

АЛГОРИТМЫ И СРЕДСТВА ФОРМИРОВАНИЯ ОНТОЛОГИИ ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ

Дидык Т.Г.¹, Рыков В.И.¹, Шаронова Ю.В.¹

¹ФГБОУ ВПО «Уфимский государственный авиационный технический университет», Уфа, Россия (450000, Уфа, ул. К. Маркса, 12), e-mail: tanayr@mail.ru

Статья посвящена описанию методологии формирования онтологии заданной предметной области. В различных предметных областях один и тот же концепт может иметь отличающиеся лингвистические объекты и несовпадающие реализации экземпляра концепта в реальном мире. Трудности поливалентного описания явлений реального мира могут быть решены методом формирования соответствующей онтологии. В качестве метода используется технология Рационального Унифицированного процесса (РУП). Онтология имеет формат OWL и строится средствами программы Protégé. Знания, содержащиеся в онтологии в виде свойств объектов, аксиом и URI ссылок, исследуются средствами приложений программы Protégé: OntoGraf, DL Query, SPARQL. Для описания методики построения онтологии используется формат языка UML. Методы и средства решения данной задачи рассмотрены для сферы недропользования на примере нефтедобычи.

Ключевые слова: онтология, онтограф, OWL-формат, Protégé, DLQuery, нефтедобыча.

ALGORITHMS AND METHODS OF THE METHODOLOGY OF CREATING ONTOLOGY OF THE GIVEN KNOWLEDGE DOMAIN

Didyk T.G.¹, Rykov V.I.¹, Sharonova J.V.¹

¹Ufa State Aviation Technical University, Russia (450000, Ufa, street K. Marks, 12), e-mail: tanayr@mail.ru

Purpose of the article is to describe the methodology of creating ontology of the given knowledge domain. In different subject areas, the same concept may have differing linguistic objects and inconsistent implementation of the instance of the concept in the real world. Difficulties polyvalent description of the phenomena of the real world can be solved by the method of formation of the ontology. Rational Unified process (RUP) is used as a method. Ontology has the OWL format and is created using Protégé system. Knowledge, contained in the ontology in the form of object properties, axioms and URI links, is investigated by Protégé tools: OntoGraf, DL Query, SPARQL. UML language is used to describe methodology of the ontology creation. Methods and means of solution of this task are considered to subsurface use by the example of oil production.

Keywords: ontology, ontograph, OWL-format, Protégé, DLQuery, oil production.

Введение

Работа посвящена описанию методологии формирования онтологии определённой предметной области. В качестве метода используется технология Рационального Унифицированного процесса (РУП) [3]. Онтология имеет формат OWL [10] и строится средствами программы Protégé [8]. Знания, содержащиеся в онтологии в виде свойств объектов, аксиом и URI ссылок, исследуются средствами приложений программы Protégé: OntoGraf, DL Query, SPARQL. Для описания методики построения онтологии используется формат языка UML [9].

Целью исследования является анализ методов онтологического моделирования и формирование алгоритмов реализации онтологии конкретной предметной области на примере области нефтедобычи.

Прогресс в области обработки цифровой информации дал сфере прикладных наук и технологий инструменты и возможности оцифровки, хранения и свободного сетевого обмена весьма большими объёмами данных.

Задача доступа к сетевым распределённым данным успешно решается в рамках технологии URI ссылок (англ. Uniform Resource Identifier) [7]. Ссылка URI является расширением идентификатора URL и может быть использована, например, для идентификации книжного издания в терминах ISBN или УДК, в отличие от идентификатора URL, указывающего, по определению, точное местонахождение существующего ресурса в сети.

При наличии доступа к ресурсам определённую проблему вызывает истолкование и использование полученных данных. Рассмотрим методы и средства решения указанной задачи для сферы недропользования на примере нефтедобычи.

Возможности оперирования цифровыми данными в области наук и технологий, описывающих Землю, реализованы во множестве интернет-источниках, библиотеках и ряде профессиональных систем типа ГИС [5]. Комплекс наук о земле и недропользовании, объединённый единым объектом исследования, весьма разнообразен в понятийном, семантическом и инструментальном аспектах. Специалисты в области геологии и, например, геофизики имеют разные цели, термины и методологии исследования одного и того же объекта – Земля.

Обычно источник научных или технологических данных ориентирован на одну или несколько семантически тесно связанных предметных областей и содержит данные, фиксированные в соответствующих терминах и форматах определённой области знания, как правило, отличных от концептуально идентичных аналогов в соседних предметных областях.

Прорывные открытия часто делаются на стыке двух или более формально не связанных наук или технологий. Естественно, что семантический разрыв (несовпадение терминов и методов изучения одного и того же явления) существенно осложняет работу исследователя или технолога. Трудности освоения смежной научной дисциплины носят не только семантический, но и инструментальный характер. Представим концепт, рассматриваемый как класс, в виде триплета [6] (рис. 1).



Рис. 1. Треугольник обозначений

В различных предметных областях объекты концепта могут иметь отличающиеся имена и (или) несовпадающие предметные реализации для различных экземпляров концепта. Трудности поливалентного описания явлений реального мира могут быть решены методом формирования соответствующей онтологии.

Указанной задаче посвящён проект организации «Консорциум Всемирной паутины» (англ. World Wide Web Consortium, W3C, сайт W3.org). В рамках проекта предлагается методология использования знаний конкретной предметной области на базе OWL-онтологии. Опишем принципиальную методику формирования онтологии.

Заметим, что термины концепт, класс, объект имеют одинаковое значения в рамках излагаемой темы и применяются в зависимости от семантики конкретного абзаца (объектный подход, онтологии и т.д.).

Онтология строится как иерархическая структура классов, связанных понятиями. Связь имеет вид «Субъект – Отношение – Объект», где субъект – это класс онтологии, отношение имеет вид «Понятие отношения + Аксиома», объектом может быть класс или объект реализации класса.

При построении онтологии выделяется следующая последовательность действий:

1. Классификация базовых понятий – имен классов: типов объектов, их характеристик, комплексов понятий с их участием. Сюда, например, входит класс нефтяное месторождение. Месторождение имеет название, год открытия, оператора разработки, расчётный запас залежей и т.д. Вводится понятие характеристики объекта, которое имеет имя «Глубина», и значения: минимальная глубина залегания, максимальная глубину залегания.

2. Отбор базовых понятий. В нашем случае это, например, месторождение, скважина, устройства бурения. Устройства бурения делятся на буровые установки, буровое оборудование, буровой инструмент, способы добычи и т.д.
3. Определение отношений. Отношение вводится для описания конкретного понятия. Например, нефтяное месторождение имеет атрибут плотность нефти. Тогда в онтологии появится описание «Месторождение имеет Плотность only Плотность». Описание истолковывается так: любое нефтяное месторождение содержит нефть только одной конкретной плотности. Понятие «имеет Плотность» универсально и может быть также отнесено к буровому раствору или другим материалам.
4. Формируется концептуальная схема онтологии как связанного комплекса концептов. Схема определяется структурой подчинения классов и системой отношений между классами.
5. Онтология дополняется предметными реализациями классов (Individuals) и данными, имеющими физический смысл.
6. Выполняется формирование лингвистической составляющей. Фиксируются синонимичные обозначения каждого понятия или значения (термины): Н=Нефтяное месторождение, НГ=Нефтегазовое месторождение, ГН=Газонефтяное месторождение, Крупнейшие=Очень большие и т.д. Описываются способы выражения отношений из онтологии в языке – типовые лексико-грамматические конструкции, для чего используется соответствующий лингвистическому анализатору формализм.

Формирование онтологии, определение методов её использования и развития является типичной задачей области информационных технологий, и вряд ли она может быть решена вне контекста данной предметной области.

Указанная проблема достаточно хорошо исследована в работах [3] и решается на основе объектного подхода [4].

Рассмотрим в качестве примера решение задачи построения онтологии области нефтедобычи. Онтология предметной области есть явное представление концептуализации предметной области и совокупность соглашений по ее описанию [2]. Эти соглашения являются результатом договоренности между специалистами, работающими в конкретной предметной области. Онтология позволит не только определить термины предметной области и их толкование, но и определить утверждения, которые ограничат смысл этих терминов

Онтология, описывающая заданную предметную область, включает в себя совокупность семантически значимых терминов и отношений, а также правила, согласно которым можно строить утверждения об элементах в данной предметной области [1].

Терминологическая система онтологии предметной области может быть представлена в следующем виде (рис. 2).

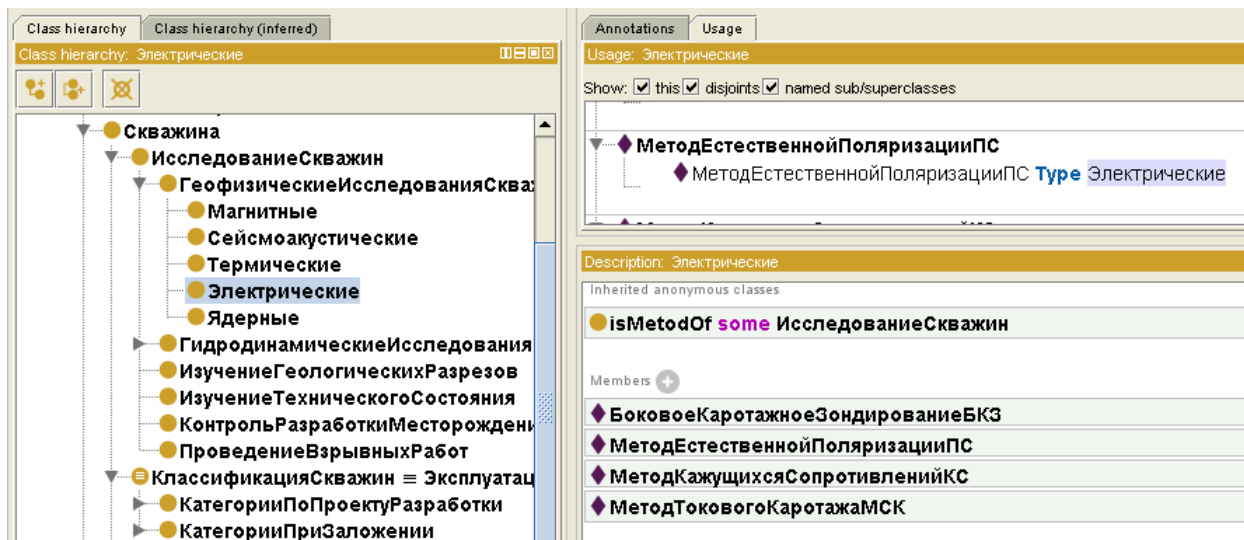


Рис. 2. Концепты онтологии и взаимосвязи между ними

Предлагается следующая последовательность решения задачи построения, использования или модификации объектной модели онтологии для единственной или связанной группы предметных областей.

1. Для определённого понятия заданной предметной области описываются содержащие его процессы. Используется технология построения UML-диаграмм Эриксона-Пенкера (рис. 3). Идентифицированные объекты используются в качестве словаря для построения элемента онтологии.



Рис. 3. Процесс исследования скважин методом каротажа

2. Далее средствами выбранной CASE-системы описываются идентифицированные объекты (или пополняется описание для объектов уже существующих в системе) и находится их место в разрабатываемой объектной структуре, описывающей заданную предметную область.
3. Пополняется при необходимости объектная диаграмма базисных типов сущностей заданной предметной области.
4. Принимается решение пополнения онтологии рассматриваемой системы, и алгоритм принятия решения по формированию класса онтологии фиксируется в виде соответствующих UML-диаграмм. На основании данных предыдущих пунктов определяются свойства нового класса онтологии, которые описываются средствами редактора онтологий в виде аксиом, использующих объекты словаря процесса и соответствующих отношений. При необходимости формируются классы описания свойств основного понятия.
5. Задаётся семантика сформированного класса онтологии в виде URI-ссылок на источники профессионального описания смысла нового класса онтологии. При необходимости описывается семантика введённых классов – свойств.

Для сформированной онтологии определяется структура и содержание основных аксиом и запросов и описывается базовая структура её концептов.

Рассмотрим фрагмент разработки онтологии на примере классификации типов скважин. На нефтяных месторождениях бурят следующие типы скважин: вертикальная, горизонтальная, наклонно-направленная. Выделенные концепты в виде онтологии представлены на рис 4. Для концепта «Тип скважины» характерны следующие свойства: типом может быть либо вертикальная, либо горизонтальная, либо наклонно-направленная. Поэтому вводим функциональное свойство «имеет тип» и описываем с его помощью следующее ограничение.

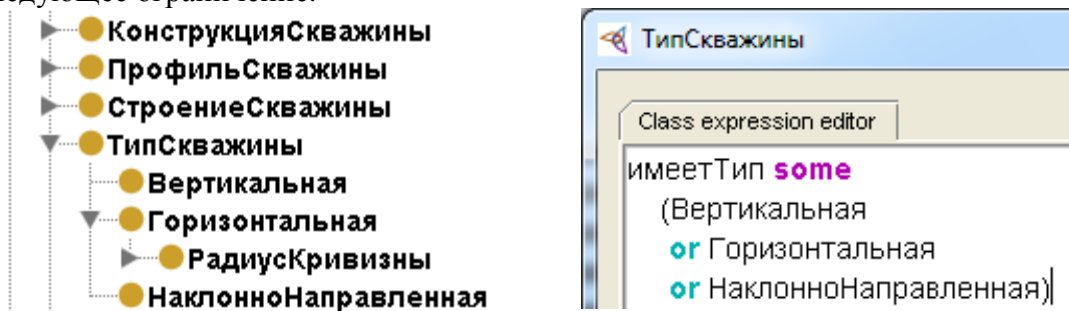


Рис. 4. Концепты онтологии и ограничения для концепта «Тип скважины»

Разработанная онтология может давать ответы на введенные разработчиком запросы, основанные на синтаксисе Манчестер OWL, извлекающие данные на основе сбора всей информации о конкретном классе, свойстве или экземпляре класса. Например, результатом выполнения запроса «какая скважина имеет минимум 2501 м», будет «сверхглубокая» (рис. 5).

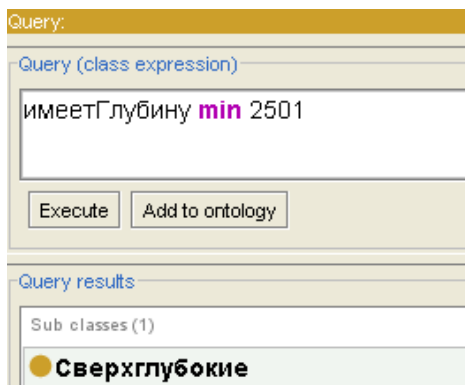


Рис. 5. Результат выполнения запроса

Финальным этапом построения и исследования полученной онтологии является определение её места в структуре базовых геоинформационных систем, содержащих данные в разрезе исследуемой предметной области.

Статья написана в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы», № гос. контракта 14.514.11.4113.

Список литературы

1. Гаврилова Т.А. Инженерия знаний для преподавателей: модели, методы, инструменты [Электронный ресурс]. – Режим доступа: <http://www.mirbis.ru/data/File/rabo/Gavrilova.pdf> (дата обращения 01.11.2013).
2. Клещев А.С., Артемьева И.Л. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия «онтология» // Научно-техническая информация, Серия. 2. – 2001. - № 2. – С. 20–27.
3. Кролл П. Rational Unified Process – это легко. Руководство по RUP / П. Кролл, Ф. Кратчен. – М.: КУДИЦ-ОБРАЗ, 2004. – 432 с.
4. Мартынов В.В., Рыков В.И., Филосова Е.И., Шаронова Ю.В. Применение методов и средств онтологического анализа для управления образовательной деятельностью // Вестник УГАТУ: Сб-к. науч. трудов. – Уфа, УГАТУ, 2012, Т.16. - №3 (48). – С. 230–234.
5. Портал доступа к картам и облегченным приложениям для нефтегазовой отрасли (по материалам Esri) [Электронный ресурс]. – Режим доступа: http://dataplus.ru/news/arcreview/detail.php?ID=10955&SECTION_ID=286 (дата обращения 27.10.2013).
6. Тамм Б.Г. и др. Анализ и моделирование производственных систем. – М.: Финансы и статистика, 1987. – 191 с.
7. Naming and Addressing: URIs, URLs, ... [Электронный ресурс]. – Режим доступа: <http://www.w3.org/Addressing/> (дата обращения 24.10.2013).
8. Protégé [Электронный ресурс]. – Режим доступа: <http://protege.stanford.edu/> (дата обращения 27.10.2013).
9. UML [Электронный ресурс]. – Режим доступа: <http://www.uml.org/> (дата обращения 27.10.2013).
10. Web Ontology Language (OWL). Guide Version 1.0 [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/2002/WD-owl-guide-20021104> (дата обращения 27.10.2013).

Рецензенты:

Черняховская Л.Р., д.т.н., профессор, ФГБОУ ВПО Уфимский государственный авиационный технический университет, г. Уфа.

Токарев Д.В., д.т.н., профессор, ФГБОУ ВПО Уфимский государственный авиационный технический университет, г. Уфа.