

Zbyszko Królikowski
prof. dr hab. inż.
Instytut Informatyki
Politechniki Poznańskiej
email: Zbyszko.Krolikowski@cs.put.poznan.pl

Poznań, 01.08.2020 r.

Ocena

dorobku naukowego i rozprawy habilitacyjnej dr Zbigniewa Marszałka **p.t. *Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych*** ***w bazach NoSQL***

Niniejsza opinia została przygotowana w odpowiedzi na pismo R-WIMiL-BOD-512-18/19 prof. dr. hab. inż. Roberta Nowickiego, Kierownika ds. Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja, Wydziału Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej w Częstochowie w związku z powołaniem mnie przez Radę Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja na recenzenta w przewodzie habilitacyjnym dr Zbigniewa Marszałka. Moja opinia, przygotowana zgodnie z obowiązującymi przepisami, dotyczy oceny osiągnięcia naukowego habilitanta dokonanej na podstawie cyklu prac stanowiących rozprawę habilitacyjną pt. *Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych w bazach NoSQL*, udokumentowanej cyklem publikacji powiązanych tematycznie, oraz oceny jego aktywności naukowej.

Dr inż. Zbigniew Marszałek jest obecnie zatrudniony w Instytucie Matematyki Wydziału Matematyki Stosowanej Politechniki Śląskiej w Gliwicach na etacie starszego wykładowcy. Stopień doktora nauk matematycznych uzyskał w roku 1986 na Wydziale Matematyczno-Fizycznym Politechniki Śląskiej w Gliwicach przedstawiając rozprawę doktorską pt. *O pewnych residualnych metodach rozwiązywania dużych układów równań liniowych*.

Przedmiotem niniejszej oceny jest przedłożona mi dokumentacja obejmująca:

- A. Rozprawę habilitacyjną pt. „*Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych w bazach NoSQL*” w postaci 14 powiązanych tematycznie następujących prac, stanowiących podstawę do oceny osiągnięcia naukowego habilitanta:

- [1] Zb. Marszałek, M. Woźniak, D. Połap: *Fully flexible parallel merge sort for multi-core architectures*. Complexity (IF 1.829, 35 pkt według wykazu MNiSW), 2018, p-ISSN 1076-2787, e-ISSN: 1099-0526, Hindawi-John Wiley & Sons, Inc., DOI: 10.1155/2018/8679579.
- [2] Zb. Marszałek: *Parallelization of modified merge sort algorithm*. (IF 1.256, 30 pkt według wykazu MNiSW), 2017, Symmetry, vol. 9 no. 9, s. 1-18. p-ISSN: 2073-8994, MDPI - Multidisciplinary Digital Publishing Institute, DOI: 10.3390/sym9090176.
- [3] Zb. Marszałek: *The Analysis of Energy Performance in Use Parallel Merge Sort Algorithms*. Journal of Information Technology and Control (IF 0.707, 40 pkt według wykazu MNiSW z 2019), Vol. 48, no. 3, 2019 s. 487-498, DOI: 10.5755/j01.itc.48.3.23696.
- [4] Zb. Marszałek: *Parallel fast sort algorithm for secure multiparty computation*. Journal of Universal Computer Science (IF 1.066, 20 pkt według wykazu MNiSW z 2017), 2018 vol. 24 no. 4, s. 488-514, p-ISSN: 0948-695X, e-ISSN: 0948-6968.
- [5] M. Woźniak, Zb. Marszałek. *Extended algorithms for sorting large data sets*. Wydawnictwo Politechniki Śląskiej, 2014, Monografia (25 pkt według wykazu MNiSW); nr 526, ISBN: 978-83-7880-224-2.
- [6] M. Woźniak, Zb. Marszałek: *Selected algorithms for sorting large data sets*. Wydawnictwo Politechniki Śląskiej, 2013, Monografia (25 pkt według wykazu MNiSW); nr 499, ISBN: 978-83-7880-110-8.
- [7] Zb. Marszałek: *Performance tests on merge sort and recursive merge sort for big data processing*. Technical Science – Uniwersytet Warmińsko-Mazurski w Olsztynie (11 pkt według wykazu MNiSW), 2018 vol. 21 nr 1, s. 19-35, p-ISSN: 1505-4675, e-ISSN: 2083-4527.
- [8] Zb. Marszałek: *Performance test on triple heap sort algorithm*. Technical Science – Uniwersytet Warmińsko-Mazurski w Olsztynie (11 pkt według wykazu MNiSW), 2017 vol. 20 nr 1, s. 49-61, p-ISSN: 1505-4675, e-ISSN: 2083-4527.
- [9] Zb. Marszałek, G. Capizzi: *Modification of parallelization of modified merge sort algorithm*. 25th ICIST - International Conference on Information and Software Technologies 2019, Vilnius, Lithuania, October 10-12, 2019, Proceedings. Eds.

- Robertas Damasevicius and G. Vasiljevienė (Eds.): ICIST 2019, CCIS 1078, pp. 428–440, 2019, DOI: 10.1007/978-3-030-30275-7_33.
- [10] Zb. Marszałek: *Modification of parallelization for fast sort algorithm*. 24th ICIST - International Conference on Information and Software Technologies 2018, Druskininkai, Lithuania, October 4-6, 2018, Proceedings. Eds. Robertas Damasevicius, Vilma Mikasyte: Springer (15 pkt według wykazu MNiSW), 2018, s. 270-278, DOI: 10.1007/978-3-319-99972-2_21.
- [11] Zb. Marszałek: *Parallelization of fast sort algorithm*. 23rd ICIST - International Conference on Information and Software Technologies 2017, Druskininkai, Lithuania, October 12-14, 2017, Proceedings Eds. Robertas Damasevicius, Vilma Mikasyte. Cham: Springer (15 pkt według wykazu MNiSW), 2017, s. 408-421, DOI: 10.1007/978-3-319-67642-5_34.
- [12] Zb. Marszałek: *Novel recursive fast sort algorithm*. 22nd ICIST - International Conference on Information and Software Technologies 2016, Druskininkai, Lithuania, October 13-15, 2016, Proceedings Eds.: Giedre Dregvaite, Robertas Damasevicius. New York: Springer (15 pkt według wykazu MNiSW), 2016, s. 344-355, DOI: 10.1007/978-3-319-46254-7_27.
- [13] M. Woźniak, Zb. Marszałek, M. Gabryel, R. Nowicki. *Modified merge sort algorithm for large scale data sets*. 12th International Conference Artificial Intelligence and Soft Computing - ICAISC 2013, Zakopane, Poland, June 9-13, 2013. Proceedings Eds. L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L.A. Zadeh, J. M. Zurada. Berlin: Springer (15 pkt według wykazu MNiSW), 2013, s. 612-622, DOI: 10.1007/978-3-642-38610-7_56.
- [14] M. Woźniak, Zb. Marszałek, M. Gabryel, R. K. Nowicki. *Preprocessing large data sets by the use of quick sort algorithm*. Knowledge, information and creativity support systems: recent trends, advances and solutions. Selected papers from KICSS'2013 - 8th International Conference on Knowledge, Information, and Creativity Support Systems, November 7-9, 2013, Kraków, Poland. Eds. Andrzej M. J. Skulimowski, Janusz Kacprzyk. Berlin: Springer (15 pkt według wykazu MNiSW), 2016, s. 111-121, Springer; Scopus. DOI: 10.1007/978-3-319-19090-7_9.

B. Wybrane publikacje habilitanta

C. Inne dokumenty, w tym:

1. Autoreferat.
2. Wykaz osiągnięć naukowych dorobku habilitacyjnego, z uwzględnieniem cyklu powiązanych tematycznie publikacji.
3. Opublikowane prace naukowe po uzyskaniu stopnia doktora.
4. Wykaz informacji osobowych oraz informacji o osiągnięciach dydaktycznych, współpracy naukowej i popularyzacji nauki.

Problematyka rozprawy habilitacyjnej i ocena osiągnięć naukowych dr Zb. Marszałka

Tematyka badawcza rozprawy habilitacyjnej dra Zbigniewa Marszałka przedstawionej w postaci cyklu prac, stanowiąca podstawę oceny jego osiągnięcia naukowego, koncentruje się na zagadnieniach, które habilitant określił jako metody sortowania przez scalanie dużych zbiorów danych w bazach NoSQL. Celem prowadzonych badań było opracowanie nowych metod równoległego sortowania przez scalanie dedykowanych dla dużych wolumenów danych, przeprowadzenie i wykonanie testów opracowanych algorytmów, przeprowadzeniu dowodu teoretycznego ich złożoności obliczeniowej oraz dokonanie analizy porównawczej z innymi metodami sortowania.

Sortowanie to jeden z podstawowych, klasycznych problemów informatyki, polegający na uporządkowaniu zbioru danych względem pewnych cech charakterystycznych każdego elementu tego zbioru. Algorytmy sortowania są stosowane w celu uporządkowania danych i umożliwienia stosowania wydajniejszych algorytmów ich wyszukiwania, w szczególności w bazach danych oraz prezentacji danych w sposób czytelniejszy dla człowieka.

Bazy NoSQL zostały stworzone z potrzeby obsługiwanie bardzo dużych wolumenów danych. Bardzo istotnym katalizatorem rozwoju baz NoSQL była gwałtownie rosnąca popularność rozwiązań z obszaru Big Data. Wykorzystanie analiz bardzo dużych wolumenów danych stało się popularne, użyteczne i często po prostu konieczne. Bazy danych NoSQL doskonale wpisują się w trend Big Data, ponieważ – w przeciwieństwie do klasycznych systemów zarządzania relacyjnymi bazami danych – pozwalają na wyszukiwania oraz szybką analizę dużych zbiorów danych niestrukturyzowanych i badanie korelacji pomiędzy nimi.

Tak więc, zagadnienia sortowania dużych zbiorów danych należą do grupy problemów ważnych i jako takie mogą z pewnością stanowić przedmiot rozprawy habilitacyjnej.

Aktualność i ważność problematyki sortowania dużych wolumenów danych w bazach NoSQL oraz zakres badań podjętych przez dra Zbigniewa Marszałka w cyklu prac składających się na jego rozprawę habilitacyjną wskazują, że habilitant podjął badania w zakresie aktualnej, ważnej i ciekawej problematyki, stwarzające szansę na uzyskanie wartościowych wyników odpowiadających wymaganiom stawianym przy przyznawaniu stopnia doktora habilitowanego w naukach technicznych z zakresu informatyki.

Rozprawa habilitacyjna dra Zbigniewa Marszałka składa się z 14 prac. W przypadku 8 z nich habilitant jest jedynym autorem, w pozostałych 6 przypadkach habilitant jest jednym z współautorów. Cztery prace wchodzące w skład rozprawy zostały opublikowane w czasopismach posiadających IF z listy A MNiSzW [1, 2, 3, 4] (*Complexity, Symmetry, Journal of Information Technology and Control, Journal of Universal Computer Science* – wszystkie *Open Access Journals*). Pozostałe prace ukazały się jako wydawnictwa lokalne, tj. monografie (2) Wydawnictwa Politechniki Śląskiej [5, 6], artykuły (2) w czasopiśmie *Technical Science* Uniwersytetu Warmińsko-Mazurskiego w Olsztynie [7, 8], w materiałach konferencyjnych opublikowanych przez wydawnictwo Springer (konferencje: *ICIST – International Conference on Information and Software Technologies* (4 prace), *International Conference Artificial Intelligence and Soft Computing* (1 praca) oraz *International Conference on Knowledge, Information, and Creativity Support Systems* (1 praca)).

Rozprawa habilitacyjna stanowiąca podstawę oceny osiągnięcia naukowego jest poświęcona, jak już wspomniałem powyżej, sortowaniu dużych wolumenów danych w bazach NoSQL. W swoich badaniach habilitant zmodyfikował, znany z literatury algorytm sortowania przez scalanie proponując nowe podejście do liniowego scalania ciągów – umożliwia ono konstrukcję różnych wersji algorytmu sortowania wykorzystującego scalanie ciągów działających na współczesnych komputerach wieloprocesorowych. Rozprawa obejmuje następujące zagadnienia: rozszerzone algorytmy sortowania przez scalenie oraz kopcowanie, zwiększenie wydajności sortowania przez kopcowanie wielodzielne, nie rekursywne sortowanie przez scalenie dla dużych zbiorów danych oraz w pełni skalowalne równoległe sortowanie przez scalenie, w tym tzw. mocne sortowanie (ang. *fast sort*). W ramach omawianego cyklu prac habilitanta, przeprowadzono również testy opracowanych algorytmów oraz dowody ich złożoności obliczeniowej, a także wykonano analizę porównawczą z innymi metodami sortowania.

Dostarczona dokumentacja dorobku dr Zb. Marszałka, tj. wykaz osiągnięć naukowych i autoreferat nie zostały uporządkowane chronologicznie, lecz z uwzględnieniem wagi punktowej publikacji. Utrudnia to ocenę cyklu prac składających się na jego rozprawę habilitacyjną.

Zdaniem autora niniejszej recenzji, pierwszy krok na drodze do opracowania nowych równoległych metod sortowania przez scalenie dużych zbiorów danych stanowiła monografia [6] (M. Woźniak, Zb. Marszałek: *Selected algorithms for sorting large data sets*).

Wydawnictwo Politechniki Śląskiej, 2013). Zawiera ona analizę porównawczą wybranych algorytmów sortowania dużych zbiorów danych, tj. sortowania przez wstawianie oraz sortowania bąbelkowego i *quick sort*. Zaproponowano tutaj również pierwsze próby zwiększenia efektywności tych algorytmów dla dużych zbiorów danych przez wprowadzenie odpowiednich modyfikacji. W pracy [13] (M. Woźniak, Zb. Marszałek, M. Gabryel, R. Nowicki. *Modified merge sort algorithm for large scale data sets*. 12th International Conference Artificial Intelligence and Soft Computing - ICAISC 2013) zostały przedstawione pierwsze wyniki badań nad nie rekursywną wersją sortowania przez scalanie dla dużych zbiorów danych. Celem tej pracy było opracowanie i implementacja nie rekurencyjnego algorytmu sortowania przez scalanie oraz wykonanie jego badań. Sortowanie przez scalanie – rekurencyjny algorytm sortowania danych, stosujący metodę *dziel i zwyciężaj* był znany w literaturze od dawna. Algorytm sortowania przez scalanie został zmodyfikowany przez habilitanta w wyniku nowego podejścia do liniowego scalania ciągów. Zaproponowane to nowe podejście umożliwia konstrukcję różnych wersji algorytmu sortowania wykorzystującego scalanie ciągów działających na współczesnych komputerach wieloprocesorowych.

Monografia [5] (*Extended algorithms for sorting large data sets*, Wydawnictwo Politechniki Śląskiej, 2014) stanowi podsumowanie pewnego etapu badań habilitanta i jego współpracowników nad rozszerzonymi algorytmami sortowania przez scalenie oraz kopcowanie (inaczej sortowanie stogowe). Prace nad rozszerzonymi metodami sortowania przez scalenie zostały zainicjowane opracowaniem trójdzielonego algorytmu sortowania przez scalenie (Zb. Marszałek, D. Połap, M. Woźniak. *On preprocessing large data sets by the use of triple merge sort algorithm*. Proceedings of International Conference on Advances in Information Processing and Communication Technology. IPCT 2014, Rome, Italy, 2014) oraz algorytmu wykorzystującego *quick sort* [14] (M. Woźniak, Zb. Marszałek, M. Gabryel, R. K. Nowicki. *Preprocessing large data sets by the use of quick sort algorithm*. KICSS'2013 – 8th International Conference on Knowledge, Information, and Creativity Support Systems, 2013). W rzeczonyj monografii [5] zostały przedstawione efektywne metody sortowania dużych zbiorów danych, które wykorzystują wielodzielność (scalanie wielu ciągów). Można je zaimplementować w systemach wieloprocesorowych. Przedstawione w [5] wyniki badań pozwalają na określenie, jaka liczba scalanych ciągów pozwala osiągać optymalne wyniki. Opracowane metody sortowania opierały się na algorytmie wyboru najmniejszego elementu z wielu stosów. Szczególnie niską złożonością obliczeniową charakteryzowały się algorytmy

sortowania scalające dwa, trzy i cztery ciągi. W wyniku prac nad scalaniem trzech stosów dokonana została modyfikacja algorytmu scalania trzech ciągów, w ten sposób, że najpierw dokonuje się scalania pierwszych dwóch stosów do tablicy tymczasowej, a następnie scala się stos z tablicy tymczasowej z pozostałym trzecim stosem.

Celem pracy [1] (Zb. Marszałek, M. Woźniak, D. Połap: *Fully flexible parallel merge sort for multi-core architectures*. Complexity, 2018) było opracowanie i zaimplementowanie nowego algorytmu równoległego sortowania przez scalanie, przeprowadzenie i wykonanie testów, przeprowadzenie dowodu jego złożoności obliczeniowej oraz dokonanie analizy porównawczej z innymi metodami. Przedstawiona w tej pracy metoda umożliwia użycie dużej liczby procesorów – stąd jej nazwa: w pełni skalowalna równoległa metoda sortowania przez scalanie (ang. *fully flexible parallel merge sort*). Metoda ta jest skalowalna, to znaczy, dostosowuje się do liczby procesorów, na których sortowanie jest równoległe wykonywane. W zaproponowanej metodzie każdy krok został podzielony na dwa etapy. W skrócie, w kroku pierwszym dokonuje się scalania każdej pary ciągów do tablicy tymczasowej. Następnie scala się wszystkie pary posortowanych ciągów z tablicy tymczasowej do tablicy wyjściowej. Scalania ciągów jest wykonane za pomocą równoległego algorytmu sortowania przez scalanie. Ponieważ wszystkie procesory mogą jednocześnie odczytać pamięć i każdy procesor wykonuje wstawianie elementu do innej komórki pamięci, stąd cały proces scalania może być wykonany równoległe.

