

Posudek oponenta bakalářské práce

Autor: **Vít Kovarovič**

Název práce: **Návrh a implementace open source knihovny pro kompresi molekulárních trajektorií**

Cílem práce bylo vytvořit open-source knihovnu implementující metodu pro kompresi dat pocházejících z molekulární dynamiky (PMC - Predictive Molecule Compression). PMC je výsledek výzkumu na Katedře informatiky a výpočetní techniky ZČU. Data z molekulární dynamiky popisují pozice atomů molekuly v časových snímcích. Metoda PMC dokáže významně redukovat počet bitů potřebných na zakódování pozic jednotlivých atomů díky využití informace o vazbách v molekule. Jedná se o ztrátovou kompresi s parametrem pro kontrolu maximální povolené odchylky souřadnic. Nedostatkem PMC byla prototypová implementace v jazyce C#, bránící širší použitelnosti metody. Předložená práce metodu implementuje do jazyka C++ a řeší použitelnost i v jazyce Python.

Předložená kvalifikační práce je vhodně organizována do kapitol. Po krátkém úvodu do problematiky a předstření obsahu autor analyzuje možnosti implementace programových knihoven vzhledem k multiplatformnosti, integraci, použitelnosti i z jiných jazyků než z C++ (pilotně Python), závislostí na knihovnách třetích stran, a licencím. Následuje kapitola s úvodem do molekulární dynamiky a shrnutím principů fungování kompresní metody PMC. Autor neopomněl analyzovat nedostatky prototypové implementace. Následuje návrh knihovny s popisem použitých technologií a návrhem rozhraní. Rozhraní je moudře rozděleno na moduly dvou kategorií - jedna pro uživatele a druhá pro vlastní fungování kompresní metody. Další část práce vhodně popisuje některé aspekty implementace (od Huffmanova kódování přes práci s bity až po formát souboru, návod k sestavení, jak se bude testovat, tvorbu dokumentace). Následuje validace kompresních výsledků formou srovnání oproti původní implementaci na reálných datech (konzistence komprimovaných a dekomprimovaných dat, rychlost běhu, a spotřeba paměti) a závěrečné shrnutí dosažených výsledků.

Práce má celkem 39 stran, z toho přibližně 30 stran původního textu, odpovídá doporučení na minimální rozsah. Knihovna byla navržena a implementována z pohledu oponenta kvalitně a je pečlivě dokumentována. Po formální úrovni je práce v pořádku, text je vhodně doplněn ilustracemi. Autor poctivě cituje 55 zdrojů včetně vědeckých článků. Všechny body zadání byly splněny.

Dotazy k práci:

1) Zdá se, že byla měřena pouze rychlost komprese, nikoliv dekomprese, a měření zahrnuje i režii práce se soubory. Jak je na tom z hlediska času rychlost komprese a dekomprese bez zahrnutí času práce se soubory? Například pokud bychom měli data (řádkově stovky MB až jednotky GB) uložená pouze v RAM. Jaká je spotřeba paměti při dekompresi?

2) Jak byla měřena paměťová náročnost staré a nové implementace (metodika měření)? Lze vyloučit, že porovnání nebylo významně ovlivněno zahrnutím už "dávno mrtvých" objektů, o které se ještě nepostaral garbage collector?

3) Zjevným nedostatkem je nutnost mít při kompresi i dekompresi všechna data v hlavní paměti. Jak by to šlo budoucna vyřešit, a je na to připravené API?

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

V Plzni 30.5.2023

Mgr. Martin Maňák, Ph.D.