

## PRETRAŽIVANJE ONTOLOŠKI HETEROGENIH TEKSTUALNIH RESURSA NA WEB-U QUERYING ONTOLOGICALLY HETEROGENEOUS TEXTUAL RESOURCES ON THE WEB

Milica Vučković, Siniša Nešković

**REZIME:** U radu se daje jedno rešenje problema distribuiranog pretraživanja ontološki heterogenih tekstualnih resursa na Web-u, zasnovanom na "peer-to-peer" (P2P) arhitekturi. Predloženo rešenje se bazira na definisanom apstraktnom modelu i na njemu zasnovanom apstraktnom upitnom jeziku AQL (Abstract Query Language). Različite lokalne ontologije koje različite autonomne grupe koriste za semantičke opise svojih dokumenata uniformno se reprezentuju preko osnovnih koncepata apstraktnog modela – tipova dokumenata i njihovih semantičkih osobina koje ih definišu. Predloženo rešenje problema pretraživanja ontološki heterogenih tekstualnih resursa na Web-u primenjeno je u novoj distribuiranoj verziji opšteg sistema za administrativno poslovanje ADMIS koji se razvija u Laboratoriji za informacione sisteme FON-a. Kao implementaciono okruženje za realizaciju ADMIS sistema koristi se Microsoft .NET implementaciono okruženje koje pruža veoma dobru podršku za XML i Web zasnovane aplikacije.

**KLJUČNE REČI:** Apstraktni upitni jezik (AQL), Apstraktni model, P2P arhitektura, Semantička interoperabilnost i integracija, Distribuirano pretraživanje resursa na Web-u.

**ABSTRACT:** This paper presents a solution for searching ontologically heterogeneous distributed textual resources on the Web, based on "peer-to-peer" (P2P) architecture. Suggested solution is founded on the Abstract Query Language (AQL), which is based on the Abstract Model. The concepts of the Abstract Model are capable to incorporate and unify different local ontologies that describe different kinds of specific local documents. An ontological ADMIS- Super peer architecture supports the new distributed version of ADMIS, a generalized administrative process management system, which is under development at the Laboratory for Information Systems, Faculty of Organizational Sciences, is presented too. The new version of ADMIS incorporates searching ontologically heterogeneous distributed textual resources based on AQL. The system is implemented in Microsoft .NET environment that provides very good support for XML and Web based applications.

**KEY WORDS:** Abstract Query Language (AQL), Abstract Model, P2P architecture, Semantic interoperability and integration, Searching distributed resources on the Web..

### 1. UVOD

Informacioni sistemi savremenih poslovnih organizacija i državnih institucija sadrže veliki broj heterogenih i distribuiranih izvora podataka. Mnogi od ovih izvora podataka projektovani su i razvijani nezavisno jedan od drugih i to na način koji najviše odgovara lokalnim potrebama njihovih korisnika. Iz takvih distribuiranih i autonomnih izvora proizilazi i njihova heterogenost koja se može razmatrati sa različitih aspekata: tehničkog ili sistemskog (različite hardverske i operativne platforme, različiti komunikacioni protokoli), sintaksnog (različiti reprezentacioni jezici i formati izvora podataka), strukturnog (heterogenost modela i šema) i semantičkog (različite interpretacije podataka).

Sa razvojem Interneta, distribuiranih softverskih infrastruktura i posebno sa pojavom Web-a zapravo su stvoreni uslovi za fizičko povezivanje ogromnog broja distribuiranih i heterogenih izvora podataka. Međutim, mogućnost dinamičkog povezivanja ogromnog broja heterogenih, autonomnih i dinamičkih izvora podataka daje dodatnu kompleksnost problemu pretraživanja i integracije podatka. Pod procesom pretraživanja ne podrazumeva se samo transparentni pristup informacijama, već i njihova integracija, tj. ekstraktovanje i kombinovanje informacija koje dolaze iz heterogenih

i distribuiranih izvora podataka. Problem integracije podataka nije lak za rešavanje, jer integracija podataka iz različitih izvora, zahteva prevazilaženje heterogenosti na različitim nivoima složenosti.

Jedan od najsloženijih problema za rešavanje u kontekstu integracije podataka, a koji proizilazi iz pristupa heterogenim izvorima podataka, je semantička heterogenost. Pod semantičkom heterogenošću podrazumevaju se razlike u semantici (značenju) podataka koje zavise od korišćene terminologije za izražavanje realnih koncepata i konteksta (domena) u kojem se interpretiraju. Razumevanje semantike podataka je neophodan uslov za otkrivanje, ekstraktovanje i kombinovanje relevantnih podataka u procesu integracije.

U dinamičkom, heterogenom i distribuiranom sistemu kao što je to Web, semantička integracija može se ostvariti samo korišćenjem apstrakcija pomoću kojih se skriva kompleksnost i implementacioni detalji različitih vrsta izvora podataka. Za ostvarivanje semantičke interoperabilnosti očigledno je da su potrebni bolji modeli i alati za opis semantičkih sadržaja izvora podataka, tehnike za automatsko ekstraktovanje i klasifikaciju (na primer, ontologije) podataka i meta podataka, skalabilne i fleksibilne interoperabilne arhitekture i tehnologije i novi semantički upitni jezici za pretraživanje distribuiranih i heterogenih izvora podataka na uniforman način.

U ovom radu prezentuje se jedno rešenje problema distribuiranog pretraživanja ontološki heterogenih tekstualnih resursa na Web-u. Pod tekstualnim resursom podrazumeva se dokument bilo koje vrste (formata) proširen sa strukturnim i semantičkim opisom njegovog sadržaja.

Rad je organizovan na sledeći način: Pregled postojećih rešenja problema distribuiranog pretraživanja daje se u drugom poglavlju ovog rada. U trećem poglavlju opisuje apstraktni model i apstraktni upitni jezik (Abstract Query Language) zasnovan na njemu. U četvrtom delu se diskutuju ontološke transformacije AQL-a, kao i transformacije iz AQL u pojedine konkretne upitne jezike. U petom poglavlju opisane su glavne komponente predložene peer-to-peer arhitekture i realizacija AQL upita. Na kraju se daje zaključak i literatura.

## 2. PREGLED POSTOJEĆIH REŠENJA

Većina postojećih informacionih resursa na Web-u standardno se prikazuju kao HTML stranice, a postojeće tehnike koje se najčešće koriste za njihovo pretraživanje zasnivaju se na leksičkim pretraživačima. Međutim, resursi se obično pamte na neki od načina (relacione baze, Web formati XML i RDF) koji im omogućava očuvanje njihove strukture i semantike nezavisno od njihove prezentacije, tj. HTML formata. Jedan od važnijih aspekata u razvoju Web-a i semantičkog Web-a je projektovanje i implementacija upitnih jezika koji su uglavnom pogodni za pretraživanje XML, RDF ili OWL resursa. Za pretraživanje XML resursa koriste se složeni XML upitni jezici sa strukturnim karakteristikama (kao što su XPath [18] i XQuery [12]). Primarni cilj semantičkih tehnika za pretraživanje heterogenih resursa na Web-u je nalaženje dokumenata koji su relevantni za krajnje korisnike. Postoji određeni broj preporuka za pretraživanje RDF resursa koje uključuju ontološke upitne jezike RQL, SquishQL, TRIPLE, OWL-QL, RDQL. I pored toga što su razvijeni novi semantički upitni jezici, jedan od ključnih problema u razvoju tehnika za semantički Web je uniformno pretraživanje i integracija semantički heterogenih izvora podataka.

Poslednjih godina problem pretraživanja i integracije semantički heterogenih resursa razmatran je u različitim distribuiranim okruženjima, uključujući integraciju podataka, semantički Web, i peer-to-peer i grid paradigme, servisno-orijentisane sisteme. Najnoviji pristupi za rešavanje problema distribuiranog pretraživanja semantički heterogenih resursa na Web-u su: pristupi koji se baziraju na globalnoj ontologiji i pristupi zasnovani na peer-to-peer paradigmi.

Većina tekućih ontoloških pristupa rešenje problema distribuiranog pretraživanja i integracije semantički heterogenih izvora podataka zasnivaju na ideji da se smisljena komunikacija između različitih autonomno razvijenih izvora podataka može realizovati korišćenjem

globalne ontologije kao modela za reprezentovanje zajedničke semantike određenog domena. Reformulacija korisničkog upita, formulisanog u terminologiji globalne ontologije, na upite nad lokalnim izvorima podataka zasniva se na definisanim semantičkim preslikavanjima između globalne ontologije i šema lokalnih izvora podataka. Za specifikaciju preslikavanja koriste se poznati LAV i GAV formalizmi [10]. Za reprezentovanje globalne ontologije koriste se klasični ontološki jezici zasnovani na deskriptivnoj logici (CARIN [5], LOOM [11]) ili novi web-ontološki reprezentacioni jezici kao što su to RDF/RDFS [9], DAML+OIL [8], OWL [3]).

U otvorenom i decentralizovanom Web okruženju veoma je teško definisati globalnu ontologiju za određeni domen. Čak i u specifičnim domenima kao što su medicina, bankarstvo i turizam, usaglašavanje semantičke heterogenosti podataka pomoću jedne ontologije nije jednostavno. Autonomne zajednice (pojedinci, grupe ili organizacije) obično koriste sopstvene lokalne ontologije, tj. skup termina i veza između njih, za konceptualizaciju određenog domena. Lokalne konceptualizacije su parcijalne (tj. obuhvataju samo deo domena) i aproksimativne (tj. koriste različite nivoe granularnosti za opis istog domena) reprezentacije određenog domena i odražavaju različite poglede autonomnih zajednica na posmatrani domen. Konkretizacije jedne lokalne ontologije pojavljuju se u različitim oblicima, kao što su to na primer različite lokalne šeme (relaciona, RDF ili XML šema), lokalni sistemi za upravljanje dokumentima ili struktura direktorijuma sistema datoteka.

