

Wpł. RDITT
26.04.2023
M. Skowron

Gdańsk, 18 kwietnia 2023

Dr hab. inż. Rafał Leszczyna
Politechnika Gdańska
Wydział Zarządzania i Ekonomii
rle@zie.pg.edu.pl

Recenzja rozprawy doktorskiej

mgr inż. Mariusza Duki

pt. „Wyznaczenie rankingu stron WWW algorytmem ISOWQ Rank”

Promotor dr hab. Artur Strzelecki, Uniwersytet Ekonomiczny w
Katowicach

1. Wprowadzenie

Niniejsza recenzja rozprawy doktorskiej, której autorem jest mgr inż. Mariusz Duka, została wykonana w oparciu o uchwałę Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej (Uchwała nr 15/2023) z dnia 14 lutego 2023 r. podpisaną przez Przewodniczącego Rady prof. dr hab. inż. Andrzeja Polańskiego. Uchwała ta wskazuje mnie jako recenzenta w komisji doktorskiej w postępowaniu w sprawie nadania stopnia doktora mgr. inż. Mariuszowi Duce.

Promotorem niniejszej rozprawy jest dr hab. Artur Strzelecki, prof. Uniwersytetu Ekonomicznego w Katowicach.

2. Podstawowe dane o kandydacie

2.1 *Data uzyskania tytułu magistra oraz nazwa jednostki organizacyjnej, w której ten tytuł został nadany*

30 czerwca 2018, Wyższa Szkoła Informatyki i Umiejętności w Łodzi, Wydział Informatyki i Zarządzania

2.2 Informacja, czy kandydat ubiegał się uprzednio o nadanie stopnia doktora, w tym - o ile wynika to z dokumentacji sprawy - informacja o przebiegu i zakończeniu wcześniejszego postępowania

Kandydat nie ubiegał się uprzednio o nadanie stopnia doktora.

2.3 Przebieg pracy naukowo-zawodowej (miejsce pracy, zajmowane stanowiska)

Od grudnia 2001, mgr inż. Mariusz Duka zajmuje stanowisko prezesa firmy Interface Poland Sp. z o.o. z siedzibą w Chorzowie. Firma świadczy m.in. usługi pozycjonowania stron internetowych.

Od września 2017, jest prezesem stowarzyszenia na Rzecz Wspierania Relacji Polsko – Łotewskich „Silesia” oraz od roku 2020 – szkoły programowania „Young Coders” z siedzibą w Rydze.

Jest autorem trzech specjalistycznych książek dotyczących programowania w PHP i SQL oraz minikomputera Onion Omega2, opublikowanych przez wydawnictwo Helion.

3. Informacje o ocenianej rozprawie doktorskiej

3.1 Tytuł rozprawy doktorskiej stanowiącej podstawę ubiegania się w aktualnym postępowaniu o nadanie stopnia doktora

Przedstawiona rozprawa doktorska nosi tytuł „Wyznaczenie rankingu stron WWW algorytmem ISOWQ Rank”.

3.2 Cele i tezy pracy kandydata

Cele pracy kandydata oraz jej główne cele zostały przedstawione w dedykowanym podrozdziale Wprowadzenia pracy. Cele pracy zostały sformułowane jako:

1. badania związane z opracowaniem i oceną skuteczności algorytmu rankingowego ISOWQ Rank, potwierdzeniem, czy istnieje dodatnia korelacja pomiędzy algorytmami ISOWQ Rank i MOZ oraz wykazanie, jaki wpływ na tę korelację mają poszczególne czynniki rankingowe, a w szczególności treść i struktura tekstu na stronie WWW;
2. implementacja algorytmu ISOWQ Rank i jego praktyczne zastosowanie.

Natomiast tezę pracy określono w sposób następujący:

„Algorytm ISOWQ Rank w sposób optymalny wyznacza ranking stron WWW przez nadanie im określonej wartości, oznaczającej ich jakość.”

3.3 Ocena układu rozprawy doktorskiej, w tym informacje o jej poszczególnych częściach składowych

Układ i zawartość rozdziałów merytorycznych jest zasadniczo poprawna. Całość rozprawy napisana jest w języku polskim.

Rozprawa doktorska składa się z listy używanych skrótów, Wprowadzenia, czterech głównych rozdziałów, Zakończenia, listy pozycji bibliograficznych, spisu ilustracji oraz list pseudokodów i kodów źródłowych.

Zarys problematyki badawczej podejmowanej przez Doktoranta przedstawiony jest w podrozdziale pt. „Motywacja badań i uzasadnienie wyboru tematu”. Autor wskazuje tam, że główną motywacją badań było opracowanie systemu umożliwiającego bezpłatną analizę techniczną i ocenę jakości strony WWW w czasie rzeczywistym, jak również próba odkrycia czynników rankingowych wpływających na ranking MOZ.

Rozdział pierwszy poświęcony jest problematyce optymalizacji dla wyszukiwarek internetowych (ang. *search engine optimization* – SEO), w tym elementom historii wyszukiwarek, rodzajom wyników wyszukiwania stron internetowych przedstawianych przez wyszukiwarki, głównym czynnościom wykonywanych w ramach SEO oraz wybranym narzędziom SEO, takich jak Google Search Console, Ahrefs, czy Semrush.

W rozdziale drugim przedstawiono podstawy działania algorytmów rankingowych, opisano wybrane prace badawcze w kierunku ustalania parametrów pozycjonowania stron internetowych oraz przedstawiono najbardziej popularne narzędzia optymalizacji dla wyszukiwarek, w tym Google PageRank, HITS, MOZ, czy Ahrefs.

Rozdział trzeci wyjaśnia sposób działania algorytmu ISOWQ, w tym uwzględniane parametry stron internetowych. Autor opisał również sposób implementacji systemu ISOWQ w języku PHP, jego architekturę, projekt bazy danych czy używane technologie. W oddzielnym podrozdziale przedstawiono rodzaje wyników dostarczanych przez system.

Rozdział czwarty poświęcony jest analizie porównawczej algorytmu ISOWQ Rank i MOZ. Opisano wybór próby wraz z kryterium wyboru. Omówiono wyniki rankingowe uzyskane dla stron należących do próby. Przeprowadzono analizę korelacji między rezultatami oceny stron za pomocą algorytmu MOZ oraz różnych konfiguracji parametrów algorytmu ISOWQ Rank (punktacją za wykorzystane technologie i pozycje rankingowe (PM), punktacją za optymalizację kodu źródłowego (PK), punktacją za treść i strukturę tekstu (PT), czy średnią arytmetyczną z PK i PT). Przedstawiono wnioski płynące z analizy.

W Zakończeniu, Autor przytacza główne obserwacje z pracy.

3.4 Zastosowane metody badawcze

Badania opisane w rozprawie doktorskiej można zakwalifikować do obszaru badań stosowanych. Stosując metody heurystyczne oraz analizę literatury autor identyfikuje parametry mające wpływ na pozycję strony internetowej w rankingach. Na tej podstawie zostaje opracowany algorytm, który określa zagregowaną wartość rankingową strony. Algorytm został zaimplementowany oraz w ramach eksperymentów poddany ocenie porównawczej z powszechnie stosowanym rozwiązaniem alternatywnym (MOZ).

3.5 Zastosowane piśmiennictwo

Bibliografia zawiera 227 pozycje. Są to zarówno artykuły i dokumenty opublikowane w ciągu ostatnich kilku (5-10) lat, jak i istotne starsze publikacje z obszarów poruszanych przez Autora. Sposób zastosowania piśmiennictwa (jak źródła wiedzy dla przytaczanych faktów) świadczy o dobrej znajomości tematyki poruszanej w pracy. Jednocześnie brakuje przedstawienia pogłębionej i systematycznej analizy literatury naukowej mającej na celu zidentyfikowanie luk i problemów badawczych. W efekcie nie uwzględniono niektórych badań, mogących mieć niepomijalne znaczenie dla problematyki poruszanej w pracy. Na przykład Giomelakisa i Veglisa pt. "Investigating Search Engine Optimization Factors in Media Websites" (Digital Journalism, 4:3, 379-400, DOI: 10.1080/21670811.2015.1046992), czy Sankpala i Patila „Rider-Rank Algorithm-Based Feature Extraction for Re-ranking the Webpages in the Search Engine” (The Computer Journal, Volume 63, Issue 10, October 2020, Pages 1479–1489, <https://doi.org/10.1093/comjnl/bxaa032>).

3.6 Stopień praktycznego zastosowania uzyskanych wyników badań

Przedstawiony w rozprawie system ISOWQ jest stosowany w praktyce w ramach projektu *ISOWQ Project* od ponad 10 lat. Za jego pomocą wykonano prawie półtora miliona analiz stron oraz przeanalizowano blisko 27 milionów witryn. Zebrane dane udostępnione są bezpłatnie na stronie internetowej projektu (www.isowq.org).

Jak napisano na stronie internetowej systemu, jego głównym celem jest wspieranie rozwoju Internetu przez podwyższanie jakości stron internetowych, w tym ich zgodności z faktami i wiarygodności. Najważniejszymi odbiorcami systemu są webmasterzy, specjaliści IT, programiści i właściciele serwisów internetowych, agencje marketingowe, czy środowiska badawcze.

Tak wysoki stopień praktycznego zastosowania opracowanego algorytmu zasługuje na pozytywną ocenę.

3.7 Stopień oryginalności rozwiązania problemu naukowego

Przedstawione rozwiązanie zawiera elementy, które cechują się oryginalnością na tle przedstawionej w pracy literatury naukowej i mogą świadczyć także o jego szerszej

oryginalności. Należy do nich opracowanie algorytmu rankingowego stron internetowych uwzględniającego bogatą gamę parametrów, w tym związanych z optymalizacją kodu źródłowego strony internetowej oraz zawartością treściową i strukturą tekstu strony. Także ocena strony za technologie i pozycje rankingowe rozszerza wskaźniki uzyskane z dwóch serwisów rankingowych o szereg dodatkowych. Wyniki cząstkowe ustalane są w określonych przez Doktoranta skalach, a całość agregowana do jednego wyniku rankingowego na podstawie autorskiej formuły (równanie (5) na stronie 51).

Przedstawione elementy wykazują potencjał oryginalności, który jednak nie został w pracy dostatecznie udokumentowany. Brakuje bezpośredniego odniesienia rozwiązania do innych prac naukowych i wskazania, które luki zostały wypełnione.

Na przykład jak parametry zidentyfikowane przez Autora odnoszą się do pracy Ziakis'a et al. pt. „Important Factors for Improving Google Search Rank” [89], Giomelakisa i Veglisa pt. Investigating Search Engine Optimization Factors in Media Websites” (Digital Journalism, 4:3, 379-400, DOI: 10.1080/21670811.2015.1046992), czy Sankpala i Patila „Rider-Rank Algorithm-Based Feature Extraction for Re-ranking the Webpages in the Search Engine” (The Computer Journal, Volume 63, Issue 10, October 2020, Pages 1479–1489, <https://doi.org/10.1093/comjnl/bxaa032>).

Także sposób ustalania równań, wag i skal w algorytmie nie został ujawniony w pracy.

3.8 Ocena części rozprawy doktorskiej dotyczącej omówienia wyników badań

Wyniki analizy porównawczej algorytmu ISOWQ Rank i MOZ zostały omówione w rozdziale czwartym rozprawy doktorskiej. Bazując na stosunkowo niewielkiej próbie statystycznej składającej się z 9-10 (w zależności od grupy testowej) stron internetowych dokonano oceny stron, a następnie obliczono współczynnik korelacji dla różnych konfiguracji wartości oceny (wspomnianych w rozdziale 3.3 niniejszej recenzji). Na podstawie wyników uzyskanych dla algorytmu MOZ i punktacji za treść i strukturę tekstu (PT), Autor wysnuwa wniosek, że czynniki wpływające na ranking MOZ mogą być podobne do czynników rankingu ISOWQ Rank. Ta część pracy odnosi się do pierwszego celu pracy Kandydata.

Zastanawia fakt doboru tak niewielkiej próby. Szczególnie w zestawieniu z informacją o milionach stron przeanalizowanych przez system ISOWQ w ciągu lat jego działania. Także kryterium doboru próby tj. strony reprezentujące różną tematykę i zawierające co najmniej 10 podstron, nie zostało uzasadnione. Kwestie te wymagałyby szerszego opisanie w pracy.

Brakuje też odniesienia się do tezy pracy, która, jak wskazano we wcześniejszym podrozdziale niniejszej recenzji wiązała się z pojęciem optymalności wyznaczania rankingu. Pojęcie to nie zostało zdefiniowane w pracy.

Drugi cel pracy, związany z implementacją algorytmu ISOWQ Rank i jego praktycznym zastosowaniem został całkowicie spełniony. Opisane rozwiązanie praktyczne zostało opisane dość szczegółowo w trzecim rozdziale rozprawy doktorskiej. Ponadto w rozprawie umieszczono ponad 40 listingów pseudokodu oraz 11 kodu źródłowego. Jak już wskazano w niniejszej recenzji, ta część pracy cechuje się znaczącym elementem praktycznym i przydatnością dla sektora gospodarczego. Zawiera też potencjał oryginalności, który należałoby jednak lepiej udokumentować. Z pewnością wartościowa byłaby szersza weryfikacja zaproponowanego rozwiązania, na przykład pod względem jego efektywności, czy innych parametrów jakościowych.

3.9 Stopień reprezentacji ogólnej wiedzy teoretycznej kandydata w dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej

Jak wspomniano wcześniej Autor wykazał dobrą znajomości tematyki poruszanej w pracy. Zaprezentowane wyniki twórcze i naukowe o istotnym walorze praktycznym potwierdzają umiejętność samodzielnego prowadzenia pracy naukowej o charakterze wdrożeniowym. Wyniki te zostały również przedstawione w artykule naukowym opublikowanym w czasopiśmie naukowym ujętym w wykazie ministerialnym.

4. Podsumowanie

Przedstawiona do recenzji rozprawa doktorska spełnia ustawowe wymagania stawiane rozprawom doktorskim.

Na tej podstawie wnioskuję o jej dopuszczenie do publicznej obrony w postępowaniu w sprawie nadania stopnia doktora.

Rafał Leszczyński
