

Posudek diplomové práce Martina Ralbovského

Využití doménových znalostí při aplikacích GUHA procedur

Jedná se o diplomovou práci kategorie „výzkumný problém“. Jejím cílem je vytvořit prostředky pro reprezentaci doménových znalostí a jejich využití v GUHA procedurách implementovaných v systému LISp-Miner. Systém LISp-Miner je akademický systém pro výzkum a výuku dobývání znalostí z databází (DZD). Dalším cílem práce je experimentální implementace těchto prostředků. Těmto cílům odpovídá i členění práce.

Práce se podrobněji zabývá dvěma typy doménových znalostí. Prvním typem je tzv. *background knowledge*, druhým studovaným typem znalostí jsou ontologie. Pro oba typy znalostí byla prostudována rozsáhlá literatura týkající se možností jejich použití v DZD. Je zkoumána a diskutována vhodnost aplikace prostudovaných způsobů práce se znalostmi pro GUHA procedury systému LISp-Miner.

Jsou navrženy nové způsoby formalizace *background knowledge* pomocí vhodných pravidel vycházející z vlastností GUHA procedur. Jsou implementovány potřebné nástroje pro konfrontaci těchto pravidel s výsledky jednotlivých GUHA procedur. K implementaci je využito vizuální prostředí Ferda vytvořené v rámci studentského softwarového projektu jehož byl diplomant vedoucím. Implementované nástroje jsou ověřeny na sadě pravidel formalizujících *background knowledge* pro reálný datový soubor z medicínské oblasti. Dále jsou rozsáhlým způsobem diskutovány možnosti využití ontologií při práci s GUHA procedurami včetně možnosti implementace ve stejném prostředí. Je uvedeno shrnutí a navrženy další směry vývoje vytvořených prostředků.

Práci považuji za velmi zdařilou. Za její klady pokládám zejména:

- komplexní přístup k řešení zadaného výzkumného problému
- zavedení pojmů pro práci s *background knowledge* (odstavec 3.2)
- implementaci validace pravidel *background knowledge* (kapitola 5) využívající vizuální prostředí vytvořené v rámci dříve obhájeného softwarového projektu
- rozsáhlou diskusi o ontologiích (hlavně kapitola 6)
- návrhy navazujících směrů vývoje.

Za nedostatky práce považuji (řídce se vyskytující) překlepy a potíže s nastavením connection string při spouštění programu. Za nedostatek, který ale jistě není nezpůsoben autorem, je třeba považovat i občasné problémy s převzatým modulem *ResultBrowser*. Uvedené nedostatky však žádným způsobem nesnižují výsledky dosažené v rámci hodnocené práce.

Diplomovou práci Martina Ralbovského doporučuji k obhajobě.

Praha, 4.9.2006


Doc. RNDr. Jan Rauch, CSc.