

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ МОВИ В РОЗРОБЦІ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Агарков М.А.

Науковий керівник – к.т.н., доц. Сумцов Д.В.

Харківський національний університет радіоелектроніки
(61174, Харків, просп. Людвіга Свободи, 51б, тел. (063) 642-61-31)

e-mail: maksym.aharkov@nure.ua

The given work is devoted to modern developments in the use of speech recognition technologies in software development, namely, automating the process of writing code with the help of the human speech apparatus of the so-called «coding by voice». It discusses the main tools for automatic recognition of speech information and analysis of further work with them. This work also discusses ready-made software products and solutions that can later serve as a support for developing own product for writing code using an HTML voice and embedding it in a text editor or HTML compiler.

В основі будь-якої мовної технології лежить так званий «engine», або ядро програми – набір даних та правил, за якими здійснюється обробка даних. Залежно від призначення цього ядра розрізняють TTS та ASR engine. TTS (Text-To-Speech) engine надає можливість синтезу мови за текстом, а ASR (Automatic Speech Recognition) engine – для розпізнавання мови. Існує кілька великих виробників, які займаються створенням ASR ядер.

Компанія SPIRIT займається створенням програмних засобів для цифрової телефонії, ущільнення мови, ідентифікації мовця для технологій VoIP та GPS. ASR engine від SPIRIT розроблений для розпізнавання мовних команд і застосовується в різних застосунках, таких як голосове управління пристроями, голосовий набір в hands-free пристроях, введення персональних ідентифікаційних кодів (PIN) в системах безпеки. Дане ядро вбудовується в будь-які DSP або RISC платформи і поставляється у вигляді об'єктного коду.

Sakrament ASR Engine – програмна розробка білоруської компанії «Сакрамент», розрахована на застосування в різних апаратних системах і програмних застосунках, що використовують технології розпізнавання мови. Заявлені характеристики:

- точність розпізнавання 95-98%;
- незалежність від диктора;
- незалежність від мови.

Однак, дана система не має можливості навчання – додаткові словники створюються за замовленням самою компанією «Сакрамент».

Sphinx – відкритий програмний продукт для розпізнавання мови. Розробка ведеться в університеті Карнегі-Меллона, продукт поширюється на умовах ліцензії Berkley Software Distribution (BSD) і доступний як для комерційного, так і для некомерційного використання. Його особливості:

- незалежність від диктора;
- розпізнавання безперервної мови;

Dragon NaturallySpeaking Preferred фірми Dragon Systems – єдина програма, яка наблизилася до того, щоб відповідати заявленим характеристикам. В цілому пакет дуже близько підходить до досягнення заявленої безпомилковості розпізнавання – 95%. Хоча пакет Dragon і поступається деяким з конкурентів у тому, що стосується переміщення по екрану, правки й форматування, він перевершує всіх у головному – здатності з першого разу правильно записувати вимовлені слова.

Прикладне програмне забезпечення розпізнавання мови в основному сконцентровано на диктуванні, дозволяючи друкувати документи, електронні листи, допомагаючи фахівцям, чия робота пов'язана з тривалими періодами друку, наприклад журналістам, програмістам, письменникам та вченим.

Розробляються спеціалізовані модулі, які можна втілювати в різні текстові редактори:

- VoiceCode – дозволяє набирати чистий програмний код за допомогою голосових команд, не торкаючись клавіатури. VoiceCode дозволяє диктувати код природним чином, при цьому автоматично перетворює людську мову в специфічні програмні функції. Програма працює лише 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 й нормалізації тексту. Модуль можна використовувати для реалізації інших мовних інтерфейсів.

- Voice Grip – додатковий макрос для редактора Emacs, створений з метою спрощення розпізнавання мови для програмістів.

- Java by voice – серія макросів для редактора Emacs, спроектовані для спрощеного введення коду мовою Java.

- Cache Pad – макрос для редактора Emacs для кешування недавно продиктованих імен функцій та змінних.

Література

1. Кузнецов В., Отт А. Автоматичний синтез мови. - Таллінн: Валгус, 1989. - 135 с.
2. Чекмарьов А. Мовні технології - проблеми та перспективи.// Компьютерра, № 49 с. 26-43, 1997 р.
3. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения. [Электронный ресурс] — 2003. - Режим доступа: <http://www.frolov-lib.ru/books/hi/index.html>