

# ТЕХНОЛОГИЯ ПРОЕКТИРОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

С.А. Самодумкин

Белорусский государственный университет информатики и радиоэлектроники,  
кафедра интеллектуальных информационных технологий

П.Бровки, 6, г. Минск, Республика Беларусь  
+375-017-293-80-92; e-mail: Samodumkin@bsuir.by

Предлагается класс интеллектуальных систем – интеллектуальные вопросно-ответные системы, построенные с использованием универсального семантического кода. Рассматриваются принципы и технологии проектирования данного класса систем.

Ключевые слова – интеллектуальная вопросно-ответная система, база знаний, технология проектирования.

## 1 ВОПРОСНО-ОТВЕТНЫЕ СИСТЕМЫ И ИХ ИНТЕЛЛЕКТУАЛИЗАЦИЯ

В настоящее время накоплены большие объемы информации, которые хранятся в компьютерных системах. Для эффективного использования накопленной информации актуальна задача развития технологий, направленных на проектирование и разработку систем информационного поиска и справочных систем. Такие системы выдают пользователю информацию, удовлетворяющую его информационной потребности [1]. Наиболее представительна на сегодняшний момент класс поисковых машин в среде Internet и справочных систем. Однако при видимой пользе таких систем обозначился и ряд проблем. Большинство поисковых машин (Google, Rambler, Yandex и др.), информационно-поисковых и справочных систем базируется на поиске релевантной информации на основе совпадения текстовых фрагментов (на уровне отдельных слов или предложений). В результате на запрос пользователя такие системы выдают большое количество ссылок, многие из которых не имеют прямого отношения к нужной теме. Как итог поиск необходимой информации отнимает все больше и больше времени в связи с растущими объемами информации, а поиск и использование нужной информации становится все более сложным, трудоемким и неэффективным [2].

Предлагаемые подходы в оптимизации информационного поиска в настоящее время лежат в плоскости разработки вопросно-ответных систем (ВОС) в которых осуществляется сопоставление вопросов пользователей с требуемой информацией. В таких системах информационная потребность пользователя формируется в виде специального запроса – вопроса, а результатом поиска является ответ, т.е. выдача компьютерной системой информации пользователю, удовлетворяющей поставленному вопросу. Концептуальной основой для формализа-

ции вопросов в ВОС является язык вопросов и эротетическая логика [3,4], что позволяет задавать вопросно-ответные отношения. Известные на сегодняшний момент ВОС способны с помощью анализатора естественного языка осуществить разбор вопроса и сопоставить ответ. Ограничением таких систем является выявление семантических отношений между объектами предметной области только в проиндексированных текстах, что накладывает следующие ограничения:

- нет возможности строго формально установить семантические отношения между объектами в тексте по причине сложности естественного языка;
- невозможно сгенерировать ответ пользователю, когда такого ответа нет в текущем информационном состоянии системы; не поддерживаются вопросы на выявление соответствий и аналогий между объектами и понятиями.

Устранение перечисленных ограничений требует создания следующего поколения ВОС – интеллектуальных вопросно-ответных систем (ИВОС). В таких системах акцент смещается с текстового представления информации на формирование и использование пространств знаний [2], что делает данный класс систем инвариантным от языка, на котором формируются вопросы.

В качестве решения предлагается использовать универсальный семантический код (Semantic Code – SC) [5] для представления знаний о предметной области, формирования вопросов и поиска или генерации ответов. Указанные обстоятельства позволяют:

- 1) формировать базы знаний о предметной области, используя универсальный семантический код;
- 2) конструировать формальные представления вопросов, используя ту же формальную модель представления знаний;
- 3) обрабатывать информационные предписания пользователя программами, ориентированными на обработку баз знаний.

Таким образом, на единой концептуальной основе формализуются знания и умения ИВОС. Интеллектуальные вопросно-ответные системы, в которых для представления знаний используется SC-код, названы вопросно-ответными sc-системами.

Целью данной работы является создание технологии проектирования интеллектуальных вопросно-ответных систем в основе которой лежат принцип компонентного проектирования такого класса систем и ориентация на широкий круг разработчиков.

