



УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У
НОВОМ САДУ



Виолета Вукобрат

Реализација система у *cloud* окружењу за сакупљање мета-података расположивих на интернету

ДИПЛОМСКИ РАД
- Основне академске студије -

Нови Сад, 2014.



УНИВЕРЗИТЕТ У НОВОМ САДУ • ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, РБР:		
Идентификациони број, ИБР:		
Тип документације, ТД:	Монографска документација	
Тип записа, ТЗ:	Текстуални штампани материјал	
Врста рада, ВР:	Завршни (Bachelor) рад	
Аутор, АУ:	Виолета Вукобрат	
Ментор, МН:	Др Јелена Ковачевић	
Наслов рада, НР:	Реализација система у <i>cloud</i> окружењу за сакупљање мета-података расположивих на интернету	
Језик публикације, ЈП:	Српски / латиница	
Језик извода, ЈИ:	Српски	
Земља публикавања, ЗП:	Република Србија	
Уже географско подручје, УГП:	Војводина	
Година, ГО:	2014.	
Издавач, ИЗ:	Ауторски репринт	
Место и адреса, МА:	Нови Сад; трг Доситеја Обрадовића 6	
Физички опис рада, ФО: <small>(поглавља/страна/ цитата/табела/слика/графика/прилога)</small>	8/36/0/7/3/0/1	
Научна област, НО:	Електротехника и рачунарство	
Научна дисциплина, НД:	Рачунарска техника	
Предметна одредница/Кључне речи, ПО:	Интернет, мултимедијални садржај, дигитална телевизија, мета-подаци	
УДК		
Чува се, ЧУ:	У библиотеци Факултета техничких наука, Нови Сад	
Важна напомена, ВН:		
Извод, ИЗ:	У овом раду представљена је реализација система за сакупљање мултимедијалног садржаја и мета-података са интернета, везаних за телевизијске емисије.	
Датум прихватања теме, ДП:		
Датум одбране, ДО:		
Чланови комисије, КО:	Председник: Др Иштван Пап	
	Члан: Др Милан Бјелица	Потпис ментора
	Члан, ментор: Др Јелена Ковачевић	



UNIVERSITY OF NOVI SAD • FACULTY OF TECHNICAL SCIENCES
21000 NOVI SAD, Trg Dositeja Obradovića 6

KEY WORDS DOCUMENTATION

Accession number, ANO :			
Identification number, INO :			
Document type, DT :	Monographic publication		
Type of record, TR :	Textual printed material		
Contents code, CC :	Bachelor Thesis		
Author, AU :	Violeta Vukobrat		
Mentor, MN :	PhD Jelena Kovačević		
Title, TI :	Realization of a system in a cloud environment for collecting metadata available on the Internet		
Language of text, LT :	Serbian		
Language of abstract, LA :	Serbian		
Country of publication, CP :	Republic of Serbia		
Locality of publication, LP :	Vojvodina		
Publication year, PY :	2014.		
Publisher, PB :	Author's reprint		
Publication place, PP :	Novi Sad, Dositeja Obradovica sq. 6		
Physical description, PD : (chapters/pages/ref./tables/pictures/graphs/appendixes)	8/36/0/7/3/0/1		
Scientific field, SF :	Electrical Engineering		
Scientific discipline, SD :	Computer Engineering, Engineering of Computer Based Systems		
Subject/Key words, S/KW :	Internet, multimedia content, digital television, metadata		
UC			
Holding data, HD :	The Library of Faculty of Technical Sciences, Novi Sad, Serbia		
Note, N :			
Abstract, AB :	This paper presents a realization of a system for collecting multimedia content and metadata from the Internet, related to television shows.		
Accepted by the Scientific Board on, ASB :			
Defended on, DE :			
Defended Board, DB :	President:	PhD Ištvan Pap	
	Member:	PhD Milan Bjelica	Menthor's sign
	Member, Mentor:	PhD Jelena Kovačević	

SADRŽAJ

1. Uvod	1
2. Teorijske osnove.....	4
2.1 Značaj ponude sadržaja	4
2.2 Vrste sadržaja na internetu.....	5
2.3 Meta-podaci i "semantički veb"	5
2.4 Izvori informacija na internetu.....	6
2.4.1 Hypertext Transfer Protocol - HTTP.....	6
2.4.2 Komunikacija sa internet servisima koji nude aplikativnu programsku spregu za pristup sadržaju	7
2.4.2.1 Extensible Markup Language - XML	7
2.4.2.2 JavaScript Object Notation - JSON.....	7
2.4.3 Komunikacija sa internet servisima koji ne nude aplikativnu programsku spregu za pristup sadržaju.....	7
2.5 Java Enterprise Edition – JEE.....	8
2.5.1 Enterprise Java Bean - EJB	9
2.6 EPG podaci o televizijskim emisijama	10
3. Analiza problema.....	11
4. Koncept rešenja	13
4.1 Izvori sadržaja na internetu	14
4.2 Servisi za obezbeđivanje sadržaja.....	15
4.3 Modul za rukovanje sakupljanjem dodatnog sadržaja	15
4.4 Modul za sakupljanje slika koje odgovaraju televizijskim emisijama.....	16
5. Programsko rešenje.....	17
5.1 Internet servisi.....	17
5.1.1 HTTP klijent modul.....	18
5.2 Servisi za obezbeđivanje sadržaja.....	18
5.3 Modul za rukovanje sakupljanjem dodatnog sadržaja	21
5.4 Modul za sakupljanje slika koje odgovaraju televizijskim emisijama.....	22

6. Ispitivanje i rezultati	24
6.1 Ispitivanje brzine sakupljanja sadržaja i slika.....	24
6.1.1 Ispitivanje brzine sakupljanja dodatnog sadržaja	24
6.1.2 Ispitivanje brzine sakupljanja slika	25
6.2 Ispitivanje mišljenja korisnika o sakupljenom sadržaju	26
7. Zaključak	28
8. Literatura	29
Prilog.....	30

SPISAK SLIKA

Slika 4.1. Arhitektura realizovanog sistema	14
Slika 5.1 Postupak pronalazjenja dodatnog sadržaja vezanog za glumce koji glume u datoj seriji ili filmu	21
Slika 6.1 Prikaz rezultata sprovedene ankete koja ispituje mišljenje korisnika o sakupljenom dodatnom sadržaju	26

SPISAK TABELA

Tabela 5.1 Metoda <i>prepareQuery</i> servisa za obezbeđivanje sadržaja.....	19
Tabela 5.2 Metoda <i>fetchContent</i> servisa za obezbeđivanje sadržaja	20
Tabela 5.3 Metoda <i>fetchContentForEvent</i> modula za rukovanje sakupljanjem sadržaja22	
Tabela 5.4 Metoda <i>getContentServiceList</i> modula za rukovanje sakupljanjem sadržaja22	
Tabela 5.5 Metoda <i>registerContentServices</i> modula za rukovanje sakupljanjem sadržaja	22
Tabela 6.1 Prosečna vremena čekanja na pristizanje rezultata od modula za sakupljanje sadržaja za određene žanrove televizijskih emisija.....	25
Tabela 6.2 Prosečna vremena trajanja sakupljanja slika za televizijske emisije za određene veličine grupa paralelnih zahteva	25

SKRAĆENICE

API – *Application Programming Interface*, aplikativna programska sprega

URL – *Universal Resource Locator*, univerzalni lokator resursa

HTTP – *Hypertext Transfer Protocol*, mrežni protokol aplikativnog nivoa

JEE – *Java Enterprise Edition*, tehnologija za razvoj poslovne programske podrške

EPG – *Electronic Program Guide*, elektronski vodič kroz program

EJB – *Enterprise Java Bean*, deo JEE tehnologije, koristi se za implementaciju poslovne logike

1. Uvod

Ovaj rad se bavi opisom sistema koji za cilj ima da ponudi okruženje za sakupljanje dodatnog sadržaja sa interneta. Polazna tačka za prikupljanje sadržaja jeste televizijska emisija prosleđena od strane korisničkog uređaja (prijemnika digitalnog TV signala). Dodatan sadržaj se sastoji od multimedijalnog sadržaja i dodatnih informacija vezanih za prosleđenu televizijsku emisiju. U ovom poglavlju biće navedeni razlozi koji su doveli do ideje za stvaranjem teme rada.

Pojava digitalne televizije donela je određena poboljšanja za krajnje korisnike u odnosu na analognu televiziju. Korišćenjem efikasnih tehnika kompresije, modulacije i drugih digitalnih obrada signala, pored poboljšanja u kvalitetu prenosa i prezentacije slike i zvuka, stvorena je prilika za razvoj niza novih usluga koje se mogu ponuditi korisnicima. Te usluge su promenile klasičan način na koji se korisnici odnose prema svojim TV prijemnicima. Uvele su različite nivoe učešća korisnika prilikom korišćenja TV prijemnika, kao što su: izbor audio trake, prevoda, pregled raznih multimedijalnih i interaktivnih sadržaja, pristup internetu, itd. U poslednjih par godina na tržištu su se pojavili tzv. pametni televizori (eng. *Smart TV*). Ovaj termin se koristi za opis trenutne težnje spajanja interneta i mrežnih tehnologija sa modernim televizijskim prijemnicima i prijemnicima digitalnog televizijskog signala (eng. *STBs - set-top boxes*). Za razliku od prethodnih generacija ovih uređaja, čiji je fokus bio na tradicionalno emitovanom sadržaju (televizijskim kanalima koji se emituju od strane televizijskih kuća), trenutna generacija ovih uređaja više pažnje posvećuje novo uvedenim pojmovima kao što su: internet televizija, OTT sadržaj (eng. *OTT – Over The Top content*), video na zahtev (eng. *VoD – Video on Demand*), itd.

Internet televizija podrazumeva digitalnu distribuciju televizijskog sadržaja preko interneta.

Pod OTT sadržajem, ili OTT uslugama misli se na one aplikacije i usluge koje koriste internet konekciju za prenos audio, video ili drugog sadržaja, ali nisu povezane sa samim provajderom te internet konekcije. Tu spadaju i VoIP (eng. *Voice over IP - VoIP*) usluge kao što su npr. Skype, Viber, Apple-ov Facetime i dr., ali su u domenu televizije značajnije usluge koje nude video sadržaj, kao npr. Netflix i Hulu.

Video na zahtev omogućava gledaocima da gledaju video sadržaj koji se emituje, ali u proizvoljno vreme, ne samo u vreme emitovanja, kao i video sadržaj koji nije u vezi sa onim što se emituje na programu.

Pomenute usluge se za funkcionisanje umnogome oslanjaju na korišćenje interneta. Jedan od prisutnih otvorenih standarda koji kombinuje internet i televiziju jeste HbbTV (eng. *HbbTV – Hybrid Broadcast Broadband Television*) [1]. Ova tehnologija koristi postojeće mrežne tehnologije za dostavljanje i prikaz sadržaja korisnicima. Međutim, sadržaj koji trenutno nude HbbTV aplikacije je ograničen na uski skup informacija koje su odabrane od strane emitera kanala.

OTT usluge i HBB tehnologije [2] jesu nešto što će pratiti intenzivan razvoj i proširenje u narednom periodu. Njihov razvoj uticaće i na tzv. “pametne televizore”. Interaktivnost zajedno sa sadržajem vezanim za televiziju ima velikog potencijala da promeni način na koji gledaoci doživljavaju televiziju u narednoj deceniji. Neki od primera navedenog sadržaja jesu dobijanje dodatnih informacija o onome što je na programu, filmovima, vestima, sportu, vremenu i sl. Odatle je i potekla ideja za stvaranjem sistema koji bi gledaocima ponudio dodatan sadržaj vezan za televizijske emisije, oblikovan u skladu sa njihovim navikama i interesovanjima.

U narednim poglavljima predstavljen je jedan način realizacije sistema u *cloud* okruženju za prikupljanje multimedijalnog sadržaja i meta-podataka sa interneta vezanih za televizijske emisije. Prilikom izrade zadatka glavni cilj bio je stvaranje sistema koji će biti lako proširiv, kako bi se omogućila ponuda što šireg opsega informacija korisnicima.

Ovaj rad se sastoji od osam poglavlja.

U prvom poglavlju dat je kratak uvod koji daje razloge dolaska na ideju za temu ovog rada.

U drugom poglavlju objašnjeni su pojmovi i principi neophodni za razumevanje koncepta i programskog rešenja, čime se bave četvrto i peto poglavlje, dok su u trećem poglavlju dati glavni problemi koji su se pojavljivali prilikom izrade zadatka.

Šesto poglavlje sadrži rezultate odabrane grupe testova koji su vršeni nad programskim rešenjem. Testovi su vršeni kako bi se prikazali funkcionalnost i performanse korišćenja realizovanog rešenja u realnim uslovima.

U sedmom poglavlju dat je kratak osvrt na ceo rad.

Osmo poglavlje sadrži spisak literature korišćene prilikom pisanja rada.

2. Teorijske osnove

U ovom poglavlju objašnjeni su pojmovi i tehnologije koji su bili osnov za izradu rešenja zadatka. Dati su opis izvora informacija na internetu i način komunikacije sa njima i opis podataka potrebnih za shvatanje načina funkcionisanja sistema. Pored toga, ukratko je objašnjena i tehnologija koja je korišćena za izradu rešenja, radi razumevanja odluka donesenih prilikom organizacije i implementacije realizovane programske podrške.

2.1 Značaj ponude sadržaja

Iako je televizija doživela značajne promene od njene pojave pa do danas, osnovni razlozi za njeno korišćenje od strane gledalaca su ostali isti – gledaoci je koriste kao vid informisanja i zabave. Sa ubrzanim razvojem interneta, televizija kao vid informisanja je u određenoj meri izgubila na značaju, ali kao vid zabave je i dalje veoma popularna. Gledanje televizije je vreme kada gledaoci žele da budu zabavljeni, pri čemu bi im zabavan sadržaj odmah bio na raspolaganju, za razliku od npr. korišćenja interneta, gde korisnici sami traže sadržaj koji ih interesuje. Tokom godina, pojavljivale su se različite tehnologije, kao na primer MHEG (eng. *Multimedia and Hypermedia Experts Group*) i HbbTV (eng. *Hybrid Broadcast Broadband Television*) koje su podigle nivo interaktivnosti između korisnika i njihovih TV prijemnika i obogatile iskustvo njegovog korišćenja. Međutim, ključ privlačenja većeg broja korisnika leži u ponudi kvalitetnog, zanimljivog sadržaja, a ne samih tehnologija. Tako i danas, prilikom predstavljanja uređaja krajnjim korisnicima, samo nabranje aplikacija koje se nalaze na korisnikovom "pametnom televizoru" njemu znači mnogo manje od nabranja mogućnosti i sadržaja koje mu te aplikacije pružaju.

2.2 Vrste sadržaja na internetu

Cilj realizacije zadanog sistema jeste ponuda obogaćenog dodatnog sadržaja gledaocima vezanog za televizijske emisije koje su na programu. Pod ovim dodatnim sadržajem podrazumevaju se npr. video snimci vezani za filmove i serije, biografije glumaca koji glume u gledanim emisijama, najnovije vesti u domenu sporta, informacije o igračima, trenutne vesti u svetu i lokalno, vremenska prognoza i sl. Dakle, sadržaj koji se nudi gledaocima svodi se na tekstove, slike, internet stranice, video i audio snimke, itd. Takođe, treba napomenuti još jedan veoma značajan vid podataka koji je potrebno obezbediti da bi se navedeni sadržaj mogao prikupiti, a to su meta-podaci, odnosno, podaci o podacima. Na primer, u meta-podatke o nekom filmu bi spadao spisak glumaca koji u njemu glume, ko je režirao film, ko je radio muziku, godina i lokacija snimanja, trenutna ocena sadržaja na osnovu glasanja skupa gledalaca (eng. *rating*) i sl. Meta-podaci predstavljaju veoma bitnu tačku prilikom sakupljanja sadržaja sa interneta. Razlog njihovog značaja leži u tome što omogućavaju dodatno sakupljanje informacija o detaljima nekog pojma i time proširuju skup sadržaja koji se dostavlja a koji je vezan za taj određeni pojam. Takođe, organizacija informacija o nekom pojmu na strukturiran način, umnogome olakšava njihovo tumačenje od strane računara, odnosno programa.

2.3 Meta-podaci i "semantički veb"

Pojam "semantički veb" (eng. *Semantic web*) odnosi se na pokret vođen od strane WWW konzorcijuma (eng. *World Wide Web Consortium - W3C*) [3], međunarodnog udruženja koje se bavi razvojem standarda i podsticanjem diskusija vezanih za veb. Ovaj pokret se zalaže za preoblikovanje trenutnog uređenja veba, kojim dominiraju nestruktuirani ili slabo strukturirani dokumenti u tzv. "mrežu podataka". Ona predstavlja zajedničko okruženje za razmenu podataka, tako organizovanih da omogućuju lakše tumačenje i obradu od strane mašina, odnosno računara. Sam termin "semantički veb" skovan je od strane Tim Berners-Li-ja (eng. *Tim Berners-Lee*). Stvaranjem ovog sistema mašinama bi se omogućilo razumevanje i reagovanje na kompleksne ljudske zahteve na osnovu njihovih značenja, što zahteva da relevantni izvori informacija budu semantički strukturirani. Način postizanja ovog cilja je proširenje mreže sadašnjih veb dokumenata, međusobno povezanih hipervezama, namenjenih za čitanje od strane ljudi, tako što bi se ubacili meta-podaci o dokumentima i njihovim međusobnim odnosima tako da se omogući njihovo čitanje i tumačenje od strane mašina. Kao primer dokumenta na vebu koji je na prvi pogled dobro strukturiran jeste stranica na Vikipediji (eng. *Wikipedia, The Free Encyclopedia*) vezana za neki pojam. Podaci na toj stranici zaista jesu u određenoj meri dobro organizovani – tekst je uglavnom podeljen na smisljena poglavlja,

