

UDK (811.172+378.147):004.7
DOI: <https://doi.org/10.17721/1728-2659.2023.34.05>

Anželika Gaidienė, humanitarinių mokslų daktarė
ORCID: 0000-0001-8775-788X
e-mail: anzelika.gaidiene@iki.lt
Lietuvių kalbos institutas, Vilnius, Lietuva

LIETUVIŲ KALBOS IŠTEKLIŲ INFORMACINĖ SISTEMA „E. KALBA“: NUO KŪRIMO ISTORIJS, PAIEŠKOS GALIMYBIŲ IKI PLĖTROS PERSPEKTYVŲ

Šiame straipsnyje pristatoma Lietuvių kalbos institute kurta Lietuvių kalbos išteklių informacinė sistema „E. kalba“ – supažindinama su „E. kalbos“ kūrimo istorija, aptariami sistemos privalumai, parodoma, kaip atlikti greitąją paiešką visuose arba pasirinktuose kalbos ištekliuose, kaip ieškoti dominančios informacijos naudojant pakaitos simbolius, glaustai pristatoma sistemos naudojimo statistika, galimos sistemos plėtros perspektyvos.

Nuo 2021 m. sistema nemokamai prieinama plačiajai visuomenei ir specialistams adresu <https://ekalba.lt>. „E. kalboje“ galima rasti ir patogiai naršyti net 39 kalbos ištekliuose, kurie suskirstyti į tris grupes: 1) e. žodynus („Bendrinės lietuvių kalbos žodynas“, „Dabartinės lietuvių kalbos žodynas“, „Lietuvių kalbos žodynas“, „Lietuvių kalbos naujažodžių duomenynas“, „Sinonimų žodynas“, „Lietuvių–anglų kalbų žodynas“, „Vokiečių–lietuvių kalbų žodynas“ ir kt.), 2) e. kartotekas („Lietuvių kalbos žodyno“ Pagrindinė ir Papildymų kartotekos, „Mįslių kartoteka“ ir kt.) ir duomenų bazes („Pavardžių duomenų bazė“ ir kt.) bei 3) e. paslaugas („Paieška Žodžių prasmų tinkle“, „E. sąvokos“, „E. pavadinimas“, „Nuomonių analizė“, „Žodžių darybos vedlys“, „Kalbos patarimai“, „Mokomieji kalbos žaidimai“ ir kt.).

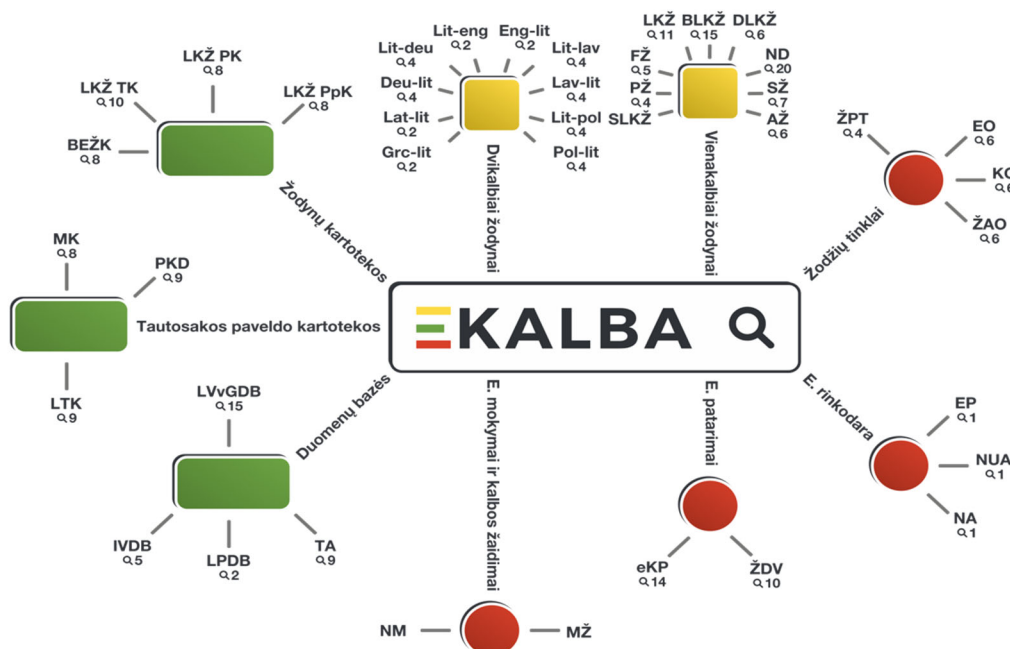
„E. kalba“ yra ne tik kalbinių išteklių lobynas, bet ir daugiakriteris paieškos įrankis – tam, kad būtų galima patogiai ieškoti dominančios informacijos, sukurta efektyvi ir įvairi paieška (greitoji, detalioji paieškos; paieška visuose ištekliuose ar kiekviename ištekliuje atskirai ir pan.). „E. kalboje“ veikia patogi paieškos rezultatų peržiūra – rezultatus galima parsisiųsti ir išsisaugoti PDF, XML formatais, rezultatais galima dalintis socialiniame tinkle „Facebook“. Sistema pritaikyta naudoti ir mobiliuosiuose įrenginiuose.

Esminiai žodžiai: Lietuvių kalbos išteklių informacinė sistema „E. kalba“, skaitmeniniai ištekliai, kalbos technologijos, paieškos galimybės.

Įvadas. Nuo 2021 m. Lietuvių kalbos išteklių informacinė sistema „E. kalba“ (toliau – „E. kalba“) nemokamai prieinama plačiajai visuomenei adresu <https://ekalba.lt>. Ši sistema kurta Lietuvių kalbos institute (plačiau žr. <http://iki.lt>) bendradarbiaujant su pažangiųjų informacinių technologijų specialistais (UAB „Proit“, UAB „Tilde“, UAB „Netcode“).

Pagrindinis šio straipsnio tikslas – pristatyti Lietuvių kalbos išteklių informacinę sistemą „E. kalba“, parodyti jos naudojimosi galimybes.

„E. kalba“ yra skaitmeninių lietuvių kalbos išteklių, susijusių tarpusavyje, lobynas. „E. kalboje“ galima rasti ir patogiai naršyti net 39 kalbos ištekliuose, kurie suskirstyti į tris grupes: e. žodynus – vienakalbius ir dvikalbius (žr. 1 pav. geltonus kvadratus), e. kartotekas ir duomenų bazes – žodynų kartotekas, tautosakos paveldo kartotekas, duomenų bazes (žr. 1 pav. žalius stačiakampius) ir e. paslaugas – žodžių tinklus, e. rinkodarą, e. patarimus, e. mokymus ir kalbos žaidimus (žr. 1 pav. raudonus apskritimus).



1 pav. Lietuvių kalbos išteklių informacinės sistemos „E. kalba“ schema
(© Anželika Gaidienė, Marius Glebus, Aurelija Tamulionienė)

Vienas iš „E. kalbos“ privalumų yra ne tik suskaitmenintų kalbos išteklių gausa, bet ir tuose ištekliuose fiksuojamas duomenų kiekis, plg.: vienakalbiai žodynai sudaro daugiau nei 500000 antraštinų žodžių; el. kartotekos – daugiau nei 530000 antraštinų žodžių; žodžių tinklas – 24000 sinsetų ir pan.

„E. kalba“ yra ne tik kalbinių išteklių lobynas, bet ir daugiakriteris paieškos įrankis. Tam, kad būtų galima patogiai ieškoti dominančios informacijos, „E. kalboje“ veikia efektyvi ir įvairi paieška, t. y. pagal įvairius kriterijus galima patogiai ir greitai ieškoti informacijos tiek visuose ištekliuose,

тік konkrečiame ištekliuje atskirai; galima informacijos ieškoti naudojant simbolius (*, ?). Taip pat „E. kalboje“ veikia patogi paieškos rezultatų peržiūra – rezultatus galima parsisiųsti ir išsisaugoti PDF, XML formatais, rezultatais galima dalintis socialiniame tinkle *Facebook*. Sistema pritaikyta naudoti ir mobiliuosiuose įrenginiuose. Sistemos kūrėjai – kalbininkai – gali patogiai dirbti vidiniame „E. kalbos“ portale – kelti naujus duomenis, taisyti esančius ir pan.

Apibendrinant galima teigti, kad „E. kalba“ atitinka tokius pagrindinius, naudotojams aktualius, kriterijus (dar žr. [4]):

- išsamumas – daugybė įvairių šaltinių ir kalbinių duomenų vienoje vietoje;
- atnaujinimas – kai kurie ištekliai nuolat pildomi, taisomi;
- patikimumas – sistema kurta mokslininkų ir kalbos technologijų specialistų;
- patogumas – paieška veikia vieno langelio principu, galima detalioji paieška;
- naujoviškumas – pažangios kalbos paslaugos, pavyzdžiui, „Nuomonių analizatorius“ naudoja pažangiausias technologijas, konkrečiai – giliojo neuroninio tinklo BiLSTM (angl. *Bidirectional Long Short-Term Memory*) architektūrą su BERT (angl. *Bidirectional Encoder Representations from Transformer*) žodžių vektorizacija ir kt.

Taigi „E. kalba“ kuria pridėdamąją mokslinę ir praktinę vertę – tarp gausybės kalbinių duomenų mokslininkai ras tyrimams naudingos medžiagos, o plačioji visuomenė galės naudotis sistema mokymo(si) procese, versle ar tiesiog kasdieniame gyvenime.

„E. kalba“ plačiai visuomenei ir mokslo bendruomenei pristatyta ne kartą. Apie „E. kalbą“ rašyta dr. Jurgitos Jaroslavienės ir dr. Ritos Miliūnaitės straipsnyje „Beribis lietuvių kalbos pasaulis skaitmeninių išteklių sistemoje „E. kalba“ [4]. „E. kalba“ pristatyta ir tarptautiniu mastu – dr. Jurgitos Jaroslavienės ir dr. Albinos Aukšoriūtės straipsnyje „Innovations and Challenges in the Digital Transformation of the Lithuanian Language Industry“ [3, p. 155–167]. Apie lietuvių kalbos naujažodžių duomenyną plačiai rašyta dr. Ritos Miliūnaitės straipsnyje „Naujažodžių pateikimas ir paieškos galimybės lietuvių kalbos naujažodžių duomenyne“ [6, p. 1–21]. Atskirai apie leksikografinius tarmių duomenis, randamus „E. kalboje“, rašyta dr. Jolitos Urbanavičienės straipsnyje „Leksikografiniai tarmių duomenys lietuvių kalbos išteklių informacinėje sistemoje „E. kalba“ [8, p. 109–121]. Daug naudingos informacijos apie Žodžių darybos vedlį pateikta dr. Daivos Murmulaitytės straipsnyje „Žodžių darybos vedlys plačiai visuomenei ir specialistams“ [7, p. 8–24].

Lietuvių kalbos išteklių informacinės sistemos „E. kalba“ kūrimo istorija. „E. kalba“ (anksčiau vadinta LKIS¹ pradėta kurti lietuvių kalbos institute 2012 m. pagal projektą „IRT sprendimų bei turinio, padedančių išsaugoti lietuvių kalbą viešojoje erdvėje, kūrimas bei galimybių naudotis jais sudarymas“. Šio projekto metu (2012–2015 m.):

- suskaitmeninti šeši vienakalbiai lietuvių kalbos žodynai: „Dabartinės lietuvių kalbos žodynas“ (vyriausiasis redaktorius dr. Stasys Keinys, 8-tas leidimas², 2021 m.), „Sinonimų žodynas“ (autorius dr. Antanas Lyberis, 2002 m.), „Antonimų žodynas“ (autorė dr. Irena Ermanytė, 2003 m.), „Frazeologijos žodynas“ (vyriausiasis redaktorius dr. Jonas Paulauskas, 2001 m.), „Palyginimų žodynas“ (autorė dr. Klementina Vosyltė, 2014 m.) ir „Sisteminis lietuvių kalbos žodynas“ (autorius dr. Jonas Paulauskas, 1987 m.);

- suskaitmeninta dešimt dvikalbių žodynų: lietuvių–anglų, anglų–lietuvių; lietuvių–latvių, latvių–lietuvių, lietuvių–lenkų, lenkų–lietuvių, lietuvių–vokiečių, vokiečių–lietuvių, lotynų–lietuvių, senosios graikų–lietuvių;

- suskaitmenintos (iš dalies) trys „Lietuvių kalbos žodyno“ kartotekos: „Pagrindinė kartoteka“ (skelbiama dalies A ir G raidžių kortelių vaizdai ir jų metaduomenys), „Papildymų kartoteka“ (skelbiama A–N, P–S, T–Z raidžių kortelių vaizdai ir jų metaduomenys) ir „Tarmių kartoteka“ (skelbiami naujausi leksikografiniai tarmių duomenys);

- suskaitmenintos trys tautosakos paveldo kartotekos: „Pokario partizanų dainų kartoteka“ (suskaitmeninta daugiau kaip 4000 partizanų dainų), „Mįslių kartoteka“ (suskaitmeninta apie 80000 mįslių) ir „Liaudies tikėjimų kartoteka“ (suskaitmeninta daugiau kaip 8000 tekstų);

- integruotos keturios elektroninės lietuvių kalbos išteklių duomenų bazės: „Lietuvos vietovardžių geoinformacinė duomenų bazė“, „Istorinių vietovardžių duomenų bazė“, „Pavardžių duomenų bazė“ ir „Tarmių archyvas“;

- atliktas lietuviško šrifto „Palemonas“ patobulinimas;

- sukurta vieša išteklių prieiga visuomenei.

Projektas vykdytas kartu su partneriais – Lietuvių literatūros ir tautosakos institutu, Vilniaus universitetu ir tuometiu Lietuvos edukologijos universitetu.

„E. kalbos“ plėtra vykdyta 2018–2021 m. pagal projektą „Lietuvių kalbos išteklių informacinės sistemos plėtra (E. KALBA)“. Įgyvendinus šį projektą:

- sukurtos septynios naujos visuomenei aktualios elektroninės paslaugos: žodžių tinklai („Paieška Žodžių prasmių tinkle“, „E. sąvokos“), e. rinkodara („E. pavadinimas“, „Naudotojų užklausų analizė“, „Nuomonių analizė“) ir e. patarimai („Žodžių darybos vedlys“, „Kalbos patarimai“);

- integruoti trys vienakalbiai žodynai: „Bendrinės lietuvių kalbos žodynas“ (vyriausioji redaktorė dr. Danutė Liutkevičienė; nuolat pildomas išteklis), „Lietuvių kalbos naujažodžių duomenynas“ (sudarytoja dr. Rita Miliūnaitė; nuolat pildomas išteklis) ir „Lietuvių kalbos žodynas“ (20 tomų, rašytas 1902–2002 m.);

- sukurti įrankiai naujai integruotų žodynų ir e. paslaugų duomenims tvarkyti;

- atnaujinta paieškos sistema;

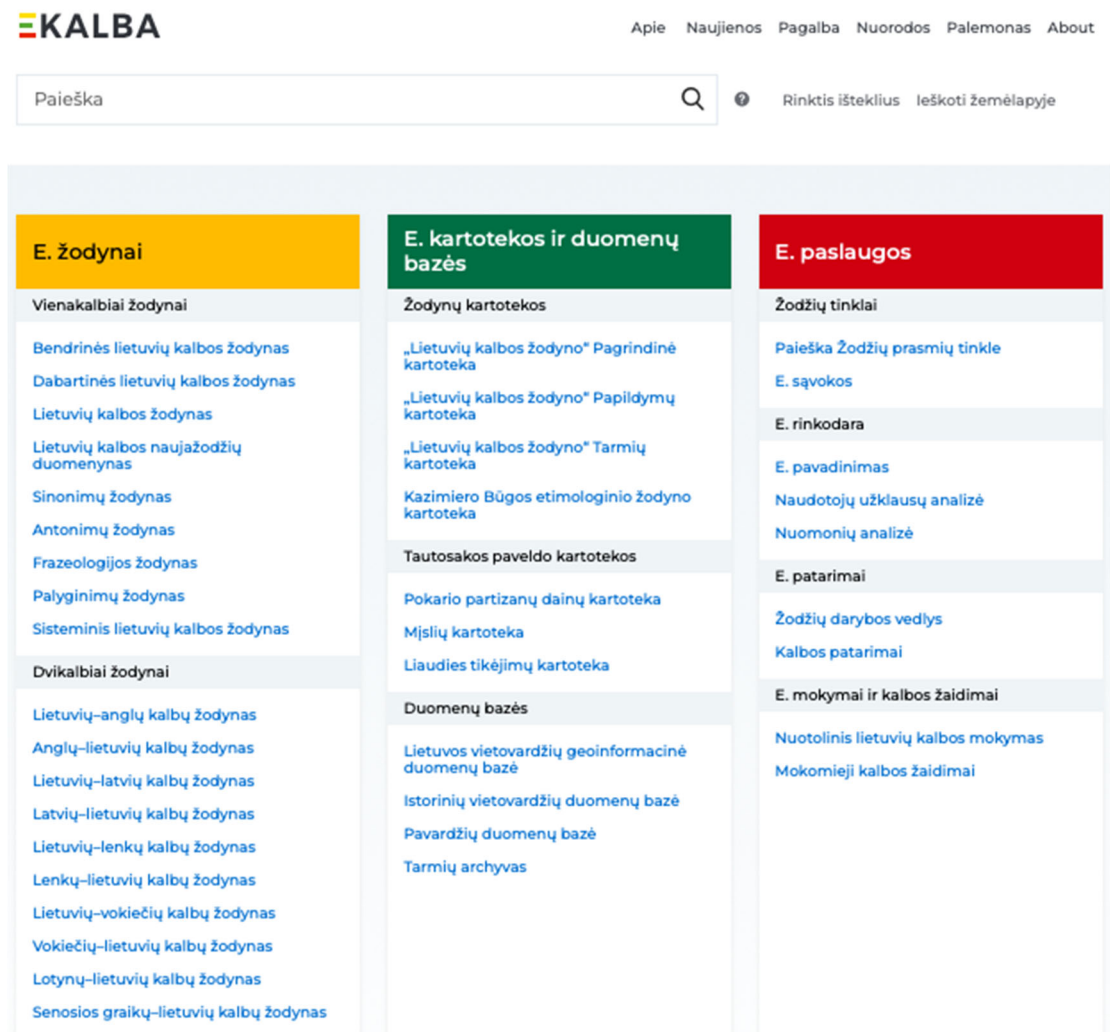
- sukurta nauja vieša išteklių prieiga visuomenei – <https://ekalba.lt> (žr. 2 pav.);

- svetainė pritaikyta mobiliesiems įrenginiams.

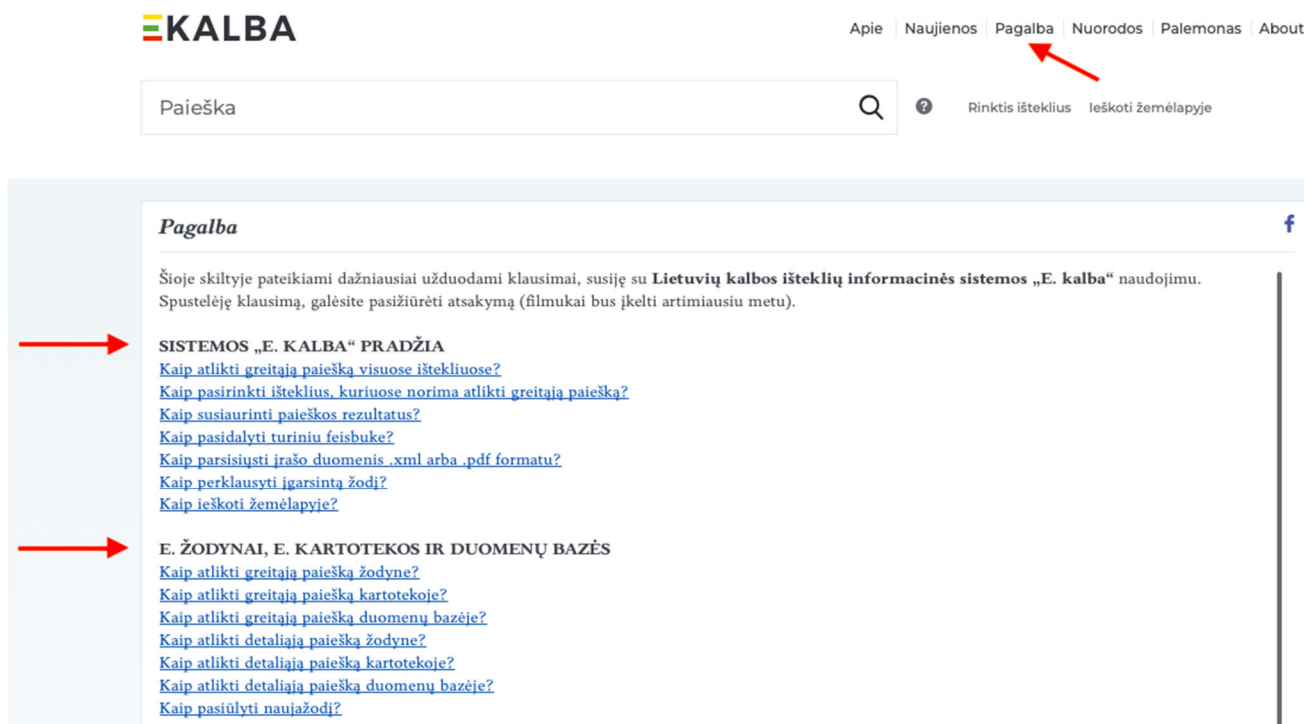
Viršutinėje „E. kalbos“ meniu juostoje (žr. 2 pav.) galima rasti svetainės lankytojams naudingos informacijos. Paspaudus *Apie* pateikiama informacija *Apie svetainę, Apie projektą, Naudojimosi taisyklės. Naujienų* skiltyje skelbiama naujausia informacija, susijusi su išteklių atnaujinimu ar pan. *Pagalbos* skiltyje (žr. 3 pav.) pateikiami dažniausiai užduodami klausimai, susiję su „E. kalbos“ naudojimosi galimybėmis, pavyzdžiui, spustelėjus klausimą *Kaip atlikti greitąją paiešką visuose ištekliuose?*, atsiveria *Youtube* kanalo langas, kuriame rodomas mokomas filmukas (žr. 4 pav.). Skiltyje *Nuorodos* galima rasti nuorodų į kitus skaitmeninius lietuvių kalbos išteklius. Paspaudus *Palemonas*, galima parsisiųsti originalų lietuvišką šriftą, skirtą plačiai visuomenei ir lituanistikos mokslo reikmėms. Skiltyje *About* pateikiama informacija apie sistemą anglų kalba.

¹ Lietuvių kalbos išteklių informacinė sistema.

² 2012–2015 m. LKIS versijoje buvo suskaitmenintas ir prieinamas 7-tas žodyno leidimas, o 2022 m. į „E. kalbą“ integruotas 8-tas žodyno leidimas.



2 pav. Lietuvių kalbos išteklių informacinė sistema „E. kalba“: pradžios puslapis



3 pav. Meniu punkto *Pagalba* fragmentas



E KALBA

Kaip atlikti greitąją paiešką visuose ištekliuose?



1. Kaip atlikti greitąją paiešką ištekliuose sistemoje E. KALBA?



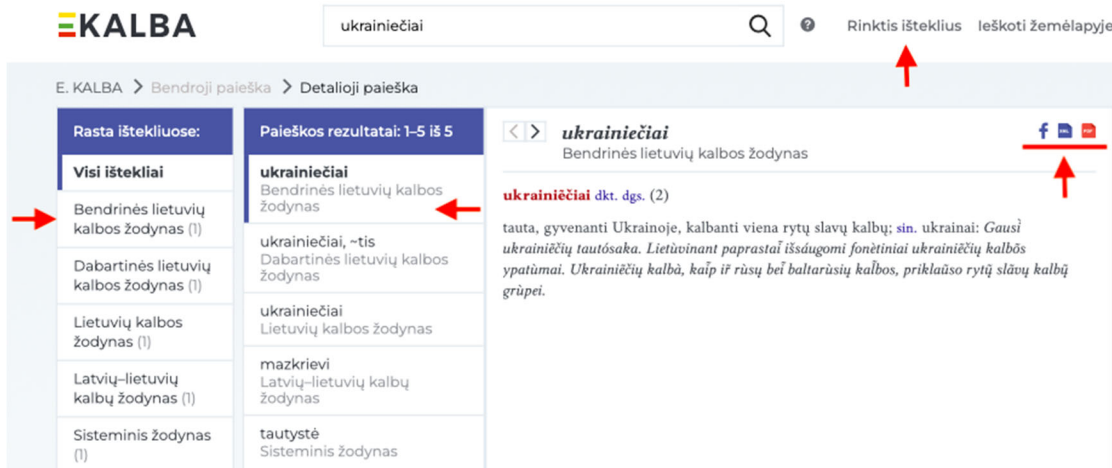
4 pav. Mokomasis filmukas *Kaip atlikti greitąją paiešką visuose ištekliuose?* (Youtube kanalas)

Paieška visuose ištekliuose arba greitoji paieška.

Paieška visuose „E. kalbos“ ištekliuose arba greitoji paieška veikia vieno langelio principu, t. y. į paieškos langelį įrašius ieškomą žodį (ar frazę), sistema pateikia rezultatus. Pavyzdžiui, įrašius žodį *ukrainiečiai* (žr. 5 pav.), pateikiamas išteklių, kuriuose rastas paieškos žodis, sąrašas (*Rasta ištekliuose*) ir penki paieškos rezultatai (*Paieškos rezultatai*). Paspaudus ant konkretaus išteklio, pavyzdžiui, „Bendrinės lietuvių kalbos žodyno“, arba iš karto ant konkretaus paieškos rezultato, pavyzdžiui, *ukrainiečiai* („Bendrinės lietuvių kalbos žodynas“), pateikiamas dominančio žodžio leksikografinis straipsnis (žr. 5 pav.). Spaudžiant viršuje esančias rodykles pirmyn (>) ir atgal (<), galima naršyti paieškos rezultatus. Dominantį leksikografinį

straipsnį galima parsisiųsti XML, PDF formatais arba „E. kalbos“ įrašų dalintis socialiniame tinkle Facebook (žr. 5 pav. viršutinį dešinę kampaną).

Svarbu atkreipti dėmesį, kad „E. kalboje“ ištekliai surikiuoti prioritetine tvarka – pirmiausia pateikiami naujausi aiškinamieji norminamieji žodynai – „Bendrinės lietuvių kalbos žodynas“ (atspindintis dabartinę bendrinės lietuvių kalbos leksiką; nuolat pildomas, atnaujinamas) ir „Dabartinės lietuvių kalbos žodynas“ (atnaujintas 2021 m.), toliau „Lietuvių kalbos žodynas“ (apimantis XVI–XX a. lietuvių kalbos leksiką), „Lietuvių kalbos naujažodžių duomenynas“ (apimantis nuo XX a. pab. viešąją vartoseną; nuolat pildomas, atnaujinamas) ir t. t. Taigi ir greitosios paieškos rezultatai rikiuojami prioritetiškai.



5 pav. Greitoji paieška: žodžio *ukrainiečiai* paieškos rezultatai

„E. kalboje“ galima atlikti greitąją paiešką ne tik visuose ištekliuose vienu metu, bet ir pasirinktuose ištekliuose. Norint ieškoti informacijos konkrečiuose ištekliuose, reikia paspausti *Rinktis išteklius* (žr. 5 pav.). Atsivėrusiame lange įrašyti ieškomą žodį ar frazę, varnele pažymėti dominančius išteklius ir atlikti paiešką (žr. 6 pav.).

„E. kalboje“ taip pat galima naršyti kiekviename ištekliuje atskirai. Pavyzdžiui, norint naršyti „Bendrinės lietuvių kalbos žodyne“, reikia paspausti jo pavadinimą (žr. 7 pav.).

Paspaudus „Bendrinės lietuvių kalbos žodyno“ pavadinimą patenkama į žodyno pasvėtainę (žr. 8 pav.), kurioje galima atlikti detaliąją paiešką pagal įvairius kriterijus, peržiūrėti žodžių sąrašą ir kt.

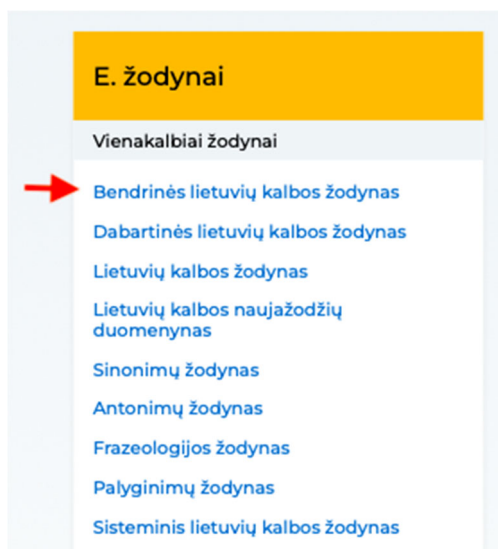
Rinktis išteklius

🔍 ?

