

dr hab. inż. Rafał Stanisławski, prof. uczelni
Katedra Informatyki
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Opolska
e-mail: r.stanislawski@po.edu.pl

Opole, 7.08.2023

POLITECHNIKA ŚLĄSKA
Biuro Rady Dyscypliny
Automatyka, Elektronika, Elektrotechnika
i Technologie Kosmiczne
wpłynęło dnia 22.08.2023
nr 28 zał.

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy:

Automatic quality analysis of textual requirements using natural language processing

Automatyczna analiza jakości specyfikacji oprogramowania za pomocą przetwarzania języka naturalnego

Autor rozprawy: mgr inż. Jerzy Kocerka

Promotor rozprawy: dr hab. inż. Adam Gałuszka, prof. uczelni

Niniejsza recenzja została opracowana na zlecenie Przewodniczącej Rady Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Śląskiej, dr hab. inż. Moniki Kwoki, prof. PŚ, z 19 maja 2023 roku.

1. Zawartość pracy i ocena formalna

Przedłożona do oceny rozprawa doktorska zawiera łącznie 121 stron podzielonych na osiem rozdziałów, literatury liczącej 127 pozycji, streszczeń przedstawionych w języku polskim i angielskim, wykazu ważniejszych oznaczeń i struktów, oraz spisu rysunków i tabel. Dysertacja została napisana w języku angielskim. Zawartość rozprawy zaprezentowano poniżej.

W Rozdziale pierwszym przedstawiono krótkie wprowadzenie do zagadnień poruszanych w pracy. Scharakteryzowano w nim kontekst pracy, zaprezentowano obszar i zakres przeprowadzonych badań, zaprezentowano cele i tezy dysertacji, oraz omówiono jej główne osiągnięcia. Ponadto w Rozdziale pierwszym scharakteryzowano strukturę pracy.

W Rozdziale drugim zawarto wprowadzenie do inżynierii wymagań. Omówiono w nim związane z inżynierią wymagań języki formalne i półformalne, języki modelowania wymagań (*Unified Modeling Language - UML*), sieci Petriego i wybrane języki strukturalne (*Attempto Controlled English, Pattern-based specifications*). W rozdziale przedstawione zostały również wybrane narzędzia i metody do analizy oraz oceny jakości wymagań, a także techniki wykrywania w nich niejasności/niejednoznaczności. Ponadto przedstawiono metody automatycznej analizy wymagań oraz omówiono standard ISO/IEC/IEEE 29148 obejmujący m. in. inżynierię wymagań. Rozdział ma charakter wprowadzający, przedstawia narzędzia wykorzystywane w inżynierii wymagań i na tym tle uzasadnia podjęcie tematu pracy.

Rozdział trzeci dokonuje przeglądu wybranych metod analizy specyfikacji wymagań oprogramowania, a w szczególności przetwarzania języka naturalnego (*Natural Language Processing*). Omówiono również metody podejmowania decyzji i klasyfikacji zastosowane do oceny wymagań oprogramowania tj. drzewo decyzyjne, losowy las decyzyjny, metoda *Gradient Boosted Trees* i algorytm k najbliższych sąsiadów. Omówiono zalety i ograniczenia poszczególnych metod oraz zaprezentowano praktyczne przykłady ich zastosowania w procesach wytwarzania oprogramowania.

W Rozdziale czwartym podjęto próbę określenia wskaźników, które mają największy wpływ na jakość wymagań oprogramowania. W tym celu przeprowadzono ankietę spośród ekspertów, której wymagania zostały opisane w języku naturalnym. W rozdziale przedstawiono zarówno metodologię badania ankietowego i doboru ocenianych kryteriów, jak również narzędzia oceny wyników ankiet. W rozdziale zawarto analizę wyników z przeprowadzonych ankiet, która przeprowadzona została dla projektu wytwarzania oprogramowania na urządzenia wbudowane w pojazdach szynowych. Wyniki posłużą jako baza do opracowania modelu jakości wymagań i systemu automatycznej ich oceny w dalszych częściach dysertacji.

Na podstawie ankiet przeprowadzonych wśród ekspertów oraz doświadczeń Autora pracy, w Rozdziale piątym przedstawiono autorski model jakości wymagań oprogramowania. Model opiera się na następujących wskaźnikach: a) obecność struktury wymagań, b) liczba błędów językowych, c) liczba odniesień w odniesieniu do obiektów znalezionych w tekście (nazwane podmioty), d) obecność w tekście subiektywnych określeń. W rozdziale przedstawiono zasady budowy poszczególnych wskaźników. Ponadto przedstawiono również wymagania wstępne dotyczące struktury przedmiotowych wymagań. Rozdział ten stanowi główny wkład teoretyczny ocenianej dysertacji.

Rozdział szósty dokonuje analizy procesu oceny ogólnej wymagań oprogramowania na podstawie wskaźników omówionych w Rozdziale piątym. Do analizy zastosowano cztery narzędzia klasyfikacyjne opisane w Rozdziale trzecim. Do oceny poszczególnych klasyfikatorów zastosowano powszechnie używany współczynnik κ Cohena. Na podstawie analizy efektywności wykazano, że wszystkie narzędzia osiągają bardzo dobre rezultaty, ale najlepsze wyniki osiągnął algorytm losowego lasu decyzyjnego i tą metodę wybrano do późniejszej implementacji. Rozdział ten, razem z Rozdziałem czwartym, stanowi główny wkład praktyczny pracy.

Rozdział siódmy prezentuje analizę efektywności autorskiej metody zautomatyzowanego narzędzia do oceny jakości wymagań oprogramowania w porównaniu z tradycyjnym procesem ręcznym wykonanym przez zespół ekspertów. W szczególności uwzględniono czas niezbędny do analizy i rozważenia sugerowanych poprawek. Przeprowadzona analiza dowodzi wysokiej efektywności zaproponowanego w pracy systemu automatycznej oceny wymagań oprogramowania, co potwierdza istotny praktyczny wkład ocenianej pracy.

W Rozdziale ósmym zostało zawarte podsumowanie i wnioski z przeprowadzonych badań, wraz z odniesieniem do postawionej tezy dysertacji. W Rozdziale przedstawiono również kierunki potencjalnych przyszłych prac Doktoranta.

Układ pracy jest poprawny. Treści dysertacji zostały logicznie podzielone na poszczególne rozdziały pracy. Układ i zakres poszczególnych rozdziałów, podrozdziałów i sekcji nie budzi żadnych wątpliwości recenzenta.

Warta szczególnego podkreślenia jest bardzo dobra strona edycyjna pracy. Rozdziały, podrozdziały, sekcje, nagłówki, stopki itp. utrzymane są tej samej konwencji, przez co praca bardzo dobrze wygląda. Rysunki są bardzo starannie przygotowane z dbałością o estetyczne i precyzyjne

przedstawienie graficzne zastosowanych algorytmów i wyników badań. Język użyty w pracy (język angielski) jest czytelny i precyzyjny, przez co pracę czyta się dobrze. Jedyną wątpliwość budzi w tym aspekcie jest dość selektywny wybór oznaczeń zestawionych na końcu dysertacji. Należy ponadto zaznaczyć, że rozprawa jest bardzo dobrze przygotowana redakcyjnie. Występujące nieliczne błędy typograficzne zostały wypunktowane w dalszej części recenzji.

2. Ocena merytoryczna pracy

Pierwsze użycie terminu *inżynieria wymagań* miało miejsce prawdopodobnie w 1964 r. na jednej z konferencji naukowych poświęconych wytwarzaniu oprogramowania, ale wszedł on do powszechnego użytku dopiero pod koniec lat 90 XX wieku. Jednak wraz ze zwiększaniem się złożoności systemów informatycznych znaczenie tego zagadnienia bardzo szybko rośnie. Obecnie to cała gałąź inżynierii systemów i oprogramowania, w której powstały cykle konferencji i czasopisma poświęcone wyłącznie temu zagadnieniu. Jest to obecnie istotny element procesu wytwarzania oprogramowania, który stanowi kluczową rolę w kosztach i jakości/niezawodności oprogramowania. Choć zwykle inżynieria systemów dotyczy produkcji różnego rodzaju oprogramowania, to również stosuje się ją w alternatywnych procesach obejmujących implementację systemów regulacji/sterowania, kontroli i innych. W tym obszarze powstało i wciąż powstaje wiele narzędzi informatycznych wspierających ten proces. Problematyka specyfikacji wymagań jest dość trudna, gdyż próbuje skwantyfikować często trudne do określenia wielkości. Dlatego ocena jakości wymagań i ich ew. modyfikacja to proces pozytywnie wpływający na koszt i jakość wytwarzanego oprogramowania. Należy zauważyć, że obecnie automatyzacja procesu oceny wymagań (w przeciwieństwie do automatyzacji procesów testowania oprogramowania) jest nieco niszowym zagadnieniem, któremu poświęconych zostało dotychczas dosłownie kilka prac naukowych. Jednak znaczna liczba cytowań tych prac oraz obecne trendy polegające na poszukiwaniu nowych obszarów dających perspektywę automatyzacji procesów, pokazując, że obrany przez Doktoranta kierunek jest interesujący i perspektywiczny. Autor opracowane narzędzia implementuje w procesie wytwarzania oprogramowania na urządzenia wbudowane stosowane w przemyśle kolejowym, gdzie kwestie niezawodności i bezpieczeństwa są kluczowe, a ponadto są regulowane dodatkowymi przepisami. Dlatego podjęcie tematu pracy w tym obszarze wydaje się być tym bardziej uzasadnione.

W związku z powyższym Autor podejmuje pracę w ważnym i aktualnym obszarze badań, który ma swoje podstawy w światowej literaturze naukowej, jak również daje możliwości osiągnięcia nowych, istotnych wyników szczególnie w ujęciu praktycznym. Recenzent umiejscawia przedmiotową dysertację w dwóch dyscyplinach naukowych tj. *Informatyka Techniczna i Telekomunikacja* i *Automatyka, Elektronika i Elektrotechnika i Technologie Kosmiczne* w dziedzinie nauk inżyniersko-technicznych. Choć sama tematyka podjęta w pracy dotyczy raczej pierwszej z w/w dyscyplin, to zastosowane narzędzia i przeznaczenie (systemy wbudowane w przemyśle kolejowym) pozwalają, zdaniem recenzenta zaliczyć ją z powodzeniem do drugiej z w/w dyscyplin.

Doktorant postawił sobie ambitny i ważny cel opracowania procedury mającej na celu automatyczną analizę i wykrywanie niejednoznacznych, nieprecyzyjnych, błędnych lub niedostatecznie opisanych wymagań, które są łatwo implementowane w środowiskach przemysłowych. Zbudowane narzędzia mogą zostać zastosowane do optymalizacji czasu i kosztu procesu modelowania wymagań. Autor przedstawił następującą treść rozprawy:

Automatyczna analiza jakości wymagań poprawia efektywność procesu tworzenia i testowania oprogramowania dla systemów wbudowanych w stosunku do metod powszechnie stosowanych w przemyśle.

Zwartość rozprawy, opisana w poprzednim punkcie recenzji, wynika bezpośrednio z postawionych celów pracy i prowadzi do dowiedzenia w/w tezy. Jako główne narzędzie do określania wymagań Doktorant dokonał słusznego wyboru języka naturalnego, który wydaje się być oczywisty w kontekście zodań systemu, jak również ma uzasadnienie w analizie literaturowej przedmiotu. Zastosowane metody klasyfikacji, jak również narzędzia statystyczne użyte przez Doktoranta w procesie ewaluacji modelu nie budzą większych wątpliwości recenzenta. Pewne pytania i uwagi dotyczące tych metod zostały zawarte dalszej części recenzji. Ostatecznie, do podstawowych osiągnięć dysertacji można zaliczyć:

- Określenie wskaźników, które mają wpływ na jakość wymagań oprogramowania w oparciu o przeprowadzoną ankietę spośród ekspertów, na przykładzie oprogramowania wbudowanego stosowanego w przemyśle kolejowym.
- Opracowanie nowego modelu jakości wymagań wykorzystujący zarówno cechy opisane dotychczas w literaturze, jak nowe wskaźniki jakości, które zostały zaproponowane przez Doktoranta, tj. liczbę odniesień do innych wymagań i strukturę wymagania.
- Przeprowadzenie obszernych badań w celu określenia generalnej jakości wymagań oprogramowania na urządzenia wbudowane. Przebadanie efektywności różnych klasyfikatorów w tym kontekście tj. drzewo decyzyjne, losowy las decyzyjny, metoda *Gradient Boosted Trees* i algorytm *k* najbliższych sąsiadów.
- Ustanowienie procesów decyzyjnych w celu ustalenia, czy wymaganie może być wykorzystane później w procesie tworzenia oprogramowania, czy też wymaga ulepszenia.
- Przeprowadzenie analizy efektywności autorskiej metody zautomatyzowanego narzędzia do oceny jakości wymagań oprogramowania w porównaniu do podejścia klasycznego.

Finalnie w ramach pracy powstało nowe, oryginalne rozwiązanie problemu jakości wymagań programowych poprzez ocenę automatyczną w oparciu o narzędzia modelowania. Pozwala to na optymalizację kosztów i czasu wytwarzania oprogramowania, co należy zaliczyć do istotnych osiągnięć praktycznych pracy.

