

Práce s InterCorpem

Adresa: <http://www.korpus.cz/intercorp/>

Odkaz: **Hledání v korpusu**

Přihlášení: Uživatelské jméno: knj
heslo: brno

Vstup: „přihlaš se“
„Nový dotaz“

Objeví se tabulka se sloupci JAZYKY a DOKUMENTY

Ve sloupci JAZYKY zatrhnout „cs“ a „de“ (popř. jiný jazyk, který Vás zajímá). Počet dokumentů se zredukuje.

Přístup k dotazům: „Jdi k dotazu“, objeví se následující maska:

Korpus: intercorp_cs		Korpus: intercorp_de	
Lem ma:	<input type="text"/>	Lem ma:	<input type="text"/>
Slovní spojení:	<input type="text"/>	Slovní spojení:	<input type="text"/>
Word Form:	<input type="text"/>	Word Form:	<input type="text"/>
	Match case: <input type="checkbox"/>		Match case: <input type="checkbox"/>
CQL:	<input type="text"/>	CQL:	<input type="text"/>
Default attribute:	<input type="text"/>	Default attribute:	<input type="text"/>
Řádků na stránce		20	
Spust'			

Dotaz: Dotaz vpisujete buď do české paralely **Korpus: intercorp_cs** nebo německé **Korpus: intercorp_de**.

Spuštění hledání: „Spust'“

Vysvětlení pojmů:

Lemma = základní tvar (např. *dům*)

Slovní spojení = kombinace slov (např. *malý dům*)

Word Form = interval (část slova) (např. *dome.**)

CQL = morfologická značka

Nejčastěji asi použijete „Word form“

Syntax dotazů: tečka „.“ = libovolný znak; hvězdička „*“ = libovolný počet znaků

např. *dome.** najde *domek, domech, domeček* atd.

Ve „slovním spojení“ se hledají po sobě jdoucí slova:

např: *mal.* dom.** najde *malý domek, malých domečků, malém domku, maličkého domku* atd.

Hledáte-li část slova a diakritické znaménko, pak mezi nimi musí být mezera!

např.: *A.*[mezerník]*!

Hledání:

Pokud zadaný jev v korpusu není nebo dotaz byl špatně položen, objeví se na horní liště červené „no lines“

Je-li výskytů hodně, hledání nějakou dobu trvá (na horní liště „stále hledám“)

Až korpus skončí hledání, objeví se na dolní liště „Export: xls“

Export:

Až korpusový manažer dokončí hledání (na horní liště zmizí „stále počítám“), objeví se na dolní liště: „Export: xls“. Soubor už tedy můžete exportovat. Je však záhodno získat ještě informace o tom, z jakých zdrojů texty pochází. **Proto nejprve na první straně konkordancí, pod horní lištou rozevřete „Zobraz možnosti“.** Tam ve sloupci „References“ zatrhnete „doc.id“ (druhá položka shora) a „Použij“.

Teprve potom exportujte:

Klikněte na xls, soubor excelu se uloží do počítače (typicky do složky „Downloads“)

Práce s excellem:

1) Výskyty jsou v excelu seřazeny stejně tak, jak se zobrazily v InterCorpu.

Pokud jste použili některou z možností („Zobraz možnosti“= tzv. metainformace), je dobré je převést do jedné řádky s paralelními konkordancemi: (pokud ne, přeskočte na bod 2) označte celý soubor a seřaďte podle prvního sloupce, řazení se záhlavím, tedy „podle intercorp cs“, řazení podle barvy textu (to je důležité!!). Pak vytvořte nový sloupec A a od 2. řádky vložte barevné metainformace.

2) Výskyty jsou v excelu seřazeny stejně tak, jak se zobrazily v InterCorpu.

Hledaný výraz však není červeně, ale je vždy v příslušné buňce oddělen znakem „rovná se“ („=“) před hledaným výrazem (KWIC) a dvěma „rovná se“ („==“) za ním.

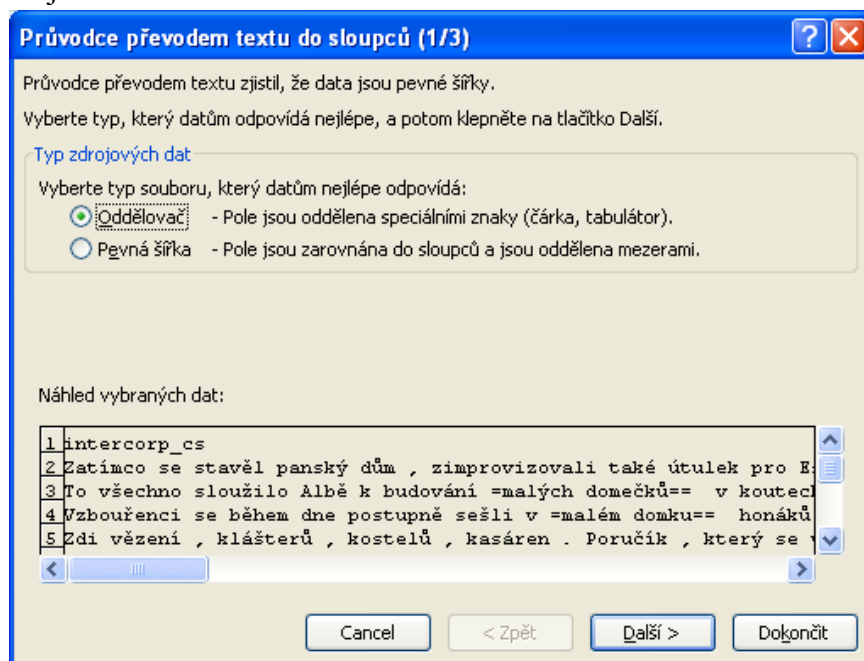
Pro jednodušší třídění doporučuji KWIC vložit do samostatného sloupce. To se dělá následovně:

Doporučuji přesunout sloupec s hledanými výrazy (=XYZ==) tak, aby byl jako sloupec poslední.

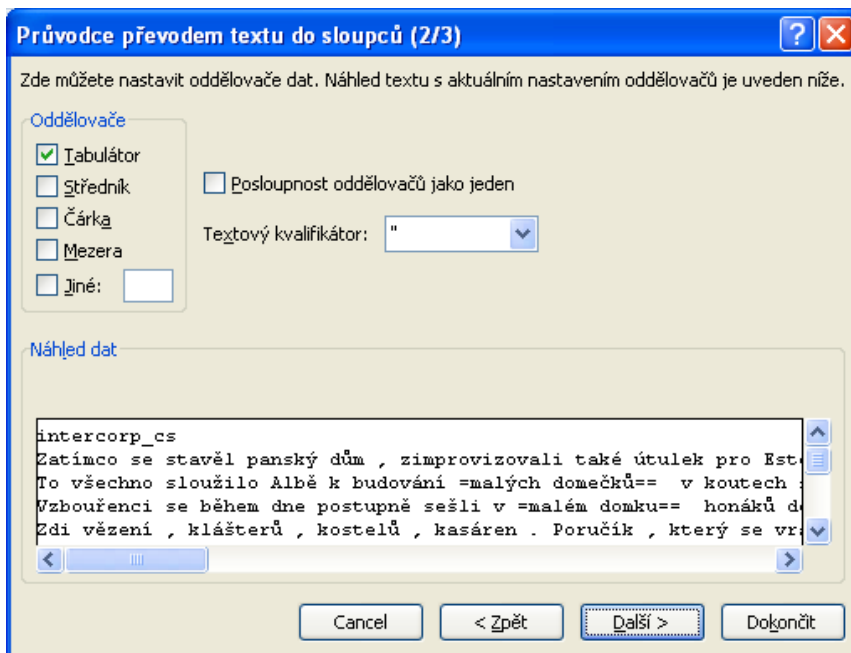
Označit příslušný sloupec

v odkazu „data“ zvolit „text do sloupců“

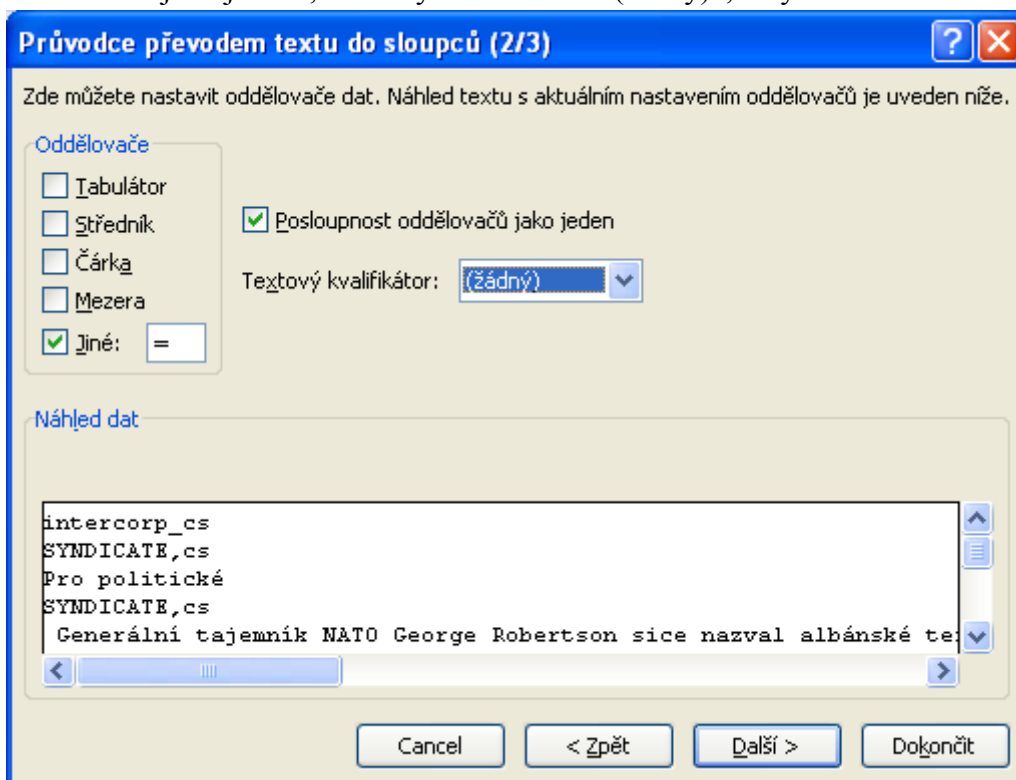
Objeví se:



Zatrhnout „Oddělovač“ (přednastavená je „Pevná šířka“) a jít na „další“



Zde odkliknout „Tabulátor“ a: zatrhnout „Jiné“, vepsat „=“, zatrhnout „Posloupnost oddělovačů jako jeden“, Textový kvalifikátor –(žádný)“, tedy takto:



jít na „Další>“ a v následujícím okně „Dokončit“.
KWIC se přesune do samostatného sloupce.