A.

V češtině je prefixace jedním z nejrozšířenějším způsobů tvoření nových sloves. Podívám-li se na slovesné prefixy, zjistíme, že v naprosté většině případů jsou „krátké“, zatímco u jmen (substantiv i adjektiv) tomu tak není.

Například:

na-brat ALE ná-bor

za-brat ALE zá-bor

při-jmout ALE pří-jem

u-jmout ALE ú-jma

vy-brat ALE vý-bor

po-hnat ALE pů-hon

do-vést ALE dů-vod

pro-vést ALE prů-vod

Přesto existují i případy, kdy se i u sloves můžeme setkat s délkou v prefixu.

1. Popiš postup, jak vyhledat v korpusu (SYN2010) data pro výzkum uvedeného jevu (tj. lemmata sloves, která začínají na řetězce shodné s dlouhými prefixy).

2. Pozorujte data a pokuste se je interpretovat.

[lemma="((ná)|(zá)|(dů)|(pů)|(prů)|(pří)|(vý)|(ú)).\*" & tag="V.\*"]

67 lemmat

Časté jsou deriváty od jmen, která mají dlouhý prefix.

Objevují se případy, kde nejde o prefix (*zářit, úpět*)*.*

O prefix slovesného tvaru jde pouze v těchto případech: *záležet, záviset, náležet, závidět*.

V ostatních případech je motivujícím slovem jméno (*důvěra→důvěřovat, závod→závodit*), popř. etymologicky se jménem souvisí (*zápasit ← držet se za pas, působit ← po sobě* viz *násobit ← dávat na sebe*).

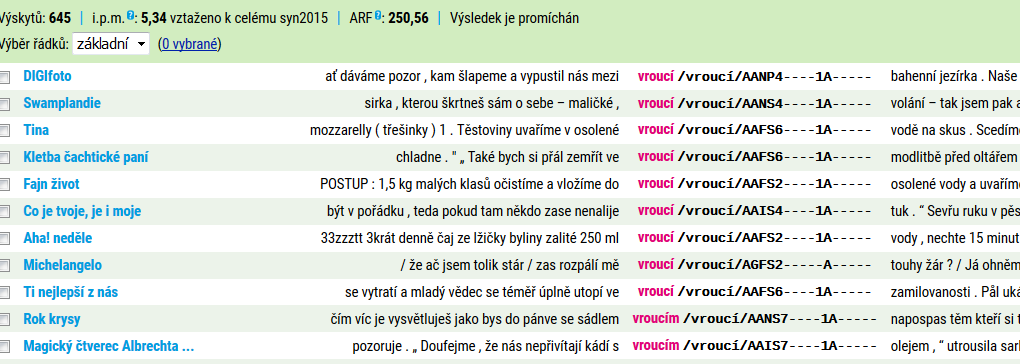
Výjimka je též *půjčit* a alternace v syntetickém futur slovesa *jít* (*půjde*).

B.

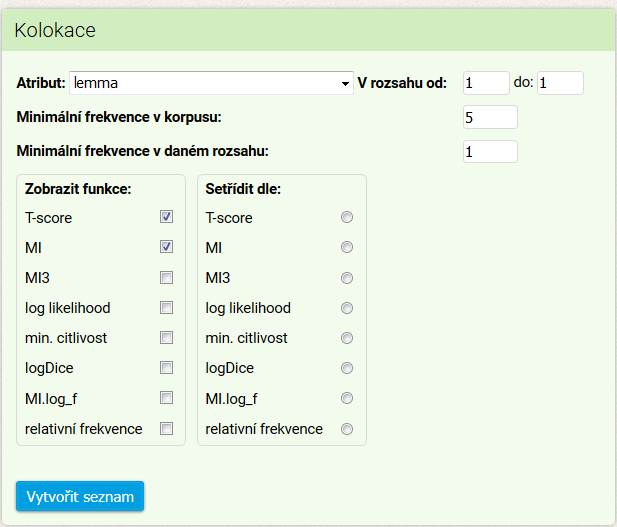
Vyhledejte v korpusu všechny výskyty adjektiva ***vroucí****.* Pozoruj je a popiš významy, které toto adjektivum může mít (alespoň dva).

Navrhni postup, jak vybrat pouze ty konkordanční řádky, na nichž má adjektivum jeden nebo druhý význam.

