

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
SISTEMINĖS ANALIZĖS KATEDRA

Vytas Kriščiūnas

**Automatizuoto vietovardžių tyrimo modelis ir
realizacija**

Magistro darbas

Darbo vadovas

doc. dr. Vitolis Sekliuckis

Kaunas, 2009

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
SISTEMINĖS ANALIZĖS KATEDRA

Vytas Kriščiūnas

**Automatizuoto vietovardžių tyrimo modelis ir
realizacija**

Magistro darbas

Recenzentas

doc. dr. Stasys Maciulevičius

2009-05-25

Darbo vadovas

doc. dr. Vitolis Sekliuckis

2009-05-25

Atliko

IFM-3/1 gr. stud.

Vytas Kriščiūnas

2009-05-25

Kaunas, 2009

Automated Model of Place-Names Research and Implementation

SUMMARY

Place-names are part of cultural heritage and national identity. It is very important to protect and analyze current and historical place-names. The objective of master's thesis was to develop and analyze software for Lithuanian Place-names research. Firstly there were other analogous Lithuanian and European systems analyzed and compared. Also the current place-names research process was analyzed. The automated model of place-names research was developed, and based on this model; the software was designed and implemented. The software results lets perform onomastic research, saving results in a way they can be restored at any time later. There was described experiment performed in Vilnius University, using developed software. The suggestions of how to improve the software and the model were made.

TURINYS

Įvadas	3
1. Vietovardžių tyrimai ir jų specifika	4
1.1. Dabartinių vietovardžių tyrimų eiga	4
1.2. Vietovardžių skaitmeninimas ir normavimas Lietuvoje	5
1.3. Egzistuojantys vietovardžių tyrimų sprendimai užsienyje ir Lietuvoje	6
1.3.1. Toponimų duomenų bazė (Toponyms Data Bank)	6
1.3.2. Šiaurės-vakarų Rusijos toponimikos tyrimų sistema (Toponymic Research System of Northwest Russia: The TORIS System).....	7
1.3.3. Joniškio rajono vietovardžių geolingvistinė duomenų bazė.....	11
1.4. Sprendimų palyginimas.....	14
1.5. Išvados ir siūlymai galimam sprendimui.	15
2. Automatizuotų vietovardžių tyrimų programinės įrangos projektas	16
2.1. Projekto prielaidos ir tikslai	16
2.2. Esminiai reikalavimai programinei įrangai	16
2.2.1. Funkciniai reikalavimai.....	16
2.2.2. Nefunkciniai reikalavimai	19
2.3. Automatizuotų vietovardžių tyrimų PĮ architektūra.....	19
2.4. Sudėtingesni projektavimo sprendimai	21
2.5. Panaudojimo atvejai	22
2.6. Duomenų bazės modelis ir aprašymas	23
2.7. Pasirinktos realizacijai priemonės.....	28
2.7.1. CakePHP karkasas.....	28
2.7.2. MySQL DBVS	29
2.7.3. Google Maps GIS.....	30
2.8. Statistinė automatizuotų vietovardžių tyrimų PĮ informacija.....	30
3. Automatizuoto vietovardžių tyrimo modelis.....	32
3.1. Dabartinio vietovardžių tyrimo modelio automatizavimo galimybės	32
3.2. Automatizuotų vietovardžių tyrimų modelis.....	32
3.3. Programinės įrangos tobulinimo galimybės	35
4. Vietovardžių tyrimų programinės įrangos eksperimentinis tyrimas	37
4.1. Vietovardžių panaudojimo mokymo procese analizė	37
4.2. Paieškos užklausų tyrimas.....	42
4.3. Rezultatų sąrašo atkūrimo iš duomenų bazės tyrimas.....	45
Išvados	47
Literatūra	48
Terminų ir santrumpų žodynas.....	50
Priedai	51

IVADAS

Vietovardžiai neatsiejama Lietuvos kultūros paveldo ir tautinio identiteto dalis [2]. Todėl labai svarbu tinkamai apsaugoti bei ištirti ne tik esamus vietovardžius, tačiau ir tuos, kurie yra paminėti istoriniuose šaltiniuose. Taip pat nagrinėti vietovardžių kaitos dinamiką.

Šio magistrinio darbo pagrindinis tikslas yra išanalizuoti dabartinių vietovardžių tyrimų procesą, pasiūlyti sprendimus pagrįstus informacinėmis technologijomis, kaip tą procesą galima būtų automatizuoti, kad tyrimus galima būtų atlikti efektyviau. Kitas svarbus sistemos kūrimo tikslas yra pasiūlyti sistemą, kuria naudotis galėtų visos Lietuvos mokslininkai vietovardžių tyrimams atlikti.

Šiame darbe analizuojamas dabartinis onomasto darbo procesas. Jį automatizuojant sukuriama dabartinis vietovardžių tyrimų modelis. Išnagrinėjus jau egzistuojančias sistemas, pateikiami esminiai siūlymai projektuojamai programinei įrangai.

Identifikuojant silpnąsias dabartinio vietovardžių tyrimų proceso dalis, sukuriama automatizuota vietovardžių tyrimų modelis. Šis modelis, nors ir nevisiškai pilnai, bet realizuojamas kuriamoje programinėje įrangoje.

Eksperimentinėje tyrimo dalyje tiriamos sudėtingos duomenų bazės užklausos ir jų veikimas. Taip pat tiriamas šiuo metu vykstantis eksperimentinis studentų mokymo procesas.

Realizuojama automatinė vietovardžių tyrimų programinė įranga. Išanalizuojamas onomasto darbo procesas naudojantis sukurta sistema. Darbo pabaigoje pateikiami pasiūlymai, kaip tą sistemą būtų galima tobulinti.

Rengiant šį darbą bendrai su kolegomis buvo parašyta publikacija „Naujų technologijų taikymas diachroninėje toponimikoje“, kuri priimta spausdinimui tarptautiniame Vilniaus universiteto leidžiamame žurnale „Respectus Philologicus“.

Sukurta programinė įranga eksperimento tvarka buvo naudojama Vilniaus universiteto studentų mokymo procese. Ją taip pat išbandė ir VU mokslininkai.

1. VIETOVARDŽIŲ TYRIMAI IR JŲ SPECIFIKA

1.1. Dabartinių vietovardžių tyrimų eiga

Šiuo metu atliekant vietovardžių tyrimus informacinėmis technologijomis naudojamosi minimaliai.

Pirmas etapas: duomenų surinkimas. Mokslininkai prieš atlikdami tyrimą patys susiranda visą tyrimams reikalingą medžiagą. Dažnai tyrimui atrinkti duomenys arba jų dalis dubliuojasi su jau anksčiau surinktais duomenimis.

Kai duomenys atrinkti, jie yra sisteminami ir normuojami: tyrimui reikalingi vietovardžiai surašomi ant popierinių kortelių, tose pačiose kortelėse surašom lingvistinė ar morfologinė informacija apie vietovardžius. Sudaroma vietovardžių kartoteka, kuri vėliau naudojama tyrimui atlikti.

Tyrimo metu kortelės iš kartotekos rūšiuojamos rankiniu būdu, atrenkant jas pagal tam tikrus kriterijus. Egzistuoja dvi strategijos: siaurinojami ir platinamoji. Siaurinojami – tai daugkartinis vietovardžių filtravimas. Naudojant platinamąją strategiją vietovardžiai atrenkami pagal labai siaurus kriterijus, taip pildant rezultatų sąrašą.

Kai atrenkami visi tyrime naudojami vietovardžiai, tyrimas aprašomas laisva forma: aprašomi visi kriterijai kuriais naudojantis buvo atrinkti duomenys, pateikiama statistika, padaromos tyrimo išvados.

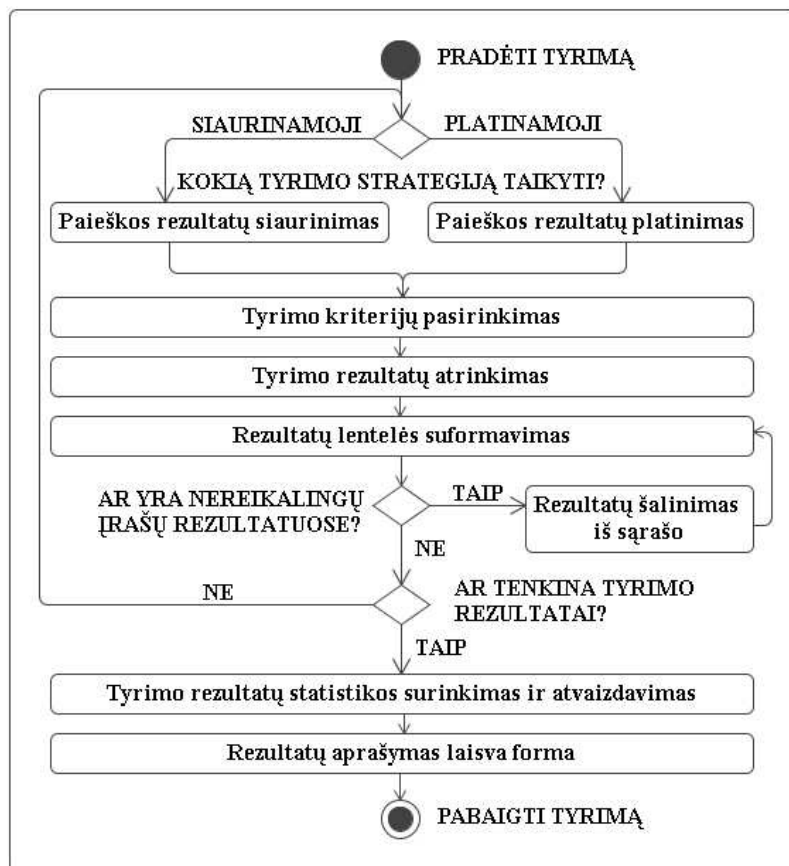
Aukščiau aprašyti tyrimo eigai galime sudaryti dabartinį vietovardžio tyrimo modelį(pav. 1).

Modelio privalumai:

- Neformalių kriterijų pasirinkimas.
- Laisva rezultatų lentelės forma.

Pagrindiniai modelio trūkumai:

- Ilgas tyrimo laikas.
- Nėra rezultatų atvaizdavimo žemėlapyje.
- Sudėtingas statistikos surinkimas ir atvaizdavimas
- Nėra galimybės atkurti prieš tai buvusio sąrašo.



Pav. 1 Dabartinis onomasto tyrimo procesas

1.2. Vietovardžių skaitmeninimas ir normavimas Lietuvoje

Informacinių technologijų panaudojimas vietovardžių tyrimuose pasaulyje pradėtas tuo pačiu metu, kai tik atsirado tam galimybės. Pirmoji mūsų aprašyta sistema pradėta kurti dar 1990 metais. Lietuvoje vietovardžių skaitmeninimas ir normavimas dažniausiai apsiriboja žodynų sudarymu, kurie dažnai yra tik popierinių leidinių analogai.

Iš rankraštinųjų istorijos šaltinių išrašyta 10 000 etninių vietovardžių. Į duomenų bazę įvesti 6 885 vietovardžiai. Sukurtos MS Access atrankos bei sujungimo užklauskos pagal duomenis iš kelių susijusių lentelių, kurios pildomos su sukurtos bazės Istoriniai etniniai vietovardžiai įvedimo modulio pagalba. Sukurtas ir tobulintas bazės peržvalgos modulio algoritmas pagal atrankos užklauskų projektą. Peržvalgos modulio algoritmas realizuotas Visual Basic priemonėmis [4].

Lietuvių kalbos instituto Vardyno skyriaus viena iš veiklos sričių yra istorinės ir dabartinės toponimikos problemų tyrimas. Visi vartotojai internete gali laisvai prieiti prie elektroninio vietovardžių žodyno. Šiuo metu žodyne yra 12 000 vietovardžių [5]. Tačiau tai tik nedidelė dalis tų (daugiau kaip 300 000), kuriuos Vardyno skyrius iš istorinių šaltinių sukaupe savo kartotekoje.

1.3. Egzistuojantys vietovardžių tyrimų sprendimai užsienyje ir Lietuvoje

Prieš įgyvendinant šį projektą būtina įvertinti, kas šiuo klausimu nuveikta Lietuvoje ir kitose Europos šalyse. Apžvelgiamos sukurtos 3 sistemos, susijusios su kuriama programine įranga.

1.3.1. Toponimų duomenų bazė (Toponyms Data Bank)

Suomijos kalbų tyrimu instituto mokslininkas Risto Miikkulainen, Tarptautiniame Onomastikos kongrese pranešė apie pradėtą kurti suomių toponimų duomenų bazę („Toponyms Data Bank“) [6]. Ši duomenų bazė leis apjungti skirtingų tipų toponiminius duomenis. Naudojantis sukurta duomenų baze Suomijos kalbų tyrimų institutas įgyvendino keltą mokslinių tiriamųjų projektų, kurių rezultatus paskelbė publikacijomis.

Vienas iš tokių projektų buvo Paikannimikartasto“ („Vietovardžių atlasas“) [8]. Projekto metu, naudojantis toponimų duomenų baze buvo sukurtas skaitmeninis vietovardžių atlasas, iliustruojantis daugiau kaip 200 labiausiai paplitusių suomių vietovardžių elementų pasiskirstymą. Atlasas nurodo koordinates, fizines savybes ir kitą aktualią lingvistinę bei geografinę informaciją.

Kitas vis dar vykdomas projektas „Paikannimikirja“ („Suomių vietovardžių knyga“). Šio projekto metu bus surinkti duomenys apie svarbiausius ir geriausiai žinomus vietovardžius. Nagrinėjamos tokios temos kaip jų kilmė ar reikšmė.

Techninė realizacija. Sistemą realizavus ir suvedus apie 100 000 vietovardžių duomenys buvo prieinami dviem būdais:

- Sukurta reliacinė duomenų bazė, o paieškom vykdyti naudojama SQL kalba.
- Sukurta elektroninė tekstinė kartoteka. Joje talpinami duomenys išsamesni nei duomenų bazėje. Paieškai vykdyti naudojamos SEARCH, AGREP, SEGREP komandos.

Informacija apie vietovardį. Šioje sistemoje numatytos galimybės labai išsamiai aprašyti ir normalizuoti vietovardžius. Duomenų bazėje vienam vietovardžiui skiriami net 32 laukai. Kaupiama informacija apie: struktūrą, sudedamąsias dalis, kilmę, susijusius vietovardžius, šaltinį, lokaciją ir kt.

1.3.2. Šiaurės-vakarų Rusijos toponimikos tyrimų sistema (Toponymic Research System of Northwest Russia: The TORIS System)

Remdamiesi suomių patirtimi, Georgy Kert iš Rusijos lingvistikos, literatūros ir istorijos instituto 1997 m. paskelbė straipsnį apie kuriamą šiaurės vakarų Rusijos toponiminių tyrimų sistemą („Toponymic Research System of Northwest Russia: The TORIS System“) [7]. Sistema nuo 2001 m. prieinama internetu: <http://toris.krc.karelia.ru/TORIS/>.

Sistemoje nėra lokalizavimo galimybės. Yra tik onomastinės analizės įrankiai.

Techninė realizacija

TORIS serverio architektūra susideda iš trijų pagrindinių elementų (žr. Pav. 2):

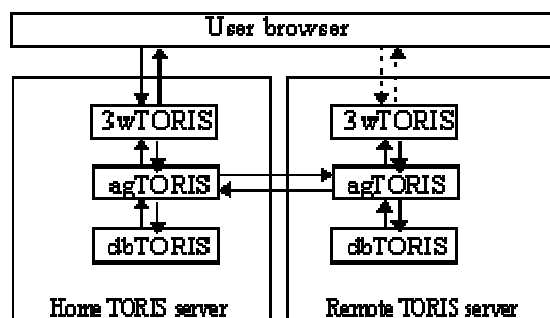
- dbTORIS yra reliacinė duomenų bazė SQL užklausų palaikymui;
- 3wTORIS yra internetinė aplikacija įvesti/išvesti informaciją;
- agTORIS yra intelektualių agentų aibė įvairioms užduotims atlikti (pagalba pildant SQL užklausų formas, autonominė paieška TORIS serveryje).

Kliento-serverio architektūros modeliui buvo naudojamas AS-modelis (Application Server), kurį sudaro trys sluoksniai:

Prezentacinė aplikacija atlikti įvedimo/išvedimo užduotis. Ji taip pat atsakinga už ryšį tarp 3wTORIS ir vartotojo naršyklės.

Dalykinė aplikacija – apjungti toponimų domenų. Ji taip pat atsakinga už ryšį tarp 3wTORIS ir agTORIS komponentų.

Resursų menedžeris – atlikti duomenų kontrolę ir saugojimą. TORIS architektūroje tai yra dbTORIS komponentas.

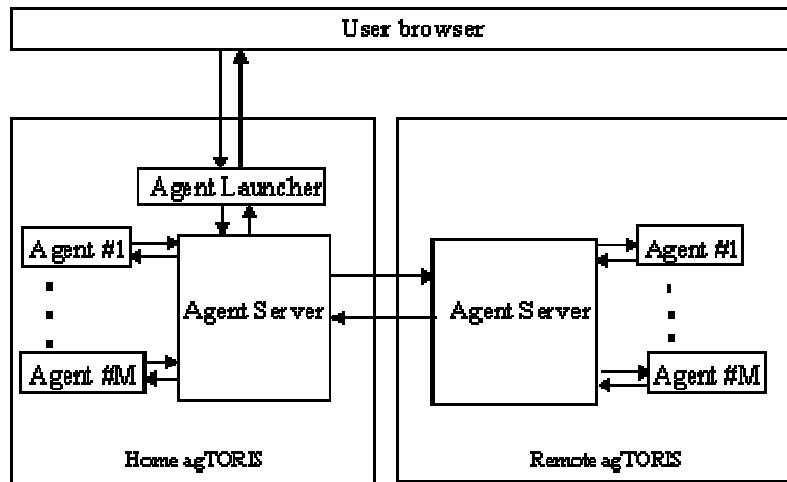


Pav. 2 TORIS architektūra. [7]

Komponentą 3wTORIS sudaro HTML šablonai ir CGI skriptai skirti vietovardžių peržiūrai, redagavimui ir įterpimui.

Pagrindinė komponento AgTORIS (žr. Pav. 3) užduotis – autonomiškai atlikti šias užduotis:

- autonomiškai atlikti paiešką TORIS serveryje,
- duomenų validavimas,
- kritinių situacijų(klaidų) valdymas.



Pav. 3 AgTORIS architektūra [7]

Paieška ir jos galimybės. TORIS sistemoje paieškos kriterijams įrašyti skirti du puslapiai. Pirmame(žr. Pav. 4) puslapyje nurodomi kriterijai apie vietovardį. Galima ieškoti pagal pavadinimo dalį, rusiška vietovardžio vardo atitikmenį, vietovardžio tipą, vietovės informaciją, kalbą, struktūrinę formulę, komponentų skaičių ir kitus.

Antrame puslapyje pažymimi sudarančių komponentų struktūrinė formulė, kilmė bei kalbos dalis.

Problem Website on Toponymy of the European North of Russia
TORIS
Toponymic Research Information System

Russian version: ISO KOI WIN DOS MAC News Databases Publications Forum Links Contact information Site map

Database "TORIS"

Data search. Step 1

Toponym

The toponym search form is divided into 3 parts. If you seek the toponym by first letters, enter the value into the first cell, if you need the words with some word-combination in the end, enter it into the third cell. The middle cell is needed for any entering word-combination into the toponym. If you are interesting any values, keep empty cells.

Example: Find toponyms, which begin with "al", end by "a:ivi" and contain "mi"

Russian name of the object

Toponym object	Republic, Region	District, City
<input type="text"/>	<input type="text"/>	<input type="text"/>

Settlements (for microtoponyms)	Representation language
<input type="text"/>	<input type="text"/>

Structural formula	The number of components	Case
<input type="text"/>	<input type="text"/>	<input type="text"/>

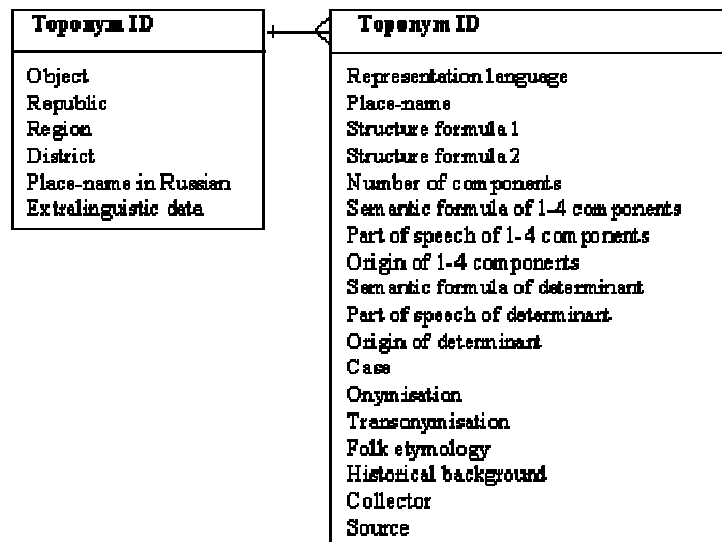
News Databases Publications Forum Links Contact information Site map

Last modified on April 10, 2003
webmaster@www.krc.kareliaru

Database "TORIS"

Pav. 4 TORIS paieškos galimybės

Informacija apie vietovardį. Vietovardžiai normalizuoti pagal specifinę projekto paskirtį ir regioną. Sistemos privalumas tai, kad pilnai išnagrinėta vietovardžių kilmė bei sudėtis. Iš viso sistemoje yra apie 2 200 vietovardžių. Reliacinės duomenų bazės schemos dalis, kurioje saugomi vietovardžio duomenys (žr. Pav. 5) rodo, kad buvo numatyta galimybė susieti to paties vietovardžio kelias skirtingas versijas skirtingomis kalbomis. Tokią duomenų bazės struktūrą galima naudoti ir istoriniams vietovardžiams susieti. Duomenų bazėje saugojama lingvistinė, kultūrinė ir istorinė informacija apie vietovardį. Pav. 6-tame paveikslėlyje pavaizduota vartotojui pateikiamos informacijos pavyzdys.



Pav. 5 TORIS reliacinės duomenų bazės struktūros dalis [7]

Database "TORIS"

Data search.

Toponym's object	Часть деревни
Republic, Region	Karelia
District	Kalevala national
Representation language	Karelian
Toponym	Šuotermä
Structural formula	C111 - Сложные. Атрибут - имя существительное в номинативе
The number of components	2
Case	НОМИНАТИВ
Attribute	
<u>1 component</u>	
Component's name	Šuo
Semantic formula	A221 - Вселенная. Ландшафт и полезные ископаемые

Pav. 6 TORIS informacija apie vietovardį

1.3.3. Jonišio rajono vietovardžių geolingvistinė duomenų bazė

Pirmoji skaitmeninė regioninės toponimijos duomenų bazė oficialiai pasirodė 2007 m. [3]. Jos kūrėjai teigia jog ji „gali pasitarnauti kaip modelis skaitmeninant ir kitų regionų ar net visos Lietuvos vietovardžius“. Išskiriamos šios pagrindinės panaudojimo galimybės:

- Vietovardžių tyrėjams (onomastams).
- Studentų mokymams, mokant juos mokslo tyrinėjimo metodų.
- Visuomeninėms reikmėms – krašto pažinimui.
- tam tikroms administracinėms regiono reikmėms.

Šioje vietovardžių duomenų bazėje yra apie 6 000 vietovardžių, priklausančių Jonišio rajono teritorijai. Visi vietovardžiai yra lokalizuoti, arba bent jau apytiksliai nustatyta jų lokacija.

Techninė realizacija. Duomenų bazei sukurti naudota atvirojo kodo programinė įranga: 1) duomenų bazės valdymo sistema pasirinkta MySQL DBVS. 2) Naudotas *Ruby on Rails* internetinių aplikacijų kūrimo karkasas. Šis karkasas naudoja MVC projektavimo šabloną. 3) Vartotojo sąsajai sukurti buvo naudojamos XHTML, CSS, Javascript ir AJAX priemonės.

Informacijos susiejimas su žemėlapiu. Sistemoje kiekvienas vietovardis yra lokalizuotas konkrečioje Jonišio rajono teritorijos vietoje. Sistemoje naudojami didelio tikslumo GIS centro sukurti, didelio tikslumo, Jonišio rajono žemėlapiai. Šie žemėlapiai yra tik popierinių žemėlapių analogai. Kiekvienas vietovardis buvo tiksliai arba apytiksliai lokalizuotas pagal šaltiniuose pateiktą informaciją.

Paieškos galimybės (žr. Pav. 7). „Šioje duomenų bazėje galima labai įvairi paieška: 1) pagal konkretų visą vietovardį (pvz., *Akmėnynė*), 2) pagal vietovardžio dalį: *als* (bus rasti visi vietovardžiai, turintys šių raidžių junginį), *^als* (bus rasti vietovardžiai, turintys šį raidžių junginį pradžioje), *ynė\$* (bus rasti visi vietovardžiai, turintys šį raidžių junginį pabaigoje), 3) pagal išskirtus požymius: šaltinį, administracinę teritoriją, toponiminę klasę, darybos būdą, darybos pamatą, 4) pagal bet kurią „Kitos informacijos“ žodį (pvz. *miestas* – bus rasti visi vietovardžiai, kurių kitoje informacijoje yra žodis *miestas*).“ (Bartkutė; Jodelis, 2007 : 3-4).

Pav. 7 Paieškos galimybės. Joniškio rajono vietovardžių DB. [3]

Informacija apie vietovardį (žr. Pav. 8). Pateikiama pakankamai nedaug susistemintos informacijos apie vietovardį. Išskiriami šie vietovardžio atributai:

- Pavadinimas
- Kirtis
- Seniūnija
- Valsčius
- Klasė
- Daryba
- Pamatas
- Lokacija
- Kita informacija.

Pav. 8 Informacija apie vietovardį. Joniškio rajono vietovardžių DB [3].

Informacija žemėlapyje (žr. Pav. 9 ir Pav. 10). Kiekvienas vietovardis lokalizuojamas atskirame lange „Padėtis žemėlapyje“. Jų vieta, remiantis šaltiniuose pateikta informacija, gali būti tiksli (vėliavėlė) arba apytiksli (rodyklės). Taip pat galima ieškoti ir kitų vietovardžių, esančių tam tikru atstumu nuo jau rasto vietos vardo.



Pav. 9 Informacija žemėlapyje: lokacija vietovėje smulkiu masteliu [3]



Pav. 10 Informacija žemėlapyje: lokacija rajone. Joniškio rajono vietovardžių DB.

1.4. Sprendimų palyginimas

Lyginant sistemas vertėtų atkreipti dėmesį, kad šios buvo kurtos skirtingu laiku, skirtingų regionų vietovardžiams. Pradžioje bus apžvelgti konkrečių sistemų privalumai bei trūkumai atskleidžiantys pagrindinius šių sistemų skirtumus.

Suomijos kalbų tyrimų instituto informacinės sistemos pagrindiniai privalumai:

- GIS palaikymas
- Daugiausiai ir išsamiausių duomenis turinti sistema.

Trūkumai:

- Neprieinama internete.
- Mažai duomenų apie sistema gautus rezultatus.

Rusijos lingvistikos, literatūros ir istorijos instituto sukurtos sistemos privalumai:

- Viešai prieinama internete.
- Vietovardžiai išskaidyti dalimis pagal jų darybą.
- Galimybė parinkti kriterijus pagal vietovardžių dalis.
- Kelios vartotojo sąsajos kalbos

Trūkumai:

- Vietovardžiai nėra lokalizuoti, neatvaizduojami žemėlapyje.
- Ne visi vietovardžių atributai gali būti naudojami kaip paieškos kriterijai, pvz. vietovardžio šaltinis.

Joniškio rajono vietovardžių geolingvistinės duomenų bazės privalumai:

- Lokalizuoti vietovardžiai.
- Dviejų lygių atvaizdavimas žemėlapyje, bei didelis žemėlapių tikslumas.
- Aplinkinių vietovardžių paieška(nurodytu atstumu).
- Lietuvių kalbos morfologijos palaikymas
- Suprojektuota darbui internete.
- Naudojami AJAX metodai.

Trūkumai:

- Šiuo metu neprieinama internete.
- Tinkama tik vieno rajono vietovardžiams.
- Nėra galimybės rezultatų sąrašė matyti ne tik vietovardžių pavadinimus.

Lentelė 1

Funktionalumas	Suomijos kalbų tyrimų instituto IS	Rusijos lingvistikos, literatūros ir istorijos instituto IS	Joniškio rajono vietovardžių geolingvistinė duomenų bazė
Lietuvių kalbos morfologijos palaikymas	–	–	+
Galimybė dirbti internete	–	+	+/-
Kartografinė eksplikacija	+	–	+
GIS palaikymas	+	–	–
Dabartinių vietovardžių morfeminė analizė	+	+	+
Dabartinių ir istorinių vietovardžių susiejimas	–	+/-	–
Automatizuotų vietovardžių tyrimų galimybės	–	–	–
Vartotojo sąsąja keliomis kalbomis	–	+	–

1.5. Išvados ir siūlymai galimam sprendimui.

Išanalizavus onomasto darbo procesą, bei jau sukurtas sistemas buvo padarytos tokios išvados:

- Nėra sukurto tinkamo sprendimo dabartiniam vietovardžių tyrimo procesui automatizuoti, sukurti sprendimai gali būti naudojami kaip papildomi įrankiai jiems atlikti.
- Sukurtos sistemos orientuotos į paprastą vietovardžio paiešką, tačiau ne į patį tyrimo procesą.
- Nė viena egzistuojanti sistema neleidžia susieti istorinio vietovardžio informacijos su dabartiniu vietovardžiu

Esminiai siūlymai naujai kuriamam sprendimui:

- Sukurti posistemę vietovardžių tyrimams atlikti. Numatyti galimybes viešinti tyrimų rezultatus.
- Integruoti GIS.
- Sudaryti galimybę susieti istorinius vietovardžius su dabartiniais. Būtina sąlyga diachroninės toponimikos tyrimams atlikti.
- Sistema turi būti prieinama internetu.

2. AUTOMATIZUOTŲ VIETOVARDŽIŲ TYRIMŲ PROGRAMINĖS ĮRANGOS PROJEKTAS

2.1. Projekto prielaidos ir tikslai

Šiuo projektu siekiama sukurti programinę įrangą vietovardžių tyrimams atlikti. Kadangi Lietuvoje neegzistuoja tinkama duomenų bazė vietovardžiams normuoti, šios duomenų bazės projektavimas ir realizavimas yra sudėtinė, ir neatsiejama šio projekto dalis. Realizuojant šį projektą, ypatingas dėmesys skiriamas sistemos tobulinimo galimybėms ateityje. Jeigu sistema bus pastoviai tobulinama, ateityje tai gali tapti onomastikos tyrimų pagrindiniu įrankiu Lietuvoje.

Kitas svarbus projekto tikslas: skatinti informacinių technologijų plėtrą humanitariniuose moksluose. Sukurta programinė įranga bus naudojama ne tik vietovardžių tyrimams tačiau ir studentų apmokymams.

2.2. Esminiai reikalavimai programinei įrangai

Išanalizavus Lietuvoje ir užsienyje sukurtus sprendimus, taip pat bendradarbiaujant su Vilniaus universiteto Kauno humanitarinio fakulteto mokslininkais, buvo suformuluoti reikalavimai projektuojamai programinei įrangai.

2.2.1. Funkciniai reikalavimai

1. Galimybė įvesti vietovės ir gramatikos informaciją dabartiniams, vietovardžiams:

1.1. vietovardžio pavadinimas;

1.2. vietovės informacija:

- tipas (pvz., *kaimas*);
- apskritis (pvz., *Kauno*);
- rajonas (pvz., *Jonavos*);
- seniūnija (pvz., *Garliava*)
- atstumas nuo seniūnijos centro (km);
- atstumas nuo parapijos centro (km);
- ar egzistuojantis;
- literatūra;
- pastabos.

1.3. gramatikos informacija:

- priešdėliai;
- šaknys;

- kamienai;
- priesagos;
- galūnės;
- tipas (*pirminis, vedinys (priešdėlinis, priešdėlėtasis, priesaginis, priesagėtasis), dūrinys (suvestinis, sudurtinis)*);
- Ar taisyklingas vietovardis?
- Taisyklingumo pastabos;
- Semantika;
- Kirčiuotė;
- Kirčiuotės pastabos;
- Etimologija;
- Liaudies etimologija.

2. Galimybė įvesti vietovės ir gramatikos informaciją istoriniams, vietovardžiams:

2.1. istorinio vietovardžio pavadinimas;

2.2. istorinė vietovės informacija:

- šaltinis;
- šaltinio puslapio numeris;
- numeris schemeje;
- vyskupija;
- dekanatas;
- parapija;
- senoji apskritis;
- senoji seniūnija;
- metai;
- vartojimo kontekstas lietuvių kalba;
- vartojimo kontekstas užsienio kalba;
- vartojimo konteksto užsienio kalba.

2.3. gramatikos informacija:

- priešdėliai;
- šaknys;
- kamienai;
- priesagos;
- galūnės.

3. Galimybė lokalizuoti dabartinius vietovardžius Lietuvos teritorijoje;

4. Galimybė atlikti paiešką tarp dabartinių ir istorinių vietovardžių pagal:

4.1. pagal konkretų dabartinį arba istorinį vietovardį (pvz., *Gegužiai*);

4.2. pagal dabartinę vietovardžio informaciją:

- tipą (pvz., *kaimas*);
- apskritį (pvz., *Kauno*);
- rajoną (pvz., *Jonavos*);
- seniūniją (pvz., *Garliava*);
- atstumą nuo seniūnijos centro (pvz., *nuo 10 iki 40 km*);
- atstumą nuo parapijos centro (pvz., *nuo 10 iki 40 km*);
- egzistavimą (pvz., *taip arba ne arba visi*);
- paminėjimą literatūros šaltiniuose (pvz., *taip arba nesvarbu*);
- literatūros raktinį žodį (pvz., *Čekiškės parapija*);
- lokalizaciją (pvz., *taip arba ne arba visi*).

4.3. pagal istorinę vietovardžio informaciją:

- šaltinį (pvz., *vietovardžių žiniaraštis 2147*);
- istorinio vietovardžio numerį schemeje (pvz., *4a*);
- vyskupiją (pvz., *Kauno*);
- dekanatą (pvz., *Raseiniu*);
- parapiją (pvz., *Čekiškės*);
- bažnytinio archyvo suskirstymo metus (pvz., *1847*);
- senąją apskritį (pvz., *Kauno*);
- senąją seniūniją (pvz., *Kauno*);
- vartojimo kontekstą lietuvių kalbą;
- vartojimo kontekstą užsienio kalbą;
- vartojimo konteksto užsienio kalbą (pvz., *lenku*);
- laikmetį (pvz., *nuo 1859 iki 1914*).

4.4. pagal gramatikos informaciją:

- priešdėlius (pvz., *pa*);
- šaknis (pvz., *geg*);
- kamienus (pvz., *geguž*);
- priesagas (pvz., *už*);
- galūnes (pvz., *iai*);
- tipą (pvz., *pirminis, vedinys, dūrinys*);
- taisyklingumą (pvz., *taip arba ne arba visi*);
- semantiką;
- kirčiuotę (pvz. 1)

- aprašytą etimologiją (pvz., *taip arba nesvarbu*);
 - aprašytą liaudies etimologiją (pvz., *taip arba nesvarbu*).
5. Prie sistemos gali prieiti tik prisijungę vartotojai. Turi būti trys vartotojų lygiai:
 - 5.1. administratoriai, turintys visas teises naudotis sistema;
 - 5.2. tyrėjai, turi visas teises naudotis sistema, išskyrus vartotojų administravimą;
 - 5.3. vartojai, galintys įvesti ir koreguoti vietovardžių informaciją.
 6. Turi būti galimybė peržiūrėti suvestą dabartinių vietovardžių informaciją;
 7. Turi būti galimybė peržiūrėti suvestą istorinių vietovardžių informaciją;
 8. Turi būti galimybė peržiūrėti lokalizuotus dabartinius vietovardžius žemėlapyje;
 9. Turi būti galimybė atlikti toponimų mokslinius tyrimus:
 - 9.1. Tyrimas gali turėti tris būsenas: nebaigtas, laukia atsakymo ir pabaigtas;
 - 9.2. Turi būti galimybė rašyti tyrimo pastabas;
 - 9.3. Tyrimus turi būti galima vykdyti iteracijomis;
 - 9.4. Kiekvienos iteracijos metu paieškos kriterijai turi būti išsaugomi;
 - 9.5. Turi būti galimybė grįžti į bet kurią buvusią iteraciją ir tęsti tyrimą kita kryptimi;
 - 9.6. Tyrimo rezultatus turi būti galima plėsti (pridėti naujus paieškos rezultatus);
 - 9.7. Turi būti galimybė atlikti paiešką tarp tyrimo rezultatų;
 - 9.8. Turi būti galimybė šalinti vietovardžius iš atfiltruotų vietovardžių sąrašo;
 - 9.9. Tyrimų rezultatų lentelė turi būti dinaminė, t.y. turi būti galima pasirinkti, kokius lentelės laukus rodyti;
 - 9.10. Turi būti galimybė išsaugoti tyrimų rezultatus;
 - 9.11. Turi būti galimybė atkurti anksčiau vykdyto tyrimo rezultatus.
 10. Sistemoje turi būti galimybė kirčiuoti lietuviškus žodžius, rašyti lietuviškais, lenkiškais, latviškais, rusiškais rašmenimis.

2.2.2. Nefunkciniai reikalavimai

1. Lengvai įskaitoma vartotojo sąsaja;
2. Vartotojas turi būti informuojamas įvykus klaidai;
3. Žemėlapyje vietovardis turi būti žymimas vėliavėle.

2.3. Automatizuotų vietovardžių tyrimų PĮ architektūra

Surinkus reikalavimus galima pradėti projektuoti programinę įrangą. Sistemos kūrimas yra iteracinis procesas, kadangi reikalavimų kaitos rizika yra didelė, o sistemos palaikymo, priežiūros ir tobulinimo poreikis ateičiai yra akivaizdus. Buvo pasirinktas MVC (modelis-peržiūra-kontroleris) projektavimo šablonas. Jis išskiria sistemą į tris lygius, kuris kiekvienas atsakingas už tam tikrą funkciją. Panašus sprendimas buvo naudotas ir TORIS

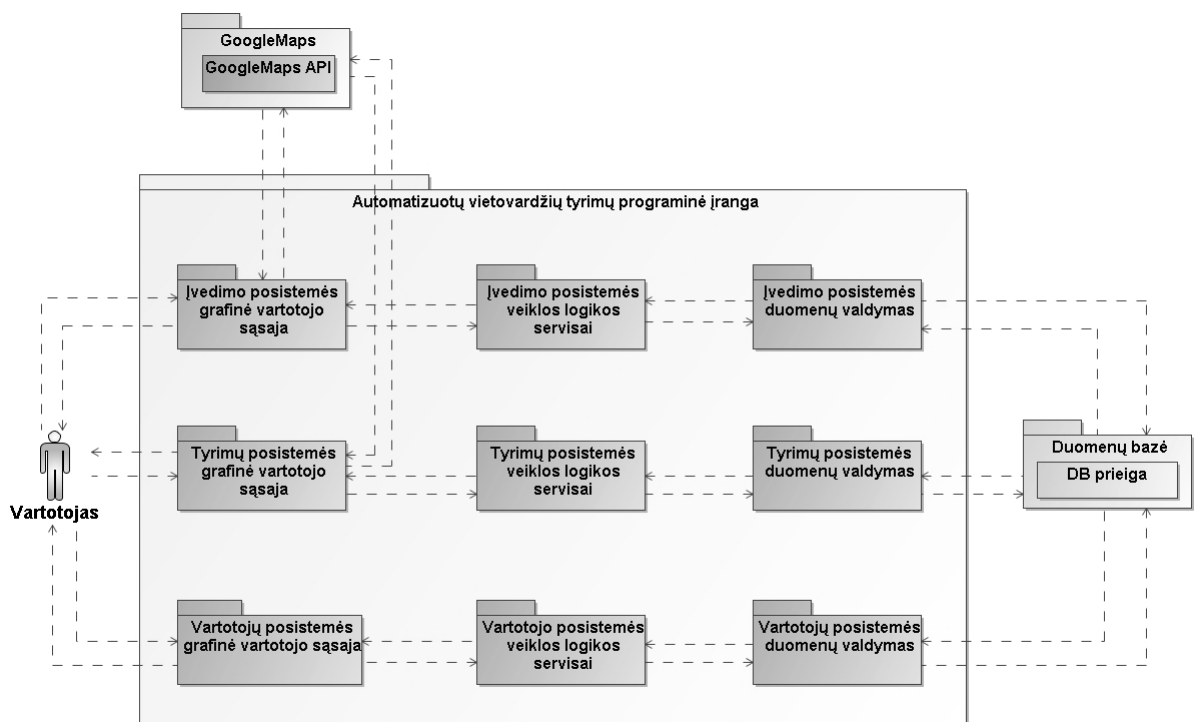
sistemoje, kur architektūra taip pat buvo trijų lygių. Modelis atsakingas už duomenų sinchronizaciją su duomenų baze. Kontroleris – už sistemos funkcinę dalį. Peržiūra – už duomenų atvaizdavimą vartotojui. MVC pagrindinė idėja yra labai paprasta: programas yra lengviau kurti, derinti ir palaikyti, jei skirtingi programinės įrangos dalių atsakomybės yra paskirstytos tarp skirtingų objektų [23].

Remiantis surinkta informacija apie projektą ir MVC architektūros principus, buvo sudaryta Automatizuotų vietovardžių tyrimų programinės įrangos architektūra (Pav. 11). Sistema yra suskaidyta į 3 posistemes:

- Vartotojų duomenų valdymo;
- Vietovardžių duomenų valdymo (įvedimo);
- Tyrimo valdymo.

Vartotojų duomenų valdymo posistemė yra atsakinga už vartotojų duomenų administravimą, jų teises. Vietovardžių duomenų valdymo posistemė atsakinga už duomenų, šiuo atveju vietovardžių ir jų atributų, administravimą ir atvaizdavimą. Taip pat duomenų valdymo posistemė vykdo paprastą vietovardžių paiešką. Tyrimų valdymo posistemė už tyrimų organizavimą, jų rezultatų saugojimą. Per Google Maps geografinę informacinę sistemą vietovardžių duomenų valdymo posistemės grafinė vartotojo sąsaja lokalizuoja vietovardį. Vietovardžių atvaizdavimas Google Maps vykdomas tyrimų posistemėje.

Programos duomenų bazėje saugoma visa informacija apie vietovardžius ir jų tyrimus.



Pav. 11 Automatizuotų vietovardžių tyrimų PĮ architektūra

2.4. Sudėtingesni projektavimo sprendimai

Toliau aprašomi kelios sudėtingesnės projektavimo problemos, būdai joms spręsti bei pasirinktas sprendimas ir jo pagrindimas.

Tyrimo rezultatų ir kriterijų saugojimas

Jeigu turimi duomenys nekistų, tiems patiems tyrimo rezultatams gauti pakaktų saugoti tik veiksmus, kuriuos atliko tyrėjas. Šiuo atveju saugojami būtų paieškos kriterijai ar šalinamų vietovardžių numeriai. Naudojant šį sprendimą būtų galima minimizuoti saugomų duomenų kiekį duomenų bazėje. Šio sprendimo realizacija būtų sąlyginai sudėtinga. Tačiau didžiausias tokio sprendimo trūkumas būtų rezultatų atkūrimo priklausomybė nuo norimo atkurti sąrašo. Kuo rezultatui pasiekti būtų naudojama daugiau žingsnių, tuo labiau būtų apkrauta DBVS bei lėtėtų atkūrimo laikas.

Kitas panašus sprendimas būtų saugoti duomenų bazės SQL užklausas, paieškomis tarp rezultatų naudojant užklausų užklausose SQL galimybę, o rezultatų papildymui naudot *UNION* sujungiant užklausas. Šio varianto realizacija supaprastėja, tačiau pagrindine problema ir toliau išlieka rezultatų atkūrimas. Taip pat didėja ir duomenų bazėje užimama vieta, nors tarp šių dviejų sprendimų įmanomi ir kompromisiniai variantai.

Sudėtingiausiai realizuojamas sprendimas būtų pagal saugomus kriterijus ir veiksmus sudaryti vieną SQL užklausą. Toks rezultatų atrinkimas reikalautų sudėtingo kriterijų formalizavimo, loginės jų analizės atkuriant rezultatų sąrašą. Taip pat būtų ribojamas sistemos funkcijų išplečiamumas ateityje. Šiuo atveju rezultatų atrinkimas pagrinde priklausytų nuo vietovardžių esančių duomenų bazėje skaičiaus.

Pasirinktas sprendimo variantas buvo saugoti visus sąrašo elementus (jų numerius). O kriterijus saugoti tik kaip papildomas pastabas tyrėjams. Tokiu būdu duomenų bazėje saugoma daugiau duomenų, ypač kai sąrašuose yra daug įrašų. Norint pasiekti minimalų išrinkimo laiką, būtina tokia lentelę indeksuoti. Naudojant šį sprendimą konkrečių rezultatų atrinkimas priklauso nuo rezultatų kiekio, bet ne nuo lentelės dydžio ar konkretaus tyrimo. Šis sprendimas reikalauja daugiau priežiūros nes egzistuoja galimybė, kad daugėjant duomenų, bei didėjant tyrimų apimčiai, bus viršytos įrangos techninės galimybės.

Kirčių saugojimas duomenų bazėje ir jų atkūrimas

Kirčiams atvaizduoti buvo pasirinkti unikodo kirčio simboliai, kurie nevisada teisingai atvaizduojami, tačiau kitas įmanomas sprendimas reikalautų įrašyti papildomą šriftą

kliento kompiuteryje, nes visus lietuvių kalboje naudojamus ženklus turi vienintelis lietuvių kalbai sukurtas šriftas – „Palemonas“ .

Norint atlikti paieškos užklausas neatsižvelgiant į kirčius, buvo būtina kirčius saugoti atskirai nuo vietovardžio pavadinimo, tačiau vietovardžiai atvaizduojami turėjo būti su kirčiais. Šiam tikslui pasiekti buvo naudojamos *CakePHP* karkaso modeliuose integruotos funkcijos: *beforeSave* ir *afterFind*. *BeforeSave* funkcija yra visada iškviečiama, prieš įrašant įrašą į duomenų bazę. *AfterFind* funkcija leidžia pakoreguot paieškos rezultatus prieš perduodant juos kontrolieriui.

Vieno vietovardžio kirčiams duomenų bazėje saugoti buvo paskirti trys baitai: po vieną kiekvienam kirčiui (reikalavimų analizės metu buvo nustatyta, kad maksimalus galimas kirčių skaičius – 3). Pirmieji du kiekvieno baito bitai buvo skirti kirčio tipui (lietuvių kalboje naudojami trys kirčio tipai). Kiti šeši bitai naudojami nusakyti vietovardžio vietą, kurioje kirtis yra dedamas. Šis sprendimas tinkamas sukirčiuoti ne ilgesnius nei 64-ių ženklų ilgio vietovardžius.

2.5. Panaudojimo atvejai

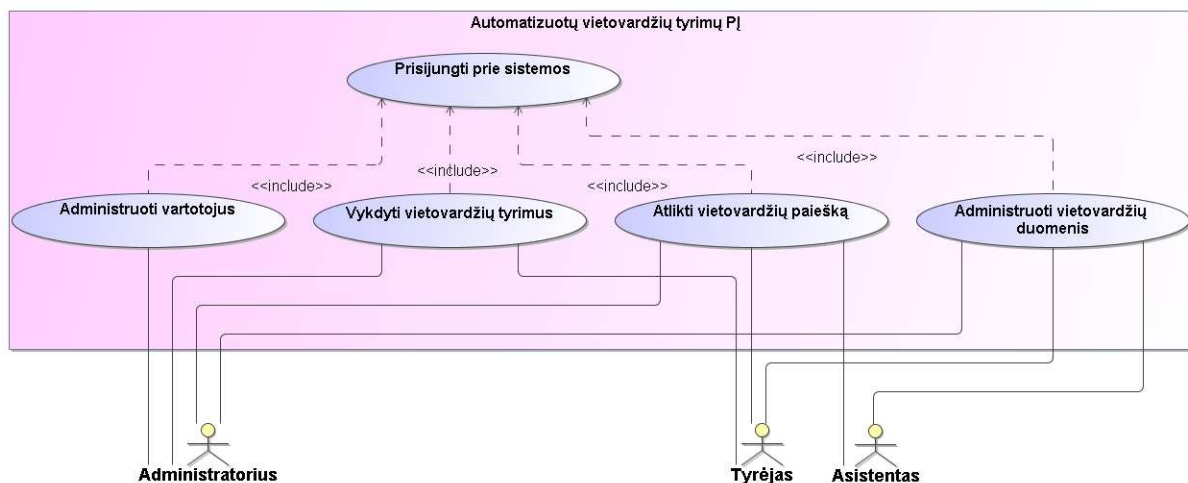
Kaip ir buvo minėta (funkcinis reikalavimas 5), sistemoje bus trys vartotojų tipai, turintys skirtingas teises. Jos yra pavaizduotos panaudojimo atvejų diagramoje (Pav. 12).

Administratoriai galės administruoti vartotojus bei vietovardžių duomenis, atlikti toponimų paiešką bei vykdyti vietos vardų tyrimus.

Tyrėjai galės administruoti vietovardžių duomenis, atlikti toponimų paiešką bei vykdyti vietos vardų tyrimus. Jiems bus draudžiama administruoti vartotojus.

Tyrėjų asistentai galės tik suvesti vietovardžių duomenis bei atlikti jų paiešką tam, kad galėtų koreguoti ir lokalizuoti savo įvestus toponimus.

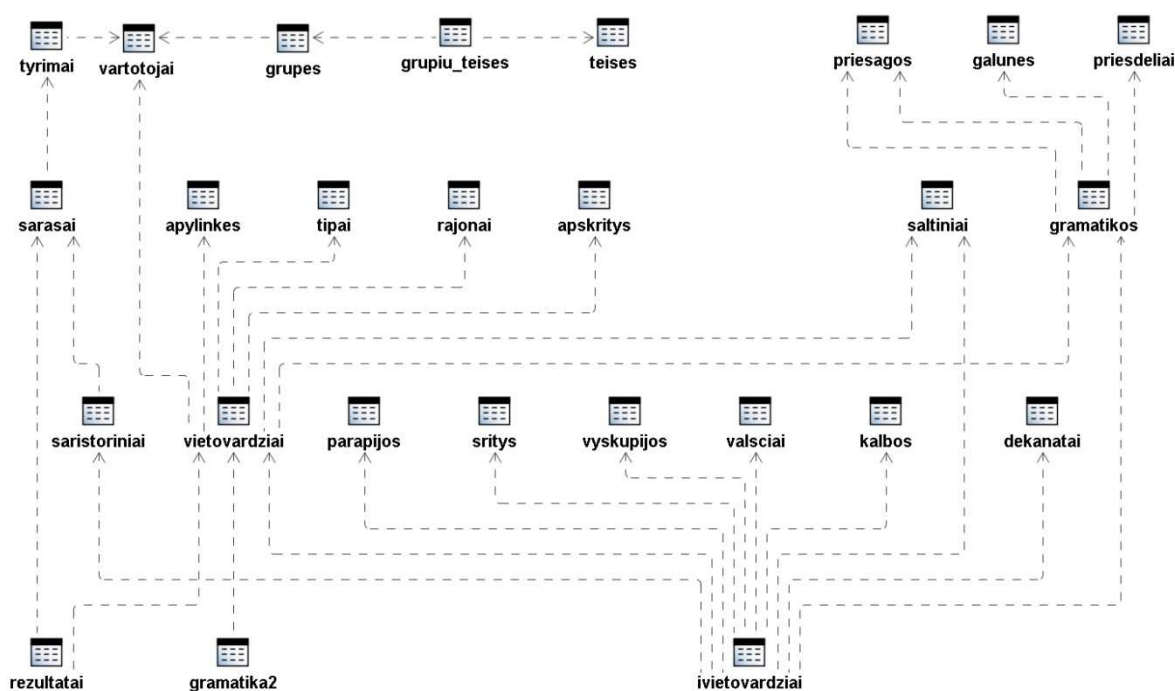
Visi vartotojų tipai savo veiksmus galės atlikti tik prisijungę prie sistemos, neprisijungus prie sistemos jokie veiksmai nėra galimi.



Pav. 12 Automatizuotų vietovardžių tyrimų PĮ panaudojimo atvejų diagrama

2.6. Duomenų bazės modelis ir aprašymas

Duomenų bazėje, kurios struktūra pavaizduota Pav. 13, saugoma visa informacija apie dabartinių ir istorinių vietovardžių vietovę, gramatiką ir lokaciją. Ten taip pat saugoma ir visa informacija apie tyrimus. Duomenų bazėje taip pat saugomi ir vartotojų duomenys. Sistemos lankstumą užtikrina vartotojų grupių teisių saugojimas duomenų bazėje. Taip ateityje nesunkiai, bus galima sukurti daugiau vartotojų tipų, turinčių skirtingas teises.



Pav. 13 Automatizuotų vietovardžių tyrimų PĮ duomenų bazės struktūra

Lentelė 2 Duomenų bazės lentelės, jų atributų tipai ir reikšmės [25].

ATRIBUTAS	TIPAS	Paskirtis
VIETOVARDZIAI		
Saugoma visa informacija apie dabartinį vietovardį.		
ID	INT	Vietovardžiui identifikuoti.
PAV	VARCHAR(64)	Vietovardžio pavadinimui.
TIPAS_ID	INT	Vietovardžio tipo identifikuoti
APSKRITIS_ID	INT	Vietovardžio apskričiai identifikuoti.
RAJONAS_ID	INT	Vietovardžio rajonui identifikuoti.
APYLINKE_ID	INT	Vietovardžio seniūnijai identifikuoti.
SALTINIS_ID	INT	Vietovardžio šaltiniui identifikuoti.
PLATUMA	DOUBLE	Vietovardžio koordinatei žemėlapyje (platuma).
ILGUMA	DOUBLE	Vietovardžio koordinatei žemėlapyje (ilguma).
ATS1	VARCHAR(10)	Atstumui nuo parapijos centro.
ATS2	VARCHAR(10)	Atstumui nuo seniūnijos centro.
EGZISTUOJA	TINYINT	Vietovardžio egzistavimui identifikuoti.
LITERATURA	TEXT	Kokioje literatūroje paminėtas vietovardis.
PASTABOS	TEXT	Pastabos apie vietovardį.
NR_SCHEMOJE	VARCHAR(5)	Vietovardžio numeriui kartografinėje schemeje.
GRAMATIKA_ID	INT	Gramatikai (morfeminiam išskaidymui) identifikuoti.
KIRTIS1	TINYINT	Vietovardžio pirmo žodžio kirčiui.
KIRTIS2	TINYINT	Vietovardžio antro žodžio kirčiui.
KIRTIS3	TINYINT	Vietovardžio trečio žodžio kirčiui.
VARTOTOJAS_ID	INT	Sistemos vartotojui identifikuoti.
GRAMATIKOS		
Saugoma informacija apie morfeminį vietovardžio išskaidymą.		
ID	INT	Vietovardžio gramatikai (morfeminiam išskaidymui) identifikuoti.
VIETOVARDIS	VARCHAR(32)	Vietovardžio pavadinimui
PRIEDELIS1	INT	Vietovardžio pirmam priešdėliui.
PRIEDELIS2	INT	Vietovardžio antram priešdėliui.
PRIEDELIS3	INT	Vietovardžio trečiam priešdėliui.
PRIESAGA1	INT	Vietovardžio pirmai priesagai.
PRIESAGA2	INT	Vietovardžio antrai priesagai.
PRIESAGA3	INT	Vietovardžio trečiai priesagai.
PIEDELIS4	INT	Vietovardžio ketvirtai priesagai.
GALUNE1	INT	Vietovardžio pirmai galūnei.
GALUNE2	INT	Vietovardžio antrai galūnei.
SAKNYS	VARCHAR(100)	Vietovardžio šaknims.
KAMIENAI	VARCHAR(100)	Vietovardžio kamienams.
GRAMATIKA2		
Saugoma informacija apie papildomą dabartinio vietovardžio gramatiką.		
ID	INT	Vietovardžio gramatikai (nemorfeminei) identifikuoti.
VIETOVARDIS_ID	INT	Vietovardžiui identifikuoti.

TIPAS	TINYINT	Ar vietovardis pirminis ar vedinys. Ir jei vedinys, tai koks.
TAISYKLINGUMAS	TINYINT	Ar vietovardis taisyklingas.
TAISYKLINGUMOP	TEXT	Pastaboms apie vietovardžio taisyklingumą.
SEMANTIKA	TEXT	Pastaboms apie vietovardžio semantiką.
KIRCIUOTE	ENUM('1', '2', '3A', '3B', '4', '4A', '4B')	Vietovardžio kirčiuotei.
KIRCIUOTESP	TEXT	Pastaboms apie vietovardžio kirčiuotę.
ETIMOLOGIJA	TEXT	Vietovardžio etimologijai (kilmei ir raidai).
LIAUDIESET	TEXT	Vietovardžio liaudies etimologijai (kilmei ir raidai).
IVIETOVARDZIAI		
Saugoma visa informacija istorinį vietovardį.		
ID	INT	Istoriniam vietovardžiui (toliau - IV) (dabartinio atitikmeniui) identifikuoti.
PAV	VARCHAR(100)	IV pavadinimui.
SALTINIS_ID	INT	IV šaltiniui identifikuoti.
VIETOVARDIS_ID	INT	Dabartiniam vietovardžiui identifikuoti.
PUSLAPIS	VARCHAR(5)	Kuriame šaltinio puslapyje IV paminėtas.
NR_SCHEMOJE	VARCHAR(12)	IV numeriui kartografinėje schemoje.
VYSKUPIJA_ID	INT	IV vyskupijai identifikuoti.
DEKANATAS_ID	INT	IV dekanatui identifikuoti.
PARAPIJA_ID	INT	IV parapijai identifikuoti.
SRITIS_ID	INT	IV sričiai identifikuoti.
VALSCIUS_ID	INT	IV valsčiui identifikuoti.
KONTEKSTAS_LT	TEXT	IV vartojimo kontekstas lietuvių kalba.
KONTEKSTAS_KT	TEXT	IV vartojimo kontekstas užsienio kalbai.
METAJ	YEAR	Kuriais metais IV priklausė vyskupijai, dekanatui, parapijai.
KALBA_ID	MEDIUMINT	IV vartojimo konteksto kalbai identifikuoti.
KIRTIS1	TINYINT	IV pirmo žodžio kirčiui.
KIRTIS2	TINYINT	IV antro žodžio kirčiui.
KIRTIS3	TINYINT	IV trečio žodžio kirčiui.
GRAMATIKA_ID	INT	Gramatikai (morfeminiam išskaidymui) identifikuoti.
APSKRITYS		
Saugoma informacija apie Lietuvos apskritis.		
ID	INT	Vietovardžio apskričiai identifikuoti.
PAV	VARCHAR(50)	Apskritis pavadinimui.
RAJONAI		
Saugoma informacija apie Lietuvos rajonus. Jie yra susieti su apskritimis.		
ID	INT	Vietovardžio rajonui identifikuoti.
APSKRITIS_ID	INT	Kuriai apskričiai priklauso rajonas identifikuoti.
PAV	VARCHAR(100)	Rajono pavadinimui.

APYLINKES		
Saugoma informacija apie Lietuvos seniūnijas. Jos yra susietos su rajonais.		
ID	INT	Vietovardžio seniūnijai identifikuoti.
RAJONAS_ID	INT	Kuriam rajonui priklauso seniūnija identifikuoti.
PAV	VARCHAR(100)	Seniūnijos pavadinimui.
TIPAI		
Saugoma informacija apie vietovardžio tipus.		
ID	INT	Vietovardžio tipui identifikuoti.
PAV	VARCHAR(100)	Tipo pavadinimui.
SALTINIAI		
Saugoma informacija apie vietovardžių paminėjimo šaltinius.		
ID	INT	Vietovardžio šaltiniui identifikuoti.
PAV	VARCHAR(100)	Šaltinio pavadinimui.
PR_METAI	INT	Nuo kada šaltinis gyvavo.
PB_METAI	INT	Iki kada šaltinis gyvavo.
VYSKUPIJOS		
Saugoma informacija apie Lietuvos vyskupijas.		
ID	MEDIUMINT	IV vyskupijai identifikuoti.
PAV	VARCHAR(32)	Vyskupijos pavadinimui.
DEKANATAI		
Saugoma informacija apie Lietuvos dekanatus.		
ID	MEDIUMINT	IV dekanatui identifikuoti.
PAV	VARCHAR(32)	Dekanato pavadinimui.
PARAPIJOS		
Saugoma informacija apie Lietuvos parapijas.		
ID	MEDIUMINT	IV parapijai identifikuoti.
PAV	VARCHAR(32)	Parapijos pavadinimui.
VALSCIAI		
Saugoma informacija apie Lietuvos valsčius.		
ID	MEDIUMINT	IV valsčiui identifikuoti.
PAV	VARCHAR(32)	Valsčiaus pavadinimui.
SRITYS		
Saugoma informacija apie Lietuvos sritis.		
ID	MEDIUMINT	IV sričiai identifikuoti.
PAV	VARCHAR(32)	Srities pavadinimui.
KALBOS		
Saugoma informacija apie vartojimo konteksto kalbas.		
ID	MEDIUMINT	Kalbai identifikuoti.

PAV	VARCHAR(32)	Kalbos pavadinimui.
PRIEDELIAI		
Saugoma informacija apie priešdėlius.		
ID	INT	Vietovardžio priešdėliui identifikuoti.
PAV	VARCHAR(100)	Priešdėlio pavadinimui.
PRIESAGOS		
Saugoma informacija apie priesagas.		
ID	INT	Vietovardžio priesagai identifikuoti.
PAV	VARCHAR(100)	Priesagos pavadinimui.
GALUNES		
Saugoma informacija apie galūnes.		
ID	INT	Vietovardžio galūnei identifikuoti.
PAV	VARCHAR(100)	Galūnės pavadinimui.
VARTOTOJAI		
Saugoma informacija apie sistemos vartotojus.		
ID	INT	Vartotojui identifikuoti.
PRISIJ_VARDAS	VARCHAR(50)	Vartotojo prisijungimo vardui.
SLAPTAZODIS	VARCHAR(32)	Vartotojo slaptažodžiui.
VARDAS	VARCHAR(50)	Vartotojo vardui ir pavardei.
E_PASTAS	VARCHAR(100)	Vartotojo slaptažodžiui.
GRUPE_ID	INT	Vartotojų grupei identifikuoti.
AKTYVUMAS	TINYINT	Ar leisti vartotojui prisijungti prie sistemos.
GRUPES		
Saugoma informacija apie sistemos vartotojų grupes.		
ID	INT(10)	Vartotojų grupei identifikuoti.
PAV	VARCHAR(50)	Vartotojų grupės pavadinimui.
TIPAS	ENUM('ALLOW', 'DENY')	Grupės teisių tipui.
GRUPIU_TEISES		
Saugoma informacija apie sistemos vartotojų grupių teises.		
GRUPE_ID	INT	Vartotojų grupei identifikuoti.
TEISE_ID	INT	Prieigos teisei identifikuoti.
TEISES		
Saugoma informacija apie sistemos prieigos teises.		
ID	INT	Prieigos teisei identifikuoti.
PAV	VARCHAR(50)	Prieigos teisės pavadinimas.
REZULTATAI		
Saugoma informacija apie sąrašui priklausančius dabartinius vietvardžius.		
ID	INT	Rezultatui identifikuoti.
SARASAS_ID	INT	Sąrašui identifikuoti.

VIETOVARDIS_ID	INT	Vietovardžiui, priklausančiam sąrašui identifikuoti.
SARISTORINIAI		
Saugoma informacija apie sąrašui nepriklausančius istorinius vietovardžius.		
ID	INT	Rezultatui identifikuoti.
SARASAS_ID	INT	Sąrašui identifikuoti.
IVIETOVARDIS_ID	INT	IV nepriklausančiam sąrašui identifikuoti.
SARASAI		
Saugoma informacija apie tyrimų medį.		
ID	INT	Tyrimo medžio šakai-sąrašui identifikuoti.
SARASAS_ID	INT	Sąrašui, iš kurio suformuotas naujas sąrašas, identifikuoti.
DATA	DATETIME	Sąrašo suformavimo datai
LAUKAI	VARCHAR(200)	Dinaminėje rezultatų lentelėje rodomiems laukams.
KRITERIJAI	TEXT	Pagal kokius kriterijus ieškota.
PASTABOS	TEXT	Pastaboms apie tyrimą.
TYRIMAI		
Saugoma bendrinė informacija apie tyrimą.		
ID	INT	Tyrimui identifikuoti.
VARTOTOJAS_ID	INT	Vartotojui, atlikusiam tyrimą, identifikuoti.
PAV	VARCHAR(30)	Tyrimo pavadinimui.
DATA	DATETIME	Tyrimo datai.
VIESUMAS	TINYINT	Ar tyrimas viešas, ar privatus.
BUSENA	TINYINT	Tyrimo būsenai
APRASYMAS	TEXT	Tyrimo aprašymui.
SARASAS_ID	INT	Paskutiniam suformuotam sąrašui identifikuoti.

2.7. Pasirinktos realizacijai priemonės

Programinei įrangai realizuoti buvo pasirinktos atvirojo kodo (nemokami) priemonės: CakePHP karkasas, parašytas PHP programavimo kalba, MySQL duomenų bazės valdymo sistema bei Google geografinė informacinė sistema. Žemiau pateikiami esminiai argumentai, kodėl buvo pasirinktos būtent tokios priemonės.

