



AGH

AGH UNIVERSITY OF KRAKÓW
FACULTY OF ELECTRICAL ENGINEERING, AUTOMATICS, COMPUTER SCIENCE AND BIOMEDICAL ENGINEERING
INSTITUTE OF APPLIED COMPUTER SCIENCE

Kraków 20 stycznia 2024

Dr. hab. inż. Jarosław Wąs, prof. AGH
Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Recenzja rozprawy doktorskiej

Recenzję pracy doktorskiej mgr inż. Anh Nguyena zatytułowanej „*Mining sequence and inter-sequence patterns in large databases*”, opracowano zgodnie z decyzją Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Wrocławskiej – zgodnie z pismem z dnia 27 listopada 2024.

Tematyka rozprawy i tytuł pracy

Tematyka rozprawy doktorskiej mgr inż. Anh Nguyena zatytułowanej „*Mining sequence and inter-sequence patterns in large databases*” dotyczy problematyki eksploracji wzorców sekwencji i inter-sekwencji w dużych bazach danych.

Praca obejmuje analizę sekwencji i inter-sekwencji w kontekście architektury i zastosowań dużych baz danych, a także propozycję własnych rozwiązań algorytmicznych i zaproponowanych metod w zakresie inter-sekwencji.

W mojej ocenie temat pracy poprawnie koresponduje z zawartymi treściami.

Charakterystyka rozprawy i ocena układu rozprawy

Praca doktorska mgr inż. Anh Nguyena składa się łącznie z pięciu rozdziałów, obejmujących: wstęp, zasadniczą część pracy, podsumowanie oraz bibliografię. Tekst pracy zajmuje 99 stron wraz z bibliografią.

AGH University of Science and Technology ,
Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
Institute of Applied Computer Science
A. Mickiewicza 30, 30-059 Kraków, Poland

WPLYNEŁO
29-01-2024

RDN-IIT/81/2024

Pierwszy rozdział pracy zatytułowany *Wstęp* poświęcony jest określeniu celów ogólnych i szczegółowych pracy, wskazaniu problemu naukowego oraz określeniu metodyki pracy z uwzględnieniem: badań literaturowych, analizy istniejących rozwiązań i opracowaniu algorytmu przewartościowania krawędzi.

Rozdział 1 zatytułowany *Introduction* przedstawia wprowadzenie do tematu, a także motywację, przedstawienie problemu oraz cel pracy.

Z kolei w rozdziale 2 *Literature Reviews* przedstawiono bieżący stan wiedzy na temat rozpatrywanego problemu z uwzględnieniem klasycznej i najnowszej literatury.

Rozdział 3 *Mining Inter-Sequence Patterns Based on Pseudo-IDList and Early Pruning Techniques* opisuje istotę problemu inter-sekwencji, przedstawia matematyczne podstawy baz i wprowadza strukturę nazwaną pseudo-IDList oraz algorytm ISP-PI oraz ISP-IC.

Rozdział 4 *methods for Mining Inter-Sequence Patterns with Constraints* przedstawia rozwiązania algorytmiczne przygotowane przez Doktoranta nazwane DBV-ISPMIC oraz DBV_ISPMIC-IMPROVING i jego zrównoleżoną wersją pDBV_ISPMIC-IMPROVING wraz z jego weryfikacją.

Ostatni z rozdziałów tzn. rozdział 5 *Conclusions and Future Works* stanowi podsumowanie odnoszące się do uzyskanych wyników. Odniesiono się do realizacji celów ogólnych i szczegółowych pracy, które zostały osiągnięte, a także przedstawiono potencjalne przyszłe prace.

W mojej ocenie układ redakcyjny pracy jest poprawny. Autor posługuje się poprawną terminologią w zakresie obszaru baz danych i, patrząc szerzej, stosowanych metod naukowych.

Należy stwierdzić, że oceniana praca jest napisana generalnie poprawnym językiem, chociaż trafiają się w niej drobne uchybienia stylistyczne. Bibliografia jest generalnie aktualna. Odwołania do pozycji literaturowych są prawidłowe.

Tytuły i kolejność rozdziałów jest również prawidłowa, zaś treść rozdziałów odpowiada wymaganiom stawianym dla prac doktorskich.

- **ocena zastosowanego piśmiennictwa**

Bibliografia składa się z sześćdziesięciu kilku pozycji związanych z tematyką pracy obejmujących zarówno podstawy tematu i nowe trendy w zakresie baz danych i ich zastosowań. Pozycje literaturowe zostały dobrane prawidłowo, biorąc pod uwagę ich istotność i aktualność. Odwołania do literatury zostały poprawnie przeanalizowane i wkomponowane w treść rozprawy.

