

dr hab. inż. Joanna Domańska
Instytut Informatyki
Teoretycznej i Stosowanej PAN
ul. Bałtycka 5, 44-100 Gliwice

Gliwice, 15 lipca 2023 r.

POLITECHNIKA ŚLĄSKA
Biuro Rady Dyscypliny
Automatyka, Elektronika, Elektrotechnika
i Technologie Kosmiczne
wpłynęło dnia 20.07.2023
nr 24 zat.

Recenzja rozprawy doktorskiej

mgr inż. Jerzego Kocerki

pt. "Automatic quality analysis of textual requirements using natural language processing"

Cel, zakres i charakter rozprawy

Przedstawiona do recenzji rozprawa doktorska została wykonana pod kierunkiem naukowym dr hab. Adama Gałuszki, profesora Politechniki Śląskiej.

Praca doktorska ma wdrożeniowy charakter, co oznacza, że praca badawcza wykonana przez doktoranta ma wymiar nie tylko teoretyczny, ale również praktyczny, wpisując się tym samym w cel wprowadzenia przez Ministerstwo Nauki i Szkolnictwa Wyższego (obecnie Ministerstwo Edukacji i Nauki) programu "Doktorat wdrożeniowy", którym jest intensyfikacja współpracy pomiędzy środowiskiem naukowym a przemysłem. Przeprowadzone w ramach pracy badania i osiągnięte wyniki są zgodne z potrzebami firmy Tritem w zakresie opracowywania innowacyjnych rozwiązań związanych z poprawą procesów testowania.

Rozprawa dotyczy problematyki zastosowania przetwarzania języka naturalnego (NLP) do automatycznej oceny jakości specyfikacji oprogramowania przemysłowego. Badania prowadzone przez doktoranta dotyczą również wpływu jakości tych wymagań na wydajność procesu tworzenia oprogramowania. Istnieją różne metody i narzędzia do analizy jakości wymagań, zazwyczaj oparte na listach zakazanych terminów i wyrażeniach regularnych. Pojawiają się również nowe narzędzia wykorzystujące metody uczenia maszynowego do analizy wymagań, ale narzędzia te skupiają się na przede wszystkim na wykrywaniu pojedynczych wskaźników. W ramach recenzowanej pracy doktorskiej doktorant przedstawia proces tworzenia systemu, który nie tylko analizuje poszczególne metryki wymagań, ale również oblicza ogólny wynik jakości wymagań. Doktorant proponuje metodologię mającą na celu analizę i detekcję niejasnych, nieprecyzyjnych, niepoprawnych lub niewystarczająco opisanych wymagań, zapisanych w języku naturalnym. Zadaniem prezentowanego procesu jest szybka identyfikacja takich wymagań, aby móc dostarczyć odpowiednie informacje zwrotne autorowi wymagania. Walidacja przedstawionego w pracy automatycznego procesu została przeprowadzona przy wykorzystaniu wymagań ocenianych przez ekspertów,

Celem recenzowanej pracy doktorskiej jest poprawa efektywności przemysłowego procesu tworzenia oprogramowania poprzez optymalizację procesu tworzenia wymagań. Efektywność definiowana jest jako odwrotność całkowitego czasu potrzebnego na sformułowanie, analizę i poprawę wymagań, a optymalizacja polega na maksymalizacji

tej funkcji efektywności. Aby osiągnąć ten cel analizowana jest możliwość zapisywania wymagań w sposób sformalizowany oraz możliwość tworzenia narzędzi wspomagających ten proces.

W ramach realizowanego celu pracy autor rozprawy sformułował tezę, która zakłada, że automatyczna analiza jakości wymagań zwiększa efektywność procesu tworzenia i testowania oprogramowania dla systemów wbudowanych w porównaniu do metod stosowanych powszechnie w przemyśle.

Zawartość rozprawy

Forma drukowana recenzowanej rozprawy obejmuje 113 stron. Praca została napisana w języku angielskim i składa się z wprowadzenia, sześciu rozdziałów, podsumowania pracy oraz wykazu literatury. Tytuł odzwierciedla zawartość pracy, a przyjęty układ pracy jest właściwy.

