УДК 004.8+655.2

АНАЛІТИЧНІ ДОСЛІДЖЕННЯ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ МОВЛЕННЯ

З. Сельменська, М. Дубневич, З. Плахтина, А. Цебрик

Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

Розглядається питання застосування систем розпізнавання мовлення для введення текстової інформації в комп'ютерні видавничі системи. Проведено аналітичні дослідження технологій розпізнавання мовлення та програмного забезпечення, що дозволяє голосове введення інформації персональним комп'ютером. Здійснено аналіз наявних методів розпізнавання мови людини. Розглянуто основні типи задач, які не можуть існувати без систем розпізнавання. Проаналізовано основні теоретичні положення про системи розпізнавання мовлення, висвітлено проблеми, які виникають під час використання цих систем у програмному забезпеченні і шляхи їх вирішення.

Ключові слова: *розпізнавання мовлення, глибинне навчання, голосове керування, голосове введення.*

У багатьох користувачів, чия діяльність пов'язана з наборами великих обсягів тексту, часто виникає бажання якось прискорити цей процес. Одні відточують метод сліпого друку, інші залучають підчитчиків (диктування, паралельний набір різних частин тексту), а хтось використовує сучасні новації, що вже стали буденністю нашого життя. Серед останніх чільне місце займає голосовий набір тексту, що дає змогу значно прискорити процес введення [3].

Сучасні технології голосового введення інформації надають користувачам безліч можливостей для оптимізації та полегшення роботи і економії часу. Нікого вже не здивує програма для перетворення голосу в текст, що набирає замість людини все те, що вона вимовляє. Розвиватися в цьому напрямку ще є куди, але і на сьогодні можна знайти цілком гідні сервіси та софт для мовного спілкування з комп'ютером. Системи розпізнавання мови оцифровують звук, що надходить з мікрофона та ідентифікують інформацію, звертаючись до наявних словників (софт може підтримувати різні мови і мати великий словниковий запас), після чого виводять на екран вже надрукований текст або задають різні команди [9].

На сьогодні величезні зусилля скеровуються на науково-дослідні розробки для вирішення проблем автоматичного розпізнавання і розуміння мови. Водночас потрібно звернути увагу на те, що в практичному використанні відсутні системи, які вважаються вершиною розвитку автоматичного розпізнавання мови.

Отже, концепція розпізнавання мовлення — здатність комп'ютерів розпізнавати та інтерпретувати мовлення — не є новою. Ця тема цікавила комп'ютерну

індустрію з моменту створення комп'ютерів. Розпізнавання голосу колись було далекою мрією, але тепер стало повсякденною реальністю. Ідея досить проста: для розпізнавання мови використовується мікрофон, підключений до комп'ютера, в якому працює програма розпізнавання мовлення. Зазвичай він збирає слова, вимовлені в мікрофон, а потім перетворює аналоговий звук голосу в цифрові дані, які потім обробляються програмою розпізнавання мови.

Проблема розпізнавання мови на сьогодні вважається надзвичайно серйозною і відіграє важливу роль у спілкуванні людини з машиною. Управління об'єктами за допомогою мови відкрило б широкі перспективи перед автоматизацією у багатьох галузях людської діяльності, відкрило б можливість спілкування з машинами, особливо користувачів персональних комп'ютерів, не знаючих мов програмування. Мовний контакт полегшує запис даних у машину, допомагає працювати людині і комп'ютеру в реальному часі: людина сказала — машина виконала.

Постановка проблеми. Повсякденне життя настільки стрімке, що інколи навіть бракує часу на виконання звичних справ на комп'ютері у звичний спосіб — за допомогою клавіатури та мишки. На зміну прийшли нові технології, що дають змогу керувати персональним комп'ютером за допомогою голосу. Програмне забезпечення для розпізнавання мовлення розроблялося, щоб забезпечити швидкий спосіб введення інформації у комп'ютер, а також може допомогти людям з обмеженими фізичними можливостями. Програми розпізнавання голосу працюють, аналізуючи звуки та перетворюючи їх у текст. Після правильного налаштування система повинна розпізнавати приблизно 95 % інформації, але за умови, що користувач чітко вимовляє слова.

Аналіз останніх досліджень та публікацій. Системи розпізнавання голосу — це обчислювальні системи, які можуть визначати мову людини із загального потоку. Ця технологія пов'язана із технологією розпізнавання мови, яка перетворює вимовлені слова в цифрові текстові сигнали шляхом проведення процесу розпізнавання мови машинами [1].

Загальновідомо, що найважливішим засобом комунікації є мова. Процес розпізнавання мови має на меті формування, сприйняття і розуміння певних мовних конструкцій [6]. У сучасному світі інформаційних технологій голосове управління породжує новий спосіб взаємодії з функціями різних пристроїв [2] . Для створення такої технології необхідно було вирішити завдання розпізнавання мови.

Проблема розпізнавання вербальної інформації — розпізнавання мови — процес перетворення мовного сигналу у цифрову інформацію (наприклад, текстові дані). У будь-якої людини є свої особливі вокальні характеристики, які визначаються індивідуальною структурою його голосового апарату. Задача розпізнавання людини за голосом полягає у виділенні з вхідного аудіопотоку людської мови, її класифікацію і розпізнавання. При цьому зазвичай вирішуються дві підзадачі: розпізнавання мовця і перевірка. Алгоритм ідентифікації мовця можна також визначити як текстозалежний і текстонезалежний. Якщо алгоритм ідентифікації мови залежить від тексту, то в ньому можна використовувати як фіксовані заздалегідь фрази, так і фрази, які генеруються системою розпізнавання. Текстонезалежні системи необхідні

для обробки довільної мови. Починаючи з 1980-х років у розпізнаванні мови найгостріше стоїть проблема наявності перешкод. Системи ефективно працюють в ідеальних умовах запису, але при цьому не справляються з фоновими шумами на кшталт звуку із сусідньої кімнати. Штучні шуми виділити цілком можливо, але важко відрізнити голос людини, який нам потрібно розпізнати, від голосу людини, що розмовляє по сусідству. Проблема завадостійкості до цього часу не вирішена [3]. Сьогодні розпізнавання мови зводиться до вирішення трьох типів завдань: 1) розпізнавання окремо вимовлених слів (використовується для мовного управління обчислювальною машиною); 2) розпізнавання злитого мовлення (має на меті перетворення в текст природної мови людини); 3) ідентифікація за зразком мови (використовується для цілей забезпечення безпеки) [10]. У процесі реєстрації користувача запам'ятовуються особливості його голосу і формується так звана мовна модель. Під час тестування виконується порівняння запропонованого зразка мови із запам'ятованою мовною моделлю користувача, а також з моделлю «самозванця», складеної на базі голосів безлічі інших людей. Якщо результат порівняння виявиться позитивним для першого випадку і негативним для другого, вважається, що тестування пройшло успішно [5].

На сьогодні на ринку представлені такі основні системи, які використовуються для автоматичного розпізнавання мови: − Dragon NaturallySpeaking − IBM ViaVoice Gold − L&H Voice Xpress Professional − Philips FreeSpeech 2000 [4]. Вони вважаються найкращими, але ні одна із них не є ідеальною, основні їхні недоліки: рівень безпомилковості розпізнавання мови не перевищує 85 %; нерівномірна якість розпізнавання; низька якість розпізнавання власних назв і скорочених слів, повільна робота в середовищі деяких програм; затрата великого часу для налаштування системи.

Натепер створено цілу низку програм, які дають змогу не набирати текст вручну, а голосом надиктовувати всю необхідну інформацію.

Поширеним прикладом є голосове введення тексту в браузері **Google Chrome - Google Web Speech.** Мобільні пристрої від компанії Apple забезпечені спеціальним додатком для розпізнавання мови на базі «голосового движка» **Siri.** Програми **RealSpeaker PRO і Speechka** так само непогано справляються зі своєю функцією розпізнавання тексту, але вони використовують розробки компанії Google. Існують і онлайн-сервіси, наприклад **Speechpad.ru**, які дають змогу набирати текст голосом. Speechpad працює тільки в браузері Google Chrome. Створено системи, що здійснюють розпізнавання ізольовано вимовлених слів з певного словника (об'ємом від 20 до 1000 слів), розпізнавання зв'язної мови, що складається зі слів вибраного словника (об'ємом до 1000 слів), розпізнавання і смислову інтерпретацію злитого мовлення на штучній або природній мовах певної предметної галузі. Розпізнавання мови — це процес, пов'язаний з фонемним перекодуванням мовного акустичного сигналу. Під час розв'язування задач розпізнавання мовлення широко використовують методи математичної статистики, граматик формальних, математичного програмування тощо.

Програма **RealSpeaker** дає змогу вводити текст будь-якої довжини за допомогою голосу. Можна використовувати онлайн або офлайн текстовий редактор

(блокнот, MS Word, Skype, Facebook, Evernote тощо); вводити текст на будь-якій з восьми мов: українська, англійська (британська і американська), французька, німецька, китайська, російська, японська та корейська. Для зміни мови необхідно перемкнути мову на самому пристрої, далі потрібно запустити програму і відкрити текстовий редактор, після чого можна форматувати звук в текст.

Intelligent Voice Operating System (IVOS) (розробник Comun X). IVOS дає змогу: а) розпізнавати мову і перетворювати її в текст в будь-якому Windows-сумісному текст-процесорі; б) керувати своїм ПК за допомогою різноманітних голосових команд, а також створювати свої власні; в) озвучувати електронні книги за допомогою зовнішніх голосових движків.

Voxx (розробник Voxx Support Team). Можливості програми нагадують IVOS (стенографування/голосові команди/читання тексту) за винятком того, що в ній є корисний бонус — озвучування кожної дії: набір тексту або відкриття файлу.

Via Voice. Функція розпізнавання мови IBM відрізняється своєю здатністю від початку, без навчання, розпізнавати до 80 % слів. Під час навчання ймовірність правильного розпізнавання підвищується до 95 %.

Розроблені спеціалізовані модулі, які можна втілювати в текстові редактори:

VoiceCode — дає змогу набирати чистий програмний код за допомогою голосових команд, не використовуючи клавіатуру. VoiceCode дозволяє диктувати код природним чином, при цьому автоматично перетворює людську мову в специфічні програмістські функції. Програма працює тільки з однією мовою програмування Python, але її можна практично без проблем адаптувати під інші мови програмування. **EmacsListen** — програмний модуль, що виконує голосові функції текстового редактора GNU Emacs. Він постачається з граматикою ShortTalk, має підтримку розпізнавання і нормалізації тексту. Модуль можна використати для реалізації інших мовних інтерфейсів. **Voice Grip** — додатковий макрос для редактора Emacs, який створений з метою спрощення розпізнавання мови для програмістів. **Java by voice** — серія макросів для редактора Emacs, які спроєктовані для спрощеного введення коду мовою Java. **Cache Pad — макрос** для редактора Emacs для кешування недавно продиктованих імен функцій і змінних. **Emacs VR Mode — макрос** для редактора Emacs, що містить функціонал «Select and say» з Dragon Naturally-Speaking [6].

Комп'ютерні системи розпізнавання мови поступово знаходять застосування не тільки в науковій сфері, а й у побутовій. Прикладом тому можуть бути офісні пакети та інше програмне забезпечення з вбудованим розпізнаванням мови для голосового введення текстової інформації. Для того щоб машина навчилась розуміти людську мову, відповідати на запитання потрібно затратити багато сил і часу, забиваючи її великим обсягом інформації тільки для того, щоб розпізнати окремі звуки. У кожного звуку складна структура, яка містить різні частоти і коливання, крім того, те саме слово різні люди вимовляють по-різному: різний тембр голосу, різні інтонації, різна чистота вимови. Скільки людей, стільки й голосів. Голос — індивідуальна ознака особистості. Щоб навчити машину впізнавати мову, її потрібно змусити прослуховувати слова, сказані як однією людиною, так і

різними людьми. Задача машини — прослухавши всі дані, взяти середні значення особливостей вимови, повністю нівелювати індивідуальність, щоб потім, почувши слово, не зробити помилку. Найбільші проблеми виникають в умовах: довільний користувач; спонтанна мова, яка супроводжується мовним «сміттям», наявність акустичних перешкод і викривлень; наявність мовних перешкод. Для спрощення процесу розпізнавання мови доцільно було б використовувати шаблони окремих звуків, які є єдині для всіх дикторів. На сьогодні таких шаблонів не існує через те, що не виявлено інформативних ознак звуків, які не залежать від характерних особливостей голосу. Тому для реалізації ефективних дикторнезалежних систем автоматизованого розпізнавання мови необхідно виділити інформативні ознаки звуків мови, розробити математичні методи їх опрацювання з метою створення єдиних для всіх дикторів шаблонів. У такому випадку система розпізнавання не буде потребувати навчання (створення набору шаблонів окремо для кожного диктора), її швидкодія збільшиться, оскільки відпаде потреба створення набору шаблонів слів і з'явиться можливість розпізнавати мову незалежно від характерних особливостей голосу диктора [11].

Аналіз технологій розпізнавання мовлення зводиться до задач та обмеження технології розпізнавання мовлення. У процесі дослідження розпізнавання природної мови досягнуто значних результатів, серед яких розробка потужних лексикографічних систем, програм для машинного перекладу, електронних словників та ін. Проте існує досі невирішена проблема, яка криється у природі людської мови. Проблема розуміння людського мовлення полягає саме у його неоднозначності. Виділяють такі види неоднозначностей [5]: синтаксична, смислова, відмінкова, референційна.

Важливою проблемою у процесі обробки природної мови є проблема синонімії, в результаті якої одне поняття може бути виражене декількома різними словами.

Вплив вищенаведених явищ є актуальним під час створення систем розпізнавання мовлення. Проблема полягає у складності встановлення конкретного відображення дійсної семантико-синтаксичної структури речення у його внутрішнє логічне уявлення, яке автоматично генерується системою [1].

Такі типи неоднозначностей можливо розв'язати за допомогою введення додаткових значень, які збільшать інформацію програми про ту чи іншу галузь. На сьогодні не існує програм, які «розуміють» усі типи неоднозначностей у великому спектрі галузей, проте є програми, що можуть коректно реагувати на неоднозначності у дуже вузьких сферах, зокрема поліграфії.

Існує дві незалежні задачі процесу розпізнавання мови — це задача локального розпізнавання мови та задача відновлення тексту неперервної мови за множиною можливих гіпотез розпізнавання.

Отже, технологія з розпізнавання мовлення вирішує такі завдання:

- синтез мовлення — озвучення/читання тексту голосом наближеним до природнього;
- розпізнавання мови — відповідає за виведення або розпізнавання тексту від сканованих документів або файлів у PDF форматі;

– генерування природної мови — конвертування комп'ютерних даних у природну мову людини;
– машинний переклад — автоматичний переклад з однієї мови на іншу;
– питально-відповідальні системи — відповіді на питання, поставлені людською мовою;
– розпізнавання/визначення теми — полягає у поділі тексту на частини та визначенням провідної теми для кожної з частин;
– інформаційний пошук — полягає у пошуку, розпізнаванні та вилученні інформації;
– отримання інформації — полягає у вилученні семантичної інформації з тексту;
– отримання зв'язків — полягає у визначенні зв'язків між об'єктами у певній частині тексту;
– спрощення тексту — зміна, розширення або інша обробка інформації для спрощення структури або граматики тексту зі збереженням суті та змісту;
– розв'язання лексичної багатоманітності — надання списку можливих значень конкретного багатозначного слова, з поміж яких можна вибрати слово відповідно до контексту;
– детектування абревіатур та заголовків;
– детектування окремих лінгвістичних одиниць;
– морфологічна декомпозиція — перетворення окремих термінів у зрозумілу форму [7].

На сьогодні відомі методи розпізнавання мови мають низку загальних властивостей, зокрема:

1) для розпізнавання використовується метод порівняння з еталонами;

2) сигнал може бути представлений у вигляді безперервної функції або ж у вигляді слова в певному кінцевому алфавіті;

3) для скорочення обсягу обчислень використовуються методи динамічного програмування.

Наявні методи вирішення основних завдань з розпізнавання мовлення поділяються на дві великі групи: непараметричні і параметричні [2].

Непараметричні методи використовують міру близькості до еталонів на множині мовних сигналів (на основі формальних граматик чи метрик). Перевагами цих методів є простота реалізації та навчання. До недоліків належить складність обчислення міри близькості, яка пропорційна квадрату довжини сигналу та великий обсяг пам'яті, необхідний для зберігання еталонів команд — пропорційний довжині сигналу і кількості команд в словнику.

Параметричні методи застосовують теорію прихованих моделей Маркова — подвійні стохастичні процеси і ланцюги Маркова. Перевагами методу прихованих моделей Маркова є швидкий спосіб обчислення значень функції відстані (ймовірності) та істотно менший об'єм пам'яті. Основними недоліками є велика складність його реалізації та необхідність використання великих фонетично збалансованих мовних корпусів для навчання параметрів.

Загалом для розпізнавання мовлення використовуються такі методи: статистичний, символічний, коннективістський, допоміжних векторів, прихована Марковська модель, Умовні випадкові поля, N-грамні моделі.

Розглянемо основні методи, які застосовують для розпізнавання мовлення [11].

Статистичний метод. В основі статистичного методу обробки природної мови лежить припущення, що зміст тексту може бути визначено за найуживанішими словами. Основним завданням цього підходу є визначення кількості повторень конкретного слова в тексті. Латентно-семантичний підхід є різновидом статистичного методу та базується на ідеї, що сукупність усіх контекстів, у яких зустрічається або не зустрічається певне слово, визначає множину взаємних обмежень для виявлення схожостей у значеннях слів. Основна проблема, з якою стикаються статистичні підходи, полягає в розгляді тексту як набору слів тез смислового зв'язку.

Лінгвістичний підхід складається з чотирьох рівнів: графематичного, морфологічного, синтаксичного та семантичного. Перший рівень полягає у виділенні окремих елементів тексту/документа, наприклад розділів, абзаців, речень тощо. Другий рівень полягає у визначенні морфологічних характеристик окремого слова. Третій рівень відповідає за визначення синтаксичної залежності слів у реченнях. Останній рівень пов'язаний зі смисловим розумінням тексту, що передбачає розробки у сфері штучного інтелекту. Дослідницькі досягнення у цій сфері є дуже обмеженими через складність людської мови [6].

Символічний метод здійснює глибинний аналіз лінгвістичних явищ та базується на явному представленні знань, що здійснюється шляхом використання добре досліджених схем представлення знань та алгоритмів, що працюють з ними. Джерелом знання про мову можуть бути словники, формули та правила, розроблені людьми.

Коннективістський метод відповідає за обробку загальних моделей з використанням конкретних прикладів мовних явищ. Найбільш значуща відмінність коннективістського підходу від інших статистичних методів полягає у поєднанні статистичних знань та різних теорій уявлень, що дають змогу працювати з логічними висновками та трансформацією логічних формул.

Метод допоміжних векторів — це диференційний метод машинного навчання, що допомагає провести класифікацію слів за категоріями. Цей метод побудований на певній множині властивостей.

Прихована марковська модель — це графічна система, у якій кожна вершина представляє собою випадкову змінну, що може набувати будь-якого значення (з певними ймовірностями) між декількома станами, породжуючи при цьому один з декількох можливих вихідних символів з кожним переходом. Множина всіх можливих станів та унікальних символів може бути великою. Ми можемо бачити вихідні дані, проте початкові стани системи є прихованими.

Умовні випадкові поля — роздільна (диференційна) модель, яка формує логістичну регресію для послідовності даних. Використовується для передбачення стану змінної, що базується на спостереженій змінній.

N-грамні моделі побудовані на послідовності з п елементів: речень, слів, букв, звуків тощо. Модель дає змогу розрахувати ймовірність появи будь-якого елемента за відомих ймовірностей появи таких попередніх елементів. Така модель зводиться до скінченної множини ймовірностей, кожну з яких може бути оцінено після обчислення повторюваності відповідних n-грам [8].

На сьогодні розпізнавання мови зводиться до вирішення трьох типів задач:

1. Розпізнавання окремо вимовлених слів.
2. Розпізнавання злитої мови.
3. Ідентифікація за зразком мови.

Розпізнавання окремих слів більшою мірою використовується для мовного управління обчислювальною машиною. Метою ж розпізнавання злитої мови є перетворення в текст звичайної мови людини [9]. Механізм розпізнавання для перших двох типів. Для цих двох типів задач механізм розпізнавання мови буде виглядати так. Тут можна виділити 4 основних модулі: модуль збору даних; екстрактор; компаратор; інтерпретатор. Модуль збору даних передбачає отримання вхідного сигналу і його попередню обробку, яка може містити автоматичний регулятор посилення, приглушення еха, виявлення присутності або відсутності мови і виявлення інтонаційного кінця фрази. Цей модуль також містить виділення відрізка мови із вхідного сигналу. Існує декілька алгоритмів визначення початку і кінця мови. В одному із них визначається деякий граничний рівень сигналу. Початкова точка мови в цьому випадку відповідає моменту, коли вхідний сигнал починає перевищувати граничний рівень, а кінцева точка — моменту, де амплітуда вхідного сигналу менша граничної. Другий метод використовує нормалізацію амплітуди вхідного сигналу відповідно з мінімальною амплітудою. Отримані нормалізовані значення зрівнюються з граничним значенням. Екстрактор виконує частотний аналіз сигналу. Акустично-фонетичний потік даних розбивається на короткі кадри, або вектори, тривалістю зазвичай приблизно 10 мс. Здебільшого для кожного кадру визначається низка параметрів, використовуючи швидке перетворення Фур'є. Крім того, можна ще використовувати й інші характеристики, наприклад спектральні. Компаратор здійснює акустичні порівняння: кожен кадр, або вектор, порівнюється з акустично-фонетичними зразками, які зберігаються в спеціальній базі даних. Водночас можуть порівнюватись як окремі фонеми, так і слова і навіть фрази. За невеликої кількості слів, використовуваних диктором, більш високу надійність і швидкість можна очікувати від розпізнавання цілих слів, але при збільшенні словника швидкість різко падає, і оптимальним стає розпізнавання окремих фонем [7].

Загалом використовується три алгоритми для розпізнавання кадрів: алгоритм динамічної трансформації шкали часу; приховане Марковське моделювання; нейронна мережа з часовою затримкою. Алгоритм динамічної трансформації шкали часу використовує оптимізаційний принцип для мінімізації кількості помилок, виникаючих під час порівняння розпізнаваного слова з еталонною моделлю. Приховане Марковське моделювання використовує ймовірнісні моделі слів. При використанні цієї технології для кожного можливого варіанта слова, яке

розпізнається, вичислюється ймовірність, потім отримані ймовірності порівнюються і вибирається слово з найбільшою ймовірністю. Нейронна мережа с часовою затримкою у випадку розпізнавання обмеженої кількості слів дає кращі результати, ніж метод прихованого Марковського моделювання. Один із методів, оснований на порівнянні фонем, використовує поняття «контекстна фонема». В цьому методі фонема розглядається в поєднанні з попередньою і наступною фонемою. Далі в процесі розпізнавання визначається фонема, яка найбільше близько відповідає тій, яка розпізнається. Інтерпретатор вирішує задачу динамічного програмування з метою знайти найкраще розбиття отриманого від компаратора алфавітного потоку на слова і фрази. Залежно від об'єму використовуваного словника і чинних синтаксичних правил застосовуються різні стратегії пошуку і відсіювання. В цьому блоці із розпізнаних фонем формуються слова, а із слів фрази. При цьому також часто використовується ймовірна система порівняння результатів. Ідентифікація за зразком мови використовується для досягнення забезпечення безпеки. Вона складається із трьох стадій: реєстрація; тестування; допуск. Механізм розпізнавання для третього типу. У процесі реєстрації користувача запам'ятовуються особливості його голосу і формується так звана мовна модель [9]. Під час тестування виконується порівняння запропонованого зразка мови із запам'ятованою мовною моделлю користувача, а також з моделлю «самозванця», складеною на базі голосів інших людей. Якщо результат порівняння виявиться позитивним для першого випадку і негативним для другого, можна вважати, що тестування пройшло успішно. Ідентифікацію по голосу можна використовувати і в поєднанні з іншими засобами забезпечення безпеки [7].

Системи розпізнавання мови можуть також поділятися на:

– диктороорієнтовані;
– дикторонезалежні.

Системи першого типу потребують наявності етапу «навчання», тобто налаштування системи на конкретного користувача, якому необхідно промовити визначений набір слів для того, щоб еталонні моделі його вимови були занесені у базу даних. Згодом під час розпізнавання мови цього користувача система опирається на еталонні моделі, які зберігаються у базі даних. У випадку виконання системи іншим користувачем необхідне повторне навчання [1].

Крім того, системи розпізнавання мови поділяються на:

– системи автоматичного розпізнавання ізольованих слів для розпізнавання вимовлених людиною команд послівно;
– системи автоматичного розпізнавання неперервного мовлення — з можливістю виділяти слова в природному неперервному потоці людської мови;
– системи розуміння мови — з елементами інтелекту, що дає змогу, по-перше, на основі змістовного аналізу більш правильно виділяти слова в потоці мови і, по-друге, зберігати інформацію в базі знань, звідки її можна витягнути для вирішення певних інтелектуальних завдань.

Основні компоненти систем розпізнавання мови:

– Графічне середовище для розробки, компіляції та оптимізації граматичних і лексичних блоків розпізнавання, перевірки і редагування лексиконів.

– Система для протоколювання діалогів з працюючою програмою з метою оцінки якості розпізнавання і налаштування системи.
– Інструмент оцінки якості роботи системи для перевірки відповідності слова, сказаного абонентом до використаної граматики.
– Система для створення «тренованих» мовних моделей, що підвищують продуктивність і пришвидшують процес розпізнавання.
– Система для розподілу багатьох паралельних запитів різних типів і прозорою інтеграцією різних мовних модулів в мережі [3].

Висновки. Подано аналіз наявних методів розпізнавання мови людини. Розглянуто основні типи задач, які не можуть існувати без систем розпізнавання. Теоретично проаналізовано моделі і методи аналізу та розпізнавання сигналів багатьох змінних. На сьогодні існує багато методів вирішення цих проблем, але жоден метод не є ідеальним, їхня точність не перевищує 85 %. Наша мета — досягти результатів, які максимально будуть наближені до ідеальних.

Можна виділити основні проблеми розробників програмного забезпечення для розпізнавання усного мовлення:
– Пересічний користувач зі спонтанною мовою.
– Наявність змінних акустичних чи мовних перешкод і спотворень.
– Необхідність попереднього налаштування системи на голос від кількох десятків хвилин до кількох годин попереднього наговорювання текстів.
– Недостатня інвестиційна база, що не дає змоги інтенсивно проводити дослідження і розробляти нові інноваційні алгоритми в мовних технологіях.

Загалом для підвищення якості мовного введення тексту на комп'ютері рекомендується сервіс або програма для обробки мови. Перетворення її в текстовий вигляд буде працювати краще, якщо забезпечити для цього всі умови, адже якість написання безпосередньо залежить від правильно налаштованого мікрофона, дикції користувача, відсутності додаткового шумового супроводу. Не варто сподіватися, що розпізнавач голосу буде коректно працювати, якщо є явні мовні дефекти. Щоб знизити кількість помилок і менше часу присвячувати коригуванню тексту, потрібно дотримуватися таких умов:
– Для коректного перетворення мови необхідна чиста вимова і відсутність сторонніх звуків. Якщо максимально чітко вимовляти слова з розстановкою розділових знаків, правити текст не доведеться надто довго.
– Перед виконанням робіт необхідно налаштувати мікрофон. Якщо немає можливості прибрати сторонні шуми, знизити його чутливість і вимовляти слова голосніше і чіткіше.
– Не потрібно вимовляти занадто довгі фрази, приправлені безліччю складних синтаксичних конструкцій.

Дотримуючись цих рекомендацій та диктувати правильно, програма буде писати текст з мінімальним вмістом помилок, що сприятливо позначиться на продуктивності праці. При цьому розглядати мовне введення як стовідсоткову альтернативу клавіатурного набору поки не доводиться, однозначно буде потрібне коригування, але багатьом користувачам така можливість полегшує повсякденні завдання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Іванов О. В. Класичний контент-аналіз та аналіз тексту: термінологічні та методологічні відмінності. *Вісник Харківського національного університету імені В. Н. Каразіна*. 2013. № 1045. С. 72.
2. Karen S. Jones. Natural language processing: a historical review. Cambridge: Computer Laboratory, University of Cambridge, 2001. P. 2.
3. Анисимов А. В., Марченко А. А. Система обработки текстов на естественном языке. *«Штучний інтелект»*. 2002. № 4. С. 157.
4. Слюсар В. И. Нейромережні технології та їх застосування НМТіЗ-2020 : збірник наукових праць XIX Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2020». Краматорськ : Донбаська державна машинобудівна академія. 2020. С. 156–162.
5. Диковицкий В. В., Шишаев М. Г. Обработка текстов естественного языка в моделях поисковых систем. *Сборник научных трудов*. 2010. С. 30.
6. Olaronke G. Iroju, Janet O. Olaleke. A Systematic Review in Natural Language Processing in Healthcare. *I.J. Information Technology and Computer Science*. 2015. #8. P. 45.
7. Оппенгейн А. В., Шафер Р. В. Цифровая обработка сигналов. Москва : Радио и связь, 1979. 347 с.
8. Ялковський А. Є. Проблеми інформатизації та управління, 3(27)-2009. 164-168 ПРОБЛЕМИ РОЗПІЗНАВАННЯ МОВИ ЛЮДИНИ.
9. Liddy E. D. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc. 2001. P.10.
10. Бабенко Т. В., Сушко С. О. Про ентропію української мови. *Науково-практичний журнал «Захист інформації»*. 2012. № 3. С. 105.
11. Кузнецов В., Отт А. Автоматический синтез речи. Таллин : Валгус, 1989. 135 с.

