



СПИРИДОНОВ
Игорь Николаевич
доктор технических наук,
профессор



ГАВРЮШИН
Сергей Сергеевич
доктор технических наук,
профессор



ЕВГЕНЕВ
Георгий Борисович
доктор технических наук,
профессор
кафедры
«Компьютерные системы
автоматизации
производства»
(МГТУ им. Н.Э. Баумана)

Многоагентная биометрическая система

И.Н. Спиридонов, С.С. Гаврюшин, Г.Б. Евгеньев

Описана методология создания интегрированных интеллектуальных биометрических систем. Представлена общая схема биометрической системы. Изложена многоагентная методология создания интегрированных интеллектуальных систем. Построена концептуальная биометрическая модель человека. Описана CASE-технология создания многоагентных систем, включая средства генерации методов агентов, а также методы извлечения знаний из данных.

Ключевые слова: биометрическая система, интеллектуальная система, мультиагентная методология, CASE-технология, база знаний, извлечение знаний из данных.

The methodology for creating integrated intelligent biometric systems is considered. An overview of the biometric system is presented. A multi-agent methodology for creating integrated intelligent systems is described. A conceptual model of a human biometrics is constructed. A CASE-technology to create multi-agent systems is described, including the generation of agents, as well as methods to extract knowledge from data.

Keywords: biometric system, intelligent system, multi-agent methodology, CASE-technology, knowledge base, knowledge extraction from data.

В конце XX века сформировалось новое направление создания современных систем защиты от несанкционированного доступа на основе использования в качестве полезной информации статических биометрических характеристик человека (БХЧ) — параметры отпечатков пальцев, изображения радужной оболочки глаза (РОГ), голоса, изображения лица, и динамических БХЧ — параметры манеры работы на клавиатуре компьютера, динамики подписи, походки, потенциально обеспечивающих возможность явной и скрытой идентификации личности.

Первые биометрические системы предназначались для обеспечения доступа к информации в ПЭВМ и банковским счетам по голосу, отпечаткам пальцев, изображениям лица и РОГ. Производители биометрических устройств справедливо полагают, что их продукция надежнее паролей и микропроцессорных карточек. Основная доля доходов приходится на биометрические технологии по отпечаткам пальцев, геометрии рук и лица. Биометрические технологии получили поддержку со стороны Microsoft, объявившей о своем намерении обеспечить поддержку биометрической верификации в различных операционных системах семейства Windows.

Помимо описанных выше задач идентификации и верификации личности БХЧ используются при определении функционального состояния человека. Под функциональным состоянием человека по-

нимают психофизиологическое явление, закономерности которого заложены в модулирующих системах головного мозга и которое проявляется на биохимическом, физиологическом, фенотипическом и поведенческом уровнях.

В последнее время обоснована взаимосвязь особенностей генотипа человека и информативных морфогенетических вариантов головы, шеи, РОГ, гребешковой кожи ладоней и др. Наибольшее практическое применение среди них получили дерматоглифический (ДФ) и иридоглифический фенотипы (ИФ). В медицине и биологии параметры ДФ и ИФ, например, используют для описания особенностей генотипа, физиологических и поведенческих реакций, симптомов наследственных и врожденных болезней, в криминалистике — для идентификации личности, в антропологии — для описания вида и т. д. В настоящее время ДФ и ИФ являются основными показателями функционального статуса человека.

Если при построении несложных систем идентификации и верификации личности достаточно использования ограниченного набора биометрических характеристик человека, то при комплексном определении функционального статуса человека необходимо построение интегрированных интеллектуальных систем, основанных на концептуальной биометрической модели человека. В этом случае результаты многовековых исследований ДФ и ИФ будут эффективно применены в медицине, генетике, антропологии, криминалистической экспертизе, профотборе и др.

Биометрические технологии

Биометрика — область знаний, изучающая методы и средства измерения и формализации персональных физических характеристик и поведенческих черт человека, а также их использование для идентификации или верификации человека и определения функционального статуса человека.

На рисунке 1 представлена общая схема биометрической системы. Первый функциональный блок обеспечивает получение данных с помощью сенсора. Сенсор — промежуточная

связь между реальным миром и системой; он должен получить все необходимые данные.

Второй блок осуществляет все необходимые предварительные процессы: он должен удалить все «лишнее» с сенсора (датчика) для увеличения чувствительности на входе (например удаление фоновых шумов при распознавании голоса).

В третьем блоке извлекаются необходимые данные. Это важный шаг, так как корректные данные нуждаются в извлечении оптимальным путем. Вектор значений с особыми свойствами используется для создания шаблона. Шаблон — синтез (совокупность) релевантных характеристик, извлеченных из источника. Элементы биометрического измерения, которые не используются в сравнительном алгоритме, не сохраняются в шаблоне, чтобы уменьшить размер файла и защитить личность регистрируемого, сделав невозможным воссоздание исходных данных по информации из шаблона.

Информация, представленная шаблоном, хранится в базе данных биометрической системы. Полученный шаблон передается к сравнителю, который сравнивает его с другими существующими шаблонами, оценивая разницу между ними с использованием определенного метода. Сравнивающая программа анализирует шаблоны с поступающими, а затем эти данные передаются для любого специализированного использования.

Анализ шаблона может производиться базой знаний без сравнения с другими шаблонами. База знаний формирует необходимое заключение.

Биометрические данные можно разделить на два основных класса:

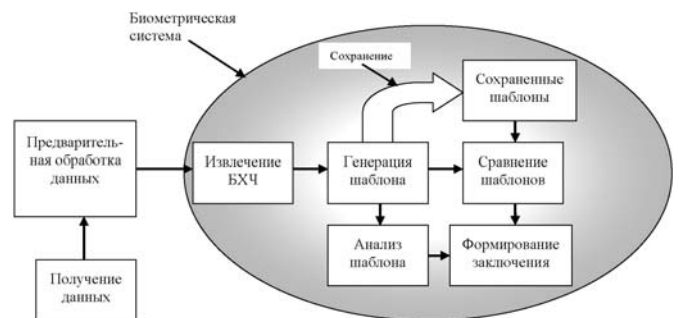


Рис. 1. Общая схема биометрической системы

1) *физиологические* — относятся к форме тела. Например, отпечатки пальцев, распознавание лица, ДНК, ладонь руки, сетчатка глаза, запах/аромат;

2) *поведенческие* связаны с поведением человека. Например, походка и голос. Для этого класса биометрии используется термин *behaviometrics*.

Выбор источника БХЧ является основной задачей при создании конкретных биометрических технологий (БТ). Идеальная БХЧ должна быть универсальной, уникальной, стабильной, собираемой. Универсальность означает наличие биометрической характеристики у каждого человека. Уникальность означает, что не может быть двух человек, имеющих идентичные значения БХЧ. Стабильность — независимость БХЧ от времени. Собираемость — возможность получения биометрической характеристики от каждого индивидуума. Обычно перечисленные четыре характеристики оцениваются по трехбалльной шкале. В таблице 1 даны суммарные оценки различных источников БХЧ.

Таблица 1

Характеристика	Балл
Видеообраз лица	9
Термограмма лица	10
Отпечаток пальца	10
Геометрия руки	9
Радужная оболочка глаза	11
Сетчатка	9
Подпись	6
Голос	6
Отпечаток губ	9
Особенности ушной раковины	8
Динамика письма	10
Походка	7

Как видно в табл. 1 из 12 характеристик половина принадлежит лицу человека (видеообраз лица, термограмма лица, радужная оболочка глаза, сетчатка, отпечаток губ и особенности ушной раковины), а две — его руке (отпечаток пальца и геометрия руки). Подпись, голос и походка имеют низшие баллы. Из числа ха-

рактеристик, не принадлежащих лицу и руке, высокий балл имеет только динамика письма.

Многоагентная методология создания интегрированных интеллектуальных систем

Чтобы вывести биометрические системы на принципиально новый уровень, способный обеспечить комплексное решение проблем биометрической технологии, необходимо построить концептуальную биометрическую модель человека.

Создание больших программных систем невозможно без использования CASE-технологий. CASE-технология — совокупность методологий анализа, проектирования, разработки и сопровождения сложных систем программного обеспечения (ПО), поддержанную комплексом взаимосвязанных средств автоматизации. CASE — компьютеризированный инструментарий разработчиков программных продуктов. Основная цель CASE состоит в том, чтобы отделить проектирование ПО от его кодирования и последующих этапов разработки, автоматизировав их в максимальной степени [1].

Большинство CASE-средств основано на парадигме методология/метод/нотация/средство. Методология определяет шаги работы и их последовательность, а также правила распределения и назначения методов. Метод — это систематическая процедура генерации описаний компонент ПО. Нотации предназначены для описания структур данных, порождающих систем и метасистем. Средства — инструментарий для поддержки методов с использованием принятой нотации. Эти инструменты поддерживают работу пользователей при создании и редактировании графического проекта в интерактивном режиме, они способствуют организации проекта в виде иерархии уровней абстракции, осуществляют генерацию ПО и используются при его тестировании.

В данной работе описывается методология, основанная на многоагентной CASE-технологии применительно к созданию интегрированных интеллектуальных биометрических систем.

Агент — *объект*, который подвергается воздействию со стороны и сам воздействует на

другие *объекты*; возникает в *среде*, где он может выполнять определенные *действия*, способен к восприятию части своей *среды*, может общаться с другими *агентами* и обладает автономным *поведением*, являющимся следствием его наблюдений, знаний и взаимодействий с другими агентами. Агент представляет собой дальнейшее развитие понятия объект. *Понятие объект* не связано с наличием *среды*, которая играет существенную роль в определении агента. *Объект*, в принципе, не требует существования себе подобных, а агент не может быть один. Таким образом, агент — это *подкласс объектов*, обладающий всеми их свойствами, но имеющий также дополнительные качества. С прагматической точки зрения агент — это *система*, обеспечивающая решение определенной задачи и действующая во взаимосвязи с сетью других агентов для решения комплексной проблемы, которое не может быть получено отдельными агентами. Агенты в *многоагентной сети* гетерогенны, т. е. принадлежат разным *классам* [1].

Агенты в интеллектуальных биометрических системах соответствуют органам человеческого тела.

Многоагентные системы (МАС) принадлежат к классу интеллектуальных систем распределенного решения задач. Их основу составляет иерархическая метасистема агентов, проектируемая сверху вниз. Применительно к интеллектуальным биометрическим системам в качестве агентов рассматриваются модели тела человека и входящих в него органов.

Формальную модель МАС можно представить так

$$MAS = (A, E, R, ORG),$$

где A — множество агентов; $E = \{e\}$ — среда, в которой находится данная МАС; R — множество взаимодействий между агентами; ORG — множество базовых организационных структур, соответствующих конкретным функциям (ролям) агентов, устанавливающая отношения между ними.

В биометрических системах множество агентов A формируется, как было упомянуто, из моделей тела человека и его органов. В качестве среды $E = \{e\}$ выступает база данных, хра-

нящая информацию о состоянии человека в разное время. Задачи, которые должна решать интеллектуальная биометрическая система, определяются методами, встроенными в модели органов человека. Методы обеспечивают решение задач идентификации или верификации человека и определения его функционального статуса.

Множество взаимодействий R между агентами определяется ребрами графа экспорта и импорта свойств органов человека, необходимых для работы методов. Эти взаимодействия носят как вертикальный, так и горизонтальный характер. Вертикальные взаимодействия осуществляются между агентами, связанными друг с другом по иерархии организационной структуры ORG , а горизонтальные — между иерархически не связанными агентами.

Наконец, организационная структура ORG в МАС представляет собой иерархическую метасистему, моделируемую И/ИЛИ графом. Связки типа И описывают отношения класса «целое—часть», а связки типа ИЛИ — отношения класса «род—вид». Связки типа И в медицинских системах описывают декомпозицию организма человека; связки типа ИЛИ — разновидности органов, например, мужских и женских органов мочеиспускания.

Для моделирования МАС в теории искусственного интеллекта (ИИ) наиболее подходят семантические сети. Семантическая сеть — структура данных, состоящая из узлов, соответствующих понятиям (органам), и связей, указывающих на взаимосвязи между узлами (органами). В объектно-ориентированном подходе этому понятию при представлении знаний практически однозначно соответствует диаграмма классов языка UML [2], которая будет описана ниже.

Недостатком традиционных диаграмм классов для моделирования МАС является отсутствие возможностей моделирования взаимодействий агентов при экспорте и импорте свойств. Такая возможность является важнейшим элементом построения интеллектуальных биометрических систем.

Согласно принципам объектного подхода, объект — это абстракция множества предметов

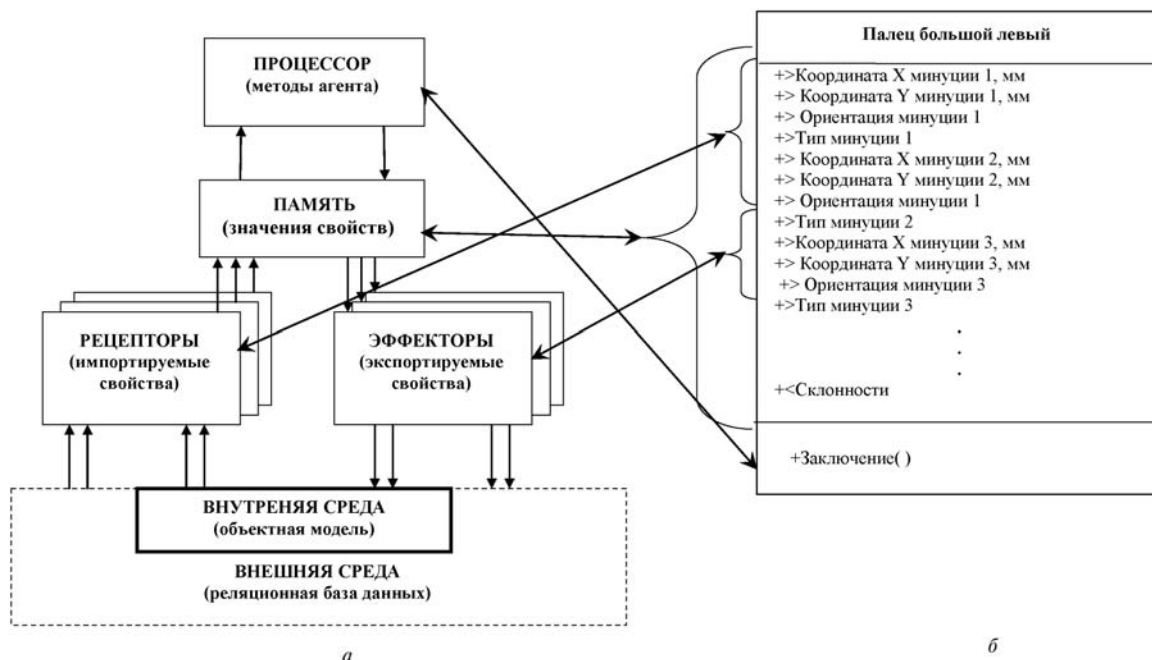


Рис. 2. Архитектура агента

реального мира, в которой: 1) все предметы множества (экземпляры) имеют одни и те же характеристики (свойства); 2) все экземпляры подчинены и согласовываются с одним и тем же набором правил поведения. Состояние объекта характеризуется перечнем его свойств и текущим значением каждого из этих свойств, а поведение объекта определяется методом, состоящим из набора правил. При объектно-ориентированном подходе к проектированию программных систем есть две относительно самостоятельных и вместе с тем тесно взаимосвязанных стороны моделирования: статическая и динамическая. Статическое моделирование определяет структуру классов и объектов, а динамическое — их поведение.

В объектно-ориентированном подходе существуют понятия класса и экземпляра. Класс — это описание множества объектов, обладающих общими атрибутами, операциями, отношениями и семантикой. Экземпляр представляет собой конкретную материализацию абстракции. Эта сущность обладает состоянием, в котором запоминаются результаты операций, реализуемых методами.

Обобщенная модель класса искусственных агентов приведена на рис. 2, а. Любой агент представляет собой открытую систему, поме-

щенную в некоторую среду. Этой средой является база данных состояний человека.

Свойства агента могут принадлежать трем различным категориям: импортируемым, экспортируемым и внутренним. Импортируемые свойства являются рецепторами агента, формирующими его систему восприятия. Экспортируемые свойства агента являются его эффекторами, функция которых состоит в воздействии на среду, т. е. на состояние человека.

Свойства агента всех трех категорий образуют его память, в которой хранится текущее состояние агента.

Процессор агента формируют его методы, обеспечивающие объединение и переработку разнородных данных, выработку соответствующих реакций на информацию о состоянии среды, принятие решений о постановлении тех или иных выводов.

В целом процессор определяет поведение агента, которое можно наблюдать, используя инспектор модели агента. С помощью модели пользователь следит за состоянием свойств агента, либо в графическом окне, с отображением результатов.

На рисунке 2, б приведена модифицированная модель класса объектов в нотации UML. Из рисунка следует, что процессор агента соот-

ветствует операциям класса, память агента — значениям атрибутов класса. Несоответствие заключается в том, что агент имеет импортируемые и экспортируемые свойства, а в нотации класса UML это не предусмотрено. Знаком «+» в UML отмечаются свойства видимые извне. Простейшим решением этой проблемы может быть пометка импортируемых свойств знаком «>», а экспортируемых — знаком «<», как это показано на рис. 2, б. При этом агент принимает форму статической объект-функции [1].

Метод описания организационной структуры тела человека

Как было отмечено выше, описание организационной структуры тела человека может быть произведено с помощью диаграмм классов объектов языка UML. На рисунке 3 представлена диаграмма классов концептуальной биометрической модели человека. Корневым объектом здесь является Организм человека. В качестве свойств он имеет: Num — кодовый номер пациента, его фамилию имя и отчество, год рождения, по которому в методе объекта вычисляется текущий возраст человека, а также пол, рост и вес. В больших диаграммах классов полный набор свойств, как правило, не отображается, а зачастую опускается.

Состав концептуальной модели задается с помощью линий с зачерненными ромбами.

Как следует из рис. 3 в концептуальную биометрическую модель человека входят базовые компоненты: лицо и кисть руки, которые могут быть подвергнуты дальнейшей декомпозиции.

Лицо включает следующие компоненты: лоб, брови, нос, уши, глаза, волосы, усы и борода.

Линии с зачерненными ромбами представляют собой отношение «целое-часть». Другой разновидностью связей являются отношения «род-вид». Они изображаются линиями со светлыми треугольниками.

Помимо перечисленных основных видов связей имеются связи зависимостей. Они отображаются пунктирными линиями, направленными на объект, от которого зависит данный.

Средство генерации компьютерной модели организационной структуры тела человека и взаимодействий между органами

В CASE-технологии помимо описанных выше методологии, метода и нотации описания организационной структуры тела человека и связей между органами необходимо включать инструментальные средства для генерации компьютерных моделей. Применительно к описы-

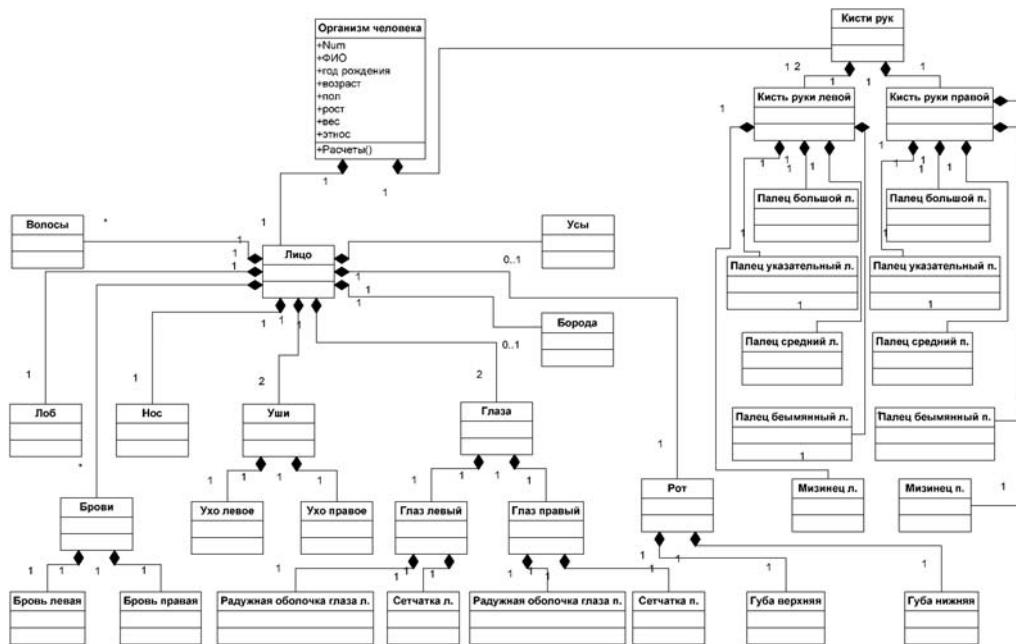


Рис. 3. ORG-диаграмма

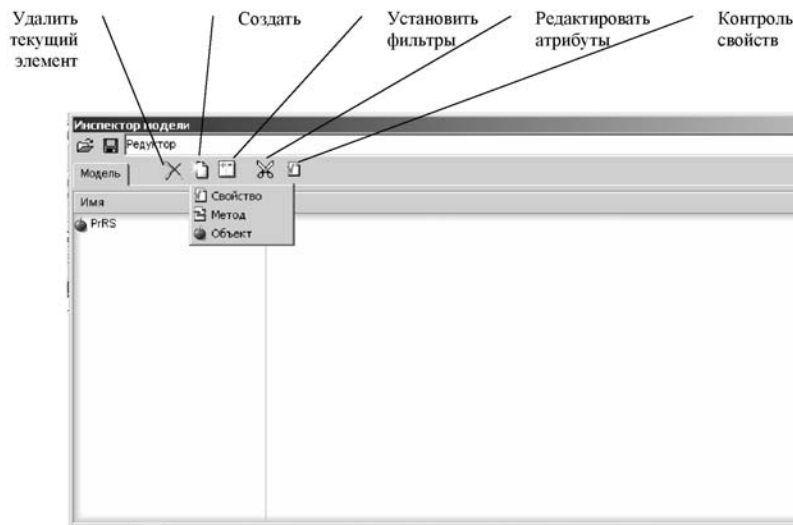


Рис. 4. Системная панель системы Sprut-X

ваемой технологии таким средством является система Sprut-X.

На рисунке 4 представлена системная панель системы Sprut-X. С ее помощью можно создавать и удалять элементы, устанавливать фильтры видимости элементов, редактировать атрибуты и контролировать свойства объектов. В выпадающем меню на рис. 4 представлен набор элементов, которыми можно оперировать: объект, свойство и метод. Тут же приведены иконки этих элементов. Имя создаваемой модели, а также кнопки ее открытия и записи расположены на верхней панели. Поле под системной панелью предназначено для отображения текущего состояния модели.

На рисунке 5 представлена сгенерированная в Sprut-X концептуальная модель организационной структуры биометрической системы, соответствующая диаграмме классов (см. рис. 3). Корневым объектом биометрической системы (BIOMETR) является Организм человека (HumanBody). Этот объект состоит из двух компонент: «Лицо» (Face) и «Кисти рук» (Hands). В состав объекта «Лицо» входят: «Лоб» (Forehead), «Нос» (Nose), «Брови» (Nose), «Уши» (Ears), «Глаза» (Eyes) и «Рот» (Mouth). В свою очередь брови, уши и глаза состоят из двух компонент — «правой» и «левой», а «Рот» из губ верхней и нижней.

В состав объекта «Кисти рук» входят: «Кисть руки левой» (Left_Hand) и «Кисть руки правой»

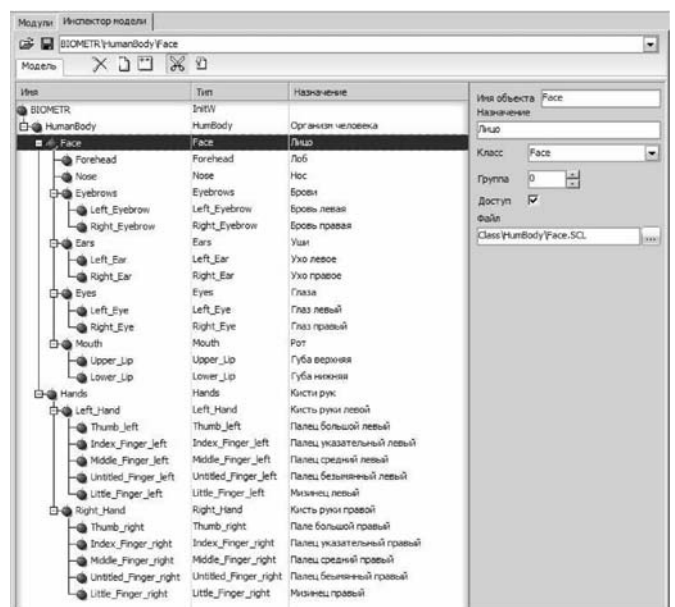


Рис. 5. Организационная биометрическая структура

(Right_Hand). Каждая кисть состоит из пяти пальцев.

На рисунке 6 представлен фрагмент организационной диаграммы со свойствами биометрической системы. В диаграмму введены свойства узоров пальцев. Свойства могут иметь три типа: целое (integer), действительное (real) и строковое (string). Для построения диаграммы связей органов по свойствам каждое из них может иметь один из следующих трех статусов: экспорт (export), импорт (import) и внутреннее (internal).

Справа на рис. 6 отображена диаграмма связей органов по свойствам (R-диаграмма). На

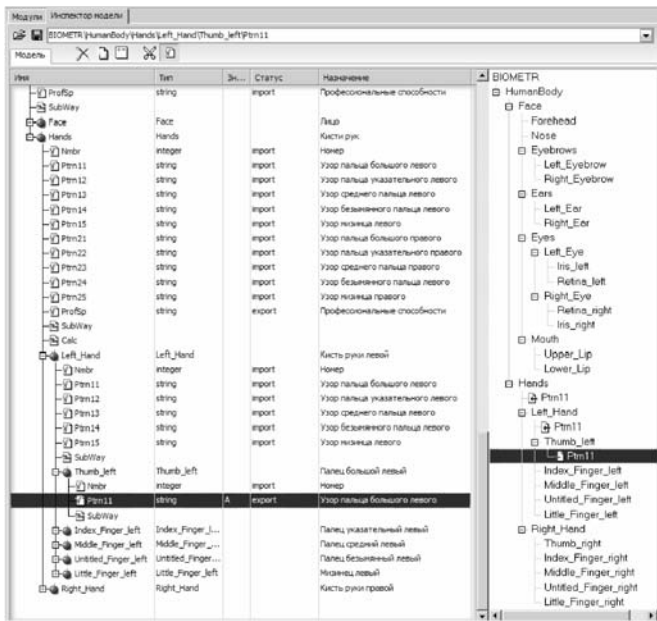


Рис. 6. Свойства и диаграмма связей свойств биометрической системы

этой диаграмме показано, что узор пальца большого левого со значением А экспортируется сначала в свойства кисти левой руки, а затем в свойства кистей рук. Эта функция является важнейшим элементом для построения интегрированных интеллектуальных систем в биометрике.

На рисунке 7 представлен фрагмент организационной диаграммы с подключенными мето-

дами. К объекту «Кисти рук» подключен метод определения склонностей человека.

Конструирование методов агентов

Описанные выше компоненты представления знаний определяют концептуальную модель организма человека и относятся к категории статических. Динамическими компонентами системы, обеспечивающими преобразование исходных данных в результирующие, являются методы агентов. Они определяют функциональность системы, которая заключается преимущественно в идентификации личности и определения ее характеристик.

Имеются два принципиально отличных способа конструирования методов агентов: процедурный, или алгоритмический, и непроедурный, или декларативный. Алгоритмический способ требует использования соответствующих языков и привлечения программистов. В интеллектуальных системах он не применяется. Декларативный способ основывается на базах знаний [3] и доступен для непрограммирующих носителей знаний.

Существует несколько источников знаний, на базе которых могут быть построены интеллектуальные системы. Знания могут содержаться в литературных источниках, методиках и головах экспертов. В этом случае базы зна-

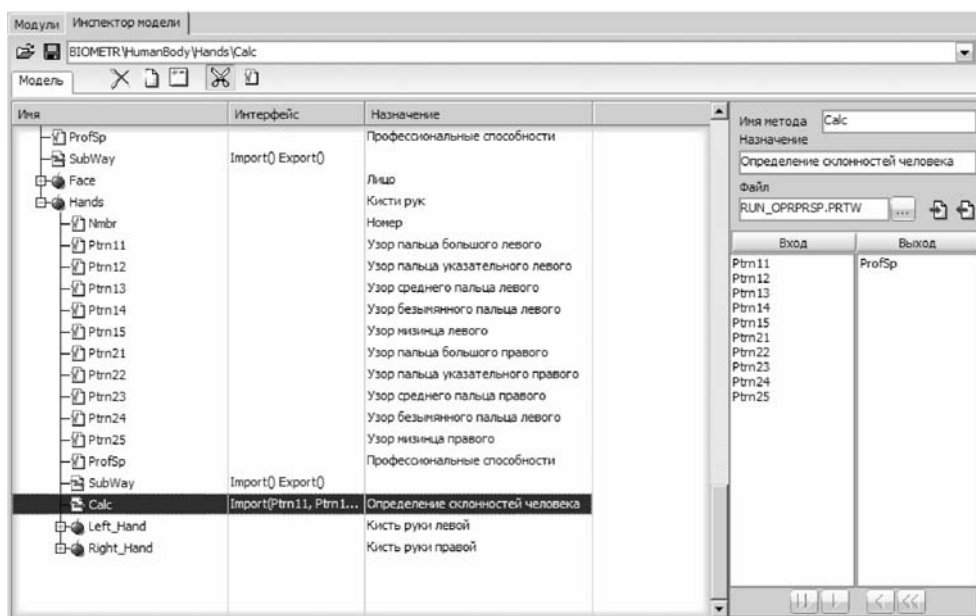


Рис. 7. Фрагмент организационной диаграммы с подключенными методами



Рис. 8. Функциональные схемы конструирования интеллектуальных систем

ний разрабатываются экспертами соответствующими инструментальными средствами [1]. На рисунке 8, а с использованием стандарта IDEF0 представлена функциональная схема создания интеллектуальной системы. Процесс ведется под управлением пользователя с использованием в качестве механизмов реализации соответствующей инструментальной среды и баз данных, в которых хранится информация о ресурсах, необходимых для принятия решений.

Другим источником знаний могут быть базы данных, в которых знания содержатся в неявном виде (рис. 8, б). Здесь возможна автоматическая генерация баз знаний с использованием средств извлечения знаний из баз фактов (knowledge discovery). Такой подход носит название интеллектуального анализа данных (ИАД) [4].

Начало развития технологии ИИ относится ко второй половине 50-х годов прошлого века, когда начались интенсивные поиски моделей и алгоритмов человеческого мышления и разработка первых программ на их основе. От теоретических поисков исследовательского характера к первым системам, обеспечивающим по-

лучение практических результатов, перешли в середине 1970-х годов, когда на смену поискам универсального алгоритма мышления пришла идея моделировать конкретные знания специалистов-экспертов. В это время появились первые системы, основанные на знаниях, или *экспертные системы* (ЭС).

При построении ЭС были учтены опыт предшествующих исследований в области ИИ, что определило успехи их практического применения. Важнейшим из этого опыта явилось положение, согласно которому мощность ЭС обусловлена, в первую очередь, мощностью базы знаний, содержащей правила принятия решений, и только во вторую очередь методами логического вывода, основанными на этих правилах. Как показало развитие ИИ важнее иметь разнообразные специальные знания, а не общие процедуры вывода. Это положение использовано в экспертном программировании, в котором программы, реализующие функции вывода, генерируются применительно к каждому конкретному набору правил, составляющих базу знаний.

Элементы, из которых складываются ЭС, представляют собой обобщенные функциональные блоки. Наиболее удачным и широко распространенным представлением функциональных блоков является стандарт IDEF0. Конструкция функционального блока приведена на рис. 9.

Каждая из четырех сторон прямоугольника имеет определенное назначение: левая — входы, правая — выходы, верхняя — управление, нижняя — механизмы. Входы представляют собой информацию, необходимую для выполнения функции, и в результате ее выполнения преобразуются в выходы. Входы показывают все характеристики, которые необходимы для выполнения функции и она не может быть вы-

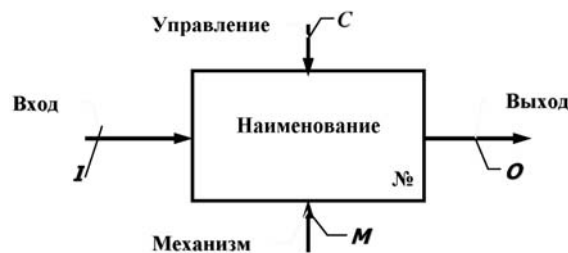


Рис. 9. Функциональный блок ЭС

полнена без получения их значений. Управление описывает условие, оказывающее влияние на выполнение функции, но само не подвергается переработке.

К нижней части изображения функционального блока присоединяются стрелки механизмов, обозначающие программное средство, которые обеспечивают выполнение функции. Входы и выходы показывают, что делается функцией, управление — почему это делается, а механизмы — с помощью чего делается.

Механизмы модулей знаний должны обеспечивать реализацию всех функций, которые могут потребоваться при формировании баз знаний. В число таких функций входят следующие основные:

- вычисление по формулам (в том числе присвоение значений переменным);
- определение значений по таблицам;
- выбор значений из баз данных;
- обновление значений в базах данных;
- занесение значений в базы данных;
- вычисление значений с использованием подпрограмм;
- вычисление значений методами, сгенерированными из модулей знаний;
- вычисление значений с помощью исполняемых exe-модулей или dll-библиотек, сгенерированных другими системами.

В продукционных системах ИИ элементом представления знаний является правило-продукция. Такое правило содержит предусловие, определяющее применимость его при конкретном состоянии переменных базы данных (если <условие>, то <действие>).

В технологии экспертного программирования правила-продукции представляются в форме обобщенных функциональных блоков, описанных выше. Они носят название модули знаний (МЗ). Неструктурированная совокупность МЗ в определенной прикладной области представляет собой базу знаний этой области, аналогичную базе знаний продукционной системы.

Наименования и имена входных, управляющих и выходных переменных МЗ должны выбираться из словаря базы знаний.

Словарь (табл. 2) представляет собой аналог списка терминов и обозначений, который час-

то помещается в начале книги. От такого списка он отличается наличием графы, определяющей тип данных. Используются данные трех типов: действительные числа (REAL), целые числа (INTEGER) и символьные переменные (STRING). Переменная любого типа может быть перечисляемой. Для каждой перечисляемой переменной должен быть составлен список допустимых значений. Например, символьные переменные узоров пальцев Ptrn11, Ptrn21, Ptrn12 и Ptrn22 могут иметь такие значения: A, AL, LA, ALW, L, LW, WL и W. Переменная «Профессиональные способности» (ProfSp) имеет следующие варианты значений: «технические», «гуманитарные», «разносторонние» и «неопределенные».

Словарь в совокупности со значениями содержащихся в нем переменных выполняет в экспертном программировании функцию рабочей памяти продукционной системы.

Таблица 2

Словарь

Имя	Наименование	Тип
Age	Возраст, лет	REAL
BMI	Индекс массы тела	REAL
Weigt	Масса тела, кг	REAL
Height	Рост, м	REAL
Ptrn11\$	Узор пальца большого левого	STRING
DI11	Дельтовый индекс пальца большого левого	INTEGER
Ptrn21\$	Узор пальца большого правого	STRING
DI21	Дельтовый индекс пальца большого правого	INTEGER
Ptrn12\$	Узор пальца указательного левого	STRING
DI112	Дельтовый индекс пальца указательного левого	INTEGER
Ptrn22\$	Узор пальца указательного правого	STRING
DI22	Дельтовый индекс пальца указательного правого	INTEGER
ProfSp\$	Профессиональные способности	STRING

Средство генерации методов агентов

При генерации методов агентов используется инструментальная среда экспертного программирования Sprut-ExPro. На рисунке 10

представлено окно словаря базы знаний этой системы.

В графу «Имя» заносятся идентификаторы переменных, которые используются затем при автоматической генерации программных средств. Графа «Наименование» служит для связи системы с человеком и в нее записываются общепринятые наименования переменных.

Для каждой перечисляемой переменной должен быть составлен список допустимых значений. Наименования этих списков заносятся в графу «Ассоциативный Список». На рисунке 10 изображено окно «Ассоциативные Списки», в котором показаны значения списка «Узор пальца». Этот список привязан ко всем переменным, содержащим описания узоров пальцев.

На рисунке 11 показано окно формирования списка для переменной «Профессиональные способности».

Формирование заголовков и типа механизмов МЗ выполняется в соответствующем окне (рис. 12).

Формирование содержания МЗ производится после нажатия на кнопку «Механизм» и зависит от его типа. В качестве примера рассмотрим формирование двух модулей с механизмами типа «Формула» и типа «Таблица».

Приведенное ниже внешнее представление модуля расчета фенотипа пальцевой дерматоглифики имеет механизм типа «Формула». Этот модуль не имеет предусловия. На основе входных переменных, содержащих дельтовые индексы пальцев обеих рук, по приведенной ниже формуле рассчитывается искомый фенотип.

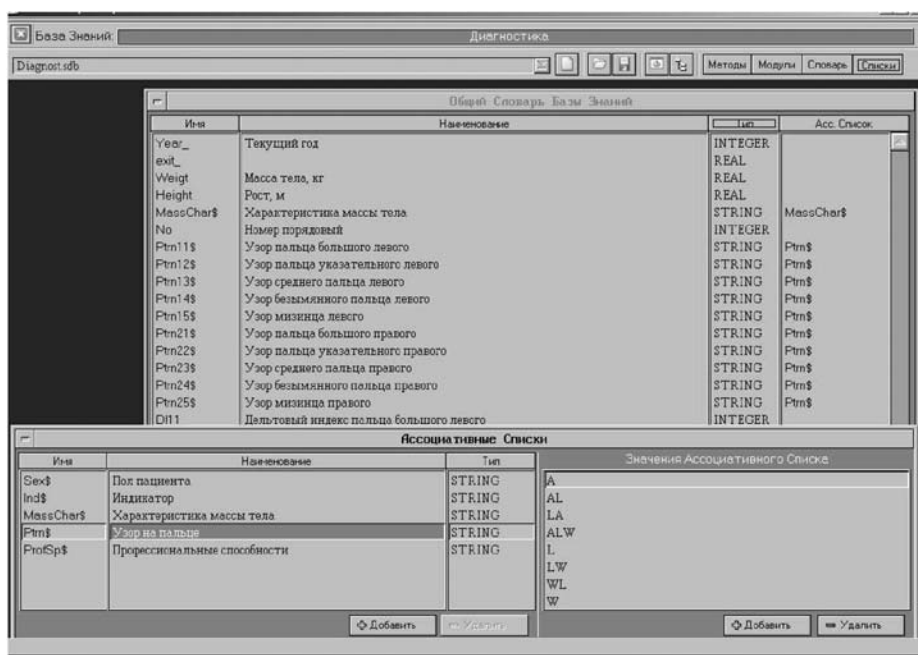


Рис. 10. Окно формирования словаря базы знаний

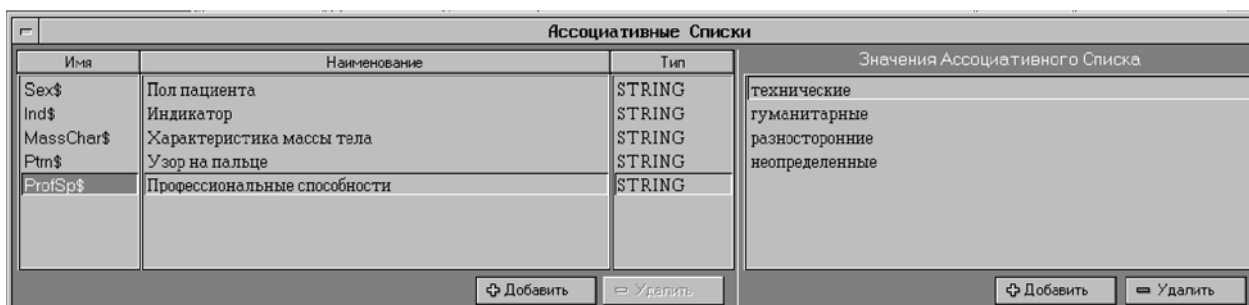


Рис. 11. Окно формирования ассоциативных списков

Имя	Наименование	Тип
RaBMI	Расчет индекса массы тела	Формула
NzChrMes	Назначение характеристики массы	Таблица
InfBody	Вывод информации о способностях	Rt-модуль
RasFntPD	Расчет фенотипа пальцевой дерматоглифики	Формула
NzDI11	Назначение дельтового индекса пальца большого левого	Таблица
NzDI12	Назначение дельтового индекса пальца указательного левого	Таблица
NzDI13	Назначение дельтового индекса среднего пальца левого	Таблица
NzDI14	Назначение дельтового индекса безымянного пальца левого	Таблица
NzDI15	Назначение дельтового индекса мизинца левого	Таблица
NzDI21	Назначение дельтового индекса большого пальца правого	Таблица
NzDI22	Назначение дельтового индекса указательного пальца правого	Таблица
NzDI23	Назначение дельтового индекса среднего пальца правого	Таблица
NzDI24	Назначение дельтового индекса безымянного пальца правого	Таблица
NzDI25	Назначение дельтового индекса мизинца правого	Таблица
NzProfSp	Назначение профессиональных способностей	Таблица

Рис. 12. Окно формирования заголовков и типа M3

МИЗ: "RasFntPD" — Расчет фенотипа пальцевой дерматоглифики
Предусловия запуска

Имя	Наименование	Тип	Условие
-----	--------------	-----	---------

Входные свойства

Имя	Наименование	Тип	Значение
DI23	Дельтовый индекс среднего пальца правого	INTEGER	
DI13	Дельтовый индекс среднего пальца левого	INTEGER	
DI24	Дельтовый индекс безымянного пальца правого	INTEGER	
DI22	Дельтовый индекс пальца указательного правого	INTEGER	
DI21	Дельтовый индекс пальца большого правого	INTEGER	
DI25	Дельтовый индекс мизинца правого	INTEGER	
DI11	Дельтовый индекс пальца большого левого	INTEGER	
DI14	Дельтовый индекс безымянного пальца левого	INTEGER	
DI12	Дельтовый индекс пальца указательного левого	INTEGER	
DI15	Дельтовый индекс мизинца левого	INTEGER	

Механизм - Формула

$$FntPD = DI11 + DI12 + DI13 + DI14 + DI15 + DI21 + DI22 + DI23 + DI24 + DI25$$

Выходные свойства

Имя	Наименование	Тип	Значение
FntPD	Фенотип пальцевой дерматоглифики	REAL	

На рисунке 13 представлено окно формирования описанного выше модуля вместе с открытым окном словаря модуля. Этот словарь представляет собой подмножество терминов из общего словаря базы знаний.

Приведенное ниже внешнее представление модуля назначения дельтового индекса пальца большого левого имеет механизм типа «Таблица». Этот модуль также не имеет предусловия. На основе входной переменной «Узор пальца большого левого» по приведенной таблице назначается дельтовый индекс этого пальца. В левом столбце таблицы указаны значения узора пальца, а в правом — соответствующие индексы.

МИЗ: "NzDI11" — Назначение дельтового индекса пальца большого левого
Предусловия запуска

Имя	Наименование	Тип	Условие
-----	--------------	-----	---------

Входные свойства

Имя	Наименование	Тип	Значение
Ptm11\$	Узор пальца большого левого	STRING	

Механизм — Таблица

Конфигурация свойств в таблице

Ptm11\$	DI11
---------	------

Таблица

A	0
AL	1
LA	1
ALW	2
L	1
LW	2
WL	2
W	2

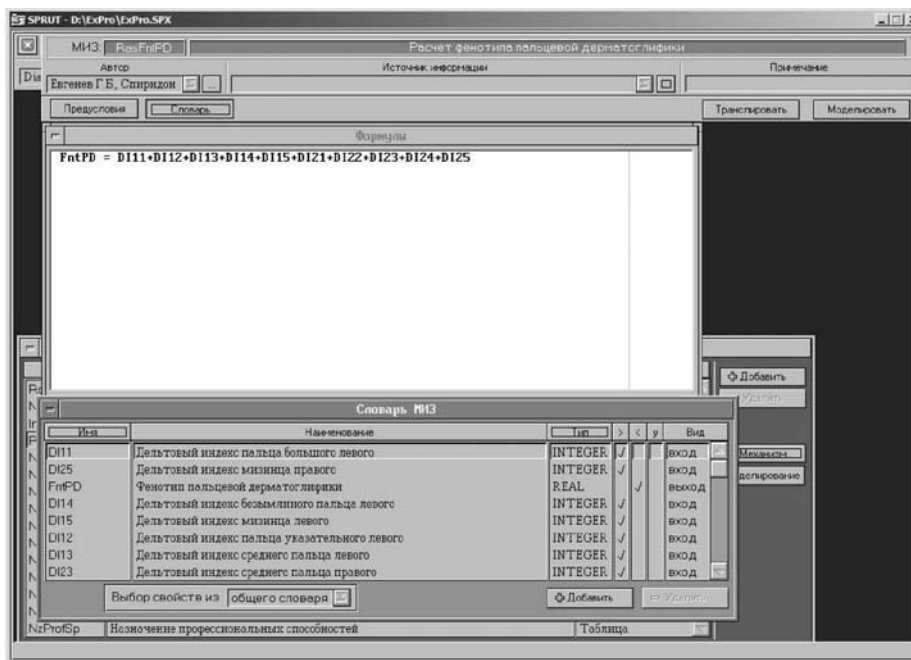


Рис. 13. Окно формирования модуля типа «Формула»

Выходные свойства

Имя	Наименование	Тип	Значение
D111	Дельтовый индекс пальца большого левого	INTEGER	

На рисунке 14 представлено окно формирования приведенного выше модуля.

Разработанные МЗ используются для генерации методов. На рисунке 15 приведено окно формирования наименований методов базы знаний.

Кнопкой «Компоновка» открывается окно формирования состава метода базы знаний (рис. 16). В левой половине открываются все наименования разработанных модулей. Из их числа кнопкой «Добавить» отбираются модули, необходимые для включения в метод.

Кнопкой «Скомпоновать метод» производится генерация метода. В левом верхнем окне (рис. 17) приводится автоматически упорядоченная последовательность модулей, обеспечивающая решение поставленной задачи. В левом нижнем окне автоматически отображаются определенные входные свойства метода, а в правом нижнем — все остальные переменные этого метода. Из их числа разработчик отмечает галочками те свойства, которые должны стать выходными. После нажатия на кнопку «Транслировать» автоматически генерируется

программа метода, которая подключается к агенту в Sprut-X (рис. 7). На рисунках 15, 16 и 17 представлен процесс генерации метода «Определение профессиональных способностей», в котором используются модули, изображенные на рис. 13 и 14.

Из приведенного описания следует, что базу знаний в Sprut-ExPro может генерировать не-программирующий эксперт.

Методы извлечения знаний из данных

Выше описаны методы и средства построения баз знаний первого типа (см. рис. 8, а), когда знания хранятся в литературных источниках и опыте человека. Другим весьма эффективным и ценным для построения медицинских интеллектуальных систем является извлечение знаний из данных (см. рис. 8, б). К этому направлению автоматической генерации баз знаний относятся методы, объединяемые под общим названием Data Mining (discovery-driven data mining) — Data Mining переводится как «добыча» или «раскопка данных». Иногда Data Mining интерпретируется более точно, как «обнаружение знаний в базах данных» (knowledge discovery in databases) и «интеллектуальный анализ данных». Возникновение этих новых терминов в 1990-х годах было связано с новым

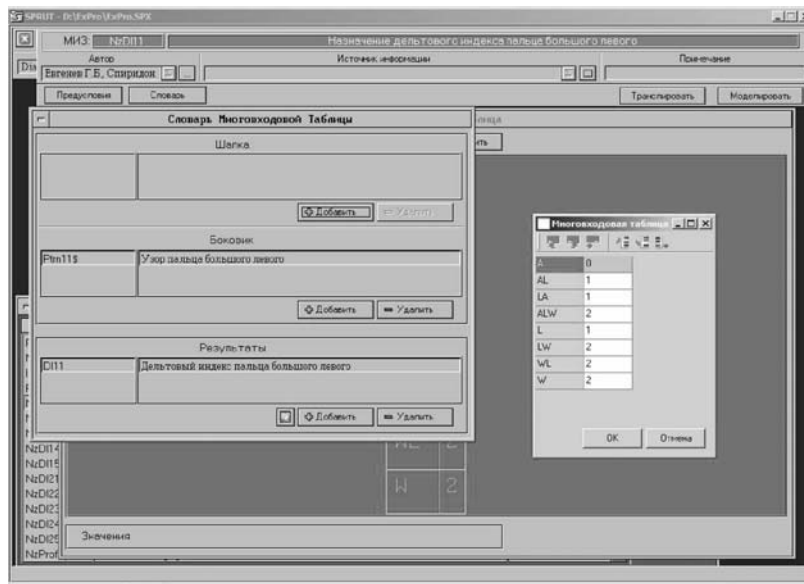


Рис. 14. Окно формирования модуля типа «Таблица»

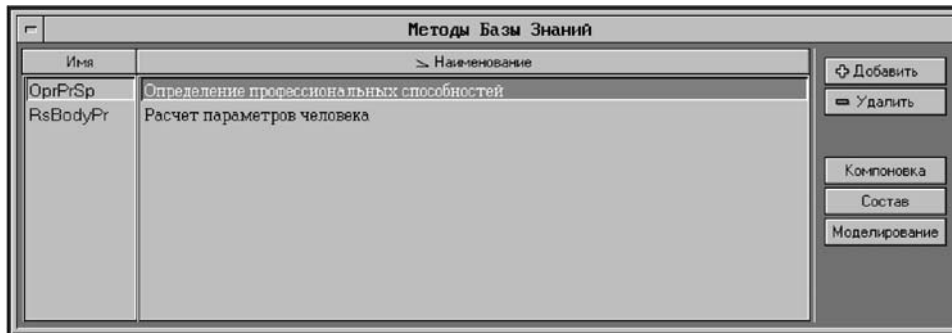


Рис. 15. Окно формирования наименований методов базы знаний

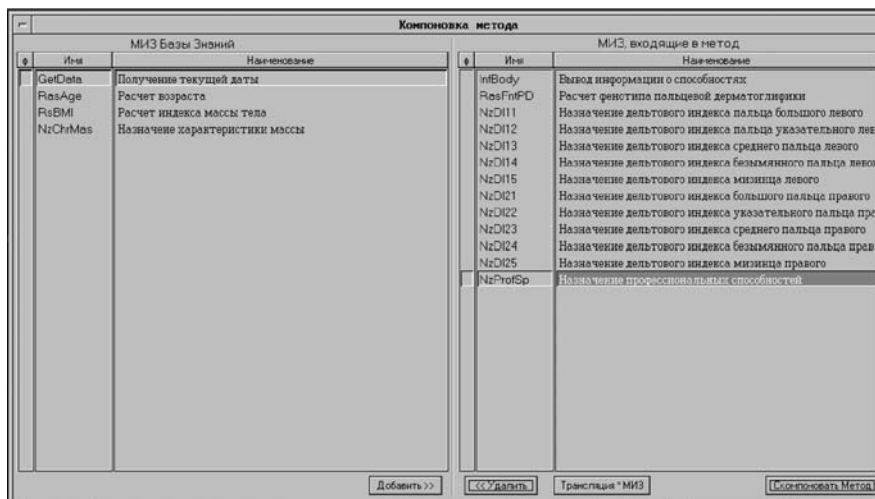


Рис. 16. Окно формирования состава метода базы знаний

витком развития средств и методов обработки данных.

В основу технологии Data Mining положена концепция шаблонов (правил), отражающих фрагменты многоаспектных взаимоотношений

данных. Эти шаблоны представляют собой закономерности, свойственные подвыборкам данных, которые могут быть компактно выражены в понятной человеку форме [4].

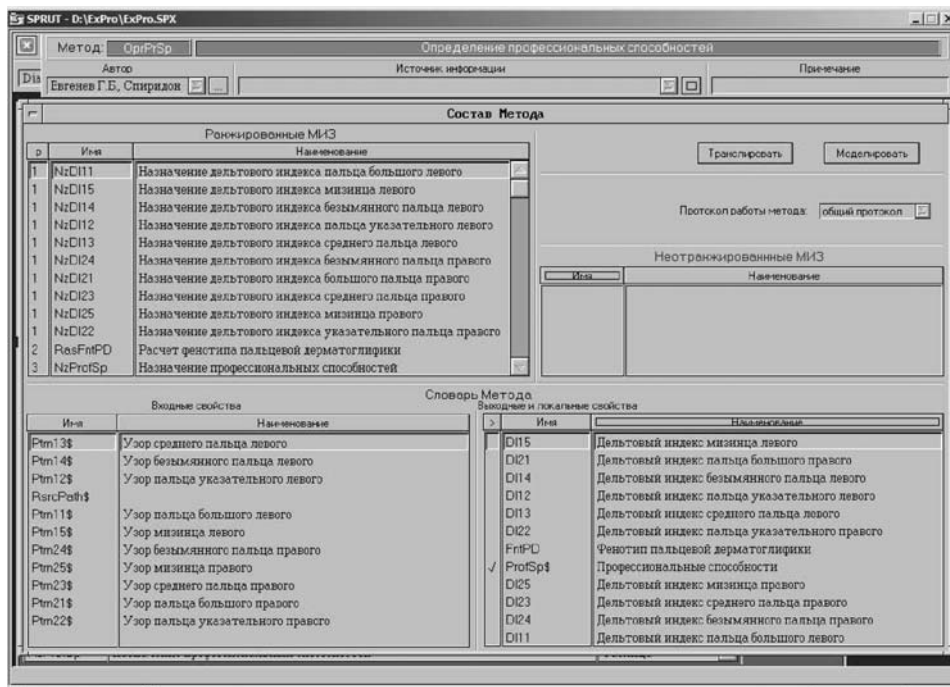


Рис. 17. Окно генерации метода «Определение профессиональных способностей»

Важное положение Data Mining — нетривиальность разыскиваемых шаблонов (правил), которые должны отражать неочевидные, неожиданные (unexpected) регулярности в данных, составляющие так называемые скрытые знания (hidden knowledge) [4].

Data Mining — это процесс обнаружения в сырых данных:

- ранее неизвестных;
- нетривиальных;
- практически полезных;
- доступных интерпретации знаний, необходимых для принятия решений в различных формах человеческой деятельности [4].

Несмотря на обилие методов Data Mining, приоритет постепенно все более смещается в сторону логических алгоритмов поиска в данных if-then правил. Результаты таких алгоритмов эффективны и легко интерпретируются [4].

Любая система, в том числе и база данных, состоит из элементов и связей между ними. Формально структуру системы можно представить в виде упорядоченной пары $S = \langle A, R \rangle$, где A — множество элементов системы, а R — множество отношений между этими элементами. Базы данных предназначены для построения компьютерных моделей сущностей. Сущность — это часть мира, выделяемая

как единое целое в течение определенного отрезка времени.

На интересующей нас сущности система S задается набором соответствующих свойств A и назначением каждому из них определенной переменной, поэтому систему всегда рассматривают не как реальную вещь, а как абстрагирование или отображение некоторых свойств сущности.

Мы рассмотрели первый компонент упорядоченной пары $S = \langle A, R \rangle$, определяющей структуру баз данных. Обратимся ко второму компоненту R , т. е. к множеству отношений между этими элементами.

Реляционный подход к построению моделей данных основывается на математическом понятии *отношение*, которое определяется следующим образом (рис. 18). Пусть даны N множеств $D1, D2, \dots, DN$, тогда R есть отношение над этими множествами, если R есть множество упорядоченных последовательностей (n -кортежей) вида $\langle d1, d2, \dots, dn \rangle$, где $d1$ — элемент из $D1$, $d2$ — элемент из $D2, \dots$ и dn — элемент из DN , то $D1, D2, \dots, DN$, называются *доменами отношения R*.

Для пояснения этих теоретических положений рассмотрим пример медицинской базы

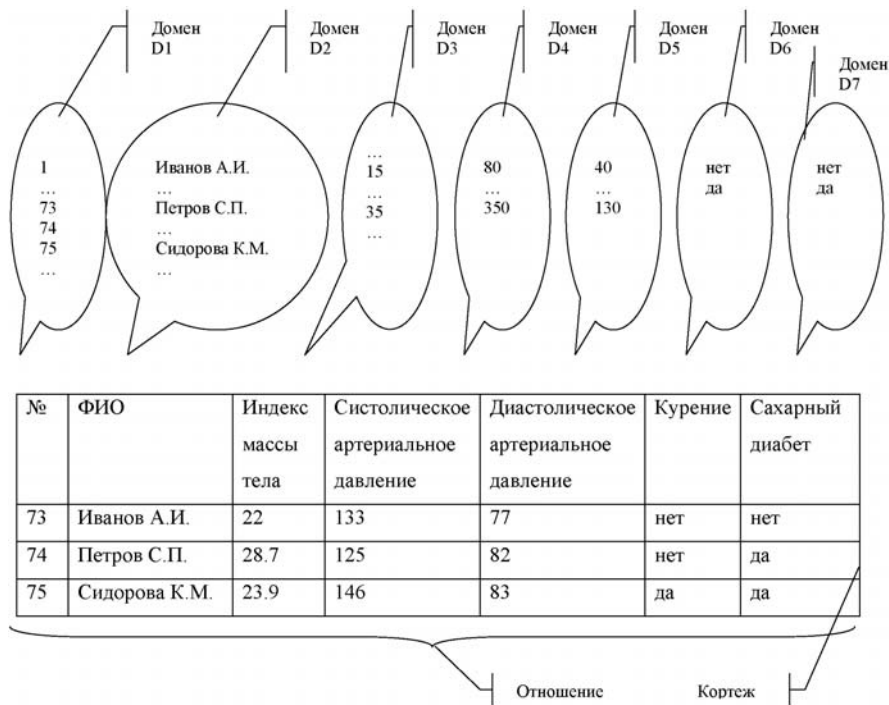


Рис. 18. Модель базы данных

данных. Эта база данных предназначена для решения задачи диагностики атеросклероза.

Применительно к данной задаче $D1=BMI$, $D2=Sys$, $D2=Dia$, ..., $D15=meanmax$ (табл. 3). Особую роль в этой задаче играет так называемая целевая переменная. При первоначальной постановке задачи роль целевой переменной играла средняя оценка экспертов $Mean_est$, которая вычислялась на основе оценок вероятности заболевания, построенных по 5-бальной шкале, данных отдельными экспертами.

Таблица 3

Имя	Наименование	Тип	Значение
BMI	Индекс массы тела	REAL	30
Sys	Систолическое артериальное давление	REAL	143
Dia	Диастолическое артериальное давление	REAL	81
Sm\$	Курение	STRING	нет
Dm\$	Сахарный диабет	STRING	да
HT_A\$	Гипертоническая болезнь с приемом лекарств	STRING	да
LVG\$	Гипертрофия левого желудочка миокарда	STRING	нет
A_CS\$	Наследственность по инфаркту миокарда	STRING	нет

A_H\$	Наследственность по гипертонической болезни	STRING	нет
A_D\$	Наследственность по сахарному диабету	STRING	да
Cho	Холестерин	REAL	230
Tg	Триглицериды	REAL	235
HDL	Альфа-холестерин	REAL	53.3
mean	Средняя толщина интимо-медиального слоя сосудистой стенки	REAL	0,76
meanmax	Максимальная толщина интимо-медиального слоя сосудистой стенки	REAL	0,87
StPrSum	Оценка степени принадлежности к отсутствию заболевания суммарная	REAL	0.4

Применительно к разработанной методике в качестве целевой переменной необходимо использовать не бальную оценку вероятности заболевания, а нормированную оценку степени принадлежности к отсутствию заболевания суммарную $D16=StPrSum$, изменяющуюся в пределах от 0 до 1. Метод расчета этой переменной будет описан ниже.

Разработанная методика основана на использовании нечетких логических переменных, *Нечеткая логическая переменная* может быть описана тройкой параметров $\langle a, X, A \rangle$, где:

a — имя нечеткой переменной;

X — универсальное множество, на котором заданы значения переменной a ;

A — нечеткое подмножество универсального множества X , для каждого элемента которого определена функция $m(x)$, задающая степень принадлежности данного элемента к множеству A .

Выше приведены компьютерные имена и общепринятые наименования нечетких переменных, необходимых для решения поставленной задачи (словарь входных переменных базы знаний). Универсальное множество, на котором заданы значения переменных, определяется их типом. Переменные действительного типа имеют нижнее и верхнее ограничения по величине.

Нечисловые (строковые) переменные в данном случае могут иметь два значения: «да» и «нет».

Для каждого элемента нечеткого подмножества необходимо определить функцию $m(x)$, задающую степень принадлежности данного элемента к множеству A . Для построения этой функции логично предположить, что числовые переменные являются случайными числами, удовлетворяющими нормальному закону распределения.

Если поставить условие, чтобы на границе отрезка, моделирующего нечеткое множество, функция принадлежности принимала значение, равное 0,03, то эта функция будет иметь вид

$$m(x) = e^{-((3.5*(x - a)/(gr - a))^2)},$$

где a — центр группирования; gr — граничное значение x (min или max).

