

## **Automatický nástroj Deriv**

*Deriv* (<http://nlp.fi.muni.cz/projekty/ajka/cjbb85/>) je nástroj vyvinutý na FI MU, který umožňuje jednoduchým způsobem vyhledávat ve strojovém slovníku automatického morfologického analyzátoru *ajka* (Sedláček, 2004) lemmata podle formálně zadaných pravidel (úvodní řetězec, koncový řetězec, morfologická značka) a vytvářet touto cestou seznamy slov, která s velkou mírou pravděpodobnosti patří k jednomu derivačnímu typu, respektive jsou tvořena jedním derivačním prostředkem. Se seznamy lze následně pracovat, je možné je prohlížet ve dvou modech, a to jako prosté seznamy nebo jako seznamy s uvedením frekvence vyhledaných jednotek v korpusu SYN2000 (viz níže). Nástroj tedy umožňuje rychlé prohledání a extrakci dat z rozsáhlého strojového slovníku českých kmenů (Osolsobě, 1996, Sedláček, 2004), je propojen s korpusem SYN2000 (100 milionů slovních tvarů), a umožňuje tak v korpusu ověřovat frekvenci jednotek zahrnutých ve strojovém slovníku.

Poznámka: Strojový slovník češtiny (Osolsobě, 1996) byl vybudován na základě hesláře SSSJČ a doplněn o řadu dalších slov na základě testování aplikací automatických morfologických analyzátorů *lemma* (Ševeček, 1996) a *ajka* (Sedláček, 2004) na korpusech češtiny (korpusey ČNK, korpusey budované na FI MU). Strojový slovník analyzátoru *ajka* zahrnuje zhruba 400 000 jednotek (kmenů), k nimž na základě formálních pravidel (deklinačních vzorů) generuje 6 milionů slovních tvarů.

## **Analýza materiálu**

Pro vyhledání informací o možných slovotvorných vztazích se používá automatický nástroj *Deriv* (webové rozhraní) následujícím způsobem.

Po přihlášení (uživatelské jméno : **Osolsobe** /heslo: **1996**) v *Deriv* zvolíme z nabídky funkci **Hledání slov podle pravidla**. Nejdříve zadáme do nabídky příkaz pro vyhledání slov končících na *hledaný řetězec (suffix)* a majících značku k.\* (číslice-slovní druh) a uložíme je do souboru (1). Pak zadáme příkaz pro vyhledání slov končících na *hledaný řetězec (suffix)* a majících značku k.\* (číslice-slovní druh) a uložíme je do souboru (2).

Vytvořené soubory (1) a (2) sloučíme a vytvoříme z nich soubor jeden (3). Na takto vzniklý soubor použijeme funkci **Hledej základová slova** a zadáme, že chceme v příslušném souboru najít dvojice slov takových, že jeden člen páru končí na *hledaný řetězec (suffix)* a druhý je slovo vzniklé odtržením *hledaného řetězce (suffixu)* a jeho nahrazením *hledaným řetězcem (suffixem)*. Dále aplikujeme funkci **Rozdělení souboru** a získáme dva soubory, jeden, který obsahuje nalezené dvojice (4) a druhý, který obsahuje zbylá slova, ke kterým se nepodařilo najít podle příslušného pravidla slovo do „páru“ (5). Soubor (4) pak obsahuje kandidáty na hledaný derivační typ. Postup opakujeme pro vytřídění dalších dvojic.

Příklad:

## Rozcestník

- [Hledání slov podle pravidla](#)
- [Práce se soubory](#)
- [Sloučení souborů](#)
- [Odhlášení](#)

## Hledání slov

Předpona:

Přípona:

Značka:

Vzory (oddělené čárkou):

Soubor:

Najdi a ulož

Vymaž formulář

[Rozcestník](#)

## Výběr souboru

Aktuální adresář: [Osolsobe](#)

Nový podadresář:

Vytvořit



šev [397]

Pouze slova označená jako:



Setřídít retrográdně

Otevř <u>í</u> s frekvencemi	Otevř <u>í</u> bez frekvencí	Přejmenuj	<u>S</u> maž
------------------------------	------------------------------	-----------	--------------

Uprav dvojice	Uprav trojice	Rozděl s <u>o</u> ubor
---------------	---------------	------------------------

Hledej základová s <u>l</u> ova	Přesuň jinam
---------------------------------	--------------

[Rozcestník](#)

## Sloučení souborů

- Osolsobe/ šev
- Osolsobe/AM

Označené sloučit do souboru:   [Rozcestník](#)

