

Prepoznavanje enot v bibliografskih podatkih

Vladimir Batagelj

UP FAMNIT Koper in IMFM Ljubljana

1336. sredin seminar

Ljubljana, 26. julij 2023

- 1 Uvod
- 2 Enote v bibliografskih podatkih
- 3 Problemi
- 4 Viri



Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Tekoča različica prosojnic (July 26, 2023 at 14:04): [PDF](#)
<https://github.com/bavla/biblio/>

UP FAMNIT, ULj FF, IMFM; triletni (oktober 2022-2025)

- 1 Poiskati zanimive primere bibliografskih storitev višje stopnje za razne vrste uporabnikov. Razvoj nekaj prototipnih rešitev.
- 2 Razvoj metod in algoritmov za kakovostno prepoznavanje bibliografskih enot (na osnovi analize bibliografskih omrežij). Ti so osnova za pridobivanje visokokakovostnih bibliografskih podatkov za nadaljnje analize.
- 3 Nadaljnji razvoj metodologij in algoritmov za analizo bibliografskih omrežij, ki temeljijo na naših preteklih raziskavah (dvovrstna omrežja, deležni pristop, časovna omrežja in časovne količine).

Dosedanje delo na analizi bibliografskih podatkov

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Bibliografska podatkovja so bogat vir zanimivih omrežij.

- ① SPC 1991, Erdos 2000 [4, 8, 17]
- ② Dva članka za IS 2002: omrežja iz besedil [10, 11]
- ③ Normalizacije: Slovenski časopisi [2], Reuters 11 september [9]
- ④ Matjaž Zaveršnik, otoki SOM, Geom [18], SPC Patenti [1]
- ⑤ Amazon [12]
- ⑥ Social networks WoS2Pajek, Vizards [19]
- ⑦ FDV, analiza slovenske znanosti
- ⑧ Doktorski (Cerinšek, Bodlaj, Praprotnik), projekt GReGAS [5, 14, 15, 13], Španci
- ⑨ COST Peere [6]
- ⑩ Daša [20, 22, 3, 7, 21, 23]
- ⑪ Nataša [24]

Viri bibliografskih podatkov

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Gradivo o pripravi bibliografskih podatkov za analize se mi je nabiralo postopoma ob reševanju sprotih težav. Prvič sem ga uredil za predavanje za doktorske študente v Uppsali leta 2016. Izpopolnjeni različici sva pripravila skupaj z Dašo (Daria Maltseva) za delavnico "Analysis of bibliographic networks" na konferenci NetGlow (Networks in the Global World) v St. Petersburgu, 4-6. julija 2018 in za 10. mednarodno poletno šolo "Analysis of Scientific Networks" v Voronovem pri Moskvi, 15-21. julija 2019.

Bibliografska omrežja

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Do bibliografskih podatkov lahko pridemo na razna načine. Na primer iz:

- seznama virov v preglednem članku ali knjigi (**ReM**, **MCO**, **CoD**)
- področnih bibliografij (**BibT_EX**)
- bibliografskih storitev: **Web of Science**, **Scopus**, **SICRIS**, **CiteSeer**, **Zentralblatt MATH**, **Google Scholar**, **DBLP Bibliography**, **US patent office**, **IMDb**, in drugih.

Iz njih lahko pridobimo različna dvovrstna omrežja. Ponavadi omrežji dela \times avtorji (**WA**) in dela \times gesla (**WK**).

Pri opisno bogatejših virih dobimo vsaj še omrežje dela \times kode (**WC**) – klasifikacija (npr. MOS) in enovrstno omrežje sklicevanj delo \times delo (**Ci**). Pri tem dela vsebujejo članke, knjige, poročila, patente, itd. Pristop je uporaben tudi za filme in glasbo.

Poleg tega dobimo še vsaj razbitje po revijah ali založbah in po letih objave ter vektor števila strani.

Enote v bibliografskih podatkih

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

O enotah v bibliografskih podatkih so začeli razmišljati že zgodaj. Med brskanjem po spletu sem naletel na zanimiv vir iz leta 1967 [16].

Opis dela v zbirki bibliografskih podatkov je praviloma veliko popolnejši kot je opis dela v seznamu virov.

To poraja težave pri pripravi podatkov za analize, pri katerih želimo upoštevati tudi sklicevanja med deli (omrežje **Ci**).

Opisi del v virih

Prepoznavanje enot

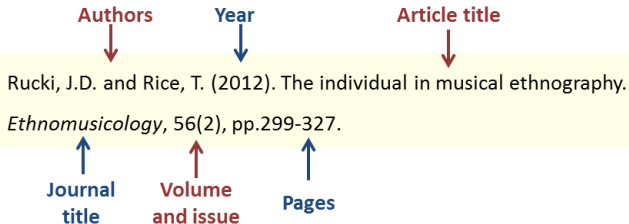
V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri



... Opisi del v virih

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

To JUUL or Not to JUUL: Health Risks Associated with e-Cigarettes and Marketing to Youth

by facing a public health crisis due to electronic cigarette use among young adults, e-cigarettes, commonly known as e-cigarettes or vape pens, are small devices that heat flavored nicotine or THC, along with other chemicals and flavors.

REFERENCE LIST

sentence case → last name and initials of the authors & = "and"

(2013). → year

Metal and silicate particles including nanoparticles → title of the article

are present in electronic cigarette cartomizer fluid and aerosol.

***PloS One*, 8(3), 1-11. <https://doi.org/10.1371/journal.pone.0057987>**

academic journal → a tool to help on and limit the sale of new smokers (Zhu et al., 2014).

volume number

issue number → new and flavors, and young and

page range

digital object identifier (DOI) → signs for product

[051670](https://doi.org/10.1371/journal.pone.0057987)

Zhu, S., Sun, J. Y., Bonneville, C., Cummins, S. E., Gamst, A., Yin, L., & Lee, M. (2014). Four signs for product

[136/tobaccocontrol-2014-](https://doi.org/10.1371/journal.pone.0057987)

Slog APA

Prepoznavanje
enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Book:

Author, A.A. (Year of Publication). *Title of work: Capital letter also for subtitle.* Location: **Publisher.**

Journal Article from a Database (without a doi):

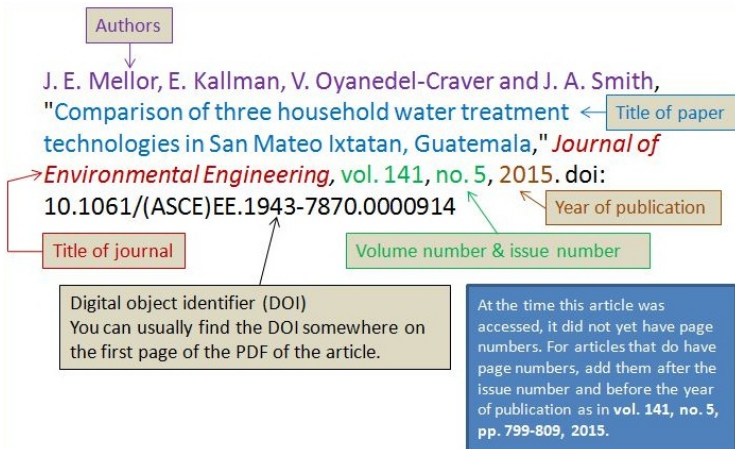
Author, A.A., & Author, B.B. (Year of Publication). *Title of article. Title of Journal, volume number (issue number), page range.* Retrieved from <http://www.someaddress.com/full/url/>

Journal Article from a Database (with a doi):

Author, A.A., & Author, B.B. (Year of Publication). *Title of article. Title of Journal, volume number (issue number), page range.* doi: 000000/000000000000

Non-periodical Web Document, Web Page, or Report (Website)

Author, A.A. & Author, B.B. (Date of publication). *Title of document.*
Retrieved from <http://www.someaddress.com/full/url/>



Vrste enot

Prepoznavanje enot

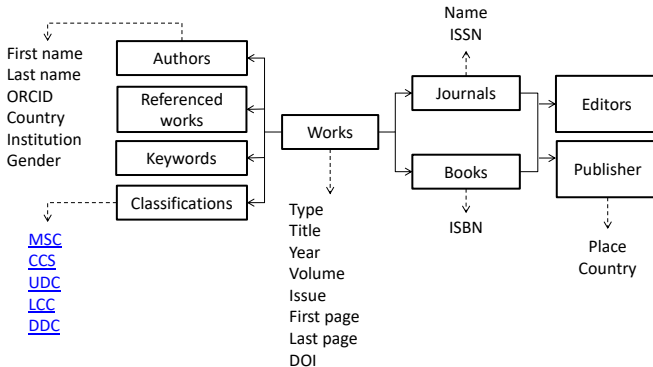
V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri



Osnovne vrste enot po IFLA-LRM [25]

Prepoznavanje enot

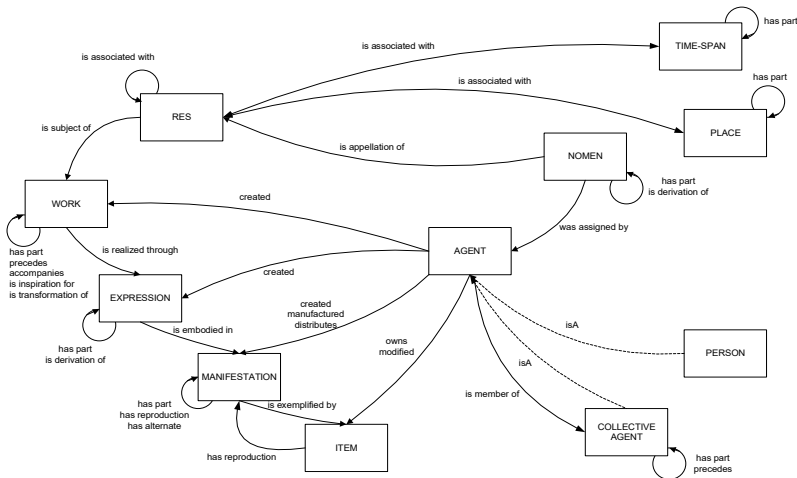
V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri



```
@Article{int:Mizuno1,
  author =      "S. Mizuno",
  title =      "An  $O(n^3L)$  algorithm using a sequence for
                linear complementarity problems",
  journal =     "Journal of the Operations Research Society of Japan",
  volume =     "33",
  year =       "1990",
  pages =      "66--75",
}

@InCollection{int:Vorst1,
  author =      "{J. G. G. van de} Vorst",
  title =      "An attempt to use parallel computing in large scale
                optimisation",
  booktitle =   "Logistics, Where Ends Have to Meet~: Proceedings of
                the Shell Conference on Logistics in Apeldoorn, The
                Netherlands, November 1988",
  editor =      "{C. F. H. van} Rijn",
  year =       "1989",
  pages =      "112--119",
  publisher =   "Pergamon Press",
  address =     "Oxford, United Kingdom",
}
```

Bib2Pajek.py

Zapis iz DBLP

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

```
<article mdate="2004-01-15" key="journals/arscom/BeinekeGL97">
<author>Lowell W. Beineke</author>
<author>Wayne Goddard</author>
<author>Marc J. Lipman</author>
<title>Graphs with Maximum Edge-Integrity.</title>
<year>1997</year>
<volume>46</volume>
<journal>Ars Comb.</journal>
<url>db/journals/arscom/arscom46.html#BeinekeGL97</url>
</article>

<inproceedings mdate="2004-12-09" key="conf/sigcse/BermanD96">
<author>A. Michael Berman</author>
<author>Robert C. Duvall</author>
<title>Thinking about binary trees in an object-oriented world.</title>
<pages>185-189</pages>
<year>1996</year>
<crossref>conf/sigcse/1996</crossref>
<booktitle>SIGCSE</booktitle>
<ee>http://doi.acm.org/10.1145/236536</ee>
<url>db/conf/sigcse/sigcse1996.html#BermanD96</url>
</inproceedings>
```

DBLP2Pajek.py

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

```

an  00549739
ai  gross.mark-d
is  ISSN 0025-5874; ISSN 1432-1823
au  Gross, Mark
py  1993
cc  *14M15 14J15
ti  Surfaces of bidegree  $(3,n)$  in  $\mathbb{P}^3$ .
ut  congruence; family of lines
so  Math. Z. 212, No.1, 73-106 (1993).
an  01488230
ai  tiras.yuecel; harmanci.abdullah; -
is  ISSN 0092-7872; ISSN 1532-4125
au  Tiras, Y. "ucel; Harmanci, Abdullah; Smith, P.F.
py  2000
cc  *13A15 13C05
ti  Some remarks on dense submodules of multiplication modules.
ut  multiplication module; dense submodule
so  Commun. Algebra 28, No.5, 2291-2296 (2000).
se  00000057 Communications in Algebra Commun. Algebra 0092-7872; 1532

