

Valodas resursi
Latvijas Universitātes
Matemātikas un informātikas institūtā

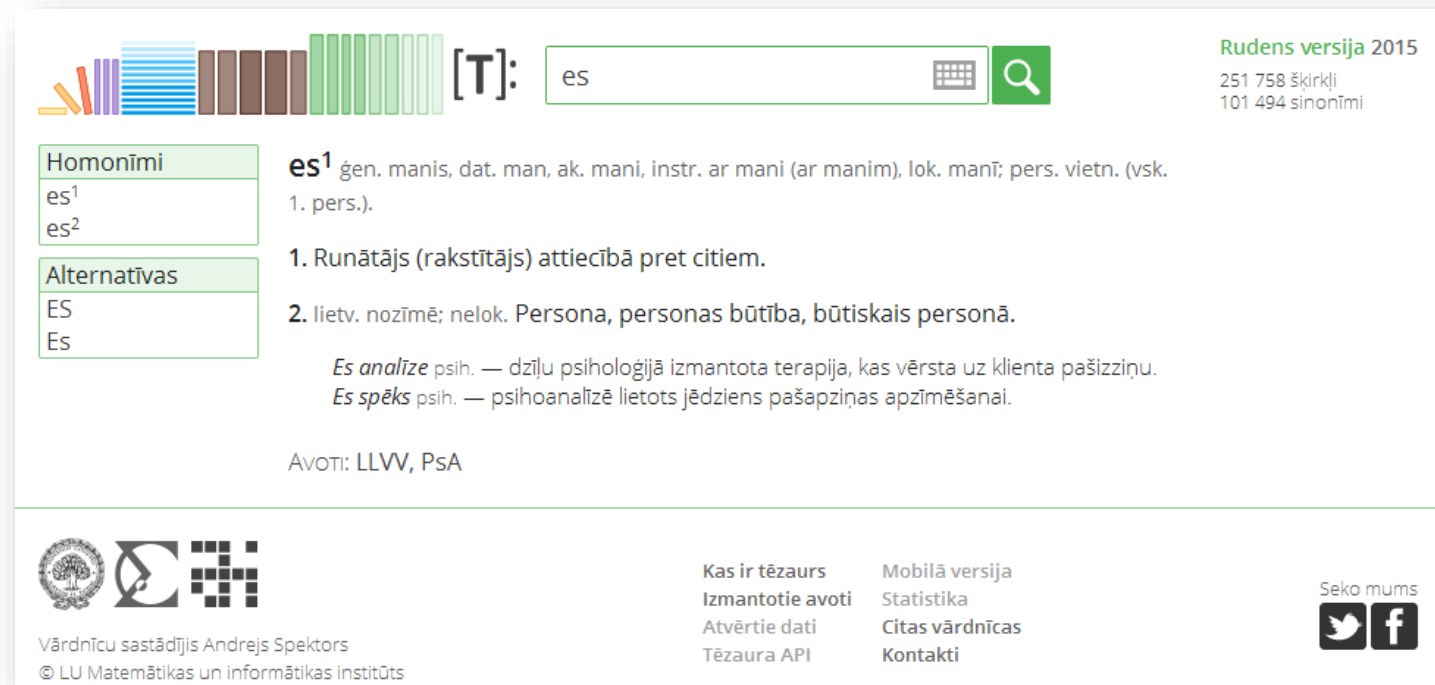
Ilze Auziņa

- LU MII Mākslīgā intelekta laboratorijā latviešu valodas resursu uzkrāšana un izstrāde norit jau vairāk nekā 25 gadus
- Nozīmīgākie resursi:
 - Elektroniskās vārdnīcas
 - Tekstu un runas korpusi

