

# Posudek oponenta diplomové práce

Autor/autorka práce: **Michal Kasal**

Název práce: **Experimentální zpracování velkých biologických dat**

## Obsah práce

Práce nejprve představuje koncept velkých dat, poté představuje nejvýznamnější biologické databáze a metody pro zpracování dat. Na klasifikaci variací části lidského genomu potom zkouší centralizovaný distribuovaný a kombinovaný způsob klasifikace. Nakonec dává doporučení jak upravit systém EEG/ERP portál, tak aby mohl implementovat podobné zpracování dat.

## Kvalita řešení a dosažených výsledků

Práce obsahuje kvalitní analytickou část, získané poznatky jsou poté aplikovány při výběru metod pro analýzu variant lidského genomu. Výsledky jsou prezentovány přehledně a zvolená metodika odpovídá zadání.

## Formální úroveň

Práce svým obsahem, více jak 80 stran, přesahuje standardně vyžadovaný rozsah, je psána na dobré úrovni, pouze s malým množstvím chyb a překlepů.

## Práce s literaturou

Autor vhodně pracuje s literaturou. Práce obsahuje téměř 90 referencí, vesměs v dobré kvalitě.

## Splnění zadání

Zadání je splněno ve všech bodech. Oproti zadání je navíc uvedeno doporučení, jak využít výsledky v EEG/ERP Portálu.

## Doplňující informace k práci

Ačkoliv je metodika vhodně zvolená a dobře popsána s ohledem na zadání je poměrně málo prostoru věnováno čistě distribuovanému zpracování, které je v práci prezentováno pouze na jednom převzatém experimentu. Není mi jasné jestli ho autor opravdu prováděl, nebo pouze převzal i s výsledky. Vzhledem k tématu velkých dat by mělo být distribuované prostředí nejvíce probrané a odzkoušené. Práci by také slušelo podrobněji se věnovat popisu metodiky s přesným popisem, jak experimenty zopakovat. Např. uvést do přílohy postup instalace potřebných nástrojů, připravit spouštěcí skripty apod. Příložené CD obsahuje pouze KNIME workspace , což je však pouze část experimentu. Doporučení pro EEG/ERP portál, zvláště v oblasti binárních dat by chtělo lépe vysvětlit. Kód 10. pouze kopíruje jeden datový proud do druhého, ale není jasné jakou to přinese výhodu z pohledu zpracování dat.

**SOUHLASÍ  
S ORIGINÁLEM**



Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
katedra informatiky a výpočetní techniky

Dotazy k práci

1. Blíže vysvětlete co přesně byla Vaše práce ve třetím experimentu s užitím čistě distribuovaného systému.
2. Vysvětlete, jak byste ukládal data z EEG/ERP portálu v distribuovaném prostředí a jaké výhody by to přinášelo a proč je současné řešení využívající databázi PostgreSQL problematické, jak sám uvádíte.

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 18.5.2016

Ing. Petr Ježek, Ph.D.



SOUHLASÍ  
S ORIGINÁLEM



Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
katedra informatiky a výpočetní techniky  
①