УДК 004.81+61

ПРОБЛЕМАТИКА МЕХАНИЗМА ФУНКЦИОНИРОВАНИЯ ИЕРАРХИЧЕСКОЙ ОНТОЛОГИИ ПОНЯТИЙ О КАЧЕСТВЕ ДЕЯТЕЛЬНОСТИ IT-СПЕЦИАЛИСТА МЕДИЦИНСКОЙ ОРГАНИЗАЦИИ

Газизова М.Д. [1], Гольдштейн С.Л. [1], Грицюк Е.М. [2], Донцов О.Г. [2], Дугина Е. А. [2]

[1] ФГАОУ ВПО УрФУ, г. Екатеринбург

[2] ГАУЗ СО МКМЦ «Бонум», г.Екатеринбург

В статье приведен обзор проблем, связанных с механизмом функционирования иерархической онтологии понятий о качестве деятельности IT-специалиста медицинской организации. Рассмотрены способы оценки программных средств для создания онтологий, методов отбора релевантных терминов, определения типов связей между терминами, опций и их контента для вершин и дуг онтологии, когнитивных маршрутов и решения задачи о качестве деятельности, описываемой создаваемой онтологией. Приведены аналоги, лучшие из которых включены в компилятивный прототип. Выдвинуты гипотезы о возможности его совершенствования.

Ключевые слова: онтологии, иерархическая структура, построение и анализ онтологий, программное средство для создания онтологий, аналоги, прототипы, гипотезы.

The problems of the functioning of hierarchical ontologies of concepts about quality of work of an it-professional medical organizations

Gazizova M. D. [1], Goldstein S. L. [1], Gritsyuk E. M. [2], Dontsov O. G. [2], Dugina E. A. [2]

[1] Urals Federal University, Ekaterinburg, Russia
[2] State Autonomic Health Institution of the Sverdlovsk Region Multiprofile Clinical Medical Center «BONUM» Ekaterinburg, Russia

The review of the problems connected with the mechanism of functioning of hierarchical ontology of concepts about quality of activity of the IT specialist of the medical organization is provided in article. Ways of assessment of software for creation of ontologies, methods of selection of relevant terms, definitions of types of communications between terms, options and their content for tops and arches of ontology, cognitive routes and the solution of a task on quality of the activity described by the created ontology are considered. Analogs, the best of which are included in a kompilyativny prototype, are given. Hypotheses of a possibility of his improvement are made.

Keywords: ontology a hierarchical structure, organization and analysis of ontologies, a software tool for creating ontologies, analogues, prototypes, hypotheses.

Введение

В настоящее время онтологии широко используют в инженерии знаний, информационном поиске, семантической интеграции информационных ресурсов и в других направлениях. Сформированные на основе онтологий когнитивные системы продуктивны в различных сферах деятельности, в том числе, в здравоохранении [1 - 4]. За несколько десятков лет этой теме посвящено большее число исследований, например [5 - 6], поскольку преимущество онтологий – их формальная структура, упрощающая компьютерную обработку, и онтологический метод представления знаний, позволяющий приложениям распознавать те семантические отличия, которые очевидны человеку, но не понятны компьютеру. Однако, несмотря на это, задача онтологического моделирования до сих пор актуальна.

Кроме построения, работа с онтологиями включает в себя наполнение вершин и дуг типа «часть-целое», а также дуг других типов, структурированной информацией. В известных решениях существует возможность писать лишь комментарий к вершине или дуге объемом от одного слова до краткого текста. Однако, способа закрепить структурированный текст нет.

В статье поставлена и решена задача анализа источников информации по теме, выбора и анализа аналогов, выхода на пакет научных прототипов, их критики и генерирования гипотез о ее парировании.

Заказчик текущей работы – учреждение системы здравоохранения ГАУЗ СО – многопрофильный клинический медицинский центр «Бонум», г. Екатеринбург.

Обзор информации и отбор аналогов

Обзор выполнен по шести темам: компьютерные программы построения онтологий, методы отбора релевантных терминов, средства спецификации опций вершин и дуг, оценки качества, настройка онтологии на специфику предметной области.

Компьютерные программы построения онтологии

Для построения онтологии, в частности, понятий о качестве деятельности IT-специалиста медицинской организации, требуется наиболее подходящая компьютерная программа, позволяющая строить, анализировать и представлять необходимую информацию. В качестве таковых нами рассмотрены следующие инструменты:

- Protégé – одно из наиболее известных программных средств для работы с онтологиями [7], написанных на языке JAVA и предназначенных для создания, редактирования, анализа и просмотра онтологий моделей предметной области, способных

генерировать формы получения знаний для введения экземпляров классов и подклассов, имеющее удобный графический пользовательский интерфейс;

- OntoEdit выполняет проверку, просмотр, кодирование и модификацию онтологий [8]. Преимущества данного программного продукта – разработка онтологий на методологической базе, с помощью логического вывода, аксиом, расширяемой структуры плагинов и т.д;

- Ontolingua предназначена для поддержки формального специфицирования задач пользователя на основе библиотеки формальных описаний фрагментов задач, моделей и понятий [9]. Пользователь системы может быть удаленным и взаимодействовать с инструментом через Internet. В работе [10] представлен обзор инструментов [7-9] для построения и отображения, объединения онтологий, а также аннотирования на основе онтологий;

- Fluent Editor 2 – инструмент для редактирования и обработки онтологий с использованием естественного языка. Редактор содержит встроенный словарь [11];

- TopBraid Composer – редактор для создания семантических приложений [12]. Статья [13] содержит краткое резюме инструментов построения онтологий [7, 11, 12] относительно их особенностей;

- SWOOP – инструмент просмотра и редактирования, с функциями автоматического разделения классов и извлечения модулей. Кроме того, имеет широкий спектр для выбора отображения объектов [14];

- KNOODL – это распределенная система, предназначенная для управления информацией, содержащая инструменты для создания, управления, анализа и визуализации описания [15];

- OntoBuilder – универсальный инструмент, поддерживающий извлечение онтологий из web-поисковых интерфейсов [16].

В таблице 1 приведены результаты нашего анализа аналогов для построения онтологий предметной области, а в таблице 2 – значения численных оценок таблицы 1.

Таблица 1

Сравнение программ автоматизированного построения онтологии

Критерии сравнения	Оценки аналогов по источникам информации:							
	[7]	[8]	[14]	[11]	[12]	[9]	[16]	[15]
стоимость	1,0	0,6	1,0	0,7	0,6	0,0	0,0	0,0
доступность	1,0	0,6	0,6	0,6	0,6	0,6	0,6	0,0
дата выпуска последней версии	1,0	0,7	0,9	1,0	0,0	0,2	0,6	0,7
возможность наполнения вершин / дуг	0,6	0,0	0,0	0,0	0,6	0,0	0,6	0,0
возможность построения иерархии	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
расширяемость	1,0	1,0	0,1	0,0	0,1	0,0	1,0	1,0
итоговая оценка	5,6	3,9	3,6	3,3	2,9	1,8	3,8	2,7

Таблица 2

Численные значения оценок

Критерий	Значения оценок
стоимость	1,0 – бесплатный программный продукт; 0,7 – бесплатно для некоммерческого использования; 0,4 – платная версия, бесплатная имеет ограничения; 0,1 – платный программный продукт; 0,0 – нет данных
доступность	1,0 – открытый исходный код; 0,6 – открытый доступ / закрытый исходный код; 0,2 – закрытый доступ; 0,0 – нет данных
дата версии	1,0 – 2016 г.; 0,93 – 2015 г.; 0,72 – 2012 г.; 0,65 – 2011 г.; 0,2 – 2002 г.
возможность наполнения вершин / дуг	1,0 – возможность наполнения вершин и дуг; 0,6 – возможность наполнения вершин или дуг; 0,1 – нет возможности наполнения; 0,0 – нет информации
возможность построения иерархической онтологии	1,0 – есть; 0,1 – нет; 0,0 – нет информации
расширяемость	1,0 – есть возможность расширения функциональности с помощью плагинов; 0,1 – нет возможности расширения функциональности с помощью плагинов; 0,0 – нет информации

В результате по совокупности показателей программное средство Protégé представляется нам предпочтительным.

Методы отбора релевантных терминов для вершин онтологии

Существует множество сервисов для определения релевантности (семантического соответствия текста ответа запросу) страницы в интернете [17 - 20], а также метод экспертной оценки [21]. В таблице 3 представлено сравнение методов отбора релевантных вершин по критериям:

- анализа текста / web-страниц – на релевантность в различных источниках информации;

- поиска по ключевым словам;

- анализа плотности ключевых слов – показатель частоты повторения ключевых слов в тексте, рассчитываемое, как процентное отношение количества ключевых слов к общему числу слов в тексте;

- доступности.

Таблица 3

Сравнение методов отбора релевантных терминов для вершин онтологии

Ссылка	Оценки по критериям:				Итоговая оценка
	анализа текста, web-страниц	поиска по ключевым словам	анализа плотности ключевых слов	доступности	
[17]	1,0	1,0	1,0	0,5	3,5
[18]	0,5	1,0	0,0	1,0	2,5
[19]	0,5	1,0	0,0	1,0	2,5
[20]	0,5	1,0	0,0	0,5	2,0
[21]	1,0	1,0	1,0	1,0	4,0

Видно, что экспертный метод [21] дает наиболее точную и развернутую оценку релевантности.

Методы определения дуг типа «часть-целое»

Поскольку выделение связей типа «часть-целое» входит в процедуру отбора релевантных вершин, использование программных средств для определения таких связей нецелесообразно, его можно провести тоже экспертной оценкой [21].

Средства спецификации опций вершин и дуг

Для описания понятий онтологии предметной области с учетом требований к расширяемости и поддерживаемости, предъявляемых к репозиториям технической информации, предпочтительны языки, синтаксис которых близок к естественному при сохранении выразительных возможностей формальной логики, такие как OWL [22]. Он поддерживается многими программными продуктами для работы с онтологиями, в том числе и Protégé. Основные элементы языка OWL: классы (owl:Class); свойства (owl:ObjectProperty или owl:DatatypeProperty) и индивиды – экземпляры классов или свойств.

Для представления частных (авторских) опций вершин и дуг может быть использован один из 6-ти способов описания также на языке OWL [23]: идентификатор класса (URI); перечисление всех экземпляров класса; ограничение на значение свойства; пересечение двух и более определений класса; объединение двух и более определений класса; дополнение (логическое отрицание) определения класса.

Оценки качества онтологии

В работе [24] представлен сравнительный анализ методов оценки качества онтологии предметной области. Программное средство EvaLexon подходит для сравнения терминов из словаря онтологии с текстами на естественном языке, на основе которых была построена онтология. Инструмент также дает оценку, которая показывает, включена ли основная часть понятий входного текста в онтологию. Таким образом формируется перечень наиболее важных понятий, которые относятся к тексту предметной области.

С точки зрения когнитивной эргономики построенной онтологии при оценке воспринимаемости онтологии можно использовать субъективные метрики [25-26]. Для принятия решений предложен алгоритм [27], объединяющий 2 этапа: использование инструментов оценки и принятие решения на основании полученных результатов.

Настройка онтологии на специфику предметной области

Проблематика настройки корпоративной информационной системы на задачи конкретного предприятия на основе полуформализованных моделей известна [28]. Первая задача в ее составе – создание словника по предметной области. Относительно деятельности IT-специалиста источниками информации могут служить, прежде всего, образовательные, профессиональные и специальные стандарты [29-31].

Пакет прототипов

На основе проведенного анализа аналогов нами предложен компилятивный прототип механизма функционирования иерархической онтологии понятий (МФИОП) о деятельности, отнесенный к 0-му рангу. Проведен сравнительный анализ возможностей и характеристик аналогов, учитывая их достоинства и недостатки. Лучшие аналоги систем МФИОП выбраны в качестве прототипов 1-го ранга. В результате анализа также представлена их критика (табл. 4).

Таблица 4
Пакет научных прототипов

Ранг	Наименование		Ссылка	Критика
0	МФИОП		[21-24]	Системно-структурная неполнота
1	Системы:	спецификации опций вершин и дуг	[22-23]	Функциональная неполнота
		оценки качества онтологии	[24-27]	
		настройки на специфику	[28-31]	

Гипотезы о предполагаемых решениях

Для развития МФИОП с учетом задачи о качестве деятельности конкретного специалиста в заданной организации предполагается осуществить следующие нововведения:

- добавить систему настройки на специфику деятельности IT-специалиста медицинской организации;
- модифицировать систему спецификации опций вершин и дуг;
- дополнить систему оценки качества онтологии.

Результаты и выводы

Проведен обзор аналогов по 6-ти темам: компьютерные программы построения онтологий, методы отбора релевантных терминов, средства спецификации опций вершин и дуг, оценки качества, настройка онтологии на специфику предметной области;

1. Предложен пакет научных прототипов
2. Сформированы гипотезы о правках научного прототипа

Сделан вывод о достаточности этого материала для моделирования предложенных решений.

Список литературы

1. Клещев А. С. Математические методы онтологий предметных областей. Часть 1. Существующие подходы к определению понятия "онтология" / А.С. Клещев, И.Л. Артемьева // Научно техническая информация. Серия 2 "Информационные процессы и системы". – 2001. – №2. – С. 20-27.

2. Hymphreys B. L. The Unified Medical Language System: An Informatics Research Colloboration / B. L. Hymphreys, D. A. B. Lindberg, H. M. Schoolman [and other] // J. Am. Medical Informatics Assoc. – 1998. – Vol. 5. – №1. – P. 1-11.

3. Брук Н. Медицинские онтологии для представления текстовой и графической информации / Н. Брук // Revista ştiinţifica a Universităţii de Stat din Moldova "Studia Universitatis". – 2013. – №2 (62). – P. 33-38.

4. Лебедев С.В. Слияние медицинских данных на основе онтологий / С.В. Лебедев, Н.А. Жукова // Онтология проектирования, Т. 7. – 2017. – № 2 (24). – С. 145–159.

5. Горшков С. Введение в онтологическое моделирование / С. Горшков. – Екатеринбург: ТриниДата, 2014-2016. – 165с.

6. Константинова Н.С. Онтологии как системы хранения знаний / Н.С. Константинова, О.А. Митрофанова. URL: http://www.sciinnov.ru/icatalog_new/index.php?action=send_att&entry_id=68352&fname =68352e2-st08_(Митрофанова О.А.).pdf.

7. Noy N. Creating Semantic Web Contents with Protégé-2000. / N. Noy, M. Sintek, S. Decker [and other] // IEEE Intelligent Systems. – 2001. – March/April. – P. 60-71.

8. Sure Y. OntoEdit: Collaborative ontology development for the Semantic Web / Y. Sure, M. Erdmann, J. Angele [and other] // In Proc. of the Inter. Semantic Web Conference. – 2002.

9. Farquhar A., Fikes R., Rice J. The ontolingua server: A tool for collaborative ontology construction / A. Farquhar, R. Fikes, J. Rice // International Journal of Human-Computer Studies. – 1997. – 46(6). – P.707–728.

10. Овдей О.М. Обзор инструментов инженерии онтологий / О.М. Овдей, Г.Ю. Проскудина. URL: http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part4.

11. FluentEditor2 //www.cognitum.eu – 25.03.2013. URL: http://www.cognitum.eu/semantics /FluentEditor2/Default.

12. TopBraid //www.w3.org – 9.11.2011. URL: http://www.w3.org/2001/sw/wiki/TopBraid.

13. Dr Abburu S. Survey on Ontology construction tools / Dr Sunitha Abburu, G. Suresh Babec // Intermotional Jornal of Scientific & Engineering Research. – 2013. – №4(6) – P.1748.

