

Komentář Odborného panelu

Ford 1. 2. Computer and Information Sciences

Vypracoval: prof. RNDr. Antonín Kučera, PhD

Datum: 28. 6. 2021

Úvodní komentář

Počítačové vědy (computer science, dále CS) se od ostatních vědních oborů odlišují specifickou rolí vysoce výběrových konferencí, které v celosvětovém měřítku slouží jako hlavní platforma pro rychlou publikaci nových výsledků. Časopisecké články v CS jsou typicky revidované a doplněné verze článků dříve publikovaných ve sbornících konferencí a vycházejí zpravidla s víceletým zpožděním. Zejména u dynamicky se rozvíjejících podoborů CS (např. umělá inteligence) není nijak neobvyklé, že časopisecké verze konferenčních článků *vůbec nevzniknou*. Tento trend lze pozorovat zejména u výzkumných skupin, které produkují špičkové výsledky na hraně současného poznání a pro přípravu časopiseckých verzí dříve publikovaných prací nemají časový prostor.

Úroveň konferencí v CS je různá. Obecně akceptovaným indikátorem kvality je jejich ohodnocení (ranking) vytvořený Australskou společností pro výzkum a výuku v počítačových vědách (The Computing Research and Education Association of Australasia, CORE). Datábase CORE je pravidelně aktualizována (naposledy v roce 2021) a v současné době indexuje 976 konferencí rozdělených do pěti kategorií A*, A, B, C, „unranked“. Nejvyšší ohodnocení A* má 7% konferencí, kde jsou každoročně publikovány nejvýznamnější objevy v dané podoblasti CS. Přitom platí, že *takřka v každé oblasti CS existují konference ranku A**. Ranking konferencí (na rozdíl od impaktního faktoru časopisů) *nezávisí na velikosti příslušné vědecké komunity*. Platí rovněž, že míra citovanosti článků publikovaných ve sbornících konferencí ranku A* je výrazně vyšší než u časopiseckých článků (přesná data a související analýzu lze nalézt např. v článku „G. Vrettas, M. Sanderson: *Conferences vs. journals in computer science*, Journal of the Association for Information Science and Technology 66(12), 2015.).

Bibliometrickou analýzu v oblasti počítačových věd je tedy žádoucí provádět na základě výstupů ve sbornících výběrových konferencí při zohlednění jejich rankingu. Analýza prováděná na základě časopiseckých výstupů a příslušných impaktních faktorů podává zkreslený obraz o úrovni výzkumu v jednotlivých oblastech CS a systematicky znevýhodňuje podoblasti CS s méně početnou vědeckou komunitou.

Podklady pro bibliometrickou analýzu na základě konferenčních výstupů je možné získat automaticky z databází CORE a DBLP Computer Science Bibliography, které jsou dostupné zdarma a pravidelně aktualizovány.

Základní otázky – obory FORD

1. Jaká je úroveň oboru v ČR v mezinárodním srovnání?

Na základě dat databáze WoS lze konstatovat, že rozložení časopiseckých výstupů mezi jednotlivými kvartily i podíl publikací v prvním decilu jsou pro výstupy vyprodukované v ČR takřka stejné, jako u zemí EU15. Čeští výzkumníci pravidelně publikují ve sbornících konferencí CORE Rank A*, pro přesnější kvantifikaci podílu těchto publikací vzhledem k celkové produkci a příslušné srovnání se zahraničím ovšem chybí potřebné datové podklady. Mimo bibliografických dat je poměrně spolehlivým indikátorem špičkového výzkumu počet získaných ERC grantů v CS panelech. Ten je za posledních pět let nulový (nositelé ERC grantů v CS, kteří na českých institucích působí, je získali během předchozího působení v zahraničí). V tomto ohledu ČR zaostává za okolními zeměmi (Rakousko, Německo, Polsko), tento fakt ovšem není specifický pouze pro CS. Míra zapojení českých vědců do mezinárodních projektů a konsorcií je obecně na solidní úrovni, nikoliv ale ve vůdčí roli.

2. Je tato úroveň adekvátně reflektována zpracovanou bibliometrickou analýzou?

Bibliometrická analýza na bázi dat z WoS a impaktních faktorů podává spíše povrchní obraz o skutečném stavu CS v České republice z důvodů popsaných v úvodním komentáři. Zejména není na bázi těchto dat možné spolehlivě identifikovat a kvantifikovat skutečně špičkové vědecké výsledky z minulých let a přiřadit je k příslušným podoborům CS. Za tímto účelem by bylo třeba analyzovat data z databáze DBLP s využitím CORE ranku konferencí.

3. Má obor vysokou úroveň své produkce v D1 a Q1 ve srovnání s EU15 a se světem?

Úroveň produkce v segmentech D1 a Q1 je s EU15 a se světem plně srovnatelná, poměrné zastoupení výsledků v prvním decilu a jednotlivých kvartilech je velmi podobné.

4. Má v daném oboru význam institut korespondujícího (reprint) autora?

V CS je zvykem uvádět autory v abecedním pořadí. V případě, kdy je podíl některých autorů na publikaci výrazně menší, mohou být zařazeni na konec seznamu. Tento jev je spíše výjimečný a nemá ustálená pravidla. Korespondující autor nemá žádné výsadní postavení, zpravidla jde o člena autorského kolektivu zodpovědného za technickou přípravu publikace.

5. Jaké je v daném oboru zastoupení výsledků vytvořených ve velkých kolaboracích?

V CS se velké kolaborace nevyskytují, nebo jen velmi výjimečně. V České republice představují publikace s alespoň 30 autory méně než polovinu procenta z celkového počtu publikací.

6. Jaký je v daném oboru rozsah mezinárodní spolupráce při tvorbě špičkových výsledků? Můžete tyto výsledky nějak blíže charakterizovat (viz konkrétní výsledky v příloze)?

Rozsah mezinárodní spolupráce při tvorbě špičkových výsledků v CS je podobný, jako v jiných oborech. Produkce špičkových výsledků není vázána na dostupnost speciálního technického vybavení, zásadní roli má ovšem odborná erudice autorského

kolektivu, která přirozeným způsobem vybízí ke spolupráci s nejlepšími zahraničními pracovišti. Situace v ČR je v tomto ohledu plně srovnatelná se situací ve světě.

- 7. Do jaké míry lze považovat výsledky klasifikované jako D1 a Q1 za domácí „knowhow (viz konkrétní výsledky v příloze)? Je tento podíl podle vašich zkušeností v souladu s praxí v zahraničí?**

Jak již bylo zmíněno v úvodní pasáži, v CS je indikátorem kvality vědeckých výsledků jejich přijetí na prestižní konferenci a řada špičkových prací není publikována časopisecky. Z tohoto důvodu nelze domácí „know-how“ v oblasti CS spolehlivě identifikovat na základě časopiseckých výsledků v D1 a Q1.

- 8. Existují WoS kategorie, které jsou v oboru nadprůměrně dobré? Můžete je stručně okomentovat (nově je vám k tomu v analýze k dispozici podpůrný graf)?**

WoS kategorie v CS jsou velmi široké a spojují jen volně související podoblasti. Určitou singularitou je kategorie „Logic“, která zahrnuje veškeré aplikace nástrojů formální logiky v CS. Zde ČR dosahuje nejvyššího podílu na celkovém počtu dosažených výsledků v rámci zemí EU15 (zhruba 5 %) a má také vyšší podíl publikací v D1 a Q2 než EU15. V ostatních kategoriích WoS je podíl ČR na počtu dosažených časopiseckých výsledků 1-3 % a podíl publikací v D1 a Q1 je menší, než u zemí EU15. Skutečná vypovídací hodnota těchto údajů je ovšem omezená z důvodů popsaných v úvodním komentáři.

- 9. Dochází k významným překryvům s jinými obory, zejména u článků v pásmech D1/Q1? Můžete to stručně okomentovat? Odpovídají publikace klasifikované jako D1 a Q1 předpokládaným obsahem a zaměřením skutečně výzkumu v oboru FORD do jakého jsou přiřazeny (viz seznam jednotlivých článků v přílohách).**

Struktura oborů OECD (FORD) definuje dvě oblasti s výrazným zastoupením CS, a to 1.2 Computer and information sciences, kde všechny příslušné WoS kategorie spadají do oblasti počítačových věd, a dále 2.2 Electrical eng., Electronic eng., Information eng., kde jen některé příslušné WoS kategorie spadají do počítačových věd. Většina publikací klasifikována jako D1 a Q1 v oboru 1.2 svým obsahem odpovídá oblasti CS. Pro obor 2.2 toto platí jen částečně.

- 10. Jaké VO jsou podle oborových bibliometrických zpráv v daném oboru nejvýznamnější z hlediska produkce v D1 a Q1? Odpovídá významný podíl na špičkových výstupech relativně velikosti VO (viz referenční údaje a vaše znalosti oboru)?**

Z hlediska produkce v D1 má v rámci ČR nejvyšší podíl ČVUT (33 %), UK (23 %) a dále s odstupem Ostravská univerzita, Masarykova univerzita, UP Olomouc a Ústav teorie informace a automatizace AV ČR (6-8 %). Údaje o velikosti VO bohužel nejsou k dispozici.

- 11. Jaký je bibliometrický profil nejdůležitějších z těchto VO ve srovnání s benchmarky (svět, EU15, ČR)? Dosahují dle bibliometrie úrovně EU15 resp. světové úrovně, případně ji převyšují?**

Podíl zastoupení článků v D1 a Q1 je u výše zmíněných institucí srovnatelný se světem a EU15.

- 12. Je podle vašeho názoru bibliometrický výkon zásadní také z hlediska mise těchto VO? (Např. některé VO věnující se aplikovanému výzkumu mohou mít významný výkon v M2, ale z hlediska jejich mise nemusí mít zásadní úlohu.)**

Kvalitní základní výzkum, jehož výstupy mají formu kvalitních publikací, je jistě součástí mise veřejných vysokých škol i ústavů akademie věd.

- 13. Respektive můžete vyzdvihnout VO, která nemá mezi top 10 institucemi tak vysoký podíl v D1/Q1 na celonárodní úrovni, ale z hlediska své velikosti ve skutečnosti převyšuje výkon větších institucí?**

Pro posouzení velikosti instituce jsou potřebné podrobné údaje o počtu a struktuře vědeckých pracovníků, které bohužel nejsou v současné době k dispozici.

- 14. Existují v oboru VO, které podle Vašeho názoru produkují významný podíl národních výsledků, přesto se mezi špičkovými VO neobjevují? Je v některých z těchto VO je významné vysoké zastoupení výsledků vytvořených ve velkých kolaboracích, v mezinárodní spolupráci a výsledků s cizím reprint autorem? A naopak, existuje v oboru špičková instituce, která výrazněji než ostatní VO tvoří vynikající výsledky vlastním přispěním?**

V seznamu špičkových VO jsou uvedeny instituce, jejichž podíl na počtu dosažených časopiseckých výsledků v rámci České republiky činí v součtu 98%. Velké kolaborace nejsou v CS obvyklé.

- 15. Můžete se pokusit vyjmenovat publikačně výrazně slabé (ve srovnání se světem) organizace a to v obou významech: hodně publikací v nízkých pásmech i málo nebo žádné publikaceu těch VO, kde by se očekávaly?**

Výrazně slabé VO v oblasti CS není možné spolehlivě identifikovat na základě dat WoS z důvodů popsáných v úvodní části.

English Summary

Unlike in most other fields of science, CS researchers view conference publication as the most prestigious and sometimes final venue to describe their work. In almost every subfield of CS, there exist flagship (CORE Rank A*) conferences where the best results achieved by the world research community are presented. Although CS journals are often willing to publish revised and full versions of conference papers, CS academics often prefer to continue their cutting-edge research without investing additional time and effort into preparing journal papers. In certain CS subfields, the best conferences have become the sole record of research advances.

Consequently, the credibility of bibliographic analysis of CS research based solely on journal publications and journal impact factors is limited. More credible results can be obtained by analyzing conference papers, taking the conference ranking into account. The required data are available freely and updated regularly (DBLP and CORE databases).