KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

EIMANTAS ŽLABYS

AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU TYRIMAS

Magistro darbas

Vadovas
Prof. dr. R. Butkienė

KAUNAS, 2014

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

EIMANTAS ŽLABYS

AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU TYRIMAS

Magistro darbas

Vadovas
Prof. dr. R. Butkienė
2014-05-

Recenzentas
doc. dr. T. Blažauskas
2014-05-

Atliko:
IFM-2/2 gr. studentas
Eimantas Žlabys
2014-05-

KAUNAS, 2014

AUTENTIŠKUMO PATVIRTINIMAS

Patvirtinu, kad įteikiamas baigiamasis magistro darbas „Automatinis frazių atpažinimas tekste sudarytų šablonų pagrindu tyrimas“:

1. Autoriaus atliktas savarankiškai, jame nėra pateikta kitų autorių medžiagos kaip savos, nenurodant tikrojo šaltinio.
2. Nebuvo to paties autoriaus pristatytas ir gintas kitoje mokymo įstaigoje Lietuvoje ar užsienyje.
3. Nepateikia nuorodų į kitus darbus, jeigu jų medžiaga nėra naudota darbe.
4. Pateikia visą naudotos literatūros sąrašą.

_____	_____	_____
(studento vardas, pavardė)	(parašas)	(data)

SANTRAUKA

Šiame darbe pateikiama teksto analizės bei daiktvardinių ir veiksmažodinių frazių radimo ir paieškos analizė. Analizuojami algoritmai kurių pagalba būtų galima atpažinti frazes. Algoritmai kurie sugebėtų surasti veiksmažodines bei daiktvardines frazes lietuviškame tekste. Taipogi analizuojami sprendimai bei algoritmai kurių pagalba galima būtų sudaryti šablonus, kurių dėka būtų įmanoma ieškoti frazių. Atlikus išsamią algoritmų bei esminių sprendimų analizę buvo nustatyta, jog nėra algoritmo kuri lietuviškame tekste, naudodamas vartotojo sudarytus šablonus, galėtu aptikti veiksmažodines ir daiktvardines frazes. Darbo tikslas - ištirti daiktavardinių bei veiksmažodinių frazių atpažinimą lietuviškame tekste. Siekiamas rezultatas – daiktavardinių bei veiksmažodinių frazių šablonų sudarymo programos prototipas bei algoritmo, kuris galėtu ieškoti frazių lietuviškame tekste naudojant sudarytus šablonus prototipas.

Suprojektuotas algoritmas leidžia aptikti įvairios struktūros daiktvardines ir veiksmažodines frazes. Leidžia surasti tam tikro norimo ilgio frazes. Suprojektuota frazių paieškos sistema leidžia laisvai sudaryti šablonus su norimomis detalėmis (kalbos dalis, linksnis, giminė, skaičius ir t.t.). Eksperimento metu buvo pasirinkti atsitiktiniai tekstai kuriose algoritmas, pagal tam tikrus sistemos sudarytus šablonus, aptiko frazes. Algoritmo surastų frazių algoritmo pagrindžiamumui, buvo atliktas eksperimentas su asmenimis, kurie patys rankiniu būdu atliko frazių paiešką tuose tekstuose. Eksperimento metu buvo palygintos frazės rastos tiek eksperto tiek sistemos. Sistemos rastų frazių skaičius ne ką skiriasi nuo eksperto rastų, tačiau sistemoje sudaryti šablonai nepilnai rado visas frazes, o jei rado tai jos būdavo nepilnos. Eksperimento metu nustatyta, jog turint gerą bei tinkamą šabloną galima aptikti geresnes bei tinkamesnes frazes.

SUMMARY

This paper provides an analysis of the text , and nouns and verbal expressions of detection and analysis of the search . Analyzing algorithms which make it possible to identify phrases. The algorithms are able to find the verb and noun Lithuanian phrases in the text. We also analyzed solutions and algorithms which permit to create templates that make it possible to search for phrases . After a detailed algorithms , and key decision- analysis it was found that there is no algorithm that Lithuanian text using user -defined templates to detect noun and verb phrases. The aim - to explore the noun and verbal phrases recognition Lithuanian text. To achieve the desired result - noun and verbal phrases in the template of the program of the prototype and the algorithm is to search for phrases in Lithuanian text using templates consisting of a prototype .

Designed by the algorithm allows to detect different types of noun and verb phrases. Let's find some desired length phrases. Allows you to easily create templates with the desired parts (parts of speech, case, gender, number , etc.) . The experiment was randomly chosen texts with the algorithm in a system- defined templates found phrases. Algorithm find phrases algorithm experiment was carried out with individuals who have to manually search for phrases by those texts. During the experiment, the algorithm has been successfully implemented and applied to the analysis of the text.

TURINYS

LENTELIŲ SĄRAŠAS

PAVEIKSLŲ SĄRAŠAS

TERMINŲ IR SANTRUMPŲ ŽODYNAS

SBVR - Semantics of Business Vocabulary and Business Rules
VT – veiklos taisyklės, tam tikrus veiklos aspektus apibrėžiantis arba apribojantis teiginys, kuris reikalingas norint įvertinti veiklos struktūrą arba valdyti/įtakoti veiklą.
TSZ – Teminės srities žodynas
DF – daiktavardinė frazė
VF – veiksmažodinė frazė

ĮVADAS

Sistemoms ir technologijoms tobulėjant – vartotojų norai didėja. Labiau norima, jog dauguma rankinio darbo galima būtų perduoti kompiuteriams. Tokie įrankiai kurie galėtų pagelbėti vartotojams greičiau ir kokybiškiau atlikti darbus, vis tobulėja ir plečiasi į rinką.

Jau nuo seno buvo norima daugiau teksto analizės įrankių kurie galėtų analizuoti ir atlikti teksto analizę. Tokie įrankiai gali labai supaprastinti ir be abejo pagreitinti teksto analizę. Tokiems įrankiams šiuo metu yra skiriamos labai didelės lėšos ir laikas. Šių įrankių paklausa visame pasaulyje yra didelė ir panaudojimo galimybės labai plačios. Įrankiai kurie sugeba analizuoti tekstus, išskirti tekstą iš nuotraukų ar perskaityti dokumentus yra labai naudingi. Kaip vienas iš pavyzdžių šiuo metu naudojamo teksto atpažinimo įrankių gali būti panaudojamas – automobilių numerių atpažinimui ar kitur. Tokios sistemos leidžia užtikrintai visą tekstą perkelti į programas kurios gali toliau analizuoti tekstą pagal vartotojo pageidavimus.

Darbo aktualumas

Norint lietuviškame tekste rasti frazes, reikia gerai išmanyti lietuvių kalbos taisykles. Teksto analizė rankiniu būdu yra sudėtingas ir daug laiko reikalaujantis procesas. Dabartiniai algoritmai bei sistemos negali iš teksto išskirti daiktavardines bei veiksmažodines frazes, naudodamiesi vartotojo sudarytais šablonais.

Tyrimo objektas - daiktavardinių bei veiksmažodinių frazių atpažinimas lietuviškame tekste.

1. Išanalizuoti:
   1.1. Esančius frazių paieškos algoritmus.
   1.2. Esančius teksto analizės įrankius.
   1.3. Frazių atrinkimo procesą.
2. Suprojektuoti automatinį frazių paieškos įrankį.
3. Realizuoti automatinį frazių paieškos įrankį.
4. Atlikti automatinio frazių paieškos eksperimentinį tyrimą.

Gauti rezultatai

Atlikus eksperimentą buvo nustatyta jog sistema sugeba atpažinti daiktavardines bei veiksmažodines frazes lietuviškame tekste naudojant šablonus.

Darbo struktūra

Darbą sudaro šeši skyriai. Pirmame skyriuje aprašoma tyrimo tikslas, sritis, objektas, uždaviniai, esami sprendimai ir analizės išvados. Antrame skyriuje aprašomas frazių paieškos posistemės specifikacija ir analizė. Trečiame skyriuje aprašomas automatinio frazių atpažinimo tekste sudarytų šablonų pagrindu sprendimo algoritmas. Ketvirtame skyriuje aprašomi automatinio frazių atpažinimo tekste sudarytų šablonų pagrindu realizacijos projekto: realizacijos, architektūros ir elgsenos modelis. Penktame skyriuje aprašomas sistemos veikimas, testavimo modelis. Šeštame skyriuje aprašomas eksperimentas ir jo rezultatas. Prieduose galima rasti frazes kurias surado bei aptiko eksperimente dalyvavę asmenys bei sistema.

1. DAIKTAVARDINIŲ IR VEIKSMAŽODINIŲ FRAZIŲ ATPAŽINIMO TEKSTE ANALIZĖ

1.1. Analizės tikslas

Analizės tikslas yra išsiaiškinti daiktavardinių bei veiksmažodinių frazių sudarymo savybes. Išnagrinėti pagrindinius daiktvardinių frazių bei veiksmažodinių frazių skirtumus ir panašumus. Analizės tikslas taipogi yra suprasti tyrimo problemą, jos priežastis ir taipogi rasti tinkamą problemos sprendimą.

1.2. Tyrimo objektas, sritis, problema

Šioje dalyje apžvelgsime pagrindinius SBVR žodyno sudarymo principus bei metodus. Detaliau paanalizuosime semantinių žodyno sudarinėjimo taisykles. Taipogi šioje dalyje bus aprašoma teksto atpažinimo principai bei eiga. Taipogi bus labiau padetalizuota daiktavardinių bei veiksmažodinių frazių atpažinimo principai bei algoritmai leidžiantys lietuviškame tekste aptikti frazes.

Šio tyrimo pagrindinė sritis yra teksto automatizavimas pritaikant naujausias technologijas ir įrankius.

Problema:

Prieš pradedant nagrinėti tyrimo objekto problemą, būtina suprasti ką reiškia sąvoka žodynas. Žodynas - yra žodžių sąrašas su jų reikšmėmis (apibrėžimais). Žodynus sudaryti yra sunku ir užima daug laiko. Sudaryti tam tikros srities specializuotus žodynus reikia tam tikrų resursų tokių kaip: specializuotos knygos, straipsniai ar kiti dokumentai. Norint šiuos dokumentus analizuoti bei sudaryti gerą specializuotą žodyną būtina gerai išmanyti tą sritį kuriai yra sudaromas žodynas. Ankščiau ir iki šių dienų sudaryti specializuotą žodyną užtrunka daug laiko bei išeikvojama daug žmogiškųjų resursų. Norint šį procesą pagreitinti, jau yra kuriamos sistemos bei algoritmai kurie gali analizuoti pateiktus tekstus ir taipogi sudaryti specializuotus žodynus. Tačiau specializuotus žodynus sudaryti Lietuvių kalbai yra sunku bei šiuolaikiniai įrankiai nepasižymi geru bei tinkamu sudarymu.

Norint gerai bei tinkamai sudaryti specializuotą žodyną būtina žinoti pagrindinius žodynų tipus. Lietuvoje bei Užsienyje vyrauja keletas pagrindinių žodynų tipų:

- Kalbos žodynai
- Dvikalbiai žodynai
- Terminų žodynai
- Semantiniai veiklos žodynai

Kalbos žodynas - yra universalus aiškinamasis, norminamasis, plačiausiam skaitytojų ratui skirtas bendrinės kalbos kodifikacijos veikalas [2].

Dvikalbiai žodynai – verstiniai dvikalbiai, trikalbiai ar daugiakalbiai žodynai. Juose ne aiškinami kurios nors vienos kalbos žodžiai, bet „išverčiami“ į kitą kalbą arba, tiksliau sakant, pateikiami vienos kalbos žodžiams kitos ar kitų kalbų atitikmenys [3].

Terminų žodynai - tai žodynai kurios nors atskiros mokslo ar technikos šakos žodžius, daugiausia terminus, fiksuojantys žodynai. Todėl jie paprastai vadinami terminologiniais (pvz., fizikos, chemijos, tekstilės, sporto, lingvistikos ir t. t. žodynai) [4].

Semantiniai veiklos žodynai - tai žodynai, kuriuos sudaro ne tik tam tikros veiklos srities žodžiai, bet ir jų junginiai, išreiškiantys tos srities sąvokas, sąvokų junginius. Šie žodynai sudaromi prisilaikant SBVR standarto ir naudojami veiklos taisyklėms aprašyti. Žodynas ir taisyklės, kurie sudaryti prisilaikant SBVR standarto, gali būti interpretuojami ir kompiuterinių programų. Tam tikrose veiklos srityse, kur yra daug taisyklių ir jos keičiasi.

Žodynų pagrindinė sudedamoji dalis yra daiktavardžiai arba daiktavardinės frazės bei veiksmažodžiai arba veiksmažodinės frazės. Šiuos junginius tekste vartotojui rankiniu būdu yra gan sudėtinga aptikti ir tai užtrunka daug laiko.

Dėl laiko taupimo sumetimų, bei greitesnio žodynų sudarymo, daiktavardinių bei veiksmažodinių frazių radimas turi būti kompiuterizuojamas. Kadangi šiuo metunėra algoritmo ir įrankio sugebančio lietuviškame tekste, pagal vartotojo nurodytus šablonus aptikti frazių, dėl to iškyla tokie tyrimo objektas bei sritis:

Tyrimo objektas:

Daiktavardinių bei veiksmažodinių frazių atpažinimas lietuviškame tekste.

Tyrimo sritis:

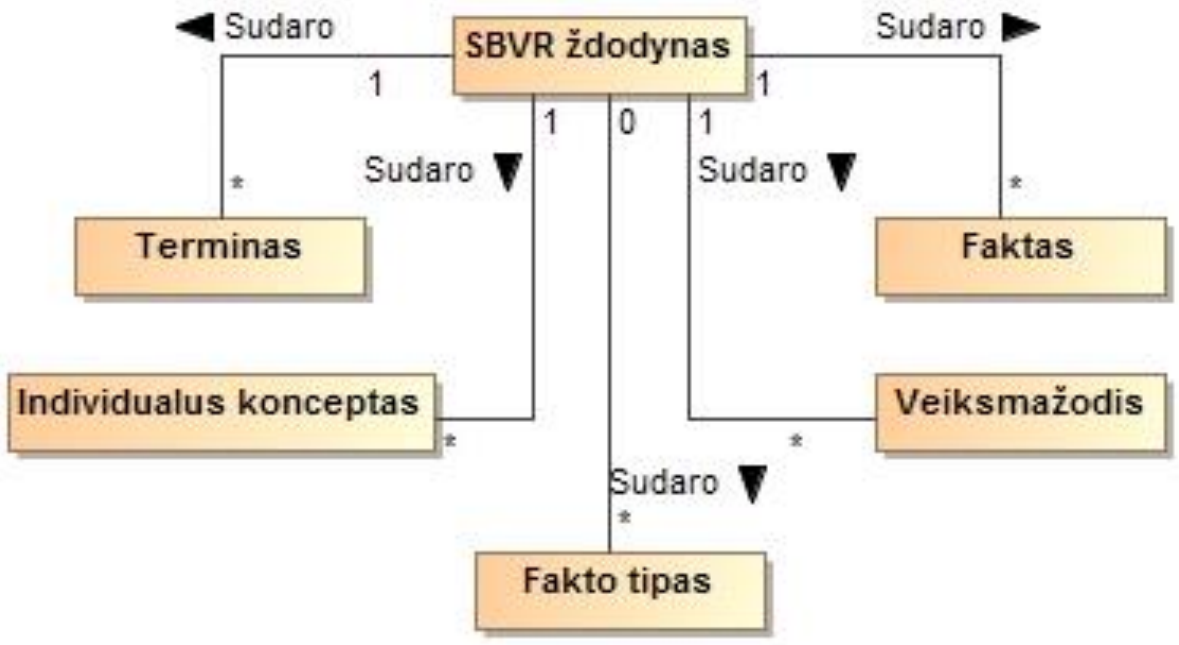
Daiktavardinių bei veiksmažodinių frazių atpažinimo lietuviškame tekste automatizavimas, automatizavimo algoritmai ir įrankiai.

1.2.1. SBVR standartas

Semantinių veiklos žodynų ir veiklos taisyklių *(Semantics of Business Vocabulary and Business Rules* – SBVR) specifikacija iš esmės apima du meta-modelius: žodyno ir taisyklių. Žodyno meta modelis tai žodynas veiklos žodynui apibrėžti(angl.), o taisyklių meta modelis - žodynas veiklos taisyklėms apibrėžti. Pastarasis naudoja ir žodyną veiklos žodynui apibrėžti.

Verslo žodynas apibrėžiamas kaip vieta, kur sukaupti visi specializuoti terminai ir koncepcijų apibrėžimai, kuriuos naudoja nagrinėjama organizacija ar bendruomenė kalbėdama ar rašydama apie savo veiklą.

SBVR žodynus sudaro tokios pagrindinės dalys: terminai, faktai, faktų tipai ir individai (1 Pav. SBVR žodyno sududenamosios dalys). Pagrindiniai SBVR struktūrizuotos kalbos elementai kurie padetalizuoja terminų bei faktų prasmę yra terminai, individualūs konceptai, klausimai, veiksmažodžiai ir raktiniai žodžiai.



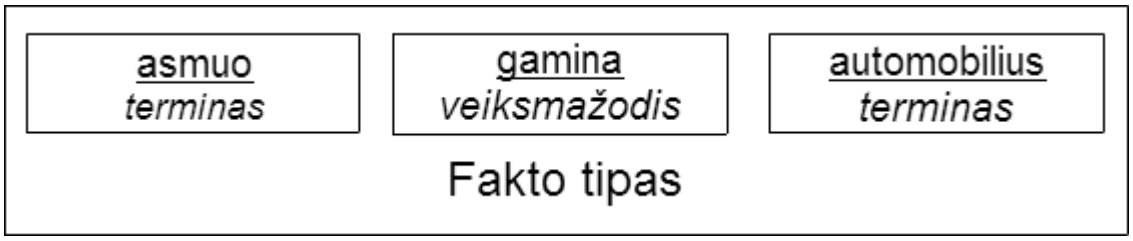
1 Pav. SBVR žodyno sududenamosios dalys

- **Terminai** – naudojami aprašyti esybėms ar rolėms, pavyzdžiui:
  universitetas, studentas, miestas, ...
- **Veiksmažodžiai** (angl. *verb*) – arba prielinksniai apibūdina terminus arba aprašo ryšius tarp terminų, pavyzdžiui:
  yra, iki, sugadinta, ...
- **Individualus konceptas** – tikriniai daiktavardžiai, reiškiantys asmenis, miestus ir kitus įvardytus objektus,
  skaičius:
  ‚Kaunas‘, ‚Vilnius‘, ‚Petras‘, ‚200‘, ...

- **Fakto tipas** – derinys vieno fakto ir kelių terminų, pavyzdžiui:
  studentai studijuoja universitete, bankas suteikia paskolą, ...

- **Faktai** sieja individualius konceptus arba individualius konceptus ir objektų tipus, pavyzdžiui:
  Jonas Jonaitis studijuoja KTU
  Jonas Jonaitis studijuoja universitete

SBVR žodyno fakto tipo sudarymą galima sudaryti pagrinde iš dviejų elementų: termino ir veiksmažodžio (2 Pav. Fakto tipo schema). Pateiktame pavyzdyje turime sakinio iškarpą iš trijų žodžių: *asmuo gamina automobilius.*



2 Pav. Fakto tipo schema

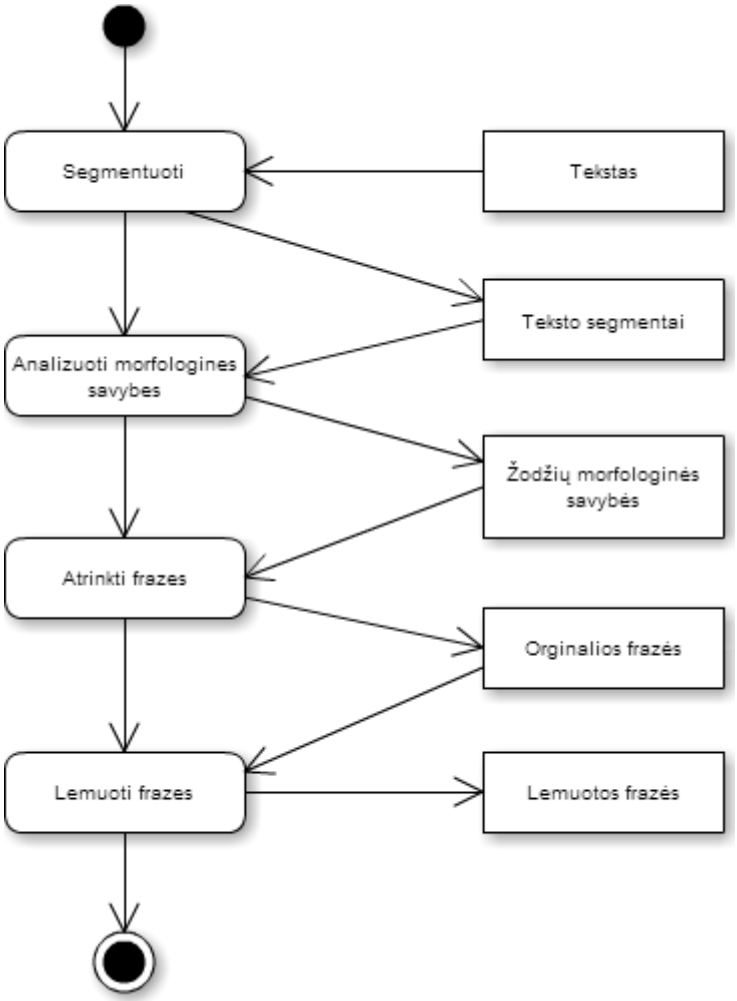
Iš šio pavyzdžio galime matyt, jog fakto tipą sudaro *veiksmažodis* kaip ryšys tarp dviejų terminų. Faktų tipo junginių gali būti įvairių, tačiau labiausiai yra naudojami:

Daiktavardis + Veiksmažodis + Daiktavardis

Dažniausiai sudarinėjant SBVR žodyną iš terminų laikomasi taisyklės, jog terminai negali būti šalia vienas kito.

1.2.2. Frazių atpažinimo procesas

Šioje dalyje apžvelgsime teksto analizavimo procesą. Plačiau panagrinėsime pagrindinius teksto analizavimo žingsnius bei atliekamus veiksmus norint rasti frazes (3 Pav. Frazių atpažinimo veiklos diagrama)



3 Pav. Frazių atpažinimo veiklos diagrama

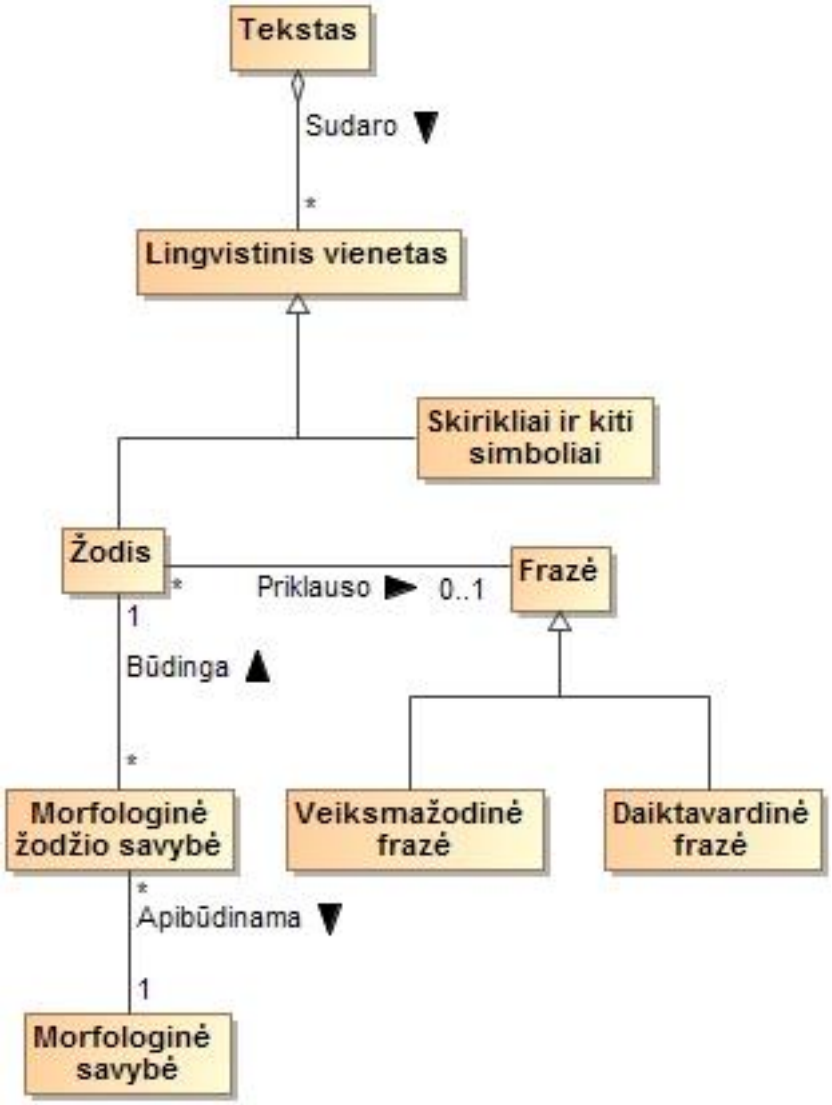
Detaliau panagrinėsime esmines teksto analizės bei frazių radimo žingsnius kurie yra pateikti (3 Pav. Frazių atpažinimo veiklos diagrama) :

- **Segmentuoti** – šiame veiksme atliekamas teksto segmentavimas į žodžius bei skirtukus. Šio žingsnio rezultatas yra žodžių bei lingvistinių vienetų sąrašas.
- **Analizuoti morfologines savybes –** šiame žingsnyje atliekamas kiekvieno žodžio ar skirtuko detalesnis analizavimas. Visi žodžiai bei skirtukai yra morfologiškai išanalizuojami t.y. išskiriamos *kalbos dalys*, *linksniai*, *giminė* ir t.t.
- **Frazių formavimas –** šiame žingsnyje yra ieškomos frazės, kurios gali tikti į žodyną. Šio žingsnio rezultatas yra visos rastos frazės iš teksto.
- **Frazių lemavimas –** šio žingsnio esmė yra iš rastų frazių sąrašo, pateikti visas frazes sulemuojame. Lemavimas yra žodžių ar žodžių junginių vertimas į bendrinę formą tarkime:
  - **Daiktavardžių bei būdvardžių lema –** paprastai šių kalbos dalių lema laikoma vienaskaitos vardininko forma, pvz:

    gražų Kauno Technologijos Universitetą → gražus Kauno Technologijos universitetas
  - **Veiksmažodžių lema –** yra tiesiog bendratis, pvz:

    skubėjo į → skubėti į

Pažvelgus į šį frazių atpažinimo procesą iš kitos perspektyvos t.y. veiklos pusės, aiškėja kitos pagrindinės teksto analizės bei žodynų esybės.



4 Pav. Frazių radimo veiklos esybės

Kaip matyti iš (4 Pav. Frazių radimo veiklos esybės) paaveiklėlo jog iš teksto pirmiausia būtina išskirti „lingvistinius vienetus“, kurie yra susideda iš dviejų dalių, t.y. žodžio bei skyriklio ir kitų simbolių. Kiekvienas žodis Lietuviu kalboje turi savo morfologines dalis, t.y. kalbos dalį, giminę, linksnį, skaičių ir t.t.. Frazę gali sudaryti daug žodžių, kadangi frazė yra žodžių junginys. Tačiau ne kiekvienas žodis gali patekti į frazės sudėti ir žodis frazėje gali būti daugiausiai vieną kartą. Frazė – baigtinę mintį reiškiantis ir antiniciškai baigtas žodžių junginys ar posakis. Tokius junginius galima skirstyti į dvi dalis t.y. Daiktavardines frazes bei veiksmažodines frazes.

1.2.3. Daiktavardinių ir veiksmažodinių frazių tipinės struktūros

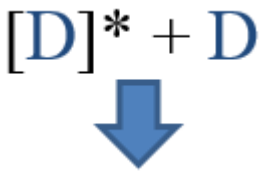
DAIKTAVARDINĖ FRAZĖ (angl. *noun phrase*) tradiciškai apibrėžiama kaip frazė, kurios pagrindinis dėmuo yra daiktavardis arba daiktavardį pakeičiantis įvardis. Iš pirmo žvilgsnio panašiai ir lietuvių kalbotyros tradicijoje apibrėžiama sąvoka „daiktavardinis junginys“, plg.: „daiktavardiniais junginiais laikomi žodžių junginiai, kurių pagrindinis dėmuo yra daiktavardis“ [5]. Tačiau kadangi tradicinėje lietuvių kalbotyroje, skirtingai nei Vakarų kalbotyroje, daiktavardiniais junginiais nelaikomi daiktavardiškieji įvardžiai ir jų junginiai (pvz., *jis, tas su geltonomis gatbanomis*), taip pat daiktavardiškai vartojami būdvardžiai (pvz., *pastarasis, aktyvioji*) bei neišplėsti daiktavardžiai (pvz., *mergaitė* ir *gėlės* sakinyje *mergaitė laistė gėles*), šiame darbe pasirinkta vartoti tarptautinį terminą

DAIKTAVARDINĖ FRAZĖ[1], kurio vadinami argumento poziciją užimantys daiktavardžiai, daiktavardiniai junginiai, daiktavardiškieji įvardžiai ir jų junginiai, daiktavardiškai vartojami būdvardžiai ir kt.

Pagrindinės daiktavardinių formas sudaro daiktavardis kartu einantis su būdvardžiu. Šių junginių kombinacijų sudaryti galima be galo daug, kadangi tiek daiktavardžiai tiek būdvardžiai gali kartotis daug kartų frazėje.

Frazių šablonų sudarymui yra naudojami raktiniai simboliai, kurie apibendrina šabloną:
D - daiktavardis
B – būdvardis
Dal - dalyvis
[ ] – privalomumas (elementas gali būti privalomas arba ne)
* - pasikartojamumas (elementas gali kartotis *n* kartų)

- Daiktavardinių frazių šablonas kurį sudaro daiktavardžiai (1 Lentelė Šablonai kurie susideda iš daiktavardžių):



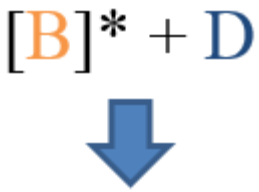
1 Lentelė Šablonai kurie susideda iš daiktavardžių

Daiktavardinės frazės struktūra	Pavyzdys
D	Kaunas
DD	technologijos universitetas
DDD	Kauno technologijos universitetas

Šios struktūros šablonas leidžia tekste aptikti bet kokio ilgio daiktavardines frazes kurios yra sudarytos iš daiktavardžių. Taipogi šis šablonas gali surasti ir vieno ilgio daiktavardinę frazę t.y. „Kaunas“ kadangi šiame šablone, pirmasis daiktavardis neprivalomas. Tačiau jei daiktavardis randamas tai jis gali kartotis *n* kartų.

[1] Iš lietuvių kalbininkų daiktavardiškai vartojamus įvardžius „daiktavardžio frazėmis“ laiko Rosinas (Rosinas 2009: 54). Mikulskas vartoja „daiktavardinės frazės“ sąvoką, kuri visiškai atitinka anglų noun phrase (Mikulskas 2006a).

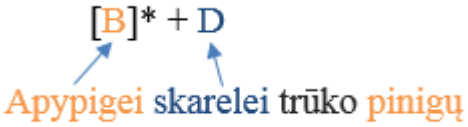
- Daiktavardinių frazių šablonas kurį sudaro būdvardis bei daiktavardis. Būdvardis gali kartotis *n* kartų, tačiau jo gali ir nebūti daiktavardinėje frazėje (2 Lentelė Šablonai kurie susideda iš būdvardžio bei daiktavardžio):



2 Lentelė Šablonai kurie susideda iš būdvardžio bei daiktavardžio

Daiktavardinės frazės struktūra	Pavyzdys
D	Kaunas
BD	naujasis kapitonas
BBD	Naujasis, paprastasis namas

Šios struktūros šablonas leidžia tekste aptikti bet kokio kiekio būdvardžio ir vieno daiktavardžio frazes. Taipogi šis šablonas gali aptikti tik vieną daiktavardį kadangi būdvardis yra neprivalomas šios struktūros elementas. Žemiau pateiktame pavyzdyje matosi kaip iš sakinio galima išskirti tokio tipo frazes:

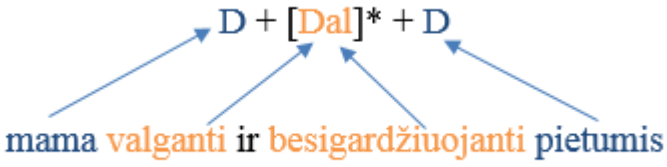


- Daiktavardinių frazių šablonas kurį sudaro du daiktavardžiai bei per vidurį esantis dalyvis (3 Lentelė Šablonai kurie susideda iš dalyvio bei aplinkui einančių daikvaradžių):

3 Lentelė Šablonai kurie susideda iš dalyvio bei aplinkui einančių daikvaradžių

Daiktavardinės frazės struktūra	Pavyzdys
DD	Technologijos Universitetas
DDalD	mama mėgdžiojanti katę
DDalDalD	mama valganti ir besigardžiuojanti pietumis

Šios struktūros šablonas leidžia tekste aptikti frazes kurios yra sudarytos iš dviejų daiktavardžių bei per vidurį einančio dalyvis. Detaliau panagrinėkime paskutinį lentelės (3 Lentelė Šablonai kurie susideda iš dalyvio bei aplinkui einančių daikvaradžių) pavyzdį:

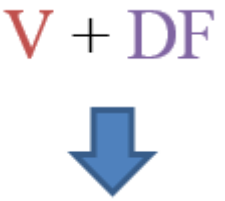


Kaip matyti iš pateikto pavyzdžio, jog tarp dviejų būdvardžių yra jungtukas „*ir*“, kuris turėtu neleisti rasti šios frazės. Tačiau pagal nutylėjimą lietuvių kalboje yra jog prieš tai ėjęs elementas ir po jo sekantis einantis elementas yra tokios pat kalbos dalies ir tarp jų yra jungtukas „*ir*“, tai galima šį jungtuką ignoruoti, tačiau įtraukti į galutinį frazės vaizdą.

VEIKSMAŽODINĖ FRAZĖ (verb phrase) susideda pagrinde iš veiksmažodžio ar veiksmažodžių junginių. Viena iš pagrindinių veiksmažodinių frazių taisyklių, jog veiksmažodis privalo eiti kartu su daiktavardine fraze, tačiau daiktavardinė frazė nepriklauso veiksmažodinei frazei.

Frazių šablonų sudarymui yra naudojami raktiniai simboliai, kurie apibendriną šabloną:
V – veiksmažodis
Prl – prielinksnis
Prv – prieveiksmis
DF – daiktavardinė frazė
[ ] – privalomumas (elementas gali būti privalomas arba ne)
* - pasikartojamumas (elementas gali kartotis *n* kartų)

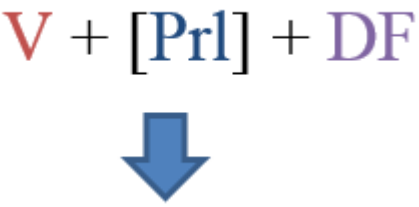
- Veiksmažodinių frazių šablonas kurį sudaro vienas veiksmažodis:



4 Lentelė Šablonų pavyzdžiai kurie susideda iš veiksmažodžio ir DF

Veiksmažodinės frazės struktūra	Pavyzdys
V + DF	turi automobilį, neša maišą, bėgo namo

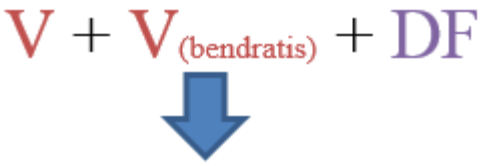
- Veiksmažodinių frazių šablonas kurį sudaro veiksmažodis ir prielinksnis:



5 Lentelė Šablonai kurie susideda iš veiksmažodžio, prielinksnio bei DF

Veiksmažodinės frazės struktūra	Pavyzdys
V + DF	turi automobilį, neša maišą, bėgo namo
V + Prl + DF	neša ant stogo, bėgo į urvą

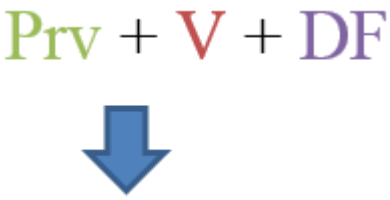
- Veiksmažodinių frazių šablonas kurį sudaro du veiksmažodžiai, bet antras veiksmažodis turi būti bendratis:



6 Lentelė Šablonai kurie susideda iš veiksmažodžio, veiksmažodžio bendraties bei DF

Veiksmažodinės frazės struktūra	Pavyzdys
V + V(bendratis) + DF	turi būti namas, gali pamatyti krūmą

- Veiksmažodinių frazių šablonas kurį sudaro prieveiksmis ir veiksmažodis:



7 Lentelė Šablonai kurie susideda iš veiksmažodžio, prieveiksmio bei DF

Veiksmažodinės frazės struktūra	Pavyzdys
Prv + V + DF	smarkiai lijo kieme, sunkiai kalba meška

1.3. Vartotojų analizė

Vartotojai atliekantys specializuotų žodynų sudarymą yra tos srities specialistai. Šios grupės vartotojų pagrindinis noras, greitai bei kokybiškai atlikti teksto analizę. Šio tipo vartotojai puikiai išmano sritį kuriai žodynas yra kuriamas bei, kur jis bus panaudojamas.

Pagrindinė vartotojų problema yra lėtas ir gan sudėtingas žodyno sudarymas bei skirtingas frazių ar faktų išskyrimas tekste. Specializuotų Lietuviškų žodynų sudarymas kompiuterių pagalba dar nėra pakankamai išvystytas, jog leistu kokybiškai išskirti daiktavardines bei veiksmažodines frazes. Vartotojai norėdami išanalizuoti didelius kiekius tekstų, privalo kaupti daug informacijos geresniems specializuotiems žodynams sudaryti.

1.4. Frazių atpažinimo įrankių analizė

Pasaulyje jau yra sukurta automatinių įrankių, kurie padeda automatiškai atrinkti terminus ir frazes. Iš jų galima sudaryti žodynus. Programos gali veikti vienoje kalboje arba gali veikti keliose kalbose. Šiame skyriuje bus apžvelgta populiariausios mokamos ir nemokamos programos skirtos ieškoti tekstuose frazių ir terminų.

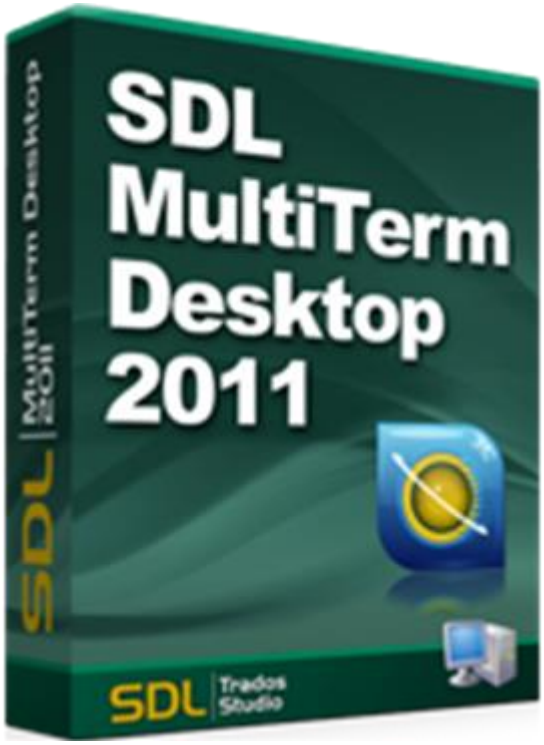
Frazių paieškos įrankių analizė:

- **SDL MultiTerm Desktop S2011**
- **SynchroTerm**
- **WordFast**
- **TexNet32**

Žemiau pateikiama, detalesnė šių įrankių analizė:

SDL MultiTerm Desktop S2011

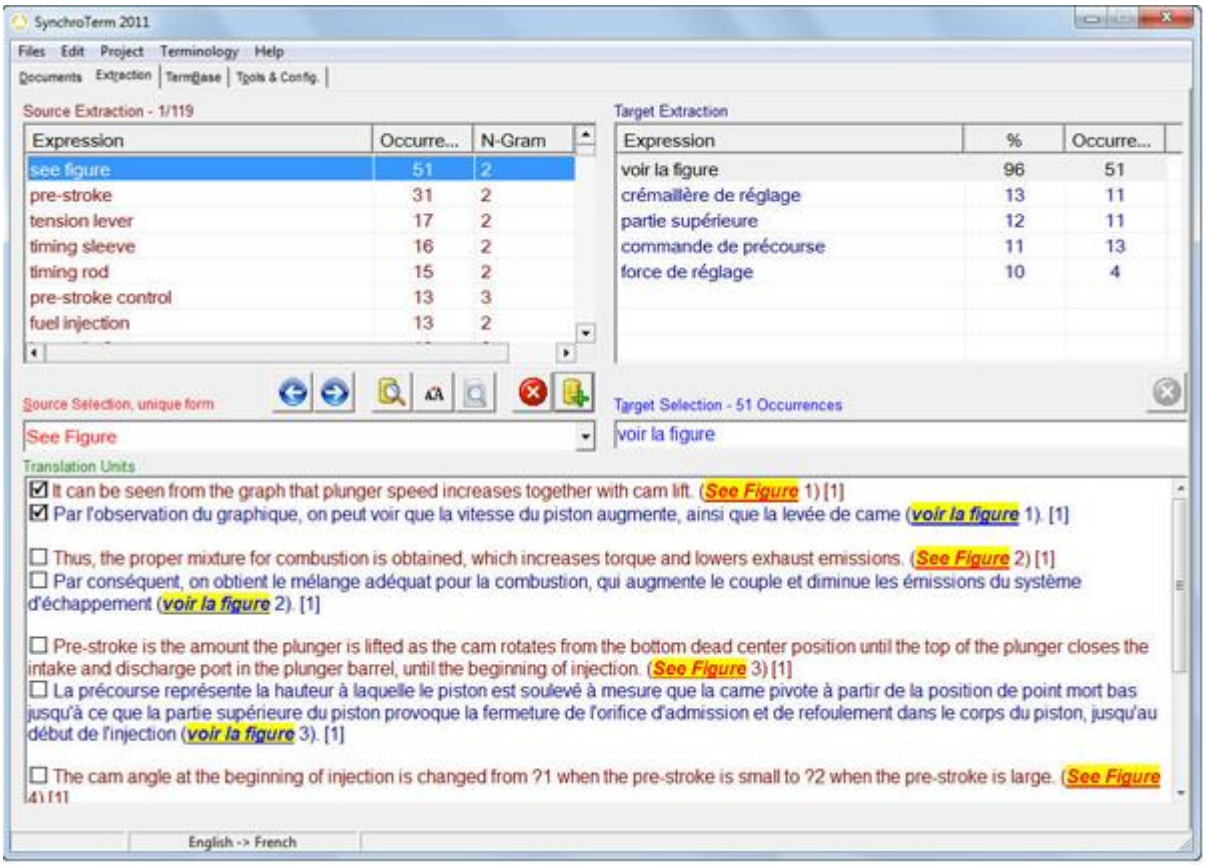
Šią mokamą (606 doleriai metams) programą sukūrė SDL kompanija ir buvo išleista į rinką 2011. Ji palaiko visas didžiąsias vakarų valstybių kalbas, norint su šia programa sėkmingai dirbti kitose kalbose reikia įdiegti papildomų priedų. SDL MultiTerm išrenka potencialius terminus, jei reikia parašo jų vertimą ir parodo galimų kandidatų sąraše (1.pav). Sąrašą galima koreguoti ir keisti. Sutvarkytą sąrašą galima išsaugoti įvairiais formatais (XML, TXT, DOC), tai pat galima saugoti ir duomenų bazėje. Saugant į duomenų bazę galima pasirinkti ar norime saugoti į egzistuojančią ar norime sukurti nauja. Galima uždėti įvairius filtrus paieškai (terminai bus ne mažiau dviejų žodžių). Programa suteikia galimybę ieškoti iš skirtingomis kalbomis parašytų tekstų [6].



5 Pav. MultiTerm Desktop 2011 [6]

SynchroTerm

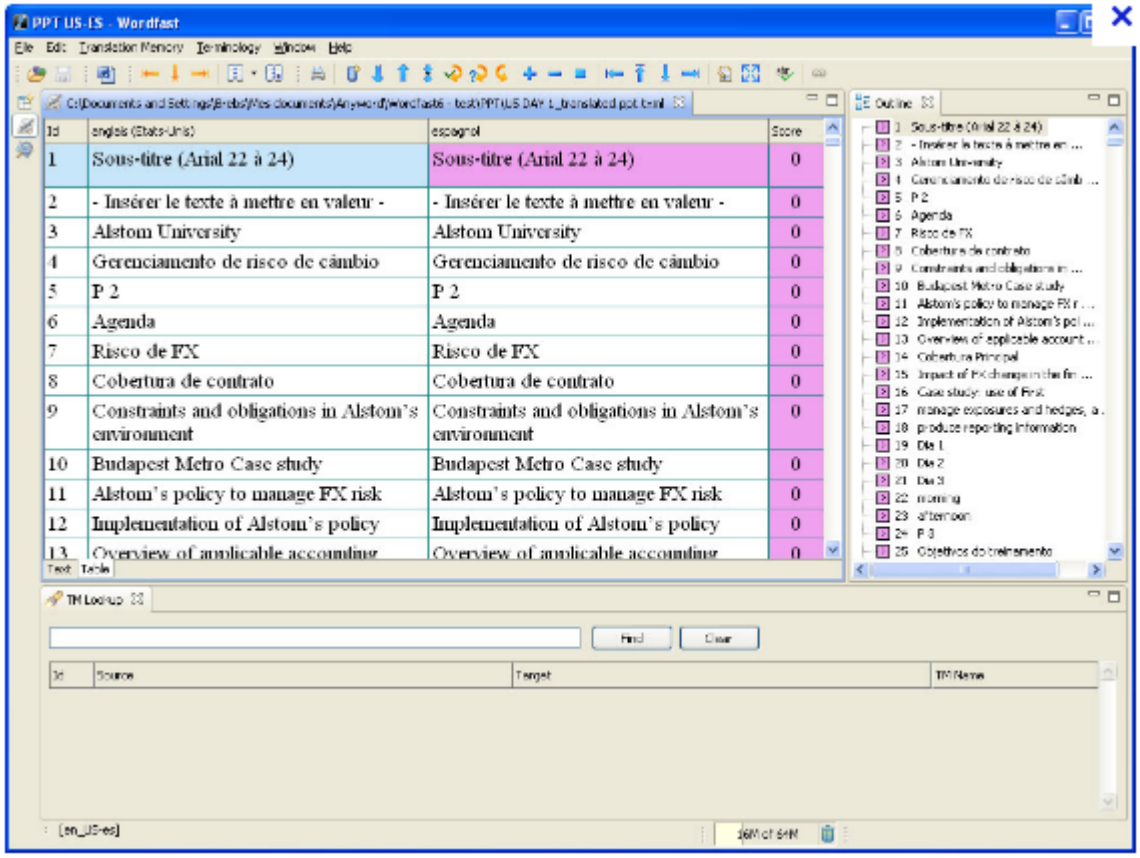
Šią mokamą (420 dolerių metams) programą išleido Terminotix kompanija. Ji palaiko didžiąsias vakarų Europos kalbas (anglų, prancūzų, ispanų ir t. t.). Galima rinkti terminus iš vieno arba kelių tekstų. Išrinktus terminus parodo sąraše, kurį galima koreguoti (2.pav). Terminus saugoma duomenų bazėse. Duomenų bazėse esančius terminus galima koreguoti, perkelti į kitą duomenų bazę arba pašalinti iš jos. Synchterm naudoja statistinius, morfologinius ir sintaksinius algoritmus rasti terminams [7].



6 Pav. SynchoTerm, teksto analizavimo įrankis

WordFast

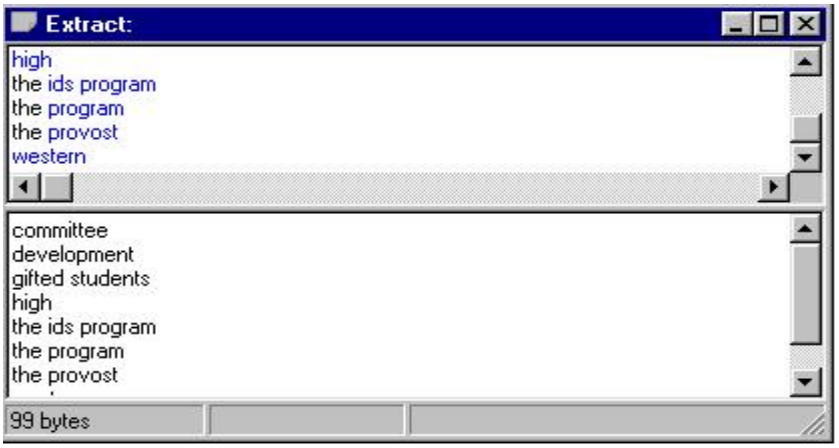
Šią mokamą (500 dolerių metams) programą sukūrė Wordfast LLC kompanija. Ši programa naudoja Microsoft Word kaip teksto modifikatorių. Ji veikia dauguma platformų (Windos, Mac os, Linux). Ji palaiko visas kalbas, kurias palaiko Microsoft Word. Į jas įeina tokios kaip kinų, japonų, korėjiečių kalbas, tai pat kalbos, kuriose rašoma iš dešinės į kairę (arabų, hebrajų). Viename žodyne gali būti iki 250 tūkstančių įrašų. Tai pat į Wordfast gali apdoroti ne tik paprastu Word dokumentus, bet tai pat ir Exel'io ar Power Point. Wordfast palaiko ir HTML dokumentus. Šioje programoje galima pasirinkti kaip norima, kad atrodytų žodynas (su aprašu ar ne). Atrinktus terminus rodo Word dokumente, kuriame galima juos koreguoti. Sudarytais žodynais gali naudotis daug vartotojų per tinklą [8].



7 Pav. Wordfast sąsaja, kurioje rodomi atrinkti terminai

TexNet32

TexNet32 yra nemokama programa, kuri padeda išrinkti žodžius, frazes. Veikia tik Windows aplinkoje. TexNet32 gali nuskaityti .txt arba .rtf formatus. Tai pat failus sukurtus su TexNetF(.doc- čia ne Microsoft Word formatas ir .net). Galima apdoroti tai pat ir HTML dokumentus. Išsaugoti galima trimis formatais .rtf, .txt arba .htm. Atrinktus žodžius galima redaguoti arba ištrinti. Po koregavimo žodžius ir terminus pateikia atskirame lange. Vienu metu apdoroja tik viena dokumentą. Atrinkimas vyksta pagal kiek daugiausia simboliu gali būti frazėje (leidžiama. kad daugiausia gali būti 10 simbolių). Tai pat žiūrima kiek kartų tekste jie pasikartojo. Ieškoma tai pat ir stop žodžių (anglų kalboje tai yra *an, the*, lietuvių kalboje būtų atitikimas - *prie*).



8 Pav. TexNet32 sąsaja, kurioje rodomi atrinkti terminai.

Išanalizuotų sistemų palyginimas

Šiame skyriuje palyginsime jau aukščiau išanalizuotas sistemas. Sistemos bus įvertintos pagal šiuos kriterijus:

1. Galimybė analizuoti Lietuvišką tekstą.
2. Galimybė pateikti lemuotas frazes.
3. Galimybė iš teksto išrinkti tam tikro ilgio ar struktūros frazes.
4. Galimybė vartotojui sudaryti paieškos tekste šablonus
5. Sistema gali analizuoti: .doc, .docx, .txt, .pdf failų formatus

Teksto analizės įrankių palyginimas pateikiamas (8 Lentelė Teksto analizės įrankių palyginimas) lentelėje.

8 Lentelė Teksto analizės įrankių palyginimas

	SDL MultiTerm Desktop S2011	SynchroTerm	WordFast	TexNet32	Projektuojama sistema
Galimybė analizuoti Lietuvišką tekstą	+	–	–	–	+
Galimybė pateikti lemuotas frazes	–	+–	–	–	+
Galimybė iš teksto išrinkti tam tikro ilgio ar struktūros frazes	–	–	–	–	+
Galimybė vartotojui sudaryti paieškos tekste šablonus	–	–	–	–	+
Sistema gali analizuoti: .doc, .docx, .txt, .pdf failų formatus	+–	+–	+–	+–	+

1.5. Frazių ir terminų atpažinimo algoritmai

Per pastaruosius metus, ryškus progresas buvo padarytas, tekstų analizavime norint iš jo išgauti visą reikiamą informaciją, kurią būtų galima panaudoti įvairiems tikslams. Pradedant nuo paprasčiausių frazių išskyrimo algoritmų kurie turi didelį įvairių simbolių ir išsireiškimų skaičių kurie reikalingi identifikuoti ir sugrupuoti tekstą taip, kad suprastų žmogus apie ką šiame tekste yra kalbama neskaičius jo.

Šioje dalyje bus kalbama apie žodynų sudarymo ir teksto supratimo algoritmus. Norint sudaryti specializuotą žodyną reikia iš teksto išskirti raktinius žodžius bei frazes, kurių pagalba bus sudarinėjamas žodynas. Dažniausiai, žodynų įrašų vardai ir aprašai gali programoms padėti sudaryti specializuotus žodynus.

Žodynų įrašai skiriasi nuo techninių vartojamų žodžių. Nes tarkime turėdami vieną žodį paprastame žodyne, tai jo prasmė gali skirtis nuo tarkime techniniame žodyne esančio žodžio.

1.5.1. Variantų atrinkimo algoritmo analizė

Šio algoritmo veikimo principas yra iš teksto išrinki ir atrinkti panašius žodžiu bei frazes ir juos suskirstyti į tam tikras dalis, iš kurių būtų galima sudaryti specializuotą žodyną. Po to kai žodžiai yra atrenkami, jie sudedami į bendro panašumo logikas, kurios tampa raktiniais žodžiais frazių paieškoje tekste. Toliau apžvelgsime kaip apskaičiuoti galimybę iš gautų žodžių sąrašo ir taip išrinkti frazes. Yra keli tipai formų suskirstyti frazes:

- Simbolių variacijos
- Variantų sudarymas
- Variantai su klaidomis
- Santraukos

Simbolių variacijos algoritmas

Algoritmo logika yra tokia, jog reikia išskirti iš žodžių įvairius skirtukus tarkime: , - , / , ir kt. Kaip žinoma lietuvių kalboje taipogi yra žodžių ar išsireiškimų kurių reikšmė yra viena tačiau jie per vidurį yra išskirti brūkšneliu, tai šis algoritmas tiesiog panaikina tarp žodyje esantį skyriklį ir jį implimentuoja kaip vieną ištisą žodį, tarkime:

„morfologinė-sintaksinė daryba“ -> *„*Morfologinė sintaksinė daryba“

Variantų sudarymo algoritmas

Šio algoritmo logika paremta panašių žodžių struktūros ieškojimu. Jis lygina žodžius su kitais tariamais panašiais žodžiais ar frazėmis nors jos gali būti ir nepanašios užrašymu tačiau prasmė gali būti tokia pati, tarkime:

Angl.:Passenger Airbag ir *passenger air bag,*

Mass Airflow Sensor ir *Mass Air Flow Sensor;*

Variantų su klaidomis algoritmas

Šis algoritmas atpažįsta nors ir klaidingai (su klaidomis) parašytas frazes ir juos prideda į sąrašą, kuris po to yra pridedamas prie variantų sudarymo algoritmo. Šio algoritmo pagrindinis tikslas yra rasti nors ir galimai klaidingas frazes, tačiau jų nepašalinant. Šis algoritmas tinka ir lietuviško teksto analizei bei frazių atrinkimo.

Santraukos algoritmas

Įvairiuose dokumentuose, straipsniuose, o ypač techninės paskirties dokumentuose, yra naudojama daug sutrumpinimų, dėl to reikia ypač atsargiai į tuos sutrumpinimus žiūrėti bei juos išskirti iš teksto. Šioje dalyje bus naudojamas metodas, kuris yra apibrėžtas šioje dalyje[17], norint atpažinti sutrumpinimus. Pvz.:

Angl.: 4H : four-wheel drive high

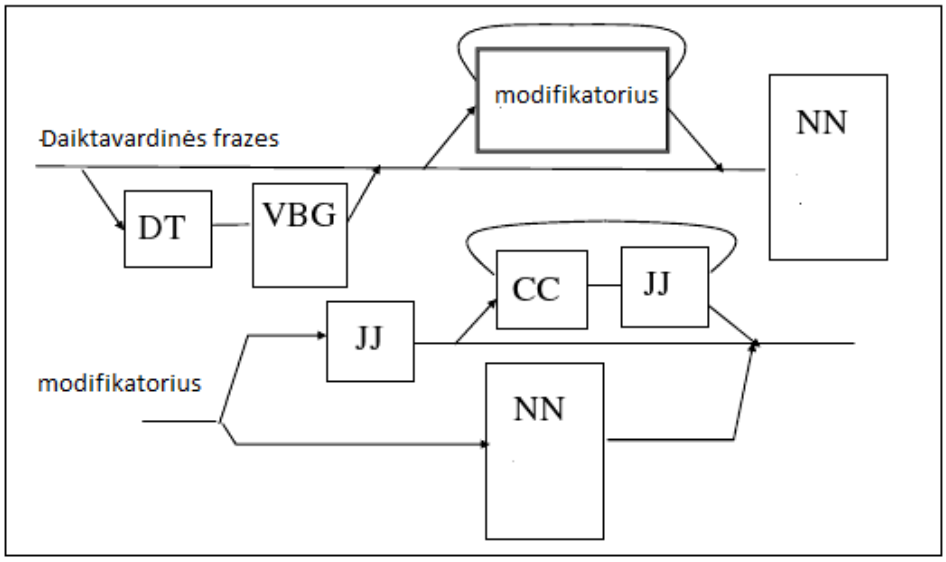
ATF : automatic transmition fluid

Liet.: KTU : Kauno Technologijos Universitetas

UAB : Uždara Akcinė Bendrovė

1.5.2. Daiktavardinių frazių atpažinimo iš sakinio struktūros algoritmas

Žodyno elementas gali būti arba daiktavardinė frazė arba veiksmažodinė frazė. Veiksmažodžius svarstome tik tokius, kurie yra ne pagalbiniai ir imame jų bendrines formas, kurios yra įtraukiamos į žodyną. Daiktavardinėms frazėms yra pristatyta struktūra pagrįsta tyrimais Juston ir Katz(1995). Žemiau pateiktame paveikslėlyje pateikiama kaip atpažįstama daiktavardinės frazes anglų kalboje (9 pav.) [9].



9 Pav. Daiktavardinių frazių struktūra [9]

DT- sprendėjas, VBG- apibudina specifinę veiksmažodžio formą, JJ- apibūdina būdvardį, CC- apibūdina jungtuką, NN- apibūdina daiktavardžius. Tokia struktūra tinka ir lietuvių kalbai. Pavyzdžiui sakinys “Mes skaičiuojame kalendorines dienas“. Frazė bus kalendorinės dienos. Arba “ Aš turiu sporto inventoriaus“ šiuo atveju frazė būtų sporto inventorius.

1.5.3. Atpažinimas kandidatų į žodyno elementus pagal jų formas

Bendra strategija atpažinti techninę terminiją , tai surasti daiktavardines fraze, o jas rasti galima pagal leksika(mokslinė, buitinė), kalbos sudėtį(veiksnys tarinys papildinys), filtrais. Atpažinimas galimo kandidato į žodyno elementus negalima padaryti tik paprastu leksikos filtru. Taipogi atpažinti žodyno elementą susiduriama su problema dėl dviprasmybių, kuri negali būti išbręsta tik leksiniu peržiūrėjimu.

Norint integruoti daugybę paruošiamųjų darbų prieš paleidžiant žodyno gavima, lingvistikos analize paremta rašyti per tekstą anotacijas(apie žodį trumpa santrauka pvz. kalbos dalis, linksnis). Principe nėra svarbu anotacijų skaičius ar tipas: žodis, vardų sąrašas, specialios frazes, gabaliukai, sakiniai visa tai įeina į anotacija. Leidžiant per lingvistikos analize , paslepia žodžių išsidėstymo tvarką. Pavyzdys žodyno elementų yra pateiktos žemiau esančioje lentelėje. A-būdvardis, N-daiktavardis ir C-jungtukas.

Structure	Terms
AN	genuine part
NN	sport utilities
AAN	heavy commercial use
ANN	rear wiper blade
NNN	emission control system
AANN	other qualified service technician
ACAN	unpaved or dusty roads
ANNN	automatic transmission fluid level
NNNN	engine oil fluid level
AANNN	new personalized oil reset percentage
AACAN	certain frontal or near-frontal collision
ACAAN	ambient and wide open throttle
NNNNN	steering wheel fan speed control

10 Pav. Kandidatų lentelė

Jeigu žodyno elementas yra naudojamas daug dažniau specifiniame dokumente nei kituose dokumentų kolekcijose, tai tikriausiai rastas elementas yra to dokumento specifinis terminas. Šis algoritmas skaičiuoja tikimybė termino pasirodymo specifiniame dokumente nei paprastame dokumentų tekstynuose(*ang.l: text corpus*). Tai užrašoma lygtimi:

$$\frac{\sum_{w_i \in T} \frac{P_d(w_i)}{P_c(w_i)}}{\mid T \mid}$$

kur T yra žodžių skaičius elemente, $P_d(w_i)$ tikimybe žodžio w_i specifiniame dokumente, $P_c(w_i)$ tikimybė žodžio w_i dokumentų rinkiniuose. Tikimybės nustatomos pagal dokumentų didžius.

1.5.4. Frazių ir terminų radimo algoritmų analizės išvados

1) Atlikta algoritmų analizė parodė jog nei vienas iš analizuotų algoritmų nesuteikia galimybės daiktavardinių bei veiksmažodinių frazių ieškoti lietuviškame tekste.
2) Algoritmų analizė jog frazes galima ieškoti naudojant frazių sudedamąsias struktūras.
3) Atlikta analizė parodė jog lietuviškam daiktvardinių ir veiksmažodinių frazių paieškai tinkamas algoritmas yra ieškoti pagal žodžių formas.
4) Lietuviško teksto analizei, t.y. daiktavardinių bei veiksmažodinių frazių paieškai atlikti, bus remiamasi žodžių formomis grįsto algoritmu. Formos kuriomis bus atpažįstamos frazės bus minimos kaip – šablonai.

1.6. Siekiamo sprendimo apibrėžimas

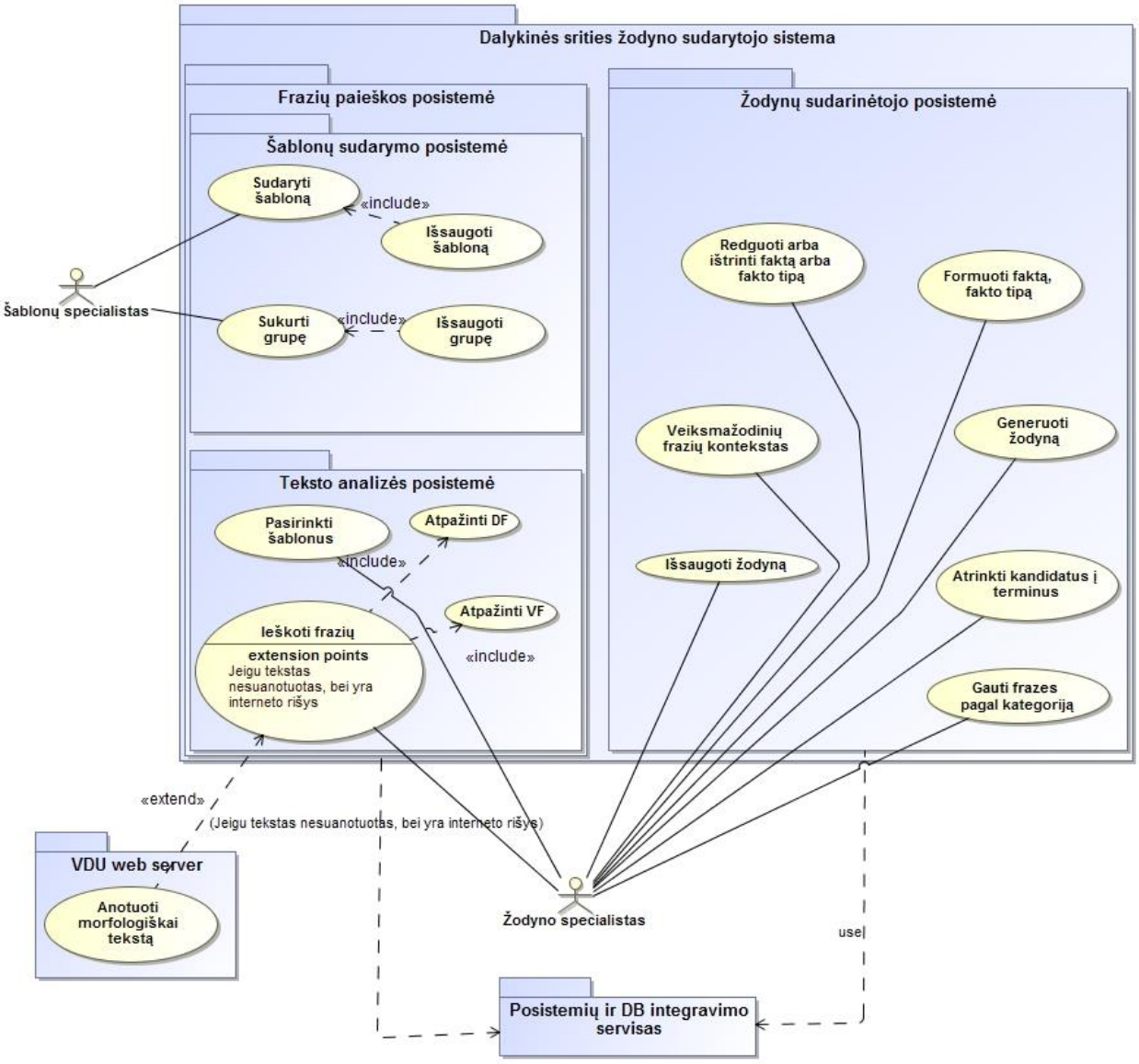
Daiktavardinių bei veiksmažodinių frazių radimo algoritmas, turi pagal vartotojo pateiktus šablonus išskirti iš teksto daiktavardines bei veiksmažodines frazes. Taipogi sistema turėtų teikti galimybę iš teksto išskirti, vardus ir pavardes, sinonimus, sutrumpinimus.

Bendruoju atveju teksto analizės sistema turi atlikti šiuos veiksmus:

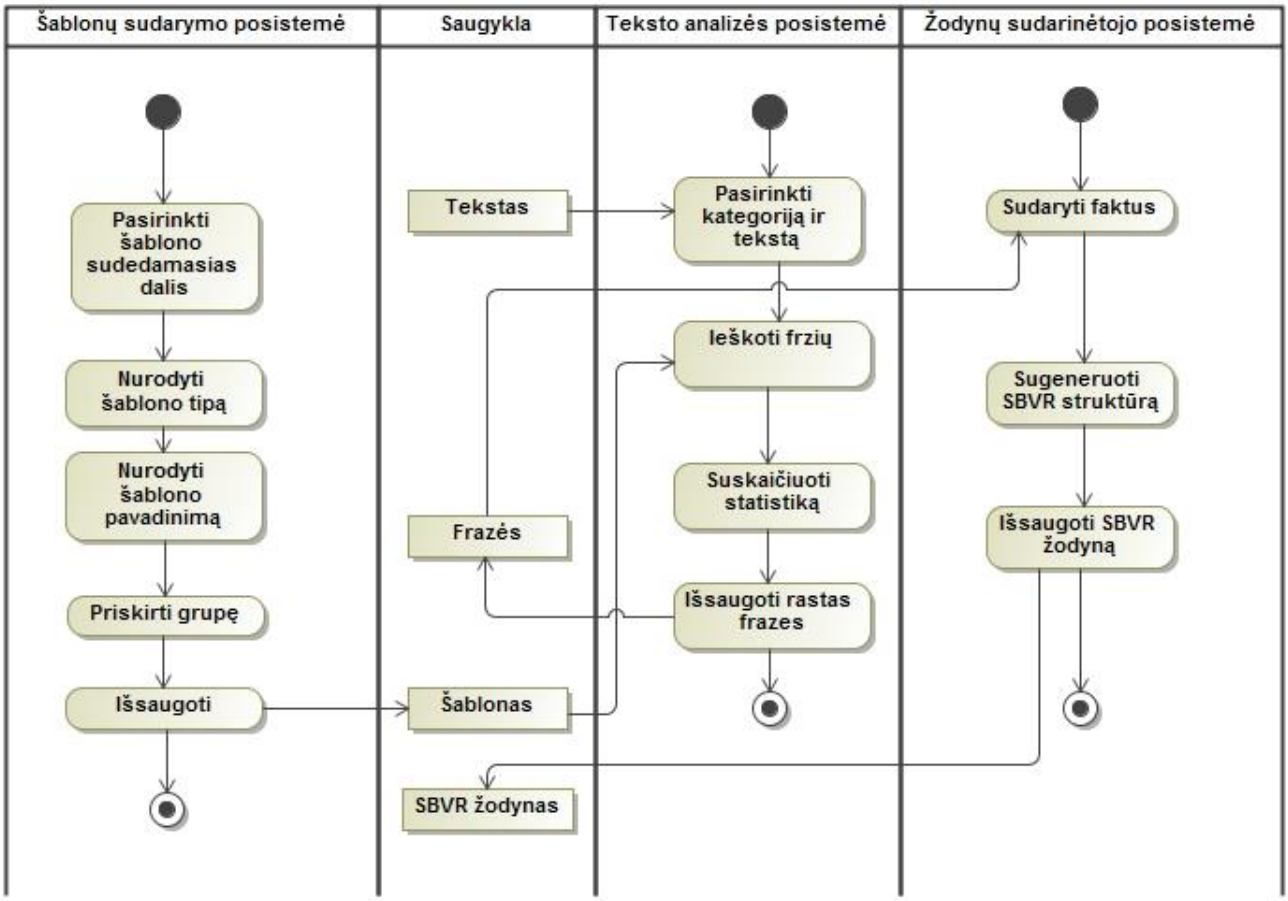
- Leisti vartotojui sudaryti bet kokio tipo šablonus;
- Pripažinti bei ieškoti frazių nepriklausomai nuo šablono sudėtingumo;
- Suteikti galimybę vartotojui peržiūrėti rastus įrašus;
- Leisti įrašus redaguoti bei juos saugoti;

1.6.1. Dalykinės srities žodyno sudarytojo sistemos panaudojimo atvejų diagrama

Panaudojimo atvejų diagramoje pavaizduojami pagrindiniai komponentai kurie sudaro dalykinės srities žodyno sudarinėtojo sistemą. Kaip matyti iš (11 Pav. Dalykinės srities žodyno sudarytojo sistemos panaudojimo atvejų diagrama) sistemą sudaro dvi pagrindinės dalys, tai yra: Frazių paieškos posistemė bei žodynų sudarinėtojo posistemė kuri skirta SBVR žodynui generuoti.



11 Pav. Dalykinės srities žodyno sudarytojo sistemos panaudojimo atvejų diagrama



12 Pav. Dalykinės srities žodyno sudarytojo sistemos veiklos diagrama

Sistemos veiklos diagramoje yra pavaizduoti pagrindiniai panaudojimo atvejų veiklos etapai (12 Pav. Dalykinės srities žodyno sudarytojo sistemos veiklos diagrama). Sistemos veiklos diagrama išskirta į keturias pagrindines t.y. šablonų sudarinėjimo posistemę, saugyklą, teksto analizės posistemę bei žodynų sudarinėjimo posistemę. „Ieškoti frazių“ panaudojimo atvejų detalesnis pavaizdavimas pateiktas 3 Pav. Frazių atpažinimo veiklos diagrama

1.6.2. Dalykinės srities žodyno sudarinėtojo funkciniai reikalavimai

Pagrindiniai sistemos funkciniai reikalavimai:

***Sudaryti šabloną*:**

Vartotojas turi turėti galimybę sudaryti šablonus, pagal kuriuos tekste ieškos frazių. Taipogi vartotojui sudarant šablonus turi būti galimybė, frazių šablonus išskirti į dvi grupes t.y. *daiktavardinių frazių šablonas* ar *veiksmažodinių frazių šablonas*.

***Išsaugoti šabloną*:**

Vartotojas turi turėti galimybę išsaugoti šablonus. Šablonus vartotojas turi galėti išsaugoti keliose vietose t.y. išsaugoti asmeniniame kompiuteryje, arba jei yra interneto ryšys išsaugoti duomenų bazėje.

***Sukurti grupę*:**

Vartotojas turi turėti galimybę sukurti grupę, į kurią gali pridėti sudarytus šablonus.

***Išsaugoti grupę*:**

Vartotojas turi galėti išsaugoti sudarytas grupes bei išsaugoti priskirtus šablonus prie grupių. Šis veiksmas yra galimas tik tada kai yra interneto ryšys, kadangi visi duomenys apie šablonus bei grupes yra saugomi duomenų bazėje.

***Gauti šablonus*:**

Vartotojas turi turėti galimybę gauti šablonus, kuriuos jis yra sudaręs.

***Ieškoti frazių*:**

Vartotojui turi būti sudaryta galimybė ieškoti frazių tekste. Frazių paieška turi būti išskaidyta į kelias grupes, tokias kaip: Vardų bei pavardžių paieška, sutrumpinimų paieška, sinonimų paieška bei daiktavardinių ir veiksmažodinių frazių paieška naudojant vartotojo sudarytus šablonus.

***Ieškoti DF (daiktavardinių frazių)*:**

Sistema turi galėti ieškoti daiktavardinių frazių, naudojant vartotojo sudarytus šablonus.

***Ieškoti VF (veiksmažodinių frazių)* :**

Sistema turi galėti ieškoti veiksmažodinių frazių, naudojant vartotojo sudarytus šablonus.

***Anotuoti morfologiškai tekstą*:**

Sistema turi pati sugebėti išanotuoti pateiktą tekstą. Sistema gali nepriimti, arba neleisti anotuoti teksto dėl interneto ryšio nebuvimo, kadangi dėl teksto anotavimo sistema kreipiasi į nutolusį VDU servisą.

***Sudaryti faktą*:**

Sistema turi galėti bei vartotojui leisti sudaryti faktus, iš rastų daiktavardinių bei veiksmažodinių frazių.

***Išsaugoti faktą*:**

Sistema turi leisti vartotojui išsaugoti sudarytus SBVR faktus, į specialų SBVR duomenų failą, kad sudarytą žodyną būtų galima naudoti ir kitose sistemose.

1.6.3. Dalykinės srities žodyno sudarinėtojo sistemos nefunkciniai reikalavimai

1.6.3.1. Reikalavimai sistemos išvaizdai

Bendri reikalavimai teminės srities žodyno sudarinėtojo vartotojo sąsajai:

- Lengvai suprantama sąsaja;
- Neįkyri sąsaja (pvz.: nereikalauja pastoviai ką nors vis patvirtinti);
- Aiškus bei patogus pagrindinis valdymo meniu;
- Sąveikaujanti sąsaja;
- Suprantama ir aiškiai išdėlioti bei paaiškinti funkcionalumo mygtukai;

1.6.3.2. Reikalavimai panaudojamumui

Panaudojamumo paprastumas (lengvumas) teminės srities žodyno sudarinėtojo kriterijai:

- Klaidų poveikių mažinimas;
- Galimybė bet kada gauti naujausius sistemos atnaujinimus;
- Procesų vykdymo vizualizacija;
- Galimybė patogiai vartotojui pateikti tiek patį šabloną, tiek rastas frazes;
- Patogus ir lengvai suprantamas šablonų kūrimo sąsaja.
- Patogi ir lengva frazių bei faktų redagavimo sąsaja;

1.6.3.3. Reikalavimai veikimo sąlygoms

Sistema turi veikti windows 32 ir windows 64 bitų operacinėse sistemose: windows xp, windows vista, windows 7, windows 8, windows 8.1.

Norint pilnai naudotis sistema, būtina turėti interneto ryšį. Interneto ryšys būtinas tam, jog sistema galėtų susisiekti su servisu, kuris aptarnauja programą duomenimis iš duomenų bazės.

1.6.3.4. Saugumas

Toliau pateikiami teminės srities žodyno sudarinėtojo sistemos saugumo aspektai:

- Konfidencialumas – sistemoje saugomi tik išanalizuotų tekstų duomenys, nesaugoma jokia kita su asmeniu ir jo tapatybę galinti atskleisti informaciją.
- Privatumas – kiekvienas vartotojas parsijunkdamas prie sistemos identifikuojamas atskirai, tad nėra galimybės pamatyti svetimo vartotojo duomenų.
- Pasiekiamumas – pagrindiniai sistemos duomenys su kuriais operuoja sistemas yra rastos frazės bei tekstai ir vartotojo sudaryti šablonai. Visi šie duomenys dėl patikimumo ir didesnio pasiekiamumo saugomi duomenų bazėje.

1.7. Analizės išvados

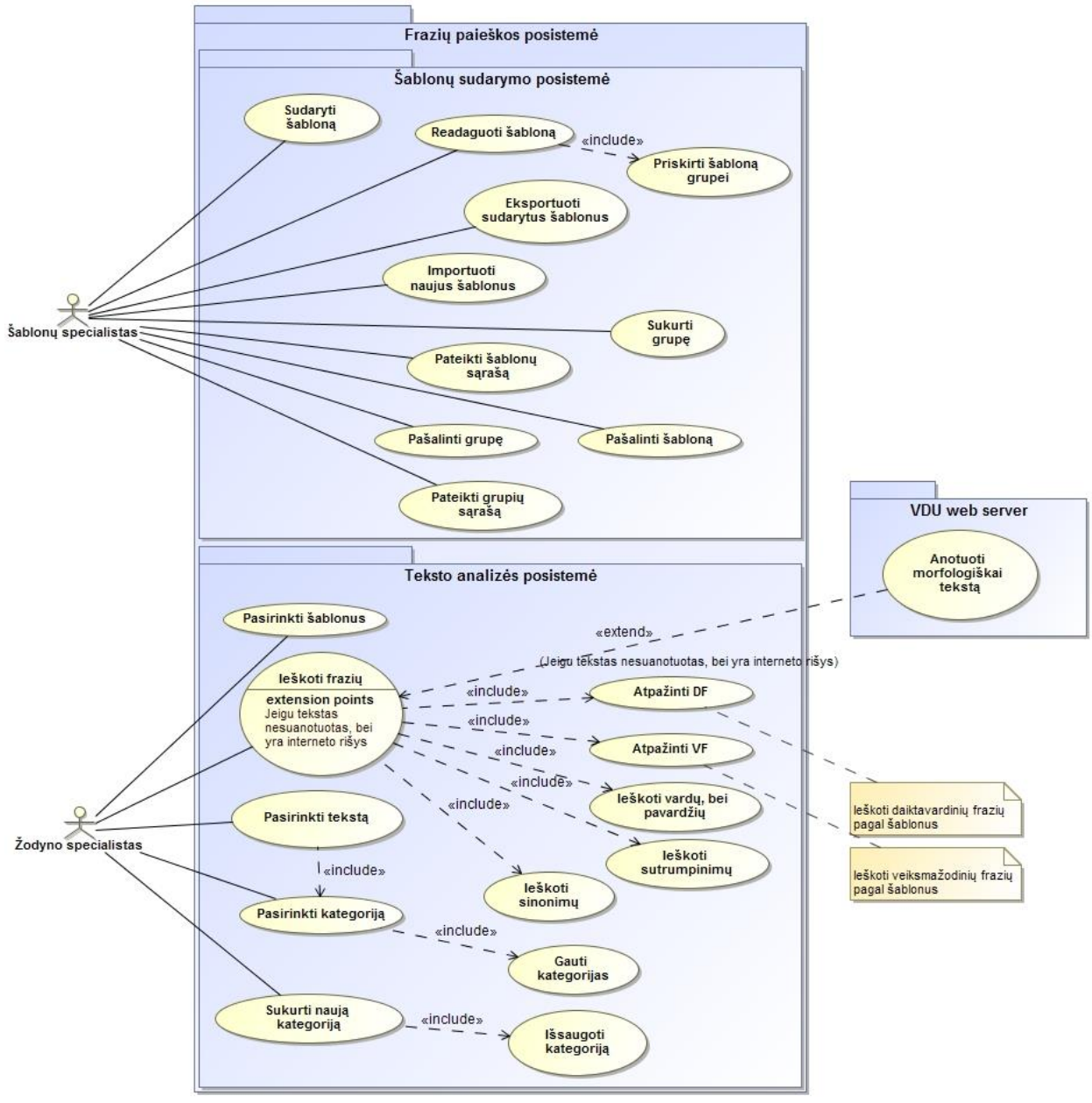
1) Atlikta analizė leidžia teigti, jog frazių ar žodžių junginių radimas tekstuose ar žodynų sudarinėjimas yra gan sudėtingas ir laiko reikalaujantis mechanizmas.
2) Atlikus panašių sistemų analizę, buvo pastebėta, jog nei viena iš analizuotų sistemų netinka lietuviško teksto analizei. Taipogi nei viena iš sistemų neturėjo galimybės pačiam vartotojui nustatyti ieškomų frazių, junginių ar kitokių struktūrų žodžių, sudaryti numanomus šablonus pagal kuriuos sistema galėtų rasti bei aptikti frazes ar junginius.
3) Iš atliktų teminės srities žodynų sudarinėtojų programų analizės galima teigti, jog lietuviško teksto žodynų sudarymas gali būti gan sudėtingas. Norėdami palengvinti frazių radimą bei pačios programos kūrimą galima bus remtis, jau esamomis kitoms kalboms skirtomis funkcijomis bei paieškos metodais.

2. FRAZIŲ PAIEŠKOS POSISTEMĖS SPECIFIKACIJA IR ANALIZĖ

2.1. Reikalavimų specifikacija

Pateiksime (11 Pav. Dalykinės srities žodyno sudarytojo sistemos panaudojimo atvejų diagrama) sistemos panaudojimo atvejų „Frazių paieškos posistemė“ detalesnį vaizdą bei panagrinėsime ir aprašysime pagrindinius funkcinius bei nefunkcinius reikalavimus.

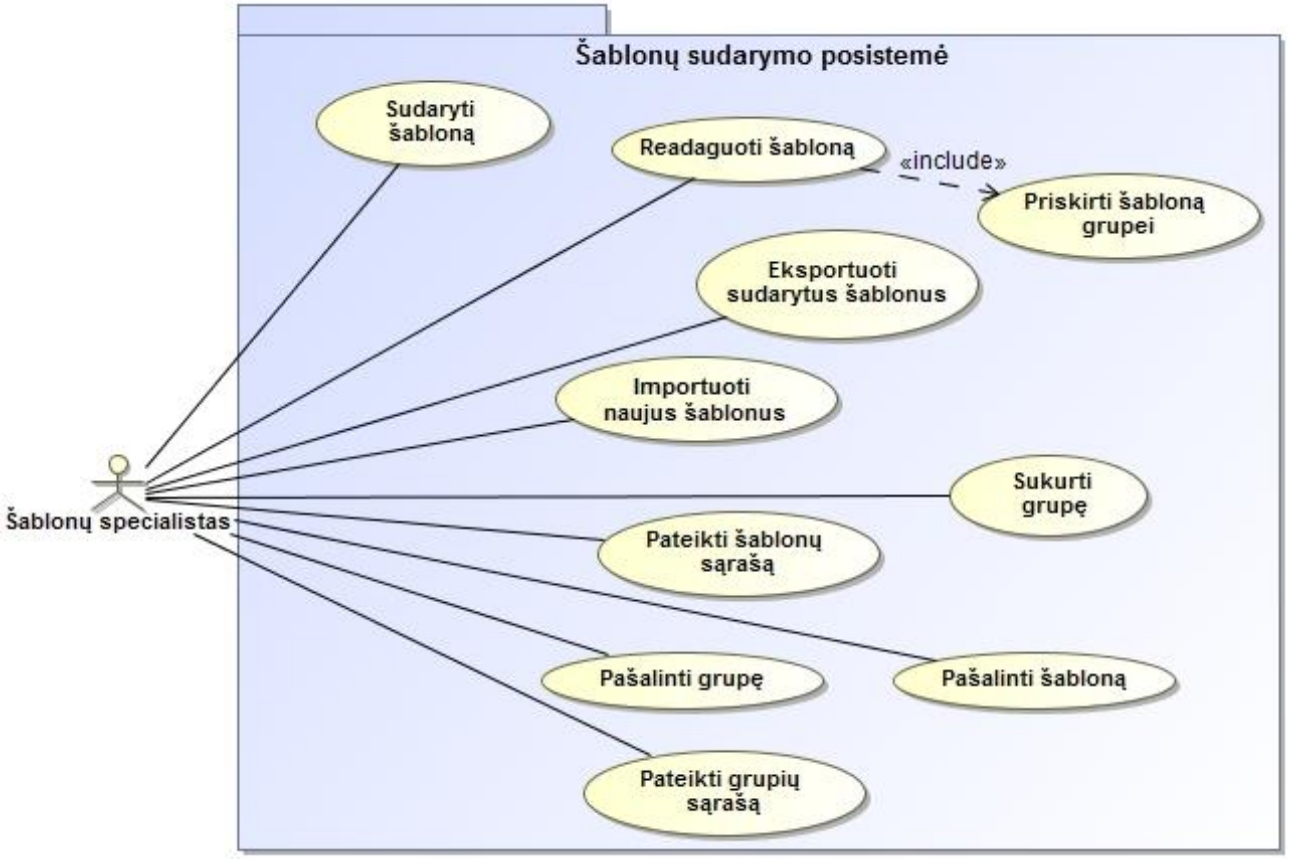
Žemiau (13 Pav. Frazių paieškos posistemės panaudojimo atvejų diagrama) pateiktoje „Frazių paieškos posistemėje“ matyti, jog ji išsiskiria į du paketus. Toliau panagrinėsime abiejų šių paketų funkcinius reikalavimus, nefunkcinius reikalavimus, taipogi sekų ir veiklos diagramas.



13 Pav. Frazių paieškos posistemės panaudojimo atvejų diagrama

2.1.1. Šablonų sudarinėjimo posistemė

Žemiau pateiktoje panaudojimo atvejų diagramoje 14 Pav. Šablonų sudarinėjimo posistemės panaudojimo atvejų diagrama matome „Šablonų sudarinėjimo posistemę“. Toliau bus pateikta pagrindinių panaudojimo atvejų, funkciniai bei nefunkciniai reikalavimai, taipogi jų sekos ir panaudojimo diagramos.



14 Pav. Šablonų sudarinėjimo posistemės panaudojimo atvejų diagrama

2.1.1.1. Šablonų sudarymo posistemės funkcijos

Sudaryti šabloną – vartotojas turi turėti galimybę sudaryti, bet kokio tipo bei sudedamųjų dalių šabloną. Šablono sudarymo metu vartotojas turi galėti, priskirti šablonui vardą, nurodyti jo tipą. Jei sistema yra prijungta prie interneto ir yra pasiekiamas servisas, vartotojas tada turi turėti galimybę sudarytus šablonus išsaugoti duomenų bazėje.

Redaguoti šabloną – vartotojui turi būti suteikta galimybė redaguoti pasirinktą šabloną. Vartotojas pasirinkęs šabloną turi galėti, keisti šablono pavadinimą, grupę kuriai priklauso ir šablono tipą.

Priskirti šabloną grupei – vartotojas turi turėti galimybę priskirti šabloną grupei. Grupės tikslas yra atskirti tam tikrus frazių paieškos šablonus, pagal vartotojo reikalavimus. Taipogi vartotojui turi būti galima tam tikras jaus sudarytas frazes priskirti ir kitoms grupėms, ar šablonus pašalinti iš kitų grupių.

Eksportuoti sudarytus šablonus – galimybė išeksportuoti pasirinktus šablonus į XML formatą.

Importuoti sudarytus šablonus – galimybė importuoti XML formate esančius šablonus.

Sukurti grupę – vartotojas turi turėti galimybę sukurti naują grupę. Jeigu sistema turi prieigą prie interneto ir taipogi servisas funkcionuoja, tai vartotojas turi galėti naujas suskurtas grupes išsaugoti duomenų bazėje.

Išsaugoti šabloną – sistema turi galėti išsaugoti vartotojo sudarytus šablonus.

Gauti grupes – sistema turi vartotojui pateikti, būtent jo sudarytas grupes bei prie grupių automatiškai priskirti šablonus.

Gauti šablonus – sistema turi vartotojui pateikti visus jo sudarytus šablonus.

Išsaugoti grupę – sistema turi galėti išsaugoti vartotojo sukurtą naują grupę.

2.1.1.2. Šablonų sudarymo posistemės nefunkciniai reikalavimai

- **Reikalavimai išvaizdai:**

Šablonų sudarinėjimo langas turi būti lengvai suprantamas bei aiškus. Vartotojui turi būti aiškiai ir suprantamai išdėstyti elementų t.y. kalbos dalių, giminių, laipsnių ir t.t. pasirinkimas bei pridėjimas į šabloną.

- **Reikalavimai saugumui:**

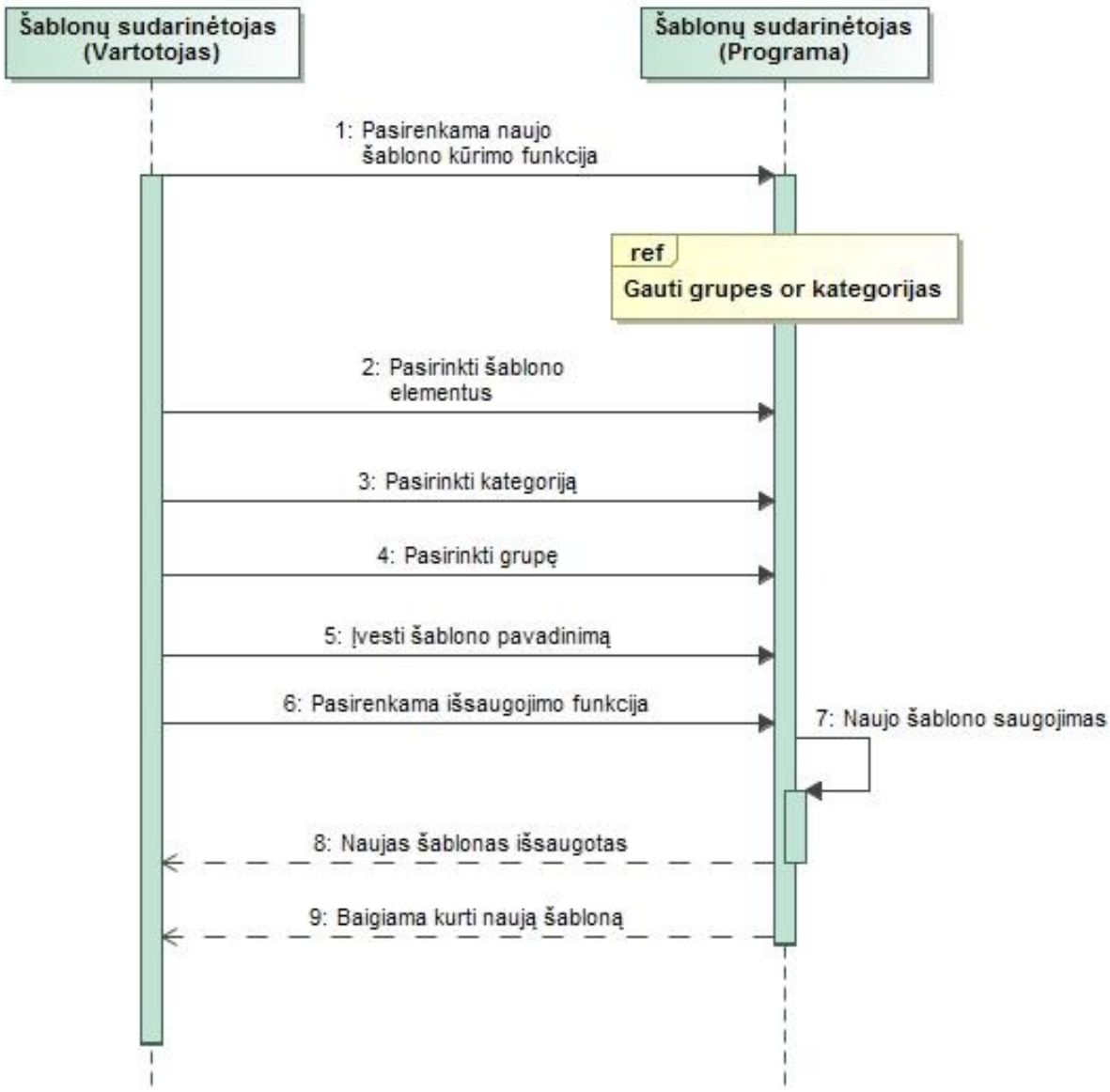
Vartotojai negali matyti vienas kito šablonų, nebent jie juos apsikeičia *eksportuodami* ir *importuodami* kitam asmeniui.

Išeksportuoti frazių šablonai privalo turėti unikalų saugumo raktą, kuris yra tinkamas tik vienam vartotojui, taip apsaugant dėl duomenų nutekėjimo kitam vartotojui.

Duomenų perdavimas iš vartotojo sistemos į duomenų saugojimo saugyklą, privalo būti saugus, kad nebūtų prarandama duomenų.

2.1.1.3. Šablonų sudarymo posistemės sekų diagramos

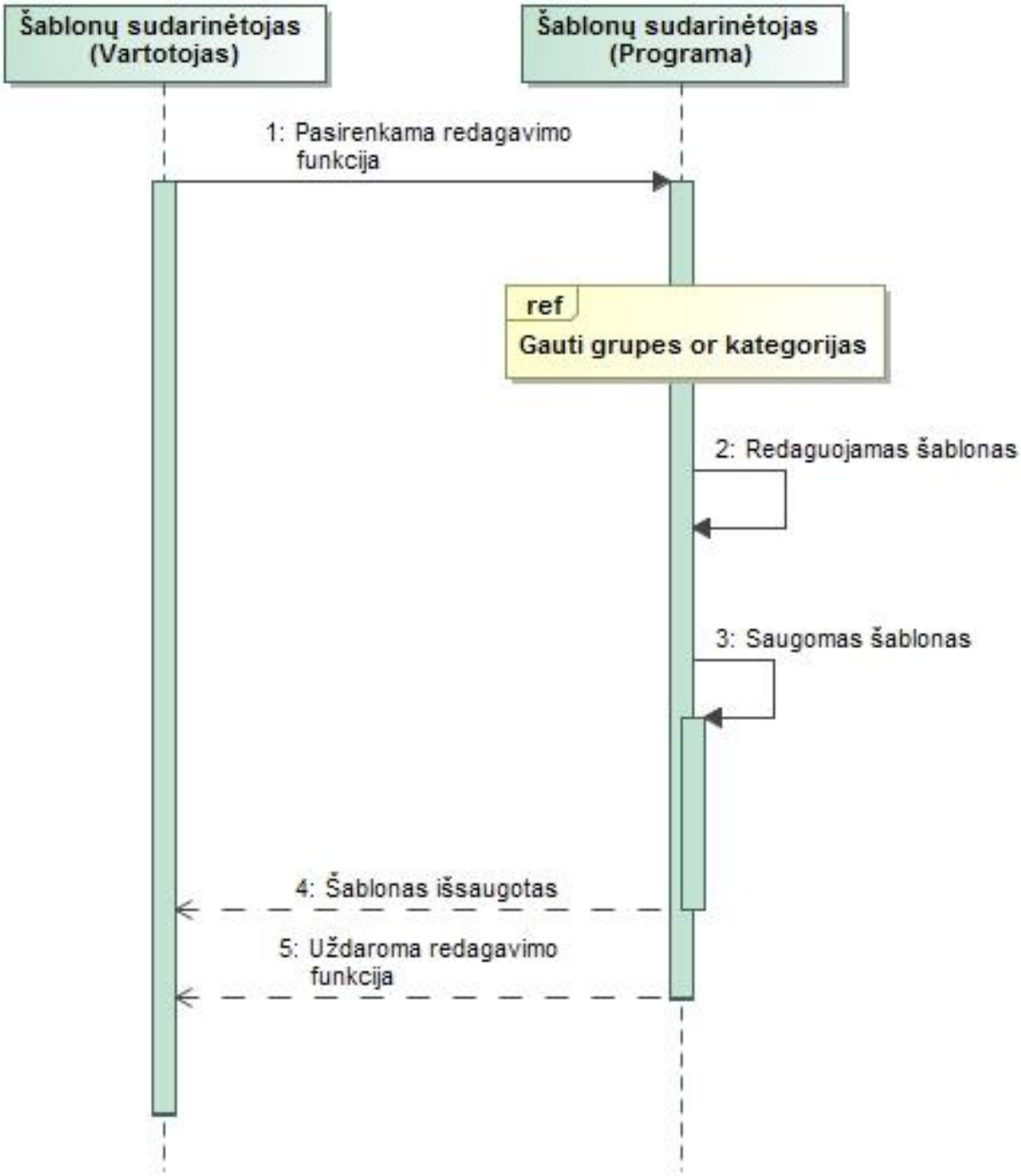
Žemiau esančioje sekų diagramoje detaliau panagrinėsime panaudojimo atvejo „*Sudaryti šabloną*“ scenarijų, kurio sekų diagrama pateikta (15 Pav. P.A. „Sudaryti šabloną“ sekų diagrama) paveiksle.



15 Pav. P.A. „Sudaryti šabloną“ sekų diagrama

Pagrindiniame sistemos lange pasirinkus funkciją frazių šablono sudarinėtojas yra atidaromas šablonų sudarinėtojo langas. Vartotojas šiame lange gali pasirinkti šablono elementus, t.y. kalbos dalį, giminę, skaičių, linksnį, taipogi pasirinkti kategoriją, įvesti šablono grupę, įvesti šablono pavadinimą. Po to gali visus šiuos pakeitimus išsaugoti. Kada naujas šablonas yra išsaugomas sistemas automatiškai išvalo visus sistemos laukus, taip paruošdama langą naujo šablono kūrimui.

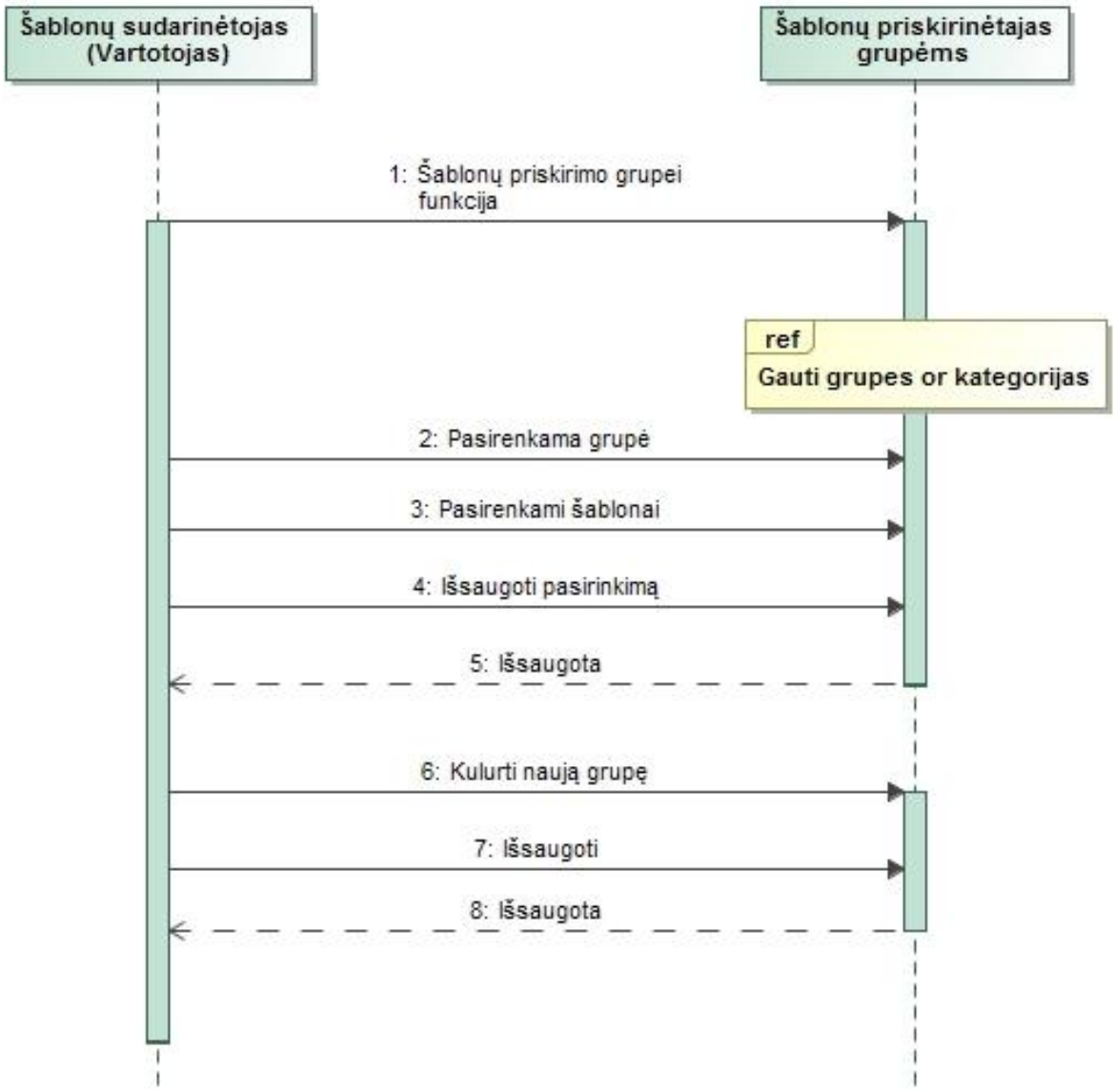
Žemiau esančiame paveiksle detaliau panagrinėsime panaudojimo atvejo „Redaguoti šabloną“ scenarijų, kurio sekų diagrama pateikta (16 Pav. P.A. „Redaguoti šabloną“ sekų diagrama) paveiksle.



16 Pav. P.A. „Redaguoti šabloną“ sekų diagrama

Šablonų sudarinėjimo lange pasirinkus redagavimo funkciją, t.y. pasirinkus tam tikrą šabloną ir nuspaudus redagavimo funkciją, vartotojui suteikiama galimybė redaguoti šabloną. Vartotas redaguojamam šablonui galima pakeisti kalbos dalį, giminę, skaičių ir kita. Taipogi suteikiama galimybė redaguoti pavadinimą, pakeisti grupę, ar šablono tipą. Vartotojui baigus redagavimą, sistema naujus pakeitimus išsaugo duomenų bazėje. Kai pakeitimai išsaugomi, sistema automatiškai užbaigia redagavimą, taip užbaigdama redagavimo funkciją.

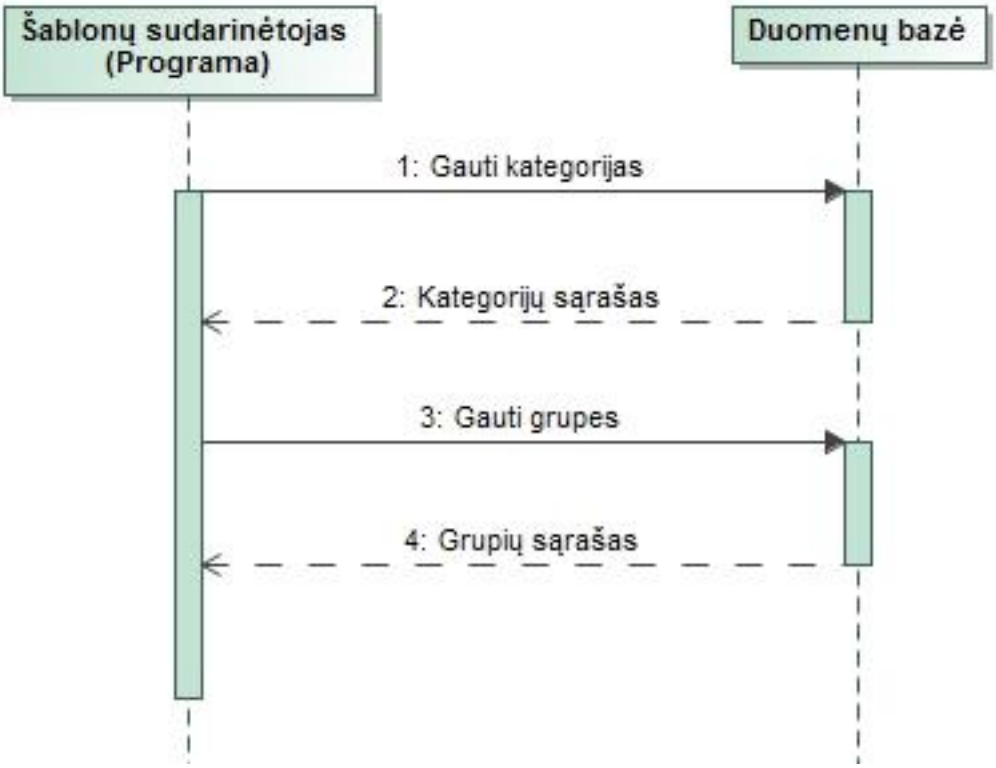
Detalesnė panaudojimo atvejo „Priskirti šabloną grupei“ sekų diagrama pateikta (17 Pav. P.A. „Priskirti šabloną grupei“ sekų diagrama) paveiksle.



17 Pav. P.A. „Priskirti šabloną grupei“ sekų diagrama

Šablonų sudarinėtojas taipogi turi galimybę šablonus priskirti tam tikroms grupėms taip juos sugrupuodamas. Pasirinkus šablonų priskyrimo funkciją vartotojui atsidarymo papildomas langas su vartotojo grupėmis ir šablonais. Vartotojas pirmiausia pasirenką grupę po to pasirenka šablonus kurios nori priskirti būtent šiai grupei. Pasirinkus atitinkamus šablonus specialiai grupei, galima šiuos duomenis išsaugoti. Dar viena papildomą funkcija šablonų priskyrimo funkcijoje yra naujo šablono sukūrimas.

Panaudojimo atvejo „*Gauti kategorijas ir grupes*“ sekų diagrama pateikiama (18 Pav. P.A. „Gauti kategorijas ir grupes“ sekų diagrama) paveiksle.

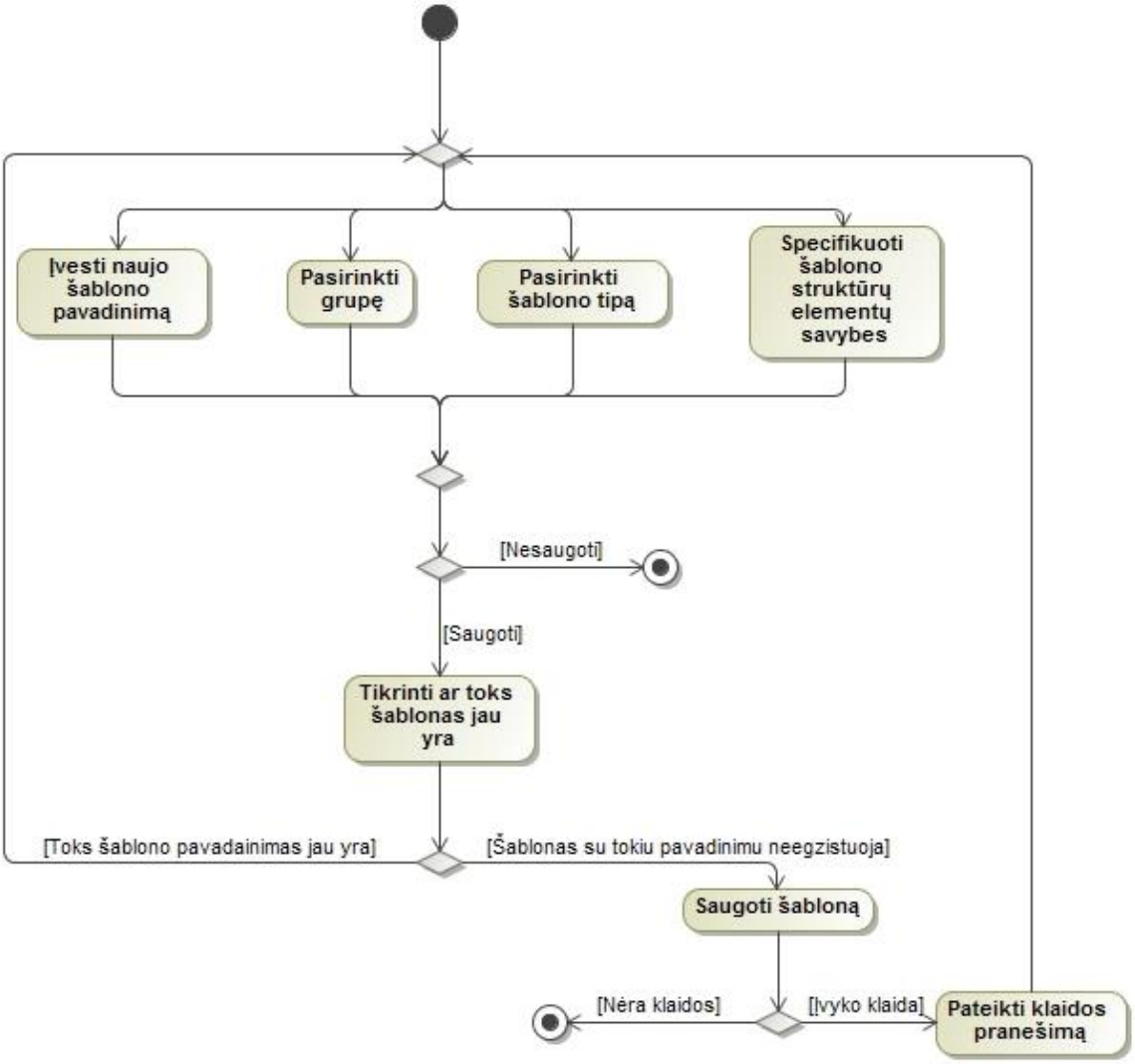


18 Pav. P.A. „Gauti kategorijas ir grupes“ sekų diagrama

Kiekvienas vartotojas turi savo sudarytas, grupes, kategorijas ir taipogi šablonus. Aukščiau pavaizduotoje veiklos diagramoje vaizduojama sekų diagrama kuri parodo kaip sistema gauna kategorijas bei grupes.

2.1.1.4. Šablonų sudarymo posistemės veiklos diagramos

Žemiau esančioje veiklos diagramoje yra vaizduojamas šablono sudarymas:

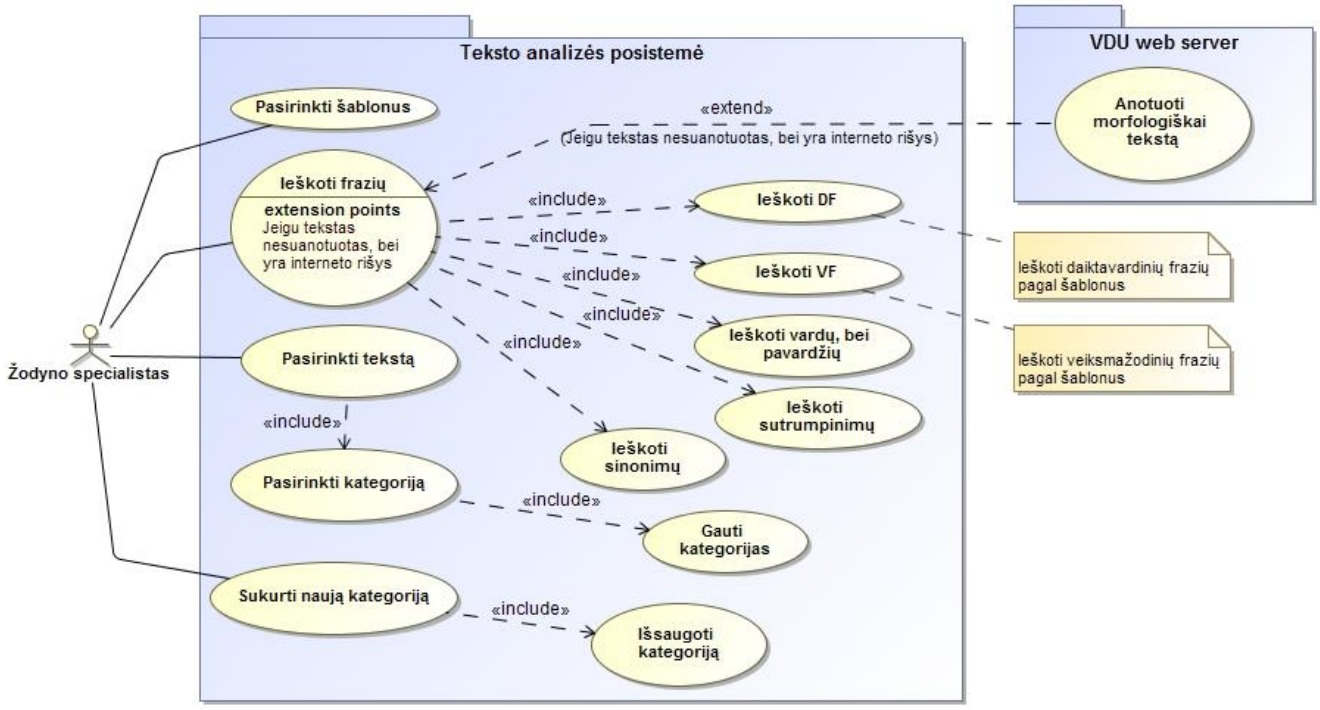


19 Pav. P.A. „Sudaryti šabloną“ veiklos diagrama

Šablonas susideda iš keturių pagrindinių dalių, t.y. pavadinimo, grupės, tipo bei struktūrų elementų savybių. Šios keturios dalys yra privalomos kuriant šablonus. Vartotojas pasirinkęs visus elementus gali jų nesaugoti, tačiau jei pasirenkama saugoti, tai sistema automatiškai ar toks šablono pavadinimas jau egzistuoja pas vartotoją. Jei šablonas su tokiu pavadinimu jau egzistuoja tai sistema jį informuoja, jog šablonas su tokiu pavadinimu jau egzistuoja. Jeigu šablonas su tokiu pavadinimu dar neegzistuoja pas vartotoją, tai sistema jį išsaugo, tačiau jei įvyksta sistemoje klaida, šablonas neišsaugomas.

2.1.2. Teksto analizės posistemė

Žemiau pateiktoje panaudojimo atvejų diagramoje (20 Pav. Teksto analizės posistemės panaudojimo atvejų diagrama) matome „Teksto analizės posistemę“. Toliau bus pateikta pagrindinių panaudojimo atvejų, funkciniai bei nefunkciniai reikalavimai, taipogi jų sekos ir panaudojimo diagramos.



20 Pav. Teksto analizės posistemės panaudojimo atvejų diagrama

2.1.2.1. Teksto analizės posistemės funkcijos

Pasirinkti šablonus – vartotojui turi būti suteikta galimybė bei įrankis kuriuo jis galėtu sudaryti šablonus. Vartotojui turi būti pilna galimybė sudaryti tam tikros sudėties šablonus su tam tikrais šablono elementais.

Ieškoti frazių – vartotojui turi būti suteikta galimybė ieškoti įvairaus tipo frazių tokių kaip: *daiktavardinės frazės*, *veiksmažodinės frazės*, *vardai bei pavardės*, *sutrumpinimai* taipogi *sinoniminių* žodžių.

Pasirinkti tekstą – vartotojui turi būti suteikta galimybė įkelti tekstą, ar tekstus kuriuos nori analizuoti. Taipogi sistema turi priimti įvairaus tipo dokumentus: *pfd.*, *doc.*, *txt.*. Prieš atliekant frazių paiešką vartotojui yra suteikiama galimybė, savo analizuojamus tekstus ar tekstą priskirti kategorijai.

Sukurti naują grupę – vartotojui turi būti suteikta galimybė susikurti naują grupę, į kurią vartotojas galės priskirti tam tikrus šablonus.

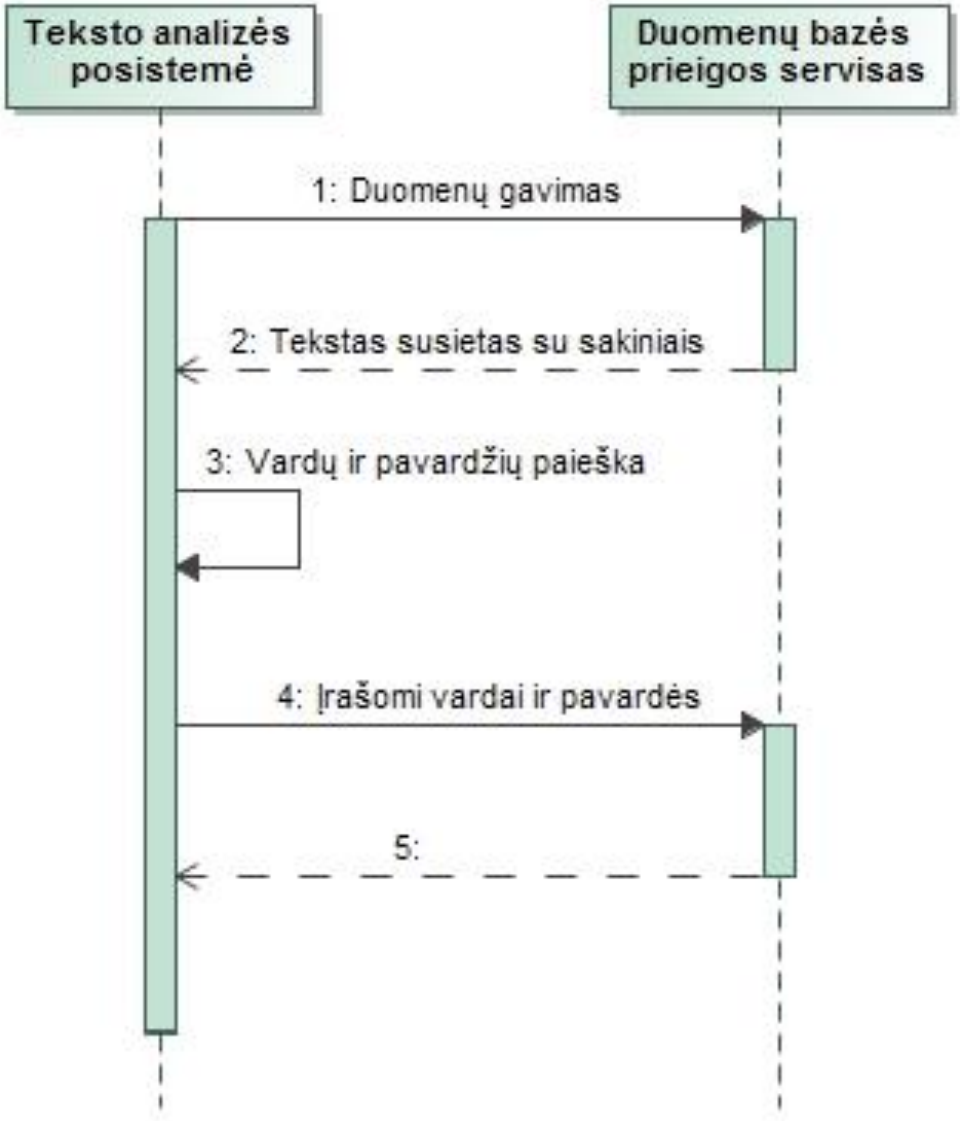
2.1.2.2. Teksto analizės posistemės sekų diagramos

Žemiau esančiame paveiksle (21 Pav. P.A. „Ieškoti frazių“ sekų diagrama) yra vaizduojamas frazių paieškos sekų diagrama, kuri apjungia šiuos pabaudos atvejus: „Frazių tarp kabučių paieška“, „Vardų pavardžių atrinkimas“, „Sinoniminių frazių paieška“, „Frazių atrinkimas pagal šablonus“.

21 Pav. P.A. „Ieškoti frazių“ sekų diagrama

Vartotojas pasirinkęs frazių atrinkimo funkciją, pradeda frazių paiešką. Šioje funkcijoje nuosekliai yra vykdomos 4 funkcijos susijusios su frazių radimu, kurių sekų diagramos yra pateiktos žemiau.

Panaudos atvejo „*Ieškoti vardų bei pavardžių*“ sekų diagrama pateikta (22 Pav. P.A. „Ieškoti vardų bei pavardžių“ sekų diagrama) paveiksle.

