

# Nadgradnja Zgodovinarskega indeksa citiranosti

Katja Meden\*, Ana Cvek†

\*Odsek za tehnologije znanja, Institut »Jožef Stefan«  
Jamova cesta 39, 1000 Ljubljana,  
katja.meden@ijs.si

†Institut za novejšo zgodovino, zunanja sodelavka  
Privoz 11, 1000 Ljubljana,  
ana.cvek3@gmail.com

## Povzetek

Začetki Zgodovinarskega indeksa citiranja segajo v leto 2003, ko so raziskovalci Inštituta za novejšo zgodovino začeli spremljati in sistematično popisovati citate za prijave projektov in programov na ARRS. Citatni indeks je v letih doživel nekaj nadgradenj, poskusov harmonizacije podatkov in prečiščevanja relacijskih baz, vendar je bilo v zadnjih letih ugotovljeno, da sistem ne zadostuje potrebam indeksatorjev in uporabnikov. Pred nadgradnjo smo izvedli analizo podatkov, kjer so se identificirale največje težave. Nadgradnja je potekala v dveh delih, v prvem delu smo nadgradili administrativni del, v drugem delu pa spletno aplikacijo. Zgodovinarski indeks citiranja je bil med nadgradnjo tehnično posodobljen in s tem oblikovan tako, da je intuitiven za indeksatorje in uporabnike.

## The Historiography Citation Index Upgrade

The beginnings of the Historiography Citation Index reach into the year 2003 when the researchers of the Institute of Contemporary History felt the urge to begin systematically indexing citations for the intent of applying for funding of research projects and programs. Over the years HCI was a subject of several upgrades and data harmonization attempts. Despite this fact, the system was proven to have several shortcomings, and thus, another system update was needed. Before the update, extensive analysis was performed to identify the most problematic aspects of the system. The upgrade was performed in two parts. The first part consisted of administration system remodeling, which was followed by a frontend redesign. With the update, HCI was technically improved, which resulted in a more intuitive system to its indexators and users.

## 1. Uvod

Ocenjevanje uspešnosti raziskovalcev na področju humanistike je v primerjavi z drugimi raziskovalnimi področji, predvsem naravoslovnimi, že od samih začetkov precej prikrajšano. Med drugim ocenjevanje temelji na frekvenci citiranosti, te podatke pa pridobimo iz različnih citatnih indeksov, kot sta na primer Web of Science (v nadaljevanju WOS) in Scopus. Monografije so primarni produkt raziskovalnega dela na področju humanistike in družboslovja (Gläzel in Schoepflin, 1999; Huang in Chang, 2008; Nederhof, 2006; Hicks, 2004). V nasprotju z vrednotenjem raziskovalne uspešnosti v naravoslovju se ta področja teže vrednotijo, predvsem zaradi dejstva, da so monografije po večini bolj obsežne kot znanstveni članki (Kousha et al., 2011), in visokih kriterijev vključevanja publikacij v obstoječe indekse citiranja, na primer WOS in Scopus. Med pomembnejše kriterije spadajo redno izhajanje serijske publikacije, jezik publikacije, recenziranost, spoštovanje mednarodnih standardov (kot so informativni naslov, povzetek, popolna bibliografska informacija za vse citirane reference), poleg pogojev pa težavo predstavlja tudi indeksiranje monografij. Obstojeci citatni indeksi se namreč bolj osredotočajo na serijske publikacije. Web of Science zajema okoli 12.000 znanstvenih revij in samo okoli 50.000 monografij, medtem ko Scopus zajema več 21.500 znanstvenih revij in 113.000 znanstvenih monografij. Število monografij v indeksu Scopus odraža večji obseg monografij v primerjavi z WOS, pa vendar monografije v primerjavi s številom znanstvenih člankov v revijah predstavljajo zgolj zanemarljiv del citatnega indeksa (Južnič, 2017). ZIC trenutno vsebuje 4.837 vseh vnosov,

od tega 2.901 vnos serijskih publikacij in 1.936 vnosov monografij in poglavij iz monografij, kar predstavlja razmerje 59,9 % serijskih publikacij ter 39,1 % monografij in poglavij iz monografij. Neenakosti pri vključevanju publikacij v citatne indekse so na Inštitutu za novejšo zgodovino skušali zamejiti že v letu 2003. Raziskovalci so začutili potrebo po spremeljanju in sistematičnem popisovanju citatov za prijave projektov in programov, kar predstavlja zametek Zgodovinarskega indeksa citiranja (v nadaljevanju ZIC). Osnovni namen je bil ustvariti bazo citatov iz slovenskih zgodovinskih monografij, osrednjih znanstvenih časopisov in revij (Lazarević in Zemljic, 2003). Začetna shema baze, ki je bila precej enostavna, je ob svojem nastanku dobro služila potrebam raziskovalcev, vendar so se sčasoma pokazale pomanjkljivosti (Pančur et al., 2014), ki so vodile v nadaljnje nadgradnje, poskuse harmonizacije podatkov in prečiščevanja relacijskih baz. Zadnja nadgradnja je potekala leta 2012 in predstavlja osnovo in temelj nadgradnje, ki je predstavljena v nadaljnjem besedilu članka.

## 2. Cilji in potek nadgradnje

Pri postopku nadgradnje smo z uporabo sodobnih tehnologij in estetsko privlačne grafične podobe žeeli preoblikovati administratorski spletni vmesnik in indeksatorju omogočiti prijazno in pregledno izkušnjo pri urejanju podatkov. Najpomembnejši cilj nadgradnje je bila postavitev ZIC kot ločene aplikacije. Ker je baza MySQL trenutno integralni del portala SIStory in se upravlja s pomočjo skupne administracije, je treba podatkovno bazo ZIC postaviti kot ločeno aplikacijo na poddomeni portala SIStory. Razlog za to je načrtovana postavitev nove digitalne knjižnice por-

tala SIStory kot samostojnega repozitorija z ločeno administracijo. Poleg ločene baze in administracije smo pri nadgradnji upoštevali naslednje sklope problemov. V prejšnji nadgradnji uvoz in izvoz podatkov nista bila mogoča, zato smo želeli to omogočiti. Prav tako smo želeli, da je spletna aplikacija narejena modularno, kar bo omogočalo dodajanje novih funkcionalnih rešitev. Pri uporabniškem vmesniku smo želeli, da je stran prijazna za mobilne obiskovalce, pri iskalniku pa smo želeli doseči hitro in pregledno iskanje po podatkih. Nadgrajeni administracijski modul naj bi omogočal enostavnejši dostop in upravljanje vseh podatkov ter z gesлом zaščiten dostop do administracije. Izbrani osnovni podatki morajo biti z ustreznim vmesnikom prosto dostopni strojnemu zajemu podatkov (Pančur, 2019b).

Pri postavljanju ciljev in procesu nadgradnje smo izhajali iz temeljnih načel Raziskovalne infrastrukture slovenskega zgodovinopisja (v nadaljevanju RI INZ), ki vključujejo uporabo uveljavljenih in razširjenih tehnologij, ki jih člani infrastrukture dobro poznajo in obvladajo (načeli enostavnosti in poznavanja), modularno nadgrajevanje obstoječih tehnologij (načelo fleksibilnosti) in uporabo odprtih ali lastniških standardov (načelo odprtosti) (Pančur in Šorn, 2019). V procesu nadgradnje smo tako uporabljali tehnologije, ki jih priporoča RI INZ (Pančur, 2019a), in upoštevajo načeli enostavnosti in poznavanja HTML5 in CSS3, najnovejše verzije PHP, MySQL, ElasticSearch engine, JavaScript in Java knjižnice. Pomemben vidik nadgradnje je tudi vidik interoperabilnosti, ki se v svojem pomenu prepleta z načelom fleksibilnosti. Fleksibilnost in interoperabilnost sistema želimo doseči z implementacijo aplikacijskega profila MODS za uvoz in izvoz metapodatkov v različnih formatih, ki podpirajo nadaljnjo diseminacijo in izmenjavo podatkov z drugimi informacijskimi sistemi. Nadgradnja je potekala v posameznih sklopih, ki so opisani v nadaljevanju besedila.

### 3. Rezultati nadgradnje

Nadgradnja je potekala v dveh delih: prvi del se nanaša na administrativni sistem SIStory. Nadgradnja v tem delu zajema preoblikovanje mask in njihovih polj, postavitev nove XML sheme po standardu MODS za uvoz in izvoz podatkov, iskalnik, ki temelji na tehnologiji Elasticsearch, ter migracije vrednosti ločenih polj Avtor(ji). Drugi del se osredotoča na nadgradnjo spletne aplikacije in uporabniškega vmesnika. Pri programski nadgradnji smo sodelovali z zunanjimi sodelavci Infrastrukture.

#### 3.1. Administrativni sistem SIStory

##### 3.1.1. Maske za vnos podatkov

Glavna spremembra v administracijskem sistemu (admin) je prehod s prej enotne maske na dve ločeni. Enotna maska je vsebovala tri razdelke: *Splošni podatki*, *Podatki o viru* in *Vsebinska obdelava*. Vnos podatkov v maske poteka ročno, podatkovna polja v enotni maski pa so bila nejasna (npr. ponavljanje polja za vnos id številke COBISS, imena avtorja idr.), nekatere tudi brez pomena za potrebe citatnega indeksa. Tako je bil na primer razdelek Vsebinska obdelava za citatni indeks povsem neuporaben, saj vsak zapis vsebuje identifikatorje s povezavami na zapise publikacij (COBISS, SIStory) s polnim metapodatkovnim opisom.

Iz enotne maske sta nastali dve neodvisni maski za vnos podatkov v ZIC V2. Iz maske za vnos publikacije sta nastali dve - *maska za vnos monografij* in *maska za vnos serijskih publikacij*, ki dovoljujeta natančnejši opis glede na publikacijo, ki jo indeksiramo. Vsaka izmed mask, tako kot v prejšnji verziji, vsebuje tudi masko za vnos citatov. Maski so bile oblikovane na podlagi zaznanih težav v prejšnjem administracijskem sistemu, o katerih so poročali indeksatorji, ter na podlagi potreb za opis določene publikacije in citatnega indeksa. Spodnja tabela (Tabela 1) prikazuje polja oziroma metapodatke za opis posameznih del in citatov.

Večina elementov, potrebnih za opis publikacij, je ostala nespremenjena. Po opravljeni analizi elementov mask smo izpostavili ključna polja za potrebe opisa publikacij in njihovih citatov. Večina polj je splošne narave (npr. avtor, naslov, leto, kraj itd.), publikacije, ki jih vnašamo (monografije in serijske publikacije), pa se med seboj razlikujejo v določenih vidikih. Ločeni maski s prilagojenimi polji omogočata (z indeksatorskega vidika) kakovostnejšo indeksacijo publikacije. Elementi so bili spremenjeni ali prilagojeni, saj določeni niso bili ažurirani (na primer element Tipologija) ali niso omogočali dovolj natančnega opisa (element Avtor). Pri poljih Avtor in Urednik smo metapodatkovno polje ločili na dve polji: Ime in Priimek. S tem smo zagotovili natančnejši, bolj strukturiran opis in posledično boljše prikazovanje podatkov. Zaradi nove strukture polja je bilo za povezovanje vrednosti polj treba opraviti migracijo vrednosti iz starih, neločenih polj v nova, strukturno ločena polja v obliki Priimek, Ime (za namek prikaza). Nekaterih elementov iz stare maske v novih maskah nismo vključili, npr. Ključne besede ali Država, saj so bili za opis publikacij v citatnem indeksu nepotrebni. Dodani so bili tudi novi elementi, ki jih starejša maska za vnos podatkov ni vsebovala, ker ti podatki še niso bili potrebni. Tu govorimo predvsem o maski za vnos serijskih publikacij in citatov, kjer smo dodali polji DOI in URL, ki omogočata enoznačno, trajno identifikacijo, prav tako pa poleg polja Sistory ID uporabniku omogočata hiter dostop do publikacije. Pri analizi obstoječih zapisov se je izkazalo, da so zapisi pomanjkljivi in neenotni. Do takšnih napak je prihajalo predvsem zato, ker indeksatorji niso imeli nobenih konkretnih navodil in so publikacije v maski (glavni vnos in citat) vpisovali po lastni presoji. Zato smo se pri nadgradnji odločili, da indeksatorjem ponudimo pomoč, ki jim bo olajšala vnos podatkov, še bolj pomembno pa je, da bi s temi navodili oz. pomočjo radi zagotovili čim bolj enotno indeksacijo ter pravilnejše in natančnejše zapise v indeksu. Ob vsakem polju se pri vseh treh maskah nahaja opis polja z navodili za vnos in primeri, ki naj bi bili indeksatorju v pomoč oz. oporo pri vpisovanju podatkov. Tu velja poudariti, da se zavedamo, da bo do napak kljub pomoči še vedno prihajalo, saj se podatki vpisujejo ročno. S tem, da nudimo navodila za vnos, poskušamo zmanjšati število pogostih napak.

##### 3.1.2. Elasticsearch iskalnik in filtriranje

Iskalnik Elasticsearch je distribucijsko, odprtokodno in analitično orodje za vse vrste podatkov, skupaj z besedilnimi, številčnimi, geoprostorskimi, strukturiranimi in nestrukturiranimi podatki, in temelji na knjižnici Lucene Apa-

Tabela 1: Metapodatki mask za vnos podatkov.

Metapodatek	min/max. št	Podatkovni tip	Maska (Mono., Serijska, Citat)	Primer
Cobiss ID	0,1	ID	M,S,C	3278924
Sistory ID	0,1	ID	M,S,C	handle.net/11686/4320
ISBN	0,1	ID	M	987-961-3421-43
ISSN	0,1	ID	S	0353-0329
Jezik	1,1	ISO639-2b	M,S	slv - slovenski
Tipologija	1,1	COBISS tipologija	M,S	1.16 - Samostojni znan. sestavek
Tip	0,1	interni seznam	M	Poglavlje v monografiji
Avtorji	1,neomejeno	niz	M,S,C	Marko Zajc
Naslov	1,1	niz	M,S,C	Slovenski intelektualci in..
Vzporedni naslov	0,1	niz	M,S	Slovenian Intellectuals...
Naslov zbornika	0,1	niz	M	Slovenija v Jugoslaviji
Naslov vira	0,1	niz	S	Prispevki za novejšo zgodovino
Uredniki	0,neomejeno	niz	M	Zdenko Čepič (ur.)
Kraj	0,1	niz	M,S,C	Ljubljana
Založba	0,1	niz	M,S,C	Založba INZ
Leto	0,1	številčna vrednost	M,S,C	2015
Letnik	0,1	številčna vrednost	S,C	57
Številka	0,1	številčna vrednost	S,C	1
Zbirka	0,1	niz;št. vrednost	M	Vpogledi; 10
Stran	0,1	št. vrednost	M,S,C	241 - 256
DOI	0,1	ID	S,C	10.1090/019339135
Baza citatov INZ	0,1	gumb	M,S	DA
Citat na strani	1,1	št.vrednost	C	34
Vir	0,1	niz	C	Prispevki za novejšo zgodovino

che. Postopek zajema vrednosti se začne s tako ime novanim *data ingestion*, v katerem so surovi podatki zajeti v iskalnik iz različnih virov. V naslednjem delu se oblikuje ElasticSearch index, ki je zbirka med seboj povezanih dokumentov. Vsak izmed dokumentov je povezan s ključi (imena, podatkovna polja ali lastnosti) in njihovimi vrednostmi (niz, številke, Boolovi operatorji, nabor vrednosti ...), indeks sam pa omogoča izvajanje kompleksnih iskalnih poizvedb.

Za implementacijo iskalnika ElasticSearch za ZIC v administrativnem sistemu podatke zajamemo iz relacijske baze, ki temelji na tehnologiji MySQL (Elastic, nd). Indeksirani ključi so v tem primeru podatkovna polja, ki bodo služila iskalnim poizvedbam, in njihove vrednosti (ki so večinoma besedilni nizi ali številčne vrednosti). Iskalnik ponuja izvajanje kompleksnih iskalnih poizvedb, ZIC uporablja funkcijo *simple string query*.

```
GET /_search
{
  "query": {
    "simple_query_string" : {
      "query": "Mojca + Šorn +
\"Življenje Ljubljjančanov
med drugo svetovno vojno\""
      "fields": ["title^5", "body"],
      "default_operator": "and"
    }
  }
}
```

Funkcija uporablja preprosto sintakszo za besedilne iskalne poizvedbe, na podlagi katere vrača iskalne rezultate z uporabo parserja.

Za iskalnik v spletni aplikaciji indeksiramo zgolj polji *Avtor* in *Naslov*, filtri v spletni aplikaciji pa imajo indeksirana polja (in njihove vrednosti) *Identifikator*, *Avtor*, *Naslov*, *Tipologija*, *Leto*, *Kraj* in *Št. citatov*. V administrativnem sistemu je bil filter nadgrajen. Prej je omogočal filtriranje po naslednjih parametrih: *Avtor*, *Leto*, *Naslov*, *Vir* in *Kraj*. Ti po mnenju indeksatorjev niso omogočali učinkovitega in natančnega iskanja zapisov znotraj baze. Novi filtri vsebujejo večje število parametrov: *Tip (monografija/serijska publikacija)*, *ID*, *Avtor*, *Naslov*, *Leto* in *Vir*. Iskalnik ElasticSearch podpira tudi funkcijo samodokončanja iskalne poizvedbe, poznano tudi pod imenom *Autocomplete* ali *Completion suggester*. Funkcija je optimizirana za hitrost tipkanja, saj se prilagaja hitrosti tipkanja iskalne poizvedbe, ki jo uporabnik vnese. Podpira izključno *type as you go* funkcijo in ni mišljena za samodejno korekcijo iskalne poizvedbe ali funkcije *Ali ste misili* (Elastic, nd). V našem primeru se na funkcijo samodokončanja, enako kot pri osnovnem iskalniku, vežeta zgolj polji *Avtor* in *Naslov*.

### 3.1.3. Uvoz in izvoz metapodatkov - MODS aplikacijski profil

XML ali eXtensible Markup Format prihaja iz družine označevalnih jezikov, kot sta SGML in HTML. Vendar pa se od omenjenih formatov razlikuje predvsem po svoji

fleksibilnosti v primerjavi s HTML omogoča oblikovanje lastnih označevalcev (angl. tag) in s tem predstavlja enega izmed najpogosteje uporabljenih standardov za izmenjavo podatkov v digitalni humanistiki (W3C, 2008). Že v prejšnjih verzijah baze je izvoz podatkov bil mogoč v formatu XML. Shema je predpostavljala lastne označevalce in ni upoštevala kateregakoli metapodatkovnega standarda. Kot je bilo že omenjeno, to pomeni zmanjšano stopnjo interoperabilnosti podatkov. Iz tega razloga smo shemo in polja prenesli v metapodatkovno shemo MODS. Postopek je vključeval tri faze:

- Pregled elementov stare sheme, ki je za svoje označevalce upoštevala imena, kot so *OpTipBiblEnote* ali *OpSistoryUrnId* del označevalca 'Op' se nanaša na publikacijo, ki jo opisujemo (Op = original publication), 'Pv' pa označuje podatke za vir publikacije, sledi interno poimenovanje polja (ki ustreza imenu polja, iz katerega vzamemo podatke).
- Preslikava internih polj (poimenovanje po meri) v metapodatkovni standard MODS in komentiranje kode (navodila za programerja, iz katerih polj v stari metapodatkovni shemi se vežejo vrednosti v nove elemente). Iz ene sheme sta nastali dve novi, upoštevali smo novo strukturo mask za vnos podatkov, tako kot smo predhodno enotno masko razdelili na masko za monografije in serijske publikacije. V aplikacijskem profilu v skupnem metapodatkovnem zapisu v formatu XML sta ločena zapisa mask definirana z elementom *mods* in identifikatorjem *ID=pub* za oznako zapisa za monografijo ali serijsko publikacijo (na primer *mods ID="pub.224"*) ali elementom *relatedItem* in identifikatorjem za oznako navedenih del, na primer *relatedItem type="references" ID="ref.1"*.
- Prenos vrednosti iz starih internih polj v polja MODS ima svoje prednosti poleg dejstva, da tako povečamo interoperabilnost svojih podatkov z drugimi sistemi, s tem pridobimo večjo strukturiranost in pogosto tudi dodatne podatke, ki jih v stari shemi ne bi mogli implementirati. Element *OpJezik* ima za svojo vrednost na primer le številčno vrednost "21", kar se navezuje na interni nekontroliran seznam jezikovnih vrednosti, novi element pa v svoji strukturi dovoljuje navedbo avtoritete in tipa poimenovanja. Tako poleg jezikovne kode pridobimo tudi podatek o standardu oziroma kontroliranem seznamu, ki je bil uporabljen, s tem pa tudi standardiziramo vrednost zapisa.

Slika 1 prikazuje strukturo in del elementov stare, interne metapodatkovne sheme.

Spodaj sta prikazana star in nov način poimenovanja ter primerjava strukture posameznega zapisa:

Interna shema ZIC (element Avtor):

```
<field name="OpAvtor0">Hadalin Jurij</field>
```

Aplikacijski profil v XML:

```
<name type="personal">
  <namePart>Priimek Ime avtorja</namePart>
```

```
<?xml version="1.0"?>
<root>
  <del>
    <field name="ID" >4916</field>
    <field name="OpTipBiblNote" ></field>
    <field name="OpTipologija" >1</field>
    <field name="OpSist" >1</field>
    <field name="OpJezik" >21</field>
    <field name="OpDrzava" >17</field>
    <field name="OpStatVtorjev" >1</field>
    <field name="OpCoblId" >3955316</field>
    <field name="OpSistoryUrnId" >0</field>
    <field name="OpAvtor1" >Hadalin Jurij</field>
    <field name="OpAvtor2" ></field>
    <field name="OpUrednik" ></field>
    <field name="OpAvtor" >Jugoslovanska molica</field>
    <field name="OpVzNoslov" ></field>
    <field name="OpVzNaslov" ></field>
    <field name="PvCoblId" >392352928</field>
    <field name="PvISSN" >978-961-6386-99-9</field>
    <field name="PvTip" >2</field>
    <field name="PvAvtor" ></field>
    <field name="PvNoslov" >Monografski blagog : materialna kultura potrošniške družbe na Slovenskem</field>
    <field name="PvPodnaslov" ></field>
    <field name="PvPodredni" ></field>
    <field name="PvVzporedniNoslov" ></field>
    <field name="PvZbirka" >Vpogledi ; 22</field>
    <field name="Pvkraj" >Ljubljana</field>
    <field name="PvLetnik" >1998</field>
    <field name="PvLeto" >2019</field>
    <field name="PvSt" ></field>
    <field name="PvStran" >151-165</field>
    <field name="KwDeskriptor1" ></field>
    <field name="KwDeskriptor2" ></field>
```

