



OSTIS-2012

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

СЕМАНТИЧЕСКАЯ ТЕХНОЛОГИЯ ПРОЕКТИРОВАНИЯ БЕЛОРУССКО- И РУССКОЯЗЫЧНЫХ ЕЯ-ИНТЕРФЕЙСОВ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

Гецевич Ю.С.*, Гецевич С.А. *, Елисеева О.Е.** , Житко В.А.***, Кузьмин А.А.**

** Объединенный институт проблем информатики
Национальной академии наук Беларуси,
г. Минск, Республика Беларусь*

Yury.Hetsevich@gmail.com

Novaeimya@gmail.com

***Белорусский государственный университет,
г. Минск, Республика Беларусь*

olae@open.by

kuzAleksAleks@gmail.com

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

zhitko.vladimir@gmail.com

В работе рассматривается семантическая технология проектирования естественно-языковых и речевых интерфейсов для интеллектуальных вопросно-ответных систем. Данная технология развивается в рамках открытого проекта OSTIS [Ostis, 2012]. Рассматривается также библиотека компонентов проектирования естественно-языкового интерфейса, возможности ее пополнения сторонними компонентами и создания новых компонентов.

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

ВВЕДЕНИЕ

В связи с динамичным развитием и распространением компьютерных систем в различных сферах деятельности человека актуальным является снижение затрат на подготовку новых пользователей. Перспективным в этом случае представляется использование привычного для пользователя естественного языка для организации диалога с компьютерной системой. Такая возможность реализуется средствами естественно-языкового пользовательского интерфейса, обладающего рядом преимуществ: минимальной подготовкой пользователя для работы с системой, простотой и высокой скоростью задания произвольных запросов к пользовательскому интерфейсу и высоким уровнем модели предметной области. Использование речевого синтеза при этом позволяет снизить нагрузку на пользователя по восприятию результатов обработки не через графический интерфейс (зрительную систему), а

посредством речи. Кроме того, языковой ввод команд и вопросов позволяет пользователю использовать эту же систему в качестве справочной и получать от неё помощь в исполнении любой деятельности (получение справки при выполнении технических работ и т.д.).

Следует отметить, что для естественно-языкового пользовательского интерфейса вопросно-ответных систем возможно использование ограниченного набора лексики и грамматики языка без серьезного ущерба функциональности вопросно-ответной системы. Ограниченный естественный язык – это подмножество естественного языка, текст на котором без каких-либо усилий воспринимается носителем исходного естественного языка, а также не требует длительного изучения для приобретения навыков составления текстов на этом языке, т.к. обладает сокращенным набором лексики и грамматики. Это позволяет снизить время обработки естественно-языковых конструкций в вопросно-ответной

системе, а также частично избежать лингвистических неоднозначностей.

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, 2012], разрабатываемой в рамках открытого проекта OSTIS. В соответствии с указанной технологией пользовательский естественно-языковой интерфейс представляет собой набор взаимодействующих между собой модулей. Такие модули будем называть ip-компонентами, продуктами интеллектуальной собственности. Все модули разрабатываемого естественно-языкового интерфейса 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.

К ip-компонентам могут относиться также сторонние разработки и системы, например, системы распознавания и синтеза речи. В соответствии с этим в число задач данной работы включена также разработка методики интеграции сторонних продуктов в качестве независимых ip-компонентов. Данная задача подразумевает разработку универсальных и специализированных языков общения для всех ip-компонентов. Одним из таких базовых языков является семантический язык вопросов.

1. Исследование естественно-языковых интерфейсов

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

Как указывалось выше, для естественно-языкового пользовательского интерфейса вопросно-ответных систем возможно использование ограниченного естественного языка без серьёзного ущерба их функциональности. Благодаря использованию продуманных лексических и грамматических ограничений, принятых в рамках ограниченного естественного языка, появляется возможность избежать многих проблем, связанных с анализом неограниченного естественного языка. При этом язык общения пользователя с компьютерной системой по-прежнему остаётся вполне естественным, а процесс реализации этого общения становится более управляемым и в значительной степени упрощается.

