

Архитектура предметно-ориентированной базы знаний интеллектуальной системы

А.С. Сигов, В.В. Нечаев, М.И. Кошкарёв

Аннотация — Рассматривается архитектура одной из компонент интеллектуальной системы – базы знаний. Целью работы является создание такой архитектуры, в которой с одной стороны объекты можно было бы классифицировать в соответствии с предметными областями. А с другой – совместить хранение этих объектов, описанных различными моделями представления знаний.

Ключевые слова — интеллектуальная система, система искусственного интеллекта, база знаний, модель представления знаний, архитектура программного средства

I. ВВЕДЕНИЕ

База знаний это один из важнейших компонентов любой интеллектуальной системы (ИС) Она выступает не только в качестве «памяти» такой системы, но и служит механизмом взаимодействия между интеллектуальной системой и её знаниями. Важнейшим этапом на пути создания ИС, как целостной системы, а базы знаний как ведущего компонента, является их архитектурное моделирование. Актуальность создания универсальной архитектуры базы знаний, ориентированной на различные предметные области для любой ИС очевидна. Такая архитектура позволит хранить данные и знания о различных предметных областях в рамках одной архитектуры ИС.

В качестве примеров таких предметных областей или сфер деятельности могут служить: обучение, научные разработки, информационные ресурсы органов государственного управления, корпоративные информационные ресурсы и т. п. Целью работы, рассматриваемой в настоящей статье, является построение архитектурной модели базы знаний ИС. Для достижения поставленной цели требуется осуществить решение следующих задач:

- определение места базы знаний в интеллектуальной системе;
- анализ существующих моделей представления знаний (МПЗ);

Статья получена 3 декабря 2014.

Исследование выполнено федеральным государственным бюджетным образовательным учреждением высшего профессионального образования "Московский государственный технический университет радиотехники, электроники и автоматики" (МГТУ МИРЭА) за счет гранта Российского научного фонда (проект №14-11-00854)

Сигов А. С., д.ф.-м.н., профессор МГТУ МИРЭА

Нечаев В. В., к.т.н, д.ф.-м.н. зав.каф. МГТУ МИРЭА (e-mail: nechaev@mirea.ru)

Кошкарёв М. И., аспирант МИРЭА (e-mail: kmi_89@mail.ru)

- выбор способов хранения информации в базе знаний;
- разработка типов связей и взаимодействий между МПЗ;
- разработка основных таблиц базы или баз данных, необходимых для хранения знаний;
- разработка модели файлового хранилища для хранения неструктурированных данных;
- разработка архитектурной модели базы знаний, включающей в себя совокупность баз данных и файлового хранилища.

Целевым результатом данной работы является архитектурная, многоуровневая модель базы знаний для хранения информации в предметно-ориентированной ИС.

II. БАЗА ЗНАНИЙ В АРХИТЕКТУРЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

База знаний является основой любой интеллектуальной системы. В первую очередь это связано с тем, что модель данных, заложенная в базу знаний, должна легко интегрироваться с моделью данных заложенных в саму систему. Чаще всего эти модели совпадают.

База знаний (БЗ) – совокупность программных средств, обеспечивающих поиск, хранение, преобразование и запись в памяти ЭВМ сложно структурированных информационных единиц (знаний)[1]. В общем виде, архитектуру интеллектуальной системы можно рассматривать как интерфейс взаимодействия пользователя с базой знаний (рисунок 1).

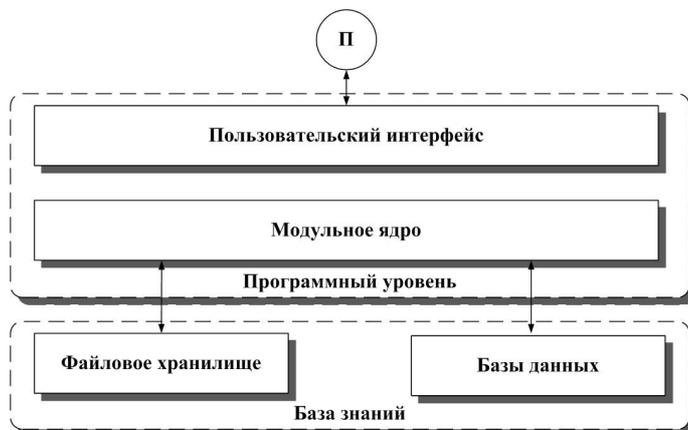


Рисунок 1 – Архитектура ИС, как интерфейс взаимодействия пользователя с базой знаний

Относительно такой модели база знаний должна решать следующие задачи:

- поиск необходимой пользователю информации (как заложенной в БЗ, так и косвенной информации, выведенной на основе существующей);
- преобразование полученной информации в модель знаний, использующуюся внутри ИС взаимодействующей с БЗ;
- своевременное обновление знаний внутри себя;
- поддержка целостности и адекватности информации.

III. КРАТКИЙ ОБЗОР СУЩЕСТВУЮЩИХ МОДЕЛЕЙ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

Согласно [2] существует три основных типа моделей представления знаний в ИС: формальные, неформальные, интегрированные, или смешанные.

Формальные модели представлены множествами сущностей (алфавит, синтаксические правила, аксиомы, правила вывода и др.). На таких моделях хорошо разработаны методы логического вывода, однако, они имеют и недостатки – отсутствие гибкости модели.

К формальным относятся такие модели, как фреймовые, семантические и продукционные. Кроме сущностей эти модели включают также и отношения между этими сущностями. Интегрированные модели содержат две или более формальных и/или неформальных моделей. Приведем краткое описание наиболее известных моделей представления знаний [2-5].

Модель представления знаний в исчислении высказываний

Исчисление высказываний является одной из самых простых формальных моделей представления знаний. Несмотря на то, что ее применение невозможно для описания сложных систем знаний. Данная модель даёт возможность понять основной механизм вывода в системах искусственного интеллекта.

Модель представления знаний в исчислении предикатов
Формирование модели предметной области с помощью метода исчисления предикатов является схожим с рассмотренным ранее методом исчисления высказываний, однако, она является более гибкой.

Продукционная модель представления знаний

Продукционная модель представления знаний построена на основе принципов построения человеческой речи. Таким образом, данная форма представления знаний является понятной и привычной для восприятия человеком, однако имеет некоторые сложности при алгоритмизации.

Продукционная модель представления знаний в общем виде соответствует следующей конструкции человеческой речи:

Если(условия), то(действие).

Онтологическая модель представления знаний

Термин “онтология”, используемый в информатике, заимствован из современной

лингвистики и является производным от древнего философского понятия “ontology”.

Онтология – это детальная формализация некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из структуры данных,

содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области.

Фреймовая модель представления знаний

Фреймовая модель представления знаний предложена Минским М.(1979), как структура знаний для восприятия пространственных сцен. Эта модель, как и семантическая сеть, имеет глубокое психологическое обоснование.

Фрейм – структура данных для представления некоторого концептуального объекта. Информация, относящаяся к фрейму, содержится в составляющих его слотах.

Принимая во внимание вышесказанное, можно предположить, что для обеспечения поддержания целостности базы знаний, содержащей несколько моделей представления знаний, необходимо создать таблицы, отражающие как сущности различных типов, так и отношения между ними.

IV. АРХИТЕКТУРА БАЗ ДАННЫХ В БАЗЕ ЗНАНИЙ ИС

В самом простом варианте система может содержать три базы данных: служебную, моделей, объектов.

В служебной базе данных хранится вся необходимая служебная информация, как о самой программе и ее модулях, так и о пользователях.

В базе данных моделей хранятся модели, описывающие различные модели представления данных, описанные в разделе II.

База данных моделей может быть представлена несколькими базами данных, однако тогда поддержку целостности между моделями необходимо будет вести на программном уровне. В данной статье база данных моделей рассматривается как единое целое.

А. Служебная база данных

Служебная база данных хранит в себе все необходимые данные о пользователях и системе. Таблицы БД сгруппированы по этому признаку (рисунок 2).

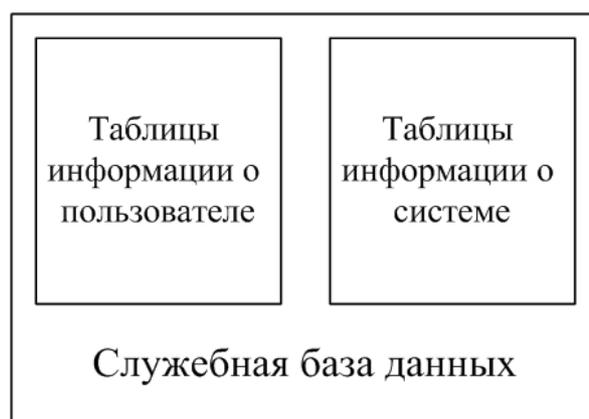


Рисунок 2 – Архитектура СБД

Служебная база данных позволяет сохранять и использовать пользовательские данные и предпочтения, а также всю необходимую информацию о состоянии системы в целом.

В. База данных моделей

Для поддержки развития и модернизации ИС необходимо продумать механизм взаимосвязи между моделями представления знаний (МПЗ).

Для этого требуется разработать механизм, уникально идентифицирующий в БЗ: сущности (объекты, связи и др.), отношения между сущностями (вхождение, расширение и др.).

Для идентификации сущностей и отношений между ними предлагается присваивать каждому типу сущностей и типу отношений между ними уникальное 32-х значное, шестнадцатеричное число вида – уникальный (УИД):

xxxx-xxxx-xxxxxxxx-xxxxxxxx-xxxx-xxxx; (1)

Тогда система сможет идентифицировать до 16^{32} сущностей каждого типа, и такое же число связей каждого типа между ними.

Такой подход позволит интегрировать модели представления знаний между собой, избегая при этом избыточности информации (рисунок 3).



Рисунок 3 – Отношения между моделями представления знаний

На рисунке 3 приняты следующие сокращения:

ИП – модель данных исчисления предикатов;

ИВ – модель данных исчисления высказываний.

Для решения рассматриваемой проблемы на уровне архитектуры модельной базы данных предлагается следующая модель архитектуры (рисунок 4).



Рисунок 4 – Архитектура БДМ

Данная модель содержит следующие группы таблиц:

- Таблицы структур и связей (основные таблицы);
- Временные таблицы (выборки из основных таблиц);
- Таблицы предметных областей;

– Дополнительные таблицы для каждой модели данных. Рассмотрим более подробно каждую из этих групп.

1) . Таблицы структур и связей

В группу данных таблиц входят следующие подгруппы таблиц: таблицы структур (сущностей), таблицы связей вхождения, таблицы связей наследования (расширения).

Таблицы структур содержат информацию о классах структур в системе. По мнению авторов, необходимо ввести в систему следующие структуры (сущности): объекты, связи, процессы, условия и свойства. Данные структуры были выбраны на основе анализа методов конфигурационного моделирования [6] и модельных представлений данных рассмотренных в п.III.

Эти структуры позволят полностью описать все основные компоненты существующих модельных представлений знаний.

Таблицы связей вхождения позволяют устанавливать связи вхождения между структурами. Это необходимо для описания сложных структур содержащих в себе подструктуры.

Таблицы связей наследования позволяют устанавливать связи наследования между структурами. Это необходимо для того чтобы избежать переизбытка информации в системе, данные о структуре с помощью этих таблиц можно представить как данные о выбранной структуре и данные взятые из родительской структуры.

2) Таблицы моделей данных

Таблицы моделей данных содержат информацию, относящуюся исключительно к конкретной модели представления знаний. Для онтологической модели это могут быть таблицы имен структур и их онтологического описания. Для фреймовой модели это таблицы связи слотов и структур и связи демонов и структур. Такая организация данных позволит логически разделить представление данных между моделями и организовать свои методы поиска для каждой модели, при этом оставляя возможность пользоваться общими методами поиска для всех моделей. Стоит отметить, что все таблицы отдельных моделей должны иметь связь с основными таблицами через ссылку на таблицы структур.

3) Таблицы предметных областей

Набор предметных областей (ПрО) представляется однонаправленным, многосвязным графом. В таком графе предметная область может иметь несколько подобластей и при этом сама являться подобластью одной или нескольких вышестоящих ПрО. Для описания такой структуры необходимо минимум две таблицы: таблица описания предметных областей, таблица связей между предметными областями. Такой подход позволяет создавать иерархии предметных областей со множественным наследованием (рисунок 5)

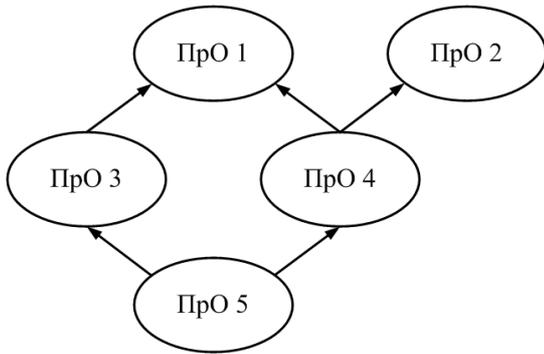


Рисунок 5 – Пример иерархии предметных областей

4) *Временные таблицы*

При инициализации системы предлагается создавать временные таблицы предметных областей наиболее часто применяемых пользователем. В данных таблицах будут находиться только структуры и связи этих предметных областей. Т.е., если база данных содержит таблицы {Str_1, ..., Str_i, ..., Str_n} и {ER_Str_1_Str_2, ..., ER_Str_m_Str_n, IR_Str_1_Str_2, ..., IR_Str_m_Str_n}, то для каждой из этих таблиц при инициализации системы формируются таблицы вида tmp_tablename, которые заполняются запросами вида INSERT INTO tmp_tablename values (SELECT * FROM tablename WHERE «Условие»). Заполнение таблиц подобным образом имеет хорошие временные показатели. При этом, подобного вида таблицы позволяют уменьшить область поиска до уровня наиболее вероятного поиска, т.е. наиболее часто необходимых предметных областей. В случае отсутствия информации во временных таблицах поиск информации будет протекать в стандартном режиме.

С. *База данных объектов*

База данных объектов содержит информацию об объектах известной системе. Данные объекты должны быть классифицированы в соответствии с модельной базой данных и содержать таблицы объектов классов, описанных в ней. При этом во избежание дублирования информации и усложнения классификации объектов, таблицы классов-родителей должны содержать информацию об объектах классов-детей, относящуюся к ним. В соответствии с изложенным выше целесообразно рассмотреть следующую фреймовую модель данных (рисунок 6).

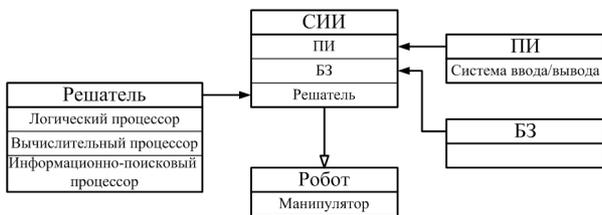


Рисунок 6 – Структуры, представленные фреймовой моделью данных

Для описания объектов такой модели представления данных в базе данных объектов создаются следующие таблицы.

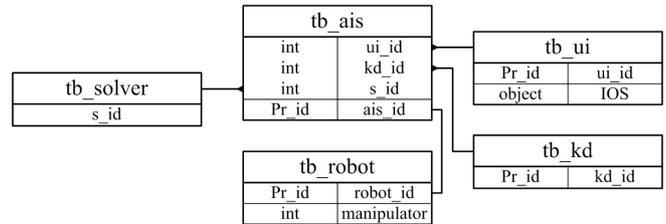


Рисунок 7 – Физическая модель БД, соответствующая фреймовой модели

Тем самым таблицы в базе данных объектов представляют собой совокупность реляционной и объектно-ориентированной моделей представления данных.

Такой подход приводит к тому, что таблицы в базе данных объектов представляют собой совокупность реляционной и объектно-ориентированной моделей представления данных. Дополнительную информацию об объектах, также как и информацию о классах, предлагается хранить в отдельном файловом хранилище. Такой информацией могут быть текстовые файлы, аудио- и видеозаписи.

D. *Общая архитектурная модель базы знаний интеллектуальной системы*

На основе проведенного выше анализа предлагается следующая архитектурная модель базы знаний интеллектуальной системы (рисунок 8).



Рисунок 8 – Архитектурная модель базы знаний ИС.

Таким образом, в настоящей работе подробно рассмотрены таблицы баз данных, входящих в представленную на рисунке 8 систему и предложены варианты оптимизации поиска по запросам на основе архитектурной модели базы знаний; оптимизация размещения файлов файлового хранилища (ФХ) базы знаний, а также предложено идентифицировать каждую сущность и связь между ними с помощью уникального идентификатора - УИДа.

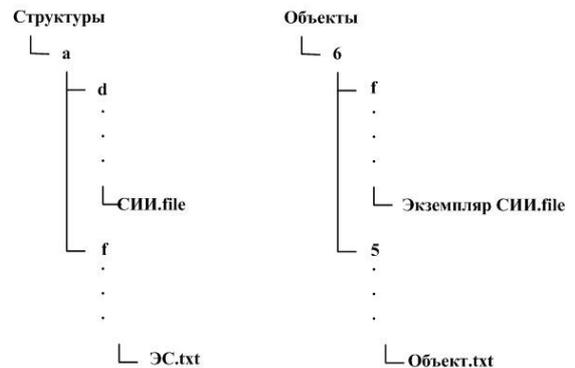


Рисунок 9 – Структура расположения файлов в файловом хранилище

Architecture subject-oriented knowledge base of intelligent system

A.S. Sigov, V.V. Nechaev, M.I. Koshkarev

Abstract – In this article Architecture of one of the components of intelligent system - a knowledge base. The aim is to create such an architecture in which on the one hand the objects could be classified according to subject areas. And on the other side it combines the storage of these objects described by the different models of knowledge representation.

Keywords – an intelligent system, the system of artificial intelligence, knowledge base, knowledge representation model, software architecture