**Projektowanie systemów z dostępem w języku naturalnym**

**Badanie dostępnych narzędzi służących do przetwarzania mowy**

Dawid Mikowski | Paweł Kolak

Spis treści

[**1.** **Wstęp** 3](#_Toc23932868)

[**2.** **Cel i opis projektu** 3](#_Toc23932869)

[**3.** **Opis wybranych narzędzi** 4](#_Toc23932870)

[**3.1.** **https://dictation.io/speech** 4](#_Toc23932871)

[**3.2.** **https://speechnotes.co/** 5](#_Toc23932872)

[**3.3.** **Microsoft Azure** 5](#_Toc23932873)

[**4.** **Badania** 6](#_Toc23932874)

[**4.1.** **Proste pojedyncze słowa** 6](#_Toc23932875)

[**4.2.** **Proste kilkuwyrazowe zdania** 7](#_Toc23932876)

[**4.3.** **Zaawansowane lingwistycznie wyrazy** 7](#_Toc23932877)

[**4.4.** **Krótkie fragmenty literatury** 8](#_Toc23932878)

[**5.** **Wnioski** 8](#_Toc23932879)

1. **Wstęp**

Proces automatycznego rozpoznawania tekstu w dzisiejszym świecie zyskuje na coraz większej popularności, a jego zastosowanie znajduje swój udział w coraz większej liczbie systemów. Istotny wzrost polarności zyskują szczególnie narzędzia i systemy, które pozwalają na komunikację z urządzeniami przy pomocy mowy. W tym celu powstaje coraz więcej narzędzi umożliwiających nie tylko przetwarzanie mowy w celu izolacji wypowiedzianych zdań, ale również poddawane są analizie i przy ich pomocy możliwe jest wprowadzanie poleceń.

Sam proces automatycznego rozpoznawania mowy polega na przyporządkowywaniu sygnału do odpowiednich klas, a w zależności od przeznaczenia proces odbywa się na różnych poziomach (na przykład: alofonów, fonemów, sylab, wyrazów lub nawet całych fraz). Zwykle proces rozpoznawania poprzedzony jest fazą uczenia. W wyniku uczenia powinniśmy otrzymać w przestrzeni parametrów rozłączne obszary odpowiadające poszczególnym klasom rozpoznawanych sygnałów.

Czerpiąc z tych faktów inspirację, zdecydowano się na przeprowadzenie badań nad narzędziami, które umożliwiają przetwarzanie mowy naturalnej na tekst maszynowy. W tym celu sprawdzenia, które z dostępnych na rynku narzędzi najlepiej radzi sobie z rozpoznawaniem mowy oraz przeprowadzono badania sprawdzające ich skuteczność. Znaleziono cztery dostępne narzędzia i zostały przetestowane pod kontem skuteczności zarówno dla mowy w języku angielskim oraz języka polskiego, jeśli narzędzie będzie na to pozwalało.

1. **Cel i opis projektu**

Jako cel projektu zdefiniowano zbadanie skuteczności narzędzi, które umożliwiają przetwarzania mowy na tekst.

Pierwszym etapem, byłoby wybranie systemów, które zostaną poddane analizie.

Wybrane narzędzia powinny spełniać następujące wymagania:

* Być darmowe
* Nie wymagające dodatkowej instalacji (preferowane narzędzia online)
* Proste w obsłudze
* Obsługiwać język polski

Przewidujemy następujące etapy przebiegu projektu:

* Prostych pojedynczych słów
* Prostych kilkuwyrazowych zdań
* Zaawansowanych lingwistycznych wyrazów
* Krótkich fragmentów literatury.

