

**ТЕХНОЛОГИЯ  
ФОРМИРОВАНИЯ БАЗ  
СОЦИАЛЬНО-  
ЭКОНОМИЧЕСКИХ ДАННЫХ**

# ПОНЯТИЕ «БАЗА ДАННЫХ»

- Базой данных является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины

# Особенности социально-экономических баз данных

- К социальным данным относятся данные о населении и социальной среде. Сведения о населении включают половые, социальные, медицинские и любые другие персональные данные об отдельных личностях, а также сводные данные о населении страны в целом и ее отдельных территориях и об определенных группах населения: пенсионерах, квартиросъемщиках, детях школьного возраста, женщинах и др.

# СОСТАВ БАЗ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ДАННЫХ

- Сведения о социальной среде включают данные о вакансиях, градостроительстве, городском хозяйстве, городском пассажирском транспорте, законах, нарушениях общественного порядка и др.
- К социальным фактографическим данным отнесены также другие данные сферы здравоохранения: о заболеваниях и их последствиях, лекарственных средствах, медицинском оборудовании и медицинских учреждениях и т.д.
- Данные о градостроительстве включают сведения о зданиях и сооружениях, дорогах, зеленых насаждениях и транспортных маршрутах города.
- Данные о городском хозяйстве содержат сведения о жилом фонде, тепло и энергосети, водопроводной сети, электро и автотранспорте, предприятиях, занятых переработкой отходов, ремонтом дорог и другими функциями коммунального хозяйства.

# Продолжение слайда

- Подавляющее большинство баз социальных данных формируется в рамках систем местных органов власти. Источниками таких данных являются, как правило, формализованные документы.
- Потребителями баз социальных данных являются население, органы управления различных уровней и сфер, начиная от жилищно-эксплуатационных контор и управлений и кончая Президентом и Федеральным Собранием, а также научные работники.
- Все данные о личностях имеют ограниченный доступ, их содержание контролируется населением, о чем во многих государствах были изданы законы.
- Базы данных о градостроительстве и городском хозяйстве создаются за рубежом при городских муниципалитетах и службах. Сведения о жилом и нежилом фонде используются налоговыми органами.
- Информация о зданиях, дорогах, зеленых насаждениях и транспорте предназначена для работников служб градостроительства и экологических служб, информация о водо-, тепло-, энергосетях города - для работников служб, управляющих этими сетями.
- В нашей стране создание баз демографических и паспортных данных о населении находится в начальной стадии.

# ПРИЗНАКИ БАЗЫ ДАННЫХ

- База данных хранится и обрабатывается в вычислительной системе. Таким образом, любые внекомпьютерные хранилища информации (архивы, библиотеки, картотеки и т. п.) базами данных не являются.
- Данные в базе данных логически структурированы (систематизированы) с целью обеспечения возможности их эффективного поиска и обработки в вычислительной системе.
- Структурированность подразумевает явное выделение составных частей (элементов), связей между ними, а также типизацию элементов и связей, при которой с типом элемента (связи) соотносится определённая семантика и допустимые операции.

# Продолжение слайда

- База данных включает метаданные, описывающие логическую структуру БД в формальном виде (в соответствии с некоторой метамodelью).
- В соответствии с ГОСТ Р ИСО МЭК ТО 10032-2007, «постоянные данные в среде базы данных включают в себя схему и базу данных. Схема включает в себя описания содержания, структуры и ограничений целостности, используемые для создания и поддержки базы данных. База данных включает в себя набор постоянных данных, определенных с помощью схемы. Система управления данными использует определения данных в схеме для обеспечения доступа и управления доступом к данным в базе данных

# РАБОТА НАД БАЗОЙ ДАННЫХ

- Работа над базой данных включает комплекс технических процедур и исследовательские задачи.
- Как правило, исходный текст статистических сборников и аналитических материалов поступает на обработку в виде текстовых файлов формата Microsoft Word, либо загружается с сайта информационного провайдера как совокупность HTML-страниц.
- Процесс приведения массивов к единому формату и интеграции в базу данных автоматизирован.

# ЭТАП 1

- На первом этапе обработки документ-первоисточник, прошедший экспертную проверку, преобразуется в формат HTML.

## ЭТАП 2

- На втором этапе сохраненный HTML-документ автоматически разделяется на самостоятельные смысловые блоки (оглавление, таблицы, графические материалы, сноски, методологические пояснения) с выделением их поисковых атрибутов и добавлением элементов пользовательского сервиса, таких как гиперссылки на данные в формате MS Excel, аннотацию документа, методологию, термины глоссария, а также обратные ссылки к оглавлению раздела. Для этих целей разработана библиотека программ-конверторов, настроенных с учетом особенностей оформления каждой группы первоисточников.

# ЭТАП 3

- Третий этап обработки предполагает автоматизированную проверку орфографии, необходимую для правильного формирования поискового индекса. После этого каждая таблица дополнительно конвертируется в формат MS Excel, материалы интегрируются в документографическую базу и поисковую систему (четвертый этап). В процессе обработки данные проходят несколько этапов экспертной проверки.

# ПРИНЦИПЫ ПОСТРОЕНИЯ БАЗ ДАННЫХ

- 1. Генератором базы является, как правило, организация-владелец документов-источников, которая может передавать функции распространения информации другим БнД.
- 2. Создание баз социальных данных необходимо производить на основе централизованного частичного финансирования при условии финансовой поддержки этих работ федеральными и местными органами управления. В первую очередь следует финансово поддержать эти работы в опорных зонах информатизации.
- 3. Для обеспечения унификации записей в базах данных и запросов целесообразно создать систему баз данных в социальной сфере, совмещенную с системой словарно-справочных служб.

# ПРОДОЛЖЕНИЕ СЛАЙДА

- 4. Основное внимание следует уделить созданию баз детальных данных об элементарных объектах, таких как отдельный человек, конкретное лекарство, конкретная квартира. Эти базы следует создавать в местах их наиболее интенсивного использования. Так, базы данных о населении следует создавать в ДЭЗах или РЭУ, отделениях милиции, центральном адресном бюро. Базу данных о квартирах и строениях следует создавать в бюро технической инвентаризации и в ДЭЗах. Сводные данные о жителях и квартирах следует хранить в базах данных районных и городских органов управления. Допускается дублирование данных при хранении, если это оправдано экономически.
- 5. В фактографических базах социальных данных велика интенсивность их обновления (15-25% в год). Вследствие этого каналы актуализации должны тщательно отработаться до начала загрузки базы данных.

# ПРИМЕР ПОСТРОЕНИЯ БАЗЫ ДАННЫХ

*«ВЫБОРЫ В  
ГОРОДСКУЮ ДУМУ»*

# ЛИСТОВКИ – КАНАЛ ПЕРЕДАЧИ ИНФОРМАЦИИ

- Одним из каналов передачи информации от кандидатов к избирателям являются печатные агитационные материалы.
- Самые распространенные из них - это листовки, где в сжатой и доступной форме приводится позиция кандидата.
- Изучение листовок как массового источника становится возможным при создании специализированной базы данных (БД). В данной работе приведены результаты формирования такой базы на основе листовок кандидатов в депутаты Тверской городской думы, выборы которых прошли в ноябре 1998 г. в г. Тверь. Было обработано 186 единиц наблюдения по 121 кандидатам.

# ЭТАПЫ ФОРМИРОВАНИЯ БАЗЫ ДАННЫХ

- При выделении градаций признаков использовались три метода измерения признаков: естественное упорядочивание, по частоте встречаемости ответов, альтернативное разбиение.
- База была разбита на три смысловых части. В первую занесены личные данные кандидатов (фамилия, имя, отчество, пол, возраст, место рождения, национальность, социальное происхождение, образование, семейное положение, место работы, служба в армии и партийность).
- Во второй части отмечены пункты программы каждого кандидата (создание новых рабочих мест, обещание постоянно встречаться со своими избирателями, организация досуга, вопросы спорта, улучшение работы ЖЭУ, проблемы молодежи, проблемы семьи, вопросы местного самоуправления, работа общественного транспорта, помощь социально незащищенным категориям населения, благоустройство города, своевременные выплаты денежных пособий, поддержка предприятий города, проблема преступности, контроль администрации города, вопросы налоговой политики, критика власти и др.).
- В третьей части зафиксированы сведения, связанные с выходными данными печатных материалов (тираж, дата выхода, цветность, наличие фотографии, форма листовки (расположение текста)), а также информация об избирательном округе (номер избирательного округа, территория и номера избирательных участков, наличие даты проведения выборов).
- В результате, сформированная БД "Выборы в городскую думу" включает в себя 70 полей и 121 запись.

# Литература

- Афанасьев В.Г. Социальная информация. М.: Наука, 1994.
- Дементьев В.В. О методике создания базы данных "выборы в городскую думу" // <http://www.hist.msu.ru/Association/НАС/aik/polit.htm>
- Иванова Е., Забелин В., Божко В. Особенности построения статистических информационных систем в государствах с рыночной экономикой. // Вопросы статистики. – 1998. – №8.
- Информатика: Учебник для вузов/ под ред. проф. Макаровой. - М.: Финансы и статистика, 2001.
- Семенов М.И. Трубилин И.Т. и др. Автоматизированные информационные технологии в экономике. М.: Финансы и статистика, 2000.