

В целом модуль выборок и запросов ориентирован на предоставление пользователю гибких возможностей для формирования запросов на основе любых имеющихся в базе данных свойств и значений единиц и их связей.

2.7. *Модуль словарей.* Центральным модулем системы «Манускрипт» является разрабатываемый модуль словарей. Модуль должен обеспечить использование единых словарей при фрагментировании и разборе текстов и рукописей.

Наиболее сложной по исполнению является система лингвистических словарей, которая ориентирована на автоматизированный морфологический анализ древних и средневековых славянских текстов. В основе грамматических словарей лежит принцип оперирования изменяемыми и неизменяемыми частями словоформы — окончаниями и основами. В системе «Манускрипт» используются два известных в настоящее время подхода к делению словоформы на эти компоненты: с сохранением в основе чередований (нормализованные основы и окончания) и без сохранения чередований (псевдоосновы и псевдоокончания). Это сделано для того, чтобы комплексно решить главную задачу автоматического морфологизатора — задачу разбора очень сильно варьирующейся графической и морфологической структуры древней славянской словоформы.

Основными возможностями модуля в настоящее время являются (а) хранение элементов словарей, связей, а также их свойств и значений, (б) редактирование элементов словарей, (в) установление связей между элементами одного словаря и элементами различных словарей, (г) установление связи между основами и их окончаниями, (д) построение парадигм на основе леммы.

3. Таким образом, в настоящее время ИПС «Манускрипт», имея развитые средства для ввода, редактирования, разбора, макетирования и публикации сложных по структуре и составу документов, а также для создания запросов и работы с выборками, удовлетворяет многим требованиям, предъявляемым к многофункциональным полнотекстовым исследовательским системам, а использование системы при лингвистических и лингвотекстологических исследованиях позволило получить материал для анализа древних славянских текстов, их фрагментов и сделать значимые фундаментальные выводы.

Описание проблемной области в интеллектуальных информационных технологиях

И. А. Барков

Ижевский государственный технический университет, Россия

The main principles of creating and using the semantic describing of the problem sphere in the intellectual informational technologies are considered.

Введение

Основной тенденцией создания современных информационных технологий, в том числе и информационных технологий обработки текстов, является попытка придать ей признаки интеллектуальности. Под интеллектуальностью в этом случае понимается обеспечение автоматизированной системы возможностями обработки смысловой, семантической информации. Традиционно в интеллектуальных информационных технологиях смысловая составляющая информация фиксируется в модели предметной области, а для реализации процессов обработки используется логический вывод. В общем случае под предметной областью памятника письменного наследия будем понимать смысловое описание изложенной в памятнике темы. Как правило, созданием описания предметной области занимается эксперт. Современный уровень развития научного и прикладного знания характерен развитой системой наук, поэтому задача описания предметной области является многопрофессиональной: один и тот же объект или явление в глазах различных специалистов получает совершенно различное смысловое описание. Различие смысловой интерпретации окружающего мира привело к дифференциации наук, усиливающейся с течением времени. Используя понятие семантического треугольника можно представить задачу описания проблемной области следующим образом (рис. 1).

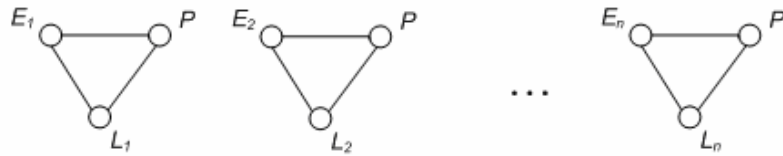


Рис. 1. Схема задачи описания проблемной области

На рис. 1 использованы следующие обозначения: P — проблемная область; E_1, \dots, E_n — эксперты; L_1, \dots, L_n — описание проблемной области. Таким образом, описание проблемной области представляет собой кортеж $L = \langle L_1, \dots, L_n \rangle$. Дифференциация наук привела к тому, что каждый специалист E_i изобрел свой язык L_i , ($i=1, \dots, n$), свою систему смысловых приемов познания мира. При создании автоматизированных систем подобное разнообразие является непожелательной роскошью. Создание единого семантического языка и инвариантных семантических приемов рассуждения становится необходимостью, какой бы трудной или даже абсурдной на первый взгляд ни выглядела эта задача. Автором разработан семантический язык и информационная технология для проблемной области «Проектирование изделий». Настоящая работа является попыткой изучения возможности трансформации полученных результатов в задачи обработки текстовой информации.

1. Основные принципы создания и использования описания проблемной области

Сформулируем основные принципы создания семантических средств.

Принцип инвариантности. Суть этого принципа заключается в том, что в дальнейшем мы будем использовать только инвариантные средства семантического описания проблемной области и использования таких средств в решении задач. Их немного, но их нужно собрать, объединить и на этой основе в дальнейшем получать новые.

Принцип языкового ядра и его профессиональных расширений. Этот принцип заключается в том, что необходимо создать инвариантный семантический язык, позволяющий построить описание базовых смысловых данных и операций. Далее необходимы средства построения профессиональных расширений базового языка. В этом случае деятельность эксперта E_i , $i=1, \dots, n$, будет заключать-

ся в создании своего профессионального языка и последующего описания своей проблемной области на созданном языке. Естественно, что главным условием построения профессиональных расширений базового языка будет использование единых алгоритмических приемов обработки, как базового языка, так и его профессиональных расширений.

Принцип исполняемости языка заключается в том, что автоматизированная система должна иметь возможность эффективной обработки профессиональных описаний проблемной области.

Принцип смыслового абстрагирования. Суть этого принципа состоит в ориентации на смысловое, семантическое описание проблемной области. В качестве базовых семантических категорий в настоящей работе выбраны свойства и особенности. Под свойством объекта, явления понимается такая черта, которая характеризует именно данный объект, данное явление. Свойство характеризует объект изолированно от внешнего мира. Особенность объекта или явления представляет собой черту, зависящую от внешнего мира. Например, диаметр является свойством трубы, а пропускная способность — особенностью; использование соединений железа для изготовления трубы характеризует свойство трубы, а коррозия — особенность. Описание свойства представляет собой отношение, заданное на внутренних элементах объекта; описание особенности представляет собой отношение, заданное на элементах объекта и элементах объектов внешнего мира. В автоматизированных системах имеются хорошие средства представления и обработки отношений.

Принцип типизации языка и универсума. Данный принцип необходим нам для реализации механизма приписывания набора свойств и особенностей объектам и явлениям проблемной области. Описание типа представляет собой перечисление определенного набора свойств и особенностей, таким образом, тип становится носителем заданного набора свойств и особенностей. Сопоставление имени объекта или явления некоторого типа представляет собой приписывание объекту заданного в типе набора свойств и особенностей. Например, типу «множество» мы можем приписать свойство различимости элементов и свойство неупорядоченности элементов. Затем, если некоторому объекту мы припишем тип «множество», то будем считать, что указанный объект становится

обладателем свойств «множество». В настоящей работе понятие «тип» отождествляется с понятием «понятие». Описание понятий можно реализовать в автоматизированной системе в виде специализированного тезауруса. Таким образом, понятие является носителем набора свойств и особенностей, а тезаурус представляет собой систему определения понятий путем перечисления свойств и особенностей.

Принцип существования представителя предметной области. Построение описания проблемной области в виде связанного множества вариантных определений понятий предполагает, что результатом будет описание класса объектов или явлений. Данный принцип утверждает возможность или даже необходимость существования конкретного объекта или явления. Тогда возникают две задачи использования описаний предметной области.

Первая, задача анализа, формулируется следующим образом: даны объект или явление; проверить, являются ли они представителями предметной области. В автоматизированной системе задача анализа решается путем проверки исполнения в предъявленном объекте свойств и особенностей, определенных в описании предметной области.

Вторая задача использования описания предметной области является задачей синтеза: при некоторых начальных условиях и заданном описании предметной области получить множество представителей предметной области. Предполагается, что в каждом элементе множества представителей присутствуют свойства и особенности, заданные в описании предметной области. В автоматизированной системе задача синтеза реализуется в виде специализированного генератора.

Принцип сочетаемости понятий. Этот принцип состоит в том, что в описание проблемной области включаются условия сочетаемости понятий. Например, понятие «воздушный шарик» сочетается с понятием «цвет», а понятие «электрон» с понятием «цвет» не сочетается; понятие «воздушный шарик» сочетается с понятием «материал», а понятие «автомобиль» с понятием «материал» не сочетается (автомобиль состоит из деталей, соответствующее каждой детали автомобиля понятие может сочетаться с понятием «материал»). Замечание: приведенные примеры на самом деле утверждают, что автору (эксперту) известны или не известны ука-

занные сочетания понятий. В общем случае можно задавать как условия сочетаемости понятий, так и условия их не сочетаемости. Сочетаемость понятий реализуется в автоматизированной системе с помощью правил сочетаемости понятий, которые представляют собой некоторые высказывания.

Принцип иерархического абстрагирования является наиболее распространенным и проверенным практическим опытом принципом. Поэтому он является инвариантным средством для многих наук. Наиболее распространенными отношениями являются: компонентная, атрибутивная, компонентно-атрибутивная, атрибутивно-компонентная сочетаемость понятий.

2. Основные средства реализации описания проблемной области и решения задач

Понятия. Понятие представляет собой мысль, отражающую в обобщенной форме предметы и явления действительности посредством фиксации их свойств и особенностей. Свойства и особенности выступают в понятии как общие и специфические признаки, соотношенные к классам предметов и явлений. Например, «крышка», «корпус», «редуктор» — это понятия, которые являются носителями определенных свойств, после включения их в конструкцию становятся дополнительно носителями компонентных свойств изделия. Аналогично, понятия «диаметр», «материал», «мощность» обозначают определенные атрибутивные свойства (параметры) элементов изделия. Будем считать указанные понятия носителями атрибутивных свойств. Совокупность компонентных и атрибутивных свойств с заданными структурными отношениями будем называть структурными свойствами. Каждое понятие будем сопровождать описанием условий сочетаемости структурных свойств, заданных на декартовом произведении множества носителей структурных свойств. Простейшим примером сочетаемости конструкторских параметров является следующий: вес изделия равен сумме весов всех компонентов. Будем называть условия сочетаемости структурных свойств предикатными свойствами.

Содержание и объем понятия. Содержание понятия фиксируется в виде описания его свойств и особенностей, а выделение объема понятия, то есть допустимых реализаций понятия, будет осуществляться процедурно путем проверки выполнения свойств и особенностей в рассматриваемых объектах. Для трактовки ука-

занной проблемы понятия в ее формально-логическом аспекте можно воспользоваться готовым арсеналом трех областей современного знания [Барков 2003]: 1) общей алгебры, 2) логической семантики, 3) математической логики.

1. Процесс образования понятия в алгебре естественно описывается в терминах гомоморфизма; разбивая интересующее нас множество объектов на классы эквивалентных в каком-либо отношении элементов, игнорируя при этом все различия между элементами класса, не интересующие нас в данный момент. Полученное множество отождествленных между собой образов исходных объектов и есть понятие. Например, во множестве «электрические машины» используя отношение эквивалентности «использующие постоянный ток» мы можем выделить класс эквивалентности: «электрические машины постоянного тока»

2. При рассмотрении семантического аспекта проблемы понятия объемом понятия называется совокупность объединяемых в это понятие объектов, а содержанием понятия — перечень свойств и особенностей, на основании которых производилось это объединение. Например, для описания смысла термина «автомобиль» можно указать его компонентные свойства (состоит из кузова, двигателя, шасси), атрибутивные свойства (характеризуется скоростью движения, количеством пассажиров, принципом работы двигателя) и множество соотношений, которые однозначно показывают, что выделенный набор указанных свойств в сочетании соответствует автомобилю, а не трактору. Чем обширнее набор свойств, тем уже класс объектов, удовлетворяющих этим свойствам, и наоборот, чем уже содержание понятия, тем шире его объем; это очевидное обстоятельство часто именуют законом обратного отношения.

3. Логико-математическую проблематику, связанную с теорией понятия, можно описать, опираясь на хорошо разработанный аппарат исчисления предикатов. Семантика этого исчисления такова, что им легко описывается субъектно-предикатная структура суждений. При этом субъекты и предикаты выступают как имена объектов: вторые самым буквальным образом, а первые, будучи переменными, «пробегают» некоторые предметные области, служащие объемом понятия. Таким образом, предикаты — это содержание понятия, а классы объектов, на которых эти предикаты

истинны, образуют объем понятия. Например, «сверхзвуковой самолет» — это такой самолет, у которого число Маха M принадлежит диапазону 1-5. В этом примере для описания содержания понятия «сверхзвуковой самолет» используется предикатное свойство, а объемом понятия являются все удовлетворяющие указанному свойству самолеты.

Описание понятий. Для создания профессиональных языков и методов работы с ними необходимо предусмотреть средства расширения семантического языка. Важнейшим инвариантным средством является теория определений. Чтобы квалифицированно применять теорию определений, мы обязаны хорошо различать, что и как можно выразить исходными средствами. Рассмотрим некоторые особенности построения определения понятий.

Определение понятия это указание или объяснение смысла обозначаемого понятие термина и (или) содержания выражаемого данным термином понятия. Этот термин называется определяемым, а совокупность действий (слов), осуществляющих его определение, — определяющим [Лингвистический 1990]. Определения, в которых и определяемое и определяющее имеют языковую природу, называются вербальными [Там же]. Посредством вербальных определений преимущественно вводятся новые термины.

Явные (эксплицитные) определения позволяют не только вводить определяемое в качестве «сокращения» для определяющего в любой контекст, но и удалять из произвольного контекста определяемое, «расшифровывая» его посредством «определяющего». В научной практике весьма распространены также неявные (имплицитные) определения, в которых определяемое непосредственно не дано, но может быть «извлечено» из некоторого контекста. Типичным примером является процесс решения системы уравнений, которая с самого начала может рассматриваться как неявное определение неизвестных.

Делению определений на виды в современной логике соответствуют синтаксические и семантические определения [Справочная 1982]. В первых определяемое и определяющее принадлежат одному семантическому уровню (значение выражений определяется через значение других выражений), во вторых определяемое и определяющее представляют собой языковые выражения разных

уровней абстракции (значение термина определяется через свойство предметов).

К определениям всех видов предъявляется ряд общих требований. Правило переводимости (или эллиминируемости) требует равнообъемности определяемого и определяющего. Правило однозначности — это требование единственности определяемого для каждого определяющего. Если определяется термин, обладающий полисемией, то необходимо определить каждое смысловое содержание данного термина. Наконец, правило отсутствия порочного круга: определяющее не должно зависеть от определяемого (негативный пример: «языкознание — это наука о языках», «язык — это объект изучения языкознания» в [Лингвистический 1990]).

Семантические структуры. Предложение и другие синтаксические единицы языка являются частью знаковой системы языка. С другой стороны они являются самостоятельными единицами и характеризуются своими семантическими признаками. Так, выделяя слова и предложения как основные знаки языка, лингвисты в то же время четко разграничивают их отличительные признаки. В [Никогосов 1982] говорится в этой связи: «Если слово именуется вещь, то высказывание квалифицирует вещь или процесс с точки зрения познания человека». В [Колшанский 1981] конкретизируется это противопоставление: «В качестве референта слова — отдельно взятый объект, в качестве референта предложения — связь между отдельными объектами или между объектом или его признаками». Исходя из этих различий, лингвисты предлагают рассматривать предложение как сложный знак по отношению к простым знакам — словам, как номинацию особого рода, денотатом которой является не предмет, а целая ситуация, факт.

Будучи языковым знаком, предложение характеризуется диалектическим единством двух таких его сторон, как означающее и означаемое, которым соответствуют синтаксический и семантический аспекты изучения. Каждый из этих двух аспектов оперирует понятием «структура». Синтаксическая структура предложения отражает его грамматические свойства. Содержательное, смысловое наполнение синтаксических моделей исследуется семантикой, которая оперирует понятием семантической структуры, представляющей обобщенные модели фактов и ситуаций, отраженных человеческим сознанием в форме предложения.

При построении семантических конструкторских структур мы будем использовать в качестве базовых свойств понятия, а правила семантической сочетаемости понятий будем рассматривать как средство структурного описания свойств объекта или явления. Автоматизированная система должна содержать два взаимосвязанных вида семантических моделей: структурную семантическую модель и общую семантическую модель. Отметим, что обе модели являются по своей сути семантическими моделями, их отличие состоит в том, что первая описывает структурные свойства, а вторая — произвольные. Структурная семантическая модель определяет возможные структурные сочетания свойств. Каждое предложение общей семантической модели должно показывать предикатные свойства в виде записи соответствующего соотношения над элементами структурной семантической модели.

Семантическая сочетаемость понятий — это инвариантный прием введения отношений между понятиями, способ определения и выявления свойств и особенностей. Правила сочетаемости понятий являются своеобразными правилами логического вывода в процедурах семантического анализа и синтеза. Рассмотрим более подробно некоторые правила семантической сочетаемости понятий.

Правила структурной сочетаемости понятий характеризуют структурные свойства объекта или явления путем описания множества допустимых сочетаний понятий.

Правило агрегации описывает отношение целого и части, приводящее к соответствующей иерархии понятий. Агрегация может означать компонентную сочетаемость понятий. Например, самолет состоит из фюзеляжа, двигателей, шасси и других частей. С другой стороны отношение понятия «самолет» с понятиями «крейсерская скорость полета», «пассажир», «горючее» — это агрегация, которая предусматривает компонентно-атрибутную сочетаемость. Правило агрегации подчеркивают индивидуальность структурных свойств. Например, «в планетарных передачах оси некоторых колес — сателлитов закреплены на вращающемся звене — водиле». В данном примере утверждается, что понятия «сателлит», «водило» являются неотъемлемой частью понятия «планетарная передача».

Правило наследования — это такое отношение между понятиями, когда структурный элемент повторяет свойства элементов других понятий. Понятие, структура и свойства которого наследуются, будем называть родовым понятием. Производное от родового понятия будем называть частным (партикулярным) понятием. Например, «в мелких редукторах (массой свыше 25 кг) на крышке корпуса необходимо предусматривать элементы для строповки». Здесь говорится о том, что понятие «мелкий редуктор» обладает теми же свойствами, что и «обычный редуктор», но при этом добавляются еще индивидуальные свойства: компонентное («включает элементы для строповки») и предикатное («масса редуктора должна быть свыше 25 кг»). Отношение наследования устанавливает между конструкторскими понятиями иерархию общего и частного.

Важнейшей разновидностью отношения наследования является отношение референциального тождества. Отношение референциального тождества понятий определяют одинаковость смыслового соответствия понятий. Например, «вал является телом вращения» — это утверждение о том, что понятия «вал» и «тело вращения» обозначают один и тот же объект действительности (референт). Однако понятие «вал» принадлежит, например, проблемной области «редукторы», а понятие «тело вращения» — проблемной области «геометрия». С понятием «вал» в описании будут связаны представления инженера-конструктора, а с понятием «тело вращения» — математика.

Правила наследования позволяют построить иерархию понятий, отражающую свойства компонент и атрибутов по отношению наследования.

Семантическая валентность. Семантическая валентность [Лингвистический 1990] представляет собой набор семантических характеристик понятия, регламентирующих способность вступать в структурные (синтаксические) связи с другими понятиями.

Рассмотрим примеры семантических валентностей.

1. Характеристика состава понятия «агрегат/блок» отражает способность понятия (структурно) присоединять другие понятия или, наоборот, присоединяться к господствующему компоненту сочетания. С помощью этой характеристики можно задать свойства: быть целым, быть частью целого, быть неделимым элементом.

