Речевые технологии на службе электронного государства

Ядигар Имамвердиев[1], Людмила Сухостат[2]
[1,2] Институт Информационных Технологий НАНА, Баку, Азербайджан
[1]*yadigar@lan.ab.az*, [2]*lsuhostat@hotmail.com*

***Аннотация* – В работе приводится обзор применения речевых технологий в интересах электронного государства. Также рассматривается интеллектуальный анализ аудиоданных как подход для работы с большими объемами речевой информации.**

Ключевые слова – речевые технологии, электронное государство, интеллектуальный анализ аудиоданных, распознавание слитной речи с большим словарем

I. Введение

В настоящее время информационные и коммуникационные технологии (ИКТ) широко применяются в государственных органах для повышения эффективности их работы, оптимизации предоставления услуг населению и бизнесу, более широкого вовлечения граждан в процессы принятия решений по разным задачам государственной жизни [1, 2].

Порталы электронного правительства используются гражданами многих стран для получения доступа к набору услуг. Многие страны реализуют программы развития электронного правительства. Основными целями этих программ являются ускорение и упрощение предоставления услуг гражданам, снижение бюрократической нагрузки благодаря современным технологическим решениям в области обработки и передачи данных.

Современные речевые технологии, как один из видов ИКТ, имеют уникальные возможности для совершенствования и повышения эффективности предоставления государственных услуг гражданам и бизнес-структурам. Ведь речь является самым простым и удобным способом обмена информацией, наиболее естественным способом взаимодействия между людьми.

Пользователи электронных услуг имеют дело с большим количеством текстовой информации для чтения и заполнения полей, что создает определенные трудности и неудобства, в частности для людей с ослабленным зрением.

Другим неудобством является процедура аутентификации пользователя. С одной стороны, сам ввод пароля не является какой-то необычной для интернет-пользователя практикой. С другой – запомнить номер достаточно трудно, а сохранять такие данные в браузере недостаточно безопасно.

Решением приведенных выше проблем является внедрение речевых технологий: синтез речи, распознавание речи и голосовая биометрическая аутентификация.

В данной статье рассматривается применение речевых технологий для порталов э-правительства. Рассматривается интеллектуальный анализ аудиоданных как подход для работы с большими объемами речевой информации, осуществляющий быстрый поиск аудио- и видеоконтента. А также приводятся возможные пути развития речевых технологий в интересах э-государства.

II. Обзор современных применений речевых технологий

В 1960-70-х годах возросло внимание исследователей к разработке систем распознавания речи.

Начиная с 80-х годов технический прогресс сделал программное обеспечение и устройства распознавания речи более функциональными и удобными для пользователя, при этом большинство современных продуктов работают с точностью до более 90%.

Распознавание речи используется в широком спектре приложений, удовлетворяя потребностям э-государства, упрощая взаимодействие с пользователями, повышая эффективность и снижая эксплуатационные затраты. Действительно, последние достижения в области распознавания речи создают динамичную среду, так как эта технология привлекает внимание тех, кто нуждается или хочет проводить вычислительные задачи, оставляя руки свободными. Так, объединение крупных словарей и распознавание слитной речи позволяют двигаться в направлении распознавания речи и занять свое место в качестве лидера в секторе высоких технологий.

Рассмотрим некоторые сферы применения речевых технологий в рамках э-государства.

A. Речевые технологии для эффективного взаимодействия с веб-порталом э-государства

Веб-технологии являются основным инструментом электронного взаимодействия государства и граждан. Электронное государство использует современные технологии Web 2.0 для предоставления электронных услуг, повышения качества деятельности государственных органов и привлечения всех граждан к разработке и обслуживанию государственной политики. Уровень развития э-государства определяется исходя из используемых веб-технологий и спектра предоставляемых э-услуг.

Одним из основных требований к веб-сайтам государственных органов является обеспечение общедоступности (Accessibility). Общедоступность направлена на снятие барьеров и обеспечение доступа к

информации каждому человеку, пользующемуся информационными технологиями. Такими барьерами могут быть, например, ухудшение и нарушение зрения.

Консорциум W3C (World Wide Web Consortium) в рамках инициативы Web Accessibility Initiative (WAI) [3] устанавливает стандарты разработки веб-ресурсов для обеспечения доступности людям с ограниченными возможностями. Одним из таких стандартов является стандарт WAI-ARIA (Accessibility Rich Internet Applications), который определяет подходы к разработке активных интернет-приложений.

Внедрение речевых технологий в веб-порталах э-государства позволит осуществлять [4]:

- голосовую навигацию;
- голосовой поиск;
- голосовой ввод данных;
- голосовую биометрическую аутентификацию пользователя.

B. Речевые технологии для обучения языку

Системы распознавания речи должны быть разработаны специально для компьютеризированного обучения произношению (Computer-assisted pronunciation training, CAPT) в целях поддержания правильной обратной связи (отклика) и в то же время получения удовлетворительных характеристик [5]. Например, системы распознавания слитной речи с большим словарем (large vocabulary continuous speech recognition, LVCSR) широко доступны для коммерческого использования, они не обязательно подходят для обучения произношению неносителя языка. Вообще системы LVCSR приспособлены для работы с широким спектром акцентов и нестандартных произношений. Они не предназначены для использования в качестве инструмента для различения фонетически похожих произношений конкретных слов. Кроме того, неправильное произношение неносителем языка может быть связано с разнообразием факторов, таких, как несовершенное понимание семантики, синтаксиса, морфологии, фонетики, эффектов коартикуляции и правил произношения.

C. Э-медицина

Распознавание речи доказало свою эффективность в оказании помощи врачам при создании электронных медицинских записей [6]. Сегодня десятки тысяч врачей используют распознавание голоса для диктовки результатов в электронных записях гораздо чаще, чем документирование результатов с помощью текстового набора или мыши. Преимущества речевых систем электронных медицинских записей включают:

- резкое снижение затрат на копирование;
- улучшение ухода за пациентами через полную документацию и быстрое получение результатов;
- сокращение времени, затрачиваемого на документирование карты пациента;
- увеличение продуктивности (работоспособности) врача.

Наличие у всех жителей электронной карты здоровья является движущей силой создания стандартов совместимости.

D. Речевые технологии в дистанционном обучении

Дистанционное обучение предполагает доставку образовательных услуг студентам, которые не присутствуют физически в учебном заведении. Это наиболее быстро растущий сегмент образования, особенно в сфере высшего образования и обучения взрослых [7].

Верификация диктора используется, чтобы убедиться, что человек, сдающий экзамен, является зарегистрированным студентом.

Аутентификация личности студента является особенно трудной в области электронного обучения, потому что учреждение никогда не может иметь прямого контакта с учеником. Необходимость решения этой проблемы в настоящее время поручена органам по сертификации. I Drive Safely и Language Testing International (LTI) – две такие компании, использующие верификацию диктора, чтобы помочь решить проблемы, связанные с аутентификацией.

Речь используется в дистанционном обучении для обучения пению, повышения грамотности, в режиме реального времени для транскрибирования телевизионных лекций, упражнений для пациентов, страдающих афазией (системное нарушение уже сформировавшейся речи [8]), и в качестве вспомогательных средств.

В Tell Me More распознавание речи является одним из компонентов педагогической системы, который предназначен для систематического улучшения произношения языка как родного.

E. Криминалистическая фоноскопия

Криминалистическая фоноскопия – это идентификация человека по голосу, она изучает звуковую информацию на звуковых, магнитных и других носителях [9].

В связи с развитием речевых технологий письменные документы чаще стали заменять видеосъемкой и звукозаписью, по которым можно установить личность человека.

Фоноскопическая экспертиза используется при расследовании уголовных дел по взяткам, коммерческому подкупу, вымогательствам, разнообразным мошенничествам и телефонному терроризму.

В методику фоноскопической экспертизы положены следующие виды анализа устной речи человека:

1) лингвистический – исследует устную речь человека, его интеллектуальные и психофизиологические особенности;

2) акустический – направлен на изучение анатомических и других особенностей определенного субъекта. Фоноскопическая экспертиза позволяет судить о физических и психических признаках человека.

III. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ АУДИОДАННЫХ

Развитие почти неограниченных объемов для хранения данных в сочетании с растущим использованием Интернета сделало возможным доступ к огромным объемам текста одним щелчком мыши. С развитием технологий еще большие объемы данных в форме речи будут получены из телевизионных передач, радио, телефонных звонков, конференций и презентаций. Интеллектуальный анализ аудиоданных (Audio mining)

превращает информацию во всех этих речевых данных в архивы, которые можно просматривать, искать и извлекать так же легко, как и текст.

Audio mining (извлечение данных из аудиозаписей), известный так же как поиск в аудиозаписях (audio searching), берет текстовый запрос и находит искомое слово или фразу в аудиофайле. Это помогает пользователям быстро находить определенные места в записанном разговоре [10].

Ряд различных терминов используется в связи с audio mining. К ним относятся: индексация аудиозаписи, фонетический поиск, фонетическое индексирование, индексация речи, аудио аналитика, анализа речи, выделение слов, информационный поиск.

Индексация аудиозаписи (audio indexing) заключается в разделении речи, записанной в аудио- или видеофайле, на слова, каждому из которых присваиваются индекс и временная метка. Затем этот индекс используется для быстрой идентификации заданного фрагмента речи (например, система Google Audio Indexing (GAUDI).

Серьезные исследования audio mining начались в конце 1970-х. Исследование ведется уже в ряде крупных школ, в том числе Университете Карнеги-Меллона, Колумбийском университете, Технологическом институте Джорджии и Университете штата Техас.

В последние годы эта технология начала предлагать приемлемый уровень производительности и точности для коммерческого использования. Продукты на основе audio mining были интегрированы в более крупные системы, потому что возможность поиска и воспроизведения контента является важным для управления большими архивами информации, такими, как медиа или системы управления контентом.

Есть два основных подхода к audio mining.

1) Текстовая индексация. Текстовая индексация, также известная как LVCSR, преобразует речь в текст, а затем идентифицирует слова в словаре, который может содержать несколько сотен тысяч записей. Если слово или имя отсутствует в словаре, система LVCSR выберет наиболее похожее слово.

LVCSR audio mining – это двухступенчатый процесс. На первом этапе (предварительная обработка или этап индексации) речь из аудиозаписи обрабатывается устройством распознавания с большим словарем для генерации поискового индексного файла. Индексный файл содержит информацию о последовательности слов, произнесенных в аудио или видеоданных.

На втором этапе (этап поиска) определяется поисковый термин (например, слово или фраза), и один или несколько индексных файлов ищут все упоминания, которые соответствуют указанному поисковому запросу. Результаты поиска могут отображаться в графическом виде или соответствующие разделы аудио или видеофайлов могут быть проиграны пользователю.

2) Фонемная индексация. Фонемная индексация в отличие от LVCSR работает только со звуками. Сначала система анализирует и идентифицирует звуки в аудиофрагменте, чтобы создать фонетический индекс. Затем он используется для преобразования поискового запроса пользователя в правильную строку фонем.

Сравнение фонетических подходов и LVCSR приводит к различным преимуществам и недостаткам.

У фонетического audio mining скорость, с которой аудиоконтент может быть индексирован, во много раз выше, чем для LVCSR подхода. На этапе поиска, однако, вычислительная нагрузка больше для фонетических систем.

Исторически сложилось так, что фундаментальные исследования распознавания речи были сосредоточены почти исключительно на оптимизации LVCSR. Основным стимулом для этого исследования были спонсируемые правительством США конкурсы, проводимые ежегодно агентством DARPA (Defense Advanced Research Projects Agency) [11]. Основной акцент этих конкурсов был направлен на улучшение точности распознавания для чтения предложенного материала из ряда стандартных источников (например, The Wall Street Journal или The New York Times).

Поиск audio mining обычно можно проводить во много тысяч раз быстрее, что делает возможным поиск больших объемов речевых данных, в зависимости от времени, которое потребуется для людей, чтобы прослушать материал.

Audio mining был также использован в качестве создания подписей (субтитров) к телевизионному и другому видео/медиа контенту. Однако более эффективным и действенным способом обнаружения начала и конца каждого слова в тексте заголовка является использование распознавания речи для автоматического выравнивания известного текста.

IV. Заключение

Рассматривая возможности развития речевых технологий в интересах электронного государства, можно выделить следующие приоритетные направления:

- разработка технологии распознавания слитной речи, включая готовые акустико-фонетические модели фонем речи для различных языков;
- создание комплекса программ восстановления искаженных и зашумленных речевых сообщений как в ограниченном (тематическом) словаре, так и смешанном;
- разработка поиска в информационных сетях речевой информации по заданным ключевым словам или проблематике с учетом современных технологий;
- создание интерпретаторов, верно передающих смысл сильно искаженного речевого сообщения;
- автоматизированное обнаружение в текстах и речевых сообщениях лингвистической информации, значимой для психофизиологической оценки и биометрического контроля.

В дальнейшем будут проводиться работы по реализации вышеперечисленных направлений с целью повышения качества услуг э-государства.

V. Благодарности

Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской

Республики – Грант № EİF-RİTN-MQM-2/İKT-2-2013-7(13)-29/18/1.

Библиография

[1] M.Warkentin, D.Gefen, P.A. Pavlou and M. Rose. Encouraging Citizen Adoption of e-Government by Building Trust, Electronic Markets, 2002, v.12, no.3, pp. 157-162.

[2] L.Carter, F.Bélanger., The Utilization of e-Government Services: Citizen Trust, Innovation and Acceptance Factors, Information Systems Journal, 2005, v.15, no.1, pp. 5–25.

[3] J.A.Qadri, M.T.Banday. Web Accessibility – A timely recognized challenge, The Business Review, 2009, v. 1,2, no.14, pp. 99–102.

[4] Повышение качества обслуживания в ЦТО МФЦ и на сайте предоставления государственных услуг. Центр речевых технологий,

[5] http://www.speechpro.ru/files/filefield_stats/1245/496/0/18bdb2613048c5698b9970e0a942e62a

[6] F.Ehsani and E.Knodt. Speech technology in computer-aided language learning, Language learning and technology, 1998, v.2, no.1, pp. 45–60.

[7] Speech Recognition: Accelerating the Adoption of Electronic Medical Records (White paper), Nuance Communications Inc., 2008.

[8] J.Markowitz, Speech for Distance Learning, Speech technology magazine, 2009, v.14, no.1, pp. 68.

[9] А.Р.Лурия. «Высшие корковые функции человека и их нарушения при локальных поражениях мозга», М.:Академпроект, 2000.

[10] А.Аленников, «Шпаргалка по криминалистике», М.:Аллель-2000, 2006.

[11] Manpreet Kaur Mand, Diana Nagpal, Gunjan, An Analytical Approach for Mining Audio Signals, International journal of advanced research in computer and communication engineering, 2013, v.2, no.9, pp. 3645–3647.

[12] L. Hirshman. Overview of the DARPA Speech and Natural Language Workshop, Proc. of the workshop on speech and natural language, 1989, pp. 1–2.