OSTIS-2013
(Open Semantic Technologies for Intelligent Systems)

УДК 621.391

СПРАВОЧНАЯ СИСТЕМА С РЕЧЕВЫМ ИНТЕРФЕЙСОМ

Житко В.А.*, Гецевич Ю.С.**, Лобанов Б.М.**

* *Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Республика Беларусь*
zhitko.vladimir@gmail.com

** *Объединённый институт проблем информатики НАН Беларуси, Минск*
Yury.Hetsevich@gmail.com
lobanov@newman.bas-net.by

Описывается модель справочной системы по правилам дорожного движения РБ с речевым пользовательским интерфейсом. Для распознавания речи используется интернет сервис Google по распознаванию речи. Разработан дополнительный компонент распознавания речи. Внесенные доработки позволили распознавать поток непрерывной речи. Анализ вопросов и генерация ответов осуществляется с помощью настраиваемого лингвистического процессора NooJ. Синтез речи по тексту осуществляется с помощью разработанной системы «МультиФон» на булоруском или русском языках.
Ключевые слова: речевой интерфейс; синтез речи; распознавание речи, справочные системы; анализ текстов.

Введение

В связи с бурным ростом рынка мобильных и встроенных приложений разработкам в сфере упрощения ввода запроса информации уделяется большое внимание. Наиболее перспективными в этой ситуации являются разработки, связанные с распознаванием и синтезом речи, – современной альтернативе ручному вводу с клавиатуры и графическому интерфейсу. Использование речевого ввода команд и вопросов позволяет запрашивать справочную информацию одновременно с выполнением основной деятельности (например, при управлении автомобилем), а использование при этом речевого синтеза позволяет также разгрузить зрительный канал восприятия информации.

Следует отметить, что для естественно-языкового интерфейса конкретной прикладной системы возможно использование ограниченной лексики и грамматики языка без серьёзного 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. Это позволяет снизить время обработки естественно-языковых конструкций в системе и частично избежать лингвистических неоднозначностей.

1. Справочная система по правилам дорожного движения

В настоящее время существует ряд справочных систем обладающих речевым интерфейсом (Siri, Voice Actions и др.), однако все они являются закрытыми коммерческими проектами и использовать их в других проектах не представляется возможным.

Целью данного проекта является разработка моделей и средств построения естественно-языкового интерфейса конкретной прикладной системы. Выбранный компонентный подход к построению пользовательского интерфейса позволит в дальнейшем свести разработку новых прикладных систем к настройке и адаптации готовых компонентов к конкретной предметной области.

В качестве предметной области для проекта были выбраны реальные правила дорожного движения (ПДД), действующие на территории РБ. Отметим, что набор ПДД характеризуются структурированностью исходных данных, что существенно облегчает решение поставленной задачи. В процессе взаимодействия с пользователем справочная система в качестве речевого ответа

производит выбор текста наиболее релевантного ПДД или задает уточняющие вопросы.

1.1. Обработка данных

Комплекс ПДД представлен в виде набора отдельных правил с дополнительным указанием ключевых слов для каждого правила. Цикл работы справочной системы начинается с ввода пользователем голосового (речевого) запроса на естественном языке. Используемый в системе компонент распознавания речи преобразует произнесённый запрос в набор наиболее вероятных текстовых сообщений. По введенному в систему запросу строится его формальное описание в памяти системы. Все предшествующие результаты анализа используются при анализе последующих запросов, что позволяет системе сохранять ход диалога с пользователем и задавать уточняющие вопросы.

Первым этапом анализа запроса пользователя является его морфологический анализ. На данном этапе для каждого слова входной фразы ставится в соответствие ее начальная форма. Для данного шага используется лингвистическая система NooJ [Silberztein, 2003].

NooJ является свободно распространяемым продуктом для формализации лингвистических данных. Система включает в себя морфологический и синтаксический анализатор, а также удобные средства для разметки корпуса вручную. В NooJ встроена система визуального написания грамматик, которая позволяет создавать различные системы анализа текста (например, для задачи фактографического поиска).

В настоящее время системой NooJ поддерживается ряд языков: арабский, армянский, болгарский, каталанский, китайский, хорватский, английский, французский, немецкий, иврит, венгерский, итальянский, польский, португальский, испанский язык, русский и белорусский языки. Ведутся работы по поддержке и других языков.

Лингвистический процессор NooJ включает несколько вычислительных средств, используемых для формализации лингвистических данных и разбора текстов.

Finite-State Transducer (FST) - граф, описывающий ряд последовательностей языковых единиц, в котором каждая такая последовательность связанна некоторым аналитическим результатом. Последовательности языковых единиц описаны во входной части FST; соответствующие результаты описаны в части продукции FST.

Как правило, синтаксические FST представляют собой последовательности слов, и производят добавление лингвистической информации (такой как синтаксическая структура). Морфологические FST представляют последовательности букв, которые описывают словоформу, и затем дополняют лексическую информацию (такую как часть речи, ряд морфологических, синтаксических и семантических меток).

В NooJ конечные автоматы (FSA), как правило, используются, чтобы определить местонахождение морфо-синтаксических шаблонов в корпусах и извлечь последовательности соответствий, чтобы построить индексы, соответствия.

Рекурсивные Сети Перехода (RTNs) - грамматики, которые содержат больше чем один граф: графы могут быть FST или FSA, а также включать ссылки на другие, включенные графы, эти графы могут в свою очередь содержать другие ссылки к тому же самому, или к другим графам. Вообще, RTNs используются в NooJ, чтобы построить библиотеки повторно используемых грамматик: разработанные простые грамматики могут быть снова использованы в более общих грамматиках, в свою очередь эти грамматики снова могут быть использованы и т.д.

Расширенные Рекурсивные Сети Перехода (ERTNs) - RTNs, которые содержат переменные, эти переменные как правило хранят части последовательностей, и затем используются в некоторых операциях (например, преобразуют их содержимое в множественное число, и т.д.), и затем уже используется при построении результатов.

NooJ включает флективный/деривационный модуль, который связан со словарями, так, чтобы он мог автоматически связывать словарные статьи с соответствующими формами, которые используются в грамматиках (эта функциональность позволяет избавляться от полных словарей формы INTEX, таких как DELAF и DELACFs).

Словари NooJ связывают каждую лексическую статью с флективной и/или деривационной парадигмой. Например, все глаголы, которые спрягаются как “aimer”, связаны с парадигмой “+FLX=AIMER”; все глаголы, которые принимают “-able” суффикс, связаны с парадигмой “+DRV=ABLE” и т.д.

Парадигмы, такие как “AIMER” или "ABBLE" описаны или графически в RTNs или в текстовых файлах.

Все это позволяет использовать NooJ как на первом этапе морфологического разбора, так и на последующем при поиске релевантного правила ПДД.

Перспективным так же выглядит использование грамматик NooJ для анализа пользовательских запросов с целью их классификации и выделения предмета вопроса. Классификация вопросов позволит учитывать особенности классов вопросов при поиске релевантных ответов.

Пример грамматики описывающий один из классов вопросов приведен на рисунке 1. Такой класс вопросов может быть использован при следующем диалоге:

- Что такое автомобиль? (выделяются следующие ключевые слова: автомобиль)

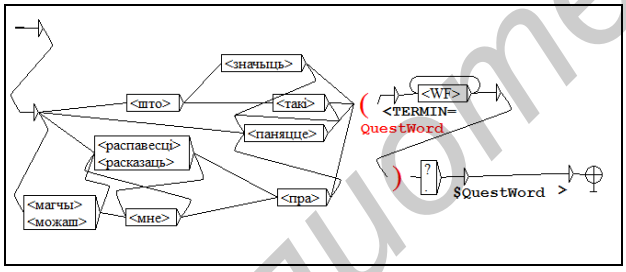
- Автомобиль — механическое транспортное средство, имеющее не менее четырех колес, расположенных не менее чем на двух осях, за исключением колесных тракторов и самоходных машин. К автомобилям приравниваются квадрициклы (мотоколяски) — четырехколесные механические транспортные средства, имеющие приводы управления автомобильного типа и массу в снаряженном состоянии не более 550 килограммов;

- А есть ли двухколесное транспортное средство? (выделяются следующие ключевые слова: транспортное средство)

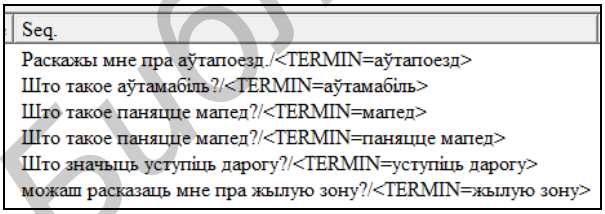
- Мотоцикл — двухколесное механическое транспортное средство с боковым прицепом или без него, приводимое в движение двигателем с рабочим объемом более 50 куб. сантиметров. К мотоциклам приравниваются трехколесные механические транспортные средства, имеющие массу в снаряженном состоянии не более 400 килограммов, а также механические транспортные средства, оборудованные двигателем с рабочим объемом до 50 куб. сантиметров, имеющие максимальную конструктивную скорость движения, определенную их технической характеристикой, более 50 км/ч;

- А что такое прицеп? (выделяются следующие ключевые слова: прицеп)

- Прицеп — транспортное средство, предназначенное для движения в составе с механическим транспортным средством;



а)



б)

Рисунок 1 – Пример анализа возможных вопросов к справочной системе: а) синтаксическая грамматика для поиска предмета вопроса; б) результат поиска предмета вопроса.

На последнем этапе – этапе поиска релевантной группы ПДД – рассчитываются коэфициены соотношения входной фразы и группы ПДД. Далее операция повторяется на множестве ПДД из выбранных групп. Если по результату поиска было найдено множество ПДД с равными коэфициентами, система строит множество ключевых слов позволяющих сделать окончательный выбор правила и просит пользователя уточнить вопрос или выбрать ключевое слово из списка. Ответ пользователя анализируется аналогичным образом, но уже с использованием прошлого результата.

Пример диалога пользователя и системы с уточняющим вопросом:

- Должен ли я уступать дорогу на перекрестке? (выделяются следующие ключевые слова: дорога, уступать, перекресток)

- Уточните вопрос или выберете одно из следующих ключевых слов: неравнозначные дороги, меняет направление, равнозначные дороги.

- Перекресток с равнозначными дорогами. (выделяются дополнительные ключевые слова: равнозначные, дорога)

- На перекрестке равнозначных дорог водитель транспортного средства, кроме трамвая, обязан уступить дорогу транспортным средствам, приближающимся справа. Этим же правилом должны руководствоваться между собой водители трамваев. На таких перекрестках водитель трамвая имеет преимущество перед другими транспортными средствами (кроме трамваев) независимо от направления его движения.

Пример диалога пользователя и системы с использованием словаря синонимов:

- Как мне развернуться? (выделяются следующие ключевые слова: развернуться (вводится синоним "разворот"))

- При повороте налево или развороте водитель транспортного средства обязан уступить дорогу встречным транспортным средствам, движущимся по равнозначной дороге прямо или направо, и попутному трамваю. Этим же правилом должны руководствоваться между собой водители трамваев. При одновременном праве на движение водители трамваев имеют преимущество перед другими транспортными средствами (кроме трамваев) независимо от направления их движения.

- Как мне развернуться на светофоре? (выделяются следующие ключевые слова: развернуться (вводится синоним "разворот"), светофор)

- При повороте налево или развороте по зеленому сигналу светофора водитель транспортного средства, за исключением трамвая, обязан уступить дорогу встречным транспортным средствам, движущимся прямо или направо, и попутному трамваю. Таким же правилом должны руководствоваться между собой водители трамваев.

1.2. Синтез речи

Для перевода текстовой информации в речевую используется два компонента, основанных на

синтезаторах речи по тексту на белоруском [Гецевич и др,. 2010] и руском [Лобанов и др, 2008] языках.

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 2. 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.

Главный модуль лингвистического процессора – текстовый процессор – 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).

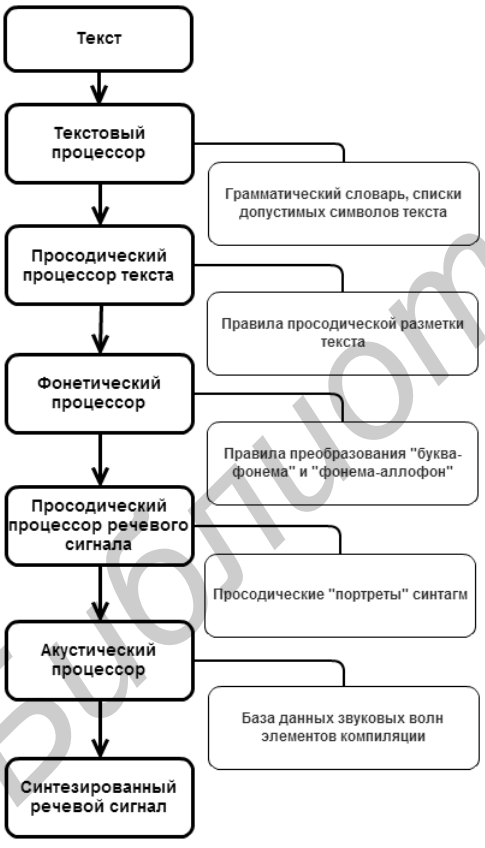


Рисунок 2 – Общая структура синтезатора

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 их лексико-грамматических характеристик. Процессор создания синтагм разделяет полученные 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.

1.3. Распознавание речи

Для распознавания речевого сигнала в системе используется компонент, в основе которого лежит сервис распознавания речи, разработанный компанией Google.

В настоящее время компания Google является лидером по предоставлению облачных технологий распознавания речи [Manjoo F., 2011]. В течение последних пяти лет активно развивалась облачная технология распознавания речи Google Voice, и к настоящему времени существуют технологии распознавания речи для большинства европейских языков, включая русский, японский и китайский языки. Одним из немаловажных компонентов системы распознавания речи Google Voice является обучающая выборка звукозаписей человеческого голоса. Для системы Google Voice источником таких записей являются различные сервисы, предоставляемые Google, использующие речевые технологии, к ним относятся система распознавания

речи и команд в системе Android, сервис диктовки писем Google Mail, телефонная справочная система Goog411 и др. [Singhal, A., 2011]. Таким образом обучающая выборка постоянно пополняется новыми образцами голосов с их особенностями как произношения, так и эффектами, вносимыми техническими особенностями записи и передачи голоса на различных устройствах. К примеру, в 2011 г. обучающая выборка для английского языка составляла примерно 230 млрд записей слов извлеченных из реальных запросов [Enge, E. 2011]. Для обработки таких объемов информации требуется около 70 лет процессорного времени, однако с использованием облачной технологии Google время сокращается до одного дня [Singhal, A., 2011]..

Для того, чтобы использовать систему распознавания речи Google Voice в рассматриваемом прикладном проекте, разработан дополнительный компонент распознавания речи. В разработанном компоненте распознавания речи выделены следующие программные блоки:

- детектор речи;
- блок сжатия и обработки речевого сигнала;
- блок сопряжения с системой Google Voice;

Детектор речи является необходимым дополнительным компонентом, обусловленным используемой архитектурой удаленного распознавания речи. В этом случае канал передачи данных (в нашем случае интернет-соединение) является «бутылочным горлышком», ограничивающим максимально возможный объем передачи данных в заданный промежуток времени. При передаче по каналу слишком длинного отрезка речевого сигнала происходит недопустимо длительная задержка ответной реакции системы распознавания. Кроме того, при этом возникает значительный риск получения ошибочных результатов распознавания. Ввод коротких отрезков речи возможен при использовании ручного стартстопного режима, предоставляемого системой Google Voice. Однако для решения поставленной здесь задачи стенографирования устной речи этот режим оказывается крайне неудобным. Обычно диктовка осуществляется короткими фразами, заканчивающимися паузами. После произнесения каждой из них пользователь, как правило, желает убедиться в правильности распознавания и при необходимости повторить ее более четко.

Задача детектора речи заключается в автоматической локализации полезного сигнала, т. е. в определении начала и конца произнесенной фразы. Это обеспечивает автоматический пофразовый ввод речи. В экспериментальной программной системе использован относительно простой, но достаточно эффективный алгоритм определения полезного потока речи, основанный на расчете усредненной текущей энергии звуковой волны входного сигнала. Перед началом работы в течение одной секунды производится замер фонового шума для определения порога срабатывания детектора речи. Порог вычисляется как произведение средней энергии фонового шума на коэффициент, заданный в настройках программы (по умолчанию равный 2,0).

Далее ведется запись в кольцевой буфер задержки сигнала (по умолчанию задержка равна 1,0 с) и одновременно вычисляется текущая средняя энергия звуковой волны входного сигнала (по умолчанию время усреднения равно 0,4 с). Если порог будет превышен, задержанный входной сигнал передается на вход компонента сжатия и обработки сигнала, а передача сигнала будет вестись до тех пор, пока значение средней энергии не станет меньше значения порога.

Полученный полезный сигнал в модуле сжатия и обработки сигнала подготавливается для отправки на удаленный сервер распознавания речи Google Voice. Для этого сигнал кодируется в открытом формате FLAC (Free Lossless Audio Codec) с заданными характеристиками (частота дискретизации 16кГц, моно). При этом происходит сжатие сигнала кодеком, что уменьшает объем передаваемых данных и, как следствие, сокращает время ожидания результата распознавания. Использование open-source-кодека имеет и другие преимущества: простота использования сторонними разработчиками, единообразие получаемого сервером формата данных. Каждый компонент программы работает асинхронно и имеет свои пулы данных для нивелирования эффектов, связанных с задержками в работе каждого компонента. Такие задержки появляются как на этапе записи и кодирования речевого сигнала в файл, так и отправки его на удаленный сервер.

На рисунке 3 описанный алгоритм более подробно представлен в графическом виде.

Рисунок 3 – Схема работы компонента распознавания речи

Заключение

Представленная модель справочной системы с разбиением на отдельные компоненты со строго заданным функционалом, позволяет упростить разработку, а в дальнейшем – сопровождение, различных систем с речевым интерфейсом.

Разбиение системы на отдельные независимые компоненты дает возможность интеграции сторонних разработок и проектов и производить интеграцию различных подходов и методов в рамках одного проекта, что позволяет эффективно использовать их лучшие стороны.

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 языковой интерфейс еще более естественным для пользователя.

Библиографический список

[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.

[Manjoo F., 2011] Manjoo, F. Now You're Talking / Farhad Manjoo // The Slate Group. – Washington Post Compa-ny, 2012. – [Electronic resource] – Mode of access : http://www.slate.com/articles/technology/technology/2011/04/now_youre_talking.single.html. – Date of access : 01.08.2012.

[Singhal, A., 2011] Singhal, A. Knocking down barriers to knowledge / Amit Singhal // Google Official Blog [Elec-tronic resource]. – 2011. – Mode of access : http://googleblog.blogspot.com/2011/06/knocking-down-barriers-to-knowledge.html. – Date of access : 01.08.2012.

[Enge, E. 2011] Enge, E. Search Algorithms with Google Director of Research Peter Norvig / E. Enge // Stone Temple Consulting [Electronic resource]. – 2011. – Mode of access : http://www.stonetemple.com/search-algorithms-with-google-director-of-research-peter-norvig. – Date of access : 01.08.2012.

[Silberztein, 2003] Silberztein, M. NooJ Manual [Electronic resource]. – 2003. – Mode of access : http://www.nooj4nlp.net/NooJManual.pdf. – Date of access : 01.07.2012.

HELP SYSTEM WITH SPEECH USER INTERFACE

V.A. Zhitko*, Y.S. Hetsevich**, B.M. Lobanov**

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

zhitko.vladimir@gmail.com

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

Yury.Hetsevich@gmail.com

lobanov@newman.bas-net.by

This work describes a model of help system with speech user interface for highway regulations. Also describes components for text-to-speech, question-answering and voice recognition. Primary goal of this work is making easy building speech user interfaces.

MAIN PART

For analyzing user questions and generating answers used linguistic processor NooJ.

Another important 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.

Additional third-party component is natural Russian voice input. For this component used Google voice recognition service with made some changes. So component can recognize non-stop user speech.

CONCLUSION

In the given paper short description of model and methods for designing and prototyping natural language interfaces with voice input, question-answering processing and output Belarusian or Russian languages.