

Úkol hledání sémantických vztahů mezi entitami na základě předloženého textu označujeme jako extrakci vztahů (relationship extraction). Metodou distant supervision, která spočívá ve spojení báze znalostí (Wikidata) a korpusu (české Wikipedie), jsme vytvořili Český dataset pro extrakci vztahů (CERED). Použitou metodiku a problémy, na které jsme narazili, důkladně rozebíráme. CERED využíváme při tréninku neuronové sítě pro extrakci vztahů. Základem této sítě je BERT – lingvistický model předtrénovaný na velkém množství prostého textu. Navržený model vyhodnocujeme na anglických datech (Semeval 2010 Task 8, TACRED) a porovnáváme jeho kvalitu s ostatními výsledky v oblasti extrakce vztahů. Přikládáme i výsledky naměřené na CEREDu.