```

ZBml.py

Prepoznavanje
enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

```
PT J
AU Dipple, H
   Evans, B
TI The Leicestershire Huntington's disease support group: a social network
   analysis
SO HEALTH & SOCIAL CARE IN THE COMMUNITY
LA English
DT Article
C1 Rehabil Serv, Troon Way Business Ctr, Leicester LE4 9HA, Leics, England.
RP Dipple, H, Rehabil Serv, Troon Way Business Ctr, Sandringham
   Suite,Humberstone Lane, Leicester LE4 9HA, Leics, England.
CR BORGATTI SP, 1992, UCINET 4 VERSION 1 0
   FOLSTEIN S, 1989, HUNTINGTONS DIS DISO
   SCOTT J, 1991, SOCIAL NETWORK ANAL
NR 3
TC 3
PU BLACKWELL SCIENCE LTD
PI OXFORD
PA P O BOX 88, OSNEY MEAD, OXFORD OX2 ONE, OXON, ENGLAND
SN 0966-0410
J9 HEALTH SOC CARE COMMUNITY
JI Health Soc. Care Community
PD JUL
PY 1998
VL 6
IS 4
BP 286
EP 289
PG 4
SC Public, Environmental & Occupational Health; Social Work
GA 105UP
UT ISI:000075092200008
ER
```

WoS2Pajek

Problemi pri pripravi bibliografskih podatkov

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Večina izvornih bibliografskih podatkov je polstrukturiranih — na voljo so v obliki zapisov iz neke baze podatkov. Izbrana polja v zapisu predstavljajo različne enote: imena ljudi, imena revij, ključne besede, ID del, države, ustanove ... Na žalost imena teh enot običajno niso shranjena v standardizirani obliki. Pri prepoznavanju posameznih enot se pojavita dve težavi

Sinonimija / sopomenke: Različni imeni določata isto enoto.

Homonimija / dvournost: Več enot z enakim imenom.

Problem 1. opisi del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

- natančnost opisa (seznam upoštevanih lastnosti del)
- popolnost opisa (ali so vnešeni vsi podatki?)
- **Slogi opisa virov**: obstaja več slogov podprtih od raznih strokovnih združenj: **MLA** (Modern Language Association of America), **APA** (American Psychological Association), **Chicago** (University of Chicago Press), **AMA** (American Medical Association), **SCE** (Council of Science Editors), **GOST** (Ruski državni standard).
- večina bibliografskih podatkovnih baz podpira vsaj en znakovni način izpisa izbranih zapisov. Pri izpopolnjevanju podatkovja, ki ga nameravamo uporabiti v analizah, je večkrat potrebno pretvarjati med različnimi oblikami zapisov.

... Problem 1. opisi del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

source

APA:

White, H. (2008). Identity and Control: How Social Formations Emerge (Second Edition). PRINCETON; OXFORD: Princeton University Press.

MLA:

White, Harrison C. Identity and Control: How Social Formations Emerge (Second Edition). Princeton University Press, 2008.

Chicago:

White, Harrison C. Identity and Control: How Social Formations Emerge (Second Edition). PRINCETON; OXFORD: Princeton University Press, 2008.

GOST:

White H. C. Identity and control. { Princeton University Press, 2008.

... Problem 1. opisi del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Graph products

Opis del v seznamu virov iz knjige ali preglednega članka običajno vsebuje naslednje sestavine:

- 1 Imena avtorjev; večkrat nepopolna (et al., samo začetnice imen)

WASSERMAN S, 1994, SOCIAL NETWORK ANAL

za

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press

- 2 Naslov
- 3 Leto ali datum objave

za članke še:

- 1 Revija
- 2 Knjiga? Letnik?
- 3 Zvezek
- 4 Strani

in za knjige: izdajatelj (podjetje, kraj)

... Problem 1. opisi del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Bibliographic data formats:

BibTex (LaTeX style):

```
@book{white2012identity,  
  title={Identity and control},  
  author={White, Harrison C},  
  year={2012},  
  publisher={Princeton University Press}  
}
```

EndNote (Clarivate Analytics):

```
%0 Book  
%T Identity and control  
%A White, Harrison C  
%D 2012  
%I Princeton University Press
```

... Problem 1. opisi del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

RIS (Research Information Systems style):

TY - BOOK
TI - Identity and Control
AU - White, Harrison C.
AB - <p>In this completely revised edition ...</p>
PB - Princeton University Press
PY - 2008
SN - 9780691137155
T1 - How Social Formations Emerge (Second Edition)
UR - <http://www.jstor.org/stable/j.ctt1r2fg1>
ER -

Problem 2: Različne kulture na posameznih področjih

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

- 1 število soavtorjev; doktorand mora imeti samostojne objave, **credit**
- 2 zapis imen avtorjev (začetnice/polna imena)
- 3 vrstni red sestavin avtorjevega imena (Francozi, Španci, Arabci, Rusi, itd., dodatki de, van, ibn, Prof, Dr, itd.).
- 4 nekatere revije imajo posebna pravila o krajšanju imen revij