## Hledání základových slov ke slovům ze souboru Osolsobe/ šev

	Předpona	Přípona	
Odvozené slovo	<input type="text"/>	<input type="text"/>	
Základové slovo	<input type="text"/>	<input type="text"/>	
Základové slovo	<input type="text"/>	<input type="text"/>	<input type="checkbox"/> Použít
Výsledek uložit do souboru:	<input type="text" value="Osolsobe/"/>	<input type="text"/>	

Najdi a ulož	Vymazat fo <u>r</u> mulář	<a href="#">Rozcestník</a>
--------------	---------------------------	----------------------------

# Rozdělení souboru Osolsobe/JEHNECI

Cílový soubor pro známá základová slova:

Cílový soubor pro neznámá základová slova:

## [Rozcestník](#)

ce  
k1gM  
t  
k5.\*mF

bránce,bránit  
chránce,chránit  
chválce,chwálit  
dohodce,dohodit  
dopravce,dopřavit  
dovozce,dovozit  
hanobce,hanobit  
hájce,hájit  
mluvce,mluvit  
obhájce,obhájit  
obránce,obránit  
ochránce,ochránit  
odvozce,odvozt  
pojistce,pojistit  
popravce,popravit  
poradce,poradit  
porobce,porobit  
proradce,proradit  
přenosce,přenosit  
přepravce,přepravit  
působce,působit  
rozhodce,rozhodit  
schránce,schránit  
smluvce,smluvit  
snoubce,snoubit  
soudce,soudit  
stráľce,stráľit  
zachránce,zachránit

## KORPUS

lemma

.\*ce

P-filtr

NNM.\*

N-filtr

.\*(a|e|i|o|u|y|á|é|ě|í|ó|ú|ů|ý)ce

lemma:	##	
zástupce	24094	
výrobce	15104	
zájemce	9579	
vůdce	6880	
soudce	5787	
tvůrce	5527	
důchodce	5157	
obránce	5050	
poradce	4328	
správce	4106	
odpůrce	3121	
obhájce	2732	
prodejce	2533	
průvodce	2527	
předchůdce	2410	
nástupce	2354	
strážce	2162	
ochránce	2103	
zastánce	1815	
nájemce	1728	
dárce	1652	
dovozce	1606	
vládce	1543	
dopravce	1448	
zákonodárce	1184	
plátce	1162	
příjemce	1160	
výherce	1120	
vývozce	1032	
žalobce	995	
původce	983	
škůdce	905	
dozorce	829	
zrádce	710	
vynálezce	610	
rádce	602	
únosce	548	
strůjce	401	
přepravce	394	
zachránce	361	
vojevůdce	346	
porotce	343	
strojvůdce	333	
Pierce	234	
svůdce	194	
rozhodce	171	
autodopravce	170	
spolutvůrce	158	
zpravodajce	124	
prvovýrobce	107	
nálezce	103	
mírotvorce	94	

vlastizrádce	75
chlebodárce	72
Terence	64
úpadce	64
odhadce	62
Lawrence	53
přímluvce	50
samosoudce	48
úpravce	47
nákupce	43
velkovýrobce	38
Peirce	33
příkazce	32
mravokárce	31
výstavce	29
samoplátce	29
zástavce	27
neplátce	23
samovládce	23
velezáradce	22
pronájemce	22
opravce	18
dohodce	17
podpůrce	16
krutovládce	15
zhoubce	14
mírotvůrce	14
převodce	13
výkupce	13
starodůchodce	12
protichůdce	12
podnájemce	11
malovýrobce	11
spoluvládce	10
divotvůrce	9
maloprodejce	9
velkoprodejce	9
přídělc	9
židobijce	8
divotvorce	8
ústavodárce	8
autovýrobce	8
přestupce	8
přechůdce	7
úplatce	7
oprávce	7
normotvůrce	6
drakobijce	6
převozce	5
ohněstrůjce	5
prapůvodce	5
ohnostrůjce	4
návodce	4
popravce	4
odvozce	4
školdozorce	3
spolustrůjce	3
nápravce	3
spolunájemce	3
hrůzovládce	2
světovládce	2

hromovládce	2	
zlolajce	2	
výzkumce	2	
vševládce	2	
Divotvůrce	2	
smíchotvorce		2
exporadce	2	
lidovládce	1	
všenápravce	1	
sektobijce	1	
zeměvládce	1	
doprovodce	1	
středovládce		1
supervýrobce		1
hmyzochránce		1
přínosce	1	
antiprůvodce		1
stranovládce		1
šéfdovozce	1	
France	1	
spoluvynálezce		1
spolupůvodce		1
bolševikobijce		1