



System pro správu publikací

Ondřej Pittl¹

1 Úvod

Související bakalářská práce si kladla za cíl vytvoření webové aplikace sdružující a spravující výsledky publikační činnosti vědců zvolených vědeckých pracovišť². Informace, se kterými aplikace pracuje, jsou získávány externě z bibliografických služeb a jsou aplikací uchovávány v databázi. Společně s vypočtenými statistikami jsou prezentovány uživateli formou seznamu autorů a seznamu jejich publikací. Publikace jsou navíc formátovány dle citační normy ČSN ISO 690:2011 (Firstová 2011) nebo BiBTeX. Obsah obou seznamů je filtrován, řazen a umožněn je i jeho export do souboru ve formátu CSV či TXT.

2 Volba dat a jejich získání

Poskytované informace nebývají napříč bibliografickými službami konzistentní, neboť každá z nich poskytuje jinou kolekci dat. Ta bývá obvykle doprovázena statistickými údaji, zejména počtem citací publikace a metrikami ohodnocujícími míru vědeckého přínosu autora, například h-index, viz Hirsh (2005), g-index, viz Egghe (2006), a i-index, viz Connor (2011).

Data, s nimiž tento systém pro správu publikací pracuje, jsou získávána dolováním webových stránek zvolených bibliografických služeb. Výběr služeb je výsledkem komplexní rešerše jejich sedmi významných zástupců³. Sledovaným faktorem rešerše byla nejen kvalita a kvantita poskytovaných informací, ale také jejich dostupnost a způsob, jakým jsou data službou indexována. Časté jsou dva druhy indexace a to manuální, kterou provádí člověk, a automatická, kterou zajišťuje autonomní algoritmus. Automaticky indexující služby obvykle disponují objemnou databází informací, zatímco služby využívající manuální indexaci se pyšní spíše důvěryhodností a správností poskytovaných informací. Pro další zpracování byly zvoleny služby obojího typu, konkrétně se jedná o automaticky indexující službu GoogleScholar⁴ a manuálně indexující služby ResearchGate⁵ a DBLP⁶.

Všechny zahrnuté služby jsou dolovány vyvinutým obecným parserem, který se řídí popisy jednotlivých služeb uvedených v konfiguračním souboru ve formátu XML. Získaná data z webových stránek daných služeb jsou uchovávána v MySQL databázi, jejíž struktura byla navržena podle informací poskytovaných službami. Tato aplikace indexuje především bibliografické údaje vztahující se k publikaci (například jméno autora či autorů publikace, její název, rok a místo vydání, ISSN a/nebo ISBN atd.) a také afiliace autorů. Nad vydolovanými daty jsou například počítány hodnoty základních metrik ohodnocujících autory, konkrétně h-index,

¹ Student bakalářského studijního programu Inženýrská informatika, obor Informatika, e-mail: pittl@students.zcu.cz. Vedoucí práce: Ing. Michal Nykl, Ph.D.

² Příkladem takového pracoviště může být Katedra informatiky a výpočetní techniky ZČU.

³ BibSonomy, CiteSeerX, DBLP, Google Scholar, Mendeley, ResearchGate a Web of Science.

⁴ Oficiální stránky: <https://scholar.google.com>

⁵ Oficiální stránky: <https://www.researchgate.net>

⁶ Oficiální stránky: <http://dblp.uni-trier.de>

g-index, i5-index a i10-index.

Jak jsem již uvedl výše, informace poskytované službami nebývají konzistentní, například i z důvodu nepřesností způsobených chybami indexace či uvedením odlišných informací (např. užitím zkratk, vynecháním slova apod.). Tento systém s nepatrnými odchylkami v názvech publikací počítá a na základě podobnosti (tj. procentuální shody) názvů publikací, množiny jejich autorů a roků vydání sdružuje příslušné záznamy publikací v databázi. Spoléhá se však na korektnost uvedených jmen autorů.

3 Vizualizace výsledků

Uchovávané informace jsou strukturovaně prezentovány uživateli formou seznamu autorů a seznamu publikací, které jsou doplněny o hodnoty vypočítaných metrik. Výpis publikací je navíc formátován v souladu s citační normou ČSN ISO 690:2011 nebo BiBTeX. Oba seznamy lze řadit a také filtrovat nastavením kritérií filtrů. Obsah filtrovaných a seřazených seznamů je možné exportovat do souboru ve formátu CSV či TXT. Do exportovaného výstupu je umožněno zahrnout rozšiřující informace (spočtené metriky, odkaz na publikaci atd.).

4 Závěr

Implementovaný systém pro správu publikací byl navržen s ohledem na snadnou rozšiřitelnost o další bibliografické služby dodržením třívrstvé architektury, využitím obecného parseru, který využívá přehledný konfigurační soubor s definicemi služeb, a správou závislostí nástrojem Composer⁷. Tato aplikace je silně závislá na zvolených bibliografických službách a jakákoliv změna z jejich strany může znemožnit proces dolování informací a tak i použitelnost aplikace.

Frontend aplikace je plně responzivní⁸, přičemž důraz byl kladen na přehlednost aplikace a jednoduchost designu. Proces dolování a následného zpracování informací je uživateli dynamicky vizualizován grafy a textovým logováním v reálném čase. Přístup na administrační stránku, umožňující ovládání tohoto systému, je podmíněn přihlášením do systému.

Vytvořená aplikace je využitelná například pro získávání informací o publikační činnosti vědeckých pracovníků na Západočeské univerzitě, ale může být využita i jinou vědeckou institucí, přičemž je otevřená dalšímu vývoji.

Literatura

Connor, J. (2011) *Google Scholar Citations Open To All* [online]. Google Inc. [cit. 2017/05/13]. Dostupné z: <https://scholar.googleblog.com/2011/11/google-scholar-citations-open-to-all.html>.

Egghe, L (2006) Theory and practise of the g-index. *Scientometrics*. 69, 1, s. 131–152. ISSN 0138-9130. doi: 10.1007/s11192-006-0144-7. Dostupné z: <https://doi.org/10.1007%2Fs11192-006-0144-7>.

Firstová, Z. (2011) *Nová citační norma ČSN ISO 690:2011* [online]. [cit. 2017-05-13]. Univerzitní knihovna Západočeské univerzity v Plzni. Dostupné z: <https://sites.google.com/site/novaiso690/home>.

Hirsh, J.E. (2005) An Index to Quantify An Individual's Scientific Output. *Proceedings of the National Academy of Sciences*. 102(46), s. 16569–16572. ISSN 1091-6490.

⁷ Oficiální stránky: <https://getcomposer.org>

⁸ Responzivita zajišťuje čitelnost na široké škále zařízení o různých rozlišeních jejich displejů.