

Opravené části jsou zvýrazněny červeně

Technická poznámka

Jednotlivé úkoly, které obsahuje tato příručka, jsou založeny na práci se 100milionovým korpusem SYN2000, v němž se jazykové jevy vyhledávají pomocí programu Bonito, **jehož autorem je Pavel Rychlý z FI MU v Brně**. Manuál k tomuto programu je na internetové adrese <http://ucnk.ff.cuni.cz/bonito/>. **Morfologickou analýzu a desambiguaci korpusu SYN2000 zpracoval Jan Hajič z MFF UK v Praze (viz příložený Popis jednotlivých pozic morfologické značky)**.

Po instalaci je jako základní atribut pro vyhledávání nastaven atribut word. To znamená, že slovní tvary je možné zadávat bez závorek a uvozovek, např. **korpus**. Z tohoto důvodu se variantní zápis [word="korpus"] v klíči neobjevuje. Z tohoto základního nastavení proto vychází i klíč k jednotlivým úkolům. Pro rychlejší orientaci je možné využít 20 základních vyhledávacích vzorů.

Základní publikací pro seznámení s korpusovou lingvistikou a její terminologií, která se v zadáních jednotlivých úkolů objevuje, je „Český národní korpus. Příručka uživatele“ (Praha 2000).

Hlavní část příručky

Strana 7 – 2. odstavec, 3. řádek shora, doplněna e-mailová adresa
(*Filozofická fakulta Karlovy univerzity, nám. J. Palacha 2, Praha 1, 110 00*, e-mail ucnk@ff.cuni.cz a web <http://ucnk.ff.cuni.cz>)

Strana 15 – Úkol 19, 4. řádek pod nadpisem, upraven text
Která slova do frekvence 10 jsou česká a která cizí?...

Strana 79 – upraveny sloupce u tabulky

	<i>brýle</i>	<i>brejle</i>
CELKEM	2280	206
beletrie (NOV, COL)	816	139
beletrie (FAC)	75	16
krásná literatura bez označení	3	
básně a písně (VER, SON)	31	13
scénáře (SCR)	8	
	<u>41 %</u>	<u>82 %</u>
ostatní	1347	38
	<u>59 %</u>	<u>18 %</u>

Strana 80 – Úkol 9, řádek pod nadpisem, doplněny horní uvozovky
V korpusu zadáme dotaz `[lemma="“.+ú.+“]` a získáme 50689 výskytů.

Strana 81 – Úkol 10, řádek pod nadpisem, opraven počet výskytů
V korpusu zadáme dotaz `[word="“.+ú“]` a dostaneme 2061 výskytů.

Strana 87 – Úkol 20, řádek pod nadpisem, odstraněna první hranatá závorka
V korpusu zadáme dotaz `{[(lemma=".*[hrk]i.*") & (word=".....")]}...`

Strana 88 – Úkol 1, 2. řádek pod nadpisem, upraveno odsazení

V korpusu zadáváme postupně dotazy

`[lemma="lov"]` – 3664 výskytů

`[lemma="hon"]` – 1468 výskytů

`[lemma="lovení"]` – 46 výskytů

`[lemma="honění"]` – 51 výskytů

Strana 91 – Úkol 6, 3. řádek pod nadpisem, doplněna hvězdička
Potom zadáme dotaz `[(lemma="proti.") & (tag="Vf.*")]` a získáme 35 výskytů.

Strana 91 – Úkol 7, řádek pod nadpisem, doplněna hvězdička

V korpusu zadáme dotaz [(lemma=".+jít") & (tag="V.*")] a získáme 211985 výskytů.

Strana 91 – Úkol 8, řádek pod nadpisem, doplněny dvě hvězdičky

V korpusu zadáme dotaz [(lemma="pod.*") & (tag="N.F.*")] a získáme 106194 výskytů.

Strana 91 - Úkol 6, 7. řádek pod nadpisem nahrazen celý takto:

proti- 35 infinitiv : 312 slovesné tvary, spolu- 2732 infinitiv : 7734 slovesné tvary

Strana 93 – Úkol 11, řádek pod nadpisem, doplnit část slova

V korpusu zadáme dotaz [lemma=".*přeškol.*"] a získáme 171 výskytů.

Strana 94 – Úkol 12, řádek pod nadpisem, doplněny dvě hvězdičky

V korpusu zadáme dotaz [lemma=".*pozi[cč].*"] a získáme 50515 výskytů.

Strana 94 – Úkol 14, velká písmena v tabulce opravena na malá

<i>péro</i>	266	pero
<i>lejstro</i>	211	úřední papír, listina
<i>cigáro</i>	121	cigareta
<i>fáro</i>	69	auto

Strana 94 – Úkol 12, řádek pod nadpisem, doplněna hvězdička

V korpusu zadáme dotaz [lemma=".*pozi[cč].*"] a získáme 50515 výskytů.

Strana 99 – Úkol 18, 3. a 4. řádek pod tabulkou, opraveny počty výskytů

[(lemma="během") | (word="průběhu")] a zůstane nám 268 výskytů, popř.

[(word="během") | (word="průběhu")] a zůstane nám 219 výskytů.

Strana 101 – Úkol 2, 2. řádek pod nadpisem odstraněna mezera

V korpusu zadáme dotaz [bcčdd'fghjklmnňpršštt'vwxyz][bcčdd'fghjklmnňpršš-
tt'vwxyz][bcčdd'fghjklmnňpršštt'vwxyz][bcčdd'fghjklmnňpršštt'vwxyz]

Strana 104 - Úkol 7, 7. řádek pod nadpisem závorka { nahrazena závorkou hranatou

[(word=".*[oO][vV][0éÉ]") & (tag="N.MP1.*") & ((lemma="politik") | (lem-

Strana 104 – Úkol 7, 3. řádek ve 2. sloupci v tabulce, opraven počet výskytů

<i>politici</i>	4211	<i>politikové</i>	825
-----------------	------	-------------------	-----

Strana 115 – Úkol 13, řádek pod nadpisem, doplněna hranatá závorka

V korpusu zadáme dotaz [(lemma="vůně") | (lemma="zápach") |

Strana 120 – Úkol 18, 5. a 6. řádek pod nadpisem, na dvou místech opraveny počty výskytů
Poměry jsou *zatleskal* : *zatleskal rukama* = 59 : 4, *zamrkala* : *zamrkala očima* = 49 : 3

Strana 126 – Úkol 9, 3. řádek pod nadpisem, doplněna kulatá závorka
(`[lemma="světle"] [pos="A"] [lemma="barva"]`) – 30 výskytů

Strana 127 – Úkol 12, řádek pod nadpisem, doplněna kulatá závorka
V korpusu zadáme dotaz (`[lemma="oblast"] [{}3] [lemma="život"]`) a získáme 37 výskytů.

Strana 140 – Úkol 16, 1-2. řádek pod nadpisem, odstraněno `within <s>`
V korpusu postupně zadáváme dotazy (`[word="Kam"] [{}1] [word="\?"]`) **within<s>**,
(`[word="Kam"] [{}2] [word="\?"]`) apod.

Strana 143 – Úkol 1, 3. řádek pod nadpisem, odstraněny uvozovky
dat takto: **.{32}**

Strana 150 – Úkol 14, 4. řádek pod nadpisem, doplněna meze
zaměstnán. Nejfrekventovanější ze spisovných tvarů je neutrum *dáno* (30x), ovšem

Strana 150 – Úkol 15, řádek pod nadpisem, doplněna kulatá závorka
V korpusu zadáme dotaz (`[lemma="lézt"] [] [pos="[NPC]"]`) a získáme 925 výskytů.

Strana 150 – Úkol 14, 5. řádek pod nadpisem, doplněna meze
ovšem mnohem frekventovanější je obecně český tvar *dany* (96x).

Strana 165 – Úkol 7, 13. a 20. řádek pod nadpisem, odstraněno „within <s>“:
`[tag="VB.....F.*" & lemma="být"] [tag="[NARDCP].*"]? [tag="Vf.*"] within`
`<s>`
`[lemma="dokud"] [tag="[NARDCP].*"]? [tag="V.....A.*"] within`
`<s>`

Strana 166 – Úkol 12, 2. řádek pod nadpisem, slovo „lemma“ nahrazeno slovem „tag“
`[tag="V.*"] [word=",\" [word="že"]`

Strana 166 – Úkol 13, 2. řádek pod nadpisem, odstraněny uvozovky, tučně
.*t.*t.*t.*t.*t.*t.*

Strana 168 – Úkol 10, řádek pod nadpisem, doplněna první kulatá závorka a meze za
druhou kulatou závorkou

V korpusu zadáme dotaz (**[tag="PP.S1.*"] [tag="V.....R.*"]**) a získáme 9442 výskytů.

VYHLEDÁVACÍ VZORY

12) kombinace lemmat

např. *domáci kočka* (tj. *domáci kočky, domáci kočce...*)

[lemma="domáci"] [lemma="kočka"]

12a vyhledání kombinace dvou lemmat a slovního tvaru

např. *měla jsem namířeno, má být namířeno* apod.

([lemma="mít"] [lemma="být"] [(tag="Vs.*") & (word="*o")])

13) kombinace lemmat s jedním vloženým slovem

např. *domáci léčení kočky*

[lemma="domáci"] [] [lemma="kočka"]

14) kombinace lemmat s více vloženými slovy

např. *domáci zvířata a kočky, domáci léčení kočky, domácím kočkám, domáci zvířata, zvláště kočky*

[lemma="domáci"] [{0,3} [lemma="kočka"]

14a přesný počet vložených slov

např. *pes, když uvidí kočku, pes, event. kočka, psi, celá smečka koček*

[lemma="pes"] [{3} [lemma="lemma="kočka"]

14b opakující se lemmata nebo tagy několikrát za sebou (např. 3 substantiva v dativu)

např. *hraběti Karlu Josefovi, panu Václavu Durychovi, doktoru Josefu Peškovi*

[tag="N...3.*"]{3}

OBSAH

Předmluva	str. 5
Práce s ČNK krok za krokem	
A. Pravopis	str. 11
B. Tvoření slov	str. 16
C. Tvarosloví	str. 22
D. Slovní zásoba	str. 28
E. Kolokace	str. 33
F. Syntax	str. 38
G. Kombinovaná zadání	str. 43
2. Co říká o různých slovech korpus a co slovníky	str. 47
3. Význam slova prozrazuje kontext	
A. Slovní tvary	str. 53
B. Lemmata	str. 54
C. Kolokace	str. 55
D. část slov	str. 56
E. Formálně podobná slova	str. 57
F. Významy slov	str. 63
4. Úkoly pro náročnější	str. 70
5. Práce se subkorpusy	str. 74
Klíč	
1A.	str.77
1B.	str. 88
1C.	str.101
1D.	str. 111
1E.	str. 122
1F.	str.134
1G.	str.143
2.	str. 153
3.	str.155
4.	str.164
5.	str.169
Slovníček hlavních termínů	str. 173
Studijní literatura	str. 176
Vyhledávací vzory	str. 177