

гибкую политику в этом вопросе. Пока эта информация присутствует в текстах документов систем по Северо-Западному и Центральному округам.

РУБРИКАТОР

В настоящее время не существует утвержденного на государственном уровне классификатора арбитражной практики, рекомендованного для использования в судах. В каждом ФАС классификация документов ведется в соответствии с собственным классификатором, степень проработки которого в большинстве случаев неглубокая. Наиболее проработанные классификаторы используют

ся в Северо-Западном и Уральском ФАС. Сделанные на их основе рубрикаторы применяются и в соответствующих системах КонсультантАрбитраж:ОкружнойВыпуск. В системах ФАС других округов будет использоваться специальный рубрикатор арбитражной практики, разработанный на основе подготовленного ЗАО "Консультант Плюс", по заказу Российского фонда правовых реформ, классификатора правовых актов, утвержденного Указом Президента РФ от 15.03.2000 № 511. Данный рубрикатор доработан специально для классификации арбитражной практики с учетом существующих классификаторов ФАС и мнений российских ученых. В перспективе планируется перевод на единый рубрикатор всех систем Консультант Арбитраж:ОкружнойВыпуск.

УДК 004.41:347.965

М. Ф. Королева

КонсультантПлюс: программная технология

Рассматривается собственный специализированный формат базы данных, разработанный КонсультантПлюс и позволяющий достичь высоких технических характеристик.

На сегодняшний день КонсультантПлюс — признанный лидер в создании баз данных, предназначенных для работы со слабо структурированной информацией, в которой преобладают текстовые поля большого объема. Во многом этому способствовала программная технология, главные составляющие которой:

- оригинальный формат базы данных, ориентированный на текстовую информацию большого объема;

- динамическая актуализация и корректировка данных и индексов;

- гипертекст, не требующий полного перекомпилирования при изменении связей или вводе новых документов;

- простой и эффективный механизм репликации баз данных;

- интерфейс: мощные поисковые и аналитические средства;

- многоплатформенная технология программного обеспечения.

1. ФОРМАТ БАЗЫ ДАННЫХ КОНСУЛЬТАНТПЛЮС

Решающим фактором, позволившим достичь высоких технических характеристик, явилась разработка собственного специализированного формата базы данных, обладающего рядом свойств, важность которых для пользователя трудно переоценить.

1.1. Переменная длина записей

В отличие от стандартных баз данных (FoxPro, dBASE), которые заранее отводят для полей пространство описанного размера, независимо от наполнения, база данных КонсультантПлюс содержит

записи переменной длины, учитывающие реальное наполнение полей. Обеспеченное таким образом компактное хранение информации не сказывается на скорости операций с базой данных, причем предусмотрена операция COMPRESS, дающая максимальное сжатие данных.

1.2. Отсутствие ограничений на размер текстового поля

Обычная длина текстовых полей (в формате dbf они называют *тетра-полями*) составляет 64 Kb. В базе данных КонсультантПлюс этот размер практически не ограничен, что является важнейшим показателем для баз данных, содержащих декларативную информацию.

1.3. Сильная степень сжатия текстовых полей

Эффективность сжатия достаточно высока: 2–3 раза в зависимости от специфики текстов и размеров документов. Если в качестве примера взять базу данных КонсультантПлюс:ВерсияПроф, то степень сжатия составляет 2,8 раза, что лишь немногим хуже, чем если паковать тексты документов, например, программой arj.exe. В процессе работы происходит динамическая распаковка, абсолютно безболезненная для работы программы и поэтому незаметная для пользователя.

1.4. Развитый инструментальный набор настроек формата записи

База данных КонсультантПлюс содержит файл настройки и обладает широкими возможностями для изменения таких различных характеристик, как:

- количество полей, способы их задания и индексации;

разделение полей на физические и виртуальные, поисковые и справочные;

способы построения результатов поиска (списков);

виды сортировки записей и т. д.

К интересным особенностям формата базы данных следует отнести возможность динамически, в уже существующей базе данных, без ее перегрузки, проводить следующие операции:

добавление нового поля или удаление существующего;

изменение типа индексации или типа задания существующего поля.

2. ДИНАМИЧЕСКИЙ МЕХАНИЗМ ДОБАВЛЕНИЯ ЗАПИСЕЙ И ИНДЕКСАЦИИ ПОЛЕЙ

В программную технологию КонсультантПлюс с самого начала был заложен механизм динамической актуализации информации в базе данных. Причем это в равной мере относится и к процессу добавления/изменения записей базы данных, и к формированию индексных файлов.

При добавлении записи в базу данных не требуется строить заново индексные файлы, в них автоматически встраиваются новые индексы, которые относятся к этой записи. При изменении записи в базе данных происходит автоматическая корректировка индексных файлов, что означает верификацию и изменение индексов, относящихся только к этой записи. Такой подход, с одной стороны, позволяет организовать максимально эффективное (оперативно, технологично, дешево) информационное сопровождение пользователей, а с другой — требует от разработчиков постоянного совершенствования структуры базы данных и алгоритмов работы с ней. Это объясняется тем, что, в отличие от систем, требующих при актуализации полной замены Информационного Банка, механизм динамической подгрузки предполагает проведение всех операций по актуализации базы данных непосредственно на компьютере пользователя, что предъявляет особенно жесткие требования к надежности и быстрдействию данных процедур.

Поэтому все алгоритмы ориентированы на текстовые поля больших размеров и обеспечивают высокую скорость всех операций с базой данных. К настоящему моменту достигнуты если не предельные, то достаточно высокие результаты.

На компьютере скромной (по современным меркам) конфигурации (486 DX4-100, 8 Мб RAM) база из 13 800 документов общим объемом около 250 Мб (в виде ASCII-текстов) индексируется DOS-программой за 1 час 20 минут, 32-разрядной Windows-программой — за 35 минут. На компьютере PentiumPro с 32 Мб RAM 32-разрядная программа КонсультантПлюс дает результат еще более впечатляющий — 15(!) минут.

При этом следует иметь в виду, что, в отличие от баз данных, использующих стандартные СУБД (FoxPro, dBASE, Clipper и т. д.), индексы строятся не только по реквизитам документа (дополнительным полям), но и по текстам документов непосредственно, т. е. фактически “ключами” являются все

слова, встречающиеся в текстах документов базы данных.

Лучшие из известных на сегодняшний день аналогичных разработок, и не только отечественных, показывают в пять раз худший результат, а большинство — проигрывают на порядок.

3. ГИПЕРТЕКСТ И КВАЗИТЕКСТ

Последнее серьезное изменение, затронувшее структуру баз данных КонсультантПлюс, было связано с совмещением технологии динамического обновления и технологии гипертекста. Для воплощения этой идеи разработчики пошли своим путем, не прибегая к уже известным решениям. Дело в том, что гипертекст, в “классическом” своем исполнении, ориентирован на полную перегрузку базы данных при введении новых или изменении старых связей между текстовыми полями. Поэтому он становится не эффективным при стремительном увеличении объемов информации и частоты ее актуализации (что характерно, в частности, для российского законодательства).

Реализованный в программных продуктах КонсультантПлюс гипертекст выгодно отличается от “классического” следующими факторами:

с учетом специфики предметной области введен принцип разделения гипертекстовых связей на прямые и обратные (респонденты и корреспонденты), причем непосредственно отражаются только прямые, а обратные строятся динамически. Для этого реализован механизм BRG (BackReferenceGeneration), который автоматически генерирует обратные связи в момент актуализации базы данных. Такой принцип повышает надежность базы данных, исключает наличие в базе данных некорректных ссылок и уменьшает вероятность внесения ошибок при обработке документов;

изменение конкретных фрагментов гипертекста не оказывает влияния не только на остальные фрагменты, но и на сами тексты документов. Для этого реализован механизм DGI (DynamicGypertextInsert), который осуществляет динамический независимый от текста ввод и обновление прямых связей. Это эффективно при актуализации базы данных как по объему информации, так и по времени;

специально для нормативной базы введена типизация связей между документами с учетом специфики. Благодаря этому можно не только перейти по ссылке из документа в документ, но и определить форму юридической взаимосвязи.

Наряду с гипертекстовыми связями, формат базы КонсультантПлюс предусматривает возможность так называемых квазитекстовых вставок. Квазитекст представляет собой отдельное поле, которое при просмотре автоматически встраивается в текстовое поле на заданные места и по своим функциям ничем не отличается от текста, в частности, может содержать гипертекстовые связи. Технология квазитекстовых полей позволяет вводить специальную комментирующую информацию, не затрагивая текстов документов, тем самым существенно (в десятки раз) уменьшая трафик при замене конечного пользователя.