

Martin Zeman: Využití ontologií pro GUHA procedury

Práce Martina Zemana je aktuálně poslední z řady prací věnujících se využití doménových znalostí pro proces dobývání znalostí z databází (DZD). Téma je zpracováváno zejména na KIZI VŠE, kde v souvislosti s ním vyšla skoro desítka odborných článků a zhruba poloviční počet diplomových prací. V současné době se výzkum shlukuje kolem dvou systémů: systému LIsp-Miner a jeho napojením na apriorní znalosti v rámci projektu SEWEBAR na straně jedné a systému Ferda na straně druhé. Práce vychází z méj diplomové práce, která představovala první návrhy na automatizaci postupů využití doménových znalostí pro DZD. Ze dvou hlavních typů doménových znalostí (*background knowledge* a ontologií) jsem se rozhodl implementovat podporu pro první typ. Do implementace podpory ontologií jsem se vzhledem k pracnosti nepouštěl, nicméně jsem navrhl přesný popis, jakým způsobem ontologie implementovat.

Martin Zeman tento popis vzal a dokázal jej přetvořit v rozumně fungující implementaci. Musel přitom řešit problémy technické i teoretické. Z technických problémů byly hlavními problémy volba ontologického jazyka (OWL Full) a propojení NET Frameworku a Javy pomocí middleware Ice v systému Ferda a implementace krabiček pro postupy využití ontologií v systému Ferda. Tyto problémy diplomant vyřešil. Teoretické problémy byly zejména reprezentace přidaných doménových znalostí ve vybraném jazyku a vymyšlení algoritmu na konstrukci atributů podle těchto informací z ontologie. Teoretické problémy diplomant taktéž zvládl.

Práce má nicméně i svoje nedostatky. Implementace obsahovala při odevzdání velké množství malých neduhů, některé z nich však autor již odstranil. Hlavním problémem stále zůstává neefektivnost počítání rozdělení kategorií v krabičce atributu vytvořeného z ontologie. Výpočet se (zřejmě v důsledku častého dotazování ontologie) na sloupcích větších domén zpomalí natolik, že Ferda po desítky minut nereaguje na uživatelské příkazy.

Na textu diplomové práce je vidět, že byl psán ve shonu na poslední chvíli před odevzdáním. Obsahuje několik věcných chyb a velkou řadu drobných nepřesností. Dle mého názoru je nedostatečná experimentální část, kde na malém počtu sestavených úloh autor vlastně prohlásí, že implementace funguje. Myslím, že by bylo vhodné provést víc experimentů a na nich provést diskusi, jakým způsobem nově implementované postupy ulehčují postup DZD a co by se dále mohlo zlepšit.

I přes výše jmenované nedostatky práci doporučuji k obhajobě.

V Praze 28.1.08

