

Федеральное агентство по образованию
Государственное образовательное учреждение высшего профессионального образования
Амурский государственный университет
(ГОУВПО «АмГУ»)

УТВЕРЖДАЮ

Зав. кафедрой ИиУС

_____ А.В. Бушманов

«__» _____ 2008 г.

Учебно-методический комплекс

по дисциплине «Базы знаний и экспертные системы»

для студентов специальности

230102 – Автоматизированные системы обработки информации и управления;

Составитель: доцент кафедры ИУС Акилова И.М.

Факультет математики и информатики

Кафедра информационных и управляющих систем

2008 г.

*Печатается по решению
редакционно-издательского совета
факультета математики и информатики
Амурского государственного
университета*

И.М.Акилова

Базы знаний и экспертные системы для специальности 230102 «Автоматизированные системы обработки информации и управления»: учебно-методический комплекс дисциплины. / Акилова И.М. – Благовещенск. Изд-во Амурского гос. ун-та, 2008. 147 с.

Пособие содержит рабочую программу, курс лекций, методические рекомендации по проведению и выполнению лабораторных работ. Составлено в соответствии с требованиями государственного образовательного стандарта.

©Амурский государственный университет, 2008

©Кафедра информационных и управляющих систем, 2008

1. РАБОЧАЯ ПРОГРАММА

По дисциплине «Базы знаний и экспертные системы»

Для специальности 230102 – «Автоматизированные системы обработки информации и управления»

Курс 4 семестр 8

Лекции 30 (час.) Экзамен -

Практические (семинарские) занятия - (час.) Зачет 8 семестр

Лабораторные занятия 15 (час.)

Самостоятельная работа 34 (час.)

Всего часов 79 час.

Составитель: доцент Акилова И.М.

Факультет математики и информатики

Кафедра Информационных и управляющих систем

2006

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ, ЕЕ МЕСТО В УЧЕБНОМ ПРОЦЕССЕ

1.1. Цели и задачи дисциплины

Цель дисциплины – получение теоретических знаний и практического опыта по использованию экспертных систем и инженерии баз знаний.

В результате изучения курса студенты должны знать основные понятия экспертных систем, основные средства проектирования и разработки баз

знаний для экспертных систем. Способы получения вывода, методы обучения и объяснения принятого решения. Владеть навыками создания экспертных систем в прикладных задачах.

Дисциплина связана с предшествующими ей дисциплинами: "Моделирование систем", "Информационная технология", "Организация вычислительных систем", "Алгоритмические языки и программирование", "Технология программирования", "Организация баз данных", "Методы искусственного интеллекта в АСОИиУ".

2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1. Федеральный компонент

Программа курса «Базы знаний и экспертные системы» составлена в соответствии с требованиями государственного образовательного стандарта специализации – Автоматизированные системы обработки информации и управления, специализации 230102, блок обще-профессиональные дисциплины ОП-Д.Р.14.

2.2. Наименование тем, их содержание, объем в лекционных часах

ТЕМАТИЧЕСКИЙ ПЛАН ЛЕКЦИОННЫХ ЗАНЯТИЙ

№ темы	Наименование темы	Кол-во часов
1	Назначение и особенности методов искусственного интеллекта для разработки экспертных систем. Определение экспертной системы.	2
2	Структура экспертных систем и основные понятия. Общие сведения об архитектуре экспертных систем.	4
3	Классификация экспертных систем. Классификация инструментальных средств. Методология разработки экспертных систем.	4
4	Этапы разработки экспертных систем. Основные направления в изучении и практическом использовании.	4
5	Обобщенная схема ЕЯ - систем. Методы реализации ЕЯ - систем. Основные классы ЕЯ - систем. Системы общения с базами данных. Обзор промышленных ЕЯ - систем.	4
6	Основные положения систем речевого общения. Фонетическая и просодическая структуры речи. Информационная структура речевого сигнала. Классификация речевых	4

	процессоров. ДП - анализатор речевых команд. Форматный синтезатор речевых сигналов. Обзор промышленных систем речевого общения.	
7	Назначение, классификация и область применения систем обработки визуальной информации. Автоматизированные системы обработки изображений. Системы анализа изображений. Пакет программ для распознавания рукописной, символьной и графической информации.	4
8	Назначение машинного перевода. Периодизация и классификация систем машинного перевода. Лингвистическое обеспечение систем машинного перевода. Три подхода задач к решению задач машинного перевода. Перспективы развития систем машинного перевода.	4
Итого		30

1. ОСОБЕННОСТИ РАЗРАБОТКИ И ИСПОЛЬЗОВАНИЯ ЭКСПЕРТНЫХ СИСТЕМ - 14 часов.

Назначение и особенности методов искусственного интеллекта для разработки экспертных систем. Определение экспертной системы. Структура экспертных систем и основные понятия. Общие сведения об архитектуре экспертных систем. Пользовательский интерфейс, механизм вывода и база знаний. Классификация экспертных систем. Классификация инструментальных средств. Методология разработки экспертных систем. Этапы разработки экспертных систем. Основные направления в изучении и практическом использовании.

2. ЕСТЕСТВЕННО - ЯЗЫКОВЫЕ СИСТЕМЫ - 4 часа.

Обобщенная схема ЕЯ - систем. Методы реализации ЕЯ - систем. Основные классы ЕЯ - систем. Системы общения с базами данных. Обзор промышленных ЕЯ - систем.

3. СИСТЕМЫ РЕЧЕВОГО ОБЩЕНИЯ - 4 часа

Основные положения. Фонетическая и просодическая структуры речи. Информационная структура речевого сигнала. Классификация речевых процессоров. ДП - анализатор речевых команд. Форматный синтезатор речевых сигналов. Обзор промышленных систем речевого общения.

4. СИСТЕМЫ ОБРАБОТКИ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ - 4 часа

Назначение, классификация и область применения. Автоматизированные системы обработки изображений. Системы анализа изображений. Пакет программ для распознавания рукописной, символьной и графической информации.

5. СИСТЕМЫ МАШИННОГО ПЕРЕВОДА - 4 часа

Назначение машинного перевода. Периодизация и классификация систем машинного перевода. Лингвистическое обеспечение систем машинного перевода. Три подхода задач к решению задач машинного перевода. Перспективы развития систем машинного перевода.

2.3. Лабораторные занятия, их содержание и объем в часах.

ТЕМАТИЧЕСКИЙ ПЛАН ЛАБОРАТОРНЫХ ЗАНЯТИЙ

№ темы	Наименование темы	Кол-во часов
1	Разработка экспертных систем, базирующихся на правилах	6
2	Разработка экспертных систем, базирующихся на логике	6
3	Создания интерфейса общения пользователя с экспертной системой	3
	Всего	15

2.4. Самостоятельная работа студентов (34 часа)

Разработка фрагментов экспертных систем по темам:

1. Придумайте систему управления лифтом, предусматривающую кнопки для трех этажей, а также для закрытия и открытия дверей, сенсорный датчик для обнаружения препятствия, возникающего при закрытии дверей, реле-таймер для фиксации времени, в течение которого двери открыты, и отдельные кнопки вызова на каждом этаже. Напишите систему продукций для управления лифтом. Пример типичной продукции:

Если срок действия таймера истек и

дверь открыта и

ничто не препятствует закрытию двери

То закрыть дверь

2. Требуется разработать фрагмент экспертной системы, предназначенной

для прогнозирования местной погоды.

3. Требуется разработать фрагмент экспертной системы, предназначенной для обнаружения того, что делать, если автомобиль не заводится.

4. Требуется разработать фрагмент экспертной системы, предназначенной для определения стратегии гоночной яхты в регате. Система должна работать в режиме реального времени на протяжении всего периода гонок.

5. Требуется разработать фрагмент экспертной системы, предназначенной для помощи отвечающему по телефону доверия, когда отвечающий должен определить яд, который мог быть принят звонящим.

6. Требуется разработать фрагмент экспертной системы, предназначенной для определения оптимального маршрута продавца в любой данный день, причем продавец должен посетить всех клиентов и израсходовать возможное количество бензина.

7. Требуется разработать фрагмент экспертной системы, предназначенной для подбора членов экипажа с учетом их совместимости и возраста. При этом учитываются индивидуальные особенности претендентов, их совместимость, пожелания тех или иных кандидатов работать вместе, а также их возраст (командир должен быть старше всех остальных членов экипажа).

8. Поиск информации в базе данных из другого модуля программы. В базу знаний занесена информация о сотрудниках:

- а) фамилия;
- б) год рождения;
- в) место рождения;
- г) национальность;
- д) профессия;
- е) место работы;
- ж) занимаемая должность.

Система должна по вводимой с клавиатуры частичной информации о сотрудниках находить в базе знаний полную информацию и выводить ее на экран либо сообщать, что необходимая информация в базе данных отсутствует.

9. Требуется разработать фрагмент экспертной системы, предназначенной для поиска свободного места в салоне самолета. Система запрашивает требования к свободному месту

- а) класс салона;
- б) у окна или прохода;
- в) для некурящих или нет

и выводит на экран номер свободного места. Программа и данные содержатся в различных модулях.

10. Требуется разработать фрагмент экспертной системы, предназначенной для диагностики неисправностей персонального компьютера. Экспертная система должна исследовать ситуацию и попытаться определить на общем уровне, допускает ли ошибки пользователь или, действительно, имеется неисправность в системном блоке, на диске, в мониторе и т.д. Возможный путь проектирования – беседа с мастером–профессионалом. При оценке ситуации подразумеваются грубые функциональные тесты, без глубокого анализа электронных элементов.

11. Требуется разработать фрагмент экспертной системы, предназначенной для подбора субоптимальной конфигурации персонального компьютера с учетом субъективных и объективных потребностей заказчика.

12. Требуется разработать фрагмент экспертной системы, предназначенной для подбора субоптимальной конфигурации локальной компьютерной сети с учетом множества эксплуатационных, финансовых и прочих важных критериев.

13. Требуется разработать фрагмент экспертной системы, предназначенной для диагностики широко распространенных заболеваний человека по совокупности симптомов. Диагностируется не менее 20 болезней с учетом 15 типовых симптомов. Каждый симптом может указывать на несколько болезней (возможно, с разной степенью уверенности).

14. Требуется разработать фрагмент экспертной системы, предназначенной для консультации в отношении покупки автомобиля с учетом

субъективных факторов, объективных потребностей и платежеспособности клиента, а также сезона и др.

15. Требуется разработать фрагмент экспертной системы, предназначенной для консультации в отношении покупки недвижимости с учетом связанных с этим важных факторов (надежность продавца, платежеспособность покупателя, страхование сделки, изменение цен и банковских процентных ставок и др.).

2.5. Вопросы к зачету

1. Назначение и особенности методов ИИ для разработки ЭС
2. Структура и режимы экспертных систем
3. Характеристики ЭС
4. Классификация ЭС
5. Классификация инструментальных средств ЭС
6. Методология разработки ЭС
7. Этапы разработки ЭС
 - этап идентификации
 - этап концептуализации
 - этап формализации
 - этап выполнения
 - этап тестирования
 - этап опытной эксплуатации
8. Взаимодействия инженера по знаниям с экспертом
9. Трудности разработки ЭС
10. Проблемы и перспективы ЭС
11. Общие сведения об архитектуре ЭС
12. Программы поиска для ЭС
13. Управление с помощью эвристик
14. Ориентированный на человека диалог
15. Архитектура для автоматического рассуждения, основанная на правилах
16. Автоматическое рассуждение

17. Особенности реализации ЕЯ – систем. Роль ИИ в данных системах
18. Методы реализации ЕЯ – систем
19. Анализаторы, используемые в ЕЯ - системах
20. Основные положения систем речевого общения. Роль ИИ в данных системах
21. Принципы построения систем речевого общения
 - анализаторы речевых сообщений
 - синтезаторы речи
22. Фонетическая и просодическая структуры речи
23. Информационная структура речевого сигнала
24. Назначение, классификация и область применения систем переработки визуальной информации. Роль ИИ в данных системах
25. Автоматизированные системы обработки изображений
26. Системы анализа изображений
27. Системы машинной графики
28. Назначение систем машинного перевода. Роль ИИ в данных системах
29. Лингвистическое обеспечение систем машинного перевода
30. Грамматики и алгоритмы систем машинного перевода
31. Фильтровой и эвристический методы в системах машинного перевода
32. Математическое и программное обеспечение систем машинного перевода
33. Перспективы систем машинного перевода

2.6. Виды контроля

Для проверки эффективности преподавания дисциплины проводится контроль знаний студентов.

При этом используются следующие виды контроля:

- текущий контроль за аудиторной и самостоятельной работой обучаемых осуществляется во время проведения аудиторных занятий посредством устного опроса;

- промежуточный контроль осуществляется два раза в семестр в виде анализа итоговых отчетов;
- итоговый контроль в виде зачета осуществляется после успешного прохождения студентами текущего и промежуточного контроля и сдачи отчета по самостоятельной работе и устного экзамена при ответах экзаменуемого на два вопроса в билете и дополнительные вопросы по желанию экзаменатора.

2.7. Требования к знаниям студентов, предъявляемые на зачете

Для получения зачета студент должен посещать занятия, студентом должны быть выполнены все лабораторные работы, самостоятельное задание и отчет к нему, а также знание теоретического материала в объеме лабораторного курса.

«Зачтено» - студент не имеет задолженностей по семестровым отчетным работам. Хорошо владеет теоретическим материалом.

«Незачтено» - студент не отчитался по семестровым отчетным работам, не знает теоретический материал.

3. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА.

Основная

1. Введение в экспертные системы. Питер Джексон. Москва. Вильямс. 2001 г.
2. Справочник. Искусственный интеллект. Книга 1: Системы общения и экспертные системы.
3. Уотермен Д. Руководство по экспертным системам./ Пер. с англ. М.: Мир, 1989г.
4. "Обработка знаний" С. Осуга. Москва "Мир" 1989
5. "Представление и использование знаний" Х.Уэно, М.Исидзука, Москва "Мир" 1989

6. "Практическое введение в технологию искусственного интеллекта и экспертных систем с иллюстрациями на Бейсике", Р.Левин. Москва, "Финансы и статистика" 1990

7. "Искусство программирования на языке ПРОЛОГ", Л.Стрейлинг, Э.Шапиро. Москва "Мир" 1990

Дополнительная

1. Братко И. Программирование на языке Пролог для искусственного интеллекта./ Пер. с англ. М.:Мир, 1990 г.

2. "Турбо-ПРОЛОГ в сжатом изложении" А.Янсон, Издательство "Мир".

3. "Интеллектуализация ЭВМ" Ю.М.Смирнов, Москва "Высшая школа", 1989

4. "Как построить свою экспертную систему" К.Нейлор, Атомиздат 1991.

5. "Экспертные системы" Р.Форсайт, Москва "Радио и связь" 1987

6. "Построение экспертных систем", Ф.Хайтес-Рот, 1987

7."Экспертные системы: концепции и примеры", Элти, 1985

8."Моделирование и Экспертные системы", 1989

9."Извлечение и структурирование знаний для ЭС", Гаврилова Т.А. 1992

Методическое обеспечение

1. Акилова И.М. Разработка экспертных систем. Практикум. Благовещенск, 2000 г.

2. Акилова И.М., Назаренко Н. В. Методы представления и обработки знаний. Практикум. Благовещенск 2002 г.

4. Необходимое техническое и программное обеспечение

Лекции проводятся в стандартной аудитории, оснащенной в соответствии с требованиями преподавания теоретических дисциплин.

Для проведения лабораторных работ необходим компьютерный класс на 12-14 посадочных мест пользователей. В классе должен быть установлен язык программирования Пролог.

5. Учебно-методическая (технологическая) карта дисциплины

Номер недели	Номер темы	изучаемые на лекции Вопросы,	Занятия		Используемые наглядные и методические пособия	Самостоятельная работа студентов		Форма контроля
			Практические	Лабораторные		Содержание	Часы	
1	2	3	4	5	6	7	8	9
1	1	1-8	-	1	1-осн.	Изучение литературы по языку программирования Пролог	16	
2	2	1-4	-		1-осн.			
3	2	1-8	-	1	1-осн.			
4	3	1-6	-		1-осн.			
5	3	1-5	-	1	1-осн.			злр.
6	4	1-4	-		1-осн.			
7	4	1-6	-	2	1-осн.			
8	5	1-5	-		1-осн.			
9	5	1-3	-	2	1-осн.	Выполнение задания для самостоятельной работы		
10	6	4-5	-		1-осн.			
11	6	1-6	-	2	1-осн.		злр.	
12	7	1-5	-		1-осн.		14	
13	7	1-2	-	3	1-осн.			
14	8	1-5	-		1-осн.	сб.	4	
15	8	1-4	-	3	1-осн.			злр.

Условные обозначения:

осн. – основная литература

злр - защита лабораторной работы

сб. - собеседование по результатам самостоятельной работы студентов

2. ГРАФИК САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Содержание	Объем в часах	Сроки и форма контроля
Изучение возможностей логического языка программирования Пролог.	10 час.	Собеседование
Составление логической программы на Прологе	24 час.	Собеседование по отчету

3. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПРОВЕДЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

По выбранной теме студенты выполняют реферативную работу.

Работа включает следующие разделы:

1. Обоснование актуальности выбранной тематики и описание целей выполнения работы.
2. Систематизация и анализ найденных в научной печати, в сети Интернет и других источниках материалов.
3. Выводы.
4. Предложения по использованию результатов работы в конкретных областях и возможные направления дальнейших исследований.

5. Составление и отладка программы.
6. Составление отчета.

4. ПЕРЕЧЕНЬ УЧЕБНИКОВ, УЧЕБНЫХ ПОСОБИЙ

4.1. Перечень основной литературы

1. Введение в экспертные системы. Питер Джексон. Москва. Вильямс. 2001 г.
2. Справочник. Искусственный интеллект. Книга 1: Системы общения и экспертные системы.
3. Уотермен Д. Руководство по экспертным системам./ Пер. с англ. М.: Мир, 1989г.
4. "Обработка знаний" С. Осуга. Москва "Мир" 1989
5. "Представление и использование знаний" Х.Уэно, М.Исидзука, Москва "Мир" 1989
6. "Практическое введение в технологию искусственного интеллекта и экспертных систем с иллюстрациями на Бейсике", Р.Левин. Москва, "Финансы и статистика" 1990
7. "Искусство программирования на языке ПРОЛОГ", Л.Стрейлинг, Э.Шапиро. Москва "Мир" 1990

4.2. Перечень дополнительной литературы

1. Братко И. Программирование на языке Пролог для искусственного интеллекта./ Пер. с англ. М.: Мир, 1990 г.
2. "Турбо-ПРОЛОГ в сжатом изложении" А.Янсон, Издательство "Мир".

3. "Интеллектуализация ЭВМ" Ю.М.Смирнов, Москва "Высшая школа", 1989
4. "Как построить свою экспертную систему" К.Нейлор, Атомиздат 1991.
5. "Экспертные системы" Р.Форсайт, Москва "Радио и связь" 1987
6. "Построение экспертных систем", Ф.Хайтес-Рот, 1987
7. "Экспертные системы: концепции и примеры", Элти, 1985
8. "Моделирование и Экспертные системы", 1989
9. "Извлечение и структурирование знаний для ЭС", Гаврилова Т.А. 1992

4.3. Перечень методической литературы

1. Акилова И.М. Разработка экспертных систем. Практикум. Благовещенск, 2000 г.
2. Акилова И.М., Назаренко Н. В. Методы представления и обработки знаний. Практикум. Благовещенск 2002 г.

5. КРАТКИЙ КОНСПЕКТ ЛЕКЦИЙ

Лекция №1 Тема: Понятие экспертной системы. Смысл экспертного анализа.

Компьютерную программу можно назвать экспертом, если:

1) если такая программа *обладает знаниями*;

2) знания, которыми обладает программа, должны быть сконцентрированы на *определенную предметную область*. Случайный набор имен, дат и мест событий и т.п. – это отнюдь не те знания, которые могут послужить основой для программы, претендующей, претендующей на способность выполнить экспертный анализ. Знания предполагают определенную организацию и интеграцию – то есть определенные сведения должны соотноситься друг с другом и образовывать нечто вроде цепочки, в которой одно звено "тащит" за собой следующее;

3) из этих знаний должно непосредственно вытекать *решение проблем*.

Экспертная система – это программа для компьютера, которая оперирует со знаниями в определенной предметной области с целью выработки рекомендаций или решения проблем.

Экспертная система может полностью взять на себя функции, выполнение которых обычно требует привлечения опыта человека-специалиста, или играть роль ассистента для человека, принимающего решение. Другими словами, система (техническая или социальная), требующая принятия решения, может получить его

непосредственно от программы или через промежуточное звено – человека, который общается с программой. Тот, кто принимает решение, может быть экспертом со своими собственными правами, и в этом случае программа может "оправдать" свое существование, повышая эффективность его работы. Альтернативный вариант – человек, работающий в сотрудничестве с такой программой, может добиться с ее помощью результатов более высокого качества. Вообще говоря, правильное распределение функций между человеком и машиной является одним из ключевых условий высокой эффективности внедрения экспертных систем.

Технология экспертных систем является одним из направлений новой области исследования, которая получила наименование *искусственного интеллекта*. Исследования в этой области сконцентрированы на разработке и внедрении компьютерных программ, способных эмулировать (имитировать, воспроизводить) те области деятельности человека, которые требуют мышления, определенного мастерства и накопленного опыта. К ним относятся задачи принятия решений, распознавания образов и понимания человеческого языка. Эта технология уже успешно применяется в некоторых областях техники и жизни общества – органической химии, поиске полезных ископаемых, медицинской диагностике. Перечень типовых задач, решаемых экспертными системами, включает:

- ◆извлечение информации из первичных данных (таких как сигналы, поступающие от гидролокатора);
- ◆диагностика неисправностей (как в технических системах, так и в человеческом организме);
- ◆структурный анализ сложных объектов (например, химических соединений);
- ◆выбор конфигурации сложных многокомпонентных систем (например, распределенных компьютерных систем);
- ◆планирование последовательности выполнения операций, приводящих к заданной цели (например, выполняемых промышленными роботами).

Характеристики экспертных систем

Экспертная система отличается от прочих прикладных программ наличием следующих признаков:

- ◆ Моделирует не столько физическую (или иную) природу определенной проблемной области, сколько *механизм мышления человека* применительно к решению задач в этой проблемной области. Это существенно отличает экспертные системы от систем математического моделирования или компьютерной анимации. Нельзя сказать, что программа полностью воспроизводит психологическую модель специалиста в этой предметной области (эксперта), но важно, что основное внимание все-таки уделяется воспроизведению компьютерными средствами методики решения проблем, которая применяется экспертом, т.е. выполнению некоторой части задач так же (или даже лучше), как это делает эксперт.
- ◆ Система, помимо выполнения вычислительных операций, формирует определенные *соображения и выводы, основываясь на тех знаниях*, которыми она располагает. Знания в системе представлены, как правило, на некотором специальном языке и хранятся отдельно от собственно программного кода, который и формирует выводы и соображения. Этот компонент программы принято называть *базой знаний*.
- ◆ При решении задач основными являются *эвристические и приближенные методы*, которые, в отличие от алгоритмических, не всегда гарантируют успех. Эвристика по существу, является *правилом влияния {rule of thumb}*, которое в машинном виде представляет некоторое знание, приобретенное человеком по мере накопления практического опыта решения аналогичных проблем. Такие методы являются *приблизительными* в том смысле, что, во-первых, они не требуют исчерпывающей исходной информации, и, во-вторых, существует определенная степень уверенности (или неуверенности) в том, что предлагаемое решение является верным.

Экспертные системы отличаются и от других видов программ из области искусственного интеллекта.

- ◆ Экспертные системы имеют дело с предметами *реального мира*, операции с

которыми обычно требуют наличия значительного опыта, накопленного человеком. Множество программ из области искусственного интеллекта являются сугубо исследовательскими и основное внимание в них уделяется абстрактным математическим проблемам или упрощенным вариантам реальных проблем (иногда их называют "игрушечными" проблемами), а целью выполнения такой программы является "повышение уровня интуиции" или отработка методики. Экспертные системы имеют ярко выраженную практическую направленность в научной или коммерческой области.

♦Одной из основных характеристик экспертной системы является ее *производительность*, т.е. скорость получения результата и его достоверность (надежность). Исследовательские программы искусственного интеллекта могут и не быть очень быстрыми, можно примириться и с существованием в них отказов в отдельных ситуациях, поскольку, в конце концов, — это инструмент исследования, а не программный продукт. А вот экспертная система должна за приемлемое время найти решение, которое было бы не хуже, чем то, которое может предложить специалист в этой предметной области.

♦Экспертная система должна обладать способностью *объяснить*, почему предложено именно такое решение, и *доказать его обоснованность*. Пользователь должен получить всю информацию, необходимую ему для того, чтобы быть уверенным, что решение принято "не с потолка". В отличие от этого, исследовательские программы "общаются" только со своим создателем, который и так (скорее всего) знает, на чем основывается ее результат. Экспертная система проектируется в расчете на взаимодействие с разными пользователями, для которых ее работа должна быть, по возможности, прозрачной.

Зачастую термин *система, основанная на знаниях* используется в качестве синонима термина *экспертная система*, хотя, строго говоря, экспертная система – это более широкое понятие. Система, основанная на знаниях, – это любая система, процесс работы которой основан на применении правил от-

ношений к символическому представлению знаний, а не на использовании алгоритмических или статистических методов. Таким образом, программа, способная рассуждать о погоде, будет системой, основанной на знаниях, даже в том случае, если она не способна выполнить метеорологическую экспертизу. А вот чтобы иметь право называться метеорологической экспертной системой, программа должна быть способна давать прогноз погоды (другой вопрос – насколько он будет достоверен).

Суммируя все сказанное, отметим — экспертная система содержит знания в определенной предметной области, накопленные в результате практической деятельности человека (или человечества), и использует их для решения проблем, специфичных для этой области. Этим экспертные системы отличаются от прочих, "традиционных" систем, в которых предпочтение отдается более общим и менее связанным с предметной областью теоретическим методам, чаще всего математическим. Процесс создания экспертной системы часто называют *инженерией знаний* и он рассматривается в качестве "применения методов искусственного интеллекта".

Базовые функции экспертных систем

Поскольку теория экспертных систем выросла из более общей концепции искусственного интеллекта, то нет ничего удивительного в том, что проблематика этих областей имеет много общего.

1. Приобретение знаний

Приобретение знаний – это передача потенциального опыта решения проблемы от некоторого источника знаний и преобразование его в вид, который позволяет использовать эти знания в программе.

Передача знаний выполняется в процессе достаточно длительных и пространственных собеседований между специалистом по проектированию экспертной системы (будем в дальнейшем называть его *инженером по знаниям*) и экспертом в определенной предметной области, способным достаточно четко сформулировать имеющийся у него опыт. По существующим оценкам, таким методом можно сформировать от двух до пяти "элементов знания" (например, пра-

вил влияния) в день. Конечно, это очень низкая скорость, а потому многие исследователи рассматривают функцию приобретения знаний в качестве одного из главных "узких мест" технологии экспертных систем.

Причинами такой низкой производительности являются:

- специалисты в узкой области, как правило, пользуются собственным жаргоном, который трудно перевести на обычный "человеческий" язык. Но смысл жаргонного "словечка" отнюдь не очевиден, а потому требуется достаточно много дополнительных вопросов для уточнения его логического или математического значения. Например, специалисты по военной стратегии говорят об "агрессивной демонстрации" иностранной военной мощи, но при этом не могут объяснить, чем такая "агрессивная" демонстрация отличается от демонстрации, не несущей угрозы;

- факты и принципы, лежащие в основе многих специфических областей знания эксперта, не могут быть четко сформулированы в терминах математической теории или детерминированной модели, свойства которой хорошо понятны. Так, эксперту в финансовой области может быть известно, что определенные события могут стать причиной роста или снижения котировок на фондовой бирже, но он ничего вам не скажет точно о механизмах, которые приводят к такому эффекту, или о количественной оценке влияния этих факторов. Статистические модели могут помочь сделать общий долговременный прогноз, но, как правило, такие методы не работают в отношении курсов конкретных акций на коротких временных интервалах.

- для того чтобы решить проблему в определенной области, эксперту недостаточно просто обладать суммой знаний о фактах и принципах в этой области. Например, опытный специалист знает, какого рода информацией нужно располагать для формулировки того или иного суждения, насколько надежны различные источники информации и как можно расчленить сложную проблему на более простые, которые можно решать более или менее независимо. Выявить в процессе собеседования такого рода знания, основанные на личном опыте и плохо поддающиеся формализа-

ции, значительно сложнее, чем получить простой перечень каких-то фактов или общих принципов.

- экспертный анализ даже в очень узкой области, выполняемый человеком, очень часто нужно поместить в довольно обширный контекст, который включает и многие вещи, кажущиеся эксперту само собой разумеющимися, но для постороннего отнюдь таковыми не являющиеся. Возьмем для примера эксперта-юриста, который принимает участие в судебном процессе. Очень трудно очертить количество и природу знаний общего рода, которые оказываются вовлечены в расследование того или иного дела.

Неудовлетворительные результаты подобных собеседований пробудили у некоторых исследователей интерес к автоматизации процесса передачи знаний специалистом машине. Одно из направлений исследований в этой области — *автоматизированное извлечение знаний* — появилось как побочный продукт в развитии систем человеко-машинного диалога. Другие исследователи полагают, что "расшить" это узкое место можно, двигаясь по пути *машинного обучения*. Идея состоит в том, чтобы машина училась решать проблемы примерно так, как учится человек.

