

Mezi Skyllou a Charybdou přegenerování a podgenerování

*Jak efektivně vyhledávat v korpusech data pro
výzkum slovo tvorby*

Klára Osolsobě
osolsobe@phil.muni.cz

Information retrieval

- Termíny, jejichž pomocí se hodnotí úspěšnost vyhledávání
- Přesnost (precision)
- Pokrytí (recall)

Positives/negatives

- Mějme vyhledávací systém a zadejme dotaz (cql dotaz v korpusovém manažeru). Výsledkem jsou data, jejichž část A) odpovídá tomu, co jsme hledali (true positives) a B) část jsou data, která přestože splňují zadání, neodpovídají tomu, co jsme hledali (fals positives).
- Zároveň ovšem nenajdeme C) data, která jsme najít chtěli (fals negatives) a také D) data, která jsme najít nechtěli (true negatives).

Pravda (+) /výsledek (AB)

- A: + + B: + -
- C: - + D: - -

Pokrytí/přesnost

- Pojmem **pokrytí (recall)** označujeme číslo vyjadřující, kolik ze všech výsledků, které se měly nalézt, bylo nalezeno $A / (A + C)$.
- **Přesnost (precision)** je podíl správně nalezených výsledků mezi všemi nalezenými, tedy $A / (A + B)$.

Efektivní kladení dotazů

- přegenerování
- podgenerování

Přegenerování

- *Formální definici (algoritmu) odpovídají jednotky, které tvoří homogenní skupinu (tu, kterou se prostřednictvím formálního zadání snažíme definovat), ale i jednotky, které jsou vůči této skupině heterogenní. Tento jev spadá na vrub obecné vlastnosti přirozeného jazyka, jíž je víceznačnost (homonymie) na všech úrovních.*

Podgenerování

- *Rubem téže mince je tzv. podgenerování, tedy případ, kdy formální zadání je vymezeno příliš úzce, takže nejsou zachyceny jednotky, které se jeho prostřednictvím snažíme definovat.*

Příklad efektivního vyhledávání jmen osob od l-ových příčestí

- maskulina životná
- skloňují se podle vzoru *pán*
- končí na *-l*

Formulace dotazu pro získání dat z korpusů

- Korpus > Implicitní atribut lemma
- *.*|.*
- P-filtr
- Korpus > Implicitní atribut tag
- NNM.*
- Konkordance > Statistiky > Frekvenční distribuce
- frekvenční distribuce **lemmat**



Hledat v korpusu

Korpus:

syn2010

Typ dotazu:

CQL

[vložit tag](#) | [vložit "within"](#) | [klávesnice](#)

CQL:

```
[lemma=".*1" & tag="NNM.*"]
```

V dotazu CQL můžete vložit další řádek stisknutím klávesy Shift+ENTER ([další tip](#))

pozorování dat

tom záleží ? " odvětil netrpělivě . " Kdyby nebylo	Sherrilla	a téhle bláznivé ženské , všechno by vyšlo . Ale
" odhaduje Skružný . Největším soupeřem jeho týmu bude tradiční	rival	Pepino Ostrava , finálové ambice mají i brněnští Bulldogs .
se živila jako překladatelka . Máte nějakého oblíbeného anglicky píšícího	spisovatele	nebo spisovatelku ? Lituju , že není přeložená moje oblíbená
zatím nedařilo plnit ani onu zákonnou devadesátidenní lhůtu a mnohým	žadatelům	přišla místo odpovědi omluva . Úředníci se hájí , že
židličku a natáhl nohy před sebe . " Vy s	Lubellem	jdete proti sobě a já jsem zrovna uprostřed . To
dost citu a ohleduplnosti . Je schopen podat svému nejlepšímu	příteli	nejnovější rostlinný alkaloid , ne ze zlé vůle , rozumíš
jím způsobené se majitel nevyviní , " shrnul . Majitel	pitbula	obviněn z utýrání kočky Z týrání zvířat byl obviněn 21letý
" vysvětlil Remeš . U F. L. Věka : tunel	buditelů	Kdyby existoval stroj času , v hostinci na náměstí v
dovolovalo vycházet jen za výjimečně krásného počasí , neměl jsem	přátel	, kteří by svými návštěvami přerušili jednotvárnost mých dnů .
někým v domě . Je tu přece ten komorník ,	Lal	Rao , kterého jsme neviděli . Podle paní Bernstonové to
nové . Jmenuje se Pistorius podle svého zakladatele , dlouholetého	ředitele	nakladatelství Mladá fronta a do letoška nakladatelství Paseka . Právě
: " Annabel Čung . " Bond se obrátil ke	Quarrelovi	: " Zavolejte Slizavce . " Quarrel volnou rukou uchopil
primitivním a kostrbatým písmem : ' Podpis čtyř - Jonathan	Small	, Mahomet Singh , Abdullah Khan , Dost Akbar .
k takovým zločinům obyčejně dochází , když jsou u moci	liberálové	. Ti dovedou poplést hlavy lidovým masám , což logicky
, koupí si miliardář nespokojené lidové masy a uplatí královského	generála	. Rychle je zinscenována vlastenecká vzpoura . Do chodu se
k nejlépe využít atraktivní brněnskou lokalitu zvanou Západní brána včera	zastupitelé	nerozhodli . Bod , který měl vyřešit , zda postavit
konce života . Jak bych se mohl s tím tvým	přítelem	seznámit ? " " Určitě bude v laboratoři . Buď

◀ široký most se sloupkovitým zábradlím a pod námi lenivě plula temná řeka . Za vodou se opět prostírala širá pustina z cihel a malty , jejíž ticho přerušovaly jen těžké pravidelné kroky strážníka a prozpěvování a křik nějaké opožděné skupinky **hýřilů** . Po obloze pomalu plynul těžký mrak a z potrhaných mračen mdle blikaly dvě tři hvězdičky . Holmes ujížděl mlčky , s hlavou skloněnou na prsa a s výrazem člověka úplně pohrouženého do svých myšlenek , a já seděl vedle ▶

Alonso Renault 2002 Rubens Barrichello Ferrari Jezdci podle počtu vítězství	Michael	Schumacher 4 Ayrton Senna 3 z dalších vítězů : Nelson
pravidelné kroky strážníka a prozpěvování a křik nějaké opožděné skupinky	hýřilů	. Po obloze pomalu plynul těžký mrak a z potrhaných
vás chci dotknout , musím říci , že pokládám dopadení	pachatele	za příliš tvrdý oříšek pro policii , a proto jsem
jde na procenta s trojčlenkou nebo integrální rovnicí , je	debil	! " zařvala jsem bezmocně . " Poslyš , "
7 definitivně s kariérou rozloučí Jiří Novák , který dostane od	pořadatelů	divokou kartu . Zlínský rodák , který však svůj tenisový
dole , my tu na vás počkáme , " odpověděl	velitel	. Mladíci rychle seskákali z koní , přivázali je a

Přegenerování

		lemma	Frekvence
1.	p/n	ředitel	24 415
2.	p/n	přítel	23 968
3.	p/n	obyvatel	19 459
4.	p/n	manžel	17 278
5.	p/n	Pavel	16 942
6.	p/n	majitel	15 626
7.	p/n	Karel	15 597
8.	p/n	král	14 744
9.	p/n	učitel	10 943
10.	p/n	podnikatel	9 283
11.	p/n	uživatel	8 373
12.	p/n	spisovatel	7 544
13.	p/n	nepřítel	7 377
14.	p/n	Michal	7 146
15.	p/n	představitel	6 912
16.	p/n	velitel	6 557
17.	p/n	zaměstnavatel	5 810
18.	p/n	pachatel	5 612
19.	p/n	generál	5 449
20.	p/n	zastupitel	5 086

Chyby v disambiguaci

Dobry , Andy , jen se nerozčiluj . " Andy brblal /brblal/NNMS1-----A----- . Prešli jsme cestu a vratili se do vozu .
jsem vysvětlení . " Dnes mám mít volno , " brblal /brblal/NNMS1-----A----- . " Sejdeme se v hospodě ? V kolik ?
ninut promluvit s Goshenem o jedné bezpečnostní záležitosti . Hackett brblal /brblal/NNMS1-----A----- a reptal , že návštěva po úředních hodinách je proti
Všechno je to zralé na vyhození z okna , " brblala /brblal/NNMS4-----A----- , když vycházela z kanceláře . " Když je den
" A moje sestra je pořád stejně hnusná , " brblal /brblal/NNMS1-----A----- Billy a , jak jsem si všimla , dál roztržitě
Bych ráda věděla , nad čím tak johohuje ! " brblala /brblal/NNMS4-----A----- Frankie a přitom pohledem vraždila laminátového piráta v životní velikosti
taky musíme bydlet s Feargalem a Majklem D , " brblal /brblal/NNMS1-----A----- Warren a pořádkoval špičkou nohy na trávníku . No dobře
" Páni , musím si sehnat další náplasti , " brblala /brblal/NNMS4-----A----- Kyra a shrnovala si ponožku , aby si ověřila rozsah
schody . " Copak je ten krám živý ? " brblala /brblal/NNMS4-----A----- otráveně , protože svinutá hadice pořád utíkala a nedovolila jí
jí tak , když je to taková ježibaba , " brblala /brblal/NNMS2-----A----- Kyra . " Hele , jestlipak víte , s jakým
. " Ty si to aspoň můžeš oholit , " brblal /brblal/NNMS1-----A----- Pete . " Ty budeš zase slavnej , " odsekla
No hele , já teda nevím . . . " brblala /brblal/NNMS4-----A----- jsem váhavě . " Možná tam budem spíš překážet .
rebarborového koláče ! " " Hráť , jo ? " brblal /brblal/NNMS1-----A----- Clark . " Copak jsme dvouletý mimina ? " "
atakďál . " " Já jim prostě nevěřím , " brblal /brblal/NNMS1-----A----- Ford a vrtěl hlavou . " Zastřílíme si zas trochu
včely , jako kdyby už teď na něj chodily zástupy brblalů /brblal/NNMP2-----A----- s požadavky na údržbu . " Ano , " řekl
jsem . " No to je vážně zajímavé , " brblala /brblal/NNMS2-----A----- nespokojeně babička . " A já myslela , že Billy
mi tedy můžete ukázat ? " " Závěry , " brblal /brblal/NNMS1-----A----- . " Vypracovaná hlášení . " " Chcete říct padělky
ti povím , fakt mě už začíná vytáčet ! " brblala /brblal/NNMS4-----A----- Sandie , když jsme se obě ploužily k ragbyovému mlýnu

Dílčí závěry

- Přegenerování
- chyby v disambiguaci
- potřeba jiného postupu při vyhledávání relevantních dat z korpusu

Zlepšení přesnosti (precision)

- Snaha eliminovat maximum false positives
- Snížit přegenerování

Analýza přegenerovaných dokladů a reformulace dotazů

- Mohou I-ová příčestí v češtině končit např. na řetězce *tel, rel, vel, ael, ...*?
- Jaké další (nezahrnuté) formální vlastnosti musí splňovat hledané jednotky?

Mohou l-ová příděstí v češtině končit např. na řetězce *tel, rel, vel, ael, ...*?

- a) mohou a v korpusu jsou doložena
- b) mohou, ale v korpusu nejsou doložena
- c) nemohou a tudíž v korpusu nejsou doložena

Analýza dat

- Korpus cql
- **[lc="((.*[atrv]el) | (.*[atrv]el[aoiy]))"]**
- P-filtr
- Korpus > Implicitní atribut tag
- **V[pq].***
- Konkordance > Statistika Frekvenční distribuce

((.*[atrɔ]el) | (.*[atrɔ]el[aoi y]))

hotelu Piccadilly , který ani zdaleka nebyl tak módní jako její špičce ještě daleko . Ale zároveň nepřestával mít na nejvýše 30 hodin týdně při šestihodinové denní pracovní době .	hotely /hoteɫ/NNIP1-----A-----	v ulicích vedoucích z Jeffersonovy třídy a t
Regionálního operačního programu Moravskoslezsko . Moravskoslezský kraj je jako úspěšný	zřeteli /zřeteɫ/NNIS6-----A-----	důvod jejich schůzky : bonbonek v podobě
To jsi ji nechal jít samotnou ? " otázal se	Novela /noveɫa/NNFS1-----A-----	umožní práci přesčas i zaměstnancům , kt
třínohou stoličku a druhou mi přišoupl nohou . " Můj	žadatel /žadateɫ/NNMS1-----A-----	o dotace použije na výstavbu jižního obchv
má tetování : ukřižovaného Krista mezi námořnickými motivy . Jeho	Cadfael /cadfael/NNMS1-----A-----	s hraným pohoršením . " Ovšem že ne ! Za
. Není se co divit , kousek odtud se rozprostírá	přítel /příteɫ/NNMS1-----A-----	hledá byt , a poněvadž jste si mi stěžoval ,
, je téměř vyloučeno , aby si z nich schopný	přítel /příteɫ/NNMS1-----A-----	byl světlolvasý a hubenější , se světlou ple
a pokud každá nově odhalená skutečnost jen posilovala v mém	favela /faveɫa/NNFS1-----A-----	zvaná " Horká díra " , jež s bohatou čtvrtí
na osobu blízkou . Autoři návrhu chtějí , aby měl	pozorovatel /pozorovateɫ/NNMS1-----A-----	neudělal přesný obraz . " " To je ale nesmys
" To je naprosto vyloučeno ! Všichni už jsme v	příteli /příteɫ/NNMS6-----A-----	presvědčení o mládežnické nevině , neztrác
prosím , " řekl Holmes . " To je můj	majitel /majiteɫ/NNMS1-----A-----	vozidla povinnost sdělit vyjmenovaným úřa
dělat se smrtí Josepha Stangersona . " " Trpělivost ,	posteli /posteɫ/NNFS6-----A-----	a je skoro půlnoc . " " Tak zítra .
z mnoha biotopů v jejich širokém areálu rozšíření . Typickými	přítel /příteɫ/NNMS1-----A-----	a spolupracovník doktor Watson , který mi
Andrey Verešové (27) , stojí její manžel ,	příteli /příteɫ/NNMS3-----A-----	, trpělivost ! Brzo poznáte , že to s ní
	obyvateli /obyvateɫ/NNMP7-----A-----	pralesa jsou dvě krásné , navzájem si velm
	podnikatel /podnikateɫ/NNMS1-----A-----	Daniel Volopich . A kvůli svému těhotenstv

Výsledek frekvenční distribuce lemmat

	lemma	pos	Frekvence
1.	p/n ředitel	N	15 764
2.	p/n Pavel	N	12 073
3.	p/n přítel	N	12 029
4.	p/n obyvatel	N	10 427
5.	p/n Karel	N	9 630
6.	p/n kostel	N	8 641
7.	p/n postel	N	8 248
8.	p/n majitel	N	7 017
9.	p/n učitel	N	4 215
10.	p/n Michael	N	4 074
11.	p/n hotel	N	3 973
12.	p/n velitel	N	3 789
13.	p/n nepřítel	N	3 232
14.	p/n podnikatel	N	2 990
15.	p/n spisovatel	N	2 976
16.	p/n pachatel	N	2 543
17.	p/n zaměstnavatel	N	2 442
18.	p/n Izrael	N	2 419
19.	p/n novela	N	2 120
20.	p/n uživatel	N	1 949

? pos

Víceúrovňová frekvenční distribuce

Frekvenční limit:

Úroveň	Atribut	Nerozlišovat velikost	Pozice	(Node) začít od ?	
1.	<input type="text" value="pos"/>	<input type="checkbox"/>	<input type="text" value="KWIC"/>	<input type="text" value="slova KWIC nejvíce vlevo"/>	-

[Přidat úroveň ↓](#)

Vytvořit frekvenční seznam

Celkem: 4 (1 str.)

	<u>pos</u>	<u>Frekvence</u>
1.	p/ n N	167 777
2.	p/ n X	2 208
3.	p/ n V	19
4.	p/ n A	5

? V

a zastrč . A kdyby nešel , tak mu holt	ustel /ustlat/Vi-S---2--A---P	peřinu na faulenc . Dobrou noc , jo ? "
... " Júj . . . Poslyš ,	Stel /stlát/Vi-S---2--A---I	, vůbec jsem si nevšimla , kolik je hodin .
spát ? V hotelu ? " Třeba si aspoň	ustel /ustlat/Vi-S---2--A---P	tady na sedačce . . . Ne , sem si
já leknutím málem vyletěla z kůže . " Čauky ,	Stel /stlát/Vi-S---2--A---I	! To jsem já , Lauren ! " začířikal mi
Řek , " prohlásil Alistair a zvedl se . "	Ustel /ustlat/Vi-S---2--A---P	mi postel , a lampu na stůl dej , přijdu
i já srdečně poděkoval . Až se dům prodá a	Tetel /tetelit/Vi-S---2--A---I	bude mít spolehlivého poručníka , až už tu Mileva nebude
cestičku před svým hrobem . Děkujeme . " Stolečku ,	ustel /ustlat/Vi-S---2--A---P	si ! u / Něco potřebuješ ? Podrž mi to
blesků všech , v hlubinách uspi světlá pacholátka a pocestnému	ustel /ustlat/Vi-S---2--A---P	suchý mech ! A propast úzkosti své , země vratká
f. du quartier (franc. [franšíz d l '	otel /otelit/Vi-S---2--A---P	, dy kartié] ; od poplatků osvobozený úřad)
, proč byli Jan van Riebeeck a Simon van der	Stel /stlát/Vi-S---2--A---I	tak hodní a lord Charles Somerset tak zlý . Také
to , abychom ho utráceli čtením podobných knih . (Stel /stlát/Vi-S---2--A---I	Pavlou : Gene . Přel . Jana Rečková . Praha
rdny , kterou založil první nizozemský koloniální správce Simon der	Stel /stlát/Vi-S---2--A---I	jako základnu East Indian Company . Vystupuji na peron zalitý
nápoje . hôtel de l ' ambassadeur (franc. [otel /otelit/Vi-S---2--A---P	d lambasader]) - příbytek velvyslance , viz místnosti
užíván pojem hôtel de l ' ambassadeur , franc. [otel /otelit/Vi-S---2--A---P	d lambasader]) - prostory , ve kterých se
*** JEN POČKEJ , TROJSKÝ ZAJÍCI !	Stel /stlát/Vi-S---2--A---I	Pavlou je v našich knihkupectvích novým jménem , nakladatelství Triton
vrchní si prozpěvuje obrať se zády ke všemu co bylo	ustel /ustlat/Vi-S---2--A---P	si na slunci a čekej závěje obrať se zády ke
naučil Prahu mluvit . . . , no nic ,	ustel /ustlat/Vi-S---2--A---P	si . " Za pár dní bylo Vladimíra všude plno
Následovaly jej například kapely Landmine Spring , Buster Nixon či	Nestel /stlát/Vi-S---2--N---I	Postel . Příznivci alternativní hudby si večer vychutnali kapelu Vladivojna
jejich okraj měl úhledně ohrnuté šedivé teplé podkolenky . "	Stel /stlát/Vi-S---2--A---I	, my na tebe čekáme , " řekl . "

