

Prof. dr hab. inż. Andrzej Czyżewski
Politechnika Gdańska, Wydział ETI
Katedra Systemów Multimedialnych

17. 09. 2024 r.

Opinia o rozprawie doktorskiej **Michała Junczyka**
pt.: "Application of speech datasets management methods for the evaluation of Automatic
Speech Recognition systems for Polish"
przygotowanej w ma Wydziale Matematyki i Informatyki Uniwersytetu im. Adama
Mickiewicza w Poznaniu pod kier. prof. dr hab. Krzysztofa Jassemę

Opinię tę opracowałem na zlecenie Dziekana Wydziału Matematyki i Informatyki UAM,
które wpłynęło do mnie w lipcu 2024 r.

Postępowanie toczy się na podstawie przepisów Ustawy z dn. 20 lipca 2018 r. Prawo o
szkolnictwie wyższym i nauce (Dz. U. z 2022 r. poz. 574 z późniejszymi zmianami). Układ mojej
opinii ma formę odpowiedzi na typowe pytania stawiane recenzentom prac kwalifikacyjnych.

1. Jakie zagadnienie naukowe/badawcze jest rozpatrywane w pracy (cel i teza rozprawy) i
czy zostało ono dostatecznie sformułowane przez autora

W rozprawie doktorskiej Michała Junczyka, zatytułowanej "Application of speech datasets
management methods for the evaluation of Automatic Speech Recognition systems for Polish,"
głównym zagadnieniem naukowym jest poprawa metod zarządzania zbiorami danych mowy i
oceny systemów automatycznego rozpoznawania mowy (ASR) dla języka polskiego. Autor dąży
do poprawy metod zarządzania zbiorami danych mowy i benchmarkingu systemów ASR,
zainspirowanych międzynarodowymi standardami. W związku z tym zrealizował następujące
zadania:

- Przeprowadzenie przeglądu dostępnych publicznie zbiorów danych mowy dla języka
polskiego i ich ocena
- Opracowanie i udostępnienie nowych, benchmarkowych zbiorów danych mowy dla języka
polskiego
- Porównanie wydajności różnych systemów ASR przy użyciu opracowanych zbiorów danych
- Organizacja otwartego konkursu dla społeczności ASR, aby promować współpracę i innowacje
w dziedzinie

Główna teza główna rozprawy zawiera hipotezę, że poprawa metod zarządzania zbiorami
danych odpowiadających sygnałowi mowy i benchmarkingu systemów ASR dla języka polskiego
może znacznie zwiększyć efektywność i wiarygodność oceny tych systemów. Autor argumentuje,
że dostępność i interoperacyjność publicznie dostępnych zbiorów danych mowy są kluczowe dla
rozwoju systemów ASR, a ich brak utrudnia weryfikację i powtórzenie testów jakościowych.
Autor uzasadnia ponadto, że opracowane przez niego zbiory danych mowy (BIGOS i PELCRA)

stanowią istotny wkład w dziedzinę, poprawiając metody zarządzania danymi i benchmarkingu systemów ASR. Teza rozprawy doktorskiej jest poparta wynikami badań, które wykazały znaczne różnice w wydajności systemów ASR w różnych scenariuszach testowych.

Zagadnienie naukowe jest jasno i precyzyjnie sformułowane w rozdziale 1, "Introduction", w którym Autor przedstawia tło problemowe, cel badań, hipotezę badawczą, cele i pytania badawcze. Autor wyjaśnia, że głównym problemem jest brak dostępnych i interoperacyjnych zbiorów danych mowy dla języka polskiego, co utrudnia ocenę systemów ASR.

Treść rozprawy jest spójna z głównym celem i tezą. Autor systematycznie przedstawia przegląd literatury, metodologię badawczą, wyniki badań i dyskusję, co potwierdza postawioną tezę.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł, w tym, literatury światowej, stanu wiedzy

W rozprawie doktorskiej Michała Junczyka, przeprowadzono szczegółową i właściwą analizę źródeł, w tym literatury światowej i stanu wiedzy w dziedzinie rozpoznawania mowy. Poniżej przedstawiam uzasadnienie:

- Rozprawa zawiera obszerny (liczący 154 pozycje) przegląd literatury, który obejmuje zarówno polskie, jak i międzynarodowe źródła. Autor omawia istniejące metody i narzędzia związane z zarządzaniem zbiorami danych mowy i oceną systemów ASR.

- W rozdziale 2, "Literature Review," autor przedstawia szczegółowy przegląd literatury, w tym benchmarkowanie systemów uczenia maszynowego, benchmarkowanie systemów ASR, metody zarządzania zbiorami danych mowy i narzędzia do oceny systemów ASR.

- Praca uwzględnia aktualny stan wiedzy w dziedzinie rozpoznawania mowy, w tym wyzwania związane z zarządzaniem zbiorami danych mowy i oceną systemów ASR. Autor omawia specyficzne wyzwania związane z językiem polskim i jego strukturą fonetyczną i morfologiczną.

- Autor przeprowadził krytyczną analizę istniejących zbiorów danych mowy i benchmarków, identyfikując ich słabości i ograniczenia. Praca zawiera również porównanie różnych systemów ASR i modeli, co pozwoliło na ocenę ich wydajności w różnych scenariuszach testowych.

- W rozdziale 4, "Results," Autor przedstawia wyniki badań, w tym wyniki ankiety dotyczącej dostępnych zbiorów danych mowy, cechy i dostępność opracowanych benchmarkowych zbiorów danych, oraz wyniki oceny systemów ASR przy użyciu opracowanych zbiorów danych.

- Praca uwzględnia współczesne standardy i praktyki w benchmarkowaniu systemów ASR, zainspirowane międzynarodowymi standardami. Autor promuje najlepsze praktyki w zarządzaniu zbiorami danych mowy i ocenie systemów ASR, co jest istotne dla rozwoju dziedziny.

Podsumowując, rozprawa doktorska Michała Junczyka przeprowadza właściwą i szczegółową analizę źródeł, w tym literatury światowej i stanu wiedzy w dziedzinie rozpoznawania mowy. Praca wykazuje się solidną metodologią badawczą i krytyczną analizą istniejących zbiorów danych i benchmarków, co jest kluczowe dla rozwoju dziedziny.

3. Czy autor rozwiązał postawione zagadnienia; czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione, czy wykazał umiejętność poprawnego przedstawienia uzyskanych przez siebie wyników?

Rozprawa doktorska Michała Junczyka, wykazuje, że Autor skutecznie rozwiązał postawione zagadnienia, zastosował właściwą metodologię i przedstawił wyniki w sposób transparentny i uzasadniony. Poniżej przedstawiam szczegółowe uzasadnienie:

- Autor zidentyfikował kluczowe problemy związane z dostępnością i interoperacyjnością publicznie dostępnych zbiorów danych mowy dla języka polskiego, co utrudnia ocenę systemów ASR.

- Praca obejmuje kompleksowe badania nad zarządzaniem zbiorami danych mowy i oceną systemów ASR, w tym przegląd dostępnych zbiorów danych, kurację i przygotowanie nowych benchmarkowych zbiorów danych, oraz porównanie wydajności różnych systemów ASR.

- Autor stworzył i udostępnił nowe, kompleksowe zbiory danych mowy (BIGOS i PELCRA), które obejmują prawie 400.000 nagrań i ponad 800 godzin mowy od 5.000 mówców.

- Powyższe zbiory danych zostały zaprojektowane, aby poprawić metody zarządzania danymi i benchmarkingu systemów ASR, zainspirowane międzynarodowymi standardami.

- Autor zastosował solidną metodologię badawczą, obejmującą survey, projektowanie i kurację benchmarkowych zbiorów danych, oraz implementację systemu do benchmarkingu systemów ASR.

- Praca zawiera szczegółowe opisy metodologii, co pozwala na powtórzenie i weryfikację wyników.

- Autor porównał wydajność 7 systemów ASR i 25 modeli przy użyciu opracowanych zbiorów danych, co pozwoliło na ocenę ich wydajności w różnych scenariuszach testowych.

- Wyniki badań wykazały znaczne różnice w wydajności systemów ASR w różnych testach.

- Autor zapewnił transparentność wyników, udostępniając wszystkie zasoby i wyniki badań, co umożliwia innym badaczom powtórzenie i weryfikację wyników.

- Praca zawiera szczegółowe opisy wyników, w tym wyniki ankiety dotyczącej dostępnych zbiorów danych mowy, cechy i dostępność opracowanych benchmarkowych zbiorów danych, oraz wyniki oceny systemów ASR.

- Autor przeprowadził dyskusję wyników, w której omówił znaczenie uzyskanych wyników i ich wpływ na praktyczne zastosowania systemów ASR w języku polskim.

- Praca wykazuje, że opracowane zbiory danych i metody benchmarkingu poprawiły efektywność i wiarygodność oceny systemów ASR.

Podsumowując, Autor rozprawy doktorskiej skutecznie rozwiązał postawione zagadnienia, zastosował właściwą metodologię i przedstawił wyniki w sposób transparentny i uzasadniony.

4. Oryginalność rozprawy, jej mocne strony, pozycja rozprawy w stosunku do stanu wiedzy, przydatność rozprawy dla nauk inżynieryjno-technicznych, w szczególności wkład do dyscypliny Informatyka.

Rozprawa doktorska Michała Junczyka, wykazuje się wysokim poziomem oryginalności, silnymi stronami i istotnym wkładem w stan wiedzy w dziedzinie rozpoznawania mowy. Poniżej przedstawiam uzasadnienie:

- Rozprawa obejmuje szeroki zakres tematów związanych z zarządzaniem zbiorami danych mowy i oceną systemów ASR, w tym przegląd literatury, projektowanie i kurację benchmarkowych zbiorów danych, oraz porównanie wydajności różnych systemów ASR.

- Praca wykazuje się solidną wiedzą teoretyczną i praktyczną w dziedzinie rozpoznawania mowy.

- Autor wnosi istotny wkład w rozwój dziedziny, poprawiając metody zarządzania danymi i benchmarkingu systemów ASR dla języka polskiego. Praca promuje najlepsze praktyki w benchmarkowaniu systemów ASR, zainspirowane międzynarodowymi standardami]

- Rozprawa zwiększa dostępność i interoperacyjność publicznie dostępnych zbiorów danych mowy, co ma praktyczne znaczenie dla zastosowań akademickich i przemysłowych.

- Rozprawa jest aktualna w kontekście bieżących wyzwań w dziedzinie rozpoznawania mowy, szczególnie w odniesieniu do języka polskiego.

- Praca uwzględnia najnowsze trendy i standardy w benchmarkowaniu systemów ASR, co ją sytuuje na czele badań w tej dziedzinie.

- Autor udostępnia nowe zbiory danych i narzędzia do ich zarządzania, co zwiększa możliwości porównawcze i walidacyjne systemów ASR.

- Praca promuje najlepsze praktyki w benchmarkowaniu systemów ASR, zainspirowane międzynarodowymi standardami. Autor zorganizował otwarty konkurs dla społeczności ASR, aby promować współpracę i innowacje w dziedzinie.

5. Uwagi polemiczne

Poniżej przedstawiam uwagi do dyskusji dotyczące rozprawy doktorskiej Michała Junczyka, które mogą być przydatne w dalszej dyskusji i ocenie pracy:

- Jakie są dalsze możliwości zastosowania tych zbiorów danych w innych obszarach, takich jak rozpoznawanie emocji w mowie lub analiza tonu głosu?

- Czy istnieją inne metody badawcze, które mogłyby być zastosowane w celu dalszego ulepszania metod zarządzania danymi i benchmarkingu systemów ASR?

- Jakie są implikacje tych wyników dla praktycznych zastosowań systemów ASR w języku polskim? Czy istnieją możliwości dalszego ulepszania tych systemów na podstawie uzyskanych wyników?

- Jakie są potencjalne obszary zastosowania opracowanych metod i narzędzi w innych dziedzinach inżynierijno-technicznych? Czy istnieją możliwości transferu tych metod do innych języków lub zastosowań?

- Na ile ważna jest transparentność w badaniach naukowych, szczególnie w dziedzinie rozpoznawania mowy? Czy istnieją inne sposoby na zapewnienie transparentności i dostępności wyników w badaniach?

- Jakie są dalsze możliwości promowania najlepszych praktyk w dziedzinie rozpoznawania mowy? Czy istnieją inne sposoby na organizację konkursów lub wydarzeń, które mogą przyczynić się do rozwoju dziedziny?

6. Uwagi szczegółowe

Strona	Uwaga
20	Podwójna kropka. Ponadto hasło ułożone poza porządkiem alfabetycznym, jak to występuje dalej.
20	powinien być przecinek zamiast drugiej kropki
20	Podwójna kropka
21	Podwójna kropka
21	Błędne umieszczenie odnośnika, powinien być w stopce, stopka pusta w ref. 3.
21	niepotrzebny znak:)
22	Podwójna kropka
23	podwójna kropka
23	Podwójna kropka
23	Podwójna kropka
23	NeMo toolkit is not limited to NLP
24	Podwójna kropka
24	Podwójna kropka
25	Podwójna kropka
25	Podwójna kropka
25	Podwójna kropka
25	Podwójna kropka
29	Powinna być wielka litera.
30	Wielka litera.
55	Wielka litera
59	Niepoprawny zapis, powinien być jeden znak "minus"
59	Niepoprawny zapis, powinien być jeden znak "minus"
69	Co oznaczają skróty L1 i L2?
73	Powinna być wielka litera.
77	Tutaj powinno być "has"
90	Powinna być wielka litera.
99	Jaka została przyjęta konwencja podpisu tabel? Powszechnie stosuje się podpis powyżej tabeli.
100	Powinna być wielka litera
114	Pomyłka drukarska, znak "falka" powinien być przed liczbą, nie nad pierwszą cyfrą.
115	Z treści nie wynika jasno, czy losowy dobór nagrań z podzbiorów był robiony raz, czy osobno dla każdego modelu.
115	W wybranych przypisach widnieją wprost adresy internetowe, w innych tylko nazwy stron, a adres ukryty jest odpowiadającym linku, przez co adresy będą widoczne tylko w wersji elektronicznej.
128	Ponownie - konwencja umieszczania podpisów tabel - w 4.5 i 4.6 są powyżej tabeli.
133	Tytuł tabeli jest pod tabelą.
133	Tab. 4.16 i 4.17 mają identyczne tytuły, zapewne błąd w edycji.
136	Czy nie powinno być nazwane jako średnia słów na sekundę oraz znaków na sekundę podobnie jak w Tab. 4.26 i 4.27? Wartości są przecież wyliczone na podstawie wielu nagrań.
142	Brak odwołania w tekście do Tab. 4.35.
142	Brak odwołania w tekście do Tab. 4.34.

143	Brak odwołania w tekście do Tab. 4.37.
143	Brak odwołania w tekście do Tab. 4.36.
144	Formalnie należałoby opisać oś y (WER(?))
150	Brak odwołania w treści do Fig. 4.13.
150	Przypis dolny na środku strony.
179	Nieprzetworzone komendy LaTeX.
180	Nieprzetworzona komenda LaTeX.
190	Niepoprawne formatowanie Tab. 5.1.
231	Czy w pustych polach nie można by podać informacji wyliczonej na podstawie analizy danego datasetu, którą autor wnikliwie przeprowadził, nanosząc dodatkowe dane w tabeli na przykład kursywą?
260	Pozycja bibliograficzna niekompletna.

