

Posudek diplomové práce

předložené na Matematicko-fyzikální fakultě

Univerzity Karlovy

Posudek oponenta

Autor:	Bc. Ladislav Láska
Název práce:	Scalable link-time optimization
Stud. program a obor:	informatika, diskrétní modely a algoritmy
Rok odevzdání:	2017
Jméno a tituly oponenta:	Mgr. Martin Mareš, Ph.D.
Pracoviště:	Katedra aplikované matematiky
Kontaktní e-mail:	mares@kam.mff.cuni.cz

Předložená práce se věnuje globálním optimalizacím v překladači, tedy takovým, které probíhají až v čase linkování programu. Obvyklou překážkou u nich bývá přílišná výpočetní náročnost algoritmů používaných pro analýzu a transformaci kódu, kvůli čemuž se některé z nich buďto vůbec nepoužívají na interprocedurální úrovni, nebo se spouštějí jen na menší části programu. Cílem práce je prozkoumat chování těchto algoritmů v překladači GCC a některé z nich zefektivnit.

První část práce představuje čtenáři problém globálních optimalizací a na základě profilování překladu několika rozsáhlých programů (například webový prohlížeč Firefox) identifikuje výkonově problematická místa. Nejvýraznějším z nich je points-to analýza (jeden z druhů alias analýzy), konkrétně její práce s množinami reprezentovanými pomocí bitmap.

Dále autor popisuje různé přístupy k alias analýze a points-to analýze a formuluje jejich požadavky na množinové datové struktury. Poté hodnotí různé reprezentace množin a dochází k tomu, že většinu požadavků splňují nejlépe pravděpodobnostní Bloomovy filtry. Těm ovšem chybí potřebná schopnost enumerace prvků množiny. Navrhuje proto novou datovou strukturu, od Bloomových filtrů odvozenou, která enumeraci podporuje.

Závěr práce je věnován modifikacím GCC, aby tuto datovou strukturu používalo, a jejich experimentálnímu hodnocení.

Práce je psána příjemnou, čtivou angličtinou, kterou trochu kazí občasné překlepy. Použité zdroje jsou korektně citovány.

Úvodní část je zpracována podrobně a činí práci srozumitelnou i pro čtenáře, jenž není odborníkem na překladače. Samotný popis datové struktury je bohužel velmi stručný a obtížně srozumitelný – jakým způsobem funguje side index a rozdělení hodnot na offset, hash a index, je nutné luštit z pseudokódu operací, jehož různé části navíc používají nekompatibilní značení.

Další slabinou práce je hodnocení navrženého řešení: V práci chybí jakýkoliv teoretický rozbor; je ovšem pravděpodobné, že by se ukázalo, že specifický případ použití struktury v GCC se značně liší od vypočteného průměrného chování. To by bylo alespoň vhodné zmínit. Experimentální hodnocení je postaveno na jediném pokusu s překladem Firefoxu, ačkoliv úvodní část zmiňovala několik různých projektů vhodných k benchmarkování překladače.

Práce se sice pokouší kvantifikovat vliv chybovosti datové struktury (způsobené randomizací) na přesnost reprezentace množin, ale mnohem zajímavější by bylo změřit, nakolik tyto nepřesnosti mají vliv na výkon přeložených programů.

Na druhou stranu oceňuji, že se autorovi povedlo analyzovat netriviální část produkčního překladače a výrazně vylepšit její výkon.

Práci proto doporučuji přijmout jako diplomovou.

V Praze dne 25. ledna 2017
Martin Mareš