

I. IDENTIFIKAČNÍ ÚDAJE

| | |
|-----------------------------------|---|
| Název práce: | Nástroj pro objevování sekvenčních motivů v RNA-Seq datech. |
| Jméno autora: | Petr Schimperk |
| Typ práce: | díplomová |
| Fakulta/ústav: | Fakulta elektrotechnická (FEL) |
| Katedra/ústav: | Katedra počítačů |
| Oponent práce: | Doc. Ing. Jiří Kléma, PhD. |
| Pracoviště oponenta práce: | Katedra počítačů FEL ČVUT |

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

| | |
|--|-------------------------|
| Zadání | průměrně náročné |
| <i>Hodnocení náročnosti zadání závěrečné práce.</i> | |
| Zadání předpokládá vytvoření dobře ovladatelného nástroje pro objevování sekvenčních motivů v RNA-Seq datech. Jde především o efektivní současné využití existujících detektorů sekvenčních motivů a přehlednou kompilaci jejich výstupů. Současně práce směřuje na RNA-Seq data. Vzhledem ke klasické úloze vyhledávání motivů v DNA sekvencích je zde třeba silněji zohlednit otázku efektivity spojenou s potenciálně velkým počtem čtení a jejich chybovostí. Při zhodnocení významu motivů hraje roli také četnost transkriptů. Motivy nalezené v čtených transkriptech jsou pro uživatele významnější. | |

| | |
|---|------------------------------------|
| Splnění zadání | splněno s menšími výhradami |
| <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i> | |
| Splnění zadání není snadné přesně hodnotit. Dobře se podařilo naplnit uživatelské hledisko. Vytvořený nástroj je přímočaře použitelný biologem, kompilace ve formě sloučených motivů zadání vyhovuje. Problematiku RNA-Seq dat ale autor v práci reflektuje jen velmi zjednodušeně a málo přehledně. Odráží se především v diskusi SAM/BAM vstupního formátu a zavedení vlastního rc-skóre, které původní skóre motivu násobí počtem čtení. Otázka efektivity není diskutována, v práci prezentovaný test na RNA-Seq data necílí, jeden z praktických experimentů ale s BAM formátem pracuje. | |

| | |
|---|------------------------|
| Zvolený postup řešení | částečně vhodný |
| <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i> | |
| Jde o bioinformatickou práci se silným důrazem na software. Softwarové řešení je v mnoha ohledech robustní a integruje všechny základní kroky pro práci s motivy. Autor ale minimálně v textu rezignuje na vědecké a analytické hledisko. Z pohledu prezentace mohla práce lépe pracovat s literaturou, přehledně popsat stávající přístupy k extrakci motivů z RNA-Seq dat a nabídnout srovnání s implementovaným postupem. Přímo z textu práce se čtenář a tím ani uživatel přímo nedozví, že hlavní pracovní tok pro RNA-Seq data počítá se SAM/BAM vstupem, extrahuje z něj konsenzus sekvencí s nenulovým počtem přiřazených čtení a v této sekvenci vyhledá příslušné motivy. Motivy pak lze prioritizovat dle výše uvedeného rc-skóre. Pro tento druh sekundárního vstupu ale nejsou vhodné všechny nástroje pro vyhledání motivů. | |

| | |
|---|------------------|
| Odborná úroveň | C - dobře |
| <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i> | |
| Práce má dobrou úroveň z pohledu softwarového. Z pohledu bioinformatického se jí daří integrovat velký počet relevantních externích nástrojů pro vyhledávání, prezentaci, porovnávání a slučování motivů. V souladu se zadáním umožňuje různé formáty vstupů včetně importu sekvenčních dat na základě identifikátorů transkriptů či genů. Kroky dotýkající se RNA-Seq dat ale autor neřeší dostatečně transparentně a pravděpodobně ani metodicky. | |

Formální a jazyková úroveň, rozsah práce

D - uspokojivě

Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku.

Práce obsahuje základní úvod do teorie sekvenčních motivů, zbytek textu je ale převážně dokumentací k vytvořené aplikaci a jejím praktickým srovnáním s paralelními frameworky. To by bylo zcela v pořádku, kdyby zadání nespecifikovalo konkrétní kroky, na které se má diplomant v pracovním toku zaměřit. Zde bych očekával, že student bude tyto kroky jasněji specifikovat, odkazovat na stávající přístupy k jejich vyřešení v dostupné literatuře, srovnávat je a řádně vyhodnocovat. Pokud nebyla data pro komparativní vyhodnocení implementace k dispozici, mělo by to být v práci zmíněno. Jazykově je text v pořádku.

Výběr zdrojů, korektnost citací

C - dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posudte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Diplomant pracuje s 28 referencemi, což odpovídá charakteru textu uvedeného v odstavci výše. Navrhoval bych aktivnější práci s literaturou, relevantních zdrojů je zcela určitě mnohem více.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Další komentář nemám.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Diplomová práce plní základní cíl zadání, implementuje a popisuje nástroj pro detekci sekvenčních motivů použitelný molekulárním biologem. Nástroj splňuje podle mého názoru všechny základní funkční požadavky. Cíle směřující na specifická RNA-Seq data jsou adresovány spíše okrajově. Nástroj vyhledání motivů v RNA-Seq datech umožňuje, nenabízí ale jeho hlubší diskusi. Jako součást diskuse u obhajoby proto navrhuji krátké demo, kde by diplomant demonstroval test splňující poslední bod zadání, tj. demonstrovat efektivitu a přehlednost vyhledání motivů v RNA-Seq datech.

K obhajobě mám dvě otázky:

1) Jaká je časová náročnost obvyklých praktických úloh vyhledání sekvenčních motivů? Jak Váš nástroj podporuje uživatele v návrhu pracovního toku proveditelného v očekávaném čase? Mám na mysli především velikost vstupních dat, resp. volbu škály použitých detekčních a sumarizačních nástrojů.

2) Které typové úlohy nebyly řešitelné nástrojem Jana Holčáka a jsou řešitelné pomocí Vašeho Programmable Motif Discovery Toolu?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm C - dobře.



POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

Datum: 31/05/2024

Podpis: