

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ INŽINERIJOS KATEDRA

Jonas Andrijauskas

**Lietuvos toponimikos tyrimų programinės įrangos
kūrimas ir tyrimas**

Magistro darbas

Darbo vadovas

doc. dr. Vitolis Sekliuckis

Kaunas, 2009

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
PROGRAMŲ INŽINERIJOS KATEDRA

Jonas Andrijauskas

**Lietuvos toponimikos tyrimų programinės įrangos
kūrimas ir tyrimas**

Magistro darbas

Recenzentas

doc. dr. Stasys Maciulevičius

2009-05-25

Darbo vadovas

doc. dr. Vitolis Sekliuckis

2009-05-25

Atliko

IFM-3/2 gr. stud.

Jonas Andrijauskas

2009-05-25

Kaunas, 2009

Lietuvos toponimikos tyrimų programinės įrangos kūrimas ir tyrimas

SANTRAUKA

Magistrinio darbo tikslas buvo sukurti ir iširti Lietuvos toponimikos tyrimų programinę įrangą. Jo metu buvo atlikta egzistuojančių analogiškų sistemų Lietuvoje ir Europoje lyginamoji analizė. Iširti bei sumodeliuoti onomasto dabartinio ir automatizuoto darbo procesai. Suprojektuota ir sukurta programinės įranga integruota su Google Maps geografinė informacinė sistema. Realizuotas mokslinių tyrimų rezultatų išsaugojimas ir atkūrimas. Programa buvo realiai naudojama eksperimentui Vilniaus universiteto studentų mokyme, o jos kokybė įvertina labai gerai.

Software for Lithuanian toponymy analysis: development and analysis

SUMMARY

The objective of master's thesis was to develop and analyze software for Lithuanian toponymy analysis. There were compared existing analogous Lithuanian and European systems. There were also analysed and designed current and automated activity processes of toponymy researcher. Designed and developed software is integrated with Google Maps GIS. It is possible to save and restore results of onomastics research. Students of Vilnius University were performing experiment on their studies with this software. The quality of software is very well.

TURINYS

Įvadas.....	1
1. Lietuvos toponimikos tyrimų problemų sprendimas.....	2
1.1. Toponimikos skaitmeninimas Lietuvoje	2
1.2. Egzistuojantys toponimikos tyrimų sprendimai Lietuvoje ir pasaulyje.....	2
1.3. Oonomasto darbo procesas ir jo tobulinimas.....	7
1.4. Sprendimų palyginimas naujai kuriamos sistemos atžvilgiu.....	10
2. Lietuvos toponimikos tyrimų programinės įrangos projektavimo svarbiausi aspektai.....	12
2.1. Projekto prielaidos ir tikslai.....	12
2.2. Esminiai reikalavimai programinei įrangai	12
2.3. Lietuvos toponimikos PĮ architektūra	15
2.4. Panaudojimo atvejai	17
2.5. Duomenų bazės modelis ir aprašymas	17
2.6. Pasirinktų priemonių sprendimo įgyvendinimui pagrindimas.....	23
2.7. Specifiniai realizacijos aspektai.....	25
2.8. Statistinė Lietuvos toponimikos PĮ informacija.....	27
3. Lietuvos toponimikos tyrimų programinės įrangos kokybės tyrimas	28
3.1. Tyrimo metodika.....	28
3.2. Kokybės vertinimas.....	29
3.3. Tolesnės plėtros galimybės.....	31
4. Lietuvos toponimikos tyrimų programinės įrangos eksperimentinis tyrimas.....	33
4.1. Vietovardžių įvedimo analizė	33
4.2. Paieškos atlikimo ir rezultatų atkūrimo laikų analizė	37
Išvados.....	42
Literatūra.....	43
Terminų ir santrumpų žodynas	45
Priedai.....	46

IVADAS

Lietuvos toponimikos tyrimai iki šiol buvo atliekami pavieniui. Onomastai savo tyrimais negalėjo tarpusavyje, nes jie arba išvis neskaitmenindavo savo tiriamosios medžiagos, arba tai darydavo tiek, kiek reikia jų pačių tyrimams. Tačiau poreikis turėti vieną bendrą sistemą, kurioje būtų vietovardžių duomenų bazė ir mokslinių tyrimų įrankiai visuomet buvo siektinas tikslas. Oficialiai pirmoji tokio tipo sistema buvo sukurta Joniškio rajono vietovardžių tyrimams atlikti [1]. Joje galima atlikti tik dabartinės toponimikos tyrimus. Jie nuo istorinės toponimikos tyrimų skiriasi tiek duomenų rinkimo, tiek ir jų apdorojimo etapuose. Būtent dėl to duomenų rinkimo analogijos tarp dabartinės ir istorinės toponimikos tyrimų yra beveik neįmanomos. Toponimikos tyrimus onomastai atlieka tik tada, kai patys susirenka jiems reikalingą informaciją. Šiam etapui jie sugaišta labai daug laiko. Kartais tas pačias kartotekas peržiūri keli skirtingi panašius tyrimus atliekantys onomastai. Bendra visų atliekamų toponimų tyrimų programinė įranga palengvintų onomastų darbą ir pagreitintų duomenų analizę bei paskatintų atlikti daugiau bendrų tyrimų, nes visi duomenys būtų toje pačioje sistemoje.

Kuriant programinę įrangą tenka dirbti su onomastais, kurie nenaudoja modernių technologijų savo darbe. Dėl šios priežasties kuriamos programinės įrangos reikalavimų išgavimas yra sudėtingas bei atsakingas procesas. Suvokiant šios programinės įrangos svarbą labai svarbu išanalizuoti, kas yra nuveikta Lietuvoje ir Europoje, ir kruopščiai išsinagrinėti onomasto darbo procesą, kad jį patobulinus ir automatizavus palengvėtų moksliniai tyrimai ir padidėtų jų sklaida.

Darbo pradžioje aptariamos egzistuojančios dabartinės toponimikos sistemos Lietuvoje ir pasaulyje. Tiriamas onomastų darbas ir sudaromas dabartinio darbo modelis, išanalizuojama, kaip jį galima automatizuoti. Pasiūlomi nauji darbo metodai. Sudaromi modeliai. Pagal surinktą informaciją ir reikalavimus yra suprojektuojama ir sukuriama programinė įranga. Ištiriama ir įvertinama jos kokybė. Sukurta programinė įranga buvo įdiegta Vilniaus universitete ir eksperimento tvarka naudojama studentų mokymo procese.

Rengiant šį darbą bendrai su kolegomis buvo parašyta publikacija „Naujųjų technologijų taikymas diachroninėje toponimikoje“, kuri 2010 m. bus išspausdinta tarptautiniame Vilniaus universiteto leidžiamame žurnale „Respectus psychologicus“ (numeris 18 (23)).

1. LIETUVOS TOPONIMIKOS TYRIMŲ PROBLEMŲ SPRENDIMAS

1.1. Toponimikos skaitmeninimas Lietuvoje

Valstybinė lietuvių kalbos komisija savo 2006 m. veiklos ataskaitoje [2] skelbia, kad iš rankraštinių istorijos šaltinių išrašyta 10 000 etninių vietovardžių. 6 885 vietovardžiai buvo įvesti į duomenų bazę. Microsoft Access duomenų bazėje buvo sukurtos paieškos užklausos pagal duomenis iš kelių lentelių, kurios pildomos per sukurtos bazės „Istoriniai etniniai vietovardžiai“ įvedimo modulį. Sukurtas ir tobulintas bazės peržvalgos modulio algoritmas pagal atrankos užklausų projektą. Peržvalgos modulio algoritmas realizuotas Visual Basic priemone. Prie šios duomenų bazės negali prieiti viešai visi, norintys ja naudotis.

Lietuvių kalbos instituto Vardyno skyrius yra bene daugiausiai nuveikęs skaitmeninant Lietuvos toponimiką. Jo viena iš veiklos sričių yra istorinės ir dabartinės toponimikos problemų tyrimas. Visi vartotojai internete gali laisvai prieiti prie elektroninio vietovardžių žodyno. Deja jame trūksta daugelio vietovardžių, kuriuos Vardyno skyrius iš istorinių šaltinių sukaupe savo kartotekoje (daugiau kaip 300 000 toponimų).

1.2. Egzistuojantys toponimikos tyrimų sprendimai Lietuvoje ir pasaulyje

Kadangi Lietuvoje kol kas nėra sukurta vietovardžių analizės informacinė sistema, būtina įvertinti, kas šiuo klausimu nuveikta kitose Europos šalyse. Apžvelgiamos sukurtos 3 sistemos, susijusios su kuriama programine įranga.

Toponimų duomenų bazė (Toponyms Data Bank)

1990 metais R.Miikkulainen (Suomijos kalbų tyrimų institutas, Helsinkis) 37-tame Tarptautiniame Onomastikos kongrese pranešė apie pradėtą kurti suomių toponimų duomenų bazę („Toponyms Data Bank“) [3], kuri leis sujungti skirtingų tipų toponiminius duomenis. Remiantis šia vietovardžių duomenų baze ir onomastinės analizės informacine sistema Suomijos kalbų tyrimų institutas įgyvendino keletą mokslinių tiriamųjų projektų, kurių rezultatus paskelbė publikacijomis. Vienas įdomiausių projektų yra pavadinimu „Paikannimikartasto“ („Vietovardžių atlasas“) [5]. Remiantis toponimų duomenų baze buvo sukurtas skaitmeninis atlasas, iliustruojantis 234 labiausiai paplitusių suomių vietovardžių elementų pasiskirstymą Suomijos teritorijoje. Atlasas nurodo koordinates, fizines savybes ir kitą aktualią toponiminę bei geografinę informaciją. Šis elektroninis atlasas taip pat įgalina atlikti įvairių vietovardžių paieškų. Šiuo metu vis dar vykdomas projektas „Paikannimikirja“

(„Suomių vietovardžių knyga“) [6], kuriuo siekiama surinkti informaciją apie geriausiai žinomus ir svarbiausius Suomijos geografinius vietovardžius, siekiant atsakyti į dažniausiai užduodamus klausimus apie jų kilmę ir reikšmę.

Deja jokia informacija apie technines galimybes nėra publikuojama. Skelbiami tiksliai rezultatai, pasiekti šia programa.

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

1995 m. norvegų tyrinėtojas E.Soderholm iš Tromso universiteto pasiūlė projektą „Toponiminis Barento regiono atlasas“ („Toponymic atlas of the Barents region“) [4]. Projekto reikalingumas buvo grindžiamas poreikiu plėsti kultūrinius, mokslinius, politinius bei ekonominius ryšius tarp Barento regiono šalių. Įkvėpti šios idėjos bei remdamiesi suomių patirtimi, Georgy Kert iš Rusijos lingvistikos, literatūros ir istorijos instituto 1997 m. paskelbė straipsnį apie kuriamą šiaurės vakarų Rusijos toponimikos tyrimų sistemą („Toponymic Research System of Northwest Russia: The TORIS System“) [4]. Sistema buvo sukurta ir prieinama Internetu 2001 m adresu <http://toris.krc.karelia.ru/TORIS/>.

Sistema nenaudoja kartografinės eksplikacijos. Yra tik onomastinės analizės įrankiai.

Techninė realizacija

TORIS serverio architektūra susideda iš trijų pagrindinių elementų (žr. 1):

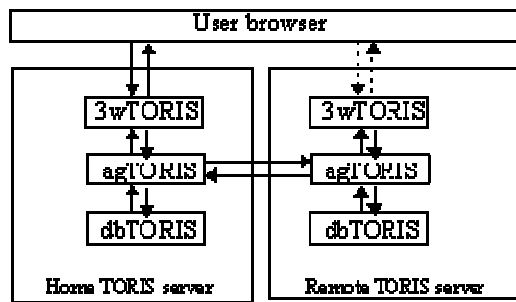
- dbTORIS yra reliacinė duomenų bazė SQL užklausų palaikymui;
- 3wTORIS yra WWW aplikacija įvesti/išvesti informaciją;
- agTORIS yra intelektualių agentų aibė įvairioms užduotims atlikti (pagalba pildant SQL užklausų formas, autonominė paieška TORIS serveryje).

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

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

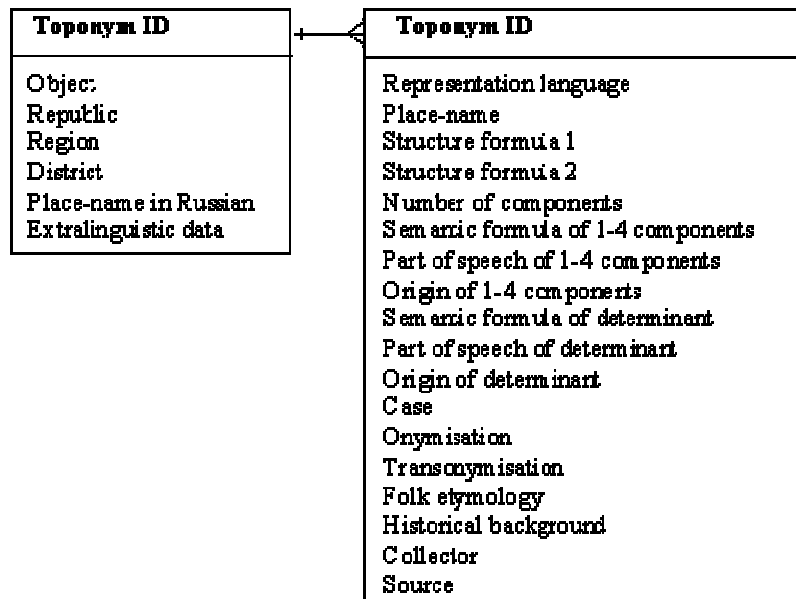
Dalykinė aplikacija – apjungti toponimų domenų. Ji taip pat atsakinga už sąveiką tarp 3wTORIS ir agTORIS komponentų.

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



1 pav. TORIS architektūra [4]

Reliacinė duomenų bazės struktūra (žr. 2 paveikslėlis): apima visą informaciją apie vietovardį: lingvistinę, kultūrinę ir istorinę .

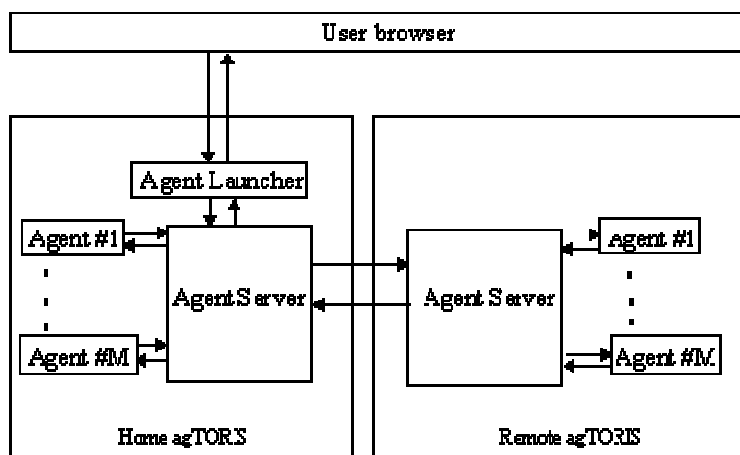


2 pav. Reliacinės duomenų bazės struktūra [4]

Komponentą 3wTORIS sudaro HTML puslapiai ir CGI skriptai tam, kad peržiūrėti, redaguoti vietovardžius ar įterpti naują toponimą.

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

- autonomiškai atlikti paiešką TORIS serveryje,
- užtikrinti teisingų duomenų įvedimą,
- pagalba atliekant užklausą ar įvedant duomenis.



3 pav. AgTORIS architektūra [4]

Joniškio rajono vietovardžių geolingvistinė duomenų bazė

Nors ji oficialiai pasirodė tik 2007 m. rugsėjo mėn, tai, anot kūrėjų, pirmoji skaitmeninė regioninės toponimijos duomenų bazė [1], kuri „gali pasitarnauti kaip modelis skaitmeninant ir kitų regionų ar net visos Lietuvos vietovardžius“. Autoriai išskiria tokias jos panaudojimo galimybes: 1) onomastams, 2) pedagoginei veiklai – darbui su studentais, juos mokant mokslinio tyrinėjimo, 3) praktinėms reikmėms – krašto pažinimui, 4) tam tikroms administracinėms regiono reikmėms. Šioje vietovardžių duomenų bazėje yra apie 6000 toponimų, priklausančių Joniškio rajono teritorijai. Vietovardžiai yra lokalizuoti.

Sistema turi ne tik lokalizuotų vietovardžių duomenų bazę, bet ir atvaizduoja juos žemėlapyje.

Techninė realizacija. Duomenų bazei sukurti naudota tik atvirojo kodo programinė įranga: 1) MySQL duomenų bazės valdymo sistema, 2) sistema suprogramuota Ruby on Rails programavimo kalba, 3) vartotojo sąsajai sukurti naudotos XHTML, CSS ir JavaScript kalbos, 4) interaktyvumui pagerinti naudoti AJAX metodai.

Informacijos susiejimas su žemėlapiu. Sistemoje kiekvienas toponimas yra lokalizuotas konkrečioje Joniškio rajono teritorijos vietoje panaudojus Lietuvos GIS centro surinktą Lietuvos geografinis žemėlapi. Dėl didelio tikslumo (10–20 m vietovėje) jis gan tinkamas yra mikrotoponimams lokalizuoti. Kiekvienas vietovardis buvo tiksliai arba apytiksliai lokalizuotas pagal šaltiniuose pateiktą informaciją.

Paieškos galimybės (žr. 4 ir 5). Paiešką galima vykdyti pagal įvairius kriterijus: 1) pagal toponimo pavadinimą 2) pagal toponimo žodžio dalį, 3) pagal šaltinį, administracinę

teritorija, darybos būdą, darybos pamatą, toponiminę klasę, 4) pagal bet kurią į lauką „Kita informacija“ įvestą žodį.

Informacija žemėlapyje (žr. 6 ir 7). Kiekvienas toponimas lokalizuojamas atskirame lange „Padėtis žemėlapyje“. Jų vieta, remiantis šaltiniuose pateikta informacija, gali būti tiksli (vėliavėlė) arba apytikslė (rodyklės). Taip pat galima ieškoti ir kitų vietovardžių, esančių tam tikru atstumu nuo jau rasto vietos vardo.

4 pav. Paieškos galimybės: paieška pagal vietovardį [1]

5 pav. Informacija apie vietovardį [1]



6 pav. Informacija žemėlapyje: lokacija vietovėje smulkiu masteliu [1]



7 pav. Informacija žemėlapyje: lokacija rajone [1]

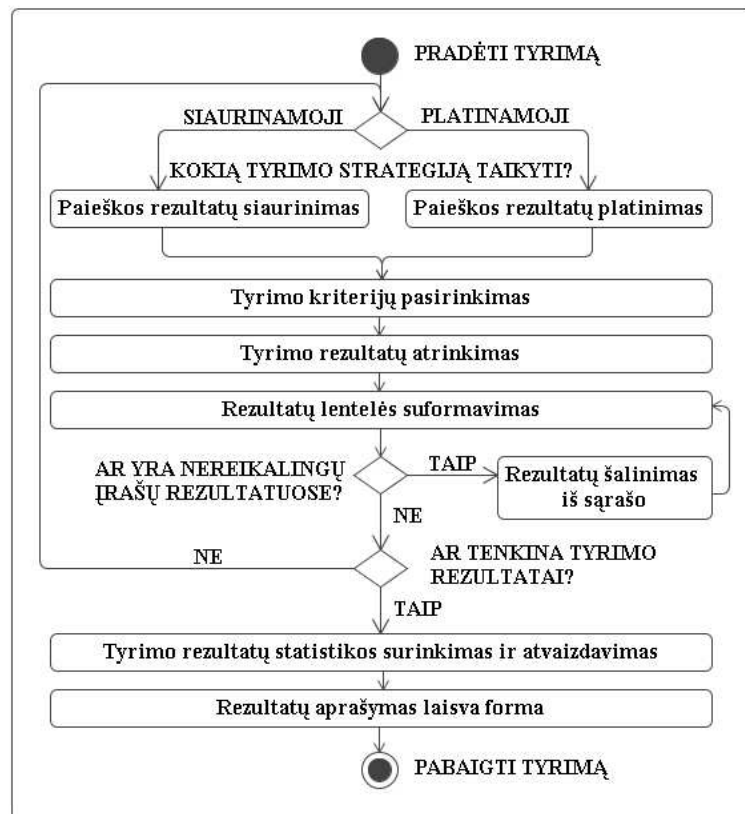
1.3. Onomasto darbo procesas ir jo tobulinimas

Išanalizavus sukurtas toponimikos tyrimų sistemas reikėjo iširti, kaip dabar dirba Lietuvos onomastai. (žr. 8).

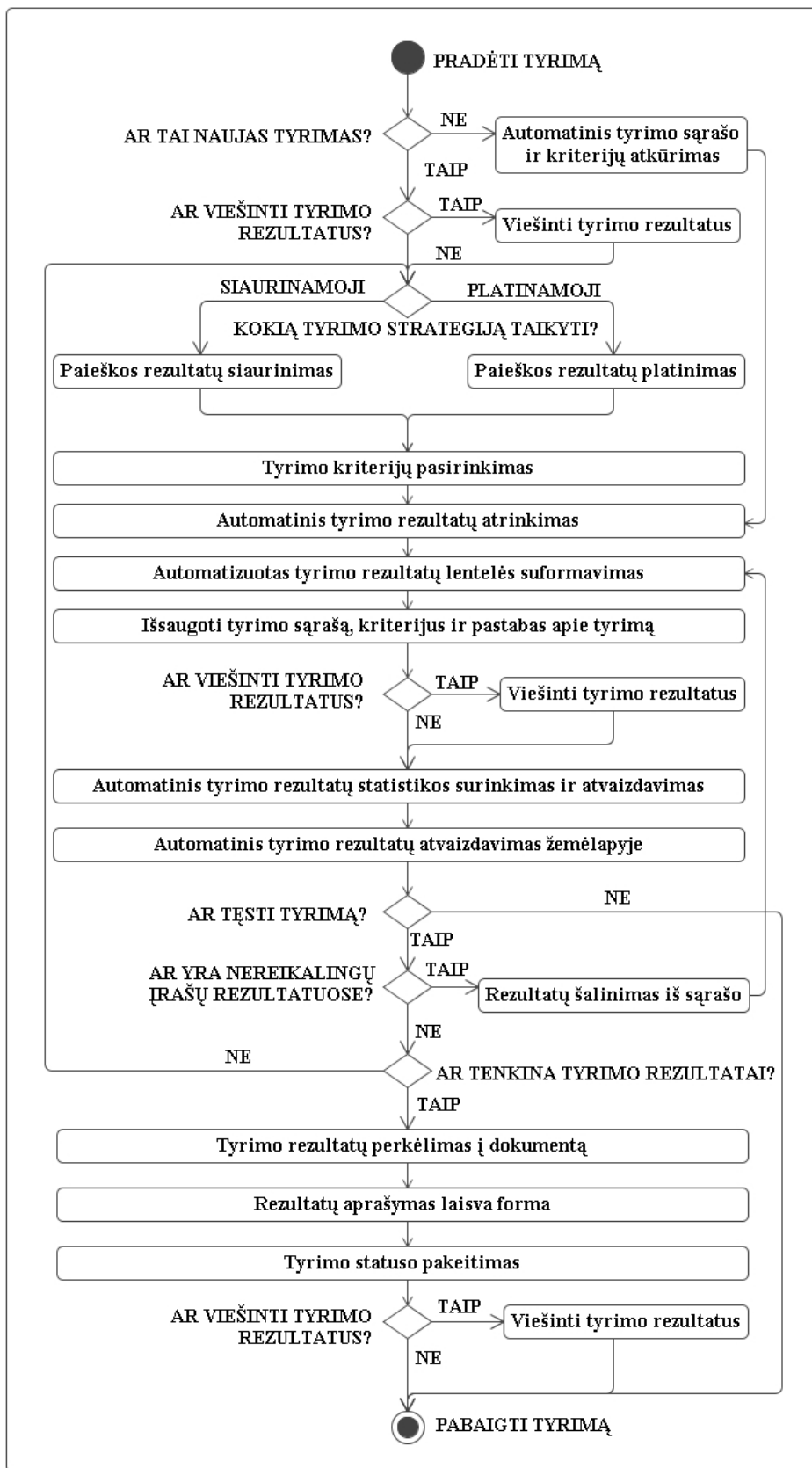
Pateikiamas minėto modelio sudėtingesnių etapų paaiškinimas:

- Rezultatų siaurinimas: paieška vietovardžių sąrašė;
- Rezultatų platinimas: sąrašo pildymas naujais vietovardžiais;
- Kriterijai: apskritis, rajonas, tipas ir t.t.;
- Rezultatų atrinkimas: reikiamų vietovardžių kortelių ištraukimas;
- Rezultatų lentelė: vietovardžių sąrašas pateiktas pasirinktų kriterijų lentele;
- Nereikalingi įrašai: onomasto euristiškai atmetami rezultatai.

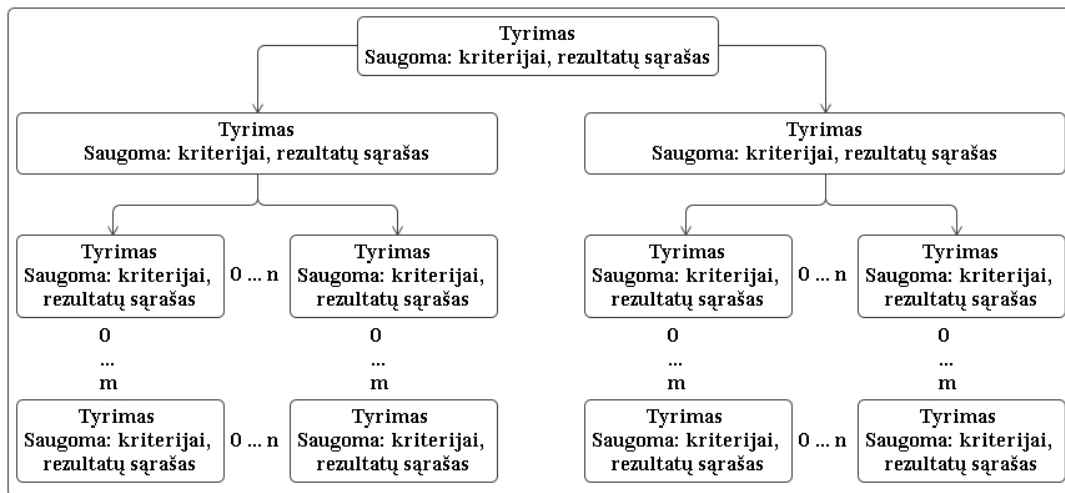
Sukūrus dabartinio onomasto darbo proceso modelį buvo galima išvelgti, kaip galima jį automatizuoti, kad jis dirbtų našiau. Tam, kad pavyktų išvengti pakartotino informacijos rinkimo būtina viską suvesti į vieną duomenų bazę. Ji turėtų būti prieinama visiems suinteresuotiems mokslininkams. Naudojant šią duomenų bazę būtų galima atlikti plačios duomenų apimties toponimikos tyrimus. Tad bendra duomenų bazė yra būtina tyrimo automatizavimo sąlyga. Taip pat buvo sudarytas automatizuotas onomasto tyrimo darbo modelis (žr. 9). Įvestos naujovės: tarpinių rezultatų (sąrašo ir kriterijų) išsaugojimas ir automatinis atkūrimas – taip įgalinamas mokslinių tyrimų tęstinumas, tyrimo rezultatų viešinimas – taip užtikrinama mokslinių tyrimų plėtra ir bendradarbiavimas tarp mokslininkų, automatinis tyrimo rezultatų atvaizdavimas žemėlapyje ir tyrimo statuso keitimas – jis gali būti nebaigtas, laukiantis atsakymo ir pabaigtas. Išsaugoti tyrimo rezultatus ir juos atkurti yra vienas esminių momentų vykdant automatizuotus toponimikos tyrimus. Tad peržvelgus automatizuotą tyrimo darbo modelį, buvo pasiūlytas automatizuotas iteracinis tyrimo modelis, kuris yra realizuojamas medžio struktūra (10). Jį realizavus onomastai galėtų ne tik išsaugoti ar pratęsti paskutinio tyrimo rezultatus, bet ir pratęsti tyrimą nuo bet kurio anksčiau gauto rezultato.



8 pav. Dabartinio onomasto darbo proceso modelis



9 pav. Automatizuotas onomasto darbo proceso modelis



10 pav. Iteracinis onomasto darbo proceso modelis

1.4. Sprendimų palyginimas naujai kuriamos sistemos atžvilgiu

Vienas iš metodų tirti vietovardžius yra struktūrinė analizė. Lietuvių kalba priklauso indoeuropiečių kalbų grupei, o suomių – fino-ugrų. Dėl šios priežasties suomių sukurta informacinė sistema negali būti prototipine kuriamajai programinei įrangai morfologijos atžvilgiu. Rusijos lingvistikos, literatūros ir istorijos instituto IS pritaikyta tik rusų kalbos morfologijai. Visgi jos modelis galėtų būti taikomas kuriamai programinei įrangai dėl praeityje itin artimų baltų ir slavų santykių. Jau egzistuojančios vietovardžių tyrimų informacinės sistemos pritaikytos dabartinių toponimų tyrimams. Ten atliekama tik dabartinių vietos vardų morfeminė analizė. Istorinės onomastikos mokslininkų noras yra turėti galimybę atlikti ir senųjų toponimų morfeminę analizę, kuri būtų siejama su dabartiniais tyrimais.

Mokslininkai, tiriantys tiek dabartinę, tiek istorinę toponimiką, lokalizuoja gyvenamųjų vietų vardus. Onomasto darbas palengvėtų, jei Lietuvos toponimų programinėje įrangoje įvesti vietovardžiai būtų pavaizduoti žemėlapyje. Šiuo metu dar egzistuojančių gyvenamųjų vietų vardų lokalizavimas nėra toks sudėtingas, tačiau norint lokalizuoti jau nebeegzistuojančius, būtina kuo tiksliau pažymėti vietovę arba istorinius žemėlapius susieti su dabartiniais žemėlapiais. Tam būtina naudoti geografinę informacinę sistemą. Iki šiol jos dar nėra viena nagrinėta IS nebuvo taikiusi.

Onomastams yra labai aktuali galimybė dirbti bet kokioje darbo vietoje, dėl to labai svarbu, kad programinę įrangą būtų galima pasiekti internetu iš bet kurio kompiuterio, prijungto prie interneto. Šiuos poreikius tenkina ir Rusijos, ir Joniško informacinės sistemos. Iki šiol sukurtose vietovardžių informacinėse sistemose nebuvo galima tyrimų vykdyti iteracijomis. O tam, kad pradėtą vieną tyrimą galėtų tęsti kitas mokslininkas, yra būtina

galimybė dar nebaigtą tyrimą viešinti kitiems mokslininkams. Šių funkcijų neturėjo dar ne viena iki šiol sukurta informacinė sistema.

Apžvelgtų informacinių sistemų ir kuriamos programinės įrangos esminių funkcijų palyginimas yra pateiktas 1.

1 lentelė Toponimikos tyrimų sprendimų palyginimas

Funkcionalumas	Suomijos kalbų tyrimų instituto IS	Rusijos lingvistikos, literatūros ir istorijos instituto IS	Joniškio rajono vietovardžių geolingvistinė duomenų bazė	Planuojama sukurti toponimų PĮ
Lietuvių kalbos morfologijos palaikymas	–	–	+	+
Galimybė dirbti internete	–	+	+	+
Atvaizdavimas žemėlapyje	+	–	+	+
GIS palaikymas	–	–	–	+
Istorinių žemėlapių integravimas	–	–	–	+
Grafinis paieškos rezultatų atvaizdavimas	–	–	–	+
Dabartinių toponimų morfeminė analizė	+	+	+	+
Istorinių toponimų morfeminė analizė	–	–	–	+
Dabartinių ir istorinių toponimų susiejimas	–	–	–	+
Tyrimų medžio sudarymas	–	–	–	+
Tyrimų viešinimo galimybė	–	–	–	+

2. LIETUVOS TOPONIMIKOS TYRIMŲ PROGRAMINĖS ĮRANGOS PROJEKTAVIMO SVARBIAUSI ASPEKTAI

2.1. Projekto prielaidos ir tikslai

Kaip ir buvo minėta, onomastai dabar pradiniam savo tyrimo etapui turi skirti daug laiko, nes duomenų surinkimo ir analizės etapai nėra automatizuoti. Tyrimų pratęsimo galimybės, kur vieno vietovardžių tyrėjo pradėtą darbą visai kita linkme galėtų tęsti kitas, išvis nėra. Dabartiniams toponimikos tyrimams sukurtų programinių įrangų pagrindu ne visada galima skaitmeninti istorinės toponimikos duomenis. Tad yra siekiama suprojektuoti ir sukurti tokią programinę įrangą, kurioje būtų siejami dabartinės ir istorinės toponimikos duomenys bei būtų galima atlikti testinius mokslinius tyrimus.

2.2. Esminiai reikalavimai programinei įrangai

Išanalizavus egzistuojančius sprendimus tiek Lietuvoje, tiek Europoje ir sudarius automatizuoto onomasto darbo modelį buvo iškelti esminiai reikalavimai pradinei programinės įrangos versijai (prototipui).

2.2.1. Funkciniai reikalavimai

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

1.1. vietovardžio pavadinimas;

1.2. vietovės informacija:

1.2.1. tipas (pvz., *kaimas*);

1.2.2. apskritis (pvz., *Kauno*);

1.2.3. rajonas (pvz., *Jonavos*);

1.2.4. seniūnija (pvz., *Garliava*)

1.2.5. atstumas nuo seniūnijos centro (km);

1.2.6. atstumas nuo parapijos centro (km);

1.2.7. ar egzistuojantis;

1.2.8. literatūra;

1.2.9. pastabos.

1.3. gramatikos informacija:

1.3.1. priešdėliai;

1.3.2. šaknys;

1.3.3. kamienai;

1.3.4. priesagos;

- 1.3.5. galūnės;
 - 1.3.6. tipas (*pirminis, vedinys (priešdėlinis, priešdėlėtasis, priesaginis, priesagėtasis), dūrinys (suvestinis, sudurtinis)*);
 - 1.3.7. Ar taisyklingas vietovardis?
 - 1.3.8. Taisyklingumo pastabos;
 - 1.3.9. Semantika;
 - 1.3.10. Kirčiuotė;
 - 1.3.11. Kirčiuotės pastabos;
 - 1.3.12. Etimologija;
 - 1.3.13. Liaudies etimologija.
2. Galimybė įvesti vietovės ir gramatikos informaciją istoriniams, vietovardžiams:
- 2.1. istorinio vietovardžio pavadinimas;
 - 2.2. istorinė vietovės informacija:
 - 2.2.1. šaltinis;
 - 2.2.2. šaltinio puslapio numeris;
 - 2.2.3. numeris schemeje;
 - 2.2.4. vyskupija;
 - 2.2.5. dekanatas;
 - 2.2.6. parapija;
 - 2.2.7. senoji apskritis;
 - 2.2.8. senoji seniūnija;
 - 2.2.9. metai;
 - 2.2.10. vartojimo kontekstas lietuvių kalba;
 - 2.2.11. vartojimo kontekstas užsienio kalba;
 - 2.2.12. vartojimo konteksto užsienio kalba.
 - 2.3. gramatikos informacija:
 - 2.3.1. priešdėliai;
 - 2.3.2. šaknys;
 - 2.3.3. kamienai;
 - 2.3.4. priesagos;
 - 2.3.5. 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;

- 4.2.1. tipą (pvz., *kaimas*);
 - 4.2.2. apskritį (pvz., *Kauno*);
 - 4.2.3. rajoną (pvz., *Jonavos*);
 - 4.2.4. seniūniją (pvz., *Garliava*);
 - 4.2.5. atstumą nuo seniūnijos centro (pvz., *nuo 10 iki 40 km*);
 - 4.2.6. atstumą nuo parapijos centro (pvz., *nuo 10 iki 40 km*);
 - 4.2.7. egzistavimą (pvz., *taip arba ne arba visi*);
 - 4.2.8. paminėjimą literatūros šaltiniuose (pvz., *taip arba nesvarbu*);
 - 4.2.9. literatūros raktinį žodį (pvz., *Čekiškės parapija*);
 - 4.2.10. lokalizaciją (pvz., *taip arba ne arba visi*).
- 4.3. pagal istorinę vietovardžio informaciją:
- 4.3.1. šaltinį (pvz., *vietovardžių žiniaraštis 2147*);
 - 4.3.2. istorinio vietovardžio numerį schemeje (pvz., *4a*);
 - 4.3.3. vyskupiją (pvz., *Kauno*);
 - 4.3.4. dekanatą (pvz., *Raseinių*);
 - 4.3.5. parapiją (pvz., *Čekiškės*);
 - 4.3.6. bažnytinio archyvo suskirstymo metus (pvz., *1847*);
 - 4.3.7. senąją apskritį (pvz., *Kauno*);
 - 4.3.8. senąją seniūniją (pvz., *Kauno*);
 - 4.3.9. vartojimo kontekstą lietuvių kalbą;
 - 4.3.10. vartojimo kontekstą užsienio kalbą;
 - 4.3.11. vartojimo konteksto užsienio kalbą (pvz., *lenku*);
 - 4.3.12. laikmetį (pvz., *nuo 1859 iki 1914*).
- 4.4. pagal gramatikos informaciją:
- 4.4.1. priešdėlius (pvz., *pa*);
 - 4.4.2. šaknis (pvz., *geg*);
 - 4.4.3. kamienus (pvz., *geguž*);
 - 4.4.4. priesagas (pvz., *už*);
 - 4.4.5. galūnes (pvz., *iai*);
 - 4.4.6. tipą (pvz., *pirminis, vedinys, dūrinys*);
 - 4.4.7. taisyklingumą (pvz., *taip arba ne arba visi*);
 - 4.4.8. semantiką;
 - 4.4.9. kirčiuotę (pvz. 1)
 - 4.4.10. aprašytą etimologiją (pvz., *taip arba nesvarbu*);
 - 4.4.11. 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. asistentai, gali 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 (kai tyrimas atiduotas kito onomasto peržiūrai) 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 plėst (pridėti naujus paieškos rezultatus);
 - 9.7. Turi būti galimybė ieškoti vietovardžių 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. Lietuvos toponimikos PĮ architektūra

Surinkus reikalavimus ir sudarius modelį galima pradėti kurti sistemą. Sistemos kūrimas yra iteracinis procesas, kadangi reikalavimų kaitos rizika yra itin didelė, o sistemos palaikymo, priežiūros ir tobulinimo poreikis ateičiai yra akivaizdus. Todėl būtina naudoti pakankamai lanksčią architektūrą. Tam geriausiai tinka MVC (modelis-peržiūra-kontroleris) architektūra. Ji išskiria sistemą į tris lygius, kuris kiekvienas atsakingas už tam tikrą funkciją. Modelis atsakingas už duomenų sinchronizaciją su duomenų baze. Kontroleris – už sistemos

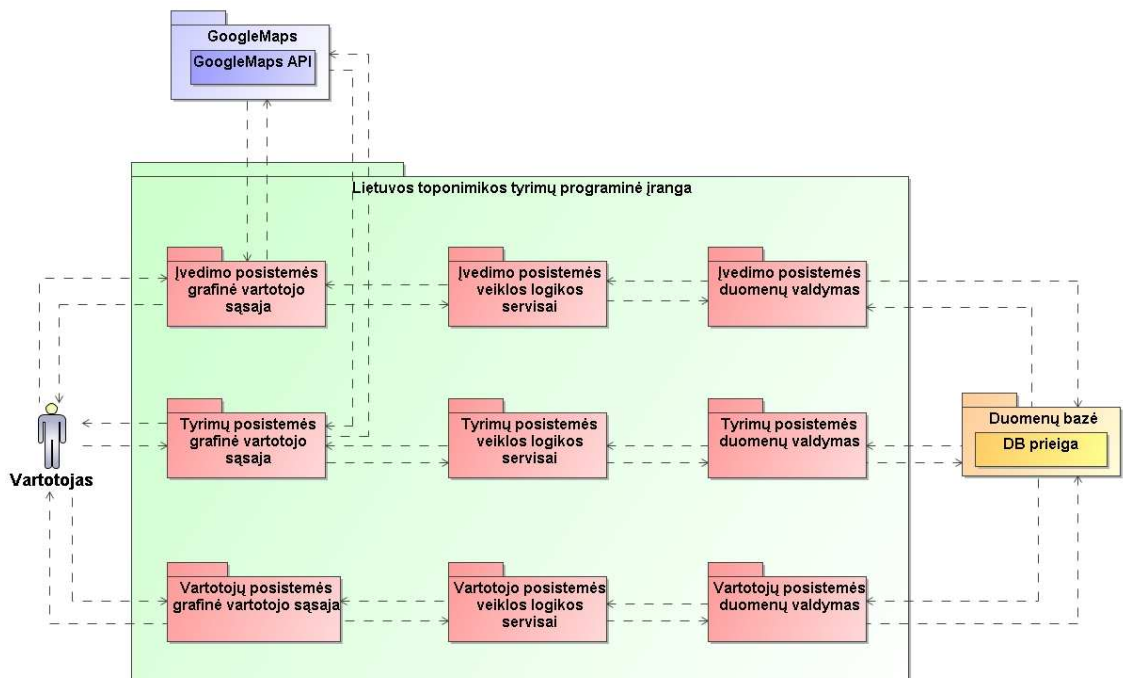
funkcinę dalį. Peržiūra – už duomenų atvaizdavimą vartotojui. MVC pagrindinė idėja yra labai paprasta: programas yra lengviau kurti, derinti (debug) ir palaikyti (maintain), jei skirtingos programinės įrangos dalių atsakomybės yra paskirstytos tarp skirtingų objektų [20].

Remiantis surinkta informacija apie projektą ir MVC architektūros principus, buvo sudaryta Lietuvos toponimikos tyrimų programinės įrangos architektūra (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 ir atvaizdavimą. Vietovardžių duomenų valdymo posistemė atsakinga už duomenų administravimą ir atvaizdavimą, o tyrimų valdymo posistemė už tyrimų organizavimą. Vietovardžių duomenų valdymo posistemės grafinė vartotojo sąsaja bendrauja su Google Maps geografinė informacinė sistema. Per ją yra lokalizuojamas vietovardis. Tyrimų valdymo posistemė per Google Maps GIS atvaizduoja žemėlapyje lokalizuotus toponimus.

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



11 pav. Lietuvos toponimikos PĮ architektūros diagrama

2.4. Panaudojimo atvejai

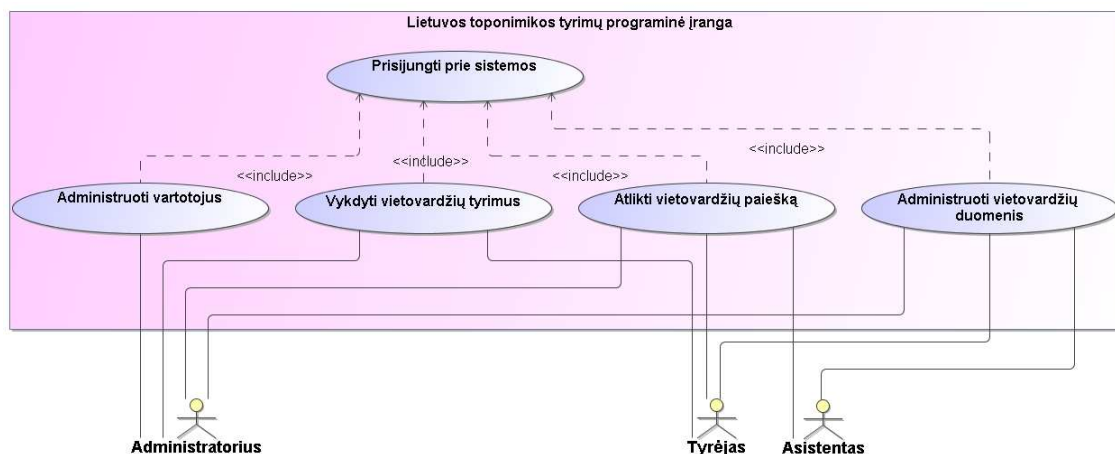
Kaip ir buvo minėta (funkcinis reikalavimas 5), sistemoje bus trys vartotojų tipai, turintys skirtingas teises. Jos yra pavaizduotos panaudojimo atvejų diagramoje (12).

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.

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

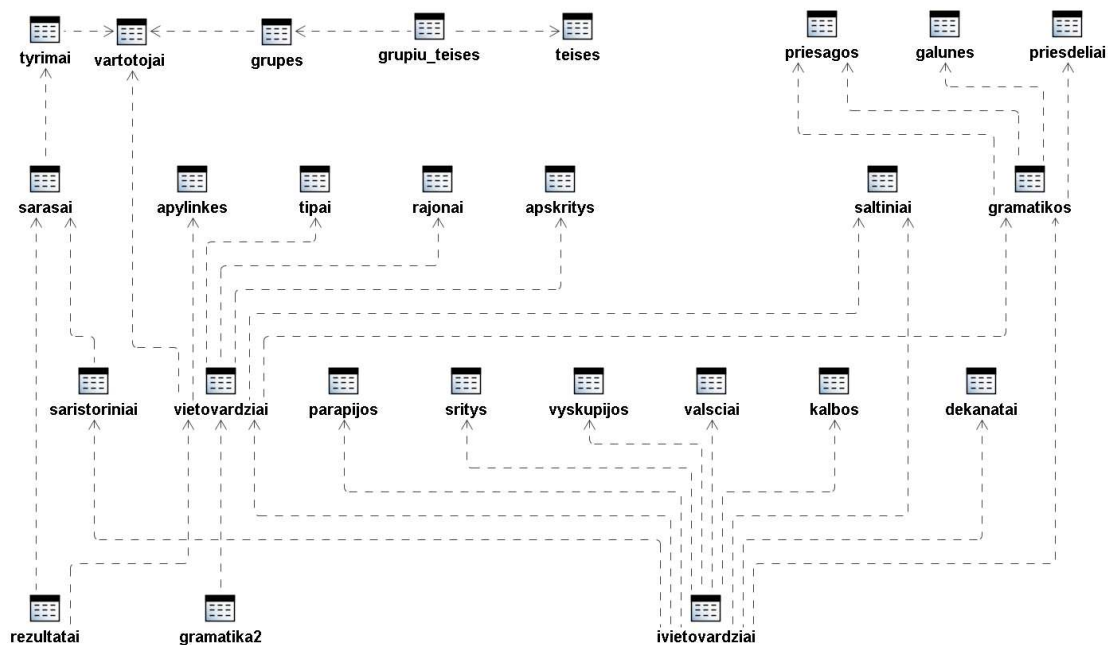
Visi vartotojų tipai savo veiksmus galės atlikti tik prisijungę prie sistemos.



12 pav. Lietuvos toponimikos PĮ panaudojimo atvejų diagrama

2.5. Duomenų bazės modelis ir aprašymas

Duomenų bazėje, kurios schema pavaizduota 13, saugoma visa informacija apie dabartinių ir istorinių vietovardžių vietovę, gramatiką ir lokaciją. Taip pat ten talpinama visa su tyrimais susijusi informacija.



13 pav. Lietuvos toponimikos tyrimų PĮ duomenų bazės schema (lentelių sąryšiai)

2 lentelė Duomenų bazės lentelės, jų atributų tipai ir reikšmės

ATRIBUTAS	TIPAS	PASKIRTIS
VIETOVARDZIAI		
Saugoma visa informacija apie dabartinį vietovardį.		
ID	INT(10)	Vietovardžiui identifikuoti.
PAV	VARCHAR(100)	Vietovardžio pavadinimui.
TIPAS_ID	INT(10)	Vietovardžio tipo identifikuoti
APSKRITIS_ID	INT(11)	Vietovardžio apskričiai identifikuoti.
RAJONAS_ID	INT(10)	Vietovardžio rajonui identifikuoti.
APYLINKI_ID	INT(10)	Vietovardžio seniūnijai identifikuoti.
SALTINIS_ID	INT(10)	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(1)	Vietovardžio egzistavimui identifikuoti.
LITERATURA	TEXT	Kokioje literatūroje paminėtas vietovardis.
PASTABOS	TEXT	Pastabos apie vietovardį.
NR_SCHEMOJE	VARCHAR(5)	Vietovardžio numeriui kartografinėje schemoje.
GRAMATIKA_ID	INT(11)	Gramatikai (morfeminiam išskaidymui) identifikuoti.
KIRTIS1	TINYINT(3)	Vietovardžio pirmo žodžio kirčiui.
KIRTIS2	TINYINT(3)	Vietovardžio antro žodžio kirčiui.
KIRTIS3	TINYINT(3)	Vietovardžio trečio žodžio kirčiui.
VARTOTOJAS_ID	INT(11)	Sistemos vartotojui identifikuoti.

GRAMATIKOS		
Saugoma informacija apie morfeminį vietovardžio išskaidymą.		
ID	INT(11)	Vietovardžio gramatikai (morfeminiam išskaidymui) identifikuoti.
PRIEDELIS1	INT(5)	Vietovardžio pirmam priešdėliui.
PRIEDELIS2	INT(4)	Vietovardžio antram priešdėliui.
PRIEDELIS3	INT(4)	Vietovardžio trečiam priešdėliui.
PRIESAGA1	INT(10)	Vietovardžio pirmai priesagai.
PRIESAGA2	INT(9)	Vietovardžio antrai priesagai.
PRIESAGA3	INT(9)	Vietovardžio trečiai priesagai.
PIEDELIS4	INT(9)	Vietovardžio ketvirtai priesagai.
GALUNE1	INT(9)	Vietovardžio pirmai galūnei.
GALUNE2	INT(9)	Vietovardžio antrai galūnei.
SAKNYS	VARCHAR(100)	Vietovardžio šaknims.
KAMIENAI	VARCHAR(100)	Vietovardžio kamienams.
GRAMATIKA2		
Saugoma informacija apie papildomą dabartinio vietovardžio gramatiką.		
ID	INT(10)	Vietovardžio gramatikai (nemorfeminei) identifikuoti.
VIETOVARDIS_ID	INT(11)	Vietovardžiui identifikuoti.
TIPAS	TINYINT(3)	Ar vietovardis pirminis ar vedinys. Ir jei vedinys, tai koks.
TAISYKLINGUMAS	TINYINT(1)	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(10)	Istoriniui vietovardžiui (toliau - IV) (dabartinio atitikmeniu) identifikuoti.
PAV	VARCHAR(100)	IV pavadinimui.
SALTINIS_ID	INT(10)	IV šaltiniui identifikuoti.
VIETOVARDIS_ID	INT(10)	Dabartiniam vietovardžiui identifikuoti.
PUSLAPIS	VARCHAR(5)	Kuriame šaltinio puslapyje IV paminėtas.
NR_SCHEMOJE	VARCHAR(12)	IV numeriui kartografinėje schemeje.
VYSKUPIJA_ID	INT(9)	IV vyskupijai identifikuoti.
DEKANATAS_ID	INT(9)	IV dekanatui identifikuoti.
PARAPIJA_ID	INT(9)	IV parapijai identifikuoti.
SRITIS_ID	INT(9)	IV sričiai identifikuoti.
VALSCIUS_ID	INT(9)	IV valsčiui identifikuoti.

KONTEKSTAS_LT	TEXT	IV vartojimo kontekstas lietuvių kalba.
KONTEKSTAS_KT	TEXT	IV vartojimo kontekstas užsienio kalbai.
METAİ	YEAR(4)	Kuriais metais IV priklausė vyskupijai, dekanatui, parapijai.
KALBA_ID	INT(9)	IV vartojimo konteksto kalbai identifikuoti.
KIRTIS1	TINYINT(4)	IV pirmo žodžio kirčiui.
KIRTIS2	TINYINT(4)	IV antro žodžio kirčiui.
KIRTIS3	TINYINT(4)	IV trečio žodžio kirčiui.
GRAMATIKA_ID	INT(11)	Gramatikai (morfeminiam išskaidymui) identifikuoti.
APSKRITYS		
Saugoma informacija apie Lietuvos apskritis.		
ID	INT(11)	Vietovardžio apskričiai identifikuoti.
PAV	VARCHAR(50)	Apskritis pavadinimui.
RAJONAI		
Saugoma informacija apie Lietuvos rajonus. Jie yra susieti su apskritimis.		
ID	INT(10)	Vietovardžio rajonui identifikuoti.
APSKRITIS_ID	INT(11)	Kuriai apskričiai priklauso rajonas identifikuoti.
PAV	VARCHAR(100)	Rajono pavadinimui.
APYLINKES		
Saugoma informacija apie Lietuvos seniūnijas. Jos yra susietos su rajonais.		
ID	INT(10)	Vietovardžio seniūnijai identifikuoti.
RAJONAS_ID	INT(11)	Kuriam rajonui priklauso seniūnija identifikuoti.
PAV	VARCHAR(100)	Seniūnijos pavadinimui.
TIPAI		
Saugoma informacija apie vietovardžio tipus.		
ID	INT(10)	Vietovardžio tipui identifikuoti.
PAV	VARCHAR(100)	Tipo pavadinimui.
SALTINIAI		
Saugoma informacija apie vietovardžių paminėjimo šaltinius.		
ID	INT(10)	Vietovardžio šaltiniui identifikuoti.
PAV	VARCHAR(100)	Šaltinio pavadinimui.
PR_METAI	INT(4)	Nuo kada šaltinis gyvavo.
PB_METAI	INT(4)	Iki kada šaltinis gyvavo.
VYSKUPIJOS		
Saugoma informacija apie Lietuvos vyskupijas.		
ID	INT(9)	IV vyskupijai identifikuoti.
PAV	VARCHAR(32)	Vyskupijos pavadinimui.
DEKANATAI		
Saugoma informacija apie Lietuvos dekanatus.		
ID	INT(9)	IV dekanatui identifikuoti.

PAV	VARCHAR(32)	Dekanato pavadinimui.
PARAPIJOS		
Saugoma informacija apie Lietuvos parapijas.		
ID	INT(9)	IV parapijai identifikuoti.
PAV	VARCHAR(32)	Parapijos pavadinimui.
VALSCIAI		
Saugoma informacija apie Lietuvos valsčius.		
ID	INT(9)	IV valsčiui identifikuoti.
PAV	VARCHAR(32)	Valsčiaus pavadinimui.
SRITYS		
Saugoma informacija apie Lietuvos sritis.		
ID	INT(9)	IV sričiai identifikuoti.
PAV	VARCHAR(32)	Srities pavadinimui.
KALBOS		
Saugoma informacija apie vartojimo konteksto kalbas.		
ID	INT(9)	Kalbai identifikuoti.
PAV	VARCHAR(32)	Kalbos pavadinimui.
PRIEDELIAI		
Saugoma informacija apie priešdėlius.		
ID	INT(10)	Vietovardžio priešdėliui identifikuoti.
PAV	VARCHAR(100)	Priešdėlio pavadinimui.
PRIESAGOS		
Saugoma informacija apie priesagas.		
ID	INT(10)	Vietovardžio priesagai identifikuoti.
PAV	VARCHAR(100)	Priesagos pavadinimui.
GALUNES		
Saugoma informacija apie galūnes.		
ID	INT(10)	Vietovardžio galūnei identifikuoti.
PAV	VARCHAR(100)	Galūnės pavadinimui.
VARTOTOJAI		
Saugoma informacija apie sistemos vartotojus.		
ID	INT(10)	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(10)	Vartotojų grupei identifikuoti.
AKTYVUMAS	TINYINT(1)	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(10)	Vartotojų grupei identifikuoti.
TEISE_ID	INT(10)	Prieigos teisei identifikuoti.
TEISES		
Saugoma informacija apie sistemos prieigos teises.		
ID	INT(10)	Prieigos teisei identifikuoti.
PAV	VARCHAR(50)	Prieigos teisės pavadinimas.
REZULTATAI		
Saugoma informacija apie sąrašui priklausančius dabartinius vietvardžius.		
ID	INT(11)	Rezultatui identifikuoti.
SARASAS_ID	INT(11)	Sąrašui identifikuoti.
VIETOVARDIS_ID	INT(11)	Vietovardžiui, priklausančiam sąrašui identifikuoti.
SARISTORINIAI		
Saugoma informacija apie sąrašui nepriklausančius istorinius vietovardžius.		
ID	INT(11)	Rezultatui identifikuoti.
SARASAS_ID	INT(11)	Sąrašui identifikuoti.
IVIETOVARDIS_ID	INT(11)	IV nepriklausančiam sąrašui identifikuoti.
SARASAI		
Saugoma informacija apie tyrimų medį.		
ID	INT(11)	Tyrimo medžio šakai-sąrašui identifikuoti.
SARASAS_ID	INT(11)	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(11)	Tyrimui identifikuoti.
VARTOTOJAS_ID	INT(11)	Vartotojui, atlikusiam tyrimą, identifikuoti.
PAV	VARCHAR(30)	Tyrimo pavadinimui.
DATA	DATETIME	Tyrimo datai.
VIESUMAS	TINYINT(1)	Ar tyrimas viešas, ar privatus.
BUSENA	TINYINT(3)	Tyrimo būsenai

APRASYMAS	TEXT	Tyrimo aprašymui.
SARASAS_ID	INT(11)	Paskutiniam suformuotam sąrašui identifikuoti.

2.6. Pasirinktų priemonių sprendimo įgyvendinimui pagrindimas

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.

2.6.1. CakePHP karkasas

MVC architektūrą palaiko CakePHP karkasas (framework), parašytas PHP programavimo kalba. Jis pasižymi šiomis savybėmis [21]:

- 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 instaliacijai nereikia jokių specialių Apache nustatymų. Instaliavus pats karkasas praneša apie trūkstamus modelius, kontrolierius ar peržiūros failus, tad aplikacijos kūrimas bei integracija su Google geografinė informacinė sistema vyksta pakankamai greitai. Kiek daugiau darbo tenka įdėti visų duomenų bazės ryšių aprašymui.

2.6.2. MySQL DBVS

Renkantis duomenų bazę, negalima apsiriboti vien savybių aibe. Reikia apsvarstyti dokumentaciją, palaikymą, kainą, stabilumą, bendrą veikimą bei greitį. Atvirojo kodo MySQL duomenų bazė buvo suprojektuota būti greita duomenų baze [7]. Kaip ir dauguma kitų duomenų bazių, ji taip pat turi galimybę saugoti geografinės informacinės sistemos duomenis [8]. Maksimalus MySQL DB dydis (neribotas), lentelės dydis (16TB) ir kiti parametrai [9] visiškai patenkina būsimos programinės įrangos poreikius. Renkantis

MySQL, itin svarbų vaidmenį vaidina labai didelė ir aktyvi duomenų bazės vartotojų ir kūrėjų bendruomenė [10].

MySQL duomenų bazė pasirenkama dar ir dėl šių kriterijų:

- Lengvai naudojama ir kuriama [11];
- Aukštos kokybės duomenų bazės administravimo ir kūrimo įrankis phpMyAdmin;
- Greitas ir stabilus veikimas [11];
- Patikimumas [11];
- Jau yra sėkmingai pritaikyta panašiam projekte [1].

2.6.3. Google Maps GIS

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

ArcGIS [12]. Mokama programinė įranga. Licencijas reikia pirkti kas metai. Pagalba bei kursai programuotojui taip pat mokami. Galinga programinė įranga, funkcionalumu galinti patenkinti kuriamos sistemos poreikius.

Yahoo Maps [13]. Nemokamas. Flash technologija paremtas žemėlapis. Yra suteikę galimybę naudotis žemėlapiu programuotojams savo svetainėje, tačiau jam vis dar stinga tarptautinio palaikymo [16].

Microsoft's Local Live Map [14]. Nemokamas. Žemėlapiu galima naudotis tik local.live.com puslapyje. Nėra galimybės įsidėti savo žemėlapių sluoksnį, pačiam žymėti vietovę, jokios pagalbos programuotojui.

Google Maps [15]. 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 bei nuolat tobulinamas. 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“ [17]. Lietuvos žemėlapiai yra pakankamai detalūs. Google Maps leidžia įsidėti ir savus žemėlapius [19], tad detalesnių ar istorinių žemėlapių įkėlimas nebūtų technologinė kliūtis. Ši geografinė informacinė sistema yra palaikoma įvairių internetinių naršyklių. Tokių kaip kaip (pateiktos minimalios versijos): 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 [18] – funkcijų rinkinį, kuris leidžia

praplėsti Google Maps funkcionalumą. Visas API funkcijas galima naudoti JavaScript skriptuose, esančiuose savo žiniatinklio svetainėje.

Taigi iš šio palyginimo galima daryti išvadą, kad Google Maps GIS yra optimalus sprendimas kuriamai programinei įrangai.

2.7. Specifiniai realizacijos aspektai

Realizuojant Lietuvos toponimų programinę įrangą teko spręsti sudėtingas technines problemas.

2.7.1. Toponimų kirčiavimas

Kiekvienas vietovardis turėjo būti kirčiuotas. Tačiau paieškoje įvedus vietovardį be kirčio jį irgi turėjo rasti. Nenorint apsunkinti onomasto darbo papildomų šriftų diegimu (pvz. Palemonas), buvo pasirinkti unikodo kirčio simboliai. Kirčius teko saugoti atskirai nuo vietovardžio pavadinimo. Taip buvo išspręsta atvaizdavimo ir paieškos problema. Realizuojant šį sprendimą buvo naudojamos CakePHP karkaso modeliuose integruotos BeforeSave ir AfterFind funkcijos. Prieš įrašant toponimą į duomenų bazę visuomet yra iškviečiam BeforeSave funkcija. Paieškos rezultatų pakoregavimui prieš perduodant juos valdikliui naudojama AfterFind funkcija. Duomenų bazėje vieno toponimo kirčiams saugoti buvo nuspręsta skirti tris baitus, po vieną kiekvienam kirčiui. Šių kirčių maksimaliai gali būti 3. Kiekvieno baito pirmi du bitai buvo naudojami kirčio tipui (lietuvių kalboje naudojami trys kirčio tipai: riestinis, kairinis ir dešininis). Likę 6 bitai skirti nusakyti, kurioje toponimo vietoje kirtis turi būti dedamas. Lietuvių kalboje nėra ilgesnio žodžio nei 64 (2^6) simboliai, todėl 6 bitų visiškai pakanka.

2.7.2. Vietovardžių paieška

Informacijos apie dabartinį ir istorinį vietovardį yra tikrai daug, kai kurie kriterijai tinka tiek dabartiniams, tiek istoriniams vietovardžiams, tad paieška pagal nurodytus kriterijus yra pakankamai sudėtinga. Žemiau pateikiama vienos dinaminės paieškos užklauso fragmentas. Ji naudojama vietovardžių paieškos lange ir šiuo konkrečiu atveju ieško vietovardžio, kurio ID yra 1958.

```
SELECT `Vietovardis`.`id`, `Vietovardis`.`pav`, .... , `Gramatika2`.`etimologija`,  
`Gramatika2`.`liaudieset` FROM `vietovardziai` AS `Vietovardis` LEFT JOIN `apylinkes`  
AS `Apylinke` ON (`Vietovardis`.`apylinke_id` = `Apylinke`.`id`) LEFT JOIN `saltiniai` AS  
`Saltinis` ON (`Vietovardis`.`saltinis_id` = `Saltinis`.`id`) LEFT JOIN `tipai` AS `Tipas` ON  
(`Vietovardis`.`tipas_id` = `Tipas`.`id`) LEFT JOIN `gramatikos` AS `Gramatika` ON  
(`Vietovardis`.`gramatika_id` = `Gramatika`.`id`) LEFT JOIN `rajonai` AS `Rajonas` ON
```

```
('Vietovardis`.`rajonas_id` = `Rajonas`.`id`) LEFT JOIN `apskritys` AS `Apskritis` ON  
(`Vietovardis`.`apskritis_id` = `Apskritis`.`id`) LEFT JOIN `gramatika2` AS `Gramatika2`  
ON (`Gramatika2`.`vietovardis_id` = `Vietovardis`.`id`) WHERE `Vietovardis`.`id` IN (1958)  
ORDER BY `Vietovardis`.`id` ASC LIMIT 20
```

2.7.3. Tyrimo rezultatų saugojimas ir atkūrimas

Tai buvo viena sudėtingiausių realizacijos dalių. Norint ją atlikti tinkamai, buvo analizuojami keli įgyvendinimo variantai [24].

1 variantas. Jeigu duomenų bazė nebūtų pildoma, tyrimo rezultatams atkurti užtektų saugoti tik onomasto atliktus veiksmus, tačiau šioje programinėje įrangoje būtų saugojami paieškos kriterijai arba šalinamų iš rezultatų sąrašo vietovardžių numeriai. Tokiu būdu būtų sumažinamas duomenų bazėje saugomų duomenų kiekis. Tačiau šis sprendimas turi nemažai trūkumų: sudėtinga realizacija, rezultatų atkūrimo laiko priklausomybė nuo norimo atkurti sąrašo, t.y. kuo daugiau veiksmų atliekama iki tam tikro rezultato, tuo labiau apkraunama DBVS ir lėtėja atkūrimo laikas.

2 variantas. Galima saugoti duomenų bazės SQL užklausas. Paieškoms tarp rezultatų būtų naudojamos užklausos SQL užklausose. Rezultatai būtų pildomi sujungiant užklausas (naudojama UNION funkcija). Nors šio varianto realizacija yra paprastesnė, tačiau vis tiek lieka keli trūkumai: didelis rezultatų atkūrimo laikas, sparčiai auga duomenų bazės dydis.

3 variantas. Šiuo atveju būtų saugomi kriterijai ir veiksmai ir pagal juos sudaroma viena SQL užklausa. Tai sudėtingai realizuojamas variantas, nes toks rezultatų atrinkimas reikalauja sudėtingo kriterijų formalizavimo, loginės jų analizės atkuriant rezultatų sąrašą bei yra priklausomas nuo duomenų bazėje esančių toponimų kiekio. Šis variantas nuo pirmojo skiriasi realizacijos sudėtingumu, nes pirmu atveju rezultatų atkūrimas vyksta etapais, o šiuo atveju – vienu metu, išanalizavus visus žingsnius.

4 variantas. Galima saugoti visus sąrašo elementus (jų numerius), o kriterijai būtų saugomi tik kaip papildomos pastabos tyrėjams. Duomenų bazėje būtų saugoma nemažai duomenų, ypač kai sąrašuose yra daug įrašų. Tam, kad rezultatų išrinkimo laikas būtų kuo mažesnis, reikia indeksuoti lentelę, kurioje jie yra saugomi. Todėl rezultatų atrinkimo laikas priklauso ne nuo lentelės dydžio ar konkretaus tyrimo, o nuo rezultatų kiekio. Dėl augančių duomenų bei tyrimų skaičiaus gali būti viršytos aparatūrinės įrangos techninės galimybės, todėl reikia vykdyti nuolatinę sistemos priežiūrą.

Tad išanalizavus visus atvejus buvo pasirinktas ketvirtas variantas. Sudarinėjant tyrimų medį, kiekviename etape reikėjo pateikti informaciją, kokie paieškos rezultatai atsispindi tam tikroje medžio šakoje. Žemiau pateikiamas tyrimo rezultatų atrinkimo užklausos fragmentas, kuri atrenka vietovardžių sąrašą pagal išsaugotus numerius 4, 13, 16, 36, 56, 57, 185, 213, 252, 258, 261, 294, 1781, 1783, 1784:

```
SELECT      `Vietovardis`.`id`,      `Vietovardis`.`pav`,      .....,
`Gramatika2`.`etimologija`,      `Gramatika2`.`liaudieset` FROM      `vietovardziai` AS
`Vietovardis` LEFT JOIN `apylinkes` AS `Apylinke` ON (`Vietovardis`.`apylinke_id` =
`Apylinke`.`id`) LEFT JOIN `saltiniai` AS `Saltinis` ON (`Vietovardis`.`saltinis_id` =
`Saltinis`.`id`) LEFT JOIN `tipai` AS `Tipas` ON (`Vietovardis`.`tipas_id` = `Tipas`.`id`)
LEFT JOIN `gramatikos` AS `Gramatika` ON (`Vietovardis`.`gramatika_id` =
`Gramatika`.`id`) LEFT JOIN `rajonai` AS `Rajonas` ON (`Vietovardis`.`rajonas_id` =
`Rajonas`.`id`) LEFT JOIN `apskritys` AS `Apskritis` ON (`Vietovardis`.`apskritis_id` =
`Apskritis`.`id`) LEFT JOIN      `gramatika2`      AS      `Gramatika2`      ON
(`Gramatika2`.`vietovardis_id` = `Vietovardis`.`id`) WHERE `Vietovardis`.`id` IN (4, 13, 16,
36, 56, 57, 185, 213, 252, 258, 261, 294, 1781, 1783, 1784) ORDER BY `Vietovardis`.`id`
ASC LIMIT 20
```

2.8. Statistinė Lietuvos toponimikos PĮ informacija

3 lentelėje pateikiama realizuotos Lietuvos toponimikos programinės įrangos statistinė informacija.

3 lentelė Lietuvos toponimikos 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) Grafinio sistemos stiliaus aprašymas.	600
3. Duomenų bazės lentelių kiekis	26

3. LIETUVOS TOPONIMIKOS TYRIMŲ PROGRAMINĖS ĮRANGOS KOKYBĖS TYRIMAS

3.1. Tyrimo metodika

Lietuvos toponimikos tyrimų programinės įrangos kokybė buvo vertinama pagal šiuos faktorius:

Programinės įrangos panaudojamumo faktoriai:

- Korektiškumas – kaip korektiškai programa atlieka savo funkcijas (pvz. įvedimą, lokalizavimą, paiešką ir tarpinių duomenų saugojimą bei atkūrimą);
- Patikimumas – pvz. ar įvedinėjant vietovardį ar atliekant tyrimą programa „neužlūš“;
- Efektyvumas – programai reikalingų išteklių įvertinimas;
- Vientisumas – ar vartotojai nesugebės sugadinti programos duomenų neteisingai ja naudodamiesi;
- Panaudojamumas – kaip onomastui yra lengva naudotis programa.

Produkto aptarnavimo faktoriai:

- Testuojamumas – kokios pastangos reikalingos, kad įsitikinti, jog programa veikia teisingai;
- Lankstumas – galimybė pakeisti ką nors programoje, kad ji labiau tenkintų onomasto reikalavimus.

Kiekvienas faktorius išanalizuotas keliais kriterijais. Įvertinimui buvo naudojama metrika, kur kiekvienam kriterijui bus priskiriamas tam tikras svertinis rodiklis (nuo 0 iki 10). Taip pat visai programai buvo paskirstyti svoriniai koeficientai, kurių suma yra 1.

Faktoriaus kokybės kriterijus įvertinimas skaičiuojamas kriterijų įvertinimų sumą dalijant iš kriterijų skaičiaus ir gautą rezultatą padauginus iš svorinio koeficiento.

Bendras programos kokybės įvertinimas gaunamas susumavus kiekvieno faktoriaus įvertinimus.

3.2. Kokybės vertinimas

4 lentelė Lietuvos toponimikos tyrimų programinės įrangos kokybės vertinimas

Kriterijus	Svorinis koeficientas	Įvertinimas
Korektiškumas		
Prie programos gali prisijungti/atsijungti 3 vartotojų tipai, turintys skirtingas teises.	0,1	10
Programa korektiškai įrašo, atnaujina ir ištrina vietovardžio informaciją.		10
Su programa atliekama vietovardžių paieška yra teisinga, rezultatai yra puslapiuojami.		10
Sistema teisingai lokalizuoja ir atvaizduoja lokalizuotus vietovardžius žemėlapyje.		10
Programa tinkamai išsaugo tyrimo rezultatus.		10
Programa teisingai atkuria bet kurios tyrimo medžio šakos rezultatus.		10
<i>Iš viso korektiškumas</i>		<i>1 (iš 1)</i>
Patikimumas		
Visos programos klaidos registruojamos ir apie jas pranešama vartotojui. Pastabos: tyrimo organizavimo modulyje ne visos klaidos registruojamos, pvz. nuėmus varnelės nuo visų rezultatų lentelės laukų – nerodomas joks pranešimas, o tiesiog atstatomi prieš tai buvę duomenys.	0,1	8
Naudojantis programa įvyksta ne daugiau kaip viena klaida per 7 dienas. Pastabos. Kadangi sistema naudojasi pakankamai daug studentų, jų didžiausio aktyvumo laikotarpiu buvo registruojama iki 4-5 klaidų (nenumatytų sistemos atvejų) per dieną, kurios buvo operatyviai šalinamos.		6
<i>Iš viso patikimumas</i>		<i>0,7 (iš 1)</i>
Efektyvumas		
Paieškos laikas, kai vienas vietovardis susietas su vienu	0,1	10

Kriterijus	Svorinis koeficientas	Įvertinimas
istoriniu vietovardžiu esant iki 10000 vietovardžių duomenų bazėje nėra didesnis nei 1 sekundė.		
Programai nereikia saugoti Lietuvos žemėlapių, taip taupant serverio vietą.		10
<i>Iš viso efektyvumas</i>		<i>1 (iš 1)</i>
Vientisumas		
Mažiausiai nusimanantys vartotojai – asistentai – negalės redaguoti ne savo įvestų duomenų bei trinti bet kokius duomenis iš sistemos.	0,2	10
Prieš šalinant duomenis iš sistemos, yra prašoma patvirtinti savo veiksmus.		10
Netyčia atlikus nereikalingas korekcijas ar šalinimus bus galima atstatyti duomenis.		10
<i>Iš viso vientisumas</i>		<i>2 (iš 2)</i>
Panaudojamumas		
Vartotojo sąsajoje vietovardžio įvedimo formoje yra aišku, kokius duomenis reikia įvesti prie kiekvieno lauko.	0,1	10
Vartotojo sąsaja įskaitoma, meniu aiškiai išdėstytas.		10
Vartotojas susipažįsta ir išmoksta naudotis pagrindinėmis programos funkcijomis per vieną darbo dieną (8 valandas).		10
<i>Iš viso panaudojamumas</i>		<i>1 (iš 1)</i>
Testuojamumas		
Programai nereikia sukonfigūruot ir paleist specialią aplinką prieš pradėdant jos testavimą.	0,1	10
Programą galima testuoti naudojant automatinius testus. Pastaba. Augant duomenų bazės įrašų skaičiui, gali tekti koreguoti automatinius testus.		9
<i>Iš viso testuojamumas</i>		<i>0,95 (iš 1)</i>
Lankstumas		
Sistema naudoja aiškią MVC architektūrą, t.y.	0,2	10

Kriterijus	Svorinis koeficientas	Įvertinimas
skirtingos programinės įrangos dalių atsakomybės yra paskirstytos tarp skirtingų objektų.		
Programos kodas dokumentuotas pagal phpDocs taisyklės [23], kad būtų galima automatiškai generuoti dokumentaciją, kuri būtina norint atlikti pakeitimus. Pastaba. Šiuo metu ne visas kodas yra dokumentuotas pagal šį standartą.		8
<i>Iš viso lankstumas</i>		<i>1,8 (iš 2)</i>
IŠ VISO (maks. 10 balų)		8,75

Atlikus skaičiavimus buvo gautas bendras programos kokybės įvertinimas – 8,75 balo iš 10-ties galimų. Tai reiškia, kad programa yra labai geros kokybės.

Kokybės faktorių įvertinimai daugmaž vienodi. Gal kiek išsiskiria patikimumo ir lankstumo faktoriai. Patikimumo faktoriaus įvertinimą daugiausiai įtakojo nemažas vartotojų skaičius, kuris naudojo sistema. Dėl šios priežasties pasireiškėdavo sistemos klaidos ar nenumatytos vietos. Iš kitos pusės - tai yra pateisinama, nes tik nenaudojantys programos vartotojai neatranda klaidų ar tobulintinų vietų. Lankstumo faktorių labiausiai įtakojo nevisiškas programinio kodo dokumentavimas pagal nurodytus standartus. Tačiau suprantant, kad norint tobulinti ir palaikyti programą būtina turėti dokumentaciją, tai turėtų artimiausiu metu išsispęsti.

3.3. Tolesnės plėtros galimybės

Kaip ir kiekvieną programinę įrangą, kaip ir šią galima tobulinti plečiant jos funkcionalumą ir patogumą vartotojui.

Ateityje į geografinę informacinę sistemą galima integruoti istorinius žemėlapius, tuo sudarant galimybes lokalizuoti ir istorinius vietovardžius. Dėl to bus galima atlikti lyginamąją analizę tarp dabartinių ir istorinių vietovardžių.

Programą bus galima papildyti naujomis tyrimo strategijomis: be dabar esančios siaurinosios ir plečiamosios papildomai gali būti rezultatų suliejimo (du skirtingų tyrimų sąrašai sujungiami į vieną bendrą.) ir susikirtimo (iš dviejų skirtingų tyrimų sąrašų bendrų elementų suformuojamas vienas sąrašas) strategijos.

Kaip ir buvo apžvelgta automatizuoto onomasto darbo modelyje, vėliau bus galima įdiegti statistikos surinkimą ir atvaizdavimą onomastui, kad jis galėtų vizualiai matyti savo tyrimo rezultatus.

Taip pat programą bus galima papildyti įrankiais, leidžiančiais ant žemėlapių vizualiai nupiešti parapijų, seniūnijų ribas taip suteikiant dar daugiau interaktyvumo darant mokslinius tyrimus.

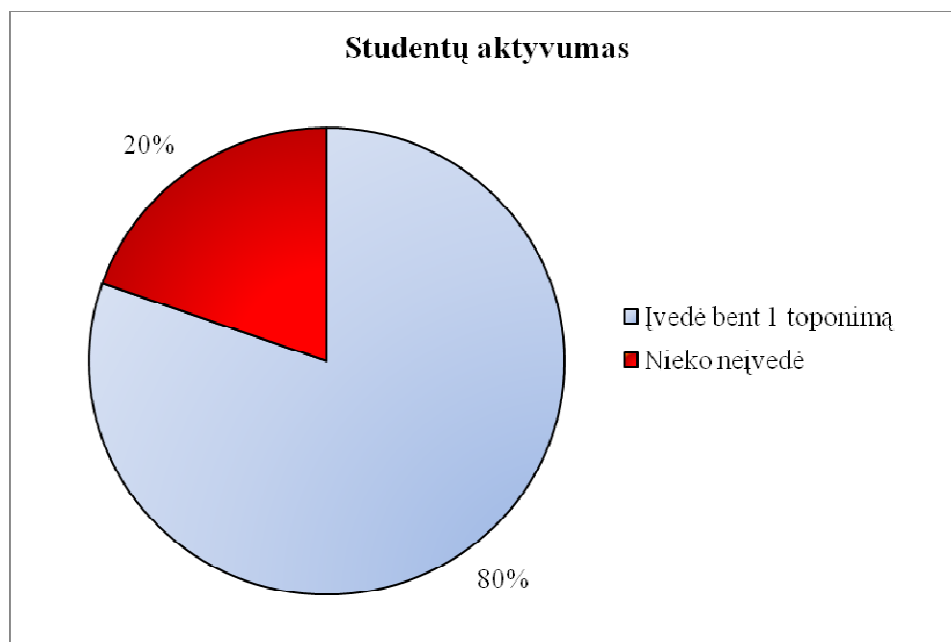
Rezultatų išsaugojimas į MS Excel ar MS Word dokumentą jau yra įtrauktas į būsimų patobulinimų sąrašą.

4. LIETUVOS TOPONIMIKOS TYRIMŲ PROGRAMINĖS ĮRANGOS EKSPERIMENTINIS TYRIMAS

4.1. Vietovardžių įvedimo analizė

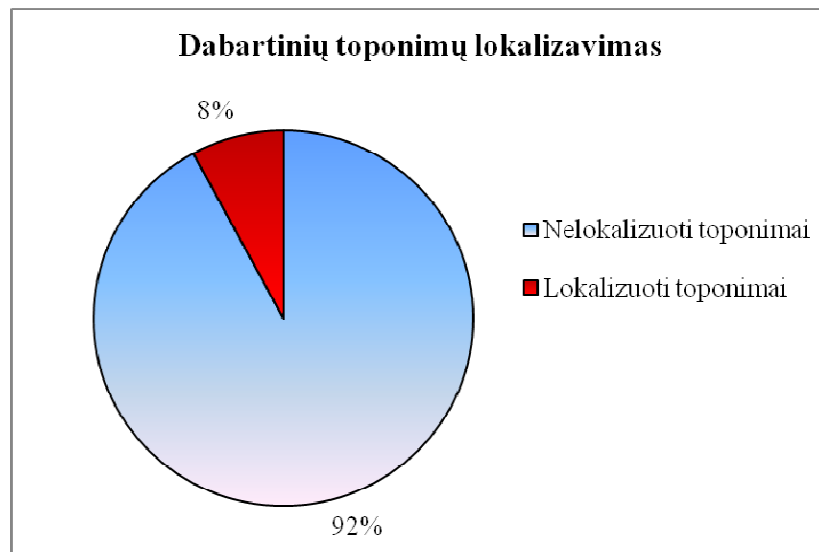
Kadangi Lietuvos programinės įrangos poreikis yra gan nemažas, eksperimentinis sistemos išbandymas buvo vykdomas sąlyginai nedidelėje vartotojų aplinkoje – Vilniaus universiteto Kauno humanitarinio fakulteto lietuvių filologijos ir reklamos studijų programos leksikologijos modulis studentai suvedinėjo informaciją apie vietovardžius į duomenų bazę.

Iš 51 sistemoje užregistruoto studento 2009 vasario – 2009 gegužės mėnesiais bent vieną vietovardį į sistemą įvedė 41 studentas (žr. 14). Kiekvienas iš jų vidutiniškai įvedė po 2,2 vietovardžio. Ir tik vienai studentei pavyko suvesti 5 toponimus. Tai rodo, kad yra labai sunku surinkti informaciją apie vietos vardą arba studentai viską darė paskutinėmis minutėmis.



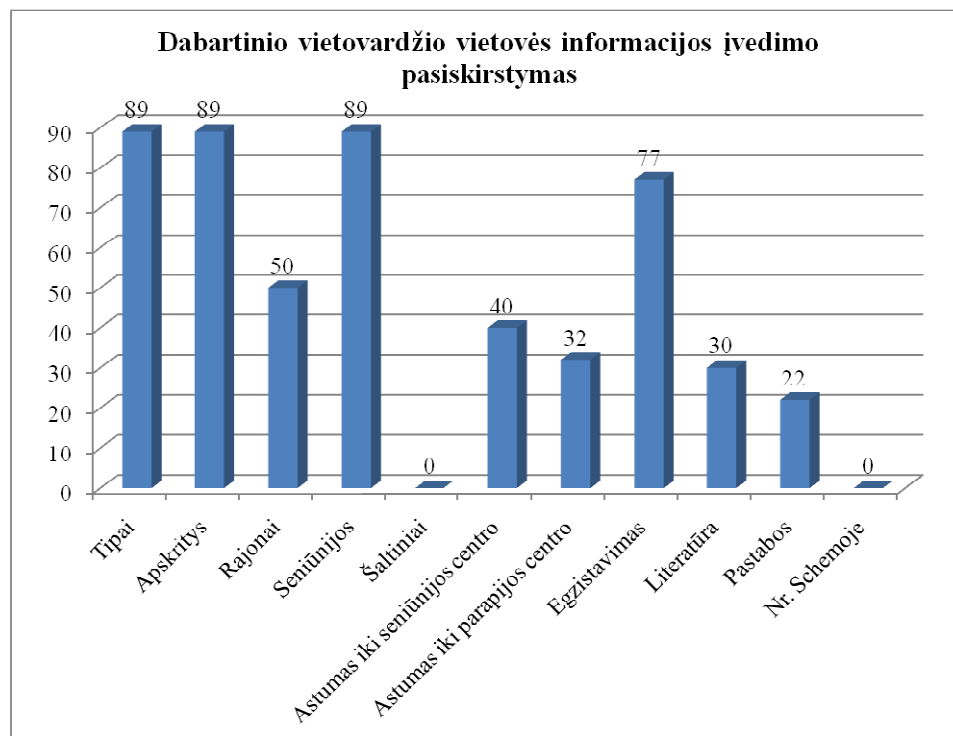
14 pav. Studentų aktyvumo diagrama

Toliau bus apžvelgti studentų darbo rezultatai. Kaip ir buvo minėta anksčiau, mokslininkams yra labai svarbu, kad jie matytų vietovardį žemėlapyje. Studentai iš 89 įvestų vietovardžių lokalizavo vos 7 (žr.). Tai rodo, kad jie gali nustatyti toponimo vietą, tačiau tai yra pakankamai sudėtingas procesas.



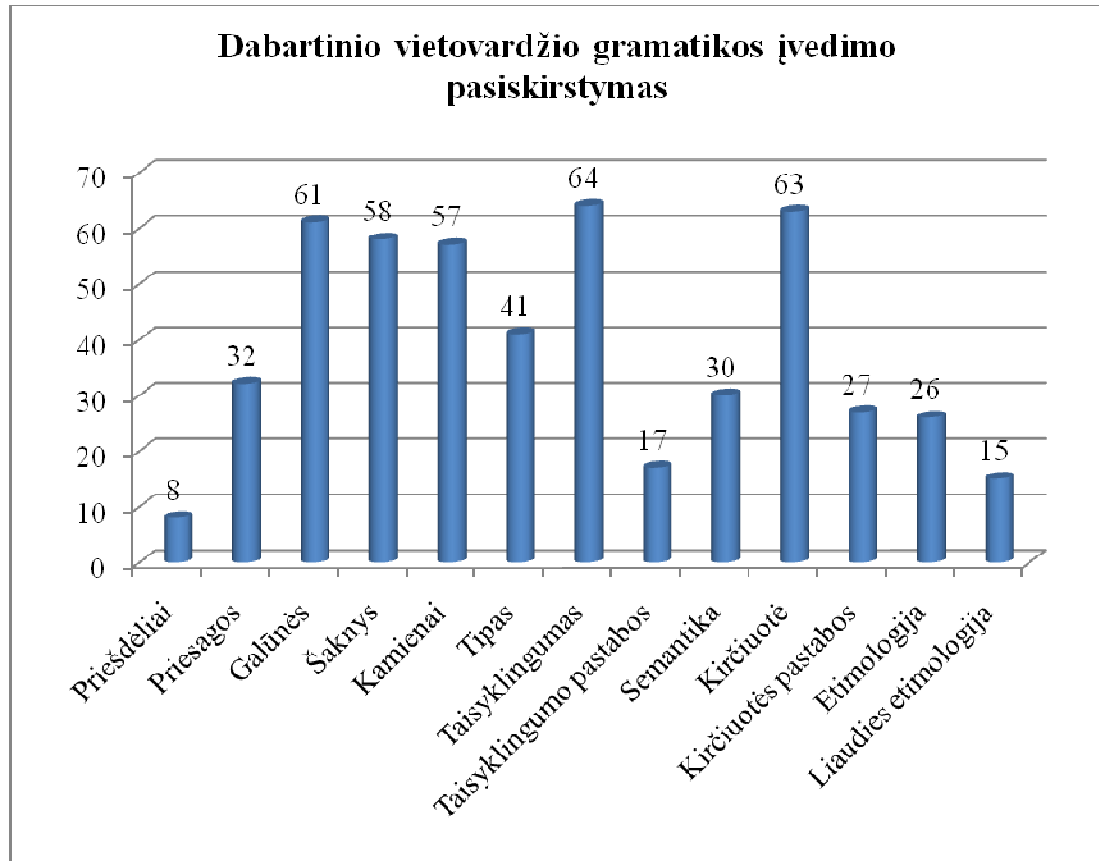
15 pav. Dabartinių toponimų lokalizavimo diagrama

Be lokalizavimo studentams taip pat reikėjo suvedinėti dabartinio vietovardžio vietovės informaciją. Kaip matome 16 jiems geriausiai sekėsi nustatinėti tipus, apskritis, seniūnijas, egzistavimą. Nė prie vieno vietovardžio nebuvo įvestas šaltinis ir numeris schemeje. Tai rodo, kad tai labai neaktuali arba sunkiai prienama informacija. Todėl pagal šiuos rezultatus bus galima patobulinti vartotojo sąsają taip, kad mažiausiai aktualios informacijos laukai būtų prienami, tik esant konkrečiam poreikiui.



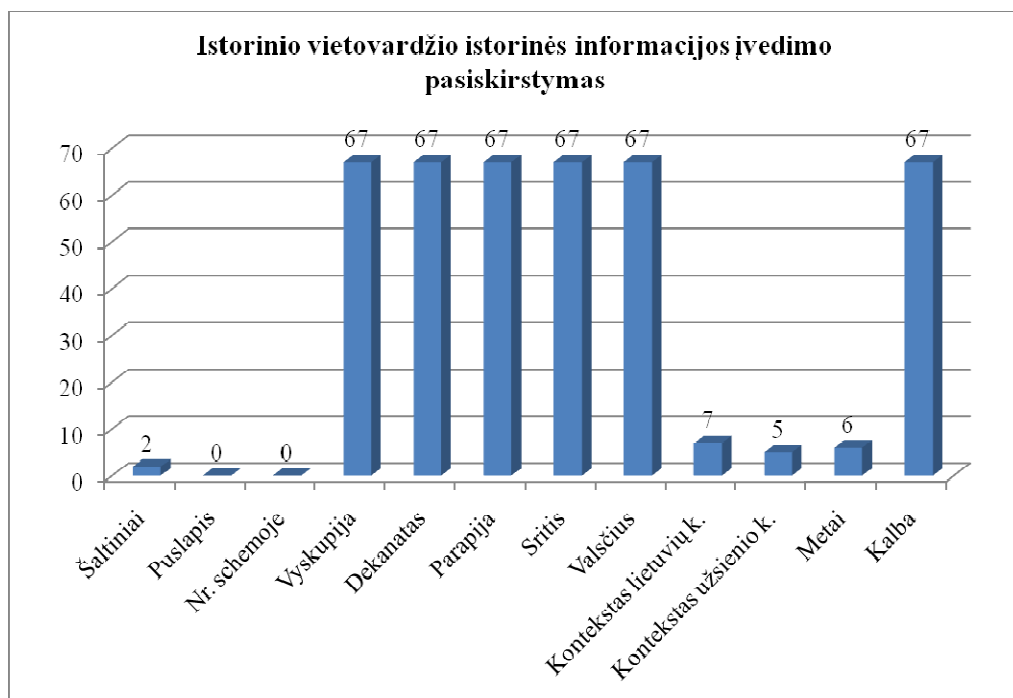
16 pav. Dabartinio vietovardžio vietovės informacijos įvedimo pasiskirstymo diagrama

Didžiąjai daliai vietovardžių (69) buvo užpildytas bent vienas gramatikos laukas. Pagal gramatikos pasiskirstymo diagramą (žr. 17) matyti, kad dauguma toponimų buvo labai elementarios sudėties (šaknis ir galūnė). Daugiausiai vietovardžių buvo nustatytas taisyklingumas ir kirčiuotė.



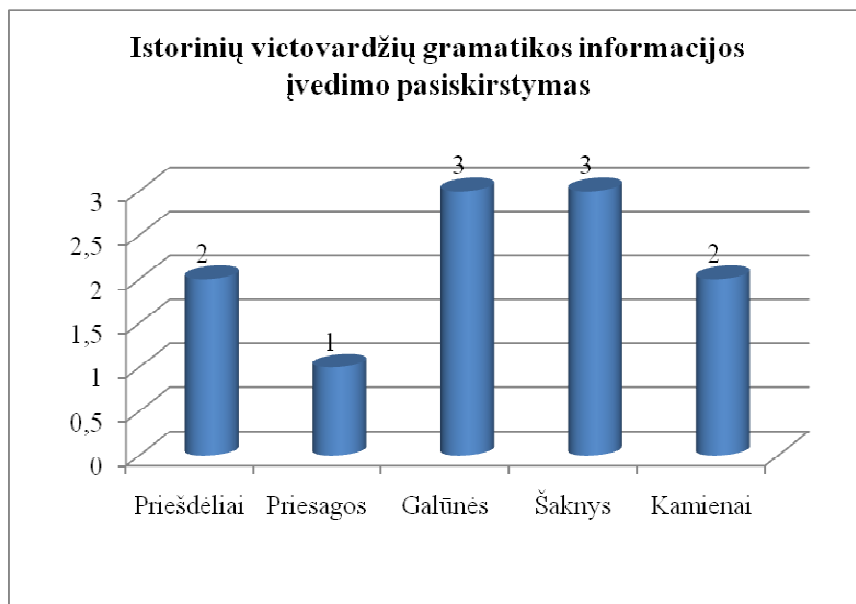
17 pav. Dabartinio vietovardžio gramatikos įvedimo pasiskirstymo diagrama

Su dabartiniais vietovardžiais buvo susieti 67 istoriniai vietovardžiai. Vienam dabartiniam toponimui įvestas tik vienas istorinis. Stebint istorinių vietovardžių istorinės informacijos įvedimo diagramą (žr. 18) išryškėja 6 pagrindiniai laukai, kurie buvo suvesti prie kiekvieno istorinio toponimo (vyskupija, dekanatas, parapija, sritis, valsčius ir kalba). Tai reiškia, kad senojo administracinio ir bažnytinio suskirstymo informacijos gausa yra lengvai prieinama. Nuostabą kelia, kad toponimo vartojimo konteksto užsienio kalbos lauką suvedė prie visų 67 įrašytų vietovardžių, tačiau patį kontekstą įrašė tik prie 5 vietovardžių. Iš to galima daryti išvadą, kad kalbos laukas buvo suprastas klaidingai ir jį reikės aiškiau priskirti prie vartojimo konteksto užsienio kalba.



