



RECENZJA

osiągnięcia naukowego dr inż. Wojciecha Kwedło nt.

„Algorytmy grupowania

danych wykorzystujące prototypy grup”

(na podstawie cyklu 5 publikacji powiązanych tematycznie

**w związku z wystąpieniem o nadanie stopnia naukowego doktora
habilitowanego w dziedzinie nauk technicznych w dyscyplinie**

„Informatyka techniczna i telekomunikacja”

**oraz ocena jego pozostałych osiągnięć naukowo-badawczych,
dydaktycznych i organizacyjnych**

1. Wstęp

Niniejsza recenzja została przygotowana w odpowiedzi na pismo Przewodniczącego Rady Naukowej Wydziału Informatyki Politechniki Białostockiej, prof. dr hab. inż. Marka Krętowskiego z dnia 16 października 2023 r. o powołaniu mnie jako recenzenta w postępowaniu habilitacyjnym dr inż. Wojciecha Kwedło.

Dr inż. Wojciech Kwedło w 1996 roku ukończył z wyróżnieniem studia magisterskie z zakresu informatyki prowadzone przez Instytut Informatyki Politechniki Białostockiej. Stopień doktora nauk technicznych w dyscyplinie informatyka uzyskał na Wydziale Informatyki tejże Uczelni w 2004 roku za obronioną z wyróżnieniem pracę poświęconą algorytmom ewolucyjnym jako narzędziu pozyskiwania wiedzy z baz danych. Od roku 1996 do chwili obecnej jego miejscem zatrudnienia jest Politechnika Białostocka, gdzie dr inż. Kwedło był najpierw asystentem (do września 2004 roku), a później adiunktem (od października 2004 roku do chwili obecnej). Oprócz wszechstronnej działalności dydaktycznej, Habilitant prowadził dostatecznie intensywną działalność naukowo-badawczą, przy czym na szczególne podkreślenie zasługuje jego zaangażowanie w realizację projektów badawczych finansowanych z różnorodnych źródeł.

2. Ocena osiągnięcia naukowego dr inż. Wojciecha Kwedło

W skład monotematycznego cyklu publikacji dokumentującego osiągnięcie naukowe nt. „Algorytmy grupowania danych wykorzystujące prototypy grup” wchodzi 5 pozycji o łącznej liczbie 59 stron. Zgodnie z autoreferatem Habilitanta, do artykułów tego cyklu zaliczono następujące prace (w przytoczonym spisie

podano współczynnik wpływu **IF** czasopisma, liczbę cytowań w bazie Web of Science (**WoS**), Scopus i Google Scholar (**GS**), przy czym w nawiasach przytoczono wartości bez autocytowań, jak również punkty MNiSW przydzielone publikacji (**PM**):

[C.1] W. Kwedlo: *A clustering method combining differential evolution with the K-means algorithm.* **Pattern Recognition Letters** 32 (12), 2011, pp. 1613–1621.

IF=1,0341, PM: 100, WoS: 91, Scopus: 119, GS:155

[C2] W. Kwedlo: *A parallel EM algorithm for Gaussian mixture models implemented on a NUMA system using OpenMP.* **Proceedings of the 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing - PDP 2014**, 2014, s. 292–298

PM: 20, WoS: 11, Scopus: 12, GS: 19

[C3] W. Kwedlo: *A new random approach for initialization of the multiple restart EM algorithm for Gaussian model-based clustering.* **Pattern Analysis and Applications** 18 (4), 2015), 757–770.

IF=1,104, PM: 70, WoS: 12, Scopus: 15, GS: 22

[C4] W. Kwedlo, P.J. Czocharński: *A hybrid MPI/OpenMP parallelization of K-means algorithms accelerated using the triangle inequality.* **IEEE Access** (7) 2019, 42280–42297

IF=3,745, PM: 100, WoS: 17, Scopus: 20, GS: 27 (wkład własny: 80%)

[C5] W. Kwedlo: *A hybrid steady-state evolutionary algorithm using random swaps for Gaussian model-based clustering.* **Expert Systems with Applications** (208) 2022, 118159

IF=8,665, PM: 140, GS:2

Wszystkie prace mają przydzielone punkty MNiSW. Dają one razem 430 punktów. Cztery artykuły zostało opublikowanych w czasopismach posiadających Impact Factor z bazy JCR z sumarycznym IF cyklu wynoszącym 14,55, przy czym trzy prace opublikowano w czasopismach należących do pierwszego, najwyższego kwartyłu Q1. Jedna praca opublikowana w materiałach znanej międzynarodowej konferencji naukowej indeksowana jest w bazie WoS. Sumaryczna liczbą cytowań prac cyklu według Web of Science wynosi 131 (126 bez autocytowań, co stanowi 96% ich ogólnej liczby). Powyższe wskaźniki stanowią podstawę do pozytywnej oceny poziomu publikacji wchodzących w skład przedstawionego cyklu od strony formalnej.

Na bardzo pozytywną ocenę zasługuje również strona stricte merytoryczna powyższego cyklu publikacji dokumentującego osiągnięcie naukowe dotyczącego zwiększenia efektywności algorytmów grupowania danych (*ang. data clustering*), stanowiącego ważną technikę uczenia maszynowego bez nadzoru, powszechnie stosowaną w eksploracji danych. Opisywany cykl publikacji związany jest z dwiema podstawowymi podejściami do grupowania danych: algorytmem *K*-średnich (*ang. K-means*), wykorzystującym sumaryczny błąd kwadratowy (*ang. SSE - sum of squared errors*) oraz grupowaniem przy pomocy modeli mieszanych rozkładów normalnych (*ang. GMM - Gaussian mixture models*)

wykorzystującym algorytm EM (*ang. expectation maximization*) do estymacji ich parametrów. Obydwie techniki są szeroko stosowane i nie bez podstaw zostały zaliczone do najważniejszych metod w eksploracji danych. W obydwu do reprezentacji grup wykorzystywane są ich prototypy. W przypadku metody K -średnich takim prototypem, zwanym centroidem, jest średnia arytmetyczna wektorów cech przypisanych do grupy. W przypadku modeli mieszanych rozkładów normalnych jako prototyp wykorzystywany jest wektor wartości oczekiwanej, będący parametrem komponentu mieszaniny o wielowymiarowym rozkładzie normalnym.

Pomimo dużej popularności, wyżej wymienione metody nie są pozbawione wad. Najważniejszą z nich jest lokalny charakter optymalizacji, co nie gwarantuje zbieżności do optimum globalnego. Najprostszą techniką redukującą negatywny wpływ tej wady jest wielokrotny start algorytmu, za każdym razem z innego (najczęściej losowo generowanego) punktu startowego. Bardziej zaawansowane metody wykorzystują algorytmy optymalizacji globalnej, takie jak algorytmy ewolucyjne, lub techniki hybrydowe łączące optymalizację globalną i lokalną. Drugą wadą omawianych algorytmów jest wysoka złożoność obliczeniowa. Wada ta jest szczególnie istotna w sytuacji, gdy algorytm optymalizacji lokalnej jest wykonywany wielokrotnie.

W publikacjach wchodzących w skład przedłożonego jednotematycznego cyklu Habilitant zaproponował oryginalne metody i algorytmy jako instrumenty zaradcze przeciwdziałające wadom dwóch powyższych technik grupowania. Publikacje te można podzielić na dwie grupy:

- Prace ukierunkowane na unikanie optimów lokalnych dzięki wykorzystaniu hybrydowych algorytmów optymalizacji globalnej [pozycje C5, C1] lub specjalnych technik inicjalizacji połączonych z wielokrotnym startem algorytmu optymalizacji lokalnej [pozycja C3].
- Prace koncentrujące się na wykorzystaniu obliczeń równoległych w celu redukcji czasu wykonania algorytmu [pozycje C4, C2].

Przegląd i analizę publikacji wchodzących w skład cyklu rozpocznę od pracy [C5], w której Habilitant zaproponował nowy hybrydowy algorytm ewolucyjny EARSEM (*ang. evolutionary algorithm with random swaps and expectation maximization*) przeznaczony do grupowania opartego na wykorzystaniu modeli mieszanych rozkładów Gaussa. Celowość jego opracowania wynika z faktu, iż na ogół lepszą alternatywą w stosunku do wielokrotnego uruchamiania klasycznego algorytmu EM, w celu uzyskania parametrów modelu mieszanego o możliwie największej wartości logarytmicznej funkcji wiarygodności, jest wykorzystanie algorytmu optymalizacji globalnej, na przykład algorytmu ewolucyjnego, a szczególnie metod hybrydowych łączących optymalizację globalną i lokalną. W metodach tych przebieg algorytmu EM staje się jednym z operatorów ewolucyjnych. Algorytm EARSEM reprezentuje sobą metodę stanu ustalonego, w każdym pokoleniu której wybiera się dwa osobniki z populacji, tworząc dwoje potomstwa za pomocą mutacji lub krzyżowania i dostrajając potomstwo za pomocą algorytmu EM maksymalizacji oczekiwań. Potomstwo konkuruje z rodzicami o przetrwanie w następnym pokoleniu. Zaproponowane podejście wykorzystuje losową zamianę, która

zastępuje wektor wartości oczekiwanych losowo wybranym wektorem cech jako operatorem mutacji. W operatorze krzyżowania losowy składnik jest kopiowany z mieszaniny źródłowej do mieszaniny docelowej. Kopiowanie w trybie krzyżowania faworyzuje składniki mieszaniny źródłowej znajdujące się z dala od składników mieszaniny docelowej. W eksperymentach obliczeniowych wykonanych przez Habilitanta na jednym rzeczywistym i 29 syntetycznych zbiorach danych zaproponowane rozwiązanie porównano z algorytmem EM z wieloma restartami, algorytmem EM z czysto losową wymianą i reprezentującym state-of-the-art. hybrydowym algorytmem ewolucyjnym bazującym na modelach mieszanych rozkładów normalnych. Osiągnięte wyniki wskazują, że przy tym samym budżecie obliczeniowym proponowana metoda zwykle zapewnia wyższą jakość modelu niż konkurencyjne algorytmy. W rezultacie, grupowania danych uzyskane zaproponowaną metodą najlepiej odpowiadają pierwotnym podziałom zbiorów danych na klasy.

W pracy [C1] rozpatrzono problem podziału zbioru danych na znaną liczbę skupień (grup) przy użyciu kryterium sumy kwadratów błędów SSE. Zaproponowano nową metodę grupowania, nazwaną DE-KM, która łączy algorytm ewolucji różnicowej DE z dobrze znaną procedurą K -średnich. W metodzie tej algorytm K -średnich jest używany do dostrojenia każdego rozwiązania kandydującego uzyskanego przez operatory mutacji i krzyżowania algorytmu DE. Dodatkowo, zaproponowano procedurę zmiany kolejności, która pozwala algorytmowi ewolucyjnemu uporać się z problemem redundantności reprezentacji. Efektywność opracowanej metody grupowania DE-KM zweryfikowano w wyczerpujących badaniach eksperymentalnych, które umożliwiły jej porównanie z efektywnością ewolucji różnicowej, globalnej metody K -średnich, genetycznego algorytmu K -średnich i dwóch wariantów klasycznego algorytmu K -średnich. Wyniki eksperymentu pokazują, że jeśli liczba K skupień jest wystarczająco duża, zaproponowana metoda DE-KM pozwala uzyskać rozwiązania o niższych wartościach kryterium SSE niż pozostałych pięć algorytmów.

W publikacji [C3] zaproponowano nową metodę inicjalizacji algorytmu EM wielokrotnego restartu dla grupowania opartego na mieszaninie modeli z rozkładem normalnym Gaussa. Metoda inicjuje losowo zarówno wektor wartości oczekiwanych, jak i macierz kowariancji składnika mieszaniny. W szczególności wektor wartości oczekiwanych jest inicjowany przez wektor cech wybrany deterministycznie z losowego podzbioru potencjalnych wektorów cech. Kryterium wyboru oparte jest na maksymalnej odległości Mahalanobisa pomiędzy wektorem kandydującym a już wybranym wektorem wartości oczekiwanych. Macierz kowariancji składnika mieszaniny jest inicjowana poprzez losowe generowanie jego wartości własnych i wektorów własnych. W przeprowadzonych eksperymentach obliczeniowych zaproponowane podejście porównano z trzema innymi metodami losowej inicjalizacji algorytmu w algorytmie EM. Eksperymenty zrealizowano na syntetycznych zbiorach danych wygenerowanych dla mieszanek rozkładów Gaussa o różnych charakterystykach, a także na czterech rzeczywistych zbiorach danych. Wyniki uzyskane na danych syntetycznych wskazują, że w przypadku dobrze rozdzielonych klastrów, dla których maksymalne nakładanie się par nie jest zbyt duże, opisana metoda daje skupienia, które lepiej odpowiadają pierwotnym podziałom danych. Eksperymenty na rzeczywistych

danych wskazują zaś, że opracowana metoda daje wyniki porównywalne z pozostałymi trzema metodami dla dwóch mniejszych zbiorów danych i znacznie lepsze dla dwóch większych zbiorów danych.

Praca [C4] stanowi pierwszą z dwóch pozycji poświęconych wykorzystaniu architektur i obliczeń równoległych w celu redukcji czasu wykonania algorytmu grupowania. Uwzględniono w niej fakt, iż standardowe sformułowanie grupowania z wykorzystaniem algorytmu K-średnich prowadzi do wykonania wielu niepotrzebnych obliczeń odległości. W związku z tym w artykule skupiono się na opracowaniu metody zrównoleglania czterech znanych algorytmów, znanych jako algorytmy Hamerly'ego, Elkana, Drake'a i Yinyang które wykorzystują nierówność trójkąta, aby uniknąć niepotrzebnych obliczeń odległości. Wychodząc naprzeciw charakterystykom współczesnych klastrów obliczeniowych, których węzły budowane są w oparciu o procesory wielordzeniowe, w pracy zaproponowano hybrydowy schemat zrównoleglania łączący standardy MPI oraz OpenMP programowania równoległego, w którym zarówno zbiór przetwarzanych danych, jak i odpowiadające im struktury danych przechowujące ograniczenia na odległości są równomiernie rozdzielone pomiędzy procesami MPI. W trakcie realizacji schematu intensywnie wykorzystywana jest operacji globalnej redukcji typu all-reduce, co korzystnie wpływa na wydajność obliczeń. W przeprowadzonych eksperymentach obliczeniowych zbadano silną skalowalność tych czterech algorytmów w porównaniu ze skalowalnością standardowego algorytmu zrównoleglonego przy użyciu tego samego podejścia. Uzyskane wyniki potwierdziły przewagę w wydajności wszystkich czterech algorytmów nad standardowym algorytmem. Na podkreślenie zasługuje również fakt, iż w badanych środowiskach obliczeniowych opracowane implementacje są bardziej wydajne w porównaniu z dwoma pakietami oprogramowania, których źródła są publicznie dostępne.

Druga praca w obszarze zrównoleglania, [C2], zawiera propozycję zrównoleglenia algorytmu EM grupowania opartego na mieszaninie modeli z rozkładem normalnym Gaussa. Jej wyróżnikiem jest orientacja na maszyny ze wspólną pamięcią o architekturze NUMA, a właściwie ccNUMA (cache-coherent non-uniform memory access). Tego typu architektury zyskały popularność w ostatnim okresie w związku z rozpowszechnieniem idei budowy procesorów wielordzeniowych w oparciu o tzw. chiplety czyli oddzielne układy krzemowe zamknięte w jednym korpusie i połączone odpowiednią siecią komunikacyjną. Tak w szczególności budowane są najnowsze procesory CPU firmy AMD, zgodne z architekturą EPYC kolejnych generacji czy też procesory firmy Ventana oparte o architekturę RISC-V. Jedną z zalet architektury ccNUMA w tym wydaniu jest większa skalowalność. Jej efektywne wykorzystanie wymaga jednak optymalizacji oprogramowania, tak aby jak najwięcej odwołań do danych było wykonywanych do pamięci lokalnej. Autorskie podejście do zrównoleglenia bazuje na odpowiednio dobranym wariancie dekompozycji danych w postaci dużych tablic przechowujących wektory cech i prawdopodobieństwa posteriori. Dodatkowo zastosowano szereg optymalizacji wykorzystujących właściwości ccNUMA, które umożliwiły wątkom optymalizację dostępu do pamięci lokalnej, w oparciu o politykę alokacji pamięci „pierwszego dotknięcia” (*ang. first-touch policy*) systemu operacyjnego Linux. Zaproponowaną metodę zaimplementowano w OpenMP i

przetestowano na 64-rdzeniowym systemie opartym na czterech procesorach AMD Opteron 6272 (o nazwie kodowej „Interlagos”). Wyniki eksperymentów pokazały, że na dużych zbiorach danych algorytm bardzo dobrze skaluje się wraz ze wzrostem liczby rdzeni, a dokonane optymalizacje dla ccNUMA znacząco poprawiły jego wydajność.

Podsumowując powyższy przegląd publikacji wchodzących w skład przedłożonego cyklu, należy stwierdzić, że zawierają one oryginalne i wartościowe wyniki dotyczące zarówno aspektów czysto algorytmicznych ulokowanych w tak ważnym obecnie obszarze jakim jest Data Science, jak i wymagającej pod wieloma względami, lecz jednocześnie bardzo nośnej aplikacyjnie, dziedzinie obliczeń wysokiej wydajności (*ang. HPC – high performance computing*) z wykorzystaniem obliczeń równoległych. Na podkreślenie zasługuje przy tym bardzo szczegółowe i spójne przedstawienie uzyskanych wyników w sposób dostatecznie jasny dla ich zrozumienia przez badaczy nie specjalizujących się w danym obszarze informatyki.

Jednocześnie nasuwają się dwie uwagi krytyczne, które nie podważają wartości i oryginalności uzyskanych rezultatów. Ich uwzględnienie pozwoliłoby jednak na ich zaprezentowanie w szerszym kontekście. Pierwsza z nich dotyczy braku odniesienia się do innych, alternatywnych podejść do rozwiązywania zagadnienia klasyfikacji i scharakteryzowania na ich tle miejsca rozpatrywanego przez Habilitanta podejścia opartego na grupowaniu danych. Wśród tych alternatywnych podejść można wymienić chociażby drzewa decyzyjne czy sieci neuronowe różnego typu. Druga z moich uwag dotyczy braku informacji o praktycznym wykorzystaniu przedstawionych przez Habilitanta propozycji algorytmicznych i implementacyjnych w konkretnych systemach informatycznych opracowanych z wydatnym udziałem dr Kwedło (patrz następny punkt mojej recenzji) i przeznaczonych głównie do przetwarzania dużych zbiorów danych medycznych.

Konkluzja oceny jednotematycznego cyklu publikacji

Podsumowując tę część recenzji, stwierdzam, iż pomimo powyższych uwag krytycznych przedłożone osiągnięcie naukowe udokumentowane jednotematycznym cyklem publikacji pod wspólnym tytułem „Algorytmy grupowania danych wykorzystujące prototypy grup” może być ocenione jako **wносящее значный вклад в развитие дисциплины научной „Информатика техническая и телекоммуникация”** w zakresie metod i algorytmów rozwiązywania złożonych zagadnień przetwarzania dużych zbiorów danych, ze szczególnym uwzględnieniem ich zrównoleglania z wykorzystaniem technologii HPC. Stwierdzam więc, że przedstawiony cykl publikacji może być podstawą do nadania dr. inż. Wojciechowi Kwedło stopnia naukowego doktora habilitowanego nauk technicznych w dyscyplinie „Informatyka techniczna i telekomunikacja”.

3. Ocena istotnej aktywności naukowej w tym realizowanej w więcej niż jednej uczelni lub instytucji naukowej, w szczególności zagranicznej

Dorobek naukowo-badawczy

Oprócz oceny samego cyklu publikacji, warto odnotować następujące statystyki dokumentujące osiągnięcia naukowo-badawcze Habilitanta po uzyskaniu stopnia doktora w postaci dorobku publikacyjnego:

- Liczba punktów wg listy MNiSW uzyskanych z cyklu publikacji zgłoszonego jako osiągnięcie naukowe Habilitanta: 410
- Sumaryczny Impact Factor dla osiągnięcia naukowego: 14,55
- Sumaryczna liczba cytowań dla osiągnięcia naukowego 131, w tym 126 bez autocytowań (96%)
- Sumaryczny Impact Factor publikacji: 26,62
- Liczba publikacji po habilitacji: 29, w tym:
 - artykuły w czasopismach znajdujących się w bazie JCR: 9
 - rozdziały w monografiach: 15
- Indeks Hirscha (Web of Science): 9
- Indeks Hirscha (Scopus): 10
- Indeks Hirscha (Google Scholar): 11
- Liczba cytowań (Web of Science): 255, w tym 238 bez autocytowań (93%)
- Liczba cytowań (Scopus): 328, w tym 303 bez autocytowań (92%)
- Liczba cytowań ((Google Scholar): 501

Po doktoracie Habilitant opublikował 29 prac. Liczba opublikowanych prac nie jest może zbyt duża, jak na okres czasu od uzyskania doktoratu (około 18 lat), ale należy uwzględnić dużą liczbę prac samodzielnych – aż 11, co bardzo pozytywnie świadczy o wkładzie własnym Kandydata. Na podkreślenie zasługuje znaczna liczba artykułów (18) opublikowanych w recenzowanych materiałach specjalistycznych konferencji międzynarodowych indeksowanych w bazie DBLP dedykowanej informatyce, w tym 4 prace w materiałach prestiżowych konferencji rangi Core A. Istotnym pozytywnym wyróżnikiem dorobku publikacyjnego Habilitanta są przy tym dostatecznie wysokie dla danego etapu kariery naukowej wartości wskaźników cytowań, tj. indeksów Hirscha i liczby cytowań, szczególnie z uwzględnieniem faktu, iż aż 93% ogólnej liczby cytowań w bazie Web of Science nie dotyczy autocytowań (dla bazy Scopus wskaźnik ten jest praktycznie taki sam). Świadczy to jednoznacznie pozytywnie o stopniu akceptacji wyników prac Habilitanta przez międzynarodową społeczność naukową i jego rozpoznawalności w tym środowisku.

Działalność naukowa Habilitanta związana była z udziałem w szeregu projektów badawczych finansowanych z różnych źródeł, takich jak:

- Projekt celowy „**CLUSTERIX – Krajowy Klaster Linuksowy (National Cluster of Linux Systems)**”, lata 2003-2006, finansowany przez MNiSW - projekt realizowany przez konsorcjum 12 jednostek KDM i MAN, mający na celu zbudowanie rozproszonego klastra składającego się z odległych geograficznie klastrów lokalnych połączonych siecią Polskiego Internetu Optycznego PIONIER; w ramach projektu Habilitant kierował zadaniem WP15 „Nowe aplikacje testowe”. Celem tego zadania było opracowanie i przetestowanie nowych aplikacji równoległych, w tym aplikacji obejmujących zasięgiem więcej niż jeden klaster lokalny, realizowanych przy pomocy biblioteki MPICH-G2 standardu MPI.
- Projekt „**MOBIT: Stworzenie referencyjnego modelu diagnostyki personalizowanej guzów nowotworowych w oparciu o analizę heterogenności guza z wykorzystaniem biomarkerów genomowych, transkryptomu i metabolomu oraz badań obrazowych PET/MRI jako narzędzia do wdrażania i monitorowania terapii zindywidualizowanej**”, lata 2015-2019, finansowany przez NCBiR, wykonawca odpowiedzialny m.in. za wykorzystanie rozproszonych metod uczenia maszynowego opartych o bibliotekę MLlib pakietu Spark we wspieraniu diagnostyki medycznej.
- Projekt „**Analiza relacji względnych w eksploracji zintegrowanych danych omicznych**”, lata 2020-2023, finansowany przez NCN w ramach konkursu OPUS 17, wykonawca odpowiedzialny za zaprojektowanie i wygenerowanie sztucznych wielowymiarowych zbiorów danych zawierających różnego rodzaju klastry, wstępną analizę statystyczną oraz eksplorację stworzonych danych.
- Projekt „**Centrum sztucznej inteligencji w medycynie przy Uniwersytecie Medycznym w Białymstoku**”, finansowany przez Ministerstwo Zdrowia (aktualnie realizowany) – wykonawca odpowiedzialny za rozwój systemu informatycznego, a także za zastosowanie metod grupowania (klasteryzacji) do analizy danych pacjentów w stanie przedcukrzycowym (wspólnie z pracownikami Uniwersytetu Medycznego).

Aktywność naukowa dr inż. Kwedło prowadzona była nie tylko w Politechnice Białostockiej (choć tam głównie była skoncentrowana), lecz również w innych instytucjach naukowych, a mianowicie:

- W *Centrum voor Wiskunde et Informatica (CWI)* w Amsterdamie, gdzie przez 2 miesiące w 1996 roku zajmował się realizacją systemów wieloagentowych opartych o modele automatów uczących się w rozproszonym środowisku Manifold rozwijanym w CWI.
- W Universität Fridericiana w Karlsruhe (TH), 1.09.1999-31.01.2000, gdzie odbywał staż naukowo-badawczy finansowany przez stypendium badawcze Niemieckiej Centrali Wymiany Akademickiej DAAD, zajmując się rozwojem systemu EDRL-MD i jego zastosowaniem do analizy danych pozyskanych z przemysłu motoryzacyjnego (firma Daimler-Chrysler, obecnie Daimler-Benz); rezultatem prowadzonych prac były dwie wspólne publikacje, w tym praca dotycząca dostosowania systemu EDRL-MD do danych z różnymi kosztami błędnych decyzji.

Ważny element dorobku Kandydata stanowi kierowanie badaniami statutowymi prowadzonymi w Katedrze Oprogramowania Wydziału Informatyki Politechniki Białostockiej w latach 2014-2022 w zakresie metod eksploracji dużych zbiorów danych, a także wdrożenie innowacyjnych technologii takich jak:

- opracowanie systemu informatycznego do gromadzenia i analizy wielkoskalowych danych medycznych, zbudowany w ramach projektu MOBIL finansowanego przez program STRATEGMED NCBiR; system ten jest wdrożony w Centrum Badań Klinicznych Uniwersytetu Medycznego w Białymstoku i obecnie jest wykorzystywany do gromadzenia i analizy danych w ramach projektu Polski Rejestr Diabetologiczny. Co ważne, system jest ciągle rozbudowywany i trwają prace nad zastosowaniem go dla innych grup pacjentów.
- budowa trzech klastrów obliczeniowych, z których dwa pierwsze używały nowatorskiej (na swoje czasy) technologii Infiniband. Habilitant był odpowiedzialny za koncepcję, projekt, wnioski do podmiotów finansujących, którymi były fundusze strukturalne UE oraz Ministerstwo Nauki i Szkolnictwa Wyższego, instalację oprogramowania i bieżącą administrację klastrów.

Ocena aktywności Habilitanta w rozpatrywanym obszarze byłaby niepełna bez uwzględnienia wygłoszenia w okresie po doktoracie 8 referatów na specjalistycznych międzynarodowych konferencjach naukowych, wykonanie aż 73 recenzji prac dla czasopism naukowych, włączając recenzje dla takich prestiżowych czasopism w dyscyplinie informatyka techniczna i telekomunikacja, jak np. IEEE Transactions on Pattern Analysis and Machine Intelligence, Artificial Intelligence in Medicine, Pattern Recognition czy Journal of Parallel and Distributed Computing, ja również członkostwo w wiodących międzynarodowych organizacjach zrzeszających badaczy w obszarze technologii ICT, takich jak IEEE (od 2011 r.) wraz z IEEE Computer Society (od 2020 r.) i IEEE Computational Intelligence Society (od 2011 r.) oraz ACM (od lutego 2021 r.), co niewątpliwie również sprzyjało zwiększeniu stopnia jego rozpoznawalności w środowisku.

Inne elementy dorobku dydaktycznego i organizacyjnego

Działalność dydaktyczna dr Kwedło związana była przede wszystkim z prowadzeniem zajęć dydaktycznych dla studentów kierunku informatyka macierzystej uczelni. Na szczególne podkreślenie zasługują przy tym takie elementy dorobku dydaktycznego jak: pełnienie funkcji koordynatora i prowadzenia wykładów i specjalistycznych zajęć laboratoryjnych pracowni specjalistyczne z przedmiotów obowiązkowych obejmujących systemy operacyjne, zaawansowane systemy operacyjne, obliczenia równoległe, oraz obieralnych, w tym dotyczących algorytmów ewolucyjnych, obliczeń z wykorzystaniem akceleratorów graficznych GPU, metod wielowymiarowej analizy danych oraz modeli i metod statystycznych. Prowadził też kilkakrotnie zajęcia z systemów operacyjnych dla studentów zagranicznych w ramach programu Erasmus+. Jego zajęcia zawsze były wysoko oceniane w anonimowych ankietach

studenckich. W szczególności, w roku akademickim 2017/2018 znalazł się pośród trzech najwyżej ocenianych przez studentów nauczycieli akademickich na Wydziale Informatyki, co znalazło swój wyraz w otrzymaniu nagrody Rektora. Pełnił także funkcję promotora 10 prac inżynierskich i 14 magisterskich. O wysokim poziomie tych prac świadczy fakt, że finalnym wynikiem dwóch z nich były wspólne publikacje w czasopiśmie IEEE Access z listy JCR.

Działalność organizacyjna Habilitanta nie jest może zbyt bogata, lecz mimo to zasługuje na generalnie pozytywną ocenę. Z dokumentacji zawartej we wniosku wynika, iż w okresie po obronie doktoratu skupiała się ona głównie na kierowaniu planowaniem, wydatkowaniem i rozliczaniem środków na badania statutowe prowadzone w latach 2024-2022 przez zespół około 15 pracowników naukowo-dydaktycznych Katedry Oprogramowania Wydziału Informatyki, działalności w Radzie Wydziału oraz w szeregu komisjach wydziałowych, z których najważniejsze to Wydziałowa Komisja ds. Oceny Nauczycieli Akademickich oraz Wydziałowa Komisja ds. Modyfikacji Programów Studiów. Brał także bardzo aktywny udział w pracach komitetu organizacyjnego międzynarodowej konferencji naukowej 13th International Conference on Parallel Processing and Applied Mathematics odbywającej się we wrześniu 2019 roku na Wydziale Informatyki Politechniki Białostockiej.

Konkluzja oceny istotnej aktywności naukowej

Po doktoracie Habilitant był współautorem lub autorem 29 publikacji, uzyskując dostatecznie wysokie dla danego etapu kariery naukowej wartości wskaźników cytowań, co potwierdza jego dobrą rozpoznawalność przez międzynarodową społeczność naukową, szczególnie z uwzględnieniem faktu, iż ponad 90% ogólnej liczby cytowań w bazach Web of Science i Scopus nie dotyczy autocytowań, uczestniczył w 4 dużych projektach finansowanych ze źródeł zewnętrznych, przy czym w ramach ogólnokrajowego projektu celowego ClusteriX kierował zadaniem WP15 „Nowe aplikacje testowe”, wygłosił 8 referatów na specjalistycznych międzynarodowych konferencjach naukowych, odznaczył się bardzo dużą liczbą recenzji prac dla czasopism naukowych, włączając recenzje dla najbardziej prestiżowych czasopism w dyscyplinie informatyka techniczna i telekomunikacja, wyróżnił się wdrożeniem innowacyjnych technologii takich jak system informatyczny do gromadzenia i analizy wielkoskalowych danych medycznych, wdrożony w Centrum Badań Klinicznych Uniwersytetu Medycznego w Białymstoku i obecnie wykorzystywany do gromadzenia i analizy danych w ramach projektu Polski Rejestr Diabetologiczny, a także był autorem trzech innowacyjnych klastrów obliczeniowych, wyróżnił się szeregiem istotnych elementów dorobku dydaktycznego, w tym prowadzeniem zajęć dla studentów zagranicznych w ramach programu Erasmus+ oraz pełnieniem funkcji promotora licznych prac inżynierskich i magisterskich, o których wysokim poziomie świadczy fakt wspólnych publikacji w renomowanym czasopiśmie, w ramach działalności organizacyjnej może pochwalić się wieloletnim zarządzaniem badaniami statutowymi prowadzonymi w Katedrze Oprogramowania oraz działalnością w Radzie Wydziału Informatyki i w szeregu komisjach wydziałowych, jak również bardzo aktywnym udziałem w pracach komitetu organizacyjnego międzynarodowej konferencji naukowej organizowanej na Wydziale.

Stwierdzam również, iż istotna działalność naukowa Kandydata była realizowana w więcej niż jednej uczelni lub instytucji naukowej, w szczególności zagranicznej.

Podsumowując w tym kontekście ocenę aktywności naukowej dr. inż. Wojciecha Kwedlo po doktoracie stwierdzam zatem, że dorobek naukowy, w tym publikacyjny i projektowo-technologiczny, a także dydaktyczny i organizacyjny spełniają wymagania stawiane w postępowaniu o nadanie stopnia doktora habilitowanego w dziedzinie nauk technicznych.

4. Podsumowanie

Przedłożone osiągnięcie naukowe udokumentowane jednotematycznym cyklem publikacji pod wspólnym tytułem „Algorytmy grupowania danych wykorzystujące prototypy grup” wnosi znaczny wkład w rozwój dyscypliny naukowej „Informatyka techniczna i telekomunikacja” w zakresie metod i algorytmów rozwiązywania złożonych zagadnień przetwarzania dużych zbiorów danych, ze szczególnym uwzględnieniem ich zrównoleglenia. Dokonania przedstawione w osiągnięciu naukowym składają się na oryginalny i twórczy wkład w obszarze informatyki technicznej, w szczególności istotny dla rozwoju Data Science oraz systemów HPC i posiadają duże znaczenie dla dalszego rozwoju tej problematyki, zarówno w aspekcie teoretycznym, jak i praktycznym.

Jednocześnie uważam że działalność naukowo-badawcza Habilitanta oraz projektowo-konstrukcyjna, jak też działalność dydaktyczna i organizacyjna w sferze nauki, spełniają wymagania stawiane w postępowaniu o nadanie stopnia doktora habilitowanego w dziedzinie nauk technicznych, zwłaszcza spełniony jest ustawowy wymóg, aby istotna działalność naukowa Kandydata była realizowana w więcej niż jednej uczelni lub instytucji naukowej, w szczególności zagranicznej.

W konkluzji stwierdzam więc, że przedłożone osiągnięcie naukowe oraz całokształt dorobku naukowo-badawczego, dydaktycznego i organizacyjnego dr inż. Wojciecha Kwedlo spełniają wymogi aktualnie obowiązującej Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki, co upoważnia mnie do wnioskowanie o nadanie mu stopnia doktora habilitowanego nauk technicznych w dyscyplinie naukowej „Informatyka techniczna i telekomunikacja.

