

**Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Амурский государственный университет»**

Кафедра Информационных и управляющих систем

**УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ
«БАЗЫ ЗНАНИЙ И ЭКСПЕРТНЫЕ СИСТЕМЫ В ИС»**

Основной образовательной программы по специальности 230201.65 – Информационные
системы и технологии

Благовещенск 2012

УМКД разработан доцентом Акиловой Ириной Михайловной

Рассмотрен и рекомендован на заседании кафедры

Протокол заседания кафедры от « _____ » _____ 2012 г. № _____

Зав. кафедрой _____ / А.В.Бушманов

УТВЕРЖДЕН

Протокол заседания УМСС специальности 230201.65 – Информационные системы и технологии

от « _____ » _____ 2012 г. № _____

Председатель УМСС _____ / В.В.Еремина /

СОДЕРЖАНИЕ

Рабочая программа учебной дисциплины	4
Краткое изложение программного материала	11
Методические указания по выполнению лабораторных работ	40
Контроль знаний	44

1. РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Амурский государственный университет»

УТВЕРЖДАЮ
Проректор по учебной работе

_____ В.В. Проказин

«__» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА

БАЗЫ ЗНАНИЙ И ЭКСПЕРТНЫЕ СИСТЕМЫ В ИС

для специальности 230201.65 «Информационные системы и технологии»

Квалификация (степень) выпускника – инженер

Курс – 3

Семестр – 6

Лекции – 36 (час.)

Зачет – 6

Лабораторные занятия – 18 (час.)

Самостоятельная работа – 54 (час.)

Общая трудоемкость дисциплины – 108 (час.)

Составитель – И.М.Акилова, доцент

Факультет математики и информатики

Кафедра информационных и управляющих систем

2012 г.

Рабочая программа составлена на основании Государственного образовательного стандарта высшего профессионального образования по специальности 230201.65 «Информационные системы и технологии»

Рабочая программа обсуждена на заседании кафедры информационных и управляющих систем

«__» _____ 20__ г., протокол № _____

Заведующий кафедрой _____ А.В. Бушманов

Рабочая программа одобрена на заседании учебно-методического совета специальности 230201.65 «Информационные системы и технологии»

«__» _____ 20__ г., протокол № _____

Председатель _____ В.В. Еремина

Рабочая программа переутверждена на заседании кафедры информационных и управляющих систем

«__» _____ 20__ г., протокол № _____

Заведующий кафедрой _____ А.В. Бушманов

СОГЛАСОВАНО
Учебно-методическое
управление

«__» _____ 20__ г.

СОГЛАСОВАНО
Председатель учебно-методического
совета факультета

_____ С.Г. Самохвалова
«__» _____ 20__ г.

СОГЛАСОВАНО
Заведующий выпускающей кафедрой

_____ А.В. Бушманов
«__» _____ 20__ г.

СОГЛАСОВАНО
Директор научной библиотеки

_____ Л.А. Проказина
«__» _____ 20__ г.

1 ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель дисциплины: формирование профессиональных компетенций в области проектирования и разработки экспертных систем; получение теоретических знаний и практического опыта по использованию экспертных систем и инженерии баз знаний.

Задачи дисциплины:

- развитие логического мышления, воспитание высокой математической культуры;
- формирование личности студента, развитие его интеллекта.
- освоение обучаемыми математическими методами и основами логического моделирования;
- на примерах математических понятий и методов продемонстрировать студентам сущность научного подхода, специфику искусственного интеллекта в области создания экспертных систем и его роль в прикладных исследованиях

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", "Технология программирования", "Организация баз данных".

Курс базируется на знании основных языков логического программирования, элементов операционных систем, математического моделирования.

В результате изучения дисциплины студент должен:

- 1) знать принципы построения экспертных систем, формирование баз знаний для экспертных систем;
- 2) уметь создавать фрагменты экспертных систем, экспертные системы, базы знаний для экспертных систем.

2 МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВПО

Программа курса «Базы знаний и экспертные системы» составлена в соответствии с требованиями государственного образовательного стандарта специальности 230201.65 – Информационные системы и технологии, блок дисциплины специализации ДС.04.

3 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 108 часов.

№ п/п	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость в часах				Формы текущего контроля успеваемости Форма промежуточной аттестации
				Лек	Пр	Лаб	Сам	
1	Назначение и особенности методов искусственного интеллекта для разработки экспертных систем. Определение экспертной системы.	6	1-2	4		2	6	Защита лаб. работы
2	Структура экспертных систем и основные понятия.	6	3-4	4		2	6	Защита лаб. работы

	Общие сведения об архитектуре экспертных систем.							
3	Классификация экспертных систем. Классификация инструментальных средств. Методология разработки экспертных систем.	6	5-6	4		2	6	Защита лаб. работы
4	Этапы разработки экспертных систем. Основные направления в изучении и практическом использовании.	6	7-9	6		2	6	Защита лаб. работы
5	Обобщенная схема ЕЯ - систем. Методы реализации ЕЯ - систем. Основные классы ЕЯ - систем. Системы общения с базами данных. Обзор промышленных ЕЯ - систем.	6	10-12	6		2	8	Защита лаб. работы
6	Основные положения систем речевого общения.	6	13-15	6		2	6	Защита лаб. работы
7	Назначение, классификация и область применения систем обработки визуальной информации.	6	16-17	4		3	8	Защита лаб. работы
8	Назначение систем машинного перевода.	6	18	2		3	8	Защита лаб. работы
9	Всего по разделам	6	1-18	36		18	54	Зачет

4 СОДЕРЖАНИЕ РАЗДЕЛОВ И ТЕМ ДИСЦИПЛИНЫ

4.1 Лекции

1. ОСОБЕННОСТИ РАЗРАБОТКИ И ИСПОЛЬЗОВАНИЯ ЭКСПЕРТНЫХ СИСТЕМ

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

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

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

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

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.

4.2 Лабораторные занятия

4.2.1 Лабораторная работа 1. Разработка экспертных систем, базирующихся на правилах

4.2.2 Лабораторная работа 2. Разработка экспертных систем, базирующихся на логике

4.2.3 Лабораторная работа 3. Создания интерфейса общения пользователя с экспертной системой

5 САМОСТОЯТЕЛЬНАЯ РАБОТА

№ п/п	Раздел дисциплины	Форма (вид) самостоятельной работы	Трудоемкость в часах
1	Разработка экспертных систем, базирующихся на правилах	Оформление отчета.	27
2	Разработка экспертных систем, базирующихся на логике	Оформление отчета.	27
3	Всего		54

6 ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Образовательный процесс по дисциплине строится на основе комбинации следующих образовательных технологий.

Интегральную модель образовательного процесса по дисциплине формируют технологии методологического уровня: модульно-рейтинговое обучение, технология поэтапного формирования умственных действий, технология развивающего обучения, элементы технологии развития критического мышления.

Реализация данной модели предполагает использование следующих технологий стратегического уровня (задающих организационные формы взаимодействия субъектов образовательного процесса), осуществляемых с использованием определенных тактических процедур:

– лекционные (вводная лекция, информационная лекция, обзорная лекция, лекция-консультация, проблемная лекция);

– практические (углубление знаний, полученных на теоретических занятиях, решение задач);

– тренинговые (формирование определенных умений и навыков, формирование алгоритмического мышления);

– активизации познавательной деятельности (приемы технологии развития критического мышления через чтение и письмо, работа с литературой, подготовка презентаций по темам домашних работ);

– самоуправления (самостоятельная работа студентов, самостоятельное изучение материала).

Рекомендуется использование информационных технологий при организации коммуникации со студентами для представления информации, выдачи рекомендаций и

консультирования по оперативным вопросам (электронная почта), использование мультимедиа-средств при проведении лекционных и практических занятий.

7 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

7.1 Оценочные средства для текущего контроля успеваемости

7.1.1 Контрольные вопросы допуска к выполнению практических работ

7.1.2 Отчеты о выполнении индивидуальных вариантов заданий практических работ

7.2 Оценочные средства для промежуточной аттестации

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

7.3 Учебно-методическое обеспечение самостоятельной работы

7.3.1 Учебное пособие по языку программирования Пролог с заданиями, контрольными вопросами и методическими указаниями по выполнению лабораторных работ

8 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. Матвеев М.Г. Модели и методы искусственного интеллекта: учеб. пособие/ М.Г.Матвеев, А.С.Свиридов, Н.А.Алейникова. – М.:Финансы и статистика: ИНФРА-М, 2008. – 448 с.
2. Рыбина Г.В. Основы построения интеллектуальных систем: учеб. пособие / Г.В.Рыбина. – М.:Финансы и статистика: ИНФРА-М, 2010. – 432 с.
3. Сергиевский Г.М., Волченков Н.Г. Функциональное и логическое программирование: учеб. пособие: М.: Академия, 2010. – 320 с.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. Голицына О.Л. Информационные системы : учеб. пособие : рек. УМО/ О.Л. Голицына, Н.В. Максимов, И.И. Попов. -М.: ФОРУМ: ИНФРА-М, 2009. -496 с.
2. Зубов А.В. Основы искусственного интеллекта для лингвистов : учеб. пособие: рек. УМО/ А. В. Зубов, И. И. Зубова. -М.: Логос, 2007. -320 с.
3. Новак, В. Математические принципы нечеткой логики / В. Новак, И. Перфильева, И. Мочкорж ; пер. с англ., под ред. А. Н. Аверкина. - М. : Физматлит, 2006. - 348 с.
4. Андрейчиков А.В. Интеллектуальные информационные системы : учебник: Рек. Мин. обр. РФ/ А. В. Андрейчиков, О. Н. Андрейчикова. -М.: Финансы и статистика, 2004. - 424 с.
5. Гаврилова, Т.А. Базы знаний интеллектуальных систем : Учеб.: Доп. Мин. обр. РФ / Т.А. Гаврилова, В.Ф. Хорошевский . - СПб. : Питер, 2000. - 382 с.
6. Акилова И.М. Методы представления и обработки знаний: Практикум/И.М.Акилова, Н.В.Назаренко. Благовещенск: Изд-во Амур. гос. Ун-та, 2002. 52с.
7. Частиков А.П. Разработка экспертных систем. Среда CLIPS : [учеб. пособие]/ А. П. Частиков, Т. А. Гаврилова, Д. Л. Белов. -СПб.: БХВ-Петербург, 2003. -606 с.
8. Статические и динамические экспертные системы: учеб. пособие: рек. Мин. обр. РФ / Э.П. Попов [и др.]. – М.:Финансы и статистика, 1996. -320 с.
9. Джексон П.Введение в экспертные системы: Учеб. пособие/П.Джексон. – 3-е изд. – М:Вильямс, 2001. -623 с.

СПРАВОЧНЫЕ ИЗДАНИЯ

1. Искусственный интеллект: справ.: в 3 кн./под ред. Э. В. Попова. – М.: Радио и связь, 1990.

ПЕРИОДИЧЕСКИЕ ИЗДАНИЯ:

1. «PC magazine Персональный компьютер сегодня»
2. «Программные продукты и системы»
3. «Информационные технологии и вычислительные системы»

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ИНТЕРНЕТ-РЕСУРСЫ

1. Свободно распространяемая версия языка программирования Пролог.

№ п/п	Наименование ресурса	Характеристика
1	http://www.iqlib.ru	Интернет библиотека образовательных изданий, В которой собраны электронные учебники, справочные и учебные пособия. Удобный поиск по ключевым словам, отдельным темам и отраслям

		знаний.
2	http://www.intuit.ru	Интернет-университет информационных технологий. В котором собраны электронные и видео-курсы по отраслям знаний
3	http://amursu.ru	Сайт АмГУ, Библиотека – электронная библиотека АмГУ

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

9.1 Лекционная аудитория

9.2 Лаборатории, оборудованные рабочими местами пользователей ЭВМ

10 РЕЙТИНГОВАЯ ОЦЕНКА ЗНАНИЙ СТУДЕНТОВ ПО ДИСЦИПЛИНЕ

В связи с подготовкой специалистов по настоящей учебной программе данный раздел не разрабатывался.

2. КРАТКОЕ ИЗЛОЖЕНИЕ ПРОГРАММНОГО МАТЕРИАЛА

Лекция №1 Тема: Назначение и особенности методов искусственного интеллекта для разработки экспертных систем. Определение экспертной системы.

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.

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.

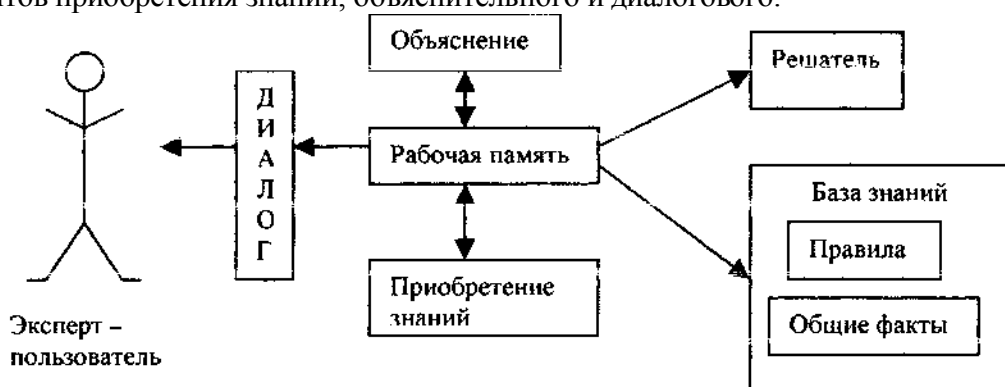
♦ Экспертная система должна обладать способностью *объяснить*, почему предложено именно такое решение, и *доказать его обоснованность*. Пользователь должен получить всю информацию, необходимую ему для того, чтобы быть уверенным, что решение принято "не с потолка". В отличие от этого, исследовательские 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 отношений 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".

Лекция № 2 Тема: Структура экспертных систем и основные понятия. Общие сведения об архитектуре 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.

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.

Лекция № 3 Тема: Классификация экспертных систем. Классификация инструментальных средств. Методология разработки 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 возможных 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

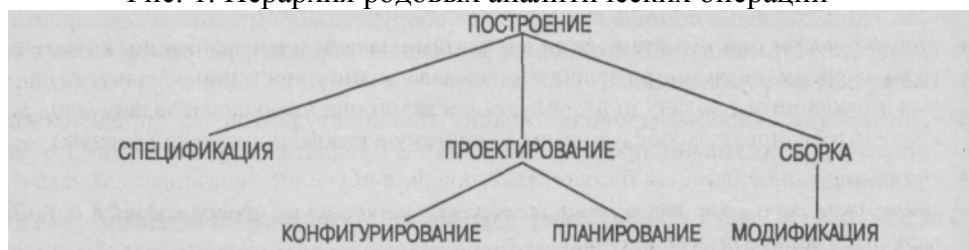


Рис. 2. Иерархия родовых синтетических операций

Лекция № 4 Тема: Этапы разработки экспертных систем. Основные направления в изучении и практическом использовании.

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. Определяются 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.

Рассмотрим 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.

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 № 5 Тема: Обобщенная схема ЕЯ-систем. Методы реализации ЕЯ-систем. Основные классы ЕЯ-систем. Системы общения с базами данных. Обзор промышленных 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 неприемлемость 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 несоответствующих представлений, а так же 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 знания; 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 № 6 ТЕМА: Основные положения 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, соотносимого 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 телефону 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 колебаний 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 полных 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 серединные, 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 серединной.

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.

Принципы построения систем речевого общения

Классификация речевых процессоров

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

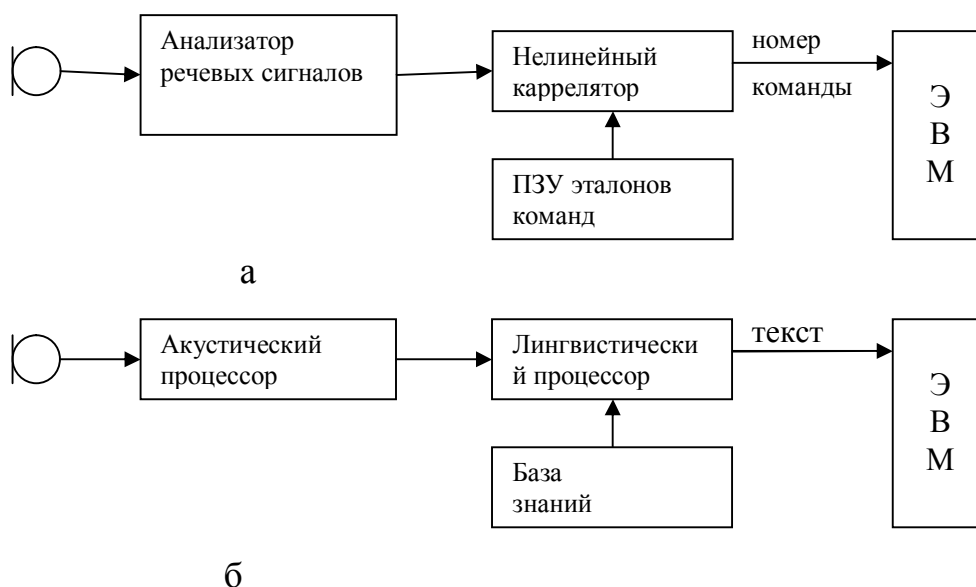
Анализаторы. Эти устройства предназначены для преобразования речевых сигналов с микрофона (информационный поток сигналов 10^5 бит/с) в последовательность цифровых кодов с существенно меньшим информационным потоком (10^4 — 10^1 бит/с) и с обязательным сохранением передачи смыслового компонента речи. Анализаторы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

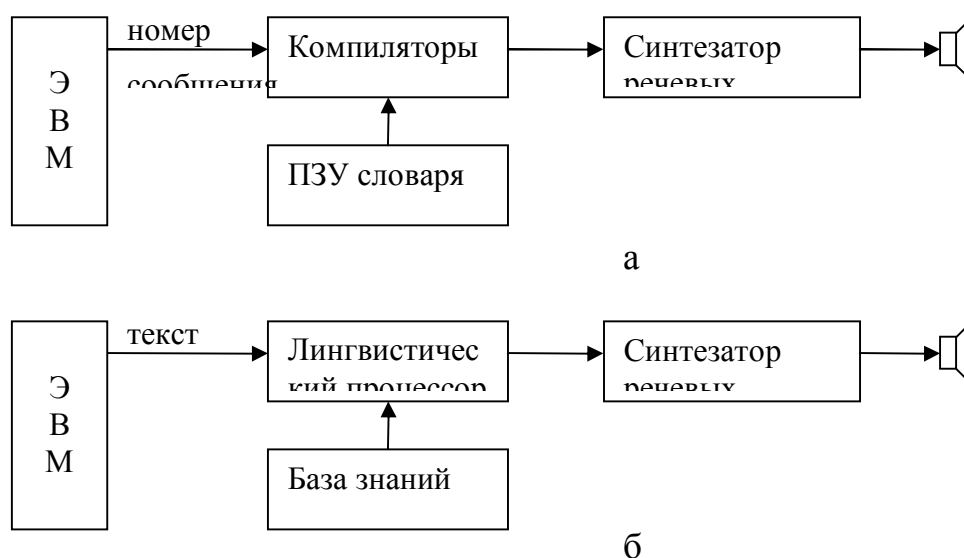
Синтезаторы речи. Эти устройства предназначены для преобразования кодовой последовательности, поступающей от ЭВМ, ПЗУ или линии связи, в непрерывный речевой сигнал. Синтезаторы подразделяются на классы и группы по тем же признакам, что и анализаторы речи. Классам анализаторов речевых сигналов и сообщений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

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

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 считываются 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.

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



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 применение 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 № 7 Тема: Назначение, классификация и область применения систем обработки 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 направлении 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 геометрического 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.

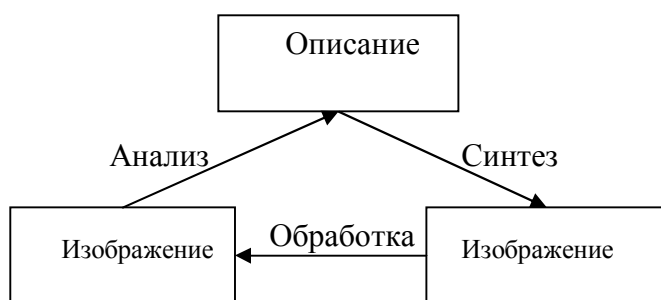


Рис. 1. Взаимосвязь направлений цифровой обработки изображений

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, ультразвукологии, инфраскопии (обработка двух- и трехмерных изображений 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 разнородных 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.

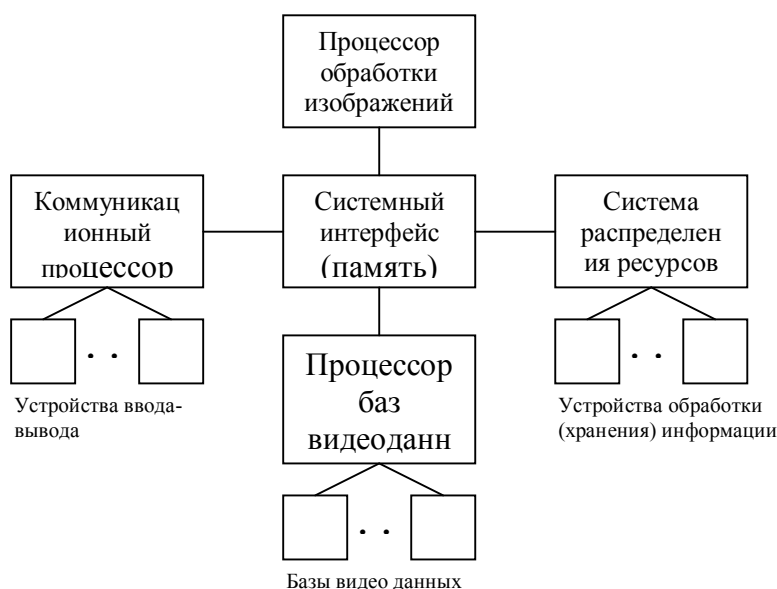


Рис.2 Структура интегрированной АСОИЗ

Работа по анализу изображения становится своего рода многократным экспериментированием с исходными данными. Характерной чертой такого эксперимента 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 содержимому).

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 изображения), 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 (эталонные, 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.

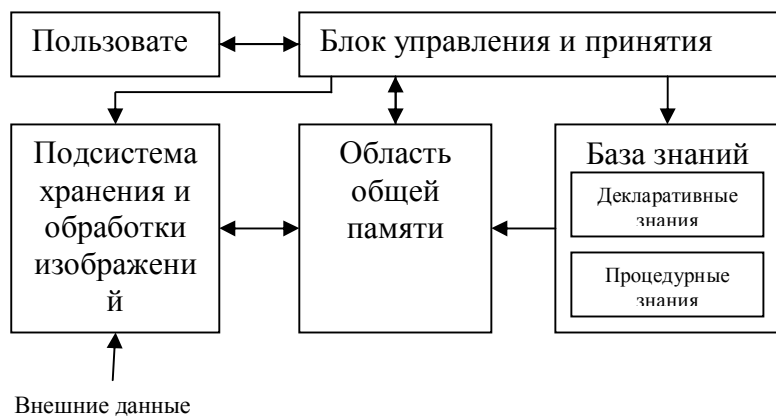


Рис. 3. Структура экспертной системы анализа изображений

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 (перепадов 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

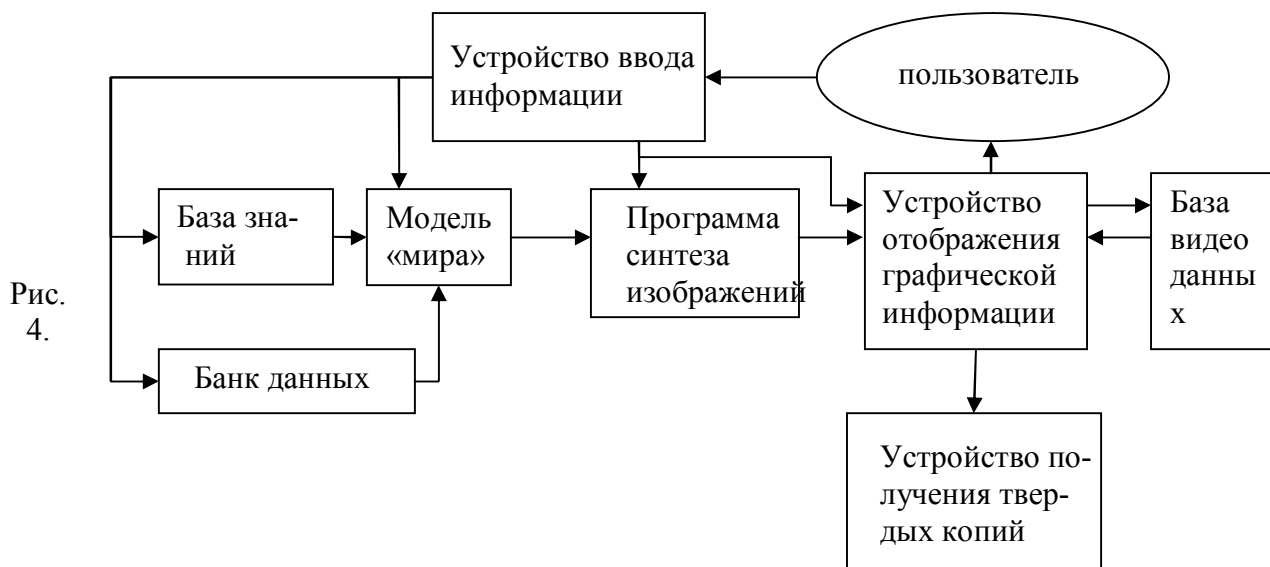
Основными областями применения систем машинной графики являются:

- 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 современных СМГ.



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 двухмерная 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.

Лекция № 8 Тема: Назначение систем 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.

2. 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.

3. 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 единый стиль 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 относятся 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:

- 1) словари;
- 2) грамматики;
- 3) 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 «собирания» 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, 1. 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 $A \rightarrow B...C$. 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 НС. 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 КС-правил. В общем виде это значит, что

правила продукций, переписываются так: $A [a] \rightarrow V[B]$, ... , $C [c]$, где малыми буквами обозначены условия, тесты, инструкции и т. д., расширяющие исходные жесткие правила и дающие грамматике гибкость и эффективность.

В грамматике обобщенных составляющих—ТСС введены метаправила, являющиеся обобщением закономерностей правил КС1.

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

Разновидностью РСПГ являются каскадные РСПГ. Каскад—это РСП, 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. РСП 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 КСГ, проверочные 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 практическими 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.

3. 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 - "Стоянка 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.

4. КОНТРОЛЬ ЗНАНИЙ 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