

Citiranje jezikoslovnih podatkov v slovenskih znanstvenih objavah: stanje in priporočila

Darja Fišer^{*†}, Jakob Lenardič^{*}, Tomaž Erjavec[†]

^{*} Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani

darja.fiser@ff.uni-lj.si, jakob.lenardic@ff.uni-lj.si

[†] Odsek za tehnologije znanja, Institut »Jožef Stefan«

tomaz.erjavec@ijs.si

Pregled predavanja

1. Uvod
2. Analiza stanja v izbranih slovenskih publikacijah
3. Smernice za boljšo prakso
4. Zaključki

Uvod

- Odprti dostop do znanstvenih publikacij in **podatkov** pospešuje inovacije, spodbuja sodelovanje, zmanjšuje podvajanje dela in omogoča dograjevanje predhodnih rezultatov raziskav ter vključevanje državljanov in družbe (European Commission, 2012)
- *Nacionalna strategija odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015-2020*
 - *Akcijski načrt izvedbe nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015—2020*
 - *Pilotni program Odprti dostop do raziskovalnih podatkov 2017—2020*
- *Resolucija o nacionalnem programu za jezikovno politiko 2014–2018*

Citiranje jezikovnih virov v Sloveniji

- Odprti viri: MULTEXT-East, JOS, SSJ ((cc)GigaFida, (cc)KRES), Janes, KAS, ...
- Raziskovalna infrastruktura za jezikovne vire in orodja **CLARIN.SI**
 - certificiran repozitorij (>100 virov, večina CC)
 - konkordančnika (>50 korpusov, prijava ni potrebna)

Dictionary of the Slovenian Language in the Works of Janez Svetokriški



“ Za citiranje vnosa uporabite naslednjo referenco ali jo izvozite v prednastavljeno obliko:

BIBTEX

CMDI

Snoj, Marko, 2006, *Dictionary of the Slovenian Language in the Works of Janez Svetokriški*, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1092>.



Delite:

CLARIN.SI Data & Tools

Avtorji	Snoj, Marko
Item identifier	http://hdl.handle.net/11356/1092
URL projekta	http://www.fran.si/139/slovar-jezika-janeza-svetokriskega
Demo URL	http://www.fran.si/139/slovar-jezika-janeza-svetokriskega

Citiranje podatkov v jezikoslovju

- Ozadje: »Citiranje je še posebej pomembno, ker je merljiv kazalec raziskovalne uspešnosti, zato bi se tudi moralo dosledno izvajati. Žal pa to ni v navadi pri citiranju publikacij o jezikovnih virih: vse prepogosto se nek vir omeni samo po imenu ali pa se v najboljšem primeru doda njegov spletni naslov, namesto da bi se v virih citiralo publikacijo, kjer je vir prvotno opisan« (Erjavec, 2009).
- Povod: sestanek interesne skupine za jezikoslovne podatke »Research Data Alliance«, Berlin 22. 3. 2018
- Cilj: preučiti citiranje uporabe jezikovnih virov v znanstvenih člankih slovenskih publikacij

V Sloveniji?

- Pregled jezikoslovnih zbornikov in revij 2013-2017:
 - Zborniki: *JTDH 2016, Obzorja 2016*
 - *Revije: Linguistica, Jezik in slovstvo, Jezikoslovni zapiski, Slavistična revija, Slovene Linguistic Studies, Slovenščina 2.0*
- Rezultati:
 - Pregledanih 751 člankov, 133 (18 %) jih omenja vire
 - Od 133 jih 25 (19 %) vira ne citira, ostali ga, na več načinov
 - Samo URL 60 (45 %), publikacijo o viru (in mogoče tudi URL) 24 (18 %)
 - Najmanj raznotero v *Slavistični reviji SR*, najbolj raznotero v *Jezikoslovnih zapiskih*
 - Najbolj pomembna so navodila avtorjem

Korpus GigaFida. Na spletu.

Krek, S., Erjavec, T., Dobrovoljc, K., Može, S., Ledinek, N. in Holz, N. (2015): Training corpus ssj500k 1.4. Dostopno prek: <http://hdl.handle.net/11356/1052>.

Austinski principi (1)

- *Slabo zastopano: izobraževanje, navodila za avtorje in uredniška politika publikacij*
 - **Pomembnost**
 - *Podatki bi morali biti legitimen rezultat raziskav in jih je obvezno citirati*
 - **Dokazi**
 - *V znanstvenih objavah, kadarkoli in kjerkoli neka trditev sloni na podatkih, bi morali biti ti podatki ustrezno citirani*
- *SICRIS, SCOPUS, COBISS*
 - **Pomembnost**
 - *Citati podatkov bi za merjenje raziskovalčeve znanstvene uspešnosti morali biti enako pomembni, kot so to citati objav*
 - **Priznanje zaslug in avtorstva**
 - *Citiranje podatkov bi moralo služiti priznavanju znanstvenih zaslug, normativnega ter pravnega avtorstva vsem, ki so prispevali k njihovi izdelavi*

Austinski principi (2)

- *CLARIN.SI, (ADP):*
 - **Nedvoumna identifikacija**
 - *Citiranje podatkov naj bi vsebovalo trajno metodo identifikacije, primerno za strojno obdelavo, mednarodno edinstveno in široko sprejeto v skupnosti*
 - **Dostop**
 - *Citiranje podatkov naj bi pripomoglo k dostopu do samih podatkov in do povezanih metapodatkov, dokumentacije, programske opreme in drugih materialov, ki so potrebni, da tako ljudje kot računalniki te podatke informirano uporabljajo*
 - **Trajnost**
 - *Enoznačni identifikatorji in metapodatki, ki opisujejo podatke, morajo biti trajni, celo bolj kot sami podatki*
 - **Specifičnost in preverljivost**
 - *Citiranje podatkov naj bi pripomoglo identifikaciji, dostopu in preverjanju specifičnih podatkov, ki podpirajo neko trditev*
 - *Citiranje ali metapodatki citiranja naj bi vsebovali podatke o izvoru in stabilnosti v zadostni meri, da omogočijo preverbo, da je specifičen časovni okvir, različica ali del podatkov, ki so bili naknadno prevzeti, enak kot podatki, ki so bili izvirno citirani*
 - **Interoperabilnost in fleksibilnost**
 - *Metode za citiranje podatkov naj bi bile fleksibilne v zadostni meri, da omogočajo različne prakse med skupnostmi, vendar se ne smejo razlikovati v tolikšni meri, da ogrozijo interoperabilnost praks citiranja podatkov med skupnostmi*

Zaključki

- preverljivost, ponovljivost in nadgrajevanje prejšnjih raziskav, temelji odprte znanosti
- ozaveščanje, izobraževanje, podpora v skupnosti
- smernice za navajanje in seveda pripravo jezikovnih virov, univerzitetni študijski programi,
- izobraževalni dogodki in spletna gradiva: CLARIN.SI (iščemo!)
- kako je z eksperimentalnim in računalniškim jezikoslovjem v Sloveniji?