- **wskazanie oraz ocena realizacji celu pracy**

Głównym celem pracy było zaproponowanie rozwiązań algorytmicznych w zakresie wzorców inter-sekwencyjnych w dużych bazach danych. Rozwiązanie stworzone przez Autora nazwane DBV-ISPMIC wraz z zaproponowaną strukturą DBV – PatternList oraz wersją równoległą nazwaną pDBV-ISPMIC, a także algorytmowi ISP-PI stanowi skuteczne rozwiązanie problemu naukowego przedstawionego w pracy. Przeprowadzone testy pozwoliły potwierdzić osiągnięcia pracy i stanowi oryginalne rozwiązanie poprawnie postawionego problemu naukowego.

Reasumując, w mojej ocenie cele pracy zostały zrealizowane w prawidłowy sposób. W mojej opinii praca stanowi więc

- **wskazanie oraz ocena zastosowanych metod badawczych**

Autor w sposób poprawny i konsekwentny przedstawił problem badawczy (w rozdziale 1) oraz przedstawił metody badawcze opierające się na krytycznej analizie problemu, określeniu podstaw matematycznych, zbudowaniu szeregu algorytmów oraz ich przetestowaniu. Ważną rolę w pracy odgrywa opisanie ograniczeń w końcowym rozdziale oraz kompleksowa analiza wyników. W mojej ocenie praca pod kątem metodyki badawczej i warsztatu naukowego został napisana w poprawny sposób.

Uwagi:

Poniżej przedstawiono uwagi dotyczące strony merytorycznej pracy:

- Autor opisuje ciekawy problem badawczy dotyczący inter-sekwencji w dużych bazach danych. W pracy przydałoby się szerzej odnieść do procesu skalowalności, gdyż to właśnie duże bazy danych będą stanowiły ważny obszar zastosowań dla prac Autora.
- Kolejna kwestia dotyczy ograniczeń (constraints). Autor odnosi się do tego problemu w kilku miejscach pracy. Zabrakło mi również szerszej dyskusji, na temat wpływu innych ograniczeń i konkretnych prac (future works), które powinny zostać w tym zakresie przeprowadzone – tym bardziej, że ostatnie lata przyniosły wiele prac w tym zakresie.
- W rozdziale 4.3.4. (strona 74) wprowadzono wersję równoległą algorytmu DBV-ISPMIC, gdzie dzieli się na kolejne zadania wyszukiwanie kolejnych wzorców. Jakie formy ewentualnej optymalizacji tego algorytmu Autor dostrzega?

- Autor jest proszony o odniesienie się do zagadnienia zakresu wartości parametru maxspan należącego do przedziału (0,13), i przedyskutowanie pod kątem możliwości zwiększenia tego zakresu.

Ogólnie praca jest napisana poprawnie od strony użytego języka czy składu. W pracy można wskazać drobne uwagi natury językowej, edytorskiej i stylistycznej:

- W pracy autor powołuje się na notację dwójkową czy dziesiętną w przypadku wzorców. W tym przypadku lepiej byłoby w tekście napisać wprost „in binary notation” czy „in decimal notation”.
- W zapisie zdarzają się niekonsekwencje w zapisie przedimków typu „a s-extension” (np. str. 34) – podczas, gdy reguły pisowni angielskiej mówią o pisowni zależnej od wymowy w zależności od spółgłoski czy samogłoski wymawianej na początku wyrazu.
- Niektóre rysunki, przykładowo dotyczące drzewa wzorców (4.2.), czy przykładów równoległego przetwarzania ISP-tree (4.3.) i ich jakość powinna być poprawiona

Opisane powyżej uwagi nie wpływają jednak negatywnie na mój pozytywny odbiór ocenianej pracy doktorskiej.

Wniosek końcowy

Rozprawa stanowi oryginalne rozwiązanie problemu naukowego i wskazuje na wysoki poziom wiedzy teoretycznej i praktycznej Kandydata w dyscyplinie naukowej Informatyka Techniczna i Telekomunikacja.

Przedstawiona praca doktorska spełnia w pełni warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (j.t. Dz.U. z 2020 r. poz. 85 z późn. zm.). W szczególności zgodnie z wymaganiami ww. ustawy oceniana rozprawa potwierdza wiedzę teoretyczną Kandydata oraz umiejętność samodzielnego prowadzenia pracy naukowej. Przedmiotem rozprawy jest oryginalne rozwiązanie problemu naukowego. W opinii recenzenta praca doktorska autorstwa mgr inż. Anh Nguyena może więc być dopuszczona do publicznej obrony.

Dr hab. inż. Jarosław Wąs Prof. AGH
jaroslaw.was@agh.edu.pl

