

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě

Univerzity Karlovy v Praze

posudek vedoucího posudek oponenta

Autor/ka: **Matěj Klonfar**

Název práce: **Strojové učení formálních jazyků**

Studijní program a obor: **Informatika, Programování**

Rok odevzdání: **2007**

Jméno a tituly vedoucího/oponenta: **RNDr. František Mráz, CSc.**

Pracoviště: **KSVI MFF UK**

	excellentní	odpovídající	slabší	nevýhovující
Náročnost zadaného tématu	X			
Míra splnění zadání			X	
Struktura textové části práce		X		
Jazyková a typografická úroveň		X		
Analýza			X	
Vývojová dokumentace		X		
Uživatelská dokumentace			X	
Kvalita zpracování softwarové části		X		
Stabilita aplikace	X			

Nejvýznamnejší klady:

- Základné rozhranie na implementáciu algoritmov strojového učenia jazykov, tzv. gramatickej inferencie, ktoré umožňuje relatívne jednoducho pridať ďalšie algoritmy a reprezentácie učených jazykov.
- Priložená aplikácia obsahuje implementáciu štyroch rôznych algoritmov učenia jazykov.

Nejzávažnejší nedostatky:

- Veľmi nekvalitná textová časť, okrem neúmerného množstva gramatických chýb má mnoho ďalších nedostatkov.:
 - práca neobsahuje zodpovedajúci popis riešeného problému. Napr. ako poznat', že algoritmus vydal správnu odpoveď.
 - V práci sú uvedené definície základných pojmov, ktoré však obsahujú mnoho chýb – napríklad prázdne slovo je označované ako znak, autor definuje abecedu, ale definíciu množiny všetkých slov nad danou abecedou vyniechal a značenie Σ^* používa, podobne definuje gramatiku, ale nedefinuje ako sa od gramatiky dostane k jazyku, autorom definované regulárne gramatiky v skutočnosti generujú lineárne jazyky, definície deterministického a nedeterministického konečného automatu sú zmotané dohromady, takže potom definícia jazyka rozpoznávaného nedeterministickým konečným automatom chýba, definície pre systémy prepisovacích pravidiel, s ktorými aplikácia pracuje sú rozdelené na dve prekrývajúce sa časti – jedna je v textovej časti a druhá v teste na priloženom CD, atď.
 - Úloha automatického učiteľa, ktorý je vlastne určený na testovanie jedného typu algoritmov učenia jazykov, nie je dostatočne vysvetlená.
 - Popis implementovaných algoritmov je veľmi stručný a je iba na priloženom CD.
 - Užívateľská príručka je len na priloženom CD. Podľa nej sa s aplikáciou takmer nedá pracovať viz. nižšie.
 - Formáty súborov, ktoré reprezentujú množiny príkladov a konečné automaty nie sú kompletne vysvetlené. Napr. do abecedy sa pridáva znak „_“ pre prázdne slovo, ale v zápise prázdneho slova ako pozitívneho alebo negatívneho príkladu sa tento znak nepoužíva (to sa dá vyčítať iba z príkladov a experimentovaním s aplikáciou). Formát súboru s konečným automatom si čitateľ musí domyslieť z príkladu, pritom nie je vysvetlené čo sa robí, keď nejaký prechod (daným symbolom abecedy) nie je uvedený. Autor naznačuje, že jeho reprezentácia umožňuje pracovať aj s nedeterministickými automatmi, ale nikde nie je vysvetlené ako zadat' viacej než jeden počiatočný stav, ...
- Chyby v implementácii:
 - užívateľské rozhranie je veľmi obmedzené a užívateľovi nedáva informácie o možnostiach jeho ďalších akcií. Napr. po načítaní konečného automatu môže užívateľ testovať prijímanie slov týmto automatom, ale táto informácia nie je ani uvedená v užívateľskej príručke, ani ju program nesignalizuje. To, že beží výpočet algoritmu sa pozná iba podľa toho, že je povolené tlačítko „Stop“, ale to je povolené aj po dobehnutí algoritmu učenia.
 - autor tvrdí, že užívateľ môže editovať množinu príkladov slov. V skutočnosti užívateľ v aplikácii môže iba pridať príklady. Nemôže príklady mazat', ani meniť ich ohodnotenie z pozitívnych na negatívne a naopak. Pri pokuse o zmenu množiny príkladov.
 - Výsledkom učenia napr. algoritmu RPNI je konečný automat. Aj pre najjednoduchšie príklady z dokumentácie (abeceda $\{a,b\}$, pozitívne príklady

- prázdne slovo, aa, b, negatívny príklad bb) výsledný automat obsahuje prechody do stavov, ktoré automat neobsahuje. Bezprostredne po naučení aplikácia správne odpovedá na dotazy o prijímaní zadaného slova naučeným automatom. Ked' sa však automat uloží a potom načíta zo súboru, tak už je to iný automat rozpoznávajúci iný jazyk.*
- Algoritmy L*, LA nie sú korektne implementované. Po tom čo sa zadá slovo ako protipríklad k navrhovanému automatu sa program znova pýta na to, či je to isté slovo prijímané alebo nie.

Další poznámky:

- Autor prácu vypracoval samostatne a odovzdanú prácu so mnou nekonzultoval. Práca Matěja Klonfara nesplňuje požiadavky kladené na bakalársku prácu, a preto ju nedoporučujem uznat' ako bakalársku prácu.

	výborně	velmi dobrě	dobře	neprospěl/a
Návrh známky				X

Datum: 4.9.2007

Podpis: