



ПРИМЕНЕНИЕ ОНТОЛОГО-ТЕЗАУРУСНОГО ПОДХОДА ПРИ РАЗРАБОТКЕ
БАЗЫ ЗНАНИЙ «БУРОВОЙ ИНСТРУМЕНТ»

Гордашник К.З.¹, Кобзев В.Г.², Кулаковский В.Н.¹, Сороченко Т.А.¹

¹ Институт сверхтвердых материалов им. В.Н. Бакуля НАН Украины,

² Харьковский национальный университет радиоэлектроники

Создание систем сбора, анализа и управления информационными массивами на основе онтолого-тезаурусного подхода является актуальным направлением научных исследований в области создания, совершенствования и применения сверхтвердых материалов в базовых отраслях промышленности. Проведенные ранее исследования предметной области «Сверхтвердые материалы» позволили разработать ее модель [1], которая допускает дополнение и расширение, обусловленные появлением новых материалов и новых взаимосвязей между ними.

Тезаурус конкретной предметной области предназначен для описания информационных ресурсов, относящихся к данной предметной области, с использованием стандартизированной терминологии и представляет собой набор ключевых понятий, связанных между собой определенными семантическими отношениями. Построение тезауруса предметной области с помощью статистических методов обработки текстов на естественном языке предполагает, что ключевые слова имеют различные распределения среди релевантных (тематика которых относится к заданной предметной области) и нерелевантных документов. При построении тезауруса различают такие этапы: 1) подготовка обучающей коллекции с двумя классами документов – релевантных и нерелевантных по отношению к заданной предметной области, 2) формирование словаря на основе документов обучающей коллекции, 3) выделение из словаря ключевых понятий предметной области, 4) построение семантических отношений между извлеченными ключевыми понятиями, 5) оценка применимости созданного тезауруса для автоматической классификации документов предметной области (проверка его валидности).

В работе [2] обоснована необходимость создания онтолого-тезаурусной системы управления знаниями в предметной области «Сверхтвердые материалы» и связанной с ней предметной области «Буровой инструмент». Принципиальной особенностью разрабатываемой системы является использование онтологического подхода к организации базы знаний, которая организует семантическую сеть понятий и относящихся к ним описаний.

Онтология представляет собой формализованное представление предметной области, которое включает словарь (или имена) указателей на термины предметной области и логические выражения, описывающие, как они соотносятся между собой [3]. В теоретико-множественной форме онтологию можно описать выражением

$$O = \langle P, R, F \rangle, \quad (1)$$

где P – конечное множество терминов (понятий, концептов) предметной области, которую представляет онтология O ; R – конечное множество отношений между терминами данной предметной области; F – конечное множество функций интерпретации (аксиоматизации), которые задаются на концептах и/или отношениях онтологии O .

Множество $P = \{P_i\}$ в выражении (1) должно быть непустым и конечным. Отдельное понятие P_i имеет собственное семантическое представление, которое связано с множеством конкретных фактов и множеством допустимых синтаксических конструкций. Элементом онтологии также является связь R_j между понятиями или группой понятий

$$(P_n, \dots, P_m) R_j (P_k, \dots, P_l).$$

Таким образом, онтология обеспечивает словарь для представления и обмена знаниями о предметной области и множество связей, установленных между терминами в этом словаре.

Процесс разработки онтологии носит итерационный характер. Понятия в онтологии должны быть близки к объектам (физическим или логическим) и отношениям в предметной области. Чаще всего, ими являются существительные (классы) и глаголы (отношения) в



Секция 1. Информационные системы и технологии: опыт создания, модели, инструменты, проблемы

предложениях, которые описывают предметную область. После определения начальной версии онтологии, ее необходимо оценить, отладить и, как правило, пересмотреть.

Для создания онтологии вначале важно получить полный список терминов, не беспокоясь о пересечении классов, которые они представляют, об отношениях между классами, о возможных свойствах классов или о том, чем являются классы. Следующие два шага – разработка иерархии классов и определение свойств классов – тесно переплетены. Обычно в иерархии дают несколько формулировок классов, затем описывают свойства этих классов и т.д. Существует несколько возможных подходов для разработки иерархии классов: нисходящая, восходящая и комбинированная разработка. Самым простым является комбинированный метод, т.к. классы, находящиеся «посередине», чаще всего являются самыми наглядными классами в предметной области. Из списка основных понятий, относящихся к предметной области, необходимо выбрать те понятия, которые описывают объекты, существующие независимо, а не понятия, которые описывают эти объекты. В онтологии такие понятия будут классами и станут точками привязки в иерархии классов. Иерархия классов организуется путем получения ответов на вопросы следующего вида: если объект является экземпляром одного класса, будет ли он обязательно (т.е. по определению) экземпляром некоторого другого класса?

Стандарт онтологического исследования IDEF5 [4] предусматривает последовательное выполнение следующих пяти действий: 1) изучение и систематизирование начальных условий, 2) сбор и накопление данных, 3) анализ данных (группировка собранных данных для облегчения построения терминологии), 4) начальное развитие (формирование) онтологии, 5) уточнение и утверждение онтологии. В IDEF5 существуют специальные графические (схематический язык SL) и текстовые (язык доработок и уточнений EL) средства построения и описания онтологий. В языке SL предусмотрены обозначения классов и отдельных элементов, обозначения взаимосвязей и изменения состояния, обозначения процессов, соединений и перекрестков, благодаря чему возможно построение целого спектра диаграмм и схем. Текстовый язык EL позволяет детально охарактеризовать элементы онтологии.

Проведенное исследование [2] с помощью языка SL позволяет создать диаграммы строгой (по формальным признакам) и естественной классификации бурового инструмента. Композиционные схемы языка SL позволяют наглядно отобразить состав объектов (буровых инструментов), относящихся к тому или иному классу. Схемы взаимосвязей в SL помогают визуализировать и изучить взаимосвязи между различными классами объектов бурового инструмента. Документирование изменения состояния или класса бурового инструмента в языке SL можно выполнить с помощью диаграммы состояния объекта.

Таким образом, применение описанных средств детального анализа, обеспечения полноты представления структуры имеющихся данных позволяет создать и описать сложную модель базы знаний предметной области «Буровой инструмент», пригодную для эффективного выбора необходимого инструмента для конкретных условий его применения.

1. Кулаковский, В.Н. Интегрированная модель предметной области «Сверхтвердые материалы» / В.Н. Кулаковский, А.А. Лебедева, И.В. Скворцов и др. // Сверхтвердые материалы. – 2009. – №5. – с.90-91. 2. Лебедева, А.А. Онтолого-тезаурусный анализ информации в области бурового инструмента из сверхтвердых материалов / А.А. Лебедева, Е.П. Поладко, Р.К. Богданов и др. // Породоразрушающий и металлообрабатывающий инструмент – техника и технология его изготовления и применения. – Сб. научных трудов. - Вып. 15. – К.: ИСМ им. В.Н. Бакуля НАН Украины, 2012. – с.42-47. 3. В.В. Карасюк, В.Г. Кобзев, Б.А. Железко. Интеллектуальные основы представления системы знаний для целей образования / Дистанционное обучение – образовательная среда XXI века: Материалы VII Международной научно-методической конференции. Минск: БГУИР, 2011. - с. 304-306. 4. IDEF5 Ontology Description Capture Method [Элек. ресурс]. – Режим доступа: <http://www.idef.com/IDEF5.htm>.