

Diplomová práce – posudek oponenta

Pavel Žoha: Algoritmy konstrukce sufixového pole

Předložená práce se věnuje porovnání vlastností různých (již existujících) algoritmů pro konstrukci sufixového pole (Manbera a Myerse, Karkkainena a Sanderse, Sewarda, Manziniho a Ferraginy), jednoho algoritmu (Ukkonen) na konstrukci sufixového stromu a běžných třídících algoritmů. Je zde uvedena motivace i teoretický základ k celé problematice konstrukce sufixového pole.

Jednotlivé algoritmy jsou v textu práce přehledně popsány, včetně důležitých implementačních detailů, nákrešů a uvedení ilustračních příkladů. Pro každý algoritmus je zde i uvedena časová a prostorová složitost. Velmi oceňuji, že autor se zabýval i kontrolou správnosti práce jednotlivých algoritmů a implementoval algoritmus na tuto kontrolu.

Text diplomové práce je srozumitelný a dobře strukturovaný, bez většího množství gramatických, typografických chyb. Faktické chyby jsem v práci nenalezl žádné. Elektronickou přílohu tvoří výborně okomentované zdrojové kódy všech uvedených algoritmů, testovací data a výsledky provedených testů.

Rozsáhlou část této diplomové práce tvoří provedené testy. Autor testoval algoritmy na různých běžně používaných formátech dat (komprimovaný soubor, binární soubor, zdrojový kód, obrázek, text, XML, DNA a dokonce i japonský a čínský text), dále na náhodně vygenerovaných datech a nakonec i na speciálních případech dat, které jsou z hlediska uvedených algoritmů zajímavé. Autor zkoumal jak se mění vlastnosti uvedených algoritmů i v závislosti na velikosti souboru. U testů je uvedena doba výpočtu algoritmu a také spotřebovaná paměť.

Výsledky jednotlivých testů jsou obsaženy ve velkém množství tabulek a výsledně jsou shrnuty v grafickém přehledu, který pro každý soubor a jeho velikost určuje nejvhodnější algoritmus. Rovněž závěry z experimentů jsou shrnuty v textu práce. Chtěl by jsme ocenit, že autor se zajímal i o rozdíly ve výsledcích při použití různých překladů a jejich nastavení. Jedinou kritiku k této práci mám tu, že možná mohla být zvolena vyšší minimální mez velikosti souboru pro testování, protože pro nejmenší velikosti souborů jsou výsledky neporovnatelné (vše menší než 1ms) a zároveň mohly být testovány i soubory větší (třeba desítky MB)

Domnívám se, že jde o velmi zdařilou práci, ve které autor prokázal, že umí nastudovat netriviální algoritmy popsané v odborné literatuře a následně je implementovat v jednotném prostředí. Experimenty pro srovnání algoritmů byly vhodně navrženy a získané výsledky jsou srozumitelně a přehledně vyhodnoceny.

Předloženou práci považuji po všech stránkách za práci splňující kritéria pro diplomové práce na MFF UK a doporučuji ji k obhajobě.

Praha, 22. 1. 2007

Mgr. Jan Lánský, KSI MFF UK

