

SOFTVERSKI SISTEM ZA OBRADU TEMA IZ ARHITEKTURE I ORGANIZACIJE RAČUNARA

Dejan Prodanović, Boško Nikolić, Jovan Đorđević, Elektrotehnički fakultet Beograd

Sadržaj — *U okviru visokog obrazovanja mnogih usmerenja oblast arhitekture i organizacije računara zauzima važno mesto. Pored mnogih osobenosti, jedna od njenih karakteristika je i da se sastoji iz velikog broja teorijskih principa i koncepata. Zato je predavačima iz ove oblasti veoma bitna svaka informacija o tome kako su studenti razumeli predavanu materiju i koje teme im predstavljaju najveći problem. Sa druge strane studentski forume i slični Web bazirani sistemi indirektno sadrže upravo takve informacije. U ovom radu je predstavljen jedan sistem koji ma inteligentan način vrši klasifikaciju i povezivanje koncepata ove oblasti iz studentskih foruma. U realnom vremenu vrši se analiza sadržaja, pronalaze i ocenjuju koncepti, i na taj način daje se profesoru prikaz najzastupljenijih i najtraženijih tema.*

Gljučne reči — *Analiza sadržaja, pomoć prilikom učenja, ocenjivanje koncepata.*

I. UVOD

OBLAST arhitekture i organizacije računara zauzima bitno mesto u nastavi mnogih visokoškolskih ustanova i kod nas i u svetu. Zato se i obraća veća pažnja na probleme koji se pojavljuju tokom predavljanja tema iz ove oblasti. Jedan od karakterističnih problema je i veliki broj apstraktnih pojmova i koncepata koji se objašnjavaju na predavanjima i vežbama [1]. Uobičajeno rešenje je da se pored standardnog pristupa nastavi organizuje i praktičan rad na laboratorijskim vežbama sa simulatorima reprezentativnih računarskih sistema [2]. Kao primer može poslužiti Elektrotehnički fakultet u Beogradu, gde je za ove potrebe razvijen niz računarskih sistema i odgovarajućih softverskih grafičkih simulatora kojima se na laboratorijskim vežbama demonstriraju odgovarajuće teme iz ovih oblasti [3].

Pored navedenih uobičajenih metoda nastave, studenti koriste i druge prednosti savremenih informacionih tehnologija i elektronskog učenja. Već uobičajena mogućnost je komunikacija sa drugim studentima i kolegama putem foruma i elektronskih razgovora vezanih za neki predmet ili oblast. Sa druge strane i profesorima je od velikog interesa da dobiju verodostojne informacije o razumevanju studenata predavane teme pre samog

polaganja ispita. Sadržajne analize prihvaćenog gradiva bi predavačima bile od značajne pomoći prilikom verifikacije ili promene sopstvenih kurseva.

Kako su forumi vezani za predmete iz računarske tehnike i informatike već realizovani i koriste se duži niz godina, na Elektrotehničkom fakultetu u Beogradu je implementiran sistem koji inteligentnim pristupom analizira sadržaj, pronalazi i ocenjuje koncepte iz oblasti arhitekture i organizacije računara u okviru foruma. Na taj način korisnik dobija detaljnu analizu tema koje su studentima od najvećeg značaja i te informacije, kao predavač, može iskoristiti za poboljšanje svojih kurseva.

U daljem tekstu dat je detaljan opis problema koji je rešavan, opis softverskog sistema koji se koristi u navedene svrhe i predstavljeni su dobijeni rezultati.

II. OPIS PROBLEMA

Problem, koji su autori sistema opisanog u radu, pokušali da reše, je kako iz grupe tema određenog foruma izdvojiti koncepte i termine koji su najčešće korišćeni. Ovakvom inteligentnom pretragom i analizom profesor bi mogao da dobije kvantitativnu i kvalitativnu ocenu šta je ono što iz njegovih kurseva najviše zanima studente. Tako bi dobio informacije do kojih ne može doći direktnim kontaktom, jer studenti putem foruma uspostavljaju drugačiju i neposredniju vrstu komunikaciju. Kao validaciju svojih pretpostavki i radi smanjenja početnog domena analize izabrana je oblast arhitekture i organizacije računara.

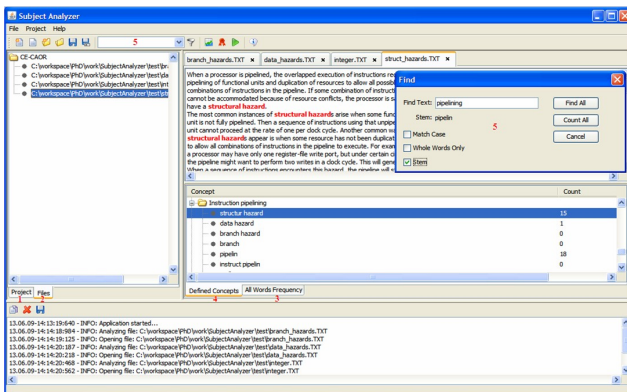
Stručna udruženja IEEE i ACM su u okviru svojih preporuka oblast arhitekture i organizacije računara podelili na osnovna poglavlja i odeljke unutar poglavlja [4]. Tako da su upravo ove preporuke autori iskoristili prilikom primene tehnika klasifikacije teksta, i to tako što su kao klasifikacione kategorije uzimali preporučene odeljke.

Sa druge strane, za sam proces klasifikacije može se primeniti neki od poznatih algoritama [4], kao što su: *Bayes formula, kNN, primena entropijskog modela, stabla odlučivanja...* Zajednička karakteristika navedenih algoritama jeste početni korak, gde se tekst transformiše iz dugačkog skupa znakova u objekte koji su pogodni za dalju obradu ili neku operaciju u okviru specifičnog posla. Tekst se može reprezentovati i kao vektor koji sadrži frekvencije reči koje se nalaze u određenom rečniku pojmova. Iz tako formiranog vektorskog prostornog modela vrši se dalja analiza.

III.OPIS SOFTVERSKOG SISTEMA

Detalniji rečnik pojmova može se formirati na osnovu postojećeg znanja ili analizom tekstova iz knjiga i skripti, koji su za odgovarajuću oblast reprezentativni, gde se izdvajaju reči sa visokom frekvencijom, formiraju složeniji pojmovi koji na jasan i nedvosmislen način opisuju pojavu u određenoj oblasti. Za ovakvu operaciju potrebno je uvesti određenu automatizaciju, pa je uočena potreba za izradom softverskog alata pomoću kojeg će se vršiti statistička analiza tekstova. Prilikom analize alat treba da omogući korisniku vizuelnu navigaciju i podešavanje svih bitnih parametara.

Na slici 1 je prikazan glavni prozor aplikacije za statističku analizu tekstova.



Sl 1. Alat za statističku analizu tekstova

Sistem omogućava:

1. Prikaz svih poglavlja, odeljaka i definisanih pojmova za odeljke u formi stabla sa mogućnošću unosa novih koncepata, kao i izmene i brisanja postojećih koncepata. Poglavlja, odeljci i koncepti su definisani u *xml* datoteci.
2. Dodavanje tekstualnih datoteka u skup za analizu tekstova.
3. Određivanje broja pojavljivanja svake reči u tekstu., gde se pod jednom reči podrazumeva se skup znakova između definisanih znakova za razdvajanje.
4. Određivanje broja pojavljivanja definisanih koncepata za određenu oblast.
5. Mogućnost pretrage po tekstu sa opcijama koje su dostupne u svim naprednim editorima teksta.

Na slici 2 prikazan je rezultat izdvajanja svih reči koje se pojavljuju u reprezentativnom tekstu. Kao primer posmatrao se tekst iz navedene oblasti arhitekture i organizacije računara i to oblast protočne obrade (pipeline), sa naglaskom na problem strukturnog hazarda.

