

# Hodnocení vedoucího závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

**Student:** Jakub Koza  
**Vedoucí práce:** RNDr. Petr Škoda, CSc.  
**Název práce:** Design and implementation of a distributed platform for knowledge discovery in big astronomical spectra archives  
**Obor:** Softwarové inženýrství (bakalářský)

**Datum vytvoření:** 3. 6. 2015

<p><i>Hodnotící kritérium:</i></p> <p><b>1. Náročnost a další komentář k zadání</b></p> <p><i>Popis kritéria:</i> Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)</p> <p><i>Komentář:</i> Podstata zadání práce spočívá na první pohled sice jen v integraci již hotových modulů, ale v podstatě jde o vytvoření zcela nového unikátního distribuovaného systému pro strojové učení rozsáhlých astronomických archivů spekter, jehož návrh je podstatně odlišný od původního VO-KOREL, na který navazuje. Pro účely získávání dat musel student vytvořit vedle FTP a HTTP klientů i speciálního klienta pro SSAP a DataLink protokol Virtuální observatoře. Aby mohla být služba integrována do infrastruktury Virtuální observatoře, musela být postavena nad protokolem UWS založeným na návrhovém schématu REST. Během řešení práce bylo třeba vedle čtení rozsáhlých anglických textů nastudovat velmi podrobně i nejnovější vlastnosti J2EE technologie a perfektně zvládnout práci s parsováním XML a JSON formátů, stejně jako s astronomickým FITS formátem. Z těchto důvodů osobně považuju zadání za velmi náročné.</p>	<p><i>Způsob hodnocení - následující škálou 1 až 5:</i></p> <p><b><u>1=mimořádně náročné zadání,</u></b> <b>2=náročnější zadání,</b> <b>3=průměrně náročné zadání,</b> <b>4=lehčí, ale ještě dostatečně náročné zadání,</b> <b>5=nedostatečně náročné zadání</b></p>
<p><i>Hodnotící kritérium:</i></p> <p><b>2. Splnění zadání</b></p> <p><i>Popis kritéria:</i> Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</p> <p><i>Komentář:</i> Realizace zadání předčila mé očekávání. Student vytvořil pružný a snadno škálovatelný systém založený na jednom aplikačním serveru, který udržuje databázi uživatelských experimentů a množství paralelně pracujících workerů na kterých kromě vlastního výpočtu probíhá i vytváření grafických vizualizací výstupu programů, jež jsou pak vtěleny do webové stránky s výsledky výpočtu. Velmi zajímavý koncept je pak univerzální worker, který si sám registruje své schopnosti bez nutnosti připravit specifické prostředí pro každý modul strojového učení. Těžištěm práce je pak objektový souborový systém mapovaný na databázi, kam se ukládají velká data (řádu jednotek miliónů spekter), které může administrovat pouze uživatel se zvláštními manažerskými rolemi.</p>	<p><i>Způsob hodnocení - následující škálou 1 až 4:</i></p> <p><b><u>1=zadání splněno,</u></b> <b>2=zadání splněno s menšími výhradami,</b> <b>3=zadání splněno s většími výhradami,</b> <b>4=zadání nesplněno</b></p>
<p><i>Hodnotící kritérium:</i></p> <p><b>3. Rozsah písemné zprávy</b></p> <p><i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.</p> <p><i>Komentář:</i> Zpráva má požadovaný rozsah – cca 42 stran textu, cca 12 stran příloh a 15 stran tiráže, rejstříků apod. Informačně je celkem bohatá a může posloužit i jako kompaktní úvod do technologie Java EE a Java Server Faces stejně jako do standardů Virtuální observatoře. Je třeba zdůraznit, že práce je psána anglicky, aby posloužila i jako metodický návod ostatním vývojářům.</p>	<p><i>Způsob hodnocení - následující škálou 1 až 4:</i></p> <p><b><u>1=splňuje požadavky,</u></b> <b>2=splňuje požadavky s menšími výhradami,</b> <b>3=splňuje požadavky s většími výhradami,</b> <b>4=nesplňuje požadavky</b></p>
<p><i>Hodnotící kritérium:</i></p> <p><b>4. Věcná a logická úroveň práce</b></p>	<p><i>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</i></p> <p>95 (A)</p>

**Popis kritéria:**

Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.

**Komentář:**

Práce je velmi přehledná a logická. Po krátkém úvodu vysvětlujícím motivace práce poskytuje v první kapitole veškeré podstatné informace o stávajícím systému VO-KOREL/VO-CLOUD a poměrně neznámé problematice Virtuální observatoře, zejména protokolu UWS založeném na návrhovém schématu REST. Krátce je shrnuto v kostce i vše podstatné o použité technologii Java EE a Java Server Faces. V druhé kapitole jsou uvedeny základní požadavky na systém (zejména funkční a nefunkční). Třetí kapitola popisuje vlastní návrh workera s automatickou registrací a čtvrtá pak design a realizace master serveru, zejména jeho úložiště souborů a protokolů pro jejich získání. Pátá kapitola stručně shrnuje plánovaná vylepšení. V přílohách jsou pak vedle slovníčku zkratk představeny i podrobné postupy instalace jak workeru tak serveru. Ačkoli se několikrát objeví zmínka o integraci modulů strojového učení pro náhodné lesy (RDF) a Kohonenovy mapy (SOM), nejsou v práci nikde explicitně ukázány vazby na praktické použití těchto modulů stejně jako podrobněji popsáno jak probíhá předzpracování získaných souborů a jejich transformace do podoby vhodné pro strojové učení. Absence podobných ukázek vazeb je i důvodem ke snížení hodnocení.

Chybí také zmínka o stránce projektu na githubu.

**Hodnotící kritérium:**

*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*

**5. Formální úroveň práce**

97 (A)

**Popis kritéria:**

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 12/2014, článek 3.

**Komentář:**

Práce je velmi příjemně čitelná, doplněná obrázky, tabulkami i UML diagramy a ukázkami kódu, sázeného neproporcionálním písmem. Terminologie je korektně použita a zkratky řádně vysvětleny. Práce je psaná pěknou srozumitelnou angličtinou, působí dojmem profesionálního dokumentu. Velmi oceňuji zařazení stručného návodu na instalaci systému. Práce je psaná pečlivě a s invencí, i když podle mého názoru mohly být uvedeny ukázky použití na příkladech současných modulů strojového učení, i když ty nejsou náplní práce studenta. Nenašel jsem žádné překlepy ani formální nedostatky.

**Hodnotící kritérium:**

*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*

**6. Práce se zdroji**

98 (A)

**Popis kritéria:**

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. 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.

**Komentář:**

Komentář: Všechny použité zdroje jsou řádně citovány a jelikož je většina dostupná jen v elektronické podobě, je správně citován i datum stavu daného linku podle doporučení pro elektronické publikování. Zásadní vliv na úspěch mělo nekolikeré osobní setkání s autorem VO-KORELu a dalšími členy vývojového týmu.

**Hodnotící kritérium:**

*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*

**7. Hodnocení výsledků, publikační výstupy a ocenění**

95 (A)

**Popis kritéria:**

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

**Komentář:**

Všechny použitý software je poskytován v rámci otevřených licencí GNU a jim podobných, stejně tak veškerá studentova práce je k dispozici pod licencí GPL. Dosažené výsledky jsou vynikající. Podařilo se realizovat unikátní řešení distribuovaného systému pro strojové učení astronomických spekter, jež umožňuje získávání dat z mnoha zdrojů i z Virtuální observatoře. Výsledky projektu budou prezentované i s názornou ukázkou funkce na fóru vývojářů Mezinárodní aliance pro Virtuální observatoř (IVOA Interop) v Sextenu v Itálii v červnu 2015 a budou náplní mé prezentace na prestižní konferenci o astronomickém SW - ADASS v Sydney v říjnu 2015. Později by se měl popis serveru stát částí článku v odborném časopise.

**Hodnotící kritérium:**

*Způsob hodnocení - nehodnotí se*

**8. Komentář o využitelnosti výsledků****Popis kritéria:**

Uvedte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uvedte možnosti využití výsledků ZP v praxi.

**Komentář:**

Výsledkem práce je nová mnohem více rozšířená verze distribuovaného systému pro strojové učení velkých astronomických archivů spekter, která bude po odladění nabízena odborné veřejnosti a jež umožní astroinformatické analýzy.

**Hodnotící kritérium:**

*Způsob hodnocení - následující škálou 1 až 5:*

## 9. Aktivita a samostatnost studenta v průběhu řešení

9a:

**1=výborná aktivita,**  
2=velmi dobrá aktivita,  
3=průměrná aktivita,  
4=slabší, ale ještě dostatečná aktivita,  
5=nedostatečná aktivita

9b:

**1=výborná samostatnost,**  
2=velmi dobrá samostatnost,  
3=průměrná samostatnost,  
4=slabší, ale ještě dostatečná samostatnost,  
5=nedostatečná samostatnost

### Popis kritéria:

Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (9a). Posuďte schopnost studenta samostatně tvůrčí práce (9b).

### Komentář:

Komentář: Student pravidelně chodil na konzultace, účastnil se diskusí přes Skype, pružně odpovídal na e-maily a zadané úkoly řešil v rámci možností včas. Aktivně si vyhledával možná řešení problému a v případě nejasností či podezření na chybu kontaktoval autory příslušných řešení. Pravidelně se účastnil schůzek vývojového týmu VO-CLOUDu na Ondřejově i v Praze. Bezprostředně reagoval na odhalené chyby v nově nasazeném systému VO-CLOUD a průběžně systém vylepšoval v souladu se stanovenými cíly. Po odevzdání práce je už např. realizováno poílání spekter protokolem SAMP.

### Hodnotící kritérium:

Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):

## 10. Celkové hodnocení

100 (A)

### Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

### Text hodnocení:

Podle mého názoru je tato práce špičková a unikátní ve světovém měřítku. Student prokázal, že se umí dobře zorientovat v nové problematice, poradit si s nestandardním problémem i dobře se domluvit se SW vývojáři na efektivním řešení. Kromě toho profesionálně zvládal SW vývoj i pomocí nejnovějších VO technologií, spolupracoval i v astronomických otázkách a prokázal velký rozhled i mimo své zaměření. Velmi mě překvapil zejména propracovaným designem a některými chytrými řešeními i poněkud náročnějších problémů. Za toto i perfektní funkčnost systému proto dávám plný počet bodů.

Podpis vedoucího práce: