Institute of Formal and Applied Linguistics (ÚFAL)
Projects Related to LINDAT/CLARIAH-CZ

Jan Hajič
LINDAT/CLARIAH-CZ in 2020/1

- LINDAT = CLARIN and DARIAH in Czechia
  - European networks for supporting research (in SSH)
    - CLARIN: Language resources and technology
    - DARIAH: Digital humanities and arts
  - LINDAT/CLARIAH-CZ: 2019-2022
    - Combines membership in EU CLARIN and DARIAH networks
    - 11 institutions in the Czech combined network
    - UK, MU, Academy of Sciences, National and Moravian Libraries, NG, NFA
    - Center for Visual History Malach: under LINDAT (& Herzl Center)
  - Provides also large part of computing infrastructure of UFAL
  - New hardware due in 2021 (OP VVV VI 2)
    - Lots of GPUs 😊
• “Internal” projects – offer datasets, annotation, curation, tools, services
• Repository development (CLARIN DSpace), HW and other IT support
• Universal Dependencies project
  – General support, official editions
  – Tools: UDPipe (now UDPipe 2), soon to have: NER
• Prague Dependency Treebank – Consolidated edition
  – Approx. 4MW, manual morphology
    • PDT, PCEDT, PDTSC, Faust; TR annotation same, a-layer: auto (PDT: manual), new morph. dictionary
• SynSemClass
  – Multilingual (cs, en so far) event type ontology
    • A.k.a synonym dictionary of verbs, with valency, linked to corpora and other lexical resources
• TEI:TOK
  – Platform for rich annotation and multimedia content
• Support for other projects (GAČR, TAČR, H2020)
• Running services (https://lindat.cz/en/services), maintaining HW and basic SW
• International cooperation (support or development)
  – ELG European Language Grid, EU H2020 Call 29a
    • „Marketplace“ for mainly commercial use of Language Technologies
    • Partners: DFKI, UK, Sheffield, ELRA, commercial partners
    • Pilot projects: organized by us (2M EUR), now 2nd Call (Oct 1)
  – Some UFAL tools provided – UDPipe, MT; harvesting repository

• SSHOC, Infrastructural H2020 project
  – Part of CLARIN ERIC participation
  – UFAL/LINDAT provides basic language tools for IR, MT for Social Surveys (COVID)

• Humaine-Al-Net: AI Center of Excellence, Call 48
  – Small participation, w/DFKI Saarbrucken: microproject (anyone…?)

• ELG and SSHOC 2019-2022, Humaine-Al-Net 2020-2023
• Two more submitted in 2020: ELG continuation, LANGEQ
Shared Tasks

- Supporting Shared Tasks by data, manpower, participation
- CoNLL Shared Task on Meaning Representation Parsing (MRP)
  - Run in 2019 (and 2020, Cz added)
  - Stephan Oepen & colleagues
  - Co-organization
- Data/repo support for WMT
The Future

• Evaluation 2021 (!!!)
  – We have to pass w/either of the two highest marks, otherwise LINDAT will not get funding 2023- (same for equipment)

• Internal LINDAT members:
  – your main work for October-November/December 2020!

• Others:
  – we might ask for help

• 80 page questionnaire, with partners, attachments, indicators, budget, …
  – Deadline: Dec. 10, 2020; interview in front of panel: May/June 2021; verdict: July 2021, full proposal for 2023-2026(9?) due in early 2022

• Extended to EHRI (Holocaust research infrastructure)
  – Natural partner for CVHM, lead: Masarykův ústav AV ČR

• Main points: SoA and new tools, equipment renewal 23-29, new resources, annotation, support to Digital Humanities, new services at partners, international cooperation (support for more projects)