

Oponentský posudek na diplomovou práci
Miroslav Blaško
Modelling n-ary relations in description logic

Práce je z oblasti ontologického inženýrství které se rozvíjí v souvislosti s vizí sémantického webu. Konkrétně, zadání (podrobně viz zadání) ukládá obeznámení se s n-árními deskripčními logikami (DL), srovnat existující přístupy i výpočtovou sílu nástrojů, navrhnout a implementovat n-ární dedukční stroj (reasoner) a ověřit algoritmy na *n-árních ontologiích vytvořených z binárních* (zvýrazněno oponent – to není nejšťastnější rozhodnutí, protože hlavní myšlenka přístupu je pro n-ární ontologie použít binární deduktory a ne nejdřív binární ontologie přeložit do n-árních...). Práce téměř vůbec nemluví o webu (natož sémantického), více řeší vnitřní problémy ontologického inženýrství. Jedna z motivací pro n-ární DL je použití pro anotaci webovských zdrojů, nicméně o anotaci se toho moc v této práci nedozvíme.

Práce je psána anglicky. Je členěna následovně. Začíná popisem základů deskripčních logik (DL), pak logiky SHOIQ a n-ární DLR a OWL-DL. Celkově je materiál čtivý a srozumitelný. Další kapitola dává dobrý přehled o současném stavu problematiky ve světě i o EU CIPHER (<http://cipherweb.open.ac.uk>, viz také <http://krizik.felk.cvut.cz/cipher/>) ve kterém se diplomant angažuje.

Další motivace je negativní hodnocení W3C dokumentu „Defining N-ary Relations on the Semantic Web“ (<http://www.w3.org/TR/swbp-n-aryRelations/>). Tvrdí se „...Document explains how to model n-ary relations by so-called reification, i.e. representing n-ary relation by concept. Such a representation is, however, weak form of reification in which n-ary relation does not have to be interpreted as a set of tuples...“. Nicméně všechny příklady v práci pracují jenom s relacema které vyhovují W3C přístupu (protože vznikly z binárních DL). Materiál W3C má ještě daleko ke standardu, patrně je třeba najít dobrou motivaci (use case) který přesvědčí tvůrce, že je potřeba model rozšířit (ani předkládaná práce to nečiní).

Od kapitoly 4 začíná vlastní obsah práce. Definuje NDL (v něčem bohatší, v něčem chudší než DLR), patrně proto že se používá v projektu CIPHER. Použitý příklad je velice jednoduchý.

Základ práce tvoří dva typy překladů (syntaxe, T boxů, A Boxů, ...) z binárních na n-ární a z n-árních do binárních DL (samozřejmě tak, aby sa zachovala splnitelnost a syntaktická korektnost (25^{15} pojem “tuple admissibility” by zasluhoval přesnější definici)). Kapitola 5 se zabývá reifikačním algoritmem (postup běžný např. při převodu ER diagramů na binární). Nicméně tady to má význam, protože pro binární DL máme deduktory. Trochu mi scházel podrobnější popis, např. že $\text{dom}(f_i)=T_n$, $\text{rng}(f_i)=T_1$. Ale prokousal jsem se tím a je to v pořádku, i dodatečné axiomy které zabezpečují korektnost překladu. Přiznám se že překlad A-boxu od Calvaneseho bych byl oželel (více se mi líbí Horrocksův přístup) a místo toho ocenil popis software (případně autorův přínos – předvedl mi ho ale popsal chabě). Tyto dvě kapitoly považuji za podstatné z hlediska celé práce.

Kapitola 6 se zabývá tvorbou testovacích dat. Byl zvolen nejproblematičtější přístup, a to překlad z binárních DL. Autor W3C na 17^{10-11} vytýká „One of disadvantages is that W3C design patterns provide diferent representation choices for semantically same relation.“ Nicméně právě tato kapitola generuje nečitelné ontologie a překlad $DL \rightarrow NDL \rightarrow DL$ produkuje různé reprezentace pro semanticky stejné relace. Další nevýhodou je, že buď generuje NDL s binárními relacemi nebo je potřeba napsat ručně XML Schema na překlad (v práci jsem to nenašel, jen při předvádění se dozvěděl).


Poslední kapitola obsahuje srovnání reifikačních algoritmů. Teoretické srovnání je zajímavé a ukazuje že A-SHIQ a A-SHIF překlady jsou efektivnější než NDL. Mnohem zajímavější je část o empirickém srovnání implementace. Poznámka o populaci instancí

59_{5-7} se neprojeví v tabulkách – tam se srovnávali DL s nejvýše 1-2 instancemi. Opět Horrocks a Blažkův algoritmus je lepší než Calvanese, a Racer je jasně lepší než Pellet a FaCT++. Schází ještě srovnání s původní složitostí ukolů pro binární zdroje.

Všechny připomínky je třeba spíš chápat jako podněty do další (PhD?) práce a v zásadě nesnižují kvalitu předkládané diplomové práce. Největším problémem práce je absolutní nedostatek dokumentace implementace (v práci není ani zmínka o tom, které části jsou implementovány). Nicméně, autor mi implementaci předvedl a on je schopen ji ovládat (mnohdy v řádkovém příkazovém režimu, pochybuji že někdo to po něm bude schopen používat). Ze zadání tedy schází i popis návrhu (bod 2 design) a už jak bylo řečeno výše anotace (bod 1. Semantic annotation creation).

Práci doporučuji k obhajobě.

V Praze 17.5.2008



Prof. RNDr. Peter Vojtáš, DrSc.
oponent

Technické poznámky:

8^{5-7}R - nepřesnosti (přebytečné \exists)

9_{7-8} - schází označení pro atomární role (stejně na 20_6), které se v dalším používá, taky mi trochu schází $\exists[\$i]R$, pak na 10_5 $\forall[\$i]R$ dává jednoprvkovou množinu...taky mi schází zmínka, že \exists resp. $\forall[\$i]R.C$ se dá definovat, a když R je binární, tak operace dávají totéž jak v binární DL).

21_2 schází číslo obrázku

41 obr.6.1 je z W3C webu, nezmněno

64 popisek k obrázku 7.7 asi nepřesný