18 pav. Istorinių vietovardžių istorinės informacijos įvedimo pasiskirstymo diagrama

Studentams išties buvo sunku suvedinėti istorinių vietovardžių gramatikos informaciją. Iš visų 67 įvestų istorinių vietovardžių tik 6 iš jų studentai nurodė bent vieną žodžio dalį (žr. 19). Įdomu tai, kad kiekvienam vietovardžiui buvo nurodomos ne visos jam priklausančios žodžio dalys. Pavyzdžiui šaknys nurodytos tik 3 iš 6 vietovardžių. Tai reiškia, kad studentai išties sunkiai atskiria istorinių vietovardžių žodžio dalis.



19 pav. Istorinių vietovardžių gramatikos informacijos įvedimo pasiskirstymo diagrama

4.1.1. Išvados

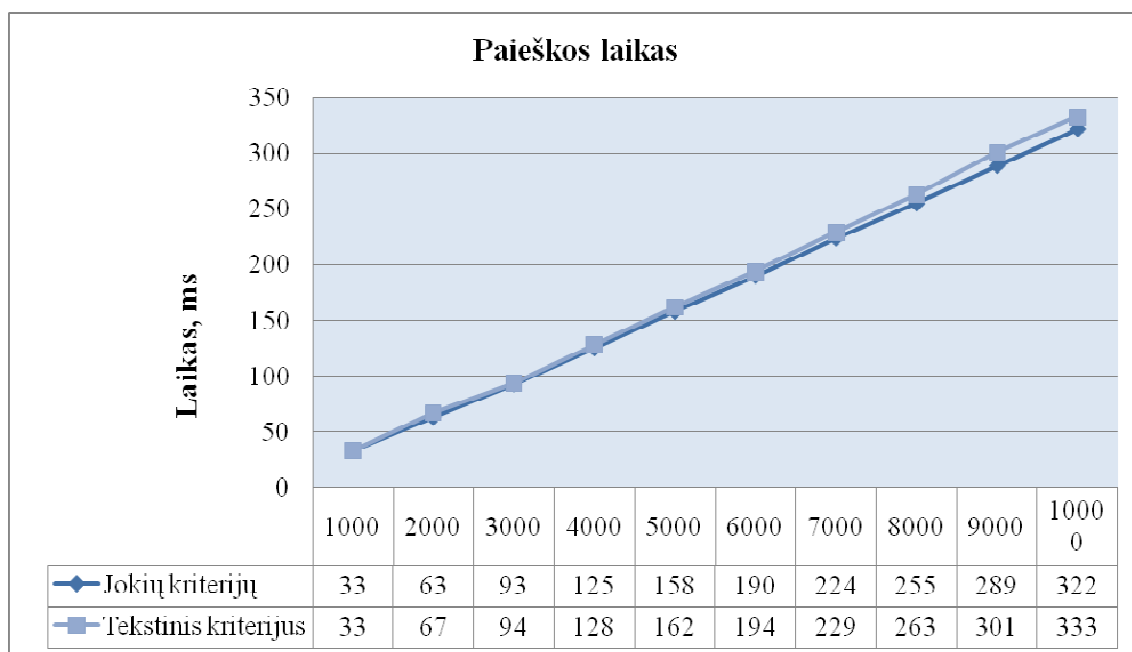
Taigi pagal studentų sistemos eksperimentinio išbandymo rezultatus matyti, kad sistemos funkciniai reikalavimai buvo išpildyti, ja galima realiai naudotis. Buvo identifikuotos vartotojo sąsajos patobulinimo galimybės. Pagal įvestų vietovardžių skaičių ir prie kiekvieno iš jų nurodytos informacijos kiekį matyti, kad duomenų įvedimas yra labai darbiui imlus procesas, tačiau studentai informaciją pildė paskutinėmis dienomis, todėl onomastams dar teks pildyti šią informaciją.

4.2. Paieškos atlikimo ir rezultatų atkūrimo laikų analizė

Sukūrus programinę įrangą itin aktualu yra iširti jos veikimą bei esminius veiksnius, kurie gali jį riboti. Į duomenų bazę surašius realius duomenis buvo pastebėta, kad paieškos laikas yra gan didelis. Dėl to padidėjus duomenų kiekiui programa būtų ribojamas jos veikimas. Toponimų paieška buvo tiriama eksperimento būdu.

Tam buvo sukurti specialūs programinės įrangos papildymai. Jie generavo dalinai atsitiktinius duomenų kiekius ir jais papildydavo duomenų bazę. PĮ buvo tiriama ne interneto serveryje, kur ji realiai veikia, o vartotojo kompiuteryje.

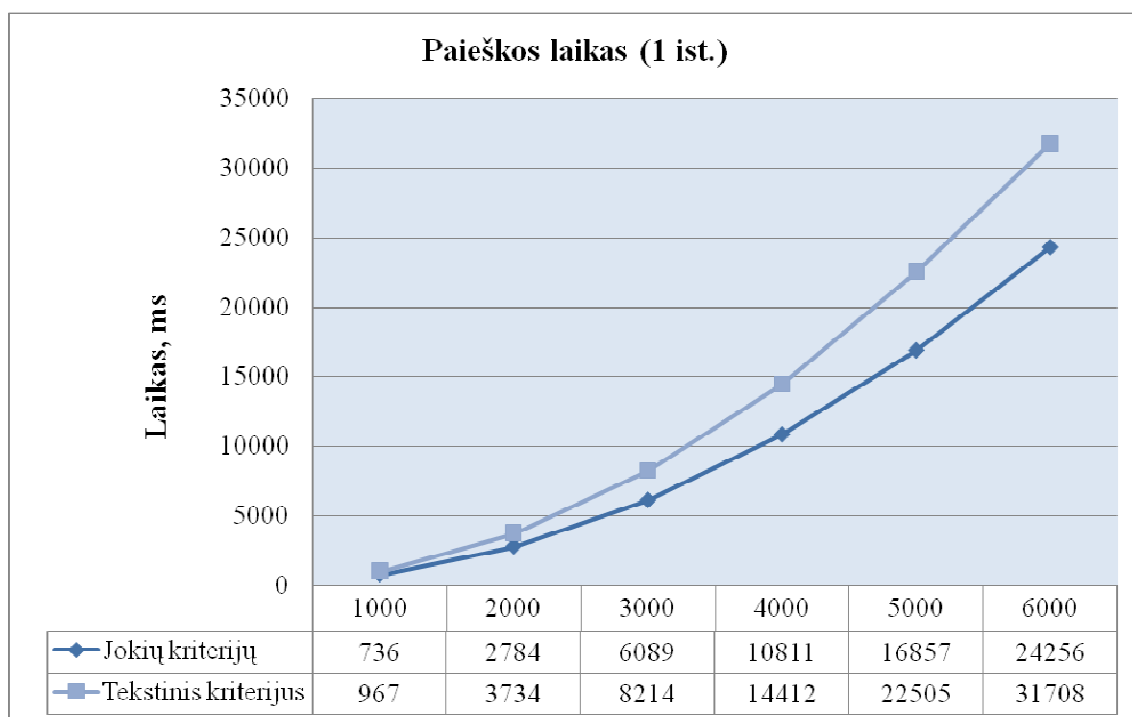
1 eksperimentas. Į duomenų bazę buvo įterpta po 1000 vietovardžių. Tuomet yra skaičiuojamas jų paieškos laikas, kai išvis nenurodomas arba nurodomas tik tekstinis paieškos kriterijus. Tuomet yra užfiksuojami užklausų vykdymo laikai. Tai kartojama 3 kartus ir gale išvedamas vidurkis. Šio eksperimento metu gauti rezultatai pavaizduoti 20 paveikslėlyje.



20 pav. Paieškos laikas, kai duomenų bazėje nėra istorinių vietovardžių.

Akivaizdu, kad paieškos laikas nuo įvestų toponimų skaičiaus priklauso tiesiškai, t.y. paieška tarp 10 000 vietovardžių užtruko iki 350 milisekundžių. Pagal šiuos rezultatus galima teigti, kad paieška tarp 100 000 vietovardžių truktų truputį ilgiau nei tris sekundes. Tad matome, jog tekstinis kriterijus praktiškai neįtakoja paieškos laiko.

2 eksperimentas. Į duomenų bazę buvo įterpta po 1000 toponimų. Kiekvienam dabartiniam vietovardžiui buvo sukurtas vienas istorinis vietovardis. Paieškos identiškos pirmajam eksperimentui: be kriterijų ir su tekstiniu kriterijumi. Tekstinis kriterijus atrenka visus dabartinius ir istorinius toponimus. Tai kartojama 3 kartus ir gale išvedamas vidurkis. Šio eksperimento metu gauti rezultatai pavaizduoti 20 paveikslėlyje.



21 pav. Paieškos laikas, kai vienas vietovardis susietas su vienu istoriniu vietovardžiu.

Eksperimentas buvo baigtas, kai programa nesugebėjo užbaigti paieškos tarp 7000 vietovardžių – tai galėjo nutikti dėl DBVS klaidos. Išanalizavus gautus rezultatus paaiškėjo, kad kreivės yra beveik idealios parabolės. Vadinasi parašytų užklausų sudėtingumas yra $\Omega(n^2)$. Toks užklausų sudėtingumas ir paieškos laikas ne tik apsunkintų vartotojų darbą, bet ir be reikalo eikvotų aparatūrinius resursus.

Užklausos naudotos abiejuose eksperimentuose [24]:

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 (
```

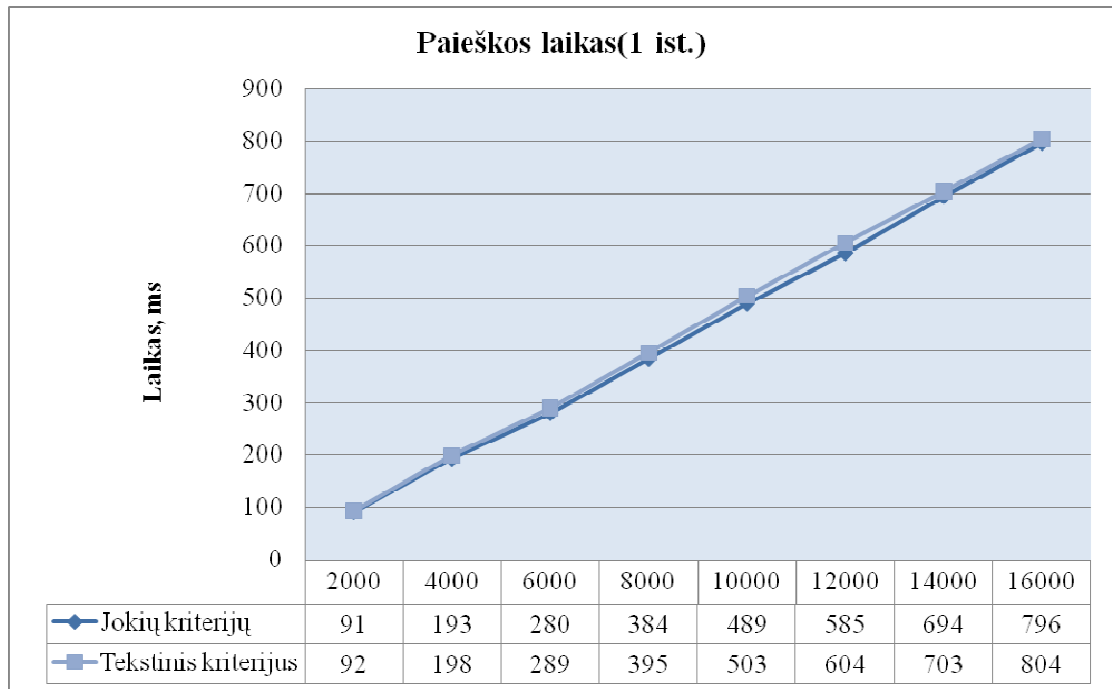
```
`vietovardis`.`gramatika_id` = `gramatikav`.`id` ) LEFT JOIN `gramatika2` AS `gramatika2`
ON ( `gramatika2`.`vietovardis_id` = `vietovardis`.`id` ) LEFT JOIN `gramatikos` AS
`gramatikai` ON ( `istorinis`.`gramatika_id` = `gramatikai`.`id` )WHERE (true)
```

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%')
```

Atlikus aukščiau pateiktų užklausų bei pirmųjų dviejų eksperimentų analizę nustatyta, kad didžiąją laiko dalį duomenų bazių valdymo sistema sugaišta sujungdama dvi lenteles: vietovardžių ir istorinių vietovardžių. Siekiant optimizuoti paieškos laiką istorinių vietovardžių lentelėje atributui *vietovardis_id* buvo sukurtas indeksas.

Trečias eksperimentas. Sukūrus indeksą istorinių vietovardžių lentelėje kartojamas antras eksperimentas. Tai kartojama 3 kartus ir gale išvedamas vidurkis. Šio eksperimento metu gauti rezultatai pavaizduoti 20 paveikslėlyje.



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

Lyginant antrojo ir trečiojo eksperimentu rezultatus akivaizdu, kad pradėjus indeksuoti *vietovardis_id* atributą paieška pradėjo veikti žymiai greičiau, o paieškos užklauskos vykdymo laukas yra tiesiškai priklausomas nuo toponimų kiekio.

Gautų rezultatų palyginimas.

5 pateikiami visų trijų eksperimentų, naudojant tekstinio kriterijaus užklauskas, rezultatai. Tai dar kartą parodo, kad tinkamai indeksavus duomenų bazės laukus paieškos laikas be kriterijų ar tik su tekstiniu kriterijumi yra tiesiškai priklausomas nuo toponimų kiekio.

5 lentelė Lietuvos toponimikos tyrimų programinės įrangos tyrimų rezultatai

Toponimų kiekis	1 eksperimentas (ms)	2 eksperimentas (ms)	3 eksperimentas (ms)
1 000	33	967	92
2 000	67	3734	198
3 000	94	8214	289
4 000	128	14412	395
5 000	162	22505	503
6 000	194	31708	604
7000	229	-	703
8 000	263	-	804
9 000	301	-	92
10 000	333	-	198

Rezultatų atkūrimas iš duomenų bazės

Šis eksperimentas atliktas norint patikrinti projektavimo metu padarytą prielaidą, kad laikas, per kurį yra atkuriami sąrašo rezultatai yra tiesiškai priklausomas nuo rezultatų sąraše skaičiaus.

Buvo išvalyta duomenų bazė ir įvesta 10 000 vietovardžių.

Eksperimentas buvo vykdomas tokia tvarka:

1. Sukuriamas naujas tyrimas;
2. Pirmą tyrimo sąrašą sudaro visi 10000 duomenų bazėje esantys toponimai;
3. Tada vykdoma siaurinama tyrimo strategija - kas kart atfiltruojama po 1000 vietovardžių.
4. Kiekvienai tyrimo šakai (nuo paskutinės iki pirmos) išmatuojamas atkūrimo užklauskos laikas.

Šio eksperimento metu gauti rezultatai pavaizduoti 20 paveikslėlyje.



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

Paieškos atlikimo ir rezultatų atkūrimo laikų analizės išvados

- Analizuojant toponimų paieškos užklauskas buvo rastos ir ištaisytos projektavimo metu padarytos klaidos.
- Tekstiniai kriterijai turi nedaug įtakos vietovardžių paieškos laikui.
- Duomenų bazės valdymo sistema didžiausią užklauskos vykdymo laiko dalį sunaudoja lentelių sujungimui.
- Patvirtinta prielaida, kad tyrimo rezultatų atkūrimo laikas tiesiškai priklauso nuo toponimų kiekio.

IŠVADOS

1. Išanalizavus egzistuojančius toponimikos tyrimų sprendimus Lietuvoje ir Europoje nustatyta, kad kuriama programinė įranga turės tokias naujas galimybes: GIS palaikymą, dabartinių ir istorinių vietovardžių susiejimą, morfeminę istorinių toponimų analizę, tyrimų tęstinumą ir jų viešinimą.
2. Ištirtas Lietuvos vietovardžių tyrėjo darbo procesas ir sukurtas originalus iteracinis modelis.
3. Sukurta Lietuvos toponimikos tyrimų programinė įranga naudoja Google Maps geografinę informacinę sistemą, kad būtų galima ateityje integruoti istorinius žemėlapius, bei lanksčią MVC architektūrą, siekiant paprasčiau ją tobulinti ir palaikyti.
4. Kaupiant onomastikos tyrimams reikalingą duomenų bazę, programinė įranga sėkmingai naudojama Vilniaus universiteto lietuvių filologijos ir reklamos studijų programos leksikologijos modulio studentų mokymo procese.
5. Sukurta automatizuota onomasto darbo vieta užtikrins mokslinių tyrimų tęstinumą ir plėtrą.

LITERATŪRA

- [1] Bartkutė N., Jodelis R. „Toponimija skaitmeniniu pavidalu: Joniškio rajono vietovardžių geolingvistinė duomenų bazė“. Konferencijos „Europos studijos: dialektai ir kalbos politika“ leidinys. 2007 m.
- [2] 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>>
- [3] 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>>
- [4] 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>>
- [5] „Paikannimikartasto“. Kotimaisten kielten tutkimuskeskus. 1997–2002 m. [Žiūrėta 2007 11 13], prieiga internete <<http://www.kotus.fi/index.phtml?s=383>>
- [6] „Paikannimikirja“. Kotimaisten kielten tutkimuskeskus. 2005-2007 m. [Žiūrėta 2007 11 13], prieiga internete <<http://www.kotus.fi/index.phtml?s=443>>
- [7] Open Source DBMS Comparison: MySQL vs. PostgreSQL [Žiūrėta 2007 11 13], prieiga internete <<http://www.tometasoftware.com/MySQL-vs-PostgreSQL.asp>>
- [8] PostgreSQL vs. MySQL vs. Commercial Databases: It's All About What You Need [Žiūrėta 2007 11 13], prieiga internete <<http://www.devx.com/dbzone/Article/20743/0/page/2>>
- [9] Comparison of relational database management systems [Žiūrėta 2007 11 13], prieiga internete <http://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems>
- [10] Reuven M. Lerner „MySQL Deserves a Double Take“. 2007 m.. [Žiūrėta 2007 11 13], prieiga internete <<http://www.linuxjournal.com/article/9571>>
- [11] John Koenig „Why MySQL sets criteria for well-tempered stacks“. 2005 m. [Žiūrėta 2007 11 13], prieiga internete <<http://www.itmanagersjournal.com/feature/5622>>
- [12] GIS programinė įranga [Žiūrėta 2007 11 13], prieiga internete <<http://www.hnit-baltic.lt/DesktopDefault.aspx?tabID=3569&alias=hnit-baltic&lang=lt-LT>>
- [13] Yahoo Maps [Žiūrėta 2007 11 13], prieiga internete <<http://maps.yahoo.com>>
- [14] Microsoft Live Search [Žiūrėta 2007 11 13], prieiga internete

- <<http://local.live.com>>
- [15] Google Maps [Žiūrėta 2007 11 13], prieiga internete
<<http://maps.google.com>>
- [16] Google Maps [Žiūrėta 2007 11 13], prieiga internete
<http://en.wikipedia.org/wiki/Google_Maps>
- [17] 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>>
- [18] Mike Williams „Google Maps API Tutorial“ [Žiūrėta 2007 11 13], prieiga internete <<http://www.econym.demon.co.uk/googlemaps/index.htm> >
- [19] Google Maps API Tutorial [Žiūrėta 2007 11 13], prieiga internete
<<http://econym.googlepages.com/custommap.htm>>
- [20] J. Dargham. MVC Extension: Evaluation According to Rating Criteria and Possible Integration. Internet and Multimedia Systems and Applications, 2007.
- [21] Wikipedia. CakePHP. [Žiūrėta 2009 04 06], prieiga internete
<<http://en.wikipedia.org/wiki/CakePHP>>
- [22] Hiddenbrains. CakePHP - Choice of Developers. [Žiūrėta 2009 04 06], prieiga internete <<http://hiddenbrains.wordpress.com/2008/12/02/cakephp-choice-of-developers/>>
- [23] Gregory Beaver. phpDocumentor QuickStart. [Žiūrėta 2009 05 05], prieiga internete
<http://manual.phpdoc.org/HTMLSmartyConverter/HandS/phpDocumentor/tutorial_phpDocumentor.quickstart.pkg.html>
- [24] V. Kriščiūnas „Automatizuotų vietovardžių tyrimų modelis ir jo realizacija“. 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.

Onomastas – onomastikos mokslininkas.

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

GIS – geografinė informacinė sistema.

Lingvistika – bendras mokslų, tiriančių natūralias kalbas, pavadinimas.

Morfologija – Lingvistikos dalis, nagrinėjanti žodžių formų struktūrą

JavaScript – tai scenarijų programavimo kalba.

CVS – versijų valdymo sistema.

SVN – versijų valdymo sistema.

MVC – Programų architektūrinis modelis.

PHP – Plačiai paplitusi dinaminė interpretuojama programavimo kalba.

IS – Informacinė sistema

PĮ – Programinė įranga

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.

DB – duomenų bazė.

DBVS – duomenų bazių valdymo sistema.

phpMyAdmin – MySQL duomenų bazės administravimo įrankis.

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

CGI (*Common Gateway Interface*) — protokolas, apibrėžiantis, kaip turi bendrauti WWW serveris ir jo vykdomos programos, skirtos iš naršyklės gautai informacijai apdoroti ir/arba dinaminiais puslapiams generuoti.

Mikrotoponimas - smulkaus geografinio objekto, žinomo nedideliame plote, pavadinimas.

PRIEDAI

1 priedas. Publikacija „Naujų technologijų taikymas diachroninėje toponimikoje“.

2 priedas. Publikacijos priėmimui spausdinti tarptautiniame Vilniaus universiteto leidžiamame žurnale „Respectus psychologicus“ pažyma.

3 priedas. Vilniaus universiteto toponimikos tyrėjų ir lietuvių kalbos reklamos studijų profesinės kalbos kultūros modulio dėstytojų doc. dr. Ilonos Mickienės ir Linos Bačiūnaitės-Lužinienės pasirašytas Lietuvos toponimikos tyrimų programinės įrangos diegimo aktas.

Jonas Andrijauskas

Kauno technologijos universitetas

Informatikos fakultetas

Studentų g. 50-411, LT-51368 Kaunas

El. p. glodas@gmail.com

Domėjimosi sritis: lingvistinių tyrimų kompiuterizavimas ir standartizavimas, interneto sistemų kūrimas.

Lina Bačiūnaitė-Lužinienė

Vilnius universitetas

Kauno humanitarinis fakultetas

Muitinės g. 8, LT-44280 Kaunas

El. p. lina.baciunaite@vukhf.lt

Domėjimosi sritis: onomastika, kalbos kultūra, leksikologija.

Vytas Kriščiūnas

Kauno technologijos universitetas

Informatikos fakultetas

Studentų g. 50-411, LT-51368 Kaunas

El. p. vytas@mail.lt

Domėjimosi sritis: lingvistinių tyrimų kompiuterizavimas ir standartizavimas, interneto sistemų kūrimas.

NAUJŲJŲ TECHNOLOGIJŲ TAIKYMAS DIACHRONINĖJE TOPONIMIKOJE

Žinių visuomenėje vis daugiau dėmesio skiriama humanitarinių ir tikslųjų mokslų dermei bei siekiama naujasias technologijas, ypač informacines, kuo plačiau taikyti lingvistikoje, todėl pastebimos skirtingų sričių specialistų bendradarbiavimo tendencijos. Lietuvoje sinchroninėje onomastikoje informacinės technologijos, nors ir labai nespārčiai, tačiau taikomos. Panašios tendencijos pastebimos ir kitose Europos šalyse, kurios šiek tiek anksčiau ėmė domėtis, kaip palengvinti, supaprastinti ir automatizuoti onomastų darbą.

Daugelis onomastikos specialistų dar ir dabar pradiniam tyrimo etapui turi skirti labai daug laiko, kadangi duomenų rinkimo ir apdorojimo stadijos nėra automatizuotos, o tęstinių tyrimų, kur vieno onomasto pradėta darba visai kita linkme galėtų tęsti kitas, iš viso nėra. Naujosios technologijos Lietuvoje iki šiol buvo taikytos tik sinchroninės toponimikos tyrimams. Sinchroniniams toponimikos tyrimams sukurtų informacinių sistemų pagrindu ne visada galima skaitmeninti diachroninės onomastikos duomenis, todėl norima sukurti tokią informacinę sistemą, kurioje būtų siejami sinchroninės ir diachroninės toponimikos duomenys.

KEY WORDS: diachroninė ir sinchroninė onomastika, toponimika, gyvenamosios vietos pavadinimas, archyvai, toponimikos informacinė sistema, automatizuotasis iteracinis tyrimo modelis, onomasto darbo modelis.

Ivadas

Naujosios technologijos vis plačiau taikomos tokiose lingvistikos mokslo šakose, kurioms anksčiau bet kokia tikslųjų mokslų invazija buvo nebūdinga. Jei fonetikoje tradicinę oscilografą jau pakeitė kiti skaitmeniniai garsų matavimo prietaisai su atmintimi (*Kas yra svarbiausia renkantis oscilografą*. Prieiga per internetą: <http://rtn.elektronika.lt/rtn/0202/kas.html>), tai leksikologijoje bet koks žodžių skaitmeninimas yra labiau pirminis nei patobulintas reiškinys. Be jau spėjusių išpopuliarėti elektroninių žodynų, kurie gana dažnai yra popierinės jų versijos analogas, norima sukurti tam tikras informacines sistemas, kurios palengvintų konkrečios humanitarinių mokslų srities specialistų darbą bei susietų atliekamus eksperimentus su kitais atliktais ar atliekamais humanitarinio pobūdžio tyrimais. Akivaizdu, kad dirbdami pavieniui nei informatikai, nei filologai tokios informacinės sistemos nesukurs. Norint pasiekti optimalų rezultatą, būtina skirtingų sričių specialistams dirbti kartu.

Šio straipsnio tikslas – nustatyti sinchroninės ir diachroninės toponimikos universalijas ir pateikti projektą, kaip automatizuoti diachroninę toponimiką tiriančio specialisto darbą.

Diachroninės toponimikos tyrimai nuo sinchroninės toponimikos tyrimų skiriasi tiek duomenų rinkimo, tiek ir jų apdorojimo etapuose. Būtent dėl šios priežasties duomenų rinkimo analogijos tarp sinchroninių ir diachroninių toponimikos tyrimų yra beveik neįmanomos. Duomenų apdorojimo etape sąsajų galima išvystyti daugiau, tačiau laikyti jas universaliomis nebūtų tikslinga.

Iki šiol Lietuvos mokslo įstaigose dirbantys tiek sinchroninę, tiek diachroninę toponimiką tiriantys onomastai dirbo pavieniui (D. Deltuvienė, I. Mickienė (Vilniaus universitetas), N. Taluntytė, R. Endzelytė (Vytauto Didžiojo universitetas), G. Blažienė (Vilniaus pedagoginis universitetas, dabar jau Lietuvių kalbos institutas), N. Bartkutė (Šiaulių universitetas), D. Kiseliūnaitė, L. Simutytė (Klaipėdos universitetas), V. Garliauskas (Vilniaus Abraomo Kulviečio vidurinė mokykla)) ir savo tyrimais negalėjo dalintis su kitais, kadangi jų tiriama medžiaga arba iš viso nebuvo skaitmeninama, arba tai buvo daroma tik tiek, kiek to reikia jų pačių tyrimams. Bendra visų atliekamų toponimų tyrimų informacinė sistema palengvintų mokslininkų darbą ir pagreitintų bet kokių duomenų filtravimą bei paskatintų atlikti daugiau komparatyvistinių tyrimų, nes visi duomenys būtų toje pačioje informacinėje sistemoje.

Valstybinės lietuvių kalbos komisijos 2006 m. veiklos ataskaitoje buvo rašoma, kad „iš rankraštinės istorijos šaltinių išrašyta 10 000 etninių vietovardžių. Į duomenų bazę įvesti 6 885 vietovardžiai. Sukurtos McAccess atrankos bei sujungimo užklauskos pagal duomenis iš kelių susijusių lentelių, kurios pildomos su sukurtos bazės Istoriniai etniniai vietovardžiai įvedimo moduliu. Sukurtas ir tobulintas bazės peržvalgos modulio algoritmas pagal atrankos užklauskų projektą. Peržvalgos modulio algoritmas realizuotas Visual Basic priemonėmis“ (*Valstybinės lietuvių kalbos komisijos 2006 m. veiklos ataskaita*. Prieiga per

internetą: <http://www3.lrs.lt/owa-bin/owarepl/inter/owa/U0162220.doc>). Šia duomenų baze iki šiol taip ir negalima viešai visiems norintiems naudotis.

Toponimikos skaitmeninimo požiūriu daugiausiai yra nuveikęs Lietuvių kalbos instituto Vardyno skyrius, kurio viena iš veiklos sričių yra dabartinės ir istorinės toponimikos problemų tyrimas. Visiems vartotojams yra laisvai prieinamas internetinis vietovardžių žodynas (*Vietovardžių žodynas*. Prieiga per internetą: <http://lkz.mch.mii.lt/Vietovardziai/>), sukurtas pagal knygą (*Vietovardžių žodynas*) išleistą jo variantą, tačiau jame tikrai nėra daugelio iš 300 000 Vardyno skyriaus kartotekoje iš istorinių šaltinių užrašytų vietovardžių.

Kadangi universitetų ir instituto mokslininkai tyrimus atlieka tik po to, kai patys susirenka tyrimui reikalingą medžiagą, tai jie duomenų rinkimo etape sugaišta daug laiko. Kartais tas pačias kartotekas peržiūrėti tenka skirtingiems panašius tyrimus atliekantiems mokslininkams. Norėdami lingvistinius istorinių toponimų duomenis papildyti ekstralingvistiniais duomenimis, tyrėjai turi domėtis istorija, etnologija, mitologija, antropologija, kultūros ir bažnyčios istorija, etc. Visi ekstralingvistiniai faktai yra kitų mokslo sričių tyrimų objektai, todėl būtų tikslinga tai tiriančius mokslininkus įtraukti į diachroninės toponimikos informacinės sistemos kūrimo komandą.

Egzistuojančios sinchroninės toponimikos informacinės sistemos ir jų nepritaikomumas diachroninės toponimikos tyrimuose

Daugelis pasaulyje egzistuojančių naujų technologijų, kurios atliekamus tyrimus padaro lengvesniais, kaskart sukuriamos vis tobulesnės naudoti, o jei iš principo naudojama priemonė yra optimali, tai tobulinamas jos dizainas, automatizuojami kai kurie elementai. Tokiais pasiekimais tikrai negali pasigirti nei vienas toponimiką tyrinėjantis onomastas, nes šiai lingvistikos šakai dar tik bandoma pritaikyti naujų technologijų galimybes.

Palyginus jau egzistuojančias onomasto darbą lengvinančias, bet tik sąlyginai informacinėmis sistemomis vadinamas, elektronines toponimų kartotekas, duomenų bazes ar MS Excel lentelėse sukurtus filtrus, matyti, kad nei vienas iš jų negali būti diachroninės toponimikos informacinės sistemos prototipas.

Kadangi Lietuvoje kol kas nėra sukurta vietovardžių analizės informacinė sistema, kuri gali patenkinti tiek sinchroninės, tiek diachroninės toponimikos tyrėjus, būtina įvertinti, kas šiuo klausimu nuveikta kitose Europos šalyse. Dar 1990 m. Suomijos kalbų tyrimų institute buvo pradėta kurti suomių toponimų duomenų bazė (MIKKULAINEN. Prieiga per internetą: <http://scripta.kotus.fi/www/artikkelit/toponyms.html>), kuri leido sujungti skirtingų tipų toponimų duomenis. Suomijoje yra apie 2,5 milijono vietovardžių, todėl pradžioje buvo nuspręsta į duomenų bazę suvesti ir lokalizuoti tik 10% šių vietovardžių. Per 6 metus suomiams pavyko suskaitmeninti 100 000 vietovardžių, o 2006 m. jų buvo į duomenų bazę įtraukta 800 000. Deja, jokios informacijos apie technines galimybes nėra publikuojama. Skelbiami tik rezultatai, pasiekti dirbant šia programa.

Remiantis šia vietovardžių duomenų baze ir onomastinės analizės informacine sistema Suomijos kalbų tyrimų institutas įgyvendino keletą mokslinių tiriamųjų projektų. Vienas iš jų nagrinėjo, kaip keitėsi

vietovardžių pavadinimai devyniuose Suomijos kaimuose nuo 1950 – 1960 m. iki 1980 m. (MIIKKULAINEN. Prieiga per internetą: <http://scripta.kotus.fi/www/artikkelit/toponyms.html>). Kitas projektas paskatino remiantis toponimų duomenų baze sukurti atlasą iliustruojantį 234 labiausiai paplitusių suomių vietovardžių elementų pasiskirstymą Suomijos teritorijoje. Atlase nurodytos koordinatės, fizinės savybės ir kita aktuali toponimų bei geografinė informacija. Šis atlasas nėra elektroninis, tačiau suteikia galimybę atlikti įvairių vietovardžių paieškų.

1995 m. norvegų tyrinėtojas E. Soderholm iš Tromso universiteto pasiūlė projektą „Toponymic atlas of the Barents region“ („Toponiminis Barento regiono atlasas“). Projekto reikalingumas buvo grindžiamas poreikiu plėsti kultūrinius, mokslinius, politinius bei ekonominius ryšius tarp Barento regiono šalių. Įkvėpti šios idėjos bei 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“). Sistema buvo sukurta 2001 m. ir nuo tada yra prieinama internetu (Toponymic Research System of Northwest Russia: The TORIS System. Prieiga per internetą: <http://toris.krc.karelia.ru/papers/source/toris98/index.phtml>), tačiau jai būdingi tik onomastinės analizės įrankiai.

Lietuvoje pirmieji skaitmeninę regioninę toponimijos duomenų bazę (BARTKUTĖ, N., JODELIS, R. Prieiga per internetą: http://www.neryte.lt/uploads/db_pranesimas_su_leidiniui.doc), kurioje suskaitmeninta 6 000 Joniškio rajono vietovardžių bei kiekvienas iš jų lokalizuotas skaitmeniniame žemėlapyje, 2007 m. sukūrė N. Bartkutė ir R. Jodelis. Ji buvo skirta atlikti sinchroninės toponimikos tyrimams. Kadangi ši sistema nenaudoja pažangios geografinės informacinės sistemos, turi labai nedaug paieškos filtrų ir jau beveik 2 metus netobulinama (šiuo metu net nepasiekiamą internetu), todėl ji negali būti naudojama kaip modelis skaitmeninant visos Lietuvos vietovardžius.

Dėl visų šių priežasčių buvo pasirinkta kurti visai kitokią informacinę sistemą, kuri tiktų ir sinchroninei, ir diachroninei toponimikai skaitmeninti, lokalizuoti ir aprašyti.

Vienas iš toponimų tyrimų metodų yra struktūrinė analizė. Kadangi lietuvių kalba priklauso indoeuropiečių kalbų grupei, o suomių – fino-ugrų, todėl suomių sukurta informacinė sistema negali būti kuriamajai prototipinė morfologijos požiūriu. Rusų sukurta sinchroninė toponimų informacinė sistema pritaikyta tik rusų kalbos morfologijai, tačiau dėl praeityje ypač glaudžių baltų ir slavų santykių, jos modelis galėtų būti taikomas ir planuojamai sukurti toponimų informacinei sistemai. Kadangi jau egzistuojančios toponimų informacinės sistemos pritaikytos sinchroniniams tyrimams, todėl ten atliekama tik dabartinių toponimų morfeminė analizė. Diachroninės onomastikos specialistai norėtų, kad būtų galima atlikti senųjų toponimų arba dabartinių toponimų senųjų lyčių morfeminę analizę, o ji būtų siejama su sinchroniniais tyrimais.

Tiek sinchroninę, tiek diachroninę toponimiką tiriantys specialistai lokalizuoja ir kartografuoja gyvenamųjų vietų vardus. Onomasto darbas palengvėtų, jei toponimų informacinėje sistemoje suskaitmeninti vietovardžiai būtų pavaizduoti žemėlapyje. Dabar dar egzistuojančių gyvenamųjų vietų vardų kartografavimas nėra toks sudėtingas, tačiau norint kartografuoti jau nebeegzistuojančius, būtina kuo tiksliau lokalizuoti vietovę arba istorinius žemėlapius susieti su dabartiniais žemėlapiais. Tam būtina naudoti

geografinę informacinę sistemą¹, kuri veikia geografiniu principu, t. y. susieta ne tik su aprašomąja, bet kartu ir su koordinuota erdvine informacija. Sistema skirta skaitmeniniams koordinuotiems erdvėje duomenims kaupti, saugoti, vaizduoti, redaguoti, integruoti bei analizuoti. Pagrindinis geografinės informacinės sistemos privalumas – operavimas erdvine (koordinuota, orientuota erdvėje) informacija, todėl informacija apie tam tikrą regioną gali būti pateikiama konkrečios teritorijos tematinų žemėlapių, atspindinčių laikmetį ar kitą kultūrinę informaciją, pavidalu (Pav. 1).



Pav. 1. Geografinės informacinės sistemos informaciniai sluoksniai

Šaltinis: sukurta autorių.

