

Hodnocení výzkumu v informatice – přístupy a problémy

Jiří Zlatuška

Fakulta informatiky Masarykovy univerzity

Abstrakt

Obecné metodologie hodnocení výzkumu a výzkumníků v informatice mají oproti starším a zavedenějším vědním disciplinám zvláštní rysy projevující se ve zvláštní úloze konferenčních příspěvků pro zveřejňování nových poznatků, ale i specifické oborové rysy výzkumu, který není čistě teoretickou disciplinou. Tento příspěvek shrnuje stávající prezentace těchto odlišností a metodologická východiska hodnocení, která pocházejí od amerického National Research Council, společnosti Computing Research Association a evropské Informatics Europe. V záběru jsou zmíněna větší dokumentovaná hodnocení výzkumu, která mohou sloužit jako případové studie praxe uplatněné praxe v této oblasti.

Obecná východiska

Informatiku lze jako vědní disciplínu nahlížet z jedné strany jako disciplínu přirozeně vyrůstající z lidského poznání, která si před sebe staví fundamentální vědecké a metodologické problémy (viz např. [Gruska2010]). Stejně se ovšem jedná o disciplínu, která je relativně mladá a vyznačující se velmi specifickými rysy vzhledem k ostatním „tradičním“ disciplinám a výrazně se od nich liší rychlostí přenosu nových poznatků do aplikací, vztahem základního a aplikovaného výzkumu i okolností, že lze řadu oblastí informatiky nahlížet jako disciplínu technickou nebo aplikovanou, jejíž výsledky se podstatným způsobem objevují jako podpůrná součást výzkumu v jiných disciplinách [IE2008].

Rychlost změn a mladost disciplíny vedla specificky k využívání prezentací na konferencích a článků v konferenčních sbornících jako médiu vědecké komunikace, které má ve srovnání s časopiseckými publikacemi větší důležitost, než je tomu v jiných disciplinách (kvantitativní podklady pro toto konstatování lze nalézt např. v [Sandstrom2007], [Moed2005], [Mattern2008]). Srovnání bibliometrických měřítek kvality publikací ve srovnání s časopisy problematické ([CoyleFreyneSmythCunningham2010], [Diaz2008]), z dalších rozdílů lze upozornit na lepší možnosti pro prezentace interdisciplinárních prací na konferencích než v tradičních časopisech [HapkernParkes2011].

Kritika bibliometrických měřítek založených na komerčních bibliometrických datech (Thomson Ruters resp. Scopus) je předmětem analýzy Mezinárodní unie matematiků [IMU2008]. Zdůrazněna je zde skutečnost, že citační statistiky nepřestávají vyšší míru přesnosti, pokud jsou nesprávně použity, a při špatném užití nebo zneužití mohou naopak mást. Objektivita citačních statistik může být v důsledku špatného pochopení jejich významu iluzorní a význam citací velmi daleko od impaktu citovaných publikací. Jednoduchost jednotlivého čísla pro hodnocení kvality může vést k velmi mělkému pochopení výzkumu a čísla nejsou sama o sobě lepší, než opodstatněná úvaha. Citační statistiky lze považovat za část procesu hodnocení kvality, nemohou ho však v žádném případě nahradit („užijeme-li pouze impaktní faktor, je to jako posuzovat zdraví pouze podle váhy posuzované osoby“). V obecnější rovině existují i kvalifikovaná varování nad proliferací „amatérské bibliografie“ vycházející z politických požadavků hodnocení jedním číslem, lákavosti zdánlivé objektivity takového přístupu pro neodborné uživatele a zdánlivé legitimizace tohoto užití vědeckou praxí, která si jím zjednodušuje práci [GlaseLaudel2007].

Hodnocení v informatice z pohledu profesních asociací

Hodnocení kvality výzkumu je s ohledem na novost informatiky jako disciplíny i potřebě hodnotit její interdisciplinární aspekty považováno za potenciálně problematickou úlohu.

První z komplexnějších studií věnovaných specificky hodnocení informatiky, v tomto případě zejména s vazbou na experimentální informatiku, je studie americké Národní akademie věd [NAS1994]. Zdůrazňuje se zde zvláštní význam komutačních artefaktů v podobě hardwarových nebo softwarových systémů či prototypů, které umožňují studovat i koncepty, jejichž čistě teoretické uchopení by bylo s ohledem na jejich složitost nemožné (složitost reálného problému může vést k tomu, že jeho abstrakce na podobu teoreticky uchopitelnou může vést ke ztrátě esence obsažené v původním systému). Zdůrazněna je i rozle dostupných technologií a instalované infrastruktury pro vytváření nových konceptů, i vzájemné ovlivňování technické realizace s hodnotou duševního výtvoru, který s ní souvisí. Negativním faktorem může být napětí mezi výzkumníky v teoretické a experimentální oblasti, v důsledku čehož nemusí být mladí pracovníci hodnoceni adekvátně, jsou-li např. za práci v experimentální informatice posuzováni v prostředí vycházejícím z paradigmat teoretického oboru. Studie celá nabízí své závěry pro užití na pracovištích, kde dochází k hodnocení pracovníků v oblasti experimentální informatiky.