## 2 КОМПОНЕНТНОЕ ПРОЕКТИРОВАНИЕ ВОПРОСНО-ОТВЕТНЫХ SC-СИСТЕМЫ

Модель вопросно-ответной sc-системы включает три компонента: базу знаний, машину обработки баз знаний и пользовательский интерфейс (рис. 1)

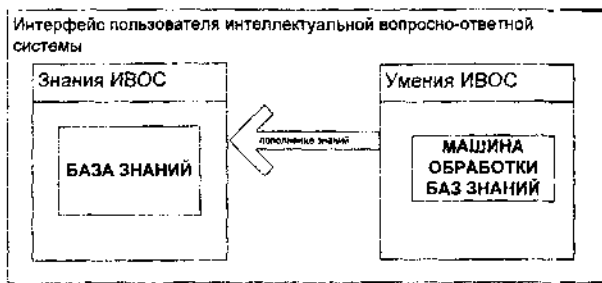


Рис. 1. Компоненты вопросно-ответных sc-систем

В базе знаний накапливаются знания ИВОС, которые формализуются разработчиком в соответствии с семантической технологией проектирования интеллектуальных систем и пополняются в процессе эксплуатации системы. Навигация по базе знаний осуществляется с помощью базовых информационно-поисковых операций машины обработки знаний, а информационная потребность пользователя формируется в виде вопросов.

Функциональные возможности интеллектуальной вопросно-ответной системы определяются теми умениями системы, которые реализуются операциями машины обработки знаний. Такие операции инициируются системой в случаях, когда в текущем информационном состоянии системы (т.е. в базе знаний) отсутствует ответ на поставленный вопрос, что позволяет сгенерировать ответ пользователю и пополнить базу знаний ИВОС новыми знаниями.

Пользовательский интерфейс управляет диалогом конечного пользователя с системой при запросах (т.е. формировании вопроса) к системе или ответах. В качестве базового способа визуализации вопросов пользователей и соответствующих им ответов выступает графическое представление семантического кода – SCg-код (Semantic Code Graphic). Для визуализации ответов пользователю в ином виде требуется разработка специализированных пользовательских интерфейсов. В зависимости от характера предметной области вопросно-ответной sc-системы проектируются картографический, естественно-языковой, речевой и другие интерфейсы.

Таким образом, состояние вопросно-ответной sc-системы определяется текущим состоянием базы знаний и операциями машины обработки баз знаний, которые осуществляют поиск ответов в базе знаний или генерацию ответов, в случае отсутствия ответов в базе знаний ИВОС.

## 3 ТЕХНОЛОГИЯ ПРОЕКТИРОВАНИЯ ВОПРОСНО-ОТВЕТНЫХ SC-СИСТЕМЫ

В технологию проектирования вопросно-ответных sc-систем включается:

1. Модель или теория вопросно-ответных sc-систем, т.е. формальное описание такого класса систем.

2. Библиотека ip-компонентов (intellectual property) вопросно-ответных sc-систем.

3. Средства автоматизации проектирования вопросно-ответных sc-систем.

4. Методика проектирования вопросно-ответных sc-систем.

5. Методика обучения проектированию вопросно-ответных sc-систем.

6. Интеллектуальная help-система.

**Модель вопросно-ответных sc-систем.** В основе модели вопросно-ответных sc-систем лежит язык вопросов. Данный язык принадлежит семейству sc-языков, на нем формируется информационная потребность пользователей данного класса систем. Для рассматриваемого класса систем *вопрос – это информационная цель*. Объектами анализа языка вопросов являются классы вопросов в соответствии с семантической типологией вопросов.

Можно выделить следующие классы вопросов.

**Класс предикатных вопросов**, т.е. вопросов, конструируемых с помощью образа (шаблона) поиска. Ответом для такого класса вопросов является вывод с точностью до изоморфизма фрагментов базы знаний ИВОС. Ответ может выводиться в общем случае в развернутом виде или иметь краткую форму вывода.

Приведем пример предикатного вопроса Q1: «Какие отношения могут связывать между собой прямые?». На рис.2. приведена формальная запись вопроса Q1 с требованием краткого ответа, а рис.3 демонстрирует ответ на поставленный вопрос, где в качестве результата указано, что прямые могут быть параллельными, пересекающимися и скрещивающимися.

