

Autoreferat

Zastosowanie użytkowe technik informatycznych i uczenia maszynowego w nowoczesnych formach komunikacji

Spis treści

1. DANE HABILITANTA.....	3
1.1. IMIĘ I NAZWISKO.....	3
1.2. POSIADANE DYPLOMY, STOPNIE	3
1.3. DOTYCHCZASOWE ZATRUDNIENIE	3
2. WSKAZANIE OSIĄGNIĘCIA WYNIKAJĄCEGO Z ART. 16 UST. 2 USTAWY Z DNIA 14 MARCA 2003 R. O STOPNIACH NAUKOWYCH I TYTULE NAUKOWYM ORAZ O STOPNIACH I TYTULE W ZAKRESIE SZTUKI	4
2.1. TYTUŁ OSIĄGNIĘCIA NAUKOWEGO	4
2.2. WYKAZ MONOTEMATYCZNYCH PUBLIKACJI WCHODZĄCYCH W SKŁAD OSIĄGNIĘCIA NAUKOWEGO:.....	4
2.3. OMÓWIENIE PRAC ZWIĄZANYCH ZE WSKAZANYM OSIĄGNIĘCIEM	7
2.3.1. <i>Wprowadzenie</i>	7
2.3.2. <i>Charakterystyka prac składających się na osiągnięcie naukowe</i>	8
2.3.3. <i>Omówienie osiągnięcia naukowego</i>	13
2.3.4. <i>Wykaz monotematycznych publikacji habilitanta sprzed obrony doktoratu</i>	18
3. OSIĄGNIĘCIA I DOROBEK NAUKOWO-BADAWCZY	21
3.1. AUTORSTWO I WSPÓLAUTORSTWO PUBLIKACJI.....	21
3.2. LICZBA CYTOWAŃ, INDEKS HIRSCHA.....	22
3.3. KRAJOWE I MIĘDZYNARODOWE PROJEKTY BADAWCZE	24
4. DZIAŁALNOŚĆ DYDAKTYCZNA I DOROBEK POPULARYZATORSKI	25
4.1. DZIAŁALNOŚĆ DYDAKTYCZNA.....	25
4.2. UDZIAŁ W KONFERENCJACH.....	28
4.3. DZIAŁALNOŚĆ RECENZENCKA I REDAKCYJNA.....	29
4.4. OPRACOWANE ROZWIĄZANIA, PROGRAMY I PATENTY	30
4.5. WSPÓLPRACA Z INNYMI NAUKOWCAMI	32
4.6. DZIAŁANIA POPULARYZATORSKIE	33
4.7. NAGRODY, GRANTY I WYRÓŻNIENIA	35

1. Dane habilitanta

1.1. Imię i nazwisko

Krzysztof Wołk

1.2. Posiadane dyplomy, stopnie

- Doktor nauk informatycznych, Polsko-Japońska Akademia Technik Komputerowych, Warszawa, 10.2016
- Studia doktoranckie, Bioinformatyka, Polsko-Japońska Akademia Technik Komputerowych, Warszawa, 02.2016
- Magister inżynier, wydział informatyki, katedra multimediiów, Polsko-Japońska Akademia Technik Komputerowych, Warszawa, 07.2011
- Inżynier, wydział informatyki, katedra multimediiów, Polsko-Japońska Akademia Technik Komputerowych, Warszawa, 02.2010

1.3. Dotychczasowe zatrudnienie

- 2017.02-obecnie, adiunkt, Polsko-Japońska Akademia Technik Komputerowych, katedra Multimediiów, wydział informatyki, Warszawa;
- 2009-obecnie, redaktor serwisu internetowego IN4.pl
- 2018-obecnie, redaktor serwisu internetowego PCLAB.pl
- 2009-2017.02, asystent, Polsko-Japońska Akademia Technik Komputerowych, katedra Multimediiów, wydział informatyki, Warszawa;
- 2014, DlaKadr.pl, trener zewnętrzny, szkoleniowiec.
- 2013-2016, asystent, Warszawska Szkoła Fotografii i Grafiki Projektowej, Warszawa;
- 2010, Agora SA, trener zewnętrzny, szkoleniowiec;
- 2010, Poczta Polska SA, trener zewnętrzny, szkoleniowiec;
- 2009-2012, własna działalność gospodarcza z branży IT.

2. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki

2.1. Tytuł osiągnięcia naukowego

Zastosowanie użytkowe technik informatycznych i uczenia maszynowego w nowoczesnych formach komunikacji

2.2. Wykaz monotematycznych publikacji wchodzących w skład osiągnięcia naukowego:

- 1) Wołk, K. [100%], (2019). Machine Learning in Translation Corpora Processing. CRC Press. Monografia 25pkt
- 2) Wołk, K. [70%], Zawadzka-Gosk, E [20%], Czarnowski, W. [10%], Deep learning and sub-word-unit approach in written art generation. In World Conference on Information Systems and Technologies (pp. 367-377). Springer, Cham. Indeksowana w Web of Science MNiSW: 15pkt
- 3) Zawadzka-Gosk, E. [50%], Wołk, K. [40%], Czarnowski, W. [10%], Deep learning in State-of-The-Art image classification exceeding 99% accuracy. In World Conference on Information Systems and Technologies (pp. 367-377). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
- 4) Wołk, K. [100%], (2018). Emergency, pictogram-based augmented reality medical communicator prototype using precise eye-tracking technology. Cyberpsychology, Behavior and Social Networking. <https://doi.org/10.1089/cyber.2018.0035>, Czasopismo z Listy A MNiSW 35pkt.
- 5) Wołk, K. [80%], & Marasek, K. [20%] (2018, September). Survey on Neural Machine Translation into Polish. In International Conference on Multimedia and Network Information System (pp. 260-272). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
- 6) Zawadzka-Gosk, E. [60%], & Wołk, K. [40%] (2018). Semantic-enabled Hybrid Genetic Disease Diagnostics in Next-Generation Sequenced Data. Computer Science, 19(2), 179. Czasopismo z listy B MNiSW 12pkt
- 7) Wołk, K. [100%] (2018, March). Mixing Textual Data Selection Methods for Improved In-Domain Data Adaptation. In World Conference on Information

- Systems and Technologies (pp. 367-377). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
- 8) Wołk, K. [60%], Glinkowski, W. [20%], & Żukowska, A. [20%] (2018, March). Enhancing the Assessment of (Polish) Translation in PROMIS Using Statistical, Semantic, and Neural Network Metrics. In World Conference on Information Systems and Technologies (pp. 351-366). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
 - 9) Wołk, K. [80%], & Wołk, A. [20%] (2018, March). Augmenting SMT with Semantically-Generated Virtual-Parallel Corpora from Monolingual Texts. In World Conference on Information Systems and Technologies (pp. 358-374). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
 - 10) Wołk, K. [70%], Zawadzka, E. [20%], & Wołk, A. [10%] (2018, March). Statistical Approach to Noisy-Parallel and Comparable Corpora Filtering for the Extraction of Bi-lingual Equivalent Data at Sentence-Level. In World Conference on Information Systems and Technologies (pp. 797-812). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
 - 11) Wołk, K. [80%], & Wołk, A. [20%] (2017). Machine enhanced translation of the Human Phenotype Ontology project. *Procedia Computer Science*, 121, 11-18 Indeksowana w Web of Science, MNiSW: 15pkt
 - 12) Wołk, K. [60%], Wołk, A. [20%], Marasek, K. [10%], & Glinkowski, W. [10%] (2017). Pictogram-based mobile first medical aid communicator. *Procedia Computer Science*, 121, 3-10. Indeksowana w Web of Science, MNiSW: 15pkt
 - 13) Zawadzka, E. [60%], Wołk, K. [30%], & Marasek, K. [10%] (2017). Automated Medical Articles Analysis for Human Genetic Diagnostic. *Procedia Computer Science*, 121, 797-802. Indeksowana w Web of Science, MNiSW: 15pkt
 - 14) Wołk, K. [70%], Wołk, A. [20%], & Glinkowski, W. [10%] (2017) A cross-lingual mobile medical communication system prototype for foreigners and subjects with speech, hearing, and mental disabilities based on pictograms., *Computational and Mathematical Methods in Medicine*, Hindawi. Czasopismo z listy A MNiSW 15pkt
 - 15) Wołk, K. [70%], Wołk, A. [20%], & Marasek K. [10%] (2017). Big Data Language Model of Contemporary Polish., *Federated Conference on Computer Science and Information Systems – LTA*, Prague, Czech Republic, 2017. Indeksowana w Web of Science, MNiSW: 15pkt

- 16) Wołk, K. [60%], Wołk, A. [25%], & Marasek K. [15%] (2017). Unsupervised tool for quantification of progress in L2 English phraseological., Federated Conference on Computer Science and Information Systems – LTA, Prague, Czech Republic, 2017. Indeksowana w Web of Science, MNiSW: 15pkt
- 17) Wołk, K. [70%], Korzinek, D. [30%] (2017). Comparison and Adaptation of Automatic Evaluation Metrics for Quality Assessment of Re-Speaking., Journal of Computer Science, vol 18 no 2, p. 129-144. Czasopismo z listy B MNiSW 12pkt
- 18) Korzinek, D. [40%], Marasek, K. [20%], Brocki, Ł. [20%], & Wołk, K. [20%] (2017). Polish Read Speech Corpus for Speech Tools and Services. Linköping Electronic Conference Proceedings, 136, 54 – 62. MNiSW: 5pkt
- 19) Wolk, K. [70%], & Marasek, K. [30%] (2017). PJIIT's systems for WMT 2017 Conference. In Proceedings of the Second Conference on Machine Translation (pp. 416-421). MNiSW: 5pkt
- 20) Wołk, K. [80%], & Wołk, A. [20%] (2017, April). Augmenting SMT with Generated Pseudo-parallel Corpora from Monolingual News Resources. In World Conference on Information Systems and Technologies (pp. 308-316). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
- 21) Wołk, K. [80%], & Wołk, A. [20%] (2017, April). Automatic Parallel Data Mining After Bilingual Document Alignment. In World Conference on Information Systems and Technologies (pp. 317-327). Springer, Cham. Indeksowana w Web of Science, MNiSW: 15pkt
- 22) Wołk, K. [70%], Korzinek, D. [20%], & Marasek, K. [10%] (2017). Semi-automatic and Human-Aided Translation Evaluation Metric (HMEANT) for Polish Language in Re-speaking and MT Assessment. In Multimedia and Network Information Systems (pp. 241-249). Springer International Publishing. Indeksowana w Web of Science, MNiSW: 15pkt
- 23) Wołk, A. [60%], Wołk, K. [30%], & Marasek, K. [10%] (2017). Analysis of Complexity Between Spoken and Written Language for Statistical Machine Translation in West-Slavic Group. In Multimedia and Network Information Systems (pp. 251-260). Springer International Publishing. Indeksowana w Web of Science, MNiSW: 15pkt
- 24) Wołk, K. [60%], Wołk, A. [20%], & Marasek, K. [20%] (2017). Implementing Statistical Machine Translation into Mobile Augmented Reality Systems. In

- Multimedia and Network Information Systems (pp. 61-73). Springer International Publishing Indeksowana w Web of Science, MNiSW: 15pkt
- 25) Wołk, K. [80%], & Marasek, K. [20%] (2017). Unsupervised Construction of Quasi-comparable Corpora and Probing for Parallel Textual Data. In Multimedia and Network Information Systems (pp. 307-320). Springer International Publishing. Indeksowana w Web of Science, MNiSW: 15pkt
- 26) Wołk, K. [80%], Marasek, K. [10%], & Wołk, A. [10%] (2016, September). Exploration for Polish-* bi-lingual translation equivalents from comparable and quasi-comparable corpora. In Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on (pp. 517-525). IEEE. Indeksowana w Web of Science, MNiSW: 15pkt

2.3. Omówienie prac związanych ze wskazanym osiągnięciem

2.3.1. Wprowadzenie

Tłumaczenie w czasie rzeczywistym, umożliwiające komunikację na całym świecie opiera się na interakcji zachodzącej między systemem a użytkownikiem. Możliwość tłumaczenia zróżnicowanych dokumentów wymaga odpowiedniego treningu systemu, modyfikacji danych i adaptacji dziedziny tłumaczeń.

Opracowanie metody pozwalającej na rozszerzenie funkcjonalności systemu daje możliwość zastosowania rozwiązania w różnych środowiskach. Połączenie metod i technik tłumaczenia maszynowego ze światem nowoczesnych technologii pozwala na tłumaczenie w czasie rzeczywistym z dokładnością umożliwiającą płynną komunikację między użytkownikiem posługującym się technologią a odbiorcą informacji. Technologia wspierana przez tłumaczenie maszynowe kreuje coraz bardziej rozbudowane narzędzia umieszczone w aplikacjach lub innych programach. W szczególności mowa tu o systemach tłumaczących mowę na mowę. Nie można także zapomnieć o innych technikach komunikacji międzyludzkich oraz komunikacji człowieka z komputerem w przypadkach różnego rodzaju chorób oraz dysfunkcji. W obu przypadkach olbrzymi wpływ na jakość działania tego typu systemów ma nie tylko sama technika i proces treningowy, ale przede wszystkim tzw. pre-processing danych treningowych, a niekiedy także post-edycje. Dla jakości ważne jest także posiadanie dobrej jakości, akuratnych technik automatycznej lub półautomatycznej ewaluacji. Wszystkie te zagadnienia zostały poruszone w kontekście języka polskiego oraz języków słowiańskich w ramach opisywanego osiągnięcia, będącego tematem tego autoreferatu.

Metoda badania i analizy korpusów porównywalnych stała się fundamentem pracy nad korpusami treningowymi (porównywalnymi i quasi-porównywalnymi), co w konsekwencji prowadziło do analizy metod statystycznych oraz metod opartych o sieci neuronowe wykorzystywanych w tłumaczeniu i uczeniu maszynowym. Te techniki pozwoliły także nie tylko na budowę korpusów równoległych na podstawie już istniejących danych, ale także na sztuczne rozszerzenie ich poprzez modele generatywne.

Zastosowanie opracowanych metod i technik charakterystycznych dla tłumaczenia i uczenia maszynowego w obszarze nowoczesnych technologii przy przetwarzaniu języka naturalnego pozwala na uzyskanie znacznie lepszych rezultatów od tych, które osiągane są bez użycia tych rozwiązań, jednocześnie zapewniając wyższą efektywność w kontekście zarówno nauki jak i biznesu.

Innym zagadnieniem związanym z komunikacją było opracowanie metody komunikacji bazującej jedynie na ruchu gałek ocznych oraz położeniu głowy. Wykorzystuje ono specjalnie zaprojektowane piktogramy, co pozwala osobom z niepełnosprawnościami ruchowymi oraz częściowo-sparaliżowanym na komunikację i interakcję z otoczeniem bez udziału osób trzecich. Sprawdza się ono także w przypadku chorób neurologicznych i zaburzeniach autystycznych.

Podsumowując stworzenie elastycznych metod, narzędzi, które mogą zostać wykorzystane w zróżnicowanych środowiskach pozwala na wielosegmentową adaptację tego typu rozwiązań. Z kolei praca nad przetwarzaniem języka naturalnego, skoncentrowana wokół dużych zbiorów danych o różnych charakterach daje możliwość poznania różnego rodzaju prawidłowości warunkujących działanie konkretnego algorytmu w konkretnym zastosowaniu. Modyfikacje jednostkowe niejednokrotnie prowadzą do stworzenia modelu sieci powiązań, która odpowiednio przeszkolona byłaby zdolna odpowiadać na odmienne potrzeby znacznej części dyscyplin naukowych.

2.3.2. Charakterystyka prac składających się na osiągnięcie naukowe

Monotematyczne publikacje wchodzące w skład osiągnięcia naukowego przedstawione zostały w układzie chronologicznym malejącym. Publikacja 1 jest rezultatem wcześniejszych dokonań i odkryć w dziedzinie tłumaczenia i uczenia maszynowego. Opisuje w sposób komplementarny współczesne techniki uczenia maszynowego używane w przetwarzaniu języka naturalnego. W szczególności mowa tu o tłumaczeniu maszynowym opartym o sieci neuronowe i techniki statystyczne, lingwistykę korpusową, techniki pre-processingu dużej ilości zasobów tekstowych, a z techniki automatycznej ewaluacji wyników tłumaczenia maszynowego. Opisuje także

potencjalne możliwości zastosowania tychże technik w innych dziedzinach nauki na przykład medycynie i genetyce.

Chronologiczny układ prac związanych ze wskazanym osiągnięciem jest obrazem rozwoju koncepcji i ewolucji badań. W wyniku prowadzonych analiz dochodziło systematycznie do rozszerzenia obszaru zastosowania projektowanych rozwiązań. Eksperymenty w tłumaczeniu maszynowym przeprowadzone przez habilitanta w 2016 roku doprowadziły do opracowania metody budowy i przeszukiwania korpusów porównywalnych. Odnajdywanie i ekstrakcja danych równoległych na poziomie zdania z danych wysoce nierównoległych takich jak korpusy porównywalne doprowadziło do rozszerzenia pola analiz o korpusy quasi-porównywalne. W roku 2017 eksperymenty w tłumaczeniu maszynowym przeprowadzone przez habilitanta pozwoliły na opracowanie efektywnej metody budowy oraz przeszukiwania korpusów quasi-porównywalnych oraz udostępnienie tychże technik w formie usługi internetowej w ramach projektu Clarin-PL.

Opracowane metody tłumaczeń maszynowych zostały przeniesione na pole mobilnych systemów rozszerzonej rzeczywistości, zwiększając zakres funkcjonalności zarówno technik jak i metod wypracowanych na potrzeby ich działania. Publikacja *Implementing Statistical Machine Translation into Mobile Augmented Reality Systems* wskazuje na rozwiązania, których dostarczają metody tłumaczenia maszynowego usprawniające procesy komunikacyjne. Interakcja użytkownik-system jest jedną z głównych interakcji występujących w XXI wieku. Możliwość tłumaczenia tekstu przy użyciu smartfona jest niezwykle pomocną funkcją, wykorzystywaną przez miliony użytkowników na całym świecie. Zakres oferowanych usług tłumaczeniowych musi być bardzo szeroki – od języka ogólnego, przez wyrażenia specjalistyczne, aż do złożonych konstrukcji technicznych. Użytkownik musi mieć możliwość tłumaczenia tekstu ze zdjęć wykonanych urządzeniem. Rozszerzenie zakresu translatoryki maszynowej o pracę w czasie rzeczywistym daje możliwość tłumaczenia dokumentów prawnych, medycznych, technicznych czy specjalistycznych z języka obcego na język użytkownika w czasie rzeczywistym.

Nowoczesne systemy tłumaczenia maszynowego (MT) nie opierają się na regułach charakterystycznych dla określonych języków, jednakże odpowiednie przeszkolenie statystycznego systemu tłumaczenia maszynowego (SMT) umożliwia poprawę jakości tłumaczeń. Poprawa jakości tłumaczenia jest jednym z warunków podniesienia poziomu komunikatywności użytkownika. Interakcja zachodząca pomiędzy tłumaczącym a systemem pozwala na rozszerzenie procesu komunikacji lub jego usprawnienie. Dobra

jakość tłumaczenia może oznaczać m.in. lepszą komunikację między pacjentem a lekarzem.

Odpowiedni trening systemów tłumaczenia maszynowego, wymagający zaawansowanej analizy różnic w budowie języka pozwala na uzyskanie wyższej jakości przekładu. Trening SMT i rozszerzenie funkcjonalności w kontekście języków słowiańskich, które są językami morfologicznie złożonymi w porównaniu z językiem angielskim daje możliwości opisane w publikacji *Analysis of Complexity Between Spoken and Written Language for Statistical Machine Translation in West-Slavic Group*. Habilitant, odpowiedzialny za przygotowanie systemu dokonał automatycznego przekładu próbek tłumaczeń, z rezultatem wskazanym w publikacji. Na tej podstawie zostały wysnute wnioski oraz zaproponowane adaptacje dla języków słowiańskich.

Doskonalenie systemu tłumaczeń pozwala na usprawnienie poziomu jakości procesu, przekładającego się na efektywną komunikację między dwoma osobami posługującymi się odmiennymi językami. Stworzenie miary oceny jakości tłumaczenia bazującej na rozwiązaniu pół-automatycznym typu HMEANT dla języka polskiego udowodniło skuteczność tego typu narzędzi, polegającą m.in. na rozwiązywaniu wielu zawiloci językowych oraz dwuznaczności. Przedstawienie i omówienie zagadnień związanych z oceną jakości tłumaczeń dla języka polskiego wraz z badaniami wykorzystującymi wskazane narzędzie stanowiło przedmiot pracy *Semi-automatic and Human-Aided Translation Evaluation Metric (HMEANT) for Polish Language in Re-speaking and MT Assessment*.

Skuteczność metod pół-automatycznych pozwoliła na wskazanie metody automatycznego pozyskiwania danych równoległych na poziomie zdania, co znalazło odzwierciedlenie w pracy *Automatic Parallel Data Mining After Bilingual Document Alignment*. Wcześniejsze prace, których współautorem jest habilitant, rozwijały i doskonaliły metody stosowane w poprawie procesu na polu tłumaczenia i uczenia maszynowego. Opracowana przez habilitanta metoda budowy oraz przeszukiwania korpusów porównywalnych (2016) i quasi-porównywalnych (2017) ewoluowała na potrzeby kolejnych projektów, rozszerzając zakres swojej funkcjonalności i stanowiąc punkt odniesienia dalszych prac przy posiadanym doświadczeniu i wynalezionych rozwiązaniach.

Trenowanie systemów i analizy możliwości generowania syntetycznych zasobów językowych doprowadziły do rozwiązania problemu braku dostępności tychże zasobów wśród języków mniej popularnych, które nie posiadają wystarczającej ilości jednostek językowych dostępnych dla naukowców. W pracy *Augmenting SMT with Generated*

Pseudo-parallel Corpora from Monolingual News Resources opisana została próba stworzenia syntetycznych zasobów językowych równoległych na poziomie zdania przy użyciu analizy semantycznej.

Prace nad modelowaniem i przetwarzaniem języka naturalnego doprowadziły do wypracowania sposobu modelowania języka polskiego technikami statystycznymi, co przedstawione zostało w *Big Data Language Model of Contemporary Polish* oraz *Polish Read Speech Corpus for Speech Tools and Services*. Prace stanowią przejaw zainteresowania habilitanta procesami modelowania i przetwarzania języka naturalnego, co zostało docenione i wyróżnione podczas **FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS – LTA 2017**. Wypracowane rozwiązania dały podłoże badań nad systemami respeakingu, które to badania zakończyły się opracowaniem przez habilitanta modelu dokonującego trafnych ocen doboru właściwego znaczenia do usłyszanego słowa oraz samych respeakerów. Zagadnienie to szerzej zostało omówione w publikacji *Comparison and Adaptation of Automatic Evaluation Metrics for Quality Assessment of Re-Speaking*, opublikowanej w 2017 roku na łamach *Journal of Computer Science*.

Modelowanie języka naturalnego poprzez trenowanie modeli języka daje możliwość stworzenia zastosowań bardziej rozwiniętych. Przykładem są prace nad systemem autonomicznej ewaluacji przyswajania przez obcokrajowców nowych kompetencji frazeologicznych, opisane w *Unsupervised tool for quantification of progress in L2 English phraseological competence*.

Coraz większe możliwości rozwoju stwarzane przez nowoczesne technologie wyposażone w odpowiedniej jakości oprogramowanie wymagają odpowiednich narzędzi, wspierających rozwój jednostki na różnym poziomie. Zaprojektowanie rozwiązania bazującego na piktogramach, służącego do komunikacji wyłącznie w domenie medycznej, stało się jednym z osiągnięć habilitanta w 2017 roku, opisanych w publikacjach: *A cross-lingual mobile medical communication system prototype for foreigners and subjects with speech, hearing, and mental disabilities based on pictograms* oraz *Pictogram-based mobile first medical aid communicator*.

Publikacje *Automated Medical Articles Analysis for Human Genetic Diagnostic* oraz *Semantic-enabled Hybrid Genetic Diagnostic in Next-Generation Sequenced Data*, prezentują sposób redukcji ilości hipotez w diagnostyce genetycznej do poziomu weryfikowalnego dla samego lekarza. Narzędzie to usprawnić ma proces diagnostyki, stanowiąc jednocześnie wyraz wpływu metod i technik wypracowanych w uczeniu maszynowym na rozwój innych dyscyplin naukowych. Artykuł *Machine enhanced*

translation of the Human Phenotype Ontology project podkreślając niebagatelne znaczenie projektu *Human Phenotype Ontology* porusza problematykę wsparcia tłumaczy i recenzentów przy użyciu technik tłumaczenia maszynowego.

Prawidłowości zachodzące pomiędzy poszczególnymi sekwencjami stanowią podstawę konstrukcji modelu statystycznego o różnym przeznaczeniu. Jedną z proponowanych funkcjonalności jest analiza danych, pozwalających na określenie poziomu ilościowego i jakościowego zasobów w danym języku. Analiza treści, pozwalająca na wykluczenie błędów jakościowych bądź innych nieprawidłowości stanowiła podstawę opracowania statystycznego modelu weryfikacji, co przedstawione zostało w pracy *Statistical Approach to Noisy-Parallel and Comparable Corpora Filtering for the Extraction of Bi-lingual Equivalent Data at Sentence-Level*.

Wielopoziomowa weryfikacja zasobów językowych, pozwalająca na osiągnięcie założonych rezultatów może zostać uzupełniona o analizę semantyczną, która pozwoli na wygenerowanie nowych danych, stanowiących zasoby językowe. Habilitant jest autorem systemu tłumaczenia maszynowego oraz dwóch algorytmów analizy semantycznej opartych o LSA oraz WordNet, pozwalających na generowanie nowych danych, co zaprezentowane zostało w pracy *Augmenting SMT with Semantically-Generated Virtual-Parallel Corpora from Monolingual Texts*.

Prace z modelami weryfikującymi dane pozwoliły na stworzenie systemu tłumaczenia maszynowego przyspieszającego prace i redukującego koszty, co przedstawione zostało w publikacji: *Enhancing the Assessment of (Polish) Translation in PROMIS Using Statistical, Semantic, and Neural Network Metrics*

Wzajemna konwergencja technologii, komunikacji i metod przetwarzania języka naturalnego, w celach usprawnienia samej komunikacji, pozwoliły habilitantowi na stworzenie specjalistycznego rozwiązania usprawniającego komunikację w pewnych grupach osób z niepełnosprawnościami ruchowymi.

Problematyka komunikacji osób niepełnosprawnych ruchowo i częściowo sparaliżowanych wydaje się trudno-rozwiązywalnym zagadnieniem. Niedowład kończyn, niemożność poruszania rękami to częste bariery uniemożliwiające tym osobom komunikację. Nierzadko częściowa niezdolność do używania aparatu mowy to problem, którego nie da się rozwiązać bez wsparcia technologicznego. Publikacja *Emergency, pictogram-based augmented reality medical communicator prototype using precise eye-tracking technology* prezentuje zaprojektowany przez habilitanta komunikator bazujący wyłącznie na ruchach gałek ocznych oraz położeniu głowy.

Prowadzone badania i analizy, zarówno przed obroną doktoratu, jak i po uzyskaniu tytułu naukowego doktora pozwoliły habilitantowi na wypracowanie wielu rozwiązań w dziedzinie przetwarzania i modelowania języka naturalnego, które usprawniają funkcjonalność takich sektorów jak medycyna czy nowoczesne technologie. Statystyczne metody zarządzania danymi oparte na są na logicznych prawidłowościach, dzięki czemu powstające algorytmy pozwalają na gromadzenie, przetwarzanie i zarządzanie danymi. Prace nad jednostkami mniejszymi niż słowa naturalnie prowadziły do rozszerzenia pola testowania określonych systemów. Wstępne analizy potwierdziły wpływ metod i technik trenowania systemów na późniejsze ich granice. Dalsze badania potwierdziły, że granice poprawności systemów zależne są od poziomu elastyczności parametrów, co z kolei przyczyniło się do opracowania modeli usprawniających trenowanie określonych systemów w różnych dziedzinach.

W oparciu o proces samodoskonalenia, stanowiący wynik badań i analiz habilitant wypracował narzędzia przydatne zarówno w lingwistyce, humanistyce cyfrowej, medycynie jak i komunikacji. Nowoczesne technologie wspierane właściwie i z funkcjonalnie zaprojektowanym oprogramowaniem mogą nie tyle usprawnić określone procesy, ile podnieść poziom jakości życia, o czym świadczyć może chociażby zaprojektowany komunikator dla osób z problemami ruchowymi.

2.3.3. Omówienie osiągnięcia naukowego

Habilitant jest autorem metod pozwalających na pozyskiwanie i analizę korpusów porównywalnych i quasi-porównywalnych. Współautor publikacji naukowych z zakresu tłumaczenia i uczenia maszynowego oraz monografii poświęconych multimediom i systemom informatycznym odpowiedzialny jest również m.in. za trenowanie systemów, modelowanie języka polskiego technikami uczenia maszynowego, opracowanie modelu dokonującego trafnych ocen re-speakingu czy wprowadzanie implementacji rozwiązań. Wykaz osiągnięć habilitanta przedstawiony został poniżej. Za punkt wyjścia przyjęto rok 2016 po obronie doktoratu:

2016:

1. Opracowanie metody budowy i przeszukiwania korpusów porównywalnych oraz jej udoskonalenie i udostępnienie w ramach narzędzia opensource;

2017:

2. Opracowanie metody budowy i przeszukiwania korpusów quasi-porównywalnych celem wydobycia danych równoległych na poziomie zdania i udostępnienie w formie narzędzia opensource i usługi www;
3. Opracowanie i implementacja metody rozpoznawania obrazu w czasie rzeczywistym przez smartfona w celu stworzenia rozwiązania pozwalającego na tłumaczenie na język ojczysty tekstu, z wykorzystaniem technologii rzeczywistości rozszerzonej;
4. Opracowanie metody i adaptacja parametrów treningu systemów tłumaczenia maszynowego w celu generowania lepszych rezultatów w przekładach między językami słowiańskimi, analiza złożoności problemu i porównanie z przekładem na język angielski;
5. Implementacja pół-automatycznego narzędzia opartego o miarę HMEANT dla języka polskiego, rozwiązującego wiele zawłości językowych i dwuznaczności w czasie ewaluacji tłumaczeń, zapewniającego znacznie większą korelację z ludzką oceną w stosunku do metryk automatycznych np. BLEU. Narzędzie zostało udostępnianie w formie opensource;
6. Opracowanie nowych modeli (dla innych języków europejskich w parze z polskim) dla metody automatycznego pozyskiwania danych równoległych na poziomie zdania, udostępnienie narzędzia oraz modeli online;
7. Opracowanie metody generowania syntetycznych zasobów językowych równoległych na poziomie zdania za pomocą analizy semantycznej wywodzących się z danych monolingwalnych;
8. Opracowanie systemów tłumaczenia maszynowego oraz ich zgłoszenie do międzynarodowej ewaluacji jakości tłumaczenia maszynowego dla kilku par języków.
9. Udostępnienie narzędzi oraz zasobów językowych wypracowanych w ramach projektu Clarin-PL, w szczególności modelu języka polskiego ograniczonego dziedzinowo dla mowy;
10. Stworzenie modelu automatycznej ewaluacji re-speakingu, silnie skorelowanego z ludzką oceną;
11. Opracowanie narzędzia dla języka polskiego i angielskiego oceniającego w sposób automatyczny i nienadzorowany kompetencje frazeologiczne osób władających językiem obcym;
12. Opracowanie współczesnego modelu języka polskiego, opartego o zasoby CommonCrawl oraz jego udostępnienie w ramach opensource. Wyniki publikacji

- zostały wyróżnione podczas konferencji 2nd International Workshop on Language Technologies and Applications (LTA);
13. Opracowanie aplikacji na zegarki inteligentne, opartej o piktogramy, będącej komunikatorem medycznym związanym z pierwszą pomocą przeznaczonym dla obcokrajowców, wymagających błyskawicznego wywiadu i pomocy medycznej. Adaptacja aplikacji na potrzeby osób z dysfunkcjami;
 14. Opracowanie metody interpolacji oraz filtracji wyników priorytetyzacji genów celem trafniejszej diagnostyki genetycznej poprzez ograniczenie liczby hipotez;
 15. Udział w projekcie Human Phenotype Ontology poprzez implementację i adaptację systemu tłumaczenia maszynowego celem przyspieszenia realizacji projektu oraz obniżenia jego kosztów.

2018:

16. Opracowanie metody poprawy jakości korpusów równoległych, automatycznie pozyskanych z danych porównywalnych i quasi-porównywalnych poprzez identyfikację i filtrację błędnych fraz za pomocą technik uczenia maszynowego;
17. Opracowanie metody oraz narzędzia pozwalającego za pomocą analizy semantycznej i danych monolingwalnych budować korpusy równoległe celem poprawy jakości tłumaczenia maszynowego dla par języków ubogich w tego typu zasoby lingwistyczne;
18. Opracowanie systemu tłumaczenia maszynowego zaadaptowanego do tłumaczenia tekstów medycznych w ramach projektu PROMIS celem obniżenia kosztów jego realizacji. System tłumaczenia maszynowego został zaimplementowany jako kolejny czynnik decyzyjny w metodologii tłumaczenia FACIT;
19. Opracowanie metodologii domenowej adaptacji danych dla języka polskiego zapewniającej lepszą jakość niż powszechnie stosowana interpolacja liniowa i logarymiczna;
20. Opracowanie metody automatycznej ekstrakcji i analizy danych w artykułach medycznych celem poprawy jakości i szybkości diagnozowania chorób genetycznych;
21. Opracowanie autorskiej metody podziału na jednostki mniejsze niż słowa, wykorzystywane w procesach tłumaczenia maszynowego oraz ich udostępnienie w formie narzędzia opensource;
22. Dokonanie systematycznego przeglądu dostępnych technik tłumaczenia maszynowego w oparciu o sieci neuronowe, ich analiza, przegląd oraz adaptacja na

potrzeby przekładu związanego z językiem polskim jako źródłem i celem tłumaczenia;

23. Opracowanie metody komunikacji bazującej jedynie na ruchu gałek ocznych oraz położeniu głowy. Wykorzystuje ono specjalnie zaprojektowane piktogramy, co pozwala osobom z niepełnosprawnościami ruchowymi oraz częściowo-sparalizowanym na komunikację i interakcję z otoczeniem bez udziału osób trzecich;

2019

24. Poprawa metody podziału tekstu na jednostki mniejsze niż słowa, uruchomienie go jako usługi narzędzia online oraz zastosowanie w dziedzinie generowania poezji i wykazanie, iż zapewnia ona rezultaty lepsze niż podział na znaki oraz byte pair encoding (bpe);

Charakterystyka statystyczna osiągnięć habilitanta, uzupełniona o opracowania dostępne w postaci publikacji przedstawiona została w Tabeli 1.

Tabela 1 Charakterystyka statystyczna osiągnięć habilitanta wyrażona w publikacjach

Osiągnięcie (nr)	Publikacja (nr)
1	26
2	25
3	24
4	23
5	22
6	21
7	20
8	19
9	18
10	17
11	16
12	15
13	14, 12
14	13
15	11
16	10
17	9
18	8
19	7
20	6
21, 22	5
23	4

	24	2
Suma:	24	24

Na 26 publikacji naukowe o charakterze badawczo-rozwojowym habilitant zyskał 24 osiągnięcia, skupione wokół konstruowania modeli pozwalających na rozwiązywanie problemów z zakresu tłumaczenia i uczenia maszynowego. Jedno z kluczowych odkryć habilitanta dotyczyło rozszerzenia funkcjonalności urządzeń mobilnych wyposażonych w wytrenowane systemy tłumaczenia maszynowego oraz systemy komunikacji oparte o piktogramy, co przyczyniło się do stworzenia rozwiązania pozwalającego na komunikację między pacjentami częściowo sparaliżowanymi lub z niepełnosprawnością ruchową. Prowadzone przez habilitanta analizy przyczyniły się do poszukiwania rozwiązań wspierających zarządzanie systemami komunikacji w różnych dziedzinach naukowych, takich jak genetyka czy medycyna.

Na płaszczyźnie poprawy zdolności komunikacyjnych jednostki posługującej się technologią wyposażoną w wytrenowane systemy tłumaczenia maszynowego habilitant odpowiedzialny był za implementację rozwiązania opartego o modele częstotliwościowe oraz miary mierzące kolokacje językowe. Podjęte działania służyły potwierdzeniu korelacji między skutecznością opracowanej metody ilościowej a oceną nauczyciela. Habilitant zajmował się badaniami nad sposobami zdobywania przez mówiących w obcym języku nowych kompetencji frazeologicznych, co przyczyniło się do dalszych analiz procesu komunikacji wśród różnych grup docelowych.

Wsparcie, którego dostarczać może metodologia tłumaczenia i uczenia maszynowego różnym gałęziom i dyscyplinom naukowym omówione zostało m.in. na przykładzie specjalistycznego komunikatora wykorzystywanego w medycynie oraz udziale w projektach około medycznych. Badania przeprowadzone przez habilitanta dowiodły zawodności większości dostępnych na rynku narzędzi tłumaczenia mowy, co doprowadziło do opracowania specjalistycznego komunikatora medycznego opartego na piktogramach. Każda przeprowadzona przez autora publikacji analiza miała na celu określenie możliwości rozwoju danej dziedziny naukowej, gałęzi czy dyscypliny bądź sektora przy uwzględnieniu metod wykorzystywanych w tłumaczeniu maszynowym.

Opracowanie modeli tłumaczenia maszynowego specyficznych dla diagnostyki genetycznej pozwoliło na wyeliminowanie większości błędów, przyczyniając się do zmniejszenia kosztów samych tłumaczeń i usprawnienia procesu leczenia. Skomplikowane terminologie medyczne wymagały narzędzia wykorzystującego techniki tłumaczenia maszynowego, co doprowadziło do rozszerzenia zakresu badań i założeń,

że techniki tłumaczenia maszynowego są w stanie poprawić jakość usług w wieku sektorach, wspierając jednocześnie procesy tłumaczenia i komunikacji w różnorodnych dyscyplinach i sub-gałęziach.

Spośród 26 prac habilitanta szczególną wartość odkrywczą posiadało 16 z nich, pozostałe koncepcje stanowią punkt odniesienia do przyszłych analiz i badań.

2.3.4. Wykaz monotematycznych publikacji habilitanta sprzed obrony doktoratu

1. Wołk, K., & Marasek, K. (2016). Translation of Medical Texts using Neural Networks. *International Journal of Reliable and Quality E-Healthcare (IJRQEH)*, 5(4), 51-66. MNiSW: 5pkt
2. Wołk, K., & Marasek, K. (2016). PJAiT Systems for the WMT 2016. In *Proceedings of the First Conference on Machine Translation*, Berlin, Germany. Association for Computational Linguistics. MNiSW: 5pkt
3. Wołk K., Rejmund E., Marasek K., „Multi-domain machine translation enhancements by parallel data extraction from comparable corpora. Poprawa jakości tłumaczenia maszynowego dla wielu domen poprzez ekstrakcję danych równoległych z korpusów porównywalnych.”, *Polskojęzyczne korpusy równoległe. Polish-language Parallel Corpora*. Warszawa: Instytut Lingwistyki Stosowanej. ISBN: 978-83-935320-4, 2016, p. 157-179 MNiSW: 2,5pkt
4. Wołk A., Wołk K., Marasek K., „Enhancements in Statistical Spoken Language Translation by De-normalization of ASR Results”, *Journal of Computers*, January 2016 Volume 11 Number 1, ISSN: 1796-203X, p. 33-40, International Academy Publishing, 2016, MNiSW: 5pkt
5. Wołk K., Marasek K., “Unsupervised comparable corpora preparation and exploration for bi-lingual translation equivalents”, *Proceedings of the 12th International Workshop on Spoken Language Translation*, Da Nang, Vietnam, December 3-4, 2015, p.118-125, MNiSW: 5pkt
6. Wołk K., Marasek K., “PJAiT Systems for the IWSLT 2015 Evaluation Campaign Enhanced by Comparable Corpora.”, *Proceedings of the 12th International Workshop on Spoken Language Translation*, Da Nang, Vietnam, December 3-4, 2015, p.101-104, MNiSW: 5pkt
7. Wołk K., Rejmund E., Marasek K., „Harvesting comparable corpora and mining them for equivalent bilingual sentences using statistical classification and analogy-based heuristics”, *Proceedings of 22nd International Symposium ISMIS 2015 Lyon, France*, October 21-23, 2015, Foundations of Intelligent System, p. 433-441, ISSN 0302-9743,

- ISSN 1611-3349 (electronic), Springer, Indeksowana w Web of Science MNiSW: 15pkt
8. Wołk K., „Noisy-Parallel and Comparable Corpora Filtering Methodology for the Extraction of Bi-Lingual Equivalen Data at Sentence Level.”, Journal of Computer Science, ISSN: 1508-2806, 2015, 16.2: 169-184. Czasopismo z Listy B MNiSW 12pkt
 9. Wołk K., Marasek K., Glinkowski W., „Telemedicine as a special case of Machine Translation.” Computerized Medical Imaging and Graphics, ISSN: 0895-6111, 2015. Czasopismo z Listy A MNiSW 25pkt
 10. Wołk K., Marasek K., „Neural-based Machine Translation for Medical Text Domain. Based on European Medicines Agency Leaflet Texts.”, Procedia Computer Science, ISSN: 1877-0509, 2015, 64: 2-9. Indeksowana w Web of Science MNiSW: 15pkt
 11. Wołk K., Marasek K., „Tuned and GPU-accelerated Parallel Data Mining from Comparable Corpora”, Lecture Notes in Artificial Intelligence, p. 32 – 40, ISBN: 978-3-319-24032-9, Springer, 2015, Indeksowana w Web of Science MNiSW: 15pkt
 12. Wołk K., Marasek K., “Polish – English Speech Statistical Machine Translation Systems for the IWSLT 2014.”, Proceedings of the 11th International Workshop on Spoken Language Translation, Tahoe Lake, USA, 2014, p. 143-149, MNiSW: 5pkt
 13. Wołk K., Marasek K., „Building Subject-aligned Comparable Corpora and Mining it for Truly Parallel Sentence Pairs.”, Procedia Technology, 18, Elsevier, p.126-132, 2014, Indeksowana w Web of Science MNiSW: 15pkt
 14. Marasek K., Wołk K., Korzinek D., Brocki Ł., “Spoken Language Translation for Polish”, Proceedings of Forum Acusticum 2014, Kraków, MNiSW: 5pkt
 15. Wołk K., Marasek K., “Polish -English Statistical Machine Translation of Medical Texts”, New Research in Multimedia and Internet Systems, Springer. 09/2014, ISSN: 1867-5662, Indeksowana w Web of Science MNiSW: 15pkt
 16. Wołk K., Marasek K., “Enhanced Bilingual Evaluation Understudy”, Lecture Notes on Information Theory, ISSN: 2301-3788, 2014, MNiSW: 5pkt
 17. Wołk K., Marasek K., “A Sentence Meaning Based Alignment Method for Parallel Text Corpora Preparation.”, Advances in Intelligent Systems and Computing volume 275, p.107-114, Publisher: Springer, ISSN 2194-5357, ISBN 978-3-319-05950-1, Indeksowana w Web of Science MNiSW: 15pkt
 18. Wołk K., Marasek K., “Real-Time Statistical Speech Translation”, Advances in Intelligent Systems and Computing volume 275, p.107-114, Publisher: Springer,

ISSN 2194-5357, ISBN 978-3-319-05950-1, Indeksowana w Web of Science
MNiSW: 15pkt

19. Wołk K., Marasek K., „Alignment of the Polish-English Parallel Text for a Statistical Machine Translation.”, Computer Technology and Application 4, Publisher: David Publishing, ISSN:1934-7332 (Print), ISSN: 1934-7340 (Online), p. 575-583, 2013
Czasopismo z Listy B MNiSW 7pkt
20. Wołk K., Marasek K., “Polish – English Speech Statistical Machine Translation Systems for the IWSLT 2013.”, Proceedings of the 10th International Workshop on Spoken Language Translation, Heidelberg, Germany, p. 113-119, 2013, MNiSW: 5pkt
21. Wołk K., “Polish to English Statistical Machine Translation.”, XV International Phd Workshop OWD 2013, Wisła, p.108-115, 2013, MNiSW: 5pkt

3. Osiągnięcia i dorobek naukowo-badawczy

3.1. Autorstwo i współautorstwo publikacji

Habilitant jest autorem 26 publikacji opublikowanych w latach 2016-2019 (okres po doktoracie), z których 19 indeksowanych jest w bazie Web of Science, 2 pochodzą z recenzowanych międzynarodowych konferencji naukowych, 2 zostały opublikowane w czasopiśmie z listy B, 2 w czasopiśmie z listy A (wg ministerialnego wykazu czasopism) oraz 1 monografii, wykorzystujących wiedzę techniczną z zakresu konstrukcji systemów teleinformatycznych. Charakterystyka publikacji z lat 2016-2018 przedstawiona została poniżej w postaci tabelarycznej (Tabela 2).

Tabela 2 Publikacje z lat 2016-2018

Rodzaj publikacji	Liczba	Punkty MNISW
Indeksowane w bazie WoS	19	285
Monografie, podręczniki	1	25
Pokonferencyjne (Proceedings - recenzowane)	2	10
Lista czasopism B	2	24
Lista czasopism A	2	50
Suma	26	394

W okresie przed doktoratem habilitant był autorem 21 publikacji, opublikowanych w latach 2013-2016, z których 7 indeksowanych jest w bazie Web of Science, 10 pochodzi z recenzowanych międzynarodowych konferencji naukowych, 2 zostały opublikowane w czasopiśmie z listy B, 1 w czasopiśmie z listy A (wg ministerialnego wykazu czasopism) oraz 1 monografii, wykorzystujących wiedzę techniczną z zakresu konstrukcji systemów teleinformatycznych. Charakterystyka publikacji z lat 2013-2016 przedstawiona została poniżej w postaci tabelarycznej (Tabela 3).

Tabela 3 Publikacje z lat 2013-2016

Rodzaj publikacji	Liczba	Punkty MNISW
Indeksowane w bazie WoS	7	105
Monografie, podręczniki	1	2.5
Pokonferencyjne (Proceedings - recenzowane)	10	50
Lista czasopism B	2	19
Lista czasopism A	1	25

Suma	21	201.5
------	----	-------

W całym okresie pracy badawczo-naukowej habilitanta opublikował on łącznie 47 publikacji w latach 2013-2019, z których 26 indeksowanych jest w bazie Web of Science, 12 pochodzi z recenzowanych międzynarodowych konferencji naukowych, 4 zostały opublikowane w czasopiśmie z listy B, 3 w czasopiśmie z listy A (wg ministerialnego wykazu czasopiśm) oraz 2 monografie, wykorzystujących wiedzę techniczną z zakresu konstrukcji systemów teleinformatycznych. Charakterystyka publikacji z lat 2013-2019 przedstawiona została poniżej w postaci tabelarycznej (Tabela 4).

Tabela 4 Publikacje z lat 2013-2019

Rodzaj publikacji	Liczba	Punkty MNISW
Indeksowane w bazie WoS	26	390
Monografie, podręczniki	2	27,5
Pokonferencyjne (Proceedings - recenzowane)	12	60
Lista czasopiśm B	4	43
Lista czasopiśm A	2	75
Suma	47	595.5

3.2. Liczba cytowań, indeks Hirscha

Mierniki statystyczne, pozwalające na ocenę jakości materiału oraz jego popularności jako dzieła o charakterze naukowym, skupiające się wokół wartości opartych na wskaźniku cytowań, indeksie Hirscha oraz wskaźniku i-10. Wartości mają charakter ogólny, informujący o liczbie cytowań wszystkich opublikowanych przez autora prac. Punktem odniesienia statystyk jest baza danych Google Scholar.

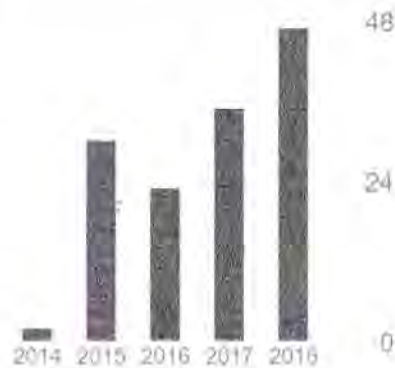
Ilustracja 1. Zrzut ekranowy Google Scholar na dzień 02.01.2019

	Wszystkie	Od 2014
Cytowania	146	146
h-indeks	6	6
i10-indeks	4	4

Tendencja wzrostowa liczby cytowań w badanym okresie świadczy o wysokim znaczeniu przeprowadzonych badań i analiz, prezentowanych w publikacjach habilitanta. Zaprezentowana ona została na ilustracji 2, na której to widać, iż w roku 2014 liczba

cytowań wynosiła 2, w 2015 wynosiła 30, w 2016 wynosiła 23, w 2017 wynosiła 35, a w 2018 wynosiła 47.

Ilustracja 2. Zrzut ekranowy Google Scholar przedstawiająca liczbę cytowań na dzień 02.01.2019.



Autocytowania uwypuklają kierunek badań autora, o czym świadczyć może publikacja habilitanta z 2016, zamieszczona w *Multimedia and Network Information Systems*, cytowana dwukrotnie (2017-2018) podczas zbliżonych jakościowo badań i analiz.

Podobne tendencje można zaobserwować w innych bazach danych. W portalu ResearchGate habilitant posiada „RG score” równy 17.09. Z kolei ilość cytowań wynosi w tej bazie 113, liczba czytań artykułów 21348, a lista rekomendacji wynosi 41. Zgodnie z portalem ResearchGate h-index wynosi 6, po usunięciu autocytowań wynosi 3.

Z kolei w bazie danych Scopus system nie zaindeksował wszystkich publikacji habilitanta, a jedynie 30. Mimo niepełnych informacji h-index wynosi 3.

Ilustracja 3. Statystyki na portalu ResearchGate na dzień 02.01.2019.



3.3. Krajowe i międzynarodowe projekty badawcze

W okresie po obronie doktoratu habilitant aktywnie uczestniczył i wciąż uczestniczy w projekcie Clarin. **Clarin-PL** – Common Language Resources & Technology Infrastructure jest projektem ogólnoeuropejskiej infrastruktury naukowej, w ramach której udostępniane są narzędzia badawcze dla głównie humanistów m. in. umożliwiające pracę z bardzo dużymi zbiorami tekstów. Habilitant w ramach projektu odpowiedzialny jest za dostarczanie rozwiązań umożliwiających automatyczne przetwarzane języka naturalnego, w szczególności w budowie korpusów oraz modelowaniu języka. Wypracowane w projekcie rozwiązania mają przysłużyć się wszystkim dyscyplinom naukowym, ze szczególnym uwzględnieniem nauk humanistycznych i społecznych.

Habilitant jest kierownikiem grantu obliczeniowego numer WCSS#27202666 przyznanego przez Wrocławskie Centrum Sietkowo-Superkomputerowe.

Habilitant obecnie jest kierownikiem B+R, projektu naukowego realizowanego przez firmę Nextsell pt. „Nextsell 2.0 - Opracowanie systemu informatycznego opartego m.in. na technologii uczenia maszynowego, wspierającego małe i średnie przedsiębiorstwa w sprzedaży ich produktów na międzynarodowych platformach e-

commerce (tzw. marketplace'ach)". Firma na chwile obecną stara się o dofinansowanie ze środków NCBiR w ramach tzw. „szybkiej ścieżki”. Przeszedł on wymogi formalne a stan rozpatrywania projektu na chwile obecną jest na etapie spotkania panelowego.

Z kolei w okresie przed doktoratem habilitant uczestniczył w projekcie Eu-Bridge 7th FR EU (grant agreement n°287658), w którym to zajmował się badaniami nad systemami tłumaczenia maszynowego pomiędzy językiem polskim a językami europejskimi. Wyniki tychże badań były co roku ewoluowane podczas konferencji IWSLT. Habilitant w tym okresie był także beneficjentem stypendium z grantu Interkadra UDA- POKL-04.01.01-00-014/10-00.

4. Działalność dydaktyczna i dorobek popularyzatorski

4.1. Działalność dydaktyczna

W ramach pracy dydaktycznej w okresie po obronie doktoratu habilitant prowadzi i prowadził zajęcia z 17 przedmiotów (wykłady oraz ćwiczenia) na wydziale Informatyki oraz Sztuki Nowych mediów na Polsko-Japońskiej Akademii Technik Komputerowych oraz Warszawskiej Szkole Fotografii i Grafiki Projektowej. Zajęcia prowadzone przez habilitanta skupiały się wokół tematów związanych z Grafiką Komputerową, Multimediami, Warsztatami Komputerowymi przygotowującymi do posługiwania się programami marki Adobe. Były to następujące przedmioty:

- Grafika Komputerowa
- Interakcja Człowiek Komputer
- Uniwersalne Techniki Programowania
- Multimedia 1
- Multimedia 2
- Grafika Multimedialna
- Laboratorium Nowych Mediów 1
- Laboratorium Nowych Mediów 2
- Laboratorium Nowych Mediów 3
- Laboratorium Nowych Mediów 4
- Kompozycja Proceduralna 1
- Kompozycja Proceduralna 2
- Warsztaty Komputerowe 1

- Warsztaty Komputerowe 2
- Podstawy informatyki dla projektantów
- Podstawy informatyki
- Zaawansowane Techniki Multimedialne

Habilitant był promotorem technicznym 40 prac magisterskich oraz 106 prac licencjackich. W roli promotora pomocniczego doktorantów opiekuje się dwiema osobami studiującymi w Polsko-Japońskiej Akademii Technik Komputerowych, a także jest mentorem pięciorga doktorantów z czego czworo jest studentami interdyscyplinarnych studiów Humanistyka Cyfrowa Instytutu Badań Literackich Polskiej Akademii Nauk i Polsko-Japońskiej Akademii Technik Komputerowych, a jeden Międzyuczelnianych Interdyscyplinarnych Studiów Doktoranckich - Psychologia & Informatyka (MISD - P&I) - ICT & Psychology w Polsko – Japońskiej Akademii Technik Komputerowych we współpracy z SWPS Uniwersytetem Humanistycznospołecznym.

Charakterystyka działalności dydaktycznej przedstawiona została w Tabeli 5. Tabela 6 zawiera specyfikę działalności dydaktycznej z wyszczególnieniem m.in. takich aspektów jak rodzaj zajęć, czas realizacji, czy rodzaj jednostki naukowej.

Tabela 5 Charakterystyka działalności dydaktycznej i naukowej po obronie doktoratu (2016-obecnie)

Charakter działalności	Ilość (os., h., szt.)
Zajęcia dydaktyczne	2,445h
Opieka naukowa nad studentami (wszystkie)	153 os.
Recenzje międzynarodowych referatów konferencyjnych oraz artykułów do czasopism	46 szt.
Programy, patenty	1*
Publikacje książkowe (wszystkie)	10

Tabela 6 Specyfika działalności dydaktycznej w Polsko-Japońskiej Akademii Technik Komputerowych po obronie doktoratu (2016-obecnie)

Nazwa przedmiotu	Charakter (Wyk, ćw, lab.)	Ilość godzin
Grafika Komputerowa	Ćwiczenia	460
Interakcja Człowiek Komputer	Ćwiczenia	367
Uniwersalne Techniki Programowania	Ćwiczenia	150
Multimedia 1	Wykład + ćwiczenia	16 + 127
Multimedia 2	Wykład + ćwiczenia	16 + 58
Grafika Multimedialna	Ćwiczenia	96
Laboratorium Nowych Mediów 1	Laboratorium	60

Laboratorium Nowych Mediów 2	Laboratorium	60
Laboratorium Nowych Mediów 3	Laboratorium	76
Laboratorium Nowych Mediów 4	Laboratorium	90
Kompozycja Proceduralna 2	Ćwiczenia	180
Warsztaty Komputerowe 2	Ćwiczenia	231
Podstawy informatyki dla projektantów	Ćwiczenia	263
Podstawy informatyki	Ćwiczenia	105
Zaawansowane Techniki Multimedialne	Wykład + ćwiczenia	30+60
	SUMA:	2445

Habilitant stale podnosi swoje kwalifikacje dydaktyczne poprzez samodoskonalenie, udział w szkoleniach, a także zdobywanie certyfikatów zawodowych. W roku 2011 odbył on **kurs pedagogiczny** w Centrum Kształcenia AGAT. Habilitant wziął udział w 5 szkoleniach:

1. EITCA Computer Graphics
2. Mac OS X Server Essentials v10.6
3. ASP.NET 4.0, Combidata Warszawa
4. Linux LPI, Akademickie Centrum Szkoleniowe PJWSTK
5. Microsoft 70-290, Akademickie Centrum Szkoleniowe PJWSTK

Na drodze samodoskonalenia habilitant przystąpił do 26 branżowych egzaminów organizowanych przez wiodące firmy z branży IT takie jak Adobe, Apple, Microsoft czy W3Schools, które zaliczył uzyskując następujące certyfikacje:

1. Adobe Certified Expert: Photoshop
2. Apple Certified Associate – Mac Integration
3. Apple Certified Specialist – Mac OS X Directory Services
4. Apple Certified Support Professional
5. Apple Certified Technical Coordinator
6. Apple Certified Specialist – Mac OS X Deployment
7. Apple Certified Specialist – Mac OS X Security and Mobility
8. Apple Certified System Administrator
9. MCSA: Windows Server 2012
10. MCSA: Windows 7
11. MCSA: Windows Server 2008
12. MCSA: Windows Server 2003
13. MCPD: Web Developer
14. MCITP: Enterprise Desktop Administrator on Windows 7
15. MCITP: Consumer Support Technician on Windows Vista

16. MCITP: Server Administrator on Windows Server 2008
17. MCTS: Windows Server Virtualization, Configuration
18. MCTS: Windows 7, Configuration
19. MCTS: .Net Framework 2.0, Web Applications
20. MCTS: SQL Server 2008, Database Development
21. MCTS: SQL Server 2008, Implementation and Maintenance
22. MCTS: .Net Framework 2.0, Windows Applications
23. MCTS: Windows Server 2008 Active Directory, Configuration
24. MCTS: Windows Server 2008 Network Infrastructure, Configuration
25. MCTS: Microsoft Windows Vista, Configuration
26. HTML 5 (z wyróżnieniem)

Habilitant stale aktualizuje skrypty laboratoryjne, a także swoją wiedzę zgodnie z trendami na dynamicznie zmieniającym się rynku IT. Pisał oraz aktualizował sylabusy do prowadzonych przez siebie przedmiotów na Wydziale Informatyki oraz Sztuki Nowych Mediów. Udziela się także jako freelancer na rynku szkoleniowym, a kilka jego szkoleń jest oferowanych w Akademickim Centrum Szkoleniowym Polsko-Japońskiej Akademii Technik Komputerowych.

4.2. Udział w konferencjach

W latach 2016-2018 (po obronie doktoratu) habilitant wygłosił referaty podczas następujących konferencji międzynarodowych:

1. The 10th edition of International Conference on Multimedia & Network Information Systems, 14-16 wrzesień 2016, Wrocław, Polska –
2. 2017 World Conference on Information Systems and Technologies, 11-13 kwietnia 2017, Porto Santo, Madera, Portugalia
3. Federated Conference on Computer Science and Information Systems – LTA, 3-6 wrzesień 2017, Praga, Czechy
4. HCist 2017 – International Conference on Health and Social Care, 8-10 październik 2017, Barcelona, Hiszpania
5. 2018 World Conference on Information Systems and Technologies, 27-29 marca 2018, Neapol, Włochy
6. The 11th edition of International Conference on Multimedia & Network Information Systems, 12-14 września 2018, Wrocław, Polska
7. Telemedycyna i eZdrowie 2018, 14-15 grudnia 2018, Warszawa, Polska

W latach 2013-2016 (przed obroną doktoratu) habilitant wygłosił referaty podczas następujących konferencji międzynarodowych:

1. XV International Phd Workshop OWD 2013, 19-22 Października 2013, Wisła, Polska
2. 10th International Workshop on Spoken Language Translation, 5-6 Grudnia 2013, Heidelberg, Niemcy
3. 2014 World Conference on Information Systems and Technologies, 15-18 Kwietnia 2014, Madera, Portugalia
4. Konferencja Telemedycyna i eZdrowie 2014, Warszawa, Polska
5. 3rd International Conference on Advancements in Information Technology, Sierpień 22-23, 2014, Dubaj, Zjednoczone Emiraty Arabskie
6. 9th International Conference on Multimedia & Network Information Systems, 17-19 Wrzesień, 2014, Wrocław
7. Polish-language Parallel Corpora seminar, Institute of Applied Linguistics, University of Warsaw, 30 marzec 2015, Warszawa, Polska
8. 18th International Conference Text, Speech and Dialogue, 14-17 Września 2015, Pilzno, Republika Czeska
9. Telemedycyna i eZdrowie 2015, Warszawa, Polska
10. HCist 2015 – International Conference on Health and Social Care, 7-9 Październik 2015, Vilamoura, Algarve, Portugalia

4.3.Działalność recenzencka i redakcyjna

Habilitant aktywnie udziela się jako recenzent do czasopism, jak i konferencji międzynarodowych. W okresie od obrony doktoratu był recenzentem w 13 czasopismach, 10 konferencjach (z czego 8 było indeksowanych w Web of Science), gdzie łącznie zrecenzował 46 prac. Czasopisma, do których recenzował:

1. Cyberpsychology, Behavior, and Social Networking (IF: 2.689, Lista A: 35pkt)
2. Journal of King Saud University – Computer and Information Sciences
3. The Knowledge Engineering Review Journal (IF: 1.07, Lista A: 25pkt)
4. Studies in Engineering and Technology Journal (IF: 1.46)
5. The Open Cybernetics and Systemics Journal
6. Computer and Information Science
7. Studies in Engineering and Technology

8. Dove Medical Press journals
9. IEEE Journal of Biomedical and Health Informatics (IF: 3.850, Lista A: 30pkt)
10. Expert Systems with Applications (IF: 3.768, Lista A: 35pkt)
11. Artificial Intelligence Review Journal (IF: 3.814, Lista A: 35pkt)
12. Journal of Natural Language Engineering Issue: Machine Translation Using Comparable Corpora (IF: 0.8, Lista A: 20pkt)
13. IEEE Transactions on Services Computing (IF: 4.418, Lista A: 45pkt)

Konferencje:

1. InterSpeech edycja 2017 oraz 2018
2. Multimedia, Interaction, Design and Innovation (MIDI) edycja 2017 oraz 2018 – indeksowane w Web of Science
3. Health and Social Care Information Systems and Technologies (HCIST) edycja 2017 oraz 2018 – indeksowane w Web of Science
4. World Conference on Information Systems and Technologies (WorldCIST) edycja 2017, 2018 oraz 2019 – indeksowane w Web of Science
5. Fuzzy Systems and Data Mining 2018 (FSDM 2018) – indeksowane w Web of Science

Dodatkowo od roku 2018 habilitant należy do grona ekspertów Narodowej Agencji Wymiany Akademickiej (NAWA).

4.4. Opracowane rozwiązania, programy i patenty

Habilitant jest twórcą m.in. wyżej opisanych rozwiązań zostały one zaimplementowane głównie w języku Python oraz udostępnione społeczności naukowców w ramach repozytorium GitHub (<https://github.com/krzwolk>). W tymże repozytorium habilitant rozwija 21 projektów. Inne zasoby takie jak wytrenowane modele również udostępnia pod adresami wyszczególnionymi w konkretnych publikacjach. Wśród autorskich aplikacji habilitanta wymienić należy:

- Aplikację służącą do analizy równoległych korpusów tekstowych pod kątem jakości danych, które można przefiltrować, jak również sprawdzająca i poprawiająca alignment na poziomie zdań.

- Aplikację służącą do budowy korpusów równoległych na podstawie danych porównywalnym i quasi-porównywalnych.
- Aplikację służącą do denormalizacji wyników automatycznego rozpoznawania mowy (ASR) oraz segmentacji tekstu na zdania celem poprawy jakości maszynowego tłumaczenia mowy oraz systemu S2S.
- Program służący do adaptacji danych równoległych lub monolingwalnych do konkretnej domeny tekstowej, używający trzech heurystyk komparacji (TF-IDF, Perplexity, oraz odległość Levenshteina).
- Program pozwalający na sprawdzenie czy w korpusie równoległym nie występują duplikaty, a w przypadku ich wykrycia – na usunięcie.
- Reimplementację narzędzia Yalign, zawierającą optymalizacje pod kątem jego wielowątkowości oraz usprawnienia związane z poprawą jakości wydobywanych danych z korpusów porównywalnych za pomocą algorytmu tuningującego. Dodatkowo zmieniono algorytm porównujący na Needleman-Wuncha wraz z jego optymalizacją na GPU.
- Aplikacja służąca do regulowego oczyszczania i poprawy jakości korpusów równoległych. Zawiera wbudowane reguły dla języka polskiego i angielskiego. Niemniej jednak adaptacja narzędzia na inne języki nie powinna stanowić problemu. Użycie narzędzia poprawia jakość tłumaczenia maszynowego.
- Serie robotów internetowych pozwalających na automatyczne przetwarzanie tekstów sieci i zapisywanie ich w formie korpusów.
- Aplikacja badająca trendy w mediach społecznościowych na podstawie analizy sentymentu i interpolacji kilku rodzajów modeli.
- Aplikacja pozwalająca na trenowanie modeli języka w oparciu o sieci neuronowe, a konkretniej o architekturę LSTM oraz modele log-bilinearne.
- Aplikacja pozwalająca na generowanie zasobów językowych na podstawie analizy semantycznej, jak i samą analizę semantyczną tekstu.
- Narzędzia automatycznie wykrywające języki w obrębie dokumentu jako preprocessing do tłumaczenia maszynowego.
- Narzędzie będące implementacją miary HMEANT dla języka polskiego.
- Aplikacja dedykowana językowi polskiemu służąca do podziału korpusów tekstowych na jednostki mniejsze niż słowa, także z możliwością ich anotacji częściami mowy oraz grupami gramatycznymi.
- Habilitant na podstawie artykułów Wikipedii opracował korpusy równoległe dla języka polskiego oraz innych języków europejskich, które to zostały użyte

podczas kampanii ewaluacyjnej IWSLT oraz udostępnione w ramach projektu OPUS¹.

Niektóre wyżej wymienione narzędzia otrzymały interfejsy graficzne i zostały uruchomione jako usługi internetowe w ramach projektu Clarin-PL. Mowa tu o narzędziu do budowania korpusów równoległych, rozpoznawaniu języka w dokumentach oraz podziału korpusów na jednostki mniejsze niż słowa.

Prace nad metodami i technikami doskonalenia systemów i generowania rozwiązań pozwalających na osiągnięcie założonych celów pozwalają na tworzenie zróżnicowanych programów i aplikacji, których konstrukcyjne podstawy mogą zostać wykorzystane na innym polu. Narzędzia usprawniające jakość tłumaczenia maszynowego bazujące na regulowym oczyszczaniu korpusów równoległych przyczyniają się do poprawy ich jakości. Zaprojektowanie rozwiązania według reguł modyfikowalnych daje możliwość rozwoju systemu i jego funkcjonalności. Wypracowane narzędzia mogą się stać podstawą badań prowadzonych przez innych naukowców w szczególności w dziedzinach humanistyki, lingwistyki, socjologii, przetwarzania języka naturalnego i nie tylko.

***Habilitant ma otwarte zgłoszenie patentowe dotyczące rozwiązania opartego o sztuczną inteligencję, rekomendującego terapię w leczeniu nadciśnienia tętniczego.**

4.5. Współpraca z innymi naukowcami

Podczas wyjazdów konferencyjnych habilitant nawiązał szereg cennych kontaktów naukowych w kraju i za granicą. Dzięki tym kontaktom został członkiem komitetu organizacyjnego konferencji (o czym w punkcie 4.6), a także nawiązał interdyscyplinarne badania naukowe między różnymi uczelniami oraz organizacjami. W zakresie przetwarzania języka naturalnego, a w szczególności wytworzenia zasobów i narzędzi wspierających lingwistów habilitant współpracował z firmą GetIT oraz NextSell. Prowadził badania we współpracy z Zakładem Genetyki Medycznej Warszawskiego Uniwersytetu Medycznego w zakresie diagnostyki chorób genetycznych za pomocą narzędzi sztucznej inteligencji. Nawiązał także współpracę z Zakładem Informatyki Medycznej i Telemedycyny Warszawskiego Uniwersytetu Medycznego, a

¹ <http://opus.nlpl.eu/Wikipedia.php>

także Polskim Towarzystwem Telemedycyny w zakresie inkorporacji technik uczenia maszynowego w interdyscyplinarnych badaniach i projektach naukowych.

Współpracuje także z Katedrą Nadciśnienia Tętniczego i Diabetologii Gdańskiego Uniwersytetu Medycznego nad systemem opartym o algorytmy sztucznej inteligencji, służącym do rekomendacji nowoczesnych postępowań medycznych w leczeniu nadciśnienia tętniczego. Rozwiązanie to zostało zgłoszone do procedury patentowej pod koniec roku 2018.

4.6. Działania popularyzatorskie

Habilitant jest aktywnym członkiem mazowieckiego oddziału Polskiego Towarzystwa Informatycznego oraz Polskiego Towarzystwa Telemedycyny i eZdrowia, gdzie w roku 2018 za swoje zasługi dla Towarzystwa dostał odznaczenie Bene Meritus.

Habilitant aktywnie uczestniczy w seminariach naukowych katedry multimediiów Polsko-Japońskiej Akademii Technik Komputerowych, które to także organizował w roku 2017. Z kolei w maju 2018 wystąpił na Seminarium „Przetwarzanie języka naturalnego”, które organizowane jest przez Zespół Inżynierii Lingwistycznej w Instytucie Podstaw Informatyki Polskiej Akademii Nauk w Warszawie. W maju 2017 wystąpił on na Warsztatach CLARIN-PL organizowanych przez Instytut Podstaw Informatyki Polskiej Akademii Nauk, gdzie w ramach swojego referatu przedstawił narzędzie umożliwiające pozyskanie równoległych segmentów tekstów z artykułów porównywalnych za pomocą skryptów w języku Python oraz udostępnienie pozyskanych tą metodą korpusów równoległych PL-*

Habilitant aktywnie uczestniczył także w organizowaniu międzynarodowych konferencji naukowych, w roku 2017, 2018 oraz 2019 był członkiem komitetu programowego konferencji Multimedia, Interaction, Design and Innovation (MIDI)². W roku 2017 poprowadził on sesję 6D na międzynarodowej konferencji Health and Social Care Information Systems and Technologies, w tym samym roku był on także przewodniczącym sesji Human-Computer Interaction na międzynarodowej konferencji World Conference on Information Systems and Technologies. W roku 2018 habilitant brał aktywny udział w przygotowaniach do międzynarodowej konferencji Telemedicine and eHealth 2018, konkretniej w przygotowaniu specjalnej sesji Polsko-Japońskiej Akademii Technik Komputerowych. Konferencja PolEval jest kampanią ewaluacyjną

² <https://fedesis.org/2018/midi/committee>

narzędzi służących przetwarzaniu języka naturalnego dla polskiego. Habilitant przygotował zestaw zadań dla uczestników tej konferencji w edycji 2019. Zajmie się także przeprowadzeniem tego konkursu i ewaluacją wyników.

Habilitant jest autorem 12 publikacji książkowych o tematyce związanej z nowoczesną informatyką oraz multimediami. Książki te zostały wyszczególnione chronologicznie poniżej, z czego cztery ostatnie są z okresu przed doktoratem, natomiast pozostałe z okresu po doktoracie:

1. Microsoft Office 2019 oraz 365 od podstaw, Styczeń 2016
2. Windows Serwer 2016, Styczeń 2019
3. Zabawa ze sztuczną inteligencją. Grudzień 2018. ISBN: 978-83-8119-372-6
4. Aplikacje mobilne oraz interaktywne komponenty na strony WWW. Grudzień 2018. ISBN: 978-83-8119-438-9
5. Nowoczesne strony WWW. HTML5, CSS3, Adobe Muse, Wordpress. Lipiec 2018. ISBN: 978-83-8119-294-1.
6. Rzeczywistość wirtualna (VR) dla każdego. Aframe i HTML 5. Lipiec 2018. ISBN: 978-83-8119-295-8.
7. Applying Comparable Corpora to Machine Translation. Lipiec 2015. ISBN: 978-3-659-76286-4
8. Designing and planning children educational applications. Designing and planning children oriented accessible applications for unsupervised and subconscious educational purposes. Wrzesień 2014. ISBN: 978-3-659-44859-1
9. Building an Internet Radio System with Interdisciplinary factored system for automatic content recommendation. Czerwiec 2013. ISBN: 978-3-659-41584-5
10. Mac OS X Server 10.8. Lipiec 2013. ISBN:978-83-7900-090-6
11. Biblia Windows Server 2012. Podręcznik administratora. Wrzesień 2012. Wydawnictwo Psychoskok. ISBN:978-83-63548-07-0
12. Praktyczny przewodnik po MAC OS X Server 10.6 oraz 10.7. Lipiec 2012. Wydawnictwo Psychoskok. ISBN:978-83-933996-8-0

Dodatkową aktywnością habilitanta jest redakcja portalu internetowego PCLab.pl jak również wolontariat w roli redaktora portali IN4.pl oraz e-biotechnologia.pl. Na tych portalach łącznie opublikował on 68 artykułów oraz poradników związanych ze sztuczną inteligencją, informatyką, bioinformatyką oraz nowoczesnymi technologiami.

Habilitant promował naukę oraz uczelnię, w której pracuje poprzez aktywny udział w dniach otwartych oraz organizację promocyjnego turnieju e-sportowego Warsaw Summoners Open zorganizowanego w Polsko-Japońskiej Akademii Technik Komputerowych oraz opatrzonego patronatem medialnym firmy ZOTAC, ElitePC oraz mediów jak in4.pl, pvp24.pl oraz komputerswiat.pl. Promuje on także uczelnię i naukę poprzez udostępnianie na zasadzie opensource wypracowanych zasobów językowych, modeli uczenia maszynowego oraz narzędzi w formie nie tylko otwartego kodu, ale także przystępnych narzędzi online, których to utrzymaniem i administracją się zajmuje.

4.7. Nagrody, granty i wyróżnienia

Habilitant w roku 2018 za swoje zasługi dla Polskiego Towarzystwa Telemedycyny i e-zdrowia otrzymał odznaczenie Bene Meritus.

Big Data Language Model of Contemporary Polish - wyróżnienie publikacji podczas konferencji Federated Conference on Computer Science and Information Systems – LTA 2017.