na samom početku dokumenta nalazi se mesto gde su nabrojana sva poglavlja i date hiperveze do njih, takođe, u okviru teksta su date hiperveze do ostalih pojmova od značaja, a na određenim stranicama čak postoji i mesto gde su izvučene najbitnije informacije (eng. *Infobox*) vezane za pojam za koji je vezana data stranica. Međutim, ova dobra organizovanost podataka možda važi za ljude, ali ne i za mašine. Naime, izvlačenje podataka sa jedne Vikipedija stranice programskim putem se svodi na tumačenje običnog teksta, bez uniformnog uređenja koje bi omogućilo neki vid automatizacije tog procesa. Sa druge strane, primer dobro organizovanog sadržaja, koji je pogodan za tumačenje od strane mašina, predstavlja struktura internet baze podataka Freebase. Freebase organizuje podatke u vidu grafa, gde su čvorovi grafa međusobno povezani različitim relacijama. Odnosno, svaki pojam na Freebase-u, sadrži određen broj osobina koje ga bliže određuju. Vrednosti tih osobina mogu biti i relacije datog pojma sa drugim pojmovima koji se nalaze u bazi, čime se ostvaruje odnos pojam-relacija-pojam. Freebase u suštini ne nudi mnogo više informacija od npr. Vikipedije, ali je njegov način organizacije podataka mnogo prirodniji za tumačenje od strane računara. To je omogućeno činjenicom da su svi pojmovi i veze između njih jasno definisani, odnosno prate unapred zadatu šemu.

2.4 Izvori informacija na internetu

Da bi se gledaocima isporučio dodatni sadržaj vezan za televizijske emisije, bilo je potrebno pronaći izvore na internetu koji ih nude. Prilikom traženja izvora prednost je data onima koji su nudili određenu aplikativnu programsku spregu (eng. *Application Programming Interface - API*) za pristup informacijama u odnosu na one koji su nudili samo sirove, neuređene podatke. Aplikativna programska sprega internet servisa za pristup podacima podrazumeva skup definisanih URL-ova (eng. *Universal Resource Locator - URL*) na koje se šalju upiti ka internet servisu, parametre upita (eng. *query parameters*) i odgovore na postavljeni upit koji imaju određeni format. Za komunikaciju klijenata sa internet servisima koristi se HTTP (eng. *Hypertext Transfer Protocol - HTTP*).

2.4.1 Hypertext Transfer Protocol - HTTP

Ovo je protokol aplikativnog nivoa koji funkcioniše po principu klijent-poslužilac modela. Način rada je takav da klijent šalje HTTP poruku zahteva poslužiocu, a poslužilac vraća poruku odgovora klijentu. Pri tome, poslužilac može da raspolaže resursima kao što su HTML (eng. *Hypertext Markup Language - HTML*) datoteke i drugi sadržaj, ili da obavlja određene funkcije u ime klijenta. Odgovor od poslužioca sadrži informacije o statusu završetka zahteva koji je postavio klijent u zaglavlju poruke, a može sadržati i zatraženi sadržaj u telu poruke.

HTTP resursi se identifikuju sa tzv. univerzalnim identifikatorima resursa (eng. *Universal Resource Identifiers - URIs*), odnosno univerzalnim lokatorima resursa (eng. *Universal Resource Locators - URLs*).

2.4.2 Komunikacija sa internet servisima koji nude aplikativnu programsku spregu za pristup sadržaju

Programska sprega za komunikaciju s klijentima koju nude ovi internet servisi se uglavnom zasniva na specificiranju parametara upita u okviru univerzalnog lokatora resursa. Ovi upiti se najčešće sastoje od specificiranja pojma koji se traži i raznih parametara filtera koji se može postaviti prilikom pretrage za tim pojmom. Takođe, potrebno je specificirati i željenu metodu, tj. akciju koja treba da se izvrši nad zadatim resursom (u većini slučajeva, željena metoda je “GET” (traženje sadržaja)). Na osnovu ovih ulaznih podataka internet servis dostavlja traženi sadržaj klijentu. Taj sadržaj ima određeni oblik koji klijent mora da zna kako da istumači. Najčešći formati koji se koriste za odgovore od strane internet servisa jesu XML (eng. *Extensible Markup Language*) i JSON (eng. *JavaScript Object Notation*). Kako su oba od navedenih formata predstave podataka široko u upotrebi, treba napomenuti da postoji niz raspoloživih programskih biblioteka koje omogućuju lakše rukovanje ovim formatima.

2.4.2.1 Extensible Markup Language - XML

XML je tekstualni format podataka koji se široko koristi za predstavljanje proizvoljnih struktura podataka, npr. u veb servisima. Osnovni termini vezani za XML dokumente jesu elementi i njihovi atributi. U okviru jednog elementa može se nalaziti više podelemenata, i sa te strane gledano, struktura XML dokumenta podseća na stablo. Detaljan opis strukture XML dokumenata nalazi se u specifikaciji [4].

2.4.2.2 JavaScript Object Notation - JSON

JSON predstavlja format za razmenu podataka, koji je jednostavan za pisanje i čitanje od strane ljudi a i za tumačenje i generisanje od strane mašina. To je tekstualni format koji je jezički potpuno nezavisan, ali koristi konvencije poznate programerima C (eng. *C programming language*) familije programskih jezika. Zasniva se na dve osnovne strukture: objektima i nizovima. Objekte čine skupovi parova imena i vrednosti.

2.4.3 Komunikacija sa internet servisima koji ne nude aplikativnu programsku spregu za pristup sadržaju

Pored internet servisa koji nude određenu aplikativnu programsku spregu za komunikaciju sa klijentima, postoje i ostali, čiji je broj mnogo veći i koji je ne nude. Kod ovih internet

servisa izvlačenje informacija se sastoji iz parsiranja HTML oblika stranice. Mane ovog formata jesu činjenice da je mnogo manje čitljiv, kako od strane ljudi, a tako i mašina, odnosno, parsiranje dokumenata u tom formatu predstavlja mnogo mukotrpniji posao od parsiranja dokumenata u prethodno spomenutim formatima. Struktura HTML dokumenata je specifična za svaku internet stranicu, nije jasno definisana niti dokumentovana, za razliku od slučaja kada postoji aplikativna programska sprega.

2.5 Java Enterprise Edition – JEE

Za realizaciju zadatka korišćena je JEE (eng. *Java Enterprise Edition*) tehnologija. JEE predstavlja Java softversku platformu inicijalno razvijenu od strane Sun Microsystems (eng. *Sun Microsystems*), a trenutno razvijanu od strane Orakl (eng. Oracle) korporacije. Ova platforma nudi aplikativnu programsku spregu za razvoj poslovne (eng. *enterprise*) programske podrške koja je prenosiva, skalabilna i kompatibilna sa postojećim tehnologijama. Pod poslovnom programskom podrškom podrazumeva se programska podrška razvijena da zadovolji potrebe neke veće organizacije, a ne individualnih korisnika. JEE proširuje standardno izdanje Java platforme (eng. *Java Standard Edition – Java SE*). Ona nudi programske sprege za objektno-relaciono mapiranje za rad sa bazama podataka, zatim sprege za rad sa distribuiranim višeslojnim aplikacijama i veb servisima, itd. Ova tehnologija je stvorena tako da zadovolji potrebe velikih, kompleksnih aplikacija, tako da i ona sama sadrži određen nivo kompleksnosti koji je u određenim slučajevima čini pravim izazovom za korišćenje.

JEE je definisana svojom specifikacijom. Da bi neki proizvod bio "u skladu sa JEE" (eng. "*Java EE compliant*"), on mora da zadovolji određene zahteve propisane specifikacijom. JEE specifikacija se sastoji od niza aplikativnih programskih sprege, kako za već postojeće tehnologije (npr. veb tehnologije), tako i za komponente koje su specifične samo za JEE platformu. Ono što zapravo implementira JEE specifikaciju jesu JEE aplikacioni poslužiooci (eng. *application servers*). Oni mogu implementirati sve aplikativne programske sprege definisane specifikacijom, ali i samo deo njih. Aplikacioni poslužilac se može posmatrati kao skup komponenti koje su na raspolaganju onima koji razvijaju programsku podršku kroz aplikativnu programsku spregu koja je definisana platformom. Jedan JEE aplikacioni poslužilac se brine o transakcijama, sigurnosti, skalabilnosti i uopšte rukovanju komponentama koje se razvijaju, kako bi se programerima omogućilo da se usredsrede na samu poslovnu logiku (eng. *business logic*), odnosno same zadatke koje treba da obavljaju aplikacije koje se razvijaju, umesto da se brinu o problemima infrastrukture i integracije.

Programi razvijani u ovoj tehnologiji se u najvećoj meri pišu u Java programskom jeziku. Sama platforma stavlja naglasak na tzv. “kodiranje po konvenciji” (eng. “*coding by convention*”). Ovo je paradigma u dizajnu programske podrške koja teži smanjenju broja odluka koje programeri moraju da donose, čime se dobija na jednostavnosti, gde nije nužno da se pritom nešto izgubi na fleksibilnosti. Jedna od primera primena ove paradigme jeste korišćenje anotacija za konfiguraciju. JEE anotacije predstavljaju meta-podatke čijim se postavljanjem iznad klasa, polja i metoda na jednostavan način aplikacionom poslužiocu dostavljaju dodatne informacije koje mogu uticati na njegovo ponašanje.

2.5.1 Enterprise Java Bean - EJB

Prilikom razvijanja poslovnih aplikacija, često se susreću iste vrste problema koje programeri moraju stalno iznova da rešavaju. Enterprise JavaBean-ovi su stvoreni da ponude standardna rešenja tih problema, kako bi programeri mogli da se koncentrišu na pravi problem koji treba da reše. Oni su definisani u EJB specifikaciji koja čini jednu od nekoliko aplikativnih programskih sprega definisanih u okviru JEE specifikacije. Ona nastoji da ponudi standardni način implementiranja poslovne logike koja se tipično nalazi u poslovnim aplikacijama. EJB-ovi predstavljaju komponente na strani poslužioca, čije su usluge na raspolaganju klijentima.

Da bi jedan EJB bio dostupan klijentima, mora postojati tzv. EJB poslužilac ili EJB kontejner. Ovaj poslužilac je sastavni deo aplikacionog poslužioca. Zadatak EJB kontejnera je da rukuje EJB-ovima, pri čemu to podrazumeva njihovo instanciranje, brigu o njihovom životnom ciklusu, sigurnosti, rukovanje transakcijama, nitima, itd. Time što je on preuzeo tu ulogu, programer je slobodan da se koncentriše na razvijanje poslovne logike, odnosno koda koji ostvaruje svrhu aplikacije.

EJB-ovi se implementiraju kao obične Java klase čije se metode prozivaju od strane klijenata. Ono što razlikuje klasu koja predstavlja EJB od obične Java klase jeste upotreba anotacija specifičnih za EJB specifikaciju. Jedne od njih se stavljaju iznad klase i daju do znanja kontejneru da je data klasa EJB. Postoji više anotacija specifičnih za EJB specifikaciju. Zbog potreba zadatka, značajno je spomenuti anotaciju “@Asynchronous” koja omogućava da se metoda iznad koje se ova anotacija stavi, poziva asinhrono, odnosno bude neblokirajuća.

Session bean je vrsta Enterprise JavaBean-a koja enkapsulira poslovnu logiku. Da bi pristupio aplikaciji koja je raspoloživa na poslužiocu, klijent proziva metode Session bean-a.

EJB-ovi pojednostavljaju razvoj velikih, distribuiranih aplikacija. Koriste se u slučajevima kada je potrebno stvoriti skalabilnu aplikaciju, npr. raširiti delove aplikacije na više mašina kako bi se prilagodila rastućem broju korisnika. Ne samo što se EJB-ovi mogu

izvršavati na različitim mašinama, nego njihova lokacija ostaje transparentna korisnicima. Takođe, oni su prenosive komponente i mogu se iskoristiti za stvaranje raznih novih aplikacija. Ukoliko koriste standardne aplikativne programske sprege (JSE i JEE), mogu se izvršavati na bilo kom JEE aplikacionom poslužiocu.

Postoje različite implementacije JEE, odnosno Java aplikacionih poslužilaca, koji su otvorenog koda (eng. *open source*). Aplikacioni poslužilac korišćen prilikom realizacije zadatka je JBoss AS, verzija 7.1.1. koji implementira JEE verziju 6 [5], odnosno EJB 3.1 specifikaciju.

2.6 EPG podaci o televizijskim emisijama

EPG (eng. *Electronic Program Guide*), odnosno elektronski vodič kroz program predstavlja osnovnu aplikaciju digitalne televizije. U najjednostavnijem obliku, ona sadrži raspored emitovanja televizijskih emisija na određenom kanalu za određeni dan. Pored imena televizijskih emisija i vremena emitovanja, može sadržati i njihov kratak opis. Malo naprednija varijanta vodiča sadrži pregled emisija koje se emituju na raznim kanalima, sa mogućnošću prebacivanja na te kanale, snimanja emisija, itd. U još naprednijoj varijanti ova aplikacija omogućava veliki izbor opcija, kao što su filtriranje po različitim kriterijumima, gledanje preporučenih emisija i sl. Jedan od načina isporuke potrebnih podataka za realizaciju EPG aplikacije jeste dostavljanje podataka o emisijama putem digitalnog transportnog toka. U slučaju DVB (eng. *Digital Video Broadcasting*) [6] transportnog toka, EPG podaci o emisijama se prenose u okviru EIT tabela (eng. *Event Information Table*). Neke od bitnijih informacija sadržanih u ovim podacima su: ime emisije, ime kanala na kom se emisija emituje, vreme početka i kraja emitovanja, trajanje, žanr, kraći i duži opis emisije, uzrast za koji je sadržaj emisije prikladan (eng. *parental rate*) itd. Ovi podaci korišćeni su kao ulaz u realizovani sistem za prikupljanje sadržaja sa interneta.

3. Analiza problema

U ovom poglavlju je ukratko opisan sam zadatak i navedeni su neki od problema koji su se pojavili prilikom njegove izrade. Nabrojani su zahtevi koji su morali biti ispunjeni kako bi se moglo realizovati rešenje zadatka.

Zadatak je bio napraviti sistem koji će pronaći dodatan multimedijalni sadržaj sa interneta vezan za televizijske emisije. Realizovani sistem treba da bude u stanju da prihvata zahteve klijenata za dodatnim sadržajem, prikupi taj sadržaj i isporuči ga klijentima. U ovom slučaju, klijenti su prijemnici digitalnog televizijskog signala.

Pronalazak izvora odgovarajućih sadržaja na internetu je bio prvi zahtev koji je bilo potrebno ispuniti. To je podrazumevalo pronalaženje raspoloživih servisa na internetu, istraživanje sadržaja koji oni nude i uslova njihovog korišćenja, kao i upoznavanje sa načinom njihovog rada, odnosno pristupom raspoloživom sadržaju. Takođe, bilo je potrebno izdvojiti manji skup servisa koji će se koristiti kao izvori za dobavljanje sadržaja, kako zbog kvaliteta sadržaja koji nude i mogućnosti pristupa tom sadržaju, tako i zbog vremenske ograničenosti izrade zadatka.

Ulazni podaci koje klijenti šalju sistemu prilikom zahtevanja sadržaja jesu EPG podaci o televizijskim emisijama, sakupljeni iz DVB prenosnog toka podataka ili sa interneta. Sledeći zahtev koji je trebalo ispuniti jeste bio pronaći način kako iskoristiti ove ulazne podatke za pretragu interneta. Ispunjenje ovog zahteva se ispostavilo da je poprilično veliki problem. Naime, glavni uzrok tome leži u činjenici da EPG podaci o televizijskim emisijama nemaju neku zajedničku strukturu, već su proizvoljne forme. Ovo važi i za emisije na istom kanalu (npr. emisije različitih žanrova), a kamoli za emisije sa različitih kanala. Takođe, dodatni problem jeste česta oskudnost ili čak i potpuno nepostojanje nekih informacija. Podaci o npr. sportskim emisijama su prilično oskudni – uglavnom se svode samo na naziv, gde su, u slučaju npr. utakmice, navedena imena dva tima koji učestvuju u utakmici, bez

ikakvog dodatnog opisa, i bez načina za određivanjem koja je utakmica u pitanju. Čak je i izvlačenje samih imena timova koji igraju problem, jer se način obeležavanja dva suparnička tima razlikuje od kanala do kanala, a neretko i od utakmice do utakmice (na istom kanalu).

Naredni problemi koji su se pojavili jesu određivanje sadržaja koji bi bio od značaja za različite vrste emisija i pronalazak sadržaja koji je posredno i neposredno vezan za datu emisiju. Na primer, u slučaju prenosa trke formule 1, sadržaj posredno vezan za tu emisiju činile bi aktuelne vesti iz sveta formule 1, a neposredno vezan sadržaj bila bi npr. dotadašnja statistika trka, podaci o učesnicima trke i rezultati date trke (ukoliko su u tom trenutku poznati). Prilikom određivanja sadržaja koji će se sakupljati za određene emisije, na donošenje odluka uticalo je više faktora: polazni podaci o emisijama, ponuda sadržaja od strane odabranih izvora na internetu, a zatim i činjenica da taj dodatan sadržaj koji se sakupi treba da se prikaže na TV prijemnicima gledalaca.

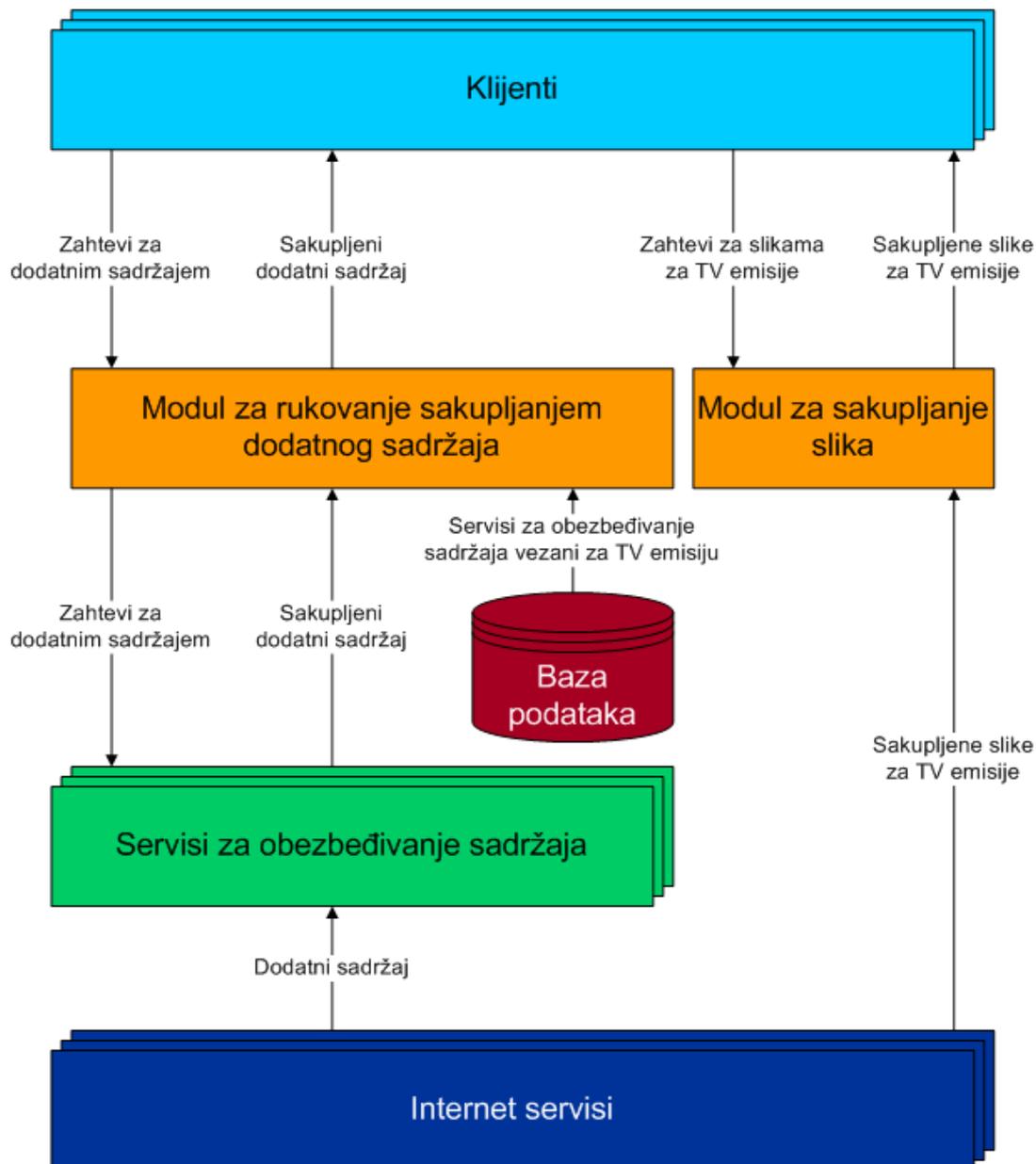
Pored dodatnog sadržaja vezanog za emisije, bilo je potrebno pronaći i slike koje im odgovaraju, radi njihovog prikaza na TV prijemniku i time omogućavanje lakšeg snalaženja korisnika. Veličina ovog problema je neposredno povezana sa kvalitetom polaznih podataka za određene emisije – što polazni podaci bolje opisuju emisiju, to je lakše pronaći sliku te emisije. Međutim, ovaj problem je samo u određenoj meri bilo moguće rešiti programski – ne postoji jednostavan način pomoću kog bi program na osnovu linka ka slici (URL-a) mogao da odredi šta je na toj slici prikazano, odnosno da li se ona odnosi na traženu televizijsku emisiju ili ne.

Nakon sakupljanja sadržaja, potrebno je bilo pronaći način njegovog slanja klijentima. Razmatrana je struktura u kojoj bi se sadržaj isporučivao klijentima, kako bi se sa jedne strane klijentima olakšao prikaz raznorodnih sadržaja na uniforman način, a sa druge strane omogućilo slanje što je veće količine multimedijalnog sadržaja vezanog za određeni pojam, odnosno televizijsku emisiju.

Na kraju, kako je broj različitih vrsta televizijskih emisija poprilično velik, a količina sadržaja na internetu još veća, u realizovanom rešenju nije bilo moguće pokriti sakupljanje sadržaja za sve emisije. Stoga, bilo je potrebno omogućiti da realizovani sistem bude relativno lako proširiv kako bi se u budućnosti mogle ubacivati nove funkcionalnosti, odnosno sakupljati nove vrste sadržaja.

4. Koncept rešenja

U ovom poglavlju je dat kratak pregled arhitekture realizovanog sistema. Dati su opisi glavnih modula i izdvojenih celina u sistemu, uz objašnjenja razloga za njihovo izdvajanje. Na slici 4.1 prikazan je izgled realizovanog sistema. Klijenti, u ovom slučaju, prijemnici digitalnog televizijskog signala (STB-ovi), šalju zahteve za prijem dodatnog sadržaja i slika vezanih za televizijske emisije. Ti zahtevi se upućuju modulu za sakupljanje sadržaja, odnosno modulu za sakupljanje slika. Modul za sakupljanje sadržaja izvlači neophodne informacije iz baze podataka kako bi mogao da prosledi zahteve za sadržajem odgovarajućim servisima za obezbeđivanje sadržaja. Servisi za obezbeđivanje sadržaja i modul za sakupljanje slika koriste internet servise kako bi dobavili odgovarajući sadržaj odnosno slike vezane za televizijske emisije. Nakon sakupljanja, sadržaj i slike se šalju klijentima kao odgovori na zahteve.



Slika 4.1. Arhitektura realizovanog sistema

4.1 Izvori sadržaja na internetu

Izvori sadržaja na internetu mogu biti obične internet stranice, internet baze podataka, pretraživači itd. Na osnovu načina pristupa sadržaju koji nude, mogu se podeliti u dve grupe:

- izvori koji nude aplikativnu programsku spregu za pristup sadržaju
- izvori koji ne nude aplikativnu programsku spregu za pristup sadržaju

Aplikativne programske sprege i sam sadržaj koji nude ovi izvori se veoma razlikuju. Razlikuju se upiti za dobijanjem sadržaja koji se šalju, mogućnosti filtriranja rezultata, formati odgovora, itd. U realizovanom sistemu bilo je potrebno omogućiti pristup sadržajima koji ovi izvori nude. Zbog navedenih razlika, nije bilo moguće realizovati jednu aplikativnu programsku spregu koja bi omogućavala pristup svim izvorima sadržaja, već su napravljeni

posebni moduli za pristup svakom od izvora. Ovi moduli nazvani su internet servisi. Prilikom odabira internet servisa koji će biti realizovani, tj. odabira izvora na internetu kojima će se obezbediti pristup, težilo se da se pokrije sadržaj od značaja za što je veći broj televizijskih emisija. Pri tome, prioritet su imale vrste emisija koje su češće na programu, kao i one za koje je bilo moguće ponuditi neki dodatan sadržaj koji bi bio zanimljiv gledaocima. Shodno tome, neki od internet servisa koji su realizovani su: servisi koji nude linkove ka video snimcima (*YouTube, TrailerAddict*), baze znanja (*Freebase, Wikipedia*), pretraživači (*Google Search Engine*), i dr.

4.2 Servisi za obezbeđivanje sadržaja

Kako bi se rešili problemi određivanja sadržaja od značaja za određene televizijske emisije realizovani su posebni moduli za prikupljanje sadržaja sa interneta. Ovi moduli su podeljeni na osnovu vrste sadržaja koje sakupljaju, pa tako postoje moduli za sakupljanje informacija o glumcima, video snimaka kao najava za filmove, sportskih vesti, vesti iz sveta fudbala, recepata za televizijske emisije o kuvanju, itd. Servisi za obezbeđivanje sadržaja koriste informacije koje nude internet servisi kako bi sakupili što bolji sadržaj. Kao ulazni podaci servisima za obezbeđivanje sadržaja, šalju se EPG informacije o televizijskoj emisiji. Izlazni podaci koje generišu ovi servisi jesu sakupljeni sadržaji.

4.3 Modul za rukovanje sakupljanjem dodatnog sadržaja

Ovaj modul predstavlja tačku preko koje realizovani sistem komunicira sa ostatkom okruženja. To okruženje čine klijenti koji traže dodatni sadržaj i programska podrška za pristup bazi podataka u kojoj se čuvaju EPG podaci o televizijskim emisijama. U te podatke spadaju i informacije o tome koji su servisi za obezbeđivanje sadržaja povezani sa odgovarajućom televizijskom emisijom. Kada klijent zatraži dodatni sadržaj za određenu televizijsku emisiju, on prosleđuje njenu identifikaciju. Na osnovu te identifikacije je potrebno pronaći traženu emisiju u bazi podataka i pronađene EPG podatke proslediti modulu za rukovanje sakupljanjem dodatnog sadržaja. Ovaj modul ima za zadatak da prosledi zahtev klijenta odgovarajućim servisima za obezbeđivanje sadržaja i dostavi im dobijene EPG informacije kao ulazne podatke. Servisi za obezbeđivanje sadržaja kojima će se proslediti zahtevi klijenata se biraju na osnovu toga da li su povezani sa datom televizijskom emisijom, što je zapisano u bazi podataka. Nakon prikupljanja sadržaja, ovaj modul treba da vrati klijentima rezultate prikupljanja, odnosno listu sadržaja koji su vratili svi servisi za obezbeđivanje sadržaja kojima su prosleđeni zahtevi.

4.4 Modul za sakupljanje slika koje odgovaraju televizijskim emisijama

Ovaj modul je izdvojen kao poseban slučaj servisa za obezbeđivanje sadržaja jer, iako mu se kao ulazni podaci šalju EPG podaci o televizijskoj emisiji kao i ostalim servisima, izlazni podaci koje ovaj modul generiše jesu linkovi ka slikama vezanim za tražene televizijske emisije.

5. Programsko rešenje

U okviru ovog poglavlja objašnjeni su detalji realizacije pojedinih modula u sistemu. Navedene su aplikativne programske sprege i primeri funkcionisanja pojedinih modula. Takođe, objašnjeni su načini prevazilaženja određenih problema koji su se pojavili prilikom realizacije.

5.1 Internet servisi

Internet servisi su realizovani kao obične Java klase (eng. *POJO – Plain Old Java Object*). Aplikativne programske sprege koje nude internet servisi se međusobno razlikuju jer se načini pristupa različitim izvorima razlikuju. Kod nekih internet servisa je napravljen odabir sadržaja kome se obezbeđuje pristup, na osnovu njegovog značaja za određene teme, kako bi se izbacivanjem stvari koje nisu od značaja pojednostavila aplikativna programska sprega koju ti internet servisi nude i time olakšalo njihovo korišćenje. Kod drugih internet servisa se omogućuje pristup svim sadržajima koje izvori nude. Informacije koje nude izvori na internetu mogu biti: tekstovi, linkovi ka slikama, linkovi ka audio i video snimcima, linkovi ka internet stranicama itd. Takođe, jedna vrsta informacija su i meta-podaci potrebni za obavljanje pretrage za gore navedenim sadržajem. Na primer, ukoliko je potrebno pronaći dodatan sadržaj vezan za film koji gledalac gleda, jedan od prvih koraka bio bi pronalazak fotografija i biografija glumaca koji glume u tom filmu i njihovo dostavljanje gledaocu. Da bi se prikupile informacije o glumcima, prvo se mora pronaći sam spisak glumaca koji glume u filmu, a tek tada se može nastaviti sa pretragom. Spisak glumaca u ovom slučaju predstavlja spomenute meta-podatke.

Za komunikaciju sa izvorima sadržaja internet servisi koriste HTTP protokol. Kako je broj zahteva koji se šalju ka izvorima sadržaja veliki, a slanje odgovora od strane izvora može da potraje, slanje zahteva se obavlja asinhrono. To znači da je moguće poslati više zahteva za

sadržajima, kako bi oni mogli da se obrađuju od strane izvora, a zatim sačekati odgovore u nekom kasnijem trenutku. Ovo je urađeno kako bi se povećala brzina izvršavanja uvođenjem određenog nivoa paralelnog slanja zahteva i čekanja na njihovu obradu od strane poslužioca.

Kako bi se obezbedila ova funkcionalnost, bilo je potrebno uvesti dodatni modul. Ovaj modul nazvan je HTTP klijent.

5.1.1 HTTP klijent modul

EJB specifikacija zabranjuje, odnosno savetuje protiv stvaranja novih niti u okviru EJB-ova, jer je predviđeno da se o tim stvarima brine i njima rukuje aplikacioni poslužilac, odnosno EJB kontejner. Međutim, kako bi se ostvarila zamisao za slanjem i paralelnim čekanjem na veći broj zahteva za sadržajima, ovo ograničenje je, u određenoj meri, prevaziđeno stvaranjem HTTP klijent modula. Ovaj modul čini nekoliko klasa, od kojih su najbitnije:

- `HttpClient` – predstavlja spregu (eng. *interface*) za korišćenje HTTP klijenta, odnosno nudi metodu za asinhrono slanje HTTP zahteva.
- `HttpClientBean` – ova klasa je implementacija prethodno spomenute sprege. Predstavlja jedan session bean. Asinhrono izvršavanje metode za slanje zahteva se postiže postavljanjem “`@Asynchronous`” anotacije iznad nje.

Metode obeležene “`@Asynchronous`” anotacijom u slučaju da imaju povratnu vrednost vraćaju implementaciju `Future` sprege. Ova sprega omogućava da se nastavi sa izvršavanjem odmah nakon poziva metode, a rezultati asinhrono metode sačekaju u nekom kasnijem trenutku pozivom metode `get()` nad prosleđenim `Future` objektom.

Korišćenjem EJB-ova, na ovaj način, omogućeno je slanje više zahteva i asinhrono čekanje na odgovore, bez narušavanja načela EJB specifikacije, jer je u ovom slučaju aplikacioni poslužilac svestan kreiranih niti i zna kako njima da rukuje. Naime, kada se prozove asinhrona metoda, EJB kontejner odmah vraća kontrolu klijentu, odnosno ne blokira ga, a nastavlja obradu iz prozване metode u posebnoj niti.

Odgovori koji stignu od izvora su najčešće u JSON ili XML formatu. Oni se nakon pristizanja preslikavaju u Java objekte, kako bi se mogli dalje koristiti u radu. Za ovo preslikavanje koriste se `java-json`, `jackson-core-asl` i `jackson-mapper-asl` biblioteke.

5.2 Servisi za obezbeđivanje sadržaja

Servisi za obezbeđivanje sadržaja realizovani su kao EJB-ovi. Svaki servis predstavlja posebnu implementaciju jedne zajedničke sprege. Ta sprega sadrži dve metode:

Prototip metode	ContentQuery prepareQuery (EPGEvent event)
Opis metode	Ova metoda služi za stvaranje upita koji će se poslati kao ulazni argument metodi za samo prikupljanje sadržaja. Upit se formira na osnovu prosleđenih EPG podataka o datoj televizijskoj emisiji. Razlog postojanja ove metode leži u želji da se u jednom momentu omogući formiranje upita ne samo na osnovu EPG podataka, već i nekih drugih, npr. informacija iz teleteksta. Takođe, u okviru ove metode se vrše odabir i priprema podataka za prosleđivanje metodi za dobavljanje sadržaja, čime se ona oslobađa brige o tome. Npr. uklanjaju se brojevi epizoda iz naziva televizijskih emisija, određuje se način formiranja upita na osnovu trajanja emisije i sl.
Parametri metode	event – nadjeni EPG podaci vezani za televizijsku emisiju za koju se traži dodatni sadržaj
Povratna vrednost	Objekat klase ContentQuery koji predstavlja formirani upit. Klasa ContentQuery sadrži polja kao što su: sam upit u tekstualnom obliku, zatim željeni jezik, trajanje, itd. Ova polja su izdvojena kao značajne informacije koje bi mogle poslužiti servisima za obezbeđivanje sadržaja prilikom odabira internet servisa za korišćenje.

Tabela 5.1 Metoda *prepareQuery* servisa za obezbeđivanje sadržaja

Prototip metode	Future<List<ContentItem>> fetchContent (ContentQuery query)
Opis metode	U okviru ove metode se na osnovu prosleđenog upita obavlja sakupljanje dodatnog sadržaja korišćenjem različitih internet servisa. Ova metoda je neblokirajuća.
Parametri metode	query – upit formiran u okviru metode <i>prepareQuery</i>
Povratna vrednost	Lista pronađenih sadržaja. Sadržaj koji se vraća klijentima koji su ga zatražili je strukturiran na određeni način, i to u obliku jedinice sadržaja (eng. <i>content item</i>). Strukturu ove jedinice sadržaja čine: ime sadržaja, tekstualni opis, link ka slici, linkovi ka video, audio snimcima, internet

stranicama, izvori datog sadržaja, itd. Ova struktura je predstavljena klasom `ContentItem`. Jedna jedinica sadržaja može da sadrži neke ili sve od navedenih elemenata. Ovo je urađeno da bi se omogućilo slanje što većeg broja informacija vezanih za određeni pojam koji ta određena jedinica sadržaja predstavlja, kao i da bi se klijentima olakšalo rukovanje primljenim sadržajem.

