

# Natural Language Processing Centre FI MU, Brno

<http://nlp.fi.muni.cz>

## 1. Obsah výzkumu

Centrum zpracování přirozeného jazyka na Fakultě informatiky Masarykovy univerzity je zaměřeno na získání teoretických i aplikovaných výsledků v oblasti zpracování přirozeného jazyka. Naše práce se týká zejména korpusové lingvistiky, lexikálních databází, reprezentace znalostí, reprezentace významu výrazů přirozeného jazyka a využití metod strojového učení pro automatické zpracování textů.

Z hlediska navazujících aplikací je výzkum zaměřen na komunikaci typu člověk-stroj a na internetové technologie soustředěné na analýzu, vytěžování a generování textů a řeči.

## 2. Klíčoví výzkumníci

prof. PhDr. Karel Pala, CSc.

doc. RNDr. Aleš Horák, Ph.D.

doc. Mgr. Pavel Rychlý, Ph.D.

RNDr. Adam Rambousek, Ph.D.

RNDr. Zuzana Nevěřilová, Ph.D.

RNDr. Pavel Šmerk, Ph.D.

RNDr. Miloš Jakubíček, Ph.D.

Mgr. et Mgr. Vít Baisa, Ph.D.

RNDr. Vojtěch Kovář, Ph.D.

## 3. Klíčové metody a technologie

Počítačové zpracování přirozeného jazyka, strojové učení, neuronové sítě, zpracování velmi velkých dat (textů), počítačová analýza textů (morfologická, syntaktická, sémantická), reprezentace a odvozování znalostí, dolování znalostí, dialogové systémy, stylometrie, analýza autorství, komunikace člověk-stroj.

## 4. Top 3 výsledky

Články:

- KILGARRIFF, Adam, Vít BAISA, Jan BUŠTA, Miloš JAKUBÍČEK, Vojtěch KOVÁŘ, Jan MICHELFEIT, Pavel RYCHLÝ a Vít SUCHOMEL. The Sketch Engine: ten years on. *Lexicography*, Springer Berlin Heidelberg, 2014, roč. 1, č. 1, pp.7-36. ISSN 2197-4292.
- MEDVEĎ, Marek a Aleš HORÁK. AQA: Automatic Question Answering System for

Czech. In Text, Speech, and Dialogue 19th International Conference, TSD 2016 Brno, Czech Republic, September 12–16, 2016 Proceedings. Switzerland: Springer International Publishing, pp.270-278, 2016.

- CVRČEK, František, Karel PALA a Pavel RYCHLÝ. Behaviour of Collocations in the Language of Legal Subdomains. In LREC 2014 - Ninth International Conference On Language Resources and Evaluation. Paris: European Language Resources Assoc-ELRA, 2014.

Aplikované výsledky:

- Slovníková platforma DEB (např. Internetová jazyková příručka, Encyklopedie češtiny, Centrální registr terminologických tezaurů, ...)
- Korpusový manažer Manatee, Sketch Engine
- Analyzátor jazyka Majka/Desamb

## 5. Top 5 projektů

- LINDAT/CLARIN - Výzkumná infrastruktura pro jazykové technologie (MŠMT 2015-2022)
- Hyperintensionální usuzování nad texty přirozeného jazyka (GAČR 2018-2020)
- Harvesting big text data for under-resourced languages (CZ-NO 2014-2017)
- Analýza přirozeného jazyka v prostředí internetu (MV 2010-2014)
- Pattern Recognition-based Statistically Enhanced MT (EU 2010-2012)

## 6. Spolupráce s průmyslem

- Lexical Computing/LCL - strategický partner FI MU, zpracování velmi velkých textových dat v mnoha jazycích, Sketch Engine
- Konica Minolta - strategický partner FI MU, analýzy a dolování z obchodních dokumentů
- Seznam.cz - jazykové zpracování češtiny