

Модели процессов извлечения, представления и использования в информационных системах

1. Извлечение информации

В процессе извлечения информации основной акцент сделан на формы и методы исследования данных, позволяющих формализовать и абстрагировано описать предметную область.

Объекты и их свойства, процессы и функции, выполняемые этими объектами или для них являются источниками данных в любой предметной области. Под предметной областью понимают часть реального мира, отражаемую в информационной базе. Например, если информационная система (ИС) предназначена для обслуживания читателей библиотеки, то предметной областью является библиотечный фонд книг и периодических изданий. Если ИС обслуживает пассажиров железной дороги, то предметной областью является система пассажирских перевозок.

Любая предметная область рассматривается в виде трех представлений:

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

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 ;
- 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.

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.

Экземпляр объекта – это конкретный элемент множества.

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.

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.

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.

2. 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. 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. Оно может быть пассивным, когда пользователю предоставляется информация по жесткому алгоритму, и активным, когда пользователь создает необходимые ему данные.

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. На каждом шаге диалога пользователю предъявляются все возможные в данный момент команды в виде наборов пунктов меню, из которого пользователь должен выбрать нужный. Такой способ общения удобен для начинающих и непрофессиональных пользователей.

Интерфейс на основе языка команд требует знания пользователем синтаксиса языка общения с компьютером. Достоинством командного языка является его гибкость и мощность.

Указанные два способа реализации интерфейса представляют собой крайние случаи, между которыми возможно существование различных промежуточных вариантов.

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. Однако доступ к объектам и их экземплярам возможен только через систему окон различных типов. Ряд окон связан с конкретным объектом. В соответствии с этим предложением в сценарии работы пользователя при информационном наполнении понятий предметной области выделяем две фазы:

- выбор окон; • работа с окнами.

Для упрощения работы окна можно группировать в соответствии с функциональными потребностями. С этой целью вводится механизм разделов, который предоставляет возможность создания иерархии функционально ориентированных разделов, в каждый из которых включается необходимый набор других разделов и окон. Посредством спецификации окон для каждого из объектов возможно указать допустимые режимы работы с экземплярами и состав видимых атрибутов с режимами работы с ними. Возможно отобрать несколько разделов и несколько окон в них одновременно.

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.

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. В связи с этим возникает проблема интеграции источников информации,

закрывающаяся в том, что, во-первых, пользователю должны предоставляться не данные, а информация в форме, максимально удобной для восприятия, во-вторых, он должен использовать единственный универсальный интерфейс, позволяющий единообразно работать с подготовленной информацией. Пассивные пользователи, называемые иногда потребителями, обладают рядом специфических качеств, связанных с отсутствием времени, желания и квалификации для более глубокого изучения используемых инструментальных средств. В этом случае алгоритм общения с системой должен быть предельно простым. Другая часть пользователей требует предоставления достаточно широкого круга средств активного влияния на выполняемые информационные процессы.

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 Холмом 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. Внутри книги содержатся ссылки на другие источники. Дальнейшее развитие гипертекст получил с появлением сети Интернет, позволившей размещать тексты на различных, территориально удаленных компьютерах. При этом требовалось дальнейшее совершенствование интерфейса, так как имеющийся не позволял представить разнообразную информацию разной природы, был ограничен и затруднен для восприятия, отсутствовал доступ множества потребителей к единому массиву структурированной информации. В результате была предложена и реализована концепция навигатора Web. Web-сервер выступает в качестве информационного концентратора, получающего информацию из разных источников и в однородном виде представляющую ее пользователю. Средства Web обеспечивают также представление информации с нужной степенью легализации 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, имеющая следующие особенности:

- информация предоставляется потребителю в виде публикаций;
- публикация может объединять информационные источники различной природы и географического расположения;
- изменения в информационных источниках мгновенно отражаются в публикациях;

- в публикациях могут содержаться ссылки на другие публикации без ограничения на местоположение и источники последних (гипертекстовые ссылки);
- потребительские качества публикаций соответствуют современным стандартам мультимедиа (доступны текст, графика, звук, видео, анимация);
- публикатор не заботится о процессе доставки информации к потребителю;
- число потенциальных потребителей информации практически не ограничено;
- публикации отражают текущую информацию, время запаздывания определяется исключительно скоростью подготовки электронного документа;
- информация, предоставленная в публикации, легко доступна благодаря гипертекстовым ссылкам и средствам контекстного поиска;
- информация легко усваивается потребителем благодаря широкому спектру изобразительных возможностей, предоставляемых Web - технологией;
- технология не предъявляет особых требований к типам и источникам информации;
- технология допускает масштабируемые решения: увеличение числа одновременно обслуживаемых потребителей не требует радикальной перестройки системы.

Контрольные вопросы

1. Какие формы исследования данных вы знаете?
2. Какова суть декомпозиции на основе объектно-ориентированного подхода?
3. Что такое инкапсуляции, полиформизм и наследование?
4. Какие существуют методы обогащения информации?
5. Что такое интерфейс и какова его роль в процессе представления и использования информации?
6. На чем основана концепция гипертекста?
7. Какие существуют виды интерфейсов?