Slika 1: Metapodatkovna polja maske za vnos podatkov pred nadgradnjo.

```
<role>
  <roleTerm type="code">cre</roleTerm>
  <roleTerm>Avtor</roleTerm>
</role>
<namePart type="family">Priimek</namePart>
<namePart type="given">Ime</namePart>
</name>
```

Interna shema ZIC (element Jezik)

```
<field name="OpJezik" >21</field>
```

Aplikacijski profil v XML:

```
<language>
  <languageTerm type="code">slv</languageTerm>
  <scriptTerm type="code">Latin</scriptTerm>
</language>
```

Interna shema ZIC (element Tipologija):

```
<field name="OpTipologija" >1</field>
```

Aplikacijski profil:

```
<classification authorityURI
  ="https://www.izum.si/">101</classification>
```

Z novim aplikacijskim profilom, ki izhaja iz metapodatkovnega standarda MODS, smo namesto internih metapodatkovnih elementov v shemi uporabili obstoječi in razširjeni metapodatkovni standard MODS. S tem smo naboljšali dve izmed temeljnih načel: poznavanje oziroma uporabo poznanih in razširjenih tehnologij ter načelo interoperabilnosti. Format XML nam namreč zagotavlja lažje izmenjevanje in diseminacijo podatkov z drugimi sistemmi.

### 3.1.4. Migracija vrednosti polj Avtorji

Enega izmed večjih problemov, ki nam ga je delno uspelo rešiti med nadgradnjo, predstavlja migracija vrednosti polja Avtor(ji) iz skupnega polja v dve ločeni. Problem je nastal zaradi neenotnega zapisa oziroma različnih oblik vrednosti Priimek in Ime (oblike: *Priimek, Ime; Ime in Priimek, Ime, Priimek ...*) ter naštevanja več avtorjev v enem polju (*Avtor1; Avtor2 ...*), ki so bili med seboj ločeni z različnimi ločili. Ta problem nam je uspelo rešiti zgolj

delno: migracija, ki je potekala strojno, je bila uspešna na poljih, ki so se med seboj ujemala, pri določenih zapisih pa to ni bilo mogoče (primer *Ime Ime, Priimek*), zato zahteva ročne popravke. Te napake bomo lahko odpravili po začetku procesa prečiščevanja baze, ki pa za zdaj še ni predviden.

### 3.2. Spletna aplikacija in uporabniški vmesnik

#### 3.2.1. Podatkovna baza vseh del in podatkovna baza vseh bibliografskih navedb

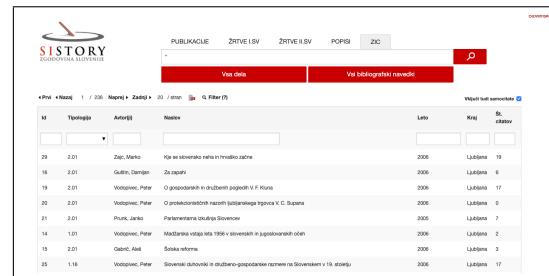
Spletna aplikacija vsebuje dve podatkovni bazi bazo *Vsa dela* in podatkovno bazo *Vse bibliografske navedbe*. Razlog za dve medsebojno ločeni bazi leži v prikazu rezultatov, še natančneje v prikazu števila prejetih citatov pri določenem zapisu. Pri izpisu rezultatov je na voljo število citatov, ki jih je določeno delo prejelo, vendar ti podatki morda niso pravilni, ker se število prejetih citatov določenega dela veže na ujemanje naslova pri glavnem vnosu (maska za vnos glavnega zapisa) in pri citatu (maska za vnos citata). Kot pa smo omenili že zgoraj, nemalokrat pride do napak. Zaradi tega je potrebna druga baza *Vse bibliografske navedbe*, po kateri je omogočeno brskanje z uporabo filtrov. Ta baza dovoljuje uporabniku dodaten in bolj natančen vpogled v citate, saj tu dejansko vidimo vse vnesene citate, indeksatorjem pa predstavlja dodatno orodje za lažje popravke že obstoječih zapisov (preglednejše iskanje zapisov slabše kakovosti).

#### 3.2.2. Prikaz iskalnih rezultatov

Iskalni rezultati so prikazani v obliki tabel, ki uporabnikom ponujajo tudi filtriranje rezultatov oziroma omogočajo oženje iskalne poizvedbe znotraj tabele. Rezultate je mogoče tudi razvrščati. Poleg filtriranja je uporabniku omogočen izvoz zadetkov na seznamu rezultatov in posameznega zadetka v formatu PDF. Za uporabnike sta prav tako pripravljeni tudi dve vrsti pomoči osnovna razlaga uporabe citatnega indeksa na prvi strani ZIC (iskanje/brskanje) in pa manjši namig pri uporabi filtrov s primeri uporabe ločil. Prikaz posameznega zapisa uporabniku dovoljuje vpogled v osnovne podatke (metapodatke dela), osnovne podatke vseh del, v katerih je bil citiran, in avtorjev seznam literature. Podatki so prikazani v dveh ločenih tabelah, *Citiran v in Seznam literature*, zapisi so med seboj povezani.

Med oblikovanjem vmesnika so v vmesnih fazah sodelovali raziskovalci/uporabniki, s katerim smo testirali odzive na novi vmesnik, novo podatkovno strukturo in nove funkcionalnosti. Največ težav je predstavljala terminologija, predvsem na podlagi dejstva, da se zgodovinarsko dajevanje terminov literature in virov precej razlikuje od pojmovanja na področju tehnologije. Nerodna poimenovanja iz prejšnje verzije vmesnika (*Avtor citira, Citiranost Avtorja*) je bilo treba nadomestiti s terminom, ki bo uporabnikom razumljiv. Kot že omenjeno, smo se na podlagi tega odločili za osnovno iskanje in dve ločeni bazi, ki sta po številnih preimenovanjih pridobili ime *Vsa dela* in *Vsi bibliografski navedki*. Čeprav sta imeni daljši, smo prednost namenili razlagi terminov, saj so uporabniki menili, da sta ti poimenovanji najbolj jasni in logični. Poleg terminologije je problem predstavljala tudi postavitev elementov na spletni strani (predvsem gumbi). Tu se je izkazalo, da je

uporabnike precej medla postavitev gumbov za obe bazi, saj so mislili, da s klikom na npr. *Vsa dela* dobijo vsa dela iskanega avtorja. Težavo smo odpravili tako, da smo ustvarili različne statične verzije uporabniškega vmesnika in s pomočjo uporabnikov določili najbolj jasno in intuitivno.



The screenshot shows a table of search results. The columns are labeled: ID, Tvorilja, Autorji, Naslov, Leto, Kraj, and Št. citatov. The data includes:

ID	Tvorilja	Autorji	Naslov	Leto	Kraj	Št. citatov
29	Zelen, Matja		Na slovensko napis in translaci zache	2006	Ljubljana	13
16	Gutten, Damjan		Za zdravje	2006	Ljubljana	6
19	Wodopivec, Peter		O preporočilih in državnih projektih v R. Kralje	2006	Ljubljana	17
20	Wodopivec, Peter		O preteklosti in načinu življenja kneza K. C. Supina	2006	Ljubljana	5
21	Prušek, Janko		Parlementarna ustupka Slovencev	2006	Ljubljana	3
14	Wodopivec, Peter		Mednarodna vstopila leta 1996 v evropske in jugoslovanske učeh	2006	Ljubljana	2
15	Gaberl, Aleš		Sloške reforme	2006	Ljubljana	3
25	Wodopivec, Peter		Slovenski duhovni in državno-gospodarske razmere na Slovenskem v 19. stoletju	2006	Ljubljana	17

Slika 2: Trenutni uporabniški vmesnik ZIC-a.

#### 3.2.3. Uporaba indeksa citiranosti

Primarni uporabniki citatnega indeksa so raziskovalci, ki lahko v sistemu enostavno preverijo št. prejetih citatov za posamezno avtorsko delo; če je to indeksirano v sistem. Poleg izpisa iz sistema SICRIS (Slovenian Current Research Information System), ki je osnova za vrednotenje znanstvene uspešnosti na posameznem raziskovalnem področju, lahko izpis iz ZIC predstavlja dodano vrednost pri prijavljanju projektov ali programov na področju humanistike. Poleg raziskovalcev si z ZIC lahko pomagajo tudi uredniki revij, ki želijo preveriti, kolikokrat so bili posamezni članki citirani, in s tem upravičijo obstoj revije. Poleg primarne naloge, ki je nudjenje vpogleda v št. prejetih citatov, pa indeks ponuja tudi druge možnosti, ki jih stari ZIC ni ponujal. Te naj bi uporabniku omogočile prijetnejšo interakcijo s sistemom. Ena izmed takšnih funkcionalnosti je npr. možnost *prijaznega kopiranja*, ki uporabniku omogoča lažje navajanje virov v svojih delih, saj ZIC ponuja skoraj popolne bibliografske podatke, ali npr. izpis št. citatov v formatu PDF ipd.

## 4. Zaključek

Sistem je bil že v svoji začetni zasnovi izjemno ambiciozen in zaradi načina objavljanja na področju zgodovinopisja izjemno potreben. Vendar je Zgodovinarski indeks citiranja zadnja leta nekoliko stagniral. Po pregledu in analizi podatkov smo ugotovili, da je nadgradnja potrebna, saj sistem ne zadostuje potrebam indeksatorjev in uporabnikov. Začeli smo nadgradnjo administrativnega dela, kjer smo preoblikovali oz. nadgradili nove maske, nadgradili metapodatkovno shemo oziroma ustvarili nov aplikacijski profil na podlagi metapodatkovnega standarda MODS, filtre in dodali pomoč indeksatorjem, ki naj bi pripomogla k poenotenim zapisom. Poleg administrativnega dela smo nadgradili tudi uporabniški vmesnik z občasnim testiranjem baze in njenih komponent z raziskovalci. Z omenjeno nadgradnjo smo rešili večino zaznanih problemov, od nejasnih in nepotrebnih polj vnosa podatkov in razčlenitve mask, ki indeksatorju omogočajo lažje in natančnejše oblikovanje zapisov, oblikovanja aplikacijskega profila MODS, ki

omogoča lažji uvoz in izvoz podatkov, do uporabniku prijaznejšega vmesnika itd. Vseh težav pa zaradi omejitev, povezanih z ročnim vnosom podatkov, ni bilo mogoče v celoti rešiti. To velja predvsem za postopek migracije polja Avtorji, kjer bo problem v celoti rešen šele po prečiščenju cele baze podatkov. Postopek prečiščenja bo pripomogel tudi k poenotenu zapisov, kar bo omogočalo, da uporabniki v sistemu pridobijo zanesljive in kakovostne informacije. Pri nadgradnji Zgodovinarskega citatnega indeksa smo dosegli zastavljeni cilje. Sistem smo tehnično posodobili in ZIC postavili kot ločeno spletno aplikacijo na poddomeni portala SIStory. Spletna aplikacija je narejena modularno, zato je mogoče dodajati nove funkcionalne rešitve, iskalnik s tehnologijo ElasticSearch pa omogoča natančnejše in preglednejše iskanje po podatkih. Z nadgradnjo smo oblikovali sistem, ki je intuitiven za indeksatorje in uporabnike, s tem pa zagotovili, da ZIC služi svojemu namenu.

## 5. Zahvala

Raziskavo je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije v okviru programa Raziskovalne infrastrukture slovenskega zgodovinopisa (I0-0013) in slovenske raziskovalne infrastrukture DARIAH SI.

## 6. Literatura

- Elastic. n.d. What is elasticsearch. <https://www.elastic.co/what-is/elasticsearch>.
- Wolfgang Glänzel in Urs Schoepflin. 1999. A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, 35(1):31–44.
- Diana Hicks, 2004. *The four literatures of social science*, pogl. Handbook of quantitative science and technology research, str. 476–496. Kluwer.
- Mu-hsuan Huang in Yu-wei Chang. 2008. Characteristics of research output in social sciences and humanities: from a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, 59(11):1819–1828.
- Primož Južnič. 2017. Bibliometrijski indikatorji. Dostopno na: <https://www.youtube.com/watch?v=19W5g1Z197I&feature=youtu.be>.
- Kayvan Kousha, Mike Thelwall in Somayeh Rezaie. 2011. Assessing the citation impact of books: The role of google books, google scholar, and scopus. *Journal of the American Society for Information Science and Technology*, 62(11):2147–2164.
- Žarko Lazarević in Igor Zemljič. 2003. Slovenski zgodovinarski indeks citiranosti - izhodišča in pomisleki. Neobjavljena dokumentacija.
- Anton Nederhof. 2006. Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1):81–100.
- Andrej Pančur, Mojca Šorn in Jurij Hadalin. 2014. Slovenski indeks citiranosti (sici): Načrt izgradnje in delovanja. Tehnično poročilo, Inštitut za novejšo zgodovino. Dostopno na: <https://www.sistory.si/11686/36153>.
- Andrej Pančur in Mojca Šorn, 2019. *Zbornik ob 60-letnici Inštituta za novejšo zgodovino*, pogl. Raziskovalna infrastruktura Inštituta za novejšo zgodovino. Inštitut za novejšo zgodovino.
- Andrej Pančur. 2019a. Preprosta raziskovalna infrastruktura za kompleksne raziskovalne podatke v humanistiksi - si4 (Simple research Infrastructure FOR complex research data in digital humanities). Neobjavljena dokumentacija.
- Andrej Pančur. 2019b. Specifikacije za izvedbo naročila izdelave Zgodovinarskega indeksa citiranosti (ZIC). Neobjavljena dokumentacija.
- W3C. 2008. Extensible markup language (xml) 1.0 (fifth edition). Tehnično poročilo, World Wide Web Consortium. Dostopno na: <https://www.w3.org/TR/xml/>.