Bon G., 1896, CROWD STUDY POPULAR

Le Bon G., 1897, CROWD STUDY POPULAR

LeBon G., 1960, CROWD STUDY POPULAR

Lebon G., 2011, PSIHOLOGIJA NARODOV

Le Bon Gustave, 1930, CROWD STUDY POPULAR

Gustave Le Bon, 1982, PSYCHOL MASSEN

Newman, M. E. (2001). Scientific collaboration networks.
II. Shortest paths, weighted networks, and centrality.
Physical review E, 64(1), 016132.

M.E.J. Newman, preceding paper, Phys. Rev. E 64, 016131 (2001).

Problem 2: Različne kulture

Različni opisi virov

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

1 Običajni opis AU (PY). TI. JI, BP-EP

Freeman, L. C., & White, D. R. (1993). Using Galois lattices to represent network data. *Sociological methodology*, 127-146.

2 Številka članka AU (PY). TI. JI, VL(IS), BP

Newman, M. E. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, 64(1), 016132.

3 Brez naslova članka AU, JI VL, IS (PY)

P. Erdos and A. Renyi, *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17 (1960).

Problem 3: Zapis podatkov

Prepoznavanje
enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Velike/male črke – pretvorba v izbrano obliko.

ASCII / Unicode

Cyrillic to Latin (Unicode, programska pretvorba v ASCII)

Prevedba je odvisna od jezika, kateremu se ime prilagaja.

Пётр Ильич Чайковский

angleško: Pyotr Ilyich Tchaikovsky

nemško: Pjotr Iljitsch Tschaikowski

francosko: Piotr Ilitch Tchaïkovski

špansko: Piotr Ilich Chaikovsky

italijansko: Pëtr Il'ič Čajkovskij

slovensko: Peter Iljič Čajkovski

Russian

Problem 4. Nepopolni podatki

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

- **manjkajoči podatki:** nekatere sestavine opisa manjkajo ali so nepopolni (dodatni avtorji, naslov, letnik, letnik, zvezek, strani itd.). Podatke ročno dopolnimo.
- **Manjkajoči podatki:** manjkajo opis del pomembnih za obravnavano temo, ki v pridobljenem podatkovju niso zajeta (v začetnem obdobju področja se je uporabljalo drugo izrazje). Če imamo omrežje sklicevanj, jih lahko odkrijemo z njegovo analizo in dodamo polne opise našemu podatkovju.

Za male bibliografije, kjer lahko pregledamo, sprejmemo in "popravimo" vsak opis dela zadošča tabelarni pristop (Excel).

V bazi podatkov so lahko tudi napake (tipkarske napake) – popravimo jih v svojem izvodu podatkov.

Problem 5. Prepoznavanje enot imena avtorjev

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Lorenzo Bartolini iz **Letters to Juliet**.

Obstaja veliko načinov pisanja imena. Nekatere baze bibliografskih podatkov določijo imenu ustreznega avtorja pri vnosu podatkov in mu pripešejo enolično oznako (DBLP, ZB, ResearcherId). Kitajci 100 priimkov. **MathSciNet**; **Orcid** - vnesi ime avtorja v iskalno okence; **Scopus**; **eLibrary** - klikni na ime avtorja.

<https://orcid.org/0000-0002-0240-9446>

https://elibrary.ru/author_items.asp?authorid=155240

Za določitev različic v osebnih imenih uredimo imena glede na priimke.

Pri avtorjih z imeni zapisanimi v različnih abecedah prevedemo zapise v izbrano abecedo ali uporabimo "slovar" enakovrednih zapisov.

Pristop **AMS**

... Problem 5. Prepoznavanje enot imena avtorjev

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

V podatkovju zbMATH: Borštnik, N. S. Mankoč; Mankoč Borštnik, N.; Mankoč-Borštnik, Norma; Mankoč Borštnik, Norma Susana; Mankoc-Borstnik, N.S.; Mankoč Borštnik, N.S..

Prav tako v zbMATH se objave avtorja Smith, John W. pojavljajo od leta 1868 do leta 2007.

V bazi MathSciNet je vsaj 623 različnih matematikov z imenom Zhang, Li.

V zbMATH je na primer O'Regan, Donal zapisan kot oregon.donal (kanonsko ime, ZBunified) in drugič o'regan.d. Včasih ni uporabljano kanonsko ime. V teh primerih naš program predela polno ime O'Regan, Donal v o'regan.d.

Podobno velja za avtorja Pečarić, Josip E, ki ima kanonsko ime pecaric.josip-e. Kadar ni uporabljeno njegovo kanonsko ime, dobimo pecaric.j in pecaric.j-e iz dveh polnih imen Pečarić, J. in Pečarić, J. E.

... Problem 5. Prepoznavanje enot imena avtorjev

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

V BibTEXovi "Computational Geometry Database" (Jones 2002) ima isti avtor sedem različnih imen Č R.S. Drysdale, Robert L. Drysdale, Robert L. Scot Drysdale, R.L. Drysdale, S. Drysdale, R. Drysdale, in R.L.S. Drysdale.

Krivoshe\u\i n, Leonid Evgen\cprime evich ima v Mathematical Reviews Database 20 različnih zapisov imena.

Poznavanje področja je potrebno, da vemo, da sta Otfried Schwarzkopf in Otfried Cheong ali Mikhail Efimovich Tylkin in Michel Marie Deza ista oseba.

Vir problemov so tudi zapisi vzhodnoevropskih imen. Najbrž sta Krachkovskij, A. P. in Krachkovskii, A. P. ista oseba.

... Problem 5. Prepoznavanje enot

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

id	canon	alter1	alter2	alter3	alter4	alter5
ORCID:0000-0002-0240-9446	Batagelj, Vladimir	Batagelj, Vlado	Batagelj, V	MR:32440	Scopus: 56037441100	
Scopus: 35615877200	Batagelj, Valentin	Batagelj, V				
ORCID:0000-0003-4467-7075	Cheong, Otfried	Schwarz kopf, Otfried	Cheong, O	Schwarzkopf, O	Scopus: 57191986875	
MR:57370	Deza, Michel-Marie	Deza, MM	Deza, M	Deza, Mikhail	Scopus: 7003745115	
MR:57370	Deza, Michel-Marie	Tylkin, Mikhail Efimovich	Tylkin, ME	Тылкин, Михаил Ефимович	Тылкин, ME	Де́за, Мишел
ORCID:0000-0002-4294-9017	Zweig, Katharina Anna	Zweig, KA	Zweig, K	Lehmann, Katharina Anna	Lehmann, K	Scopus 592816 2000
eLib:696348	Maltseva, Daria	Мальцева, Дарья Васильевна	Мальцева, ДВ	Мальцева, Д	Maltseva, DV	
eLib:155240	Maltseva, Diana	Мальцева, Диана Васильевна	Мальцева, ДВ	Мальцева, Д	Maltseva, DV	

Problem 5. Prepoznavanje enot imena del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Imena ISI: pojavljajo se v polju CR v WoS

LEFKOVITCH LP, 1985, THEOR APPL GENET, V70, P585

in imajo naslednjo zgradbo

AU + ', ' + PY + ', ' + SO[:20] + ', V' + VL + ', P' + BP

V WoSu ima lahko isto delo več imen ISI.

GRANOVET.MS, 1973, AM J SOCIOL, V78, P1360

GRANOVETTER M, 1983, SOCIOLOGICAL THEORY, V1, P203

BORGATTI SP, 2002, UGINET WINDOWS SOFTW

BORGATTI S, 1999, UCINET V USERS GUIDE

CANTANZARO M, 2005, PHYS REV E, V71, UNSP 027103

CANTAZARO M, 2005, PHYS REV E, V71, UNSP 056104

CATANZARO M, 2005, PHYS REV E 2, V71, ARTN 056104

Problem 5. Prepoznavanje enot imena del

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Za povečanje natančnosti smo v programu WoS2Pajek uvedli **kratka imena** v obliki:

```
LastNm[:8] + ' ' + FirstNm[0] + '(' + PY + ')' + VL + ':'
```

Na primer: LEFKOVIT L(1985)70:585.

Iz priimkov z VAN, DE, itd. so izločeni presledki.

Problem 5. Prepoznavanje enot imena revij in knjig

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

V WoSovih opisih virov najdemo na primer naslednja imena revij:
NUCLEIC ACIDS RES, NUCL ACIDS RES, NUCLEIC ACIDS RES S,
NUCLEIC ACIDS RES S2, NUCL ACID RES, NUCL ACIDS RES S2,
NUCL ACIDS S SER, NUCL ACIDS RES S, NUCL AC RES,
NUCLEIC ACIDS RES S1, Nucleic Acids Res, NUCL ACIDS RES S1
in
Q J R MET SOC, Q J R METEOROL SOC, Q J ROY METEOR SO
S1, Q J ROY METEOR SOC, Q J ROY METEOR SOC B, QUART
J ROY METEOR S, QUART J ROY METEOROL, QUART J ROY
METEOROL SOC, QUART J ROYAL METEOR.

Ali pripadajo eni sami reviji, ali večim?

Problem 5. Prepoznavanje enot imena revij in knjig

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

Ustvarimo ključe iz prvih (dveh) črk imena revij in uredimo pare (ključ, revija). Po "čiščenju" seznama revij v podatkovju zbMATH se je število revij zmanjšalo z 3158 na 2665.

Podatki o knjigah. Pomagamo si lahko z International Standard Serial Number **ISSN**; Digital Object Identifier **DOI**; International Standard Book Number **ISBN**.

Pri prepoznavanju revij si lahko pomagamo z Global Serials Directory (Ulrichsweb 2013) in Journal Abbreviation Sources (JAS 2013) ter druge vire na spletu.

Problem 5. Prepoznavanje enot gesla

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri

Včasih so gesla (ključne besede) navedena v podatkih, sicer jih lahko poskusimo pridobiti iz besedila (naslov, povzetek). Pri določanju gesel v besedilu izpustimo *nepomembne besede* (stopwords). Sopomenke prepoznamo z geslenjem (lemmatization) ali/in uporabo slovarjev. Večbesedna gesla.

Na primer izrazi 'function', 'map', 'mapping', and 'transformation' so v matematiki enakovredni. Podobno je, kadar imamo opravka z več jeziki. To lahko razrešimo z uporabo slovarjev.

Drug vir sopomenk so slovnična pravila jezika. Tako na primer se v angleščini 'go' pojavlja v besedilu kot 'go', 'goes', 'gone', 'going', and 'went'. Problem razrešujemo z geslenjem, ki ga podpirajo knjižnice za analizo besedil, kot sta NLTK (Bird et al. 2009; Perkins 2010) in MontyLingua (Liu 2004).

[Github Bavla/biblio](#)

Acknowledgments

Prepoznavanje
enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J5-2557, J1-2481 and J5- 4596), and prepared within the framework of the COST action CA21163 (HiTEc).



Vladimir Batagelj.

Efficient algorithms for citation network analysis, 14 Sep 2003.

URL: <https://arxiv.org/abs/cs/0309023>,

[doi:https://doi.org/10.48550/arXiv.cs/0309023](https://doi.org/10.48550/arXiv.cs/0309023).



Vladimir Batagelj.

Pajek, 2001.

Dagstuhl seminar.

URL:

<http://vlado.fmf.uni-lj.si/pub/networks/doc/dagstuhl/kazalo.htm>.



Vladimir Batagelj.

On fractional approach to analysis of linked networks.

Scientometrics, 123(2):621–633, 2020.

[doi:10.1007/s11192-020-03383-y](https://doi.org/10.1007/s11192-020-03383-y).



Vladimir Batagelj.

Some mathematics of network analysis, January 1991.

Network seminar, Department of Sociology, University of Pittsburgh.

URL:

<http://vlado.fmf.uni-lj.si/pub/networks/data./cite/report.pdf>.

References II

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri



Vladimir Batagelj and Monika Cerinšek.
On bibliographic networks.

Scientometrics, 96(3):845–864, 2013.

[doi:10.1007/s11192-012-0940-1](https://doi.org/10.1007/s11192-012-0940-1).



Vladimir Batagelj, Anuška Ferligoj, and Flaminio Squazzoni.

The emergence of a field: A network analysis of research on peer review.

Scientometrics, 113(1):503–532, 2017.

[doi:10.1007/s11192-017-2522-8](https://doi.org/10.1007/s11192-017-2522-8).



Vladimir Batagelj and Daria Maltseva.

Temporal bibliographic networks.

J. Informetr., 14(1):Article No. 101006, 2020.

[doi:{10.1016/j.joi.2020.101006}](https://doi.org/10.1016/j.joi.2020.101006).



Vladimir Batagelj and Andrej Mrvar.

Some analyses of erdos collaboration graph.

Social Networks, 22:173–186, 2000.

References III

Prepoznavanje enot

V. Batagelj

Uvod

Enote v bibliografskih podatkih

Problemi

Viri



Vladimir Batagelj and Andrej Mrvar.

Density based approaches to network analysis: Analysis of reuters terror news network.

In *Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003)*, August 27, 2003.

URL:

<http://www.cs.cmu.edu/~dunja/LinkKDD2003/papers/Batagelj.pdf>.



Vladimir Batagelj, Andrej Mrvar, and Matjaž Zaveršnik.

Network analysis of dictionaries.

In *Proceedings B of the 5th International Multi-Conference IS'2002 / Language Technologies*. Ljubljana, 2002.

URL: <http://nl.ijs.si/isjt02/zbornik/sdjt02-24abatagelj.pdf>.



Vladimir Batagelj, Andrej Mrvar, and Matjaž Zaveršnik.

Network analysis of texts.

In *Proceedings B of the 5th International Multi-Conference IS'2002 / Language Technologies*. Ljubljana, 2002.

URL: <http://nl.ijs.si/isjt02/zbornik/sdjt02-24bbatagelj.pdf>.

References IV

Prepoznavanje
enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri



Vladimir Batagelj, Kejžar Nataša, and Simona Korenjak-Černe.
Analysis of the customers' choice networks: An application on amazon
books and cds data.
Metodološki zvezki/Advances in Methodology and Statistics, 4(2):191–204,
2007.



Vladimir Batagelj and Selena Praprotnik.
An algebraic approach to temporal network analysis based on temporal
quantities.
Soc. Netw. Anal. Min., 6(1):28:1–28:22, 2016.
[doi:10.1007/s13278-016-0330-4](https://doi.org/10.1007/s13278-016-0330-4).



Jernej Bodlaj and Vladimir Batagelj.
Network analysis of publications on topological indices from the web of
science.
Molecular informatics, 33(8):514–535, 2014.
[doi:10.1002/minf.201400014](https://doi.org/10.1002/minf.201400014).



Monika Cerinšek and Vladimir Batagelj.
Network analysis of Zentralblatt MATH data.
Scientometrics, 102(1):977–1001, 2015.
[doi:10.1007/s11192-014-1419-z](https://doi.org/10.1007/s11192-014-1419-z).

References V

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri



Ann T. Curran and Henriette D. Avram.
The identification of data elements in bibliographic records, 1967.
URL: <https://apps.dtic.mil/sti/citations/AD0666447>,
[doi:AD0666447](https://doi.org/10.26434/chemrxiv-2019-08-01).



Jerry Grossman.
The Erdős number project, 2022.
URL: <https://sites.google.com/oakland.edu/grossman/home/the-erdoes-number-project>.



B. Jones.
Computational geometry database, 2002.
URL: <ftp://ftp.cs.usask.ca/pub/geometry/>.



Hugo Liu.
Montylingua: A free, commonsense-enriched natural language understander for english (version 2.1), 2004.
URL: <http://alumni.media.mit.edu/~hugo/montylingua/>.

References VI

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri



Daria Maltseva and Vladimir Batagelj.
Social network analysis as a field of invasions: Bibliographic approach to
study SNA development.

Scientometrics, 121(2):1085–1128, 2019.

[doi:10.1007/s11192-019-03193-x](https://doi.org/10.1007/s11192-019-03193-x).



Daria Maltseva and Vladimir Batagelj.
imetrics: the development of the discipline with many names.

Scientometrics, 125(1):313–359, 2020.

[doi:10.1007/s11192-020-03604-4](https://doi.org/10.1007/s11192-020-03604-4).



Daria Maltseva and Vladimir Batagelj.
Towards a systematic description of the field using keywords analysis: Main
topics in social networks.

Scientometrics, 123(1):357–382, 2020.

[doi:10.1007/s11192-020-03365-0](https://doi.org/10.1007/s11192-020-03365-0).



Daria Maltseva and Vladimir Batagelj.
Journals publishing social network analysis.

Scientometrics, 126(1):3593–3620, 2021.

[doi:10.1007/s11192-021-03889-z](https://doi.org/10.1007/s11192-021-03889-z).

References VII

Prepoznavanje enot

V. Batagelj

Uvod

Enote v
bibliografskih
podatkih

Problemi

Viri



Nataliya Matveeva, Vladimir Batagelj, and Anuška Ferligoj.
Scientific collaboration of post-soviet countries: the effects of different
network normalizations.
Scientometrics, 128(1):4219–4242, 2023.



Maja Žumer.
Ifla library reference model (ifla lrm): Harmonisation of the frbr family.
Knowl. Org., 45(4):310–318, 2018.
[doi:10.5771/0943-7444-2018-4-310](https://doi.org/10.5771/0943-7444-2018-4-310).