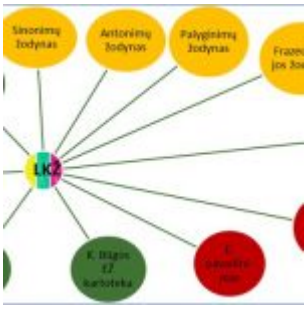


„Lietuvių kalbos žodyno“ ir jo kartotekų transformacijos skaitmeniniame amžiuje

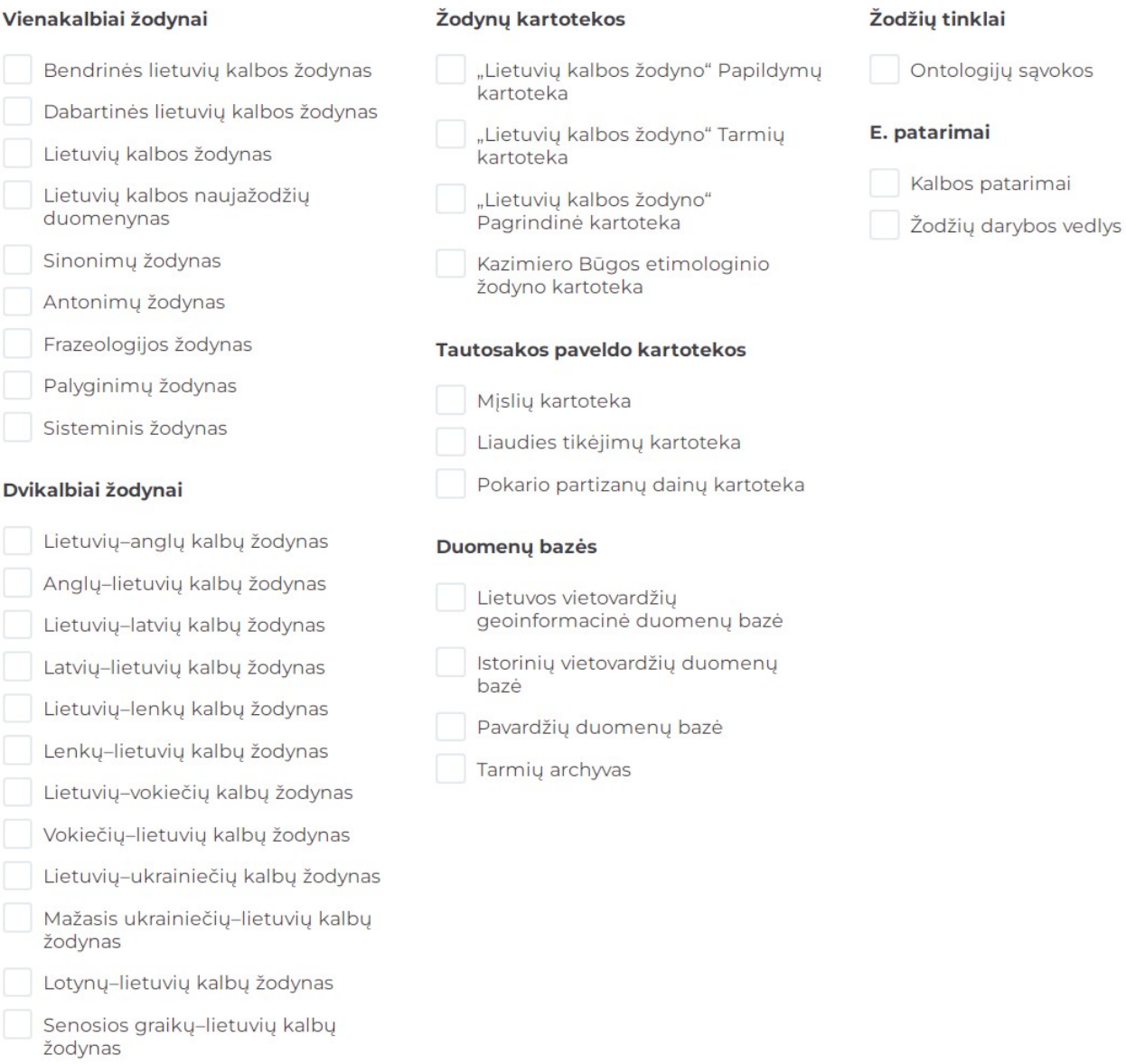
Nuo 2021 m. lietuvių kalbos vartotojai Lietuvoje ir užsienyje gali naudotis modernizuota lietuvių kalbos išteklių informacine sistema „E. kalba“ (https://ekalba.lt/), sukurta Lietuvių kalbos institute (apie šią sistemą jau buvo rašyta „Pasaulio lietuvyje“, žr. Urbanavičienė 2022).

Jolita **URBANAVIČIENĖ**
Lietuvių kalbos institutas

Priminsime, kad „E. kalba“ buvo kuriama dviem etapais. 2012–2015 m. Lietuvių kalbos institutas kartu su partneriais – Lietuvių literatūros ir tautosakos institutu, Vilniaus universitetu ir tuomečiu Lietuvos edukologijos universitetu – sukūrė Lietuvių kalbos išteklių informacinę sistemą (LKIIS), kurią sudarė 11 suskaitmenintų vienakalbių ir dvikalbių žodynų, 5 kartotekos, elektroninės šių išteklių duomenų bazės ir jų valdymo įrankiai[1]. Antrasis sistemos kūrimo etapas vyko 2018–2021 m. LKIIS‘o modernizavimas[2] plėtotas dviem kryptimis: kiekybiškai, t. y. integruojant naujus išteklius, ir kokybiškai, t. y. plečiant sistemos funkcionalumus. Lietuvių kalbos išteklių informacinės sistemos kokybine plėtra siekta įgyvendinti tris funkcionalumus: 1) padidinti lietuvių kalbos elektroninių išteklių prieinamumą; 2) užtikrinti didesnį išteklių patogumą; 3) turimus išteklius panaudoti išvestiniams produktams kurti, t. y. sukurti išteklių tarpusavio integralumą. Modernizuojant „E. kalbą“ į sistemą įtraukti trys itin svarbūs žodynai, kurie iki šiol internete buvo randami tik atskirose svetainėse: 1)

