

System automatycznego rozpoznawania mowy i jego możliwe zastosowania w nowoczesnej kancelarii prawniczej

W związku z badaniami nad przetwarzaniem mowy oraz języka naturalnego prowadzonymi przez Laboratorium Zaawansowanych Technologii Języka i Mowy w ścisłej współpracy z Wolters Kluwers Polska oraz Polską Platformą Bezpieczeństwa Publicznego mam przyjemność zasygnalizować Państwu ukazanie się w najbliższych dwóch latach systemu rozpoznawania mowy polskiej w zakresie dyktowania pism prawniczych, w tym procesowych. Zwracam się jednocześnie z uprzejmą prośbą o wzięcie udziału w nagraniach wypowiedzi specjalistycznych na potrzeby projektowanego systemu rozpoznawania mowy.

Narzędzia wykorzystujące techniki rozpoznawania mowy - stan obecny

Dyktowanie treści wypowiedzi i jej automatyczny zapis jest niewątpliwie najbardziej efektywnym i elastycznym przekazem informacji. Przy dzisiejszym poziomie techniki bezpośrednie dyktowanie komputerowi tekstów swobodną mową, wprowadzanie ciągów poleceń za pomocą głosu, czy też dostęp do informacji za pomocą automatycznych systemów telefonicznych staje się coraz bardziej powszechne. Można obecnie wskazać pewną liczbę profesjonalnych systemów rozpoznawania mowy funkcjonujących dla różnych języków, np. wdrożony do transkrypcji tekstów prawniczych system *Dragon Naturally Speaking Legal 9* dla języka angielskiego. System ten przetwarza tekst na mowę z szybkością 160 wyrazów na minutę, z poprawnością rzędu 90-95% i zapisuje wynik w dowolnym edytorze tekstowym. Dodatkowo automatycznie formatuje formuły prawnicze i określone wzorce dokumentów. System ten nie jest dostępny dla języka polskiego z uwagi na zbyt wysokie koszty implementacji. Dla języka polskiego podobny system nie został do tej pory opracowany, co wynika między innymi z faktu, że język polski jest pod wieloma względami trudniejszy w komputerowej obróbce, niż język angielski.

Perspektywa rozwoju systemu dyktowania dla języka polskiego

W ramach projektu: *Zintegrowany system automatycznej konwersji mowy polskiej na tekst oparty na modelu językowym stworzonym w środowisku analizy i obiegu dokumentów prawniczych w zakresie bezpieczeństwa wewnętrznego*, podjęliśmy wyzwanie stworzenia systemu dyktowania mowy dla języka polskiego z silną determinacją i zaangażowaniem w kierunku rozwiązania istniejących problemów i pozytywnego zakończenia prowadzonych badań. Można zatem z pewnym optymizmem sądzić, iż system do automatycznego zapisywania (tekstowego) dyktowanej mowy polskiej powinien się pojawić na rynku w perspektywie najbliższych kilku lat. Znaczącym atutem jest wykorzystanie w projekcie oprogramowania Systemu Informacji Prawnej LEX Wydawnictwa Wolters Kluwer Polska - przyczyni się ono do dodatkowego wyspecjalizowania tworzonego systemu dla sądów, kancelarii i ewentualnie biur podatkowych.

W harmonogramie projektu przewidziano opracowanie automatycznego systemu rozpoznawania mowy głównie dla dedykowanej grupy użytkowników - na potrzeby kancelarii prawnych, sądownictwa oraz organów dochodzeniowo-śledczych.

Kształt tworzonej bazy słownej na potrzeby systemu Automatycznego Rozpoznawania Mowy (ARM) oraz jej reprezentatywność w sposób zasadniczy wpłynie na jakość dyktowania. **Konieczny więc będzie dźwiękowy zapis mowy pochodzący od kilkudziesięciu specjalistów, potencjalnych użytkowników systemu - czyli prawników w kancelariach oraz w sądach. W celu osiągnięcia reprezentatywności bazy w zakresie dialektów, konieczne będą również nagrania w innych miastach polskich.**

Udział użytkowników systemu w procesie badawczym

Zwracamy się do Państwa z uprzejmą prośbą o wzięcie udziału w nagraniach wypowiedzi na potrzeby bazy danych tworzonej w ww. projekcie. Państwa udział będzie stanowił ważny wkład w rozwój projektowanego systemu. Poniżej zamieszczono krótki opis typowej sesji nagraniowej. W razie jakichkolwiek wątpliwości chętnie udzielimy szerszych wyjaśnień.

- **Czas trwania sesji nagraniowej:** nagranie głosu jednej osoby, to maksymalnie 1 godzina, do sesji należy doliczyć pół godziny dla osoby nagrywającej na czas przygotowania stanowiska nagraniowego
- **Warunki lokalowe:** ciche pomieszczenie biurowe, najlepiej jeśli jest wyposażone w wykładzinę, meble, dobrze jeśli ściany i okna są przynajmniej w pewnym stopniu wytłumione (szafy, półki, firanki).
- **Narzędzia:** są zapewniane przez nas - komputer przenośny oraz mikrofony przywozi ze sobą osoba nagrywająca
- **Opis typowej sesji nagraniowej:** Sesja nagraniowa dla jednej osoby trwa przeciętnie około 45 minut, godziny.
 - Osoby nagrywane są proszone o odczytanie około trzydziestu zdań lub fraz o tematyce prawniczej lub sądowniczej oraz listy około dziesięciu krótkich fraz: skrótów, skrótowców lub terminów specjalistycznych.
 - Dodatkowo mówca zostanie poproszony o podyktowanie krótkiego tekstu, przygotowanego wcześniej przez siebie. Tekst powinien mieć około 20-30 zdań i być typowym tekstem, który mógłby być dyktowany w codziennej pracy kancelarii prawniczej.
 - Mówcy odczytują teksty z monitora komputera przenośnego. Po odczytaniu wszystkich zdań (30 zdań, 10 fraz) z puli mowy czytanej, na monitorze zostanie wyświetlony komunikat: „A teraz proszę podyktować uprzednio przygotowany tekst”, po którym mówca powinien powoli i starannie podyktować własną wypowiedź.
 - Nagranie rejestrowane jest za pomocą dwóch mikrofonów: małego mikrofonu nagłownego oraz mikrofonu stojącego na stole w odległości około pół metra od mówcy.
 - Osoba przeprowadzająca nagranie odpowie na wszelkie wątpliwości osoby nagrywanej, odnośnie sposobu czytania, odpowiedzi na pytania itp. oraz w razie pilnej potrzeby przerwie na krótki czas nagranie.
- Za udział w nagraniu każdy prawnik otrzyma „**podarunek**” dyplom, oraz **dwa bony** do naszej księgarni, każdy o wartości 75 zł (bony ze względów podatkowych otrzyma w dwóch ratach - jeden przy nagraniu, drugi pocztą wraz z dyplomem)
- Nagrany tekst podlega analizie fonetycznej oraz językowej (statystyka występowania słów), w żaden sposób nie jest wykorzystywany „zewnętrznie”

Prof. dr hab. inż. Stefan Grocholewski

Kierownik Projektu: Zintegrowany system automatycznej konwersji mowy polskiej na tekst oparty na modelu językowym stworzonym w środowisku analizy i obiegu dokumentów prawniczych w zakresie bezpieczeństwa wewnętrznego

Poznańskie Centrum Superkomputerowo-Sieciowe

Laboratorium Zaawansowanych Technologii Języka i Mowy

stefan.grocholewski@cs.put.poznan.pl, <http://www.speechlabs.pl> tel. (061) 6682151