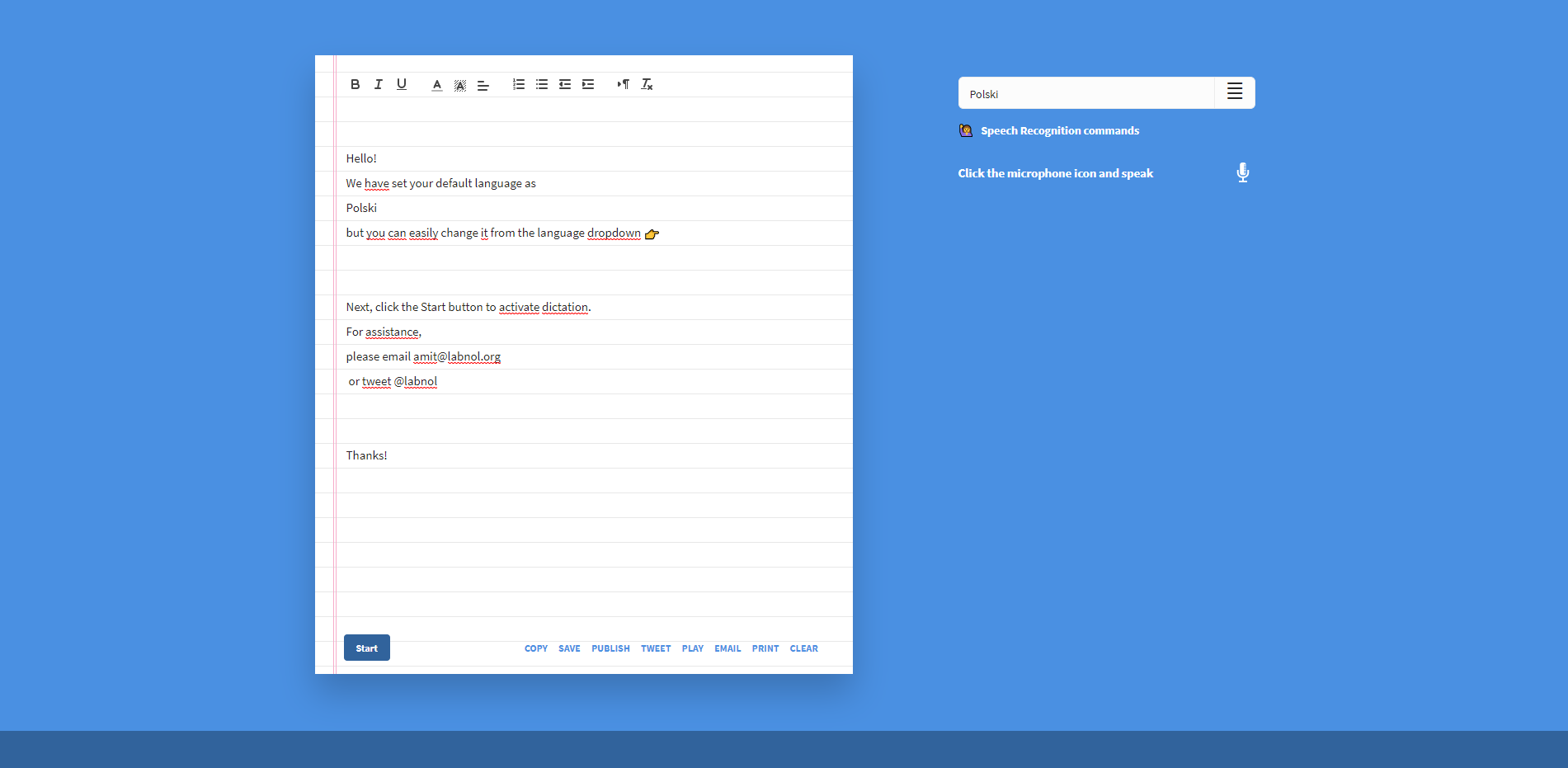
Głównym celem badań jest sprawdzenie jak badane narzędzia radzą sobie z językiem polskim, który ze względu na bardzo skomplikowaną gramatykę i fakt, że należy do języków fleksyjnych sprawia trudności w automatycznym przetwarzaniu mowy.

Możliwymi kandydatami na narzędzie które zostanie wykorzystane do testowania jego  skuteczności:

* https://dictation.io/speech
* https://speechnotes.co/
* Microsoft Azure

Na testowany dźwięk nałożone będą różne stopnie zakłóceń o różnym poziomie natężenia.

