

Latviešu valodas leksiskais tīkls (Latvian Wordnet)

CLARIN prakiskais seminārs 08.12.2022

**Pēteris Paikens,
Laura Rituma**

Kas ir WordNet?

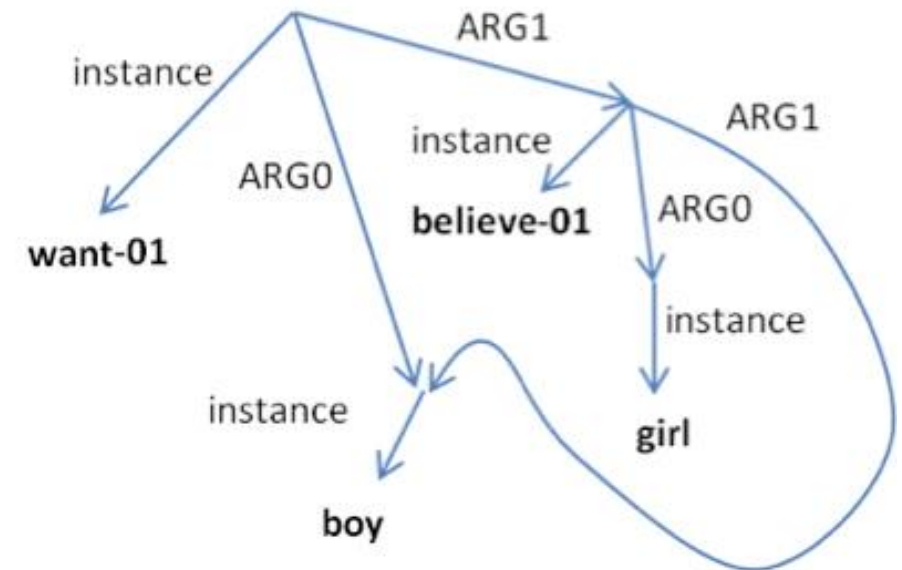
Apjomīgs, leksiski semantisks resurss kurā apkopota strukturēta informācija par

- lietvārdu,
- darbības vārdu,
- īpašības vārdu,
- apstākļa vārdu

nozīmēm un to savstarpējām semantiskajām attiecībām (Fellbaum 1998).

Kāpēc WordNet ir noderīgs resurss?

- Dabiskās valodas apstrāde (NLP) (piem. noskaņojuma analīze);
- vārdu nozīmju nošķiršana (automātiska);
- teksta nozīmes analīze (semantic parsing);
- abstraktu nozīmju atveide;
- mašīntulkošana;
- teksta anotēšana.



Kāpēc WordNet ir noderīgs resurss?

- Labāka, “dziļāka” vārdnīca - “trīs vienā”:
 - skaidrojošā vārdnīca, kas nodala un izskaidro nozīmes;
 - sinonīmu vārdnīca, kas norāda līdzīgus un saistītus vārdus;
 - tulkojošā vārdnīca ar piesaisti citām valodām (caur PWN).
- Resurss valodas apguvējiem
- Palīgs tulkošanā - tulkojumi “nozīmes līmenī” ir kvalitatīvi citādāki.
- Kad būs vārdu nozīmju nošķiršana - iespēja meklēt korpusā pēc nozīmes.

Dažādu valodu WordNet

- Princeton WordNet (1990) – prototips, līderis; 117 000 *synset* pašreizējā v3.1
- Open Multilingual Wordnet – apvieno 34 valodu brīvi pieejamos WordNet;
- PlWordNet (2009) – viens no lielākajiem tīkliem pasaulē (178 000 šķirklū);
- EuroWordNet (1999) – 7 Eiropas valodas + angļu;
(pašreiz wordnet kaut kādā apjomā ir visām ES valodām);
- IndoWordNet (2006) – 18 valodas, 18 institūti, 3 apakšprojekti + angļu.

Latviešu valodas WordNet:

LUMII AiLab projekta "Latviešu valodas WordNet un vārdu nozīmju nošķiršana" (Izp-2019/1-0464) mērķi:

- izveidot leksisko tīklu biežāk lietoto patstāvīgo vārdu nozīmēm (biežumu nosakot pēc korpusa);
- izveidot starpvalodu saites starp latviešu valodas un Prinstonas angļu valodas wordnet sinonīmu kopām.

Veidots uz skaidrojošās tiešsaistes vārdnīcas "Tēzaur.lv" bāzes.

<https://wordnet.ailab.lv>



Leksiskā tīkla izveides gaita:

- rīku izveide un integrācija platformā Tezaurs.lv;
- vārdu atlase, nozīmju revīzija un piemēru atlase;
- manuāla vārdu semantisko saišu marķēšana;
- starpvalodu semantisko saišu kandidātu automātiskā iegūšana;
- manuāla starpvalodu saišu marķēšana;
- marķējuma vienprātības novērtēšana.

Tēzaurs



- Tiešsaistes skaidrojošā vārdnīca
- Mašīnlasāms resurss
- Satur vairāk nekā 350 000 ierakstu no apmēram 280 dažādiem avotiem
- ~ 16 tūkst. lietotāju dienā (2022.gada septembris)
- 1 milj.+ šķirklju skatījumu mēnesī (2022.gada septembris)



[T]:

vārds vai frāze



Pieslēdzies kā **laura**

Datubāze: tezaurs_dv
Lietotne: 2021-03-21 19:44

☐ Rādīt visas apakšnozīmes ☐ Rādīt visus piemērus

spēlēt

spēlēt 2. konjugācijas darbības vārds; transitīvs

1. Veikt noteiktu darbību kopumu (spēli), kam ir sacensības pazīmes un ar ko cenšas sasniegt vēlamu rezultātu, izmantojot prasmes, iemaņas, arī apstākļu nejaušu sakritību; gūt prieku, izklaidēties.

☒ Rādīt apakšnozīmes ☐ Rādīt piemērus

1.1. Veikt darbību kopumu (spēli) ar mērķi sagādāt prieku, izklaidēšanos, kam parasti raksturīga iztēlē radīta situācija un darbības objekti, kādu norišu, cilvēku, dzīvnieku u. tml. atdarināšana; rotaļāties.

2. Atveidot, tēlot (lomu drāmas daiļdarbā vai filmā); īstenot skatītāju priekšā (iestudētu izrādi).

☒ Rādīt apakšnozīmes ☐ Rādīt piemērus

Spēlēt kumēdiju novecojis — tēlot teātra izrādē.

Spēlēt teātri — 1. piedalīties teātra izrādēs, tēlojot lomas. 2. izlikties.

2.1. nāvēstā nozīmē izlikties (par ko) izturēties tā, ka iznaužas (kāda

spēlēt -ēju, -ē, -ē, pag. -ēju, tr. darb.

1. Piedalīties (kā) spēlē (1); nodarboties ar (kā) spēli (1).

S. futbolu, volejbolu.

Iemācīties s. šahu.

S. dambreti, novusu, skvošu.

S. kārtis.

Beigt s. tenisu.

S. bumbu – ar roku vai kāju skarot (bumbu), virzīt (to kur).

2. Piedalīties (kā) spēlē (2); nodarboties ar (kā) spēli (2).

Komentārs

MLVV

LLVV

Šķirkļa vārds

spēlēt

☒ pēc lemmas

☐ pēc vārda formas

Korpuss

Līdzsvarotais 2018

Daiļliteratūra

- Un pa ielu **spēlēdami**, dziedādami nāk jauni puiši un viņu vidū liela, skaista meita — melnie, vilņainie mati pār pleciem, seja balta, lūpas sarkanas kā tā saule. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Dižistabā vīri laikam **spēlēja** kārtis. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Bet nu papucis istabā sāka kliegt par blēdīšanos un piespēlēšanu, un Voldis blāva, ka pats nemāk **spēlēt** un blēdis, gāzās krēsli, šūkāja soļi, kaut kas mīksti būksķēja. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Tur Jančelis un Bille **spēlēja** "pašiem savu māju", kad citu nebija sētā. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Nu varētu **spēlēt** nekāpšanu uz strīpām visu ceļu, bet acis, asaru aizvilktas, neļāvās saskatīt, kur strīpas, kur nē. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Kad viņi ar Jančeli **spēlētu** aiz raspuškas pašiem savu māju, pusdienās viņiem allaž būtu kūku atgriezumi. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Kovbojus **spēlēt**! ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe
- Viņai bija licies, ka Eskopam arī būs prieks **spēlēt** kovbojus, bet Eskops iekoda. ✨
1 1.1 2 2.1 3 3.1 3.2 3.3 3.4 mwe

<

1

2

3

4

5

...

21

>

1

Beigt rediģēt

Pievienot šķirklī

Aizvērt wordnet

Publiskā versija

Klasiskā versija

Labošanas vēsture

Vaicājumi

Zīņojumi

Semantisko saišu veidi latviešu leksiskajā tīklā

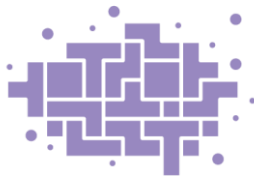
- Sinonīmija. Sinonīmu kopa (*synset*) – tīkla pamatvienība. *ēka, celtnē*
- Hiponīmija (virsjēdzieni un apakšjēdzieni): *koks - kļava*
- Meronīmija (daļas un veselā attiecības): *koks - zars, auto - motors*
- Antonīmija: *liels - mazs, pirkt - pārdot*
- Gradācijas jēdzienu grupa: *gigantisks, milzīgs, liels, neliels, mazs, sīciņš*
- Aptuvenā sinonīmija: *lēdija - dāma*
- Saistīts ar: *dzeltens - krāsa*

+ starpvalodu saites
(Princeton WordNet)

(Lokmane et al. 2021)

Sastatīšana ar Prinstonas WordNet (angļu valodā)

- Notiek sinonīmu kopu līmenī: (*piedzēries, ķitē*) → (*drunk, plastered*)
- Tiek lietotas 3 starpvalodu saites:
- tieša atbilsme: (*teātris*) → (*theatre, theater*)
- LV šaurāka nozīme: (*gājiens*) un (*brauciens*) → (*trip*)
- LV plašāka nozīme: (*pilsēta*) → (*town*) un (*city, metropolis, urban center*)



Marķēšanas rīks saišu veidošanai

☒ Rādīt visas saites

Sinonīmu kopa



izpildīt_{2,2}, spēlēt₃, atskaņot₁

izpildīt_{2,2} Atveidot ar mūzikas instrumentu (instrumentālu skaņdarbu); spēlēt, atskaņot.

spēlēt₃ Atskaņot skaņdarbu ar mūzikas instrumentu, instrumentiem vai iekārtu.

atskaņot₁ Izpildīt (skaņdarbu).

SINONĪMU VĀRDNĪCA

spēlēt – muzicēt, atskaņot, strinkšķināt, strinkšināt, trinkšķināt, trinkšināt, staipīt, virgāt, lirināt, pļerkstināt, bungot, plinkšķināt, klimpināt, klimšķināt, raudināt, čīgāt, džigāt, vilēt, zāgēt, ērgelēt, plēst vajā, raut vajā

TULKOJUMI

izpildīt – effectuate, effect, perform, accomplish, implement, comply, administer, satisfy, fulfil, fill, discharge, execute, fulfil, execute, accomplish, carry out, realize, comply with, perform, make good, attend to, attend on, fill in, fill out

spēlēt – play, game, finger, pick, play, act, perform, feature in

SAITES

ĀRĒJĀS SAITES:

[LV plašāka nozīme](#)

[\(v\) play](#)

play on an instrument; "The band played all night long"

SINONĪMU KOPAS HIPERONĪMI (VIRSJĒDZIENI):

muzicēt₁

muzicēt₁ Nodarboties ar mūziku, spēlējot kādu instrumentu, dziedot; atskaņot skaņdarbu.

SINONĪMU KOPA SAISTĪTA AR:

uzstāties₁

uzstāties₁ Izpildīt skaņdarbu, daiļdarbu, dejas u. tml. (publikas priekšā).

SINONĪMU KOPAS HIPONĪMI (APAKŠJĒDZIENI):

pavadīt₆

pavadīt₆ joma: mūzika Atskaņot pavadījumu; vadīt pavadījuma atskaņojumu.

Meklēt

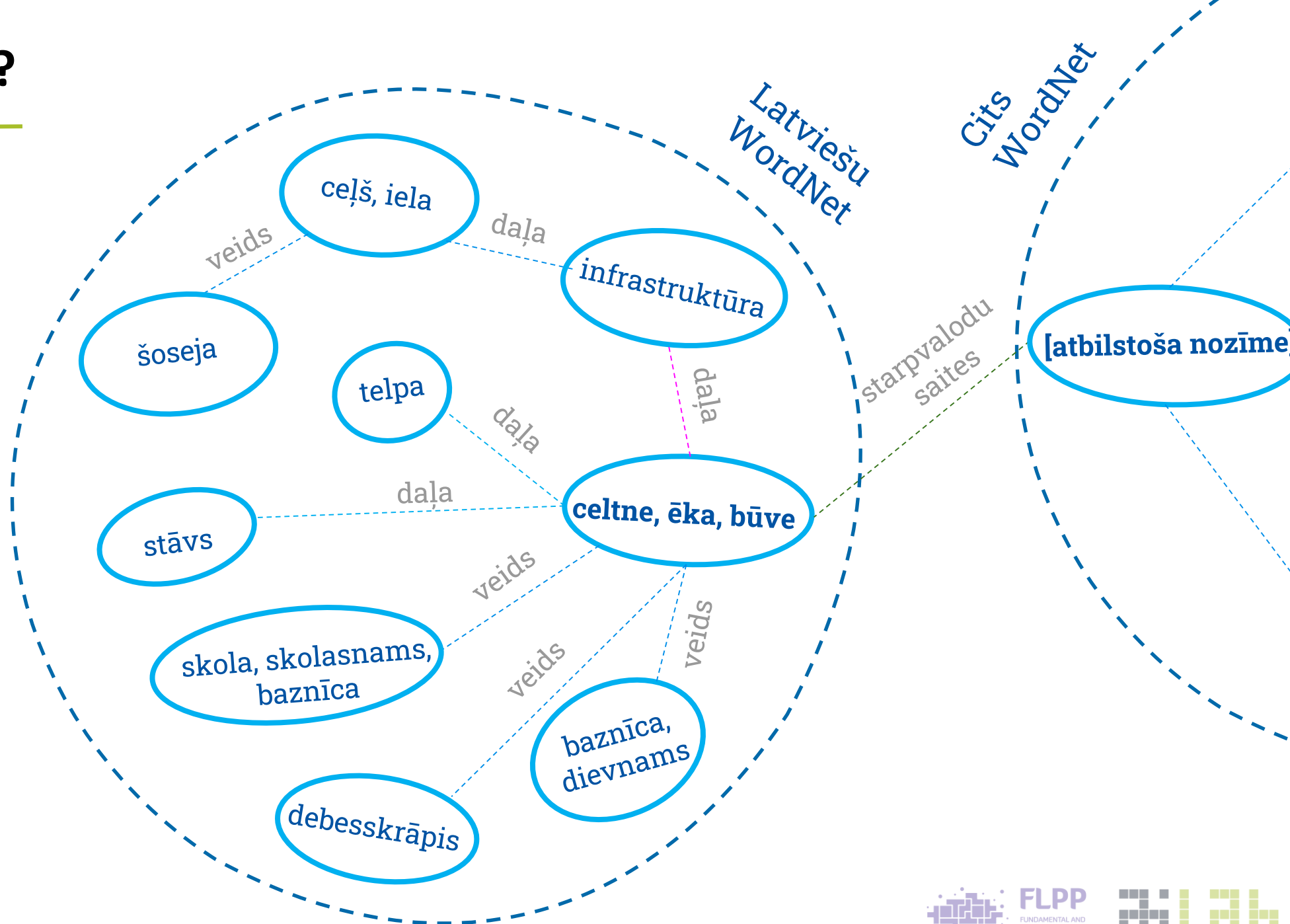
☒ Tikai nozīmes ☐ Ar apakšnozīmēm ☐ Angliski ☐ Angliski*

Kā tas “izskatās”?

sinsets

=

mazākā wordnet
vienība



Homonīmi
liels ¹
liels ²
Apkaime
lielrendenieki
lielrendnieki
lielrozes
Lielrūjene
lielrunātājs
lielrūpniecība
lielrūpnieks
liels
lielsādža
lielsaimniecība
lielsaimnieks
lielsaišķis
lielsaitis
lielsakta
Lielsala

liels¹

liels [liēls] īpašības vārds   Locīšana

<https://tezaurs.lv/liels>

1. Tāds, kas pārsniedz pēc izmēriem ko citu līdzīgu; pretstats: mazs.

- ✓ Piemēri *Kad iebraucu Amerikā, pirmais, ko ievēroju, bija skaistās, lielās amerikāņu automašīnas.*
- ✓ Saistītās nozīmes *brangs, dižens, dižs, ievērojams, pamatīgs, prāvs*
- ✓ Stabili vārdu savienojumi *Lieli stipri.*
- ✓ Tulkojumi *large, big, great*

1.1. parasti kopā ar: prievārds "par" Tāds, kas pārsniedz vajadzīgo izmēru (piemēram, par apģērbu). 

- ✓ Piemēri *Kleita viņai bija par lielu, laikam kāda vecāka māsa no tās izaugusi un Ērika vēl neieaugusi.*
- ✓ Saistītās nozīmes *paliels*
- ✓ Tulkojumi *outsize, outsized, oversize, oversized*

1.2. formā: noteiktā galotne Tāds, kas pārsniedz parasto, piemēram, savas dzimtas sugu, šķirņu, noteiktas priekšmetu grupas, lielumu.  

- ✓ Stabili vārdu savienojumi *Lielais grozītājs. Lielais lakats.*
- ✓ Tulkojumi *major*

1.3. Tāds, kas pārsniedz parasto attālumu starp sākuma un beigu punktu (par kustībām).

- ✓ Piemēri *Tālāk speriet četrus lielus soļus uz priekšu.*
- ✓ Saistītās nozīmes *plašs*

2. Tāds, kas pārsniedz pēc skaitliskā apjoma ko citu līdzīgu.

- ✓ Piemēri *Ap sirdi kļuva vieglāk, un es mammas, vecmammās un lielā draugu pulka priekšā jau labprāt piekritu visiem laulības noteikumiem.*
- ✓ Saistītās nozīmes *krietns, labs, prāvs, pamatīgs*
- ✓ Stabili vārdu savienojumi *Liela nauda.*
- ✓ Tulkojumi *large, big*

https://wordnet.ailab.lv

Latvian WordNet



Data



Demo



List of Words



Publications

Words processed in the construction of Latvian WordNet

Word	Links to Tēzaurs.lv	Part of Speech
būt	https://tezaurs.lv/būt:1	verb
gads	https://tezaurs.lv/gads:1 https://tezaurs.lv/gads:2	noun
tikt	https://tezaurs.lv/tikt:1 https://tezaurs.lv/tikt:2	verb
laiks	https://tezaurs.lv/laiks:1 https://tezaurs.lv/laiks:2	noun
valsts	https://tezaurs.lv/valsts:1 https://tezaurs.lv/valsts:2	noun
darbs	https://tezaurs.lv/darbs:1	noun
liels	https://tezaurs.lv/liels:1 https://tezaurs.lv/liels:2	noun, adjective
cilvēks	https://tezaurs.lv/cilvēks:1	noun
jauns	https://tezaurs.lv/jauns:1	adjective
vēl	https://tezaurs.lv/vēl:1 https://tezaurs.lv/vēl:2	particle, adverb

Showing 1 to 10 of 1136 results

Previous

1

2

3

...

114

Next

Resursa izveides rezultāti:

- semantikas un leksikogrāfijas teorijas attīstīšana;
- pieejami dati valodas apguvējiem un tulkotājiem;
- automātiskās nozīmju nošķiršanas eksperimenti;
- pieejami mašīnlasāmi dati datorlingvistikas rīku attīstīšanai.

Tālākā attīstība

- Projektu pieteikumi tālākai attīstībai
 - LZP “Latviešu valodas skaitļojamo leksisko resursu attīstīšana dabiskās valodas sapratnei un tekstradei”
 - VPP Digital Humanities
- Leksiskā tīkla paplašināšana ar pusautomātiskām metodēm
 - "Liels un labs WordNet"
- Automātiskās nozīmju nošķiršanas eksperimenti

Atsauces

- Felbaum, C (ed.) (1998) Wordnet: An Electronic Lexical Database. MIT press. Massachusetts, USA.
- EuroWordNet. Building a multilingual database with wordnets for several European languages. Pieejams: <https://archive.illc.uva.nl/EuroWordNet/>
- IndoWordNet. A Wordnet Of Indian Languages. Pieejams: <https://www.cfilt.iitb.ac.in/indowordnet/>
- Lokmane, I. and Rituma, L. (2021). Verba nozīmju nošķiršana: teorija un prakse; verb sense distinction: theory and practice. *Valoda: Nozīme un forma*, 12: 142. – 162.lpp.
- Lokmane, I., Rituma, L., Stade, M., and Klints, A. (2021). The Latvian WordNet and word sense disambiguation: Challenges and findings. *Electronic lexicography in the 21st century (eLex 2021) Post-editing lexicography*, 232.–246.lpp.
- Open Multilingual Wordnet. Pieejams: <http://compling.hss.ntu.edu.sg/omw/>
- Paikens, P., Grasmanis, M., Klints, A., Lokmane, I., Pretkalniņa, L., Rituma, L., Stāde, M., Strankale, L. (2022). Towards Latvian WordNet. *Proceedings of The 13th Language Resources and Evaluation Conference (LREC)*. 2808. - 2815.lpp.
- Paikens, P., Pretkalniņa, L., Rituma, L.(2022). Towards Words Sense Disambiguation for Latvian. *Baltic Journal of Modern Computing*, 10(3), 402. - 408.lpp.
- PlWordnet. A Large Network of Words. Pieejams: <http://plwordnet.pwr.wroc.pl/wordnet/>
- Strankale, L. and Stade, M. (2022). Automatic word sense mapping from Princeton WordNet to Latvian WordNet. In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence*, volume 1, 478 – 485.lpp.
- Spektors, A., Pretkalniņa, L., Grūzītis, N., Paikens, P., Rituma, L., Saulīte, B. (2020). Tēzaurs.lv 2020.
- WordNet. A Lexical Database of English. Pieejams: <https://wordnet.princeton.edu>

Paldies!

Rakstiet - peteris@ailab.lv



Projekts "Latviešu valodas WordNet un vārdu
nozīmju nošķiršana" (Nr. lzp-2019/1-0464)

