



УДК 004.822:514

### ТЕХНОЛОГИЯ КОМПОНЕНТНОГО ПРОЕКТИРОВАНИЯ БАЗ ЗНАНИЙ НА ОСНОВЕ УНИФИЦИРОВАННЫХ СЕМАНТИЧЕСКИХ СЕТЕЙ

Давыденко И.Т.

*Белорусский государственный университет информатики и радиоэлектроники,  
г. Минск, Республика Беларусь*

**ir.davydenko@gmail.com**

В работе рассматривается состав семантической технологии проектирования баз знаний интеллектуальных систем. Данная технология ориентирована на семантическое представление знаний, расширение контингента разработчиков баз знаний и сокращение сроков проектирования.

**Ключевые слова:** база знаний, интеллектуальная система, онтология, технология проектирования.

#### ВВЕДЕНИЕ

В настоящее время все более актуальной становится задача эффективного информационного обеспечения научной и производственной деятельности, связанная с бурным ростом объемов информации в различных отраслях знаний. Данная задача, как правило, рассматривается в контексте создания хранилищ знаний и их систематизации и структуризации с целью облегчения их обработки.

База знаний является одним из ключевых компонентов интеллектуальных систем различного назначения [Гаврилова и др., 2001], [Гаврилова, 2008], [Хорошевский, 2008]. Разработка этого компонента является трудоемким и продолжительным процессом.

При разработке баз знаний важно обеспечить не только возможность хранения знаний и навигации по ней, но и возможность работы над созданием и изменением базы знаний распределенным коллективом разработчиков.

На сегодняшний день существует ряд проблем, в области формирования баз знаний [Ефименко, 2011]:

- доступность семантического контента;
- доступность баз знаний и средств их разработки;
- эволюция баз знаний;
- масштабируемость баз знаний;
- мультязычность баз знаний;
- стабильность баз знаний;
- визуализация баз знаний.

Эти проблемы обусловлены следующими недостатками существующих технологий проектирования баз знаний:

- сложность языков представления знаний;
- неоднородность и ограниченность моделей представления знаний;
- ограниченность или отсутствие средств верификации и отладки баз знаний;
- плохая отчуждаемость баз знаний в силу привязки к инструментарию;
- ограниченность или отсутствие возможности интеграции баз знаний, что влечет за собой ограничения на расширение базы знаний, отсутствие общих стандартов совместимости уже разработанных фрагментов баз знаний [Ивашенко, 2012].

Все эти недостатки порождают еще одну проблему – недостаточное количество квалифицированных инженеров баз знаний, которая возникает из-за высоких требований к разработчикам.

В качестве решения вышеуказанных проблем предлагается технология компонентного проектирования баз знаний, основанная на унифицированных семантических сетях с базовой теоретико-множественной интерпретацией. Данная технология представляет собой комплекс моделей, инструментальных средств и методов проектирования баз знаний.

В основе предлагаемой технологии лежат следующие основные принципы массовой семантической технологии проектирования интеллектуальных систем OSTIS (Open Semantic Technology for Intelligent Systems) [OSTIS, 2012]:

- поэтапное эволюционное проектирование баз знаний на основе быстрого прототипирования;
- ориентация на коллективное проектирование баз знаний в рамках Open Source проекта;
- ориентация на семантическое представление знаний;
- унификация моделей баз знаний интеллектуальных систем;
- модульное проектирование на основе библиотек типовых многократно используемых компонентов.

Технология проектирования баз знаний представляет собой комплекс моделей, инструментальных средств и методов проектирования баз знаний.

Предлагаемая технология имеет следующую структуру:

- унифицированная семантическая модель представления знаний
- семантическая модель базы знаний;
- библиотека многократно используемых компонентов баз знаний;
- средства проектирования семантических моделей баз знаний;
- методика проектирования семантических моделей баз знаний.

С формальной точки зрения, технология проектирования материализуется в виде интеллектуальной метасистемы для поддержки проектирования баз знаний.

## 1. Унифицированная семантическая модель представления знаний

В качестве формальной основы проектируемых логико-семантических моделей баз знаний интеллектуальных систем используются графодинамические модели специального вида – семантические модели представления и обработки знаний, в основе которых лежат унифицированные семантические сети с базовой теоретико-множественной интерпретацией. Основным способом кодирования информации для таких сетей является SC-код (Semantic Code) [Голенков, 2012].

Все элементы (атомарные фрагменты) семантической сети являются знаками различных сущностей. Такими сущностями могут быть всевозможные внешние описываемые объекты, а также различные множества, состоящие их элементов (атомарных фрагментов) этой же семантической сети.

База знаний интеллектуальной системы, представленная в виде корректно построенной семантической сети, полностью исключает дублирование информации в рамках такой базы знаний.

Представление знаний в виде унифицированных семантических сетей позволяет существенно упростить процедуру ассоциативного доступа к различным видам фрагментов хранимой базы

знаний, а также существенно расширить типологию запросов к базе знаний [Голенков, 2012].

## 2. Семантическая модель базы знаний

В основе разработки баз знаний с помощью технологии OSTIS лежит четкое разделение процесса проектирования формального описания семантической модели разрабатываемой базы знаний от процесса реализации (интерпретации) этой модели на той или иной платформе [Голенков, 2012]. Данный факт позволяет обеспечить кросс-платформенную разработку интеллектуальных систем.

Всю семантическую сеть (максимальную семантическую сеть), хранимую в семантической памяти абстрактной логико-семантической модели интеллектуальной системы, будем называть **абстрактной семантической моделью базы знаний** этой интеллектуальной системы.

Семантическая модель базы знаний интеллектуальной системы представляет собой формальную трактовку семантического пространства, которое известно интеллектуальной системе в текущий момент времени.

Эффективность интеллектуальной системы в первую очередь определяется объемом и качеством содержащихся в них формализованных экспертных знаний, как декларативных (теоретических), так и процедурных (практических навыков).

База знаний должна содержать в себе всю информацию, необходимую агентам, работающим над семантической памятью, для организации коллективной деятельности по решению задач, с которыми должна справляться интеллектуальная система.

Для расширения разнообразия видов знаний, хранимых в базе, необходимым этапом в разработке семантической модели базы знаний является ее структуризация.

Понятие базы знаний тесно связано с понятием предметной области. Соотношение между базой знаний и описываемой ею предметной областью задает семантику базы знаний интеллектуальной системы.

Рассмотрение структуры базы знаний во взаимосвязи с предметной областью позволяет рассматривать исследуемые объекты на разных уровнях детализации. Детализацию рассмотрения исследуемых объектов можно осуществлять как в рамках исходной предметной области, так и в системе самостоятельных, но связанных между собой предметных областей.

При переходе от предметной области к ее модели, представленной в виде семантической сети, выполняются следующие условия:

- каждому элементу предметной области взаимно однозначно соответствует обозначающий его элемент семантической сети;

- каждому сигнатурному элементу предметной области взаимно однозначно соответствует либо обозначающий его ключевой узел семантической сети, либо обозначающий элемент алфавита семантической сети.

Первым и важнейшим этапом проектирования семантической модели базы знаний является уточнение структуры описываемой предметной области или нескольких взаимосвязанных предметных областей. Под уточнением структуры предметной области понимается явное выделение *класса исследуемых объектов, класса вторичных объектов*, построенных на основе исследуемых, *класса вспомогательных объектов*, через связи с которыми описываются некоторые характеристики исследуемых объектов, *отношения, связи которых связывают только исследуемые объекты между собой*, а также *отношения, связи которых связывают исследуемые объекты со вспомогательными*.

Рассмотрение базы знаний с позиции ее соотношения с предметной областью позволяет рассматривать исследуемые объекты на различных уровнях детализации:

- классификация класса исследуемых объектов по различным признакам;
- классификация самих исследуемых объектов, т.е. рассмотрение структур взаимосвязанных частей этих объектов;
- рассмотрение связей исследуемых объектов со вспомогательными объектами, не входящими в класс исследуемых объектов.