1. **Opis wybranych narzędzi**
   1. **https://dictation.io/speech**

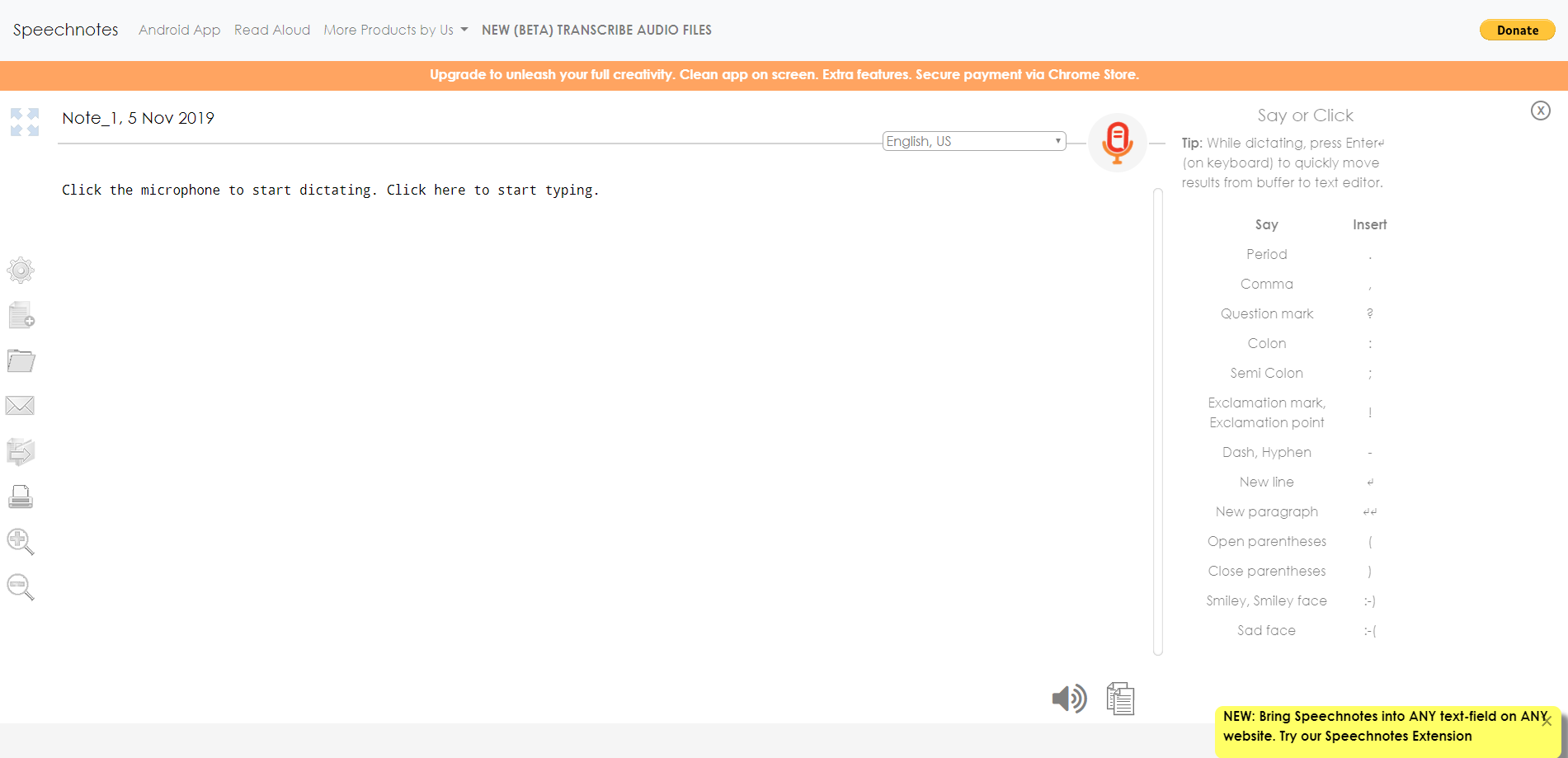
****

Według twórców, Dictation, które wykorzystuje rozpoznawanie mowy, przynosi korzyści uczniom ze specjalnymi potrzebami i trudnościami w uczeniu się, które przeszkadzają im w mowie i pisowni. Ułatwia pisanie osobom cierpiącym na zespół cieśni nadgarstka lub powtarzające się obrażenia (RSI) w dłoniach lub nadgarstkach.

Podobnie jak w przypadku poprzedniego narzędzia, stylistyka Dictation, kieruje się minimalistyką i prostotą. Na stronie nie znajdziemy nic, poza krótką instrukcją, polem do wyboru języka oraz polu, w którym zapisany zostanie wynik. Obsługa tego narzędzia jest najbardziej intuicyjna ze wszystkich.

