

Recenzja

osiągnięcia naukowego

dra inż. Wojciecha Macyny

w związku z postępowaniem habilitacyjnym prowadzonym przez Politechnikę Wrocławską,
Rada Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja

1 Informacje Ogólne

Recenzja została sporządzona na podstawie pisma Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Wrocławskiej oraz dostarczonych materiałów obejmujących autoreferat, wykaz osiągnięć naukowych oraz kopii publikacji.

Wojciech Macyna uzyskał tytuł magistra inżyniera informatyki na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej w roku 1997, a następnie stopień doktora nauk technicznych w zakresie informatyki (dyscyplina informatyka) na tym samym wydziale w roku 2004. Od roku 2005 jest pracownikiem naukowo-badawczym Politechniki Wrocławskiej (asystent, adiunkt, starszy wykładowca), w tym od 2021 jest zatrudniony jako adiunkt w Katedrze Podstaw Informatyki na Wydziale Informatyki i Telekomunikacji.

Tematyka prac badawczych od początku dotyczyła systemów baz danych, zarządzaniem danymi oraz mechanizmów przechowywania i manipulacji danymi, z wykorzystaniem nowoczesnych urządzeń sprzętowych. Jest to powiązane ściśle z działalnością dydaktyczną, dotyczącą baz danych, systemów informatycznych, wytwarzania oprogramowania, programowania aplikacji oraz zarządzania projektami informatycznymi. Autor posiada też doświadczenie przemysłowe, poprzez pracę w kilku firmach informatycznych, w tym Comarch w latach 2005-2009, oraz realizację projektów badawczo-wdrożeniowych w ramach Dolnośląskich Bonów na Innowacje.

2 Ocena osiągnięcia naukowego

Zgłoszonym osiągnięciem habilitacyjnym jest jednotematyczny cykl 12 publikacji zatytułowany: „Przechowywanie oraz manipulowanie danymi z uwzględnieniem nowoczesnych rodzajów pamięci”. Osiągnięcie to składa się z trzech części: (1) dotyczy badań związanych z przechowywaniem danych na pamięci flash, (2) badań dotyczących indeksowania na drzewie LSM, oraz (3) badań związanych z indeksowaniem na pamięci PCM.

WPLYNĘŁO

17-11-2025

RDNDO3/224/2025

2.1 Przegląd osiągnięć

2.1.1 Badania związane z przechowywaniem danych na pamięci flash

Prace badawcze, opublikowane w latach 2011-2020 artykułach [A1], [A2], [A3], [A4], [A5], [A6], [A7] dotyczą dostosowania systemów przechowywania danych i zarządzania bazami danych do lepszego wykorzystania pamięci flash. Jest to rodzaj pamięci obecnie stosowany w dyskach SSD, charakteryzujący się innymi ograniczeniami niż dyski twarde, co wynika z jej architektury sprzętowej, na przykład usuwania bloków pamięci, różnic w kosztach energetycznych między zapisem a odczytem, oraz ograniczonej żywotności. Powoduje to konieczność dostosowania struktur danych oraz algorytmów do tej charakterystyki.

Pierwszym z osiągnięć w tym zakresie jest zaproponowanie liczników probabilistycznych do zliczania operacji usuwania danych z bloków pamięci flash. Jest to istotne z powodu ograniczonej żywotności takich bloków, które przestają działać po zbyt dużej liczbie operacji usunięć. Ponieważ nie potrzebne jest dokładne zliczanie tych operacji, autor zaproponował liczniki przybliżone oparte na liczbach losowych, co znacząco zmniejsza rozmiar takich liczników, bez istotnej straty w jakości ich działania. Zostało to pokazane teoretycznie oraz potwierdzone eksperymentalnie poprzez zaimplementowany symulator pamięci flash.

Kolejnym osiągnięciem jest opracowanie metod tworzenia indeksów w systemach baz danych, uwzględniających charakterystykę pamięci flash, w której przechowywane są pliki bazy danych. Autor opracował metody przechowywania wielkości zagregowanych związanych ze strukturą R-drzewa, będącego jedną z odmian indeksu przestrzennego, tak aby były one dostosowane do pamięci flash. Przechowywanie wartości zagregowanych w osobnym miejscu pamięci w stosunku do zawartości R-drzewa. Wiąże się to z problemem minimalizacji liczby zapisywanych bloków pamięci przy wyborze rozmieszczenia wartości zagregowanych, co jest problemem NP-trudnym. Zaproponowane algorytmy aproksymacyjne pozwoliły na zwiększenie wydajności metody poprzez istotne zmniejszenie liczby operacji zapisu do pamięci w porównaniu do wcześniejszych metod. Kolejnym elementem prac było opracowanie indeksu klastrowego dla baz danych z wykorzystaniem zaproponowanej nowej struktury danych zwanej FA-drzewem. Ma ono na celu zmniejszenie liczby kosztownych operacji zapisu i usuwania danych z pamięci flash. Dodatkowo, autor zaproponował nowe metody tworzenia indeksu wtórnego częściowego, dostosowane do pamięci flash, dzięki czemu czas odpowiedzi na zapytania może być skrócony dwukrotnie w porównaniu do tradycyjnych metod. Jako ostatni element tych prac, autor opracował nową koncepcję przechowywania kolumnowych baz danych na pamięci flash, poprzez stworzenie struktury CF-drzewa, na której oparto przechowywanie danych, oraz formalnie oszacowano jej efektywność.

2.1.2 Indeksowanie na drzewach LSM

Prace badawcze opublikowane w latach 2022-2024 w artykułach [A8], [A9] i [A10] dotyczą indeksu wtórnego w bazach danych typu klucz-wartość opartych na drzewach LSM. Drzewa takie stosowane są w systemach baz NoSQL takich jak Hbase, MongoDB, czy Cassandra i służą do przechowywania danych w plikach zorganizowanych w wiele poziomów hierarchii.

Pierwszym z osiągnięć w tym zakresie jest opracowanie metody ułatwiającej ładowanie wielu elementów indeksu wtórnego na raz (bulk load) do drzewa LSM znajdującego się na pamięci flash. Zaletą takiego podejścia jest unikanie kosztownego dla pamięci flash łączenia poziomów, przy nieznacznym wydłużeniu wyszukiwania elementów na danym poziomie. Efektywność działania metody potwierdzona została w sposób eksperymentalny dla różnego zestawu danych oraz zapytań.

Kolejnym osiągnięciem jest opracowanie metody częściowego indeksowania dla baz danych opartych o drzewo LSM, wykorzystując fakt że dane są przechowywane w wielu niedużych plikach. Zaproponowane rozwiązanie wykorzystuje wielowątkowość oraz implementację opartą o bazę danych LevelDB. Ewaluacja eksperymentalna posłużyła do porównania zaproponowanej metody z metodą wykorzystującą indeksowanie całkowite w zależności od ilości danych, zakresu zapytań oraz liczby wątków.

Trzecim osiągnięciem w tym zakresie jest zaproponowanie hierarchicznych filtrów Blooma do wyszukiwania elementów niekluczowych w bazach danych opartych o drzewo LSM. Opracowane zostały algorytmy wyszukiwania wartości nieposortowanych oraz tworzenia hierarchii filtrów Blooma przy łączeniu poziomów drzewa LSM. Eksperymenty przeprowadzone z użyciem bazy LevelDB pokazały znaczącą redukcję czasu przeszukiwania w porównaniu do podstawowej metody.

2.1.3 Indeksowanie na pamięci PCM

Prace badawcze opublikowane w latach 2018-2022 w artykułach [A11] i [A12] dotyczą wykorzystania w bazach danych pamięci typu PCM, która jest adresowana bajtowo i charakteryzuje się szybkim odczytem, ale za to wolniejszym zapisem.

Pierwszym osiągnięciem jest zaproponowanie i implementacja mechanizmów przechowywania zagregowanych wartości w R-drzewach przy użyciu pamięci PCM. Algorytm szacuje koszty operacji odczytu na podstawie hierarchicznej struktury drzewa. Przeprowadzone eksperymenty pokazały że metoda redukuje ponaddwukrotnie liczbę operacji zapisu w porównaniu do podstawowej metody FIFO.

Drugim osiągnięciem w tym zakresie jest zaproponowanie metody częściowego indeksowania na pamięci PCM, dla zapytań o zakres danych. Zaproponowano też nowy indeks oparty o BB+ drzewo oraz opracowano symulator pamięci PCM, przy pomocy którego oceniono wydajność algorytmów. Eksperymenty pokazały że nowy częściowy indeks jest wydajniejszy od indeksów znanych z literatury.

2.2 Ocena wkładu do dyscypliny

Przedstawione osiągnięcia autora oceniam na podstawie ich związku z dyscypliną, oryginalności, jakości metodologicznej oraz wpływu.

Po pierwsze, autor niewątpliwie jest specjalistą w tematyce baz danych oraz wykorzystania nowych rozwiązań sprzętowych do przechowywania danych w pamięci dyskowej, w szczególności typu flash i PCM. Jest to istotna dziedzina informatyki, zaś autor jawi się jako ekspert zarówno w zakresie elementów algorytmicznych i struktur danych wykorzystywanych

przez nowoczesne systemy baz typu SQL i NoSQL, ich odwzorowania na architekturę sprzętową oraz na praktyczne zastosowania w różnych rodzajach zapytań. Choć zakres czasowy cyklu publikacji jest szeroki: obejmuje lata 2011 – 2024, to tematyka artykułów była aktualna w momencie ich publikacji.

Po drugie, autor proponuje niewątpliwie rozwiązania oryginalne i dobrze osadzone w aktualnym na moment publikowania stanie wiedzy, a zaprezentowane prace odwołują się do istotnych prac z literatury w dziedzinie, a tym dotyczące baz danych (VLDB, USENIX, SIGMOD) oraz sprzętu (ISSCC, MICRO, IEDM). Wartościowym elementem prac jest połączenie aspektów algorytmicznych i częściowo teoretycznych, z dostosowaniem proponowanych rozwiązań do charakterystyki sprzętowej, w szczególności pamięci flash czy PCM. Proponowane rozwiązania są poddawane ewaluacji, pokazując zalety w porównaniu do bazowych istniejących aktualnie rozwiązań w dziedzinie.

Po trzecie, pozytywnym elementem wkładu do dyscypliny jest moim zdaniem wysoka jakość przedstawionych publikacji naukowych pod względem metodologicznym. Każdy z artykułów przedstawia dobrze określony problem, jego analizę od strony teoretycznej czy formalnej, w tym w niektórych przypadkach twierdzenia z dowodami, zaproponowane rozwiązania w postaci schematów oraz algorytmów, a także ilustracyjnych przykładów, wreszcie eksperymenty przy użyciu symulacji lub środowisk sprzętowych, z wynikami pokazującymi działanie zaproponowanych algorytmów i rozwiązań. W artykułach widać dużą staranność w prowadzeniu badań i w formie prezentacji osiągniętych wyników. Jedynym zauważalnym niedociągnięciem jest brak analizy błędów pomiarowych, co powinno być standardowym elementem każdej pracy wykorzystującej analizę wydajności systemów komputerowych, których złożoność powoduje naturalny niedeterminizm i fluktuacje wyników.

Po czwarte, pewną, choć niedoskonałą, miarą znaczenia osiągnięć naukowych jest ranga czasopism i konferencji, na których zostały one opublikowane, a także liczba cytowań wskazująca na wpływ tych prac i zainteresowanie nimi przez środowisko naukowe. Artykuły prezentowane były na konferencjach o randze co najwyżej B wg CORE, żadna z prac nie uzyskała więcej niż 70 pkt MNiSW, co jest poniżej oczekiwań. W przypadku autora, liczba cytowań według bazy Google Scholar: 68, h-index:4 (na dzień sporządzenia recenzji), a według bazy Scopus: 25, h-index:2 (na dzień złożenia wniosku), co jest wynikiem raczej niewielkim. Pytaniem pozostaje, dlaczego autor, choć jest dobrze zorientowany w swojej dziedzinie, nie podjął próby publikacji na choć jednej konferencji o wyższej randze, co dałoby szansę na większą rozpoznawalność i wpływ osiągnięć. Pozytywnie należy ocenić publikacje na konferencji Database and Expert Systems Applications (DEXA), będącej uznaną konferencją w dziedzinie.

Podsumowując, ocena wskazanych osiągnięć naukowych jest pozytywna – choć nie są one przełomowe i ich wpływ jest trochę ograniczony, stanowią znaczący wkład w rozwój dyscypliny.

3 Ocena pozostałych aktywności

Autor jest aktywnym członkiem społeczności naukowej, co przejawia się m. in. udziałem w międzynarodowych (w tym IEEE Big Data, DEXA, International Conference on Management of Data SIGMOD/PODS) i krajowych konferencjach naukowych (m.in. BDAS) – są to uznane

konferencje, choć nie wiodące w dziedzinie. Część prac powstała w wyniku współpracy międzynarodowej (USA, Francja) z Houston University of Technology i ISAE ENSMA w Poitiers. Autor odbył też krótką wizytę naukową w Xidian University w Chinach. Był też redaktorem sekcji tematycznej, współorganizatorem warsztatów naukowych oraz szkolenia na konferencji IEEE Big Data. O aktywności świadczy też, że był zapraszany do komitetów programowych kilkunastu konferencji międzynarodowych.

Należy podkreślić duże zaangażowanie dydaktyczne autora, w tym promotorstwo ponad 100 prac inżynierskich i ponad 30 magisterskich, prowadzenie przedmiotów związanych z bazami danych, nauką programowania i inżynierią oprogramowania, a także kursów z zastosowań SQL dla osób z firm zewnętrznych. W autoreferacie wymienione są także udział w projektach dydaktycznych oraz opieka nad praktykami studenckimi. Był także promotorem pomocniczym pracy doktorskiej mgr inż. Michała Kukowskiego, z którym opublikował kilka prac wchodzących w skład osiągnięcia habilitacyjnego.

Autor posiada też cenne doświadczenie praktyczne i projektowe, powiązane tematycznie z pracami naukowymi. W ramach pracy w firmie Comarch oraz współprac z innymi firmami zrealizowane były m. in. projekty dotyczące systemu zarządzania dla małych i średnich przedsiębiorstw, hurtownie danych, architektury systemu ERP, system do transformacji danych obiektów muzealnych do formatu RDF oraz prototyp systemu do zdalnego zarządzania treściami muzycznymi. Część z tych prac realizowana była w ramach programu Dolnośląski Bon na Innowacje.

4 Konkluzja

Przedstawione osiągnięcia naukowe dra inż. Wojciecha Macyny zawarte w jednotematycznym cyklu 12 publikacji dotyczących przechowywania oraz manipulowania danymi z uwzględnieniem nowoczesnych rodzajów pamięci, w mojej opinii stanowią znaczący wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. Prace te stanowią ciekawe i oryginalne połączenie elementów algorytmicznych i struktur danych wykorzystywanych przez nowoczesne systemy baz typu SQL i NoSQL, z ich odwzorowaniem na architekturę sprzętową oraz pokazują praktyczne zastosowania zaproponowanych rozwiązań w różnych rodzajach zapytań.

Ocena pozostałej działalności naukowej, dydaktycznej i projektowej jest również pozytywna, gdyż autor wykazuje się współpracą zagraniczną, jest aktywny na międzynarodowych konferencjach, posiada duży dorobek dydaktyczny, a także doświadczenie praktyczne w realizacji projektów informatycznych przy współpracy z firmami.

Podsumowując, na podstawie oceny otrzymanej dokumentacji, wnioskuję o dopuszczenie dra inż. Wojciecha Macyny do dalszych etapów postępowania w sprawie nadania stopnia doktora habilitowanego.

 Digitally signed by
Maciej Malawski
Date: 2025.11.15
20:25:51 CET



Raport weryfikacji

Wynik

Wynik przetwarzania
Skrót pliku z podpisem
Nazwa pliku z podpisem
Identyfikator weryfikacji

WAŻNY/PRAWIDŁOWY

Zakończono
298bf885bec423808888d6575c90ae86190e96f0
WojciechMacyna-recenzja-MMalawski.pdf
B3F681339293DBA77014A2C6A08504481CD742EC

Lista podpisów

Status weryfikacji podpisu
Data i czas wykonania podpisu
Data i czas weryfikacji podpisu
Format podpisu
Poziom podpisu
Certyfikat podpisującego

WAŻNY/PRAWIDŁOWY

2025-11-15 20:25:51 GMT+01:00
2025-11-17 09:15:05 GMT+01:00
PAdES
T (Podpis ze znacznikiem czasu)
kwalifikowany

Lista znaczników czasu

Walidacja znacznika czasu
Status certyfikatu
Data i czas wykonania podpisu
Podpisujący

WAŻNY/PRAWIDŁOWY

WAŻNY
2025-11-15 20:25:53 GMT+01:00
OID.2.5.4.97=VATPL-5170359458, CN=Certum QTST 2017, O=Asseco Data
Systems S.A., C=PL

Ścieżka certyfikacji

Certyfikat podpisującego

Walidacja certyfikatu
Status certyfikatu
Wystawiony dla
Wystawiony przez
Ważny od
Ważny do
Data wystawienia CRL/OCSP
Numer seryjny

WAŻNY/PRAWIDŁOWY

WAŻNY
C=PL, surname=Malawski, givenName=Maciej, CN=Maciej Malawski
CN=Certum QCA 2017, O=Asseco Data Systems S.A., C=PL
2025-06-12 08:06:45 GMT+02:00
2027-06-12 08:06:45 GMT+02:00
2025-11-17 09:05:03 GMT+01:00
312fc44023af9eb24d949bdb319f39bd

Certyfikat wystawcy

Walidacja certyfikatu
Status certyfikatu
Wystawiony dla
Wystawiony przez
Ważny od
Ważny do
Numer seryjny

WAŻNY/PRAWIDŁOWY

WAŻNY
CN=Certum QCA 2017, O=Asseco Data Systems S.A., C=PL
CN=Narodowe Centrum Certyfikacji, O=Narodowy Bank Polski, C=PL
2017-03-15 11:17:34 GMT+01:00
2028-03-16 00:59:59 GMT+01:00
7c8490c833d798a784dbfb588768233f6ebbf47f