

Hodnocení vedoucího diplomové práce

Autor práce: **Bc. Vítězslav Vacek**

Název práce: **BodyInNumbers a prostředky sémantického webu**

Aktivita studenta, spolupráce s vedoucím

Aktivita studenta nebyla v zimním semestru vůbec dobrá. Neplnil předem dané termíny a jeho práce nebyla úplně samostatná. Potom došlo ke zvratu v letním semestru a student začal přistupovat k práci zodpovědně a pilně. Pravidelně docházel na konzultace, které byly určeny pro vyjasnění dotazů a problémů. Dodržoval většinu daných termínů kontrol přírůstků diplomové práce, ale finální práci mi dal k dispozici pár dní před odevzdáním. Případné nedostatky v diplomové práci svědomitě opravil. Navržené řešení důkladně a svědomitě otestoval. Z výše uvedených důvodů hodnotím aktivitu studenta a spolupráci se mnou známkou **velmi dobře**.

Původnost práce a práce související

Systém, se kterým se měl autor seznámit, implementoval student Pavel Šnejdar v rámci diplomové práce „Vytvoření aplikace pro sběr medicínských dat v rámci projektových dnů“. Komunikace se serverem probíhá pomocí RestAPI, veškerá data jsou posílána a přijímána v datovém formátu JSON nebo jsou dostupná ke stažení v CSV formátu. Výsledné řešení autora obsahuje naměřená data ze systému BodyInNumbers.

Kvalita řešení

Autor zmapoval termíny používané při sběru heterogenních dat v rámci systému BiN. Výstupem je terminologie BiN (138 termů, z toho definováno autorem 88 termínů a 50 termínů převzato z ostatních zdrojů), která je uvedena v příloze E diplomové práce. Značnou výhodou pro něj bylo i zúčastnění se sběru dat na projektových dnech Dny otevřených dveří na FAV a Dny vědy a techniky. Dále prostudoval prostředky sémantického webu a znalosti aplikoval na tvorbu ontologie popisující doménu BiN a na realizaci úložiště trojic. Při tvorbě ontologie autor dodržel pravidla pro znovupoužití již nadefinovaných termínů z jiných publikovaných ontologií. Nakonec provedl publikaci ontologie v systému BioPortal pro sdílení ontologií. Autor implementoval modul v Pythonu do infrastruktury BiN pro konverzi naměřených dat z formátu CSV do RDF formátů (RDF/XML, Turtle, N-Triple). Před realizací úložiště trojic si otestoval několik řešení splňující požadavky pro systém BiN (Jena, GraphDB a Virtuoso) a vybral Virtuoso Open Source Edition. Pro testování byly využity 3 soubory, kde největší obsahoval 7 809 410 trojic. Úložiště trojic nasadil na počítač v rámci infrastruktury BiN a naplnil daty z Dnů vědy a techniky 2017. Nakonec znovu otestoval načítání různých podporovaných RDF formátů (RDF/XML, Turtle, N-Triples a Notation3). Správnost realizovaného řešení byla následně ověřena autorem vytvořenými SPARQL dotazy. Výsledky mi byly prezentovány a jsem s nimi spokojen.

Po formální stránce je kvalita práce vyhovující. Práce je logicky strukturovaná. Práce však obsahuje typografické chyby, gramatické a stylistické chyby, ale vzhledem k rozsahu práce (cca 50 normostran textu práce) jsou v toleranci.

V práci mi chybí konverze dat z formátu JSON do JSON-LD a pro testování jednotlivých úložišť trojic bych navíc zvolil soubor s desítkami miliónů trojic, a proto kvalitu řešení hodnotím známkou **velmi dobře**.

Využitelnost dosažených výsledků

Hlavním cílem této práce bylo popsání domény BodyInNumbers prostředky sémantického webu. Byla vytvořena ontologie, která je dostupná na stránkách BioPortálu na adrese: <https://bioportal.bioontology.org/ontologies/BIN>. Autor dodržuje mezinárodní základní pravidla pro tvorbu ontologie. Výsledek práce je použitelný pro vědecké komunity, které podporují sběr heterogenních dat a prostředky sémantické webu (RDF, OWL).

Splnění zadání

Diplomová práce ve značném rozsahu spočívala v seznámení se s doménou BodyInNumbers, definování terminologie experimentů a navržením ontologie. Dále autor realizoval úložiště trojic, které naplnil daty z projektových Dnů vědy a techniky 2017. Tato část byla pracná a přinesla řešení pro neuroinformatickou skupinu na KIVu. Výsledným řešením je Virtuoso jako nejlepší možnost pro ukládání trojic.

Zadání práce považuji za splněné s výhradami ke kvalitě řešení (viz výše) a počáteční aktivitě studenta.

Otázky k obhajobě


Nejsou.

Závěrečné hodnocení

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

V Plzni 3. 6. 2019


Ing. Petr Brůha


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