- *Tēzauris*

- <http://tezaurs.lv>

- Rudens versija 2015



The screenshot shows the Tezaurs.lv website interface. At the top, there is a search bar with the text 'es' and a search icon. To the right of the search bar, it says 'Rudens versija 2015' and '251 758 šķirkļi' and '101 494 sinonīmi'. Below the search bar, there are two columns of results. The left column is titled 'Homonīmi' and lists 'es¹' and 'es²'. The right column is titled 'Alternatīvas' and lists 'ES' and 'Es'. Below these columns, there is a detailed definition for 'es¹' in the genitive case, followed by two numbered points: '1. Runātājs (rakstītājs) attiecībā pret citiem.' and '2. lietv. nozīmē; nelok. Persona, personas būtība, būtiskais personā.' There are also two sub-definitions: '*Es analīze* psih. — dziļu psiholoģijā izmantota terapija, kas vērsta uz klienta pašizziņu.' and '*Es spēks* psih. — psihoanalīzē lietots jēdziens pašapziņas apzīmēšanai.' At the bottom left, there are logos for the Latvian Language Academy and the Institute of Mathematics and Informatics, along with the text 'Vārdnīcu sastādījis Andrejs Spektors' and '© LU Matemātikas un informātikas institūts'. At the bottom right, there are links for 'Kas ir tēzauris', 'Izmantotie avoti', 'Atvērtie dati', 'Tēzaura API', 'Mobilā versija', 'Statistika', 'Citas vārdnīcas', and 'Kontakti'. There are also social media icons for Twitter and Facebook with the text 'Seko mums'.



- LLVV – Latviešu literārās valodas vārdnīcas elektroniskā versija
 - <http://www.tezaurs.lv/llvv>

LETONIKA Vārds:

> 64 200 šķirklju
07.12.2011.

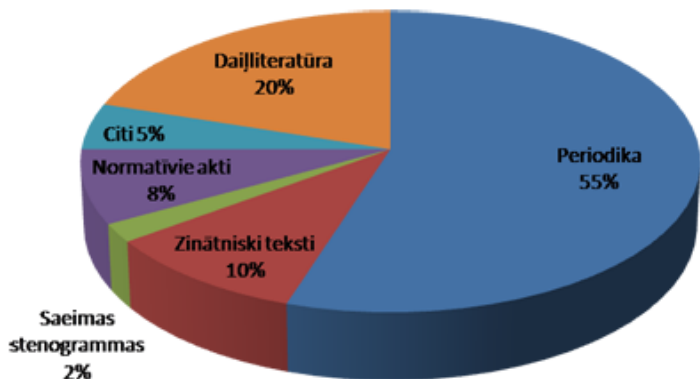
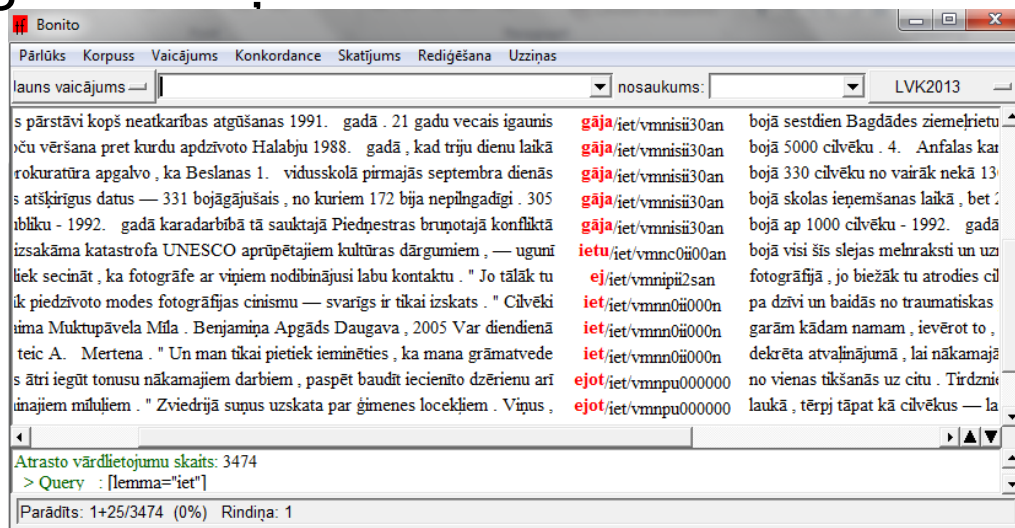
meklēt vārdus, kas sākas ar...
 piedāvāt arī ortogrāfiski līdzīgos vārdus
 nerādīt citātus

