

О создании системы специализированных рубрикаторов по приоритетным направлениям науки, техники и технологий

*Ефременкова В.М.
(ВИНИТИ РАН)*

*Кириллова О.В., Кузнецов А.Ю., Лебедев В.В.
(НП НЭИКОН)*

Представленные на портале НП НЭИКОН специализированные рубрикаторы, охватывающие тематику пяти приоритетных направлений развития науки и технологий, разработаны по заказу Минобрнауки РФ в рамках ГК № 07.551.11.4002 от 29 июня 2011 г. по теме «Поддержка и расширение системы обеспечения новыми информационными технологиями участников Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 гг.»». Создание рубрикаторов, в первую очередь, преследовало решение прагматических задач, связанных с возможностью их использования для автоматического индексирования больших универсальных информационных массивов и поддержки информационного обеспечения с помощью включенных в рубрикаторы ключевых слов для формирования типовых запросов и осуществления поиска необходимой информации в этих массивах.

Необходимо заметить, что первый опыт создания рубрикатора подобного типа имел место при выполнении проекта по заказу Минобрнауки РФ в рамках ФЦП «Развитие инфраструктуры наноиндустрии в Российской Федерации на 2008-2011 гг.», предусматривавшего разработку автоматизированной информационной системы «Специализированный информационно-библиографический ресурс» (СИБР). Разрабатывавшаяся АИС была предназначена для создания инфраструктуры предоставления информационных услуг участникам наносети (НС) и другим пользователям по навигации и доступу к отечественным и зарубежным ресурсам в области нанотехно-

логий и включала, в т.ч., разработку рубрикатора по нанотехнологиям.

После завершения проекта, СИБР и рубрикатор, предназначенный для индексирования документов, включенных в эту систему, не получили развития. Корректировка и совершенствование Нанорубрикатора и поддержка СИБР осуществлялись уже в рамках проекта «Поддержка и расширение системы обеспечения новыми информационными технологиями...» [1].

Составителями каждого тематического рубрикатора являются специалисты в соответствующих областях знаний, одновременно с этим они же являются информационными специалистами, имеющими большой опыт научно-информационной деятельности, связанной с использованием различных классификационных систем при формировании информационных ресурсов. При решении задач загрузки разработанных рубрикаторов в систему привлекались информационные специалисты в области формирования и использования информационных систем и ресурсов.

Созданию новых рубрикаторов предшествовала работа по анализу возможностей отражения тематики приоритетных направлений средствами таких известных классификационных систем, как УДК, ГРНТИ, ОКВЭД, РФФИ, INSPEC и др. Новая структура рубрикаторов, представляющих иерархическую трех-четырёх уровневую классификационную схему, была разработана на основе изучения информации о современных тенденциях развития каждого приоритетного направления и в соответствии с логикой представления области знания специалистами-разработчиками. Для всех рубрикаторов была предложена унифицированная структура формального представления в едином формате составляющих элементов рубрикаторов, упрощающая загрузку рубрикаторов в систему и их использование.

Для рубрикаторов была предложена единая унифицированная структура, которая включает:

- цифровой индекс — каждый уровень рубрикатора представлен двузначным кодом, разделяемым точкой;
- заглавие рубрики, подрубрики на русском и английском языках;

- код(ы) рубрикатора ГРНТИ, соответствующие тематике рубрики нового рубрикатора;
- две группы ключевых слов на русском и английском языках;
- ключевые слова для каждой рубрики или подразубрики, непосредственно отражающие тематику приоритетного направления и позволяющие причислять индексируемые документы к тематике данного конкретного рубрикатора;
- ключевые слова, отражающие тематику данной конкретной рубрики или подразубрики, раскрывающие содержание описываемых в них объектов, процессов, явлений, способов и методов создания, применения и т.д., но относящиеся, в т.ч., кроме данной тематической области, к другим областям знаний.

Во вторую группу ключевых слов могут быть отнесены также термины, которые в одинаковой степени отражают основную тему рубрикатора, но включены во вторую группу для удобства их объединения с ключевыми словами первой группы с целью обеспечения точности индексирования документов по данной тематике.

Объединение двух групп ключевых слов в поисковые запросы (поисковые предписания) с помощью булевых операторов «AND» и «OR» (редко «NOT») позволяют однозначно приписать индексируемые документы к тематике рубрикатора и представляемого им приоритетного направления. В то же время в рубриках и подразубриках вторая группа ключевых слов может отсутствовать, если первая группа однозначно раскрывает содержание рубрики.

Ключевые слова указанных двух типов используются, в первую очередь, для автоматизированного поиска, отбора и индексирования документов информационного массива по каждой рубрике. Для этой цели к каждой рубрике составляются поисковые предписания с применением булевых операторов, позволяющих использовать и связывать термины рубрикатора, причем в поисковое предписание могут быть включены ключевые слова как из данной, конкретной рубрики, так и из других рубрик. Такая методика позволяет не дублировать ключевые

слова при наполнении рубрикатора, а, в случае необходимости их дублирования, делать на них ссылку в другом разделе.

Разработанная методика и предложенная для построения рубрикаторов унифицированная структура легли в основу АРМ ведения рубрикаторов (<http://rubric.peicon.ru>), позволяющего в едином формате, в сетевом режиме вести рубрикатор — дополнять новыми ключевыми словами, удалять ключевые слова в случае необходимости, вводить рубрики ГРНТИ и т.д. Таким образом, система рубрикаторов не является статичным, законченным ресурсом, а предполагает дальнейшее развитие, ведение, дополнение и корректировку. Возможно также дополнение системы другими рубрикаторами, построенными по такому же принципу, как и представленные.