Praca [2] (Zb. Marszałek: *Parallelization of modified merge sort algorithm*, 2017) poprzedzała przedstawioną wyżej pracę [1] w dorobku habilitanta i stanowiła punkt wyjścia do opracowania wspomnianej metody *fully flexible parallel merge sort*. W ramach omawianej pracy habilitant określił sposób zrównoleglenia zmodyfikowanego algorytmu sortowania przez scalanie oraz wykonał analizę porównawczą z innymi metodami sortowania.

W pracach [4, 11] (Zb. Marszałek: *Parallel fast sort algorithm for secure multiparty computation*, Journal of Universal Computer Science, 2017) (Zb. Marszałek: *Parallelization of fast sort algorithm*, 23rd ICIST - International Conference on Information and Software Technologies 2017) przedstawiono zrównolegloną metodę tzw. mocnego sortowania (ang. *fast sort*). Prezentowany w pracy zrównoleglony algorytm mocnego sortowania o złożoności czasowej $O(n)$ przy użyciu $n/3$ procesorów, może być również wykorzystany przy ograniczonej liczbie procesorów. Zrównoleglenie sortowanych ciągów zostało wykonywane przy wykorzystaniu tymczasowej tablicy, rozpoczynając równoległe scalenie ciągów od indeksu pierwszego scalanego ciągu. Przeprowadzone badania dla 100 próbek losowych

ciągów, dla każdego z wymiarów 100, 1000, 10000, 100000, 1000000, 10000000 i 100000000 elementów na wejściu potwierdziły teoretyczną złożoność czasową i stabilność metody dla dużych zbiorów danych. Powiększenie zadania sortowania 10 krotnie zwiększa czas sortowania 10 razy dla dużych zbiorów danych. Badania wykazały, że metoda mocnego sortowania działa w krótszym czasie przy zastosowaniu wielu procesorów dla wymiarów zadania większego niż 10000. Zaproponowana metoda charakteryzuje się dużą stabilnością dla dowolnej liczby użytych procesorów.

Kolejna grupa, tj. [3, 7, 8] w ramach prac powiązanych tematycznie, stanowiących podstawę osiągnięcia naukowego habilitanta to prace poświęcone ocenie efektywności zaproponowanych nowych metod sortowania dużych wolumenów danych.

W pracy [3] (Zb. Marszałek: *The Analysis of Energy Performance in Use Parallel Merge Sort Algorithms*. Journal of Information Technology and Control, 2019) przeprowadzono analizę teoretyczną efektywności oraz praktyczną analizę porównawczą poboru energii podczas realizacji zrównoleglonej zmodyfikowanej metody sortowania przez scalanie i zrównoleglonej metody mocnego sortowania – przy czym jak już wspomniano powyżej, ta ostatnia metoda została przedstawiona w pracach [4, 11]. Proponowana metoda bazuje na modelu maszyny PRAM – *Parallel Random Access Machine*, która pozwala na swobodny dostęp do pamięci każdemu procesorowi przyłączonemu do szyny danych.

Celem pracy [7] (Zb. Marszałek: *Performance tests on merge sort and recursive merge sort for big data processing*. Technical Science – Uniwersytet Warmińsko-Mazurski w Olsztynie, 2018) była ocena efektywności dwóch wersji algorytmu sortowania przez scalanie, tj. z rekursją i bez. Praca licząca 35 stron w dużej mierze jest poświęcona omówieniu obu algorytmów, natomiast wyniki badań efektywnościowych – uwypuklonych w tytule pracy – przedstawiono na 3 stronach, przy czym są to w zasadzie badania identyczne w swoich założeniach jak te prezentowane w pracach [4 i 11]. Wyniki przeprowadzonych badań porównawczych metod nie rekursywnej i rekursywnej sortowania przez scalania wskazują, że wersja nie rekursywna jest o 60% szybsza od metody rekursywnej.

W pracy [8] (Zb. Marszałek: *Performance test on triple heap sort algorithm*. Technical Science – Uniwersytet Warmińsko-Mazurski w Olsztynie, 2017) przedstawiono wyniki badań nad możliwością zwiększenia wydajności sortowania przez kopcowanie wielodzielne oraz testy porównawcze nowego algorytmu sortowania przez kopcowanie z klasycznym algorytmem. Według habilitanta (strona 6 Autoreferatu), przeprowadzone badania wykazały, że algorytm kopcowania ma zbyt dużą stałą czasową i nie spełnia

pokładanych w nim nadziei. Autor dochodzi do tych wniosków w roku 2017 i według podanych przez niego informacji skupia się na algorytmach sortowania przez scalenie, a w pracy [13] z roku 2013 zostały już przedstawione pierwsze wyniki badań nad nie rekursywną wersją sortowania przez scalenie dla dużych zbiorów danych. Jest to zaskakująca niespójność.

