

Impact Analysis of Croatian Language as the Publication Language on the Research Paper Visibility and Citation Potential in Computer Science and Engineering

Analiza utjecaja hrvatskog jezika kao jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva

Boris Tomaš

Trinity College Dublin

School of Computer Science and Statistics, CONNECT Centre

Dublin 2, Ireland D02 W272

boris.tomas@tcd.ie

Abstract. *This paper presents an extensive bibliometric analysis of the impact of the language of publication on the visibility and citation of scientific papers in computer science and engineering, with a particular focus on Croatian. Using data from the OpenAlex database, papers were analysed by language and country of origin. The results show that English-language papers achieve several-fold higher citation rates and global visibility, while Croatian-language papers are limited to a regional audience and have negligible international impact. The analysis suggests that publishing in Croatian is not effective for researchers who aim to contribute to global science, especially in computer science and engineering. Possible causes are discussed, including the language barrier, limited accessibility, and potentially lower quality of papers published in Croatian.*

Keywords. language of publication, citations, computer science, engineering, bibliometrics, OpenAlex, Croatian language

Sažetak. *Ovaj rad donosi opsežnu bibliometrijsku analizu utjecaja jezika publikacije na vidljivost i citiranost znanstvenih radova u području računalnih znanosti i inženjerstva, s posebnim naglaskom na hrvatski jezik. Korištenjem podataka iz baze OpenAlex, analizirani su radovi prema jeziku i zemlji porijekla. Rezultati pokazuju da radovi na engleskom jeziku ostvaruju višestruko veću citiranost i globalnu vidljivost, dok su radovi na hrvatskom jeziku ograničeni na regionalnu publiku i imaju zanemariv međunarodni utjecaj. Analiza sugerira da objavljivanje na hrvatskom jeziku nije učinkovito za znanstvenike koji žele doprinijeti svjetskoj znanosti, osobito u računalnim znanostima i inženjerstvu. U radu se raspravlja o mogućim uzrocima, uključujući jezičnu barijeru, ograničenu dostupnost i potencijalno nižu kvalitetu radova na hrvatskom jeziku.*

Ključne riječi. jezik publikacije, citiranost, računalne znanosti, inženjerstvo, bibliometrija, OpenAlex, hrvatski jezik

1 Introduction

Computer science and engineering represent one of the most dynamic and internationalised areas of contemporary science (Douglas et al., 2010; Garousi & Mäntylä, 2016; Kelly, 2015; Rajkumar et al., 2024). Owing to intensive global collaboration, rapid knowledge exchange, and constant technological progress, the international visibility and citation of scientific papers are key indicators of the impact and relevance of researchers and institutions. In this context, the language of publication becomes a decisive factor for the global reach of research results. Although a significant num-

1 Uvod

Računalne znanosti i inženjerstvo predstavljaju jedno od najdinamičnijih i najinternacionaliziranih područja suvremene znanosti (Douglas i dr., 2010; Garousi & Mäntylä, 2016; Kelly, 2015; Rajkumar i dr., 2024). Zbog intenzivne globalne suradnje, brze razmjene znanja i stalnog tehnološkog napretka, međunarodna vidljivost i citiranost znanstvenih radova ključni su pokazatelji utjecaja i relevantnosti istraživača i institucija. U tom kontekstu, jezik publikacije postaje odlučujući faktor za globalni doseg rezultata istraživanja. Iako se u Hrvatskoj i dalje objavljuje značajan broj radova

ber of papers in Croatia are still published in Croatian, particularly in domestic journals and conference proceedings, numerous bibliometric analyses indicate the limited international impact of such papers (Garousi & Mäntylä, 2016; Nurzhanov et al., 2021). The aim of this paper is to quantitatively analyse the relationship between language of publication and citation in computer science and engineering, using the latest data from the OpenAlex¹ database of scholarly works.

1.1 Description of the OpenAlex database and reasons for its use

OpenAlex is an open, global bibliographic database that indexes scientific papers, journals, authors, and institutions worldwide (OurResearch, 2024; Priem et al., 2022). Created in 2022 as a project of the non-profit organisation OurResearch, it succeeded the Microsoft Academic Graph, and today contains more than 260 million works from all scientific fields, including computer science and engineering. Unlike commercial databases such as Web of Science and Scopus, OpenAlex is completely free, open, and transparent; all data and program code are available under an open licence. The database allows searching, analysis, and data export via a web interface, REST API, or by downloading periodic snapshots of the entire database. OpenAlex aggregates data from numerous sources (Crossref, ORCID, arXiv, PubMed, and others), covering a wide spectrum of publication types (articles, conference papers, books, or dissertations) and provides rich metadata, including citations, links between authors, institutions, topics, and sources of funding. In this study OpenAlex was used for several key reasons:

- **Openness and accessibility:** Unlike closed commercial databases, OpenAlex enables free access to and export of data without licensing restrictions, which allows full transparency and reproducibility of the analysis.
- **Wide coverage:** OpenAlex indexes a considerably larger number of works, including publications in different languages and from less represented regions, which is important for analysing national and regional patterns in science.
- **Rich and connected metadata:** The database supports linking works with authors, institutions, topics, and countries, which is crucial for advanced bibliometric analyses and visualisations of scientific output and collaboration. OpenAlex provides detailed classification by scientific fields, including computer science and engineering, and enables filtering and analysis by language, country, and publication type, which is necessary for this research.

¹openalex.org

na hrvatskom jeziku, posebice u domaćim časopisima i zbornicima konferencija, rezultati brojnih bibliometrijskih analiza ukazuju na ograničen međunarodni utjecaj takvih radova (Garousi & Mäntylä, 2016; Nurzhanov i dr., 2021). Cilj ovog rada je kvantitativno analizirati odnos između jezika publikacije i citiranosti u području računalnih znanosti i inženjerstva, koristeći najnovije podatke iz OpenAlex¹ baze znanstvenih radova.

1.1 Opis baze podataka OpenAlex i razlozi korištenja

OpenAlex je otvorena, globalna bibliografska baza podataka koja indeksira znanstvene radove, časopise, autore i institucije iz cijeloga svijeta (OurResearch, 2024; Priem i dr., 2022). Nastala je 2022. godine kao projekt neprofitne organizacije OurResearch, nasljeđujući Microsoft Academic Graph, te danas sadrži više od 260 milijuna radova iz svih znanstvenih područja, uključujući računalne znanosti i inženjerstvo. Za razliku od komercijalnih baza poput Web of Science i Scopus, OpenAlex je potpuno besplatan, otvoren i transparentan; svi podaci i programski kod dostupni su pod otvorenom licencom. Baza omogućuje pretraživanje, analizu i izvoz podataka putem web sučelja, REST API-ja ili preuzimanjem periodičnih snimki cijele baze. OpenAlex agregira podatke iz brojnih izvora (Crossref, ORCID, arXiv, PubMed i drugi), pokrivajući širok spektar publikacija (članci, konferencijski radovi, knjige ili disertacije) i pruža bogate meta-podatke, uključujući citate, povezanost autora, institucija, tema i izvora financija. U ovom istraživanju OpenAlex je korišten iz nekoliko ključnih razloga:

- **Otvorenost i dostupnost:** Za razliku od zatvorenih komercijalnih baza, OpenAlex omogućuje slobodan pristup i izvoz podataka bez licencijskih ograničenja, što omogućuje potpunu transparentnost i ponovljivost analize.
- **Široka pokrivenost:** OpenAlex indeksira znatno veći broj radova, uključujući publikacije na različitim jezicima i iz manje zastupljenih regija, što je važno za analizu nacionalnih i regionalnih obrazaca u znanosti.
- **Bogati i povezani meta-podaci:** Baza omogućuje povezivanje radova s autorima, institucijama, temama i zemljama, što je ključno za napredne bibliometrijske analize i vizualizacije znanstvene produkcije i suradnje. OpenAlex pruža detaljnu klasifikaciju po znanstvenim područjima, uključujući računalne znanosti i inženjerstvo, te omogućuje filtriranje i analizu prema jeziku, zemlji i tipu publikacije, što je nužno za ovo istraživanje.

¹openalex.org

Thanks to these characteristics, OpenAlex enabled a comprehensive, open, and reproducible bibliometric analysis with an emphasis on the impact of publication language and geographical origin of works.

1.2 Structure of the paper

The paper is organised as follows: the *Introduction* presents the motivation, aims, and relevance of the topic and describes the OpenAlex database. The *Review of previous research* summarises relevant literature and prior findings on the impact of language of publication on citation in computer science and engineering. The *Methodology* describes the data sources, data collection and processing, and the analytical methods used. The *Results* section presents the main findings, including the number of papers and citations by language, the share of Croatian-language works, the geographical origin of papers, and trends over time. The *Discussion* analyses and interprets the results, discusses possible causes and limitations, and provides recommendations. The *Conclusion* summarises the main conclusions and suggests directions for future research.

2 Review of previous research

Computer science and engineering are characterised by an exceptionally high level of internationalisation, which is supported by the fact that the leading journals, conferences, and repositories are almost exclusively in English (Douglas et al., 2010; Rajkumar et al., 2024). According to Kelly (Kelly, 2015), in computer science and software engineering a significant share of citations comes from conference papers and articles in English compared with other scientific fields. Garousi et al. (Garousi & Mäntylä, 2016) show that the language of publication significantly affects citation even within the same journals or disciplines. Papers in English have a several-fold higher probability of being cited, which is associated with greater accessibility, a broader reach, and better indexing in international databases. Similar trends are observed in engineering, where international collaboration increases impact (Brito et al., 2018; Yildirim et al., 2024).

3 Methodology

3.1 Data sources and processing

Data were retrieved from the OpenAlex database of scientific works, using an API query that enables grouping of results by language and country of origin. All works classified in computer science and engineering were analysed. For each language, the total number of works, the total number of citations, and the distribution of works by country were recorded. OpenAlex assigns a work's country based on the major

Zahvaljujući navedenim karakteristikama, OpenAlex je omogućio provedbu opsežne, otvorene i ponovljive bibliometrijske analize s naglaskom na utjecaj jezika publikacije i zemljopisnog porijekla radova.

