

Kiedy szukamy znaczenia jakiego słowa w języku obcym, zazwyczaj sięgamy po słownik dwujęzyczny, który najczęściej jest listą słów i ich odpowiedników w języku obcym. Jednak wraz z rozwojem technologii nowe zasoby językowe, oferujące bardziej obszerne spektrum informacji oraz wszechstronnie opracowane dane językowe, stały się coraz powszechniejsze. W proponowanym projekcie wykorzystamy jeden z nowoczesnych zasobów - tzw. wordnety. Wordnet można scharakteryzować jako obszerną elektroniczną bazę znaczeń słów połączonych ze sobą za pomocą rozbudowanej siatki relacji leksykalno-semantycznych, które wyrażają różne rodzaje powiązań między znaczeniami słów. Dla przykładu, wyrazom samochód, auto i wóz przyporządkowane są w wordnecie jednostki leksykalne {samochód 1}, {auto 1} i {wóz 1}, które razem tworzą zbiór wyrażający to samo pojęcie, czyli tzw. synset. Ten synset jest połączony relacją hiponimii (podrzędności) z synsetem {autobus 1}, który zbudowany jest z tego samego z nim jednostki leksykalnej, oraz relacją hiperonimii (nadrzędności) z synsetem {pojazd drogowy 1}. W trakcie prowadzonych przez nas badań skupimy się na porównaniu Słownosieci - wordnetu języka polskiego oraz WordNetu pryncetonskiego - wordnetu języka angielskiego. Porównywane wordnety są dwoma niezależnymi modelami reprezentacji leksyki języka polskiego i angielskiego. Celem przeprowadzonych badań, które naszym zdaniem będą przydatne zarówno tłumaczom, leksykografom jak i uczącym się języka angielskiego, jest określenie stopnia przystawalności danych językowych zebranych w obu wordnetach. W szczególności interesujące dla nas będzie zbadanie jakiego rodzaju luki leksykalne i rozbieżności mogą występować pomiędzy powiązaniem zasobami Słownosieci i WordNetu pryncetonskiego oraz jakie mogłyby być przyczyny ich występowania (np. różnicami kulturowymi, różnicami typologicznymi między językiem polskim i angielskim, różne sposoby budowy Słownosieci i WordNetu pryncetonskiego). Ponadto po zbadaniu stopnia przystawalności zasobów obu wordnetów, wykorzystamy zdobyte wiedzę i informacje do zaprojektowania metod łączenia wordnetów na poziomie jednostek leksykalnych. Takie połączenie wordnetów będzie nowym w skali światowej, gdyż obecnie wszystkie wordnety łączone są ze sobą na poziomie synsetów. Tym samym, w obszarze językoznawstwa komputerowego, nasze badania znacznie przyczynią się do rozwoju metod pozyskiwania informacji z tekstów dwujęzycznych.