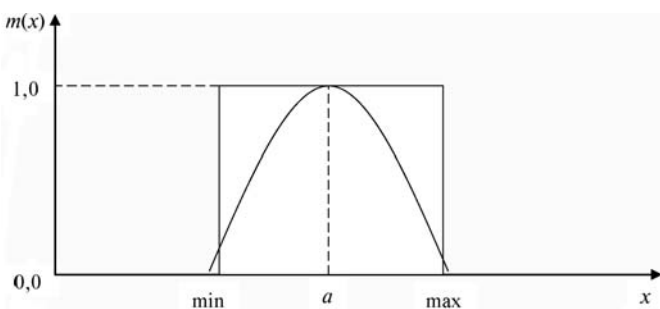


Рис. 19. Функция принадлежности свойства

Граничные значения определены при построении базы данных, а в качестве центра группирования логично принять величину, соответствующую здоровому человеку. В этом случае функция принадлежности является асимметричной и складывается из двух кривых:

- 1) для $x_{\min} \leq x < a$ $m_1(x) = e^{-((3.5*(x - a)/(x_{\min} - a))^2}$;
- 2) для $a \leq x < x_{\max}$ $m_2(x) = e^{-((3.5*(x - a)/(x_{\max} - a))^2}$.

Функции принадлежности рассчитываются для всех числовых элементов базы данных. Для нечисловых элементов они назначаются экспертами. Суммарная степень принадлежности диагноза рассчитывается по формуле (рис. 20)

$$\text{StPrSum} = K_1*m_1 + K_2*m_2 + K_3*m_3 + K_4*m_4 + \dots$$

Здесь коэффициенты K_i определяются на основе обучения системы специалистами по множеству фактических значений степени уверенности в диагнозе специалиста или консилиума специалистов StPrSum на базе значений m_i .

Степени принадлежности для нечисловых атрибутов назначаются по таблицам, в которых эти величины задаются в зависимости от значений этих атрибутов.

На основе полного набора правил, описанного выше типа, при заданных значениях переменных в табл. 3, соответствующих конкретному пациенту, определяются степени принадлежности для всех этих переменных. Затем можно рассчитать суммарную степень принадлежности, оценивающую здоровье данного пациента. записи (кортеже). Искомыми величинами являются коэффициенты k_{nm} . Номера записей могут выбираться произвольно в пределах числа имеющихся записей в базе данных.

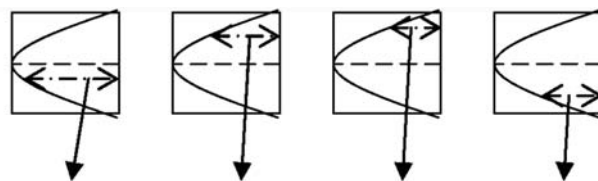


Рис. 20. Функция принадлежности диагноза

Число выбираемых записей должно равняться количеству искомых коэффициентов.

Сами искомые коэффициенты в простейшем случае могут определяться как среднее по данным N решений систем уравнений, например

$$K_{-1} = (k_{1i} + \dots + k_{1j} + \dots + k_{1k})/N.$$

В более сложном варианте может быть построен алгоритм для поиска наилучших значений коэффициентов. Наилучшие значения могут определяться как обеспечивающие минимальное среднеквадратичное отклонение от значений целевых переменных в базе данных.

Интегрированная интеллектуальная биометрическая система

На основе описанной CASE-технологии разработан демонстрационный прототип интегрированной интеллектуальной биометрической системы. Вверху экрана расположены закладки, соответствующие концептуальной модели организма человека (см. рис. 5). На первом уровне в их состав включены корневой агент «Организм человека» и входящие в него системы первого уровня, введенные в данный вариант системы: «Лицо» и «Кисти рук». Вначале выбирается закладка «Организм человека». С левой стороны располагаются кнопки

«Инспектор» и «Диагноз». С помощью кнопки «Инспектор» можно просматривать и устанавливать свойства агента, выбранного закладкой. На рисунке 21 представлено окно «Инспектор» свойств человека. Числовые свойства вводятся непосредственно, а перечисляемые символьные — с помощью меню. На рисунке 22 представлено окно системы с результатами расчета свойств человека по заданным параметрам. При этом выбран объект «Кисти рук», вследствие чего произошла автоматическая перестройка закладок — открылись закладки объектов, входящих в состав «Кисти рук»: «Кисть руки правой» и «Кисть руки левой».

Как следует из рис. 7, к объекту «Кисти рук» привязан метод определения способностей человека. При нажатии на кнопку «Диагноз» выполняется расчет с использованием описанной выше базы знаний. При заданных исходных данных получен ответ: «Профессиональные способности — технические».

Заключение

Разработана программная платформа для создания интегрированных интеллектуальных систем в биометрии. В качестве методологической основы платформы использована системология. Концептуальная биометрическая модель

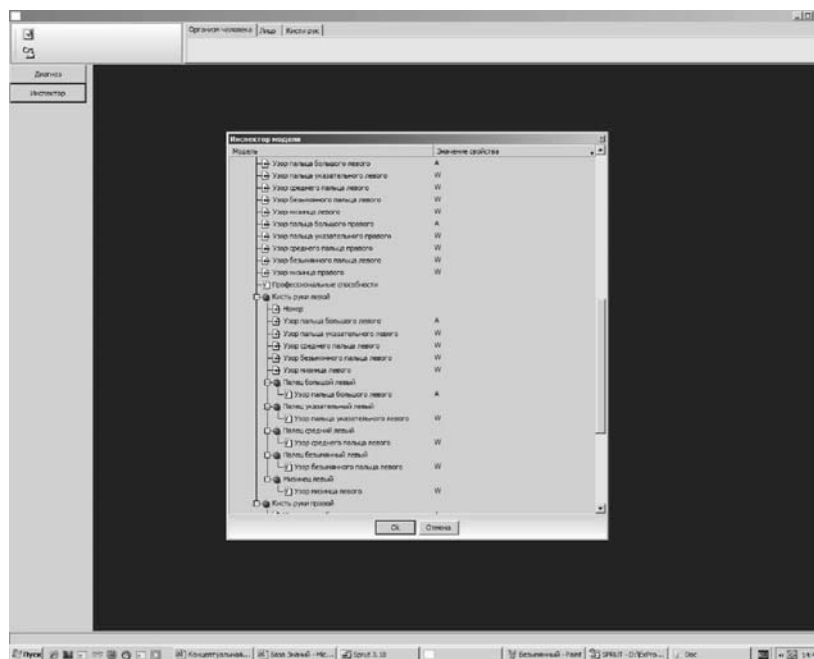


Рис. 21. Окно системы «Инспектор» свойств человека

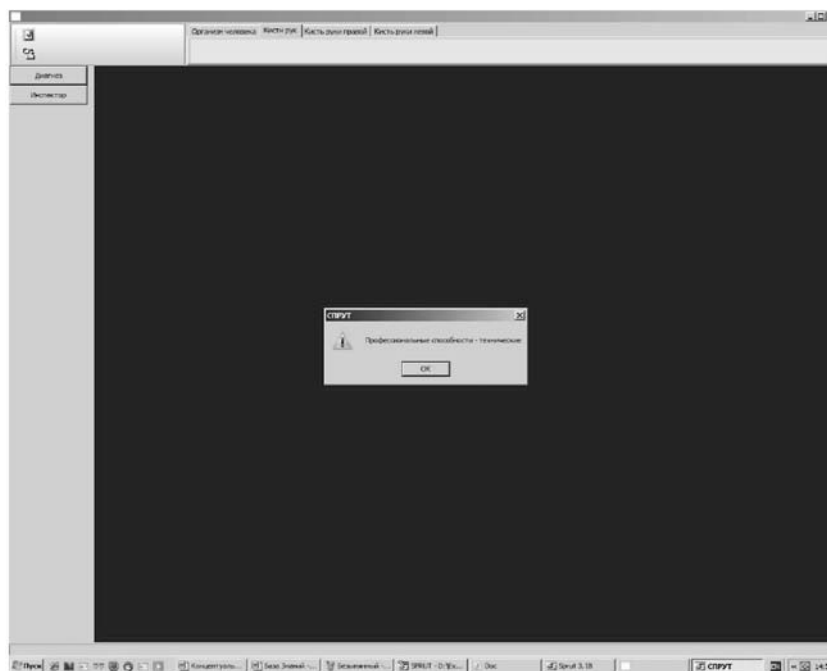


Рис. 22. Окно системы с результатами расчета свойств человека

построена с применением методов создания многоагентных систем и инструментальной среды Sprut-X. Предложены два способа создания методов агентов с использованием инструментальной среды Sprut-ExPro: полуавтоматический и автоматический. В первом способе эксперт описывает свои знания в виде продукционных правил. Второй основан на интеллектуальном анализе данных.

Построен демонстрационный прототип интегрированной интеллектуальной биометрической системы с реализацией метода определения склонностей человека. Дальнейшее развитие системы заключается в наполнении

демонстрационного прототипа новыми методами.

Литература

1. *Евгенов Г.Б.* Интеллектуальные системы проектирования. М.: Изд-во МГТУ им. Н.Э. Баумана. 2009. 320 с.
2. *Буч Г., Рамбо Дж., Джекобсон А.* Язык UML. Руководство пользователя / Пер. с англ. Слинкин А.А. М.: ДМК Пресс; СПб.: Питер, 2004. 432 с.
3. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.
4. *Дюк В.А.* Технологии Data Mining в медико-биологических исследованиях // *Новости искусственного интеллекта.* 2004. № 3. С. 49–57.

Статья поступила в редакцию 06.03.2012