„Bendrinės lietuvių kalbos žodynas“ (BLKŽ, https://ekalba.lt/bendrines-lietuviu-kalbos-zodynas/); 2) *„Lietuvių kalbos naujažodžių duomenynas“* (NžD, https://ekalba.lt/naujazodziai/); 3) *„Lietuvių kalbos žodynas“* (LKŽ, https://ekalba.lt/lietuviu-kalbos-zodynas/). Kuriamą svetainę „E. kalba“ siekta padaryti kuo intuityvesnę, patogesnę naudoti, atitinkančią tiek Lietuvoje, tiek už jos ribų esančių vartotojų lūkesčius. Pvz., visi „E. kalboje“ esantys lietuvių kalbos ištekliai pritaikyti naudoti tiek kompiuteriuose, tiek ir mobiliuosiuose įrenginiuose; sukurta galimybė popierinių kartotekų korteles padidinti per visą įrenginio ekraną, kad būtų patogu perskaityti; skaitmeninių kortelių duomenis galima kopijuoti ar atsisiųsti **pdf* formatu; atliekant paieškas numatyta galimybė raides su lietuviškais ar kitų kalbų diakritikais surinkti naudojant „E. kalboje“ įdiegtas klaviatūras, esančias šalia paieškos laukelių, ir kt.

Šiuo metu lietuvių kalbos išteklių informacinę sistemą „E. kalba“ sudaro 35 ištekliai[3], kurie suskirstyti į tris dalis: „E. žodynai“, „E. kartotekos ir duomenų bazės“ ir „E. paslaugos“ (žr. 1 pav.).



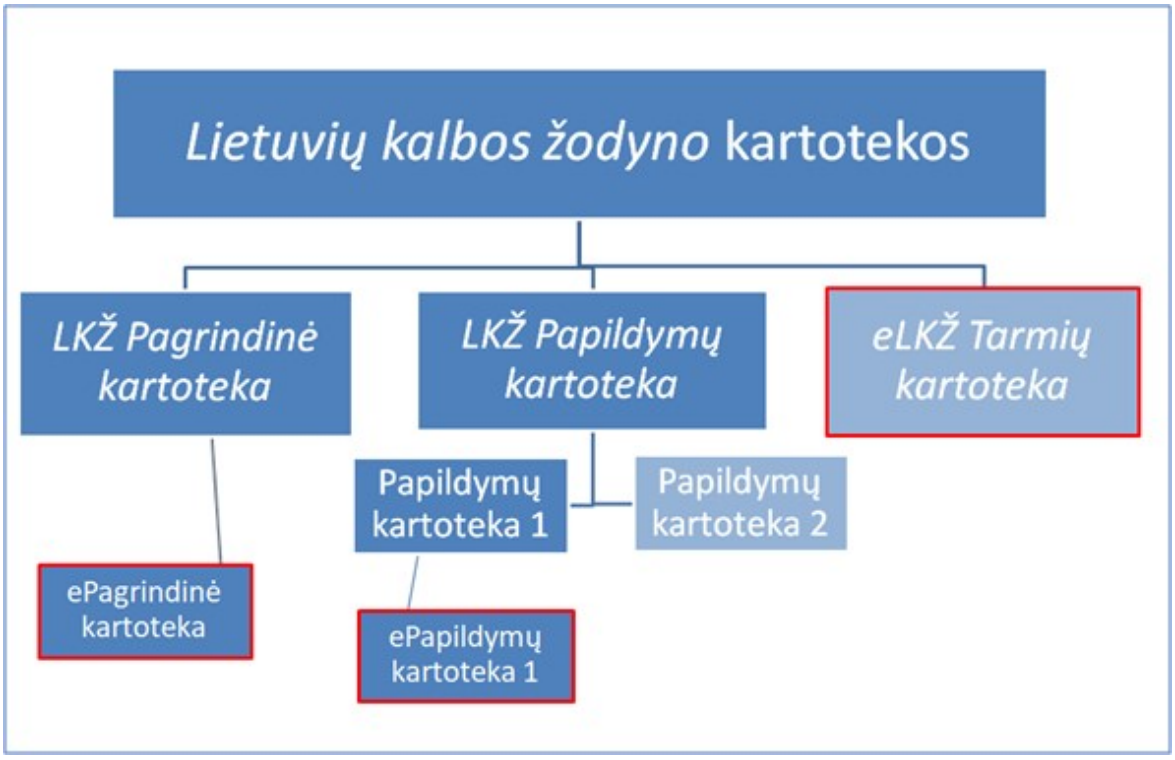
Lietuvių kalbos išteklių informacinė sistema „E. kalba“ (https://ekalba.lt/)

Visose šiose dalyse reikšmingą vietą užima „Lietuvių kalbos žodyno“ – milžiniškos apimties lietuvių kalbos tezauro – duomenys. Šis žodynas yra didžiausias sistemos „E. kalba“ išteklius. Į „E. kalbą“ integruotas ne tik LKŽ tekstas, bet ir suskaitmenintos LKŽ kartotekos, be to, sukurtos elektroninės šių išteklių duomenų bazės bei jų paieškos sistemos. Šiame straipsnyje paanalizuosime, kokią įtaką lietuvių kalbos išteklių informacinei sistemai padarė „Lietuvių kalbos žodyno“ (toliau – LKŽ) duomenų integravimas, taip pat aptarsime LKŽ leksikografinių duomenų pritaikymą kuriant bendrą lietuvių kalbos žodžių tinklą bei išvestinius kalbos produktus. LKŽ medžiaga buvo naudojamasi rašant kitus į „E. kalbą“ įtrauktus žodynus bei kuriant „E. paslaugų“ produktus – „E. pavadinimą“ ir „Žodžių prasmių tinklą“. Atskirai pristatysime kiekvieną iš šių išteklių bei jų santykį su LKŽ.

„Lietuvių kalbos žodynas“ ir jo kartotekos. „Lietuvių kalbos žodynas“ yra baigtinis XVI–XX a. lietuvių kalbos leksikos šaltinis, naujais kalbos faktais jis nebepildomas ir neatnaujinamas. Tačiau Lietuvių kalbos institute saugomos keturios su „Lietuvių kalbos žodynu“ susijusios kartotekos, kurių duomenys gali papildyti LKŽ straipsnius ir kurios „E. kalbos“ dėka randamos vienoje sistemoje:

1) *LKŽ Pagrindinė kartoteka* (LKŽ PgK, https://ekalba.lt/lkz-pagrindine-kartoteka/); 2) *LKŽ Papildymų kartoteka* (LKŽ PpK, https://ekalba.lt/lkz-papildymu-kartoteka/); 3) *LKŽ Tarmių kartoteka* (LKŽ TK, https://ekalba.lt/lkz-tarmiu-kartoteka/); 4) *Kazimiero Būgos etimologinio žodyno kartoteka* (BgEK, https://ekalba.lt/BugosEtimologinioZodynoKartoteka/).

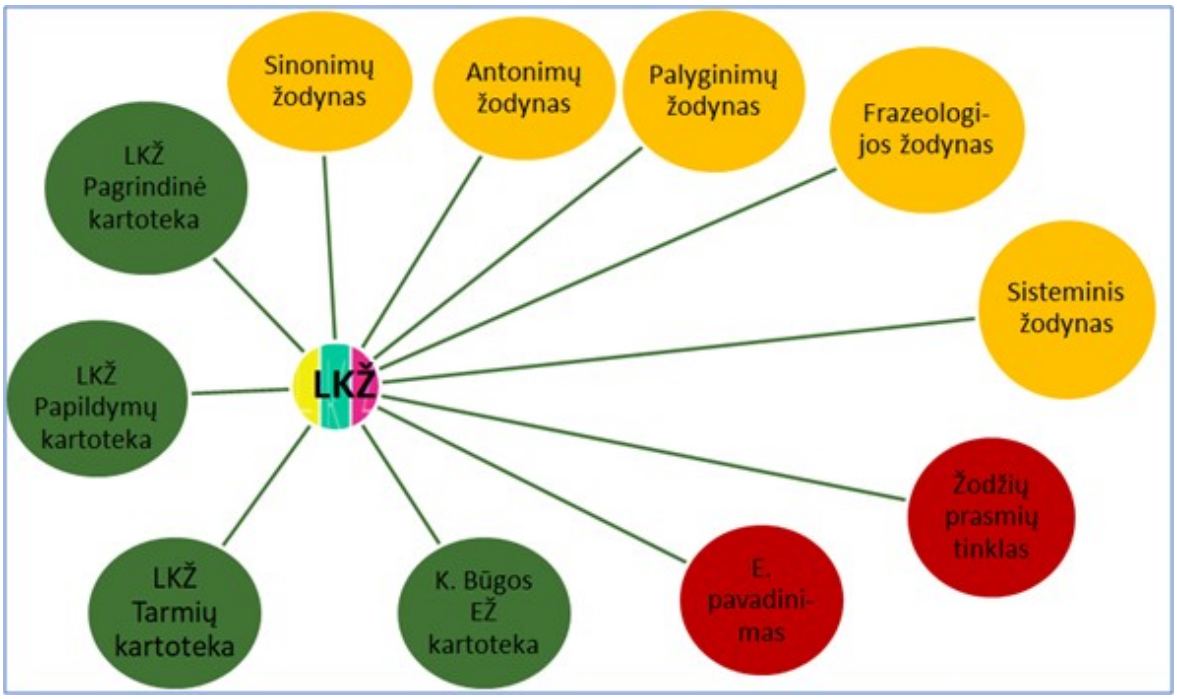
Šios kartotekos skiriasi savo apimtimi, medžiagos pobūdžiu, duomenų struktūra ir jų pateikimo principais. Schemoje (žr. 2 pav.) šviesesne spalva pažymėtos LKŽ kartotekos, kurios tebėra pildomos – tai *LKŽ Tarmių kartoteka* ir 2-oji *LKŽ Papildymų kartoteka* (dar kitaip – *Naujausių tarminių papildymų kartoteka*). Paveikslėlyje raudonai apvestos kartotekos, kurių visos ar dalis kortelių yra nuskenuotos ir randamos „E. kalboje“ (*LKŽ Pagrindinė kartoteka*, *1-oji LKŽ Papildymų kartoteka*). Prie pastarųjų priskirtina ir *LKŽ Tarmių kartoteka* – pirmoji elektroninė LKŽ kartoteka, kuri neturi popierinių kortelių vaizdo, o visa surinkta medžiaga saugoma skaitmeniniu pavidalu.



