

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА КОМП'ЮТЕРНА ТЕХНІКА

УДК 681.3.019:621.39

М. М. Биков, канд. техн. наук, доц.; В. В. Ковтун, канд. техн. наук, доц.;
Н. Г. Савінова, асп.

АНАЛІЗ СТАНУ ПРОБЛЕМИ РОЗРОБКИ ЕФЕКТИВНИХ СИСТЕМ ПОШУКУ КЛЮЧОВИХ СЛІВ

Проведено аналіз програмних засобів і технологій, які застосовуються для створення систем розпізнавання мови. Основний акцент зроблено на програмних засобах, які використовуються для створення систем пошуку ключових слів. Проведено аналіз недоліків існуючих програмних засобів і технологій та виявлено шляхи підвищення якості систем пошуку ключових слів.

Вступ

Розвиток і утвердження у всьому світі технологій автоматичного розпізнавання мови (АРМ) зумовили виникнення додаткових напрямків і застосувань, що ґрунтуються на досягненнях цих технологій. Одним з нових напрямків, у якому застосовується інструментарій АРМ, є автоматичний аналіз неперервної розмови. В цьому напрямку зазвичай вирішуються задачі таких типів: розпізнавання мови, аналіз мови і добування знань, та низка інших. Однією з найважливіших проблем цього напрямку, що вимагає застосування інструментарію АРМ, є пошук і розпізнавання ключових слів в мовному сигналі. Пошук ключових слів (KWS – Keyword Spotting) є одним із ефективних способів автоматичного пошуку необхідного фрагмента фонограми в величезних звукових базах чи звукових потоках. Проблема пошуку ключових слів в злитій мові актуальна в системах безпеки, телефонних сервісах, а також в інших прикладних задачах. Її розв'язання дозволить автоматично здійснювати ретельний контроль за можливими загрозами національній безпеці, загрозами функціонуванню і діяльності установ і людей, а також здійснювати мовне фільтрування (наприклад, нецензурної лексики).

Метою роботи є аналіз сучасного стану проблеми побудови KWS систем і визначення задач, розв'язання яких дозволило б підвищити ефективність їх роботи.

Порівняльний аналіз існуючих KWS систем

На сьогодні існує декілька підходів до пошуку ключових слів:

- KWS, що ґрунтуються на розпізнаванні слів;
- KWS, що ґрунтуються на розпізнаванні послідовності фонем;
- KWS, що ґрунтуються на розпізнаванні зливої мови [1].

При цьому для співставлення мовних фрагментів використовуються методи динамічного програмування і статистичні методи, що ґрунтуються на прихованих Марковських моделях. Концепція KWS має ряд особливостей в порівнянні з методами, вживаними в областях розпізнавання мови. Під час використання технології пошуку доводиться мати справу зі зливою мовою двох або більшої кількості дикторів. Розмова може йти на довільні теми, які можуть мати необмежений словниковий склад (при цьому як диктор, так і його мова можуть змінюватися). У цьому випадку для пошуку з використанням шаблонів необхідно мати еталони довільного характеру.

До основних критеріїв оцінки ефективності систем пошуку ключових слів можна віднести імовірність правильного розпізнавання та вартість.

На сьогодні відомо багато KWS систем, в яких використовуються різні технології пошуку

слів. Наведемо дані про принципи їх реалізації на прикладі деяких зразків таких систем.

Технологія пошуку ключових слів у фонограмах мови для російської мови — Voice Digger (Компанія «Центр Речевих Технологій»), основана на неперервному розпізнаванні мови, що реалізується на акустичних прихованих Марковських моделях. Ключові слова задаються у вигляді звичайного набраного тексту, за яким система будує приховану Марківську модель кожного слова. Крім того в системі задається так звана фонова модель — модель загальної мови. При побудові моделей ключових слів використовується транскриптор російської мови та акустичні моделі фонів для російської мови. На виході Voice Digger надає посилання на звуковий документ і місце розташування шуканого слова або словосполучення. Метод є досить швидким і не залежить від словника. Рівень помилки становить 8 %. Voice Digger є першим у світі комерційним продуктом подібного класу для російської мови.

Така технологія використовується для швидкого і ефективного пошуку заданої інформації у звукових архівах без попереднього прослуховування. Вона дає посилання на звуковий документ і місце розташування знайденого в ньому слова чи словосполучення, вирішує проблему автоматичної мовної аналітики і дозволяє автоматично визначати тему повідомлення та класифікувати повідомлення на теми. Технологія є швидкою і забезпечує високий рівень надійності пошуку заданих слів у фонограмах мови або звукових архівах в автоматичному режимі. Після відповідних доопрацювань можливе застосування технології і для інших мов [2].

Із недоліків Voice Digger як комерційної системи, можна виділити обмеженість вхідних аудіо форматів та недосконалість словника словосполучень для пошуку, в якому користувачу необхідно дотримуватись чітких правил та вимог, що для сучасної системи такого рівня є суттєвим недоліком.

Система пошуку ключових слів Санкт-Петербурзького державного університету телекомунікацій ім. Проф. М. А. Бонч-Бруєвича є багаторівневою і складається з декількох модулів. Виділення інформативних ознак сигналу проводиться на базі апарату лінійного передбачення. Пошук за словником еталонів виконується з використанням динамічного програмування [3]. Тим не менш у цій системі відсутні відомості про системний підхід, які не дають можливість зробити висновки про ефективність системи в цілому.

Система пошуку ключових слів Білоруського державного університету основана на решітках складів. Останнім часом досить широкого розповсюдження набула ідея побудови і використання решітки фрагментів мови для вирішення різних задач. Для пошуку ключових слів кожен вузол решітки асоціюється з моментом часу мови. Головна перевага цього методу полягає в тому, що він має велику гнучкість: навіть якщо фонема ключового слова не є кращою гіпотезою між вузлами решітки, вона все одно зберігається в результаті розпізнавання. Результат пошуку не залежить від словника розпізнавача, оскільки пошук можна організувати для будь-якої фонемної послідовності ключового слова [4]. Для цієї системи відсутнє готове для тестування та випробування програмне забезпечення, тому всі характеристики системи носять лише теоретичний характер, які обов'язково потрібно перевірити на практиці надалі.

Phonexia. У цій технології використовується акустичний підхід (ключове слово шукається з використанням шаблонів), оснований на штучних нейронних мережах. Основною перевагою такого підходу є його швидкодія.

Мови системи: англійська, російська, чеська, словацька, угорська та польська. Необмежена кількість ключових слів [5, 6]. На практиці система показала себе дуже дикторозалежною. Порівняно висока ефективність пошуку забезпечується лише в ручному режимі. А можливість застосування програмного модуля для української та російської мов можливо лише після попередньої домовленості із виробником, що не дало змогу в повному об'ємі провести його тестування.

В таблиці наведено порівняльну характеристику зазначених систем пошуку ключових слів.

Проведений детальний аналіз роботи вказаних KWS систем, опис якого виходить за рамки цієї роботи, дав можливість зробити висновки про наявність низки проблем, розв'язання яких дозволило б підвищити ефективність їх роботи.

Порівняльна характеристика систем пошуку ключових слів

Назва системи пошуку ключових слів	Країна, місто	Точність розпізнавання	Розмір словника	Вартість
Voice Digger	Росія, Санкт-Петербург	рівень помилки 8 %	не вказано	75000 тис. руб.
KWSS Санкт-Петербурзького державного університету телекомунікацій	Росія, Санкт-Петербург	відсоток правильно знайдених слів 90 %	30 слів	некомерційний продукт
KWSS Білоруського державного університету	Білорусь, Мінськ	86,2 %	не вказано	некомерційний продукт
Phonexia	Чехія, Брно	не вказано	необмежена	не вказано

Висновки

Незважаючи на досягнуті успіхи цих систем, дуже актуальними на сьогодні залишаються роботи, спрямовані на підвищення точності, швидкості і надійності систем пошуку ключових слів.

В результаті порівняння і аналізу можна зробити висновок, що на сьогодні невирішеними залишаються такі проблеми розробки ефективних систем пошуку ключових слів як:

- розробка ефективної стратегії автоматизованого розпізнавання (пошуку) ключових слів;
- розробка методів і моделей сегментації звукового потоку на інформативні елементи;
- побудова моделей часових параметрів інформативних елементів мовних сигналів;
- розробка покращених процедур розпізнавання методом динамічного розпізнавання з використанням реперних точок;
- удосконалення методів опису мовного сигналу за рахунок використання нечітких параметрів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Автоматический поиск ключевых слов в непрерывном потоке речи на основе технологии «распознавание через синтез» [Електронний ресурс] / Диалог 2006 : труды международной конференции / В. В. Киселёв, А. О. Таланов, М. Ю. Татарникова, Ю. Ю. Хохлов. — Режим доступа : <http://www.dialog-21.ru/dialog2006/materials/pdf/Kiselev.pdf>.
2. Voice Digger [Електронний ресурс] : Компанія «Центр Речевых Технологий». — Режим доступу : <http://www.speechpro.ru/product/recognition/asr/voice-digger/docs>. — Назва з екрану.
3. Распознавание отдельных слов в разговорной речи [Електронний ресурс] : VI Всероссийская межвузовская конференция молодых ученых: тез. докл. / К. К. Гладышев // ИТМО. — СПб, 2009.
4. Поиск ключевых слов с использованием решетки слогов [Електронний ресурс] : международная конференция по компьютерной лингвистике 2009 / Р. М. Алиев, Янь Цзинбинь, И. Э. Хейдоров — Режим доступа : <http://www.dialog-21.ru/dialog2009/materials/html/01.htm>.
5. Phonexia Keyword Spotting [Електронний ресурс] : Компанія «Phonexia» — Режим доступу : http://www.phonexia.com/docs/white/AcousticKWS_v3.pdf. — Назва з екрану.
6. Phonexia Keyword Spotting [Електронний ресурс] : Компанія «Phonexia» — Режим доступу : <http://www.phonexia.com/download/demo-kws>. — Назва з екрану.

Рекомендована кафедрою комп'ютерних систем управління

Стаття надійшла до редакції 20.05.11
Рекомендована до друку 21.11.11

Биков Микола Максимович — професор; **Ковтун В'ячеслав Васильович** — доцент; **Савінова Наталія Геннадіївна** — аспірант.

Кафедра комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця