

Recenzja rozprawy doktorskiej magistra Dawida Jurkiewicza pt.:

Novel Methods and Datasets for Intelligent Document Processing

Rozprawa doktorska została przygotowana na Wydziale Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu, pod kierownictwem dr hab Filipa Gralińskiego, Profesora UAM. Rozprawa Doktorska Pana Dawida Jurkiewicza należy do dziedziny nauk ścisłych i przyrodniczych, do dyscypliny naukowej *Informatyka*. Rozprawa jest przygotowana w języku angielskim. Jej część merytoryczna, zawierająca wkład Doktoranta w dyscyplinę Informatyka, składa się z 6 publikacji wieloautorskich.

Rozdział pierwszy przedstawia ogólną motywację do prowadzonych badań, omawia wyznaczone cele badawcze, oraz podsumowuje – w postaci Tabeli – listę publikacji wchodzących w skład Rozprawy, ich punktacje MEiN, liczbę cytowań i merytoryczny wkład Doktoranta w każdą z nich.

Rozdziały 2-7 to sześć publikacji, których Doktorant był współautorem.

W skład Rozprawy wchodzi również dwa załączniki. W Załączniku A znajduje się lista konkursów i projektów, z którymi powiązane jest sześć publikacji stanowiących część merytoryczną Rozprawy. Natomiast Załącznik B zawiera, osobiście podpisane, deklaracje wkładu merytorycznego każdego ze współautorów do każdej z publikacji, które wchodzi w skład Rozprawy.

Publikacje, z których składa się Rozprawa, dotyczą dwu zagadnień z obszaru tzw. inteligentnego przetwarzania dokumentów (ang. *intelligent document processing*): (1) rozpoznawania zakresu (ang. *span identification*) – trzy prace (Rozdziały 2-4), oraz (2) rozumienia zawartości dokumentów (ang. *document understanding*) – trzy prace (Rozdziały 5-7). Omówię teraz w skrócie ich zawartość.

1. *Contract Discovery: Dataset and a Few-Shot Semantic Retrieval Challenge with Competitive Baseline* (3 autorki i 7 autorów; Rozdział 2). Artykuł dotyczy rozpoznawania, które fragmenty tekstu dokumentu dotyczą „zadanego” tematu/zagadnienia. W przypadku omawianego tekstu, uczenie na podstawie niewielkiej liczby przykładów (ang. *few shot-learning*) oznacza, że tworzony system otrzymuje ograniczoną liczbę przykładowych wyrażenia prawniczych (ang. *legal clauses*) i ma nauczyć się odnajdywać podobne wyrażenia w tekście. W wyniku przeprowadzonych badań stwierdzono, że popularne rozwiązania, bazujące na zastosowaniu encoderów nie radzą sobie zbyt dobrze z rozważanym problemem. Równocześnie, zaproponowane w tekście, rozwiązania bazujące na zastosowaniu modelu języka działają znacznie lepiej. W uzupełnieniu do tekstu artykułu dostępne są również: zbiór danych oraz „prawnicze modele językowe”.

2. *Dynamic Boundary Time Wrapping for Sub-sequence Matching with Few Examples* (4 autorów; Rozdział 3). W tekście tym przedstawiono metodę wyszukiwania fragmentu tekstu w długim ciągu czasowym. Wyszukiwany tekst powinien być podobny do zbioru krótszych ciągów (które są danymi wejściowymi). W zaproponowanym podejściu przykłady zapytań, w niezmienionej postaci, wykorzystywane są równocześnie. Opracowana metoda wykorzystuje technikę dynamicznego dopasowania czasu i jest dostosowana do wieloetapowych zadań wyszukiwania na podstawie przykładu. Wyniki zostały ocenione na podstawie dwu (różnych) problemów z dziedziny przetwarzania języka naturalnego. Pokazują one, że dla niewielkiej liczby przykładów, zaproponowana metoda

przewyższa wartości bazowe i wyniki otrzymane w poprzednio stosowanych podejściach, albo przynajmniej osiąga porównywalne wyniki.

3. *ApplicaAI and SemEval-2020 Task 11: On RoBERTa-CRF, Span CLS and Whether Self-Training Helps Them* (1 autorka i 3 autorów; Rozdział 4). W artykule przedstawiono system, który zwyciężył w zadaniu: Klasyfikacja Techniki Propagandowej (TC) oraz system, który zajął drugie miejsce w zadaniu: Identyfikacja Rozpiętości Propagandowej (SI), zorganizowanymi w ramach konkursu SemEval-2020. Celem zadania TC była identyfikacja, na podstawie fragmentu tekstu propagandowego, zastosowanej w nim specyficznej techniki propagandowej. Celem zadania SI było odnalezienie, w dostarczonym do systemu tekście, jego fragmentów, które zawierają przynajmniej jedną technikę propagandową. Obydwa rozwiązania, opracowane przez zespół, którego członkiem był Doktorant, wykorzystywały technikę uczenia częściowo-nadzorowanego (ang. *semi-supervised learning*) polegającego na samokształceniu (ang. *self-training*). Warto zwrócić uwagę na fakt, że wprawdzie Conditional Random Fields (CRF) są rzadko używane w modelach językowych opartych na transformerach, to do zadania SI zastosowano właśnie architekturę RoBERTa-CRF. Natomiast do zadania TC zaproponowano zestaw modeli opartych na architekturze RoBERTa, z których jeden wykorzystuje warstwę Span CLS. Oprócz opisu stworzonych systemów zbadano wpływ poszczególnych decyzji architektonicznych, i schematów szkoleniowych, na otrzymane wyniki eksperymentalne. Ponadto omówiony został wpływ modeli szkoleniowych na wyniki, które zostaną otrzymane w przypadku niższego „budżetu obliczeniowego”. Jest to bardzo pożądanym krok w kierunku tworzenia systemów należących do klasy tzw. „oszczędnej sztucznej inteligencji” (ang. *frugal AI*). W omawianej pracy została również zaprezentowana całościowa analiza błędów.

Do grupy tekstów dotyczących rozumienia dokumentów należą:

4. *Going Full-TILT Boogie on Document Understanding with Text-Image-Layout Transformer* (1 autorka i 5 autorów; Rozdział 5). Tekst dotyczy rozumienia języka naturalnego wykraczającego poza typowe dokumenty „czysto tekstowe”. W tym kontekście zaproponowana została sieć neuronowa TILT, która równocześnie uczy się informacji o układzie, cechach wizualnych i semantyce tekstu. W proponowanym podejściu wykorzystywany jest dekodery, który jest zdolny do ujednoczenia różnorodnych problemów związanych z językiem naturalnym. W szczególności (wizualny) układ tekstu reprezentowany jest jako „bias” i uzupełniony kontekstowymi informacjami wizualnymi, podczas gdy centralną rolę odgrywa wstępnie wytrenowany transformer encoder-decoder. Zaproponowane podejście pozwala na osiągnięcie najlepszych (jak do tej pory) wyników w wydobywaniu informacji z dokumentów (nie składających się tylko z tekstu), i w odpowiadaniu na pytania wymagające zrozumienia układu tekstu na „stronie” (DocVQA, CORD, SROIE).

5. *DUE: End-to-End Document Understanding Benchmark* (1 autorka i 6 autorów; Rozdział 6). Celem pracy jest przedstawienie (i uzasadnienie) nowatorskiego zestawu dokumentów o bogatym układzie (benchmark nazwany DUE), pozwalający na porównywanie skuteczności podejść do zagadnienia rozumienia tekstów w przypadku gdy na wejściu pojawia się nie tylko strumień znaków, ale również informacja o układzie tekstu „na stronie”. W skład benchmarku wchodzi zarówno powszechnie dostępne, jak i celowo przeformułowane zbiory danych, oddające kompleksowość i różnorodność potencjalnych rzeczywistych scenariuszy. Benchmark obejmuje możliwości testowania efektywności rozwiązań dotyczących, między innymi, wizualnej odpowiedzi na pytania, wyodrębniania kluczowych informacji, czy też maszynowego czytania ze zrozumieniem. Przedstawione zadania dotyczą różnych dziedzin, układów dokumentów – w tym dokumentów obejmujących tabele, wykresy, listy, czy też infografiki. Ponadto w tekście przedstawiono również wartości bazowe (na moment publikacji benchmarku) dla efektywności rozwiązywania wybranych zagadnień. Tekst zawiera również analizę

najnowszych osiągnięć w zakresie modelowania języka wykorzystującego informacje o układzie tekstu (ang. *layout-aware language modelling*) i podsumowanie najważniejszych wyzwań w rozważanym obszarze badań.

6. *ICDAR 2023 Competition on Document Understanding of Everything (DUDE)* (11 autorów; Rozdział 7). W artykule przedstawiono wyniki konkursu ICDAR 2023, pt. „Dokumentowe zrozumienie wszystkiego”. W ramach konkursu wprowadzony został nowy zbiór danych – DUDE, który składa się z 5 tys. bogatych wizualnie, wielobranżowych, wielodomenowych, wielostronicowych dokumentów różnego pochodzenia i mających bardzo różne daty powstania. Ponadto zdefiniowano 40 tys. „pytań testowych” dotyczących zawartości dokumentów. Konkurs miał formę pojedynczego zadania, z wieloetapowym protokołem oceny, który oceniał możliwości zaproponowanych modeli w kilku etapach, poprzez testowanie uogólnień dotyczących wcześniej niewidzianych pytań i dziedzin. Dodatkowo, w ramach prac badawczych zespołu przygotowującego konkurs, skonstruowano nowy, niezależny zestaw testów diagnostycznych do szczegółowej analizy wydajności modeli uczestniczących w konkursie. Ponadto przeprowadzono całościową analizę wyników uzyskanych różnymi metodami. Na podstawie przeprowadzonej analizy stwierdzono, że zbiór danych DUDE ma szansę stać się istotnym, mogącym przetrwać wiele lat jako niezmienny punkt odniesienia, benchmarkiem dla testowania wyników badań na rozumieniem dokumentów.

Zgodnie z dokumentacją przedstawiona na stronie 20 Rozprawy, publikacje 1-4 oraz 6 maja 140 punktów parametrycznych MEiN, natomiast publikacja 5 ma 200 punktów. Oznacza to, że *wszystkie artykuły zostały opublikowane w bardzo dobrych czasopismach lub materiałach bardzo dobrych konferencji*. Oznacza to również, że można uznać, że ***zawartość merytoryczna każdego z nich została odpowiednio zweryfikowana i wnosi istotny wkład w dyscyplinę naukową Informatyka***.

W tym momencie należy więc odpowiedzieć na pytanie, czy zebrane w Rozprawie teksty są „dostatecznie bliskie tematycznie” aby można było je potraktować jako „całościowe osiągnięcie”. Pytanie to jest szczególnie zasadne, albowiem prace dotyczą dwu obszarów inteligentnego przetwarzania dokumentów. Jednakże, na podstawie przeprowadzonej analizy wskazanych sześciu tekstów stwierdzam, że ***spełniają one kryteria bycia całościowym osiągnięciem naukowym***. Co więcej, łącząc to stwierdzenie z tezą z poprzedniego akapitu można jednoznacznie stwierdzić co następuje.

Przedstawiona do oceny Rozprawa Doktorska spełnia warunki określone w Art. 187 Ustawy z dnia 20 lipca 2018 r./Prawo o szkolnictwie wyższym i nauce /(Dz.U. z 2022 r. poz. 574 z późn. zm.), oraz uzasadnia nadanie stopnia naukowego doktora w dyscyplinie informatyka.

Pozostaje więc ostatnie pytanie, które jest kluczowe i na które nie ma jednoznacznej odpowiedzi w przygotowanych przez Doktoranta materiałach. Chodzi o wkład naukowy poszczególnych autorek i autorów (wszystkie prace są wieloautorskie), w tym o wkład samego Doktoranta. Niezaprzeczalnie, w Tabeli 1.1, dla każdej pracy, wchodzącej w skład Rozprawy, opisany został szczegółowy wkład merytoryczny Doktoranta, ale otwartą pozostaje kwestia, jak ten wkład ma się do pracy pozostałych autorek i autorów, i do całościowego nakładu pracy wymaganego do przygotowania każdej z publikacji. Ponadto otwartą pozostaje kwestia, kto był/była pomysłodawcą/pomysłodawczynią zaproponowanych w każdej z prac rozwiązań całościowych, jak i pod-rozwiązań będących częścią poszczególnych etapów pracy naukowej. W tym kontekście, dodatkowa informacja dostępna jest w Załączniku B. Jednakże również tam nie ma szczegółowych informacji dotyczących tego jaki był *wkład procentowy* każdego z autorek i autorów. Co więcej, ponieważ listy autorek i autorów nie są alfabetyczne, można by dojść do wniosku, że osoby na pierwszym miejscu (pierwszych miejscach)

miały „ważniejszy wkład” w daną pracę. Tak więc, bardzo trudną jest odpowiedź na pytanie: czy można powiedzieć, że to właśnie wkład Doktoranta jest tym który jest znaczącym i powinien być „nagrodzony” dopuszczeniem do obrony doktoratu, na podstawie materiału zawartego w ocenianej Rozprawie.

W tym kontekście, w moim dorobku naukowym zdecydowana większość prac to prace wielo-autorskie. Nie tylko jestem zwolennikiem pracy zespołowej, ale uważam również, że jeśli chodzi o informatykę, to nie ma już dzisiaj zbyt wiele miejsca na „indywidualizm naukowy”. Ponadto, będąc liderem kilku zespołów projektowych zdaję sobie bardzo dobrze sprawę z tego jak trudno jest opisać wkład autorski gdy badania były prowadzone (i wynikające z nich publikacje były przygotowywane) przez skutecznie współpracujący ze sobą zespół. Patrząc na problem z perspektywy cotygodniowych wielogodzinnych spotkań zespołu projektowego jestem wręcz gotów stwierdzić, że rozpisywanie wkładu pracy pomiędzy poszczególne osoby nie ma większego sensu. W większości przypadków nie da się, na przykład, powiedzieć, że to dana osoba (jedno-osobowo) miała dany pomysł (który później był rozwinięty i zrealizowany przez zespół badawczy).

Tak więc, po dogłębnym zapoznaniu się z zawartością Załącznika B i skonfrontowaniem go z moim wieloletnim (i wielokrotnym) doświadczeniem bycia liderem, współ-liderem i członkiem zespołu projektowego stwierdzam co następuje.

Przedstawiona do oceny Rozprawa Doktorska spełnia warunki określone w Art. 187 Ustawy z dnia 20 lipca 2018 r./Prawo o szkolnictwie wyższym i nauce /(Dz.U. z 2022 r. poz. 574 z późn. zm.), oraz uzasadnia nadanie Panu Jurkiewiczowi stopnia naukowego doktora w dyscyplinie informatyka.

W związku z tym wnioskuję o dopuszczenie Pana Magistra Dawid Jurkiewicz do następnych kroków postępowania w sprawie nadania stopnia doktora, w tym do Obrony Pracy Doktorskiej.