

УДК 622.276:004.82

РАЗРАБОТКА ТЕХНОЛОГИИ ОНТОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ ОПИСАНИЯ ПРОЦЕССОВ НЕФТЕДОБЫЧИ

В. В. МАРТЫНОВ¹, Л. П. ФАНДРОВА², Е. И. ФИЛОСОВА³, Ю. В. ШАРОНОВА⁴

¹martynov@rb.ru, ²fandrova@yandex.ru, ³filosova@yandex.ru, ⁴hedviga@mail.ru

ФГБОУ ВПО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступила в редакцию 11.11.2013

Аннотация. Представлена информационная технология описания свойств скважин при нефтедобыче в виде онтологии и применения полученных результатов для решения задачи эффективного поиска научной информации по указанному направлению в интернет-источниках.

Ключевые слова: онтология; база знаний; объектно-ориентированный подход; нефтедобыча; концептуальный граф; DL-логика.

Междисциплинарный характер и быстрое развитие нефтедобывающей отрасли приводит к существенному росту объема научно-технической информации. Трудности адекватного описания взаимозависимости и взаимодействия комплексных структур при недропользовании – области и проблемы терминологической путаницы, вызванной особенностями языка описания задач в конкретной предметной области, существенно усложняют проблему поиска требуемой информации. Как следствие, требуется тематическая, терминологическая и структурная систематизации данных.

В статье рассматриваются вопросы формирования онтологии, позволяющей организовать систему доступа к информационным ресурсам, описывающей методы и технологии, используемые в организациях, участвующих в исследованиях и разработках в области нефтедобычи.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОПИСАНИЯ ОБЪЕКТОВ НЕФТЕДОБЫЧИ

Для поиска информации используются специализированные поисковые серверы: поисковые системы, тематические интернет-каталоги, системы мета-поиска и т.д. Несмотря на существенные сдвиги в области машинной интерпретации и реализации поисковых запросов, основ-

ной проблемой при обращении к интернету является либо пустой список, либо поиск вручную требуемых данных в необозримом объеме найденных записей.

Основной причиной вышеприведенного результата является неэффективная работа поисковых систем, не учитывающих семантику запросов, так как основным методом поиска является использование технологии индексации словарного состава источников при полном отсутствии средств анализа смысла представленной в них информации [1].

Релевантная модель оценки соответствия содержания исследуемого документа поисковому запросу во многих случаях не справляется с решением задачи распознавания и поиска омонимов (грамматических и особенно лексических), синонимов и многозначных слов. Это обусловлено тем, что в основу релевантной модели поиска заложен лингвистический подход и ряд оценочных синтетических критериев, например, положение слов на странице, которые не могут быть распознаны без понимания смысла поискового запроса.

С другой стороны, в рамках теории семантического пространства (пространства понятий) оценка информационного соответствия между одним документом (поисковым запросом) и другим (исследуемым документом) есть проекция пространства понятий запроса на пространство понятий документа. Чем существеннее данная проекция, тем больше смысл исследуемого документа соответствует смыслу поискового запроса [2]. Семантический поиск определяется как поиск по содержательным аспектам

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации по государственному контракту от 10.10.2013 г. № 14.514.11.4098 в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы».

всех компонентов документов электронной библиотеки.

Существующая система использования природных ресурсов в России находится в ведении Министерства природных ресурсов и экологии Российской Федерации. Функции по оказанию государственных услуг и управлению государственным имуществом в сфере недропользования осуществляет Федеральное агентство по недропользованию (Роснедра). Перечисленные ресурсы содержат уникальную подборку материалов в виде статей, баз данных и каталогов, словарей и т.д. Однако описание объектов нефтедобычи достаточно затруднено по ряду причин. Сложность задач, связанных с описанием скважин при нефтедобыче, заключается в том, что их физические и механические свойства зависят не только от химического и фазового составов пластов, но и от особенностей процессов, протекающих в них.

В рамках стандартного подхода описать объекты нефтедобычи достаточно сложно. Рассмотрим объект «скважина при нефтедобыче», представленный на рис. 1. На схеме представлены различные методы и свойства описываемого объекта. Однако отсутствует информация о методах исследования и особенностях его структуры. Высокая структурная сложность такого объекта не позволяет полностью и доступно описать конкретную скважину при нефтедобыче в рамках традиционной многомерной схемы.



Рис. 1. Схема концепта

В определенной степени можно решить проблему качественного описания свойств и структуры скважины путем использования соответствующего тезауруса. Тезаурус – это особая разновидность словаря, в котором указаны семантические отношения между лексическими единицами [3]. В отличие от толкового словаря, тезаурус позволяет выявить смысл не только с помощью определения, но и посредством соотнесения слова с другими понятиями, благодаря чему может использоваться в системах искусственного интеллекта. Структура тезауруса может быть выстроена на основе концептуального графа, который отражает семантику предметной области в виде понятий и отношений и является

одним из наиболее удобных способов представления знаний [4]. Концептуальный граф или концепт-карта – это графический инструмент для организации и представления знаний. Концепт-карты можно отнести к так называемым «легким» онтологиям, у которых отсутствует аксиоматическая составляющая. Онтология концепт-карты $Ok = (C, R)$, где C – конечное множество концептов предметной области, которую представляет онтология, R – множество отношений между концептами [5].

В концептуальном графе роль вершин выполняют понятия, формируемые тезаурус, а связи описывают отношениями между понятиями (рис. 2).

Описание скважины при нефтедобыче по представленной схеме, с сожалением, не позволяет в полной мере представить свойства объекта в области недропользования, так как свойства скважин на концептуальном графе выделены в отдельный класс и не «привязаны» по сути к процессам, происходящим при бурении и исследовании скважин. Кроме этого, представленные в виде класса свойства с течением времени могут изменяться, следовательно, представленная технология визуализации отношений между различными концептами недостаточна. Существует необходимость управлять созданием, распространением, обработкой и использованием структуры описания сложных объектов в рассматриваемой предметной области.

Одним из вариантов решения задачи описания системы недропользования является построение онтологии конкретной предметной области, например, касающейся нефтедобычи с целью моделирования изучаемого объекта в виде формализма. Онтология предметной области есть явное представление концептуализации данной области, совокупность соглашений по ее описанию [6]. 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 сформировать утверждения, которые ограничат смысл этих терминов.

ПРИМЕНЕНИЕ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ПОДХОДА К ОНТОЛОГИЧЕСКОМУ ИНЖИНИРИНГУ

Формальная модель онтологии состоит из терминов (понятий), их определений и атрибутов, а также связанных с ними аксиом и правил вывода.

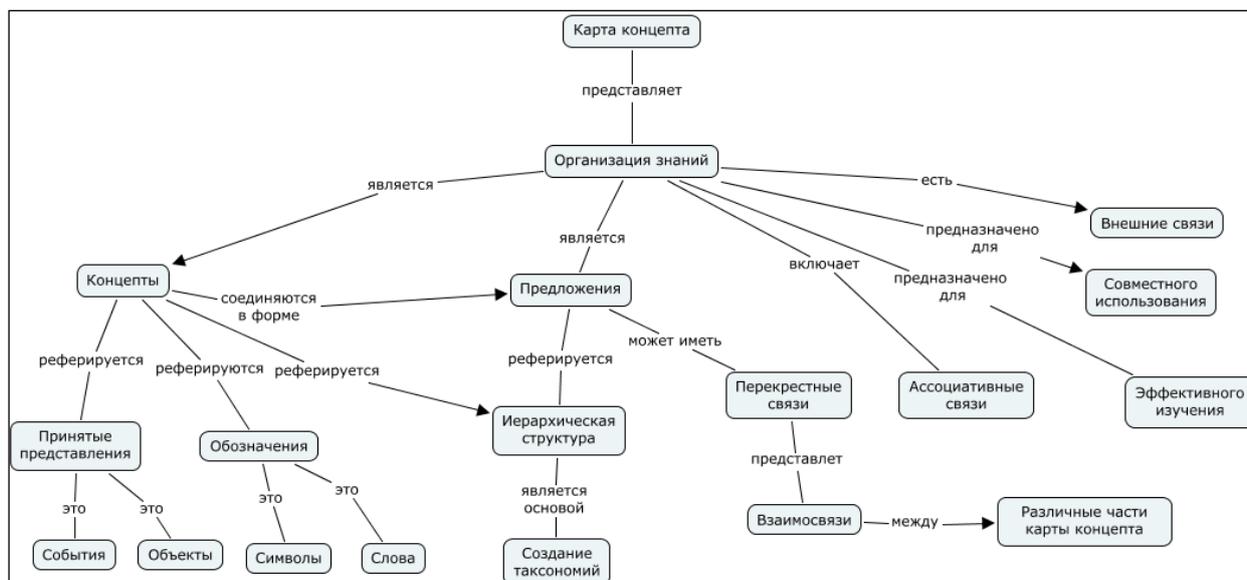


Рис. 2. Концептуальный граф

Формальная модель онтологии $O = \langle K, R, F \rangle$ – это упорядоченная тройка конечных множеств [8], где

K – концепты предметной области (ПрО), которую описывает онтология O ;

R – отношения между концептами заданной ПрО;

F – функции интерпретации, заданные на терминах и/или отношениях онтологии O .

Визуальные методы проектирования онтологий способствуют более быстрому и полному пониманию структуры знаний предметной области, что особенно ценно для исследователей, осваивающих новую предметную область.

Для реализации системы описания объектов нефтедобычи предлагается применить объектно-ориентированный подход (ООП) [7], сущность которого состоит в накоплении и многократном использовании автономных и отлаженных единиц описания материалов и технологий, оформленных в виде классов. Важным свойством ООП является универсальность, то есть любые данные можно представить как объекты и обрабатывать единообразно.

Объектно-ориентированный подход используется как часть общей методологии реализации и развития информационных систем в рамках рационального унифицированного процесса (РУП). Применение технологии онтологического инжиниринга на базе сквозного объектно-ориентированного подхода в рамках РУП позволяет определить функции системы описания

и визуализации структуры объектов области недропользования на примере нефтедобычи.

Заказчик, представленный специалистом или группой специалистов в области недропользования, в понятных ему терминах формирует комплекс требований на языке объектно-ориентированного моделирования, реализованный, например, средствами программы моделирования Enterprise Architect. Исполнитель в рамках комплекса предъявленных требований строит объектную модель области недропользования на примере нефтедобычи, которая описывает набор концептов, необходимый для однозначного понимания требований Заказчика. Так как отбор концептов осуществляется в диалоге между Исполнителем и Заказчиком, то процесс построения концептуального графа и модели онтологии имеет циклический характер, вызванный процедурой взаимного пополнения объема знаний.

В результате Исполнитель формирует требуемое описание функций онтологии в терминах Заказчика. Затем Исполнитель разрабатывает тезаурус предметной области в виде концептуального графа с согласованием у Заказчика семантических взаимосвязей между отобранными объектами. Средством создания онтологий может служить редактор Protege версии 4.1 и выше. Опишем технологию разработки онтологий объектов и процессов нефтедобычи в виде модели, представленной на рис. 3.

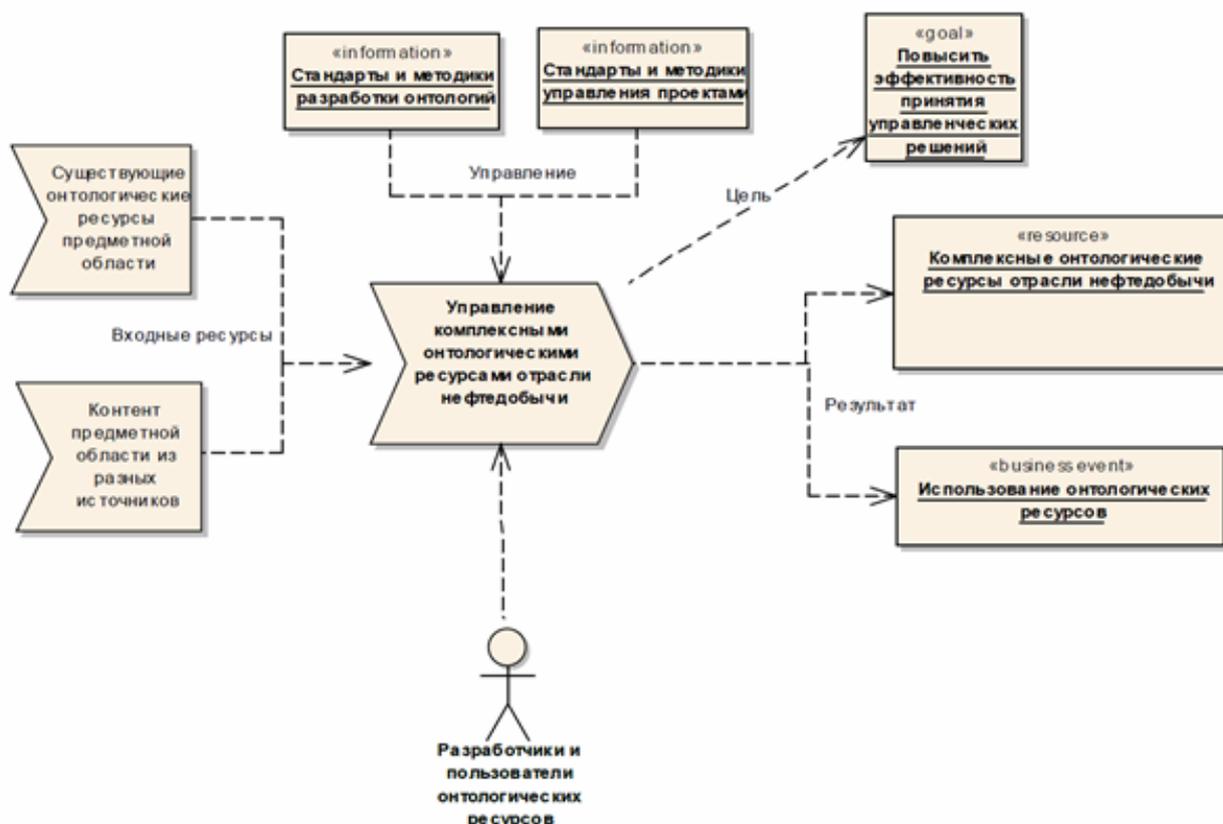


Рис. 3. Концептуальная диаграмма управления комплексными онтологическими ресурсами процесса нефтедобычи

СРЕДСТВА ОПИСАНИЯ КОНЦЕПТОВ ОНТОЛОГИИ

Используя онтологию как базовую модель знаний, можно описать систему оперирования этими знаниями на основе какой-либо формальной системы.

В созданной онтологии аксиомы (ограничения) задаются по правилам дискреционных логик первого порядка для каждого класса. На рис. 4 показаны вхождения концепта «Самотлорское» в онтологию и аксиомы, связывающие его с концептами «Глубина залегаения», «Нефтегазоносный бассейн», «Запасы нефти», «Год открытия».

Дескриптивная логика DL [4] как формальная система представляет собой следующее множество:

$$DL = \langle A, P, T \rangle,$$

где A – множество аксиом, P – множество правил вывода и T – множество теорем.

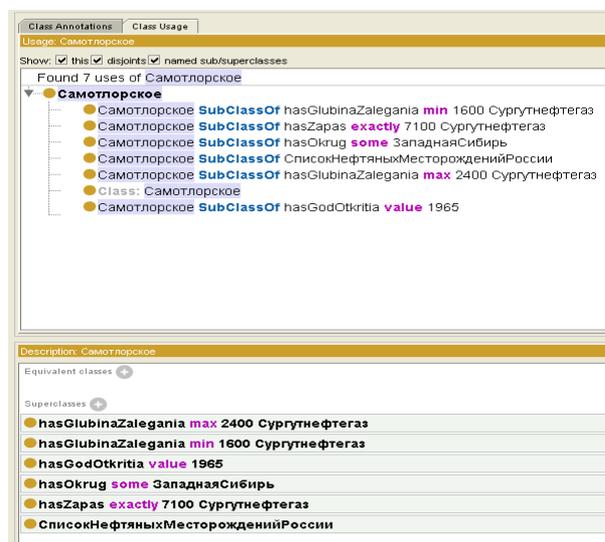


Рис. 4. Набор аксиом для концепта «Самотлорское месторождение»

Множество аксиом A описывает определение концептов (терминологические аксиомы) и аксиомы существования отношений между концептами. Для онтологии «Нефтедобыча» применялись следующие схемы отображения:

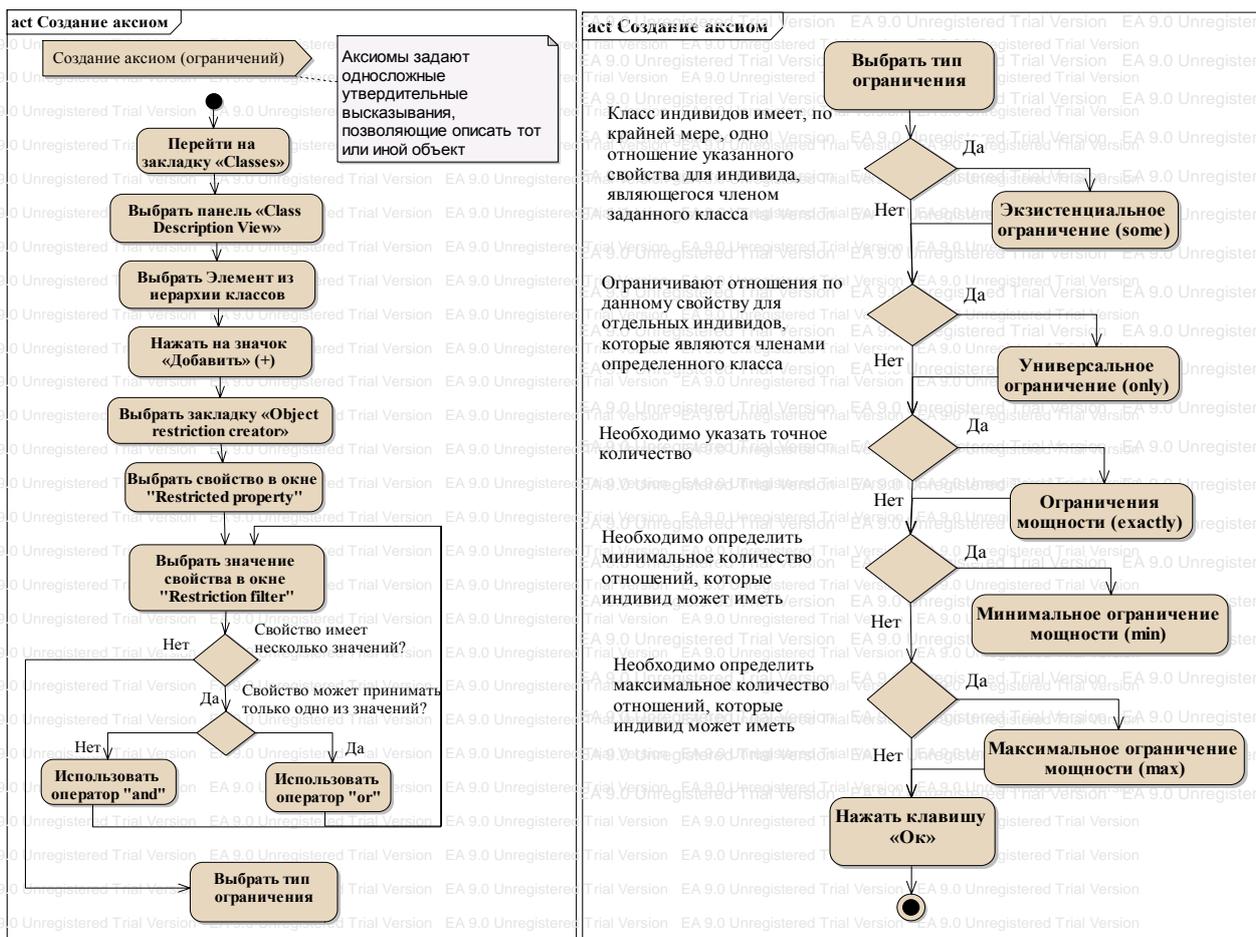


Рис. 5. Алгоритм создания аксиом

1. Всеобщий класс – элемент Thing – имеет семантику Δ ;
2. Класс – элемент Class – имеет семантику $C \subseteq \Delta$;
3. Объединение нескольких свойств – элемент and – имеет семантику $C_3 = C_1 \cap C_2$;
4. Принадлежность хотя бы одному из нескольких свойств – элемент or – имеет семантику $C = \{ C_1, C_2, \dots \}$;
5. Свойство принадлежности классу одного из подклассов – элемент some – имеет семантику $C = \{ p \mid \exists q \langle p, q \rangle \in R \wedge q \in C_1 \}$.
6. Свойство принадлежности классу только одному из подклассов – элемент only – имеет семантику $C_3 = C_1 \cup C_2$.
7. Свойство эквивалентности двух концептов – элемент \equiv – имеет семантику $C_1 = C_2$;
8. Минимальное допустимое значение – элемент min – имеет семантику $C = \{ p \mid \# \{ q \mid \langle p, q \rangle \in R \} \geq n \}$;

9. Максимальное допустимое значение – элемент max – имеет семантику $C = \{ p \mid \# \{ q \mid \langle p, q \rangle \in R \} \leq n \}$ и др.

Алгоритм создания аксиом можно представить в виде диаграммы деятельности, представленной на рис. 5.

Работа с онтологической базой знаний подразумевает использование правил логического вывода. Для этого существует специальный инструмент – блок логического вывода (reasoner). Эта подсистема в Protege позволяет проверить истинность, ложность или неопределенность некоторой аксиомы в рамках заданной онтологии. Также он используются для построения полного набора аксиом на базе существующей онтологии и вывода неявных знаний по рассматриваемой предметной области.

Так как онтология включает не только словарь терминов и понятий предметной области, но и логические выражения (связи), описывающие соотношения друг с другом, то она счита-

