



Posudek oponenta závěrečné práce

Student: Martin Melichar
Oponent práce: Mgr. Jan Starý, Ph.D.
Název práce: Systém pro správu elektronických verzí literárních děl
Obor: Web a multimédia

Datum vytvoření: 28. 5. 2018

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání
Popis kritéria: 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.)	
Komentář: Jde o jednoduchou webovou aplikaci pro správu dokumentů.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: 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.	
Komentář: Zadání je splněno, ale dosti primitivním způsobem (podrobně viz níže).	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Rozsah práce je odpovídající. Zbytečná je asi ukázka kódu na konci 3.1.2 (jak podle ID pojmenovat podadresář a zapat do něj *.jpg), ukázka ve 3.2.1 (jak v PHP zavolat new PDO("sqlite")), ukázka ve 3.2.11 (jak pod "public/images" vytvořit podadresář s pětičíselným jménem) a screenshot A.1 (přihlášení: email, heslo).	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	65 (D)
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:

Aplikace má spravovat XML dokumenty, práce tedy začíná analýzou různých XML editorů. Zmiňuje jich pět, ale nezabývá se podporu XML ve standardních editorech (vim, emacs, ...). Z uvedených například EditPadPro údajně "explicitně nepodporuje XML", přestože seznam na TEI, který práce cituje, vyzdvihuje jeho XML content folding. Žádný ze zmíněných editorů se dále nepoužije.

V odstavci 1.2 se tvrdí, že klient-server architektura "chrání data uložením na server, což je jedna z nejlepších metod ochrany".

V přehledu databází (1.2.2) se tvrdí, že u MariaDB je "ochrana dat na špičkové úrovni", aniž by se řeklo, co se tím myslí, a co mají ostatní DB za problém s "ochranou dat". Podobně MySQL "funguje i při výpadku internetu", a má server "oddělený od vývojového prostředí" (?), zatímco PostgreSQL "je vyvíjen pouze komunitou". Ústav pro českou literaturu, která data poskytuje, sám používá MySQL, nicméně zvolena je SQLite, protože pro chod aplikace "nebude potřeba posílání několika dotazů současně".

Odstavec 1.2.3 popisuje elektronické formáty knih, od historických až po současný TEI (Text Encoding Initiative). Ten je "vybrán po diskuzi s pracovníky ÚČL", ale ve skutečnosti se nakonec nepoužije.

V návrhu databáze jsou autoři a vydavatelé jedna a tatáž entita, v téže tabulce authors_publishers. Toto sloučení je naprosto umělé: autor má vždy prázdnou "korporaci", a vydavatel nikdy nemá "pseudonym". Tímto způsobem lze jakékoli dvě struktury prohlásit za totožné. Ve vazbě na konkrétní dílo se pak pochopitelně rozlišuje extra atributem, zda se jedná o autora, nebo nakladatele.

Heslo uživatele se ukládá jako SHA256 hash, ke kterému se přidává ještě salt "pro větší bezpečnost"?

Podle kapitoly 3 má ÚČL "stávající díla v databázi." Co je to za databázi, jaká je její struktura, a jak koresponduje s naší strukturou, se nedozvíme, ani proč pomocí Python scriptů provádíme "import", když databáze již existuje.

Podle 3.2 se do odevzdání "nevyjasnila sada tagů použitých pro formátování textu". Ná základě čeho tedy máme podporovat použitý formát? Podle 3.2.12 je aplikace dokonce "nezávislá na zvoleném formátu". V Závěru se pak píše, že "díla měla různé tagy s různými atributy", nicméně se "úspěšně sjednotila a importovala". Které tagy a atributy to byly, které z nich bylo potřeba "sjednotit" a proč, nebo jak přesně jsou tedy ve stávající databázi sjednoceny, se nedozvíme.

Screenshoty v příloze A zřejmě ukazují jakousi předběžnou verzi: v uživatelském menu je "něco" a "něco" (str. 41), všichni vydavatelé patří ke korporaci "Tiskány a.s." (str. 42). Obrázek A.8 na str. 44 (porovnání skenu a elektronické verze) ukazuje jinou část skenu a jinou v xml textu.

Hodnotící kritérium:

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

5. Formální úroveň práce

70 (C)

Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Komentář:

Překlepů není více než je obvyklé, ani jazykových prohrěšků, i když některé jsou strašlivé ("jsem se s ním setkal").

Hodnotící kritérium:

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

6. Práce se zdroji

70 (C)

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:

Literaturu tvoří vesměs odkazy na použité technologie. K výběru DB lze ale zřejmě použít lepší zdroje než blog, a bez odkazů na it-joker.cz (2009) a stackoverflow (technikálie Pythonu) by se seznam obešel.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

55 (E)

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:

Výsledkem je webová aplikace pro správu XML dokumentů, které tagují staré básnické sbírky. Tato aplikace v zásadě funguje, ale v detailech je mnoho nedostatků.

Hledání v autorech nefunguje ("Borovský" nenajde K.H.Borovského). Fulltextové hledání rovněž nefunguje ("církev" nenajde nic, přestože v Borovského sbírce se vyskytuje). Rok vydání může být v budoucnosti, stránek může být minus pět. Autorovi nelze přímo vytvořit pseudonym - je třeba nejprve vytvořit pseudonym jako samostatného autora. (Přesně s tím měli testovací uživatelé problém, podle 4.1.1). Přílohy k danému dílu (typicky scany stránek) lze měnit pouze na konci. Například nahrát jiný scan strany 7 lze jen tehdy, pokud je to poslední strana (nebo smazat a znovu nahrát všechny následující). "Vícenásobné řazení" (?) ve skutečnosti řadí jen podle jednoho sloupce.

Celá XML editace se po předchozí rešerši formátů a editorů redukuje na textarea v otevřeném okně browseru (!). To lze těžko považovat za jakoukoli "podporu" XML formátu nebo jeho editace. Formát se ani nevaliduje: obsah textového pole se prostě uloží do DB (a počet stran v metadatach se nezmění, i když se změnil). Navíc je taková editace u delších textů (což sbírka básní typicky je) zoufale pomalá: na odsazení řádku lze čekat i jednotky sekund.

Jednotkové testy testují pouze HTML response code a slovo nebo dvě z použité šablony. Samotná funkcionality tím není otestována téměř nijak. Jeden z testů například testuje, zda GET request na "/content: vrací html 200. Zda se z databáze získal seznam děl, zda se správně zobrazil atd, se nijak netestuje: samotné 200 OK je splněním testu. Jinými slovy, testuje se jen to, že stránka existuje. Podobně test loginu zkoumá jen to, že na POST login (email, heslo) přijde odpověď 302 (Found). Zda je uživatel nyní skutečně nalogován, se netestuje nijak. Podobně správný výpis seznamu sbírek mezi lety 1950 a 2000 se testuje tak, že obsahuje slovo "Rukopisy".

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

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

Komentář:

Vytvořená aplikace je použitelná, ale nedovedu si představit, že by ji někdo používal - především díky absenci jakékoli skutečné podpory pro XML formát, ve kterém data jsou, a frustrující pomalosti "editoru" (kterýkoli jiný je lepší).

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

Jak jste z editorů na <https://wiki.tei-c.org/index.php/Editors> vybral pět zmíněných?
Oč přesně je bezpečnější SHA256 se solením oproti SHA256?
Proč je v ukázce A.4 autor (Karel Babánek) vypsán dvakrát?

Hodnotící kritérium:

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

10. Celkové hodnocení

51 (E)

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:

Aplikaci podobného rozsahu a kvality lze s nasazením použitých frameworků vyrobit za dva dny.

Podpis oponenta práce: