

PRIEM, J. et al., 2010. *Altmetrics: A manifesto* [online]. [cit. 2020-11-12]. Dostupné na: <http://altmetrics.org/manifesto/>

PRIEM, J., 2013. Beyond the paper. In: *Nature* [online]. Vol. 495, s. 437-440 [cit. 2020-11-12]. ISSN Dostupné na: <https://www.nature.com/articles/495437a>

RINGELHAN, S., WOLLERSHEIM, J. a I.M. WELPE, 2015. I like, I cite? Do facebook likes predict the impact of scientific work? In: *PLoS One* [online]. Vol. 10, No. 8 [cit. 2020-11-12]. Dostupné

na: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134389>

Mgr. Michaela Melicherová, PhD.

michaela.melicherova@uniba.sk

(Knižničné a edičné centrum,
Fakulta matematiky, fyziky a informatiky,
Univerzita Komenského v Bratislave)

VÝSLEDKY SEO ANALÝZY VZORKY SVETOVÝCH DSPACE REPOZITÁROV

Cieľom predloženého článku je stručne načrtnúť potrebu SEO (Search Engine Optimization) procesov nad online dostupnými informačnými systémami, medzi ktoré patria aj digitálne knižnice a inštitucionálne repozitáre. V kontexte článku chápeme SEO ako periodický proces, ktorého cieľom je snaha o zvyšovanie viditeľnosti a vyhľadateľnosti sprístupňovaného obsahu. Na vybranej vzorke registrovaných svetových DSpace repozitárov ilustrujeme aktuálny stav v tejto oblasti a formulujeme závery užitočné pre ďalší výskum v tejto oblasti.

ÚVOD K SEO A JEJ POUŽITIU NAD INŠTITUCIONÁLNYM REPOZITÁROM

Problematika tzv. SEO (Search Engine Optimization) je v prostredí internetu široko diskutovaná, a to najmä v období posledných rokov, kedy sa internet neustále a masívne rozvíja. Každým rokom pribúdajú miliardy nových webových stránok. Podľa slov viacerých autorov (Ziakis, Vlachopoulou, Kyrkoudis, Karagkiozidou 2019), môže byť pre používateľov veľmi náročné extrahovať z tohto množstva elektronických informačných zdrojov užitočné, relevantné informácie a efektívne tak uspokojiť informačné potreby. Potreba jednoduchého a efektívnejšieho vyhľadávania tak umožnila vznik a rozvoj vyhľadávacích (ná)strojov (v angl. *search engines*). Aj z toho dôvodu je nepochybne v tomto smere vhodná, ba priam až nutná určitá optimalizácia. Na myslí máme už spomenutý akronym SEO, ktorý sa do slovenčiny voľne prekladá ako optimalizácia (webu) pre vyhľadávače. V skratke ju možno vo všeobecnosti popísať ako proces neustáleho prispôsobovania webových stránok s cieľom zlepšenia ich viditeľnosti medzi inými odkazmi zobrazenými po zadaní kľúčového slova, resp. skupiny slov do rozličných internetových vyhľadávacích strojov, akými sú dnes predovšetkým Google, Bing alebo Yahoo,

prípadne v ČR obľúbený seznam.cz (Kubíček, Linhart 2010; Andrejko, 2017). Automatizované mechanizmy vyhľadávacích robotov prechádzajú im viditeľný webový korpus a indexujú jednotlivé stránky, resp. objekty. Tieto entity následne daný stroj pomocou rôznych zložitých metrik a algoritmov ohodnotí a zoradí do podoby rebríčka. Výsledné poradie je zobrazené na základe takto vypočítaného výsledku vzhľadom na konkrétne zadaný dopyt, resp. požiadavku. Lepšie umiestnenie v zoznamoch výsledkov je zvyčajne sponzorované, okrem iného, aj nárastom tzv. „organickej“ návštevnosti daného webu, ktorej zdrojom je práve vyhľadávač, pričom v tomto prípade nemá na pozíciu v poradí zobrazených výsledkov vplyv platená služba (Andrejko, 2017).

Proces spojený s aplikáciou SEO môže byť uplatniteľný ako nad komerčnými, tak i akademickými webovými stránkami. Môžu nimi byť webové stránky univerzít, akademických knižníc či rozličných elektronických informačných systémov. Autori Giannakoulou, Konstantinou, Koutsompolis et al. (2019) vo svojej štúdii preukázali, že komplexne vzatá kvalita webových stránok je kauzálne spätá práve s úrovňou SEO.

V našom článku sa zameriavame bližšie na aplikáciu SEO v akademickom prostredí – konkrétne nad inšti-

tucionálnymi repozitármi. Tieto elektronické systémy disponujú webovým rozhraním, prostredníctvom ktorého sa určitý digitálny obsah (resp. i metadáta) prezentuje ako fyzickým návštevníkom webu, tak i vyhľadávacím robotom indexujúcich viditeľný obsah. Pri otázkach viditeľnosti digitálneho obsahu zohráva významnú úlohu i pozícia daného webu (v našom prípade repozitára) vo výsledkoch konkrétneho vyhľadávania – v bežných vyhľadávacích nástrojoch, akým je napr. Google. V tomto bode možno namietat, že vedecký obsah býva zväčša indexovaný primárne vo vedeckých databázach, no viditeľnosti, vyhľadateľnosti (a napokon i citovanosti) daného vedeckého obsahu môže výrazne pomôcť i optimalizovaná prezentácia repozitára voči vyhľadávacím strojom. Netreba zabúdať ani na elektronické zdroje, akými sú Google Scholar či Microsoft Academics, pri ktorých môže mať optimalizácia (t. j. SEO) výrazný vplyv na web obsahujúci vedecký obsah.

Autori Baka a Leyni (2017) vo svojej štúdii taktiež zdôrazňujú, že úspech alebo neúspech akademickej inštitúcie, ako aj jej publikovaného vedeckého obsahu do značnej miery závisí i na spôsobe jeho prezentácie v elektronickom webovom priestore. Autori dodávajú (tamtiež), že každá vedecká inštitúcia napreduje s úspechom len vtedy, keď sa správnym spôsobom prezentuje i v online priestore a primerane interaguje so zvyškom spoločnosti. Výskumní vedeckí pracovníci teda nevyhnutne potrebujú efektívnym spôsobom sprístupňovať a komunikovať svoje vedecké výsledky – k čomu im môžu dopomôcť i správne optimalizované webové rozhrania digitálnych repozitárov.

METODIKA SEO ANALÝZY DSPACE REPOZITÁROV V ROAR

Primárnym a dlhodobým zámerom nášho výskumu je pokúsiť sa stanoviť rámcovú metodiku vedúcu k úspešnému zvyšovaniu kvality SEO parametrov nad DSpace repozitármi. DSpace, ako cieľový objekt záujmu, sme si zvolili z toho dôvodu, že aktuálne i dlhodobo stojí na vrchole rebríčka popularity softvérov určených na podporu digitálnych knižníc a repozitárov. Podľa aktuálnych údajov (august 2020), dostupných z registra OpenDOAR, získava DSpace podiel až 40 % (Jisc, 2020). Jednou z našich prvých vedeckých úloh, stanovených vo vyššie uvedenom kontexte, bolo exaktne zmerať úroveň SEO parametrov svetových DSpace repozitárov. Pri stanovení finálneho zoznamu systémov sme vychádzali z dostupných štatistík registra ROAR (The Registry of Open Access Repositories), ktorý predstavuje alternatívu k už spomenutému autoritatívnemu registru OpenDOAR. ROAR sme sa rozhodli použiť