Na tuto analýzu navázal metodický materiál profesní společnosti Computing Research Association „Best Practices Memo“ [CRA1999] vycházející z problematického užití přístupů „publish or perish“ a důrazu na publikování v časopisech a ignorování průkazné nové hodnoty, kterou přinášejí experimentální a inženýrské přístupy v informatice. Jednostrannost hodnocení podle časopiseckých publikací autoři označují za škodlivou nejen pro kariérní postup konkrétních pracovníků, ale pro informatiku jako disciplínu, zdůrazněn je i sociální kontext „impaktu“ informatických výsledků daný skutečností, že informatické artefakty nejsou jevy existující samy o sobě v přírodě, i okolností, že se při posuzování impaktu nejde omezit jen na okolnost novosti výsledku, ale toho, zda je výsledek v nějakém smyslu „lepší“ než stávající znalosti. Důraz je kladen na okolnost, že zatímco standardní publikace jsou jedním z indikátorů akademického úspěchu, musí být brány v úvahu i jiné formy zveřejnění nových poznatků na konferencích a posuzování artefaktů jinými odborníky včetně uvažování numerických metrik typu počtu stažení softwaru, počet uživatelů, počet přístupů na webové stránky, atd.

Zatím nejúplnějším materiálem popisujícím požadavky na korektní hodnocení výzkumu v informatice je materiál „Research Evaluation for Computer Science“ [IE2008] vypracovaný společností Informatics Europe sdružující pracoviště zabývající se akademickým výzkumem v informatice. Shrnutí širší pracovní verze obsahuje deset konkrétních požadavků a doporučení pro korektní hodnocení informatiky:

1. Informatika je svébytnou disciplínou kombinující charakteristiky přírodních a technických věd. Hodnocení výzkumníků musí být přizpůsobeno těmto zvláštnostem informatiky.
2. Význačným rysem zveřejňování poznatků v informatice je důležitost konferencí, z nichž některé jsou vysoce selektivní, stejně jako knih. Časopisecké publikace, byť mohou být důležité pro hlubší pojednání o některých tématech, se nevyznačují větší prestiží než špičkové konference nebo knihy.
3. Důležitá část informatiky produkuje jako výsledky artefakty jiné než časopisecké publikace, zejména se to týká softwarových systémů. Při hodnocení impaktu mohou být tyto artefakty stejně důležité jako publikace.
4. Publikační kultura běžná v informatice nerozlišuje mezi pořadím autorů a to by proto nemělo být bráno v úvahu při hodnocení přínosu jednotlivého výzkumníka.

5. Počty publikací, ať již vážené nebo bez váhy, nesmějí být užívány jako indikátor hodnoty výzkumu. Jsou mírou produktivity, nikoli však impaktu nebo kvality.
6. Numerické hodnoty impaktu jako např. počty citací mají své místo, ale nesmějí být užity jako jediný zdroj hodnocení. Jakékoli použití těchto nástrojů musí být předmětem lidského filtru interpretace dat za účelem vyhnoutí se mnoha možným zdrojům chyb. Musí být doplněno peer review a hodnocením impaktu přínosů jiných než jen publikací.
7. Jakékoli evaluační kritérium, zejména pokud dává kvantitativní výsledek, musí být založeno na jasných a zveřejněných kritériích.
8. Numerické indikátory nesmí být užity pro srovnání výzkumníků v různých disciplínách.
9. Pro hodnocení publikací a citací je ISI Web of Science pro většinu oblastí informatiky neadekvátní a nesmí být užíván. Nedokonalými nicméně preferovanými alternativami jsou Google Scholar, CiteSeer a (potenciálně) ACM Digital Library.
10. Evaluační kritéria musí být sama o sobě předmětem zhodnocení a revize.

Neadekvátnosti ISI Web of Science se na konkrétních údajích věnuje [Mattern2008]. Nedostatečné rozlišení informatiky a komunikačních technologií, zpracování signálů nebo výpočtových věd způsobuje nepoužitelnost seznamu „250 nejcitovanějších výzkumníků v informatice“ podobně jako je tomu u „top 20 citovaných článků z informatiky“ v databázi Scopus (nevyhovujícímu záběru ve Scopus je rovněž věnována poznámka [Meyer2011] jednoho z autorů zprávy Informatics Europe), která doplňuje dřívější kritiku ISI Web of Science i o Scopus a podstatě vylučuje tyto dvě komerční databáze z použitelnosti pro hodnocení v informatice. Na obdobná faktor upozorňuje [Diaz2008] v souvislosti s prázdným průnikem mezi-dvaceti výzkumníky s nejvyšším hodnotami H-faktorem v informatice a nositeli Turingovy ceny.

Výše zmíněné přístupy k hodnocení vyzházejí vesměs z hodnocení vhodných pro posuzování jednotlivých informatiků, nikoli celých pracovišť. Informatics Europe deklarovala snahu vytvořit de facto agenturu pro posuzování kvality informatických pracovišť, tzv. Department Evaluation Initiative [IE2011]. Deklarovaným cílem je posuzování kvality informatických pracovišť z vícedimenzionálního pohledu, kde „výzkum“ není omezen jen vlastními výzkumnými výsledky, ale také managementem výzkumu, výzkumnou politikou a vybavením. Integrovanou složkou posuzování kvality je výchova PhD studentů a společenská

relevance prováděného výzkumu. Cílem této iniciativy je poskytnout nestrannou a průhlednou metodu posuzování výzkumných aktivit pracovišť prostřednictvím analýzy, která pomůže identifikovat silné stránky, příležitosti a potřebná zlepšení. Tento přístup by měl vyvážit praxi, ve které se posuzování svěřá agentuře, která nemá předpoklady posoudit informatiku jako specifickou oblast. Řídící výbor této iniciativy pracuje ve složení:

- Fausto Giunchiglia - University of Trento, Italy,
- Manuel Hermenegildo - Technical University of Madrid, Spain,
- Jeff Magee - Imperial College, UK,
- John Mylopoulos - University of Toronto, Canada,
- Manfred Nagl - Aachen University, Germany,
- Joseph Sifakis - CNRS, France,
- Letizia Tanca - Politecnico di Milano, Italy,
- Jan van Leeuwen - University of Utrecht, The Netherlands,
- Peter Widmayer - ETH Zurich, Switzerland..

Případové studie existujících hodnocení

Kromě výše zmíněných přístupů profesních společností se lze zmínit o existujících případech specifického posuzování infromatických pracovišť, kde dokumentace z příslušného procesu poskytuje dostatečný vhled do celého postupu.

Rozsáhle popsané je hodnocení finských pracovišť provedené Academy of Finland v roce 2007, viz [AF2007].

Hodnocení prováděné v rámci britského Research Assesment Excercise je pro RAE 2008 popsáno v dokumentu [RAECS2008] .

Přípravy nového systému hodnocení pro všechny obory ve Velké Británii, tzv, Research Excellence Framework (REF), jsou dosud jen v přípravné fázi, a není zde dopsud k dispozici závazně přijatá konečná podoba hodnocení, které se plánuje na rok 2014. Ke dni 5. října 2011 má být ukončen sběr připomínek k systému hodnocení pro jednotlivé subpanely, který byl zveřejněn 4. srpna 2011. Informatika tvoří subpanel 11 hlavního panelu B (viz [REF2011]). U citačních dat se předpokládá jejich využití u 12 subpanelů včetně informatiky z celkem 36, vesměs však jen jako podklad pro peer review, nikoli jeho náhradu (viz též bod 63 citovaného návrhu). Informatika si mezi zmíněnými 12 panely drží zvláštní postavení v tom, že se zde explicitně předpokládá užití Google Scholar (bod 62 citovaného návrhu), pro srovnání lze

uvést, že panel v matematice nebude citační data využívat vůbec (bod 64) . Obecné požadavky uvedené v [IE2008] tento návrh metodiky REF 2014 splňuje nepochybně lépe, než dosavadní aktualizace metodiky hodnocení výzkumu užívané v ČR.

Postup hodnocení informatiky na jedné univerzitě je obsažen v hodnocení z Univerzity v Eindhovenu [Eindhoven2009].

Příklad komplexního hodnocení pracovišť realizujících výzkumné doktorské programy je v publikaci [NRC1995], kde je hodnoceno 108 amerických univerzit na bázi publikační i reputační.

Oborová specifika se promítají i do postupů vytvářejících žebříčky pracovišť, Žebříček nejlepších světových pracovišť v informatice [QSR], jak ho sestavila společnost QS Quacquarelli Symonds, staví na metodice zohledňující zvláštnosti informatiky v různé skladbě ukazatelů, jak je popsáno v [QSM].