Пользователь, впервые столкнувшись с естественно-языковым интерфейсом, может иметь

завышенные или заниженные ожидания от такой формы взаимодействия с компьютерной системой. Сравнительный анализ типов пользовательских интерфейсов (основанных на формах, с формальным языком запросов, графические пользовательские интерфейсы и пр.) показывает, что при построении пользовательских интерфейсов с использованием естественного языка у разработчиков преобладает желание максимально приблизить интерфейс к потребностям неподготовленного пользователя. Это несколько поднимает планку требований к степени дружелюбности и надёжности (безотказности) естественно-языковых пользовательских интерфейсов, поскольку пользователь, впервые столкнувшись с системой, "понимающей" естественный язык, слабо представляет реальные возможности системы. При этом его ожидания к степени понимания естественного языка могут отличаться от реальных способностей системы в обе стороны. Пользователь может спрашивать систему о том, чего она "не знает", а может "по привычке" использовать простейшие формулировки запросов.

Для решения обозначенной выше проблемы на начальных этапах развития естественно-языкового интерфейса может быть использована специально предусмотренная обратная связь с пользователем. В рамках такой обратной связи пользователю, задающему вопрос системе, предоставляется возможность видеть результат разбора этого запроса и его представление на формальном языке. Таким образом, пользователь, методом «проб и ошибок», во время работы с вопросно-ответной системой, может приспособиться (научиться) более эффективно использовать все возможности естественно-языкового интерфейса. В качестве целевого формального семантического языка используется специализированный предметно-независимый язык вопросов. Указанный 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 естественно-языкового диалога может разрешаться за счет того, что предметная и лингвистическая базы знаний создаются в рамках единого информационного пространства, единой предметной области. У каждого понятия и отношения в этом случае в предметной базе знаний имеются соответствующие естественно-языковые идентификаторы (слова естественного языка), и это составляет первичную связь предметных и лингвистических знаний. Построение более развернутых связей является задачей разработчика лингвистической базы знаний конкретной вопросно-ответной системы для конкретной предметной области.

2. Моделирование естественно-языкового интерфейса

Естественно-языковой пользовательский интерфейс в рамках разрабатываемой технологии компонентного проектирования рассматривается нами как специализированная интеллектуальная система, обеспечивающая диалог между прикладной вопросно-ответной системой и пользователем в заданной предметной области. Таким образом, будучи интеллектуальной системой, естественно-языковой интерфейс, как и любая другая подобная система, в качестве основных своих компонентов имеет базу знаний, машину обработки знаний и пользовательский интерфейс.

Определим здесь следующие основные функциональные возможности естественно-языкового интерфейса:

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

Указанный функционал определяет структуру интеллектуальной системы естественно-языкового интерфейса, которая включает следующие составляющие:

- пользовательский интерфейс, посредством которого происходит ввод сообщений пользователя и вывод ответа системы пользователю;
- трансляторы естественно-языковых запросов на внутренний sc-язык вопросов;
- трансляторы конструкций sc-языка (sc-конструкций) на естественный язык.

Общая структура интеллектуальной системы (ИС) естественно-языкового пользовательского интерфейса вопросно-ответных систем представлена на рисунке 1.



Рисунок 1 - Структура ИС естественно-языкового пользовательского интерфейса

2.1. База знаний естественно-языкового интерфейса

База знаний ИС естественно-языкового пользовательского интерфейса подразделяется на лингвистическую и предметную базы знаний. Предметная база знаний используется вопросно-ответной системой для поиска и генерации ответов по запросам пользователя. Кроме того, естественно-языковой интерфейс использует эту базу знаний для решения ряда задач, связанных с разрешением неоднозначности естественного языка, т.к. предметная база знаний однозначно задает контекст диалога пользователя с вопросно-ответной системой [Сулейманов, 2011].

Структура базы знаний естественно-языкового интерфейса представлена на рисунке 2.



Рисунок 2. Структура базы знаний естественно-языкового интерфейса

Лингвистическая база знаний содержит формальное описание используемого естественного языка, привязку лексики к предметной базе знаний, спецификации семантических языков.

Для описания знаний о каждом из уровней естественного языка (см. ниже описание структуры базы знаний) используется специализированный семантический язык представления лингвистических знаний. Спецификация данного языка также входит в состав лингвистической базы знаний. На практике лингвистическая база знаний может не включать все уровни описания естественного языка, т.к. использование ограниченного языка для общения пользователя с системой может дать более качественный результат, нежели использование неограниченного естественного языка [Byron Long, 1994].

2.1.1. Лингвистическая база знаний естественного языка

Лингвистическая база знаний естественного

языка имеет следующую логическую структуру, соответствующую уровням его формального описания:

- лексика и фразеология – включает описание лексического и фразеологического состава русского языка, а также некоторые закономерности словообразования;

- фонетика – включает описание звукового состава современного литературного русского языка и основные звуковые процессы, протекающие в языке;

- орфоэпия – включает описание норм современного русского литературного произношения;

- графика – описание русского алфавита, а также соответствий букв звукам;

- орфография – включает описание основных принципов русского написания – морфологические, фонетические, традиционные;

- морфемика и словообразование – содержит описание морфемного состава отдельных слов и основных способов образования новых слов, в том числе:

- морфологический;
- морфолого-синтаксический;
- лексико-семантический;
- лексико-синтаксический;
- грамматика – содержит описание правил морфологии и синтаксиса:
 - морфология – описание грамматических категорий и грамматических форм слов;
 - синтаксис – описание основных синтаксических единиц: словосочетания и предложения, виды синтаксических связей, типы предложений и их структуры;
 - пунктуация – включает описание совокупности правил расстановки знаков препинания;
 - текст, признаки, характеристики – описание совокупности правил построения текстов и их характеристики.

Как указывалось выше, для описания знаний каждого из перечисленных уровней используется специализированный формальный семантический язык представления лингвистических знаний, спецификация которого входит в состав лингвистической базы знаний.

2.1.2. Грамматический словарь русского языка

Грамматический словарь русского языка является важным компонентом базы знаний естественно-языкового интерфейса. Грамматические знания о естественном языке, записанные в лингвистической базе знаний, используются для грамматического анализа естественно-языкового текста вопроса, а также для синтеза ответов пользователю. 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 осуществляется так, как показано на рисунке 3.

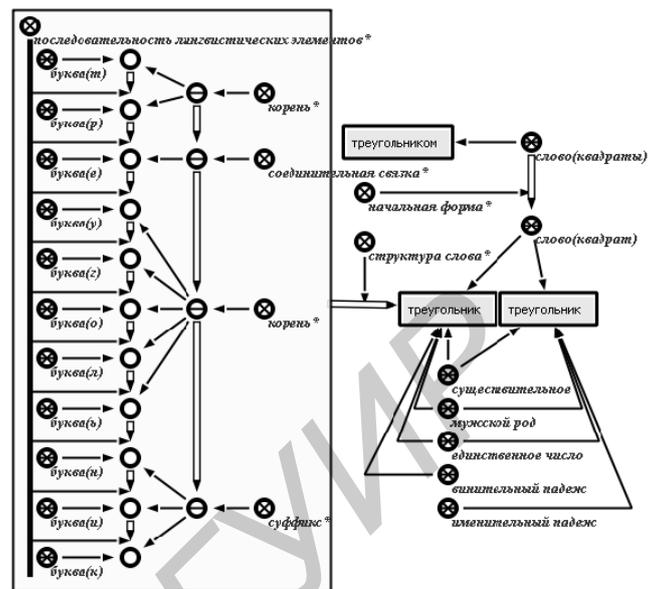


Рисунок 3. Запись морфологической информации в лингвистической базе знаний

При описании морфологической информации каждому слову ставятся в соответствие грамматические категории (число, род, склонение и т.д.). Для этого данное слово включается во множество конкретной грамматической категории (например, как показано на рис.3, единственное число, мужской род и т.д.). Каждое конкретное слово включено во множество его грамматических синонимов. Это позволяет быстро получать доступ ко всем наборам грамматических категорий при грамматическом анализе. Для каждого слова строится его разбиение на морфемы, а также эквивалентная запись в семантической сети.

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

2.2. Машина обработки знаний

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, например, трансляция фактографических знаний по предметной области в текст на естественном языке. Задачей анализаторов является анализ фрагментов знаний и выявление ранее неизвестных фактов, например, анализ запроса пользователя,

биссектрисы треугольников?» в семантически полном (развернутом) полужформальном виде будет выглядеть следующим образом: «Какие утверждения являются общими для понятия треугольник и понятия биссектриса?». Из примера видно, что в процессе развертки произошли следующие подстановки: высказывание «*обладать свойствами*» было развернуто в «*иметь общие утверждения*», «*свойства*» – в «*утверждения*», «*биссектрисы*» – в «*понятие биссектриса*», «*треугольники*» – в «*понятие треугольник*». Ответ на данный вопрос будет следующим: «У любого треугольника все биссектрисы пересекаются в одной точке»

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.

Полученный от предметной справочной системы ответ транслируется в текст на ограниченном естественном языке, который в результате выводится пользователю [Карпилович, 1997] [Попов, 1982].

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 [Мельчук, 1974]. Для этого необходимо добавить лингвистические знания о специфических для этой предметной области понятиях.

3.1. Схема взаимодействия компонентов ея-интерфейса

Если в системе используется только один естественный язык, то работу компонентов можно представить в виде последовательной обработки входящего сообщения. Если же в системе используются несколько языков или несколько компонентов, выполняющих одни и те же функции, процесс работы естественно-языкового интерфейса

начинает ветвиться.

В разрабатываемый прототип добавлена поддержка белорусского языка. По ряду технических причин существующий анализатор входящих сообщений не имел возможности обрабатывать тексты на белорусском языке. Поэтому в систему был добавлен простой, но универсальный компонент обработки входящих сообщений. Новый компонент может обрабатывать как белорусскоязычный, так и русскоязычный тексты на основе анализа состава и порядка слов в сообщении. В проведенных экспериментах было рассмотрено поведение системы в присутствии двух компонентов схожего функционала.

Цикл работы системы представлен на рисунке 4.



Рисунок 4. Схема работы компонентов ея-интерфейса с несколькими трансляторами

Из рисунка 4 видно, что пока один из трансляторов не создаст в системе рему и тему заданного пользователем вопроса, трансляция самого вопроса не будет произведена. Таким образом пользователь увидит ответ на свой вопрос только тогда, когда предметная интеллектуальная вопросно-ответная система не обработает входящий запрос, что произойдет только после того как один из анализаторов и трансляторов его не сгенерируют. Можно сказать, что в системе происходит свободная конкуренция функционально подобных компонентов. Те компоненты, которые «не успели» дать результат, при его появлении в памяти системы завершают свою текущую задачу.

3.2. Речевые компоненты

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 органом 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 и устройства вывода – речевым синтезом, повысив тем самым комфорт и скорость ввода и вывода, кажется очень привлекательной. Однако до последнего времени самыми распространёнными примерами использования звука в интерфейсе можно считать озвучивание некоторых операций с использованием слов либо звуков из стандартного набора. В настоящее время 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 – 1980-х гг.), и графического интерфейса (1980 – 2000-е гг.), будущее принадлежит комплексному пользовательскому интерфейсу, позволяющему задействовать, кроме зрения разные органы чувств человека, в первую очередь – его слух [Лобанов, 2006]. 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, подключённых к интернету. Это обеспечивает огромный рынок для будущих систем автоматического диалога.

В свете всех перечисленных тенденций вполне закономерно и актуально 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 неизбежно растёт число элементов графического меню, что в свою очередь замедляет время поиска нужной опции и заставляет отвлечься от искомого объекта [Landauer et al., 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.

3.3. Разработка компонента речевого ввода информации на основе пакета НТК

Как уже было сказано, небезосновательным выглядит факт, что после интерфейса командной строки и графического интерфейса будущее принадлежит комплексному пользовательскому интерфейсу, позволяющему задействовать, кроме зрения, разные органы чувств человека, в первую очередь – его слух.

В свете всех перечисленных тенденций, вполне закономерно и актуально выглядит разработка речевых способов управления 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 неизбежно растёт число элементов графического меню, что в свою очередь замедляет время поиска нужной опции и заставляет отвлечься от искомого объекта [Landauer et al., 1985].

Одной из важнейших задач данной работы является выбор подходящей технологии и методики для создания модуля распознавания слитной русской и белорусской речи, призванного обеспечить использование голосового ввода для реализации запросов к базам знаний. При этом акцент ставится на разработку именно относительно автономного компонента, который мог бы быть использован разработчиком, не имеющим высокой квалификации в сфере обработки сигналов или теории Скрытых Марковских Моделей (СММ). Такая подсистема должна стать частью целой технологии проектирования интеллектуальных

систем, разрабатываемой в рамках открытого проекта OSTIS.

Второй отличительной чертой является возможность распознавания белорусской речи, что является весьма актуальной проблемой из-за малого количества аналогов. Причин того, что так немного внимания уделяется включению белорусского языка в интерфейсы различного рода цифровых устройств разные. Однако мы не сомневаемся в необходимости развития белорусскоязычных способов взаимодействия и делаем конкретные шаги в этом направлении.

Теория СММ была выбрана в качестве методологической основы для создания модуля распознавания речевых запросов. В качестве набора инструментов, реализующего все основные функции и алгоритмы, был использован пакет НТК (<http://htk.eng.cam.ac.uk/> - The Hidden Markov Model Toolkit)

3.3.1. 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:

Русский:

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

Ответ: Это треугольник со сторонами а, b и углом С, 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.

Белорусский:

Вопрос: Як выглядае трохвугольнік?

Вопрос: Вывесці трансляцыю паняцця трохвугольнік.

Вопрос: Якое паняцце з'яўляецца надмноствам паняцця трохвугольнік?

И т. д.

Стоит отметить, что фразы выбирались по

следующему критерию: слова, входящие в запросы, должны состоять по возможности из большого количества звуков, чтобы повысить качество распознавания.

3.3.2. Выбор способа подготовки данных для создания моделей

В качестве вариантов, как для русскоязычного, так и для белорусскоязычного распознавателей рассматривались системы со следующими характеристиками: монофонные без разметки обучающих данных, монофонные с разметкой обучающих данных и системы на основе связанных трифонов, где обучение проводилось без разметки обучающих данных. Критерием для сравнения выступали среднее время распознавания на один фрейм, точность распознавания по фразам, а также точность распознавания по отдельным словам. Обучение осуществлялось на одинаковом ограниченном наборе из 50 фраз на русском языке и 73 фраз для белорусского аналога, содержащих в сумме 217 и 320 слов соответственно. Результаты представлены в сводной таблице 1 для обоих языков.

Таблица 1 - Оценка характеристик систем распознавания

Русский язык		
Среднее время распознавания на один фрейм, с/фрейм	Точность распознавания фраз, %	Точность распознавания слов, %
<i>Система на основе монофонов. Обучение без разметки данных</i>		
0.011396	74.00	94.97
<i>Система на основе связанных трифонов. Обучение без разметки данных</i>		
0.011802	72.00	93.53
<i>Система на основе монофонов. Обучение с разметкой данных</i>		
0.011626	98.00	99.57
Белорусский язык		
<i>Система на основе монофонов. Обучение без разметки данных</i>		
0.011567	77.00	95.60
<i>Система на основе связанных трифонов. Обучение без разметки данных</i>		
0.011912	74.00	94.40
<i>Система на основе монофонов. Обучение с разметкой данных</i>		
0.011803	98.00	99.65

На основе выше приведённой информации выбор был сделан в пользу системы на основе монофонов, для создания которой использовались данные, размеченные вручную. Решающим фактором стал высокий процент точности распознавания по фразам при приемлемом времени распознавания. Такие результаты можно объяснить следующим образом: исходя из условия ограниченности словаря и возможных последовательностей слов, целесообразным выглядит разметка ограниченного количества обучающих файлов вручную, с другой стороны. Это же ограничение препятствует созданию качественных трифонов, тем более связанных.

3.3.3. Алгоритм создания системы СММ для распознавания

Полный цикл создания хорошо обученных СММ включает два основных этапа: подготовка данных для обучения и непосредственно самообучение (Рисунок. 5).

Подготовка данных.

Шаг 1 -- Грамматика. Первым шагом явилось создание грамматики каждого из запросов. Грамматика в данном случае представляет собой сеть, которая включает строгие последовательности слов, допустимых для распознавания.

Шаг 2 -- Словарь. Спускаясь ниже по иерархии от фраз к отдельным словам, был создан упорядоченный словарь слов, входящих в запросы. Закрытый словарь предоставляет возможность создания уникальной фонетической транскрипции, адаптированной для конкретного диктора и учитывающей региональные особенности произношения некоторых слов русского и белорусского языков. Особенно это касается специфических звуков, характерных исключительно для белорусского языка: «дз», «дж», «ць» и т. д.

Шаг 3 -- запись данных. На данном этапе осуществляется создание набора файлов формата wav, содержащих несколько вариантов (не менее трёх) записей базовых слов, произнесённых диктором. В качестве инструмента использована функция HSLab из пакета НТК, которая позволяет не только записывать данные, но и размечать их по содержащимся фонемам. Всего для обучения использовалось 113 фраз, состоящих в сумме из 613 слов.

Шаг 4 -- кодирование данных. Финальным шагом в подготовке данных является обработка речевых сигналов и преобразование их в последовательности векторов признаков. В данной работе в качестве таких векторов были использованы кепстральные коэффициенты шкалы мел-частот [Oppenheim et al., 2004] [Опенгейм и др., 1979].

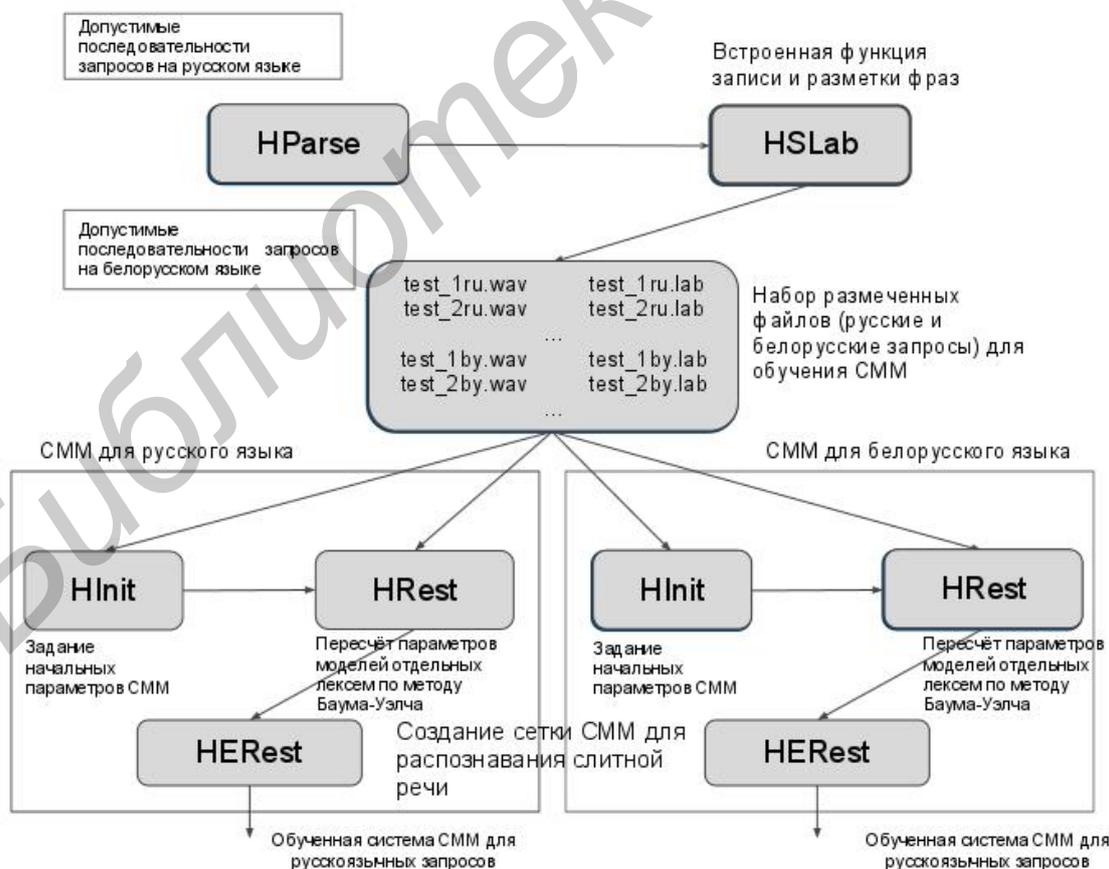


Рисунок 5. Алгоритм создания набора СММ для распознавания запросов на русском и на белорусском языках

Обучение

С этого момента начинается создание набора обученных СММ, эмиссионные вероятности в которых описываются смесью плотностей распределения Гауссовых случайных величин [Rabiner, 1988].

Шаг 5 -- Создание начальных монофонов. Первым шагом в создании системы СММ является определение модели прототипа. Для системы, основанной на фонемах, хорошей топологией является лево-правая схема с тремя состояниями [Бовбель и др., 2008]. Начальные значения для модели каждой фонемы рассчитывались на основании обучающих данных с помощью алгоритма Витерби [Бовбель и др., 1998]. Программным средством здесь выступила функция HInit.

Шаг 7 -- Пересчёт коэффициентов монофонов. Параметры модели каждой фонемы пересчитываются по методу Баума-Уэлча (алгоритм прямого-обратного хода или алгоритм максимизации правдоподобия) [Бовбель и др., 2008]. Пересчёт производится с помощью функции HRest.

Шаг 8 -- Создание системы СММ для распознавания фраз. Последним шагом является корректировка параметров моделей фонем, но уже не по отдельности, а в связке друг с другом в контексте как отдельных слов, так и целых предложений. Такая интегрированная система создаётся функцией HERest.

3.3.4. Ключевые особенности компонента речевого ввода

В заключение отметим, что результатом работы стало решение конкретной задачи, а именно создания модуля распознавания речи для осуществления голосовых запросов к интеллектуальной базе знаний. Цель была достигнута, что стало следствием решения следующих подзадач:

- выбран способ реализации системы с учетом требований к точности и времени распознавания;
- создан прототип модуля, который способен с высокой точностью распознавать широкий спектр запросов к базе знаний по геометрии. При этом в модуле реализована двуязычность, позволяющая делать устные запросы как на русском, так и на белорусском языках;
- выработана методика создания модуля распознавания как для русскоязычных, так и для белорусскоязычных фраз, которая может быть использована разработчиком, который не является специалистом в сфере речевых технологий, для создания модуля распознавания запросов к базам знаний в любых других сферах помимо геометрии.

Кроме того, стоит отметить, что благодаря большому проценту протяжных гласных в белорусском языке, потенциально системы распознавания, созданные для фраз белорусского

языка, обладают большей точностью распознавания при прочих равных условиях.

3.4. Разработка компонента речевого вывода информации на основе синтезатора русского и белорусского языков

Для перевода текстовой информации в речевую из вопросно-ответной системы используется два ир-компонента, основанных на синтезаторах речи по тексту на белорусской [Гецевич и др., 2010] и русской речи [Лобанов и др., 2008]. Данные ир-компоненты реализованы в качестве сторонних подключаемых модулей. В результате такого подключения к системе синтезаторов речи по тексту пользователь получает озвученный ответ от системы. Таким образом, система естественно-языкового интерфейса становится 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 для каждого языка. 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 5. 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 5. 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.

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

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, ненужных 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 происходить преобразование PDF в простой текст.

Главный модуль лингвистического процессора управляет другими его модулями и контролирует процесс преобразования в них текста в последовательность синтагм. Процессор слов определяет возможные лексико-грамматические характеристики слова (последовательности символов, отделенных пробелами и знаками препинания). Лексико-грамматический процессор определяет лексико-грамматические характеристики слова на основе вариантов, предложенных предыдущим процессором и лексико-грамматических характеристик других слов в тексте. Дополнительные плагины производят обработку специальных выражений (например, чисел, сокращений, дат, времени и др.) в орфографические слова. Процессор выражений расставляет ударения в словах, создаёт фонетические слова через присоединение к словам предлогов и частиц, заменяет конкретные выражения результатами обработки из плагинов. Процессор сборки словосочетаний соединяет отдельные слова в группы, исходя из лексико-грамматических характеристик этих слов, чтобы словосочетание не было правильно разбито на синтагмы. Процессор создания синтагм разделяет слова и словосочетания на возможные синтагмы, с указанием интонационного типа синтагмы в

зависимости от знаков препинания и лексико-грамматических характеристик слов.

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, определяемые 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 в формат MP3).

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

ЗАКЛЮЧЕНИЕ

Семантическая технология компонентного проектирования естественно-языковых интерфейсов вопросно-ответных систем предоставляет единую основу для анализа естественно-языковых текстов, начиная с ввода текста, морфологического, синтаксического, семантического анализа и генерации текста. Кроме того, указанная технология 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-компоненты системы распознавания и синтеза речи по тексту предоставляют конечному пользователю возможность устно задавать вопрос и слышать ответ на него от системы, а не просто вводить вопрос через клавиатуру и читать ответ с экрана компьютера. Это делает естественно-языковой интерфейс еще более естественным для пользователя.

Возможность интеграции сторонних разработок и проектов в качестве внешних ip-компонентов позволяет производить интеграцию различных подходов и методов в рамках одного проекта, что

позволяет эффективно использовать их лучшие стороны.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

[Ostis, 2012] Open Semantic Technology for Intelligent Systems. [Электронный ресурс]. – 2012. – Режим доступа: <http://www.ostis.net/>. – Дата доступа: 01.02.2012

[Сулейманов, 2011] Сулейманов Д.Ш. Двухуровневый лингвистический процессор ответных текстов на естественном языке / Сулейманов Д.Ш. // сборник трудов Международной научно-технической конференции OSTIS-2011, Минск, 2011 / БГУИР – Минск, 2011

[Byron Long, 1994] Byron Long, Natural Language as an Interface Style / Byron Long // Dynamic Graphics Project Department of Computer Science University of Toronto, 1994.

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

[Карпилович, 1997] Алгоритмы порождения предложений естественного языка (обзор и анализ) / Т.П. Карпилович. – Минск, 1977.

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

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

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

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

[Бовбель и др., 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, 1988] Rabiner L. A. Tutorial on Hidden Markov Models and Selected Applications in Speech / L. A. Rabiner // Recognition. IEEE Press, 1988. pp. 257—286.

[Oppenheim et al., 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 и др., 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 и др., 1979] Опенгейм, А. В. Цифровая обработка сигналов / А. В. Опенгейм, Р. В. Шафер. М., 1979.

[Гецевич и др., 2010] Гецевич, Ю.С. Система синтеза белорусской речи по тексту / Ю.С. Гецевич, Б.М. Лобанов. Речевые технологии. – 2010. – № 1. – С. 91-100.

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

SEMANTIC TECHNOLOGY DESIGN NL INTERFACES FOR QUESTION ANSWERING SYSTEMS

Y.S. Hetsevich*, S.A. Hetsevich*, O.E. Yeliseyeva**,
V.A. Zhitko***, A.A. Kuzmin**,

* *United Institute of Informatics Problems of the
National Academy of Sciences of Belarus, Minsk,
Republic of Belarus*

Yury.Hetsevich@gmail.com, Novacimya@gmail.com

** *Belarusian State University, Minsk, Republic
of Belarus*

olae@open.by, kuzaleksaleks@gmail.com

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

zhitko.vladimir@gmail.com

INTRODUCTION

Description of technology to design a natural language interfaces for intelligent systems which is based on semantic networks (sc-technology) [Ostis, 2012], applying to a prototype of natural language interface for question answering intelligent system on the geometry is presented. Model includes speech-to-text and text-to-speech Belarusian and Russian subsystems, the generation of responses in the form of natural language and formal text. Also article include the description methods for linking various components of the system and description methods for development of new components.

MAIN PART

Given technology is based on semantic technology designing user interfaces for intelligent systems, which is part of OSTIS technology. It means that natural language user interface is also an intelligent system with its' own knowledge base, knowledge production machine and user interface. But domain intelligent system and natural language interface aren't separated systems; they are integrated systems with shared memory and union knowledge bases.

Additional third-party component is natural Belarusian and Russian voice input. This component requires high performance, so it built in a traditional technology. But it doesn't mean that it not require semantic technology for improving functionality. Some data needed for recognizing human voice could be storage as knowledge in knowledge base or even generate from knowledge.

Another important third-party component is voice output. This component is top required for making natural language interface friendlier for users. Also, as previous component, it built in a traditional technology. But some tasks are hard to solve in traditional ways, for example correct detect the stress in words, sometimes for this need understand meaning of text content, and this could be solve by semantic approaches.

CONCLUSION

In the given paper short description of semantic technology to designing and prototyping natural language interfaces with voice input and output in Belarusian and Russian languages. Also described basics of designing linguistic knowledge bases for Russian language, it can be expand for various languages, which will be integrate in domain knowledge base in the future.