z toho dôvodu, že umožňuje jednoduchým spôsobom vyexportovať celosvetový zoznam repozitárov využívajúcich systém DSpace do podoby excelovskej tabuľky. V integrovanom vyhľadávači v rámci ROAR sme 11. marca 2020 zadali zvolené selekčné parametre (akákoľvek krajina pôvodu, akýkoľvek typ repozitára, použitý softvér DSpace). Výsledkom bolo získanie CSV súboru obsahujúceho záznamy dovedna o 1977 DSpace repozitároch umiestnených po celom svete. Otestovať v rozumnom čase tak vysoký počet systémov je nespĺniteľná úloha. Preto sme sa rozhodli vybrať náhodným výberom vzorku a súbor skúmať štatistickými metódami. Pomocou online kalkulačára¹ sme určili, že pri veľkosti populácie 1977 objektov a intervale spoľahlivosti 95 % (alfa 0,05), marginálnej chybovosti maximálne 10 % budeme potrebovať vybrať z tohto konkrétneho základného súboru náhodnú vzorku obsahujúcu minimálne 92 systémov. Vzorku sme získali opakujúcim volaním excelovskej funkcie [randbetween (1,1977)], ktorá zabezpečila náhodný výber 92 systémov z celkového súboru obsahujúceho 1977 riadkov (1 záznam o repozitári na 1 riadku). Pri výbere vzorky sme automaticky kontrolovali výskyt možných duplicit, aby sa nestalo, že niektorý zo systémov otestujeme viackrát. Takto ošetrovanú a vybranú vzorku sme podrobili bližšiemu skúmaniu. Z množstva údajov dostupných v tabuľke sme zamerali konkrétne len stĺpce obsahujúce informáciu o názve repozitára a jeho domovskej webovej stránke (URL adresa). URL jednotlivých repozitárov sme postupne otestovali prostredníctvom troch nezávislých online SEO audítorských nástrojov – ktorých cieľom je merať a vyhodnocovať hodnoty desiatok SEO parametrov testovaného webu a z nich v závere určiť finálne skóre v rozsahu 0-100 bodov. Zisk väčšieho počtu bodov predstavuje lepší výsledok. Nástroje sme zvolili tak, aby výsledkom testovania každého systému boli tri numerické údaje, t. j. trojica bodového skóre v rozsahu 0 –100 (spoločne s aritmetickým priemerom) získaná z troch testovacích nástrojov: Seo-checker², SEO Site Checkup³ a Woorank⁴. Vo všetkých prípadoch ide o komplexné nástroje, ktoré hodnotia parametre webu podstatné pre webové vyhľadávače, ako aj používateľskú skúsenosť – ide napríklad o štruktúru a vlastnosti webu, rýchlosť jeho načítania, úroveň integrovaných metadát, počet a smerovanie odkazov, kvalitu integrovaných webových komponentov a technológií, úroveň bezpečnosti webu, jeho optimalizáciu voči sociálnym sieťam, mobilným zariadeniam a mnoho ďalších atribútov. Každý z testovacích nástrojov vo výsledku zobrazil ako čiastkové, parciálne výsledky, tak i sumárne bodové skóre, ktoré sme zapisovali postupne do tabuľky. Technické

detaily, ako aj doplnujúce informácie nájde čitateľ na webových stránkach daných nástrojov.

Testovanie prebiehalo postupne v období od 11. marca 2020 do 6. apríla 2020. Dôvodom tohto pomerne širokého časového rozpätia boli rôzne funkčné obmedzenia testovacích nástrojov, ktoré umožňovali vykonať len limitovaný počet testov každý deň.

VÝSLEDKY

Prvým zarážajúcim výsledkom nášho testovania bolo zistenie, že z 92 testovaných digitálnych repozitárov DSpace bolo až 31 systémov sieťovo nedostupných, resp. nefunkčných, no v priebehu testovacieho obdobia 4 sa systémy stali opäť funkčnými. Ku všetkým nefunkčným webstránkam systémov sme sa pokúša-

li pravidelne opätovne pristupovať v snahe zistiť, či bola ich funkčnosť medzičasom obnovená alebo nie. K 6. aprílu 2020, na konci fázy testovania, sme celkovo evidovali až 27 nefunkčných systémov, čo predstavuje viac ako 29 % z celkového počtu. Vysokú mieru nefunkčných systémov, na ktoré odkazuje register ROAR, pripisujeme neaktuálnosti sprístupňovaných záznamov. Zvyšných 65 funkčných repozitárov, resp. ich webových rozhraní sme podrobili bližšiemu skúmaniu v zmysle už predloženej metodiky. Neprinášame však konkrétne čiastkové výsledky pre konkrétne systémy. V tabuľke 1 uvádzame len vypočítané aritmetické priemery nameraných hodnôt a ďalšie vybrané štatistické ukazovatele. Vyššie bodové skóre predstavuje lepší výsledok.

	Získané body (max 100b)			
	Seo site checkup	Seo-checker	Woorank	Celkový priemer
Priemerné získané skóre v danom teste (65 repozitárov)	53,47	48,08	49,22	50,26
Smerodajná odchýlka	9,31	9,29	10,27	9,62
Medián	54	46,7	52	50,90
Modus	52	40	54	48,67
Minimum	39	21,7	33	31,23
Maximum	76	65,9	69	70,30

Tabuľka 1 Priemerné získané skóre (65 funkčných DSpace repozitárov)

Z tabuľky 1 je zrejmé, že svetové DSpace repozitáre, resp. ich webové rozhrania dosiahli v priemere doslova len priemerné bodové skóre (50,26b zo 100b možných, medián 50,90b) a v jednotlivých troch parciálnych testoch podávali navzájom porovnateľné a vyrovnané výsledky. Hoci v tabuľke 1 nie je nasledovný údaj uvedený, viac ako 60b získali v priemere z troch testov len 3 z celkového počtu 65 funkčných a testovaných DSpace repozitárov. Už tieto výsledky napovedajú, že úroveň SEO optimalizácie nie je vysoká. Aby sme získali lepšiu predstavu o výpovednej hodnote získaných údajov, rozhodli sme namerané výsledky

priamo porovnať s výsledkami SEO testov prázdnej, čistej inštalácie najnovšej verzie DSpace (6.3)⁵ s prostredím XMLUI – t. j. šlo o prázdny repozitár, navyše bez akejkoľvek dodatočnej optimalizácie, mimo najzákladnejších nastavení (hostname,...). Výsledky komparácie prinášame v tabuľke 2 nižšie.

Pri pohľade na priemerné hodnoty skóre v tabuľke 2 zreteľne vidíme, že priemerná kvalitatívna úroveň merateľných SEO parametrov svetových DSpace repozitárov je len o 7,62 % vyššia oproti absolútne neoptimalizovanej a prázdnej inštancii DSpace repozitára.

	Získané body (max 100b)			
	Seo site checkup	Seo-checker	Woorank	Priemer
Priemerné získané skóre (vzorka 65 repozitárov)	53,47	48,08	49,22	50,26
Čistá, neoptimalizovaná inštalácia DSpace 6.3	58	50,1	32	46,7
				+7,62%

Tabuľka 2 Komparácia výsledkov s čistou inštaláciou DSpace

ZÁVER

Objektívne a komplexne zhodnotiť kvalitu SEO parametrov akéhokoľvek webového rozhrania je pomerne náročné. Aj z toho dôvodu sme sa rozhodli pri skúmaní využiť tri online nástroje, ktoré nám napokon dopomohli k formovaniu určitého náčrtu aktuálneho stavu SEO nad množinou svetových DSpace repozitárov registrovaných v registri ROAR. Prekvapilo nás vysoké percento (29 %) nedostupných systémov, čo pri autoritatívnom registri považujeme za problém pravdepodobne plynúci z neaktuálnosti metadát.

Konštatujeme, že téma SEO je značne technicky orientovaná, pretože na výsledky testov orientovaných týmto smerom preukázateľne vplyvajú desiatky rôznych faktorov a nie je ani v možnostiach tohto článku ich detailne priblížiť. Miesto rozsiahleho technického pozadia skúmaného problému sme preto pristúpili priamo k predloženiu a stručnému zhodnoteniu získaných kritických výsledkov.