ется базой знаний специального вида, которую можно применять самостоятельно в рамках рассматриваемой предметной области. Для ее использования пользователь должен сформулировать запрос из существующих в базе знаний понятий и связей между ними, переводя его на формальный язык логики, и получить однозначный ответ, если таковой существует.

В Protege отбирать необходимые данные можно тремя способами:

1. Средством построения онтологических графов OntoGraf для создания простых запросов типа «вывести всех индивидов определенного класса»;

2. С помощью закладки DL Query, позволяющей получать сведения из проекта по всем экземплярам классов, которые удовлетворяют интересующим критериям;

3. Посредством SPARQL-запросов, используемых для поиска информации в онтологии по более сложным критериям (свойствам объекта, меткам, комментариям).

Фрагмент приведенной онтологии позволяет комплексно описать явные и неявные знания по рассматриваемой области нефтедобычи (рис. 6).

Онтология объектов и процессов нефтедобычи позволяет средствами логических запросов получать новые знания в указанной области. Результаты могут быть оформлены в виде новых объектов как следствие визуального анализа структуры предметной области, либо как результаты выполнения семантических запросов.

ЗАКЛЮЧЕНИЕ

В данной статье представлены результаты поисковой НИР по теме «Структуризация, комплексный анализ и представление алгоритмов онтологической организации комплекса научных, технических и технологических знаний в области недропользования на примере нефтедобычи», выполненной в рамках Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы», номер государственного контракта 14.514.11.4098.

Созданная на основе представленного подхода онтология области «Нефтедобыча» обеспечивает возможность публичного, отраслевого, или корпоративного оперирования информационными ресурсами с большим количеством данных. С ее помощью организована система доступа к пространственным ресурсам, описывающая методы и технологии, используемые в организациях, участвующих в исследованиях и разработках в области недропользования для кластеризации тематических направлений исследований и выявления эффективных, конкурентоспособных и патентоспособных научных разработок.

Сформированная онтология является основой для построения web-сайта онтологии организации пространственных данных в области недропользования, являющегося информационным зеркалом указанной онтологии. Данный сайт позволит реализовать интернет-навигацию по разработанной базе знаний.

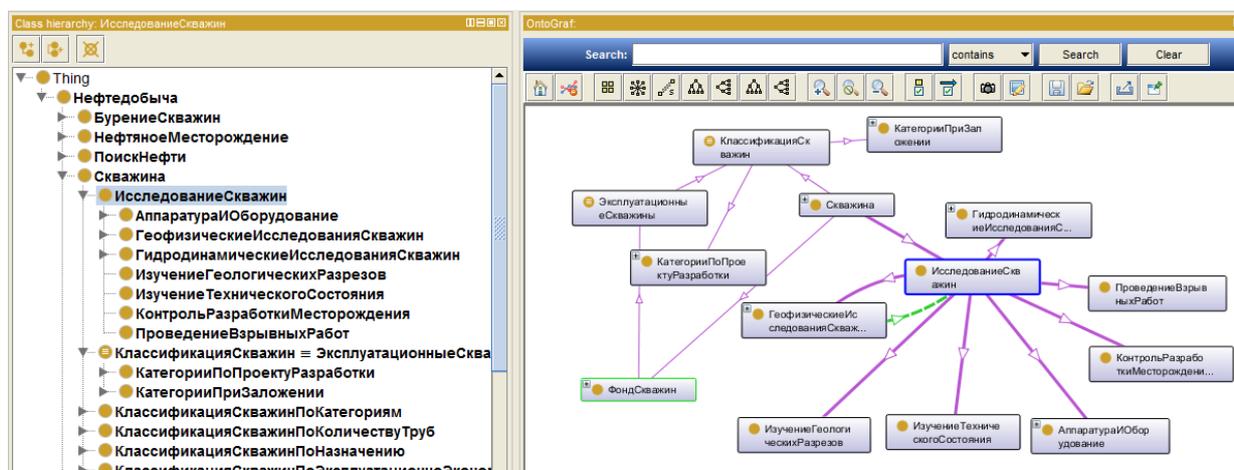


Рис. 6. Концепты онтологии и взаимосвязи между ними

СПИСОК ЛИТЕРАТУРЫ

1. **Боровикова О. И., Загоруйко Ю. А.** Организация порталов знаний на основе онтологий // Компьютерная лингвистика и интеллектуальные технологии: тр. Междунар. сем. «Диалог'2002». М.: Наука, 2002. Т. 2. С. 76–82.
2. **Зацман И. М.** Семантический поиск научной информации: неоднородные коммуникативные компоненты и цветовая палитра объектов поиска // Компьютерная лингвистика и интеллектуальные технологии: тр. Междунар. сем. «Диалог'2002». М.: Наука, 2002. Т. 2. С. 214–227.
3. **Большой** энциклопедический словарь (БЭС) [Электронный ресурс]. URL: <http://www.slovopedia.com> (дата обращения 27.10.2013).
4. **Шустова Д. В.** Разработка тезауруса прикладной онтологии. // Аспирантский вестник Поволжья. 2010. № 1–2. С. 108–110.
5. **Муромцев Д. И.** Концептуальное моделирование знаний в системе Smart Tools. СПб.: СПбГУ ИТМО, 2009. 83 с.
6. **Гаврилова Т. А.** Инженерия знаний для преподавателей: модели, методы, инструменты. [Электронный ресурс]. URL: <http://www.mirbis.ru/data/File/rabo/Gavrilova.pdf> (дата обращения 10.10.2013).
7. **Мартынов В. В., Рыков В. И., Филосова Е. И., Закиева Е. Ш.** Методы объектного подхода в управлении учебным процессом. Уфа: УГАТУ, 2010. 167 с.

ОБ АВТОРАХ

МАРТЫНОВ Виталий Владимирович, проф., зав. каф. экон. информатики. Дипл. инж.-мех. (Марьинск. политехн. ин-т, 1981). Д-р техн. наук авт. сист. управления (УГАТУ, 2000). Иссл. в обл. проектир. инф. систем.

ФАНДРОВА Людмила Петровна, доц. каф. экон. информатики. Дипл. инж.-системотехн. (УГАТУ, 1985). Кан. техн. наук по электротехн. комплексам и системам (УГАТУ, 2003). Иссл. в обл. разработки онтологич. систем.

ФИЛОСОВА Елена Ивановна, доц. каф. экон. информатики. Дипл. инж.-прогр. (УГАТУ, 1993). Кан. техн. наук по упр. в соц.-экон. системах (УГАТУ, 2008). Иссл. в обл. проектир. соц.-экон. систем.

ШАРОНОВА Юлия Вениаминовна, доц. каф. экон. информатики. Дипл. инж.-прогр. (УГАТУ, 1993). Кан. соц. наук по социологии управления (БГУ, 2006). Иссл. в обл. управления соц.-экон. системами.

METADATA

Title: Development of the technology of ontological analysis to describe the processes of oil production.

Authors: V. V. Martynov¹, L. P. Fandrova², E. I. Filosova³, J. V. Sharonova⁴

Affiliation: Ufa State Aviation Technical University (UGATU), Russia.

Email: martynov@rb.ru, fandrova@yandex.ru, filosova@yandex.ru, hedviga@mail.ru.

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 17, no. 5 (58), pp. 188-194, 2013. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: Information technology is presented describing the properties of wells in oil production in the form of ontologies and use the results to solve the problem of effective search of scientific information in the indicated direction to Internet sources.

Key words: ontology; knowledge base; object-oriented approach; oil production; conceptual graph; DL-logic.

References (English transliteration):

1. O. I. Borovikova, Y. A. Zagorulkov, "The organization of knowledge portals based on ontologies," *Computational Linguistics and Intelligent Technologies: Proc. Int. Seminar Dialog'2002*, vol. 2, pp. 76-82, Moscow: Nauka, 2002,
2. I. M. Zatsman, "Semantic search of scientific information: heterogeneous communication components and color objects of the search," *Computational Linguistics and Intelligent Technologies: Proc. Int. Seminar Dialog'2002*, vol. 2, pp. 214-227, Moscow: Nauka, 2002.
3. *Large Encyclopedic Dictionary* (2013, Oct. 27). *Semantic search* [Online], (in Russian). Available: <http://www.slovopedia.com>
4. D. V. Shustova, "Development of a thesaurus application on-pathology," *Journal of Postgraduate Volga*, no. 1-2, pp. 108-110, 2010.
5. D. I. Muromtsev, *Conceptual modeling of knowledge in the Smart Tools*, (in Russian). St. Petersburg: St. Petersburg State University of Mechanics and Optics, 2009.
6. T. A. Gavrilova, (2013, Oct. 10). *Semantic search* [Online], (in Russian). Available: <http://www.mirbis.ru/data/File/rabo/Gavrilova.pdf>.
7. V. V. Martynov, V. I. Rykov, E. I. Filosova, and E. I. Zakieva, *The methods of the object approach in the management of the learning process*, (in Russian). Ufa: Ufa State Aviation Technical Univ., 2010.

About authors:

MARTYNOV, Vitaliy Vladimirovich, Prof., Dept. of Economic Informatics. Dipl. engineering-mechanics inequality (Mar'ins'kyi. watered. Inst, 1981). Dr. of Tech. Sci. (UGATU, 2000).

FANDROVA, Lyudmila Petrovna, docent, Dept. of Economic Informatics. Dipl. Systems Engineer (USATU, 1985). Cand. of Tech. Sci. (UGATU, 2003).

FILOSOVA, Elena Ivanovna, docent, Dept. of Economic Informatics. Dipl. Software Engineer (USATU, 1993). Cand. of Tech. Sci. (UGATU, 2008).

SHARONOVA, Julia Veniaminovna, docent, Dept. of Economic Informatics. Dipl. Software Engineer (USATU, 1993). Cand. soc. science (BSU, 2006).