

Е.М. Бениаминов

## Библиотеки онтологий в Веб. Состояние и перспективы

*Дается обзор использования компьютерных онтологий для связывания данных в Интернете и представления знаний. Рассматривается тенденция организации библиотек и серверов онтологий для совместной коллективной разработки онтологий и их использования. Библиотеки онтологий рассматриваются как открытый ресурс в Веб. Особое внимание уделяется проблемам формирования библиотек онтологий, их отладки и направлениям развития в этой области.*

**Ключевые слова:** онтологии, библиотека онтологий, представление знаний, связанные данные, открытый язык

### ВВЕДЕНИЕ

Термин «онтология» в информационных технологиях стал очень популярным. Почти на всех последних конференциях в области ИТ был раздел, посвященный онтологиям. Компьютерные технологии, в которых используются онтологии, становятся ключевыми в развитии современного Веба, аналитики и управления большими данными, машинного обучения, в приложениях в области медицины, науки, образовании, торговле, производстве, в музейном деле и т.д.

Последний наш обзор [1] по технологиям с использованием компьютерных онтологий был написан 9 лет назад, в котором давались следующие определения.

- Онтологии представляют собой спецификации на формальном языке, в которых фиксируются договоренности группы специалистов о том, что как называется в их области, какими свойствами обладают и каким соотношениям удовлетворяют.

- На логическом уровне каждой онтологии соответствует некоторая теория (сигнатура+аксиомы), а иногда и некоторая фиксированная модель (множества+операции+отношения). Вопросы к онтологии интерпретируются как запросы к соответствующей ей теории (модели).

- Онтологии, как правило, строятся по модульному принципу: при определении новой онтологии могут использоваться уже построенные ранее.

- Онтологии должны быть удобны для понимания специалистами и интерпретироваться компьютерными системами при использовании.

В настоящей статье мы продолжаем придерживаться этих определений. Обзор [1] заканчивался призывом к созданию открытых библиотек онтологий в Веб с открытыми языками формирования онтологий и запросов к ним.

За последние 10 лет появилось много новых библиотек и серверов онтологий. Некоторые из них спе-

циализированы и разрабатываются целенаправленно в ведомствах, фирмах и объединениях. Некоторые онтологии становятся ведомственными стандартами, например в медицине [2, 3], в музейном деле [4], в госучреждениях [5].

Уже в 2012 г. появился хороший обзор [6], название которого можно перевести как «Где публиковать и находить онтологии? Обзор онтологических библиотек». В нем приводится сравнение нескольких библиотек онтологий и показывается, что многие из них разработаны с разными целями использования, а также приводятся некоторые общие принятые требования к библиотекам онтологий в Веб.

### ОНТОЛОГИИ ДЛЯ ОРГАНИЗАЦИИ СВЯЗАННЫХ ДАННЫХ В ВЕБ

Особое место занимают библиотеки онтологий, используемые в проекте открытых связанных данных, который был объявлен одним из создателей Веба Тимом Бернерсом-Ли в 2001 г. В эмоциональной видео-лекции «От гипертекстовой организации страниц и серверов открытых энциклопедий к открытым и связанным данным в Веб» [7], в которой говорится, о развитии Веба в направлении открытых связанных данных, Тим Бернерс-Ли призывает участников Веба открывать свои данные (базы данных) для общего пользования, обеспечивая доступ к данным через онтологии, описывающие схемы их базы данных. При этом он на примерах демонстрирует общую выгоду от использования открытых данных через связывание данных в общую систему. Естественно, что онтологиям и библиотекам онтологий в этой системе отводится особая роль по обеспечению согласования данных из разных источников и по добыче знаний из больших объемов данных [8].

DBpedia [9] – это проект, направленный на извлечение структурированной информации из данных, созданных в рамках проекта Википедия, и

публикации её в виде доступном под свободной лицензией наборов данных. Проект был отмечен Тимом Бернерсом-Ли, как один из наиболее известных примеров реализации концепции связанных данных. Как отмечается в [9], по состоянию на ноябрь 2016 г. базы данных DBpedia описывали уже более 6,6 млн сущностей, из которых 4,9 млн имеют аннотации, 1,9 млн имеют географические координаты. В целом, 5,5 млн источников Интернета расклассифицированы в соответствии с онтологией DBpedia, состоящей, в том числе, из 1,5 млн персоналий, 440 тыс. географических объектов, 139 тыс. музыкальных альбомов, 111 тыс. фильмов, 21 тыс. видеоигр, 286 тыс. организаций.

Как отмечено разработчиками DBpedia в [10], онтология DBpedia была создана вручную на основе наиболее часто используемых информационных подразделов в Википедии. Онтология в настоящее время охватывает 685 классов, которые образуют иерархию и описываются 2795 различными свойствами.

С выпуском DBpedia 3.5 была открыта библиотека Wiki [11], с помощью которой пользователи могут самостоятельно редактировать онтологии DBpedia и сопоставлять ее элементы с элементами Википедии приблизительно так же, как производится редактирование страниц в Википедии.

Выпуск DBpedia 2016-10 состоит из 13 млрд (2016-04: 11,5 млрд) информации в виде троек RDF, из которых 1,7 млрд (2016-04: 1,6 млрд) были извлечены из английского издания Wikipedia, 6,6 млрд (2016-04: 6 млрд) были извлечены из других изданий на разных языках изданий и 4,8 млрд (2016-04: 4 млрд) из Wikipedia Commons и Wikidata.

## ТИПЫ И ЗАДАЧИ БИБЛИОТЕК ОНТОЛОГИЙ

В настоящее время библиотеки онтологий созданы для разных целей и, таким образом, обеспечивают разные функциональные возможности, имеют очень разный объем и могут быть специализированы по предметным областям. Поэтому пользователь, который хочет повторно использовать онтологию из множества сотен или тысяч онтологий, доступных в Интернете, должен не только использовать библиотеку онтологий, но и иметь средства поиска и анализа библиотек, чтобы разобраться в этом разнообразии и выбрать, какую онтологическую библиотеку использовать.

Как отмечено в [6], существуют онтологии, которые используются для описания данных в социальных сетях; есть онтологии, которые становятся стандартом для описания продуктов и услуг коммерческих организаций. Некоторые сообщества специалистов в конкретных областях деятельности достигли соглашения о применяемых у них онтологиях, обеспечив высокий уровень повторного использования этих конкретных онтологий. Например, многие биомедицинские исследователи используют генную онтологию для аннотирования своих данных. Точно так же сложился стандарт онтологии в области музейного дела. В таких областях создаются отдельные серверы и программные средства для коллективного

формирования и поддержки стандарта онтологий в своих областях.

Есть более закрытые для редактирования библиотеки онтологий, например, библиотека онтологий SWEET [12] для наук об окружающей среде, разработанная в одной из лабораторий NASA. Онтологии SWEET написаны на языке OWL и являются общедоступными. SWEET-2.3 имеет модульную организацию и содержит около 6000 концепций в 200 отдельных онтологиях. Большинство пользователей использует эти онтологии как онтологии среднего уровня, и пополняет их онтологиями прикладной области для удовлетворения потребностей конечных пользователей.

Особый интерес представляет программная система OntoWiki [13], которая свободно распространяется и служит средством создания библиотек онтологий и связанных данных в стиле семантической Wiki. Это веб-приложение, написанное на PHP и использующее базу данных MySQL для коллективного использования. Система служит редактором онтологий на основе форм взаимодействия с пользователями. Имеется форма для формирования запросов на языке SPARQL к построенным и внешним онтологиям и модуль построения ответов на запросы.

Рассмотренные примеры библиотек онтологий показывают следующие проблемы, которые приходится решать при создании библиотек:

- большие онтологии и большие библиотеки онтологий;
- формирование сложных систем онтологий требует соответствующих средств опробования и отладки онтологий;
- для сложных онтологий полностью отделить не процедурные и процедурные знания не удастся (эффективность использования онтологий, прагматика);
- поддержка модульности построения онтологий и использования библиотек онтологий при создании новых онтологий;
- учет контекстности онтологий в библиотеке и взаимной противоречивости онтологий в различных контекстах;
- проблема интеграции онтологий, представленных на разных языках в разных логиках и моделях.

Для изучения последней проблемы некоторая инициативная группа создала из разнородных онтологий и отображений между ними пример библиотеки Open Ontology Repository [14]. Для согласования таких онтологий группа предлагает использовать современные алгебраические подходы к онтологиям.

## БИБЛИОТЕКИ ОНТОЛОГИЙ С ИСПОЛЬЗОВАНИЕМ ФРАГМЕНТОВ ЕСТЕСТВЕННЫХ ЯЗЫКОВ ФОРМИРОВАНИЯ ОНТОЛОГИЙ И ЗАПРОСОВ К НИМ

Одна из важных проблем взаимодействия с библиотеками онтологий – это представление онтологий пользователям в удобном и понятном виде. Другая важная проблема – это программная поддержка процесса формирования новых онтологий специалистами в различных областях знаний, обеспечивающая использо-

вание уже существующих в библиотеке онтологий и процесс отладки проектируемой онтологии.

Решение этих проблем имеет несколько направлений. Первое – это использование графических редакторов для представления и формирование онтологий. На этом направлении все ещё лидирует по числу использований система Protégé [15], развивающаяся в соответствии с пожеланиями объединения пользователей Protégé. В настоящее время Protégé поставляется и в виде веб-сервера. При этом есть сервер, на котором хранятся онтологии, разработанные пользователями Protégé в виде библиотеки OWL-файлов [16]. Отладка сформированных онтологий производится с помощью программ логического вывода (резонеров). Они выделяют классы, которые не могут иметь экземпляры (пустые классы). С помощью резонеров строятся также ответы на запросы к онтологиям. Запросы могут быть написаны, например, на языке SPARQL.

Второе направление – это использование в библиотеке онтологий-шаблонов для формирования новых онтологий. К таким системам относится библиотека онтологий ODP [17]. Онтологии-шаблоны представляют собой фрагменты онтологий, хранящиеся в библиотеке и используемые в других онтологиях в виде модулей. В этом подходе предполагается построение больших онтологий из большого числа отлаженных онтологий-шаблонов и их согласований в рамках новой. Отладка онтологий производится по результатам тестирования по SPARQL запросам к проектируемой онтологии.

Третье направление – это использование ограниченного формализованного (контролируемого) фрагмента естественного английского языка для представления онтологий. К таким системам относится, например [18], в состав которой входит редактор онтологий Fluent Editor [19], позволяющий описывать онтологии на ограниченном фрагменте естественного английского языка. При этом онтология отображается графически и на языке OWL с использованием языка Semantic Web Rule Language (SWRL). В системе также есть Reasoner, с помощью которого проверяется правильность построения онтологии и строятся ответы на вопросы.

Онтологии, написанные на таком естественном языке, все равно читаются и понимаются с большим трудом. Писать на естественном языке специализированные онтологии также неестественно, как писать математические тексты на естественном языке без математических обозначений.

Следует заметить, что все эти направления не являются абсолютно новыми. Все это уже было (за исключением связанных данных) в системе Ontolingua, созданной и открытой для работы в Веб лабораторией KSL Стэнфордского университета в 1995 г. (система не поддерживается с 2010 г.). Более того, в системе Ontolingua была разработана громадная библиотека онтологий по многим областям знания и получена возможность создания на этой основе онтологий задач с помощью компьютерного моделирования.

## ПЕРСПЕКТИВЫ БИБЛИОТЕК С ОТКРЫТЫМ ЯЗЫКОМ ФОРМИРОВАНИЯ ОНТОЛОГИЙ И ЗАПРОСОВ

Опыт проектирования онтологий показывает, что на пользовательском уровне применение графических методов представления онтологий и использование формализованного фрагмента естественного языка явно недостаточно. Почти в каждой области знаний при работе с онтологиями и их проектировании у пользователей возникает потребность в разработке и использовании специализированных для предметной области формализованных языков. Это ярко проявляется, когда представляются онтологии из математических областей знаний, но и в других областях знаний без использования языка предметной области текст онтологий становится трудночитаемым для специалистов. В предметных областях специализированный язык является более важным, чем естественный. А так как науки развиваются, развиваются языки, то и в системах представления онтологий должны быть средства ввода и использования новых языковых конструкций. Таким образом, хотелось бы, чтобы у пользователей для представления онтологий была возможность воспользоваться средствами открытого языка, и для формирования такого языка была соответствующая компьютерная поддержка.

Итак, будем исходить из следующих положений при развитии систем проектирования библиотек онтологий:

- так как онтология есть фиксация в формальном виде договоренностей группы специалистов определенной области о системе используемых ими понятий, их свойствах и аксиомах, то каждая система онтологий имеет смысл только для группы людей, принимающих эти договоренности (социальный характер онтологий);
- так как в онтологиях фиксируются договоренности специалистов, представлять онтологии должны специалисты предметных областей, поэтому язык представления онтологий должен быть удобен для этих специалистов;
- в каждой области знания при формировании понятий этой области формируются специализированные языки для работы с этими понятиями. Поэтому язык представления онтологий должен быть открытым для пользователей с возможностью его настройки для данной предметной области. При этом внутреннее представление онтологий должно быть стандартизованным для компьютерного использования и межмашинного обмена;
- так как науки и определенные представления в областях знаний меняются, то в компьютерных системах онтологий требуются средства поддержки целостности данных библиотеки онтологий при изменениях в языке и постепенном накоплении онтологий.

В связи с этим, предлагаются три принципа построения библиотек онтологий нового типа.

1. Онтологии строятся в стиле Wikipedia с поддержкой модульности, коллективной работы,

версий и системы согласований (лучшие образцы WebProtégé и OntoWiki).

2. В системе поддерживается среда открытого языка работы с онтологиями, который, по мере пополнения базы онтологий, формируется самими пользователями.

3. Вместе с текстом онтологии в системе формируется внутреннее представление онтологии, которое используется при семантическом анализе выражений языка, при формировании ответов на запросы к онтологии и её отладке. При межмашинном обмене онтологиями и при использовании онтологий в приложениях возможен обмен в некотором стандарте, например в OWL.

В Российском государственном гуманитарном университете на кафедре математики, логики и интеллектуальных систем разработан прототип системы для формирования библиотек онтологий с открытым языком представления онтологий [20]. Ведется сервер проекта [21]. Программы проекта, разработанные на основе программных средств Drupal и Visual Prolog 5.2, доступны на GitHub [22] под открытой лицензией GNU. Имеется руководство пользователя системой ЭЗОП [23].

## ЗАКЛЮЧЕНИЕ

Представленный в статье обзор библиотек онтологий в Веб, позволяет утверждать, что при построении библиотек все ещё актуальными являются следующие задачи:

- использование Web 2.0-технологии для создания социальных сетей и сред в Web, для формирования, наполнения и использования самими пользователями библиотек онтологий;
- открытый язык представления онтологий для пользователя и стандартный – для внутреннего представления;
- предоставление пользователям Веб удобных средств модульного (с использованием чужих модулей) формирования внутреннего (семантического) представления данных своих страниц, онтологий схем своих баз данных и языка запросов к страницам и данным.

Полезно также использование алгебраического подхода к моделированию онтологий, как к средству для интеграции разнородных онтологий.

## СПИСОК ЛИТЕРАТУРЫ

1. Бениаминов Е.М. Некоторые проблемы широкого внедрения онтологий в ИТ и направления их решений // Труды Симпозиума «Онтологическое моделирование». – М.: ИПИ РАН, 2008. – С.71-82. – URL: <http://beniaminov.rsuh.ru/BeniaminovOntoNew.pdf> (дата обращения: 31.12.2018).
2. Открытые биомедицинские онтологии – URL: [https://ru.wikipedia.org/wiki/Открытые\\_биомедицинские\\_онтологии](https://ru.wikipedia.org/wiki/Открытые_биомедицинские_онтологии) (дата обращения: 31.12.2018).
3. The National Center for Biomedical Ontology. – URL: <https://www.bioontology.org/> (дата обращения: 31.12.2018); The OBO Foundry. – URL: <http://obofoundry.org/> (дата обращения: 31.12.2018).
4. The CIDOC Conceptual Reference Model (CRM). – URL: <http://www.cidoc-crm.org/> (дата обращения: 31.12.2018).
5. Ontologies for e-Government. – URL: <http://www.oegov.us/> (дата обращения: 31.12.2018).
6. d'Aquin M., Noy N.F. Where to Publish and Find Ontologies? A Survey of Ontology Libraries // Web Semantics: Science, Services and Agents on the World Wide Web. – 2012. – Vol. 11. – P. 96-111. – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293483/> (дата обращения: 31.12.2018).
7. Berners-Lee T. The next web // TED2009. – URL: [https://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web](https://www.ted.com/talks/tim_berniers_lee_on_the_next_web) (дата обращения: 31.12.2018).
8. Abele A., McCrae J.P., Buitelaar P., Jentzsch A., Cyganiak R. Linking Open Data cloud diagram 2017. – URL: <http://lod-cloud.net> (дата обращения: 31.12.2018); Linked Data. – URL: [https://en.wikipedia.org/wiki/Linked\\_data](https://en.wikipedia.org/wiki/Linked_data); <http://linkeddata.org/> (дата обращения: 31.12.2018).
9. DBpedia. – URL: <http://wiki.dbpedia.org/>; <http://wiki.dbpedia.org/datasets/dbpedia-version-2016-10> (дата обращения: 31.12.2018).
10. Ontology of DBpedia. – URL: <http://wiki.dbpedia.org/services-resources/ontology> (дата обращения: 31.12.2018).
11. DBpedia Mappings Wiki. – URL: [http://mappings.dbpedia.org/index.php/Main\\_Page](http://mappings.dbpedia.org/index.php/Main_Page) (дата обращения: 31.12.2018).
12. Semantic Web for Earth and Environmental Terminology (SWEET). – URL: <https://sweet.jpl.nasa.gov> (дата обращения: 31.12.2018).
13. Semantic Data Wiki and Linked Data Publishing Engine. – URL: <https://ontowiki.net/>; <https://docs.ontowiki.net/> (дата обращения: 31.12.2018).
14. Open Ontology Repository (OOR) Initiative - Home Page. – URL: <http://www.oor.net/>; <http://ontologyforum.org/index.php/OpenOntologyRepository>; <http://ontolog.cim3.net/wiki/OpenOntologyRepository.html#nid17YN> (дата обращения: 31.12.2018).
15. A free, open-source ontology editor and framework for building intelligent systems Protégé. – URL: <https://protege.stanford.edu/> (дата обращения: 02.02.2018).
16. Protege\_Ontology\_Library. – URL: [https://protegewiki.stanford.edu/wiki/Protege\\_Ontology\\_Library](https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library) (дата обращения: 02.02.2018).
17. Ontology Design Patterns.org (ODP). – URL: <http://ontologydesignpatterns.org> (дата обращения: 02.02.2018).
18. Cognitum. – URL: <http://www.cognitum.eu/> (дата обращения: 02.02.2018).
19. Fluent Editor for PC. – URL: <http://www.cognitum.eu/semantics/FluentEditor/> (дата обращения: 02.02.2018).
20. Web-сервер онтологий системы ЭЗОП. – URL: <http://ontoserver.rsuh.ru/ezop/> (дата обращения: 02.02.2018).
21. Сервер проекта системы ЭЗОП. – URL: <http://ezop-project.ru/> (дата обращения: 02.02.2018).

22. Тексты программ системы ЭЗОП. – URL: [https://github.com/beniaminov/ezop\\_server](https://github.com/beniaminov/ezop_server); <https://github.com/beniaminov/WebEzop> (дата обращения: 02.02.2018).
23. Бениаминов Е.М. Работа в системе коллективного формирования библиотек онтологий ЭЗОП (руководство пользователя). – М.: РГГУ (препринт), 2015. – 49 с. – URL: [http://beniaminov.rsuh.ru/User\\_guideEzop.pdf](http://beniaminov.rsuh.ru/User_guideEzop.pdf) (дата обращения: 02.02.2018).

*Материал поступил в редакцию 05.02.18.*

#### **Сведения об авторе**

**БЕНИАМИНОВ Евгений Михайлович** – доктор физико-математических наук, профессор, заведующий кафедрой математики, логики и интеллектуальных систем Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российский государственный гуманитарный университет» (РГГУ), Москва  
e-mail: [ebeniamin@yandex.ru](mailto:ebeniamin@yandex.ru)