

Лекция № 7

Информационные технологии.

Модели процессов извлечения, обработки данных, хранения, представления и использования в информационных системах

*Кафедра прикладной
математики и кибернетики
СибГУТИ*

Юрий Иванович Молородов

yumo@ict.sbras.ru

Содержание

0. Замечания о новой CMS.

<http://docs.jelastic.com/calipso>

1. Извлечение информации.

2. Обработка информации.

3. Хранение информации.

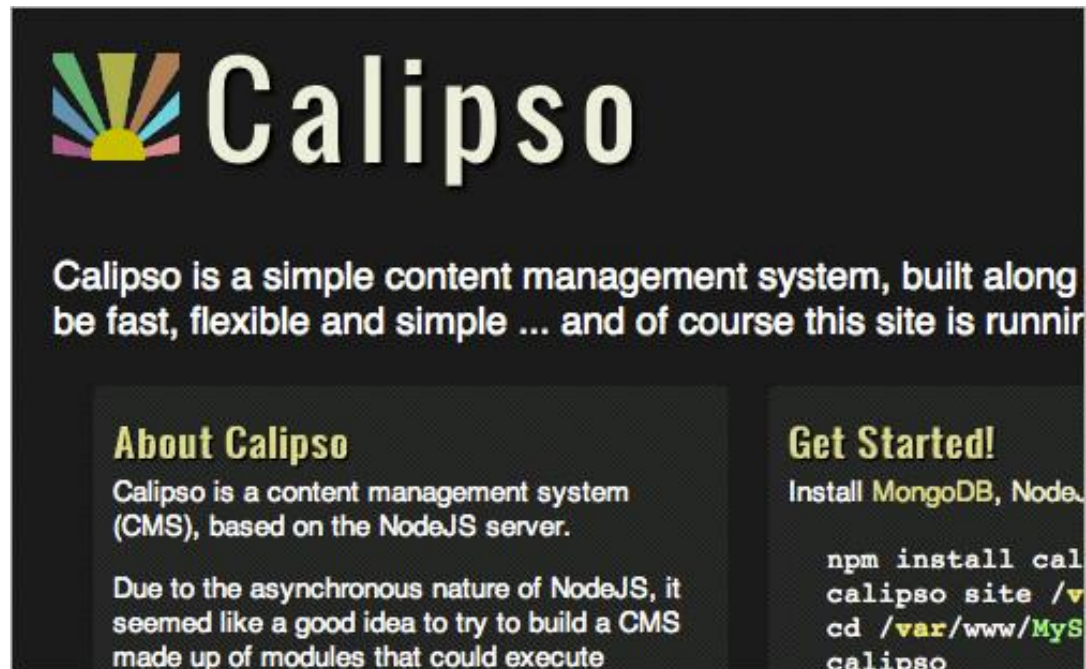
4. Представление и использование информации.

Calipso: экзотическая CMS на новомодных технологиях

<http://habrahabr.ru/post/128601/>

Система CMME (Content Management Made Easy) - система управления контентом с открытым исходным кодом, простая в использовании, не имеет больших системных требований с сохранением гибкости.

CMME написана на PHP5 и работает на файлах (не нуждается в базе данных).



Calipso: экзотическая CMS на новомодных технологиях

Система, написана на серверном JavaScript (для node.js).

Пригодна для виртуального хоста в минимальной конфигурации.

Отсутствует PHP. Calipso хранит свои данные не в MySQL, а в MongoDB. Это способствует минимальности, скорострельности и простоте обслуживания.

Перспективность оценил создатель Wordpress Мэтт Мюлленберг.

«Я смотрел новую, использующую node.js, CMS называется Calipso и одну на Ruby, называется Jekyll. Неплохие системы.»

Calipso поддерживает все необходимые функции, которых админы ожидают от CMS: регистрация пользователей, управление контентом, смена тем, кроссбраузерность.



Тип	документо-ориентированная СУБД
Разработчик	10gen
Написана на	C++
Операционная система	кроссплатформенное программное обеспечение
Первый выпуск	2009
Последняя версия	2.6.3 (19 июня 2014) ^[1]
Лицензия	GNU AGPL (СУБД) и Apache License (драйверы)
Сайт	mongodb.org

MongoDB (*humongous* — огромный) — документо-ориентированная система управления базами данных (СУБД) с открытым исходным кодом, не требующая описания схемы таблиц. Написана на языке C++. Управляет наборами JSON-подобных документов, хранимых в двоичном виде в формате BSON. Хранение и поиск файлов в MongoDB происходит благодаря вызовам протокола GridFS. Подобно другим документо-ориентированным СУБД, MongoDB не является реляционной СУБД.

Calipso: экзотическая CMS на новомодных технологиях

Node.js - создана Райаном Далем в 2009 году.

Относительно новая технология, завоевавшая в последнее время большую популярность среди веб-разработчиков.

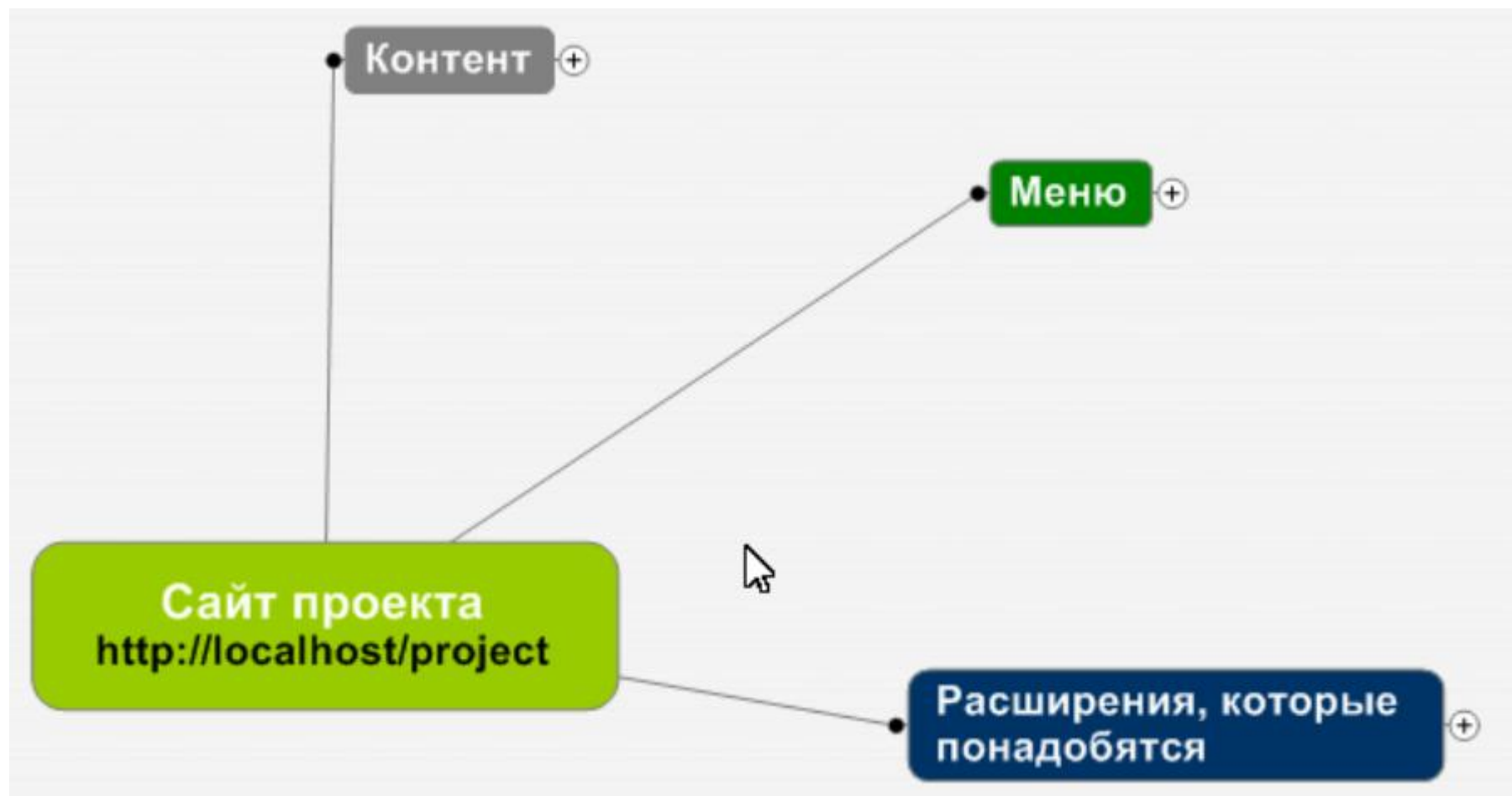
Не все знают, что на самом деле Node.js это серверная JavaScript среда, использующая асинхронную событийную модель.

Это среда, которая предназначена для написания масштабируемых, высокопроизводительных приложений. Почти как Ruby's Event Machine или Python's Twisted, но гораздо глубже — JavaScript реализовывает цикл событий, на уровне самого языка, а не как отдельная библиотека.

Главное преимущество этой системы — высочайшая модульность. Немало модулей уже разработано участниками сообщества. При желании научиться писать собственные модули не сложно - имеется простой шаблон.

У Calipso есть и немало недостатков, большинство из которых связано с тем, что ее развитие несколько месяцев назад по каким-то причинам прекратилось. Во-первых, крайне мало документации. Этот недостаток можно (хотя и не без труда) восполнить просто анализируя код. Во-вторых, устанавливать Calipso следует на несколько устаревшую версию node.js, а именно на 0.4.x. Все попытки установить эту CMS поверх node 0.6.x или хотя бы 0.5.x не удались.

Однако, его шаблоны весьма просто устроены, их легко править вручную.



О продукте Ссылка Отзывы Контакты

SYREX DUMPER -

БЫСТРЫЙ И УДОБНЫЙ БЕКАП

Рассылка с подарками
Ваша имя на русском:
Ваш Email адрес:
Подписаться

Главная меню

- О продукте
- Возможности
- Сравнение
- Технические требования
- Скачать
- Документация
- Отзывы
- Вопросы и ответы
- История версий
- Спонсоры

Новости

Обновление бэкапа
Syrex Dumper 1.0.8

Возможности

Изначально Syrex Dumper создавался для работы с большими базами данных (сотни мегабайт для него не проблема), при этом он должен был быть максимально быстрым, компактным и удобным.

Основные возможности:

- создание резервной копии и восстановление базы данных MySQL без использования сторонних программ, таких как mysqldumper;
- работа с базами данных любых размеров (от нескольких килобайт до сотен мегабайт), в связи с этим вся работа с файлами бэкапа осуществляется по FTP, но download возможен и с помощью менеджера загрузки (Download Manager, Kget и др.);
- поддержка двух форматов сжатия файлов* (Gzip и Bzip2), а также разной степени сжатия;
- поддержка фильтров для таблиц, с их помощью легко можно выбрать нужные таблицы;
- работает на Windows и Linux;
- высокая скорость работы (в 2 раза быстрее, чем phpMyAdmin**);
- оригинальный алгоритм парсинга данных благодаря которому дампер обгоняет даже некоторые mysql-приложения, не говоря уже о скриптах (в 4 раза быстрее, чем phpMyAdmin**);
- индикатор выполнения задачи***, благодаря чему всегда видно на каком этапе работы находится скрипт;
- очень компактный (всего 25 КБ);
- легко настраивается;
- сохраняются последние настройки, что удобно при частом использовании;
- имеется собственная система авторизации, несколько вариантов авторизации, в том числе пользовательская;
- небольшой расход памяти при работе;
- совместимость с дампами phpMyAdmin и mysqldumper;

* — если соответствующие PHP-модули установлены;
** — данные из нашего сравнительного тестирования, с результатами которого можно ознакомиться здесь;
*** — при восстановлении БД прогресс показывается только для своих файлов, т.е. файлов созданных дампером;

Интер сайта: 4100, Все права защищены (с)2008.



Шаблон

Извлечение информации

Объекты и их свойства, процессы и функции, выполняемые этими объектами или для них являются источниками данных в любой предметной области.

Любая предметная область рассматривается в виде трех представлений:

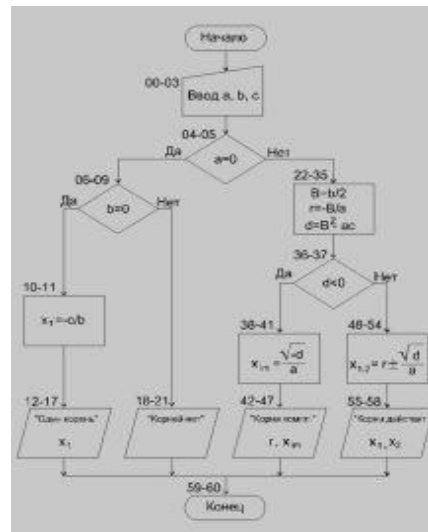
- реальное представление предметной области;
- формальное представление предметной области;
- информационное представление предметной области.

Извлечение информации

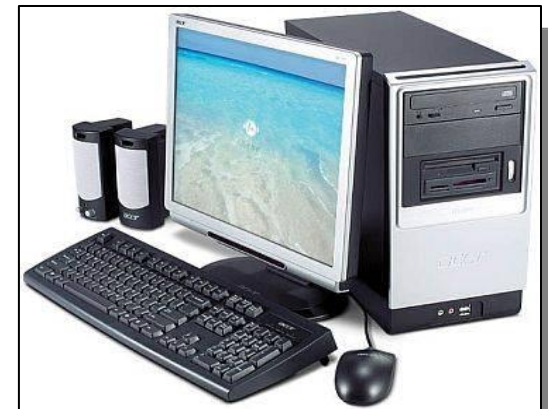
Реальное представление
предметной области



Формальное
представление
предметной области



Информационное
представление
предметной области



Извлечение информации

При извлечении информации важное место занимают различные формы и методы исследования данных:

- ✓ поиск ассоциаций, связанных с привязкой к какому-либо событию;
- ✓ обнаружение последовательностей событий во времени;
- ✓ выявление скрытых закономерностей по наборам данных, путем определения причинно-следственных связей между значениями определенных косвенных параметров исследуемого объекта (ситуации, процесса);
- ✓ оценка важности (влияния) параметров на развитие ситуации;
- ✓ классифицирование (распознавание), осуществляемое путем поиска критериев, по которым можно было бы относить объект (события, ситуации, процессы) к той или иной категории;
- ✓ кластеризация, основанная на группировании объектов по каким-либо признакам;
- ✓ прогнозирование событий и ситуаций.

Извлечение информации

Декомпозиция – научный метод, использующий структуру задачи и позволяющий заменить решение одной большой задачи решением серии меньших задач.

Декомпозиция на основе объектно-ориентированного подхода основана на выделении следующих основных понятий: объект, класс, экземпляр.

Объект – это абстракция множества предметов реального мира, обладающих одинаковыми характеристиками и законами поведения. Объект характеризует собой типичный неопределенный элемент такого множества. Основной характеристикой объекта является состав его атрибутов (свойств).

Атрибуты – это специальные объекты, посредством которых можно задать правила описания свойств других объектов.

Извлечение информации

Экземпляр объекта – это конкретный элемент множества. Например, объектом может являться государственный номер автомобиля, а экземпляром этого объекта – конкретный номер.

Класс – это множество предметов реального мира, связанных общностью структуры и поведением. Элемент класса – это конкретный элемент данного множества.

Понятия полиморфизма и наследования определяют эволюцию объектно-ориентированной системы, что подразумевает определение новых классов объектов на основе базовых.

Полиморфизм – способность объекта принадлежать более чем одному типу.

Наследование выражает возможность определения новых классов на основе существующих с возможностью добавления или переопределения данных и методов.

Извлечение информации

Инкапсуляция – это процесс отделения друг от друга элементов объекта, определяющих его устройство и поведение; инкапсуляция служит для того, чтобы изолировать контрактные обязательства абстракции от их реализации.

Среди методов обогащения информации различают структурное, статистическое, семантическое и прагматическое обогащения.

Структурное обогащение предполагает изменение параметров сообщения, отображающего информацию в зависимости от частотного спектра исследуемого процесса, скорости обслуживания источников информации и требуемой точности. При **статистическом обогащении** осуществляют накопление статистических данных и обработку выборок из генеральных совокупностей накопленных данных.

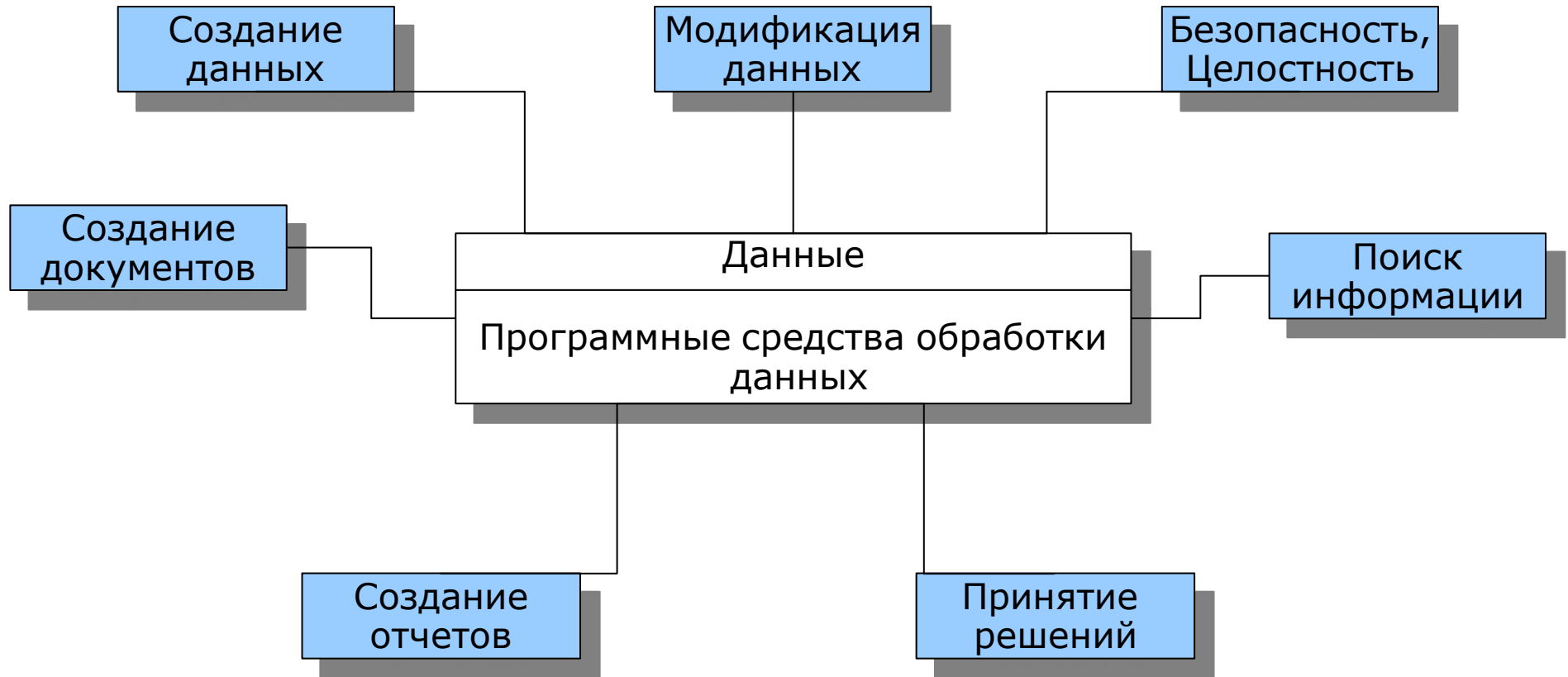
Извлечение информации

Семантическое обогащение означает минимизацию логической формы, исчислений и высказываний, выделение и классификацию понятий, содержания информации, переход от частных понятий к более общим. В итоге семантического обогащения удастся обобщенно представить обрабатываемую либо передаваемую информацию и устранить логическую противоречивость в ней.

Прагматическое обогащение является важной ступенью при использовании информации для принятия решения, при котором из полученной информации отбирается наиболее ценная, отвечающая целям и задачам пользователя.

