

Словацький національний корпус

Slovenský národný korpus

Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied



Slovenský národný korpus

Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied

→English

- [Korpusy a databázy SNK](#)
- [Registrácia, anotácie a citovanie korpusu](#)
- [Nástroje SNK](#)
- [Prednášky a workshopy](#)
- [Poskytovatelia textov](#)
- [Elektronické slovníky a iné lingvistické zdroje](#)
- [Projekty a granty SNK](#)
- [Knižné tituly SNK](#)
- [Bibliografia SNK](#)
- [Konferencie SLOVKO](#)
- [Zahraničné korpusy](#)
- [Bližšie o korpusoch](#)
- [Kontakty a profily](#)

 Hľadaj v korpuse

Slovenský národný korpus (SNK) je elektronická databáza primárne obsahujúca slovenské texty od r. 1955 z rôznych štýlov, žánrov, vecných oblastí, regiónov a pod. v rozsahu poskytnutom autormi a majiteľmi autorských a/alebo distribučných práv na základe licenčnej zmluvy. Texty a slová v korpuse sú obohatené o jazykové informácie a predstavujú referenčný materiálový zdroj poznatkov o slovenčine a jej reálnom používaní, ktoré sa z korpusu získavajú pomocou špecializovaných vyhľadávacích nástrojov. Korpus nie je elektronická knižnica, ani **nenahrádza kodifikačné príručky**.

V korpusových dátach sa dajú vyhľadávať jazykové informácie na výskumné, učebné a iné výlučne nekomerčné ciele. Plný prístup do hlavného korpusu, podkorpusov a ďalších databáz sa získava bezplatne po [zaregistrovaní](#), jednoduché vyhľadávanie v dvoch základných korpusoch je možné bez registrácie prostredníctvom [WWW.rozhrania](#).

Novinky

- 2019-07-30: [Knižné novinky](#): Dynamické javy v súčasnej slovenčine a jej výskume
- 2019-07-17: Sprístupnená prvá verzia [slovensko-španielskeho paralelného korpusu](#)
- 2019-02-01: Zverejnený špecializovaný korpus textov [štátnej správy](#)
- 2019-01-31: Zverejnený špecializovaný korpus textov z projektu [OpenData](#)
- 2019-01-28: Sprístupnená tretia verzia databázy [Archívu nárečí SNK](#)
- 2018-12-31: Rozšírenie [Slovenskej terminologickej databázy](#) o nové termíny z oblasti **ekonómie**
- 2018-12-19: Sprístupnená nová verzia [Korpusu nárečí SNK](#)
- 2018-12-13: Sprístupnená nová verzia beletristických textov [slovensko-českého paralelného korpusu](#)
- 2018-12-13: Sprístupnená nová verzia [slovensko-latinského paralelného korpusu](#)
- 2018-12-03: Sprístupnená prvá verzia [slovensko-poľského paralelného korpusu](#)
- 2018-11-21: [Knižné novinky](#): **Frekvenčný slovník** korpusu slovenských textov

nxt_1.4.4.zip

OANC_GrAF.zip

Лекція_11.doc

Показати все



Введіть пошуковий запит тут



21:16
26.11.2019

Словацький національний корпус - це електронна база даних, що містить тексти словацької мови з 1955 року і охоплює широкий спектр мовних стилів, жанрів, областей, регіонів тощо. База даних містить словацькі тексти в реальному світі, розширені додаванням різної мовної інформації, отриманої спеціалізованим корпусом інструментів для запитів.

Словацький національний корпус - це проект, спрямований на створення електронних мовних ресурсів словацької мови (паралельні корпуси, розмовний корпус, діалектний корпус, історичний корпус, лексикографічні бази даних). Проект також спрямований на цифровізацію лінгвістичних досліджень, проведених у Словацькому національному корпусному відділі Інституту лінгвістики Штура Словацької академії наук у Братиславі. Проект підтримується Міністерством культури та Міністерством освіти, науки, досліджень та спорту Словацької Республіки.

Query


veda

Hľadať

v korpuse

prim-6.0-public-all ▾

áäčďéěííľńóôřřšťúúůýžžAAČĎĚĚÍĹLNŃOOÖRRŠŤTUŮŮÝŽ[]^*.,:;'"'\ [lemma=""]

[Nastavenia](#) • [Info](#) • [O korpuse](#) •  

| | | |
|----------------------------|-----|--|
| 13384 | () | Sandová . Čo by to bola za literárna veda alebo literárna kritika , keby silou - |
| 46420 | () | aby sa ako synonymá striedali termíny veda , vedný odbor , náuka a vedná disciplína |
| 140476 | () | račiansky , ale Biňa — biňanský ; veda — vedný , ale právo — právny ; žena |
| 171743 | () | . Štvrtá kapitola nesie titul Národ a veda . Autor sám ako vedec sa na pôde vedy |
| 171780 | () | vystupuje proti téze , že existuje iba " veda na Slovensku ") . Autorova argumentácia |
| 276114 | () | zámená , ale aj termíny ako vzdelanie , veda , kultúra atď .) , alebo je ich význam |
| 296124 | () | okolnostiam , v ktorých sa ocitla naša veda , nezastavila a že jazykovedci opätovne |
| 331263 | () | Nógrádmegye , obsahujúca v kapitole Literatúra , veda a umenie 174 významných osobností Novohradu |
| 336553 | () | slovo gréckeho pôvodu a pomenúva sa ním veda skúmajúca vzájomné vzťahy medzi živými |
| 404040 | () | trestné právo — trestnoprávny , literárna veda — literárnovedný . zoštieňovanie |
| 0..9/12221 | | (10) |

[nasledujúci](#) → [posledný](#)

Veda - «Наука»

Query



myšlienka

Hľadať

v korpuse

prim-6.0-public-all ▾

áäčďěéííľŕńóóorřšťúúüýžÁĀČĎĚĚÍĹĽŇŇŌŌŔŔŠŤUŮŮÝŽ[]^*.,:;'"\"' [lemma=""]

[Nastavenia](#) • [Info](#) • [O korpuse](#) •  

| | | |
|------------------------|-----|--|
| 13500 | () | z nich nesmiali ? Je to naozaj trúfalá myšlienka ponižovať vlastný národný jazyk , |
| 28454 | () | štúrovskú kodifikáciu nášho jazyka (myšlienka F . Mika) . Váha spisovateľovho slova |
| 58206 | () | 1987 , s . 452) . Znesiteľná nie je ani myšlienka primitívnej amerikanizácie slovenskej |
| 95826 | () | 749) . Tu sa však ponúka na zváženie myšlienka , či by sa zložené podstatné meno uznášania schopnosť |
| 103057 | () | kufre a podlomili sa mi kolena . Prvá myšlienka — nazad na stanicu ! Ale od slabosti |
| 175450 | () | doba nie je ešte pravá , lebo zdanlivá myšlienka sa dostane do rozporu so samým jazykom |
| 241857 | () | z poľštiny sa zasa nájde formulácia myšlienka , že zem obehovala dookola svojho slnka |
| 389874 | () | zborníka a uvádza sa iba strana , kde sa istá myšlienka , téza alebo tvrdenie vyskytuje . V úvodzovkách |
| 479447 | () | najmä z tej príčiny , že nám chýba aj myšlienka) , tak je pravda i to , že niekedy to |
| 525331 | () | Je to vlastne inými slovami vypovedaná myšlienka o spätosti vývinu slovnej zásoby s rozvojom |

0..9/24420 (10)

[nasledujúci](#) → [posledný](#)

