



OSTIS-2011

(Open Semantic Technologies for Intelligent Systems)

УДК 004.8

СЕМАНТИЧЕСКАЯ ТЕХНОЛОГИЯ КОМПОНЕНТНОГО ПРОЕКТИРОВАНИЯ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ИНТЕРФЕЙСА ИНТЕЛЛЕКТУАЛЬНЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

В.А. Житко (*minsk.nes@gmail.com*)

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

В. Н. Вяльцев (*vltsvn@gmail.com*)

Ю. С. Гецевич (*mix1122@gmail.com*)

*Объединенный институт проблем информатики НАН Беларуси,
г.Минск, Республика Беларусь*

А.А. Кузьмин (*kuzaleksaleks@gmail.com*)

Белорусский государственный университет, г.Минск, Республика Беларусь

Аннотация: в данной статье рассматривается семантическая технология проектирования естественно-языковых интерфейсов для интеллектуальных вопросно-ответных систем. Данная технология развивается в рамках открытого проекта OSTIS [OSTIS, 2010]. Также рассматривается библиотека компонентов проектирования естественно-языкового интерфейса, ее пополнение, как сторонними компонентами, так и создание новых компонентов.

Ключевые слова: анализ естественно-языковых текстов, анализ речевого ввода, естественно-языковой интерфейс, речевой синтез, семантическая технология, технология проектирования

Введение

В связи с динамичным развитием и распространением компьютерных систем возникает необходимость в снижении расходов на подготовку новых пользователей. Привлекательно, в этом случае, выглядит использование привычного для пользователя языка для организации его диалога с компьютерной системой. Такая возможность реализуется средствами естественно-языкового пользовательского интерфейса, обладающего рядом преимуществ: минимальной подготовкой пользователя необходимой для работы с системой, простотой и высокой скоростью задания произвольных запросов к пользовательскому интерфейсу и высоким уровнем модели предметной области. Использование речевого синтеза позволит снизить нагрузку на пользователя, а также выполнять действия под руководством интеллектуальной системы, например решение задачи на построение геометрических чертежей.

Целью данной работы является создание семантической технологии компонентного проектирования естественно-языковых интерфейсов интеллектуальных вопросно-ответных систем. Такая технология основывается на семантической технологии компонентного проектирования пользовательских интерфейсов [OSTIS, 2010], разрабатываемой в рамках открытого проекта OSTIS. Таким образом, пользовательский естественно-языковой интерфейс будет представлять собой набор взаимодействующих между собой модулей. Такие модули мы

будем называть ip-компоненты, т.е. компоненты интеллектуальной собственности. Все модули располагаются в библиотеке совместимых ip-компонентов. Такая библиотека осуществляет хранение ip-компонентов, предоставляет разработчику доступ к информации о хранимых компонентах, а также средства поиска нужных компонентов.

1. Анализ естественно-языковых интерфейсов

Использование естественного языка для организации диалога пользователя с компьютерной системой сопряжено с рядом проблем: неоднозначность естественного языка, несоответствие возможностей реализации естественно-языкового интерфейса ожиданиям пользователя.

Для естественно-языкового пользовательского интерфейса вопросно-ответных систем возможно использование ограниченного набора лексики и грамматики, без серьёзного ущерба функциональности вопросно-ответной системы. В таком случае стоит говорить о ограниченном подмножестве естественного языка. Ограниченный естественный язык – это подмножество естественного языка, текст на котором, без каких либо усилий, воспринимается носителем исходного естественного языка, а так же не требует длительного изучения для приобретения навыков составления текстов этого языка, однако обладает сокращенным набором лексики и грамматики. Это позволит снизить время обработки естественно-языковых конструкций, а также устранить часть лингвистической неоднозначности.

Пользователь, впервые столкнувшись с естественно-языковым интерфейсом, может иметь завышенные или заниженные ожидания от естественно-языкового пользовательского интерфейса. Сравнительный анализ типов пользовательских интерфейсов (основанных на формах, с формальным языком запросов, графические пользовательские интерфейсы) показывает, что в целях построения естественно-языковых пользовательских интерфейсов превалирует желание максимально приблизить интерфейс к потребностям неподготовленного пользователя. Это несколько поднимает планку требований к дружелюбности и надежности естественно-языковых пользовательских интерфейсов, поскольку пользователь, впервые столкнувшись с системой, понимающей естественный язык, слабо представляет возможности системы. При этом ожидания к степени понимания естественного языка может отличаться от реальных способностей системы в обе стороны - т.е. пользователь может спрашивать систему о том, чего она "не знает", а может "по привычке" использовать простейшие формулировки запросов.

Для решения этой проблемы может быть использована «обратная связь». Пользователь, задавая вопрос программной системе, имеет возможность видеть результат разбора своего запроса и представление его на формальном языке. Таким образом, пользователь, методом «проб и ошибок», во время работы с вопросно-ответной системой, приспосабливается более эффективно использовать все возможности естественно-языкового интерфейса. В качестве целевого семантического языка используется специализированный предметно независимый язык вопросов. Язык вопросов обладает большей семантической мощностью, по сравнению с языками запросов к базам данных и знаний, что упрощает анализ естественно-языковых пользовательских запросов.

Семантическая (смысловая) неоднозначность может решаться за счет того, что предметная и лингвистическая база знаний находятся в едином информационном пространстве. У каждого понятия и отношения в предметной базе знаний есть естественно-языковые идентификаторы, что составляет первичную связь предметных и лингвистических знаний. Построения развернутых подобных связей является задачей разработчика лингвистической базы знаний конкретной вопросно-ответной системы. Возможность задействовать предметные знания при разборе естественно-языковых конструкций позволяет решать проблемы смысловой неоднозначности. К таким знаниям относятся область действия отношений, та область понятий, на которой задано данное отношение, классификация понятий и отношений, и др.

2. Общая модель естественно-языкового интерфейса

Естественно-языковой пользовательский интерфейс, в рамках разрабатываемой технологии, рассматривается как специализированная интеллектуальная система, обеспечивающая диалог между прикладной вопросно-ответной системой и пользователем. Являясь интеллектуальной

системой, естественно-языковой интерфейс включает в себя базу знаний, машину обработки знаний и пользовательский интерфейс.

Функционал естественно-языкового интерфейса можно выделить следующий:

- возможность ввода сообщения в естественно-языковой форме посредством текста или речи; трансляция сообщения пользователя на внутренний язык интеллектуальной системы;
- трансляция ответа системы в тексты естественного языка в форме текста или речи.

Данный функционал естественно-языкового интерфейса интеллектуальной справочной системы определяет структуру интеллектуальной системы естественно-языкового интерфейса. Структура естественно-языкового интерфейса должна состоять из следующих составляющих: пользовательский интерфейс, посредством которого происходит ввод сообщений пользователем и вывод ответа системы пользователю; трансляторы естественно-языковых запросов на sc-язык вопросов; трансляторы sc-конструкций на естественный язык.

Общая структура естественно-языкового пользовательского интерфейса представлена на рисунке 1.

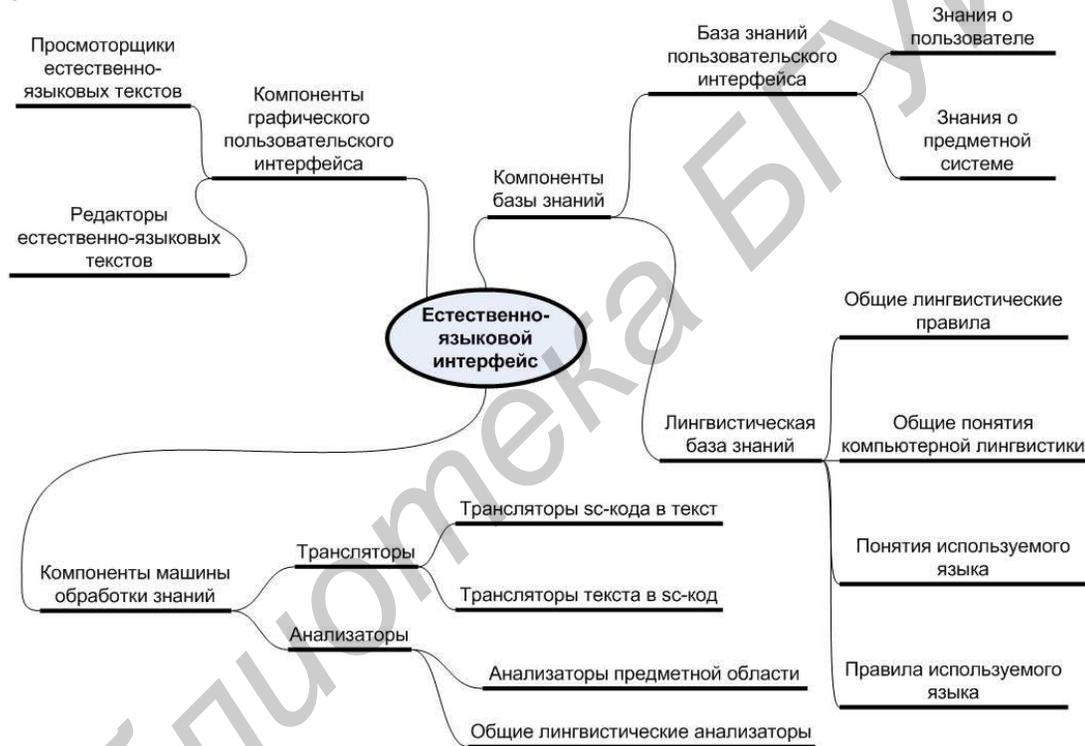


Рисунок 1 - Структура естественно-языкового пользовательского интерфейса.

Машина обработки знаний естественно-языкового интерфейса включает в себя операции обеспечивающие обработку различных лингвистических конструкций, перевод внешних языков на семантические языки интеллектуальной системы и обратно.

Все компоненты машины обработки знаний естественно-языкового интерфейса можно разделить на трансляторы и анализаторы. Задачей трансляторов является перевод знаний из одного языка представления знаний в другой, к примеру, транслятор фактографических знаний по предметной области в текст на естественном языке. Задачей анализаторов является анализ фрагментов знаний и выявления ранее неизвестных фактов, к примеру, анализатор запроса пользователя направленный на поиск цели и задачи вопроса.

Трансляторы естественно-языкового интерфейса, в минимальной конфигурации, должны включать в себя транслятор ограниченного естественного языка на язык вопросов и транслятор фактографических знаний на ограниченный естественный язык.

Трансляторы естественно-языкового пользовательского интерфейса должны осуществлять обработку информационных структур полагаясь исключительно на описание синтаксиса

используемого языка и описание семантики предметной области [Апресян, 1995]. Описание синтаксиса используемого языка записывается в лингвистической и предметной базах знаний. В лингвистической базе знаний так же должны быть описаны специфические для данной области знания о языке. Такой фрагмент базы знаний может представлять собой отдельный ир-компонент, т.к. он представляет ценность только в совокупности с предметной и лингвистической базой знаний.

Цикл работы естественно-языкового интерфейса начинается с ввода пользователем сообщения на естественном языке, посредством ввода текста или голосового ввода. По введенному в систему тексту строится его формальное отображение в памяти системы. Все предшествующие результаты анализа используются в анализе последующих запросов, что позволяет системе сохранять ход диалога с пользователем и разрешать спорные моменты, связанные с использованием одних и тех же терминов в разных предметных областях.

Первым этапом анализа запроса пользователя является морфологический анализ. На данном этапе для каждой словоформы, в формальной записи предложения, строятся отношения в соответствии значений её грамматических категорий, т.е. совокупности грамматических значений (род, падеж, склонение и т.д.). Грамматические категории могут присутствовать в базе знаний естественно-языкового интерфейса явно (словоформа присутствует в словаре с описанием её грамматических категорий), либо грамматическая категория может выявляться на основании морфологических правил используемого естественного языка. Такой подход позволяет разработчику, при проектировании естественно-языкового интерфейса, выбирать между быстродействием и объемом проектируемой базы знаний.

На этапе лексического анализа строятся отношения синтаксических связей внутри предложения, выделяются главные и второстепенные члены предложения, выявляется тип предложения и т.д. Такой анализ протекает поэтапно “наращивая” формальное описание структуры предложения, используя информацию, полученную на этапе морфологического анализа. На этом этапе используются лексические и синтаксические правила используемого языка. В общем случае, такие правила описывают критерии существования синтаксических связей в предложении, при анализе предложения проверяется соответствие таким критериям, и в зависимости от результата создаются либо удаляются соответствующие связи в предложении. Правила синтаксического разбора предложения в лингвистической базе знаний записываются, как приведено на рисунке 2.

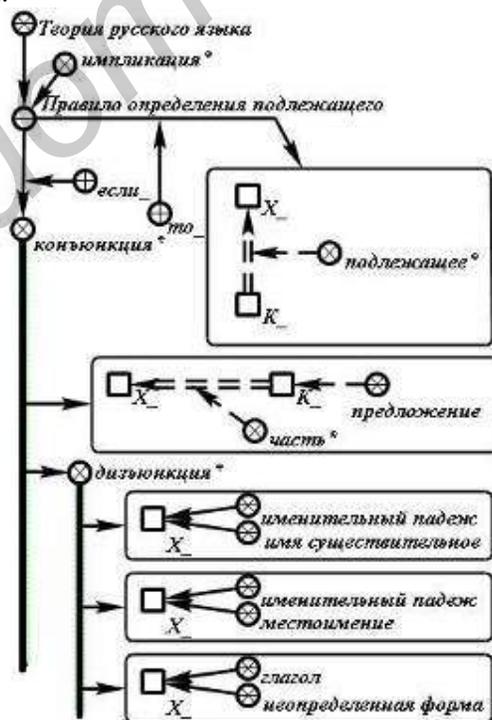


Рисунок 2 - Запись синтаксического правила в лингвистической базе знаний.

Результатом этапа лексического анализа является формальное описание лексического разбора предложения в памяти естественно-языкового интерфейса. На следующем этапе, семантический анализ, строятся отношения эквивалентности элементов предложения и узлов в памяти системы.

Данный этап является наиболее сложным этапом обработки естественно-языкового текста. В рамках используемой технологии, семантический анализ, представляет собой обработку семантической сети, отражающую результат анализа естественно-языкового текста на всех предыдущих этапах, а так же присутствующих в системе знаний о языке и предметной области основной системы. На этом этапе происходит соотношение лингвистической конструкций и конструкций в памяти системы, для выявления отношения соответствия, эквивалентности и т.д.

Результатом семантического анализа является конструкция запроса к системе. Сложность в получении такого результата заключается в семантической неполноте исходного естественно-языкового вопроса. Например, вопрос «В какой класс многоугольников входит понятие квадрата?» в развернутом виде будет выглядеть следующим образом: «Какое понятие является подмножеством понятия многоугольника и надмножеством понятия квадрат?». Из примера видно, что, в процессе развертки, произошли следующие подстановки: высказывание «входит» было развернуто в «является подмножеством», «квадрата» - «понятие квадрата», «класс многоугольников» - «понятие, являющееся подмножеством понятия многоугольника», «многоугольник» - «понятие многоугольника».

Вопросы, сгенерированные естественно-языковым интерфейсом, обрабатывает универсальный решатель вопросно-ответных систем, являющийся частью вопросно-ответной системы по соответствующей предметной области. Стоит заметить, что универсальный решатель использует знания и лингвистической базы знаний, таким образом, система может отвечать на вопросы связанные с используемым естественным языком.

3. Библиотека совместимых ip-компонентов естественно-языкового пользовательского интерфейса

Главным элементом семантической технологии компонентного проектирования естественно-языковых пользовательских интерфейсов является библиотека совместимых ip-компонентов. Такая библиотека включает в себя лингвистические базы знаний по разным языкам, различные трансляторы и анализаторы естественных текстов, элементы пользовательского интерфейса. Это позволяет проектировать естественно-языковые интерфейсы комбинируя уже существующие компоненты, выбирая нужные лингвистические базы знаний, анализаторы и синтезаторы. Задачей разработчика в данном случае является привязка естественно-языкового интерфейса к предметной области интеллектуальной системы. Для этого необходимо добавить лингвистические знания о специфических для этой предметной области понятиях.

3.1. Грамматический словарь русского языка

Грамматический словарь русского языка является важным компонентом естественно-языкового интерфейса. Грамматические знания о естественном языке, записанные в лингвистической базе знаний, используются для грамматического анализа естественного текста, а также для его синтеза. Эта часть лингвистической базы знаний выделяется в отдельный ip-компонент и может использоваться в качестве предметной базы знаний по грамматике русского языка. Запись морфологической информации о конкретном слове в лингвистической базе знаний записывается, как показано на рисунке 3.

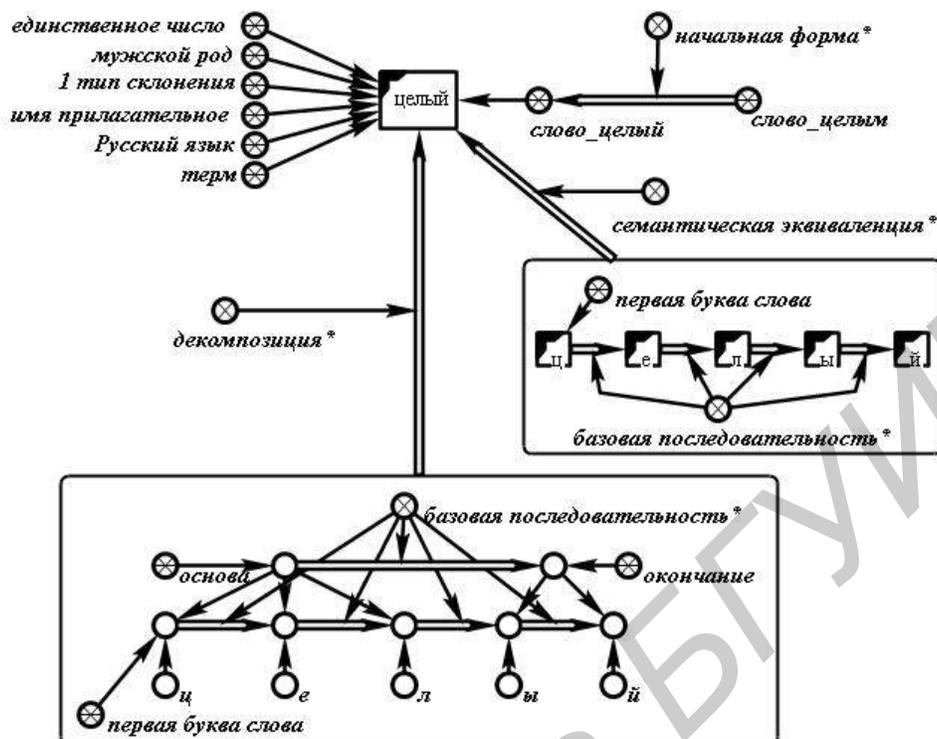


Рисунок 3 - Запись морфологической информации в лингвистической базе знаний.

Для каждого слова ставятся в соответствие грамматические категории (число, род, склонение и т.д.). Для этого данное слово включается во множество конкретной грамматической категории (единственное число, мужской род и т.д.). Каждое конкретное слово включено во множество его грамматических синонимов. Что позволяет быстро получать доступ ко всем наборам грамматических категорий при грамматическом анализе.

Для каждого слова стоит его разбиение на морфемы, а также эквивалентная запись в семантической сети. Первое используется при словоизменении.

На множестве слов задаются отношения синонимии, антонимии и пр. Что может использоваться при семантическом анализе естественно-языкового текста.

3.2. Речевой синтезатор русского и белорусского языков

Одним из *ip*-компонентов, расширяющих функционал естественно-языкового интерфейса, является синтезатор речи по тексту. Данный *ip*-компонент реализован в качестве стороннего подключаемого модуля. Отметим что, внутренняя работа компонента не связана с семантическими сетями и взаимодействие с системой происходит с помощью трансляторов, но в результате такого подключения к системе синтеза речи по тексту пользователь получит озвученный ответ от системы. Таким образом, система естественно-языкового интерфейса станет еще более «естественной» для пользователя.

Синтез устной речи по тексту осуществляется на основе лексико-грамматического анализа входного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования. Орфографический текст поступает на вход синтезатора и далее подвергается последовательной обработке рядом специализированных процессоров в соответствии с общей структурой синтезатора речи по тексту, представленной на рисунке 4. Синтезатор включает модули: текстовый процессор, просодический процессор текста и сигнала, фонетический процессор и акустический процессор. Каждый из этих модулей поддерживается наборами соответствующих баз данных и правил.



Рисунок 4 - Структура системы синтеза речи по тексту

Входной орфографический текст подвергается ряду последовательных обработок в каждом из процессоров. Текстовый процессор обрабатывает входной орфографический текст в следующей последовательности: очистка текста, преобразование знаков (аббревиатур, сокращений, чисел и др.), расстановка словесных ударений и грамматических признаков словоформ.

Преобразованный текст поступает на входы просодического, а затем фонетического процессоров. В результате работы просодического подпроцессора для текста, текст разделяется на синтагмы, акцентные единицы, далее он разделяется на элементы акцентных единиц: интонационное предъядро, ядро и заядро. И наконец, просодический подпроцессор для языкового сигнала - устанавливает в соответствии с базой данных просодических "портретов" синтагм значимости амплитуды, продолжительности фонем и частоту основного тона для каждого элемента акцентных единиц.

Задача фонетического процессора состоит в преобразования орфографического текста в фонемный, а также в генерации позиционных и комбинаторных аллофонов.

Акустический процессор на основе информации про то, какие необходимо синтезировать аллофоны, а также какие просодические характеристики должны быть приписанные каждому аллофону, генерирует речевой сигнал путем компиляции отрезков натуральных звуковых волн соответствующих аллофонов и мультифонов. Таким образом, входной текст преобразуется в речевой сигнал.

На рисунке 5 представлена структура системы синтеза речи по тексту.

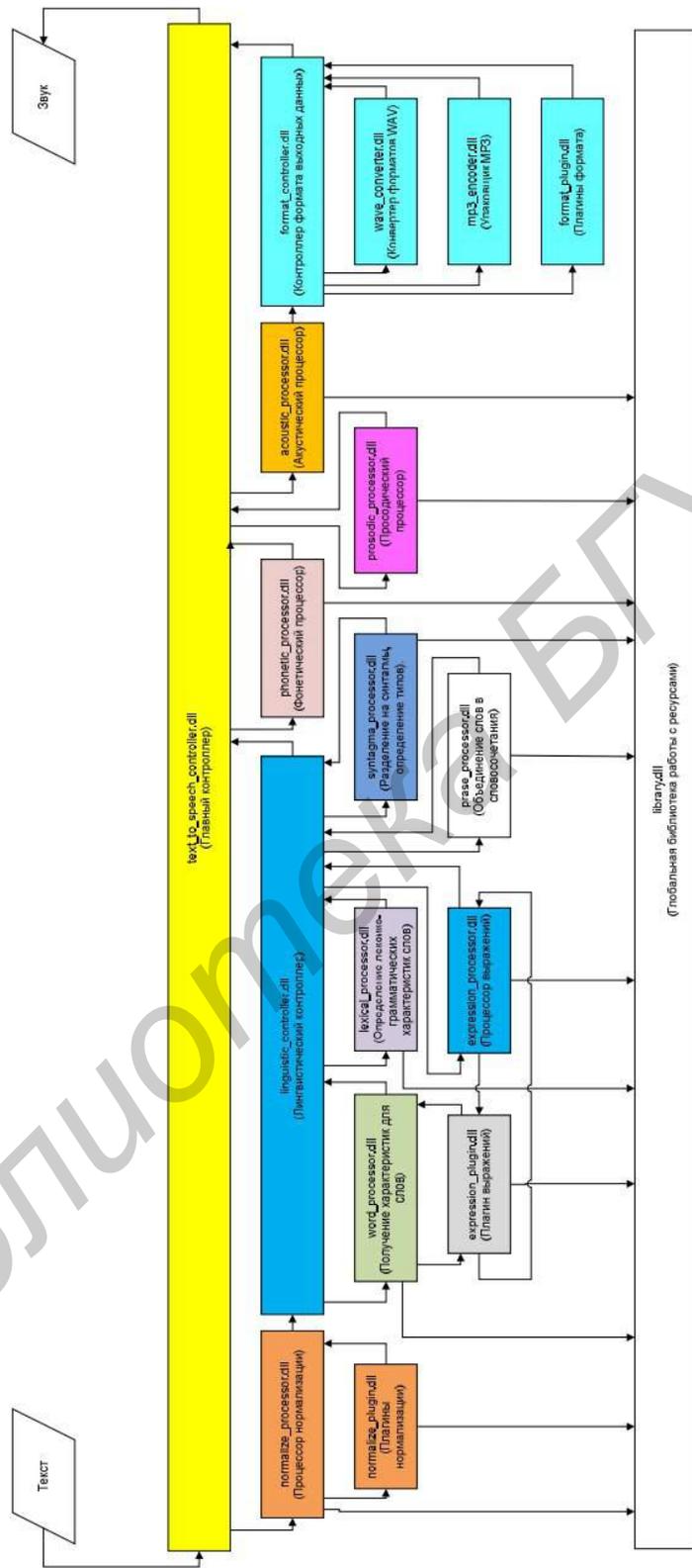


Рисунок 5 - Проектная структура системы синтеза речи по тексту

Разработанная архитектура системы синтеза речи по тексту содержит в себе пятнадцать модулей, описание которых приводится ниже.

Главный модуль системы синтеза речи по тексту управляет работой всех остальных модулей, определением списка и загрузкой плагинов к остальным модулям, передачей данных между модулями.

Процессор нормализации текста перед лингвистической обработкой, производит удаление из текста символов, не нужных для синтеза речи, удаляет случайное дублирование знаков препинания, заменяет похожие символы на один из них. Подключаемые плагины производят дополнительное преобразование поступающих данных перед подачей их на нормализацию, так, может быть, например, преобразование PDF в простой текст.

Главный модуль лингвистического процессора управляет другими модулями лингвистического процессора и контролирует преобразование этими модулями текста в последовательность синтагм. **Процессор слов** определяет возможные лексико-грамматические характеристики слова (последовательности символов, отделенных пробелами и, возможно, содержащими пробел). **Лексико-грамматический процессор** определяет лексико-грамматические характеристики слова на основе вариантов, предложенных предыдущим процессором и лексико-грамматических характеристик других слов в тексте. **Плагины дополнительных выражений** производят обработку специальных выражений (например, чисел или сокращений) и преобразование их в их словесное выражение. **Процессор выражений** расставляет ударения в словах, присоединяет к словам предлоги и частицы, заменяет выражения, обрабатываемые плагинами на результат обработки. **Процессор сборки словосочетаний** соединяет отдельные слова в словосочетания, исходя из лексико-грамматических характеристик этих слов для недопущения разделения слов границей синтагмы. **Процессор деления на синтагмы** разделяет поток словосочетаний из предыдущего процессора на синтагмы с указанием интонационного типа синтагмы в зависимости от знаков препинания и лексико-грамматических характеристик слов в словосочетаниях.

Фонетический процессор производит преобразование последовательности букв, из которых состоит синтагма в последовательность фонем.

Просодический процессор производит определение просодических характеристик (частоты основного тона, длительности, амплитуды сигнала) для каждой фонемы в последовательности исходя из интонационного контура, определяемого типом синтагмы.

Акустический процессор соединяет аллофоны, определяемых фонемами, изменяет просодические параметры аллофонов, формирует звуковой сигнал. **Контроллер преобразования звуковых форматов** управляет плагинами, преобразующими звук. Такие плагины производят преобразование звука в различные форматы (изменение частоты дискретизации, точности передачи звука, упаковка MP3).

Таким образом, спроектированная архитектура позволит разработать качественно новый синтезатор речи по тексту с высокой степенью «лингвистического понимания» входного текста и генерацией речи для самого широкого круга потребителей.

3.3. Разработка компонента речевого ввода информации на основе пакета НТК

Сегодняшний информационный мир предоставляет огромные возможности. Тем не менее, взаимодействие между человеком и компьютерным устройством по-прежнему базируется в основном на различных формах графического интерфейса. Но для повсеместной интеграции в нашу повседневную жизнь необходимы новые виды взаимодействия.

Если учесть, что после зрения вторым по объёму принимаемой информации органам чувств является слух, то вполне разумно выглядит использование речевой коммуникации в качестве дополнения к графическим интерфейсам. В сфере разработки альтернативных способов взаимодействия речевая коммуникация играла главенствующую роль на протяжении последних десятилетий. Идея дополнить традиционные устройства ввода голосовым контролем, повысив тем самым комфорт и скорость ввода, кажется очень привлекательной. Однако до последнего времени самыми распространенными примерами использования звука в интерфейсе можно считать озвучивание некоторых операций набором слов либо звуков из стандартного набора. В наши дни наблюдается тенденция к ускорению прогресса в этой сфере по следующим причинам. Во-первых, речевые технологии достигли нового уровня развития. В основном за

счёт использования метода стохастического моделирования. Во-вторых, по сравнению с другими альтернативными видами взаимодействия человека с компьютером, речь стала приоритетной.

Таким образом, очевидным выглядит факт, что после интерфейса командной строки (1960 – 80-е г.г.), и графического интерфейса (1980 – 2000-е г.г.), будущее принадлежит комплексному пользовательскому интерфейсу, позволяющему задействовать, кроме зрения, разные органы чувств человека, в первую очередь – его слух. Многие факты говорят в пользу этого утверждения:

- речь – наиболее популярная форма коммуникации между людьми;
- нет необходимости в непосредственном контакте при взаимодействии, поскольку микрофон и динамики могут располагаться на расстоянии;
- руки и глаза остаются свободными, что делает голосовой интерфейс приоритетным в некоторых ситуациях, таких как, например, процесс вождения транспортного средства либо параллельного использования нескольких приложений одновременно;
- на данный момент в мире существует порядка 1,3 миллиарда мобильных телефонов, что в пять раз превышает количество компьютеров, подключенных к интернету. Это обеспечивает огромный рынок для будущих систем автоматического диалога.

В свете всех перечисленных тенденций вполне закономерным и актуальным выглядит разработка речевых способов управления и для поддержки средств навигации и поиска в семантических сетях. Как правило, от таких приборов требуется, с одной стороны, обеспечивать обработку большого объёма запросов, а с другой – в процессе использования как можно меньше отвлекать пользователя от объекта поиска. При этом в неизбежно растёт число элементов графического меню, что в свою очередь замедляет время поиска нужной опции и заставляет отвлекаться от искомого объекта [Landauer, 1985].

Речевой способ взаимодействия – наиболее естественный интерфейс для общения человека с человеком. Это влечет за собой простоту изучения и использования речевого интерфейса при взаимодействии с базами знаний. Обсуждаемый вид взаимодействия позволяет освободить руки и глаза человека при подаче команд. Пользователь может осуществлять запросы (устно отдавая соответствующие команды) в процессе работы, передвижения или манипуляции другими объектами. Дополнительный комфорт интерфейса становится следствием отсутствия необходимости в специальных устройствах, таких как мышь, палочка или перчатки данных. Таким образом, широкому кругу пользователей, включая пожилых людей и инвалидов, удобно приспосабливаться к речевому интерфейсу.

Целью данной работы является создание модуля распознавания слитной речи призванного обеспечить использования голосового ввода для осуществления взаимодействия с интеллектуальной вопросно-ответной справочной системой. Основной подход подобных интеллектуальных систем основывается на семантической технологии компонентного проектирования интеллектуальных справочных систем [OSTIS, 2010], разрабатываемой в рамках открытого проекта OSTIS.

Проблемы создания подобного речевого дополнения к интерфейсу традиционны для систем распознавания речи: человеческая речь представляет из себя нестационарный сигнал весьма сложный для анализа, большая зависимость физических характеристик сигнала от диктора (состояние голоса, тембр голоса, гендерная принадлежность диктора, язык ввода и т.д.)

К способам, призванным решить эти проблемы относится исторически первый подход к распознаванию речи, основанный на технологии динамическом искривлении времени (Dynamic time warping (DTW)), а также практически заменившая его в последствии технология, основанная на скрытых марковских моделях (СММ). По причине высокой эффективности распознавание речи с использованием математического аппарата СММ реализовано во многих приложениях таких как пакет НТК, система Sphinx, а также отдельный модуль среды инженерного программирования MatLab -- НММ Toolbox.

Именно СММ были выбраны как теоретическая основа для создания модуля распознавания речевых запросов. В качестве набора инструментов, реализующего все основные функции и алгоритмы был использован пакет НТК. Данное приложение помимо распознавания отдельных слов позволяет на основе возможных последовательностей членов предложения реализовывать интерпретацию слитной речи, что очень хорошо подходит для ограниченного набора запросов.

3.3.1. Задача обработки сигналов. Для успешного решения проблемы распознавания речевых сигналов большую роль играет выделение их информативных признаков. Они должны удовлетворять нескольким критериям [Бовбель, 1998] [Rabiner, 1993]:

быть легко вычисляемыми;

сохранять всю необходимую информацию, которая содержится в сигнале;

в режиме обучения на основании обучающего множества векторов признаков создавать модель, которая будет более общей, чем обучающий материал.

В качестве таких признаков использовались кепстральные параметры, т. к. они наилучшим образом подходят для работы с речевыми сигналами и являются широко распространённым представлением спектральных характеристик в задачах обработки подобных сигналов [Орпенгейм, 2004]. В данной работе использовались следующие шаги расчёта мел-частотных кепстральных коэффициентов [Бовбель, 2008]:

Предисказание. На первом этапе речевой сигнал обрабатывался цифровым фильтром первого порядка. Целью этого фильтра является усиление энергии на высоких частотах, которая обычно уменьшается при генерировании речевого сигнала.

Сегментация. На втором шаге сигнал сегментировался на статистически однородные блоки. Во время формирования вектора признаков предполагалось, что сигнал можно рассматривать как квазистационарный процесс, состоящий из последовательности фреймов, которые могут обрабатываться независимо.

Взвешивание. Выигрышным является умножение каждого кадра на весовую функцию для минимизации нежелательных концевых эффектов. В данной работе в качестве такой функции использовалось окно Хемминга.

Спектральный анализ. После разбиения сигнала на фреймы, вычислялось БПФ [Опенгейм, 1979]. После получения спектра по причине его симметричности используется только половина получившихся данных.

Мел-шкала. Слуховая система человека не является линейной относительно спектра аудио сигнала и практика показывает, что использование такого нелинейного подхода способна увеличить качество распознавания. Наибольший интерес представляет не детальное представление сигнала, а форманты -- спектральные максимумы. Именно они дают основную (необходимую) информацию для распознавания речи и верификации дикторов. Для выделения формант производится сглаживание спектра путём применения набора полосовых фильтров. Для локализации фильтров использовалась мел-шкала.

Операция логарифмирования. Поскольку воспринимаемая громкость сигнала приблизительно логарифмическая, то на выходе каждого рассчитывается усреднённое значение E_j и выполняется операция логарифмирования: $m_j = \log E_j$.

Дискретное косинусное преобразование (ДКП). Полученные коэффициенты получают достаточно сильно коррелированы, таким образом очень важным является использование кепстрального преобразования для последующего использования в СММ с диагональными матрицами ковариации.

3.3.2. Тестовые диалоги для интерфейса семантических баз знаний. При создании прототипа речевого интерфейса для осуществления различного рода запросов в семантических базах знаний были выбраны ряд тестовых диалогов пользователя с системой с целью выбора необходимых в общении слов и фраз для последующего создания СММ с соответствующими параметрами. Примером такого рода диалогов могут служить следующие ситуации:

Вопрос: Что это такое (В окне геометрического редактора выделяется некоторая фигура)?

Ответ: Это треугольник со сторонами a, b и углом C, равным 45 градусам, между ними.

Вопрос: Как они связаны (В окне системы выделяются понятия треугольника и тригонометрии)?

Ответ: Эти два понятия связаны в теореме синусов и теореме косинусов.

Вопрос: Что из этого следует?

Ответ: Это дает возможность расчета численных характеристик конкретного треугольника.

Вопрос: Приведите пример.

Ответ: Зная длины двух сторон и значение угла между ними, можно рассчитать длину третьей стороны.

Вопрос: Приведете примеры другой теоремы.

Ответ: Сумма углов треугольника равна 180 градусам.

И т. д.

Из существующего ограниченного набора запросов во-первых: был сформирован набор базовых слов, сгруппированный по назначению:

• Язык вопросов:

- Вопрос
- Ответ
- Что это такое?
- ...

• Предметная область:

- Треугольник
- Произвольный треугольник
- Прямоугольный треугольник
- Вершина
- ...

Во-вторых, сформированы стандартные синтаксические последовательности этих слов для перехода от распознавания отдельных слов к распознаванию слитной речи.

Алгоритм создания набора СММ для распознавания

Полный цикл создания хорошо обученных СММ включает два основных этапа: подготовка данных для обучения и непосредственно самообучение (Рис. 6). Во многом успех распознавания зависит именно от качества обучающей информации. Поэтому этой части работы уделялось особенно много внимания.

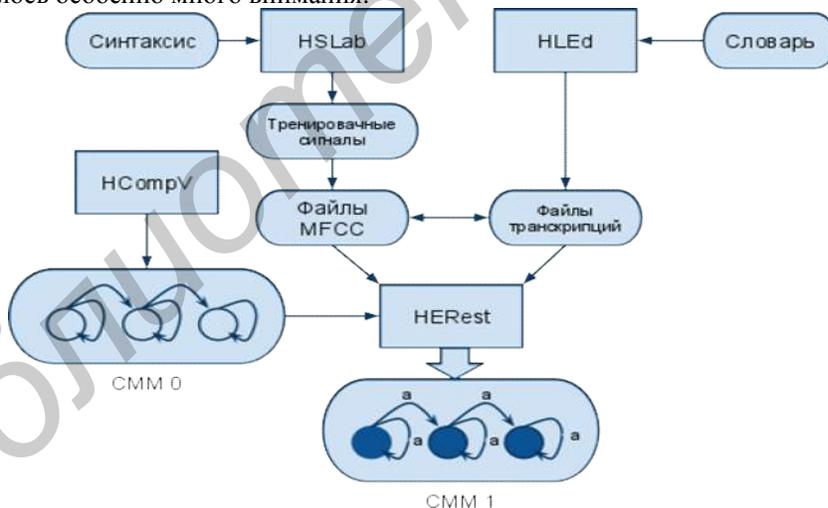


Рисунок 6. Алгоритм создания набора СММ для распознавания

Подготовка данных.

Шаг 1 -- Грамматика. Первым шагом было создание грамматики. Так как система должна распознавать запросы к базе знаний, то распознаватель обязан идентифицировать последовательность ключевых слов, выстроенных согласно грамматической последовательности в запросах.

Шаг 2 -- Словарь. Здесь первым делом создавался отсортированный список всех необходимых для распознавания слов. Далее запись каждого слова расширилась фонетической транскрипцией.

Шаг 3 -- запись данных. На данном этапе осуществлялось создание набора файлов формата .wav, содержащих несколько вариантов (не менее трёх) записей базовых слов, произнесённых диктором. В качестве инструмента выступала функция HSLab из пакета НТК, которая позволила не только записать данные, но и разметить их по содержащимся фонемам. Следует отметить, что чем короче слово, тем большее количество вариантов требуется для более качественного распознавания.

Шаг 4 -- Генерация файла транскрипций. Для обучения наборы СММ каждому файлу параметров тренировочных данных должна быть поставлена в соответствие фонетическая транскрипция. Этот процесс был автоматизирован за счёт использования инструмента HLEd, который с помощью словаря транскрипций слов создаёт файлы с расширением .mlf (Maser Label File) соответственно для каждого файла содержащего запись речевого сигнала.

Шаг 5 -- кодирование данных. Финальным шагом в подготовке данных является обработка речевых сигналов и преобразование их в последовательности векторов признаков. Как и было упомянуто выше в данной работе в качестве таких векторов были использованы кепстральные коэффициенты шкалы мел-частот. Обработка сигнала производилась согласно алгоритму, описанному в разделе “Задача обработки сигналов”.

Помимо 13 кепстральных коэффициентов набор параметров также содержал 13 дельта-коэффициентов и 13 коэффициентов ускорения. Таким образом, размерность вектора признаков составила величину, равную 39. Для хранения векторов признаков использовался специальный формат .MFCC (Mel Frequency Cepstral Coefficients -- от англ. Кепстральные коэффициенты шкалы мел-частот).

Обучение.

С этого момента начинается создание набора хорошо обученных СММ, все вероятности в которых описываются единичными Гауссианами.

Шаг 6 -- Создание начальных монофонов. Первым шагом в создании СММ является определение модели прототипа. Для системы основанной на звуках хорошей топологией является лево-правая с тремя состояниями. Инструмент HCompV просматривает файлы с параметрами сигналов и рассчитывает одинаковые начальные средние и дисперсии для всех гауссианов модели.

Шаг 7 -- Пересчёт коэффициентов. Параметры модели затем пересчитывались по методу Баума-Уэлча (алгоритм прямого-обратного хода или алгоритм максимизации правдоподобия). Пересчёт производился с помощью функции HERest.

В итоге были созданы три комплекта Моделей: монофонные, трифонные, а также трифонные со связанными состояниями. Каждый из комплектов применяется для распознавания в прототипе системы.

Заключение

В заключение следует отметить, что за счёт использования передовых теорий и инструментов в сфере распознавания слитной речи, удалось создать надёжный распознаватель речевых сигналов, что существенным образом расширило и преобразило интерфейс взаимодействия пользователя с базой знаний в целом. Созданное дополнение качественным образом изменяет способ, а следовательно и эффективность взаимодействия пользователя с поисковой системой. Позволяет сделать работу с приложением более естественной для человека.

Заключение

Семантическая технология компонентного проектирования естественно-языковых интерфейсов вопросно-ответных систем предоставляет единую основу для анализа естественно-языковых текстов, начиная с ввода текста, морфологического, лексического, семантического анализа и генерации текста. Так же, позволяет устанавливать отношения между эквивалентными лингвистическими конструкциями и структурой внутренней памяти системы.

Библиотека совместимых ip-компонентов естественно-языкового пользовательского интерфейса позволяет разработчику проектировать естественно-языковой интерфейс посредством компоновки уже готовых компонентов, включая лингвистические базы знаний, трансляторы и элементы пользовательского интерфейса. Ip-компонент системы синтеза речи по тексту позволяет пользователю слышать ответ от системы на поставленный вопрос, а не

просто читать его с экрана компьютера. Это делает естественно-языковой интерфейс еще более естественным для пользователя.

Возможность интеграции сторонних разработок и проектов в качестве внешних ир-компонентов позволяет производить интеграцию различных подходов и методов в рамках одного проекта, что позволяет использовать их лучшие стороны.

Библиографический список

[Апресян, 1995] Апресян, Ю. Д. Избранные труды, том I. Лексическая семантика / Ю. Д. Апресян – Москва: Школа «Языки русской культуры», 1995.

[Апресян, 1995] Апресян, Ю. Д. Избранные труды, том II. Интегральное описание языка и системная лексикография / Ю. Д. Апресян – Москва: Школа «Языки русской культуры», 1995.

[Апресян, 1989] Апресян, Ю. Д. Лингвистическое обеспечение системы ЭТАП-2 / Ю. Д. Апресян [и др.] – Москва: Наука, 1989

[OSTIS, 2010] Open Semantic Technology for Intelligent Systems. [Электронный ресурс]. – 2010. - Режим доступа: <http://www.ostis.net/>. – Дата доступа: 30.11.2010

[Landauer, 1985] Landauer T.K. Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width / Т.К. Landauer, D.W. Nachbar. New York : ACM, 1985.

[Бовбель, 1998] Бовбель Е. И. Статистические методы распознавания речи: скрытые марковские модели / Е. И. Бовбель, И. Э. Хейдоров // Зарубежная радиоэлектроника. Успехи современной электроники. 1998. № 3. С. 36–54.

[Rabiner, 1993] Rabiner, L. R. Fundamentals of Speech Recognition / L. R. Rabiner, B. N. Juang. New Jersey, 1993.

[Oppenheim, 2004] Oppenheim, A. V. From Frequency to Quefrancy: A History of the Cepstrum / A. V. Oppenheim, R. W. Schafer // IEEE Signal Processing Magazine. 2004. Vol. 21. P. 95--106

[Бовбель, 2008] Бовбель Е. И. Скрытые марковские модели и машины на опорных векторах от теории к практике / Е. И. Бовбель, И. Э. Хейдоров, Ю. В. Пачковский. Минск БГУ, 2008

[Опенгейм, 1979] Опенгейм, А. В. Цифровая обработка сигналов / А. В. Опенгейм, Р. В. Шафер. М., 1979

[Белнап, 1981] Белнап, Н. Логика вопросов и ответов / Н. Белнап, С. Стил. - М., 1981.

[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.

[Kowalsky, 1975] Kowalsky, R. A. Proof procedure using connection graphs / R. A. Kowalsky // Journal of the ACM. – 1975 – № 22(4).

[Лобанов, 2008] Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи / Минск: Белорусская наука, 2008. – 344 с.: ил.

[Карпов, 1992] Карпов В.А. Язык как система / Минск.: Выш. шк., 1991.

[Карпилович, 1977] Алгоритмы порождения предложений естественного языка (обзор и анализ) / Т.П. Карпилович. – Минск, 1977.

[Кречетова, 1999] Кречетова, Т.В. Формальный аппарат лингвистических описаний для систем понимания текста на естественном языке / Т.В. Кречетова // сборник трудов 12 Международной научной конференции, Новгород, 1999 / Новгород. гос. ун-т. – Новгород, 1999

[Кустова, 2005] Кустова, Г.И. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы / Г.И. Кустова, О.Н. Ляшевская, Е.В. Падучева, Е.В. Рахилина // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. – Москва, 2005.

[Мельчук, 1974] Мельчук, И.А. Опыт теории лингвистических моделей «Смысл-Текст». Семантика, Синтаксис / И.А. Мельчук. – Москва, 1974

[Попов, 1982] Попов, Э.В. Общение с ЭВМ на естественном языке / Э.В. Попов. – Москва: Наука, 1982.

[Шенк, 1987] Шенк, Р. Обработка концептуальной информации: пер. с англ. / Р. Шенк. – Москва: Мир, 1987.

[Головня, 2007] Головня А.И. Омонимия как системная категория языка / Минск: Издательский центр БГУ, 2007.

[Лобанов и др., 2006] Лобанов, Б. М. Речевой интерфейс интеллектуальных систем: учебное пособие / Б. М. Лобанов [и др.]. – Минск: БГУИР, 2006.

[Голенков, 2001] Голенков, В. В. Представление и обработка знаний в графодинамических ассоциативных машинах / В. В. Голенков [и др.]. – Минск: БГУИР, 2001.

[Голенков, 2001] Голенков, В. В. Программирование в ассоциативных машинах / В.В. Голенков [и др.]. – Минск: БГУИР, 2001.