

Тема 5. Экспертные системы

На начальных этапах развития искусственный интеллект подвергался жесткой критике и одним из доводов был тот, что практической пользы от него нет, он занимается игрушками. Экспертные системы одними из первых доказали практическую пользу этого научного направления, принося в начале 80-х годов XX века коммерческую прибыль своим создателям.

Термин "системы, основанные на знаниях" (knowledge-based systems) появился в 1976 году одновременно с первыми системами, аккумулирующими опыт и знания экспертов.

Экспертные системы - это прикладные системы искусственного интеллекта, в которых база знаний представляет собой формализованные эмпирические знания высококвалифицированных специалистов (экспертов) в какой-либо узкой предметной области, они аккумулируют эти знания и тиражируют их для консультации менее квалифицированных специалистов.

В 1965 году Э. Фейгенбаум (бывший студент Герберта Саймона), Б. Бьюкенен (философ по образованию) и Д. Ледерберг (лауреат Нобелевской премии в области генетики) начали работы над первой экспертной системой DENDRAL. В 1969 году была создана математическая экспертная система MACSYMA В. Мартином и Д. Мозесом. Первая экспертная система для медицинской диагностики была создана в 1973 году Э. Шортлиффом и называлась MYCIN, она повлекла за собой разработку первого командного интерпретатора EMYCIN (В. Милле, Шортлифф и Бьюкенен). В 1976 году Дуда и Харт начали работу над экспертной системой PROSPECTOR, предназначенной для разведки полезных ископаемых, в 1984 году система точно предсказала существование месторождения молибдена, оцененного в многомиллионную сумму.

Эти экспертные системы, разработанные в 60-70-х годах стали классическими. По происхождению, предметным областям и по преемственности применяемых идей, методов и инструментальных программных средств их можно разделить на несколько семейств.

1. **МЕТА-DENDRAL.** Система DENDRAL позволяет определить наиболее вероятную структуру химического соединения по экспериментальным данным (масс-спектрографии, данным ядерном магнитного резонанса и др.). Meta-DENDRAL автоматизирует процесс приобретения знаний для DENDRAL. Она генерирует правила построения фрагментов химических структур.

2. **MYCIN-EMYCIN-TEIREIAS-PUFF-NEOMYCIN.** Это семейство медицинских ЭС и сервисных программных средств для их построения.
3. **PROSPECTOR-KAS.** PROSPECTOR - предназначена для поиска (предсказания) месторождений на основе геологических анализов. KAS - система приобретения знаний для PROSPECTOR.
4. **CASNET-EXPERT.** Система CASNET - медицинская ЭС для диагностики выдачи рекомендаций по лечению глазных заболеваний. На ее основе разработан язык инженерии знаний EXPERT, с помощью которой создан ряд других медицинских диагностических систем.
5. **HEARSAY-HEARSAY-2-HEARSAY-3-AGE.** Первые две системы этого ряда являются развитием интеллектуальной системы распознавания слитной человеческой речи, слова которой берутся из заданного словаря. Эти системы отличаются оригинальной структурой, основанной на использовании доски объявлений- глобальной базы данных, содержащей текущие результаты работы системы. В дальнейшем на основе этих систем были созданы инструментальные системы HEARSAY-3 и AGE (Attempt to Generalize-попытка общения) для построения экспертных систем.
6. Системы **AM** (Artificial Mathematician - искусственный математик) и **EURISCO** были разработаны в Станфордском университете доктором Дугласом Ленатом для исследовательских и учебных целей. В систему AM первоначально было заложено около 100 правил вывода и более 200 эвристических алгоритмов обучения, позволяющих строить произвольные математические теории и представления. EURISCO – это развитие системы AM, с ее помощью в военно-стратегической игре, проводимой ВМФ США, была разработана стратегия, содержащая ряд оригинальных тактических ходов.

Кроме разработки самих экспертных систем исследователи занялись созданием инструментальное средство для экспертных систем: в 1983 году компания IntelliCorp создала KEE, а в 1985 году агентство NASA выпустило первую версию CLIPS.

Экспертные системы быстро завоевали позиции на информационном рынке и получили широкое распространение. Уже в 1987 году опрос пользователей, проведенный журналом "Intelligent Technologies" (США), показал, что примерно:

- 25% пользователей используют ЭС;
- 25% собираются приобрести ЭС в ближайшие 2—3 года;
- 50% предпочитают провести исследование об эффективности их

использования.

В России в исследования и разработку ЭС большой вклад внесли работы Д. А. Поспелова (основателя Российской ассоциации искусственного интеллекта и его первого президента), Э. В. Попова, В. Ф. Хорошевского, В. Л. Стефанюка, Г. С. Осипова, В. К. Финна, В. Л. Вагина, В. И. Городецкого и многих других.

Экспертные системы 60-90-ых годов являются первым поколением экспертных систем, для них характерно:

- 1) знаниями системы являются только знания эксперта, накопление знаний не предусматривается;
- 2) методы представления знаний позволяют описывать лишь статические предметные области;
- 3) модели представления знаний ориентированы на простые предметные области.

Развиваясь, экспертные системы вышли за эти рамки. Принципы представления знаний в экспертных системах второго поколения изменились:

- 1) используются не поверхностные знания, а более глубинные;
- 2) для представления знаний привлекаются средства и методы других направлений искусственного интеллекта, например, нейронных сетей;
- 3) системы имеют динамической базы знаний.

Появление интернета не могло не повлиять на развитие экспертных систем. Возможность получать знания через сеть и извлекать знания из сети не могла не быть использована разработчиками. Поэтому сейчас развиваются распределенные и web-ориентированные экспертные системы.

Сейчас количество экспертных систем исчисляется тысячами и десятками тысяч. В развитых зарубежных странах сотни фирм занимаются их разработкой и внедрением в различные сферы жизни.

В качестве современных ЭС можно назвать быстродействующую систему OMEGAMON (фирма Candle, с 2004 г. IBM) для отслеживания состояния корпоративной информационной сети и G2 (фирма Gensym) - коммерческую экспертную систему для работы с динамическими объектами.

Экспертные системы используют в тех случаях, когда недостаточно экспертов, в опасных (вредных) для них условиях, в процессе обучения. Экспертные системы решают задачи, для решения которых отсутствуют четкие алгоритмы решения.

Модель экспертных систем

Экспертные системы работают в диалоговом режиме (отвечают на поставленные пользователем вопросы), при этом, они должны уметь объяснять, откуда получено то или иное решение, любая экспертная система содержит как минимум пять компонентов или подсистем (рис. 30).

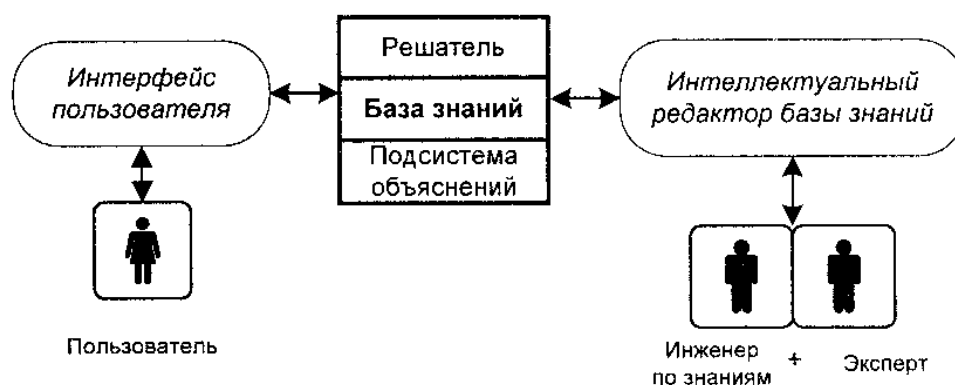


Рис. 30. Базовая структура экспертной системы

Пользователь экспертной системы - специалист предметной области, для которого предназначена система. Обычно его квалификация недостаточно высока и поэтому он нуждается в помощи и информационной поддержке своей деятельности.

Инженер по знаниям — специалист в области искусственного интеллекта, работающий с экспертами и формирующий базу знаний. Синонимы: *когнитолог, инженер-интерпретатор, аналитик.*

Интерфейс пользователя — комплекс программ, реализующих диалог пользователя с ЭС как на стадии ввода информации, так и при получении результатов.

База знаний — ядро экспертной системы, совокупность знаний предметной области, записанная на машинный носитель в форме, понятной эксперту и пользователю (обычно на некотором языке, приближенном к естественному).

Решатель — программа, моделирующая ход рассуждений эксперта на основании знаний, имеющихся в базе знаний. Синонимы: *дедуктивная машина, машина вывода, блок логического вывода.*

Подсистема объяснений — программа, позволяющая пользователю получить ответы на вопросы: "Как была получена та или иная рекомендация?" и "Почему система приняла такое решение?" Ответ на вопрос "как" — это трассировка всего процесса получения решения с указанием использованных фрагментов базы знаний, т. е. всех шагов цепи умозаключений. Ответ на вопрос "почему" — ссылка на умозаключение,

непосредственно предшествовавшее полученному решению, т. е. отход на один шаг назад. Развитые подсистемы объяснений поддерживают и другие типы вопросов.

Интеллектуальный редактор базы знаний — программа, представляющая инженеру по знаниям возможность создавать базу знаний в диалоговом режиме. Включает в себя систему вложенных меню, шаблонов языка представления знаний, подсказок и других сервисных средств, облегчающих работу с базой.

Описанная структура является базовой и может расширяться (рис. 31).



Рис. 31. Структура экспертной системы

Обозначенные штриховой линией подсистемы моделирования внешнего мира, интерфейс с внешним миром, система датчиков – необходимы для экспертных систем реального времени для получения данных и их интерпретации. Экспертные системы могут накапливать опыт в виде прецедентов (уже разрешенных ситуаций), которые сохраняются в базе знаний и используются в дальнейшем.

Блок алгоритмических методов решения включает в себя все вычислительные операции и алгоритмы, реализуемые методами традиционного программирования, интегрированные в экспертную систему. Объединение в рамках экспертной системы методов традиционного

программирования и искусственного интеллекта позволяет значительно повысить эффективность и качество принимаемых решений.

Специфика предметной области, для которой строится система, отображается описывается не только в базе знаний, но и в подсистеме «Контекст предметной области», которая позволяет более наглядно представить входную и выходную информацию в принятом для конкретной предметной области виде.

Современные информационные системы часто используют архитектуру клиент-сервер, позволяющую строить распределенные и сетевые приложения. При клиент-серверной архитектуре экспертной системы на стороне клиента находятся интерфейсы пользователя, эксперта, внешней среды и система датчиков. Остальные блоки располагаются на серверной стороне.

Классификация экспертных систем и оболочек экспертных систем

Существующее множество экспертных делится на несколько классов (рис.32) по различным критериям.

По назначению выделяют системы общего назначения, которые претендуют на универсальность в решении задач (CASNET), специализированные, решающие конкретную задачу (1-st Clas, Элис) или ориентированные на определенную предметную область (MYCIN, MACSYMA, МОДИС, ДИАГЕН, INTERNIST-I).

По критерию взаимодействия с внешней средой различают статические системы, в которых есть только интерфейс пользователя, а механизма взаимодействия с внешним миром, например, через датчики, отсутствует. Динамические экспертные системы с помощью встроенных интерфейсов получают информацию с внешних датчиков или других устройств. Квазидинамические экспертные системы получают информацию об изменении ситуации во внешней среде через заданный промежуток времени (больше нескольких секунд).

Экспертные системы разрабатываются для различных ЭВМ и различаются по *аппаратно-программной платформе*. Они разрабатываются и эксплуатируются на персональных (PROSPECTOR), на символьных (Picon), на мини- (СПЭИС) и на суперкомпьютерах (ЭКСПЕРТИЗА).

По степени интеграции делят экспертные системы на автономные программные комплексы (ДАМП), которые работают самостоятельно, или экспертная система может быть частью более общей системы (интегрированные системы) или же быть звеном в цепочке программ,

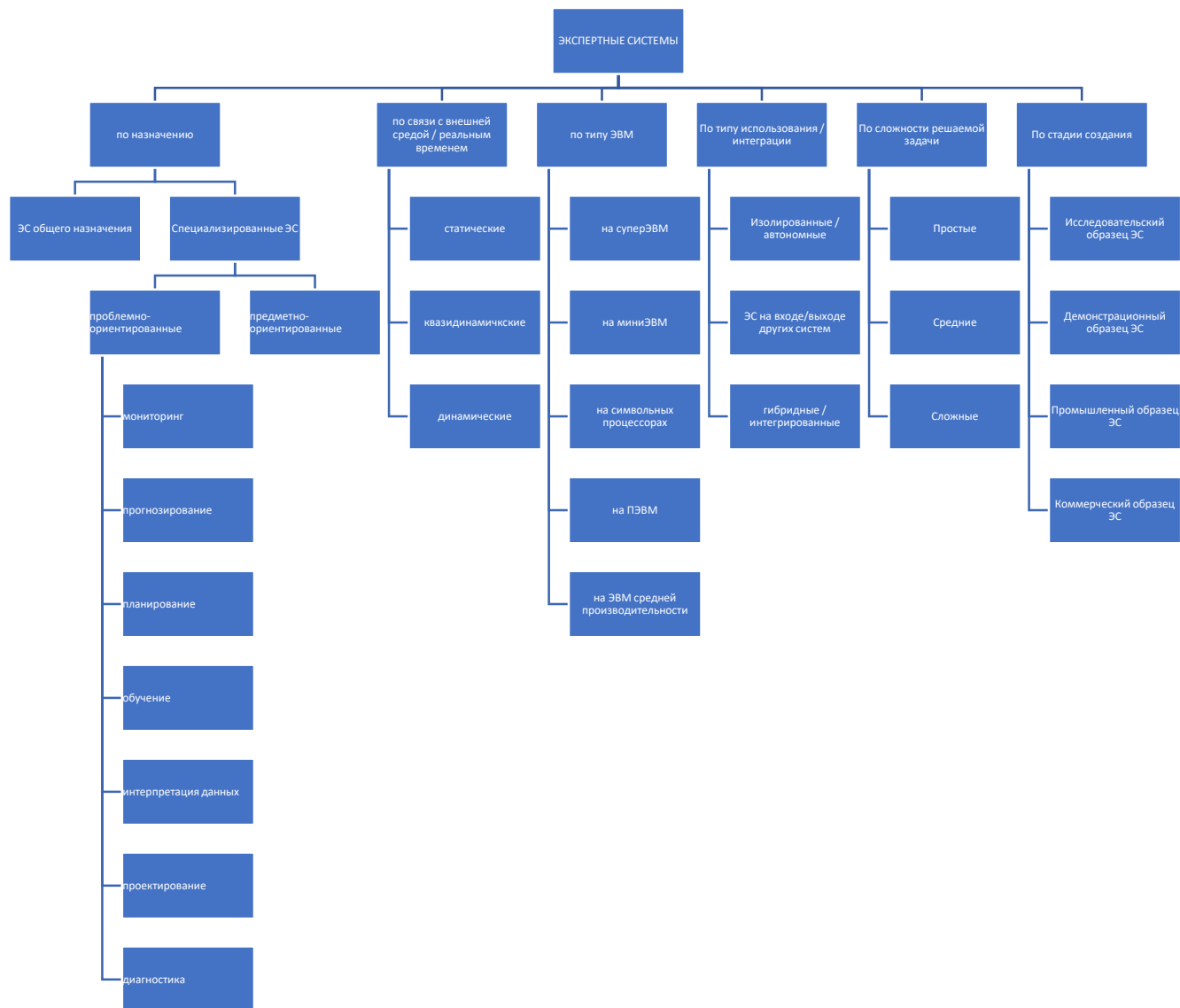


Рис 32. Классификация экспертных систем

обрабатывающих информацию с общей целью, например управление предприятием.

В зависимости от *размера базы знаний* выделяют простые экспертные системы - до 1000 простых правил (GUIDON, Плотина), - средние - от 1000 до 10000 структурированных правил (XCON, GOSSEYN, ДИАГЕН) - и сложные - более 10000 структурированных правил.

Экспертные системы различаются *по стадии существования*, т.е. по степени завершенности системы. Первая стадия существования экспертной системы – это исследовательский образец - разрабатывается 3-6 месяца с минимальной базой знаний (SYSTEM-D, SYN), вторая - демонстрационный образец - разрабатывается 6-12 месяца (THYROID MODEL), третья - промышленный образец - разрабатывается 1-1,5 года с полной базой знаний (PUFF, FOSSIL) и последняя – коммерческий образец - разрабатывается 1,5-3 года с полной базой знаний (KNEECAP, MACSYMA).

Средства разработки экспертных систем

Существующие средства разработки экспертных систем можно разделить на 3 класса (рис. 33). *Традиционные языки программирования* (C++, Java, Delphi) позволяют построить экспертные системы «с нуля» для конкретной задачи или предметной области, обеспечив хорошие показатели качества и необходимую функциональность системы, но на разработку требуются значительные временные и финансовые ресурсы. Так создают экспертные системы любой стадии существования, в особенности, коммерческие системы, продажа которых возместит затраты.

Языки искусственного интеллекта (LISP, PROLOG, Рефал) были разработаны специально для представления знаний, построение с их помощью экспертных систем позволяет более легко оперировать экспертными знаниями, но ограничивают способ их представления структурой языка. С помощью языков искусственного интеллекта создаются исследовательские и демонстрационные образцы.

Следующий класс средств построения экспертных систем – специальный программный инструментарий – ориентирован только на создание интеллектуальных информационных систем и делится на два подкласса: оболочки и среды разработки интеллектуальных систем.

Среды разработки являются программными комплексами, позволяющими строить системы из отдельных готовых блоков, на их основе создаются демонстрационные и промышленные образцы экспертных систем.

Оболочка экспертных систем - инструментальное средство для проектирования и создания экспертных систем. В состав оболочки входят средства проектирования базы знаний с различными формами представления знаний и выбора режима работы решателя задач. Для конкретной предметной области инженер по знаниям определяет нужное представление знаний и стратегии решения задач, а затем, вводя их в оболочку, создает конкретную экспертную систему.



Рис.33. Классификация инструментальных средств разработки ЭС

Применение оболочки позволяет достаточно быстро и с минимальными затратами создать исследовательскую, демонстрационную или промышленную экспертную систему. Оболочки можно классифицировать следующим образом (рис.34). По степени обработанности выделяют экспериментальную (GPSI), исследовательскую (Expert) и коммерческую (EXSYS) оболочки.

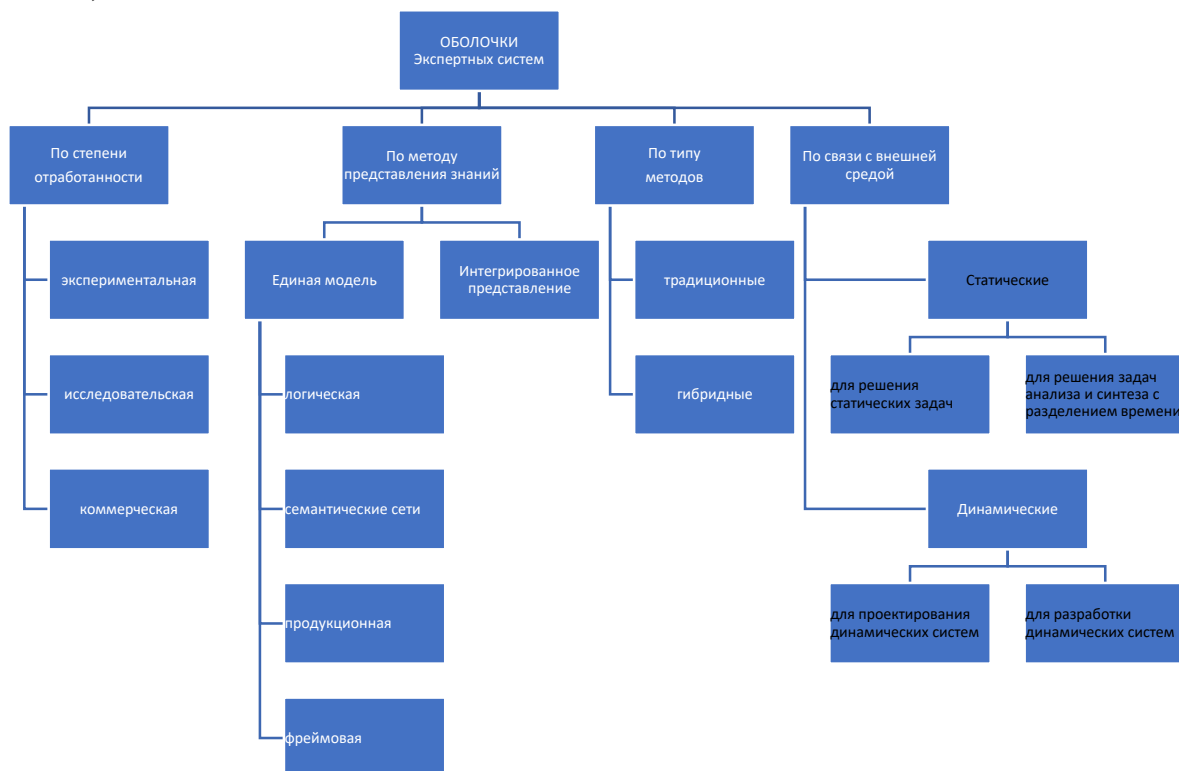


Рис.34 . Классификация оболочек экспертных систем

Знания в базе могут быть представлены одним способом (EMYCIN, CLIPS) - семантической сетью, продукциями, фреймами и т.д.) или же несколькими (MINEVRA, EsWin), для создания более полной, гибкой и наглядной модели предметной области.

Используемые в оболочке методы могут быть традиционными (CubiCalc, NEXPERT, Алеф) - алгоритмы, деревья вывода и т.д. - и гибридными (FuzzyCLIPS, MultiNeuron) - совместно с традиционными используются нейронные сети, нечеткая логика и т.д.

Существуют статические оболочки, предназначенные для решения статических задач (1-st Clas, Элис). Они характеризуются использованием поверхностной технологии, общих правил и поиска решения от цели к данным, применяются для решения задач анализа.

Статические оболочки, предназначенные для решения задач анализа и синтеза с разделением времени (KAPPA, Clips), используют глубокий и структурный подходы, осуществляют поиск решений - от цели к данным и от данных к цели.

Оболочки для проектирования динамических систем (Framework, NExpert) применяют поверхностный подход, принимают решения на основе правил общего вида.

Оболочки для разработки динамических систем (G2, Rethink, RkWorks) имеют подсистему моделирования, планировщика решений, используют смешанную технологию, правила общего и частного вида, решение задачи анализа и синтеза в реальном времени.

EMYCIN – первая оболочка, основанная на MYCIN. Принципы, разработанные для PROSPECTOR, были использованы при создании таких систем, как KAS, SAGE, SAVOIR.

Изменение принципов построения ведет к развитию инструментария. Поэтому оболочки прошли тот же эволюционный путь, что и ЭС. Современные оболочки предлагают следующие возможности (в каждой конкретной оболочке представлены частично):

- гибридное представление знаний (EsWin);
- выбор из нескольких стратегий вывода (G2, CLIPS);
- подключение библиотек и других систем (ACTIVATION FRAMEWORK);
- архитектура на основе «доски объявлений» (HEARSAY-III);
- архитектура «клиент-сервер» (JESS);
- интеграция в интернет / интранет (Egg2Lite, Exsys Corvid);
- графический интерфейс (WindExS, WxCLIPS);

- подсистема моделирования (G2);
- модульное построение системы (ReThink, G2);
- визуализация структуры БЗ (W.E.S.T.) и т.д.