Na ovakav način dobija se kvantitativna ocena pojavljivanja određenih koncepata. Ali, to je samo jedna mogućnost pristupa i mogu se uočiti neki nedostaci.

Word	Stemmed Word	Count
hazard	hazard	15
structural	structur	15
instruction	instruct	15
pipeline	pipelin	11
processor	processor	11
1		10
cycle	cyd	10
instructions	instruct	10

Sl 2. Prikaz pojavljivanja najfrekventnijih reči u reprezentativnom tekstu koji opisuje problem strukturnog hazarda

Posmatrajući entitete kao gradivne i funkcionalne jedinice nekog sistema, uočava se da svaki entitet može imati više podentiteta ili nadeniteta, biti u relaciji sa drugim entitetima, imati odgovarajuća svojstva (atribute) i nad njim se mogu definisati određene operacije. Stoga sledi zaključak da se entiteti, kao i realicije između njih mogu sadržajnije opisati uvođenjem baze znanja putem ontologija.

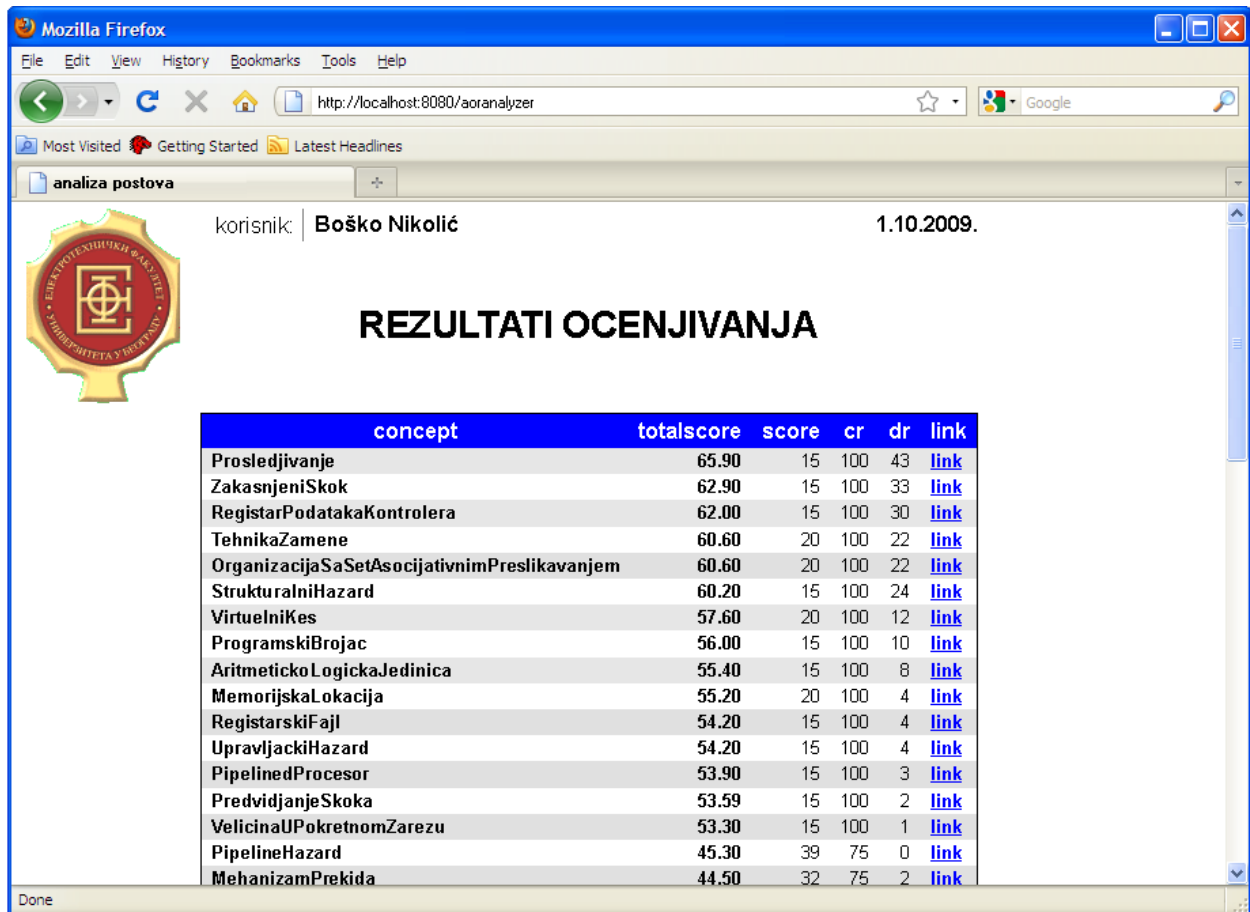
Korišćenjem alata *Protégé* [7] dobijena je ontologija u domenu arhitekture i organizacije računara. Prilikom realizacije su korišćeni rezultati dobijeni analizom svih tekstova iz materijala sa predavanja, vežbi na tabli i u laboratoriji, koristeći se alatom za statističku analizu tekstova. Kao najistaknutiji *markup* jezik za definisanje ontologija, korišćen je *OWL* (*Web Ontology Language*).

Ontologija sadrži koncepte iz domena arhitekture i organizacije računara i sačinjena je od sledećih klasa u korenu hijerarhije:

- MemoryArchitecture,
- ProcessorArchitecture,
- ComputerArchitecture,
- IOSystemArchitecture,
- Bit,
- ElementsOfMemoryArchitecture,
- ElementsOfProcessorArchitecture,
- ElementsOfIOSystemArchitecture,
- InterfaceAndCommunication,
- Modul,
- ParallelismLevel,
- ComputerOrganization,
- PipelineHazard,
- Data,
- Programme,
- Signal

IV.REALIZACIJA SOFTVERSKOG SISTEMA

Razvoj Interneta doprineo je velikoj ekspanziji web tehnologija, odnosno programskog jezika Java kao njihove okosnice. Višeslojne web aplikacije donele su brojne pogodnosti u razvoju savremenih informacionih sistema kako za inženjere softvera tako i za krajnje korisnike.



concept	totalscore	score	cr	dr	link
Prosledjivanje	65.90	15	100	43	link
ZakasneniSkok	62.90	15	100	33	link
RegistarPodatakaKontrolera	62.00	15	100	30	link
TehnikaZamene	60.60	20	100	22	link
OrganizacijaSaSetAsocijativnimPreslikavanjem	60.60	20	100	22	link
StrukturalniHazard	60.20	15	100	24	link
VirtuelniKes	57.60	20	100	12	link
ProgramskiBrojac	56.00	15	100	10	link
AritmetickoLogickaJedinica	55.40	15	100	8	link
MemorijskaLokacija	55.20	20	100	4	link
RegistarskiFajl	54.20	15	100	4	link
UpravljackiHazard	54.20	15	100	4	link
PipelineProcesor	53.90	15	100	3	link
PredvidjanjeSkoka	53.59	15	100	2	link
VelicinaUPokretnomZarezu	53.30	15	100	1	link
PipelineHazard	45.30	39	75	0	link
MehanizamPrekida	44.50	32	75	2	link

Sl 3. Prikaz rezultata ocenjivanja postova

Nakon izvršene identifikacije korisniku je omogućen uvid u rezultate ocenjivanja, kao što je prikazano na slici 3. Za svaki ocenjeni koncept definisane su u tabeli su definisane sledeće kolone: naziv klase (koncepta), ukupna ocena za koncept i hiperlink. Nakon aktivacije hiperlinka dobija se skup postova koji je rezultat unije dva skupa:

1. U prvom skupu nalaze se postovi u kojima se pojavljuje koncept.
2. Drugi skup čine postovi koji su u relaciji sa postovima iz prvog skupa.

Funkcionalnom dekompozicijom posmatranog sistema, mogu se izdvojiti sledeće faze: *formiranje i korišćenje rečnika, čitanje ontologije, parsiranje dokumenta, pronalaženje i ocenjivanje koncepata*, od kojih prve dve čine inicijalni, a poslednje dve jedan izvršni ciklus u radu softverskog sistema.

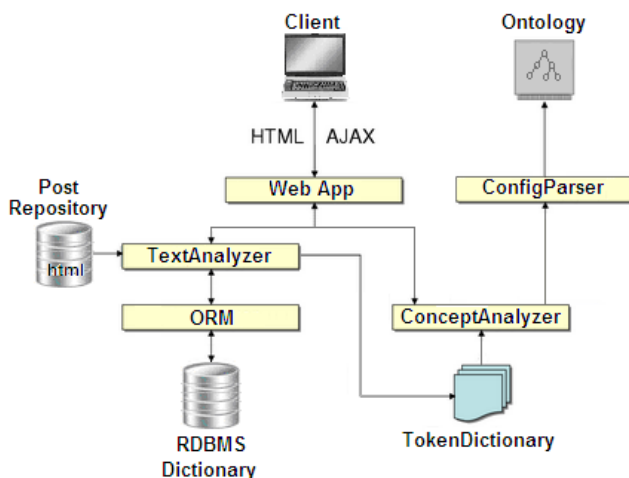
Posmatrajući semantička podstabla u ontologiji, uočeni su sledeći parametri koji utiču na ocenu:

1. **Koncentracija koncepta (score)** predstavlja meru prisutnosti koncepta i njegovih potomaka u semantičkom podstablu.
2. **Relevantnost koncepta(cr)** predstavlja meru zastupljenosti svih njegovih potomaka u dokumentu.
3. **Relevantnost dokumenta (dr)** predstavlja statističko prisustvo koncepta u dokumentu.

Kao primer ocenjivanja može se pomstarati sledeći slučaj:

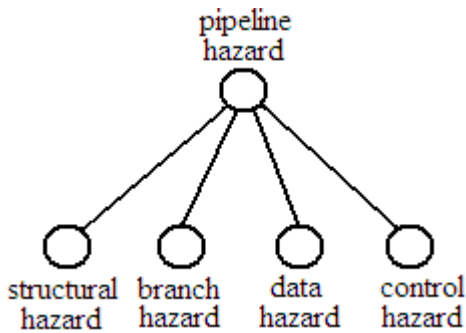
“Da li neko može da mi objasni korak 4 u primeru iz skripte, gde je ilustrovano preklapanje između *data hazarda* i *control hazarda*? Hvala.”

Na slici 5, prikazano je podstablo iz ontologije u kojem su opisane vrste hazarda kod procesora sa pipeline arhitekturom.



Sl. 4. Softverski sistem

Na slici 4, prikazani su svi bitni softverski entiteti i repozitorijumi koji se koriste u radu softverskog sistema.



Sl 5. Vrste hazarda kod procesora sa pipeline arhitekturom definisani u ontologiji

Koncepti koji su u ontologiji definisani kao potomci koncepta *pipeline hazard* i pojavljuju se u tekstu, prikazani su u *italic* stilu. U tabeli 1 prikazani su parametri koji utiču na formiranje konačne ocene, gde koeficijenti srazmere imaju sledeće vrednosti: $w_{sc} = 0.2$, $w_{cr} = 0.5$, $w_{dr} = 0.3$

Tabela 1: Rezultati ocenjivanja

concept	w_c	nsub	nsc	nntokens	sc	cr	dr	totalscore
<i>data hazard</i>	20	1	1	11	20	100	9.1	56.73
<i>control hazard</i>	20	1	1	11	20	100	9.1	56.73
<i>structural hazard</i>	0	1	0	11	0	0	0	0
<i>branch hazard</i>	0	1	0	11	0	0	0	0
<i>pipeline hazard</i>	0	5	2	11	10	40	0	22

V. ANALIZA REZULTATA

Iz rezultata mogu se uočiti ocene za one koncepte koji se pojavljuju u tekstu, kao i za one koncepte koji su u hijerarhiji iznad njih nezavisno od pojavljivanja. U datom primeru, superkoncept se ne pojavljuje u tekstu, ali je normalno očekivati da će biti ocenjen, jer su i strukturni hazard i data hazard tipovi pipeline hazarda, naravno nižom ocenom jer postoji više vrsta pipeline hazarda. Ocenjene koncepte ne treba posmatrati individualno, već na nivou grupe. Analizirajući rezultate prikazane na slici 5, mogu se uočiti koncepti iz oblasti memorijske hijerarhije i arhitekture procesora sa protočnom obradom, što su inače oblasti koje se izučavaju na predmetu Arhitektura i organizacija računara 1 na Elektrotehničkom fakultetu u Beogradu, pa su rezultati očekivani.

VI. ZAKLJUČAK

Pored standardnog pristupa u realizaciji nastave, razvoj savremenih informacionih tehnologija u oblasti elektronskog učenja zahteva dodatni pristup. Sadržajne analize studentskih mišljenja bi bile od velike pomoći predavačima prilikom verifikacije ili izmene kurseva.

Sistem opisan u ovom radu koristi se u okviru nastave iz više predmeta Arhitekture i organizacije računara na

Elektrotehničkom fakultetu u Beogradu. Forumi koji postoje u okviru ovih predmeta sadrže pitanja i odgovore koje studenti postavljaju u okviru međusobne komunikacije. Sistem je konfigurisan na osnovu tekstova predavanja i vežbi iz ovih predmeta. Na osnovu ove konfiguracije sistem izvršava analizu datih postova. Koncepti koji se najviše puta pojavljuju i koji su najbolje ocenjeni se prezentuju predavačima. Takođe, predavači imaju mogućnost interaktivne upotrebe sistema, tako što u svakom trenutku mogu da pogledaju postove vezane za određene koncepte.

LITERATURA

- [1] Computing Curricula – Computer Engineering, IEEE Computer Society and ACM, [Online] Available: <http://www.eng.auburn.edu/ece/CCCE/CCCE-FinalReport-2004Dec12.pdf>
- [2] John L. Hennessy, David A. Patterson, Computer Architecture A Quantitative Approach Fourth Edition.
- [3] J. Djordjevic, B. Nikolic, A. Milenkovic, Flexible Web-based Educational System for Teaching Computer Architecture and Organization, IEEE Transactions on Education, Vol. 48, No. 2, pp. 264–273, May 2005.
- [4] Christopher D. Manning, Hinrich Schütze, Foundations of Statistical Natural Language Processing, The MIT Press Cambridge, Massachusetts London, England
- [5] Stuart J. Russell, Peter Norvig, Artificial intelligence: A Modern Approach
- [6] Mark Watson, Practical Artificial Intelligence Programming With Java
- [7] Protégé, Ontology creation tool, <http://protege.stanford.edu>

ABSTRACT

Computer architecture and organization plays an important role in many fields of higher education. An important feature of this area is that it comprises many theoretic principles and concepts. Hence the lecturers need every information on how the students comprehend covered matter and what the biggest issues are. On the other hand, student forums and other web-based systems can indirectly provide just the kind of information needed. This paper presents a software system that uses intelligence to classify concepts of this area in student forums and link them. The real-time content analysis is done, concepts are found and evaluated and lecturers are presented with the most frequent topics.

SOFTWARE SYSTEM FOR PROCESSING COMPUTER ORGANIZATION AND ARCHITECTURE TOPICS

Dejan Prodanović, Boško Nikolić, Jovan Đorđević