

SEMINAIRE

СЕМИНАР

SEMINAR

СТАТИСТИЧЕСКАЯ КОМИССИЯ и
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ
КОМИССИЯDistr.
GENERALCES/SEM.43/28
4 April 2000КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ
СТАТИСТИКОВ

RUSSIAN

ONLY

Семинар по интегрированным статистическим
информационным системам и связанным
с ними вопросам (ИСИС-2000)
(Рига, Латвия, 29-31 мая 2000 года)

Тема IV: Совершенствование стратегий распространения данных

Использование новых технологий для публикации статистической информации**Вспомогательный доклад**

Документ представлен Агентством Республики
Казахстан по статистике ¹

I. Введение

1. Одним из основных принципов совершенствования публикации и распространения статистической информации является поиск новых возможностей и методов, обеспечивающих быстрое действие и эффективность предоставления информации пользователям. С этой целью в информационном вычислительном центре агентства Республики Казахстан по статистике постоянно ведутся работы по освоению современных информационных технологий и на их основе внедряются новые методы электронной публикации информации.

2. Так в последние годы в ИВЦ была освоена технология создания и распространения информации на CD-ROM носителях, осуществлена публикация статистических данных в Интернете и выполнена основная часть работ по созданию системы хранения и управления документами, предназначенных для электронной публикации.

II. Опыт публикации статданных на CD-диске.

3. В Агентстве РК по статистике разработана информационно-поисковая система «Предприятия Казахстана», функционирующая на CD-ROMе и предназначенная для распространения в среде бизнеса. Информация, предоставляемая данной системой, может широко использоваться в деловой среде для быстрого поиска клиентов, партнеров по бизнесу, а также производителей определенного вида продукции, она нужна как предпринимателям малого и среднего бизнеса, так и руководителям крупных предприятий.

4. Основой информационно-поисковой системы «Предприятия Казахстана» является база данных, установленная на CD-ROM, включающая в себя информацию из Административного регистра юридических лиц. База данных содержит все ныне действующие и вновь зарегистрированные предприятия сектора нефинансовых корпораций (около 120000 предприятий). Информационно-поисковая система включает

¹ Авторы: А. Косьяненко, Т. Кучина

в себя программное обеспечение поиска данных, базу данных и ряд классификаторов, широко используемых в нашей стране, и созданных на основе международных стандартов и правил.

5. Система предоставляет пользователю несложную структуру меню и подменю, которая дает возможность осуществлять разнообразный поиск данных:

- Поиск партнеров-клиентов по выпускаемой продукции, возможность поиска по классификатору продукции КПЕС, по коду продукции или ключевому слову.
- Поиск партнеров-клиентов по виду деятельности, поиск по классификатору экономической деятельности ОКЭД и по коду вида деятельности.
- Поиск партнеров-клиентов по другим признакам. Этот режим включает в себя поиск по территории, размеру предприятия и дате регистрации.

6. После того как пользователь, воспользовавшись одним из режимов поиска, выбирает интересующие его предприятия, ему предлагается следующая информация о предприятии:

- ◆ Код предприятия, наименование предприятия, адрес, телефон;
- ◆ Основной вид деятельности, размер по численности, страны партнеры, дата регистрации.

7. По желанию может быть осуществлена печать выходного отчета. Кроме того, пользователь может полностью просмотреть весь список предприятий с кратким описанием его атрибутов.

8. После приобретения диска, пользователь устанавливает программное обеспечение с диска на свой компьютер, базу данных он не может скопировать на свой диск, поэтому при создании запроса идет постоянное обращение к диску. Данная система может использоваться практически на любом компьютере, работающем под управлением Windows95, Windows 98, Windows NT. Программное обеспечение разработано на языке системы разработки программно-технологических средств DELPHI 4.0.

9. Тем, кто подписался на получение диска на год, в течение года ежемесячно предоставляется на диске обновленная база данных. Ежемесячные изменения составляют примерно 10% от объема всей базы.