P-filtr V[qp].*

- Pouze **Vi.***
- **Většinou chyby v disambiguaci/slovníku**

[vložit tag](#) | [vložit "within"](#) | [klávesnice](#) | [předchozí dotazy](#)

```
[tag="V[qp].*"]|
```



Dotaz | Korpusy | **Uložit** | Konkordance | Filtr

Výskytů: 0 | i.p.m. [?]: 0,00 (vztaženo k celému syn2010) | Výsledek je promíchán

Výsledky

- Výsledkem frekvenční distribuci **lemmat** je prázdný seznam.
- Tudíž by mělo platit **b)** nebo **c)**.

Platí b) nebo c)?

- b) mohou, ale v korpusu nejsou doložena
- c) nemohou a tudíž v korpusu nejsou doložena

Co stojí zpravidla před tvarovou koncovkou *-l*?

- Všechny samohlásky a souhlásky (lépe všechny grafémy).
- Pouze některé samohlásky a souhlásky (lépe všechny grafémy).

Implicitní atribut tag

$V[pq][MI]S.*$

	lc	Frekvence
1.	p/n byl	240 868
2.	p/n řekl	104 307
3.	p/n měl	104 290
4.	p/n mohl	46 771
5.	p/n nebyl	30 142
6.	p/n chtěl	29 144
7.	p/n začal	27 742
8.	p/n musel	23 183
9.	p/n dostal	22 283
10.	p/n uvedl	21 593
11.	p/n přišel	19 881
12.	p/n viděl	19 792
13.	p/n neměl	19 024
14.	p/n dal	18 717
15.	p/n zeptal	18 148
16.	p/n věděl	16 432
17.	p/n stal	15 888
18.	p/n dodal	15 445
19.	p/n nemohl	14 987
20.	p/n říkal	14 104

Které se vyskytují u tvarů sloves nejčastěji?

- [aá] : začal, dál
- [ěe] : měl, (šel)
- [i] : prohlásil
- [u] : ????, (tisknul)
- [y] : byl
- [dhk] : uvedl, mohl, řekl
- [???) : (kopl)

Které se nevyskytují vůbec?

Predikce:

- Dlouhé samohlásky?
- Samohláska **o**?
- Další souhlásky?

Zdůvodnění z gramatiky

- Kmenotvorné přípony kmene minulého jsou pouze některé krátké vokály kromě *stál*, *bál se se* dlouhé *á* infinitivní v I-ové přičestí pravidelně krátí a totéž platí o dlouhém *í*.
- III. třída slovesná s otevřeným kořenem (*krýt*) pravidelně krátí kořenový vokál v I-vých přičestích.
- I. a II. třída slovesná s uzavřeným kořenem bez kmenotvorné přípony /s 0 kmenotvornou příponou (*nést*, *péci*, *tisknout*).
- Kmenotvorná přípona -e- a kořenový vokál e.

Kmenotvorná přípona *-e-*

- I. třída (*umřel*)
- IV. třída (*trpěl, sázel*)

Kořenový vokál *e*

- III. třída (*kryl*)

Co v češtině nutně předchází pře *e*?

- Před KmV -e-: [cčjlrššzž]: *kácel, klečel, nezapíjel, mlel, bolel, musel, slyšel, odřel, vytvářel, zmizel, ležel*
- Před KoV -e-: [lsz]: *zaklel, nasel, zel*

Empirie a systém jazyka

- Aby mohlo být substantivum konvertovaným l-ovým příčestím, pak
- může před / předcházet pouze [aáeěiuý];
- před / nemůže předcházet [bpfvmdtnkghr]e;
- před / nemůže předcházet [aáeěííóóuůyý][aáeěiuý].

Chybná desambiguace:

Jaké další (nezahrnuté) formální vlastnosti musí splňovat hledané jednotky?

- `[lc=".*[aáeěiuý]l((ové) | (ů) | (ům) | (ech) | (e) | (ovi))"]`
- N-filtr
- `[lc=".*[bpfvmdtnkg hr]el((ové) | (ů) | (ům) | (ech) | (e) | (ovi))"]`
- N-filtr
- `[lc=".*[aáeěéííoóuůýý][aáeěiuý]l((ové) | (ů) | (ům) | (ech) | (e) | (ovi))"]`
- P-filtr
- `[tag="N.M.*"]`
- Konkordance > Statistiky > Frekvenční distribuce

Výsledky vyhledávání v SYN2000

, ale nezjistili nic nového a také paní Smithová o	manželovi /manžel1/NNMS6-----A-----	nic neví . Co nevidět dojdou k názoru , že
divokou bitvu mezi zhruba tisícovkou policistů a šesti sty pravicových	radikálů /radikál/NNMP2-----A-----	. Její následky nelze v betonovém městě přehlédn
hottsreicha Sigismonda z Ormsteinů , velkovévodu z Cassel-Falsteinu a dědičného	krále /král/NNMS4-----A-----	české země . " " Zajisté ovšem chápete , "
na nich postaveny paláce , pevnosti či výletní sídla dávných	králů /král/NNMP2-----A-----	a královen , nebo - a to nejčastěji - skrývají
, ať mají na sobě jakýkoli kostým . Vždycky bývali	králové /král/NNMP1-----A-----	a šašci , i milenci a matky stále zůstávají ,
v Československu . Tato zpráva nepochybně potvrzovala výhrady aparátu vůči	Nezvalovi /Nezval/NNMS3-----A-----	a mohla pomoci oslabit některé nepohodlné tóny
222 miliónů . " Velmi výrazně se projevilo společné zdanění	manželů /manžel/NNMP2-----A-----	, " řekl Právu ředitel plzeňského finančního úřadu
dechnutí , jen výstup kratičký mu dopřává , by na	krále /král/NNMS4-----A-----	si hrál a strašil , zabíjel svým pohledem . .
řekl Holmes . " Je to malý dáreček od českého	krále /král/NNMS2-----A-----	jako pozornost za pomoc , kterou jsem mu posky
výjimka . Autory všech předchozích knih a článků o Karlu	Krylovi /Kryl/NNMS6-----A-----	překonává především množstvím faktů , jejich pře
zcela novou kvalitu , na osamostatněném veletrhu budou mít logističtí	profesionálové /profesionál/NNMP1-----A-----	koncentrované možnosti . CEMAT nabídne i mno
na průmyslové zboží . Na své si tak přijdou domácí	kutilové /kutil/NNMP1-----A-----	a řemeslníci . Opilý řidič havaroval , je těžce zran
třeba Kanibalové ? Zní to povědomě ? Aby ne !	Kanibalové /kanibal/NNMP1-----A-----	jsou totiž považováni za jeden z nejrozšířenějších
slavnostnímu otevření nové budovy opery v Oslu . Vedle norského	krále /král/NNMS2-----A-----	Haralda V. nechyběla německá kancléřka Angela
vyjedou jako obvykle do parku , 'řekla paní Irena	manželovi /manžel/NNMS3-----A-----	a odjela . Nic víc jsem neslyšel . Rozjeli jsme
před čtrnácti dny se hřbitovem v Hostíně u Vojkovic prohnali	vandalové /vandal/NNMP1-----A-----	a zničili , co se dalo . Policisté už zjistili
května bude hlavním bodem oslav svěcení pramenů a průvod družiny	krále /král/NNMS2-----A-----	Karla IV . Už od půl desáté vystoupí na Mlýnské
pravidelné kroky strážníka a prozpěvování a křik nějaké opožděné skupinky	hýřilů /hýřil/NNMP2-----A-----	. Po obloze pomalu plynul těžký mrak a z potrh

frekvenční distribuce lemmat

25.	p/n	čumil	98
26.	p/n	konzul	95
27.	p/n	federál	93
28.	p/n	kanibal	85
29.	p/n	Michal	83
30.	p/n	Dale	80
31.	p/n	debil	72
32.	p/n	Cale	69
33.	p/n	kutil	68

kutil

na průmyslové zboží . Na své si tak přijdou domácí	kutilové /kutil/NNMP1-----A-----	a řemeslníci . Opilý řidič havaroval , je těžce zraněn
integrováným akumulátorem . (www . blackanddecker . com)	Kutilové /kutil/NNMP1-----A-----	uvítají i vyznamenaný jednovrstvý krycí lak Aura , vyvinutý společností
najitelé nadnárodních zbrojařských komplexů . Proti nim stojí parta šikovných	kutilů /kutil/NNMP2-----A-----	trvale udržitelného rozvoje . V nespravedlivém , globalizovaném světě
potrvá v Litoměřicích do neděle . Zahrádkáři , chalupáři i	kutilové /kutil/NNMP1-----A-----	mají v areálu výstaviště z čeho vybírat . Většinu pavilónů
" " To je ta černá skvrna na pověsti všech	kutilů /kutil/NNMP2-----A-----	a bývalý pubertální děs neboli tvůj otec ? " šeptla
vytřídí jen dva . ČESKÉ BUDĚJOVICE Češi jsou stále národem	kutilů /kutil/NNMP2-----A-----	, kteří raději spotřebiče donekonečna opravují , než aby se
dokáže říct s horkým bramborem v puse . . Digitální	kutilové /kutil/NNMP1-----A-----	V Brně začala přehlídka digitální kultury Zoom 2005 Poslední týden
čerpady překypuje více , než dobře udržovaná zahrádka květy .	Kutilové /kutil/NNMP1-----A-----	, kteří chtějí něco ušetřit a přesto mít ve svém
patřily panu Jonesovi - " " Tisíc užitečných rad domácím	kutilům /kutil/NNMP3-----A-----	" " , " " Každý sobě zedníkem " "
nesnesitelný . Takto však , čistě omyt , byl jsem	kutilovi /kutil/NNMS3-----A-----	vítán . Předvedl mi všechny své mašiny a jejich rozmanité
a bydlení - Euroregion Hobby 2007 . Prodejní akce určená	kutilům /kutil/NNMP3-----A-----	, zahrádkářům a chalupářům začala ve čtvrtek 26 . dubna
jakési vznešenosti . Jako u bubnovacího stroje , jenž tomu	kutilovi /kutil/NNMS3-----A-----	posloužil nejspíš jako model , visel Greff a jeho protiváha
než krouceniny v mozku . My Němci jsme už zkrátka	kutilové /kutil/NNMP1-----A-----	. Dělejte proti tomu něco ! Bebra : Teď i
nastěhovali sem , byl odhodlaný stát se vzorem všech domácích	kutilů /kutil/NNMP2-----A-----	. Jenom jsem si dělala starosti , že nakonec budeme
benzinových pump budou u silnic vodní stanice . Osobně našim	kutilům /kutil/NNMP3-----A-----	věřím . Pokud jsou v krizi , udělají pro sebe
(do-it-yourself , tedy udělej si sám nebo volně přeloženo	kutilové /kutil/NNMP1-----A-----	, pozn. red.) , tak malé a střední firmy
pět prodejen , do kterých denně zamíří tisíce profesionálů a	kutilů /kutil/NNMP2-----A-----	. Správa takové organizace je složitá , zvláště když velká
Českém Krumlově Lubomír Veselý . Poznamenal , že počet těchto	kutilů /kutil/NNMP2-----A-----	se podle údajů z policejních kontrol snížil , ale nikdo
Kovljevové , který byla doma obviněna z napomáhání loupežní	Kutilové /kutil/NNMP1-----A-----	Sloupek Vždycky jsem byl proti tomu , aby se seriózní
Do té doby zůstane bez světla a topení . Nebezpeční	kutilové /kutil/NNMP1-----A-----	Po zásahu naskakuje skupina do aut a odjíždí k dalšímu
zatáhnout i mnozí tatínkové , přičemž vyzkoušejí své schopnosti domácích	kutilů /kutil/NNMP2-----A-----	- mohou nově vymalovat dětský pokojík nebo třeba zhotovit kolébku
Českém Krumlově Lubomír Veselý . Poznamenal , že počet těchto	kutilů /kutil/NNMP2-----A-----	se podle údajů z policejních kontrol snížil , ale nikdo
zařizována , proto podle architektů zvláště v menších bytech díky	kutilům /kutil/NNMP3-----A-----	vznikají Potěmkinovy vesnice : vánoční výzdoba totiž ledacos zakryje

Čumil/ ? čuměl

policejní mašinérie , obojí totiž příliš dobře znala : morbidní	čumilové /čumi1/NNMP1-----A-----	toužili aspoň na okamžik zahlédnout zkrvavenou mrtvolu , policis
jsme se přepravili davem nakupujících , obchodníků i všelijakých podezřelých	čumilů /čumi1/NNMP2-----A-----	, kteří mlsně okukovali naše kufrы a batohy , bezmála
že odlet vrtulníku zhruba za pět minut přiláká ještě víc	čumilů /čumi1/NNMP2-----A-----	. Listy rotoru se dál s pleskáním pomalu otáčely ,
Bourne trhl hlavou doleva , kde viděl za okraj řady	čumilů /čumi1/NNMP2-----A-----	. Naskytl se mu nejodpornější výjev večera plného násilí .
obdivuje můj vyvinutý nosánek , česky zvaný frňák . Takových	čumilů /čumi1/NNMP2-----A-----	jsem už potkal za život dost . Kašlu na něj
! Vykusala jsem z kabiny a vystoupila na palubu .	čumilovi /čumi1/NNMS3-----A-----	se náhle vrátila schopnost pohybu . " Madam ! "
přítom se sem teď od rána do večera táhnou davy	čumilů /čumi1/NNMP2-----A-----	. Kdeko najednou hrozně touží aspoň si náš motel vyfotit
telefon . Otočil jsem se a pomalu kráčel k hloučku	čumilů /čumi1/NNMP2-----A-----	. Vrhli kritický , nepřátelský pohled na můj zřízený obličej
tím ? " Za ním se shromáždil ještě větší hlouček	čumilů /čumi1/NNMP2-----A-----	než posledně . Cara některé obličeje poznává , tváří se
strategie . Záměr mu zkaltí nejen dav náhle se objevivších	čumilů /čumi1/NNMP2-----A-----	, ale i televizní hlasatelka Agáta (Anny Dupereyová)
68 . ulici . Naproti přes ulici se shromáždil hlouček	čumilů /čumi1/NNMP2-----A-----	, ale na téhle straně udržoval uniformovaný polda občany v
hmotou , jakou byla studená plasma . Ani tady nechyběli	čumilové /čumi1/NNMP1-----A-----	a radilové . Je to zvláštní druh fanoušků , kteří
lodí zboží všeho druhu za " vydatné pomoci " všelijakých	čumilů /čumi1/NNMP2-----A-----	s rukama v kapsách či s lahví pálenky a taxikáři
ze dveří . " Teď už víš , co se	čumilům /čumi1/NNMP3-----A-----	stává , " prohlásil jsem . Mezitím se trochu vzpamatovala
na poštu další , okamžitě se zařadil do toho davu	čumilů /čumi1/NNMP2-----A-----	. Společenství jim dávalo sílu , aby neodkrouhli šéfy .
zvládnout tenhle případ . Přemítal , zda někdo z těch	čumilů /čumi1/NNMP2-----A-----	už Nettii identifikoval . Jestliže ano , tak se to
odpoledne , vpustíte dovnitř nějaké dva tisíce nebo i víc	čumilů /čumi1/NNMP2-----A-----	, které vyvolávací správně navnadí , aby vyhodili dva dolary
Dvojí klapnutí se odrazilo ozvěnou od klenby a pár posledních	čumilů /čumi1/NNMP2-----A-----	, kteří setrvali dosud ve dveřích , rychle zmizelo .
žalostnější než Gogolův hlavní aktér . . . A náhodní	čumilové /čumi1/NNMP1-----A-----	budou říkat , proboha , cožpak je masopust . .

čmucha1

přeložila si dolní ret přes horní . " Hele , **čmucha1** /čmucha1/NNMS5-----A----- , dám ti radu . Vypadni odtud , dokud se pozornost . Řekl tiše : " Co se děje , **čmucha1** /čmucha1/NNMS5-----A----- ? Žádnej kšeftík na ty zatracený lodi ? " "

koncem a smyčkou . Proč sem jenom lez . Ti **čmucha1** /čmucha1/NNMP1-----A----- mu všechno přišpendlí a ještě se o něm ledacos dozví chce možná skončit . Pro začátek na to nasadím pár **čmucha1** /čmucha1/NNMP2-----A----- a uvidíme , s čím přijdou . Budu chtít mluvit " . Cítil však , že Dewey a někteří další **čmucha1** /čmucha1/NNMP1-----A----- téhle historce moc nevěří , jako by se jim zdálo když se před pár dnama kousek od něho ti dva **čmucha1** /čmucha1/NNMP1-----A----- schovávali . Jenomže , milej pane , on má voči bez kňourání a naříkání a běhání za psychiatra . Kapradinovi **čmucha1** /čmucha1/NNMP1-----A----- jsou všichni blbci , ale někteří z nich - a během týdne po svém příjezdu vysloužili všeobecnou nevráživost a přezdívku **čmucha1** /čmucha1/NNMP1-----A----- . Další z jejich charakteristických vlastností však byla houževna pěstí do břicha . Ten menší ji znásilnil , řekla **čmucha1** /čmucha1/NNMP3-----A----- z oddělení vnitřních záležitostí . Ten se širokými rameny se další tah , ten , co udělá , jakmile ti **čmucha1** /čmucha1/NNMP1-----A----- odejdou . " Poslal jste mu šek na sto tisíc . Ať už tyto instituce vedl jeden z těch kapradinových **čmucha1** /čmucha1/NNMP2-----A----- nebo celý jejich výbor , bylo jasné , že je Schucmanové . Svou verzi Markova vyprávění jsem kvůli té bandě **čmucha1** /čmucha1/NNMP2-----A----- uschoval na bezpečné místo , kde ji nikdo nenajde . . Díky . " " Koukej mě zásobovat novinkami , **čmucha1** /čmucha1/NNMS5-----A----- - tohle má totiž být můj případ . " "

Sam . " Kampak ? Na lov ? Koukej , **čmucha1** /čmucha1/NNMS5-----A----- , tobě nechutná naše strava a já bych se nijak k velikému vzteku posádky lehkého letounu - pokud to byli **čmucha1** /čmucha1/NNMP1-----A----- . Žijeme krásnou dobu , uvažoval Golan . Lidé jako nějaký únik z místnosti , který dokážou najít dva obyčejní **čmucha1** /čmucha1/NNMP1-----A----- ? Neřekl to ovšem nahlas . Ale tím , že nenosí tlumiče , blesklo mu hlavou . " Ty zasranej **čmucha1** /čmucha1/NNMS5-----A----- , nadělal jsi mi tolik nepříjemností , že bych ti novinářů ? " Šéf KGB pokrčil rameny . " Ti **čmucha1** /čmucha1/NNMP1-----A----- se dostanou všude . " Ano , pokud jim někdo smyčkou . Ksaku , proč sem jenom lezl . Ti **čmucha1** /čmucha1/NNMP1-----A----- mu všechno přišpendlí a ještě se o něm ledacos dovědí

? mýval

ruch nočních tvorů -- třepotání křídel netopýrů a šustot **mývalů** /mýval/NNMP2-----A----- , kradmé , tiché kroky skunků . Měl vyprahlé hrdlo
dítě před třiceti čtyřiceti lety běžně vyrůstalo s teoretickou znalostí **mývalů** /mýval/NNMP2-----A----- , svišťů , deněk , pytlonošů , stromů hickory ,
podobně savci zvlášť menší druhy , jako jsou nosálové , **mývalové** /mýval/NNMP1-----A----- , fretky hlodavci psouni , pískomilové a různé druhy myši
to myslel celý večer , v noci se mu o **mývalovi** /mýval/NNMS6-----A----- zdálo a druhý den odpoledne žadonil tak dlouho , až
do pasti chytila jiná zvířata , než měla , například **mývalové** /mýval/NNMP1-----A----- a skunci , a umístění v blízkosti tábora podobné tvory
Podle B. Rensche (1957 , 1958) opice , **mývalové** /mýval/NNMP1-----A----- a různí ptáci upřednostňovali pravidelné objekty před nepravidelnými , s
tom nic ani bývalý proděkan plzeňských práv a známý chovatel **mývalů** /mýval/NNMP2-----A----- Milan Kindl , který hřímá na svých webových stránkách ,
kávy . Když konečně dopluli k obchodu , ti dva **mývalové** /mýval/NNMP1-----A----- je pozdravili svým zvláštním brucením , a biskup je nakrmil
totiž bydlel kromě dvou lidí také pes a několik medvědů **mývalů** /mýval/NNMP2-----A----- . Když v roce 2002 byt s rozlohou 120 metrů
a podkládají další vrstvou tkaniny . al návštěva Když odešli **mývalové** /mýval/NNMP1-----A----- Majitelka atypicky velkého městského bytu v domě z třicátých let
" PLYNOVÝ ÚTOK ! " řvali jsme , až vyděšení **mývalové** /mýval/NNMP1-----A----- rozsypali naše čokoládové tyčinky . " BUĎTE ZTICHA ! "
na palouku u naší chaty za New Yorkem , o **mývalech** /mýval/NNMP6-----A----- , o psech , kteří tam někomu místnímu patřili ,
pro čilé malpy riskantní záležitost , protože nosálové - příbuzní **mývalů** /mýval/NNMP2-----A----- - jsou zvířata vybavená mohutnými čelistmi a dvakrát větší než
papiříčkám s vysokým obsahem capsaicinu se vyhýbají například lišky a **mývalové** /mýval/NNMP1-----A----- . Ptáci , kteří nemají ty správné receptory pro pálivé
výsledek . Papoušci , holubi , krysy , delfíni , **mývalové** /mýval/NNMP1-----A----- a samozřejmě primáti jsou schopni rozlišit zrakové a sluchové podněty
se tu prý ještě najde . Krysy . Veverky . **Mývalové** /mýval/NNMP1-----A----- . " " Já jde . Já podívá , "
Vztekle jsem vzal pušku a baterku a zabil prvních šest **mývalů** /mýval/NNMP2-----A----- , které jsem našel . O nic lépe jsem se
krmili jsme ho tajně , protože se to nemělo . **Mývalové** /mýval/NNMP1-----A----- jsou škodná . Ale my krmili a on nám věřil
zatímco jejím manžel pomáhal Jimovi s naftou . Dva malí **mývalové** /mýval/NNMP1-----A----- , kterým vydávali zvláštní bučivém zvuky , prosili Marka o

? bacil

k celému syn2010) | ARF^u : 2,30 | Výsledek je promíchán

Mávnete - li nad prvními příznaky nachlazení rukou , dáte	bacilům /bacil/NNMP3-----A-----	šanci . Velmi brzy si najdou cestu do dalších oblastí
- pozn. překl . Bezpečnější sex - jak se vyhnout	bacilům /bacil/NNMP3-----A-----	lásky Pohlavně přenosnými nemocemi je ohrožen každý z nás .
- v džusech a ovoci - zvýší vaši odolnost proti	bacilům /bacil/NNMP3-----A-----	. Cvičení zvyšuje teplotu těla a je proto přírodním "
bázi , aby ses vyhnula rozedření pochvy . Odřeniny usnadňují	bacilům /bacil/NNMP3-----A-----	proniknout tam , kde nemají co dělat . Pošli svého
musím upozornit , že prostředí v blízkosti prádla je plné	bacilů /bacil/NNMP2-----A-----	" . Prohlídka se pak zastavila . Gestapáci : "
všechny do Lucerny ? Když jsme měli výročí pětadvaceti let	Bacilů /bacil/NNMP2-----A-----	, pozval jsem všechny . Na jevišti jsme měli troje

? *serval*

- poslední čtyři v letošním roce na jaře . Malých

servalů /serval/NNMP2-----A-----

se tedy druhý den po porodu uja

gepard udržuje rovnováhu . &figure; Skoč si pro oběd !

Servalové /serval/NNMP1-----A-----

loví celou řadu tvorů . Tento níz

Zoo : prohlídka pavilónů a galerie 11.20 - křtiny mláďat

servalů /serval/NNMP2-----A-----

11.45 - oběd na Staré radnici 13

fňukal

nevyňikal v žádném sportu . Vlastně patřil do skupiny "	fňukalů /fňuka1/NNMP2-----A-----	, cvoků a tloušťků " , kteří všechny míčové hry
kráček s potravinami , jehož majitel dělá navíc i službu	fňukalům /fňuka1/NNMP3-----A-----	. Někdo si přijde postěžovat , jiný se potřebuje svěřit

Ruční analýza dat 627 lemmat

- Budeme zvlášť analyzovat apelativa a propria
- P-filtr
- **[lemma= "[[:upper:]].*"]**
- **459 lemmat mezi nimiž budou propria (Nezval)**
- **627-459=168**
- **Čumil, kutil, mýval, ožrala, čmuchal, břídil, narval, šibal, patolízal, šfoural, žvanil, slídil, střádal, packal, reptal, bacil, loudal, hýřil, rejpal, chlubil, tlachal, rýpal, všeuměl, kecal, kýval, serval, mazal, loudil, radil, brblal, všudybyl, prášil, škrabal, bouchal, skuhral, fňukal, šukal, cmrndal, cintal, koktal, hloubal, šeptal, čmáral, hrabal, škudlil, klábosil, mlsal, zlobil.**
- **41+3**

Zlepšení pokrytí (recall)

- Snaha dohledat false negatives
- Snížit podgenerování

Jak získat další doklady

- Vrátime se zpět před krok, na němž jsme užili tag
- P-filtr
- **[tag="X.*"]**

1542 ?lemmat

Celkem: 1542 (31 str.)

		lemma	Frekvence
1.	p/n	Kennerdale	249
2.	p/n	Générale	111
3.	p/n	Abagnale	75
4.	p/n	urgalové	73
5.	p/n	Cantabile	52
6.	p/n	urgalů	52
7.	p/n	Calovi	52
8.	p/n	odstále	47
9.	p/n	freestyle	47
10.	p/n	fatale	45
11.	p/n	Kyle	42
12.	p/n	Ferndale	41
13.	p/n	Émile	32
14.	p/n	Freestyle	31
15.	p/n	nále	30
16.	p/n	plížilů	30
17.	p/n	Urgalové	30
18.	p/n	eldilové	28
19.	p/n	Module	25
20.	p/n	Scarsdale	24

Další doklady

- N-filtr
- **[lemma= "[[:upper:]].*"]**
- 707 ?lemmat
- plížil, plazil, ?řinčila, ?muchal, čadil, remcal, brouzdal, ?ospalové, kukal, babral, drnkal, kousal.
- 9

plížilů

eno	Seznam slov...	F [?] : 1,02 Výsledek je promíchán
	podívala na Zlomený štít . " Tady máš veletucet vajíček plížilů /plížilů/x@-----	Zebu . Patří imperátorovi . Za dvaasedmdesát otáček se vrátí
	Potěšená kapsa a inženýr Kovoohýbač společně vyvinuli stroj vyrábějící maso plížilů /plížilů/x@-----	, které bylo od pravého k nerozeznání . " Připadá
	Kromě tří párů , pestře zbarvených , ale naprosto tupých plížilů /plížilů/x@-----	chovaných na maso různých chutí , měla i několik cvičených
	zmizela . Patrně je mění s ostatními klany za maso plížilů /plížilů/x@-----	, které jim imperátor ponechává . " " Vítejte ,
	Vyměníme s nimi generátor na výrobu výživné pasty za stádečko plížilů /plížilů/x@-----	. " " Na to by nepřistoupili , " namítl
	se nerozteču . " Čtenář dopisů se zabýval jedním z plížilů /plížilů/x@-----	, kteří mu zbyli a který byl zjevně nemocný .
	jako když se blíží divoký mršt , takže jeho stádo plížilů /plížilů/x@-----	nic neohrožovalo . Rozhodně to však bylo něco nového a
	na maso různých chutí , měla i několik cvičených užitkových plížilů /plížilů/x@-----	vyšlechtěných pro vyšší inteligenci . Teď , když přinesla Puřka
	ty špatné zprávy . " " Odmítli převzít nová vejce plížilů /plížilů/x@-----	. " " Takže jsi rozsekala vůdce klanu a několik
	loď snášela na povrch Vejce , nešťastnou náhodou zahynulo několik plížilů /plížilů/x@-----	Zebu , které střežil klan . Ochotně imperátorovi za jejich
	po staré cestě k západnímu pólu , minul velké stádo plížilů /plížilů/x@-----	. Byla s nimi Skvrnitý vršek se studenty , kteří
	, " řekl , když si prohlédl vlnící se stádo plížilů /plížilů/x@-----	. " Dal bych si jednoho z těch pruhovaných .
pu .	Vedoucímu elektromagnetiků/1 připomínal jednu komickou ráci domácích plížilů /plížilů/x@-----	, kteří měli srst tak dlouhou , že potřebovali robota
	Sotva otevřel hlavní dveře , obklopila ho smečka šťastně frkajících plížilů /plížilů/x@-----	. Byla mezi nimi dokonce i Lassie , která se
	skály . " Vezmu si 132 kusů od každého druhu plížilů /plížilů/x@-----	pro imperátora , " řekla Výběřcí . " Zbytek vám
	rozvlnily , když si všimla klanových obydlí a ohrad plných plížilů /plížilů/x@-----	. " Načasování je skoro dokonalé , " poznamenal Kůroplaz
	! " Zlomený štít shromáždila svůj klan , posbírali vejce plížilů /plížilů/x@-----	a zamířili na východ k vesnici klanu . Čtenář dopisů
	ho mnohem víc , než by ho bylo z veletuctu plížilů /plížilů/x@-----	. " Výběřcí se zarazil a vlnění jejích očních stopek

Závěry

- redukce počtu lemmat (z 3023 na 168) pro ruční analýzu (o 94%)
- efektivní vyhledání maximálního počtu jednotek (srv. Štícha 2011)
- minimalizace problému přegenerování i podgenerování
- 44 dokladů – zachycených AMA, +9 (20%) bez AMA
- redukce rozsahu ruční práce (nebezpečí chyb z nepozornosti)

ú 26. 10.

- Pokuste se využít nástroj morfio k hledání dalších dokladů v korpusu SYN2015

Děkuji Vám za pozornost !