REFERENCES

1. Ivanov, O. V. (2013). Klasychnyi kontent-analiz ta analiz tekstu: terminolohichni ta metodolohichni vidminnosti: Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina, 1045, 72 (in Ukrainian).
2. Karen S. Jones. (2001). Natural language processing: a historical review. Cambridge: Computer Laboratory, University of Cambridge, 2 (in English).
3. Anisimov, A. V., & Marchenko, A. A. (2002). Sistema obrabotki tekstov na estestvennom jazyke: «Shtuchnij intelekt», 4, 157 (in Russian).
4. Sliusar, V. Y. (2020). Neiromerezhni tekhnolohii ta yikh zastosuvannia NMTiZ-2020 : zbirnyk naukovykh prats XIX Mizhnarodnoi naukovoi konferentsii «Neiromerezhni tekhnolohii ta yikh zastosuvannia NMTiZ-2020». Kramatorsk : Donbaska derzhavna mashynobudivna akademiia, 156–162 (in Ukrainian).
5. Dikovickij, V. V., & Shishaev, M. G. (2010). Obrabotka tekstov estestvennogo jazyka v modeljah poiskovyh sistem: Sbornik nauchnyh trudov, 30 (in Russian).
6. Olaronke G. Iroju, & Janet O. Olaleke. (2015). A Systematic Review in Natural Language Processing in Healthcare: I.J. Information Technology and Computer Science, 8, 45 (in English).
7. Oppengejn, A. V., & Shafer, R. V. (1979). Cifrovaja obrabotka signalov. Moskva : Radio i svjaz', 347 s.

8. Yalkovskyi, A. Ye. Problemy informatyzatsii ta upravlinnia, 3(27)-2009 164-168. PROBLEMY ROZPIZNAVANNIa MOVY LIuDYNY (in Ukrainian).
9. Liddy, E. D. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc. 2001, 10 (in English).
10. Babenko, T. V., & Sushko, S. O. (2012). Pro entropiiu ukrainskoi movy: Naukovo-praktychnyi zhurnal «Zakhyst informatsii», 3, 105 (in Ukrainian).
11. Kuznecov, V., & Ott, A. (1989). Avtomaticheskij sintez rechi. Tallin : Valgus (in Russian).

doi: 10.32403/2411-3611-2021-2-40-16-27

ANALYTICAL RESEARCH OF SPEECH RECOGNITION TECHNOLOGIES

Z. Selmenska, M. Dubnevich, Z. Plakhtyna, A. Tsebrik

Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine
zorselm@gmail.com

The problem of speech recognition is nowadays considered very serious and plays an important role in human-machine communication. Controlling objects with the help of a language would open up wide prospects for automation in many fields of human activity, opening up the possibility of communicating with machines, especially for users of personal computers who do not know the programming language. Everyday life is so fast-paced that sometimes there is not even enough time to do the usual things on the computer in the usual way – with a keyboard and a mouse. New technology has come to the forefront, allowing you to operate your personal computer using your voice. Speech recognition software is developed to provide a quick way to input information into a computer and can also help people with physical disabilities. Computer-based language recognition systems are gradually finding applications in the scientific area and in the domestic sphere. Modern voice input technology provides users with many possibilities to optimize and facilitate their work and save time.

The application of speech recognition systems for the input of textual information into computer publishing systems is considered. The paper also conducts analytical research on speech recognition technologies and software that allows voice input of the information by personal computers. An analysis of existing methods of human speech recognition is presented. The main types of tasks which cannot exist without recognition systems are considered. The basic theoretical positions of speech recognition systems are analysed, the problems that arise when using these systems in the software and ways to solve them are highlighted.

Keywords: *speech recognition, deep learning, voice control.*

Стаття надійшла до редакції 27.09.2021.
Received 27.09.2021.