

ОРГАНИЗАЦИЯ ПЕРСОНАЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА НА ОСНОВЕ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ АВТОРА

Е.И. Моисеев^{1,2}, А.А. Муромский¹, Н.П.Тучкова¹, А.М. Меденников¹

¹ Федеральное государственное бюджетное учреждение науки Вычислительный центр им. А. А. Дородницына РАН

² Федеральное государственное образовательное учреждение высшего профессионального образования "Московский государственный университет имени М.В.Ломоносова", факультет ВМиК

ВВЕДЕНИЕ

Одним из первых, кто поставил вопрос о проблеме информационного обмена в научной среде, был профессор математики университета Торонто Кеннет Мэй (Kenneth O. May). Ему же принадлежит авторство одного из первых цифровых проектов по созданию словарей и тезаурусов по математике совместно с издательством McGraw-Hill, поддержанных канадским советом по культуре в 70-х г. прошлого столетия. Ранее тезаурусы рассматривались, в основном, как инструменты для языковых словарей, например, тезаурус английских слов Роже [1], который переиздается уже больше столетия (на сегодня продано 32 миллиона экземпляров). К.О. Мэй выступил на международной конференции в Лос-Аламосе, посвященной компьютерной истории с докладом: «Перспектива для программистов» (Historiography: A perspective for computer scientists), где рассказал о своих исследованиях в области истории математики (с периода 1800 г.). Основной тезис его выступления: «Вы не можете сделать научной работы и систематически работать ни в какой области без словарей в этой области» [2] сохранил актуальность и приобрел новое практическое значение в эпоху интернет-технологий и междисциплинарных исследований.

Действительно, о каких исследованиях может идти речь, если нет ясности в терминологии, а публикации ищутся в интернете по ключевым словам. В настоящий период активной информатизации общества возникла реальная проблема использования терминологии предметной области и систематизации публикаций в целях организации (систематизации) собственного, индивидуального информационного цифрового пространства. Данные о персоне: авторе публикации, участнике научного коллектива и проекта, члене административной структуры и пр. многократно дублируются, обрастая неточностями. Такая же картина с публикациями, которые помещаются в бесчисленное количество информационных систем, в то время как сам автор, зачастую не всегда имеет к ним доступ в силу финансовых или иных организационных проблем, возникающих при работе с библиотечными ресурсами и системами цитирования. Все это усугубляется ростом информационного потока в интернет, нарастанием информационного шума при запросах и как следствие, увеличением времени, необходимого для поиска и обработки актуальной информации, снижением конкурентоспособности научных исследований.

Один из подходов, который позволяет подойти к решению проблем информационного поиска и обмена информацией в электронной среде, основан на представлении терминологии предметной области в виде тезауруса. Свои первые словари и тезаурусы по математике К.О. Мэй записывал на разлинованных карточках и переносил на перфокарты [3, 4]. Современная реализация этого подхода основывается на использовании понятия онтологии предметной среды [5]. В настоящей работе предлагается ограничиться вопросом представления предметной области автора [6] и описанием персонального информационного пространства, под которым понимается совокупность информационных элементов и инструментов, позволяющих автору распоряжаться своими публикациями (накапливать, редактировать, создавать, систематизировать), накапливать публикации и ссылки других авторов, исходя из собственных интересов и потребностей [7], организовывать и участвовать в рассылке, подписке, рецензировании публикаций и др.

ОСНОВНЫЕ ТЕРМИНЫ И ПОНЯТИЯ

Для описания персонального информационного пространства автора (ПерИнфо) требуется определить такие данные для конкретного участника информационной среды, как: тезаурус [1, 8], предметная область [7, 9], онтология, информационная среда, публикации, персоны, подписка, рассылка (и некоторые другие по запросу).

Ключевые слова - множество ключевых слов (КС) из публикаций автора, соавторов, рецензентов, рецензируемых работ, цитируемых и/или внесенный в список, как интересные для автора (ПублИнт).

Тезаурус – разрабатывается тезаурус адресата (ТА). ТА включает словарь ключевых слов с парадигматическими связями, из публикаций автора, соавторов, рецензентов, рецензируемых работ, цитируемых и/или внесенный в список ПублИнт.

Предметная область – предметная область автора (ПОавт). ПОавт задается на основе множества КС и технологии описания онтологии предметной области.

Онтология – онтология предметной области автора и объединение онтологий различных авторов, для которых можно считать, что они работают в одной предметной области, исходя из списков их ключевых слов.

Информационная среда – информационная среда автора (ИСА) совокупность информации и инструментов ее обработки. ИСА состоит из множества публикаций, персон (авторов публикаций), множества ключевых слов (и других вторичных документов публикаций), инструментов описания онтологии ПОавт, ТА, инструментов коммуникации (подписка на публикации, рассылка информации о публикациях) и других, задаваемых индивидуально.

Публикации – публикации автора (ПублА), публикации соавторов (ПублСА), цитируемые публикации (ПублЦит), публикации по ПОавт из списка, формируемого автором по подписке, или из списка ПублИнт.

Персоны – авторы (А), соавторы (СА) и другие участники коммуникации (УК).

Подписка – сервис индивидуальной подписки на публикации из ПОавт по КС, реализованный в коммуникационной системе для А.

Рассылка – сервис индивидуальной рассылки ПублА.

СТРУКТУРА И ФУНКЦИИ

Состав информационной среды автора, структура информации, функции автора, участников информационной среды и коммуникационной среды определяются целевым назначением разрабатываемого ресурса.

Под описанием функций персоны (адресата) подразумевается список возможных действий, поддерживаемых системой, для зарегистрированного пользователя (автора).

Автор работает с информацией и ему доступны:

- ввод информации – создание нового раздела информации;
- редактирование существующего раздела информации;
- подписка на информацию:
 - по ключевым словам,
 - по авторам,
 - по названию;
- рецензирование;
- -научная переписка и совместное редактирование,
 - «информация» в данном случае:
 - другой адресат,
 - другая предметная область,
 - смежная предметная область,
 - публикации,
 - вторичная информация о публикациях,
 - отчет по публикациям,
 - корректура.

Основное назначение предложенного описания предметной области адресата на основе технологии тезаурусного представления знаний – создание условий для поиска контактов и публикаций, подготовки публикаций, рецензий и отчетов по тематике адресата – научного работника. Необходимость такого сетевого ресурса определяется тем, что «жизнь» публикации во многом переместилась в интернет. В качестве примера представления таксономических отношений рассмотрим статью лексико-семантического указателя (ЛСУ) по ОДУ и терминологическое обозначение уравнения с частными производными математической физики и смежных областей.

Пример 1. Статья ЛСУ (класс СЛОВАРЬ°, значение слота°°), экземпляр:

- ° ОДУ первого порядка, разрешенное относительно производной, именное
- °° Абеля ОДУ первого рода
- °° Абеля ОДУ второго рода
- °° Вернуллы ОДУ
- °° Дарбу ОДУ
- °° Муанье ОДУ
- °° Пирсона ОДУ
- °° Риккати ОДУ обобщенное
- °° Риккати ОДУ общее
- °° Риккати ОДУ специальное
- °° Риккати ОДУ каноническое
- °° Эйлера ОДУ первого порядка
- °° Якобы ОДУ первого порядка
- ° Коши задача ОДУ первого порядка, разрешенного относительно производной

Пример 2. Уравнения с частными производными:

класс СЛОВАРЬ°, АВТОР°, ПУБЛИКАЦИЯ°, СЛОТЫ°°

- Уравнение Лаврентьева–Бицадзе

◦

$$u_{xx} + \text{sign } y \cdot u_{yy} = 0$$

- Сабитов К.Б.
- Уравнения математической физики
- Учебное пособие для вузов
- М.: Высшая школа.
- 2003
- 255

ВОПРОСЫ РЕАЛИЗАЦИИ

Вопросы реализации включают объектное и функциональное описание системы. Описание классов, схемы взаимодействия, связей, платформы, коммуникации [10]. Для объектного описания используется технология задания онтологии ПОавт. За основу берется список ключевых слов со связям в виде ТА. Коммуникации в системе определяются как функции рассылки и подписки.

ЗАКЛЮЧЕНИЕ

Известно, что в настоящее время много усилий направлено на то, чтобы включить отдельного автора (персону) и результаты его исследований (публикации, доклады и др.) в различные профессиональные и маркетинговые базы данных в интернет. Что дает автору на практике такая, часто навязанная, услуга? Иногда, если подписка была осуществлена целенаправленно, предоставляется полезная информация, (например, система СОЦИОНЕТ [11]), но в большинстве случаев мы получаем поток спама, в котором нет времени и возможности разобраться. Таким образом, полезная и бесполезная информация борется в информационном пространстве каждого пользователя за право быть востребованной. Для научных публикаций и их использования особенно важно иметь актуальную информацию и желательно в систематизированном виде, пригодную для изучения и цитирования. Однако, сам автор и исследователь, как пользователь библиотечной, поисковой системы или системы цитирования, как правило, лишен возможности упорядочить необходимую информацию. В работе предлагается использовать ресурсы самого автора, как создателя собственной предметной области, и на основе ее описания организовать персональное информационное пространство.

Работа поддерживается РФФИ (проект 13-07-00334).

ЛИТЕРАТУРА:

1. Roget P.M. Roget's thesaurus of English words and phrases. Longman, London, 1852, Penguin Books Ltd; Revised edition, 1998, 1140 p.
2. May K.O. Historiography: A Perspective for Computer Scientists. Invited address to International Research Conference on the History of Computing, June 10, 1976, Los Alamos.
3. Enros P.C. May Dictionary and Thesaurus Projects, Scope and Status. Report Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, 1980.
4. Lewis A.C. Kenneth O. May and information retrieval in mathematics // *Historia Mathematica* V. 31, Issue 2, May 2004, P. 186–195.
5. <http://www.w3.org/standards/semanticweb/ontology> (последнее обращение 17.04.2014)
6. Моисеев Е.И., Муромский А.А., Тучкова Н.П., Меденников А.М. Описание научных исследований и представления предметных областей в коммуникационной среде на основе контролируемой лексики // «Научный сервис в интернет: все грани параллелизма». Труды международной конференции», 23 - 28 сентября 2013 г. Новороссийск. М.: Изд-во МГУ, 2013. 589 с. (с. 520-521)
7. Моисеев Е.И., Муромский А.А., Тучкова Н.П. О представлении и поиске научных результатов современными средствами в электронной информационно-поисковой среде. Отв. ред. Моисеев Е.И. М.: МАКС Пресс. 2009. 100 с.
8. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Тезаурус информационно-поисковый по предметной области: обыкновенные дифференциальные уравнения. М.: МАКС Пресс, 2005, 116 с.
9. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Об онтологии научного информационного пространства. Сообщения по прикладной математике ВЦ РАН. М: ВЦ РАН. 2013, 48 с. http://www.ccas.ru/depart/gusev/newtex/Pr2013_mmt.pdf
10. Моисеев Е.И., Муромский А.А., Тучкова Н.П. Об онтологии адресата в научно-информационной среде. // Информационные и математические технологии в науке и управлении /Труды XIX Байкальской Всероссийской конференции «Информационные и математические технологии в науке и управлении». Часть III. -Иркутск: ИСЭМ СО РАН, 2014. 245 с. (с. 158-166)
11. Паринов С.И. Соционет.ру как модель информационного пространства 2-го поколения// Информационное общество. 2001. № 1. С. 43-45. <http://socionet.ru/portal.html> (последнее обращение 15.05.2014)