CJBB84
Morfologie a korpus

10.00-11.30 G13

Triviální a netriviální vyhledávání substantiv podle vzoru

- Morfologické značkování neobsahuje informace o skloňovacím typu
- Je možné vyhledat v korpusech lemmata skloňovaná podle určitého vzoru?

Formální vlastnosti substantiv a vzor

- Zakončení a slovní druh (*kos, sál, pila, žeň, chudě*)
- Slovníkové informace: lemma, rod, tvar genitivu
- Rod a forma - zakončení (*kroj* x *zbroj, kůň* x *tůň*)
- 1. a 3. pozice (pražský systém)
- Atributy k a g (brněnský systém)

Substantiva (N/k1)

- Rod (tag: pozice 3/g)
- Zakončení lemmatu

Slovní formulace

- Najděte substantiva skloňovaná podle vzoru *žena.*
- 1. substantiva
- 2. feminina
- 3. lemma končí na *a*

Cql dotaz

- [tag="NNF.*" & lemma=".*a"]

vysoko nad jejich hlavami zašustili jako badmintonový míček , sklopili	letky/letka/NNFP4-----A-----	a vystoupali do nějaké jir
Představoval si jeho tělo bez života a krev vytékající do	vody/voda/NNFS2-----A-----	s octem . A s lítostí pom
? Kolašin potřebuje leccos . Vodu , kanalizaci , zdroje	elektřiny/elektřina/NNFP4-----A-----	a tepla . Musíme se post
není břesk hesel , ale náročná kázeň a skromnost .	Cifry/cifra/NNFP4-----A-----	bych seškrtal , uklízečky
neodolají mé horlivosti a stáhnou se pryč . Slunce barví	krajinu/krajina/NNFS4-----A-----	rezavě rudě . Namlouván
trochu sebrat , nemůže si přijít a po tak dlouhé	době/doba/NNFS6-----A-----	se vnutit k Petře na návš
krajany pracující v ČR , poslední dobou pracovala u montážní	firmy/firma/NNFS2-----A-----	. Měla osmiletou dceru .
. A sedmnáctka , jménem Goethova , přináší pohled na	lavičku/lavička/NNFS4-----A-----	vytesanou do skály , kde
zavolaly hasiče , dívku našli až policejní potápěči po několika	hodinách/hodina/NNFP6-----A-----	. Do bývalého kaolinovéh
sleva žádný vliv a jen 1 % z nich se	slevám/sleva/NNFP3-----A-----	záměrně vyhýbá . Špatná
až 240 km/h , 60metrový volný pád , restaurace ,	prodejny/prodejna/NNFP1-----A-----	suvenýrů , expozice o his

Počet lemmat (podle frekvence) skloňovaných podle vzoru žena v korpusu SYN2015

Frekvenční limit: 0 Nastavit limit

Celkem: 20043 (401 str.)

	Filtr	lemma	Frekvence
1.	p/ n	doba	108 001
2.	p/ n	ruka	89 653
3.	p/ n	strana	79 801
4.	p/ n	žena	75 329
5.	p/ n	hlava	72 080
6.	p/ n	cesta	68 611
7.	p/ n	voda	62 399
8.	p/ n	hodina	56 758
9.	p/ n	cena	52 219
10.	p/ n	Praha	52 029
11.	p/ n	řada	50 694
12.	p/ n	škola	50 466
13.	p/ n	firma	49 113
14.	p/ n	koruna	45 414
15.	p/ n	otázka	39 932
16.	p/ n	kniha	39 619
17.	p/ n	skupina	39 589
18.	p/ n	rodina	38 284
19.	p/ n	matka	37 123
20.	p/ n	změna	36 652

Lemmata rozpoznaná automatickou morfologickou analýzou

- Nerozpoznaná lemmata
- [tag="X.*" & lemma=".*([ayěeubdfghklmnprstvz]|ou|ách|ám|ami)"]

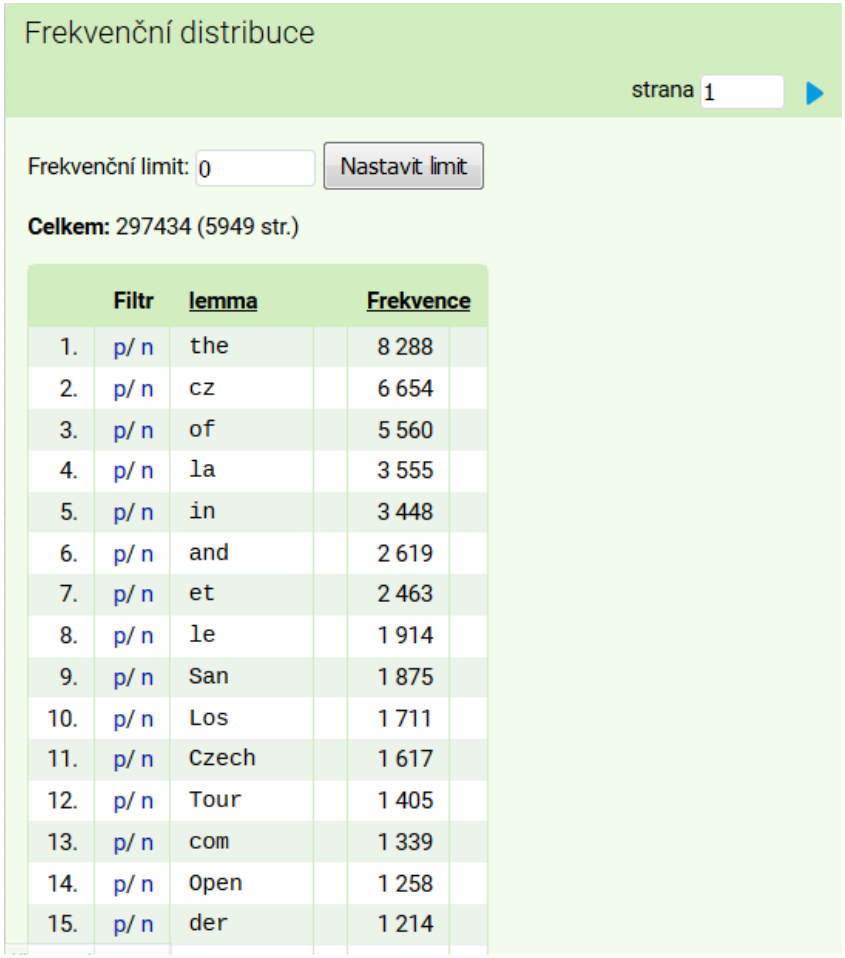
Výsledky

Výskytů: **861 434** | i.p.m. : **7 134,1** (vztaženo k celému "omezeni/syn2015") | ARF : **363 169,21** | Výsledek je promíchán 1 / 21 536

Výběr řádků: základní

Procházka amazonským pr...	cesty na nějaký náhradní transport . Samotné soužití s domácími	castañeros/castañeros/XX----------	bylo nac
Aleje české a moravské kra...	) v Ilford u Londýna (1848) nebo v	Olmstedově/olmstedově/X@----------	plánu pr
Sport	výběrem legendy Zinedina Zidana ! Pražský tým se jmenuje FC	Hunters2Dreams/hunters2dreams/X@----------	a je tvoi
Pátý elefant	přišla odpověď , rozhlédli se oba trollové kolem , spatřili	Tračníka/tračníka/X@----------	a pomal
K moři	, Bára jí telefonuje a Matti jí nedovolí za ní	zaject/zaject/X@----------	, dokud
Svět kuchyní	) , cena 3 320 Kč , SIKO 7/ Komfortním zařízením	Hot/hot/XX----------	Water D
Skřípavý smích Jeana Anou...	. Generál přijímá návštěvy své bývalé milenky , herečky Mélusine	Melita/melita/X@----------	, a jejího
Reflex	. Překvapila mě síla mé tety Hany , doma říkáme	Handulinky/handulinky/X@----------	, Hnátov
Křižácké zlato	co to je . Jedna dávná legenda vyprávěla o obrech	Kablunatech/kablunatech/X@----------	opásany
Ohniska napětí v postkoloni...	„ etnických čtvrtí " , obývaných Somálci , Afary ,	Oromy/oromy/X@----------	, Amhar
Předkové	pozice je v zásadě pozitivní a kritická , zdůrazňuje spíše	oportunismusdávných/oportunismusdávných/X@----------	hominin

Frekvenční analýza



Frekvenční distribuce

strana 1

Frekvenční limit: 0 Nastavit limit

Celkem: 297434 (5949 str.)

	Filtr	lemma	Frekvence
1.	p/ n	the	8 288
2.	p/ n	cz	6 654
3.	p/ n	of	5 560
4.	p/ n	la	3 555
5.	p/ n	in	3 448
6.	p/ n	and	2 619
7.	p/ n	et	2 463
8.	p/ n	le	1 914
9.	p/ n	San	1 875
10.	p/ n	Los	1 711
11.	p/ n	Czech	1 617
12.	p/ n	Tour	1 405
13.	p/ n	com	1 339
14.	p/ n	Open	1 258
15.	p/ n	der	1 214

Zjednodušení dotazu

- [tag="X.*" & lemma=".*([ayěeu]|ou|ách|ám|ami)"]

: **390 279** | i.p.m. : **3 232,16** (vztaženo k celému "omezeni/syn2015") | ARF : **174 380,31** | Výsledek je promíchán

ádků: základní

Mladá fronta Dnes	, " řekl Váňa České televizi . V sedle třináctiletého	Tiumena/tiumena/X@----------	abso
Technik	zatížení . Panel řídicího systému RCS 05 . Kvalitní MIG	Double/double/XX----------	Pulse
Pavilon z oblaků	kouř . „ Mám hlad , " dal se slyšet	Marume/marume/X@----------	. „ To
Mladá fronta Dnes	při každém dalším průchodu turnikety bude porovnávat obsluha . Firma	Melida/melida/X@----------	převz
Johann Sebastian Bach	novou formu (putující chorálová melodie v kompaktní větě „	Jesu/jesu/X@----------	, mei
Lidové noviny	portrét , " dodal Mikulka . Jakub . Vítězný portrét	Spodinka/spodinka/X@----------	POSL
Velká kniha římských detek...	" Bohové ! " vykřikl Arpocras . Jediným hrozivým pohledem	Pudenta/pudenta/X@----------	umlč
Marianne	Průzračná , čistá a vonící až kovově . Laine de	verre/verre/X@----------	, Serg
Vesmír	časové období . 1) Bagemihl B . , Biological	Exuberance/exuberance/X@----------	. Ani
Marianne	povrch leštěný , keramický střep , 1967 Kč/m2 , Mondo	Ceramica/ceramica/X@----------	Doko
Nymburský deník	z řecké ligy , ale i Eurocupu a Euroligy .	Zasvou/zasvou/X@----------	dosa
Maxim	a možná i o tom napíšeme . Naprosto skvělý je	Range/range/X@----------	Rove
Deník ctihodné prostitutky	, zašli si na kávičku nebo na zákusek k „	Vasiliu/vasiliu/X@----------	" , za
Johann Sebastian Bach	Zelenky a Giovanniho Alberta Ristoriho i koncertního mistra Johanna Georga	Pisendela/pisendela/X@----------	. Bac

Příliš mnoho dat

Frekvenční limit: 0 Nastavit limit

Celkem: 147526 (2951 str.)

	Filtr	lemma	Frekvence
1.	p/ n	the	8 288
2.	p/ n	la	3 555
3.	p/ n	le	1 914
4.	p/ n	Santa	1 017
5.	p/ n	Jeanie	1 000
6.	p/ n	me	857
7.	p/ n	du	822
8.	p/ n	Zoey	780
9.	p/ n	One	713
10.	p/ n	die	618
11.	p/ n	League	590
12.	p/ n	mobile	574
13.	p/ n	Daily	539
14.	p/ n	Core	530
15.	p/ n	Energy	527

Ještě zjednodušíme

- [tag="X.*" & lemma="(ách|ám|ami)"]

(vztaženo k celému "omezeni/syn2015") | ARF : **1 130,19** | Výsledek je promíchán 1 / 61

, roste hojně v druhotných vysazených lesích , zejména v	akátinách/akátinách/X@----------	, indikuje d
, 382 stran LEVICOVÝ INTELEKTUÁL . China Miéville patoí k	hvizdám/hvizdám/X@----------	soueasné f
let života . Připojuji se k početným gratulantům a upřímně	blahoželám/blahoželám/X@----------	. Vyprošuji
hroznu pinotu s kamenitým a nesmírně oživujícím , sametovým ,	kyselinkami/kyselinkami/X@----------	však vyváž
módy za neuvěřitelně nízké ceny – Primark , známý českým	shopperkám/shopperkám/X@----------	hlavně z Lo
: " Ráno s Piotrem Brožynou ve Wilanowě a v	Lazienkách/lazienkách/X@----------	. Piotr je je
okolí , můžete si objednat „ vitaminové balíčky “ ve	freshbedýnkách/freshbedýnkách/X@----------	, které dora
ještě nevyprchal . Ten smrad mě dovádí k šílenství .	Hadami/hadami/X@----------	se vrátí z p
které báječně osvěží váš interiér ! Tip Květináče s kvetoucími	pokojovkami/pokojovkami/X@----------	umístěte d
: s primární aminokyselinou modrofialový produkt Ruhemanova violeť , s	imonoskupinami/imonoskupinami/X@----------	pak vytváří
s námi sedí pod akácií před vesnicí , se jmenuje	Lkichami/lkichami/X@----------	. U Sambur
nicméně že nehodlám odejít jen tak beze všeho . Théodore	Zami/zami/X@----------	, kterému p
Nastoupil jsem na metro . Když jsem dorazil , vyděšená	Hadami/hadami/X@----------	seděla na c
společně vyřeší . Už si na to zvykli . Tím	senám/senám/X@----------	podařilo oc

Alespoň něco

	Filtr	lemma	Frekvence
1.	p/ n	Hadami	157
2.	p/ n	ami	72
3.	p/ n	ách	23
4.	p/ n	senám	18
5.	p/ n	Yami	16
6.	p/ n	ám	16
7.	p/ n	narozkám	14
8.	p/ n	Áách	14
9.	p/ n	pastrami	14
10.	p/ n	vámvámVám	12
11.	p/ n	Zami	12
12.	p/ n	klinoformami	11
13.	p/ n	mikropilotami	10
14.	p/ n	Tami	10
15.	p/ n	nami	9
16.	p/ n	serpentýnách	8
17.	p/ n	kuwách	8
18.	p/ n	Forgách	8
19.	p/ n	Zátorách	7
20.	p/ n	martenskách	7

Vzory zjistitelné analogicky

- U kterých vzorů kombinací formy lemmatu a morfologické informace o rodu získáme jednoznačný výsledek?
- U kterých je postup složitější?

vzory s komplikovanějším postupem

- *pán* a *muž*
- *hrad* a *stroj*
- *píseň* a *kost*
- *moře* a *kuře* POZOR na dvě grafické varianty *[eě]*

Úkol na 17. 10. 2018

- Popiš postup vyhledávání substantiv skloňovaných podle vzorů *pán/muž.*
- Všímej si, jak se chovají obojetné souhlásky *[bfmpvlsz].*
- Lze mezi obojetnými souhláskami najít nějakou skupinu, která by byla typickým zakončením pouze tvrdých vzorů?
- Existují nějaké postupy, jak formálně odlišit maskulina zakončená na obojetné souhlásky, které se mohou vyskytovat v zakončení maskulin obou vzorů?