

Posudek vedoucího diplomové práce **Jindřicha Šedka: Algoritmy nad rozšířeným sufixovým polem**

Sufixový strom je datová struktura, která umožňuje efektivní řešení řady vyhledávacích problémů, její nevýhodnou je ovšem značná prostorová náročnost. Sufixové pole je naproti tomu prostorově úsporná datová struktura, která však nemá tak široké aplikace. Jistý kompromis mezi oběma strukturami představuje sufixové pole, vybavené dodatečnými informacemi. Předložená práce zkoumá, nakolik jsou algoritmy pro sufixový strom přenositelné na tuto úspornější datovou strukturu. Autor prostudoval známé aplikace sufixových stromů, zvolil několik typických reprezentantů, a ty se pokusil převést nad rozšířené sufixové pole bez výrazného zvýšení časové složitosti. Součástí práce je též implementace zvolených algoritmů nad oběma datovými strukturami a experimentální porovnání časové i prostorové náročnosti, jak na standardním datovém korpusu, tak na náhodně generovaných datech.

Na textu práce oceňuji kultivovaný výklad a srozumitelné vyjadřování, velmi dobrou formální úroveň ruší jen poněkud nesourodé obrázky (str. 9, 15, 37, 50). Úvodní kapitoly práce jsou věnovány zavedení nezbytných pojmů, které autor ilustruje na příkladech. Definice některých datových struktur (sufixový strom, str. 9) by mohla být i o něco formálnější, pro pochopení dalšího výkladu je však zvolený popis dostačující. Následuje výklad algoritmů, převzatých z doporučené literatury, které popisují simulaci průchodu sufixovým stromem pomocí rozšíření sufixového pole. Po stručném popisu implementace datové struktury, testovacích dat a podmínek, za nichž byly prováděny experimenty, se dostáváme ke kapitole 6, která tvoří jádro práce. Autor zde vybral pět reprezentativních vyhledávacích problémů, pro něž je popsáno efektivní řešení pomocí sufixového stromu, a navrhuje algoritmy jejich řešení, postavené na rozšířeném sufixovém poli. Při tom částečně vychází z doporučené literatury (Aboulheda et al.), přidává ale i svůj vlastní malý přínos. U každého problému je kromě implementačních poznámek též podrobný popis výsledků experimentů na reálných i náhodných datech.

Nejzajímavější částí práce jsou zřejmě výsledky experimentů a jejich interpretace. Autor zde musel vyřešit problém, jaké parametry vstupních dat vlastně sledovat: pro každý typ problému totiž přichází v úvahu minimálně délka vstupního textu a velikost abecedy, a dále parametry specifické pro řešenou úlohu. Slovní vyhodnocení ilustrují grafy, zaměřené hlavně na závislost délky výpočtu na velikosti abecedy. Tabulky s podrobnými výsledky naměřených parametrů, která zahrnují prostorovou náročnost implementovaných datových struktur, jsou uvedeny v příloze textu. Elektronická příloha práce obsahuje kromě zdrojových kódů v jazyce C s dokumentací též použitá testovací data a výsledky měření.

Dle mého soudu jde o zdařilou práci, která poskytuje cenné svědectví o efektivitě rozšířeného sufixového pole. Její autor zde prokázal schopnost zvládnutí všech podstatných fází vlastního výzkumu: studium relevantních výsledků z literatury, jistou invenci pro návrh vlastního řešení, ale též schopnost navržený postup implementovat a experimentálně vyhodnotit. Práce zahrnuje i malý vlastní teoretický přínos při přenosu algoritmů pro sufixový strom na rozšířené pole, a poměrně zajímavou interpretaci výsledků experimentů, které by mohly být přínosné i pro širší publikum, možná by se tedy dalo uvažovat i o jejich publikaci.

Pokud jde o celkové hodnocení práce, domnívám se, že její obsah splňuje všechny požadavky, uvedené v pokynech pro vypracování tohoto tématu, a proto mohu doporučit, aby byla přijata jako práce diplomová.

V Praze dne 15. května 2009

Tomáš Dvořák