

Nový český slovník pro kontrolu pravopisu

LinuxDays 2019

Stanislav Horáček

česká skupina kolem LibreOffice
The Document Foundation

6. října 2019

Úvod

český slovník pro kontrolu pravopisu

Hunspell

Úvod

český slovník pro kontrolu pravopisu

Hunspell

slovník pod licencí GNU GPL

Toto je český slovník pro kontrolu pravopisu založený na českém slovníku pro ispell, verze z 29. 10. 2006, který vytvořil Petr Kolar spolu s desítkami dalších přispěvatelů.

Úvod

Český tvarotvorný slovník

slovní zásoba

- Masarykova univerzita
- únor 2019, licence public domain (= CC0)
- analýza jazykového korpusu
- podstatná jména, přídavná jména, slovesa
- ~60 000 základních tvarů
- github.com/plin/slovník

Slovníková data na Wikidatech

rozhraní

- oddělený prostor
- databáze slovní zásoby
- lexémy – tvary, významy, příznaky
- ~3000 základních tvarů
- během roku 2018, licence CC0
- stabilní, univerzální
- www.wikidata.org/wiki/Wikidata:Lexicographical_data/cs

Úvod

Český tvarotvorný slovník
+ slovníková data z Wikidat
= české CC0 slovníky

Úvod


Český tvarotvorný slovník
+ slovníková data z Wikidat
= české CC0 slovníky

experimentální!

Jak vyzkoušet

rozšíření pro LibreOffice

extensions.libreoffice.org/extensions/czech-cc0-dictionaries-ceske-cc0-slovniky

Log in Register

[Home](#) [Extensions](#) [Templates](#) [Events](#)

[Home](#) [Extensions](#) [Czech CC0 dictionaries / České CC0 slovníky](#)

Czech CC0 dictionaries / České CC0 slovníky

Czech spell check dictionary licensed under the Creative Commons CC0 License / Slovník kontroly pravopisu pro češtinu zveřejněný pod licencí Creative Commons CC0 👍 13 👎 1

Project Description

The dictionary uses data from the [Czech morphological dictionary](#) created by the Masaryk University in Brno and from [Czech Wikidata lexemes](#).

Repository of this dictionary can be found at [GitLab](#).

The dictionary is considered as experimental (many words missing), with a lot of space for improvements.

For more details and instructions how to contribute, see the webpage [ceskeslovniky.cz](#).

Slovník využívá data [Českého tvaroslovného slovníku](#) vytvořeného na Masarykově univerzitě v Brně a [česká slovníková data z Wikidat](#).

Repozitář s tímto českým slovníkem je k dispozici na [GitLabu](#).

Slovník je třeba považovat za experimentální (mnoho slov chybí) a je zde značný prostor pro jeho vylepšování.

Podrobnosti a návod jak se zapojit naleznete na stránkách [ceskeslovniky.cz](#).

Project Resources

- [External Project Page](#)

Screenshot



Install Instructions

To install an extension, follow these steps:

- Download an extension and save it anywhere on your computer.
- In LibreOffice, select Tools --> Extension Manager from the menu bar.
- In the Extension Manager dialog click Add.

Jak vyzkoušet

doplněk pro software Mozilly

gitlab.com/strepon/czech-cc0-dictionaries/tree/master/mozilla

Experimental Czech spell check dictionary for Mozilla applications

This is an experimental dictionary extension for Firefox. For common use please install the [dictionary from addons.mozilla.org](https://addons.mozilla.org).

How to build

Simply run the `make` command which will yield and `.xpi` file into the `target` sub-folder.

How to install

The self-build extension is not signed and Firefox will reject to install it directly from file. However you can install it temporarily (until restart) to *debug it*.

1. Open add-ons manager (`Ctrl+Shift+A`).
2. Click the *gear* button in the top right corner.
3. Select *Debug Add-ons (Ladění doplňků)*.
4. Click *Load Temporary Add-on (Načíst dočasný doplněk)* button.
5. In file picker select the `.xpi` file from the `target` sub-folder.