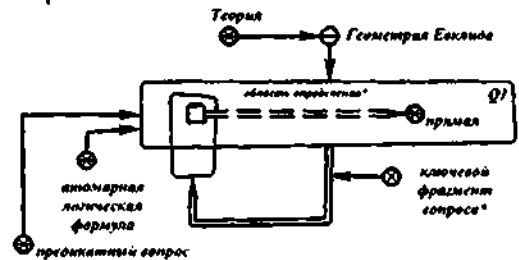


Рис.2. Формальная запись вопроса Q1

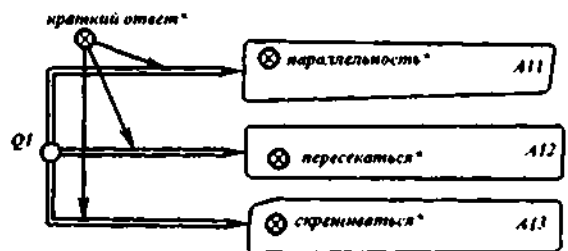


Рис.3. Результат ответа на вопрос Q1 в краткой форме

Предикатный вопрос может и не быть атомарной формулой. Вопрос Q2: «Приведите пример геометрической фигуры G<sub>i</sub>, удовлетворяющей следующим свойствам: существуют две точки t<sub>1</sub> и t<sub>2</sub> принадлежащие фигуре G<sub>i</sub> такие, что 1) каждая точка, принадлежащая фигуре G<sub>i</sub>, лежит между точками t<sub>1</sub> и t<sub>2</sub>; 2) каждая точка, лежащая между точками t<sub>1</sub> и t<sub>2</sub>, принадлежит фигуре

$G_i$  — является предикатным, но не может быть задан в виде атомарной формулы. На рис. 4. приведена формальная запись вопроса Q2. Ответом на этот вопрос являются конкретные геометрические фигуры, являющиеся отрезками (поскольку свойства запрашиваемых геометрических фигур в точности соответствуют определению отрезка).

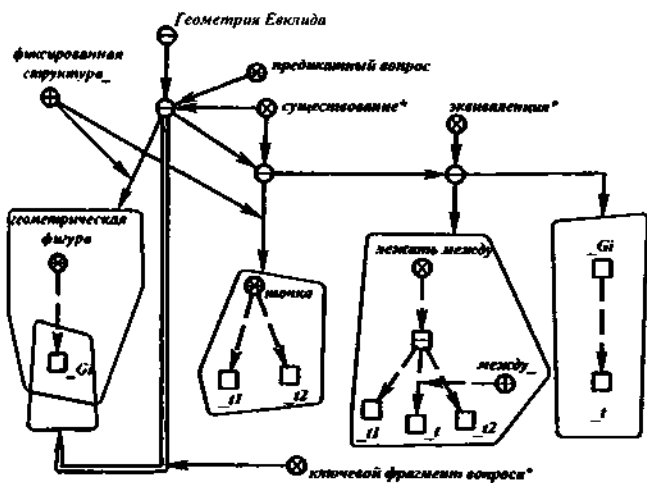


Рис. 4. Формальная запись вопроса Q2

Класс вопросов, требующих специализированной процедуры поиска в зависимости от семантического класса вопроса. Для данного класса вопросов ответ имеет сложную процедуру локализации, однако ответ явно имеется в текущем состоянии системы. В данный класс вопросов входят следующие запросы:

- запрос типа «Что это такое?», который запрашивает информацию об объекте (понятии);
- запрос типа «Как связаны?», устанавливающий связи между объектами (понятиями);
- запрос на сопоставление двух объектов (понятий), устанавливающий сходство и различия между объектами или понятиями;
- запрос типа «Почему справедливо данное высказывание?»;
- запросы типа «Где?», «Когда?» на установление обстоятельств положения и времени;
- запрос нахождения способа решения «С помощью каких формул можно найти решение?».

Приведем этапы поисковой операции, дающей ответ на вопрос «В чем аналогия между двумя понятиями?».

Шаг 1. Найти в базе знаний связки тех отношений, в которые входят сравниваемые понятия.

Шаг 2. Выделить пары связок одинаковых отношений сравниваемых объектов. Если таких пар связок не найдено, то аналогии между понятиями не установлено.

Шаг 3. Для каждой выделенной пары связок одинаковых отношений при условии наличия пересечения множеств, являющихся вторыми компонентами связок, добавить результат пересечения и пару связок в формируемый ответ.

На рис. 5. приведена формальная запись вопроса Q3: «С помощью каких формул можно вычислить площадь треугольника?»

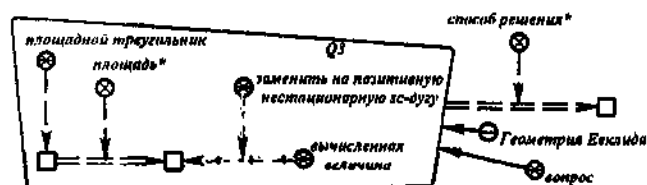


Рис. 5. Формальная запись вопроса Q3

Класс вопросов, требующих формирования числовых значений в соответствии с исходными данными. Для выдачи ответа сначала осуществляется информационный поиск в базе знаний. При его успешности выводится ответ пользователю, иначе система находит способ решения и на основе заложенных в нее умений (операциями машины обработки баз знаний) решает задачу, выдает ответ пользователю и полученный результат помещает в базу знаний, тем самым пополняя базу знаний новыми знаниями. В качестве примера приведем вопрос Q4 (рис. 6.): «Чему равна площадь треугольника, если его стороны равны 3, 4 и 5 соответственно?»

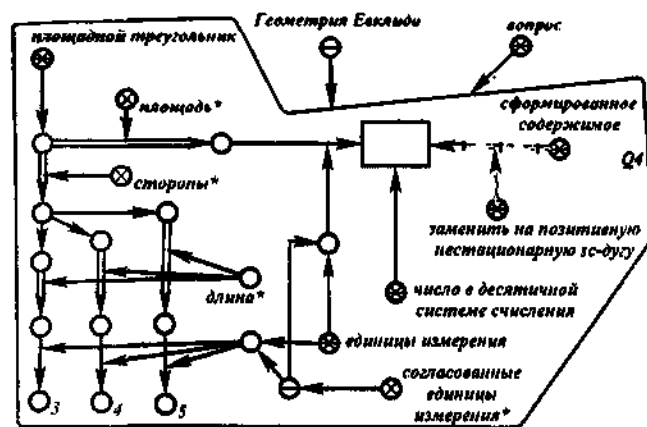


Рис. 6. Формальная запись вопроса Q4

Отметим, что в формулировке вопроса присутствует декларативная и императивная часть вопроса. Декларативная часть вопроса включает контекст вопроса, а императивная — требование вопроса.

Библиотека ip-компонентов вопросно-ответных sc-систем. Важным аспектом предлагаемой технологии является открытый код проектируемых систем. Данный факт позволяет при проектировании прикладных систем использовать уже готовые фрагменты ИВОС или их части. Это позволяет, во-первых, увеличивать функциональные возможности и по знаниям системы, и по ее умениям; во-вторых, интегрировать разработанные системы для смежных предметных областей (например, система по геометрии может состоять из двух систем — ИВОС по планиметрии и стереометрии и т.д.). ИВОС, спроектированные в соответствии с данной технологией и оформленные по соответствующим правилам помещаются в

библиотеку и доступны разработчикам ИВОС. Кроме того, в библиотеку входит онтология языка вопросов.

Средства автоматизации проектирования вопросно-ответных *vs*-систем включают инструментарий по сборке ИВОС.

Методика проектирования вопросно-ответных *vs*-систем.

1. *Технико-экономическое обоснование проектируемой ИВОС.* Обосновывается актуальность разработки ИВОС по заданной предметной области; создается портрет пользователя системы (указание категории пользователей системы; функциональных возможностей, предоставляемых системой конкретной категории пользователей); рассматриваются аналоги систем, решающие проблемы в заданной предметной области, их сравнение с разрабатываемой системой по различным критериям (функциональные возможности, многообразие видов знаний, обработка знаний и т.д.); указываются преимущества (достоинства) разрабатываемой прикладной системы по отношению к аналогам; оцениваются трудозатраты по реализации (изготовлению) системы.

