

**UKORP (cvicny)**

**jméno**

**jméno:**

hodnocení: otázky 1–7 za 3 body; otázky 8–12 za 1 bod

*praktická část*

1.

V korpusu SYN2015 vyhledejte adjektiva vyskytující se za předložkou „pod“ (bez ohledu na velikost písmen). Uveďte tři nejfrekventovanější lemmata, které se za touto předložkou vyskytují

dotaz:

lemmata:

2.

V korpusu SYN2015 vyhledejte všechny tvary (bez ohledu na velikost písmen) adverbia „pěkně“, zjistěte jejich frekvence:

dotaz:

výsledek:

3. V korpusu SYN2015 vyhledejte všechna slova, která začínají na „post-“, a končí na „-ost“ (bez ohledu na velikost písmen); vypište všechny nalezené slovní tvary

dotaz:

výsledek:

4. Na základě korpusu SyD zjistěte, kdy začal převažovat slovní tvar „nést“ nad slovním tvarem n“nésti“

5. Prostřednictvím aplikace KWords určete klíčová slova z textu (viz www):

*Případ se stal ve středu ve tři hodiny a deset minut odpoledne. Lupič počkal, až z obchodu odejde poslední zákazník, a vešel dovnitř. Obešel pult, aby se dostal až ke květinářce, a vytáhl z kapsy revolver. „Dej mi tisícovku!“ křikl na ženu.*

*„Nejdřív jsem se hrozně lekla. Hlavou mi blesklo, co když opravdu vystřelí,“ vypráví přepadená květinářka Hana Klímová.*

*Líčí, že lupič byl vysoký, měl vousy, jaké dnes mladí nosí, a nevypadal nijak zanedbaně.*

*„Najednou jsem si řekla: Máš dvě zdravé ruce a děláš tohle? Ty hajzle! V pravé ruce jsem zrovna držela kytici, kterou jsem vážala, tak jsem levačkou popadla to první, co mi přišlo pod ruku,“ líčí.*

*Chytla zásobník na lepicí pásku vážící přes kilogram. „Hodila jsem ho po něm, jenže levou rukou jsem se těsně netrefila a rozbila keramický šálek na stoličce vedle něj,“ vypráví.*

*„On viděl, že z toho nic nebude, tak strčil zbraň zpátky do kapsy a odcházel. Já šla i s tou kytkou za ním, abych viděla, kudy jde, a mohla to říct policistům,“ dodává Klímová.*

*Před policejním autem si klekl na chodník a dal ruce za hlavu*

*Muž se nejprve ukryl do vedlejší ulice a vyčkával. Po chvíli jeli kolem policisté. „Právě vezli nějakou paní z výslechu,“ říká květinářka.*

*Jak lupič zahlédl policejní auto, zamával na ně, klekl si na silnici, kolt položil vedle sebe a dal ruce za hlavu.*

*Až na služebně se žena dozvěděla, že neměl opravdovou zbraň, ale hračku – kapslový revolver. To ale při přepadení nemohla vědět.*

*„Šestadvacetiletý muž měl pozitivní dechovou zkoušku na alkohol,“ sdělil teplický policejní mluvčí Daniel Vítek.*

*Policie pachatele obvinila z loupeže, za niž mu hrozí až deset let vězení. Z obavy, aby v minulosti třikrát odsouzený muž nepokračoval v trestné činnosti, zaslal vyšetřovatel státnímu zástupci podnět k podání návrhu na vzetí obviněného do vazby.*

Parametry vyhledávání:

- a) referenční korpus: PUB
- b) stop list (tj. do analýzy nezahrnovat): zájmena, předložky, spojky, čísla
- c) metoda: log-likelihood
- d) hladina významnosti: 0,001
- e) minimální frekvence: 3
- f) procento klíčových slov: všechny signifikantní typy

Výsledek: uveďte jak frekvenci v textu, tak v referenčním korpusu

6. V korpusu SYN2015 vyhledejte slova začínající na „roz-“ (bez ohledu na velikost písmen), musí jít pouze o adverbia; napište 3 nejfrekventovanější

dotaz:

frekvence:

7. V korpusu SYN2015 vyhledejte nejsilnější kolokace lemmatu „číst“ – musí jít jen o pravé kolokáty do vzdálenosti 2, min. frekvence v korpusu je 5, minimální frekvence kolokátů je také 5. Vypište 5 lemmat, která tvoří nejsilnější kolokaci podle míry logDice.

***teoretická část (označte správnou odpověď)***

**8. Korpusová data se skládají**

- a) výhradně z autentických textů
- b) výhradně ze speciálních textů vytvořených lingvisty pro účely jazykové analýzy
- c) z autentických textů pouze spisovného jazyka
- d) z autentických textů i ze speciálních textů vytvořených lingvisty pro účely jazykové analýzy

**9. Diachronní jazykový korpus češtiny, které je v Českém národním korpusu**

- a) je morfologicky označovaný, nikoliv však lemmatizovaný
- b) je lemmatizovaný, nikoliv však morfologicky označovaný
- c) je morfologicky označovaný i lemmatizovaný
- d) není morfologicky označovaný ani lemmatizovaný

**10. Ve v následující posloupnosti slov „hezká, hezký, hezké, nejhezčí, dcera“**

- a) je 5 lemmat, 5 slovních tvarů (types), 5 tokenů
- b) jsou 3 lemmata, 5 slovních tvarů (types), 4 tokeny
- c) jsou 3 lemmata, 4 slovní tvary (types), 5 tokenů
- d) jsou 2 lemmata, 5 slovních tvarů (types), 5 tokenů

### **11. Paralelní korpusy**

- a) jsou složeny z originálů a překladů stejných textů
- b) jsou složeny z dialogických mluvených textů jednoho jazyka
- c) jsou složeny ze spisovných a nespisovných variant sledovaných výrazů daného jazyka
- d) jsou složeny z historických textů a jejich současných ekvivalentů daného jazyka

### **12. Kolokace je**

- a) specifický druh lemmatizace zaměřený na ustálené spojení dvou nebo více slov
- b) typ syntaktické anotace v korpusu SYN2015
- c) ustálené, syntagmatické spojení dvou (nebo víc) slovních tvarů v blízkém kontextu
- d) asociační míra používaná v SYN2015