How to report issues

- If you are not happy with the results, please check this repository root folder on how to improve the data.
- If you have some issues with installing the extension into Firefox, please contact [Mozilla.cz](https://www.mozilla.cz). Do not forget to include a link to this repository, or explicitly mention you have issues with this **experimental dictionary**.

Jak vyzkoušet

kontrola na ceskeslovniky.cz

Kontrola slov [O slovníku](#)

České CC0 slovníky

Kontrola slov

Slovo linux days

Zkontrolovat

Slovo	Ve slovníku	Návrhy na opravu
linux	není	linu, inulin
days	není	davy

Schází ve slovníku správné slovo, nebo jste v něm naopak našli chybné? [Slovník můžete vlastnoručně vylepšit!](#)

Co znamená CC0?

CC0 je licence, díky níž lze slovníky libovolně využívat, a to jak pro komerční, tak pro nekomerční účely.

Ukázky

Konference LinuxDays 2019 znamená desítky přednášek a workshopů, stánky mnoha různých projektů (Fedora, Gentoo, openSUSE, ypsfree.cz, Python, Mozilla, brmlab, OpenAlt a mnoho dalších), LPI certifikace, mnoho zajímavých osobností a mnoho nových informací. **Vstup je volný.**

Kdy a kde...

Letošní ročník se uskuteční **5. a 6. října 2019** v Praze v areálu ČVUT v Dejvicích na Fakultě Informačních Technologii.

LinuxDays...

LinuxDays je česká linuxová konference, která se odehrává každoročně v Praze. Tato konference vznikla v roce 2012, aby nahradila zrušenou konferenci LinuxExpo. Za sedm let se z LinuxDays stala největších akce svého druhu v Česku.

Komunita...

Když se o zrušení LinuxExpa dozvěděla česká linuxová komunita, zorganizovala se a jelikož všichni chtěli konferenci typu LinuxExpo (ale ještě lepší a dokonalejší), rozhodla se udělat si konferenci sama a po svém. Kdo je součástí této komunity? Začali jsme jako skupina lidí, která se vídá na různých konferencích v Čechách a díky tomu jsou mezi organizátory konference LinuxDays zástupci různých otevřených komunit.

Ukázky

Za karmínové soumraky viditelné i v Česku může sopečný prach z Kurilských ostrovů

Výrazné červenofialové barvy, které v poslední době doprovázejí i v Česku soumraky nebo rozbřesky, způsobuje prach ze sopky na Kurilských ostrovech i erupce dalších vulkánů na zeměkouli. Fotogenický jev mohou lidé pozorovat na obloze ještě několik týdnů, možná i měsíců.

Při pohledu na karmínové nebe mají pozorovatelé dojem, že kdesi v dálce hoří. Je to způsobeno tím, že světlo paprsků se opírá o sopečný prach, a ten jejich zář rozptylem světla zvýrazňuje.

„Tyto syté odstíny v obloze na skloncích dne a noci mají poněkud více zajímavý původ, než by se mohlo běžně zdát. Stojí za nimi prach ze sopky na Kurilských ostrovech, který se větrnými poryvy neustále rozměšňuje vysoko v atmosféře nad severní zemskou polokoulí,“ napsal astronom a fotograf Petr Horálek na astronomickém informačním webu www.astro.cz.

Prach pochází ze sopky Raikoke na Kurilských ostrovech, která leží v neobydlené oblasti Ochotského moře, přibližně 950 kilometrů východně od ruského pobřeží. Měří asi 2,5 kilometru na šířku a sahá zhruba 550 metrů nad úroveň moře.

„Silnou sérii erupcí, které vyvrcholily 22. června 2019 ve čtyři hodiny tamního času, se do vzduchu dostalo několik milionů tun sopečného popela včetně oxidu síry. Sopečný oblak se díky opravdu velkému uvolnění energie dosápal do výšky nad 17 kilometrů nad zemí a začal putovat východním až severovýchodním směrem. O rozptýl prachu se pak postaraly větrné poryvy ve vrchní troposféře a stratosféře, kde se prach rozkládá doposud,“ vysvětlil Horálek.

Ukázky

Šli jsme oba ruku v ruce oba zachránění
šli jsme ruku v ruce v otevřeném snění
za město kde počínaly Košiče
z dálky mávaly nám noční vějíře
nad kiosky smutku tance alkoholů
šli jsme ruku v ruce nemluvíce spolu
bylo tu však něco těžkého co drtí
smutek stesk a úzkost z života i smrti

Odemkl jsem dveře rozžal svítiplyn
veďa na nocleh svůj pouliční stín
řek jsem pane pro nás pro oba to stačí
nebylo tu však už stínu po mém hráči
či to byl jen přízrak nebo sebeklam?
stál jsem nad svým každodenním lůžkem sám

bylo tu však něco těžkého co drtí
smutek stesk a úzkost z života i smrti

Usedl jsem za stůl nad kupy svých knih
pozoruje oknem padající sníh
pozoruje vločky jak své věnce vijí

se svou věčně chimérickou nostalgií
píják nezachytitelných odstínů
píják světél potopených do stínů
píják žen jichž poslouchají sny a hadi
píják žen jež pochovávají své mládí
píják krutých hazardních a krásných žen
píják rozkoše a zkrvavělých pěn
píják všeho krutého co štve a drtí
píják hrůz a smutku z života i smrti

Řekl jsem si zapomeň už na stíny
otvíraje týden staré noviny
kde jsem v pachu novinářské černi tona
uzřel velkou podobiznu Edisona
byl tu jeho nejnovější vynález
seděl v taláru jak středověký kněz
bylo tu však něco krásného co drtí
odvaha a radost z života i smrti.

Srovnání úspěšnosti

procentuální podíl slov označených jako chybná

	2019.06	2019.08	2019.10	GNU GPL
Dobrodružství Sherlocka Holmese	9,02	7,40	7,02	2,63
Evangelium podle Jana	7,46	5,50	4,91	0,67
LibreOffice Writer: Praktický průvodce	6,00	5,39	4,76	3,33
R.U.R.	16,82	12,49	12,03	8,37
Ústava České republiky	7,89	6,52	6,23	0,90
program LinuxDays	34,05	33,12	32,97	29,10

Jak vylepšit

- doplňování slov na Wikidata
-
-
-
-
-
-

Jak vylepšit

nový lexém na Wikidatech

www.wikidata.org/wiki/Special:NewLexeme

Special page

Create a new Lexeme

By clicking "Create", you agree to the [terms of use](#), and you irrevocably agree to release your contribution under the [Creative Commons CC0 License](#).

Create a new Lexeme

Lemma

 *

Language of Lexeme

 *

Lexical category

 *

Jak vylepšit

šablony pro různé slovní druhy

tools.wmflabs.org/lexeme-forms/

Wikidata Lexeme Forms [Documentation](#) [Wikimedia Toolforge](#) [Source code](#)

české podstatné jméno (rod střední)

1. pád, jednotné číslo

To je mé .

2. pád, jednotné číslo

Má strach z mého .

3. pád, jednotné číslo

Dej to mému .

4. pád, jednotné číslo

Vidím jedno .

5. pád, jednotné číslo

Kam kráčíš, ?

6. pád, jednotné číslo

Pověz mi něco o tvém .

7. pád, jednotné číslo

Seznámil jsem se s tvým .

1. pád, množné číslo

To jsou má .

2. pád, množné číslo

Má strach z mých .

Jak vylepšit

- doplňování slov na Wikidata
- import z Tvarotvorného slovníku
- import z Wikislovníku, Wikidat
- vzory do Hunspellu
- nové šablony
- nahlášení ze softwaru
- kampaň pro určité texty

Shrnutí

nový český slovník kontroly pravopisu

- licence CC0
- experimentální
- Tvarotvorný slovník a Wikidata
- budoucnost?
- ceskeslovniky.cz
- gitlab.com/strepon/czech-cc0-dictionaries
- lokalizace@cz.libreoffice.org