

Masarykova univerzita  
Filozofická fakulta

# Dotazy v CQL (pro Sketch Engine)

*PLINo21: Sémantická analýza v praxi*

# Obsah

Úvod .....	3
I. Co je to CQL .....	4
II. Pojmy .....	4
III. Vyhledávání v CQL.....	5
III.I. Regulární výrazy .....	6
III.II. Morfologické značky pro češtinu.....	11
IV. Zdroje .....	18

# Úvod

Tématem naší práce je vyhledávání v CQL se zaměřením na české korpusy. Můžete se dočíst, co vlastně CQL znamená nebo kde a jak ho používat. Naleznete zde i krátkou exkurzi do tohoto vyhledávání, která by pro vás mohla být přínosná a užitečná, pokud se s touto problematikou teprve seznamujete.

Naším cílem je tedy přiblížit vám význam CQL a jeho využití.

## I. Co je to CQL

CQL je zkratkou pro Corpus Query Language, což je formální jazyk umožňující složité dotazy. Tento jazyk byl vyvinut v roce 1990 na univerzitě ve Stuttgartu a jeho rozšířená verze z května 2013 je použita ve Sketch Engine.

## II. Pojmy

**token** = pozice v korpusu (slovo i interpunkce)

**atribut** = rozlišovací rys

**řetězec** = posloupnost znaků

**regulární výraz** = řetězec popisující celou množinu řetězců

**operátor** = metaznak, který má svoji specifickou funkci v regulárním výrazu

**kvantifikátor** = speciální operátor, který definuje určitý způsob opakování

### III. Vyhledávání v CQL

Základní dotazy v CQL se skládají z atributu a hodnoty ve tvaru [atribut = "hodnota"].

V korpusech lze zvolit implicitní atribut i samostatně:

IMPLICITNÍ ATRIBUT	VÝZNAM IMPLICITNÍHO ATRIBUTU
<b>word</b>	<i>hledáme zadaný výraz, kterým je slovo v jakémkoli tvaru</i>
<b>lemma</b>	<i>hledáme základní slovní tvar</i>
<b>tag</b>	<i>hledáme výraz pomocí morfologických značek a regulárních výrazů</i>
<b>lc</b>	<i>hledáme výraz bez ohledu na velikost písmen</i>

V korpusu vyhledáváme například výskyty slovního tvaru *auto*, kde *word* je atributem a *auto* hodnotou tohoto atributu:

**[word = "auto"]** - vyhledá pouze výskyty slovního tvaru *auto*.

Pokud ovšem zvolíme implicitní atribut *word*, stačí pak zadat jen hledaný výraz do uvozovek:

**"auto"** - vyhledá totéž, co předchozí příklad.

### III. I. Regulární výrazy

Při hledání v CQL nám často usnadňuje práci zapsání dotazu jako *regulární výraz*. Pro správné používání regulárních výrazů je nejprve třeba seznámit se s jejich *speciálními znaky* (tzv. metaznaky).

#### Přehled speciálních znaků:

<b>OPERÁTORY</b>	
.	<i>libovolný znak</i>
( )	<i>skupina znaků (např. při opakování)</i>
<b>KVANTIFIKÁTORY</b>	
*	<i>opakovat nejméně okráť, nejvíce neomezeno</i>
+	<i>opakovat nejméně jednou, nejvíce neomezeno</i>
?	<i>opakovat okráť nebo 1krát</i>
{A}	<i>opakovat právě Akrát</i>
{A,B}	<i>opakovat nejméně Akrát, nejvíce Bkrát</i>
<b>DALŠÍ OPERÁTORY</b>	
(?)	<i>nezáleží na velikosti znaku (např. a/A) alternativa k atributu lc</i>
 a b	<i>nebo vybírání buď a nebo b</i>
[ ] [abc]	<i>nebo (zapsáno jinak) vybírání buď a, b nebo c</i>
- [a-d]	<i>od do a, b, c nebo d</i>
^ [^ab]	<i>negace (na klávesnici: AltGr + š, Alt + 94) vybírání cokoliv kromě a, b</i>
<s>	<i>začátek struktury/věty</i>
</s>	<i>konec struktury/věty</i>
<s/>	<i>celá struktura/věta</i>
&	<i>konjunkce (a zároveň)</i>
!	<i>negace</i>
meet	<i>speciální typ kontextu</i>
union	<i>spojí výsledky vyhledané pomocí meet</i>
containing	<i>struktura obsahující zadaný dotaz</i>
within	<i>vyhledávání uvnitř zvolené struktury</i>
1:[ ] 2:[ ] & 1.c=2.c	<i>shoda podle morfologické značky (c ... pád)</i>

Prvním takovým základním metaznakem je *tečka* “.”, která nahrazuje právě jeden *libovolný znak* (písmena, číslice i různé další znaky):

“**l.p**” - vyhledá např. slova *lup, líp, lep*.

