



Politechnika  
Wrocławska

Prof. dr hab. inż. Halina Kwaśnicka  
Katedra Inteligencji Obliczeniowej  
Wydział Informatyki i Zarządzania  
Politechnika Wrocławska

Wyb. Wyspiańskiego 27, 50 370 Wrocław  
Tel: (48)(71) 320 35 34, Fax: (48)(71) 321 10 18  
E-mail: halina.kwasnicka@pwr.edu.pl

06 marca 2019r.

## Recenzja rozprawy doktorskiej

Tytuł rozprawy: SUMMARIZATION OF POLISH PRESS ARTICLES USING COREFERENCE

Autor rozprawy: mgr Mateusz Kopeć

Przewód doktorski: w dziedzinie nauk matematycznych, w dyscyplinie Informatyka

Promotor rozprawy: dr hab. Agnieszka Mykowiecka

Promotor pomocniczy: dr Maciej Ogrodniczuk

Recenzja wykonana jest na zlecenie Rady Naukowej Instytutu Podstaw Informatyki PAN, pismo z 15 listopada 2018r.

### 1. Obszar problemowy pracy

Zasoby dokumentów zgromadzonych przez różne instytucje naukowe, ustawodawcze czy wydawnictwa wiadomości internetowych, są ogromne. Wyszukując dokumenty na zadany temat otrzymujemy zwykle ich dużą liczbę. Podobnie, gdy otwieramy serwis informacyjny, widzimy dużo różnych dokumentów. Czas przeglądania całych dokumentów, zwłaszcza jeśli nie są krótkie, jest stosunkowo długi. Przegląd krótkich, ale informatywnych streszczeń pozwoliłby w znacznie krótszym czasie wybrać interesujące dokumenty do zapoznania się z nimi w całości. Rolą streszczenia jest zatem pomoc użytkownikowi w wyborze dokumentu zgodnego z oczekiwaniami i zainteresowaniami. Temat automatycznego generowania streszczeń dokumentów jest zatem ważny z praktycznego punktu widzenia. Czy jest to temat bardzo interesujący również z naukowego punktu widzenia, łatwo się przekonać choćby poprzez sprawdzenie liczby publikacji naukowych na ten temat w renomowanych czasopismach naukowych. Zapytanie w postaci dwóch słów „summarization text” do bazy Web of Science Core Collection zwraca łącznie 2 292 dokumenty naukowe, z czego 35 jest z datą 2019 rok, dodanie do zapytania słowa „automatic” daje odpowiedź w postaci 795 dokumentów, w tym 4 z 2019 roku. Wyszukiwarka Scholar Google zwraca 111 000 linków na to samo zapytanie. Recenzowana praca dotyczy wymienionego wyżej obszaru. Autor w rozdziale 2.3 opisuje zadanie generowania streszczeń, które chce skutecznie rozwiązać w ramach tej pracy. Wcześniej, w rozdziale 1.2 Autor stawia **tezę doktoratu**, że **automatyczna analiza koreferencji może być wykorzystana z sukcesem do automatycznego tworzenia streszczeń polskojęzycznych artykułów prasowych**. „Z sukcesem” oznacza tu, że wynikowe streszczenia będą lepsze niż te, generowane przez inne, aktualne metody.

Autor skupia się na automatycznym tworzeniu streszczeń pojedynczych dokumentów (artykułów prasowych online), o konfigurowalnej długości, informatywnych, dedykowanych dla różnych odbiorców (nie dla specyficznego użytkownika). Autor uzasadnia celowość wykorzystania koreferencji w metodzie generowania streszczeń. Z uwagi na brak odpowiedniego korpusu tekstów polskojęzycznych, w ramach doktoratu opracowano taki zasób, jest on teraz publicznie dostępny.

Podsumowując tę część rozprawy stwierdzam, że Autor **podjął ważny i aktualny problem naukowy. Przeprowadził analizę istniejących podejść i możliwych rozwiązań, zaproponował autorskie rozwiązania, oraz przeprowadził badania eksperymentalne,** pozwalające ocenić opracowane rozwiązania.

## 2. Kompozycja i zawartość pracy

Praca składa się z wstępu (numerowanego jako rozdział 1.), siedmiu kolejnych rozdziałów (łącznie z podsumowaniem) oraz trzech załączników. Na początku dokumentu są spisy rysunków i tabel oraz spis listingów (przykładowy plik tekstowy korpusu; przykładowy plik z pięcioma streszczeniami; przykład ‘optymalnego streszczenia’ wg miary  $ROUGE_1$ ; przykład dokumentu zawierającego wynik działania tagera). Na końcu, po załącznikach, zamieszczona jest bibliografia. Załączniki, jak sama nazwa wskazuje, to dodatki umieszczane na końcu całej pracy, a bibliografia jest niezbędną częścią rozprawy, powinna być po podsumowaniu pracy, ale przed załącznikami.

We wstępie Autor zamieścił **motywacje** do zajęcia się realizowanym tematem, następnie **przedstawił tezę** swojego doktoratu z komentarzem na temat miar oceny jakości streszczeń, po czym omówił krótko organizację rozprawy doktorskiej. Brakuje mi już na tym etapie zdefiniowania głównych zadań badawczych (i innych), których realizacja jest niezbędna do ‘wykazania’ zaprezentowanej tezy doktoratu. **Sformułowane jednoznacznie cel i zakres pracy oraz pytania badawcze** mówią znacznie więcej niż teza, pokazują zakres i stopień trudności pracy badawczej, niezbędnej do zrealizowania doktoratu. Porządkują też pośrednio strukturę pracy.

Rozdział drugi prezentuje podstawowe pojęcia związane z realizowanym tematem. Omawia rolę streszczeń i ich typy. Kolejne sekcje omawiają tematykę koreferencji, wskazując na problem podmiotu domyślnego w języku polskim. Zgodnie z moją wiedzą, praca ta jest pierwszą próbą włączenia podmiotu domyślnego do realizowanego zadania. Autor wspomina o sposobie oceny nowego systemu streszczającego dokumenty tekstowe, poprzez porównanie streszczeń automatycznie tworzonych z tworzonymi ręcznie, przez ludzi.

Rozdział trzeci to obecny w rozprawach naukowych przegląd literatury, prezentujący stan aktualny badań w danym obszarze, tu: automatycznego tworzenia streszczeń. Autor zaczyna od prezentacji istniejących podejść. Krótko pokazuje historyczne tło, po czym przechodzi do tego zadania dla języka polskiego. Omawia tu poszczególne podejścia do generowania streszczeń dokumentów polskojęzycznych, następnie podejścia wielojęzyczne. Kończy tę część prezentacja prostego, podstawowego algorytmu. Szkoda, że nawet ten podstawowy, prosty ale stosunkowo silny algorytm nie jest zaprezentowany w postaci pseudokodu.

Druga część tego rozdziału omawia tworzenie streszczeń z wykorzystaniem koreferencji oraz występujące problemy, np. streszczenie może zawierać wzmiankę (zaimek, podmiot domyślny) która może być niezrozumiała bez wystarczającego kontekstu. Dalej są krótko omawiane kolejno prace, począwszy od najstarszych – nie jest to częsty sposób prezentacji przeglądu literatury. Rozdział pokazuje, że Doktorant zapoznał się dobrze z istniejącą literaturą przedmiotu. W podsumowaniu rozdziału Autor zwraca uwagę, że algorytmy dla języka polskiego, z pewnymi wyjątkami, nie są publicznie dostępne, co utrudnia ich ocenę. Dostępne oceny systemów np. Świetlickiej i Dudczaka są nieporównywalne z racji sposobu ich oceny. Problemem też jest dostęp do dobrych gotowych narzędzi użytecznych w wykrywaniu koreferencji. Problem automatycznego streszczania jest trudny w ogólności, tym bardziej dla języka polskiego. Praca skupia się na selekcji treści dokumentu do streszczenia (rozdział 5.) oraz poprawie streszczenia poprzez włączenie podmiotu domyślnego (rozdział 6.). Odnoszę wrażenie, że brak na początku pracy jasno wypunktowanego celu i zakresu pracy oraz wynikających z tego zadań badawczych, powoduje, że Autor ponownie wraca tu do wskazania zadania rozwiązywanego w ramach doktoratu.

Do oceny zaproponowanych metod konieczne było wykonanie znacznie większego zbioru polskich streszczeń. Tej tematyce poświęcony jest rozdział 4., który omawia POLISH SUMMARIES CORPUS, udostępniony publicznie, zaprezentowany wcześniej w publikacji współautorstwa Doktoranta i promotora pomocniczego. Wykonanie korpusu jest zadaniem przede wszystkim bardzo czasochłonnym, ale też niełatwym, wymagającym wiedzy i uwagi. W pracy omówiono istniejące korpusy i ich ograniczenia, zdefiniowano wymagania w stosunku do konstruowanego korpusu. Omówiono sposób tworzenia oraz wybrane charakterystyki korpusu. W podsumowaniu zwrócono uwagę na problem przy stosowaniu miary *ROUGE*, wynikający z wolnego szyku wyrazów w streszczeniach. Jednakowe informatywnie streszczenia mogą otrzymać inne wartości miary *ROUGE*, ponieważ opiera się ona na współwystępowaniu *n*-gramów. W kolejnym rozdziale Autor prezentuje własne podejście do oceny streszczeń, radzące sobie z tym problemem (*ROUGE-M*).

Rozdział piąty jest bardzo ważny, ponieważ prezentuje sedno doktoratu – podejścia do generowania streszczeń wykorzystujące informacje o koreferencjach, co było tezą pracy. Rozdział zaczyna się od omówienia wstępnego badania, którego celem była weryfikacja, czy istnieje korelacja pomiędzy fragmentami tekstu wybranymi jako ważne do streszczeń ‘wydobywczycy’ (*extractive*) przez ludzi, a obecnością relacji koreferencji pomiędzy wybranymi fragmentami tekstu. Pozytywna odpowiedź dowodziłaby, że ludzie wybierają częściej fragmenty z koreferencjami. Wyniki prac były publikowane w 2015 roku, ich konkluzją było to, że warto korzystać z informacji o koreferencjach pomiędzy zdaniem, nawet jeśli one byłyby identyfikowane nieprecyzyjnie. Pozostaje problem, jak wykorzystać takie informacje.

Druga sekcja piątego rozdziału prezentuje poprawioną wersję pierwszego podejścia (opublikowanego w 2015r.). Na tym etapie pojawiło mi się pytanie: ile różnych wersji Doktorant prezentuje w dysertacji? Kolejny problem mam z wstępem do sekcji 5.2. Autor mówi o zaimplementowanym „narzędziu EMILY”. Doktorat to praca naukowa, zdecydowanie wolałabym, aby Autor prezentował opracowane metody, algorytmy, a nie zaimplementowane narzędzia. Narzędzia (zaimplementowane metody) służą do eksperymentalnej weryfikacji ich

działania oraz do udostępniania ich innym użytkownikom. Jeśli chodzi o osiągnięcie naukowe, to ważna jest opracowana metoda i jej właściwości. Jako pierwszy, przedstawiony jest zachłanny algorytm, przyjmujący dwa parametry: tekst, którego streszczenie ma być tworzone oraz rozmiar streszczenia, zdefiniowany jako procentowy rozmiar tekstu źródłowego. Sekcja 5.2.2, zatytułowana „Greedy summarization algorithm” kończy się fragmentem „When all the points, calculated according to the procedure presented above, are assigned to each unit in a single text, the unit scores are standardized, i.e. scaled to have mean of 0 and a standard deviation of 1. Following the application of such standardization in every training text, each training example (i.e. a unit) has its gold score attached.” (pokręślenia moje). Pytanie do Autora tekstu: czy zna metody prezentacji algorytmów? Czy to dobry zwyczaj prezentować je w postaci takiego słownego opisu? Prezentacja algorytmu powinna umożliwić czytelnikowi własne zaimplementowanie tegoż algorytmu. Śmiem wątpić, czy opis w sekcji 5.2.2 pozwala na to. Jestem negatywnie zaskoczona sposobem prezentacji algorytmów przez Autora dysertacji. Zapowiadana na początku rozdziału szczegółowa prezentacja algorytmu, w mojej ocenie nie jest szczegółową prezentacją algorytmu, a swoistym ‘esejem’. W sekcji 5.2.3 Autor omawia narzędzia, które będą wykorzystywane w czasie wstępnego przetwarzania tekstu. Wiemy, że ‘preprocessing’ jest w całości dokonywany automatycznie, wymienione i krótko przedstawione są wykorzystywane narzędzia, ale klarownego potoku przetwarzania, całej sekwencji nie ma jasno zaprezentowanej. Kolejna sekcja mówi o cechach jednostki (*unit*) – wybór zestawu cech wydaje się rozsądny, ale nie ma pogłębionej dyskusji tego problemu w pracy. Autor testował podejście z dwiema definicjami ‘jednostek’ (klauzula i zdanie) oraz z zastosowaniem regresji liniowej i drzewa regresji. Nie rozumiem, dlaczego Autor korzystał z domyślnych wartości parametrów implementacji zawartych w środowisku WEKA. Z uwagi na nieco lepsze wyniki regresji liniowej, pan Kopeć postanowił ją stosować do ewaluacji w dalszej pracy. Kolejno, Autor prowadził badania pod kątem korelacji cech związanych z koreferencjami a wynikiem wzorcowym (*gold unit scores*). Konkluzją tych badań jest, że jednostka informatywna powinna zawierać długie wzmianki w sensie liczby znaków, ale nie w sensie liczby słów. W dalszej części rozdziału przedstawione są miary oceny systemów generujących streszczenia pojedynczych dokumentów *ROUGE*. Pierwsza z nich, zliczająca współwystępowania *n*-gramów w generowanym streszczeniu i streszczeniu wzorcowym (*gold summary*), zapisana jest formalnie w postaci wzoru (str. 99). Pozostałe są zdefiniowane opisowo. Autor dołączył do oceny miarę *ROUGE-M<sub>n</sub>* – miara oceny systemu dla każdego tekstu, jako maksymalna wartość uzyskana dla pojedynczego ręcznego streszczenia (odpowiedni wzór na str. 100).

Brakuje mi zwięzłego i jednocześnie szczegółowego przedstawienia autorskiej metody. Można w zgrubny sposób przedstawić cały potok przetwarzania, a potem omówić szczegółowo metody działające na poszczególnych etapach. Poprawne zaimplementowanie autorskiej metody oraz metod referencyjnych do celów porównawczych nie jest zadaniem naukowym, wartym szczegółowego omówienia, jest jednak niezbędne, aby zrealizować zadania typowo naukowe. Autorski system jest porównywany z czterema innymi, opisanymi krótko w pracy. Porównania były wykonane dla streszczeń o długościach 20%, 10% i 5%. Ocena metod jest realizowana wg miary *ROUGE<sub>n</sub>* i *ROUGE-M<sub>n</sub>* dla *n*=1, 2, 3. Wyniki zestawione są w tabelach. Widzimy w nich autorski system EMILY w wersji EMILY-C i EMILY-S. Czytając ten fragment

należy się domyślić ich znaczenia, są one wyjaśnione akapit niżej, pod tabelami. Wyniki systemu EMILY nie są imponujące w porównaniu do wyników czterech konkurentów. W związku z tym, autor, obserwując siłę prostego algorytmu bazowego BASELINE stwierdza, że sensowne jest wprowadzenie do niego modyfikacji, aby polepszyć jego działanie. Następnie, Autor analizuje górną granicę miary ROUGE – teoretycznie możliwy maksymalny wynik. Analiza ta jest dość ciekawa, konkluzja mówi, że teoretycznie możliwe jest uzyskanie znacznie wyższej oceny niż najlepsze spośród otrzymanych w zaprezentowanych badaniach. Problemem jednak może być czytelność streszczenia. Kolejnym spostrzeżeniem Autora jest fakt, że dla streszczeń ‘wydobywczych’ popularna miara ROUGE ma mniej wad niż dla streszczeń abstrakcyjnych. Systemy ekstrakcyjne (wydobywcze), nie mogą tworzyć niezrozumiałych opisów, ponieważ zdania źródłowe mają poprawną składnię. W oparciu o wydajność ludzkich anotorów, każdy wynik powyżej 60% dla metryki  $ROUGE_1$  należy uznać za sukces.

W sekcji 5.4 Autor rozważa próbę adaptacji systemów state-of-the-art dla języka angielskiego. Wykorzystano dwa systemy, oba są nowsze niż EMILY, obecne w literaturze przedmiotu. Są to: nienadzorowany algorytm grafowy TEXTRANK (wielojęzyczny) oraz anglojęzyczny POINTERGENERATOR – implementacja głębokich rekurencyjnych sieci neuronowych. Oba systemy zostały zaadaptowane do tworzenia streszczeń dokumentów w języku polskim. Problemem z zastosowaniem uczenia głębokiego jest najprawdopodobniej za mały korpus dokumentów i streszczeń dla języka polskiego.

Korzystając z doświadczenia z systemem EMILY, Autor proponuje system NICOLAS, stawiając przed nim wymagania, aby był lepszy dla języka polskiego od systemu ŚWIETLICKA dla streszczeń o długości 20% tekstu źródłowego. Został opracowany na podstawie dokładnej analizy działania swego poprzednika. Jako dane wejściowe system przyjmuje tekst źródłowy i rozmiar streszczenia (procentowo). Etap ‘preprocessingu’ jest taki sam jak w systemie EMILY. Pierwszym spostrzeżeniem podczas analizy działania EMILY był fakt, że nie wszystkie łańcuchy koreferencji są jednakowo ważne. Dlatego podjęto próbę klasyfikacji wzmianek (*mentions*) jako ważne, bądź nie, z punktu widzenia tworzenia streszczenia. W tym celu należało zaproponować metodę identyfikacji ‘dobrych wzmianek’. Jako dobre wzmianki oznaczono te, które wystąpiły w wielu ‘wzorcowych’ (*gold*) streszczeniach. Na tej podstawie wyuczono klasyfikator binarny, zastosowano lasy losowe uczone na 562 cechach. Autor testował wiele klasyfikatorów, moje wątpliwości budzi fakt, że wszystkie z domyślnymi wartościami parametrów. Można było zbadać wrażliwość na parametry choćby najlepszego klasyfikatora i dostroić go do realizowanego zadania. Dokonano selekcji istotnych cech zdań pod kątem oceny przydatności każdego zdania do wykorzystania w streszczeniu. Wybrano klasyfikator spośród trzech testowanych, każdy z domyślnymi wartościami parametrów.

Pozytywnie odnotowuję fakt, że Autor zaprezentował w czytelny sposób ogólny schemat przetwarzania w systemie NICOLAS (rys. 5.4). Porównanie systemu NICOLAS z BASELINE, ŚWIETLICKA, TEXTRANK i POINTERGENERATOR (te dwa ostatnie dostosowane do przetwarzania języka polskiego), dla streszczeń długości 20%, wg miar  $ROUGE_n$  i  $ROUGE-M_n$  dla  $n=1, 2, 3$  pokazuje wyższość NICOLASA. W konkluzji Autor wymienia możliwości dalszego polepszenia tego systemu.

Rozdział szósty prezentuje dalszy rozwój systemu NICOLAS. Rozwój polega na automatycznym włączeniu podmiotu domyślnego, co skraca tekst streszczenia i zwiększa jego płynność. W tej części pracy również nie ma ogólnej procedury działania i opisów poszczególnych etapów, lecz całość jest podana w postaci opisowej. Sekcja 6.3 jest zatytułowana „Algorithm”, gdzie Autor informuje, że do wskazania, która wzmianka (*mention*) może być usunięta, stosuje klasyfikator binarny, a model zaczyna się od ekstrakcji cech. Cechy te są wymienione w tabeli, następnie, za pomocą prób z kilkoma klasyfikatorami ze środowiska WEKA, z domyślnymi wartościami parametrów, wybrano model bazowy do stosowania w tej części pracy. Przeprowadzono eksperymenty oceniające działanie zaproponowanego rozwiązania, wysnuto poprawne konkluzje.

Rozdział siódmy jest zatytułowany „Implementation”. Wg mojej oceny nie wnosi on wiele do osiągnięć naukowych, pokazując poprawne inżynierskie podejście do wytwarzania oprogramowania. Mógłby być załącznikiem do dysertacji, bo jest użyteczny dla potencjalnych użytkowników wytworzonego oprogramowania.