Wprowadzenie - oprócz zakresu i celu rozprawy - zawiera również poprawnie sformułowaną tezę pracy oraz opis wkładu własnego autora. Wprowadzenie zamyka opis struktury pracy. Autor rozprawy numeruje rozdziały wliczając do numeracji Wprowadzenie.

Rozdział drugi opisuje stan wiedzy dotyczący inżynierii wymagań. W rozdziale tym autor zawarł przegląd dotychczasowych podejść do formalizacji i półformalizacji języka stosowanego do opisu wymagań. Autor opisuje różne metody formalnej i półformalnej reprezentacji wymagań, zaczynając od najszerzej stosowanego Języka Modelowania Zunifikowanego (UML). Wyszczególnione zostały zalety metody reprezentacji formalnej, przede wszystkim dostarczenie spójnego i ustandaryzowanego opisu wymagań. Tego typu podejścia nie są powszechnie stosowane przede wszystkim ze względu na czasochłonny proces uczenia notacji. W rozdziale podkreślono zalety używania języka naturalnego, takie jak możliwość skupienia się na kluczowych aspektach projektu, łatwiejsze zrozumienie przez nie-technicznych ludzi oraz łatwiejsze zarządzanie wymaganiami. Rozdział zawiera również opis narzędzi i metod stosowanych do analizy różnych miar jakości wymagań oraz opis standardu ISO/IEC/IEEE 29148 związanego z inżynierią wymagań w cyklu życia zarówno oprogramowania, jak i całych systemów.

W rozdziale trzecim zawarto opis metod analizy specyfikacji wymagań. Rozdział podzielono na dwie części. Pierwsza część dotyczy przetwarzania języka naturalnego (NLP), natomiast druga metod podejmowania decyzji. W ramach części pierwszej przedstawiono dwa główne aspekty NLP: analizę składniową oraz semantyczną. Opisało metodę kodowania wektorów o wysokiej wymiarowości, tzw. "Bloom embeddings", która znalazła zastosowanie w różnych dziedzinach uczenia maszynowego, w tym w modelowaniu języka. Jej zaletą jest przede wszystkim oszczędność zarówno pamięci, jak i czasu obliczeniowego. Opisało również tzw. osadzenia słów czyli reprezentacje wektorowe, jako popularny sposób reprezentowania słów w przetwarzaniu języka naturalnego. W części drugiej omówiono proces podejmowania decyzji, przybliżając takie algorytmy uczenia maszynowego jak: Drzewo decyzyjne, Las losowy, Drzewa wzmacniane gradientowo oraz metodę K najbliższych sąsiadów.

Rozdział czwarty opisuje proces ewaluacji 112 wymagań napisanych w języku naturalnym. Wymagania te były związane z projektem przemysłowym dotyczącym kontroli procesu hamowania pojazdów kolejowych, natomiast ekspertami oceniającymi byli doświadczeni inżynierowie. Zdefiniowano konkretne kryteria oceny wymagań, które obejmują m.in. subiektywne i niejednoznaczne sformułowania, brak odniesień do innych wymagań czy dokumentów, błędy ortograficzne lub gramatyczne. Celem opi-

sanych w tym rozdziale badań był stworzenie punktu odniesienia dla przyszłych ocen automatycznych narzędzi analizy jakości wymagań. Dane zebrane podczas tej oceny posłużą do oceny zdolności tych narzędzi do precyzyjnego wykrywania wymagań, które wymagają poprawy, oraz do prawidłowego klasyfikowania wymagań poprawnych.

Rozdział piąty opisuje automatyczny system oceny jakości wymagań. Struktura wymagań jest w postaci Given-When-Then (Zakładając, Kiedy, Wtedy), co ułatwia ich zrozumienie i sprawdzenie, czy zostały spełnione. Do sprawdzania błędów językowych wymagań został użyty LanguageTool - wielojęzyczne narzędzie do sprawdzania pisowni i gramatyki. W rozdziale wskazano jakie metryki związane z błędami językowymi zostaną użyte do obliczenia ogólnego wskaźnika jakości: liczba błędów ortograficznych, liczba błędów gramatycznych oraz liczba błędów stylistycznych (przecinki, kropki, przerwy w zdaniach itp.). W rozdziale zawarto również opis procesu wykorzystania narzędzi uczenia maszynowego do rozpoznawania nazwanych bytów (Named-entity recognition (NER)) w kontekście specyficznych wymagań dotyczących systemu transportu kolejowego. Zostało to zrealizowane przy wykorzystaniu biblioteki spaCy.

Rozdział szósty zawiera przedstawienie ogólnej oceny jakości wymagań przy wykorzystaniu metryk opisanych w rozdziale piątym. Klasyfikacji dokonano przy użyciu metod ML opisanych w rozdziale czwartym. Wynikiem klasyfikacji jest przypisanie do danego wymagania jednej z czterech grup jakości zgodnie ze skalą, na której przeprowadzono ocenę ekspertów, opisaną w rozdziale czwartym. W rozdziale opisano również zastosowaną metodę walidacji krzyżowej.

Rozdział siódmy dotyczy efektywności procesu tworzenia oprogramowania. Rozdział ten ocenia skuteczność ulepszenia procesu tworzenia wymagań dla oprogramowania wbudowanego poprzez użycie automatycznego narzędzia do oceny ich jakości, w porównaniu do tradycyjnej ręcznej metody przeglądu. Badanie polegało na porównaniu wyników ręcznego przeglądu wymagań oprogramowania z wynikami automatycznej oceny za pomocą narzędzia, ze szczególnym naciskiem na czas potrzebny na analizę i rozważenie sugerowanych poprawek.

Podsumowanie pracy zawiera skrócony opis uzyskanych w ramach pracy wyników wraz z propozycjami dalszych badań oraz zastosowaniem praktycznym. Propozycje przyszłych badań obejmują m.in. rozszerzenie badania o więcej wymagań dotyczących oprogramowania oraz poprawę klasyfikacji poprzez rozszerzenie modelu jakości o nowe metryki.

Bibliografia zawarta w rozprawie składa się z 127 pozycji. Cytowane w tekście rozprawy pozycje oraz analiza ich zawartości potwierdzają dostateczną znajomość stanu wiedzy doktoranta - związanego z tematyką rozprawy.

Ocena rozprawy

W ramach pracy doktorskiej doktorant rozważa możliwość utworzenia automatycznego narzędzia wspierającego proces tworzenia wymagań dla oprogramowania wbudowanego celem poprawy efektywności tego procesu. Autor rozprawy zaproponował rozwiązanie, które polega na automatycznym ocenianiu wymagań zaraz po ich stworzeniu. Jego podejście ma na celu zmniejszenie potrzeby przełączania kontekstu i poprawę efektywności całego procesu. Tworzone w ramach pracy narzędzie ma na celu dostarczenie oceny jakości i wykrywania błędów w wymaganiach, oferując natychmiastowe informacje zwrotne dla ich autorów. Narzędzie składa się z modułów odpowiedzialnych za przetwarzanie tekstu, wykrywanie błędów językowych, analizowanie i wykrywanie

nazwanych obiektów i odniesień oraz wykrywanie słabych i subiektywnych stwierdzeń. Skuteczność narzędzia potwierdzono na zestawie danych z 112 wymagań, ocenianych wcześniej przez ekspertów. Wyniki sugerują, że narzędzie może być wykorzystywane do klasyfikacji wymagań na te, które mogą być później wykorzystane w procesie tworzenia oprogramowania oraz te, które wymagają poprawienia przez autora. Narzędzie wykazało również potencjał do znacznego skrócenia czasu spędzanego na poprawianiu wymagań, co skutkuje skróceniem czasu tworzenia oprogramowania oraz poprawą jego jakości.

Teza pracy jest oryginalna, a uzyskane i opisane w pracy wyniki potwierdzają jej prawdziwość. Jak wspomniano wyżej - zebrane dane z procesu eksperckiej oceny wymagań zostały wykorzystane do porównania wyników uzyskanych z utworzonego w ramach pracy doktorskiej narzędzia do automatycznej analizy jakości wymagań z oceną doświadczonych ekspertów, a więc umożliwiły ocenę przydatności takiego narzędzia - co potwierdza weryfikację tezy rozprawy doktorskiej.