III. Опыт публикации статинформации в Интернете.

10. Одним из главных достоинств интернет-сетей является возможность электронной публикации документов с обеспечением быстрого и легкого доступа к ним пользователей. Специалистами ИВЦ Агентства РК по статистике разработан и внедрен Web-сайт «Статистика», включающий в себя информацию по всем отраслям статистики. На Web-сайте размещены все материалы, выпускаемые Агентством Республики Казахстан по статистике – ежегодники, сборники, справочники, журналы, бюллетени и экспресс-информация. Информация загружается в сайт согласно утвержденному графику, из каждого управления статистики в определенные сроки на Web-сайт поступают новые данные.

11. В качестве стандартов для метаданных, сопровождающих статистические данные, используется Генеральная система распространения данных (GDDS), которая является стандартом МВФ и способствует распространению надежных, полных и своевременных экономических, финансовых и социально-демографических данных.

12. В будущем, по мере необходимости, количество Web-сайтов будет увеличиваться, т.к. будут создаваться дополнительные специализированные сайты по определенной теме в каждом управлении Агентства РК по статистике.

13. Web-страница Агенства РК по статистике доступна пользователю как в сети **Intranet**, так и в сети **Internet**. Пользователю, получившему доступ к статданным, представленным на Web-странице в виде таблиц, позволяется скопировать таблицу в виде файла на свой компьютер, распечатать ее или отправить таблицу по электронной почте.

IV. Системы управления документами, предназначенных для публикации (СУДП)

14. В настоящее время в системе Агентства РК по статистике ведутся работы по созданию системы управления документами, подготовленных для публикации, которая включает в себя ведение крупного хранилища документов и их архивов.

15. Полнофункциональная СУДП позволяет не только организовывать сложные маршруты движения документов корпоративной сети, но и контролировать их подготовку, вести аудит, оптимизировать загрузку документов в хранилище и управлять публикацией документов.

16. Ядром такой системы является структурированное хранилище (с виртуальными тематическими подборками, перекрестными ссылками, аудитом работ с документами), обладающее малым временем доступа к документу, снабженное развитыми средствами поиска и публикации документов, поддерживающее работу с нетрадиционными носителями информации. Последнее важно, как для организации публикации документов из хранилища документов, так и для создания крупных архивов, которые экономически невыгодно держать на жестких дисках, поэтому для этих целей будут использоваться магнитооптические библиотеки, CD-ROM, стримеры.

17. В централизованное хранилище будут поступать только те документы, которые предназначены для публикаций. При этом будет возможность организовать автоматическую публикацию документов из хранилища в Интернете, на CD-ROM и дискеты или автоматическую отправку определенных документов по электронной почте и т.д.

18. Одной из главных задач такой системы является проблема организации данных и метаданных таким образом, чтобы они удовлетворяли требованиям конечных пользователей. Такая система должна обладать гибкостью с тем, чтобы в случае необходимости можно было осуществить ее реорганизацию и обновление документов. Предварительно требуется тщательно изучить потребности внешних и внутренних клиентов с тем, чтобы система могла удовлетворить любые запросы пользователей.

19. В настоящее время в ИВЦ ведется работа по внедрению такого проекта. На первом этапе были выполнены работы по подготовке основных каталогов и стандартов, необходимых для создания СУДП (системы управления документами, предназначенных для публикаций). Прежде всего были определены главные требования к публикуемым документам:

- ◆ Разработка системы каталогов документов;
- ◆ Определение системы стандартов документов, предназначенных для загрузки в хранилище документов;
- ◆ Описание процедуры подготовки документов (в т.ч. групповой или коллективной), предназначенных для публикации;
- ◆ Определение стандарта метаданных, сопровождающих статистические данные.

20. Далее была определена схема движения документов и выполнено описание основных процедур тестирования, согласования и преобразования документов в стандартный формат для загрузки в хранилище документов. Описание последовательности и содержания процедур:

- (i) Определение и описание возможных источников документов. В первую очередь таковыми являются сгенерированные отчеты, полученные в результате отработки программного комплекса сбора и обработки первичной статистической отчетности. Данные отчеты формируются в виде документов, предназначенных для публикации;
- (ii) Определение стадий дополнительной обработки полученных документов;
- (iii) Тестирование документов, предназначенных для загрузки в хранилище документов;
- (iv) Включение этапов согласования и утверждения документов;
- (v) Процесс загрузки документов в хранилище;
- (vi) Организация публикации документов;
- (vii) Процедура передачи документов в архив.

