



**KAUNO TECHNOLOGIJOS UNIVERSITETAS  
INFORMATIKOS FAKULTETAS**

**Marius Viliokaitis**

**OPERATYVIAI RENKAMŲ DUOMENŲ TYRIMAS IR  
APIBENDRINIMAS**

**MAGISTRO DARBAS**

**Vadovas**

Doc. dr. Eimutis Karčiauskas

**KAUNAS, 2015**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**INFORMATIKOS FAKULTETAS**

**OPERATYVIAI RENKAMŲ DUOMENŲ TYRIMAS IR**  
**APIBENDRINIMAS**

**MAGISTRO DARBAS**

**Vadovas**

Doc. dr. Eimutis Karčiauskas

**Recenzentas**

Doc. dr. Lina Nemuraitė

**Projektą atliko**

Marius Viliokaitis

**KAUNAS, 2015**

# TURINYS

<b>1.</b>	<b>IŽANGA.....</b>	<b>7</b>
1.1.	DOKUMENTO PASKIRTIS .....	7
1.2.	SANTRAUKA .....	7
<b>2.</b>	<b>ANALITINĖ DALIS .....</b>	<b>7</b>
2.1.	ĮVADAS .....	7
2.2.	TIKSLAS .....	8
2.3.	IŠKILUSIOS PROBLEMOS.....	8
2.4.	PROJEKTO KŪRIMO PAGRINDIMAS.....	8
2.5.	REALIZAVIMO GALIMYBĖS.....	9
2.5.1.	<i>BBC weather</i> .....	9
2.5.2.	<i>National Weather service</i> .....	9
2.5.3.	<i>Weather spark</i> .....	9
2.5.4.	<i>WeatherBug</i> .....	9
2.5.5.	<i>HAMweather</i> .....	10
2.5.6.	<i>Open Weather Map</i> .....	10
2.5.7.	<i>WUnderground</i> .....	10
2.5.8.	<i>JS Charts</i> .....	10
2.5.9.	<i>Website Template</i> .....	10
2.6.	VEIKIMO PRINCIPAS.....	10
2.6.1.	<i>Duomenų surinkimas</i> .....	10
2.6.2.	<i>Surinkėjo darbas</i> .....	11
2.6.3.	<i>Surinkėjų tipai</i> .....	13
2.6.4.	<i>Gero rinkimo algoritmo ypatybės</i> .....	13
2.7.	IŠVADOS .....	13
<b>3.</b>	<b>PROJEKTINĖ DALIS.....</b>	<b>14</b>
3.1.	SISTEMOS VEIKLOS SFERA .....	14
3.1.1.	<i>Sistemos funkcijos</i> .....	14
3.1.2.	<i>Panaudojimo atvejų sąrašas</i> .....	14
3.2.	FUNKCINIAI REIKALAVIMAI IR REIKALAVIMAI DUOMENIMS .....	17
3.2.1.	<i>Funkciniai reikalavimai</i> .....	17
3.2.2.	<i>Reikalavimai duomenims</i> .....	24
3.3.	NEFUNKCINIAI REIKALAVIMAI .....	24
3.3.1.	<i>Reikalavimai sistemos išvaizdai</i> .....	24
3.3.2.	<i>Reikalavimai panaudojamumui</i> .....	25
3.3.3.	<i>Reikalavimai vykdymo charakteristikoms</i> .....	25
3.3.4.	<i>Reikalavimai veikimo sąlygoms</i> .....	25
3.3.5.	<i>Reikalavimai sistemos priežiūrai</i> .....	26
3.3.6.	<i>Reikalavimai saugumui</i> .....	27
3.4.	VEIKLOS KONTEKSTAS .....	27
3.5.	SISTEMOS STATINIS VAIZDAS .....	28
3.6.	SISTEMOS DINAMINIS VAIZDAS .....	29
3.6.1.	<i>Sekų diagramos</i> .....	29
3.6.2.	<i>Būsenų diagramos</i> .....	31
3.6.3.	<i>Veiklos diagramos</i> .....	34
3.7.	DUOMENŲ VAIZDAS .....	36
<b>4.</b>	<b>TYRIMO DALIS .....</b>	<b>36</b>
4.1.	TIKSLAS .....	36
4.2.	TYRIMO REIKŠMĖ.....	36
4.3.	TYRIMO EIGA .....	37
4.4.	TYRIMO IŠVADOS.....	39
<b>5.</b>	<b>EKSPERIMENTINĖ DALIS .....</b>	<b>39</b>
5.1.	TIKSLAS .....	39
5.2.	EKSPERIMENTO APRAŠAS.....	39

5.2	EKSPERIMENTAS 1 .....	39
5.3.	EKSPERIMENTAS 2 .....	41
5.4.	EKSPERIMENTO REZULTATAI.....	42
<b>6.</b>	<b>IŠVADOS.....</b>	<b>43</b>
<b>7.</b>	<b>LITERATŪRA .....</b>	<b>44</b>
<b>8.</b>	<b>TERMINŲ IR SANTRUMPŲ ŽODYNAS.....</b>	<b>45</b>
<b>9.</b>	<b>PRIEDAI .....</b>	<b>46</b>
9.1.	EKSPERIMENTO REZULTATŲ APIBENDRINIMAS.....	46

## PAVEIKSLĖLIŲ SĄRAŠAS

1. PAV. CRAWLER'IO ARCHITEKTŪRA.....	11
2. PAV. SURINKĖJO VEIKSMŲ SEKA.....	12
3. PAV. SISTEMOS RIBOS .....	14
4. PAV. PRADINIS DUOMENŲ MODELIS .....	24
5. PAV. KONTEKSTO DIAGRAMA.....	28
6. PAV. SISTEMOS IŠSKAIDYMAS Į PAKETUS .....	29
7. PAV. PRISIJUNGIMO PA SEKŲ DIAGRAMA .....	30
8. PAV. PASIRINKTI MIESTĄ PA SEKŲ DIAGRAMA .....	30
9. PAV. PERŽIŪRĖTI STATISTIKĄ PA SEKŲ DIAGRAMA.....	31
10. PAV. FILTRUOTI STATISTIKĄ PA SEKŲ DIAGRAMA.....	31
11. PAV. PRISIJUNGIMO PA BŪSENŲ DIAGRAMA .....	32
12. PAV. PASIRINKTI MIESTĄ PA BŪSENŲ DIAGRAMA .....	32
13. PAV. PERŽIŪRĖTI STATISTIKĄ PA BŪSENŲ DIAGRAMA .....	33
14. PAV. FILTRUOTI STATISTIKĄ PA BŪSENŲ DIAGRAMA.....	33
15. PAV. PRISIJUNGTI PA VEIKLOS DIAGRAMA .....	34
16. PAV. PASIRINKTI MIESTĄ PA VEIKLOS DIAGRAMA .....	34
17. PAV. PERŽIŪRĖTI STATISTIKĄ PA VEIKLOS DIAGRAMA .....	35
18. PAV. FILTRUOTI STATISTIKĄ PA VEIKLOS DIAGRAMA .....	35
19. PAV. DUOMENŲ BAZĖS MODELIS .....	36
20. PAV. KAUNO TEMPERATŪRA VASARIO 16 DIENAI .....	37
21. PAV. KAUNO TEMPERATŪRA VASARIO 16 NAKČIAI .....	37
22. PAV. BBC DIENOS IR NAKTIES ILGALAIKIAI TIKSLUMAI .....	41
23. PAV. INTELICAST DIENOS IR NAKTIES ILGALAIKIAI TIKSLUMAI.....	41
24. PAV. OWM DIENOS IR NAKTIES ILGALAIKIAI TIKSLUMAI.....	42
25. PAV. WWO DIENOS IR NAKTIES ILGALAIKIAI TIKSLUMAI.....	42
26. PAV. BBC DIENOS TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI .....	46
27. PAV. BBC NAKTIES TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	46
28. PAV. INTELICAST DIENOS TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	47
29. PAV. INTELICAST NAKTIES TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	47
30. PAV. OWM DIENOS TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI .....	48
31. PAV. OWM NAKTIES TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	48
32. PAV. WWO DIENOS TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	49
33. PAV. WWO NAKTIES TEMPERATŪRŲ PROGNOZIŲ TIKSLUMAI.....	49

## SUMMARY

By collecting information we can make assumptions, foresee possible events and even conjecture certain objects or exams for approval. Even if information collecting is versatile and technological possibilities are developing, it grows in public gathering and presentment as well. However, not all information given is appropriate and useful. Therefore information is collected and re-processed as many times as necessary for aims of the project.

This information is presented in many formats and not always from primary sources, making it difficult to analyze and present the results. Even information gathering from several sources can vary cardinally and be completely incompatible. Since information collected from different sources cannot match, it requires the gathering of big quantities of information as quickly as possible for processing purposes, analysis and distinguishment of the most suitable interface connections.

Because the project is designed to collect long-term data it needs to analyze possible changes in source system and atleast slightly protect it from data loss. Information gathering must be protected from possible service errors to avoid interrupting the information collection.

In order to obtain more precise data and to make it as accurate as possible, the information must be collected for a long time, without interruptions. In order to get more precise data minimal time of data gathering should be at least three months.

# 1. IŽANGA

## 1.1. Dokumento paskirtis

Dokumentas yra ruošiamas kaip baigiamojo magistrinio darbo dalis. Siekiama skaitytojus supažinti su kurta sistema, jos problemomis ir paskirtimi. Šiame dokumente pateikiama analitinė dalis suteikia informacijos apie projektuojamos sistemos ir problemų sprendimo būdus, pristatomas projektas. Pateikiamas kuriamai sistemai iškeltas uždavinys, peržvelgiamos projektavimo metu iškilusios problemos. Taip pat apžvelgiami galimi sprendimo būdai. Projektinėje dalyje aprašomos sukurtos programinės įrangos techninės-projektinės dokumentacijos esminės dalys, pateikiami iškelti reikalavimai sistemai ir panaudojimo atvejai. Tiriamojoje ir eksperimentinėse dalyse pateikiama informacija tiriant ir analizuojant duomenis, gautus naudojant sistemą.

## 1.2. Santrauka

Informacijos rinkimas ir pateikimas nuo senų senovės domina žmones ir tai nuolatos plėtojama. Rinkdami informaciją galime daryti prielaidas, spelioti galimus įvykius ar net daryti hipotezes tam tikriems objektams ar tyrimams patvirtinti. Nors informacijos rinkimas yra įvairiapusiškas, plėtojantis technologinėms galimybėms vis labiau plinta viešas informacijos rinkimas ir pateikimas. Tačiau informacija ne visada pateikiama tikslinga ir naudinga, todėl ši informacija vėl renkama ir vėl apdorojama tiek kartų, kiek reikia projekto tikslams įgyvendinti.

Šio projekto atžvilgiu informacija pateikiama daugeliu formatų ir ne visada iš pirmųjų šaltinių, todėl sunku ją analizuoti ir pateikti norimą rezultatą. Netgi renkant iš kelių šaltinių informaciją gali kardinaliai skirtis ir būti visiškai nesuderinama. Kadangi net surinkus informaciją iš kelių šaltinių ji gali neatitikti, reikia surinkti kuo daugiau informacijos per kuo trumpesnę laiką, ją apdoroti, analizuoti ir atskirti tinkamiausius sąsajos ryšius.

Kadangi projektas skirtas ilgalaikiam duomenų rinkimui turi išanalizuoti galimus pokyčius šaltinių sistemos ir bent minimaliai apsaugoti nuo informacijos praradimų. Informacijos rinkimas turi būti apsaugotas nuo galimų serviso darbo nesklaidumų, kad nenutrūktų informacijos rinkimas.

Siekiant gauti kuo tikslesnius duomenis ir padaryti kuo tikslesnes darbo išvadas ir analizę informacija turi būti renkama ilgą laiką, be pertraukų. Minimalus informacijos rinkimo laikotarpis siekiant išgauti tikslesnius duomenis turėtų būti bent tris mėnesius.

# 2. ANALITINĖ DALIS

## 2.1. Įvadas

Informacijos rinkimas ir pateikimas nuo senų senovės domina žmones ir tai nuolatos plėtojama. Rinkdami informaciją galime daryti prielaidas, spelioti galimus įvykius ar net daryti hipotezes tam tikriems objektams ar tyrimams patvirtinti. Nors informacijos rinkimas yra įvairiapusiškas, plėtojantis technologinėms galimybėms vis labiau plinta viešas informacijos rinkimas ir pateikimas. Tačiau informacija ne visada pateikiama tikslinga ir naudinga, todėl ši informacija vėl renkama ir vėl apdorojama tiek kartų, kiek reikia projekto tikslams įgyvendinti.

Informacijos kiekiai ir paklausa nuolat didėja, todėl reikalingi vis tobulesni informacijos rinkimo ir saugojimo įrankiai. Kadangi ši paklausa vis didėja rinkoje plėtoti šį projektą yra labai naudinga. Taip pat informacijos rinkimas nėra pakankamas indėlis į tikslingų duomenų gavimą, todėl ją reikia analizuoti ir atrinkti naudingus šaltinius.

Šiomis dienomis galite pasiekti informaciją pateikiamą internete bet kokiomis priemonėmis, turinčiomis interneto prieigą tikslinga orientuotis į informacijos pateikimą skirtą visoms platformoms. Papildomai, siekiant palengvinti vartotojų paieškas, galima siūlyti vis tobulėjančias technologijas, pavyzdžiui – vietovės nustatymo galimybę. Tai padėtų ir pagreitintų informacijos gavimą reikiamai auditorijai su reikiamais parametrais.

## 2.2. Tikslas

Projekto tikslas yra ištirti duomenų rinkimo metodus ir pritaikyti žinias kuriant duomenų rinkimo sistemą, siekiant analizuoti gautus rezultatus ir jų tikslumą. Sukūrus tokią sistemą ją turėtų būti nesunku pritaikyti kelioms sritims. Sistemos siekis yra sukurti sistemą skirtą didelio informacijos kiekio rinkimui ilgą laiką ir jos apdorojimui, pateikiant vartotojui prieinamu būdu.

Pasirinkta projekto kryptis (orų prognozės) yra labai aktuali šių dienų pasauliui, ypač keliautojams. Ne visi informaciniai šaltiniai pateikia tikslią informaciją, o analizę apie praėjusių metų vidutinę temperatūrą reikiamoje vietoje pasilieka tik sau. Vartotojui nepaleikama galimybė pačiam apgalvoti ar tikrai gali būti tokia temperatūra tokiu metu. Jam pateikiama tik artėjančių dienų prognozė.

Igyvendinus šį projektą vartotojai turėtų jaustis saugiau dėl galimų kelionių į svetimą šalį, galėtų atskirti kuris šaltinis yra tikslesnis ir patys pasirinkti sau malonesnę informacijos tiekėją. Pritaikius šią sistemą ir įvairiems įrenginiams galima būtų ja naudotis iš bet kurios pasaulio vietos ir turėti orų statistiką pasirinktai vietai.

## 2.3. Iškilusios problemos

Siekiant įgyvendinti šią sistemą ir padaryti ją kuo imanoma universalesnę susiduriama su nemažais problemų. Pati didžiausia problema – skirtingos tiekėjų struktūros. Kadangi ne visi tiekėjai teikia galimybę skaityti duomenis patogiais formatais, tenka kurti skirtingus sprendimus. Netgi jei tiekėjai gražina patogiu apdoroti formatu – beveik visada išskaidymas skiriasi. Tuomet tenka prie jo prisitaikyti ir jį išanalizuoti.

Dar viena labai didelė problema – atsakymo netikslumai. Kadangi sistema siekiama rinkti didelius kiekius, neretai tiekėjų serveriai nespėja pateikti visų rezultatų ar dalies jų. Tokiu atveju gaunamas tuščias arba klaidos atsakymas. Sistema turi toliau veikti nekreipiant dėmesio į tokius trikdžius.

Taip pat galima ir ganėtinai maža, bet svarbi problema – duomenų dubliavimasis. Kartais dėl laiko juostos skirtumo gražinami duomenys jau yra duomenų bazėje, todėl sistema turi gebėti juos apdoroti.

## 2.4. Projekto kūrimo pagrindimas

Didžioji dauguma šių dienų informacinių sistemų naudoja standartinius vieno tiekėjo duomenų šaltinius ir taip sulaukia daug neatitikimų. Lyginant tų pačių duomenų atitikmenis skirtingose sistemose dažnai pastebima didelių skirtumų. Tokiose sistemose ne vien duomenys skiriasi, bet ir tikslumas gali skirtis – vienoje sistemoje tam tikri duomenys tikslesni nei kitoje. Ši tendencija nuolatos kinta, kadangi duomenų tiekėjai patys ne visada gali pateikti tikslius duomenis. Tokio neatitikimo ir netikslumo problema yra tokia, kad vartotojas niekada negali tiksliai žinoti reikiamos informacijos ir todėl nukenčia sistemos patikimumas.

Taip pat dar viena priežastis yra duomenų dingstamumas. Dauguma sistemų sistemų nesaugo duomenų ilgą laiką, duomenys yra panaikinami iškart, kai jie nebereikalingi, todėl neįmanoma daryti tikslingų ataskaitų. Nerenkant duomenų sunkiau sudaryti tikslingesnius šios dienos duomenis, nes neįmanoma atrinkti kuris šaltinis ar tiekėjas yra naudingesnis ir pateikia tikslesnę informaciją.

Viena didžiausių problemų tokių sistemų kūrime yra kaina. Kiekviena sistema ir duomenų rinkimas yra unikalūs, todėl reikiamos informacijos įsisavinimas, projektavimas ir paruošimas gavybai yra ganėtinai ilgas procesas. Siekiant sistemai dirbti tikslingai reikalingi specialistai, kurie per šį ganėtinai ilgą procesą yra brangiai apmokami.

Dar viena problema yra duomenų saugyklos. Kadangi sistemos būna didelės, o informacijos srautai dar didesni, duomenų bazė ir sistemos sąveikavimas su ja turi būti kuo geresnis. Kad tai būtų galima pasiekti informacijos centrai turi būti tam paruošti ir samdomi specialistai jiems palaikyti.

Būtent dėl šių priežasčių pradėta projektuoti kuo universalesnės duomenų rinkimo sistemos prototipas. Sistemą siekiama sukurti kuo universalesnę ir našesnę, pradžioje pasirenkant vieną duomenų kryptį.



Sistema turėtų būti kuo universalesnė vartojimo atžvilgiu, taip pagreitindama ir palengvindama naujų šaltinių pritaikymą esamiems komponentams. Tai turėtų sumažinti išlaidas skirtas kiekvieno naujo komponento gamybai ir pritaikymui, bei sumažinti darbą siekiant išgauti reikiamą informaciją ilgą laikotarpį.

## **2.5. Realizavimo galimybės**

Interneto platybėse apstu įvairių šaltinių, tačiau ne visi yra tokie paprasti ar tiesiog nemokami. Siekiant palengvinti informacijos išgavimą naudojamos aplikacijų programavimo sąsajos (API). Šiame skyriuje bus aptartos kelios pradinės API, kurių, siekiant gauti tikslesnius įvertinimus, gali daugėti. Taip pat jei API netenkins reikalavimu gali būti atsisakoma.

### **2.5.1. BBC weather**

Vienos geriausių ir populiariausių pasaulio naujienų agentūros orų prognozių pateiktis. Žinių tarnyba pateikia ganėtinai tikslią informaciją kas tris valandas bet kuriam pasirinktam miestui. Puslapis pateikia galimybę peržiūrėti orų prognozes kelioms dienoms į priekį. Taip pat pateikiama informacija apie matomumą ir drėgmę ore.

Nors puslapis ir informacija pateikiama sklandžiai, greitai ir maloniai vartotojui pasigendama orų prognozių analizės praėjusiom dienom. Taip pat kartais neteisingai nustatomos oro sąlygos (rodoma debesuota, nors sninga).

### **2.5.2. National Weather service**

Nacionalinio orų centro prognozės, lyginant su BBC pateikiamomis, yra žymiai tikslesnės. Pateikiama ne tik galimybė peržiūrėti orų sąlygas esamu metu ar artėjančioms dienoms, čia galima peržiūrėti praėjusių mėnesių orų prognozių tikslumus, šilumos pasiskirstymą, vėjo gūsius ir panašiai. Puslapis informatyvus ir kupinas informacijos, kuri padėtų atlikti analizę.

Siekiant sukurti kuo išsamesnę ir tikslesnę analizę atliekant šį projektą, reikėtų ligiuotis į tokio lygio informacijos pateikimo puslapius. Analizuojant šį informacijos pateikimo puslapį nerasta jokios vartotojui malonios pateikties. Dauguma analizės ir atvaizdavimo formų yra ganėtinai nesuprantamos ir perpildytos informacijos.

### **2.5.3. Weather spark**

Tai vienas iš maloniausių ir išsamiausių informacijos apie orus pateikčių. Informacija pateikiama sklandžiai, suprantamai ir maloniai vartotojui. Atliekama analizė gali būti pasirenkama mėnesių, metų, dienų ar savaitės tikslumu. Informacija pateikiama tiksliai ir išsamiai. Pateikiamas pilnas grafai pasirinktu laikotarpiu, netgi pateikiama animuota informacija apie pasirinkto pasaulio taško temperatūrą, vėjo ar drėgmės pokyčius minučių tikslumu.

Nors visa tai pateikiama sklandžiai ir patraukliai pasigendama tikslumų analizės ir pačių orų prognozių tikslumo. Taip pat naujam vartotojui pateikiama per daug neaiškios informacijos ir prarandama daug laiko analizuojant pačią sistemą, o ne pateikiamus duomenis. To vertėtų vengti kuriamame produkte.

### **2.5.4. WeatherBug**

Pasaulinė orų duomenų bazė skirta programuotojams. Šis API leidžia kūrėjams pasiekti duomenų centrą ir naudoti gautus duomenis savom reikmėm. Programuotojams suteikiama galimybė ieškoti vietovių ne tik pagal adresą, bet ieškoti ir individualių stočių. Taip pat teikiama galimybė gauti nuolatos atnaujinamus orus (Live). Kaip ir didžioji dauguma nemokamų paslaugų tiekėjų reikalingas registravimasis.

Per dieną leidžiamų užklausų – 5000. Per valandą 400.

### **2.5.5. HAMweather**

Tai turbūt viena paprasčiausių ir maloniausių programuotojui skirtų API. API kūrėjai pateikia nemokamas paslaugas duodami tikslius programinius kodus reikalingus išgauti reikiamai informacijai keliomis programavimo kalbomis. Taip pat suteikia galimybę užklausas pateikti saugesniu "https" protokolu. Šis tiekėjas kūrėjams (developer) nemokamai duoda 15 dienų, savaitės orų prognozes, saulės kilimo ir leidimosi laikus, audrų, potvynių, gaisrų ir žemės drebėjimų informaciją.

Per diena leidžiamų užklausų – 750.

### **2.5.6. Open Weather Map**

Paprastas API skirtas specialiai kūrėjams. Nors ir šio informacijos centro tik pradinė stadija, jis jau siūlo ganėtiniai išpūdingus skaičius. Pasak OpenWeatherMap puslapio kūrėjus informacija renkama iš virš 40,000 oro stotelių. Informacija gražinama JSON, XML ar HTML formatais. Informacija gaunama pagal miesto pavadinimą, geografines coordinates arba miesto ID. Pateikiama 5 ir 14 dienų orų prognozė.

Nemokamos versijos maksimalus užklausų skaičius per dieną – 40 milijonų. Per minutę – iki 30 tūkst.

### **2.5.7. WUunderground**

Kaip ir dauguma (galbūt net visi) API informacija gražinama XML arba JSON formatais. Šis tiekėjas taip pat teikia informacija apie orus kiekvieną valandą. Papildomai siulomos papildomos bibliotekos ar integracijos su kitais šaltiniais.

Užklausų per dieną – 500, per minutę – 10.

### **2.5.8. JS Charts**

Javascript kalba parašytas komponentas skirtas komponentas. Tai diagramų ir grafų generavimo yrankis skirtas paprastam ir lengvam garfikų kūrimui. Šis komponentas lengvai kuria grafikus iš gerai žinomų „XML“ ar „JSON“ duomenų paketų.

Šis komponentas smarkiai pagreitintų programinės vartotojų terpės kūrimą. Naudojant komponentą išgautus duomenis iš duomenų bazės galima perkonvertuoti į grafą, taip sukuriant malonų ir suprantamą vaizdą.

### **2.5.9. Website Template**

Nors tai ir minimalus komponentas vartotojo ar administratoriaus sąsajai kurti, tai sparčiai pagreitintų darbą kuriant malonią vartotojo aplinką. Dauguma internete esančių šablonų jau turi reikiamas komponentes (prisijungimo langas, meniu, registracija, administratoriaus panele, t.t.).

Šio komponento dėka laikas būtų sutaupomas, nes nereikėtų pačiam kurti vartotojo sąsajos griaučių ir daugiau laiko skirti sistemos tobulinimui ir universalumui.

## **2.6. Veikimo principas**

### **2.6.1. Duomenų surinkimas**

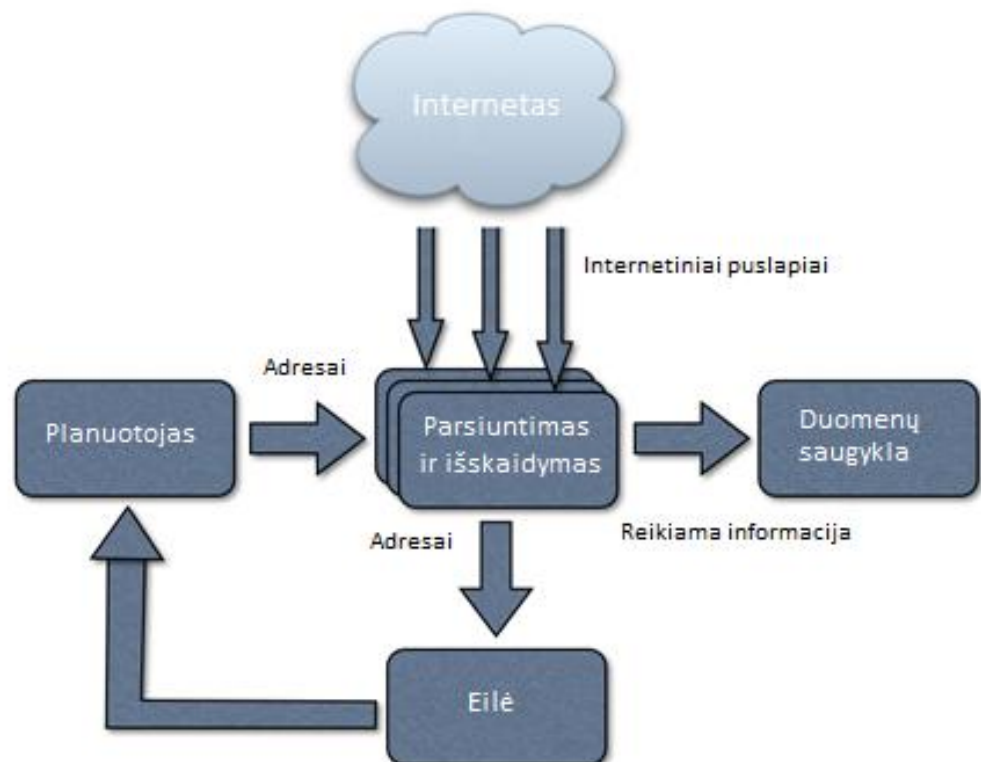
Puslapių šliaužiotojas (crawler) – tai programa ar programinis scenarijus skirtas naršyti internetiniuose puslapiuose pagal tam tikras pateiktas taisykles ir surinkti reikiamą informaciją. Kadangi crawler'iai geba vaikščioti per aprašytus puslapius ir rinkti reikiamą informaciją, tai tampa svarbia dalimi paieškos sistemoms. Nuolatos daugėjant interneto puslapių auga ir surinkėjų paklausa. Duomenų surinkėjui vaikstant po atitinkamus internetinius adresus surenkami tolimesni adresai – šakos. Šakos gaunamos užėjus į puslapį ir surinkus tame puslapyje esančius kitus adresus, kurie yra pridedami prie tolimesnės sekos. Ši seka vadinama surinkėjo riba. Pagal tam tikras taisykles ir nuostatas adresai pateikti sekoje yra aplankomi atskirai. Kiekvienas puslapis yra parsisiunčiamas ir saugomas paieškos varikliuko duomenų bazėje. Vėliau šie puslapiai yra patalpinami į eilę ir yra

paruošti pateikimui betkuriam paieškos vartotojui. Ieškomi adresai ir susiję failai gali būti pasiekiami betkuriuo laiko momentu. Pasitelkiant atitinkamus algoritmus puslapių šliaužiotojai gali rasti tinkamus adresus paieškom ir juo vėliau naudoti. Tačiau tai reikalauja labai didelių duomenų bazių, gebančių apdoroti didelius duomenų kiekius.

## 2.6.2. Surinkėjo darbas

Žemiau pateiktame paveiklėlyje atvaizduojama standartinė surinkėjo struktūra. Pabrėžtini trys komponentai: šakos – tai sąrašas neaplankytų adresų, kurie turi būti aplankyti, puslapio siuntėjas – tai komponentas parsiončiantis puslapio turinį, ir saugykla – tai vieta, kur saugomi rinkėjo parsionti puslapiai. Trumpai apie kiekvieną dalį:

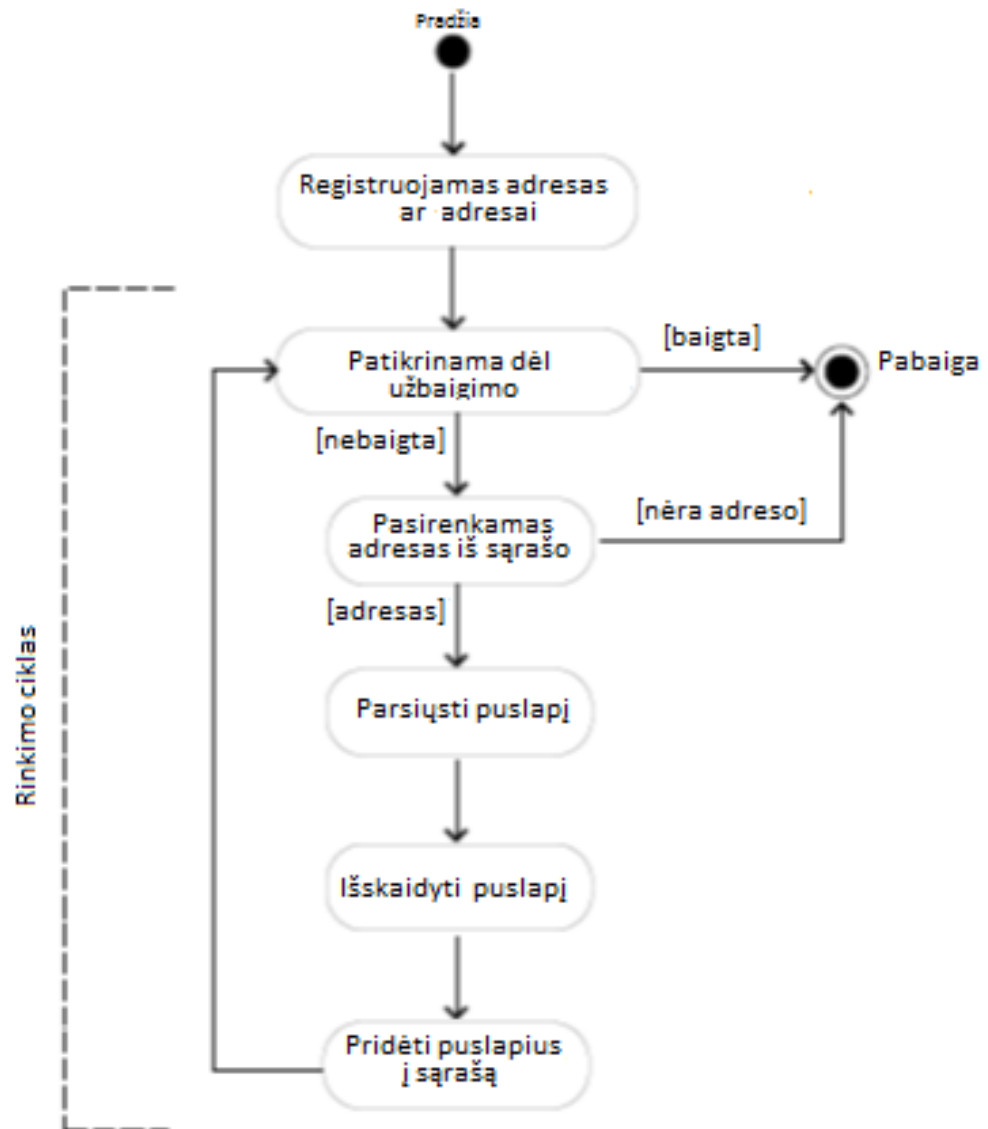
- **Šakos/Planuotojas** – tai sąrašas neaplankytų internetinių adresų. Rinkėjas pasiema adresą iš šakų sąrašo ir jį aplanko, surenka informaciją ir gražina. Paprastai – tai tiesiog adresų sąrašas. Jeigu lankomame puslapyje randamas adresas, kurio dar nėra sąrašė – jis pridedamas į eilės galą. Rinkėjas atlieka darbą tol kol šis sąrašas yra pilnas arba dėl kitų pašalinių dalių.
- **Puslapio siuntėjas** – šios dalies pagrindinė atsakomybė yra atsisiųsti puslapį iš adreso, kurį gauname iš šakų sąrašo. Siuntėjui reikalingas HTTP klientas, tam kad galėtų pateikti HTTP užklaudas ir perskaityti atsakymus. Reikia nepamiršti, kad būtinos ribotas laikas užklaudoms. Tai reikalinga tam, kad nebūtų švaistomas laikas gaunant milžiniškus failus ar renkant informaciją iš lėtų serverių.
- **Duomenų saugykla** – saugykla yra atsakinga už didelį duomenų kiekį, gautą iš puslapio siuntėjo. Ji turi gebėti saugoti ir valdyti didelį kiekį duomenų. Šioje saugykloje laikomas tik standartinis HTML kodas, visi kiti failai yra praleidžiami. Teoriškai tai beveik taspats kaip visos kitos duomenų saugyklos, tačiau rinkėjui skirta saugyklai nereikalingas didelis funkcionalumas, sąsajos ar svarbus vardinimas. Ji saugo gautus puslapius kaip atskirus failus.



1. pav. Crawler'io architektūra

Surinkėjo veikimas eina tokia seka:

- Registruojamas adresas ar adresai
- Išsaugomas adresas į šakų sąrašą
- Pasirenkamas adresas iš sąrašo
- Parsiunčiamas parinktas puslapis
- Išskaidomas puslapis ieškant papildomų adresų
- Neaplankyti adresai pridedami į šakų sąrašą.
- Baigus vėl vykdomas antrasis punktas ir kartojamas veikimo principas.



2. pav. Surinkėjo veiksmų seka.

Taip pat surinkėjas palaiko strategijų rinkinį:

- **Pasirinkimo strategija** – parinkimas kuriuos puslapius rinkti.
- **Pakartotinio apsilankymo strategija** – apibūdina kada reikia pakartotojinai patikrinti adresą siekiant atrasti pakitimus.
- **Mandagumo strategija** – kaip reikia rinkti informacija, kad neapkrautume puslapių, iš kurių ją renkame.
- **Paralelinė strategija** – teikia kaip valdyti naudojamus rinkėjus.

### 2.6.3. Surinkėjų tipai

Priklausomai nuo puslapio, kurį aplankome, jo tipo ar pasiekimo galimybių galime surinkėjus suskirstyti į kelis tipus. Keletas jų yra:

- **Didėjantis** – šis tipas yra skirtas atnaujinti jau egzistuojančius, pasiūstus puslapius. Atsižvelgiant į puslapio atsinaujinimo dažnumą nustatoma kaip dažnai reikia rinkti informaciją, taip palaikoma naujausia duomenų versija. Taip pat nuolatos keičiami mažiau svarbūs puslapiai į svarbesnius.
- **Orientuotas** – šis tipas skirtas siūsti tik tuos puslapius, kurie yra panašaus tipo. Rinkėjas renka informaciją orientotą vienai temai ar sričiai. Šis tipas dar vadinamas – temos tipu, dėl savo veikimo principo. Orientuotas rinkėjas nustato svarbą ir kryptį, nustato kaip susijusi tikrinamo puslapio informacija su reikiama tema ir tolimesnius veiksmus. Didžiausias šio tipo pliusas yra tas, kad žiūrint ekonomiškai, tai paspartina siuntimus ir sumažina tinklo apkrovą.
- **Paslėptas** – didelė dalis informacijos yra paslėpta duomenų bazėse, todėl ją galima pasiekti tik per tam tikras užklausas ar formas. Dauguma rinkėjų tiesiog eina per puslapius ir renka informaciją laisvai prieinamą visiems, be didesnių pastangų, praleidžiant puslapius pasiekiamus pildant formas ar atliekant paieškas. Šis rinkėjų tipas skirtas būtent rinkti informaciją, paslėptą už formų ar papildomų paieškų.
- **Lygiagretusis** – kadangi nuolat daugėja internetinių puslapių ir informacijos jose sunku gauti visą ar dalį puslapio vienu metu. Tam kuriamas lygiagretusis rinkėjas, naudojantis kelis rinkimo procesus lygiagrečiai, siekiant paspartinti gavimo procesą.

### 2.6.4. Gero rinkimo algoritmo ypatybės

Kadangi rinkėjų tipų ir galimybių yra daug, būtina išskirti ypatybes, apibūdinančias gerus algoritmus:

- **Greitis** (*Speed*) – įprastas rinkėjas gali surinkti apie 86,400 puslapių per dieną, todėl būtinas gebėjimas greitai apdoroti didelius kiekius duomenų.
- **Mandagumas** (*Politeness*) – kaip jau buvo minėta rinkėjas turi būti mandagus serverio atžvilgiu. Rinkėjas turėtų daryti pertraukas tarp užklausių, kad nebūtų apkrautas šaltinio serveris.
- **Praleidžiamas turinys** (*Excluded Content*) – kaikiuriuose puslapiuose yra paliekami failai „robot.txt“, kuriuose pateikiama informacija apie rinkimo ribojimus šiame puslapyje. Rinkėjas (ar jo kūrėjas) turi atsižvelgti į tokius reikalavimus.
- **Besidubliuojantis turinys** (*Duplicate content*) – rinkėjas turėtų gebėti atpažinti besikartojantį turinį skirtinguose šaltiniuose.
- **Nuolatinis rinkimas** (*Continuous Crawling*) – nuolatos rinkti tų pačių puslapių informaciją yra nenaudingas procesas. Reikėtų rinkti tik atnaujintus puslapius.
- **Šlamšto atmetimas** (*Spam Rejection*) – rinkėjas turėtų ignoruoti ir praleisti adresus, kurie vedą į reklaminius ar nenaudingus puslapius ir juos įtraukti į juodąjį sąrašą.

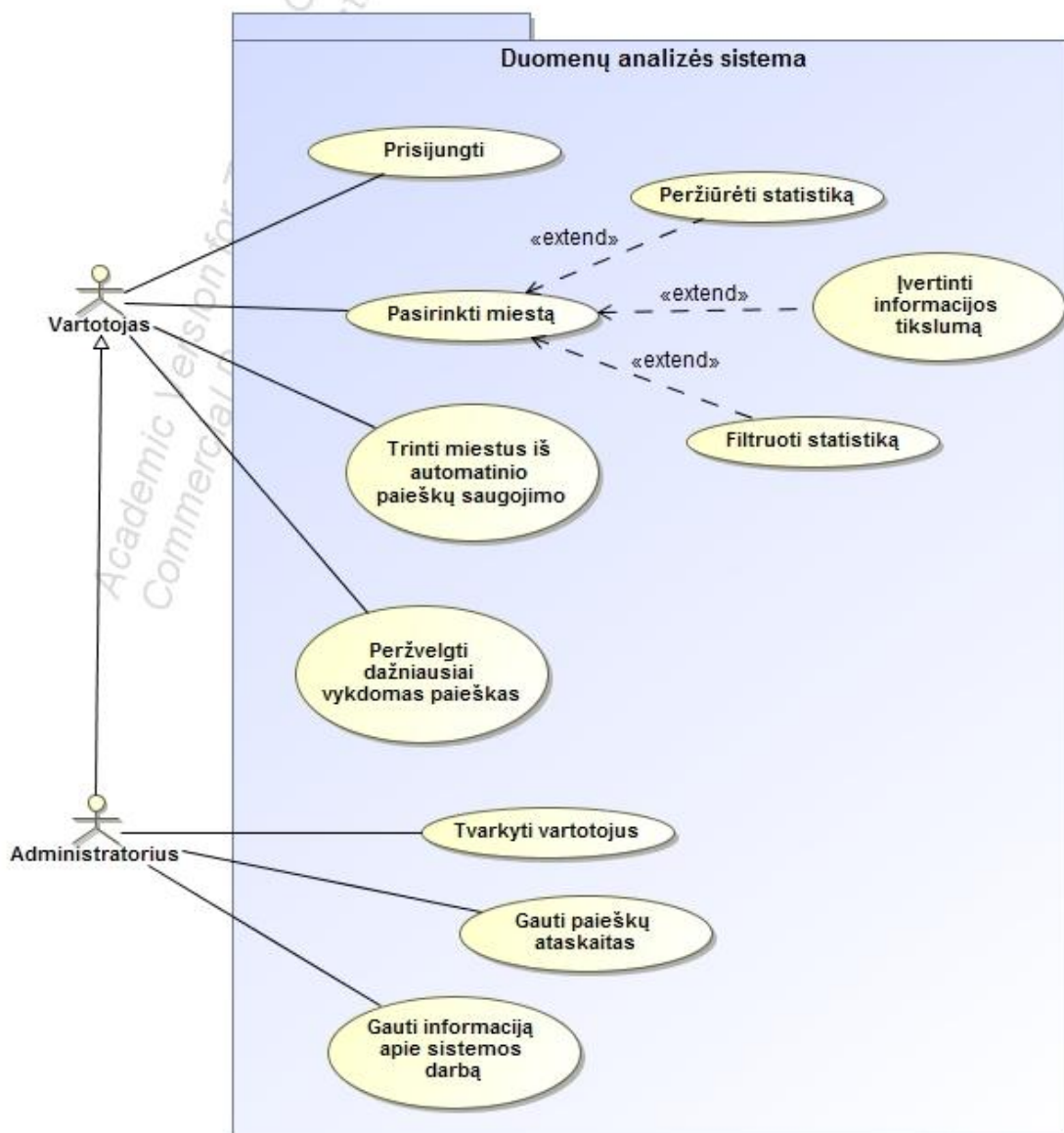
### 2.7. Išvados

- Kuriant sistemą ir siekiant gauti tikslesnius duomenis reikia daugiau nei dviejų šaltinių.
- Gaunama informacija turi būti be perstojo renkama, kad vartotojas bet kuriuo metu galėtų peržiūrėti naujausią galimą informaciją.
- Sistema turi gebėti tvarkytis su iškilusiomis grėsmėmis arba jas apeiti.
- Saugumas turi būti iškart po informacijos rinkimo svarbumo skalėje.
- Surinkėjai turi palaikyti bent kelias gero rinkimo algoritmo savybes.

### 3. PROJEKTINĖ DALIS

#### 3.1. Sistemos veiklos sfera

##### 3.1.1. Sistemos funkcijos



3. pav. Sistemos ribos

##### 3.1.2. Panaudojimo atvejų sąrašas

###### Panaudojimo atvejis 1: Prisijungti

**Aktorius:** Vartotojas

**Aprašas:** Procesas skirtas autorizuoti vartotoją ir suteikti jam teises

**Sužadinimo sąlyga:** Vartotojas bando atlikti veiksmus, kuriems reikalingas autorizavimas.

**Po sąlygos:** Vartotojas autorizuotas ir gavo teises naudotis sistemos paslaugomis.

**Panaudojimo atvejis 2: Pasirinkti miestą**

**Aktorius:** Vartotojas

**Aprašas:** Išgauti iš duomenų bazės duomenis apie miestą

**Prieš sąlyga:** Serveris veikia, sistema paruošta darbui.

**Sužadinimo sąlyga:** Vartotojas pasirenka arba įveda dominantį miestą.

**Po sąlygos:** Vartotojui pateikiama informacija pagal jo pageidaujamus kriterijus.

**Panaudojimo atvejis 3: Peržiūrėti statistiką**

**Aktorius:** Vartotojas

**Aprašas:** Procesas gauti statistinę informaciją apie pasirinktą miestą pagal tam tikrus kriterijus

**Sužadinimo sąlyga:** Vartotojas pareiškia norą matyti detalią statistiką.

**Prieš sąlyga:** Serveris veikia, sistema paruošta darbui. Parinktas miestas.

**Po sąlygos:** Vartotojui pateikiama informacija pagal jo pageidaujamus kriterijus.

**Panaudojimo atvejis 4: Įvertinti informacijos tikslumą**

**Aktorius:** Vartotojas

**Aprašas:** Procesas skirtas įvertinti sistemos darbą, tikintis tobulinimo.

**Prieš sąlyga:** Serveris veikia, sistema paruošta darbui. Vartotojas stebi statistiką.

**Sužadinimo sąlyga:** Vartotojas pareiškia norą įvertinti informacijos statistiką.

**Po sąlygos:** Vartotojo įvertinimas išsaugomas į duomenų bazę.

**Panaudojimo atvejis 5: Filtruoti statistiką**

**Aktorius:** Vartotojas

**Aprašas:** Procesas skirtas panaikinti arba patikslinti gautą informaciją.

**Prieš sąlyga:** Serveris veikia, sistema paruošta darbui. Vartotojas stebi statistiką.

**Sužadinimo sąlyga:** Vartotojas pareiškia norą gauti tikslesnę statistiką.

**Po sąlygos:** Vartotojui pateikiama tikslesnę informacija.

**Panaudojimo atvejis 6: Trinti miestus iš automatinio paieškų saugojimo.**

**Aktorius:** Vartotojas

**Aprašas:** Procesas skirtas panaikinti nereikalingas paieškas iš automatinio saugojimo sąrašo.

**Prieš sąlyga:** Serveris veikia, sistema paruošta darbui. Vartotojas prisijungęs.

**Sužadinimo sąlyga:** Vartotojas nori pašalinti miestus iš automatinio saugojimo sąrašo.

**Po sąlygos:** Vartotojo automatinio saugojimo sąrašas pakeistas.

**Panaudojimo atvejis 7: Peržvelgti dažniausiai vykdomas paieškas.**

**Aktorius:** Vartotojas

**Aprašas:** Procesas skirtas peržvelgti visų vartotojų dažniausiai vykdomas paieškas.

**Prieš sąlygą:** Serveris veikia, sistema paruošta darbui.

**Sužadinimo sąlyga:** Vartotojas pasirenka dažniausių paieškų sąrašą.

**Po sąlygos:** Vartotojui pateikiamas dažniausiai ieškomų miestų sąrašas.

**Panaudojimo atvejis 8: Tvarkyti vartotojus.**

**Aktorius:** Administratorius

**Aprašas:** Procesas administruoti ir reguliuoti esamus vartotojus.

**Prieš sąlygą:** Serveris veikia, sistema paruošta darbui. Administratorius prisijungęs.

**Sužadinimo sąlyga:** Administratorius pasirenka vartotojų administravimą.

**Po sąlygos:** Vartotojų sąrašas atnaujintas.

**Panaudojimo atvejis 9: Gauti paieškų ataskaitas.**

**Aktorius:** Administratorius

**Aprašas:** Procesas peržvelgti ir padaryti išvadas pagal tai kokios paieškos dažniausiai vykdomos.

**Prieš sąlygą:** Serveris veikia, sistema paruošta darbui. Administratorius prisijungęs.

**Sužadinimo sąlyga:** Administratorius pasirenka ataskaitų gavimą.

**Po sąlygos:** Administratorius gauna reikiamą informaciją.

**Panaudojimo atvejis 10: Informaciją apie sistemos darbą.**

**Aktorius:** Administratorius

**Aprašas:** Procesas skirtas peržvelgti ir analizuoti sistemos darbą. Galimų problemų stebėjimas. Tai palengvins sistemos vystymąsi.

**Prieš sąlygą:** Serveris veikia, sistema paruošta darbui. Administratorius prisijungęs.

**Sužadinimo sąlyga:** Administratorius pasirenka ataskaitų gavimą.

**Po sąlygos:** Administratorius gauna reikiamą informaciją.



## 3.2. Funkciniai reikalavimai ir reikalavimai duomenims

### 3.2.1. Funkciniai reikalavimai

Reikalavimas #:	1	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	1
Aprašymas	Sistema turi leisti prisijungti vartotojams				
Pagrindimas	Vartotojų teisių gavimas ir papildomų funkcijų vykdymas				
Šaltinis	Vartotojas				
Tikimo kriterijus	Galima bus prisijungti ir pagal vartotojo lygi gauti atitinkamas teises				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	2	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	1
Aprašymas	Sistema turi leisti vartotojams pakeisti arba priminti slaptažodį				
Pagrindimas	Vartotojas gali užmiršti slaptažodį.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas pamiršęs slaptažodį gales prašyti priminimo				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	3	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	1
Aprašymas	Sistema turi leisti vartotojams susikurti naują paskyrą				
Pagrindimas	Vartotojas gali aplankyti sistemą pirmą kartą.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas gales užsiregistruoti ir naudotis papildomomis funkcijomis				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	4	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	2
Aprašymas	Sistema turi leisti pasirinkti miestą				
Pagrindimas	Vartotojas turi turėti galimybę atlikti miesto paiešką				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas atlieka paiešką ir gauna duomenis pagal pasirinktą miestą				
Užsakovo tenkinimas	4	Užsakovo netenkinimas	5		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	5	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	2
Aprašymas	Sistema turi leisti keisti miesto pasirinkimą				
Pagrindimas	Vartotojas turi turėti galimybę atlikti miesto paiešką				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas gali atlikęs paiešką pakeisti miestą				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	3		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	6	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	3
Aprašymas	Vartotojas turi turėti galimybę peržiūrėti miesto statistiką				
Pagrindimas	Miesto statistikos peržiūrėjimas ir duomenų atvaizdavimas				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas gali matyti detalią statistiką				
Užsakovo tenkinimas	4	Užsakovo netenkinimas	5		
Priklausomybės	4	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	7	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	4
Aprašymas	Vartotojas turi turėti galimybę vertinti duomenų tikslumą ir atvaizdavimo patogumą				
Pagrindimas	Sistema turi būti lengvai suprantama				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas įvertina sistemą				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	3
Priklausomybės	6			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	8	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	4
Aprašymas	Vartotojas turi turėti galimybę parašyti komentarą apie sistemą				
Pagrindimas	Sistema turi būti vertinama tiksliai				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas įveda komentarą				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	3
Priklausomybės	6			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	9	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	5
Aprašymas	Sistema turi teikti galimybę vykdyti filtravimą				
Pagrindimas	Siekimas gauti tikslesnius reikiamus duomenis				
Šaltinis	Vartotojas				
Tikimo kriterijus	Duomenys filtruojami tiksliai				
Užsakovo tenkinimas	4			Užsakovo netenkinimas	4
Priklausomybės	6			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	10	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	6
Aprašymas	Sistema turi leisti vartotojam redaguoti savo ieškojimo sąrašus				
Pagrindimas	Galimybė naikinti nebereikalingus miestus				
Šaltinis	Vartotojas				
Tikimo kriterijus	Duomenys ištrinami				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	1
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	11	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	7
Aprašymas	Sistema turi leisti peržvelgti visų vartotojų dažniausiai ieškomų miestų ir kriterijų sąrašą.				
Pagrindimas	Galimybė peržiūrėti kitų vartotojų paieškas				
Šaltinis	Vartotojas				
Tikimo kriterijus	Matomos dažniausiai vykdomos paieškos				
Užsakovo tenkinimas	2			Užsakovo netenkinimas	1
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	12	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	8
Aprašymas	Sistema turi leisti matyti visus vartotojus				
Pagrindimas	Galimybė peržiūrėti ir pasirinkti reikiamą vartotoją				
Šaltinis	Administratorius				
Tikimo kriterijus	Matomi visi vartotojai				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	13	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	8
Aprašymas	Sistema turi leisti blokuoti vartotoją				
Pagrindimas	Galimybė užblokuoti kenkėjišką vartotoją				
Šaltinis	Administratorius				
Tikimo kriterijus	Pasirinkus vartotojo blokavimą – jis nebegali prisijungti, išvedamas pranešimas.				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	3		
Priklausomybės	12	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	14	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	8
Aprašymas	Sistema turi leisti vartotoją ištrinti				
Pagrindimas	Galimybė ištrinti neaktyvius vartotojus				
Šaltinis	Administratorius				
Tikimo kriterijus	Pasirinkus vartotojo trinimą – duomenys užsaldomi.				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	3		
Priklausomybės	12	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	15	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	8
Aprašymas	Sistema turi leisti suteikti vartotojui aukštesnes teises				
Pagrindimas	Galimybė vartotojus paversti administratoriais ir atvirkščiai				
Šaltinis	Administratorius				
Tikimo kriterijus	Pasirinkus vartotoją jam pasirenkamas naujas teisių lygis.				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	3		
Priklausomybės	12	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	16	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	9
Aprašymas	Sistema turi teikti ataskaitas				
Pagrindimas	Galimybė gauti ataskaitas apie vykdomas paieškas ir duomenų reikiamumą				
Šaltinis	Administratorius				
Tikimo kriterijus	Gaunamos suformuotos ataskaitos.				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

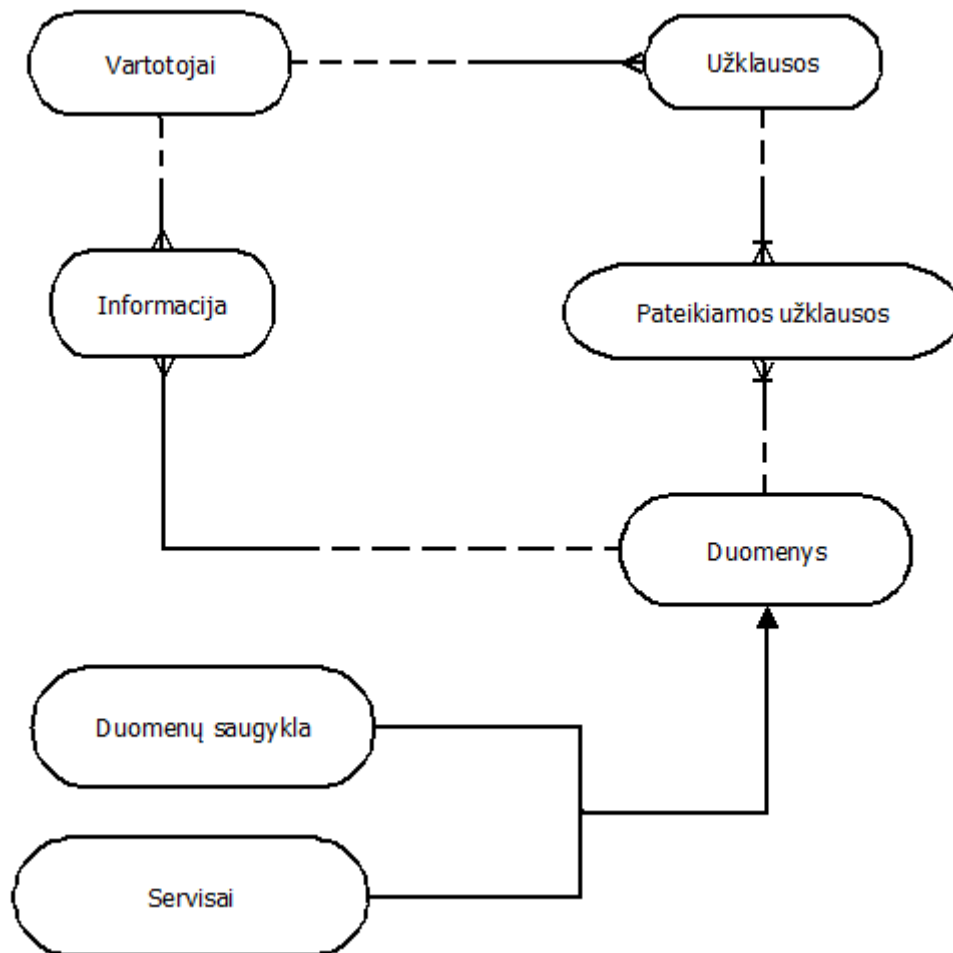
Reikalavimas #:	17	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	9
Aprašymas	Sistema turi keisti ataskaitas pagal pasirinktus kriterijus				
Pagrindimas	Galimybė keisti gaunamą ataskaitą pasirinkus atitinkamus kriterijus				
Šaltinis	Administratorius				
Tikimo kriterijus	Gaunamos ataskaitos pagal reikiamus kriterijus.				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	16			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	18	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	10
Aprašymas	Sistema turi teikti ataskaitas apie duomenų rinkimo procesą				
Pagrindimas	Galimybė stebėti sistemos darbo procesą				
Šaltinis	Administratorius				
Tikimo kriterijus	Gaunamos ataskaitos pagal reikiamus kriterijus.				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	19	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	10
Aprašymas	Sistema turi teikti galimybę peržvelgti kiekvienos dienos rinkimų ataskaitas				
Pagrindimas	Galimybė peržvelgti kaip duomenys renkami kiekvieną dieną.				
Šaltinis	Administratorius				
Tikimo kriterijus	Gaunamos ataskaitos pagal reikiamus kriterijus.				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	18			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	20	Reikalavimo tipas:	9a	Įvykis/panaudojimo atvejis #:	10
Aprašymas	Sistema turi pateikti klaidingų dalių informaciją				
Pagrindimas	Galimybė greitai eliminuoti susidariusias klaidas				
Šaltinis	Administratorius				
Tikimo kriterijus	Gaunami klaidų parametrai ir jų sukelėjai.				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	3
Priklausomybės	18			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.2.2. Reikalavimai duomenims



4. pav. Pradinis duomenų modelis

### 3.3. Nefunkciniai reikalavimai

#### 3.3.1. Reikalavimai sistemos išvaizdai

Reikalavimas #:	21	Reikalavimo tipas:	10	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi būti paprasta, neperkrauta ir lengvai suprantama				
Pagrindimas	Sistema naudosis įvairaus patyrimo lygio vartotojai				
Šaltinis	Vartotojas				
Tikimo kriterijus	Vartotojas be papildomų apmokymų naudojami sistema.				
Užsakovo tenkinimas	2	Užsakovo netenkinimas	5		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				



### 3.3.2. Reikalavimai panaudojamumui

Reikalavimas #:	22	Reikalavimo tipas:	11a	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi padėti supaprastinti duomenų išgavimą, nepadarrant klaidų				
Pagrindimas	Dėl vartotojų nepatyrimo klaidos ir nesugebėjimas naudotis duomenimis.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Duomenų išgavimas sumažintas, surenkant duomenis į vieną vietą. Klaidų skaičius tai atliekant sumažintas iki minimalaus.				
Užsakovo tenkinimas	2	Užsakovo netenkinimas	4		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.3.3. Reikalavimai vykdymo charakteristikoms

Reikalavimas #:	23	Reikalavimo tipas:	12	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Visi skaitiniai duomenis verčiami į tiksluosius skaičius				
Pagrindimas	Siekiant paprastumo skaičiai apvalinami iki sveikojo skaičiaus.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Visi skaičiai bus suapvalinti iki sveikojo skaičiaus.				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	2		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.3.4. Reikalavimai veikimo sąlygoms

Reikalavimas #:	24	Reikalavimo tipas:	13b	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi atnaujinti ryšį nutrūkus interneto ryšiui.				
Pagrindimas	Galimam ryšio nutrūkimui turi būti sudarytos sąlygos atsinaujinti.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Sugrįžus ryšiui – tęsiamas darbas.				
Užsakovo tenkinimas	3	Užsakovo netenkinimas	3		
Priklausomybės	Nėra	Konfliktai	Nėra		
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	25	Reikalavimo tipas:	13b	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi veikti visuose įrenginiuose				
Pagrindimas	Sistema naudotis gali įvairiais įrenginiais				
Šaltinis	Vartotojas				
Tikimo kriterijus	Sistemos dizainas prisitaiko prie įrenginio				
Užsakovo tenkinimas	1			Užsakovo netenkinimas	5
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.3.5. Reikalavimai sistemos priežiūrai

Reikalavimas #:	26	Reikalavimo tipas:	14d	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi veikti vienodai ir tikslingai, nepriklausomai nuo įrenginio				
Pagrindimas	Sistema gali naudotis įvairiais įrenginiais				
Šaltinis	Vartotojas				
Tikimo kriterijus	Sistemos dizainas prisitaiko prie įrenginio				
Užsakovo tenkinimas	3			Užsakovo netenkinimas	2
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	27	Reikalavimo tipas:	14d	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema turi veikti tolygiai. Neturi kilti problemų sistemos veikimui dėl vieno ar kelių tiekėjų pokyčių.				
Pagrindimas	Vienas ar keli servisai gali neveikti. Tai turi nekeisti sistemos darbo.				
Šaltinis	Vartotojas				
Tikimo kriterijus	Neveikiant vienam ar keliems servisams sistema toliau veikia.				
Užsakovo tenkinimas	4			Užsakovo netenkinimas	5
Priklausomybės	Nėra			Konfliktai	Nėra
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.3.6. Reikalavimai saugumui

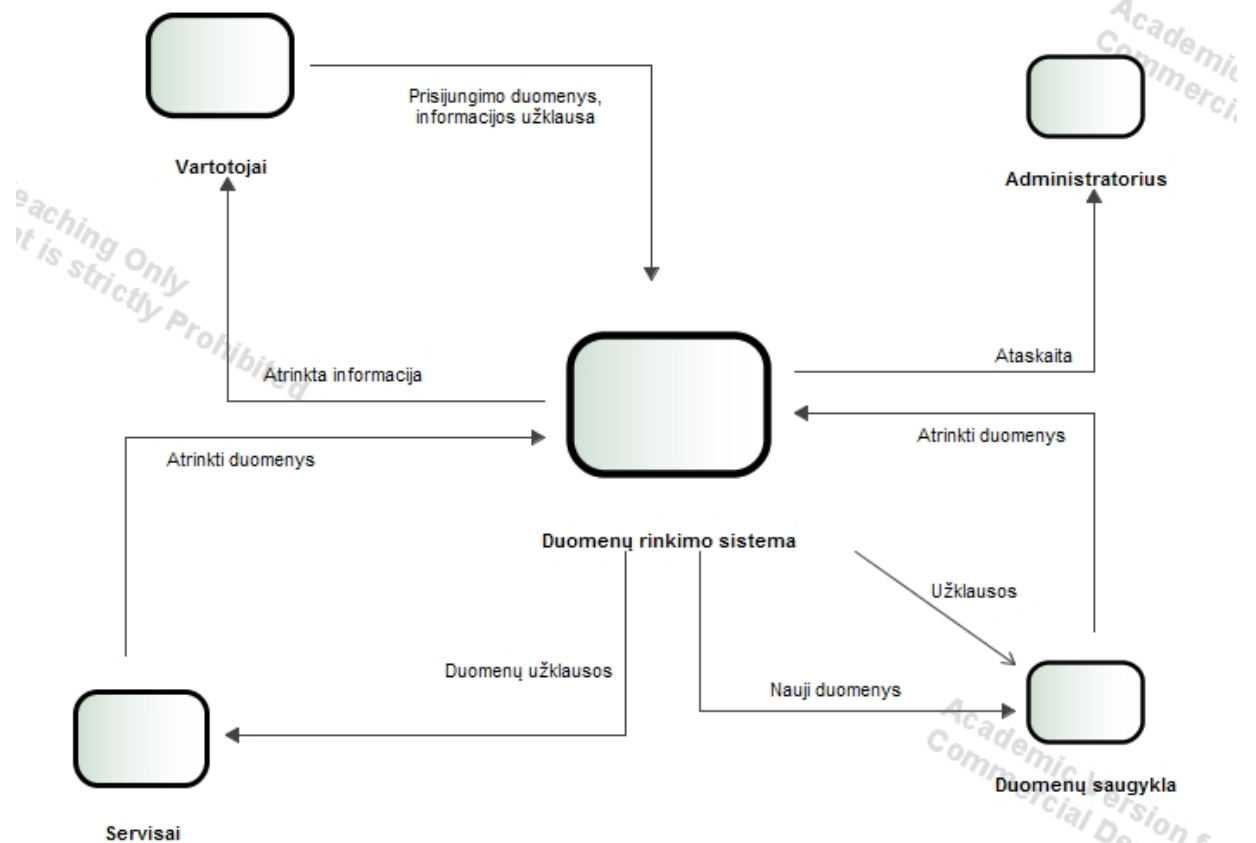
Reikalavimas #:	28	Reikalavimo tipas:	15a	Įvykis/panaudojimo atvejis #:	1-10
Aprašymas	Sistema neleis prisijungti neautorizuotiems asmenims				
Pagrindimas	Sistema apsaugos nuo neteisėto prisijungimo				
Šaltinis	Vartotojas				
Tikimo kriterijus	Įvedus neteisingus duomenis pranešama, kad prisijungti nepavyko.				
Užsakovo tenkinimas	3		Užsakovo netenkinimas	2	
Priklausomybės	Nėra		Konfliktai	Nėra	
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

Reikalavimas #:	29	Reikalavimo tipas:	15a	Įvykis/panaudojimo atvejis #:	8
Aprašymas	Administratorius galės blokuoti kenkėjiškus vartotojus				
Pagrindimas	Dėl galimų bandymų sugadinti sistemą administratorius turės galimybę blokuoti kenkėjiškus vartotojus				
Šaltinis	Administratorius				
Tikimo kriterijus	Administratorius užblokavo vartotoją				
Užsakovo tenkinimas	3		Užsakovo netenkinimas	2	
Priklausomybės	Nėra		Konfliktai	Nėra	
Papildoma medžiaga					
Istorija	Užregistruotas 2014 gegužės 10d.				

### 3.4. Veiklos kontekstas

Žemiau pateiktoje diagramoje matomas sistemos veiklos kontekstas. Diagramą galima išskaidyti į keturias pagrindines funkcijas:

- Vartotojai – vartotojas pateikia užklausą veiksmui atlikti, sistema apdoroja užklausą, patikrina jos tikrumą ir gražina sugeneruotą atsakymą.
- Administratorius – administratorius nors ir turi tas pačias funkcijas kaip vartotojai, yra atskirtas dėl galimybės gauti ataskaitas apie sistemos darbą.
- Servisai – tai sąveikavimo su informacijos tiekėjais dalis. Sistema tam tikrais laiko etapais kreipiasi į servisus užklaudama tam tikros informacijos. Tuo tarpu sistema gražina informaciją pagal paduotus reikalavimus.
- Duomenų saugykla – visiem prieš tai minėtiems veiksmams atlikti reikalinga duomenų saugykla. Vartotojui ar administratoriui paduodant užklausą duomenys yra atrenkami iš duomenų bazės ir gražinami tolimesniam apdorojimui. Tuo tarpu gražinta informacija iš servisų yra įrašoma į šią duomenų saugyklą.



5. pav. Konteksto diagrama

Žemiau pateiktoje lentelėje aprašomi veiklos padalinimai.

1. lentelė Veiklos padalinimas

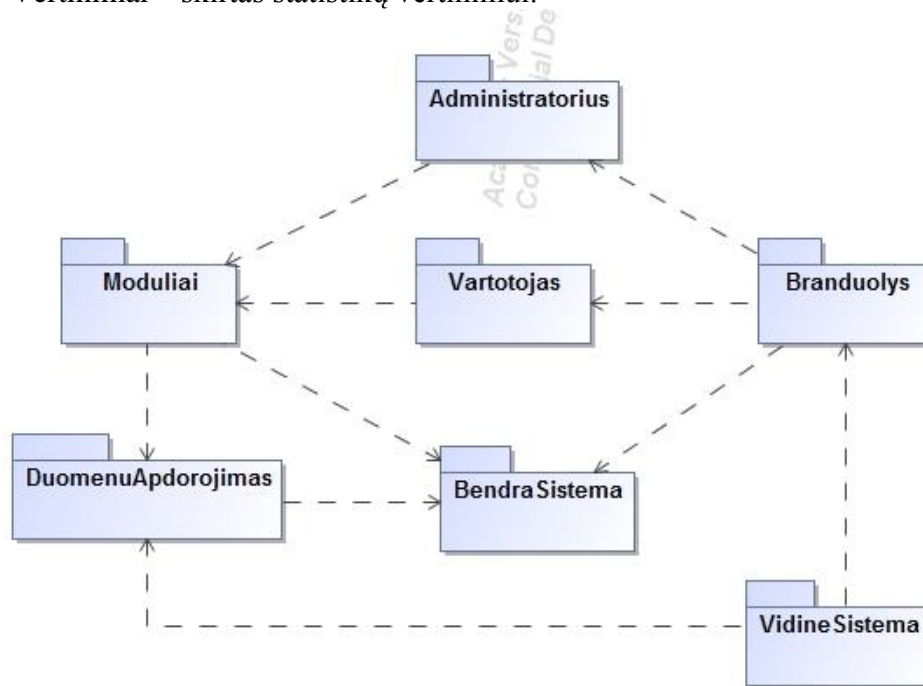
Nr.	Įvykis	Įeinantys/išeinantys srautai
1.	Informacijos pareikalavimas	Vartotojo duomenys, informacijos užklausa (in)
2.	Atrinktos informacijos gražinimas	Atrinkta informacija (out)
3.	Ataskaitos perdavimas administratoriui	Ataskaita (out)
4.	Reikiamų duomenų užklausa	Užklausa duomenys(out)
5.	Pagal užklausą atrinkti duomenys	Atrinkti duomenys (in)
6.	Naujai surinkti serviso duomenys	Nauji duomenys (out)
7.	Naujų duomenų pareikalavimas	Duomenų užklausa (out)
8.	Serviso surinkti duomenys	Atrinti duomenys (in)

### 3.5. Sistemos statinis vaizdas

Žemiau esančiame paveikslėlyje pateikiamas sistemos išskaidymas į septynis pagrindinius paketus:

- Vidinė sistema – skirta duomenų gavimui iš servisų.
- Branduolys – šis komponentas reikalingas susieti vartotojus su sistema.
- Administratorius – tai paketas skirtas atlikti administratoriaus funkcijas.
- Vartotojas – tai paketas skirtas vartotojo funkcijom atlikti.
- Duomenų apdorojimas – tai darbo su duomenų baze paketas.
- Bendra sistema – tai paketas, kuris naudojamas visoje sistemoje.
- Moduliai – šis paketas skirtas tam tikrom papildomom funkcijom atlikti ir gali būti išskaidytas į keturis sub-paketus:
  - Ataskaitos - skirtas ataskaitų kūrimui ir pateikimui.
  - Filtrai – skirtas duomenų filtravimui pagal pateiktus kriterijus.

- Miestai – skirtas duomenų išgavimui apie norimą miestą.
- Vertinimai – skirtas statistikų vertinimui.



6. pav. Sistemos išskaidymas į paketus

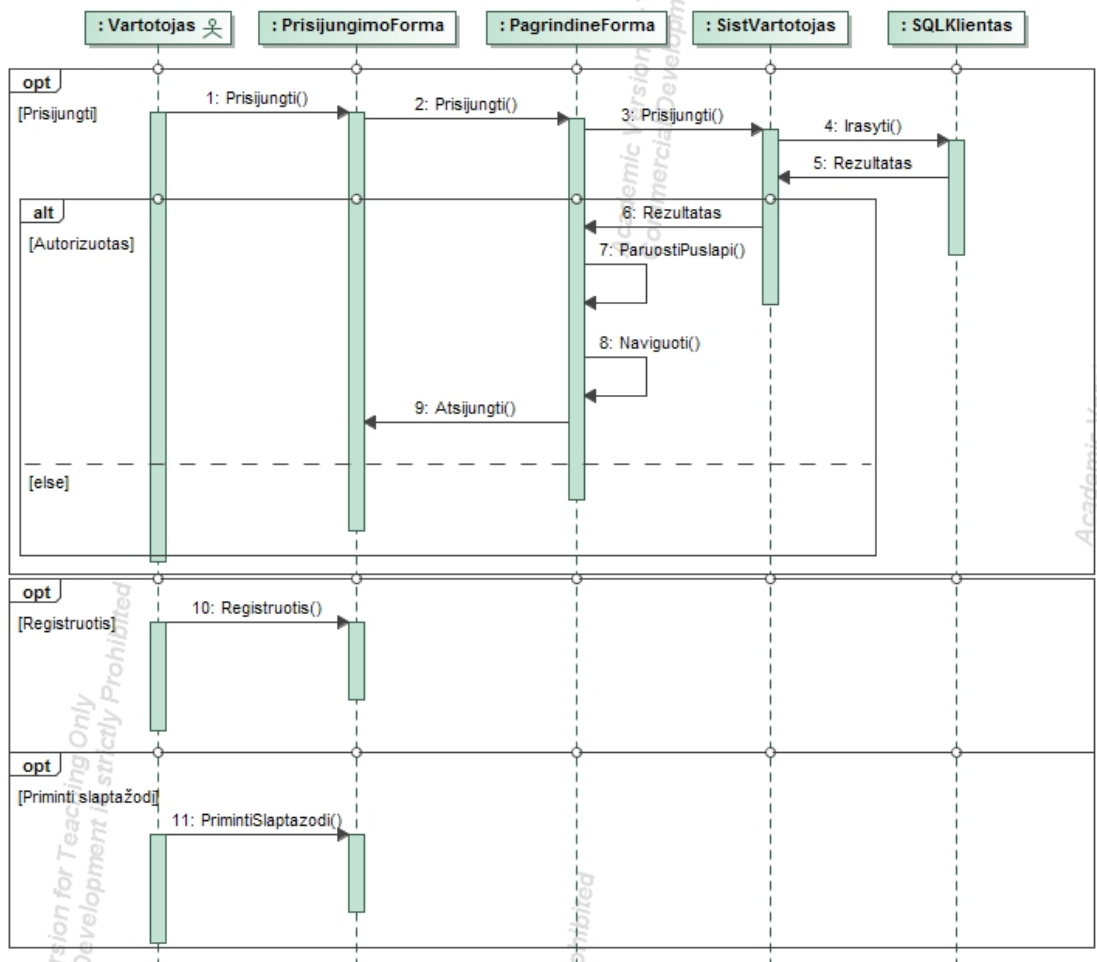
### 3.6. Sistemos dinaminis vaizdas

Šiame skyriuje vaizduojamos vienos veiksmų sekos diagramos. Seka buvo parinkta svarbumo atžvilgiu ir ji susideda iš:

- Vartotojo prisijungimas
- Miesto pasirinkimas
- Informacijos peržvelgimas
- Informacijos filtravimas

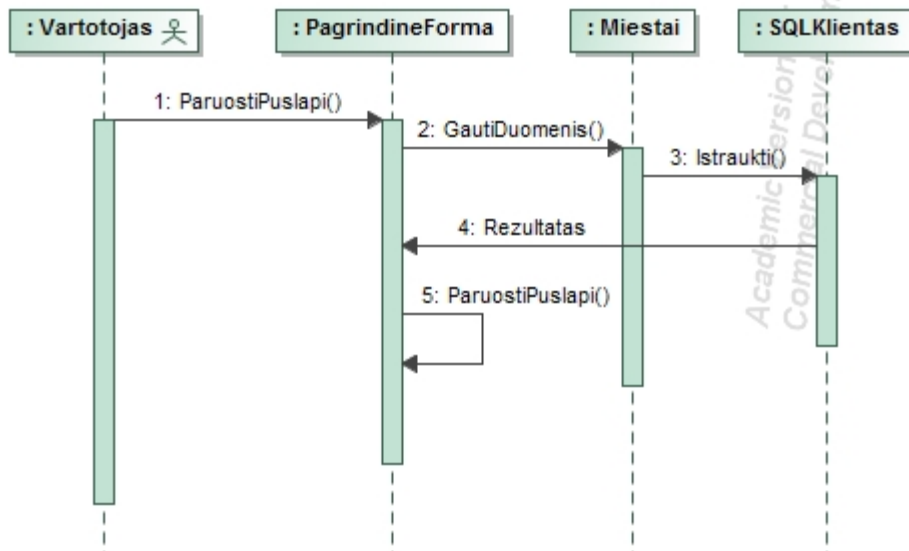
#### 3.6.1. Sekų diagramos

Šioje diagramoje matomas vartotojo prisijungimas. Taip pat sistema suteikia galimybę vartotojui registruotis ar prisiminti slaptažodį.



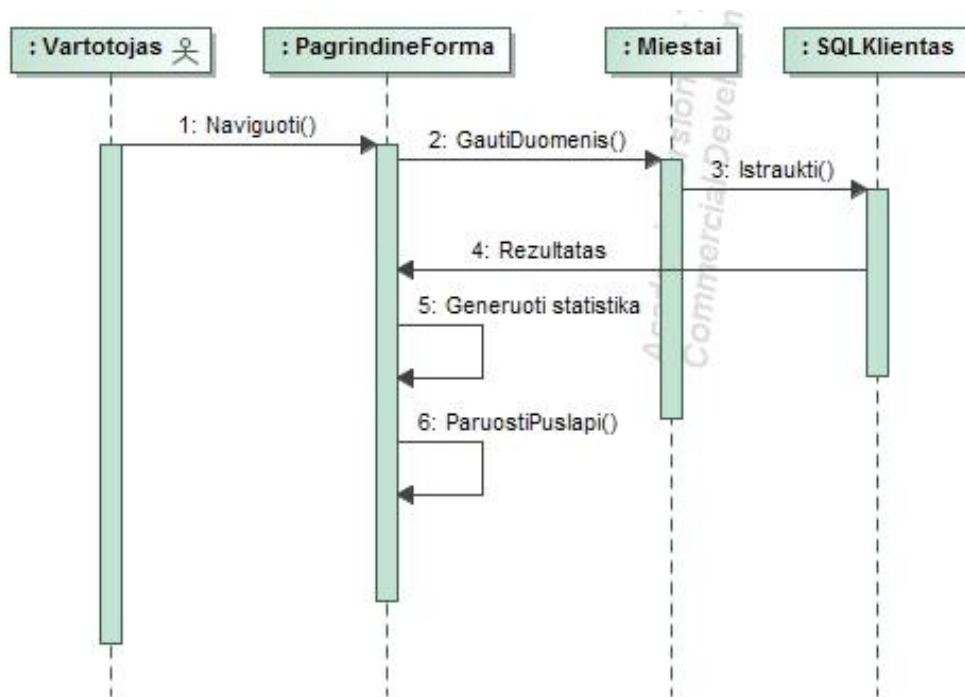
7. pav. Prisijungimo PA sekų diagrama

Šioje diagramoje pateikiama sistemos veiksmų seka vartotojui pasirinkus miestą.



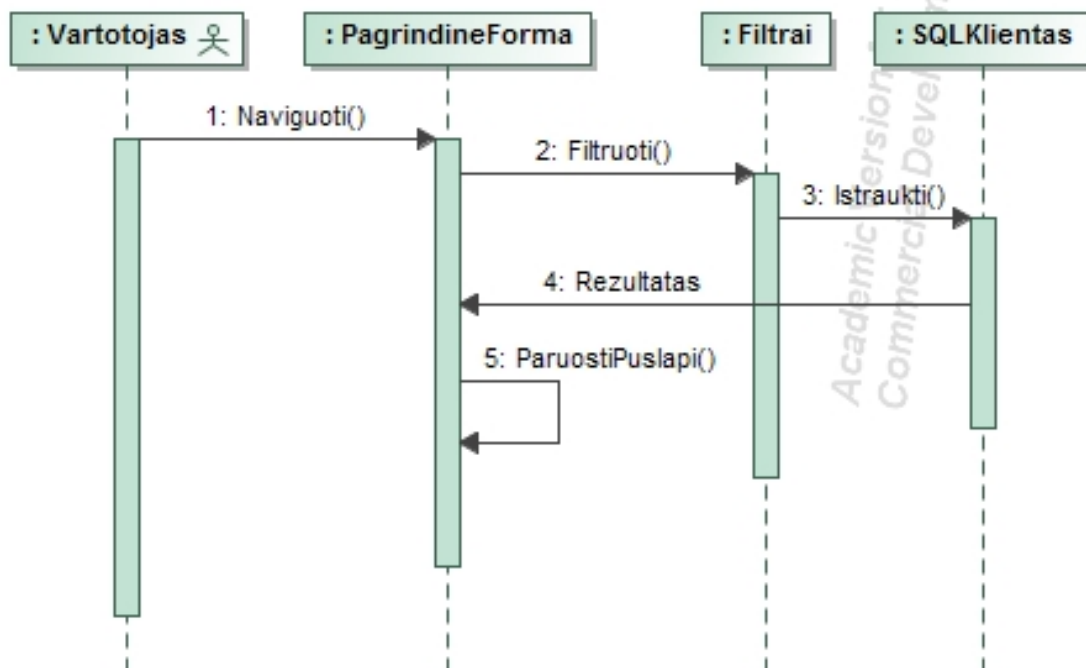
8. pav. Pasirinkti miestą PA sekų diagrama

Šioje diagramoje pateikiama sistemos veiksmų seka vartotojui pateikiant informaciją peržiūrėti.



9. pav. Peržiūrėti statistiką PA sekų diagrama

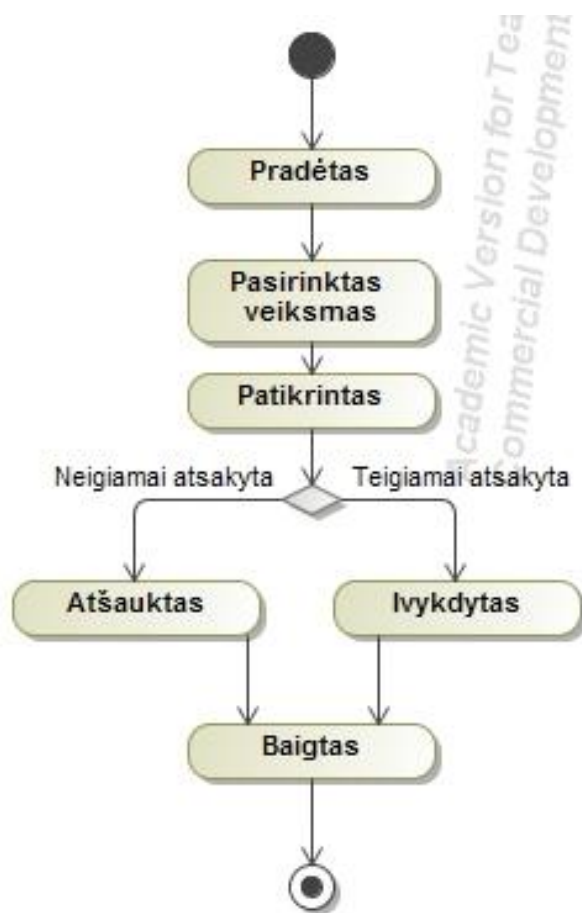
Šioje diagramoje pateikiama sistemos veiksmų seka vartotojui filtruojant gautą informaciją.



10. pav. Filtruoti statistiką PA sekų diagrama

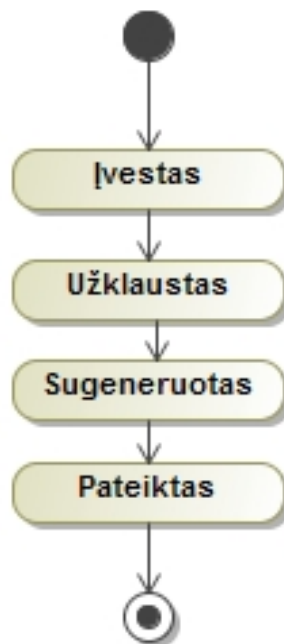
### 3.6.2. Būsenų diagramos

Šioje diagramoje matomas vartotojo prisijungimas. Pateikiamos visos sistemos būsenos atliekant prisijungimą.



11. pav. Prisijungimo PA būsenų diagrama

Šioje diagramoje pateikiamos sistemos būsenos vartotojui pasirinkus miestą.



12. pav. Pasirinkti miestą PA būsenų diagrama

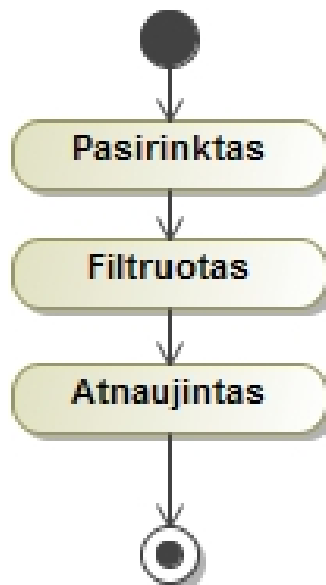


Šioje diagramoje pateikiamos sistemos būsenos vartotojui pateikiant informaciją.



**13. pav.** Peržiūrėti statistiką PA būsenų diagrama

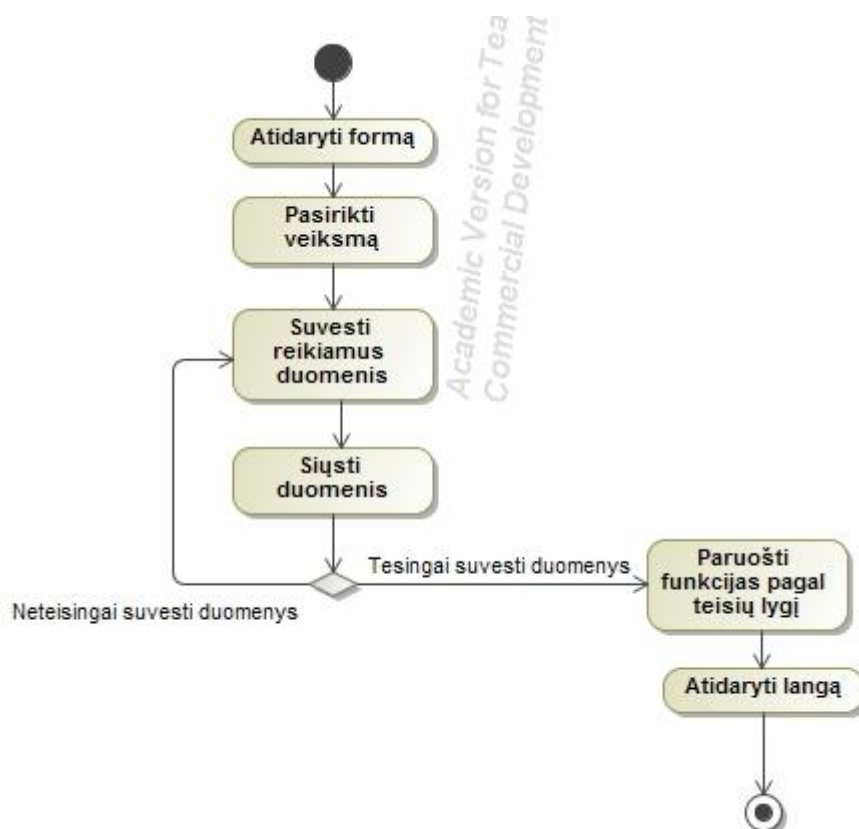
Šioje diagramoje pateikiamos sistemos būsenos vartotojui filtruojant gautą informaciją.



**14. pav.** Filtruoti statistiką PA būsenų diagrama

### 3.6.3. Veiklos diagramos

Šioje diagramoje matomas vartotojo prisijungimas. Pateikiama sistemos veikla atliekant prisijungimą.



15. pav. Prisijungti PA veiklos diagrama

Šioje diagramoje pateikiama sistemos veikla vartotojui pasirinkus miestą.



16. pav. Pasirinkti miestą PA veiklos diagrama

Šioje diagramoje pateikiama sistemos veikla vartotojui pateikiant informaciją.



17. pav. Peržiūrėti statistiką PA veiklos diagrama

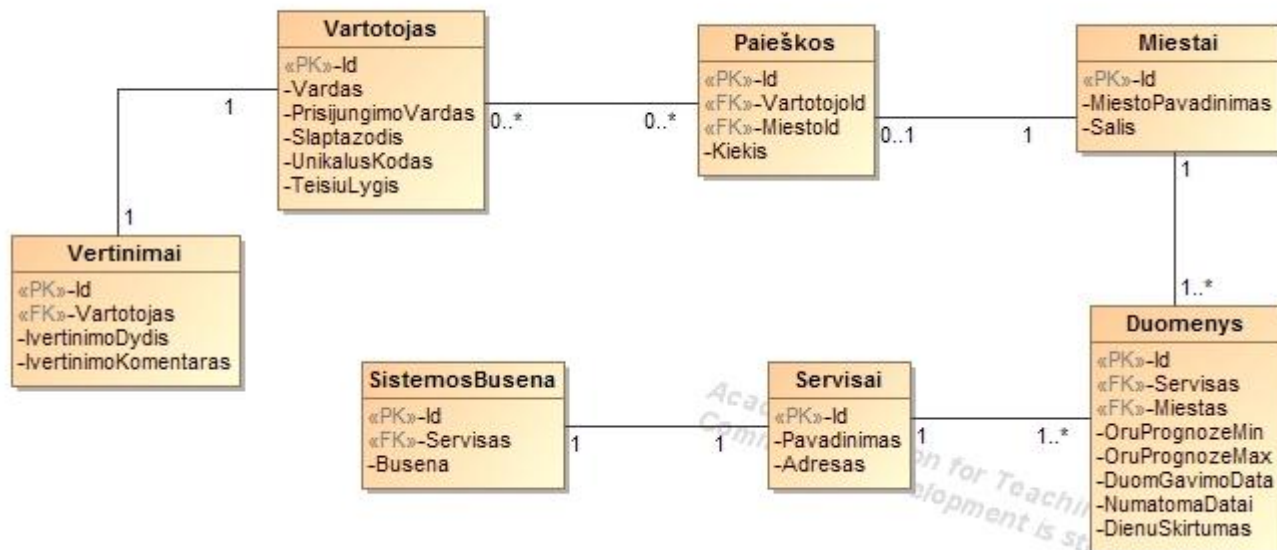
Šioje diagramoje pateikiama sistemos veikla vartotojui filtruojant gautą informaciją.



18. pav. Filtruoti statistiką PA veiklos diagrama

### 3.7. Duomenų vaizdas

Šiame skyriuje pateiktoje diagramoje vaizduojamas sistemos duomenų vaizdas. Visi duomenys yra tarpusavyje susiję vienaip ar kitaip. Pavyzdžiui – duomenys apie miestą yra susiję su vartotoju per vartotojo paieškas ir t.t.



19. pav. Duomenų bazės modelis

## 4. TYRIMO DALIS

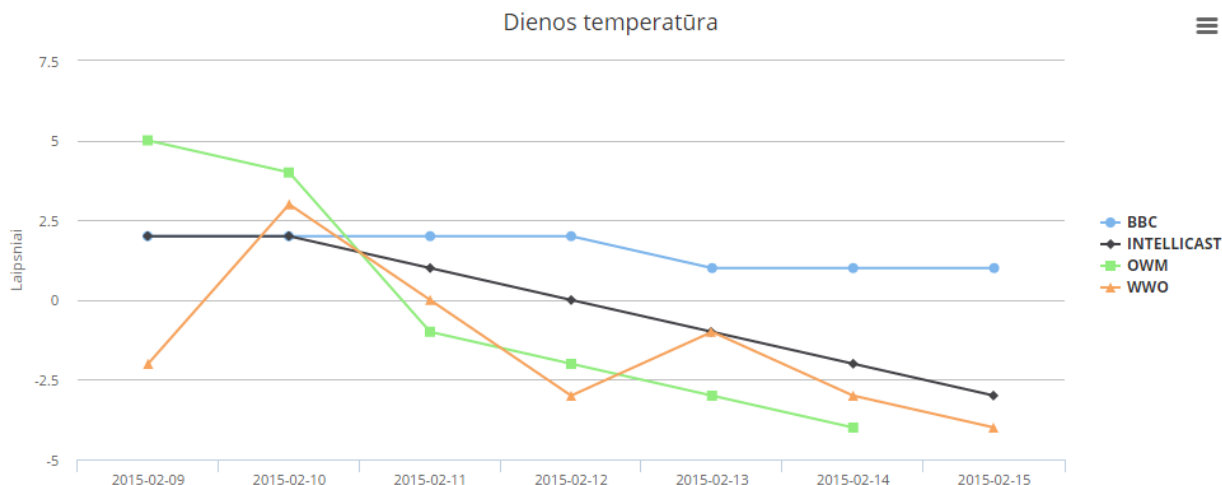
### 4.1. Tikslas

Renkami duomenys dažnai būna netikslūs, ypač jeigu skirtingi tiekėjai pateikia spėjimus. Šiuo tyrimu siekiama apskaičiuoti galimas paklaidas ir vidurkius. Šie skaičiavimai padės nustatyti tikslesnes orų prognozes.

### 4.2. Tyrimo reikšmė

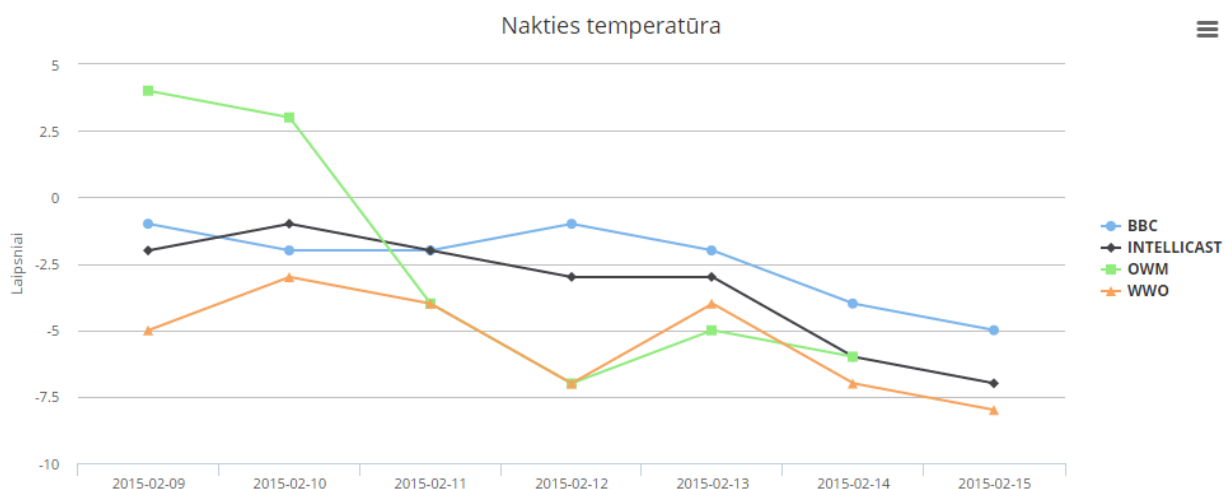
Sistemai renkant ir pateikiant duomenis buvo pastebėti dideli temperatūrų nuokrypiai. Nuokrypių pavyzdžiai dieną ir naktį yra pateikti žemiau esančiuose paveikslėliuose. Tai puikiai atvaizduoja gautų duomenų netikslumus. Šiuo tyrimu siekiama atrinkti tikslesnius temperatūrų duomenis ir išanalizuoti tiekėjų informacijos pastovumą. Tyrimui buvo parinktas Kauno miestas.

## Kaunas



20. pav. Kauno temperatūra vasario 16 dienai

## Kaunas



21. pav. Kauno temperatūra vasario 16 nakčiai

Kaip iš diagramų matoma nuokrypiai yra ganėtinai dideli todėl būtinas tolimesnis analizavimas. Apskaičiavus minimalius ir maksimalius nuokrypius žinosime pagrindinius režius tarp kurių yra tiksliausia data. Taip pat turėtų būti apskaičiuotas temperatūrų vidurkis, parodantis tikslesnę temperatūrą.

### 4.3. Tyrimo eiga

Pirmiausia pasirenkamos kelios atsitiktinės datos. Siekiant neapkrauti sistemos duomenų apdorojimo, tačiau turėti pakankamai duomenų išvadoms, pasirinktos penkios datos: šių metų vasario 16, kovo 11, kovo 30, balandžio 15, gegužės 1. Kiekvienai datai sugeneruojamos dienos ir nakties diagramos. Kiekvienai dienai generuojamos septynios prognozės, tai yra – kiekviena diena renkamos prognozės septynioms ateinančiom dienom, todėl analizei imamos orų prognozės šiai dienai pranašautos per praėjusias septynias dienas.

Kiekvienai dienai generuojamos diagramos, vaizduojančios temperatūras. Sugeneravus diagramas informacija atrenkama ir surašoma į lenteles skirtas kiekvienam tiekėjui, apskaičiuojami vidurkiai.

**2. lentelė** BBC Temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Nakties min	Nakties max	Nakties vidurkis
2015-02-16	1	2	1.57	-5	-1	-2.43
2015-03-11	10	7	8	-1	5	2.43
2015-03-30	8	9	8.29	2	3	2.43
2015-04-15	6	10	8.29	1	3	2
2015-05-01	13	15	14.14	4	8	6.43

**3. lentelė** Intellicast temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Nakties min	Nakties max	Nakties vidurkis
2015-02-16	-3	2	-0.14	-7	-1	-3.43
2015-03-11	7	8	7.43	1	4	2.86
2015-03-30	8	10	8.57	3	6	4.43
2015-04-15	9	12	11	1	4	2.71
2015-05-01	13	16	14.67	4	7	6.17

**4. lentelė** Open weather map temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Nakties min	Nakties max	Nakties vidurkis
2015-02-16	-4	5	-0.17	-7	4	-2.5
2015-03-11	4	8	5.86	-2	4	0.86
2015-03-30	7	10	9	2	9	5.33
2015-04-15	5	12	9.67	6	9	7.83
2015-05-01	9	15	12.33	4	13	6.33

**5. lentelė** World Wether Online temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Nakties min	Nakties max	Nakties vidurkis
2015-02-16	-4	3	-1.43	-8	-3	-5.43
2015-03-11	6	9	7.8	1	3	2.2
2015-03-30	8	12	9.17	3	4	3.67
2015-04-15	10	15	11.71	1	6	4
2015-05-01	14	18	16.71	2	7	5.71

Surašius pasirinktų dienų duomenis ir apskaičiuavus vidurkius matomas didelis temperatūrų pokytis bandant pranašauti orus. Tam skaičiuojamas prognozių temperatūros vidurkis, kuri padeda nustatyti tikslesnę temperatūrą.

#### 4.4. Tyrimo išvados

Tyrimo metu atliktame pasirinktų datų tyrime buvo nustatyti smarkūs temperatūrų šuoliai tarp mažiausios ir didžiausios temperatūrų. Siekiant nustatyti tikslesnes prognozes buvo apskaičiuotas visų dienų temperatūrų vidurkis. Buvo pastebėta dažnai pasitaikanti savybė – dienos vidurkis dažniau yra arčiau maksimalios temperatūros, o nakties arčiau minimalios.

### 5. EKSPERIMENTINĖ DALIS

#### 5.1. Tikslas

Eksperimento pagrindinis tikslas – patikrinti sistemos veikimą su realiais duomenimis. Tyrime parinktos ir apskaičiuotos prognozių reikšmės sugretinamos su realiomis tos dienos prognozėmis, pateiktomis orų prognozių dienoraščio - <http://www.gismeteo.ru/diary/>.

#### 5.2. Eksperimento aprašas

Eksperimentas atliekamas dviem etapais:

1. Pirmasis eksperimentas skirtas patikrinti duomenų tikrumą su duomenimis, naudotais tyriamajame darbe.
2. Antrasis eksperimentas – kiekybinis eksperimentas. Tikrinamas duomenų tikslumas didesniame laiko tarpe.

#### 5.2 Eksperimentas 1

Duomenų rinkimas yra labai nepastovus procesas, duomenys nuolat kinta, jų pateikimo struktūra nepastovi. Tuo labiau jeigu duomenys yra spėjami jų tikrumą nustatyti darosi vis sudėtingiau. Dėl tokių problemų patikimus tiekėjus atrinkti darosi vis sunkiau ir sunkiau. Kad aptikti tinkamus tiekėjus neužtenka patikrinti jų teikiamų paslaugų ar duomenų kokybę, reikia patikrinti ir atitikimą su realiais duomenimis. Šiame eksperimente tikrinami darbe parinktų keturių tiekėjų pateikiama informacija ir jos tikslumas. Analizuojama informacija – tai dienos, kurios jau praėjo ir turime realią informaciją kokia temperatūra iš tikrųjų buvo. Tačiau dėl galimų matavimo prietaisų vietovės duomenys gali būti taip pat klaidingi.

Eksperimente naudojamos tos pačios datos, parinktos sistemos tyrime, aprašytame šio dokumento 4 skyriuje. Kiekviena parinkta data patikrinama orų prognozių dienoraštyje ir palyginama su numatytomis prognozėmis. Patikrinama ar reali temperatūra yra tarp spėjamų temperatūrų rėžių. Taip pat apskaičiuojamas nuokrypis nuo apskaičiuoto vidurkio.

6. lentelė BBC Temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Tikra	Nakties min	Nakties max	Nakties vidurkis	Tikra
2015-02-16	1	2	1.57	-2	-5	-1	-2.43	-5
2015-03-11	10	7	8	7	-1	5	2.43	5
2015-03-30	8	9	8.29	9	2	3	2.43	6
2015-04-15	6	10	8.29	12	1	3	2	7
2015-05-01	13	15	14.14	16	4	8	6.43	12

7. lentelė Intellicast temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Tikra	Nakties min	Nakties max	Nakties vidurkis	Tikra
2015-02-16	-3	2	-0.14	-2	-7	-1	-3.43	-5

2015-03-11	7	8	7.43	7	1	4	2.86	5
2015-03-30	8	10	8.57	9	3	6	4.43	6
2015-04-15	9	12	11	12	1	4	2.71	7
2015-05-01	13	16	14.67	16	4	7	6.17	12

**8. lentelė** Open weather map temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Tikra	Nakties min	Nakties max	Nakties vidurkis	Tikra
2015-02-16	-4	5	-0.17	-2	-7	4	-2.5	-5
2015-03-11	4	8	5.86	7	-2	4	0.86	5
2015-03-30	7	10	9	9	2	9	5.33	6
2015-04-15	5	12	9.67	12	6	9	7.83	7
2015-05-01	9	15	12.33	16	4	13	6.33	12

**9. lentelė** World Wether Online temperatūrų lentelė

Data	Dienos min.	Dienos max	Dienos vidurkis	Tikra	Nakties min	Nakties max	Nakties vidurkis	Tikra
2015-02-16	-4	3	-1.43	-2	-8	-3	-5.43	-5
2015-03-11	6	9	7.8	7	1	3	2.2	5
2015-03-30	8	12	9.17	9	3	4	3.67	6
2015-04-15	10	15	11.71	12	1	6	4	7
2015-05-01	14	18	16.71	16	2	7	5.71	12

Eksperimento metu gautuose rezultatuose matome, kad iš tiekėjų gauti rezultatai kartais neatitinka realių temperatūrų, taip pat pastebėta tendencija, kad reali temperatūra yra artimesnė maksimaliai temperatūrai, nei vidurkinei. Tikslumo rezultatai pateikti žemiau esančioje lentelėje. Rezultatų grafines išraiška pateikta prieduose, 9.1 skyriuje.

Eksperimentas parodė, kad tiekėjai nėra labai tikslios informacijos tiekėjai. Tačiau gauta informacija suteikia daug žinių apie informacijos nuokrypius. Galima daryti prielaidas apie kiekvina tiekėją ir šio eksperimento rezultatus pritaikyti tolimesniam sistemos plėtojimui.

Siekiant išsiaiškinti kaip tiksliai tiekėjai pateikia informaciją buvo sudaryta tikslumų įvertinimo lentelė, pateikta žemiau. Kadangi datos buvo parinktos penkios, maksimalus dienos ir nakties įvertinimas yra penki balai.

**10. lentelė** Tikslumų įvertinimo lentelė

Tiekėjas	Dienos tikslumas	Nakties tikslumas
BBC	2/5	2/5
Intellicast	5/5	2/5
OWM	4/5	4/5
WVO	5/5	1/5



Atlikus tikslumo vertinimą nustatyta, kad tiksliausiai dienos prognozes pateikia Intellicast ir World Weather Online tiekėjai. Nakties duomenis tiksliausiai pateikia Open weather map tiekėjas. Tuo tarpu bendras duomenų prognozės tikslumas geriausiai atspindimas Open weather map tiekėjo pateiktuose duomenyse. Prasčiausiai pasirodė BBC tiekėjas. Galime daryti prielaidą, kad siekiant išgauti tiksliausius duomenis šiai dienai reikėtų imti duomenis dienos temperatūrai iš Intellicast arba WWO, o nakties duomenim iš OWM tiekėjų. Taip būtų gaunami tiksliausi duomenys.

### 5.3. Eksperimentas 2

Eksperimentas atliekamas siekiant sužinoti pateiktų duomenų tikslumo vertinimą lyginant su realiais duomenimis. Pasirinkta 60 dienų pradedant vasario šešioliktąja. Šis eksperimentas parodo kaip per ilgą laiko tarpą teisingai nuspėja tiekėjai galimą temperatūrą. Kiekvienai parinktai datai paimamos septynių dienų prognozės ir apskaičiuojamos minimalios ir maksimalios reikšmės. Temperatūros lyginimas atliekamas naudojant paprastąją formulę:

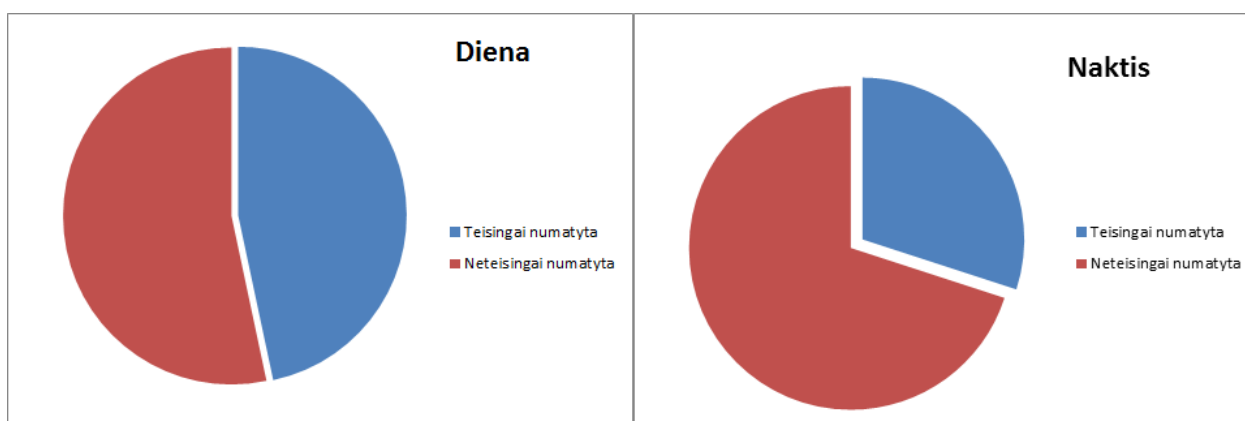
$$\text{MinTemp} < \text{RealTemp} < \text{MaxTemp}$$

MinTemp – minimali dienos ar nakties temperatūra

RealTemp – reali dienos ar nakties temperatūra, patikrinta „Gismeteo“ dienoraštyje

MaxTemp – maksimali dienos ar nakties temperatūra.

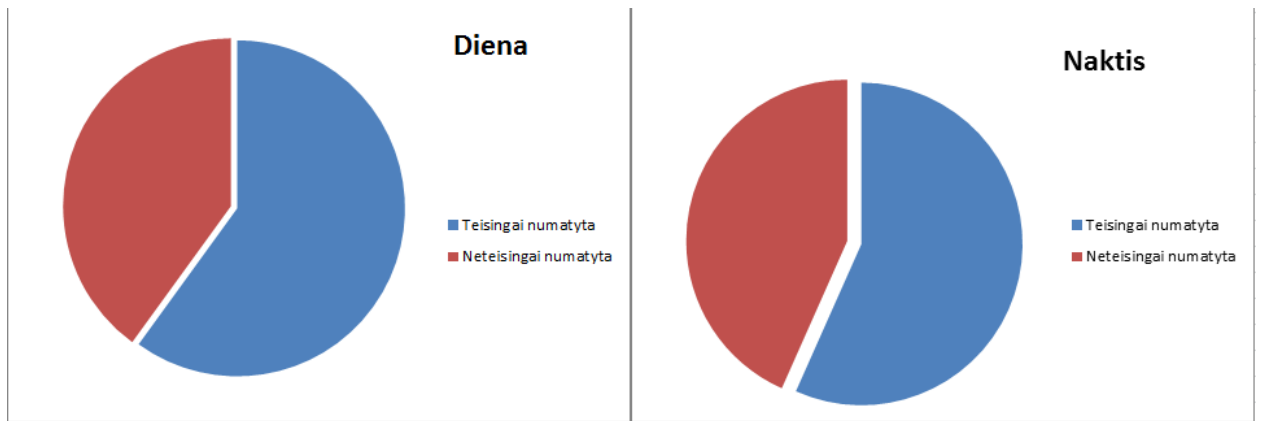
Apskaičiuoti ir susumuoti rezultatai diagramos formomis pateikia žemiau esančiose diagramose.



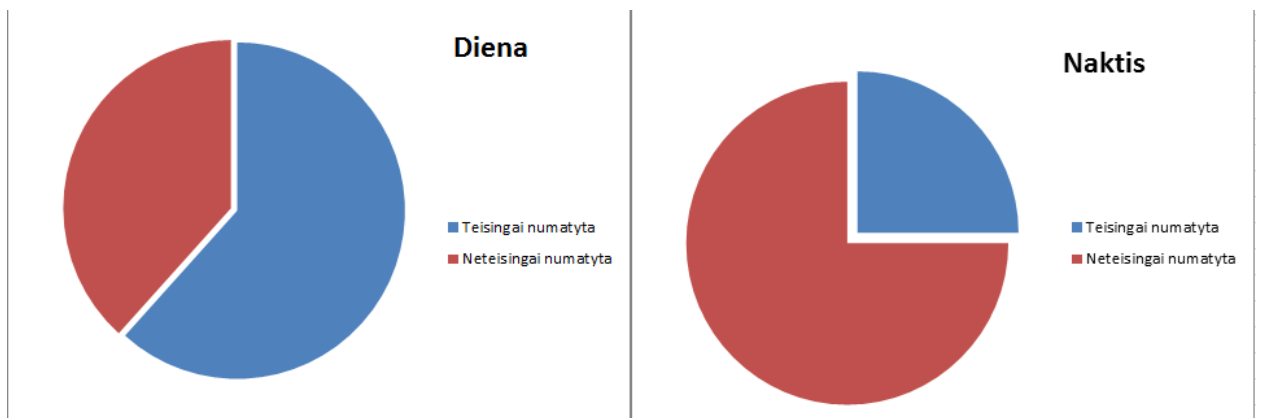
22. pav. BBC dienos ir nakties ilgalaikiai tikslumai



23. pav. Intellicast dienos ir nakties ilgalaikiai tikslumai



24. pav. OWM dienos ir nakties ilgalaikiai tikslumai



25. pav. WWO dienos ir nakties ilgalaikiai tikslumai

#### 5.4. Eksperimento rezultatai

Atlikus eksperimentus ir išanalizavus duomenis buvo nustatyti dideli neatitikimai tarp realių temperatūrų pastebėtos tendensijos tarp duomenų pateikimo. Eksperimente nr.1 prasčiausiai pasirodė BBC surinkęs 4 balus iš 10 galimų, tuo tarpu OWM surinko daugiausiai – 8 iš 10 galimų.

Skirstant į atskirus dienos ir nakties matavimus, atskirų dienų matavimuose geriausiai pasirodė WWO ir Intellicast tiekėjai, pateikę tiksliausias dienos temperatūras, tuo tarpu nakties – OWM, surinkęs 4 balus iš 5 galimų.

Apžvelgiant ilgalaikius temperatūrų matavimus geriausiai pasirodė OWM informacijos tiekėjas surinkęs 72 balus iš galimų 120 balų. Prasčiausiai vėlgi pasirodė BBC surinkęs 46 balus iš 120 galimų balų.

Apibendrinant gautus rezultatus, detalai pateiktus pirmame ir antrame eksperimentuose, priimta, kad prasčiausiai visuose matavimuose pasirodė BBC tiekėjas. Intellicast, nors ir pasirodęs gerai atrinktų datų tikslumo matavime, ilgalaikiame eksperimente pasirodė prastai. Tačiau geriausiai atrinktų datų eksperimente pasirodęs OWM ilgalaikiame tyrime parodė geriausią rezultatą.

Priimta išvada, kad siekiant rasti tiksliausią tiekėją neužtenka pasiremti vienu matavimu. Šiuose eksperimentuose nustatytas tiksliausias tiekėjas – Open Weather Map.

## 6. IŠVADOS

1. Išanalizavus sistemos poreikį ir esamus sisteminius sprendimus buvo sukurta duomenų rinkimo sistema skirta orų prognozių rinkimui. Sistema sukurta atitinkanti geras algoritmų savybes ir pasirinkus orientuotą rinkėjo tipą.
2. Projektavimo metu suprojektuota sistema, gebanti rinkti, apdoroti ir patogiu formatu pateikti duomenis reikiamai temai. Apžvelgti funkciniai ir nefunkciniai reikalavimai, sudarytos reikiamos sekų ir veiklos diagramos. Sudaryti duomenų sąryšiai. Projektavimo etape gautos diagramos buvo panaudotos kuriant sistemą.
3. Tyrimo metu ilga laiką rinkti duomenys atsitiktiniu būdu išrinkti ir patikrinti. Pasirinktos kelios datos ir patikrintas gautų duomenų tikslumas. Išanalizuoti rezultatai parodė tiekėjų duomenų neatitikimus ir nuokrypius.
4. Eksperimento metu surinkti duomenys buvo patikrinti su realiais duomenimis, pateiktais internetinio orų dienoraščio. Buvo nustatytas didelis neatitikimas tarp tikrų duomenų ir prognozių. Nustatyti geriausi ir praščiausi tiekėjai tolimesniam programos tobulinimui.
5. Eksperimento metu dviem būdais buvo ištestuoti tiekėjų pateikiami duomenys ir nustatytas naudingiausias tiekėjas ir duomenų tikslumas.

## 7. LITERATŪRA

- [1] GNU General Public License, prieiga internete  
<https://www.gnu.org/licenses/licenses.html>
- [2] BBC weather, prieiga internete  
<http://www.bbc.com/weather/>
- [3] National weather service, prieiga internete  
<http://www.weather.gov/>
- [4] Weather spark, prieiga internete  
<https://weatherspark.com/>
- [5] Weather bug, prieiga internete  
<http://weather.weatherbug.com/>
- [6] HAMweather, prieiga internete  
<http://www.hamweather.com/>
- [7] OpenWeatherMap, prieiga internete  
<http://openweathermap.org/>
- [8] WUnderground prieiga internete  
<http://www.wunderground.com/>
- [9] Effective Page Refresh Policies For Web Crawlers, prieiga internete  
<http://oak.cs.ucla.edu/~cho/papers/cho-tods03.pdf>
- [10] Optimal Robot Scheduling for Web Search Engines, prieiga internete  
<https://hal.inria.fr/file/index/docid/73372/filename/RR-3317.pdf>
- [11] User-Centric Web Crawling, prieiga internete  
<http://wwwconference.org/proceedings/www2005/docs/p401.pdf>
- [12] Optimal Crawling Strategies for Web Search Engines, prieiga internete  
<http://www.columbia.edu/~js1353/pubs/wolf-www02.pdf>
- [13] Recrawl Scheduling Based on Information Longevity, prieiga internete  
<http://infolab.stanford.edu/~olston/publications/www08.pdf>
- [14] Estimating Frequency of Change, prieiga internete  
<http://oak.cs.ucla.edu/~cho/papers/cho-freq.pdf>
- [15] Study of Web Crawler and its Different Types, prieiga internete  
<http://files.figshare.com/1717744/A016160105.pdf>
- [16] Web Crawler: Extracting the Web Data, prieiga internete  
<http://www.ijcttjournal.org/Volume13/number-3/IJCTT-V13P128.pdf>

## **8. TERMINŲ IR SANTRUMPŲ ŽODYNAS**

UI – vartotojo sąsaja

UX – vartotojo sąsajos patogumas

Vidinė sistema – tai sistemos dalis, skirta rinkti informaciją iš servisų. Ši dalis vartotojui yra nepasiekiamo.

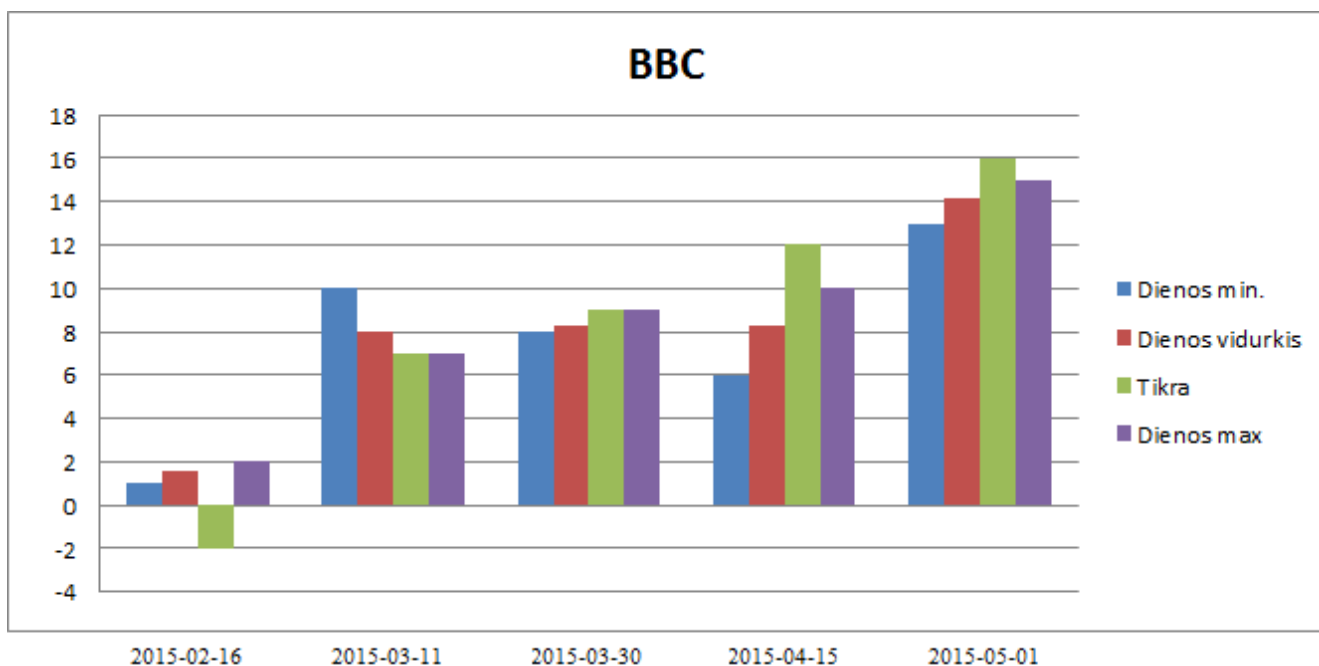
Servisas – tai nuotolinis serveris, teikiantis informaciją ar jos paketą

IS – informacinė sistema

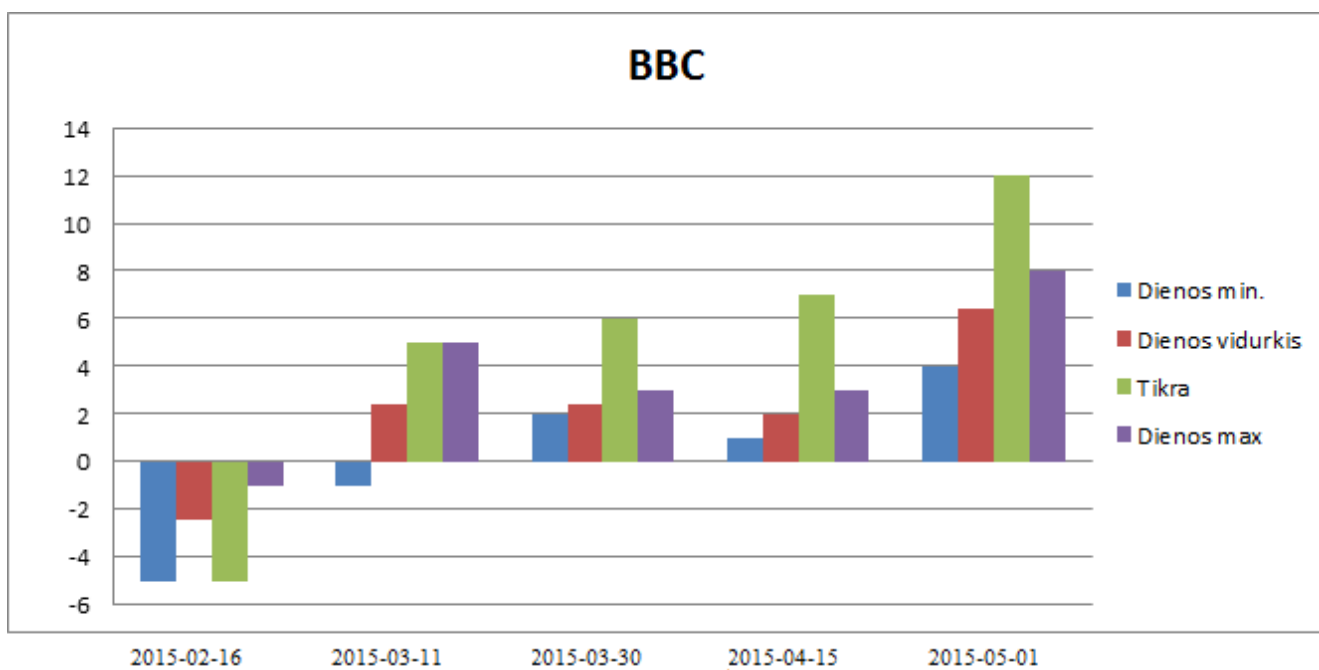
Crawler – šliužiotojas, tai programinė įranga skirta rinkti informacijai internete, surinkėjas.

