



**19. konferenca  
Dnevi slovenske informatike**

**nove spletne tehnologije:  
MOŽNOST AVTOMATSKEGA USTVARJANJA  
METAPODATKOV V KNJIŽNICAH**

*Zoran Krstulović, Matjaž Kragelj*

*NARODNA IN UNIVERZITETNA KNJIŽNICA*

17. 04. 2012



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

### USTVARJATI, TO POMENI DVAKRAT ŽIVETI (*Albert Camus*)

- Knjižno gradivo predstavlja enega od temeljev kulturne dediščine
- **95%** vseh elektronskih publikacij dostopnih na portalu Digitalne knjižnice Slovenije ([www.dlib.si](http://www.dlib.si)) **ni** bibliografsko obdelanih
- Digitalna knjižnica Slovenije vsebuje več kot **500.000** enot publikacij.



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

### PRIMERI gradiva:

- Laibacher Zeitung (časopis) – 58.500+ števil
- Ljubljanski Zvon (literarni in znanstveni list) – 11.000+ člankov
- Dom in svet (literarni mesečnik) – 16.000+ člankov

ZA VSE TRI „časopise“ velja:

zgolj **en zapis** v knjižničnem katalogu COBISS za vse pripadajoče publikacije časopisa



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

---

### CILJI

- Opremiti vse publikacije digitalne knjižnice s **pripadajočim** UDK vrstilcem
- Opremiti **poljubno** digitalno publikacijo v **poljubnem jeziku** z enim ali več UDK vrstilcem



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

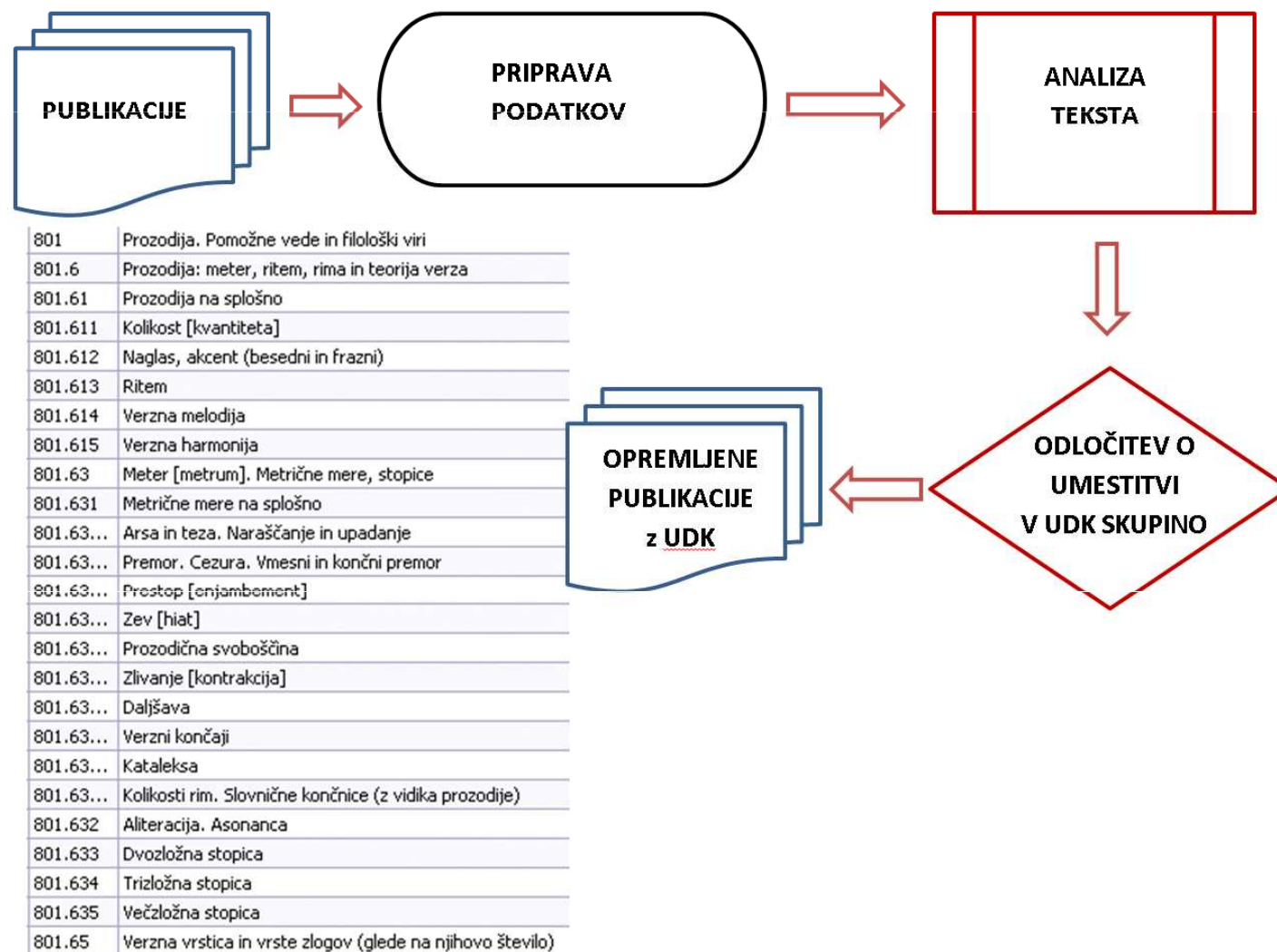
---

### NAČRT

- Določiti učno in testno množico publikacij
- Izgradnja odločitvenega drevesa UDK vrstilcev  
(*enakomerna porazdelitev, imenska lestvica*)
- Lematizacija in prevod besedila
- Izločitev nepomembnih („stop words“) besed
- Umeščanje v pravo UDK skupino/skupine  
(*odločitveno drevo, Naivni Bayes, k-najbližjih sosedov*)



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah





## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

---

### REZULTAT

Delujoč servis za AVTOMATSKO umeščanje digitaliziranih in digitalno rojenih publikacij v UDK skupine



## Možnost avtomatskega ustvarjanja metapodatkov v knjižnicah

---

### ZAKAJ?

- Vsebina starega gradiva iskalna zgolj po polnem besedilu
- Pomoč uporabnikom pri preiskovanju „starejšega“ in „neobdelanega“ gradiva
- UDK je najbolj razširjen in uporabljan klasifikacijski sistem na svetu
- Pomoč bibliotekarjem pri označevanju *novih* publikacij v sistem Univerzalne decimalne klasifikacije.





***Hvala za vašo pozornost !***

*Vprašanja?*

*Pripombe?*

*Predlogi?*