние, динамика и дает прогноз основных социально-экономических индикаторов. В каждом номере – хроника событий экономики. Результаты конъюнктурных опросов предприятий. Самые свежие данные. Аналитический материал представлен с использованием таблиц, графиков и диаграмм.

Каталог Агентства «Роспечать» – индекс 79275
Объединенный каталог «Пресса России» – индекс 40548

Координаты редакции:
117312 Москва, ул. Вавилова, 7, офис 203
E-mail: id.hse@mail.ru Тел./факс: (495) 772-9571

РАСПРОСТРАНЯЮТСЯ ПО РОССИИ И ДРУГИМ СТРАНАМ СНГ. ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ – НА САЙТЕ: www.hse.ru