U drugoj vrsti pristupa koja se zasniva na distribuiranoj peer-to-peer paradigmi ne pretpostavlja se egzistencija jedne globalne ontologije, već se rešenje problema distribuiranog pretraživanja i integracije podataka iz različitih izvora koji koriste autonomno razvijene lokalne šeme za organizovanje svojih podataka, zasniva na semantičkim preslikavanjima između lokalnih ontologija. U takvom scenariju, upit se može posmatrati kao zahtev, formulisan nad lokalnom ontologijom, za pronalaženjem semantički povezanih informacija iz udaljenih ontologija.

Dva osnovna problema koja su predmet najživljih tekućih istraživanja i koja se vezuju za semantička pretraživanja u distribuiranom i heterogenom okruženju zasnovanom na P2P paradigmi, su: (i) otkrivanje i specifikacija preslikavanja između heterogenih lokalnih ontologija i (ii) korišćenje postojećih semantičkih preslikavanja za reformulaciju upita, formulisanim nad datom lokalnom ontologijom, na jedan ili više upita nad udaljenim ontologijama.

Određena rešenja za problem distribuiranog pretraživanja ontološki heterogenih Web resursa data su u projektima Edutella [4, 13] i Piazza [6, 7] u kojima se kao P2P infrastruktura koristi nova klasa semantičkih P2P mreža, tzv. P2P mreže zasnovane na meta šemama, i u kojima se kombinuju tehnologije semantičkog Web-a, P2P sistema i baza podataka.

### 3. APSTRAKтни MODEL I APSTRAKтни UPITNI JEZIK

Rešenje problema distribuiranog pretraživanja ontološki heterogenih tekstualnih resursa na Web-u bazira se na apstraktnom modelu i na njemu zasnovanom apstraktnom upitnom jeziku AQL (Abstract Query Language).

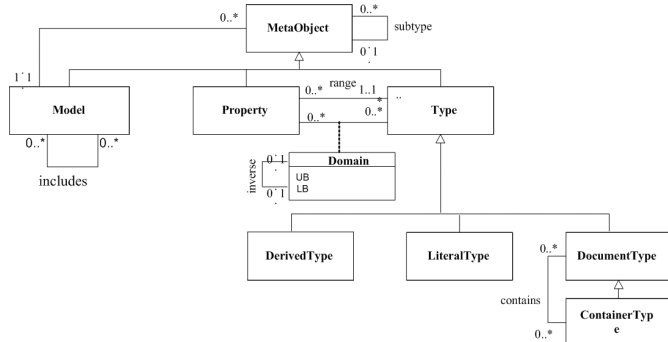
#### 3.1. Apstraktni model

Za uniformno reprezentovanje različitih lokalnih konceptualizacija koristi se apstraktni model, definisan u [16]. Apstraktni model preko svojih osnovnih koncepata – tipova i njihovih semantičkih osobina koje ih definišu, na uniforman način opisuje različite lokalne ontologije koje različite autonomne grupe koriste za semantičke opise svojih dokumenata. Na ovaj način eliminišu se sintaksne i strukturne heterogenosti različitih konkretnih modela podataka u kojima se smeštaju lokalne kolekcije dokumenata.

Na Slici 1. su prikazani fundamentalni koncepti apstraktnog modela u UML notaciji i to sa aspekta semantičkog opisa dokumenata.

Jedan od osnovnih koncepata u ovom modelu je koncept tipa dokumenta (*DocumentType*) koji predstavlja klasu dokumenata koji imaju ista semantička svojstva. Takva semantička svojstva, po kojima se dokumenti datog tipa mogu opisivati, odnosno po kojima se kasnije mogu pretraživati, reprezentovana su u modelu pomoću klase *Property*. U cilju podrške generalne notacije resursa koji imaju ista semantička svojstva uvedena je klasa *Type*, koja je modelovana kao generalizacija klasa *DocumentType* i *LiteralType*. Klasa *Type* koristi se kao domen (*domain*) i kao kodomen (*range*) za vrednosti osobina (*property*). Domen specificira klasu na čije vrednosti se može primeniti određena osobina. Modelom je podržana mogućnost specifikacije više domena za jednu posmatranu osobinu. U zavisnosti od toga da li je kodomen posmatrane osobine literal ili klasa, osobina se može razmatrati kao prost atribut ili binarna asocijacija.

Binarna asocijacija definiše u stvari dva preslikavanja koja su međusobno inverzna (asocijacija **inverse** na klasi **Domain**). Jedna od bitnih karakteristika binarne asocijacije je i kardinalnost njenih preslikavanja. Za svako pres-



Slika 1. – Osnovni koncepti apstraktnog modela

likavanje i tip kao domen u takvom preslikavanju, moguće je specificirati kardinalnost preslikavanja preko atributa **UB** i **LB** (na klasi **Domain**).

Pošto svi definisani tipovi i osobine pripadaju nekom ontološkom modelu, uvedena je klasa *Model*. Model je ontološka definicija koja definiše ontološku, ravnu strukturu preko pripadajućih tipova i njihovih osobina. Klasa *Meta Object* generalizuje sve opisane koncepte, uključujući i sam *Model*. Asocijacija *Subtype* uvedena je da bi se omogućile specijalizacije koncepata modela. Svaki model može da uključuje više drugih modela (asocijacija *includes* na klasi *Model*), tako da ontološke definicije drugih modela postaju dostupne specificiranom modelu. S druge strane, model može da bude uključen u više drugih modela.

#### 3.2. Apstraktni upitni jezik

Na osnovu definisanog apstraktnog modela definisan je apstraktni upitni jezik AQL. AQL je upitni jezik sa bogatim semantičkim konceptima i namenjen je za distribuirano pretraživanje semantički heterogenih resursa, na uniforman način, jedinstvenim jezikom.

Model apstraktnog upitnog jezika AQL [17] definisan je nad osnovnim konceptima datog apstraktnog modela – tipovima dokumenata i njihovim semantičkim osobinama koje ih definišu. Pošto se semantičke osobine koriste se za opis sadržaja dokumenata određenog tipa, onda je moguće takva dokumenta pretraživati preko AQL upita na osnovu njihovog semantičkog opisa.

AQL upiti, koji se definišu nad apstraktnim modelom, kao i rezultati dobijeni njihovom evaluacijom, reprezentuju se preko AQL modela. Svaki AQL upit specificira se skupom promenljivih, uslovom pretraživanja i rezultatom koji se želi dobiti. AQL upiti formiraju se na osnovu više vrsta izraza, koji se mogu kombinovati u složenije izraze. Za konstruisanje složenih upita, AQL koristi **select-from-where** izraz. AQL upit se u stvari sastoji iz tri elementa: Deklaracije promenljivih u AQL upitu, Selekcije promenljivih upita i Projekcije rezultata upita.

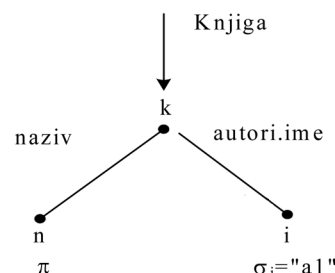
Promenljive u AQL upitu deklariraju se u **from** klauzuli. U izrazima za deklaraciju specificiraju se i ontološke putanje pomoću kojih se promenljive upita povezuju. U AQL upitu, promenljive i ontološke putanje koje ih povezuju formiraju stablo upita. U korenu stabla AQL upita je tzv. korena ili slobodna promenljiva. U ostalim čvorovima stabla nalaze se tzv. vezane promenljive koje imaju ulogu vezivanja promenljivih sa odgovarajućim ontološkim izrazima putanja. Ne koristi se operacija eksplicitnog spajanja, već se spajanje specificuje definisanim promenljivih upita.

**Where** klauzula je opciona i specificira logička ograničenja na vrednosti promenljivih definisanih u **from** klauzuli. Logički uslov sastoji se iz jednog ili više prostih predikata koji se povezuju u složenije pomoću logičkih operacija AND ili OR.

Type
Knjiga
Autor

Property	Domain	Range
naslov	Knjiga	string
autori	Knjiga	Autor
ime	Autor	string
knjige	Autor	Knjiga

**Q<sub>1</sub>:** **select** n  
**from** k **in** Knjiga, n **in**  
 k.naslov, i **in** k.autori.ime  
**where** contains(i, "a1")



Slika 2. – Ontologija Knjiga, AQL upit Q<sub>1</sub> i stablo AQL upita

**Select** klauzula definiše projekciju nad promenljivama koje su od interesa za rezultat upita. Na ovaj način specificira se koje promenljive i u kom redosledu će biti prikazane u rezultatu upita. Rezultat AQL upita je skup n-torki oblika  $\{[c_i, c_j, \dots, c_m]\}$ , gde su  $c_i, c_j, \dots, c_m$  instance promenljivih u **select** klauzuli AQL upita. Podrazumeva se da se u rezultat upita, za svaku n-torku koja predstavlja vrednosti atributa (promenljivih) o dokumentu koji zadovoljava uslov upita, uključuje i referenca (URL-Uniform Resource Locator) na pun sadržaj takvog dokumenta.

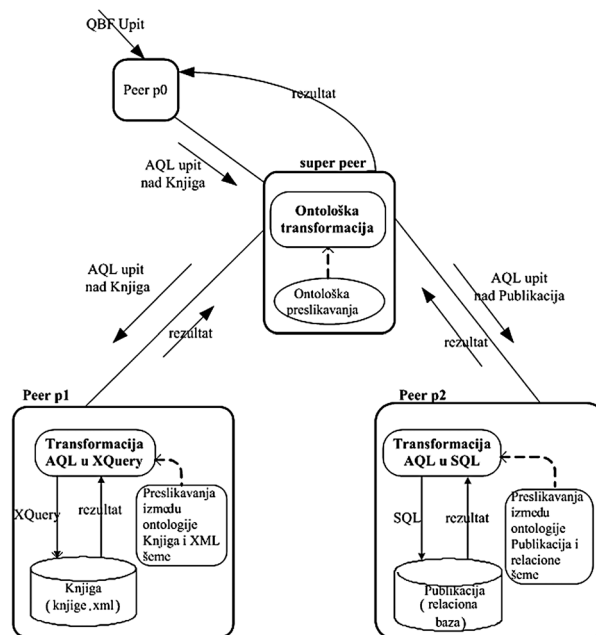
Na slici 2. dat je primer semantičkog opisa (preko osnovnih koncepata apstraktnog modela) ontologije *Knjiga*, primer AQL upita Q<sub>1</sub> koji je definisan nad opisanom ontologijom i u kojem je postavljen zahtev za nalaženje naslova svih knjiga koje je objavio autor «a1»

Notacija “.” (tzv. DOT notacija) se koristi za povezivanje delova izraza putanje i označava sintaksnu skraćenicu za operaciju implicitnog spajanja (*implicit join*) levog i desnog dela izraza putanje.

AQL predstavlja originalni upitni jezik za koji su definisane dve vrste sintakse. Konvencionalna sintaksa AQL upitnog jezika, veoma je slična SQL-u i OQL-u čime se značajno olakšava rad mnogim korisnicima koji su navikli na ovakav tip upitnih jezika. Takođe je definisana i XML šema AQL-a što implicira da se AQL upit može reprezentovati kao XML dokument u skladu sa definisanom XML šemom. Prednosti koje proizilaze iz predstavljanja modela AQL-a preko XML šeme su višestruke. XML je standardni, platformski nezavisan jezik i omogućava laku distribuciju AQL upita, kao XML dokumenta, preko Web-a. Definisana XML šema AQL-a pojednostavljuje ontološke transformacije AQL-a, kao i prevođenje AQL u konkretne upitne jezike.

**4. TRANSFORMACIJE AQL UPITA**

AQL je upitni jezik namenjen za distribuirano pretraživanje ontološki heterogenih resursa u P2P okruženju. U takvom scenariju, AQL upit, koji korisnik formuliše u terminologiji svoje lokalne ontologije, razmatra se kao zahtev za nalaženje semantički povezanih dokumenata iz udaljenih ontologija i interpretaciju dobijenih rezultata u lokalnoj ontologiji nad kojom je upit postavljen. Međutim, odgovor na postavljeni AQL upit moguće je dobiti samo pretraživanjem stvarnih dokumenata koja su smeštena u konkretnim modelima podataka (relacionom, objektnom, XML ili RDF) na različitim čvorovima u P2P mreži. Za ovakvu vrstu pretraživanja moraju se koristiti konkretni upitni jezici



Slika 3. – Dve transformacije AQL upita

(SQL, XQuery, RQL) koji su definisani nad odgovarajućim konkretnim modelima podataka. Zbog toga se za evaluaciju postavljenog AQL upita nad apstraktnim modelom zahtevaju dve transformacije:

- 1) Ontološka transformacija AQL upita, definisanom nad apstraktnim modelom, iz jedne ontologije u drugu
- 2) Transformacija AQL upita u konkretni upitni jezik nad konkretnim modelom podataka

Primer na Slici 3. na uprošćen način ilustruje neophodnost obe vrste transformacija za evaluaciju distribuiranog AQL upita u P2P okruženju.

Postupak za ontološku transformaciju AQL upita iz jedne ontologije u drugu, zasniva se na specifičnim semantičkim preslikavanjima između semantički ekvivalentnih koncepata različitih lokalnih ontologija [17]. Semantička preslikavanja reprezentovana su pomoću ontoloških izraza putanja.

Transformacija AQL upita nad apstraktnim modelom u upit nad konkretnim modelom je moguća, ako je poznata transformacija iz apstraktnog modela u konkretni. Pravila preslikavanja između apstraktnog modela i **XML šeme**, odnosno apstraktnog modela i **relacione šeme** i kao i postupak za transformaciju AQL upita u konkretni upitni jezik SQL, odnosno XQuery dati su u [17].

Predstavljanje modela AQL upitnog jezika preko XML šeme pojednostavljuje ontološke transformacije AQL upita, kao i prevođenje AQL upita u konkretne upitne jezike, jer se svodi na standardni rad sa XML-om. Kao primer za realizaciju prevođenja AQL-a u konkretni upitni jezik, u dat je implementacija transformacije AQL upita u SQL upit, zasnovana na jednoj definisanoj XSLT transformaciji nad XML šemom AQL-a [17].

## 5. ARHITEKTURA I REALIZACIJA DISTRIBUIRANOG PRETRAŽIVANJA

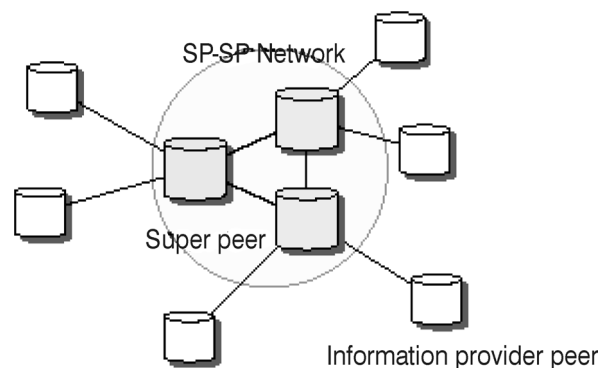
Peer-to-peer paradigma predstavlja distribuirani sistem kao dinamičku mrežu čvorova (peer-ova). P2P ("peer-to-peer") tehnologije [2], danas privlače enormno veliku pažnju, jer su veoma podesne za ad hoc kolaboracije i podelu resursa u distribuiranom okruženju. Kao posledica njihove fleksibilnosti, skalabilnosti i adaptivnosti P2P rešenja postala su veoma pogodna za prevazilaženje mnogih poznatih nedostataka iz klijent/server arhitekture kao i za dinamičke promene u Internet okruženju. P2P tehnologija definiše se kao mreža autonomnih čvorova (peer-ova) koji direktno razmenjuju podatke i servise sa drugim čvorovima. P2P tehnologije uvode interesantnu paradigmu decentralizacije zajedno sa samo-organizovanjem autonomnih peer-ova. Jedna od najčešće isticanih karakteristika P2P tehnologija je postojanje simetrične komunikacije između peer-ova: svaki čvor u mreži ima ulogu i klijenta i servera.

U novije vreme pojavila se nova klasa P2P mreža, tzv. P2P mreže zasnovane na meta šemama, koje kombinuju tehnologije semantičkog Web-a, P2P sistema i baza podataka. Čvorovi u mreži eksplicitno koriste meta šeme za opis svojih sadržaja. P2P mreže zasnovane na meta šemama apstrahuju tehničku infrastrukturu i topologiju i omogućavaju efikasno korišćenje integracionih tehnologija i strategija za rutiranje kontekstno-specifičnih upita.

Na osnovu detaljne analize novijih P2P arhitektura zasnovanim na meta šemama definisana je specifična ADMIS-SuperPeer arhitektura kao distribuirano P2P okruženje za pretraživanje semantički heterogenih resursa. Predložena arhitektura u potpunosti podržava evaluaciju distribuiranih AQL upita.

### 5.1. ADMIS-SuperPeer arhitektura

Super peer arhitektura kombinuje elemente hibridnih i potpuno decentralizovanih P2P sistema, tako da se kombinuje efikasnost centralizovanog pretraživanja sa autonomijom i robustnošću distribuiranog pretraživanja. ADMIS-SuperPeer arhitektura sastoji se iz dve vrste čvorova: super peer čvorova i običnih čvorova (peer-ova).



Slika 4. – Super peer arhitektura

Super peer je čvor koji se ponaša kao centralizovani server za manji skup čvorova-klijenata, koji su zajedno sa super peer-om organizovani u specifične domenske klustere. Super peer je odgovoran za ontološke transformacije AQL upita, rutiranja AQL upita do selektovanih čvorova u mreži kao i za procesiranje dobijenih rezultata od njih.

Svi obični ADMIS čvorovi imaju istu funkcionalnost i fizičku arhitekturu. Obični ADMIS čvorovi smeštaju i opisuju svoje lokalne kolekcije dokumenata korišćenjem autonomno razvijenih lokalnih šema (relacionih, objektnih, RDF ili XML šema). Lokalne ontologije (šeme) koje pojedini čvorovi koriste za semantičke opise sadržaja svojih dokumenata, uniformno su opisane preko osnovnih koncepata (tipova i njihovih semantičkih osobina) apstraktnog modela. Na ovaj način korisniku je omogućeno da postavlja distribuirane AQL upite u terminologiji svoje lokalne ontologije (tj. lokalne ontologije

datog čvora na koga je priključen), a da se pri tom pretražuje po svim dostupnim resursima u mreži bez obzira na ontologiju, šemu ili model koji su korišćeni za njihovo smeštanje i semantički opis.

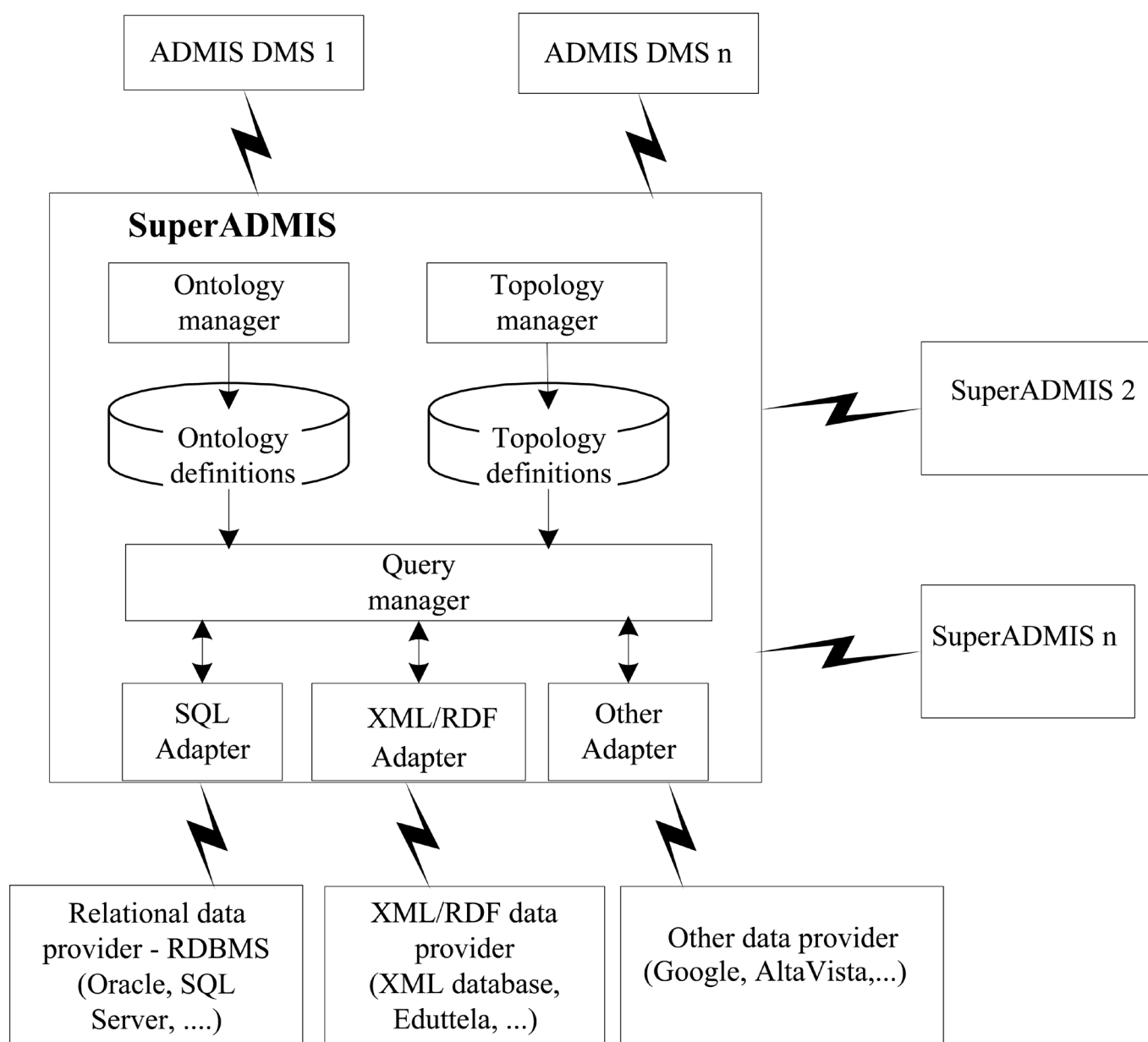
Transformacija AQL upita u konkretni upit nad lokalnim modelom ADMIS čvora, izvršava se na specifičanom ADMIS čvoru na osnovu definisanih preslikavanja između apstraktnog modela i lokalne šeme, koju čvor koristi za organizovanje svojih dokumenata. Odgovori koji se dobijaju kao rezultati izvršavanja konkretnih upita nad lokalnim kolekcijama dokumenata ADMIS čvorova, prosleđuju se do super peer-a koji vrši njihovo uniranje.

Na Slici 5. prikazane su glavne komponente arhitekture ADMIS super peer-a. Osnovne komponente ove arhitekture su «Ontology manager» i «Topology

Manager» koji rade nad ontološkom i topološkom metabazom, respektivno, kao i niz adaptera koji omogućavaju izvršenje AQL upita nad različitim modelima podataka.

### 5.2. Realizacija distribuiranog pretraživanja

Predloženo rešenje problema pretraživanja ontološki heterogenih tekstualnih resursa na Web-u primenjeno je u distribuiranoj verziji opšteg sistema za administrativno poslovanje ADMIS [14, 15], koji se razvija u Laboratoriji za informacione sisteme FON-a. Kao implementaciono okruženje za realizaciju ADMIS sistema koristi se Microsoft .NET implementaciono okruženje koje pruža veoma dobru podršku za XML i Web zasnovane aplikacije. Ovo okruženje sadrži koncept



Slika 5. – Arhitektura Super ADMIS čvora

Dataset-a, kao centralnu komponentu za XML. Dataset je u osnovi XML šema. AQL model je opisan Dataset-om, a u okviru komponente koja manipuliše Dataset-om implementirana je funkcionalnost pretraživanja. Nad ovom komponentom definisan je i izložen web servis za pretraživanje, koji kao svoj ulazni argument primi AQL u formi XML dokumenta, a kao povratnu vrednost servis vraća rezultat upita, opet kao XML dokument. Definisani servis obezbeđuje sintaksu i semantiku za transparentni standardni interfejs za distribuirano pretraživanje heterogenih resursa smeštenim na čvorovima u ADMIS-SuperPeer arhitekture.

Sakrivanje detalja formiranja upita i njegovog predstavljanja kao XML dokumenta realizovana je preko grafičkog korisničkog interfejsa, koji omogućava korisnicima da u «Query by Form» načinu, postavljaju AQL upite u svojim lokalnim ontologijama. Na ovaj način korisnicima je omogućeno da na transparentan način, (bez poznavanja apstraktnog modela i sintakse i semantike AQL upitnog jezika) formulišu AQL upite u svojim lokalnim ontologijama, a dobijene rezultate distribuiranog pretraživanja prikazu u terminologiji lokalne ontologije u kojoj je AQL upit i postavljen.

