

## OPONENTNÍ POSUDEK ZÁVĚREČNÉ KVALIFIKAČNÍ PRÁCE

**Autor závěrečné práce:** Bc. Michal Štrick

**Název práce:** Sběr, interpretace a katalogizace volně dostupných online dat

**Oponent práce** Ing. Martin Řimnáč, Ph.D.

**Pracoviště oponenta** Ústav informatiky, AV ČR, v.v.i.

A. Kvalita abstraktu, klíčová slova odpovídají náplni práce . . . . .	Velmi dobře mínus (2-)
B. Rozsah a zpracování rešerše . . . . .	Velmi dobře (2)
C. Řešení práce po teoretické stránce . . . . .	Výborně (1)
D. Vhodnost, přiměřenost použité metodiky . . . . .	Výborně (1)
E. Úroveň zpracování výsledků a diskuse . . . . .	Výborně (1)
F. Vlastní přínos k řešené problematice . . . . .	Výborně mínus (1-)
G. Formulace závěru práce . . . . .	Výborně mínus (1-)
H. Splnění zadání (cílů) práce . . . . .	Splněno
I. Skladba, správnost a úplnost citací literárních údajů . . . . .	Výborně (1)
J. Typografická a jazyková úroveň (vč. pravopisu) . . . . .	Velmi dobře (2)
K. Formální náležitosti práce . . . . . (struktura textu, řazení kapitol, přehlednost ilustrací)	Velmi dobře (2)

Komentáře či připomínky:

Jednotlivé pasáže práce vykazují rozkolísanou úroveň, zejména úvodní pasáže jsou napsány ne zcela srozumitelným způsobem. V rešeršní části jsou některé výroky nepřesné, v kapitole 2.3 Shluková analýza se i vzájemně vylučují (příslušnost objektu do shluku).

V sekcích Technické prostředky a Realizace je uveden duplicitní text (kapitoly 3.2.1 a 4.1.2 a popis htmlUnit), kapitolu 4.2 NoSQL databáze bych osobně uvítal v sekci Technické prostředky, v Realizaci bych případně očekával popis vlastní implementace API pro NoSQL databázi. Podobně popis programu Pajek (kapitola 5.2) bych přesunul do sekce Technické prostředky, neb není vlastní implementací autora, ale pro vizualizaci použitým prostředkem.

Po stránce typografické bych předložené práci vytkl nízkou kvalitou přiložených obrázků.

V práci postrádám diskuzi o konfigurovatelnosti a škálovatelnosti modulů (autor sám uvádí, že při změně rozvržení analyzovaných stránek je nutné částečně změnit plugin, není však patrné, zda do konfigurace nebo do zdrojového kódu). Zatímco pro získání dat z citačních databází je vyžadováno použití funkcionality poskytovatele dat (javascript) vedoucí na specifické způsoby zpracování, nejsou diskutovány ani citovány technologie obecnějších postupů při extrakci dat z webových stránek (popř. jejich škálovatelnost), zde použitelné pro data z fulltextových vyhledávačů.

...pokračuje na straně 2



**Celkové zhodnocení:**

Předložená práce popisuje aktuální problematiku, autor prokázal schopnost samostatné tvůrčí práce; správně pro realizaci zvolil modulární architekturu, která umožňuje případné rozšiřování práce o další funkcionalitu. Po stránce zvolených technologií a implementace nemám výhrad.

Práce má praktické použití zejména při jednorázovém získávání dat (kampaň), u kontinuálního sledování může být její nasazení komplikovanější.

**Otázky k obhajobě:**

1. Rozveďte problematiku použití mezipaměti (cache) pro zpracování výsledků. Jaké jsou výhody a nevýhody jejího použití, případně jak se mohou projevit v kontextu vaší práce.
2. V krátkosti diskutujte vztah funkcionálních a nefunkcionálních požadavků na výsledný software s volbou programovacího jazyka.
3. V krátkosti diskutujte pravidla pro volbu typu úložišť dat, zejména z pohledu architektury (relační databáze, NoSQL databáze, XML, JSON atd.)



**Celková klasifikace:**

Práce splňuje požadavky na udělení akademického titulu, a proto ji doporučuji k obhajobě

Navrhuji tuto práci klasifikovat stupněm **Velmi dobře (2)**

V Praze

dne 03.06.2014

Podpisem současně potvrzuji, že nejsem v žádném osobním vztahu k autorovi práce

.....  
podpis oponenta