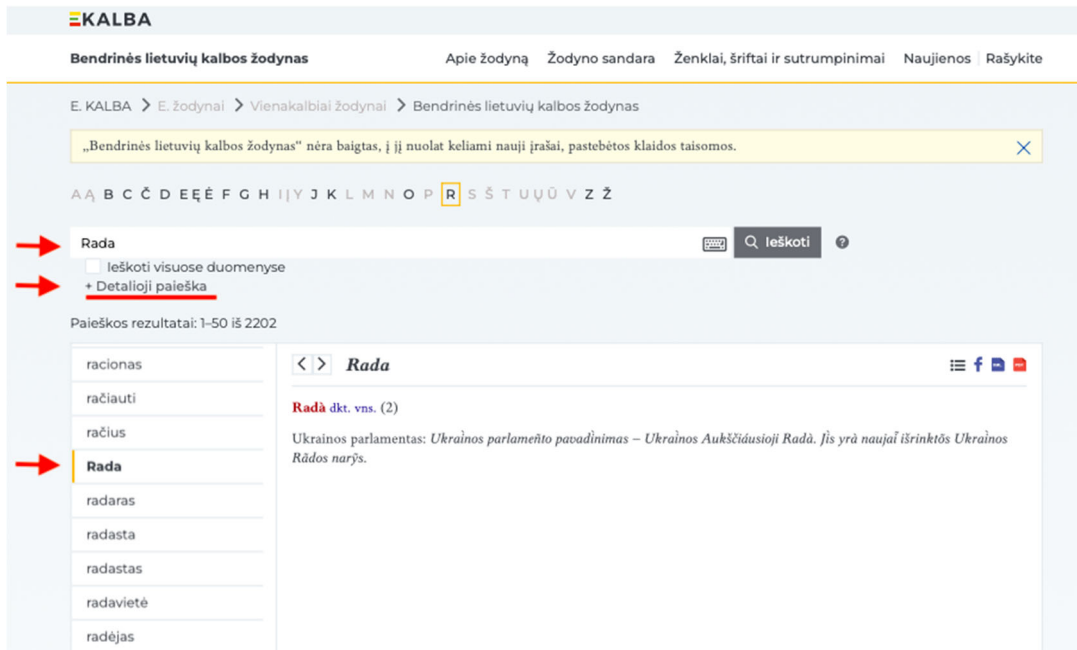
Rinktis išteklius: Pažymėti visus

| | | |
|--|---|---|
| <p>Vienakalbiai žodynai</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Bendrinės lietuvių kalbos žodynas <input type="checkbox"/> Dabartinės lietuvių kalbos žodynas <input checked="" type="checkbox"/> Lietuvių kalbos žodynas <input type="checkbox"/> Lietuvių kalbos naujažodžių duomenynas <input type="checkbox"/> Sinonimų žodynas <input type="checkbox"/> Antonimų žodynas <input type="checkbox"/> Frazeologijos žodynas <input type="checkbox"/> Palyginimų žodynas <input type="checkbox"/> Sisteminis žodynas <p>Dvikalbiai žodynai</p> <ul style="list-style-type: none"> <input type="checkbox"/> Lietuvių–anglų kalbų žodynas <input type="checkbox"/> Anglų–lietuvių kalbų žodynas <input type="checkbox"/> Lietuvių–latvių kalbų žodynas <input type="checkbox"/> Latvių–lietuvių kalbų žodynas <input type="checkbox"/> Lietuvių–lenkų kalbų žodynas <input type="checkbox"/> Lenkų–lietuvių kalbų žodynas <input type="checkbox"/> Lietuvių–vokiečių kalbų žodynas <input type="checkbox"/> Vokiečių–lietuvių kalbų žodynas <input type="checkbox"/> Lotynų–lietuvių kalbų žodynas <input type="checkbox"/> Senosios graikų–lietuvių kalbų žodynas | <p>Žodynų kartotekos</p> <ul style="list-style-type: none"> <input type="checkbox"/> „Lietuvių kalbos žodyno“ Papildymų kartoteka <input type="checkbox"/> „Lietuvių kalbos žodyno“ Tarmių kartoteka <input type="checkbox"/> „Lietuvių kalbos žodyno“ Pagrindinė kartoteka <input type="checkbox"/> Kazimiero Būgos etimologinio žodyno kartoteka <p>Tautosakos paveldo kartotekos</p> <ul style="list-style-type: none"> <input type="checkbox"/> Mjėslų kartoteka <input type="checkbox"/> Liaudies tikėjimų kartoteka <input type="checkbox"/> Pokario partizanų dainų kartoteka <p>Duomenų bazės</p> <ul style="list-style-type: none"> <input type="checkbox"/> Lietuvos vietovardžių geoinformacinė duomenų bazė <input type="checkbox"/> Istorinių vietovardžių duomenų bazė <input type="checkbox"/> Pavardžių duomenų bazė <input type="checkbox"/> Tarmių archyvas | <p>Žodžių tinklai</p> <ul style="list-style-type: none"> <input type="checkbox"/> Ontologijų sąvokos <p>E. patarimai</p> <ul style="list-style-type: none"> <input type="checkbox"/> Kalbos patarimai <input type="checkbox"/> Žodžių darybos vedlys |
|--|---|---|

6 pav. Atsivėrusio lango „Rinktis išteklius“ pavyzdys



7 pav. Kaip patekti į „Bendrinės lietuvių kalbos žodyno“ pasvėtainę?



8 pav. „Bendrines lietuvių kalbos žodyno“ pasvetainė: žodžio *Rada* paieškos rezultatas

Paieška naudojant simbolius. „E. kalboje“ ieškant dominančios informacijos galima pasinaudoti patarimais, t. y. paspausti klaustuko simbolį (žr. 9 pav.).

Iš 9 pav. matyti, kad paspaudus klaustuko simbolį, iššoka informacija, kaip naudojant žvaigždutę (*), klaustuką (?), kabutes ieškoti dominančios informacijos, pvz.:

1) gali būti ieškoma tiksliai tokio žodžio, koks įrašytas paieškos langelyje;

2) viršutinės kabutės (iš klaviatūros) naudojamos ieškant tikslios frazės, t. y. kelių žodžių junginio;

3) žvaigždutė (*) fragmento pradžioje, viduryje ar pabaigoje reiškia, kad toje vietoje galima bet kokia arba jokia raidžių seka;

4) klaustukas (?) naudojamas, kai jo vietoje gali būti bet kokia viena raidė.



9 pav. „E. kalbos“ langas: kaip atlikti paiešką naudojant simbolius

Pavyzdžiui, į paieškos langelį įrašius *ukrain**, gaunami tokie rezultatai: *Ukraina, ukrainai, ukrainas, ukrainiečiai, ukrainietybė, ukrainietis, ukrainietiška* ir kt.; įrašius **ynas*: *adresynas, agrastynas, agurkynas, ..., brūzgynas, daigynas, ..., duomenynas* ir kt.; įrašius *gal*a*: *galaktika, galanterija, galerija, galia* ir kt. Į paieškos langelį įrašius fragmentą su klaustuku, pvz., *k?lti*, gaunami tokie rezultatai: *kalti, kelti, kilti, kult* ir kt.; įrašius *sl?ti*: *šlapti, šlėkti, šlieti, šlioti, šlubti, šluoti* ir kt.; įrašius *?auti*: *gauti, kauti, mauti, rauti, šauti* ir kt.

„E. kalbos“ naudojimo statistika. „E. kalba“ ir joje pateikiami ištekliai itin aktualūs ir pravartūs visiems lietuvių kalbą vartojantiems ar norintiems jos išmokti žmonėms,

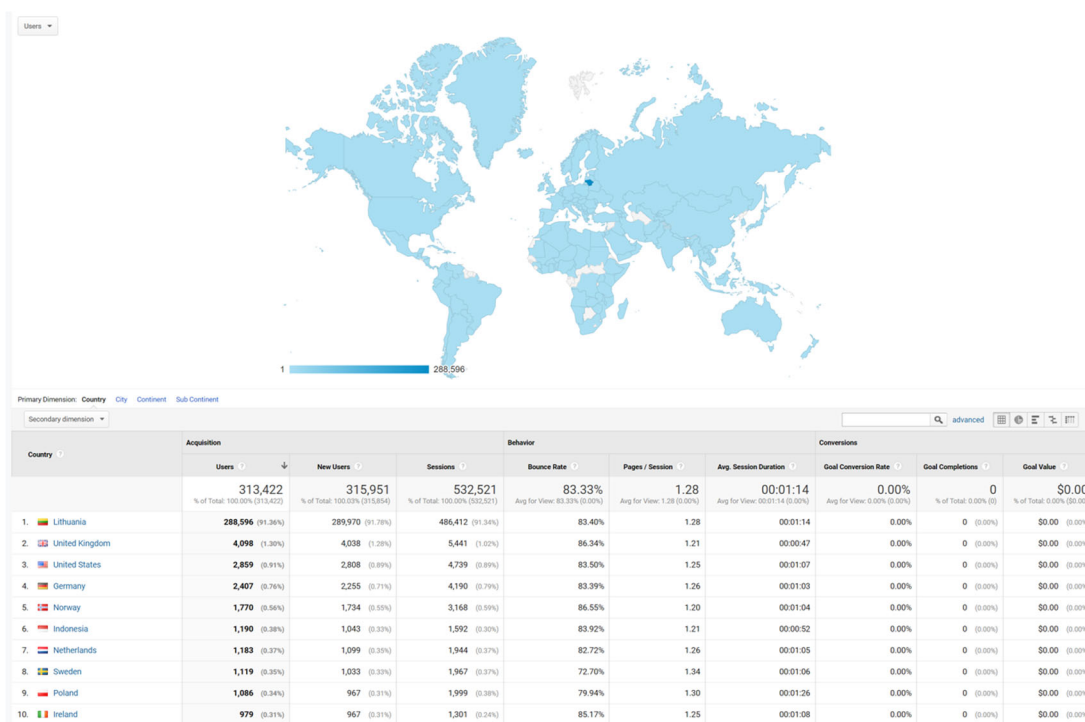
taip pat mokslui ir švietimui, valstybės institucijoms ir verslui. Tai atsispindi ir analizuojant „E. kalbos“ naudojimo statistiką (žr. 10 pav.) – vidutiniškai per mėnesį sistema naudojami apie 26000 unikalių naudotojų, užmezgama apie 44000 seansų. Iš grafiko matyti (žr. 10 pav.), kad maždaug nuo 2022 m. spalio mėnesio „E. kalbos“ naudotojų skaičius kyla. Matyt, tai susiję su didesne „E. kalbos“ sklaida. Įdomu tai, kad analizuojant sistemos statistiką galima penkių minučių laikotarpiu matyti unikalų naudotojų skaičių. Pavyzdžiui, 10 pav. dešinėje pusėje (mėlynoje zonoje) matyti, kad naudojimo statistikos peržiūros metu unikalų naudotojų skaičius siekė 33.



10 pav. Google Analytics statistikos įrankis: „E. kalbos“ naudojimo statistika (2022 m. kovo 14 d. – 2023 m. kovo 14 d. laikotarpis)

Iš 11 pav. matyti, kad daugiausia prie „E. kalbos“ jungiamasi iš Lietuvos (288596 unikalūs naudotojai per metus), antroje vietoje – iš Jungtinės Karalystės (4098 unikalūs naudotojai per metus), trečioje ir ketvirtoje vietose – iš Jungtinių Amerikos Valstijų (2859 unikalūs naudotojai per metus) ir iš Vokietijos (2407 unikalūs naudotojai per metus). Iš kitų šalių – Norvegijos, Nyderlandų ir kt. –

jungiamasi kiek rečiau. Šalys, iš kurių dažniausiai jungiamasi prie „E. kalbos“, nėra atsitiktinės. „Užsienio reikalų ministerijos ir Migracijos departamento duomenimis, apie 461 tūkst. lietuvių gyvena užsienyje. Daugiausia Jungtinėje Karalystėje (33 proc.), Vokietijoje (13 proc.), Norvegijoje (13 proc.) ir Airijoje (10 proc.)“ [1].



11 pav. Google Analytics statistikos įrankis: „E. kalbos“ naudojimo statistika pagal šalis (2022 m. kovo 14 d. – 2023 m. kovo 14 d. laikotarpis)

Baigiamosios pastabos. „E. kalba“ – kas toliau? Kaip teigiama „Lietuvių kalbos plėtros skaitmeninėje terpėje ir kalbos technologijų pažangos 2021–2027 m. gairėse“ [5], „Vertinant nacionalinių kalbos technologijų ir duomenų infrastruktūrų aktualumą ir perspektyvas, būtina užtikrinti nuolatinę jų plėtrą [...]“. Europos kalbų lygybės (angl. *European Language Equality*, <https://european-language-equality.eu>) projekto metu (2021–2022 m.) atlikti tyrimai rodo, kad nuo 2012 m. Lietuvoje padaryta didžiulė pažanga plėtojant įvairius skaitmeninius kalbos išteklius ir

paslaugas (plačiau žr. [2, p. 184–208]). Nepaisant to, kad esama sukurta nemažai lietuvių kalbos skaitmeninių išteklių, atsižvelgiant į kalbos technologijų ir visuomenės poreikius, būtina tokias sistemas kaip „E. kalba“ pildyti naujais skaitmeniniais ištekliais, vadinasi, kurti ir atnaujinti vienakalbius ir dvikalbius žodynus, leksikonus, gausinti tekstynus, ontologijas, žodžių tinklus, kurti mašininiam vertimui reikalingus keliakalbius lygiagrečiuosius tekstynus, anotuotus garsynus ir kt.

„E. kalbos“ plėtra galėtų vykti keliomis kryptimis:

1) atnaujinamas kokybinis sistemos lygmuo – integruojami nauji ištekliai (pavyzdžiui, „Vietovardžių žodynas“), baigiami skaitmeninti jau įkelti ištekliai (pavyzdžiui, „Lietuvių kalbos žodyno“ kartotekos), plėtojami esantys ištekliai (pavyzdžiui, „Žodžių tinklas“, „E. sąvokos“), kuriamos naujos visuomenei ir specialistams aktualios e. paslaugos;

2) tobulinama sistema – kuriamos papildomos paieškos galimybės, išteklių tarpusavio sąsajos, modernizuojamas kalbinių duomenų archyvavimas, pastebėtų sistemos klaidų taisymas ir pan.;

3) sudaroma galimybė perpanaudoti duomenis, t. y. užtikrinamas duomenų ir technologijų atvirumas;

4) plėtojamos integracijos su kitomis nacionalinėmis ir tarptautinėmis sistemomis;

5) atnaujinama techninė įranga ir rūpinamasi jos priežiūra.

Taigi „E. kalbos“ kūrėjai tikisi, kad vykdant sistemos plėtrą būtų prisidedama prie spartesnės kalbos technologijų pažangos.