Tabela 5.2 Metoda `fetchContent` servisa za obezbeđivanje sadržaja

Jedan od primera pronalaženja dodatnog sadržaja je traženje informacija o glumcima koji glume u prosleđenoj seriji ili filmu. Prvi korak je pokušaj pronalaženja spiska glumaca. Za ovo se koristi Freebase baza podataka, odnosno internet servis koji pruža aplikativnu programsku spregu ka ovom izvoru. Ukoliko se pronađe spisak glumaca, dalje se traže njihove slike, biografije i internet stranice. Početna tačka za traženje slika jeste IMDb baza podataka. Da bi se došlo do IMDb stranice traženog glumca, potrebno je poslati upit Google pretraživaču, u formi “ime i prezime glumca + imdb”, i onda na osnovu dobijenih rezultata doći do tražene stranice (ukoliko ona postoji). Ovo se radi jer se URL-ovi stranica glumaca na IMDb-u sastoje od posebnih identifikatora, a ne imena glumaca ili nečeg sličnog čime bi se automatski mogao formirati željeni URL, odnosno lako doći do tražene IMDb stranice. Ukoliko se pronađe željena stranica, sa nje se izvlače informacije poput linka ka slici i biografije osobe. Ukoliko ne postoji slika za datog glumca/glumicu, šalje se upit Google pretraživaču slika, i zatim se kao link ka slici uzima prvi dobijeni rezultat. Na kraju, proverava se da li postoji stranica posvećena datom glumcu/glumici na Vikipediji i ukoliko postoji, taj link se upisuje u polje linka ka internet stranici u okviru objekta klase `ContentItem` koja predstavlja datog glumca/glumicu. Opisani postupak prikazan je na slici 5.1., pri čemu se podrazumeva da se koraci nakon pronalaženja spiska glumaca obavljaju za svakog glumca/glumicu iz spiska.



Slika 5.1 Postupak pronalazjenja dodatnog sadržaja vezanog za glumce koji glume u datoj seriji ili filmu

5.3 Modul za rukovanje sakupljanjem dodatnog sadržaja

Modul za rukovanje sakupljanjem sadržaja je realizovan kao EJB. U okviru ovog modula čuva se lista raspoloživih servisa za obezbeđivanje sadržaja. Sadrži sledeće metode:

Prototip metode	<pre>Future<List<CMoreContent>> fetchContentForEvent(EPGEvent event, String preferredLanguage)</pre>
Opis metode	U okviru ove metode se na osnovu prosleđenih podataka o televizijskoj emisiji iz baze podataka izvlači lista servisa za obezbeđivanje sadržaja vezanih za datu emisiju. Nakon toga, pronađeni servisi se asinhrono prozovu i zatim se sačekaju rezultati sakupljanja sadržaja svakog od njih. Ova metoda je asinhrona (neblokirajuća).
Parametri metode	<p>event – EPG podaci vezani za televizijsku emisiju za koju se traži dodatan sadržaj</p> <p>preferredLanguage – željeni jezik, odnosno jezik koji ima prioritet prilikom traženja sadržaja. Tako, ukoliko je moguće pronaći dodatan sadržaj na željenom jeziku, to će se i uraditi i taj sadržaj će biti</p>

	dostavljen klijentu. Ukoliko nije moguće, dostaviće mu se sadržaj na onom jeziku na kom ga je bilo moguće pronaći (u trenutnoj implementaciji, u pitanju je engleski jezik).
Povratna vrednost	lista sakupljenih sadržaja od svih servisa za obezbeđivanje sadržaja koji su bili prozvani

Tabela 5.3 Metoda *fetchContentForEvent* modula za rukovanje sakupljanjem sadržaja

Prototip metode	List<String> getContentServiceList()
Opis metode	Ova metoda vraća spisak servisa za obezbeđivanje sadržaja koji su trenutno raspoloživi u sistemu.
Povratna vrednost	Lista raspoloživih servisa za obezbeđivanje sadržaja.

Tabela 5.4 Metoda *getContentServiceList* modula za rukovanje sakupljanjem sadržaja

Prototip metode	void registerContentServices()
Opis metode	U okviru ove metode vrši se registracija svih servisa za obezbeđivanje sadržaja, odnosno, zapisivanje svih raspoloživih servisa u strukturu podataka, sa kojom se kasnije rukuje. Ova metoda se proziva pri inicijalizaciji modula za rukovanje sakupljanjem dodatnog sadržaja.

Tabela 5.5 Metoda *registerContentServices* modula za rukovanje sakupljanjem sadržaja

5.4 Modul za sakupljanje slika koje odgovaraju televizijskim emisijama

Modul za sakupljanje slika koje odgovaraju televizijskim emisijama nastoji da sakupi što je bolje slike za prosleđene televizijske emisije. Za sakupljanje slika koriste se Google pretraživač (njegova mogućnost pretraživanja slika, a ne internet stranica) i internet baza podataka vezanih za filmove i serije IMDb (eng. *Internet Movie Database*). Sakupljanje slika se u određenoj meri razlikuje za različite televizijske emisije, i to po njihovim žanrovima. Naime, empirijskim putem zaključeno je da se u slučaju npr. vesti, dobijaju bolji rezultati slika ukoliko se slike traže na osnovu imena televizijske emisije zajedno sa imenom kanala na kom se ta emisija pušta, dok za neke druge žanrove to nije slučaj. U slučaju npr. filmova i

televizijskih serija, prva tačka za traženje odgovarajuće slike jeste IMDb internet stranica. Ukoliko se na njoj ne pronađe slika za traženi film ili seriju, prelazi se na korišćenje Google pretraživača.

6. Ispitivanje i rezultati

Nad realizovanim sistemom sprovedene su dve vrste ispitivanja. Prvo ispitivanje se bavi proverom brzine dobijanja sadržaja i slika vezanih za televizijske emisije. Drugo ispitivanje čini anketa sprovedena nad grupom ispitanika koja za cilj ima da pokaže mišljenje korisnika TV prijemnika o sakupljenom dodatnom sadržaju.

6.1 Ispitivanje brzine sakupljanja sadržaja i slika

U cilju ispitivanja brzine sakupljanja sadržaja napisana je aplikacija u okviru koje se šalju zahtevi za sakupljanjem sadržaja i zahtevi za sakupljanjem slika za televizijske emisije. Zahtevi za sakupljanje sadržaja su bili upućeni modulu za rukovanje sadržajem, a zahtevi za slikama modulu za sakupljanje slika. Prilikom ispitivanja brzina sakupljanja sadržaja, posmatrano je vreme potrebno za dobavljanje sadržaja vezanog za svaku televizijsku emisiju ponaosob, jer to odgovara načinu zahtevanja dodatnog sadržaja od strane korisnika. Sa druge strane, prilikom ispitivanja brzine sakupljanja slika, posmatrano je vreme potrebno za dobavljanje slika za više televizijskih emisija odjednom. Ovo takođe odgovara standardnom načinu korišćenja modula za sakupljanje slika.

6.1.1 Ispitivanje brzine sakupljanja dodatnog sadržaja

Ispitivanje je vršeno nad skupom od 200 televizijskih emisija koje su povezane sa realizovanim servisima za obezbeđivanje sadržaja. U pitanju su servisi koji pokrivaju sledeće žanrove: filmovi i serije (eng. *Movie/Drama*), vesti (eng. *News*) i sport (eng. *Sports*). Vršene su tri iteracije slanja zahteva za sadržajem. Prosečna vremena čekanja na odgovore, odnosno pristizanje rezultata od modula za sakupljanje sadržaja prikazana su u tabeli 6.1.

Žanr televizijske emisije	Prosečno vreme čekanja na odgovor [ms]
Filmovi/serije	1467.3
Vesti	123.1
Sport	1787.7

Tabela 6.1 Prosečna vremena čekanja na pristizanje rezultata od modula za sakupljanje sadržaja za određene žanrove televizijskih emisija

Vremena sakupljanja sadržaja zavise od više faktora. U njih spadaju: količina sadržaja koji se sakuplja, broj internet servisa koji se koriste prilikom sakupljanja, kao i broj skokova sa jednog na drugi servis. Međutim, glavni uzrok za smanjenje brzine sakupljanja sadržaja jeste vreme za koje poslužilac određenog internet servisa odgovori na poslat zahtev. Ovo vreme zavisi od stanja u mreži, iznenadnih prekida veze, opterećenosti poslužioca i drugih faktora na koje realizovani sistem ne može uticati. Može se primetiti da je prosečno vreme čekanja na dodatni sadržaj manje od dve sekunde. Iako se može smatrati da je ta vrednost prihvatljiva, povećanje brzine sakupljanja sadržaja je ipak neophodno. Ono se može postići zapisom sakupljenog sadržaja u *cache* memoriju, radi bržeg pristupa.

