

Костенко Л.И.

*ЦНБ им. В.И. Вернадского Национальной
АН Украины*

Автоматизированная система оперативной подготовки обзоров прессы

Представлена система, обеспечивающая оперативную аналитико-синтетическую переработку входного документального потока библиотеки, последующую кумуляцию полученной информации в виде обзорно-аналитических материалов и их передачу соответствующим службам органов государственной власти и управления по телекоммуникационным каналам связи.

В настоящее время возрастает актуальность проблемы научно-информационного обеспечения органов законодательной, исполнительной и судебной власти. Библиотекам как старейшим социальным институтам, обеспечивающим кумуляцию, обработку, хранение и использование документированных знаний, надлежит сыграть заметную роль в ее решении.

Для организации научно-информационного обеспечения управленческих структур в Центральной научной библиотеке им. В.И. Вернадского Национальной АН Украины создана и развивается автоматизированная система текущего информирования. Эта система обеспечивает оперативную аналитико-синтетическую переработку входного документального потока библиотеки, последующую кумуляцию полученной информации в виде обзорно-аналитических материалов, и их передачу соответствующим службам органов государственной власти и управления по телекоммуникационным каналам связи. Информационные ресурсы, накапливаемые в системе в процессе ее функционирования, используются и в библиотеке. Они содействуют удовлетворению потребности читателей и абонентов в аналитической информации.

Источниковой базой системы является входной документальный поток ЦНБ, составляющий более 150 тыс. документов, в числе которых 100 тыс. книг, 50 тыс. номеров журналов и 2 тыс. комплектов газет. Тематика новых поступлений в ЦНБ универсальна. Библиотека достаточно полно комплектуется изданиями Украины и получает до 35 тыс. документов из стран

ближнего и дальнего зарубежья. Кроме того, ЦНБ как депозитарий ООН на Украине получает все публикации этой организации.

Технологический цикл текущего информирования управленческих структур включает следующие процессы:

выявление среди новых поступлений документов, содержание которых может содействовать оптимизации управленческих решений (такowymi являются материалы по вопросам национальной безопасности, внутренней и внешней политики, экономики, экологии, демографии и статистики);

интеллектуальная обработка документов и подготовка их рефератов с необходимыми сопровождающими данными для обеспечения реализации современной информационной технологии обзорно-аналитической деятельности библиотеки;

ввод текстов рефератов, библиографических описаний первоисточников и набора поисковых терминов в базы данных с последующим проведением формально-логического и семантического контроля введенной информации;

формирование с заданной периодичностью обзорно-аналитических материалов (упорядоченных текстовых файлов с ежедневными обзорами прессы; ежемесячные электронные реферативные журналы, составляемые на основе обработки журналов и книг);

передача подготовленных материалов, а также пополнений баз данных по телекоммуникационным каналам связи внешним абонентам, в числе которых Администрация Президента Украины и Верховный Совет Украины.

По изложенной информационной технологии ЦНБ, начиная с четвертого квартала 1992 г., поставляет управленческим структурам ежедневные обзоры прессы. Необходимо подчеркнуть, что разработка и создание автоматизированной системы оперативной подготовки этих обзоров были проведены в течение одного месяца. Формирование текущего обзора прессы осуществляется в условиях жестких временных ограничений, обусловленных необходимостью его получения к 16 часам. В хронологическом плане технологический цикл подготовки этого обзора протекает следующим образом. После получения утренней почты сектор аналитико-синтетической переработки газетных публикаций ЦНБ (в его составе 10 специалистов) приступает к выявлению материалов для текущего информирования, их реферированию, индексированию и вводу в базы данных на четырех ПЭВМ IBM PC. В 14.30 час. ввод данных прекращается и проводится редактирование информации. В

15.30 час. информация текущего дня, введенная в разные ПЭВМ, объединяется и информационно-поисковая система автоматически формирует итоговый текстовый файл, упорядоченный по тематическому и географическому признакам, а также по алфавиту названий газетных статей. До передачи файла управленческим структурам он редактируется средствами текстового редактора "Lexicon" (выделяются заголовки, подчеркиваются наиболее важные разделы и т. п.). К 16 часам текущий обзор прессы объемом 14—18 страниц поступает на телекоммуникационный узел ЦНБ и с этого времени становится доступным ее удаленным абонентам.

Некоторые абоненты получают информацию самостоятельно, другим ЦНБ пересылает текстовые файлы с обзорами прессы по электронной почте, отдельным абонентам (например, Администрации Президента Украины) передает непосредственно на их телекоммуникационные узлы.

Кроме центральных органов государственной власти и управления обзоры прессы поступают и в ряд областных администраций. При этом посредниками выступают областные научные библиотеки, которые проводят работу по их информационно-аналитическому обеспечению. Вначале эти библиотеки были только пассивными потребителями информации, однако затем они стали партнерами ЦНБ в подготовке обзоров. Такое сотрудничество существенно повысило информативность обзоров, так как ранее они отражали весьма ограниченный круг сведений регионального характера (из-за значительных задержек в доставке газет из областей в Киев и старения содержащейся в них информации, ее включение в ежедневные реферативные обзоры прессы представлялось нецелесообразным).

Исходные материалы из областей поступают в ЦНБ по телекоммуникационным каналам связи в виде текстовых файлов, которые вводятся в базу данных с рефератами газетных статей с помощью конвертора Fangorn, обеспечивающего преобразование "текстовый файл — формат ISO-2709" (указанный формат рекомендован Международной организацией по стандартизации для использования в качестве коммуникативного формата при обмене библиографической информацией на магнитных носителях. Он поддерживается всеми информационно-поисковыми системами, при разработке которых соблюдались международные стандарты и рекомендации).

После передачи текстового файла с текущим обзором прессы на телекоммуникационный узел ЦНБ работа с записями базы дан-

ных продолжается. Они кумулируются в еженедельном массиве, который затем вторично просматривается. При этом удаляется часть записей, информация в которых дублируется. Если, например, на Украине с визитом находился государственный деятель другой страны и в нескольких ежедневных обзорах отражались его переговоры с должностными лицами Украины, то в базе данных оставляется только итоговая запись о состоявшемся визите. Затем уточняются классификационные индексы и предметные рубрики, которые предназначены для тематического поиска информации. После этого еженедельный массив подвергается процедуре автоматической проверки корректности данных (в системе предусмотрен контроль на обязательное наличие информации в полях, содержащих название статьи и газеты, сверка записей на дублетность, а также проверка поисковых терминов по словарю (инвертированному файлу) основной базы данных (основной считается та, в которой кумулируется итоговая информация).

В начале 1995 г. основная база данных с рефератами газетных статей насчитывала более 10 тыс. записей. Ее информационные ресурсы доступны как управленческим структурам, так и читателям библиотеки. Для обеспечения поиска в этой БД разработано специальное лингвистическое обеспечение, которое основывается на фасетных классификациях. Предусмотрены хронологический, географический и тематический фасеты, а также фасет персоналий, разделенный на фасет лиц (государственных и политических деятелей, предпринимателей и т. д.) и фасет организаций (высших органов власти, министерств, политических партий и т. д.). Хронологический фасет ведется с точностью до месяца, географический — включает названия стран и городов.

Для реализации тематического фасета в ЦНБ разработан специальный классификатор. Его структура отражает основные аспекты деятельности органов власти в их иерархической взаимосвязи. Лексическая база классификатора отвечает терминосистеме существующих классификационных схем (УДК, ББК, Рубрикатор ГАСНТИ). Классификационный индекс состоит из цепочки арабских цифр, разделенных точками. В общем случае он имеет три уровня иерархии. Верхний уровень содержит девять тематических групп: нормативные документы, внутренняя политика, внешняя политика, военная политика, экономическая и социальная политика, отраслевая экономика, социальная инфраструктура, комплексные проблемы (экология, статистика и демография, чрезвычайные ситуации), персоналии.

В качестве базового программного обеспечения автоматизированной системы оперативной подготовки обзоров прессы выбран ППП CDS/ISIS/M. При этом в подсистеме ввода информации используется версия 3.0 данного пакета, а в подсистеме поиска — версия 3.06 (интерфейс Heurisko). Универсальность и широкие функциональные возможности CDS/ISIS не потребовали от создателей системы проведения каких-либо программных доработок. Ввод информации, ее редактирование, формирование упорядоченных по трем ключам сортировки текстовых файлов, формально-логический и семантический контроль корректности данных осуществляются средствами CDS/ISIS, входящими в стандартный комплект поставки.

Стандартными средствами реализована и подсистема информационного поиска, который предполагает, прежде всего, проведение поиска по фиксированному набору полей. В базе данных с рефератами газетных статей предусмотрено семь полей. Они сгруппированы в блоки адресного и тематического поиска. Первый включает поля персоналий-лиц, персоналий-организаций, названий газет и хронологических рубрик (последние, как уже отмечалось, ведутся с точностью до месяца). Блок тематического поиска содержит поля классификационных индексов, терминов (словесных определений этих индексов) и географических рубрик.

При поиске по классификатору информационно-поисковая система предоставляет пользователю возможность просмотра классификационных индексов вместе с их расшифровками и выбора интересующего раздела базы данных. Особенностью реализации этого режима поиска является то, что в самой библиографической записи указанные словесные определения отсутствуют. CDS/ISIS добавляет их при выводе на экран дисплея, обращаясь для этого к подмножеству записей базы данных, в которые занесен классификатор. Указанные записи имеют номера с 1 по 999, при этом номер записи совпадает с цифровым значением классификационного индекса.

Аналогично реализуется и поиск по терминам, которых также в явном виде нет в записи. Информационно-поисковая система сама определяет классификационный индекс, соответствующий заданному термину, а затем выявляет записи, в поисковых образах которых находится установленный системой индекс.

Указанные операции стали возможными благодаря наличию в языке форматирования печати CDS/ISIS функции REF, позволяющей объединить данные, которые фактически содержатся в

разных библиографических записях (в нашем случае добавлять словесные определения классификационных индексов, которые занесены в первую тысячу записей БД).

Дальнейшее развитие системы направлено на создание средств информационного анализа рефератов газетных публикаций. Следует отметить, что основная сложность решения этой проблемы находится в области постановки задач анализа и разработки алгоритмов их решения, т. е. вне программного инструментария. Однако и последний не является пассивным компонентом средств информационного анализа. Например, программная утилита статистической обработки данных в среде CDS/ISIS, созданная в ГПНТБ России, инициировала постановку ряда задач по выявлению латентных связей между лицами и организациями, фигурирующими в газетных статьях. Открытость же программной системы CDS/ISIS, а также наличие в ее составе языка высокого уровня CDS/ISIS Paskal обеспечивают возможность применения этой системы и для решения других задач информационного анализа.

В целом же из опыта использования ППП CDS/ISIS в реально функционирующей автоматизированной системе оперативной подготовки обзоров прессы можно сделать два вывода:

данный пакет позволяет в сжатые сроки создавать автоматизированные системы с типовым набором пользовательских функций (клавиатурный и пакетный ввод данных, адресный и тематический поиск информации, формирование выходных текстовых файлов заданной структуры и др.);

заложенные в CDS/ISIS возможности прикладного развития обеспечивают "живучесть" этого программного инструментария. Он, безусловно, найдет применение и в информационных системах нового поколения, предполагающих использование баз знаний и средств информационного анализа.