Odkazy:

1. **[AF2007]** Computer Science Research in Finland 2000-2006, Academy of Finland, 8/2007
<http://www.aka.fi/Tiedostot/Tiedostot/Julkaisut/8.07%20Computer%20Scienceverkko.pdf>
2. **[CRA1999]** David Patterson, Lawrence Snyder, Jeffrey Ullman: Best Practices Memo Evaluating Computer Scientists and Engineers For Promotion and Tenure, Computing Research News, September 1999, Special Insert, pp. A,B.
3. **[CoyleFreyneSmythCunningham2010]** Lorcan Coyle, Jill Freyne, Barry Smyth, and Pdraig Cunningham: Relative Status of Journal and Conference Publications in Computer Science, Communications of the ACM Vol. 53(11) , November 2010, 124-132
4. **[Diaz2008]** Josep Diaz: Evaluation in Computer Science, talk at U. Geneve, December 2008, <http://www.lsi.upc.edu/~diaz/geneve.pdf>
5. **[Eindhoven2009]** research Evaluation Computer Science 2009, University of Eindhoven, 2009 <http://www.win.tue.nl/cwb/oz-vis-inf-2009.pdf>
6. **[GlaserLaudel2007]** J. Gläser, G. Laudel: The Social Construction of Bibliometric Evaluations, in R. Whitley and J. Gläser (eds.). The Changing Governance of the Sciences, Springer Science+BusinessMedia D.V., 2007, pp. 101-127.

7. **[HapkpernParkes2011]** Joseph Y. Halpern, David C. Parkes: Journals for Certification, Conferences for Rapid Dissemination. Rethinking the role of journals in computer science. Communications of the ACM, Vol. 54 (8), 36-38
8. **[IE2008]** Bertrand Meyer, Christine Choppy, Jørgen Staunstrup, Jan van Leeuwen; Research Evaluation for Computer Science, Vol. 52(4), April 2009, 31-34
9. **[IE2011]** Informatics Europe Department Evaluation Initiative, http://www.informatics-europe.org/department_evaluation/
10. **[IMU2008]** The International Council for Industrial and Applied Mathematics, the Mathematical Union and the Institute for Mathematical Statistics: Citation Statistics, June 2008, <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>
11. **[Meyer2011]** Bertrand Meyer: Scopus's view of computer science research, July 2011, <http://informaticseurope.wordpress.com/2011/07/28/scopuss-view-of-computer-science-research/>
12. **[Mattern2008]** Friedmann Mattern: Bibliometric Evaluation of Computer Science – Problems and Pitfalls, European Computer Science Summit 2008 (ECSS 2008), <http://www.vs.inf.ethz.ch/publ/slides/Bibliometry-ECSS-Summit-08.pdf>
13. **[Moed2005]** Henk F. Moed: Citation Analysis in Research Evaluation, Springer 2005
14. **[NAS1994]** Committee on Academic Careers for Experimental Computer Scientists, National Research Council: Academic Careers for Experimental Computer Scientists and Engineers, National Academies Press, 1994, <http://www.nap.edu/catalog/2236.html>
15. **[NRC1995]** Marvin L. Goldberger, Brendan A. Maher, Pamela Ebert Flattau, eds.: Research-Doctorate Programs in the United States. Continuity and Change, NAP 1995. <http://www.nap.edu/readingroom.php?book=researchdoc>
16. **[Gruska2010]** A Perception of Informatics, Academia Europeana, April 2010, http://www.ae-info.org/attach/Acad_Main/Sections/Informatics/Informatics_and_Activities/10-05-05-Gruska-Informatics.pdf
17. **[QSM]** QS: QS World University Rankings, Subject Tables Methodology, Version 0.3, March 2011, <http://www.topuniversities.com/sites/www.topuniversities.com/files/articles/subject-rankings-methodology.pdf>
18. **[QSR]** QS: 2011 QS World University Rankings by Subject: Computer Science & Information Systems Rankings, <http://www.topuniversities.com/university-rankings/world-university-rankings/2011/subject-rankings/engineering/computer-science>

19. **[RAE2008]** HEFCE: UOA 23 Computer Science and Informatics, HEFCE, 2007, <http://www.rae.ac.uk/pubs/2006/01/docs/f23.pdf>
20. **[REF2014]** HEFCE: Draft for consultation: Main Panel B, HEFCE, srpen 2011, http://www.hefce.ac.uk/research/ref/pubs/2011/03_11/03_11b.pdf
21. **[Sandstrom2007]** Ulf Sandström: A Metric for Academic Research, WCU-2, Shanghai 2007, http://ed.sjtu.edu.cn/wcu-2/ppt/Shanghai_1nov2007.ppt