## 9. PRIEDAI

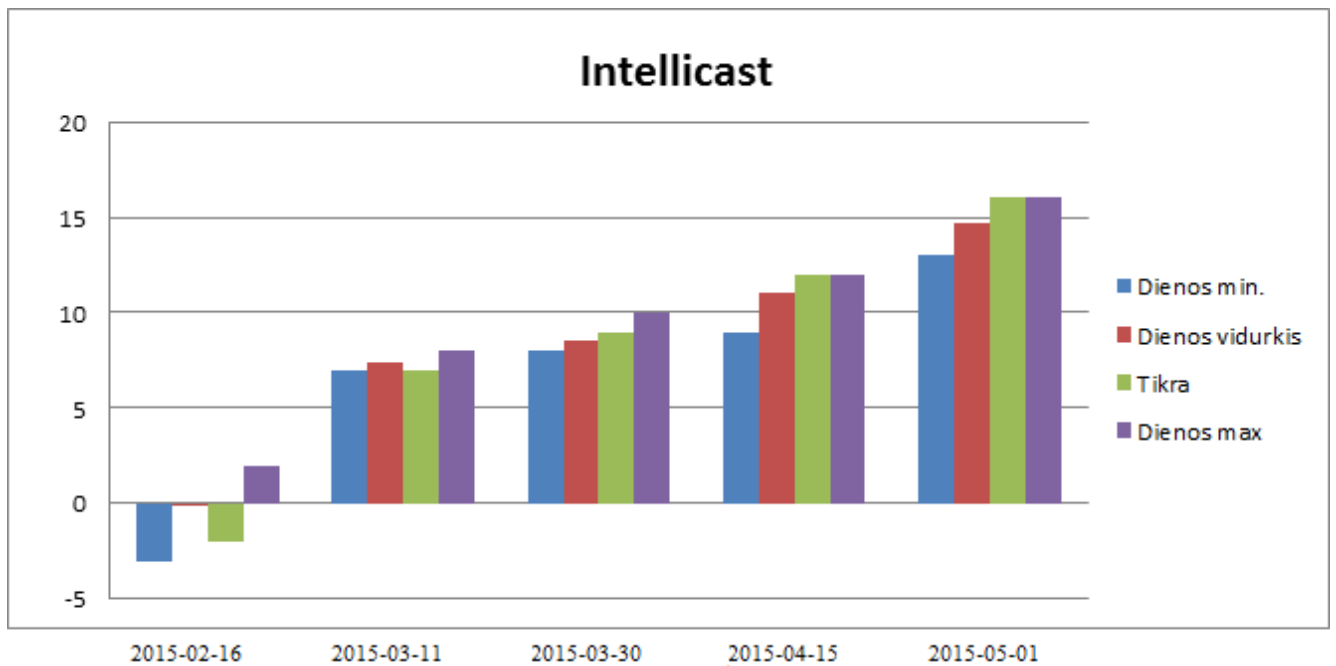
### 9.1. Eksperimento rezultatų apibendrinimas



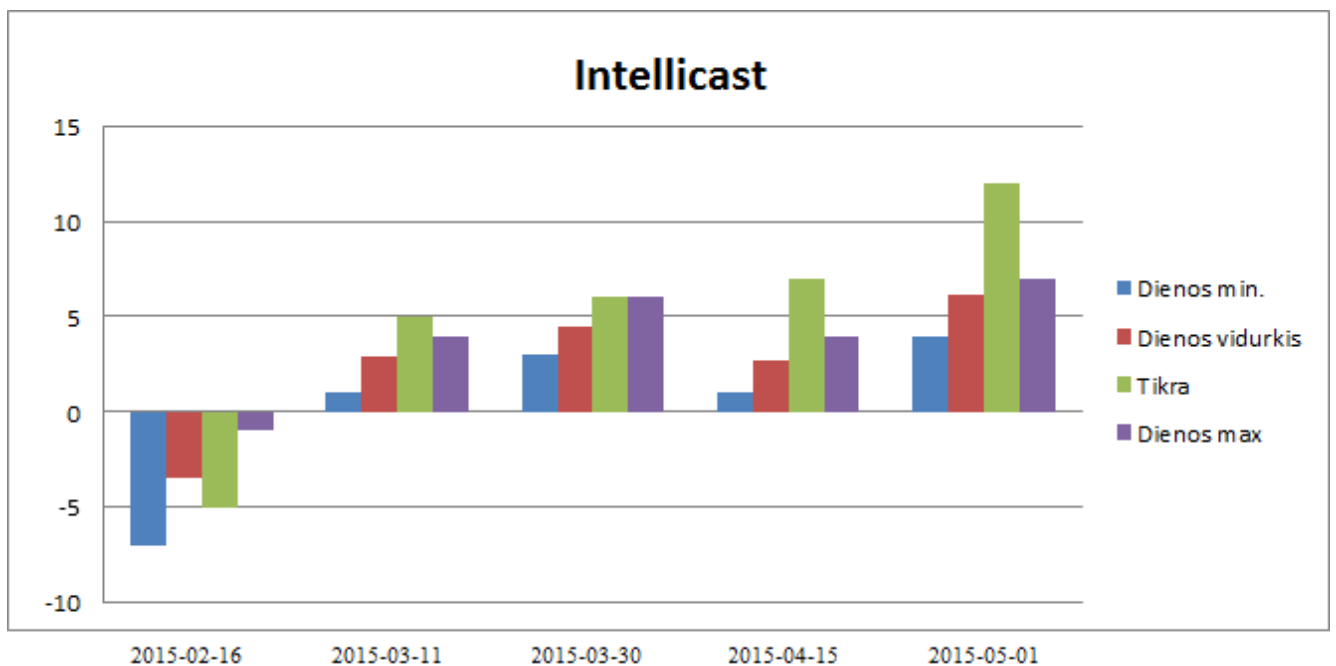
26. pav. BBC dienos temperatūrų prognozių tikslumai



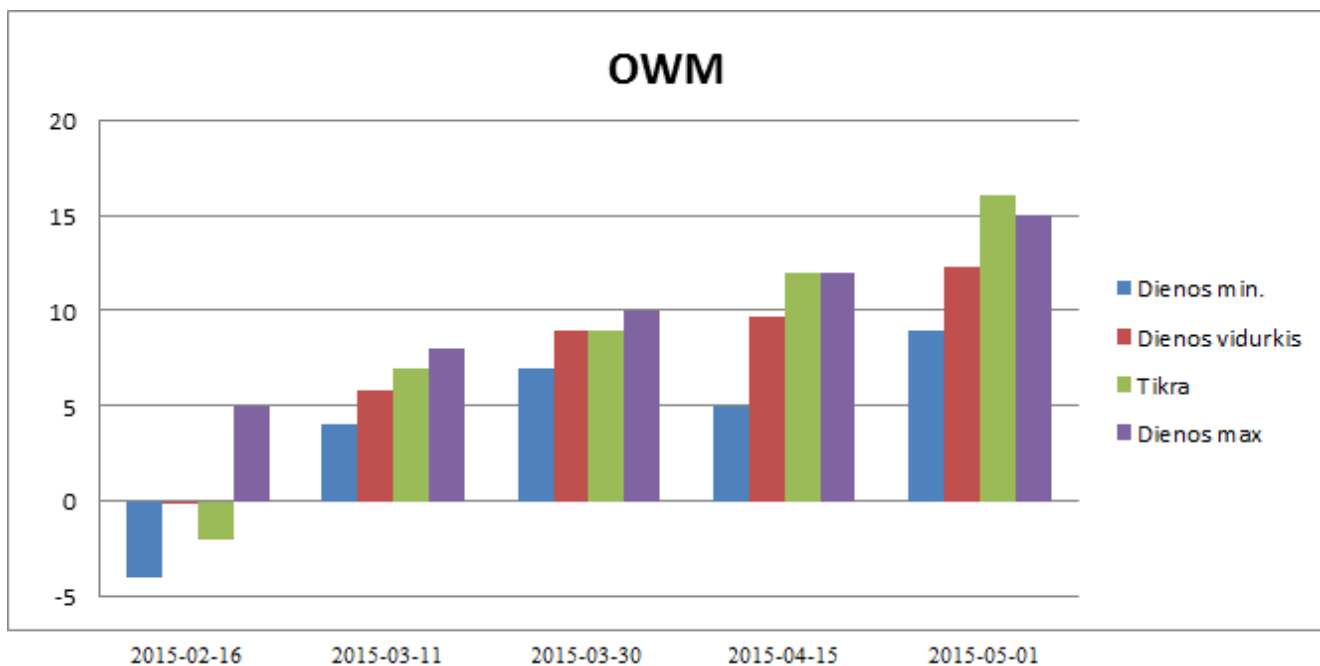
27. pav. BBC nakties temperatūrų prognozių tikslumai



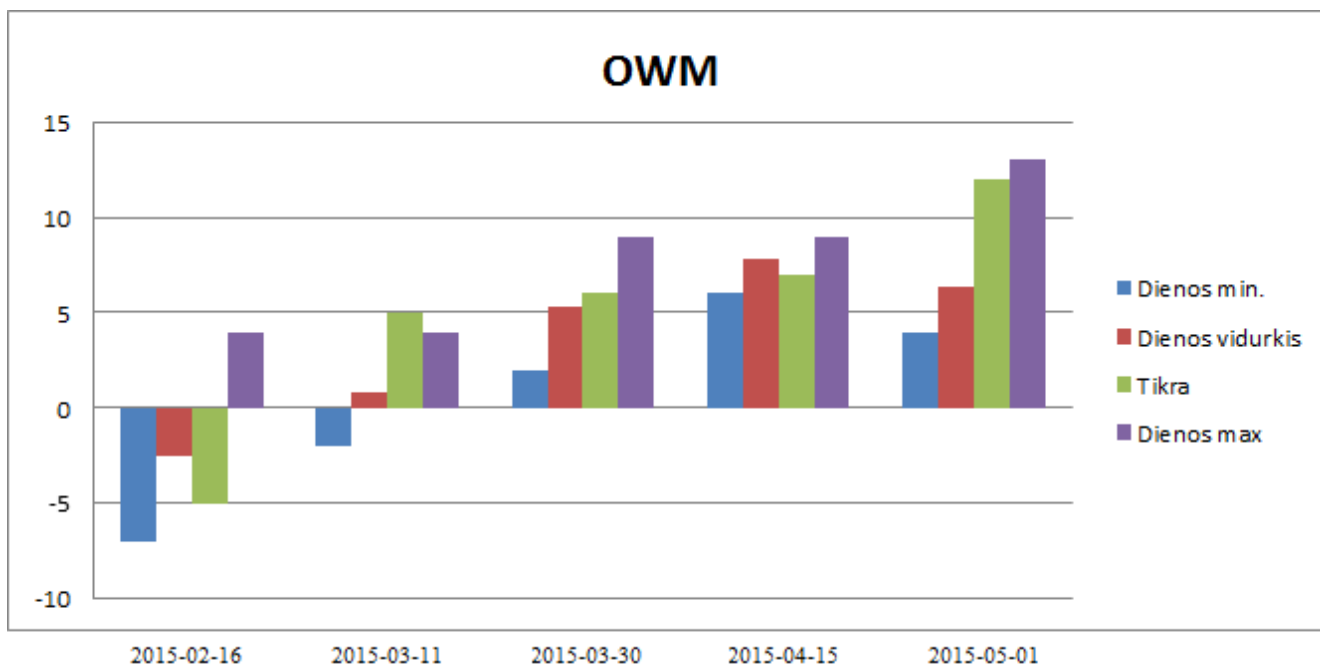
**28. pav.** Intellicast dienos temperatūrų prognozių tikslumai



**29. pav.** Intellicast nakties temperatūrų prognozių tikslumai

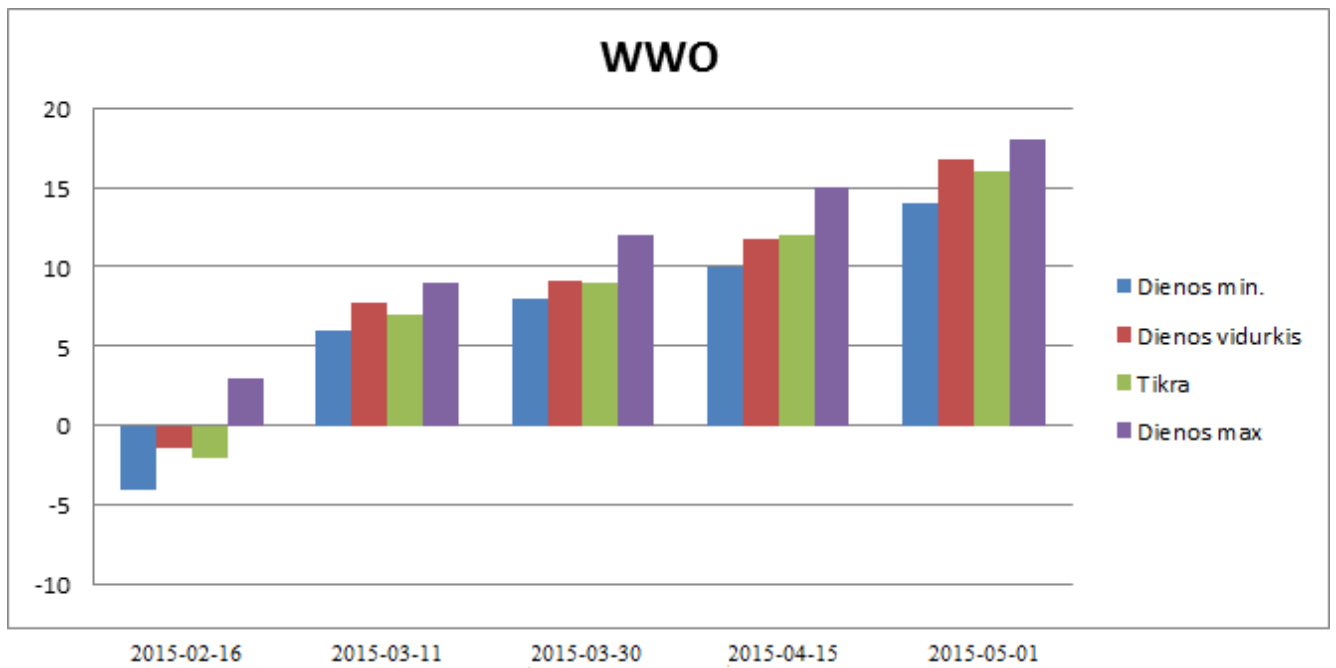


**30. pav.** OWM dienos temperatūrų prognozių tikslumai

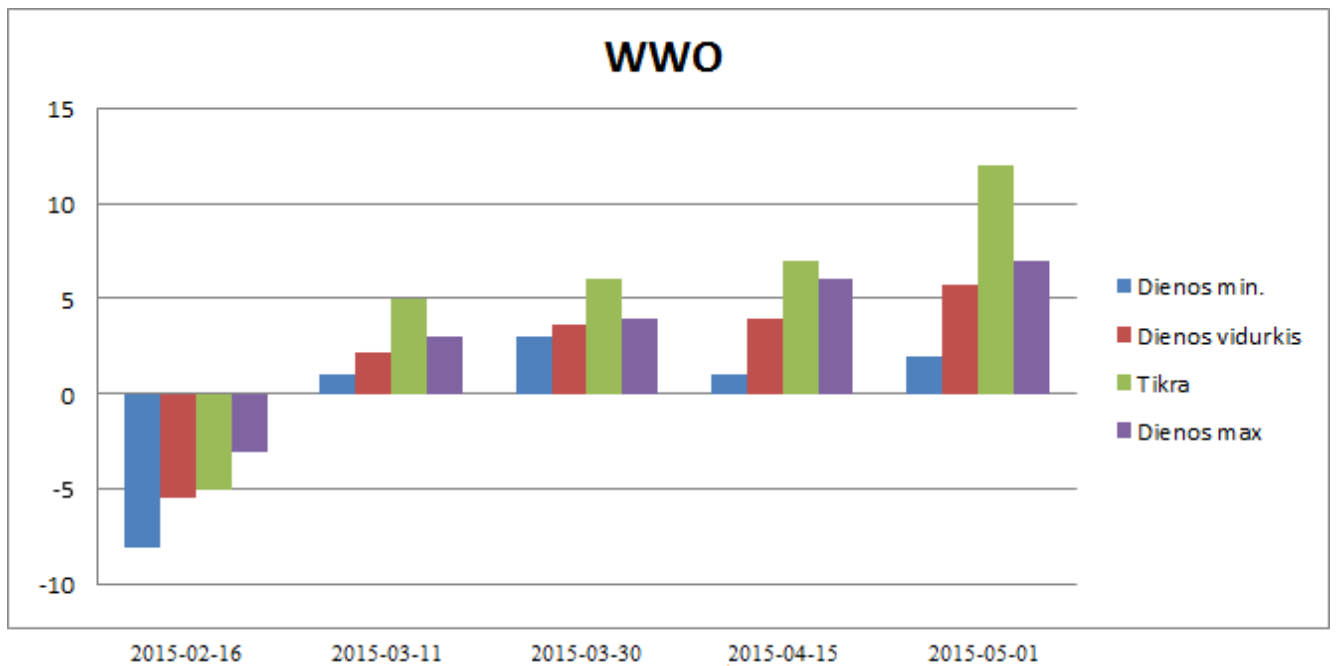


**31. pav.** OWM nakties temperatūrų prognozių tikslumai





32. pav. WWO dienos temperatūrų prognozių tikslumai



33. pav. WWO nakties temperatūrų prognozių tikslumai