7. Podsumowanie

Rozprawa doktorska Michała Junczyka, zatytułowana "Application of speech datasets management methods for the evaluation of Automatic Speech Recognition systems for Polish," prezentuje się jako praca o odpowiedniej jakości, charakteryzująca się oryginalnością, solidną metodologią badawczą i istotnym wkładem w dziedzinę rozpoznawania mowy.

Autor stworzył nowe, kompleksowe zbiory danych mowy (BIGOS i PELCRA) dla języka polskiego, co stanowi istotny wkład w rozwój dziedziny. Praca obejmuje przegląd dostępnych zbiorów danych mowy, kurację i przygotowanie nowych benchmarkowych zbiorów danych, oraz porównanie wydajności różnych systemów ASR. Badania wykazały znaczne różnice w wydajności systemów ASR w różnych scenariuszach testowych. Autor udostępnił wszystkie zasoby i wyniki badań, zapewniając transparentność i możliwość weryfikacji wyników. Rozprawa ma praktyczne zastosowania w dziedzinie rozpoznawania mowy, poprawiając metody zarządzania danymi i benchmarkingu systemów ASR. Praca promuje najlepsze praktyki w benchmarkowaniu systemów ASR, zainspirowane międzynarodowymi standardami.

Rozprawa doktorska prezentuje aktualny stan wiedzy teoretycznej i praktycznej w dyscyplinie Informatyka, w obszarze rozpoznawania mowy. Rozprawa obejmuje szeroki zakres tematów związanych z rozpoznawaniem mowy, w tym zarządzanie zbiorami danych mowy, ocena systemów ASR, oraz benchmarking. Autor przedstawia teoretyczne podstawy benchmarkingu systemów uczenia maszynowego i systemów ASR, co wskazuje na solidną wiedzę teoretyczną w tych obszarach. Praca zawiera szczegółową metodologię badawczą, w tym przegląd dostępnych zbiorów danych mowy dla języka polskiego, projektowanie i konserwację benchmarkowych zbiorów danych, oraz implementację systemu do benchmarkingu systemów ASR. Autor stosuje różne metody analityczne i narzędzia, co świadczy o jego umiejętnościach w zakresie projektowania i prowadzenia badań naukowych.

Rozprawa wnosi istotny wkład w dyscyplinę Informatyka, poprzez stworzenie i udostępnienie szeroko zakrojonych i dobrze udokumentowanych zbiorów danych mowy dla języka polskiego, co zwiększa możliwości porównawcze i walidacyjne systemów ASR.

W recenzji sformułowałem pytania i wylistowałem szereg uwag szczegółowych, jednak te elementy nie wpływają w zasadniczym stopniu na moją pozytywną ocenę pracy doktorskiej. W Bibliografii nie dostrzegłem odniesień do własnych publikacji Autora (poza wzmianką o Hugging

Face paper i GitHub dataset), zaś przeszukiwanie repozytoriów publikacyjnych pozwoliło mi odnaleźć jedynie jego artykuł jednoautorski, zatytułowany „Framework for Curating Speech Datasets and Evaluating ASR Systems: A Case Study for Polish” opublikowany jako preprint w repozytorium arXiv w dniu 18 lipca 2024 r.

Wniosek

Rozprawa pana mgr inż. Michała Junczyka została zrealizowana w sposób odzwierciedlający wymagane kwalifikacje jej Autora, wystarczający nakład pracy badawczej, implementacyjnej i eksperymentalnej. W mojej opinii treść rozprawy mgr inż. Michała Junczyka spełnia wymogi Prawa o Szkolnictwie Wyższym i Nauce, z dnia 20 lipca 2018 r. (Dz. U. z 2022 r. poz. 574 z późniejszymi zmianami), stawiane kandydatom do stopnia naukowego doktora.