Zbiór prac składających się na rozprawę habilitacyjną Pana dra Zbigniewa Marszałka jest tematycznie spójny i dotyczy problematyki sortowaniu dużych wolumenów danych. Uzyskane przez habilitanta wyniki w rozprawie habilitacyjnej uważam za wartościowe i ciekawe poznawczo. Za podstawową kontrybucję habilitanta należy uznać, w moim przekonaniu, następujące wyniki:

- (1) opracowanie nowych rozszerzonych wersji algorytmów sortowania przez scalenie oraz kopcowania ze zwiększoną wydajnością przez wykorzystanie kopcowania wielodzielnego;
- (2) opracowanie nowej metody nie rekursywnego sortowanie przez scalenie dla dużych zbiorów danych;
- (3) opracowanie w pełni skalowalnego algorytmu równoległego sortowania przez scalenie, w tym tzw. mocnego sortowanie (ang. *fast sort*);
- (4) przeprowadzenie dowodów złożoności obliczeniowej opracowanych algorytmów;
- (5) wykonanie analizy porównawczej efektywności zaproponowanych metod z innymi algorytmami.

Opracowane metody, w większości przypadków, jak wynika z przedstawionych prac, zostały zaimplementowane i zweryfikowane na drodze eksperymentów obliczeniowych. Obok skalowalności, zmniejszenie czasu realizacji czasu sortowania jest podstawową zaletą zaproponowanych metod dla dużych wolumenów danych. Co więcej, mają one z góry zdefiniowany czas działania i nie zawierają krytycznych ustawień ciągów. Na podstawie wyników badań empirycznych można stwierdzić, że zarówno mocne sortowanie jak i zmodyfikowana metoda algorytmu przez scalenie w wersji równoległej mogą być z powodzeniem stosowane do sortowania dużych wolumenów danych. Tak więc, proponowane metody mają walor nie tylko teoretyczny, ale mogą mieć również istotny walor praktyczny – jednak autor niniejszej recenzji postrzega ten walor inaczej niż habilitant.

Autor niniejszej opinii ma dwie uwagi w odniesieniu do przedstawionego cyklu prac składających się na rozprawę habilitacyjną. Pierwsza wątpliwość dotyczy właśnie aspektu praktycznego proponowanych rozwiązań. Tytuł ocenianego osiągnięcia naukowego –

przypomnijmy: *Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych w bazach No SQL* – odnosi proponowane w rozprawie rozwiązania do baz No SQL. Habilitant w autoreferacie stwierdza, cytując: „W celu dokonania wyboru algorytmu sortowania do zrównoleglenia obliczeń zostały przeprowadzone badania nad złożonością czasową używanych algorytmów w bazach No SQL”. Niestety w żadnej z prac habilitanta takiej analizy nie znalazłem. W pracach [7, 12] przedstawiono bardzo ogólną – niestety nie do końca przeze zrozumianą – architekturę przetwarzania Big Data w bazach NoSQL (w [7]: rys. 1, strona 21), jako możliwe miejsce wykorzystania proponowanych algorytmów. To samo w pracy [8] – rys. 1, strona 51. W [11] na rys. 1 strona 409, BigData i bazy NoSQL występują w powiązaniu z strukturami tabelarycznymi wykorzystywanymi w relacyjnych bazach danych, i ich implementacją na poziomie fizycznym w postaci baz kolumnowych.

W pracach [1, 2, 3] przyjmuje się model maszyny PRAM (*Parallel Random Access Machine*), jako podstawę równoległych realizacji proponowanych algorytmów, a w całym cyklu, (w szczególności w [4, 5, 6, 9, 10]) wspomina się o możliwości wykorzystania tych algorytmów w bazach NoSQL, przyjmując jednocześnie model maszyny wieloprocessorowej PRAM, który w tych bazach wykorzystywany nie jest.

Bazy NoSQL to systemy rozproszone wykorzystujące jako architekturę klastry serwerów (bazy klucz-wartość, rodzina kolumn bazy). W bazach tych nie występuje schemat danych oraz nie ma wymagania aby obiekty pamiętane w bazie były jednorodne pod względem struktury. Bazy NoSQL są bardzo szybkie w kontekście przetwarzania Big Data, m.in. dlatego, że jedynymi operacjami oferowanymi w odniesieniu do danych, jest ich zapis (operacja PUT) lub wyszukanie za pomocą klucza i jednego żądania GET. W takim środowisku (czytaj w bazach NoSQL) proponowanie takich czy innych algorytmów sortowania jest nieuzasadnione. Jedynym systemem, w którym na poziomie systemowym wykonywana jest operacja sortowania jest MongoDB – klauzula *\$orderby* w języku zapytań opartym na notacji JSON. Można natomiast rozważać wykorzystanie proponowanych algorytmów sortowania na poziomie aplikacji, które z systemem bazy danych NoSQL współpracują, bowiem w tych systemach interpretacja zawartości bazy jest bowiem zadaniem aplikacji. Odrębną kwestią jest to, czy te aplikacje będą pracować na maszynach o modelu PRAM. Wątpliwości budzi również rozmiar danych analizowanych w eksperymentach przeprowadzonych przez habilitanta – 100 mln elementów w sortowanym zbiorze danych to jeszcze nie Big Data. Przyjmuje się, że w zależności od branży Big Data może oznaczać rozmiar terabajtów (10^{12} bajtów) lub petabajtów (10^{15}). Podsumowując, tytuł ocenianego osiągnięcia naukowego *Nowe równoległe metody sortowania przez scalanie dużych zbiorów*

danych w bazach No SQL jest nieuzasadniony – jest swoistym nadużyciem metodycznym. Jak już to tytuł ten powinien brzmieć: *Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych*.

Druga uwaga, de facto marginalna, odnosi się do wspomnianej powyżej niespójności w chronologii badań i wniosków z nich wynikających.

Powyższe uwagi, które mają drugorzędny charakter w stosunku do oceny merytorycznej rozprawy habilitacyjnej, nie wpływają na moją pozytywną ocenę rozprawy habilitacyjnej jako oryginalnego osiągnięcia naukowego, stanowiącego znaczący wkład w rozwój problematyki sortowania dużych wolumenów danych.

Ocena dorobku naukowego nie objętego rozprawą habilitacyjną, tzn. ocena aktywności naukowej dr Zbigniewa Marszałka

Prace dra Zbigniewa Marszałka, nie wchodzące w skład ocenianego osiągnięcia naukowego, obejmują nieco szerszą tematykę niż te ujęte w ramach tego osiągnięcia. Nie mniej, zdecydowana większość prac dotyczy, głównie, problematyki sortowania dużych zbiorów danych. Do tego nurtu zaliczyłbym prace habilitanta dotyczące nie rekursywnej wersji trójdzielnego algorytmu sortowania przez kopcowanie, szybkiego sortowania z losowym wyborem elementu środkowego, sortowania bąbelkowego i algorytmu sortowania przez przestawianie. Ponadto, dr Zbigniew Marszałek posiada w swoim dorobku prace dotyczące: przetwarzania dźwięków i obrazów oraz programowania w języku C++.

Dr Zbigniew Marszałek jest autorem (współautorem) 29 publikacji (w tym 14 prac złożonych jako rozprawa habilitacyjna) po uzyskaniu stopnia doktora. W chwili składania dokumentów przez habilitanta wskaźniki dotyczące IF, indeksu Hirscha i cytowań były następujące:

- (1) suma wskaźników IF wynosi 8.698,
- (2) łączna suma punktów MNiSzW dla osiągnięcia naukowego wynosi 140 (2019),
- (3) liczba cytowań wg. Web of Science: 79, Scopus: 150, Google Scholar: 235.

Najczęściej cytowaną publikacją dr Zb. Marszałka jest *“Modified Merge Sort Algorithm for Large Scale Data Sets”* (21 razy). Z 15 prac nie objętych rozprawą habilitacyjną dwie zostały opublikowane w czasopismach posiadających IF z listy A MNiSzW (*Sensors, Journal of Universal Computer Science – Open Access Journals*).

Pozostałe prace ukazały się jako wydawnictwa lokalne, tj. *Int. J. Math. Comput. Sci.* 2014, Zeszyty Naukowe i skrypty Wydawnictwa Politechniki Śląskiej oraz w materiałach konferencyjnych lub warsztatów międzynarodowych: *Asia-Pacific Conference on Computer-Aided System Engineering. APCASE 2015, International Conference on Advances in Information Processing and Communication Technology - IPCT 2014, KICSS'2013 International conference, ICIST 2014 IEEE Symposium Series on Computational Intelligence, IEEE SSCI 2014 Federated Conference on Computer Science and Information Systems, 2018*).

Te dane wskazują na aktywną działalność naukową i publikacyjną dra Zbigniewa Marszałka w okresie ostatnich 6 lat jego pracy zawodowej – stopień doktora nauk matematycznych habilitant uzyskał w roku 1986.

Podsumowując, dorobek naukowy dra Zbigniewa Marszałka, stanowiący podstawę do oceny jego aktywności naukowej, jest w miarę różnorodny i dotyczy dyscypliny informatyki (metody sortowania dużych wolumenów danych oraz przetwarzanie dźwięków i obrazów). Generalnie, dorobek naukowy Pana dr Zbigniewa Marszałka oceniam pozytywnie. Spełnia on, w moim przekonaniu w podstawowym zakresie, kryterium aktywności naukowej, na co składa się w szczególności wartość merytoryczna i oryginalność prac.

Trzeba jednak zaznaczyć, że znaczna część prac stanowiących dorobek habilitanta ukazała się w czasopiśmie z grupy *Open Access Journals* oraz materiałach konferencji informatycznych o relatywnie niskiej randze. To w moim przekonaniu zmniejsza szanse habilitanta na upowszechnienie jego wyników i automatycznie zmniejsza potencjalną liczbę przyszłych cytowań (indeks H).

Powyższe uwagi krytyczne mają charakter ogólny i nie dotyczą wartości merytorycznej prac Pana dr Zbigniewa Marszałka. Odnoszą się one raczej do „sposobu” uprawiania nauki i upowszechniania wyników swoich prac, i nie wpływają na moją pozytywną ocenę dorobku naukowego Pana dr Zbigniewa Marszałka.

Działalność dr Zbigniewa Marszałka na rzecz środowiska naukowego, jego pozycja w środowisku naukowym oraz współpraca z otoczeniem społeczno – gospodarczym

Dr Zbigniew Marszałek wykonał 50 recenzji dla wielu międzynarodowych czasopism naukowych, m. in., *Sensors, Symmetry, Algorithms, Entropy, Electronics, International Journal of Distributed Systems*. Wygłosił 11 referatów naukowych na międzynarodowych i krajowych konferencjach naukowych: *24-th International Conference on Information and Software Technologies; International Conference ICIST 2014, 2016, 2017, 2018, 2019, FedCSIS 2018, Symposium for Young Scientists in Technology, Engineering and Mathematics - SYSTEM 2016, International Conference on Advances in Information Processing and Communication Technology. IPCT 2014, Rome, Italy, IEEE Symposium Series on Computational Intelligence - IEEE SSCI 2014, Orlando, USA*.

Był członkiem komitetu organizacyjnego *International Conference on Information Technology – IVUS 2017* oraz przewodniczącym sesji na *Intelligent Systems and Software Engineering Advances at ICIST 2016, 2018, 2019, VI Konferencja Zastosowań Matematyki w Technice, Informatyce i Ekonomii ZAMA '16*.

Innym aspektem działalności dr Zbigniewa Marszałka jest jego udział jako wykonawcy w realizacji uczelnianych projektów badawczych Politechniki Śląskiej tj. *Rektorski Grant Projakościowy II stopnia - 2018, Rektorski Grant Habilitacyjny - 2018, Uczelniany program SYNERGIA na Politechnice Śląskiej w ramach projektu NAWA*.

Jest członkiem *Polskiego Towarzystwa Matematycznego*. W latach 2018-19 dwukrotnie odbywał staże naukowe na University of Catania, Department of Electronics, Engineering and Informatics, Catania, Italy, jako “*invited research professor*”. Dr Zb. Marszałek był ekspertem *Polskiej Agencji Przedsiębiorczości* w latach 2009 - 2015 – w ramach tej działalności wykonał recenzje wielu projektów międzynarodowych i krajowych, które zostały zrealizowane na terenie Unii Europejskiej i Polski. Był i jest nadal organizatorem *Ogólnopolskiego Konkursu Wiedzy Matematyczno - Informatycznej Algorytmion* (od roku 2010).

Wniosek końcowy

Podsumowując, uważam, że zarówno osiągnięcie naukowe dr Zbigniewa Marszałka, ocenione na podstawie zgłoszonego przez habilitanta zbioru 14 powiązanych tematycznie publikacji „*Nowe równoległe metody sortowania przez scalanie dużych zbiorów danych w bazach NoSQL*”, jak i jego aktywność naukowa, oceniona na podstawie prac nie wchodzących w skład rozprawy habilitacyjnej, spełniają wymogi stawiane przez Ustawę

z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce*, odnośnie nadawania stopnia naukowego doktora habilitowanego. W związku z powyższym wnoszę o przyjęcie przez Radę Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Wydziału Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej w Częstochowie przedłożonej rozprawy habilitacyjnej, w postaci cyklu powiązanych publikacji, i dopuszczenie dr Zbigniewa Marszałka do dalszych etapów postępowania habilitacyjnego.

J. Kwiek