

# Персональная электронная библиотека

Броди Д.

Основное правило использования компьютерной техники состоит в том, чтобы она не мешала людям выполнять обычную работу с информацией на бумажных носителях. Учёный-гуманитарий или специалист по общественным наукам с годами вырабатывают индивидуальный у каждого набор привычек работы: читают ли они, пишут или наводят определенный порядок в своих рукописях и в своих библиотеках — книгах и журналах.

Система пользования документами должна подготовить условия для работы в новой, то есть в электронной персональной библиотеке. Конечно, не следует питать иллюзии, полагая, что электронными документами могут быть вытеснены книги или журналы. При более интенсивном общении с текстом пользователь персонального компьютера (ПК) обычно стремится получить доступ к бумагам, чтобы улучшить их эстетические качества в электронной форме, а также для того, чтобы обрабатывать, комментировать, индексировать, перерабатывать, цитировать, дополнять их и т. д.

Работа в обычном кабинете, без большого количества компьютеров, — электронная персональная библиотека — неоценимая поддержка для преподавателей, исследователей, редакторов, переводчиков или архивистов.

Я встречал философов, которые утверждали, что им нужны лишь голова, бумага и авторучка, чтобы работать. И всё-таки человеку нужна комната или, по крайней мере, пара полок с книгами и бумагами, находящимися на расстоянии вытянутой руки: специальная и реферативная литература, записные книжки, толковые и двуязычные словари, списки рукописей, своих и чужих, например своих коллег, ученические работы, собственные недописанные тексты...

Многие преподаватели и учёные используют свою домашнюю библиотеку приблизительно таким образом: поступающие книги и журналы, фотокопии, доклады и рукописи размещаются на полках или в ящиках стола, а компьютерные документы размещаются на магнитных или оптических носителях (дискеты, жесткие диски или сетевые серверы, компакт-диски).

Кроме собственной библиотеки, пользователь обычно держит на своём столе и полках вспомогательный материал, необходимый для задач исследования и обучения, и все это в виде неоконченных рукописей и фрагментов, заметок и выдержек на бумажных или магнитных носителях.

Обычно имеется также доступ к внешним источникам, таким, как библиотеки и архивы, базы данных, которые не могут быть доступны в любую минуту, а требуют поддержки почты или телефонной сети. Самый важный материал можно подключить к домашней библиотеке путем архивации, фотокопирования или сохранения файлов. Из библиотеки выходят бумажные носители или файлы с рукописями статей и книги, доклады, исследовательский материал и т.д., некоторые экземпляры которых обычно присоединяются к собственным домашним библиотекам.

При работе над этими документами «подручная библиотека» — это основа основ. Это относится не только к информации на бумажных носителях, но и к компьютерному архиву. Эта база данных локально расширена настолько, чтобы пользователь (или рабочая группа) мог распоряжаться материалом примерно так же, как и книжными полками. Самая важная черта локальных баз «подручной библиотеки» — доступность. Документы должны быть доступны в разных планах, размещены на блоках памяти в собственных компьютерах или в локальной сети, для обработки собственными фаворитными программами (программами обработки текста, обработки иллюстраций, программами калькуляции). Они должны быть дополнены и оснащены пометками и перекрёстными ссылками — это так называемый гипертекст.

Собрание документов и материалов должно быть организовано так, чтобы пользователь мог без особых усилий понять общее содержание или карты баз данных, легко проверить не только отдельные документы, но и структуру этих баз (это должно быть так же легко, как и перебирать книги на полках или сортировать бумаги в папках).

Я выбрал метафору «подручная библиотека» именно потому, что хочу провести параллель с тем, как именно преподаватель или исследователь обычно использует книжные полки или стол в своём кабинете, не имея в виду, что собрание информационных материалов должно быть небольшим или что должно быть мало документов. Виртуальная база данных может вместить больше, чем бумажная, и можно представить себе, что тот же пользователь (или рабочая группа) получают в своё распоряжение большое число баз данных, приспособленных для особых профессиональных задач. Материал в локальные базы данных можно передать из разных мест и через многие носители: на компакт-дисках, дискетах или в книжных томах, через компьютерную сеть или из книг и журналов через сканеры в комбинации с программой оптического распознавания текстов и обработки изображения. Здесь же концентрируются собственные произведения и работы коллег.

Кроме того, материалы могут быть взяты из внешних источников, под которыми я подразумеваю материалы, которые не могут быть сразу интегрированы в локальные базы. Например, подобный материал должен быть обработан группами исследователей, издателей или библиотекарей. Уже сейчас учёные определённой отрасли науки могут сделать свои доклады или изыскания публично доступными через международные университетские сети (за исключением случаев, когда из-за противоречий в авторском праве разных стран усложняются научные контакты). Сейчас, без сомнения, продаётся всё больше компьютерной литературы и дешевле всего стоят компакт-диски.

Таким образом, всё больше информации появляется на внешних носителях, с которых пользователь может выбирать материал для того, чтобы пополнить им свои локальные базы данных. Каждый, кто пытается использовать подобный материал, сталкивается с трудностями, связанными с различиями форматов. Получить документы на бумажных носителях не представляет никакой сложности, обычно используются форматы ASCII или Post-Script или другие знакомые форматы. Но когда речь идет о комплекте простых или сложных документов, то требуется много времени и компьютерных знаний, чтобы адаптировать их к своим компьютерам или программам. При конвертации форматов файлы часто деформируются. Речь идёт о главах, рубрикации, пометках, особых языковых символах и т. д. Формат SGML, например, позволяет без особых проблем сохранить структуру текстов при их переводе через границы, независимо от программных продуктов, оперативных систем, названий файлов, национальных стандартов и т. п. (мы имеем в виду западноевропейские языки). Важно, чтобы локальные базы данных были интегрированы с материалом во время работы.

Таким образом, пользователь может редактировать, комментировать, объединять материал и т. д. Находятся возможности объединить информационный материал собственного производства с локальными базами данных, процесс становится более сложным, и вряд ли его можно полностью автоматизировать. Из этой рабочей среды материал поставляется разными способами: напрямую, на простых пишущих машинках, в виде файлов систем TEX или LATEX, фотокопированием, для распространения на магнитных носителях или через компьютерную сеть путем электронной почты, системы конференций или ftp.

Тесное сотрудничество между преподавателями и библиотекарями желательно. В ближайшем будущем преподаватели и учёные, работая в своей электронной персональной библиотеке, будут связаны с библиотеками школы или института и смогут получать оттуда информацию и взамен поставлять свои доклады и отчёты, которые станут публично доступными документами. Обязательные экземпляры, которые рассылаются по научным библиотекам, должны в определённых случаях производиться в цифровой форме. Компьютерные документы должны быть в формате, доступном для каталогизации и автоматизации.

Документ, который поступает или выходит из библиотеки, должен поставляться в конвертируемом формате, а именно в таком, чтобы его могли использовать в разных компью-

терах и при помощи разных программ. Очень важно при этом, чтобы структура документа и его основные показатели были сохранены (текст, примечания, рубрикация, ссылки на источники, автор, заголовки и другие библиографические данные). Пользователю важно получить документ в таком виде, чтобы его можно было легко редактировать. Всё это позволяет сделать стандарт SGML.

Для подобной системы документации следует соблюдать целый перечень требований, В системе SGML Darc, которая используется в Королевском технологическом институте, мы пытаемся удовлетворить большинство следующих требований:

**1. Максимально облегчить пользователям работу с документами** — чтение документов на электронных или бумажных носителях, редактирование этих документов, их объединение или, наоборот, извлечение отдельных фрагментов из данного материала; распечатку или подготовку файлов для фотонаборного оборудования; поиск, компиляцию, содержательный, семантический, синтаксический или лексический анализ документов и т.д.; максимально приспособить конфигурацию системы для потребностей пользователя.

**2. Документ должен быть конвертируемым.** Пользователи должны тратить минимум усилий, умея при этом работать, например, с избранными программами или с графическим редактором в своей операционной системе (обычно в MS-Windows, Unix или Macintosh System), а документы должны легко конвертироваться в другую среду.

Система должна поддерживать распространение материалов, например, источники текстов сами должны быть доступны и не произведены в компилированные версии, для которых нужны особые программы.

Здесь потребности авторов сталкиваются с коммерческими интересами. По вполне объяснимым причинам производители программных продуктов стремятся к тому, чтобы потребители предпочитали покупать только их программы. Поэтому авторы, которые хотят опубликовать свои материалы в компьютерной форме, например на компакт-дисках, встречаются с предложениями той или иной компании за тысячу крон предоставить CD-ROM, позволяющий интегрировать эти материалы в систему, дающую возможности для индексации, редактирования, поиска и т. д. Таким образом дискредитируется идея о конвертируемости документа. Это в принципе то же самое, как такой пример: писатель пришёл в издательство, чтобы опубликовать своё произведение, а ему в ответ издатели заявляют, что они напечатают его книгу невидимыми чернилами их собственного изготовления, а для читателей сделают специальные очки за тысячу крон.

Названия документов обычно содержат ссылку на ту программу, в которой они созданы, например, имеют расширение doc для файлов Word или xls для Excel. Затем автор определяет свое собрание документов по имени машины или операционной системы: например, автор работает на Unix или в Windows, или в Macintosh. В большинстве случаев было бы лучше разместить документы в компьютерном центре. Современная техника позволяет обработать один и тот же документ или собрание документов разными способами, при помощи различных программ и на разнообразных машинах. Для гуманитария или исследователя в области общественных наук — это естественный способ работы. Например, у меня есть источник или интервью, которые я намерен обработать разными вариантами, но я не хочу, чтобы мои документы были приспособлены только к моей операционной системе или для одной-единственной программы.

**3. Система должна поддерживать сотрудничество в группах редакторов и исследователей, работающих с её документами.** Многие сотрудники должны, иметь возможность пользоваться одними и теми же документами или локальной базой данных. Здесь очень важен контроль принадлежности, как каждого отдельного документа, так и организации всей базы документов. В редакционной группе или группе исследователей главный редактор или руководитель научной группы может, в частности, организовать всю документальную базу и

менять все её содержание. В то время как другие сотрудники с более ограниченными возможностями имеют право редактировать отдельные документы, или, например, только читать определенные документы и лишь для себя использовать перекрестные ссылки и создавать свои персональные картотеки.

**4. Система должна поддерживать обработку (а не только, как уже говорилось, презентацию) документов.** Некоторые читатели, возможно, недоумевают, почему слово «мультимедиа» не представлено в данной статье. И, возможно, кому-то покажется, что чересчур большое внимание уделяется теме обработки текстов.

Я думаю, что развитие базовых систем обработки текста и иллюстраций не должно оказаться в тени более перспективного направления — презентации конечного продукта ПК. Термин «мультимедиа» означает, что компьютеры представляют текст, иллюстрации, анимационные фильмы, аудио- и видеоприложения, которые с помощью соответствующих средств можно вводить в компьютер или показывать. Но сегодня все стремятся к тому, чтобы создавать цифровые системы, которые приближаются по качеству к студийной аудио- и видеопродукции. Иными словами, желательно одновременно использовать множество каналов с видео- и аудиоинформацией, которая может быть обработана и собрана в единую программу. Подобный способ, однако, редко применяется. Чтобы создавать полноценные мультимедийные программы, требуется всё-таки высокий уровень компетентности, серьёзные технические знания и в особенности знания о форматах, об изобразительных и музыкальных возможностях, знания, более основательные, чем, например, требуемые при производстве музыкальной видеопродукции.

Информация, построенная на основе аудио- и видеоматериала, требует для своего воспроизведения времени и читается не так, как текст и обычная графика, к тому же при этом затруднён поиск информации, конвертация в другие форматы и т.д. Без сомнения, со временем и непрофессионалам станет легче редактировать и комбинировать цифровую звуковую информацию и цифровые подвижные иллюстрации.

Звук, записанный в цифровом формате, требует примерно 5 Мбайт в минуту для воспроизведения в режиме моно и 10 Мбайт в минуту — для стерео, а одна минута цифровых цветных видеоприложений требует не менее 1,5 Гбайт памяти и соответствующего оснащения компьютера. И графика, и другие иллюстрации могут использоваться без проблем благодаря новым программам архивации, позволяющих добиться приемлемого визуального качества с соотношением архивации 1:50 или более, что значительно снижает требования к объёму памяти (например, JPEG от Joint Photographic Expert Group — стандарт сжатия, оставляющий только информацию, важную для восприятия, и отсекающий малозначущую информацию, такую, как монохромные поверхности и пр.).

Однако если базовые проблемы не найдут удовлетворительного решения, другими словами, базы данных текстов и иллюстраций высокого качества не будут доступны ни учителям, ни ученикам в каждодневной работе, то даже мультимедийные компьютеры не смогут им помочь. Здесь требуются большие инвестиции в систему использования документов — и от промышленных предприятий, и от государственных структур.

**5. Система должна помогать пользователю редактировать, упорядочивать и просматривать материал в локальных базах данных.** Следует прежде всего проанализировать проблему просмотра, когда каждый пользователь подбирает нужный ему материал для использования на своем компьютере. Проблема просмотра актуальна даже по отношению к отдельным документам.

Работая с картотекой в библиотеке, пользователю приходится тратить определенные усилия в процессе поиска нужной книги. Кроме того, он должен узнать место этой книги на полке. Только взяв книгу в руки, можно выяснить, много ли информации она содержит.

Я приблизительно знаю по опыту, как определить год издания книги, найти пометки или комментарии. Раскрывая книгу, я, конечно, автоматически запоминал, что и где находится в

ней. В процессе чтения я знаю, на каком его этапе я нахожусь: в начале книги, в середине или в конце. Мы оставляем закладку в книжке на той странице, к которой хотим вернуться, делаем разные пометки на её полях, подчеркиваем в ней что-либо, просто загибаем страницу или нечаянно проливаем кофе на неё — есть целый комплекс визуальных и тактильных ориентиров, которые позволяют получить информацию. Нужны новые способы её получения. Чтобы электронная информация стала доступна, нужны новые пути, которые помогут и писателям и читателям ориентироваться в ней, просматривать и систематизировать имеющийся материал.

Представлять сейчас информацию на бумажных носителях немодно. Прошли века, прежде чем типографские традиции презентации текста обрели форму. Первые поколения книгоиздателей имитировали почерк, и это считалось престижным, а сейчас средневековые рукописи кажутся нам чересчур массивными. Традиция делить текст на фрагменты — инновация, как и деление на главы и использование принципа рубрикации, разные стили, нумерация страниц, пометки или особые знаки. Сегодня мы ждём читабельности текста.

Что касается электронной информации, то соответствующее развитие формообразующих традиций уже налицо. Система должна снабжать пользователя вспомогательными средствами, такими, как карты к базам данных, «линки» гипертекстов, то есть перекрёстные ссылки или указатели к разным информационным массивам. Линки гипертекстов часто используются как указатели, помогающие найти материалы, находящиеся вне данного текста. На мой взгляд, очень важно также развивать средства, помогающие просматривать каталоги материалов. Электронная картотека должна не только помогать разобраться в базе документов, но и давать возможность перемещать информацию, например, так же, как и я передвигаю книги на книжных полках. Разные пользователи или группы пользователей должны иметь возможность по-разному использовать одну и ту же базу данных, каждый в соответствии со своими задачами. Один и тот же документ может быть размещён в разных директориях (например, когда делаешь копии одного и того же документа для разных задач).

Множество программистов сейчас предпочитают принцип «информации на кончиках пальцев», по выражению Билла Гейтса, или «Seamless information Environment» (бесшовная информационная среда): пользователь должен найти необходимое, не задумываясь о том, как организована информация. Этот принцип применим, когда речь идёт о поиске отдельных документов или фактов, но в своей собственной библиотеке пользователю хочется контролировать состояние документов.

**6. Система должна поддерживать обработку больших и сложных документов, а также собрание документов.** Следует устранить угрозу уничтожения отдельных документов и собрания документов. Система должна позволять работать с большими документами, насчитывающими сотни или тысячи страниц, и с базой документов, охватывающей много тысяч документов. Время ожидания не должно превышать время возникновения мысли.

**7. Одно и то же собрание документов должно использоваться на разных уровнях.** И новички, и специалисты, и школьники, и школьные учителя, и учёные должны с помощью разных фильтров и других средств использовать одну и ту же базу документов и перемещать материалы в свою персональную электронную библиотеку. Можно создавать отдельные издания для разных групп пользователей из одного и того же материала.

И школьники, работающие с первой главой романа А. Стриндберга<sup>1</sup> «Красная комната», и доктора литературоведения должны иметь доступ к полному виртуальному изданию собрания сочинений Стриндберга с комментариями и алфавитными указателями. Разумеется, новичкам и специалистам требуются разные тексты, но самое приемлемое решение, которое гарантирует большие издания и широкое распространение — собрать, например, эти тексты на компакт-дисках.

<sup>1</sup> Стриндберг Август (1849–1912) — шведский драматург и прозаик, родоначальник современного шведского литературного языка.

Вряд ли есть какие-нибудь причины (за исключением чисто коммерческих), чтобы производить компьютерные двойники словарей или учебных пособий, которые предназначены только для студентов или школьников. Я уже однажды высказывал мысль о том, что не только ученики, но и учителя должны иметь возможность использовать компьютерную литературу для школ. Я твёрдо убеждён, что все, даже ученики, по большому счёту оказываются в выигрыше, если учителя используют компьютерную систему и компьютерные программы в своей собственной работе, в процессе обучения, в подготовке курса, в планировании лекций.

Есть шанс обработать материал и обогатить им эту систему.

Здесь напрашивается сравнение с сущностью библиотеки: представьте себе, как было бы уныло, если бы школьные библиотеки содержали только примитивные учебники, из которых учителям нечего было бы почерпнуть. К тому же ученики перенимают от учителей опыт общения с компьютером как с рабочим инструментом.

Сейчас многие издательства планируют использовать свои базы данных, обрабатывают печатные энциклопедии и выпускают книги для определенных групп: например, «Всё о собаках» или «Всё об Африке» и т. п. При распространении дешёвых магнитных или оптических носителей вряд ли есть причины (опять же кроме коммерческих) для подобного опыта. Лучше произвести всю энциклопедию и дать возможность пользователю сориентироваться в ней и выбрать нужные данные, например о собаках.

Общий принцип таков: издание должно содержать как можно более исчерпывающий материал, даже если он будет использован не полностью. К переводу можно приложить работу на языке оригинала (например, в Европе существует традиция книгоиздания на параллельных языках, и эти издания становятся все более привычными). Короткое предисловие к тексту даёт представление о теме.

Большой выигрыш получается тогда, когда специалисты и новички могут использовать одни и те же базы данных, но на разных уровнях. Это пока удается редко. Например, аптека берёт лекарства для покупателей у фармацевтической фирмы-производителя. При этом на функциональной клавиатуре, подключённой к компьютеру, пользователь может нажать на любую кнопку, например «Головная боль» или «изжога», и получить полезные советы об имеющихся лекарствах. Другой пример — система ADB, которую используют многие частнопрактикующие стоматологи, в том числе при ведении журналов приёма пациентов. Вероятно, многие группы пользователей получают доступ к разным фрагментам одного и того же материала. В аптеке и персонал, и пациенты могут пользоваться одной базой данных. А у стоматолога пациент может получить выписку из своего журнала, дополненную переводами сложных терминов.

**8. Система должна быть адаптирована к расширению и модификации.** Система должна иметь средства для работы с различными видами мультимедийной информации и поддерживать операции импорта и экспорта файлов других систем.

**9. Простое оснащение.** Пользователь должен иметь возможность, не обладая доступом к новейшим рабочим станциям или техническому персоналу, читать, писать и редактировать документ, а также работать со всей базой документов или выписками из неё. Работа должна быть оформлена настолько просто, чтобы при необходимости пользователь мог взять её с собой на дачу или передать своим коллегам, находящимся в другом месте. А более развитые системы обработки документов предназначены прежде всего для крупных предприятий или администраций. В первую очередь, гуманитарии ограничены тем, что многие компьютеры в университетах и высших школах приближаются к тому, что американцы называют административным компьютерингом. А академический компьютеринг, напротив, предполагает более подвижный способ работы.

**10. Система должна уметь использовать структурированные документы.** В структурированных документах имеют значение не только содержание, тексты и иллюстрации, но и способ, с помощью которого содержание организовано по главам и рубрикам, с разного рода текстовыми документами для различных типов содержания (текст, сноски, цитаты, комментарии, фамилия и имя автора, способ печати, год издания и пр.), а также характер истолкования того или иного структурированного документа.

### **SGML (Standard Generalized Markup Language) — язык для описания форматов компьютерных документов**

В описание документа необходимо ввести пометки процедуры и описывающие пометки. Пометка процедуры говорит о том, как та или иная информация может использоваться, например, что компьютер или писатель или фотомагазин должны делать с текстовым файлом. Каждый, кто использует ссылки в рукописи, знает, что это означает. Определённая ссылка может, например, означать, что определённый фрагмент или фраза должны быть набраны 12-м кеглем и курсивом. В этом случае пометка процедуры содержит инструкцию к наборному цеху.

Описывающая пометка даёт справку о том, какие элементы содержит текст. Некоторые фразы, имеющие тот же типографский дизайн, например набранные курсивом 12-м кеглем, можно выделить ещё и другими типами элементов, например: 1-я и 2-я рубрики — то, что автор хотел бы подчеркнуть, 3-я рубрика — слово на иностранных языках,

4-я — название книги. Четыре типа элементов могут таким образом использоваться в одном типе текста (12 кегль, гарнитура «курьер», *курсив*).

Традиции шведского типографского набора могут варьироваться. Цитата порой выносится на левое поле, порой набирается меньшим кеглем. Во французских изданиях комментарии, относящиеся к цитате, высказывания нередко располагаются непосредственно перед цитатой. А в английской традиции книгопечатания автор цитаты часто указывается непосредственно после цитаты.

При использовании описывающих маркировок пользователь заботится в принципе не только о подобных эстетических задачах, но довольствуется тем, что ставит пометку <q> перед цитатой или </q> после нее. Затем следует решить задачу, как пользователь при помощи типографских форматирующих программ и доступных возможностей шрифта может набрать необходимый текст.

В качестве образца я предлагаю международный стандарт SGML, который был выпущен в октябре 1986 г. Стандартное пособие Чарльза Гольдфарба «Карманный словарь SGML (Clanderon Press Oxford, 1990) одобрено Международной организацией стандартов (ISO). За основу этого стандарта был взят стандарт GML, который Ч. Гольдфарб в конце 1960-х гг. разрабатывал для IBM. Идея заключалась в том, что документ (это относится к юридическим документам) должен быть помечен единым способом так, чтобы размещаться в различных системах для хранения, форматирования и обработки информации.

Современная техника открывает определённые возможности для того, чтобы обработать один и тот же документ или одно и то же собрание документов множеством способов и с помощью многих программ на разного типа машинах. Подобное развитие предполагает общий язык описания документов, своего рода эсперанто, чтобы не зависеть от типа машин, программных продуктов и национальных способов пометок. Один и тот же документ может быть использован в машинах SUN, Macintosh или машинах с DOS-Windows и обработан разными программами.

В этой области ведётся активная международная работа, связанная с развитием SGML и целым рядом «дочерних стандартов», таких, как DSSSL и Hy Time. DSSSL (Document Style Semantic and Specification Language — обычно её называют «дизелем») управляет типографским форматированием, то есть стремится к тому, чтобы все тексты были приспособлены для разных средств массовой информации с разным оснащением. DSSSL предполагает, что

документ зафиксирован в формате SGML и поддерживает стандарт Draft International Standard (на август 1991 г.).

Ну Time (ISO 10744), который был одобрен в апреле 1992 г., после рекордно короткого срока подготовки, является первым международным стандартом для оформления гипертекстовых и мультимедийных документов (а точнее, это зависящие от времени документы, содержащие и звук и движущиеся картинки).

Основа Ну Time — стандарт SGML. Эта стандартизация открывает новые возможности для того, чтобы преодолеть любые различия между странами, между личными компьютерами и программами, между уровнями образования и проблемами, между исследователями и обучающимися, между компьютерными пользователями-энтузиастами и прочими — то, что сегодня препятствует объединению усилий для создания и внедрения содержательной базы данных хорошего качества.

В стандарт SGML (ISO 8879) включен Reference Concrete Syntax, который является исходным пунктом для переводов между разными компьютерными средами, национальными языками, форматами и т.д. В этот синтаксис входит также способ пометки (именуемый Base Character Set), который согласуется со стандартом ISO 646 (7-битовый стандарт, известный как IRV—International Reference Version). Как и ASC-формат (так называемый 0–127), он положен в основу множества других международных текстовых стандартов.

При помощи 128 знаков можно изобразить практически любые буквы и знаки европейских языков. Самые популярные знаки помечаются особым сочетанием клавиш. SDML не содержит правил кодирования документов, а представляет собой международное соглашение о том, что помеченные этим форматом документы сообщают пользователям информацию о доступе к тексту (для машины или человека).

То, что SGML — язык описания, означает, что SDML не сообщает о том, какие именно типографские знаки используются. При чтении или фотокопировании SGML — документы можно перевести в такой формат программ, которые следят за тем, чтобы перевести его в желаемую форму.

Речь идёт не о том, что обычный пользователь должен употреблять формат SGML. Можно взять версию, подходящую для той или иной программы (с буквой Ё, например). Альтернатива есть: документы SDML конвертируются в другие программы и там обрабатываются. Многие из наиболее популярных текстовых программ могут получить поддержку SGML.

Документ SGML состоит из трех частей. Первая часть — декларация SGML, которая разъясняет, каким способом можно построить или модифицировать ISO-8879. Например, пометка может быть длиной в 128 знаков, а это означает отказ от формата ISO-8879, который содержит всего 8 знаков. Вторая часть — DTD (Document Type Definition), который предписывает, как должны помечаться подобные документы. В формате DTD, можно пометить деловые письма, романы, научные монографии и т. д. Третья часть — это само содержание документа. Нет необходимости в том, чтобы декларация SGML и DTD сопровождали каждый отдельный документ, но они должны быть доступны системе, в которой пользователь работает с документами.

Несколько слов о международном проекте TEI (Text Encoding Initiative), который взял на себя задачу выполнить рутинную работу кодирования гуманитарной и общественной литературы и в этом смысле объединяет гуманитариев и учёных-обществоведов из разных стран в различные рабочие группы.

Цель TEI — «сформулировать и аккумулировать принципы обработки научных текстов литературного, лингвистического, исторического и других гуманитарных направлений».

### **Некоторые преимущества документов SGML:**

*Описание даёт однозначное толкование:* программа или пользователи могут однозначно истолковать документ, подготовленный в формате SGML, иными словами, можно избежать того, что один и тот же знак имеет разные смыслы и представляет разные пометки (например, курсив, цитата и т.д.)

*Есть рекомендации, как можно использовать документ:* документы можно конвертировать в программы редактирования текстов, формировать оригинал-макеты для типографий, адаптировать для текстовой базы данных или базы данных другой структуры, использовать для строгого поиска (для того, чтобы найти отдельные слова или фразы) и/или для структурного поиска (при котором язык поиска использует отношения между элементами, на которых строится документ или база документов) и т.д.

*Документы конвертируемы:* SGML-документ в принципе независим от платформ и программ и может к ним адаптироваться; отсутствуют ограничения, обусловленные рамками национального языка, традиций маркировки (когда речь идёт о западных странах) или типа клавиатуры.

*Документы допускают обработку:* документ можно в принципе редактировать с помощью простейших программ и печатать на любом печатающем устройстве (вначале даже возникла идея приспособить SGML-формат для пишущих машинок).

SGML-формат понятен для простых пользователей: пометка мнемонична и непосредственно примыкает к тексту. В этом важное отличие, например, от множества типовых текстовых файлов, которые сопровождаются кодами и указателями, понятными для машин, но неясными почти для всех пользователей.

Очевидно, что пользователь не должен применять формат SDML к своим документам вручную. Естественный способ работы — использовать программы, позволяющие конвертировать другие форматы в SGML-формат. Но каждая пометка позволяет не очень опытным пользователям понимать ее смысл.

*SDML-документы — текстовые файлы,* которые можно использовать как базы данных: можно в одном документе или собрании документов найти все элементы определённого рода, например, литературные рефераты или все основные рубрики, все цитаты на греческом, которые содержат определенные слова или удовлетворяют другим критериям; можно экспортировать из документа все фрагменты, содержащие имя писателя, и потом делать с ними всё что угодно и т.д.

Весьма привлекательные возможности для библиотекарей и других пользователей заключаются в том, что TEI описывает всю информацию, которая требует каталогизации и библиографической обработки согласно международным стандартам и эта информация должна быть доступна перед использованием SGML-документа.

Такие элементы, как фамилия, имя писателя, редактора, заголовок, место и дата публикации, издательство и другие данные оформлены как часть документа. Тем самым каждый текст снабжается информацией о том, как провести полную каталогизацию и получить доступ к библиографическим источникам, которые очень важны для учеников или преподавателей.

*Любой документ должен быть использован вновь* для самых разных целей: документы живут дольше операционных систем и программ, в которых они созданы.

В заключение вспомним Альда Мануция, который был книгопечатником в Венеции в конце 1400-х — начале 1500-х гг. и издал множество произведений античных писателей в «формате времени» (то есть в виде квадратных томиков, которые можно было положить в карман), грамотно набранных стандартизированным и легко читаемым шрифтом и изданных большим для того времени тиражом (тысяча экземпляров); в течение веков эти красивые книжечки считались самыми безупречными изданиями трудов латиноязычных авторов.

Я считаю, что будущая компьютерная литература представляет ту же самую счастливую комбинацию доступности, популярности и высокого качества. Именно таким условиям и удовлетворяет международный язык электронных документов — SGML.

*Перевод со шведского Катарини Мурадян*