

- V čem spočívají základní metody poznávání fungování jazyka?
- Jaký je rozdíl mezi sbírkou textů a korpusem v moderním slova smyslu?
- Je počet slov v korpusu objektivním měřítkem pro hodnocení jeho kvality?
- Proč nelze užívat texty na internetu tímž způsobem jakým se využívají jazykové korpusy?
- Co je to vertikála?
- Jaké jsou čtyři hlavní rysy korpusu v moderním slova smyslu.
- Jmenuj nějaké typy strukturních značek.
- Co je to OCR?
- Vysvětli rozdíl mezi transkripcí a transliterací.
- Proč nejsou mluvené korpusy přepsány fonetickou transkripcí?
- Charakterizujte pojetí synchronního korpusu řady SYN
- K čemu mohou sloužit autorské korpusy?
- Uveďte důvodu malého rozsahu korpusů mluveného jazyka?
- Existují korpusy dialektologické?
- Je možné zkoumat synchronní stav jazyka starších jazykových období?
- Která pracoviště zpřístupňují diachronní data?
- Lze zkoumat česká diachronní data pomocí stejných nástrojů (manažeru kontext) jako data synchronní?
- Existují lemmatizovaná diachronní data?
- Je možno zohlednit časové hledisko u specializovaných korpusů (vývoj jazyka jedince, vývoj podoby žánru)?
- Vyjmenuj jednotlivé kroky automatického zpracování textu, jejichž výsledkem je lemmatizovaný a morfologicky taggovaný text.
- Jak zjistíme velikost korpusu?
- Uveďte příklady, kdy jeden slovní tvar lze interpretovat vícero lemmaty.
- Uveďte příklady, kdy jeden slovní tvar lze interpretovat jediným lemmatem, ale má více významu gramatických kategorií pádu a čísla.
- V jakém formátu se zobrazují výsledky vyhledávek v korpusu?
- Co jsou to regulární výrazy? Uveď příklad.
- Jak se jmenuje jazyk, ve kterém se zadávají složitější dotazy do korpusového vyhledávače.
- Vyjmenuj důležité funkce korpusového manažeru.
- Lze ve všech nástrojích, které nabízí rozhraní Kontext, pracovat se zadáním dotazu pomocí regulárních výrazů?

- Vyber si jeden z nástrojů nabízených rozhraním KonText, který jsi vyzkoušel, a popiš úkol, který tě zaujal.
- Které z nástrojů nabízených rozhraním KonText nabízejí data čerpaná z paralelního korpusu?
- Jaká jsou omezení funkcí, které porovnávají data z psaných a mluvených korpusů?
- Který byl první mluvený korpus zveřejnění ÚČNK?
- Která další pracoviště se podílí na budování korpusů mluveného jazyka zveřejněných ÚČNK?
- V jaké podobě jsou uložena data mluvených korpusů (typy transkripce, přístup k nahrávkám)?
- Které publikace vznikly na základě mluvených korpusů češtiny?
- Které z mluvených korpusů jsou lemmatizované a morfologicky označované?
- Pro který typ korpusu je důležitá deduplikace textů.
- Existují paralelní korpusy, v nichž by byly paralelně texty jednoho jazyka?
- Jaké typy textů (žánry) tvoří převážně texty v korpusu InterCorp?
- Vyjmenuj důležité konference specializované na KL
- Vyjmenuj alespoň dvě gramatiky založené na korpusech
- Vyjmenuj důležité časopisy zaměřené na KL