

Системно-комплексный подход при накоплении знаний в предметной области биомедицины

В.В. Баранюк, В.В. Нечаев, О.С. Крылова

Аннотация – С точки зрения системно-комплексного подхода в статье описаны вопросы представления, накопления, объединения и дальнейшего использования знаний в области биомедицины. При этом затронута актуальность проработки и использования подходов, обеспечивающих пертинентность информационных ресурсов. В данном случае под пертинентностью понимается соответствие найденных информационных ресурсов информационным потребностям пользователя, иногда независимо от того, как точно эта потребность выражена в тексте запроса. В статье определены и проанализированы основные задачи, направленные на создание онтологической модели соответствующей предметной области, выполняющей роль концептуальной семантической модели при автоматизированном пополнении хранилища данных информационными ресурсами, пертинентными относительно запросов, связанных с узлами онтологической модели. Рассмотрен мировой опыт накопления знаний в области биологии за счет разработки нескольких сотен онтологий, которые используются для описания и интеграции знаний, а также вывода новых знаний по различным направлениям, включая анатомию, биохимию, биологические процессы, функции и последовательности, заболевания, окружающую среду, экспериментальные доказательства, фенотип, белки, таксономии и др. Представлены предложения по обеспечению ссылочной связанности объектов онтологической модели, а также по разработке механизмов использования онтологии в задачах биомедицины.

Ключевые слова – онтология, предметная область, онтологическая модель, представление знаний, пертинентные информационные ресурсы, потребности пользователя, качество поиска, биомедицина.

ВВЕДЕНИЕ

В различных научных источниках в настоящее время сконцентрировано большое количество публикаций в области биомедицины, накоплены огромные объёмы ценного материала, работа с которым затруднена вследствие его разрозненности, разноформатности, а

также отсутствия чёткой систематизации, обусловленной спецификой предметной области, сформированной на стыке различных наук и направлений. Такая информация включает накопленные сведения и исследования из различных предметных областей: общая медицина, ветеринария, стоматология, гистология, эмбриология, также фундаментальных наук, таких как химия, биология, биохимия, генетика, анатомия, физиология, микробиология, биомедицинский инжиниринг (рисунок 1). Именно поэтому в рамках развития биомедицины как раздела медицины, с теоретических позиций исследуется организм человека, его строение и функции, патологические состояния, методы их диагностики, коррекции и лечения [1].

При этом возникает задача обеспечения интероперабельности знаний, которая предполагает, что знания, сформированные при решении одной задачи, были бы пригодны при решении других задач в рамках исследуемой предметной области. Кроме этого, при проведении исследований в области биомедицины при работе с информационными ресурсами остается актуальной задача проработки и использования подходов, отвечающих пертинентным информационным потребностям специалистов (рисунок 2). В данном случае под пертинентностью понимается соответствие найденных информационных ресурсов информационным потребностям пользователя, иногда независимо от того, как точно эта потребность выражена в тексте запроса [2].

Именно поэтому в рамках формирования системно-комплексного подхода к представлению, накоплению, объединению и дальнейшему использованию знаний в области биомедицины необходимо решение следующих задач:

- создание модели представления знаний;
- автоматизированное пополнение хранилища информации информационными ресурсами, пертинентными относительно запросов, связанных с узлами модели;
- обеспечение ссылочной связанности объектов модели;
- разработка механизмов использования базы знаний в задачах биомедицины.

Статья получена 30.10.2018 г.

К.т.н., с.н.с., В.В. Баранюк, МИРЭА (e-mail: valentina_bar@mail.ru).

К.т.н., В.В. Нечаев, МИРЭА (e-mail: nechayev@mirea.ru).

О.С. Крылова, МИРЭА (e-mail: mail.olga.krylova@yandex.ru).



Рис. 1. Предметная область



Рис. 2. Релевантные информационные ресурсы

Для решения указанных задач, в том числе представления знаний, интеграции информационных ресурсов в области биомедицины и семантического поиска, в качестве модели представления знаний целесообразно использовать онтологии. Онтология – соглашение об общем использовании понятий, которое содержит средства представления предметных знаний. Она может рассматриваться как определенное описание взгляда на мир в конкретной сфере интересов, который состоит из набора терминов и правил использования этих терминов, ограничивающих их значение в рамках конкретной предметной области [3]. Применительно к сфере информационных технологий онтология рассматривается как подробная формализация области знаний с помощью некоторой концептуальной схемы.

При использовании онтологии исследователь может получать информационные ресурсы в области биомедицины, семантически релевантные его запросам. Следует отметить, что кроме этого использование онтологий позволяет соответствующим программным средствам (интеллектуальным агентам) автоматически (без участия человека) определять смысл терминов, использованных при описании ресурсов, и сопоставлять его со смыслом поставленной задачи [4]. Онтологии имеют важное прикладное значение: они используются при решении задач искусственного интеллекта, задач программной инженерии и др.

Наиболее интересные идеи и решения часто возникают на стыке наук, поэтому исследователю в области биомедицины важно использовать знания, накопленные по различным научным направлениям. В настоящее время в области биологии разработано несколько сотен онтологий, которые можно использовать для описания и интеграции знаний, а также вывода новых знаний. Наиболее масштабный пример – проект Open Biological Ontologies (OBO) – Открытые биомедицинские онтологии. В настоящее время в OBO описано более 70 онтологий по различным направлениям, включая анатомию, биохимию, биологические процессы, функции и последовательности, заболевания, окружающую среду, экспериментальные доказательства, фенотип, белки, таксономии и др. [5]:

- CHEBI: Chemical Entities of Biological Interest: химические объекты биологического интереса;
- GO: Gene Ontology GO: геновая онтология;
- PATO: Phenotypic Quality Ontology PATO: фенотипическая онтология качества;
- PRO: Protein Ontology PRO: онтология протеина;
- XAO: Xenopus Anatomy Ontology XAO: онтология анатомии Xenopus;
- ZFA: Zebrafish Anatomy Ontology ZFA: онтология анатомии;
- и другие.

Следует отметить, что основные из разработанных онтологий составлены на английском языке, что не даёт возможности использовать их в исходном виде, а несогласованность перевода может усугубить некорректность описания предметной области,

вероятность которой всегда и так высока из-за различий в интерпретации терминов и определений.

Тем не менее прямым путём решения задачи интеграции знаний из большого количества предметных (тематических областей) является объединение различных онтологий или рубрикаторов конкретных областей под единым узлом «биомедицина». Большими недостатками такой обобщенной онтологии являются разные подходы к формированию уже существующих отдельных онтологий и использование разной терминологии, зачастую не позволяющей корректно связать узлы онтологий.

Формирование масштабной онтологической модели для предметной области «биомедицина» является сложной и нетривиальной задачей. Однако, осуществляя её создание по схеме «сверху-вниз» и охватывая все большее количество семантических узлов из различных областей знаний, возникает потребность использования системно-комплексного подхода к представлению знаний в области биомедицины, формированию тематических областей, соответствующих знаниям в области биологии, химии, генетики, физиологии и др. При этом правила формирования отдельных тематических областей определяются в соответствии с целями использования онтологий «Биомедицина».

Следует отметить, что к традиционным наукам и направлениям периодически добавляются новые. Например, биомеханика и бионика, которые изучают свойства живых организмов и органов, особенности строения и жизнедеятельности организмов с целью создания новых механизмов, приборов, бионических систем и др. Это приводит к совершенствованию и расширению онтологий, появлению новых веток, узлов и связей.

В вопросе автоматизированного пополнения хранилища информации информационными ресурсами, пертинентными относительно запросов, связанных с узлами онтологической модели, одним из наиболее важных аспектов является классификация информационных ресурсов (категоризация, рубрикация) – определение ресурса в одну или несколько тематик (категорий, рубрик) на основании его содержания. Для автоматизированного пополнения хранилища информационных ресурсов необходимо использование механизмов автоматической классификации [6]. При пополнении хранилища информационными ресурсами из информационных систем, содержащих медицинские данные, следует обратить внимание и на правовые вопросы [7].

Обеспечение ссылочной связанности объектов и узлов онтологической модели является ресурсозатратным процессом и требует от специалиста знаний предметной области разрабатываемой онтологии. Именно поэтому для оптимизации работы специалиста-предметника целесообразно использование механизмов, обеспечивающих нахождение связей между объектами онтологий, основанных на поиске и анализе имеющихся в определениях терминов-концептов текстовых ссылок на другие термины-концепты. Термины в тексте могут

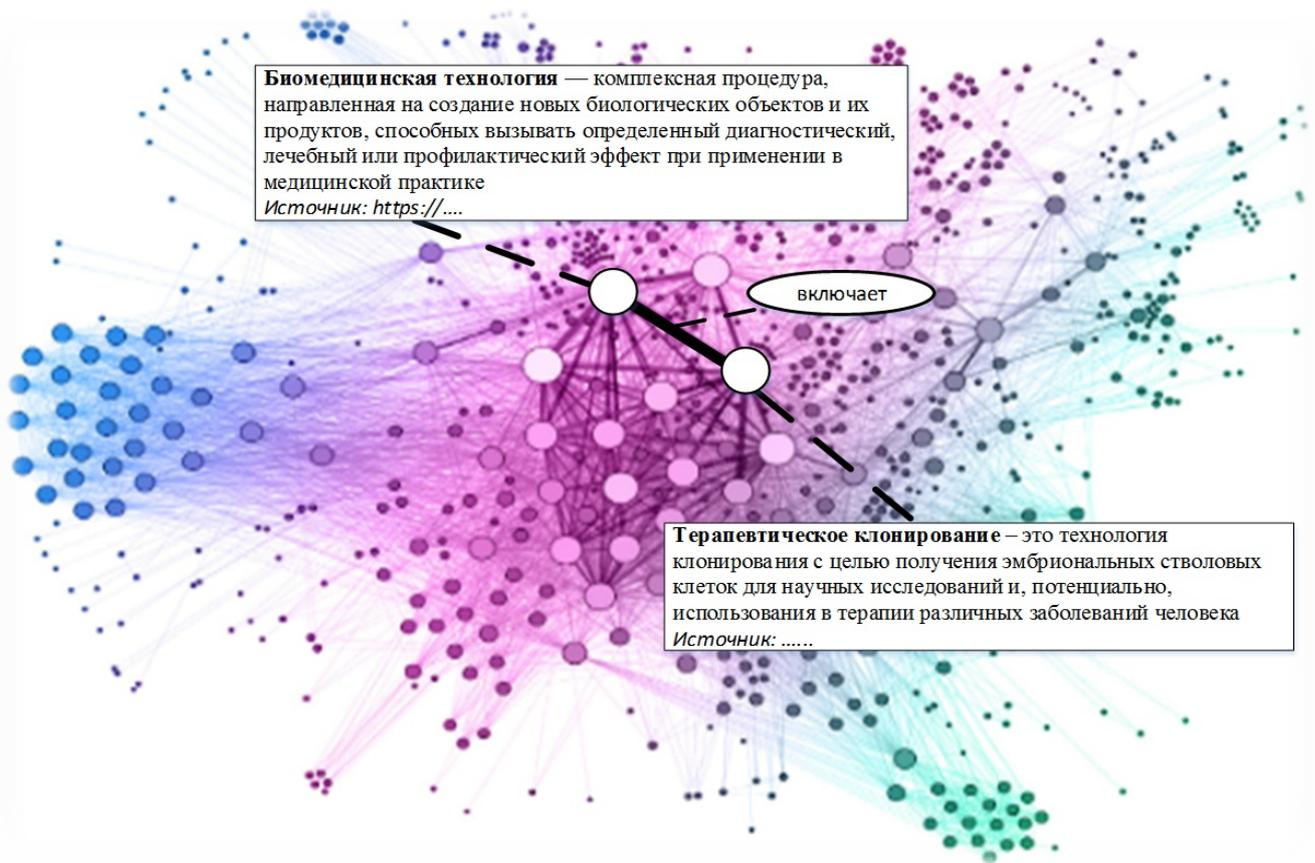


Рис. 3. Пример визуализации онтографа

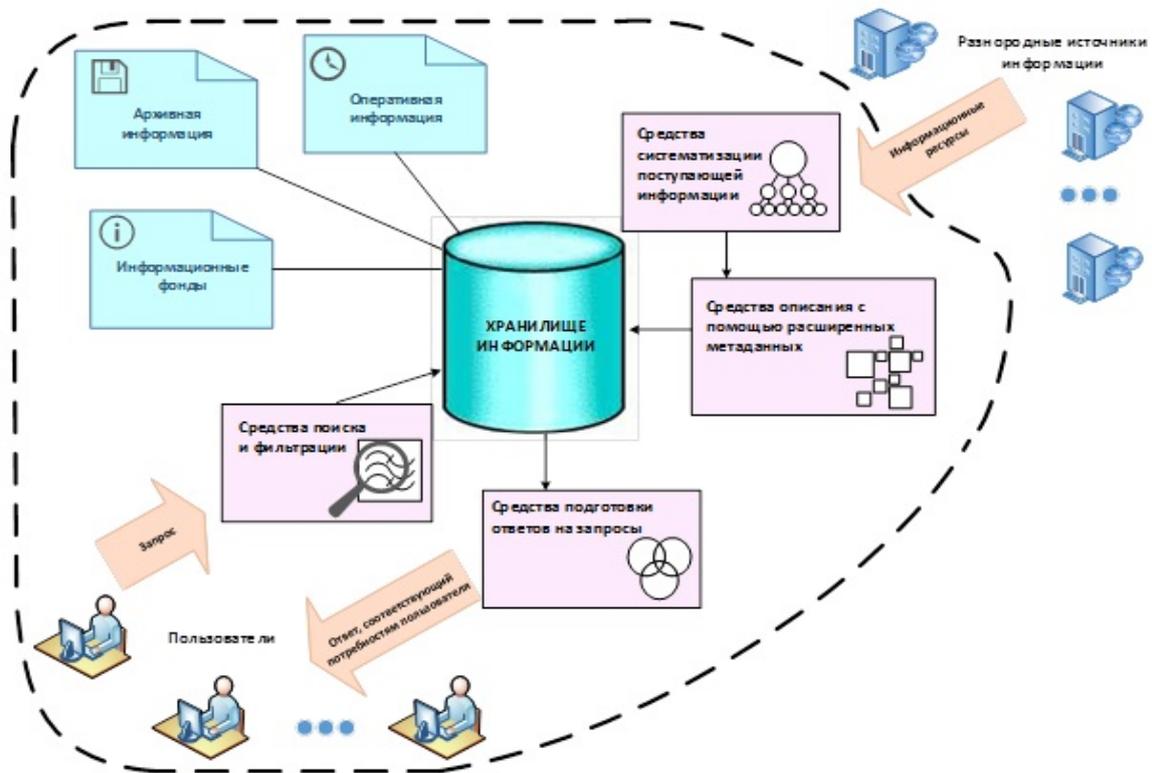


Рис. 4. Концептуальная схема получения пертинентной информации

быть употреблены в различном числе или падеже в зависимости от контекста. Очевидно, что в таком случае необходимо осуществлять поиск не по точному имени термина, а с учётом различных языковых форм. В рамках разработки подобных механизмов основными задачами являются выделение неизменяемой части термина (основа слова) и поиск вхождений термина в текст статьи с учётом различных форм его изменяемой части (окончания).

В качестве основного механизма использования базы знаний в задачах биомедицины предлагается визуализация онтологической модели, отображающая совокупность семантических схем, в графической форме представляющих парадигматические отношения между дескрипторами. Следует отметить, что вследствие обширности предметной области отображение онтографа также будет очень большим, поэтому должна быть предусмотрена возможность просмотра отдельных областей, относящихся к выбранным пользователем понятиям.

Графическое представление онтологической модели биомедицинской области (рисунок 3) позволяет наглядно отобразить:

- место термина в системе понятий рассматриваемой предметной области;
- ссылки и связи данного термина;
- привязку термина к источникам (информационным ресурсам) и др.

Исследователь с помощью онтологической модели области биомедицины сможет:

- при просмотре тезауруса «путешествовать» по предметной области, находя интересующие его термины;
- осуществлять поиск по интересующим терминам, а также по близким понятиям или терминам;
- осуществлять навигацию по онтографу с определением информационных ресурсов, соответствующих интересующему понятию [8].

Концептуальная схема получения пертинентной информации с применением технологий систематизации представлена на рисунке 4.

ЗАКЛЮЧЕНИЕ

Системно-комплексный подход к представлению, накоплению и объединению знаний в области биомедицины посредством формирования онтологии обеспечивает нахождение специалистами пертинентных информационных ресурсов и использование знаний из разных предметных областей для проведения исследований по изобретению новых лекарственных препаратов, более глубокому пониманию механизмов, лежащих в основе болезни, и многое другое, что в целом способствует усовершенствованию подходов к профилактике, диагностике и лечению заболеваний.

БИБЛИОГРАФИЯ

- [1] Биотехнологии. URL: <http://mfina.ru/biotexnologii> (дата обращения: 03.08.2018).
- [2] Sigov A.S., Nechaev V.V., Baranyuk V.V., Koshkarev M.I., Smirnova O.S., Melikhov A.A., Bogoradnikova A.V. Architecture of domain-specific data warehouse for bionic information resources. Ecology, environment and conservation. Vol. 21 Nov. 2015 Suppl. Issue; Page No. 181 – 186.
- [3] Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – Спб.: Питер, 2001.
- [4] Построение онтологии предметной области. Воронин А.А. Запорожский институт государственного и муниципального управления. URL: <http://works.tarefer.ru/69/100462/index.html> (дата обращения: 23.08.2018).
- [5] Н.Л. Подколотный, О.А. Подколотная. Онтологии в биоинформатике и системной биологии. Вавиловский журнал генетики и селекции. 2015, 19(6):652-660.
- [6] Sigov A., Baranyuk V., Nechaev V., Smirnova O., Melikhov A., Approach for forming the bionic ontology. XII международный симпозиум «Интеллектуальные системы» (INTELS'2016). М.: РУДН, 07.10.2016, 5 с.
- [7] Мартынова Т.Л. Правовое обеспечение жизненного цикла государственных систем // В сборнике развитие российского права: новые контексты и поиски решения проблем. III Московский юридический форум. X Международная научно-практическая конференция: в 4 частях. 2016. С.399 – 405.
- [8] Sigov A., Baranyuk V., Nechaev V., Melikhov A., Smirnova O. Improving the quality of bionic resource retrieval by visualizing a specific bionic-oriented thesaurus. XII международный симпозиум «Интеллектуальные системы» (INTELS'2016). М.: РУДН, 07.10.2016, 6 с.

System-integrated approach in the accumulation of knowledge in the subject field of biomedicine

V.V. Baranjuk, V.V. Nechaev, O.S. Krylova

Abstract – The article describes, in terms of the system-integrated approach, the presentation, accumulation, combination and further use of knowledge in the biomedical sphere. The relevance of the elaboration and use of approaches ensuring the pertinence of information resources is affected. Pertinence means the correspondence of the information resources found to the information needs of the user, sometimes regardless of how exactly this need is expressed in the request text. In the review, the main tasks were identified and analyzed, aimed at creating an ontological model of the relevant subject area, playing the role of a conceptual semantic model with automated replenishment of data storage with information resources pertinent to queries related to nodes of the ontological model. The world experience in the accumulation of knowledge in the field of biology is considered through the development of several hundred ontologies that are used to describe and integrate knowledge, as well as to derive new knowledge in various areas, including anatomy, biochemistry, biological processes, functions and sequences, the environment, experimental evidence, phenotype, etc. Proposals are proposed to provide a reference link between objects of the ontological model, as well as the development of mechanisms for the use of ontology in biomedical tasks.

Keywords – ontology, subject area, ontological model, knowledge representation, pertinent information resources, user demands, search quality, biomedicine.