Pomerne nízky priemerný bodový rozdiel nameraný medzi výsledkami vzorky DSpace repozitárov a čistou inštaláciou DSpace 6.3 (tabuľka 2) poukazuje na skutočnosť, že väčšina registrovaných DSpace repozitárov z našej vzorky je len veľmi slabo (ak vôbec) optimalizovaná voči vyhľadávacím strojom. S istou pravdepodobnosťou môžeme túto vlastnosť rozšíriť i na celkový súbor repozitárov založených na DSpace a registrovaných v ROAR. Ide o zarážajúce zistenie, najmä, ak uvažíme aj skutočnosť, že i prázdna, „default“ inštalácia spĺňa kritériá troch SEO testovacích nástrojov v priemere už na 46,7 % (46,7b zo 100b možných). V kontexte nastoleného problému teda môžeme konštatovať, že minimálne v prípade DSpace repozitárov (z ROAR) vnímame vysoký potenciál a priestor pre možnú zmenu k lepšiemu nastaveniu webových roz-

hraní a uplatneniu SEO procesov. Vzhľadom na získané výsledky SEO testov teda intenzívne pociťujeme nutnosť zamerať sa na zlepšovanie odpovedajúcich parametrov nad digitálnymi repozitármi, pričom v tomto smere finalizujeme ďalšie kroky potrebného výskumu. Naším cieľom je vo výsledku priniesť odbornej verejnosti ucelenú metodiku, pomocou ktorej bude možné signifikantným spôsobom docieľiť radikálne zvýšenie kvalitatívnej úrovne väčšiny sledovaných SEO parametrov nad DSpace repozitármi. Tento prvotný cieľ by však bol celkom zbytočný, ak by sa jednalo len o prosté „tabuľkové“ zvyšovanie hodnôt, bez sledovania ďalšieho dopadu takejto optimalizácie na samotné repozitáre. Dlhodobý cieľ nášho výskumu teda spočíva, okrem iného, i v preukázateľnom zvýšení organickej návštevnosti repozitárov vplyvom uplatňovania SEO nad webovým rozhraním (DSpace) repozitárov. Potenciálnym dôsledkom navyšovania návštevnosti je i zvýšenie viditeľnosti vedeckého obsahu, jeho citovanosti a pod. Tejto i pridruženým, sekundárnym témam sa plánujeme venovať v ďalších publikovaných článkoch spätých so SEO v prostredí digitálnych repozitárov.

Poznamky:

¹ <https://www.calculator.net/sample-size-calculator.html>

² Viac informácií je dostupných na adrese: <https://suite.seotesteronline.com/seo-checker>

³ Viac informácií je dostupných na adrese: <https://seositecheckup.com/>

⁴ Viac informácií je dostupných na adrese: <https://woorank.com>

⁵ Verzia DSpace 7.0 nebola v čase testovania a písania článku ešte oficiálne vydaná pre produkčné prostredia. 2.7.2020 bola vydaná verzia DSpace 7.0 Beta 3 (Lyrasis 2020)

Zoznam bibliografických odkazov

ANDREJKO, Ján. 2017. *Čo je SEO* [online]. 2017-10-23 [cit. 2020-08-10]. Dostupné na:

<https://visibility.sk/blog/co-je-seo-prakticka-definicia/>

BAKA, Ahmad Bakeri Abu; LEYNI, Nur. 2017. Webometric Study of World Class Universities Websites. In: *Qualitative and Quantitative Methods in Libraries*, [S.l.], s. 105-115, 2017-07 [cit. 2020-05-19]. ISSN 2241-1925. Dostupné na: <http://qqml-journal.net/index.php/qqml/article/view/367>

GIANNAKOULOPOULOS, Andreas.; KONSTANTINOOU, Nikos.; KOUTSOMPOLIS Dimitris et al. 2019. Academic Excellence, Website Quality, SEO Performance: Is there a Correlation? In: *Future internet*. 2019-11, iss. 11, vol. 11. 242. DOI:10.3390/fi11110242

JISC. 2020. *OpenDOAR Statistics* [online]. 2020 [cit. 2020-08-15]. Dostupné na: https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

KUBÍČEK, Michal; LINHART, Jan. 2010. *333 tipu a triku pro SEO*. Brno: Computer Press. ISBN 978-80-251-2468-0

LYRASIS. 2020. *7.x Release Goals* [online]. 2020 [cit. 2020-08-20]. Dostupné na: <https://wiki.lyrasis.org/display/DSPACE/Dspace+7+Release+Goals>

PhDr. Matúš Formanek, PhD.

matus.formanek@fhv.uniza.sk

(Katedra mediamatiky a kultúrneho dedičstva, Fakulta humanitných vied, Žilinská univerzita v Žiline)