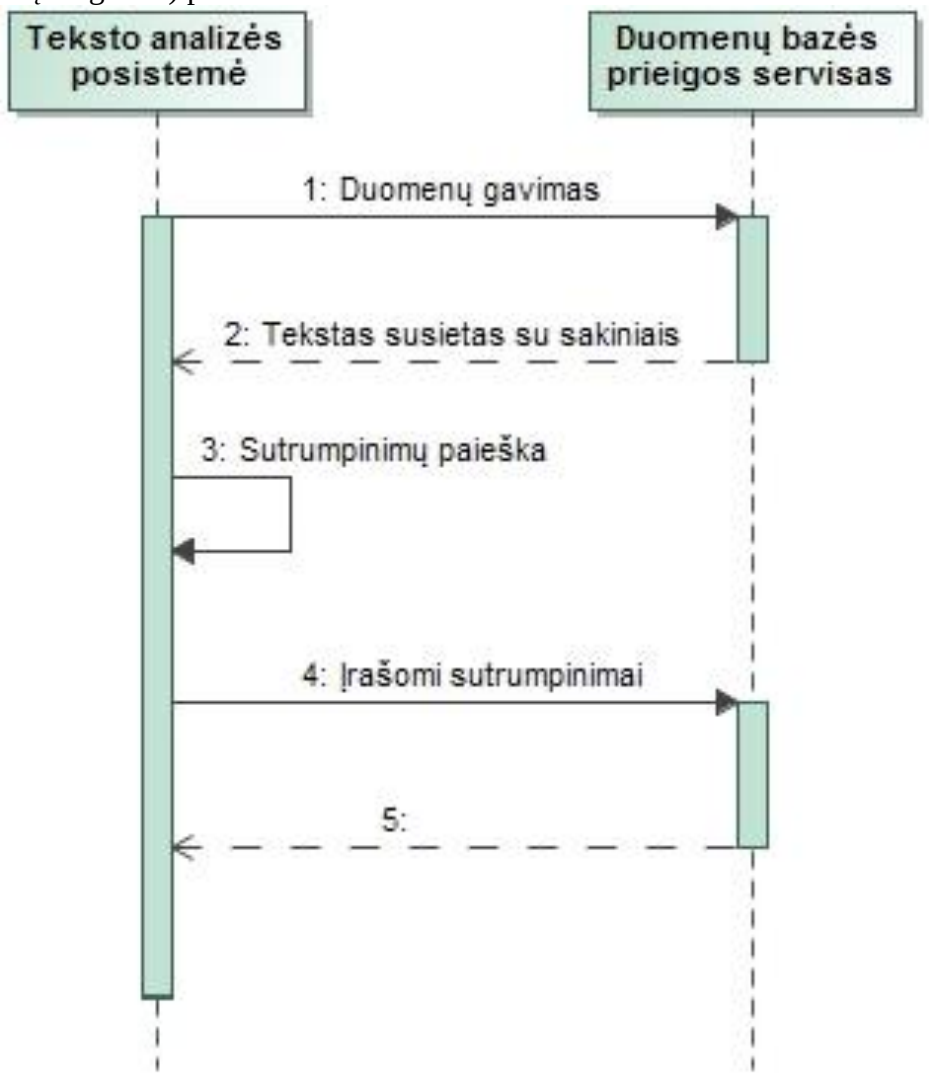


22 Pav. P.A. „Ieškoti vardų bei pavardžių“ sekų diagrama

„Ieškoti vardų bei pavardžių“ pirmiausia iš duomenų bazės atsisiunčia duomenis kuriose yra tekstas bei žodžiai susieti su sakiniais. Toliau vyksta paieška vardų ir pavardžių.

Baigus paiešką visi rasti vardai bei pavardės išsaugomos duomenų bazėje.

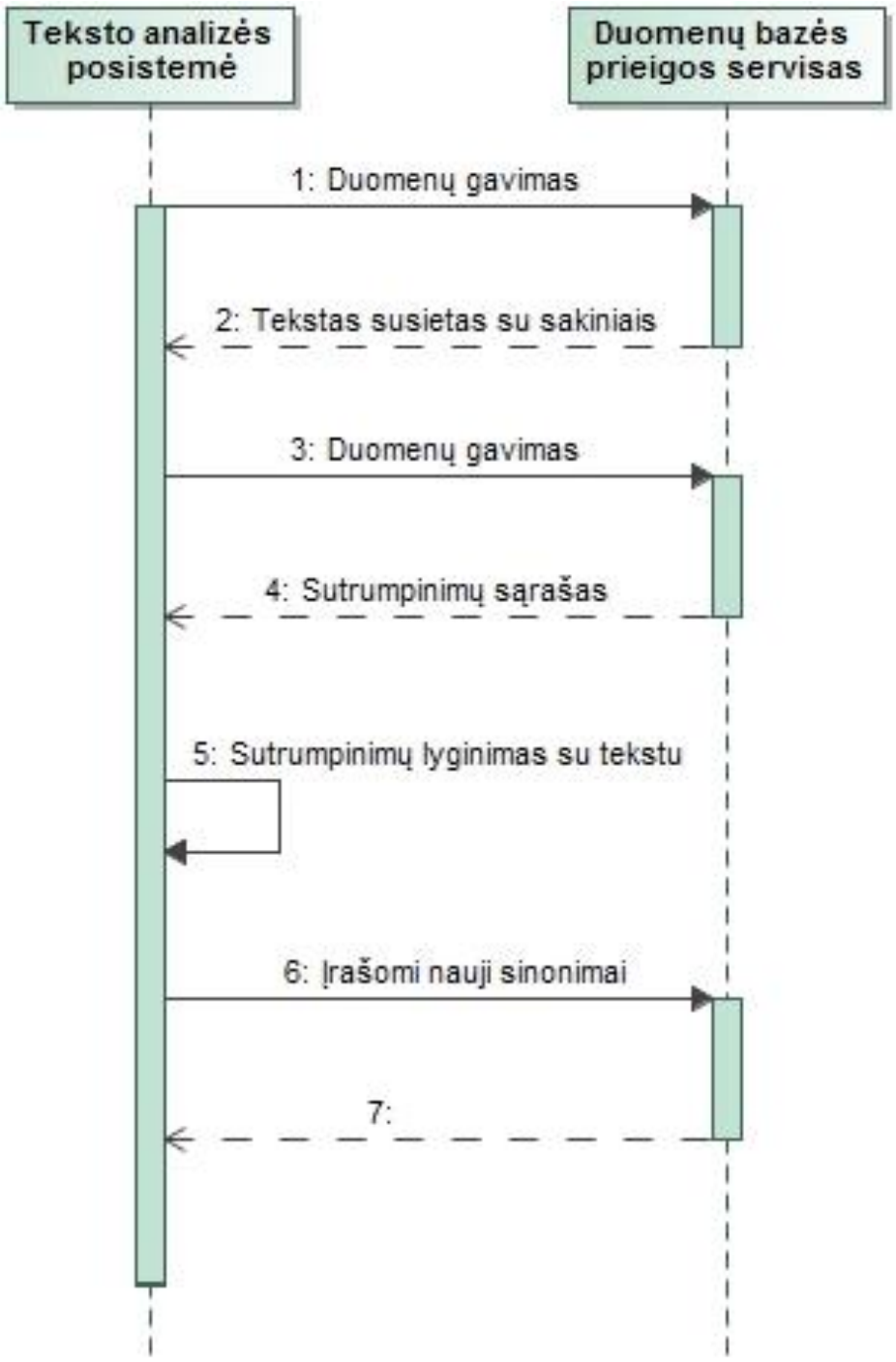
Panaudos atvejo „*Ieškoti sutrumpinimų*“ sekų diagrama pateikta (23 Pav. P.A. „Ieškoti sutrumpinimų“ sekų diagrama) paveiksle.



23 Pav. P.A. „Ieškoti sutrumpinimų“ sekų diagrama

„Ieškoti sutrumpinimų“ pirmiausia iš duomenų bazės atsisiunčia duomenis kuriose yra tekstas bei žodžiai susieti su sakiniais. Po to vyksta sutrumpinimų (KTU, VDU ir t.t.) paieška pagal joje esančias taisykles iki tol kol baigiasi duomenys, kurios atsisiuntė iš lokalios duomenų bazės. Baigus paiešką visi rasti sutrumpinimai išsaugomi duomenų bazėje.

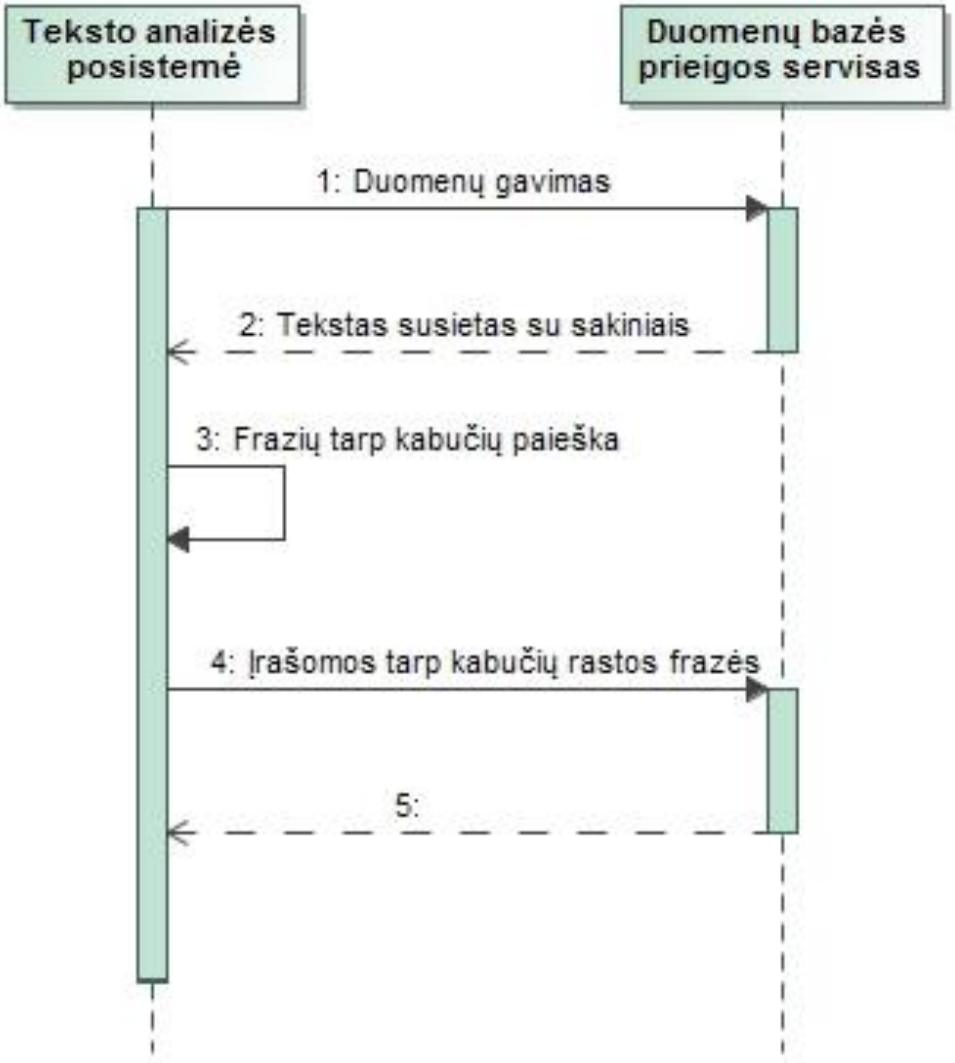
Panaudos atvejo „*Ieškoti sinonimų*“ sekų diagrama pateikiama (24 Pav. P.A. „Ieškoti sinonimų“ sekų diagrama) paveiksle.



24 Pav. P.A. „Ieškoti sinonimų“ sekų diagrama

Pirmiausia yra gaunamas, tekstas susietas su sakiniais toliau gaunamas sutrumpinimų sąrašas, tada lyginama sutrumpinimai su tekstu, bigę lyginti įrašome sinonimus į duomenų bazę.

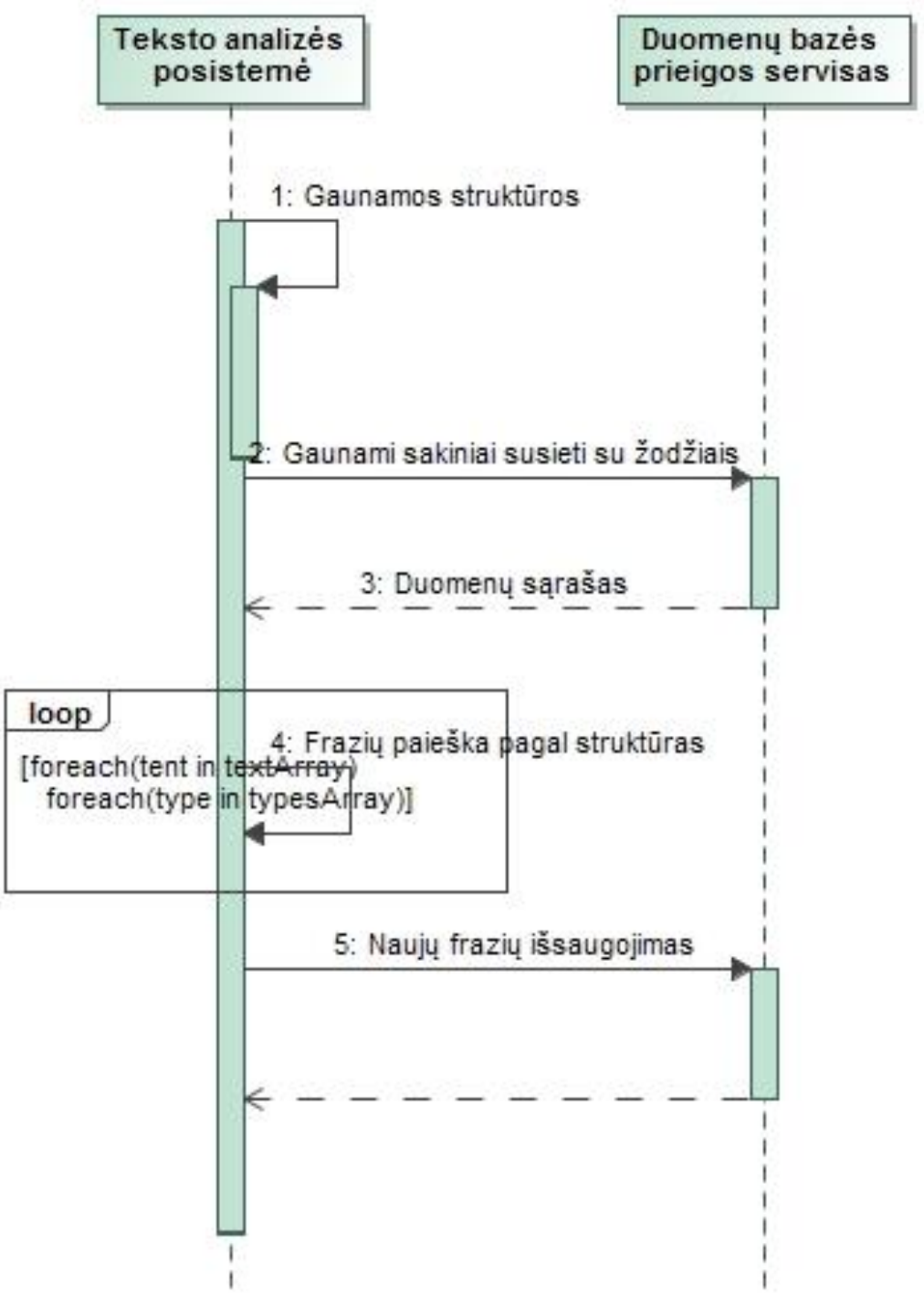
Panaudos atvejo „*Ieškoti frazių tarp kabučių*“ sekų diagrama pateikiama (25 Pav. P.A. „Ieškoti frazių tarp kabučių“ sekų diagrama) paveiksle.



25 Pav. P.A. „Ieškoti frazių tarp kabučių“ sekų diagrama

Pirmiausia yra gaunamas, tekstas susietas su sakiniais, toliau ieškome frazių tarp kabučių. Pabaigus ieškoti įrašomos tarp kabučių rastos frazės į duomenų bazę.

Detaliau panagrinėta frazių paieškos sekų diagrama pateikta (26 Pav. „Frazių paieška pagal šablonus“ sekų diagrama) paveiksle. Joje pavaizduota pagrindinė frazių paieška pagal šablonus seka, tarp „*Teksto analizės posistemės*“ ir „*Duomenų bazės prieigos serviso*“.

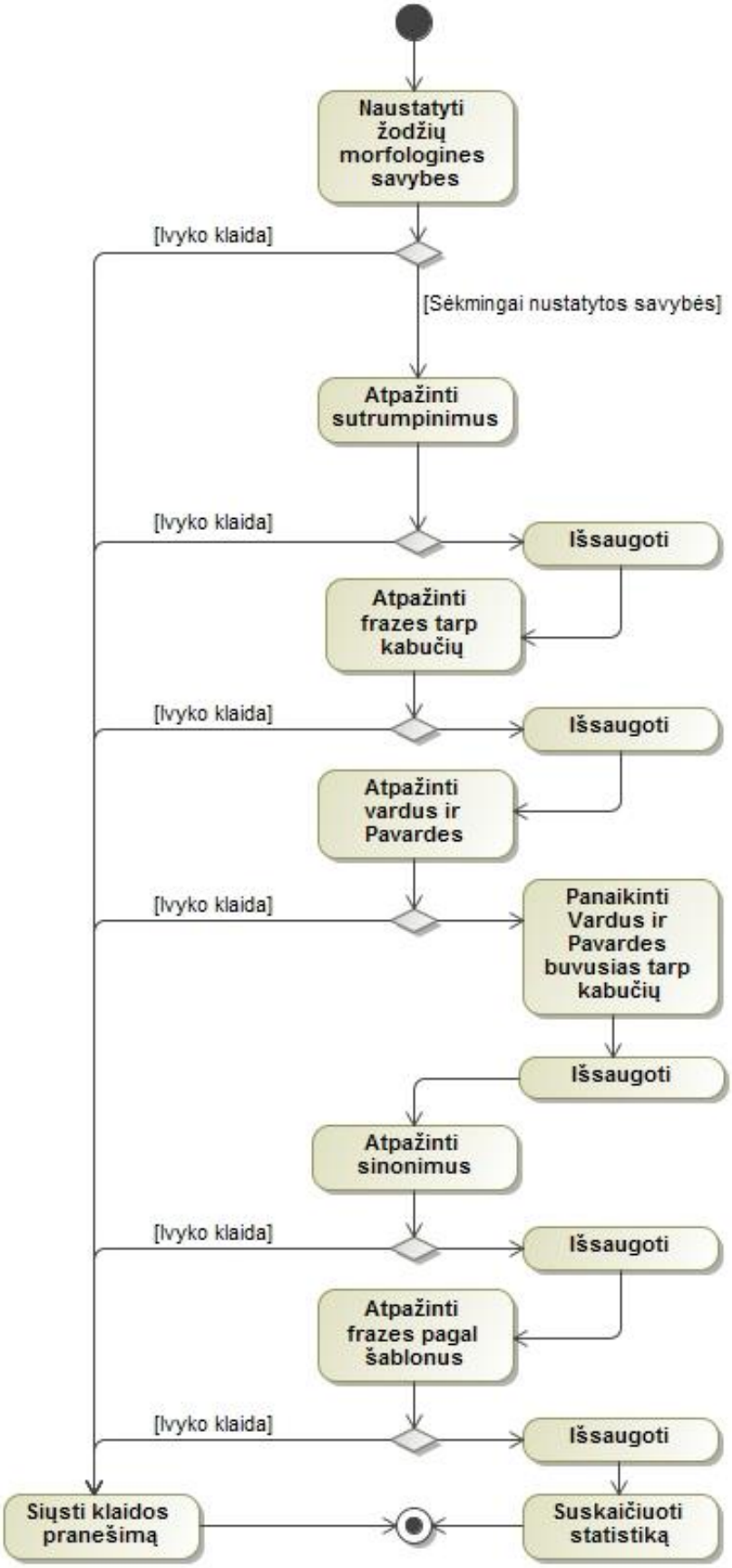


26 Pav. „Frazių paieška pagal šablonus“ sekų diagrama

Frazių paieškos pagal šablonus funkcija pirmiausia iš frazių struktūrų duomenų failo pasiima visas rastas struktūras. Vėliau einama prie sekančio žingsnio „Naujų struktūrų sudarymas“ kuriame neišbaigtos frazių struktūros yra papildomomis, kad užpildytu visą pateiktos struktūros aibę. Vėliau gaunamos iš duomenų bazės duomenys t.y. sakiniai susieti su žodžiais, kurie vėliau yra analizuojami. Prieš paskutinis žingsnis frazių ieškojime tai pagal pateiktus šablonus, tekste ieškoti frazių atitikmenų. Ciklas vyksta tol kol ne iki galo yra išanalizuotas tekstas. Vėliau visos rastos frazės išsaugomos duomenų bazėje.

2.1.2.3. Teksto analizės posistemės veiklos diagramos

Žemiau esančioje veiklos diagramoje yra vaizduojama frazių radimas:

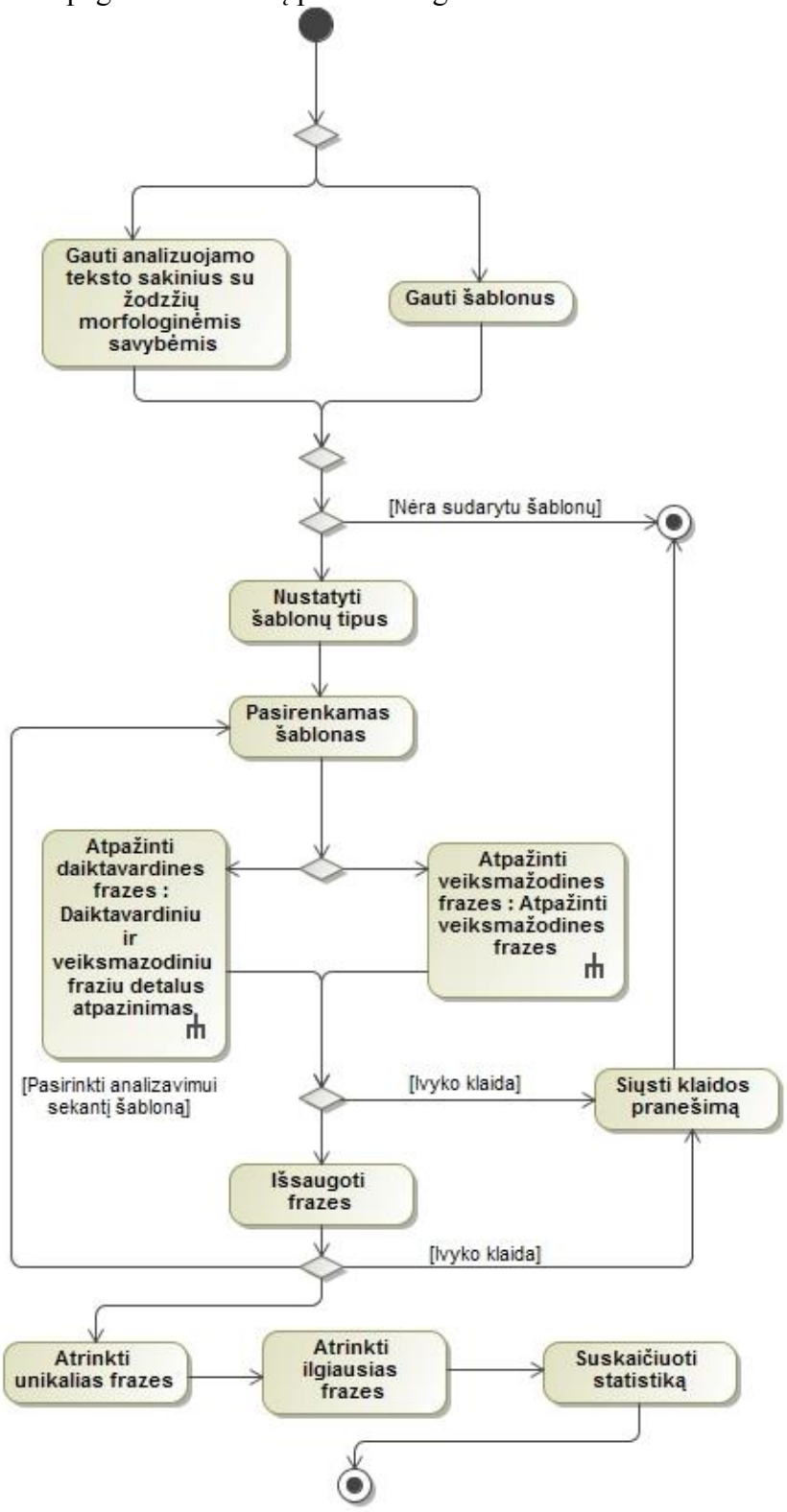


27 Pav. „Frazių radimo algoritmo“ veiklos diagrama

3. AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU SPRENDIMO ALGORITMAS

3.1. Daiktavardinių bei veiksmažodinių frazių paieškos veikimo aprašymas

Žemiau pateiktame (28 Pav. „Frazių radimo naudojant šablonus algoritmas“ veiklos diagrama) paveiksle, kuriame matome pagrindines frazių paieškos algoritmo veiksmai.



28 Pav. „Frazių radimo naudojant šablonus algoritmas“ veiklos diagrama

Veiklos diagramoje (28 Pav. „Frazių radimo naudojant šablonus algoritmas“ veiklos diagrama) pavaizduota esminė frazių paieškos dalis, kuri yra susieta su šablonų gavimu bei jau rezultatų išsaugojimu. Algoritmas prieš pradėdamas teksto analizę pirmiausias susirenka vartotojo sudarytus bei nurodytus šablonus ir taipogi gauna analizuojamo teksto žodžius su jų morfologinėmis savybėmis. Jei vartotojas nebuvo sudaręs jokių šablonų tai algoritmas nieko neieško ir baigia paieška. Jei šablonų buvo tai sistema automatiškai nustato jų ilgį pagal kurį po to atliekama analizė tam tikrame algoritme. Kai sistema pasiima pirmą šabloną iš vartotojo sudarytų šablonų ir nustato šablono tipą, t.y. „*daiktavardinių frazių*“ šablonas ar „*veiksmažodinių frazių*“ šablonas, automatiškai pereina į atitinkamą veiksmą. Daiktavardinių frazių paieškos veiksmas padetalizuotas pateiktas (**Error! Reference source not found.**) veikos diagramoje, veiksmažodinių frazių paieškos veiklos diagrama pateikta (29 Pav. Veiksmažodinių frazių atpažinimo veiklos diagrama) paveiksle. Jei sistemoje neįvyko klaidų beiškant frazių, tai sistema išsaugo visas rastas frazes ir pereina prie kito šablono. Kai sistema išanalizavo tekstą su visais frazių radimo šablonais, tai pereina prie rastų frazių apdorojimo veiksmo. Frazių papildomas apdorojimas susideda iš dviejų pagrindinių veiksmų:

- „*Atrinkti unikalias frazes*“ – šiame veiksme programa automatiškai išrenka tik unikalias frazes.
- „*Atrinkti ilgiausias frazes*“ – šiame veiksme yra išrenkamos ilgiausios frazės iš visų pasikartojančių ar įeinančių į ta frazę elementų, pvz.:
  Frazių paieškos algoritmas rado tokias frazes: *namas, didelis, raudonas namas, didelis raudonas namas*. Algoritmas automatiškai išrenka tik ilgiausias tinkančias frazes, šiuo atveju liktų tik „*didelis raudonas namas*“ frazė.

Papildomas veiksmas po frazių apdorojimo vyksta „*Suskaičiuoti statistiką*“. Šiame veiksme programa suskaičiuoja kiek iš viso buvo rasta šių frazių tekste.

3.1.1. Šablonų sudedamosios dalys

Šabloną, kaip ir minėta ankščiau privalo sudaryti būtent keturios dalys. Pagrindinė ir esminė šablonų sudedamoji dalis yra patys šablono elementai. Šabloną gali sudaryti 5 šablono elementai, kurie kiekvienas iš jų atskirai gali turėti savo dedamąsias dalis.

Žemiau pateiktame paveikslėlyje matyti pats duomenų modelis. Kaip atrodo, tiek visas ieškomos frazės struktūros šablonas, tiek jį sudarantys frazės struktūros šablono elementai:

Frazės struktūros šabloną gali sudaryti nuo vieno iki penkių frazės struktūros šablono elementų. Kiekvieną šablono elementą sudaro 10 dedamųjų:

- Kalbos dalis – žodžio kalbos kalbos dalis pvz.: *daiktavardis, būdvardis* ir t.t.
- Giminė – žodžio giminė pvz.: *moteriška, vyriška*.
- Linksnis – žodžio linksnis
- Laipsnis – žodžio laipsnis
- Asmuo – žodžio asmuo
- Laikas – žodžio laikas
- Pasikartojamumas – nusako ar šis šablono elementas gali kartotis, tol kol kitas elementas neatitinka jo struktūros.
- Privalomumas –nusako ar šis elementas yra privalomas šablone.
- Pažymimasis daiktavardis – nurodo ar šiame elemente esantis daiktavardis yra pažymimasis.
- Pažymimojo daiktavardžio dalis – nurodo kokie kiti šablono elementai yra priklausomi prie pažymimojo daiktavardžio.

Žemiau pateiktoje (9 Lentelė Kalbos dalių morfologinės savybės) lentelėje yra pateiktos lietuvių kalbos dalių, morfologinės kaitymo savybės. Kokioms kalbos dalims būdingas linksnis, gimine ir t.t.

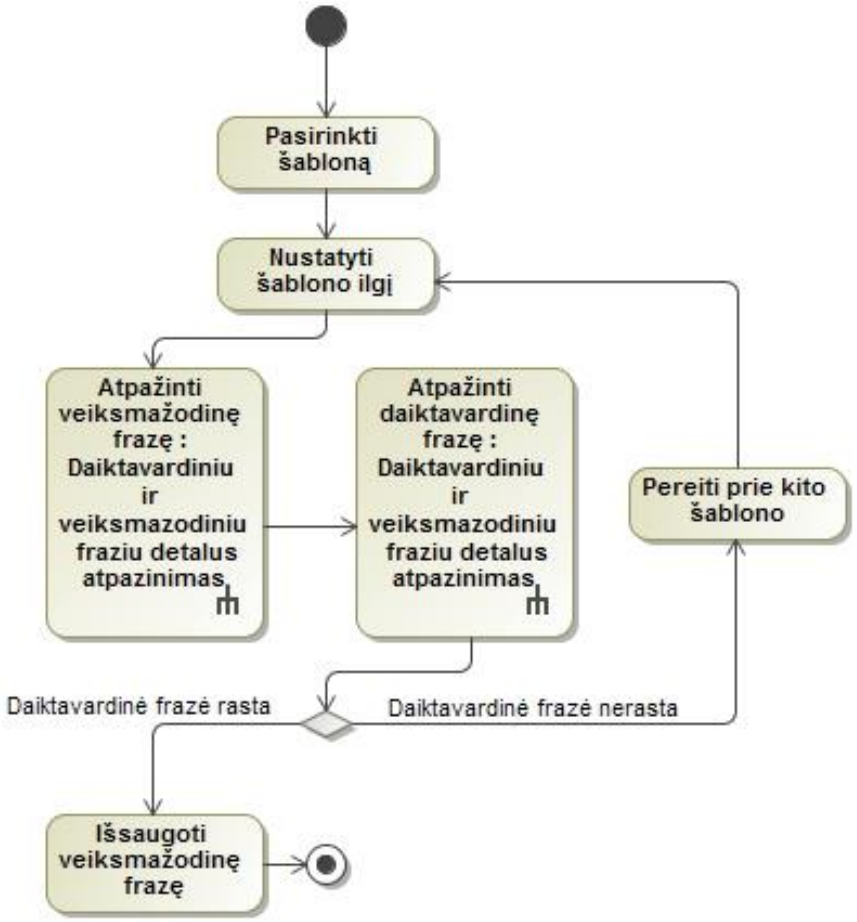
9 Lentelė Kalbos dalių morfologinės savybės

Kalbos dalis	*Morfologinės savybės*
Daiktavardis	Skaičiais, linksniais, giminėmis
Būdvardis	Skaičiais, linksniais, giminėmis
Veiksmažodis	Asmenimis, skaičiais, laikais, nuosakomis
Prieveiksmis	----------
Skaitvardis	Skaičiais, linksniais, giminėmis
Įvardis	Giminėmis, linksniais, skaičiais
Prielinksnis	----------
Jungtukas	----------
Jaustukas	----------
Ištiktukas	----------
Dalelytė	----------

3.1.2. Frazių paieška naudojant šablonus

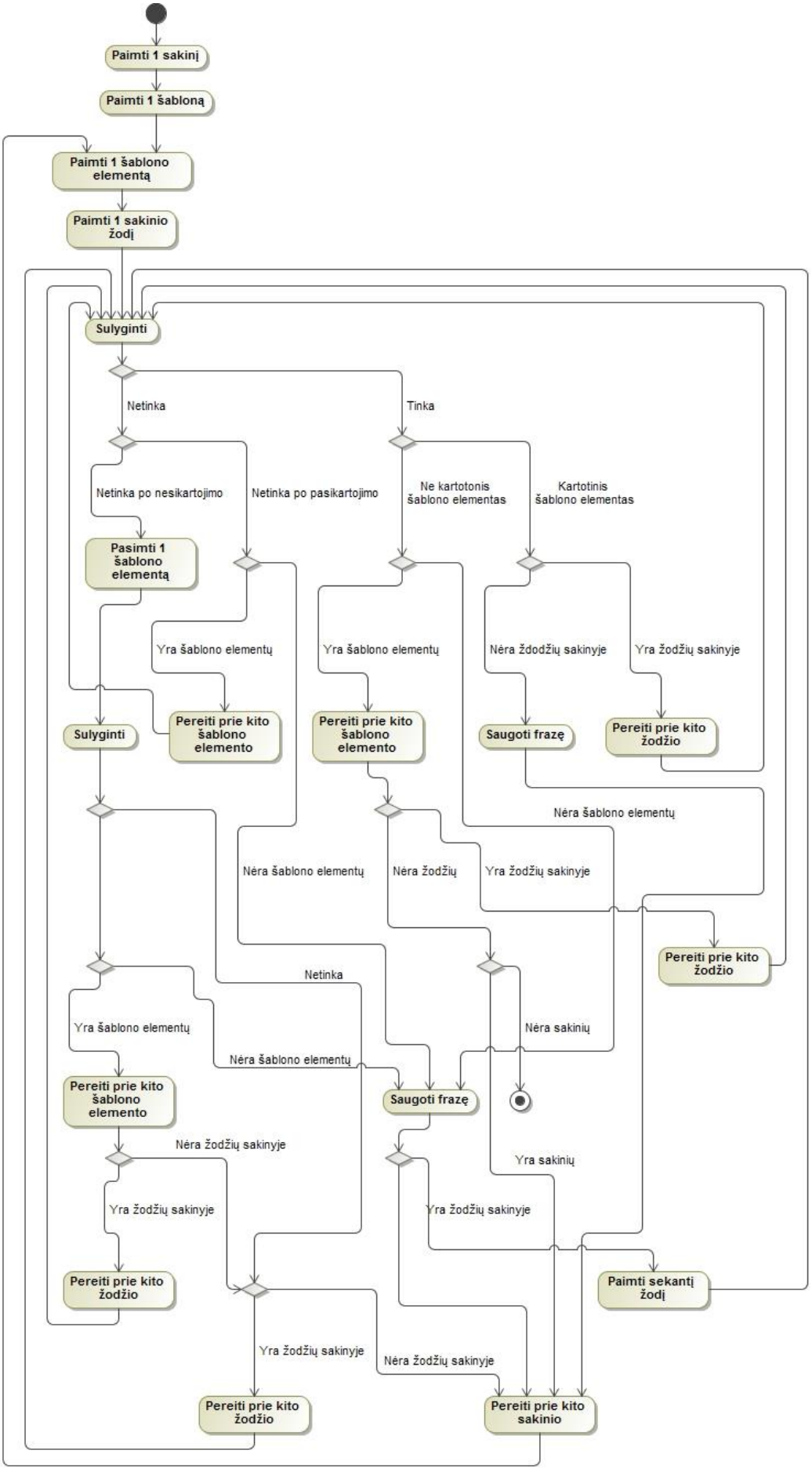
Kaip ir kiekvieni tekste esantys žodžiai taip ir šablono elementai turi tuos pačius elementus t.y. *kalbos dalį*, *giminę*, *skaičių* ir t.t. Teksto analizės metu frazės ieškomos nagrinėjant po vieną sakinį tikrinant ar esantys žodžiai su savo kalbos dalimis atitinka frazės struktūros šablono elementus. Jei visi esantys frazės struktūros elementai atitinka tekste esančių žodžių kalbos dalims ir eilės tvarka, tai šis žodžių junginys rastas tekste traktuojamas kaip frazė.

Veiklos diagramos „*Frazių radimo naudojant šablonus algoritmas*“ kuris yra pateiktas (28 Pav. „Frazių radimo naudojant šablonus algoritmas“ veiklos diagrama) paveiksle veiksmo „*Atpažinti veiksmažodines frazes*“ detalesnė veiklos diagrama pateikta žemiau (28 Pav. „Frazių radimo naudojant šablonus algoritmas“ veiklos diagrama) paveiksle.



29 Pav. Veiksmažodinių frazių atpažinimo veiklos diagrama

Žemiau pateiktame (30 Pav. Detali daiktvardinių veiksmažodinių frazių paieškos algoritmo veiklos diagrama) paveiksle pavaizduota frazių paieškos veiklos diagrama. Ši veiklos diagrama detalizuoja pagrindinį frazių paieškos veiklos algoritmą, pagal kurį sistema ieško frazių naudodama šablonus. Šis algoritmas yra naudojamas tiek daiktavardiniams šablonams, tiek veiksmažodiniams.



30 Pav. Detali daiktvardinių veiksmažodinių frazių paieškos algoritmo veiklos diagrama

3.1.3. Algoritmo taikymo prielaidos ir situacijos

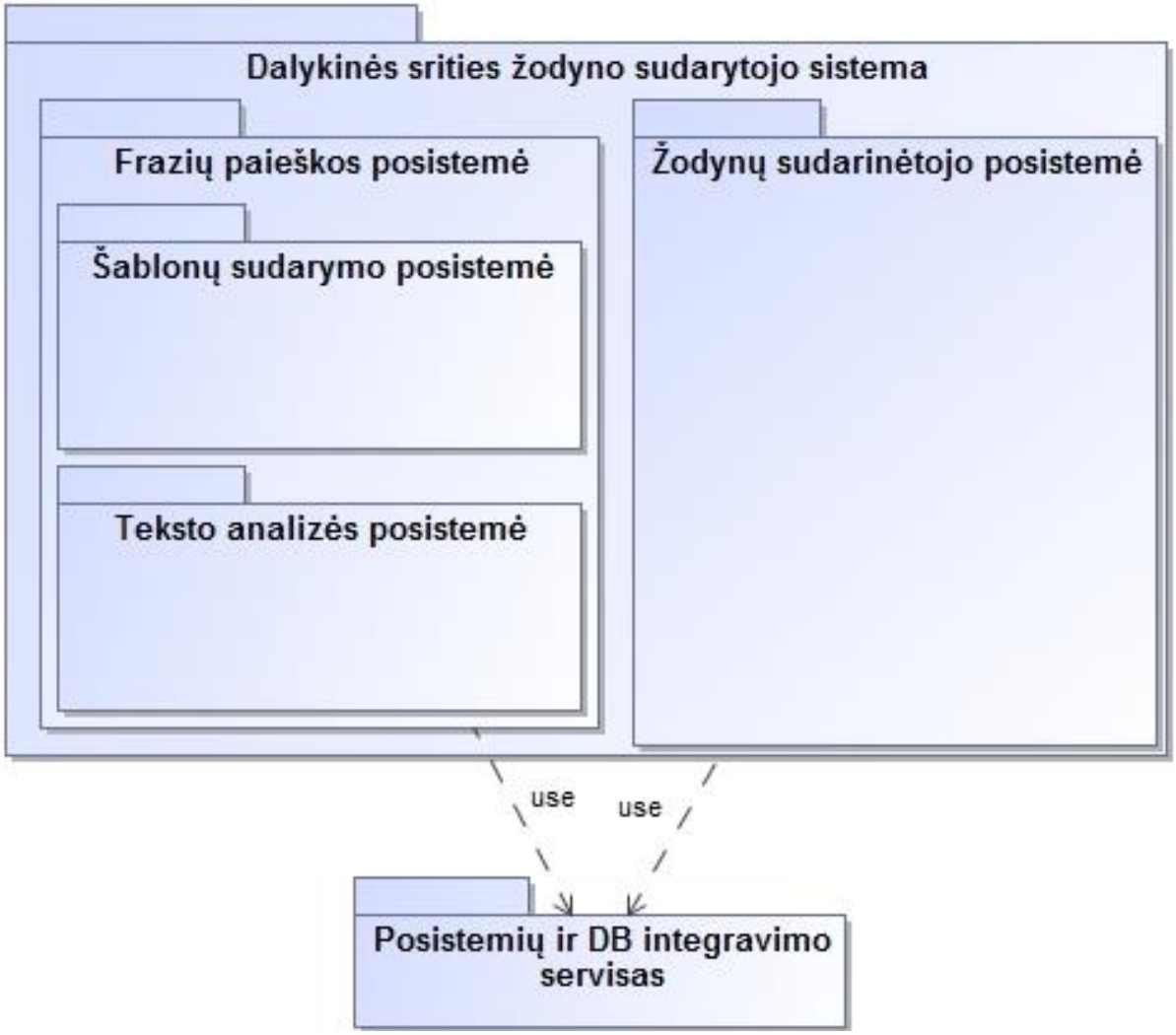
Šioje dalyje pateikiamos pagrindinės prielaidos, kurios gali lemti geresnių bei tinkamesnių frazių radimui. Taipogi pateikimas situacijos dėl kurių vartotojas gali būti apribotas.

- Vartotojas gali nenorėti vesti visų įmanomų frazių formų į sistemą, dėl to galimybė pasirinkti *pasikartojamumą* ir *privalomumą* frazės struktūros elementuose, leis jam laisviau ir lengviau sudaryti šablonus.
- Norint rasti kuo tinkamesnes frazes, būtina gerai ir kruopščiai sudaryti frazės struktūrų šablonus.
- Sudarinėjant šablonus ir juos skirstant į grupes bus galima rasti specifines, vartotojo norimas frazes, pagal jo užduotus frazių struktūrų šablonus.
- Algoritmas yra pritaikytas ieškoti penkių ilgio frazių, kadangi *praplėčiamumas* ir *pasikartojamumas* leidžia praplėsti paiešką nuo penkių iki dvidešimties žodžių frazėje.

4. AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU REALIZACIJOS PROJEKTAS

4.1. Sistemos architektūra

Šioje dalyje pateikimas pagrindinis sistemos architektūrinis sprendimas. Žemiau esančiame (31 Pav. Dalykinės srities žodyno sudarytojo sistemos architektūrinis sprendimas) paveiksle pateikimas sistemos architektūrinis sprendimas tik iš pagrindinių komponentų.

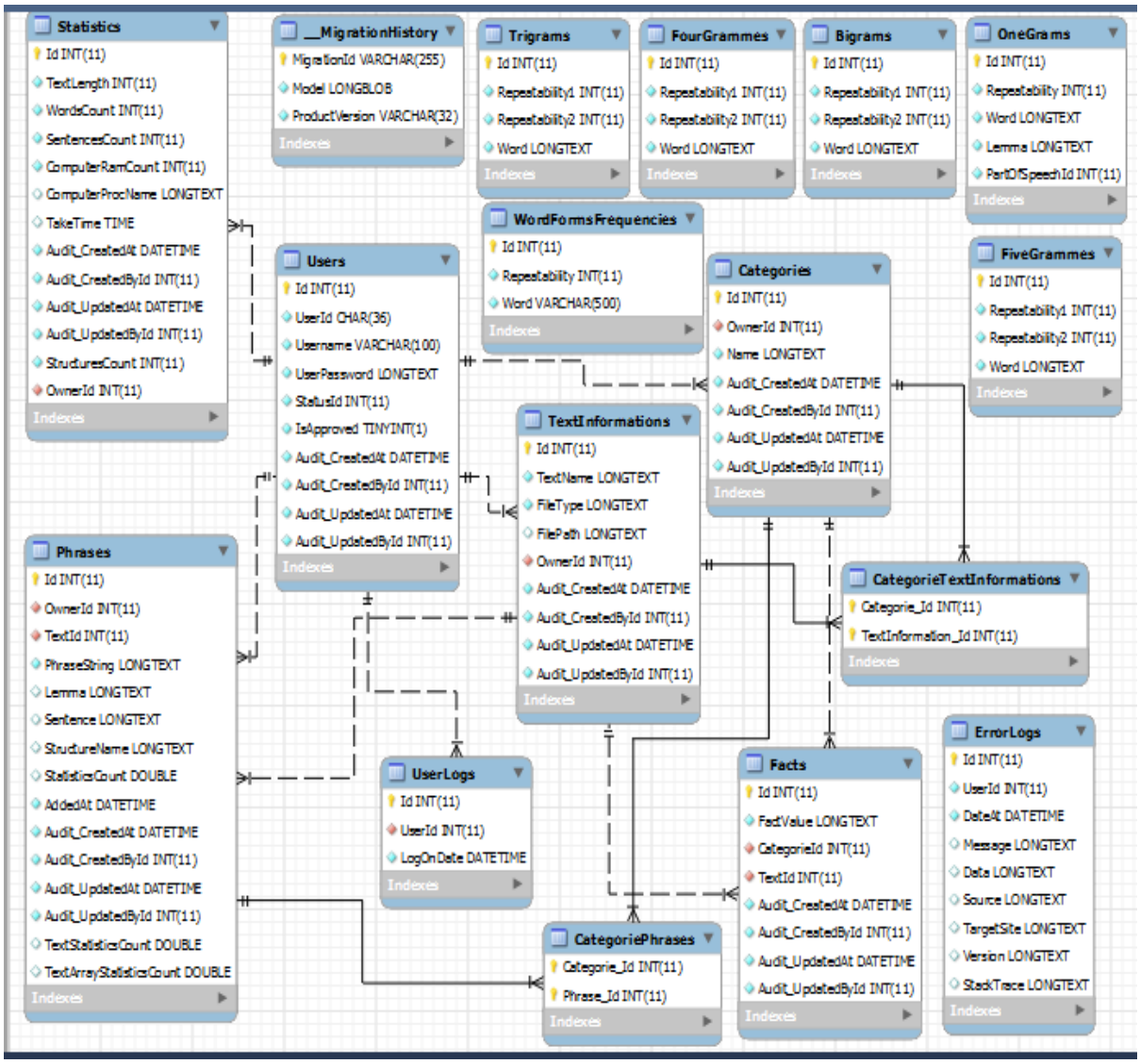


31 Pav. Dalykinės srities žodyno sudarytojo sistemos architektūrinis sprendimas

Dalykinės srities žodyno sudarytojo sistemą sudaro du pagrindiniai komponentai (posistemės), t.y. „*Frazių paieškos posistemė*“ ir „*Žodynų sudarinėtojo posistemė*“. Frazių paieškos posistemėje yra dar dvi papildomos posistemės kurios yra atsakingos už frazių šablonų sudarymą bei teksto analizės dalį. Šablonų sudarymo posistemės detalesnis vaizdas pateiktas (14 Pav. Šablonų sudarinėjimo posistemės panaudojimo atvejų diagrama) paveiksle, taipogi teksto analizės posistemės detalesnis vaizdas pateikiamas (20 Pav. Teksto analizės posistemės panaudojimo atvejų diagrama) paveiksle.

4.1.1. Duomenų bazės schema

Šioje duomenų bazėje yra saugoma visos vartotojų sudarytos frazės, vartotojo kategorijos, grupės. Duomenų bazėje taipogi saugoma visa informacija apie vartotojus, kaupiama jų statistika. Taipogi šioje duomenų bazėje yra sukaupta statistika iš kitų tekstų, kuri yra naudojama vartotojo rastų frazių statistikai skaičiuoti.

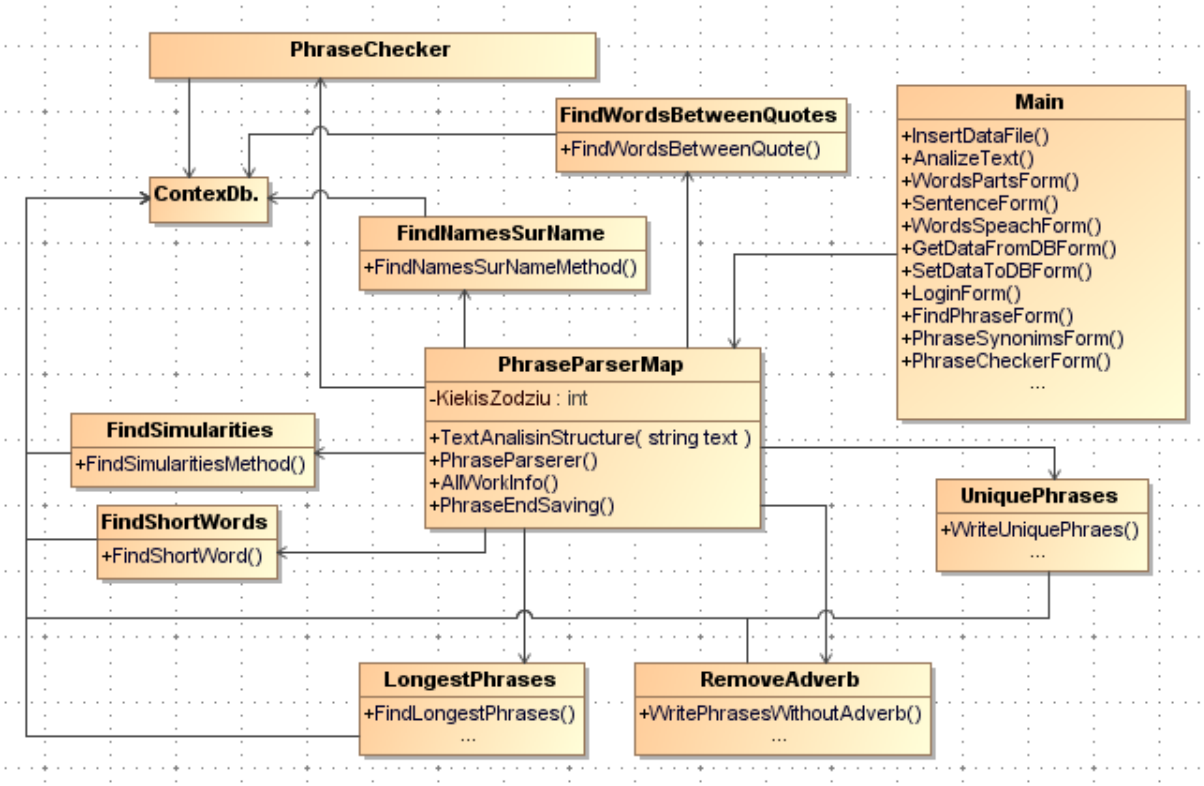


32 Pav. Pagrindinės duomenų bazės schema

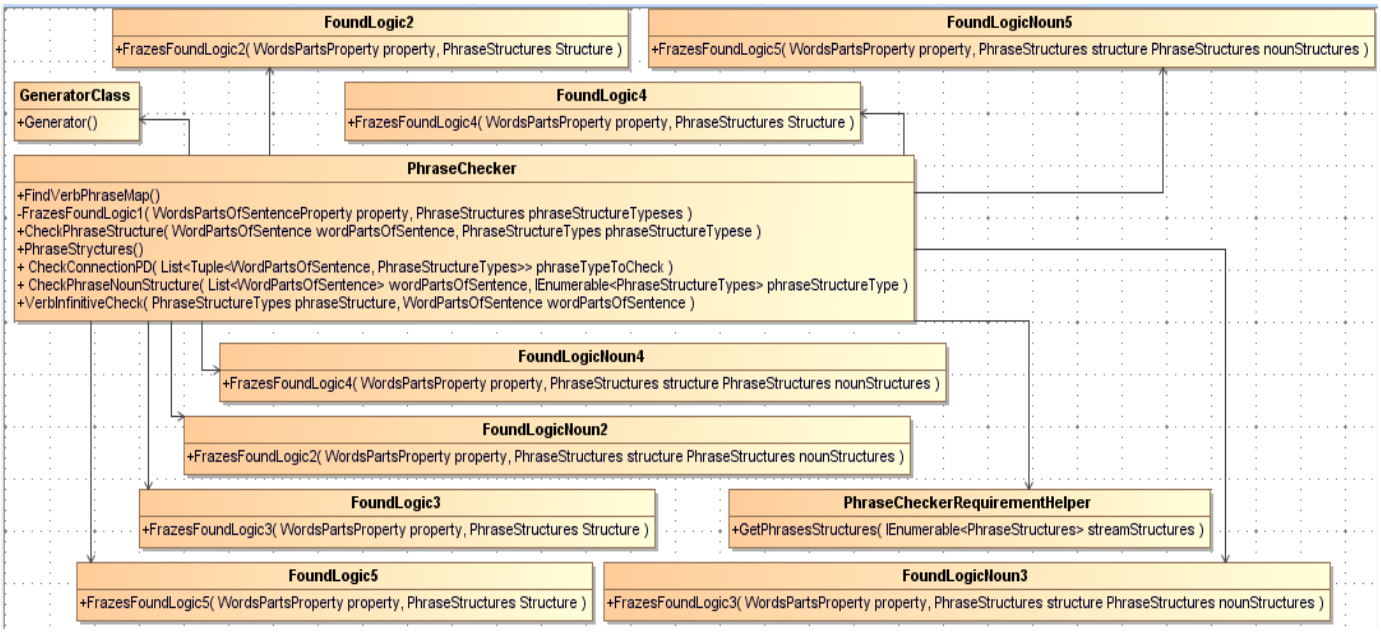
10 Lentelė Pagrindinės duomenų bazės schemos lentelių aprašai

Lentelės pavadinimas	Aprašymas
Users	Saugoma informacija apie vartotojus, kurie gali prisijungti prie serviso
PhraserTable	Saugomos visos frazės, kurias atsiuntė vartotojai
UsersLogs	Saugoma informacija apie vartotoju prisijungimus
Categories	Saugomos informacijos apie kategorijas
Statistics	Saugoma visa statistika
Facts	Saugomi vartotojo sudaryti faktai
ErrorLogs	Saugomi gauti klaidos pranešimai

4.2. Frazių paieškos posistemės detalus projektas



33 Pav. Frazių radimo klasių daigrama



34 Pav. Frazių radimas naudojant šablonus klasių diagrama

Klasifikacija: Paketas.
Apibrėžimas: Pakete pateikiamos klasės kurios atsakingos už frazių radimą.
Atsakomybės: Paketas atsako už frazių radimą.
Struktūra: Komponentą sudaro klasės aprašytos pakete „Teksto analizės posistemė“ ir pateiktos diagramoje.
Sąveikavimas: Frazių radimo paketas naudojasi klaidos paketu.
Resursai: Paketas naudoja lokalia DB.
Skaičiavimai: Skaičiavimai detalizuojami paketo klasių metodų aprašymuose.
Sąsaja/eksportas: Public metodai.

Klasė „FindWordsBetweenQuotes“

Apibrėžimas: Suranda frazes, kurios yra tarp kabučių.
Atsakomybės: Surasti frazes, kurios parašytos tarp kabučių.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *FindWordsBetweenQuotes* naudojasi *contexDB* metodais, o *FindWordsBetweenQuotes* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindWordsBetweenQuote*

Klasės metodai

11 Lentelė FindWordsBetweenQuote metodas.

Public FindWordsBetweenQuote	
Atsakomybės	Surasti frazes, kurios yra parašytos tarp kabučių.
Skaičiavimai	Nagrinėja tekstą sakiniais ir ieško frazių tarp kabučių.
Sąsaja/eksportas	-
Išimtys	Klaidos atveju įrašas nėra įrašomas ir tęsiama paieška

Klasė „ FindNamesSurName “

Apibrėžimas: Suranda vardus ir pavardes.
Atsakomybės: Surasti vardus ir pavardes iš teksto.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *FindNamesSurName* naudojasi *contexDB* metodais, o *FindNamesSurName* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindNamesSurNameMethod*

Klasės metodai

12 Lentelė FindNamesSurNameMethod metodas.

Public FindNamesSurNameMethod	
Atsakomybės	Suranda vardus ir pavardes iš teksto.
Skaičiavimai	Nagrinėdami tekstą, pagal sakinius, ieškome žodžių kurie eina greta ir yra iš didžiųjų raidžių.
Sąsaja/eksportas	*DataBaseContext*
Išimtys	Jai paskutinis žodis arba sakinys iš 1 žodžio.

Klasė „ FindShortWords“

Apibrėžimas: Suranda sutrumpinimus.
Atsakomybės: Surasti sutrumpinimus iš teksto(pvz. KTU).
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *FindShortWords* naudojasi *contexDB* metodais, o *FindShortWords* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindShortWord*

Klasės metodai

13 Lentelė FindShortWord metodas.

Public FindShortWord	
Atsakomybės	Surasti sutrumpinimu, tokius kaip KTU
Skaičiavimai	Ieškomi žodžiai kurie yra sudaryti vien iš didžiųjų raidžių.
Sąsaja/eksportas	Rasti sutrumpinimai įrašomi į duomenų bazę naudojant *WhritePhrase* metodą.
Išimtys	-

Klasė „FindSimularities“

Apibrėžimas: Suranda sinoniminius žodžius.
Atsakomybės: Suranda sinoniminius žodžius.
Struktūra: Klasės struktūra pateikta klasių diagramoje.

Sąveikavimas: *FindSimularities* naudojasi *contexDB* metodais, o *FindSimularities* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindSimularitiesMethod*

Klasės metodai

14 Lentelė FindSimularitiesMethod metodas.

Public FindSimularitiesMethod	
Atsakomybės	Suranda sinoniminius žodžius tarp sutrumpintų frazių ir teksto.
Skaičiavimai	Ieško atitikmenų tekste su sutrumpintomis frazėmis.
Sąsaja/eksportas	Duomenų įrašymui į lokalią duomenų bazę *contexDB*
Išimtys	-

Klasė „LongestPhrases“

Apibrėžimas: Suranda ilgiausias frazes.
Atsakomybės: Palikti ilgiausias frazes ir pašalinti trumpesnes, kurios įeina į ilgesne frazes.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *LongestPhrases* naudojasi *contexDB* metodais, o *LongestPhrases* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindLongestPhrases*

Klasės metodai

15 Lentelė FindLongestPhrases metodas.

Public FindLongestPhrases	
Atsakomybės	Ilgiausių frazių radimas
Skaičiavimai	Randamos ilgiausios rastos frazės sakinyje.
Sąsaja/eksportas	Paimamos rastos frazės iš lokalios duomenų bazės. Įrašomos tik ilgiausios rastos frazės į *contexDB.*
Išimtys	-

Klasė „ RemoveAdverb“

Apibrėžimas: Pašalina frazių priekyje einančius jungtukus ,prielinksnius.
Atsakomybės: Pašalina frazių priekyje einančius jungtukus, prielinksnius.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *RemoveAdverb* naudojasi *contexDB* metodais, o *RemoveAdverb* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *WritePhrasesWithoutAdverb*

Klasės metodai

16 Lentelė WritePhrasesWithoutAdverb metodas.

Public WritePhrasesWithoutAdverb	
Atsakomybės	Pašalinti frazių priekyje einančius jungtukus, prieveiksmius.
Skaičiavimai	Iš visų rastų frazių pašalinami pirmi žodžiai kurie yra arba jungtukai, arba prieeiksmiai.
Sąsaja/eksportas	-
Išimtys	-

Klasė „ UniquePhrases “

Apibrėžimas: Suranda iš visų rastų frazių tik unikalias frazes.
Atsakomybės: Suranda iš visų rastų frazių tik unikalias.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *UniquePhrases* naudojasi *contexDB* metodais, o *UniquePhrases* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *WriteUniquePhraes*

Klasės metodai

17 Lentelė WriteUniquePhraes metodas.

Public WriteUniquePhraes	
Atsakomybės	Surasti iš visų surastų frazių tik unikalias frazes
Skaičiavimai	Gaunamos visos frazes, einama po vieną fraze ir ieškoma ar tokia jau yra sąraše, jai yra nerašome jai nėra įrašoma i nauja unikaliu frazių sąrašą.
Sąsaja/eksportas	-
Išimtys	-

Klasė „ PhraseChecker“

Apibrėžimas: Ieško daiktavardinių ir veiksmažodinių frazių pagal šablonus.
Atsakomybės: Ieško daiktavardinių ir veiksmažodinių frazių pagal šablonus
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *PhraseChecker* naudojasi *contexDB, FoundLogic2, FoundLogic3, FoundLogic4, FoundLogic5, GeneratorClass, FoundLogicNoun2, FoundLogicNoun3, FoundLogicNoun4, FoundLogicNoun5, PhraseCheckerRequirementHelper* metodais, o *PhraseChecker* metodais naudojasi *PhraseParserMap*.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FindVerbPhraseMap, FrazesFoundLogic1, CheckPhraseStructure, CheckPhraseStructure*

Klasės metodai

18 Lentelė FindVerbPhraseMap metodas.

Public FindVerbPhraseMap	
Atsakomybės	Ieško daiktavardiniu ir veiksmažodinių frazių.
Skaičiavimai	Gaunami pateikti šablonai taipogi sakiniai su tam sakiniui priklausančiais žodžiais ir jų kalbos dalimis. Tikriname sakiniais tekstą naudojant gautus šablonus (daiktavardinį ar veiksmažodinį šabloną)
Sąsaja/eksportas	-
Išimtys	Nėra sudaryta šablonų.

19 Lentelė FrazesFoundLogic1(WordsPartsOfSentenceProperty property, PhraseStructures phraseStructureTypeses) metodas.

Public FrazesFoundLogic1(WordsPartsOfSentenceProperty property, PhraseStructures phraseStructureTypeses)	
Atsakomybės	Suranda daiktavardines frazes iš vieno šablono elemento
Skaičiavimai	Ieškoma sakinyje pagal einančių žodžių kalbos dalis aitkmenų šablone
Sąsaja/eksportas	Įrašomos rastos frazės į sąrašą.
Išimtys	Jai nėra sudarytų šablonų, skaičiavimai nevyksta.

20 Lentelė CheckPhraseStructure(WordPartsOfSentence wordPartsOfSentence, PhraseStructureTypes phraseStructureTypese) metodas.

Public CheckPhraseStructure(WordPartsOfSentence wordPartsOfSentence, PhraseStructureTypes phraseStructureTypese)	
Atsakomybės	Patikrina ar žodis atitinka frazes šablono elementą.
Skaičiavimai	Patikrina gauto žodžio struktūra su frazių šablono struktūra.
Sąsaja/eksportas	Grąžina *true* – jei elementas tinkamas, arba *false* – jai elementas netinkamas.
Išimtys	-

21 Lentelė PhraseStryctures metodas.

Public PhraseStructures	
Atsakomybės	Gauna visas struktūras, kurias sudarė vartotojas, jai vartotojas nesudarė fazių struktūrų, tai nauji šablonai sugeneruojami automatiškai sistemos.
Skaičiavimai	-
Sąsaja/eksportas	Gražina *PhraseStructures* sąrašą.
Išimtys	-

22 Lentelė CheckConnectionPD(List<Tuple<WordPartsOfSentence, PhraseStructureTypes>> phraseTypeToCheck) metodas.

Public CheckConnectionPD(List<Tuple<WordPartsOfSentence, PhraseStructureTypes>> phraseTypeToCheck)	
Atsakomybės	Patikrina ar daiktavardis yra pažymimasis.
Skaičiavimai	Patikrina atėjusioje struktūroje ar yra pažymimasis daiktavardis bei prie PD yra elementai kurie turi jį atitikti. Tikrinama ar žodžių masyvas atitinka pažymimojo daiktavardžio ir jį atitinkančių elementų kombinaciją. Jai *true* – rastas pažymimasis daiktavardis, jei *false* – ieškoma toliau pagal sekančius žodžius, tikrinant pagal PD.
Sąsaja/eksportas	Grąžina *true* arba *false*
Išimtys	Jai šablone nėra pažymimojo daiktavardžio, šis metodas nevykdomas.

23 Lentelė VerbInfinitiveCheck(PhraseStructureTypes phraseStructure, WordPartsOfSentence wordPartsOfSentence)metodas.

Public VerbInfinitiveCheck(PhraseStructureTypes phraseStructure, WordPartsOfSentence wordPartsOfSentence)	
Atsakomybės	Patikrina ar veiksmažodis yra bendratis.
Skaičiavimai	-
Sąsaja/eksportas	Grąžina *true* arba *false, DataBaseContext*
Išimtys	Jai ne bendratis, tačiau kalbos dalis yra *Veiksmažodis,* tai iš duomenų bazės randama bendratis atitinkantį veiksmažodžio žodį.

Klasė „ FoundLogic2“

Apibrėžimas: Ieško daiktavardinių frazių, kurie sudaryti iš dviejų elementų.
Atsakomybės: Surasti daiktavardines frazes sudarytas iš dviejų elementų.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *PhraseChecker* naudojasi *FoundLogic2* metodais.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FrazesFoundLogic2*

Klasės metodai

24 Lentelė FrazesFoundLogic2metodas.

Public FrazesFoundLogic2(WordsPartsProperty property, PhraseStructures Structure)	
Atsakomybės	Ieško daiktavardinių frazių kurios atitinka dviejų elementų sudarytus šablonus.
Skaičiavimai	Gauna sakinio žodžius su jų kalbos dalimis, vėliau tikrinama ar šios kalbos dalys atitinka pateiktą daiktavardinės frazės šablono struktūrą.
Sąsaja/eksportas	Grąžina rastų frazių sąrašą.
Išimtys	-

Klasės *FrazesFoundLogic3, FrazesFoundLogic4, FrazesFoundLogic5* logika bei metodai atitinka *FrazesFoundLogic2* logika, tik ieškoma pagal atitinkamai ilgesnius šablonus.

Klasė „ GeneratorClass “

Apibrėžimas: Sugeneruoja standartinius frazių radimo šablonus.
Atsakomybės: Sugeneruoja standartinius frazių radimo šablonus.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *PhraseChecker* naudojasi *FoundLogic2* metodais.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: Generator

Klasės metodai

25 Lentelė Generator metodas.

Public Generator	
Atsakomybės	Sugeneruoti standartinių šablonų sąrašą.
Skaičiavimai	Pagal tam tikras taisykles sugeneruoja frazių šablonus.
Sąsaja/eksportas	Gražina frazių šablonus.
Išimtys	-

Klasė „PhraseCheckerRequirementHelper“

Apibrėžimas: Generuoti naujas frazių struktūras pagal vartotojo šablonus.
Atsakomybės: Generuoti naujas frazių struktūras pagal vartotojo šablonus.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *PhraseChecker* naudojasi *PhraseCheckerRequirementHelper* metodais.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *GetPhrasesStructures*

Klasės metodai

26 Lentelė GetPhrasesStructures metodas.

Public GetPhrasesStructures(IEnumerable<PhraseStructures> streamStructures)	
Atsakomybės	Sugeneruoti naujus frazių šablonų struktūras.
Skaičiavimai	Iš pravalomų ir neprivalomų kalbos dalių elementų sudaromos naujos visos įmanomos frazių struktūros.
Sąsaja/eksportas	Grąžinama visos galimos frazių struktūros.
Išimtys	-

Klasė „ FoundLogicNoun2 “

Apibrėžimas: Ieško visų veiksmažodinių frazių sudarytų iš dviejų elementų.
Atsakomybės: Surasti visas veiksmažodines frazes sudarytas iš dviejų elementų.
Struktūra: Klasės struktūra pateikta klasių diagramoje.
Sąveikavimas: *PhraseChecker* naudojasi *FoundLogicNoun2* metodais.
Resursai: Klasė nenaudoja jokių papildomų resursų.
Skaičiavimai: Skaičiavimai pateikti klasės metodų aprašymuose
Sąsaja/eksportas: Metodai: *FrazesFoundLogic2*

Klasės metodai

27 Lentelė FoundLogicNoun2 metodas.

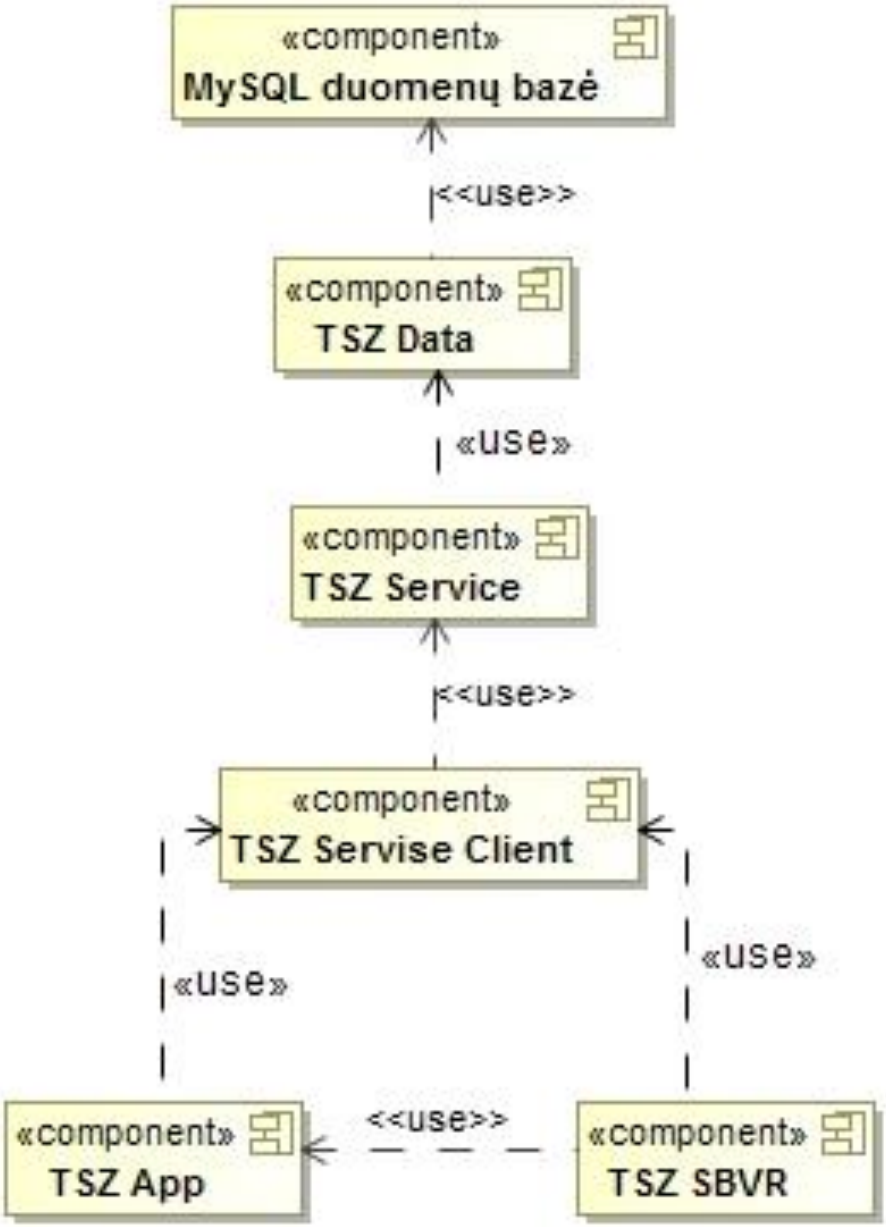
Public FoundLogicNoun2(WordsPartsProperty property, PhraseStructures structure PhraseStructures nounStructures)	
Atsakomybės	Ieško veiksmažodinių frazių kurios atitinka dviejų elementų sudarytus šablonus.
Skaičiavimai	Gauna sakinio žodžius su jų kalbos dalimis, vėliau tikrinama ar šios kalbos dalys atitinka pateiktą veiksmažodinės frazės šablono struktūrą.
Sąsaja/eksportas	Rastų frazių sąrašas.
Išimtys	-

Klasės *FoundLogicNoun3, FoundLogicNoun 4, FoundLogicNoun 5* logika bei metodai atitinka *FoundLogicNoun2* logika, tik ieškoma pagal atitinkamai ilgesnius šablonus.

4.3. Realizacijos modelis

4.3.1. Programinių komponentų architektūra

Žemiau pateiktame (35 Pav. Programų komponentų architektūra) paveiksle pateikta programos komponentų sąveika tarpusavyje.



35 Pav. Programų komponentų architektūra

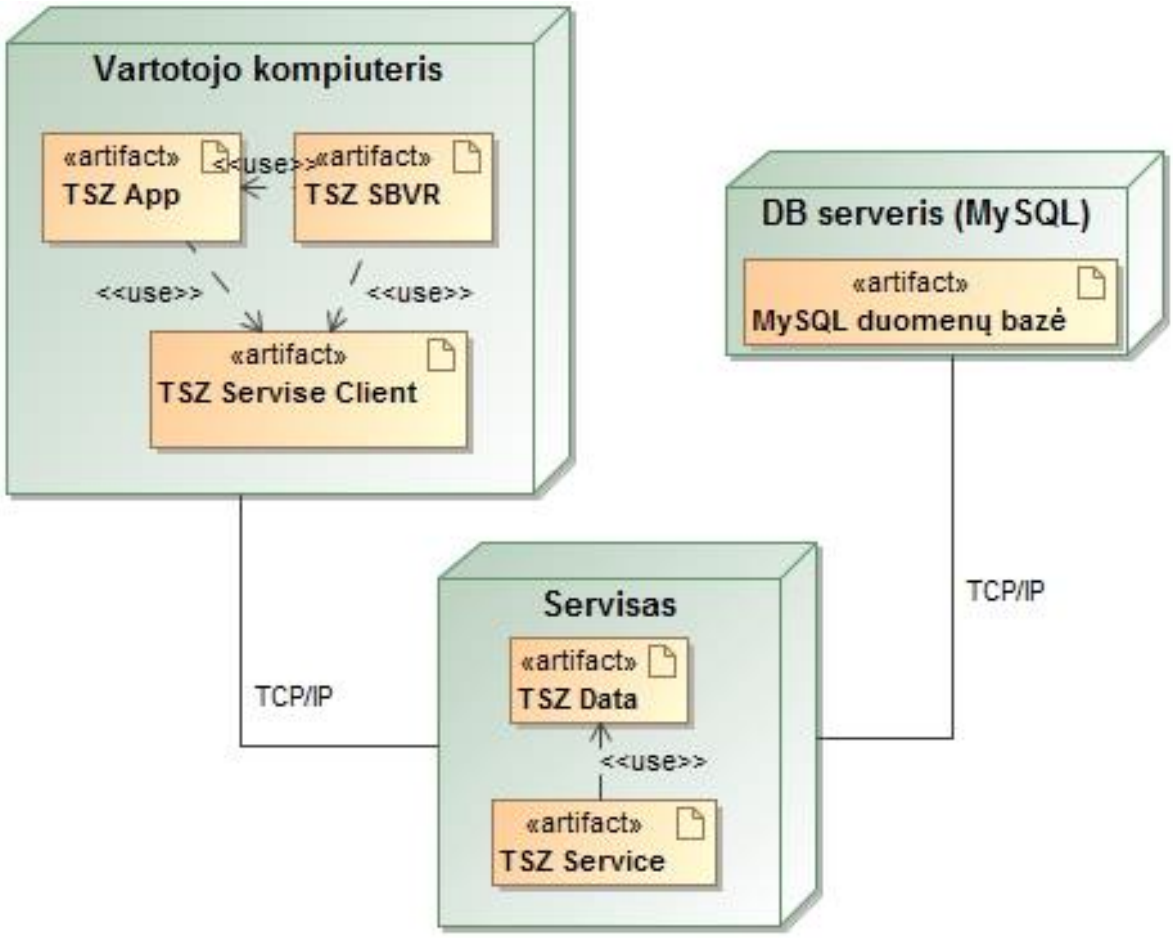
Sistemą sudaro šeši komponentai. Kiekvieno komponento trumpas aprašymas pateiktas (28 Lentelė Komponentų paaiškinimų lentelė) lentelėje.

28 Lentelė Komponentų paaiškinimų lentelė

Komponento pavadinimas	Paaiškinimas
TSZ App	Komponentas kris atlieka pagrindinį programos veikimą: kuria šablonus, ieško frazių ir t.t.
TSZ SBVR	Komponentas atsakingas už SBVR žodyno sudarymą,
TSZ Service Client	Komponentas atsakingas už bendravimą su servisu bei duomenų siuntimą per TCP/IP.
TSZ Service	Komponentas atliekantis statistikos analizę, duomenų išsaugojimą.
TSZ Data	Komponentas kuris atsakingas už duomenų išsaugojimą į duomenų bazę.
MySQL duomenų bazė	Komponentas skirtas duomenims saugoti (duomenų bazė).

4.3.2. Diegimo modelis

Žemiau pateiktame (36 Pav. Sistemos diegimo modelis) paveiksle vaizduojamas pagrindinis komponentų išsidėstymas diegimo būsenoje. Komponentai: *TSZ App, TSZ SBVR, TSZ Service Client* yra diegiami pas vartotoją į kompiuterį. Komponentai: *TSZ Data*, *TSZ Service* yra diegiami atskirai į nutolusį servisą. Kaip jau minėta komponentų architektūros dalyje, *TSZ Service Client* ir *TSZ Service* bendrauja TCP/IP protokolu, per internetą. Komponentas *MySQL duomenų bazė* yra duomenų bazė kurioje saugomi duomenys iš vartotojo kompiuterio, kurie pateikiami per servisą. Servisas su duomenų baze bendrauja internetu TCP/IP protokolu.



36 Pav. Sistemos diegimo modelis

5. AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU REALIZACIJA

5.1. Realizacijos ir veikimo aprašymas

5.1.1. Darbo pradžia

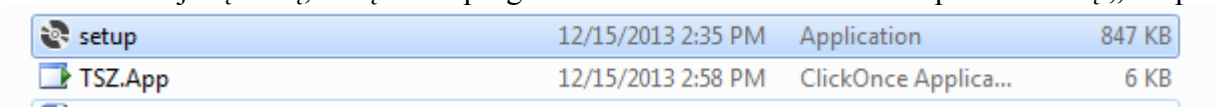
Prieš pradedant darbą sistemoje - Teminės srities žodyno sudarinėtojuje (toliau TSZ), reikia instaliuoti papildomus priedus, jai jų nėra jau instaliuota kompiuteryje:

- .NET 4 – 4,5
- Windows PowerPack
- SQL ServerPack

Pradėjus instaliuotis, mūsų sistemos instaliacijos vedlys pats pamatys, kokių priedų trūksta sistemoje, kad TSZ veiktų tinkamai ir korektiškai sistemoje.

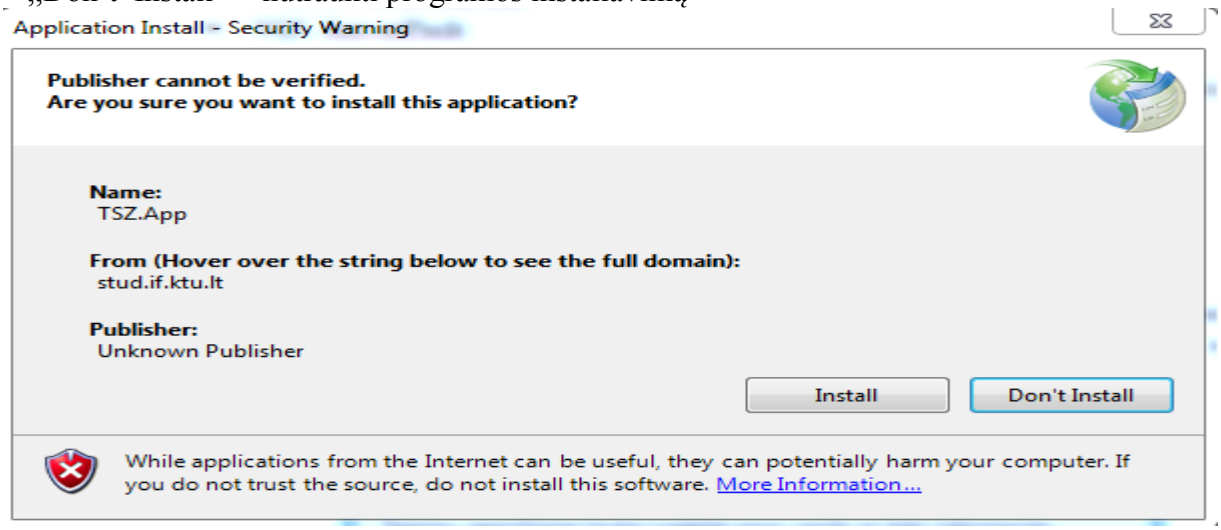
5.1.1.1. Instaliavimas

Nuėjus į vietą, kur įkėlėte programos instaliavimo failus reikia paleisti failą „setup“.



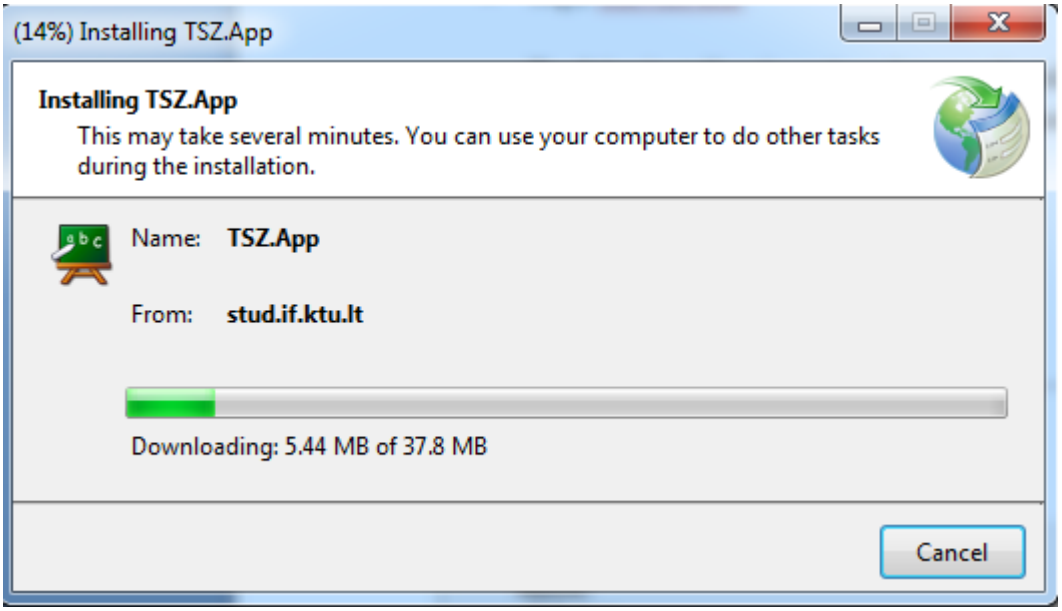
37 Pav. Instaliavimo failai

Atsiradusiame naujame lange (62 pav.) galimi du variantai : „Install“ ir „Don’t Install“.
„Install“ – skirtas toliau tęsti programos instaliavimą
„Don’t Install“ - nutraukti programos instaliavimą



38 Pav. Instaliavimo langas

Pasirinkus „Install“ toliau atsiranda instaliavimo langas (63 pav.), kuriame matome instaliavimo progresą.

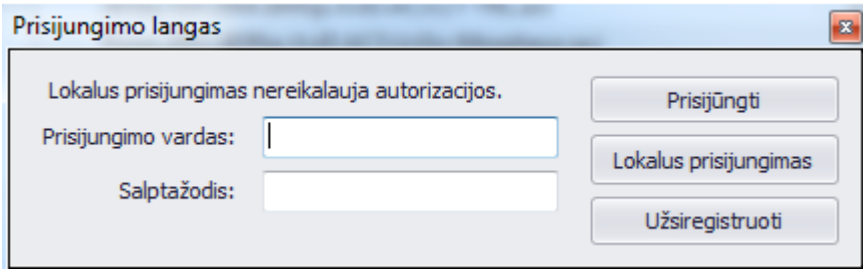


39 Pav. Instaliavimo progreso langas

Baigus instaliaciją, programa pirmą kartą pasileidžia automatiškai.

5.1.1.2. Prisijungimas

Kiekvieną kartą paleidžiant programa atsiranda prisijungimo langas (61 pav.). Prisijungimo lange galima pasirinkti, kuriame režime norima dirbti. Galima suvesti prisijungimo vardą ir slaptažodį ir paspaudus mygtuką „Prisijungti“ įgalinti funkcijas skirtas darbui su servisu. Lokalus prisijungimas turi visas funkcijas, išskyrus bendravimą su servisu. Užsiregistruoti skirta, naujam vartotojui prisiregistruoti prie serviso.

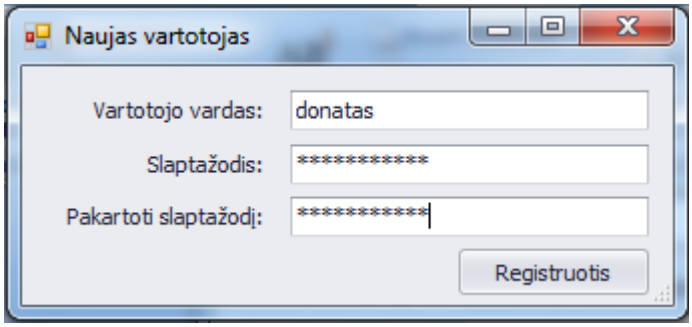


40 Pav. Prisijungimo langas

5.1.1.3. Registracija

Registracija skirta vartotojams, kurie norės naudotis programos servisu. Prisijungimo lange (65 pav.) reikia paspausti mygtuką „Užsiregistruoti“. Atsiradusiame naujame lange reikia užpildyti tokius laukus:

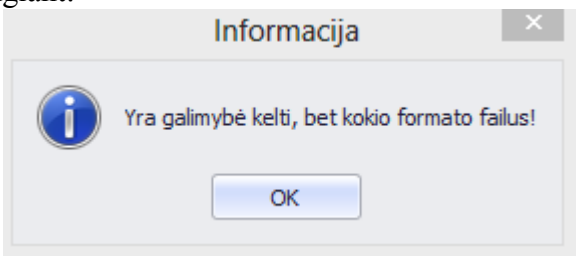
- Vartotojo vardas – vartotojo vardas, su kuriuos bus prisijungiama prie programos serviso.
- Slaptažodis – vartotojo slaptažodis, su kuriuo bus jungiamasi prie programos serviso
- Pakartoti slaptažodį – patikrinti ar vartotojas gerai suvedė slaptažodį



41 Pav. Registracijos langas

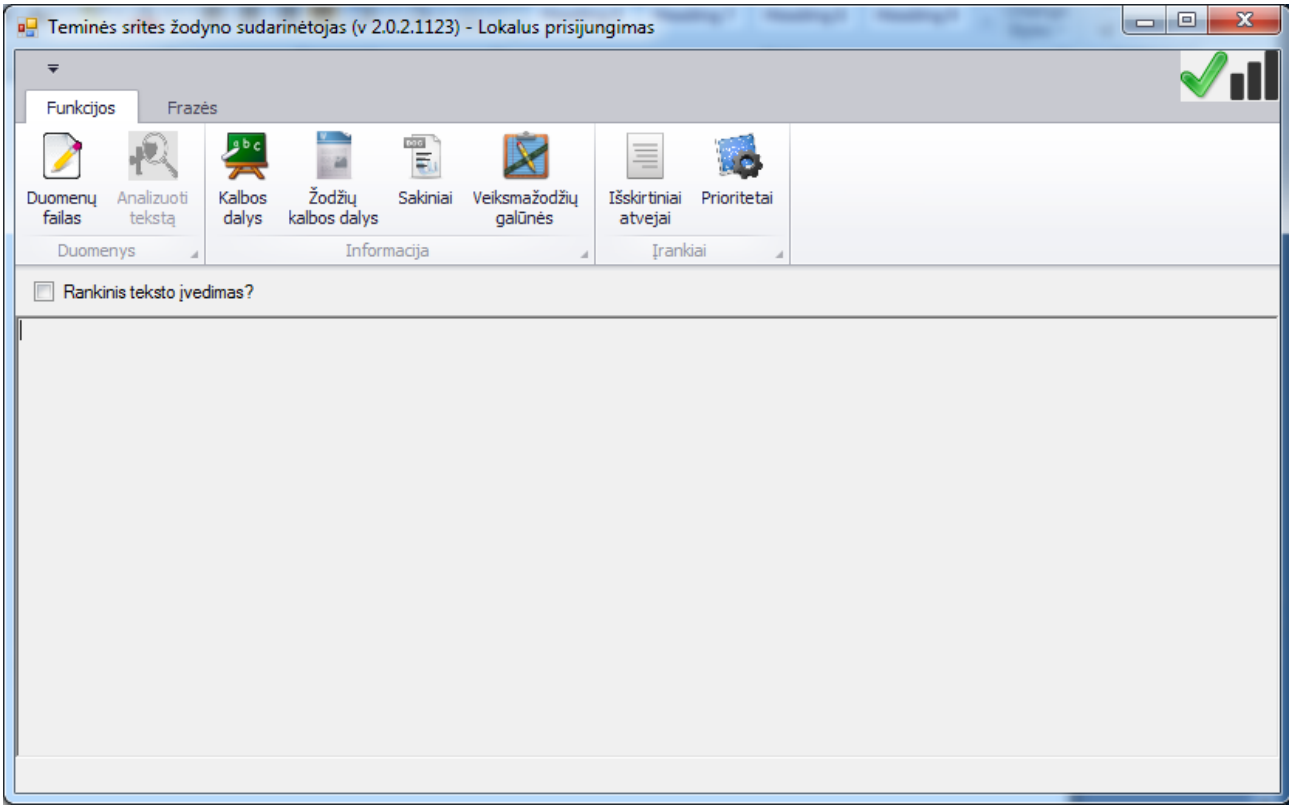
5.1.1.4. Sistemos informacija

Apie naujas sistemos funkcijas, ar naujus priedus, vartotojas yra supažindinamas, prieš sistemai įsijungiant:



42 Pav. Sistemos informacinis langas

5.1.2. Sistemos langas



43 Pav. Pagrindinis programos langas

Šis langas yra pagrindinis vartotojo funkcionalumo langas, kuriame jis galės pilnai valdyti TSZ sistemą.

5.1.2.1. Informaciniai ženklai

TSZ sistema turi du pagrindinius sistemos būklės nusakymo elementus:

- Serviso jungiamumo:

  Ryšys su servisu yra: 

  Ryšio su servisu nėra: 

- Interneto jungiamumo:

  Internetas yra: 

  Interneto nėra: 

Nesant ryšio su servisu nėra galimybės rastas frazes išsaugoti sistemos pagrindiniame serveryje. Nesant ryšio su internetu nėra galimybės kelti bet kokio tipo tekstų, kadangi nėra galimybės susisiekti su pagrindiniu VDU servisu, kuris galėtu morfologiškai išanotuoti įkeltą tekstą.

5.1.2.2. Sistemos funkciniai laukai

Pagrindinis sistemos langas yra suskirstytas į tris pagrindines dalis:

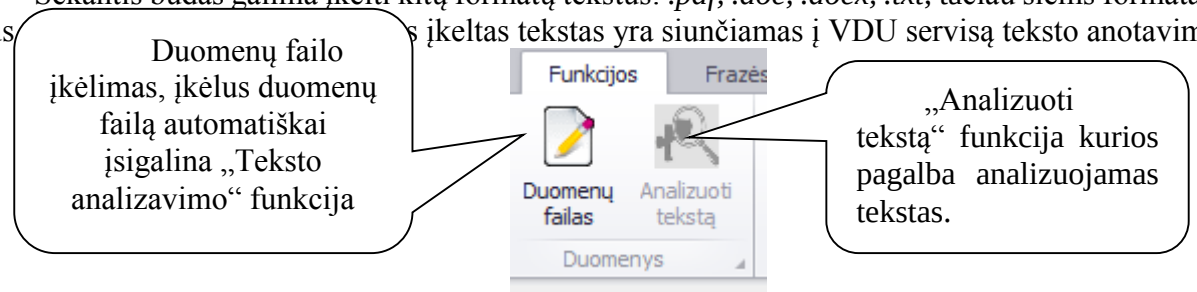
- Funkcijos – čia pateiktos pagrindinės naudojamos funkcijos, kurios yra dar papildomai suskirstytos į tris pogrupius: *Duomenys, Informacija, Įrankiai.*
- Frazės – čia pateiktas funkcijų sąrašas kurių pagalba galima valdyti frazių paiešką, tvarkyti rastas frazes, ar tiesiog peržiūrėti rastų frazių statistika (informacija: norint peržiūrėti frazių statistika būtina burėti ryšį su servisu). Ši dalis suskirstyta taipogi į tris pogrupius: *Frazės, Frazių statistika, Sinchronizavimas.*
- Teksto įvedimo langas.

5.1.2.3. Duomenų pateikimas sistemai

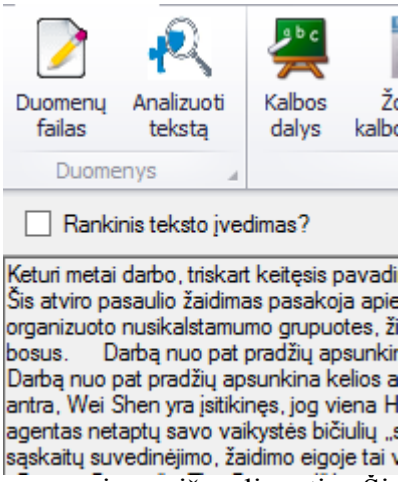
Sistemai pateikti duomenis galima pateikti keliais būdais.

Pirmas būdas yra tiesiog sistemai pateikti suanotuotą duomenų failą. Šitokiam duomenų failo tipui nereikia interneto prisijungimo.

Sekantis būdas galima įkelti kitų formatų tekstus: *.pdf, .doc, .docx, .txt*, tačiau šiems formatams įkėlus s įkeltas tekstas yra siunčiamas į VDU servisą teksto anotavimui.



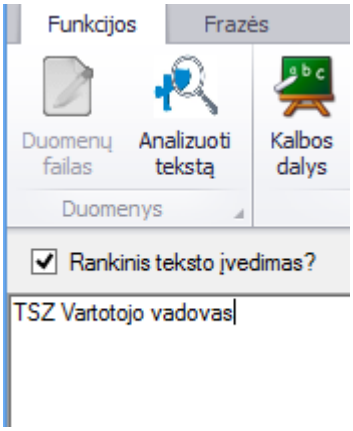
Įkėlus duomenų failą, galima pasirinkti iš karto analizuoti tekstą, arba galima šiek tiek vėliau, nuspaudus „Analizuoti tekstą“ mygtuką.



Sėkmingai įkėlus tekstą sistema įsigalina „Analizuoti tekstą“ mygtukas, taipogi analizuojamas tekstas yra pateikimas vartotojui, peržiūrėti žemiau esančiame lange.

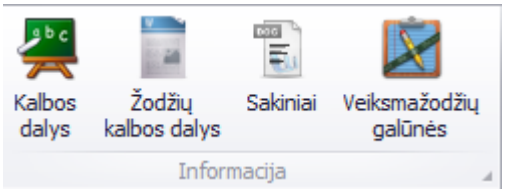
Trečias būdas tiesiog įgalinus „Rankinio teksto“ vedimo funkciją, galima tekstą suvesti ranka. Tačiau čia kaip ir antru būdu privalomas internetas teksto suanotavimui VDU servise.

Įgalinus rankinio teksto vedimą, galima parašyti tekstą, kurį norima išanalizuoti. Šio teksto analizavimui atlikti galima nuspaudus „Analizuoti funkcija“.

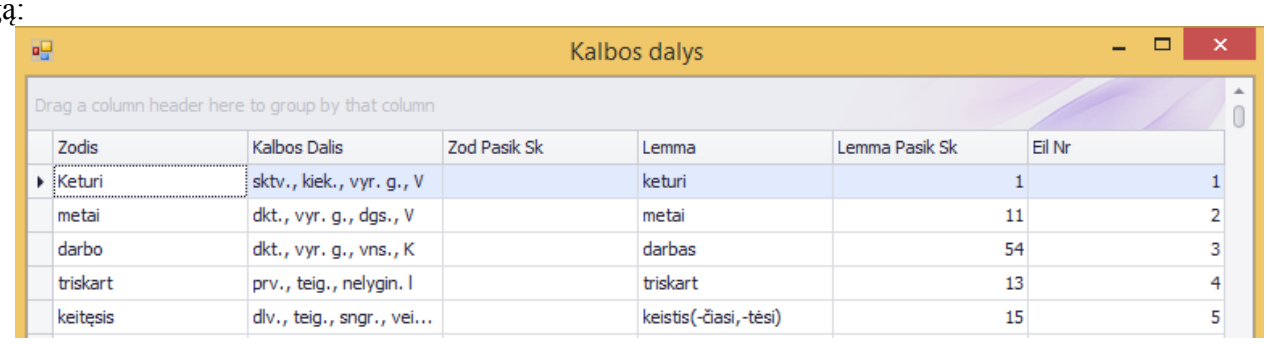


5.1.2.4. Informacijos pateikimas

Šioje dalyje galima peržvelgti analizuojamo teksto: kalbos dalis, išdetalizuotas žodžių kalbos dalis, atskirtus sakinius bei veiksmažodžių galūnes, tiesiog nuspaudus ant jų pasirinkimo mygtuko.



Tarkime pasirinkus „Kalbos dalys“ išvystame langą:



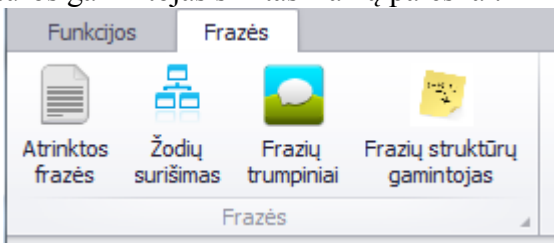
Zodis	Kalbos Dalis	Zod Pasik Sk	Lemma	Lemma Pasik Sk	Eil Nr
Keturi	sktv., kiek., vyr. g., V		keturi	1	1
metai	dkt., vyr. g., dgs., V		metai	11	2
darbo	dkt., vyr. g., vns., K		darbas	54	3
triskart	prv., teig., nelygin. I		triskart	13	4
keitęsis	dlv., teig., sngr., vei...		keistis(-čiasi,-tėsi)	15	5

44 Pav. Kalbos dalių langas

- Žodis – tekste esantis žodis.
- Kalbos dalis – rasto žodžio kalbos dalis
- Lema – žodžio veiksmažodinė forma
- Lemos pasikartojimo skaičius – Tekste rastų vienodų lemų pasikartojimo skaičius.
- Eil Nr – Žodžio eilės numeris

5.1.3. Frazės

Iš meniu pasirinkę „Frazės", gauname sąrašą įrankių skirtų dirbti su frazėmis. Pirmi trys įrankiai (Atrinktos frazės, Žodžių surišimas, Frazės-sinonimai) skirti darbui su surastomis frazėmis. Frazių struktūros gamintojas skirtas frazių paieškai.

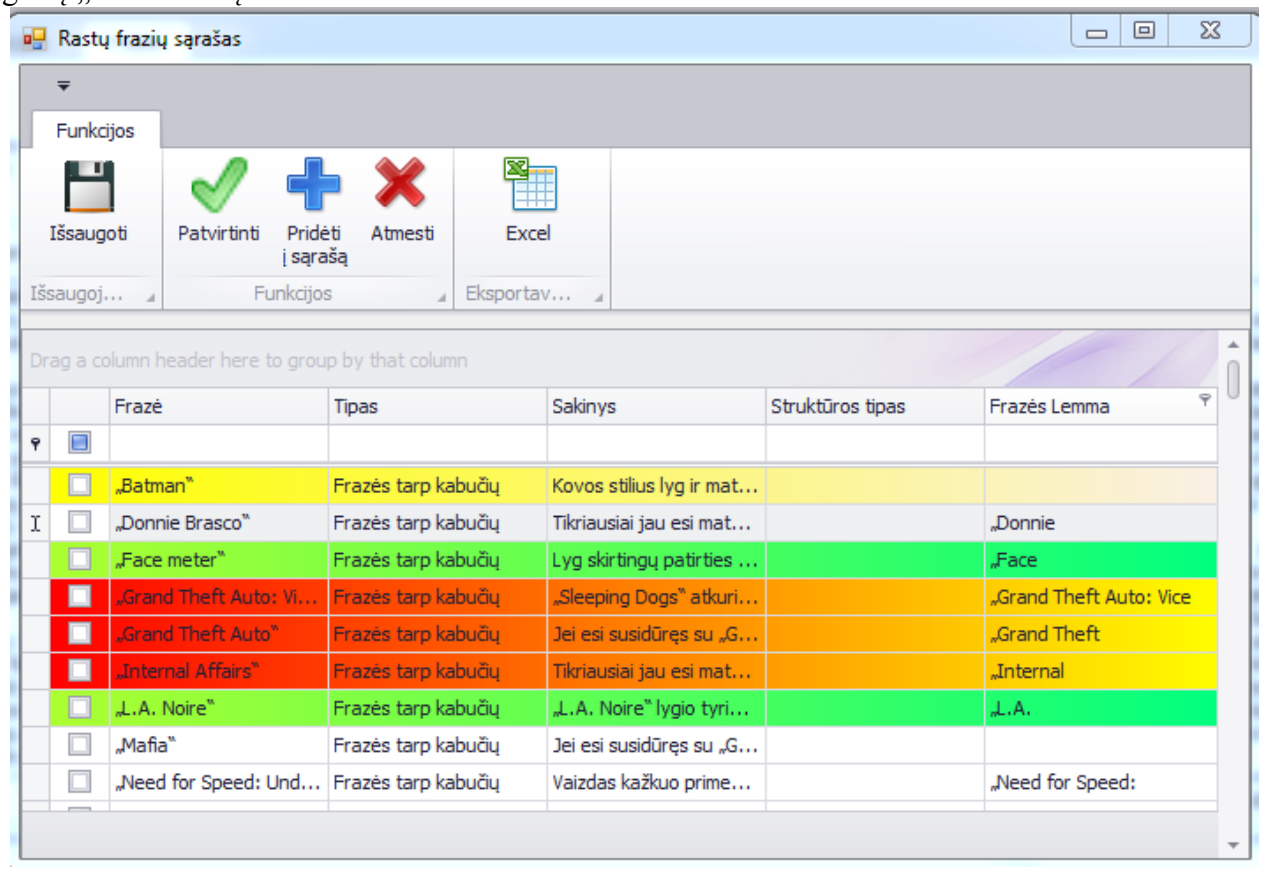


45 Pav. Frazių menių lango pasirinkimai

- Atrinktos frazės – atvaizduoti rastas frazes.
- Žodžių surišimas – leidžia padaryti faktus pvz. „Sleeping dog yra žaidimas"
- Frazė sinonimai – atvaizduoja rastu sinonimus pvz. KTU – Kauno Technologijos Universitetas
- Frazių struktūrų gamintojas – leidžia sudaryti frazių šablonus, juos redaguoti .

5.1.3.1. Atrinktos frazės

Atrinktose frazėse matome informaciją apie frazę (frazės tipas, sakinys, struktūros tipas, kuri rado frazę, frazės lemos). Yra penkios galimybės: išsaugoti, patvirtinti, pridėti į sąrašą, atmesti arba išsaugoti į „Excel" failą.



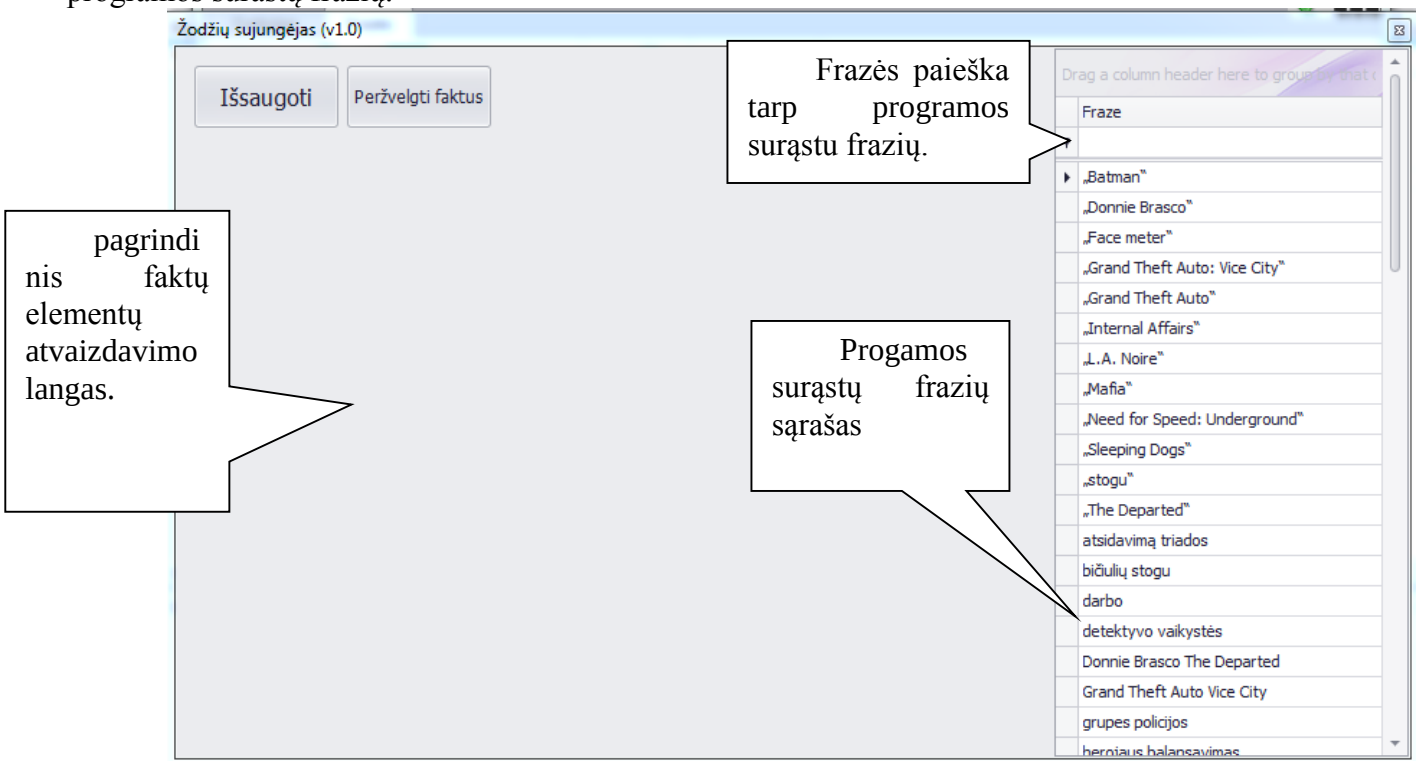
46 Pav. Atrinktų frazių pagrindinis langas

- Išsaugoti – išsaugo patvirtintas frazes ir panaikina atmestas frazes iš rastų frazių sąrašo.
- Patvirtinti – patvirtinti, kad programos surasta frazė yra tinkama. Patvirtintos frazės atvaizduojamos frazių sąraše žalia spalva.

- Pridėti į sąrašą – galima į programos surastų frazių sąrašą pridėti naują frazę. Frazių sąraše pridėtos frazės atvaizduojamos geltona spalva.
- Atmesti – leidžia atmesti surastas frazes kaip netinkamas. Frazių sąraše jos atvaizduojamos raudona spalva.

5.1.3.2. Žodžių surišimas

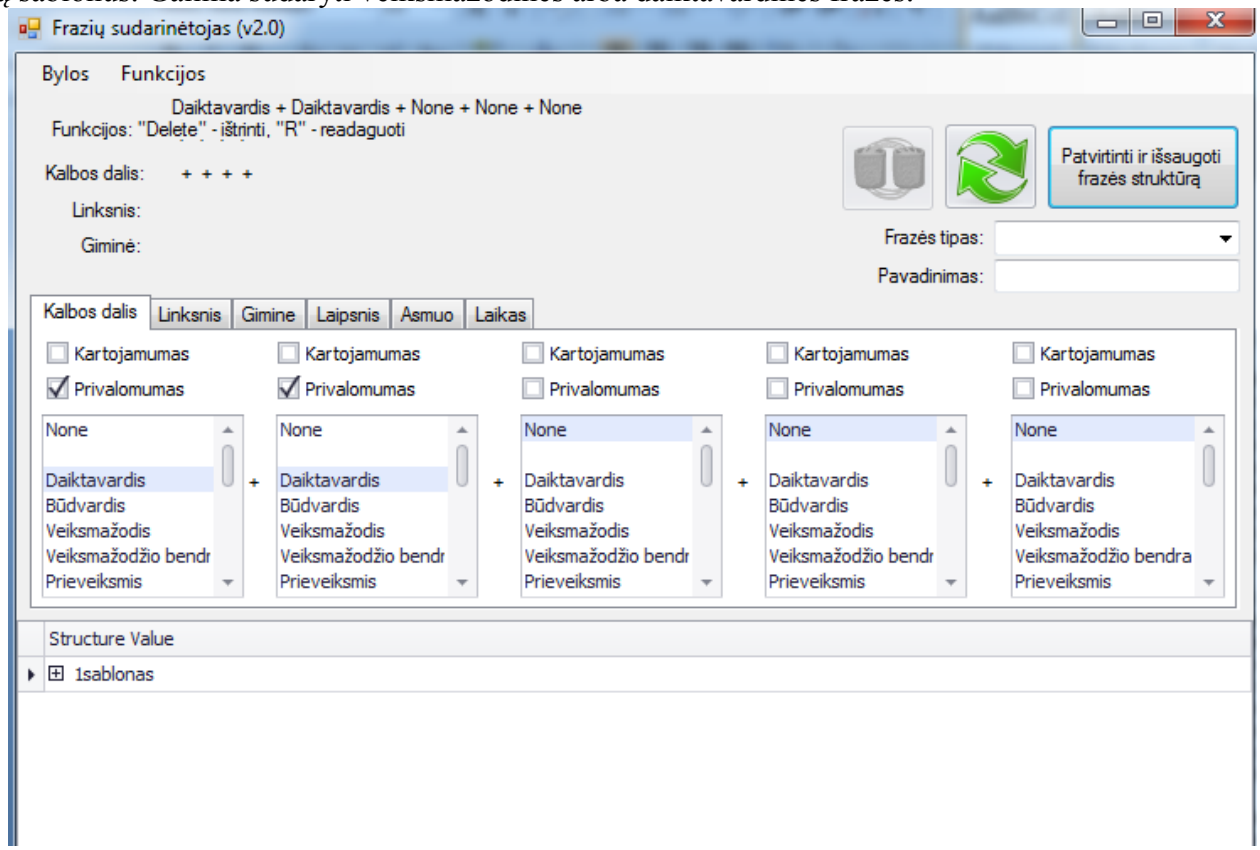
Galima rastas frazes tarpusavyje sujungti ir padaryti kaip faktą pvz. „Sleeping dog yra žaidimas.“. Faktą galima sudaryti iš programos surastų frazių arba įvesti savo. Faktai sudaryti iš elementų. Fakto elementas yra įrašas, tiek iš frazių sąrašo, tiek įvestas. Yra frazės paieška tarp programos surastų frazių.



47 Pav. Žodžių sujungėjo langas

5.1.3.3. Frazių šablonų gamintojas

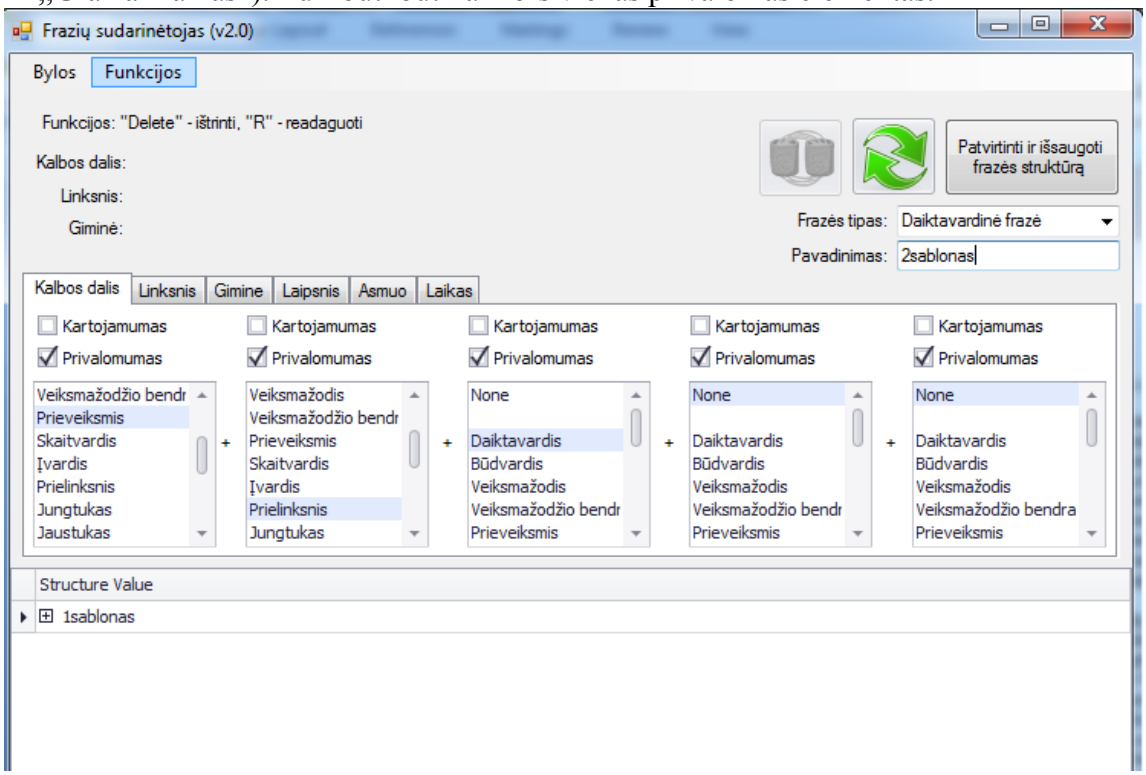
Žemiau yra pavaizduotas frazių šablono sudarytojo langas. Šiame lange yra galimybė sudaryti frazių šablonus. Galima sudaryti veiksmažodines arba daiktavardines frazes.



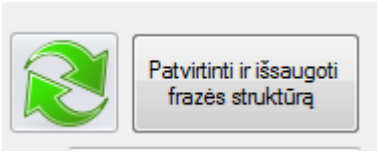
48 Pav. Frazių sudarinėtojo langas

5.1.3.3.1. Naujo šablono įvedimas

Galima pasirinkti, kad šablone butų nurodyta kalbos dalis, linksnis, giminė, laipsnis, asmuo, laikas. Programa automatiškai atmeta neleistinus pasirinkimus pvz. pasirinkus pirmame elemente būdvardis. Šiam elementui nebus leista pasirinkti laipsnio, asmens ir laiko. Iš viso gali būti nurodyti penki elementai. Tai pat gali būti nurodyta ar elementas daug kartų pasikartoja (šablonas yra būdvardis ir pažymėta ,,Kartojamumas“, sekantis elementas yra daiktavardis, sakinys yra ,,Rudas , piktas šuo bėga”, tai šablonas ras ,,Rudas piktas šuo”). Yra galimybė pažymėti privalomumą (Šablonas sudarytas iš trijų elementų prieveiksmis privalomas, prielinksnis neprivalomas ir daiktavardis privalomas. Du sakiniai ,,Gražiai į namus parėjau“ ir ,,Gražiai namas atrodo”. Šis šablonas ras dvi frazes ,,Gražiai į namus“ ir ,,Gražiai namas“). Turi būti būtinai nors vienas privalomas elementas.



49 Pav. Frazių sudarinėtojo langas

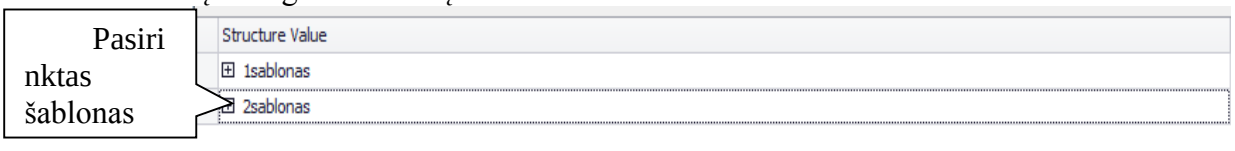


Norėdami išsaugoti šabloną reikia paspausti mygtuką ,,Patvirtinti ir išsaugoti frazės struktūrą”.

Skirta dar kartą išnagrinėti įkeltą tekstą su esančiomis ir naujai įvestais šablonais.

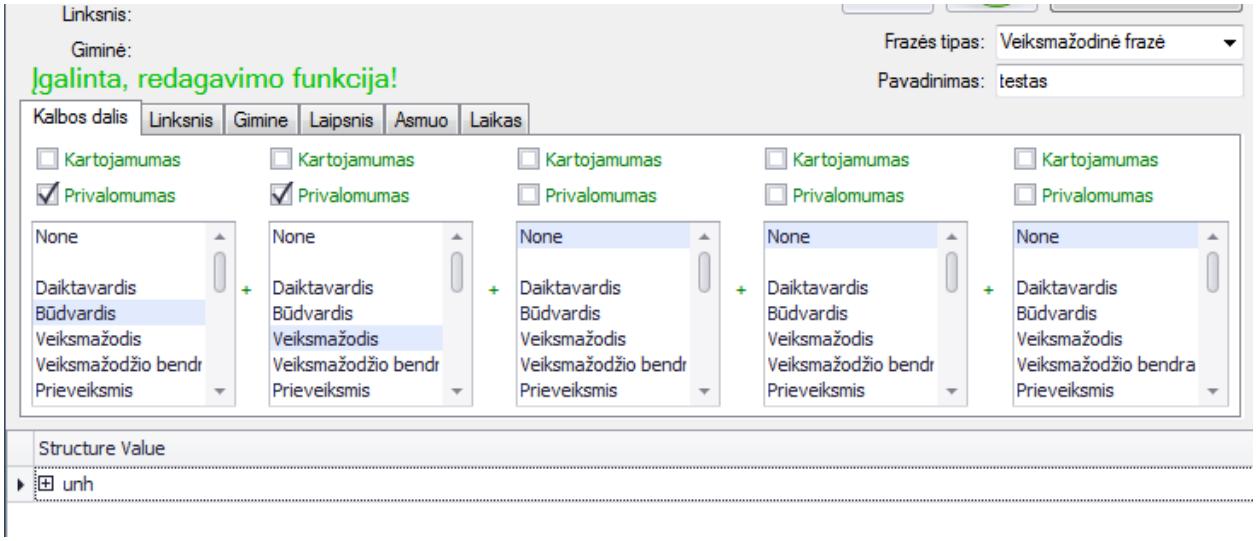
5.1.3.3.2. Šablono redagavimas

Yra galimybė išsaugotus šablonus redaguoti. Reikia iš išsaugotų šablonų sąrašo pasirinkti norimą redaguoti šabloną.



50 Pav. Frazių šablonų sąrašas

Pasirinkus norima šabloną, reikia paspausti klaviatūros klavišą ,,R“

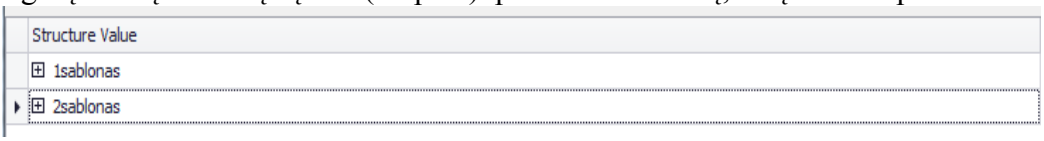


51 Pav. Frazių sudarinėtojo langas redagavimo režime

Jeigu esate redagavimo režime atsiranda užrašas „Įgalinta, redagavimo funkcija!“. Tada galima keisti šablono elementų reikšmes. Baigus keisti, kad išsaugotų pakeitimus reikia paspausti mygtuką „Patvirtinti ir išsaugoti frazės struktūrą”.

5.1.3.3.3. Šablono ištrynimas

Yra galimybė nebereikalingą šabloną pašalinti, kad pagal jį būtų nebeieškoma frazių. Reikia iš išsaugotų frazių šablonų sąrašo (87 pav.) pasirinkti šabloną, kurį norima pašalinti



52 Pav. Frazių šablonų sąrašas

Reikia paspausti klaviatūros mygtuką „Delete“. Pažymėtas frazės šablonas yra pašalinamas.

