

**Recenzja rozprawy doktorskiej**  
*Artura Nowakowskiego*  
**zatytułowanej:**  
*Quality Optimization Methods*  
*in Neural Machine Translation Systems*

## **1. Problem badawczy i jego znaczenie**

Najważniejszym problemem omawianym w rozprawie są metody optymalizacji jakości w systemach neuronowego tłumaczenia maszynowego (ang. *Neural Machine Translation*, NMT). Neuronowe tłumaczenie maszynowe jest rodzajem tłumaczenia maszynowego wykorzystującego sieci neuronowe do tłumaczenia tekstu z jednego języka na inny. Zakres rozprawy jest zgodny z najnowszymi osiągnięciami w tej dziedzinie dotyczącymi poprawy jakości i wydajności tłumaczenia.

Jednym z dwóch znaczących ostatnich postępów jest wykorzystanie mechanizmu uwagi w architekturze Transformer (Vaswani i in., 2017). Innym jest wykorzystanie wstępnie wytrenowanych modeli, które są później dostrajane do konkretnego zadania.

Pomimo tych postępów, bardziej specyficzne scenariusze wymagają dodatkowych optymalizacji, takich jak włączenie ludzkiej wiedzy (np. ręcznie tworzonych leksykonów) do modeli NMT, zwłaszcza w odniesieniu do tłumaczeń specyficznych dla danej dziedziny, co należy do kontrybucji niniejszej rozprawy.

Problemy rozwiązywane w ramach rozprawy dotyczą nie tylko nowych metod optymalizacji jakości w NMT, ale celem rozprawy jest również zademonstrowanie ich zastosowania w rzeczywistych systemach tłumaczenia maszynowego. Praca jest wynikiem doktoratu przemysłowego realizowanego we współpracy firmy Poleng i Uniwersytetu Adama Mickiewicza. Wyniki zostały wdrożone w firmie Poleng w systemie nEYron (Implementacja i wdrożenie systemu MT dla dużej firmy audytorskiej i konsultingowej) oraz POLENG MT (Adaptacyjna platforma MT).

## **2. Wkład autora**

Rozprawa składa się z czterech artykułów o charakterze naukowym i trzech artykułów opisujących prace badawczo-rozwojowe.

Artykuły o charakterze naukowym przedstawiają nowatorskie pomysły na optymalizację jakości NMT w różnych scenariuszach: tłumaczenie między parami języków o niskich zasobach, włączenie terminologii specyficznej dla danej dziedziny, wykorzystanie modeli szacowania jakości tłumaczenia w celu poprawy NMT, tłumaczenie na poziomie dokumentu i tłumaczenie jednostek nazwanych.

Artykuły opisujące prace badawczo-rozwojowe opisują rzeczywiste zastosowania: związane z firmą Poleng i z systemem MT będącym wynikiem projektu badawczo-rozwojowego.

Poniżej podsumowuję wkład naukowy i/lub badawczo-rozwojowy poszczególnych artykułów:

### **1. Neural Machine Translation with Inflected Lexicon**

Artykuł opisuje integrację wiedzy ludzkich ekspertów, ręcznie stworzonego leksykonu. Eksperymenty dotyczyły ograniczeń leksykalnych dla tłumaczenia pomiędzy językami angielskim i polskim. W artykule przedstawiono metodę opartą na algorytmie dekodowania z ograniczeniami w celu obsługi fleksyjnych form haseł leksykalnych, przy czym dane trenujące i architektura modelu pozostają niezmiennicze. W artykule zaproponowano nowe metryki oceny związane ze zdolnością metody do obsługi ograniczeń.

### **2. Approaching English-Polish Machine Translation Quality Assessment with Neural-based Methods**

Artykuł został opublikowany w ramach materiałów konferencyjnych PolEval 2021, kampanii ewaluacyjnej narzędzi do przetwarzania języka naturalnego dla języka polskiego. Artykuł dotyczy oceny jakości tłumaczenia maszynowego. Obecne metryki stosowane do tego zadania, takie jak COMET (Rei i in., 2020), to duże modele językowe (LLM) dostrojone na podstawie ludzkich ocen do oceny tłumaczenia maszynowego. Nie zależą one od jednego dosłownego ludzkiego tłumaczenia referencyjnego jako złotego standardu, a zatem są bardziej niezawodne. Te oparte na sieciach neuronowych modele szacowania jakości mogą być wykorzystywane do wyboru najlepszej hipotezy ze zbioru wielu hipotez, w sposób symulujący ludzką ocenę.

Artykuł zawiera opis eksperymentów dotyczących wspólnego zadania oceny jakości tłumaczenia maszynowego PolEval 2021 w dwóch wariantach: oceny jakości opartej na referencjach i bez referencji.

Autor zastosował techniki takie jak stopniowe odmrażanie i dyskryminacyjne wskaźniki uczenia się w celu optymalizacji jakości.

### **3. Adam Mickiewicz University's English-Hausa Submissions to the WMT 2021 News Translation Task**

Artykuł dotyczy problemu tłumaczenia maszynowego dla odległych języków o niskich zasobach, w szczególności kierunków tłumaczenia angielski-Hausa i Hausa-angielski. W artykule przedstawiono wiele eksperymentów wykorzystujących techniki, takie jak uczenie transferowe z pary języków o wysokich zasobach, iteracyjne tłumaczenie wsteczne w celu rozszerzenia danych trenujących oraz filtrowanie danych w celu zmniejszenia ich zaszumienia.

### **4. Adam Mickiewicz University at WMT 2022: NER-Assisted and Quality-Aware Neural Machine Translation**

W artykule przedstawiono zgłoszenia naukowców z Uniwersytetu im. Adama Mickiewicza do ścieżki WMT 2022 General MT Task, pokazując ich udział i osiągnięcia w kierunkach tłumaczenia ukraiński ↔ czeski. Języki te są blisko spokrewnione, ale mają niewielkie zasoby. Autorzy wykorzystują ważony zespół czterech modeli opartych na architekturze Transformer, która jest najnowocześniejszym modelem neuronowego tłumaczenia maszynowego.

W artykule opisano wiele nowatorskich metod optymalizacji jakości dla NMT. W szczególności są to uczenie transferowe (z pary języków o wysokich zasobach), technika zaszumionego tłumaczenia wstecznego w celu rozszerzenia korpusów trenujących, wykorzystanie modelu na poziomie dokumentu do trenowania i modelu na poziomie zdania do dekodowania i generowania listy  $n$

najlepszych, połączenie tak uzyskanej listy  $n$  najlepszych z listą  $n$  najlepszych uzyskaną przez model wytrenowany na poziomie zdania (w celu tłumaczenia wielu zdań jednocześnie), tłumaczenie wspomagane NER (rozpoznawaniem jednostek nazwanych) i wreszcie zastosowanie modeli szacowania jakości i dekodowania minimalnego ryzyka Bayesa w celu ponownego uszeregowania listy  $n$  najlepszych i wybrania najlepszej hipotezy mierzonej metryką oceny COMET. Wyniki automatycznej oceny wskazują, że zaproponowane systemy zajmują pierwsze miejsce w konkursie w obu kierunkach tłumaczenia.

### **5. A Neural Translator Designed to Protect the Eastern Border of the European Union**

W artykule przedstawiono system opracowany dla Państwowej Straży Granicznej jako moduł projektu badawczo-rozwojowego "Zaawansowana analiza Internetu wspierająca wykrywanie grup przestępczych" (nazwa skrócona: AI Searcher; projekt nr DOB-BIO9/19/01/2018), finansowanego przez Narodowe Centrum Badań i Rozwoju. Nowością jest wykorzystanie ręcznie tworzonych leksykonów.

### **6. nEYron: Implementation and Deployment of an MT System for a Large Audit & Consulting Corporation**

Artykuł opisuje wdrożenie systemu MT w polskim oddziale EY Global Limited.

### **7. POLENG MT: An Adaptive MT Platform**

Artykuł przedstawia POLENG MT, platformę MT, która może być używana jako aplikacja internetowa w chmurze lub rozwiązanie lokalne.

Główny wkład w ramach rozprawy obejmuje:

- algorytm włączania fleksyjnych ograniczeń leksykalnych,
- analizy eksperymentalne dotyczące neuronowych modeli oceny jakości tłumaczenia maszynowego,
- konceptualizacja i implementacja różnych technik dotyczących wstępnego przetwarzania danych i etapów filtrowania danych,
- konceptualizacja i implementacja różnych technik optymalizacji wydajności, takich jak iteracyjne tłumaczenie wsteczne, PB-SMT, uczenie transferowe, dekodowanie uwzględniające jakość i łączenie modeli.

## **3. Poprawność**

W przedstawionych artykułach nie zidentyfikowano żadnych istotnych błędów.

Moje główne zastrzeżenia dotyczące poprawności, kompletności lub spójności są następujące:

1. W niektórych artykułach brakuje szczegółowego wyjaśnienia metodologii oceny i wskaźników stosowanych do oceny jakości tłumaczeń. Korzystne byłoby zamieszczenie większej ilości informacji na temat konfiguracji ewaluacji. Analiza modeli bazowych jest czasami krótka i brakuje kompleksowej analizy lub porównania ich wydajności. Dalsze szczegóły dotyczące mocnych i słabych stron tych modeli bazowych wzbogaciłyby wkład naukowy.

2. W niektórych artykułach brakuje przedstawienia motywacji lub wyjaśnienia wybranych wskaźników. Choć wspominają o różnych zastosowanych metrykach, takich jak BERTScore, Bleu i TransQuest, nie dostarczają wystarczających wyjaśnień ani szczegółów na ich temat. Czytelnicy mogą nie znać wszystkich tych wskaźników, więc korzystne byłoby przedstawienie krótkiego opisu każdego z nich i sposobu w jaki przyczyniają się one do oceny jakości tłumaczenia.
3. Czasami brakuje szczegółów opisu konfiguracji eksperymentalnej i opis danych, dotyczący na przykład rozmiaru zbioru trenującego, czasu trwania trenowania oraz sprzętu / oprogramowania używanego do szkolenia i oceny jest ograniczony. Szczegóły te mają kluczowe znaczenie dla odtwarzalności i zrozumienia warunków eksperymentalnych. Na przykład pewne kwestie dotyczą niepełnego opisu podziału zbioru danych: jeden z artykułów krótko wspomina o podziale zbioru danych na zbiór "dev-0" i dwa zbiory testowe ("test-A" i "test-B"), ale nie dostarcza informacji o rozmiarze lub konkretnym składzie tych zbiorów. Pomocne byłoby uwzględnienie pewnych konkretnych cech danych w opisie eksperymentów.
4. W artykułach związanych z udziałem w wyzwaniach brakuje porównania z innymi uczestnikami.
5. W niektórych artykułach brakuje jasnej dyskusji na temat ograniczeń lub potencjalnych braków proponowanego podejścia. Cenne byłoby odniesienie się do wszelkich słabości lub wyzwań napotkanych podczas eksperymentów.

#### **4. Wiedza kandydata**

Kandydat posiada doskonałą znajomość teorii i know-how w dziedzinie przetwarzania i rozumienia języka naturalnego, która wchodzi w zakresu dyscypliny informatyka. Kandydat korzysta z różnych technik opisanych w literaturze, wskazując na odniesienia literaturowe i wykazując się biegłością w korzystaniu z różnych narzędzi i zestawów narzędzi. Zademonstrował tę wiedzę i umiejętności w różnych zadaniach, takich jak modelowanie języka lub przetwarzanie wstępne. Jako przykład omówię rozdział 5.2 artykułu zatytułowanego "Adam Mickiewicz University at WMT 2022: NER-Assisted and Quality-Aware Neural Machine Translation". Rozdział zatytułowany "Data" dotyczy przetwarzania wstępnego. Kandydat wspomina o wielu zastosowanych technikach i narzędziach, takich jak OpusFilter (Aulamo i in., 2020) do czyszczenia i filtrowania zdań, filtrowanie wykrywania języka w oparciu o Fast-Text (Joulin i in., 2016), usuwanie duplikatów, stosowanie heurystyki opartej na długości zdania, stosowanie skryptów wstępnego przetwarzania Moses (Koehn i in., 2007) w celu normalizacji interpunkcji, usunięcie znaków niedrukowalnych, tokenizację tekstu na jednostki podwyrazowe przy użyciu SentencePiece (Kudo & Richardson, 2018) z algorytmem modelu językowego unigramu (Kudo, 2018). Ta głęboka wiedza pozwoliła kandydatowi dokonać zwycięskich zgłoszeń do wyzwań w zakresie WMT.

#### **5. Inne uwagi**

Obecnie, wraz z rozwojem dużych modeli języka, obszary pracy doktorskiej stają się kluczowymi obszarami naukowymi i technologicznymi. Doktorant zaprojektował i opracował różnorodne rozwiązania oraz wykazał się różnorodnymi umiejętnościami związanymi z projektowaniem, budową i wdrażaniem Neuronowych Systemów Tłumaczenia Maszynowego.

## 6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami)<sup>1</sup> moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak X)

<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

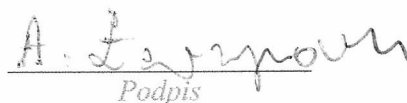
B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Ponadto, biorąc pod uwagę osiągnięcia kandydata w zakresie wyzwań w obszarze neuronowych systemów tłumaczenia maszynowego **rekomenduję wyróżnienie rozprawy doktorskiej.**

  
Podpis

<sup>1</sup> <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276>