6.1.2 Ispitivanje brzine sakupljanja slika

Zahtevi za sakupljanjem slika za televizijske emisije su slani paralelno, u grupama od po 50, 100 i 200. Nakon poslate grupe zahteva, čekalo se na njihove odgovore, a zatim se isti proces ponavljao sa sledećom grupom. Sprovedeno je više iteracija opisanog ispitivanja. Prosečna vremena sakupljanja slika za različite veličine grupa emisija prikazana su u tabeli 6.2.

Broj paralelnih zahteva za sakupljanjem slika	Prosečno vreme sakupljanja slika [ms]
50	3532.8
100	4468.6
200	5535.4

Tabela 6.2 Prosečna vremena trajanja sakupljanja slika za televizijske emisije za određene veličine grupa paralelnih zahteva

U realizovanom sistemu, sakupljanje slika se obavlja slanjem grupa paralelnih zahteva. Čekanje na rezultate poslanih zahteva se takođe vrši u grupama. Odnosno, prvo se paralelno pošalje jedna grupa zahteva, pa se zatim sačekaju odgovori na njih, a nakon njihovog pristizanja se prelazi na slanje sledeće grupe zahteva itd. Prilikom čekanja na odgovore na

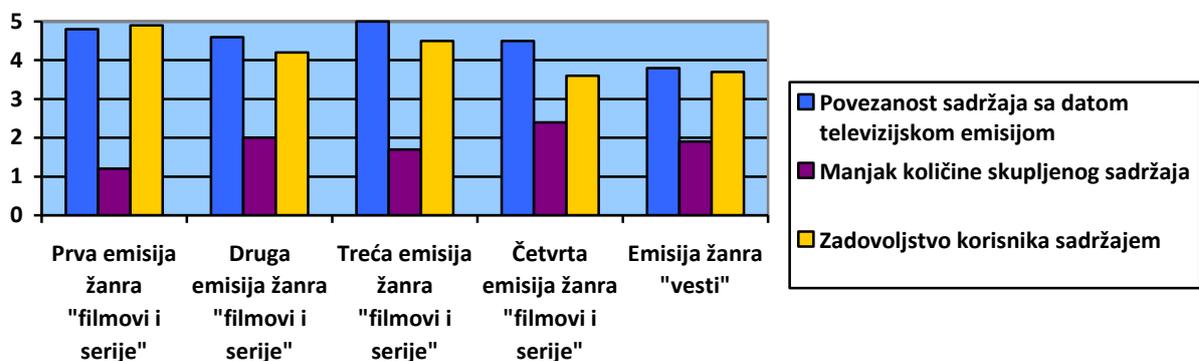
poslate zahteve, vreme čekanja zavisi od najsporije obrade zahteva. Ono zavisi od udaljenog poslužioca koji obrađuje zahteve. Iz rezultata sprovedenog ispitivanja može se primetiti da vremena čekanja na odgovore u slučaju različitih veličina grupa ne rastu linearno sa povećanjem broja zahteva u grupi. Stoga se može zaključiti da se povećanjem broja paralelno poslanih zahteva skraćuje ukupno vreme čekanja na odgovore, odnosno povećava se brzina, jer je u slučaju veće grupe zahteva potrebno poslati manji broj grupa da bi se pronašle slike za televizijske emisije. Ono što ograničava broj zahteva u okviru jedne grupe jeste broj niti kojima aplikacioni poslužilac istovremeno može rukovati. Taj broj se može promeniti u konfiguraciji poslužioca, međutim, pri tome se mora voditi računa o korišćenju resursa.

6.2 Ispitivanje mišljenja korisnika o sakupljenom sadržaju

Prilikom ispitivanja mišljenja korisnika o sakupljenom dodatnom sadržaju, posmatrane su sledeće osobine:

- povezanost sadržaja sa datom televizijskom emisijom
- količina skupljenog sadržaja
- zadovoljstvo korisnika sadržajem

Posmatran je sadržaj vezan za četiri emisije žanra „filmovi i serije“ i jedne emisije žanra „vesti“. Anketa je sprovedena nad deset ispitanika i njen primer se nalazi u prilogu. Ispitanici su ocenjivali tačnost navedenih tvrdnji. Rezultati ankete su prikazani na slici 6.1.



Slika 6.1 Prikaz rezultata sprovedene ankete koja ispituje mišljenje korisnika o sakupljenom dodatnom sadržaju

Sa slike se može videti da su se dobile relativno visoke ocene za povezanost sadržaja sa datom televizijskom emisijom, što pokazuje da trenutna realizacija sistema obezbeđuje sadržaj većinom usko povezan za date televizijske emisije. Prosečna ocena prilikom ispitivanja ove osobine je 4.54 od mogućih 5, te se može zaključiti da postoji mesta za napredak. Taj napredak se može postići poboljšanjem algoritama za pretragu, pronalaskom

dodatnih izvora sadržaja na internetu i boljim iskorišćenjem polaznih podataka o televizijskim emisijama. Što se tiče količine sakupljenog sadržaja, prosečna ocena iznosi 1.84, pri čemu više ocene označavaju nezadovoljstvo manjkom sakupljenog sadržaja, dok niže ocene označavaju zadovoljstvo korisnika količinom sadržaja. Na osnovu dobijene ocene, može se zaključiti da sakupljena količina dodatnog sadržaja nije mala, ali da se može još povećati. Prosečna ocena zadovoljstva korisnika sakupljenim sadržajem iznosi 4.18. Iz toga se može zaključiti da je potrebno još raditi na poboljšanju ovog aspekta rada sistema.

Prilikom tumačenja rezultata sprovedene ankete treba imati u vidu relativno mali broj ispitanika, ispitivanih osobina i različitih televizijskih emisija. Iako bi se povećanjem ovog broja dobili realniji rezultati, dobijeni rezultati se ipak mogu iskoristiti kao okvirni pokazatelj mišljenja korisnika. Takođe, mogu se primetiti razlike u ocenama sadržaja sakupljenog za emisije različitih žanrova. Time se mogu odrediti pravci daljeg razvoja sistema, odnosno odlučiti se za delove sistema na čijem će se poboljšanju raditi u budućnosti.

7. Zaključak

U ovom radu predstavljena je realizacija programske podrške u *cloud* okruženju za prikupljanje multimedijalnog sadržaja i meta-podataka sa interneta. Polazna tačka za prikupljanje sadržaja jesu podaci o televizijskoj emisiji koji su prosledjeni od strane korisničkog uređaja (prijemnika digitalnog TV signala).

S obzirom na veliki broj različitih vrsta televizijskih emisija i multimedijalnih sadržaja koji postoje, trenutna implementacija programske podrške ne pokriva sve slučajeve, odnosno, nije obezbeđeno sakupljanje sadržaja za sve televizijske emisije.

Međutim, prilikom realizacije sistema vodilo se računa da postoji mogućnost relativno lakog proširenja postojećih funkcionalnosti. Sistem ima modularnu arhitekturu, gde najbitnije module čine internet servisi i servisi za obezbeđivanje sadržaja. Ovi moduli se mogu lako dodavati u sistem, čime se, u slučaju internet servisa, omogućava pristup novim izvorima sadržaja na internetu, a u slučaju servisa za obezbeđivanje sadržaja, omogućava sakupljanje nove vrste dodatnog sadržaja.

U daljem radu je planirano da se radi na:

- automatskom prepoznavanju raspoloživih servisa za obezbeđivanje sadržaja u sistemu
- sakupljanju sadržaja za više vrsta televizijskih emisija dodavanjem novih servisa za obezbeđivanje sadržaja
- pronalaženju novih izvora sadržaja na internetu
- poboljšanju postojećih servisa za obezbeđivanje sadržaja
- poboljšanju načina sakupljanja slika vezanih za televizijske emisije
- povećanju brzine odgovora klijentima zapisivanjem sakupljenog sadržaja u *cache* memoriju

8. Literatura

- [1] ETSI TS 102 796: Hybrid Broadcast Broadband TV; V1.1.1.
- [2] P. Podhradsky, "Evolution trends in hybrid broadcast broadband TV", E-LMAR, 2013 55th International Symposium.
- [3] WWW konzorcijum – W3C, <http://www.w3.org/>, učitano 27.06.2014.
- [4] XML specifikacija, <http://www.w3.org/TR/2008/REC-xml-20081126/>, učitano 25.06.2014.
- [5] JEE 6, <http://docs.oracle.com/javase/6/tutorial/doc/>, učitano 27.06.2014.
- [6] ETSI TS 102 809: Digital Video Broadcasting (DVB); Signalling and carriage of interactive applications and services in hybrid broadcast / broadband environments; V 1.1.1.

Prilog

Primer sprovedene ankete:

Na pitanja odgovarajte ocenama od 1 do 5, gde:

- 1 – slažem se
- 2 – delimično se slažem
- 3 – ne znam
- 4 – delimično se ne slažem
- 5 – ne slažem se

1. Sakupljeni sadržaj sa interneta je vezan za datu televizijsku emisiju.
2. Sakupljena je prevelika količina dodatnog sadržaja.
3. Sakupljeni sadržaj je nešto što bih voleo/volela da vidim kao dodatni sadržaj vezan za datu emisiju.