„Lietuvių kalbos žodyno“ kartotekų struktūra

Visos šios kartotekos kartu sudėjus sudaro daugiau kaip 5,2 mln. leksikografinių vienetų ir gali būti laikomos elektroniniu duomenynu, kuriame greitai ir patogiai galima gauti patikimų empirinių duomenų. O kadangi „Lietuvių kalbos žodynas“ ir su juo susijusios kartotekos randami vienoje sistemoje, vartotojai gali nesunkiai susirasti reikiamo žodžio LKŽ straipsnį bei jį papildyti kartotekų medžiaga.

„Lietuvių kalbos žodynas“ ir vienakalbiai žodynai. „E. žodynų“ dalyje vartotojai gali rasti devynis vienakalbius lietuvių kalbos žodynus – be anksčiau minėtų „Lietuvių kalbos žodyno“, „Bendrinės lietuvių kalbos žodyno“, „Dabartinės lietuvių kalbos žodyno“, „Lietuvių kalbos naujažodžių duomenyno“, čia esama „Sinonimų žodyno“ (SinŽ, https://ekalba.lt/sinonimu-zodynas/), „Antonimų žodyno“ (AntŽ, https://ekalba.lt/antonimu-zodynas/), „Frazeologijos žodyno“ (FrzŽ, https://ekalba.lt/frazeologijos-zodynas/), „Palyginimų žodyno“ (PlgŽ, https://ekalba.lt/palyginimu-zodynas/), „Sisteminio lietuvių kalbos žodyno“ (SisŽ, https://ekalba.lt/sisteminis-lietuviu-kalbos-zodynas/perziura).



„Lietuvių kalbos žodyno“ (LKŽ) sąsajos su kitais žodynais, kartotekomis ir e. paslaugomis

Pastarieji vienakalbiai specialieji žodynai parengti remiantis kitų žodynų medžiaga, iš jų svarbiausias – „Lietuvių kalbos žodynas“ bei gausūs jo kartotekų duomenys. Kai kurie žodynai (pvz., „Frazeologijos žodynas“) sudaryti remiantis LKŽ principais, nes žodynų autoriai (Jonas Paulauskas, Antanas Lyberis, Irena Ermanytė, Klementina Vosylytė) rašė ir redagavo LKŽ tekstą, taigi priklausė

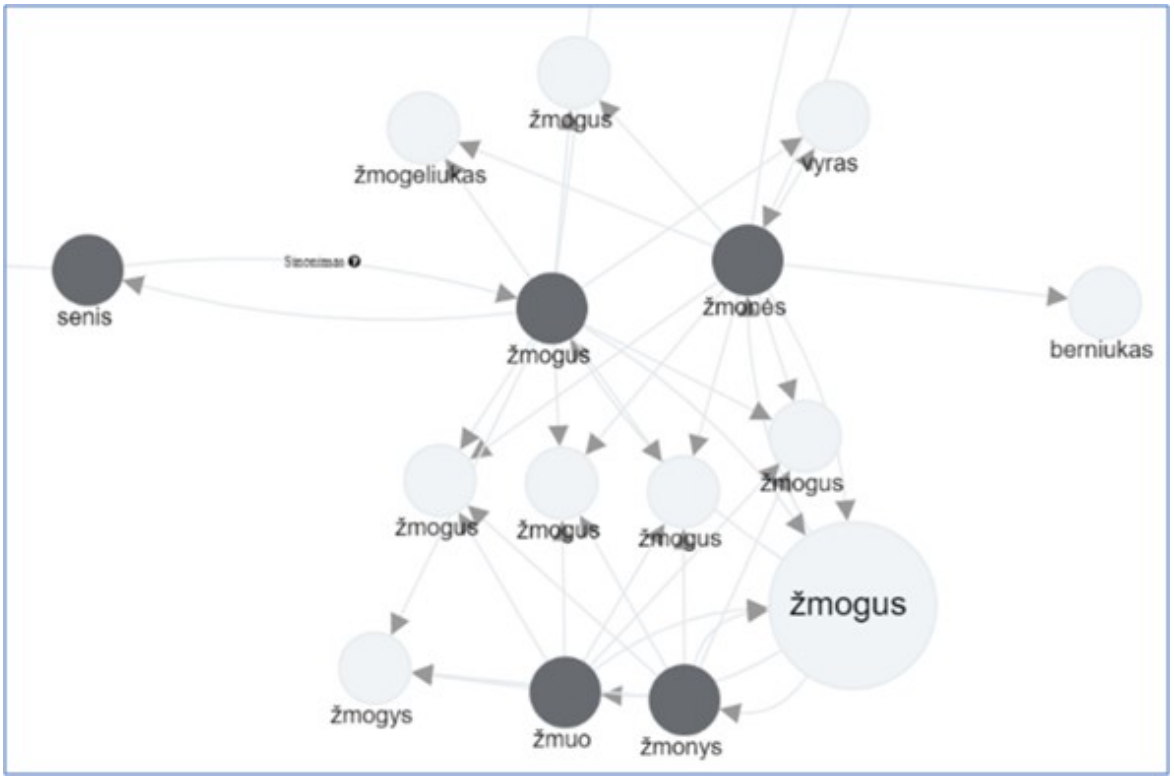
„Lietuvių kalbos žodyno“ leksikografinei mokyklai, kuri ilgainiui suformavo ir lietuvių leksikografijos tradiciją.

„Lietuvių kalbos žodynas“ ir elektroninės paslaugos. Sistemoje „E. kalba“ randamos septynios elektroninės paslaugos: „Žodžių prasmių tinklas“ (susideda iš e. paslaugų „Paieška Žodžių prasmių tinkle“ ir „E. sąvokos“), „E. rinkodara“ (susideda iš e. paslaugų „E. pavadinimas“, „Naudotojų užklausų analizė“ ir „Nuomonių analizė“) bei „E. patarimai“ (susideda iš e. paslaugų „Žodžių darybos vedlys“ ir „Kalbos patarimai“). „Lietuvių kalbos žodyno“ medžiaga naudojama atliekant paiešką „Žodžių prasmių tinkle“ bei kuriant „E. pavadinimą“.

