LangTech and DigiLing projects

Zdeněk Žabokrtský

Perlová Voda, 18th September 2018

LangTech

- LANGuage TECHnology (computational aspects of NLP)
- an EU-financed project (European Structural and Investment Funds through Czech OPVVV)
- the main aim: modernize our PhD program
- 3 key areas: DL, multilinguality, (big) data processing
- duration 2017-2022
- budget 9+ MKč
- team: Ondřej Bojar, Jan Hajič, Jindřich Helcl, Jana Hamrlová, Jindřich Libovický, David Mareček, Pavel Pecina, Martin Popel, Rudolf Rosa, Dan Zeman, Zdeněk Žabokrtský,

LangTech: progress in 2017-2018

- 10+ modernized or newly created courses
- elaborated web-publishing infrastructure, graphical templates for web and slides . . .
- five invited foreign academic guests
- five stays abroad of ours
- new accredditation application prepared and submitted
- http://ufal.mff.cuni.cz/langtech

LangTech: covered courses

- NPFL117 Deep Learning seminar
- NPFL118 NLP on Computational Cluster
- NPFL120 Multilingual Natural Language Processing
- NPFL116 Compendium of Neural Machine Translation
- NPFL068 Stat. Methods in Natural Language Processing II
- NPFL070 Language Data Resources
- NPFL087 Statistical Machine Translation
- NPFL092 Technology for NLP
- NPFL097 Selected Problems in Machine Learning
- NPFL114 Deep Learning

- course modernization to be continued
- selected PhD students to be sent to research stays abroad
- the biggest challenge:
  - the main project output (in terms of project evaluation): 20 new PhDs enrolled by 2021
  - ÚFAL master program insufficient for feeding ÚFAL PhD program
  - new PhD students badly needed

- 2016-2019 EU funded project
- partners: Ljubljana, Mainz, Leeds, Zagreb, Prague, Brno
- on-line course materials, incl. audio/video recordings, interactive quizzes, hands-on activities
- Moodle used as the main platform

DigiLing courses

- Variability of languages in space and time
  - taught by Magda Ševčíková, Anja Nedoluzhko and Šárka Zikánová
  - 350 slides in 28 short lectures in 10 learning units
- Introduction to text processing
  - Lucie Lukešová (Chlumská) and Pavel Vondřička (FF UK)
- Post-Editing Machine Translation
- Introduction to Python for linguists
- Computational lexicology and lexicography
- Localization Tools
- Mining and Managing Multilingual Terminology
- Introduction to SketchEngine

- one-week summer school in June 2019
- selected learning units from each course to be tested on students
- 30 students expected