1.2 Struktura rada

Rad je organiziran na sljedeći način: *Uvod*, donosi motivaciju, ciljeve i relevantnost teme te opisuje korištenu bazu podataka OpenAlex. *Pregled dosadašnjih istraživanja* donosi pregled relevantne literature i dosadašnjih rezultata o utjecaju jezika publikacije na citiranost u računalnim znanostima i inženjerstvu. *Metodologija* opisuje izvore podataka, način prikupljanja i obrade podataka te korištene analitičke metode. *Rezultati* prikazuju glavne nalaze analize, uključujući broj radova i citiranost po jeziku, udio hrvatskog jezika, geografski izvor radova te trendove kroz vrijeme. *Diskusija* analizira i interpretira dobivene rezultate, raspravlja o mogućim uzrocima i ograničenjima te daje preporuke. *Zaključak* sažima glavne zaključke rada i predlaže smjernice za buduća istraživanja.

2 Pregled dosadašnjih istraživanja

Računalne znanosti i inženjerstvo karakterizira izuzetno visok stupanj internacionalizacije, čemu doprinosi i činjenica da su vodeći časopisi, konferencije i repozitoriji gotovo isključivo na engleskom jeziku (Douglas i dr., 2010; Rajkumar i dr., 2024). Prema Kelly (Kelly, 2015), u računalnim znanostima i softverskom inženjerstvu značajan broj citata dolazi iz konferencijskih radova i članaka na engleskom jeziku u odnosu na druga polja znanosti. Garousi i sur. (Garousi & Mäntylä, 2016) pokazuju da jezik publikacije značajno utječe na citiranost čak i unutar istih časopisa ili disciplina. Radovi na engleskom jeziku imaju višestruko veću vjerojatnost da budu citirani, što je povezano s većom dostupnošću, širim dosegom i boljom indeksiranošću u međunarodnim bazama. Slični trendovi uočeni su i u inženjerstvu, gdje međunarodna suradnja povećavaju utjecaj (Brito i dr., 2018; Yildirim i dr., 2024).

3 Metodologija

3.1 Izvori podataka i obrada

Podaci su preuzeti iz baze znanstvenih radova OpenAlex, koristeći API upit koji omogućuje grupiranje radova po jeziku i zemlji porijekla. Analizirani su svi radovi iz područja računalnih znanosti i inženjerstva. Za svaki jezik zabilježen je ukupan broj radova, ukupan broj citata te distribucija radova po zemljama. OpenAlex dodjeljuje zemlju rada na temelju porijekla većine autora ili na temelju zemlje porijekla prvog autora rada.

ity of authors' affiliations or, alternatively, the first author's country of origin.

3.2 Examples of code used for the analysis

For the purposes of this research, bespoke tools and scripts were used to download, process, and analyse data from OpenAlex database. The code snippets shown provide transparent insight into the methods used and facilitate the reproducibility and extensibility of the analysis. Below are the key parts of the code used to retrieve data on the language of publication, number of works, citations, and geographical origin, as well as to perform relevant statistical analyses.

The result of executing Code 1 is a file with all languages in the OpenAlex database, including the number of works written in the corresponding language. These data are available in the open repository here: github.com/boristomas/R1Podaci/blob/main/jezici.csv

The result of executing Code 2 is a file with data on each language, including citations and fields. These data are available in the open repository here: github.com/boristomas/R1Podaci/blob/main/podaci.csv In both code snippets, field identifiers 17 and 22 denote computer science and engineering.

4 Results

4.1 Number of papers by language in computer science and engineering

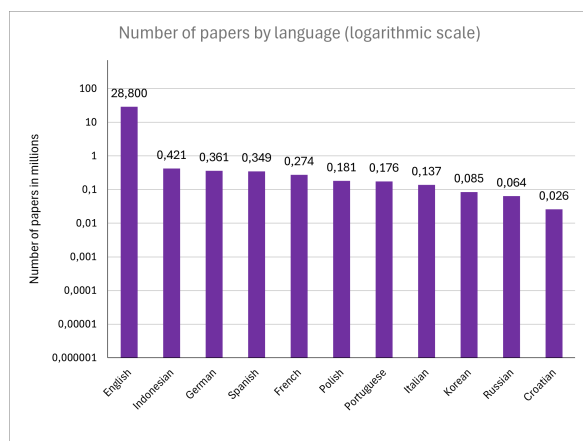


Figure 1. Number of papers by language (top 10 languages and Croatian) in computer science and engineering.

Fig. 1 shows the number of papers by language for the ten most represented languages in computer science and engineering, including Croatian as the eleventh (last column). It is clearly visible that English absolutely dominates, with more than 28 million papers, while all other languages trail significantly, including Croatian (25,810 papers). Croatian ranks 11th by number

3.2 Primjeri kôda korištenih za analizu

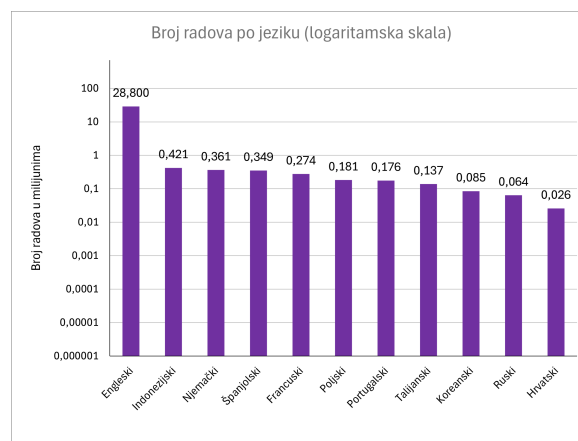
Za potrebe ovog istraživanja korišteni su vlastiti programski alati i skripte za preuzimanje, obradu i analizu podataka iz baze znanstvenih radova OpenAlex. Prikazani primjeri kôda omogućuju transparentan uvid u korištene metode te olakšavaju ponovljivost i proširivost analize. U nastavku su detaljno opisani ključni dijelovi kôda koji su korišteni za dohvat podataka o jeziku publikacije, broju radova, citiranosti i zemljopisnom podrijetlu radova, kao i za izvođenje relevantnih statističkih analiza.

Rezultat izvršavanja kôda 1 je datoteka sa svim jezicima u OpenAlex bazi podataka uključujući i broj radova napisanih na odgovarajućem jeziku. Ovi podaci su dostupni na otvorenom repozitoriju ovdje: github.com/boristomas/R1Podaci/blob/main/jezici.csv

Rezultat izvršavanja Kôda 2 je datoteka s podacima o svakom jeziku o citiranosti i područjima. Ovi podaci su dostupni na otvorenom repozitoriju ovdje: github.com/boristomas/R1Podaci/blob/main/podaci.csv U oba kôda, oznaka polja 17 i 22 su za računalne znanosti i inženjerstvo.

4 Rezultati

4.1 Broj radova po jeziku u računalnim znanostima i inženjerstvu



Slika 1. Broj radova po jeziku (top 10 jezika i hrvatski jezik) u računalnim znanostima i inženjerstvu.

Na slici 1 prikazan je broj radova po jeziku za deset najzastupljenijih jezika u području računalnih znanosti i inženjerstva, uključujući i hrvatski jezik kao jedanaesti (zadnji stupac). Jasno je vidljivo da engleski jezik apsolutno prevladava, s više od 28 milijuna radova, dok su svi ostali jezici značajno iza, uključujući i hrvatski jezik (25810 radova). Hrvatski jezik je 11. po broju

of papers among all languages in the database (55). For readability, a logarithmic scale was used; absolute counts are printed above each column.

4.2 Citation trends by language

The analysis of average citation by language, shown in Fig 2, further confirms the dominant position of English. English has an average of 11.2 citations per paper, whereas Croatian has only 0.42, more than 26 times lower. A higher citation rate is notable for French-language papers, which is likely a result of the relatively small proportion of proficient English speakers in the Francophone world. According to (First, 2025), France ranks only 49th in English proficiency, while Croatia is in a high 5th place.

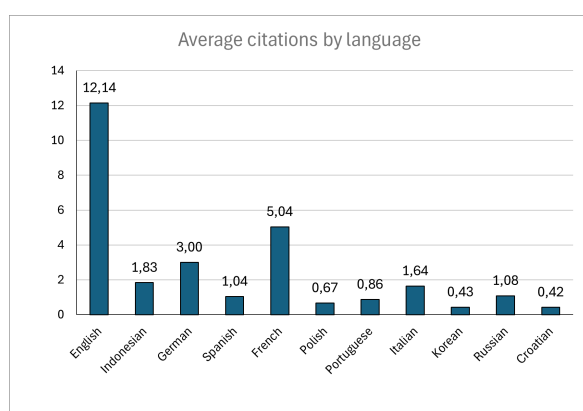


Figure 2. Average citations per paper by language (top 10 languages and Croatian) in computer science and engineering.

4.3 Share of Croatian in the total number of papers and citations

The share of Croatian-language papers in the total number of papers is just 0.08% (25810 papers), while the share of citations is only 0.003% (10838 citations) (Fig 3).

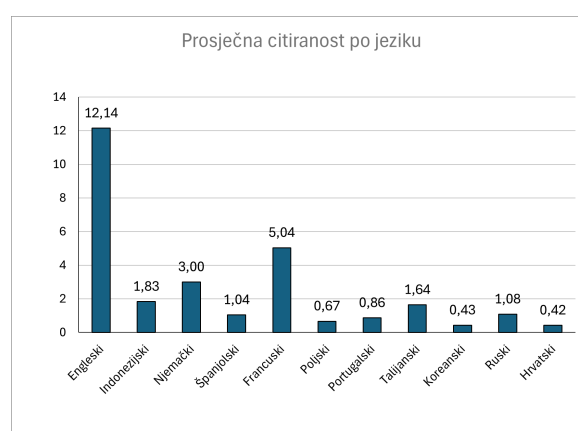
4.4 Countries of origin of Croatian-language papers

Table 1 shows the top 20 countries from which Croatian-language papers originate. Although most works are from Croatia, a significant number come from Serbia, Bosnia and Herzegovina, and Slovenia, but also from countries outside the region such as the USA, Germany, the UK, and China.

radova od svih jezika u bazi podataka (55). Zbog lakše čitljivosti korištena je logaritamska skala, broj radova je izražen kao apsolutni broj iznad svakog stupca.

4.2 Trendovi citiranosti po jeziku

Analiza prosječne citiranosti po jeziku, prikazana na slici 2, dodatno potvrđuje dominantnu poziciju engleskog jezika. Engleski jezik ima prosječnu citiranost od 11.2 po radu, dok hrvatski ima samo 0.42, što je više od 26 puta manje. Primjetna je veća citiranost radova na francuskom jeziku što je vjerojatno rezultat malog udjela stanovnika koji znaju engleski jezik u francuskom govornom području. Prema (First, 2025) Francuska je tek 49. zemlja po broju i kvaliteti govornika na engleskom jeziku, a Hrvatska je na visokom 5. mjestu.



Slika 2. Prosječna citiranost po jeziku (top 10 jezika i hrvatski jezik) u računalnim znanostima i inženjerstvu.

4.3 Udio hrvatskog jezika u ukupnom broju radova i citata

Udio radova na hrvatskom jeziku u ukupnom broju radova iznosi samo 0.08% (25810 radova), dok udio citata iznosi tek 0.003% (10838 citata) (Slika 3).

4.4 Zemlje izvora radova na hrvatskom jeziku

Tablica 1 prikazuje prvih 20 zemalja iz kojih dolaze radovi na hrvatskom jeziku. Iako je većina radova iz Hrvatske, značajan broj dolazi iz Srbije, Bosne i Hercegovine, Slovenije, ali i iz zemalja izvan regije poput SAD-a, Njemačke, UK-a i Kine.

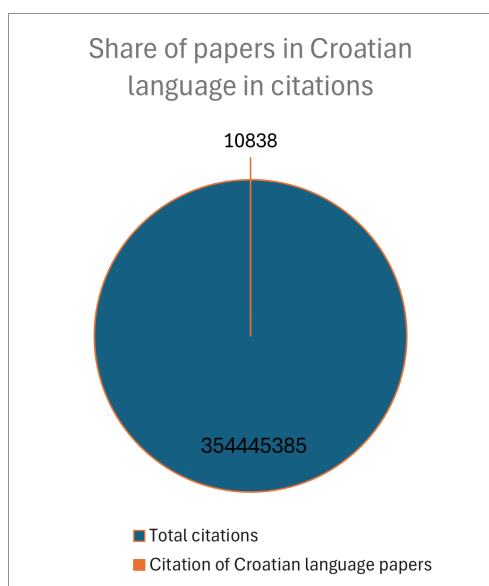


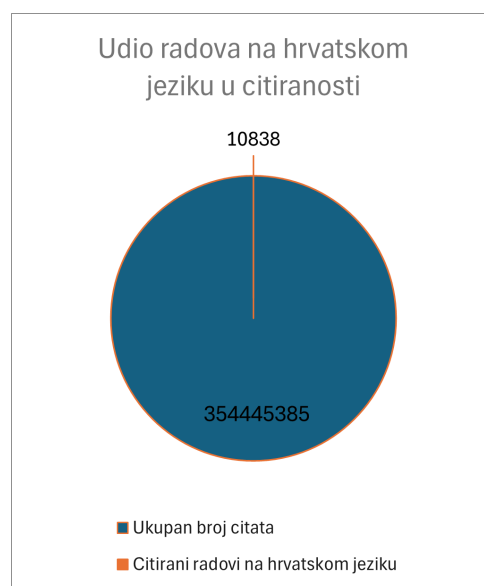
Figure 3. Share of Croatian-language papers in the total number of citations.

Table 1. Top 20 countries of origin of Croatian-language papers

Country	Number of papers
Croatia	2,495
Serbia	683
Bosnia and Herzegovina	221
Slovenia	171
USA	91
Czechia	60
Poland	53
Slovakia	53
Türkiye	52
Germany	46
UK	43
Ukraine	31
Austria	28
China	26
Hungary	25
Spain	24
Peru	19
France	18
Italy	18
Montenegro	18
Russia	18

4.5 Trends in publications and citations over time

Over the last 15 years, the number of English-language papers in computer science and engineering has been continuously growing, while the number of Croatian-language papers is stagnating or declining (Fig 4). Citation rates for English-language papers also show an upward trend, while Croatian-language papers remain



Slika 3. Udio radova na hrvatskom jeziku u ukupnom broju citata.

Tablica 1. Prvih 20 zemalja iz kojih dolaze radovi na hrvatskom jeziku

Zemlja	Broj radova
Hrvatska	2,495
Srbija	683
Bosna i Hercegovina	221
Slovenija	171
SAD	91
Češka	60
Poljska	53
Slovačka	53
Turska	52
Njemačka	46
UK	43
Ukrajina	31
Austrija	28
Kina	26
Mađarska	25
Španjolska	24
Peru	19
Francuska	18
Italija	18
Crna Gora	18
Rusija	18

4.5 Analiza trendova publikacija i citiranosti kroz vrijeme

Kroz posljednjih 15 godina, broj radova na engleskom jeziku u računalnim znanostima i inženjerstvu kontinuirano raste, dok broj radova na hrvatskom jeziku stagnira ili opada (Slika 4). Citiranost radova na engleskom jeziku također pokazuje uzlazni trend, dok radovi na hrvatskom jeziku ostaju na vrlo niskim vrijednostima.

very low. These trends further emphasise the importance of the choice of publication language for international visibility and impact.

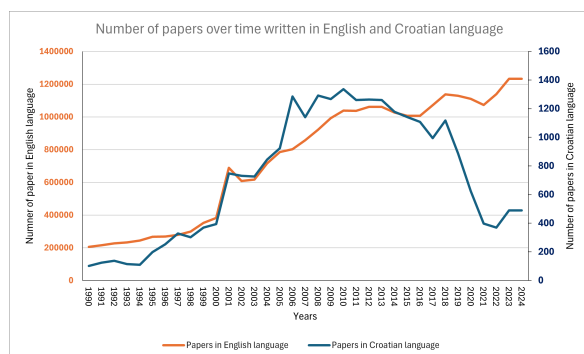
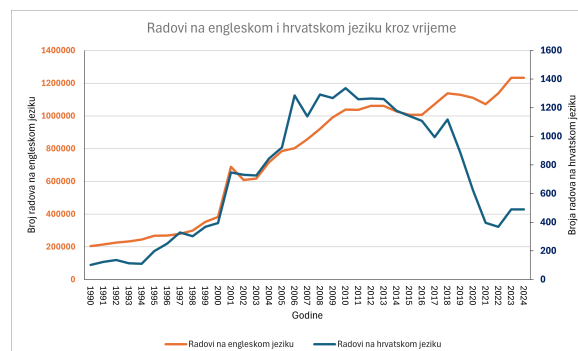


Figure 4. Papers published in English and Croatian over time.

5 Discussion

The results clearly show that English is not only dominant in terms of the number of papers and citations, but also in the international dispersion of authors and institutions. The data show an interesting and significant contribution of Indonesian, which stems from OpenAlex's strategy to minimise geo-discrimination. It is also noteworthy that Chinese (Mandarin) does not dominate by number of works. The reason lies in the fact that the Chinese government invests in and financially supports publication of papers in English (MoChridhe, n.d.). China has recognised the importance of English in science and, through such measures, strives to ensure and maintain its leading scientific position. Croatian, although with a smaller total number of works, shows an interesting regional and diaspora component. A significant number of Croatian-language papers come from Serbia, Bosnia and Herzegovina, and Slovenia, indicating a shared scientific space across the former Yugoslavia. Works from countries outside the Croatian language's regional environment (for example, China) are an interesting phenomenon that may be the result of university promotion policies. However, the average citation of Croatian-language papers remains low, suggesting limited international visibility. This limitation may be due to the smaller number of journals indexed in international databases, the language barrier, and perceptions regarding the importance of national languages in science. The analysis suggests that publishing in Croatian is less effective for researchers aiming to contribute to global science, especially in computer science and engineering. This analysis did not include a qualitative assessment of the papers, but it can be assumed that works published in Croatian are of less general interest, as indicated by lower citation. However, a further reason may simply be that Croatian-language papers are accessible to a smaller population

Ovi trendovi dodatno naglašavaju važnost odabira jezika publikacije za međunarodnu vidljivost i utjecaj.



Slika 4. Objavljeni radovi na engleskom i hrvatskom jeziku kroz vrijeme.

5 Diskusija

Rezultati jasno pokazuju da je engleski jezik ne samo dominantan po broju radova i citata, već i po međunarodnoj disperziji autora i institucija. Analiza podataka pokazuje zanimljiv i značajan doprinos indonezijskog jezika, razlog tome leži u strategiji OpenAlex organizacije da minimizira geo-diskriminaciju. Također je interesantna činjenica da kineski jezik (mandarinski) ne dominira u broju radova. Razlog tome leži u činjenici da kineska vlada ulaže i financijski podupire objavu radova na engleskom jeziku (MoChridhe, bez datuma). Kina je prepoznala važnost engleskog jezika u znanosti te svojim potezima nastoji osigurati i zadržati vodeću poziciju u znanosti. Hrvatski jezik, iako s manjim ukupnim brojem radova, pokazuje zanimljivu regionalnu i dijasporu komponentu. Značajan broj radova na hrvatskom jeziku dolazi iz Srbije, Bosne i Hercegovine i Slovenije, što ukazuje na postojanje zajedničkog znanstvenog prostora na području bivše Jugoslavije. Radovi iz zemalja koje su izvan regionalnog okruženja hrvatskog jezik, poput na primjer Kina su interesantna pojava koja je možda rezultat sveučilišnih politika uvjetovanja za napredovanja. Međutim, prosječna citiranost radova na hrvatskom jeziku ostaje niska, što sugerira da je njihova međunarodna vidljivost ograničena. Ova ograničenost može biti posljedica manjeg broja časopisa indeksiranih u međunarodnim bazama, jezične barijere, ali i percepcije o važnosti nacionalnih jezika u znanosti. Analiza sugerira da objavljivanje na hrvatskom jeziku manje učinkovito za znanstvenike koji žele doprinijeti svjetskoj znanosti, osobito u računalnim znanostima i inženjerstvu. Analiza nije uključivala kvalitativnu procjenu radova, no može se pretpostaviti da su radovi objavljeni na hrvatskom jeziku manje interesantni, što je pokazano i manjom citiranošću. Međutim, razlog manje citiranosti može ležati i u činjenici da su radovi na hrvatskom jeziku dos-

of researchers who understand Croatian, which reduces the likelihood of being cited. Limited availability and the language barrier are key factors influencing this result, consistent with other studies that examine differences in citation practices between English and Croatian authors (Varga & Gradečak-Erdeljić, 2017). A potentially significant source of error in this research is the limited dataset within the OpenAlex database.

6 Conclusion

The analysis clearly shows that, in computer science and engineering, publishing in English is essential in order to achieve international visibility and scientific impact. Croatian-language papers remain restricted to the regional scientific community and have a limited contribution to global science. It is recommended that researchers in these fields publish their results in English, particularly if they aim to contribute to the global scientific community. The national language may make sense for professional, educational, or popular science works, but not for scientific publications that seek international relevance. A particularly unfavourable effect on the internationalisation of Croatian science is the requirement, in some Croatian universities' promotion criteria, to publish in Croatian. Such requirements significantly hinder the career plans of visiting scientists who do not speak Croatian but wish to spend part of their scientific career in the Republic of Croatia. Future research should analyse the quality of language in Croatian-language papers produced outside the regional environment; one possible hypothesis is that such papers are originally written in another language and machine-translated into Croatian. Furthermore, future research should analyse in more detail the impact of international collaboration and the diaspora on Croatian-language scientific production, as well as possibilities for improving the visibility of domestic journals through open access and inclusion in international databases.

tupni manjoj populaciji znanstvenika koji razumiju hrvatski jezik pa samim time je manje vjerojatno da će biti citirani. Ograničenost dostupnosti i jezična barijera ključni su faktori koji utječu na ovaj rezultat, što je u skladu s nalazima drugih studija koje analiziraju razlike u citatnim praksama između engleskih i hrvatskih autora (Varga & Gradečak-Erdeljić, 2017). Potencijalno značajan izvor pogreške u ovom istraživanju može biti i ograničen set podataka u OpenAlex bazi podataka.

6 Zaključak

Analiza rada jasno pokazuje da je u području računalnih znanosti i inženjerstva objavljivanje na engleskom jeziku nužno za postizanje međunarodne vidljivosti i znanstvenog utjecaja. Radovi na hrvatskom jeziku ostaju ograničeni na regionalnu znanstvenu zajednicu te imaju ograničen doprinos svjetskoj znanosti. Preporučuje se da znanstvenici iz ovih područja svoje rezultate objavljuju na engleskom jeziku, osobito ako im je cilj doprinijeti globalnoj znanstvenoj zajednici. Nacionalni jezik može imati smisla za stručne, nastavne ili popularno-znanstvene radove, ali ne i za znanstvene publikacije koje ciljaju na međunarodnu relevantnost. Naročito nepovoljan efekt na internacionalizaciju hrvatske znanosti ima uvjetovanje napredovanja znanstvenika na sveučilištima u Republici Hrvatskoj objavom radova na hrvatskom jeziku. To uvjetovanje značajno otežava karijerni plan gostujućih znanstvenika koji ne znaju hrvatski jezik, a žele dio znanstvene karijere provesti u Republici Hrvatskoj. Buduća istraživanja bi trebala analizirati kvalitetu jezika kod radova na hrvatskom jeziku koji su napravljeni izvan regionalnog okruženja; moguća pretpostavka je da se radi o radovima koji su napisani originalnim jezikom te su strojno prevedeni na hrvatski jezik. Nadalje, buduća istraživanja trebala bi i detaljnije analizirati utjecaj međunarodne suradnje i dijaspor na znanstvenu produkciju na hrvatskom jeziku, kao i mogućnosti za unaprijeđenje vidljivosti domaćih časopisa kroz otvoreni pristup i uključivanje u međunarodne baze podataka.

References / Literatura

- Brito, C., Ciampi, M., Santos, H., Vasconcelos, R., & Amaral, L. (2018, June). *Internationalization and Its Impact in Engineering Programs* (tech. rep.). EasyChair. <https://doi.org/10.29007/jx9r>
- Douglas, S., Farley, A., Lo, G., Proskurowski, A., & Young, M. (2010). Internationalization of computer science education. *Proceedings of the 41st ACM technical symposium on Computer science education*, 411–415. <https://doi.org/10.1145/1734263.1734404>
- First, E. (2025). EF EPI - EF English Proficiency Index. A Ranking of 116 Countries and Regions by English Skills. <https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2024/ef-epi-2024-english.pdf>
- Garousi, V., & Mäntylä, M. V. (2016). Citations, research topics and active countries in software engineering: A bibliometrics study. *Computer Science Review*, 19, 56–77. <https://doi.org/10.1016/j.cosrev.2015.12.002>
- Kelly, M. (2015). Citation Patterns of Engineering, Statistics, and Computer Science Researchers: An Internal and External Citation Analysis across Multiple Engineering Subfields. *College Research Libraries*, 76, 859–882. <https://doi.org/10.5860/crl.76.7.859>
- MoChridhe, R. (n.d.). The Hidden Language Policy of China's Research Evaluation Reform. <https://blog.westminster.ac.uk/contemporarychina/the-hidden-language-policy-of-chinas-research-evaluation-reform/>
- Nurzhanov, C., Pidlisnyuk, V., Naizabayeva, L., & Satymbekov, M. (2021). Research and trends in computer science and educational technology during 2016–2020: Results of a content analysis. *World Journal on Educational Technology: Current Issues*, 13, 115–128. <https://doi.org/10.18844/wjet.v13i1.5421>
- OurResearch. (2024). OpenAlex Documentation [Available at: <https://docs.openalex.org/>].
- Priem, J., Piwowar, H., Orr, R., et al. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts [Available at: <https://openalex.org/about>].
- Rajkumar, S., Sujay, D., Sangeetha, R., Sudhakar, G., Muralikrishna, K., & Kumaraswamy, D. (2024). An Overview of Computer Science and Engineering and Its Latest Technologies. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2, 683–698. <https://doi.org/10.47392/IRJAEH.2024.0099>
- Varga, M., & Gradečak-Erdeljić, T. (2017). English and Croatian citation practices in research articles in applied linguistics: a corpus-based study. *Kalbotyra*, 70, 153–183. <https://doi.org/10.15388/Klbt.2017.11199>
- Yildirim, G., Pan, X., & Rahman, A. (2024). Fluid Mechanics Education in Engineering: Challenges and Trends from a Bibliometric Analysis. *Proceedings of the 3rd International Conference on Advancements in Engineering Education*. <https://doi.org/10.71427/iCAEED2025/24>

Appendix / Prilozi

Appendix 1 / Prilog 1

Code 1. Analysis of languages and countries of origin / Kod 1: Analiza jezika i zemalja porijekla

```
import requests
import pandas as pd
from urllib.parse import urlparse
API_URL = "https://api.openalex.org/works?group_by=
language&per_page=200&filter=type:types/
article,primary_topic.field.id:
fields/17|fields/22"
def extract_language_code(key_url):
    if not key_url:
        return "unknown"
    parsed = urlparse(key_url)
    return parsed.path.split('/')[1] if '/' in parsed.path else parsed.path
def fetch_language_groups(api_url):
    response = requests.get(api_url)
    response.raise_for_status()
    data = response.json()
    groups = data.get('group_by', [])
    results = []
    for group in groups:
        lang_code = extract_language_code(
            group.get('key'))
        lang_name = group.get('key_display_name', 'Unknown')
        count = group.get('count', 0)
        results.append({
            'language_name': lang_name,
            'language_code': lang_code,
            'count': count
        })
    return results
def main():
    language_data = fetch_language_groups(API_URL)
    df = pd.DataFrame(language_data)
    print(df.to_string(index=False))
```

Appendix 2 / Prilog 2

Code 2. Data analysis / Kod 2: Analiza podataka

```

import pandas as pd
import time
import requests

languages_df = pd.read_csv('jezici.csv', sep=';')
def get_works_and_citations(language_code):
    base_url = "https://api.openalex.org/works"
    params = {
        "filter": f"type:types/article,primary_topic.field.id:fields/17|fields/22,
                    language:languages/{language_code}",
        "sort": "publication_year:desc",
        "per_page": 200,
        "apc_sum": "false",
        "cited_by_count_sum": "true"
    }
    response = requests.get(base_url, params=params)
    if response.status_code != 200:
        print(f"Error fetching data for {language_code}: {response.status_code}")
        return 0, 0
    data = response.json()
    total_works = data.get('meta', {}).get('count', 0)
    total_citations = data.get('meta', {}).get('cited_by_count_sum', 0)
    return total_works, total_citations
def get_country_distribution(language_code):
    base_url = "https://api.openalex.org/works"
    params = {
        "group_by": "authorships.countries",
        "filter": f"type:types/article,primary_topic.field.id:fields/17|fields/22,
                    language:languages/{language_code}",
        "per_page": 200
    }
    response = requests.get(base_url, params=params)
    if response.status_code != 200:
        print(f"Error fetching data for {language_code}: {response.status_code}")
        return {}
    data = response.json()
    country_groups = data.get('group_by', [])
    country_distribution = {group['key_display_name']: group['count'] for group in
                           country_groups}
    return country_distribution
results = []
for idx, row in languages_df.iterrows():
    lang_name = row['language_name']
    lang_code = row['language_code']
    if not isinstance(lang_code, str) or lang_code.lower() in ['none', '', 'nan']:
        print(f"Skipping {lang_name}: invalid language code.")
        continue
    print(f"Processing {lang_name} ({lang_code})...")
    works, citations = get_works_and_citations(lang_code)
    countries = get_country_distribution(lang_code)
    results.append({
        "language_name": lang_name,
        "language_code": lang_code,
        "total_works": works,
        "total_citations": citations,
        "countries": countries
    })
    time.sleep(1)
results_df = pd.DataFrame(results)
results_df.to_csv("podaci.csv", index=False)

```