



**prof. dr hab. Jerzy Tyszkiewicz**  
**Instytut Informatyki**  
**Uniwersytet Warszawski**

Warszawa, 10/05/2015

**Recenzja rozprawy doktorskiej**  
**„Odpowiadanie na pytania w języku polskim z użyciem głębokiego**  
**rozpoznawania nazw”**  
**autorstwa mgr. Piotra Przybyły**

## **Wstęp**

Niniejszą recenzję piszę z pozycji człowieka, który nie pracuje naukowo w głównym nurcie automatycznego przetwarzania języka naturalnego, ani nie jest znawcą tematyki automatycznego odpowiadania na pytania. Jestem za to wikipedystą, o czym Rada mogła nie wiedzieć, gdy powoływała mnie na recenzenta, a co może mieć jakiś wpływ na mój punkt widzenia na rozprawę. Jako zatem fachowiec niepełny, z konieczności nie będę się specjalnie wdawał w kwestie techniczne i dyskusję detali poszczególnych rozwiązań, postaram się za to spojrzeć na sprawę z szerszej perspektywy.

## **Zawartość merytoryczna rozprawy i jej ocena**

Recenzowana rozprawa składa się z następujących rozdziałów:

1. Wprowadzenie
2. Zadanie odpowiadania na pytania
3. Budowa systemu RAFAEL
4. Głębokie rozpoznawanie nazw
5. Ewaluacja
6. Wnioski

Celem, który autor sobie postawił, było stworzenie systemu RAFAEL, automatycznie odpowiadającego na pytania sformułowane w języku polskim, bez ograniczenia dziedziny, jednak z ograniczeniem do pytania o proste fakty. Zasobem wiedzy systemu miała być polskojęzyczna Wikipedia.

Cel ten mieści się w polu badawczym „Question answering”, które jest na świecie rozwijane od początku lat 60. ubiegłego wieku. Badania o takim charakterze nakierowane na języki fleksyjne, w tym polski, były dotąd prowadzone wyraźnie mniej intensywnie niż dla angielskiego. Historia tych badań jest wyczerpująco omówiona w pierwszym i drugim rozdziale rozprawy.

Osiągnięcie zaplanowanego celu wymagało zaprojektowania architektury całego systemu, złożonej z wielu modułów realizujących zadania cząstkowe, oraz ich dość skomplikowanych interakcji.



Architektura ta jest, jak zauważa sam autor, dość typowa, a wartość naukowa projektu jest zawarta w poszczególnych modułach, przy czym nie jest równo między nie rozłożony.

Sam autor w tytule rozprawy wskazuje, że bardzo ważnym jej elementem jest głębokie rozpoznawanie nazw (DeepER). W dużym skrócie, ta metoda, zaproponowana przez doktoranta, jest stosowana w momencie, gdy system dokonał już analizy pytania oraz wykonał zapytanie, uzyskując znacznie zredukowany zbiór dokumentów (w tym wypadku: artykułów Wikipedii), w których należy oczekiwać pożądanego odpowiedzi. Z analizy pytania pochodzi typ oraz potencjalne synsety poszukiwanej jako odpowiedź nazwy. Wówczas wszystkie wzmianki w wyselekcjonowanych artykułach źródłowych, które dotyczą bytów o poszukiwanym typie również są oznakowywane ich potencjalnymi synsetami. Następnie badanie dostępnej w sieci WordNet relacji hiperonimii pozwala stwierdzić, które ze wzmianek dotyczą bytów, które można również zakwalifikować do synsetów bytu poszukiwanego jako odpowiedź. Metoda ta pozwala precyzyjnie wybierać spośród wielu potencjalnie pasujących do pytania wzmianek te, których szansa na bycie poprawną odpowiedzią jest największa.

Pomysł ten brzmi prosto, ale jego realizacja była tak naprawdę skomplikowana i pracochłonna. Niezbędne było na przykład identyfikowanie synsetów na poziomie analizy pytania i analizy wstępnie wybranych dokumentów. Kluczem było tutaj stworzenie bazy danych bytów, pochodzącej z analizy fraz definicyjnych z artykułów Wikipedii. To dzięki niej możliwe było oznakowywanie pytania i wzmianek synsetami. W dotychczasowych systemach QA te zadania nie były wykonywane. W tym krótkim opisie widać podstawową trudność badań, prowadzonych przez doktoranta: nawet naturalne, zdroworozsądkowe pomysły wymagają stworzenia mechanizmu gromadzącego informacje niezbędne do ich zastosowania, a często rozbijają się o brak odpowiedniego zasobu.

Ażeby sprawdzić sensowność tego pomysłu, autor porównał wykonaną implementację z analogicznym rozwiązaniem, w którym w miejsce DeepER wykorzystane zostały dwa różne standardowe narzędzia realizujące prostsze zadanie Named Entity Recognition (NER). Poprawa wyników systemu była wyraźna, wskazując na trafność pomysłu oraz umiejętną jego realizację w praktyce.

Z pomysłów zawartych w rozprawie bardzo sposobną mi się mechanizm automatycznego sprawdzania poprawności odpowiedzi na pytania, również oparty na mechanizmie DeepER. Znakomicie usprawnia on eksperymenty, poprzez dokonanie sprawdzenia, czy wyselekcjonowana przez system jako odpowiedź wzmianka dotyczy tego samego bytu, który znajduje się w (ręcznie stworzonej) tabeli odpowiedzi na pytania. W tym momencie potrzebna jest ostrożność, bo z jednej strony RAFAEL wykorzystuje mechanizm DeepER do selekcji odpowiedzi spośród wzmianek, a z drugiej strony ten sam mechanizm ocenia poprawność tego wyboru. Na szczęście interweniuje tu dostarczona z zewnątrz poprawna odpowiedź, dzięki czemu nie mamy do czynienia z „byciem sędzią we własnej sprawie”.

Nie zdziwiłbym się, gdyby to właśnie ten mechanizm kontroli poprawności stał się powodem zainteresowania DeepER ze strony społeczności QA.

Największym sukcesem doktoranta jest jednak to, że RAFAEL powstał jako całość, działa ze skutecznością na dobrym światowym poziomie, mimo tego, że język polski stawia ze względu na swoją złożoność strukturalną większe przeszkody niż wiele innych.

Rozprawa napisana jest bardzo starannie w warstwie językowej. W trakcie kilkukrotnej lektury zdołałem zauważyć tylko jedną drobną pomyłkę, przy czym nie udało mi się jej ponownie odnaleźć w trakcie pisania recenzji.

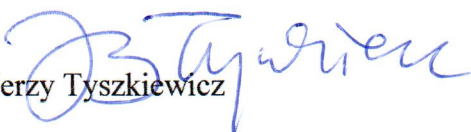
#### **Uwagi polemiczne i krytyczne**

Wykresy na stronie 101 są bardzo trudne do analizy. Każdy z nich jest sporządzony w innej skali pionowej i niezmiernie trudno jest porównywać prezentowane przez nie rezultaty.

Jako wikipedysta, jestem świadom wielu braków i niedociągnięć Wikipedii. Bardzo jestem ciekaw, czy i w jakim stopniu błędy językowe obecne w artykułach wpływają na proces odpowiadania na pytania. Wydaje się, że dla systemów odpowiadających na pytania, marzenie o bezbłędności językowej źródeł jest całkowitą utopią. Jasne jest też, że w miarę coraz dokładniejszej analizy treści źródłowych, niezbędnej dla poprawiania skuteczności, błędy takie będą coraz większym obciążeniem. Chciałbym się dowiedzieć, na ile rozwiązania zaimplementowane przez autora są odporne na niepoprawne formy fleksyjne, składniowe i interpunkcyjne. Szkoda, że testy systemu RAFAEL nie objęły tego zagadnienia.

### **Konkluzja**

**Rozprawa mgr. Piotra Przybyły spełnia wszelkie formalne i zwyczajowe wymogi stawiane rozprawom doktorskim i wnioskuje o dopuszczenie jej autora do dalszych etapów przewodu doktorskiego.**

  
Jerzy Tyszkiewicz