

## Recenzja rozprawy doktorskiej

mgra inż. Bartosza Wojciecha Dobrzyńskiego  
z tytułu:

Wielowymiarowa eksploracja repozytoriów programowych w zakresie raportów zgłoszeń oraz ich obsługi

Promotor prof. dr hab. inż. Janusz Sosnowski

### 1. Problem badawczy i jego znaczenie

Tematyka rozprawy mgra inż. Bartosza Wojciecha Dobrzyńskiego dotyczy wielowymiarowej analizy i eksploracji danych rejestrowanych w repozytoriach programowych dostarczanych przez narzędzia klasy ITS (Issue Tracking Systems) dostępne w takich platformach jak JIRA, GitHub, czy Bugzilla lub Redmine. Narzędzia tego typu wspierają cały cykl życia projektu, więc mogą dotyczyć różnych aspektów inżynierii oprogramowania od tworzenia serwisu internetowego i raportowania błędów poprzez wspomaganie zarządzania projektami i nadzorowanie pracy zespołów aż do monitoringu procesów w całych organizacjach. Takie narzędzia są użyteczne i wręcz konieczne w czasach szybko zmieniającej się rzeczywistości biznesowej i błyskawicznego dostępu do informacji, które wymuszają konieczność podejmowania szybkich decyzji. Dlatego ich znaczenie stale wzrasta i stają się jednym z głównych elementów procesu wytwarzania.

Recenzowana rozprawa koncentruje się na określeniu aktywności projektantów oraz ich zaangażowaniu w proces wytwarzania, a także na wyznaczeniu korelacji pomiędzy atrybutami zgłoszenia i modyfikacjami tworzonego kodu systemu czy aplikacji. Doktorant rozpatruje statystyki charakterystyczne dla dostarczanych zgłoszeń oraz zajmuje się wielowymiarową analizą procesu obsługi zgłoszeń w celu zapewnienia wiarygodnego wyszukiwania możliwych anomalii. Poza tym stara się wskazać potencjalne kierunki optymalizacji procesu raportowania i obsługi zgłoszeń, by zminimalizować nakłady przeprowadzenia takiej oceny w wybranych etapach życia wytwarzanej aplikacji. Tego typu działania mają istotne znaczenie dla zapewnienia niezawodnego i ekonomicznego projektowania i testowania systemów i aplikacji informatycznych.

Po krótkim wprowadzeniu do analizowanej problematyki Doktorant formułuje tezę rozprawy doktorskiej oraz przedstawia konieczne zadania badawcze, których wyniki powinny potwierdzić jej zasadność. Podjęte przez Niego zadania dotyczą opisu i eksploracji zawartości informacyjnej repozytoriów zgłoszeń, zaproponowanie modelu grafowej obsługi zgłoszeń (IHG – Issue Handling Graph) oraz metod analizy zarejestrowanych w takich repozytoriach wielkości parametrów i komunikatów (tekstów). Ekstrakcja gromadzonych danych oraz przeprowadzone analizy dla tych danych wspierane są przez zbudowane przez Autora rozprawy narzędzie IssueAnalyzerTool oraz skrypty pomocnicze, które wywołują opracowane przez Autora rozprawy algorytmy takich analiz. W rozprawie wykorzystano też nowe modele klasyfikacji oraz analizy semantycznej, które wymagały odpowiedniej transformacji opisów tekstowych zawartych w zgłoszeniach.

Sposób przeprowadzania badań jak i zastosowana metoda analizy zgromadzonych danych w repozytoriach nie budzą zastrzeżeń. W rozprawie doktorskiej wykorzystano odpowiedni aparat matematyczny związany z eksploracją danych, analizą tekstów, teorią grafów oraz budową algorytmów. Nie mam więc wątpliwości, że recenzowana rozprawa ma charakter naukowy, a jej wyniki są wartościowe, bo przede wszystkim odnoszą się do praktycznych zastosowań. Należy też podkreślić, że zakres tematyczny rozprawy mieści się w dyscyplinie Informatyka Techniczna i Telekomunikacja.

## 2. Osiągnięcia badawcze

W procesie wytwarzania oprogramowania istotną rolę odgrywa informacja o jego przebiegu oraz o postępie w realizacji kolejnych zadań. Źródłem takich informacji mogą być różnego typu zgłoszenia generowane przez członków zespołów projektowych (reporterów) i umieszczane w repozytoriach zgłoszeń. Format i treść tych zgłoszeń może być różnorodne w zależności od śledzonych parametrów procesu wytwarzania oraz od typu zgłaszanych sytuacji, a nawet od wykorzystywanych narzędzi wspomagających tego typu czynności. Tego typu zgłoszenia są przechowywane w tzw. repozytoriach zgłoszeń i wymagają odpowiedniej reakcji zespołu projektowego. Podejmowane reakcje również są rejestrowane i jest tworzona historia takich wszystkich działań. W ten sposób są gromadzone interesujące dane, które mogą być wykorzystane do analizy procesu wytwarzania aplikacji informatycznej. Z tego punktu widzenia istotna jest zakres rejestrowanych informacji jak i sam proces obsługi zgłoszeń, by wyłuskać i przeanalizować istotne dane mające wpływ na przebieg realizacji projektu. Rozważania Doktoranta dotyczyły właśnie analizy wybranych danych zawartych w repozytoriach zgłoszeń, jak też oceny jakości metody obsługi tych zgłoszeń. Wyjściem do Jego rozważań były praktyczne systemy wspomagające tego typu działania, a także ocena ich możliwości, a w dalszej perspektywie rozszerzenie ich funkcjonalności w celu gromadzenia bardziej adekwatnych danych i skutecznej ich analizy dotyczącej jakości procesu wytwarzania aplikacji i systemów informatycznych.

Do osiągnięć Doktoranta w obszarze analizy danych zawartych w repozytoriach oraz oceny metod obsługi zgłoszeń należy zaliczyć:

1. Opracowanie profili statystycznych zawartości repozytoriów mogących ujawniać niedoskonałości raportowania projektu, a także profili aktywności uczestników (aktorów) projektu tzn. wyróżnić członków zespołu o mniejszym zaangażowaniu w projekcie, bądź jak najbardziej aktywnych jego członków.
2. Przebadanie wpływu eksploracji opisów oraz komentarzy zgłoszeń, a w szczególności różnych konfiguracji klasyfikatorów oraz cech danych wejściowych na dokładność dokonywanej klasyfikacji (algorytm klasyfikacji - Alg. 4), co stanowi wyznacznik jakości opisów zgłoszeń w repozytorium.
3. Rozwinięcie metodyki wielowymiarowej analizy obsługi zgłoszeń umożliwiającej przeprowadzenie dokładnej ewaluacji procesu obsługi zgłoszeń w różnych perspektywach, np. ogólnej – uwzględniającej wszystkie typy zgłoszeń lub szczegółowej – koncentrując się na konkretnych, takich jak typ zgłoszeń, grupy reporterów czy priorytet zgłoszeń. Proponowana metodyka postępowania bazuje na wprowadzonym grafowym modelu IHG oraz zestawie oryginalnych algorytmów, takich jak budowanie tego grafu (Alg. 7), filtracja wierzchołków i krawędzi grafu (Alg. 8, Alg. 9), a także generowania pliku stanów unikalnych ścieżek oraz wyszukiwania pętli (Alg.10 - Alg. 12). Dzięki tym algorytmom jest możliwe wyliczanie profili wydajnościowych, czasowych i strukturalnych, a także wyznaczenie możliwych anomalii. Pozwala to również na przeprowadzenie porównań, np. pomiędzy procesami obsługi dla



różnych priorytetów zgłoszeń czy różnych typów zgłoszeń, a nawet fragmentów realizowanych projektów.

4. Implementacja opracowanych algorytmów jako narzędzia IssueAnalyzerTool wykorzystanego do analizy wybranych sytuacji pojawiających się w procesie wytwarzania oprogramowania. Pozyskana w ten sposób wiedza może też zostać wykorzystana do optymalizacji prac związanych z obsługą zgłoszeń oraz poprawą konfiguracji systemu raportowania zgłoszeń (ITS). W szczególności umożliwia określanie „wąskich gardeł, zidentyfikowanie trudno dostępnych stanów oraz przejść między nimi, jak również wskazanie stanów o wysokim czasie obsługi, ścieżek z wieloma stanami i dużą liczbą komentarzy czy pętli stanów ścieżek. Co więcej świadomość tych anomalii daje szansę, w uzgodnieniu z całym zespołem projektowym, na podjęcie działań zaradczych minimalizujących wpływ takich sytuacji na przebieg procesu wytwarzania.

Są to istotne osiągnięcia badawcze Doktoranta, zrealizowane na bazie dostępnych praktycznych danych eksperymentalnych. Dane te pochodzą z wybranych projektów, których realizacja była raportowana przy wykorzystaniu narzędzi klasy ITS – JIRA. Dzięki tym narzędziom wyniki procesu gromadzenia danych udokumentowane zostały w odpowiednich repozytoriach. Jednak dostęp do nich jak i ich interpretacja nie jest łatwa. Zależy od kontekstu ich rejestracji (typu czy miejsca zgłoszenia), a także od wykorzystania indywidualnego żargonu, czy nieznanymi skrótów w polach tekstowych, a także popełniania typowych ludzkich błędów. Poza tym istnieją też techniczne (wydajnościowe czy funkcjonalne) ograniczenia wykorzystywanych narzędzi co ogranicza zakres gromadzonych danych. Zatem konieczne było krytyczne spojrzenie Doktoranta na sposób gromadzenia takich danych oraz interpretację ich znaczenia, jak też sprawdzenie ich przydatności w całościowej i szczegółowej ocenie realizowanych projektów.

Doktorant przeanalizował zagregowane dane literaturowe dotyczące projektów open source, takich jak: Apache Casandra (cas), Apache Spark (spark), Apache Flink (flink), Apache Ignite (ignite), Mozilla Tunderbird (moz), Red Hat Enterprise Linux 8 (red), a także projektu P1, który jest projektem komercyjnym zrealizowanym z istotnym udziałem Doktoranta. Porównywał rozkłady wartości kategorii i typów atrybutów zgłoszeń, analizował korelacje między atrybutami i typami zgłoszeń, rozpatrywał zapewnienie optymalnej konfiguracji narzędzia, dokonał analizy aktywności użytkowników repozytoriów oraz ewaluację procesu obsługi zgłoszeń, w tym ocenę długu błędów (błędów istotnych, ale odłożonych do późniejszej analizy). Różne spostrzeżenia odniósł do konkretnych projektów co jest istotną wartością badawczą z punktu widzenia inżynierii oprogramowania.

Przedstawione wyniki tego typu analiz pośrednio odnoszą się i potwierdzają tezę rozprawy doktorskiej zdefiniowanej następująco: *Analiza i ocena procesu wytwarzania oprogramowania wymaga opracowania reprezentatywnych modeli oraz metod eksploracji danych z repozytoriów zgłoszeń i repozytoriów kodu. Uwzględnienie różnych poziomów ekstrakcji i agregacji informacji oraz perspektyw obserwacji, poszerza zakres przedmiotowy monitorowania projektu i ułatwia identyfikację niedoskonałości.* Pewne uwagi krytyczne dotyczące zakresu stosowalności tej tezy zostały podane w następnym rozdziale.

### 3. Uwagi krytyczne

Zauważalnym mankamentem rozprawy doktorskiej jest ograniczenie się jej Autora do wykorzystania możliwości istniejących narzędzi klasy ITS, bez próby zdefiniowania uogólnionego modelu procesu tworzenia, rejestracji i obsługi zgłoszeń związanych z problemami realizacji typowego projektu informatycznego. To umożliwiłoby porównanie zakresu badań zawartych w rozprawie z



niezbędnym zakresem prac możliwych do wykonania. Cenne natomiast jest wykorzystanie modelu grafowego, umożliwiającego takie uogólnione podejście w przypadku analizy obsługi zgłoszeń.

Poza tym w rozprawie zdawkowo potraktowano opis analizowanych projektów jak i szczegóły implementacyjne narzędzia przeznaczonego do tworzenia repozytorium danych. Szersza charakterystyka analizowanych projektów oraz podanie szczegółów implementacyjnych proponowanego narzędzia wskazywałyby na możliwości dalszego rozwoju repozytorium zgłoszeń oraz dokładną ocenę możliwości i kompletności zastosowanej metody analizy.

Dla większej czytelności rozprawy doktorskiej przydałby się też podanie koncepcji i założeń dotyczących badań eksperymentalnych oraz metod analizy uzyskanych wyników. Bez próby dokonania takich uogólnień Doktorant w zasadzie koncentruje się na konkretnej analizie wybranych parametrów (ale czy najważniejszych?) opisujących niektóre aspekty procesu wytwórczego. Do takiej analizy zorientowanej na pojedyncze projekty, dostosowuje metody oceny bez uwzględnienia innych aspektów inżynierii oprogramowania również istotnych z punktu rejestracji nieodpowiednich zdarzeń czy podejrzanych sytuacji. Co prawda przy zaprezentowaniu algorytmów Doktorant skoncentrował się na konkretnej i trafnie dobranej strukturze danych, ale taka struktura może nie obejmować wszystkich możliwych przypadków. Konieczne byłoby więc potwierdzenie istnienia lub choćby zasugerowanie potrzeby opracowania standardu opisu zgłoszeń i danych zawartych w repozytoriach dotyczących procesu śledzenia i raportowania realizowanych projektów, co w przyszłości ułatwiłoby szerszą ich analizę. Z drugiej jednak strony zakres gromadzenia różnego typu zgłoszeń może być bardzo obszerny, podobnie różnorodność wymaganych metod do analizy wyników zawartych w repozytorium zgłoszeń, co wymagałoby znacznego wysiłku, przewyższającego zrealizowanie jednej rozprawy doktorskiej. Jednak brak dyskusji na temat kompletności repozytorium danych oraz oceny zakresu funkcjonalności systemów wspomagania procesów śledzenia czynności projektowych jest pewnym mankamentem tej rozprawy. Poza tym zauważa się brak w rozprawie wykazu najważniejszych oznaczeń wraz z krótkim ich opisem, co także zmniejsza czytelność recenzowanej rozprawy.

Jak wspomniano powyżej, czego Doktorant też jest świadomy, w praktyce inżynierii oprogramowania może pojawić się znacznie więcej różnego typu sytuacji istotnych dla przebiegu procesu wytwarzania aplikacji informatycznej niż te które uwzględniono w ocenianej rozprawie doktorskiej. Przykładem mogą być choćby takie przypadki jak: niskie kompetencje uczestników zespołów projektowych, ograniczone możliwości dostępnych narzędzi projektowania bądź nieadekwatny wybór metod wytwarzania czy zarządzania procesem projektowania, a także pojawiające się niesprzyjające okoliczności utrudniające działania. Mają one również istotny wpływ na przebieg i efekty realizowanego projektu. Jeśli więc zakres monitorowania występujących sytuacji może być znacznie szerszy niż ten, który został przedstawiony w ocenianej rozprawie doktorskiej, to ogólnie sformułowana teza rozprawy wydaje się być zbyt daleko idąca, powinna jedynie odnosić się do przypadków rozpatrzonych w tej rozprawie?

Co więcej w rozprawie nie podano jakie metodyki wytwarzania były wykorzystywane przy realizacji i analizie projektów informatycznych. Wspomniano jedynie o metodyce SCRUM, dla której zaprezentowane metody analizy repozytorium zgłoszeń dobrze sprawdziły się przy unikalnej ocenie efektywności sprintów. Interesujące byłoby odniesienie się do innych metodyk projektowania wykorzystujących również metody śledzenia i raportowania procesu wytwarzania aplikacji, zwłaszcza do takich jak DevOps czy ostatnio bardzo popularnej metody DevOps+AI, które to najczęściej są stosowane w środowiskach chmury obliczeniowej.

Na zakończenie opisu uwag krytycznych chciałbym jednak podkreślić, że mają one charakter dyskusyjny i nie przekreślają wartości i znaczenia badań i wyników szczegółowych zawartych w recenzowanej rozprawie doktorskiej.

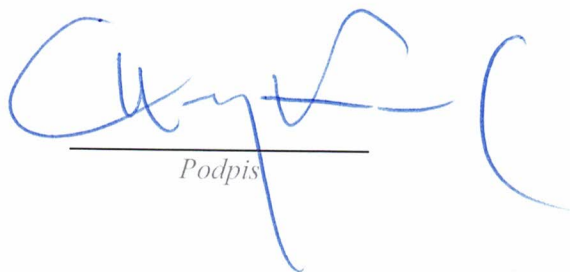
#### 4. Podsumowanie

Doktorant wykazał się dużymi umiejętnościami budowy i wykorzystania narzędzi śledzenia i raportowania procesów wytwarzania projektów informatycznych. W kilku rozdziałach swojej rozprawy odniósł się też do teoretycznych i praktycznych osiągnięć światowych dotyczących rozpatrywanej problematyki. To potwierdza dobry stan wiedzy Doktoranta w tym zakresie. Na podkreślenie zasługuje również właściwie dobrana literatura – 108 pozycji, w tym 4 z nich zostały przygotowane przez Doktoranta (współautorstwo, pozycje 10, 11, 14, 40). Najistotniejsza jego publikacja: *Analysing problem handling schemes in software projects. Information and Software Technology*, 2017.

Reasumując stwierdzam, co następuje:

1. Doktorant rozpatrzył bardzo istotny, ale zarazem bardzo trudny i złożony problem naukowy jakim jest usprawnienie śledzenia i raportowania procesu realizacji projektów informatycznych,
2. Opracował oryginalne i przydatne algorytmy opisu i analizy zarejestrowanych danych liczbowych i testowych gromadzonych w tzw. raportach zgłoszeń,
3. Dokonał oceny kilku reprezentatywnych przypadków (user cases) na podstawie dostępnych (open data) i własnych (komercyjnych) danych oraz zaproponował rozsądne usprawnienia procesu śledzenia i raportowania realizacji projektów informatycznych.

Biorąc powyższe pod uwagę oraz uwzględniając wymagania zdefiniowane przez odpowiednią Ustawę o stopniach naukowych i tytule naukowym, stwierdzam, że moja ocena rozprawy jest zdecydowanie pozytywna i proponuję dopuszczenie mgr inż. Bartosza Wojciecha Dobrzyńskiego do dalszych etapów przewodu doktorskiego.



Podpis

