

НОВОЕ В ТЕХНОЛОГИИ БИБЛИОТЕЧНО-ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

УДК 681.3-181.4(083.4)

Розеншильд Л.К., Мильграм В.Д., Смирнов И.Ф.

Составление рабочих таблиц УДК на персональном компьютере

Рассказано о методике машинного составления таблиц классификации и системе их записи в виде баз данных.

Методика составления рабочих (проблемно-ориентированных) таблиц классификации (РТК) достаточно известна [1—3]. Те, кто занимался этой работой, хорошо знают ее объем и трудоемкость: критическая оценка существующего текста, совмещение логики отрасли знания с логикой системы классификации, отбор лексики по теме, построение индексов для сложных понятий средствами грамматики используемого языка, составление ссылок и отсылок, методические примечаний и алфавитно-предметного указателя (АПУ).

Сложность такого процесса побудила нас к разработке методики машинного составления рабочих таблиц классификации с целью упрощения составления АПУ.

Ст. научный сотрудник Библиотеки по естественным наукам Российской академии наук И.Ф. Смирнов выбрал для этой цели персональные компьютеры типа РС/АТ/ХТ стандартной конфигурации с жестким диском типа "винчестер" от 20 МБ и более и разработал АРМ. В.Д. Мильграм, директор Биоинформсервиса, разработал автоматизированную систему составления, ввода и вывода на печать рабочих таблиц или любых их фрагментов в режиме баз данных, создания текстовых файлов, включая текст таблиц и АПУ.

Благодаря системе записи таблиц в виде баз данных можно осуществлять автоматизированную систематизацию и научное редак-

тирование систематического каталога с дальнейшей переиндексацией (при необходимости) информационных массивов.

При подготовке к работе с документами рабочих таблиц УДК составителю на экране предлагается следующий шаблон:

_____ Система введения УДК _____
_____ Ввод
_____ Поиск документов УДК
_____ Импорт текстового файла
_____ Завершение работы с УДК

Используя различные позиции этого шаблона, можно выполнять предложенные выше операции. При этом в режиме ввода документов рабочих таблиц УДК шаблон принимает такой вид:

_____ Документ рабочей таблицы УДК _____
№ документа
Индекс УДК
Значение (запись класса)
Дополнительная информация (методические примечания)
Ссылки
Ответственный (составитель)

Шаблон ввода снабжен подсказками, определяющими единственную операцию — ввод документов: ввод, запоминание и выход из шаблона. В этом режиме заполняются поля всех перечисленных позиций, после чего информация вводится в память машины.

По завершении ввода переходят в режим поиска документов, цель которого — дополнение введенной информации, редактирование документов таблиц, систематизация и (или) предметизация первичной или вторичной информации, редактирование или ресистематизация систематического каталога или тематических картотек.

При этом шаблон для поиска и коррекции документов УДК, появляющийся на экране, выглядит следующим образом:

_____ Поиск значений УДК-схемы _____
№ документа
Индекс
Значение
Дополнительная информация
Ответственный

Как видно, этот шаблон предоставляет места для индекса, записи формулировки рубрики, методических указаний и ссылок. В нем содержатся также сведения о порядковом номере и составителе.

Подсказки к этому шаблону таковы:

поиск,
перечень всех документов,
перечень найденных по теме документов,
первый найденный документ,
выход из режима поиска.

Пользуясь этим шаблоном, можно осуществить поиск по любому заданному признаку:

по порядковому номеру документа,
по индексу УДК или начальному фрагменту индекса,
по одному или двум ключевым словам или их фрагментам.

Поиск можно уточнять, варьируя длину фрагмента ключевого слова, индекса и т. п. Это дает возможность получения релевантных ответов (на один запрос машина выдает все имеющиеся документы в различных аспектах темы):

число документов в таблицах, число найденных документов;
список всех документов, имеющихся в системе, как в режиме баз данных, так и в режиме списка;

число всех документов, содержащих ключевое слово или его фрагмент, отвечающих заданной теме (или заданному фрагменту класса);

список всех документов на заданную тему (или заданного класса) с представлением его в режиме списка или базы данных (т. е. набора документов, включающего первый и последующие документы списка).

Элементы представляемого машиной единичного документа УДК имеют такой же вид, как и шаблон поиска. Его особенностью является информация об объеме открываемого им массива (1 из 1, или 1 из 285).

Кроме того, этот шаблон дает возможность "перелистать" нужный информационный массив, так как среди команд подсказки есть команды "предыдущий документ", "последующий документ".

Найденные документы УДК можно редактировать, заполняя незаполненные поля, дополняя или изменяя информацию.

Как уже упоминалось, манипулируя поисковыми приемами, можно непосредственно с экрана вести систематизацию и тематическое

(научное) редактирование систематического каталога или любого другого индексированного массива информации в пределах тематики таблиц, а также дополнять таблицы новыми понятиями за счет систематизируемых документов.

При корректировке или дополнении РТК составителю предоставляется шаблон другого формата, позволяющий:

найти нужную карточку по любому из необходимых признаков или их комбинации;

отсортировать по рабочему номеру или тематическим признакам любые фрагменты РТК;

объединять РТК или их фрагменты по различным отраслям знаний, дисциплинам.

Индексы записываются в соответствии с их расположением в традиционных таблицах, каталогах или СИФах. Сложность индекса определяется необходимостью однозначной передачи с его помощью содержания понятий, включаемых в рубрики.

Запись класса ("значение", "формулировка рубрики") составляется в соответствии с текстом "Полных таблиц УДК", дополненных однородной или близкой современной лексикой. Запись ведется с соблюдением некоторых условностей, в дальнейшем необходимых для составления АПУ, а также машинного поиска понятий.

Технология машинного составления таблиц требует максимально сокращенной записи и минимального количества знаков препинания. В этом смысле при машинной записи таблиц приходится нарушать некоторые из привычных правил составления РТК (применение косых скобок и т. п.).

Параллельно с составлением РТК машина составляет и перечни предметных рубрик (ключевых слов), служащие основой составления АПУ (при последующей доработке программы предусматривается полная автоматизация этого процесса). Из перечней ключевых слов с помощью специальной программы удаляются все незначимые слова и знаки препинания. Чистота перечней предметных рубрик обеспечивается системой стоп-слов, включающих все знаки препинания, служебные и неинформативные слова, которые не имеют поисковой ценности. Составляется и редактируется общий перечень ключевых слов, дописываются индексы. Все понятия, включенные в поле "значение", автоматически включаются в АПУ.

Адекватное составление перечней предметных рубрик производится путем соединения слов, входящих в понятие, нижней соедини-

тельной чертой, благодаря чему понятие читается машиной как единое слово. Для обеспечения более полного представления понятий в АПУ формулировка записи класса выполняется, начиная с ключевого слова. Это делает необходимой запись понятий, включенных в таблицы, как в прямой, так и в инвертированной форме.

Поиск дает возможность найти документ или перечень документов по одному или нескольким признакам: порядковому номеру, индексу или его фрагменту, ключевому слову или его фрагменту, а также по ссылкам. Можно получить перечень документов, содержащих определенные ключевые слова, что очень удобно при проверке отражения понятий, а также при составлении ссылок и редактировании систематического каталога.

Особое значение в этой работе мы придаем однозначности формулировок записей класса, ссылочному аппарату и методическим примечаниям, оговаривающим сферу применения класса и возможность его дальнейшей детализации.

Распространено мнение, что при хорошем АПУ, включающем исчерпывающие гнездовые рубрики, ссылки не нужны. Возможно, это в какой-то мере оправдывает себя в АПУ к полным таблицам, однако при составлении РТК отсутствие ссылок может поставить работающего систематизатора перед сложным поиском дополнительных "аспектов", которые необходимо отразить в полном индексе.

Примером отбора поисковых признаков при составлении индексационного описания документов послужил поиск альтернативных признаков, используемый при определении живых организмов, минералов и прочих объектов природы по принятым для этого справочным пособиям — определителям, дающим последовательное описание признаков изучаемого объекта наряду с исключением тех признаков, которых объект не имеет.

Альтернативность признаков благодаря уточнениям, содержащимся в формулировках, исключает или сводит к минимуму возможность появления ошибок, например:

"Стебель опущенный" — "Стебель гладкий",

"Прикорневые листья собраны в розетку" — "Прикорневая розетка отсутствует" и т. п.

Находя нужные и исключая ненужные признаки, исследователь получает единственно правильное решение.

При составлении рабочих таблиц также необходимо ориентироваться на предоставление пользователю максимальной возможности

выбора, которая во многом определяется однозначностью формулировки рубрики, что далеко не всегда имеется в исходных таблицах. Не всегда предоставляется она и гнездовыми рубриками АПУ к таблицам, поскольку последние, как правило, объединяют различные аспекты одного и того же понятия, тогда как сами понятия, описывающие какой-либо предмет или явление, могут иметь различное положение в алфавитном ряду, например:

гибридомы — моноклональные антитела — терапия рака.

Цель предоставления пользователю максимума альтернатив достигается составлением исчерпывающих перечней ссылок.

Наличие исчерпывающего перечня ссылок дает исключительно полный и синхронно предоставляемый выбор аспектов систематизации документов. Совокупность неформально составленных ссылок (альтернатив) придает системность не только РТК, но и составленным на их основе систематическим каталогам и картотекам.

Программа машинного составления РТК была опробована при составлении проблемно-ориентированных таблиц по иммунологии, физиологии нервной системы, природоведению и охране природы, вычислительной технике, водоподготовке и водоснабжению и др.

В результате внедрения метода и анализа проблемно-ориентированных таблиц мы пришли к выводу об их наибольшей эффективности в случае комплексности темы, ее построения на различных классах УДК или в случае сложной, неоднозначной тематики с развивающимся лексическим составом и обширной синонимией.

Именно поэтому в дальнейшем мы планируем продолжение и развитие разработок РТК по принципу баз данных для ряда тем, в том числе по Полному варианту класса 51 Математика, ряду тем физиологии (Иммунология, Частная физиология систем, органов, тканей, Нервная система), общей биологии (Биотехнология, Радиобиология, Биохимия) и некоторых других.

СПИСОК ЛИТЕРАТУРЫ

1. Павловская Н.С. Рабочие схемы классификации и отраслевые таблицы УДК//Науч. и техн. 6-ки СССР. — 1975. — Вып.4 (136). — С. 30—32.

2. Порядок, правила и методы составления рабочих таблиц классификации по УДК: Метод. указания//ГПНТБ СССР; Сост. С.Г. Дуплева, Л.В. Лобова, О.Л. Орлинкова — М., 1984. — С. 1—16.

3. Составление рабочих таблиц классификации (РТК) на основе УДК: Инструкция//Регламентирующие документы по работе с систематическим каталогом: Положения и инструкции//Б-ка по естеств. наукам АН СССР.— М., 1986. — С. 65—72.