Vyhledáme v korpusu lemma *vroucí.*

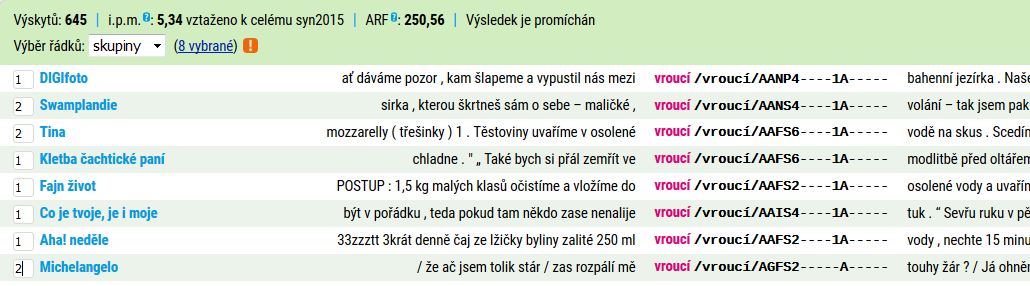


Dále se zaměříme na kolokace vpravo (adjektivum rozvíjí substantivum ve funkci shodného přívlastku, který v češtině typicky stojí před substantivem, které rozvíjí): *voda, oselený, polévka, vývar, kotel, …* ALE *přání, láska, hlas, přátelství, vztah, slovo, polibek*.





Pomocí Výběru řádků označíme skupinu 1 a 2.



C.

V korpusu SYN nejsou slovní tvary *mudla/mudlové* (viz věty níže) rozpoznány automatickou morfologickou analýzou.

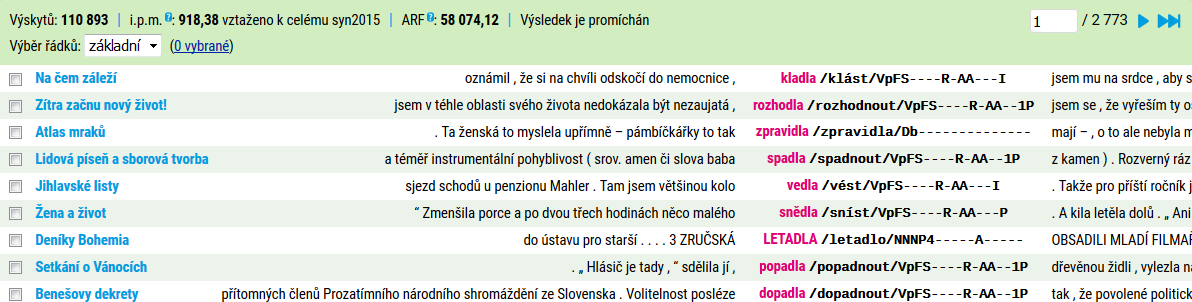
Popište, jak byste určili lemma a tag, aniž byste museli vědět, co slovní tvar označuje.

A *mudla* zůstává - stejně jako v Harrym Potterovi - nezasvěcen tajnému vědění

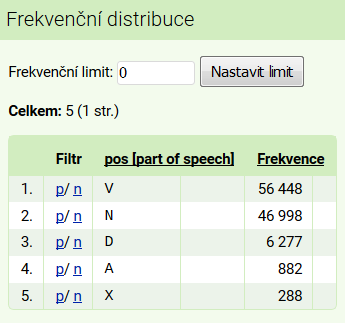
A teď mají tu možnost i malí *mudlové*.

Předpokládáme-li, že lemmatizujeme tvar bez kontextu, pak můžeme postupovat následovně:

[lc="(.\*dla)|(.\*dlové)"]



Frekvence slovních druhů:



1. sloveso: mudla/mudnout/(Vp-SF)|(Vp-PN) jako rozhodla/rozhodnout
2. substantivum: mudla/mudlo/NNN(S2|P[145]) jako divadla/divadlo, mudla/mudl/NNMS[24] jako pudla/pudl, mudla/mudla/NN[MF]S1 jako modla/modla nebo Kudla/Kudla, mudlové/mudla/NNMP[145] jako Kudlové/Kudla,

mudlové/mudl/NNMP[145] jako pudlové/pudl,

1. adjektivum mudlové/mudlový/AAFS[236] …. jako nudlové/nudlový.

S desambiguací – pravidlovou:

mudla/mudla|mudl/NNMS1

mudlové/mudla|mudl/NNMP[145]

D.

Jaké lemma a pos (slovní druh) mohou mít tučně vyznačená slova.

*To má být* ***výchovné****?*

***Místo*** *vašich schůzek sis měl lépe vybírat.*

***Jí je*** *špatně.*

***Vezme*** *ho domů.*

***výchovné/výchovné/NNIS1|výchovné/výchovný/AANS1***

***místo/místo/NNNS4|místo/místo/R---2***

***jí/ona/P.FS3***

***jí/jíst/VB-S---3***

***vezme/vzít/ VB-S---3***

***vezme/vézt/Vi-P---1***