

# Hodnocení inforatického výzkumu

Radim Bělohlávek

31. května 2010

**Abstrakt.** Studie se zabývá problematikou hodnocení inforatického výzkumu. Shrnuje poznatky, ke kterým se dospělo v literatuře publikované na toto téma, přináší další názory, formuluje zásady, kterými by se hodnocení inforatického výzkumu mělo řídit, a upozorňuje na možná úskalí různých přístupů k hodnocení. Studie dále analyzuje metodiku hodnocení výzkumu a vývoje používanou v České republice a na příkladech ukazuje její nedostatky v bodování publikací.

## 1. Hodnocení výzkumu

Hodnocení kvality výzkumu v jakémkoli oboru je citlivé téma. Jedním aspektem je sociální význam takového hodnocení. Hodnotíme výsledky lidí, jejichž náplní práce je vytvářet nové poznatky, vymýšlet nové a lepší. Tito lidé — vědci — patří zpravidla k intelektuální elitě společnosti, požívají ve společnosti vážnosti, a tak se jich hodnocení výsledků jejich činnosti, ať pozitivní či negativní, z pochopitelných důvodů silně dotýká.

Druhým aspektem je ekonomický význam hodnocení. V mnoha zemích, včetně České republiky, má hodnocení kvality výzkumu ekonomické důsledky, které zásadním způsobem ovlivňují životy lidí.

Hodnocení kvality výzkumu se používá například v následujících situacích:

- při kariéerním postupu v akademické sféře (obhajoby doktorských prací, habilitace, profesury), kdy součástí materiálů dodaných uchazečem je jeho odborný životopis, tj. zejm. seznam výsledků jeho výzkumu,
- při přijímání na místa výzkumných pracovníků, vysokoškolských učitelů apod.,
- při udělování grantů (finančních podpor) jednotlivcům i výzkumným týmům,
- při udělování akreditací, tj. oprávnění uskutečňovat studijní programy,
- při dělení finančních prostředků v rámci dané organizační jednotky (přidělování peněz fakultám v rámci vysoké školy, katedrám v rámci fakulty, jednotlivcům v rámci katedry),
- při dělení finančních prostředků ze státního rozpočtu mezi vysoké školy, výzkumné ústavy a případně jiné výzkumné instituce.

Ekonomické dopady hodnocení výsledků výzkumu jsou tedy značné a s postupem doby se ve většině vyspělých zemí zvětšují. Otázka, jak se pozná kvalitní výzkum, je ale delikátní a neexistuje na ni jednoduchá odpověď. Není proto divu, že o způsobu hodnocení výzkumu se vedou bouřlivé diskuze.

Je třeba poznamenat, že spíše než o systému hodnocení je správné hovořit o systémech hodnocení. Z různých důvodů, z nichž některé zmíníme později, se totiž systém vhodný pro hodnocení kolektivů, popř. institucí, může ukázat zcela nevhodným pro hodnocení jedinců a naopak. Známým příkladem je hodnocení pomocí tzv. impact factor (IF) časopisu, jehož použití se pro hodnocení jedinců nedoporučuje [Am00]. Na druhou stranu ale zkušenost ukazuje, že vedoucí pracovníci mají tendenci používat systém hodnocení, podle kterého byly jejich organizační jednotce přiděleny finanční prostředky, i pro hodnocení jim podřízených jednotek a jedinců. To ukazuje potřebu jisté provázanosti systémů hodnocení používaných na různých úrovních. Není například představitelné, aby systémy hodnocení používané na různých úrovních byly v hrubém rozporu (např. aby to, co je hodnoceno jako kvalitní v jednom systému bylo hodnoceno jako nekvalitní nebo neutrální v druhém systému a naopak).

Potřeba kvalitního hodnocení výsledků výzkumu je obzvláště markantní ve společnostech, kde převážná část prostředků na výzkum pochází ze státního rozpočtu. Takovou zemí je i Česká republika. V České republice byly v posledních několika letech zahájeny reformy systému financování výzkumu a vývoje. Bouřlivá debata se vede zejména o vhodnosti metodiky hodnocení výsledků výzkumu a vývoje, podle které se začaly výzkumným institucím přidělovat prostředky tzv. institucionální podpory. Tato metodika má vážné nedostatky, které jsou diskutovány na různých fórech. Poznamenejme, že podobné reakce vyvolávají snahy o zavedení systémů hodnocení výzkumu i v jiných zemích (viz např. dopis [Let08] významných profesorů informatiky švýcarských univerzit protestujících proti systému založeném na použití bibliometrických dat pro hodnocení výkonu švýcarských univerzit).

Přestože nelze očekávat, že jakákoli metodika podobná svým významem metodice hodnocení výsledků výzkumu a vývoje vznikne napoprvé v podobě všem vyhovující, ukazuje se, že problém spočívá do jisté míry v tom, že v době přípravy nebyla metodika potřebným způsobem diskutována. Nedostatky, které metodika obsahuje, jsou jedním z důvodů vzniku této studie.

Lze předpokládat, že mezi výhradami směřovanými k vládní metodice budou jak ty, které budou oborově neutrální, tj. shodnou se na nich vědci ze všech oborů, tak výhrady oborově specifické, tj. výhrady vznášené zástupci jen určitých oborů. V této studii nám jde především o poslední jmenované výhrady. Konkrétně nám jde o část vládní metodiky, zabývající se hodnocením publikací. Ukážeme, že tato část metodiky má z pohledu oboru informatika vážné nedostatky.

Kromě konkrétních výhrad k vládní metodice, které doložíme příklady, ve studii shrneme specifika oboru informatika a zformulujeme zásady, které by měly být každým systémem hodnocení informatického výzkumu respektovány. Studie by tak měla být užitečná při dalších jednáních o metodice hodnocení výsledků výzkumu a vývoje i jako podpůrný materiál pro přípravu jiných systémů hodnocení informatického výzkumu.

## **2. Informatický výzkum a jeho specifika**

V literatuře existují studie, které upozorňují na to, že informatika je oborem s jistými specifiky, které musí být při hodnocení informatického výzkumu respektovány. Informatiku jako obor lze definovat mnoha způsoby. V [De89] je informatika definována následovně:

Computer science and engineering is the systematic study of algorithmic processes—their theory, analysis, design, efficiency, implementation, and application—that describe and transform information. The fundamental question underlying all of computing is, What can be (efficiently) automated. This discipline was born in the early 1940s with the joining together of algorithm theory, mathematical logic, and the invention of the stored-program electronic computer.

The roots of computing extend deeply into mathematics and engineering. Mathematics imparts analysis to the field, engineering imparts design. The discipline embraces its own theory, experimental method, and engineering, in contrast with most physical sciences, which are separate from the engineering disciplines that apply their findings ... The science and engineering are inseparable because of the fundamental interplay between the scientific and engineering paradigms within the discipline.