5.2. Testavimo modelis

Pagrindinis testavimo tikslas yra pateikti ir įrodyti, jog parašyta programa veikia tinkamai ir kode ar klasėse ar metoduose nėra klaidų. Mažai šansų, jog ištestavus programą galima sakyti, jog programa nuo šiol veiks tinkamai ir gerai, tačiau testavimas sumažina blogo kodo riziką bei padidina programos efektyvumą bei lankstumą. Taipogi sistemos eksploatavimo metu kūrėjai įsipareigoja taisyti programos klaidas.

Remiantis testavimo planu pagrindiniai testavimo objektai bus klasės bei juose esantys metodai. Taipogi nemaža dalis testavimo apims ryšių testavimui tarp klasių bei metodų. Taipogi atliekant teminės srities žodyno testavimą, bus siekiama patikrinti ir užtikrinti, kad sistema atitinka sistemos specifikaciją.

5.2.1. Testavimo apimtis

Teminės srities žodyno sudarinėtojo programą galima suskaidyti į tokius elementus:

- Vartotojo sąsaja
- Sistemos moduliai
- Stambesnieji sistemos moduliai (tai sistemos moduliai kurie yra sudaryti iš kelėtos ar daugiau smulkesnių sistemos modulių)

Sistemos testavimui bus taikomi šie testavimo metodai: vienetų testavimas, integracijos testavimas ir aukšto lygio testavimas. Toliau pateiksime kurie sistemos elementai bus testuojami kiekvieno testavimo etapo metu:

Vienetų testavimas:

- Sistemos moduliai
- Integruoti sistemos moduliai

Šiame etape naudojamas juodos dėžės testavimas.

Integracijos testavimas:

- Vartotojo sąsaja
- Sistemos moduliai
- Integruoti sistemos moduliai

Aukšto lygio testavimo:

- Vartotojo sąsaja
- Sistemos moduliai

5.2.2. Testavimo strategija

5.2.2.1. Vienetų testavimas

Vienetų testavimo etape bus testuojami atskiri teminės srities žodyno sudarinėtojo moduliai ir paketai. Bus taikomas baltos dėžės testavimo metodas, kai testuotojas gali disponuoti programos kodu. Pirmiausiai bus išsiaiškinta, kokius duomenis reikia paduoti į atitinkamą programos modulį ir bus nustatomi laukiami rezultatai. Tada automatizuotai bus paduodami testiniai duomenys ir tikrinami rezultatai. Testavimo rezultatą matysime testavimo įrankio išvedimo lange, kuriame bus parašytas atitinkamas rezultatas, ar šis modulis praėjo testą ar ne.

Vienetų testavimas tai pats pirmas testavimo etapas, kurį atlieką patys programuotojai. Jie turi išsiaiškinti ar modulis veikia tinkamai. Jai modulis veikia, ne tinkamai, tai programuotojai privalo ištaisyti šias klaidas kuo skubiai, nes šio testavimo metu atsiradusios klaidos, gali pakenkti tolimesniam sistemos veikimui.

5.2.2.2. Aukštesnio lygmens testavimas

Šiame etape bus testuojami nefunkciniai sistemos reikalavimai. Teminės srities žodyno sudarinėtojo vienas svarbiausių nefunkcinių reikalavimų yra efektyvus kompiuterio resursų panaudojimas, taipogi lengva bei taipogi intuityvi bei turinti pakankamai funkcionalumo vartojo sąsaja. Greitaveikos reikalavimai yra aprašyti sistemos specifikacijoje.

5.2.3. Testavimo resursai

Programiniai testavimo resursai:

- Teminės srities žodyno sudarinėtojas
- *UnitTest* įrankis, testams realizuoti

Žmogiškieji resursai:

- Programuotojas/testuotojas

5.3. Testavimo duomenys ir rezultatai

Pagrindiniai sistemos testavimo elementai:

- Vardų ir pavardžių atpažinimo algoritmas;
- Sutrumpinimų radimo algoritmas;
- Sakinių išskaidymo algoritmas;
- Frazių radimo pagal šablonus (daiktvardinių frazių, veiksmažodinių frazių) algoritmai;

5.3.1. Vardų ir pavardžių algoritmo testavimas

Šio algoritmo testavimui buvo pasirinkti penki sakiniai kuriuose yra įvairaus tipo vardų bei pavardžių pvz.: Eimantas Žlabys, E.Žlabys, E.Ž., Valdas Mykolas Brazauskas ir t.t.

Žemiau pateiktuose lentelėse matyti gauti rezultatai iš įvairaus tipo sakinių.

1) *Kaip direktorių yra sake V.Adamkus man reikia kad perimtu E.Žlabys visa valdymą*
   Gauti rezultatai: *V.Adamkus E.Žlabys*
2) *Kaip ir V.Adamkus, direktorių yra sake V.A. man reikia kad e.z. perimtu B.A. visa valdyma*
   Gauti rezultatai: *V.Adamkus V.A. B.A.*
3) *Kaip direktorių yra sake V.A. man reikia kad e.z. perimtu B.A. visa valdymą*
   Gauti rezultatai: *V.A. B.A.*
4) *Kaip direktorių yra sake Valdas Adamkus man reikia kad perimtu visa valdymą*
   Gauti rezultatai: *Valdas Adamkus*
5) *Kaip direktorių yra sake Valdas Mykolas Brazauskas man reikia kad perimtu visa valdymą A.B.*
   Gauti rezultatai: *Valdas Mykolas Brazauskas A.B.*

5.3.2. Sutrumpinimų radimo algoritmo testavimas

Šio algoritmo testavimui buvo pasirinkti įvairaus tipo sutrumpinimai kurie gali pretenduoti į sutrumpinimus, tačiau pridėti ir klaidingi elementai:

Sakinys: *KTU pats geriausias universitetas lyginant su VDU VGTU V ar Vu kitais Aukštaisiais LIETUVOS Universitetais*

Gauti rezultatai: KTU VDU VGTU LIETUVOS

5.3.3. Sakinių išskaidymo algoritmas

Šio algoritmo tikslas yra tinkamai išskirti sakinius iš teksto. Šio algoritmo testavimui buvo pasirinkti penki sakinių tipai kurie gali dažniausiai būti vartojami tekste. Taipogi prie kiekvieno sakinio prikabinamos įvairaus tipo galūnės kuriomis gali užsibaigti sakiniai (29 Lentelė Galimi sakinių pabaigos simboliai):

29 Lentelė Galimi sakinių pabaigos simboliai

SAKINIO PABAIGOS ŽENKLAS	PAAIŠKNIMAS
.	Taškas
?	Klaustukas
!	Šauktukas
...	Daugtaškis

Sakiniai su kuriais buvo atliekami testavimai:

Šios sutarties sąlygos taikomos abonementui, fotogalerijai, (kartu visos paminėtos paslaugos vadinamos susijusiai programinei įrangai, žiniatinklio svetainei arba paslaugai (apskritai – paslaugoms)
Šios sutarties sąlygos taikomos „Microsoft! Hotmail“, „Microsoft.?!?!?! SkyDrive“, „Microsoft“ abonementui, „Windows Live Messenger“, „Windows“ fotogalerijai, „Windows Movie Maker“, „Microsoft Mail Desktop“, „Windows Live Writer“ (kartu visos paminėtos paslaugos vadinamos firminėmis „Microsoft“ paslaugomis), taip pat „Bing“, MSN, „Office.com“ ir bet kokiai kitai su šia sutartimi susijusiai programinei įrangai, žiniatinklio svetainei arba paslaugai (apskritai – paslaugoms)
Kaip direktoriu yra sake „ Man reikia kad V. Adamkus perimtu visa valdyma“
1. V. Adamkus visada daug padedavo saliai
26. V. Adamkus teigė: „visada Ž. Eimantas bus prezidentas“ nuo 1989 metų

5.3.4. Frazių radimo pagal šablonus algoritmų testavimai

Šio algoritmo testavimui buvo pasirinktas vienas pagrindinis sakinys, kuriame yra įvairaus tipo kalbos dalių. Visi šablonai buvo generuojami automatiškai iš trijų elementų: Daiktavardžio, būdvardžio bei skaitvardžio. Iš šių trijų elementų buvo generuojamos nuo dviejų iki penkių ilgio šablonų elementai, kurie sudarydavo šabloną.

Testavimo elementai sudarė tokius atvejus:

- Visi šablono elementai yra daiktavardžiai;
- Šablono elementai maišosi tarpusavyje;

Veiksmažodinėms frazėms pagal šablonus testuoti buvo pasirinktas sakinys:
Studentai labai skubėjo į gražų Kauno Technologijos Universitetą.

Šį sakinį sudaro dvi dalys t.y. daiktvardinė bei veiksmažodinė frazė. Testavimui buvo pasirinkti keli tipai daiktvardinių bei veiksmažodinių frazių šablonai:

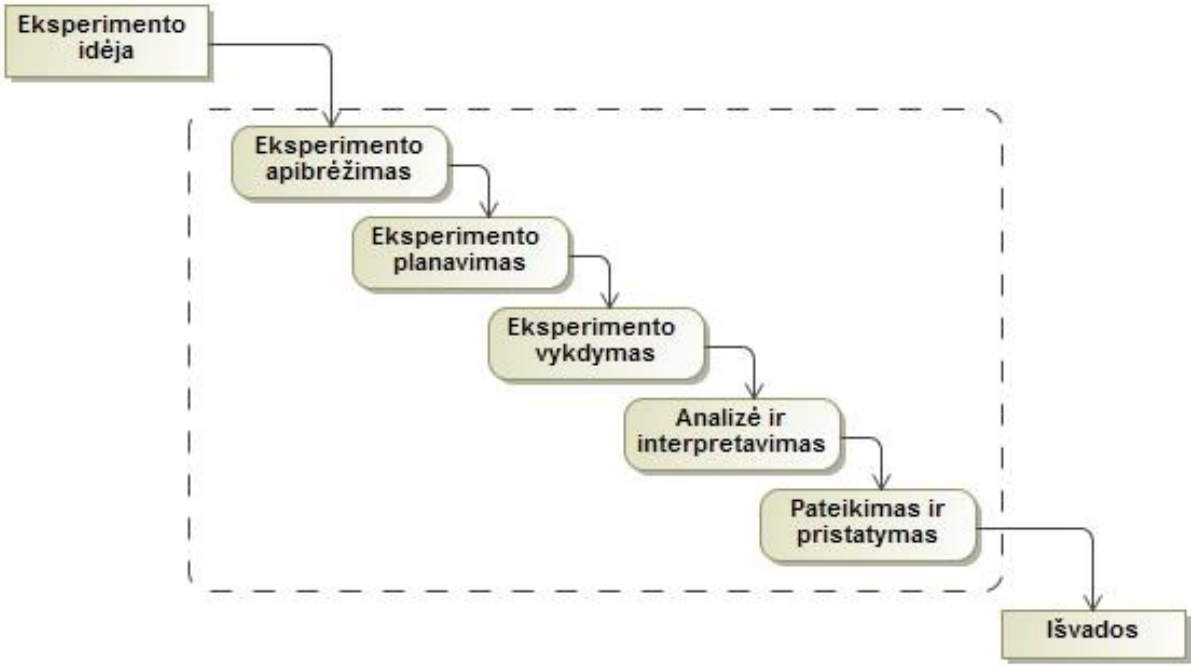
Veiksmažodinei frazei rasti: [Prieveiksmis] + Veiksmažodis + [Prielinksnis] + DF
Daiktavardinei frazei rasti: [Būdvardis] + Daiktavardis*

Gauti rezultatai:
Sistema išanalizavus sakinį pateikia šiuos rezultatus:

Rezultatas	*Tipas*
gražų Kauno Technologijos Universitetą	DF
Kauno Technologijos Universitetą.	DF
skubėjo į	VF

6. EKSPERIMENTINIS AUTOMATINIO FRAZIŲ ATPAŽINIMO TEKSTE SUDARYTŲ ŠABLONŲ PAGRINDU SISTEMOS TYRIMAS

Žemiau (53 Pav. Eksperementavimo procesas) pateikiamas eksperimentavimo proceso schema kuria remiantis bus vykdomas eksperimentas.



53 Pav. Eksperementavimo procesas

6.1. Eksperimento planas

Eksperimento tikslas – pateikiant sistemai bei asmenims kurie dalyvauja bandyme tam tikrus tekstus, patikrinti programos algoritmo veikimą.

Eksperimentą atlieka eilinis vartotojas kuris išanalizuoja tekstus rankiniu būdu, po to tekstus duoda apdoroti programai. Taipogi eksperimente dalyvauja ekspertas kuris taipogi tuos tekstus išanalizuoja rankiniu būdu.

Eksperimento grėsmės:

- Dėl taisyklių, kurių pagalba algoritmas aptinka frazes iš vartotojo sudarytų šablonų neatitikimo.
- Esant netiksliam šablonui galimi nekorektiški rezultatai.
- Dėl temos naujumo informacinių sistemų srityje ir informacijos stokos algoritmo eksperimento rezultatai nebus visiškai korektiški.

Eksperimente naudojami įrankiai:

- Algoritmo sugebandžio naudojant šablonus tekste aptikti daiktavardines bei veiksmažodines frazes prototipas

Eksperimente naudojami duomenys:

- Tam tikri atrinkti tekstai.
- Atrinkti bei sudaryti šablonų tipai, kurie bus naudojami sistemai aptikti frazes

6.2. Eksperimento rezultatai

Šiame skyriuje pateikiami eksperimento rezultatai kurie gaunami naudojant frazių pagal šablonus atpažinimo algoritmą ir taipogi asmens, kuris atlieka eksperimentą rankiniu būdu rezultatai.

Eksperimentui atlikti sistemoje naudojamų šablonų sudėtis ir elementų kiekiai yra nekintantys. Eksperimentui atlikti sistemoje pasirinkti penki šablonai, kurie galimi dažniausiai atspindi frazę.

Žemiau pateiktoje (30 Lentelė Sistemos ekeperimente naudojamų šablonų sąrašas) lentelėje išrašyti visų šablonų sudedamosios dalys:

30 Lentelė Sistemos ekeperimente naudojamų šablonų sąrašas

Šablonas	*Tipas*	*Paaiškinimas*
[Sk][B][D*](k)D*	Daiktvardinė frazė	Šabloną sudaro skaitvardis neprivalomas, būdvardis, kuris gali kartotis tačiau jis neprivalomas, daiktavardis kilmininko linksnyje ,kuris gali kartotis tačiau neprivalomas ir daiktavardis
Dal[B][D*](k)D*	Daiktvardinė frazė	Šabloną sudaro dalyvis, būdvardis, kuris gali kartotis tačiau jis neprivalomas, daiktavardis kilmininko linksnyje, kuris gali kartotis tačiau neprivalomas ir daiktavardis
DVbendr	Daiktvardinė frazė	Šabloną sudaro daiktavardis ir veiksmažodžio bendratis
V[Prl]	Daiktvardinė frazė	Šabloną sudaro veiksmažodis ir prielinksnis, kuris yra neprivalomas
VVbenr[Prl]	Veiksmažodinė frazė	Šabloną sudaro veiksmažodis, veiksmažodžio bendratis ir prielinksnis, kuris yra neprivalomas

Kadangi sistema automatiškai, nepriklausomai ar šablonai yra sudaryti ar ne ji ieško tekste vardų ir pavardžių, sutrumpinimų. Taigi prie šio eksperimento prisidės ne tik šablonų rastos frazės tačiau ir automatiškai tekste rastos frazės.

Algoritmui bei asmeniui atliekančiam eksperimentą yra pateikiami penki tekstai kurie yra bendros tematikos – politika. Šie tekstai yra patalpinti internete adresu: https://eimzlab.stud.if.ktu.lt/Tekstai%20analizei/.

Teksto analizei bei frazių radimui buvo iškelti kriterijai:

- Veikimo tikslumas (palyginimas sistemos ir asmens atliekančio eksperimentą)
- Kiek frazių buvo rasta tekste
- Kiek frazių yra korektiškų kiek nekorektiškų
- Asmens atlikusio eksperimentą bei sistemos rastų frazių palyginimas

6.2.1. Sukurtos sistemos eksperimento rezultatai

Sistemos algoritmas tekstuose surado 296 frazes. Žemiau esančioje (31 Lentelė Sistemos kiekvieno šablono rastų frazių skaičius) lentelėje yra pateikiamos tikslesnės kiekvieno šablono aptiktų frazių skaičius:

31 Lentelė Sistemos kiekvieno šablono rastų frazių skaičius

Šablono pavadinimas	*Rastų frazių skaičius*
[Sk][B][D*](k)D*	151
Dal[B][D*](k)D*	43
DVbendr	10
V[Prl]	62
VVbenr[Prl]	6
Viso:	272

Kaip matyti iš lentelės frazių atpažinimo naudojant šablonus algoritmas aptiko 238 frazes kurios tenkino sudarytus šablonus. Kitas 24 frazes sistema aptiko automatiškai ir juos priskyrė vardams bei pavardėms ir sutrumpinimams:

32 Lentelė Ne šablonų rastų razių skaičius

Šablono pavadinimas	*Rastų frazių skaičius*
Sutrumpintos frazės	1
Vardai bei pavardės	23
Viso:	24

Žemiau esančioje (33 Lentelė Sistemos rastų frazių kiekiai kiekviename tekste) lentelėje yra pateikiamas frazių skaičius rastas tam tikrame tekste. Taipogi 1priede pateikiamos visos sistemos rastos frazės:

33 Lentelė Sistemos rastų frazių kiekiai kiekviename tekste

Teksto pavadinimas	*Rastų frazių skaičius*
bals-Pasaulis-090907-2.txt	61
bals-Lietuva-091220-111.txt	27
bals-Lietuva-091213-104.txt	48
bals-Lietuva-091203-94.txt	81
bals-Lietuva-091124-85.txt	79
Viso:	296

6.2.2. Eilinio vartotojo atlikusio eksperimentą rezultatai

Asmuo atlikęs teksto analizę yra eilinis vartotojas, kuris sugeba tekste aptikti frazes, taipgi žino pagrindines kalbos dalis, tekstuose surado 74 frazes. Žemiau esančioje (34 Lentelė Eilinio vartotojo rastų frazių skaičius kiekviename tekste) lentelėje yra pateikiama kiek eilinis vartotojas aptiko frazių kiekviename tekste, vartotojo rastos frazės pateikiamos (2 priede):

34 Lentelė Eilinio vartotojo rastų frazių skaičius kiekviename tekste

Teksto pavadinimas	*Rastų frazių skaičius*
bals-Pasaulis-090907-2.txt	20
bals-Lietuva-091220-111.txt	11
bals-Lietuva-091213-104.txt	19
bals-Lietuva-091203-94.txt	14
bals-Lietuva-091124-85.txt	10
Viso:	74

6.2.3. Eksperto atlikusio eksperimentą rezultatai

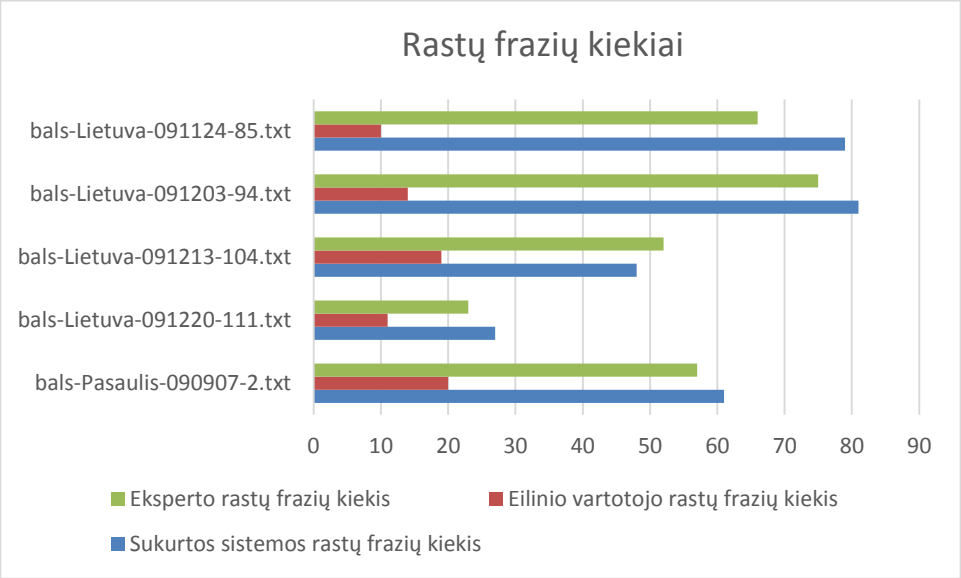
Asmuo atlikęs teksto analizę yra ekspertas, kuris gerai žino Lietuvių kalbos tekstų struktūra. Suvokia bei sugeba sudaryti bei rasti frazes. Taipogi asmuo sugeba ir taiko gramatines kalbos dalis frazėms aptikti. Ekspertas tekstuose surado 293 frazių. Žemiau esančioje (35 Lentelė Eksperto rastų frazių skiačius kiekviename tekste) lentelėje yra pateikiamos kiek kiekviename tekste ekspertas patiko frazių:

35 Lentelė Eksperto rastų frazių skiačius kiekviename tekste

Teksto pavadinimas	*Rastų frazių skaičius*
bals-Pasaulis-090907-2.txt	77
bals-Lietuva-091220-111.txt	23
bals-Lietuva-091213-104.txt	52
bals-Lietuva-091203-94.txt	75
bals-Lietuva-091124-85.txt	66
Viso:	293

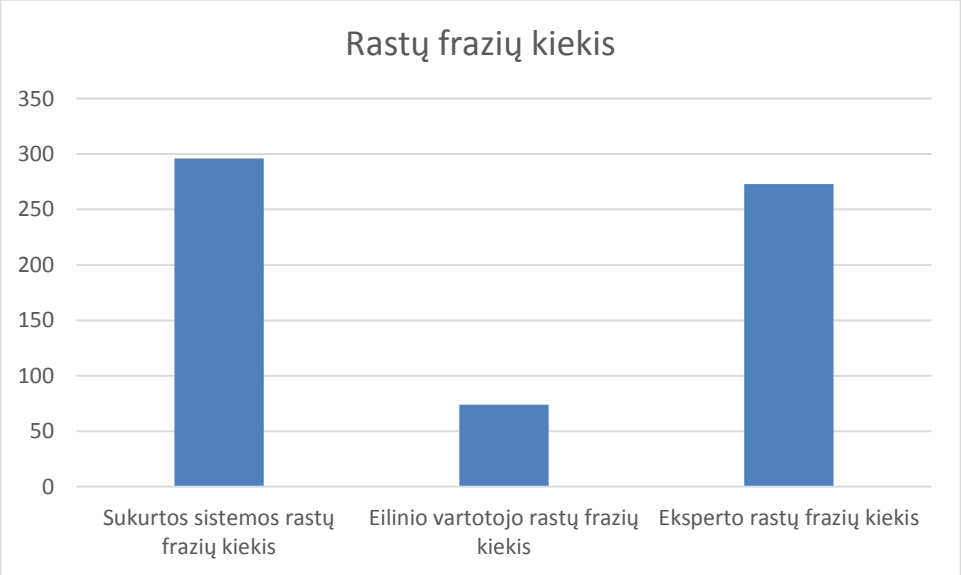
6.3. Sprendimo savybių analizė, kokybės kriterijų įvertinimas

Atlikus eksperimentus iš sistemos pusės bei asmenų kurie atliko teksto analizę, galime matyti, jog rastų frazių skaičius skiriasi. Žemiau pateiktoje diagramoje (54 Pav. Rastų frazių kiekiai) matyti kaip skiriasi frazių skaičius tam tikrame tekste.



54 Pav. Rastų frazių kiekiai

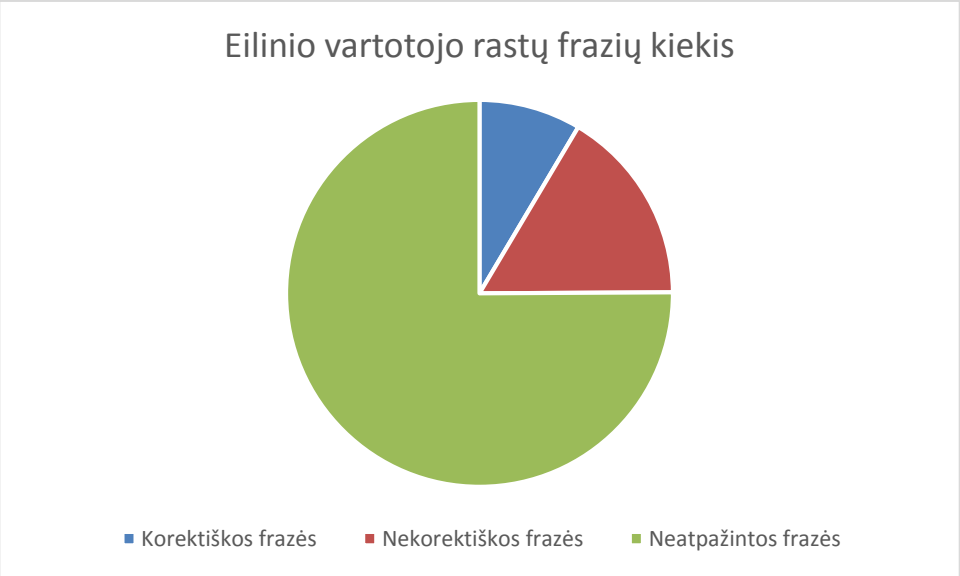
Iš viso rastų frazių kiekiai pateikti (55 Pav. Visuose tekstuose rastų frazių kiekiai) diagramoje:



55 Pav. Visuose tekstuose rastų frazių kiekiai

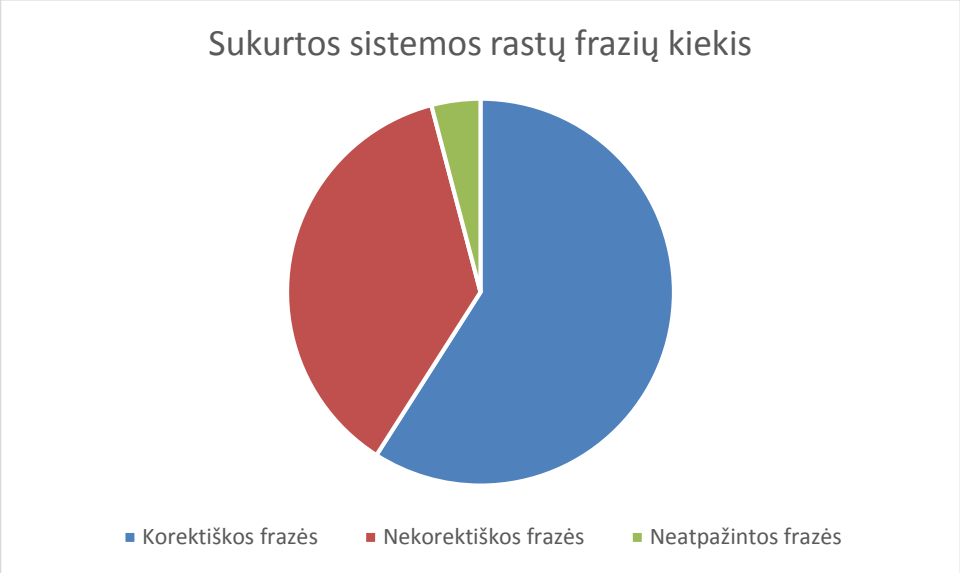
Žemiau pateiktose diagramose pavaizduoti kiek sukurtos sistemos (57 Pav. Sukurtos sistemos rastų frazių palyginimas su eksperto rastomis frazėmis) ir eilinio vartotojo (56 Pav. Eilinio vartotojo rastų frazių palyginimas su eksperto rastomis frazėmis) lyginant su ekspertu rastų frazių skirtumai. Lyginant rastas frazes buvo iškelti trys kriterijai, pagal kuriuos buvo matuojama:

- Korektiškos frazės – frazės rastos tiek eksperto tiek sukurtos sistemos yra vienodos;
- Nekorektiškos frazės – frazės rastos sukurtos sistemos nepilnai ar dalinai atitinka frazes rastas eksperto;
- Neatpažintos frazės – frazės rado ekspertas, tačiau sukurta sistemą jų nerado;



56 Pav. Eilinio vartotojo rastų frazių palyginimas su eksperto rastomis frazėmis

Kaip matyti iš 54pav. jog vartotojo rastų frazių skaičius lyginant su eksperto ganėtinai skiriasi. Vartotojas korektiškai rado tik 25 frazes lyginant su eksperto rastomis frazėmis. Nekorektiškai rastos, ar iš dalies panašios frazės su eksperto sudaro 48 frazes. Tuo tarpu nerastų frazių skaičius siekia 220, t.y. vartotojas neišskyrė frazių, kurias rado ekspertas.

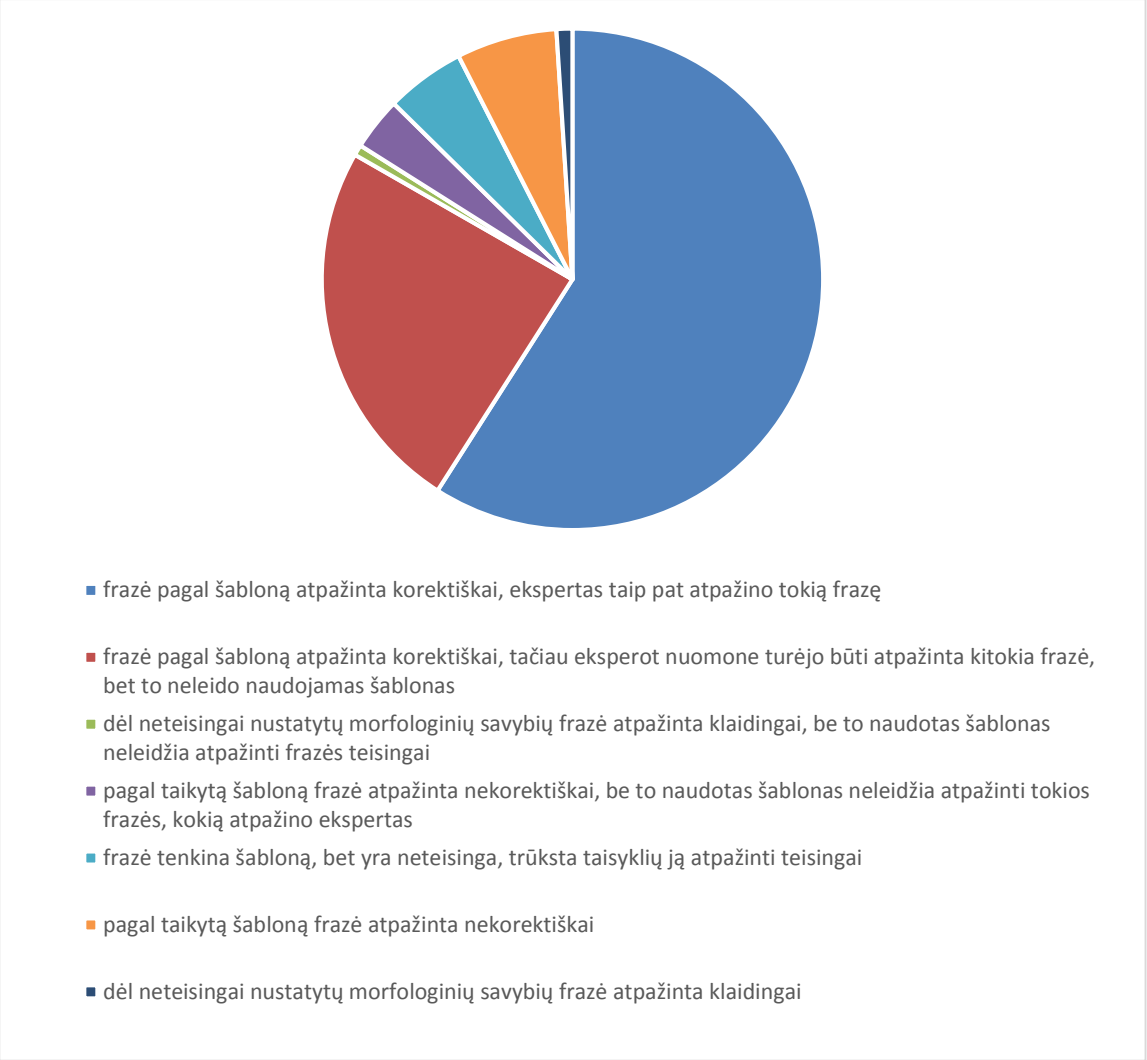


57 Pav. Sukurtos sistemos rastų frazių palyginimas su eksperto rastomis frazėmis

Iš 55pav. matyti jog sukurta sistema surado daugiau nei pusę, t.y. 173 frazes kurias surado ir ekspertas. Neatpažintų frazių pasitaikė tik 12, t.y. ekspertas rado tačiau sistema nerado. Nekorektiškai rastų frazių skaičius siekia 108, taip gali nutikti dėl nepilnų ar netinkamų šablonų kurie buvo sudaryti sistemai.

Žemiau pateiktoje diagramoje mes sukurtos sistemos rastas frazes išskaidėme į kitus septynis kriterijus, pagal kuriuos galime matyti sistemos rastų frazių tikslumą žiūrint pagal eksperto rastas frazes. Iškelti kriterijai:

- frazė pagal šabloną atpažinta korektiškai, ekspertas taip pat atpažino tokią frazę
- frazė pagal šabloną atpažinta korektiškai, tačiau eksperto nuomone turėjo būti atpažinta kitokia frazė, bet to neleido naudojamas šablonas
- dėl neteisingai nustatytų morfologinių savybių frazė atpažinta klaidingai, be to naudotas šablonas neleidžia atpažinti frazės teisingai
- pagal taikytą šabloną frazė atpažinta nekorektiškai, be to naudotas šablonas neleidžia atpažinti tokios frazės, kokią atpažino ekspertas
- frazė tenkina šabloną, bet yra neteisinga, trūksta taisyklių ją atpažinti teisingai
- pagal taikytą šabloną frazė atpažinta nekorektiškai
- dėl neteisingai nustatytų morfologinių savybių frazė atpažinta klaidingai



58 Pav. Sukurtos sistemos atpažintų frazių detalesnis pjūvis

36 Lentelė Sistemai keltų kriterijų sąrašas bei rezultatai

Kriterijus	
frazė pagal šabloną atpažinta korektiškai, ekspertas taip pat atpažino tokią frazę	173
frazė pagal šabloną atpažinta korektiškai, tačiau eksperto nuomone turėjo būti atpažinta kitokia frazė, bet to neleido naudojamas šablonas	71
dėl neteisingai nustatytų morfologinių savybių frazė atpažinta klaidingai, be to naudotas šablonas neleidžia atpažinti frazės teisingai	2
pagal taikytą šabloną frazė atpažinta nekorektiškai, be to naudotas šablonas neleidžia atpažinti tokios frazės, kokią atpažino ekspertas	10
frazė tenkina šabloną, bet yra neteisinga, trūksta taisyklių ją atpažinti teisingai	15
pagal taikytą šabloną frazė atpažinta nekorektiškai	19
dėl neteisingai nustatytų morfologinių savybių frazė atpažinta klaidingai	3

Kaip matyti ir iš aukščiau esančios (36 Lentelė Sistemai keltų kriterijų sąrašas bei rezultatai) lentelės bei 56pav. jog sistemos rastos frazės ir eksperto rasto frazės daugiau nei pusė sutampa lyginant su visomis rastomis frazėmis. Pirmąjį kriterijų, t.y. *frazė pagal šabloną atpažinta korektiškai, ekspertas taip pat atpažino tokią frazę* atitiko 173 frazės. Toliau kitos sistemos frazės išsibarstė į kitus kriterijus.

Sistemos frazių radimo korektiškumą galima padidinti sudarant tinkamesnius šablonus. Sudarant šablonus būtina remtis lietuvių kalbos žiniomis, norint patikslinti šablono dedamąsias. Dėl neteisingų, ar nepilnų sistemos šablonų 71 frazė buvo rasta nepilna, lyginant eksperto. Tokių rastų frazių pavyzdžiai pateikti žemiau esančioje (37 Lentelė Nepilnų frazių lyginant su eksperto sąrašas) lentelėje:

37 Lentelė Nepilnų frazių lyginant su eksperto sąrašas

Sistemos rasta frazė	*Eksperto rasta frazė*
švietimo	Šiaulių švietimo darbuotojų
žadama	žadama išmokėti
paramos ir labdaros fondo Kalėdų šventėje	surengtoje Almos Adamkienės paramos ir labdaros fondo Kalėdų šventėje
pirmoji šalies ponia	buvusi pirmoji šalies ponia
prezidento	prezidento Valdo Adamkaus žmona Alma

Norint pagerinti automatinį sistemos frazių radimą, reikia sudaryti tinkamus bei gerus šablonus. Esant netinkamiems, ar ne pilniems šablonams, aptikti reikiamas frazes sudėtinga. Norint sudaryti gerus šablonus reikia gerai išmanyti Lietuvių kalbos gramatiką, žinoti pagrindines daiktavardinių bei veiksmažodinių frazių sudarinėjimo taisykles.

Taipogi sukurtos sistemos nepilnų frazių radimo aspektas yra nepilnas sistemos išbaigtumas. Algoritmų nepilnumas, ne visų taisyklių įgyvendinimas. Aptikti sudėtingesnės struktūros frazės yra ypač sudėtinga, kadangi sunku sudaryti jiems šablonus.

7. IŠVADOS

1) Atlikus panašių sistemų analizę, buvo pastebėta, jog nei viena iš analizuotų sistemų netinka Lietuviško teksto analizei. Taipogi nei viena iš sistemų neturėjo galimybės pačiam vartotojui nustatyti ieškomų frazių, junginių ar kitokių struktūrų žodžių, sudaryti numanomus šablonus pagal kuriuos sistema galėtų rasti bei aptikti frazes ar junginius.
2) Atlikta analizė leidžia teigti, jog frazių ar žodžių junginių radimas tekstuose ar žodynų sudarinėjimas yra gan sudėtingas ir laiko reikalaujantis mechanizmas.
3) Iš atliktų teminės srities žodynų sudarinėtojų programų analizės galima teigti, jog Lietuviško teksto žodynų sudarymas gali būti gan sudėtingas. Norėdami palengvinti frazių radimą bei pačios programos kūrimą galima bus remtis, jau esamomis kitoms kalboms skirtomis funkcijomis bei paieškos metodais.
4) Nustačius pagrindinius tikslus, algoritmo funkcionalumus ir prototipo viziją buvo nustatytos pagrindinės algoritmo veiklos detalės, suprojektuotas ir atvaizduotas algoritmo funkcionalumas grafiškai bei aprašyti algoritmo veikimo principai.
5) Projektavimo metu buvo apibrėžti pagrindiniai algoritmo prototipo reikalavimai. Sukurtos taisyklės pagal kurias algoritmas vykdo numatytus funkcionalumus. Taip pat aprašyti bei atvaizduoti algoritmo veikimo žingsniai kuriais remiantis algoritmas pateikia rezultatus.
6) Atlikus eksperimentą buvo nustatyta jog sistema gali aptikti daugiau nei pusę frazių lyginant su eksperto rastomis. Eksperimento metu buvo prieita išvados jog norint geriau ir tinkamiau rasti frazes, reikia geriau sudaryti šablonus, atitinkamai parinkti tinkamesnius šablono elementus.
7) Remiantis darbo analizės, projektavimo ir algoritmo aprašymo medžiaga galime teigti, jog turime visapusišką įrankio prototipą, kuris gali pagreitinti bei padėti lengviau ir tiksliau rasti daiktvardines ir veiksmažodines frazes lietuviškame tekste. Algoritmas žymiai paspartina bei pagerina SBVR žodynų sudarymo eigą, taipogi pagerintų šių žodynų kokybę bei tikslumą.
8) Programos realizavimas parodė, jog frazių paieškos algoritmą naudojant šablonus įmanoma sukurti, o testavimas - , kad sukurtas bei realizuotas algoritmas veikia tinkamai, nes tenkina iškeltus reikalavimus.

8. LITERATŪRA

[1] VDU, „http://tekstynas.vdu.lt/,“ 1998-2013. [Tinkle]. Available: http://tekstynas.vdu.lt/page.xhtml;jsessionid=5A6DA6C0EC9EAD7A1CAFD94713D3D9F8?id=morphological-annotator. [Kreiptasi 13 balandžio 2014].

[2] „dz.lki.lt,“ 2012-2013. [Tinkle]. Available: http://dz.lki.lt. [Kreiptasi 09 balandžio 2014].

