



VÝZKUMNÉ ZAMĚŘENÍ A POSLÁNÍ

- Výzkumná infrastruktura zajišťovaná Ústavem Českého národního korpusu FF UK a Ústavem teoretické a počítačové lingvistiky FF UK.
- Projekt Český národní korpus (ČNK), založený v roce 1994, se zaměřuje na kontinuální mapování češtiny ve všech dostupných dimenzích (časové, regionální i žánrové).
- ČNK vytváří a zpřístupňuje rozsáhlé elektronické sbírky textů (jazykové korpusy), které slouží jako báze pro jakýkoli jazykově orientovaný výzkum.
- Vedle toho se věnuje rozvoji metodologie empirického lingvistického výzkumu a vývoji nástrojů na vytěžování jazykových korpusů.
- ČNK od roku 2012 slouží jako výzkumná infrastruktura pro mnoho oblastí společenských a humanitních věd (zejm. pro lingvistiku, psychologii, sociologii, historii, počítačové zpracování přirozeného jazyka ad.).
- Pro svoji rozsáhlou a kvalitně budovanou základnu je ČNK vyhledávaným partnerem pro zahraniční vědeckou spolupráci.
- ČNK se zaměřuje i na poskytování poradenství, vytváření analýz pro vědecké i popularizační účely, poskytování dat pro výzkum češtiny i dalších jazyků pro srovnání s ní a automatické zpracování jazykových dat.

NEJVĚTŠÍ ŘEŠENÉ PROJEKTY

- Projekt ČNK poskytuje svoje data a nástroje všem registrovaným uživatelům prostřednictvím portálu www.korpus.cz.
- Kromě obecných nástrojů na práci s rozsáhlými jazykovými korpusy (KonText) na portálu zájemci naleznou i specializované nástroje, např. SyD pro analýzu variant, Treq pro vyhledávání překladových ekvivalentů nebo KWords pro zjišťování klíčových (prominentních) slov v textech.
- ČNK úzce spolupracuje např. s nakladatelstvím Fraus na vytvoření pomůcek pro výuku češtiny a pravidelně organizuje praktické semináře a workshopy zaměřené na práci s korpusy pro veřejnost i odborníky z řad učitelů, překladatelů nebo redaktorů.

NEJVĚTŠÍ DOSAŽENÉ ÚSPĚCHY

- Vytvoření a zveřejnění největší a nejpestřejší veřejně přístupné databáze českých textů, která je volně dostupná pro badatelské a pedagogické účely.
- Vytvoření sady nástrojů pro práci s rozsáhlými korpusy (KonText, SyD, KWords, Treq).
- Vývoj unikátních nástrojů pro počítačové zpracování českých textů.
- Zařazení projektu ČNK na cestovní mapu výzkumných infrastruktur na období 2016–2022.

PUBLIKACE

- Bartoň, T. a kol.: *Statistiky češtiny*. Nakladatelství Lidové noviny, Praha 2009.
- Cvrček, V. a kol.: *Mluvnice současné češtiny*. Nakladatelství Karolinum, Praha 2010.
- Čermák, F., Křen, M. (eds.): *A Frequency Dictionary of Czech: Core Vocabulary for Learners*. Routledge, London 2011.
- Čermák, F., Rosen, A.: *The case of InterCorp, a multilingual parallel corpus*. *International Journal of Corpus Linguistics*, 2012, 17(3), 411–427.
- Čermák, F., Křen, M. (eds.): *Frekvenční slovník češtiny*. Nakladatelství Lidové noviny, Praha 2004.
- Čermák, F., Cvrček, V.: *Slovník Bohumila Hrabala*. Nakladatelství Lidové noviny, Praha 2009.

