**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**

Wraz z rozwojem technologii charakter kontaktu człowieka ze sterowanym procesem technicznym nabiera cech inteligentnego dialogu, nie jest to już tylko prymitywne sterowanie. Także system podaje człowiekowi wysoce przetworzoną i opracowaną informację o swoim stanie. W rezultacie dostępne dotąd środki techniczne jak klawiatury, dźwignie czy pokrętła przestają odpowiadać stawianym przed nimi wymaganiom. Rozwiązaniem tego problemu może stać się sygnał mowy, który to niewątpliwie jest najbardziej naturalnym i najszybszym sposobem komunikacji między ludźmi.

Najpopularniejsze podejście akustyczno-fonetyczne bazuje na teorii akustyki i fonetyki oraz na postulatach zakładających istnienie skończonej liczby jednostek fonetycznych w mowie. Jednostki te można scharakteryzować przy pomocy zestawu cech zawartych w przebiegu czasowym sygnału mowy lub jego spektrum. Pierwszymi etapami rozpoznawania przy tym podejściu są segmentacja oraz klasyfikacja. Polegają one na podziale sygnału na dyskretne fragmenty i sklasyfikowaniu ich jako jednostkę fonetyczną lub grupę jednostek, przydzielając im odpowiednią etykietę lub etykiety. Drugi etap rozpoznawania polega na zdeterminowaniu konkretnego słowa (lub grupy słów) z uzyskanego w pierwszym etapie ciągu etykiet odpowiadającym jednostkom fonetycznym. Słowa te muszą należeć do słownika systemu oraz spełniać określoną gramatykę języka.

Bazując na tych faktach postanowiliśmy przeprowadzić badania efektywności na narzędziach, które umożliwiają przetwarzanie mowy naturalnej na tekst. Wyselekcjonowane zostały dwa narzędzia na których na których zostały przeprowadzane badania poprawności działania.

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.

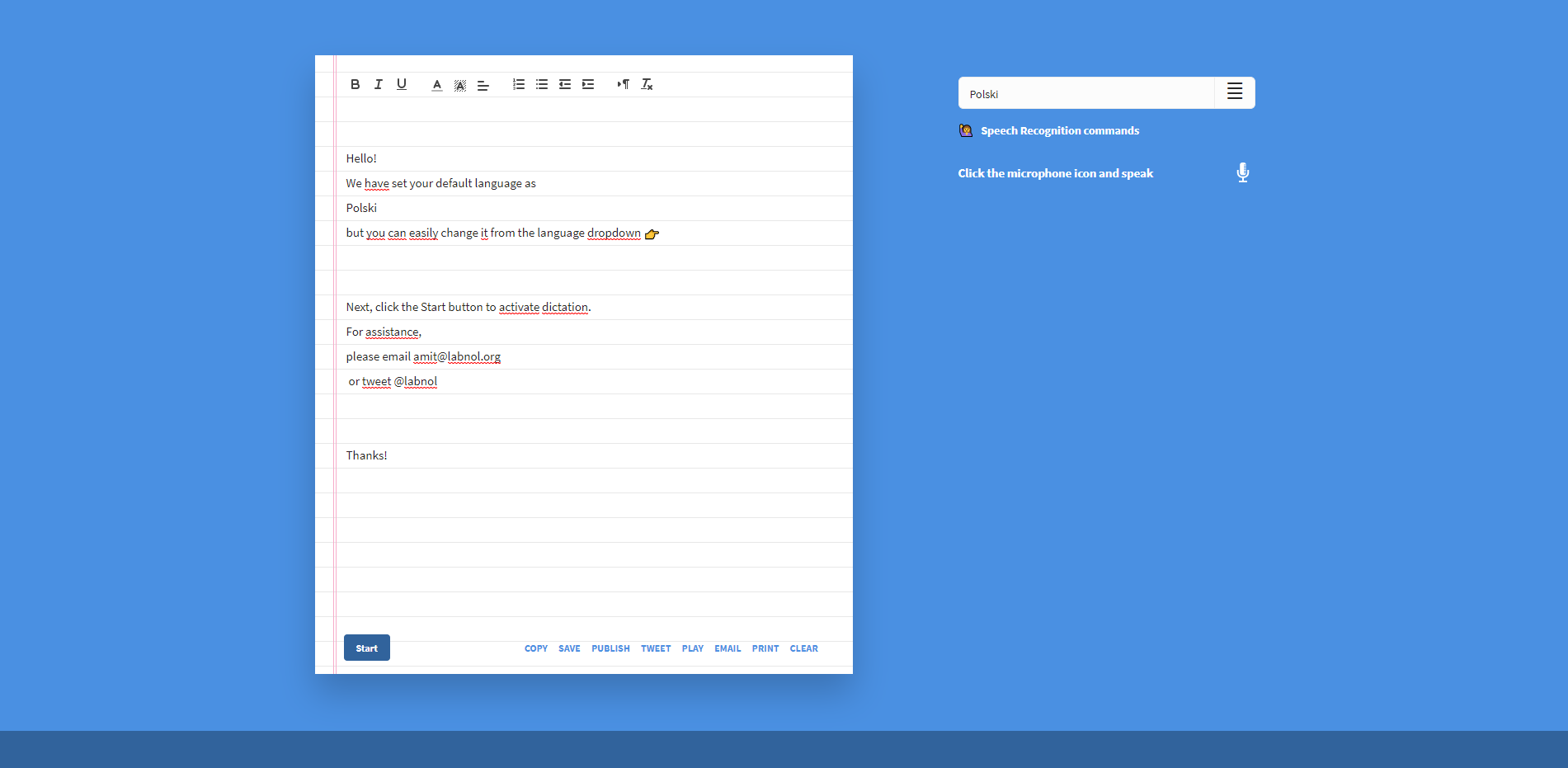
Głównym celem badań jest sprawdzenie jak badane narzędzia radzą sobie z językiem polskim, który ze względu na bardzo skomplikowaną gramatykę i fakt, że należy do języków fleksyjnych sprawia trudności w automatycznym przetwarzaniu mowy.

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

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

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

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

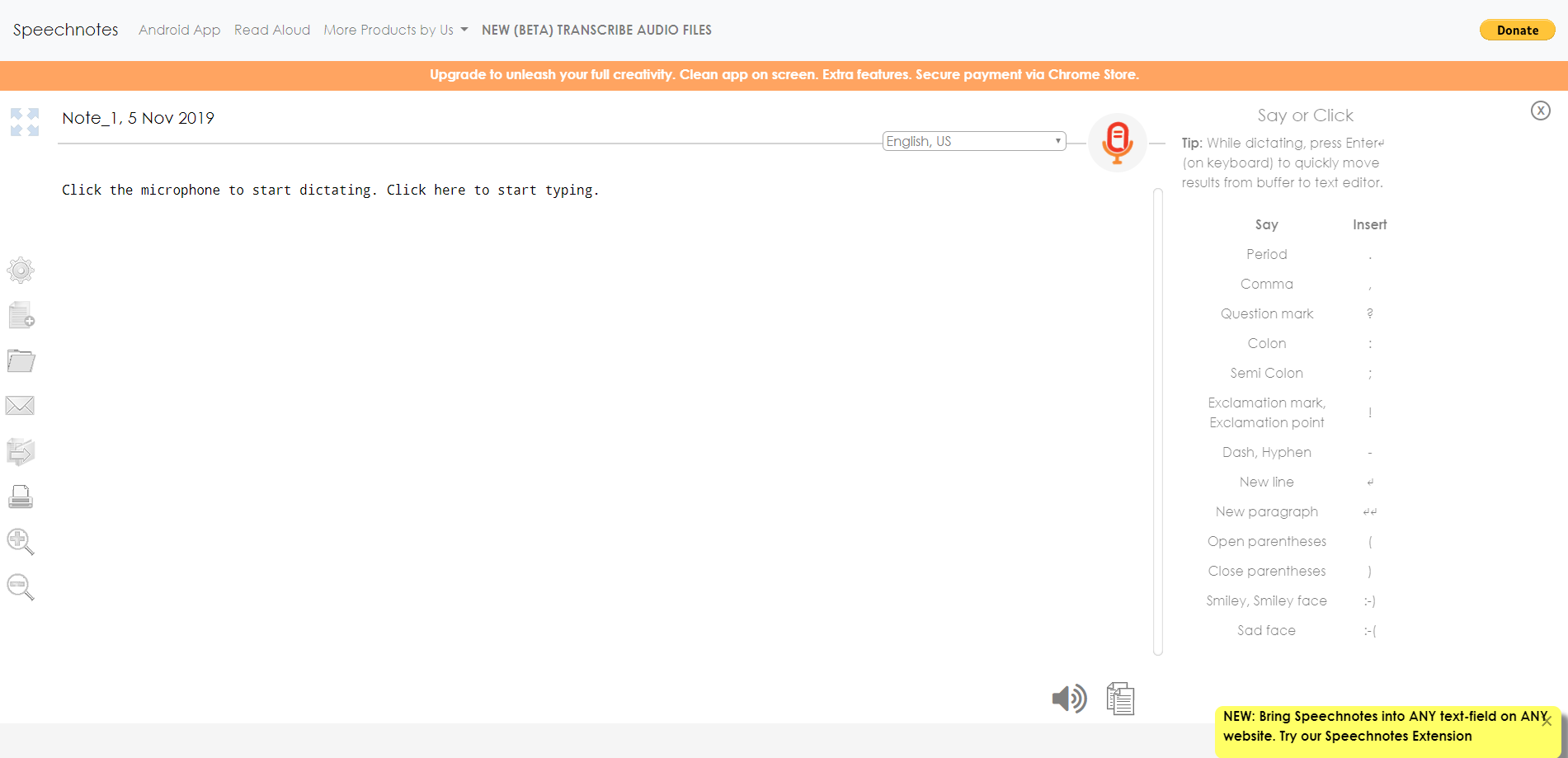
****

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 minimalistyka 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.

Do korzystania z urządzenia potrzebna 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.

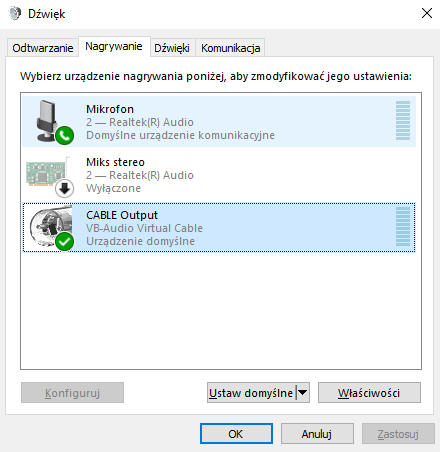
SpeechNotes opiera się na zaawansowanych technologiach rozpoznawania mowy firmy Google. W rzeczywistości cała mowa jest wysyłana do Google, tam jest interpretowana za pomocą wydajnych serwerów równoległych i algorytmów, i wysyłana z powrotem do SpeechNotes jako strumień możliwych wyników transkrypcji. 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 poziom dokładności, powinien wynosić powyżej 90%.

1. **Badania**

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

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

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 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, że będą 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.

Rysunek 1 - Pomiar prostych słów (Dictation.io)

Rysunek 2- Pomiar prostych słów (Speechnotes)

Rysunek 3 - Porównanie efektywności obu narzędzi

W pierwszym teście oba narzędzia sprawdziły się z ponad 94% efektywnością. Dictation.io sprawdziło się dokładnie tak samo dobrze i w przypadku głosu żeńskiego jak i męskiego, Speechnotes natomiast wykazał perfekcyjny wynik dla głosu męskiego oraz niewielką ilość pomyłek dla żeńskiego.

Najtrudniejszym słowem do rozpoznania okazało się słowo „Wszystko”

* 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 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, zdania zostały pozbawione znaków interpunkcyjnych, ponieważ, te narzędzia ich nie rozpoznają.

Rysunek 4 - Pomiar zdań (Dictation.io)

Rysunek 5 - Pomiar zdań (Speechnotes)

Rysunek 6 - Porównanie efektywności obu narzędzi

Jak widać na wykresach jeżeli chodzi o rozpoznawanie całych zdań oba narzędzia sprawdzają się dokładnie tak samo efektywnie. Każde z nich miało większe problemy z rozpoznaniem mowy żeńskiej niż męskiej.

* 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.

Rysunek 7 - Pomiar trudnych słów (Dictation.io)

Rysunek 8 - Pomiar trudnych słów (Speechnotes)

Rysunek 9 - Porównanie efektywności obu narzędzi

Zaawansowane lingwistycznie wyrazy sprawiły dużo większy problem dla narzędzi niż proste słowa oraz zdania. Procent poprawnego rozpoznania mieści się w zakresie 84%-86% jednak nie jest to zły wynik biorąc pod uwagę że słowa wykorzystane do badania nie są często spotykane.

