РАЗШИРЯВАНЕ ВЪЗМОЖНОСТИТЕ НА ИНФОРМАЦИОННА СИСТЕМА ЗА РАБОТА С ГЛАС

Веска Михова

Гл. ас. д-р в катедра „Информационни технологии и комуникации“, УНСС
e-mail: vmihova@unwe.bg

Резюме

Този научен доклад разглежда възможностите и перспективите за разширяване на функционалността на информационни системи, които използват гласов интерфейс. Гласовите технологии са доказали своята значимост в съвременния свят, предоставяйки удобен начин за комуникация между хората и машините. Докладът обръща внимание на иновативни решения и технологии, които могат да разширят приложенията и възможностите на такива системи.

Abstract

This scientific report examines the possibilities and prospects for expanding the functionality of information systems that use a voice interface. Voice technologies have proven their relevance in the modern world, providing a convenient way for humans and machines to communicate. The report highlights innovative solutions and technologies that can extend the applications and capabilities of such systems.

Ключови думи: гласови технологии, гласово разпознаване, гласов синтез
JEL: O30, C88

Въведение

Разширяването на възможностите на информационна система за работа с глас е важна стъпка в напредъка на технологичния свят и подобряването на взаимодействието между хората и компютрите. Включването на гласови интерфейси и обработка на гласова информация може да увеличи удобството и ефективността на множество приложения и системи.

Системите за работа с глас са внушителен пример за интелигентни информационни системи, които се основават на гласови команди и разпознаване на гласова информация. Те не само улесняват ежедневния живот на потребителите, но също така имат голям потенциал в разнообразни области като медицината, образованието, бизнеса и социалната интеграция на хора с увреждания. Този доклад изследва настоящите и бъдещи възможности за разширяване на функционалността на тези системи.

Предизвикателство в тази област, което е извън обхвата на доклада, е важната тема свързана със сигурност и поверителност на информацията. Трябва да се обърне особено внимание на сигурността на гласовата информация и личните данни на потребителите, особено при съхранението и обработката на чувствителни гласови записи.

Гласово разпознаване и синтез

Гласово разпознаване

Гласовото разпознаване е технология, която позволява компютърните системи да разпознават и интерпретират човешка реч. Това представлява процес на анализиране и преобразуване на гласовите сигнали в текст или друг вид машинно обработваем изход.

Използването на технологии за разпознаване на реч може да позволи на информационната система да преобразува гласови команди или гласови записи в текст, който след това може да бъде обработван и анализиран от информационни системи ([1], [2]).

Изследванията в областта на машинното обучение са позволили на системите за разпознаване на реч да стават по-точни и адаптивни. Интегрирането на напреднали алгоритми и модели може да увеличи точността при разпознаването на глас. Разширяването на способностите за работа с различни езици може да повиши достъпността и употребата на гласовите системи по света.

Гласовото разпознаване е напреднала технология със значителен потенциал за решаване на различни предизвикателства и за предоставяне на редица ползи в различни области. Някои от основните цели и приложения на гласовото разпознаване включват:

- Подобряване на потребителското изживяване;
- Автоматизиране на процеси и контрол на устройства чрез гласови команди;
- Облекчаване на функционални ограничения на хора със зрителни, моторни или други увреждания;
- Улесняване транскрипцията на медицински записи и подпомагане здравеопазването.
- Подобряване мултиезичната комуникация и взаимодействие с технологията.
- Използване за биометрична идентификация и аутентикация на потребители.
- Използване в образователни приложения, които улесняват ученето и комуникацията.
- Гласовото разпознаване стимулира разработката на нови услуги и приложения, които подобряват ежедневния живот и бизнес процесите.

Освен всичко това, гласовото разпознаване продължава да е фокус в иновациите в областта на изкуствения интелект и машинното обучение, като целта е постигане на по-голяма точност и ефективност в разпознаването на различни езици и диалекти.

Гласов синтез

Гласовият синтез е технология, която се използва за генериране на изкуствено създадени гласови сигнали или реч посредством компютърни програми и системи. Този процес включва преобразуване на текст или друга форма на символна информация в аудио формат, който след това може да бъде изговорен или чут от хора чрез високоговорители или слушалки.

Гласовият синтез имитира човешкия глас и може да се използва за множество цели, включително:

- Предоставяне на аудио книги и текстове в говорима форма за лица със зрителни увреждания или с дислексия.
- Разработка на гласови асистенти и виртуални говорители, които да предоставят информация и да извършват задачи посредством комуникация с потребителите.
- Гласови обявления и реклами.
- Интерактивни гласови системи за обучение и образование.
- Аудио-синтез на данни и информация за потребление на медицински съвети, новинарски приложения и др.

Съвременните системи за гласов синтез се базират на компютърни алгоритми и изкуствени невронни мрежи, които могат да имитират различни интонации, тонове и емоции в човешкия глас, правейки ги по-натурални и приятни за слушателя.

Преобразуването на текстова информация в аудио в една система може да бъде реализирано посредством гласов синтезатор с API за гласов синтез. Интегрирането на API за гласов синтез може да предостави информация на потребителите на една система чрез глас. Синтезът на реч може да бъде използван, за да позволи на системата да отговаря на потребителите с гласови отговори. Този процес включва преобразуване на текст в артикулирана реч.

Вграждането на емоционален израз в гласовия синтез може да направи комуникацията по-човешка и емоционално адекватна. Комуникацията между потребителите и системите може да бъде улеснена и чрез възможността на потребителите да избират различни гласове за системата, както и персонализация на гласовите настройки, тон и скорост на речта, които отговарят на техните предпочитания ([3], [4])

Инструменти и библиотеки за гласово разпознаване

Гласовото разпознаване е процес на трансформиране на изречение, произнесено на човешки глас, в текст. Това е ключова технология, използвана в гласови асистенти, системи за навигация, софтуер за транскрипция и други приложения.

Водещите технологии и инструменти за гласово разпознаване притежават характеристики като:

- Висока точност при разпознаването на речта, както и поддръжка на различни езици, акценти, а също така и на диалекти.
- Позволяват обработка на аудио данни в реално време и поддържат паралелна обработка за бърз отговор.
- Разполагат с вградени професионални модели за различни области като медицина, финанси и други.
- Позволяват създаването на персонализирани модели, които се обучават да разпознават специфични термини и изрази.
- Различни ценови планове, включително безплатен план с ограничен брой безплатни заявки в месец и тарифи за по-големи обеми.

Водещи инструменти и библиотеки за гласово разпознаване

Google Cloud Speech-to-Text:

- Платформа за гласово разпознаване, предоставяща висока точност и поддръжка на различни езици. Интегрира се с облачните услуги на Google.
- Предлага безплатен тестов период с кредит за потребление на услугите на Google Cloud. Може да се използва безплатен период, след което се налага разплащане според тарифите на Google Cloud.

Microsoft Azure Speech Services:

- Облачна услуга за гласово разпознаване от Microsoft. Предлага възможности за разпознаване на реч, конверсия на реч в текст и обратно, както и аудио анализ.
- Лесно може да се интегрира с други услуги в облачната платформа на Azure, като Azure Cognitive Services и Azure Logic Apps.
- Azure Speech SDK предоставя безплатен план с ограничен брой заявки в месец. Има и допълнителни тарифи за по-големи обеми.

IBM Watson Speech to Text:

- Система за гласово разпознаване, използваща технологии на изкуствения интелект. Предоставя API за интеграция в различни приложения.
- IBM Watson предлага безплатен план с ограничение на броя на заявките в месеца. Необходима е регистрация и използване на ключове за достъп.

Amazon Transcribe:

- Услуга за автоматично гласово разпознаване от Amazon Web Services. Поддържа различни езици и предоставя възможности за транскрипция на аудиофайлове.

Dragon NaturallySpeaking:

- Продукт на Nuance Communications, предлагащ точно и бързо гласово разпознаване. Използва се в медицински, правни и други професионални среди.

Някои компании предоставят безплатни API услуги за гласово разпознаване с определени ограничения. Важно е да се отбележи, че безплатните планове обикновено имат ограничен брой заявки в месеца или предоставят базови функционалности. Няколко такива безплатни API услуги:

Wit.ai:

- Wit.ai е платформа с отворен код, управлявана от Facebook, за обработка на естествен език (NLP), която предоставя API за гласово разпознаване и обработка на текст.
- Предоставя безплатен план с ограничение на броя на заявките и обема на обучение.

PocketSphinx:

- PocketSphinx е библиотека с отворен код за гласово разпознаване, предназначена за вградени системи и устройства с ограничени ресурси. Това е част от проекта CMU Sphinx (Carnegie Mellon University Sphinx), който се занимава с разработването на технологии за автоматично разпознаване на реч.
- Проектиран да бъде лек и ефективен, подходящ за вграждане в устройства с ограничени ресурси, като например мобилни устройства и вградени системи. Подходящ за вграждане в умни устройства, които използват гласови команди за управление.
- Позволява офлайн работа, като не изисква постоянна връзка с интернет.

CMU Sphinx (Sphinx-4):

- Система за гласово разпознаване, разработена от Карнеги-Мелън Университет. Sphinx-4 е с отворен код и предоставя гъвкавост, но разработчиците трябва да имат познания в Java за оптимално използване. Предоставя инструменти за изследователи и разработчици.

SpeechRecognition (Python библиотека):

- Проста и лесна за използване библиотека за гласово разпознаване в програмен език Python.
- Тази библиотека използва различни back-end двигатели за гласово разпознаване, като Google Web Speech API, Sphinx, Microsoft Bing Voice Recognition и други.
- SpeechRecognition се предоставя под лиценза Apache 2.0, който позволява свободно използване и промяна на кода.

Kaldi:

- Отворена платформа за гласово разпознаване, разработена от университета в Джонс Хопкинс. с акцент върху иновации и ефективност в обучението на модели.

Mozilla DeepSpeech:

- DeepSpeech на Mozilla е проект с отворен код, насочен към разработването на система за автоматично разпознаване на реч (ASR).
- Използва DeepSpeech модел, базиран на рекурентни невронни мрежи (RNN), което позволява на системата да улавя сложни зависимости в речта.
- Тази система е разработена от Mozilla и използва DeepSpeech архитектурата, която е основана на рекурентни невронни мрежи (RNN) и е направена с отворен код с цел предоставяне на свободно налични средства за гласово разпознаване.
- Може да се използва с обучен модел или да се обучи собствен модел.

Тези технологии и инструменти предоставят широк спектър от възможности за разработка на приложения, които използват гласово разпознаване, независимо дали става въпрос за създаване на гласов асистент, транскрипция на аудио, или други приложения ([5], [6]).

Инструменти и библиотеки за гласов синтез

Гласовият синтез е процесът на генериране на изкуствени гласове или реч чрез компютърни програми и системи ([7], [8]). Тази технология има различни приложения, включително гласови асистенти, аудио книги, системи за навигация и други. Някои от водещите технологии и инструменти за гласов синтез са:

Google Text-to-Speech:

- Облачна услуга от Google, която преобразува текст в гласов сигнал. Предлага различни гласове и поддържа многоезичие.

Amazon Polly:

- Услуга за гласов синтез от Amazon Web Services. Предоставя голям набор от гласове и опции за персонализация.

Microsoft Azure Text to Speech:

- Облачна услуга, предоставена от Microsoft Azure, която преобразува текст в гласов сигнал. Поддържа няколко езика и гласове.

IBM Watson Text to Speech:

- Система, използваща технологии на изкуствения интелект, която преобразува текст в реч. Предоставя възможности за персонализация на гласовете.

Nuance Communications:

- Компания, която предлага различни продукти за гласов синтез, включително Dragon NaturallySpeaking и други инструменти за гласова комуникация.

Festival Speech Synthesis System:

- Отворен проект, който предоставя гласов синтез в реално време. Често се използва в научни и образователни приложения.

CereProc:

- Компания, специализирана в гласов синтез, предлагаща персонализирани гласове и услуги за клиенти.

Neospeech:

- Компания, която предлага системи за гласов синтез със синтетични гласове, които са близки до човешките.

ResponsiveVoice:

- Уеб базиран инструмент за гласов синтез, който може да бъде лесно интегриран в уеб сайтове и приложения.

eSpeak:

- Компактен и отворен инструмент за гласов синтез, който се фокусира върху простота и ефективност.

Гласовият синтез се развива бързо, и тези технологии предоставят широк спектър от възможности за приложения в различни области, включително виртуални асистенти, образование, развлечения и много други ([9]).

Заключение

Разширяването на възможностите на информационна система за работа с глас може да подобри потребителския опит, да увеличи ефективността и да създаде нови възможности за бизнеса.

Технологичните иновации, обсъдени в настоящия доклад, предоставят основа за бъдещето на гласовите системи и тяхната роля в съвременния свят.

Този доклад е само начало на разглеждането на темата за системи за работа с глас и може да послужи като отправна точка за допълнителни изследвания и иновации в тази област.

Литературни източници

1. Neha Jain, Somya Rastogi, 2019, Speech recognition systems - a comprehensive study of concepts and mechanism, Acta Informatica Malaysia (AIM), https://actainformaticamalaysia.com/archives/AIM/1aim2019/1aim2019-01-03.pdf
2. Dong Yu , Li Deng, 2015, Automatic Speech Recognition A Deep Learning Approach, Springer, https://link.springer.com/book/10.1007/978-1-4471-5779-3
3. Sabato Marco Siniscalchi, Torbjørn Svendsen, Chin-Hui Lee, 2014, An artificial neural network approach to automatic speech processing, Neurocomputing Volume 140, 22 September 2014, Pages 326-338.
4. Andreas M. Klein, Kristina Kölln, Jana Deutschländer & Maria Rauschenberger, 2023, Design and Evaluation of Voice User Interfaces: What Should One Consider?
5. International Conference on Human-Computer Interaction, HCII 2023: Design, Operation and Evaluation of Mobile Communications pp 167–190.
6. Plamen Milev, "Approach for Analysis and Comparison of Search Query Results in Web Publications, 11th International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2021), November 25-26th, 2021, UNWE, Sofia, Bulgaria.
7. Митко Радоев, 2021, Съвременни тенденции в развитието на базите от данни, Икономически алтернативи, https://www.unwe.bg/doi/alternativi/2021.1/ISA.2021.1.01.pdf
8. Monika Tsaneva, 2019, A practical approach for integrating heterogeneous systems - Бизнес управление.
9. Geno Stefanov, Maria Marzovanova, Building IoT Solution for Better University Using IBM Watson IoT Platform 2017, ICAICTSEE - 2017, http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2017.pdf
10. E. Karkalikova, A. Murdjeva, Organization of Data in Data Lake – Real-Life Practice, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.