В зависимости от исследуемых объектов можно говорить о достаточно богатой типологии предметных областей. Можно выделить следующие классы предметных областей:

- предметная область, описывающая *теоретико-множественные характеристики и связи* заданного семейства объектов. Такие предметные области, в частности, могут быть онтологиями других предметных областей;
- *терминологическая онтология* – это класс предметных областей, для каждой из которых объектами исследования являются термины (словосочетания), соответствующие различным элементам описываемой предметной области;
- *предметная область, являющаяся логическим описанием некоторой предметной области* – это класс предметных областей, для каждой из которых объектами исследования являются всевозможные логические формулы, описывающие причинно-следственные закономерности, интерпретируемые на некоторой описываемой предметной области;
- *логическая система понятий, описываемых в заданной формальной теории*. Эта предметная метаобласть выделяет класс понятий, не определяемых в заданной формальной теории, и связывает каждое определяемое понятие с теми понятиями, на основе которых оно определяется;

- *логическая система утверждений заданной формальной теории*. Эта предметная метаобласть выделяет класс аксиом для заданной формальной теории, каждой теореме ставит в соответствие одно из её доказательств (основное доказательство) и связывает каждую теорему со всеми теми утверждениями и определениями, которые используются в основном доказательстве этой теоремы;

- *предметная область вопросов и информационных задач* – это класс предметных областей, для каждой из которых объектами исследования являются вопросы, информационные задачи, обобщенные вопросы и обобщенные информационные задачи, задаваемые по отношению к некоторой описываемой предметной области, а также соответствующие им способы решения информационных задач (то есть различные программы).

Семантическая структура базы знаний интеллектуальной системы трактуется в рамках технологии проектирования баз знаний интеллектуальных систем как *иерархическая система взаимосвязанных между собой предметных областей*, которые представляются в базе знаний.

На множестве предметных областей могут быть заданы следующие отношения: включение, объединение, пересечение, декомпозиция, гомоморфизм, изоморфизм, теоретико-множественная онтология, логическое описание, логическая онтология. Исходя из этого, мы можем рассматривать некую *метаобласть*, объектами исследования которой являются всевозможные предметные области.

Таким образом, семантическая структура базы знаний представляет собой иерархическую систему описываемых ею предметных областей, надстраиваемых над заданной основной предметной областью.

Построение семантической структуры базы знаний требует не только явного представления спецификации каждой описываемой предметной области в виде формального текста, но и явного описания всевозможных связей между этими предметными областями.

### 3. Библиотека многократно используемых компонентов баз знаний

В целях сокращения времени процесса проектирования семантических моделей баз знаний интеллектуальных систем необходимо создать библиотеку многократно используемых семантически совместимых компонентов баз знаний. На основе этой библиотеки разработана методика компонентного проектирования баз знаний.

К основным типам компонентов баз знаний, хранящихся в библиотеке относятся:

- онтологии различных предметных областей, которые могут быть самыми различными по содержанию, однако должны быть семантически совместимыми;

- базовые фрагменты теорий, соответствующие различным уровням знания пользователя, начиная от базового школьного до профессионального;

- семантические окрестности различных объектов;

- спецификации формальных языков описания различных предметных областей.

Для обеспечения семантической совместимости таких компонентов баз знаний, которые являются унифицированными семантическими моделями, необходимо

- согласовать семантику всех используемых ключевых узлов;

- согласовать глобальные идентификаторы ключевых узлов, используемых в разных компонентах. После этого интеграция всех компонентов, входящих в состав библиотеки, и в любых комбинациях осуществляется автоматически, без вмешательства разработчика.

Для включения компонента в библиотеку необходимо его специфицировать по следующим критериям:

- предметная область, описание которой содержится в компоненте;

- класс (тип) компонента базы знаний;

- состав базы знаний;

- количественные характеристики ключевых узлов базы знаний;

- информация о разработчиках базы знаний;

- дата создания базы знаний;

- информация о верификации базы знаний;

- версия компонента базы знаний;

- условия распространения компонента базы знаний;

- сопровождающая информация.

#### **4. Средства проектирования семантических моделей баз знаний**

При коллективном инжиниринге знаний любая база знаний рассматривается как результат консенсуса группы специалистов о модели некоторой области знаний, что должно быть обеспечено соответствующими средствами проектирования [Ефименко, 2011].

Для эффективной организации проектирования баз знаний интеллектуальных справочных систем необходимо включать в состав основной системы в качестве подсистем следующие компоненты:

- интеллектуальную подсистему поддержки проектирования и сопровождения системы;

- help-систему информационного обслуживания разработчиков баз знаний;

- интеллектуальную систему автоматизации проектирования;

- интеллектуальную подсистему управления проектированием и сопровождением системы.

К средствам проектирования также относятся:

- средства верификации баз знаний, включающие пополняемую библиотеку команд и операций верификации баз знаний;

- средства анализа качества баз знаний, позволяющие определить такие характеристики баз знаний как полнота, связность, информативность;

- средства редактирования баз знаний, решающие проблему синхронизации редактирования семантической модели базы знаний и соответствующего фрагмента ее исходного текста несколькими разработчиками;

- средства поддержки коллективной разработки баз знаний.

#### **5. Методика проектирования семантических моделей баз знаний**

В основе комплексной методики проектирования семантических моделей интеллектуальных систем лежат два аспекта: эволюционное проектирование и коллективное проектирование интеллектуальных справочных систем.

В рамках семантической технологии проектирования баз знаний выделяют три основных метода разработки баз знаний.

1. Тест-ориентированный метод разработки баз знаний. Данный метод основывается на реализации стартового диалога с конечным пользователем системы. На первом этапе проектирования базы знаний исследуются запросы пользователей к проектируемой базе знаний и составляется тестовый сборник вопросов, охватывающий вопросы пользователей по рассматриваемой предметной области, что предполагает выделение семантически полного набора вопросов, ответы на которые должны содержаться в первой версии базы знаний. На следующем этапе все вопросы систематизируются и объединяются в классы. Далее на все вопросы, входящие в тестовый сборник записываются ответы, затем они переводятся на формальный язык. В процессе формализации ответов на вопросы выделяются ключевые узлы описываемой предметной области. Таким образом формируется первая версия базы знаний. Затем сборник вопросов расширяется и база знаний развивается эволюционно.

Такой метод проектирования баз знаний достаточно удобен при разработке баз знаний систем справочного назначения. К достоинствам данного подхода можно отнести эволюционное развитие базы знаний, которое позволяет быстро получить первый прототип.

2. Метод поэтапной формализации, основанный на семантической структуризации исходных документов. Данный подход к разработке баз знаний

основывается на использовании в качестве источника информации для формализации некоторого документа (книги, справочника, учебника, энциклопедии и т.п.).

В качестве подходов к декомпозиции процесса проектирования можно выделить следующие:

- выделение ветвей проектирования, соответствующих содержательной декомпозиции материала исходного документа (разделение на главы, параграфы, части и т.п.) вплоть до атомарных разделов. В данном случае каждый атомарный раздел документа соответствует атомарному разделу проектируемой базы знаний.

- проектирование семантической модели базы знаний заключается в детальной структуризации базы знаний, т.е. рассматривать структуру базы знаний, как иерархическую систему взаимосвязанных друг с другом предметных областей, представляемых в базе знаний. При таком рассмотрении процесса проектирования модели базы знаний можно выделить отдельные направления работ, соответствующие различным видам структуризации базы знаний, таких как построение теоретико-множественной онтологии описываемой предметной области, построение логической онтологии предметной области, построение онтологии системы высказываний предметной области, построение терминологической онтологии предметной области, и т.д. (направления структуризации знаний, хранящихся в базе знаний рассматривались выше).

3. Метод компонентного проектирования баз знаний, основанный на модели глобального семантического пространства человеческих знаний.

Суть данного подхода заключается в использовании библиотеки многократно используемых компонентов баз знаний, пополняемой авторами баз знаний. Основной интеграции для всех компонентов служит унифицированная модель представления знаний с использованием SC-кода [Голенков, 2011]. Достоинством данного подхода является максимальное сокращение сроков разработки базы знаний. Фактически процесс разработки базы знаний сводится к выбору инженером необходимых компонентов баз знаний из библиотеки, а затем их интеграции в единое семантическое пространство (принципы и подходы к интеграции знаний рассмотрены в работе [Ивашенко, 2013]). Недостающие фрагменты базы знаний разрабатываются одним из приведенных выше методов.

## ЗАКЛЮЧЕНИЕ

В данной работе рассмотрена технология проектирования баз знаний интеллектуальных систем, основанная на семантических сетях. К особенностям приведенной технологии можно отнести:

- разнообразие видов хранимых знаний;

- наличие унифицированных средств структуризации базы знаний;
- наличие средств навигации по базе знаний;
- обеспечение коллективного проектирования баз знаний соответствующими средствами;
- обеспечение компонентной разработки баз знаний;
- наличие средств анализа качества баз знаний;
- мощные средства редактирования баз знаний.

Результаты, приведенные в работе, апробируются в рамках открытого проекта OSTIS [OSTIS, 2012].

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

[Башмаков, 2003] Башмаков, А. И. Разработка компьютерных учебников и обучающих систем. / Башмаков А. И., Башмаков И. А., М.: Информационно-издательский дом "Филинь", 2003. — 616 с.

[Башмаков, 2004] Башмаков, А.И. Интеллектуализация как средство повышения доступности технологий разработки компьютерных средств обучения / А.И. Башмаков // Образовательная среда сегодня и завтра: материалы Всероссийской научно-практической конференции [Электронный ресурс]. — 2004. - Режим доступа: <http://www.ict.edu.ru/vconf/index.php>. — Дата доступа: 10.10.2011

[Бениаминов, 2008] Бениаминов, Е.М. О построении Web-сервера в стиле Semantic Wiki с открытым контекстным языком представления и запросов/Е. М. Бениаминов// КИИ-2008. Труды конференции. Т 2, С. 15-21

[Голенков, 2011] Голенков, В.В., Гулякина Н.А. Принципы построения массовой семантической технологии компонентного проектирования интеллектуальных систем. — В кн Междунар. научн.-техн. конф. . «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2011). Материалы конф. [Минск, 10-12 февр. 2011 г.]. — Минск: БГУИР, 2011, с. 21-59.

[Голенков, 2012] Голенков, В.В., Гулякина Н.А. Графодинамические модели параллельной обработки знаний: принципы построения, реализации и проектирования. — В кн Междунар. научн.-техн. конф. . «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2012). Материалы конф. [Минск, 16-18 февр. 2012 г.]. — Минск: БГУИР, 2012.

[Гаврилова и др., 2001] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. Учебник / Гаврилова Т.А. [и др.]; — СПб.: Изд-во «Питер», 2001.

[Гаврилова, 2008] Гаврилова, Т. А. Визуальные методы работы со знаниями: попытка обзора / Т. А. Гаврилова, Н. А. Гулякина // Искусственный интеллект и принятие решений, 2008, № 1, С. 15-21

[Грибова, 2011] Грибова, В. В. Облачная платформа для разработки и управления интеллектуальными системами / В. В. Грибова, А. С. Клещев, Д. А. Крылов, Ф. М. Москаленко, С. В. Смагин, В. А. Тимченко, М. Б. Тютюнник, Е. А. Шалфеева //Материалы международной научно-технической конференции OSTIS-2011, 5-14 стр. Минск БГУИР

[Давыденко, 2012] Давыденко И. Т. Комплексная методика проектирования семантических моделей интеллектуальных справочных систем Интеллектуальная справочная система по геометрии / И. Т. Давыденко //Материалы международной научно-технической конференции OSTIS-2012, Минск БГУИР

[Ефименко, 2011] Ефименко, И. В., Хорошевский, В. Ф. Онтологическое моделирование экономики предприятий и отраслей современной России: Часть 1. Онтологическое моделирование: подходы, модели, методы, средства, решения : препринт WP7/2011/08 (ч. 1) / И. В. Ефименко, В. Ф. Хорошевский ; Нац. исслед. ун-т «Высшая школа экономики». — М. : Изд. дом Высшей школы экономики, 2011. — 76 с.

[Загорулько и др., 2008] Загорулько, Ю.А., Боровикова, О.И. Подход к построению порталов научных знаний / Ю.А. Загорулько [и др.] // Автометрия. – 2008 – № 1, Т. 44, – С. 100–110.

