

# Webometria v univerzitnom prostredí (alebo o čom vypovedajú slovenské univerzitné weby)

Eubica JEDLIČKOVÁ, Slovenská poľnohospodárska knižnica pri SPU v Nitre

*World Wide Web je v súčasnosti kľúčovým miestom rozvoja, resp. rozširovania akademických, vedeckých a vzdelávacích kompetencií univerzity. Web môže atraktívnym spôsobom priblížiť univerzitu tak študentom, ako aj vedcom či ďalším záujemcom, ktorí hľadajú informácie o možnostiach štúdia, vedeckých aktivitách inštitúcie, zameraní výskumu, kvalite výučby a podobne. Možnosti a potenciál webu sú predmetom výskumu z rôznych uhlov pohľadu. Pozornosť sa sústreďuje na hodnotenie webových prezentácií univerzít ako indikátorov edukačných a vedeckých aktivít, vzájomný vzťah vedeckých výstupov a úrovne webových sídel, alebo na porovnávanie bibliometrických indikátorov s úrovňou webových sídel univerzít. Najznámejšou aktivitou v tomto smere je Ranking Web of World Universities, webometrické hodnotenie, ktoré 2x ročne realizuje Cybermetrics Lab. v Španielsku na základe vlastných metodologických. Iniciatíva sa orientuje na kvantitatívne analýzy internetu a obsahu na webe s prioritným zameraním na komunikáciu akademického a vedeckého obsahu, pričom indikátory sa snažia korelovať s tradičnými scientometrickými a bibliometrickými indikátormi. Príspevok prezentuje výsledky analýzy vybranej skupiny webových sídel slovenských univerzít, ktorá bola spracovaná v rámci špecializovanej štúdie použitím vlastnej (pôvodnej) metodiky založenej na prirodzenom crawlingu cez natívne rozhrania nad databázami vyhľadávacích strojov (Google, Google Scholar, Yahoo, Bing).*

*Webometria, webometrické analýzy, univerzity, Slovenská republika*

## Úvod

Sieť – termín, ktorý za posledných 40 rokov nadobudol špecifický význam a stal sa symbolom podstaty našej doby. „Pavučina bez pavúka“, entita s výrazne emergentným charakterom, zásadným spôsobom ovplyvňuje a modifikuje sociálnu komunikáciu až natolko, že vstupuje do oblastí sociálnych deviácií až patologických vzťahov a javov. Globálne elektronické komunikačné prostredie modifikovalo všetky prvky komunikačného modelu. Informačné entity novej kvality a komunikáty ako nositelia obsahu nie sú však jednoduchou retrokonverziou tradičného informačného priestoru a jeho prostriedkov. Ako zásadný moment vývoja sa javí kvalitatívne nový stupeň sieťového zdieľania dát a prechod na nové formy počítačovo sprostredkovanej komunikácie. Nové informačné médium ponúka nové dimenzie informačným štruktúram, ktoré prezentujú zásadne vždy technickými rámcami determinované nástroje spracovania, organizácie a komunikácie obsahu. Takouto štruktúrou je hypertext. Až po opustení intranetovej úrovne počítačovo sprostredkovanej komunikácie a zdieľania dát získava hypertextový model organizácie a diseminácie informácií nový rozmer, ktorý umožňuje naplnenie ideí metaobsahu a nadtextovej obsahovej súvislosti. Hypertext ako nový spôsob organizácie a prezentácie informácií sa až prostredníctvom webu stáva globálne využívanou komunikačnou bázou.

Nové informačné prostredie vytvára okrem iného tlak na vývoj a použitie relevantných kritérií a nástrojov, ktoré umožnia „merať“ a hodnotiť komunikáciu tak, ako je to štandardom v tradičnom komunikačnom prostredí. Napriek tvrdeniu, že elektronické publikovanie a elektronický text nie sú jednoduchou retrokonverziou klasického komunikačného reťazca,

hľadajú sa možnosti a inšpirácie v „historických koreňoch“ a prostredí tlačených informačných zdrojov. Východiskom je bibliometria a jej metódy kvantitatívneho hodnotenia. Na pôde informačnej vedy sa postupne začínajú formovať aktivity zamerané na analýzu a hodnotenie prostredia webu. Už v roku 1990 prezentuje William Paisley názor, že prechod vedeckej komunikácie do elektronického prostredia a nárast objemu publikovaných dokumentov nájde odraz v potrebe nasadenia relevantných nástrojov, ako ich používa bibliometria pri hodnotení tradičnej tlačenej formy komunikácie informácií a poznatkov (Paisley, 1990). Kvantitatívne metódy bibliometrie prenikajú do virtuálneho priestoru a slovník sa obohacuje o termíny cybermetria, netomeria, webometria a ďalšie.

### **Web a jeho hodnotenie**

V roku 1997 Tomas L. Almind a Peter Ingwersen po prvý raz predstavili webometriu ako spôsob skúmania celej sieťovo-založenej komunikácie prostredníctvom infrometrických a ďalších kvantitatívnych metód. *Web nie je nič iné ako citačná sieť, v ktorej sú tradičné informačné entity a citácie na ne naviazané, nahradené webovými stránkami. Tieto stránky sú informačnými entitami na webe a hypertextové odkazy vystupujú ako citácie* (Almind – Ingwersen, 1997).

Výskum hypertextových odkazov, ako jednej z oblastí štúdia internetu, sa v časovom horizonte posledných 20 rokov realizoval prierezo vo v rámci mnohých disciplín a svoju reflexiu našiel aj v podobe rôznorodosti pohľadov na podstatu samotného predmetu výskumu. V najvšeobecnejšej rovine však možno konštatovať, že všetky prístupy, resp. východiská k štúdiu hypertextovej štruktúry webu (odkazov) možno rozdeliť do dvoch základných skupín: je to sieťová analýza (hypertext network analysis – HNA) a webometria. Čo však obidve skupiny spája, je aplikácia predominantne kvantitatívnych metód interpretácie, resp. sumarizácie dát vydolovaných z hypertextových odkazov webu.

P. Ingwersen v roku 1998 vyvinul ako jeden z kvantitatívnych indikátorov založený na myšlienke merania priemernej hodnoty frekvencie odkazov na webe *Web Impact Factor (WIF)*. Stal sa z neho kvantitatívny nástroj pre hodnotenie, kategorizáciu a porovnávanie webových sídel. Od vzniku prešiel mnohými kritickými prehodnoteniami, vzniklo množstvo štúdií, ktoré sa pokúšali a pokúšajú overiť jeho použiteľnosť pri meraní a hodnotení kvality webových entít. Meranie WIF sa robí na základe rozdelenia odkazov do troch základných skupín: odchádzajúcich odkazov, tzv. outlinks, odkazoch prichádzajúcich z iných sídel, tzv. inlinks (resp. backlinks) a odkazov v rámci danej weboje entity, tzv. self-links (Ingwersen 1998).

Najznámejšou aktivitou webometrického hodnotenia v súčasnom akademickom priestore je *Ranking Web of World Universities*, webometrické hodnotenie, ktoré 2x ročne realizuje Cybermetrics Lab. v Španielsku na základe vlastných metodologických postupov (Aguillo – Ortega – Fernández, 2008). Iniciatíva sa orientuje na kvantitatívne analýzy internetu a obsahu na webe s prioritným zameraním na komunikáciu akademického a vedeckého obsahu, pričom indikátory sa snažia korelovať s tradičnými scientometrickými a bibliometrickými indikátormi. Oficiálne pôsobí od roku 2004 a výsledky sú publikované každý polrok. Hlavným cieľom, deklarovaným samotnými riešiteľmi, je okrem analýz a hodnotení inštitúcií, pôsobenie na akademickú komunitu v zmysle pochopenia významu publikovania na webe nielen vo význame jednoduchej diseminácie obsahu, ale ako nástroja na hodnotenie a meranie vedeckých aktivít a vplyvu na celú komunitu.

Hodnotené indikátory:

- veľkosť sídla (size – S): počet stránok (Yahoo, Google, Live Search, Exalead),
- vizibilita (visibility – V): počet unikátnych externých odkazov typu inlinks (Yahoo),

- obsah s pridanou hodnotou podľa typového rozšírenia zdrojového súboru (rich files – R): publikačné aktivity v. formáty – PDF, DOC, PPT, PS,
- akademický obsah (scholar – Sc): počet stránok odkazovaných v špecializovanej databáze Google Scholar.

Zastúpenie indikátorov pri hodnotení:

S: 20% - R: 15% - V: 50% - Sc.: 15%

### **Webometrická analýza slovenských univerzít**

V rámci špecializovanej štúdie sme sa zamerali na zhodnotenie aplikácie hypertextu pri diseminácii obsahu v rámci vybranej skupiny webových sídel univerzít v SR prostredníctvom webometrickej analýzy odkazov. Hodnotenie sa realizovalo na úrovni parciálnych štúdií a hodnotení, ktoré sú reprezentované skupinou čiastkových cieľov.

#### **• Čiastkový cieľ 1**

Analýza hypertextových odkazov kolekcie webových sídel univerzít v SR v prostredí vybranej skupiny vyhľadávacích strojov; zhodnotenie veľkosti webových sídel a topológia časti univerzitného webu SR (verejné vysoké a štátne vysoké školy); geografická špecifikácia. Úloha: Charakteristika a topológia vybranej časti univerzitného webu podľa vybraných kritérií: všeobecná topológia, geografická/regiónálna topológia.

Na základe získaných dát boli realizované:

- spracovanie incidenčnej matice pre všetky uzly siete (20 univerzít),
- výpočet veľkosti sídla, mohutnosti uzla (veľkosť vektora) pre každý uzol zvlášť a spracovanie topologickej tabuľky,
- vytvorenie neorientovaného grafu odkazov podľa incidenčnej matice,
- spracovanie geografickej topológie podľa definovaných regiónov.

V prostredí vyhľadávacích strojov Yahoo, Google a Bing boli zadané úlohy indexovať referencie z domény A na doménu B. Spracovanie surových dát: incidenčná matica, výpočet mohutnosti uzla, interpretácia na úrovni dendrogramu, (agnes), krabicového (boxplot) a hviezdicového (starplot) grafu.

#### **• Čiastkový cieľ 2**

Analýza hypertextových odkazov kolekcie webových sídel slovenských univerzít v prostredí špecializovanej databázy Google Scholar; vizibilita domén vybranej skupiny slovenských univerzít z pohľadu diseminácie vedeckého/akademického obsahu. Úloha: Viditeľnosť webových sídel univerzít v databáze Google Scholar, zastúpenie odkazovaných zdrojov podľa obsahu/dokumentu.

#### **• Čiastkový cieľ 3**

Analýza hypertextových odkazov kolekcie webových sídel slovenských univerzít a vybranej skupiny generických domén a sub-domén z univerzitného prostredia v zahraničí.

Do štúdie práce bola zahrnutá skupina 17 verejných vysokých škôl a 3 štátnych vysokých škôl, ktoré patria do systému vysokých škôl poskytujúcich vysokoškolské vzdelávanie na území Slovenskej republiky. Skúmaná množina teda predstavuje – s vylúčením zahraničných vysokých škôl – 66,6% podiel z celkového počtu vysokých škôl v Slovenskej republike a 85%, resp. 100% podiel zo skupiny verejných vysokých škôl, resp. štátnych vysokých škôl. Do štúdie neboli zahrnuté umelecké vysoké školy, súkromné vysoké školy a zahraničné inštitúcie. Dáta boli primárne spracované na úrovni absolútnych hodnôt, následne sa pre potreby ďalšieho spracovania dáta normalizovali tak, aby boli eliminované riziká súvisiace s nekonzistenciou súboru. Štandardne bola použitá log-normalizácia, ktorá umožňuje bez deformácií (resp. so zanedbateľnými posunmi) komparáciu a interpretáciu výsledkov. Tento postup bol nevyhnutný z dôvodu nulových hodnôt a veľmi širokého rozptylu hodnôt. Relatívne hodnoty umožňujú eliminovať niektoré vyššie uvedené riziká. Pre zber dát bola použitá metóda manuálneho vyhľadávania na webových stránkach ako aj

indexovanie webových stránok pomocou vybranej skupiny vyhľadávacích strojov (Google, Yahoo, Bing) a špecializovanej databázy (Google Scholar). Nakoľko jednotlivé vyhľadávacie stroje používajú v rámci operátorov odlišnú syntax, výsledky sú prezentované a diskutované v kontexte týchto diferencií. Syntax dopytu je popísaná pri každom konkrétnom prípade. V rámci manuálneho crawlingu boli navštevované webové sídla univerzít (vstupným uzlom bola domovská stránka) a prostredníctvom navigácie, ktorú konkrétne webové sídlo používa, resp. prostredníctvom vyhľadávania (ak je dostupné na stránke) sa v závislosti na definovaných kritériách monitoroval publikovaný obsah.

### *Spracovanie dát*

V procese základného spracovania súborov surových dát bo li použité metódy viackriteriálnej analýzy a teórie grafov. Viackriteriálny prístup umožňuje spracovanie a interpretáciu heterogénnych dát. Pre spracovanie základných tabuľkových prehľadov bol použitý relevantný editor. Vizualizácia výsledkov využíva teóriu grafov ako jednu z metód operačného výskumu. Grafické výstupy boli spracované v prostredí aplikačného softvéru R.

### **Záver**

Výsledky práce potvrdili skutočnosť, že aj slovenské vysokoškolské, akademické prostredie si uvedomuje zásadný význam a úlohu webu pri sprístupňovaní výsledkov svojej činnosti a na zviditeľnenie vo virtuálnom informačnom a komunikačnom priestore. Manažment týchto inštitúcií si osvojuje fakt, že „reklama“ na webe sa stáva nevyhnutnou, univerzity vyvíjajú svoju činnosť v komerčných trhových vzťahoch. Webové sídla hodnotenej časti univerzitnej siete predstavujú rôznorodú množinu prístupov, aplikácií a komunikovaného obsahu a v kontexte štúdie, heterogénnu množinu aplikácií pri dokumentácii a sprístupňovaní tak intelektuálnej produkcie a vedeckovýskumných výstupov, ako aj širšieho obsahového rámca – edukačné aktivity a pôsobenie.

Závery poukazujú na výrazné rozdiely v prístupoch k využívaniu hypertextu v prostredí slovenskej univerzitnej komunity. Prevaha - a to výrazná – odkazov na úrovni internej navigácie v rámci webového sídla redukuje hypertext na úroveň jednoduchého ukazovateľa smeru. Motivácia a etika odkazovania na partnerov vo vlastnej sieti je veľmi nízka. Výrazný je podiel tzv. administratívneho obsahu. Nezodpovedajúca možnosťiam a najmä dosiahnutým výsledkom, podľa nášho názoru, je prezentácia akademického vedeckovýskumného obsahu, edičných a publikačných aktivít, resp. sa tak „javí“ z dôvodu používanej navigácie. Niekedy prevažujú, alebo sú v nevyváženej miere používané rôzne grafické prvky. Skvalitniť možno vyhľadávanie. Potrebné je minimalizovať výskyt neaktuálneho obsahu alebo chýb z dôvodu odstránenia zdrojového dokumentu, ale nie odkazu na dokument na všetkých miestach jeho ďalších výskytov na webových stránkach v rámci sídla. Dokumenty v prípade potreby viacnásobného odkazovania, neduplikovať, ale skutočne odkazovať na jeden primárny zdroj.

Použité nástroje a postupy zberu a dobývania základnej množiny sú iba jednou z alternatív. Objektívnejšie a obsahovo bohatšie výsledky by bolo možné získať, podľa nášho názoru, metódou analýzy transakčných logov. Tento postup by si však vyžadoval širšiu pracovnú skupinu a spoluprácu na úrovni administrátorov domén. Získané výsledky by mohli byť východiskom diskusií s realizačným tímom, ktorý pracuje na svetovom hodnotení univerzít, nakoľko hodnotenia, podľa nášho názoru, určitým spôsobom skresľujú stav slovenského univerzitného webu. Univerzity musia nájsť vo virtuálnom svete svoje miesto a byť nielen veľké rozsahom publikovaných dokumentov, ale hlavne viditeľné vďaka publikovanému obsahu, ktorý zaujme a bude z univerzity formovať autoritu „siete bez pavúka“. Tvorba webu je taktiež o motivácii odkazovania a v tomto smere majú, podľa nášho názoru, slovenské univerzitné weby výrazné rezervy až nedostatky. Bolo by vhodné, aby sa zvýšila početnosť

odkazov na partnerské inštitúcie a medzi univerzitami navzájom. Neodkazovať iba na zahraničné zdroje, ale zmysluplne odkazovať na partnerov vo vlastnej časti siete.

Prieskum bol prvou štúdiou zameranou na kvantitatívnu analýzu časti slovenského univerzitného webu a to prostredníctvom vlastnej metodiky a postupov. O získaných výsledkoch možno ďalej diskutovať a najmä zopakovať postupy a na základe získaných výsledkov overiť možnosť navrhnutých postupov pre podobné hodnotenia z perspektívneho hľadiska. Použité postupy ponúkajú možnosť komparácie s výsledkami, ktoré by boli získané v rámci iných prieskumov, nakoľko pri tvorbe metodiky sa zohľadňovali práve aspekty opakovateľnosti a verifikovateľnosti. Zatiaľ možno komparovať iba s výsledkami publikovanými Cybermetrics Lab. No a kvantitatívny výskum by mal ísť ruka v ruke s výskumom kvalitatívnym.

### **PodĎakovanie**

Autorka na tomto mieste vyjadruje svoje poďakovanie Mgr. R. Muchovi za jeho pomoc pri zbere a spracovaní dát, metodickú pomoc a podporu, ako aj dr. Isidrovi Aquillovi za spoluprácu a konzultácie v prípravných fázach projektu.

### **Literatúra**

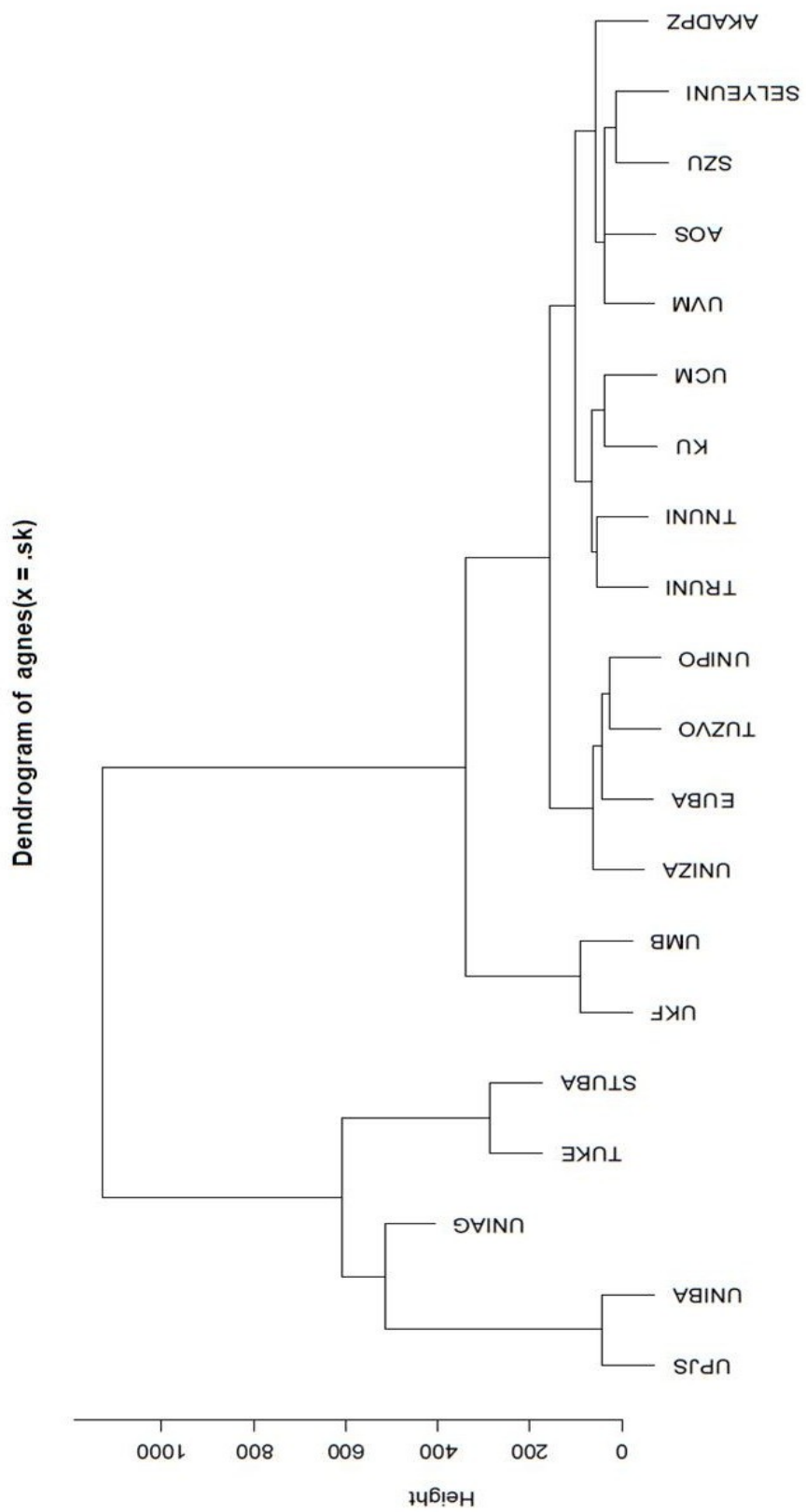
- AGUILLO, I.F. – ORTEGA, J.L. – FERNÁNDEZ, M. 2008. Webometrics ranking of world universities : introduction, methodology, and future developments. In *Higher education in Europe*, vol. 33, July-October 2008, no. 2/3, p. 233-244.
- ALMIND, T. – INGWERSEN, P. 1997. Informetric analyses on the World Wide Web : methodological approaches to „webometrics“. In *Journal of Documentation*, vol. 53, 1997, no. 4, p. 404-426.
- BJÖRNEBORN, L - INGWERSEN, P. 2004. Toward a basic framework for webometrics. In *Journal of the American Society for Information Science and Technology*, vol. 55, 2004, no.14, p. 1216-1227.
- INGWERSEN, P. 1998. The calculation of Web impact factors. In *Journal of Documentation*, vol. 54, 1998, no. 2, p. 236-243.
- MAKULOVÁ, S. 2006-2010. SEO optimalizácia. In *ELET newsletter* [online], 2006-2010 [cit. 2010-01-20]. Dostupný na internete: <<http://www.elet.sk/?seo-optimalizacia-1>>.
- MAKULOVÁ, S. 2006. Analýza faktorov vplyvujúcich na nájdiťnosť webových sídiel. In *Information Use in Information Society : International Conference. Bratislava, Slovakia, October 10-11, 2006*. Bratislava : Centrum VTI SR, 2006, s. 109 – 118.
- ORTEGA, J.L. – AGUILLO, I.F. 2009. Mapping world-class universities on the web. In *Information Procesing and Management*, vol. 45, 2009, p. 272-279.
- THELWALL, M. 2006. Interpreting Social Science Link Analysis Research: A Theoretical Framework. In *Journal of the American Society for Information Science and Technology*, vol. 57, 2006, no. 1, p. 60-68.
- THELWALL, M. 2004. *Link analysis: an information science approach*. Amsterdam : Elsevier Academic Press, 2004. 269 p. ISBN 0-12-088553-0.

### **Kontakt**

PhDr. Ľubica JEDLIČKOVÁ, PhD.

Slovenská poľnohospodárska univerzita – Slovenská poľnohospodárska knižnica pri SPU,  
Štúrova 51, 949 59 Nitra

E-mail: [lubica.jedlickova@uniag.sk](mailto:lubica.jedlickova@uniag.sk)



Obr. 1 – Dendrogram hodnotených domén na úrovni vzorky – kritérium stupeň zložitosti uzla