Dictation, wykorzystuje potężny mechanizm rozpoznawania mowy Google (ten sam co poprzednie narzędzie), który umożliwia transkrypcję głosu na tekst, a wyniki otzymywane są niezwykle szybko. Potrzeban jest jedynie przeglądarka Google Chrome z systemem Windows, Mac OS X lub Linux. Dyktowanie działa również w Chrome na Androida, ale zalecane jest używanie przeglądarki komputerowej, aby zwiększyć wydajność. Narzędzie nie jest kompatybilne z iPhonem ani iPadem.

* 1. **https://speechnotes.co/**

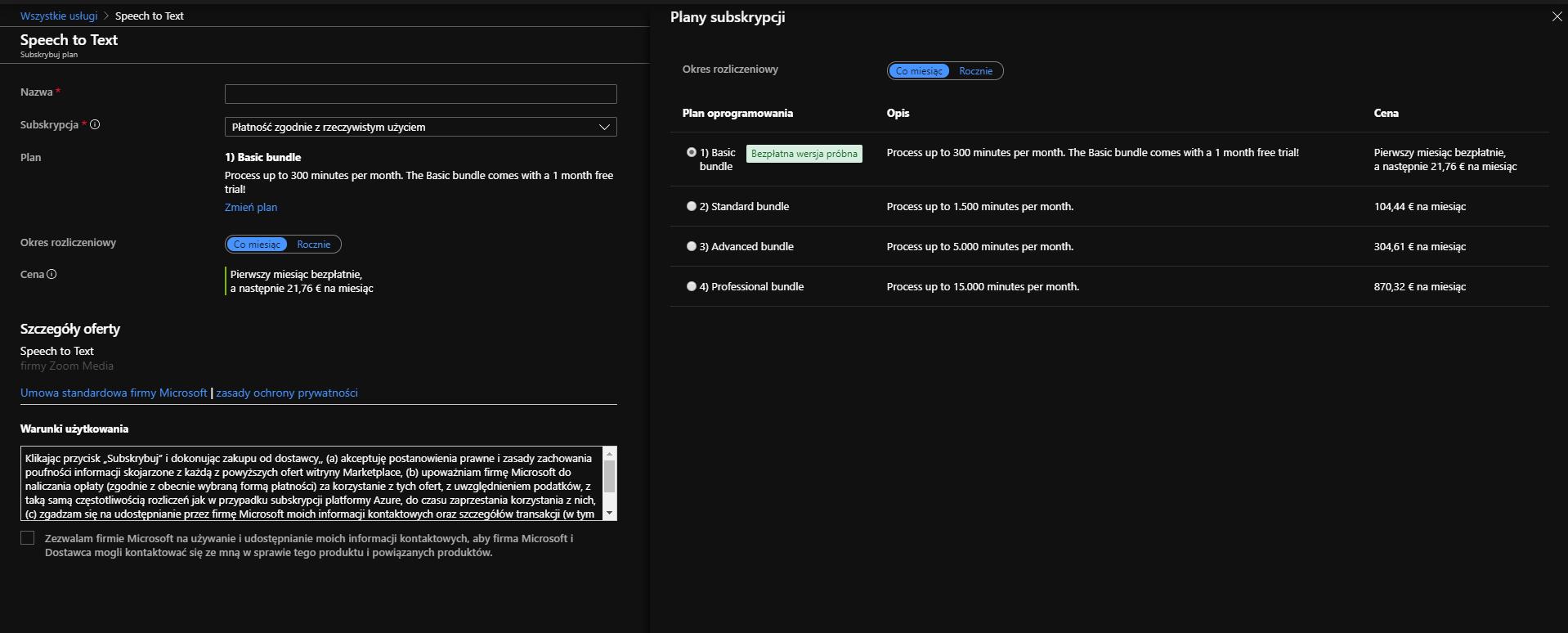


SpeechNotes jest notesem online, z możliwością „notowania” głosem. Powstał po to, aby przyspieszyć zapisywanie notatek. Według twórców, wygodniejszą i szybszą metodą jest przetwarzanie języka na tekst, niż zapisywanie treści, przy pomocy klawiatury. Jest to całkowicie darmowa platforma.

Wygląd narzędzia, jest schludny i minimalistyczny, żeby nie odciągać użytkowników od tego, co jest najważniejsze, od „notowania głosem”. Dostępna jest również wersja mobilna, niemniej jednak, wygląd ona bardzo podobnie i opiera się na dokładnie tym samym mechanizmie przetwarzania mowy.

SpeechNotes opiera się na zaawansowanych technologiach rozpoznawania mowy firmy Google. W rzeczywistości cała mowa jest wysyłana do Google, tam jest interpretowana za pomocą wydajnych serwerów równoległych i algorytmów, i wysyłana z powrotem do SpeechNotes jako strumień możliwych wyników transkrypcji. Dzięki prawidłowej obsłudze tych wyników i zestawu poleceń dla silników „Speech-To-Text” witryna jest w stanie osiągnąć wyniki, które nie tracą na dokładności, nawet w porównaniu z najbardziej profesjonalnym i drogim oprogramowaniem dostępnym na rynku. Mimo to, narzędzie nie rozpoznaje znaków interpunkcyjnych. Oczekiwany poziomu dokładności, powinien wynosić powyżej 90%.

* 1. **Microsoft Azure**

****

1. **Badania**

Badania zostały podzielone na cztery części, które zostały opisane rozdziale 2. W tym miejscu podsumowane zostaną rezultaty badań. Dla języka polskiego w badaniu brały udział następujące narzędzia:

* https://speechnotes.co/
* https://dictation.io/speech
* https://voicenotebook.com/

W celu zapewnienia jak największej stabilności wyników próbki głosu zostały nagrane przy pomocy oprogramowania VB-Audio. Oprogramowanie to umożliwia tworzenie wirtualnych głośników oraz wirtualnych mikrofonów, z których dźwięk można przekierowywać bezpośrednio na inne wirtualne lub fizyczne urządzenie. Próbki dźwiękowe zostały najpierw nagrane a następnie były odtwarzane jako dane wejściowe do programów rozpoznawania mowy. Próbki zostały nagrane dla dwóch głosów męskiego i żeńskiego. Każda próbka w trakcie badania była puszczana 6 razy w celu ustabilizowania wyniku.

* 1. **Proste pojedyncze słowa**

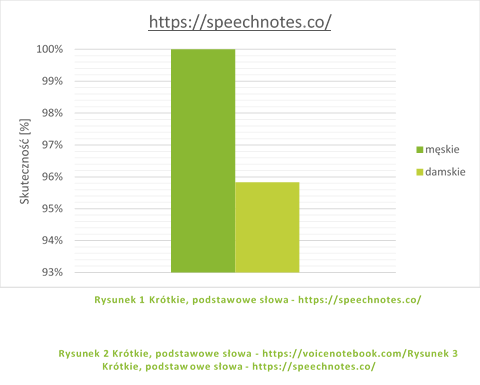
Słowa były wybierane z listy najpopularniejszych słów języka polskiego. Wszystkie wybrane znajdują się w pierwszych 300 pozycjach wspomnianej listy.

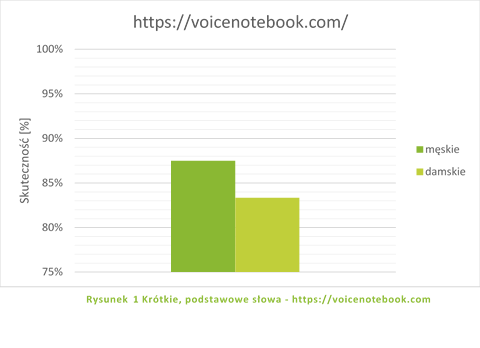
Do badań wybrano następujące słowa:

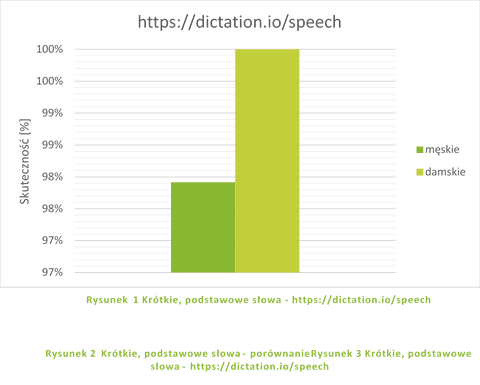
* się
* tak
* ale
* już
* może
* potem
* trzeba
* matka
* szybko
* wszystko
* człowiek
* tymczasem