[Клещев и др., 2001a] Клещев А.С., Артемьева И.Л. Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия "онтология" // НТИ. Серия 2 "Информационные процессы и системы", 2001, № 2

[Клещев и др., 2001b] Клещев А.С. Артемьева И.Л. Математические модели онтологий предметных областей. Часть 2. Компоненты модели // НТИ. Серия 2 "Информационные процессы и системы", 2001, № 3

[Клещев и др., 2001c] Клещев А.С. Артемьева И.Л. Математические модели онтологий предметных областей. Часть 3. Сравнение разных классов моделей онтологий // НТИ. Серия 2 "Информационные процессы и системы", 2001, № 4

[Ивашенко, 2012] Ивашенко, В. П. Семантические модели и средства интеграции и отладки баз знаний / В. П. Ивашенко // Материалы международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» – Минск, 2012. – С. 193-204

[Ивашенко, 2013] Ивашенко В.П. Унифицированное представление и интеграция знаний/ В. П. Ивашенко // Материалы международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» – Минск, 2013.

[Хорошевский, 2008] 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.

[ЭЗОП, 2007] Web-сервер онтологий системы ЭЗОП [Электронный ресурс]. – Режим доступа: <http://ezor-project.ru/drupal5/>. – Дата доступа: 14.11.2012.

[OSTIS, 2012] Проект OSTIS [Электронный ресурс]. Минск, 2012. – Режим доступа: <http://ostis.net/>. – Дата доступа: 11.12.2012.

## COMPONENT TECHNOLOGY OF DESIGN OF KNOWLEDGE BASES BASED ON UNIFIED SEMANTIC NETWORKS

Davydenko I.T.

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

**ir.davydenko@gmail.com**

### INTRODUCTION

This paper considers the composition of semantic technology knowledge base design of intelligent systems. This technology is focused on the semantic representation of knowledge, the expansion contingent development of knowledge bases and reducing the time of design.

### MAIN PART

The technology of design knowledge bases is a set of models, tools, and methods for knowledge bases designing.

The proposed technology has the following structure: unified semantic model of knowledge representation; semantic model of the knowledge base; library of reusable components of knowledge bases; design tools of semantic models of knowledge bases; the methods of designing semantic models of knowledge bases.

As a formal framework designed logical and semantic models of knowledge bases of intelligent systems using a semantic model of representation and processing of knowledge, which are based on semantic networks with a basic set-theoretic interpretation.

The unified semantic model of knowledge representation is in fact a formal way of describing the semantics of different kinds of knowledge. Semantic model of a knowledge base of intelligent system is a formal interpretation of the semantic space, which is known for the intelligent system at the current time.

To enhance the diversity of the knowledge stored in the knowledge base, a necessary step in the development of the semantic model knowledge base is its structuring.

The semantic structure of a knowledge base of intelligent system is treated in the technology knowledge base design of intelligent systems as a hierarchy of inter-related subject areas, which are submitted to the knowledge base.

In order, to reduce the time of the design process of semantic models of knowledge bases of intelligent systems, it's needed to create a library of reusable components is semantically compatible knowledge bases.

Designing tools are also include: tool of verification of knowledge bases, including the growing library of commands and operations verification of knowledge bases; tools for analysis of quality of knowledge bases for determining properties such as completeness of knowledge bases, connectivity, information; editing knowledge bases, the crucial problem of synchronization editing semantic model knowledge base and the corresponding fragment of the source code by multiple developers; tools to support the collaborative development of knowledge bases.

As a part of the semantic technology design knowledge bases there are three main methods of development of knowledge bases:

1. Test-driven development method knowledge bases.
2. The method of gradual formalization, based on the semantic structuring of the source documents.
3. The method of the component database design knowledge based on the idea of a global semantic space of human knowledge.

### CONCLUSION

In the given paper short description of component technology of design of knowledge bases based on unified semantic are present. The article describes the main parts of this technology: the model of representation of knowledge, the model of knowledge base, tools and methods of designing of knowledge bases based on the unified semantic networks.

Given results are tested in an open project OSTIS [OSTIS, 2012].