

Prof. dr hab. inż. Andrzej Czyżewski  
Politechnika Gdańska, Wydział ETI  
Katedra Systemów Multimedialnych

19. 01. 2022 r.

Opinia o rozprawie doktorskiej mgr inż. **Tomasza Stanisławka**  
pt.: „Ekstrakcja informacji z dokumentów o bogatej strukturze graficznej”  
przygotowanej w Pol. Warszawskiej pod kier. dr hab. inż. Przemysława Biecka, prof. uczelni

Rozprawa składa się z części wprowadzającej o objętości ok. 25 stron tekstu wraz z pięcioma rysunkami, siedmioma tablicami oraz z wykazu literatury obejmującego 58 pozycji. Tę część uzupełnia wykaz dorobku naukowego, związanego z tematem rozprawy, który obejmuje cztery współautorskie publikacje konferencyjne. W przypadku dwóch z nich, mgr inż. Tomasz Stanisławek jest pierwszym autorem, w przypadku dwóch pozostałych, trzecim współautorem. Autor rozprawy wspomina na zakończenie powyżej wspomnianego podsumowania także o sześciu innych publikacjach. Nie są one bezpośrednio związane z tematem rozprawy, ale wchodzi w skład jego ogólnego dorobku publikacyjnego.

W odniesieniu do publikacji wchodzących w skład rozprawy, ich łączna punktacja według listy MEiN wynosi 620 (3 x 140 +200), mają one jednak we wszystkich przypadkach licznych współautorów (w liczbie od 5 do 8), stąd też niezbędne okazało się załączenie oświadczeń współautorów, w których zgodnie określają oni rolę mgr inż. Tomasza Stanisławka jako autora wiodącego dwóch publikacji, których jest on pierwszym autorem (rola pomysłodawcy, autora metody, implementacji, głównego eksperymentatora). W trzeciej wymienionej publikacji jego rola określana jest w oświadczeniach jako m. in. uczestnika dyskusji, osoby pracującej nad zbiorami danych i benchmarkami. Jako współautor czwartej, ostatniej, z wchodzących w skład rozprawy publikacji, mgr inż. T. Stanisławek ponownie wymieniany jest jako inicjator głównej idei, współautor koncepcji i metodyki, eksperymentator. Tym niemniej, dorobek wynikający z opublikowania wspomnianych czterech referatów, według załączonych oświadczeń, w znacznym stopniu dzieli się pomiędzy licznych współautorów. Pozostałą część rozprawy stanowią kopie wspomnianych czterech publikacji konferencyjnych z lat 2019-2021.

Dalsza część mojej opinii zredagowana jest w formie odpowiedzi na pytania stosowane zwyczajowo w toku oceny rozpraw naukowych.

1. Jaki problem naukowy (teza) został rozwiązany i przedstawiony w rozprawie?

Praca dotyczy dziedziny przetwarzania języka naturalnego. Jest to jeden z centralnych problemów współczesnej informatyki technicznej, którego rozwiązywanie w ogólności powinno umożliwić postęp w automatycznym przetwarzaniu dokumentów. Autor podkreśla, że typowe dokumenty składają się nie tylko z warstwy tekstowej, ale także z włączonych w nią treści innej natury, np. obrazów. W związku z powyższym, zagadnienie to komplikuje się oraz wypada zauważyć, że przynosi w pewien sposób

ograniczone korzyści praktyczne, zwłaszcza gdy podchodzi się do niego bez użycia wiedzy związanej z automatyzacją semantycznego opisu obrazów. W tym kontekście można zauważyć dążenie przez Doktoranta do wyabstrahowania stosunkowo wąskiego podejścia do ekstrakcji informacji z dokumentów o bogatej formie, gdyż w pracy nie wspomina się na przykład o dynamicznie rozwijających się technikach automatycznej semantycznej adnotacji obrazów .

W rozprawie nie sformułowano w sposób wyraźny tez, które byłyby formalnie udowodniane. Wprawdzie obecnie coraz częściej spotyka się rozprawy odbiegające od klasycznego stylu hipoteza-teza-dowód, ale taki sposób podejścia do konstruowania rozprawy doktorskiej ogranicza możliwości wnikliwej oceny składających się na nią osiągnięć. Z kolei zaletą wąskiego spojrzenia na zagadnienie naukowe jest spójny i dość przejrzysty charakter narracji, co ma zastosowanie do opracowanego przez Doktoranta przewodnika po publikacjach. Aktualna wartość inżynierska udokumentowanych prac jest znaczna, aczkolwiek z uwagi na szybką ewolucję algorytmów i zbiorów tekstów, wkład naukowy, mający głównie charakter tymczasowych przyczynków, może nie okazać się trwały. Nie zmienia to oceny, że przedstawione koncepcje, potwierdzone eksperymentalnie odzwierciedlają podejście do zagadnienia nie tylko inżynierskie, ale również naukowe.

2. W jaki sposób doktorant rozwiązał problem, jakich użył metod i jakich to wymagało umiejętności?

Z oświadczeń licznych współautorów referatów przedstawionych na czterech konferencjach wynika, że Doktorant inicjował, inspirował te referaty, bądź konsultował ich treści w toku dyskusji, a ponadto wykazał się m. in. kompetencjami z zakresu gromadzenia i obróbki danych testowych, implementacji oprogramowania do przetwarzania tego typu danych. Umiejętności tych nie sposób precyzyjnie ocenić, ani też zakwestionować na podstawie dokumentacji ograniczonej do stosunkowo krótkich treści wieloauorskich referatów, których objętości zawierają się w przedziale 10-15 stron wydruku przy współautorstwie 5-8 osób.

3. Na czym polega oryginalny dorobek autora i jakie jest jego znaczenie poznawcze lub przydatność praktyczna dla nauki bądź techniki?

Referat konferencyjny, zatytułowany "Named Entity Recognition--Is there a glass ceiling?" prezentuje szczegółową analizę typów błędów występujących w nowoczesnych metodach uczenia maszynowego. Pokazano w nim słabe i mocne strony modeli: Stanford, CMU, FLAIR, ELMO i BERT, a także ich wspólne ograniczenia. Zaproponowano nowe techniki poprawy adnotacji, procesów treningowych oraz sprawdzania jakości i stabilności modelu. Prezentowane wyniki oparte są na zbiorze danych CoNLL 2003 dla języka angielskiego. Rezultatem tego jest nowa, wzbogacona semantyczna adnotacja błędów dla tego zbioru danych oraz nowe zestawy danych diagnostycznych.

Kolejny referat konferencyjny zatytułowany “Kleister: Key Information Extraction Datasets Involving Long Documents with Complex Layouts” podkreśla znaczenie zadania ekstrakcji kluczowych informacji w procesie przetwarzania języka naturalnego i prezentuje dwa nowe zbiory danych nazwane: Kleister NDA i Kleister Charity. Obejmują one kolekcję obszernych formalnych dokumentów w języku angielskim. W tych zbiorach danych system przetwarzania języka naturalnego ma za zadanie znaleźć lub wnioskować o różnych typach podmiotów, wykorzystując zarówno tekstowe, jak i strukturalne cechy dokumentów. Autor wraz ze współpracownikami przetestował kilka systemów bazowych z dziedziny ekstrakcji informacji kluczowych (Flair, BERT, RoBERTa, LayoutLM, LAMBERT), co pokazało, że opracowane zbiory danych są stosunkowo trudne do analizy przy użyciu znanych modeli. Opracowane zbiory dokumentów zostały udostępnione środowisku naukowemu.

Trzeci referat, zatytułowany “DUE: End-to-End Document Understanding Benchmark” dotyczy rozumienia dokumentów o na tyle bogatych układach, że pozostaje ono trudnym zadaniem w opinii środowisk badawczych, które zajmują się przetwarzaniem języka naturalnego. Kwantyfikację postępu w tej dziedzinie utrudnia brak powszechnie akceptowanego benchmarku, stąd motywacja do wprowadzenia opisanego w pracy benchmarku pod nazwą Document Understanding Evaluation (DUE), składającego się z dostępnych i przeformułowanych zbiorów danych, pozwalających porównywać możliwości systemów w warunkach działania na rzeczywistych dokumentach. Zproponowany benchmark obejmuje zadania ekstrakcji kluczowych informacji oraz maszynowego czytania ze zrozumieniem dokumentów tekstowych, wzbogaconych o liczne elementy dodatkowe, inne niż tekst. Referat w sposób systematyczny prezentuje i porównuje aktualnie dostępne bazy danych i najnowsze osiągnięcia w modelowaniu języka z uwzględnieniem układu graficznego. Benchmarki oraz implementacje referencyjne zostały udostępnione do wykorzystania przez środowisko naukowe.

Ostatni za zaprezentowanych referatów nosi tytuł “LAMBERT: Layout-Aware Language Modeling for Information Extraction”. Przedstawiono w nim nowe, proste podejście do problemu rozumienia dokumentów, w których nietrywialny układ wpływa na lokalną semantykę. Model został oceniony w zadaniu ekstrakcji informacji z wykorzystaniem publicznie dostępnych zbiorów danych. Wykazano, że opracowany model osiąga wyższą wydajność na zbiorach danych składających się z dokumentów bogatych wizualnie, a także przewyższa bazowy model RoBERTa na dokumentach o płaskim układzie. Rozwiązanie to zajęło pierwsze miejsce w rankingu publicznym dla zadania Ekstrakcja kluczowych informacji z zestawu danych SROIE, a ponadto zostało wyróżnione jako Best Industry Related Paper Award na konferencji ICDAR 2021 w Lozannie.

4. Jaka jest szansa dalszego wykorzystania wyników rozprawy?

Potrzeba rozwiązywania problemów wynikających z ekstrakcji dokumentów o bogatej strukturze graficznej (ang. Visually Rich Documents, VRDs) oraz opracowywania rozwiązań uwzględniających aspekty struktury dokumentu mają znaczenie dla rozwoju technologicznego podmiotów, w których obieg informacji odbywa się poprzez różnego rodzaju dokumenty elektroniczne. Bezpośrednio wykorzystane przez społeczność naukową, zajmującą się przetwarzaniem języka naturalnego mogą być opracowane i udostępnione publicznie zbiory tekstów. Na platformie hostingowej oprogramowania Github Doktorant i współpracownicy umieścili implementacje referencyjne. Szansę wykorzystania ma ponadto opracowany benchmark pod nazwą Document Understanding Evaluation (DUE).

5. Jakiej wiedzy, umiejętności oraz kompetencji i na jakim poziomie nabył doktorant w wyniku realizacji rozprawy?

Analiza przewodnika po publikacjach i treści załączonych publikacji pozwala zauważyć, że Autor posiadał umiejętności z zakresu badań literaturowych, tworzenia i usprawniania algorytmów, programowania komputerowego, budowy i adnotowania zbiorów danych i tekstów wykorzystywanych do analizy algorytmicznej.

6. Czy rozprawa obejmuje najnowsze osiągnięcia nauki i świadczy o znajomości współczesnej literatury z dyscypliny naukowej, której dotyczy?

Rozprawa jest ściśle powiązana z nową i z najnowszą literaturą dotyczącą zagadnień przetwarzania języka naturalnego. Autor wykazał się orientacją na temat nowoczesnych algorytmów i zbiorów testowych. W publikacjach udokumentowane są eksperymenty porównawcze, oparte na zastosowaniu najnowszych algorytmów ekstrakcji informacji tekstowych z dokumentów, w tym dokumentów o bogatej strukturze.


7. Czy doktorant w trakcie pracy nad rozprawą wykazał się kompetencjami społecznymi niezbędnymi do prowadzenia pracy naukowo-badawczej?

Zespołowy charakter publikacji w pewnym stopniu utrudnia śledzenie indywidualnego wkładu Doktoranta, jednak należy uznać, że współczesne zaawansowane projekty informatyczne w częstych przypadkach wymagają pracy zespołowej, do której prowadzenia konieczne jest posiadanie przez członków zespołu określonych kompetencji społecznych. Biorąc udział w zespołowej pracy eksperymentalnej i w opracowaniu wieloautorskich publikacji, Doktorant w trakcie pracy nad rozprawą wykazał się kompetencjami społecznymi niezbędnymi do prowadzenia pracy naukowo-badawczej.

## Wniosek

Rozprawa pana mgr inż. Tomasza Stanisławka została zrealizowana w sposób odzwierciedlający wymagane kwalifikacje jej Autora, wystarczający nakład pracy badawczej, implementacyjnej i eksperymentalnej, jak również jego potwierdzony oświadczeniami współautorów referatów udział w opublikowaniu wyników na wysokopunktowanych konferencjach.

W mojej opinii treść rozprawy mgr inż. Tomasza Stanisławka spełnia zatem wymogi Prawa o Szkolnictwie Wyższym i Nauce, z dnia 20 lipca 2018 r. (Dz. U. 30. 08. 2018 r. Poz. 1668), stawiane kandydatom do stopnia naukowego doktora.

A handwritten signature in black ink, appearing to be 'J. C. M.' or similar, written in a cursive style.