

Tato dizertační práce se zabývá automatickou extrakcí sémantických vztahů z textů. Jsou zkoumány jak metody strojového učení tak i pravidlový přístup. Pro každý přístup se experimentálně ověřuje vliv různých nástrojů pro zpracování přirozeného jazyka na úspěšnost systému. V práci implementujeme systém REextractor, a vylepšujeme několik nástrojů pro zpracování přirozeného jazyka: od větné segmentace a tokenizace až ke automatickým syntaktickým parserům. Kromě toho, představujeme Český korpus právních textů s několika vrstvami anotací. Korpus byl použit na trénování a testování všech komponent systému REextractor. Výsledky a nástroje v předkládané práci byli publikovány jako nové zdroje v projektu Sémantického webu.