

Masarykova univerzita
Filozofická fakulta
Ústav českého jazyka
Český jazyk se specializací počítačová lingvistika

Veronika Kalivodová

**Tvorba uživatelského manuálu pro
DEBDict a Deriv**

Bakalářská diplomová práce

Vedoucí práce: Mgr. Hana Žižková
2017

*Prohlašuji, že jsem diplomovou práci vypracovala
samostatně s využitím uvedených
zdrojů a literatury.*

Poděkování

Za pomoc při psaní bakalářské diplomové práce bych ráda poděkovala mnoha lidem. Největší a nejsrdečnější dík patří mé vedoucí práce, Mgr. Haně Žižkové, za cenné rady, společnou práci a oporu při psaní, kdy už jsem nevěděla, jak pokračovat.

Dále bych ráda poděkovala RNDr. Adamovi Rambouskovi, Ph.D., a doc. PhDr. Kláře Osolsobě, Dr., za konzultace a vstřícný přístup ke všem mým dotazům k programům, k nimž jsem vytvářela manuály.

Velké díky patří i Tině Heindlové a Lukáši Noskovi za přečtení prvotních manuálů, stovky cenných připomínek a hodiny trpělivého zkoušení mého postupu. Také bych ráda poděkovala své rodině za trpělivost s mými náladami při psaní práce, když všechno nešlo tak, jak mělo.

Klíčová slova

DEBDict, Deriv, styl, text, morfologie, lexikologie, syntax, doporučení, manuál

Keywords

DEBDict, Deriv, style, text, morphology, lexicology, syntax, recommendation, user manual

Anotace

V této bakalářské diplomové práci jsou popsána doporučení, jimiž by se autor manuálu měl řídit, pokud si přeje, aby jeho text byl účelný. V první části jsou popsány aspekty a kompozice prostědělovacího stylu a popisného slohového postupu, které jsme v druhé části specifikovali na napsání manuálu. Třetí část práce je zaměřena na vlastní manuály, při nichž jsme postupovali dle doporučení.

Abstract

This bachelor thesis describes the recommendations which the author of the manual should follow if he wishes his text to be useful. The first part describes aspects of the style and composition of the simple style of communication and the descriptive style approach, which we have specified in the second part of writing the manual. The third part of the thesis focuses on our own manuals where we followed the recommendations.

Obsah

Úvod	3
1 Co je to manuál	4
1.1 Typy manuálů	4
2 Kompozice	5
2.1 Slohový postup a funkční styl	5
2.2 Morfologie	6
2.2.1 Podstatná jména	6
2.2.2 Přídavná jména	7
2.2.3 Zájmena	7
2.2.4 Číslovky	7
2.2.5 Slovesa	7
2.2.6 Příslovce	8
2.2.7 Předložky	8
2.2.8 Spojky	8
2.2.9 Částice	9
2.2.10 Citoslovce	9
2.3 Lexikum	9
2.4 Syntax	10
2.5 Shrnutí	11
3 Tvorba textu manuálu	12
3.1 Předpoklad	12
3.2 Morfologická stránka	12
3.3 Lexikální stránka	13
3.4 Syntaktická stránka	13
4 Doporučení	14
5 DEBDict	15
5.1 Výkladové slovníky	15
5.1.1 Příruční slovník jazyka českého	15
5.1.2 Slovník spisovného jazyka českého	16
5.1.3 Slovník spisovné češtiny	16
5.2 Slovníky současného jazyka	17
5.2.1 Akademický slovník cizích slov	17
5.2.2 Slovník české frazeologie a idiomatiky	17
5.2.3 Slovník českých synonym	18
5.3 Český WordNet	18

5.4	Morfologický analyzátor <i>ajka</i>	19
5.5	CIA World Factbook	19
6	Manuál pro DEBDict	20
6.1	Přístup k obecnému prohlížeči slovníků	20
6.2	Přihlášení	20
6.3	Dotaz	22
6.3.1	Volba elektronických zdrojů	22
6.3.2	Typ dotazu	22
6.4	Výstup	23
6.4.1	Výstup hledání ve slovnících	23
6.4.2	Výstup hledání v dalších dostupných zdrojích	26
6.5	Rozšířené vyhledávání	34
7	Deriv	36
8	Manuál pro Deriv	37
8.1	Přístup k webovému rozhraní Deriv	37
8.2	Přihlášení	37
8.3	Dotaz	38
8.3.1	Volba vyhledávání	38
8.3.2	Typ dotazu	38
8.3.3	Tagy	42
8.3.4	Regulární výrazy	50
8.4	Práce s obsahem souborů, se soubory a s adresáři	51
8.4.1	Práce s obsahem souborů	51
8.4.2	Práce s adresáři	59
8.4.3	Práce se soubory	61
	Závěr	63
	Použitá literatura	64

Úvod

S manuály se během života setká každý. Ovšem ne všechny jsou napsané tak, aby byly na první pohled srozumitelné. Někdy je musíme číst vícekrát, abychom pochopili, co nám chtěl autor sdělit, a v nemálo situacích manuálu nerozumíme vůbec, protože je napsán pro čtenáře, u nichž se předpokládají znalosti oboru, z něhož produkt pochází.

Cílem předkládané bakalářské práce bylo vytvoření uživatelských manuálů pro nástroje DEBDict a Deriv. Přečetli jsme nespočet manuálů s různým zaměřením, ať už šlo o složení židle či skříně, kuchyňský recept nebo použití počítačového programu. Zkusili jsme se jimi řídit a psali jsme si poznámky o tom, jak jsme manuálu rozuměli, co porozumění usnadňovalo a co jej naopak znesnadňovalo. Na základě této analýzy a s pomocí literatury jsme vymezili funkční styl, slohový postup, základní prvky morfologie, lexikologie a syntaxe manuálu. S těmito podklady jsme dále pracovali a na jejich bázi jsme vymezili složky, které by měly být obsažené v manuálu. Tyto složky jsme pro jednoduchost uspořádali do základního doporučení.

Podle zjištěných a sepsaných doporučení jsme v další části své bakalářské práce vypracovali manuály ke dvěma programům využívaným převážně ve studijních oborech Český jazyk a literatura a Český jazyk se specializací počítačová lingvistika. Tyto dva manuály jsme dali přečíst dvěma lidem, kteří nejsou z oboru, abychom si ověřili, jestli jsou manuály srozumitelné. Na základě jejich poznámek jsme jednotlivá doporučení pro psaní manuálu pozměnili a upřesnili.

1 Co je to manuál

Příruční slovník jazyka českého definuje manuál jako příruční zápisní knihu, převážně úřední, kde úředníci zapisovali příjmy. Tentýž slovník také uvádí, že se jedná o zápisky o práci v pivovaru, tedy o popis každodenní činnosti v pivovaru. *Slovník spisovného jazyka českého* se rovněž přiklání k definici příruční knihy, obyčejně úřední. A podle *Slovníku cizích slov* jde o knihu k rychlému zapisování, z něhož byly zápisky přepisovány do úředních knih.

V této bakalářské práci však budeme manuál chápat jako návod, popis instrukcí nějakého postupu, ať už jde o pracovní postup, nebo jakýkoliv popis děje.

1.1 Typy manuálů

Po přečtení mnoha různých manuálů¹ jsme je rozdělili na tři typy. První je psaný, druhý pak obrázkový a třetí kombinovaný.

Psaný manuál je takový, který je čistě psaný. Každý krok je popsán slovy a výsledek rovněž. Může se jednat o souvislý text nebo o výpis v bodech. Neobsahuje žádné obrázky nebo schémata. Tento typ se převážně používá u receptů.

Obrázkový manuál zase obsahuje pouze obrázky každého kroku bez textu, popřípadě s minimálním textem. Obrázkový návod se převážně používá k zobrazení postupu seskládání věcí v domácnosti, například skříně, stolu a podobně. U obrázkových manuálů se používají číslice nebo písmena například k označení šroubů a míst, kam patří. Označení směru šroubování a otáčení se místo slov používají schémata a šipky.

Kombinovaný manuál je, jak už název napovídá, kombinací obou předchozích manuálů. Jedná se o text doprovázený obrázky. Kombinovaný manuál se používá v oborech různého zaměření.

¹Prošli jsme například manuály k mikrovlné troubě, některé podkapitoly z knihy *Jak pracuje internet*, která je psaná formou manuálu, a knihu receptů.

2 Kompozice

Základní výstavba textu je proces tvorby jazykového projevu uspořádáním všech jeho složek do uspořádaného celku (Čechová et al., 2008). V této kapitole se budeme věnovat stylistickým vlastnostem při tvorbě manuálu. Nejdříve si vymezíme funkční styl a slohový postup, a podle tohoto vymezení pak sepíšeme, jaké stylové normy má text obsahovat.

V jednotlivých částech podkapitol se nejdříve seznámíme s tím, co o funkčním stylu, slohovém postupu, morfologii, lexikologii a syntaxi píše odborná literatura. Tato fakta se v jistých bodech rozcházejí, proto si v následující kapitole 3 shrneme nejdůležitější zásady při psaní manuálu, ze kterých budeme vycházet při tvorbě manuálu.

2.1 Slohový postup a funkční styl

Ke správnému napsání manuálu je třeba nejprve vybrat příslušný slohový postup. Slohový postup je metoda, jak ztvárnit téma s ohledem k záměru autora a k cíli projevu. Cílem manuálu je podat co nejpodrobnější popis toho, jak něco udělat (Čechová et al., 2008). Z toho vyplývá, že se k tvorbě manuálu nejvíce hodí popisný slohový postup.

Popisný slohový postup může být nekohezní, to znamená, že nemusí být formální (tím se blíží informačnímu slohovému postupu). Podávané informace mohou být voleny a vyjadřovány subjektivně. Jednotlivá fakta popisu se často podávají formou výčtu. Očekává se řád v pořadí informací. Popisný slohový postup je rozmanitý a ustálily se jeho různé modifikace v závislosti na funkčním stylu a na cíli komunikace. (Čechová et al., 2008)

Popisný slohový postup umožňuje vytvářet různé typy textů. Existují různé typy modifikace popisu. Manuál zařazujeme do popisu děje. Popis děje musí vyjádřit dějovou linii, je časově nezakotvený a činnosti se mohou libovolně opakovat. Kromě manuálu sem patří i konkrétní útvary, tím je myšleno statický popis ve formě výčtu pomůcek, pracovních potřeb a podobně. (Čechová et al., 2008)

Funkční styl je zobecnění okruhu výrazových prostředků plnících jednu konkrétní nebo přibližně stejnou funkci. Pomíjí se individualita, kterou do textu vnáší autor, a pozornost se soustřeďuje na vyšší míru abstrakce. Primárně máme šest funkčních stylů. Popisný slohový postup patří do prostěsdělovacího funkčního stylu, ale popis děje můžeme zařadit i do odborného stylu. V naší práci nebudeme používat odborný styl, protože předpokládáme, že následující manuály mohou číst neoboroví čtenáři či začínající studenti v oboru.

Prostěsdělovací styl se vymezuje spontánností, komunikační a fatickou funkcí (je sdělovací a kontaktová). Zaměřuje se na příjemce. Základem je běžně mluvený

jazyk. Struktura textu je volná, ale u popisného slohového postupu se očekává řád podávaných informací.

Jak jsme již zmínili, popisný postup můžeme zařadit jak do odborného stylu, tak i do prostěsdělovacího stylu. Manuál má blíže k odbornému stylu, je promyšlený a strukturovaný. V našich manuálech jsme využili prvky obou těchto stylů. Z odborného stylu jsme převzali promyšlenost a strukturovanost, bez nichž nelze napsat manuál. V manuálu se zaměřujeme na příjemce, proto jsme text obohatili o komunikaci a kontaktovost.

2.2 Morfologie

Styl projevu je ovlivněn rozložením a využitím slovních druhů. K statičnosti textu přispívají jmenné konstrukce. S obsahovou a myšlenkovou náročností roste podíl použitých sekundárních předložek a spojek. Expresivní charakter komunikátu dodávají částice, citoslovce a zájmena. Některá plnovýznamová slovesa ztrácejí svůj význam a přecházejí k částicím. Lexikálněsémantické potřeby rozhodují o využití adverbíí (Čechová et al., 2008).

2.2.1 Podstatná jména

Nominativ jmenovací je prostředek užívaný převážně v administrativě a odborných projevech, ale postupně se šíří i do běžnědorozumivací sféry. Jedná se o vyjádření substantiva ve spojení se shodným adjektivem. V běžné mluvě se spíše používají substantiva bez použití obecného jména.

U maskulin typu pán, muž a u názvů neosobních v prvním pádě množného čísla bývají stylově rozlišeny dvojtvary *-i* a *-ové*. Koncovka *-i* je charakteristická pro tvary neutrální a hovorové. Tvary s koncovkou *-ové* mají knižní charakter nebo se užívají jako tvarové varianty, kdy u koncovky *-i* docházelo ke změně kmenové souhlásky.

Různou frekvenci mají rodové varianty přejatých slov. U neuter cizího původu, která mají v předložkových pádech rozšířený kmen, se vyskytuje změna rodu a slova nabývají v běžném projevu charakter maskulin. Tento jev se může nacházet i u jiných zdomácnělých výrazů, dokonce i u termínů používaných na vysokých školách.

Ve sféře běžnědorozumivací roste potřeba počešťování slov. Zdomácnělé podoby mají větší potenci stát se hovorovými a neutrálními slovy, pokud neexistuje adekvátní výraz domácího původu.

Nejčastějšími pády jsou nominativ, akuzativ a genitiv. V běžných projevech klesá výskyt genitivu adverbálního, který je po slovesech nahrazován akuzativem nebo předložkovým spojením.

2.2.2 Přídavná jména

Přídavná jména mají v běžné komunikaci tendenci k unifikaci bez ohledu na druh. Unifikaci také podléhají plurálová adjektiva, která neutralizují rod i číslo. Adjektiva a verba, jež se shodují se středním rodem v množném čísle, používají unifikované tvary. Místo tvarů středního rodu se používají tvary mužského neživotného rodu a ženského rodu.

Jmenné tvary přídavných jmen v přísudku mají zpravidla knižní ráz. Do prostředředlovacího stylu také řadíme nekodifikované, dnes převážně nářeční tvary, morfologizované jevy hláskoslovné *-ej* namísto *-ý*.

2.2.3 Zájmena

Časté je v běžnědorozumivacím stylu použití zájmena *ten* i složené výrazy zájmenné podoby bez opakování podstatného jména.

2.2.4 Číslovky

Stylově rozlišovány jsou některé dublety číslovek, respektive jsou to varianty číslovek *dva*, *tři* a *čtyři*.

Číslovka *dva* má jako jediná kodifikovaný tvar *-ma* ve 3. a v 7. pádě. Tvary s *-mi* v tomto případě nemají stylovou hodnotu, může se jednat jen o nevědomé dodržování shody nebo o snahu dodržovat spisovnost, protože koncovka *-ma* uživateli signalizuje nespisovnost (Čechová et al., 2008).

U číslovek *tři* a *čtyři* ve druhém pádě jsou dnes kodifikovány také tvary *-ech*. V 2. pádě tak dochází k dubletě tvarů *tří* a *třech*, kdežto v 6. pádě zůstává pouze *třech*. V běžné mluvě se dubletní tvary z 2. pádu mylně používají i v 6. pádu ze snahy spisovného vyjádření.

U těchto číslovek se používá kodifikovaná forma *-mi* v 7. pádě. Ve spojení s podstatnými jmény označujícími částí lidského těla se však používá duálová koncovka *-ma* (Internetová jazyková příručka, 2017).

Do běžnědorozumivací sféry z nářečí proniká kvantitativní podoba *-í* do 1. pádu, vztahuje-li se číslovka k životnému substantivu.

2.2.5 Slovesa

Slovesa jsou frekventovaná v prostředředlovacím stylu a dodávají textu dynamiku.

Rozlišují se slohové varianty starší, které postupně ustupují a nabývají platnost knižní, a mladší, prostředky hovorové a neutrální.