2. *Тестовый список вопросов и задач, которые должны решаться в проектируемой ИВОС.* Приводится полный тестовый список вопросов, на которые система может дать ответ, перечень задач, которые система сможет решить, а также необходимые в системе операции и вид пользовательского интерфейса (как осуществляется диалог с пользователем в системе).

3. *Демонстрационный список вопросов и задач, решаемых ИВОС.* Устанавливается многообразие видов используемых знаний: факты, связи, закономерности, причинно-следственные связи и т.д.; разнообразие типов вопросов: навигационно-поисковые вопросы, вопросы на связь между понятиями и т.д. Данный список должен продемонстрировать использование универсальных логических методов обработки информации, например, сведения задач к подзадачам).

4. *Содержательная структура проектируемой базы знаний ИВОС.* Первый прототип ИВОС.

5. *Интеграция первого прототипа проектируемой базы знаний с онтологиями используемых языков* – получение второго прототипа проектируемой базы знаний ИВОС. Список «интересных» дополнительных вопросов, на которые система получила возможность отвечать.

6. *Интеграция второго прототипа проектируемой базы знаний с онтологией предметной области* – получение третьего прототипа проектируемой базы знаний ИВОС. Список «интересных» дополнительных вопросов, на которые система получила возможность отвечать.

7. *Наполнение базы знаний* – получение четвертого прототипа проектируемой базы знаний ИВОС. Список «интересных» дополнительных вопросов, на которые система получила возможность отвечать.

8. *Проектирование операций машин обработки баз знаний (формирование умений проектируемой ИВОС)* – получение пятого прототипа проектируемой ИВОС. Список «интересных» дополнительных вопросов, на которые система получила возможность отвечать.

9. *Проектирование интеллектуального пользовательского интерфейса* – получение шестого прототипа проектируемой ИВОС.

Методика обучения проектированию вопросно-ответных *vs*-систем. Включение данной методики в состав технологии позволяет в процессе проектирования ИВОС повысить квалификационные требования разработчиков прикладных ИВОС.

Интеллектуальная *help*-система прикладной ИВОС. Данная система обеспечивает помощь пользователю в процессе эксплуатации ИВОС.

#### 4 АПРОБАЦИЯ

Предложенная в работе технология апробирована в учебном процессе кафедры интеллектуальных информационных технологий Белорусского государственного университета информатики и радиоэлектроники. На протяжении трех семестров студенты специальности «Искусственный интеллект» разрабатывают прототип ИВОС по выбранной предметной области. На первом этапе ими создается прототип системы, в котором ответы возможны только на вопросы, если ответы на них явно присутствуют в базе знаний ИВОС. На втором этапе появляется возможность генерировать ответ, используя умения системы, а на третьем этапе – общение пользователя с помощью внешних интерфейсов (естественно-языковой, картографический и др.) в зависимости от предметной области. Прототипы разработанных ИВОС совершенствуют по указанным в технологии компонентам студенты последующего года набора, тем самым дополняя функциональные возможности уже разработанных прикладных вопросно-ответных *vs*-систем.

Работа выполнена при поддержке БРФФИ – РФФИ (грант № Ф08Р-137).

#### ЛИТЕРАТУРА

- [1] Захаров, В. П. Информационные системы (документальный поиск). / В. П. Захаров. – СПб. : Изд-во СПбГУ, 2002. – 188 с.
- [2] Хорошевский, В.Ф. Пространства знаний в сети Интернет и Semantic Web (Часть 1) / В. Ф. Хорошевский // Искусственный интеллект и принятие решений. – 2008. – № 1. – С.80-97.
- [3] Белнап, Н. Логика вопросов и ответов / Н. Белнап, С. Стил. – М., 1981.
- [4] Финн, В.К. К формальному определению информационно-поисковой системы / В. К. Финн // Научно-техническая информация. – Сер. 2. – 1981. – № 5.
- [5] Представление и обработка знаний в графодинамических ассоциативных машинах / В.В. Голенков [и др.]; под ред. В.В. Голенкова. – Минск : БГУИР, 2001. – 412 с.