2.7.1. CakePHP karkasas

MVC architektūrą palaiko CakePHP karkasas (framework), parašytas PHP programavimo kalba. Jis pasižymi šiomis savybėmis [24]:

- Lengvai įdiegiami pagalbininkai (helpers) HTML, formoms, puslapiavimui, AJAX, Javascript;
- Tvarkinga URL hierarchija;

- Integruoti sesijos, užklausų palaikymo komponentai;
- Pagalbinės klasės darbui su failais, katalogais, masyvais ir t.t.;
- Pritaikomumas kiekvienai pasaulio kalbai (visiškas UTF-8 palaikymas);
- Lankstus kešavimas: naudojamos memcache, apc, xcache aplikacijos sukurtos sistemos darbo paspartinimui;
- Palaikomas ir PHP4, ir PHP5;
- Išsami pagalba kūrėjams internete.

CakePHP karkaso veikimui nebūtini specialūs Apache nustatymai. Instaliavus pats karkasas praneša apie trūkstamus sistemos komponentus, ar neteisingus duomenų bazės parametrus. Aplikacijos kūrimas bei integracija su Google geografine informacine sistema vyksta pakankamai greitai. Norint išplėsti Karkaso galimybes, būtina aprašyti reliacinius ryšius tarp duomenų bazės elementų.

2.7.2. MySQL DBVS

Renkantis duomenų bazę, negalima apsiriboti vien savybių aibe. Reikia apsvarstyti dokumentaciją, palaikymą, kainą, stabilumą, bendrą veikimą bei greitį. Kadangi projektuojamai programinei įrangai netinka mokamos alternatyvos, be to pasirinkta programavimo kalba PHP buvo renkama tarp MySQL ir PostgreSQL. Buvo pasirinkta MySQL duomenų bazė. Atvirojo kodo MySQL duomenų bazė buvo suprojektuota būti greita duomenų baze [10]. Maksimalus MySQL DB dydis (neribotas), lentelės dydis (16TB) ir kiti parametrai [11] patenkina projektuojamos programinės įrangos poreikius. Renkantis MySQL, itin svarbų vaidmenį vaidina labai didelė ir aktyvi duomenų bazės vartotojų ir kūrėjų bendruomenė [12].

MySQL duomenų bazė pasirenkama dar ir dėl šių kriterijų:

- Lengvai naudojama ir kuriama [13],[14];
- Daug labiau paplitusi, todėl nesunku rasti talpinimo serverį;
- Greitas ir stabilus veikimas [14];
- Patikimumas [14];
- Jau yra sėkmingai pritaikyta panašiam projekte [3].

Pasirinkus MySQL kaip DBVS, nemažiau svarbus aspektas yra pasirinkti duomenų saugojimo modulį. Dažniausiai vartojami moduliai MyISAM ir InnoDB. Dėl geresnio resursu išnaudojimo, bei mažesnių serverio apkrovimo, taip pat dėl galimybės naudoti tekstinius duomenų bazės laukus, buvo pasirinktas MyISAM, nors InnoDB ir leidžia

kurti ryšius tarp lentelių, yra realizuotos transakcijų ir trigerių galimybės. Ryšiai tarp lentelių iš dalies bus realizuoti CakePHP karkaso pagalba.

2.7.3. Google Maps GIS

Renkantis geografinę informacinę sistemą atlikta pagrindinių žemėlapių produktų analizė ir pritaikomumas darbui buvo paremti galimybių, kainos ir patogumo kriterijais:

ArcGIS [15]. Mokama programinė įranga. Metus galiojančios licenzijos bei brangūs apmokymo kursai. Funkcionalumu lenkia visus nemokamus sprendimus.

Yahoo Maps [16]. Nemokamas Flash technologija paremtas žemėlapis. Yra galimybė naudotis žemėlapiu programuotojams savo svetainėje, tačiau jam vis dar stinga tarptautinio palaikymo [19].

Microsoft's Local Live Map [17]. Panašios galimybės kaip ir Google Maps, tačiau Lietuvos žemėlapių detalumas ir tikslumas nepakankamas.

Google Maps [18]. Nemokamas. Atvirojo kodo programinė įranga. Išsami dokumentacija, galimybė naudoti žemėlapi savo svetainėje. Galimybė kurti savus žemėlapius, žymėti vietovę. Lengva integracija į kuriamą programinę įrangą.

Google Maps sistemoje įvairių regionų detalumas yra skirtingas, tačiau duomenys yra pastoviai atnaujinami. Jis priklauso to, kokio populiarumo ir dydžio yra miestas ar vietovė. „Daugelis pasaulio vietų nufotografuota 15 m raiška, didesni miestai 0,61 m raiška ir tik nedaugelis vietų nufotografuota 0,15 m raiška“ [20]. Lietuvos žemėlapiai yra pakankamai detalūs, be to galima tikėtis ateityje jų detalumas tik didės. Google Maps leidžia įsidėti ir savus žemėlapius [22]. Ši geografinė informacinė sistema yra palaikoma visų šiuo metu plačiau naudojamų internetinių naršyklų. Palaikomo naršyklės (pateikta minimali naršyklės versija) : Firefox 0.8+, Mozilla 1.4+, Netscape 7.1+, Internet Explorer 5.5+, Safari 1.2+, Opera 7+. Google Maps vartotojams nekomerciniams tikslams siūlo naudoti Google Maps API [21] – funkcijų rinkinį, kuris leidžia praplėsti Google Maps funkcionalumą. Visas API funkcijas galima naudoti Javascript skriptuose.

Palyginus visas nagrinėtas sistemas buvo pasirinkta Google Maps sistema. Pagrindinis kriterijus: vienintelė šiuo metu tinkama nemokama sistema.

2.8. Statistinė automatizuotų vietovardžių tyrimų PĮ informacija

Lentelėje pateikiama realizuotos automatizuotų vietovardžių tyrimų programinės įrangos statistinė informacija.

Lentelė 3 Automatizuotų vietovardžių tyrimų PĮ statistika

Pavadinimas	Kiekis
1. Programinės įrangos klasės	45
2. Programinės įrangos kodo eilutės	8200
a) Vartotojų duomenų valdymo posistemė;	450
b) Vietovardžio duomenų valdymo posistemė;	4400
c) Tyrimo valdymo posistemė;	2750
d) Grafinis sistemos aprašymas.	600
3. Duomenų bazės lentelių kiekis	26

3. AUTOMATIZUOTO VIETOVARDŽIŲ TYRIMO MODELIS

3.1. Dabartinio vietovardžių tyrimo modelio automatizavimo galimybės

Dabartinis vietovardžių tyrimo procesas gali būti automatizuotas taip, kad ženkliai sutaupytų laiką ir sąnaudas jam atlikti. Toliau apžvelgsime tyrimo etapus ir galimybes, panaudojant informacines sistemas, sumažinti tyrimo sąnaudas.

Tyrimo medžiagos surinkimas ir vietovardžių normavimas.

Šiuo metu beveik visa medžiaga tyrimams yra popieriniame pavidale. Todėl jos surinkimas užima daug laiko, ne visada reikalinga tyrimo medžiaga yra lengvai prieinama, dažnai skirtingi tyrimui reikalingi šaltiniai išsibarstę po visą Lietuvą. Norint išvengti pakartotino šaltinių rinkimo ir analizės, būtina juos tinkamai normalizuoti, suvedant į bendrą duomenų bazę, prieinamą visiems tos srities mokslininkams. Palaipsniui pildant duomenų bazę, atsiras galimybė atlikti vietovardžių tyrimus naudojant tokias duomenų apimtis kurios iki šiol nebuvo naudojamos tokiems tyrimams atlikti.

Bendros duomenų bazės sukūrimas yra būtina tyrimo automatizavimo sąlyga.

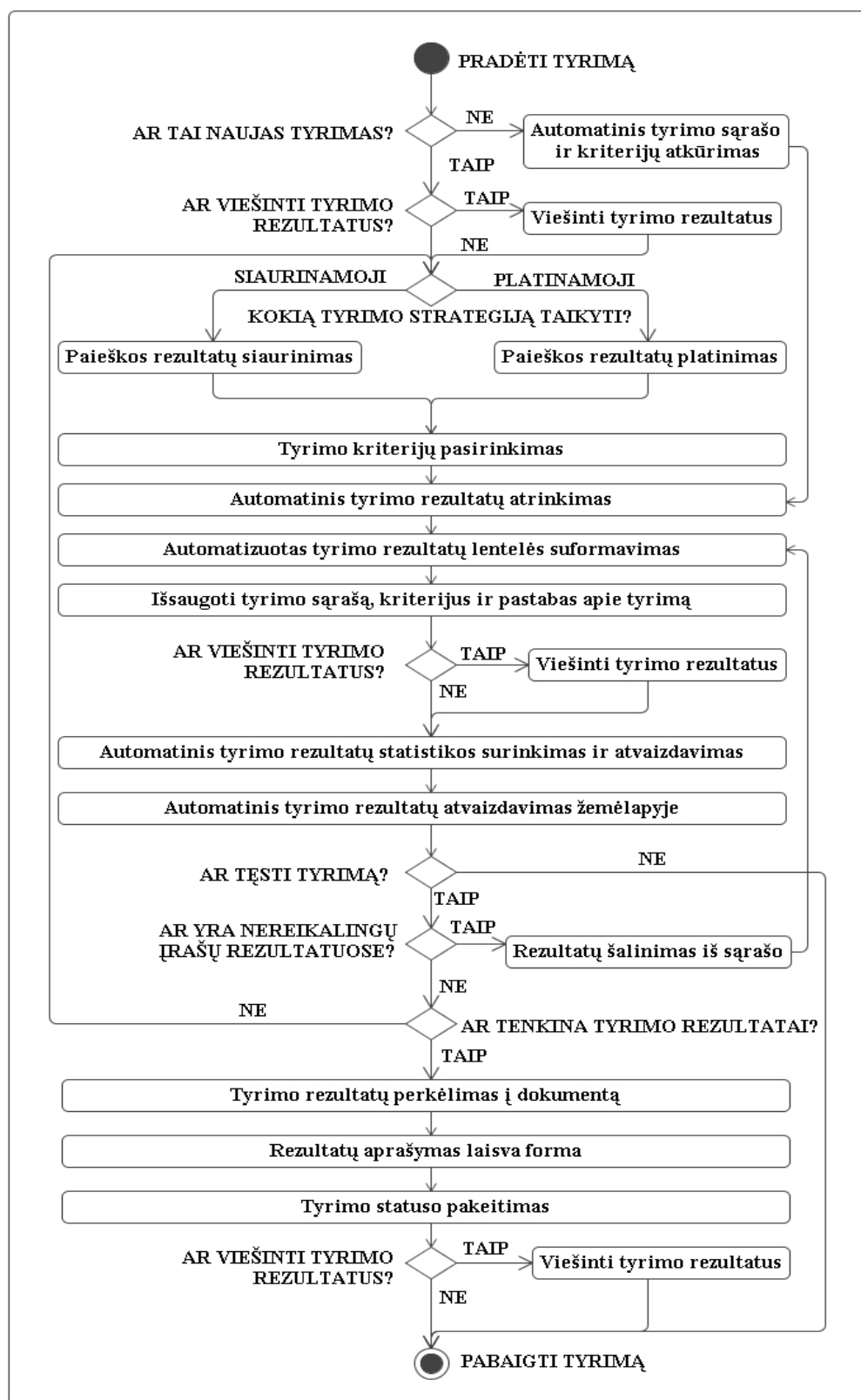
Tyrimo eiga.

Sudarius tokią duomenų bazę, labai daug laiko sutaupoma automatizuojant paieška tarp duomenų. Paieškos automatizavimui būtina sukurti įrankius, kurie atliktų paiešką pagal kuo daugiau kriterijų. Iki šiol tokios paieškos užimdavo daug laiko, nes buvo atliekamos rankiniu būdu, iš kartotekos suvedant duomenis. Reikia pažymėti, kad rankiniu būdu atliekamą paiešką nevisada gali atitikti formaliai aprašyti paieškos kriterijai. Todėl atlikus automatinę paiešką dažnai reikia peržiūrėti ar nėra ten nereikalingų rezultatų. Taip pat egzistuoja rizika, kad kai kurie klaidingai normalizuoti vietovardžiai nebus įtraukti į sąrašą.

Tyrimo tarpinių rezultatų išsaugojimas ir tęstinumas vienas svarbiausių aspektų atliekant automatizuotus vietovardžių tyrimus. Ši galimybė neegzistuoja jokiose kitose nagrinėtose sistemose. Naudojant pasiūlytą sprendimą mokslininkai ne tik galėtų išsaugoti tyrimo rezultatus, tačiau bet kuriuo metu tyrimą tęsti nuo anksčiau gautų rezultatų. Taip pat mokslininkai gali tęsti viešus vienas kito tyrimus, kas skatina mokslininkų bendradarbiavimą tyrimo metu.

3.2. Automatizuotų vietovardžių tyrimų modelis

Atsižvelgiant į automatizavimo galimybes buvo sudarytas automatizuotų vietovardžių tyrimų modelis(žr. Pav. 14).



Pav. 14 Automatizuotų vietovardžių tyrimų modelis

Žemiau esančioje lentelėje išsamiau paaiškintos tyrimą sudarančios dalys.

Lentelė 4 Lietuvos toponimikos tyrimų PĮ statistika

Dalis	Paaiškinimas
-------	--------------

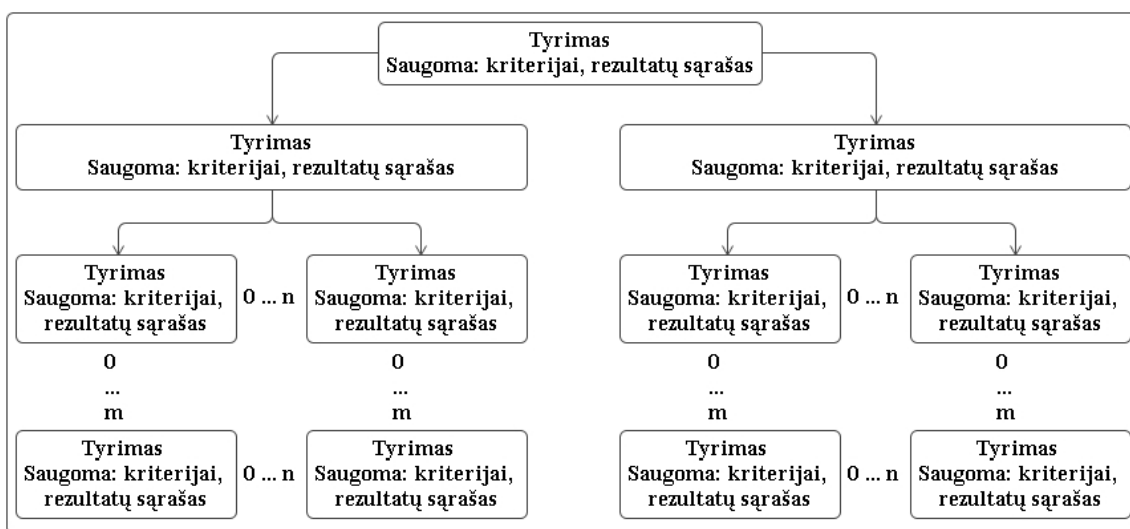
Automatinis tyrimo sąrašo ir kriterijų atkūrimas	Duomenų bazėje saugomas visas sąrašas, jis neatkuriamas pagal kriterijus. Kriterijai atkuriami tik tam, kad galima būtų matyti kokie paieškos kriterijai jau taikyti konkrečiam sąrašui gauti.
Viešinti tyrimo rezultatus	Viešinamas tyrimas. Nuo to laiko jį gali peržiūrėti ir kiti tyrėjai. Galima ir redagavimo galimybė.
Paieškos rezultatų siaurinimas	Paieškos strategija, kai ieškoma tarp jau išrinktų vietovardžių.
Paieškos rezultatų platinimas	Paieškos strategija, kai esamą sąrašą papildo naujos paieškos rezultatai.
Tyrimo kriterijų pasirinkimas	Paieškos kriterijų pasirinkimas. Visi paieškos kriterijai sujungiami loginiu operatoriumi „IR“.
Automatinis tyrimo rezultatų atrinkimas	Vykdoma paieška.
Automatizuotas tyrimo lentelės sudarymas	Lentelėje atvaizduojami tyrėjo nurodyti vietovardžių atributai.
Išsaugoti tyrimo sąrašą, kriterijus ir pastabas apie tyrimą	Gautas naujas sąrašas išsaugomas duomenų bazėje. Saugomi paieškos duomenys bei papildomos tyrėjo pastabos ir išvados.
Automatinis tyrimo rezultatų statistikos surinkimas ir atvaizdavimas	Grafikų sudarymas.
Automatinis tyrimo rezultatų atvaizdavimas žemėlapyje	Surandami visi rezultatų sąrašė esantys lokalizuoti vietovardžiai, jie atvaizduojami žemėlapyje.
Rezultatų šalinimas iš sąrašo	Tyrėjai savo nuožiūra šalina rezultatus iš sąrašo.
Tyrimo rezultatų perkėlimas į dokumentą	Tyrimo rezultatai išsaugojami tekstiniame dokumente
Rezultatų aprašymas laisva forma	Laisva forma aprašoma tyrimo eiga, rezultatai išvados.
Tyrimo statuso pakeitimas	Tyrimo pažymimas kaip baigtas.

Automatizuotas tyrimo modelis leidžia vietovardžių tyrimą skaidyt į kelias atskiras tyrimo šakas, jas atskirai analizuoti. Tokiu būdu sudaromas tyrimo medis (žr. Pav. 15). Tyrimo medžio esmė yra tai, kad bet kurio tyrimo momentu galima tęsti tyrimą nuo bet kurio to tyrimo taško. Pagal saugomus kriterijus, visada galima nustatyti, kaip gautas vienas ar kitas rezultatų sąrašas. Tai labai svarbų tęsiant tyrimą po didelės pertraukos. Be to tyrėjas ir pats gali prirašyti, savo pastabas prie kiekvieno rezultatų sąrašo.

Pagrindiniai automatizuotų vietovardžių tyrimų ir dabartinių vietovardžių tyrimų modelio skirtumai:

- Rezultatų atkūrimas. Bet kuriuo metu galima atkurti bet kurią tyrimo būseną
- Tyrimo viešinimas. Galima paviešinti tyrimą, jo rezultatais naudotis, ir jį tęsti galės visi tyrėjai.

- Tyrimo kriterijai galimi tik formalūs. Vienos paieškos metu visi kriterijai sujungti loginiu „IR“ pagrindu.
- Suformuojama lentelė su iš anksto tyrėjo pasirinktais rodomais laukais.
- Automatizuotas statistikos surinkimas ir atvaizdavimas galimas kiekviename tyrimo žingsnyje. Dabartinių tyrimų modelyje, šis punktas atliekamas tyrimo gale, nes kiekvienam žingsniui yra per sudėtingas.
- Šiame modelyje atsiranda atvaizdavimas žemėlapyje, tai labai svarbus šio modelio privalumas, leidžiantis greitai nustatyti tiriamų vietovardžių pasiskirstymą teritorijoje, jeigu jie lokalizuoti.



Pav. 15 Tyrimų medis

Sukurtas automatizuotų vietovardžių tyrimo modelis leidžia tyrimus atlikti greičiau, automatizuodamas daug laiko dabar užimančius procesus. Taip pat, dėl galimybės atkurti ankstesnius tyrimo rezultatus, bei dėl neribojamo duomenų kiekio, galima atlikti tokius tyrimus kokių iki šiol buvo neįmanoma atlikti.

3.3. Programinės įrangos tobulinimo galimybės

Didėjant vartotojų bei duomenų kiekiui, atsiranda poreikis tą PI tobulinti, įvedant daugiau funkcijų, gerinant vartotojo sąsają. Taip pat atsirandant naujoms interneto technologijoms, galimi ir architektūriniai pokyčiai. Daugumą šių dalykų yra neįmanoma numatyti šiuo metu, tačiau jau dabar galima numatyti keletą patobulinimų, išplėsiąčių programinės įrangos galimybes:

- Sukurti statistikos surinkimo ir grafinio atvaizdavimo tyrimo rezultatams posistemę.

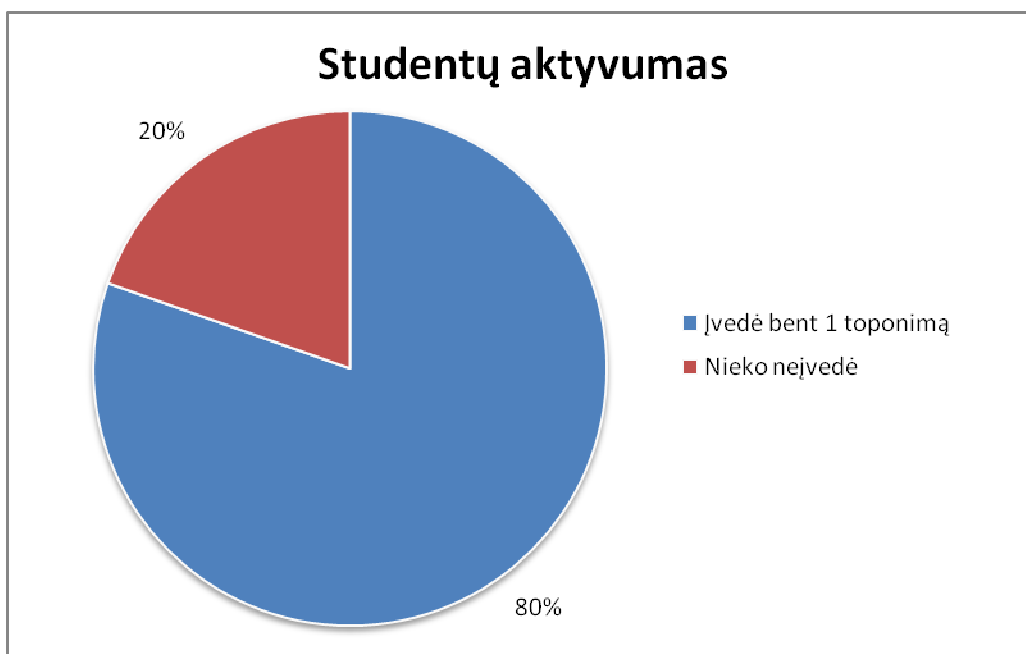
- Galimybė vietovardį lokalizuoti ne tik kaip tašką žemėlapyje, bet ir liniją arba plotą.
- Žemėlapyje atvaizduoti kelias tyrimo šakas, palyginti jų teritoriją. Tyrimo rezultatų sąrašo teritorijas galima apibrėžti.
- Patobulinti automatizuoto tyrimo modelį, leidžiant pagrindinius aibių veiksmus (sajunga, sankirta, skirtumas) tarp skirtingų tyrimų rezultatų aibių.
- Sukurti mokslininkų bendradarbiavimui skirtas priemones.
- Praplėsti tyrimo viešinimo galimybes: leisti viešinti tyrimą tik nurodytiems tyrėjams, paskirti jiems redagavimo ar tik peržiūros teises.
- Integruoti istorinius žemėlapius į GIS sistemą.
- Galimybė išsaugoti rezultatų lentelę MS Word ar MS Excel failų formatais.

4. VIETOVARDŽIŲ TYRIMŲ PROGRAMINĖS ĮRANGOS EKSPERIMENTINIS TYRIMAS

4.1. Vietovardžių panaudojimo mokymo procese analizė

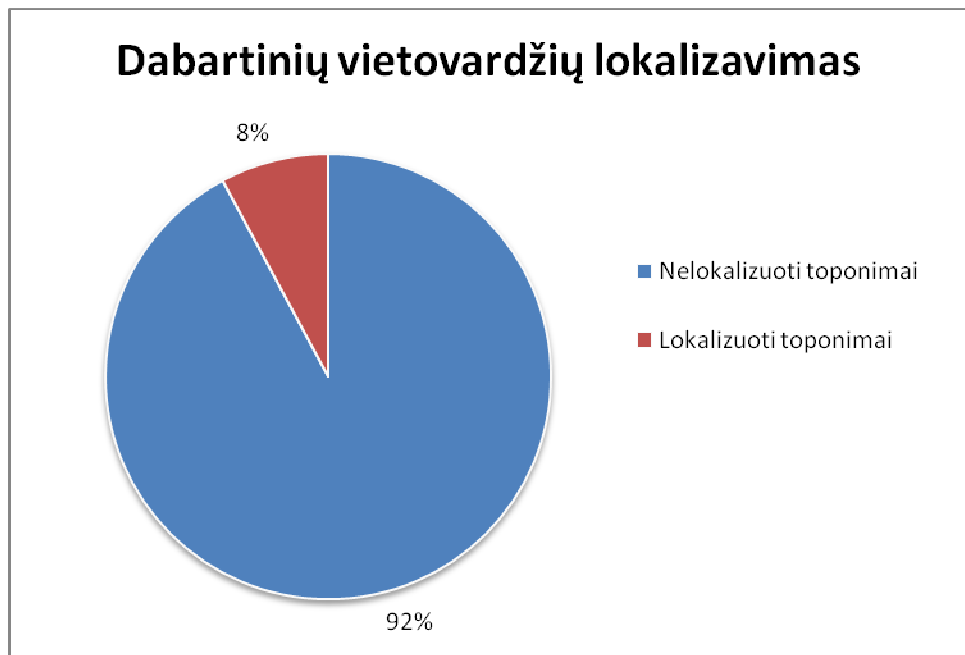
Tyrime dalyvavo VU Lietuvių kalbos reklamos studijų profesinės kalbos kultūros modulio studentai. Jie suvedinėjo informaciją apie vietovardžius į duomenų bazę. Šio darbo rašymo metu tyrimas dar nebuvo baigtas, tačiau esami rezultatai jau suteikia informacijos apie naudojimąsi sukurta PĮ.

Sistemoje buvo užregistruotas 51 studentas. 2009 m. vasario – gegužės mėnesiais, nors viena vietovardį įvedė 41 studentas (žr. Pav. 16). Vidutiniškai vienas studentas buvo įvedęs po 2.2 vietovardžio. Daugiausiai – 5, įterpė tik vienas tyrime dalyvavęs studentas.



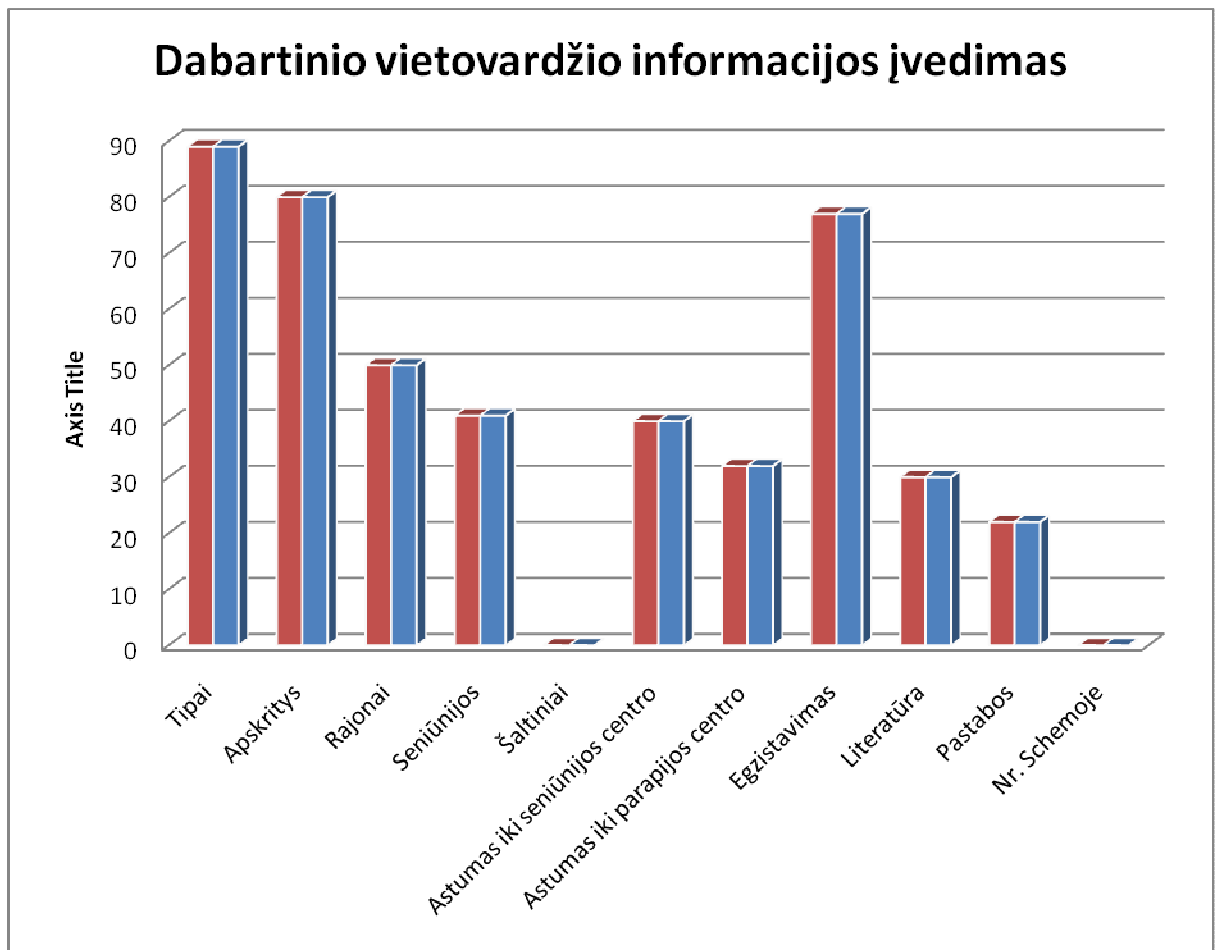
Pav. 16 Studentų aktyvumo diagrama

Vienas svarbiausių vietovardžio atributų yra jo lokacija. Iš visų 89 įvestų vietovardžių lokalizuoti pavyko vos 7 vietovardžius (žr. Pav. 17). Tai rodo, kad vietovardžio lokalizavimas studentams yra viena iš sudėtingiausių užduočių, tačiau įmanoma.



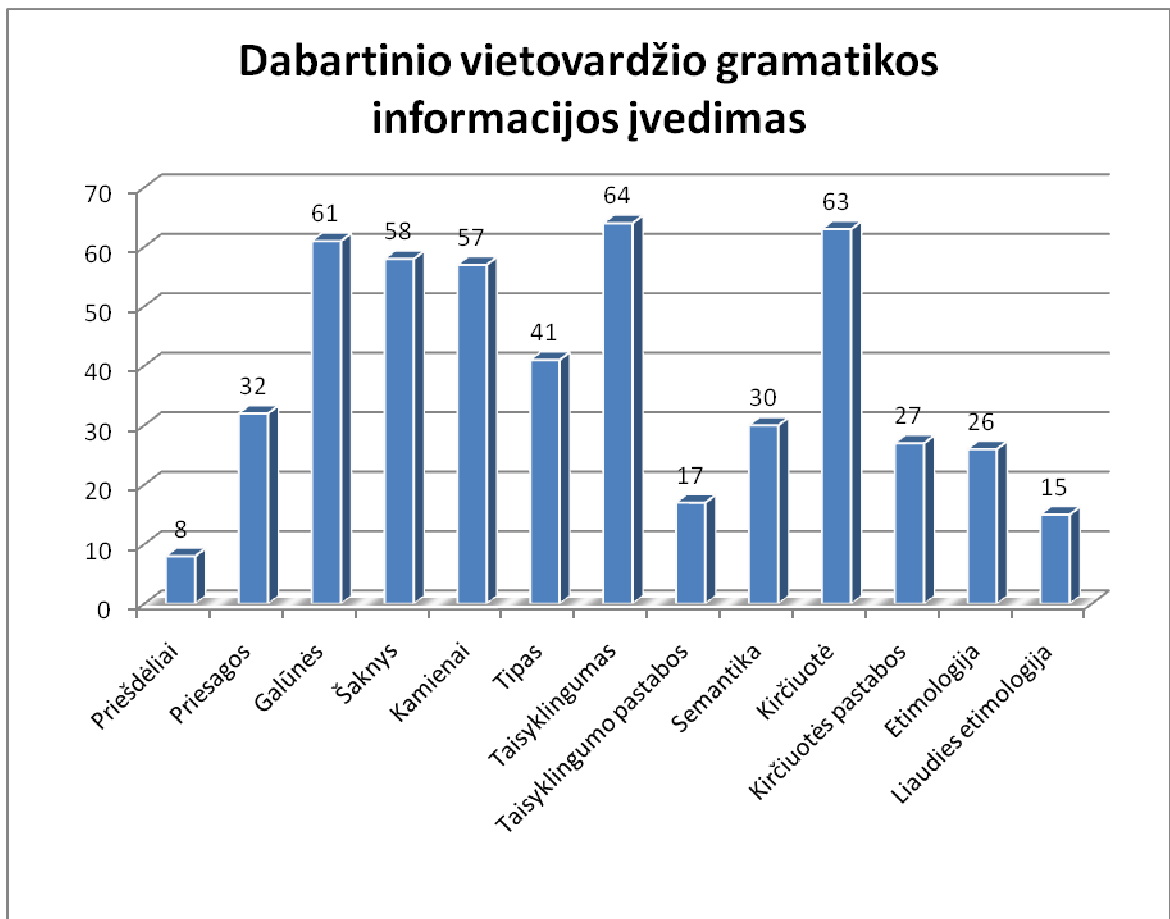
Pav. 17 Dabartinių toponimų lokalizavimo diagrama

Informacijos apie dabartinį vietovardį įvedime (žr. Pav. 18), studentai geriausiai įvedinėjo vietovardžio tipą, egzistavimą ir vietovės informaciją. Beveik visi studentai nurodė apskritį, psė jų nustatė rajoną ar seniūniją. Labai svarbus šaltinio atributas nebuvo įvestas nė vienam vietovardžiui. Tai gali būti užduoties formulavimo klaida, tačiau gal būt šaltinio atributą reikėtų labiau pabrėžti ir vartotojo sąsajoje. Gali būti jog daugelis studentų supainiojo šaltinį su literatūros atributu. Kitas neinvestas atributas Nr. schemeje. Šis atributas skirtas tik iš specifinių šaltinių suvedamiems vietovardžiams.