Viena iš pageidaujamų onomastų toponimikos tyrimų informacinių sistemų funkcijų yra galimybė dirbti bet kokioje darbo vietoje, todėl labai svarbu, kad informacine sistema būtų galima naudotis internetu ir ji būtų pritaikyta įvairioms naršyklėms. Iki šiol sukurtose toponimų informacinėse sistemose nebuvo galima pradėto vieno tyrimo tęsti kitam mokslininkui, todėl lingvistiniai darbai stokojo profesionalių ekstralingvistinių įžvalgų, o istorikų, kultūrologų, archeologų tyrimai nevisada remiasi kalbiniu palikimu. Šios problemos nebebūtų tokios opios, jei dar nebaigtas tyrimas galėtų būti viešinamas kitiems mokslininkams.

Lentelė Nr. 1 Toponimikos tyrimų informacinių sistemų funkcijos

Funkcionalumas	Suomijos kalbų tyrimų instituto IS	Rusijos lingvistikos, literatūros ir istorijos instituto IS	Joniškio rajono vietovardžių geolingvistinė DB	Planuojama sukurti toponimų IS
Lietuvių kalbos morfologijos palaikymas	–	–	+	+
Galimybė dirbti internete	–	+	+	+
Kartografinė eksplikacija	+	–	+	+
GIS palaikymas	+	–	–	+
Istorinių žemėlapių integravimas	–	–	–	+
Grafinis paieškos rezultatų atvaizdavimas	–	–	–	+
Dabartinių toponimų morfeminė analizė	+	+	+	+
Istorinių toponimų morfeminė analizė	–	–	–	+
Dabartinių ir istorinių toponimų susiejimas	–	–	–	+

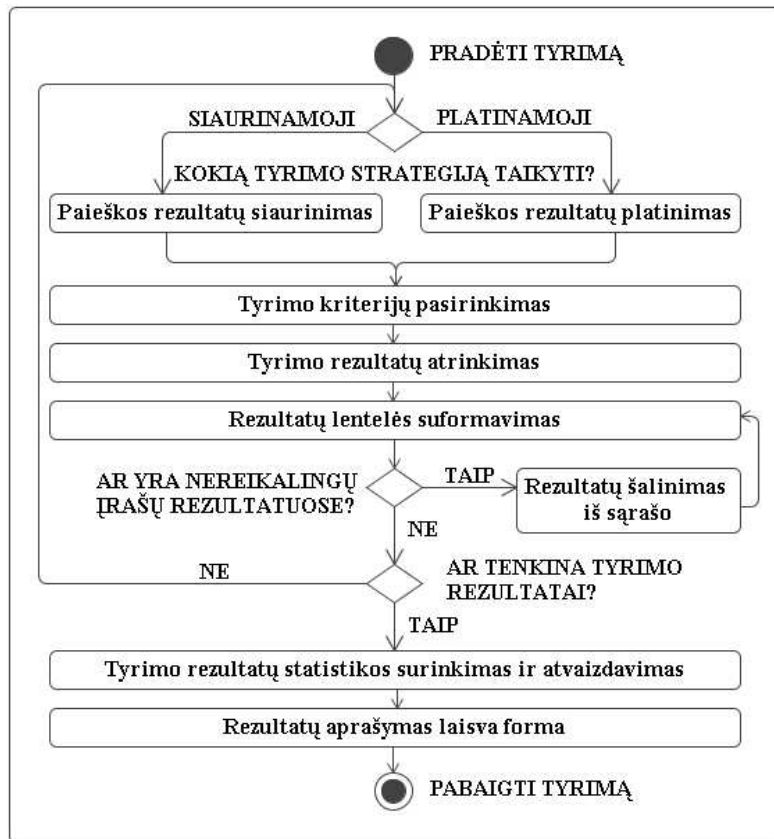
¹ Kartografinė eksplikacija planuojamoje sukurti toponimų informacinėje sistemoje turėtų remtis <http://maps.google.com/>.

Funkcionalumas	Suomijos kalbų tyrimų instituto IS	Rusijos lingvistikos, literatūros ir istorijos instituto IS	Joniškio rajono vietovardžių geolingvistinė DB	Planuojama sukurti toponimų IS
Tyrimų medžio sudarymas	—	—	—	+
Tyrimų viešinimo galimybė	—	—	—	+

Automatizuotojo iteracinio tyrimo modelio taikymas diachroninėje toponimikoje

Šiuo metu beveik visa diachroninės onomastikos tyrimų medžiaga yra pateikiama popieriniu pavidalu, todėl jos nors ir pats paprasčiausias suskaitmeninimas užima daug laiko. Be to ne visada reikalinga tyrimo medžiaga yra lengvai prieinama, dažnai skirtingi tyrimui reikalingi šaltiniai išsibarstę po Lietuvą arba net už jos dabartinės teritorijos ribų. Norint išvengti pakartotinio šaltinių rinkimo ir analizės, būtina juos tinkamai suskaitmeninti, suvedant į bendrą duomenų bazę, prieinamą visiems tos srities mokslininkams. Pamažu papildant duomenų bazę vis naujais objektais, bus sudaryta galimybė atlikti plačios duomenų apimties mokslinius tyrimus. Bendros duomenų bazės sukūrimas ir yra būtina tokio tipo tyrimo automatizavimo sąlyga.

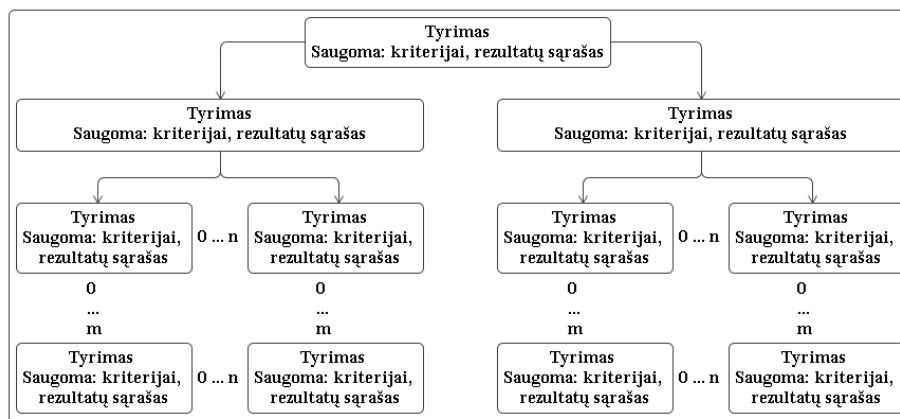
Sukūrus tokią duomenų bazę būtų labai daug laiko sutaupoma automatizuojant duomenų paiešką pagal kuo daugiau kriterijų. Iki šiol tokios paieškos užimdavo daug laiko, nes buvo atliekamos rankomis (Pav. 2). Reikia pažymėti, kad rankomis atliekamą paiešką nevisada gali atitikti formaliai aprašyti paieškos kriterijai, todėl išdiferenciabus automatiškai dažnai reikia peržiūrėti, ar ten nėra nereikalingų rezultatų (onomasto euristiškai atmetami rezultatai). Taip pat egzistuoja rizika, kad kai kurie klaidingai aprašyti vietovardžiai nebus įtraukti į sąrašą.



Pav. 2. Dabartinis onomasto darbo proceso modelis

Šaltinis: sukurta autorių.

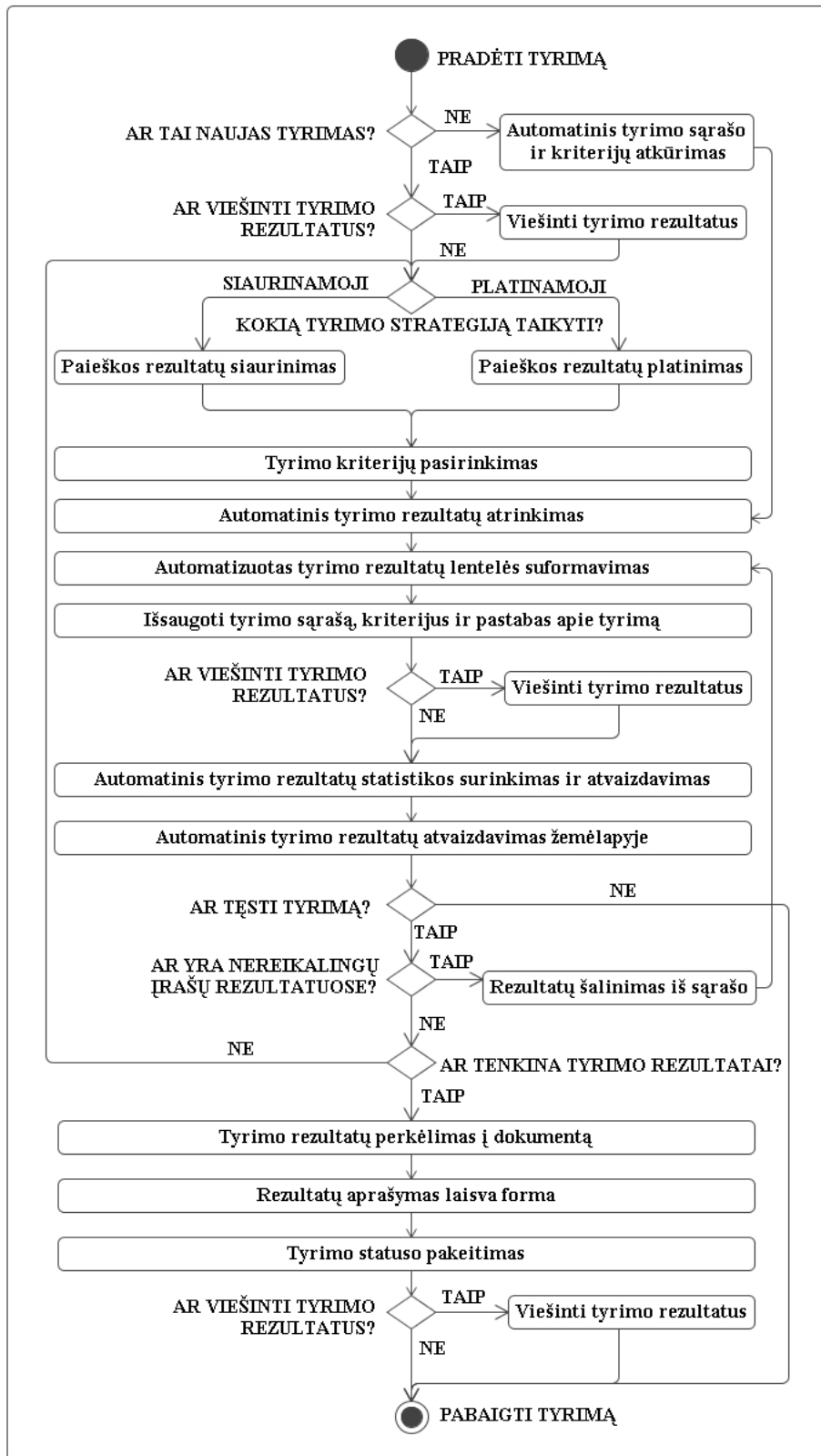
Tyrimo tarpinių rezultatų išsaugojimas ir testinumas vienas – iš svarbiausių aspektų, atliekant automatizuotus vietovardžių tyrimus. Ši galimybė neegzistuoja jokiose kitose nagrinėtose sistemose. Taikant siūlomą sprendimą, mokslininkai ne tik galėtų išsaugoti tyrimo rezultatus, tačiau bet kuriuo metu tyrimą tęsti nuo anksčiau gautų rezultatų, kadangi informacinė sistema galėtų atkurti net kelis anksčiau nurodytus kriterijus bei pagal tuos kriterijus gautus rezultatus (Pav. 3). Taip pat galima tęsti viešuosius vienas kito tyrimus, o tai skatintų mokslininkų bendradarbiavimą jau atliekamų tyrimų metu.



Pav. 3. Iteracinis onomasto darbo proceso modelis

Šaltinis: sukurta autorių.

Kiekviena 3 pav. medžio šaka atspindi automatizuotą onomasto darbo proceso modelį, kuriame dėl anksčiau išvardintų priežasčių būtų taikomos tokios onomastų darbo naujovės: 1) tarpinių rezultatų (sąrašo ir kriterijų) išsaugojimas ir automatinis atkūrimas, kuris skatintų mokslinių tyrimų tęstinumą; 2) tyrimo rezultatų viešinimas, kuris užtikrintų mokslinių tyrimų plėtrą ir mokslininkų bendradarbiavimą; 3) automatinis tyrimo rezultatų vaizdavimas žemėlapyje; 4) tyrimo statuso keitimas (jis gali būti nebaigtas, laukiantis atsakymo ir baigtas) (Pav. 4).



Pav. 4. Automatizuotas onomasto darbo proceso modelis

Šaltinis: sukurta autorių.

Toponimų informacinės sistemos duomenys leistų nesunkiai išfiltruotus rezultatus pateikti lentele, kartografuoti ir apdoroti statistiškai, todėl toponimikos specialistui būtų daug lengviau aprašyti gautus

rezultatus. Norint automatizuoti ir palengvinti onomasto darbą, būtina įgyvendinti šiuos kuriamos informacinės sistemos prototipo reikalavimus:

- Sudaryti galimybę įvesti vietovės ir gramatikos informaciją tiek dabartiniams, tiek istoriniams vietovardžiams. Svarbu, kad istorinių vietovardžių informacija (lingvistiniai ir ekstralingvistiniai dalykai) būtų tiesiogiai susieta su dabartine gyvenamosios vietos vardo lytimi.

- Sukurti paiešką pagal vietovės ir/ arba gramatikos atributus tarp dabartinių ir istorinių vietovardžių. Svarbu, kad vietovardis būtų surandamas ne tik pagal dabar egzistuojančią pavadinimo lytį, bet ir pagal skirtingus morfologinius vienetus.

- Pateikti dinaminę paieškos rezultatų lentelę, kuri leistų parodyti konkrečiam onomasto tyrimui reikalingus duomenis.

- Suteikti galimybę tyrimą plėsti (turimo sąrašo papildymas naujais gyvenamųjų vietų vardais) arba siaurinti (pasirinkti, pvz.: laiko, lokalizacijos, etc. kriterijus).

- Lokalizuoti toponimą žemėlapyje, pasirenkant vieną arba kelis geografinės informacijos sluoksnius.

- Tyrimų tęstinumas suteiktų galimybę greičiau atlikti naują tyrimą, kadangi kaskart nereiktų pradėti nuo pradžių.

- Kirčiuoti lietuviškus žodžius, rašyti lietuviškais, lenkiškais, latviškais, rusiškais rašmenimis.

Išvados

1. Kiekvienas tyrimas turi tris pagrindines stadijas: duomenų rinkimo, duomenų apdorojimo, išvadų ir tendencijų pateikimo, tačiau nuo paties tyrimo priklauso, kuri iš šių stadijų užtrunka ilgiausiai. Tiek diachroninės, tiek sinchroninės onomastikos tyrimams būdinga ilgai truncančios duomenų rinkimo ir apdorojimo stadijos, todėl norima sukurti tokią informacinę sistemą, kuri palengvintų onomasto darbą.

2. Diachroninę toponimiką tiriantis specialistas dirbdamas kartu su informacinės technologijas išmanančiu žmogumi gali sukurti informacinę sistemą, kuri automatizuotų daugelį dabar dar rankomis atliekamų onomasto darbų, iš kurių laiko atžvilgiu ilgiausiai trunka rūšiavimas pagal pasirinktus kriterijus. Naujųjų technologijų taikymo humanitariniuose moksluose pavyzdžių atsiranda vis daugiau, tačiau konkrečių onomastikai skirtų duomenų bazių ar informacinių sistemų yra vos kelios (Lietuva, Rusija, Suomija) ir tos pačios negalėtų būti prototipinės dėl to, kad yra skirtos konkrečiam tyrimui, nebetobulinamos arba nepritaikytos lietuvių kalbos sistemai.

3. Kuriama informacinė sistema būtų daug funkcionalesnė ir turėtų keletą funkcijų (integruoti istoriniai žemėlapiai, senųjų vietovardžių lyčių morfeminė išraiška, sietina su dabartine toponimo lytimi, tyrimo medis, tyrimų viešinimas), kurių nėra nei vienoje dabar egzistuojančioje toponimų informacinėje sistemoje.

4. Kuriama toponimų informacinė sistema ne tik sietų diachroninių ir sinchroninių toponimų lyčių sąrašus, morfologinius ir etimologinius duomenis, bet ir suteiktų galimybę kitų sričių mokslininkams pateikti daug ekstralingvistinių duomenų, iš kurių daugelis galėtų būti kartografuojami skirtingų informacinių sluoksnių žemėlapiuose. Kadangi informacinė sistema visada galėtų būti papildoma naujais duomenimis,

todėl prie šios sistemos vienu metu galėtų laisvai prisijungti ir naujus tyrimus atlikti, arba tęsti savo ar kolegų pradėtus eksperimentus, daugiau nei vienas mokslininkas. Tokie tyrimai skatintų skirtingų sričių specialistų (lingvistai, istorikai, etnologai, kultūrologai, sociologai, etc.) bendradarbiavimą.

Literatūra

1. BARTKUTĖ, N., JODELIS, R. Toponimija skaitmeniniu pavidalu: Joniškio rajono vietovardžių geolingvistinė duomenų bazė. *Europos studijos: dialektai ir kalbos politika*. Prieiga per internetą: http://www.neryte.lt/uploads/db_pranesimas_su_leidiniui.doc [Žiūr. 2009-04-15].
2. *Kas yra svarbiausia renkantis oscilografą*. Prieiga per internetą: <http://rtn.elektronika.lt/rtn/0202/kas.html> [Žiūr. 2009-03-18].
3. MIIKKULAINEN, R. *Toponyms Data Bank*. Prieiga per internetą: <http://scripta.kotus.fi/www/artikkelit/toponyms.html> [Žiūr. 2009-03-20].
4. MIIKKULAINEN, R. *Toponyms Data Bank*. Prieiga per internetą: <http://scripta.kotus.fi/www/artikkelit/toponyms.html> [Žiūr. 2009-03-20].
5. *Toponymic Research System of Northwest Russia: The TORIS System*. The Institute of Linguistics, Literature and History of Russia. Prieiga per internetą: <http://toris.krc.karelia.ru/papers/source/toris98/index.phtml> [Žiūr. 2007-11-13].
6. *Valstybinės lietuvių kalbos komisijos 2006 m. veiklos ataskaita*. Prieiga per internetą: <http://www3.lrs.lt/owa-bin/owarepl/inter/owa/U0162220.doc> [Žiūr. 2009-03-19].
7. *Vietovardžių žodynas*. Internetinis leidimas. Vilnius: Lietuvių kalbos institutas, 2007. ISBN 978-9955-704-23-2. Prieiga per internetą: <http://lkz.mch.mii.lt/Vietovardziai/> [Žiūr. 2009-03-19].
8. *Vietovardžių žodynas*. Vilnius: Lietuvių kalbos institutas, 2002. ISBN 5-420-01497-1.



**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



**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 Jono Andrijausko sukurta „Lietuvos toponimikos 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