## 6. ZAKLJUČAK

U radu je prezentirano jedno rešenje problema distribuiranog pretraživanja ontološki heterogenih tekstualnih resursa na Web-u, koje je primenjeno u distribuiranoj verziji opšteg sistema za administrativno poslovanje ADMIS koji se razvija u Laboratoriji za informacione sisteme FON-a. Razvijeni softverski sistem se bazira na izloženoj super peer arhitekturi u kome je, poštujući autonomiju svakog čvora u mreži, omogućeno postavljanje upita korišćenjem lokalnih, korisniku dobro poznatih ontologija, dok se pretraživanje vrši po svim dostupnim resursima bez obzira na ontologiju, šemu ili model koji su korišćeni za njihov semantički opis i memorisanje.

Prikazana je specifikacija i implementacija apstraktnog upitnog jezika AQL (Abstract Query Language) za pretraživanje dokumenata u semantički i platformski heterogenom distribuiranom okruženju. AQL je razvijen u Laboratoriji za informacione sisteme u okviru ADMIS-a (Sistem za upravljanje administrativnim poslovima), koji sadrži dve osnovne komponente: Sistem za upravljanje poslovnim procesima i Sistem za upravljanje dokumentima. AQL značajno unapređuje funkcije Sistema za upravljanje dokumentima jer omogućuje da se semantički i platformski različite baze dokumenata pretražuju zajednički, na uniforman način, jedinstvenim jezikom. Korišćenjem istih principa koji su ovde izloženi, AQL je uspešno primenjen i na pretraživanje drugih vrsta resursa (baza poslova, baza poruka) smeštenim u čvorovima distribuirane P2P mreže.

## 7. LITERATURA

- [1] Brickley, D., Guha, R.V., *Resource Description Framework (RDF) Schema Specification 1.0.*, W3C Candidate Recommendation, 2000.
- [2] Crowcroft, J., Pratt, I., *Peer to peer: Peering into the future*, in E. Gregori, G. Anastasi, and S. Basagni, editors, *Advanced Lectures on Networking*, NETWORKING 2002 Tutorials, volume 2497 of *Lecture Notes in Computer Science*, Springer-Verlag, May 2002.
- [3] Dean, M., Connolly, D., van Harmelen, F., et al., *OWL Web Ontology Language 1.0 Reference*, W3C Working Draft 29 July 2002. <http://www.w3c.org/TR/owl-ref/>
- [4] *The Edutella Project*, 2002. <http://edutella.jxta.org/>
- [5] Goasdoue, F., Lattes, V., Rousset, M.C., *The use of CARIN language and algorithms for Information Integration: the PICSEL project*, *International Journal on Cooperative Information Systems* 9 (4), 2000.
- [6] Halevy, A., Ives, Z., Suciu, D., Tatarinov, I., *Schema mediation in peer data management systems*, *Proceedings of the 19th International Conference on Data Engineering*, Bangalore, India, March 2003.
- [7] Halevy, A., Ives, I., Mork, P., Tatarinov, I., *Piazza: Data Management Infrastructure for Semantic Web Applications*, in *Proceedings of the Twelfth International World Wide Web Conference*, 2003.
- [8] Horrocks, I., *DAML+OIL: a Description Logic for the Semantic Web*, *IEEE Data Engineering Bulletin*, 25(1):4-9, 2002.
- [9] Lassila, O., Swick, R., *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendation, 1999.
- [10] Lenzerini, M., *Data Integration : A Theoretical Perspective*, in *Proc. of the ACM Symposium on Principles of Database Systems (PODS)*, pages 233-246, Madison, Winsconsin, USA, June 2002.
- [11] MacGregor, R., *Using a description classifier to enhance deductive inference*, *Proceedings of the Seventh IEEE Conference on AI Application* (pp. 93-97), Florida.
- [12] Marchiori, M., *XML Query (XQuery)*, World Wide Web Consortium, September 23, 2003. <http://www.w3.org/XML/Query>
- [13] Nejd, W., Wolpers, M., Siberski, W., Schmitz, C., Schlosser, M., Brunkhorst, I., Löser, A., *Super-Peer-Based Routing and Clustering Strategies for RDF-Based Peer-To-Peer Networks*, in *Proceedings of the Twelfth International World Wide Web Conference*, 2003.
- [14] Nešković, S., Lazarević, B., Soldar, G., *ADMIS: Generalized Administrative System*, in *Proceedings of the IASTED Conference*, Innsbruck, Austria, 2003.
- [15] Nešković, S., Vučković, M., Babarogić, S., Radetić, A., *Programski sistem i softverske komponente za formiranje i pretraživanje baza dokumenata*, *Zbornik radova INFOFEST 99*, Budva, 1999.
- [16] Nešković, S., Vučković, M., *Sistem za pretraživanje ontološki heterogenih tekstualnih resursa na Web-u zasnovan na 'peer-to-peer' arhitekturi*, *Zbornik radova INFOFEST 03*, Budva, 2003. (nagrađen kao najbolji autorski rad)
- [17] Vučković, M., *Upitni jezik za pretraživanje strukturiranog teksta*, *Doktorska disertacija*, FON, Beograd, 2005.
- [18] *XML Path Language (XPath) 2.0*, W3C Working Draft (20 December 2001), <http://www.w3.org/TR/xpath20>



Milica Vučković, FON, Laboratorija za IS  
Oblast interesovanja: projektovanje IS



Siniša Nešković, FON, Laboratorija za IS  
Oblast interesovanja: projektovanje IS