



# **SPECIFIČNOSTI DIGITALIZACIJE IZDANJA LEKSIKOGRAFSKOGA ZAVODA MIROSLAV KRLEŽA**

Irina Starčević Stančić i Cvijeta Kraus

Lexikografski zavod Miroslav Krleža  
15. i 16. listopada 2020.



... VIŠE OD 400 SVEZAKA ENCIKLOPEDIJA,  
LEKSIKONA, RJEČNIKA I DRUGIH IZDANJA ...





# PORTAL DIGITALIZIRANIH IZDANJA

Rad na digitalizaciji izdanja Zavoda započeo je 2009. projektom izradbe portala znanja (na adresi [enciklopedija.lzmk.hr](http://enciklopedija.lzmk.hr)); obuhvaća sadržaj 10 digitaliziranih izdanja s više od 110 000 natuknica.

Enciklopedija Miroslava Krleže, Filmska enciklopedija, Filmski leksikon, Hrvatski biografski leksikon, Hrvatski obiteljski leksikon, Istarska enciklopedija, Medicinski leksikon, Nogometni leksikon, Pomorski leksikon, Tehnički leksikon

LEKSIKOGRAFSKI ZAVOD MIROSLAV KRLEŽA  
[enciklopedija.lzmk.hr](http://enciklopedija.lzmk.hr)



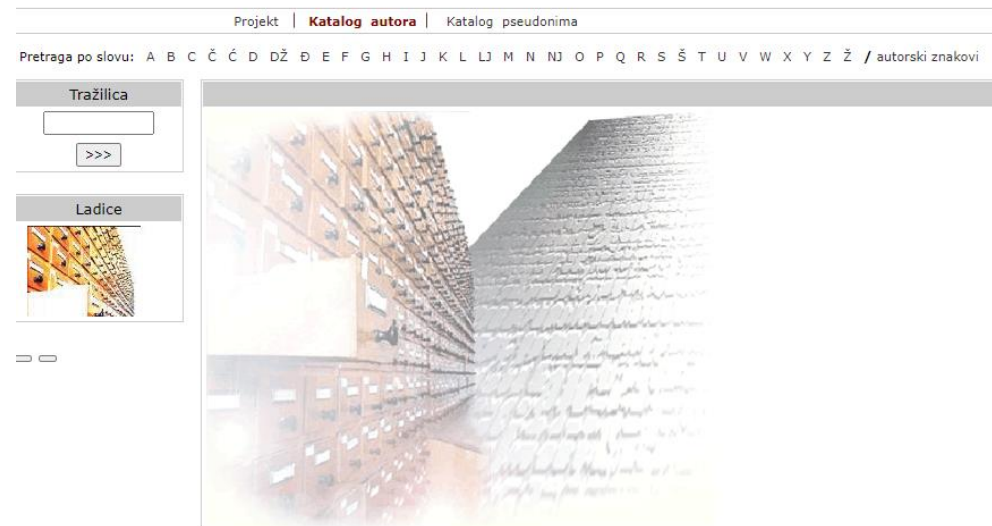


# DIGITALIZIRANI KATALOG

Digitalizirani katalog projekt je digitalizacije 10 milijuna listića Kataloga retrospektivne bibliografije članaka, a dostupan je javnosti preko mrežne stranice (na adresi katalog.lzmk.hr).

Obuhvaća skenirane listiće autorskoga kataloga te kataloga šifri i pseudonima i pretraživ je kao u ladicama izvornoga kataloga. Sadržaj skeniranih kartica unesen je u mrežnu aplikaciju. Cilj je da katalog bude pretraživ po autoru, predmetima i strukama.

## Digitalizirani Katalog



# PROJEKTI DIGITALIZACIJE

Digitalizacija enciklopedijskog procesa (strategija digitalnog razvoja Zavoda) kao temelj za promjene u izdavanju digitalnih izdanja vs. digitalizacija sadržaja tj. izdanja kao zaštita građe i podloga za bazu leksikografskih sadržaja

Cilj je sva izdanja prebaciti u digitalni oblik – dostupna i pretraživa u sustavu portala (repozitorij leksikografskog sadržaja) uz ostale digitalne sadržaje

Digitalizacija je sastavni dio strategije LZMK 2015-2020., ali ne postoji definiran plan i program digitalizacije koji je neophodan za cjelokupan digitalni razvoj – jasna strategija i definirani ciljevi, detaljan i razrađen plan, definirana metodologija rada koja uvažava korake u procesu digitalizacije

# PROJEKTI DIGITALIZACIJE ...

Uspješnost provedenih programa i projekata digitalizacije ovisi o sposobnosti organiziranja, provođenja i kontrole svih procesa digitalizacije, a upravljanje procesom digitalizacije zahtijeva određena znanja i vještine koje je potrebno stalno usavršavati

U projektima digitalizacije voditi se međunarodnim smjernicama, dobroj praksi i načelima te nacionalnom strategijom za digitalizaciju kulturne baštine

- Upute za planiranje i vođenje projekta digitalizacije u okviru Nacionalnog programa digitalizacije (2006): “Digitalizacija se odvija u okviru primjereno planiranih i vođenih projekata, sukladno poznatim načelima i prioritetima i prema utvrđenim normama.”

# PROJEKTI DIGITALIZACIJE ...

Ovise o specifičnosti enciklopedijske i leksikografske građe (atlas, enciklopedije, leksikoni, rječnici) ali i o količini/obimu i vrsti sadržaja, tisku, kulturnoj vrijednosti, dostupnosti, tehnologiji, financijskim mogućnostima

Izvode se u okviru vlastitih financijskih i kadrovskih mogućnosti, financiraju se kroz natječaje Ministarstva kulture RH (Mate Ujević, Pomorska enciklopedija), istražuju se mogućnosti financiranja kroz fondove ili programe EU

U sklopu programa digitalizacije građe LZMK od 2009. godine napravljeno je cca 70 000 digitalnih preslika

Planirana je digitalizacija 20 izdanja tiskanih do 2000. godine koja nisu izvorno nastala u digitalnom obliku, ukupno 92 556 stranica

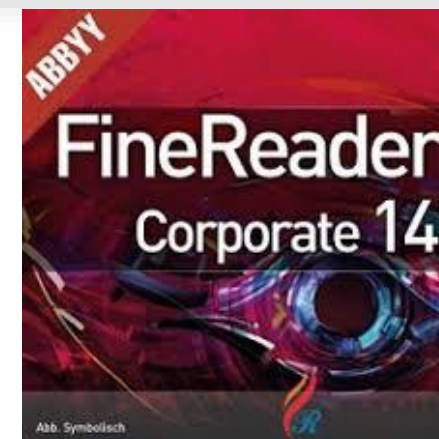


# PROCES DIGITALIZACIJE

Digitalizacija knjižnih izdanja provodi se skeniranjem građe, njezinom obradom s pomoću programa za optičko prepoznavanje znakova (OCR), korekturom teksta i njegovim unosom u aplikaciju za mrežno objavljivanje izdanja

1 skener (Microtek Artix DI 2020) s uvlakačem  
4 licence za Abbyy FineReader

Izdanja izrezana u listove skeniraju se, pohranjuju na server u slikovnom formatu (arhiviraju), a onda se dodatno obrađuju programom za optičko prepoznavanje znakova (OCR – Optical Character Recognition) da bi se dobio tekst



# PROCES DIGITALIZACIJE ...

Program automatski pretvara sliku u tekst, a po potrebi se elementi na stranici mogu označiti ručno (tekst, slika, tablica ...)

Nakon što program pretvori sliku u tekst potrebno je tekst provjeriti i ispraviti eventualne greške koje je program napravio i ujednačiti s originalnim tekstom izdanja = korektura teksta

Prilikom obrade teksta OCR programom potrebno je obratiti pažnju na određene dijelove teksta zbog mogućnosti pogrešnog prepoznavanja znakova

Obrada dobivenog teksta za unos u bazu izdanja (označavanje slika, formula, uputnica, tablica i posebnih znakova), njihovo prilagođavanje za prikaz na internetu



# PROCES DIGITALIZACIJE ...

## Izazovi/specifičnosti

1. ZNAKOVI UNUTAR TEKSTA (slova sa znakovima svojstvena drugim jezicima (ö; á; ...), nazivi natuknica, imena autora, literature, datumi unutar teksta, veličina i stilovi slova: podebljano, kurziv, razmaci)
2. SLIKE UNUTAR TEKSTA / LEGENDE SLIKA
  - slike se označavaju izdvojeno od teksta, a potpisi ispod slika se označavaju kao tekstualni dio
3. TABLICE
  - OCR program daje mogućnost oblikovanja tablica kao u originalnom tekstu
  - program ih automatski pročita i analizira, a naknadno ih je moguće dodatno urediti
4. FORMULE
  - dodatno se uređuju kao slike ili programima za pisanje matematičkih formula
  - primjer Tehničke enciklopedije gdje su označene kao slike, a tekst TE je u pretraživom pdfu



# ZAKLJUČAK

Digitalizacija u svrhu prezentacije, dostupnosti i povezivanja digitalnih i digitaliziranih sadržaja izdanja Leksikografskoga zavoda = podloga za leksikografski rad i izgradnju digitalnog repozitorija

Kreiranje digitalnog pretraživog i povezanog enciklopedijskog i leksikografskog sadržaja osnova je za jačanje uloge Zavoda u mrežnome okruženju

Preduvjet je definiran plan s razrađenim postupcima digitalizacije uklopljen u cjelokupnu strategiju digitalnog razvoja Zavoda

Digitalizacija jača mrežnu vidljivost izdanja Zavoda, omogućuje ažurniju i bržu izradu novih izdanja i povezivanje s drugim kulturnim, umjetničkim i znanstvenim institucijama