

Review of Master's Thesis

Student: Fülöp Tibor, Ing.

Title: Computational Methods for Annotation Analysis of Genetic Variations (id 17653)

Reviewer: Bendl Jaroslav, Ing., UIFS FIT VUT

1. **Assignment complexity** **more demanding assignment**
Zadání považuji za studijně i implementačně náročné. Řešené téma je značně vzdálené od jakékoliv látky probírané během studia a návrh vyžadoval nemalé úsilí. Implementační náročnost je ukryta v obtížnosti manipulace s velkými daty.
2. **Completeness of assignment requirements** **assignment fulfilled**
Zadání bylo splněno v celém rozsahu.
3. **Length of technical report** **in usual extent**
Rozsah práce je v obvyklém rozmezí.
4. **Presentation level of technical report** **90 p. (A)**
Technická zpráva má logickou strukturu a poměr teoretického úvodu, rozboru zadání a prezentace vlastního řešení je vyvážený. Jedinou připomínka ke struktuře směřuji ke kapitole 5, ve které se prolíná popis návrhu s implementací, přičemž bych považoval za logičtější tyto části vzájemně oddělit.
5. **Formal aspects of technical report** **85 p. (B)**
Technická zpráva je psaná v anglickém jazyce. Text sice obsahuje řadu drobných chyb (občasné chybné použití členů či nesprávná větná stavba), ale ty nejsou na závadu srozumitelnosti textu. V kapitole 4 nejsou některé rovnice a proměnné dostatečně popsány.
6. **Literature usage** **85 p. (B)**
Výběr zdrojů je k řešenému tématu relevantní a publikace jsou z textu vhodně odkazovány. Čerpáno je především z velmi kvalitních časopiseckých publikací. V případě dvou odborných článků nemá citační záznam uvedené jméno časopisu. Převzaté obrázky mohly mít často lepší zdroj než [online / wiki], zvláště když některé z citovaných článků z časopisů dané ilustrace mají.
7. **Implementation results** **95 p. (A)**
Realizační výstup je dvojitý: jedná o systém pro import dat z integrovaných zdrojů do databází vyvinutého nástroje a dále pak o nástroj samotný. Ten má podobu webové služby implementované prostřednictvím frameworku Shiny jazyka R. Nástroj jsem otestoval a shledal plně funkční. Celkový rozsah zdrojového kódu je přibližně 2000 řádků v 20 souborech. Funkcionalita jednotlivých souborů je dobře komentována v textu diplomky, huře pak na úrovni samotného kódu.
8. **Utilizability of results**
Vyvinutý nástroj najde široké uplatnění v cílové komunitě a předpokládám, že bude i publikován v časopise s impakt faktorem. **Jelikož nástroj vznikl jako školní dílo, fakulta by si měla pohlídat uvedení afiliace na případném článku a zajistit i vykazání do RIV jako software.**
9. **Questions for defence**
 - V práci je uvedena teorie k více metodám využitelným pro analýzu obohacení. Jaké důvody vedly k výběru hypergeometrického testu?
 - Ačkoliv počet anotací je relativně velký (38), chybí v nich výsledky prediktorů pracujících nad nekódujícími mutacemi. Tyto by bylo možné získat z Annovaru, který byl rovněž popsán v teoretické části. Proč byla dána přednost platformě Get-Evidence?
10. **Total assessment** **92 p. excellent (A)**
Předložená diplomová práce je inovativní, z vědeckého pohledu zajímavá a připravená k publikaci. Zároveň je dotažená do podoby webové služby s přehledným rozhraním umožňujícím využití širokou vědeckou komunitou. Oceňuji i kvality anglicky psané technické zprávy, ve které se však objevilo menší množství nepřesností a typografických chyb.

In Brno 5. June 2015

.....
signature

