

Recenzja pracy doktorskiej mgr. inż. Arkadiusza Janza  
„Metoda rozpoznawania semantycznej charakterystyki słów z  
wykorzystaniem lingwistycznie wzbogaconej reprezentacji tekstu”

Krzysztof Jassem

7 sierpnia 2022

## 1 Wstęp

Celem niniejszej recenzji jest stwierdzenie, czy rozprawa doktorska mgr. inż. Arkadiusza Janza „Metoda rozpoznawania semantycznej charakterystyki słów z wykorzystaniem lingwistycznie wzbogaconej reprezentacji tekstu” spełnia wymagania Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2003 Nr 65, poz. 595, ze zm.).

## 2 Ocena formalnej strony rozprawy

### 2.1 Ocena układu pracy

W standardowym opisie rozwiązania informatycznego oczekuje się rozdzielenia opisu danych i opisu przetwarzania. W recenzowanej pracy opisy przeplatają się. Podrozdział 3.1 nosi tytuł „Repozytoria znaczeniowe”, a rozdział 6 zatytułowany jest „Zasoby semantyczne” (są to moim zdaniem określenia synonimiczne).

Układ pracy powinien zapewniać ciągłość wypowiedzi. Tymczasem w rozprawie autorska metoda propagacji prawdopodobieństwa opisana jest w rozdziale 5, a jej ewaluacja – w rozdziale 7 (obecność rozdziału szóstego zaburza zatem ciągłość wypowiedzi).

W rozprawie doktorskiej oczekuje się zazwyczaj wyraźnego rozdzielenia opisu zstanego stanu badań od prezentacji rozwiązań autorskich. W recenzowanej pracy ten rozdział jest nieostry. Prezentacja badań własnych przeplata się z opisem zstanego stanu wiedzy oraz z formalizacją problemu badawczego. Na przykład w podrozdziale 5.1 następuje nawrót do formalnego zdefiniowania problemu z podrozdziału 1.4, choć rozdział 5 ma być w zamyśle poświęcony opisowi rozwiązania autorskiego.

W rozprawie brakuje spisu treści.

W mojej ocenie układ pracy nie został dogłębnie przemyślany.

### 2.2 Ocena stosowanego aparatu naukowego

W podrozdziale 1.4 podjęto próbę formalnego zdefiniowania problemu „rzutowania tekstu na bazę wiedzy”. W tym celu nieskomplikowany proces oznaczania wyrazów tekstowych ich interpre-

WPLYNĘŁO

01-09-2022

RDN-IT/227/2022

tacjami semantycznymi opisano za pomocą skomplikowanego aparatu funkcji wielu zmiennych. We wprowadzonym aparacie matematycznym popełniono błąd odwrócenia dziedziny i przeciwdziedziny funkcji  $\phi$ .

Matematyczna formalizacja problemu ujednoznaczniania nabiera sensu dopiero wtedy, gdy problem ten staramy się rozwiązać metodami numerycznymi, na przykład za pomocą sieci neuronowych. Matematyczny opis takiego podejścia znajduje się w rozdziale 5.1. Moim zdaniem podrozdział 5.1 powinien następować bezpośrednio po podrozdziale 1.4, aby uzasadnić sens wprowadzania aparatu matematycznego.

### 2.3 Ocena spójności terminologicznej

W pracach z nauk ścisłych oczekuje się spójności terminologicznej: jeden byt w obrębie całej pracy powinien być określany tym samym terminem. W recenzowanej rozprawie kluczowe pojęcie określone jest na wiele sposobów:

- klasa semantyczna
- kategoria semantyczna
- tag
- tak znaczeniowy
- sens
- znaczenie z repozytorium
- znaczenie z repozytorium znaczeniowego
- charakterystyka semantyczna
- znaczenie
- jednostka znaczeniowa
- charakterystyka semantyczna.

W różny sposób nazywane są byty, które ulegają ujednoznacznianiu. Są to wymiennie: wyrazy tekstowe, wyrażenia lub jednostki leksykalne.

W obrębie pracy następujące określenia stosowane są wymiennie:

- wyznaczanie charakterystyki semantycznej słów
- rozpoznawanie charakterystyki semantycznej słów
- rzutowanie tekstu na bazę wiedzy
- ujednoznacznianie znaczeń słów

- ujednoznacznianie sensów słów
- WSD.

Od pracy naukowej oczekuje się, aby kluczowe terminy zostały **zdefiniowane**. „Definicja jest zdaniem złożonym z 2 członów: definiowanego (łac. definiendum) i definiującego (łac. definiens), połączonych spójnikiem definicyjnym („jest to”, „to tyle, co” itp.)” (według Encyklopedii PWN).

Nie znalazłem w pracy definicji terminu zawartego w temacie: „charakterystyka semantyczna słów”. Nie znalazłem w pracy definicji pojęcia „rozpoznawanie charakterystyki semantycznej słów”. W podrozdziale 1.5 możemy dowiedzieć się, że automatyczne ujednoznacznianie znaczeń słów jest szczególnym przypadkiem wyznaczania semantycznej charakterystyki słów w tekstach, ale nie można tego uznać za formalną definicję pojęcia kluczowego dla rozprawy.

Można natomiast w pracy odnaleźć zajmującą razem z przykładem pół strony 18. definicję pojęcia *leksem*, które to pojęcie nie jest w pracy przywoływane (W przeciwieństwie do innych pojęć lingwistycznych, jak *lemat*, *pseudomorfem*, *subwyraz*, które są przywoływane, ale nie są zdefiniowane).

Uważam, że recenzowana rozprawa nie zapewnia spójności terminologicznej wymaganej od pracy naukowej.

## 2.4 Ocena stosowanego języka

Oczekuje się, że w języku naukowym unika się stwierdzeń nieprecyzyjnych. W recenzowanej pracy wielokrotnie występuje określenie „pewnego rodzaju”. Często stosowanym przymiotnikiem, określającym metody naukowe, jest wyraz „miękki”. Stosowanie tego typu określeń w pracy naukowej jest przyznaniem się do nieumiejętności precyzyjnego sformułowania problemu.

Prace z dziedziny informatyki napisane w języku polskim borykają się z problemem tłumaczenia terminów angielskich, które nie posiadają ugruntowanych polskich odpowiedników. W tym aspekcie trudno stwierdzić, czy dane tłumaczenie terminu jest poprawne, można tylko niekiedy zasugerować inną możliwość. Moje sugestie dotyczą tłumaczenia pojęć: neural na *neuronowy* (a nie neuronalny), oraz *word embedding* na reprezentację wektorową słowa a nie – osadzenie słowa.

Język pracy jest hermetyczny, nadmiernie skomplikowany, w pracy jest wiele błędów językowych. Dla przykładu dwa pierwsze zdania ze strony 69 brzmią następująco:

1. „Adekwatna do problemu eksploracja i łączenie z kontekstowymi osadzeniami jest istotnym zagadnieniem badawczym”.

W zdaniu popełniono następujące błędy:

- sztuczna składnia (powinno być: „Eksploracja adekwatna do problemu”);
- zbyteczne zapożyczenie z języka angielskiego („eksploracja”);
- brak dopełnienia dla rzeczownika pochodzącego od czasownika, wymagającego dopełnienia (eksploruje się coś);

- brak jednego z dwóch wymaganych dopełnień dla rzeczownika odczasownikowego „łączenie” (łączy się coś z czymś);
  - stosowanie niewyjaśnionego terminu „kontekstowe osadzenie” (w rozprawie zdefiniowano termin podobny, ale nie identyczny: „osadzenia słów”);
  - brak nowej informacji w zdaniu – stwierdzenie, że coś jest „istotnym zagadnieniem badawczym” nie wnosi wartości do pracy naukowej.
2. „Omówione w niniejszym podrozdziale metody nie rozwiązują jednak problemu, gdyż nie rozważają wykorzystania dalszych powiązań między znaczeniami (lub rozważają je w nieprecyzyjny sposób) i koncentrują się głównie na relacjach paradygmatycznych.”
- błąd niepoprawnej referencji – nie wiadomo, do czego odwołuje się wyraz „problemu” (w zdaniach poprzedzających nie ma mowy o jakimkolwiek problemie);
  - błąd niepoprawnej personifikacji (metody nie mogą nic rozważać, bo nie są ludźmi);
  - błąd semantyczny – nie można rozważać czegoś (np. podjęcia jakiejś decyzji) w sposób precyzyjny lub nieprecyzyjny;
  - stosowanie niewyjaśnionego terminu „relacje paradygmatyczne”.

Stwierdzam, że formalna strona praca pozostawia wiele do poprawy.

### 3 Ocena przeprowadzonych eksperymentów

Głównym osiągnięciem autora jest opracowanie metody „kontekstowej propagacji prawdopodobieństwa” w celu ujednoznacznienia wyrazów przy użyciu sieci typu WordNet. W rozdziale 7. opisano eksperymenty, których celem była optymalizacja skuteczności działania tej metody. Przeprowadzono 10 eksperymentów, w których na różnych zbiorach testowych porównywano skuteczność poszczególnych wariantów rozwiązania autorskiego z algorytmami referencyjnymi. Wykazano, że zaproponowana metoda wraz z jej usprawnieniami jest osiągnięciem, które podnosi poziom stanu badań w dziedzinie.

### 4 Ocena dorobku naukowego

W bibliografii znajduje się sześć pozycji, w których doktorant jest pierwszym autorem. Dwie z tych publikacji są wysoko punktowane na liście ministerialnej (odpowiednio 140 i 70 punktów). Wydaje mi się, że ułożenie publikacji w „zbiór opublikowanych i powiązanych tematycznie artykułów naukowych” (cytat z Ustawy: Prawo o szkolnictwie wyższym”) znacznie ułatwiłoby ocenę rozprawy.

### 5 Podsumowanie recenzji

W podsumowaniu recenzji odwołuję się do poszczególnych ustępów ustawy o rozprawie doktorskiej.

1. ust. 1. „Rozprawa doktorska, przygotowywana pod opieką promotora albo pod opieką promotora i promotora pomocniczego, powinna stanowić oryginalne rozwiązanie problemu naukowego lub oryginalne rozwiązanie problemu w oparciu o opracowanie projektowe, konstrukcyjne, technologiczne, lub oryginalne dokonanie artystyczne, oraz wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej lub artystycznej oraz umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej.”

Autor posiada ogólną wiedzę teoretyczną w dziedzinie, co wykazuje w części II rozprawy.

Przedmiotem recenzowanej rozprawy jest oryginalne rozwiązanie problemu ujednoznaczniania sensów wyrazów metodą propagacji prawdopodobieństwa. Wykazano wartość naukową metody autorskiej poprzez porównanie skuteczności jej działania z innymi współczesnymi metodami dla tego zadania na dostępnych publicznie danych testowych.

2. ust. 2. „Rozprawa doktorska może mieć formę maszynopisu książki, książki wydanej lub spójnego tematycznie zbioru rozdziałów w książkach wydanych, spójnego tematycznie zbioru artykułów opublikowanych lub przyjętych do druku w czasopismach naukowych, określonych przez ministra właściwego do spraw nauki na podstawie przepisów dotyczących finansowania nauki, jeżeli odpowiada warunkom określonym w ust. 1.”

Recenzowana rozprawa jest samodzielną monografią naukową o objętości 150 stron, składającą się z trzech części i ośmiu rozdziałów.

3. ust. 6. Rozprawa doktorska powinna być opatrzona streszczeniem w języku angielskim, a rozprawa doktorska przygotowana w języku obcym również streszczeniem w języku polskim. W przypadkach, gdy rozprawa doktorska nie ma formy pisemnej powinna być opatrzona opisem w języku polskim i angielskim.”

Rozprawa napisana jest w języku polskim i zawiera streszczenia zarówno w języku polskim, jak i angielskim.

Ustawa nie określa poprawności formalnej jako warunku koniecznego rozprawy doktorskiej. Dlatego stwierdzam, że recenzowana rozprawa spełnia wymagania Ustawy w stosunku do rozpraw doktorskich.

Kayntd Jansen