14. Swoop // Онтологический редактор. URL: http://www.softpedia.com/get/Internet/Other-Internet-Related/MIND-lab-SWOOP.shtml.

15. Knoodl // Онтологический редактор. URL: http://www.knoodl.com.

16. Ontobuilder // Онтологический редактор. URL: https://ontobuilder.soft112.com.

17. Majento // Анализ релевантности текста страницы. URL: http://www.majento.ru/ index.php?page=seo-analize/text-analize/index.

18. Megaindex // Оценка релевантности страницы. URL: http://audit.megaindex.ru/audit /relevant_page/getpage.

19. PR-CY // Аудит сайта. URL: http://pr-cy.ru/audit/.

20. SeoLib // Подбор релевантных страниц. URL: https://seolib.ru/tools/site/relevant/.

21. Отношение «часть – целое» в темпоральной/событийной онтологии // habr.com. URL: https://habr.com/post/276987/.

22. Буракова, Е. Е. Языки описания онтологий для технических предметных областей [Текст] / Е. Е. Буракова, Н. М. Боргест, М. Д. Коровин // Вестник Самарского государственного аэрокосмического университета им. С. П. Королева. - Библиогр. - 2014. - № 3. - С. 144-158.

23. OWL Web Ontology language язык представления онтологий в Web // Электронная библиотека. URL: http://kursak.net/owl-web-ontology-language-yazyk-predstavleniya-ontologij -v-web/.

24. Пращук А. В. Анализ методов и алгоритмов оценки качества онтологии предметной области / А. В. Паращук, А. А. Рыбанов // VIII Международная студенческая электронная научная конференция «Студенческий научный форум» - 2016.

25. Гаврилова Т. А. Субъективные метрики оценки онтологий. / Т. А. Гаврилова, В.А. Горовой, Е.С. Болотникова // Знания - Онтологии - Теории: Материалы Всероссийской конференции с международным участием ЗОНТ - 09. - Новосибирск - 2009. - С. 178-187.

26. Гаврилова Т.А. Оценка когнитивной эргономичности онтологии на основе анализа графа / Т. А. Гаврилова, В. А. Горовой, Е. С. Болотникова // Журнал РАН "Искусственный интеллект и принятие решений", 2009. № 03. – С. 33-41.

27. Мохов В. А. Интегрированный алгоритм когнитивной оценки и выбора оптимального варианта онтологической модели / В.А. Мохов, Н.Н. Сильнягин // Электронный журнал «Инженерный вестник Дона». – 2011. – № 11. URL: http://ivdon.ru/magazine/archive/ n4y2011/600

28. Гольдштейн С. Л. Настройка корпоративных информационных систем на задачи предприятия / С. Л. Гольдштейн, И. В. Кашперский // ГОУ ВПО "Уральский гос. технический ун-т - УПИ", НП "Уральский межакадемический союз". - Екатеринбур: Форт Диалог-Исеть, 2006. – 148 с.

29. Федеральный государственный образовательный стандарт среднего профессионального образования по специальности 09.02.04 Информационные системы (по отраслям): утвержден приказом Министерства образования и науки РФ от 14 мая 2014 г. № 525 / Министерство образования и науки Российской Федерации. – Москва: 2014.

30. Белоусова И.Д К вопросу о согласовании требований к содержанию профессиональной подготовки на основе онтологической модели / И.Д. Белоусова, Л.В. Курзаева, А.М. Агдавлетова // Современные наукоемкие технологии. – 2015. – № 11. – С. 67-70.

31. Профессиональные стандарты в области информационных технологий / АП КИТ. М., 2008. – 615 с.

Газизова Мария Дамировна - студент-бакалавр, гр. Фт-440007 ФГАОУ ВО «УрФУ имени первого Президента России Б.Н.Ельцина», г.Екатеринбург, ул.Мира, 19, тел.8 906-809-92-37, e-mail: Mgazizova5@gmail.com