В настоящее время в системе представлены следующие рубрикаторы:

1. Нанотехнологии и наносистемы;
2. Технологии живых систем;
3. Энергетика и энергосбережение;
4. Рациональное природопользование;
5. Информационно-телекоммуникационные системы.

Логика построения и структура классификационных таблиц новых рубрикаторов позволяет последовательно раскрыть содержание основных понятий каждого из рассматриваемых приоритетных направлений, показать их роль в формировании этих междисциплинарных направлений.

Глубина четырех уровневого рубрикатора определялась необходимостью наиболее полного и точного описания содержания каждой дисциплины и ее спецификой — разные рубрики развиты на разную глубину.

Первый уровень построен по классическому принципу развития классификационных систем, заключающемуся в последовательности развития содержания рассматриваемой дисциплины от общих вопросов к теоретическим (если таковые имеются в данном направлении) и далее (для «Нанорубрикатора» и рубрикатора «Технологии живых систем») — к специальным проблемам, связанным со спецификой изучаемых нано- или биологических объектов, их структурой, свойствами, происходя-

щими в них процессами и явлениями, методами получения, исследования, измерения и сертификации, и с переходом к характерным особенностям нано- или биотехнологий. В рубрикаторе «Технологии живых систем» особое внимание уделено вопросам новых поколений лекарственных средств и разработке биоматериалов. В отличие от «Нанорубрикатора», в рубрикаторе по технологиям живых систем разделы даны не по отраслям клинической медицины (кардиология, хирургия, гастроэнтерология и т.д.), а по болезням, на которые нацелены основные исследования. Кроме того, добавлен еще раздел по бионике (конструирование технических устройств, в том числе и из биомолекул, по принципу структур и процессов живой природы). В обоих рубрикаторах включены разделы по внедрению нанотехнологий или технологий живых систем в промышленности и в других отраслях экономики.

Логика построения и структура классификационной таблицы рубрикатора «Энергетика и энергосбережение» включает традиционные принципы построения классификационных систем, но здесь основаниями деления являются не изучаемые объекты, а специальные вопросы, связанные со спецификой отрасли, основу которой составляют «топливные энергетические ресурсы» традиционных и нетрадиционных источников энергии; далее — к рассмотрению видов энергосистем и способов прямого преобразования энергии; проблем энергосбережения, методов измерения и сертификации, и с переходом к задачам использования энергии и применения в различных отраслях народного хозяйства. В разработанном рубрикаторе три основных ключевых блока: топливные энергетические ресурсы; виды энергосистем; энергосбережение.

Особенностью рубрикатора «Рациональное природопользование» является отражение двух равнозначных тематических направлений: природные ресурсы и рациональное их использование; проблема мониторинга и прогнозирования состояния окружающей среды, ее устойчивого развития, включая вопросы экологической безопасности, предупреждения и ликвидации чрезвычайных ситуаций, влияние состояния окружающей среды на человека.

Общий принцип построения основных разделов рубрикатора «Информационно-коммуникационные системы» — от аппаратного обеспечения к программному, и затем — к алгоритмам и искусственному интеллекту. В каждом из этих больших разделов сначала идут теоретические основы, затем практические реализации и применения.

Более подробно логика построения рубрикаторов приведена в предисловиях к каждому разделу.

Для отражения содержания и использования рубрикаторов в системе автоматического индексирования для каждой рубрики и подрубрики был предложен аппарат ключевых слов на русском и английском языках, расширяющий и уточняющий ее содержание. В перечень ключевых слов были включены все возможные синонимы и антонимы, однокоренные и разнокоренные слова, относящиеся к одному понятию.

При построении структуры рубрикаторов и составления перечня ключевых слов, как на русском, так и на английском языке, принималась во внимание частотность того или иного термина в информационных потоках. Частотность определялась путем анализа результатов последовательных поисков в российских ресурсах и в БД Scopus, в поисковые предписания которых включались анализируемые термины. Такой непосредственный контроль терминов потоками документов по результатам информационных запросов в итоге приводит к постоянному дополнению и модификации рубрикаторов в процессе их эксплуатации.

Кроме того, было проведено сопоставление представленных выше рубрикаторов с рубриками рубрикатора ГРНТИ, УДК и классификатора ОКВЭД. В результате разработана «Таблица соответствия кодов Государственного классификатора научной и технической информации (ГРНТИ) и Общероссийского классификатора видов экономической деятельности (ОКВЭД)».

В настоящее время рубрикаторы загружены в систему и не представлены едиными файловыми информационными блоками. Сложная структура рубрикаторов в части логики использования ключевых слов требует их адаптации для упрощения их понимания пользователями.

Литература

1. А.Ю. Кузнецов, Л.Б. Борисова, В.М. Ефременкова, О.В. Кириллова, Т.А. Пронина. Разработка рубрикатора для «Специализированного информационно-библиографического ресурса» (СИБР) в области нанотехнологий // Российские нанотехнологии, 2011. — т. 6. — № 5-6. — С. 16-22, http://nanorf.ru/events.aspx?cat_id=223&d_no=3424.