[3] „Gerpkime žodį,“ [Tinkle]. Available: http://ualgiman.dtiltas.lt/leksikografija_ir_zodynu_tipai.html. [Kreiptasi 09 balandžio 2014].

[4] „Gerpkime žodį,“ [Tinkle]. Available: http://ualgiman.dtiltas.lt/leksikografija_ir_zodynu_tipai.html. [Kreiptasi 09 balandžio 2014].

[5] Labutis, Labutis, 1998.

[6] SDL, „http://www.translationzone.com/,“ SDL, 2011. [Tinkle]. Available: http://www.translationzone.com/products/sdl-multiterm/desktop/. [Kreiptasi 13 balandžio 2014].

[7] Terminotix, „http://www.terminotix.com/,“ Terminotix, [Tinkle]. Available: http://www.terminotix.com/index.asp?name=SynchroTerm&content=item&brand=4&item=7&lang=en. [Kreiptasi 13 balandžio 2014].

[8] WordFast, „http://www.wordfast.net/,“ Wordfast LLC, [Tinkle]. Available: http://www.wordfast.net/index.php?whichpage=products&lang=engb. [Kreiptasi 13 balandžio 2014].

[9] R. J. B. a. B. K. B. Younga Park, „Automatic Glossary Extraction: Beyond Terminology Indentification,“ p. 7.

[10] E. Vyšniauskas ir L. Nemuraitė, „Transforming Ontology Representation from OWL to Relational Database,“ *Information Technology and Control*, t. 35A, nr. 3, p. 333–343, 2006.

[11] K. Masiulis ir A. Krupavičius, Valstybės tarnyba Lietuvoje: praeitis ir dabartis: kolektyvinė monografija, Vilnius: Praction, 2007, p. 430.

[12] V. Biržiška, „Spaudos draudimo klausimai,“ *Kultūra,* nr. 5, pp. 249-235, 1929.

[13] I. Valiulytė, „Išlaidos krašto apsaugai, jų pagrįstumas ir tikslingumas,“ vasaris 2000. [Tinkle]. Available: http://www.sociumas.lt. [Kreiptasi 12 gruodžio 2001].

[14] D. U. Library, „IEEE Citation style guide,“ 2009. [Tinkle]. Available: http://libraries.dal.ca/content/dam/dalhousie/pdf/library/Style_Guides/IEEE_Citation_Style_Guide.pdf. [Kreiptasi 11 04 2013].

[15] R. Gradauskas, „Hibridinis velomobilis,“ įtraukta *Transporto priemonės - 99*, Kaunas, 2000.

9. PRIEDAI

9.1. 1 priedas. Sistemos rastos frazės

Phrase	*StructureName*	*TextName*
10 ame	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
10 procentų	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
12	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
13	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
16 kvadratinių metrų kambarys	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
17	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
250	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
41 metų	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
58 metų vyras	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
6 kvadratinius metrus	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
6240 litų baudą	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
76 metų namo savininkė	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
7800 litų bauda	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
A.Adamkienei	Vardai pavardės	bals-Lietuva-091220-111.txt
A.Pukelio	Vardai pavardės	bals-Lietuva-091203-94.txt
A.Pukelis	Vardai pavardės	bals-Lietuva-091203-94.txt
aiškino	V[Prl]	bals-Lietuva-091124-85.txt
algų	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Alma Adamkienė	Vardai pavardės	bals-Lietuva-091220-111.txt
Almos Adamkienės	Vardai pavardės	bals-Lietuva-091220-111.txt
AP.	Vardai pavardės	bals-Pasaulis-090907-2.txt
Apygardos teismą	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
Apygardos teismo teisėjų kolegija	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
apylinkės	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
apspjovęs	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
Arūną Pukelį	Vardai pavardės	bals-Lietuva-091203-94.txt
asilu	DVbendr	bals-Lietuva-091203-94.txt
atlikta operacija	Dal[b*][D*](k)D	bals-Lietuva-091220-111.txt
atlyginimą	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Atlyginimus	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
atostogų	DVbendr	bals-Lietuva-091213-104.txt
atsakomybės	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
atsisakė	V[Prl]	bals-Pasaulis-090907-2.txt
atsisakymą pateikti	DVbendr	bals-Pasaulis-090907-2.txt
atsiųstas ženklas	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
atstovų	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
atšventusio tauragiškio mašiną	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
atvira liepsna	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Audi" Švinius	Vardai pavardės	bals-Lietuva-091203-94.txt
aukšto	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
avansus	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
baigęs prezidentas	Dal[b*][D*](k)D	bals-Lietuva-091220-111.txt
baltas	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
batus	DVbendr	bals-Pasaulis-090907-2.txt

biudžetinėms įstaigoms	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
biurokratinių reikalavimų	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
būtina	Dal[b*][D*](k)D	bals-Pasaulis-090907-2.txt
buvę baldai	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
buvo	V[Prl]	bals-Lietuva-091203-94.txt
buvusi	Dal[b*][D*](k)D	bals-Lietuva-091220-111.txt
Buvusio kandidato teigimu	Dal[b*][D*](k)D	bals-Pasaulis-090907-2.txt
dabartiniu Niujorko meru	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
darbuotojos	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
darželio problema	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
degusio	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
delspinigių skaičiavimo	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
didesnių problemų	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
dirbs be	V[Prl]	bals-Lietuva-091213-104.txt
dolerių bauda	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
du metrus	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
dviejų nepilnamečių valdininko vaikų akivaizdoje	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
finansinę padėtį	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
gaisrą	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Gaisrininkai	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Gaisro metu	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
gali pasiguosti	VVbendr[prl]	bals-Lietuva-091203-94.txt
gatvės	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Gatvės muzikantas	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
gavęs	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
gavo	V[Prl]	bals-Lietuva-091124-85.txt
gavo	V[Prl]	bals-Lietuva-091213-104.txt
gyvenimo	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
grasinimas sudeginti	DVbendr	bals-Lietuva-091203-94.txt
grasinimus	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
gruodį	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
guli	V[Prl]	bals-Lietuva-091220-111.txt
ir gėjumi	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
ir po	V[Prl]	bals-Lietuva-091124-85.txt
ir skrybėlę	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
ir visuomeninio transporto naujovių įgyvendinimo	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
išdegė	V[Prl]	bals-Lietuva-091124-85.txt
išgelbėjo per	V[Prl]	bals-Lietuva-091124-85.txt
iškoneveikė	V[Prl]	bals-Lietuva-091203-94.txt
išnaudotas metų mokos fondas	Dal[b*][D*](k)D	bals-Lietuva-091213-104.txt
išpuolio viešoje	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
išteisintas	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
išvadino	V[Prl]	bals-Lietuva-091203-94.txt
išvakarėse	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
išvežtas	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
įspūdžio	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
įstaigų vadovėms	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt

įtakos vietos politikams	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
įteikė	V[Prl]	bals-Lietuva-091213-104.txt
yra	V[Prl]	bals-Lietuva-091203-94.txt
yra	V[Prl]	bals-Pasaulis-090907-2.txt
yra apie	V[Prl]	bals-Lietuva-091213-104.txt
jaučiasi	V[Prl]	bals-Lietuva-091220-111.txt
JAV	Sutrumpintos frazės	bals-Pasaulis-090907-2.txt
jubiliejų	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
kadenciją	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
kaltinusį prokurorą	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
kambario	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
kambaryje	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
kandidatūrą	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
kartus	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
kaubojiškus batus	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
ketina konkuruoti su	VVbendr[prl]	bals-Pasaulis-090907-2.txt
ketinimų įgyvendinti	DVbendr	bals-Lietuva-091124-85.txt
kilęs gaisras	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
kilo nuo	V[Prl]	bals-Lietuva-091124-85.txt
konstatavo	V[Prl]	bals-Lietuva-091203-94.txt
kreipėsi	V[Prl]	bals-Lietuva-091203-94.txt
kreipėsi į	V[Prl]	bals-Lietuva-091213-104.txt
lapkritį	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
lašu anot	V[Prl]	bals-Pasaulis-090907-2.txt
liepos mėnesį	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
liepsnas	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
ligoninėje	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
ligoninę	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
likimo	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
liko be	V[Prl]	bals-Lietuva-091213-104.txt
lyties	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
LNK Žinios.	Vardai pavardės	bals-Lietuva-091220-111.txt
Maiklu Blumbergu	Vardai pavardės	bals-Pasaulis-090907-2.txt
Manoma	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
mažesnėmis galimybėmis	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
mediniame gyvenamajame name	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Mercedes Benz" Tauragėje	Vardai pavardės	bals-Lietuva-091203-94.txt
mero postą	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
metu	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
metų birželio pradžioje	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
metų viduryje	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
miesto biudžete	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Miesto darželių vedėjos	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
miesto vadovo rinkimuose	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
mirė	V[Prl]	bals-Lietuva-091124-85.txt
mokesčių	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
mokos fondus	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
namas	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt

naujienų agentūra	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
neabejoja	V[Prl]	bals-Lietuva-091203-94.txt
nebeliko	V[Prl]	bals-Lietuva-091213-104.txt
nebesitiki	V[Prl]	bals-Lietuva-091213-104.txt
nebuvo	V[Prl]	bals-Lietuva-091124-85.txt
necenzūriniais žodžiais	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
nemokamų atlyginimų	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
nemokamų atostogų	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Neoficialiais duomenimis	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
nepadarė	V[Prl]	bals-Lietuva-091203-94.txt
nepadarė	V[Prl]	bals-Pasaulis-090907-2.txt
nepasirodė	V[Prl]	bals-Lietuva-091220-111.txt
nepavyko išvengti	VVbendr[prl]	bals-Lietuva-091203-94.txt
nerealus ir pareigūnams	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
Nerijui Gudui	Vardai pavardės	bals-Lietuva-091124-85.txt
nesirinko	V[Prl]	bals-Lietuva-091213-104.txt
neskirs be to	V[Prl]	bals-Lietuva-091213-104.txt
nespėjo	V[Prl]	bals-Lietuva-091124-85.txt
nesusitaupė	V[Prl]	bals-Lietuva-091213-104.txt
nesuspėjo	V[Prl]	bals-Lietuva-091124-85.txt
neužgesintos nuorūkos vyras	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
Niekas	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
Niujorko gatvėse	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
Niujorko mero rinkimuose	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
norėjęs	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
nuardė	V[Prl]	bals-Lietuva-091124-85.txt
nuimta	Dal[b*][D*](k)D	bals-Lietuva-091213-104.txt
numestos	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
Nuogas kaubojus	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
nusikaltimais	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
nusikaltimą	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
nusikaltimo	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
nusiminęs	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
nusprendė pasitraukti iš	VVbendr[prl]	bals-Lietuva-091124-85.txt
nusprendęs	Dal[b*][D*](k)D	bals-Pasaulis-090907-2.txt
nustebino	V[Prl]	bals-Lietuva-091203-94.txt
nutarė	V[Prl]	bals-Pasaulis-090907-2.txt
nuteistas	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
nutrūko	V[Prl]	bals-Lietuva-091124-85.txt
pabandžiusį	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
pagrindo	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
paguldyta	Dal[b*][D*](k)D	bals-Lietuva-091220-111.txt
pakeitė	V[Prl]	bals-Lietuva-091124-85.txt
palaikantis prokuroras	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
pamatė iš	V[Prl]	bals-Lietuva-091124-85.txt
paramos ir labdaros fondo Kalėdų šventėje	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
pareigybines instrukcijas	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
Paryžiaus	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt

parodė	V[Prl]	bals-Lietuva-091124-85.txt
pasibaigusios liūdnos istorijos pradžia	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
pasirinko	V[Prl]	bals-Pasaulis-090907-2.txt
pasirodo	V[Prl]	bals-Pasaulis-090907-2.txt
pasiūlymų	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
paskelbtas Klaipėdos apygardos teismo sprendimas	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
Pastaruoju sprendimu	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
pastatas	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
patikino	V[Prl]	bals-Lietuva-091203-94.txt
pavadavo	V[Prl]	bals-Lietuva-091220-111.txt
pavargo nuo	V[Prl]	bals-Pasaulis-090907-2.txt
pažymą	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
pinigų	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
pinigų savivaldybė	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
pirmoji šalies ponia	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
pirmoms gaisrinėms	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Planus nusižudyti	DVbendr	bals-Lietuva-091124-85.txt
planų dalyvauti	DVbendr	bals-Pasaulis-090907-2.txt
planų įgyvendinti	DVbendr	bals-Lietuva-091124-85.txt
policininkams	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
prabangų visureigį	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
pradėjo degti	VVbendr[prl]	bals-Lietuva-091124-85.txt
praneša	V[Prl]	bals-Pasaulis-090907-2.txt
pranešė	V[Prl]	bals-Lietuva-091220-111.txt
pranešė	V[Prl]	bals-Pasaulis-090907-2.txt
pranešimą	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
pravarde	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
prezidento	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
priėmė	V[Prl]	bals-Pasaulis-090907-2.txt
priešgaisrinė gelbėjimo tarnyba	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
priešgaisrinės gelbėjimo tarnybos	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
priešgaisrinės priežiūros poskyrio viršininkui	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
privalės sumokėti už	VVbendr[prl]	bals-Pasaulis-090907-2.txt
profesinių sąjungų susivienijimas	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
R. A. Velionės	Vardai pavardės	bals-Lietuva-091124-85.txt
R.Matemaitį Švinius	Vardai pavardės	bals-Lietuva-091203-94.txt
Raimondą Matemaitį.	Vardai pavardės	bals-Lietuva-091203-94.txt
rankos	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
reikalavimus	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Rinkimams R.Dž.	Vardai pavardės	bals-Pasaulis-090907-2.txt
Robertas Džonas Berkas	Vardai pavardės	bals-Pasaulis-090907-2.txt
rudenį	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
sako	V[Prl]	bals-Lietuva-091213-104.txt
santuokų	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
saugumo	DVbendr	bals-Pasaulis-090907-2.txt
sausį	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt

savivaldybės administraciją	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
sekmadienio popietė	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
sekmadienį	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
sekmadienį	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
senutėle	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
sirgusi	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
skirtos bausmės	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
skundais	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
sprendimą	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
stogo dangos	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
stovintį medinį namą	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
sugriebė	V[Prl]	bals-Lietuva-091203-94.txt
sumažino	V[Prl]	bals-Lietuva-091213-104.txt
Suprato	V[Prl]	bals-Lietuva-091124-85.txt
surengtoje Almos	Dal[b*][D*](k)D	bals-Lietuva-091220-111.txt
Susirado	V[Prl]	bals-Lietuva-091124-85.txt
sustabdė be	V[Prl]	bals-Lietuva-091203-94.txt
sustabdžiusio patrulio	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
sutaupė	V[Prl]	bals-Lietuva-091203-94.txt
svarstoma	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
šarvojimo salę	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Šeštadienį Prezidentūroje	Vardai pavardės	bals-Lietuva-091220-111.txt
Šiaulių miesto vaikų darželių auklėtojos	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
šūkį	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
šventės šeimininkė	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
švietimo	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Tačiau A.Pukeliui	Vardai pavardės	bals-Lietuva-091203-94.txt
tapo	V[Prl]	bals-Pasaulis-090907-2.txt
Tauragės apskrities viršininko pavaduotoją	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
tauragiškio automobilį	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
teigė	V[Prl]	bals-Pasaulis-090907-2.txt
teigia	V[Prl]	bals-Lietuva-091213-104.txt
teistą	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
tituluojamas nuogu kaubojumi	Dal[b*][D*](k)D	bals-Pasaulis-090907-2.txt
trumpikes	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
tūkstantį litų	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
tuometį rajono vicemerą	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
turizmo	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
turįs	Dal[b*][D*](k)D	bals-Pasaulis-090907-2.txt
ugniagesiai	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
ūkinius pastatus	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
Užblokavęs kelią	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
V. A. Biržų	Vardai pavardės	bals-Lietuva-091124-85.txt
V.Adamkus.	Vardai pavardės	bals-Lietuva-091220-111.txt
Vabalninke	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
Vakar Apygardos	Vardai pavardės	bals-Lietuva-091203-94.txt
Valdo Adamkaus	Vardai pavardės	bals-Lietuva-091220-111.txt

Valstybinės priešgaisrinės priežiūros	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
valstybinį kaltinimą	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
vasarį	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
Velionės kūnas	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
vertinamas Tauragės verslininkas	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
Vidutinis ikimokyklinių įstaigų darbuotojų atlyginimas	[sk][b*][d*](k)D	bals-Lietuva-091213-104.txt
viešojo saugumo	[sk][b*][d*](k)D	bals-Pasaulis-090907-2.txt
viešosios tvarkos pažeidimo	[sk][b*][d*](k)D	bals-Lietuva-091203-94.txt
virstančius dūmus	Dal[b*][D*](k)D	bals-Lietuva-091124-85.txt
virvė	DVbendr	bals-Lietuva-091124-85.txt
virvę	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
žadama	Dal[b*][D*](k)D	bals-Lietuva-091213-104.txt
žinomą Arūną	Dal[b*][D*](k)D	bals-Lietuva-091203-94.txt
žmona	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt
žmonai	[sk][b*][d*](k)D	bals-Lietuva-091124-85.txt
žmoną	[sk][b*][d*](k)D	bals-Lietuva-091220-111.txt

9.2. 2 priedas. Eilinio vartotojo rastos frazės

Phrase	***Text***
gaisras	bals-Lietuva-091124-85.txt
Kilo	bals-Lietuva-091124-85.txt
vyras	bals-Lietuva-091124-85.txt
gyvenimo	bals-Lietuva-091124-85.txt
Virve	bals-Lietuva-091124-85.txt
šarvojimo sale	bals-Lietuva-091124-85.txt
Valstybinės priešgaisrinės priežiūros poskyrio viršininkui	bals-Lietuva-091124-85.txt
Nerijui Gudui	bals-Lietuva-091124-85.txt
norėjo pasikarti	bals-Lietuva-091124-85.txt
ženklas	bals-Lietuva-091124-85.txt
Klaipėdos apygardos teismo	bals-Lietuva-091203-94.txt
Švinius	bals-Lietuva-091203-94.txt
sutaupė	bals-Lietuva-091203-94.txt
pinigų	bals-Lietuva-091203-94.txt
Arūną Pukelį	bals-Lietuva-091203-94.txt
kaltinusį	bals-Lietuva-091203-94.txt
prokurorą	bals-Lietuva-091203-94.txt
viešosios tvarkos	bals-Lietuva-091203-94.txt
A.Pukelis	bals-Lietuva-091203-94.txt
Tauragės verslininkas	bals-Lietuva-091203-94.txt
Tauragės apskrities viršininko pavaduotoją	bals-Lietuva-091203-94.txt
Raimondą Matemaitį	bals-Lietuva-091203-94.txt
necenzūriniais žodžiais	bals-Lietuva-091203-94.txt
6240 litų baudą	bals-Lietuva-091203-94.txt
Auklėtojos	bals-Lietuva-091213-104.txt
liko	bals-Lietuva-091213-104.txt
atlyginimo	bals-Lietuva-091213-104.txt

Šiaulių ikimokyklinių įstaigų darbuotojos	bals-Lietuva-091213-104.txt
avansus	bals-Lietuva-091213-104.txt
nebesitiki	bals-Lietuva-091213-104.txt
darželio problema	bals-Lietuva-091213-104.txt
savivaldybė	bals-Lietuva-091213-104.txt
pinigų	bals-Lietuva-091213-104.txt
neskirs	bals-Lietuva-091213-104.txt
nesirinko	bals-Lietuva-091213-104.txt
nemokamų atostogų	bals-Lietuva-091213-104.txt
nesusitaupė	bals-Lietuva-091213-104.txt
pinigų	bals-Lietuva-091213-104.txt
įstaigoms mokos fondus	bals-Lietuva-091213-104.txt
sumažino	bals-Lietuva-091213-104.txt
Šiaulių švietimo darbuotojų profesinių sąjungų susivienijimas	bals-Lietuva-091213-104.txt
ikimokyklinių įstaigų darbuotojų atlyginimas	bals-Lietuva-091213-104.txt
delspinigių	bals-Lietuva-091213-104.txt
Alma Adamkienė	bals-Lietuva-091220-111.txt
paguldyta	bals-Lietuva-091220-111.txt
ligoninę	bals-Lietuva-091220-111.txt
Prezidentūroje	bals-Lietuva-091220-111.txt
labdaros fondo	bals-Lietuva-091220-111.txt
Kalėdų šventėje	bals-Lietuva-091220-111.txt
nepasirodė	bals-Lietuva-091220-111.txt
šventės šeimininkė	bals-Lietuva-091220-111.txt
atlikta	bals-Lietuva-091220-111.txt
operacija	bals-Lietuva-091220-111.txt
pirmoji šalies ponia	bals-Lietuva-091220-111.txt
Nuogas kaubojus	bals-Pasaulis-090907-2.txt
Niujorko mero rinkimuose	bals-Pasaulis-090907-2.txt
nutarė nedalyvauti	bals-Pasaulis-090907-2.txt
Gatvės muzikantas	bals-Pasaulis-090907-2.txt
Robertas Džonas Berkas	bals-Pasaulis-090907-2.txt
nuogu kaubojumi	bals-Pasaulis-090907-2.txt
Buvusio kandidato	bals-Pasaulis-090907-2.txt
biurokratinių reikalavimų	bals-Pasaulis-090907-2.txt
finansinę padėtį	bals-Pasaulis-090907-2.txt
baltas trumpikes,	bals-Pasaulis-090907-2.txt
kaubojiškus batus	bals-Pasaulis-090907-2.txt
skrybėlę	bals-Pasaulis-090907-2.txt
Maiklu Blumbergu	bals-Pasaulis-090907-2.txt
Niekas nepadarė daugiau su mažesnėmis galimybėmis	bals-Pasaulis-090907-2.txt
miesto vadovo rinkimuose	bals-Pasaulis-090907-2.txt
mokesčių	bals-Pasaulis-090907-2.txt
turizmo	bals-Pasaulis-090907-2.txt
vienos lyties atstovų santuokų	bals-Pasaulis-090907-2.txt
viešojo saugumo	bals-Pasaulis-090907-2.txt
visuomeninio transporto naujovių įgyvendinimo	bals-Pasaulis-090907-2.txt

9.3. 3 priedas. Eksperto rastos frazės

Phrase	
250	bals-Pasaulis-090907-2.txt
vienos lyties atstovų santuokų	bals-Pasaulis-090907-2.txt
dėvėdamas baltas trumpikes	bals-Pasaulis-090907-2.txt
biurokratinių reikalavimų	bals-Pasaulis-090907-2.txt
dabartiniu Niujorko meru	bals-Pasaulis-090907-2.txt
dolerių bauda	bals-Pasaulis-090907-2.txt
atsisakymą pateikti pažymą apie jo finansinę padėtį	bals-Pasaulis-090907-2.txt
Gatvės muzikantas	bals-Pasaulis-090907-2.txt
ir skrybėlę	bals-Pasaulis-090907-2.txt
ir visuomeninio transporto naujovių įgyvendinimo	bals-Pasaulis-090907-2.txt
kandidatūrą	bals-Pasaulis-090907-2.txt
kaubojiškus batus	bals-Pasaulis-090907-2.txt
liepos mėnesį	bals-Pasaulis-090907-2.txt
mažesnėmis galimybėmis	bals-Pasaulis-090907-2.txt
mero postą	bals-Pasaulis-090907-2.txt
miesto vadovo rinkimuose	bals-Pasaulis-090907-2.txt
pasiūlymų dėl mokesčių,	bals-Pasaulis-090907-2.txt
naujienų agentūra	bals-Pasaulis-090907-2.txt
Niekas	bals-Pasaulis-090907-2.txt
Niujorko gatvėse	bals-Pasaulis-090907-2.txt
Niujorko mero rinkimuose	bals-Pasaulis-090907-2.txt
Nuogas kaubojus	bals-Pasaulis-090907-2.txt
pažymą	bals-Pasaulis-090907-2.txt
sekmadienį	bals-Pasaulis-090907-2.txt
sprendimą	bals-Pasaulis-090907-2.txt
šūkį	bals-Pasaulis-090907-2.txt
turizmo	bals-Pasaulis-090907-2.txt
viešojo saugumo	bals-Pasaulis-090907-2.txt
būtina įvykdyti	bals-Pasaulis-090907-2.txt
Buvusio kandidato teigimu	bals-Pasaulis-090907-2.txt
nusprendęs kandidatuoti mero rinkimuose	bals-Pasaulis-090907-2.txt
tituluojamas nuogu kaubojumi	bals-Pasaulis-090907-2.txt
turįs daug pasiūlymų dėl mokesčių	bals-Pasaulis-090907-2.txt
atsisakymą pateikti pažymą apie jo finansinę padėtį	bals-Pasaulis-090907-2.txt
planų dalyvauti	bals-Pasaulis-090907-2.txt
JAV	bals-Pasaulis-090907-2.txt
atsisakė	bals-Pasaulis-090907-2.txt
yra	bals-Pasaulis-090907-2.txt
Paskutiniu lašu	bals-Pasaulis-090907-2.txt
nepadarė	bals-Pasaulis-090907-2.txt
nutarė nedalyvaut	bals-Pasaulis-090907-2.txt
pasirinko	bals-Pasaulis-090907-2.txt
pasirodo	bals-Pasaulis-090907-2.txt
pavargo nuo	bals-Pasaulis-090907-2.txt
praneša	bals-Pasaulis-090907-2.txt
pranešė	bals-Pasaulis-090907-2.txt

priėmė	bals-Pasaulis-090907-2.txt
tapo	bals-Pasaulis-090907-2.txt
teigė	bals-Pasaulis-090907-2.txt
AP.	bals-Pasaulis-090907-2.txt
Maiklu Blumbergu	bals-Pasaulis-090907-2.txt
Rinkimams	bals-Pasaulis-090907-2.txt
Robertas Džonas Berkas	bals-Pasaulis-090907-2.txt
ketina konkuruoti su	bals-Pasaulis-090907-2.txt
privalės sumokėti už	bals-Pasaulis-090907-2.txt
R.Dž.Berkas	bals-Pasaulis-090907-2.txt
R.Dž.Berko	bals-Pasaulis-090907-2.txt
16 kvadratinių metrų kambarys	bals-Lietuva-091124-85.txt
17.13 val.	bals-Lietuva-091124-85.txt
58 metų vyras V. A.	bals-Lietuva-091124-85.txt
6 kvadratinius metrus stogo dangos	bals-Lietuva-091124-85.txt
Aiškino	bals-Lietuva-091124-85.txt
atvira liepsna	bals-Lietuva-091124-85.txt
Biržų priešgaisrinė gelbėjimo tarnyba	bals-Lietuva-091124-85.txt
Biržų priešgaisrinės gelbėjimo tarnybos Valstybinės priešgaisrinės priežiūros poskyrio viršininkui Nerijui Gudui	bals-Lietuva-091124-85.txt
buvę baldai	bals-Lietuva-091124-85.txt
buvo išvežtas į	bals-Lietuva-091124-85.txt
buvo pasiruošęs.	bals-Lietuva-091124-85.txt
Buvo taip nusiminęs	bals-Lietuva-091124-85.txt
degė	bals-Lietuva-091124-85.txt
per du metrus nuo degusio namo stovintį medinį namą.	bals-Lietuva-091124-85.txt
Gaisras	bals-Lietuva-091124-85.txt
Gaisrininkai	bals-Lietuva-091124-85.txt
Gaisro metu	bals-Lietuva-091124-85.txt
gavo	bals-Lietuva-091124-85.txt
gyvenimo	bals-Lietuva-091124-85.txt
iš kambario virstančius dūmus	bals-Lietuva-091124-85.txt
Išdegė	bals-Lietuva-091124-85.txt
Išgelbėjo	bals-Lietuva-091124-85.txt
įgyvendinti nespėjo	bals-Lietuva-091124-85.txt
įgyvendinti nesuspėjo	bals-Lietuva-091124-85.txt
kambaryje	bals-Lietuva-091124-85.txt
Ketinimų	bals-Lietuva-091124-85.txt
kilęs gaisras	bals-Lietuva-091124-85.txt
Kilo	bals-Lietuva-091124-85.txt
kilo nuo	bals-Lietuva-091124-85.txt
liepsnas	bals-Lietuva-091124-85.txt
likimo atsiųstas ženklas	bals-Lietuva-091124-85.txt
Manoma	bals-Lietuva-091124-85.txt
mirė	bals-Lietuva-091124-85.txt
namas	bals-Lietuva-091124-85.txt
nebuvo	bals-Lietuva-091124-85.txt

norėjęs pasikarti	bals-Lietuva-091124-85.txt
nuardė	bals-Lietuva-091124-85.txt
numestos neužgesintos nuorūkos	bals-Lietuva-091124-85.txt
nusprendė pasitraukti iš	bals-Lietuva-091124-85.txt
nutrūko	bals-Lietuva-091124-85.txt
Pakeitė	bals-Lietuva-091124-85.txt
pamatė	bals-Lietuva-091124-85.txt
Paryžiaus gatvės 10 - ame name	bals-Lietuva-091124-85.txt
parodė	bals-Lietuva-091124-85.txt
pasibaigusios liūdnos istorijos pradžia	bals-Lietuva-091124-85.txt
pastatas	bals-Lietuva-091124-85.txt
pirmoms gaisrinėms	bals-Lietuva-091124-85.txt
Planus nusižudyti	bals-Lietuva-091124-85.txt
Planų	bals-Lietuva-091124-85.txt
pradėjo degti	bals-Lietuva-091124-85.txt
pranešimą apie gaisrą vieno aukšto mediniame gyvenamajame name	bals-Lietuva-091124-85.txt
rūko	bals-Lietuva-091124-85.txt
sekmadienio popietė	bals-Lietuva-091124-85.txt
sekmadienį	bals-Lietuva-091124-85.txt
sirgusi 76 metų namo savininkė R. A.	bals-Lietuva-091124-85.txt
Suprato	bals-Lietuva-091124-85.txt
Susirado	bals-Lietuva-091124-85.txt
Svarstoma	bals-Lietuva-091124-85.txt
šarvojimo salę	bals-Lietuva-091124-85.txt
ugniagesiai	bals-Lietuva-091124-85.txt
Vabalninke	bals-Lietuva-091124-85.txt
Velionės kūnas	bals-Lietuva-091124-85.txt
virvė	bals-Lietuva-091124-85.txt
Virvę	bals-Lietuva-091124-85.txt
vyras	bals-Lietuva-091124-85.txt
žmonai	bals-Lietuva-091124-85.txt
Pastaruoju sprendimu	bals-Lietuva-091203-94.txt
41 metų	bals-Lietuva-091203-94.txt
6240 litų baudą gavęs A.Pukelis	bals-Lietuva-091203-94.txt
A.Pukelio įtakos vietos politikams	bals-Lietuva-091203-94.txt
A.Pukelis	bals-Lietuva-091203-94.txt
A.Pukeliui	bals-Lietuva-091203-94.txt
Apygardos teismą	bals-Lietuva-091203-94.txt
Apygardos teismo teisėjų kolegija	bals-Lietuva-091203-94.txt
apspjovęs “Audi”	bals-Lietuva-091203-94.txt
asilu	bals-Lietuva-091203-94.txt
atsakomybės dėl išpuolio viešoje vietoje prieš Tauragės apskrities viršininko pavaduotoją Raimondą Matemaitį	bals-Lietuva-091203-94.txt
birželio 12 d.	bals-Lietuva-091203-94.txt
Birželį	bals-Lietuva-091203-94.txt
buvo	bals-Lietuva-091203-94.txt
buvo skirta	bals-Lietuva-091203-94.txt

dviejų nepilnamečių valdininko vaikų akivaizdoje	bals-Lietuva-091203-94.txt
gėjumi	bals-Lietuva-091203-94.txt
grasinimas sudeginti vieno tauragiškio automobilį	bals-Lietuva-091203-94.txt
sustabdžiusio patrulio ūkinius pastatus	
Išgelbėjo	bals-Lietuva-091203-94.txt
iškoneveikė	bals-Lietuva-091203-94.txt
išteisintas dėl viešosios tvarkos pažeidimo	bals-Lietuva-091203-94.txt
išvadino	bals-Lietuva-091203-94.txt
išvakarėse	bals-Lietuva-091203-94.txt
įspūdžio	bals-Lietuva-091203-94.txt
yra	bals-Lietuva-091203-94.txt
jubiliejų atšventusio tauragiškio mašiną	bals-Lietuva-091203-94.txt
kartus	bals-Lietuva-091203-94.txt
konstatavo priešingai	bals-Lietuva-091203-94.txt
kreipėsi	bals-Lietuva-091203-94.txt
Neabejoja	bals-Lietuva-091203-94.txt
necenzūriniais žodžiais	bals-Lietuva-091203-94.txt
nekaltas	bals-Lietuva-091203-94.txt
nepadarė	bals-Lietuva-091203-94.txt
nepavyko išvengti	bals-Lietuva-091203-94.txt
Neteko	bals-Lietuva-091203-94.txt
nuo didesnių problemų	bals-Lietuva-091203-94.txt
nusikaltimais kaltinusį prokurorą	bals-Lietuva-091203-94.txt
nusikaltimą	bals-Lietuva-091203-94.txt
nusikaltimo	bals-Lietuva-091203-94.txt
nustebino	bals-Lietuva-091203-94.txt
nuteistas 7800 litų bauda už grasinimus	bals-Lietuva-091203-94.txt
policininkams	
Paauklėjo	bals-Lietuva-091203-94.txt
pabandžiusį aplenkti tuometį rajono vicemerą	bals-Lietuva-091203-94.txt
R.Matemaitį	
pagrindo	bals-Lietuva-091203-94.txt
pareigybines instrukcijas.	bals-Lietuva-091203-94.txt
pareigūnams	bals-Lietuva-091203-94.txt
gali pasiguosti	bals-Lietuva-091203-94.txt
paskelbtas Klaipėdos apygardos teismo sprendimas	bals-Lietuva-091203-94.txt
Patikino	bals-Lietuva-091203-94.txt
per du mėnesius po nuosprendžio įsiteisėjimo	bals-Lietuva-091203-94.txt
Pinigus	bals-Lietuva-091203-94.txt
Pinigų	bals-Lietuva-091203-94.txt
pora tūkstančių litų mažesnė	bals-Lietuva-091203-94.txt
prabangų visureigį “Mercedes-Benz”	bals-Lietuva-091203-94.txt
prieštaringai vertinamas	bals-Lietuva-091203-94.txt
privalės sumokėti	bals-Lietuva-091203-94.txt
R.Matemaitį	bals-Lietuva-091203-94.txt
senutėle “Audi 80”	bals-Lietuva-091203-94.txt
smakro	bals-Lietuva-091203-94.txt
su skundais dėl šių metų birželio pradžioje Klaipėdos	bals-Lietuva-091203-94.txt
rajono apylinkės teismo skirtos bausmės	

Sugriebė	bals-Lietuva-091203-94.txt
sustabdė	bals-Lietuva-091203-94.txt
Sutaupė	bals-Lietuva-091203-94.txt
Švinius	bals-Lietuva-091203-94.txt
Tauragėje	bals-Lietuva-091203-94.txt
Tauragės verslininkas A.Pukelis	bals-Lietuva-091203-94.txt
teistą Šviniaus pravarde žinomą Arūną Pukelį	bals-Lietuva-091203-94.txt
už grasinimus policininkams	bals-Lietuva-091203-94.txt
už rankos	bals-Lietuva-091203-94.txt
Užblokavęs kelią verslininkas	bals-Lietuva-091203-94.txt
Vaikai	bals-Lietuva-091203-94.txt
Valstybei	bals-Lietuva-091203-94.txt
valstybinį kaltinimą palaikantis prokuroras.	bals-Lietuva-091203-94.txt
Vicemerą	bals-Lietuva-091203-94.txt
vicemero posto	bals-Lietuva-091203-94.txt
A.Adamkienei	bals-Lietuva-091220-111.txt
Alma Adamkienė	bals-Lietuva-091220-111.txt
atlikta operacija	bals-Lietuva-091220-111.txt
buvusi pirmoji šalies ponia	bals-Lietuva-091220-111.txt
esą	bals-Lietuva-091220-111.txt
guli	bals-Lietuva-091220-111.txt
jaučiasi	bals-Lietuva-091220-111.txt
kadenciją baigęs prezidentas V.Adamkus.	bals-Lietuva-091220-111.txt
ligoninėje	bals-Lietuva-091220-111.txt
ligoninę	bals-Lietuva-091220-111.txt
LNK Žinios.	bals-Lietuva-091220-111.txt
Neoficialiais duomenimis	bals-Lietuva-091220-111.txt
nepasirodė	bals-Lietuva-091220-111.txt
paguldyta į	bals-Lietuva-091220-111.txt
Pavadavo	bals-Lietuva-091220-111.txt
pranešė	bals-Lietuva-091220-111.txt
prezidento Valdo Adamkaus žmona Alma	bals-Lietuva-091220-111.txt
Prezidentūroje	bals-Lietuva-091220-111.txt
surengtoje Almos Adamkienės paramos ir labdaros fondo Kalėdų šventėje	bals-Lietuva-091220-111.txt
Šeštadienį	bals-Lietuva-091220-111.txt
Šventėje	bals-Lietuva-091220-111.txt
šventės šeimininkė	bals-Lietuva-091220-111.txt
Žmoną	bals-Lietuva-091220-111.txt
10 procentų	bals-Lietuva-091213-104.txt
Algų	bals-Lietuva-091213-104.txt
Atlyginimus už lapkritį	bals-Lietuva-091213-104.txt
atlyginimų	bals-Lietuva-091213-104.txt
Auklėtojos	bals-Lietuva-091213-104.txt
Avansus	bals-Lietuva-091213-104.txt
biudžetinėms įstaigoms	bals-Lietuva-091213-104.txt
Darželio problema	bals-Lietuva-091213-104.txt
darželių auklėtojos	bals-Lietuva-091213-104.txt
delspinigių skaičiavimo	bals-Lietuva-091213-104.txt

dirbs be	bals-Lietuva-091213-104.txt
Esą	bals-Lietuva-091213-104.txt
Gavo	bals-Lietuva-091213-104.txt
gruodį	bals-Lietuva-091213-104.txt
išnaudotas metų mokos fondas	bals-Lietuva-091213-104.txt
įstaigų vadovėms	bals-Lietuva-091213-104.txt
įteikė	bals-Lietuva-091213-104.txt
Yra	bals-Lietuva-091213-104.txt
yra apie	bals-Lietuva-091213-104.txt
kreipėsi į	bals-Lietuva-091213-104.txt
Lapkritį	bals-Lietuva-091213-104.txt
liko be	bals-Lietuva-091213-104.txt
Metu	bals-Lietuva-091213-104.txt
metų viduryje	bals-Lietuva-091213-104.txt
miesto biudžete	bals-Lietuva-091213-104.txt
Miesto darželių vedėjos	bals-Lietuva-091213-104.txt
mokos fondus	bals-Lietuva-091213-104.txt
nebeliko	bals-Lietuva-091213-104.txt
Nebesitiki	bals-Lietuva-091213-104.txt
nemokamų atlyginimų	bals-Lietuva-091213-104.txt
nemokamų atostogų	bals-Lietuva-091213-104.txt
nesirinko	bals-Lietuva-091213-104.txt
neskirs	bals-Lietuva-091213-104.txt
nesusitaupė	bals-Lietuva-091213-104.txt
nuimta po	bals-Lietuva-091213-104.txt
pinigų	bals-Lietuva-091213-104.txt
reikalavimus	bals-Lietuva-091213-104.txt
Rudenį	bals-Lietuva-091213-104.txt
Sako	bals-Lietuva-091213-104.txt
sausį	bals-Lietuva-091213-104.txt
Savivaldybė	bals-Lietuva-091213-104.txt
savivaldybės administraciją	bals-Lietuva-091213-104.txt
sumažino	bals-Lietuva-091213-104.txt
Šiaulių ikimokyklinių įstaigų darbuotojos	bals-Lietuva-091213-104.txt
Šiaulių miesto vaikų darželių auklėtojos	bals-Lietuva-091213-104.txt
Šiaulių švietimo darbuotojų profesinių sąjungų susivienijimas	bals-Lietuva-091213-104.txt
Teigia	bals-Lietuva-091213-104.txt
tūkstantį litų	bals-Lietuva-091213-104.txt
vasarį	bals-Lietuva-091213-104.txt
Vidutinis ikimokyklinių įstaigų darbuotojų atlyginimas	bals-Lietuva-091213-104.txt
vieną atlyginimą	bals-Lietuva-091213-104.txt
žadama išmokėti	bals-Lietuva-091213-104.txt