W rozdziale ósmym Autor podsumowuje swoją pracę. Wskazuje, że przy krótkich streszczeniach trudno jest ‘pobić’ algorytm podstawowy. Wynika to ze specyfiki internetowych artykułów prasowych, bowiem na początku posiadają one krótką informację o artykule, mogącą służyć jako krótkie streszczenie. Dalej Autor wskazuje główne osiągnięcia i możliwe kierunki kontynuacji prac.

Przed bibliografią umieszczone są załączniki (które powinny być po bibliografii). Pracę kończy obszerna bibliografia, cytowane prace są różnorodne – publikacje w czasopismach naukowych i konferencjach oraz dysertacje magisterskie i doktorskie. Nie mam zastrzeżeń do doboru literatury, jest bardzo obszerna i na temat.

Podsumowując tę część recenzji stwierdzam, że **praca jest obszerna, skonstruowana w sposób prezentujący chronologicznie rozwój prac w ramach doktoratu**. Praca pokazuje **dużą wiedzę Doktoranta** oraz **sposób prowadzenia badań, jest to bardzo logiczny cykl**: analiza tego co jest znane, pomysł, realizacja i badania, analiza wyników i konkluzje wskazujące na kierunki dalszych prac, prowadzących do lepszych rezultatów. Mimo, że praca napisana jest nie w taki sposób, w jaki ja bym ją widziała, to pokazuje dużą zawartość merytoryczną wykonanej pracy, w tym naukowej, nie tylko implementacyjnej związanej z wytworzeniem systemu komputerowego. Pokazuje też naukowe podejście do realizowanego zadania.

### 3. Oryginalność i waga osiągnięć

Obszar badań wyznaczony tezą zawartą w dysertacji jest **ważny i aktualny z naukowego punktu widzenia**. **Aspekt praktyczny** jest również obecny i ważny, efekty pracy mogą służyć do automatycznego streszczania dokumentów prasowych. Praca zawiera oryginalne elementy naukowe, są to:

- Analiza teoretyczna górnej granicy jakości streszczenia mierzonego miarą *ROUGE*;
- Identyfikacja słabości miary *ROUGE* i propozycja miary radzącej sobie z problemem – *ROUGE-M*;

- System *NICOLAS*, w którym połączono analizę leksykalną z danymi o koreferencjach, system produkuje streszczenia ‘wydobywcze’, działa dobrze dla streszczeń o długości 20%;
- Metoda poprawy streszczeń poprzez włączenie podmiotów domyślnych.

Wartym podkreślenia jest fakt uwzględnienia w badaniach porównawczych najlepszych systemów generujących streszczenia tekstów anglojęzycznych, po ich dostosowaniu. Godną podkreślenia cechą doktoratu jest też upublicznienie powstałego oprogramowania. Podobnie, ważnym elementem jest zbudowany i udostępniony korpus *POLISH SUMMARIES CORPUS*.

Nie mam wątpliwości, że wykonana w ramach doktoratu praca jest istotnym wkładem w rozwój metod NLP dla języka polskiego (potencjalnie również dla niektórych innych języków). Pan Kopec jest współautorem siedmiu prac w bazie *Web of Science Core Collection* oraz 9 prac w bazie *DBLP (Computer Science Bibliography)* (stan na początek marca 2019).

#### 4. Uwagi i problemy do dyskusji

W mojej ocenie, rozprawa doktorska napisana jest jako sprawozdanie z wykonanych prac – od wczesnych prób po najbardziej zaawansowany system *NICOLAS* i *NICOLAS+ZERO*. Swoją drogą, nazwa *NICOLAS+ZERO* pojawia się pierwszy raz (poza spisem treści) w tabeli 6.5 i jej podpisie, przynajmniej ja nie zauważyłam, by termin ten był gdzieś wcześniej wyjaśniony. W pracy brakuje jasno sformułowanego celu i zakresu pracy, nie są sformułowane zadania badawcze, niezbędne do realizacji postawionego celu. Nie ma sformułowanych pytań badawczych. Praca zawiera sformułowaną tezę, która, podobnie jak w większości doktoratów z Informatyki, zwłaszcza z aspektem praktycznym, jest weryfikowalna tylko w jedną stronę. Bo jak wykazać, że czegoś nie można zrobić, osiągnąć... (a ‘czego’, to zależy od tematu pracy). Jeśli Autorowi się nie uda czegoś osiągnąć, to nie oznacza, że „nie można”. Teza brzmi, że automatyczna identyfikacja koreferencji może być użyta do tworzenia automatycznych systemów tworzących streszczenia polskojęzycznych tekstów prasowych, w efekcie system będzie lepszy od aktualnie istniejących. Czy można wykazać, że jest fałszywa? W ogólności jestem przeciwna stawianiu tez w tego typu doktoratach, zamiast tego preferuję jasno zdefiniowane cel i zakres pracy, postawione pytania badawcze. Z tego wynikają niezbędne zadania do wykonania, z których te naukowe stanowią podstawę do uzyskania stopnia doktora.

Praca jest bronią w dziedzinie nauk matematycznych, w dyscyplinie Informatyka. Mam pytanie do Doktoranta: co to jest algorytm (w sensie informatycznym) i jakie są sposoby prezentacji algorytmów? Praca pisana jest opisowo – wyjątek stanowią tabele i wykresy prezentujące wyniki badań eksperymentalnych. Przez brak bardziej formalnego opisu algorytmów / metod, praca wiele razy odsyła do sekcji o zadanych numerach, co utrudniało mi czytanie pracy i analizę wkładu Autora. Praca ma być bronią w obszarze nauk matematycznych, ale nie zawiera żadnej formalnej definicji, analizy proponowanych metod np. pod kątem złożoności obliczeniowej, ani innych, formalnych elementów. Autor mógłby pokusić się o wykazanie czy zaproponowana miara *ROUGE-M* jest metryką czy tylko miarą. To tylko przykłady możliwości wykazania się umiejętnością posługiwania się formalizmem.

Eksperymenty są przedstawione opisowo, bez wyodrębnionego protokołu badawczego czy planu eksperymentów. Oczywiście, wiele informacji na temat eksperymentów można znaleźć w części opisowej pracy, co pozwala stwierdzić, że eksperymenty są poprowadzone poprawnie.

Praca jest bardzo obszerna, gdyby była napisana bardziej formalnie, to mogłaby być czytelniejsza i jej objętość mogłaby być mniejsza.

Zastanawia mnie, dlaczego Autor stosował metody ze środowiska WEKA z domyślnymi wartościami parametrów. Nie znalazłam w pracy choćby wzmianki na temat wrażliwości stosowanych metod na wartości parametrów.

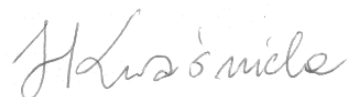
Innym elementem, który mnie zaskoczył, to sposób omówienia literatury. Oczekiwałamby spojrzeń 'z góry' na literaturę przedmiotu, zamiast prezentacji poszczególnych pozycji literaturowych. Pozytywnym elementem jest podsumowanie części przeglądowej.

Mój niedosyt po przeczytaniu pracy wynika głównie ze sposobu, w jaki Autor przedstawił swoją pracę. Merytorycznie oceniam wysoko to, co Doktorant zrealizował. Praca napisana jest starannie, bez usterek językowych i redakcyjnych.

## 1. Konkluzja

Podsumowując stwierdzam, iż przedłożona mi do recenzji rozprawa, której autorem jest mgr Mateusz Kopeć, zawiera **oryginalne i ważne osiągnięcia w obszarze przetwarzania języka naturalnego (NLP), w szczególności języka polskiego w zakresie automatycznego streszczania internetowych dokumentów prasowych**. Autor wykazał się **dużą wiedzą w tematyce rozprawy, umiejętnością pracy naukowej oraz znajomością metod badawczych**. Osiągnięte wyniki świadczą o dobrym **przygotowaniu Autora do pracy naukowej**.

**Recenzowana praca spełnia wymagania ustawowo stawiane rozprawom doktorskim, zatem wnoszę o to, by mgr Mateusz Kopeć został dopuszczony do publicznej obrony.**



Halina Kwaśnicka