

Методологические подходы к построению информационно-справочных систем по истории науки*

© В.Б.Баракхин, А.М.Федотов

Институт вычислительных технологий СО РАН
bar@ict.nsc.ru, fedotov@ict.nsc.ru

Аннотация

В работе описана методика построения информационно-справочных систем по истории науки. Основные принципы построения таких систем состоят в следующем.

Информация группируется вокруг персон, при этом подробно структурируются биографические данные в плане хронологии, географии и т.п. Библиографический список в описании персоны включает в себя, наряду с публикациями данного ученого, и публикации о нем самом. Четко отражена связь научной деятельности исследователя с формализованным описанием предметной области, в которой этот исследователь работал.

Приведенное описание включает в себя информационную модель справочника, особенности реализации подсистем создаваемой информационной системы, а также основные виды информационных запросов пользователя, необходимых для полноценной работы с системой.

1 Введение

С самого начала научных изысканий человечества каждое новое поколение исследователей пытается сохранить и преумножить полученные знания, передав их потомкам. История науки отражает открытие неизвестных ранее явлений и законов природы, рождение новых понятий, терминов, теорий, их эволюцию и влияние на жизнь человека. При этом не стоит забывать, что науку создают не абстрактные теории, а конкретные живые люди, чьи идеи так или иначе разрабатываются и поныне. «Мы стоим на плечах гигантов», — образно говорил Ньютон. Сохранение и культурного наследия выдающихся

предшественников — наш моральный долг перед их памятью.

Вполне очевидно, что наиболее удобной с точки зрения пользователя будет электронная форма справочника, позволяющая просматривать информацию о выдающихся исследователях, их результатах и публикациях, дающая возможность посредством пользовательских запросов получать хронологический, географический, тематический и прочие срезы описания предметной области или комбинацию этих срезов.

Поскольку творцами науки являются отдельные выдающиеся личности, информация в справочных системах по истории науки должна группироваться вокруг персон, при этом требуется подробное структурирование биографических данных в плане хронологии, географии и т.п. Разумеется, биографии ученых немыслимы без библиографического списка, который включает в себя, наряду с публикациями данного ученого, и публикации о нем самом. Наконец, необходимо четко отразить связь научной деятельности исследователя с формализованным описанием предметной области, в которой этот исследователь работал.

В настоящее время в Интернете имеется целый ряд справочников, содержащих биографии деятелей различных областей наук. Наиболее полно и качественно, на наш взгляд, представлены биографии математиков. Это объясняется, по меньшей мере, двумя причинами. Во-первых, основная масса ученых, занимающихся информационными технологиями, имеет математическое образование, и вполне естественно, что специалисты, создающие такие справочники (а подобная работа в большинстве случаев основана на энтузиазме небольших групп ученых), предпочитают работать с историей той области фундаментальной науки, которая им хорошо знакома. Во-вторых, формализованное описание предметных областей с учетом исторической ретроспективы, необходимое для адекватного отражения деятельности исследователей прошедших эпох, для многих разделов естественных наук затруднено тем, что естественнонаучные теории постоянно развиваются, причем нередко ранее принятые концепции

впоследствии полностью отвергаются, вследствие чего построение формальных моделей эволюции естественнонаучных теорий является сложной проблемой, решение которой, по-видимому, не допускает сколько-нибудь общих подходов. В этом смысле математика является счастливым исключением, поскольку она оперирует, согласно терминологии Канта, априорными синтетическими суждениями, что обеспечивает последовательное (практически без отвержения полученных ранее результатов) развитие математической науки.

Среди представленных в Интернете коллекций биографий математиков можно выделить портал истории математики, разработанный в шотландском Университете св. Андрея [1]. Здесь представлена наиболее полная коллекция биографий. Ресурс обладает развитыми поисковыми возможностями, имеет сортировку по годам жизни математика, а также по стране его рождения смерти, а также содержит хронологию основных математических открытий. Однако, к сожалению, данный портал не содержит ссылок на электронные публикации трудов, а также классификатора предметной области, что значительно снижает его справочную ценность.

Также стоит упомянуть проект The Mathematics Genealogy Project [2], разрабатываемый Университетом Северной Дакоты. Ресурс призван собрать информацию обо всех математиках, когда-либо получивших степень доктора математических наук. Обширная база включает информацию о диссертационной работе математика, название университета присвоившего степень, а также указывается предметная область, с использованием классификатора предметной области MSC2000, список учеников и список работ. Однако более полные биографические материалы берутся с описанного выше портала Университета св. Андрея. К недостаткам ресурса относится скудность предоставляемой информации (лишь некоторые статьи имеют указания на предметную область деятельности данного математика, а часть статей вообще не снабжена информацией даже о годе получения степени), а также то, что российские математики представлены в небольшом количестве (из общего числа 89300 количество упомянутых российских математиков — 1296).

Из российских ресурсов наиболее близкий аналог данного проекта — математический портал math.ru [3], разрабатываемый при поддержке Отделения математических наук Российской академии наук и Московского центра непрерывного математического образования. Посвященный истории математики раздел содержит, помимо подробного текста биографий как исторических, так и современных математиков, их библиографию. Однако поисковые возможности портала не поддерживают атрибутивный поиск и ограничены лишь поиском по фамилии.

Таким образом, ни один из имеющихся биографических порталов в области математики не

способен достаточно адекватно отразить персонифицированную ретроспективу развития математической науки. Порталы с подобными свойствами отсутствуют и для других разделов естественных наук.

Целью данной работы является описание методики построения информационно-справочных систем по истории науки, удовлетворяющих сформулированным выше требованиям. Это описание включает в себя информационную модель справочника, особенности реализации подсистем создаваемой информационной системы, а также основные виды информационных запросов пользователя, необходимых для полноценной работы с системой.

2 Основные коллекции системы

При создании информационной модели справочника нужно учитывать следующее соображение: стремление к максимальной полноте представления информации о каждом ученом может привести к неоправданной задержке работы над справочником. С другой стороны, ценность справочника для пользователя состоит в широте охвата биографий, позволяющей получить более или менее целостное представление о состоянии описываемой области науки в определенную эпоху, в том числе в конкретной стране, или же проследить за развитием какого-либо раздела науки. Ввиду этого, наряду с полной версией модели мы будем приводить и варианты ее упрощения, предназначенные для ускоренного первоначального наполнения справочника.

Справочник должен содержать следующие основные коллекции:

- а) персоны (анкетные данные);
- б) публикации, так или иначе связанные с персонами, представленными в справочнике;
- в) формализованное описание предметной области.

Кроме того, при создании подробной версии справочника целесообразно создание вспомогательных коллекций:

- учебные заведения;
- научные организации;
- научные сообщества (академии, редколлегии и т.п.);
- награды и премии

и др.

Документы из коллекции «Персоны» связываются с документами из прочих коллекций с использованием той или иной модели связей, позволяющей адекватно описывать различные аспекты соответствующих межличностных отношений. Достаточно универсальный характер имеет, например, модель направленных связей [4], в которой каждый тип связей обладает определенным набором атрибутов, характеризующих данное отношение. Отметим, что в роли значений

атрибутов связи могут выступать документ, входящие в ту или иную коллекцию. Так, при установлении связи между *персоной* и полученным этой *персоной результатом предметной области* в качестве атрибутов связи могут выступать атрибуты *публикации*, содержащей данный результат.

2.1 Представление анкетной информации о персонах

Необходимый набор информации о представленной в справочнике персоне должен включать в себя ее основные анкетные данные: ФИО, дата и место рождения, для умерших — дата и место смерти, а также названия страны (стран), с которой принято связывать профессиональную деятельность данной персоны.

Расширенный набор анкетной информации включает сведения о полученном образовании, местах работы, членстве в научных сообществах, наградах. Первоначально такие сведения могут быть частью текстового описания деятельности персоны, однако по мере развития системы информация подобного рода будет представлена посредством установления связи персоны с соответствующими вспомогательными коллекциями. В качестве атрибутов связей выступают даты событий, а также занимаемые должности (в широком смысле, включающем иерархический статус в академиях, редколлегиях и т.п.).

Каталожная карточка персоны может включать ссылку на файл с ее портретом.

2.2 Библиографическая информация

Библиографический каталог системы включает в себя описание публикаций, так или иначе связанных с персонами, представленными в справочнике. Можно выделить следующие виды публикаций:

- научные публикации, относящиеся к представленной в справочнике отрасли науки;
- прочие научные публикации;
- научно-популярные публикации;
- прочие публикации;
- интервью;
- биографии персоны;
- прочие публикации о персоне.

При этом одна публикация может относиться к разным видам: например, публикация об одной персоне может одновременно быть научно-популярной публикацией другой персоны, представленной в справочнике.

По мере развития системы возможно разбиение некоторых из перечисленных разделов на подразделы. Так, научные публикации могут подразделяться на монографии, статьи, учебники и т.п.

Описание публикации представляет собой, как минимум, библиографическую ссылку, оформленную в соответствии со стандартом. Для классификации публикаций крайне желательно

применять тот классификатор предметной области, который используется при создании ее тезауруса, о чем подробно сказано в п. 2.3.

В исключительных случаях, касающихся важнейших публикаций минувших эпох, описание может быть ограничено автором, названием и годом выпуска, при этом, по возможности, параллельно указывается современное общедоступное издание данной работы.

2.3 Тезаурус предметной области

Формализованное описание предметной области целесообразно представлять в виде тезауруса, т.е. словаря, содержащего все лексические единицы информационно-поискового языка данной области – дескрипторы (вместе с ключевыми словами, которые считаются синонимами этих дескрипторов), причем дескрипторы в словаре систематизированы по смыслу, а смысловые связи между ними эксплицитно выражены [5].

Нами разработана [6] технология создания такого тезауруса, позволяющая провести его построение с минимальным привлечением специалистов-экспертов по конкретным разделам математики, освободив их от значительного объема рутинной работы. Особенность предложенной технологии состоит в том, что в качестве списка ключевых слов и словосочетаний используется Предметный указатель «Математической энциклопедии» [7], содержащий более 27 тысяч терминов. В него входят названия всех статей энциклопедии, понятия (термины) определения которых приведены в статьях, а также упоминаемые в статьях наиболее важные результаты. В качестве дескрипторов из Предметного указателя выделяются названия статей «Математической энциклопедии», после чего устанавливаются связи между дескрипторами и ключевыми словами, входящими в состав Предметного указателя, которые встречаются в соответствующей статье.

Связи между дескрипторами и ключевыми словами установлены в соответствии с рекомендациями схемы Zthes [8], предназначенной для работы с использованием протокола Z39.50. Заметим, что данная схема предполагает использование весьма ограниченного набора типов отношений между терминами. Это сделано преднамеренно для большей совместимости.

Итак, между терминами устанавливаются следующие типы связей:

- **BT** – связь с родительским термином, т.е. с термином более широкого смысла;
- **NT** – связь с дочерним термином, т.е. с термином более узкого смысла. Связь BT-NT является взаимно обратной;
- **USE** – связь с термином, который используется вместо этого;
- **UF** – взаимно обратная связь USE;
- **RT** – связь, определяющая связанный термин;

- **LE** – связь между лингвистически эквивалентными терминами.

Связи VT и NT, а также USE и UF взаимно-обратны. Связи RT, а также LE симметричны.

Кроме того, производится определение типа термина в соответствии с рекомендациями Zthes, получившими развитие в [9]. Выделяются следующие типы терминов:

- **TT** – термин верхнего уровня, т.е. термин, не имеющий связанных терминов более широкого класса (терминов с типом связей VT);
- **NT** – не термин верхнего уровня, т.е. дескриптор, имеющий связи типа VT;
- **ND** – не основной термин;
- **NL** – фиктивный термин, т.е. термин, не используемый для индексации документов, но включенный в иерархию, чтобы указать логический базис раздела классов.

Определение типа терминов позволяет существенно упростить работу с тезаурусом.

Термины, входящие в тезаурус, должны быть классифицированы с использованием того или иного классификатора предметной области. Разумеется, классификатор должен быть широко распространенным и пользоваться признанием среди специалистов в данной предметной области, при этом весьма желательно, чтобы структура классификатора была достаточно простой. Названия разделов классификатора также включаются в тезаурус в качестве дескрипторов.

На первом этапе работы возможно ограничиться классификацией дескрипторов. Между дескрипторами, являющимися названиями разделов классификатора, и прочими дескрипторами устанавливаются связи вида NT, RT, LE, при этом для классификации следует использовать, по возможности, разделы максимально низкого уровня. После того, как дескриптор будет классифицирован, ключевым словам, связанным с ним отношениями VT, USE, RT, LE, приписывается тот же классификационный номер, что и дескриптору. Впрочем, это не исключает такой ситуации, что если дескриптор отнесен к классу не самого низкого уровня, то при последующей работе эксперта термины, связанные с дескриптором отношениями VT и USE, могут быть отнесены к классу более низкого уровня. В этом случае указанные термины сами станут дескрипторами.

Следует оговориться, что при создании тезаурусов большого объема, охватывающих историческую ретроспективу развития той или иной науки, в них можно включать устаревшие понятия (в том числе отвергнутые современной наукой), снабжая их соответствующей пометкой.

Наконец, термины тезауруса могут быть снабжены краткими словарными статьями, посвященными этим терминам.

3 Установление связей между коллекциями

Важнейшим этапом создания справочника является установление связей между основными коллекциями. Связь между персонами и публикациям вряд ли способна вызвать особые проблемы; в качестве атрибута этих связей достаточно указывать публикации из списка, приведенного в п. 2.2.

Менее тривиальным является установление связей между персонами и терминами из тезауруса. Практически любой термин, входящий в тезаурус предметной области, был включен в научный оборот благодаря тому, что соответствующему объекту или явлению была посвящена научная публикация, имеющая одного или нескольких авторов. Таким образом, в качестве атрибутов связи между персоной и термином из тезауруса могут выступать атрибуты соответствующей публикации (по крайней мере, год ее выхода в свет). Разумеется, чем меньше объем имеющегося тезауруса, тем более общий характер имеют его термины, и каждый из них в принципе может быть связан с большим числом персон.

Установление связей между персонами и терминами тезауруса (которые обязательно имеют классификационный признак классификатора предметной области) способствуют, в частности, наделению этими классификационными признаками персон. Решение этой задачи только на основании публикаций персоны не всегда возможно, ибо запись о публикации в каталоге системы не всегда может иметь классификационные признаки, тем более относящиеся к основному классификатору справочника (каковым по определению является классификатор, использованный при создании тезауруса).

Поскольку в процессе развития системы она дополняется вспомогательными коллекциями: учебные заведения, научные организации и научные сообщества (академии, редколлегии и т.п.) соответствующего профиля, награды и премии и др., то установление связей между персонами и элементами соответствующих коллекций способно повысить справочную ценность системы. Непременным (но, разумеется, не всегда единственным) атрибутом этих связей являются даты соответствующих событий.

4 Особенности представления информации

Очевидно, что одна из основных возможностей, предоставляемых пользователю при работе со справочником, состоит в просмотре наиболее полной информации о том или ином деятеле науки, включая библиографический список, в том числе полные электронные версии отдельных публикаций.

Однако наличие только перечисленных возможностей не позволяет считать справочник

полноценной информационно-поисковой системой, ибо еще в монографии [5] подчеркнуто, что устройства и машины, предназначенные лишь для отыскания документов по известным адресам их хранения, информационно-поисковыми системами не являются. Поэтому справочник должен быть снабжен развитым пользовательским интерфейсом, позволяющим осуществлять сложные запросы, посредством которых пользователь сможет получить представление о тех или иных аспектах развития рассматриваемой предметной области. Можно выделить следующие основные срезы информации:

- 1) тематический,
- 2) хронологический,
- 3) географический.

Таким образом, пользователь получает возможность изучить хронологию развития выбранного раздела предметной области (или представлений об ее отдельном понятии), в том числе в сужении на отдельную взятую страну, или же составить представление о состоянии соответствующей науки в тот или иной период времени.

5 Заключение

Представленная технология носит достаточно универсальный характер и может быть применена при создании информационно-справочных систем, относящихся практически к любой отрасли естественных наук. В частности, она используется нами в процессе разработки справочника «Математика в личностях» [10].

Литература

- [1] The MacTutor History of Mathematics archive. <http://www-history.mcs.st-and.ac.uk>
- [2] The Mathematics Genealogy Project. <http://www.genealogy.ams.org>
- [3] Математический портал. <http://math.ru/history/people/>
- [4] Баракхнин В.Б., Леонова Ю.В. Информационная модель отношений между документами в информационной системе. В *Вычислительные технологии*, том 10, специальный выпуск, стр. 129-137, 2005.
- [5] Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. М:Наука, 1968.
- [6] Баракхнин В.Б. Разработка тезауруса предметной области «Математика». В *Вычислительные технологии*, том 8, *Региональный вестник Востока*, N 3 (19), совместный выпуск, часть 1, стр. 111-115, 2003.
- [7] Математическая энциклопедия в 5 томах. М.: Советская энциклопедия, 1977-1985.
- [8] Zthes: a Z39.50 Profile for Thesaurus Navigation. <http://lcweb.loc.gov/z3950/agency/profiles/zthes-04.html>
- [9] Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. Новосибирск: Изд-во ИВТ СО РАН, 2004.
- [10] Баракхнин В.Б., Григорьева Я.И. Биографо-библиографический справочник «Математика в личностях». В *Труды международной конференции «Вычислительные и информационные технологии в науке, технике и образовании»*. Казахстан, Павлодар, 20-22 сентября 2006 г., том I, стр. 220-227.

Methodological approach for developing informational-reference systems on history of science

V.B.Barakhnin, A.M.Fedotov

This article describes the methodic of developing informational-reference systems on history of science. The main principles of this methodic are the following:

Information is grouped around persons, at that detailed biographical data is classified by chronological view, geographical view, etc. Bibliographical list of person includes either publication by this scientist or publications about this scientist. A connection between scientific activity of researcher with formal description of object domain where this researcher worked is clearly shown.

This description includes the informational model of the reference book, subsystem's realization features of the informational system, and also main types of user informational requests which are required for full-featured work with the system.

* Работа выполнена при частичной поддержке РФФИ: проекты 06-07-89060, 06-07-89038, 07-07-00271, президентской программы "Ведущие научные школы РФ" (грант № НШ-9886.2006.9) и интеграционных проектов СО РАН.