Pomimo pewnych (w większości drobnych) uwag recenzenta do niektórych wyników prezentowanych w dysertacji, które zawarto w części czwartej niniejszej recenzji, należy jednoznacznie stwierdzić, że wszystkie wyżej wymienione osiągnięcia są znaczące i stanowią istotny wkład Kandydata w rozwój dyscypliny *Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne* oraz finalnie doprowadziły do osiągnięcia celów pracy. Ponadto otrzymane w dysertacji wyniki jednoznacznie udowadniają postawioną tezę.

Generalnie praca ma charakter praktyczny. Główny wkład Kandydata stanowią Rozdziały czwarty, piąty, szósty i siódmy. Pewnych nowych wyników teoretycznych należy upatrywać w rozdziale Rozdział piątym, natomiast Rozdziały czwarty, szósty i siódmy mają charakter praktyczny.

3. Analiza źródeł, pozycja rozprawy, znaczenie wyników Autora, umiejętność przedstawiania wyników

Motywacja dla podjęcia tematu rozprawy wynikała z dobrze przeprowadzonej przez Autora analizy literatury przedmiotu, liczącej 127 pozycji. Dzięki szerokiej analizie literaturowej został poprawnie odzwierciedlony aktualny stan wiedzy na temat wszystkich zagadnień podejmowanych w pracy. Recenzentowi zabrakło kilku odniesień do pozycji bibliograficznych, które zostaną opisane w kolejnej części recenzji.

Pozycja rozprawy w stosunku do stanu wiedzy reprezentowanej w literaturze światowej jest dobra. Jest on współautorem jednego artykułu w czasopiśmie indeksowanym na liście JCR, tj. *Advances in Science and Technology Research Journal* oraz jednego referatu na konferencji międzynarodowej *IEEE International Conference on Methods & Models in Automation Robotics*. Ponadto jest współautorem jednego artykułu na stronie korporacji *National Instruments*. Dorobek publikacyjny Doktoranta spełnia wszystkie wymagania ustawowe stawiane rozprawom doktorskim, aczkolwiek aktywność publikacyjną Kandydata uplasować należy poniżej przeciętnej. Ponadto pewne wątpliwości budzi dobór czasopisma *Advances in Science and Technology Research Journal* do publikowanego materiału. Jest to czasopismo interdyscyplinarne, ale raczej nie jest skierowane na zagadnienia rozpatrywane w pracy. Dlatego publikacja w tym czasopiśmie może ograniczać dostęp do potencjalnych czytelników. Brak dużej aktywności publikacyjnej Autora można tłumaczyć charakterem dysertacji, który jest praktyczny.

Autor posiadał umiejętność poprawnego, przekonującego i precyzyjnego przedstawiania uzyskanych przez siebie wyników. Zarówno część merytoryczną rozprawy, jak również artykuły cechuje zwięzłość, jasność języka oraz precyzja.

4. Wady rozprawy, słabe strony, uwagi i pytania

Należy podkreślić, że poziom merytoryczny ocenianej dysertacji jest dobry i recenzent nie dostrzegł w niej istotnych wad i niedostatków. Jednak warto również w niej wskazać na pewne, w większości drobne uchybienia oraz niejasności. Uwagi zostały wypunktowane i podzielone na dwie grupy. a) uwagi dyskusyjne i pytania oraz b) drobne uchybienia typograficzne i stylistyczne.

Uwagi dyskusyjne i pytania:

- 1) W Rozdziale czwartym Autor dokonuje analizy ankiety przeprowadzonej spośród specjalistów w zakresie oceny wymagań oprogramowania. W pracy Recenzent nie zauważył szczegółowego opisu procesu przeprowadzenia ankiety, a proces ten może mieć duże znaczenie dla reprezentatywności otrzymanych wyników. Stąd pojawiają się pytania a) jak została dobrana grupa reprezentatywna? b) czy przeprowadzone zostały badania pilotażowe?
- 2) W pracy Autor przyjął współczynnik κ Cohena jako wskaźnik zgodności wyników. Recenzent jednak nie znalazł wyjaśnienia wyboru akurat tego wskaźnika. Stąd pytanie dlaczego Autor zastosował akurat ten wskaźnik, a nie np. *Percent Agreement*, lub *Intraclass Correlation Coefficient*?
- 3) W Rozdziale szóstym został przedstawiony szereg interesujących wyników dotyczących analizy efektywności poszczególnych narzędzi klasyfikacyjnych. Szczególnie interesujące wydają

się wyniki przedstawione w tablicach pomyłek (Tabele 6.2, 6.3, 6.5, 6.6, 6.8, 6.9, 6.11, 6.12). Autor skomentował otrzymane wyniki jedynie ilościowo. A przyjrzenie się uważnie tym wynikom mogą prowadzić do ciekawych wniosków, które wymagałyby potwierdzenia w bardziej szczegółowym badaniu:

a) Zwykle w klasyfikacji z podziałem na trzy klasy, klasa 2 jest w zasadzie agregatem klas 1 i 2 z podziału na cztery klasy. Dla metod *decision tree*, *Random Forest* i *Gradient Boosted Trees* orzynałibyśmy wyniki zgodne w 100%, w przypadku metody *k Nearest Neighbors* otrzymalibyśmy 2/3 różnice.