21. На следующем этапе были определены требования к создаваемому хранилищу документов:

- ◆ Проведение мониторинга среди пользователей (какие серии документов пользуются наибольшим спросом, в каком виде они должны храниться (таблицы, графики, диаграммы и пр.)).
- ◆ Определение структуры хранилища данных.
- ◆ Описание структуры репозитория метаданных, содержащего описание всех документов в хранилище документов.
- ◆ Определение схемы формирования хранилища данных (описание очередности загрузки потоков документов).
- ◆ Описание процедур предоставления документов внутренним пользователям (описание методов доступа и поиска различных документов);
- ◆ Описание процесса архивизации документов.
- ◆ Описание процедур публикации документов для внешних и внутренних клиентов.
- ◆ Описание процедур дальнейшего использования документов (публикация, передача в архив).
- ◆ Определение системы ограничений и контроля доступа к документам.

22. В дальнейшем осуществляется непосредственно создание и внедрение основных подсистем СУДП (системы управления документов, предназначенных для публикаций) и приложений, необходимых для ее функционирования:

- ◆ Проектирование репозитория метаданных, в котором хранятся каталоги документов, ссылки на документы и пути доступа к документам;
- ◆ Настройка ПО промежуточного слоя, обеспечивающего доступ к данным и поддержку распределенного режима работы (использование технологий Internet/Intranet), решение проблем администрирования, резервного копирования, восстановления и защиты данных.
- ◆ Внедрение методов интеграции СУДП (системы управления документов, предназначенных для публикаций) с офисными и прикладными программами;
- ◆ Разработка клиентских приложений публикации документов в Интернете, Интранете, на CD-ROM, дискетах, принтерах.
- ◆ Внедрение средств автоматизации процессов координации действий сотрудников, ведущих подготовку документов, с помощью встроенных средств автоматизации деловых процессов.
- ◆ Внедрение средств шифрования и электронно-цифровой подписи;
- ◆ Утверждение графика загрузки документов с указанием сроков и последовательности процессов загрузки, определение приоритетности источников загрузки.

23. В ходе работы приходится решать ряд проблем, связанных с выбором программных средств, их управлением, настройкой и внедрением. Перечислим некоторые из них:

Проблемы администрирования

24. Одной из существенных проблем, выявленных при внедрении СУДП (системы управления документов, предназначенных для публикаций) являются сложность текущего сопровождения и администрирования хранилища. Система управления СУДП должны включать в себя средства, помогающие администратору решать самые важные задачи, в том числе создавать и копировать документы, производить настройку тиражирования, импорт/экспорт данных, управлять правами пользователей и т.п.

Проблема извлечения документов из различных источников.

25. Организация извлечения документов из различных оперативных баз документов, полученных в результате работы программного комплекса сбора и обработки первичной статистической отчетности, их тестирования, интеграции и помещения в хранилище может потребовать значительных усилий и затрат, сбоев в периодичности поставки документов по причине их неготовности и т.д. При этом должна быть предусмотрена возможность извлекать и преобразовывать данные практически из

любых источников. Задания по процессу переноса и преобразованию документов, могут включать в себя множество последовательных шагов, поэтому они должны быть оформлены в виде пакетов, которые могут выполняться автоматически по расписанию.

Выполнение запросов пользователей

26. Должна быть предусмотрена улучшенная обработка запросов, обеспечивающая оптимизацию и эффективное выполнение сложных запросов, позволяющих извлекать различные документы из хранилища документов.

Организация хранилища метаданных

27. Большое внимание должно быть уделено одному из важных компонентов СУДП каким является хранилище метаданных. Здесь хранится разнообразная информация, в которой содержатся данные об источнике, приемнике, схеме преобразования и перемещения документов, о компонентах программного обеспечения и об их отношениях, а также описательная информация о системе каталогов документов, системе стандартов метаданных и так далее. Данная информация должна быть максимально структурирована и доступна всем.

28. Таким образом, создание СУДП позволит хранить статистические данные в электронном виде, легко управлять процессами пополнения, обновления, архивирования документов, в частности, осуществлять проверку качества поступающих данных и своевременности их поступления, а также осуществлять управление распространением данных через различные каналы и средства связи. Пользуясь такой системой можно легко и уверенно осуществлять любую публикацию документов в электронном виде.