V často citované práci [Pa99] upozorňující na specifické rysy informatiky autoři zdůrazňují, že informatici nemohou být při kariérních postupech posuzováni podle kritérií obvyklých v jiných oborech. Poměrně podrobnou analýzu problematiky hodnocení informatického výzkumu představuje [Me09]. Dalšími pracemi, které se hodnocením informatického výzkumu zabývají, jsou [Di], [Ma02], [Ma08], [Mo07]. Uvedené práce se shodují v tom, že výzkum v informatice má specifické rysy a že systémy hodnocení

kvality výzkumu, které jsou použitelné v jiném oboru nebo oborech, zejména systémy založené na scientometrických ukazatelích, mohou být pro hodnocení inforatického výzkumu nevhodné.

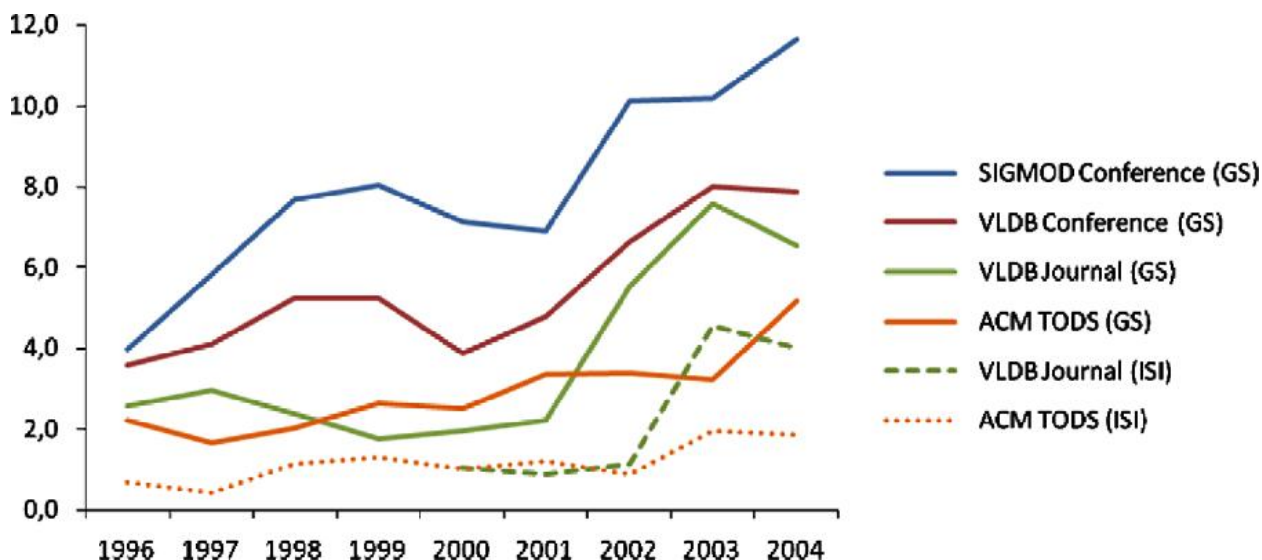
V citovaných pracích i v inforatické komunitě panuje shoda na tom, že důležitými specifickými rysy inforatického výzkumu jsou:

- Konference jsou v inforaticce významným publikačním fórem. Špičkové konference jsou vysoce selektivní (tzv. acceptance rate — procento přijatých příspěvků — se pohybuje kolem 10-25%) a jejich recenzní řízení se řídí stejnými standardy jako recenzní řízení špičkových časopisů. Publikace ve sbornících špičkových konferencí nejsou považovány za méně prestižní než publikace ve špičkových časopisech. V některých oblastech inforaticky jsou za nejméně významnější publikační fóra považovány časopisy, v některých oblastech jsou to sborníky konferencí. Je však třeba upozornit na to, že inforatických konferencí je velké množství a že mnohé z nich jsou nekvalitní.
- Bibliometrické údaje získávané z databází Web of Science společnosti Thomson Reuters jsou pro hodnocení inforatického výzkumu nevhodné, protože tyto databáze nepokrývají dobře fóra, kde je publikován inforatický výzkum.
- Inforaticka má široký záběr a je značně různorodá. Zahrnuje např. definování pojmů a dokazování teorémů, návrh algoritmů, návrh a tvorbu hardwarových a softwarových systémů, testování, experimenty. Jednotlivé disciplíny inforaticky se navzájem značně liší v metodách výzkumné práce i způsobu sdělování výsledků — některé disciplíny jsou v tomto směru blízké přírodním vědám a matematice, některé technickým vědám. Porovnávat scientometrické ukazatele napříč inforatickými disciplínami nelze.

Podívejme se na tyto rysy podrobněji.

### Konference jako významná publikační fóra v inforaticce

Význam konferencí jako významných publikačních fór v inforaticce doložíme příkladem. Následující graf, převzatý z [Ra08], ukazuje vývoj IF vybraných časopisů a konferencí podle dat z Web of Science (ISI) a Google Scholar (GS).



Poznamenejme, že graf porovnává IF špičkových inforatických konferencí a časopisů v oblasti databázových systémů, a to časopis ACM Transactions on Database Systems (ACM TODS), časopis The VLDB Journal, konferenci VLDB Conference a konferenci ACM SIGMOD (SIGMOD Conference). Je zřejmé,

že pokud budeme měřit význam publikačního fóra velikostí IF, vychází uvedená konference jako významnější fóra.

### **Databáze bibliometrických údajů**

V mnoha oborech jsou za vhodné databáze pro získávání bibliometrických údajů považovány databáze Web of Science společnosti Thomson Reuters. Jak bylo uvedeno výše, tyto databáze nepokrývají dobře fóra, na kterých je publikován inženýrský výzkum. Z pohledu informatiky je vhodné využívat kromě Web of Science i alternativní databáze, zejména následující:

- IEEE Xplore,
- ACM Digital Library,
- DBLP,
- INSPEC,
- The Collection of Computer Science Bibliographies,
- Google Scholar,
- Microsoft Academic Search (Libra)
- CiteSeer,
- Scirus.

Tyto databáze jsou odborníky v informatice běžně používány.

### **Různorodost informatiky, problém vymezení oblasti, kde lze porovnávat kvantitativní ukazatele**

Tento problém je v inženýrské komunitě znám a je diskutován např. v [My01]. Projevuje se například v systémech hodnocení, založených na posuzování kvality časopisů podle databáze Web of Science společnosti Thomson Reuters. Struktura databáze nabízí jednoduchou metodu hodnocení kvality časopisu podle tzv. impact factor (IF). IF je ukazatel, který vypovídá o citovanosti prací v daném časopise publikovaných (IF časopisu za rok 2008 je roven podílu  $C/N$ , kde  $C$  je počet citací prací, které vyšly v daném časopise v letech 2006 a 2007, a  $N$  je počet všech prací, které vyšly v daném časopise v letech 2006 a 2007; viz např.

[http://thomsonreuters.com/products\\_services/science/free/essays/impact\\_factor/](http://thomsonreuters.com/products_services/science/free/essays/impact_factor/)). Samotná hodnota IF se pro hodnocení kvality časopisu použít nedá, protože IF časopisů různých oborů mají různě vysoké hodnoty (IF nejlepších časopisů o fyzice jsou např. nižší než IF nejlepších časopisů o medicíně, absolutní porovnávání IF je tedy „míchání hrušek s jablky“). Doporučuje se tedy porovnávat pouze IF časopisů ze stejného oboru. Otázkou však je, jak definovat obor, v rámci kterého lze IF smysluplně porovnávat. Databáze Web of Science tento problém řeší zavedením tzv. subject categories. Subject categories představují jakési obory, resp. podobory, které jsou definovány tím, jaké časopisy obsahují. Subject categories nejsou disjunktní, tj. daný časopis může patřit do jedné nebo více subject categories. Web of Science umožňuje setřídít časopisy v rámci jedné subject category podle jejich IF. Výsledný seznam lze interpretovat jako žebříček kvality časopisů dané oblasti (subject category). Pořadí časopisu v tomto žebříčku, popř. normované pořadí (dané absolutním pořadím a počtem časopisů v žebříčku), lze chápat jako míru kvality časopisu, pomocí které lze mezioborově srovnávat: Který ze dvou časopisů je kvalitnější, fyzikální časopis C1 nebo chemický časopis C2? Návod popsáný výše říká: „Urči normované pořadí C1 v subject category, do které patří, a normované pořadí C2 v subject category, do které patří, a porovnej tato pořadí. Časopis s nižším pořadím je kvalitnější.“ Pokud patří časopis do více subject categories, je možné zvolit např. průměrné pořadí přes všechny subject categories, do kterých patří. (Metoda vycházející z této myšlenky je použita ve metodice hodnocení výsledků výzkumu a vývoje v ČR, která je diskutována níže.)

Problém nastane, pokud je daná subject category nevhodně definována, např. pokud je příliš široká a zahrnuje časopisy, ve kterých publikují komunity s různými publikačními zvyklostmi, různou obvyklou

dobou recenze apod. Extrémním případem by byla subject category „přírodní vědy“, v rámci které by pak docházelo k „míchání hrušek s jablky“. K takovému míchání ale může docházet v menší míře i v případě menších subject categories, zejm. pokud obor, pro který byly subject categories vytvořeny, je vnitřně různorodý, široký a jeho členění není ustáleno. Informatika takovým oborem je. Podívejme se proto na subject categories relevantní oboru informatika a na to, jaké časopisy se v rámci těchto subject categories porovnávají.

Pro obor informatika jsou relevantní následující subject categories:

subject category	počet časopisů v subject category
Computer Science, Artificial Intelligence	94
Computer Science, Cybernetics	17
Computer Science, Hardware & Architecture	45
Computer Science, Information Systems	99
Computer Science, Interdisciplinary Applications	94
Computer Science, Software Engineering	86
Computer Science, Theory & Methods	84

Uvedme dva příklady. Podívejme se na subject category Computer Science, Interdisciplinary Applications a na některé časopisy, které obsahuje. J. Combinatorial Optimization (IF 0.701) je na 74. místě z 94 časopisů této subject category. V hodnocení založeném na pořadí v subject categories tedy časopis J. Combinatorial Optimization vychází jako výrazně podprůměrný časopis. Přitom jde o prestižní časopis a tento časopis není zařazen do žádné jiné subject category. Problém je, že v dané subject category je tento časopis porovnáván s celou řadou časopisů, se kterou by se porovnávat neměl. Uvedme několik z těch, které se v dané subject category umístily nad J. Combinatorial Optimization:

vybrané časopisy subject category Computer Science, Interdisciplinary Applications	pořadí časopisu v této subject category podle IF
J. Chemical Information and Modeling	2
Scientometrics	13
SAR & QSAR in Environmental Research	15
Computers & Education	17
Computer Physics Communication	18
J. Molecular Modeling	19
Computers and Chemical Engineering	28
Computers and Electronics in Agriculture	36
Computers in Biology and Medicine	37
Speech Communication	38
Computer Methods and Programs in Biomedicine	41

Computers & Geosciences	42
Engineering with Computers	44
Electronic Commerce Research and Applications	45
CIS-Computers Informatics Nursing	59
Medical Informatics and the Internet in Medicine	65
Computer Music J.	67
Computer Aided Civil and Infrastructure Engineering	70
Int. J. Modern Physics C	71
Social Science Computer Review	73
<b>J. Combinatorial Optimization</b>	<b>74</b>

V druhém příkladě se podívejme na časopis Journal of ACM, který bývá považován za jeden z nejvýznamnějších v oboru informatika. Patří do následujících subject categories:

subject category	počet časopisů v subject category	pořadí J. ACM při sestupném uspořádání podle IF
Computer Science, Information Systems	99	19
Computer Science, Hardware & Architecture	45	10
Computer Science, Software Engineering	86	9
Computer Science, Theory & Methods	84	12

Např. v kategorii Computer Science, Information Systems je na 4. místě časopis J. of Chemical Information and Modeling a na 5. místě časopis J. of American Medical Informatics Association, které by s časopisem J. ACM pravděpodobně nikdo nesrovnával. Na 2. místě v této subject category je časopis MIS Quarterly, časopis zaměřený zejména na problémy managementu IT, který je obsahově i svou čtenářskou obcí značně odlišný od J. ACM.

Je třeba dodat, že s diskutovaným problémem se potýkají i jiné obory, zejména ty, které jsou na hranici klasických oborů, jako např. biochemie nebo biofyzika. Otázkou, jejíž zodpovězení není snadné, je, zda informatiku problém nevhodně zvolených subject categories postihuje přibližně stejně nebo více než jiné obory.

Dalším aspektem, který musíme brát při posuzování nedostatků hodnocení založeného na posuzování kvality časopisu podle jeho pořadí v nevhodně zvolených subject categories, je úroveň, na které je hodnocení použito. Pokud budeme hodnotit jedince, jde jistě o závažný nedostatek systému hodnocení. Pokud budeme hodnotit větší jednotky, tento nedostatek se nemusí jevit jako závažný. Při hodnocení velkých institucí, např. vysokých škol, se pak tento nedostatek může jevit jako nepatrný, jako daň za případné výhody daného systému hodnocení.

### 3. Dva přístupy k hodnocení výzkumu, jejich výhody a nevýhody

V principu lze uvažovat dva systémy hodnocení výsledků výzkumu a jejich kombinaci. Prvním je systém hodnocení komisí odborníků. Druhým je automatický bodový systém hodnocení. Každý systém hodnocení by měl splňovat dva základní požadavky. Zaprvé, systém musí být realizovatelný, tak aby podle něho bylo vůbec možné hodnotit. Zadruhé, systém musí být dostatečně přesný, tj. musí dávat výsledky odpovídající skutečnosti. Systém hodnocení komisí třiceti špičkových odborníků může být vhodný pro výběr kandidáta na udělení prestižního ocenění udělovaného jednou za rok, ale bude pravděpodobně zcela nevhodný pro posuzování kvality výzkumu čtyřiceti univerzit, protože posoudit čtyřicet univerzit je nad možnosti každé takové komise. Systém, který bude hodnotit tak, že bodové hodnocení vědce bude rovno počtu jeho publikací je snadno realizovatelný, ale je zcela nepřesný, protože počet publikací o kvalitě nevyovídá nic.

#### Systém hodnocení komisí odborníků

Tento systém je používán zejména pro hodnocení jednotlivců (habilitační komise, komise pro tenure apod.). Jeho výhodou je přesnost. Odborníci dokážou vnímat různé aspekty kvality výzkumu, např. ty, na které bodový systém nepamatuje a zůstává k nim slepý. Mezi nevýhody tohoto systému patří:

- možná zaujatost při hodnocení, prostor pro prosazování vlastních zájmů,
- náročnost (komise se musí sejít a hodnotit „ručně“).

Obě nevýhody se projeví zejména při posuzování na vyšší úrovni, například při posuzování výzkumu institucí. Takového posuzování by se muselo kvůli časové náročnosti účastnit více komisí, což s sebou přináší problém „kalibrace“ hodnocení: jak zajistit, aby různé komise měřily stejným metrem, tj. uplatňovaly stejná kritéria, navíc stejným způsobem?

#### Automatický bodový systém hodnocení

Systém přiřadí hodnocenému objektu za každý jeho výsledek výzkumu jistý počet bodů podle tabulky. Součet těchto bodů představuje bodové ohodnocení hodnoceného objektu. Příkladem tohoto systému je metodika hodnocení výsledků výzkumu a vývoje [Met09]. Výhody takového systému jsou:

- objektivita (na každý hodnocený objekt jsou uplatňována stejná kritéria, a to stejným způsobem),
- snadno realizovatelný, a tedy levný, za předpokladu, že seznam výsledků je k dispozici ve vhodné formě (např. v databázi).

Hlavní nevýhody takového systému jsou:

- nepřesnost,
- vede k chování popsanému Campbellovým, popř. Goodhartovým zákonem.

Systém je nepřesný zejména proto, že scientometrické údaje, používané v takových systémech, jsou schopny postihnout kvalitu výzkumu pouze orientačně. Jsou schopny „odlišit zrna od plev“, ale nejsou schopny jemnějšího porovnávání. Poznamenejme však, že [Re07] popisuje systém automatického hodnocení jehož výsledky byly v souladu s výsledky manuálním hodnocením.

Možná závažnějším nedostatkem je jev popsaný Campbellovým, popř. Goodhartovým zákonem.

Campbellův zákon ([http://en.wikipedia.org/wiki/Campbell%27s\\_Law](http://en.wikipedia.org/wiki/Campbell%27s_Law)):

"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

Goodhartův zákon ([http://en.wikipedia.org/wiki/Goodhart%27s\\_law](http://en.wikipedia.org/wiki/Goodhart%27s_law)):

„Once a social or economic indicator or other surrogate measure is made a target for the purpose of conducting social or economic policy, then it will lose the information content that would qualify it to play such a role.“

K jevu dochází stručně řečeno proto, že lidé optimalizují své chování vzhledem k automatickému bodovému systému. Systém sám tedy nejen měří, ale i ovlivňuje výsledky výzkumu. Některé negativní projevy tohoto jevu (bez uvedení, že jde o projevy jednoho jevu) jsou např. zmiňovány ve známé kritice automatických bodových systémů [Pa07]. Parnas konkrétně zmiňuje, že takový systém podporuje povrchní výzkum, podporuje opakované uvádění výsledků v publikacích, podporuje publikování částečných a nehotových výsledků, podporuje štěpení prací do tzv. MPI (minimum publishable increment), podporuje vytváření podezřelých workshopů, lidé jezdí na konference přednášet, a ne poslouchat, hlavní snahou vědce se stává „udat publikaci“ apod.

Uvedme nyní a na příkladu doložme další projev diskutovaného jevu. Jde o ovlivňování hodnoty IF časopisů. V podmínkách, kdy IF časopisu má významný ekonomický vliv, lze očekávat, že různé zájmové skupiny se budou snažit IF „svého časopisu“ ovlivnit. Takovými zájmovými skupinami mohou být

- vydavatel časopisu, pro kterého je vyšší IF předpokladem pro to, že o časopis bude vyšší zájem a vydavatel tak zvýší zisk z prodeje časopisu a zlepší své jméno na trhu časopisů,
- editor (šéfredaktor), pro kterého může být motivací odborná prestiž (být v čele významného časopisu) i finanční ohodnocení (vydavatel platí lépe editora úspěšného časopisu),
- autoři, kteří do časopisu pravidelně přispívají, pro které je motivací výhoda plynoucí z toho, že publikují v časopise s vyšším IF.

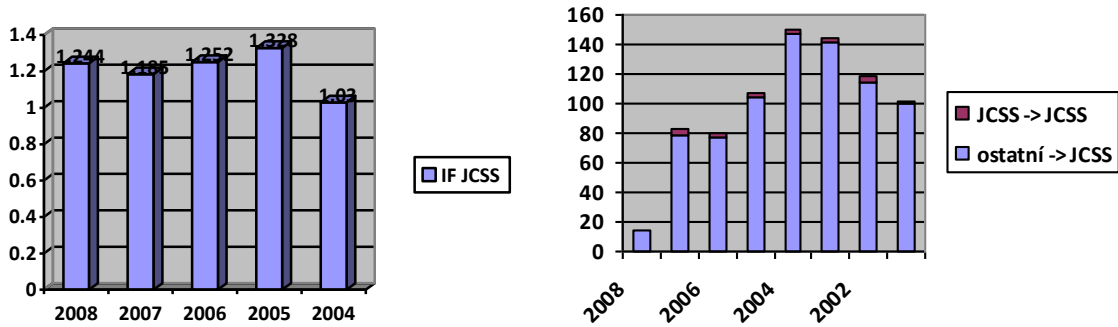
Dnes je veřejným tajemstvím, že někteří vydavatelé časopisů uvádějí ve smlouvách s editory podmínky, podle kterých výše odměny editora závisí i na hodnotě IF daného časopisu. Vydavatelé navíc vybízejí editory, aby přijímali opatření, která zvýší IF časopisu. Příkladem je uzavřít dohodu s editorem oborově příbuzného časopisu a doporučovat autorům jednoho časopisu, aby citovali práce, které vyšly v posledních dvou letech v druhém časopise, a naopak. Dalším příkladem je vyzývat autory článků zaslaných do daného časopisu, aby citovali práce daného časopisu („Článek může být přijat, pokud jako autor splníte následující podmínky ...“, mezi podmínkami je „Ukázat, že práce je obsahem relevantní tomuto časopisu, nejlépe uvedením citací alespoň dvou prací, které v minulých dvou letech v tomto časopise vyšly“).

Lze předpokládat, že výše popsané umělé zvyšování IF časopisu se bude dít zejména u časopisů nižší kvality (vysoký odborný kredit editora prestižního časopisu působí jako pojistka). Do jisté míry lze popsané manipulace s IF, přesněji řečeno podezření na tyto manipulace, zjistit z databáze Web of Science. Pro daný časopis lze totiž zobrazit vývoj jeho IF a také strukturu citací (lze zjistit procento celkového počtu citací, které práce v daném časopise získaly díky citacím uvedeným ve stejném časopise). Poznamenejme však, že IF informatických časopisů v posledních letech obecně rostly, např. díky zkráceným dobám recenzních řízení. Odhalování zmíněných jevů z dat, bez znalosti konkrétního případu, tedy není snadné.

Uvedme několik příkladů (data převzata z Web od Science, 2010). Graf vlevo ukazuje vývoj IF daného časopisu. Graf vpravo ukazuje celkový počet citací prací z daného časopisu rozdělený na autocitace (vrchní část sloupce, tj. citující práce vyšla také v daném časopise) a ostatní citace (spodní část sloupce, tj. citující práce vyšla v jiném časopise).

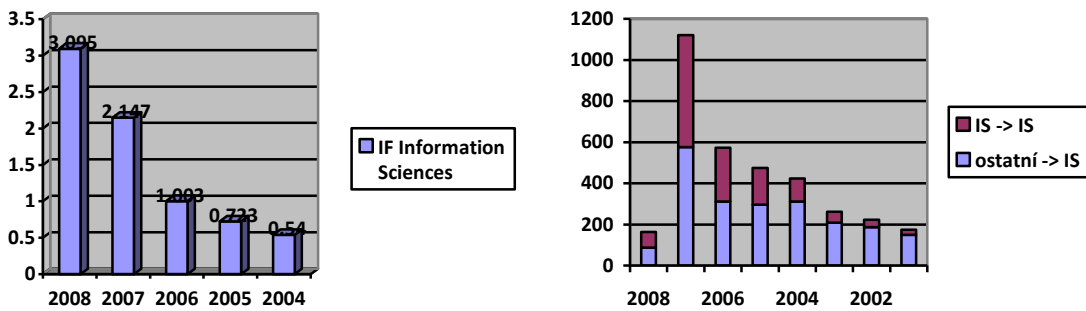


Journal of Computer and System Sciences (Elsevier)



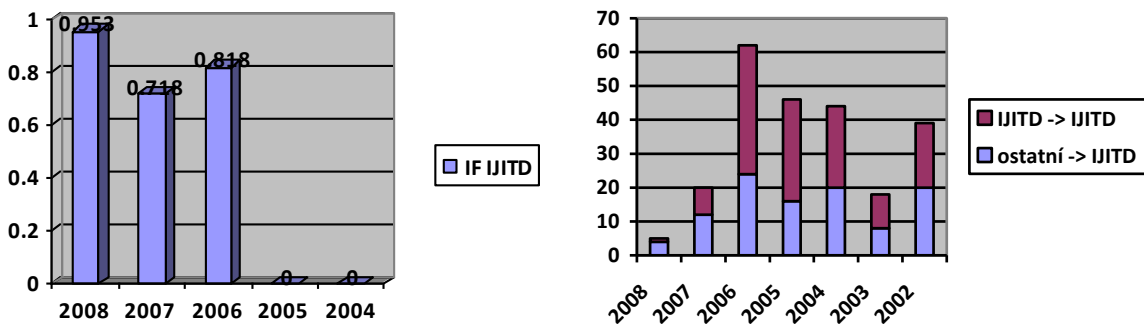
Časopis J. Computer and System Sciences je tradiční časopis s kvalitně obsazenou redakční radou, který považován za prestižní. Jeho IF je v čase stabilní a ze grafu struktury citací je vidět, že většina prací z časopisu je citována jinými časopisy.

Information Sciences (Elsevier)



Z grafu vývoje IF časopisu Information Sciences je patrný výrazný nárůst v několika posledních letech. Z grafu struktury citací lze vyčíst, že tento nárůst je zčásti způsoben tím, že výrazně přibýlo citací pracemi publikovanými v tomto časopise. To je dáno tím, že editor vybízí autory prací v recenzním řízení, aby citovali práce, které v tomto časopise vyšly.

Int. J. Information Technology & Decision Making (World Scientific)



Časopis Int. J. Inf. Technology & Decision Making je novým časopisem, u kterého je z grafu patrné, že většina citací tohoto časopisu je provedena pracemi publikovanými také v tomto časopise. Zda je to náhoda nebo výsledek politiky redakční rady, není autorovi této studie známo.

#### 4. Metodika hodnocení výsledků výzkumu a vývoje v České republice

Pro hodnocení výsledků výzkumu a vývoje v České republice je v současné době používána Metodika hodnocení výsledků výzkumu a vývoje. Tato metodika se postupem času vyvíjí, její podoba pro hodnocení v roce 2009 je [Met09]. Metodika je dílem Komise pro hodnocení výsledků výzkumu a vývoje a dalších poradních orgánů Rady pro výzkum, vývoj a inovace. V současné době je používána pro stanovení výše institucionální podpory výzkumným organizacím v České republice. Protože hodnocení výsledků výzkumu a vývoje je citlivé téma a protože se jedná o přidělování značného objemu finančních prostředků ze státního rozpočtu, stala se metodika předmětem zájmu médií v České republice. Tento zájem je umocněn bouřlivými debatami o její kvalitě. Pro ilustraci uvedme [Nov10]. Na metodiku existují dva navzájem protichůdné názory. První říká, že metodika je v principu správná, ale je třeba ji upravit. Druhý říká, že metodika je v principu špatná a je třeba ji zrušit. V této studii se k těmto názorům nevyjadřujeme. Uvedme jen, že zastánci upravené metodiky uvádějí na podporu svého názoru fakt, že metodika přináší do hodnocení jistou míru objektivitu a její nahrazení jiným systémem, např. hodnocením panelem expertů bude z mnoha důvodů problematické.

Základním mechanismem metodiky je přiřazení bodů dosaženým výsledkům výzkumu a vývoje. Metodika definuje několik typů takových výsledků a počty bodů, jimiž jsou jednotlivé typy výsledků ohodnoceny. Podíl výzkumné organizace na celkovém počtu bodů dosažených výzkumnými organizacemi v České republice určuje podíl této organizace na celkovém objemu finančních prostředků institucionální podpory přidělovaných výzkumným organizacím v České republice.

Hlavní výhrada k metodice [Met09], která byla prezentována na různých fórech, spočívá v tom, že metodika přiznává významný počet bodů tzv. „měkkým výsledkům“, tj. výsledkům, které mohou být uplatněny poměrně snadno (bez důkladného recenzního řízení, např. pouhou registrací popř. i bez jakékoli registrace). Tato skutečnost může vést k tomu, že výzkumné organizace začnou produkovat velké množství „měkkých výsledků“, pro jejichž produkci budou motivovány vidinou snadného zisku bodů. Na tento nedostatek jako jeden z prvních upozorňoval a příklady ho dokládal Tomáš Opatrný, viz např. [Op09]. Nedostatek je to zjevný a argumenty, které ho odhalují, jsou obecně platné.

Naším cílem je zabývat se tím, do jaké míry je metodika z pohledu oboru informatika dobře navržena v oblasti „tvrdých výsledků“, a to zejména publikací. Závěry provedené analýzy mohou být užitečné při provádění případné revize vládní metodiky, ale mohou být užitečné i při jiných příležitostech. Přestože byla metodika navržena pro hodnocení výkonu výzkumných organizací a přestože tvůrci metodiky zdůrazňují, že dělení získaných prostředků institucionální podpory v rámci jednotlivé výzkumné organizace je plně v kompetenci dané organizace, používá se z pochopitelných důvodů metoda dělení určená metodikou (popř. modifikace takové metody) vedoucími pracovníky výzkumných institucí i pro hodnocení menších jednotek (součástí výzkumných institucí, např. fakulty a katedry vysokých škol) a pro dělení prostředků mezi tyto menší jednotky. Závěry této studie mohou být užitečné při jednáních o systému hodnocení těchto menších jednotek.

## Výsledky a body dle vládní metodiky pro rok 2009

Následující tabulka uvádí uplatnitelné výsledky a jejich bodová ohodnocení podle vládní metodiky [Met09]. Vztahuje se na výsledky uplatněné od roku 2008 včetně (pro další informace odkazujeme na [Met09]).

	druh výsledku	body
J <sub>imp</sub>	článek v impaktovaném časopise	<b>10 až 305</b> (10+295*faktor)
J <sub>imp</sub>	článek v prestižním impaktovaném časopise (Nature, Science, Proc. Natl. Acad. Sci. USA)	<b>500</b>
J <sub>neimp</sub>	článek v recenzovaném časopise	<b>12</b> (časopis světově uznávané databáze, obory NRRE) <b>8</b> (časopis světově uznávané databáze, ostatní obory) <b>10</b> (časopis seznamu recenzovaných periodik, obory NRRE) <b>4</b> (časopis seznamu recenzovaných periodik, ostatní obory)
B	odborná kniha	<b>40</b> (obory NRRE, všechny jazyky) <b>40</b> (ostatní obory, světový jazyk) <b>20</b> (ostatní obory, ostatní jazyky jazyk)
C	kapitola v knize	<b>BB*FK</b> , kde BB je bodové hodnocení knihy jako výsledku B FK je faktor kapitoly (podíl počtu stran kapitoly a počtu stran knihy)
D	článek ve sborníku	<b>8</b>
P	patent	<b>500</b> (evropský nebo mezinárodní patent, USA, Japonsko) <b>200</b> (český nebo národní patent) <b>40</b> ostatní
Z	poloprovoz, ověřená technologie, odrůda, plemeno	<b>100</b>
F	užitný vzor, průmyslový vzor	<b>40</b>
G	prototyp, funkční vzorek	<b>40</b>
H	poskytovatelem realizované výsledky	<b>40</b>
N, L	certifikované metodiky a postupy, specializované mapy s odborným obsahem	<b>40</b>
R	software	<b>40</b>
V	výzkumná zpráva, která je výsledkem obsahujícím utajované informace	<b>50</b>

### Vysvětlivky:

- J<sub>imp</sub> jsou publikace v časopisech uváděných v databázích Web of Science společnosti Thomson Reuters (SCI-EXPANDED, SSCI, A&HCI, IC, CCR-EXPANDED).
- Koeficient faktor se při výpočtu bodového hodnocení J<sub>imp</sub> stanoví podle vzorce faktor = (1 - N) / (1 + (N / 0,057)), kde N je normované pořadí časopisu stanovené podle vzorce N = (P-1)/(P<sub>max</sub>-1), kde P je pořadí časopisu v daném oboru podle JCR (Journal Citation Report) v sestupném uspořádání dle impakt faktoru (IF), P<sub>max</sub> je celkový počet v daném oboru dle JCR.
- sborník u výsledku D musí být evidován v databázi Conference Proceedings Citation Index – Science nebo Social Science & Humanities společnosti Thomson Reuters

## **Příklady informatiků a jejich hodnocení podle vládní metodiky**

Podívejme se, jak dopadne hodnocení vládní metodikou, použijeme-li ji pro hodnocení existujících informatiků. Jde nám o ilustraci faktu, že výsledky hodnocení podle současné vládní metodiky skutečně mohou být výrazně v rozporu s tím, jak by hodnotil odborník. Doložení tohoto jevu na významnějším vzorku dat je nad rámec této studie. Bylo by užitečné takové doložení provést, lze však předpokládat, že závěr bude obdobný, jako závěr uvedený níže.

Jde nám o zodpovězení následujících otázek:

1. Je možné, že špičkový informatik světového jména dopadne v hodnocení vládní metodikou hůře než průměrný nebo podprůměrný informatik?
2. Je pravda, že při porovnání informatika s vědcem z jiného oboru exaktních věd dopadne, za předpokladu, že oba jsou „stejně dobří“, informatik hůře?

Pro výpočet bodů byly použity hodnoty z Web of Science z května 2010 a údaje z informačního systému Ministerstva školství, mládeže a tělovýchovy ČR. Dohledání údajů bylo v některých případech obtížné a výsledné počty bodů pravděpodobně nejsou zcela přesné. Pro naše účely, tj. pro ilustraci důsledků současné podoby vládní metodiky [Met09], je však přesnost dostatečná.

### **Otázka 1: špičkový vs. průměrný (podprůměrný) informatik**

Jedním z cílů vládních reforem v oblasti podpory vědy a výzkumu v České republice je podpora excelence. Podporou excelence se rozumí podpora špičkového výzkumu a utlumení podpory výzkumu průměrného a podprůměrného. Následující příklad ukazuje, že v oboru informatika současná podoba vládní metodiky excelenci nepodporuje, resp. má vážné nedostatky.

Vybrali jsme 3 špičkové vědce-informatiky (dva z oblasti zpracování a analýzy dat, jednoho logika) a jednoho informatika z České republiky, který publikuje na průměrně kvalitních fórech. (Pro náš argument bychom mohli provést jiný výběr, tento konkrétní výběr nehraje významnou roli.) Ohodnotili jsme výsledky těchto čtyř informatiků podle vládní metodiky [Met09] ve vybraném tříletém období (2003-2005 nebo 2006-2008, aby období bylo reprezentativní). Získané výsledky jsou následující.

Ronald Fagin

IBM Almaden Research Center

<http://www.almaden.ibm.com/cs/people/fagin/>

2006-2008 – hodnocené období 1

J<sub>imp</sub>: 199 b. (ACM Trans. Database Systems 2x 90b., SIAM J. Discrete Mathematics 19 b.)

D: 16 b. (ACM PODS 2x)

celkem: 215 b.

2003-2005 – hodnocené období 2

J<sub>imp</sub>: 167 b. (Theor. Comp. Sci. 18 b., ACM Trans. Database Systems 90 b., SIAM J. Discrete Mathematics 19 b., J. Computer and System Sciences 40 b.)

D: 48 b. (ACM PODS 3x, VLDB 1x, WWW Conference 1x, ACM SIGMOD 1x)

celkem: 215 b.

Johannes Gehrke

Cornell University, NY, USA

<http://www.cs.cornell.edu/johannes/>

2006-2008

J<sub>imp</sub>: 27 b. (Data Knowl. Eng. 27 b.)

D: 248 b. (ICDE 8x, SIGMOD 11x, VLDB 5x, CIDR 1x, ICDM 2x, ACM PODS 2x, WWW 1x, EDBT 1x,)

celkem: 275 b.

Gerhard Jäger

Universitaet Bern, Švýcarsko

<http://www.iam.unibe.ch/til/staff/jaeger>

2003-2005 – hodnocené období

J<sub>imp</sub>: 64 b. (Math. Logic. Quarterly 8.55, Annals of Pure and Applied Logic 2x 28 b.)

D: 0 b.

celkem: 64 b.

Informatik z České republiky (skutečná osoba, kvůli diskrétnosti neuvádíme konkrétní údaje)

2006-2008 – hodnocené období

J<sub>imp</sub>: 21 b. (podprůměrné časopisy)

D: 408 b. (51 příspěvků ve sbornících konferencí, naprostá většina průměrných nebo podprůměrných)

celkem: 429 b.

Všichni tři zahraniční odborníci dopadli výrazně hůře než informatik z ČR s podprůměrnými publikacemi. Je zřejmé, že hlavní příčinou je výše zmíněné nevhodné bodové ohodnocení článků ve sbornících konferencí. Informatik mnoha průměrnými publikacemi tak získal více bodů než J. Gehrke, který publikoval zcela výjimečně vysoký počet článků ve sbornících nejprestižnějších konferencí (ACM SIGMOD, ACM PODS, VLDB, ICDM, ICDE apod.), jejichž míra přijetí (acceptance rate) se pohybuje mezi 10% až 25%. Body za publikace ve sbornících konferencí tedy musí zohledňovat kvalitu konferencí.

### **Otázka 2: informatik vs. vědec z jiného oboru exaktních věd**

Vybrali jsme špičkového fyzika z ČR a získali následující bodové ohodnocení jeho výsledků.

Fyzik z České republiky (skutečná osoba, kvůli diskrétnosti neuvádíme konkrétní údaje)

2006-2008 – hodnocené období

J<sub>imp</sub>: 1655 b. (Phys. Rev. Lett. 5x 98 b., Phys. Rev. A 10x 95 b., Quantum Information & Computation 1x 88 b., Nature Physics 127 b.)

D: 16 b. (2)

celkem: 1671 b.

Fyzik z ČR získal mnohem vyšší počet bodů než každý ze tří informatiků zmíněných výše. Je jistě obtížné posuzovat, zda tento fyzik je ve svém oboru lepší než jsou ve svém oboru uvedení informatici. Získal však zhruba 6krát více bodů než J. Gehrke, a to považujeme za zcela neadekvátní. Opět je zřejmé, že na vině je nevhodné bodové hodnocení článků ve sbornících konferencí.

### **Další otázky**

Jak bylo zmíněno výše a jak ukázaly uvedené příklady, hodnocení článků ve sbornících konferencí stejným počtem bodů bez ohledu na to, o jakou konferenci se jedná, je pro obor informatika nevhodné a zkreslující. Zdá se, že pokud jde o hodnocení „tvrdých výsledků“, zejm. publikací, je tento nedostatek z pohledu oboru informatika hlavním nedostatkem vládní metodiky.

Další otázkou, jejíž pečlivé posouzení a prozkoumání na příkladech přesahuje rámec této studie, je otázka, zda rozdělení časopisů do tzv. subject categories ve Web of Science nezvýhodňuje nebo nediskriminuje některé oblasti informatiky. Jde o to, že body přidělené článku v časopise podle vládní metodiky záleží na průměrném normovaném pořadí časopisu ve všech subject categories, do kterých je daný časopis zařazen. Protože stanovení normovaného pořadí vychází z uspořádání časopisů podle IF v jednotlivých subject categories, nedostatek vládní metodiky z pohledu oboru informatika vyplývá z faktu, že subject categories jsou nevhodně navrženy, jak bylo ukázáno v kapitole 2.

## 5. Závěry

Závěry, ke kterým jsme v této studii dospěli, lze shrnout do následujících bodů.

1. Informatika je samostatná disciplína s širokým záběrem. Informatický výzkum má rysy výzkumů v přírodních i technických vědách. Hodnocení výzkumu v informatice musí brát v potaz specifické rysy této disciplíny.
2. Specifickým rysem informatiky je význam publikací ve sbornících konferencí. Špičkové informatické konference jsou vysoce selektivní a publikace v nich jsou recenzovány stejně důkladně jako publikace ve špičkových časopisech. Publikace ve sbornících špičkových informatických konferencí jsou považovány za stejně prestižní jako publikace ve špičkových časopisech.
3. Povaha výzkumu v jednotlivých informatických disciplínách je různorodá. Zahrnuje např. definování pojmů a dokazování teorémů, návrh algoritmů, návrh a tvorbu hardwarových a softwarových systémů, testování, experimenty. Porovnávat scientometrické ukazatele napříč informatickými disciplínami nelze.
4. Hodnocení kvality výzkumu v informatice pomocí scientometrických ukazatelů je možné, ale musí brát v potaz specifické rysy této disciplíny. Pokud je to možné, mělo by součástí hodnocení kvality výzkumu v informatice být hodnocení komisí odborníků.
5. Metodika hodnocení výsledků výzkumu a vývoje v České republice má z pohledu oboru informatika, i pokud jde jen o výsledky typu publikace, vážné nedostatky. Hlavní nedostatek je fakt, že bodové hodnocení publikací ve sbornících konferencí naprosto neodpovídá významu těchto publikací.
6. V posledních letech je problematika hodnocení informatického výzkumu tématem diskutovaným v mnoha zemích. Pro další debaty o hodnocení informatického výzkumu v České republice bude užitečné tyto debaty sledovat, zapojit se do nich a vyměňovat si zkušenosti.

## 6. Náměty pro pokračování této studie

1. Vytvořit stručný dokument analyzující současnou vládní metodiku hodnocení výsledků výzkumu a vývoje z pohledu oboru informatika. Dokument by měl poukazovat na nedostatky a navrhnout řešení.
2. Podrobněji zmapovat systémy hodnocení informatického výzkumu používaných ve vyspělých zemích, zejm. v Evropě, a to na různých úrovních hodnocení. Autor této studie kontaktoval kolegy z Austrálie, Belgie, Francie, Polska, Německa, Rakouska, Ruska a Španělska a získal různé podklady. Jejich zpracování je nad rámec této studie.
3. Vytvořit dokument analyzující, jak s použitím dostupných informačních zdrojů zajistit přiměřenější hodnocení publikací ve sbornících informatických konferencí.

4. Upřesnit, popř. rozšířit seznam příkladů informatiků, na kterých je možné testovat vhodnost systému hodnocení informatického výzkumu, zejm. systému založeného na scientometrických údajích.

## Reference

- [De89] Denning P. J., Comer D. E., Gries D., Mulder M. C., Tucker A., Turner A. J., and Young P. R.: Computing as a discipline. *Comm. ACM* 32 (1)(1989), 9-23.
- [Am00] Amin M., Mabe M.: Impact factors: Use and abuse. *Perspectives in Publishing*, October 2000, No. 1, pp. 1-6.
- [De89] Denning P. J., Comer D. E., Gries D., Mulder M. C., Tucker A., Turner A. J., and Young P. R.: Computing as a discipline. *Comm. ACM* 32 (1)(1989), 9-23.
- [De10] De Baets B.: Personal communication, May 2010.
- [Di] Diaz J.: Evaluation in Computer Science. <http://www-lsi.upc.es/~diaz/geneve.pdf>,
- [Ge96] Geist R., Chetuparambil M., Hedetniemi S., Turner A. J.: Computing research programs in the U.S. *Comm. ACM* 39(12)(1996), 96-99.
- [Gl04] Glass R. L., Ramesh V., Vessey I.: An analysis of research in computing disciplines. *Comm. ACM* 47(6)(2004), 89-94.
- [Ku10] Kuznetsov S.: Personal communication, May 2010.
- [Let08] Letter to CRUS regarding Bibliometric Evaluation, 2008, <http://www.sarit.ch/crusletter/index.html>.
- [Ma02] Mattern F.: Zur Evaluation der Informatik mittels bibliometrischer Analyse. *Informatik-Spektrum*, Vol. 25, No. 1, pp. 22-32, 2002.
- [Ma08] Mattern F.: Bibliometric Evaluation of Computer Science – Problems and Pitfalls. *European Computer Science Summit – ECSS 2008*, 9-10 Oct. 2008, Zurich.
- [Met09] Metodika hodnocení výsledků výzkumu a vývoje v roce 2009, Úřad vlády ČR, č.j. 08724/09-RVV, [http://www.vyzkum.cz/storage/att/CDDC542199F1640B59A7D1E841B7151C/Metodika%202009\\_schvaleno.pdf](http://www.vyzkum.cz/storage/att/CDDC542199F1640B59A7D1E841B7151C/Metodika%202009_schvaleno.pdf)
- [Me09] Meyer B., Choppy C., Staunstrup J., van Leeuwen J.: Research evaluation for computer science. *Comm. ACM* 52(4)(2009), 31-34. Nezkrácená verze dostupná jako dokument komise Research Evaluation Committee of Informatics Europe.
- [Mo07] Moed H. F., Visser M. S.: Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study. *Research Report to the Council for Physical Sciences of the Netherlands Organisation for Scientific Research (NWO)*, CWTS Report 2007-01, xvi+101 stran.
- [My01] Mylonopoulos N. A., Theoharakis V.: Global perceptions of IS journals. *Comm. ACM* 44(9)(2001), 29-33.
- [Nov10] Vybrané novinové články na téma financování výzkumu a vývoje v ČR z roku 2010: [http://www.lidovky.cz/zlepsovaky-berou-penize-spickove-vede-dy0-/ln\\_veda.asp?c=A100215\\_132133\\_ln\\_veda\\_lv](http://www.lidovky.cz/zlepsovaky-berou-penize-spickove-vede-dy0-/ln_veda.asp?c=A100215_132133_ln_veda_lv), <http://www.lidovky.cz/prezije-ceska-veda-kafemlejnek-dz0->

[/ln\\_noviny.asp?c=A100220\\_000095\\_In\\_noviny\\_sko&klic=235624&mes=100220\\_0](#),  
<http://www.lidovky.cz/vylepsime-kafemlejnek-slibuji-radni-d79->  
[/ln\\_noviny.asp?c=A100525\\_000102\\_In\\_noviny\\_sko&klic=237142&mes=100525\\_0](#)

[Op09] Opatrný T.: Financování aplikovaného výzkumu, aneb jak drůbež elektrárnou omračovati.  
[http://Blog.iHNed.cz/c3-38997070-YOpatr\\_d-financovani-aplikovaneho-vyzkumu-aneb-jak-drubez-elektrarnou-omracovati](http://Blog.iHNed.cz/c3-38997070-YOpatr_d-financovani-aplikovaneho-vyzkumu-aneb-jak-drubez-elektrarnou-omracovati)

[Pa07] Parnas D. L.: Stop the numbers game. Comm. ACM 50(11)(2007), 19-21.

[Pa99] Patterson D., Snyder L., Ullman J.: Evaluating Computer Scientists and Engineers For Promotion and Tenure. Computing Research news, Sep. 1999, pp. A-B.

[Re07] Ren J., Taylor R. N.: Automatic and versatile publications rating for research institutions and scholars. Comm. ACM 50(6)(2007), 81-85.

[Ti95] Tichy W. et al.: Experimental evaluation in computer science: a quantitative study. J. Systems and Software 28(1)(1995), 9-18.

[VZ] <http://vedazije.cz/node/1432>

Autor:

prof. RNDr. Radim Bělohávek, DSc.

Katedra informatiky

Přírodovědecká fakulta

Univerzita Palackého v Olomouci

17. listopadu 12

771 46 Olomouc

e-mail: [radim.belohlavek@upol.cz](mailto:radim.belohlavek@upol.cz)

web: <http://belohlavek.inf.upol.cz/>