

СПИСОК ЛИТЕРАТУРЫ

1. Положение о лингвистическом обеспечении Государственной автоматизированной системы научно-технической информации.— М.: ГКНТ, 1986.
2. Антопольский А. Б. Лингвистическое обеспечение электронных библиотек.— М.: Информрегистр, 2003.— С. 56-66.
3. ГОСТ 7.49-84 СИБИД. Рубрикатор ГАСНТИ. Структура, правила использования и ведения.— М.: Изд-во стандартов, 1984.
4. ГОСТ 7.77-98 СИБИД. Межгосударственный рубрикатор научно-технической информации. Структура, правила использования и ведения.— Минск: Изд-во стандартов, 1998.
5. ISO 15836:2003 Information and documentation — The Dublin Core metadata element set.— Genève, 2003.
6. ГОСТ 7.70-2003 СИБИД. Описание баз данных и машиночитаемых информационных массивов. Состав и обозначение характеристик.— М.: Изд-во стандартов, 2003.
7. ГОСТ ИСО 8601-2001 СИБИД. Представление дат и времени дня. Общие требования.— Минск: Изд-во стандартов, 2001.
8. ГОСТ 7.66-92 СИБИД. Индексирование документов. Общие требования к координатному индексированию.— М.: Изд-во стандартов, 1992.
9. ГОСТ 7.9-95 (ИСО 214-76) СИБИД. Реферат и аннотация. Общие требования.— М.: Изд-во стандартов, 1996.
10. ISO 214:1976 Documentation — Abstracts for publications and documentation.— Genève, 1976.

Материал поступил в редакцию 05.01.04.

УДК 502/504:004.738.5

Н. В. Аввакумова, Т. Б. Плескачева, И. И. Потапов,
Н. М. Чернавская

Информационно-поисковая система “Экологическая информация” на сайте Интернета*

Рассматриваются вопросы поиска информации узко-специализированной направленности, решение которых предлагается в создании информационно-поисковой системы “Экологическая информация”, включающей рефераты статей, монографий, материалов конференций по ряду актуальных проблем охраны окружающей среды. Представленная в системе база данных может служить первичным источником для специалистов, работающих в области эколого-экономических проблем. Встроенный в систему механизм поиска обеспечивает быстрый доступ к необходимой пользователю информации.

Проблемы охраны окружающей среды (ОС) из года в год становятся все более актуальными. Даже при катастрофическом спаде промышленного производства, который наблюдается на территориях бывшего СССР, ситуация с техногенным загрязнением не улучшается, а становится все более угрожающей.

В ряде случаев это объясняется отсутствием экологических программ, принятие которых тормозится как на уровне субъектов хозяйствования — промышленных предприятий и организаций из-за недостаточного финансирования, выделяемого на эти цели, так и по причинам объективного характера, связанным со сложностью расчетов эколого-экономической оптимизации деятельности промышленного предприятия и невозможностью прогнозирования экологических последствия такой деятельности.

В этой ситуации исследовательские коллективы экологов и экономистов должны располагать необходимым объемом справочно-статистической и научной информации.

Недостаток специализированных периодических эколого-экономических изданий, несомненно, затрудняет получение такой информации в

нужном объеме и в достаточно короткие сроки. Компенсация этого недостатка осуществляется с помощью баз данных (отечественных и в большей степени зарубежных), содержащих большой объем разнородного материала. Пользователь вынужден искать необходимый ему материал среди огромного количества информации, в основном не относящейся к сфере его профессиональных интересов. Для получения информации по конкретной, достаточно узкой тематике необходимо составление запросов на профессиональном уровне, что не всегда доступно обыкновенному пользователю — сотрудникам как научных учреждений (университетов, научно-исследовательских институтов и др.), так и общественных организаций (комитетов, советов и др.).

При попытке использования имеющихся баз данных возникает ряд трудностей не только технического, но и семантического характера, т. е. при объединении данных из разных источников необходимо идентифицировать используемую терминологию, поскольку одни и те же понятия могут быть представлены разными терминами и, наоборот, разные понятия имеют одни и те же имена.

*Работа выполнена при поддержке РФФИ (грант № 02-06-80190).

При отсутствии унифицированной терминологии в поисковую систему должна быть заложена достаточно большая семантическая и терминологическая информация о значении каждого ключевого понятия для выявления терминологического несоответствия, что необходимо и для правильной интерпретации запросов.

Значительно больший интерес для пользователей представляют тематические базы данных, которые прошли предварительную обработку специалистами в выбранной области знания. Обработка включает единообразное описание предметных ключевых понятий, отражающих основные аспекты содержания публикации, составление узкоспециального тематического рубрикатора и рубрицирование в соответствии с ним содержимого базы данных. Проводимая подготовка материала на этапе ввода данных должна значительно облегчать пользователю тематический поиск нужной информации.