Literatūros sąrašas

1. *Europos migracijos tinklas*. (2023). – URL: <https://123.emn.lt/#emigracija>
2. Gaidienė, A., Tamulionienė, A. (2022). European Language Equality in the Digital Age: the Case of Lithuania. *Acta Linguistica Lithuanica* 86, p. 184–208. – URL: <https://journals.iki.lt/actalinguisticalithuanica/article/view/2164/2283>
3. Jaroslaviene, J., Auksoriute, A. (2020). Innovations and Challenges in the Digital Transformation of the Lithuanian Language Industry. *Language and Economy: Language Industries in a Multilingual Europe*. Tõnu Tender, Ludwig M. Eichinger (eds.), European Federation of National Institutions for Language (EFNIL), p. 155–167.
4. Jaroslaviene, J., Miliūnaitė, R. (2020). Beribis lietuvių kalbos pasaulis skaitmeninių išteklių sistemoje „E. kalba“. *It*, 2020–11–21. – URL: <https://pasauliolietuvis.lt/beribis-lietuviu-kalbos-pasaulis-skaitmeniniu-istekliu-sistemoje-e-kalba/>
5. *Lietuvių kalbos plėtos skaitmeninėje terpėje ir kalbos technologijų pažangos 2021–2027 m. gairės*. (2020). – URL: <https://www.e-tar.lt/portal/lt/legalAct/71152ab0e0ee11ebb74de75171d26d52>
6. Miliūnaitė, R. (2018). Naujažodžių pateikimas ir paieškos galimybės Lietuvių kalbos naujažodžių duomenyne. Bendrinė kalba 91, p. 1–21. – URL: http://www.bendrinekalba.lt/Straipsniai/91/Miliunaite_BK_91_straipsnis.pdf
7. Murmulaitytė, D. (2022). *Žodžių darybos vedlys* plačiamai visuomenei ir specialistams. *Lietuvių kalba*, 17, p. 8–24. – URL: <https://www.zurnalai.vu.lt/lietuviu-kalba/article/view/31437/30396>
8. Urbanavičienė, J. (2022). Leksikografiniai tarmių duomenys lietuvių kalbos išteklių informacinėje sistemoje „E. kalba“. *Lietuvių kalba*, 17, p. 109–121. – URL: <https://www.zurnalai.vu.lt/lietuviu-kalba/article/view/31442/30408>

Надійшла до редколегії 05.05.23

Anželika Gaidienė, Dr. Sci. (Human.)
ORCID: 0000-0001-8775-788X
e-mail: anzelika.gaidiene@iki.lt
Institute of the Lithuanian Language, Vilnius, Lithuania

LITHUANIAN LANGUAGE RESOURCES INFORMATION SYSTEM "E. KALBA": FROM THE HISTORY, SEARCH POSSIBILITIES TO DEVELOPMENT PERSPECTIVES

This article presents the Lithuanian Language Resources Information System "E. kalba" ("E-Language"), which was developed at the Institute of the Lithuanian Language. The article therefore introduces the history of the development of "E. kalba", discusses the advantages of the system, shows how to perform a quick search in all or selected language resources, how to search for information of interest using wildcards, and briefly presents the statistics on the use of the system. The paper concludes with an introduction to the possible development perspectives of the system.

From 2021, the system is freely available to the professionals and general public at <https://ekalba.lt>. The "E. kalba" provides access to 39 language resources, which are divided into three major groups: 1) e-dictionaries ("Dictionary of the Standard Lithuanian Language", "Dictionary of the Modern Lithuanian Language", "Dictionary of the Lithuanian Language", "Database of Lithuanian Language Neologisms", "Dictionary of Synonyms", "Lithuanian-English Dictionary", "German-Lithuanian Dictionary", etc.), 2) e-catalogues ("Main Catalogue of the Dictionary of the Lithuanian Language", "Catalogue of New Additions of the Dictionary of the Lithuanian Language", "Catalogue of Riddles", etc.), databases ("Database of Surnames", etc.), and 3) e-services ("WordNet search", "E-terms", "E-names", "Sentiment Analysis", "Word Formation Guide", "Language Advice", "Educational Language games", etc.).

In addition to being a treasure trove of linguistic resources, "E. kalba" is also a multi-criteria search tool, with an effective and sophisticated search mechanism (quick and detailed search, search across all resources or within each resource, etc.). "E. kalba" offers a user-friendly way to view search results – results can be downloaded and saved in PDF and XML formats, and results can be shared on "Facebook". The system is also mobile-friendly.

Keywords: Lithuanian Language Resources Information System "E. kalba", digital resources, language technologies, search possibilities.

Анжеліка Гайдєнє, д-р гуман. наук
ORCID: 0000-0001-8775-788X
e-mail: anzelika.gaidiene@iki.lt
Інститут литовської мови, Вільнюс, Литва

ІНФОРМАЦІЙНА СИСТЕМА РЕСУРСІВ ЛИТОВСЬКОЇ МОВИ "E. KALBA": ВІД ІСТОРІЇ, ПОШУКОВИХ МОЖЛИВОСТЕЙ ДО ПЕРСПЕКТИВ РОЗВИТКУ

Представлена Інформаційна система ресурсів литовської мови "E. kalba" ("E-мова"), яка була розроблена в Інституті литовської мови. Описується історія створення й розвитку системи "E. kalba", окремо розглядаються її переваги, демонструються пошукові можливості у всіх або обраних мовних ресурсах, пояснюється процес пошуку необхідної інформації за допомогою символіє підстановки, стисло наводиться статистичні відомості використання системи. Представлено огляд можливих перспектив розвитку інформаційної системи "E. kalba".

З 2021 р. інформаційна система є доступною та безкоштовною для фахівців і широкої громадськості за електронною адресою <https://ekalba.lt>. Вона надає доступ до 39 мовних ресурсів, що поділяються на три основні групи: 1) електронні словники ("Словник стандартної литовської мови", "Словник сучасної литовської мови", "Словник литовської мови", "База даних неологізмів литовської мови", "Словник синонімів", "Литовсько-англійський словник", "Німецько-литовський словник"), 2) електронні каталоги ("Основний каталог 'Словника литовської мови'", "Каталог нових додатків 'Словника литовської мови'", "Каталог загадок"), бази даних ("База даних прізвищ"), і 3) електронні сервіси ("Пошук в WordNet", "Електронні терміни", "Електронні назви", "Аналіз настрою", "Посібник зі словотвору", "Мовні поради", "Навчальні мовні ігри" тощо).

Інформаційна система "E. kalba" є не лише скарбницею мовних ресурсів, а й пошуковим інструментом з ефективним і складним механізмом пошуку (швидкий і детальний пошук, пошук у всіх ресурсах або в одному з них тощо). "E. kalba" пропонує зручний спосіб перегляду результатів пошуку – їх можна завантажити і зберегти у форматах PDF і XML, а також поділитися ними в соціальній мережі "Facebook". Інформаційна система пристосована також для мобільних пристроїв.

Ключові слова: Інформаційна система ресурсів литовської мови "E. kalba", цифрові ресурси, мовні технології, можливості пошуку.