

Streszczenie

Dostęp do informacji jest kluczowy dla dzisiejszego świata. To dzięki dostępowi do nich, możliwy jest nasz rozwój. Największym źródłem informacji jest Internet. Umożliwia on łatwy, szybki i tani dostęp do zasobów wiedzy z każdej dziedziny życia. Możliwość oceny danych, a następnie powiązania ich w celu późniejszego przetwarzania stanowi jednak problem. W związku z ogromną ilością nowych treści nieuniknionym jest pojawianie się w nich błędów. Tematyka sposobu korekty tekstu jest szeroko znana. Niestety nadal brakuje metod potrafiących dokonać jego automatycznej poprawy, biorąc pod uwagę jego semantykę.

Niniejsza rozprawa podejmuje wskazany powyżej problem przedstawiając możliwości wykorzystania nowego rozwiązania. Celem nadrzędnym pracy było zbudowanie innowacyjnego mechanizmu służącego do efektywnego, asocjacyjnego gromadzenia, kompresowania, a następnie przetwarzania tekstu, tak aby stworzyć algorytmy do jego późniejszej, semi-automatycznej korekty. Zaprojektowano specjalistyczny graf lingwistyczny, do którego budowy wykorzystano teksty pochodzące z różnych źródeł. Następnie jego działanie zweryfikowano poprzez eksperymenty pokazujące interesujące właściwości zaproponowanego rozwiązania. Należą do nich:

- korekta tekstu za pomocą zaimplementowanej metody pobudzeń asocjacyjnych;
- porównanie korekty tekstu z istniejącymi aplikacjami dla języka angielskiego i języka polskiego;
- porównanie korekty tekstu dla dzieł literackich.

Algorytmy oparte o zaproponowaną strukturę grafu zostały zatem przetestowane, a otrzymane rezultaty okazały się zadowalające. W badaniach porównano i potwierdzono skuteczność stworzonych metod do korekty tekstu wprowadzonego z różnego rodzaju błędami, jak również w kontekstowym zapamiętywaniu nauczonych wypowiedzi. Algorytmy były w stanie znaleźć, przetworzyć i powiązać relacje pomiędzy kolejnymi słowami w taki sposób, aby „zrozumieć” czytane zdanie w danym kontekście słownym.

W rezultacie algorytmy kontekstowej korekty tekstu wraz z zaproponowanym modelem zapisu zdań stały się warstwą logiki dla opracowanej aplikacji internetowej, w której użytkownik może wprowadzać tekst z błędami, a program, wykorzystując opracowane algorytmy, jest w stanie dokonać jego semi-automatycznej korekty. Specjalistyczny graf lingwistyczny może zatem w przyszłości być wykorzystywany dla różnych języków naturalnych stanowiąc, podstawę do tworzenia nowych rozwiązań korekty tekstu.