[Vārdnīcas uzbūve](#)
[Šķirklja piemērs](#)
[Apzīmējumi](#)
[Saišinjumi](#)
[Avoti](#)

Latviešu literārās valodas vārdnīca
Projektu finansē Valsts pētījumu programma "Letonika" (2006-2009)
un "Nacionālā identitāte" (2010-2011)
Iestrādes atbalstījis Valsts kultūrkapitāla fonds
© LU [Matemātikas un informātikas institūts](#)
© LU Latviešu valodas institūts
© Apgāds "Zinātne", 1972-1996

- MEV – Mīlenbaha un Endzelīna Latviešu valodas vārdnīca
 - <http://www.tezaurs.lv/mev/>

- Līdzsvarots mūsdienu latviešu valodas tekstu korpus (LVK 2013)
 - Publiski pieejams <http://www.korpus.lv>
 - Ap 4,5 miljoni vārdlietojumu
 - Automātiski morfoloģiski marķēts

Bonito interface showing a search for the lemma "iet" in the LVK2013 corpus. The search results display a list of words with their morphological tags and a snippet of text from the corpus. The text snippet includes the following words: "s pārstāvi kopš neatkarības atgūšanas 1991. gadā . 21 gadu vecais īgaunis", "oču vēršana pret kurdu apdzīvoto Halabju 1988. gadā , kad triju dienu laikā", "rokuratūra apgalvo , ka Beslanas 1. vidusskolā pirmajās septembra dienās", "š atšķirīgus datus — 331 bojāgājušais , no kuriem 172 bija nepilngadīgi . 305", "ibliki - 1992. gadā karadarbībā tā sauktajā Piedņestras bruņotajā konfliktā", "izsakāma katastrofa UNESCO aprūpētajiem kultūras dārgumiem , — uguni", "liek secināt , ka fotogrāfe ar viņiem nodibinājusi labu kontaktu . " Jo tālāk tu", "lk piedzīvoto modes fotogrāfijas cinismu — svarīgs ir tikai izskats . " Cilvēki", "uma Muktopāvela Mīla . Benjamiņa Apgāds Daugava , 2005 Var diendienā", "teic A. Mertena . " Un man tikai pietiek ieminēties , ka mana grāmatvede", "s ātri iegūt tonusu nākamajiem darbiem , paspēt baudīt iecienīto dzērienu arī", "majiem mīļajiem . " Zviedrijā suņus uzskata par ģimenes locekļiem . Viņus ,", "gāja/iet/vmnisi30an", "gāja/iet/vmnisi30an", "gāja/iet/vmnisi30an", "gāja/iet/vmnisi30an", "ietu/iet/vmnc0ii000an", "ej/iet/vmnpü2san", "iet/iet/vmnn0ii000n", "iet/iet/vmnn0ii000n", "iet/iet/vmnn0ii000n", "ejot/iet/vmnpu000000", "ejot/iet/vmnpu000000".

Atrasto vārdlietojumu skaits: 3474
> Query : [lemma="iet"]
Parādīts: 1+25/3474 (0%) Rindīņa: 1



- Sintaktiski un semantiski marķēti korpusi
 - Sintaktiski marķēts latviešu valodas korpuss (*Treebank*)
 - Biežāk lietoto latviešu valodas verbu gramatiskās un semantiskās valences korpuss
- Dažādi specializētie korpusi, piemēram,
 - Latvijas Republikas Saeimas sēžu stenogrammu korpuss
 - <http://saeima.korpuss.lv>
 - Latviešu valodas seno tekstu korpuss
 - <http://senie.lv>
 - Lietuviešu-latviešu-lietuviešu paralēlo tekstu korpuss (LiLa)
 - <http://www.korpuss.lv/lila>
 - Mūsdienu latgaliešu tekstu korpuss (MuLa)

- Runas atpazīšanas korpusi (Ortogrāfiski marķēti latviešu valodas runas korpusi)
 - Ortogrāfiski marķēti dati (100 stundas)
 - Fonētiski marķēti dati (4 stundas)
 - Līdzsvarots
 - Izstrādes laiks – 2013. g.
 - Nav publiski pieejams

Apjoms	~100 h
Vārdu skaits	~72.5 tūkst.
Vārdlietojumu skaits	~837 tūkst.
Runātāju skaits	1 851
Vīrieši	1 016
Sievietes	835

- Latviešu valodas runas korpus (LaRKO)
 - Publiski pieejams – <http://larko.ailab.lv>
 - Apjoms: 8 stundas ortogrāfiski marķētu datu
 - Izstrādes laiks – 2014. gads

Par korpusu Statistika Ortogrāfiskā transkripcija Meklēt korpusā Ielogoties

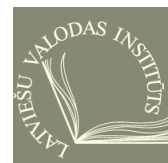
Latvija 🔍 Meklēt! ⌵ Filtrs 🗨 Palīdzība

« 1 »

...āte (.hh) (ē) ka tā noteikti ir jānovērš (0.52) (ē) Rīga un	Latvija	ir viens vienots veselums (.hh) otrs ir tas ka mēs esam (ē...
...un diriģentiem un tai pašā laikā jūs vienmēr pieminat to ka	Latvija	šeit ir īpaša un lieliska un brīnišķīgs orķestris...
...odegvielas piejaukums (.h) kas degvielu sadārdzināja bet ko	Latvija	solījusi turpināt (.h) tomēr atbalstot tikai degvielas patēr...
...tromobiļu ieviešanā jau dzirdēts ne vienu reizi vien tikmēr	Latvija	ir iesācējs šajā jomā (.h) ar elektrību darbināmas automašīn...
...audzi cilvēki visā pasaulē (.h) pēc tam meklēja kur ir tāda	Latvija	atrada kaut ko lasīja par val- pa mūsu valsti vi- vispār uzz...
...lāniem veidot Eirozonas stabilitātes mehānismu (.h) lai arī	Latvija	nav Eiropas vienotās valūtas telpas dalībniece izmaiņas Saei...
... uzņēmumu radītājs (.h) Dendrolait [1 , en , Dendrolight]	Latvija	valdes priekšsēdētājs Jānis Vašuks...
...s uzņēmumu radītājs (.) Dendrolait [1 , en , Dendrolight]	Latvija	valdes priekšsēdētājs Jānis Vašuks (.h) par to kā inovatīvas...
...sām piecdesmit trīs skolām nav pareizi tik mazai valstij kā	Latvija	neapšaubāmi ir (.h) jākoncentrē investīcijas...



VYTAUTO DIDŽIOJO
UNIVERSITETAS
MCMXXII



Latviešu valodas
aģentūra



Latvijas Zinātnes padome



VALSTS
PĒTĪJUMU PROGRAMMA
NACIONĀLĀ
IDENTITĀTE



Latvia
Lithuania
cross border
cooperation programme
2007 - 2013

bringing neighbours closer



part financed by
European Regional
Development Fund



ERAF

KOMPETENCES
CENTRS

EIROPAS REĢIONĀLĀS
ATTĪSTĪBAS FONDS

IEGULDĪJUMS TAVĀ NĀKOTNĒ!



EIROPAS SAVIENĪBA