2. Представление знаний

Представление знаний— еще одна функция экспертной системы. Теория представления знаний — это отдельная область исследований, тесно связанная с философией формализма и когнитивной психологией. Предмет исследования в этой области — методы ассоциативного хранения информации, подобные тем, которые существуют в мозгу человека. При этом основное внимание, естественно, уделяется логической, а не биологической стороне процесса, опуская подробности физических преобразований.

В 70-х годах исследования в области представления знаний развивались в направлениях раскрытия принципов работы памяти человека, создания теорий извлечения сведений из памяти, распознавания и восстановления. Некоторые из достигнутых в теории результатов привели к созданию компьютерных программ, которые моделировали различные способы связывания понятий (*кон-*

цептов). Появились компьютерные приложения, которые могли некоторым образом отыскивать нужные "элементы" знания на определенном этапе решения некоторой проблемы. Со временем психологическая достоверность этих теорий отошла на второй план, а основное место, по крайней мере с точки зрения проблематики искусственного интеллекта, заняла их способность служить инструментом для работы с новыми информационными и управляющими структурами.

В области экспертных систем представление знаний интересует нас в основном как средство отыскания методов формального описания больших массивов полезной информации с целью их последующей обработки с помощью *символических вычислений*. *Формальное описание* означает упорядочение в рамках какого-либо языка, обладающего достаточно четко формализованным *синтаксисом* построения выражений и такого же уровня семантикой, увязывающей смысл выражения с его формой.

Символические вычисления означают выполнение нечисловых операций, в которых могут быть сконструированы символы и символьные структуры для представления различных концептов и отношений между ними. В области искусственного интеллекта ведется интенсивная работа по созданию *языков представления*. Под этим термином понимаются компьютерные языки, ориентированные на организацию описаний объектов и идей, в противовес статическим последовательностям инструкций или хранению простых элементов данных. Основными критериями доступа к представлению знаний являются логическая адекватность, эвристическая мощь и естественность, органичность нотации. Эти термины, скорее всего, нуждаются в пояснениях.

♦ *Логическая адекватность* означает, что представление должно обладать способностью распознавать все отличия, которые вы закладываете в исходную сущность. Например, невозможно представить идею, что каждое лекарство имеет какой-либо побочный нежелательный эффект, если только нельзя будет провести отличие между *предназначением* конкретного лекарственного препарата и его *побочным эффектом* (например,

аспирин усугубляет язвенную болезнь). В более общем виде выражение, передающее этот эффект, звучит так: "каждое лекарство обладает нежелательным побочным эффектом, специфическим для этого препарата".

♦ *Эвристическая мощьность* означает, что наряду с наличием выразительного языка представления должно существовать некоторое средство использования представлений, сконструированных и интерпретируемых таким образом, чтобы с их помощью можно было решить проблему. Часто оказывается, что язык, обладающий большей выразительной способностью в терминах количества семантических отличий, оказывается и больше сложным в управлении описанием взаимосвязей в процессе решения проблемы. Способность к выражению у многих из найденных формализмов может оказаться достаточно ограниченной по сравнению с английским языком или даже стандартной логикой. Часто уровень эвристической мощьности рассматривается по результату, т.е. по тому, насколько легко оказывается извлечь нужное знание применительно к конкретной ситуации. Знать, какие знания более всего подходят для решения конкретной проблемы, — это одно из качеств, которое отличает действительно специалиста, эксперта в определенной области, от новичка или просто начитанного человека.

♦ *Естественность нотации* следует рассматривать как некую добродетель системы, поскольку большинство приложений, построенных на базе экспертных систем, нуждается в накоплении большого объема знаний, а решить такую задачу довольно трудно, если соглашения в языке представления слишком сложны. Любой специалист скажет вам, что при прочих равных характеристиках лучше та система, с которой проще работать. Выражения, которыми формально описываются знания, должны быть по возможности простыми для написания, а их смысл должен быть понятен даже тому, кто не знает, как же компьютер интерпретирует эти выражения. Примером может служить *декларативный* программный код, который сам по себе дает достаточно четкое представление о про-

цессе его выполнения даже тому, кто не имеет представления о деталях реализации компьютером отдельных инструкций.

За прошедшие годы было предложено немало соглашений, пригодных для кодирования знаний на языковом уровне. Среди них отметим *порождающие правила, структурированные объекты и логические программы*. В большинстве экспертных систем используется один или несколько из перечисленных формализмов, а доводы в пользу и против любого из них до сих пор представляют собой тему для оживленных дискуссий среди теоретиков.

3. Управление процессом поиска решения

При проектировании экспертной системы серьезное внимание должно быть уделено и тому, как осуществляется доступ к знаниям и как они используются при поиске решения. Знание о том, какие знания нужны в той или иной конкретной ситуации, и умение ими распорядиться — важная часть процесса функционирования экспертной системы. Такие знания получили наименование *метазнаний* — т.е. знаний о знаниях. Решение нетривиальных проблем требует и определенного уровня *планирования и управления* при выборе, какой вопрос нужно задать, какой тест выполнить, и т.д.

Использование разных стратегий перебора имеющихся знаний, как правило, оказывает довольно существенное влияние на характеристики эффективности программы. Эти стратегии определяют, каким способом программа отыскивает решение проблемы в некотором пространстве альтернатив. Как правило, не бывает так, чтобы данные, которыми располагает программа работы с базой знаний, позволяли точно "выйти" на ту область в этом пространстве, где имеет смысл искать ответ.

Большинство формализмов представления знаний может быть использовано в разных *режимах управления*, и разработчики экспертных систем продолжают экспериментировать в этой области.

4. Разъяснение принятого решения

Вопрос о том, как помочь пользователю понять структуру и функции некоторого сложного компонента программы, связан со сравнительно новой обла-

стью взаимодействия человека и машины, которая появилась на пересечении таких областей, как искусственный интеллект, промышленная технология, физиология и эргономика. На сегодня вклад в эту область исследователей, занимающихся экспертными системами, состоит в разработке методов представления информации о поведении программы в процессе формирования цепочки логических заключений при поиске решения.

Представление информации о поведении экспертной системы важно по многим причинам.

◆ *Пользователи*, работающие с системой, нуждаются в подтверждении того, что в каждом конкретном случае заключение, к которому пришла программа, в основном корректно.

◆ *Инженеры, имеющие дело с формированием базы знаний*, должны убедиться, что сформулированные ими знания применены правильно, в том числе и в случае, когда существует прототип.

◆ *Экспертам в предметной области* желательно проследить ход рассуждений и способ использования тех сведений, которые с их слов были введены в базу знаний. Это позволит судить, насколько корректно они применяются в данной ситуации.

◆ *Программистам*, которые сопровождают, отлаживают и модернизируют систему, нужно иметь в своем распоряжении инструмент, позволяющий заглянуть в "ее нутро" на уровне более высоком, чем вызов отдельных языковых процедур.

◆ *Менеджер системы*, использующей экспертную технологию, который в конце концов несет ответственность за последствия решения, принятого программой, также нуждается в подтверждении, что эти решения достаточно обоснованы.

Способность системы объяснить методику принятия решения иногда называют *прозрачностью* системы. Под этим понимается, насколько просто персоналу выяснить, что делает программа и почему. Эту характеристику системы следует рассматривать в совокупности с режимом управления, поскольку

ку последовательность этапов принятия решения тесно связана с заданной стратегией поведения.

Отсутствие достаточной прозрачности поведения системы не позволит эксперту повлиять на ее производительность или дать совет, как можно ее повысить. Прослеживание и оценка поведения системы — задача довольно сложная и для ее решения необходимы совместные усилия эксперта и специалиста по информатике.

Лекция № 2 Тема: Приобретение знаний экспертными системами

Термин *приобретение знаний* носит обобщенный характер и совершенно нейтрален к способу передачи знаний. Например, передача может осуществляться с помощью специальной программы, которая в процессе обработки большого массива историй болезни устанавливает связь между симптомами и заболеваниями. А вот термин *извлечение знаний* относится именно к одному из способов передачи знаний – опросу экспертов в определенной проблемной области, который выполняется аналитиком или *инженером по знаниям*. Последний затем создает компьютерную программу, представляющую такие знания (или поручает это кому-нибудь другому, обеспечивая его всей необходимой информацией).

Этот же термин применяется и для обозначения процесса взаимодействия эксперта со специальной программой, целью которого является:

- ◆ извлечь каким-либо систематическим способом знания, которыми обладает эксперт, например, предлагая эксперту репрезентативные задачи и фиксируя предлагаемые способы их решения;
- ◆ сохранить полученные таким образом знания в некотором промежуточном виде;
- ◆ преобразовать знания из промежуточного представления в вид, пригодный для практического использования в программе, например в набор порождающих правил.

Преимущество использования такой программы – снижение трудоемкости процесса, поскольку перенос знаний от эксперта к системе осуществляется в один прием.

Теоретический анализ процесса приобретения знаний

В лекции 1 отмечалось, что при извлечении знаний в ходе опроса экспертов за рабочий день удастся сформулировать от двух до пяти "эквивалентов порождающих правил". Причин такой низкой производительности несколько:

- ◆ прежде чем приступить к опросу экспертов, инженер по знаниям, который не является специалистом в данной предметной области, должен потратить довольно много времени на ознакомление с ее спецификой и терминологией; только после этого процесс опроса может стать продуктивным;
- ◆ эксперты склонны думать о знакомой им области не столько в терминах общих принципов, сколько в терминах отдельных типических объектов, событий и их свойств;
- ◆ для представления специфических знаний о предметной области нужно подобрать подходящую систему обозначений и структурную оболочку, что само по себе является непростой задачей.

Как известно, любую сложную задачу лучше всего разбить на подзадачи, и именно так мы поступим с задачей приобретения знаний.

Стадии приобретения знаний

В работе [Buchanan et al, 1983] предлагается выполнить анализ процесса приобретения знаний в терминах модели процесса проектирования экспертной системы (рис. 1).

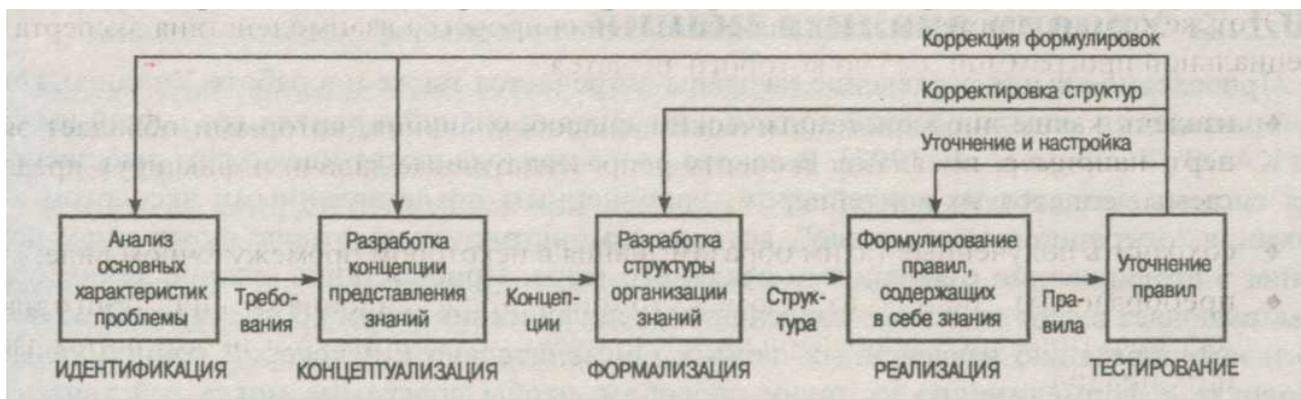


Рис. 1. Стадии приобретения знаний

(1) **Идентификация.** Анализируется класс проблем, которые предполагается решать с помощью проектируемой системы, включая данные, которыми нужно оперировать, и критерии оценки качества решений. Определя-

ются ресурсы, доступные при разработке проекта, — источники экспертных знаний, трудоемкость, ограничения по времени, стоимости и вычислительным ресурсам.

(2)**Концептуализация.** Формулируются базовые концепции и отношения между ними. Сюда же входят и характеристика различных видов используемых данных, анализ информационных потоков и лежащих в их основе структур в предметной области в терминах причинно-следственных связей, отношений частное/целое, постоянное/временное и т.п.

(3)**Формализация.** Предпринимается попытка представить структуру пространства состояний и характер методов поиска в нем. Выполняется оценка полноты и степени достоверности (неопределенности) информации и других ограничений, накладываемых на логическую интерпретацию данных, таких как зависимость от времени, надежность и полнота различных источников информации.

(4)**Реализация.** Преобразование формализованных знаний в работающую программу, причем на первый план выходит спецификация методов организации управления процессом и уточнение деталей организации информационных потоков. Правила преобразуются в форму, пригодную для выполнения программой в выбранном режиме управления. Принимаются решения об используемых структурах данных и разбиении программы на ряд более или менее независимых модулей.

(5)**Тестирование.** Проверка работы созданного варианта системы на большом числе репрезентативных задач. В процессе тестирования анализируются возможные источники ошибок в поведении системы. Чаще всего таким источником является имеющийся в системе набор правил. Оказывается, что в нем не хватает каких-то правил, другие не совсем корректны, а между некоторыми обнаруживается противоречие.

Как видно из рис. 1, проектирование экспертной системы начинается с анализа класса проблем, которые предполагается решать с помощью этой системы. Было бы ошибкой приступать к проектированию системы, заранее задавшись

определенной концепцией или определенной структурной организацией знаний. Весьма сомнительно, чтобы тот вариант концепции или тот способ организации идей, которым мы задались априори, взяв за основу предыдущие разработки, был приложим и к новой предметной области.

Уровни анализа знаний

Приведенное выше разделение на этапы встречается также и в работе Уилинги, который разработал моделирующий подход к инженерии знаний в рамках созданной им среды KADS (1992). В основе этого подхода лежит идея о том, что экспертная система является не контейнером, наполненным представленными экспертом знаниями, а "операционной моделью", которая демонстрирует некоторое нужное нам поведение в столкновении с явлениями реального мира. Приобретение знаний, таким образом, включает в себя не только извлечение специфических знаний о предметной области, но и интерпретацию извлеченных данных применительно к некоторой концептуальной оболочке и формализацию их таким способом, чтобы программа могла действительно использовать их в процессе работы.

В основу оболочки KADS положено пять базовых принципов.

- 1) Использование множества моделей, позволяющее преодолеть сложность процессов инженерии знаний.
- 2) Четырехуровневая структура для моделирования требуемой *экспертности* — набора качеств, лежащих в основе высокого уровня работы специалистов.
- 3) Повторное использование родовых компонентов модели в качестве шаблонов, поддерживающих нисходящую стратегию приобретения знаний.
- 4) Процесс дифференциации простых моделей в сложные.
- 5) Важность преобразования моделей экспертности с сохранением структуры в процессе разработки и внедрения.

Рассмотрим подробно два первых принципа.

Главным мотивом создания оболочки KADS было преодоление сложности знаний. На сегодняшний день у инженеров по знаниям имеется возможность исполь-

зовать при построении экспертных систем множество самых разнообразных методов и технологий. Однако при этом остаются три основных вопроса:

- ◆ *определение проблемы*, которую необходимо решить с помощью экспертной системы;
- ◆ *определений функций*, которые возлагаются на экспертную систему применительно к этой проблеме;
- ◆ *определение задач*, которые необходимо решить для выполнения возложенной функции.

Первый из принципов, положенных в основу KADS, состоит в том, что оболочка должна содержать множество частных моделей, помогающих найти ответ на эти вопросы. Примерами таких моделей могут служить:

- ◆ *организационная модель* "социально-экономической среды", в которой должна функционировать система, например финансовые услуги, здравоохранение и т.п.;
- ◆ *прикладная модель* решаемой проблемы и выполняемой функции, например диагностика, планирование расписания работ и т.д.;
- ◆ *модель задач*, демонстрирующая, как должна выполняться специфицированная функция, для чего производится ее разбиение на отдельные задачи, например сбор данных о доходах, формирование гипотез о заболеваниях.

Между этой терминологией и той, которой пользовался Бучанан, нет прямого соответствия, но можно сказать, что организационная и прикладная модели аналогичны стадии идентификации в предложенной Бучананом структуре.

В подходе, который реализован при создании KADS, стадия "концептуализации" разбивается на две части: *модель кооперации*, или *коммуникации*, и *модель экспертности*. Первая отвечает за декомпозицию процесса решения проблемы, формирование набора простейших задач и распределение их между исполнителями, в качестве которых могут выступать и люди, и машины. Вторая модель представляет процесс, который обычно называется извлечением знаний, т.е. анализ разных видов знаний, которые эксперт использует в ходе решения проблемы.

Кроме указанных, в состав оболочки KADS входит еще и *модель проектирования*, включающая технологии вычислений и механизмы представления знаний, которые могут быть использованы для реализации спецификаций, сформулированных предыдущими моделями.

На первый взгляд кажется, что представленный выше анализ в какой-то степени смазывает отличие между стадиями концептуализации и формализации. Можно, конечно, и возразить, что стадия формализации представляет собой просто более детальную проработку концепций и отношений, выявленных на ранних стадиях. Модель проектирования частично включает то, что в прежней схеме было отнесено к стадии реализации, но она не предполагает создание выполняемой программы.

В своей ранней работе Уилинга немного по-другому проводил разграничение между уровнями анализа (1986). Он рассматривал четыре уровня анализа.

- ◆ *Концептуализация знаний*. На этом уровне предполагалось формальное описание знаний в терминах принципиальных концепций и отношений между концепциями.
- ◆ *Уровень эпистемологического анализа*. Целью такого анализа было выявление структурных свойств концептуальных знаний, в частности таксономических отношений.
- ◆ *Уровень логического анализа*. Основное внимание уделялось тому, как строить логический вывод в данной предметной области на основе имеющихся знаний.
- ◆ *Уровень анализа внедрения*. Исследовались механизмы программной реализации системы.

В более поздней разработке три первых уровня включены в состав модели экспертности, а уровень анализа внедрения — в модель проектирования. Четырехуровневая структура KADS согласуется с предложенной Кленси схемой разделения знаний различного вида в соответствии с их ролью в процессе решения проблем. В частности, знания, касающиеся конкретной предметной области, теперь разделены на знания более высокого уровня (знания, относящиеся к по-

строению логического вывода в этой предметной области), знания выбора решаемых задач и знания стратегии решения задач.

Эти уровни знаний представлены в табл. 1.

Категория знаний	Организация	Виды знаний
Стратегическая	Стратегии	Планы, метаправила
Задача	Задачи	Цели, управляющие термы, структуры задач
Логический вывод	Структура логического вывода	Источники знаний, метаклассы, схема предметной области
Предметная область	Теория предметной области	Концепции, свойства, отношения

Стратегический уровень управляет процессом выполнения задач, использующих при решении проблем методы логического вывода, подходящие для конкретной предметной области, и знания из этой области.

Описанная схема дифференциации знаний приводит к довольно простой архитектуре экспертной системы. В частности, оказывается, что даже в рамках традиционной архитектуры, предполагающей наличие базы знаний и машины логического вывода, можно неявным образом включить задачи и стратегии и в структуру знаний о предметной области, и в механизм построения логических заключений. Явное выделение этих задач и стратегий является главным моментом как в процессе приобретения знаний, так и в процессе проектирования структуры экспертной системы.

Онтологический анализ

Александр и его коллеги предложили еще один уровень анализа знаний, который получил название онтологического анализа. В основе этого подхода лежит описание системы в терминах сущностей, отношений между ними и преобразования сущностей, которое выполняется в процессе решения некоторой задачи. Автор указанной работы используют для структурирования знаний о предметной области три основные категории:

- ◆ *статическая онтология* — в нее входят сущности предметной области, их свойства и отношения;
- ◆ *динамическая онтология* — определяет состояния, возникающие в процессе

решения проблемы, и способ преобразования одних состояний в другие;

♦ *эпистемическая онтология* — описывает знания, управляющие процессом перехода из одного состояния в другое.

В этой схеме просматривается совершенно очевидное соответствие с уровнями *концептуализации знаний* и *эпистемологического анализа* в структуре, предложенной в уже упоминавшейся работе. Но на нижних уровнях — *логического анализа* и *анализа внедрения* — такое соответствие уже не просматривается. Онтологический анализ предполагает, что решаемая проблема может быть сведена к проблеме поиска, но при этом не рассматривается, каким именно способом нужно выполнять поиск. Примером практического применения такого подхода является система OPAL.

Рассматриваемая схема онтологического анализа выглядит довольно абстрактной, но ее ценность в том, что она упрощает анализ плохо структурированных задач. Каждый, кто сталкивался с выявлением знаний в процессе опроса человека-эксперта, знает, как трудно найти подходящую схему организации таких знаний. Чаще всего в таких случаях говорят: "Давайте воспользуемся фреймами или системой правил", откладывая таким образом выбор подходящего метода реализации на будущее, когда природа знаний эксперта станет более понятна.

Оболочки экспертных систем

На раннем этапе становления экспертных систем проектирование каждой очередной системы начиналось практически с нуля, в том смысле, что проектировщики для представления знаний и управления их применением использовали самые примитивные структуры данных и средства управления, которые содержались в обычных языках программирования. В редких случаях в существующие языки программирования включались специальные языки представлений правил или фреймов.

Такие специальные языки, как правило, обладали двумя видами специфических средств:

♦ *модулями* представления знаний (в виде правил или фреймов);

♦ *интерпретатором*, который управлял активизацией этих модулей.

Совокупность модулей образует базу знаний экспертной системы, а интерпретатор является базовым элементом машины логического вывода.

Эти компоненты могут быть повторно используемыми, т.е. служить основой для создания экспертных систем в разных предметных областях. Использование этих программ в качестве базовых компонентов множества конкретных экспертных систем позволило называть их *оболочкой* системы.

Лекция № 3 Тема: Оболочки экспертных систем. Система EMYCIN

Примером такой оболочки может служить система EMYCIN, которая является предметно-независимой версией системы MYCIN, т.е. это система MYCIN, но без специфической медицинской базы знаний. По мнению разработчиков, EMYCIN вполне может служить "скелетом" для создания консультационных программ во многих предметных областях, поскольку располагает множеством инструментальных программных средств, облегчающих задачу проектировщика конкретной экспертной консультационной системы. Она особенно удобна для решения дедуктивных задач, таких как диагностика заболеваний или неисправностей, для которых характерно большое количество ненадежных входных измерений (симптомов, результатов лабораторных тестов и т.п.), а пространство решений, содержащее возможные диагнозы, может быть достаточно четко очерчено.

Некоторые программные средства, впервые разработанные для EMYCIN, в дальнейшем стали типовыми для большинства оболочек экспертных систем. Среди таких средств следует отметить следующие.

-Язык представления правил. В системе EMYCIN такой язык использует систему обозначений, аналогичную языку ALGOL. Этот язык, с одной стороны, более понятен, чем LISP, а с другой — более строг и структурирован, чем тот диалект обычного английского, который использовался в MYCIN.

-Индексированная схема применения правил, которая позволяет группировать правила, используя в качестве критерия группировки параметры, на которые ссылаются эти правила.

-Использование обратной цепочки рассуждений в качестве основной стратегии управления. Эта стратегия оперирует с И/ИЛИ-деревом, чьи листья представляют собой данные, которые могут быть найдены в таблицах или запрошены пользователем.

- Интерфейс между консультационной программой, созданной на основе EMYCIN, и конечным пользователем. Этот компонент оболочки обрабатывает все сообщения, которыми обмениваются пользователь и т.п.

- Интерфейс между разработчиком и программой, обеспечивающий ввод и редактирование правил, редактирование знаний, представленных в форме таблиц, тестирование правил и выполнение репрезентативных задач.

Значительная часть интерфейса реализуется отдельным компонентом EMYCIN — программой TEIRESIAS. Эта программа представляет собой "редактор знаний", который упрощает редактирование и сопровождение больших баз знаний. Редактор проверяет синтаксическую корректность правил, анализирует взаимную непротиворечивость правил в базе знаний и следит за тем, чтобы новое правило не являлось частным случаем существующих. Противоречие возникает, когда два правила с одинаковыми антецедентами имеют противоречивые консеквенты. Одно правило является частью другого в том случае, когда совокупность условий антецедента одного правила представляет собой подмножество совокупности условий другого правила, а их консеквенты одинаковы. Но в состав TEIRESIAS не включены знания о какой-либо конкретной предметной области или о стратегии решения проблем, которая может быть использована в проектируемой экспертной системе.

Используемые в системе методы синтаксического анализа могут быть применены к правилам, относящимся к любой предметной области.

Сопровождение и редактирование баз знаний с помощью программы TEIRESIAS

Как правило, человек-эксперт знает о той предметной области, в которой он является специалистом, гораздо больше, чем может выразить на сло-

вах. Вряд ли можно добиться от него многого, задавая вопросы в общем виде, например: "Что вам известно об инфекционных заболеваниях крови?" Гораздо продуктивнее подход, реализованный в про грамме TIRHSIAS, который предполагает вовлечение эксперта в решение несложных репрезентативных задач из определенной предметной области и извлечение необходимых знаний в процессе такого решения.

Задавшись определенным набором базовых правил, представляющих прототип экспертной системы, TEIRESIAS решает в соответствии с этими правилами какую-нибудь из сформулированных репрезентативных проблем и предлагает эксперту критиковать результаты. В ответ эксперт должен сформулировать новые правила и откорректировать введенные ранее, а программа отслеживает внесенные изменения, анализирует их на предмет сохранения целостности и непротиворечивости всего набора правил, используя при этом *модели правил*. В процессе анализа используется обобщение правил различного вила.

Модели правил являются, по существу метаправилами, поскольку они предназначены для выработки суждений о правилах, а не об объектах предметной области приложения. В частности, в программе TEIRESIAS имеются метаправила, относящиеся к атрибутам правил объектного уровня. Такие правила обращают внимание пользователя на то, что в данных обстоятельствах целесообразно сначала исследовать определенные параметры, а уж затем в процессе отладки набора правил пытаться отслеживать влияние других параметров.

Существуют также средства, помогающие эксперту добавить новые варианты типов данных. Ошибки, которые обычно возникают при решении подобных задач, состоят в том, что новые типы данных имеют структуру, не согласующуюся с типами, уже существующими в системе.

Абстрактные данные, которые используются для формирования новых экземпляров структур данных, называются *схемами*. Эти схемы представляют собой обобщенные описания типов данных, точно так же, как структуры данных являются обобщениями конкретных данных. Следовательно, схемы также могут

быть организованы в виде иерархической структуры, в которой каждая схема, во-первых, наследует атрибуты, ассоциированные с ее предшественницей в иерархии, а во-вторых, имеет еще и собственные дополнительные атрибуты.

В таком случае процесс создания нового типа данных включает прослеживание пути от корня иерархии схем к той схеме, которая представляет интересующий нас тип данных. На каждом уровне имеются атрибуты, которые нужно конкретизировать, причем процесс продолжается до тех пор, пока не будет конкретизирована вся структура. Отношения между схемами в иерархии определяют последовательность выполнения задач обновления структур данных в системе.

Таким образом, в программе TLIRHSIAS можно выделить три уровня обобщения:

- знания об объектах данных, специфические для предметной области;
- знания о типах данных, специфические для метода представления знаний;
- знания, независимые от метода представления.

В составе TEIRESIAS имеются и средства, которые помогают оболочке EMYCIN следить за поведением экспертной системы в процессе применения набора имеющихся правил.

-Режим объяснения (EXPLAIN). После выполнения каждого очередного задания консультации — система дает объяснение, как она пришла к такому заключению. Распечатываются каждое правило, к которому система обращалась в процессе выполнения задания, и количественные параметры, связанные с применением этого правила, в том числе и коэффициенты уверенности.

-Режим тестирования (TEST). В этом режиме эксперт может сравнить результаты, полученные при протоне отлаживаемой программы, с правильными результатами решения этой же задачи, хранящимися в специальной базе данных, и проанализировать имеющиеся отличия. Оболочка EMYCIN позволяет эксперту задавал системе вопросы, почему она пришла к тому или иному заключению и почему при этом не были получены извест-

ные правильные результаты.

- *Режим просмотра (REVIEW)*. В этом режиме эксперт может просмотреть выводы, к которым приходила система при выполнении одних и тех же запросов и библиотеки типовых задач. Это помогает просмотреть эффект, который дают изменения, вносимые в набор правил в процессе наладки системы. В этом же режиме можно проанализировать, как отражаются изменения в наборе правил на производительности системы.

Система EMYCIN была одной из первых попыток создать программный инструмент позволяющий перенести архитектору экспертной системы, уже эксплуатируемой в одной предметной области, на другие предметные области. Опыт, полученный в процессе работы с EMYCIN, показал, что те инструментальные средства, которые были включены в состав EMYCIN, пригодны для решения одних проблем и мало что дают при решении других.

Использование опроса экспертов для извлечения знаний в системе COM-PASS

Для переключения номеров в телефонной сети используется довольно сложная система, которая может занимать большую часть здания телефонной станции. Основная задача при обслуживании системы переключений минимизировать число вызовов, которые необходимо перебросить на запасные маршруты из-за неисправности основных линий подключений, и быстро восстановить работу всей системы. Неисправность линий подключения может быть вызвана отказом каких-либо электронных схем, обеспечивающих связь между парой абонентов.

В процессе работы в системе переключения непрерывно выполняется самотестирование. При этом проверяется, нет ли разрыва в цепях короткого замыкания, замедления срабатывания переключающих схем и т.д. При возникновении каких-либо нестандартных ситуаций система самотестирования формирует соответствующее сообщение. Причина появления неисправности в системе переключения может быть выявлена только на основании множества таких сообщений, причем на помощь приходит опыт специалистов-экспертов. Эти со-

общения поступают в экспертную систему COMPASS, которая может предложить провести какой-либо специальный дополнительный тест или заменить определенный узел в системе (реле или плату). Система разработана компанией GTE и эксплуатируется во множестве ее филиалов (1990).

Накопление в системе знаний экспертов осуществлялось в процессе опроса. Эксперты описывали применяемые ими эвристические способы поиска неисправности, а инженеры по знаниям формулировали их в виде правил "если ... то". Затем эксперты повторно анализировали результаты формализации и проверяли, насколько эти правила согласуются с их опытом и интуицией. При обнаружении разночтений инженеры по знаниям изменяли формулировку правил и совместными усилиями с экспертами добивались, чтобы правила были приемлемыми.

Обычно правила вводились в систему в виде одного или нескольких производящих правил на языке КЕЕ. Сформулированные на английском языке правила накапливались в библиотеке "документированных знаний", которая являлась одним из компонентов комплекта документации экспертной системы. Эта библиотека помогала сохранить "первоисточник знаний", что очень помогло в процессе настройки и опытной эксплуатации системы.

В процессе приобретения знаний большое внимание, по крайней мере, на первых порах, уделялось моделированию применения правил при поиске неисправностей "вручную", т.е. с помощью карандаша и бумаги. Цикл приобретения знаний при разработке системы COMPASS включал следующие этапы.

(1) В процессе собеседования с экспертом извлечь определенные знания.

(2) Задокументировать извлеченные знания.

(3) Проверить новые знания:

- попробовать применить их на разных наборах данных;
- смоделировать вручную, к каким результатам приведет использование этих знаний;
- сравнить результаты моделирования с теми, которые должны получить-

ся по мнению эксперта;

- если результаты отличаются, то определить, какие именно правила и процедуры внесли "наибольший вклад" в это отличие; вернуться к п. (1) и выяснить у эксперта, как следует скорректировать подозрительное правило или процедуру.

Графически циклическая процедура приобретения знаний представлена на рис. 1.

После того как объем накопленных знаний превысит некоторый минимум, можно проверять работу системы на практике. При этом между этапами документирования и проверки знаний появляется еще один — внедрение знаний в систему. После этого можно проверять адекватность новых знаний не только моделированием вручную, но и выполнением программы на разных наборах входных данных. Конечно, анализ и сравнение результатов при этом усложняются, поскольку на ошибки в процессе формализации могут накладываться и ошибки реализации правил в работающей программе.

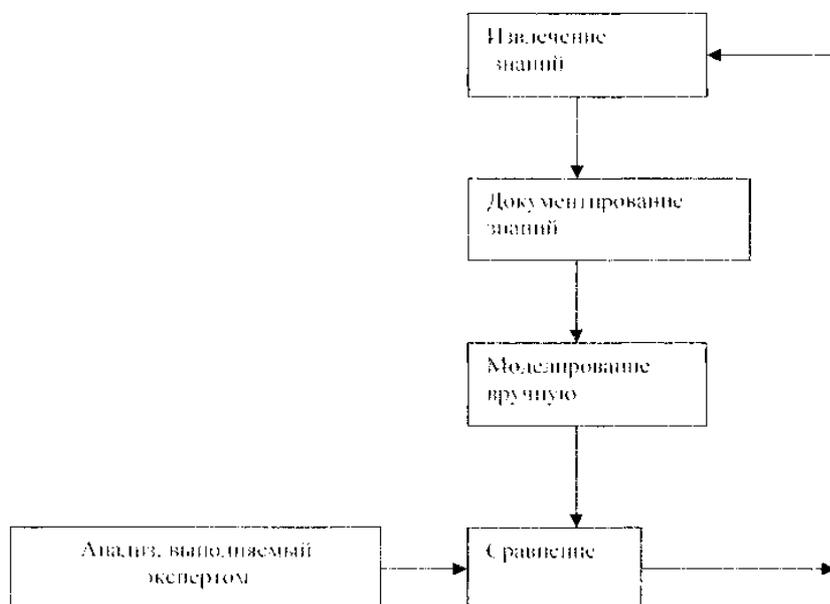


Рис. 1.

Преро (Prerau), ведущий разработчик системы, отметил, что по мере накопления опыта в процессе извлечения знаний инженеру по знаниям легче было общаться с экспертами. Последние постепенно освоились с методикой формализации знаний в виде правил, а инженер по знаниям достаточно глубо-

ко ознакомился со спецификой предметной области. Такое сближение "стилей мышления" можно было рассматривать как признак успешного хода работы над проектом. Определенную помощь в этом, по наблюдению Преро, сыграло совместное участие инженера по знаниям и эксперта в ручном моделировании процесса принятия решений на основе полученных знаний и последующей проверке результатов.

Автоматизация процесса извлечения знаний в системе OPAL

Проект COMPASS можно считать одним из наиболее ярких примеров использования традиционной методики приобретения знаний, базирующейся на соответствующим образом организованном опросе экспертов. Такая методология "выросла" из предложенной Пьюэллом и Саймоном методики *анализа протокола*. Остановимся на проекте OPAL, в котором использована другая методика, отличающаяся от традиционной в двух важных аспектах.

-Эта методика ориентирована на частичную автоматизацию процесса извлечения знаний в ходе активного диалога интервьюируемого эксперта с программой.

Методика приобретения знаний предполагает использование стратегии, направляемой знаниями о предметной области.

Программа OPAL, пытается "вытянуть" из пользователя как можно больше деталей, касающихся представления знаний и их использования. OPAL не является программой общего назначения. Она разработана специально для диагностики онкологических заболеваний и предназначена для формирования правил принятия решений на основе полученных от эксперта знаний о планах лечения в том или ином случае.

Программа OPAL, упрощает процесс извлечения знаний, предназначенных для использования в экспертной системе ONCOCIN. Последняя формирует план лечения больных онкозаболеваниями и заинтересована в использовании модели предметной области для получения знаний непосредственно от эксперта с помощью средств графического интерфейса. Понятие *модель предмет-*

ной области можно трактовать в терминах знаний различного вида, которыми обладает эксперт.

Независимо от того, о какой конкретной предметной области идет речь, игре в шахматы или медицинской диагностике, всегда существуют некоторые предварительные условия или предварительный опыт, которыми должен обладать субъект или техническая система, чтобы воспринимать знания об этой предметной области. Если речь идет об игре в шахматы, то по крайней мере нужно знать правила этой игры. Применительно к медицинской диагностике нужно иметь представление о пациентах, заболеваниях, клинических тестах и т.п. Этот вид *фоновых*, или *фундаментальных*. знаний иногда в литературе по экспертным системам называют *глубокими знаниями*, противопоставляя их *поверхностным знаниям*, которые представляют собой хаотичный набор сведений о связях "стимул-реакция".

OPAL представляет собой программу извлечения знаний, которая обладает некоторыми фундаментальными знаниями в области терапии онкологических заболеваний. Программа использует эти базовые знания в процессе диалога с экспертом для извлечения дополнительных, более детальных знаний. Знания о предметной области нужны программе и для того, чтобы преобразовать информацию, полученную с терминала в процессе диалога, в исполняемый код — порождающие правила или таблицу состояний. Такая комбинация процесса наращивания знаний и их компиляции является одной из наиболее привлекательных возможностей той методологии построения экспертных систем, которая положена в основу системы OPAL. Графически основная идея представлена на рис. 2, где на человека-эксперта возлагается задача расширения и уточнения модели предметной области. Эта модель затем компилируется в программу, состоящую из процедур и порождающих правил. Поведение программы снова анализируется экспертом, который при необходимости вносит коррективы в модель и замыкает таким образом цикл итеративного процесса.



Рис. 2. Процесс приобретения знаний с использованием модели предметной области

В экспертной системе ONCOCIN используются три разных метода представления знаний:

- иерархия объектов, представляющая протоколы и их компоненты, в частности медикаменты;
- порождающие прибила, которые связаны с фреймами и формируют заключения о значениях медицинских параметров в процессе уточнения плана;
- таблицы конечных состояний представляют собой последовательности терапевтических курсов (назначение и использование этих таблиц будет описано ниже).

Включение в систему ONCOCIN нового протокола влечет за собой формирование иерархии, которая представляет его компоненты, связывание подходящих порождающих правил с новыми объектами и заполнение таблицы конечных состояний, которая определяет порядок назначения определенных компонентов курса лечения. Программа OPAL формирует элементы нового протокола в процессе "собеседования" с экспертом с помощью средств графического интерфейса. При этом полученные знания преобразуются сначала в промежуточную форму представления, а затем транслируются в формат, используемый в системе ONCOCIN. На последней стадии формируются соответствующие порождающие правила. Для упрощения реализации промежуточных стадий, трансляции и формирования порождающих правил в программе OPAL используется модель предметной области лечения онкологических заболеваний.

В модели предметной области можно выделить четыре основных аспекта, которые явились следствием применения онтологического анализа.

- *Сущности и отношения.* Сущностями в этой предметной области являются элементы (компоненты) курса лечения — назначаемые медикаменты. Эти сущности образуют часть статической онтологии предметной области. Большая часть знаний о предметной области касается атрибутов альтернативных медикаментов, например доз и их приема. Отношения между элементами курса лечения довольно запутаны в том смысле, что они связывают различные уровни спецификации в плане лечения. Так, медикаменты могут быть частью химиотерапии, а химиотерапия может быть частью протокола.

- *Действия в предметной области.* При заданных отношениях между элементами для уточнения плана приема медикаментов потребуется обращение к перечню планов. Другими словами, уточнение плана является неявным в иерархической организации сущностей предметной области. Таким образом, модель предметной области в OPAL позволяет сконцентрировать основное внимание на задачах, а не на используемых методах поиска. Однако может потребоваться изменить планы для отдельных пациентов, например изменить дозировку или заменить один препарат другим. Такие концепции, как изменение дозировки или замена препаратов в курсе лечения, образуют часть *динамической онтологии* предметной области.

- *Предикаты предметной области.* Этот аспект модели касается условий, при которых обращаются к модификации назначенного плана лечения. Сюда могут входить результаты лабораторных анализов и проявления у пациента определенных симптомов (например, токсикоз на определенные препараты). Такие знания образуют часть *эпистемической онтологии* предметной области, т.е. эти знания направляют и ограничивают возможные действия. На уровне реализации правила, изменяющие курс лечения, основываются на этих условиях. Такие предикаты появляются в левой части порождающих правил

ONCOCIN. Подобное правило подключается к объекту в иерархии планирования таким образом, что оно применяется только в контексте определенной препарата или определенного курса химиотерапии в конкретном протоколе.

Процедурные знания. Поскольку планы курса лечения предполагают определенное расписание приема назначенных пациенту препаратов, знания о способе реализации протокола составляют существенную часть модели предметной области. Эти знаки позволяют программе OPAL извлекать информацию, которая в итоге направляется в таблицы конечных состояний, описывающие возможные последовательности этапов курса терапии, и таким образом образуют другую часть *эпистемической онтологии* предметной области. На уровне реализации программа OPAL использует для описания таких процедур специальный язык программирования, который позволяет эксперту представлять достаточно сложные алгоритмы, манипулируя пиктограммами на экране дисплея.

Используя эту модель, программа OPAL может извлекать и отображать в разной форме знания о планах лечения — в виде пиктограмм, представляющих отдельные элементы плана, формуляра, заполненного информацией об отдельных препаратах, в виде предложений специального языка, представляющих процедуры, связанные с реализацией плана лечения.

Сущности и отношения между ними вводятся с помощью экранных формуляров, в которых пользователь выбирает элементы из меню. Затем заполненный формуляр преобразуется в фрейм, причем отдельные поля формуляра образуют слоты фрейма, а введенные в них значения — значения слотов (*заполнители слотов*). Эти новые объекты затем автоматически связываются с другими объектами в иерархии. Например, медикаменты связываются с объектами курсов химиотерапии, компонентами которых они являются.

Операции предметной области также вводятся с помощью заполнения экранных формуляров. В этом случае формуляр представляет собой пустой шаблон плана, в котором представлены поля для назначения расписания приема препаратов, а меню возможных действий включает такие операции, как *изменение дозировки, временное прекращение приема* и т.д. Поскольку

список возможных действий довольно короткий, эта методика позволяет эксперту достаточно легко ввести нужную последовательность операции.

Процесс приобретения знания в значительной мере облегчается при использовании *языков визуального программирования*. Графический интерфейс позволяет пользователю создавать пиктограммы, представляющие элементы плана, и формировать из них графические структуры. Расставляя такие элементы на экране и вычерчивая связи между ними, пользователь формирует мнемоническую схему управления потоками, которая обычно представляется в виде программы на каком-нибудь языке программирования.

На последующих этапах такие программы преобразуются в таблицы конечных состояний, хорошо известные специалистам в области теории вычислительных машин. Для любого текущего состояния системы такая таблица позволяет определить, в какое новое состояние перейдет система, получив определенный набор входных сигналов, и какой набор выходных сигналов при этом будет сформирован.

Приобретение новых знаний на основе существующих

В ходе экспериментов по созданию интеллектуальных обучающих систем на основе технологии экспертных систем исследователи пришли к более глубокому пониманию того, какими видами знаний пользуется эксперт в процессе решения проблем. При создании инструментальных средств общего назначения, аналогичных EMYCIN и предназначенных для построения широкого класса экспертных систем, разработчики столкнулись с интересной проблемой: как преобразовать знания, имеющие отношение к любой проблемной области, во фреймы или порождающие правила.

Такие попытки заставили исследователей глубже проанализировать роль знаний о предметной области и специфических для нее правил логического вывода, в частности рассмотреть их с точки зрения разных стилей рассуждения, характерных для разных областей.

Программа извлечения знаний нуждается в некоторых базовых знаниях о той предметной области, в которой специализируется интервьюируемый эксперт. И точно такими же знаниями должен обладать инженер по знаниям. Только в этом случае он сможет достичь взаимопонимания и диалога с экспертом.

Вряд ли стоит надеяться на то, что со временем появится такая методика извлечения знаний у эксперта, которая будет одинаково эффективна в любой предметной области. Знания, которыми нужно обладать для того, чтобы воспринимать новые знания, можно рассматривать как метазнания. В основном к ним относятся знания о структуре и стратегии, включая информацию о методах классификации явлений и сущностей в определенной предметной области (например, заболеваний) и способах выбора альтернативных действий (например, курсов терапии). Существуют также и отдельные знания, необходимые для того, чтобы объяснить, почему получено именно такое, а не иное решение проблемы.

Извлечение знаний посредством опроса экспертов на основе модели предметной области — отнюдь не последнее слово в автоматизации этого процесса.

Лекция № 4 Тема: Классификация задач экспертных систем

В сборнике статей, опубликованном под общей редакцией Хейеса-Рота [1983], была предложена классификация экспертных систем, которая отражает специфику задач, решаемых с помощью этой технологии. С тех пор эта классификация неоднократно критиковалась различными авторами, в основном из-за того, что в ней были смешаны разные характеристики, а это привело к тому, что сформулированные категории нельзя рассматривать как взаимно исключающие. Тем не менее мы кратко представим эту классификацию и будем рассматривать ее как отправную точку для дальнейшего совершенствования.

**Интерпретирующие системы* предназначены для формирования описания ситуаций по результатам наблюдений или данным, получаемым от различного рода сенсоров. Типичные задачи, решаемые с помощью интерпретирующих систем,— распознавание образов и определение химической структуры вещества.

**Прогнозирующие системы* предназначены для логического анализа возмож-

ных последствий заданных ситуаций или событий. Типичные задачи для экспертных систем этого типа — предсказание погоды и прогноз ситуаций на финансовых рынках.

* *Диагностические системы* предназначены для обнаружения источников неисправностей по результатам наблюдений за поведением контролируемой системы (технической или биологической). В эту категорию входит широкий спектр задач в самых различных предметных областях — медицине, механике, электронике и т.д.

* *Системы проектирования* предназначены для структурного синтеза конфигурации объектов (компонентов проектируемой системы) при заданных ограничениях. Типичными задачами для таких систем является синтез электронных схем, компоновка архитектурных планов, оптимальное размещение объектов в ограниченном пространстве.

* *Системы планирования* предназначены для подготовки планов проведения последовательности операций, приводящей к заданной цели. К этой категории относятся задачи планирования поведения роботов и составление маршрутов передвижения транспорта.

* *Системы мониторинга* анализируют поведение контролируемой системы и, сравнивая полученные данные с критическими точками заранее составленного плана, прогнозируют вероятность достижения поставленной цели. Типовые области приложения таких систем — контроль движения воздушного транспорта и наблюдение за состоянием энергетических объектов.

* *Наладочные системы* предназначены для выработки рекомендаций по устранению неисправностей в контролируемой системе. К этому классу относятся системы, помогающие программистам в отладке программного обеспечения, и консультирующие системы.

* *Системы оказания помощи при ремонте* оборудования выполняют планирование процесса устранения неисправностей в сложных объектах, например в сетях инженерных коммуникаций.

* *Обучающие системы* проводят анализ знаний студентов по определенному предмету, отыскивают пробелы в знаниях и предлагают средства для их ликвидации.

* *Системы контроля* обеспечивают адаптивное управление поведением сложных человеко-машинных систем, прогнозируя появление возможных сбоев и планируя действия, необходимые для их предупреждения. Областью применения таких систем является управление воздушным транспортом, военными действиями и деловой активностью в сфере бизнеса.

Как уже упоминалось, множество исследователей отмечали наличие ряда существенных недостатков в приведенной классификации. Рейхгелт и Ван Гармелен обратили внимание на то, что некоторые из категорий в ней перекрываются или включают друг друга [1986]. Например, категорию *системы планирования* в этой классификации вполне можно рассматривать как составную часть категории *системы проектирования*, поскольку планирование можно трактовать как *проектирование последовательности операций* (на это, кстати, обратили внимание и авторы классификации [1983]). Кленси также задался вопросом: "Является ли автоматизация программирования проблемой *планирования* или *проектирования*!" [1985]. Совершенно очевидно, что подобное замечание можно высказать и по отношению к таким категориям, как *диагностические системы*, *системы мониторинга*, *системы оказания помощи при ремонте* и *обучающие системы*.

Кленси предложил альтернативный метод классификации, взяв за основу набор *родовых операций*, выполняемых в рассматриваемых системах. Вместо того чтобы пытаться разделить анализируемые программы решения проблем по признакам особенностей тех проблем, на решение которых они ориентированы, он предложил поставить во главу угла те виды операций, которые выполняются по отношению к реальной обслуживаемой системе (механической, биологической или электрической).

Кленси предложил разделять *синтетические операции*, результатом которых является изменение структуры (*конструкции*) системы, и *аналитические*

операции, которые *интерпретируют* характеристики и свойства системы, не изменяя ее как таковую. Эта обобщенная концепция может быть конкретизирована, в результате чего построена иерархическая схема видов операций, выполнение которых может быть затребовано от программы. На рис. 1 и 2 представлены такие иерархические схемы для аналитических и синтетических операций.



Рис. 1. Иерархия родовых аналитических операций



Рис. 2. Иерархия родовых синтетических операций

Классификация методов решения проблем

Классификация — это одна из наиболее распространенных проблем в любой предметной области. Например, эксперты в области ботаники или зоологии первым делом пытаются определить место в существующей таксономии для вновь открытого растения или животного. Как правило, система классов имеет явно выраженную иерархическую организацию, в которой подклассы обладают определенными свойствами, характерными для своих суперклассов, причем классы-соседи на одном уровне иерархии являются взаимно исключающими в отношении наличия или отсутствия определенных наборов свойств.

1. Эвристическое сопоставление

Кленси отметил, что одна из важнейших особенностей классификации состоит в том, что эксперт *выбирает* категорию из ряда возможных решений, которые можно заранее перечислить. Когда мы имеем дело с простыми вещами или явлениями, то для их классификации вполне достаточно бросающихся в глаза свойств объектов. Это позволяет почти мгновенно сопоставлять данные и категории. В более сложных случаях таких лежащих на поверхности

свойств может оказаться недостаточно для того, чтобы правильно определить место объекта в иерархической схеме классификации. В этом случае нам остается уповать на тот метод, который Кленси назвал *эвристической классификацией*. Суть его состоит в установлении неиерархических ассоциативных связей между данными и категориями классификации, которое требует выполнения промежуточных логических заключений, включающих, возможно, и концепции из *другой* таксономии.

На рис. 3 показаны три основных этапа выполнения эвристической классификации: абстрагирование от данных, сопоставление абстрактных категорий данных с абстрактными категориями решений (утолщенная стрелка) и конкретизация решения. Рассмотрим их по очереди.

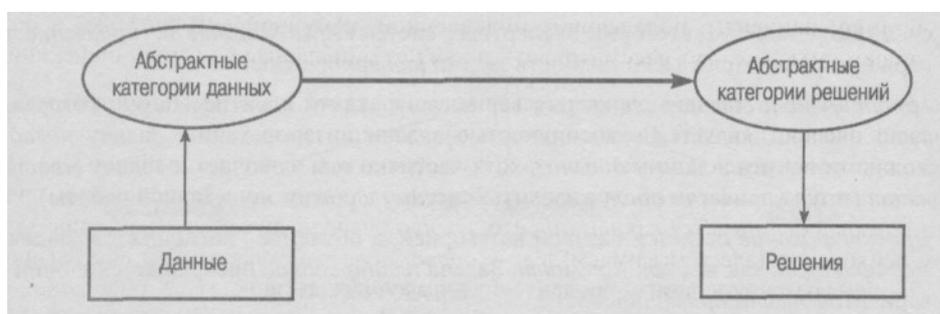


Рис. 3. Структура логических связей при эвристической классификации

**Абстрагирование от данных.* Часто бывает полезно абстрагироваться от данных, характеризующих конкретный случай. Так, при диагностировании заболевания за частую важно не столько то, что у пациента высокая температура (скажем, 39.8°), а то, что она *выше нормальной*. То есть врач обычно рассуждает в терминах *диапазона* температур, а не в терминах *конкретного ее значения*.

**Эвристическое сопоставление.* Выполнить сопоставление первичных данных в конкретном случае и окончательного диагноза довольно трудно. Гораздо легче сопоставить более абстрактные данные и достаточно широкий класс заболеваний. Например, повышенная температура может служить индикатором лихорадки, наводящей на мысль о инфекционном заражении. Данные "включают" гипотезы, но на относительно высоком уровне абстракции. Такой процесс сопоставления имеет ярко выраженный эвристический характер, поскольку соответствие между данными и гипотезами на любом

уровне не бывает однозначным и из общего правила может быть множество исключений. Анализ данных, которые "вписываются" в определенную абстрактную категорию, просто позволяет отбирать решения, лучше согласующиеся с абстрактами решений.

* *Конкретизация решений.* После того как определена абстрактная категория, которая сужает пространство решений, нужно определить в этом пространстве конкретные решения-кандидаты и каким-то образом их ранжировать. Это может потребовать дальнейших размышлений, в которые включаются уже количественные параметры данных, или даже сбора дополнительной информации. В любом случае целью этой процедуры является отбор "соревнующихся" гипотез в пространстве решений и последующее их ранжирование — сортировка по степени правдоподобия.

Кленси различает три варианта построения абстрактных категорий данных.

**Определительный.* В этом варианте в первую очередь рассматриваются характерные признаки класса объектов, и он во многом напоминает таксономический подход в ботанике и зоологии.

**Количественный.* В этом варианте абстрагирование выполняется исходя из количественных характеристик, как это было сделано в упоминавшемся выше примере с температурой пациента.

* *Обобщение.* Этот вариант основывается на иерархии характерных свойств. Например, пациенты, обладающие подавленной иммунной активностью, в более общем смысле могут рассматриваться как потенциальные носители инфекции.

На рис. 4 представлена эвристическая классификация в контексте программы медицинской диагностики MYCIN.

Исходными являются данные анализа крови пациента (количество лейкоцитов). Сначала выполняется *количественное* абстрагирование от конкретного значения этого показателя, который оценивается как *низкий*, что, в свою очередь, является характерным признаком лейкопении (здесь мы имеем дело с

определяющим вариантом абстрагирования). *Обобщение* лейкопении — подавленная иммунная активность, а обобщение последней — повышенная склонность к переносу инфекции (т.е. такие пациенты более подвержены воздействию различных микроорганизмов). Повышенная склонность к переносу инфекции является уже *родовой* категорией и наводит на мысль о наличии инфекции, вызванной граммотрицательными микроорганизмами (т.е. инфекции, связанной с *определённым классом* бактерий). Затем это родовое решение *конкретизируется* и предполагается, что источником инфекции являются бактерии E.Coli.

В системе MYCIN сопоставление данных и абстрактных категорий решений выполняется с помощью порождающих правил, а эвристическая природа такого сопоставления выражается коэффициентами уверенности. Эти коэффициенты можно рассматривать как заложенную в порождающее правило меру "строгости" соответствия между предпосылкой и выводом. Другие правила затем будут уточнять выполненное сопоставление и таким образом "подстраивать" коэффициент уверенности.

2. Общность эвристической классификации

Интересной особенностью методики анализа, предложенной Кленси, является то, что она подходит для большого спектра экспертных систем. Оказывается, что, несмотря на различие областей применения, многие экспертные системы функционируют, в принципе, одинаково. В своей статье Кленси продемонстрировал применение предложенной методики на множестве систем, помимо MYCIN.

Нужно отметить, что предложенная Кленси методика анализа фундаментальных аспектов решения проблем с помощью технологии экспертных систем не является единственной. Например, получило дальнейшее развитие понятие *родовой задачи*. Родовая задача представляет собой, по сути, спецификацию задачи, включающую описание различных форм знаний о предметной области и их организации для выполнения задачи вручную и набор режимов выполнения задачи.

Примерами таких задач являются иерархическая классификация, сопоставление или оценка гипотез и передача информации, направляемая знаниями.

Задача *иерархической классификации* включает отбор гипотез из иерархически организованного пространства альтернатив, а затем уточнение этих гипотез с учетом имеющихся данных. Гипотезами могут быть, например, заболевания или неисправности оборудования. Каждая такая гипотеза может быть порождена имеющимися данными, но прежде чем развивать ее дальше, необходимо выполнить определенные проверки или уточнить гипотезу таким образом, чтобы она соответствовала более широкому набору имеющихся фактов.

Под *сопоставлением гипотез* понимается выполнение количественной оценки свидетельств в пользу достоверности гипотез или того, насколько полным является соответствие между имеющимися данными и гипотезами. В такую оценку иногда включается априорная вероятность гипотез, поиск возможных конкурирующих гипотез и т.п. Все операции, связанные с обработкой коэффициентов уверенности в системе MYCIN, можно рассматривать как реализацию механизма родовых задач этого вида.

Передача информации, направляемой знаниями, проявляется в виде неочевидных логических связей, которые трудно классифицировать, но которые характерны для способа мышления эксперта. Например, врач не оставит без внимания тот факт, что пациент недавно перенес хирургическую операцию, поскольку, скорее всего, он подвергся анестезии, а это может косвенно сказаться на формировании диагноза. В логических рассуждениях такого типа используются "фоновые" знания о предметной области, а не те порождающие правила, с которыми мы связываем свое представление об экспертной системе медицинской диагностики.

Итак, одно из отличий между подходами Кленси и Чандрасекарана состоит в том, что для первого характерно стремление разбить решение проблемы на мелкие абстрактные категории. Подход Кленси приводит к поглощению этапа сопоставления гипотез эвристической классификацией, в то время как Чандрасекаран делает акцент на том, что обработка гипотез может рассматри-

ваться как самостоятельная родовая задача со своими собственными правами вне контекста классификации. Таким образом, появляется проблема, до какого уровня "зернистости" целесообразно доводить абстрагирование и какие критерии следует выбрать для оценки этой "грануляции".

Лекция № 5

Тема: Назначение и особенности методов искусственного интеллекта для разработки экспертных систем

Знания, которыми обладает специалист в какой-либо области (дисциплине), можно разделить на формализованные (точные) и неформализованные (неточные). *Формализованные знания* формулируются в книгах и руководствах в виде общих и строгих суждений (законов, формул, моделей, алгоритмов и т. п.), отражающих универсальные знания. *Неформализованные знания*, как правило, не попадают в книги и руководства в связи с их конкретностью, субъективностью и приблизительностью. Знания этого рода являются результатом обобщения многолетнего опыта работы и интуиции специалистов. Они обычно представляют собой многообразие эмпирических (эвристических) приемов и правил.

В зависимости от того, какие знания преобладают в той или иной области (дисциплине), ее относят к формализованным (если преобладают точные знания) или к неформализованным (если преобладают неточные знания) описательным областям. *Задачи*, решаемые на основе точных знаний, называют *формализованными*, а задачи, решаемые с помощью неточных знаний - *неформализованными*. (Речь идет не о неформализуемых, а о неформализованных задачах, т. е. о задачах, которые, возможно, и формализуемы, но эта формализация пока неизвестна).

Традиционное программирование в качестве основы для разработки программы использует алгоритм, т. е. формализованное знание. Поэтому до недавнего времени считалось, что ЭВМ не приспособлены для решения неформализованных задач. Расширение сферы использования ЭВМ показало, что неформализованные задачи составляют очень важный класс задач, вероятно значительно больший,

чем класс формализованных задач Неумение решать неформализованные задачи сдерживает внедрение ЭВМ в описательные науки. Основной задачей информатики является внедрение ее методов в описательные науки и дисциплины. На основании этого можно утверждать, что исследования в области ЭС занимают значительное место в информатике.

Особенности неформализованных задач:

- алгоритмическое решение задачи неизвестно (хотя, возможно, и существует) или не может быть использовано из-за ограниченности ресурсов ЭВМ (времени, памяти);
- задача не может быть определена в числовой форме (требуется символьное представление);
- цели задачи не могут быть выражены в терминах точно определенной целевой функции.

Как правило, неформализованные задачи обладают неполнотой, ошибочностью, неоднозначностью и (или) противоречивостью знаний (как данных, так и используемых правил преобразования)

Экспертные системы не отвергают и не заменяют традиционного подхода к программированию, они отличаются от традиционных программ тем, что ориентированы на решение неформализованных задач и обладают следующими особенностями:

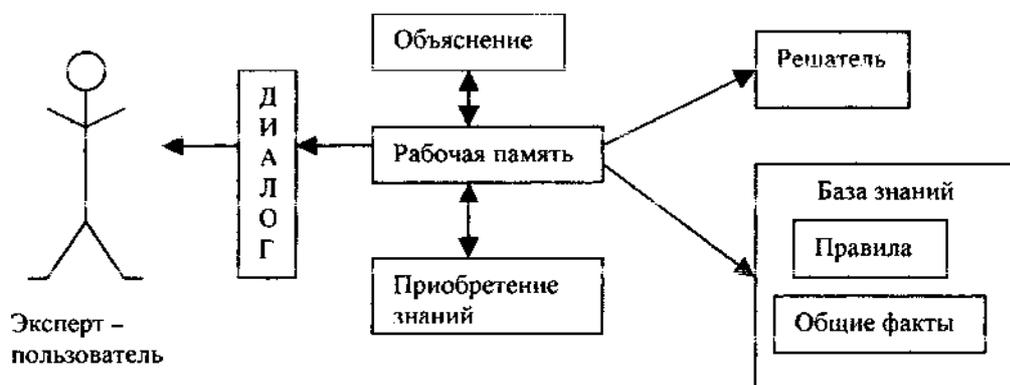
- алгоритм решений не известен заранее, а строится самой ЭС с помощью символических рассуждений, базирующихся на эвристических приемах;
- ясность полученных решений, т. е. система «осознает» в терминах пользователя, как она получила решение;
- способность анализа и объяснения своих действий и знаний;
- способность приобретения новых знаний от пользователя-эксперта, не знающего программирования, и изменения в соответствии с ними своего поведения;
- обеспечение «дружественного», как правило, естественно-языкового (ЕЯ) интерфейса с пользователем.

Обычно к ЭС относят *системы, основанные на знаниях*, т. е. системы, вычислительная возможность которых является в первую очередь следствием их наращиваемой базы знаний (БЗ) и только во вторую очередь определяется используемыми методами. *Методы инженерии знаний* (методы ЭС) в значительной степени инвариантны тому, в каких областях они могут применяться. Области применения ЭС весьма разнообразны: военные приложения, медицина, электроника, вычислительная техника, геология, математика, космос, сельское хозяйство, управление, финансы, юриспруденция и т. д. Более критичны методы инженерии знаний к типу решаемых задач. В настоящее время ЭС используются при решении задач следующих типов:

- принятие решений в условиях неопределенности (неполноты);
- интерпретация символов и сигналов;
- предсказание;
- диагностика;
- конструирование;
- планирование;
- управление;
- контроль.

Структура и режимы экспертных систем.

Типичная ЭС состоит из следующих основных компонентов: решателя (интерпретатора), рабочей памяти (РП), называемой также базой данных (БД), базы знаний (БЗ), компонентов приобретения знаний, объяснительного и диалогового.



База данных предназначена для хранения исходных и промежуточных данных решаемой в текущий момент задачи. Этот термин совпадает по названию, но не по смыслу с термином, используемым в информационно-поисковых системах (ИПС) и системах управления базами данных (СУБД) для обозначения всех данных (и в первую очередь не текущих, а долгосрочных), хранимых в системе.

База знаний в ЭС предназначена для хранения долгосрочных данных, описывающих рассматриваемую область (а не текущих данных), и правил, описывающих целесообразные преобразования данных этой области.

Решатель, используя исходные данные из РП и знания из БЗ, формирует такую последовательность правил, которые, будучи примененными к исходным данным, приводят к решению задачи.

Компонент приобретения знаний автоматизирует процесс наполнения ЭС знаниями, осуществляемый пользователем-экспертом.

Объяснительный компонент объясняет, как система получила решение задачи (или почему она не получила решения) и какие знания она при этом использовала, что облегчает эксперту тестирование системы и повышает доверие пользователя к полученному результату,

Диалоговый компонент ориентирован на организацию дружелюбного общения со всеми категориями пользователей, как в ходе решения задач, так и приобретения знаний, объяснения результатов работы

В разработке ЭС участвуют представители следующих специальностей:

- эксперт в той проблемной области, задачи которой будет решать ЭС;
- инженер по знаниям – специалист по разработке ЭС;
- программист - специалист по разработке инструментальных средств (ИС).

Необходимо отметить, что отсутствие среди участников разработки инженера по знаниям (т. е. его замена программистом) либо приводит к неудаче процесс создания ЭС, либо значительно удлиняет его.

Эксперт определяет знания (данные и правила), характеризующие проблемную область, обеспечивает полноту и правильность введенных в ЭС знаний.

Инженер по знаниям помогает эксперту выявить и структурировать знания, необходимые для работы ЭС, осуществляет выбор того инструментального средства, которое наиболее подходит для данной проблемной области и определяет способ представления знаний в этом инструментальном средстве, выделяет и программирует (традиционными средствами) стандартные функции (типичные для данной проблемной области), которые будут использоваться в правилах, вводимых экспертом.

Программист разрабатывает ИС, содержащее в пределе все основные компоненты ЭС, осуществляет сопряжение ИС с той средой, в которой оно будет использовано.

Экспертная система работает в двух режимах: приобретения знаний и решения задач (режим консультации или режим использования ЭС).

В *режиме приобретения знаний* общение с ЭС осуществляет через посредничество инженера по знаниям эксперт. Эксперт описывает проблемную область в виде совокупности данных и правил. Данные определяют объекты, их характеристики и значения, существующие в области экспертизы. Правила определяют способы манипулирования данными, характерные для рассматриваемой проблемной области. Эксперт, используя компонент приобретения знаний, наполняет систему знаниями, которые позволяют ЭС в режиме решения самостоятельно (без эксперта) решать задачи из проблемной области.

Важную роль в режиме приобретения знаний играет объяснительный компонент. Именно благодаря ему эксперт на этапе тестирования локализует причины неудачной работы ЭС, что позволяет эксперту целенаправленно модифицировать старые или вводить новые знания. Обычно объяснительный компонент сообщает следующее: как правила используют информацию пользователя; почему использовались или не использовались данные или правила; какие были сделаны выводы и т. п. Все объяснения делаются, как правило, на ограниченном естественном языке или языке графики.

Отметим, что режиму приобретения знаний при традиционном подходе к разработке программ соответствуют этапы алгоритмизации, программирования

и отладки, выполняемые программистом. Таким образом, в отличие от традиционного подхода разработку программ осуществляет эксперт (с помощью ЭС), не владеющий программированием, а не программист.

В *режиме консультации* общение с ЭС осуществляет конечный пользователь, которого интересует результат и (или) способ получения решения. Пользователь в зависимости от назначения ЭС может не быть специалистом в данной проблемной области, в этом случае он обращается к ЭС за советом, не умея получить ответ сам, или быть специалистом, в этом случае он обращается к ЭС, чтобы либо ускорить процесс получения результата, либо возложить на ЭС рутинную работу. Термин «пользователь» является многозначным, так как кроме конечного пользователя применять ЭС может и эксперт, и инженер по знаниям, и программист. Поэтому, когда хотят подчеркнуть, что речь идет о том, для кого делалась ЭС, используют термин «конечный пользователь».

В режиме консультации данные о задаче пользователя обрабатываются диалоговым компонентом, который выполняет следующие действия:

- распределяет роли участников (пользователя и ЭС) и организует их взаимодействие в процессе кооперативного решения задачи;
- преобразует данные пользователя о задаче, представленные на привычном для пользователя языке, во внутренний язык системы;
- преобразует сообщения системы, представленные на внутреннем языке, в сообщения на языке, привычном для пользователя (обычно это ограниченный естественный язык или язык графики).

После обработки данные поступают в РП. На основе входных данных из РП, общих данных о проблемной области и правил из БЗ решатель (интерпретатор) формирует решение задачи.

В отличие от традиционных программ ЭС в режиме решения задачи не только исполняет предписанную последовательность операций, но и предварительно формирует ее. Если ответ ЭС не понятен пользователю, то он может потребовать объяснения, как ответ получен.

ХАРАКТЕРИСТИКИ ЭС.

Экспертные системы как любой сложный объект можно определить только совокупностью характеристик

Выделим следующие характеристики ЭС:

- 1) назначение;
- 2) проблемная область;
- 3) глубина анализа проблемной области;
- 4) тип используемых методов и знаний;
- 5) класс системы;
- 6) стадия существования;
- 7) инструментальные средства (ИС).

Перечисленный набор характеристик не претендует на полноту (в связи с отсутствием общепринятой классификации), а определяет ЭС как целое, не выделяя отдельных компонентов (способ представления знаний, решения задачи и т. п.).

Назначение определяется следующей совокупностью параметров: цель создания ЭС—для обучения специалистов, для решения задач, для автоматизации рутинных работ, для тиражирования знаний экспертов и т. п.; основной пользователь—не специалист в области экспертизы, специалист, учащийся.

Проблемная область может быть определена совокупностью параметров: предметной областью и задачами, решаемыми в предметной области, каждый из которых может рассматриваться с точки зрения как конечного пользователя, так и разработчика ЭС.

С точки зрения пользователя, предметную область можно характеризовать описанием области в терминах пользователя, включающим наименование области, перечень и взаимоотношение подобластей и т. п., а задачи, решаемые существующими ЭС, — их типом. Обычно выделяют следующие *типы задач*:

интерпретация символов или сигналов — составление смыслового описания по входным данным;

предсказание—определение последствий наблюдаемых ситуаций;

диагностика—определение неисправностей (заболеваний) по симптомам;

конструирование—разработка объекта с заданными свойствами при соблюдении установленных ограничений:

планирование — определение последовательности действий, приводящих к желаемому состоянию объекта;

слежение — наблюдение за изменяющимся состоянием объекта и сравнение его показателей с установленными или желаемыми;

управление—воздействие на объект для достижения желаемого поведения.

С точки зрения разработчика целесообразно выделять *статические* и *динамические предметные области*. Предметная область называется статической, если описывающие ее исходные данные не изменяются во времени (точнее, рассматриваются как не изменяющиеся за время решения задачи). Статичность области означает неизменность описывающих ее исходных данных. При этом производные данные (выводимые из исходных) могут и появляться заново, и изменяться (не изменяя, однако, исходных данных). Если исходные данные, описывающие предметную область, изменяются за время решения задачи, то предметную область называют динамической. Кроме того, предметные области можно характеризовать следующими аспектами: числом и сложностью сущностей; их атрибутами и значениями атрибутов; связностью сущностей и их атрибутов; полнотой знаний; точностью знаний (знания точны или правдоподобны: правдоподобность знаний представляется некоторым числом или высказыванием).

Решаемые задачи, с точки зрения разработчика ЭС, также можно разделить на статические и динамические. Будем говорить, что ЭС решает *динамическую* или *статическую* задачу, если процесс решения задачи изменяет или не изменяет исходные данные о текущем состоянии предметной области.

В подавляющем большинстве существующих ЭС исходят из предположения статичности предметной области и решают статические задачи, будем называть такие ЭС *статическими*. ЭС, которые имеют дело с динамическими предметными областями и решают статические или динамические задачи, будем называть *динамическими*. В настоящее время начинают появляться первые динамические ЭС. Решение многих важнейших практических неформализованных за-

дач возможно *только с помощью динамических, а не статических ЭС*. Следует подчеркнуть, что на традиционных (числовых) последовательных ЭВМ с помощью существующих методов инженерии знаний можно решать только статические задачи, а для решения динамических задач, составляющих большинство реальных приложений, необходимо использовать специализированные символьные ЭВМ.

Решаемые задачи, кроме того, можно характеризовать следующими аспектами:

- числом и сложностью правил, используемых в задаче;
- связностью правил;
- пространством поиска;
- количеством активных агентов, изменяющих предметную область;
- классом решаемых задач.

По степени сложности выделяют *простые и сложные правила*. К сложным относят правила, текст записи которых на естественном языке занимает 1/3 страницы и больше. Правила, текст которых занимает менее 1/3 страницы, относят к простым.

По степени связности правил задачи делят на связные и малосвязные. К связным относятся задачи (подзадачи), которые не удастся разбить на независимые задачи. Малосвязные задачи удастся разбить на некоторое количество независимых подзадач.

Степень сложности задачи определяется не просто общим количеством правил данной задачи, а количеством правил в ее наиболее связанной независимой подзадаче.

Пространство поиска может быть определено по крайней мере тремя под-асpekтами: размером, глубиной и шириной. Размер пространства поиска дает обобщенную характеристику сложности задачи. Выделяют малые (до $10!$ состояний) и большие (свыше $10!$ состояний) пространства поиска. Глубина пространства поиска характеризуется средним числом последовательно применяемых правил, преобразующих исходные данные в конечный результат, ширина

пространства—средним числом правил, пригодных к выполнению в текущем состоянии.

Количество активных агентов существенно влияет на выбор метода решения. Выделяют следующие значения данного аспекта: ни одного агента, один агент, несколько агентов.

Класс решаемых задач характеризует методы, используемые ЭС для решения задачи. Данный аспект в существующих ЭС принимает следующие значения: задачи расширения, доопределения, преобразования. Задачи расширения и доопределения являются статическими, а задачи преобразования—динамическими.

К задачам расширения относятся задачи, в процессе решения которых осуществляется только увеличение информации о предметной области, не приводящее ни к изменению ранее введённых данных, ни к выбору другого состояния области. Типичной задачей этого класса являются задачи классификации.

К задачам доопределения относятся задачи с неполной или неточной информацией о реальной предметной области, цель решения которых—выбор из множества альтернативных текущих состояний предметной области того, которое адекватно исходным данным. В случае неточных данных альтернативные текущие состояния возникают как результат ненадежности данных и правил, что приводит к многообразию различных доступных выводов из одних и тех же исходных данных. В случае неполных данных альтернативные состояния являются результатом доопределения области, т. е. результатом предположений о возможных значениях недостающих данных.

К задачам преобразования относятся задачи, которые осуществляют изменения исходной или введённой ранее информации о предметной области, являющиеся следствием изменений либо реального мира, либо его модели.

Большинство существующих ЭС решают задачи расширения, в которых нет ни изменений предметной области, ни активных агентов, преобразующих предметную область. Подобное ограничение неприемлемо при работе в динамических областях.

По степени сложности структуры ЭС делят на поверхностные и глубинные. *Поверхностные ЭС* представляют знания об области экспертизы в виде правил (условие → действие). Условие каждого правила определяет образец некоторой ситуации, при соблюдении которой правило может быть выполнено. Поиск решения состоит в выполнении тех правил, образцы которых сопоставляются с текущими данными (текущей ситуацией в РП). При этом предполагается, что в процессе поиска решения последовательность формируемых таким образом ситуаций не оборвется до получения решения, т. е. не возникнет неизвестной ситуации, которая не сопоставится ни с одним правилом. *Глубинные ЭС*, кроме возможностей поверхностных систем, обладают способностью при возникновении неизвестной ситуации определять с помощью некоторых общих принципов, справедливых для области экспертизы, какие действия следует выполнить.

По типу используемых методов и знаний ЭС делят на традиционные и гибридные. *Традиционные ЭС* используют в основном неформализованные методы инженерии знаний и неформализованные знания, полученные от экспертов. *Гибридные ЭС* используют и методы инженерии знаний, и формализованные методы, а также данные традиционного программирования и математики.

Совокупность рассматриваемых выше характеристик позволяет определить особенности конкретной ЭС. Однако пользователи зачастую стремятся охарактеризовать ЭС каким-либо одним обобщенным параметром. В этой связи говорят о поколениях ЭС. В настоящее время выделяют ЭС первого и второго поколения. По нашему мнению, целесообразно говорить о трех поколениях ЭС. К *первому поколению* следует относить статические поверхностные ЭС, ко *второму* - статические глубинные ЭС (иногда ко второму поколению относят гибридные ЭС), а к *третьему*— динамические ЭС (вероятно, они, как правило, будут глубинными и гибридными).

В последнее время выделяют два больших класса ЭС (существенно отличающихся по технологии их проектирования), которые условно называют простыми и сложными ЭС. *Простая ЭС* может быть охарактеризована следующими значениями основных параметров:

- поверхностная ЭС;
- традиционная ЭС (реже гибридная);
- выполненная на персональной ЭВМ (ПЭВМ);
- коммерческая стоимостью от 100 до 25 тыс. дол.;
- стоимость разработки от 50 тыс. до 300 тыс. дол.;
- время разработки от 3 мес. до одного года (при использовании развитых инструментальных средств);
- от 200 до 1000 правил.

Сложная ЭС может быть охарактеризована следующими значениями параметров:

- глубинная ЭС;
- гибридная ЭС;
- выполненная либо на символьной ЭВМ, либо на мощной универсальной ЭВМ, либо на интеллектуальной рабочей станции;
- коммерческая стоимость от 50 тыс. до 1 млн. дол.;
- средняя стоимость разработки 5-10 млн. дол.;
- время разработки от 1 до 5 лет;
- от 1500 до 10 тыс. правил.

Стадия существования характеризует степень проработанности и отлаженности ЭС. Обычно выделяют следующие стадии:

- демонстрационный прототип;
- исследовательский прототип;
- действующий прототип;
- промышленная система;
- коммерческая система.

Демонстрационным прототипом называют ЭС, которая решает часть требуемых задач, демонстрируя жизнеспособность метода инженерии знаний. При наличии развитых ИС для разработки демонстрационного прототипа требуется в среднем примерно 1—2 мес., а при отсутствии—12—18 мес. Демонстрационный

прототип работает, имея в БЗ 50—100 правил. Развитие демонстрационного прототипа приводит к исследовательскому прототипу.

Исследовательским прототипом называют систему, которая решает все требуемые задачи, но неустойчива в работе и не полностью проверена. На доведение системы до стадии исследовательского прототипа уходит 3—6 мес. Исследовательский прототип обычно имеет в БЗ 200—500 правил, описывающих проблемную область.

Действующий прототип надежно решает все задачи, но для решения сложных задач может потребоваться чрезмерно много времени и (или) памяти. Для доведения системы до стадии действующего прототипа требуется 6—12 мес., при этом количество правил в БЗ увеличивается до 500—1000.

Экспертная система, достигшая *промышленной стадии*, обеспечивает высокое качество решений всех задач при минимуме времени и памяти. Обычно процесс преобразования действующего прототипа в промышленную систему состоит в расширении БЗ (до 1000—1500 правил) и переписывании программ с использованием более эффективных ИС, например в перепрограммировании на языках низкого уровня. Для доведения ЭС от начала разработки до стадии промышленной системы требуется 1—1,5 года.

Обобщение задач, решаемых ЭС на стадии промышленной системы, позволяет перейти к стадии *коммерческой системы* — к системе, пригодной не только для собственного использования, но и для продажи различным потребителям. Для доведения системы до коммерческой стадии требуется 1,5—3 года и 0,3-5 млн. дол. При этом в БЗ системы 1500—3000 правил.

Инструментальные средства определяют *программные* и *аппаратные* средства, используемые в рассматриваемой ЭС.

Лекция № 6

Тема: Особенности реализации ЕЯ-систем

В конце 60-х годов в исследованиях по ИИ сформировалось самостоятельное направление, получившее название "обработка естественного языка".

Задачей данного направления является исследование методов и разработка систем, обеспечивающих реализацию процесса общения с ЭВМ на естественном языке.

Сложность создания средств общения, предназначенных для конечных пользователей, обусловлена в значительной степени отсутствием единой теории языкового общения, охватывающей все аспекты взаимодействия коммуникантов.

Низкая эффективность, а часто неприемственность традиционных средств общения в большинстве случаев вызвано тем, что в них не учитываются важнейшие особенности процесса общения, направленного на удовлетворение реальных информационных потребностей пользователя (ИПП). Эти особенности независимо от специфики решаемых пользователями задач сводятся к следующим:

- 1) *Изменяемость*. ИПП не может быть заранее чётко определена в спецификациях на разработку системы общения. ИПП неизбежно изменяется в ходе разработки и эксплуатации системы.
- 2) *Несовпадение взглядов на мир*. Представления, имеющиеся у пользователя и системы о языке общения и проблемной области, относительно которых ведётся общение, могут не совпадать. Исходя из этого, процесс общения должен предусматривать разъяснения смысла неизвестных терминов, обнаружение и устранение несовпадающих представлений, а также предупреждение ошибочных толкований, т.е. установление общих точек зрения на обсуждаемые в процессе общения сущности.
- 3) *Связность общения*. Процесс общения не может быть ограничен обменом изолированными парами "вопрос-ответ", т.к. в большинстве реальных случаев ИПП не может быть выражена в виде одного вопроса (предложения). Часто требуется определить ситуацию, в которой возникла ИПП, т.е. предпослать запросу на решение некоторой задачи контекст, в котором эту задачу надо решать. Кроме того, процесс удовлетворения ИПП - решение некоторой задачи, в большинстве реальных предложений

требует взаимодействия, основанного на смешанной инициативе участников. Поэтому процесс общения должен иметь сложную разветвлённую структуру и состоять из обмена связанными высказываниями.

4) *"Неправильность" высказываний пользователя.* Для выражения ИПП пользователь может применять как "правильные" предложения, т.е. такие, которые будут однозначно поняты и верно обработаны системой, так и "неправильные". Неправильности могут быть вызваны, во-первых, тем, что пользователь обычно не в состоянии учесть все ограничения системы общения в частности её возможностей и знаний; во-вторых, использованием умолчаний, характерных для естественного общения и допускающих неоднозначное толкование высказываний, и, в-третьих, отклонением предложений от грамматической нормы.

Недостатки традиционных средств общения потребовали создание средств нового поколения, которые должны быть способны настраиваться на РГПП и адаптироваться к их изменению, представлять и объяснять свою точку зрения на проблемную область, а также учитывать точку зрения пользователя, поддерживать связный диалог и уметь обрабатывать "неправильные" высказывания. Разработка этих средств ведётся в настоящее время по двум основным направлениям:

1) Направление, развиваемое преимущественно специалистами по системам обработки данных, заключается в повышении уровня и увеличении непроцедурности формализованных языков общения.

2) Развивается в рамках ИИ и предполагает использование конечными пользователями для взаимодействия с ЭВМ естественного языка, семантически и прагматически ограниченного проблемной областью, относительно которой ведётся общение. ЕЯ-системы разработаны в рамках второго направления.

Чтобы быть полноправным участником общения ЕЯ-система должна выполнять некоторые обязательные функции. К этим функциям относятся:

1) введение диалога - определение его структуры и той роли, которую система и пользователь выполняют на текущем шаге диалога;

- 2) понимание - преобразование поступающих от пользователя высказываний на ЕЯ в высказывания на языке внутреннего представления;
- 3) обработка высказываний - формирование или определение знаний на решение задач на данном шаге диалога;
- 4) генерация - формирование выходных высказываний на ЕЯ.

В соответствии с высказанными функциями общения схема ЕЯ- системы может быть представлена:

- 1) Диалоговый компонент.

В связи с тем, что возможности существующих ЕЯ- систем не позволяют им самостоятельно формировать целесообразное поведение, в систему обычно вводится информация, определяющая общую и тематическую структуры диалога. По структуре и текущему состоянию диалога диалоговый компонент формирует (если инициатива принадлежит системе) или определяет (если инициатива принадлежит пользователю) задание, выполняемое системой на текущем шаге.

Ведение диалога выполняется по одной из двух схем: диалог ведёт пользователь, диалог ведёт система. Для системы весь диалог сводится к выработке реакций на текущие высказывания пользователя.

Вторая задача диалогового компонента вызвана тем, что реакции одного участника могут не соответствовать ожиданиям другого. Формирование перехвата происходит в тех случаях, когда система определяет, что текущая ситуация не соответствует ситуации, предусмотренной структурой диалога. Если же перехват инициативы осуществляет пользователь, то задача системы - обработать его, т.е. распознать наличие перехвата инициативы, определить новую цель (тему), на которую перешёл пользователь, и перейти на структуру диалога, соответствующую новой теме.

- 2) Компонент понимания высказываний предназначен для выделения смысла входного высказывания и выражения этого смысла на внутреннем языке системы. На этапе анализа выделяются описания сущностей, упомянутых во входном высказывании, выделяются свойства этих сущностей и отношения

между ними. Анализ выполняется отдельным блоком-анализатором. Анализаторы различаются по ряду параметров основные из них:

- тип анализируемых предложений,
- выделяемые описания сущностей;
- глубина проникновения в смысл;
- используемые для анализа средства

Интерпретация заключается в отображении входного высказывания на знания системы. Основными задачами интерпретации являются:

- буквальная интерпретация высказывания в контексте диалога; (состоит в том, чтобы, учитывая контекст диалога, идентифицировать образы тех сущностей области интерпретации, которые имел в виду говорящий)

- интерпретация высказывания на намерения говорящего. (состоит в том, чтобы, применяя имеющиеся у системы методы вывода, определить, как обрабатываемое высказывание соотносится с целями и планами участников общения)

3) Компонент генерации высказываний решает в соответствии с результатами, полученными остальными компонентами системы, две основные задачи:

- генерация смысла является сложной и мало изученной. Тип высказывания зависит от состояния системы и результатов, полученных предыдущими компонентами. Так, если система должна генерировать ответ на вопрос, то необходимо определить по состоянию системы, будет ли ответ прямой или косвенный.

- вторая задача состоит в синтезе естественно-языкового выражения, соответствующего внутреннему представлению выходного высказывания. Данная задача подразделяется на этапы семантического, синтаксического и морфологического синтеза. Сложность задачи синтеза определяется требованиями к естественности и выразительной мощности выходных высказываний.

Для понимания принципов построения ЕЯС важен вопрос об используемых в системе знаниях. Знания ЕЯС можно классифицировать собственно зна-

ния; способ представления знаний. К основным видам знаний относятся факты и операционные знания.

Способ представления знаний включает два аспекта: способ организации знаний и модель представления.

Лекция № 7 ТЕМА: Теоретические основы систем речевого общения.

Основные положения

В системах искусственного интеллекта с элементами естественно-языкового (ЕЯ) общения обычно предполагается, что в качестве средства общения используется письменная речь. Это не всегда удобно, а во многих случаях и неэффективно. Использование устной речи как средства общения позволяет почти на порядок повысить скорость ввода информации, разгрузить зрение и освободить руки, осуществить речевое общение на значительном расстоянии и по телефону. Если не затрагивать общих проблем ЕЯ-общения, которые связаны с пониманием речи, с созданием преобразователей «смысл—текст» и «текст—смысл», то узко специальными проблемами, стоящими перед разработчиками систем речевого общения (СЮ), становятся проблемы создания преобразователей «текст — речевой сигнал» и «речевой сигнал — текст». Первая из них называется проблемой *синтеза речи*, вторая — *анализа и распознавания речи*.

В системах ЕЯ-общения под текстом обычно понимают *орфографический*, или буквенный (как пишется), текст, в СРО—*фонемный* (как слышится). В создании преобразователей орфографического текста в фонемный и наоборот не существует особых проблем, хотя сложность таких преобразователей для разных языков будет различной (ср. русский и английский). Поэтому применительно к СЮ можно ограничиться проблемами разработки преобразователей «цепочка фонем—речевой сигнал» и «речевой сигнал—цепочка фонем». В речи фонема выступает в двух аспектах. С одной стороны, это элементарная смыслоразличительная единица письменной речи, с другой — абстрактное обозначение конкретного звука устной речи. Отметим сразу, что не существует счетного множества, а тем более одного-единственного звука речи, соотноси-

мого с фонемой. Если учесть, что на речевой звук налагается множество экстралингвистических факторов: индивидуальные особенности речи и голоса, эмоциональное и физиологическое состояния говорящего, электроакустические характеристики среды и тракта передачи, а также шумы, помехи и искажения, то сложность проблемы, стоящей перед создателями СРО, пожалуй, трудно переоценить.

Современные исследования в области СРО начаты в индустриально развитых странах в начале 60-х годов. Первые промышленные СРО появились в конце 70-х годов. К настоящему времени созданы разнообразные СРО для разных сфер применения. Это связано с осознанием потенциальными потребителями преимуществ СРО:

- удобство, простота и естественность процедуры общения, требующей минимума специальной подготовки;
- возможность использования для связи с ЭВМ обычных телефонных аппаратов и существующей телефонной сети;
- устранение ручных манипуляций с одновременным увеличением скорости ввода информации (в 3—5 раз по сравнению с клавиатурным вводом) и разгрузка зрения при получении информации.

Первое и второе преимущества с наибольшим эффектом проявляется в автоматизированных системах управления (АСУ) предприятия, организации или отрасли. Однако пока АСУ выполняет в основном информационные функции. С внедрением в АСУ речевой технологии общения ЭВМ станет по-настоящему активным звеном управления. Круг пользователей системы в силу простоты и естественности общения с ЭВМ неизмеримо расширится. При этом человеко-машинное взаимодействие будет осуществляться с помощью местной или городской телефонной сети на значительном расстоянии. В любое время большая группа пользователей сможет одновременно обращаться к ЭВМ за получением разного рода нормативно-справочной информации. Имея необходимую базу знаний — модель каждого конкретного производства, система сможет по запросу пользователя выдать необходимую консультацию по телефо-

ну в случае возникновения трудностей в той или иной сфере производства. Кроме того, система сможет выполнять функции диспетчера, обращаясь по телефону в необходимое время к конкретным работникам за получением данных о ходе производства, предупреждая их о наступлении критических ситуаций, напоминая о сроках поставки изделий или выполнения разного рода обязательств, а также осуществлять учет и контроль. Несомненно, во всех этих случаях эффективная работа СРО возможна лишь при условии, что она является составной частью систем искусственного интеллекта.

Третье свойство — разгрузка зрения и рук — важно для создания нового поколения систем оперативного человеко-машинного управления сложными объектами. К ним относятся прежде всего системы управления движением, энергетическими установками и другие АСУ ТП, САПР и АРМ.

Основой для разработки современных СРО является лингвоакустическая и информационная теории речеобразования и восприятия речи. Лингвистическая теория рассматривает фонетические и просодические характеристики речи, акустическая — акустические характеристики (признаки) фонем и просодем, информационная — структуру речевого сигнала. Эффективность СРО тем выше, чем полнее реализованы в ней принципы функционирования естественной речевой системы человека. При этом не обязательно, чтобы искусственные СРО копировали структурные особенности работы естественной системы. Важно, чтобы СРО как можно ближе была в функциональном отношении к естественной.

Фонетическая и просодическая структуры речи

Минимальной смысловой единицей речи является *фонема*. В русском языке 42 фонемы, из них 6 гласных и 36 согласных. В других языках число фонем и соотношение гласных и согласных различно. Например, в английском языке 20 гласных (из них 5 дифтонгов) и 24 согласных, во французском — 16 гласных и 20 согласных.

Акустические характеристики каждой фонемы обусловлены артикуляционными особенностями ее образования: местом и способом. Под *местом образования* понимается положение сужений артикуляционного тракта в ядре фонем,

определяющее его конфигурацию (артикуляционную статику) и в конечном итоге — его резонансные свойства.

Способ образования характеризует энергетические и динамические особенности артикуляции фонемы. По способу образования фонемы делятся на взрывные, аффрикаты, щелевые, дрожащие, носовые, боковые, плавные и гласные. Эти группы фонем различаются в основном, динамикой движения артикуляторов. Кроме того, взрывные, аффрикаты и щелевые дополнительно различаются по признаку наличия-отсутствия колебания голосовых связок (звонкие-глухие).

В потоке речи в зависимости от конкретных условий акустико-артикуляционные характеристики фонем изменяются, что приводит к появлению оттенков фонем — *аллофонов*. Аллофоны подразделяются на комбинаторные и позиционные. *Комбинаторные аллофоны* обусловлены соседством данной фонемы с другими фонемами и образуются вследствие наложения в потоке речи артикуляции одного звука на другой—так называемого эффекта коартикуляции. *Позиционные аллофоны* обусловлены положением фонемы в слове или фразе по отношению к ударному слогу, концу и началу слова и т. д.

Учет этих и других факторов позволяет дать оценку общему числу аллофонов гласных $N_g=480$ и согласных $N_c=8880$ которая, указывая порядок этого числа, наглядно иллюстрирует степень вариантности фонем в речевом потоке.

Другой класс лингвистических понятий, учет которых исключительно важен при создании высококачественных СРО, это интонация и ударение, определяющие *просодию речи*. Интонация и ударение играют важную роль при восприятии речи человеком. По ним определяются коммуникативная направленность высказывания, логический смысл, выделение главного и общего (рема и тема), осуществляется вычленение семантически связанных отрезков речи и объединение речевых элементов внутри этих отрезков. Зачастую в зависимости от интонационного оформления логически идентичные высказывания могут иметь различное семантическое значение. Физически интонация и

ударение реализуются совокупностью акустических средств - просодических характеристик речи, к которым относятся:

Мелодика - изменение частоты основного тона голоса;

Ритмика - текущее изменение длительности звуков и пауз;

Энергетика - текущее изменение интенсивности звука.

Процесс чтения текста человеком предполагает наличие автоматической процедуры формирования основного тона, интенсивности звука, длительности звуков и пауз на основе анализа входного текста. Это значит, что преобразование орфографического текста в последовательность фонем должно сопровождаться выделением информации, необходимой для задания просодических характеристик речевых сигналов. Для этой цели текст анализируется и по определенным правилам разбивается на следующие основные элементы: фраза, синтагма, акцентная группа, фонетическое слово. Эти элементы маркируются соответственно фразовым, синтагматическим, групповым и словесным ударениями. Каждой синтагме присваивается один из возможных интонационных типов. Основными интонационными типами являются восклицательно-повествовательный, вопросительный. Соответствует также большое число модальных и эмоциональных оттенков этих основных интонационных типов.

Минимальным элементом, из совокупности которых складывается интонация синтагмы, фразы и текста в целом, является акцентная группа. Под *акцентной группой* понимается одно или несколько полнозначных слов синтагмы, объединенных общим просодическим (мелодическим, ритмическим и энергетическим) контуром, привязанным к единому групповому ударению. Число акцентных групп не всегда совпадает с числом фонетических слов в синтагме. Выделение слова в акцентной группе делает словесное ударение главным или основным, другие слова в акцентной группе получают слабое или не основное ударение. Основное ударение чаще всего получает первое слово акцентной группы. Это позволяет сформулировать простое правило установления границ акцентной группы: левая граница совпадает с началом

слова, имеющего групповое ударение, правая - с началом слова следующей акцентной группы.

Акцентные группы ранжируются относительно границ синтагмы на конечные, начальные и срединные, которые вносят различный по значимости вклад в формирование просодического контура синтагмы. Основное разнообразие мелодических контуров реализуется на конечной акцентной группе, существенно меньшее - на начальной и срединной.

Мелодика, ритмика и энергетика акцентной группы задаются нормированными значениями частоты, длительности и интенсивности на трех ее участках: ядре, предъядре и заядре. *Ядром* акцентной группы является ударный слог, отмеченный знаком группового ударения, *предъядром* и *заядром* — соответственно предшествующие ему и следующие за ним фонемы акцентной группы.

Как показали экспериментальные исследования просодических контуров естественной речи и опыты по восприятию интонации и ударения синтетических стимулов, интонационный тип фразы определяется положением максимума мелодического контура относительно главноударного гласного. Положению максимума в центре гласного соответствует интонация побуждения или восклицания, сдвигу его влево—интонация завершения или утверждения, а сдвигу вправо— интонация незавершенности или вопроса. Другие параметры мелодического контура—значение максимума и его ширина—более важны при передаче эмоциональных оттенков. Положение ударения в слове, синтагме или фразе определяется в большей степени длительностью ударного гласного и в меньшей его интенсивностью. Положение ударения в слове узнается в 100 % случаев, если длительность ударного гласного превышает длительность безударного более чем в 1,5 раза.

Акустические характеристики фонем

Акустические характеристики фонем тесно связаны с артикуляционными особенностями их образования — местом и способом.

Информационная структура речевого сигнала

Для передачи любых сообщений требуется вполне определенная структура системы связи. Всякая техническая система связи включает такие функциональные элементы, как источники информации, кодирующие устройства, переносчики, модуляторы, системы разделения каналов, демодуляторы, декодеры и др. Естественные системы связи, в частности устная речь, не являются в этом смысле исключением. Основной особенностью речевого сигнала, вытекающей из такого рассмотрения является его полиинформативность и полимодуляционность.

Генерирование речевого сигнала связано с функционированием трех источников звука - переносчиков: тонального, шумового и импульсного.

При передаче информации в фонемном составе речи осуществляется непрерывный процесс перестройки речевого тракта. Это приводит к непрерывному изменению формы мгновенных спектров речевого сигнала—модуляция формы спектров—и его средней мощности—амплитудная модуляция. При перестройке речевого тракта осуществляется также фазовая модуляция, хотя полной ясности ее роли в восприятии речевых звуков нет. Кроме того, процесс передачи информации о фонемном составе связан с постоянной сменой комбинаций включения переносчиков—манипуляция переносчиков—и с изменением частоты основного тона на смычках звонких взрывных звуков—частотная модуляция. Информация об интонационной окраске речи, а также о физическом и эмоциональном состоянии, об индивидуальных особенностях голоса и характеристике электроакустической среды передается в основном путем модуляций частоты основного тона — частотная модуляция—и общего уровня звуков—амплитудная модуляция. В целом можно сделать несколько важных выводов:

- в процессе речевого общения осуществляется параллельная передача различных видов информации;
- для передачи информации каждого вида одновременно используются несколько видов модуляции;
- ни один из видов модуляции (кроме манипуляции переносчиков) не используется только для передачи одного вида информации.

Учитывая указанные особенности формирования речевого сигнала, можно сформулировать вполне определенные требования к схеме его обработки в приемном устройстве (анализаторе).

1. При анализе заданного элемента информационной структуры должна осуществляться демодуляция (детектирование) речевого сигнала по каждому виду модуляции, посредством которой ведется его передача.

2. Результат детектирования по каждому виду модуляции должен быть инвариантен относительно остальных видов модуляции. Возможным методом достижения такого рода инвариантности является осуществление предварительной нормализации речевого сигнала.

3. Если с помощью данного вида модуляции одновременно осуществляется передача еще из других элементов информационной структуры, то полученный в результате демодуляции сигнал должен быть подвергнут дальнейшим операциям разделения.

Принципы построения систем речевого общения

Классификация речевых процессоров

В общем случае СРО строятся на базе специализированных речевых процессоров двух основных типов: анализаторов и синтезаторов.

Анализаторы. Эти устройства предназначены для преобразования речевых сигналов с микрофона (информационный поток сигналов 10^5 бит/с) в последовательность цифровых кодов с существенно меньшим информационным потоком (10^4 — 10^1 бит/с) и с обязательным сохранением передачи смыслового компонента речи. Анализаторы подразделяются на два основных класса: анализаторы сигналов и анализаторы сообщений.

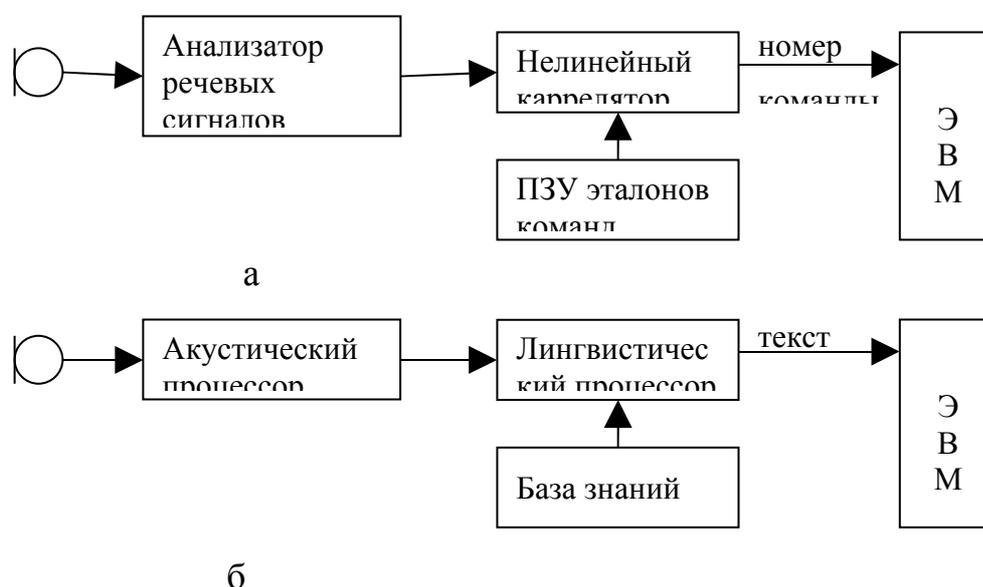
В анализаторах сигналов сокращение информационного потока достигается только за счет учета акустических и статистических характеристик речевого сигнала без обращения к его смысловой функции.

В анализаторах речевых сообщений (распознавателях) осуществляется сжатие информационного потока за счет введения операции распознавания

смысловых элементов речи (фразы, слова, морфемы, фонемы). Анализаторы речевых сообщений, в свою очередь, подразделяются на две группы: анализаторы ограниченного словаря и универсальные.

Анализаторы ограниченного словаря ориентированы на распознавание заданного конкретной задачей числа речевых команд (обычно порядка 10^2), т. е. на идентификацию одной из произнесенных речевых команд словаря в виде номера команды (рис. 1,а). Распознавание осуществляется путем нелинейного во времени сопоставления эталонов команд с произносимой командой и выбора наиболее схожего эталона. В большинстве существующих анализаторов ограниченного словаря формирование эталонов осуществляется в процессе обучения на используемый словарь команд и голос диктора. Чаще всего процесс обучения состоит в однократном прочтении оператором всего словаря команд.

Рис. 1. Схема анализаторов речевых сообщений ограниченного словаря (а) и универсального (б)



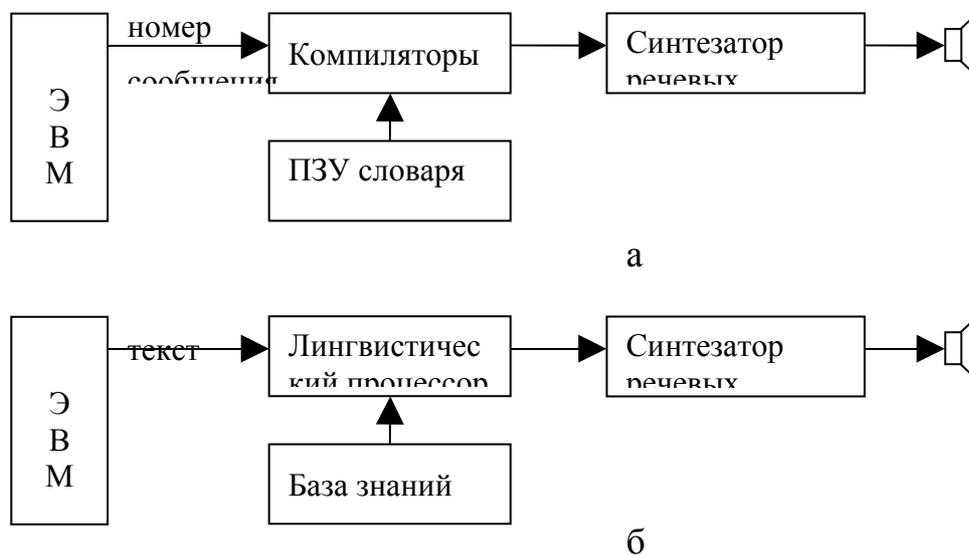
Еще одним ограничением большинства современных анализаторов этого типа является требование изолированного произнесения речевых команд, т. е. с паузами между словами от 0,3 до 1 с. Распознавание слитной речи даже ограниченного словаря— пока нерешенная научная проблема.

Универсальные анализаторы ориентированы на текущее распознавание полного набора смысловых элементов речи (фонем или морфем), из которых может быть составлено и в конечном счете распознано любое слово или слитно произнесенное речевое сообщение (рис. 1,б). Распознавание осуществляется лингвистическим процессором по правилам, заложенным в базе знаний.

Синтезаторы речи. Эти устройства предназначены для преобразования кодовой последовательности, поступающей от ЭВМ, ПЗУ или линии связи, в непрерывный речевой сигнал. Синтезаторы подразделяются на классы и группы по тем же признакам, что и анализаторы речи. Классам анализаторов речевых сигналов и сообщений соответствуют синтезаторы речевых сигналов и сообщений

Синтезаторы сообщений делятся по аналогичным признакам на синтезаторы ограниченного словаря — компиляторы и универсальные.

Рис. 2. Схемы синтезаторов речевых сообщений ограниченного словаря (а) и универсального (б)



В *компиляторах* (рис. 2,а) любое сложное речевое сообщение может быть получено путем компиляции (простого соединения) элементов речи. Элементы речи начитываются диктором, соответствующие им сигналы компрессируются тем или иным способом, кодируются и записываются в ПЗУ. При синтезе речевого сообщения из ПЗУ закодированные речевые элементы считыва-

ются в нужной последовательности и одновременно восстанавливается речевой сигнал. Очевидной простотой компиляционного метода и его технической реализации объясняется большое количество сообщений о проектируемых и законченных разработках компиляторов к их применению.

Удовлетворительный по качеству компиляционный синтез речи возможен лишь при использовании в качестве элементов речи отдельных фраз либо словоформ, подставляемых в определенное место стандартной фразы. Попытки добиться высококачественного синтеза произвольного текста простой компиляцией словоформ, слогов или аллофонов не привели к положительным результатам. Все эти элементы речи тесно связаны внутри фразы. В слитной речи не существует аналогов этих элементов, произнесенных изолированно, и наоборот, речь из изолированно произнесенных элементов звучит ненатурально. В связи с этим подготовка словаря в компиляторах представляет собой самостоятельную и сложную проблему. Для успешного ее решения в каждом конкретном случае применения синтезатора многие фирмы за рубежом идут на создание специализированных центров проектирования речи, оснащенных соответствующим оборудованием и персоналом лингвистов-прикладников.

При разработке *универсальных синтезаторов* речевых сообщений стремятся получить функциональную модель речеобразования, адекватную реально существующим языковым и акустическим явлениям. На входе такой модели орфографический или фонемный текст произвольного содержания, на выходе — звучащая речь (рис. 2,6). По своему существу разрабатываемые в рамках данного подхода синтезаторы являются кибернетической функциональной моделью чтения текста человеком. В базе знаний синтезатора хранится не только информация об элементарных единицах речи (эталон фонем и интоном), но и алгоритмические правила их модификации в зависимости от конкретного контекста звуковой реализации. Процесс применения этих правил к эталонам фонем и интоном для входного синтезируемого текста реализуется лингвистическим процессором. К настоящему времени качество речи и стоимость универсальных синтезаторов достигли коммерчески приемлемых показателей, и они начинают оказывать серьезную конкуренцию компиляторам в силу

простоты применения, малого расхода памяти на элемент речи, неограниченности состава словаря синтезируемых сообщений.

Лекция № 8 Тема: Теоретические аспекты машинного перевода

Назначение машинного перевода

Машинный перевод (МП), или автоматический перевод (АП),—интенсивно развивающаяся область научных исследований, экспериментальных разработок и уже функционирующих систем (СМП), в которых к процессу перевода с одного естественного языка (ЕЯ) на другой привлекается ЭВМ. СМП открывают быстрый и систематический доступ к информации на иностранном языке, обеспечивают оперативность и единообразие в переводе больших потоков текстов, в основном научно-технических. Работающие в промышленном масштабе СМП опираются на большие терминологические банки данных и, как правило требуют привлечения человека в качестве пред-, интер- или постредактора. Современные СМП, в особенности те, которые опираются при переводе на базы знаний в определенной предметной области, относят к классу систем искусственного интеллекта (ИИ).

Основные сферы использования МЦ

1. В отраслевых службах информации при наличии большого массива или постоянного потока иноязычных источников. Если СМП используются для выдачи сигнальной информации, постредактирование не требуется.

В крупных международных организациях, имеющих дело с многоязычным политематическим массивом документов. Таковы условия работы в Комиссии Европейских сообществ в Брюсселе, где вся документация должна появляться одновременно на девяти рабочих языках. Поскольку требования к переводу здесь высоки, МП нуждается в постредактировании.

В службах, осуществляющих перевод технической документации, сопровождающей экспортируемую продукцию. Переводчики не справляются в требуемые сроки с обширной документацией (так, спецификации к самолетам и другим

сложным объектам могут занимать до 10000 и более страниц). Структура и язык технической документации достаточно стандартны, что облегчает МП и даже делает его предпочтительным перед ручным переводом, так как гарантирует

всего массива. Поскольку перевод спецификаций должен быть полным и точным, продукция МП нуждается в постредактировании.

4. Для синхронного или почти синхронного перевода некоторого постоянного потока однотипных сообщений. Таков поток метеосводок в Канаде, который должен появляться одновременно на английском и французском языках.

Помимо практической потребности делового мира в СМП, существуют и чисто научные стимулы к развитию МП: стабильно работающие экспериментальные системы МП являются опытным полем для проверки различных аспектов общей теории понимания, речевого общения, преобразования информации, а также для создания новых, более эффективных моделей самого МП.

С точки зрения масштаба и степени разработанности СМП можно разбить на три основных класса: промышленные, развивающиеся и экспериментальные.

Лингвистическое обеспечение систем машинного перевода

Процесс МП представляет собой последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке, который должен максимально воссоздавать смысл и, как правило, структуру исходного текста, но уже средствами выходного языка. К лингвистическому обеспечению СМП относится весь комплекс собственно лингвистических, металингвистических и так называемых «экстралингвистических» знаний, которые используются при таком преобразовании.

В классических СМП, осуществляющих не прямой перевод по отдельным предложениям (пофразный перевод), каждое предложение проходит последовательность преобразований, состоящую из трех частей (этапов): анализ —> трансфер (межъязыковые операции)—>синтез. В свою очередь, каждый из этих этапов представляет собой достаточно сложную систему промежуточных преобразований.

Цель этапа анализа построить структурное описание (промежуточное представление, внутреннее представление) входного предложения, | Задача этапа трансфера (собственно перевода)—преобразовать структуру входного предложения во внутреннюю структуру выходного предложения. К этому этапу отно-

сятся и замены лексем входного языка их переводными эквивалентами (лексические межъязыковые преобразования). Цель этапа синтеза—на основе полученной в результате анализа структуры построить правильное предложение выходного языка.

Лингвистическое обеспечение стандартной современной СМП включает:
словари;

грамматики;

формализованные промежуточные представления единиц анализа на разных этапах преобразований.

Помимо стандартных, в отдельных СМП могут иметься и некоторые нестандартные компоненты. Так, экспертные знания о ПО могут задаваться с помощью специальных концептуальных сетей, а не в виде словарей и грамматик.

Механизмы (алгоритмы, процедуры) оперирования с имеющимися словарями, грамматиками и структурными представлениями относят к математико-алгоритмическому обеспечению СМП.

Одно из необходимых требований к современным СМП—высокая модульность. С лингвистически содержательной точки зрения это означает, что анализ и следующие за ним процессы строятся с учетом теории лингвистических уровней. В практике создания СМП различают такие уровни анализа:

- досинтаксический анализ (в него входит морфологический анализ - МорфАн, анализ оборотов, неопознанных элементов текста и др.);

- синтаксический анализ СинАн (строит синтаксическое представление предложения, или СинП); в его пределах может выделяться ряд подуровней, обеспечивающих анализ разных типов синтаксических единиц;

семантический анализ СемАн, или логико-семантический анализ (строит аргументно-предикатную структуру высказываний или другой вид семантического

представления предложения и текст);

концептуальный анализ (анализ в терминах концептуальных структур, отражающих семантику ПО). Этот уровень анализа используется в СМП, ориентированных на очень ограниченные ПО. По сути дела, концептуальная структура является проекцией схем ПО на лингвистические структуры, часто даже не

на семантические, а на синтаксические. Только для очень узких ПО и ограниченных классов текстов концептуальная структура совпадает с семантической; в общем случае полного совпадения не должно быть, так как текст подробнее любых концептуальных схем.

Синтез теоретически проходит те же уровни, что и анализ, но в обратном направлении. В работающих системах обычно реализован только путь от СинП до цепочки слов выходного предложения.

Лингвистическое разграничение разных уровней может проявляться также в разграничении используемых в соответствующих описаниях формальных средств (набор этих средств задается для каждого уровня отдельно). На практике часто задаются отдельно лингвистические средства МорфАн и совмещаются средства СинАн и СемАн. Но разграничение уровней может оставаться только содержательным при использовании в их описаниях единого формализма, пригодного для представления информации всех выделяемых уровней.

С технической точки зрения модульность лингвистического обеспечения означает отделение структурного представления фраз и текстов (как текущих, временных знаний о тексте) от «постоянных» знаний о языке, а также языковых знаний • от знаний ПО; отделение словарей от грамматик, грамматик - от алгоритмов их обработки, алгоритмов « от программ. Конкретные соотношения различных модулей системы (словари—грамматики, грамматик — алгоритмы, алгоритмы — программы, декларативные — процедурные знания и др.), включая распределение лингвистических данных по уровням,— это то основное, что определяет специфику СМП.

Словари. Словари анализа, как правило, одноязычные. Они должны содержать всю информацию, необходимую для включения данной лексической единицы (ЛЕ) в структурное представление. Часто разделяют словари основ (с морфолого-синтаксической информацией: часть речи, тип словоизменения, подкласс, характеризующий синтаксическое поведение ЛЕ и т. п.) и словари словозначений, содержащие семантическую и концептуальную информацию: семантический класс ЛЕ, семантические надежи (валентности), условия их реализации во фразе и т. д.

Во многих системах разделены словари общеупотребительной и терминологической лексики. Такое разделение дает возможность при переходе к текстам другой предметной области ограничиваться лишь сменой терминологических словарей. Словари сложных ЛЕ (оборотов, конструкций) образуют обычно отдельный массив, словарная информация в них указывает на способ «собира-ния» такой единицы при анализе. Часть словарной информации может задаваться в процедурной форме, например, многозначным словам могут сопоставляться алгоритмы разрешения соответствующего типа неоднозначности. Новые виды организации словарной информации для целей МП предлагают так называемые «лексические базы знаний». Наличие разнородной информации о слове (называемой лексическим универсумом слова) приближает такой словарь, скорее к энциклопедии, чем к традиционным лингвистическим словарям.

Грамматики и алгоритмы. Грамматика и словарь задают лингвистическую модель, образуя основную часть лингвистических данных. Алгоритмы их обработки, т. е. соотнесения с текстовыми единицами, относят к математико-алгоритмическому обеспечению системы.

Разделение грамматик и алгоритмов важно в практическом смысле тем, что позволяет менять правила грамматики, не меняя алгоритмов (и соответственно программ), работающих с грамматиками. Но далеко не всегда такое разделение возможно. Так, для системы с процедурным заданием грамматики и тем более с процедурным представлением словарной информации такое разделение нерелевантно. Алгоритмы принятия решений в случае недостаточной (неполнота входных данных) или избыточной (вариантность анализа) информации в большой мере эмпиричны, их формулировка требует лингвистической интуиции. Задание общего управляющего алгоритма, ведающего порядком вызова разных грамматик (если их несколько в одной системе), также требует лингвистического обоснования. Тем не менее существующая тенденция — отделить грамматики от алгоритмов так, чтобы все лингвистически содержательные сведения задавались в статической форме грамматик, а алгоритмы сделать настолько абстрактными, что они смогут вызывать и обрабатывать разные лингвистические модели.

Наиболее четко разделение грамматик и алгоритмов наблюдается в системах, работающих с контекстно-свободными грамматиками (КСГ), где модель

языка— грамматика с конечным числом состояний, а алгоритм должен обеспечить для произвольно взятого предложения дерево его вывода по правилам грамматики, и если таких выводов несколько, то перечислить их. Такой алгоритм, представляющий собой формальную (в математическом смысле) систему, называется анализатором. Описание грамматики служит для анализатора, обладающее универсальностью, таким же входом, как и анализируемое предложение. Анализаторы строятся для классов грамматик, хотя учет специфических особенностей грамматики может повысить эффективность анализатора.

Грамматики синтаксического уровня — наиболее разработанная часть и с точки зрения лингвистики, и с точки зрения их обеспечения формализмами.

Основные типы грамматик и реализующих их алгоритмов:

цепочечная грамматика фиксирует порядок следования элементов, т. е, линейные структуры предложения, задавая их в терминах грамматических классов слов (артикль+существительное+предлог} или в терминах функциональных элементов (подлежащее+сказуемое);

грамматика составляющих (или грамматика непосредственно составляющих— НСГ) фиксирует лингвистическую информацию о группировке грамматических элементов, например, именная группа (состоит из существительного, артикля,

прилагательного и других модификаторов), предложная группа (состоит из предлога и именной группы) и т. д. до уровня предложения. Грамматика строится как набор правил подстановки, или исчисление продукций вида $A \rightarrow B \dots C$. НСГ

представляют собой грамматики порождающего типа и могут использоваться как при анализе, так и при синтезе: предложения языка порождаются многократным применением таких правил;

- грамматика зависимостей (ГЗ) задает иерархию отношений элементов предложения (главное слово определяет форму зависимых). Анализатор в ГЗ основан на идентификации хозяев и их зависимых (слуг). Главным в предложении является глагол в личной форме, так как он определяет число и характер зависимых существительных. Стратегия анализа в ГЗ— сверху вниз: сначала

идентифицируются хозяева, затем слуги, или снизу вверх: хозяева определяются процессом подстановки;

- категориальная грамматика Бар-Хиллела—это версия грамматики составляющих, в ней только две категории — предложения S и имени p . Остальные определяются в терминах способности комбинироваться с этими главными в структуре HC . Так, переходный глагол определен как $n \setminus S$, так как он сочетается с именем и слева от него, образуя предложение S .

Существует много способов учета контекстных условий: грамматики метаморфозы и их варианты. Все они являются расширениями KC -правил. В общем виде это значит, что правила продукций, переписываются так: $A [a] \rightarrow B[B], \dots, C [c]$, где малыми буквами обозначены условия, тесты, инструкции и т. д., расширяющие исходные жесткие правила и дающие грамматике гибкость и эффективность.

В грамматике обобщенных составляющих— TCC введены метаправила, являющиеся обобщением закономерностей правил $KC1$.

В грамматиках расширенных сетей переходов— RCP предусмотрены тесты и условия к дугам, а также инструкции, которые надо выполнить в случае, если анализ пошел по данной дуге. В разных модификациях RCP дугам может присываться вес, тогда анализатор может выбирать путь с наибольшим весом. Условия могут разбиваться на две части: контекстно-свободные и контекстно-зависимые.

Разновидностью $RCPG$ являются каскадные $RCPG$. Каскад—это RCP , снабженная действием $1\text{шшш}1$. Это действие вызывает остановку процесса в данном каскаде, запоминание информации о текущей конфигурации в стеке и переход к более глубокому каскаду с последующим возвратом в исходное состояние. RCP обладает рядом возможностей трансформационных грамматик. Она может использоваться и как генерирующая система.

Метод анализа с помощью граф-схемы позволяет сохранить частичные результаты и представить варианты анализа.

Новым и сразу завоевавшим популярность методом грамматического описания является лексико-функциональная грамматика ($ЛФГ$). Она устраняет необходимость трансформационных правил. Хотя $ЛФГ$ основывается на KCG , прове-

рочные условия в ней отделены от правил подстановки и «решаются» как автономные уравнения.

Унификационные грамматики (УГ) представляют собой следующий после граф-схем этап обобщения модели анализа: они способны воплощать грамматики различных видов. УГ содержит четыре компонента: пакет унификации, интерпретатор для правил и лексических описаний, программы обработки направленных графов, анализатор с помощью граф-схемы. УГ объединяют грамматические правила со словарными описаниями, синтаксические валентности с семантическими.

Центральной проблемой любой системы анализа ЕЯ является проблема выбора вариантов. Для ее решения грамматики синтаксического уровня дополняются вспомогательными грамматиками и методами разбора сложных ситуаций. В НС-грамматиках применяют фильтровый и эвристический методы. Фильтровый метод состоит в том, что сначала получают все варианты анализа предложения, а затем отбраковывают те, которые не удовлетворяют некоторой системе условий-фильтров. Эвристический метод с самого начала строит лишь часть вариантов, более правдоподобных с точки зрения заданных критериев. Использование весов для отбора вариантов является примером применения эвристических методов в анализе.

Семантический уровень гораздо меньше обеспечен теорией и практически разработками. Традиционной задачей семантики считается снятие неоднозначности синтаксического анализа - структурной и лексической. Для этого используется аппарат селективных ограничений, который привязан к рамкам предложений, т. е. вписывается в синтаксическую модель.

Наиболее распространенный тип СемАн основан на так называемых падежных грамматиках. В основе грамматики—понятие глубинного, или семантического, падежа. Падежная рамка глагола является расширением понятия валентность: это набор смысловых отношений, которые могут (обязательно или факультативно) сопровождать глагол и его вариации в тексте. В пределах одного языка один и тот же глубинный падеж реализуется разными поверхностными предложно — падежными формами. Глубинные падежи в принципе позволяют

выходить за рамки предложения, а выход в текст означает переход к семантическому уровню анализа.

Поскольку семантическая информация в отличие от синтаксической, опирающейся в первую очередь на грамматики, сосредоточена в основном в словарях, в 80-е годы интенсивно разрабатываются грамматики, позволяющие «лексикализовать» КСГ. Ведется разработка грамматик, основанных на исследовании свойств дискурса.

Лингвистическое обеспечение этапов трансфера и синтеза. Синтез в стандартных СМП трехчастного типа устроен проще, чем анализ, если он обеспечивается соответствующим входом, т. е. если после работы трансфера построено одно правильное дерево фразы. Используется та же лингвистическая модель, что и в анализе, в виде порождающей грамматики, программа синтеза работает как детерминированный преобразователь, необходимые для построения выходной словоформы грамматические категории определяются свойствами подчиняющего узла в дереве предложения. Одноязычные словари синтеза содержат всю информацию, которая позволяет синтезировать требуемую словоформу; в случае, если таковая не может быть построена, предлагаются замены (синтаксические или лексические дериваты).

Одна из задач синтеза - линейаризация узлов дерева, определяющая нужный порядок слов в синтезируемом предложении. Стандартным решением вопроса о порядке слов является передача на этап синтеза информации о взаимном расположении слов (или групп, конструкций) в исходном предложении с учетом тех регулярных изменений, которые производятся на этапе трансфера (например, изменения порядка следования компонентов группы «существительное+прилагательное» при французско-русском переводе). Более тонкое решение этой проблемы требует учета таких категорий, как темо-ремное членение, модально-временная рамка, семантика артиклевых употреблений и других явлений семантического уровня.

Правильность дерева, построенного на этапе трансфера, или выбор одного из нескольких вариантов, если они остались, проверяется теми же методами, что и на этапе анализа.

Таким образом, решение всех содержательных вопросов, которые могут повлиять на построение хорошего выходного предложения, остаются в компетенции этапов анализа и трансфера (межъязыковых операций).

Компонент трансфера задается двумя частями:

словарь трансфера, или переводной (межъязыковой) словарь, это двуязычный словарь лексических замен. Как правило, он однонаправлен: ЛЕ (слово-значение) входного языка->ЛЕ выходного языка, поскольку для каждого слова (с нерасчлененными значениями) существует несколько переводов в зависимости от контекста; возможны, впрочем, и обратимые переводные словари

словарь структурных соответствий, или грамматика трансфера.

Работа алгоритма преобразования состоит в сравнении структур, вычислении нужных значений и замене структуры входного предложения структурой выходного. Эти две части связаны, так что собственно перевод состоит иногда в замене целых поддеревьев (слова с его структурным контекстом) их лексическими и структурными эквивалентами.

Для некоторых языковых пар структурное и лексическое соответствия достаточно прямые (переводится слово а слово, конструкция—в конструкцию), но и при этом степень неоднозначности перевода отдельных слов высока. В случае очень разных языков (японский и любой из европейских) грамматика трансфера сложнее. Один из способов задания грамматики трансфера—контрастивная грамматика, представляющая собой собрание грамматических явлений, поверхностное выражение которых в данной паре языков различно,

Один из трудных вопросов этапа трансфера—определение порядка применения правил. Чтобы разгрузить этот громоздкий этап, можно либо строить более глубокие структуры при анализе, позволяющие раньше выявлять и снимать смысловую неоднозначность, либо часть задач передать синтезу, который должен будет выбирать одно из нескольких деревьев.

Лекция № 9 Тема: Особенности построения и использования систем переработки визуальной информации

Назначение, классификация и области применения

В различных областях науки и техники ощущается рост потребностей в переработке, анализе и отображении визуальной, информации. В этом направле-

нии принято выделять три основных типа задач: собственно *обработку изображений*, когда и исходные данные, и результаты обработки представляются в изобразительной форме; *анализ* (интерпретацию, распознавание или "понимание") изображений, когда входные данные являются изображением, а результат представляется в неизобразительной форме, например в виде текстового описания наблюдаемой сцены; *синтез изображений* (машинную графику), когда на входе имеется описание (алгоритм построения) изображения, а на выходе по нему строится само изображение. Взаимосвязь трех перечисленных типов задач показана на рис. 1.

Обработка изображений связана с преобразованием изобразительной информации вновь в изобразительную форму. Примером здесь может служить устранение искажений или дефектов на изображении, улучшение качества получаемой визуальной информации путем повышения контраста, подчеркивание контуров объектов и т. д.

Задачей анализа изображений является получение из изображения, поданного на вход системы, неизобразительного описания. Это описание может быть различных уровней общности— от простого указания номера или имени класса, к которому относится анализируемое изображение, до подробной характеристики наблюдаемой сцены с указанием отдельных объектов и отношений между ними. В последнем случае обычно говорят о «понимании» изображений. Иногда называют анализ изображений распознаванием, хотя, этот термин имеет более узкое значение, относясь преимущественно к идентификации отдельных объектов на изображениях. К типичным задачам анализа можно отнести распознавание рукописных или печатных знаков, дешифрацию аэрофотоснимков, анализ наблюдаемых сцен.

Синтез изображений обычно отождествляют с машинной графикой. В настоящее время синтезируются на ЭВМ не только «графические» картины, наоборот, синтез все больше претендует на создание полноцветных реалистических изображений по их описаниям в неизобразительной форме. Сюда относятся и системы имитации визуальной обстановки на тренажерах, и системы гео-

метрического моделирования в САПР, и системы компьютерного киноискусства. Наряду с этим сохраняются и традиционные приложения машинной графики: вывод информации на экран дисплея в виде диаграмм, графиков; построение чертежей и т.д.

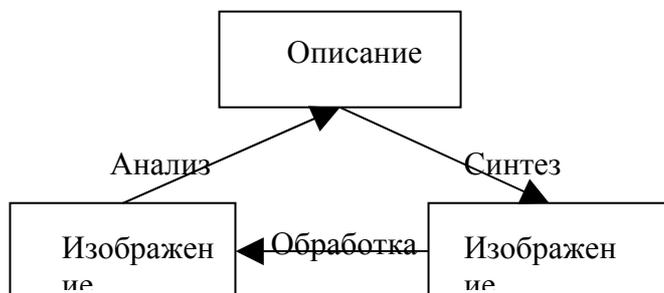


Рис. 1. Взаимосвязь направлений цифровой обработки изображений

Интерактивной графикой называют направление, предполагающее не только синтез изображения машиной, но и ввод пользователем в систему данных в графической форме, например, с помощью светового пера или манипулятора типа "мышь".

Системы обработки и анализа видеоданных, которые обычно работают со входной информацией в виде полутоновых или многоспектральных изображений, в нашей стране обычно объединяют под названием *автоматизированных систем обработки изображений* (АСОИЗ). В настоящее время системы, решающие достаточно сложные и «интеллектуальные» задачи анализа изображений, принято выделять в отдельный класс систем анализа изображений (САИ). Системы, задачами которых являются синтез изображений и обработка графической информации, называют *системами машинной графики* (СМГ); употребляется также термин «системы обработки графической информации» (СОГИ). Следует отметить, что такое деление достаточно условно, и за последнее время наметилась устойчивая тенденция к сближению этих систем. Все чаще АСОИЗ оперируют графической информацией, решая в том числе и задачи синтеза, а СМГ применяются для построения полутоновых изображений с привлечением различных методов их обработки.

Степень «интеллектуализации» АСОИЗ и систем машинной графики может быть различной в зависимости от уровня представления и использования в

этих системах знаний специалистов по обработке изображений и специалистов предметных областей. Она колеблется от возможности применения набора утилит (операций) обработки, анализа или синтеза изображений, которые в грубом приближении можно интерпретировать как записанные в процедурной форме записи специалистов, до поддержки специальных баз знаний в развитых экспертных системах анализа изображений.

Рассмотрим системы трех классов:

- традиционные «мало интеллектуальные» АСОИЗ, без понимания состава, функций и особенностей построения которых трудно представить себе уровень развития интеллектуальных систем;
- системы анализа видеоданных, основанные на знаниях, которые образуют специальный класс экспертных систем анализа изображений;
- интеллектуальные системы машинной графики.

Автоматизированные системы обработки изображений

Одно из направлений наиболее широкого практического применения АСОИЗ— обработка результатов дистанционных исследований. Среди многочисленных развивающихся прикладных областей здесь можно упомянуть: геологию (исследование и поиск природных и ископаемых ресурсов); сельское и лесное хозяйство (обнаружение аномалий, слежение за состоянием участков, предсказание урожая); анализ состояния окружающей среды и климата; метеорологию (предсказание погоды, анализ воздушных течений и температурных полей); гидрологию и океанографию (наблюдение за водными ресурсами, течениями, распределениями температуры, планктона, льдов); картографию (составление фотографических и других карт, накопление картографических данных, территориальный анализ); слежение за объектами на земной поверхности (транспорт, судовождение и т. д.).

Быстро развивается направление обработки биомедицинских данных. Здесь АСОИЗ находят применение в микробиологии и цитологии (анализ снимков с микроскопов, анализы крови, хромосомный анализ), рентгенологии, ультразвукографии, инфраскопии (обработка двух - и трехмерных изоб-

ражений внутренних органов, анализ распределения температур тела, томография), сравнительных исследованиях (идентификация отпечатков пальцев, идентификация состояний больных). Недостаточно развито пока применение АСОИЗ в производственных системах, однако потребности автоматизации управления и производства должны привести к быстрому росту числа применений АСОИЗ в таких областях, как материаловедение (обнаружение дефектов, анализ и предсказание свойств материалов, структурный анализ), сборочные работы (системы технического зрения для роботов и гибких автоматизированных производств), системы ориентации и анализа обстановки в реальном времени (для различных видов транспорта, летательных и космических аппаратов).

Исследовательские АСОИЗ создаются обычно как «универсальные», не привязанные жестко ни к какому классу видеоданных. В то же время данные конкретных предметных областей имеют свою специфику и особенности, что необходимо требует проблемной ориентации АСОИЗ, разрабатываемых для прикладных применений. Первые такие системы были специализированными, предназначенными для решения конкретных узких задач с помощью ограниченного набора алгоритмов. С развитием методов обработки изображений, удешевлением аппаратуры, появлением новых технологий (и первую очередь СБИС) и технических средств (микропроцессоров, интеллектуальных видеотерминалов, памяти большой емкости и т. п.) проявилась устойчивая тенденция к переходу от систем специализированных к системам проблемно-ориентированным, призванным решать некоторый класс задач в рамках данной предметной области.

Проблемно-ориентированные прикладные АСОИЗ используют, как правило, достаточно богатый арсенал различных методов и средств обработки данных, анализируют различную входную информацию, активно используют возможности человека. Поэтому в таких системах должны быть функциональные различные блоки и подсистемы, обеспечивающие не только решение собственно задач обработки изображений, но и ввод и обработку других типов

данных и знаний, накопление, хранение и отображение результатов, организацию взаимодействия с пользователем и другими системами. Наличие различных компонентов, работающих совместно для достижения общей цели, есть признаки превращения современных АСОИЗ в *интегрированные системы*.

Внутренняя интеграция затрагивает различные компоненты АСОИЗ: средства и методы обработки, сами обрабатываемые данные.

Интеграция средств обработки означает объединение в системе различных универсальных и/или специализированных устройств, имеющих возможность параллельной работы над данными.

Интеграция методов обработки означает наличие в системе функциональных подсистем и обрабатывающих алгоритмов для решения широкого круга задач и достижения результатов различными путями (возможно, параллельными). При этом к результату и приходят различными доступными путями, применяя разные методы к одним и тем же данным и интегрируя (на основе опыта и интуиции исследователя или заложенных в систему знаний) получаемую выходную информацию.

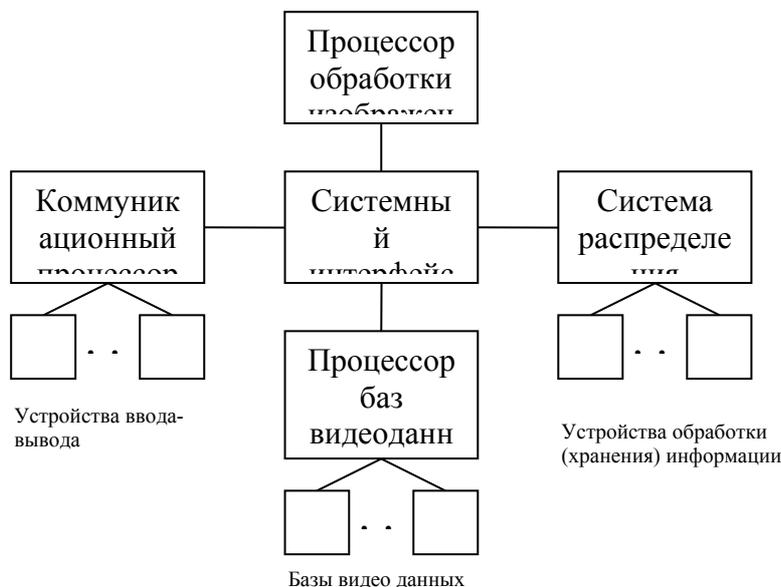


Рис.2 Структура интегрированной АСОИЗ

Работа по анализу изображения становится своего рода многократным экспериментированием с исходными данными. Характерной чертой такого экспе-

римента является совместное применение методов обработки изображений, математической статистики, искусственного интеллекта, анализа данных и других, причем в обработку по возможности вводятся все доступные данные, включая промежуточные результаты. Появляется необходимость организации эффективного обмена данными в системе, создания развитых средств хранения и обеспечения доступа к данным (включая базы видеоданных). В настоящее время это уже характерно для многих АСОИЗ, применяемых в дистанционных исследованиях.

Ключевым звеном АСОИЗ обычно считается процессор(ы), производящий обработку изображений. В первых* системах в качестве обрабатывающего процессора часто применялись крупные универсальные ЭВМ, работающие в режиме разделения времени. Это решение оказалось малоэффективным, поскольку и системы команд, и операционные системы таких ЭВМ не ориентированы на задачи обработки изображений. В архитектуре АСОИЗ 70-х годов наряду с появлением специализированных процессоров обработки изображений получило широкое распространение «монопольное использование средних и малых ЭВМ, ориентация которых на решение задач видеоданных достигалась разработкой соответствующего программного обеспечения. В настоящее время широкое распространение получили так называемые рабочие станции обработки изображений и системы, построенные на базе персональных ЭВМ.

Наличие одного процессора редко удовлетворяет требованиям по производительности системы, особенно при больших потоках обрабатываемых данных. Крупные АСОИЗ строятся на основе специальных многопроцессорных конфигураций для параллельной обработки элементов одного или нескольких изображений, при этом упомянутые средние или малые ЭВМ обычно переходят в разряд управляющих.

Другими архитектурными компонентами АСОИЗ являются:

-устройства ввода видеоданных в систему (ТВ-камеры, дальномеры, малокадровые сканеры, фототелеграфные аппараты и т. д.);

-устройства отображения (ТВ-мониторы, растровые и векторные дисплеи, устройства вывода на твердые носители - бумагу и фотопленку, графопостроители);

- запоминающие устройства (на гибких и жестких магнитных дисках, оптических дисках; устройства расширения оперативной памяти для хранения изображений).

В настоящее время в интегрированных системах аппаратно реализуются не только параллельные процессоры обработки изображений, но и другие подсистемы, ранее традиционно реализуемые программным путем. Прежде всего это средства доступа к различным видам памяти и средства, поддерживающие принятие решений. Особенно большую пользу приносят такие реализации при решении задач распознавания образов, интерпретации видеоданных, анализа сцен. Фактически аппаратные средства здесь поддерживают функционирование базы данных и базы знаний, но на качественно ином уровне. Основные решаемые задачи для базы видеоданных — это поиск эталонных объектов или их отдельных частей (видеопризнаков), поиск необходимых изображений по их идентификаторам. Чтобы решить эти задачи, разрабатываются специализированные процессоры баз видеоданных и массивы ассоциативной памяти с возможностью контекстного доступа к информации (адресация по содержанию).

Важное направление развития программного обеспечения АСОИЗ—создание эффективных систем хранения и поиска видеоданных (баз и банков видеоданных). Потребность в базах видеоданных ощущается в области дистанционных исследований, поэтому именно здесь можно проследить тенденции развития программного обеспечения этих систем.

Все чаще в состав программного обеспечения АСОИЗ включаются и средства ведения баз знаний. Пока это преимущественно специализированные подсистемы для использования базы знаний о конкретной предметной области (например, о поиске месторождений определенного типа на аэрофотоснимке), но в скором времени можно ожидать появления инструментальных средств для

разработки и наполнения знаниями экспертных систем анализа визуальной информации.

Системы анализа изображений

Выделяют три основных класса задач, решаемых САИ (системы анализа изображений), которые определяют их различное назначение:

1. Распознавание объектов. В системах этого типа решаются задачи локализации определенных объектов в поле изображения, измерения заданных параметров этих объектов, принятия решения по полученным результатам измерений. Примерами могут служить системы медицинской диагностики, где необходимо локализовать участки изображения, указывающие на отклонение от нормы, определить степень отклонения; системы контроля протекания процессов (например, сгорания топлива в камере котельной); системы локализации дефектов печатных плат и др.

2. Содержательная интерпретация изображений. К системам, в которых решается эта задача, относится большинство САИ. Сюда входят системы управления движением робота по поступающей видеоинформации, интерпретации аэрофотоснимков, анализа снимков и сопоставления их с картой местности и др. В отличие от систем первого типа в данном случае задача заключается в полной содержательной интерпретации снимка, «понимании» отношений, связывающих объекты реального мира, представленные своими проекциями на двухмерном изображении.

3. Получение справочной информации в архивах видеоданных. Задачей этих систем является выдача информации пользователю о наличии объектов определенного типа на изображениях, хранящихся в архиве видеоданных, его параметрах, отношений к другим объектам и т. п. Запрос может быть представлен как текстовое описание объекта или как видео образ. К системам такого типа относятся различные информационно-поисковые системы в картографии, метеорологии, криминалистике и т. п.

Для решения указанных задач САИ должны обеспечивать выполнение ряда функций, среди которых основными являются следующие.

1.Хранение и поиск формализованных знаний экспертов предметной области и обработки изображений, формализованными знаниями являются установленные факты, закономерности, количественные соотношения, видеоданные (эталоны, прототипы объектов), а также правила, определяющие последовательность и состав операции над данными для достижения заданных целей (например, для получения проекции трехмерного тела на плоскость).

2.Выполнение операций обработки изображений (контрастирование, устранение шума, выделение контурных линий и т. п.).

3.Выполнение операций логических выводов на основе формализованных знаний и результатов обработки видеоданных, построение и проверка гипотез.

4.Анализ промежуточных результатов работы системы и принятие решений о дальнейших действиях, невозможности решения задачи, необходимости дополнительной информации и др.

5.Анализ запросов пользователя, поддержка диалога, документирование действий системы и формирование пояснений пользователю.

6.Пополнение баз знаний и видеоданных системы новыми сведениями, данными.

Конкретная САИ не обязательно выполняет все перечисленные функции.

Современные системы анализа изображений довольно разнообразны по своей структуре и организации взаимодействия ее составных частей.

На рис. 3 приведена упрощенная структурная схема ЭСАИ в ее наиболее полном варианте, соответствующая системе содержательной интерпретации изображений.

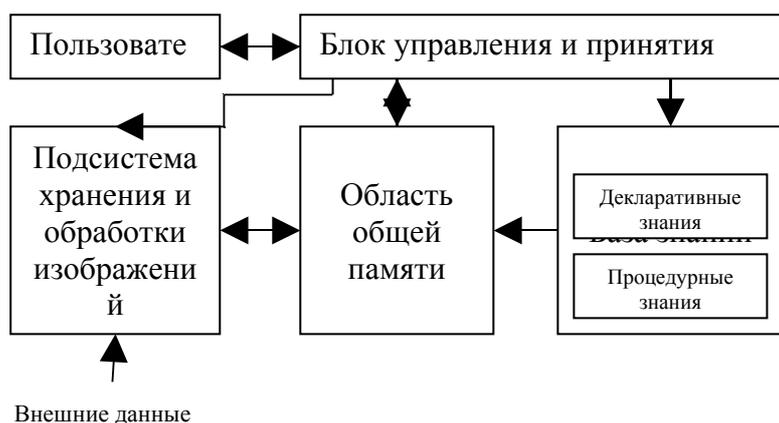


Рис. 3. Структура экспертной системы анализа изображений

Подсистема хранения и обработки изображений содержит программные и аппаратные модули для выполнения различных операций над видеоданными, поиска изображений по видео- и текстовому запросу, а также архив изображений. Важной особенностью подсистемы является модульный принцип построения, благодаря которому система может собирать технологические цепочки последовательностей операций для решения конкретных задач. Решению этой задачи могут предшествовать операции поиска заданного изображения в базе видеоданных.

В базе знаний системы хранятся формализованные знания специалистов предметной области и обработки изображений. Условно в ней можно выделить два компонента: декларативные и процедурные знания.

Декларативные знания отражают сведения об объектах, их составе, взаимном расположении, характере появления, описывают отношения между объектами и т. п. Выделяют три типа описания объектов: контекстное, визуальное и характеризующее повторяемость объекта в сцене. Следует отметить, что для большинства ЭСАИ существенная часть декларативных знаний связана с пространственными, количественными, классификационными отношениями объектов, а такие отношения, как временные, причинно-следственные, пока еще слабо используются,

Особенно следует отметить отношение типа обобщение-спецификация, используемое в большинстве современных ЭСАИ и отражающее иерархию абстракций -последовательности представлений объектов от примитивов (пере-

падов яркостей, однородных участков и т. п.) до абстрактных понятий (здание, дорога, автомобиль и т. п.). Построение иерархий абстракций является ключевым моментом воспроизведения процесса восприятия. Считается, что окончательная интерпретация изображения должна выполняться системой на верхних понятийных уровнях иерархии.

Второй основной частью базы знаний являются процедурные знания, показывающие, какие действия, в какой последовательности нужно предпринимать системе при выполнении определенных условий, отражающих сложившуюся ситуацию анализа данных, формализованных знаний и результатов обработки изображений. Процедурные знания используются для выполнения двух функций: пополнения описания сложившейся ситуации новыми фактами из базы знаний; построения гипотез и выбора конкретных действий по преобразованию полученных результатов и переходу к другому состоянию решения задачи.

Процедурные знания обычно представляются в виде правил, разделенных на две части: условие выполнения действия и назначаемое действие (последовательность действий) в случае выполнения условия. Интерпретация объектов может считаться состоятельной, если выполняются все предусмотренные между ними отношения, и т. д.

Область общей памяти, называемая также доской объявлений (blackboard) или базой данных, является частью системы, обеспечивающей передачу результатов работы всех подсистем друг другу. Обычно область общей памяти представляет собой реляционную базу данных, куда в ходе решения задачи записываются промежуточные интерпретации участков изображений (карты, платы интерпретации), установленные факты, количественные соотношения и другие результаты. В область общей памяти могут заноситься также пояснения действий ЭСАИ, с тем чтобы пользователь мог при необходимости оценить рациональность поведения системы, выявить причину принятия того или иного решения.

Блок управляющих процедур и принятия решений является координирующим центром системы, анализирующим запросы пользователя, сложившуюся ситуацию анализа, определяющим контекст и стратегию решения. Информация о промежуточных результатах анализа поступает в блок из области общей памяти. По результатам анализа состояния области общей памяти (сложившейся ситуации) блок управляющих процедур и принятия решений определяет состоятельность полученной интерпретации, генерирует гипотезы (о принадлежности объектов определенным классам, о возможности нахождения других объектов и т. п.) и «поручает» другим модулям системы выполнение конкретных действий. Окончательное решение также принимается этой подсистемой.

Процесс решения задачи содержательной интерпретации изображения является многоступенчатым, итеративным. На начальной стадии обычно выполняется предварительная сегментация изображения.

Следующим этапом является получение начальной интерпретации результатов предварительной сегментации. Это включает в себя измерения определенных параметров (например, площадей или периметров полученных областей с одинаковой яркостью) и генерацию гипотез (с помощью сопоставления результатов измерений с информацией из базы знаний).

Все результаты интерпретации записываются в область общей памяти. Изменения состояния этой области означают, что процесс интерпретации не закончен. Блок управляющих процедур и принятия решений анализирует состояние области и вырабатывает новую последовательность действий. Процесс решения задачи обычно считается завершенным, если не происходят существенные изменения в области общей памяти (на карте интерпретации).

Обычно в ЭСАИ используются две глобальные стратегии решения задач интерпретации изображений (а также их комбинации): управление целью (называемое также управлением сверху вниз или обратным выводом) и управление данными (называемое также управлением снизу вверх или прямым выводом). В первом случае последовательность действий выбирается исходя из

поставленной цели. Такая стратегия соответствует продвижению сверху вниз по иерархии абстракций: от понятия объекта к его геометрическим примитивам на изображении. Вторая стратегия соответствует движению в обратном направлении по иерархии абстракций. Последовательность действий определяется на основе полученных результатов обработки данных.

Большинство разработчиков ЭСАИ придерживаются мнения, что обе стратегии должны присутствовать в решении задачи системой. Переключение стратегий остается нерешенной проблемой, так же как и нет единого мнения о том, в каких случаях система должна выбирать ту или иную стратегию,

Системы машинной графики

Основными областями применения систем машинной графики являются:

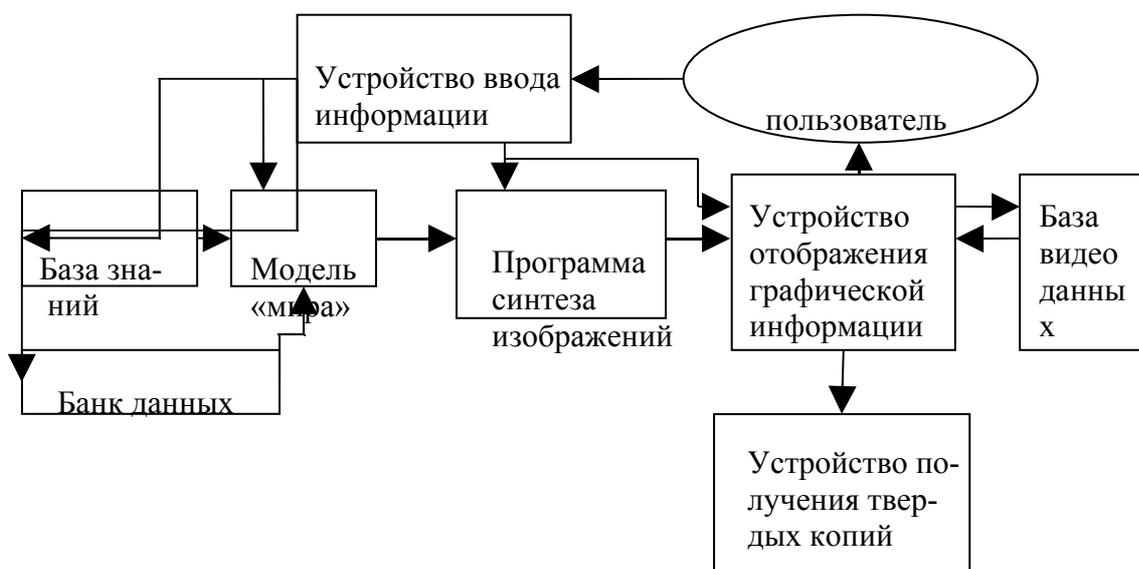
- автоматизация проектирования (построение чертежей, проекций, планов, двух- и трехмерных визуальных моделей деталей и объектов);
- дизайнерские работы в таких областях, как автомобиле- и самолетостроение, архитектура, пошив одежды;
- картография (составление, корректировка, хранение и генерация карт);
- автоматизация исследований (отображение в визуальной форме результатов экспериментов, вычерчивание графиков, гистограмм, диаграмм и т. д.);
- организация эффективного взаимодействия человека с ЭВМ (графические языки, интерактивная графика);
- синтез картин и сцен (для тренажеров самолето- и автовождения, в машинной живописи, в компьютерных играх);
- синтез динамических картин (компьютерная мультипликация и кинематография).

Машинная графика стала обязательной частью любой интеллектуальной системы, реализованной на ЭВМ. Это связано прежде всего с тем, что скорость переработки человеком информации, представленной в графической форме, выше по сравнению со скоростью переработки алфавитно-цифрового набора данных. Особенно заметен этот эффект на этапе восприятия информации,

так как при этом максимально активизируется ассоциативное мышление человека. Наиболее ярко преимущества графического представления данных проявляются в интерактивных системах машинной графики, подразумевающих активную обратную связь с пользователем.

Состав типичной СМГ представлен на рис. 4. Схематически работу системы можно представить следующим образом. На основании базы знаний и банка данных ЭВМ генерирует модель «мира». Затем эта модель преобразуется программой синтеза изображения в удобную для восприятия человеком форму. Сформированная таким образом информация выдается на устройство отображения графической информации, вызывая на нем появление соответствующего изображения, которое может быть задокументировано с помощью устройства получения твердых копий. На этапе окончательного формирования изображения в него могут включаться готовые фрагменты из банка видеоданных. Пользователь через устройство ввода информации может формировать и дополнять базу знаний и банк данных, корректировать модель «мира» и программу синтеза изображения, а также производить непосредственные операции над изображением (вырезание и перемещение фрагментов, их склейку, масштабирование, раскраску и т. д.), воздействуя через сервисные программы непосредственно на буфер устройства отображения.

Модель «мира» является центральным звеном практически всех современ-



ных СМГ.

Рис. 4. Структура системы машинной графики

По способу описание объектов модели «мира» можно разделить на модели, задающие объект в виде набора данных и в виде операции его построения. Сама операция построения может быть представлена в процедурной форме или задаваться структурой данных. Способ описания, использующий полностью детерминированное задание всех исходных данных, характеризующих объект, основывается на концепции имитации процесса фотографирования реальной сцены. Это приводит к необходимости однозначного задания всех объектов модели «мира» в момент ее «фотографирования» воображаемым фотоаппаратом. Сейчас чаще используется задание объектов «мира» в виде операции их построения. В этом случае объект синтезируется непосредственно в процессе построения. Это значительно облегчает задачу разработчика модели, так как вместо детального описания объекта он задает лишь закон его построения (например, в виде закона распределения вероятности реализации конкретных фрагментов объекта). Кроме геометрического описания объектов в модели «мира» хранятся и другие параметры, необходимые для синтеза изображения (например, цвета поверхностей объектов).

При создании модели «мира» используются база знаний и банк данных. База знаний вводит фактически законы и ограничения, действующие при создании объектов «мира». Банк данных позволяет «записывать» модель «мира» стандартными данными, необходимыми для ее работы. Банк данных и база

знаний могут постоянно изменяться и дополняться пользователем в процессе эксплуатации системы.

Сформированную таким образом модель «мира» специальная программа преобразует в массив чисел, соответствующий изображению. Наиболее сложной задачей такого рода является синтез реалистичного изображения трехмерной сцены. Большой прогресс в этой области был достигнут к началу 80-х годов. К настоящему времени в процессе синтеза цветного трехмерного реалистичного изображения можно выделить следующие этапы:

- удаление скрытых поверхностей, т. е. поверхностей или их частей, которые не видны из точки нахождения воображаемого наблюдателя. Этот этап может быть выполнен, например, с помощью «алгоритма Z-буфера» или «алгоритма художника»;
- закраска видимых поверхностей с учетом свойств самой поверхности (цвет, фактура, отражательная способность), ее относительного местоположения, ориентации, а также свойств источников света и других поверхностей;
- выравнивание интенсивностей соседних поверхностей с помощью алгоритмов закраски Гуро или Фонга.

В последнее время благодаря быстрому удешевлению аппаратной части СМГ возрос интерес к синтезу динамических изображений в реальном масштабе времени, т. е. к так называемой компьютерной мультипликации. Динамическая графика находит применение в компьютерных играх, тренажерных системах, АСНИ. При синтезе динамических объектов, движущихся по законам модели «мира», программа синтеза изображения кроме перечисленных выше операций выполняет также ряд геометрических преобразований, соответствующих моделируемому движению.

Банк видеоданных служит для воспроизведения готовых изображений и их фрагментов. Он может постоянно пополняться новыми данными, полученными как извне, так и в процессе функционирования самой системы.

В качестве базовых для синтеза изображения могут использоваться универсальные ЭВМ практически всех классов — от персональных компьютеров до суперЭВМ. Простейшими графическими средствами снабжены и современные бытовые компьютеры. Для синтеза реальных трехмерных сцен необходимы 32-разрядные ЭВМ.

Разработка графических программ на языках высокого уровня началась в конце 60-х годов. При этом широко использовалась идеология «виртуальных» устройств, обеспечивающая независимость программного обеспечения высокого уровня от особенностей внешних устройств конкретной вычислительной системы. Программист пользовался лишь координатами виртуального экрана, а программы нижнего уровня автоматически переводили их в координаты реального экрана.

Приблизительно в этот же период времени были разработаны два пакета графических программ, доступные для всех типов дисплеев. Это трехмерная система Core Graphics System и двумерная система Graphical Kernel System, принятая Международной организацией по стандартизации. Однако у этих пакетов есть существенный недостаток. Они разрабатывались еще до того, как стали отдавать предпочтение растровым дисплеям, поэтому при использовании этих пакетов невозможно в полной мере учитывать преимущества пикселей.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

ЛАБОРАТОРНАЯ РАБОТА № 1. **РАЗРАБОТКА ЭКСПЕРТНЫХ СИСТЕМ, БАЗИРУЮЩИХСЯ НА ПРАВИЛАХ.**

Цель

1. Ознакомление с принципами построения экспертных систем.
2. Изучение структуры экспертных систем, базирующихся на правилах.
3. Построение простейшей экспертной системы, базирующейся на правилах.

Общие сведения

Экспертная система - это компьютерная программа, созданная для выполнения тех видов деятельности, которые под силу только человеку-эксперту, – например, проектирования, планирования, постановки диагноза, перевода, реферирования, ревизии, выдачи рекомендаций. Сферы применения экспертных систем – бизнес, проектирование, исследования, управление.

Программы ЭС обычно работают таким способом, который воспринимается как “интеллектуальный”, т. е. они имитируют образ действий человека-эксперта.

Эти программы специфичны, поскольку, как правило, используют механизм автоматического рассуждения (вывода) и так называемые слабые методы – такие как поиск, или эвристика. Они существенно отличаются от точных и хорошо аргументированных алгоритмов и не

похожи на математические процедуры большинства традиционных разработок.

Основными компонентами экспертных систем являются:

- база знаний (БЗ), содержащая формализованное описание методов и знаний, привлекаемых при решении задач из области применения экспертных систем;

- механизм вывода (МВ), содержащий формализованное описание правил извлечения знаний из БЗ;

- система пользовательского интерфейса (СПИ), осуществляющая передачу знаний от МВ к пользователю.

В процессе работы экспертной системы (консультации) входные данные сопоставляются с данными из БЗ. Результатом сопоставления является утвердительный или отрицательный ответ. В экспертных системах, базирующихся на правилах, утвердительный ответ является результатом наличия в БЗ соответствующего продукционного правила. Выбор и активизацию продукционного правила реализует интерпретатор МВ. В каждом цикле работы интерпретатора (называемом распознавание – действие) производятся следующие действия:

- образец правила сопоставляется с элементами данных из БЗ;

- если можно активизировать более одного правила, то для выбора правила используется механизм разрешения конфликта (здесь не рассматривается);

- применяется выбранное правило.

Пример реализации экспертной системы выбора породы собаки, базирующейся на правилах, приведен в прил. 1 (LAB01.PRO).

В программу включен дополнительный раздел **database**, содержащий определение предикатов динамической базы данных (БД).

Запись данных в БЗ производится стандартным предикатом

asserta (Факт),

в результате активизации которого указанный в скобках факт будет

добавлен в начало БД.

Удаление фактов из БД производится стандартным предикатом **retract (Факт),**

в результате активизации которого из БД удаляется первый факт, отождествленный с фактом, указанным в скобках.

При сопоставлении правил с содержимым БД используется стандартная функция отрицания (not), считающаяся выполненной успешно, если заданный в ней атом представляет собой цель, которая не достигается.

Задание к лабораторной работе

1. Провести тестирование программы LABO1.PRO (см. прил. 1).
2. Изменить программу LABO1.PRO так, чтобы перед окончанием работы выводилось содержимое БД, например, путем использования предиката вида

```
wr_bd:- dpositive(P,Q),  
       write("dpositive(",P," ",Q,")"), nl,  
       fail.
```
3. Изменить программу LABO1.PRO так, чтобы она обеспечивала распознавание животных в соответствии с правилами, приведенными в прил. 2.

Порядок выполнения задания

1. Загрузить Турбо- Пролог.
2. Загрузить программу LABO1.PRO и убедиться в правильности ее работы.
3. Внести требуемые изменения.

ЛАБОРАТОРНАЯ РАБОТА № 2 **РАЗРАБОТКА ЭКСПЕРТНЫХ СИСТЕМ, БАЗИРУЮЩИХСЯ НА ЛОГИКЕ**

Цель

1. Изучение структуры экспертных систем, базирующихся на логике.
2. Построение простейшей экспертной системы, базирующейся на логике.

Общие сведения

Структура экспертной системы, базирующейся на логике, аналогична структуре экспертной системы, базирующейся на правилах - БЗ состоит из утверждений в виде предложений логики предикатов; МВ реализует процесс распознавание – действие; СПИ выполняет те же функции, что и в системах, базирующихся на правилах.

Пример экспертной системы по породам собак, базирующейся на логике, приведен в прил. 1 в виде программы на Турбо-Прологе.

Программа выдает начальное меню, предлагая пользователю выбор между **consultation** (консультацией) и **exit from the system** (выходом из системы). Если пользователь выбирает консультацию, то между пользователем и системой происходит диалог. Затем пользователю сообщается результат. Результатом является либо выбранная порода, либо сообщение **Sorry. I can't help you** (Извините, я не могу вам помочь).

БЗ содержит утверждения логики предикатов, которые представлены либо в форме **rule** (правило), либо в форме **cond** (условие). В форме rule хранятся данные о породе; в форме cond-атрибуты (условия), характеризующие породу. Данные (ответы), получаемые от пользователя, динамически записываются в БД в форме предикатов **yes** (да) и **no**(нет).

МВ организован следующим образом: в результате активизации правила **go** осуществляется просмотр утверждений из БД rule и cond для выяснения существования или отсутствия подходящих значений данных. С этой целью вызывается правило **check** (проверка). Это правило содержит трассу номеров правил, номера условий и классифицированные объекты в БЗ. Оно пытается сопоставить объекты, классифицированные

при помощи номеров условий. Если сопоставление происходит, то в программу добавляются сопоставленные значения и продолжается сопоставление с новыми данными, полученными от пользователя. Если сопоставления не происходит, МВ останавливает текущий процесс и выбирает другую трассу. Поиск и сопоставление продолжаются до тех пор, пока не исчерпаны все возможности. По завершении вывода go через интерфейс передает результаты пользователю.

СПИ состоит из трех частей: в первой содержатся правила для организации меню и уничтожения соответствующего окна после выбора пользователем предлагаемой ему программной функции: либо проведение консультации, либо выход из системы; вторая обеспечивает вывод списка собак и инициализацию процесса поиска и сопоставления по образцу; третья запрашивает и получает ответы (yes или no) от пользователя.

Задание к лабораторной работе

1. Провести тестирование программы LABO2.PRO (см. прил. 1).
2. Изменить программу LABO2.PRO так, чтобы она обеспечивала распознавание животных в соответствии с правилами, приведенными в прил. 2.

Порядок выполнения задания

1. Загрузить Турбо- Пролог.
2. Загрузить программу LABO2.PRO и убедиться в правильности её работы.
3. Внести требуемые изменения.

ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ НА РАЗРАБОТКУ ФРАГМЕНТОВ ЭС

1. Придумайте систему управления лифтом, предусматривающую кнопки для трех этажей, а также для закрытия и открытия дверей, сенсорный датчик для обнаружения препятствия, возникающего при

закрытии дверей, реле–таймер для фиксации времени, в течение которого двери открыты, и отдельные кнопки вызова на каждом этаже. Напишите систему продукций для управления лифтом. Пример типичной продукции:

Если срок действия таймера истек и
дверь открыта и
ничто не препятствует закрытию двери

То закрыть дверь

2. Требуется разработать фрагмент экспертной системы, предназначенной для прогнозирования местной погоды.

3. Требуется разработать фрагмент экспертной системы, предназначенной для обнаружения того, что делать, если автомобиль не заводится.

4. Требуется разработать фрагмент экспертной системы, предназначенной для определения стратегии гоночной яхты в регате. Система должна работать в режиме реального времени на протяжении всего периода гонок.

5. Требуется разработать фрагмент экспертной системы, предназначенной для помощи отвечающему по телефону доверия, когда отвечающий должен определить яд, который мог быть принят звонящим.

6. Требуется разработать фрагмент экспертной системы, предназначенной для определения оптимального маршрута продавца в любой данный день, причем продавец должен посетить всех клиентов и израсходовать возможное количество бензина.

7. Требуется разработать фрагмент экспертной системы, предназначенной для подбора членов экипажа с учетом их совместимости и возраста. При этом учитываются индивидуальные особенности претендентов, их совместимость, пожелания тех или иных кандидатов работать вместе, а также их возраст (командир должен быть старше всех остальных членов экипажа).

8. Поиск информации в базе данных из другого модуля программы. В

базу знаний занесена информация о сотрудниках:

- а) фамилия;
- б) год рождения;
- в) место рождения;
- г) национальность;
- д) профессия;
- е) место работы;
- ж) занимаемая должность.

Система должна по вводимой с клавиатуры частичной информации о сотрудниках находить в базе знаний полную информацию и выводить ее на экран либо сообщать, что необходимая информация в базе данных отсутствует.

9. Требуется разработать фрагмент экспертной системы, предназначенной для поиска свободного места в салоне самолета. Система запрашивает требования к свободному месту

- а) класс салона;
- б) у окна или прохода;
- в) для некурящих или нет

и выводит на экран номер свободного места. Программа и данные содержатся в различных модулях.

10. Требуется разработать фрагмент экспертной системы, предназначенной для диагностики неисправностей персонального компьютера. Экспертная система должна исследовать ситуацию и попытаться определить на общем уровне, допускает ли ошибки пользователь или, действительно, имеется неисправность в системном блоке, на диске, в мониторе и т.д. Возможный путь проектирования – беседа с мастером–профессионалом. При оценке ситуации подразумеваются грубые функциональные тесты, без глубокого анализа электронных элементов.

11. Требуется разработать фрагмент экспертной системы, предназначенной для подбора субоптимальной конфигурации персонального

компьютера с учетом субъективных и объективных потребностей заказчика.

12. Требуется разработать фрагмент экспертной системы, предназначенной для подбора субоптимальной конфигурации локальной компьютерной сети с учетом множества эксплуатационных, финансовых и прочих важных критериев.

13. Требуется разработать фрагмент экспертной системы, предназначенной для диагностики широко распространенных заболеваний человека по совокупности симптомов. Диагностируется не менее 20 болезней с учетом 15 типовых симптомов. Каждый симптом может указывать на несколько болезней (возможно, с разной степенью уверенности).

14. Требуется разработать фрагмент экспертной системы, предназначенной для консультации в отношении покупки автомобиля с учетом субъективных факторов, объективных потребностей и платежеспособности клиента, а также сезона и др.

15. Требуется разработать фрагмент экспертной системы, предназначенной для консультации в отношении покупки недвижимости с учетом связанных с этим важных факторов (надежность продавца, платежеспособность покупателя, страхование сделки, изменение цен и банковских процентных ставок и др.).

16. Требуется разработать фрагмент экспертной системы, предназначенной для выработки рекомендаций по обустройству дорожной сети (установка дорожных знаков, ограждений, нанесение разметки и т.п.). Вид ДТП - наезды на пешеходов, неожиданно выходящих на проезжую часть из-за стоящих транспортных средств.

Возможные мероприятия:

- установка знака 3.27 - "Остановка запрещена";
- установка знака 3.28 - "Стоянka запрещена";
- замена знака 3.28 знаком 3.27;
- обеспечение видимости знака в дневное и ночное время;
- установка дорожного ограждения перильного типа;

- организация поблизости пешеходного перехода;
- организация поблизости оборудованной стоянки транспортных средств;
- принятие мер по соблюдению водителями требований знака.

При выработке рекомендаций следует принять во внимание, не является ли участок дороги, где совершаются наезды, местом притяжения транспортных средств (например, рядом находится крупное учреждение) и/или пешеходов (остановка общественного транспорта, универмаг, кинотеатр и т.п.).

Дорожные знаки должны размещаться с учетом их наилучшей видимости участниками дорожного движения как в светлое, так и в темное время суток, при этом они не должны закрываться от участников дорожного движения какими-либо препятствиями – зелеными насаждениями, мачтами наружного освещения и т.п.. На участках дорог без стационарного освещения следует применять знаки со световозвращающей поверхностью, а на участках со стационарным освещением – знаки с внутренним и внешним освещением. В последнем случае возможна также установка знаков со светоотражающей поверхностью, если обеспечена их видимость с расстояния не менее 100 м.

В соответствии с правилами дорожного движения остановка и стоянка запрещены (даже при отсутствии соответствующего знака) на пешеходных переходах и ближе 5 м перед ними, а также ближе 15 м от остановочных площадок или указателей остановки общественного транспорта, если это создает помехи их движению.

ПРИЛОЖЕНИЕ 1

Текст программы работы №1 (LABO1.PRO).

/*	Пример экспертной системы	*/
/*	базирующейся на правилах	*/
/*	Эксперт по породам собак	*/
domains		

database

dpositive(symbol, symbol)

dnegative(symbol, symbol)

predicates

do_expert_job

do_consulting

ask(symbol, symbol)

dog_is(symbol)

it_is(symbol)

positive(symbol, symbol)

negative(symbol, symbol)

remember(symbol, symbol, symbol)

clear_facts

goal

do_expert_job.

Clauses

/* Система пользовательского интерфейса */

do_expert_job:-

makewindow(1,7,7, "Экспертная система",1,16,22,58),

nl,write("*****
*****"),

nl,nl,

write(" Добро пожаловать в ЖИВОТНУЮ экспертную
систему! ;)"),

nl,nl,write(" Эта система легко определит название животного
по "),

```

nl,write(" его признакам.          "),
nl,write(" Отвечайте на вопросы : 'Y'(Да) или 'N'(Нет).  "),
nl,write("*****"),
*****),

nl,nl,
do_consulting,
nl,nl,
clear_facts,
write("Нажмите пробел."),nl,
readchar(_),
removewindow,
exit.

do_consulting:-dog_is(X),!,nl,
write("Похоже, что это - ", X, ".").
do_consulting:-nl, write("Извините, но я ничем не могу вам
помочь."), nl,
write("И          вообще, где вы видели такое животное ?..").
ask(X,Y):- write(" Вопрос: ",X," ",Y," ? "),
readln(Reply),
remember(X,Y,Reply).

/*    Механизм вывода    */
positive(X,Y):-dpositive(X,Y),!.
positive(X,Y):-not(negative(X,Y)),!, ask(X,Y).
negative(X,Y):-dnegative(X,Y),!.
remember(X,Y,y):-asserta(dpositive(X,Y)).
remember(X,Y,n):-asserta(dnegative(X,Y)), fail.
clear_facts:-retract(dpositive(_, _)), fail.
clear_facts:-retract(dnegative(_, _)), fail.

```

/* Продукционные правила */

```
dog_is("Английский бульдог):-  
    it_is("короткая шерсть"),  
    positive(has,"рост меньше 55 см"),  
    positive(has,"низкопосаженный хвост"),  
    positive(has,"хороший характер"),!.  
dog_is("Гончая):-  
    it_is("короткая шерсть"),  
    positive(has,"рост меньше 55 см"),  
    positive(has,"длинные уши"),  
    positive(has,"хороший характер"),!.  
dog_is ("Дог):-  
    it_is("короткая шерсть"),  
    positive(has,"низкопосаженный хвост"),  
    positive(has,"хороший характер"),  
    positive(has,"вес больше 5 кг"),!.  
dog_is("Американская гончая):-  
    it_is("короткая шерсть"),  
    positive(has,"рост меньше 75 см"),  
    positive(has,"длинные уши"),  
    positive(has, "хороший характер"),!.  
dog_is("Коккер-спаниель):-  
    it_is("длинная шерсть"),  
    positive(has,"рост меньше 55 см"),  
    positive(has,"низкопосаженный хвост"),  
    positive (has,"длинные уши"),  
    positive (has,"хороший характер"),!.  
dog_is("Ирландский сеттер):-  
    it_is("длинная шерсть"),  
    positive(has,"рост меньше 75 см"),
```

```

        positive(has,"низкопосаженный хвост"),
        positive(has,"длинные уши"),!.
dog_is ("Колли):-
    it_is("длинная шерсть"),
    positive(has,"рост меньше 75 см"),
    positive(has,"низкопосаженный хвост"),
    positive(has,"хороший характер"),!.
dog_is("Сенбернар):-
    it_is("длинная шерсть"),
    positive(has,"низкопосаженный хвост"),
    positive(has,"хороший характер"),
    positive(has,"вес больше 5 кг"),!.
it_is("короткая шерсть):-
    positive(has,"короткая шерсть"),!.
it_is("длинная шерсть):-
    positive (has,"длинная шерсть"),!.
/*    конец программы        */

```

Текст программы работы №2 (LABO2.PRO).

```

/*    Пример экспертной системы,        */
/*    базирующейся на логике.           */
/*    Эксперт по породам собак         */
domains
    conditions = bno*
    rno,bno,fno =integer
    category = symbol

database
/* предикаты базы данных */

```

```
rule(rno,category,category,conditions)
cond(bno,symbol)
yes(bno)
no(bno)
topic(symbol)
```

predicates

```
/* предикаты системы пользовательского интерфейса */
```

```
do_expert_job
show_menu
do_consulting
process(integer)
info(category)
goes(category)
listopt
erase
clear
eval_reply(char)
```

```
/* предикаты механизма вывода */
```

```
go(category)
check(rno,conditions)
inpo(rno,bno,string)
do_answer(rno,string,bno,integer)
```

goal

```
do_expert_job.
```

clauses

```
/* база знаний */
```

```
topic("dog").
```

topic("Короткошерстная собака").

topic("Длинношерстная собака").

rule(1,"dog","Короткошерстная собака",[1]).

rule(2,"dog","Длинношерстная собака",[2]).

rule(3,"Короткошерстная собака","Английский бульдог",
[3,5,7]).

rule(4,"Короткошерстная собака","Гончая", [3,6,7]).

rule(5,"Короткошерстная собака","Дог", [5,6,7,8]).

rule(6,"Короткошерстная собака","Американская гончая",
[4,6,7]).

rule(7,"Длинношерстная собака","Коккер-спаниель", [3,5,6,7]).

rule(8,"Длинношерстная собака","Ирландский сеттер", [4,6]).

rule(9,"Длинношерстная собака","Колли", [4,5,7]).

rule(10,"Длинношерстная собака","Сенбернар", [5,7,8]).

cond(1,"Короткая шерсть").

cond(2,"Длинная шерсть").

cond(3,"Рост меньше 55 см").

cond(4,"Рост меньше 75 см").

cond(5,"Низкопосаженный хвост").

cond(6,"Длинные уши").

cond(7,"Хороший характер").

cond(8,"Вес более 5 кг").

/* Система пользовательского интерфейса */

do_expert_job:-

makewindow(1,7,7,"DOG EXPERT SYSTEM",0,0,25,80),

show_menu,

nl,write("Press spase bar."),

```
readchar(_),
```

```
exit.
```

```
show_menu:-
```

```
write("                "),nl,  
write("*****"),nl,  
write("*      DOG EXPERT      *"),nl,  
write("*                *"),nl,  
write("*  1. Consultation      *"),nl,  
write("*                *"),nl,  
write("*  2. Exit the system   *"),nl,  
write("*                *"),nl,  
write("*****"),nl,  
write("                "),nl,  
write("Please enter your choice: 1 or 2: "),nl,  
readint(Choice),  
process(Choice).
```

```
process(1):-do_consulting.
```

```
process(2):-removewindow, exit.
```

```
do_consulting:-goes(Mygoal),go(Mygoal),!.
```

```
do_consulting:-nl,write("Sorry, I can't help you."),  
clear.
```

```
do_consulting.
```

```
goes(Mygoal):-clear,clearwindow,nl,nl,
```

```
write("                "),nl,
```

```
write(" WELCOME TO THE DOG EXPERT SYSTEM "),nl,
```

```
write("                "),nl,
```

```
write(" This is a dog identification system."),nl,
```

```
write(" To begin the process of choosing a "),nl,
```

```
write(" dog, please type in 'dog'. If you "),nl,
```

```
write(" wish to see the dog types, please "),nl,
```

```

write(" type in a question mark (?).      "),nl,
write("                                "),nl,
readln(Mygoal),
info(Mygoal),!.
info("?"):-clearwindow,
write("Reply from the KBS."),nl,
listopt,nl,
write("Please any key."),
readchar(_),
clearwindow,
exit.
info(X) :- X >< "?".
listopt :-
write("The dog types are: "),nl,nl,
topic(Dog),
write("      ",Dog),nl,fail.
listopt.
inpo(Rno,Bno,Text) :-
write("Question :-",Text," ? "),
makewindow(2,7,7,"Response",10,54,7,20),
write("Type 1 for 'yes': "),nl,
write("Type 2 for 'no' : "),nl,
readint(Response),
clearwindow,
shiftwindow(1),
do_answer(Rno,Text,Bno,Response).
eval_reply('y') :-
write("I hope you have found this helpful !").
eval_reply('n') :-
write("I am sorry I can't help you !").

```

```

go(Mygoal) :-
    not(rule(_,Mygoal,_,_)),!,nl,
    write("The dog you have indicated is a(n) ",Mygoal,"."),nl,
    write(" Is a dog you would like to have (y/n) ?"),nl,
    readchar(R),
    eval_reply(R).

```

/* МЕХАНИЗМ ВЫВОДА */

```

go(Mygoal) :-
    rule(Rno,Mygoal,Ny,Cond),
    check(Rno,Cond),
    go(Ny).
check(Rno,[Bno|Rest]) :-
    yes(Bno),!,
    check(Rno,Rest).
check(_,[Bno|_]) :- no(Bno),!,fail.
check(Rno,[Bno|Rest]) :-
    cond(Bno,Text),
    inpo(Rno,Bno,Text),
    check(Rno,Rest).
check(_,[]).
do_answer(_,_,_,0) :- exit.
do_answer(_,_,Bno,1) :-
    assert(yes(Bno)),
    shiftwindow(1),
    write(yes),nl.
do_answer(_,_,Bno,2) :-
    assert(no(Bno)),
    write(no),nl,
    fail.

```

erase :- retract(_),fail.

erase.

clear :-

retract(yes(_)),

retract(no(_)),

fail,!.
clear.

/* конец программы */

ПРИЛОЖЕНИЕ 2

Система правил для экспертной системы распознавания животных

ПРАВИЛО 1

ЕСЛИ животное имеет волосяной покров

ТО это животное – млекопитающее

ПРАВИЛО 2

ЕСЛИ животное дает молоко

ТО это животное – млекопитающее

ПРАВИЛО 3

ЕСЛИ животное имеет перья

ТО это – птица

ПРАВИЛО 4

ЕСЛИ животное может летать

И откладывает яйца

ТО это животное – птица

ПРАВИЛО 5

ЕСЛИ животное ест мясо

ТО это животное – хищник

ПРАВИЛО 6

ЕСЛИ животное имеет острые зубы

И животное имеет когти

И его глаза смотрят вперед

ТО это животное – хищник

ПРАВИЛО 7

ЕСЛИ животное является млекопитающим

И имеет копыта

ТО это животное – парнокопытное

ПРАВИЛО 8

ЕСЛИ животное является млекопитающим

И жует жвачку

ТО это животное – парнокопытное

ПРАВИЛО 9

ЕСЛИ животное является млекопитающим

И это животное – хищник

И это животное желто-коричневого цвета

И это животное имеет темные пятна

ТО можно предположить, что это животное – гепард

ПРАВИЛО 10

ЕСЛИ животное является млекопитающим

И это животное – хищник

И это животное желто-коричневого цвета

И это животное имеет темные полосы

ТО можно предположить, что это животное – тигр

ПРАВИЛО 11

ЕСЛИ животное является парнокопытным

И имеет длинную шею

И имеет длинные ноги

И имеет черные пятна

ТО можно предположить, что это животное – жираф

ПРАВИЛО 12

ЕСЛИ животное является парнокопытным

И имеет черные полосы

ТО можно предположить, что это животное – зебра

ПРАВИЛО 13

ЕСЛИ животное является птицей

И не может летать

И имеет длинную шею

И имеет длинные ноги

И имеет черно-белую окраску

ТО можно предположить, что это животное – страус

ПРАВИЛО 14

ЕСЛИ животное является птицей

И не может летать

И может плавать

И имеет черно-белую окраску

ТО можно предположить, что это животное – пингвин

ПРАВИЛО 15

ЕСЛИ животное является птицей

И хорошо летает

ТО можно предположить, что это животное – альбатрос

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОРГАНИЗАЦИИ МЕЖСЕССИОННОГО КОНТРОЛЯ ЗНАНИЙ СТУДЕНТОВ

1. Межсессионная аттестация студентов проводится дважды в семестр на 7 и 13 неделях 5-го семестра.
2. Аттестационная оценка складывается из оценок, полученных по результатам промежуточного тестирования.
3. Организация аттестации студентов, проводится в соответствии с положением АмГУ о курсовых экзаменах и зачетах.

6. Экспертная система, которая решает часть требуемых задач, демонстрируя жизнеспособность метода инженерии знаний, называется:

- 1) демонстрационным прототипом;
- 2) исследовательским прототипом;
- 3) действующим прототипом;
- 4) коммерческой системой.

7. Экспертная система, участвующая в подготовке всего набора необходимых документов на создание “объектов” с заранее определенными свойствами, классифицируется как:

- 1) обучающая система;
- 2) диагностирующая система;
- 3) проектирующая система;
- 4) планирующая система.

8. ... предназначены для формирования описания ситуаций по результатам наблюдений или данным, получаемым от различного рода сенсоров:

- 1) интерпретирующие системы;
- 2) прогнозирующие системы;
- 3) диагностические системы;
- 4) системы проектирования.

9. Системы этого типа создаются на основе какой-нибудь экспертной системы, достаточно хорошо зарекомендовавшей себя на практике, называются:

- 1) средой программирования, поддерживающей несколько парадигм;
- 2) оболочкой экспертной системы;
- 3) дополнительным модулем;
- 4) языком программирования высокого уровня.

10. Этап, на котором определяются способы представления всех видов знаний, основные понятия, способы интерпретации знаний, моделируется работа системы, называется:

- 1) этапом идентификации;
- 2) этапом концептуализации;
- 3) этапом формализации;
- 4) этапом тестирования.

11. Для достижения эффективного функционирования экспертной системы необходимо осуществить:

- 1) переформулирование понятий и требований;
- 2) переконструирование представлений;
- 3) усовершенствование прототипа;
- 4) структурирование знаний.

12. Чтобы быть полноправным участником общения система переработки визуальной информации должна выполнять некоторые обязательные функции. Одной из таких функций является:

- 1) локализация определенных объектов в поле изображения и измерение заданных параметров этих объектов;
- 2) преобразование поступающих от пользователя высказываний на естественном языке в высказывания на языке внутреннего представления;
- 3) последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке;
- 4) сжатие информационного потока за счет введения операции распознавания смысловых элементов речи.

13. Процесс представляет собой последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке, который должен максимально воссоздавать смысл и, как правило, структуру исходного текста, но уже средствами выходного языка и относится к:

- 1) естественно-языковым системам;
- 2) системам распознавания речи;
- 3) системам переработки визуальной информации;
- 4) системам машинного перевода.

14. Системы обработки и анализа видеоданных, которые обычно работают со входной информацией в виде полутоновых или многоспектральных изображений объединяют под названием:

- 1) интерактивная графика;
- 2) автоматизированные системы обработки изображений;
- 3) системы обработки графической информации;
- 4) системы технического зрения для роботов.

15. Унификационные грамматики, надежные грамматики и грамматика трансфера – это программное обеспечение

- 1) ЕЯ-систем;
- 2) систем распознавания речи;
- 3) систем переработки визуальной информации;
- 4) систем машинного перевода.

16. Функция от любого числа аргументов, принимающая истинностные значения: истинно (1) и ложно (0). Аргументы принимают значения из произвольного, конечного или бесконечного множества D , называемого предметной областью, называется:

- 1) правильно построенной формулой;
- 2) предикатом;
- 3) атомом;
- 4) пропозициональной формой.

17. Укажите правильный вариант высказывания “Если у меня будет отпуск летом и я куплю автомобиль и катер, то поеду отдыхать в Крым или на Кавказ”, записанного с использованием логики предикатов первого порядка:

- 1) $(a \vee b \vee c) \leftrightarrow (d \vee e)$.
- 2) $(a \wedge b \wedge c) \rightarrow (d \wedge e)$.
- 3) $(a \vee b \wedge c) \leftrightarrow (d \vee e)$.
- 4) $(a \wedge b \wedge c) \rightarrow (d \vee e)$.

18. Укажите правильный вариант высказывания “Некоторые элементарные частицы имеют положительный заряд” записанного с использованием логики предикатов первого порядка:

- 1) $(\forall X)(S(X) \vee P(X))$
- 2) $(\forall X)(S(X) \wedge P(X))$
- 3) $(\exists X)(S(X) \wedge P(X))$
- 4) $(\exists X)(S(X) \vee P(X))$

19. Восприятие фактов посредством полученной извне информации о некотором явлении с уже имеющимися данными, накопленными опытным путем или полученными в результате вычислений. Когда человек попадает в новую ситуацию, он вызывает из своей памяти основную структуру, называемую:

- 1) правилом продукции;
- 2) логической единицей;
- 3) фреймом;
- 4) семантической сетью.

20. Заключение одного правила является посылкой другого правила, и можно установить третье правило с посылкой из первого правила и заключением из второго. Другими словами это есть:

- 1) модус поненс;
- 2) гипотетический силлогизм;
- 3) предложение Хорна;
- 4) ядро правила.

Утверждено на заседании кафедры
«23» октября 2006 г.
«УТВЕРЖДАЮ» _____

Кафедра ИУС
Заведующий кафедрой
А.В.Бушманов

Инструкция: Тест состоит из 20 заданий, время тестирования – 40 минут.

ВАРИАНТ № 2

1. Экспертные системы отличаются от других видов программ из области искусственного интеллекта тем, что:

- 1) являются системами, процесс работы которых основан на применении правил отношений к символическому представлению знаний;
- 2) имеют дело с предметами реального мира, операции с которыми обычно требуют наличия значительного опыта, накопленного человеком;
- 3) являются программами, специализирующимися на определенных задачах из представленного перечня;
- 4) являются системами, требующими принятия решения, причем могут получить его непосредственно от программы или через промежуточное звено.

2. Устройство экспертной системы, которое используя исходные данные из рабочих памяти и знания из базы знаний формирует такую последовательность правил приводящую к решению задачи, называется:

- 1) компонент приобретения знаний;
- 2) объяснительный компонент;
- 3) решатель;
- 4) диалоговый компонент.

3. Экспертные системы, которые представляют знания об области экспертизы в виде правил и определяют образец некоторой ситуации, называются:

- 1) глубинными;
- 2) традиционными;
- 3) гибридными;
- 4) поверхностными.

4. Если наряду с наличием выразительного языка представления знания существует некоторое средство использования представлений, сконструированных и интерпретируемых таким образом, чтобы с их помощью можно было решить проблему, это означает:

- 1) побочный эффект;
- 2) естественность нотации;
- 3) логическую адекватность;
- 4) эвристическую мощь.

5. Множество решений, которые удовлетворяют условию проверить, не является ли образовавшееся состояние конечным решением, иногда называют:

- 1) пространством решений;
- 2) алгоритмом поиска в ширину;
- 3) алгоритмом поиска в глубину;
- 3) комбинаторным взрывом.

6. Инструментальные средства, которые тщательно проверены, имеют документацию и поддерживаются разработчиком, но действуют неэффективно, называются:

- 1) исследовательскими инструментальными средствами;
- 2) экспериментальными инструментальными средствами;
- 3) коммерческими инструментальными средствами;
- 4) настраиваемыми оболочками.

7. Экспертная система, позволяющая предсказывать последствия некоторых событий или явлений на основании анализа имеющихся данных, классифицируется как:

- | | |
|-----------------------------|----------------------------|
| 1) диагностирующая система; | 2) прогнозирующая система; |
| 3) управляющая система; | 4) проектирующая система. |

8. ... предназначены для логического анализа возможных последствий заданных ситуаций или событий:

- | | |
|------------------------------|----------------------------|
| 1) интерпретирующие системы; | 2) прогнозирующие системы; |
| 3) диагностические системы; | 4) системы проектирования. |

9. Инструментальные средства этой категории избавляют разработчика от необходимости углубляться в детали реализации системы - способы эффективного распределения памяти, низкоуровневые процедуры доступа и манипулирования данными, называются:

- 1) средой программирования, поддерживающей несколько парадигм;
- 2) оболочкой экспертной системы;
- 3) дополнительным модулем;
- 4) языком программирования высокого уровня.

10. Этап, который разделяют на извлечение знаний из эксперта; организацию знаний, обеспечивающую эффективную работу системы; представление знаний в виде, понятном экспертной системе, называется:

- | | |
|--------------------------|-----------------------------|
| 1) этапом выполнения; | 2) этапом тестирования; |
| 3) этапом идентификации; | 4) этапом концептуализации. |

11. В процессе циклического прохождения через этапы выполнения и тестирования для отладки правил и процедур вывода необходимо выполнять:

- 1) переформулирование понятий и требований;
- 2) переконструирование представлений;
- 3) усовершенствование прототипа;
- 4) структурирование знаний.

12. Чтобы быть полноправным участником общения ЕЯ-система должна выполнять некоторые обязательные функции. Одной из таких функций является:

- 1) локализация определенных объектов в поле изображения и измерение заданных параметров этих объектов;
- 2) преобразование поступающих от пользователя высказываний на естественном языке в высказывания на языке внутреннего представления;
- 3) последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке;
- 4) сжатие информационного потока за счет введения операции распознавания смысловых элементов речи.

13. Процесс представляет собой последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке, который должен максимально воссоздавать смысл и, как правило, структуру исходного текста, но уже средствами выходного языка и относится к:

- | | |
|--|---------------------------------|
| 1) естественно-языковым системам; | 2) системам распознавания речи; |
| 3) системам переработки визуальной информации; | |
| 4) системам машинного перевода. | |

14. Системы обработки и анализа видеоданных, которые обычно работают со входной информацией в виде полутоновых или многоспектральных изображений объединяют под названием:

- 1) интерактивная графика;
- 2) автоматизированные системы обработки изображений;
- 3) системы обработки графической информации;
- 4) системы технического зрения для роботов.

15. Направление, предполагающее не только синтез изображения машиной, но и ввод пользователем в систему данных в графической форме, называют:

- 1) машинной графикой;
- 2) интерактивной графикой;
- 3) системой анализа изображений;
- 4) компьютерной графикой.

16. Всякое высказывание, составленное из некоторых исходных высказываний посредством логических операций, называют:

- 1) формулой алгебры логики;
- 2) правильно построенной формулой;
- 3) предикатом;
- 4) атомом.

17. Укажите правильный вариант высказывания “Если у меня будет свободное время и я сдам экзамены по математике и физике, то поеду отдыхать в Крым или на Кавказ”, записанного с использованием логики предикатов первого порядка:

- 1) $(a \vee b \vee c) \leftrightarrow (d \vee e)$.
- 2) $(a \wedge b \wedge c) \rightarrow (d \wedge e)$.
- 3) $(a \vee b \wedge c) \leftrightarrow (d \vee e)$.
- 4) $(a \wedge b \wedge c) \rightarrow (d \vee e)$.

18. Укажите правильный вариант высказывания “Некоторые рабы являются китами” записанного с использованием логики предикатов первого порядка:

- 1) $(\forall X)(S(X) \vee P(X))$
- 2) $(\forall X)(S(X) \wedge P(X))$
- 3) $(\exists X)(S(X) \wedge P(X))$
- 4) $(\exists X)(S(X) \vee P(X))$

19. Какой из компонентов не является основным в продукционной системе:

- 1) набор правил;
- 2) рабочая память;
- 3) логическая единица;
- 4) механизм логического вывода.

20. Структура для представления знаний в виде узлов, соединенных дугами, называется:

- 1) правилом продукции;
- 2) логикой предикатов первого порядка;
- 3) фреймом;
- 4) семантической сетью.

АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Тестовые задания по проверке остаточных знаний по дисциплине
«Базы знаний и экспертные системы»
для специальностей 230201 – «Информационные системы и технологии»

Утверждено на заседании кафедры
«23» октября 2006 г.
«УТВЕРЖДАЮ» _____

Кафедра ИУС
Заведующий кафедрой
А.В.Бушманов

Инструкция: Тест состоит из 20 заданий, время тестирования – 40 минут.

ВАРИАНТ № 3

1. Экспертная система отличается от прочих прикладных программ наличием следующего признака:

- 1) обладает способностью объяснить, почему предложено именно такое решение, и доказать его обоснованность;
- 2) обладает знаниями, а не способностью просто выполнять некоторый алгоритм;
- 3) процесс работы основан на применении правил отношений к символическому представлению знаний;
- 4) «общается» только со своим создателем, который и так знает, на чем основывается ее результат.

2. Структура в экспертной системе, предназначенная для хранения исходных и промежуточных данных решаемой в текущий момент задачи, называется:

- 1) информационно-поисковой единицей;
- 2) базой данных;
- 3) базой знаний;
- 4) системой управления базами данных.

3. Экспертные системы, которые обладают способностью при возникновении неизвестной ситуации определять с помощью некоторых общих принципов последовательность действий, которые необходимо выполнять, называются:

- 1) глубинными;
- 2) традиционными;
- 3) гибридными;
- 4) поверхностными.

4. ... следует рассматривать как некую добродетель системы, поскольку большинство приложений, построенных на базе экспертных систем, нуждается в накоплении большого объема знаний, а решить такую задачу довольно трудно, если соглашения в языке представлены слишком сложны:

- 1) побочный эффект;
- 2) естественность нотации;
- 3) логическую адекватность;
- 4) эвристическую мощьность.

5. Алгоритм, который может быстрее найти решение, особенно, если при его выполнении используются эвристики для выбора очередной ветви, называется:

- 1) пространством решений;
- 2) алгоритмом поиска в ширину;
- 3) алгоритмом поиска в глубину;
- 3) комбинаторным взрывом.

6. Экспертная система, которая надежно решает все задачи, но требует много времени и (или) памяти, называется:

- 1) демонстрационным прототипом;
- 2) исследовательским прототипом;
- 3) действующим прототипом;
- 4) коммерческой системой.

7. Экспертная система, которая реализует процесс соотнесения объекта с некоторым классом объектов и (или) обнаруживает неисправности в некоторой системе, классифицируется как:

- 1) прогнозирующая система;
- 2) диагностирующая система;
- 3) проектирующая система;
- 4) планирующая система.

8. ... предназначены для обнаружения источников неисправностей по результатам наблюдений за поведением контролируемой системы:

- 1) интерпретирующие системы;
- 2) прогнозирующие системы;
- 3) диагностические системы;
- 4) системы проектирования.

9. Средства этой категории включают несколько программных модулей, что позволяет пользователю комбинировать в процессе разработки экспертной системы разные стили программирования, называются:

- 1) средой программирования, поддерживающей несколько парадигм;
- 2) оболочкой экспертной системы;
- 3) дополнительным модулем;
- 4) языком программирования высокого уровня.

10. Этап, на котором эксперт и инженер по знаниям эксплицируют основные понятия, отношения и характеристики, необходимые для описания процесса решения задачи, называется:

- 1) этапом идентификации;
- 2) этапом концептуализации;
- 3) этапом формализации;
- 4) этапом тестирования.

11. Возврат от этапа тестирования на этап формализации приводит к пересмотру выбранного ранее способа представления знаний. Данный цикл называют:

- 1) переформулированием понятий и требований;
- 2) переконструированием представлений;
- 3) усовершенствованием прототипа;
- 4) структурированием знаний.

12. Чтобы быть полноправным участником общения система речевого общения должна выполнять некоторые обязательные функции. Одной из таких функций является:

- 1) локализация определенных объектов в поле изображения и измерение заданных параметров этих объектов;
- 2) преобразование поступающих от пользователя высказываний на естественном языке в высказывания на языке внутреннего представления;
- 3) последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке;
- 4) сжатие информационного потока за счет введения операции распознавания смысловых элементов речи.

13. Процесс представляет собой последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке, который должен максимально воссоздавать смысл и, как правило, структуру исходного текста, но уже средствами выходного языка и относится к:

- 1) естественно-языковым системам;
- 2) системам распознавания речи;
- 3) системам переработки визуальной информации;
- 4) системам машинного перевода.

14. Системы обработки и анализа видеоданных, которые обычно работают со входной информацией в виде полутоновых или многоспектральных изображений объединяют под названием:

- 1) интерактивная графика;
- 2) автоматизированные системы обработки изображений;
- 3) системы обработки графической информации;
- 4) системы технического зрения для роботов.

15. Лингвистическая и информационная теория образования и восприятия речи являются основой для:

- 1) ЕЯ-систем;
- 2) систем распознавания речи;
- 3) систем переработки визуальной информации;

4) систем машинного перевода.

16. Задавая фиксированное множество правил вывода, можно рассматривать следующее семейство проблем: исходя из выбранного множества ППФ применением некоторого числа раз правил вывода, можно получить заранее заданную ППФ. Процедура, которая из одной или нескольких ППФ производит другие ППФ называется:

- | | |
|---------------------|---------------------|
| 1) доказательством; | 2) преобразованием; |
| 3) упрощением; | 4) правилом вывода. |

17. Укажите правильный вариант высказывания “Если у меня будет свободное время, то я почитаю книгу или посмотрю телевизор” записанного с использованием логики предикатов первого порядка:

- | | |
|-------------------------------------|---------------------------------------|
| 1) $a \leftrightarrow (b \vee c)$. | 2) $a \rightarrow (b \vee c)$. |
| 3) $a \rightarrow (b \wedge c)$. | 4) $a \leftrightarrow (b \wedge c)$. |

18. Укажите правильный вариант высказывания “Некоторые спортсмены являются мастерами спорта” записанного с использованием логики предикатов первого порядка:

- | | |
|------------------------------------|------------------------------------|
| 1) $(\forall X)(S(X) \vee P(X))$ | 2) $(\forall X)(S(X) \wedge P(X))$ |
| 3) $(\exists X)(S(X) \wedge P(X))$ | 4) $(\exists X)(S(X) \vee P(X))$ |

19. Детерминированная продукция в правой части ядра которой указываются возможности выбора, оценивающиеся специальными весами выбора (в качестве таких весов могут использоваться вероятностные оценки, лингвистические оценки, экспертные оценки и т.п.), называется:

- | | |
|--------------------|-----------------------|
| 1) однозначной; | 2) альтернативной; |
| 3) не однозначной; | 4) не альтернативной. |

20. В семантических сетях используются четыре основных типа объектов. Один из них, определенный как “сведения об абстрактных или физических объектах предметной области и задается множеством доменов (параметров или констант)”, называется:

- | | |
|--------------|--------------|
| 1) понятия; | 2) события; |
| 3) свойства; | 4) значения. |

АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Тестовые задания по проверке остаточных знаний по дисциплине
«Базы знаний и экспертные системы»
для специальностей 230201 – «Информационные системы и технологии»

Утверждено на заседании кафедры
«23» октября 2006 г.
«УТВЕРЖДАЮ» _____

Кафедра ИУС
Заведующий кафедрой
А.В.Бушманов

Инструкция: Тест состоит из 20 заданий, время тестирования – 40 минут.

ВАРИАНТ № 4

1. Экспертная система отличается от прочих прикладных программ наличием следующего признака:

- 1) демонстрирует свои знания, которые сконцентрированы на определенную предметную область;

- 2) обладает знаниями, а не способностью просто выполнять некоторый алгоритм;
- 3) берет на себя функции, выполнение которых обычно требует привлечение опыта человека-специалиста или играет роль ассистента для человека, принимающего решение
- 4) при решении задач основными являются эвристические и приближенные методы, которые не всегда гарантируют успех.

2. Автоматизирует процесс наполнения экспертной системы знаниями:

- 1) решатель;
- 2) диалоговый компонент;
- 3) компонент приобретения знаний;
- 4) компонент обработки знаний.

3. Экспертные системы, которые используют и методы инженерии знаний, и формализованные методы, и данные традиционного программирования и математики, называются:

- 1) глубинными;
- 2) традиционными;
- 3) гибридными;
- 4) поверхностными.

4. ... означает, что представление знаний должно обладать способностью распознавать все отличия, которые Вы закладываете в исходную сущность:

- 1) естественность нотации;
- 2) логическая адекватность;
- 3) побочный эффект;
- 4) эвристическая мощьность.

5. Множество решений, которые удовлетворяют условию проверять, не является ли образовавшееся состояние конечным решением, иногда называют:

- 1) пространством решений;
- 2) алгоритмом поиска в ширину;
- 3) алгоритмом поиска в глубину;
- 3) комбинаторным взрывом.

6. Инструментальные средства, которые созданы для решения узких специфических задач, работают медленно и неэффективно, называются:

- 1) исследовательскими инструментальными средствами;
- 2) экспериментальными инструментальными средствами;
- 3) коммерческими инструментальными средствами;
- 4) настраиваемыми оболочками.

7. Экспертная система, в которой используются модели поведения реальных объектов с тем, чтобы логически вывести последствия планируемой деятельности, классифицируется как:

- 1) управляющая система;
- 2) прогнозирующая система;
- 3) проектирующая система;
- 4) планирующая система.

8. ... предназначены для структурного синтеза конфигурации объектов при заданных ограничениях:

- 1) интерпретирующие системы;
- 2) прогнозирующие системы;
- 3) диагностические системы;
- 4) системы проектирования.

9. Средства этой категории представляют собой автономные программные модули, предназначенные для выполнения специфических задач в рамках выбранной архитектуры системы решения проблем, называются:

- 1) средой программирования, поддерживающей несколько парадигм;
- 2) оболочкой экспертной системы;
- 3) дополнительным модулем;
- 4) языком программирования высокого уровня.

10. Этап, на котором экспертная система занимается решением всех возможных задач при работе с различными пользователями, называется:

- 1) этапом выполнения;
- 2) этапом тестирования;
- 3) этапом формализации;
- 4) этапом опытной эксплуатации.

11 Если возникшие проблемы серьезны, то после неудачи на этапе тестирования может потребоваться возврат на этапы концептуализации и идентификации. В этом случае речь идет о ... , используемых в системе:

- 1) переформулировании понятий и требований;
- 2) переконструировании представлений;
- 3) усовершенствовании прототипа;
- 4) структурировании знаний.

12. Чтобы быть полноправным участником общения система машинного перевода должна выполнять некоторые обязательные функции. Одной из таких функций является:

- 1) локализация определенных объектов в поле изображения и измерение заданных параметров этих объектов;
- 2) преобразование поступающих от пользователя высказываний на естественном языке в высказывания на языке внутреннего представления;
- 3) последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке;
- 4) сжатие информационного потока за счет введения операции распознавания смысловых элементов речи.

13. Процесс представляет собой последовательность преобразований, применяемых к входному тексту и превращающих его в текст на выходном языке, который должен максимально воссоздавать смысл и, как правило, структуру исходного текста, но уже средствами выходного языка и относится к:

- 1) естественно-языковым системам;
- 2) системам распознавания речи;
- 3) системам переработки визуальной информации;
- 4) системам машинного перевода.

14. Системы обработки и анализа видеоданных, которые обычно работают со входной информацией в виде полутоновых или многоспектральных изображений объединяют под названием:

- 1) интерактивная графика;
- 2) автоматизированные системы обработки изображений;
- 3) системы обработки графической информации;
- 4) системы технического зрения для роботов.

15. Диалоговый компонент, компонент понимания высказываний, компонент генерации высказываний - составляющие

- 1) ЕЯ-систем;
- 2) систем распознавания речи;
- 3) систем переработки визуальной информации;
- 4) систем машинного перевода.

16. Если некоторая последовательность резолюций, применяемых к исходному множеству предложений E и множеству резольвент, полученных в процессе резолюции, приводит к пустому предложению, то множество E является:

- 1) выполнимым;
- 2) невыполнимым;
- 3) пустым;
- 4) пополняемым.

8.	1	2	3	4
9.	2	4	1	3
10.	3	1	2	4
11.	4	3	2	1
12.	1	2	4	3
13.	4	4	4	4
14.	2	2	2	2
15.	4	2	2	1
16.	2	1	4	2
17.	4	4	2	2
18.	3	3	3	2
19.	3	3	2	1
20.	2	4	1	4

9. КОНТРОЛЬНЫЕ ВОПРОСЫ К ЗАЧЕТУ

1. Назначение и особенности методов ИИ для разработки ЭС
2. Структура и режимы экспертных систем
3. Характеристики ЭС
4. Классификация ЭС
5. Классификация инструментальных средств ЭС
6. Методология разработки ЭС
7. Этапы разработки ЭС
8. Этап идентификации
9. Этап концептуализации
10. Этап формализации
11. Этап выполнения
12. Этап тестирования
13. Этап опытной эксплуатации
14. Взаимодействия инженера по знаниям с экспертом
15. Трудности разработки ЭС
16. Проблемы и перспективы ЭС
17. Общие сведения об архитектуре ЭС
18. Программы поиска для ЭС
19. Управление с помощью эвристик
20. Ориентированный на человека диалог

21. Архитектура для автоматического рассуждения, основанная на правилах
22. Автоматическое рассуждение
23. Особенности реализации ЕЯ – систем. Роль ИИ в данных системах
24. Методы реализации ЕЯ – систем
25. Анализаторы, используемые в ЕЯ - системах
26. Основные положения систем речевого общения. Роль ИИ в данных системах
27. Принципы построения систем речевого общения
28. анализаторы речевых сообщений
29. синтезаторы речи
30. Фонетическая и просодическая структуры речи
31. Информационная структура речевого сигнала
32. Назначение, классификация и область применения систем переработки визуальной информации. Роль ИИ в данных системах
33. Автоматизированные системы обработки изображений
34. Системы анализа изображений
35. Системы машинной графики
36. Назначение систем машинного перевода. Роль ИИ в данных системах
37. Лингвистическое обеспечение систем машинного перевода
38. Грамматики и алгоритмы систем машинного перевода
39. Фильтровой и эвристический методы в системах машинного перевода
40. Математическое и программное обеспечение систем машинного перевода
41. Перспективы систем машинного перевода

10. КАРТА КАДРОВОЙ ОБЕСПЕЧЕННОСТИ ДИСЦИПЛИНЫ

Лектор – доцент Акилова Ирина Михайловна.

Руководитель практических работ – доцент Акилова Ирина Михайловна.

Руководитель лабораторных работ – доцент Акилова Ирина Михайловна.

СОДЕРЖАНИЕ

1. Рабочие программы

3

2. График самостоятельной работы студентов	14
3. Методические рекомендации по проведению самостоятельной работы студентов	14
4. Перечень учебников, учебных пособий	15
5. Краткий конспект лекций	17
6. Методические указания по выполнению лабораторных работ	111
7. Методические указания по организации межсессионного контроля знаний студентов	132
8. Фонд контрольных заданий для оценки качества знаний по дисциплине	133
9. Контрольные вопросы к зачету	145
10. Карта кадровой обеспеченности дисциплины	147

Ирина Михайловна Акилова,
доцент кафедры ИиУС АмГУ

Учебно-методический комплекс по дисциплине «Базы знаний и экспертные системы»

Изд-во АмГУ. Подписано к печати
Тираж Заказ

Формат 60x84/16. Усл. печ. л .