Slovesa 3. třídy používají pro 3. osobu přítomnosti neutrální koncovku *-í*, *-í*, kdežto tvary zakončené na *-u*, *-ou* jsou hovorové. Hovorové tvary se vyskytovaly nejen v mluveném projevu, ale i v uměleckých pracích, publicistice a umělecké

literatuře. Díky tomu u těchto sloves probíhá proces neutralizace. Domácí slova označující všední děj s koncovkou *-u* lze považovat za základní.

Slovesa 1. třídy, která mají koncovku *-i*, *-í*, jsou slohově příznaková.

Infinitivní koncovka *-ct* je oproti koncovce *-ci* neutrální. Podobné je to u infinitivní koncovky *-ti*, která je knižní a neutrální je dnes pouze *-t*.

Trpný rod z hlediska stylové hodnoty má zvrtný tvar neutrální a hovorový charakter. Opisný tvar má charakter knižní.

Podmínka splnitelná v minulosti se vyjadřuje kondicionálem minulým, ale v běžných projevech se vyjadřuje kondicionálem přítomným.

Přechodníky, přítomný a minulý, se v prostěsdělovacím stylu nevyskytují, stejně tak i tvary předminulého času.

2.2.6 Příslovce

Příslovce nahrazují víceslovná spojení příslovečná, což vede k ekonomičnosti vyjádření a tím i k zvýšenému užívání nejen v běžné mluvě, ale i v publicistice.

Jak v běžné mluvě, tak v publicistice se vyskytují příslovečné okazionalismy (příležitostná pojmenování), které se lexikalizují a frazeologizují.

2.2.7 Předložky

Často se zaměňují předložkové pády *mimo* a *kromě*. Předložka *mimo* se pojí se 4. pádem a předložka *kromě* se užívá s 2. pádem.

K záměně dochází i u předložek *díky* a *kvůli*, které se hojně zaměňují, protože si uživatel jazyka nedokáže uvědomit správný význam slova. Předložka *díky* vyjadřuje příčinu něčeho pozitivního, žádoucího. Kdežto předložka *kvůli* vyjadřuje něco negativního, nežádoucího.

Z hlediska frekvence dochází k pohybu v užití některých předložek. Zejména se jedná o hojně používanou předložku *na* na úkor předložek *v*, *k*, *pro* a jiných.

2.2.8 Spojky

Většina spojek je slohově neutrální, avšak některé výrazy jsou příznakové pro určitý slohový typ.

V mluvených projevech se za sebou hromadí více spojovacích výrazů, zatímco v psaných projevech se někdy vypouštějí, aby se zabránilo opakování vztažných výrazů.

Významově synonymní spojky se liší podle frekvence, ale spíše se uplatňuje individuální zvyk mluvčího. Nelze tedy říct, že spojka *a* se používá v jistých typech textu a spojka *i* zase v jiných (Čechová et al., 2008).

2.2.9 Částice

V běžné řeči je projev prokládán tzv. výplňkovými slovy (tak, no tak, ...) a jinými částicemi a zvuky.

Příznačné pro projevy v běžnědorozumívacím stylu, které jsou neveřejné neoficiální, tedy jde o denní komunikaci, je také partikulizace, tj. přechod různých slov, slovních spojení, popř. vět k částicím.

2.2.10 Citoslovce

Citoslovce nejsou vhodné v žádných typech textů vyjma soukromého projevu běžné mluvy. V uměleckých textech se citoslovce používají k oživení textu.

Tak jako u částic se u i citoslovců setkáme s přechodem slov, slovních spojení a vět k citoslovcům. Tento proces se nazývá interjekcionalizace. Nejčastěji k tomuto jevu dochází u vokativu substantiv nebo substantivní spojení.

2.3 Lexikum

Slovní zásoba jazyka je bohatší než jiné jazykové roviny. Je to dáno nejen počtem jednotek, které tvoří slovník jazyka, ale i vztahy mezi nimi. Základní jednotka slovní zásoby mající formu i význam se nazývá lexém.

V rámci lexikálního významu se rozlišuje složka pojmová a pragmatická. Pojmová složka udává, že plnovýznamový lexém označuje určitý denotát. Je možné ho vztáhnout k nějakému referentu. Neplnovýznamový lexém může mít pojmovou složku až ve spojení s autosémantickými slovy. Pragmatická složka lexikálního významu ukazuje na to, jaké stanovisko k výrazu uživatel zaujímá. Pragmatické složky významu jsou méně stálé a podléhají např. dobovým změnám a hodnotícím postojům.

Lexém, který je spisovný, frekventovaný, stylově a dobově nezabarvený a neexpresivní se pokládá za lexém nepříznakový (neutrální). Tyto lexémy jsou v textech všestranně využívány. Ostatní lexémy jsou příznakové. Příznakové lexémy mohou patřit do stylové vrstvy určitých typů textu, ale jejich užití je limitováno a ustáleno.

Lexikální stavba v prostěsdělovacím stylu je charakterizována takzvanou vrstvou kolokvialismů. Kolokvialismy jsou hovorová slova a frazémy, která stojí na pomezí spisovnosti, jsou to slova řadící se i do obecné češtiny (Jedlička et al., 1970). Jedná se o výrazy, které vznikly mimo spisovný jazyk, ale postupně do něj pronikly.

Slovní zásoba je charakteristická preferencí slov s malým obsahem a velkým rozsahem. Hojně se používají částice, citoslovce, zájmena a kontaktní slova. Popisný slohový postup používá vyšší podíl substantiv a adjektiv, což posiluje

věcnost popisu (Bečka et al., 1992). V prostěsdělovacích projevech se nepoužívají knižní výrazy, zatímco neologismy jsou běžné.

Projevy náležející do běžně dorozumívací oblasti jsou realizovány buď běžně mluveným jazykem, nebo spisovným jazykem, a to i formou hovorovou. Někteří mluvčí užívají jazykového útvaru místně, územně omezeného, dialektu, obecné češtiny či městské mluvy. (Jedlička et al., 1970)

Běžně dorozumívací projevy jsou realizovány různými formami národního jazyka, kam patří i hovorová čeština spisovná. Hovorovou češtinou rozumíme stylovou variantu uvnitř spisovného jazyka, nikoli formy užívané v stylové oblasti běžně dorozumívací. (Jedlička et al., 1970)

V popisu pracovního postupu je zapotřebí vyznačit následnost dějů. K tomu slouží zájmenná adverbia nebo přesnější časové údaje. Autor manuálu si může zvolit všeobecný podmět, podmět neosobní nebo podmět v 1. osobě jednotného čísla. Obecně je popis vyjádřen jako osobní zkušenost. (Bečka et al., 1992)

2.4 Syntax

Syntax se zabývá skladbou věty, to znamená jejími formálními prostředky výstavby, zapojením jednotek do jazykového a situačního kontextu a vztahy mezi sémantickou a formální stránkou. Syntax dává velké možnosti pro stylizaci, protože na úrovni usouvztažňování pojmenovacích jednotek je otevřena cesta k tvořivosti. Stylizaci textu umožňují stylémy, které utváří styl.

Prostěsdělovací styl je charakterizován převážně jako běžná denní komunikace. Je to oblast spontánních mluvených sdělení. Většina dialogů je přímým kontaktem malého počtu recipientů a psanost je v nich jen minimální. Projev je pevně spojen s komunikační situací a pro pochopení textu je důležitá i nonverbální komunikace, jako je mimika, gesta atd.

Konstitující vlastností prostěsdělovacího komunikátu je mluvenost. V současnosti se i tento styl začíná postupně dostávat do psané podoby, jako jsou například osobní dopisy, SMS, e-maily a podobně. Tyto psané projevy jsou spontánní, kontaktné a převážně soukromé. (Čechová et al., 2008)

Podle členitosti se věty rozlišují na typy dvojčlenné a jednočlenné (Jedlička et al., 1970). Dvojčlenná věta svojí mluvnickou stavbou jazykově vyjadřuje formu myšlení. Tyto věty jsou časté v psaných projevech odborného stylu. Vedle dvoučlenných vět se v hovorovém stylu vyskytují i jednočlenné věty jmenné.

Věty jednočlenné se dělí na jmenné a slovesné. Jednočlenná věta jmenná je prostředkem citového zabarvení výpovědi. Její mluvnická stavba má jeden základní větný člen, nemůže být rozvinutá a má nižší sdělnou hodnotu. Pro svoji stručnost, zkratkovost, neúplnost a neurčitost vyjádření a citovou zabarvenost se jednočlenné jmenné věty uplatňují nejen v běžné komunikaci, kde jsou přirozené,

ale i v uměleckých a publicistických textech. Jednočlenná věta slovesná je slohově neutrální, ale může mít i určité zabarvení.

V pevně zakotvených projevech může v konkrétní situaci vznikat neúplné, eliptické vyjadřování. Vznikají tak slovní a větné elipsy a neúplné věty, časté zejména v dialozích.

V prostěsdělovacím stylu se využívají věty jednoduché a souvětí jak souřadná, tak podřadná. Souvětí jsou v mluveném projevu jednodušší než v psaném, tam se používají spíše složitější věty.

U popisu děje se občas objevují i nevlastní věty vztažné s absolutním *co*. Pravidelně se opakující děje jsou spjaté a vyjadřují popisnou linii výčtovou. (Bečka et al., 1992)

2.5 Shrnutí

Jak můžeme vidět, v různých bodech se vlastnosti textu prostěsdělovacího stylu a popisného slohového postupu rozchází, protirečí si nebo jsou opomíjené. Proto si jednotlivá fakta shrneme v následující kapitole a na základě zaměření textu si vymezíme nejdůležitější zásady při psaní manuálu.

3 Tvorba textu manuálu

Jak jsme již zmínili manuál může být čistě psaný nebo doprovázený obrázky. Čistě obrázkové manuály (manuály obsahují schémata) ve své práci pomineme, protože neobsahují žádný text. V této kapitole si vyčleníme jednotlivá doporučení k tomu, aby byl manuál pochopitelný pro každého uživatele.

3.1 Předpoklad

Když chceme psát nějaký text, začínáme s předpokladem. Předpokládáme, že text bude obsahovat jisté body, předpokládáme jeho délku a zejména pak to, že text budou číst čtenáři ze stejného oboru jako pisatel.

Tyto předpoklady jsou potřebným prvním krokem, bez nich bychom nikdy nezačali psát. Manuály jsou psány převážně tak, aby byly srozumitelné čtenářům, kteří se orientují ve stejném oboru, kam spadá i popsaný produkt. Například máme návod na počítačový software (program), v textu převažují počítačové termíny a zkratky srozumitelné pouze lidem, kteří se tímto odvětvím zabývají. To samé platí i pro kuchařský návod, kde jsou specifické termíny pro různé druhy nádobí a podobně.

Na tomto místě bychom chtěli upravit předpoklad odbornosti čtenáře. Nemůžeme předpokládat, že návody budou číst pouze osoby, které je potřebují v nějakém pracovním odvětví či zaměření. Musíme zvážit možnost, že se manuál dostane do rukou laika či samouka, který bude chtít produkt využívat.

3.2 Morfologická stránka

Při psaní manuálu bychom měli převážně používat autosémantické slovní druhy. Jsou to slova, která jsou plnovýznamová. K plnovýznamovým slovním druhům patří podstatná jména, přídavná jména, slovesa, příslovce a někdy se k těmto slovním druhům řadí i zájmena a číslovky. Naproti tomu synsémantika nesou význam až ve spojení s autosémantikou. K neplnovýznamovým slovním druhům patří předložky, spojky, citoslovce a částice.

Tvary plnovýznamových slovních druhů by se v návodu měly objevovat ve spisovné formě s rodovou shodou. Slovesa dodávají textu dynamiku. Pro zachování čtivosti a srozumitelnosti je možné u popisu postupu použít i více hovorových tvarů sloves. Pasivní konstrukce lze použít, zatímco přechodníky a podmiňovací způsob se do takového textu nehodí.

Žádný text se neobejde bez předložek a spojek, ale není vhodné jich používat příliš, jinak bude text hůře čitelný. Také tyto slovní druhy nesmějí být přehlíženy, aby byl text srozumitelný. Částice a citoslovce se v psaném textu moc neobjevují. Citoslovce se však často používají v obrázkových manuálech, kde označují

nějaký zvuk. Jak víme, obrázkové manuály neobsahují mnoho textu, a proto jsou zdlouhavé slovní popisy nahrazeny stručnými citoslovci.

Po všech těchto vymezeních bychom neměli zapomínat i na pravopisná pravidla. Musíme se vyhnout překlepům a psát gramaticky správně.

3.3 Lexikální stránka

Co se týče slovní zásoby, musíme dbát na to, aby zvolené výrazy byly v celém návodu soudržné. Preferujeme slova s malým obsahem a velkým rozsahem.

Text by měl obsahovat spíše plnovýznamové lexémy, které jsou nepříznačkové. Příznakové jevy jsou pro manuál nutností, ale neměly by být používány nadměrně. Charakteristické je užití kolokvialismů, které stojí na pomezí spisovnosti, a neologismů. Měli bychom se vyvarovat užití knižních výrazů, které jsou v návodu zcela nemístné. Rovněž bychom se měli vyvarovat terminologii, která se může čtenáři zdát složitá.

Návod jsme zařadili do prostěsdělovacího stylu a popisného slohového postupu. Obě tyto stylové oblasti jsou specifické svojí hovorovou češtinou, dokonce i dialekty, obecnou češtinou a městskou mluvou. I přesto, že manuál zařazujeme do těchto oblastí, měli bychom používat převážně spisovný jazyk. Do spisovného jazyka patří i hovorový jazyk. Vyhýbat se musíme obecné češtině a teritoriálním dialektům. Tyto prvky nejsou v manuálu vhodné.

Podle Bečky si pisatel může zvolit všeobecný podmět, podmět neosobní nebo podmět v 1. osobě singuláru. Doporučujeme použít spíše 2. osobu plurálu. Užití 1. osoby říká, jak činnost provádí sám pisatel, kdežto použití 2. osoby v imperativu vyjadřuje jistý příkaz, který pokud se neudělá, nebude věc fungovat. Množné číslo použijeme na důkaz zdvořilosti a nikoliv z předpokladu, že manuál budou číst dvě a více osob.

3.4 Syntaktická stránka

Skladba vět ve vymezené oblasti má charakter jednoduchých a neúplných vět. V návodu by tyto věty být neměly. Naopak bychom měli používat věty jednoduché a souvětí jak souřadná, tak podřadná, která nejsou příliš rozvinutá, tj. jednoduchá souvětí.

Texty by měly být stručné, protože pokud bude text příliš rozsáhlý je pravděpodobné, že jej uživatelé nebudou číst. Se stručností rovněž souvisí i srozumitelnost. Pokud text bude natolik stručný, až nebude pochopitelný, naše práce byla zbytečná. Takové texty nikdo číst nebude. Jednotlivé činnosti by měly na sebe navazovat a tvořit souvislý text, který nepřeskakuje důležité činnosti.

4 Doporučení

V kapitole výše jsme podrobněji sepsali jednotlivá doporučení. Pro lepší přehlednost je shrneme v jednotlivých bodech.

Domníváme se, že manuál by měl po morfologické stránce:

- obsahovat plnovýznamová slova
- být gramaticky správný
- vyvarovat se užívání částic a citoslovcí

Manuál by měl z lexikální stránky:

- užívat 2. osobu množného čísla
- používat spisovnou češtinu

Ze syntaktické stránky by text měl:

- být stručný
- být srozumitelný
- být souvislý
- navazovat v jednotlivých činnostech
- používat jednoduché věty a souvětí

5 DEBDict

DEBDict je nástroj platformy DEB, vyvinutý RNDr. Adamem Rambousekem, Ph.D., na Fakultě informatiky Masarykovy univerzity (FI MU), poskytující přístup ke slovníkům českého jazyka, morfologickému analyzátoru češtiny *ajka*, českému WordNetu (umožňujícím vyhledávat sémantické sítě) a CIA World Factbooku (v němž nalezneme všechny země světa) (Rambousek et al., 2010a). Zpočátku byl nástroj doplňkovou aplikací internetového prohlížeče Mozilla Firefox (rozšíření prohlížeče již není aktuální). Z rozšíření Firefoxu se stala webová aplikace.

Z podporovaných slovníků si můžeme vybrat mezi *Akademickým slovníkem cizích slov*, *Slovníkem spisovné češtiny*, *Příručním slovníkem jazyka českého*, *Slovníkem spisovného jazyka českého*, *Slovníkem české frazeologie a idiomatiky I–IV* (rozdělený na slovesnou a neslovesnou část) a *Slovníkem českých synonym* (Rambousek et al., 2010b). Uvedené zdroje jsou dostupné pro všechny registrované uživatele aplikace.

5.1 Výkladové slovníky

5.1.1 Příruční slovník jazyka českého

Příruční slovník jazyka českého (PSJČ) je slovníkem velkého rozsahu skládajícím se z osmi dílů v devíti svazcích, protože čtvrtý díl je rozdělen do dvou svazků. Obsahuje 250 000 hesel a jeho rozsah je celkem 10 824 stran. Nikdy nebyl pojat jako kodifikační, jde o deskriptivní (popisný) slovník, který se zaměřuje na upevnění spisovné slovní zásoby. PSJČ představuje podrobný lexikografický popis slovní zásoby. Stavba hesel vychází ze základního významu, zachycuje polysémii hesla a obsahuje propracované definice významů. Další podstatnou částí hesla jsou doklady (především z beletrie, která byla u dobrých autorů považována za nositele kvalitního spisovného jazyka) uváděné z velké části formou citátů. Snaha o ujasnění spisovnosti a nespisovnosti přispěla k tomu, že tento slovník přinesl první významný pokus o stylistické hodnocení lexikálních jednotek.

Přípravné lexikografické práce na slovníku byly započaty ve 20. letech 20. století. Získané materiály byly ověřovány se slovníky jiného druhu, např. překladovými slovníky Josefa Jungmanna. Práce se zúčastnili přední čeští jazykovědci a na tvorbě hesel se podílelo 26 autorů. Přípravné práce skončily v roce 1934 a u příležitosti stého výročí vydání překladového slovníku Josefa Jungmanna vyšel v únoru 1935 první sešit *Příručního slovníku jazyka českého*. Dále byly vydávány knižní verze.

Slovník vydávalo Státní nakladatelství v Praze v letech 1935–1957. Do elektronické verze jej převedl Ústav pro jazyk český AV ČR ve spolupráci s Fakultou

informatiky Masarykovy univerzity v roce 2007.

Hlavními redaktory výše uvedených svazků byli akademik Bohuslav Havránek, univ. prof. dr. Oldřich Hujer, prof. PhDr. Emil Smetánka, univ. prof. dr. Vladimír Šmilauer, prof. PhDr. Miloš Weingarta a dr. Alois Získal (*Příruční slovník jazyka českého*, 2008).

5.1.2 Slovník spisovného jazyka českého

Slovník spisovného jazyka českého (SSJČ) obsahuje 192 908 zpracovaných slov ve čtyřech svazcích, a proto se řadí ke slovníkům středního typu. Svým rozsahem je menší než PSJČ a zároveň plní funkci prakticky využitelného slovníku, tudíž jde o slovník kodifikační (jak po stránce gramatické a pravopisné, tak i výslovnostní).

Přípravné práce započaly v 50. letech 20. století. Slovník se opíral o materiály získané při práci na PSJČ a zachycuje terminologii. Menší rozsah vedl k hutnosti heslového odstavce, proto nedokládá význam slov, ale jen typická slovní spojení. SSJČ přináší podrobnou stylistickou charakteristiku daných slov. Slovník naznačuje slovesné vazby a zachycuje vidové dvojice. Na rozdíl od PSJČ je slovník hnízdován, to znamená, že do heslového odstavce zařazuje zdobněliny, předponová slovesa, k přídavným jménům příslovce atd.

Kromě prvního knižního vydání (1960–1971) byl zároveň vydán v sešitech (1958–1971), vyšlo 40 sešitů. Ve čtvrtém díle jsou doplňky 984 nově zařazených slov a opravy. V roce 1989 slovník vyšel ve druhém vydání, tentokrát v osmi svazcích. Na elektronické podobě se podílel Ústav pro jazyk český AV ČR s Fakultou informatiky Masarykovy univerzity v roce 2002.

Hlavními redaktory byli akademik Bohuslav Havránek, PhDr. Jaromír Bělič, DrSc., PhDr. Miloš Helcl, CSc., prof. dr. Alois Jedlička, prof. dr. Václav Křístek, CSc., a prof. PhDr. František Trávníček. Rukopis slovníku připravili členové lexikografického oddělení Ústavu pro jazyk český ČSAV (Havránek, 1989).

5.1.3 Slovník spisovné češtiny

Slovník spisovné češtiny pro školu a veřejnost (SSČ) je slovník s menším rozsahem a kodifikačním statutem. Obsahuje 45 366 heslových slov a 62 872 vyznačených významů. Každé heslo obsahuje poučení o pravopisu (také o skloňování a časování) a výslovnosti slov. Výběr materiálu se přesouvá (na rozdíl od PSJČ a SSJČ) k publicistice. Zvláštní pozornost pak věnuje dubletním tvarům. Ke slovníku jsou připojeny přílohy podávající přehled o tvoření slov v češtině, soupisy rodných jmen a příjmení a zeměpisných jmen, jakož i běžných zkratk a značek.

Počátek práce na SSČ datujeme do 70. let. Tento slovník byl zamýšlen a také vydán jako jednosvazkový. První vydání vyšlo v roce 1978, druhé opravené a doplněné vydání pak v roce 1994, třetí opravené vydání 2003 a další vydání byla

v letech 2004, 2005 a 2009. Nakladatelství LEDA připravilo elektronické vydání v roce 2005.

Hlavními redaktory slovníku byli PhDr. Josef Filipec, CSc., prof. PhDr. František Daneš, DrSc., PhDr. Jaroslav Machač (1. vydání) a Vladimír Mejstřík, prom. ped. (2. a 3. vydání) (Mejstřík et al., 2003).

5.2 Slovníky současného jazyka

5.2.1 Akademický slovník cizích slov

Akademický slovník cizích slov (ASCS) je speciální normativní slovník, jenž obsahuje 100 000 významů slov, citátových spojení a běžných zkratk a značek cizího původu, je rozdělený do dvou dílů.

Jednotlivá slova uvádějí poučení o pravopisu, výslovnosti a významu, dále pak sdělují původ slova, tvaroslovnou a slovnědruhovou charakteristiku, stylové zařazení a kontextové užití. Slovník zachycuje dynamiku ve vývoji české slovní zásoby, jejího obohacování, posunu ve stylistickém hodnocení slov přejatých a významové změny. Reaguje na zásadní proměny odehrávající se v rozvíjející domácí slovní zásobě a hlavně na ty, které se vztahují ke konkrétním vědním oborům, disciplínám a řemeslům.

V roce 1995 vyšla knižní verze v nakladatelství Academia, posléze v roce 1997 vyšel dotisk. V novějších vydáních byl přejmenován na *Nový akademický slovník cizích slov* (NASCS). V roce 1999 vznikla v nakladatelství LEDA elektronická verze 1.0 s názvem *Velký slovník cizích slov* (VSCS) a v roce 2005 vyšel slovník ještě jednou jako verze 2.0, která už není součástí nástroje platformy DEB.

ASCS napsal kolektiv autorů pod vedením doc. PhDr. Věry Petráčkové, CSc., a prof. PhDr. Jiřího Krause, DrSc. (Petráčková et al., 1997).

5.2.2 Slovník české frazeologie a idiomatiky

Slovník české frazeologie a idiomatiky I–IV (SČFI) popisuje českou frazeologii. Slovník je dílem vypracovaným týmem lingvistů z Filozofické fakulty Univerzity Karlovy a v menší míře i Ústavu pro jazyk český AV ČR (Čermák et al., 2009). Dílo usiluje o všestranný popis variant frazémů, které jsou užívány v současné době.

Slovníkové heslo podrobně člení kontext užívání frazému od jeho významu a funkce přes příklady, synonymní nebo opozitní řady až po přímý odkaz k onomaziologickému slovníku. Frazémy s vyšší frekvencí mají uvedené ekvivalenty ve čtyřech jazycích. Každý díl je vybaven odbornou studií té oblasti, již popisuje.

První díl slovníku se jmenuje *Přirovnání* a věnuje se nevětným výrazům, tedy materiálu majícímu v sobě spojku jak/jako. Do doby vydání tohoto slovníku ne-

byla přirovnání jako celek zpracována. První díl poprvé vyšel v roce 1983. Druhý díl, nazývaný se *Výrazy neslovesné*, se zabývá několika oblastmi frazeologie, a to výrazy jmennými, adverbialními a všemi typy gramatickým frazémů. První výtisk proběhl v roce 1988. Třetí díl s titulem *Výrazy slovesné* zahrnuje všechny slovesné frazémy, které jsou zároveň kolokačními frazémy, nezahrnují frazémy, jež mají v sobě sloveso. Můžeme v něm listovat od roku 1994. Poslední čtvrtý díl *Výrazy větné* se orientuje na ustálené výrazy s podobou věty. Tento díl je založený na *Českém národním korpusu* (ČNK). Vyšel v roce 2009 a tehdy byla zároveň první tři díla doplněna o řadu hesel z ČNK. Všechna díla slovníku byla ve stejném roce vydána elektronicky nakladatelstvím LEDA, které vydávalo i tištěné verze. V DEBDictu je použita verze přímo od autora.

Autory slovníku jsou prof. PhDr. František Čermák, DrSc., doc. PhDr. Jiří Hronek, CSc., a PhDr. Jaroslav Machač (Čermák et al., 2009).

5.2.3 Slovník českých synonym

Slovník českých synonym (SČS) je praktická příručka obsahující něco málo přes 20 000 heslových slov. Je určena spíše pro rodilé mluvčí, kteří dokáží rozlišit nepatrné významové rozdíly mezi slovy.

Heslová slova jsou řazena abecedně a u každého jsou uvedeny jeho synonymní výrazy, tj. slova a slovní spojení se stejným či podobným významem, které lze libovolně nahradit. Synonymní řady začínají nejobecnějším a neutrálním slovem, vedle něho se pak řadí podobné a blízké výrazy. Tvary sloves se uvádějí s valencemi a pokud možno v obou videch. U frekventovaných slov jsou uváděny jejich antonyma.

Slovník poprvé vyšel v roce 1994. O dva roky později, tj. v roce 1996, nakladatelství Lidové noviny vydalo druhé opravené vydání a 2000 vydalo třetí doplněnou verzi slovníku. V roce 2004 nakladatelství zveřejnilo elektronickou verzi tohoto slovníku. Nástroj platformy DEB používá verzi přímo od autora.

Rukopis slovníku lektoroval a k vydání doporučil doc. PhDr. František Čermák, DrSc. (Pala a Všíanský, 1996). Autory jsou doc. PhDr. Karel Pala, CSc., a Jan Všíanský.

5.3 Český WordNet

WordNet je lexikální databáze, která uspořádává jazykové jednotky podle sémantických vztahů. Heslem této databáze je synset neboli synonymická řada (pokud se dvě slova objeví vedle sebe, jsou synonymní). Jedno slovo může mít i více významů, tj. může být synonymní s více slovy, v tom případě je dané slovo polysémické.

WordNet tvoří dvě roviny. Synsety jsou jednou a jejich provázání je druhou rovinou. Až jejich sémantické vztahy tvoří celkový WordNet. Některé příklady sémantických vztahů jsou synonymie, antonymie, meronymie a holonymie, hyperonymie a hyponymie.

Verze 2.1 použitá v DEBDictu obsahuje 117 597 synonymických řad, z čehož 81 426 tvoří synsety podstatných jmen, 13 650 synsety sloves, 18 877 synsety přídavných jmen a 3664 synsety příslovcí (Čapek, 2006).

5.4 Morfologický analyzátor *ajka*

Morfologický analyzátor ajka byl vyvinut na Fakultě informatiky MU Mgr. Radkem Sedláčkem, Ph.D., podle popisu segmentace českých slov z disertační práce doc. PhDr. Kláry Osolsobě, Dr. Základem algoritmu je členění slova na kmenový základ, intersegment a koncovku. Morfologický analyzátor pracuje se strojovým slovníkem kmenových základů, definičním souborem koncovkových množin a vzorů. (Sedláček, 1999)

Analyzátor je efektivní jak při morfologické analýze, tak při generování správných gramatických tvarů. *Ajka* musí pro každou odchylku (nepravidelnost) vytvořit zvláštní vzor, což vedlo k vytvoření nového analyzátoru *majka*, který není součástí DEBDictu. Současný počet analyzovaných tvarů je 1839 vzorů. (Husárová, 2015)

5.5 CIA World Factbook

První verze *World Factbook* byla publikována knižně v roce 1980 a obsahovala 165 zemí na 225 stranách. Od té doby vychází téměř každý rok a byla přejmenována na *The World Factbook* (CIA, 2013). První verze obsahovala informaci o počtu obyvatelstva, o jeho hospodářství, vládě, vodstvu, komunikaci a o obranných silách. Téměř do každého vydání jsou přidávány nové státy (nově vzniklé rozpadem mocností) a výše uvedené položky obměňovány. V roce 1997 společnost *Central Intelligence Agency* (CIA) představila první elektronickou verzi *Factbooku*. Použitá verze v DEBDictu je z roku 2009. V současné době obsahuje přes 250 států, včetně velkých vodních ploch.

6 Manuál pro DEBDict

6.1 Přístup k obecnému prohlížeči slovníků

DEBDict je nástroj, který umožňuje prohlížet slovníky v elektronické formě. Tato možnost vám dává příležitost podívat se do slovníků, aniž byste museli chodit do knihovny.

Pro přístup je nutné, abyste měli své vlastní přihlašovací údaje. Sice můžete použít tzv. demo verzi, ovšem nebudete mít přístup ke všem základním zdrojům, které jsou zpřístupněny registrovaným uživatelům.

Pro získání vlastních údajů, a tudíž plnému přístupu, musíte vyplnit formulář (NLPCentrum) (do kolonky: „Co budete využívat“ napište DEBDict), který následně vytisknete, podepíšete a odešlete na adresu Masarykovy univerzity. Je možné oskenovaný formulář zaslat i elektronicky e-mailem. Poté stačí počkat na přidělení uživatelských údajů a poté se směle pustit do užívání prohlížeče slovníků.

Prohlášení pro dočasné uživatele programů a dat Centra zpracování přirozeného jazyka (NLPCentrum) při Fakultě informatiky Masarykovy univerzity

 English/anglicky

Jméno a příjmení:

E-mail:

Organizace/pracoviště:

Adresa:

Co budete využívat:

Po vyplnění formuláře stiskněte , podepíšte jej a pošlete na adresu uvedenou níže.

Potvrzuji svým podpisem, že souhlasím s níže uvedenými podmínkami užívání programů a dat (korporusů, slovníků, ...) NLPCentra:

- Při využívání dat získaných z NLPCentra nebo z výstupů programů NLPCentra uvedu jako jejich zdroj příslušný produkt NLPCentra.
- Citovaný výstup nesmí být delší než 200 slov souvislého textu.
- Kopii (nebo elektronickou podobu) práce, ve které byla použita data či výstupy z produktů NLPCentra, zařadu po jejím uveřejnění k archivaci do NLPCentra.
- Přístup k programům a datům je umožněn pouze pro vědecké a výzkumné účely; a nelze je bez předchozí smlouvy používat ke komerčním účelům.
- Právo užívat programy a data NLPCentra je nepřenosné.
- NLPCentrum při Fakultě informatiky Masarykovy univerzity si vyhrazuje právo kdykoliv přístup k programům a datům omezit či zrušit.

v dne

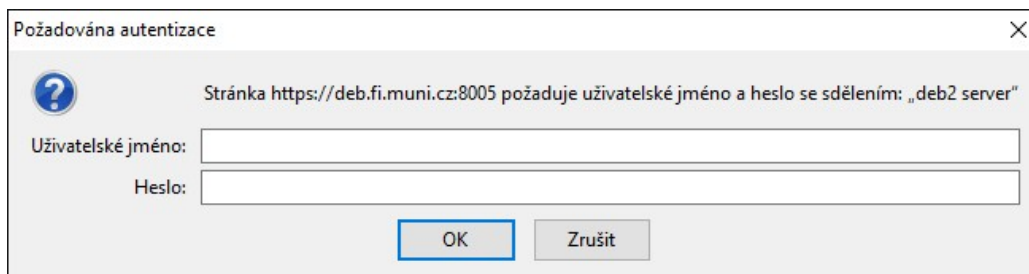
Obrázek 1: Formulář pro přístup k programům z NLPCentra

6.2 Přihlášení

Původní rozšíření prohlížeče Mozilla Firefox již nefunguje, s touto změnou odpadá nutnost instalace. K přístupu k prohlížeči slovníků je třeba použít webovou aplikaci, do níž se přihlásíte pod svými přihlašovacími údaji, získanými zasláním prohlášení, viz kapitola 6.1.

Otevřete si kterýkoli internetový prohlížeč (nemusí se jednat pouze o Mozillu Firefox). Najed'te na stránky nástroje DEBDict (Rambousek, 2010c). Při načítání

bude aplikace požadovat napsání uživatelského jména a hesla. V demo verzi je uživatelské jméno i heslo *demo*. Po jejich zadání se objeví titulní strana.



Obrázek 2: Přihlášení

Po přihlášení se zobrazí hlavní strana programu. V horním levém rohu, pod lištou záložek, máte rozbalovací nabídku s aktuálním vybraným elektronickým zdrojem, vedle něhož se nachází dotazovací řádek (textové pole) k hledání požadovaného slova. Pod nabídkou se slovníky je připraveno pole pro seznam nalezených slov. Zbytek místa v nástroji je pro vyhledaná heslová slova.

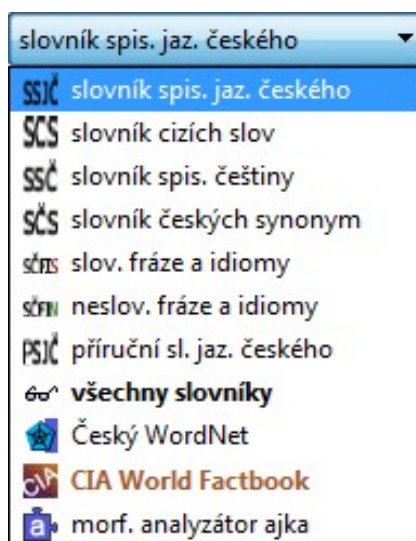


Obrázek 3: Hlavní strana programu

6.3 Dotaz

6.3.1 Volba elektronických zdrojů

Při přihlášení je hned na začátku vybrán jako výchozí *Slovník spisovného jazyka českého*. Pro volbu jiného slovníku musíte rozkliknout rozbalovací nabídku v levém horním rohu a vybrat si jiný slovník, popřípadě jiný dostupný elektronický zdroj.

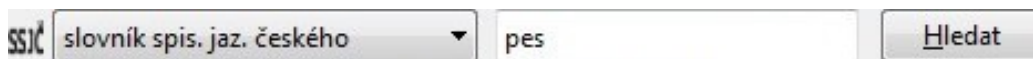


Obrázek 4: Výběr slovníku

6.3.2 Typ dotazu

Dotaz do dotazovacího řádku napište v základním tvaru, tedy jako byste jej hledali ve slovníku, tzn. jména budou v prvním pádě jednotného čísla a slovesa v infinitivu. To platí nejen pro slovníkovou část, ale i pro zbytek dostupných databází.

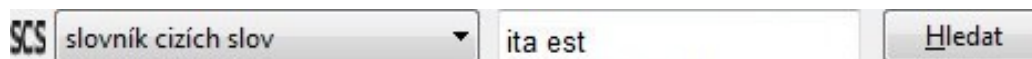
Veškeré slovníky (vyjma SCS), WordNet a morfologický analyzátor *ajka* mají zcela stejné vyhledávání. Například potřebujete-li vyhledat ve slovníku SSJČ význam slova *pes*, napište hledané slovo do dotazovacího řádku následovně:



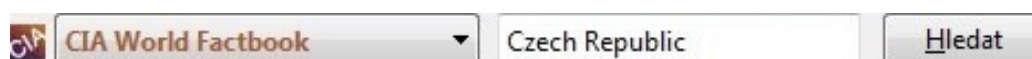
Obrázek 5: Ukázka dotazu v SSJČ

Při výběru *Slovníku cizích slov* i *CIA World Factbooku* bude dotazované slovo taktéž v prvním pádě čísla jednotného. Navíc výraz musí být v příslušném jazyce,

příčemž země hledáte pouze v angličtině (protože se jedná o anglický program). U SCS jako příklad vyzkoušíte vyhledat latinské sousloví *ita est* a u *Factbooku* Českou republiku pod anglickým názvem *Czech Republic*.



Obrázek 6: Ukázka dotazu v SCS



Obrázek 7: Ukázka dotazu v CIA World Factbook

Poté klikněte na políčko *Hledat* a aplikace pošle dotaz serveru², který vám na výstup zobrazí hledaný výraz. U všech zdrojů probíhá vyhledávání stejně.

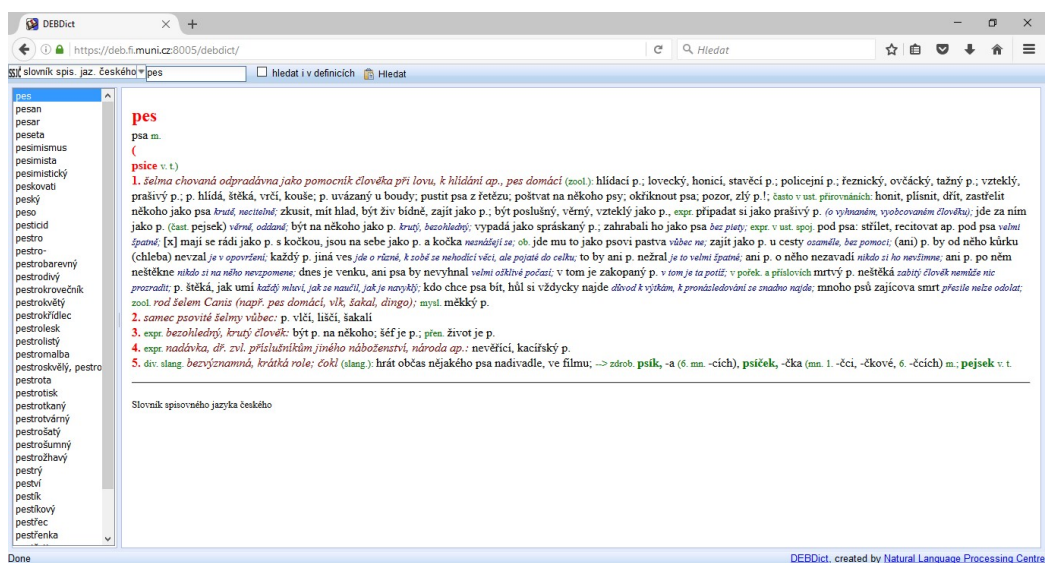
6.4 Výstup

6.4.1 Výstup hledání ve slovnících

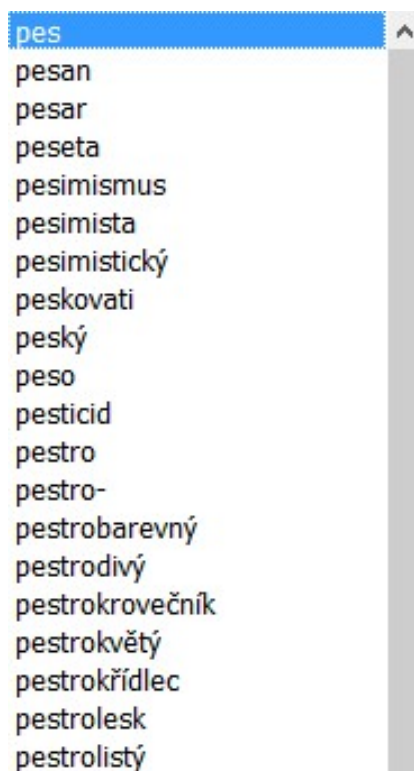
Jednotlivé slovníky

Výstup ve slovníkové části vypadá totožně u kteréhokoli slovníku vyjma obsahu. Obrázek č. 8 ukazuje pohled na celou obrazovku. Pod nabídkou se slovníky je obdélníkové pole se seznamem nalezených slov, která mají na začátku výrazu hledaný pojem. (Se seznamem se v doplňkové aplikaci mohlo dále pracovat, uložit a následně provádět analýzy. Ve webové aplikaci to již není možné.) Zbytek místa je určen pro definice ze slovníku.

²Každý slovník má jinou strukturu a server za vás vyřeší, pro který slovník je slovo určeno. Proto si nemusíte pamatovat jednotlivá dotazování ke každému slovníku.



Obrázek 8: Výstup vyhledávacího slova *pes* v SSJČ



Obrázek 9: Seznam slov v levém poli

pes

psa m.

(

psice v. t.)

1. *šelma chovaná odpradávná jako pomocník člověka při lovu, k hlídání ap., pes domácí (zool.):* hlídací p.; lovecký, honiči, stavěcí p.; policejní p.; řeznický, ovčácký, tažný p.; vztekdlý, prašivý p.; p. hlídá, štěká, vrčí, kouše; p. uvázaný u boudy; pustit psa z řetězu; poštvat na někoho psy; okřiknout psa; pozor, zlý p.!, často v ust. přirovnáních: honit, plísňit, dít, zastřelit někoho jako psa *krutě, nechtělně*; zkusit, mít hlad, být živ bičně, zajít jako p.; být poslušný, věrný, vztekdlý jako p., *expr. připadat si jako prašivý p. (o vyhlášeném, vyobcovaném člověku)*; jde za ním jako p. (*čast. pejsek*) *věrně, oddaně*; být na někoho jako p. *krutě, bezohledně*; vypadá jako spráskaný p.; zahrabali ho jako psa *bez piety*; *expr. v ust. spoj. pod psa: stílet, recitovat ap. pod psa velmi špatně*; [x] mají se rádi jako p. s kočkou, jsou na sebe jako p. a kočka *nasnájejí se*; ob. jde mu to jako psovi *pastva vůbec ne*; zajít jako p. u cesty *osaměle, bez pomoci*; (ani) p. by od něho kůrku (chleba) *nevzal je v opovržení*; každý p. jiná ves *jde o různé, k sobě se nehodící věci, ale pojeté do celku*; to by ani p. *nežral je to velmi špatně*; ani p. o něho *nezavádí někdo si ho nevíme*; ani p. po něm *neštěkne někdo si na něho nezapomene*; dnes je venku, ani psa by *nevyhnal velmi ošklivé počasí*; v tom je *zakopaný p. v tom je ta potě; v poček. a příslovích mrtvý p. neštěká zabíjí člověk nemůže nic prozradit*; p. štěká, jak umí *každý mluví, jak se naučil, jak je navyklý*; kdo chce psa bit, hůl si *vždycky najde důvod k výtčám, k pronásledování se snadno najde*; mnoho psů *zajicova smrt přestě nelze odolat*; zool. rod šelem *Canis (např. pes domácí, vlk, šakal, dingo); myst. měkký p.*
2. *samec psovitě šelmy vůbec: p. vlčí, liščí, šakal*
3. *expr. bezohledný, krutý člověk: být p. na někoho; šéf je p.; přen. život je p.*
4. *expr. nadávka, dř. zvl. příštušníkům jiného náboženství, národa ap.: nevěřící, kacírský p.*
5. *div. slang. bezvýznamná, krátká role; čokl (slang): hrát občas nějakého psa nadvadle, ve filmu; -> zdrob. psík, -a (6. mn. -cích), psiček, -čka (mn. 1. -čci, -čkové, 6. -čcích) m.; pejsek v. t.*

Slovník spisovného jazyka českého

Obrázek 10: Definice slova *pes*

ita est

(1) kniž. tak jest, souhlasí, ano

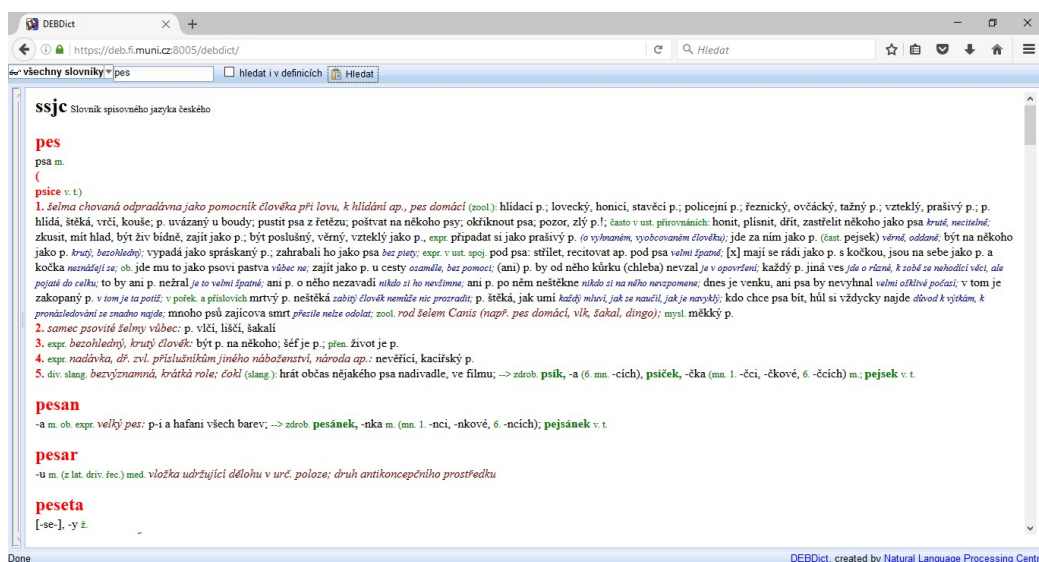
Slovník cizích slov

Obrázek 11: Definice latinského sousloví *ita est*

Všechny slovníky

U zvolení nabídky *všechny slovníky* server vypíše ze všech slovníků a českého WordNetu nejen hledaný výraz, ale i další slova, která v jednotlivých slovnících naleznete v seznamu na levém okraji.

Např. jste hledali význam slova *pes* ve všech slovnících a na výstup jste dostali u SSJČ kromě hledaného výrazu i pojmy jako *pesan*, *pesar*, *peseta* a *pesík*. Aplikace nevytvořila seznam, místo toho je vypsala do stejné úrovně s hledaným slovem. Pokud je v některém slovníku umožněn výběr z několika slov utvořených z hledaného výrazu, dostanete na výčet pouhých pět.



Obrázek 12: Výstup vyhledávacího slova *pes* ve všech slovnících

6.4.2 Výstup hledání v dalších dostupných zdrojích

Český WordNet

Výstup WordNetu je poněkud jiný než výstup slovníkové části. Nedostanete klasickou definici hledaného pojmu. Na obrázku č. 13 můžete vidět výsledek hledání v českém WordNetu.

Nahoře máte síť slov, kterou si můžete představit jako strom synsetů, v jehož kořeni je nejobecnější pojem a dále v listech ty nejspecifičtější (Pala a Ševeček, 1999). Při vyhledávání slova *pes* je nejspecifičtější pojem *psovitá šelma:1* a přes jednotlivé výrazy se postupně dostáváte k nejobecnějšímu pojmu *entita:1*. Síť slov může obsahovat i synonymní synset. Obrázek č. 13 obsahuje synonymní synset

organismus:1, bytost:1, forma života:1, což znamená, že *organismus* ve významu *1* je synonymní s výrazem *bytost* ve významu *1* i s *formou života* ve významu *1*.

Pod sítí slov jsou obecné informace ke slovu, jako jsou synonyma a slovní druh. Novinkou pro WordNet bylo udělení jedinečného identifikačního čísla (ID), které sloužilo pro vytváření ekvivalencí mezi synsety cizích jazyků. ID se začalo používat až ve WordNetu 1.5. Dalším vylepšením bylo rozdělení synsetů do sad základních pojmů (BCS – base concept set). Základní pojmy jsou vysoko v hierarchii a mají hodně potomků. Prvotní členění bylo na 3 skupiny. První skupinu tvořily subjekty vyznačující se podle způsobů konceptualizace (původ, forma, složení a funkce), druhá skupina se klasifikovala pomocí situačního typu a situační složky a třetí skupinou byly nepozorovatelné problémy, výroky či duševní subjekty (Vossen et al., 1998). Toto členění poprvé použili v EuroWordNetu a později v bylo BalkaNetu) doplněné o necelých 3 700 synsetů. Přesné rozčlenění naleznete na stránkách The Global WordNet Association (Weisscher, 2013). Rozšířené členění se používá v jakémkoliv cizojazyčném WordNetu.

Sémantickou relaci chápeme jako přiřazení hledaného výrazu k hyperonymii (slovo nadřazené) a popř. označuje i holonymii (vztah mezi slovem označujícím celek a slovem označujícím část celku). Hyperonymum k hledanému výrazu *pes* je *psovitá šelma* a holonymum je *smečka* a *rod Vlk* (latinsky *Canis*).

Po sémantické relaci následuje seznam hyponym (slova významově podřízená) a v tomto případě i meronym (části celku). Hyponyma jsou slova významově podřazená. V příkladu jste hledali slovo *pes*, jehož hyponyma jsou *puhl*, *mopslík* atd. Meronyma jsou části celku, slovo *prut* je zde míněno jako ocas, což je část každého psa. *Prut* je tedy část psa.

entita:1>objekt:1>živoucí entita:1>organismus:1, bytost:1, forma života:1>tvor:1, zvíře:1,
živočich:1>strunatec:1>obratlovec:1>savec:1>placentál:1>masožravec:1>psovitá šelma:1

Synonyma: pes:1, psice:1

Slovní druh: **podstatné jméno (n)**

ID: ENG20-02001223-n

BCS: 3

Sémantické relace:

-->> [hypernym] psovitá šelma:1

-->> [holo_member] rod Vlk:1, rod Canis:1

-->> [holo_member] smečka:1

hyponym: 10

<<-- [mero_part] prut:1

<<-- [hyponym] mopslík:1

<<-- [hyponym] hafan:1

<<-- [hyponym] špic:1

<<-- [hyponym] puď:1

<<-- [hyponym] trpasličí pes:1, trpasličí psík:1

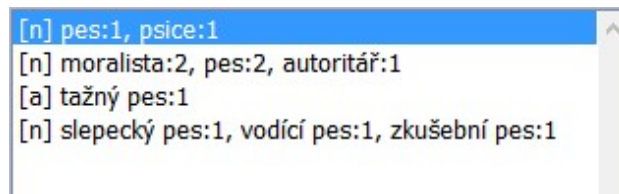
<<-- [hyponym] lovecký pes:1

<<-- [hyponym] pracovní pes:1

<<-- [hyponym] psisko:1, čokl:1

Obrázek 13: Výstup z českého WordNetu

Seznam v levé části nabízí všechny synsety mající hledaný výraz. Tento seznam má na prvním místě zmíněno, o jaký slovní druh se jedná, a není členěn podle sítí slov, nýbrž podle synonym.



Obrázek 14: Seznam v levém poli

Morfologický analyzátor *ajka*

Ajka umožňuje segmentovat slova. Tato funkce je efektivní jak při morfologické analýze, tak při generování správných gramatických tvarů. Výsledek vyhledávání se zobrazí v tabulce, kterou můžete dále rozkliknout na segmentaci slova, kategorii slova a skloňování (značené *).

Při hledání slova *pes* morfologický analyzátor na výstupu nabídl kromě hledaného výrazu i slovo *peso*. Důvod tohoto výběru je popsán níže v odstavci zabývajícím se skloňováním.

Analyzovaný tvar: pes			
Základní tvar	Segmentace	Číslo vzoru	Kategorie
pes (*)	=p=es==	1136-pes	k1gMnSc1
peso (*)	=pes===	232-barvivo	k1gNnPc2

Obrázek 15: Výstup v morfologickém analyzátoru *ajka*

Segmentace výrazu se dělí na prefix, kmenový základ, intersegment, koncovku a postfix. U základního tvaru je segmentace p-es, kde prefix, koncovka a postfix je nulový.³

³Kmenový základ je ten, který zůstává u všech tvarů stejný. Např. u slova *pes* jsou skloňované tvary: *pes*, *psa*, *psovi*, ..., proto kmenový základ je pouze *p-*, které nezměnilo svoji pozici v žádném skloňování.

Popis segmentace

=p=es==

Prefix	Kmenový základ	Intersegment	Koncovka	Postfix
-	p	es	-	-

Segmentuj tvar:

[Morfologická analýza - interaktivní režim](#) [Morfologická analýza - dávkový režim](#)

Obrázek 16: Segmentace slova *pes*

Kategorie je složena ze zkratk gramatických kategorií (GK). Při rozkliknutí dostanete podrobný popis kategorie, kde zkratky (nazvané *Hodnoty*) jsou podrobněji popsány. Malé písmeno vždy značí GK a velké písmeno, popřípadě číslo, určuje realizaci GK. Zkratky jsou použity kvůli úspornosti a přehlednosti. Kdyby tabulka obsahovala celistvý výpis gramatických kategorií, nebyla by přehledná.

Jak můžete vidět na obrázku č. 17, podrobný popis zkratky *k1* je následující. Písmeno *k* označuje GK slovní druh a *1* je realizací prvního slovního druhu, a to podstatného jména (substantivum). Totéž platí i u nečíselné hodnoty *nS*, kde malé *n* přiřazuje slovo ke GK číslo a velké *S* realizuje singulár, tedy jednotné číslo.

Popis značky

k1gMnSc1

Hodnota	Gramatická kategorie	Popis
k1	Slovní druh	Substantiva
gM	Rod	Mužský životný
nS	Číslo	Singulár
c1	Pád	Nominativ

Popiš značku:

[Morfologická analýza - interaktivní režim](#) [Morfologická analýza - dávkový režim](#)

Obrázek 17: Značka slova *pes*

Na obrázku č. 18 a č. 19 je rozkliknuté skloňování obou slov, které demonstruje, proč analyzátor uvedl výběr mezi slovem *pes* a *peso*. Tvar *pes* se nachází jak v prvním pádě jednotného čísla u výrazu *pes*, tak i v druhém pádě množného čísla u slova *peso*. Tudíž pokud se hledaný tvar objeví i v některém odvozeném tvaru jiného slova, automaticky jej *ajka* vydá na výstup.

Odvozené tvary ke slovu "pes"

Rod: Mužský životný			
Pád	Singulár	Pád	Plurál
1	pes	1	psi, psové
2	psa	2	psů
3	psu, psovi	3	psum , psům
4	psa	4	psy
5	pse	5	psové , psi
6	psu, psovi	6	psech, psách
7	psem	7	psy, psama

Obrázek 18: Skloňování slova *pes*

Odvozené tvary ke slovu "peso"

Rod: Střední			
Pád	Singulár	Pád	Plurál
1	peso	1	pesa
2	pesa	2	pes
3	pesu	3	pesům
4	peso	4	pesa
5	peso	5	pesa
6	pesu	6	pesech, pesách
7	pesem	7	pesy, pesama

Obrázek 19: Skloňování slova *peso*

CIA World Factbook

Elektronický zdroj CIA slouží jako atlas informací k jednotlivým státům. Do dotazovacího panelu jste napsali *Czech Republic* (česky Česká republika) a server vyhledal veškeré známé informace o České republice. Jelikož se musíte ptát anglicky, jsou i vyhledaná data v angličtině. Kromě informací, které jsou popsány v kapitole 5.5, můžete vpravo nahoře vidět vlajku dané země a dole mapu, kde se stát geograficky nachází.

Czech Republic



[map](#)

Introduction

Background Following the First World War, the closely related Czechs and Slovaks of the former Austro-Hungarian Empire merged to form Czechoslovakia. During the interwar years, the new country's leaders were frequently preoccupied with meeting the demands of other ethnic minorities within the republic, most notably the Sudeten Germans and the Ruthenians (Ukrainians). After World War II, a truncated Czechoslovakia fell within the Soviet sphere of influence. In 1968, an invasion by Warsaw Pact troops ended the efforts of the country's leaders to liberalize Communist party rule and create "socialism with a human face." Anti-Soviet demonstrations the following year ushered in a period of harsh repression. With the collapse of Soviet authority in 1989, Czechoslovakia regained its freedom through a peaceful "Velvet Revolution." On 1 January 1993, the country underwent a "velvet divorce" into its two national components, the Czech Republic and Slovakia. The Czech Republic joined NATO in 1999 and the European Union in 2004.

•

•

•

Transnational Issues

Disputes - international In February 2005, the ICJ refused to rule on the restitution of Liechtenstein's land and property assets in the Czech Republic confiscated in 1945 as German property; individual Sudeten Germans seek restitution for property confiscated in connection with their expulsion from Czechoslovakia after World War II; Austrian anti-nuclear activists have revived blockades of the Czech-Austrian border to protest operation of the Temelin nuclear power plant in the Czech Republic

Illicit drugs transshipment point for Southwest Asian heroin and minor transit point for Latin American cocaine to Western Europe; producer of synthetic drugs for local and regional markets; susceptible to money laundering related to drug trafficking, organized crime



Obrázek 20: Výstup z CIA World Factbook

6.5 Rozšířené vyhledávání

U obecného prohlížeče slovníků máte dvě možnosti rozšířeného vyhledávání. První možnost umožňuje vyhledat přesně hledaný výraz a druhá použije retrográdní seřazení.

Volbu, kdy potřebujete vyhledat přesné znění hledaného výrazu, většinou použijete při hledání ve výběru *všechny slovníky*. Jak již bylo zmíněno, při této volbě se vám budou mezi hledaný pojem plést i další slova, která mají na začátku výrazu hledané spojení. Pokud se chcete těchto slov zbavit, stačí před slovo v dotazovacím řádku napsat !.



Obrázek 21: Rozšířené vyhledávání přesného tvaru

Jak můžete vidět, na obrázku č. 22 je vyhledané slovo *pes* ve *všech slovnících* bez rušivých elementů. Tento výraz můžeme použít i na jednotlivé slovníky, kde je výklad stejný, jen s malým rozdílem, a to tím, že v levém postranním sloupci bude vypsán pouze hledaný výraz.

SSC Slovník spisovné češtiny

pes

. psa m

1. *šelma ochočená k hlídání, lovu ap.* hlídač, lovecký, ovčácký pes, pozor, zlý pes!, pustit psa z řetězu,

2. *samec psovitě šelmy liščí pes*, je to pes, nebo fenka?,

3. *expr. bezohledný, krutý člověk být (na někoho) pes*, život je pes, [x] ani pes po něm neštěkne, *expr. je zapomenut*; být na někoho jako pes, zlý; být poslušný, věrný jako pes, honit, zastřelit někoho jako psa, *bez citu*; každý pes jiná ves, (*pořek.*) (*o nesourodých věcech*); (ani) pes by od něho kůrku nevezal, *expr. je v opovržení*; pes, který štěká, nekouše, (*pořek.*); práce ap. pod psa, *hanl. velmi špatná*; vypadat jako spráskaný pes, *být schliplý n. unavený*; (hledat, kde je) zakopaný pes, *hlavní potíž, překážka*; (venku je,) že by ani psa nevyhnal, *velmi špatně počasí*;

psice - z *fena (svno)*.

psík -a

psiček -čka

pejsek -ska m (mn. 1. -ci, -kově, 6. -cich) zdrob. *expr.*

psjc Příruční slovník jazyka českého

pes, *psa* m. *domácí šelma chovaná člověkem k hlídání, k lovu a pod.* Za zavřenými vraty ozval se štěkot psa. *Jir.* Děti i dva velicí psové, všecko vyběhlo přede dveře, vítat babičku. *Něm.* Kmet staropražský, Přivítan jménem, musel třikrát nosit prašivého psa po celém náměstí. *Pal.* Tu se Pyskáček rozložil a poštal na něj domácího velkého psa. *Erb.* Dělalí pány chodice s puškou na rameni a se psem stavěcím v patách. *Baar.* Potkal zuřivého ovčáckého psa. *Herb.* Při kramářských karách zaštěkal hlídač pes. *Win.* *Ustál sp.* Karabiny byly rezavé jako pes. *Ros.* Ale cakra, cakra to nádoby, zednický nádoby, mám já už taky hlavu, aby ji hodil psu *nestojí za nic.* *Čap. Ch.* Život je krušný, lidé v něm zkusi jako psi. *Šmil.* Na mou duši, pracovat budu jako pes, dostanu-li se tam. *Klost.* Mám hlíd jako pes *velký.* *J. Šir.* Ohlíží se jako mlsný pes. *Prasek.* Jako pes věrná [je Katka]. *Jir.* Pravice jeho se zdvihla a rozkazně ukázala ku dveřím. *Poslušná* jako pes zdvihla jsem se z pohovky. *Ner.* Všichni jej znali mírného, pokorného jako psa po výprasku. *R. Svob.* Bartoň jako zbitý pes šoural se ke dveřím *bážílivě.* *Hol.* Musíš pamatovat na sebe, jako pes živ být nemůžeš *biďně.* *Rais.* Zahynu tu někde jako pes, sipal jsem, skřípaje zuby *biďně, bez pomoci.* *Svob.* Zabít ho chtěli jako psa. *Jir.* Zastřelte mě jak psa, jestli věrným nebudu. *Don.* Řekněte mi, proč jsou ti šefové na člověka jako psi *krutí.* *Maň.* Což mohu jít, když na mne voláš jako na psa *hrubě, neslušně?* *R. Svob.* Litáš celý den po závodě, od dělníka k dělníkovi, a honíš je do práce jako psy *krutě, necitelně.* *Svob.* Někdy si na něj také houknu, a to jde, jakby ho pes kousl, ani nemukne. *Něm.* Hube sem dostal jak

Obrázek 22: Výstup vyhledávání přesného tvaru

Retrográdní řazení je takové řazení, v němž se slova neřadí podle abecedy od začátku, ale naopak od konce slova. Museli byste procházet celý slovník, a to by zabralo mnoho času. Ovšem v DEBDictu taková slova můžete snadno nalézt jednoduchým způsobem. Stačí před slovo napsat znaménko -.

Obrázek 23: Retrogradní řazení

Vyhledaná slova jsou *herpes*, *ostropes* a původně hledané slovo *pes*. U jednotlivých slovníků to funguje stejně jako u obyčejného vyhledávání, jen jsou slova na výstupu seřazena od konce.

herpes

-etu m. (6. j. -u) (z řec.) med. opar

ostropes

-psu m. statná bylina s listy ukončenými silným ostnem a s nafialovělými květy; bot. rod *Onopordon*: o. trubil

pes

psa m.

(

psice v. t.)

1. šelma chovaná odpraďadna jako pomocník člověka při lovu, k hlídání ap., *pes domácí* (zool.): hlídací p.; lovecký, honič, stavčí p.; policejní p.; řeznický, ovčácký, tažný p.; vztekly, prašivý p.; p. hlídá, šteká, vrčí, kouše; p. uvázaný u boudy; pustit psa z řetězu; poštvat na někoho psy; okřiknout psa; pozor, zlý p.!; často v ust. přirovnáních: honit, plisnit, dřít, zastřelit někoho jako psa *krutě, necitelně*; zkusit, mít hlad, být živ bídně, zajít jako p.; být poslušný, věrný, vztekly jako p., expr. připadat si jako prašivý p. (o vyhraněném, vyobcovaném člověku); jde za ním jako p. (čast. pejsek) *věrně, oddaně*; být na někoho jako p. *krutě, bezohledně*; vypadá jako spráskaný p.; zahrabali ho jako psa *bez přetv.*; expr. v ust. spoj. pod psa: strílet, recitovat ap. pod psa *velmi špatně*; [x] mají se rádi jako p. s kočkou, jsou na sebe jako p. a kočka *nezná její se.* ob. jde mu to jako psovi pastva *vůbec ne*; zajít jako p. u cesty *osaměle, bez pomoci*; (ani) p. by od něho kůrku (chleba) *nevzal je v opovržení*; každý p. jiná ves *jde o řízení, k sobě se nehodící věci, ale pojaté do celku*, to by ani p. *nežral je to velmi špatně*; ani p. o něho *nezavádí nikdo si ho nevímeš*; ani p. po něm *neštěkne nikdo si na něho nevzpomene*; dnes je venku, ani psa by *nevyhnal velmi ošklivé počasí*; v tom je *zakopaný p.* v tom je *ta potíže*; v potek. a příslovích *mrtvý p. neštěká zabítý člověk nemůže nic prozradit*.

p. šteká, jak umí *každý mluví, jak se naučil, jak je navykly*; kdo chce psa *bít, hůl si vždycky najde důvod k výtčám, k pronásledování se snadno najde*; mnoho psů *zajícova smrt přestle nelze odolat*; zool. rod

šelem Canis (např. *pes domácí, vlk, šakal, dingo*); mysl. měkký p.

2. samec *psovitě šelmy vůbec*: p. vlčí, liščí, šakali

3. expr. *bezohledný, krutý člověk*: být p. na někoho; šéf je p.; *přem.* život je p.

4. expr. *nadávka, dř.* zvl. *přislušníkům jiného náboženství, národa ap.*: nevěřící, kacířský p.

5. div. slang. *bevyzýnamná, krátká role*; *čokl* (slang): hrát občas nějakého psa *nadivadle*, ve filmu; -> zřáb. *psik*, -a (6. mn. -cích), *psiček*, -čka (mn. 1. -čci, -čkové, 6. -čcích) m.; *pejssek* v. t.

Obrázek 24: Výsledek retrogradního řazení

7 Deriv

Webové rozhraní Deriv (Šmerk, 2010) je nástroj vyvinutý na FI MU RNDr. Pavlem Šmerkem, Ph.D., a Centrem zpracování přirozeného jazyka ve spolupráci s lingvisty. Tento nástroj umožňuje zpracovávat derivační vztahy v českém jazyce. Je propojen s již uvedeným DEBDictem a korpusem SYN2000 obsahujícím 100 milionů textových slov.

Nástroj vyhledává v morfologickém analyzátoru *ajka* lemmata a podle pravidel (morfologická značka, úvodní a koncový řetězec) vytváří seznamy slov, která jsou kandidáty na slova utvořená stejným způsobem (patří k jednomu derivačnímu typu). Se seznamy je možné dále pracovat nebo je upravovat, ukládat a třídit do různých souborů, s nimiž se také může manipulovat. Existuje i možnost prohlížet je ve dvou módech buď jako prosté seznamy, nebo seznamy s uvedením frekvencí vyhledaných jednotek (Osolsobě, 2008).

8 Manuál pro Deriv

8.1 Přístup k webovému rozhraní Deriv

K přístupu je nutné, abyste měli své vlastní přihlašovací údaje. Pro získání údajů, musíte vyplnit formulář (NLPCentrum) (do kolonky: „Co budete využívat“ napíšete Deriv). Formulář následně vytisknete, podepíšete a odešlete na adresu Masarykovy univerzity. Je možné oskenovaný formulář zaslat i elektronicky e-mailem. Poté stačí vyčkat na přidělení uživatelských údajů.

Jak můžete vidět, přístup k jakémukoliv nástroji zveřejněnému Centrem zpracování přirozeného jazyka při FI MU, je stejný, liší se pouze uvedením požadovaného nástroje v kolonce „Co budete využívat“.

Prohlášení pro dočasné uživatele programů a dat Centra zpracování přirozeného jazyka (NLPCentrum) při Fakultě informatiky Masarykovy univerzity

 [English/anglicky](#)

Jméno a příjmení:
E-mail:
Organizace/pracoviště:
Adresa:
Co budete využívat:

Po vyplnění formuláře stiskněte , podepíšte jej a pošlete na adresu uvedenou níže.

Potvrzuji svým podpisem, že souhlasím s níže uvedenými podmínkami užívání programů a dat (korpusů, slovníků, ...) NLPCentra:

- Při využívání dat získaných z NLPCentra nebo z výstupů programů NLPCentra uvedu jako jejich zdroj příslušný produkt NLPCentra.
- Citovaný výstup nesmí být delší než 200 slov souvislého textu.
- Kopii (nebo elektronickou podobu) práce, ve které byla použita data či výstupy z produktů NLPCentra, zašlu po jejím uveřejnění k archivaci do NLPCentra.
- Přístup k programům a datům je umožněn pouze pro vědecké a výzkumné účely a nelze je bez předchozí smlouvy používat ke komerčním účelům.
- Právo užívat programy a data NLPCentra je nepřenositelné.
- NLPCentrum při Fakultě informatiky Masarykovy univerzity si vyhrazuje právo kdykoliv přístup k programům a datům omezit či zrušit.

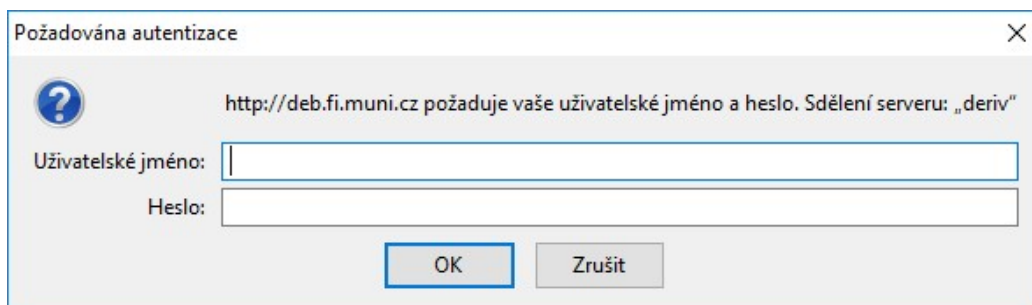
V dne

Obrázek 25: Formulář pro přístup k programům z NLPCentra

8.2 Přihlášení

Deriv je webové rozhraní, proto odpadá instalace. Stačí pouze požádat o přístupové heslo správce Derivu podle postupu popsaného v kapitole 8.1.

Otevřete si kterýkoli internetový prohlížeč. Najedte na stránky webového rozhraní Deriv (Šmerk, 2010). Při načítání bude aplikace požadovat zadání uživatelského jména a hesla. Do patřičných kolonek zadejte své údaje. Po kliknutí na tlačítko *OK* se načte titulní strana.



Obrázek 26: Přihlášení

Titulní strana Derivu je velice jednoduchá. V levé horní části, pod adresou stránky, je řádek s možností vyhledávání a s prací s uloženým obsahem, který vás po rozkliknutí navede na vámi vybranou položku. Na tomtéž řádku na konci (úplně vpravo) je uživatelské jméno s položkou *odhlásit se*. Zbylé místo je prázdné.



Obrázek 27: Titulní strana Derivu

8.3 Dotaz

8.3.1 Volba vyhledávání

Nejdříve si musíte vybrat, jaký typ hledání chcete používat. Máte na výběr mezi hledáním slovních tvarů, dvojic, trojic a n-tic, kde za n můžete dosadit číslo od 1 do 9. Je zbytečné používat poslední volbu na n-tice tvořené jedním až třemi slovy, protože tyto volby jsou přednastavené v předchozích volbách.

hledání slovních tvarů, dvojic, trojic, -tic

Obrázek 28: Volba vyhledávání

8.3.2 Typ dotazu

Dotaz v Derivu je složitější, protože nestačí napsat hledaný výraz do dotazovacího řádku. Ve webovém rozhraní se místo slov používají značky a regulární

výrazy (RE). U značek se používají tagy k definování hledaného výrazu. Tag je morfologická značka přiřazená ke každému slovnímu tvaru. Tyto tagy jsou popsány v kapitole 8.3.3 a regulární výrazy v kapitole 8.3.4.

Slovní tvar

Při výběru *slovních tvarů* budete hledat výrazy skládající se z jednoho slova, které odpovídá tagu a regulárnímu výrazu. Nemusíte používat obě položky hledání. Tagy jsou nutné, proto se jim nevyhnete, ale nemusíte specifikovat RE. Tuto možnost využijete u takových slov, která nemají gramatické kategorie, jako citoslovce a zkratky. Nespecifikovat regulární výraz můžete i u dalších slovních druhů, ovšem výsledky budou mít mnoho jednotek nebo se stane, že tag bude tak obecný, že program nevyhledá žádné výrazy. Jako příklad bez specifikování RE si můžete zkusit vyhledat předložky.

Hledání

značka RE

hledat i mezi hovorovými tvary (přesněji: tvary se značkou obsahující atribut w, v naprosté většině ale jen s hodnotou H, byť asi 2000 tvarů má jinou hodnotu)

hledat i mezi vlastními jmény (přesněji: tvary obsahujícími i jiný znak, než jen malé písmeno, v naprosté většině ale jen s velkým písmenem na začátku, byť v datech je třeba i AIDS, PhDr., SMSka, tie-breacích atp., celkem asi 1200 tvarů)

uložit do souboru ukládat namísto lemmat přímo slovní tvary

Obrázek 29: Hledání předložek

Na ukázkou použití RE můžete vyhledat slova, která odpovídají tagu *k1gMnSc1* a regulárnímu výrazu *^ná.*í\$*. Popis tagu je následující. Hledáte podstatná jména rodu mužského životného v čísle jednotném a v prvním pádě, která začínají na *ná-* a končí na *-í*.

Hledání

značka RE

hledat i mezi hovorovými tvary (přesněji: tvary se značkou obsahující atribut w, v naprosté většině ale jen s hodnotou H, byť asi 2000 tvarů má jinou hodnotu)

hledat i mezi vlastními jmény (přesněji: tvary obsahujícími i jiný znak, než jen malé písmeno, v naprosté většině ale jen s velkým písmenem na začátku, byť v datech je třeba i AIDS, PhDr., SMSka, tie-breacích atp., celkem asi 1200 tvarů)

uložit do souboru ukládat namísto lemmat přímo slovní tvary

Obrázek 30: Hledání slov začínajících na *ná-* a končící na *-í*

Po specifikování tagu, popř. regulárního výrazu, si můžete zvolit, zdali chcete hledat i mezi hovorovými tvary nebo vlastními jmény. Pokud máte zájem mezi nimi vyhledávat, stačí na příslušný čtverec kliknout.

Hledání se musí vždy uložit do souboru, proto je nezbytné, abyste zapsali název souboru, do něhož hledaný obsah uložíte. Možnost *vyhledat* příslušné tvary vyhledá a uloží je do zvoleného souboru. Možnost *vyhledat a otevřít* rovnou vyhledaná slova zobrazí na výstup.

1	<input type="checkbox"/>	ad
2	<input type="checkbox"/>	během
3	<input type="checkbox"/>	bez
4	<input type="checkbox"/>	beze
5	<input type="checkbox"/>	blízko
6	<input type="checkbox"/>	cestou
7	<input type="checkbox"/>	daleko
8	<input type="checkbox"/>	dík
9	<input type="checkbox"/>	díky
10	<input type="checkbox"/>	dle
11	<input type="checkbox"/>	dnem
12	<input type="checkbox"/>	do

Obrázek 31: Výstup vyhledávání předložek

1	nádhončí
2	náhončí
3	nákupčí
4	námluvčí
5	návodčí

Obrázek 32: Výstup vyhledávání mužská slova začínající na *na-* a končící na *-í*

N-tice

Volba *n-tic* umožňuje vyhledávat *n-tice* slov. Totéž platí i u *dvojic* a *trojic*, proto tato část kapitoly bude platit pro zbylé typy vyhledávání s menším rozdílem při vyhledávání. Rozdíl je ten, že vyhledá dvojice slov a trojice slov.

Na následujícím příkladu můžete vyhledat účelová slova končící na *-cí* odvozená od infinitivu. První značka je tag označující slovesa v infinitivním tvaru *k5.*mF*. Druhá značka *k2.*gMnSc1d1* pak určuje přídavné jméno v mužském životném rodu, v jednotném čísle, v prvním pádě a v pozitivu, což je první stupeň přídavného jména. Regulární výraz určuje pouze to, že slovo musí končit na *-cí*. Položky za druhou značkou určují pravidlo, ze kterého se slovo odvodí. V tomto případě chcete u slovesa infinitivní koncovku *-t* nahradit koncovkou *-cí*. Přitom jste dále aplikovali operaci *k*.⁴

1: značka RE

2: značka RE odvodit od 1. slova: nahrazované nahrazující

a přitom aplikovat:

hledat i mezi hovorovými tvary (přesněji: tvary se značkou obsahující atribut *w*, v naprosté většině ale jen s hodnotou **H**, byť asi 2000 tvarů má jinou hodnotu)

hledat i mezi vlastními jmény (přesněji: tvary obsahujícími i jiný znak, než jen malé písmeno, v naprosté většině ale jen s velkým písmenem na začátku, byť v datech je třeba i *AIDS, PhDr., SMSka, tie-breacích* atp., celkem asi 1200 tvarů)

uložit do souboru ukládat namísto lemmat přímo slovní tvary

Obrázek 33: Hledání účelových slov *-cí*

⁴Po rozkliknutí odkazu „více“ si můžete vybrat ze seznamu operací tu, kterou chcete při derivaci provést. Operace se bude provádět před nahrazením vámi vybrané koncovky za nahrazující určenou koncovku. Za vokál (V) se považuje i dvojnáček **ou** a za konsonant (C) i dvojnáček **ch**.

Program nabízí možnost, jež vyplní formulář podle zkráceného výrazu ve tvaru např. $t\$/k5.*mF > c\acute{i}/k2.*gMnSc1d1/k$. Jednotlivé položky jsou od sebe odděleny znaky $>$ a $/$, jejichž jednotlivé části jsou vyplněny do formuláře. První položka ($t\%$) bude vyplněna do obdélníku nahrazované, druhá ($k5.*mF$) do značky u prvního slova, třetí položka ($c\acute{i}$) do kolonky nahrazující a regulárního výrazu druhého slova, čtvrtá ($k2.*gMnSc1d1$) do značky u druhého slova a pátá položka (k) určuje typ operace při rozkliknutí odkazu „více“. Po kliknutí na *vyplnit* se formulář vyplní stejně jako na obrázku č. 34.

Vyplnit formulář podle

Obrázek 34: Formulář pro zkrácený výraz

8.3.3 Tagy

V následujících tabulkách jsou definovány tagy s popisy značek. (Šmerk, 2006) Malé písmeno vždy značí atribut a velké písmeno nebo číslice realizuje atribut – udává mu hodnotu. Řazení atributů pro každý slovní druh je sepsán v příslušné tabulce. Některé atributy lze vynechat, jiné ovšem ne. Např. u substantiva můžeme vynechat atribut x , w a z , co je však důležité, je atribut g , n a c .