Sistemos „E. kalba“ elektroninės paslaugos

Pasaulyje kuriami žodžių prasmių tinklai vertinami kaip elektroninio žodyno, tezauro ir duomenų bazės derinys, kuris yra universalesnis už popierinį žodyną, nebespėjantį atspindėti greitai kintančios kalbos situacijos. Sistemoje „E. kalba“ esantis lietuvių kalbos **„Žodžių prasmių tinklas“** (https://ekalba.lt/zodziu-prasmiu-tinklas/?p=1) – tai unikalus skaitmeninis išteklius, kuriame suformuoti prasminiai ryšiai tarp žodžių iš „Lietuvių kalbos žodyno“, taip pat iš kitų „E. kalbos“ išteklių: „Bendrinės lietuvių kalbos žodyno“, „Dabartinės lietuvių kalbos žodyno“, „Lietuvių kalbos naujažodžių duomenyno“, „Sinonimų žodyno“, „Antonimų žodyno“ ir „Frazeologijos žodyno“. Lietuvių kalbos žodžių prasmių tinklo 10 000 sinsetų sujungti su Princetono anglų kalbos žodžių tinklo (angl. *Princeton WordNet*) sinsetais (tereikia atsidarius konkretų sinsetą paspausti mygtuką „EN“). Minėtina, kad Princetono universitete sukurtas WordNet’as tapo pamatine struktūra kitų kalbų žodžių tinklams (žr. Princeton WordNet 2023). Sistemoje „E. kalba“ įdiegtas „Žodžių prasmių tinklas“ yra trečioji lietuvių kalbos WordNet’o versija[4]. Kaip ir kitose versijose, šiame žodžių tinkle sinsetai į tinklą sujungti hierarchiniais prasminiais ryšiais: 1) horizontaliaisiais (sinonimų ir antonimų), 2) vertikaliaisiais (hiperonimų ir hiponimų, holonimų ir meronimų), 3) leksiniais (frazeologizmų ir vedinių). Tačiau pirmą kartą į lietuvių kalbos žodžių prasmių tinklą įtrauktos visos kalbos dalys

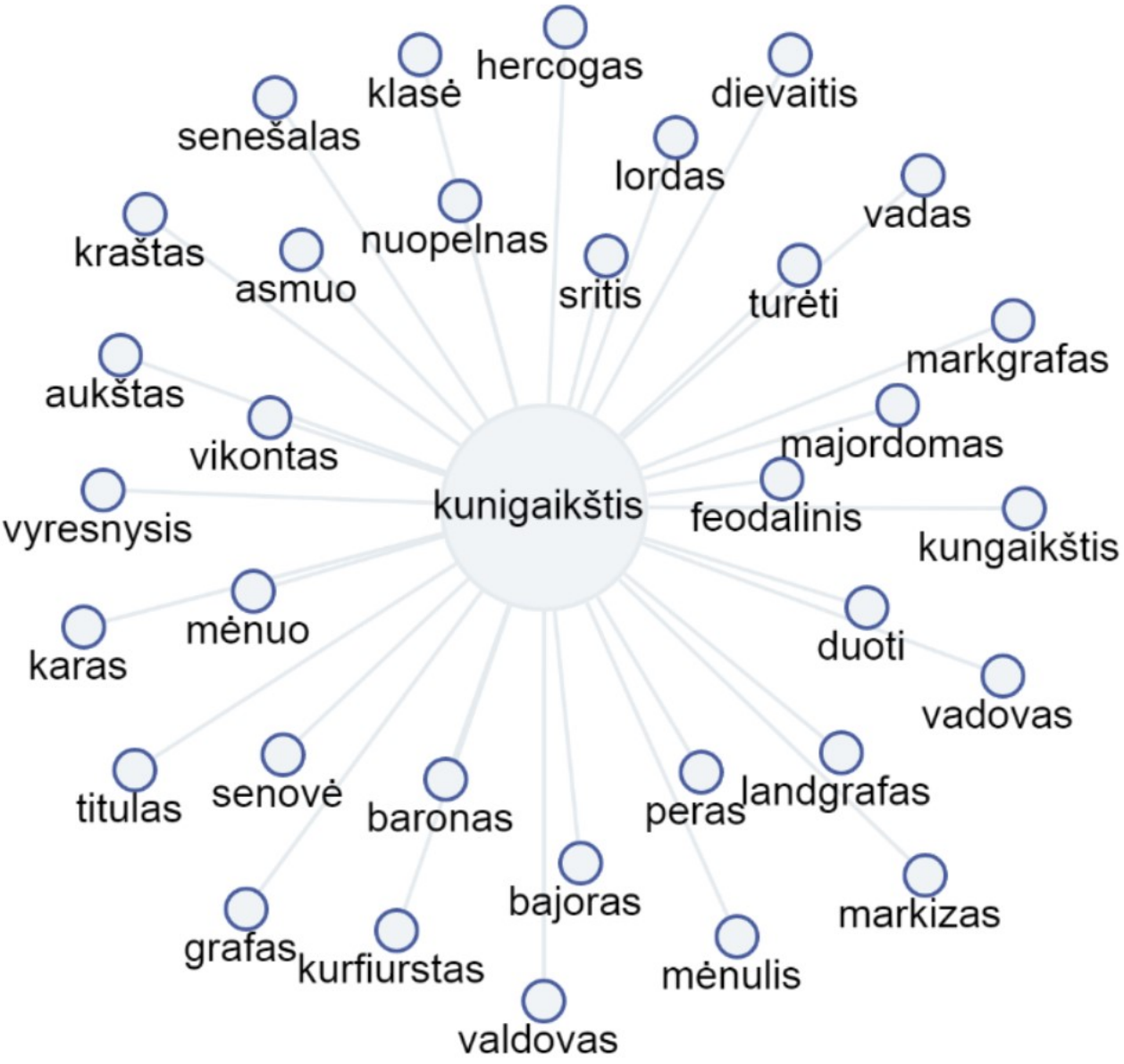
(daiktavardis, veiksmažodis, būdvardis, prieveiksmis, prielinksnis, skaitvardis, įvardis, dalelytė, jungtukas, jaustukas, ištiktukas). „Žodžių prasmių tinkle“ vartotojai gali rasti informacijos, kokiais semantiniais ir leksiniais ryšiais rūpimas žodis susijęs su kitais lietuvių kalbos žodžiais. Informacija pateikiama ir vizualiai – grafe (šis terminas apibūdina abstrakčią struktūrą, aprašančią rinkinį objektų, kurių kai kurios poros susietos prasminiais ryšiais). Grafo pavyzdys pateiktas 5 pav.



Daiktavardžio žmogus („savos tautos ar tikybos asmuo, vietos gyventojas, lietuvis“) grafas „Žodžių prasmių tinkle“ . Balti apskritimai žymi sinsetus, pilki apskritimai – sinsetą sudarančius žodžius.

Dar viena elektroninė paslauga, kurioje naudojamasi „Lietuvių kalbos žodyno“ medžiaga – **„E. pavadinimas“**. Ja naudojantis galima susikurti lietuvišką įmonės, gaminio, paslaugos ar kitą rūpimą pavadinimą: tereikia į paieškos laukelį įrašyti reikšminį žodį (žodžius), kuris apibūdina rūpimą veiklą, gaminį, paslaugą, sritį ar kt., ir sistema pasiūlys įprastų, retesnių, naujų ar netikėtų žodžių ar žodžių junginių, įvairiais prasminiais ryšiais susijusių su pateikta užklausa. Užklausos rezultatai kuriami remiantis šių „E. kalbos“ išteklių duomenimis: 1) „Žodžių prasmių tinklu“ (ŽPT), 2) trijų ontologijų (ekonomikos, kompiuterinės technikos, žmogaus anatomijos) duomenimis, 3) „Bendrinės lietuvių kalbos žodynu“ (BLKŽ), 4) „Dabartinės lietuvių kalbos žodynu“ (DLKŽ), 5) „Lietuvių kalbos žodynu“ (LKŽ),

6) „Sinonimų žodynu“ (SinŽ), 7) „Frazeologijos žodynu“ (FrzŽ), 8) „Sisteminiu lietuvių kalbos žodynu“ (SisŽ), 9) „Naujažodžių duomenynu“ (NdŽ). Užklausos rezultatai pateikiami vizualiai, vadinamojoje „nuotaikų lentoje“ (žr. 6 pav.).

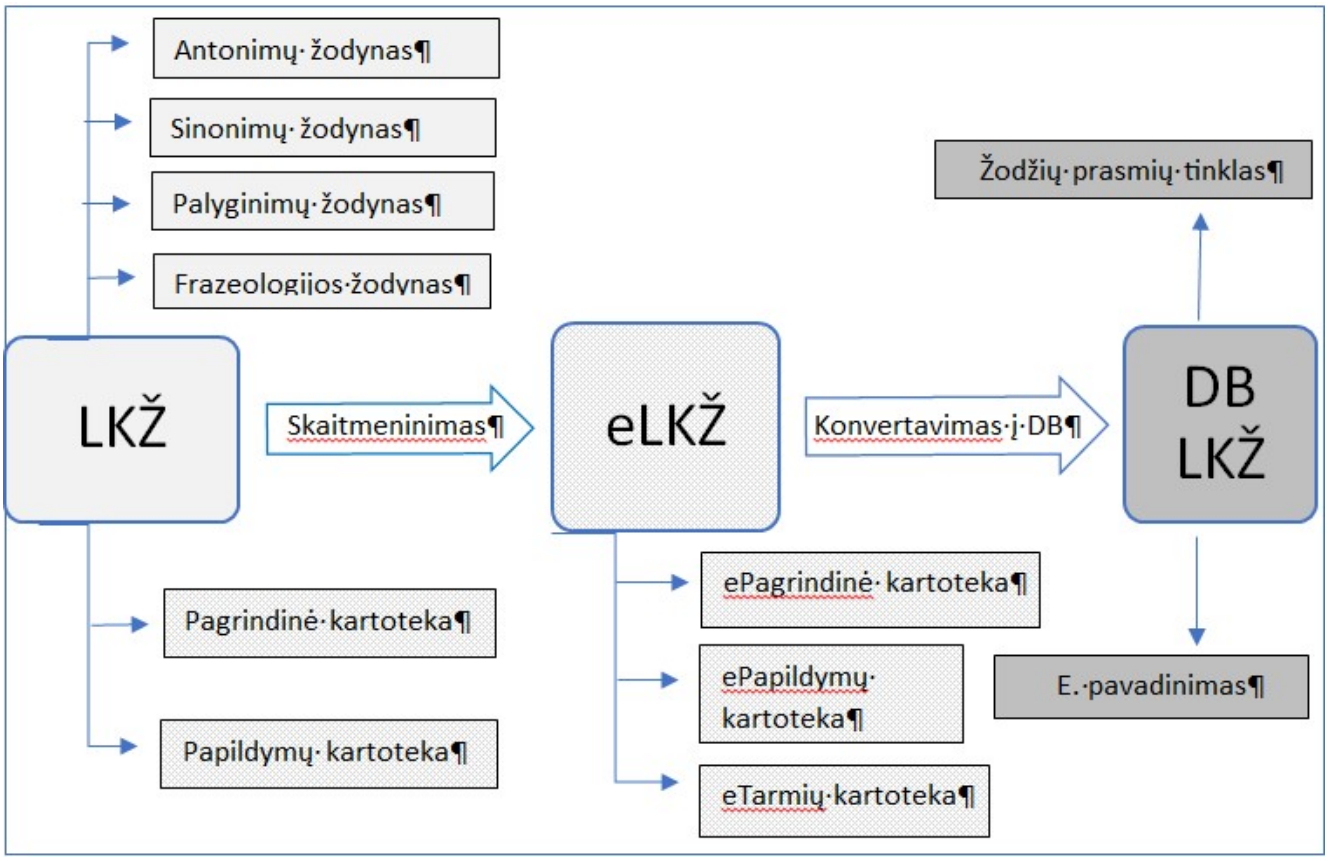


Paslauga „E. pavadinimas“: daiktavardžio kunigaikštis prasminiai ryšiai

Susikūrusiems įmonės ar prekės ženklo pavadinimą pateikiamos nuorodos į Registrų centro ir Lietuvos Respublikos valstybinio patentų biuro svetaines, kuriose galima pasitikrinti, ar toks pavadinimas dar nėra registruotas. Taip pat galima rasti nuorodas į Simbolinių pavadinimų darymo taisykles Valstybinės lietuvių kalbos komisijos svetainėje.

* * *

Apžvelgę lietuvių kalbos išteklių informacinės sistemos „E. kalba“ elektroninius produktus ir paslaugas galime konstatuoti, kad juos kuriant nebūtų buvę įmanoma panaudoti „Lietuvių kalbos žodyno“ duomenų masyvo, jei LKŽ nebūtų patyręs kelių virsmo etapų (žr. 7 pav.). XXI a. pirmaisiais dešimtmečiais įvyko dvi „Lietuvių kalbos žodyno“ transformacijos: 1) LKŽ skaitmeninimas, kuris pagerino lietuvių kalbos išteklių prieinamumą; ir 2) LKŽ duomenų bazės sukūrimas, kuris leido LKŽ medžiagos pagrindu kurti išvestinius el. produktus („Žodžių prasmių tinklą“, „E. pavadinimą“ ir kt.).



„Lietuvių kalbos žodyno“ transformacijos skaitmeninėje eroje

Lietuvių kalbos išteklių informacinės sistemos „E. kalba“ sukūrimas naudingas lietuvių kalbos vartotojams Lietuvoje ir svetur, daro teigiamą poveikį pačiai lietuvių kalbai: padeda užtikrinti visavertį lietuvių kalbos vartojimą skaitmeninėje terpėje, įtvirtina lietuvių kalbos statusą informacinėje visuomenėje, gausina skaitmeninius kalbos išteklius, padeda plėtoti kalbos technologijas, kurti viešąsias paslaugas, stiprina Lietuvos visuomenės ir išeivijos ryšius, mažina lietuviškai kalbančios bendruomenės atskirtį globalioje žinių visuomenėje, padeda kitakalbiams lengviau integruotis į Lietuvos visuomenę. Visa tai atitinka Lietuvių kalbos plėtros skaitmeninėje terpėje ir kalbos technologijų pažangos 2021–2027

metų gairėse (žr. Gairės 2021–2027) numatytus uždavinius.

Tolesnė sistemos „E. kalba“ plėtra galėtų būti siejama ne tik su naujų išteklių integracija bei inovatyvių el. produktų ar paslaugų kūrimu, bet ir su integruotų išteklių pildymu, pvz., milžiniškų LKŽ kartotekų suskaitmeninimas ir įtraukimas į „E. kalbą“ reikšmingai palengvintų visuomenės prieigą prie lietuvių kalbos leksikografinių išteklių.

Šaltiniai ir literatūra

AntŽ – Ermanytė I. *Antonimų žodynas.* Vilnius: Lietuvių kalbos instituto leidykla, 2003. Prieiga internete: https://ekalba.lt/antonimu-zodynas/.

BgEK – *Kazimiero Būgos etimologinio žodyno kartoteka.* Vilnius: Lietuvių kalbos institutas, 2015. doi.org/10.35321/BugosEtimologinioZodynoKartoteka/ Prieiga internete: https://ekalba.lt.

BLKŽ – *Bendrinės lietuvių kalbos žodynas.* Vilnius: Lietuvių kalbos institutas, 2013–2024. Prieiga internete: https://ekalba.lt/bendrines-lietuviu-kalbos-zodynas/.

DLKŽ – *Dabartinės lietuvių kalbos žodynas.* Vilnius: Lietuvių kalbos institutas, 2021. Prieiga internete: https://ekalba.lt/dabartines-lietuviu-kalbos-zodynas/.

2024. kalba – *Lietuvių kalbos išteklių informacinė sistema „E. kalba“.* Vilnius: Lietuvių kalbos institutas, 2015–2024. Prieiga internete: https://ekalba.lt/.
2025. pavadinimas – *Elektroninė paslauga „E. pavadinimas“.* Vilnius: Lietuvių kalbos institutas, 2021. doi.org/10.35321/e-pavadinimas. Prieiga internete: https://ekalba.lt/.

FrzŽ – Ermanytė I., Kažukauskaitė O., Naktinienė G., Paulauskas J. (red.), Šimėnaitė Z., Vilutytė A., *Frazeologijos žodynas.* Vilnius: Lietuvių kalbos institutas, 2001. Prieiga internete: https://ekalba.lt/frazeologijos-zodynas/.

Gairės 2021–2027: *Lietuvių kalbos plėtros skaitmeninėje terpėje ir kalbos technologijų pažangos 2021–2027 metų gairės.* Prieiga internete: https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/911407f20ee911ebbedbd456d2fb030d.

Garabik R., Pileckytė I. 2013: From Multilingual Dictionary to Lithuanian WordNet. – *Natural language Processing, Corpus Linguistics, E-Learning: Seventh International Conference Bratislava* (Slovko 2013). Slovakia: Bratislava, 74–80.

LKŽ – *Lietuvių kalbos žodynas*, I–XX, 1941–2002. Vilnius: Lietuvių kalbos institutas, 2018. Prieiga internete: www.lkz.lt; https://ekalba.lt/lietuviu-kalbos-zodynas/.

LKŽ TK – *„Lietuvių kalbos žodyno“ Tarmių kartoteka.* Vilnius: Lietuvių kalbos institutas, 2015. doi.org/10.35321/lkz-tarmiu-kartoteka. Prieiga internete: https://ekalba.lt.

LKŽ PgK – *„Lietuvių kalbos žodyno“ Pagrindinė kartoteka.* Vilnius: Lietuvių kalbos institutas, 2015. doi.org/10.35321/lkz-pagrindine-kartoteka. Prieiga internete: https://ekalba.lt.

LKŽ PpK – *„Lietuvių kalbos žodyno“ Papildymų kartoteka.* Vilnius: Lietuvių kalbos institutas, 2015. doi.org/10.35321/lkz-papildymu-kartoteka. Prieiga internete: https://ekalba.lt.

MULŽ – *Mažasis ukrainiečių–lietuvių kalbų žodynas / Малий українсько–литовський словник* / Rengė: Aurelija Gritėnienė, Svitlana Hrycenko, Ihor Koroliov, Oksana Nika, Zinaida Pacholok. – Vilnius: Lietuvių kalbos institutas, 2022. Prieiga internete: https://ekalba.lt/mazasis-ukrainieciu-lietuviu-kalbu-zodynas/.

NžD – *Lietuvių kalbos naujažodžių duomenynas.* Vilnius: Lietuvių kalbos institutas, 2015. Prieiga internete: https://ekalba.lt/naujazodziai/.

PlgŽ – Vosylytė K. *Palyginimų žodynas.* Vilnius: Lietuvių kalbos institutas, 2014. Prieiga internete: https://ekalba.lt/palyginimu-zodynas/.

Princeton WordNet – *A Lexical Database for English „WordNet“.* Princeton, New Jersey: Princeton university, 2023. Prieiga internete: https://wordnet.princeton.edu.

SinŽ – Lyberis A. *Sinonimų žodynas.* Vilnius: Lietuvių kalbos institutas, 2002. Prieiga internete: https://ekalba.lt/sinonimu-zodynas/.

SistŽ – Paulauskas J. *Sisteminis lietuvių kalbos žodynas.* Vilnius: Mokslas, 1987. Prieiga internete: https://ekalba.lt/sisteminis-lietuviu-kalbos-zodynas/.

Urbanavičienė Jolita, Lietuvių kalbos išteklių informacinė sistema „E. kalba“ – visuomenei atviri žodynai ir kartotekos. – *Pasaulio lietuvis,* 2022. Prieiga internete: https://pasauliolietuvis.lt/lietuviu-kalbos-istekliu-informacine-sistema-e-kalba-visuomenei-atviri-zodynai-ir-kartotekos/.

Vitkutė-Adžgauskienė D., Dainauskas J. J., Amilevičius D., Utka A. 2015: Lietuvių kalbos žodžių tinklas – LitWordNet. – *Darbai ir dienos* 64, 101–114.

ŽPT – *Žodžių prasmių tinklas.* Vilnius: Lietuvių kalbos institutas, 2021. doi.org/10.35321/zodziu-prasmiu-tinklas. Prieiga internete: https://ekalba.lt.

Straipsnių ciklas „Lietuvių kalbos naujovės ir įdomybės“. Straipsnį rėmė Valstybinė lietuvių kalbos komisija.

Jei norėtumėte publikuoti visą straipsnį ar jo dalį, prašom nurodyti informacijos šaltinį ir autorius.

[1] Buvo vykdomas projektas „IRT sprendimų bei turinio, padedančių išsaugoti lietuvių kalbą viešojoje erdvėje, kūrimas bei galimybių jais naudotis sudarymas“ (kodas NR. VP2-3.1-IVPK-12-K-01-005), finansuotas pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinė visuomenė visiems“ įgyvendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“.

[2] Projektas „Lietuvių kalbos išteklių informacinės sistemos plėtra (E. kalba)“ (Nr. 02.3.1-CVPA-V-527-01-0006), finansuotas iš Europos Sąjungos struktūrinių fondų lėšų.

[3] Naujausias išteklius, įtrauktas į sistemą 2024 m., yra „Mažasis ukrainiečių–lietuvių kalbų žodynas / Малий українсько–литовський словник“ (MULŽ), kurį parengė Aurelija Gritėnienė, Svitlana Hrycenko, Ihor Koroliov, Oksana Nika, Zinaida Pacholok. Prieiga internete: https://ekalba.lt/mazasis-ukrainieciu-lietuviu-kalbu-zodynas/.

[4] Pirmąją WordNet’o versiją lietuvių kalbai 2013 m. sukūrė Slovakijos mokslų

akademijos Ludovito Sturo lingvistikos institutas (Garabik, Pileckytė 2013). Antroji LitWordNet’o versija (plėtojanti slovakų variantą) sukurta 2015 m. Vytauto Didžiojo universitete (Vitkutė-Adžgauskienė, Dainauskas, Amilevičius, Utka 2015, 103).