2. Характеристика облигаторности понятия «обязательное/факультативное» означает возможность присутствия обозначаемого понятием денотата в объекте. Например, такие части изделия как «ручка», «строповочный узел», «транспортное крепление» в одних случаях включаются в изделие, а в других — нет.

3. Характеристика зависимости понятия «зависимое / независимое» обозначает обязательность (не обязательность) совместного использования структурных свойств. Простейшим примером является использование болта и гайки, которые в отдельности являются бесполезными элементами конструкции изделия.

Семантические валентности понятия дают разнообразный набор приемов для описания свойств изделия в целом и его составных частей.

Правила общей сочетаемости понятий. Представляют собой предикатные свойства, задаваемые путем описания произвольных соотношений на множестве понятий, которые должны выполняться в проблемной области.

Решение задач анализа и синтеза. Описание проблемной области, построенное по указанным ранее принципам, представляет собой систему логических соотношений, характеризующих структурные и предикатные свойства объектов и явлений. В этом случае содержанием задач анализа и синтеза становится поиск решения системы логических соотношений. Этот прием рассматривается как инвариантный способ решения семантически различных профессиональных задач.

Заключение

Переход к интеллектуальным информационным технологиям обработки текстов требует моделирования в явном виде определяемой текстом проблемной области. Для построения описания проблемной области предложена концепция семантики, которая ориентирована на детальное описание свойств и особенностей класса объектов или явлений. Предложенная концепция позволяет получить инвариантные средства построения базирующихся на едином семантическом языке профессиональных расширений, а также инвариантный способ решения различных профессиональных задач анализа и синтеза.

Список литературы

- Барков 2003 — Барков, И. А. Теория конструкторской семантики / И. А. Барков. — Ижевск : Изд-во ИжГТУ, 2003. — 360 с.
- Лингвистический 1990 — Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева. — М. : Сов. энциклопедия, 1990. — 685 с.
- Справочная 1982 — Справочная книга по математической логике : в 4 ч. / под ред. Дж. Барвайса. — М. : Наука, Главная редакция физико-математической литературы, 1982.
- Никогосов 1982 — Никогосов, С. Л. Лингвистическая семантика и логика / С. Л. Никогосов // Семантический компонент в системах автоматического понимания текстов. — М. : ВЦП, 1982. — С. 73–78.
- Колшанский 1981 — Колшанский, Г. В. Компоненты структуры текста / Г. В. Колшанский // Лингвистика в высшей школе. — Вып. X. — М. : 1981. — С. 2–13.

Лексико-семантический анализ рукописной книжности XVIII — начала XX вв. старообрядцев Прикамья

Г. П. Волгирева
Пермский государственный университет, Россия

This paper considers Cyrillic Old Believer's book-knowledge of the XVI — begining XX centuries in the Perm region from the point of view of sources. The author analyzes in details some spiritual, spatial and temporal aspects of the traditional mentality. One of the special points of her interest are questions of enterprise and buisness represented in the Old Believers educational Psalter.

Пермская губерния в начале XX века по количеству проживающих в ней старообрядцев, как известно, занимала первое место. По официальным данным, в ней проживало 216 441 человек, из которых 153 070 принадлежали к согласию «часовенных» [Поздеева 1995: 8]. Данные количественные показатели извлечены из «Статистических сведений о старообрядчестве» (на 1 января 1912 г.), изданных в 1913 г. Департаментом духовных дел МВД. Общее количество старообрядцев (без единоверцев) для всей России в этом издании определено как 2 206 621 человек. Вместе с тем в научной литературе еще в конце XIX века для определения фактического количества старообрядцев утвердилось представление о необходимости увеличения официальных данных практически в 11 раз [Поздеева 1995: 7].

Для определения исторической роли старообрядцев, хранителей старины, необходимо знать «количество той силы, которою раскол владеет» [Поздеева 1995: 6]. Но старообрядцы, как показывают источники, обладали не только «силой количественной», но прежде всего — и это главное — силой духовной. Как определить высоту и степень духовной силы? Для этого необходимо включить в спектр анализа новый источниковый материал, а именно весь круг календарно-уставного чтения старообрядцев — как литурги-