Je důležité si předem uvědomit, že vyhledávání v korpusu neoperuje s konkrétními písmeny a čísly, ale obecně se znaky zabírajícími právě jednu pozici. Proto písmeno “ch” (charakteristické pro češtinu) zastupuje znaky dva, stejně tak například číslo 12345 je znaků pět.

Tečku je možné použít i několikrát za sebou. Např. pokud zadáme mezi písmena *p* a *t* pět teček:

“**p.....t**” - vyhledá např. *pořídít, proudit, projekt, produkt*.

Abychom tečky nemuseli počítat, je možné použít tzv. *kvantifikátory* a určit tak jejich počet. Pokud chceme zadat konkrétní množství, použijeme za tečkou (obecně za znakem či skupinou znaků, které chceme násobit) číslo ve složených závorkách:

“**p.{5}t**” - vyhledá totéž co předešlý příklad “**p.....t**”.

Je také možné zadat opakování pomocí intervalu. Například pokud chceme, aby bylo slovo 5 až 7 písmen dlouhé, tedy aby začínalo na *p*, končilo na *t* a uprostřed se vyskytovalo 3 až 5 libovolných znaků, dotaz zapíšeme takto:

“**p.{3,5}t**” - vyhledá např. *pocít, použít, projekt*.

Pro skupinu znaků:

“**(ko){1,2}s**” - vyhledá slova *kos* nebo *kokos*.

V případě, že nechceme určit přesný počet opakování znaku či skupiny znaků, použijeme *speciální kvantifikátory*. Pro názorný příklad si jako základ zvolme sloveso *jít*.

*Kvantifikátor \**, tzv. *Kleeneho hvězdičku*, použijeme ve chvíli, kdy nám nezáleží na množství, stejně tak na tom, jestli se (libovolný) znak vůbec objeví.

“**.\*jít**” - vyhledá např. *jít, ujít, vyjít, sejít, přijít, odejít, vzejít, nenajít, předejít*.

Pokud potřebujeme, aby se znak objevil alespoň jednou, použijeme speciální *kvantifikátor +*:

“**+.jít**” - vyhledá např. *ujít, vyjít, sejít, přijít, odejít, vzejít, nenajít, předejít*.

Do třetice máme k dispozici speciální *kvantifikátor* ?, kterým jakýkoli znak nebo skupinu znaků vyhledáváme buď právě jednou, nebo vůbec:

“**.?jít**” - vyhledá např. *jít, ujít*.

Při vyhledávání v korpusu často usnadní práci využití různých způsobů dotazů, vyjadřujících disjunkci. Tedy existuje několik způsobů, jak zapsat v rámci jednoho dotazu více variant. Napomůžou nám k tomu speciální znaky | nebo [ ].

Pro hledání dvou či více zcela odlišných slov, použijeme svislé znaménko:

“**muž|žena**” - vyhledá výskyty slov *muž, žena*,

“**já|ty|on**” - vyhledá výskyty slov *já, ty, on*.

Hledáme-li varianty uvnitř slova, pak je třeba použít závorky:

- v případě jednotlivých hlásek závorky hranaté:

“**muž[ie]**” - vyhledá výskyty slovních tvarů *muži* a *muže*,

“**dám[ayu]**” - vypíše výskyty slov: *dáma, dámy* a *dámu*,

- v případě skupin hlásek závorky kulaté:

“**muž(ovi|ům)**” - vyhledá výskyty slovních tvarů *mužovi* a *mužům*,

“**muž(i|ovi|ům|em)**” - vyhledá výskyty slovních tvarů *muži, mužovi, mužům, mužem*.

Dotazování na více variant je možné zadávat i obecněji, a to v případě, chceme-li vybrat jedno písmeno z určité části abecedy anebo použitím negace vypsát pouze znaky, které jako jediné nevyhledáváme.

*Operátorem* - vymezíme rozsah písmen v abecedě (nebo rozsah číslic). Použijeme jej uvnitř hranatých závorek:

“**[b-d]o**” - vyhledá např. *bo, co, do* (ale už ne např. *ho, jo, po*).

Pro negaci použijeme znak ^ na začátku řetězce uvnitř hranatých závorek. Názorným příkladem nám mohou být slova nezačínající na tvrdý konsonant (souhlásku), za kterými se vyskytuje “y”:

“**[^hkrdtn].\***” - vyhledá např. *byt, systém, myslím, využití*.

Můžeme také vyhledávat slovní spojení, mezi nimiž je určitý počet dalších tokenů:



**“auto.\*” [] {2} “stojí”** - vyhledá slova začínající na *auto* a slova *stojí*, mezi kterými jsou přesně dva tokeny, např.: ...*jedním z autorů, které stojí za to...*

Nebo můžeme zadat dotaz tak, aby se mezi dvěma zadanými výrazy nacházely nula až dva další tokeny:

**“auto.\*” [] {0, 2} “stojí”** - vyhledá to samé co předchozí příklad + výsledky s žádným nebo jedním dalším tokenem, např.: ...*osobní automobily stojí ve...; ...rychlostí automatického ošetření stojí algoritmus...*

Chceme-li hledat slova v určitém kontextu, použijeme takovéto zadání dotazu:

**“(meet [tag="k1.\*"] [tag="k5.\*"] -3 3)”** - vyhledá podstatná jména, která jsou v kontextu +-3 tokeny od slovesa (atribut *tag*).

Pomocí *union* můžeme takto hledané výsledky spojovat:

**(union (meet [tag="k1.\*"] [tag="k5.\*"] -3 3) (meet [tag="k2.\*"] [tag="k5.\*"] -2 2))**  
- vyhledá všechna podstatná jména v kontextu +-3 tokeny od slovesa a zároveň všechna přídavná jména v kontextu +-2 tokeny od slovesa.

Pokud chceme vyhledat slovní spojení v rámci jedné struktury, můžeme použít například zadání dotazu v podobě:

**"auto.\*" []\* "stojí" within <s/>** - vyhledá řetězec slov začínající slovem na *auto* a končící slovem *stojí* v rámci jedné věty, mezi nimiž je libovolný počet dalších tokenů (atribut *word*).

Hledáme-li strukturu/větu, obsahující námi zadaná slova (tokeny), použijeme operátor *containing*:

**<s/> containing "auto.\*" []\* "stojí"** - vyhledá celou strukturu/větu obsahující slovo začínající na *auto*, po kterém mohou (ale nemusí) následovat tokeny libovolného množství, a slovo *stojí*.

V korpusu je možné vyhledávat tokeny, které se určitým jevem shodují:

**1:[word = "pane"] 2:[] & 1.c = 2.c** - vyhledá slovní spojení shodující se v pádě, kde prvním výrazem je slovo *pane*, např. *pane ministře, pane kolego...*  
Shodu zajišťuje *c* v části dotazu *1.c = 2.c*,  
(vizte následující kapitola *III.II. Morfologické značky pro češtinu*),

**1:[word = "d.\*"] 2:[] & 1.c = 2.c** - vyhledá slovní spojení shodujícím se v pádě, kde prvním výrazem je slovo začínající na písmeno *d*, např. *dosahuje jen, druhé přičce, duchovní hudbu.*

Negaci pomocí *vykřičníku "!"* využijeme v případě, chceme-li negovat jiný operátor. Např. *within!, containing!* nebo *!=*. V tomto případě se význam a funkce těchto operátorů neguje.

### III. II. Morfologické značky pro češtinu

Morfologické značky vyjadřují gramatické kategorie slovních tvarů.

Při zadávání dotazů pomocí morfologických značek volíme atribut *tag* a dotaz pak zapíšeme do uvozovek, např. “**k1gMnPc4**”. Nebo do dotazového řádku napíšeme dotaz v celé podobě, a to [**tag = “k1gMnPc4”**].

#### Poziční systém morfologických značek pro češtinu podle slovního druhu:

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k1</b>	<i>substantivum</i>
2. pozice	Speciální vzor	<b>xP</b> <b>xF</b>	<i>půl</i> <i>rodina (příjmení)</i>
3. pozice	Rod	<b>gM</b> <b>gI</b> <b>gN</b> <b>gF</b>	<i>mužský životný</i> <i>mužský neživotný</i> <i>střední</i> <i>ženský</i>
4. pozice	Číslo	<b>nS</b> <b>nP</b> <b>nD</b>	<i>jednotné</i> <i>množné</i> <i>duál</i>
5. pozice	Pád	<b>c1</b> <b>c2</b> <b>c3</b> <b>c4</b> <b>c5</b> <b>c6</b> <b>c7</b>	<i>první</i> <i>druhý</i> <i>třetí</i> <i>čtvrtý</i> <i>pátý</i> <i>šestý</i> <i>sedmý</i>
6. pozice	Stylistický příznak tvaru	<b>wB</b> <b>wH</b> <b>wN</b> <b>wR</b> <b>wZ</b>	<i>básnický</i> <i>hovorově</i> <i>nářečně</i> <i>řidčeji</i> <i>zastarale</i>
7. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	<i>tvar s příklonným “-s”</i> <i>kondicionál (“aby”, “kdyby”, “by”)</i> <i>zkratka</i>

Příklad morfologické značky pro substantivum “psy”:

“**k1gMnPc4**” - substantivum rodu mužského životného v množném čísle 4. pádu,

“**k1gMnPc7**” - substantivum rodu mužského životného v množném čísle 7. pádu.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k2</b>	<b>adjektivum</b>
2. pozice	Negace	<b>eA</b> <b>eN</b>	<i>afirmace</i> <i>negace</i>
3. pozice	Rod	<b>gM</b> <b>gI</b> <b>gN</b> <b>gF</b> <b>gR</b>	<i>mužský životný</i> <i>mužský neživotný</i> <i>střední</i> <i>ženský</i> <i>rodina (příjmení)</i>
4. pozice	Číslo	<b>nS</b> <b>nP</b>	<i>jednotné</i> <i>množné</i>
5. pozice	Pád	<b>c1</b> <b>c2</b> <b>c3</b> <b>c4</b> <b>c5</b> <b>c6</b> <b>c7</b>	<i>první</i> <i>druhý</i> <i>třetí</i> <i>čtvrtý</i> <i>pátý</i> <i>šestý</i> <i>sedmý</i>
6. pozice	Stupeň	<b>d1</b> <b>d2</b> <b>d3</b>	<i>pozitiv</i> <i>komparativ</i> <i>superlativ</i>
7. pozice	Stylistický příznak tvaru	<b>wB</b> <b>wH</b> <b>wN</b> <b>wR</b> <b>wZ</b>	<i>básnický</i> <i>hovorově</i> <i>nářečně</i> <i>řidčeji</i> <i>zastarale</i>
8. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	<i>tvar s příklonným "-s"</i> <i>kondicionál ("aby", "kdyby", "by")</i> <i>zkratka</i>

Příklad morfologické značky pro adjektivum "dobří":

"**k2eAgMnPc1d1**" - adjektivum afirmační rodu mužského životného čísla množného 1. pádu pozitiva,

"**k2eAgMnPc5d1**" - adjektivum afirmační rodu mužského životného čísla množného 5. pádu pozitiva.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k3</b>	<b>zájmeno</b>
2. pozice	Druh zájmena	<b>xP</b> <b>xO</b> <b>xD</b> <b>xT</b>	<i>osobní</i> <i>přivlastňovací</i> <i>ukazovací</i> <i>vymezovací</i>

3. pozice	Druh zájmena	yF yQ yR yN yI	reflexivní tázací vztažné záporné neurčité
4. pozice	Osoba	p1 p2 p3	první druhá třetí
5. pozice	Rod	gM gI gN gF gR	mužský životný mužský neživotný střední ženský rodina (příjmení)
6. pozice	Číslo	nS nP	jednotné množné
7. pozice	Pád	c1 c2 c3 c4 c5 c6 c7	první druhý třetí čtvrtý pátý šestý sedmý
8. pozice	Stylistický příznak tvaru	wB wH wN wR wZ	básnický hovorově nářečně řidčeji zastarale
9. pozice	Typ tvaru	zS zY zA	tvar s příklonným "-s" kondicionál ("aby", "kdyby", "by") zkratka

Příklad morfologické značky pro zájmeno "toto":

**"k3gNnSc1"** - zájmeno středního rodu čísla jednotného 1. pádu,

**"k3gNnSc4"** - zájmeno středního rodu čísla jednotného 4. pádu.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	k4	<b>číslovka</b>
2. pozice	Druh číslovky	xC xO xR	základní řadová druhá
3. pozice	Druh číslovky	yN yI	záporná neurčitá
4. pozice	Rod	gM gI gN gF	mužský životný mužský neživotný střední ženský

5. pozice	Číslo	<b>nS</b> <b>nP</b>	<i>jednotné</i> <i>množné</i>
6. pozice	Pád	<b>c1</b> <b>c2</b> <b>c3</b> <b>c4</b> <b>c5</b> <b>c6</b> <b>c7</b>	<i>první</i> <i>druhý</i> <i>třetí</i> <i>čtvrtý</i> <i>pátý</i> <i>šestý</i> <i>sedmý</i>
7. pozice	Stylistický příznak tvaru	<b>wB</b> <b>wH</b> <b>wN</b> <b>wR</b> <b>wZ</b>	<i>básnický</i> <i>hovorově</i> <i>nářečně</i> <i>řidčeji</i> <i>zastarale</i>
8. pozice	Terminál v gramatice	<b>tA</b> <b>tB</b> <b>tC</b> <b>tD</b> <b>tE</b> <b>tF</b> <b>tI</b> <b>tJ</b> <b>tK</b> <b>tL</b> <b>tM</b> <b>tN</b> <b>tO</b> <b>tP</b> <b>tQ</b> <b>tR</b> <b>tS</b> <b>tT</b> <b>tU</b> <b>tV</b> <b>tW</b> <b>tX</b> <b>tY</b> <b>tZ</b>	<i>terminál A (vyjadřuje zřetel)</i> <i>terminál B</i> <i>terminál C (vyjadřuje příčinu)</i> <i>terminál D (příslouce modální)</i> <i>terminál E</i> <i>terminál F</i> <i>terminál I</i> <i>terminál J</i> <i>terminál K</i> <i>terminál L (vyjadřuje místo)</i> <i>terminál M (vyjadřuje způsob)</i> <i>terminál N</i> <i>terminál O</i> <i>terminál P</i> <i>terminál Q (vyjadřuje míru)</i> <i>terminál R</i> <i>odpovídá Q@ (příslouce stavové)</i> <i>odpovídá O@ (vyjadřuje čas)</i> <i>odpovídá L@</i> <i>odpovídá jedno</i> <i>odpovídá sto</i> <i>odpovídá dvě</i> <i>odpovídá stě</i> <i>odpovídá tři/čtyři</i>
9. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	<i>tvar s příklonným "-s"</i> <i>kondicionál ("aby", "kdyby", "by")</i> <i>zkratka</i>

Příklad morfologické značky pro číslovku "několik":

"**k4c1**" - číslovka v 1. pádě,

"**k4c4**" - číslovka ve 4. pádě.

Příklad morfologické značky pro číslovku “stí”:

“**k4gMnPc1**” - číslovka rodu mužského životného množného čísla v 1. pádě,

“**k4gMnPc5**” - číslovka rodu mužského životného množného čísla v 5. pádě.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k5</b>	<i>sloveso</i>
2. pozice	Negace	<b>eA</b> <b>eN</b>	<i>afirmace</i> <i>negace</i>
3. pozice	Vid	<b>aP</b> <b>aI</b>	<i>perfektivum (= dokonavé)</i> <i>imperfektivum (= nedokonavé)</i>
4. pozice	Typ (Mód)	<b>mF</b> <b>mI</b> <b>mR</b> <b>mA</b> <b>mN</b> <b>mS</b> <b>mD</b> <b>mB</b>	<i>infinitiv</i> <i>indikativ přítomný</i> <i>imperativ</i> <i>příčestí činné (minulé)</i> <i>příčestí trpné</i> <i>přechodník přítomný (současnost)</i> <i>přechodník minulý (dřívější děj)</i> <i>indikativ futura</i>
5. pozice	Osoba	<b>p1</b> <b>p2</b> <b>p3</b>	<i>první</i> <i>druhá</i> <i>třetí</i>
6. pozice	Rod	<b>gM</b> <b>gI</b> <b>gN</b> <b>gF</b>	<i>mužský životný</i> <i>mužský neživotný</i> <i>střední</i> <i>ženský</i>
7. pozice	Číslo	<b>nS</b> <b>nP</b>	<i>jednotné</i> <i>množné</i>
8. pozice	Stylistický příznak tvaru	<b>wB</b> <b>wH</b> <b>wN</b> <b>wR</b> <b>wZ</b>	<i>básnický</i> <i>hovorově</i> <i>nářečně</i> <i>řidčeji</i> <i>zastarale</i>
9. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	<i>tvar s příklonným “-s”</i> <i>kondicionál (“aby”, “kdyby”, “by”)</i> <i>zkratka</i>

Příklad morfologické značky pro sloveso “rob”:

**k5eAaImRp2nS** - sloveso afirmační nedokonavé v imperativu 2. osoby čísla jednotného.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k6</b>	<i>příslovce</i>
2. pozice	Negace	<b>eA</b> <b>eN</b>	<i>afirmace</i> <i>negace</i>
3. pozice	Druh zájmeného příslovce	<b>xD</b> <b>xT</b>	<i>ukazovací</i> <i>vymezovací</i>
4. pozice	Druh zájmeného příslovce	<b>yQ</b> <b>yR</b> <b>yN</b> <b>yI</b>	<i>tázací</i> <i>vztažné</i> <i>záporné</i> <i>neurčité</i>
5. pozice	Stupeň	<b>d1</b> <b>d2</b> <b>d3</b>	<i>pozitiv</i> <i>komparativ</i> <i>superlativ</i>
6. pozice	Stylistický příznak tvaru	<b>wB</b> <b>wH</b> <b>wN</b> <b>wR</b> <b>wZ</b>	<i>básnický</i> <i>hovorově</i> <i>nářečně</i> <i>řidčeji</i> <i>zastarale</i>
7. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	<i>tvar s příklonným "-s"</i> <i>kondicionál ("aby", "kdyby", "by")</i> <i>zkratka</i>

Příklad morfologické značky pro příslovce "nemálo":

**k6eNd1** - příslovce negované pozitivu.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k7</b>	<i>předložka</i>
2. pozice	Pád	<b>c1</b> <b>c2</b> <b>c3</b> <b>c4</b> <b>c5</b> <b>c6</b> <b>c7</b>	<i>první</i> <i>druhý</i> <i>třetí</i> <i>čtvrtý</i> <i>pátý</i> <i>šestý</i> <i>sedmý</i>

Příklad morfologické značky pro předložku "kolem":

**k7c2** - předložka v 2. pádě.

Příklad morfologické značky pro předložku "při":

**k7c6** - předložka v 6. pádě.



POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k8</b>	<b>spojka</b>
2. pozice	Druh spojky	<b>xC</b> <b>xS</b>	souřadící podřadící
3. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	tvar s příklonným "-s" kondicionál ("aby", "kdyby", "by") zkratka

Příklad morfologické značky pro spojku "ani":

**k8xC** - spojka souřadící.

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>k9</b>	<b>částice</b>
2. pozice	Typ tvaru	<b>zS</b> <b>zY</b> <b>zA</b>	tvar s příklonným "-s" kondicionál ("aby", "kdyby", "by") zkratka

POZICE	KATEGORIE	ZNAČKA	VYSVĚTLIVKA
1. pozice	Slovní druh	<b>ko</b>	<b>citoslovce</b>

POZICE	ZNAČKA	VYSVĚTLIVKA
1. pozice	<b>kI</b>	<b>interpunkce</b>
2. pozice	<b>x.</b> <b>x,</b> <b>x"</b> <b>x(</b> <b>x)</b> <b>x~</b>	.!? ,:; " ' " " " " ( { [ < ) } ] > ~ \$ % ^ & - _ + = \   / # atd.

## IV. Zdroje

LEXICAL COMPUTING, Ltd. *Sketch Engine* [online]. 2003–2013 [cit. 2014-01-12]. Dostupné z: <http://www.sketchengine.co.uk/documentation/wiki/WikiStart>

### Regulární výrazy:

Corpus Querying: Corpus Query Language (CQL). LEXICAL COMPUTING, Ltd. *Sketch Engine* [online]. 2003–2013 [cit. 2014-01-11]. Dostupné z: <https://www.sketchengine.co.uk/documentation/wiki/SkE/CorpusQuerying>

Základy regulárních výrazů. PECKA, Miroslav. *Regulární výrazy* [online]. 2005–2013 [cit. 2014-01-11]. Dostupné z: <http://www.regularnivyrazy.info/regularni-vyrazy-zaklady.html>

### Morfologické značky:

JAKUBÍČEK, Miloš, Vojtěch KOVÁŘ a Pavel ŠMERK. Czech Morphological Tagset Revisited. In Horák, Rychlý. *Proceedings of Recent Advances in Slavonic Natural Language Processing 2011*. Brno: Tribun EU, 2011. s. 29-42, 14 s. ISBN 978-80-263-0077-9.

Morfologický analyzátor ajka – interaktivní režim [online]. [cit. 2014-1-11]. Dostupné z: <http://nlp.fi.muni.cz/projekty/wwwajka/WwwAjkaSkripty/morph.cgi?jazyk=0>