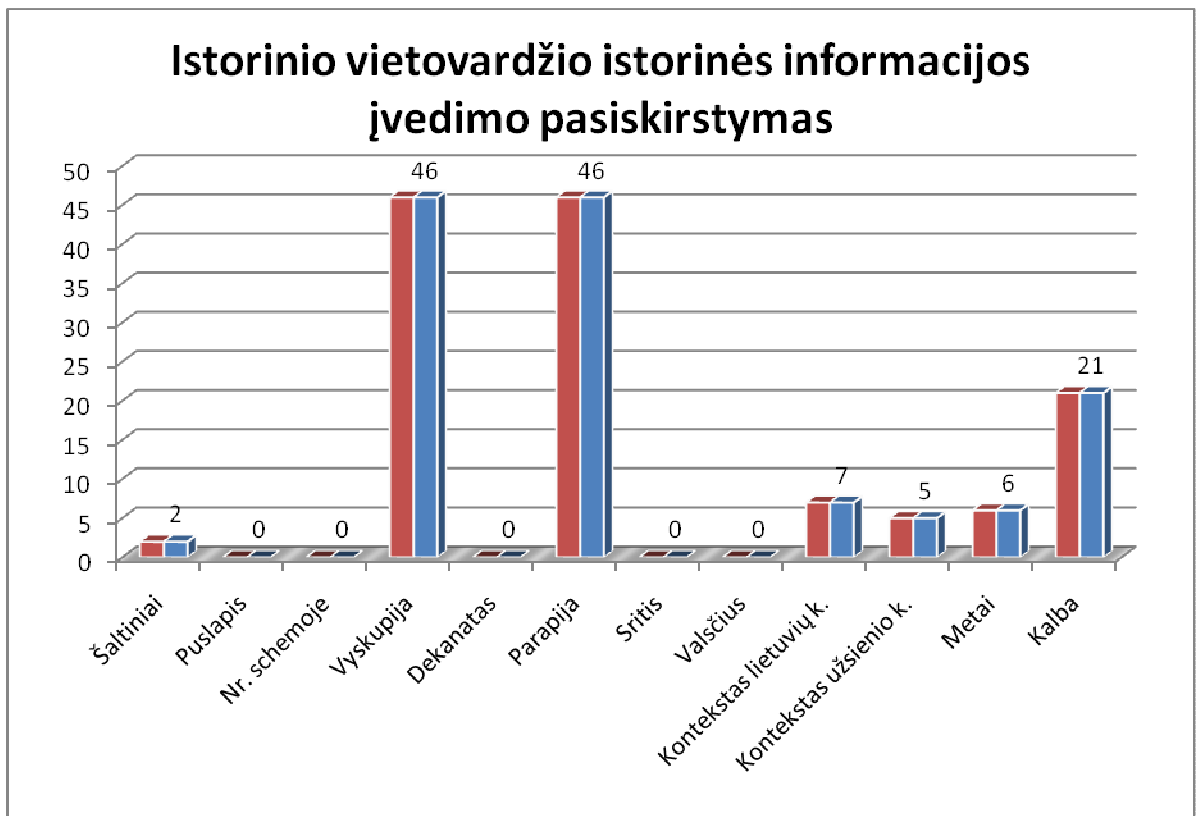
Pav. 18 Dabartinio vietovardžio informacijos įvedimo diagrama

Gramatikos informacija apie dabartinį vietovardį (žr. Pav. 19) buvo įvedama pakankamai išsamiai. Iš visų 64 vietovardžių, net 63 turėjo nustatytą kirčiutę. Tai galima paaiškinti dėstomo modulio specifika. Beveik visi įvesti vietovardžiai buvo tinkamai sukirčiuoti. Taisyklingumo atributas yra privalomas, ir jei nenurodoma kitaip, visi vietovardžiai pažymimi, kaip taisyklingi. Studentai taip pat gana tiksliai suvedinėjo gramatikos informacija, apie 60 vietovardžių buvo išskaidyta dalimis. Tačiau skirtingas skaičius vietovardžių su nurodytomis šaknimis, kamienais ir galūnėmis rodo, kad kai kurie vietovardžiai yra išnagrinėti ne iki galo.



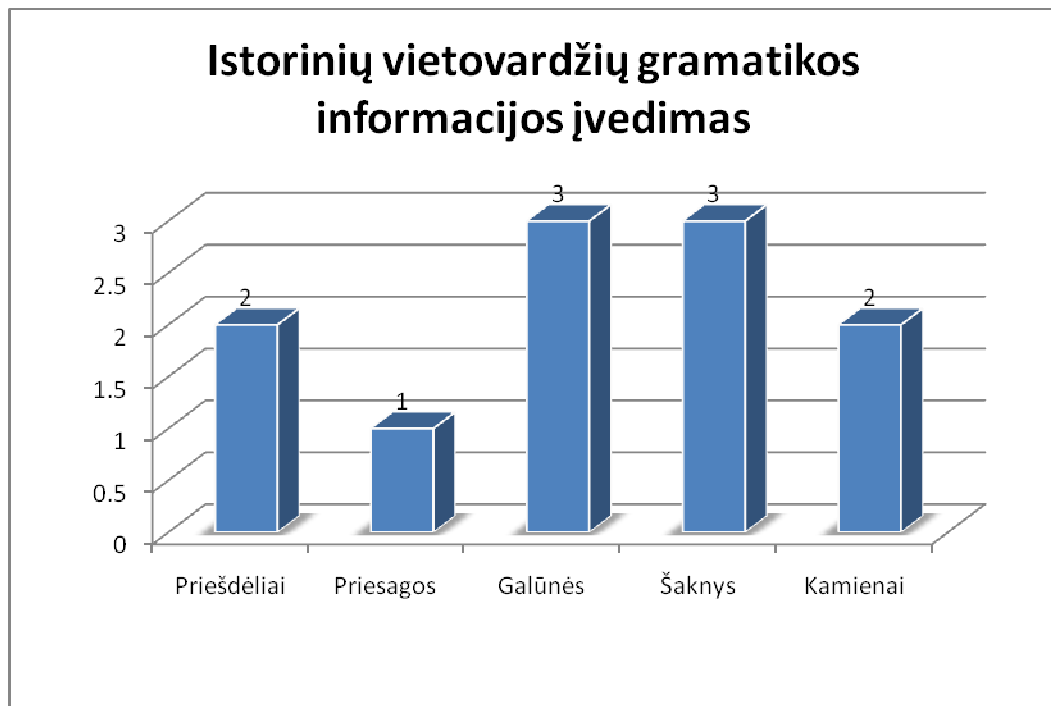
Pav. 19 Dabartinio vietovardžio gramatikos įvedimo diagrama

Su dabartiniais vietovardžiais buvo susieti 67 istoriniai vietovardžiai. Su kai kuriais iš jų, buvo susieti ne vienas, o daug istorinių vietovardžių. Tačiau apie pusę vietovardžių neturėjo istorinio atitikmens. Geriausiai istoriniams vietovardžiams buvo nustatinėjama parapija. Tai rodo, šio atributo svarbumą istorinių vietovardžių tyrimuose. Vyskupijos atributas taip pat buvo nurodytas prie tų pačių istorinių vietovardžių, tačiau šiuo atveju jis nėra labai svarbus, nes beveik visi įvesti vietovardžiai priklausė Kauno vyskupijai.



Pav. 20 Istorinių vietovardžių istorinės informacijos įvedimo pasiskirstymo diagrama

Skirtingai nuo dabartinių vietovardžių, studentai nesuvedė daug gramatinės informacijos apie vietovardžius. Nors iš viso šešiams vietovardžiams buvo pažymėta bent kažkokia gramatikos informacija, tačiau vos vienas vietovardis išnagrinėtas išsamiai.



Pav. 21 Istorinių vietovardžių gramatikos informacijos įvedimo diagrama

Išvados.

Sistema tinkama vietovardžių įvedimui. Tačiau tyrimo metu numatyta nemažai vartotojo sąsajos pakeitimų, kad duomenis būtų įmanoma įvesti lengviau. Vartotojo sąsajoje būtina pabrėžti reikalingiausius atributus, o ne taip reikalingus atvaizduoti taip, kad netrukdytų įvedimo procesui. Mažiau reikalingus atributus leisti įvesti, tik jei to nori pats vartotojas.

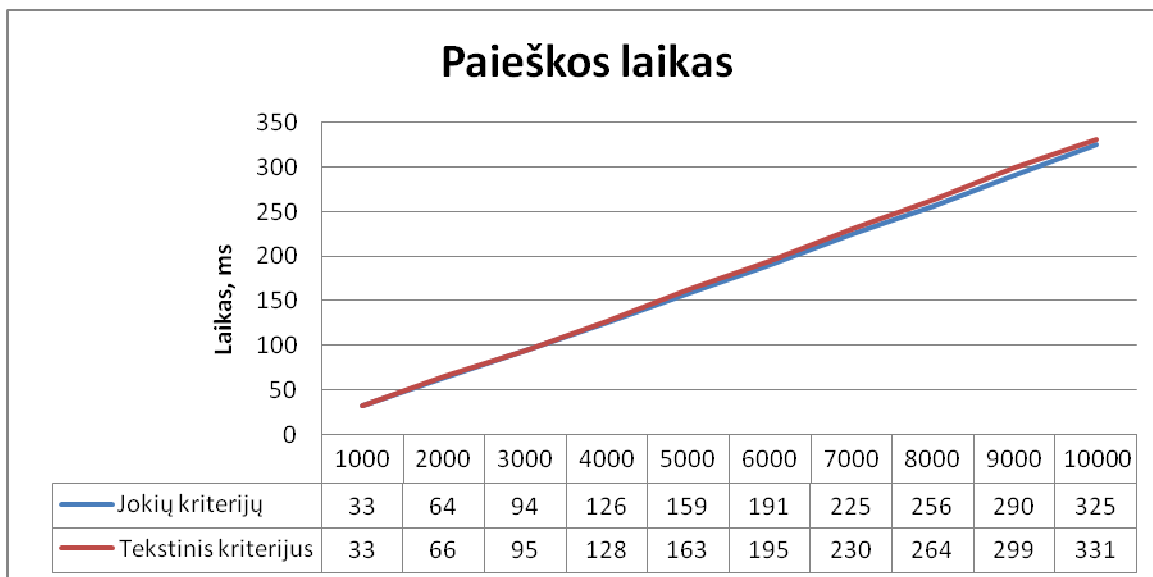
4.2. Paieškos užklausų tyrimas

Realizavus sistemą, labai svarbu numatyti jos galimybes, ir pagrindinius jas ribojančius veiksnius. Įkėlus turimus realius duomenis, pastebėta, kad paieškos laikas, nors ir priimtinas, tačiau pakankamai didelis. Esant didesniems duomenų kiekiams tai ribotų sistemos galimybes. Todėl nuspręsta jį iširti eksperimento būdu.

Eksperimentams realizuoti buvo sukurta specialūs programinės įrangos papildymai, kurie leido generuoti atsitiktinių duomenų kiekius, ir jais papildydavo duomenų bazę. Programinė įranga buvo tiriama ne realioje savo veikimo aplinkoje (interneto serveryje), tačiau vartotojo kompiuteryje, įrašius serverio programinę įrangą.

1 eksperimentas. Į duomenų bazę, įterpti po 1000 vietovardžių, vykdomas jų paieškos laikas, kai paieškos kriterijus nenurodomas, arba kai nurodomas tekstinis paieškos kriterijus. Fiksuojami užklausų vykdymo laikai.

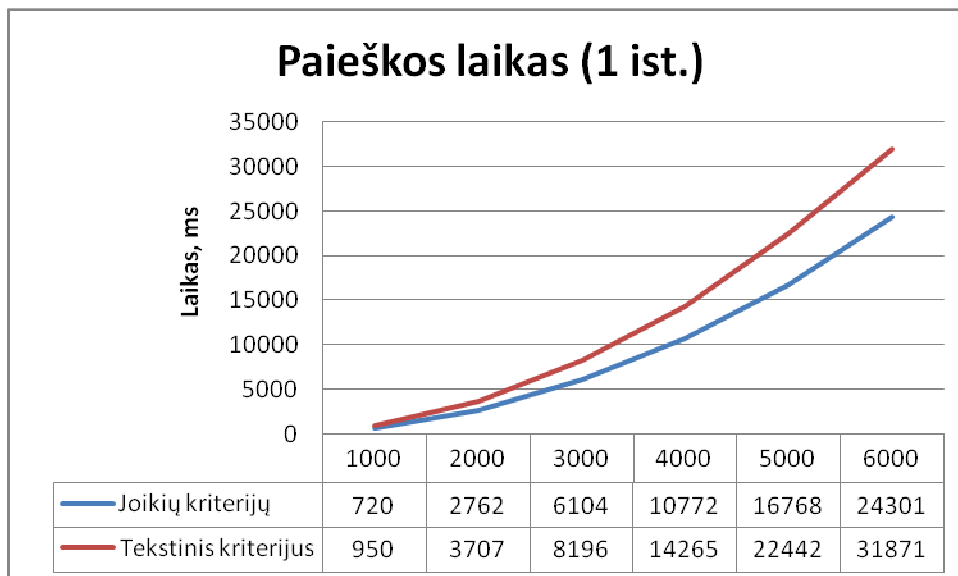
Gauti eksperimento rezultatai pavaizduoti žemiau esančiame grafike:



Pav. 22 Paieškos laikas, kai duomenų bazėje nėra istorinių vietovardžių.

Iš grafiko matyti, kad paieškos laikas tiesiškai priklauso nuo įvestų vietovardžių kiekio. Paieška tarp 10 000 vietovardžių užtruko vos trečdalį sekundės. Įvertinus šiuos aspektus, galima teigti, jog paieška tarp 100 000 vietovardžių truktų šiek tiek daugiau nei tris sekundes. Iš šio eksperimento galima daryti išvadą, kad tekstinis kriterijus minimaliai įtakoja paieškos laiką.

2 eksperimentas. Į duomenų bazę po 1000 bus sugeneruojami vietovardžiai, kiekvienam vietovardžiui taip pat bus sukurtas 1 istorinis vietovardis. Atliekamos taip pat dviejų rūšių paieškos: be kriterijų, ir su tekstiniais kriterijais. Tekstinis kriterijus parenkamas toks, kad atrinktų visus vietovardžius, taip pat ir istorinius. Eksperimento rezultatai žemiau esančiame grafike:



Pav. 23 Paieškos laikas, kai vienas vietovardis susietas su vienu istoriniu vietovardžiu.

Šis eksperimentas kiek netikėtai buvo baigtas, kai DBVS strigdavo bandant paiešką tarp 7000 vietovardžių. Tam galbūt turėjo įtakos aparatinės įrangos konfigūracija, tačiau akivaizdu, kad toks laiko didėjimas greitai pasiektų ir galingesnės įrangos ribas. Išnagrinėjus gautas kreives paaiškėjo, kad kreivės beveik idealios parabolės, taigi naudojamų užklausų sudėtingumas yra $\Omega(n^2)$. Toks užklausų sudėtingumas ir PĮ ribotumas yra nepatenkinamas PĮ vartotojams.

Užklausos naudotos abiejuose eksperimentuose:

Be kriterijų:

```
SELECT `vietovardis`.`id`,`istorinis`.`id` FROM `vietovardziai` AS
`vietovardis` LEFT JOIN `ivietovardziai` AS `istorinis` ON (
`istorinis`.`vietovardis_id` = `vietovardis`.`id` ) LEFT JOIN `gramatikos`
AS `gramatikav` ON ( `vietovardis`.`gramatika_id` = `gramatikav`.`id` )
LEFT JOIN `gramatika2` AS `gramatika2` ON ( `gramatika2`.`vietovardis_id` =
`vietovardis`.`id` ) LEFT JOIN `gramatikos` AS `gramatikai` ON (
`istorinis`.`gramatika_id` = `gramatikai`.`id` )WHERE (true)
```

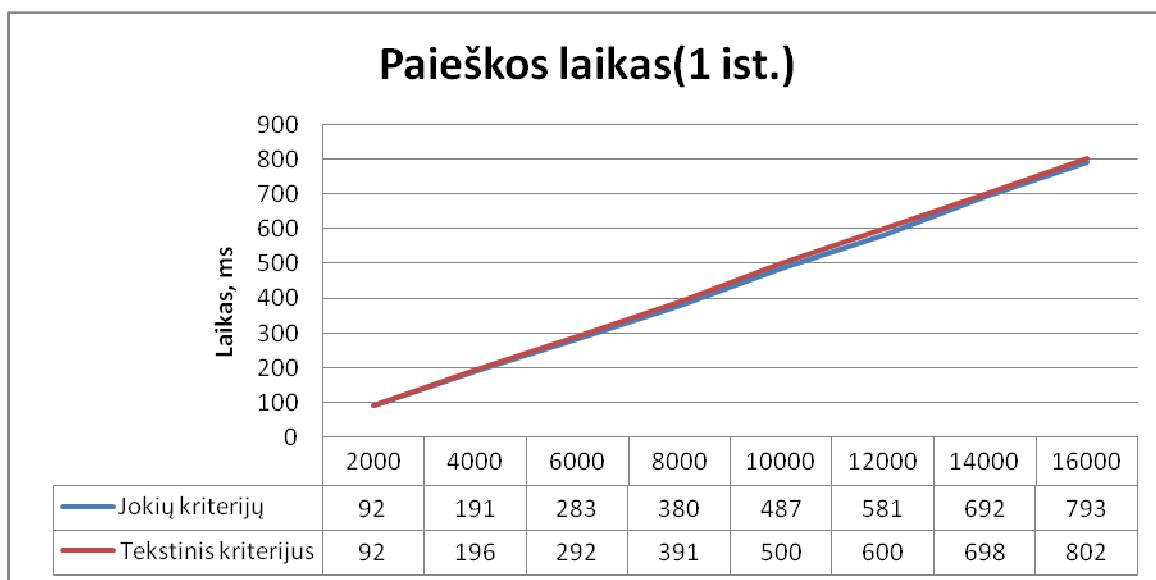
Tekstinis kriterijus:

```
SELECT `vietovardis`.`id`,`istorinis`.`id` FROM `vietovardziai` AS
`vietovardis` LEFT JOIN `ivietovardziai` AS `istorinis` ON (
`istorinis`.`vietovardis_id` = `vietovardis`.`id` ) LEFT JOIN `gramatikos`
AS `gramatikav` ON ( `vietovardis`.`gramatika_id` = `gramatikav`.`id` )
```

```
LEFT JOIN `gramatika2` AS `gramatika2` ON ( `gramatika2`.`vietovardis_id` = `vietovardis`.`id` ) LEFT JOIN `gramatikos` AS `gramatikai` ON ( `istorinis`.`gramatika_id` = `gramatikai`.`id` )WHERE ( `vietovardis`.`pav` LIKE '%a%')or ( `istorinis`.`pav` LIKE '%a%')
```

Išanalizavus aukščiau pateiktas užklaudas, taip pat pirmų dviejų eksperimentų rezultatus, nustatėme kad daugiausiai laiko DBVS sugaišta sujungdama dvi lenteles: vietovardžių, ir istorinių vietovardžių. Norint sutrumpinti šiam procesui skirtą laiką istorinių vietovardžių lentelėje, atributui *vietovardis_id* sukuriamas indeksas (naudojant MySQL su MyISAM saugojimo sistema neįmanoma nurodyti ryšių tarp lentelių, todėl indeksas nebuvo sukurtas).

Trečias eksperimentas. Pradėjus indeksuoti istorinių vietovardžių lentelėje *vietovardis_id* atributą, atliekamos tos pačios užklaudos. Gauti rezultatai grafike žemiau:



Pav. 24 Paieškos laikas, kai vienas vietovardis susietas su vienu istoriniu vietovardžiu, įvestas *vietovardis_id* indeksavimas.

Iš pateikto grafiko galima pastebėti, kad pradėjus indeksuoti *vietovardis_id* atributą paieška pradeda veikt daug greičiau, ir paieškos užklaudos vykdymo laukas yra tiesiškai priklausomas nuo vietovardžių kieki

Gautų rezultatų palyginimas.

Žemiau esančioje lentelėje palyginami visų trijų eksperimentų rezultatai, naudojant tekstinio kriterijaus užklaudas:

Lentelė 5 Lietuvos toponimikos tyrimų PĮ statistika

Vietovardžių kiekis	1 eksperimentas (ms.)	2 eksperimentas (ms)	3 eksperimentas (ms)
1 000	33	950	47
2 000	66	3 707	92
3 000	95	8 196	144
4 000	128	14 265	196
5 000	163	22 442	244

6 000	195	31 871	292
8 0000	264	-	391
10 000	331	-	500

Išvados.

- Tiriant paieškos užklausas buvo rastos ir ištaisytos kai kurios projektavimo metu padarytos klaidos.
- Tekstiniai kriterijai labai nežymiai įtakoja paieškos laiką.
- Lentelėms sujungti DBVS naudoja didžiausią užklausos vykdymo laiko dalį.

4.3. Rezultatų sąrašo atkūrimo iš duomenų bazės tyrimas.

Ši eksperimentas atliktas norint įsitikinti projektavimo metu padaryt prielaida, kad veikimo laikas skirtas sąrašo rezultatų atkūrimui yra tiesiškai priklausomas nuo rezultatų sąrašo skaičiaus.

Rezultatų atkūrimas buvo tikrinamas išvalius duomenų bazę įvedus 10 000 vietovardžių. Eksperimento eiga:

- Sukuriamas naujas tyrimas
- Pirmą tyrimo sąrašą sudaro visi 10000 duomenų bazėje esančių vietovardžių.
- Po to atliekamas siaurinamas tyrimas kas kart atfiltruojant po 1000 vietovardžių.
- Kiekvienai tyrimo šakai(nuo paskutinės iki pirmos) išmatuojamas atkūrimo užklausos laikas.

Gauti rezultatai:



Pav. 25 Paieškos laikas, kai vienas vietovardis susietas su vienu istoriniu vietovardžiu, įvestas *vietovardis_id* indeksavimas.

Iš gautų rezultatų matosi, kad anksčiau daryta prielaida, jog tyrimo rezultatų atkūrimo laikas priklauso nuo vietovardžių kiekio tiesiškai, pasitvirtino.

IŠVADOS

- Dabartinis vietovardžių tyrimo procesas yra neefektyvus, ir neišnaudoja informacinių technologijų teikiamų galimybių.
- Šiuo metu Lietuvoje neegzistuoja sistema tinkanti istorinių vietovardžių tyrimams.
- Sukurtas automatizuoto iteracinio vietovardžių tyrimo modelis.
- Programinė įranga sėkmingai realizuota naudojantis nemokamomis atviro kodo priemonėmis.
- Sukurta programinė įranga tinkama tiek tyrimams atlikti, tiek praktiniam studentų apmokymui.
- Automatizuotų vietovardžių tyrimų programinė įranga, sukurta remiantis MVC projektavimo šablonu, bus nesunkiai palaikoma ir tobulinama.
- Tyrimo metu buvo nustatytos PĮ tobulinimo galimybės.
- Sukurta programine įranga, kaip praktine mokomąja medžiaga, sėkmingai naudojosi Vilniaus universiteto studentai.

LITERATŪRA

- [1] Margarita Jankauskaitė „Būti ar nebūti lyčių lygybei Lietuvoje?“ [Žiūrėta 2007 11 13], prieiga per internetą: <<http://www.lygybe.lt/ci.admin/Editor/assets/buti-nebuti-lyciu-lygybe.margarita.doc>>
- [2] VDU Č. Milošo slavistikos centro parengto projekto „Lietuvos pilietinės tapatybės animacija iššūkio sąlygomis“ aprašymas. 2006.
- [3] Bartkutė N., Jodelis R. „Toponimija skaitmeniniu pavidalu: Joniškio rajono vietovardžių geolingvistinė duomenų bazė“. Konferencijos „Europos studijos: dialektai ir kalbos politika“ leidinys. 2007 m.
- [4] Valstybinės lietuvių kalbos komisijos 2006 m. veiklos ataskaita. [Žiūrėta 2009 03 19], prieiga per internetą: <http://www3.lrs.lt/owa-bin/owarepl/inter/owa/U0162220.doc>
- [5] Vietovardžių žodynas. Pratarinė. [Žiūrėta 2009 05 04], prieiga internete <<http://lkz.mch.mii.lt/Vietovardziai/Pratarne.htm>>
- [6] Toponyms Data Bank. The Research Institute for the Languages of Finland [Žiūrėta 2007 11 13], prieiga internete <<http://kaino.kotus.fi/www/artikkelit/toponyms.html>>
- [7] Toponymic Research System of Northwest Russia: The TORIS System. The Institute of Linguistics, Literature and History of Russia . [Žiūrėta 2007 11 13], prieiga internete <<http://toris.krc.karelia.ru/papers/source/toris98/index.phtml>>
- [8] „Paikannimikartasto“. Kotimaisten kielten tutkimuskeskus. 1997–2002 m. [Žiūrėta 2007 11 13], prieiga internete <<http://www.kotus.fi/index.phtml?s=383>>
- [9] „Paikannimikirja“. Kotimaisten kielten tutkimuskeskus. 2005-2007 m. [Žiūrėta 2007 11 13], prieiga internete <<http://www.kotus.fi/index.phtml?s=443>>
- [10] Open Source DBMS Comparison: MySQL vs. PostgreSQL [Žiūrėta 2007 11 13], prieiga internete <<http://www.tometasoftware.com/MySQL-vs-PostgreSQL.asp>>
- [11] Comparison of relational database management systems [Žiūrėta 2007 11 13], prieiga internete <http://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems>
- [12] Reuven M. Lerner „MySQL Deserves a Double Take“. 2007 m.. [Žiūrėta 2007 11 13], prieiga internete <<http://www.linuxjournal.com/article/9571>>
- [13] Why MySQL is popular [Žiūrėta 2007 11 13], prieiga internete <<http://jdk.phpkid.org/2005/08/30/why-mysql-is-popular>>
- [14] John Koenig „Why MySQL sets criteria for well-tempered stacks“. 2005 m. Žiūrėta

- 2007 11 13], prieiga internete < <http://www.itmanagersjournal.com/feature/5622>>
- [15] GIS programinė įranga [Žiūrėta 2007 11 13], prieiga internete < <http://www.hnit-baltic.lt/DesktopDefault.aspx?tabID=3569&alias=hnit-baltic&lang=lt-LT>>
- [16] Yahoo Maps [Žiūrėta 2007 11 13], prieiga internete <<http://maps.yahoo.com>>
- [17] Microsoft Live Search [Žiūrėta 2007 11 13], prieiga internete <<http://local.live.com>>
- [18] Google Maps [Žiūrėta 2007 11 13], prieiga internete <<http://maps.google.com>>
- [19] Google Maps [Žiūrėta 2007 11 13], prieiga internete
<http://en.wikipedia.org/wiki/Google_Maps>
- [20] Justina Bartkutė „Naujųjų informacinių technologijų - geografinių informacinių sistemų panaudojimas kultūros paveldo turiniui perteikti“. 2007 m. [Žiūrėta 2007 11 13], prieiga internete <<http://mokslasplius.lt/multimedija/node/61>>
- [21] Mike Williams „Google Maps API Tutorial“ [Žiūrėta 2007 11 13], prieiga internete
<<http://www.econym.demon.co.uk/googlemaps/index.htm> >
- [22] Google Maps API Tutorial [Žiūrėta 2007 11 13], prieiga internete
<<http://econym.googlepages.com/custommap.htm>>
- [23] J. Dargham. MVC Extension: Evaluation According to Rating Criteria and Possible Integration. Internet and Multimedia Systems and Applications, 2007.
- [24] Wikipedia. CakePHP. [Žiūrėta 2009 04 06], prieiga internete
<http://en.wikipedia.org/wiki/CakePHP>
- [25] J. Andrijauskas „Lietuvos toponimikos tyrimų programinės įrangos kūrimas ir tyrimas“. Magistro darbas, 2009 m.

TERMINŲ IR SANTRUMPŲ ŽODYNAS

Toponimas – vietovės pavadinimas, vietovardis.

Onomastika – mokslas tiriantis tikrinius vardus. Skiriamos dvi šakos: antroponimika (tirianti asmenvardžius) ir toponimika (tirianti toponimus).

Toponimika – mokslas tiriantis toponimus.

Diachroninė toponimika – toponimikos sritis, tirianti istorinius vietovardžius.

Onomastas – onomastikos mokslininkas.

Lingvistika – bendras mokslų, tiriančių natūralias kalbas, pavadinimas.

VLKK – Valstybinė lietuvių kalbos komisija

SQL – Struktūrizuota užklausų kalba (Structured Query Language). Tai universali duomenų bazių valdymo sistemoms sukurta kalba.

GIS – geografinė informacinė sistema.

Morfologija – Lingvistikos dalis, nagrinėjanti žodžių formų struktūrą

JavaScript – tai scenarijų programavimo kalba.

MVC – Programų architektūrinis modelis.

PHP – Plačiai paplitusi dinaminė interpretuojama programavimo kalba.

IS – Informacinė sistema

PĮ – Programinė įranga

phpMyAdmin – MySQL duomenų bazės administravimo įrankis.

HTML – (*Hypertext Markup Language* „Hiperteksto žymėjimo kalba“) – tai kompiuterinė žymėjimo kalba, naudojama pateikti turinį internete.

API (*Application Programming Interface*) – aplikacijų programavimo sąsaja, kurią suteikia kompiuterinė sistema, biblioteka ar programa tam, kad programuotojas per kitą programą galėtų pasiekti jos funkcionalumą ar apsikeistų su ja duomenimis.

WWW (*World Wide Web*) – interneto dalis, resursai, kuriuos internete galima pasiekti naudojant URL.

URL – vieningas resursų identifikatorius; universalus išteklių adresas.

UML (*Unified Modeling Language, Vieninga modeliavimo kalba*) – modeliavimo ir specifikacijų kūrimo kalba, skirta specifiuoti, atvaizduoti ir konstruoti objektiškai orientuotų programų dokumentus.

DB – duomenų bazė.

AJAX – Asinchroninis Javascript ir XML programavimas.

PRIEDAI



**VILNIAUS UNIVERSITETO
KAUNO HUMANITARINIS FAKULTETAS**

Kodas 211950810; Muitinės g. 8, LT-44280 Kaunas; tel.: +370 37 46 55 40, faks.: +370 37 42 32 22

Kauno technologijos universitetui

DĖL LIETUVOS TOPONIMIKOS TYRIMŲ PROGRAMINĖS ĮRANGOS

Magistranto Vyto Kriščiūno sukurta „Automatizuotų vietovardžių tyrimų programinė įranga“ buvo sėkmingai įdiegta ir naudojama Vilniaus universiteto Kauno humanitarinio fakulteto lietuvių filologijos ir reklamos studijų programos leksikologijos modulio mokymo procese bei onomastiniuose tyrimuose.

VU Kauno humanitarinio fakulteto lektorė

Lina Bačiūnaitė-Lužinienė

VU Kauno humanitarinio fakulteto docentė

Ilona Mickienė



**VILNIAUS UNIVERSITETO
KAUNO HUMANITARINIS FAKULTETAS**

Kodas 211950810; Muitinės g. 8, LT-44280 Kaunas; tel.: +370 37 46 55 40, faks.: +370 37 42 32 22

PAŽYMA
2009-05-22

Straipsnis „Naujų technologijų taikymas diachroninėje toponimikoje“ priimtas spausdinti VU Kauno humanitarinio fakulteto tęstiniame leidinyje „Respectus Philologicus“ Nr. 18 (23).

Respectus Philologicus redakcija