

Gabriela Wawakowska

Rozpoznawanie jednostek nazwanych i ekstrakcja informacji z dokumentów różnego typu

Streszczenie

Rozprawa doktorska prezentuje nowatorskie wykorzystanie metod rozpoznawania jednostek nazwanych i ekstrakcji informacji podczas przetwarzania dokumentów różnego typu. Praca składa się z czterech artykułów naukowych, które zostały opublikowane i zaprezentowane na konferencjach o zasięgu międzynarodowym.

Rozdział 1 opisuje problem badawczy, motywację i uzyskane efekty, a także strukturę i zakres rozprawy. Zawiera również przegląd oraz krótkie podsumowanie załączonych artykułów. Każdy opis poprzedzony jest informacją o autorach, miejscu i typie prezentacji oraz o wkładzie autora rozprawy.

Rozdziały 2 i 3 przedstawiają prace badawcze związane z wykorzystaniem metod rozpoznawania jednostek nazwanych, które posłużyły jako część rozwiązania problemów zdefiniowanych w konkursach organizowanych w ramach międzynarodowych konferencji. Rozdział 2 zawiera opis systemu tłumaczenia powstałego w ramach konferencji WMT 2022. W rozdziale 3 zaprezentowano nowe modele lematyzacji jednostek nazwanych, które zostały zastosowane w rozwiązaniu konkursu organizowanego w ramach warsztatu Slavic NLP 2023.

Rozdziały 4 i 5 skupiają się na artykułach prezentujących modele sieci neuronowych powstałe w ramach prac wdrożeniowych. Rozdział 4 opisuje model TILT powstały w ramach prac nad ekstrakcją informacji z dokumentów o dwuwymiarowej strukturze (warstwa tekstowa i wizyjna). W rozdziale 5 przedstawiono model STable będący rozwinięciem modelu TILT i służący do ekstrakcji danych tabelarycznych.

Na końcu pracy znajdują się załączniki, w których zawarte są trzy certyfikaty otrzymane od organizatorów konferencji ICDAR 2019, WMT 2022 i Slavic NLP 2023, a także pierwsze strony dwóch uzyskanych patentów związanych z modelami TILT i STable. Jako ostatnie zamieszczone zostały deklaracje o wkładzie współautorów każdego z artykułów przedstawionego w rozprawie.

Gabriela
Wawakowska