

## Abstrakt

Univerzální sémantický jazyk (USJ) je semi-formalizovaný způsob zápisu znalostí (systém pro reprezentaci znalostí). Myšlenka USJ byla rozvinuta Vladimírem Smetáčkem v 80. letech při pracích na systému SĚMAN (Univerzální sémantický analyzátor). Tento systém byl využíván pro automatizovanou extrakci klíčových slov v tehdejší informačním centru ČSSR. Avšak se zánikem centra v 90. letech byl systém SĚMAN ztracen.

Tato dizertace oživuje myšlenku USJ v novém kontextu automatizované obsahové analýzy. Nejdříve prezentujeme historický kontext a problémy spojené s reprezentací znalostí, sémů, sémantických polí, sémantických primitivů a univerzálií. Dále je představena metodika kvantitativní obsahové analýzy na příkladu tří klasických aplikací. Podrobně popíšeme architekturu nové aplikace, která byla vyvinuta speciálně pro potřeby evaluace USJ. Program může fungovat jako nástroj pro klasickou obsahovou analýzu, avšak obsahuje i nástroje pro zpracování přirozeného jazyka (NLP) a využívá algoritmů pro vyhledávání kolokací. Tyto byly upraveny pro potřeby vyhledávání vazeb mezi sémantickými anotacemi.

Jednotlivé součásti programu jsou podrobeny praktickým testům. Subsystem pro vyhledávání vzorů v textech je porovnán s existujícím extraktorem klíčových slov. Mechanismus pro překlad do sémantických kódů je otestován na příkladu automatické klasifikace dokumentů a speciální pozornost věnujeme problémům mnohoznačného překladu. Poslední část evaluace se zabývá schopnostmi systému objevit významné vazby mezi sémantickými anotacemi v textu.