Myšlienka – думка, идея

На початку 1990-х у галузі словацької лінгвістики сфера корпусної лінгвістики не була задіяна. Дж. Містрік (*Encyklopédia jazykovedy*, 1993) передбачив істотний вплив різних факторів на словацьку мову та її дослідження. Такі фактори, як комп'ютеризація, розвиток корпоративних та мовних технологій, також вплинули на формування Словацького національного корпусу. Внутрішні умови - необхідність матеріалу для вивчення мови та складання нового одномовного словника словацької мови також мають великий вплив на його становлення. Словацький національний корпус був заснований у 2002 році за підтримки Міністерства освіти, Міністерства культури та Словацької академії наук. SNC складається з декількох проектів, головним чином орієнтованих на лінгвістичні дослідження та викладання мови. В даний час корпус prim-6.1 містить 829 мільйонів лексем. Документи - тексти в корпусі містять багатий опис метаданих, включаючи детальну стильову та жанрову анотацію та морфологічну анотацію. Є багато споріднених корпорацій, наприклад вручну морфологічно позначений корпус, словацький WebCorpus, корпус юридичних текстів, корпус розмовної словацької мови, кілька паралельних корпорацій (словацько-французька, словацько-російська, словацько-чеська, словацько-англійська, словацько-латинська). Окремі проекти - словацький аналізатор морфології, база даних словацької термінології, словацька WordNet, корпус історичної словацької мови. Словацьких мовних ресурсів цілком достатньо для базових мовних досліджень, але НЛП для словацької потребує подальшої підтримки.

Текстовий корпус словацької мови продовжував своє зростання до 2002 року, коли це стало зрозумілим

що виникла потреба у новій версії корпусу, яка була б репрезентативною, зазначається

і доступний професійній спільноті. 13 лютого 2002 р. Уряд словацької мови

Республіка схвалила проект Словацького національного корпусу та проект комп'ютеризації лінгвістичного

дослідження в Словаччині. З цього часу Словацький національний корпус застосовував системний підхід у Росії

створення бажаного корпусу текстів словацької мови

динамічні / моніторингові vs статичні;

дослідницькі vs ілюстративні;

загальномовні vs спеціалізовані (діалектні, термінологічні, дитячої мови тощо);

синхронні vs діахронні;

усної мови vs писемні vs мішані корпуси;

повнотекстові vs фрагментні;

оригінальні vs перекладні;

мономовні vs полімовні;

паралельні vs порівняльні;

неанотовані vs анотовані;

Коли ми говоримо про Словацький національний корпус, ми маємо на увазі фактично дві різні речі.

Перший - Словацький національний корпус, як кафедра Інституту мовознавства ім. Л.Н. Штура брати участь у кількох різних проектах, серед них також текстовий корпус словацької мови (проект корпусу) назва Словацький національний корпус. Однак його має бути достатньо зрозуміло з контексту, якщо ми говоримо про відділ чи про корпус. Наші нинішні плани передбачають створення текстового корпусу, що складається з текстів, опублікованих протягом 1955–2005 рр. Нижня межа продиктована суттєвою реформою правопису, проведеною у 1953 році (і дати два роки для «заспокоєння» правопису). Ми очікуємо, що більшість текстів може охопити більшу частину за цей часовий інтервал, ретельно вибираючи тексти, щоб підтримувати рівномірний розподіл часу та жанру, тоді як решта корпусу охоплюватиме більш пізні дати, коли електронні версії текстів легко доступні. Бажаний розмір корпусу - 200 мільйонів слів, які ми припускаємо збирати протягом 2005 р. На момент написання корпусу міститься 187 мільйонів слів, але є сильно незбалансований, що складається здебільшого з журналістських текстів. Як мінімальна вимога потенційних користувачів, корпус повинен бути лематизований та містити інформацію про морфологію, а також бібліографічну анотацію. Ми розраховуємо створити «серцевину», що складається з лематизованих та морфологічно анотованих текстів уручну приблизно 1 мільйон слів, які можуть бути використані для підготовки морфологічних аналізаторів та інших НЛП інструменти для використання з рештою корпусу. Доступ до всього корпусу (за винятком текстів із занадто забороненими авторськими ліцензіями) є публічно доступний безкоштовно в Інтернеті в обмін на просту процедуру реєстрації.