

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК 002.53:681.3.016

Л. Г. Тупчиенко-Кадырова ИСПОЛЬЗОВАНИЕ ГИПЕРТЕКСТОВОЙ ТЕХНОЛОГИИ В БИБЛИОТЕКЕ

Рассматриваются возможность использования гипертекстовой технологии в библиотеках, этапы их автоматизации и изменения роли каждого компонента системы «библиотека» в связи с автоматизацией.

Автоматизированные гипертекстовые системы (АГТС) — это новая технология работы с информацией (ее обработки, организации и использования). Они относятся к системам искусственного интеллекта (ИИ), так как моделируют некоторые процессы мышления человека, например, память.

Под гипертекстом (подробно [1—5]) понимают систему, состоящую из фрагментов текста, которые связаны между собой ссылками. Системы, в которых кроме текста используются звук, графика, видео и т. д., называются гипермедиа или системами гиперзаписи [6]. Пробразами гипертекстовых систем являются энциклопедии, словари.

К настоящему времени создано довольно много ГТС, экспериментальных и коммерческих. Условно их можно разделить на две группы: к ГТС первого уровня отнесем те, в которых очень простая система ссылок между терминами словаря и фрагментами текста. А к ГТС второго уровня — те, которые для своего построения используют сложную иерархическую систему ассоциативных связей или типов отношений, как для связи терминов в тезаурусе, так и в текстовых фрагментах.

Последовательность создания ГТС описана в работах [1, 2, 7, 8].

Сначала — сбор информации для решения определенной задачи или по определенной теме. Отбираются документы-источники, параллельно готовится словарь терминов — тезаурус — для данной области знания, на основе которого из источников выделяются фрагменты текста и связываются ссылками с терминами тезауруса. Для ГТС второго уровня составляется иерархическая система типов отношений, которая и связывает между собой термины в тезаурусе и основные термины в текстовых фрагментах.

Затем система корректируется, дополняется недостающими ссылками и фрагментами текста.

В библиотеке ГТС применяются для разных целей, в первую очередь — как поисковые системы. Преимущество их перед традиционными информационно-поисковыми системами в том, что каждый пользователь, в соответствии со своей индивидуальностью, проходит собственный путь поиска, при этом система подсказывает и дальнейшие решения, и возможные ответвления в «пути».

Еще одно направление применения ГТС — создание на основе одной системы различных выходных документов (рефератов, статей и др.) по разным темам и

аспектам. Делается это так: при просмотре ГТС нужные фрагменты текста отмечаются — «запоминаются», затем они располагаются по порядку, текст корректируется, дополняется и документ готов. Преимущества такой системы: гибкость и открытость (т. е. можно добавлять фрагменты текста и ссылки, менять структуру системы).

ГТС также эффективны для решения исследовательских задач. Пока из-за нехватки технических средств доступ к таким системам читателей очень ограничен, но для библиотечных задач, например, для проектирования автоматизированной системы (АС) библиотеки, социологических исследований и др. — такие системы можно создавать. Преимущества их, кроме уже названных, в использовании неформализованных знаний.

Создавать системы гиперзаписи гораздо сложнее, потому что нужны дополнительные технические средства: графический дисплей с высокой разрешающей способностью, видеотерминал для работы с видео- и оптическими дисками, устройство речевого ввода-вывода для работы с аудиовизуальной информацией, устройство «мышь» для графики, лазерный принтер. Но зато возможности такой системы разнообразнее просто гипертекстовой.

Немного о конкретных реализациях подобных систем. Например, пакет программ HyperCard фирмы Apple. На его основе создана система «Путешествие по библиотеке» [9], которая не только показывает расположение библиотеки фирмы, ее отделов (на картах разного масштаба), но и позволяет узнать о направлениях работы, просматривать фонды, проводить поиск (например, конкретный журнал можно найти по любому слову из названия, по тематике, просмотреть содержание этого журнала), послать заказ в соответствующий отдел. При желании можно получить информацию о работниках отделов: их фото, рассказ о себе, своей работе (в будущем эти рассказы можно будет услышать). Создаются базы данных (БД) по истории самой фирмы Apple. Преимущество этой системы — ее комплексный характер.

Эти же программы HyperCard использованы для создания системы «Электронный блокнот» [10]. Это поисковая система по этическим проблемам, поднятым Конституцией США. В ней две части: первая содержит цитаты статей Конституции; вторая — энциклопедия терминов, которые соотносятся с Конституцией. Просматривая энциклопедию, необходимые термины отмечаются, по ним затем можно вести поиск в собрании цитат, либо, просматривая цитаты, можно вызы-

вать отмеченные термины. Выбранные цитаты «запоминаются» в «электронном блокноте». В дальнейшем система была дополнена готовыми БД, соответственно добавились термины и цитаты. Связи в системе осуществляются через так называемые «кнопки» в виде пиктограмм из рисунков и слов (например, «?» или рисунок книги с подписями «блокнот» или «энциклопедия», открытая книга означает, что Вы находитесь в этой подсистеме). Подобные системы экономят время и силы исследователя (в том числе, библиотекаря).

В основу разработок ГТС второго уровня положен смысловой анализ естественных текстов. В [1, 11] предлагаются следующие его этапы: первый — структуризация знаний, второй — систематизация. На первом проводится морфологический и синтаксический анализ, разделение на предметы и отношения, построение системы предметных связей, определение типов отношений, классификация типов отношений. На втором этапе знания систематизируются по определенной иерархии: тематической, фреймовой, по «матрице представления системы» [11, 12] или по какой-либо иной.

Логико-смысловой метод обработки естественных текстов М. М. Субботина реализован в экспериментальной системе ГИПЕРЛОГ во ВНИИПВТИ [13]. Система основывается на массиве высказываний, каждому из которых присваивается номер по порядку ввода в массив, указываются номера высказываний, с которыми имеется логическая связь, и степень близости этой связи [7]. Цель системы — автоматизированным путем вывести рассуждение.

Какое же место займут ГТС на разных этапах автоматизации библиотеки?

Рассмотрим сначала возможные этапы автоматизации библиотечных процессов.

На первом этапе создаются библиографические БД. Таким образом библиотека удовлетворяет потребности читателя в информации о документах. Преимущества перед традиционной технологией: быстрый поиск, большое количество признаков поиска, одноразовый ввод и многоаспектное использование информации, т. е. повышается эффективность.

Затем встает вопрос о том, чтобы читатель получал и первичную информацию в машиночитаемой форме. Итак, на втором этапе автоматизации библиотек будут создаваться полнотекстовые БД, в основном фактографические и, возможно, экспертные системы (ЭС) [14, 15], а также ГТС первого уровня по наиболее актуальным темам, например, БД или ГТС по законодательству, гиперсистема — гид по библиотеке с информацией о расположении отделов, об их деятельности и работниках. Второй этап — переходный от БД к базам знаний (БЗ). Их отличие в том, что БД предполагает хранение неизменяемых данных, а БЗ — изменяемых и неизменяемых [16].

Но ни сейчас, ни потом просто переносить содержание каждого документа в память машины невозможно, так как это будет стоить колоссальных усилий и затрат, да это и не рационально. Ведь в различных документах повторяются не только буквы, слова, словосочетания, которые связываются между собой правилами грамматики и синтаксиса, но и единицы смысла: предметы и отношения между ними (мы их называем фактами, мыслями, гипотезами и т. д.).

Таким образом, встает проблема оптимизации организации текстов, т. е. сжатия информации до букв и звуков и развертывания до аннотаций, рефератов, до воссоздания полного текста конкретного документа. Это — третий этап автоматизации. На этом этапе будут создаваться системы ИИ, в том числе ГТС второго уровня, а также различные «каркасы» для размещения информации, например «матрица представления системы» [11].

Вернемся, однако, к библиотеке, где все это, возможно, будет происходить. Для того чтобы ясно представить себе предмет, нужно определить его отношения с другими предметами (на том же уровне иерархии и/или выше — ниже), посмотреть на него с разных точек зрения. Тогда можно увидеть, что библиотека как один из социальных институтов также представляет собой систему. Рассмотрим библиотеку как систему.

Цель библиотеки — посредничество между документом и читателем, отсюда вытекают ее главные задачи: хранение, обработка и доведение до читателя информации о документах и самих документах. Средствами реализации этих задач являются три главных компонента системы: Документ — Библиотекарь — Читатель. Эти задачи осуществляются в таких главных процессах, как организация (ведение) библиотечного фонда (в том числе комплектование и хранение), обработка документов (интеллектуальная и техническая), обслуживание читателя (запись в библиотеку, собственно обслуживание).

Каким образом изменится роль каждого компонента системы в связи с автоматизацией?

На первом этапе АС почти полностью повторяют традиционные процессы, связанные с документом и читателем: запись, регистрация, обслуживание читателя; комплектование документов, их обработка (описание) для электронного каталога, поиск библиографической информации.

Таким образом в машиночитаемую форму переходит не сам документ, а информация о нем. Библиотекарь по-прежнему является посредником между документом и читателем, обеспечивая комплектование, хранение, обработку и доведение информации до читателя. Однако процесс обработки изменяется: информация не дублируется, а вводится один раз с дальнейшим многоаспектным ее использованием, при этом повышаются требования к тому, кто ее обрабатывает и к точности выполнения им операций. Роль читателя остается прежней, он — заказчик документов (поиск по электронному каталогу ведет сначала с помощью библиотекаря-консультанта, а затем — самостоятельно). Но некоторые читатели будут участвовать в процессе обработки документов и как критики существующей поисковой АС, и как специалисты в определенной области, которых привлекают к составлению стандартных списков ключевых слов и/или тезаурусов.

На втором этапе автоматизации перестраивается вся система в целом. В машиночитаемую форму переводится не только библиографическая, но и первичная информация — иногда документы целиком, а чаще — отдельные его фрагменты. Таким образом, начинается процесс разрушения целостности отдельного документа и одновременно интеграция всеобщего знания. Роль читателя возрастает и как заказчика тем для полнотекстовых БД, и как эксперта для составления ЭС, и как активного пользователя всех систем, в том числе ГТС, где он может создавать собственную ГТС и выходные оригинальные документы на ее основе. Таким образом, читатель все больше становится пользователем. Роль библиотекаря как посредника между заказом и выдачей документов уменьшается. Зато обработка документов станет более интеллектуальной и ответственной. Например, при создании ЭС с экспертом «разговаривает» библиотекарь-посредник, так называемый когнитолог или инженер по знаниям, который получает знания от эксперта и организует их [16].

На третьем этапе автоматизации документ переходит в машиночитаемую форму и будет претерпевать разрушение до отдельных фрагментов текста и в меньшей степени воссоздание до первоначального вида, так как читатель-пользователь чаще нуждается не в самом документе, а в информации, которую тот содержит, и не во всей, а в какой-то ее части. В единой АС чита-