

## Review of Bachelor's Thesis

**Student:** Kmenta Martin  
**Title:** Improvements of the ASMA Tool for Analysis of String Manipulating Programs via Symbolic Automata (id 25149)  
**Reviewer:** Lengál Ondřej, Ing., Ph.D., DITS FIT BUT

- 1. Assignment complexity** **more demanding assignment**  
Zadání bylo **obtížnější**: student se musel seznámit s pokročilou teorií konečných a symbolických automatů (včetně např. použití relace simulace pro redukci počtu stavů automatů), abstraktním regulárním model checkingem a technikami pro verifikaci programů s řetězci.
- 2. Completeness of assignment requirements** **assignment fulfilled**  
Zadání bylo splněno.
- 3. Length of technical report** **in usual extent**  
Rozsah technické zprávy je v obvyklém rozmezí.
- 4. Presentation level of technical report** **70 p. (C)**  
Rozsah a návaznost kapitol je na dobré úrovni. K prezentační stránce technické zprávy mám několik výtek:

- Definicím v kapitole 2 chybí důležité části a obsahují chyby, např. v definici konečného automatu není definován jeho *běh*, konečné automaty jsou zavedeny bez *epsilon* pravidel a následně je definována jejich podtřída bez *epsilon* pravidel, definice *úplného* automatu obsahuje chybu, definice *symbolického automatu* pracuje s nedefinovanými funkcemi *DomainElements* a *Predicates*, apod.
- Kapitola 3 by si zasloužila lepší popis analyzační metody nástroje ASMA, např. jakým způsobem jsou reprezentovány množiny konfigurací programu, jak se dělá abstrakce, jak vypadají vstupní programy s příkladem, atp.
- Kapitola 4 se zabývá optimalizací reverze operace konkatenace. Kapitola relativně detailně popisuje čtyři různé zvolené přístupy, tak, jak byly postupně navrhovány a systematicky testovány. Tady mám hlavní výtku ke způsobu, jak jsou popisovány algoritmy. Jejich popisy obsahují buď příliš velkou míru abstrakce (např. "foreach state in run do: if state is not noted" v Algoritmu 2, kde "run" ani "noted" nejsou definované) nebo příliš nízkou, skoro na úrovni zdrojového kódu (např. "Algebra.AreEquivalent(Tmove.label, Smove.label)" v Algoritmu 7 by šlo zapsat např. jako "Tmove <=> Smove" nebo indexovaný for cyklus na řádce 10 v tom samém algoritmu by šel zapsat jinak, vzhledem k tomu, že s indexující proměnnou se v cyklu nepracuje).
- V kapitole 5 popisující experimenty mi chybí srovnání velikosti výstupu determinizace a minimalizace vůči redukci pomocí simulace či bisimulace, v kapitole jsou uvedeny jen časy. Líbí se mi, že jsou porovnány všechny možné optimalizace, ale přijde mi, že testovací sada 14 benchmarků je relativně malá a chtělo bych jich mít víc. V Tabulce 5.6 zvolil jiný způsob výpočtu výsledného zrychlení, např. zrychlení z 63 493 ms na 4 829 ms bych prezentoval spíše jako zrychlení 13x než jako zrychlení o 92%.

- 5. Formal aspects of technical report** **90 p. (A)**  
Typografie bez větších připomínek, snad jen v kapitole 5 by šlo posunout obrázky/tabulky tak, aby úplně na konci kapitoly nebylo pět za sebou bez textu. V práci se vyskytují neodkazované obrázky, např. Figure 2.3 (existující reference na Figure 2.3 má správně být na Figure 2.2).

Práce je psaná solidní čitelnou angličtinou s menším množstvím chyb.

- 6. Literature usage** **60 p. (D)**  
V práci **zcela chybí** popis state of the art v oblasti analýzy programů s řetězci. Počet citovaných zdrojů je spíše menší, ale pokrývá do velké části zkoumanou problematiku (modulo chybějící SOTA). Převzaté části jsou odkazovány a odlišeny, ale v práci se příliš často vyskytují fráze jako "These techniques are described in more detail in the thesis written by Michal Kotoun". Citace [11] je neúplná. U citace [10] bych místo "et al." uvedl posledního zbývajících spoluautora (T. Vojnara).
- 7. Implementation results** **70 p. (C)**  
Demonstrováné řešení je funkční, nástroj ASMA byl doplněn o v práci rozebíranou funkcionalitu. Při demonstraci jsem měl pocit, že se student opravdu jen zaměřil na části, které měl optimalizovat, a moc se nezabýval tím, co,

jak a proč nástroj vlastně má dělat a dělá, což je trochu škoda.

**8. Utilizability of results**

Výsledky jsou zajímavé a očekávám, že budou součástí budoucí publikace na mezinárodním fóru, případně budou sloužit jako základ pro budoucí výzkum.

**9. Questions for defence**

1. Jaké je srovnání velikosti automatu získaných pomocí determinizace a minimalizace a redukovaných pomocí simulace/bisimulace?
2. Jaké jsou další plány s použitím simulace pro redukci automatů v nástroji ASMA? Bude zcela zahozena na úkor determinizace a minimalizace?
3. Jakým způsobem plánujete využít v nástroji zmiňované antichainy?
4. Jaké jsou další plány s nástrojem ASMA?
5. Jak se nástroj ASMA srovná s jinými nástroji pro analýzu programů s řetězci?

**10. Total assessment**

**73 p. good (C)**

Martin Kmenta se ve své práci zabíral optimalizací nástroje pro regulární model checking programů s řetězci. Podařilo se mu navrhnout a naimplementovat optimalizace, které v mnoha případech znatelně zkracují čas analýzy a mohly by sloužit jako základ dalšího výzkumu. Kvalitu práce bohužel sráží její horší prezentace v technické zprávě a menší sada benchmarků, než by stačilo na řádné experimentální vyhodnocení. Z těchto důvodů, rozepsaných detailně výše, hodnotím bakalářskou práci Martina Kmenty stupněm **dobře (C)**.

In Brno 16 May 2022

Lengál Ondřej, Ing., Ph.D.  
reviewer