Комплексное решение на основе речевых технологий для президентской библиотеки имени Б. Н. Ельцина

Integrated Solution Based on Speech Technologies for Boris Eltsyn Presidential Library

Комплексне рішення на основі мовних технологій для президентської бібліотеки імені Б. М. Єльцина

А. Ю. Поляков
ООО «Центр речевых технологий», Санкт-Петербург, Россия

Д. В. Дырмовский, А. В. Рыбаков
филиал ООО «Центр речевых технологий», Москва, Россия

Alexey Polyakov
«Speech Technologies Center» Company, St. Petersburg, Russia

Dmitry Dyrmovsky, Alexey Rybakov
Division of «Speech Technologies Center» Company, Moscow, Russia

О. Ю. Поляков
ТОВ «Центр мовних технологій», Санкт-Петербург, Росія

Д. В. Дірмовський, О. В. Рибаков
філія ТОВ «Центр мовних технологій», Москва, Росія

ООО «Центр речевых технологий» (ЦРТ) – российский лидер в области разработки электронной техники и программного обеспечения для высококачественной записи, обработки и анализа звуковой информации.
В докладе изложены основные моменты комплексного решения, предложенного для Президентской библиотеки им. Б. Н. Ельцина, в виде аппаратно-программного комплекса, реализующего возможности самых передовых технологий в области высококачественной записи, обработки, анализа, распознавания и синтеза речи.

«Speech Technologies Center» is the leading Russia's company in development of electronic facilities and software for high-quality recording, processing and analysis of audio information. The key points of the integrated solution proposed for Boris Eltsyn Presidential Library are described in this paper. This soft-hardware solution realizes the advantages of the most advanced technologies in the field of high-quality recording, processing, analysis, recognition and synthesis of speech.

ТОВ «Центр мовних технологій» (ЦМТ) – російський лідер у галузі розробки електронної техніки та програмного забезпечення для високоякісного запису, обробки та аналізу звукової інформації. У доповіді викладено основні моменти комплексного рішення, запропонованого для Президентської бібліотеки ім. Б. М. Єльцина, у вигляді апаратно-програмного комплексу, що реалізує можливості найбільш передових технологій у галузі високоякісного запису, обробки, аналізу, розпізнавання та синтезу мови.

С момента создания в 1990 году ЦРТ производит и поставляет продукцию для потребителей, в деятельности которых особое значение придается качественной передаче, регистрации и обработке речевой информации. Научные исследования являются одним из приоритетных направлений деятельности компании. В настоящее время 25% сотрудников компании – это высококвалифицированные ученые, работающие по четырем основным направлениям: автоматическое распознавание речи, синтез речи, верификация/идентификация личности диктора и шумоочистка. Среди заказчиков ЦРТ: Администрация Президента РФ, Аппараты Правительства РФ, Совета Федерации и Государственной Думы Федерального Собрания РФ, , органы исполнительной и законодательной власти субъектов Российской Федерации.

По прогнозам экспертов в ближайшие 5 лет темпы роста рынка речевых технологий будут более чем в два раза выше темпов роста рынка информационных технологий в целом. На сегодняшний день проблемы связанные с автоматизацией перевода огромного количества хранящейся в

библиотеках аналоговой информации в цифровую форму, поиск необходимых медиаматериалов, повышение качества аудиофайлов является очень актуальными. Отдельно необходимо отметить проблему доступности библиотек для людей с ограниченными возможностями по зрению. Развитие информационно-коммуникационных технологий открывает для современной библиотеки новые возможности по накоплению, каталогизации, хранению и поиску библиотечной информации. Особую роль среди них играют технологии, в основе которых лежит запись, обработка и анализ речевой информации. Среди них:

1. Технология синтеза естественной русской речи по тексту. С помощью данной технологии люди с ограниченными возможностями (например, инвалиды по зрению) могут получить доступ и прослушивать текстовые материалы, содержащиеся в библиотеках.

2. Технология поиска ключевых слов (введенных с клавиатуры компьютера) и музыкальных фрагментов в потоке аудиоинформации. Поиск нужного звукового материала среди сотен фонограмм в попытке отыскать нужную фразу или слово без применения данной технологии слишком утомителен. Применение данной технологии существенно сокращает по времени и упрощает поиск – ввел слово (звуковой фрагмент) и получил результат.

3. Технология шумоочистки речевых сигналов. Архивные аудиозаписи могут быть плохого качества и установить по ним содержание разговора практически невозможно. Для звукозаписей, так же как и для книг нужно проводить реставрацию. Специально для этого разработан единственный в мире аппаратно-программный комплекс (АПК), обеспечивающий комплексное исследование аналоговых и цифровых фонограмм речи, а также проверку подлинности фонограмм.

4. Технология выделения и сравнения биометрических признаков речи. На основе данной технологии создана система экспресс-исследований фонограмм речи и система доступа к информационным ресурсам. Данная технология позволяет автоматизировать процессы ведения пользовательской картотеки и учета пользователей с помощью их голоса, а также проводить поиск искомого диктора в архиве мультимедийной информации.

Опыт разработки и внедрения таких технологий позволил ЦРТ предложить Президентской библиотеке имени Б. Н. Ельцина комплексное решение в виде аппаратно-программного комплекса. Данный проект будет реализован в 2009 году совместно с компанией ОАО «Сатурн», г. Москва.

, Комплекс решает следующие задачи:

1. Перевод в цифровую форму и централизованное хранение мультимедиа материалов.
2. Каталогизация и ведение надежного, долговременного архива оцифрованных записей с возможностью архивирования и экспорта необходимой информации на внешние носители в стандартных форматах операционной системы Windows.
3. Оперативная подготовка текстовых расшифровок любых аудиозаписей, экспортированных в систему.
4. Расширенный интеллектуальный поиск в архиве аудиозаписей по следующим речевым признакам:
    - наличие в фонограмме заданного набора ключевых слов (набор ключевых слов вводится с клавиатуры);
    - наличие в фонограммах фраз на русском языке;
    - присутствие в записи голоса искомого диктора из заданного набора (вне зависимости от языка произношения);
    - наличие в фонограмме заданной музыкальной композиции или её фрагмента.
1. Преобразование текстовых записей в речь с помощью технологии автоматического синтеза речи и запись полученного звукового файла в стандартный формат операционной системы Windows.
2. Разделение хранящихся в архиве фонограмм на фонограммы, содержащие речевую и не речевую информацию.
3. Расширенное управление правами пользователей.
4. Интерактивный WEB доступ к поиску, отображению информации и управлению комплексом.

АПК функционирует следующим образом. Накопленные архивные аналоговые записи воспроизводятся с помощью имеющейся у библиотеки аппаратуры воспроизведения и через коммутационные блоки поступают на блок оцифровки информации, выполняющий преобразование поступившей информации в цифровую форму универсального формата. Оператор блока оцифровки

может выполнять визуальный и акустический контроль поступающей информации, контролировать и изменять параметры преобразования, а также присваивать поступающим данным текстовые комментарии для повышения эффективности дальнейшего поиска. Оцифрованные данные через блок обработки информации, предназначенный для улучшения качества записанного сигнала путем устранения шумов в полезном сигнале, сохраняются в центральном блоке системы, где происходит их индексация, каталогизация и регистрация в реляционной базе данных.

Пользовательский интерфейс АПК предоставляет оператору интерактивный доступ к архиву центрального блока системы в виде WEB страниц, доступных для просмотра стандартным программным обеспечением Internet Explorer.

По запросу Оператора на поиск информации центральный блок с помощью блока обработки и поиска информации осуществляет анализ архивных аудиозаписей, выделяет существенные речевые признаки, сравнивает их с параметрами, заданными в поиске, и возвращает результат Оператору системы.

Найденные по результатам поиска данные могут быть отображены на экране, отправлены на печать, воспроизведены или экспортированы на внешние носители.

В случае, если необходимо составить текстовую расшифровку найденных фонограмм, Оператор системы может отправить подлежащую расшифровке фонограмму в блок обработки аудио(видео) информации, результат работы которого также сохраняется на центральном блоке системы.

Безопасность АПК обеспечивает расширенная система управления правами пользователей. Администратор АПК определяет пользователей или группу пользователей, которые могут регистрироваться в системе, и задает набор ресурсов центрального блока, к которым могут получить доступ зарегистрированные пользователи.

АПК включает в себя следующие блоки:

Блок воспроизведения аудио (видео) информации с различных носителей

Блок воспроизведения аудио (видео) информации предназначен для воспроизведения аудио (видео) информации с различных носителей сигнала, имеющихся у библиотеки.

Блок состоит из модуля воспроизведения с микрофона или конферецсистемы и модуля воспроизведения материалов, хранящихся в архиве.

Блок оцифровки аудио (видео) информации

Данный блок предназначен для высококачественной оцифровки аудио (видео) информации, поступающей из «Блока воспроизведения аудио (видео) информации с различных носителей».

В состав блока входят:

1. Станции оцифровки и записи.
2. Автоматизированное рабочее место (АРМ) «Менеджер записи».

Станции оцифровки и записи осуществляют оцифровку аналогового аудио сигнала, временное накопление и передачу в Центральный блок для последующего хранения.

АРМ «Менеджер записи» позволяет управлять процессом оцифровки на 2 станциях записи одновременно, контролировать качество поступающего на вход сигнала (с помощью встроенных в программное обеспечение средств визуального и акустического контроля) и изменять параметры преобразования.

Блок подготовки аудио (видео) информации

Данный блок предназначен для первичной обработки оцифрованных аудио (видео) данных, поступающих из «Блока оцифровки аудиоинформации».

Блок состоит из следующих модулей:

1. Модуль шумоочистки мультимедиа информации.
2. Модуль выделения речевых фрагментов в мультимедиа информации.
3. Модуль деления речевого сигнала, содержащегося в мультимедиа информации, на разных дикторов.
4. Модуль диагностики подлинности фонограмм.

Модули реализованы в виде «АРМ подготовки аудио-видео информации», включающего в себя набор программно-аппаратных средств для проведения шумоочистки, восстановления и цифровой обработки фонограмм.

В состав программно обеспечения блока подготовки аудио (видео) информации входит:

1. Программное обеспечение (ПО) выделения в мультимедиа информации языковой принадлежности.
2. ПО распознавания речи.
3. ПО автоматизированного поиска медиафрагментов.
4. ПО анализа и визуализации звуковых сигналов.
5. ПО шумоочистки звуковых сигналов в реальном времени.
6. ПО сегментации звуковых сигналов.
7. ПО диагностики подлинности фонограмм

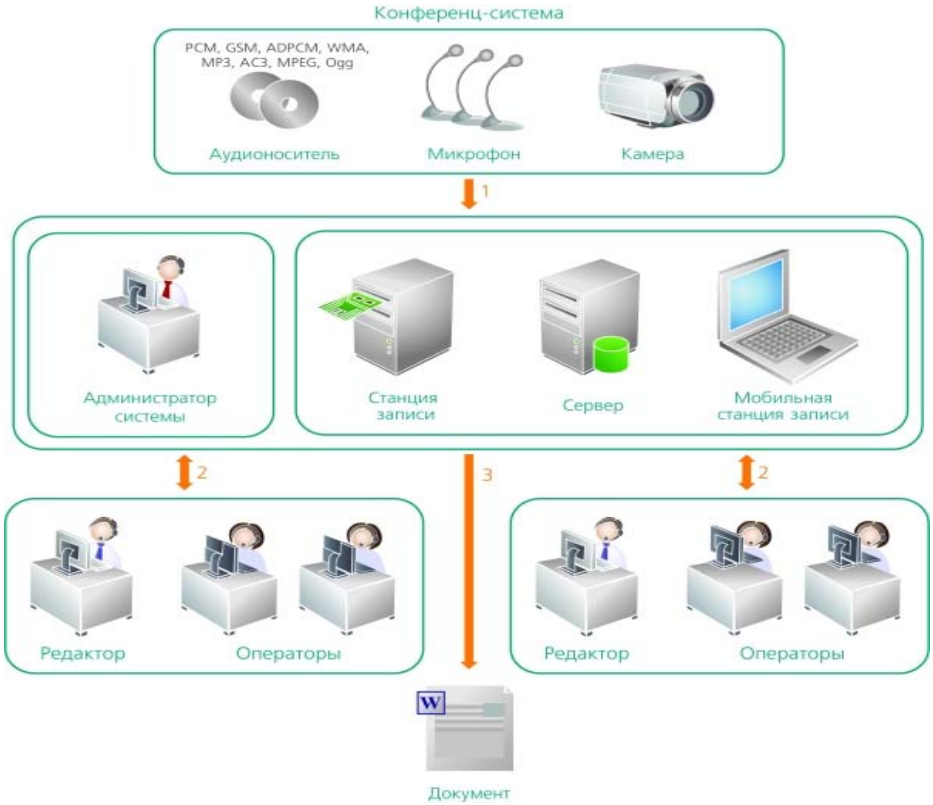
Блок обработки аудио (видео) информации

Данный блок предназначен для обработки подготовленной аудио (видео) информации и перевода ее в текстовый вид.

Модуль реализован в виде комплекса «Нестор-видео», в составе:

1. 2 АРМ менеджера.
2. 14 АРМ «Оператор».
3. 1 АРМ «Архиватор».

В принцип работы данного блока положен метод независимой квазипараллельной обработки фрагментов аудио (видео) данных различными операторами. Общая схема работы комплекса «Нестор-видео» показана на рисунке.



Общая схема работы комплекса «Нестор-видео»

Исходные данные разбиваются на фрагменты заданной длительности и передаются на обработку Операторам, каждый из которых осуществляет подготовку части итогового документа. Подготовленные фрагменты автоматически «склеиваются» Менеджером в итоговый документ, который может быть экспортирован на внешние носители с помощью АРМ «Архиватор». В случае

необходимости (например, подготовки документа к официальной публикации) Менеджер может передать первую версию («сырую» стенограмму) итогового документа на дальнейшую обработку: редактирование и корректуру. Результат работы Блока обработки сохраняется в Центральном блоке системы.

Функциональные возможности Блока обработки аудио (видео) информации:

1. Рассылка фрагментов звуковых программ мероприятий (фонограмм) на рабочие места менеджеров и операторов-стенографистов.
2. Выполнение независимой многостадийной обработки текстов стенограмм, предусматривающей следующие стадии: стенографирование (расшифровка), редактирование, корректура.
3. Повышение скорости подготовки стенограмм мероприятий посредством применения метода независимой квазипараллельной обработки несколькими операторами последовательных фрагментов текста исходной фонограммы мероприятия и их автоматической «склейки» в итоговый текстовый файл.
4. Ведение архива синхронизированных звуковых и текстовых файлов.
5. Идентификация диктора по голосу.
6. Оперативный поиск подготовленных текстов стенограмм мероприятий и соответствующих им фонограмм.
7. Возможность просмотра подготовленного текста стенограммы с одновременным озвучиванием указанных фрагментов.
8. Вывод подготовленных текстов стенограмм на печать.

Блок обработки и поиска (сортировки) архивных материалов

Данный блок предназначен для работы с архивом оцифрованного аудио (видео) материала и обеспечивает возможности поиска нужного материала по определенным параметрам, в частотности, по диктору, по ключевым словам, по интересующему музыкальному аудио фрагменту, по языку говорящего диктора.

Блок состоит из следующих модулей:

1. Модуль автоматической сортировки мультимедиа информации по языковой принадлежности содержащейся в ней речи и разбиения медиафайлов, содержащих речь на нескольких языках, на одноязыковые фрагменты.
2. Модуль автоматической идентификации диктора в мультимедиа информации вне зависимости от национальной принадлежности.

Модуль реализован на базе система автоматизации фоноучетов и экспресс-исследований фонограмм речи «Трал-М», которая основана на уникальности геометрии речевого тракта каждого человека. Используются спектрально-формантный метод и метод статистики основного тона.

1. Модуль автоматического анализа мультимедиа информации для поиска в ней интересующих («ключевых») слов (словосочетаний) с элементами технологии распознавания речи.

Модуль разработан на основе технологии Voice Digger, запатентованной специалистами ЦРТ.

2. Модуль автоматического преобразования мультимедиа информации в требуемые форматы.
3. Модуль преобразования текста в речь на основе технологии синтеза русской речи.

Модуль разработан на основе технологии «Живой голос». Технология позволяет озвучивать любые, в том числе нестандартные тексты (SMS, электронные письма, Интернет-форумы и т. п.) таким образом, что у слушателя складывается ощущение, что он слышит естественный человеческий голос. Текст может быть прочтен различными голосами синтеза: мужским или женским и может быть настроен под предпочтение слушателя (голос, громкость, темп, и т. д.). Каждый голос основан на использовании речевой базы диктора объемом около 10 часов речи, размеченной на 9 уровнях, включающих текстовую расшифровку, разметку на слова, слоги, аллофоны, паузы, маркеры словных и фразовых ударений, типы интонации, неречевые явления и другие фонетические явления.

Для правильного интонирования и определения места ударения в словах разработан мощный модуль автоматической обработки русского текста, использующий морфологический, синтаксический и семантический виды анализа. Использование данного модуля, также как и столь объемные и

тщательно размеченные голосовые базы, делают «VitalVoice» уникальной технологией синтеза русской речи.

Для того чтобы синтезированная речь звучала натурально, решен целый комплекс задач, связанных как с обеспечением естественности голоса на уровне плавности звучания и интонации, так и с правильной расстановкой ударений, расшифровкой сокращений, чисел, аббревиатур и специальных знаков с учетом особенностей грамматики русского языка.

1. Модуль анализа мультимедиа информации для поиска в ней интересующих медиафрагментов.

Модуль разработан на основе технологии Jingle Tracker поиска звуковых фрагментов в звуковом потоке или файлах, созданной и запатентованной специалистами ЦРТ.

В состав блока входя два АРМ обработки и поиска (сортировки) архивных материалов и система экспресс-исследований фонограмм речи (поиск фонограммы по имеющемуся образцу речи диктора).

Модуль ограничения прав доступа к информационным ресурсам

Модуль реализует технологию идентификации по голосу Voice Key, которая основана на уникальности геометрии речевого тракта каждого человека. В Voice Key используется спектрально-формантный метод, базирующийся на различных спектральных характеристиках речи разных людей. Достоинствами технологии являются:

- отсутствие возможности делегировать полномочия другому лицу путем передачи пароля, ключа, RFID карты;
- отсутствие возможность потери или кражи ключа: Ваш голос — Ваш ключ;
- увеличение количества пользователей в системе не влечет за собой дополнительных затрат;
- высокоинтеллектуальная система, адаптирующаяся к изменяющемуся голосу человека, автоматическое обновление речевых эталонов («ключей»);
- высокая надежность и стабильность работы системы.

Центральный блок системы

Данный блок является ключевым звеном системы и обеспечивает хранение мультимедиа информации, результатов ее обработки различными компонентами и поиска требуемой информации. Центральный блок представляет собой два сервера БД (основной и резервный).

Информация, поступающая из «Блока оцифровки аудиоинформации», сохраняется в первоначальном виде. Все виды обработки производятся в других блоках с копиями оцифрованных данных. Центральный блок системы является центральным звеном, координирующим совместную и устойчивую работу всех блоков.

Компания Центр речевых технологий располагает всеми необходимыми технологиями и программным обеспечением для создания и внедрения всех элементов АПК, которые ведутся с начала 2008 года. Окончание работ планируется в конце 2009 года. В итоге автоматизация процессов оцифровки, хранения, обработки, сортировки и интеллектуального поиска мультимедиа материалов будет осуществляться не только с помощью компьютерных технологий на базе распространенной в настоящее время операционной системы Microsoft Windows, но и с применением самых передовых отечественных технологий в области высококачественной записи, обработки, распознавания и синтеза речи.