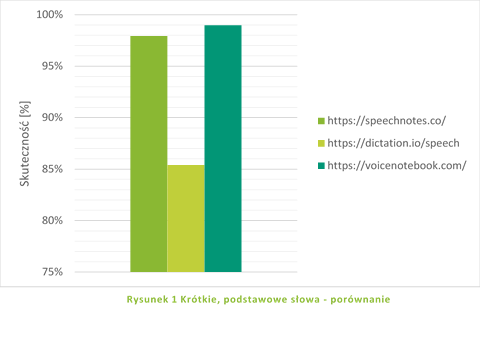
Zgodnie z założeniami każde z trzech wybranych narzędzi zostało poddanie próbie 6‑krotnie, zarówno głos męski i żeński powtórzono trzykrotnie, dla wyeliminowana błędów.

Miarą z jaką mierzono skuteczność narzędzi w tej próbie zdecydowano się, że oceniane będą w sposób zero-jedynkowy, to jest, dla w całości poprawnie przetworzonych słów uznawano, że narzędzie poradziło sobie, w innym wypadku zaliczono je jako niepowodzenie.



****

****

****

* 1. **Proste kilkuwyrazowe zdania**

Do badań wybrano następujące zdania:

* ala ma kota
* lubię placki
* jaka jest pogoda
* ile masz lat
* jak się masz
* muszę powiesić pranie
* Tomek lubi biegać rankiem
* jutro idę na zakupy
* ile kosztuje bilet miesięczny
* Styczeń to najzimniejszy miesiąc
* Jutro rano idę na uczelnię
* Kasia je śniadanie wczesnym rankiem

Każde ze zdań było wypowiadane, w taki sposób, że było słuchać ich odosobnienie. Zgodnie z założeniami każde z trzech wybranych narzędzi zostało poddanie próbie 6‑krotnie, to jest zarówno głos męski i żeński powtórzono trzykrotnie, dla wyeliminowana błędów.

Miarą mierzącą skuteczność wypowiedzianych zdań było to, czy całe zdanie zostało prawidłowo przetłumaczone. Są to krótkie zdania, dlatego zdecydowaliśmy się na takie porównanie. W celu ułatwienia porównania, wszystkie wyrazy w zdaniu składały się z małych liter, a zdania zostały pozbawione znaków interpunkcyjnych, ponieważ, te narzędzia, nie rozpoznają tych znaków.

* 1. **Zaawansowane lingwistycznie wyrazy**

Do badań wybrano następujące słowa:

* egzaltacja
* pauperyzacja
* emancypacja
* abominacja
* abstrahować
* afirmacja
* dekapitacja
* efemeryczny
* egzemplifikacja
* immanentny
* kognitywistka
* konfabulacja

Każde ze słów było kilkukrotnie przećwiczone przez lektorów, aby możliwie dokładnie i płynnie wypowiedzieć skomplikowane słowa. Zgodnie z założeniami każde z trzech wybranych narzędzi zostało poddanie próbie 6 krotnie, to jest zarówno głos męski i żeński powtórzono trzykrotnie, dla wyeliminowana błędów i uśrednienia wyników.

Miarą mierzącą skuteczność wypowiedzianych słów było poprawne ich przetłumaczenie z mowy na tekst. Miara była zero-jedynkowa. Dokładnie taka sama miara, jak w przypadku pojedynczych, normlanych słów.

* 1. **Krótkie fragmenty literatury**

Do ostatniego badania, wybraliśmy dłuższy fragment tekstu z książki „Alicja w krainie czarów” (Załącznik nr1). Jak wcześniej wspomniano, narzędzia, których użyliśmy, nie potrafiły określić interpunkcji, dlatego nie będzie ona brana pod uwagę.

Złożoność tekstu, jego długość oraz fakt, że programy nie potrafiły korzystać ze znaków przystankowych, zadecydowało o wyborze miary poprawności tekstu. Użyliśmy dwóch miar. Pierwszej, która pozwoliła określić ilość słów w tekście, bez względu na to, czy są to słowa poprawne, oraz drugiej. Dla drugiego pomiaru wybraliśmy, trzy zdania z całego tekstu.

Każde zdanie (z trzech wybranych) zostało sprawdzone, pod kątem skuteczności, przetworzenia mowy na tekst.

1. **Wnioski**