

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ДЛЯ МОНИТОРИНГА НАУЧНОЙ ДЕЯТЕЛЬНОСТИ СОТРУДНИКОВ ВУЗА

Л. Р. Сулейменова, С. Ж. Рахметуллина

*Восточно-Казахстанский государственный технический университет
имени Даулета Серикбаева, 070010, Усть-Каменогорск, Казахстан*

УДК 004.42

С каждым годом быстрым темпом растет объем информации, относящейся к научно-исследовательской деятельности. С ростом объемов информации актуальной задачей является повышение эффективности использования этих информационных ресурсов. Были разработаны множество моделей для решения этой проблемы. В этой статье, после сравнения некоторых из этих моделей, была выбрана модель SWRC.

Ключевые слова: онтологическая модель научной деятельности, ЕНИП, Dublin Core, CERIF, SWAN, SPAR, SWRC.

Введение

Управление вузом ориентировано на достижение университетом стратегических целей. Принятие рациональных и обоснованных решений в процессе управления вузом возможно только на основе обработки больших объемов различной информации с применением информационных технологий. Эти данные помогают организации в сборе идей и видений, распознавании и использовании возможностей и угроз для доступа к конкурентным преимуществам. Обработка больших объемов данных предполагает создание единого интерфейса для некоторой совокупности независимых друг от друга источников данных. Одной из основных проблем является гетерогенность обрабатываемых данных. Их гетерогенность проявляется на различных уровнях: системном, синтаксическом, структурном, семантическом, а также применении различных форматов и моделей представления данных. Некоторые источники имеют свойство расширяться со временем. Эти проблемы делают задачу интеграции данных трудоёмкой, необходимо учитывать не только структурные и синтаксические различия информационных ресурсов, а также семантику. Авторы статьи [1] предлагают сосредоточить внимание на семантической обработке данных, а не на обработке данных. В Послании Президента РК Н. Назарбаева народу Казахстана «Третья модернизация Казахстана: глобальная конкурентоспособность» 2017 года были обозначены важные приоритеты в области развития, в том числе образования. Для дальнейшего развития республики он делает ставку на развитие научного потенциала на базе вузов. Поэтому создание структурной базы знаний с учетом особенностей данных о образовательно-научной деятельности ППС для принятия эффективных решений необходимо для каждой организации образования.

1 Онтологии

Многие страны и вузы разработали свои системы управления научной информацией на основе SEmabtic Web. Semantic Web предназначен для создания стандартного представления, которое может обеспечить значимые связи между различными наборами данных для содействия интеграции и коммуникации. Как семантический инструмент, онтология определяется как структурированные знания в определенной области, которое разрабатывается по концептам этой предметной области и отношениями между ними. Онтология — это четкое и формальное определение базы знаний, состоящее из понятий (или классов), ролей (или свойств) между экземплярами и ограничениями, с набором элементов и индивидами или экземплярами, которые определяют базу знаний [2]. Данное понимание полностью согласуется с известным определением онтологии Тома Грубера [3]. Существует множество моделей описывающих научную деятельность организации. В

этих моделях учитывается публикационная активность и индексы цитирования, также патенты и авторские свидетельства, участие в НИОКР. Были рассмотрены существующие онтологические модели описывающие научные труды ЕНИП, Dublin Core, CERIF, SWAN, SPAR, SWRC.

ЕНИП

В рамках российской инициативы по организации Единого Научного Информационного Пространства (ЕНИП) была разработана модель ЕНИП для интеграции разнородных научных информационных и программных ресурсов отдельных научных учреждений, предоставлении пользователям более эффективных средств интеграции и поиска информации, научной коммуникации, сотрудничества и совместной работы. Система ЕНИП работает такими ресурсами, как персоны, публикации, организации, подразделения и проекты [4]– [5].

Dublin Core

Dublin Core относится к классу онтологии, которые не в прямую описывают научную деятельность, а к классу онтологии, которые используются при описании результатов публикационной деятельности. Стандарт метаданных Dublin Core — это набор, предназначенный для описания широкого спектра ресурсов. Он был разработан в ответ на необходимость улучшения поиска информационных ресурсов, особенно во Всемирной паутине [6].

CERIF

CERIF (Common European Research Information Format) это стандарт для управления и обмена исследовательскими данными, то есть информацией об исследователях, организациях, проектах, результатах и финансировании, вытекающих из исследовательского процесса. Он предоставляет модель данных, которая может использоваться для описания области исследования, включая отношения между составными частями, и как они изменяются со временем [7]. CERIF начали разрабатывать в начале 1980-х годов, который исследовал связывание баз данных исследовательской информации. Информация об исследовании — это информация об исследовательских организациях, таких как люди, проекты, организации, публикации, патенты, продукты, финансирование или оборудование и т.д. и отношения между ними. Со временем модель была расширена и является открытой. Информационные системы позволяют структурировать, хранить, поддерживать, обмениваться, получать доступ, распространять или оценивать содержащуюся в них информацию. CERIF — мощный инструмент для создания масштабируемых и ориентированных на качество информационных систем. Здесь большое внимание уделяется описанию процессов исследовательской деятельности и фиксации хозяйственной стороны данного процесса — формальное фиксирование проведенных измерений, используемое и закупаемое оборудование, место проведения исследований, размер исследовательской группы, описание навыков каждого из сотрудников, финансирование и премирование, участие в подотчетных конференциях, публикации трудов и их цитирование, отслеживание формальных результатов исследований. Классы имеют большое количество связей для детального описания исследовательской деятельности [8].

SWAN

SWAN (Semantic Web Applications in Neuromedicine) представляет собой набор онтологических модулей для представления научного дискурса в биомедицинских исследованиях [9]. Данная онтология также как и Dublin Core относится к классу онтологии, которые описывают публикационную деятельность. Изначально онтология SWAN была разработана для удовлетворения требований проекта SWAN. Проект SWAN направлен на разработку практической, общей, семантически-структурированной основы для научных рассуждений в области биомедицины, в частности в нейро-медицине. Онтология рассуждений в основе SWAN — это онтология о том, что сказано, а не о согласованных объективных фактах. Онтология SWAN была интегрирована с Science Collaboration Framework [SCF]. Она предназначена для поддержки междисциплинарных ученых в публикации, аннотации, обмене и обсуждении контента, таких как статьи, перспективы, интервью и новости, а также утверждение личных биографий и научных интересов. Эти веб-материалы затем могут быть связаны с внешними гетерогенными хранилищами знаний ресурсов. Существующая архитектура позволяет заинтересованным сторонам принимать значимые компоненты. Онтология SWAN для представления научного дискурса, довольно широко, во многих других областях науки, обеспечивая при

этом важный уровень интеграции со всеми приложениями на основе информации SWAN. Онтология SWAN написана в OWL. Подъязык OWL-DL был выбран так, как он предлагает большую выразительность, не теряет вычислительной полноты и разрешимости (все вычисления завершаются за конечное время) систем рассуждений. Резонеры полезны для выявления ошибок и несоответствий в онтологии, а также ошибок и противоречий в наборах данных. Более того, в OWL онтологии могут быть модульными. Онтология SWAN организована в трех типах модулей: Basic: базовые модули представляют собой стандартный блок онтологии. Они охватывают темы, которые являются достаточно общими для включения в каждое распределение онтологии; Extension: модули расширения охватывают темы, которые могут быть связаны только с некоторыми областями науки; Distribution: дистрибутивы представляют собой модули, которые включают в себя все базовые модули и расширения, необходимые для обслуживания определенного домена [10].

SPAR

SPAR (Semantic Publishing and Referencing Ontologies) — это набор из 8 независимых онтологий OWL 2 DL для создания машиночитаемых метаданных в RDF для всех аспектов семантической публикации и перекрестных ссылок: описание документов, идентификаторы библиографических ресурсов, типы Цитат, библиографические ссылки, роли агентов, библиометрические данные. Все онтологии SPAR могут использоваться вместе или независимо друг от друга. Он обеспечивает достаточно детальную категоризацию отношений, которые могут возникать между научными материалами в электронном виде, и воплощающими их связями. Все онтологии SPAR были разработаны в соответствии со следующими принципами:

- решение требований участников;
- должен существовать обширный диалог с издателями и членами академических сообществ с целью уточнения их требований;
- интерактивные онтологические модули;
- минимальные логические ограничения;
- логические ограничения, например ограничения домена и диапазона для свойств, должны добавляться только там, где они строго необходимы;
- повторное использование существующих словарей [11].

SWRC

Онтология SWRC (Semantic Web Research Community) описывает ключевые сущности в научном сообществе. Трудно найти четкую документацию по SWRC онтологии. Онтология SWRC в целом моделирует ключевые субъекты, относящиеся к типичным научным сообществам и отношениям между ними. Текущая версия онтологии включает в общей сложности более 70 понятий в таксономии и более 70 свойств объекта. Все объекты обогащены дополнительной информацией. SWRC включает в себя шесть концепций высшего уровня, а именно концепции Person, Publication, Event, Organization, Topic и Project [12].

2 Сравнительный анализ

Анализируя выше описанные онтологические модели, которые описывают научные труды или научную деятельность для разработки системы мониторинга наиболее подходят онтологии CERIF и SWRC. Данные онтологии имеют много общих классов. Обе онтологии являются открытыми. Но онтология CERIF описывает научно-исследовательские работы с точки зрения финансирования и, следовательно, отчетности, а онтология SWRC направлена на описание с позиции общих характеристик научно-образовательной деятельности.

Основные преимущества SWRC:

- выразительность;
- возможность задания порядка следования авторов публикации;

- его сущности и атрибуты охватывают большинство объектов и атрибутов, представляющие интерес для научно-образовательной организации;
- SWRC обеспечивает межоперационную и легкую интеграцию данных между различными приложениями, например между различными порталами. Сравнительный анализ моделей приведен в таблице.

Таблица 1: Сравнение онтологических моделей

Модели	Назначение	Основные классы				
		Организация	Персона	Публикации	События	Проекты
ЕНИП	Поддержка и управление научной информацией	+	+	+	-	+
SPAR	Детальная категоризация отношения между научными материалами (цитирование, библиографические ссылки и т.п.)	+	+	+	-	-
SWAN	Детальная категоризация отношения между научными материалами (цитирование, библиографические ссылки и т.п.)	-	+	+	-	-
Dublin Core	Поиск информационных ресурсов	-	-	+	-	-
CERIF	Поддержка и управление научной информацией	+	+	+	+	+
SWRC	Поддержка и управление научной информацией	+	+	+	+	+

Главное преимущество SWRC в том, что там уже имеются классы и свойста, описывающие не только научную, но и образовательную деятельность. Онтология SWRC наиболее полно подходит для описания деятельности вуза. Данная модель будет дополнена некоторыми классами и свойствами, которые описывают образовательную деятельность сотрудника, а именно читаемые дисциплины их цикл и содержание. Так как, анализе деятельности ППС главным фактором, который оказывает влияние на научный результат является не только количественный объем нагрузки, также качественный состав читаемых дисциплин.

3 Результаты

Достоинствами онтологии являются их потенциальные свойства для решения таких задач, как формализация, интеграция, анализ данных и обмен знаниями и их повторное использование. Для создания модели позволяющей оценивать научные достижения сотрудников SWRC будет расширена с добавлением новых классов, свойств и отношений.

Заключение

В статье рассматриваются различные модели, описывающие научные труды и научную деятельность. Проведено сравнение между онтологиями CERIF и SWRC, определены их схожие свойства и отличительные особенности каждой онтологии. Для создания онтологической модели за основу была выбрана модель SWRC.

Список литературы

- [1] J. Brank, M. Grobelnic, and D. Mladenic, "A survey of Ontology evaluation techniques," in In Proceedings of the Conference on Data Mining and Data Warehouses(SiKDD 2005), Ljubljana, Slovenia, 2005.
- [2] Q. Lu, "OntoKBEval: a support tool for OWL Ontology evaluation," Unpublished master's thesis, Concordia University, Montreal, Quebec, Canada, 2006.
- [3] Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. — Спб.: Питер, 2000. — 384 с.
- [4] А.А.Бездушный, А.Н.Бездушный, В.А.Серебряков, В.И.Филипсов. Интеграция метаданных Единого Научного Информационного Пространства РАН. Вычислительный центр РАН. Москва, 2006, 238 с.
- [5] Кулагин М.В., Серебряков В.А. Информационное пространство РАН (Проекты и реализация, 1998-2013) // Научный сервис в сети Интернет: труды XVIII Всероссийской научной конференции (19–24 сентября 2016 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2016. — С. 194–222.
- [6] DCMi Home: Dublin Core® Metadata Initiative (DCMI). <http://dublincore.org/>
- [7] An introduction to CERIF: http://www.ukoln.ac.uk/rim/documents/Introduction_to_CERIF_1.0.pdf
- [8] CERIF-1.6 EuroCRIS: <http://www.eurocris.org/cerif/feature-tour/cerif-16> (дата обращения 23 мая 2017 г.)
- [9] P. Ciccarese, ed., Semantic Web Applications in Neuromedicine (SWAN) Ontology. W3C Interest Group Note 20 October 2009, <http://www.w3.org/2001/sw/hcls/notes/swan/>
- [10] Semantic Web Applications in Neuromedicine (SWAN) Ontology: <https://www.w3.org/TR/hcls-swan/>
- [11] Галимов А.А. Разработка онтологической модели публикаций // NB: Кибернетика и программирование. — 2015. — № 2. — С.98–106.
- [12] York Sure, Stephan Bloehdorn, Peter Haase, Jens Hartmann, Daniel Oberle: The SWRC Ontology - Semantic Web for Research Communities. Portuguese Conference on Artificial Intelligence (EPIA) 2005: 218–231

Сулейменова Лаура Разметоллановна — докторант 2 года обучения Восточно-Казахстанского государственного технического университета имени Даулета Серикбаева;
e-mail: laurasruk@mail.ru;

Разметуллина Сауле Жадыгеровна — к.т.н., Директор Департамента стратегии развития и мониторинга Восточно-Казахстанского государственного технического университета имени Даулета Серикбаева;
e-mail: SRakhmetullina@ektu.kz.

Дата поступления — 30 мая 2017 г.