

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERIŲ KATEDRA

Aurelija Kazlauskienė

**SEMANTINIŲ METADUOMENŲ IŠGAVIMO IR
APRAŠYMO METODIKOS TYRIMAS**

Magistro darbas

Vadovas
doc. dr. V. Kiauleikis

KAUNAS, 2006

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
KOMPIUTERIŲ KATEDRA

**SEMANTINIŲ METADUOMENŲ IŠGAVIMO IR
APRAŠYMO METODIKOS TYRIMAS**

Magistro darbas

Vadovas
doc. dr. V.Kiauleikis

Recenzentas
doc. dr. D. Rubliauskas

Atliko
IFM 0/4 gr. stud.
A. Kazlauskienė

KAUNAS, 2006

Turinys

Summary	5
1. Įvadas	8
2. Analizė	9
2.1. Analizės tikslas	9
2.2. Tyrimo sritis, objektas ir problema	9
2.3. Aplinkos (organizacijos veiklos, jos dalies, funkcijos, proceso ir pan.) analizė	11
2.4. Vartotojų analizė	11
2.5. Problemos sprendimo metodų literatūros šaltiniuose analizė	11
2.5.1. Metaduomenų aprašymo metodikų analizė	12
2.5.2. Semantikos kategorijų analizė	17
2.5.3. Semantinių asociacijų analizė	19
2.5.4. Darbo su semantiniais metaduomenimis etapai	20
2.6. Panašių sistemų (Lietuvos ir tarptautiniu mastu) analizė	21
2.6.1. Semagix FREEDOM sistemos apžvalga	22
2.7. Architektūros ir galimų įgyvendinimo priemonių variantų analizė	25
2.8. Darbo tikslas ir siejami privalumai	26
2.9. Kompiuterizuojamos sistemos funkcijos	27
2.10. Nefunkciniai reikalavimai ir apribojimai	28
2.11. Rezultato kokybės kriterijai	29
2.12. ISO 9126 standartas sistemos kokybei įvertinti	31
2.13. Analizės išvados	32
3. Sistemos reikalavimų specifikacija	33
4. Projektas	45
4.1. Projekto tikslas	45
4.2. Panaudojimo atvejų realizacijos ir analizės klasių diagramos	46
4.2.1. Sistemos vartotojų paslaugų trasų diagrama (vartotojo interfeiso navigavimo planas)	46
4.2.2. Veiklos paslaugų klasių trasų diagrama	47
4.2.3. Duomenų paslaugų klasių diagrama	47
4.2.4. Dalykinės srities esybių klasių diagrama	47
4.2.5. /Dirbti su žodynu/ realizacija ir analizės klasės	49
4.2.6. /Dirbti su failais/ realizacija ir analizės klasės	53
4.2.7. /Ištraukti duomenis/ realizacija ir analizės klasės	57
4.2.8. /Palyginti/ realizacija ir analizės klasės	60
4.2.9. /Rodyti metaduomenis/ realizacija ir analizės klasės	61
4.3. Duomenų bazės schema ir lentelių specifikacijos	65
4.4. Realizacijos modelis (programinių komponentų architektūra, įdiegimo modelis)	66
5. Eksperimentinis sistemos tyrimas	67
5.1. Eksperimentinio diegimo aprašymas	67
5.2. Sistemos naudojimo instrukcija	68
5.3. Sistemos veikimo ir savybių analizė	71
5.4. Kokybės kriterijų vertinimas pagal ISO/IEC 9126 kokybės modelį	73
5.4.1. Funkcionalumas	73
5.4.2. Patikimumas	73
5.4.3. Panaudojamumas	74
5.4.4. Efektyvumas	74
5.4.5. Pernešamumas	74

5.4.6.	Priderinamumas.....	74
6.	Išvados.....	75
7.	Literatūra	77

Summary

The main problem in the semantic technology today is semantic metadata extraction and description. There is no general method used to extract any type of semantics. The main objective of this research is to analyze existing methods that are used for describing and extracting semantic metadata and to design a metadata extraction system for Lithuanian.

The quality analysis of designed system is carried out according to ISO/IEC 9126 standard. The objective of this standard is to provide a framework for the evaluation of software quality. ISO/IEC 9126 does not provide requirements for software, but it defines a quality model, which is applicable to every kind of software.

Paveikslų sąrašas

1 pav.	Metaduomenų tipai	9
2 pav.	Organizacijos veiklos tikslų diagrama.....	11
3 pav.	Šaltinio modelis.....	13
4 pav.	DCMI aprašymo modelis.....	14
5 pav.	Duomenų modelis pagal RDF.....	16
6 pav.	RDF sintaksė.....	16
7 pav.	RDF ir DC terminologijos sutapatinimas.....	17
8 pav.	Darbo su semantiniais duomenimis etapai	20
9 pav.	Esybių vaizdavimas RDF.....	21
10 pav.	Semantinių metaduomenų apdorojimo technologijos.....	22
11 pav.	Semagix FREEDOM komponentai	23
12 pav.	Architektūros pavyzdys	25
13 pav.	Semantinės asociacijos.....	25
14 pav.	Detalesnė sistemos architektūra.....	26
15 pav.	Projektuojamos sistemos RDF schema.....	26
16 pav.	Nedetalizuota sistemos panaudojimo atvejų diagrama	27
17 pav.	Metaduomenų kokybės kriterijai įvairiose duomenų gyvavimo ciklo fazėse	28
18 pav.	Metaduomenų kokybės kriterijai vartotojo – sistemos kūrėjo atžvilgiu.....	29
19 pav.	Papildyta panaudojimo atvejų diagrama	33
20 pav.	Rasti sinonimą.....	34
21 pav.	Ištraukti sinonimą	35
22 pav.	Pašalinti sinonimą	36
23 pav.	Surasti failą.....	38
24 pav.	Pridėti failą.....	39
25 pav.	Pašalinti failą.....	40
26 pav.	Rodyti metaduomenis	41
27 pav.	Palyginti	42
28 pav.	Metaduomenų ištraukimas.....	43
29 pav.	RDF schema	45
30 pav.	Vartotojų paslaugų trasų diagrama	46
31 pav.	Veiklos paslaugų klasių trasų diagrama	47
32 pav.	Duomenų paslaugų klasė	47
33 pav.	Dalykinės srities klasių trasų diagrama	47
34 pav.	Patikslinta dalykinės srities klasių trasų diagrama	48
35 pav.	Dalykinės srities klasių diagrama	48
36 pav.	Analizės klasių diagrama	49
37 pav.	Trasų diagrama PA Dirbti su žodynu	49
38 pav.	PA Ištraukti sinonimą realizacija	50
39 pav.	Sekų diagrama PA Ištraukti sinonimą.....	50
40 pav.	PA Pašalinti sinonimą realizacija	51
41 pav.	Sekų diagrama PA Pašalinti sinonimą.....	51
42 pav.	PA Surasti sinonimą realizacija	52
43 pav.	Sekų diagrama PA Surasti sinonimą	52
44 pav.	Analizės klasių diagrama	53
45 pav.	Trasų diagrama PA Dirbti su failais	53
46 pav.	PA Pridėti failą realizacija	54
47 pav.	Sekų diagrama PA Pridėti failą.....	54
48 pav.	PA Pašalinti failą realizacija	55
49 pav.	Sekų diagrama PA Pašalinti failą	55
50 pav.	PA Surasti failą realizacija.....	56
51 pav.	Sekų diagrama PA Surasti failą	56
52 pav.	Analizės klasių diagrama	57
53 pav.	Trasų diagrama Ištraukti duomenis	57
54 pav.	PA Ištraukti duomenis realizacija	58
55 pav.	Sekų diagrama PA Ištraukti duomenis	59
56 pav.	Analizės klasių diagrama	60
57 pav.	Trasų diagrama PA Palyginti	60

58 pav.	PA Palyginti realizacija.....	60
59 pav.	Sekų diagrama PA Palyginti	61
60 pav.	Analizės klasių diagrama	61
61 pav.	Trasų diagrama PA Rodyti metaduomenis.....	61
62 pav.	PA Rodyti metaduomenis realizacija	62
63 pav.	Sekų diagrama PA Rodyti metaduomenis	62
64 pav.	Detali vartotojų paslaugų klasių diagrama	63
65 pav.	Veiklos paslaugų klasių diagrama	63
66 pav.	Detali duomenų paslaugų klasių diagrama	64
67 pav.	Detali sistemos veiklos paslaugų klasių diagrama	64
68 pav.	Duomenų bazės schema	65
69 pav.	Lentelė Objektai	65
70 pav.	Lentelė Sinonimai	65
71 pav.	Lentelė MetaData	65
72 pav.	Lentelė Grupės	66
73 pav.	Komponentų diagrama	66
74 pav.	Įdiegimo diagrama	66
75 pav.	Failų pasirinkimo langas	68
76 pav.	Metaduomenų išgavimo langas	69
77 pav.	Objekto pavadinimo įvedimo langas	69
78 pav.	Duomenų bazės langas.....	69
79 pav.	Sinonimų peržiūros langas.....	70
80 pav.	Sinonimo įvedimo langas.....	70
81 pav.	Metaduomenų radimo tyrimo algoritmas	71
82 pav.	Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų žodžių kiekio	72

Lentelių sąrašas

1 lentelė.	Metaduomenų modeliai	12
2 lentelė.	DCMI elementai	12
3 lentelė.	Semantikos galimybių priklausomybė nuo semantikos tipo	17
4 lentelė.	Semagix Freedom žinių šaltiniai	23
5 lentelė.	Metaduomenų kokybės matavimo kriterijai	28
6 lentelė.	Duomenų reikšmių kokybės kriterijaus atributai	29
7 lentelė.	Duomenų vaizdavimo kokybės kriterijaus atributai	30
8 lentelė.	Duomenų modelio kokybės kriterijaus atributai	30
9 lentelė.	Duomenų architektūros kokybės kriterijaus atributai	31
10 lentelė.	PA Rasti sinonimą specifikacija.....	34
11 lentelė.	PA Įtraukti sinonimą specifikacija	35
12 lentelė.	PA Pašalinti sinonimą specifikacija	37
13 lentelė.	PA Surasti failą specifikacija	38
14 lentelė.	PA Pridėti failą specifikacija.....	39
15 lentelė.	PA Pašalinti failą specifikacija.....	40
16 lentelė.	PA Rodyti metaduomenis specifikacija	41
17 lentelė.	PA Palyginti specifikacija	42
18 lentelė.	PA Ištraukti duomenis specifikacija.....	44
19 lentelė.	Sistemos realizacijos priemonės.....	67
20 lentelė.	Tyrimui naudotos aparatūrinės ir programinės įrangos sąrašas.....	67
21 lentelė.	Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų žodžių kiekio	72
22 lentelė.	Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų dokumentų kiekio	72

1. Įvadas

Daug šiuolaikinių organizacijų susiduria su viena ir ta pačia problema – duomenų valdymas. Kompanijos, naudojančios geras duomenų administravimo ir metaduomenų valdymo strategijas, eliminuoja dažną kitų kompanijų situaciją – „duomenų daug, bet žinių mažai“.

Sintaksinis ir struktūrinis duomenų integravimas jau yra pakankamai išnarinėtas, taigi šiame darbe bus nagrinėjamos semantinių duomenų išgavimo, aprašymo ir panaudojimo technologijos bei priemonės.

Duomenų kodavimo metodikoms įvertinti pasirinktos, ištirtos ir palygintos RDF ir Dublin Core metaduomenų kodavimo technologijos, atlikta semantikos kategorijų bei semantinių asociacijų analizė.

Daugiausiai darbų semantinių metaduomenų išgavimo ir apdorojimo srityje yra nuveikusi Semagix korporacija. Siekiant rasti geriausią sprendimą kuriamos programos architektūrai ir veikimo principams darbe bus išanalizuotas Semagix FREEDOM sistemos veikimas.

Tikimasi darbe sukurti semantikos metaduomenų išgavimo sistemą ir šių metaduomenų panaudojimo metodiką duomenų semantinei analizei.

Sukurtos programų sistemos įvertinimui pasirinktas ISO/IEC 9126 kokybės modelis. Šis modelis nekelia reikalavimų programinei įrangai, tačiau gali būti taikomas įvertinti bet kokios programinės įrangos kokybei. ISO/IEC 9126 standartas pateikia kokybės vertinimo charakteristikas, tačiau nenurodo kokiomis priemonėmis turi būti atliktas vertinimas.

Šiame darbe taip pat atliktas eksperimentinis efektyvumo tyrimas. Modifikuotas sistemos kodas taip, kad galima būtų išmatuoti laikines metaduomenų paieškos trukmės charakteristikas.

2. Analizė

2.1. Analizės tikslas

Analizės tikslas yra apibrėžti duomenų semantikos aprašymo metaduomenimis problemą, parodyti jos aiškinimą dabartinėje mokslinėje literatūroje ir panagrinėti sprendimus, kuriuos siūlo šiais klausimais dirbančios kompanijos.

2.2. Tyrimo sritis, objektas ir problema

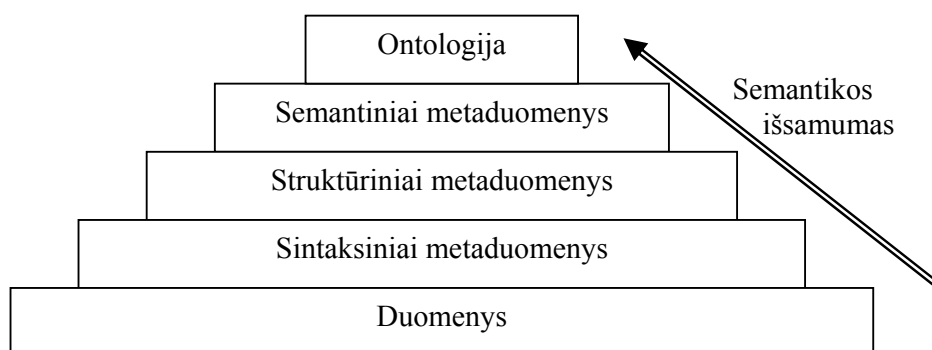
Didžiausias su duomenų integravimu siejamas darbas yra jų sintaksinis ir semantinis suderinimas su bet kuria aplikacija. Duomenų integravimo uždavinys susideda iš dviejų pagrindinių komponentų: duomenų transportavimo ir duomenų suderinamumo užtikrinimo. Dėl didelių investicijų į duomenų perdavimo infrastruktūrą, transportavimo problema yra minimizuota ir šiuo metu yra tik paprastas inžinerinis uždavinys.

Šiandien kalbant apie duomenų suderinamumą pagrinde koncentruojamasi į klausimus, susijusius su sintakse, taigi semantikos problema tarsi yra ignoruojama. Įprasta manyti, kad visi duomenys būtinai turi turėti reikšmes, tačiau nereikėtų pamiršti, kad iš pradžių reikia apibrėžti tų reikšmių kitimo sritį – išanalizuoti duomenis apie pačius duomenis.

Duomenų integravimo sritis apibrėžia sintaksinį, struktūrinį ir semantinį duomenų integravimą. Sudėtingiausia ir mažiausiai tyrinėta sritis yra integravimas semantiniu aspektu.

Darbai sprendžiant duomenų integravimo klausimą (semantikos integravimas, semantiniai duomenų tipai, semantikos suteikimas duomenų elementui) yra laikomi labai brangiais, tačiau be kurių apseiti jau negalima.

Tam, kad geriau suprastume problemos esmę, pabandydysime patyrinėti metaduomenų tipus.



1 pav. Metaduomenų tipai

Pirmame paveiksle pavaizduoti duomenų tipai ir jų išsidėstymas didėjant semantikos išsamumui [1]. Ontologijos nusako terminų, naudojamų tam tikrame tekste, reikšmę arba reikšmių sritį. Jų naudojimas yra laikomas pačiu efektyviausiu metodu, skirtu darbui su semantiniais metaduomenimis.

Ontologijos pavyzdžiai: Anatomija, Diagnostika ir t.t.

Semantiniai duomenys: iš ontologijos išplaukiantys metaduomenys pvz.: Miestas – Klaipėda, Organas – Kepenys, Patologija ir t.t.

Struktūriniai metaduomenys pvz.: dokumento struktūra ir t.t.

Sintaksiniai metaduomenys pvz.: kalba, formatas, dokumento ilgis, sukūrimo data, kodavimo būdas ir t.t.

Duomenys – struktūrizuoti, pusiau struktūrizuoti arba nestruktūrizuoti.

Pagrindinė problema yra ta, kad dar nėra bendro standarto semantiniams duomenims aprašyti, be to šių metaduomenų išgavimas iš teksto yra ligi šiol neišspręstas uždavinys.

Taigi, ateityje bus kaupiami ne tik tradiciniai metaduomenys, bet ir apibrėžtos tų duomenų reikšmės ir vaidmenys organizacijos veikloje. Naujas – semantinis požiūris į organizacijų duomenis leis kurti „protingesnius“ verslo procesus ir aplikacijas.

Pagrindiniai semantinės technologijos komponentai [2]:

1. Klasifikacija
2. Metaduomenys
3. Ontologijos

Klasifikacija – tai pateiktų duomenų klasifikavimas pagal turinį.

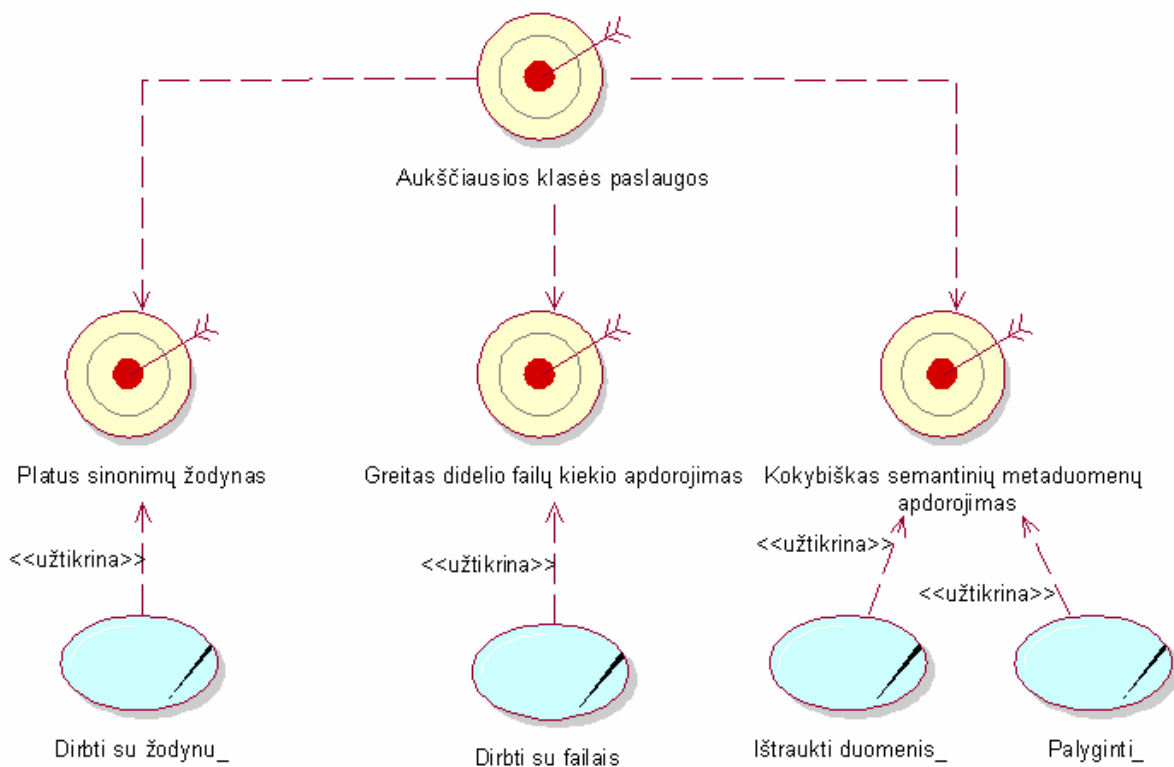
Klasifikacija būtina, jei turime didelius kiekius duomenų. Jau yra sukurta daug klasifikavimo metodų: statistinė analizė ir atitikimo šablonams tikrinimas (Joachims, 1998), taisyklėmis paremti metodai (Ipeirotis, 2000), lingvistinė analizė (Losee, 1995), tikėtinumo metodai naudojantys Bayenso teoriją (Cheeseman ir Stutz., 1996) ir mašininio mokymo metodai (Sebastiani, 2002) tame tarpe ir paremti Paslėptais Markovo Modeliais¹ (Frasconi et al., 2002). Be to, ontologijų naudojimu paremti metodai gali žymiai pagerinti klasifikavimo rezultatus [2].

Semantiniai metaduomenys – tai duomenys tiesiogiai arba netiesiogiai susieti su duota turinio dalimi. Semantiniai metaduomenys suteikia neįkainojamą pagalbą klasifikavimo metodams.

Ontologija apibrėžiama, kaip žinojimo sritis. Ji nusako terminų, naudojamų tam tikrame tekste, reikšmę arba reikšmių sritį. Ontologijų sudarymo srityje dirba tokios organizacijos kaip IBM, Semagix ir t.t.

¹ Paslėpti Markovo Modeliai – Hidden Markov Models

2.3. Aplinkos (organizacijos veiklos, jos dalies, funkcijos, proceso ir pan.) analizė



2 pav. Organizacijos veiklos tikslų diagrama

Pagrindinis organizacijos veiklos tikslas – suteikti aukščiausios klasės semantinių metaduomenų išgavimo/apdorojimo paslaugas. Siekiama kiek galima daugiau praplėsti sinonimų žodyną, kad padidinti metaduomenų išgavimo efektyvumą, taip pat minimizuoti informacijos apdorojimo laiką. Palyginimų sistema turėtų surasti net ir tolimiausią ryšį tarp pateiktų duomenų.

2.4. Vartotojų analizė

Su kuriama metodą naudojančia sistema dirbs įvairių sistemų administratoriai ir paprasti vartotojai.

2.5. Problemos sprendimo metodų literatūros šaltiniuose analizė

2.5.1. Metaduomenų aprašymo metodikų analizė

Bendrai priimtose metodikos semantinių duomenų aprašymui dar nėra, todėl norėdami rasti tinkamiausią, panagrinėsime kelias esamas metaduomenų aprašymo metodikas. Pirmoje lentelėje pateikta keletas metaduomenų modelių ir juos atitinkančių kodavimo pavyzdžių.

1 lentelė

Metaduomenų modeliai	
Metaduomenų modelis	Kodavimo pavyzdžiai
Dublin Core	RDF/XML, HTML 4.0, DCSV (Dublin Core Structured Value)/XML, ISO 19118/XML, ...
ISO 19115	ISO 19118/XML, GML 3.0, ...
OpenGIS	GML 1.0 – 3.0, legacy DDLs, CORBA IDL, ...
ISO TC211	ISO 19118/XML
Unidata VDC	VDC/XML
ESML	ESML

Detaliau panagrinėsime kelis iš išvardintų metaduomenų modelių, patyrinėsime jų panaudojimo aspektus.

Dublin Core metaduomenų standartas

Dublin Core metaduomenų standarto elementai pateikiami 2 lentelėje [10].

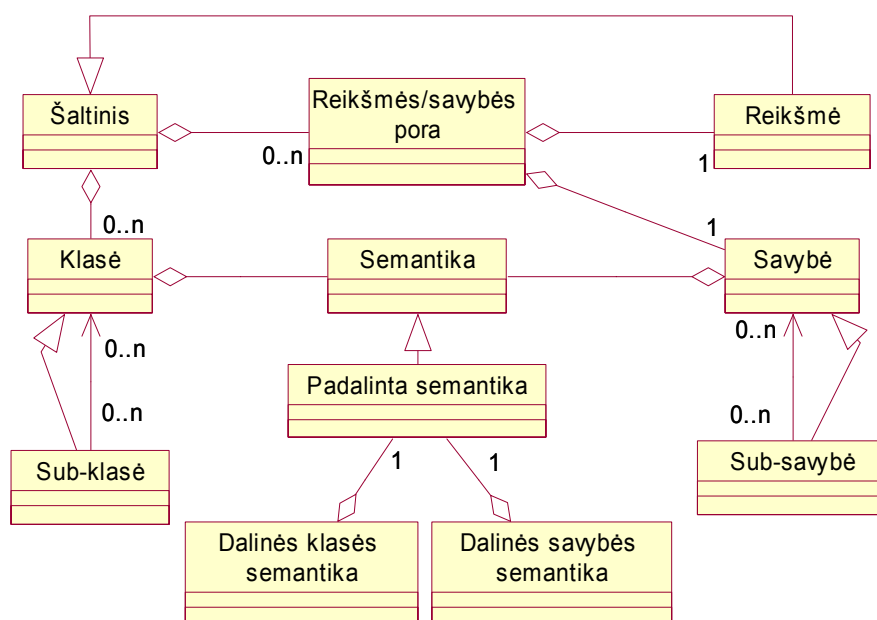
2 lentelė

DCMI elementai	
Elementas	Aprašymas
Pavadinimas	Vardas, priskirtas šaltiniui ²
Pagalbininkas (Contributor)	Esybė, atsakinga už šaltinio turinio pakitimus
Kūrėjas	Esybė, atsakinga už šaltinio turinio sudarymą
Paskelbėjas (Publisher)	Esybė, atsakinga už šaltinio paskelbimą
Tema ir raktiniai žodžiai	Šaltinio turinio tema
Aprašymas	Šaltinio turinio aprašymas
Data	Data susijusi su tam tikru šaltinio gyvavimo ciklo etapu
Šaltinio tipas	Šaltinio turinio žanras, pobūdis
Formatas	Fizinis ar skaitmeninis šaltinio paskelbimas (manifestation)
Šaltinio identifikatorius	Nedviprasmiškas ryšys su šaltiniu duotame kontekste

² Čia ir kitur, jei neparašyta kitaip šaltinis yra angliško žodžio „resource“ vertinys

Kalba	Šaltinio intelektualinio turinio kalba
Ryšys	Ryšys su susijusiu šaltiniu
Šaltinis (Source)	Ryšys su šaltiniu iš kurio gaunamas dabartinis šaltinis
Apimtis (Coverage)	Šaltinio turinio apimti ir tikslas
Teisių valdymas	Informacija apie teises dokumento keitimui

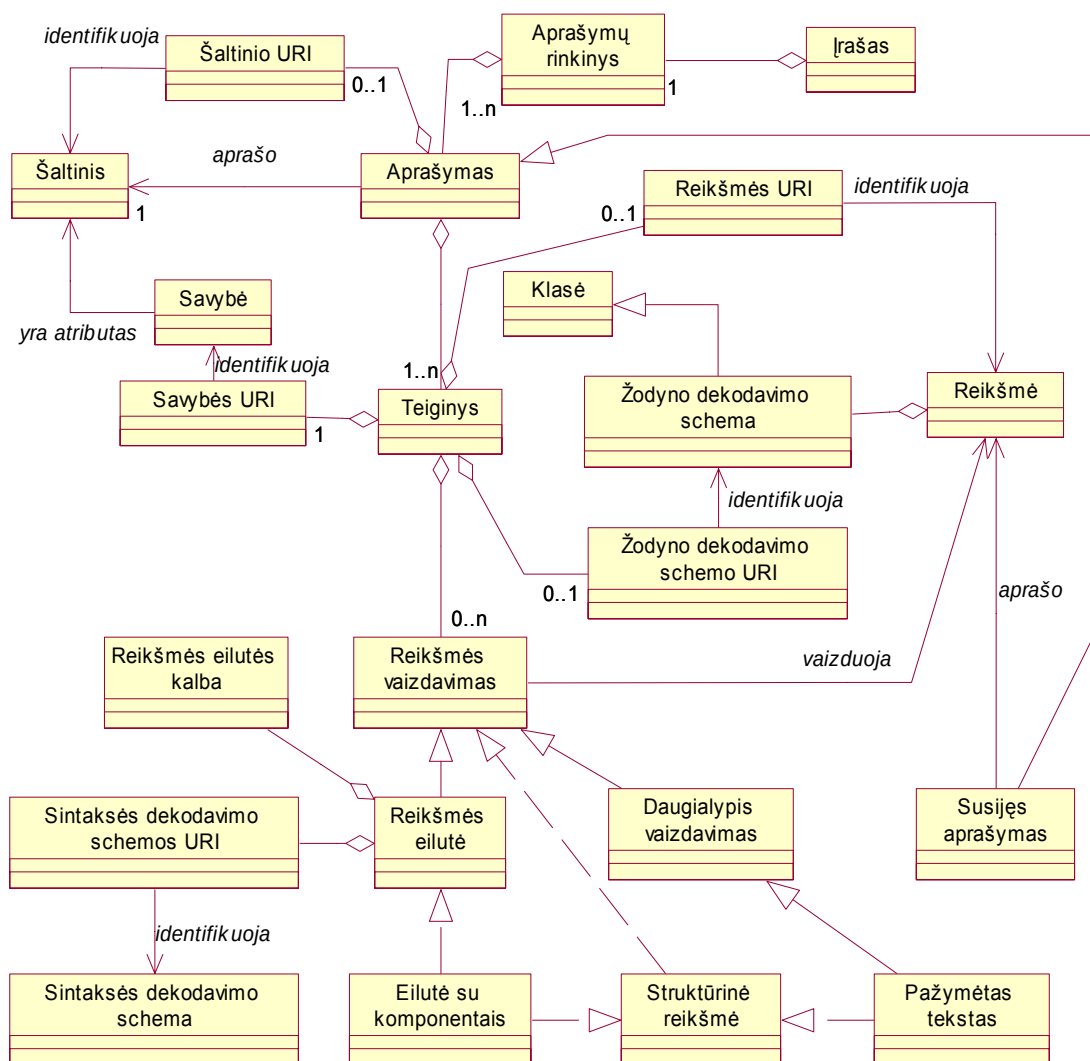
4 ir 5 paveiksluose pateiktos abstraktaus Dublin Core metaduomenų modelio klasių diagramos skirtos šaltiniui ir metaduomenų reikšmei aprašyti [10]. Iš šių modelių aišku, kad kiekvienas DCMI (Dublin Core Metadata Initiative) elementas aprašo vieną šaltinį.



3 pav. Šaltinio modelis

Abstraktaus DCMI metaduomenų modelio, pavaizduoto 3 pav. komponentų charakteristikos [10]:

- Kiekvienas *šaltinis* turi 0 arba daugiau *savybių/reikšmių porų* (property/value pairs).
- Kiekviena *savybių/reikšmių pora* sudaryta iš vienos *savybės* ir vienos *reikšmės*.
- Kiekviena *reikšmė* yra *šaltinis* (fizinė arba konceptuali esybė susijusi su savybe).
- Kiekvienas *šaltinis* gali būti vienos ar daugiau *klasių* narys.
- Kiekviena *savybė* ir *klasė* turi apibrėžtą *semantiką*.
- Kiekviena *klasė* gali turėti vieną ar daugiau *sub-klasių*.
- Kiekviena *savybė* gali būti susijusi su viena ar daugiau *smulkesnių savybių* (sub-property).



4 pav. DCMI aprašymo modelis

4 paveiksle pavaizduoto abstraktaus DCMI metaduomenų aprašymo modelio komponentų charakteristikos [10]:

- *Aprašymas* sudaromas iš vieno ar kelių *teiginių* (apie vieną ir tik vieną *šaltinį*) ir 0 arba 1 šaltinio URI (URI – nuoroda, kuri identifikuoja nagrinėjamą *šaltinį*).
- Kiekvienas *teiginys* iliustruoja *savybės/reikšmės porą* ir yra sudarytas iš *savybės URI*³ (URI – nuoroda identifikuojanti *savybę*), 0 ar 1 *reikšmės URI* (URI – nuoroda identifikuojanti

³ URI – Uniform Resource Identifier

savybės reikšmę), 0 ar 1 žodyno kodavimo *schemas URI* (URI – nuoroda identifikuojanti *reikšmės klasę*) ir 0 ar daugiau *reikšmės vaizdavimų*.

- *Reikšmės vaizdavimas* gali būti *reikšmės eilutės* (value string) arba *daugialypio vaizdavimo* (rich representation) formos.
- Kiekviena *reikšmės eilutė* yra paprasta, žmogaus įskaitoma eilutė, reprezentuojanti *šaltinį*, kuris yra *savybės reikšmė*.
- Kiekviena *reikšmės eilutė* gali būti susijusi su *sintaksės kodavimo schemas URI*, kuris identifikuoja *sintaksės kodavimo schemą* (syntax encoding scheme).
- Kiekviena *reikšmės eilutė* gali būti susijusi su *reikšmės eilutės kalba*, kuri yra ISO kalbos žyma (tag).
- Kiekvienas *daugialypis vaizdavimas* yra tam tikras *sužymėtas tekstas*, paveiksliukas, video ar audio ir t.t. arba šių dalykų kombinacija reprezentuojanti *šaltinį*, kuris yra *reikšmės savybė*.
- Kiekviena reikšmę gali būti kelių *susijusių aprašymų* tema.

Metaduomenų kodavimas pagal RDF (Resource Description Framework)

Metaduomenims koduoti, perduoti ir apdoroti dažnai naudojamas RDF standartas. Šis standartas suteikia galimybę kelioms aplikacijoms keistis mašinoms suprantama informacija. RDF metaduomenys gali būti naudojami įvairiose srityse pvz.: šaltinių radime (siekiant patobulinti paieškos mechanizmus), katalogizavime (turinio skirstymui į kategorijas) ir t.t. [11]

RDF duomenys susideda iš mazgų (nodes) ir savybių/reikšmių porų. Mazgu gali būti bet koks šaltinis (serveris, puslapis ir bet kas kam galima priskirti URI).

Šio standarto ypatybės yra šios [11]:

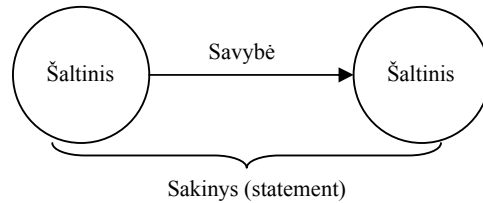
- Tai patobulintas PICS, HTML ir XML;
- Mašinoms suprantami duomenys;
- Palaiko struktūrines reikšmes;
- Palaiko metaduomenų autentifikavimą;
- Suteikia bazę įvairiems aprašymams.

RDF technologijos komponentai [11]:

- Formalus duomenų modelis;
- Sintaksė duomenų mainams;
- Schemas modelis;

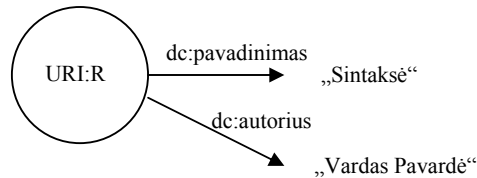
- Sintaksė mašinoms-suprantamoms schemoms;
- Užklausų ir profilių (profiles) protokolai.

RDF duomenų modelis:



5 pav. Duomenų modelis pagal RDF

RDF sintaksė (pavyzdys):



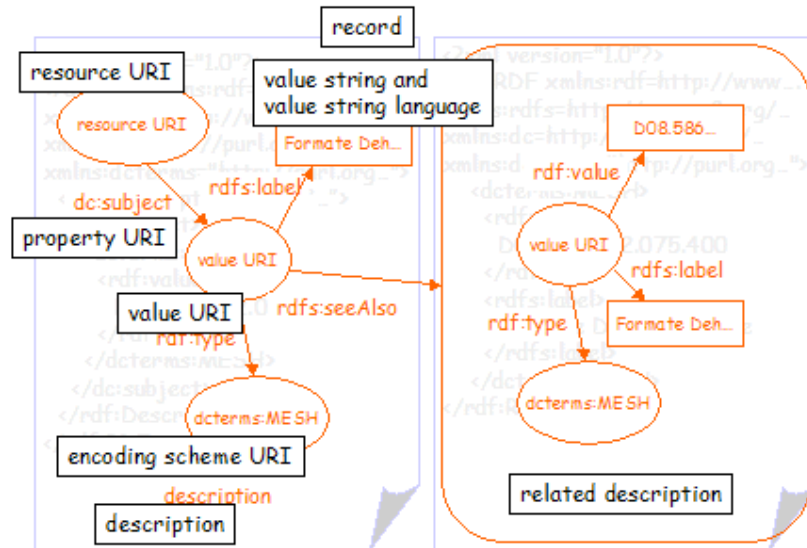
6 pav. RDF sintaksė

RDF apibendrinimas:

- Aprašo metaduomenų kodavimą, perdavimą ir apdorojimą;
- Kyla problemos, kai duomenys agreguojami iš daugelio šaltinių.

RDF ir Dublin Core metaduomenų aprašymų palyginimas

Norint parinkti tinkamesnį metaduomenų kodavimo modelį buvo sulygintos RDF ir Dublin Core terminologijos (7 pav.) [10]. Pastebėta, kad beveik visi Dublin Core metaduomenų modelio elementai turi savo atitikmenis RDF metamodelyje.



7 pav. RDF ir DC terminologijos sutapatinimas

2.5.2. Semantikos kategorijų analizė

Semantika gali būti skirstoma į tris kategorijas [7]:

- Numanoma semantika (implicit semantics) – daugiausiai naudojama internete, gali būti lengvai ištraukiama. Šis semantikos tipas dažniausiai naudojamas duomenų rinkinių apdorojimui, siekiant išgauti semantiką, kurią galima atvaizduoti formalia kalba.
- Formali semantika (formal semantics) – ontologijų forma. Tokią semantiką naudojantys reprezentavimo mechanizmai turi aiškią semantikos interpretavimo technologiją. Šie reprezentavimo mechanizmai suteikia aplikacijoms automatizuoto sprendimų priėmimo galimybę.
- Galinga semantika (powerful semantics) – neraiškos aibės (Fuzzy) ar KR (knowledge representation) mechanizmų forma.

Semantikos galimybių priklausomybė nuo semantikos tipo pavaizduota 3 lentelėje [7].

3 lentelė

Semantikos galimybių priklausomybė nuo semantikos tipo

	<i>Galimybės</i>	<i>Numanoma semantika</i>	<i>Formali semantika</i>	<i>Galinga semantika</i>
Išplėtimo fazė	Automatinis ar pusiau automatinis ontologijų generavimas	Žodžių paplitimo ir išsidėstymo tekste analizė, siekiant perprasti sistematiką ir ontologijas (Kashyap 2003)		Neraiškių aibių arba tikimybinio grupavimo naudojimas siekiant perprasti sistematiką ir ontologijas

	Nestruktūrizuoto turinio anotavimas šiomis ontologijomis (rezultatas – semantiniai metaduomenys)	Žodžių paplitimo ir hypernuorodų struktūros analizė, norint susieti sąvokų vardus iš ontologijos su resursais ir ryšiais tarp jų (Naing 2002)		Neraiškių aibių arba tikimybinių grupavimo naudojimas siekiant perprasti sistematiką ir ontologijas, arba neraiškių aibių ontologijų naudojimas
	Dviprasmiškumo pašalinimas	Grupavimo metodų arba SVM (Support Vector Mashines) naudojimas dviprasmiškumo pašalinimui (Han 2004)	Ontologijų naudojimas dviprasmiškumo pašalinimui	Neraiškių aibių ir KR mechanizmų naudojimas ontologijoms, kurios gali būti naudojamos dviprasmiškumo pašalinimui atvaizduoti
	Skirtingų schemų ir ontologijų naudojimas semantikos integravimui	Ontologijų išplėtimo naudojimas jų integravimui (Wang 2004)	Schema (diagramomis) paremti integravimo metodai (Castano 2001)	
	Semantinių metaduomenų „praturtinimas“	Anotuotų resursų analizavimas sąsajoje su ontologija tam, kad praplėsti semantinius metaduomenis (Hammond 2002)		Šitoks praturtinimas gali reikšti anotavimą naudojant neraiškių aibių ontologijas
Panaudojimo fazė	Sudėtingų užklausų apdorojimas		Užklausų validavimo hipotezė (Sheth 2003)	
	Atsakymų į klausimus sistemos	Žodžių pasikartojimo ir kitų CL metodų naudojimas klausimo ir atsakymo šaltinių analizei (Ramakrishnan 2004)	Formalių ontologijų naudojimas [Atzeni 2004]	Tikrumo lygmenų atsakymuose numatymas remiantis neraiškių aibių sąvokomis ar tikimybe
	Sąvokomis paremta paieška	Žodžių susijusių su sąvokomis paplitimo resursuose analizė	Hyponimijos, Hyperonimijos ir Partonomijos (Hypernymy,	

			Hyponymi, Partonomy) naudojimas paieškai pagerinti (Townley 2000)	
	Ryšų ir struktūros tyrimas	Pusiau struktūrizuotų duomenų saugyklų analizė struktūros išgavimui (Kuramochi and Karypis 2004)	Ontologijų naudojimas reikšmingų struktūrų išgavimui [Aleman-Meza 2003]	
	Žinomo turinio paieška	Žodžių pasikartojimo ir kitų CL metodų naudojimas resursų atitinkančių paieškos frazė analizei	Formalios ontologijos naudojimas paieškos praplėtimui	Neraiškių aibių ir KR mechanizmų naudojimas konteksto reprezentavimui
	Dinaminės vartotojo sąsajos		Ontologijų naudojimas dinaminiam vartotojų sąsajų performavimui (Quan and Karger 2004)	
	Domėjimusi paremto turinio formavimas	Turinio analizė siekiant identifikuoti turinio sąvokas atitinkančias domėjimosi sritis	Vartotojo profilis turės su juo susietą ontologiją susijusią su domėjimosi srities sąvokomis	
	Navigacinė ir tiriamoji paieška (Guha 2003)	Navigacinei paieškai reikia išanalizuoti nestruktūrizuotą turinį	Atradimų stiliaus užklauskos pusiau struktūrizuotiems duomenims (Anyanwu and Sheth 2002)	Neraiškių aibių palyginimas tiriamosios paieškos rezultatams

2.5.3. Semantinių asociacijų analizė

RDF semantinių metaduomenų modelyje esybės reikšmė yra nustatoma pagal tai, kaip ji siejasi su kitomis esybėmis. Kiekvienas ryšys tarp esybių yra vadinamas semantine asociacija. Kiekviena semantinė asociacija susideda iš tarpinių esybių ir ryšių (esybių savybių).

Semantinių asociacijų formalizavimas [12]:

1 taisyklė

Semantinis susietumas. Dvi esybės e_1 ir e_2 yra semantiškai susijusios, jei RDF grafe egzistuoja seka $e_1, P_1, e_2, P_2, e_3, \dots, e_{n-1}, P_{n-1}, e_n$ kur $e_i, 1 \leq i \leq n$ yra esybės, o $P_j, 1 \leq j \leq n$ yra savybės ir dvi esybės e_i ir e_{i+1} sieja ryšys P_j . Esybių ir savybių seka nusako *semantinį kelią*.

2 taisyklė

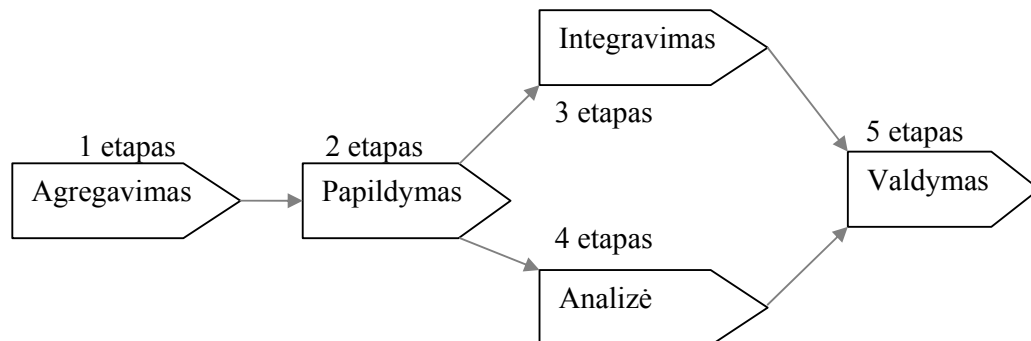
Semantinis panašumas. Dvi esybės e_1 ir f_1 yra semantiškai panašios, jei egzistuoja du semantiniai keliai $e_1, P_1, e_2, P_2, e_3, \dots, e_{n-1}, P_{n-1}, e_n$ ir $f_1, Q_1, f_2, Q_2, f_3, \dots, f_{n-1}, Q_{n-1}, f_n$ semantiškai jungiantys e_1 su e_n ir f_1 su f_n (nurodyta tvarka) ir kiekviena esybių savybių pora P_i ir $Q_i, 1 \leq i \leq n$ tenkina vieną iš sąlygų: $P_i = Q_i, P_i \subseteq Q_i, Q_i \subseteq P_i$.

3 taisyklė

Semantinė asociacija. Dvi esybės e_x ir e_y yra sujungtos semantine asociacija, jei jos yra semantiškai susijusios arba semantiškai panašios.

2.5.4. Darbo su semantiniais metaduomenimis etapai

Skiriami penki pagrindiniai darbo su semantiniais metaduomenimis etapai [4]. Šie etapai pavaizduoti 8 paveiksle.



8 pav. Darbo su semantiniais duomenimis etapai

Toliau pateikiamas 8 paveiksle pavaizduotų etapų aprašymas:

1. Pirmame etape nustatoma struktūrizuotų ir nestruktūrizuotų duomenų saugojimo vieta.
2. Antrame etape turinys klasifikuojamas naudojant pažangias klasifikavimo technologijas ir semantiškai apdorojamas (nustatoma apie ką yra tie duomenys).
3. Trečiame etape nustatoma su koku dar turiniu yra susiję duotieji duomenys.
4. Ketvirtame etape nustatoma kokie konkrečiai duomenys reikalingi vartotojui.

5. Penktame etape pasirenkama tinkama valdymo strategija.

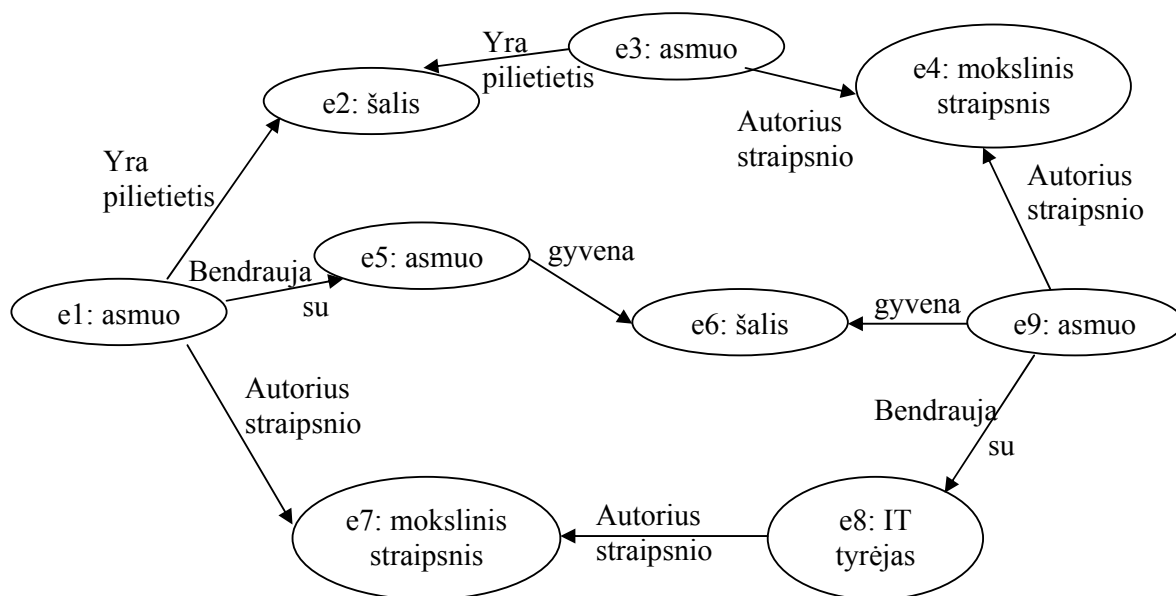
Su semantiniais metaduomenimis dirbančių organizacijų tyrimai leidžia įvardinti keturis pagrindinius uždavinius, kuriuos turi atlikti semantinius metaduomenis apdorojanti sistema [5]:

1. Automatiškai apdorojama ir apibrėžiama bei tekste surandama reikiama informacija (nepriklausomai nuo šaltinio tipo ir formato);
2. Turinys automatiškai suskirstomas į kategorijas;
3. Automatiškai sukuriami reikiamos semantinės asociacijos (paaiškinimas 2.5.3. skyriuje);
4. Semantinių užklausų mechanizmas pagal užduotus kriterijus randa reikiamą informaciją.

2.6. Panašių sistemų (Lietuvos ir tarptautiniu mastu) analizė

Šiuo metu yra sukurtos kelios semantinių metaduomenų išgavimo ir apdorojimo technologijos ir priemonės, tačiau jos skirtos tik internetinėm technologijom. Pagrindinės firmos dirbančios šioje srityje yra Semagix, Ontoprise ir Network Interference. Šių firmų darbo dėka yra sukurta keletas repozitorių, saugančių iš internetinių puslapių išgautus semantinius metaduomenis, tačiau jų veikimo principai nėra viešai skelbiami. Pagrindiniai repozitoriai yra TAP KB (saugo informaciją susijusią su įvairiais autoriais, sportu, įmonėmis ir t.t.), SWETO (saugo 800 000 įrašų ir 1,5 milijono ryšių tarp jų), WebFountain ir Semagix Freedom [3].

Visos šios semantinių metaduomenų apdorojimo priemonės duomenų aprašymui naudoja RDF schemą (3 pav.) ir yra paremtos ontologijų naudojimu [3].



9 pav. Esybių vaizdavimas RDF

Pagrindinės technologijos skirtos darbui su semantiniiais metaduomenimis:

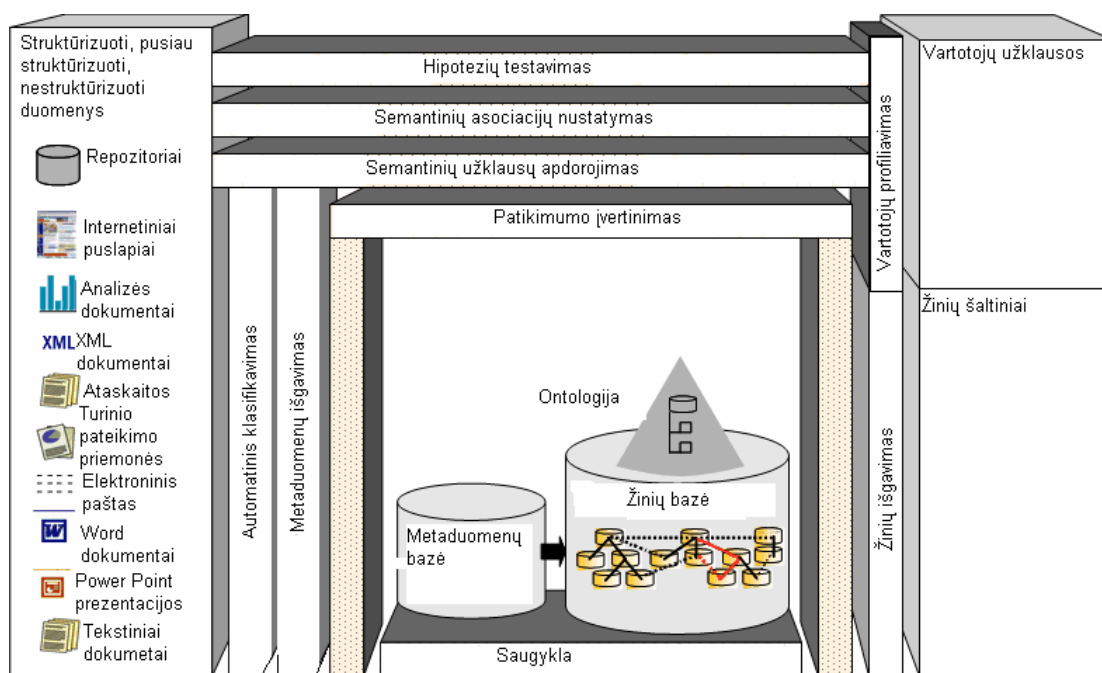
1. Sparčiai besivystančios (esamos) technologijos

- Žinių išgavimas
- Metaduomenų išgavimas
- Automatinis klasifikavimas

2. Naujausios (dar kuriamos) technologijos

- Semantinių asociacijų nustatymas
- Semantinių užklausų apdorojimas
- Patikimumo tikrinimas
- Vartotojų profiliavimas
- Hipotezių testavimas

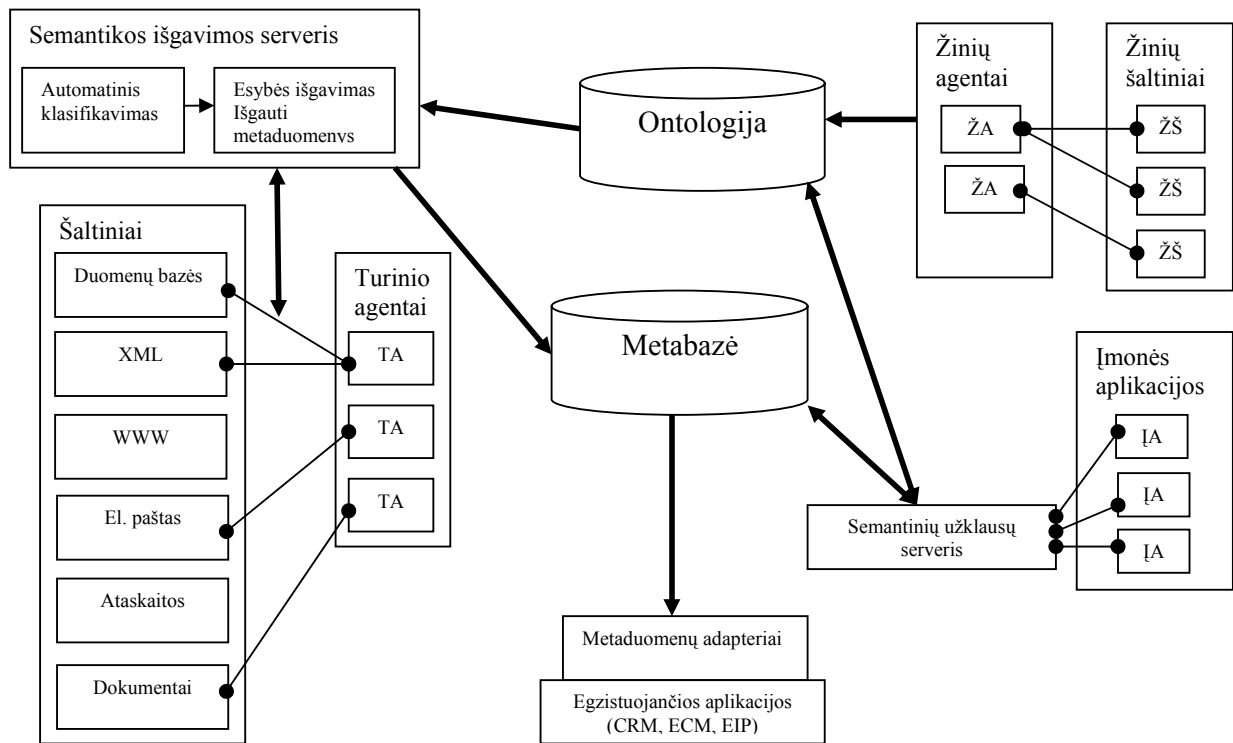
Semantinių metaduomenų sistemoje naudojamų technologijų pasiskirstymas pavaizduotas 6 paveiksle.



10 pav. Semantinių metaduomenų apdorojimo technologijos

2.6.1. Semagix FREEDOM sistemos apžvalga

Geriausius darbo semantikos srityje rezultatus yra pasiekusi kompanija yra Semagix. Šios kompanijos produktas – programa FREEDOM plačiai naudojama JAV įstaigose, oro uostuose. Pagrindiniai produkto komponentai pavaizduoti 11 paveiksle.



11 pav. Semagix FREEDOM komponentai

Semagix Freedom sistema naudojama nustatyti žinių šaltinius iš struktūrizuotų, pusiau struktūrizuotų ir nestruktūrizuotų duomenų išgauna reikiamus semantinius metaduomenis. Sistema nėra universali, ji pritaikyta informacijai apie Politikus, Įmones, Žmones ir Teroristines organizacijas išgauti. Freedom pagrindu naudojama oro uostuose siekiant patikrinti ar bilietą rezervuojantis asmuo nėra susijęs su kokia nors teroristine organizacija.

Sistemos Semagix Freedom naudojami žinių šaltiniai pateikti 4 lentelėje.

4 lentelė

Semagix Freedom žinių šaltiniai

Office of Foreign Assets Control (OFAC)	Capital Advantage (CA)
Federal Bureau of Investigation (FBI)	The Interdisciplinary Center (ICT)
Central Intelligence Agency (CIA)	Federation of American Scientists (FAS)
Data supplied from NASA (DPL)	Hoover's (H)
ZDNet (ZD)	Market Guide (MG)

Toliau pateikiami iš žinių šaltinių išgaunamos esybių klasės ir ryšiai, naudojami kaip semantinių metaduomenų kategorijos tolimesniam sistemos darbui (semantinių metaduomenų išgavimui iš pateiktų šaltinių).

Asmuo (OFAC, FBI, DPL)**-politikas (OFAC, FBI, CIA, CA)**

politikas susijęs su politine Organizacija

politikas užimama Postą

-teroristas (OFAC, FBI, DPL)

teroristas priklauso teroristinei organizacijai

teroristas minimas stebimų asmenų/organizacijų sąrašė

-kompanijos vadovas (MG)

Kompanijos vadovas vadovauja Įstaigai

Asmens adresas (OFAC, FBI)**Asmens gimimo data (OFAC, FBI)****Vieta****-Organizacija įsikūrusi Vietoje (H, OFAC)****-vietovėje praktikuojama Religija (CIA)****-organizacijos valdžia mieste (H)****Įvykis/Dalykas****-įvykis (ICT)**

Teroristinė organizacija dalyvavusi teroristiniame išpuolyje (ICT)

-politinė organizacija (CIA, CA)

Politinės organizacijos vieta tarp valdžios institucijų

Politinė organizacija susijusi su organizacija

- stebimų asmenų/organizacijų sąrašas (OFAC, FBI, DPL)

Teroristų Organizacija yra stebimų asmenų/organizacijų sąrašė (OFAC, FBI, DPL)

-organizacija (OFAC, FBI, FAS, ICT, CA, CIA)

organizacija yra stebimų asmenų/organizacijų sąrašė

organizacija yra suborganizacijos dalis

-įstaiga

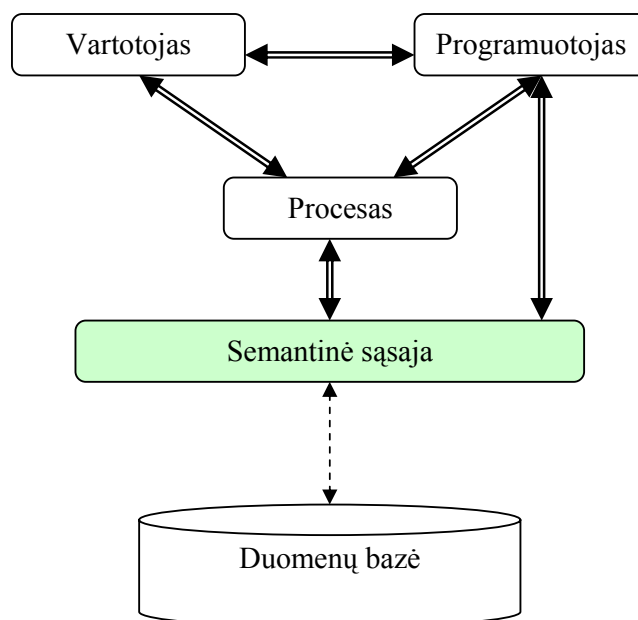
Įstaigos gaminamas produktas (ZD)

Įstaigos padaliniai (MG)

Įstaiga dirba pramonės šakoje (H)

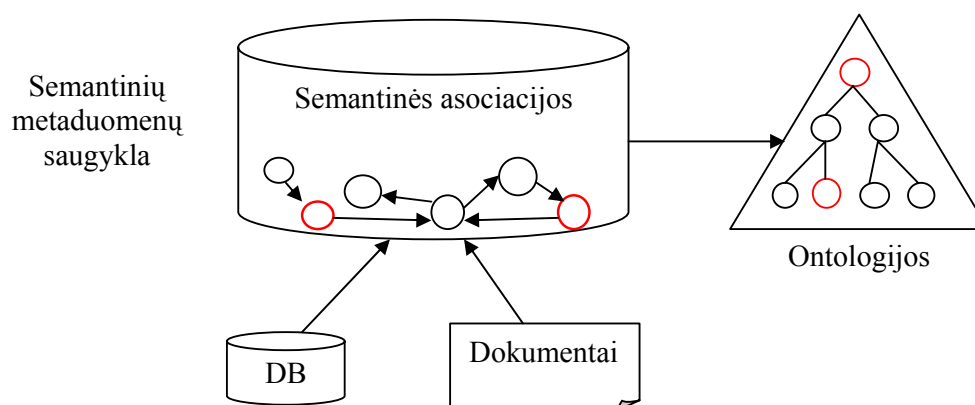
2.7. Architektūros ir galimų įgyvendinimo priemonių variantų analizė

Išanalizavus esamą sistemą, pasirinkta kuriamos sistemos architektūra pateikta 12 paveiksle. Kuriama sistema galės naudotis trijų tipų vartotojai: paprasti vartotojai, programuotojai ir kiti procesai. Detalesnė sistemos architektūra pateikta 14 paveiksle.



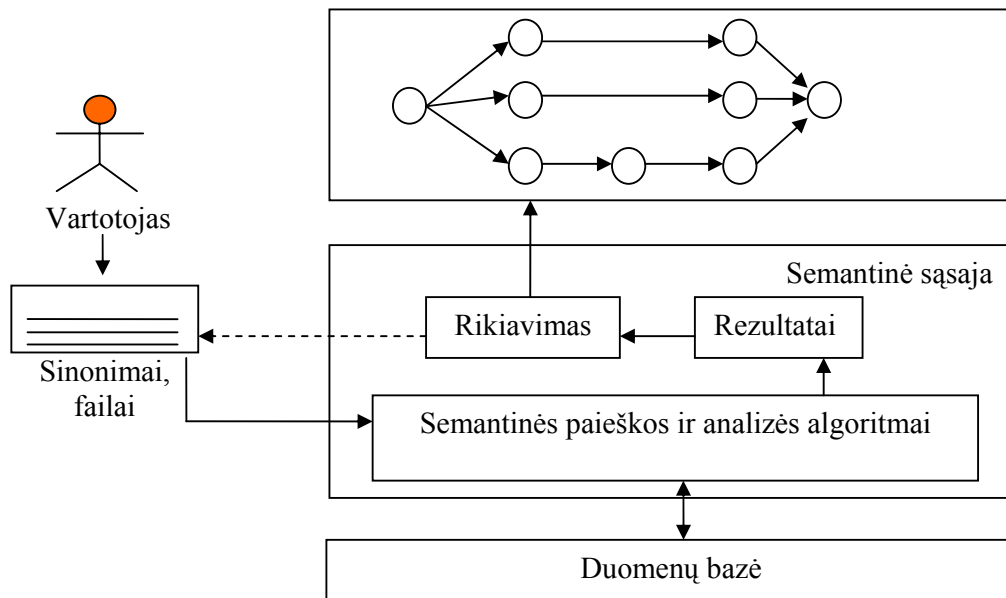
12 pav. Architektūros pavyzdys

13 paveiksle pateiktas semantinių asociacijų tarp dviejų raudonai pažymėtų įrašų pavyzdys. Kaip matome, tarp pateiktų dviejų įrašų nėra tiesioginio ryšio, taigi siekiant šį ryšį rasti mums reikalingas labai lankstus traktavimo būdas.



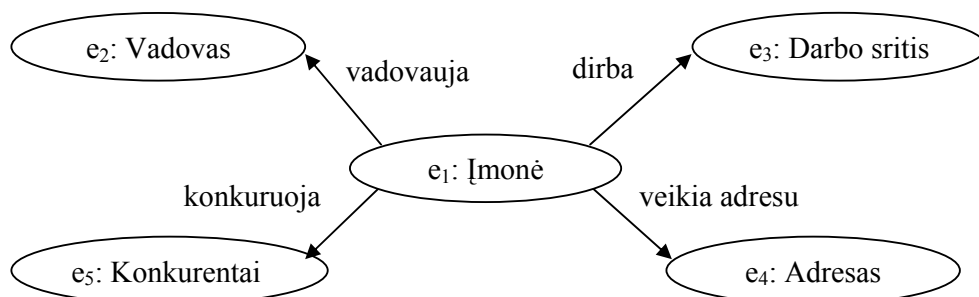
13 pav. Semantinės asociacijos

Detalesnė metaduomenų išgavimo ir apdorojimo sistemos architektūra:



14 pav. Detalesnė sistemos architektūra

Vartotojas užpildo tam tikrą ieškomo turinio aprašymą ir semantinių užklausų, bei analizės algoritmų dėka išrenka tam tikrus duomenis, kurie surikiuoti grąžinami vartotojui. Kuriamos semantinių metaduomenų apdorojimo sistemos naudojama RDF schema:



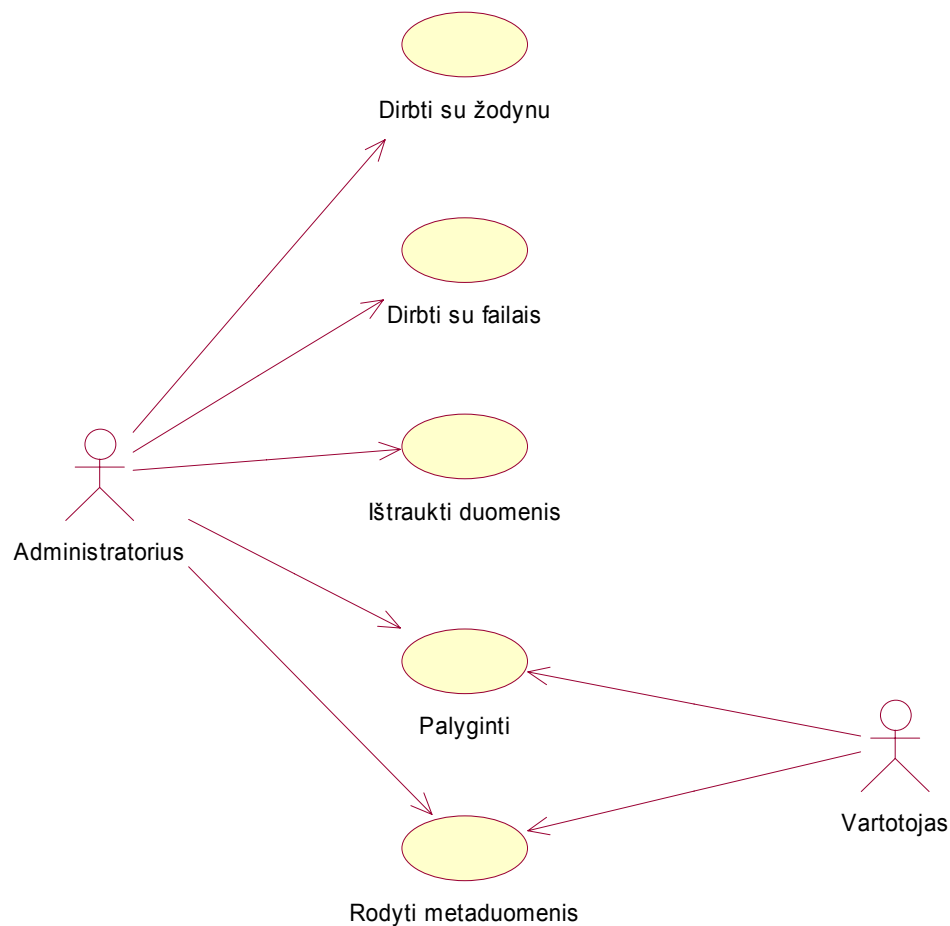
15 pav. Projektuojamos sistemos RDF schema

2.8. Darbo tikslas ir siekiami privalumai

Siekama darbe sukurti semantinių metaduomenų saugojimo, apdorojimo ir ištraukimo sistemą. Stengiamasi sukurti efektyvų semantinių metaduomenų išgavimo algoritmą, minimizuoti šios operacijos laiką ir kiek įmanoma praplėsti sinonimų biblioteką.

2.9. Kompiuterizuojamos sistemos funkcijos

Sistemos, naudojančios kuriamą metodą pradinės panaudojimo atvejų specifikacijos:

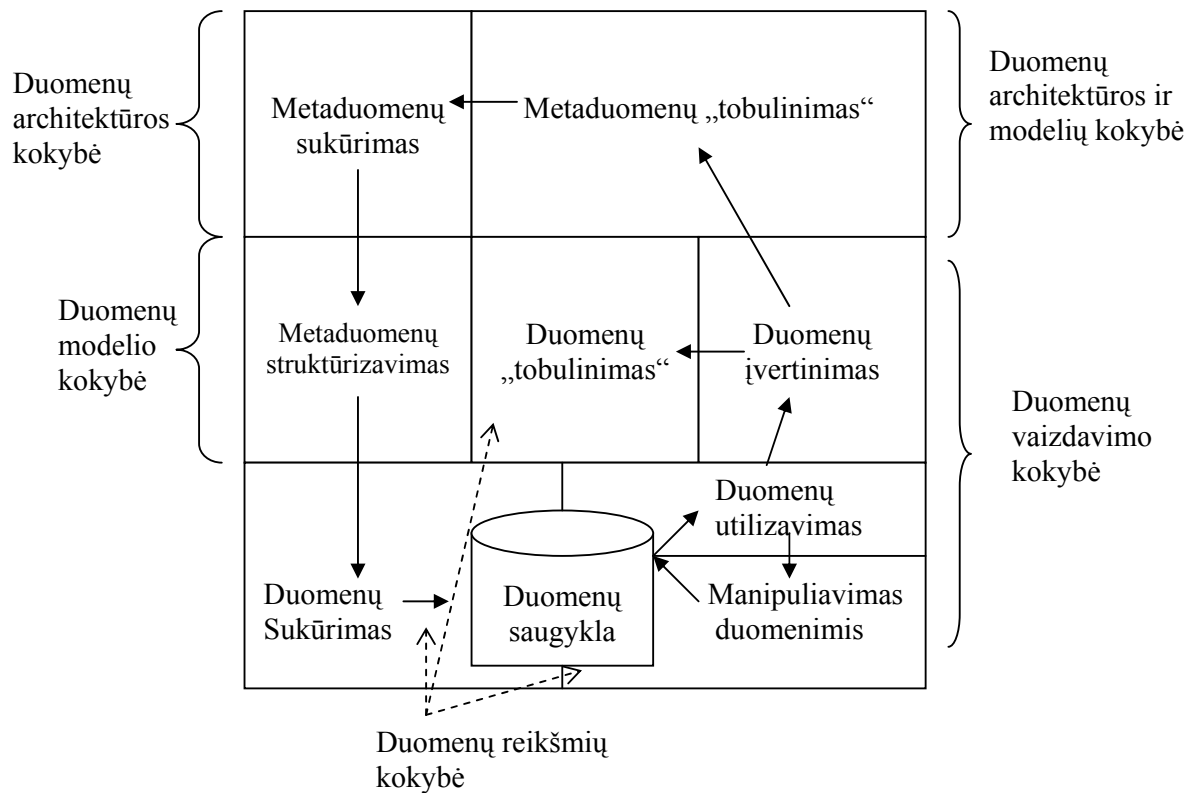


16 pav. Nedetalizuota sistemos panaudojimo atvejų diagrama

Kuriama sistema vykdo penkias pagrindines funkcijas: Dirbti su žodynu, Dirbti su failais, Ištraukti duomenis, Rodyti metaduomenis ir Palyginti. Darbo su žodynu metu vartotojui bus suteikta galimybė įvesti sinonimus susijusius su tam tikra metaduomenų sritimi. Darbo su failais metu pasirenkami reikiami failai iš kurių bus išgaunami semantiniai metaduomenys. Išgauti metaduomenys atvaizduojami duomenų bazės lange. Funkcija Palyginti suranda ryšį (semantinę asociaciją) tarp dviejų vartotojo įvestų užklausų.

2.10. Nefunkciniai reikalavimai ir apribojimai

Reikalavimai duomenų kokybei pavaizduoti 17 paveiksle [8].



17 pav. Metaduomenų kokybės kriterijai įvairiose duomenų gyvavimo ciklo fazėse

5 lentelė

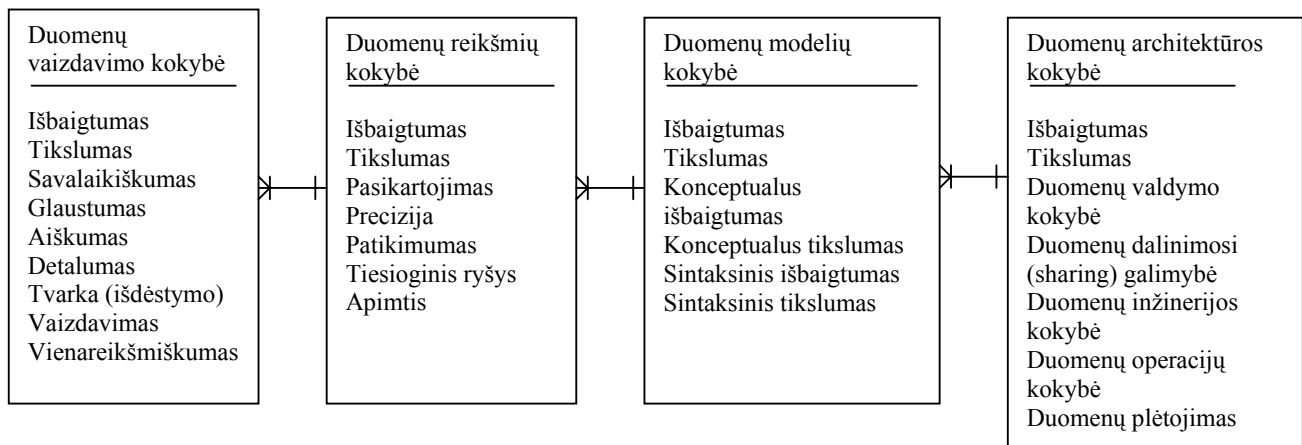
Metaduomenų kokybės matavimo kriterijai

		Metaduomenys			Duomenys				
Kriterijus	Aprašymas	Tobulini mas	Kūrimas	Struktūrizavimas	Kūrimas	Manipuliacijavimas	Tobulini mas	Utilizavimas	Įvertinimas
Duomenų Architektūros kokybė	Svarbi duomenų kūrimo ir tobulinimo fazėse	+	+						+
Duomenų modelio kokybė	Svarbi metaduomenų struktūrizavimo ir tobulinimo fazėse	+		+					+
Duomenų reikšmių	Svarbi duomenų kūrimo, manipuliavimo ir tobulinimo				+	+	+		+

kokybė	fazėse								
Duomenų vaizdavimo kokybė	Svarbi duomenų utilizavimo fazę							+	+

2.11. Rezultato kokybės kriterijai

Metaduomenų kokybės kriterijai vartotojo – sistemos kūrėjo atžvilgiu [8]:



18 pav. Metaduomenų kokybės kriterijai vartotojo – sistemos kūrėjo atžvilgiu

Duomenų reikšmių kokybės kriterijus remiasi Redman (1992) ir Levitin (1995) studijomis. Šio kokybės kriterijaus atributai aprašyti 6 lentelėje [8].

6 lentelė

Duomenų reikšmių kokybės kriterijaus atributai

Atributas	Aprašymas
Išbaigtumas	Atributai yra priskirti visoms to reikalaujančioms esybėms
Tikslumas	Duomenų reikšmės yra be klaidų
Pasikartojimas	Duomenų reikšmės yra atnaujinamos
Precizija	Turi būti nustatytas duomenų reikšmių specifikuojimo detalumas
Patikimumas	Tik tam tirkų kategorijų darbuotojai gali keisti/išsaugoti duomenis
Tiesioginis ryšys	Išsaugoti duomenys yra tiesiogiai susiję su vartotojo reikalavimais
Apimtis	Negali būti per daug ar per mažai duomenų

Duomenų vaizdavimo kokybės kriterijus paremtas Tufte (1990) tyrimais, o vėliau ir Redman (1992), bei Fox (1994). Kriterijaus atributai pavaizduoti 7 lentelėje [8].

7 lentelė

Duomenų vaizdavimo kokybės kriterijaus atributai

Atributas	Aprašymas
Išbaigtumas	Vartotojui vaizduojami visi jo reikalavimus atitinkantys duomenys
Tikslumas	Vartotojui vaizduojami duomenys yra be klaidų
Savalaikiškumas	Duomenys vartotojui turi būti vaizduojami tada, kada to reikia
Glaustumas	Vartotojui vaizduojami duomenys yra susiję su vartotojo reikalavimais. Svarbu, kad duomenų netrūktų
Aiškumas	Duomenų vaizdavimo forma turi būti aiški, lengvai suprantama vartotojui
Detalumas	Vaizduojamų duomenų detalumas turi atitikti vartotojo reikalavimus
Tvarka	Duomenys turi būti vaizduojami vartotojui patogiausią tvarka
Vaizdavimas	Duomenys turi būti vaizduojami formatu, labiausiai atitinkančiu vartotojo kompetenciją
Vienareikšmiškumas	Vartotojui vaizduojami duomenys negali reikalauti kažkokio interpretavimo

Duomenų modelio kokybės kriterijus remiasi Reingruber ir Gregory (1994) studijomis. Prie šių studijų taip pat prisidėjo ir Fox (1994) su Levitin (1993). Kriterijaus atributai aprašyti 8 lentelėje [8].

8 lentelė

Duomenų modelio kokybės kriterijaus atributai

Atributas	Apibūdinimas
Išbaigtumas/Tikslumas	Modelį sudaro komponentai reprezentuojantys realaus pasaulio sistemos komponentus. Modelio komponentai – esybės, atributai ir aprašymai turi tiksliai reprezentuoti veikiančią sistemą.
Konceptualus išbaigtumas	Modelis turi būti pakankamai išsamus, platus.
Konceptualus tikslumas	Modelių komponentų aprašymas ir ryšiai tarp esybių turi būti be klaidų.
Sintaksinis išbaigtumas	Čia reikia apsvarstyti klausimus „ar modeliuotojai įtraukė į modelius visą reikiamą informaciją?“, „ar modeliuose esančios informacijos užtenka?“
Sintaksinis tikslumas	Modeliuotojai turi nuosekliai ir teisingai pritaikyti modeliavimo techniką esybių, atributų ir ryšių aprašymui.

Duomenų architektūros kokybės kriterijų nagrinėjo Spewack (1993). Atributai aptarti 9 lentelėje [8].

9 lentelė

Duomenų architektūros kokybės kriterijaus atributai

Atributas	Apibūdinimas
Išbaigtumas	Architektūra yra tinkama naudoti, visos reikiamos funkcijos veikia
Tikslumas	Informacija aprašanti architektūrą yra tiksliai reprezentuojama naudojant reikiamą metodologiją
Duomenų valdymo kokybė	Duomenys yra valdomi paskirstant centralizuotų resursų valdymą
Duomenų dalinimosi galimybė	Galimybė dalintis (share) duomenimis organizacijos viduje
Duomenų plėtojimas	Duomenų struktūra turi būti lanksti, tai pagrinde svarbu fiziniam architektūros įgyvendinimui

2.12. ISO 9126 standartas sistemos kokybei įvertinti

Sistemos kokybės tyrimas bus atliekamas pagal ISO 9126 standarte numatytus kokybės vertinimo kriterijus. Šio standarto tikslas – apibrėžti programinės įrangos kokybės vertinimo šabloną. ISO/IEC 9126 standartas nekelia reikalavimų programinei įrangai, tačiau pateikia kokybės modelį, kuris gali būti taikomas bet kokiai programinei įrangai. Standartas pateikia 6 kokybės charakteristikas ir rekomenduoja kiekvieną jų skaidyti į subcharakteristikas [7]:

1. Funkcionalumas (*Functionality*)
 - 1.1. Tinkamumas (*Suitability*)
 - 1.2. Tikslumas (*Accurateness*)
 - 1.3. Sąveika (*Interoperability*)
 - 1.4. Suderinamumas (*Compliance*)
 - 1.5. Sauga (*Security*)
2. Patikimumas (*Reliability*)
 - 2.1. Branda (*Maturity*)
 - 2.2. Klaidų tolerancija (*Fault tolerance*)
 - 2.3. Atstatomumas (*Recoverability*)
3. Panaudojamumas (*Usability*)
 - 3.1. Suprantamumas (*Understandability*)

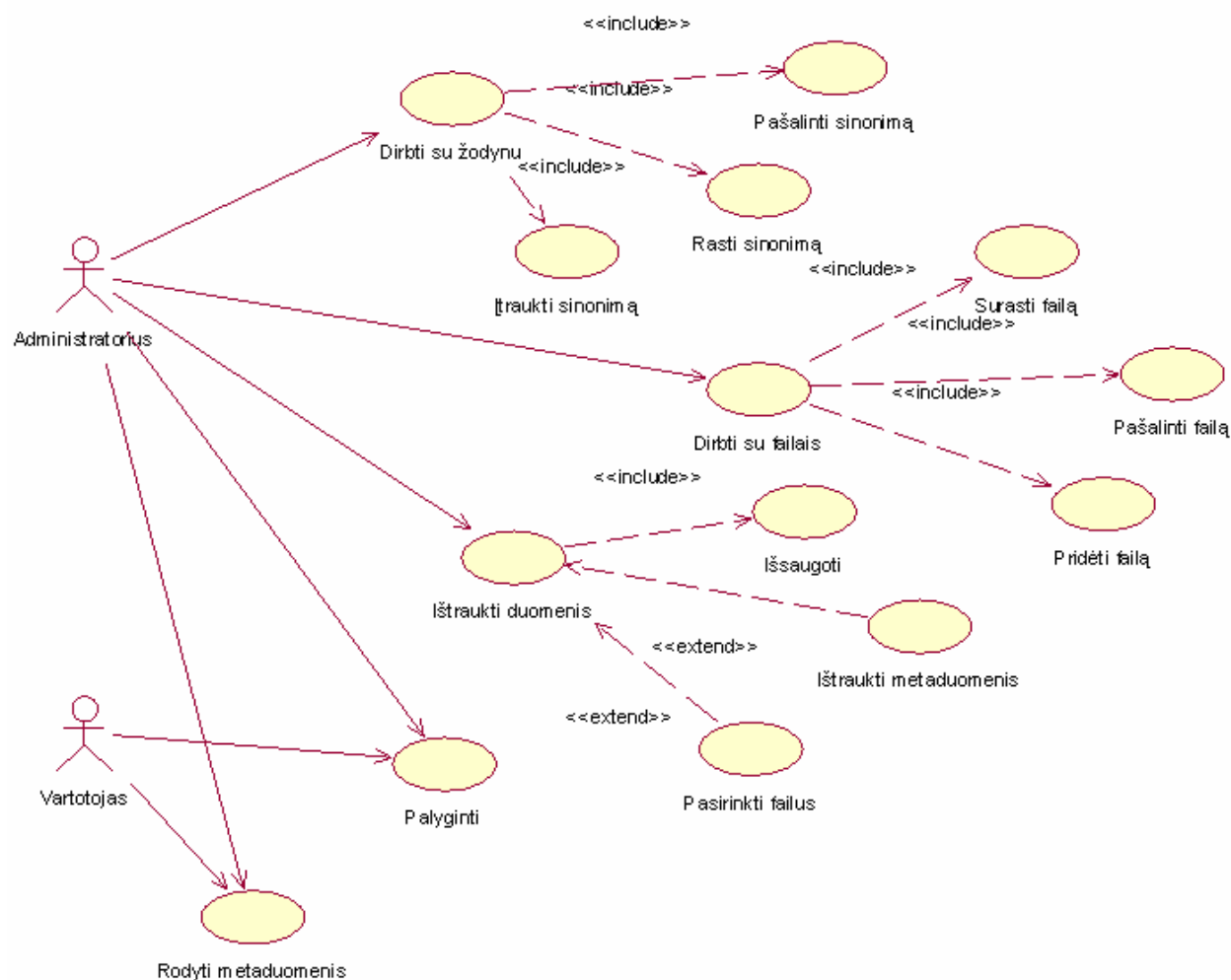
- 3.2. Apsimokymas (*Learnability*)
- 3.3. Valdomumas (*Operability*)
- 4. Efektyvumas (*Efficiency*)
 - 4.1. Laikinė charakteristika (*Time behavior*)
 - 4.2. Resursai (*Resource behavior*)
- 5. Pornešamumas (*Portability*)
 - 5.1. Adaptacija (*Adaptability*)
 - 5.2. Įdiegiamumas (*Installability*)
 - 5.3. Pakeičiamumas (*Replaceability*)
- 6. Priderinamumas (*Maintainability*)
 - 6.1. Analizuojamumas (*Analyzability*)
 - 6.2. Modifikuojamumas (*Changeability*)
 - 6.3. Stabilumas (*Stability*)
 - 6.4. Testuojamumas (*Testability*)

2.13. Analizės išvados

Darbe buvo apžvelgtos problemos susijusios su duomenų semantikos aprašymu, parodyta kurioje vietoje atsiduria semantikos problema einant link ontologijos.

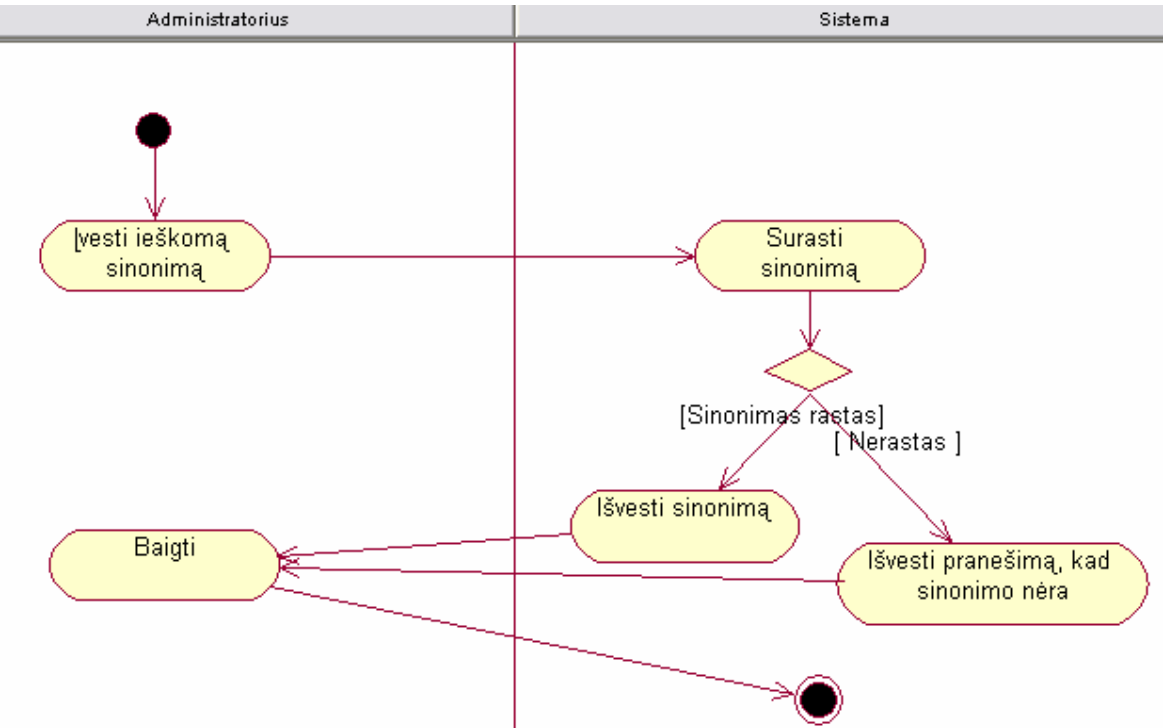
Išnagrinėjus prieinamus literatūros šaltinius galima daryti išvadą, kad šiuo metu nėra standartų aprašyti duomenų semantikai taip, kad galima būtų nestruktūrizuotuose duomenyse programiškai aptikti reikalingą informaciją semantinė prasme. Yra kompanijos, kurios užsiima semantikos metaduomenų generavimu iš nestruktūrizuotų ir pusiau struktūrizuotų duomenų, tačiau neaiškūs teoriniai principai ir praktinės galimybės. Norint sukurti sistemą, kuri analizuotų nestruktūrizuotus duomenis ir formuotų atributus reliacinei duomenų bazei pagal sukurtą schemą reikia atlikti šių kompanijų siūlomų produktų galimybių analizę ir apibrėžti jų taikymo sritį.

3. Sistemos reikalavimų specifikacija



19 pav. Papildyta panaudojimo atvejų diagrama

Veiklos diagrama panaudojimo atvejui **RASTI SINONIMĄ**:



20 pav. Rasti sinonimą

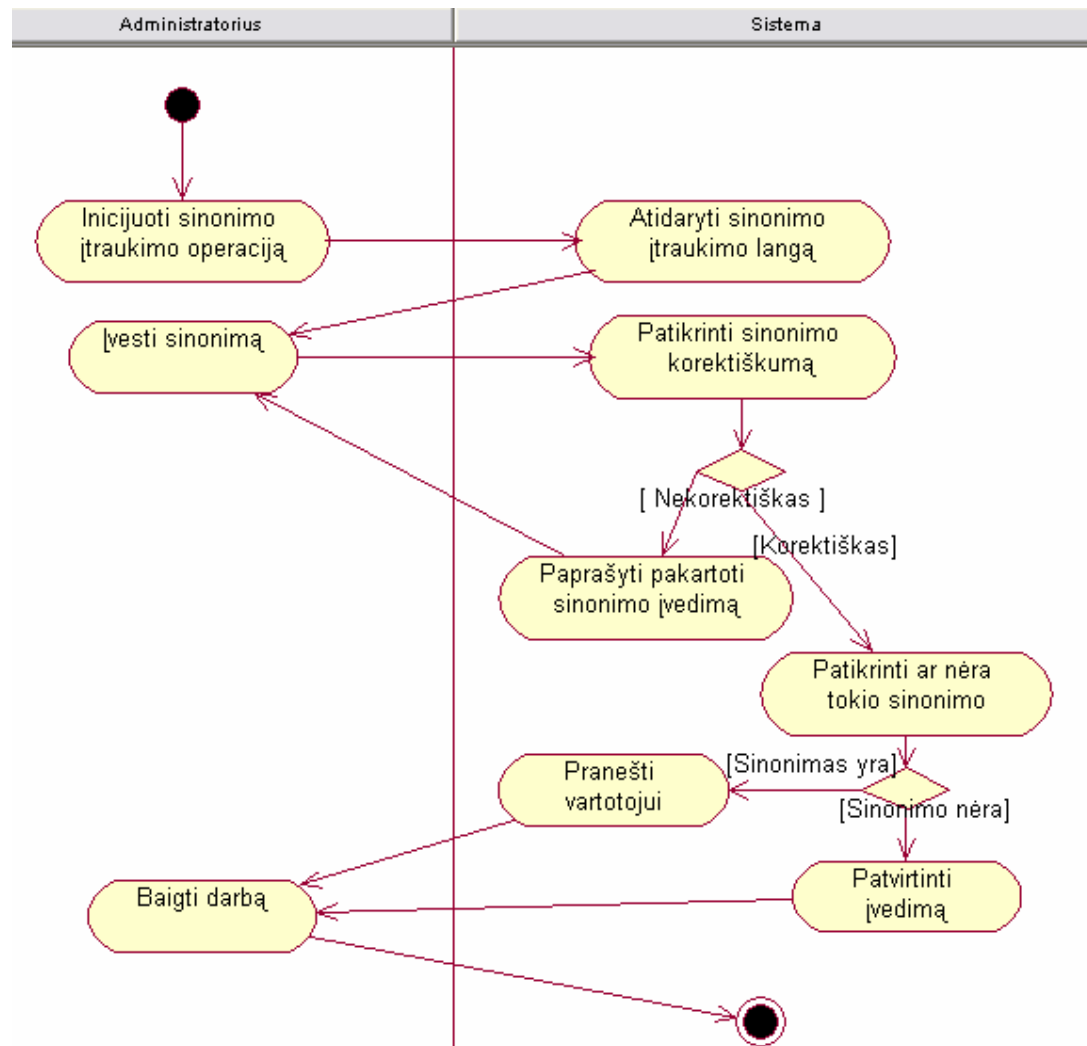
Panaudojimo atvejo specifikacija:

10 lentelė

PA Rasti sinonimą specifikacija	
Panaudojimo atvejis	Rasti sinonimą
Nr.	N1
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	Sinonimas turi būti įvestas
Pagrindinis įvykių srautas	Sistemos reakcija
1.0.Įvedamas ieškomas sinonimas	1.1. Sinonimas surandamas 1.2. Sinonimas išvedamas administratoriui
Po-sąlyga	Sinonimas rastas
Alternatyvos	1.1.a. Sinonimas nerandamas 1.2.a. Adminitratoriui pranešama, kad sinonimas nerastas
Vykdymo variantai	Administratorius įveda ieškomą sinonimą
Veiklos taisyklės	Sinonimas bus rastas, jei jis yra anksčiau įvestas į sinonimų žodyną
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju

Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Įeina į pagrindinį PA “Dirbti su žodynu” (ryšio tipas <<include>>)
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veiklos diagrama panaudojimo atvejui **ĮTRAUKTI SINONIMĄ**:



21 pav. Įtraukti sinonimą

Panaudojimo atvejo specifikacija:

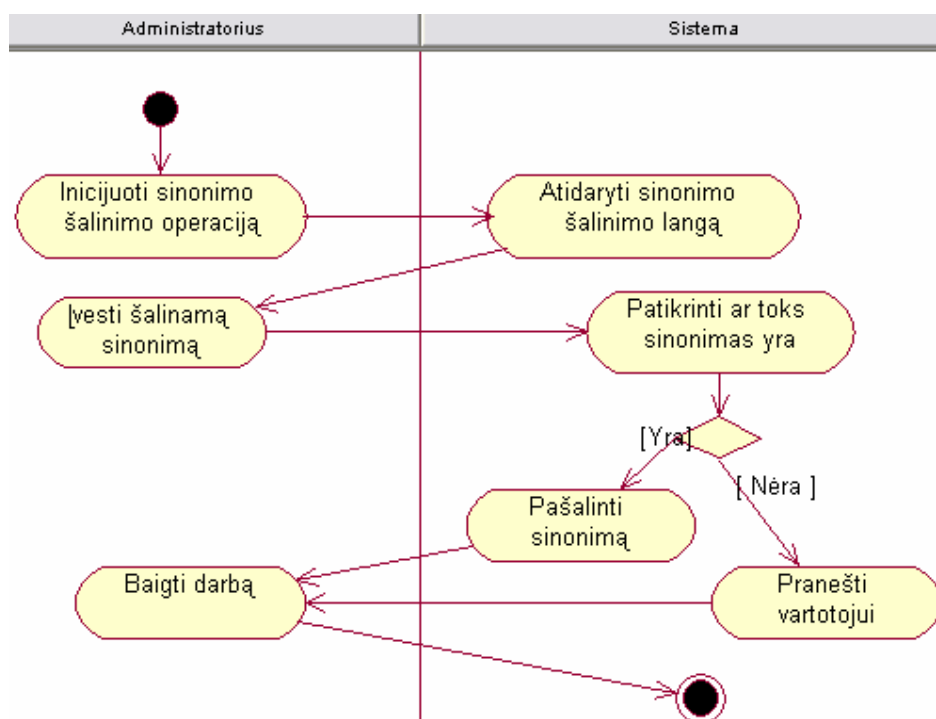
11 lentelė

PA Įtraukti sinonimą specifikacija

Panaudojimo atvejis	Įtraukti sinonimą
Nr.	N2
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema

	(SMAS)
Prieš-sąlyga	-
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedamas sinonimas	1.1 Patikrinamas sinonimo korektiškumas 1.2 Patikrinama ar tokio sinonimo dar nėra 1.3 Patvirtinamas įvedimas
Po-sąlyga	Sinonimas įvestas
Alternatyvos	1.1.a. Sinonimas nekorektiškas, todėl prašoma pakartoti 1 žingsnį 1.3.a. Toks sinonimas jau yra, todėl apie tai pranešama administratoriui
Vykdomo variantai	Administratorius įveda naują sinonimą
Veiklos taisyklės	Sinonimas bus įrašytas, jei jis įvestas korektiškai ir jei tokio sinonimo dar nėra
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Įeina į pagrindinį PA "Dirbti su žodynu" (ryšio tipas <<include>>)
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veklos diagrama panaudojimo atvejui **PAŠALINTI SINONIMĄ**



22 pav. Pašalinti sinonimą

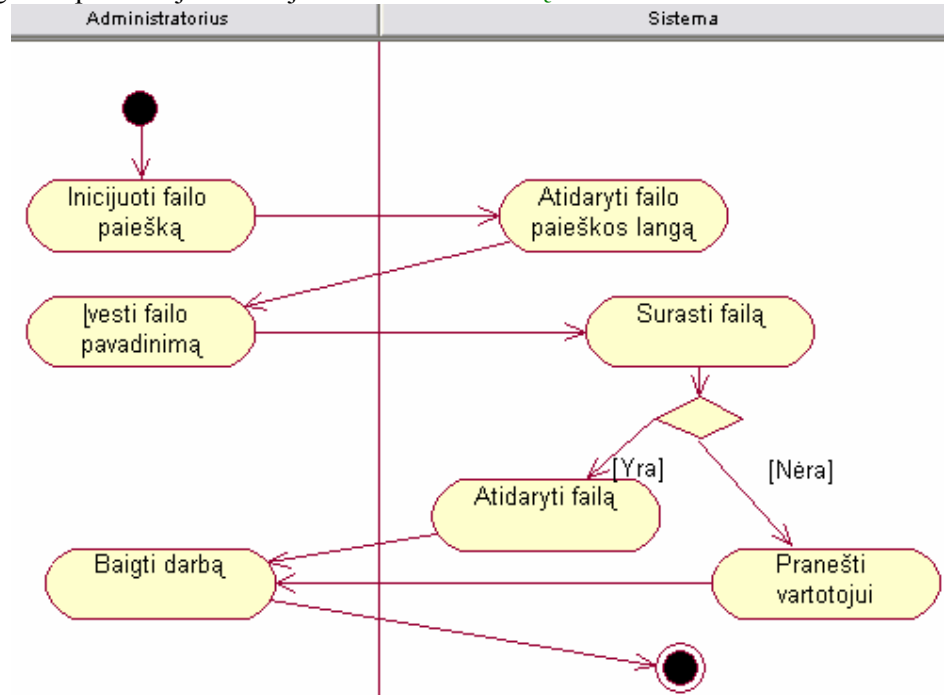
Panaudojimo atvejo specifikacija:

12 lentelė

PA Pašalinti sinonimą specifikacija

Panaudojimo atvejis	Pašalinti sinonimą
Nr.	N3
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	Toks sinonimas turi būti įvestas
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedamas sinonimas	1.1. Patikrinama ar toks sinonimas yra 1.2. Sinonimas pašalinamas
Po-sąlyga	Sinonimas pašalintas
Alternatyvos	1.2.a. Tokio sinonimo nėra, todėl apie tai pranešama administratoriui
Vykdyto variantai	Administratorius pašalina sinonimą
Veiklos taisyklės	Sinonimas bus pašalintas, jei prieš tai buvo įvestas į sinonimų žodyną
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Įeina į pagrindinį PA "Dirbti su žodynu" (ryšio tipas <<include>>)
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veklos diagrama panaudojimo atvejui **SURASTI FAILĄ**



23 pav. Surasti failą

Panaudojimo atvejo specifikacija:

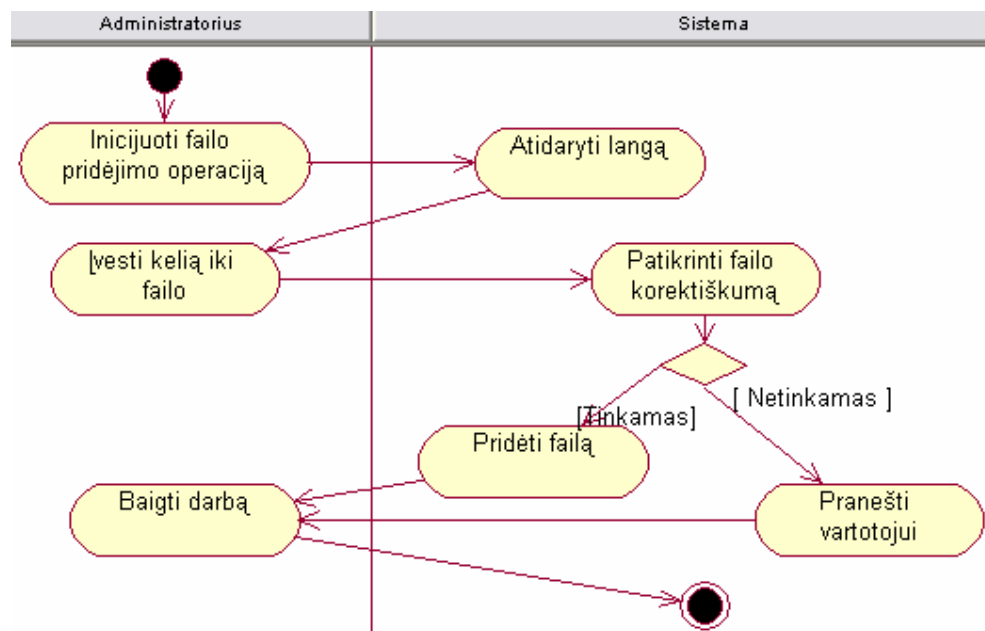
13 lentelė

PA Surasti failą specifikacija

Panaudojimo atvejis	Surasti Failą
Nr.	N4
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	Toks failas turi būti įvestas į failų rinkinį
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedamas ieškomo failo pavadinimas	1.1. Tirkinama ar toks failas yra rinkinyje 1.2. Failas paryškinamas rinkinyje
Po-sąlyga	Failas surastas
Alternatyvos	1.2.a. Tokio failo nėra, todėl apie tai pranešama administratoriui
Vykdomo variantai	Administratorius suranda ieškomą failą
Veiklos taisyklės	Failas bus surastas, jei prieš tai buvo įvestas į failų rinkinį
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Įeina į pagrindinį PA “Dirbti su failais” (ryšio tipas <<include>>)
Sudarė	A. Kazlauskienė

Sudarymo data	2005.05.03
---------------	------------

Veiklos diagrama panaudojimo atvejui **PRIDĖTI FAILĄ**



24 pav. Pridėti failą

Panaudojimo atvejo specifikacija:

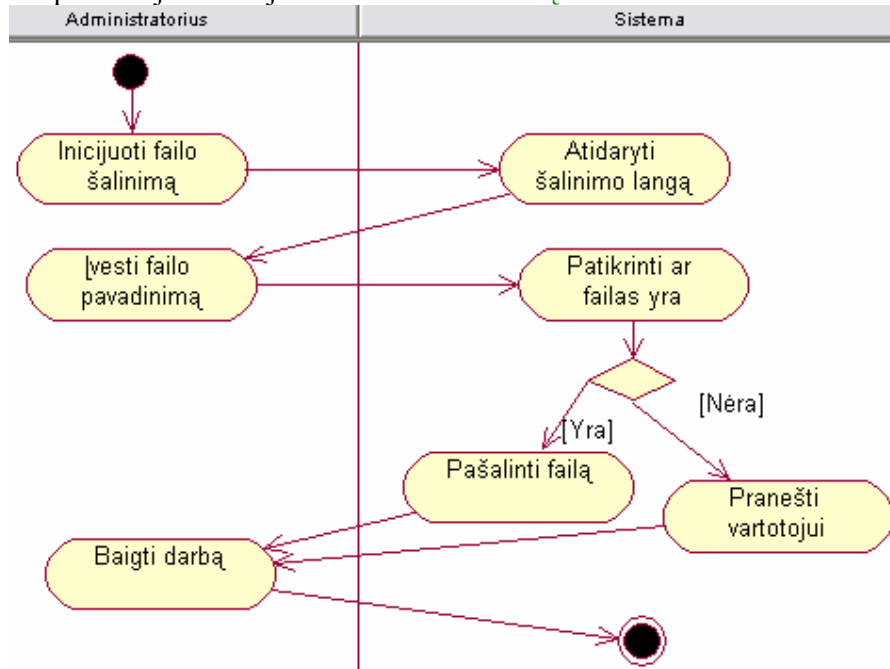
14 lentelė

PA Pridėti failą specifikacija

Panaudojimo atvejis	Pridėti Failą
Nr.	N5
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	-
Pagrindinis įvykių srautas	Sistemos reakcija
1. Nurodomas kelias iki pridedamo failo	1.1. Tikrinama ar failas yra korektiškas 1.2. Failas pridedamas
Po-sąlyga	Failas pridėtas
Alternatyvos	1.2.a. Failas nekorektiškas, todėl apie tai pranešama administratoriui
Vykdymo variantai	Administratorius prideda failą
Veiklos taisyklės	Failas bus pridėtas, jei jo struktūra yra korektiška
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Įeina į pagrindinį PA "Dirbti su failais" (ryšio

	tipas <<include>>)
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veiklos diagrama panaudojimo atvejui **PAŠALINTI FAILĄ**



25 pav. Pašalinti failą

Panaudojimo atvejo specifikacija:

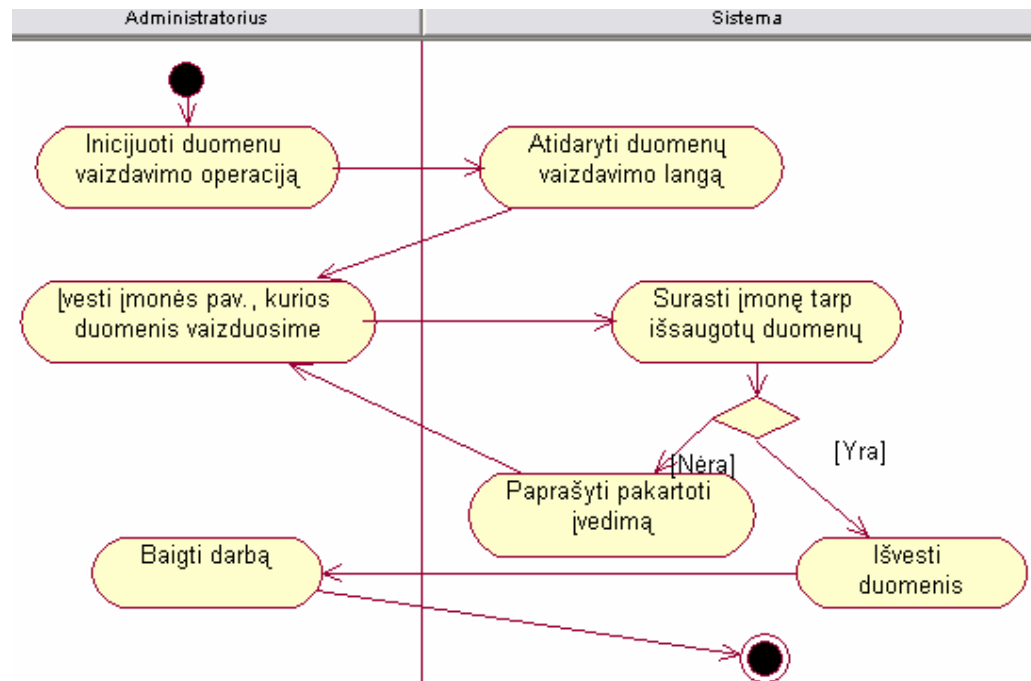
15 lentelė

PA Pašalinti failą specifikacija

Panaudojimo atvejis	Pašalinti Failą
Nr.	N6
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	-
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedamas šalinamo failo pavadinimas	1.1. Tikrinama ar toks failas yra 1.2. Failas pašalinamas
Po-sąlyga	Failas pašalintas
Alternatyvos	1.2.a. Failo nėra, todėl apie tai pranešama administratoriui
Vykdymo variantai	Administratorius pašalina failą iš rinkinio
Veiklos taisyklės	Failas bus pašalintas, jei prieš tai buvo įvestas
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra

Ryšiai su kitais PA	1. Įeina į pagrindinį PA “Dirbti su failais” (ryšio tipas <<include>>)
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veiklos diagrama panaudojimo atvejui **RODYTI METADUOMENIS**



26 pav. Rodyti metaduomenis

Panaudojimo atvejo specifikacija:

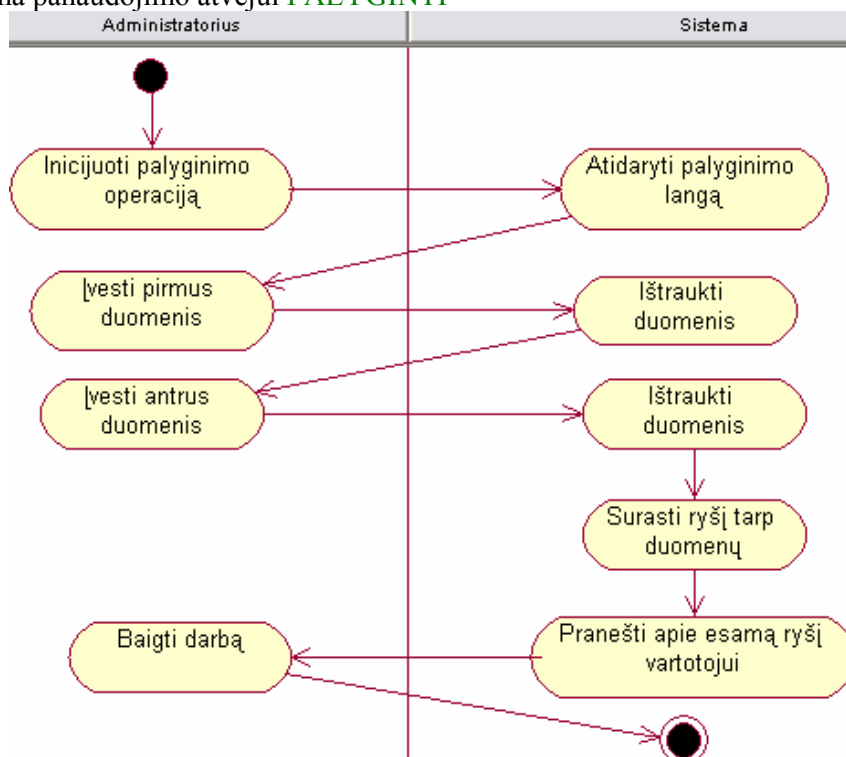
16 lentelė

PA Rodyti metaduomenis specifikacija

Panaudojimo atvejis	Rodyti metaduomenis
Nr.	N7
Aktoriai	Administratorius, Vartotojas
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	Metaduomenys turi būti ištraukti
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedamas įmonės, kurios metaduomenų ieškome pavadinimas	1.1. Tikrinama ar tokios įmonės metaduomenys yra ištraukti ir išsaugoti 1.2. Duomenys išvedami
Po-sąlyga	Metaduomenys išvesti
Alternatyvos	1.2.a. Duomenų nėra, todėl apie tai pranešama vartotojui
Vykdymo variantai	Administratorius arba vartotojas gali peržiūrėti metaduomenis apie dominančią įmonę
Veiklos taisyklės	Metaduomenys bus išvesti, jei prieš tai buvo ištraukti ir išsaugoti

Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	-
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veiklos diagrama panaudojimo atvejui **PALYGINTI**



27 pav. Palyginti

Panaudojimo atvejo specifikacija:

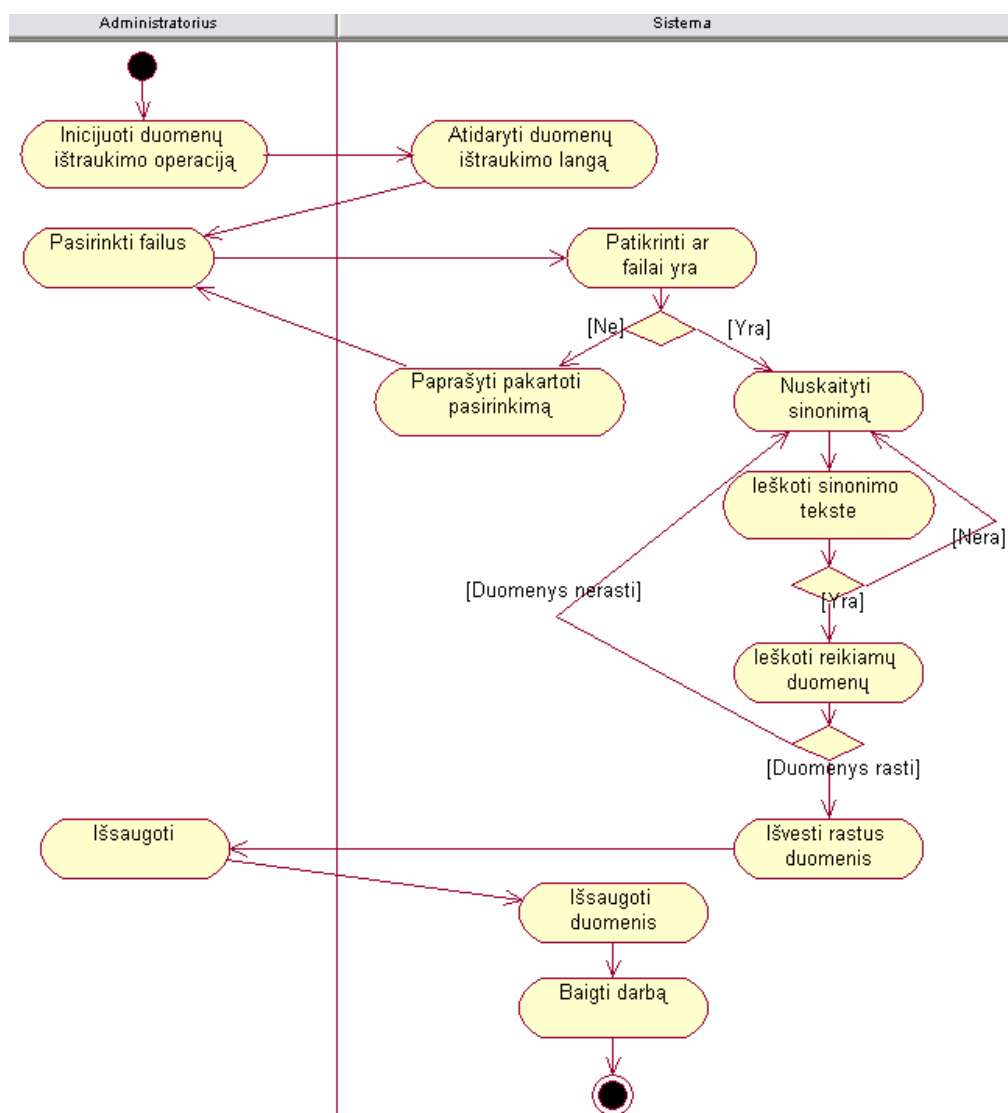
17 lentelė

PA Palyginti specifikacija

Panaudojimo atvejis	Palyginti
Nr.	N8
Aktoriai	Administratorius, Vartotojas
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	Metaduomenys turi būti ištraukti ir išsaugoti
Pagrindinis įvykių srautas	Sistemos reakcija
1. Įvedami pirmi duomenys 2. Įvedami antri duomenys	1.1. Tikrinamas duomenų korektiškumas 2.1. Tikrinamas duomenų korektiškumas 2.2. Ištraukiami duomenys iš duomenų bazės 2.3. Surandamas ryšys tarp duomenų 2.4. Vartotojui pranešama apie esantį ryšį
Po-sąlyga	Duomenys sulyginti

Alternatyvos	2.1.a. Duomenys nekorektiški, paprašoma pakartoti įvedimą 2.2.a. Duomenys nekorektiški, paprašoma pakartoti įvedimą
Vykdymo variantai	Administratorius arba vartotojas gali rasti ryšį tarp dviejų lyginamų žodžių
Veiklos taisyklės	Duomenys bus sulyginti, jei prieš tai buvo ištraukti ir išsaugoti
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	Apima PA „Ištraukti duomenis“
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

Veiklos diagrama panaudojimo atvejui **IŠTRAUKTI DUOMENIS**



28 pav. Metaduomenų ištraukimas

Panaudojimo atvejo specifikacija:

18 lentelė

PA Ištraukti duomenis specifikacija

Panaudojimo atvejis	Ištraukti duomenis
Nr.	N9
Aktoriai	Administratorius
Sistema	Semantinių metaduomenų apdorojimo sistema (SMAS)
Prieš-sąlyga	-
Pagrindinis įvykių srautas	Sistemos reakcija
1. Pasirenkami failai duomenų ištraukimui	1.1. Tikrinama ar toks failas yra 1.2. Nuskaitomas sinonimas 1.3. Ieškoma sinonimo tekste 1.4. Ieškomi reikiami duomenys 1.5. Ištraukti metaduomenys išvedami 1.6. Duomenys išsaugomi
Po-sąlyga	Iš duotų failų ištraukti semantiniai metaduomenys
Alternatyvos	1.2.a. Failo nėra, todėl prašoma pakartoti pasirinkimą 1.4.a. Sinonimo tekste nėra, todėl kartojamas 1.2. žingsnis su kitu sinonimu 1.5.a. Tekste reikiamų duomenų nėra, todėl kartojamas 1.2. žingsnis su kitu sinonimu
Vykdyto variantai	Administratorius atlieka semantinių metaduomenų ištraukimo operaciją
Veiklos taisyklės	Metaduomenys bus ištraukti, jei pateikti korektiški failai ir juose randami reikiami duomenys
Specialūs (nefunkciniai) reikalavimai	Patogi sąsaja su vartotoju
Kitos sistemos su kuriomis sąveikauja sistema	Nėra
Ryšiai su kitais PA	1. Į šį panaudojimo atvejį įeina PA: Pasirinkti failus, Ištraukti metaduomenis, Išsaugoti
Sudarė	A. Kazlauskienė
Sudarymo data	2005.05.03

4. Projektas

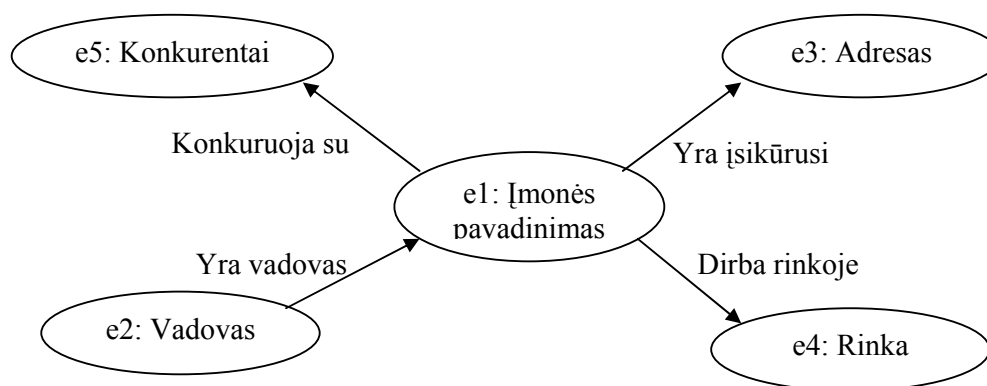
4.1. Projekto tikslas

Projektuojama sistema pagal užduotus raktažodžius iš pateiktų straipsnių ištraukia sintaksinę ir semantinę informaciją, tiesiogiai ir netiesiogiai susijusią su ieškoma sritimi. Taip pat ištraukiami duomenys, kurių vartotojas neprašė rasti, bet jie yra labai svarbūs (rastų duomenų pasikartojimas tekste duomenų patikimumui įvertinti).

Sistema pritaikyta konkrečiai darbo sričiai – verslo įmonių informacijos apdorojimui, tačiau lengvai gali būti perorientuota kitokiam darbui.

Projektuojama sistema iš pateiktų failų ištraukia informaciją susijusią su verslu: įmonės pavadinimą, vadovo vardą bei pavardę, įmonės adresą, rinkos sritį ir konkurentų sąrašą. Pakeitus sinonimų žodyno sudėtį, sistemą galima pritaikyti bet kokiems poreikiams.

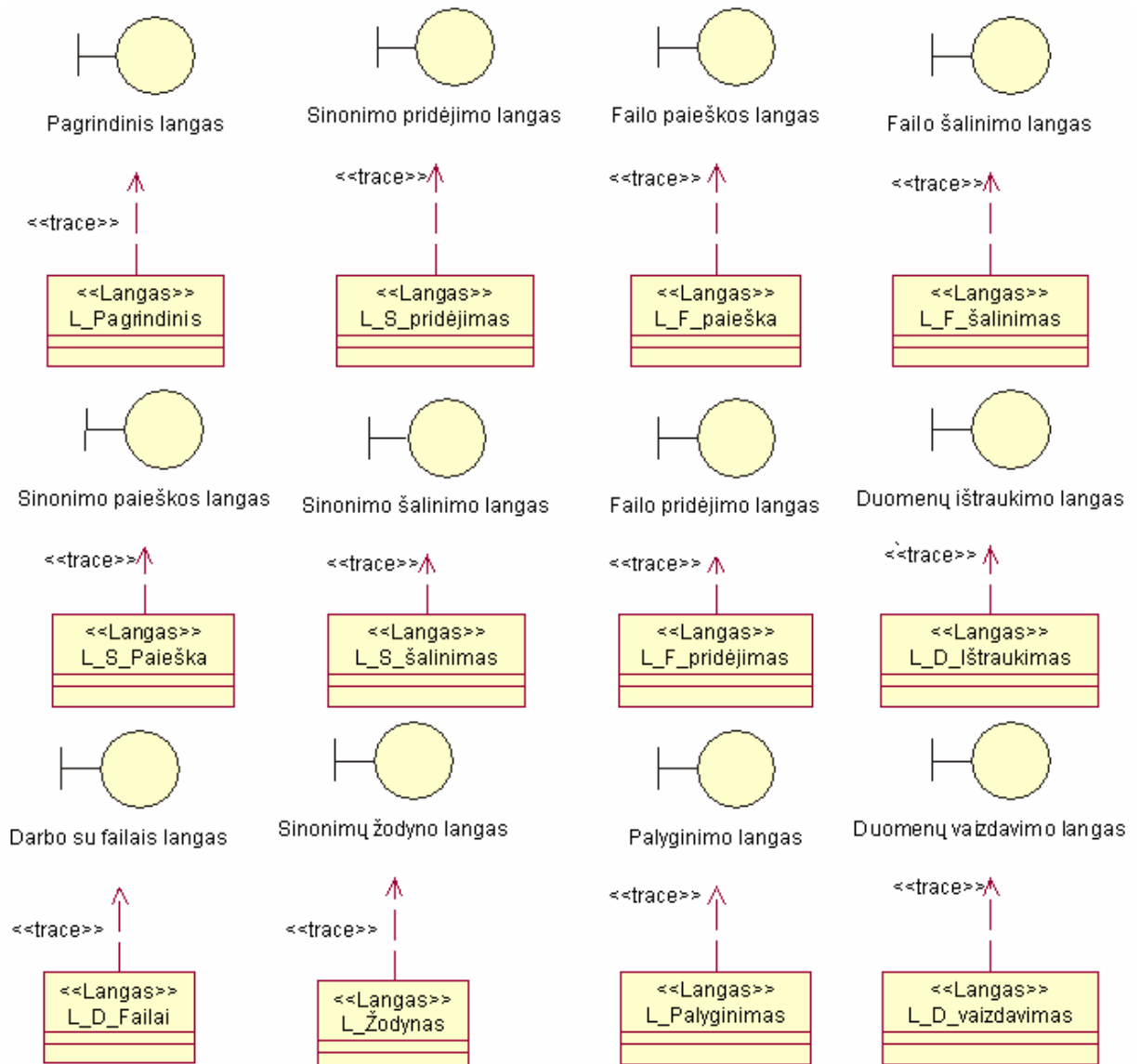
Ištraukiamų metaduomenų RDF schema:



29 pav. RDF schema

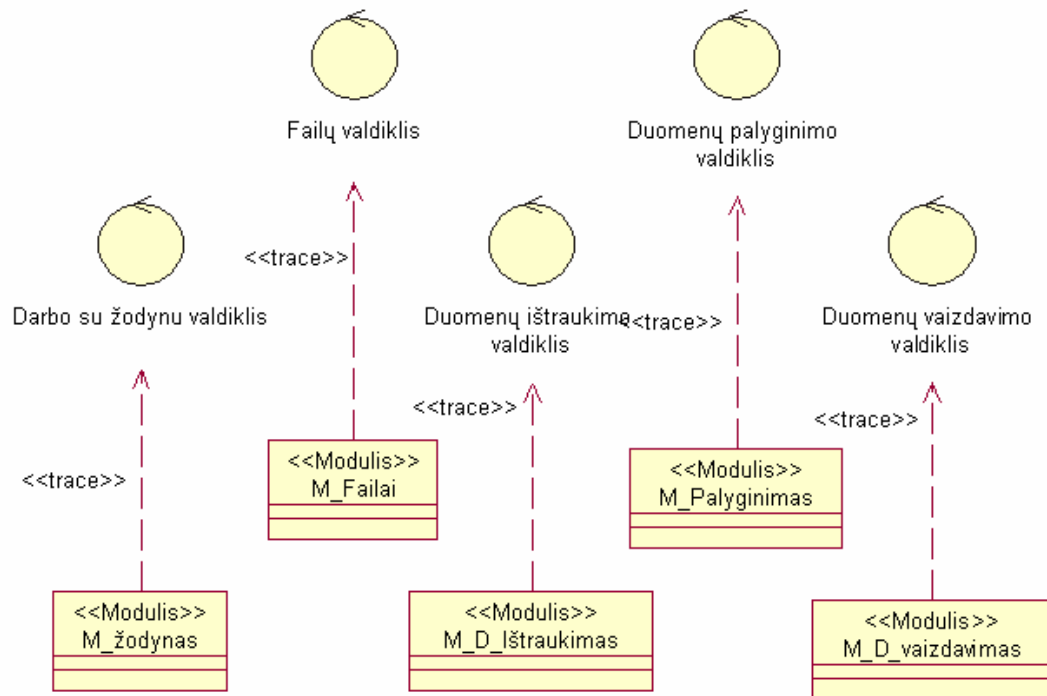
4.2. Panaudojimo atvejų realizacijos ir analizės klasių diagramos

4.2.1. Sistemos vartotojų paslaugų trasų diagrama (vartotojo interfeiso navigavimo planas)



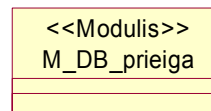
30 pav. Vartotojų paslaugų trasų diagrama

4.2.2. Veiklos paslaugų klasių trasų diagrama



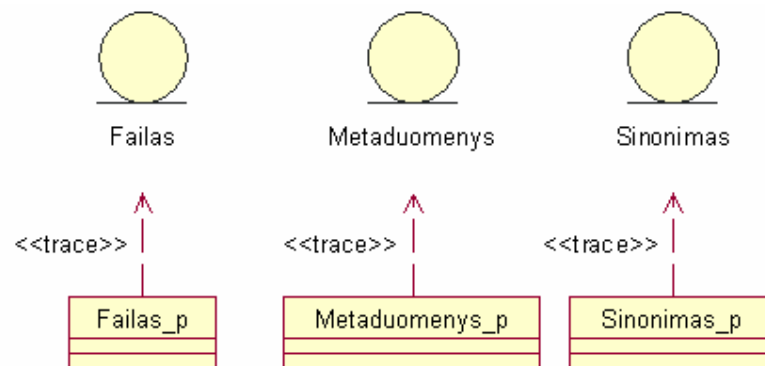
31 pav. Veiklos paslaugų klasių trasų diagrama

4.2.3. Duomenų paslaugų klasių diagrama

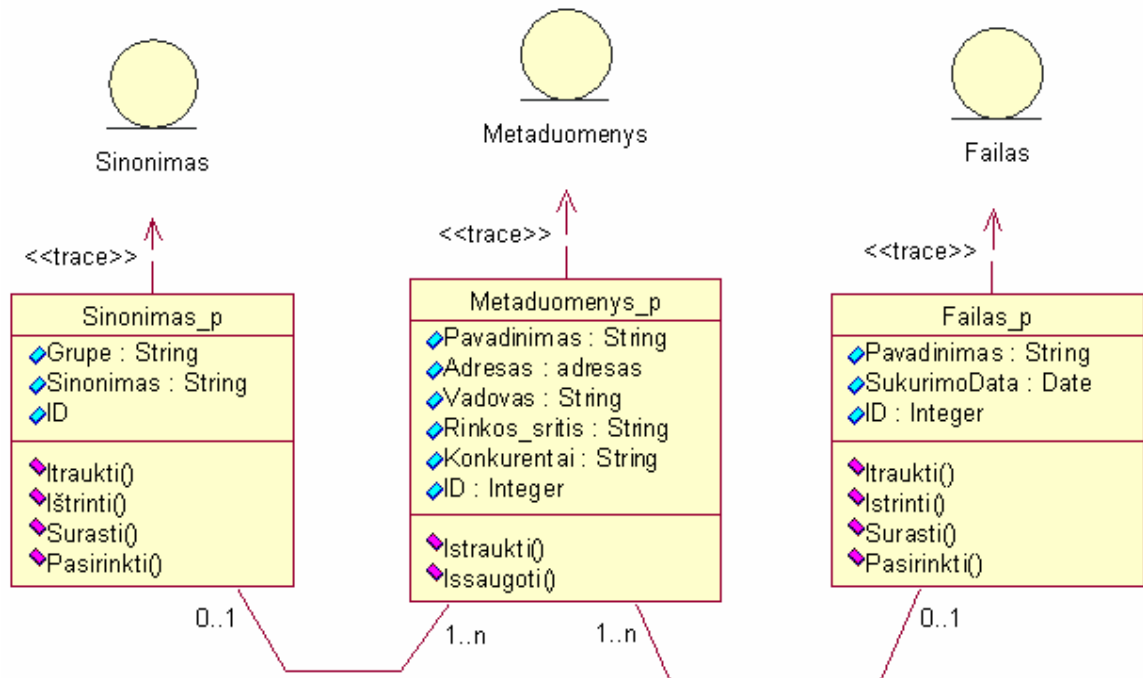


32 pav. Duomenų paslaugų klasė

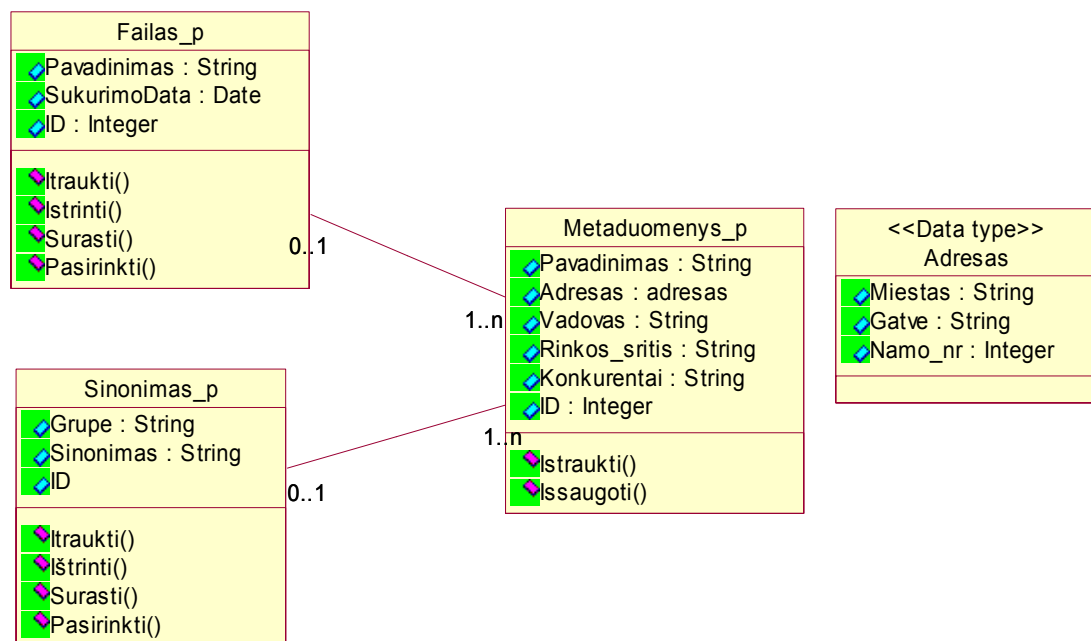
4.2.4. Dalykinės srities esybių klasių diagrama



33 pav. Dalykinės srities klasių trasų diagrama

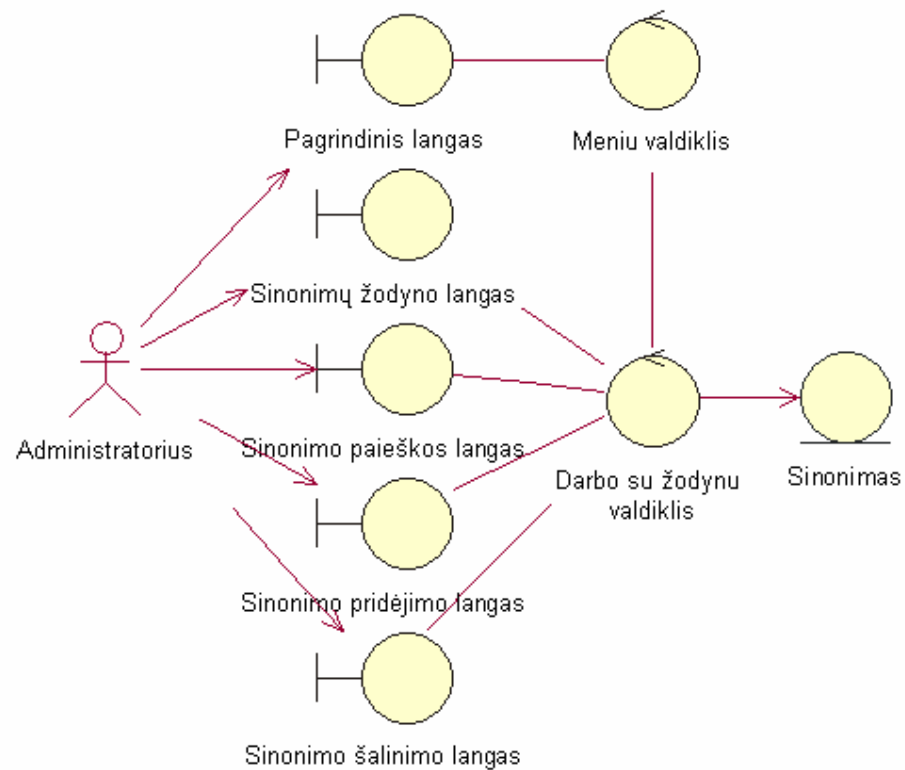


34 pav. Patikslinta dalykinės srities klasių trasų diagrama

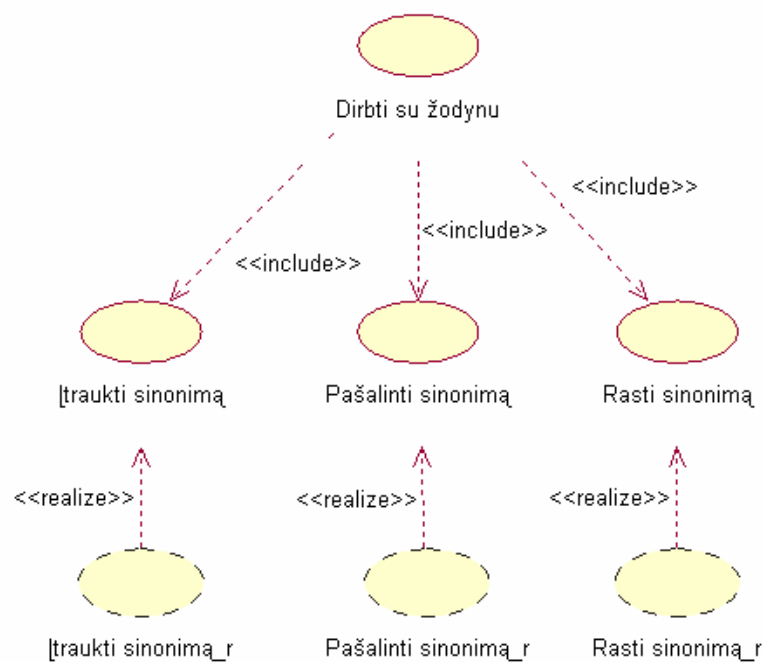


35 pav. Dalykinės srities klasių diagrama

4.2.5. /Dirbti su žodynu/ realizacija ir analizės klasės



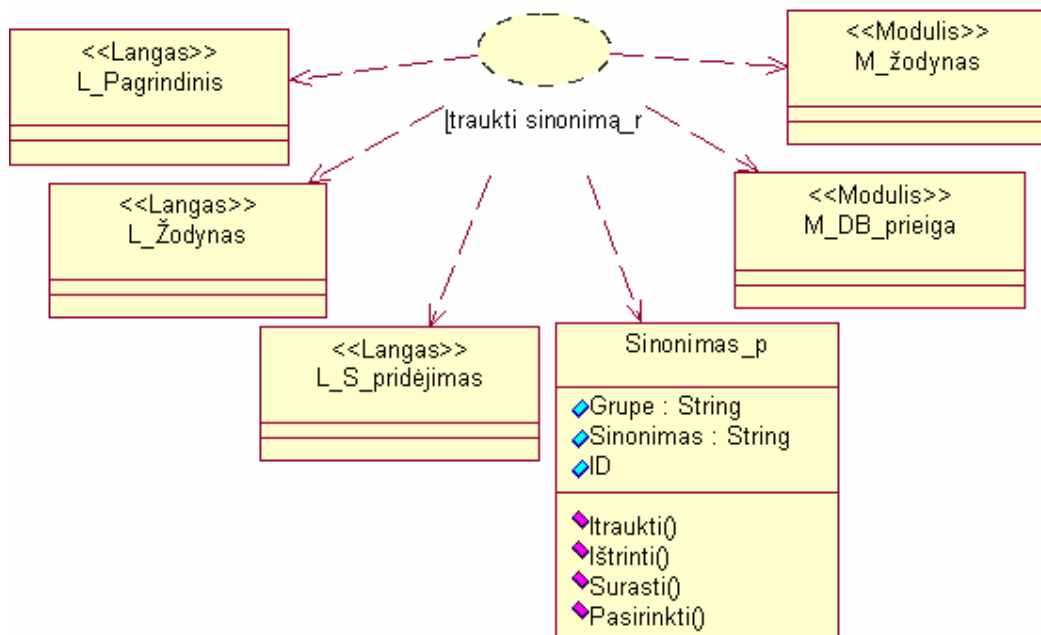
36 pav. Analizės klasių diagrama



37 pav. Trasų diagrama PA Dirbti su žodynu

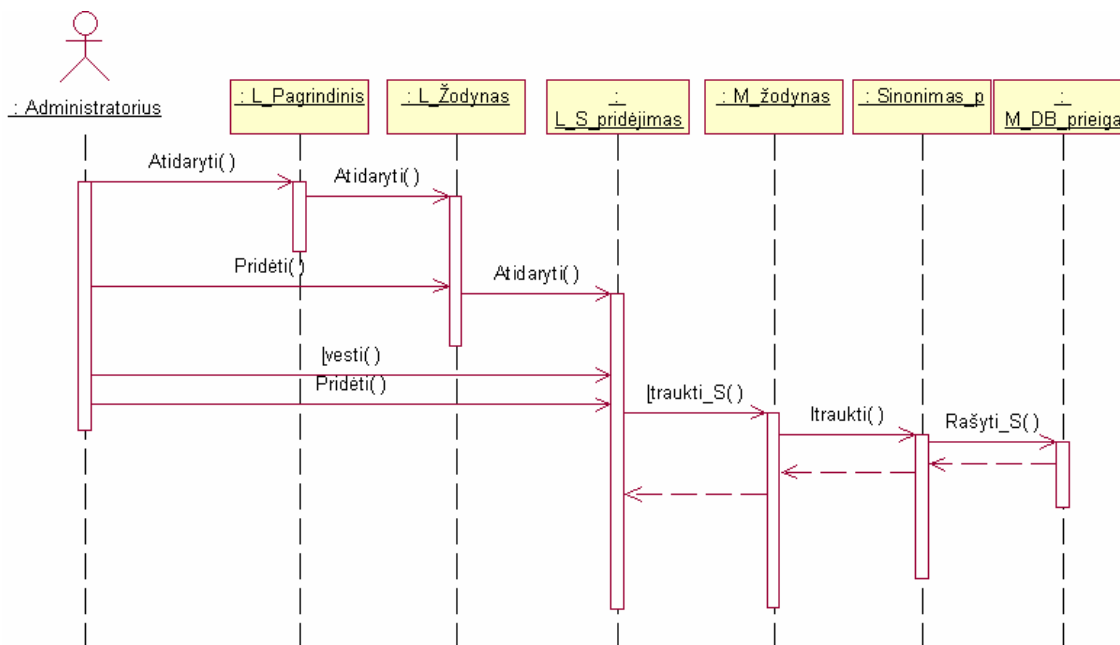
Kadangi panaudojimo atvejų “Įtraukti sinonimą”, “Rasti sinonimą” ir “Pašalinti sinonimą” sąveikos diagramos bus labai panašios, šiems trims panaudojimo atvejams pakaks vieno – sinonimų registravimo valdiklio.

Panaudojimo atvejo Įtraukti sinonimą realizacija:



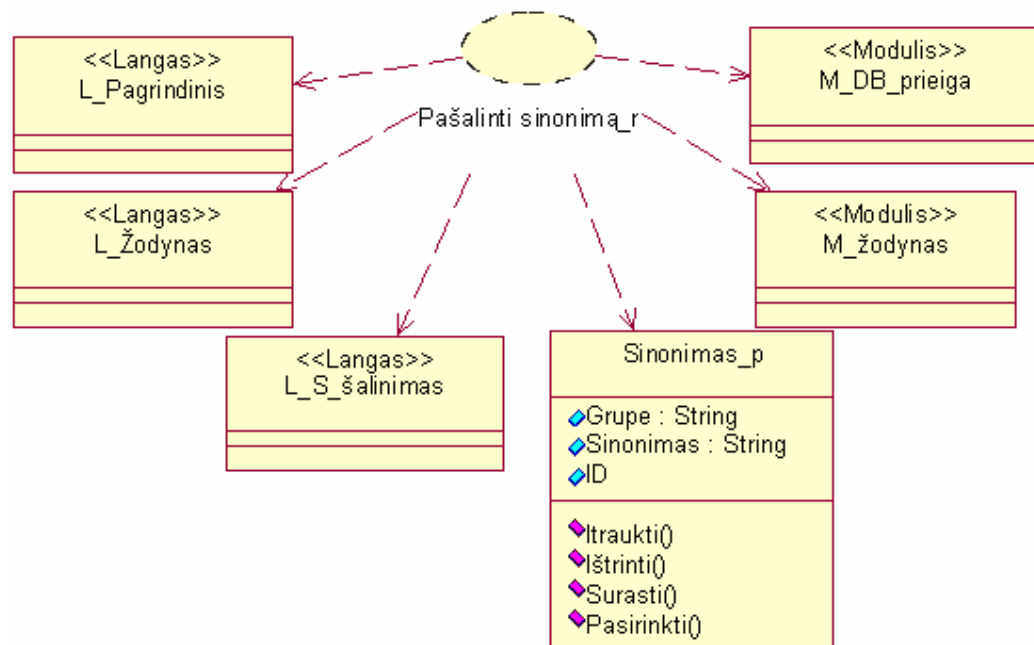
38 pav. PA Įtraukti sinonimą realizacija

Sekų diagrama panaudojimo atvejui Įtraukti sinonimą:



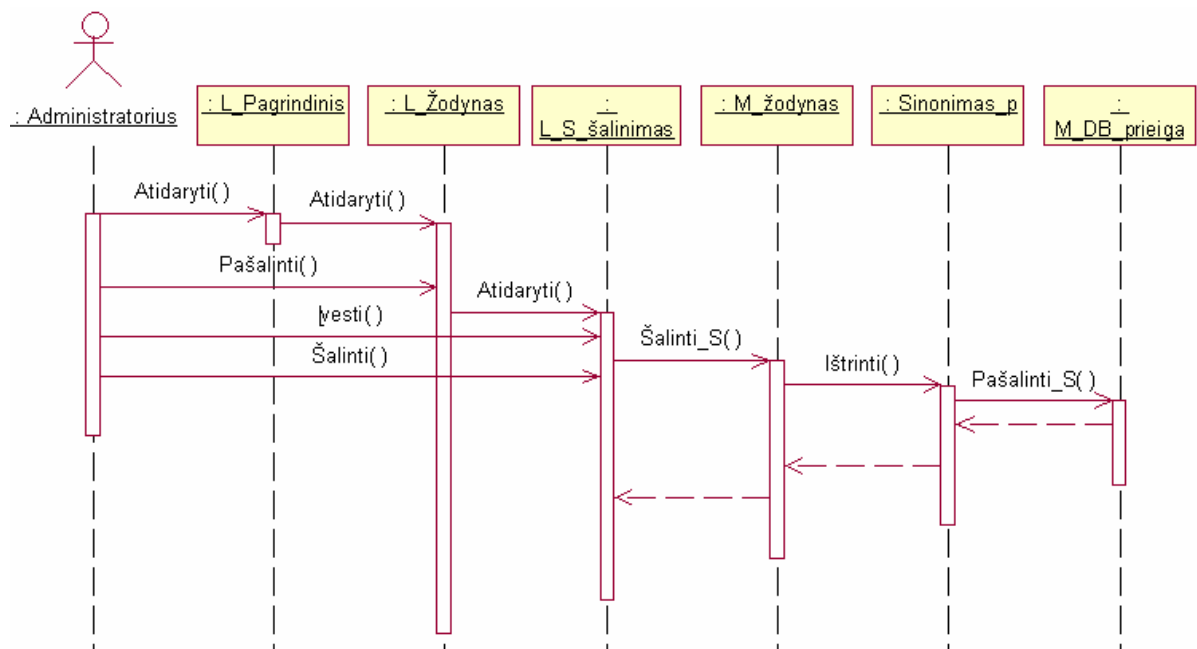
39 pav. Sekų diagrama PA Įtraukti sinonimą

Panaudojimo atvejo Pašalinti sinonimą realizacija:



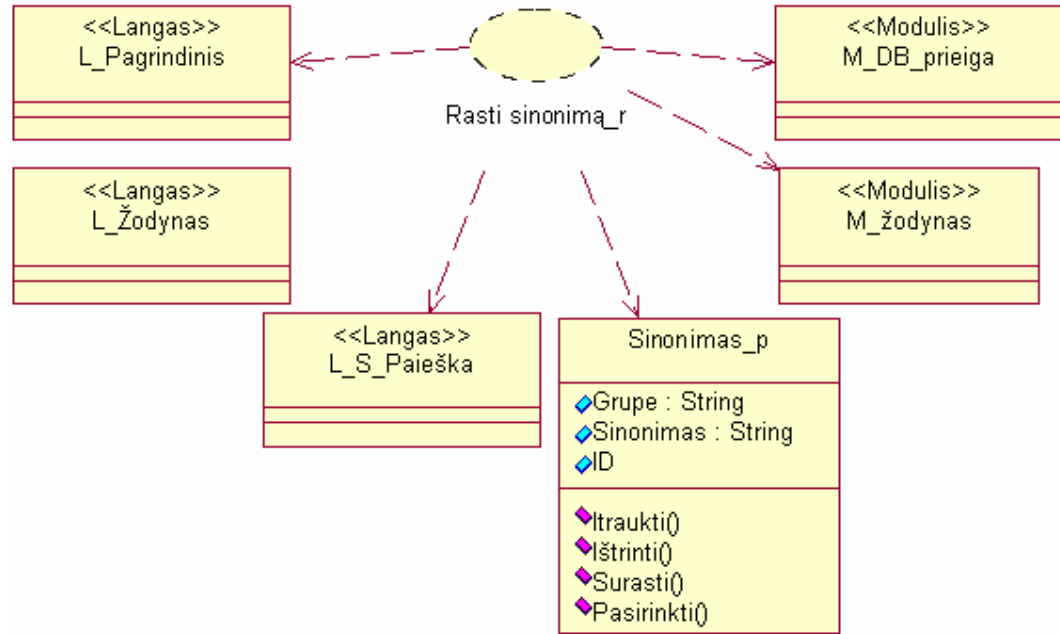
40 pav. PA Pašalinti sinonimą realizacija

Sekų diagrama panaudojimo atvejui Pašalinti sinonimą:



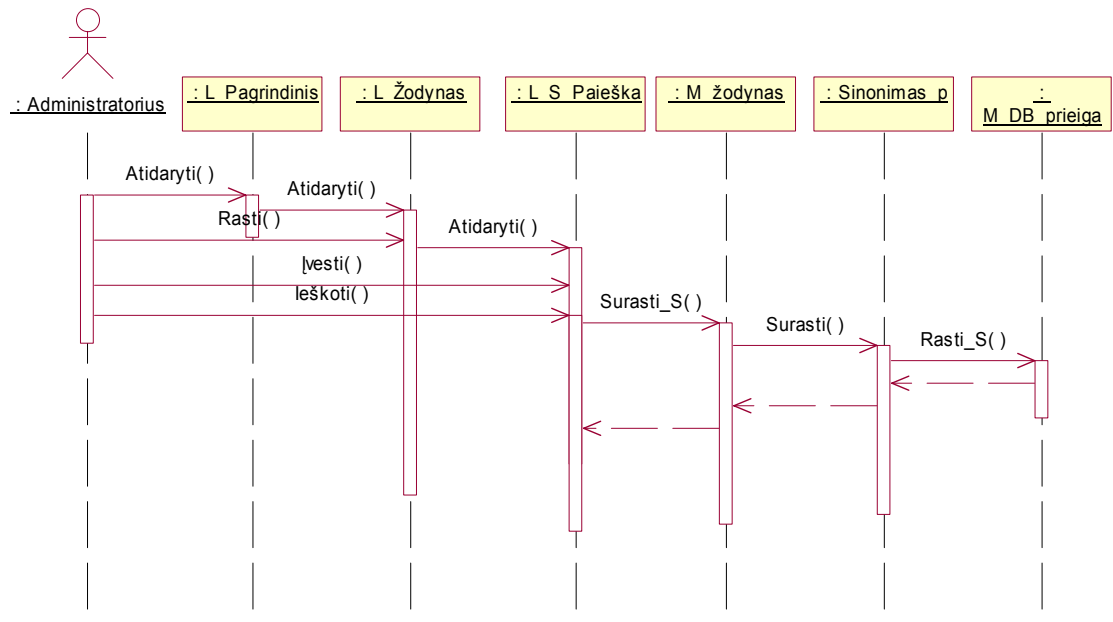
41 pav. Sekų diagrama PA Pašalinti sinonimą

Panaudojimo atvejo Surasti sinonimą realizacija:



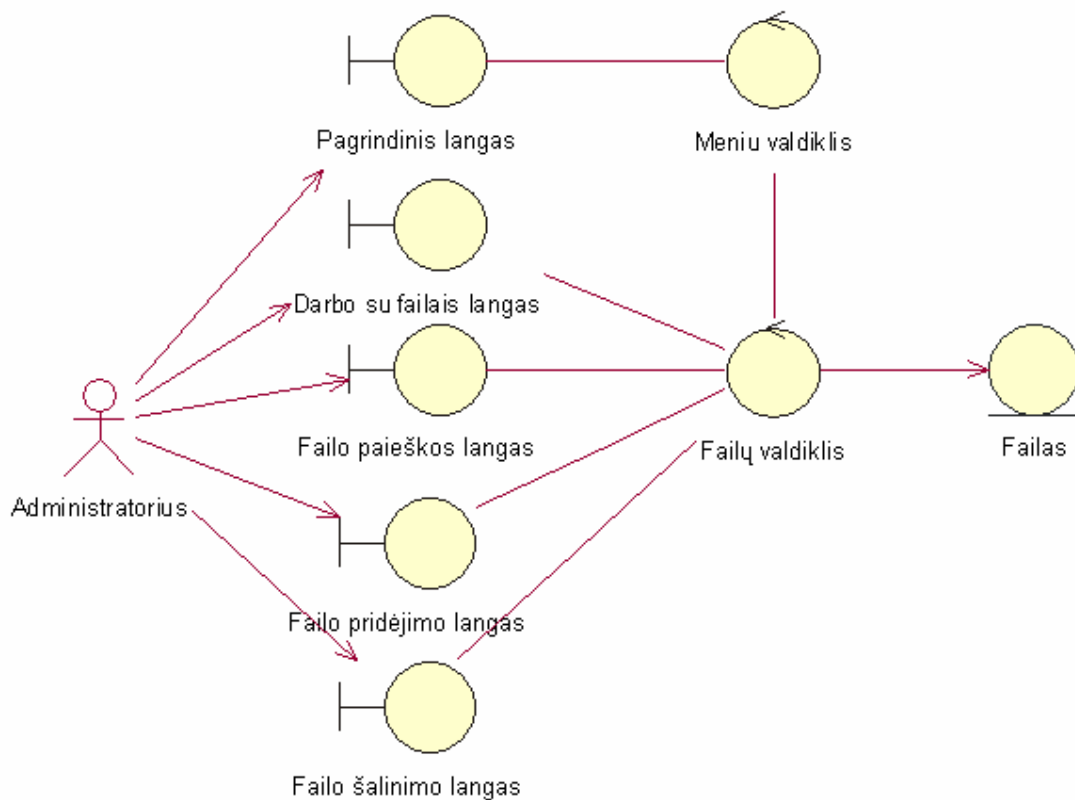
42 pav. PA Surasti sinonimą realizacija

Sekų diagrama panaudojimo atvejui Surasti sinonimą:



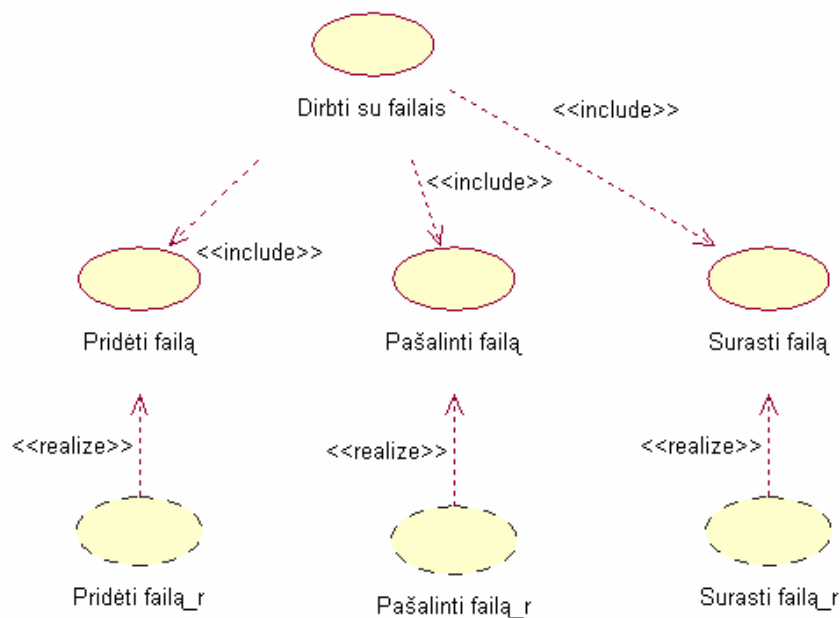
43 pav. Sekų diagrama PA Surasti sinonimą

4.2.6. /Dirbti su failais/ realizacija ir analizės klasės



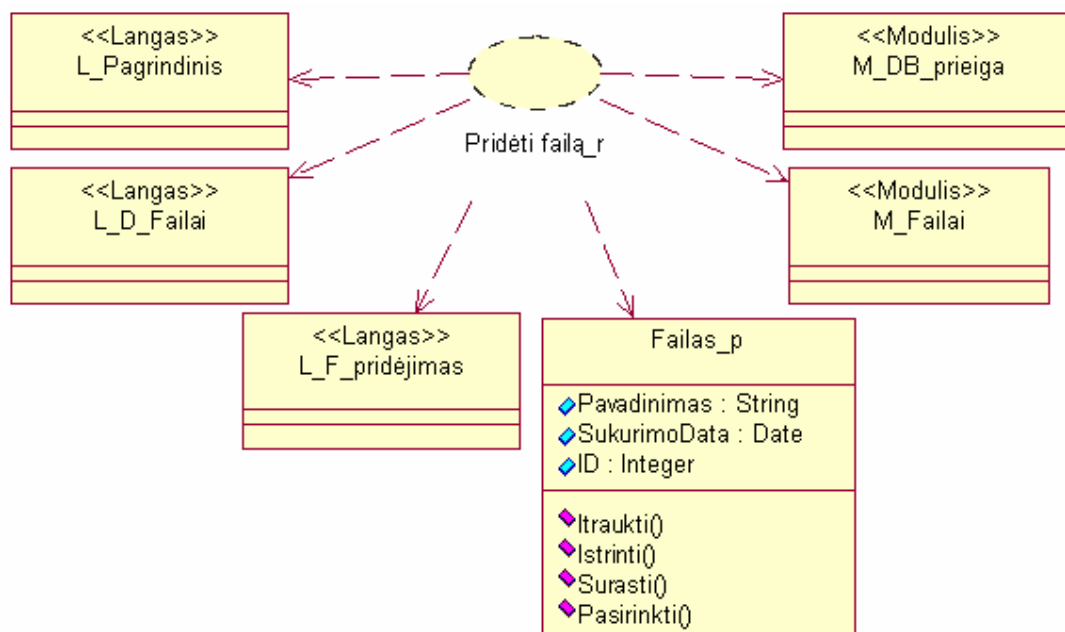
44 pav. Analizės klasių diagrama

Kadangi panaudojimo atvejų “Pridėti failą”, “Surasti failą” ir “Pašalinti failą” sąveikos diagramos bus labai panašios, šiems trims panaudojimo atvejams pakaks vieno – failų registravimo valdiklio.



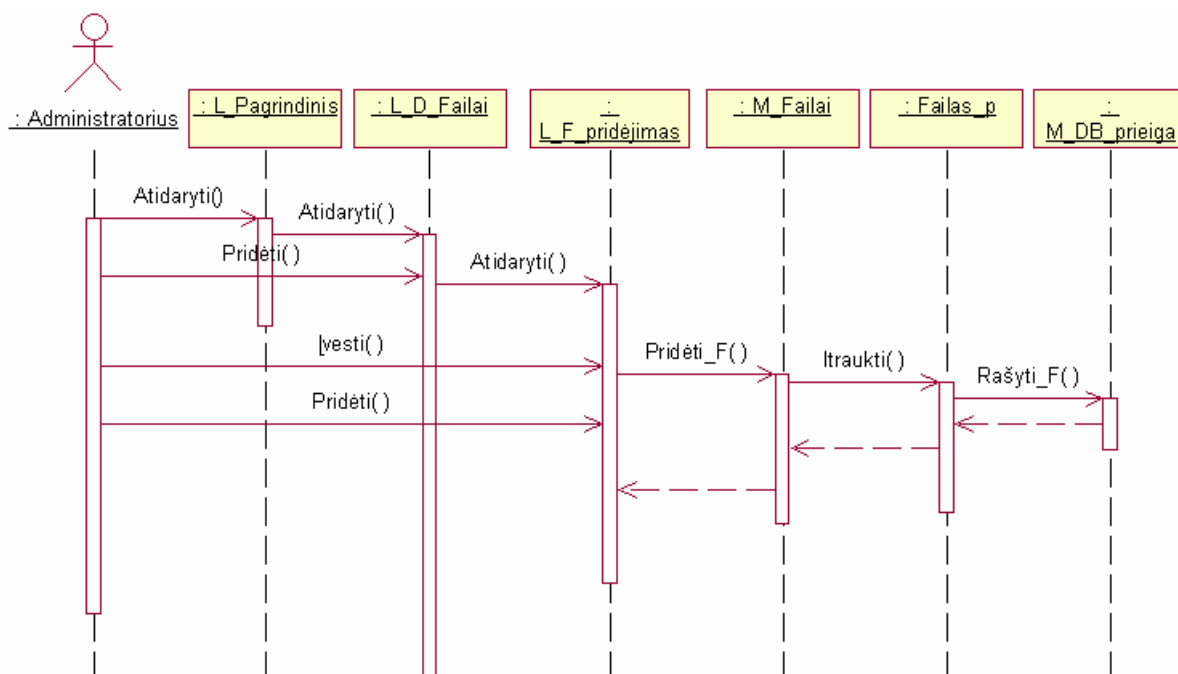
45 pav. Trasų diagrama PA Dirbti su failais

Panaudojimo atvejo Pridėti failą realizacija:



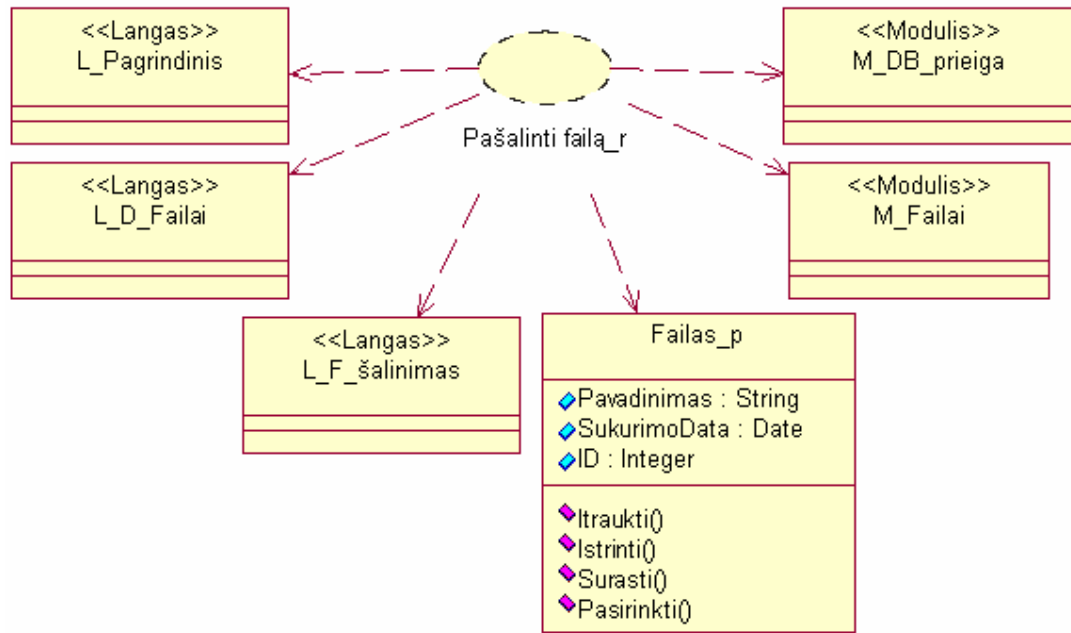
46 pav. PA Pridėti failą realizacija

Sekų diagrama panaudojimo atvejui Pridėti failą:



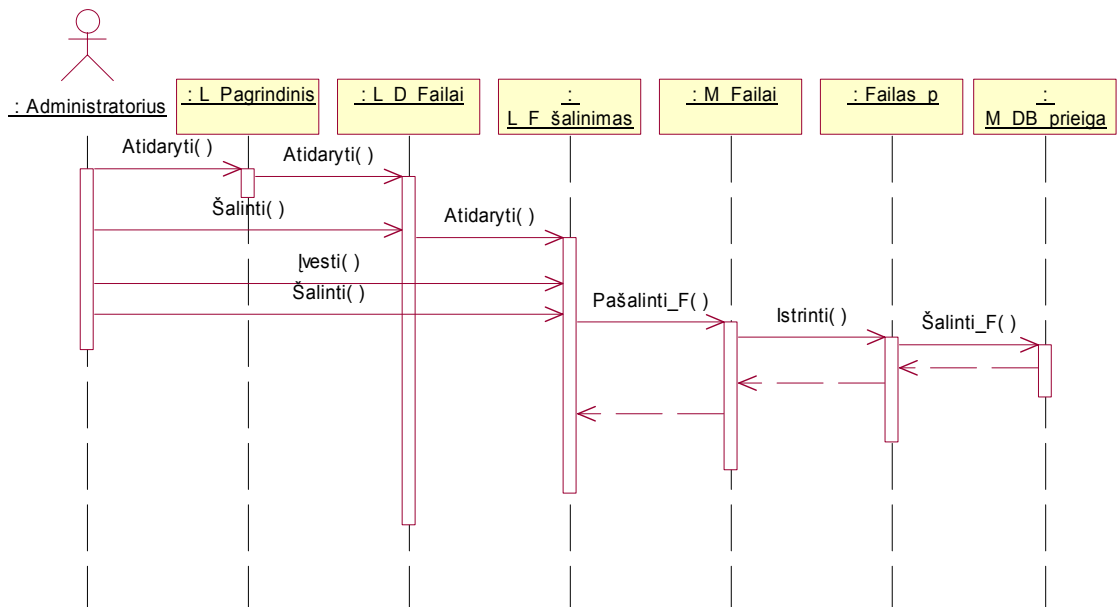
47 pav. Sekų diagrama PA Pridėti failą

Panaudojimo atvejo Pašalinti failą realizacija:



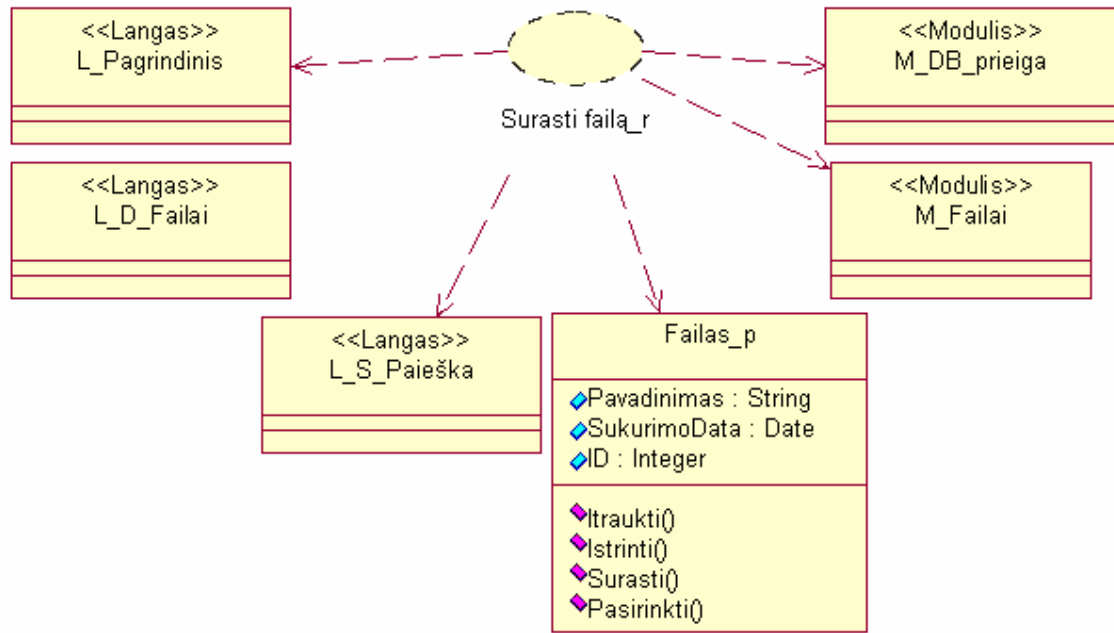
48 pav. PA Pašalinti failą realizacija

Sekų diagrama panaudojimo atvejui Pašalinti failą:



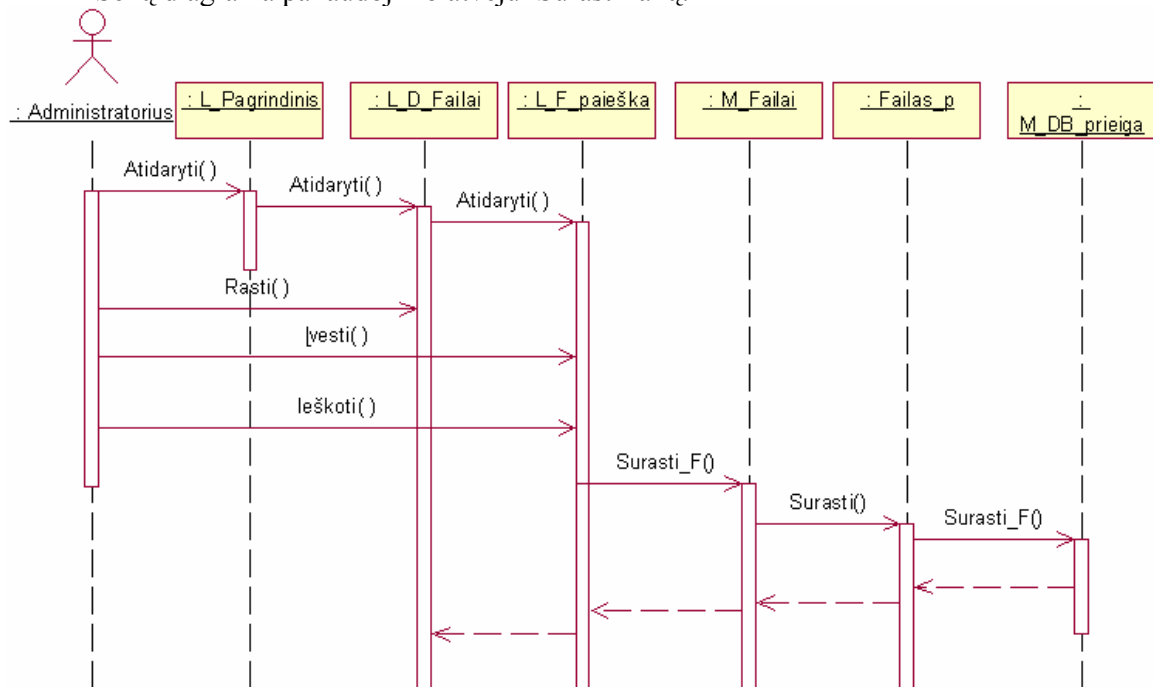
49 pav. Sekų diagrama PA Pašalinti failą

Panaudojimo atvejo Surasti failą realizacija:



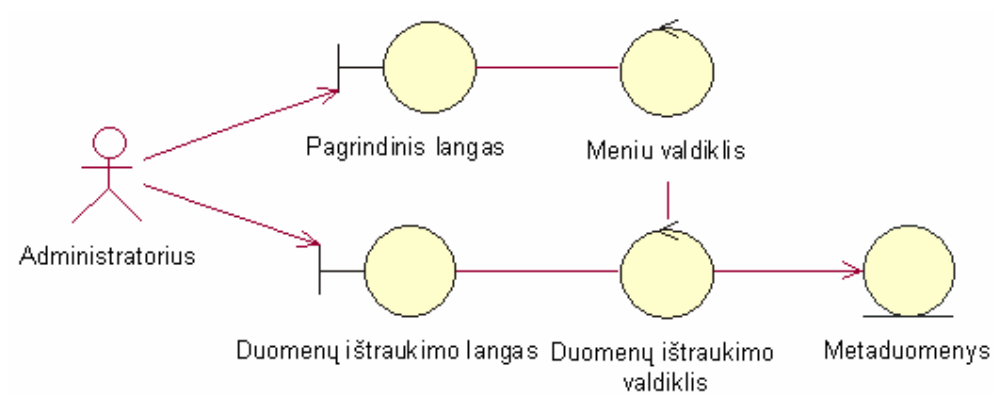
50 pav. PA Surasti failą realizacija

Sekų diagrama panaudojimo atvejui Surasti failą:

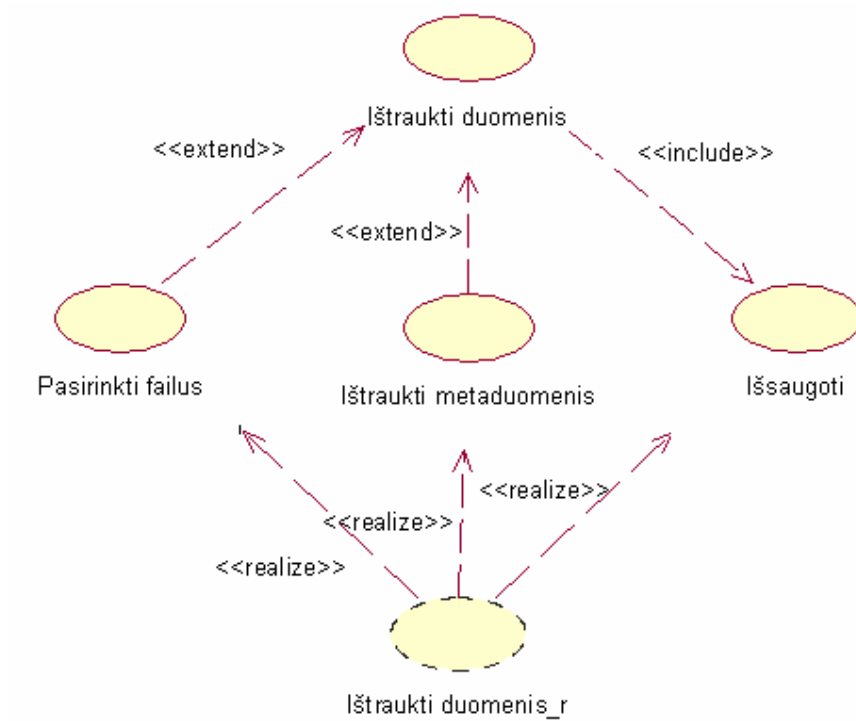


51 pav. Sekų diagrama PA Surasti failą

4.2.7. /Ištraukti duomenis/ realizacija ir analizės klasės

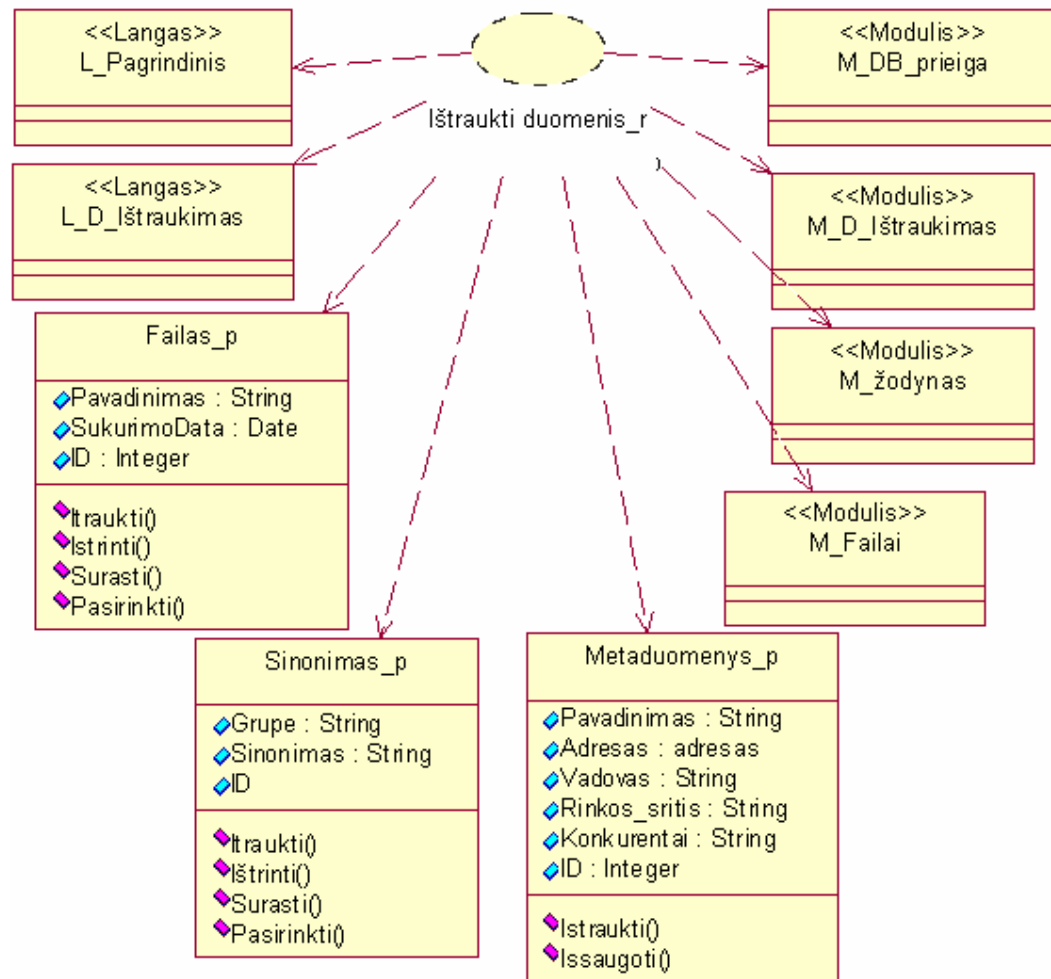


52 pav. Analizės klasių diagrama



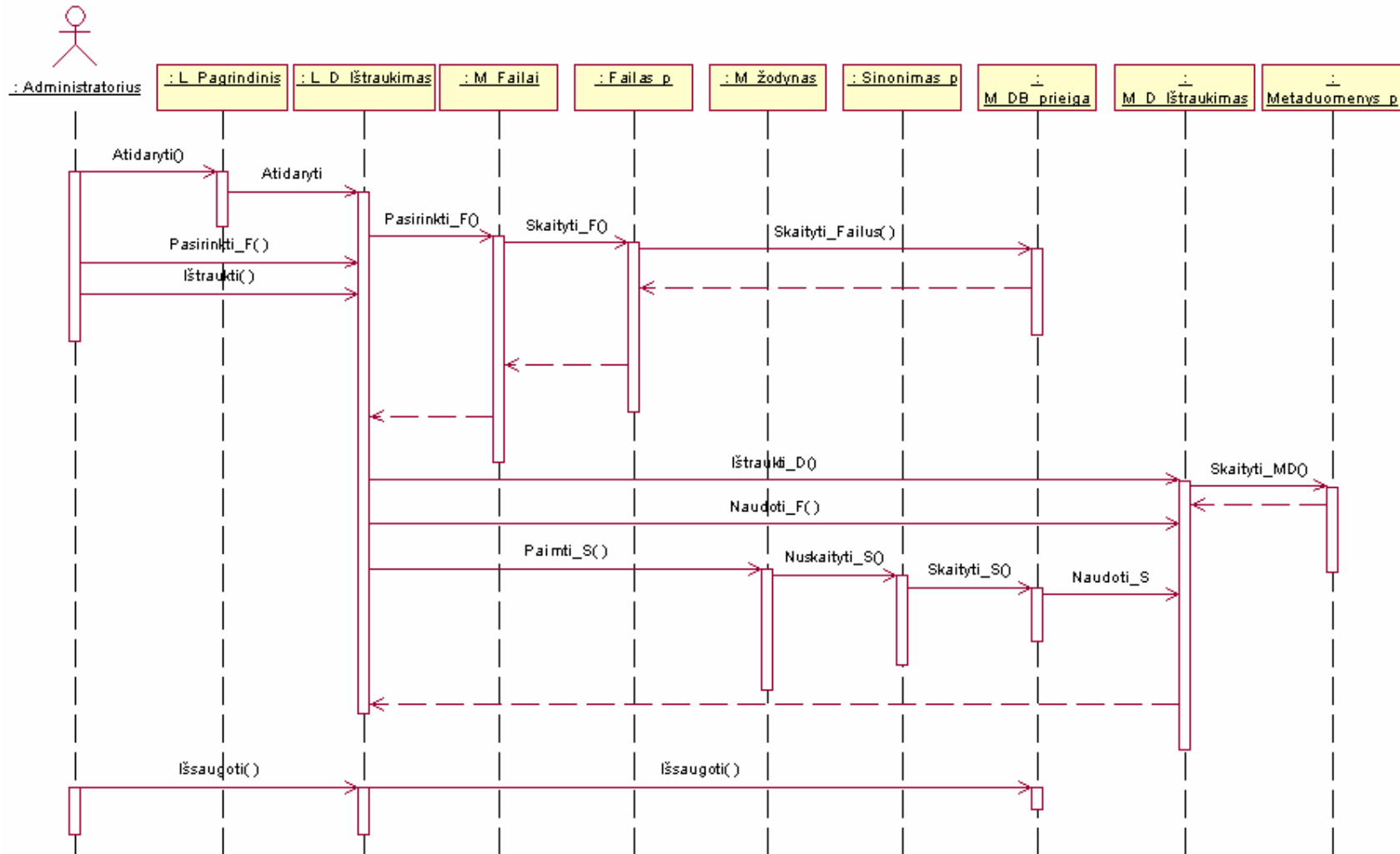
53 pav. Trasų diagrama Ištraukti duomenis

Panaudojimo atvejo Ištraukti duomenis realizacija:



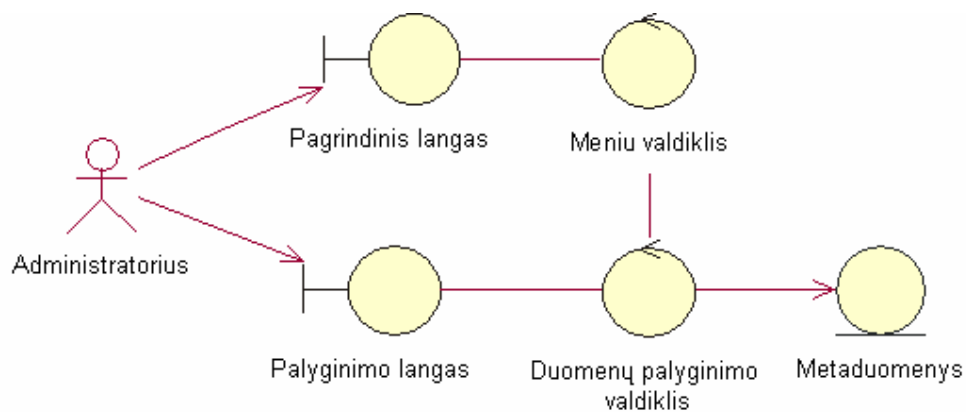
54 pav. PA Ištraukti duomenis realizacija

Sekų diagrama panaudojimo atvejui Ištraukti duomenis:

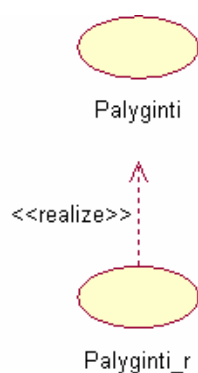


55 pav. Sekų diagrama PA Ištraukti duomenis

4.2.8. /Palyginti/ realizacija ir analizės klasės

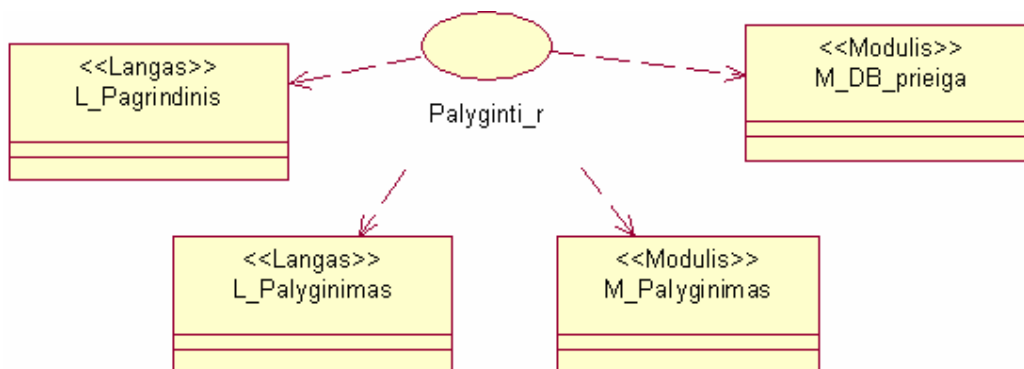


56 pav. Analizės klasių diagrama



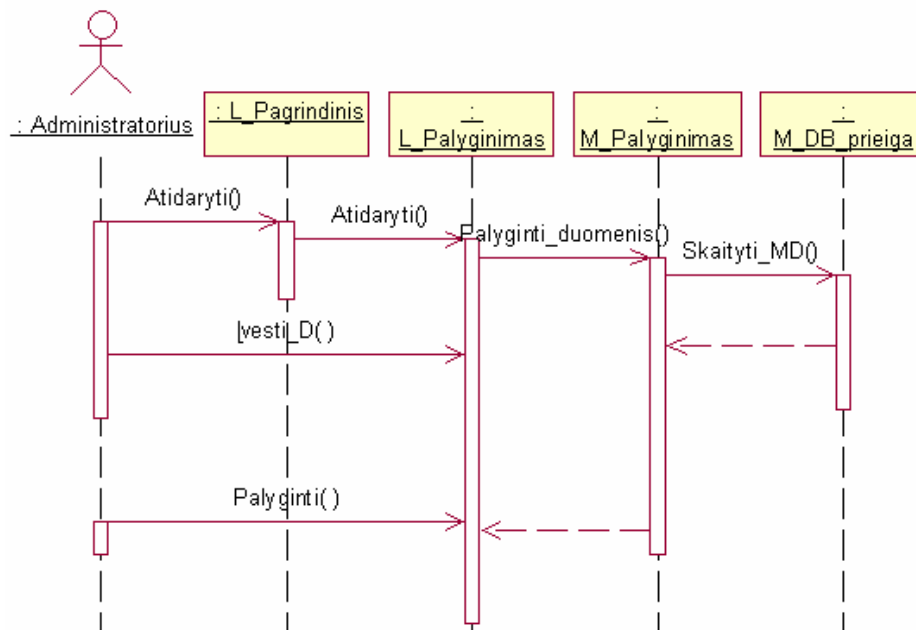
57 pav. Trasų diagrama PA Palyginti

Panaudojimo atvejo Palyginti realizacija:



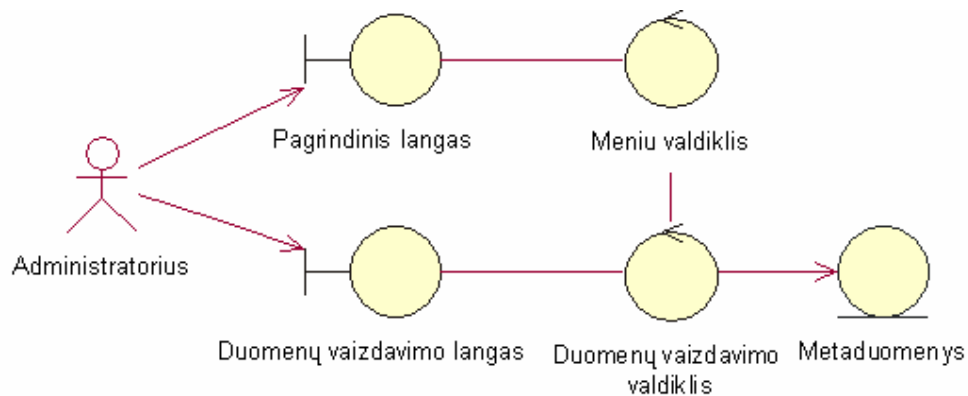
58 pav. PA Palyginti realizacija

Sekų diagrama panaudojimo atvejui Palyginti:

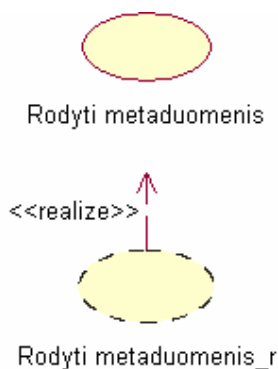


59 pav. Sekų diagrama PA Palyginti

4.2.9. /Rodyti metaduomenis/ realizacija ir analizės klasės

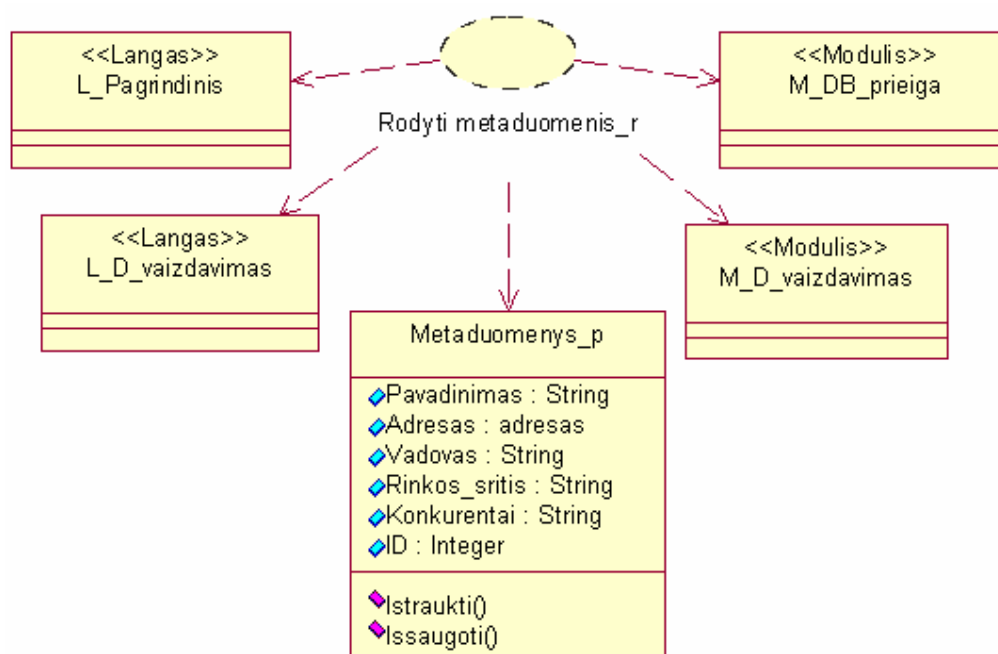


60 pav. Analizės klasių diagrama



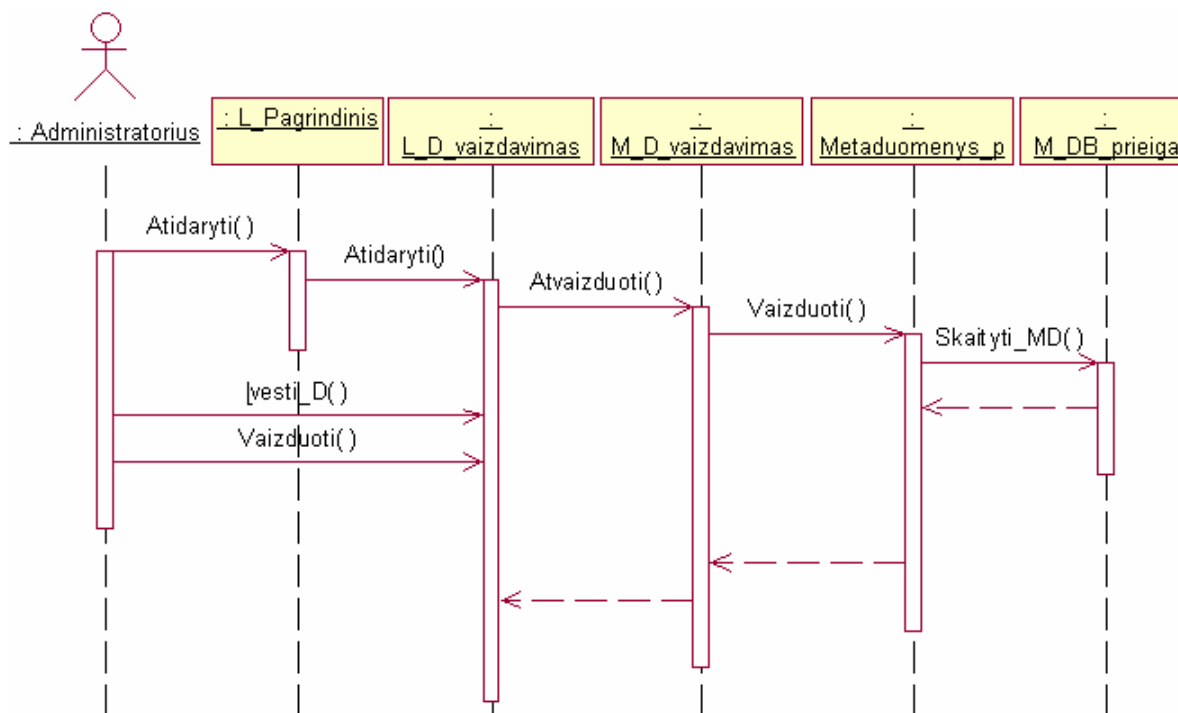
61 pav. Trasų diagrama PA Rodyti metaduomenis

Panaudojimo atvejo Rodyti metaduomenis realizacija:

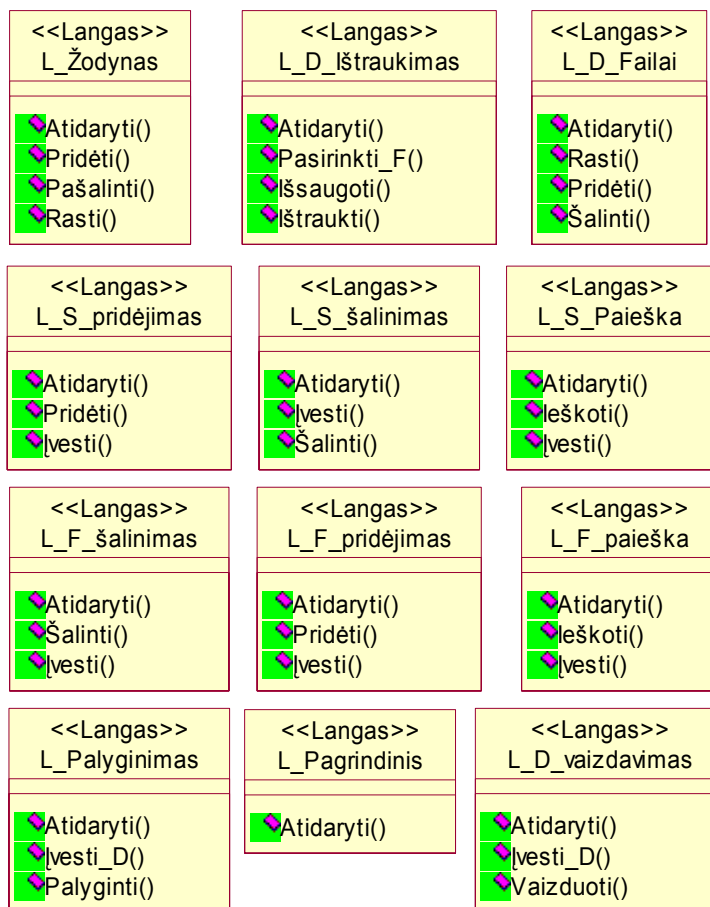


62 pav. PA Rodyti metaduomenis realizacija

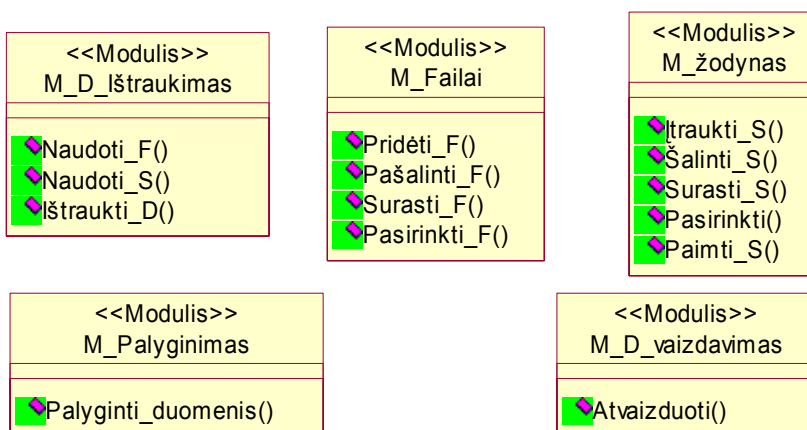
Sekų diagrama panaudojimo atvejui Rodyti metaduomenis:



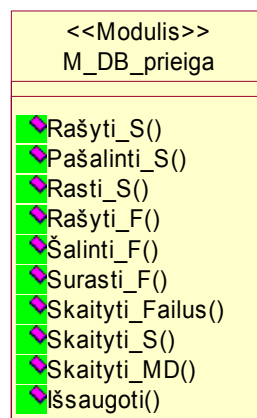
63 pav. Sekų diagrama PA Rodyti metaduomenis



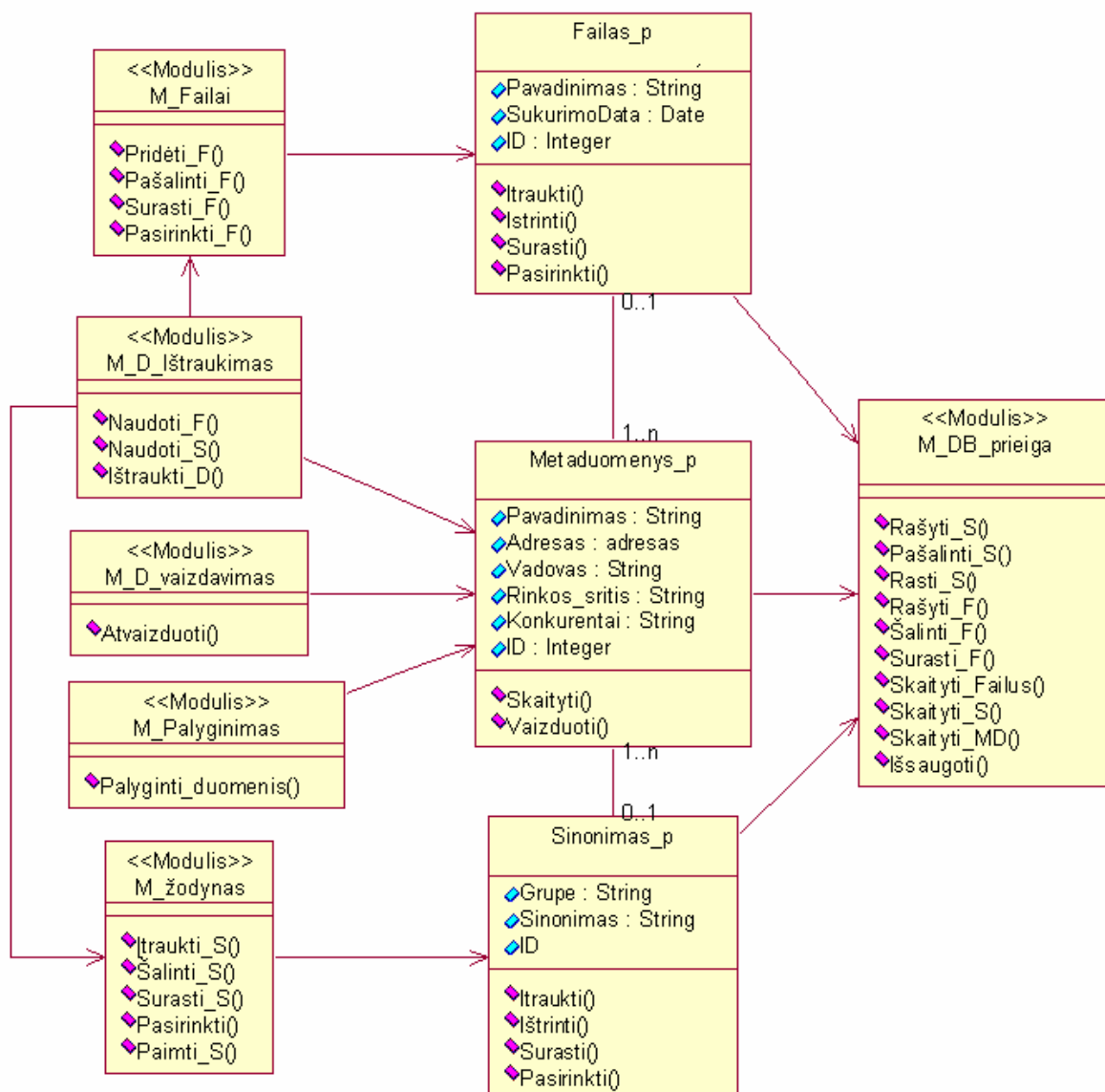
64 pav. Detali vartotojų paslaugų klasių diagrama



65 pav. Veiklos paslaugų klasių diagrama



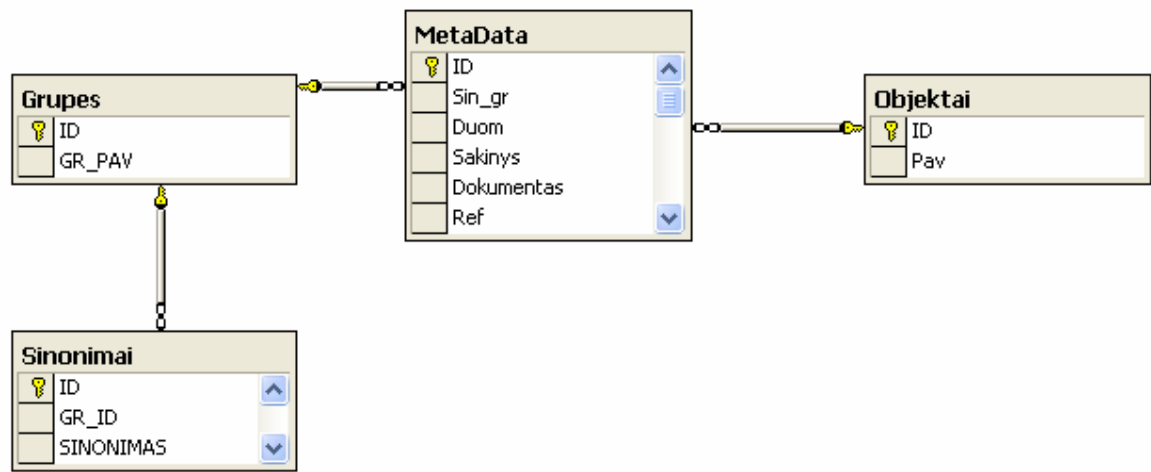
66 pav. Detali duomenų paslaugų klasių diagrama



67 pav. Detali sistemos veiklos paslaugų klasių diagrama

4.3. Duomenų bazės schema ir lentelių specifikacijos

Duomenų bazės schema:



68 pav. Duomenų bazės schema

Lentelė Objektai:

	Column Name	Data Type	Length	Allow Nulls
🔑	ID	int	4	
	Pav	varchar	100	✓

69 pav. Lentelė Objektai

Lentelė Sinonimai:

	Column Name	Data Type	Length	Allow Nulls
🔑	ID	int	4	
	GR_ID	int	4	
	SINONIMAS	varchar	50	

70 pav. Lentelė Sinonimai

Lentelė MetaData:

	Column Name	Data Type	Length	Allow Nulls
🔑	ID	int	4	
	Sin_gr	int	4	✓
	Duom	varchar	100	✓
	Sakinys	varchar	500	✓
	Dokumentas	varchar	300	✓
	Ref	int	4	✓

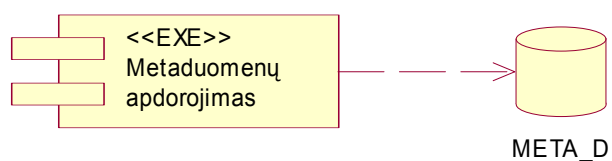
71 pav. Lentelė MetaData

Lentelė Grupės:

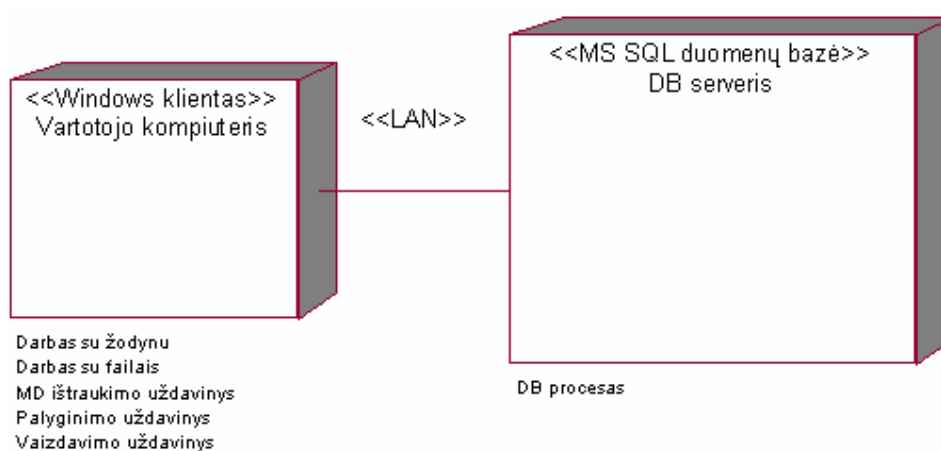
	Column Name	Data Type	Length	Allow Nulls
?	ID	int	4	
	GR_PAV	varchar	50	

72 pav. Lentelė Grupės

4.4. Realizacijos modelis (programinių komponentų architektūra, įdiegimo modelis)



73 pav. Komponentų diagrama



74 pav. Įdiegimo diagrama

5. Eksperimentinis sistemos tyrimas

Sistemos realizacijai panaudotos priemonės pateiktos 19 lentelėje.

19 lentelė

Sistemos realizacijos priemonės

Programavimo aplinka	Delphi 6 Enterprise
Duomenų bazių valdymo serveris	MySQL 4.0.15
Duomenų bazių tvarkymo priemonės	EMS MySQL Manager 2
	Borland Database Desktop

Sistemos realizacijai buvo panaudota objektiškai orientuota programavimo kalba Delphi ir Borland Delphi 6 vizualaus kūrimo aplinka. Šio pasirinkimo privalumai:

- Kalba yra objektiškai orientuota, tai palengvina tolesnius sistemos modifikavimo ir tobulinimo darbus.
- Didelis komponentų pasirinkimas užtikrina patogų ir greitą vartotojo sąsajos kūrimą.
- Galima importuoti tipų bibliotekas (*type library*) ir naudotis kitų programinių produktų sąsajomis.
- Borland BDE variklis leidžia efektyviai dirbti su dbf failais.
- Galima naudotis standartinėmis winapi funkcijomis.

Duomenų bazės realizacijai pasirinktas MySQL serveris dėl pakankamai aukštos kokybės ir žemos kūrimo kainos.

5.1. Eksperimentinio diegimo aprašymas

Tyrimas buvo atliekamas su 20 lentelėje išvardinta aparatūrine ir programine įranga.

20 lentelė

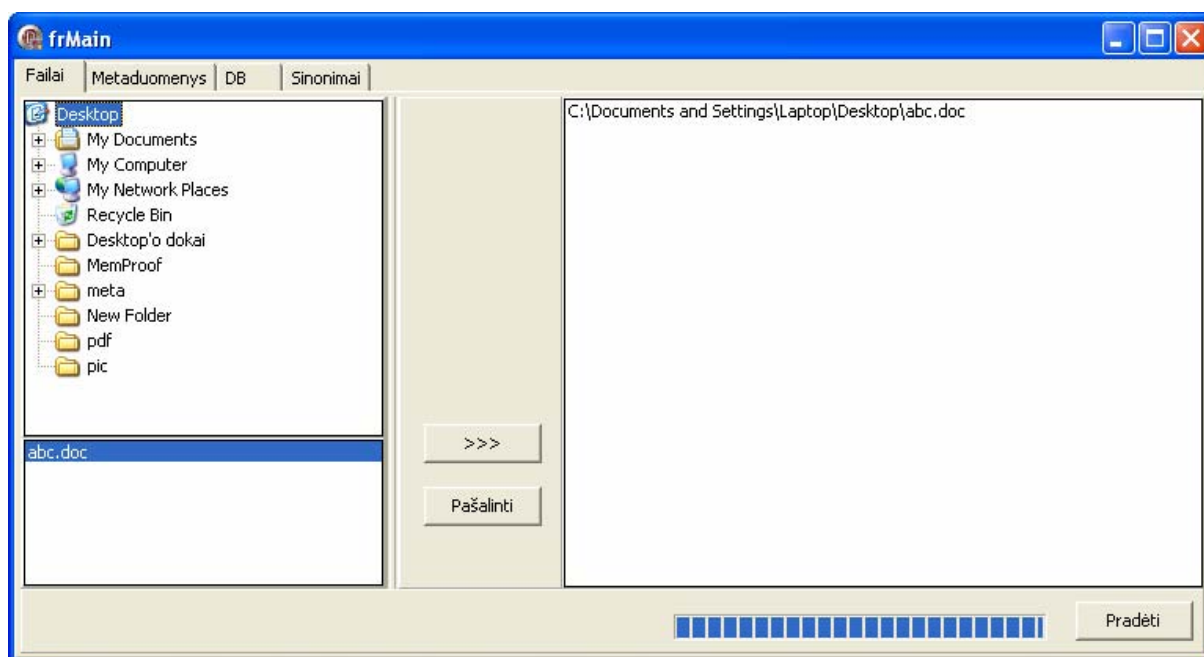
Tyrimui naudotos aparatūrinės ir programinės įrangos sąrašas

Aparatūrinė įranga	
Procesorius	AMD Athlon XP2000+ 1,66 GHz
Pagr. plokštė	Biostar M7VIT
Atmintis	PQI DDR 333 MHz, 512 MB
Kietasis diskas	Maxtor 120 GB, 8 MB buf., 7200 rpm UDMA 133
Programinė įranga	

Operacinė sistema	MS Windows XP SP2
DBVS	MySQL 4.0.15
ODBC tvarkyklė	MyODBC 3.51.06

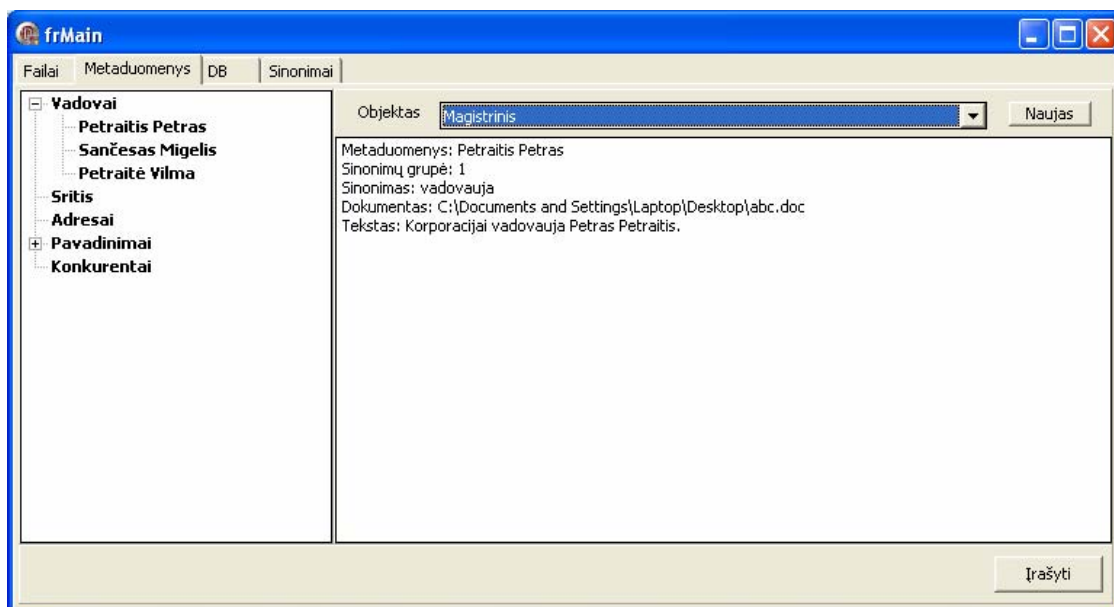
5.2. Sistemos naudojimo instrukcija

Atidarius programą Meta matomas pagrindinis langas (75 pav.). Jis yra suskirstytas į keturias sritis: Failai (75 pav.), Metaduomenys (76 pav.), DB (77 pav.), Sinonimai (79 pav.). Metaduomenų išgavimo darbas pradedamas Failų pasirinkimo lange. Pasirinkus visus norimus failus, iš kurių bus išgaunami metaduomenys spaudžiamas mygtukas „Pradėti“.

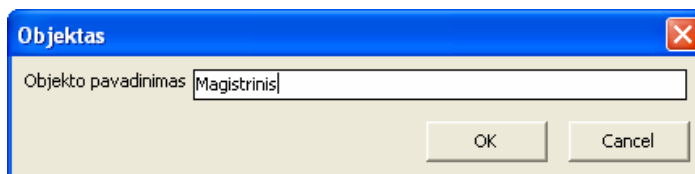


75 pav. Failų pasirinkimo langas

Išgauti semantiniai metaduomenys atvaizduojami Metaduomenų lange (76 pav.). Metaduomenų grupių sąrašą praskleidus kiekvieną grupę atvaizduojami rasti duomenys. Paspaudus ant konkrečios metaduomenų reikšmės dešinėje esančioje srityje atvaizduojama visa su šia reikšme susijusi informacija. Norint šią informaciją įrašyti į duomenų bazę, reikia pasirinkti objektą kuriam tie duomenys bus priskirti. Jei reikiamo objekto sąrašą nėra įvedamas naujas pavadinimas (pasirinkus mygtuką Naujas). Objekto pavadinimo įvedimui atsidaro langas pateiktas 77 paveiksle. Duomenys į DB įrašomi paspaudus mygtuką „Įrašyti“

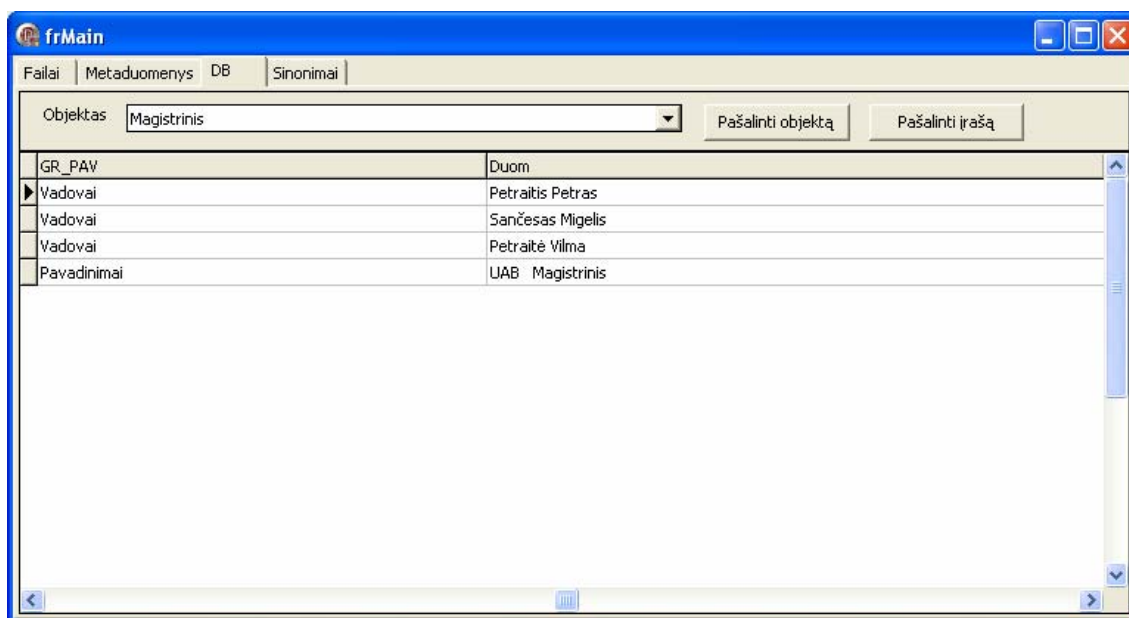


76 pav. Metaduomenų išgavimo langas



77 pav. Objekto pavadinimo įvedimo langas

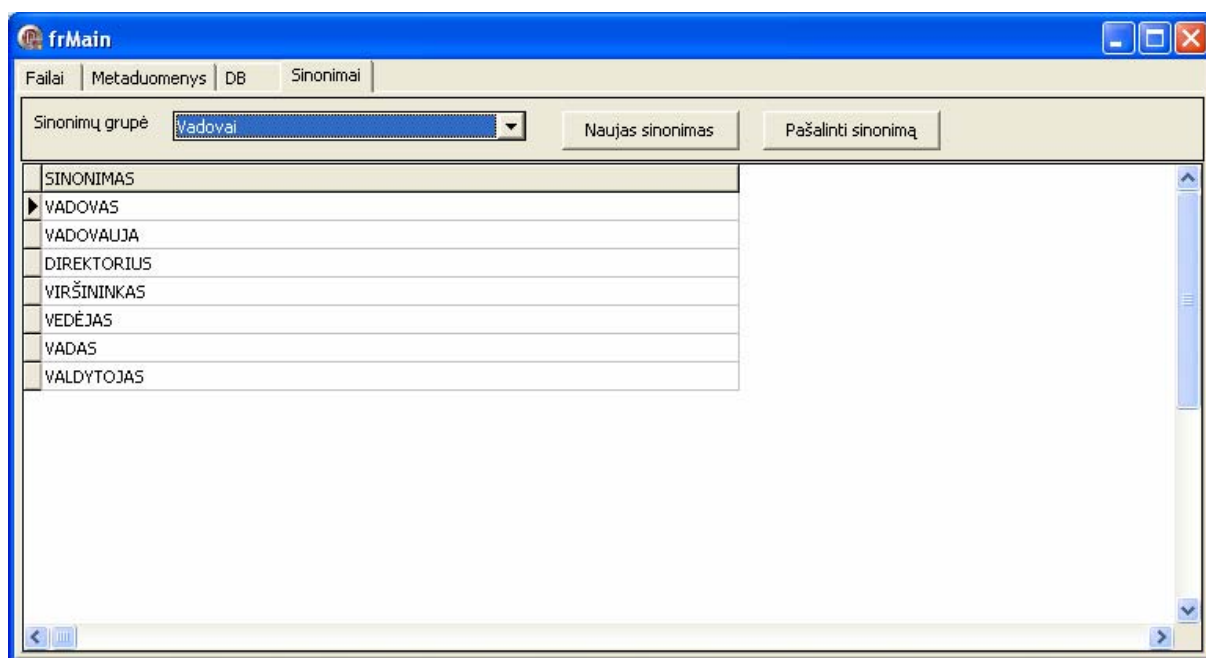
Duomenų bazėje esantys duomenys atvaizduojami 78 paveiksle pateiktame lange.



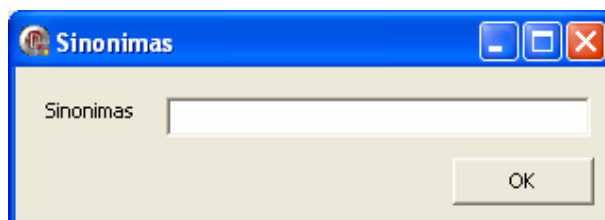
78 pav. Duomenų bazės langas

Pasirenkame objekto pavadinimą ir su juo susiję semantiniai metaduomenys atvaizduojami žemiau esančiame sąraše. Iš duomenų bazės galime pašalinti atskirus įrašus apie objektą (mygtukas „Pašalinti įrašą“) arba visus su juo susijusius įrašus („Pašalinti objektą“).

Sinonimų lange atvaizduojami ir redaguojami su tam tikra grupe susiję sinonimai. Šioje sistemoje yra penkios sinonimų grupės: Vadovai, Sritis, Adresai, Pavadinimai ir Konkurentai. Tekste ieškoma su ta sritimi susijusio sinonimo ir jį suradus ieškoma konkrečios informacijos. Naujos sinonimų grupės pridėti negalima. Sinonimas pridedamas pasirinkus mygtuką „Naujas sinonimas“ (atsidaro 80 paveiksle pateiktas langas) ir pašalinamas pasirinkus mygtuką „Pašalinti sinonimą“. Šalinimo veiksmui atlikti turi būti sąraše pažymėtas šalinamas sinonimas.



79 pav. Sinonimų peržiūros langas



80 pav. Sinonimo įvedimo langas

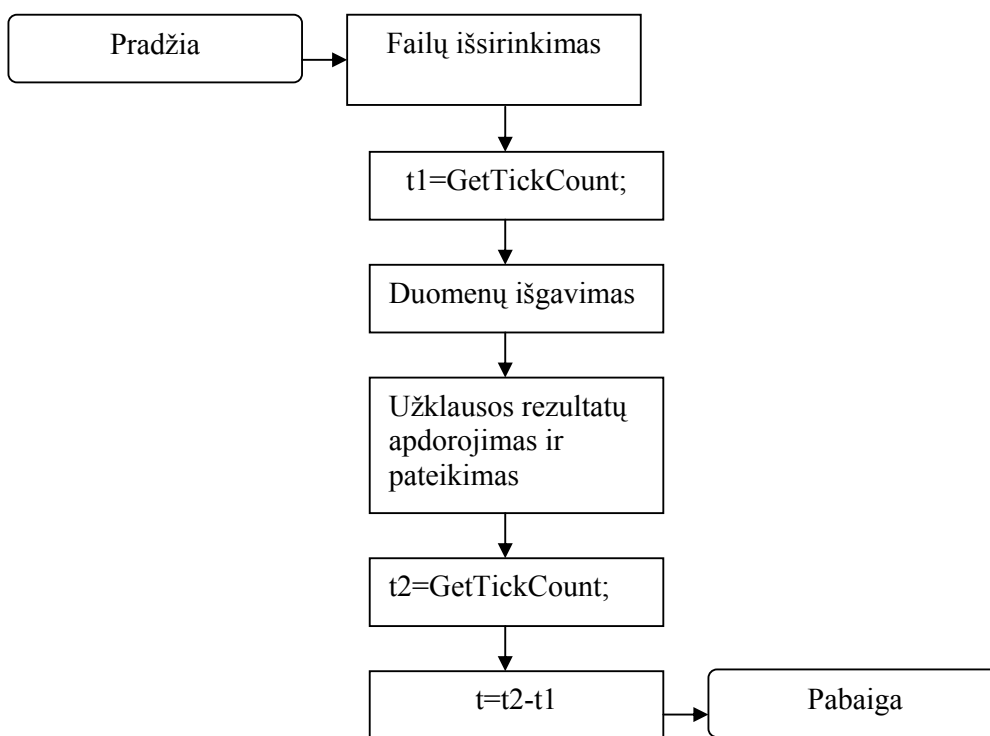
5.3. Sistemos veikimo ir savybių analizė

Sistemos veikimo sritis yra ganėtinai ribota, norint ją praplėsti reikalingas programos kodo pakeitimas ir lituanisto konsultacija.

Semantinių metaduomenų išgavimo sistema skirta informacijos apie verslo įmones (objektus) analizei ir organizavimui. Šioje sistemoje yra penkios sinonimų grupės: Vadovai, Sritis, Adresai, Pavadinimai ir Konkurentai. Duotuose MS Word tipo dokumentuose ieškoma kiekvienai sričiai priskirtų sinonimų ir juos suradus ieškoma konkreti informacija (pavyzdžiui vadovo vardas ir pavardė).

Semantinių metaduomenų paieškos algoritmo efektyvumas nustatomas matuojant duomenų radimo laikinę charakteristiką.

Algoritmas naudotas tyrimui:



81 pav. Metaduomenų radimo tyrimo algoritmas

Tiriant užklauso vykdymo laikus į programos kodą buvo įterpti procesoriaus laiko nuskaitymo operatoriai (*GetTickCount*). Kaip matyti iš 81 pav. laikas matuojamas prieš užklauso vykdymą ir po rezultatų pateikimo formoje. Šių dviejų laikų skirtumas atspindi pilną užklauso vykdymą.

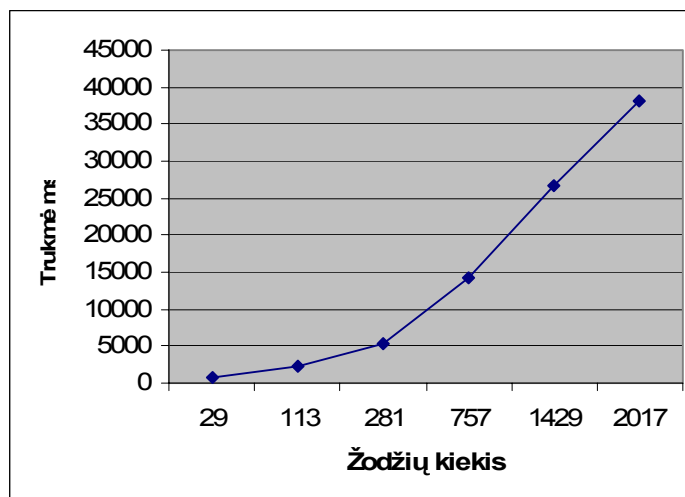
vykdymo laiką ir rezultatų apdorojimo laiką. Matavimas atliekamas su skirtingais apdorojamų duomenų kiekiais.

21 lentelė

Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų žodžių kiekio

Apdorojamų žodžių kiekis	Metaduomenų išgavimo trukmė
29	811 ms
113	2284 ms
281	5268 ms
757	14120 ms
1429	26738 ms
2017	38105 ms

Šio tyrimo rezultatai taip pat pateikti grafiškai 82 paveiksle.



82 pav. Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų žodžių kiekio

Iš 21 lentelės ir 82 paveikslo matome, kad didėjant apdorojamų žodžių skaičiui, proporcingai didėja ir metaduomenų išgavimo trukmė t.y. gaunama tiesinė priklausomybė.

22 lentelė

Metaduomenų išgavimo trukmės priklausomybė nuo apdorojamų dokumentų kiekio

Žodžių kiekis viename dokumente	Apdorojamų dokumentų kiekis	Metaduomenų išgavimo trukmė
2912	1	57623 ms
29	10	140352 ms

Iš 22 lentelės matome, kad sistema didžiąją laiko dalį sugaišta atidarydama dokumentus ir kiekvienam iš jų kurdama COM objektą.

5.4. Kokybės kriterijų vertinimas pagal ISO/IEC 9126 kokybės modelį

Bendram kokybės įvertinimui pasirinktas ISO/IEC 9126 kokybės modelis.

Pagal šį modelį yra 6 kokybės vertinimo charakteristikos [7]:

1. Funkcionalumas (*Functionality*)
2. Patikimumas (*Reliability*)
3. Panaudojamumas (*Usability*)
4. Efektyvumas (*Efficiency*)
5. Pernešamumas (*Portability*)
6. Priderinamumas (*Maintainability*)

5.4.1. Funkcionalumas

- Tinkamumas (*Suitability*) – sukurta programinė įranga neužtikrina pilno reikalavimuose numatyto funkcionalumo. Nerealizuota palyginimų dalis.
- Tikslumas (*Accuratness*) – programinė įranga neatlieka jokių skaičiavimo ar optimizavimo uždavinių, todėl tikslumas vertintas tik pagal duomenų išrinkimo funkcijas. Duomenys programoje pateikiami teisingi ir tikslūs (be iškraipymų).
- Sąveika (*Interoperability*) – programinė įranga importuoja duomenis iš Word failų.
- Suderinamumas (*Compliance*) – jokių standartų ar reikalavimų su kuriais turėtų būti suderinama ši programinė įranga nenumatyta.
- Sauga (*Security*) – nenumatyta, kad šioje programinėje įrangoje naudojami duomenys turėtų būti apsaugoti. Duomenys skirti viešam naudojimui.

5.4.2. Patikimumas

- Branda (*Maturity*) – tikslesniam brandos įvertinimui reikalingas ilgesnis darbo su programine įranga stebėjimas. Per neilgą darbo su programine įranga laikotarpį didesnių programos nesėkmių (*failure*) nepastebėta.
- Klaidų tolerancija (*Fault tolerace*) – sistema tikrina pasirinktų dokumentų formatą ir jei jis nepriimtinas, apie tai praneša vartotojui.
- Atstatomumas (*Recoverability*) – atstatomumo tyrimas nebuvo atliekamas, kadangi šios galimybės sistemoje nenumatyta.

5.4.3. Panaudojamumas

- Suprantamumas (*Understandability*) – vartotojo sąsajai paaiškinti yra parašyta vartotojo dokumentacija.
- Apsimokymas (*Learnability*) – atlikus testuotojų apklausą nustatyta, kad vartotojai nesunkiai išmoko naudotis programine įranga. Paruošta vartotojo dokumentacija buvo pakankamai išsami.
- Valdomumas (*Operability*) – daugumą sistemos funkcijų galima atlikti pelės pagalba. Tai vienas patogiausių valdymo būdų.

5.4.4. Efektyvumas

- Laikinė charakteristika (*Time behavior*) – rezultatai pateikti 5.3 skyriuje.
- Resursai (*Resource behavior*) – sistema apkrauna kompiuterio resursus minimaliai:
Procesoriaus apkrovimas: iki 11 %
Operatyvinė atmintis: apie 8 MB
Klientinės programos dydis: 1.13 MB

5.4.5. Pernešamumas

- Adaptacija (*Adaptability*) – programų sistema kurta konkrečiai aplinkai, todėl norint pritaikyti ją skirtingai aplinkai reikalingos papildomos priemonės ir veiksmai.
- Įdiegiamumas (*Installability*) – kartu su programų sistema pateikiamos visos reikalingos priemonės pilnam įdiegimui. Įdiegimas detalai aprašytas vartotojo dokumentacijoje.
- Pakeičiamumas (*Replaceability*) – sistemoje nenumatyta galimybė keisti tam tikras posistemes kitomis.

5.4.6. Priderinamumas

- Analizuojamumas (*Analyzability*) – pagal sisteminius pranešimus nėra sudėtinga išanalizuoti, kurioje sistemos dalyje gali būti klaidų. Techninių peržvalgų metu nustatyta kad programos kodas yra dokumentuotas, kintamųjų vardai prasmingi.
- Modifikuojamumas (*Changeability*) – sistema sukurta objektiškai, todėl keisti tam tikras kodo dalis nėra sudėtinga.
- Stabilumas (*Stability*) – kodo pakeitimas gali iššaukti sistemos nestabilumą.
- Testuojamumas (*Testability*) – testuoti sistemą nėra sudėtinga.

6. Išvados

1. Išanalizavus įvairius literatūros šaltinius, galima daryti išvadą, kad semantinių metaduomenų išgavimo technologijos tik pradedamos kurti, todėl nėra toli pažengusios. Pagrindinė problema susijusi su semantikos išgavimu yra ta, kad nestruktūrizuoti, pusiau struktūrizuoti ir struktūrizuoti duomenys yra analizuojami redakciniu požiūriu. Naudojant tokias technologijas vartotojas negali surasti konkrečių jį dominančių duomenų, jam pateikiama begalė informacijos taigi pačiam tenka pasirinkti tinkamą.
2. Palyginus pagrindines metaduomenų kodavimo technologijas RDF ir Dublin Core galima teigti, kad visos Dublin Core metamodelį sudarančios esybės ir ryšiai turi savo atitikmenis ir RDF metamodelyje. RDF technologija yra labiau išvystyta ir plačiau naudojama, taigi ji labiau tinkama semantinių metaduomenų aprašymui. Naudojant RDF kodavimą kiltų mažiau problemų susijusių su duomenų integravimu.
3. Tam, kad susipažinti su egzistuojančiomis semantinėmis technologijomis darbe išanalizuotas Semagix FREEDOM sistemos veikimas, trumpai apžvelgtos šios korporacijos sukurtos technologijos ir metodai. Detali informacija apie konkrečias technologijas ir metodus nėra viešai pateikiama, todėl nebuvo įmanoma atlikti pilnos jų analizės. Iš surinktos informacijos galima spręsti, kad esamos technologijos nėra universalios, jos skirtos tik konkreta turinio semantiniams metaduomenims išgauti ir apdoroti.
4. Siekiant išsiaiškinti semantinių metaduomenų susiejimo ypatumus išanalizuotos semantinių asociacijų kūrimo taisyklės. Remiantis šiomis taisyklėmis darbe suprojektuota sistema turėtų suteikti galimybę surasti net ir tolimiausią ryšį (jei tik toks yra) tarp dviejų semantinių esybių.
5. Suprojektuota ir dalinai realizuota sistema darbui su semantiniais metaduomenimis, pateikiamais lietuvių kalba. Atlikus šį darbą padaryta išvada, kad kol nėra nustatytos bendros sakinio rašymo konstrukcijos, labai sudėtinga išgauti reikiamus duomenis. Taigi galima daryti išvadą, kad semantinių metaduomenų išgavimo problemą turi spręsti ne tik IT, bet ir lingvistikos specialistai. Problemą galima būtų eliminuoti, jei egzistotų sakinio konstrukcijos taisyklės.
6. Siekiant įvertinti sistemos darbo efektyvumą ištirtas metaduomenų išgavimo trukmės matavimas. Ištirta išgavimo trukmės priklausomybė nuo apdorojamo žodžių bei dokumentų kiekio. Gauta, kad didėjant apdorojamų žodžių kiekiui viename dokumente, proporcingai didėja ir metaduomenų išgavimo trukmė. Tyrimo metu, taip pat nustatyta, kad sistema

metaduomenų išgavimo metu didesnę laiko dalį skiria dokumentų atidarymui, o ne metaduomenų paieškai.

7. Kitos darbe realizuotos sistemos charakteristikos (funkcionalumas, patikimumas, efektyvumas, pernešamumas ir priderinamumas) ištyrinėtos panaudojant ISO/IEC 9126 kokybės modelį.

7. Literatūra

- [1] David McGoveran. Enterprise Integrity: Data Integration, Part 1. Iš Business Integration Journal [interaktyvus]. 2002 birželis [žiūrėta 2004-12-01]. Prieiga per internetą: <http://www.bijonline.com>

- [2] David McGoveran. Enterprise Integrity: Data Integration, Part 2. Iš Business Integration Journal [interaktyvus]. 2003 balandis [žiūrėta 2004-12-01]. Prieiga per internetą: <http://www.bijonline.com>

- [3] David McGoveran. Enterprise Integrity: Data Integration, Part 3. Iš Business Integration Journal [interaktyvus]. 2002 kovas [žiūrėta 2004-12-01]. Prieiga per internetą: <http://www.bijonline.com>

- [4] Amit Sheth, McLean. Ontology Driven Integration and Analysis for Semantic Applications iš Ontology and Semantic Web Technical Exchange Meeting MITRE [interaktyvus], 2003 06 12 [žiūrėta 2004-10-28]. Prieiga per internetą <http://lstdis.cs.uga.edu/library/publications/>

- [5] Mark Fisher, Amit Sheth Semagix Inc. Semantic Enterprise Content Management [interaktyvus] 2003 [žiūrėta 2004-10-29]. Prieiga per internetą <http://lstdis.cs.uga.edu/library/publications/>

- [6] Finkenstein, Peter H. Aiken. Building Corporate Portals with XML. 1999

- [7] Amit Seth. Semantics for the Semantic Web: The Implicit, the Formal and the Powerful. Iš Journal on Semantic Web & Information Systems [interaktyvus]. 2005 sausis – kovas [žiūrėta 2005-02-23]. Prieiga per internetą: <http://lstdis.cs.uga.edu/library/publications/>

- [8] ISO 9126: The Standard of Reference. Iš HCI Bibliography : iš CSE – Centre for Software Engineering [interaktyvus]. 2004, kovas [žiūrėta 2005-08-20], prieiga per internetą: <http://www.cse.dcu.ie/essiscope/sm2/9126ref.html>

- [9] Finkenstein, Peter H. Aiken. Building Corporate Portals with XML. 1999

- [10] Andy Powell. DCMI abstract model. Iš Dublin Core Metadata Initiative [interaktyvus]. 2005-03-07 [žiūrėta 2005-11-15], prieiga per internetą: <http://dublincore.org/documents/abstract-model/>

- [11] Dan Brickley. RDF Schema Specification. Iš W3C [interaktyvus]. 2004-02-04 [žiūrėta 2005-11-15], prieiga per internetą: <http://www.w3.org/TR/rdf-schema/>

- [12] Christian Halaschek-Wiener. A Flexible Approach For Ranking Complex Relationships On The Semantic Web. Iš Journal on Semantic Web & Information Systems [interaktyvus]. 2005 sausis – kovas [žiūrėta 2005-10-03]. Prieiga per internetą: <http://lsdis.cs.uga.edu/library/publications/>