

Лабораторная работа № 1
EXCEL – УНИВЕРСАЛЬНАЯ СИСТЕМА ОБРАБОТКИ ДАННЫХ

Целью работы является изучение возможностей прикладного программного обеспечения для организации данных и их анализа.

1. Теоретическое введение

Базы данных как средство хранения и обработки информации используются повсеместно. Благодаря специальным алгоритмам хранения и поиска информации базы данных позволяют быстро получать необходимую информацию, анализировать ее, формировать отчеты. Офисный пакет MS Office имеет в своем составе специальное приложение для создания больших, со сложной структурой, баз данных – Access, однако для создания небольших баз данных вполне можно использовать Excel, поскольку это универсальная система обработки данных, позволяющая сортировать, фильтровать, систематизировать и структурировать информацию. К тому же это и мощный аналитический инструмент с прекрасным набором функций для работы с различными данными, выполнения сложных расчетов и визуализации.

Анализ предметной области – это первый шаг этапа системного анализа, с которого начинается разработка программной системы. Разработчик должен научиться:

- понимать язык, на котором говорят заказчики;
- выявить цели их деятельности;
- определить набор решаемых ими задач;
- определить набор сущностей, с которыми приходится иметь дело при решении этих задач.

В ходе этой лабораторной работы не стоит задача полноценного освоения всех этапов проектирования и построения БД, однако даже для создания учебной БД необходимо понимание целей и задач обработки информации в рамках выбранной предметной области.

Инфологическое моделирование

Предметная область – часть реального мира, подлежащая изучению с целью организации управления и, в конечном итоге, автоматизации.

База данных (БД) – это средство накопления и организации больших массивов информации об объектах некоторой предметной области. Она должна отражать текущие данные о предметной области, накапливать, хранить информацию и предоставлять различным категориям пользователей быстрый доступ к данным.

Информационная система (ИС) – это совокупность тем или иным способом структурированных данных (базы данных) и комплекса аппаратно-программных средств для хранения данных и манипулирования ими.

Специфика предметной области может оказывать существенное влияние на характер функционирования проектируемой информационной системы, выбор метода представления знаний, способов рассуждения о знаниях, и т. д. Предметная область сильно влияет на все аспекты проекта: требования к системе, взаимодействие с пользователем, модель хранения данных, реализацию и т.д.

Анализ предметной области, позволяет выделить ее сущности, определить первоначальные требования к функциональности и определить границы проекта.

Анализ предметной области при создании полноценной информационной системы разбивается на три фазы:

- анализ концептуальных требований и информационных потребностей;
- выявление информационных объектов и связей между ними;
- построение концептуальной модели предметной области и проектирование концептуальной схемы ИС.

Первая фаза анализа предметной области представляют собой список вопросов, которые разработчик получает в диалоге с будущими пользователями этой ИС.

Вторая фаза анализа предметной области состоит в выборе информационных объектов, задании необходимых свойств для каждого объекта, выявлении связей между объектами, определении ограничений, накладываемых на информационные объекты, типы связей между ними, характеристики информационных объектов.

Заключительная фаза анализа предметной области состоит в проектировании ее информационной структуры или концептуальной модели. Концептуальная модель

включает описания объектов и их взаимосвязей, выявляемых в результате анализа данных. Концептуальная модель применяется для структурирования предметной области с учетом информационных интересов пользователей системы. Она является представлением точки зрения пользователя на предметную область и должна быть стабильной т.е. неизменной. Хотя возможно ее наращивание и включение дополнительных данных.

Инфологическая модель (информационно-логическая модель) – ориентированная на человека и не зависящая от типа СУБД модель предметной области, определяющая совокупности информационных объектов, их атрибутов и отношений между объектами, динамику изменений предметной области, а также характер информационных потребностей пользователей.

Основные подходы к созданию инфологической модели предметной области:

- **Функциональный** подход к проектированию БД. Этот метод реализует принцип "от задач" и применяется тогда, когда известны функции некоторой группы лиц и/или комплекса задач, для обслуживания информационных потребностей которых создаётся рассматриваемая БД.
- **Предметный** подход к проектированию БД. Предметный подход к проектированию БД применяется в тех случаях, когда у разработчиков есть чёткое представление о самой предметной области и о том, какую именно информацию они хотели бы хранить в БД, а структура запросов не определена или определена не полностью. Тогда основное внимание уделяется исследованию ПО и наиболее адекватному её отображению в БД с учётом самого широкого спектра информационных запросов к ней.
- Проектирование с использованием метода "сущность-связь". Метод "сущность-связь" (entity–relation, ER–method) является комбинацией двух предыдущих и обладает достоинствами обоих. Этап инфологического проектирования начинается с моделирования предметной области. Проектировщик разбивает её на ряд локальных областей, каждая из которых (в идеале) включает в себя информацию, достаточную для обеспечения запросов отдельной группы будущих пользователей или решения отдельной задачи (подзадачи). Каждое локальное представление моделируется отдельно, затем они объединяются.

Основными составными элементами инфологической модели являются сущности (информационные объекты), связи между ними и их атрибуты (свойства).

Сущность – любой различимый объект (объект, который мы можем отличить от другого), информацию о котором необходимо хранить в базе данных. Сущностями могут быть люди, места, самолеты, рейсы, вкус, цвет и т.д. Необходимо различать такие понятия, как тип сущности и экземпляр сущности. Понятие тип сущности относится к набору однородных личностей, предметов, событий или идей, выступающих как целое. Экземпляр сущности относится к конкретной вещи в наборе. Например, типом сущности может быть ГОРОД, а экземпляром – Москва, Киев и т.д.

Атрибут – поименованная характеристика сущности. Его наименование должно быть уникальным для конкретного типа сущности, но может быть одинаковым для различного типа сущностей (например, ЦВЕТ может быть определен для многих сущностей: СОБАКА, АВТОМОБИЛЬ, ДЫМ и т.д.). Атрибуты используются для определения того, какая информация должна быть собрана о сущности. Примерами атрибутов для сущности АВТОМОБИЛЬ являются ТИП, МАРКА, НОМЕРНОЙ ЗНАК, ЦВЕТ и т.д. Здесь также существует различие между типом и экземпляром. Тип атрибута ЦВЕТ имеет много экземпляров или значений: Красный, Синий, Банановый, Белая ночь и т.д., однако, каждому экземпляру сущности присваивается только одно значение атрибута.

Абсолютное различие между типами сущностей и атрибутами отсутствует. Атрибут является таковым только в связи с типом сущности. В другом контексте атрибут может выступать как самостоятельная сущность. Например, для автомобильного завода цвет – это только атрибут продукта производства, а для лакокрасочной фабрики цвет – тип сущности.

Ключ – минимальный набор атрибутов, по значениям которых можно однозначно найти требуемый экземпляр сущности. Минимальность означает, что исключение из набора любого атрибута не позволяет идентифицировать сущность по оставшимся. Для сущности Расписание ключом является атрибут Номер_рейса или набор: Пункт_отправления, Время_вылета и Пункт_назначения (при условии, что из пункта в пункт вылетает в каждый момент времени один самолет).

Связь – ассоциирование двух или более сущностей. Если бы назначением базы данных было только хранение отдельных, не связанных между собой данных, то ее структура могла бы быть очень простой. Однако одно из основных требований к организации базы данных – это обеспечение возможности отыскания одних сущностей по значениям других, для чего необходимо установить между ними определенные связи. А так как в реальных базах данных нередко содержатся сотни или даже тысячи сущностей,

то теоретически между ними может быть установлено более миллиона связей. Наличие такого множества связей и определяет сложность инфологических моделей.

Связь один к одному (1:1) предполагает, что в каждый момент времени одному экземпляру информационного объекта А соответствует не более одного экземпляра информационного объекта В и наоборот. Такие информационные объекты, легко объединяются в один объект, ключевым становится ключевое поле любого исходного объекта.

При связи один ко многим (1:M) одному экземпляру информационного объекта А соответствует 0, 1 или более экземпляров объекта В, но каждый экземпляр объекта В связан не более чем с 1 экземпляром объекта А. В такой связи имеют место иерархические групповые отношения между экземплярами разных типов. При этом один информационный объект определяется как главный, а другой – как подчиненный.

Связь многие ко многим (M:M) предполагает, что в каждый момент времени одному экземпляру информационного объекта А соответствует 0, 1 или более экземпляров объекта В и наоборот. Такие отношения можно охарактеризовать как сетевые.

В рамках лабораторных работ по курсу ТОИ для построения инфологической модели будет использоваться предметный подход без формализации классическими методами, т.к. в простых случаях описание предметной области допускается на естественном языке с поддержкой любыми средствами визуализации, но при условии однозначной интерпретации информации.

В процессе выбора предметной области и формирования «легенды» рекомендуется ориентироваться на бизнес-процессы, реализуемые в рамках OLTP (Online Transaction Processing)-систем. **OLTP**-системы используются для ввода, первичного контроля, структурированного хранения и обработки информации (операций, документов) в режиме реального времени.

Транзакция (англ. transaction, от лат. transactio – соглашение, договор) – минимальная логически осмысленная операция, которая имеет смысл и может быть совершена только полностью.

Создание БД в Excel

Реляционная база данных – база данных, основанная на реляционной модели данных. Это совокупность взаимосвязанных таблиц, каждая из которых содержит информацию об объектах определенного типа. Строка таблицы содержит данные об одном объекте (например, товаре, клиенте), а столбцы таблицы описывают различные характеристики этих объектов – атрибутов (например, наименование, код товара, сведения о клиенте). Записи, т. е. строки таблицы, имеют одинаковую структуру – они состоят из полей, хранящих атрибуты объекта. Каждое поле, т. е. столбец, описывает только одну характеристику объекта и имеет строго определенный тип данных. Все записи имеют одни и те же поля, только в них отображаются различные информационные свойства объекта.

Обычная двухмерная таблица Excel, созданная с соблюдением определенных правил так же является таблицей базы данных. Столбцы таблицы Excel – это поля базы данных, а строки – это записи базы данных. Поле (столбец) содержит информацию об одном признаке для всех записей базы данных. Запись (строка) состоит из нескольких (по числу полей) разнообразных информационных сообщений (признаков), характеризующих один объект базы данных. Иногда БД в Excel называют списком данных.

Расширяемая база данных – это таблица, в которую постоянно добавляются новые записи (строки) информации. При этом названия и количество полей (столбцов) остаются неизменными.

Правила формирования БД:

- Область для размещения данных должна иметь прямоугольную форму.
- Первая строка БД должна содержать названия полей, описывающие содержимое соответствующего столбца. В БД не должно быть одинаковых названий полей.
- Каждый столбец (поле) должен содержать данные одного типа (поля в Excel могут содержать данные одного из следующих типов – числовой, текстовый, дата/время, логический, гиперссылка, вычисляемый).

- Каждая последующая строка базы данных должна содержать хотя бы одну заполненную данными ячейку, т.е. список не должен содержать пустых строк и столбцов.
- Использовать объединение ячеек внутри списка недопустимо.
- Вокруг базы данных ячейки должны быть пустыми, либо это должен быть край рабочего пространства листа. То есть, если на листе кроме БД есть и другие данные, то БД должна быть отделена от них, по крайней мере, одной пустой строкой и одним пустым столбцом.
- Диапазону с данными, определяющему БД, желательно присвоить имя и объявить его таблицей. (Главная | Форматировать как таблицу)

2. Порядок выполнения работы

1. Определите предметную область, проблему, для решения которой необходим анализ данных средствами MS Office.
2. Определите информационные объекты, их атрибуты, типы значений, связи.
3. Предложите визуальное описание предметной области.
4. На основе разработанной инфологической модели сформируйте структуру списка, определив, что будет отображать запись, укажите названия полей.
5. Результаты выполнения пунктов 1 – 5 занесите в Отчет.
6. Оформите список в соответствии с правилами формирования БД.
7. Подумайте, будет ли в вашей БД информация, которую было бы проще вводить используя раскрывающиеся списки? Это не только ускоряет работу по наполнению БД, но и является средством контроля вводимых данных.

имя ▾	пол ▾
Анна	ж
Саша	м
Женя	м
Валя	ж
	м

8. Заранее создайте на отдельном листе диапазоны данных для таких списков, скопировав названия соответствующих полей и записав под ними нужные значения.

9. Для того, чтобы данные из этих диапазонов стали доступны в формулах для выбора им нужно присвоить имена. Для этого выделите диапазон, вызовите контекстное меню и выберите «Имя диапазона...»
10. Чтобы не изменять диапазон в формуле выбора попробуйте создать динамически именованные диапазоны (функция СМЕЩ).
11. Для подключения раскрывающихся списков в основной таблице БД находясь в ячейке, где ввод должен осуществляться из фиксированного числа значений, выберите Данные | Работа с данными | Проверка данных | вкладка Параметры – тип данных: Список, источник данных: = имя диапазона.
12. Заполните БД. Внесите не менее 30 записей. Все записи должны семантически соответствовать предметной области.
13. Обратите внимание на подписи листов, название книги. Убедитесь, что имена соответствуют содержанию.
14. Выберите из БД записи, удовлетворяющие каким-нибудь двум условиям. Скопируйте выборку на новый лист. Дайте листу имя. Вставьте на этот лист описание условий формирования выборки. (Принт-скрин, Надпись, SmartArt)
15. На листе с основной БД познакомьтесь с возможностями Автофильтра, Расширенного фильтра, Настраиваемой сортировкой. Результаты работы поместите на отдельные листы, дав им имя, и так же как в п.14 укажите условия формирования выборки.

3. Контрольные вопросы

1. Дайте определение информационному объекту.
2. Что подразумевается под связями между объектами, какие типы связей бывают?
3. Что такое ИС и БД?
4. Может ли одно и то же понятие выступать как сущностью, так и атрибутом?
5. Зачем присваивается имя диапазону с данными?
6. Как создать динамически именованный диапазон? Для чего он создается?

4. Литература

1. Карпова, И.П. Базы данных: Учебное пособие / И.П. Карпова. - СПб.: Питер, 2013. - 240 с.
2. Материалы информационного портала <http://al-vo.ru>
3. Материалы информационного портала <http://excel2.ru>