

## Porovnání informačních zdrojů Web of Science a SCOPUS

Ivana Kadlecová, Hana Tomanová, Jaroslav Meixner  
Knihovna Akademie věd České republiky, Praha  
kadlec.i@seznam.cz

INFORUM 2008: 14. konference o profesionálních informačních zdrojích  
Praha, 28. - 30.5. 2008

### **Abstrakt**

Předmětem tohoto příspěvku je porovnání dvou největších dostupných citačních zdrojů, Web of Science (WOS) a SCOPUS. Napsáno i řečeno bylo již mnohými specialisty mnohé, oba produkty porovnávali specialisté z různých oborů a jak je uvedeno v citované literatuře, z domova i ze zahraničí. Oba výše jmenované a jistě i další informační zdroje slouží široké vědecké a odborné veřejnosti k vyhledávání – autorů, publikací, konkrétních článků, plných textů, pokud jsou propojeny, indikátorů, citací, adres atd. To však není vše, co lze z těchto informačních zdrojů vytěžit.

V úvodním příspěvku této sekce (P. Ráb) odezněla věta:

**... výkon ve vědě a výzkumu má vždy dvě složky (zastoupené v různém poměru): kvalitativní (= *co* bylo vykonáno) a kvantitativní (= *kolik* bylo vykonáno). Zatímco hodnocení první složky je nutno provádět peer review systémem, pro druhou složku jsou široce využívány databáze vědeckých informací a indikátory, které z nich lze odvodit.**

Databáze tedy slouží i jako podklad pro kvantitativní hodnocení výkonu ve vědě.

**...databáze vědeckých informací a případné z nich odvoditelné indikátory nejsou (nebo by neměly být, pokud neuvažujeme „lidský faktor“ , tj. chyb učiněné vkladateli dat) ovlivněny subjektivně a proto tyto soubory vědeckých a odborných informací jsou téměř vždy *kostrou a základem hodnotících systémů. Jejich předností je přehlednost, snadnost komprimace informací, číselná vyjádření.***

**Web of Science** je součástí produktu Web of Knowledge (WoK), potomkem původní tištěného Science Citation Index, který v roce 1960 přivedl na svět ve Filadelfii v USA Eugen Garfield v ústavu, který založil - Institute of Scientific Information a který postupným prodejem a slučováním nese nyní název Thomson – Reuter. V české národní multilicenci je obsažen rovněž Journal Citation Reports, který býval dříve nedílnou součástí tištěného citačního indexu, pravidelně jako číslo 19 v každém ročníku.

**SCOPUS** je produktem evropského obřího vědeckého vydavatelství ELSEVIER, které v roce 2004 vytvořilo vskutku konkurenční - evropské - prostředí zařitému a klasickému americkému produktu.

Porovnávání obou produktů je problematické, protože každý z obou byl založen v jiném období a měl různě dlouhé vývojové období za velice odlišných technologických podmínek.

Z uživatelského pohledu lze vynechat minulost obou produktů a konstatovat současné možnosti či omezení z pohledu uživatele i pro potřeby hodnocení.

**...vědecká a výzkumná výkonnost = články, přednášky, knihy, citace, aplikace**

Uvedené kritérium splňují oba zdroje (bez monografií); pokud v budoucnu dojde k úměrnému hodnocení konferenčních příspěvků, nikoliv jen k sledování s minoritním bodovým ohodnocením, pak bude další součástí Web of Knowledge, ISI Proceedings, určitě důležitým zdrojem informací k hodnocení.

### **Druhový obsah a rozsah**

#### **WOS**

#### **SCOPUS**

Periodika 9 325 titulů	Periodika 16 150 titulů
Ročenky („Book series“) cca 150 titulů	Ročenky („Book series“) cca 200 titulů
<i>V dalších zdrojích</i>	Obchodní literatura 600 publikací
Konferenční příspěvky 5 mil záznamů	Konferenční příspěvky 520 titulů
Patenty?	Patenty?

**Počet záznamů 39,6 mil.**

**Počet záznamů 33 mil.**

### **Předmětový obsah a rozsah**

#### **WOS**

#### **SCOPUS**

Science (SCI-Expanded) 6650 titulů	Science 14 700 publikací
Social Science (SSCI) 2020 titulů	Social Science 2 850 publikací
Arts & Humanities (A&HCI) 1170 titulů	Arts & Humanities 220 publikací

Je třeba poznamenat, že se jedná o tzv. aktivní tituly, tedy pouze o takové, které nepřestaly v průběhu času vycházet. Vzhledem k různé skladbě sledovaných dokumentů hovoří WoS o *titulech*, SCOPUS udává počty *publikací*.

U konferenčních příspěvků a patentů je nesnadné porovnání. *Web of Knowledge* obsahuje samostatnou databázi *ISI Proceedings*, v letošním roce provázanou s **WoS**, která od roku 1990 obsahuje zhruba 5 mil. záznamů. Patenty ze 41 světových patentových úřadů obsahuje *Derwent Innovation Index*, rovněž provázaný s **WoS**. **SCOPUS** udává materiál z přibližně 520 konferenčních sborníků bez udání rozsahu, k patentovým informacím umožňuje přístup prostřednictvím vyhledávače *Scirus* do volně přístupných dat čtyř světových patentových úřadů.

## Zařazování časopisů

### WOS

### SCOPUS

Časopis je zahrnut do WoS tehdy, pokud má prokazatelně určitou odezvu, tedy časopis musí být citován. <a href="http://scientific.thomsonreuters.com/mjl/selection/#jsc">http://scientific.thomsonreuters.com/mjl/selection/#jsc</a>	Pro zařazení titulu do zpracování jsou určena následující kritéria: 1. Články musí mít anglický abstrakt 2. Časopis musí vycházet pravidelně 3. Práce musí být recenzovány.
--	--

*...viditelnost v oboru a ambice = světová, evropská, národní, lokální úroveň*

### WOS

### SCOPUS

Preference angloamerické produkce	Zaměřuje se i na evropskou a jinou literaturu
-----------------------------------	---

## Retrospektiva

### WOS

### SCOPUS

Science 1900 -	Life & Health Science 1966-
Social Science 1956 -	Engineering 1970 -
Arts & Humanities 1975 -	Environmental Science 1980 -
	Biological & Agricultural Sci 1994 -
	Social Science 1996 -
	Physics, Chemistry and Mathematics 1996 -

## České časopisy v databázi

### WOS

### SCOPUS

Science Edition 22 titulů	89 titulů
Social Science Edition 4 tituly	

## Indikátory

### WOS

### SCOPUS

H-index. IF, další	H- index
--------------------	----------

S ohledem na délku retrospektivy má H-index ve **WoS** obvykle vyšší hodnotu. Je nezbytné uvádět vždy, ze kterého zdroje H-index pochází. Impact Factor (IF) je počítán pouze ve WoS, Hodnota indikátorů se mění v čase, u IF každým rokem a u H-Indexu průběžně.

*... důležitější je hodnocení trendů, než okamžitý stav*

Obě databáze nabízejí nejrozumnější možnosti vypracování rešerší a následně porovnat vývoj v čase: příkladem jsou analýzy, které zveřejnil rektor UK v časopise Univerzity Karlovy FORUM nebo autoři Ráb a Kadlecová z AV ČR v Akademickém Bulletinu, obě práce vyšly ve druhé polovině roku 2007. Není cílem tohoto příspěvku zabývat se obsahem analýz, ale jisté je, že podklady k nim byly čerpány v obou případech z Web of Science. V AV ČR byly k podkladům pro evaluaci Výzkumných záměrů použity další zdroje, a sice SCOPUS a ASEP/ARL, ale to už je jiná historie. Oba zdroje nabízejí analýzu výsledků rešerší, rozdíl spočívá v délce retrospektivy a v oboru.

### Vyhledávací prostředí, omezení a problémy

**WoS** nepřipouští diakritiku, **SCOPUS** ji umožňuje; tím se vyhledávání ve SCOPUSU stává snazší, zejména pro naše české prostředí. Vzhledem k tomu, že ve SCOPUSU jsou zařazeny též časopisy v českém jazyce, bývá i afilace uvedena v češtině. Určité problémy nastávají při vyhledávání autora článku i při hledání celého souboru prací dle názvu instituce.

Případy stejných příjmení a dokonce i křestních jmen u stejného oboru lze řešit zpřesněním dotazu, spolupráci s autory je třeba vyžádat v případech, kdy autor střídal pracoviště a má různé adresy u publikací, někdy autor(ka) změní v průběhu své kariéry příjmení. To jsou jen příklady, které ztěžují vyhledávání, ale jsou snadno řešitelné. Chybným vyhledáním může být ohroženo hodnocení autora, ať už nižším počtem prací, následně nižším počtem citací nebo zkráceným H-indexem.

#### **Daleko závažnější problém může nastat při vyhledání instituce.**

Pokud uživatelé vyhledávají databázi podle názvu instituce, narazí na problém výskytu alternativních názvů instituce. Například Elsevier uvádí, že v databázi SCOPUS se vyskytuje v záznamech více než 1700 variant názvů pro Massachusetts Institute of Technology. Každé variantě názvu instituce v databázi přitom odpovídá alespoň jeden záznam dokumentu. Výsledek vyhledávání je ovlivněn kvalitou údajů v afilačních polích databází (např. zkrácené názvy institucí, lokační údaje, podrobnost údajů).

#### **Vyhledávání instituce - WOS**

Pro zjištění publikační činnosti konkrétní instituce v databázi Web of Science je nutné identifikovat množinu záznamů dokumentů příslušejících dané instituci na základě adresy autorů.

Prostřednictvím vyhledávacího rozhraní Web of Science je možné zadávat kombinace rešeršních dotazů v poli **AD (Address)**:

- při využití podpolí pole AD (zejména OG, SG a CU) a operátorů **SAME, AND, OR, NOT**
- při využití standardizovaných zkratk **Address Abbreviations** pro formulaci názvu instituce v dotazu
- při využití nástroje pro analýzu výsledků dotazu **Analyze Results**, zejména jeho funkce pro seřazení záznamů podle názvu instituce

Vzhledem k tomu, že není k dispozici abecední index všech variant názvů institucí v databázi, je obtížné podchytit při formulaci dotazu všechny možné varianty názvu pro konkrétní instituci. Pro názvy institucí z ČR existuje různé varianty názvů v češtině i angličtině.

Pole AD se skládá z podpolí: OG (Organization), SG (Suborganization), SA (Street Address), CI (City), PS (Province/State), CU (Country), ZP (Zip/Postal Code). Všechna podpole nemusí být v záznamu vyplněna a podpole se mohou i opakovat.

#### Příklady záznamů:

**Address:** Acad Sci Czech Republic, Inst Nucl Phys, Prague 25068, Czech Republic

**Address:** Acad Sci Czech Republic, Inst Macromol Chem, Dept Struct Anal, CR-16206 Prague 6, Czech Republic

**Address:** Biol Ctr Acad Sci Czech Republ, Inst Plant Mol Biol, Dept Plant Virol, Ceske Budejovice 37005, Czech Republic

Příklad dotazu: **AD=(Inst Comp Sci SAME Czech Republic)**,

Příklad výsledného záznamu - pole AD:

**Addresses:** Konar, O (reprint author), Acad Sci Czech Republic, Inst Comp Sci, Pod Vodarenskou Vezi 2, Prague 18207 8, Czech Republic  
Acad Sci Czech Republic, Inst Comp Sci, Prague 18207 8, Czech Republic

Výraz pro instituci - pole OG, podmínka CU=Czech Republic	Počet záznamů v db WOS (stav 04/2008)
Acad Sci Czech Republ	26783
Acad Sci Czech Republic	2242
Czechoslovak Acad Sci	758
ASCR	741
AS CR	612
AV CR	451
Acad Sci	278
CAS	210
AVCR	71

Nejčastější varianty názvu instituce – Akademie věd ČR

Uvedené ukázky nevypovídají o nízké kvalitě databáze, ale o špatném uvedení adresy autora, mnohdy způsobené redakcí.

V případě zadání dotazu ve formě **AD=(Inst Comp Sci AND Czech Republic)** jsou výsledkem dotazu všechny záznamy z dotazu **AD=(Inst Comp Sci SAME Czech Republic)**, ale také záznamy, které mají výrazy **Inst Comp Sci** a **Czech Republic** v různých záznamech v poli AD.

Příklad výsledného záznamu - pole AD:

**Addresses:** Borodziuk, S (reprint author), Acad Sci, Joint Res Lab, PALS, Za Slovankou 3, Prague 18221 8, Czech Republic  
Acad Sci, Joint Res Lab, PALS, Prague 18221 8, Czech Republic  
Acad Sci, Inst Phys, Prague 18221 8, Czech Republic  
Acad Sci, Inst Plasma, Prague 18221 8, Czech Republic  
Czech Tech Univ, Fac Nucl Sci & Phys Engrn, Prague 11519 1, Czech Republic  
Warsaw Univ Technol, Inst Comp Sci, PL-00665 Warsaw, Poland

Address Abbreviations: WoS využívá řadu ustálených zkratk pro často opakující se výrazy z pole Address (např. *Acad* pro Academy). Seznam používaných zkratk je na adrese [http://images.isiknowledge.com/help/WOS/h\\_adabrv.html](http://images.isiknowledge.com/help/WOS/h_adabrv.html).

## Vyhledávání instituce - SCOPUS

SCOPUS Affiliation Identifier (Identifikátor afilací) je novým softwarovým nástrojem určeným pro identifikaci záznamů dokumentů v databázi SCOPUS, které přísluší ke konkrétní instituci. Identifikátor afilací by měl umožnit snadnější a časově méně náročné

vyhledávání záznamů instituce. Podle oznámení Elseviera identifikátor afiliací využívá sofistikované algoritmy k automatickému sjednocení a deduplikaci názvů institucí. Každé instituci (afilaci) ve SCOPUSu je následně přiděleno jednoznačné identifikační číslo ID. Pokud je zadán název instituce, objeví se v odpovědi nabídka **Find unmatched affiliations** a v ní je obsažen soubor souvisejících adres, takže je poměrně snadné sestavit seznam odpovídajících názvů a vyloučit chybné nebo dokonce kuriózní. Vzhledem k tomu, že Affiliation Identifier byl uveden do plného provozu teprve před několika dny, byly dosavadní zkušenosti s identifikací názvů institucí nepříliš dobré a hledání komplikované. Po zavedení této novinky se SCOPUS velmi rychle staví do pozice rovnocenného partnera.

## Závěr

Pro hodnocení vědeckých článků je typ dokumentu  $J_{imp}$  zásadním a vysoce bodovaným výsledkem. Pro tento účel se WoS stává nenahraditelným zdrojem – o tom není pochyb. Stejně tak je nenahraditelný, aspoň v nejbližších letech, pro vyhledávání hodnot **IF** a pro svou dlouhou retrospektivu i pro citovanost článků. **SCOPUS** je výborným zdrojem pro porovnání výsledků v evropském prostředí, což je dáno počtem sledovaných titulů časopisů. V Metodice hodnocení výsledků výzkumu, vývoje a inovací Rady pro výzkum, vývoj a inovace Vlády ČR se pro příští rok objevuje termín  $J_{neimp}$  - článek zařazený v seznamu neimpaktovaných recenzovaných periodik vydávaných v České republice (zveřejněno na [www.vyzkum.cz](http://www.vyzkum.cz)) nebo **je tento časopis evidován v některé ze světově uznávaných databází**. K hodnocení je třeba přistupovat z více pohledů, čerpat z více zdrojů a využívat možnost kontroly. Jediný zdroj sebelepší kvality již v současné době nestačí.

ŠILHÁNEK, Jaroslav. Orientační testování báze dat SCOPUS. *Ikaros* [online]. 2007, roč. 11, č. 1 <http://www.ikaros.cz/node/3800>

JASCO, Peter. 2005. As we may search – Comparison of major features of the Web of Science, SCOPUS, and Google Scholar citation-based and citation-enhanced databases. *Current Science* [online]. November 10, 2005, vol. 89, no. 9 <http://www.ias.ac.in/currsci/nov102005/1537.pdf>

LaGUARDIA, Cheryl. 2005. E-Views and Reviews: SCOPUS vs. Web of Science. *Library Journal* [online]. Jan. 15, 2005 <http://www.libraryjournal.com/article/CA491154.html%22>

DEIS, Louise F.; GOODMAN, David. 2005. Web of Science (2004 version) and SCOPUS. *The Charleston Advisor* [online]. Jan. 2005, vol. 6, no. 3 <http://charlestonco.com/comp.cfm?id=43>

DEIS, Louise F.; GOODMAN, David. 2006. Update on Scopus. *The Charleston Advisor*. Jan. 2006, vol. 7, no. 3 <http://www.charlestonco.com/comp.cfm?id=55>

Scopus Affiliation Identifier Takes Research to the Next Level. *Inside Scopus : news for librarians* [online]. 2008, is. 7 [cit. 2008-04-20]. Dostupný z WWW: <http://www.info.scopus.com/is/7/2/>