

Alexander Kuzmin Relační GUHA procedury

Původní česká metoda GUHA je historicky první metodou explorační analýzy dat a předstihla vývoj i možnosti soudobé výpočetní techniky o dvě desetiletí. Metoda definuje obecný rámec pro získávání nových zajímavých znalostí z dat, nejpoužívanějším typem znalostí v posledních letech je získávání asociačních pravidel. V původní verzi fungovalo dobývání asociačních pravidel pouze nad jednou tabulkou. Již v sedmdesátých letech ale přišel doc. J. Rauch s nápadem přidat ke stávajícímu možnost dobývání nad více tabulkami svázanými pomocí relačí. S nápadem však vyvstaly teoretické i implementační problémy, které se v plné šíři podařilo vyřešit až v diplomové práci „Relační GUHA procedury“.

Autorovi se podařilo navrhnout a implementovat relační rozšíření GUHA procedur 4FT a SD4FT v prostředí Ferda. Součástí řešení měl být i expertní systém. Později se upustilo od expertního systému a přistoupilo se řešení aktuálního problému implementace některých typů atributů do prostředí Ferda. Autor tedy implementoval i moduly pro atributy a zásadním způsobem vylepšil stávající modul pro úpravu kategorií a intervalu atributu.

S úrovní implementace jsem spokojen. I když byla počáteční implementace atributů na začátku plná chyb, autor dokázal tyto chyby odstranit. Z možných variant realizace hypotézového atributu byla i podle mého názoru vybrána ta nejlepší. Moduly poskytující bitové řetězky se dobře hodí do celkového návrhu prostředí Ferda. Jedinou slabinou zůstává nemožnost zapojení více hypotézových atributů za sebou. Tento fakt autor v textu popisujícím implementaci obchází s tvrzením, že by stejně zapojení dvou hypotézových atributů za sebou nemělo dobrý smysl a že by to trvalo neúměrně dlouho. S argumentací souhlasím, lepší by však bylo, kdyby se nejdříve zkusilo zapojení dvou hypotézových atributů a potom by se z experimentů vyvozovaly závěry, popřípadě kdyby se do textu napsalo, proč to stávající implementace neumožňuje.

Text diplomové práce se mi taktéž zdá jako zdařilý. Autor přehledným a věcně správným způsobem uvádí čtenáře do problematiky. Jednotlivé části na sebe dobře navazují. Celá práce je prokládána příklady vysvětlujícími náročnější konstrukce, což je cenné zejména u ukázek relačních hypotéz či zapojení nových modulů do Ferdy.

Jediným vážným nedostatkem diplomové práce je její rozsah. Pro posouzení množství vykonané práce je důležité si uvědomit, co všechno již bylo vyzkoumáno či implementováno ještě před zahájením práce. Použitá teorie o relačním dobývání znalostí asociačních pravidel pomocí metody GUHA byla již prozkoumaná J. Rauchem a T. Karbanem a uvedena v několika publikovaných pracích, nejkompaktněji v dosud neodevzdané disertační práci T. Karbana. Co se týče implementace, autor nemusel implementovat samotné GUHA procedury, pouze relační rozšíření stávajících modulů GUHA procedur v prostředí Ferda. GUHA procedury byly dobře navrženy a implementovány v prostředí Ferda v rámci diplomové práce T. Kuchaře.

Autor měl tedy všechny podklady připravené a implementoval relační rozšíření procedur, což spočívalo v návrhu tří nových modulů a jejich napojení na stávající moduly, modifikaci některých sdílených knihoven, domyšlení implementačních detailů a samotnou implementaci nových modulů. Toto množství práce dle mého názoru neodpovídá rozsahu diplomové práce. Proto byly přidány ještě moduly pro atributy, což situaci vylepšuje, nicméně i tak si myslím, že se diplomová práce řadí k těm méně pracným. Na druhou stranu musím poznamenat, že

kvalita implementace i textu je dobrá a výsledkem je ucelený výtvor, který může být téměř bez zásahů dále používán a prezentován před vědeckou veřejností.

Nedostatek programátorské práce nevnímám jako autorovu vinu, spíše to slouží ke cti těm, kteří pro autora připravili dobré výchozí podmínky. Zlepšení rozsahu mohl autor docílit větším počtem experimentů a testů hypotézových atributů. V rámci práce byly provedeny pouze pilotní testy zabývající se výkonností. Nesmírně zajímavé by bylo zkoumat sémantiku relačních hypotéz a právě to je směr, kde by práce mohla nabídnout mnohem více.

I když je diplomová práce „Relační GUHA procedury“ relativně malého rozsahu, jedná se o ucelené, dobře zpracované i implementované dílo. Zároveň je toto dílo významné jako historicky první dostupná funkční implementace relačních GUHA procedur. Proto práci doporučuji k obhajobě.

