

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ

ТЕМА ЛЕКЦИИ: «Разработка и применение экспертных систем. Документальные ИПС как предшественники экспертных систем»

**Массель Л.В., д.т.н., профессор
кафедры Автоматизированных систем
факультета Кибернетики ИрГТУ**

ЧЕЛОВЕЧЕСКАЯ И ИСКУССТВЕННАЯ КОМПЕТЕНТНОСТЬ: ДОСТОИНСТВА ИСКУССТВЕННОЙ КОМПЕТЕНТНОСТИ

Человеческая	Искусственная
1. Непрочная	1. Постоянная
2. Труднопередаваемая	2. Легкопередаваемая
3. Непредсказуемая	3. Устойчивая
4. Дорогая	4. Приемлемая по затратам

ЧЕЛОВЕЧЕСКАЯ И ИСКУССТВЕННАЯ КОМПЕТЕНТНОСТЬ: НЕДОСТАТКИ ИСКУССТВЕННОЙ КОМПЕТЕНТНОСТИ

Человеческая	Искусственная
<ol style="list-style-type: none">1. Творческая2. Приспосабливающаяся3. Использующая чувственное описание4. Широкая по охвату5. Использующая общедоступные знания	<ol style="list-style-type: none">1. Запрограммированная2. Нуждается в подсказке3. Использующая символный ввод-вывод4. Узконаправленная5. Использующая специализированные знания

ОСНОВНЫЕ ВИДЫ ДЕЯТЕЛЬНОСТИ ЭКСПЕРТНЫХ СИСТЕМ

- Интерпретация (описание ситуации по информации, поступающей от датчика)
- Прогноз (описание вероятностных последствий заданной ситуации)
- Диагностика (выявление причин неправильного функционирования системы по результатам наблюдения)
- Проектирование (построение конфигурации объектов при заданных ограничениях)
- Планирование (определение последовательности действий)
- Наблюдение (сравнение с ожидаемыми результатами)
- Отладка (составление «рецептов» устранения неправильного функционирования системы)
- Ремонт (выполнение последовательности действий, предписанных инструкцией)
- Обучение (диагностика, отладка обучаемых)
- Управление (поведение системы как целого)

ОБЛАСТИ ПРИМЕНЕНИЯ ЭКСПЕРТНЫХ СИСТЕМ:

- Военное дело
- Технология
- Инженерное дело
- Информатика
- Компьютерные системы
- Космические системы
- Математика
- Медицина
- Метеорология
- Промышленность
- Сельское хозяйство
- Управление процессами
- Физика
- Химия
- Электроника
- Юриспруденция

ОТЛИЧИЕ ЭКСПЕРТНОЙ СИСТЕМЫ ОТ ТРАДИЦИОННЫХ КОМПЬЮТЕРНЫХ ПРОГРАММ

- *Экспертная система* – это набор **компьютерных программ**, которые решают задачи в интересующей нас предметной области и содержат *компонент решения проблемы и компонент поддержки* .
- Компонент поддержки позволяет пользователю работать с другой программой и может включать специальные отладочные средства для работы с программой. Этот компонент помогает разработчику тестировать программу, пользователю – взаимодействовать с экспертной системой, эксперту – модифицировать данные знания экспертной системы. Также компонент поддержки может включать различные средства ввода-вывода информации, в том числе, о ходе работы системы, и средства объяснения.
- В архитектуре экспертной системы компонент решения проблемы образуют база знаний и машина вывода; компонент поддержки - подсистема объяснения, подсистема приобретения знаний, интерфейс пользователя.

ОТЛИЧИЕ ЭКСПЕРТНОЙ СИСТЕМЫ ОТ ТРАДИЦИОННЫХ КОМПЬЮТЕРНЫХ ПРОГРАММ

Под экспертной системой понимается программа, обладающая следующими свойствами:

- Компетентность, т.е. ЭС должна:
 - достигать экспертного уровня решений
 - быть «умной» (делать выводы, давать объяснения и др.)
 - адекватно реагировать на запросы
- Символьное рассуждение, т.е. ЭС должна:
 - представлять знания в символьном виде
 - уметь переформулировать символьные знания
- Глубина, т.е. ЭС должна:
 - уметь работать в предметной области, содержащей трудные задачи
 - уметь использовать сложные правила
- Самосознание, т.е. ЭС должна:
 - уметь исследовать свои рассуждения
 - объяснять свои действия

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА РАЗРАБОТКИ ЭКСПЕРТНЫХ СИСТЕМ

Существуют *два класса инструментальных средств разработки экспертных систем:*

- Пустые экспертные системы
- Оболочки экспертных систем
- В *пустых экспертных системах* БЗ не заполнена. При ее использовании специалисту необходимо спроектировать БЗ и внести ее в экспертную систему в соответствии с правилами. Ее особенностью является то, что она может быть эффективна лишь для одного класса задач, т.к. принятые в системе способы представления знаний и рассуждений ориентированы на этот класс задач.
- *Оболочки экспертных систем* являются инструментальным средством для проектирования и создания экспертных систем. В ее состав входят: средства проектирования БЗ с различными формами представления знаний и выбора режима работы решения задач (этот класс инструментальных средств более универсальный)

