

## МЕДИЦИНСКИЕ ОНТОЛОГИИ ДЛЯ ПРЕДСТАВЛЕНИЯ ТЕКСТОВОЙ И ГРАФИЧЕСКОЙ ИНФОРМАЦИИ

**Наталья БРУК**

*Институт математики и информатики Академии наук Молдовы*

### ONTOLOGII MEDICALE PENTRU REPREZENTAREA DATELOR DE TIP TEXT ȘI DE TIP GRAFIC

În articol sunt cercetate probleme privind crearea ontologiilor, cu ajutorul cărora sunt descrise atât date de tip text, cât și de tip grafic. Se face o trecere în revistă a unei serii de ontologii medicale și a celor de imagini, care ilustrează diversitatea de domenii și de mijloace de lucru cu ele.

**Cuvinte-cheie:** ontologii medicale, cunoștințe, text structurat, imagini.

### MEDICAL ONTOLOGIES FOR REPRESENTATION OF TEXTUAL AND GRAPHICAL INFORMATION

The present paper deals with the creation of ontologies that describe both text and graphic information. The overview of some medical ontologies and image ontologies is presented. This overview illustrates the diversity of domains and tools to deal with them.

**Keywords:** medical ontologies, knowledge, structured text, images.

#### 1. Введение

Последние годы онтологии (как метод представления информации) являются объектом пристального внимания в области искусственного интеллекта. Их создают, используют и изучают. Их используют при разработке программных средств (в том числе систем, основанных на знаниях) как часть результата деятельности аналитика. Они обеспечивают повторную используемость информации и обеспечивают уточнение смысла терминов, с помощью которых передается некоторая информация. На их основе (с их помощью) могут формироваться базы знаний для различных интеллектуальных систем, в частности – экспертных [1].

Онтологии можно применять для представления знаний о терминах предметной области и отношениях между ними.

Сам термин заимствован из философии и обозначает науку, описывающую формы бытия и то, как они соотносятся между собой [2].

На сегодняшний день в информатике нет четкого определения понятия «онтология». Существует множество определений этого термина. Например:

- 1) Онтология – артефакт, структура, описывающая значения элементов некоторой системы [3].
- 2) Онтология – это спецификация концептуализации [4].
- 3) Онтология – это попытка всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из иерархической структуры данных, содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области [5].

Можно выделить следующие преимущества онтологий перед другими способами представления знаний:

1. Четкое и наглядное формулирование знаний из выбранной области;
2. Возможность повторного использования;
3. Интероперабельность (способность продукта или системы, интерфейсы которых полностью открыты, взаимодействовать и функционировать с другими продуктами или системами без каких-либо ограничений доступа и реализации).

Онтологии применяются в искусственном интеллекте, семантических сетях и технологии программирования как форма представления знаний о реальном мире или его части. Они стали привилегированным и почти неизбежным средством для представления и использования знаний и данных. Это относится ко многим областям, особенно к области здравоохранения. Медицина является одной из наиболее обширных областей применения онтологий. В разделе 2 рассмотрены общие принципы создания медицинских онтологий, которые проиллюстрированы примерами из раздела 3. Оставаясь в

рамках той же предметной области, мы на конкретных примерах показали подходы к созданию онтологий знаний, представленных изображениями. Интерес к этому типу онтологий обоснован широким применением методов медицинской имажистики как в процессе диагностики, так и в лечении больных.

## 2. Онтологии в медицине

Использование онтологий в медицине до недавних пор было главным образом сосредоточено на представлении и (пре)образовании медицинской терминологии [6]. Медики создали свои собственные специализированные языки и словари, что помогает им рационально накапливать знания и иметь доступ к основным медицинским знаниям и к информации, связанной с пациентом. Такая терминология, оптимизированная для обработки человеком, характеризуется значительным количеством скрытых знаний. С другой стороны, медицинские информационные системы однозначно должны уметь обращаться со сложными и детальными медицинскими понятиями (возможно отображёнными на разных языках). Очевидно, что это сложная задача, и она требует глубокого анализа структуры и понятий предметной области. Но она может быть решена с помощью построения онтологий медицинской предметной области для представления систем медицинской терминологии.

Приложения, основанные на онтологиях, также были разработаны и в области обработки естественного языка, используемого в медицине.

Использование онтологий в медицине имеет определённые преимущества. Они могут помочь в разработке более мощных и интероперабельных информационных систем в данной области. Онтологии могут также облегчить органам здравоохранения передачу, повторное использование и распределение данных о пациентах. Также, они могут предоставлять критерии, основанные на семантике, чтобы поддерживать различные статистические группы (множества) для различных целей. Возможно, самым главным плюсом онтологий для системы здравоохранения является их способность поддерживать необходимую интеграцию знаний и данных [6].

## 3. Примеры медицинских онтологий

Рассмотрим несколько примеров медицинских онтологий, призванных проиллюстрировать разнообразие предметных областей и используемых средств для работы с ними.

1) Предметно-ориентированная онтология «Фундаментальная модель анатомии» (ФМА) (FMA Ontology - the Foundational Model of Anatomy Ontology) [7].

ФМА – это свободное, доступное для общего пользования программное обеспечение. Фундаментальная модель анатомии (ФМА) – развивающийся компьютеризированный источник знаний для био-медицинской информатики, он связан с представлением классов или типов и отношений, необходимых для символического представления фенотипической структуры человеческого тела в форме, понятной для людей, и к тому же управляемой и интерпретируемой автоматизированными системами.

ФМА является онтологией предметной области, которая представляет собой связную совокупность формальных декларативных знаний об анатомии человека. Её онтологическая структура может быть применена ко всем другим биологическим видам.

Онтология Фундаментальной модели анатомии является одним из информационных ресурсов, интегрированных в распределённую Информационную систему анатомии, разработанную и поддерживаемую Отделом структурной информатики Вашингтонского университета.

Она содержит около 75 000 классов, более 120 000 терминов, более 2 100 000 экземпляров отношений, образуя, таким образом, одну из наиболее представительных онтологий данной предметной области.

Она снабжена веб-браузером (FME – Foundational Model Explorer), который в удобной для пользователя форме выдаёт информацию, относящуюся к определённому понятию. Например, если выбрать в качестве части тела понятие “желчный пузырь”, то будет выведено дерево подчинённых ему понятий, выдано определение, синонимы, переводы на шести языках (базовым языком является английский), указано, составной частью чего это понятие является (билиарной системы; панкреатобилиарной системы; группы внутренних органов, принадлежащих брюшной полости; группы внутренних органов брюшной полости женщины; группы внутренних органов брюшной полости мужчины; группы внутренних органов брюшной полости). Кроме того, перечислены понятия, являющиеся составной частью данного (дно жёлчного пузыря, тело жёлчного пузыря, шейка желчного пузыря и т. д.). Приводятся

сведения о лимфатическом дренаже, существовании размера, массы, границ, формы. Каждое из надпонятий или субпонятий, в свою очередь, может быть афишировано аналогичным образом.

Другим полезным средством является Emily Lite — графический интерфейс для формирования запросов к FMA. Запросы структурированы в тройку “субъект, соотношение, объект”, где, по крайней мере, одна из этих компонент неизвестна. При этом «субъект» и «объект» являются классами в FMA, а соотношение выбирается из выпадающего списка. Например, нужно узнать: «К какому типу органов относится жёлчный пузырь?». Форма запроса выглядит таким образом: (Gallbladder, is type of, <unknown>) (рис. 1).

Рис.1. Пример запроса к FMA.

В результате выполнения запроса получим следующее: «Organ with organ cavity» (рис.2).



Рис.2. Результат выполнения запроса к FMA.

Отметим, что при разработке FMA использовался редактор онтологий Protégé.

## 2) Онтология рака (Cancer ontology) [8].

Тезаурус NCI (National Cancer Institute Thesaurus – Тезаурус Национального Института рака) представляет собой справочную терминологию и биомедицинскую онтологию, используемые как в системах Национального Института рака США, так и в других системах. В этот тезаурус входит лексика оказания медицинской помощи, прикладных и фундаментальных исследований, публичная информация и распорядительная деятельность. Тезаурус NCI предоставляет определения, синонимы и другую информацию о почти 10000 раковых заболеваний и заболеваниях, связанных с раком, 8000 монотерапии и комбинированной терапии, а также широкий круг других вопросов, связанных с раком и биомедицинскими исследованиями. Он поддерживается многопрофильной группой редакторов, которые добавляют около 900 новых записей каждый месяц.

Тезаурус в настоящее время содержит более 34000 понятий, структурированных в 20 таксономических деревьев.

В рамках NCI Тезаурус используется для обеспечения терминологической поддержки обществу веб-порталу Института (<http://cancer.gov>), многочисленным порталам, поддерживающим консорциум и другие коллективы исследователей, и используется в CaCORE (CaCORE – среда для стандартного онтологического представления рака, которая обладает механизмом создания интероперабельных

биомедицинских информационных систем), в качестве семантической основы для метаданных и объектов, которые формируют инфраструктуру, на основе которой строятся порталы Центра биоинформатики Национального Института рака (<http://ncicb.nci.nih.gov>). Тезаурус опубликован под открытой лицензией на контент во многих форматах, включая OWL.

В процессе разработки используется Среда для разработки терминологии компании Apelon.

Тезаурус NCI, как и описанная ранее онтология «Фундаментальная модель анатомии», разработан с помощью редактора онтологий Protégé.

### 3) Генная онтология (Gene ontology) [9]

Проект «Генная онтология» является результатом совместных усилий по решению необходимости последовательного описания генных продуктов в различных базах данных. Он включает в себя множество баз данных (FlyBase (*Drosophila*), the *Saccharomyces* Genome Database (SGD) and the Mouse Genome Database (MGD) и т. д.) и несколько крупных хранилищ данных геномов растений, животных и микроорганизмов.

Данный проект – основной в области биоинформатики, целью которого является стандартизация представления атрибутов генов и продуктов генов по видам и базам данным. Проект предоставляет как терминологию для описания характеристик продуктов гена и аннотированные данные продуктов гена Консорциума генной онтологии, так и инструментарий для обработки этих данных и доступа к ним.

Проект «Генная онтология» является онтологией определяемых терминов, представляющих собой свойства продукта гена. Данная онтология охватывает три области: клеточный компонент (части клетки или её внеклеточная среда), молекулярное функционирование (элементарная деятельность продукта гена на молекулярном уровне, такая, как связывание или катализ), биологический процесс (наборы молекулярных событий с определённым началом и концом, относящиеся к функционированию интегрированных живых единиц: клеток, тканей, органов и организмов). Например, продукт гена *цитохром с* можно описать с помощью термина молекулярной функции – активность оксидоредуктазы, терминами биологического процесса – окислительное фосфорилирование и индукция гибели клеток, и терминами клеточного компонента – матрикс митохондрии и митохондриальная внутренняя мембрана.

«Генная онтология» создана как направленный ациклический граф, и у каждого термина есть определяющие отношения к одному или нескольким другим терминам в одной и той же области, а иногда и в других областях.

Словарь «Генной онтологии» включает в себя термины, применимые к прокариотам и эукариотам, к одноклеточным и многоклеточным организмам.

При разработке Генной онтологии использовался редактор онтологий OBO-Edit, который распространяется бесплатно, с открытым кодом, и создан специально для этого проекта. Эта онтология снабжена также поисковым средством AmiGO. С помощью AmiGO можно вести поиск одного и более продуктов гена и просматривать их аннотации. Отметим, что проект находится в свободном доступе в сети Интернет.

### 4. Онтологии изображений

В предыдущих разделах были рассмотрены подходы к созданию онтологий для случая, когда знания были заданы посредством (структурированного) текста. В качестве предметной области была выбрана медицина, поскольку именно здесь онтологии нашли наиболее широкое применение. Ситуация обретает отличительные черты, когда знания представлены изображениями. Краткую характеристику нескольких подходов при создании онтологий для описания изображений дадим на нескольких примерах, тоже из области медицины.

Разработка прототипа базы данных, доступной через сеть, для управления изображениями, полученными в ходе экспериментов в биомедицинской исследовательской лаборатории, при изучении факторов, контролирующих развитие катаракты, представлена в [10].

На основе создаваемой онтологии, которая была разработана для описания экспериментальных данных и протоколов, используемых в лаборатории, хранилище изображений позволяет членам лаборатории группировать изображения по нескольким атрибутам. Использование онтологий для разработки этого и других инструментов способствует взаимосвязи между инструментами и в конечном итоге – обмену данными с другими исследователями.

Онтология «Эксперимент», которая реализуется в Protégé, формализует типы данных биомедицинского эксперимента, связанного с катарактой. Среди этих типов данных находится представление фотографии цифровой щелевой лампы, сделанные в ходе нескольких экспериментов. Например, образы, созданные во время экспериментов по исследованию роли SPARC белка в катаракте показывают степень развития катаракты у экспериментальных животных.

В [11] описана база данных BioImage Database. BioImage Database — управляемая онтологией база данных изображений для многомерных изображений биологических образцов. Онтология, являющаяся центральной частью этой базы данных, написана на OWL-DL и разработана с использованием Protégé OWL plug-in. Все взаимодействия пользователя с базой данных осуществляются посредством этой онтологии. Ввод метаданных вручную упрощён динамическим созданием из лежащей в основе онтологии простой веб-формы подчинённых пользовательских интерфейсов. Усовершенствованный поисковый интерфейс обеспечивает достоверный поиск соответствующих изображений.

### 5. Онтология органов абдоминальной области

Рассмотрим также подход к созданию онтологии ультразвуковых изображений, полученных в процессе диагностики органов гепато-панкреато-билиарной системы [12].

Для построения онтологии ультразвуковых изображений используется база знаний системы поддержки принятия решений в области ультразвуковой диагностики органов брюшной полости SonaRes [13]. База знаний создана в ходе многолетнего сотрудничества медиков-экспертов с информатиками и представляет собой формальное представление опыта и знаний медиков в данной области [14].

При построении онтологии ультразвуковых изображений, следуя проекту системы SonaRes, в первую очередь рассматривается абдоминальная зона, для которой характерно не только наличие большого количества органов, но также и большая степень их взаимного влияния. Изучение влияния органов друг на друга с помощью построенной нами онтологии является одной из задач, которую стоит рассматривать в перспективе. Для её реализации необходимо создать онтологию ультразвуковых изображений хотя бы для нескольких органов абдоминальной области.

Решение задачи построения онтологии ультразвуковых изображений (УИ) основано на том, что все УИ органов пациентов имеют вербальные (словесные) описания. Эти описания осуществляются с помощью медицинских терминов – понятий и характеристик, между которыми существуют связи и зависимости. Таким образом, задача построения онтологии УИ превращается в задачу построения онтологии медицинских описаний УИ (ОУИ), которые связаны с соответствующими УИ (рис. 3).

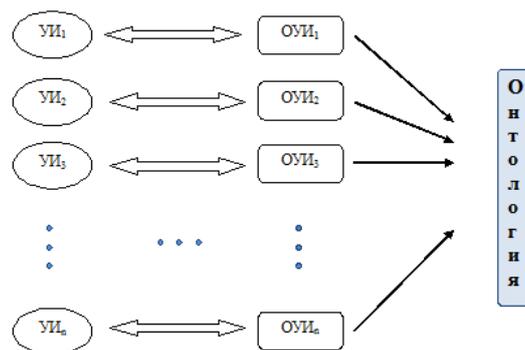


Рис.3. Соответствие между ультразвуковыми изображениями и их описаниями в построении онтологии УИ.

Следуя этому подходу, мы разработали онтологию патологий желчного пузыря, что дало возможность установить базовую структуру, а также принципы выделения классов, подклассов, слотов и т. п. Эта технология в дальнейшем будет применена к разработке онтологии других органов.

### 6. Выводы

Проведенный анализ медицинских онтологий показал, что такая форма представления знаний удобна для описания как текстовой, так и графической информации. При этом для онтологий, рассчитанных на широкий круг пользователей, необходим также удобный пользовательский интерфейс,

предназначенный как для простого обзора имеющихся данных, так и для задания более сложных запросов, как в случае Фундаментальной модели анатомии. Отметим также широкое распространение редактора Protégé, который практически доминирует в указанной предметной области.

Анализ форм и средств создания онтологий позволил разработать и применить принципы построения онтологии ультразвуковых изображений.

#### Библиография:

1. ШАЛФЕЕВА, Е.А. Классификация структурных свойств онтологий. В: *Искусственный интеллект*, 2005, №3, с.67-77.
2. MICHAEL, K. Smith, Chris Welty and Deborah L. McGuinness, Eds. OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/owl-guide/>.
3. ДОБРОВ, Б.В. и др. Онтологии и тезаурусы: модели, инструменты, приложения. Серия: *Основы информационных технологий*. Москва: Бинوم. Лаборатория знаний, 2009. 176 с. <http://www.intuit.ru/department/expert/ontoth/1/>.
4. GRUBER, T.R. The role of common ontology in achieving sharable, reusable knowledge bases. In: J. A. Allen, R. Fikes, and E. Sandewell, editors, *Principles of Knowledge Representation and Reasoning – Proceedings of the Second International Conference*, pp. 601-602. Morgan Kaufmann (1991).
5. *Ontology*. [http://en.wikipedia.org/wiki/Ontology\\_%28information\\_science%29](http://en.wikipedia.org/wiki/Ontology_%28information_science%29).
6. DOMENICO M. PISANELLI. *Ontologies*. <http://www.openclinical.org/ontologies.html>.
7. *FMA Ontology*. <http://sig.biostr.washington.edu/projects/fm/Members.html>.
8. *Cancer ontology*. <http://www.mindswap.org/2003/CancerOntology/>.
9. *Gene ontology*. <http://www.geneontology.org/>.
10. Fong, Christine and Rosse, Cornelius and Clark, J. I. and Shapiro, Linda G and Brinkley, James F (2004) An Ontology-based Image Repository for a Biomedical Research Lab. In: *Proceedings, MedInfo 2004*, pages 1598, San Francisco, CA., <http://sigpubs.biostr.washington.edu/archive/00000144/>.
11. CATTON, C., SPARKS, S. and SHOTTON, D.M. (2005) *The Image Store ontology and the BioImage Database: semantic web tools for biological research images*. EWIMT 2005, 30 Nov – 1 Dec 2005, London.
12. ГАЙНДРИК, К.В., КОЖОКАРУ, С.К., БРУК, Н.Л., МАГАРИУ, Г.А., ВЕРЛАН, Т.Б. Построение онтологии ультразвуковых изображений в области ультразвуковой диагностики: Материалы международной научно-технической Интернет-конференции “*Информационные системы и технологии*”, т. 1, с. 16-22. Орёл, 2011.
13. COJOCARU, S., GAINDRIC, C. Clinical Decision Support System SONARES. In: *International Journal Information Theories and applications*, 2008, vol. 15, no.2, p.137-143.
14. BURTSEVA, L., COJOCARU, S., GAINDRIC, C., JANTUAN, E., POPCOVA, O., SECRIERU, I., SOLOGUB, D. *SONARES - A Decision Support System in Ultrasound Investigations*. In: CSJM, 2007, no.2 (44), vol. 15, p.153-177.

Prezentat la 14.03.2013