

Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: David Hoksza

Jméno a příjmení autora práce: Petr Škoda

Název práce: Návrh efektivní generické molekulární reprezentace

Předložená multidisciplinární práce z oblasti chemické informatiky se zabývá vývojem nové metody pro určení podobnosti chemických sloučenin. Na rozdíl od stávajících metod je navrhovaná metoda generická v tom smyslu, že je parametrizovatelná prakticky libovolnými fyzikálně-chemickými deskriptory.

Největší problém práce je jazyková kvalita. Jedná se o text v anglickém jazyce a je znát, že toto není autorův rodný jazyk. Na mnoha místech je špatně shoda přísudku s podmětem, překlepy, špatně používané fráze nebo slova. Na druhou stranu je třeba oddělit kvalitu jazyka a kvalitu obsahu. Zde si myslím, že předložený text přesahuje standard diplomových prací především v kapitolách 1-4. Tyto hrají roli rešerše oblasti chemické informatiky se zaměřením na tzv. virtuální screening, který je hlavní oblastí uplatnění implementované metody. Sekce 5 popisuje standardizovaný benchmark, který se používá pro vyhodnocení kvality podobnostních měř v chemické informatice. Součástí kapitoly je i přehled možných způsobů vyhodnocení kvality měř s analýzou jejich kladů a záporů. Následující dvě sekce pak popisují implementovanou metodu z pohledu algoritmického (kapitola 6) a z pohledu implementačního (kapitola 7). Experimentální vyhodnocení je pak prezentováno v kapitole 8. Zde se autor zaměřil jak na předvedení vlivu různých parametrizací, tak na srovnání se stávajícími metodami. Srovnání se stávajícími metodami mohlo být dle mého názoru podrobnější. Navíc autor srovnává pouze se 2 nejlépe skórujícími metodami, které aktuálně existují, a výsledek pak tolik "nevyznívá". Nicméně práce, na které své experimenty zakládá, obsahuje srovnání více než deseti metod, které pak všechny skórují na daných datových sadách hůře než předkládaná metoda. Implementovaná podobnostní metoda je generická, umožňující vkládat jako parametr prakticky libovolné deskriptory, nicméně výběr těchto deskriptorů již nebyl obsahem této práce. Experimentální sekce obsahuje několik typických deskriptorů, které tento fakt pouze podtrhují. Některé kombinace deskriptorů nejsou příliš vhodné, protože např. dobře nepopisují variabilitu v konkrétní datové sadě.

Rozhodnutí psát text v anglickém jazyce má myslím dopad i na vnímání celé práce, nicméně samotná vyvinutá metoda je velice kvalitní s potenciálem do budoucna a už nyní je srovnatelná s nejlepšími metodami v oblasti chemické informatiky, konkrétně virtuálního screeningu.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

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

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

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. 5. 2014

Podpis: