

# Posudek oponenta diplomové práce

Autor/autorka práce: Bc. Ondřej Kaas

Název práce: Programové vybavení pro analýzu geografických dat založenou na shlukování

Práce se věnuje velmi aktuálnímu tématu zpracování velkých sad prostorových dat. Úvod práce je nicméně velmi obecný a postrádá jakoukoliv zmínku o motivaci ke zpracování tématu práce. Teoretická část práce (kapitola 2) je přiměřeně obsáhlá a shrnuje základní poznatky využívané v dalších částech práce. Kapitola 3 *Definice úkolu a návrh řešení* popisuje navržené úpravy stávající knihovny pro shlukování, ale není z ní zřejmá slibovaná definice úkolu. Následující kapitola podrobně popisuje technické řešení shlukování včetně navržených úprav. Diskuse k práci je obsažena v kapitole 5 *Experimenty a výsledky*. Zde jsou popsány potíže způsobené charakterem vstupních dat a výsledky experimentů, které diplomant podnikl pro zlepšení výsledků. Pro možnost zopakovat provedené experimenty by bylo vhodné uvést hodnoty všech použitých parametrů shlukování (např. váhy souřadnic). Hodnocení výsledků shlukování by kromě potřebného času a počtu bodů v jednotlivých úrovních mělo obsahovat také objektivní hodnocení výsledků shlukování (například pomocí celkové ceny shlukování). Z provedených modifikací stávajících knihoven a provedených experimentů je zřejmé, že vytvořené programové vybavení zohledňuje přinejmenším některé požadavky uživatelů zpracovávajících data z laserového skenování.

Použitá literatura je v práci zpravidla vhodně citována. Výjimkou je například kapitola 2.4.7 *Formát LAS*, kde zcela chybí použitý zdroj. Vyhledávání v seznamu literatury je pro čtenáře obtížné. Bývá zvykem seznam literatury seřadit. Pro diplomovou práci by jistě bylo vhodnější používat hodnotnější literaturu než například texty k přednáškám, obzvláště pokud se jedná o obecná témata typu metrický prostor.

Grafická úroveň práce je velmi dobrá. Text je vhodně členěn a doplněn obrázky. V celém textu je ale bohužel mnoho překlepů a chyb (např. na str. 20 „rozvodní sítě“, str. 21 „standart“, str. 33 a 34 „čtyřúhelník“, strana 37 „indexi“). Odkazy na obrázky uprostřed věty by neměly začínat velkým „O“.

Na mnoha místech práce lze nalézt nevhodné, nepřesné či jazykově nesprávné formulace, například:

- Str. 22 uprostřed – „Zpřesňující lokalizace je proces hledání identických bodů to znamená nalezení bodů v mračně korespondující s polohopisnými body.“
- Str. 22 dole – „Díky této enormní velikosti se mračna rozdělí na menší krychle, které se samostatně lokalizují.“ (Mračna se rozdělují nikoliv díky jejich velikosti, ale kvůli jejich velikosti.)
- Str. 22 dole – „Pro zachování přesnosti se požaduje na 100 m dat z laserového skenování minimálně 1 identický bod.“ (Je zvláštní měřit data na metry.)
- Str. 23 uprostřed – „Jednotlivé soubory obsahují řádově menší počet bodů, a tedy i několik úrovní detailu vstupních dat.“ (Jednotlivé soubory obsahují pouze jednu úroveň detailu. Řádově menší počet bodů je v souboru s následující vyšší úrovní detailu.)

- Str. 25 nahoře – Uvedené intervaly nemají jednotky, v číslech jsou použity desetinné tečky. Navíc má autor pravděpodobně na mysli nikoliv rozdíly, ale jejich absolutní hodnoty.

Z textu práce bohužel není zcela zřejmé, kam směřuje, tzn. čím se vyvíjený software odlišuje od stávajícího, v čem by měl být přínos práce. To by mělo být uvedeno v úvodu práce a/nebo podrobněji na začátku kapitoly 3 *Definice úkolu a návrh řešení*. V práci například není (kromě jediné věty v Závěru) vysvětlení, k čemu je dobré shlukování, které zvláštním způsobem nakládá s tzv. polohopisnými body. Text práce bez jakéhokoli vysvětlení zcela opomíjí 1. a 3. zásadu pro vypracování obsaženou v zadání práce, a to přesto, že diplomant svoji práci s uživateli konzultoval.

#### Doplňující informace k práci

Přirozeným požadavkem uživatelů pravděpodobně bude srovnatelnost výsledků shlukování bez ohledu na rozdělení dat do souborů. Výpočty s normalizovanými souřadnicemi však mohou vést, v závislosti na rozsahu souřadnic bodů v jednom souboru, ke značně rozdílným výsledkům. K nepřirozeným výsledkům může vést také normalizace v případě bytí jen jednoho souboru, který obsahuje body kolem linie rovnoběžné s jednou ze souřadnicových os. V tomto případě je sice možné nepřirozenost odstranit vhodnou volbou vah, topu se ale pouze napravuje „deformace“ způsobená normalizací. Navíc tento postup předpokládá velmi poučeného uživatele.

#### Dotazy k práci

- 1) Vztah (3.1) na str. 25 zřejmě nesplňuje trojúhelníkovou nerovnost a tedy není metrikou? Jaký to může mít vliv na funkci použitého algoritmu, který pravděpodobně použití metriky předpokládá?
- 2) Proč byla pro shlukování s polohopisnými body (kapitola 5.2) použita pouze malá část bodů? Můžete nějak porovnat výsledky shlukování bez polohopisných bodů a s nimi?
- 3) Nebylo by jednodušší řešit zobrazení naskenovaných bodů v okolí identických bodů nikoliv jako součást shlukování, ale až v prohlížeči „rozbalením“ několika nejbližších shluků v okolí identických bodů?
- 4) Na str. 25 nahoře je uvedeno: „Větší rozdíl v z-souřadnici polohopisných bodů lze odstranit normalizací souřadnic.“ Jak normalizace souřadnic pomůže?

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 5. 6. 2015



Ing. Radek Fiala, Ph.D.