



Nacionalna konferencija sa međunarodnim učešćem  
Fakultet tehničkih nauka u Čačku, 20-22. Septembar 2013.

National Conference with international participation  
Faculty of technical sciences in Cacak, 20-22. September 2013.

UDK: 004.4:37

Stručni rad

## OBRADA REZULTATA TESTOVA ZNANJA XML INTEGRACIJOM PODATAKA

### PROCESSING KNOWLEDGE TEST DATA BY XML DATA INTEGRATION

Veljko Aleksić<sup>1</sup>, Đorđe Damnjanović<sup>1</sup>, Vesna Marković<sup>2</sup>, Milica Vučetić<sup>1</sup>

<sup>1</sup>Fakultet tehničkih nauka u Čačku

<sup>2</sup>OŠ „Kralj Aleksandar I“, Gornji Milanovac

<sup>1</sup>veljko.aleksic@ftn.kg.ac.rs

**Apstrakt:** Integracija podataka iz više izvora zahteva poznavanje rada sa različitim modelima podataka, šemama, upitima i interfejsima. Današnji standard za upravljanje strukturiranim i polustrukturiranim podacima je XML. U radu je prikazana osnova obrade rezultata testova znanja XML integracijom podataka.

**Ključne reči:** Integracija podataka, XML, testovi znanja.

**Abstract:** Multiple source data integration demands working with different data models, schemas, queries and interfaces. Today's standard for managing structured and semi-structured data is XML. This paper presents basic processing of the knowledge test results by XML data integration.

**Key words:** Data integration, XML, Knowledge Test.

#### 1. TESTOVI ZNANJA

Ocenjivanje je kvalitativna analiza čiji je cilj da odredi stepen ostvarenosti određenih ciljeva [1]. Predstavlja razvrstavanje odgovora u kvalitativne kategorije, a zatim određivanje njihove kvantitativne vrednosti (ocene). To je proces proveravanja i merenja postignuća učenika i donošenja pedagoških odluka. Predstavlja sastavni deo i nastave i učenja, a ne samo kulminaciju nastavnog procesa. Prvi korak u funkcionalnom školskom ocenjivanju jeste definisanje kriterijskog obrasca (zahteva situacije) i očekivanog reperotara ponašanja, koje učenik treba da usvoji da bi ispunio zahtev.

Bjekić [1] navodi da je test merni instrument sastavljen od niza zadataka ili problema, sistematski odabranih, pomoću kojih se na objektivni način, na izazvanom uzorku ponašanja, ispituju/mere sposobnosti, osobine ličnosti i znanje pojedinca. Osnovni kriterijum diferenciranja testova je stepen objektivnosti. U nastavi mogu da se koriste standardizovani testovi znanja (tzv. pravi testovi) i nestandardizovani testovi znanja, koji

se nazivaju i nizovima zadataka objektivnog tipa. Nastavnici najčešće koriste ove nizove zadataka objektivnog tipa koje samostalno kreiraju.

Prema sadržaju testovi su klasifikovani u tri osnovne grupe: testovi sposobnosti, testovi ličnosti, testovi znanja. Posebnu grupu čine testovi gotovosti ili spremnosti (test gotovosti za polazak u školu).

**Testovi znanja**, ponekad nazivani i nastavni testovi, jesu testovi pomoću kojih se određuje koliko je znanja pojedinac stekao kroz određenu aktivnost ili određeni period učenja. Sastoje se od niza zadataka, datih u posebnim oblicima, kojima se meri znanje [1]. Testovi znanja su objektivni instrumenti ocenjivanja postignuća. Omogućavaju ispitivanje većeg broja učenika. Zahtevaju vešto sastavljanje jer je njihovo korišćenje opravdano samo ako zadovoljavaju sve metrijske karakteristike testova. Međutim, u praksi se često primenjuju nevešto sastavljeni i nediskriminativni testovi.

Strukturirana priroda podataka prikupljenih testovima znanja predstavlja dobru kvalitativnu osnovu potencijalne upotrebe XML-a za njihovo smeštanje i obradu u elektronskom obliku.

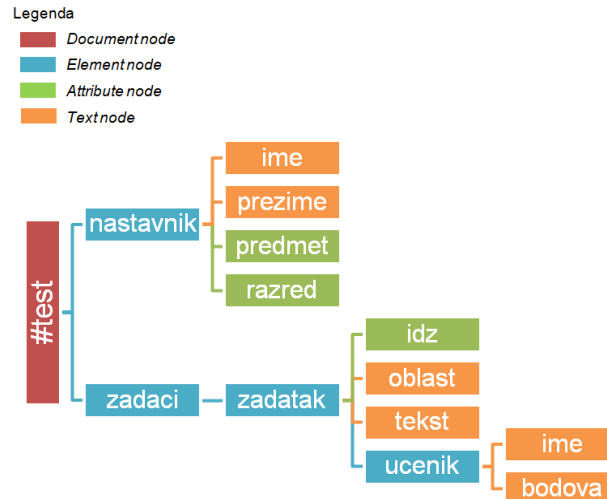
## 2. XML MODEL PODATAKA

XML predstavlja anagram izraza Extensible Markup Language. Markerski jezik označava vrstu jezika koji opisuje strukturu i sadržaj dokumenta, specifično onog koji sadrži podatke. Prema Bezi [2] markerski označava dodatnu tekstualnu sintaksu koja se može koristiti za opis formatiranja, akcija, strukturnih informacija, tekstualne semantike, atributa i sl. Izraz proširiv (eng. extensible) odnosi se na mogućnost proširenja i izmene strukture. Stoga XML predstavlja markerski jezik koji može biti proširen i prilagođen potrebama kreatora dokumenta i podacima nad kojim se vrši akvizicija. XML model podataka ne treba posmatrati kao tradicionalnu reprezentaciju XML podataka u ASCII fajlovima (sintaksno veoma sličnih HTML-u). Čak i XML dokumenti nisu obavezno kreirani u ASCII formatu.

Apstraktni model podataka XML instance je stablo (eng. *tree*) koje se sačinjeno od čvorišta (eng. *nodes*). Prema Meju [3] ovaj model razlikuje vrste čvorova, kao što su *document nodes* (ulazne tačke stabla), *element nodes* (čvorovi unutar stabla), *text nodes* (lišće) i *attribute nodes* (druga vrsta lišća). Ova struktura može se "preslikati" na strukturu podataka prikupljenih testovima znanja, u kojoj odgovarajući *element node* predstavlja zadatak, *text node* njegov sadržaj, a *attribute node* čini dati odgovor.

DOM (eng. *Document Object Model*) API predstavlja specifikaciju apstraktnog tipa podataka koji ovaj model podataka implementira. Mej [3] navodi da je svaka XML instanca povezana jedinstvenim *Document node*-om koji služi za „pristup“ i sadrži neku metainformaciju o dokumentu i *element node* (*unique child*) koji je *root node* dokumenta. *Element node* ima *tag* (ime) i sadrži uređenu listu podređenih *element node*-ova i *text node*-ova. Dodatno, ova čvorišta mogu imati i neuređen set *Attribute node*-ova povezanih sa njima (sadrže ime, tip i vrednost). XML stablo je rekurzivno pa je svaki

element node sa svojim granama podstablo. *Text node*-ovi sadrže samo tekst i nemaju dalje grananje, niti atribute.



Slika 1. Prikaz primera stabla testa znanja

XML procesor ima ulogu posrednika prilikom očitavanja i izmene podataka smeštenih u XML dokumentu. Procesor je odgovoran za razdvajanje markera od sadržaja, dobro formiranje dokumenta uz određenu validaciju, obezbeđenje identifikacije relacija u okviru sadržaja i odgovarajućih aplikacija za rad sa podacima. Ovo je obezbeđeno putem API-a koji predstavlja set funkcija koje podržavaju komunikaciju odgovarajućih delova dokumenta i softvera.

Salminen [4] navodi da postoje dva glavna protokola koja koriste XML procesori. Prvi posmatra XML dokument kao linearnu strukturu formiranu od markera i sadržaja. Alternativni dokument posmatra kao hijerarhijsku strukturu koja se poklapa sa ugnježdenom prirodom markera, sa sadržajem koji se nalazi na različitim mestima najnižih nivoa hijerarhije. Hijerarhijska struktura odgovara strukturi testova znanja.

### 3. INTEGRACIJA PODATAKA

Iako je u praksi razlikovanje informacija i podataka često nedovoljno, jasno je da je definisanje podatka jednostavnije od definisanja informacije koja zahteva viši nivo apstrakcije. Blu [5] navodi da smo imali dve decenije u kojima smo bili fokusirani isključivo na obradu podataka, nakon kojih smo dve decenije fokusirani na informacione tehnologije, a sada smo usmereni na znanje. Postoji jasna razlika između pojmova podatak, informacija i znanje. Informacija predstavlja postavljanje podatka u smisaon okvir. Znanje je mogućnost korišćenja te informacije.

Metapodaci su „podaci o podacima“. Metapodaci su informacije o organizaciji podataka, različitim domenima podataka i relacijama među njima [2]. Razlikujemo dve vrste metapodataka: opisne i semantičke. Prema Bezi, povraćaj informacija se bavi reprezentacijom, smeštanjem, organizacijom i pristupom informacionim elementima. Cilj sistema za povraćaj informacija je obezbeđenje pristupanja korisniku interesantnim informacijama. Sa druge strane sistemi za povraćaj podataka određuju koji objekti iz skupa zadovoljavaju jasno definisane uslove.

Integracija podataka predstavlja problematiku kombinovanja podataka smeštenih na različitim izvorima i njihovo jedinstveno predstavljanje korisniku [6]. Njena primena relevantna je za skladišta podataka, geografske informacione sisteme, elektronsku trgovinu i slične aplikacije. Sistemi za integraciju podataka najčešće su određeni arhitekturom baziranoj na globalnoj šemi koja obezbeđuje integrisan pregled sa više izvora. Skraćenica za ove sisteme je CDIS (eng. *Central Data Integration Systems*).

Centralna integracija podataka izdvaja dva ključna faktora: modeliranje sistema i obradu upita. Prema Ulmanu [7] modeliranje relacija između izvora i globalne šeme može se vršiti kroz dva pristupa:

- GaV (eng. *Global-as-View*) oslikavanjem globalne šeme u smislu izvora podataka;
- LaV (eng. *Local-as-View*) koji zahteva definisanje globalne šeme nezavisno od izvora, a relacije između globalne šeme i izvora se utvrđuju definisanjem svakog izvora kao pogleda na globalnu šemu.

Pretraga na web-u i sličnim sistemima za povraćaj informacija prema Kalanu [8] uobičajeno je bazirana na modelu jedne tekstualne baze podataka, u kojoj su dokumenti kopirani na centralnu bazu indeksirani sa mogućnosti pretrage. Ipak, neke informacije nisu dostupne ovim modelom iz više razloga (veličina, ograničenja i sl.). Alternativu predstavlja model sa više baza podataka u kome centralni sajt umesto smeštanja kopija dokumenata prevodi informacije korisnika u upite nad više izvora. Ovaj model proučava metapretraga (eng. *metasearch*) ili DIR (eng. *Distributed Information Retrieval*). U nekim su aspektima integracija podataka i metapretraga ekvivalenti, ali dok metapretraga cilja zadovoljenju potrebe korisnika za informacijom iz nestruktuiranih ili polustruktuiranih izvora, integracija podataka usmerena je ka struktuiranim izvorima, poput testova znanja.

U klasičnoj literaturi vezanoj za integraciju podataka do pre 15-ak godina fokus za upite i mapiranje bio je na relacionom modelu. Krajem 90-ih godina prošlog veka istraživači su interesovanje usmerili ka XML-u. Ovaj model postao je de-fakto standard za pronalaženje i razmenu podataka, pa je bio idelan za interoperabilne sisteme. Danas XML i njegovi jezici za upite predstavljaju osnovni interfejs web servisa, XML orjentisanih baza podataka i dosta drugih primena.

Integracija podataka iz različitih izvora XML-a sa sobom nosi određene probleme. Tus [9] navodi da je prvi mapiranje šeme. Šema u kojoj su uslovi izraženi upitom (globalna šema) mora biti mapirana na šemu/šeme izvora nad kojim se vrši upit. Najjednostavniji

pristup je kroz odgovarajuće atribute, u smislu da svojstva atributa jedne reprezentacije odgovara svojstvima atributa druge. Međutim ovo je jako kompleksno kada su XML reprezentacije različito strukturirane, iako je koncept mapiranja semantički isti. U primeru zasnovanom na radu koji je predstavio Halevi [10], ilustrovani su neki od problema mapiranja XML šema. Primer DTD koda dat je na narednoj slici.

Izvor1.xml DTD: odgovori zadatak* pitanje ucenik* ime odgovor* bodovi	Izvor2.xml DTD: ucenici ucenik* ime test* pitanje odgovor
--	---

Slika 2. Različiti načini mapiranja XML šeme

Primer prikazuje kako jednostavna šema kojom opisujemo date odgovore i zadatke može imati različite oblike. Težina kreiranja mape među njima zavisi od cilja mapiranja. Ukoliko je potrebna jednostavna migracija (prevođenje podataka iz jedne šeme u drugu) tada bi se ona mogla ostvariti jednostavnim template-om. Međutim, ukoliko je potrebno vršiti upite nad njima, potrebna je daleko kompleksnija strategija.

#### 4. OBRADA PODATAKA

XSLT je jezik standardizovan od W3C-a za transformaciju XML dokumenata. On je komponenta XML Stylesheet jezika koja se može nezavisno koristiti, specijalno za integraciju podataka. Transformacija u XSLT-u koja se naziva stylesheet opisuje pravila transformacije izvornog XML dokumenta u određeni. Kada se uzorak (eng. pattern) poklopi sa elementom izvornog XML stabla, odgovarajućim template-om se generiše XML kôd određeni dokumenta. Trenutna verzija XSLT 2.0 (iako preporučena od strane W3C-a) i dalje nije podržana od strane browser-a, stoga je i dalje u širokoj upotrebi XSLT 1.0. Mogućnosti u transformaciji XML dokumenata XSLT čini prirodnim izborom u integraciji podataka. Kada se mapiraju heterogene XML šeme, XSLT stylesheet-ovi se mogu ručno kodirati ili polu-automatski generisati kako bi izvršili konverziju.

XPath obezbeđuje sredstvo za lociranje specifičnog čvora u stablu XML dokumenta. Dizajniran je za rad sa XSLT-om. Izrazi ovog jezika mogu se koristiti za manipulaciju stringovima, numeričke kalkulacije i bulovu algebru. Glavna primena je u adresiranju delova XML dokumenta slično URL-ovima kojim se olakšava navigacija kroz sadržaj.

Relacije između XPath čvorova jednostavno definišu precizne i očigledne relacije među čvorovima XML dokumenta. Izrazi su bazirani na putanji unutar XML dokumenta i slični su navigaciji direktorijumima pod UNIX-om.

Atomička (tekstualna) vrednost nema relaciju sa drugim čvorovima, i to je vrednost tekstualnog ili čvora atributa. Ona predstavlja stvarnu vrednost podatka, ne metapodatak.

Sledi primer upita nad XML fragmentom dokumenta *test* čija je struktura prikazana kódom:

```
<?xml version="1.0" encoding="utf-8"?>
<?xml:stylesheet type="text/xsl" href="format.xsl" ?>
<test>
  <nastavnik>
    <ime>Veljko</ime>
    <prezime>Aleksic</prezime>
    <predmet predmet="TIO"></predmet>
    <razred razred="VII"></razred>
  </nastavnik>
  <zadaci>
    <zadatak idz="1">
      <oblast>Energetika</oblast>
      <tekst>Generator je vrsta elektricne masine</tekst>
      <ucenik>
        <ime>Petar Markovic</ime>
        <bodova>1</bodova>
      </ucenik>
      <ucenik>
        <ime>Mirko Dobric</ime>
        <bodova>0</bodova>
      </ucenik>
      <ucenik>
        <ime>Slavko Petrovic</ime>
        <bodova>1</bodova>
      </ucenik>
    </zadatak>
    <zadatak idz="2">
      <oblast>Saobracaj</oblast>
      <tekst>Cetvorotaktni motor je motor SUS</tekst>
      <ucenik>
        <ime>Petar Markovic</ime>
        <bodova>1</bodova>
      </ucenik>
      <ucenik>
        <ime>Mirko Dobric</ime>
        <bodova>1</bodova>
      </ucenik>
      <ucenik>
        <ime>Slavko Petrovic</ime>
        <bodova>0</bodova>
      </ucenik>
    </zadatak>
  </zadaci>
</test>
```

Elementarni prikaz podataka izvršen je .xls skriptom koja kao rezultat daje prikaz imena učenika i ostvarenih bodova po zadacima:

```

<HTML xmlns:xsl="http://www.w3.org/TR/WD-xsl">
<BODY>
<TABLE CELLPADDING="5" CELLSPACING="1" BORDER="1">
<TH BGCOLOR="silver">Zadatak</TH>
<TH BGCOLOR="silver">Ucenik</TH>
<TH BGCOLOR="silver">Bodovi</TH>
<xsl:for-each select="//ucenik">
<TR>
<TD><xsl:value-of select="..@idz"/></TD>
<TD><xsl:value-of select="ime"/></TD>
<TD><xsl:value-of select="bodova"/></TD>
</TR>
</xsl:for-each>
</TABLE>
</BODY>
</HTML>

```

Narednom .xml skriptom određen je ukupan broj ostvarenih bodova na testu:

```

<xsl:stylesheet version="2.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:fn="http://www.w3.org/2005/xpath-functions"
xmlns:xdt="http://www.w3.org/2005/xpath-datatypes">
<xsl:template match="/">
<HTML>
<BODY>
Ukupno bodova = <xsl:value-of select="sum(//bodova)" group-by="//zadatak"/>
</BODY>
</HTML>
</xsl:template>
</xsl:stylesheet>

```

## 5. ZAKLJUČAK

Integracija podataka različitih heterogenih izvora je tradicionalan problem baza podataka i veštačke inteligencije. Iako je XML nesporno poboljšao interoperabilnost nad podacima i metapodacima u digitalnom okruženju, novi načini prikaza informacija danas predstavljaju ontologije, namesto šema. Uprkos činjenici da je XML proistekao iz tehnologije stare preko 30 godina, jasno je da danas predstavlja centralni deo integracije podataka. Ipak, nije jasno gde je limit prednosti korišćenja struktuiranih podataka, niti kakvi pristupi nas očekuju u bliskoj budućnosti.

Obrada rezultata testova znanja korišćenjem XML integracije podataka pokazala se ostvarivom, uz naglasak da je za kvalitetniju obradu podataka ipak potrebno koristiti neki od razvijenih sistema za integraciju poput XSquare, Enosys ili Nimble.

**LITERATURA**

- [1] Bjekić, D., Papić, Ž.M., *Ocenjivanje u srednjem stručnom obrazovanju*, Ministarstvo prosvete i sporta, Beograd, 2005.
- [2] Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison-Wesley-Longman Publishing, 1999.
- [3] May, W., *A Logic-Based Approach to XML Data Integration*, Institut für Informatik, Freiburg, 2005.
- [4] Salminen, A., Tompa, F., *Communicating with XML*, Springer, New York, 2011.
- [5] Blue, A., *Davis drives home the importance of being knowledge based*, Information Outlook, Alexandria, 1997.
- [6] Lenzerini, M., *Data Integration: A Theoretical Perspective*, PODS, Madison, 2002.
- [7] Ullman, J., *Information Integration Using Logical Views*, ICDT, Delphi, 1997.
- [8] Callan, J., *Distributed information retrieval*, Advances in information retrieval, Kluwer Academic Publishers, Boston, 2000.
- [9] Tous, R., *Data integration with XML and Semantic Web Technologies*, Doctorate in Computer Science and Digital communication, Barcelona, 2006.
- [10] Halevy, Y., Ives, Z., Mork, P., Tatarinov I., Data management infrastructure for semantic web applications, World Wide Web Conference, Budapest, 2003.