<i>k1</i>	<i>Podstatné jméno</i>
x	Speciální rod
P	půl
g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský
R	rodina (příjmení)
n	Číslo
S	jednotné
P	množné
D	duál
R	hromadné označení členů ro- diny (Novákovi)
c	Pád
1	první pád
2	druhý pád
3	třetí pád
4	čtvrtý pád
5	pátý pád
6	šestý pád
7	sedmý pád
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale
z	Typ tvaru
S	tvar s příklonným -s

k2	<i>Přídavné jméno</i>
e	Negace
A	afirmace
N	negace
g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský
n	Číslo
S	jednotné
P	množné
D	duál
c	Pád
1	první pád
2	druhý pád
3	třetí pád
4	čtvrtý pád
5	pátý pád
6	šestý pád
7	sedmý pád
d	Stupeň
1	pozitiv
2	komparativ
3	superlativ
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale
z	Typ tvaru
S	tvar s příklonným -s

k3	Zájmeno
x	Druh zájmena (x)
P	osobní
O	přivlastňovací
D	ukazovací
T	vymezovací
y	Druh zájmena (y)
F	reflexivní
Q	tázací
R	vztažné
N	záporné
I	neurčité
p	Osoba
1	první osoba
2	druhá osoba
3	třetí osoba
X	první nebo druhá nebo třetí
g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský

n	Číslo
S	jednotné
P	množné
D	duál
c	Pád
1	první pád
2	druhý pád
3	třetí pád
4	čtvrtý pád
5	pátý pád
6	šestý pád
7	sedmý pád
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale
z	Typ tvaru
S	tvar s příklonným -s

k4	Číslovka
x	Druh číslovky (x)
C	základní
O	řadová
R	druhová
G	gramatika
H	gramatika
y	Druh číslovky (y)
N	záporné
I	neurčité
g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský
n	Číslo
S	jednotné
P	množné
D	duál
c	Pád
1	první pád
2	druhý pád
3	třetí pád
4	čtvrtý pád
5	pátý pád
6	šestý pád
7	sedmý pád
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale

t	Terminál v gramatice
A	terminál A
B	terminál B
C	terminál C
D	terminál D
E	terminál E
F	terminál F
I	terminál I
J	terminál J
K	terminál K
L	terminál L
M	terminál M
N	terminál N
O	terminál O
P	terminál P
Q	terminál Q
R	terminál R
S	terminál S
T	terminál T
U	terminál U
V	terminál V
W	terminál W
X	terminál X
Y	terminál Y
Z	terminál Z
z	Typ tvaru
S	tvar s příklonným -s

k5	Sloveso
e	Negace
A	afirmace
N	negace
a	Vid
P	perfektum
I	imperfektum
B	obouvidové
m	Typ (Mód)
F	infinitiv
I	indikativ přítomny
R	imperativ
A	příčestí činné (minulé)
N	příčestí trpné
S	přechodník přítomný (současnost)
D	přechodník minulý (dřívější děj)
B	indikativ futura
p	Osoba
1	první osoba
2	druhá osoba
3	třetí osoba

g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský
n	Číslo
S	jednotné
P	množné
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale
z	Typ tvaru
S	tvar s příklonným -s

k6	Příslovce
e	Negace
A	afirmace
N	negace
g	Rod
M	mužský životný
I	mužský neživotný
N	střední
F	ženský
x	Druh zájmenného příslovce (x)
D	ukazovací
T	vymezovací
M	způsobové
S	stavové
y	Druh zájmenného příslovce (y)
Q	tázací
R	vztažné
N	záporné
I	neurčité
d	Stupeň
1	pozitiv
2	komparativ
3	superlativ
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale
z	Typ tvaru
S	tvar s příklonným -s

k7	<i>Předložka</i>
c	Pád
1	první pád
2	druhý pád
3	třetí pád
4	čtvrtý pád
5	pátý pád
6	šestý pád
7	sedmý pád

k8	<i>Spojka</i>
x	Druhy spojek
C	souřadící
S	podřadící
z	Typ tvaru
S	tvar s příklonným -s

k9	<i>Částice</i>
z	Typ tvaru
S	tvar s příklonným -s

k0	<i>Cítoslovce</i>
-----------	--------------------------

kA	<i>Zkratky</i>
-----------	-----------------------

<i>kY</i>	<i>by, aby, kdyby</i>
m	Vztah ke slovesnému módu
C	kondicionál
p	osoba
1	první osoba
2	druhá osoba
3	třetí osoba
n	Číslo
S	jednotné
P	množné
w	Stylistický příznak tvaru
A	archaismus
B	básnický
C	pouze v korpusech
E	expresivně
H	hovorově
K	knižně
O	oblastně
R	řidčeji
Z	zastarale

8.3.4 Regulární výrazy

Regulární výrazy se používají pro formální definice hledaných jednotek. Ty základní RE jsou uvedeny v následující tabulce:

znak	implementace
.	(tečka) libovolný znak
^	libovolné opakování předcházejícího znaku
\$	konec řetězce
?	předchozí znak (část řetězce vyznačený kulatými závorky) může a nemusí být přítomen
+	libovolné opakování, alespoň jednou
[abc]	jeden z vyjmenovaných znaků
[[:upper:]]	velké písmeno
[[:lower:]]	malé písmeno
[^ abc]	libovolný znak mimo vyjmenované
(a b)	první nebo druhý znak
{n,m}	předcházející znak se má opakovat nejméně n-krát a maximálně m-krát

Kvůli přehlednosti a tím i zkrácení RE byly zavedeny následující zkratky, které umožní ušetřit místo, a zabraňují tak nechtěné chybě při implementaci.

znak	implementace
\$C	libovolný konsonant i s ch
\$V	libovolný vokál včetně ou
\$L	libovolný dlouhý vokál i s ou
\$S	libovolný krátký vokál

Deriv obsahuje i pokročilejší konstrukce, s nimiž se většina uživatelů nesetká, proto zde nebudou vypsány. Tyto pokročilejší konstrukce jsou složité a jejich znaky i funkce jsou sepsány v samotném programu, z něhož jsou převzaty i výše uvedené regulární výrazy.

8.4 Práce s obsahem souborů, se soubory a s adresáři

Nástroj po kliknutí na *vyhledat* automaticky extrahuje příslušná data ze slovníku a uloží je do příslušného souboru, v němž s nimi můžete dále pracovat. Při kliknutí na *vyhledat a otevřít* se extrahovaná data uloží a rovnou otevřou.

8.4.1 Práce s obsahem souborů

Tato možnost nabízí hned několik funkcí, s kterými můžete pracovat. Pokud kliknete na *obsahem souborů*, zobrazí se seznam uložených souborů. Soubory můžete zobrazit, upravit, seřadit retrográdně a vybrat pouze slova označená ručním tříděním.

- Citoslovce (1, 1027)
- k1gMnSc1 (1, 15825)
- ná_i M (1, 5)
- ná_i S (1, 84)
- ná_i Ž (1, 2)
- Předložky (1, 204)
- účelová -cí (2, 1741)
- účelová slova -cí (2, 1911)
- Zkratky (1, 113)

obsah (*jen pro n-tice*) nebo otevřít
 zobrazit seřazené retrográdně vybrat pouze slova neoznačená jako (více poznámek spojit spojkou)

Obrázek 35: Práce s obsahem souborů

Soubor obsahuje v závorce dvě čísla. To první určuje, o jaký typ hledání se jedná. Číslo 1 označuje hledání slovního tvaru, 2 pak dvojici, 3 trojici atd. Druhé číslo je počet vyhledaných výrazů.

Zobrazení s/bez frekvencí

Pokud chcete zjistit, jaká data se objevila na výstupu nástroje, musíte si data v souboru zobrazit. Máte dvě možnosti, jak obsah otevřít. Jednou z nich je otevřít bez frekvencí. Klikněte na kolečko příslušného souboru a poté na políčko *bez frekvencí*. Pokud máte velký soubor, musíte chvíli počkat, než se zpracuje. Na příklad si vyberete soubor s názvem „ná_ní M“. Tento soubor vznikl z hledaného tagu *k1gMnSc1* a regulárního výrazu $\hat{n}á$. *í\$ z kapitoly 8.3.2. Každé slovo má před sebou číslo seřazení podle abecedy a prázdný obdélník. Obdélníkové pole bude vysvětleno později.

1	<input type="checkbox"/>	nádhončí
2	<input type="checkbox"/>	náhončí
3	<input type="checkbox"/>	nákupčí
4	<input type="checkbox"/>	námluvčí
5	<input type="checkbox"/>	návodčí

Obrázek 36: Zobrazení bez frekvence

Deriv umožňuje zobrazit slovo v aplikaci DEBDict. Stačí na příslušné slovo kliknout, např. na náhončí, a budete přesměrováni do DEBDictu. Ovšem, musíte mít své vlastní přihlašovací údaje DEBDict. Postup k získání údajů je totožný s postupem popsáním v kapitole 8.1, jen do kolonky „Co budete používat“ napíšete DEBDict.

ssjc Slovník spisovného jazyka českého

náhončí

-ho m.

1. *honec 2, nadháněč (zvěře): kruh n-ch*

2. (též *nádhončí Havl.*) *zprav. hanl. kdo agitačně pro někoho pracuje n. něco někomu opatřuje; korteš: placený n.; dělat n-ho; n. kapitalistů*

SSC Slovník spisovné češtiny

náhončí

-ho m (2. mn. -ich) *hanl. kdo pracuje pro něčí zájmy agent (syno), 2 náhončí nacionalistických stran,*

Obrázek 37: Zobrazení hesla ve slovníku

Tentýž soubor můžete otevřít i s frekvencemi. Postup je stejný jako u výběru bez frekvencí, jen kliknete na políčko *s frekvencemi*. Pokud soubor obsahuje mnoho výrazů, musíte chvíli počkat, než se zpracují.

1	nádhončí (0, 0, 1, 2)
2	náhončí (11, 23, 335, 595)
3	nákupčí (66, 159, 1839, 3760)
4	námluvčí (0, 1, 9, 19)
5	návodčí (3, 3, 284, 678)

Obrázek 38: Zobrazení s frekvencí

Výstup je skoro stejný jako u předchozího postupu, jen každé slovo navíc obsahuje frekvenci v korpusu SYN2000 (první dvě čísla v závorkách) a v korpusu CzTenTen (poslední dvě čísla v závorkách). I zde platí propojení slova se slovníky v DEBDictu.

Zůstaneme u příkladu *náhončí* (11, 23, 335, 595), kde bylo demonstrováno propojení se slovníky. Frekvence označují počet výskytů v daných slovnících. První dvě čísla 11 a 23 jsou propojena s korpusem SYN2000. První číslo v závorce je počet výsledků korpusového dotazu [word=„náhončí“]. Korpus vyhledá pouze slova ve tvaru uvedeném v uvozovkách.

kteří nedávno bojovali za práva městská , stali se nejhrolivější	náhončí	hofmistroví ; pak to , že právě ti nejbohatší radní
jsi rolničky chromých rukou , tak týrán bezmocí objímat ,	náhončí	vlnek Jakube . . . !) Skloněné to nevadilo
hele , " řekl muž v montgomeráku , " pan	náhončí	. " - " Třeba , " povídám , "
provizorem , courtierem , (fr .) - volební	náhončí	, korteš > svého volebního okruhu ; nikdy neztrácí se
a hrál si na milostpána , maje při tom své	náhončí	, podobné kvítky , jako byl sám . Přistižen na

Obrázek 39: Zobrazení hesla v korpusu SYN2000

Druhé číslo pak obsahuje počet výskytů korpusového dotazu [lemma=„náhončí“]. Tento typ dotazu nehledá ve větách přímo uvedená slova, ale jejich lemmata. Proto výsledek obsahuje i slova v jiných pádech než prvních. U tohoto korpusu není nutná registrace, pokud se spokojíte s jednoduchým vyhledáváním. Pro další funkce, včetně uložení, se musíte zaregistrovat.

Ale , jak tomu bývá , hrabě se pomocí svých	náhončích	dověděl o tomto podniku tak brzy , že včas ještě
kteří nedávno bojovali za práva městská , stali se nejhrolivější	náhončí	hofmistrovi ; pak to , že právě ti nejbohatší radní
jsi rolníčky chromých rukou , tak týrán bezmocí objímat ,	náhončí	vlnek Jakube . . . !) Skloněné to nevadilo
Heideggerem , odporuje - li běžnému mínění o Nietzschem jako	náhončím	(Antreiber) vůle k moci , k politice násilí
lesy prošípovány americkými parašutisty , kteří mě stopují pomocí indiánských	náhončích	a číhají v korunách vysokých borovic (pohupujících se nad

Obrázek 40: Zobrazení hesla v korpusu SYN2000

Druhá dvě čísla 335 a 595 jsou propojena s korpusem CzTenTen. Třetí číslo v pořadí je počet výsledků dotazu [word=„náhončí“] v korpusu CzTenTen.

Čechách s takovým úspěchem, vyslala Nova své náhončí , aby nakoupili práva na všechny možné i
Ona je skutečně jeho hlavní múzou. A že náhončí její - Čas. A nemyslím to morbidně. Naopak
Za deset minut začínáme!" vyhání filmový náhončí komparsisty v barvách Číny a Francie do
zla" a podobně jsou lexikem komunistických náhončí ne z normalizace, ale přímo z pekla 50.
privatizaci a klesla spotřeba, rozjeli se náhončí na obce a přesvědčovali občany aby si topili
této prodejny do spárů profesionálů. Nejsem náhončí , ale rád doporučím někoho, kde vše funguje
300 metrů od vchodu se nás ujal první „ náhončí “. Posadil nás na bus do centra Caracasu
Hrnc. Chtěli jsme do Maracay, tak nás jeden náhončí vzal na cestu, stopl nám bus a týpkovi

Obrázek 41: Zobrazení hesla v korpusu CzTenTen

Poslední, čtvrté číslo je počet výsledků stejného dotazu [lemma=„náhončí“] jako v korpusu SYN2000. Korpus CzTenTen je přístupný pouze registrovaným uživatelům.

Čechách s takovým úspěchem, vyslala Nova své náhončí , aby nakoupili práva na všechny možné i
Ona je skutečně jeho hlavní múzou. A že náhončí její - Čas. A nemyslím to morbidně. Naopak
natě a les i pole hemžili se v tuto dobu náhončimi a myslivci. </p><p> Zvířátka dávno již o
Za deset minut začínáme!" vyhání filmový náhončí komparsisty v barvách Číny a Francie do
zla" a podobně jsou lexikem komunistických náhončí ne z normalizace, ale přímo z pekla 50.
privatizaci a klesla spotřeba, rozjeli se náhončí na obce a přesvědčovali občany aby si topili
Svěřovat se kartáčkám, různým guru a jejich náhončím je hra s ohněm: Vše, co řeknete, může být

Obrázek 42: Zobrazení hesla v korpusu CzTenTen

Upravit soubor

Možnost upravit obsah souboru můžete využít pouze u n-tic, když si myslíte, že je ve slově překlep nebo byste slovo vytvořili jinak atd. Počítejte s tím, že pokud špatně opravíte správné slovo, nemusíte dostat výše uvedené reakce s propojenými aplikacemi.

Upravený soubor uložíte do téhož souboru nebo jej můžete přejmenovat.

1733	žehnat	žehnací
1734	želat	želací
1735	želírovat	želírovací
1736	ženit	ženící
1737	žhavit	žhavicí
1738	žihat	žihací
1739	žilkovat	žilkovací
1740	žlábkovat	žlábkovací
1741	žvýkat	žvýkací

uložit do souboru Kalivodova/ -cí (oprava)

Obrázek 43: Upravit soubor

Retrogradní řazení

Retrogradní řazení je takové řazení, při kterém se slovo neseřazuje podle abecedy od začátku, ale naopak od konce slova.

1	<u>ad</u>	1	<u>chyba</u>
2	<u>během</u>	2	<u>zespoda</u>
3	<u>bez</u>	3	<u>via</u>
4	<u>beze</u>	4	<u>zprostředka</u>
5	<u>blízko</u>	5	<u>vyjma</u>
6	<u>cestou</u>	6	<u>na</u>
7	<u>daleko</u>	7	<u>kontra</u>
8	<u>dík</u>	8	<u>podlevá</u>
9	<u>díky</u>	9	<u>straněvá</u>
10	<u>dle</u>	10	<u>skrzevá</u>
11	<u>dnem</u>	11	<u>podlivá</u>
12	<u>do</u>	12	<u>stranivá</u>
13	<u>doprostřed</u>	13	<u>protivá</u>
14	<u>dovnitř</u>	14	<u>naprotivá</u>
15	<u>chyba</u>	15	<u>zprotiva</u>
16	<u>jménem</u>	16	<u>mezivá</u>
17	<u>k</u>	17	<u>místová</u>
18	<u>ke</u>	18	<u>za</u>
19	<u>kol</u>	19	<u>zeza</u>
20	<u>kolem</u>	20	<u>poza</u>

Obrázek 44: Seřazení podle abecedy a retrogradní seřazení

Ruční třídění

Ruční třídění se používá, když v automaticky extrahovaných datech potřebujete vyznačit buď chybné tagy, nebo potřebujete jen určitý typ slov ze souboru. Např. mezi předložkami chcete vyznačit pouze dvou- a více slabičná slova. Proto si do obdélníku před slovem napíšete poznámku. Většinou to bývá nějaké písmeno. Interpunkční znaménka typu mezera, dvojtečka a čárka jsou ignorována.

1		<u>ad</u>
2	y	<u>během</u>
3		<u>bez</u>
4	y	<u>beze</u>
5	y	<u>blízko</u>
6	y	<u>cestou</u>
7	y	<u>daleko</u>
8		<u>dík</u>
9	y	<u>díky</u>
10		<u>dle</u>
11		<u>dnem</u>
12		<u>do</u>
13	y	<u>doprostřed</u>
14	y	<u>dovnitř</u>
15	y	<u>chyba</u>
16	y	<u>jménem</u>
17		<u>k</u>
18		<u>ke</u>
19		<u>kol</u>
20	y	<u>kolem</u>

Obrázek 45: Ruční třídění

Po vyznačení všech dvouslabičných předložek soubor znovu uložíte, nejlépe s jiným názvem. Tříděný soubor poznáte tak, že v závorce u souboru je třetí položka s dvojtečkou a písmenem. Písmen tam může být víc, podle toho, kolik jich použijete na ruční třídění.

○ Předložky1 (1, 204, :y)

Obrázek 46: Označený soubor z ručního třídění

Zobrazit ručně tříděný soubor můžete dvěma dalšími způsoby, nepočítaje možnost otevření s frekvencí a bez frekvence. První možností je zobrazit výrazy s poznámkou, v tomto případě s poznámkou y.

vybrat pouze slova neoznačená jako

Obrázek 47: Vybrat slova označená jako y

Výše provedená volba vyhledá předložky, které mají jen dvou- a více slabičné výrazy.

1	y	během
2	y	beze
3	y	blízko
4	y	cestou
5	y	daleko
6	y	díky
7	y	doprostřed
8	y	dovnitř
9	y	chyba
10	y	jménem
11	y	kolem

Obrázek 48: Výstup označených slov z ručního třídění s poznámkou

Další způsob zobrazí výrazy bez poznámek, v tomto případě jednoslabičné předložky.

vybrat pouze slova neoznačená jako

Obrázek 49: Vybrat slova neoznačená jako y

1	<input type="checkbox"/>	ad
2	<input type="checkbox"/>	bez
3	<input type="checkbox"/>	dík
4	<input type="checkbox"/>	dle
5	<input type="checkbox"/>	dnem
6	<input type="checkbox"/>	do
7	<input type="checkbox"/>	k
8	<input type="checkbox"/>	ke
9	<input type="checkbox"/>	kol

Obrázek 50: Výstup ozačených slov z ručního třídění bez poznámky

8.4.2 Práce s adresáři

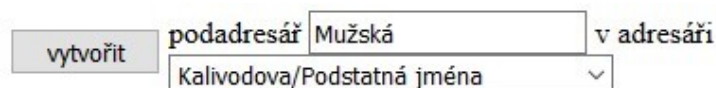
Pokud chcete používat podadresáře, musíte si je nejdříve vytvořit. V horní nabídce klikněte na (práce s) *adresáři*. Jeden adresář máte automaticky vytvořený od chvíle, kdy jste se registrovali, a obsahuje vámi zvolené přístupové jméno. Vytvořit nový podadresář je jednoduché. Stačí napsat název nového adresáře do obdélníkového pole a kliknout na *vytvořit*.

Práce s adresáři

podadresář

Obrázek 51: Vytvořit adresář

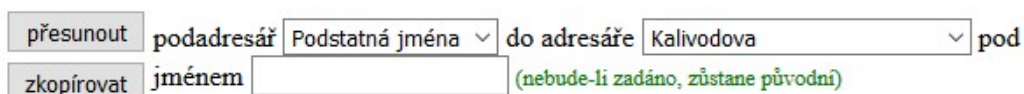
Podadresář pak má tvar „UživatelskéJméno/Podstatná jména“. Můžete vytvářet i podadresáře podadresářů.



vytvořit podadresář Mužská v adresáři Kalivodova/Podstatná jména

Obrázek 52: Tvorba podadresáře

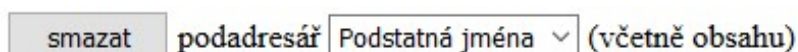
Tvar podadresáře pak bude „UživatelskéJméno/Podstatná jména/Mužská“. Podadresáře lze přesouvat, kopírovat a mazat. Přesouvání a kopírování podadresáře je totožné. Vyberete si, který podadresář chcete přesunout či zkopírovat do kterého adresáře a zdali chcete změnit původní název podadresáře. Pokud chcete nechat původní jméno, kolonka *pod jménem* bude prázdná, když však chcete soubor přejmenovat, stačí do kolonky *pod jménem* napsat nový název souboru. Poté stačí kliknout na *přesunout*, pokud chcete přesunout podadresář, nebo *zkopírovat* pro zkopírování podadresáře.



přesunout podadresář Podstatná jména do adresáře Kalivodova pod jménem (nebude-li zadáno, zůstane původní)

Obrázek 53: Přesouvání a kopírování podadresáře

Podadresáře můžete i smazat. Zvolíte si, který podadresář chcete smazat, a kliknete na položku *smazat*. Smazání podadresáře způsobí i smazání jeho obsahu.



smazat podadresář Podstatná jména (včetně obsahu)

Obrázek 54: Smazání podadresáře

8.4.3 Práce se soubory

Se soubory lze dále pracovat. Jednotlivé soubory můžete v rámci adresářů přesouvat, stejnorodé slučovat soubory do jednoho, přejmenovávat, ukládat a mazat.

Práce se soubory

- cí (oprava) (2, 1741)
- Citoslovce (1, 1027)
- klgMnSc1 (1, 15825)
- ná_i M (1, 5)
- ná_i S (1, 84)
- ná_i Ž (1, 2)
- Předložky (1, 204)
- účelová -cí (2, 1741)
- účelová slova -cí (2, 1911)
- Zkratky (1, 113)

<input type="button" value="sloučit"/>	do	<input type="text" value="Kalivodova/"/>
	souboru	<input type="checkbox"/> pouze spojit a seřadit (případné shodné n-tice budou ponechány vícekrát)
<input type="button" value="přesunout"/>	do	<input type="text" value="Kalivodova"/>
<input type="button" value="zkopírovat"/>	adresáře	
<input type="button" value="přejmenovat"/>	na	<input type="text"/> (jen pro jednotlivý soubor)
<input type="button" value="uložit"/>	na počítač uživatele	(jen pro jednotlivý soubor)
<input type="button" value="porovnat"/>		(jen pro dva soubory)
<input type="button" value="smazat"/>		

Obrázek 55: Práce se soubory

Pokud máte dva či více souborů, které jsou podobné, a chcete je dát dohromady, použijte funkci *sloučit*. Klikněte na čtverečky příslušných souborů a vyberte adresář, do něhož se soubor uloží. Musíte napsat název nového souboru, který vznikne. Jakmile kliknete na *sloučit*, program spojí dva či více souborů s tím, že shodné výrazy ze souborů napíše jen jednou. Nástroj také může soubory jen spojit a seřadit podle abecedy, aniž by shodné výrazy odstraňoval. Tuto volbu zvolíte kliknutím příslušného čtverce na konci řádku.

účelová -cí (2, 1741)
 účelová slova -cí (2, 1911)
 Zkratky (1, 113)

do pouze
 souboru **spojit a setřídít** (případně shodné n-tice budou ponechány vícekrát)

Obrázek 56: Sloučení souborů

Funkce *přejmenovat* umožňuje přejmenovat soubory, jejichž názvy nevyovídají o povaze analyzovaných dat. Označte si příslušný soubor a do kolonky u možnosti přejmenování napište nový název. Následně klikněte na *přejmenovat* a název souboru se přepíše. Tuto funkci lze použít pouze na jednotlivý soubor.

na (jen pro jednotlivý soubor)

Obrázek 57: Přejmenování souborů

Porovnání souborů slouží k porovnání dvou vybraných souborů. Ze seznamu jako příklad vyberte soubory, v nichž jsou uloženy výrazy začínající na *na-* a končící na *-í* v mužském a středním rodě. Označte příslušné soubory a klikněte na tlačítko *porovnat*. Nástroj vezme soubory a porovná je. Na výstup vypíše zeleně slova, která se objevila v prvním souboru a v druhém ne, a červeně slova, která jsou v druhém souboru a v prvním ne. Před dvojtečkou se nachází příslušné poznámky z ručního třídění, pokud jste ho použili.

Porovnání souborů Kalivodova/ná_í M a Kalivodova/ná_í S

:nádhončí
 :náhončí
 :nákupčí
 :námluvčí
 :návodčí
 :náboženství
 :nábožnůstkářství
 :nábřeží
 :nábytkářství

Obrázek 58: Porovnání dvou souborů

Závěr

Cílem naší bakalářské práce bylo vytvoření uživatelského manuálu v českém jazyce pro nástroje DEBDict a Deriv. Vytvořili jsme doporučení, kterého by se měl autor jakéhokoliv manuálu držet, aby byl jeho text čtivý, pochopitelný a stručný. Těchto doporučení jsme se drželi při psaní vlastních manuálů, které jsou součástí bakalářské diplomové práce.

Snažili jsme se naše manuály napsat tak, aby byly pochopitelné a mohl je používat kdokoliv bez ohledu na vzdělání či obor. Nechali jsme s našimi manuály pracovat dva lidi, kteří o programech nevěděli zhora nic, a s uspokojením jsme usoudili, že naše doporučení jsou správná.

Manuál k DEBDictu bude ve zkrácené podobě přiřazen na hlavní stránku nástroje DEBDict, kde si ho každý uživatel bude moci přečíst. Manuál k nástroji Deriv bude přeposlán vyučujícím, kteří tento nástroj využívají ve výuce, aby jej mohli použít a poskytnout ve svých hodinách.

Použitá literatura

- BEČKA, Josef V. et al., 1992. *Česká stylistika*. Praha: Nakladatelství Československé akademie věd. ISBN 80-200-0020-8.
- CIA, 2013. *Central Intelligence Agenci: The World Factbook* [online]. Washington, D.C.: Central Intelligence Agency [cit. 2017-01-05]. Dostupné z: <https://www.cia.gov/library/publications/the-world-factbook/>
- ČAPEK, Tomáš, 2006. *Systém pro částečné sémantické značkování volného textu* [online]. IS MU [cit. 2016-11-22]. Dostupné také z: https://is.muni.cz/th/3923/fi_m/diplomka.pdf. Diplomová práce. Masarykova univerzita. Vedoucí práce Lukáš Svoboda.
- ČECHOVÁ, Marie et al., 2008. *Současná stylistika*. Praha: Nakladatelství Lidové noviny. ISBN 978-80-7106-961-4.
- ČERMÁK, František et al., 2009. *Slovník české frazeologie a idiomatiky*. Praha: Leda. ISBN 978-80-7335-215-8.
- EDDINGS, Joshua et al, 1995. *Jak pracuje internet*. Brno: UNIS publishing. ISBN 1-56276-192-7.
- HAVRÁNEK, Bohuslav (ed.), 1989. *Slovník spisovného jazyka českého*. 2., nezměněné vyd. Praha: Academia, 8 sv.
- HUSÁROVÁ, Dagmar, 2015. *Určování tvarotvorných vzorů neznámých slov* [online]. IS MU [cit. 2016-11-22]. Dostupné z: https://is.muni.cz/auth/th/333279/ff_m/diplomova_prace.pdf. Magisterská diplomová práce. Masarykova univerzita. Vedoucí práce Marek Grác.
- Internetová jazyková příručka* [online], 2017. Ústav pro jazyk český AV ČR [cit. 2017-04-25]. Dostupné z: <http://prirucka.ujc.cas.cz/>
- JEDLIČKA, Alois et al., 1970. *Slovník české stylistiky*. Praha: Státní pedagogické nakladatelství.
- MEJSTRÍK, Vladimír et al., 2003. *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem ministerstva školství, mládeže a tělovýchovy České republiky*. 3., opr. vydání. Praha: Academia. ISBN 80-200-1080-7.
- NLPCentrum. *Prohlášení pro dočasné uživatele programů a dat Centra zpracování přirozeného jazyka* [online]. [cit. 2017-01-05]. Dostupné z: <https://nlp.fi.muni.cz/prohlaseni/>

- OSOSLOBĚ, Klára, 2008. Automatický nástroj Deriv [online]. *IS MU* [cit. 2017-01-05]. Dostupné z:
https://webcache.googleusercontent.com/search?q=cache:NKK3gzU4a3AJ:https://is.muni.cz/el/1421/jaro2008/CJBB84/Automaticky_nastroj_Deriv.doc+&cd=1&hl=cs&ct=clnk&gl=cz&client=firefox-b
- PALA, Karel a Jan VŠIANSKÝ, 1996. *Slovník českých synonym. 2., opr. vyd.* Praha: Nakladatelství Lidové noviny. ISBN 80-710-6059-3.
- PALA, Karel a Pavel ŠEVEČEK, 1999. Česká lexikální databáze typu wordnet (v rámci projektu eurowordnet-2). *Sborník prací filosofické fakulty brněnské univerzity* [cit. 2017-01-05]. ISSN 0231-7567. Dostupné z:
https://digilib.phil.muni.cz/bitstream/handle/11222.digilib/101535/A.Linguistica_47-1999-1_5.pdf?sequence=1
- PETRÁČKOVÁ, Věra et al., 1997. *Akademický slovník cizích slov: [A-Ž]*. Dotisk. Praha: Academia. ISBN 80-200-0607-9.
- Příruční slovník jazyka českého (1937—1957)*, 2008 [online]. Praha: Ústav pro jazyk český [cit. 2017-01-05]. Dostupné z: <http://bara.ujc.cas.cz/psjc/>
- RAMBOUSEK, Adam et al., 2010a. *DEBII — Dictionary Editor and Browser: DEBDict — obecný prohlížeč slovníků* [online]. Brno: Centrum zpracování přirozeného jazyka [cit. 2017-01-05]. Dostupné z:
<http://deb.fi.muni.cz/debdict/index-cs.php>
- RAMBOUSEK, Adam, 2010b. *Lexikografická platforma na vývoj slovníkových aplikací* [online]. IS MU [cit. 2016-11-22]. Dostupné z:
https://is.muni.cz/auth/th/60380/fi_m/dp.pdf. Diplomová práce. Masarykova univerzita. Vedoucí práce Aleš Horák.
- RAMBOUSEK, Adam, 2010c. *DEBDict — obecný prohlížeč slovníků* [online]. Brno: Centrum zpracování přirozeného jazyka [cit. 2017-01-05]. Dostupné z:
<https://deb.fi.muni.cz:8005/debdict/>
- SEDLÁČEK, Radek, 1999. *Morfologický analyzátor češtiny* [online]. Brno: Centrum zpracování přirozeného jazyka [cit. 2016-11-22]. Dostupné z:
<https://nlp.fi.muni.cz/projekty/ajka/ajka.pdf>. Diplomová práce. Masarykova univerzita. Vedoucí práce Pavel Rychlý.
- ŠMERK, Pavel, 2006. *ajka tagset* [online]. Brno: Centrum zpracování přirozeného jazyka [cit. 2017-01-05]. Dostupné z:
<https://nlp.fi.muni.cz/projekty/ajka/tags.pdf>

ŠMERK, Pavel, 2010. *Deriv* [online]. Brno: Centrum zpracování přirozeného jazyka [cit. 2017-01-05]. Dostupné z: <http://deb.fi.muni.cz/deriv/>

VOSSSEN, Piek et al., 1998. *The EuroWordNet Base Concepts and Top Ontology* [online]. Researchgate [cit. 2016-11-22]. Dostupné z: https://www.researchgate.net/profile/Nicoletta_Calzolari/publication/228594694_The_EuroWordNet_Base_Concepts_and_Top_Ontology/links/09e4150604a1a45d3d000000/The-EuroWordNet-Base-Concepts-and-Top-Ontology.pdf

WATTOVÁ, Fiona, 2004. *Kuchařka pro začátečníky*. Praha: Albatros nakladatelství, a. s. ISBN 80-00-01317-7.

WEISSCHER, Anja, 2013. *GWA Base Concepts*. [online] [cit. 2017-01-05]. Dostupné z: <http://globalwordnet.org/gwa-base-concepts/>