

Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: David Hoksza

Jméno a příjmení autora práce: Jakub Velkoborský

Název práce: Hierarchical visualization of the chemical space

Předložená diplomová práce se zaměřuje na problém vizualizace chemického prostoru, resp. molekul v tomto prostoru a na rozdíl od většiny existujících přístupů cílí na hierarchickou vizualizaci za pomoci konceptu tzv. molekulárních scaffoldů, které lze zjednodušeně popsat jako postupné kontrakce hran molekulárního grafu. Jedná se o silně mezioborovou práci, kdy student musel nastudovat problematiku definice, reprezentace a vizualizace molekul a scaffoldů (chemie) a následně navrhnout a implementovat přístup k vizualizaci velkých souborů molekulárních dat v tomto prostoru ve formě funkční aplikace (informatika).

Po seznámení se s problémem vizualizace molekul a pochopení motivace (kapitola *Introduction*) bylo třeba analyzovat problematiku reprezentace molekul a popis scaffoldů, manipulaci s nimi a možnosti různých přístupů k vizualizaci (kapitola *Background and Terminology*). Na základě této analýzy bylo rozhodnuto o vizualizaci datových sad pomocí scaffoldů molekul, které tvoří intuitivní abstrakci konkrétních molekul, na pozadí chemického prostoru definovaného databází Pubchem, největší volně přístupnou databází malých sloučenin. Tato data tvořila pak pozadí na základě kterého můžou být nová vizualizována formou treemapy (kapitola *Design Choices*). Aby toto bylo možné bylo třeba celou databázi zpracovat (kapitola *Data Extraction and Processing*), tj. stáhnout všechny molekuly, vytvořit z nich scaffoldy všech úrovní definovaných v předchozím kroku a tyto efektivně reprezentovat, aby tato informace byla použitelná následně v rámci interaktivní vizualizace. Tento úkol byl značně netriviální vzhledem k faktu, že Pubchem obsahuje 91 milionů sloučenin obsahujících různá nevyčištěná data a typy molekul, které bylo třeba vzít v úvahu. Výsledkem byly scaffoldové statistiky Pubchemu. Následuje kapitola *Implementation* popisující výsledné softwarové řešení, konkrétně klient-server aplikaci, která umožňuje vizualizaci vlastních datových sad na pozadí Pubchemu. To není jednoduchý problém, protože se jedná o interaktivní webovou aplikaci, kdy práce s velkými daty je netriviální. Nicméně výsledná aplikace umožňuje vizualizovat datové sady o desítkách tisíců sloučenin. To je ztíženo navíc tím, že na straně serveru je třeba pracovat s hierarchií vytvořenou v předchozím kroku, která sama o sobě má velikost několika GB. Kapitola *Results* pak popisuje scaffoldovou hierarchii získanou analýzou Pubchemu a vizualizační aplikaci jako takovou.

Celkově práce obsahuje dva hlavní výsledky: scaffoldovou analýzu Pubchemu a nový vizualizační přístup implementovaný v plně funkční aplikaci. Je třeba zdůraznit, že samotná scaffoldová analýza Pubchemu je relativně komplikovaný problém a oba výsledky mají silný publikační potenciál. Scaffoldový analyzátor je navíc napsaný jako samostatná aplikace, kterou lze aplikovat na jiné datové sady, případně i s jinak definovanou scaffoldovou hierarchií.

V současné době byla aplikace veřejně zpřístupněna a poskytnuta k testování skupině v Ústavu organické chemie a biochemie AV ČR.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input checked="" type="checkbox"/>
---	---

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

V Praze dne: 12. 8. 2016

Podpis: