



- ленных систем обработки данных: Дис. д-ра техн. наук. МГТУ им. Н.Э. Баумана, 17.10.96 г.
4. An Environment for the Design and Performance Evaluation of Portable Parallel Software: Final EDPEPPS Simulator: Report / Center for Parallel Computing University of Westminster. EDPEPPS/41. London, July 1997.
  5. *Емельянов С.В., Ларичев О.И.* Многокритериальные методы принятия решений. М.: Знание, 1985. (Новое в жизни, науке, технике. Сер. Математика, кибернетика; № 10).
  6. *Трахтенгерц Э.А.* Компьютерная поддержка принятия решений. М.: СИНТЕГ, 1998.
  7. *Липаев В.В.* Системное проектирование сложных программных средств для информационных систем. М.: СИНТЕГ, 1999.
  8. Приоритетные системы обслуживания / *Гнеденко Б.В., Даниелян Э.А., Димитров Б.Н и др.* М.: Изд-во МГУ, 1973.
  9. *Тельнов Ю.Ф.* Интеллектуальные информационные системы в экономике. М.: СИНТЕГ, 1998.
  10. *Григорьев Ю.А., Плутенко А.Д.* Жизненный цикл проектирования распределенных баз данных. Благовещенск: Изд-во Амурского гос. ун-та, 1999.
  11. *Кузнецов С.Д.* Стандарты языка реляционных баз данных SQL: краткий обзор // Системы управления базами данных (М.). 1996. № 2.
  12. *Жигалов В.* Как нам обустроить поиск в сети? // Открытые системы (М.). 2000. № 12.
  13. *Штоян Д.* Качественные свойства и оценки стохастических моделей. М.: Мир, 1979.
  14. *Герман О.В.* Введение в теорию экспертных систем и обработку знаний. Минск.: ДизайнПРО, 1995.

УДК 025.4.03

© 2001 г. **Е.А. Нурминский**, д-р физ.-мат. наук  
(Институт автоматизи и процессов управления ДВО РАН,  
Владивосток)

## **2KWIC-СИСТЕМА ГЕНЕРАЦИИ ЭЛЕКТРОННЫХ KWIC-КАТАЛОГОВ**

В статье описывается программная система, позволяющая из файлов стандартных MARC-записей генерировать электронный каталог.

### **Введение**

Поиск необходимой информации по ключевым словам в библиографических базах данных значительно облегчается, если эти слова демонст-

рируются "в контексте", подсказывающем действительное значение термина.

Здесь описана программная система **2kwic**, с помощью которой из файлов стандартных MARC-записей генерируется электронный каталог, в котором:

- ключевые слова определяются автоматически,
- поиск по ключевым словам осуществляется в контексте.

Полезной особенностью системы является реализация удаленного WWW-доступа к каталогизируемым материалам. Система создает собственную базу данных на сервере каталога и отвечает на запросы удаленных Интернет-пользователей, используя эту базу данных и двухуровневую систему меню.

### Функциональные возможные системы

Поиск необходимой информации по ключевым словам в библиографических базах данных значительно облегчается, если эти слова демонстрируются "в контексте", демонстрирующем действительное значение термина. Именно на этом и основана идея KWIC ( Key Words In Context ) каталогов – весьма эффективных средств индексации небольших и средних коллекций. Точнее, в этих каталогах использованы две основные идеи:

- автоматической генерации списка ключевых слов;
- индексации источников по ключевым словам и демонстрация их "в контексте".

В этих каталогах информация, в какой-то степени отображающая содержание книги, статьи, отчета и пр., разбивается на слова, которые по определению считаются ключевыми. Чаще всего для этого используется наименование книги, например, книга Жюль Верна "История великих путешествий" дает три ключевых слова: *История, великих и путешествий*.

Второй этап – индексация – заключается в том, что источник (в нашем примере книга Жюль Верна) заносится в KWIC-каталог под этими ключевыми словами, причем они демонстрируются в контексте:

История	<i>великих</i>	путешествий
	<i>История</i>	великих путешествий
История великих	<i>путешествий</i>	

что позволяет, например, легко отличить эту книгу от "Истории города Глупова" М.Е. Салтыкова-Щедрина или "Истории Тома Джонса-найде-ныша" Филдинга, а также от сонма других *Историй*, которые обрушатся на вас при поиске по этому ключевому слову.

Сам KWIC-каталог представляет собой гипертекстовый документ, просматриваемый пользователями удаленно через Интер- или Интранет.

Современные информационные технологии позволяют при этом легко

получить и полные библиографические данные издания. С каждой записью в KWIC-каталоге связывается гиперссылка, которая при ее активизации извлекает всю информацию, имеющуюся по данному изданию в MARC-записи библиографической базы данных.

Эта информация в текстовом виде поступает на компьютер пользователя или библиотечного работника и позволяет сделать окончательный вывод о соответствии издания запросу пользователя, определить местонахождение источника и т.п. Ниже приведен пример вывода на компьютер пользователя при активизации одной из ссылок в KWIC-каталоге Центральной научной библиотеки ДВО РАН, содержащей ключевое слово *История*.

При этом часть KWIC-каталога выглядела следующим образом:

*История Древнего мира Античность*

*История культуры Китая*

*История биологии с древнейших времен*

*История и культура чукчей*

*История математики без границ*

*История русской духовной культуры*

После активизации ссылки пользователь видит на своем экране следующую информацию:

История и культура чукчей: Ист.-этногр. очерки / АН СССР.

Ин-т истории, археологии и этнографии народов Дальнего Востока ДВНЦ, Ин-т этнографии им. Н. Н. Миклухо-Маклая ; Под общ. ред. Крушанова А. И.

Л. : Наука. Ленингр. отделение, 1987

286,1с.

40.00р.

Библиогр. в конце ст.

Этнография Чукотки

оленоводство юкагиры

902.7(57) / И 90

902.7(57)

03.61.91 rugasnti

Институт истории, археологии и этнографии народов Дальнего Востока(Владивосток)

Состав выводимой информации определяется заполнением MARC-записей и может быть весьма обширным.

Кстати, эта же самая MARC-запись была включена в KWIC-каталог и в другом месте:

...

Материальная

Древняя

культура Дагестана албанского времени

культура Ширака

Коряки Материальная	культура и социальная организация
<i>История и</i>	<i>культура чукчей</i>
Язык миф	культура Символы жизни и жизнь символа
История и	культура народов Дальнего Востока Росс
...	

что увеличивает шансы нахождения необходимой информации.

### Использование системы

Система реализована по WWW-технологии и содержит информационные массивы под управлением распространенного web-сервера (Apache).

В принципе используются лишь базовые функции сервера и не должно возникать проблем с переносом системы под управление какого-либо иного Интернет-сервера. Для размещения каталога выделяется соответствующий объем дисковой памяти сервера и организуется гиперссылка из корневой страницы сервера. Клиенты получают доступ к информации по http-протоколу, используя стандартные браузеры типа Netscape Navigator, Internet EXplorer и др. Ответы на их запросы поступают в виде динамически генерируемых HTML-страниц.

Исходным материалом для построения KWIC-каталогов являются библиографические базы данных в MARC-формате (USMARC).

При работе с MARC-форматом мы ориентировались на описание, подготовленное группой специалистов ГПНТБ России. Краткая версия этого описание приведена, например, в [1].

Хотя существуют различные версии этого формата (USMARC, UNIMARC, RUSMARC и пр.), а практика заполнения полей меняется в зависимости от библиотеки или программной системы, различия между форматами и базами данных, как показал наш опыт, в рамках системы **2kwic** достаточно легко преодолимы.

Такой формат библиографических данных, как правило, может быть получен из любой современной библиотечной системы в виде экспортируемых файлов. В настоящее время на сервере Дальневосточного отделения РАН ([www.dvo.ru](http://www.dvo.ru)) в качестве эксперимента подготовлены и размещены 2 KWIC-каталога: Центральной научной библиотеки (ЦНБ) ДВО РАН и Научно-технической библиотеки (НТБ) ОАО «Дальзавод».

Этот сервер входит в базовую российскую компьютерную сеть научных и образовательных организаций, доступен из любой точки российского и мирового Интернета, имеет хорошую связность с основными провайдерами сетевых информационных услуг во Владивостоке.

Каталог ЦНБ содержит лишь новые поступления в библиотеку приблизительно за последние 3 года, каталог НТБ интересен тем, что содержит постатейную роспись ряда журналов, пользующихся достаточной популярностью.

В таблице приведены сводные характеристики обоих каталогов по состоянию на май 2000 г.

Характеристика	ЦНБ ДВО РАН	НТБ ОАО «Дальзавод»
Количество MARC-записей	6908	9579
Объем исходной информации (байт)	5140480	2900265
Количество KWIC-записей	30778	46953
Объем базы данных (байт)	4768624	3027399
Объем гипертекста (байт)	5442707	8181760
Время генерации	6'11"	5'54"

Времена генерации KWIC-каталогов даны для средней производительности персонального компьютера (433 Mhz Celeron, 64 Mb оперативной памяти).

#### ЛИТЕРАТУРА

1. *Основные положения формата MARC для библиографических данных/* Пер. с англ. и науч. ред. С.А. Беляниной, И.В. Мельник, А.Н. Пака, Р.А. Кругловой; под общ. ред. д. чл. ПК по UNIMARC Я.Л. Штернберга. М.: ГПНТБ России, 1997.