Таким образом, *процесс извлечения* информации связан с переходом от реального представления предметной области к его описанию в формальном виде и в виде данных, которые отражают это представление

Обработка информации



Обработка информации

Обработка информации состоит в получении одних «информационных объектов» из других «информационных объектов» путем выполнения некоторых алгоритмов и является одной из основных операций, осуществляемых над информацией, и главным средством увеличения ее объема и разнообразия.

На самом верхнем уровне можно выделить числовую и нечисловую обработку. При числовой обработке используются такие объекты, как переменные, векторы, матрицы, многомерные массивы, константы и т.д. При нечисловой обработке объектами могут быть файлы, записи, поля, иерархии, сети, отношения и т.д.

Обработка информации

- С точки зрения реализации на основе современных достижений вычислительной техники выделяют следующие виды обработки информации:
- последовательная обработка, применяемая в традиционной фоннеймановской архитектуре ЭВМ, располагающей одним процессором;
- параллельная обработка, применяемая при наличии нескольких процессоров в ЭВМ;
- конвейерная обработка, связанная с использованием в архитектуре ЭВМ одних и тех же ресурсов для решения разных задач, причем если эти задачи тождественны, то это последовательный конвейер, если задачи одинаковые – векторный конвейер.

Обработка информации

Принято относить существующие архитектуры ЭВМ с точки зрения обработки информации к одному из следующих классов (Классификация параллельных архитектур по Флинну).

Архитектуры с одиночным потоком команд и данных (SISD).

Традиционная архитектура фон Неймана + КЭШ + память + конвейеризация

Архитектуры с одиночными потоками команд и данных (SIMD).

Особенностью данного класса является наличие одного (центрального) контроллера, управляющего рядом одинаковых процессоров.

SIMD (Single Instruction Multiple Data).

MIMD (Multiple Instruction Multiple Data).

Обработка информации

Архитектуры с множественным потоком команд и одиночным потоком данных (MISD). Один из немногих – систолический массив процессоров, в котором процессоры находятся в узлах регулярной решетки, роль ребер которой играют межпроцессорные соединения. К классу MISD ряд исследователей относит конвейерные ЭВМ, однако это не нашло окончательного признания, поэтому можно считать, что реальных систем – представителей данного класса не существует.

Архитектуры с множественным потоком команд и множественным потоком данных (MIMD). К этому классу могут быть отнесены следующие конфигурации: мультипроцессорные системы, системы с мультобработкой, вычислительные системы из многих машин, вычислительные сети.

Обработка информации

Создание данных, как процесс обработки, предусматривает их образование в результате выполнения некоторого алгоритма и дальнейшее использование для преобразований на более высоком уровне.

Модификация данных связана с отображением изменений в реальной предметной области, осуществляемых путем включения новых данных и удаления ненужных.

Контроль, безопасность и целостность направлены на адекватное отображение реального состояния предметной области в информационной модели и обеспечивают защиту информации от несанкционированного доступа (безопасность) и от сбоев и повреждений технических и программных средств.

Обработка информации

Поддержка принятия решения является наиболее важным действием, выполняемым при обработке информации.

Создание документов, сводок, отчетов заключается в преобразовании информации в формы, пригодные для чтения как человеком, так и компьютером. С этим действием связаны и такие операции, как обработка, считывание, сканирование и сортировка документов.

При преобразовании информации осуществляется ее перевод из одной формы представления или существования в другую, что определяется потребностями, возникающими в процессе реализации информационных технологий.

Обработка информации

В зависимости от степени информированности о состоянии управляемого процесса, полноты и точности моделей объекта и системы управления, взаимодействия с окружающей средой, процесс принятия решения протекает в различных условиях:

1. Принятие решений в условиях определенности. В этой задаче модели объекта и системы управления считаются заданными, а влияние внешней среды – несущественным. Поэтому между выбранной стратегией использования ресурсов и конечным результатом существует однозначная связь, откуда следует, что в условиях определенности достаточно использовать решающее правило для оценки полезности вариантов решений, принимая в качестве оптимального то, которое приводит к наибольшему эффекту.

2. Принятие решений в условиях риска. В отличие от предыдущего случая для принятия решений в условиях риска необходимо учитывать влияние внешней среды, которое не поддается точному прогнозу, а известно только вероятностное распределение ее состояний. В этих условиях использование одной и той же стратегии может привести к различным исходам, вероятности появления которых считаются заданными или могут быть определены.

3. Принятие решений в условиях неопределенности. Как и в предыдущей задаче между выбором стратегии и конечным результатом отсутствует однозначная связь. Кроме того, неизвестны также значения вероятностей появления конечных результатов, которые либо не могут быть определены, либо не имеют в контексте содержательного смысла.

4. Принятие решений в условиях многокритериальности. В любой из перечисленных выше задач многокритериальность возникает в случае наличия нескольких самостоятельных, не сводимых одна к другой целей. Наличие большого числа решений усложняет оценку и выбор оптимальной стратегии. Одним из возможных путей решения является использование методов моделирования.

Обработка информации

Экспертная система пользуется знаниями, которыми она обладает в своей узкой области, чтобы ограничить поиск на пути к решению задачи путем постепенного сужения круга вариантов. Для решения задач в экспертных системах используют:

- метод логического вывода, основанный на технике доказательств, называемой резолюцией и использующей опровержение отрицания (доказательство «от противного»);
- метод структурной индукции, основанный на построении дерева принятия решений для определения объектов из большого числа данных на входе;
- метод эвристических правил, основанных на использовании опыта экспертов, а не на абстрактных правилах формальной логики;
- метод машинной аналогии, основанный на представлении информации о сравниваемых объектах в удобном виде, например, в виде структур данных, называемых фреймами.

Обработка информации

Источники «интеллекта», проявляющегося при решении задачи, могут оказаться бесполезными либо полезными или экономичными в зависимости от определенных свойств области, в которой поставлена задача. Исходя из этого, может быть осуществлен выбор метода построения экспертной системы или использования готового программного продукта.

Процесс выработки решения на основе первичных данных, можно разбить на два этапа: выработка допустимых вариантов решений путем математической формализации с использованием разнообразных моделей и выбор оптимального решения на основе субъективных факторов.

Информационные потребности лиц, принимающих решение, во многих случаях ориентированы на интегральные технико-экономические показатели, которые могут быть получены в результате обработки первичных данных, отражающих текущую деятельность предприятия.

Для поддержки принятия решений обязательным является наличие следующих компонент:

- обобщающего анализа;
- прогнозирования;
- ситуационного моделирования.

Аналитические системы поддержки принятия решений (СППР) позволяют решать три основных задачи: ведение отчётности, анализ информации в реальном времени (OLAP) и интеллектуальный анализ данных.

Обработка информации

OLAP (On-Line Analytical Processing) – сервис представляет собой инструмент для анализа больших объемов данных в режиме реального времени. Взаимодействуя с OLAP-системой, пользователь сможет осуществлять гибкий просмотр информации, получать произвольные срезы данных, и выполнять аналитические операции детализации, свертки, сквозного распределения, сравнения во времени.

В зависимости от функционального наполнения интерфейса системы выделяют два основных типа систем поддержки принятия решений : EIS и DSS.

EIS (Execution Information System) – информационные системы руководства предприятия. Эти системы ориентированы на неподготовленных пользователей, имеют упрощенный интерфейс, базовый набор предлагаемых возможностей, фиксированные формы представления информации.

Обработка информации

DSS (Decision Support System) – полнофункциональные системы анализа и исследования данных, рассчитанные на подготовленных пользователей, имеющих знания как в части предметной области исследования, так и в части компьютерной грамотности.

PS (Presentation Services) – средства представления. Обеспечиваются устройствами, принимающими ввод от пользователя и отображающими то, что сообщает ему компонент логики представления PL, плюс соответствующая программная поддержка.

PL (Presentation Logic) – логика представления. Управляет взаимодействием между пользователем и ЭВМ. Обработывает действия пользователя по выбору альтернативы меню, по нажатию кнопки или выбору элемента из списка.

Обработка информации

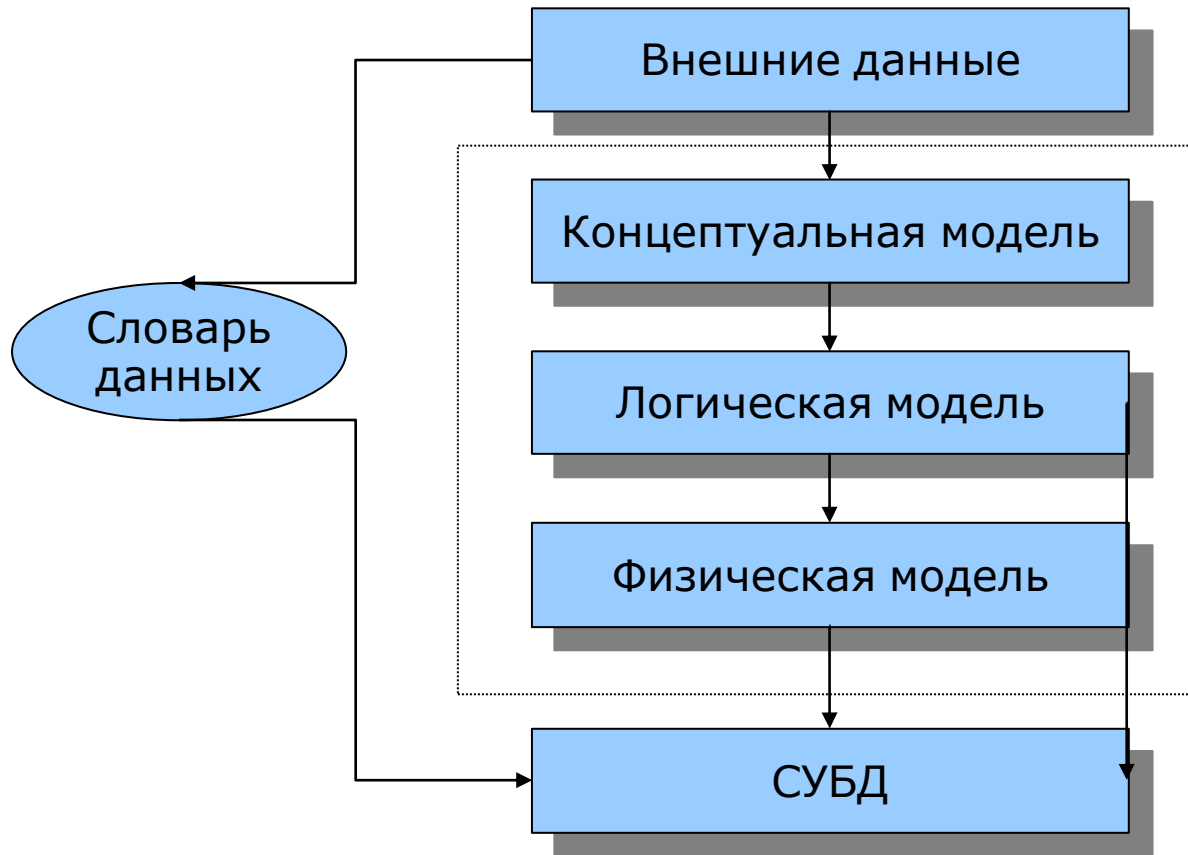
BL (Business or Application Logic) – прикладная логика. Набор правил для принятия решений, вычислений и операций, которые должно выполнить приложение.

DL (Data Logic) – логика управления данными. Операции с базой данных (SQL-операторы SELECT, UPDATE и INSERT), которые нужно выполнить для реализации прикладной логики управления данными.

DS (Data Services) – операции с базой данных. Действия СУБД, вызываемые для выполнения логики управления данными, такие как манипулирование данными, определения данных, фиксация или откат транзакций и т.п. СУБД обычно компилирует SQL-приложения.

FS (File Services) – файловые операции. Дисковые операции чтения и записи данных для СУБД и других компонент. Обычно являются функциями ОС.

Хранение информации



Хранение информации

Хранение и накопление являются одними из основных действий, осуществляемых над информацией и главным средством обеспечения ее доступности в течение некоторого промежутка времени.

База данных может быть определена как совокупность взаимосвязанных данных, используемых несколькими пользователями и хранящихся с регулируемой избыточностью.

Банк данных – система, представляющая определенные услуги по хранению и поиску данных определенной группе пользователей по определенной тематике.

Система баз данных – совокупность управляющей системы, прикладного программного обеспечения, базы данных, операционной системы и технических средств, обеспечивающих информационное обслуживание пользователей.

Хранение информации

Хранилище данных (ХД – используют также термины Data Warehouse, «склад данных», «информационное хранилище») – это база, хранящая данные, агрегированные по многим измерениям.

Альтернативой хранилищу данных является концепция витрин данных (Data Mart).

Витрины данных – множество тематических БД, содержащих информацию, относящуюся к отдельным информационным аспектам предметной области.

Хранение информации

Еще одним важным направлением развития баз данных являются репозитории. Репозитарий, в упрощенном виде, можно рассматривать просто как базу данных, предназначенную для хранения не пользовательских, а системных данных.

Каждый из участников действия (пользователь, группа пользователей, «физическая память») имеет свое представление об информации.

По отношению к пользователям применяют трехуровневое представление для описания предметной области:

- *концептуальное,*
- *логическое и*
- *внутреннее (физическое).*

Хранение информации

Концептуальный уровень связан с частным представлением данных группы пользователей в виде внешней схемы, объединяемых общностью используемой информации. Каждый конкретный пользователь работает с частью БД и представляет ее в виде внешней модели. Этот уровень характеризуется разнообразием используемых моделей (модель «сущность – связь», ER-модель, модель Чена), бинарные и инфологические модели, семантические сети.

Логический уровень является обобщенным представлением данных всех пользователей в абстрактной форме. Используются три вида моделей: иерархические, сетевые и реляционные.

Сетевая модель является моделью объектов-связей, допускающей только бинарные связи «многие к одному» и использует для описания модель ориентированных графов.

Хранение информации

Иерархическая модель является разновидностью сетевой, являющейся совокупностью деревьев (лесом).

Реляционная модель использует представление данных в виде таблиц (реляций), в ее основе лежит математическое понятие теоретико-множественного отношения, она базируется реляционной алгебре и теории отношений.

Физический (внутренний) уровень связан со способом фактического хранения данных в физической памяти ЭВМ. Основными компонентами физического уровня являются хранимые записи, объединяемые в блоки; указатели, необходимые для поиска данных; данные переполнения; промежутки между блоками; служебная информация.

Хранение информации

По наиболее характерным признакам БД можно классифицировать следующим образом:

по способу хранения информации:

- интегрированные;
- распределенные;

по типу пользователя:

- монопользовательские;
- многопользовательские;

по характеру использования данных:

- монопользовательские;
- предметные.

Хранение информации

В настоящее время при проектировании БД используют два подхода.

Первый из них основан на стабильности данных, что обеспечивает наибольшую гибкость и адаптируемость к используемым приложениям. Применение такого подхода целесообразно в тех случаях, когда не предъявляются жесткие требования к эффективности функционирования (объему памяти и продолжительности поиска), существует большое число разнообразных задач с изменяемыми и непредсказуемыми запросами.

Второй подход базируется на стабильности процедур запросов к БД и является предпочтительным при жестких требованиях к эффективности функционирования, особенно это касается быстродействия.

Хранение информации

Другим важным аспектом проектирования БД является проблема интеграции и распределения данных. Распределение данных по месту их использования может осуществляться различными способами:

1. *Копируемые данные.* Одинаковые копии данных хранятся в различных местах использования, так как это дешевле передачи данных. Модификация данных контролируется централизованно;
Подмножество данных. Группы данных, совместимые с исходной базой данных, хранятся отдельно для местной обработки;
Реорганизованные данные. Данные в системе интегрируются при передаче на более высокий уровень;
Секционированные данные. На различных объектах используются одинаковые структуры, но хранятся разные данные;

Хранение информации

Данные с отдельной подсхемой. На различных объектах используются различные структуры данных, объединяемые в интегрированную систему;

Несовместимые данные. Независимые базы данных, спроектированные без координации, требующие объединения.

Существуют два основных направления реализации СУБД:

программное и аппаратное.

Хранение информации

Программная реализация (в дальнейшем СУБД) представляет собой набор программных модулей, работает под управлением конкретной ОС и выполняет следующие функции:

- описание данных на концептуальном и логическом уровнях;
- загрузку данных;
- хранение данных;
- поиск и ответ на запрос (транзакцию);
- внесение изменений;

обеспечение безопасности и целостности.

Обеспечивает пользователя следующими языковыми средствами:

- языком описания данных (ЯОД);
- языком манипулирования данными (ЯМД);
- прикладным (встроенным) языком данных (ПЯД, ВЯД).

Хранение информации

Аппаратная реализация предусматривает использование так называемых *машин баз данных* (МБД). Их появление вызвано возросшими объемами информации и требованиями к скорости доступа. Слово «машина» в термине МБД означает вспомогательный периферийный процессор. Термин «компьютер БД» – автономный процессор баз данных или процессор, поддерживающий СУБД.

Основные направления МБД:

- параллельная обработка;
- распределенная логика;
- ассоциативные ЗУ;
- конвейерные ЗУ;
- фильтры данных и др.

Хранение информации

Проектирование БД можно объединить в *четыре этапа*.

На этапе *формулирования и анализа требований*

устанавливаются цели организации, определяются требования к БД. Эти требования документируются в форме, доступной конечному пользователю и проектировщику БД. Обычно при этом используется методика интервьюирования персонала различных уровней управления.

Этап *концептуального проектирования* заключается в описании и синтезе информационных требований пользователей в первоначальный проект БД. Результатом этого этапа является высокоуровневое представление информационных требований пользователей на основе различных подходов.

В процессе *логического проектирования* высокоуровневое представление данных преобразуется в структуре используемой СУБД. Полученная логическая структура БД может быть оценена количественно с помощью различных характеристик (число обращений к логическим записям, объем данных в каждом приложении, общий объем данных и т.д.).

14.10.2016

Хранение информации

Проектирование БД можно объединить в *четыре этапа*:

- 1.этап формулирования и анализа требований;*
- 2.этап формулирования и анализа требований;*
- 3.этап логического проектирования;*
- 4.этап физического проектирования.*

Этап **концептуального проектирования** устанавливаются цели организации, определяются требования к БД. Эти требования документируются в форме, доступной конечному пользователю и проектировщику БД. Обычно при этом используется методика интервьюирования персонала различных уровней управления. Этап **концептуального проектирования** заключается в описании и синтезе информационных требований пользователей в первоначальный проект БД. Результатом этого этапа является высокоуровневое представление информационных требований пользователей на основе различных подходов.

Хранение информации

В процессе *логического проектирования* высокоуровневое представление данных преобразуется в структуре используемой СУБД. Полученная логическая структура БД может быть оценена количественно с помощью различных характеристик (число обращений к логическим записям, объем данных в каждом приложении, общий объем данных и т.д.).

На этапе *физического проектирования* решаются вопросы, связанные с производительностью системы, определяются структуры хранения данных и методы доступа.

Взаимодействие между этапами проектирования и словарной системой необходимо рассматривать отдельно. Процедуры проектирования могут использоваться независимо в случае отсутствия словарной системы.

Предназначение склада данных – информационная поддержка принятия решений, а не оперативная обработка данных. Потому база данных и склад данных не являются одинаковыми понятиями.

Хранение информации

Основные принципы организации хранилищ данных следующие.

Предметная ориентация. В оперативной базе данных обычно поддерживается несколько предметных областей, каждая из которых может послужить источником данных для ХД.

Средства интеграции. Приведение разных представлений одних и тех же сущностей к некоторому общему типу.

Постоянство данных. В ХД не поддерживаются операции модификации в смысле традиционных баз данных.

Хронология данных. Благодаря средствам интеграции реализуется определенный хронологический временной аспект, присущий содержимому ХД.

Хранение информации

Основные функции репозитариев:

- парадигма включения/выключения и некоторые формальные процедуры для объектов;
- поддержка множественных версий объектов и процедуры управления конфигурациями для объектов;
- оповещение инструментальных и рабочих систем об интересующих их событиях;
- управление контекстом и разные способы обзора объектов репозитария;
- определение потоков работ.

Хранение информации

Рассмотрим кратко основные направления научных исследований в области баз данных:

- развитие теории реляционных баз данных;
- моделирование данных и разработка конкретных моделей разнообразного назначения;
- отображение моделей данных, направленных на создание методов их преобразования и конструирования коммутативных отображений, разработку архитектурных аспектов отображения моделей данных и спецификаций определения отображений для конкретных моделей данных;
- создание СУБД с мультимодельным внешним уровнем, обеспечивающих возможности отображения широко распространенных моделей;
- разработка, выбор и оценка методов доступа;

Хранение информации

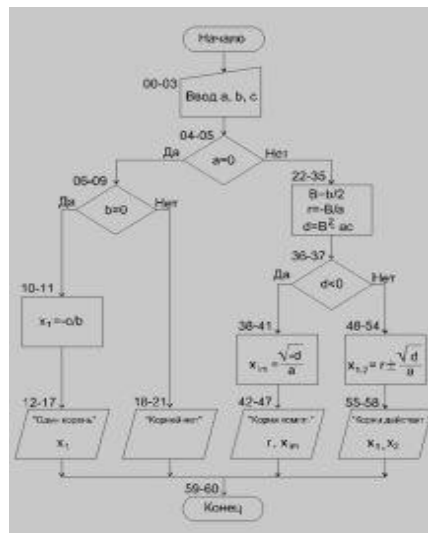
- создание самоописываемых баз данных, позволяющих применять единые методы доступа для данных и метаданных;
- управление конкурентным доступом;
- развитие системы программирования баз данных и знаний, которые обеспечивали бы единую эффективную среду как для разработки приложений, так и для управления данными;
- совершенствование машины баз данных;
- разработка дедуктивных баз данных, основанных на применении аппарата математической логики и средств логического программирования, а также пространственно-временных баз данных;
- интеграция неоднородных информационных ресурсов.

Извлечение информации

Реальное представление
предметной области



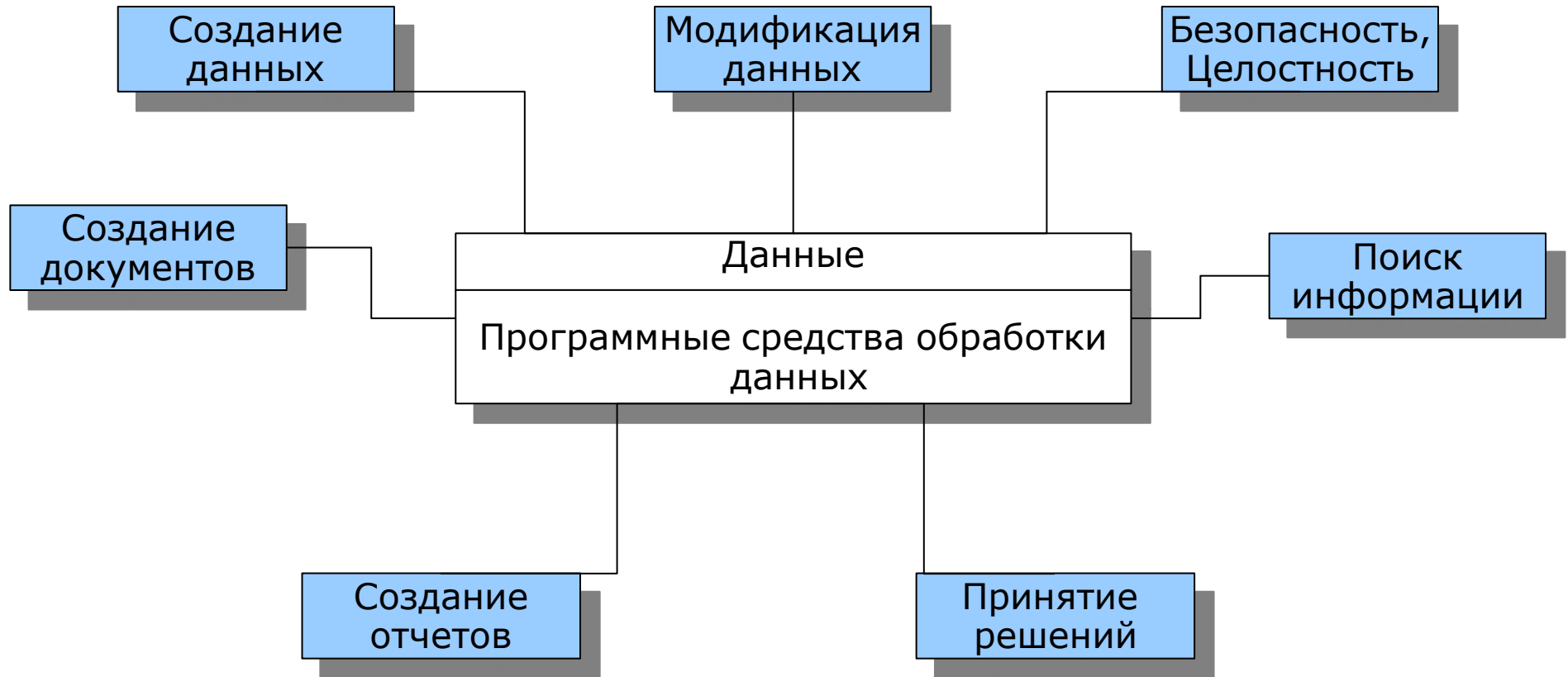
Формальное
представление
предметной области



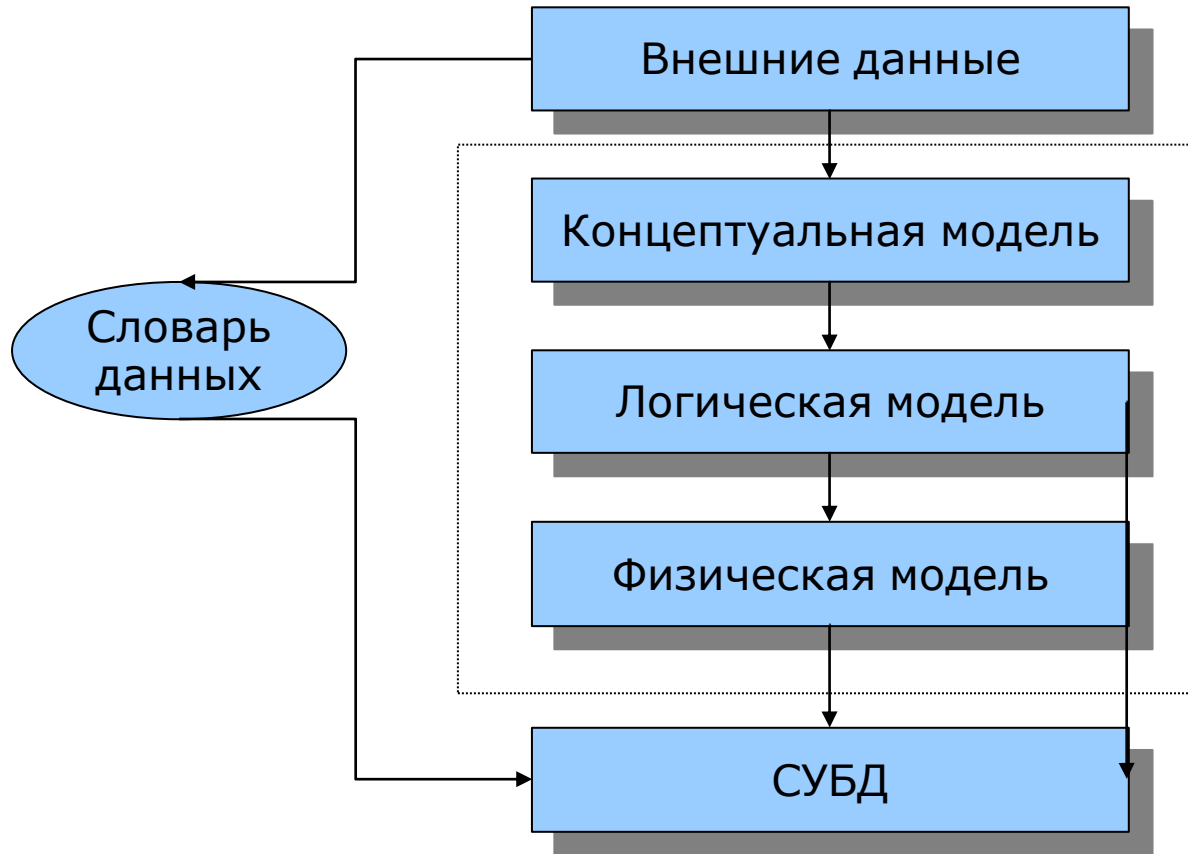
Информационное
представление
предметной области



Обработка информации



Хранение информации



Представление и использование информации



Представление и использование информации

В условиях использования информационных технологий функции распределены между человеком и техническими устройствами.

При анализе деятельности человека наибольшее значение имеют

эргономические (инженерно-психологические) и

психологические (социально-психологические) факторы.

Эргономические факторы позволяют, во-первых, определить рациональный набор функций человека, во-вторых, обеспечить рациональное сопряжение человека с техническими средствами и информационной средой.

Психологические факторы имеют большое значение, так как внедрение информационных технологий в корне изменяет деятельность человека. Наряду с положительными моментами, связанными с рационализацией деятельности, предоставлением новых возможностей, возникают и негативные явления.

Представление и использование информации

Это может быть вызвано различными факторами: психологическим барьером, усложнением функций, другими субъективными факторами (условиями и организацией труда, уровнем заработной платы, результативностью труда, изменением квалификации).

При работе в среде информационных технологий человек воспринимает не сам объект, а некоторую его обобщенную информационную модель, что накладывает особые требования на совместимость пользователя с различными компонентами информационных технологий.

Важным признаком, который необходимо учитывать при разработке и внедрении информационных технологий является отношение человека к информации.

Представление и использование информации

Основной задачей операции представления информации пользователю является создание эффективного интерфейса в системе «человек—компьютер». При этом осуществляется преобразование информации в форму, удобную для восприятия пользователя.

Среди существующих вариантов интерфейса в системе «человек—компьютер» можно выделить два основных типа: на *основе меню* («смотри и выбирай») и на *основе языка команд* («вспоминай и набирай»).

Интерфейсы типа меню облегчают взаимодействие пользователя с компьютером, так как не требуют предварительного изучения языка общения с системой.

Интерфейс на основе языка команд требует знания пользователем синтаксиса языка общения с компьютером.

Достоинством командного языка является его гибкость и мощность.

Представление и использование информации

Технология представления информации должна давать дополнительные возможности для понимания данных пользователями, поэтому целесообразно использование графики, диаграмм, карт.

Пользовательский интерфейс целесообразно строить на основе концептуальной модели предметной области, которая представляется совокупностью взаимосвязанных объектов со своей структурой. В сценарии работы пользователя при информационном наполнении понятий предметной области выделяем две фазы:

- выбор окон;
- работа с окнами.

Представление и использование информации

Таким образом, фаза выбора объектов должна поддерживаться следующими функциями:

- работой с общим каталогом окон в главном разделе;
- созданием нового раздела;
- удалением раздела;
- редактированием описания раздела;
- передачей определений и окон между разделами;
- движением по иерархии разделов;
- отбором разделов для работы;
- отбором окон для работы.

Представление и использование информации

Позиции окон могут быть связаны с другими окнами через соответствующие команды из типового набора. По существу спецификация окон задает сценарий работы с экземплярами. Окно – средство взаимосвязи пользователя с системой. Окно представляется как специальный объект. Проектирование пользовательского интерфейса представляет собой процесс спецификации окон.

Примером оконного интерфейса является интерфейс MS Windows, использующий метафору рабочего стола и включающий ряд понятий, близких к естественным (окна, кнопки, меню и т.д.).

Пользователь информационной системы большей частью вынужден использовать данные из самых разных источников: файлов, баз данных, электронных таблиц, электронной почты и т. д. При этом данные имеют самую различную форму: текст, таблицы, графика, аудио- и видеоданные и др.

Представление и использование информации

Этим требованиям удовлетворяет Web-технология. Развитие средств вычислительной техники привело к ситуации, когда вместо традиционных параметров – *производительность, пропускная способность, объем памяти*, узким местом стал *интерфейс* с пользователем. Первым шагом на пути преодоления кризисной ситуации стала концепция гипертекста, впервые предложенная Теодором Хольмом Нельсоном. По своей сути гипертекст – это обычный текст, содержащий ссылки на собственные фрагменты и другие тексты.

Аналогом гипертекста можно считать книгу, оглавление которой по своей сути представляет ссылки на главы, разделы, страницы. Дальнейшее развитие гипертекст получил с появлением сети Интернет, позволившей размещать тексты на различных, территориально удаленных компьютерах.

Представление и использование информации

Web-сервер выступает в качестве информационного концентратора, получающего информацию из разных источников и в однородном виде представляющем ее пользователю. Средства Web обеспечивают также представление информации с нужной степенью детализации с помощью Web-навигатора. Таким образом web – это инфраструктурный интерфейс для пользователей различных уровней. Несомненным преимуществом Web-технологии является удобная форма предоставления информационных услуг потребителям независимо от платформы и содержимого.

Лекция окончена!

Благодарю за внимание!