b) Klasy otrzymane przy użyciu różnych klasyfikatorów wydają się być bardzo do siebie podobne. Nasuwa się pytanie na ile pomyłki klasyfikacji są powtarzalne dla różnych klasyfikatorów? Np. czy pomyłka w 'Predicted 3/True 4' w Tabeli 6.3 to ten sam wynik co 'Predicted 3/True 4' w tabelach 6.9 i 6.13. Dokładna analiza tych wyników pozwoliłoby na potencjalne wyodrębnienie sytuacji, dla których zastosowana w dysertacji metodologia nie jest wiarygodna i nie powinna być stosowana.

- 4) W Tabeli 6.11 jest błąd. Nie zgadza się sumaryczna liczba wyników.
- 5) Autor przeprowadził dokładną analizę literaturową w odniesieniu do narzędzi, które wykorzystał w pracy. Jednak zabrakło w nim odwołań do kilku innych prac, które były poświęcone podobnej tematyce. Wymienić tu można m. in.:
 - a) Ferrari A., et al.: Detecting requirements defects with NLP patterns: an industrial experience in the railway domain. *Empirical Software Engineering* 23(1) (2018), DOI:10.1007/s10664-018-9596-7.
 - b) Verma K., Kass A.: Requirements Analysis Tool: A Tool for Automatically Analyzing Software Requirements Documents. 7th International Semantic Web Conference (ISWC 2008). *Lecture Notes in Artificial Intelligence* 5318, pp.751-763.
- 6) Pozycja [121] w spisie literatury odwołuje się do Wikipedii. Ponieważ źródła Wikipedii nie są recenzowane, nie powinno się do nich odnosić w pracach naukowych.

Drobne uchybienia typograficzne i stylistyczne:

- Strona iv, linia 16. Błąd stylistyczny z zdaniu.
- Punkt 3.2.4 vs. 6.4. Brak w konsekwencji na nazewnictwie '*K Nearest Neighbors*'.
- Strona 46 ostatnia linia. Brak kropki na końcu zdania.
- Strona 47 nad równaniem (4.1). Nr. równania powinien być podawany w nawiasach.
- Strona 53 linia 11. Brak kropki na końcu zdania.
- Strona 72 równanie (6.3). Równanie jest redundantne, ponieważ zostało już wprowadzone we wzorze (4.1).
- Generalna uwaga. Brak w konsekwencji w interpunkcji przy wypunktowywaniu (czasem średniki, czasem bez interpunkcji. czasem przecinki).

Należy podkreślić, że przedstawione powyżej uwagi, mają w większości charakter dyskusyjny i w związku z tym nie obniżają one pozytywnej oceny pracy.

5. Podsumowanie recenzji i wniosek końcowy

Recenzowana rozprawa doktorska stanowi, zdaniem recenzenta, oryginalne rozwiązanie ważnego problemu naukowego oraz wykazuje dużą ogólną wiedzę teoretyczną i aplikacyjną Kandydata w dyscyplinie naukowej *Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne*, a także Jego umiejętność samodzielnego prowadzenia pracy naukowej. Zatem stwierdzam, że **rozprawa mgr inż. Jerzego Kocerki spełnia** warunki określone w ustawie Ustawie o stopniach i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz.U. z 2017 r. poz. 1789) oraz zgodnie z Ustawą z 3 lipca 2018 r. – Przepisy wprowadzające ustawę - Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2018 r. poz. 1669 z późn. zm.) w dziedzinie nauk inżyniersko-technicznych, w dyscyplinie *Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne*.

W związku z powyższym, uwzględniając oryginalność rozwiązania problemu naukowego przedstawionego w rozprawie, specjalistyczną wiedzę Kandydata w dyscyplinie *Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne* oraz umiejętność samodzielnego prowadzenia pracy naukowej **wnoszę o dopuszczenie rozprawy doktorskiej mgr inż. Jerzego Kocerki do publicznej obrony.**