* 1. **Krótkie fragmenty literatury**

Do ostatniego badania wybraliśmy dłuższy fragment książki „Innowatorzy”. Użyte narzędzia nie potrafią określić interpunkcji dlatego nie będzie ona brana pod uwagę.

Sprawdzany Tekst:  
*„Większość innowacyjnych rozwiązań ery cyfrowej jest owocem współpracy. Powstały przy udziale wielu fascynujących postaci, zarówno tych po prostu bardzo pomysłowych, jak i kilku prawdziwych geniuszy. Książka którą czytasz opowiada właśnie o tych pionierach hakerach, wynalazcach i przedsiębiorcach.”*

Pierwsze badanie polegało na podziale tekstu na zdania tak jak było to sprawdzane w poprzednich badaniach jednak tekst czytany był w całości po czym następowało sprawdzenie poprawności pojedynczych zdań. Tekst czytany był 6 razy trzykrotnie przez kobietę i trzykrotnie przez mężczyznę.

|  |  |
| --- | --- |
| Zdania | Procent rozpoznanych słów |
| Większość innowacyjnych rozwiązań ery cyfrowej jest owocem współpracy | |
| Damski | 100% |
| Męski | 100% |
| Powstały przy udziale wielu fascynujących postaci zarówno tych po prostu bardzo pomysłowych jak i kilku prawdziwych geniuszy | |
| Damski | 0.833% |
| Męski | 0.666% |
| Książka którą czytasz opowiada właśnie o tych pionierach hakerach wynalazcach i przedsiębiorcach | |
| Damski | 100% |
| Męski | 0.833% |

Dictation.io

|  |  |
| --- | --- |
| Zdania | Procent rozpoznanych słów |
| Większość innowacyjnych rozwiązań ery cyfrowej jest owocem współpracy | |
| Damski | 100% |
| Męski | 100% |
| Powstały przy udziale wielu fascynujących postaci zarówno tych po prostu bardzo pomysłowych jak i kilku prawdziwych geniuszy | |
| Damski | 0.66% |
| Męski | 0.66% |
| Książka którą czytasz opowiada właśnie o tych pionierach hakerach wynalazcach i przedsiębiorcach | |
| Damski | 0.833% |
| Męski | 0.833% |

Speechnotes

Wyniki pokazują że narzędzie dictation.io sprawdza się lepiej w przypadku głosu żeńskiego niż Speechnotes. Dla głosu męskiego sprawdzają się one dokładnie tak samo.

Drugie badanie polegało na sprawdzeniu stosunku poprawnie rozpoznanych wyrazów do wszystkich wyrazów w tekście. Wybrany przez nas tekst składa się z 37 słów a ponieważ testy zostały wykonane 6 razy to liczba wyrazów równa jest 222.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Narzędzie |  | Liczba wyrazów rozpoznanych | | Liczba wyrazów wzoru | | Skuteczność | |
| https://speechnotes.co/ | | |  | | | | |
| Damski | | | 213 | | 222 | | 95.94% |
| Męski | | | 198 | | 222 | | 89.18% |
| https://dictation.io/speech | | |  | | | | |
| Damski | | | 215 | | 222 | | 96.84% |
| Męski | | | 204 | | 222 | | 91.89% |

Wyniki drugiego badania dają podobne wnioski do pierwszego to jest oba narzędzia sprawdzają się bardzo dobrze jednak lepiej wychodzi im rozpoznawanie głosu żeńskiego. W teście narzędzie dictation.io dało minimalnie lepsze wyniki.

1. **Wnioski**

W trakcie realizacji projektu sprawdzono skuteczność dwóch narzędzi umożliwiających przetwarzanie mowy naturalnej na tekst. Oba z nich są darmowymi, narzędziami online, które pozwalają na przeprowadzenie przetwarzania dla bardzo wielu różnorakich języków, w tym także języka polskiego.

Narzędzia te posiadają bardzo prosty w obsłudze interfejs i zostały stworzone jako rodzaj głosowego notatnika. W wyniku przeprowadzonych badań wynika, że radzą sobie dość dobrze z dyktowaniem słów, prostych zdań, czy też bardziej złożonych tekstów.

**Pytania:**

**Jak mierzyć poprawność dłuższego tekstu?**

**Zakłócenia, jak dobierać, filtr czy szum przy nagrywaniu?**

**Ilości pomiarów?**