

## КЛАССИФИКАЦИОННАЯ СРЕДА ДЛЯ СИСТЕМАТИЗАЦИИ И ПОИСКА ИНФОРМАЦИОННЫХ РЕСУРСОВ ПО НАУКЕ И ТЕХНИКЕ

### CLASSIFICATION MEDIUM FOR SYSTEMIZING AND RETRIEVING SCIENTIFIC AND TECHNOLOGICAL INFORMATION RESOURCES

*В. Н. Белоозеров*

*О. А. Антошкова*

*А. В. Шапкин (mailto:viniti.ru)*

*Всероссийский институт научной и технической информации РАН, Москва*

Описаны работы по созданию навигационной системы по информационным ресурсам на основе сопоставления применяемых в них классификаций знания. Создана база рубрик семи классификационных систем. Созданы и внедрены инструментальные средства установления семантических связей рубрик

Наличие единой системы упорядочения и поиска материалов делает распределенное в пространстве множество информационных ресурсов общедоступным фондом знаний. В настоящее время таким общим инструментом для локальных баз данных по науке и технике является Государственный рубрикатор научно-технической информации. Однако он представляет собой достаточно грубую классификацию (3 иерархических уровня), которая позволяет производить систематизацию и поиск только по достаточно широким тематическим классам, что не всегда удовлетворяет пользователей, если их информационная потребность имеет конкретный узконаправленный характер. В библиотечной практике знания систематизированы по международной Универсальной десятичной классификации или по отечественной Библиотечно-библиографической классификации. Эти классификационные инструменты охватывают тематику всего объективного знания. В системе подготовки кадров научных работников идентификация квалификации ведется по другой классификационной системе (Номенклатура специальностей научных работников ВАК).

Объединение этих частей в единую классификационную систему позволит обеспечить более точный поиск информации в ресурсах Государственной системы научно-технической информации и в итоге повысит степень информационной обеспеченности научных работ.

Работы по построению системы классификационных схем проводятся в ВИНТИ с 2000 г. В настоящее время:

- разработана модель представления классификаций в реляционной базе данных (включая предметное описание рубрик на основе нормализованного словаря дескрипторов);
- Создана база данных, включающая полный набор рубрик следующих классификаций:
  - 1) Государственный Рубрикатор НТИ (ГРНТИ),
  - 2) Рубрикатор отраслей знания ВИНТИ,
  - 3) Регистр информационных продуктов ВИНТИ,
  - 4) Номенклатура специальностей научных работников ВАК,
  - 5) Приоритетные направления фундаментальных исследований РАН (ПНФИ),
  - 6) Mathematics Subject Classification (MSC),
  - 7) Универсальная десятичная классификация (УДК).
- отработаны и внедрены инструментальные средства ведения классификационных схем в общей базе данных,
- составлены списки ключевых слов к рубрикам Рубрикатора ВИНТИ, полученные при обработке поисковых образов документов из БД ВИНТИ;
- получен опыт интеллектуального установления семантических связей между рубриками разных классификаций.

Результаты этих работ используются в повседневной практике.

Задача построения навигационной системы по информационным ресурсам в области Науки, техники и образования в качестве одного из важных аспектов включает создание системы взаимосвязанных классификационных схем. Для этого естественно использовать имеющиеся разработки с некоторым смещением акцентов.

Если раньше упор делался на обеспечение производственного процесса ВИНТИ: выпуск РЖ, ведение Рубрикатора и Регистра продуктов, то теперь настало время развить модели и средства, которые были получены в ходе исследовательских проектов по загрузке системы классификаций и по предметному описанию

классификационных рубрик. Главной же особенностью настоящего этапа является перенос усилий на содержательную сторону сопоставления классификационных систем.

Для сопоставления классификаций в единой базе данных разработаны и внедрены новые инструментальные средства установления взаимосвязей на основе аппарата отображения рубрик (MAP-технология). Эти средства основаны на реляционной модели данных и обеспечивают формальную целостность и непротиворечивость взаимных связей. Основу модели составляет теоретико-множественное представление отображения рубрики одной рубрикации на другую рубрикации на базе общего списка квалификаторов, используемых в формулах отображения.

Разработка системы квалификаторов отображения рубрик находится в стадии опытной эксплуатации. Она основана на следующих источниках:

- сопоставление Номенклатуры ВАК с ГРНТИ;
- отображение Рубрикатора информационных изданий ВИНТИ на классификационную схему MSC;
- статистическая обработка баз данных ВИНТИ для выявления связей Номенклатуры ВАК и МПК с Рубрикатором отраслей знания ВИНТИ;

Полная таблица отображения Номенклатуры ВАК на ГРНТИ использует развитую систему квалификаторов, которая может быть принята как рабочий вариант для создания программного обеспечения работы с MAP-объектами. Обозначения и значение квалификаторов показаны в Табл. 1. Близкая система квалификаторов используется при сопоставлении Рубрикатора ВИНТИ с Математической предметной классификацией (MSC), показанная в Табл. 2.

Разработанное программное обеспечение реализовано в клиентском приложении «Программа ведения Рубрикатора. Версия 3.0.2.9», которое установлено на рабочих местах в корпоративной сети ВИНТИ.

В ходе этого исследования были проведены следующие работы:

- проработаны и опробованы на практике разные технологии сопоставления классификаторов: интеллектуальная работа экспертов, загрузка отображений на основе снятия статистики с баз данных и др.;
- составлены таблицы соответствия авторитетных классификационных систем: ГРНТИ → УДК и УДК → ГРНТИ, ГРНТИ → Номенклатура ВАК, и Номенклатура ВАК → ГРНТИ, Рубрикатор ВИНТИ (раздел математики) → MSC, МПК → Рубрикатор ВИНТИ.
- предложено и опробовано несколько систем квалификаторов отображения рубрик независимых классификаций;
- практические результаты работы экспертов и программистов загружены в единую систему классификационных, включая отображение рубрик раздела «Математика» на MSC, отображение рубрик МПК на Рубрикатор ВИНТИ, отображение рубрик ВАК на Рубрикатор ВИНТИ;
- исследованы технологические аспекты решения задачи предметного описания рубрик.

Основываясь на исходной постановке задачи и полученных результатах, планируются следующие направления развития работ:

- разработка полного набора таблиц отображения рубрик всех рассматриваемых классификационных систем в обе стороны;
- включение в систему новых классификационных систем, применяемых в области науки, техники, образования и экономики (Классификатор РФФИ, Библиотечно-библиографическая классификация, Десятичная классификация Дьюи, Общий классификатор видов экономической деятельности и др.);
- развитие модели и формального аппарата построения отображений – с привлечением данных, полученных при включении в систему новых классификаций;
- разработка программно-технологических средств построения предметных описаний рубрик и предметное индексирование классификационных группировок сопоставляемых систем с целью обеспечения предметного поиска в пространстве тематических классификаций;
- разработка в ядре навигационной системы метода ранжирования отображений – с учетом глубинных связей рубрик, предметных описаний, уровней значимости квалификаторов;
- построение тезауруса классификационных рубрик, отражающего онтологию понятий науки, техники и образования, и внедрение его в технологию поиска информации через интернет-портал ВИНТИ.

Знак	Значение квалификатора и примеры
=	<p><b>Полное соответствие.</b>            Когда рубрика полностью соответствует специальности, в правой колонке обычно указывается один код без каких-либо дополнительных символов. Однако из-за эффекта множественной локализации тематики в ГРНТИ может оказаться, что релевантная документация находится также и в других рубриках. Тогда они также указываются в таблице со знаком равенства.</p>
&	<p><b>Частичный охват.</b>            Если рубрика покрывает только часть тематики специальности, то перед ней ставится знак <b>&amp;</b> (<b>логическое И</b>), поскольку обычно в этих случаях указывается две или несколько рубрик, перечисление которых уместно связывать союзом «и». Как правило в правой колонке таблицы указывается перечень рубрик, в совокупности покрывающих тематику специальности.            Пример:            ВАК 01.01.04 Геометрия и топология            соответствует ГРНТИ &amp;27.19 Геометрия, &amp;27.21 Топология            В правой колонке записано: &amp;27.19 &amp;27.21</p>
<	<p><b>Включение.</b>            Если рубрика содержит основной материал специальности, но и большой объем материала, выходящего за рамки специальности, то перед ней ставится знак &lt; (больше). (В данном случае его чтение отличается от арифметической традиции, поскольку широкая часть знака направлена к символу охватываемого множества, которое действительно «больше» сравниваемой специальности).            Пример:            ВАК 02.00.13 Нефтехимия            соответствует ГРНТИ &lt;61.51 Технология переработки нефти и газа</p>
&<	<p><b>Пересечение.</b>            В тех случаях, когда рубрика содержит часть материалов, необходимых для данной специальности, но также и заметное количество заведомо иррелевантных документов, она включается в таблицу с <b>комбинацией знаков &amp;&lt;</b>. Эти знаки предупреждают, что в данной рубрике содержатся только некоторые материалы данной специальности и что есть много посторонних материалов. Эта рубрика указывается в таблице обычно в сопровождении других более релевантных рубрик.            Пример:            ВАК 02.00.01 Неорганическая химия            соответствует ГРНТИ &amp;&lt;61.31 Технология неорганических веществ и продуктов            В таблице указано: =31.17 &amp;&lt;61.31</p>
≈ & ≤	<p><b>Приблизительное совпадение.</b>            В тех случаях, когда нужно отметить, что рассогласование тематики рубрики и специальности не велико, употребляются соответственно следующие знаки: <b>≈, &amp;, ≤</b>.            Пример:            ВАК 01.04.01 Приборы и методы экспериментальной физики            соответствует ГРНТИ ≈29.03 Общие проблемы физического эксперимента</p>
≠ << >>	<p><b>Сильное рассогласование.</b>            В тех случаях, когда не удается найти в ГРНТИ удовлетворительно соответствующую специальности рубрику, указываются рубрики, имеющие хоть что-то общее со специальностью. Их коды сопровождаются знаками <b>≠ (не равно), &lt;&lt; (много больше) и &gt;&gt; (много меньше)</b>.            Пример:            ВАК 03.00.20 Гельминтология            соответствует ГРНТИ &lt;&lt;34.33.15 Зоология беспозвоночных</p>
><	<p><b>Произведение.</b>            Если к специальности относятся только те документы, которые входят сразу в две рубрики, то их коды соединяются <b>комбинацией знаков &gt;&lt;</b>.            Пример:            ВАК 05.08.06 Физические поля корабля, океана, атмосферы и их взаимодействие            соответствует логическому произведению рубрик ГРНТИ: 73.34.37&gt;&lt;30.51.33&gt;&lt;37.25.15, где:            73.34.37 Судовождение            30.51.33 Динамика атмосферы и океана            37.25.15 Общая динамика океана и взаимодействие океана с атмосферой</p>

Таблица 1. Обозначения и значения квалификаторов соответствия классов ГРНТИ и Номенклатуры ВАК

Знак	Значение квалификатора и примеры
=	<p><b>Рубрики ВИНТИ и MSC тождественны</b></p> <p>Пример.  <u>271.33 = 45xx</u>  271.33. Интегральные уравнения  <b>Тождественна</b>  45XX. Integral equations</p>
≈	<p><b>Рубрики ВИНТИ и MSC почти тождественны</b>  (имеется ввиду глубокое пересечение; рубрики пересекаются, но ни одна из них не содержит другую)</p> <p>Пример.  <u>271.31.15 ≈ 35Axx + 35Bxx + 35Cxx+35Dxx</u>  271.31.15. Общая теория дифференциальных уравнений с частными производными  <b>Почти тождественна объединению</b>  35Axx General theory  35Bxx Qualitative properties of solutions  35Cxx Representation of solutions  35Dxx Generalized solutions of partial differential equations</p>
>	<p><b>Рубрика ВИНТИ шире рубрики MSC</b></p> <p>Пример.  <u>271.39 &gt; 46XX + 47XX + 43XX</u>  271.39. Функциональный анализ  <b>По существу шире объединения рубрик MSC</b>  46XX Functional analysis  47XX Operator theory  43XX Abstract harmonic analysis</p>
H	<p><b>Не существует прямого соответствия, но можно отнести к более общей рубрике MSC</b>  (в этом случае дается ссылка на раздел более верхнего уровня или на «None of the above, but in this section»)</p> <p>Пример.  <u>271.39.27.19 H 47Hxx</u>  271.39.27.19. Нелинейные функционалы  <b>прямого соответствия нет, но можно отнести к более общей</b>  47Hxx, Nonlinear operators and their properties</p>
<	<p><b>Рубрика ВИНТИ уже рубрики MSC</b></p> <p>Пример:  <u>271.19.17.19 &lt; 55Pxx + 55Sxx</u>  271.19.17.19 Теория гомотопий: основные задачи  <b>уже, чем объединение рубрик MSC</b>  55Pxx Homotopy theory  55Sxx Operations and obstructions</p>
Π	<p><b>Пересечение рубрик ВИНТИ и MSC</b>  (имеется ввиду пересечение меньшей части объемов рубрик))</p> <p>Пример:  <u>271.41.19.17.27 Π 65N25</u>  271.41.19.17.27 Задачи с собственными параметрами  <b>пересекается с</b>  65N25, Eigenvalue problems</p>

Таблица 2. Обозначения и значения квалификаторов соответствия классов Рубрикатора ВИНТИ и MSC