

Někdo rozumí počítačům. Počítače nerozumí nikomu. Naučíme je to?

# Corpus Engineering

---

*Otázka:* **Vejde se celá knihovna do počítače?**

*Odpověď:* **Ano.**

---



Textový korpus je soubor souvislých textů. Lingvisté na něm sledují, jak se jazyk proměňuje. Na velkých korpusech, kde už je statistický soubor dostatečný, se počítačové programy učí, jak lidé jazyk užívají. Korpusy, které vytváříme nebo spravujeme v CZPJ, obsahují miliardy slov. Pro srovnání: nejdelší román (zapsaný v Guinnessově knize rekordů) má cca 1 200 000 slov.

Centrum zpracování přirozeného jazyka  
Fakulta informatiky  
Masarykova univerzita

