

Stručni skup „Knjižnični podaci“: interoperabilnost, povezivanje i razmjena

Nestandardizirani podaci u
standardiziranom kontekstu:
primjena programskoga jezika Awk u cilju
sintaktičke interoperabilnosti podataka

Ivan Kapec

Nacionalna i sveučilišna knjižnica

Kruno Peter

Nastavni zavod za javno zdravstvo „Dr. Andrija Štampar“

Ideja

- rad s građom posebne vrste
- neoperabilnost podataka
- ograničenost resursa

Tekstualna građa – nestandardizirana

- pisma, članci, leci, skice, književna djela, koncepti, dnevnici, novinski izvještaji, itd.
- rukopisi, strojopisi, kopije, faksimilne kopije, strojopisi s rukopisnim bilješkama, tiskopisi, računalni ispisi

Organizacija podataka

- proračunske tablice (Excel, Calc)
- tiskani katalogi
- kataložni listići

Problem

- tablice su dostupne su samo lokalno
- ograničen oblik izrade izvještaja
- nefleksibilnost
- neoperabilnost
- nemogućnost integracije u veće sustave

Proces obrade je i proces organizacije

- tijekom pregleda građe stvaraju se okvir njene organizacije
- potreban je fleksibilni sistem unosa podataka koji se može lako preoblikovati u tijeku procesa obrade

Manipulacija podataka

1. izrada tiskanog kataloga
2. izrada bibliografije, brošura
3. izrada kataloga izložbe ili legendi za izložbe
4. fleksibilna izrada raznih vrsta popisa
5. integracija u veće sustave (online katalog, digitalna knjižnica, skupni katalog, itd.)

Primjena Awka u cilju sintaktičke interoperabilnosti podataka

- pohrana podataka u CSV format – tekstualni format
- generiranje tekstualnog izvještaja
- generiranje izvještaja u formi Web-stranice (HTML format)
- pohrana podataka iz CSV formata u XML format (MARCXML)

Tekstualni format podataka

Proračunska
tablica

	A	B	C	D	E	F	G	H	I
1	Broj	Autor	Naslov	Mjesto	Vrijeme	Opseg	Dimenzije	Napomena o načinu izrade	Napomena o sadržaju
2	1	Maren Köster-H	Nejednaka trojka	[s. l.]	27. 4. 1990.	2 l.	28,5 x 21 cm	strojopis s rukopisnim bilješkama	strani tisak
3	2	Tihomil Radja	Kakav priključak na Europu	[s. l.]	[1990.]	8 l.	razl. vel.	faksimil rukopisa	članak
4	3	Jakša Kušan	Predsjednik buduće hrvatske vlade	[s. l.]	[1990.]	2 l.	28,5 x 21 cm	rukopis - autograf	
5	4	Vain Packe	Sonata za solo flautu	Santa Cruz	2011.	5 l.	28 x 23 cm	rukopis - autograf	note

CSV format
(engl. Comma
Separated
Values)

Broj;Autor;Naslov;Mjesto;Vrijeme;Opseg;Dimenzije;Napomena o načinu izrade;Napomena o sadržaju
1;Maren Köster-Hetendorf;Nejednaka trojka;[s. l.];27. 4. 1990.;2 l.;28,5 x 21 cm;strojopis s rukopisnim bilješkama;strani tisak
2;Tihomil Radja;Kakav priključak na Europu;[s. l.];[1990.];8 l.;razl. vel.;faksimil rukopisa;članak
3;Jakša Kušan;Predsjednik buduće hrvatske vlade;[s. l.];[1990.];2 l.;28,5 x 21 cm;rukopis - autograf;
4; Vain Packe ;Sonata za solo flautu;Santa Cruz; 2011.;5 l.;28 x 23 cm;rukopis - autograf;note

Generiranje tekstualnog izvještaja Awkom

<p>Awkova skripta (naredba: <code>awk -f skripta.awk podaci.csv > izvjestaj.txt</code>)</p>	<pre>BEGIN { FS = ";" } NR > 1 { print "" print "Br. " \$1 "." print " " \$3 " / " \$2 " , " \$4 " , " \$5 " ; " \$6 " , " \$7 napomena = " , " \$9 if (\$9 == "") napomena = "" print " " "(" \$8 napomena ")" }</pre>
<p>Tekstualni izvještaj (izvjestaj.txt)</p>	<p>Br. 1. Nejednaka trojka / Maren Köster-Hetzendorf, [s. l.], 27. 4. 1990.; 2 l., 28,5 x 21 cm (strojopis s rukopisnim bilješkama, strani tisak)</p> <p>Br. 2. Kakav priključak na Europu / Tihomil Radja, [s. l.], [1990.]; 8 l., razl. vel. (faksimil rukopisa, članak)</p> <p>Br. 3. Predsjednik buduće hrvatske vlade / Jakša Kušan, [s. l.], [1990.]; 2 l., 28,5 x 21 cm (rukopis - autograf)</p> <p>Br. 4. Sonata za solo flautu / Vain Packe , Santa Cruz, 2011.; 5 l., 28 x 23 cm (rukopis - autograf, note)</p>

Generiranje izvještaja u formi Web-stranice

Awkova skripta	<pre>BEGIN { FS = ","; print "<!doctype html>" print "<html>" print "<head>" print " <title>Knjižnična građa</title>" print "<style>html { color: DarkBlue; font-family: sans-serif }</style>" print "</head>" print "<body>" }</pre>	<pre>NR > 1 { print "<p>" "" "Br. " \$1 "" "</p>" print "<p>" "&nbsp;&nbsp;&nbsp;" \$3 "/" \$2 ", " \$4 ", " \$5 "; " \$6 ", " \$7 "</p>" napomena = ", " \$9 if (\$9 == "") napomena = "" print "<p>" "&nbsp;&nbsp;&nbsp;" "(" \$8 napomena ")" "</p>" }</pre>	<pre>END { print "</body>" print "</html>" }</pre>
Web-stranica (HTML format)	 <p>The screenshot shows a web browser window with the title "Knjižnična građa". The address bar displays "file:///C:/src/Leci/gradja.html". The page content is as follows:</p> <ul style="list-style-type: none">Br. 1 Nejednaka trojka / Maren Köster-Hetzendorf, [s. 1.], 27. 4. 1990.; 2 l., 28,5 x 21 cm (strojopis s rukopisnim bilješkama, strani tisak)Br. 2 Kakav priključak na Europu / Tihomil Radja, [s. 1.], [1990.]; 8 l., razl. vel. (faksimil rukopisa, članak)Br. 3 Predsjednik buduće hrvatske vlade / Jakša Kušan, [s. 1.], [1990.]; 2 l., 28,5 x 21 cm (rukopis - autograf)Br. 4 Sonata za solo flautu / Vain Packe , Santa Cruz, 2011.; 5 l., 28 x 23 cm (rukopis - autograf, note)		

Pohrana podataka iz CSV formata u XML format

Awkova skripta

```
BEGIN {  
  FS = ";;";  
  
  print "<?xml version=\"1.0\" encoding=\"UTF-8\"?>"  
  print "<collection xmlns=\"http://www.loc.gov/MARC21/slim\">"  
}
```

```
NR > 1 {  
  print " "<record>"  
  print " "<datafield tag=\"245\" ind1=\"1\" ind2=\"0\">"  
  print " "<subfield code=\"a\">"  
  print " "$3  
  print " "</subfield>"  
  print " "</datafield>"  
  print " "<datafield tag=\"300\" ind1=\" \" ind2=\" \">"  
  print " "<subfield code=\"a\">"  
  print " "$8  
  print " "</subfield>"  
  print " "<subfield code=\"b\">"  
  print " "$9  
  print " "</subfield>"  
  print " "</datafield>"  
  print " "</record>"  
}
```

```
END {  
  print "</collection>"  
}
```

XML format
(MARCXML) (prikaz
grananja s jednim
razvijenim zapisom)

```
<?xml version="1.0" encoding="UTF-8"?>  
- <collection xmlns="http://www.loc.gov/MARC21/slim">  
  - <record>  
    - <datafield ind2="0" ind1="1" tag="245">  
      <subfield code="a"> Nejednaka trojka </subfield>  
    </datafield>  
    - <datafield ind2=" " ind1=" " tag="300">  
      <subfield code="a"> strojopis s rukopisnim bilješkama  
      </subfield>  
      <subfield code="b"> strani tisak </subfield>  
    </datafield>  
  </record>  
  + <record>  
  + <record>  
  + <record>  
</collection>
```

Prednosti primjene Awka u obradi tekstualnih podataka

- besplatan i dostupan za razne operacijske sustave
- lagana i brza instalacija – jedna izvršna datoteka
- jednostavnost sintakse; minimum koda
- rapidni razvoj; izrada prototipova

Kapec, I. Peter, K. Nestandardizirani podaci u standardiziranom kontekstu: primjena programskoga jezika Awk u cilju sintaktičke interoperabilnosti podataka

© CC BY 3.0 HR

STRUČNI SKUP KNJIŽNIČNI PODACI: INTEROPERABILNOST, POVEZIVANJE I RAZMJENA, 28.-29.11.2017.,
NACIONALNA I SVEUČILIŠNA KNJIŽNICA U ZAGREBU