

Vyhledávání v Českém národním korpusu – termíny:

Viz též

http://wiki.korpus.cz/doku.php/manual:zobrazeni_dotazu

Každému slovu jsou přiděleny tzv. **poziční atributy** = vlastnosti, podle kterých je můžeme hledat.

Základními pozičními atributy jsou:

- **word** - slovní tvar (např. *kočkou, modrou, píše*)
- **lemma** - základní (slovníkový) tvar (např. *kočka, modrý, psát*)
- **tag** - značka zachycující morfologickou (příp. i jinou, obvykle gramatickou) informaci

Příklad:

kočkou – word/lc: *kočkou*

lemma: *kočka* (word/lc *kočkou* patří k lemmatu *kočka*)

tag: podstatné jméno, rod ženský, jednotné číslo, 7. pád (NNFS7.*)

(N = podstatné jméno, N – obyčejné, tj. bez další zvláštní charakteristiky, F = femininum, S = singulár, 7 = 7. pád; značka „*“ znamená, že další gramatické vlastnosti nepotřebujeme)

modrou – word/lc: *modrou*

lemma: *modrý* (word/lc *modrou* patří k lemmatu *modrý* – nikoliv *modrá*!!!)

tag: přídavné jméno, rod ženský, jednotné číslo, 7. pád (AAFS7....1A.*) nebo 4. pád (AAFS4....1A.*)

(A = přídavné jméno, A – obyčejné, tj. bez další zvláštní charakteristiky, F = femininum, S = singulár, 7 = 7. pád // 4 = 4. pád; pak 4 nedefinované pozice, 1 = 1. stupeň, A = bez negace, značka „*“ znamená, že další gramatické vlastnosti nepotřebujeme)

píše – word/lc: *píše*

lemma: *psát* (word/lc *píše* patří k lemmatu *psát*)

tag: – sloveso, 3. osoba jednotného čísla přítomného času, činný rod, kladný tvar (VB.S...3P.AA.*)

(V = verbum/sloveso, B – tvar přítomného nebo budoucího času, „“ = pozice není definována, S = singulár, pak 3x nedefinovaná pozice, 3 = 3. osoba, P = prézens, pak nedefinovaná pozice, A = aktivum, A = kladný tvar)

Toto značkování umožní vyhledat právě ten výraz, který potřebujeme, nebo právě ty tvary slov (např. dativ podstatných jmen rodu ženského, všechny přechodníky přítomné apod.

Pokud chceme vyhledávat s využitím atributu tag, volíme vždy typ dotazu CQL.

Postup při „vkládání tagu“ – viz soubor „tag_vyhledavani“.