Inteligentní knihovna - INTLIB

Existuje SW projekt, který:

- je **zajímavý**?
- může být **rozšířený do diplomek**?
- má potenciál **komerčního využití**?



Ano! INTLIB je, může i má!

Co je INTLIB?

INTLIB je SW projekt, který umožní chytré vyhledávání v zadané množině dokumentů, ve kterých jsou reprezentovány objekty a vazby mezi nimi z určité předem dané domény. Příkladem jsou legislativní dokumenty, tj. jednotlivé zákony, případně omezené na určitou oblast (stavebnictví, daně apod.), v nichž jsou definovány pojmy, povinnosti a práva, a ty jsou dále rozšiřovány, upravovány a vzájemně se ovlivňují.

V projektu využijeme existující metody počítačové lingvistiky (morfologie, větný rozbor, atd.) a rozšíříme je o práci s řízenými slovníky (taxonomie, thezaury, ontologie). Cílem je v dokumentech indexovat objekty a vztahy mezi nimi posané řízenými slovníky. Vybudujeme tak silnější vyhledávací nástroj než současné metody fulltextového vyhledávání. Součástí projektu bude také vybudování intuitivního GUI pro dotazování a procházení se sofistikovanou nápovědou pro upřesňování dotazů a také pro editaci řízených slovníků.

Výsledky projektu budou **komerčně využity v projektu aplikovaného výzkumu** v oblasti vyhledávání v legislativě, který probíhá na KSI MFF UK. Např. to bude znamenat prohledávání:

- bez ohledu na tvarosloví (např. smlouva, smlouvy, smlouvách),
- v rámci nadřazených/podřazených výrazů (např. smlouva, nájemní smlouva),
- z hlediska významu textu (např. odstavec, který pojem definuje/rozšiřuje/využívá),
- s možností upozornění na připravovanou změnu v zákoně (uživatel specifikuje oblasti, které ho zajímají v případě změny).

Co ty diplomky?

Typická diplomka, která naváže na INTLIB, se zaměří na vybraný modul (např. lingvistické zpracování textu, indexace, dotazování nebo doplňování slovníků), zanalyzuje možná vylepšení, navrhne optimální, implementuje a experimentálně zhodnotí vůči původnímu modulu. Díky modularitě se nebude zaobírat ostatními částmi, uživatelským rozhraním apod., ale může se zaměřit přímo na jádro konkrétního problému.

Kontakt:
Irena Mlýnková [mlynkova@ksi.mff.cuni.cz], Martin Nečaský [necasky@ksi.mff.cuni.cz]
Místnost 209