

## Opakování

Zopakujte si typy dotazů k vyhledávání jazykových jevů.

- Lemmata pro vyhledávání pravopisných variant: nejvhodnější je např. dotaz typu lemma – **bal.n** (tj. balon/balón), kdy místo variantního písmenka volíme tečku. A pozor, pak musíme zadat frekvenci slovních tvarů, frekvence lemmat často nepomůže.
- Dotazy typu „CQL“: pro vyhledávání gramatických tvarů jsou výhodnější vyhledávání konkrétního tvaru, protože v češtině je výrazná tvarová homonymie. Proto např. tvar genitivu singuláru substantiva kámen hledáme dotazem: [(tag="N..S2.\*")(lemma="kámen")], protože tvary *kamene*, *kameni* i *kamenu* mohou být i v jiných pádech.

A pozor, vždy je vhodné tvar, o němž víme, že je homonymní s tvarem jiného pádu, zkontrolovat.

- Spojení slov: nejvýhodnější jsou dotazy CQL, kdy vedle sebe spojí např. lemma a tag ze slov, která slovní spojení tvoří: [tag="A.\*"] [lemma="vrána"] (černá vrána, bílá vrána, malá vrána...). Pak totiž zjistíme frekvenci celých spojení jednoduše jako „frekvenci lemmat“.

Nové úkoly

1. a) Zjistěte, která písmena můžete doplnit do slov *sl-nout*, *vl-s*, *l-žiny*, *pl-tký*

b) Zjistěte, jaký význam mají nalezená slova podle užití v textech.

Pokud nedokážete význam vysledovat z textu, využijte Slovník spisovného jazyka českého (<https://ssjc.ujc.cas.cz/>)

2. Zjistěte, zda se vyskytují z následujících dvojic obě slova, nebo jen jedno. Vysvětlete též jejich význam podle užití v textech, nebo využijte využití Slovník spisovného jazyka českého (<https://ssjc.ujc.cas.cz/>).

sýřit – sítit, sýřený – šířený, sýření – šíření, sýřenina – šířenina

3. Zjistěte, v jakých podobách je doloženo vlastní jméno obce *KOSTELEC NAD ČERNÝMI LESY*.

Doporučení: použijte dotaz typu CQL, jako kdybyste hledali např. název Jaroměřice nad Rokytnou.

Postup\_

1) [lemma=".aroměřice"] [lc="nad"] [lc="rokytnou"]

Atribut „lc“ nahrazuje atribut „word“. Rozdíl je v tom, že „word“ odlišuje velikost písmen, kdežto „lc“ nikoliv.

2) Frekvence – vlastní

Frekvenční limit:

| Úroveň | Atribut                            | Nerozlišovat velikost                   | Pozice                            | (Node) začít od                                       |
|--------|------------------------------------|---|-----------------------------------|---|
| 1.     | <input type="text" value="lemma"/> | <input type="checkbox"/>                | <input type="text" value="Node"/> | <input type="text" value="slova KWIC nejvíce vlevo"/> |
| 2.     | <input type="text" value="word"/>  | <input type="checkbox"/> nezaškrtovat – | <input type="text" value="1R"/>   | <input type="text" value="slova KWIC nejvíce vlevo"/> |

|    |      |  |    |                          |  |
|----|------|--|----|--------------------------|--|
|    |      | velikost   |    |                          |  |
| 3. | word | <input type="checkbox"/> nezaškrtovat –<br>chceme rozlišovat<br>velikost | 2R | slova KVIC nejvíce vlevo |  |

Dostanete pak následující frekvenční seznam:

Celkem: 4 položky (1 stránka)

|  | Filter                | lemma      | word | word     | Freq |  |
|--|-----------------------|------------|------|----------|------|--|
|  | <a href="#">p</a> / n | Jaroměřice | nad  | Rokytnou | 9779 |  |
|  | <a href="#">p</a> / n | Jaroměřice | NAD  | ROKYTNOU | 849  |  |
|  | <a href="#">p</a> / n | Jaroměřice | nad  | rokytnou | 6    |  |
|  | <a href="#">p</a> / n | jaroměřice | nad  | Rokytnou | 1    |  |

Podobně zkuste ověřit pravopis jména KOSTELCE NAD ČERNÝMI LESY.

**4. Obtížnější úkol: Zjistěte, zda se v následujících spojeních vlastní jména častěji skloňují, nebo zda častěji zůstávají nesklonná.**

Navrhněte postup, jak by to bylo možné zjistit.

města Brna, městě Brně, městu Brnu, městem Brnem – města Brno, městě Brno, městu Brno...

města Ostravy, městě Ostravě, městu Ostravě... – města Ostrava, městě Ostravě, městu Ostrava...

města Pardubic, městě Pardubicích, městu Pardubicím... – města Pardubice, městě Pardubice, městu Pardubice

**5. Pokuste se vymyslet úkol z pravopisu a z tvarosloví, který byste zadali žákům na střední škole. Měl by to být takový úkol, který by bylo možné řešit vyhledáváním v Českém národním korpusu (a který jsme ještě nedělali).**