СТАДИИ РАЗРАБОТКИ ЭКСПЕРТНЫХ СИСТЕМ

Прототип	Полнота охвата	Точность (качество)	Объем БЗ	Срок реализации
Демонстрационный	Три задачи	Низкое	50-100 правил	2-3 месяца
Исследовательский	Почти полная	Низкое	200-500 правил	1-2 года
Действующий	Полная	Хорошая, низкое быстродействие	500-1000 правил	2-3 года
Промышленный	Полная	Высокое и быстрое	1000-1500 правил	2-4 года
Коммерческий	Полная расширенная	Высокое и быстрое	До 3000 правил	?

ПРЕДШЕСТВЕННИКИ ЭКСПЕРТНЫХ СИСТЕМ:

ДОКУМЕНТАЛЬНЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ

- Информационные системы делят на фактографические (ориентированные на работу со структурированными данными - фактами) и документальные (ориентированные на работу со слабо структурированными или неструктурированными данными – текстами).
- Первые системы автоматизированного информационного поиска документов появились еще в 60-х годах, развитые коммерческие информационно-поисковые системы, ориентированные на накопление и обработку текстовых документов, получили распространение лишь в конце 80-х – начале 90-х годов.
- *Документальные информационно-поисковые системы* можно считать прообразом экспертных систем

ДОКУМЕНТАЛЬНЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ СИСТЕМЫ

- В фактографических ИС единичным элементом данных, имеющих отдельное смысловое значение, является запись, образуемая конечной совокупностью полей-атрибутов. В документальных ИС единичным элементом данных является неструктурированный на более мелкие элементы *документ*.
- Основной задачей документальных ИС является накопление и предоставление пользователю документов, содержание, тематика, реквизиты и т. п. которых адекватны его информационным потребностям. Поэтому документальную ИС определяют как единое хранилище документов с инструментарием поиска и отбора необходимых документов. Поисковый характер документальных ИС исторически определил еще одно их название – *информационно-поисковые системы (ИПС)*.

СПЕЦИФИКА ДОКУМЕНТАЛЬНЫХ ИПС

- Поиск информации (данных) осуществляется и в фактографических ИС. Специфика документальных ИПС заключается в том, что они удовлетворяют информационные запросы пользователя, предоставляя ему документы, в которых содержится интересующая пользователя информация.
- В зависимости от особенностей реализации хранилища документов и механизма поиска документальные ИПС можно разделить на две группы:
 - системы на основе индексирования;
 - семантически-навигационные системы.

- В семантически-навигационных системах документы, помещаемые в хранилище, оснащаются специальными навигационными конструкциями, соответствующими смысловым связям (ссылкам) между различными документами или отдельными фрагментами одного документа. Такие конструкции реализуют некоторую *семантическую (смысловую) сеть* в базе документов. Способ и механизм выражения информационных потребностей в подобных системах заключаются в явной навигации (перенаправлении) пользователя по смысловым отсылкам между документами. Такой подход реализуется в гипертекстовых ИПС.

СИСТЕМЫ НА ОСНОВЕ ИНДЕКСИРОВАНИЯ

- Процесс отображения документа в поисковое пространство называется индексированием.
- В системах на основе индексирования исходные документы помещаются в базу без какого-либо дополнительного преобразования, но при этом смысловое содержание каждого документа отображается в некоторое поисковое пространство.
- Иначе говоря, каждый документ описывается набором ключевых слов (словосочетаний) – дескрипторов. Этот набор дескрипторов называется поисковым образом документа (ПОД). Для поиска документа пользователь формирует свой набор дескрипторов, формируя поисковый образ запроса (ПОЗ) к базе документов.
- Документальная ИС на основе определенных критериев и способов ищет документы, для которых ПОД совпадают или близки ПОЗ, и выдает соответствующие документы (или их адреса – ссылки на них). Соответствие найденных документов запросу пользователя называется *релевантностью*.

ИНФОРМАЦИОННОЕ ОПОВЕЩЕНИЕ В ДОКУМЕНТАЛЬНЫХ ИС

Еще одна функция документальных ИС – задачи информационного оповещения по всем новым поступающим в систему документам, соответствующим заранее определенным информационным потребностям пользователя. Эти потребности отражаются в поисковое пространство в виде так называемых поисковых профилей пользователей. Информационно-поисковая система по мере поступления и индексирования новых документов сравнивает ПОД с поисковыми профилями пользователей и принимает решение о соответствующем оповещении.

Поисковое пространство, отображающее ПОД и реализующее механизмы информационного поиска документов, так же, как и в СУБД фактографических систем, строится на основе языков документальных баз данных, называемых информационно-поисковыми языками (ИПЯ). По аналогии с языками баз данных фактографических систем в ИПЯ выделяют структурную и манипуляционную составляющие (в СУБД – языки описания данных и языки манипулирования данными).

- Структурная составляющая
 - в семантически-навигационных ИПС – гипертекстовые технологии;
 - в ИПС на основе индексирования – индексные указатели:
 - информационно-поисковые каталоги;
 - тезаурус;
 - генеральный указатель.
- Поисковая (манипуляционная составляющая)
 - Дескрипторные языки
 - Семантические языки
 - предикатные
 - реляционные

СТРУКТУРНАЯ СОСТАВЛЯЮЩАЯ ИПЯ

Структурная составляющая ИПЯ семантически-навигационных систем реализуется в виде техники смысловых отсылок в текстах документов и специальном навигационном интерфейсе по ним и в настоящее время представлена гипертекстовыми технологиями.

Структурная составляющая ИПЯ документальных ИПС на основе индексирования реализуется индексными указателями в форме *информационно-поисковых каталогов, тезаурусов и генеральных указателей.*

Информационно-поисковые каталоги представляют собой *классификационную систему знаний по определенной предметной области*. Смысловое содержание документа в информационно-поисковых каталогах отражается тем или иным *классом каталога*, а *индексирование* документов заключается в *присвоении* каждому документу специального кода (*индекса*) соответствующего по содержанию класса (классов) каталога и создания на этой основе специального индексного указателя (пример – УДК (универсальная десятичная классификация)).

Тезаурус представляет собой специальным образом организованную совокупность основных лексических единиц (понятий) предметной области и описание семантических отношений между лексическими единицами, не зависящими от контекста. Независимость от контекста обозначает обобщенность (абстрагированность) смысловых отношений, например, отношения «род-вид», «предмет-целое», «субъект-объект-средство-место-время_действия». Так же, как и в информационно-поисковых каталогах, в информационно-поисковом пространстве отражается не весь текст документа, а только выраженное средствами тезауруса смысловое содержание документа.

ГЕНЕРАЛЬНЫЙ УКАЗАТЕЛЬ

Генеральный указатель (глобальный словарь-индекс) – представляет собой *перечисление всех слов (словоформ), имеющих в документах хранилища, с указанием (отсылками) координатного местонахождения каждого слова (№ документа - № абзаца - № предложения - № слова)*. Индексирование нового документа производится через дополнение *координатных отсылок тех словоформ генерального указателя, которые присутствуют в новом документе*. Так как поисковое пространство в таких системах отражает полностью весь текст документа, то такие системы получили название полнотекстовых ИПС.

ДЕСКРИПТОРНЫЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ

Поисковая (манипуляционная) составляющая ИПЯ реализуется дескрипторными и семантическими языками запросов.

В *дескрипторных языках* документы и запросы представляются наборами некоторых лексических единиц (слов, словосочетаний, терминов) – дескрипторов, не имеющих между собой связей (не имеющих грамматики). Таким образом, каждый документ или запрос представлен некоторым набором дескрипторов (ПОД или ПОЗ). Поиск осуществляется через сопоставление этих наборов.

СЕМАНТИЧЕСКИЕ ИНФОРМАЦИОННО-ПОИСКОВЫЕ ЯЗЫКИ

Семантические языки содержат грамматические и семантические конструкции для описания смыслового содержания документов и запросов.

Семантические языки делятся на две большие группы:

- предикатные языки
- реляционные языки

ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ ДОКУМЕНТАЛЬНЫХ ИПС

Полнота информационного поиска (R).

$R=A/C$, где A – число найденных документов, соответствующих информационным потребностям пользователя; C – общее число документов в системе, соответствующих информационным потребностям пользователя.

Точность информационного поиска (P).

$P=A/L$, где A – число найденных документов, соответствующих информационным потребностям пользователя, L – общее число документов, выданных в ответ на запрос пользователя.

Информационный шум системы.

$K=(L-A)/L$, где K – коэффициент информационного шума, определяется отношением числа нерелевантных документов $(L-A)$, выданных в ответе пользователю, к общему числу документов L .

В идеале полнота информационного поиска и точность информационного поиска должны приближаться к единице, хотя на практике их значения колеблются в пределах от 60 до 90%.

ВОПРОСЫ К ЛЕКЦИИ

1. Человеческая и искусственная компетентность: достоинства и недостатки
2. Основные виды деятельности ЭС
3. Области применения ЭС
4. Отличие ЭС от обычных компьютерных программ
5. Инструментальные средства и стадии разработки ЭС
6. Документальные информационно-поисковые системы как прообраз экспертных систем
7. Семантически-навигационные системы и системы на основе индексирования
8. Информационно-поисковые языки: структурная и манипуляционная составляющая
9. Структурная составляющая ИПЯ: информационно-поисковые каталоги, тезаурус и генеральный указатель
10. Манипуляционная составляющая ИПЯ: дескрипторные и семантические языки (предикатные и реляционные)
11. Показатели эффективности функционирования документальных ИС: полнота и точность поиска, информационный шум