

Posudek oponenta diplomové práce

Student: Kotraš Martin, Bc.

Téma: Detekce vizuálních vzorů ve webových stránkách (id 24460)

Oponent: Bartík Vladimír, Ing., Ph.D., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**

Cílem diplomové práce bylo navrhnout, implementovat a otestovat aplikaci pro extrakci dat z webových stránek pomocí metody využívající detekci vizuálních vzorů. K tomu bylo potřeba nastudovat framework FitLayout, problematiku extrakce dat z webu nebo problematiku ontologií, pomocí kterých jsou zadávány vstupy pro extraktor. Student také musel sám navrhnout některé součásti samotné extrakční metody.
- 2. Splnění požadavků zadání** **zadání splněno**

Všechny body zadání byly splněny.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
- 4. Prezentační úroveň předložené práce** **92 b. (A)**

Technická zpráva jako celek je zpracována pečlivě a detailně. Jak teoretická část, tak i část popisující jednotlivé kroky algoritmu extrakce, jsou napsány srozumitelně, snad jen v některých obtížnějších pasážích mohlo být více praktických příkladů, které by danou část demonstrovaly, celkově však žádné větší výhrady nemám a přínos studenta je zřejmý. Také rozsahy a návaznosti kapitol jsou zvoleny správně.
- 5. Formální úprava technické zprávy** **93 b. (A)**

Formální úprava celé technické zprávy je vysoké úrovně, text je bez jazykových chyb a jen s minimem překlepů, typografická stránka je také bez zjevných nedostatků.
- 6. Práce s literaturou** **90 b. (A)**

K porušení citační etiky nedošlo, práce v mnohém navazuje na práci doc. Burgeta a jiných diplomantů, studentův přínos je však zřejmý a převzaté věci jsou řádně citovány a odkazovány.
- 7. Realizační výstup** **90 b. (A)**

Realizačním výstupem je konzolová aplikace provádějící extrakci dat ze zadané webové stránky. Vstupní informace jsou ve formě ontologie, běžný uživatel by se tedy musel seznámit s touto problematikou. Výsledky extrakce jsou ověřeny celou řadou experimentů, které prokazují její použitelnost a některé horší výsledky jsou zdůvodněny. Jediné, co mi trochu chybí, je srovnání s jiným existujícím řešením (pokud něco srovnatelného existuje).
- 8. Využitelnost výsledků**

Výsledek je zcela jistě přínosem, zejména pro další výzkum v oblasti extrakce informace z webu na UIFS.
- 9. Otázky k obhajobě**
 - Existuje nějaký algoritmus, který by se dal s vaším řešením porovnat? Pokud ano, stručně ho porovnejte s vaším řešením.
 - Bylo by možné nahradit vstup ve formátu ontologie jiným, pro běžného uživatele jednodušším vstupem?
- 10. Souhrnné hodnocení** **91 b. výborně (A)**

Jde o nadprůměrnou diplomovou práci zahrnující novou metodu pro extrakci informace z webu, která byla implementována a experimentálně ověřena. Navrhuji hodnocení stupněm A.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 2. června 2022

Bartík Vladimír, Ing., Ph.D.
oponent