

О ПОДХОДАХ К КЛАССИФИКАЦИИ ЭЛЕКТРОННЫХ КОЛЛЕКЦИЙ НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ

Ермолов П. П.^{1,2}

¹Севастопольский национальный технический университет,
кафедра радиотехники и телекоммуникаций

ул. Университетская, 33, г. Севастополь, 99053, Украина

²Крымский научно-технологический центр им. проф. А. С. Попова

а/я 10, г. Севастополь, 99057, Украина

e-mail: 10.99057@gmail.com

Аннотация — Изложены два подхода к классификации электронных коллекций научных информационных ресурсов: по специализации (структура метаданных, уровень специализации, объем ресурса) и по принадлежности к категории «база знаний» (наличие/отсутствие отношений между элементами предметной области и/или функций интерпретации). Приведены примеры, иллюстрирующие оба подхода. Отмечены проблемы в создании узкоспециализированных электронных коллекций.

I. Введение

Концепция Web 2.0 (реализация принципа «коллективного знания») и ее развитие — Web 3.0 (привлечение профессионалов для реализации этого принципа) изменяет многие области человеческой деятельности, в том числе и технологию научных исследований. В дополнение к электронным ресурсам с многомиллионными базами данных создаются узкоспециализированные порталы научных знаний с существенно меньшими по объему, но более специализированными и структурированными базами данных, создаваемые отдельными исследователями или группой исследователей. Для создания электронных коллекций научных информационных ресурсов интенсивно разрабатывается широкий спектр инструментальных средств, в числе которых рассмотрены в [1] средства Greenstone, LibOnline, ИРБИС, EPrints, АБИС, Virtua, ИСХИ-М. В цитируемой работе детально описаны технические особенности систем: схема метаданных, поддерживаемые форматы документов, лицензия (свободная/платная) и проч.

Вследствие бурного развития указанного направления становится актуальной классификация создаваемых электронных коллекций научных информационных ресурсов, по крайней мере, по двум дополнительным критериям:

1. По специализации, которая включает в себя такие характеристики, как структура метаданных, уровень специализации и объем ресурса;

2. По принадлежности к категории «база знаний» (наличие/отсутствие отношений между элементами предметной области и/или функций интерпретации).

Существующие классификации (см., напр. [2]), не отражают перечисленные выше характеристики и учитывают только такие параметры, как *физическая природа (материальная конструкция), знаковая природа, целевое назначение, жанр (виды документов и т. п.)*.

Целью настоящей работы является представление подхода к классификации электронных коллекций научных информационных ресурсов по специализации и принадлежности к категории «база знаний».

II. Основная часть

В основу предлагаемой классификации по специализации положено то, что электронные коллекции информационных ресурсов создаются, как пра-

вило, двумя различными категориями специалистов: информационных порталов (библиотек), с одной стороны, или/и исследовательских групп или отдельными исследователями по узкому направлению исследований с другой. В соответствии с этим предлагается следующий вариант классификации:

Табл. 1. Классификация электронных коллекций научных информационных ресурсов по специализации

Уровень специализации коллекции	Узкоспециализированные коллекции	Специализированные коллекции	Универсальные коллекции
Составители коллекции	Специалисты исследовательских групп	Специалисты информационных порталов (библиотек) или/и специалисты исследовательских групп	Специалисты информационных порталов (библиотек)
Структура метаданных	Специализированная	Стандартная или/и специализированная	Стандартная или/и специализированная
Объем коллекции	до 1000	от 1000 до 10000	свыше 10000

Среди стандартных систем метаданных наиболее распространенными и признанными являются системы семейства MARC, Дублинское ядро, а также система метаданных LOM [2]. Не останавливаясь на особенностях каждой из систем, отметим, что каждая из них имеет свои особенности (и различия), а также то, что в среде разработчиков и пользователей такого рода систем одной из самых актуальных проблем является проблема интеграции описаний ресурсов, представленных в этих системах метаданных [2].

Специализированные системы метаданных по определению не могут быть универсальными, хотя и содержат в себе элементы стандартных систем.

Основной характеристикой, которая дает основание отнести электронную коллекцию научных информационных ресурсов к категории базы знаний, является использование онтологического подхода при построении коллекции. При этом в схеме метаданных должны быть формализованы не только понятия предметной области, но, как минимум, отношения между этими понятиями.¹

¹ Формальное описание онтологии, которое считается классическим, содержит конечное множество понятий предметной области, конечное множество отношений между понятиями и конечное множество функций интерпретации [3].

В этом смысле коллекции всех уровней специализации могут обладать или не обладать признаками базы знаний. Проиллюстрируем сказанное характерными примерами.

1. Научная электронная библиотека eLibrary.ru. Универсальная коллекция. Объем — более 14 млн. экземпляров. Составители — специалисты информационного портала. Обладает всеми признаками базы знаний, в т. ч. функциями интерпретации. Так, в частности, для организаций, которые представлены авторами публикаций, может быть сформирован полный список публикаций, список статей, цитирующих эти публикации, а также статистическая информация, позволяющая проанализировать публикационную активность организации по различным критериям.

2. Специализированная коллекция «Влияние миллиметровых волн КВЧ-диапазона на биологические объекты» [3]. Объем коллекции — около 4 тыс. экземпляров. Составители — специалисты Центральной библиотеки Пушинского научного центра Российской академии наук. В качестве оболочки ресурса используется СУБД Microsoft Office Access. Набор метаданных включает в себя следующее: фамилии авторов, название статьи, год издания, научное предприятие, где проводились изыскания, количество библиографических ссылок. В «ручном» режиме проводится статистический анализ ресурса [3]. Коллекция не может быть отнесена к категории базы знаний.

3. Археологический портал знаний sati.archaeology.nsc.ru. Специализированная коллекция. Ресурс содержит 4269 объектов и 15363 отношений. На портале представлены знания об археологии и этнографии как научной дисциплине, ее подразделах, методах и техниках исследования, используемым терминам и понятиям, а также информация о персоналиях исследователей, сообществах и организациях, включенных в процесс исследования. Структуризация знаний и данных выполнена на основе онтологии, информация на портале представлена в виде взаимосвязанных информационных объектов. Доступ к знаниям и данным портала осуществляется путем навигации по дереву понятий онтологии и через средства поиска — простого (по ключевым словам) и расширенного (с использованием понятий онтологии и заданием ограничений на значения их атрибутов и связей). При навигации по portalу обеспечивается возможность просмотра описания интересующего ресурса, выбора определенного класса информационных объектов, просмотра и фильтрации списков конкретных информационных объектов, навигации по ним и информационным объектам. Портал является классическим примером знаниевого построения коллекции научного информационного ресурса.

В настоящее время, к сожалению, практика создания узкоспециализированных электронных коллекций научных информационных ресурсов практически отсутствует. Можно назвать две основные причины такого положения: недостаточное внимание «узких» специалистов к разработке методологических основ разрабатываемого направления, с одной стороны, и недостаточная осведомленность в существовании технологий создания такого рода ресурсов [4], не требующих дополнительного привлечения специалистов по инженерии знаний, с другой. В качестве примера создания такого рода методологии можно привести работу автора в области историографии развития радиотехнологий [5]. Проблема актуальности разработки массовой технологии проектирования интеллектуальных систем для научных исследований была поднята в работе [6], но она далеко не исчерпана.

III. Заключение

В докладе изложены два подхода к классификации электронных коллекций научных информационных ресурсов: по специализации (структура метаданных, уровень специализации, объем ресурса) и по принадлежности к категории «база знаний» (наличие/отсутствие отношений между элементами предметной области и/или функций интерпретации). Приведены примеры, иллюстрирующие оба подхода. Отмечены проблемы в создании узкоспециализированных электронных коллекций. Материал доклада будет полезен как специалистам информационных порталов (библиотек), так и исследовательских групп.

IV. Благодарности

Автор выражает благодарность к. т. н. Загорульکو Ю. А., ознакомившемуся с материалом доклада и сделавшему ряд уточнений.

V. Список литературы

- [1] *Игнатова И. Г., Жданова И. В., Соколова Н. Ю., Шевнина Ю. С.* Инструментальные средства создания и сопровождения электронных коллекций информационных ресурсов // Наука и образование : электронное научно-техническое издание. 2009. № 5. С. 2—2.
- [2] *Антопольский А. Б., Ауссем В. И.* Типология информационных ресурсов в стандартных системах метаданных : анализ и проблемы интеграции // Информационные ресурсы России. 2006. № 5. С. 3—6.
- [3] *Никоненко А. А.* Обзор баз знаний онтологического типа // Искусственный интеллект. 2009. № 4. С. 208—219.
- [4] *Загорульکو Ю. А.* Технология построения порталов научных знаний : опыт применения, проблемы и перспективы // 21-я Международная Крымская конференция «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2011) : материалы конф. в 2 т. (Севастополь, 12—16 сент. 2011 г.). Севастополь : Вебер, 2011. Т. 1. С. 51—54.
- [5] *Ермолов П. П.* Предметные онтологии в общей и региональной историографии развития радиотехнологий // 20-я Международная Крымская конференция «СВЧ-техника и телекоммуникационные технологии» (КрыМиКо'2010) : материалы конф. в 2 т. Севастополь, 13—17 сент. 2010 г. Севастополь : Вебер, 2010. Т. 1. С. 73—78.
- [6] *Голенков В. В., Гулякина Н. А.* Принципы построения массовой семантической технологии компонентного проектирования интеллектуальных систем // Материалы Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2011). Минск, 10—12 февраля 2011 г. С. 21—58.

ON THE APPROACHES TO THE CLASSIFICATION OF COMPUTER-AIDED COLLECTION OF INFORMATION RESOURCES

Yermolov P. P.,^{1,2}

¹Sevastopol National Technical University, Radioengineering & Telecommunication dept 33, Universitetskaya str, Sevastopol, 99053, Ukraine

²Popov Crimean Scientific and Technology Center P. O. Box 10, Sevastopol, 99057, Ukraine e-mail: 10.99057@gmail.com

Abstract — Two approaches taken to classifying the computer-aided collections of information resource: one — according to specialization (structure of metadata, specialization level, resource volume) and the other — in terms of belonging to the “knowledge base” category (presence/lack of relationships between the elements of a subject area and/or the interpretation functions) are presented. Examples illustrating both of the approaches are cited. The problems with the creation of strictly specialized computer-aided collections are considered.