Praca jest napisana w sposób jasny i czytelny, nie zawiera zauważalnych literówek i innych błędów edycyjnych - widać staranność doktoranta w tej kwestii, co stanowi zaletę pracy. Jedynym poważniejszym niedociągnięciem edycyjnym jest według mnie kwestia niepotrzebnego powtarzania tych samych wzorów:

- wzory 2.1 i 4.1 są identyczne - w rozdziale czwartym powinien znaleźć się odnośnik do wzoru wprowadzonego w rozdziale drugim;
- to samo dotyczy wzorów 6.1, 6.2 i 6.3 - powinny w pracy znaleźć się odnośniki do wzorów wprowadzonych w rozdziale drugim i piątym;
- wzór 7.2 wprawdzie był wprowadzony wcześniej, ale poprzedza go prawidłowe odwołanie do rozdziału pierwszego.

Przedstawione w niniejszej recenzji uwagi i spostrzeżenia nie wpływają na pozytywną ocenę merytorycznej części pracy oraz nie podważają dobrej oceny umiejętności autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników.

Oczekuję jednakże od doktoranta ustosunkowania się do poniższych uwag podczas publicznej obrony pracy doktorskiej:

- kryterium oceny "Lack of References" skupia się na braku odniesień do innych wymagań lub dokumentów, ale nie uwzględnia sytuacji, gdy takie odniesienie jest niepoprawne lub błędne. Błędne odniesienie do innego wymagania lub dokumentu może prowadzić do nieporozumień, które mogą wpływać na ogólną jakość wymagań oprogramowania, podobnie jak brak takich odniesień. Z tego powodu, oceniając wymagania, warto również sprawdzić poprawność wszelkich odniesień, czy są one aktualne i czy prowadzą do odpowiedniego dokumentu lub wymagania. Czy było to uwzględnione w pracy, a jeśli nie to dlaczego;
- autor rozprawy wykorzystuje do analizy lingwistycznej tekstu LanguageTool, które jest znanym wielojęzycznym narzędziem do sprawdzania ortografii i gramatyki. Doktorant wspomina na str. 58 rozprawy, że LanguageTool jest oparte na algorytmach AI. Ponieważ LanguageTool bazuje głównie na zdefiniowanych regułach i wzorcach, brakuje w pracy omówienia, do czego są wykorzystywane algorytmy AI w wersji LanguageTool, która była używana w niniejszej pracy;
- brakuje w pracy szerszego uzasadnienia wykorzystania RapidMiner;

- brakuje w pracy omówienia funkcji straty wykorzystanej w algorytmie Drzew wzmacnianych gradientowo (Gradient Boosting Trees).

Wnioski końcowe

Mgr inż. Jerzy Kocerka przedstawił rozprawę doktorską zawierającą oryginalne rozwiązania w zakresie wspierania procesu tworzenia wymagań dla oprogramowania wbudowanego, które skutkują poprawą efektywności tego procesu. Doktorant do rozwiązania problemu użył właściwych metod. Jako, że rozprawa ma charakter wdrożeniowy, ważne jest również praktyczne zastosowanie badania, które polega na integracji z istniejącymi procesami tworzenia wymagań w firmie Tritem.

Rozprawa doktorska dotyczy ważnego problemu oraz stanowi wartościowe osiągnięcie naukowe Autora. Dysertacja doktorska powinna również potwierdzać ogólną teoretyczną wiedzę Doktoranta w danej dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej. Według mojej opinii przedstawiona do recenzji praca spełnia wymienione wymagania.

Biorąc pod uwagę powyższe stwierdzam, że niniejsza rozprawa spełnia wymogi ustawowe stawiane pracom doktorskim w *dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne*, określone w art.13.1 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. 2017 poz. 1789) i wnioskuję o dopuszczenie Autora rozprawy do publicznej obrony.