В настоящее время в связи с широким использованием глобальной информационной сети большое значение приобретает применение новых технологий для представления информационного материала в Интернет.

НИВЦ МГУ совместно с сотрудниками ТвГУ создана Информационно-поисковая система "Экологическая Информация" (ИПСЭИ-1.01), включающая информацию, необходимую специалистам, работающим в эколого-экономической области, для решения задач, увязывающих экологические и экономические характеристики с целью оптимизации дальнейшего развития региона без ущерба окружающей среде и здоровью населения. Главными направлениями решения двуединой задачи — обеспечение экономического развития региона с сохранением экологического благополучия территории — является использование различного рода моделей и данных биомониторинга, описывающих процессы, происходящие в окружающей среде под влиянием техногенного воздействия.

Предлагаемая нами информация на сайте Интернет представлена в виде аннотаций и рефератов монографий, заявок, статей периодических изданий (российских и зарубежных) и докладов на международных и всероссийских конференциях по следующим темам:

1. Эколого-экономическая деятельность предприятий.

2. Моделирование экосистем, экологических процессов, технологических процессов.

3. Биологические методы в экологии.

Предполагается введение новых информационных тем, в частности, мы планируем с 2004 г. открыть рубрику "Природные и техногенные катастрофы и чрезвычайные ситуации".

База данных включает материалы, опубликованные в открытой печати в последние годы (2000–2002 гг.), и ежеквартально пополняется текущей информацией.

Информационно-поисковая система "Экологическая Информация" (ИПСЭИ-1.01) представлена на сайте Ecoinformatika.narod.ru. Рост числа обращений к сайту показывает актуальность выбранных для экспозиции тем, и побуждает искать новые, более эффективные решения для представленной аналогичной информации в сети Интернет.

При разработке ИПСЭИ-1.01 использовалось приложение Microsoft FrontPage2002, входящее в

пакет Office XP и представляющее собой надежное функциональное средство создания сайтов Интернет и управления ими.

Особое внимание было уделено разработке системы поиска необходимой пользователю информации. В настоящее время поиск по сайту может осуществляться по следующим критериям: тема, ключевое слово, авторы (русские, зарубежные). Словарь ключевых слов был составлен на основе семантического и синтаксического анализа научной информации, входящей в рефераты. В результате этой работы были отобраны наиболее часто встречающиеся понятия и слова в материалах эколого-экономической направленности. При введении в базу нового объема данных существующий словарь постоянно анализируется и пополняется.

Кроме вышеперечисленного, за счет того, что сайт располагается на сервере Narod.ru, при поиске нужной информации можно также использовать ряд возможностей поисковой системы Яндекс, к некоторым из них относятся следующие:

1) независимо от того, в какой форме вы употребили слово в запросе, поиск учитывает все его формы по правилам русского языка;

2) существует возможность поиска по точной словоформе;

3) можно задать или исключить слова или комбинацию слов из запроса, которые обязательно должны быть найдены или соответственно исключены из поиска;

4) можно найти документы, содержащие любое из выбранных вами слов, что удобно при поиске одних и тех же понятий, которые могут иметь различные определения;

5) можно задать расстояние между соседними словами, то же самое относится и к предложениям.

При поиске для каждого найденного документа Яндекс вычисляет величину релевантности содержания этого документа поисковому запросу. Список найденных документов перед выдачей пользователю сортируется по этой величине в порядке убывания. Релевантность документа зависит от ряда факторов, в том числе от частотных характеристик искомых слов, веса слова или выражения, близости искомых слов друг к другу в тексте документа и т. д. Пользователь может повлиять на порядок сортировки, используя операторы веса и уточнения запроса.

Существует также возможность задать поиск не только по сайту Ecoinformatika.narod.ru, но и в самой базе "Яндекс", которая насчитывает более 20 000 различных документов.

В настоящее время функционирует ИПСЭИ-1.01 (1-я версия, 1-е издание ИПСЭИ). Каждое следующее издание будет периодически, приблизительно раз в квартал, пополняться новыми данными. Для новой версии системы будут использованы, в сочетании с традиционной HTML-страницей, прогрессивные технологии, такие, например, как расширяемый язык разметки XML, что сможет обеспечить идеальное решение при обработке все более расширяющегося потока и одновременно увеличивающейся сложности эколого-экономической информации, размещаемой в глобальной сети Интернет.

Материал поступил в редакцию 15.12.03.