

WELCOME ADDRESS AND INTRODUCTION TO THE WORKSHOP

Jan Hajič, ÚFAL MFF UK & LINDAT/CLARIAH-CZ

May 3, 2022



WELCOME (VIRTUALLY) TO PRAGUE FOR THE 3RD ELRC CZ WORKSHOP

- Charles University (Institute of Formal and Applied Linguistics - ÚFAL)
 - Supporter of statistical MT development since the 1990s
 - EU projects EuroMatrix, EuroMatrixPlus, T4ME, QTLeap, QT21, Faust, HimL, ... now ELG
 - Data collection
 - now distributed through the LINDAT/CLARIAH-CZ Research Infrastructure, hosted by UFAL
 - Language Technology research and development
 - Machine translation (multiple winner of the WMT competitive shared task)
 - CUBBITT system by M. Popel featured in Nature Communications
 - Basic text analytics (winner of a CoNLL shared task with the UDPipe system)
 - Language Resources – since 1990s
 - Prague Dependency Treebanks, Universal Dependencies publisher and maintainer
 - 300+ other datasets

LINDAT/CLARIAH-CZ REPOSITORY

- Preservation
- Identification
- FAIRness
- Single sign-on
- Open licences
- Certified by CTS, CLARIN, OpenAire
- Depositing service (free for SSH)

Central Repository for data, software and metadata in language technologies and humanities

https://lindat.mff.cuni.cz/repository/xmlui/discover

LINDAT CLARIAH-CZ

Search Catalogue Education Projects Tools Services About

LINDAT/CLARIAH-CZ Repository Home / Search

universal dependencies Search

Advanced Search

Limit your search

Author Subject Rights Language (ISO) Type Contain Files Community

Showing 1 through 10 out of 658 results

1 2 3 > 66

Corpus LINDAT / CLARIAH-CZ

Deep Universal Dependencies 2.8
(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2021-06-17)

Author(s):
Zeman, Daniel and Drohanova, Kira

This item contains 1 file (1.13 GB). Publicly Available

Corpus LINDAT / CLARIAH-CZ

Deep Universal Dependencies 2.9
Linguistics Consortium /

Author(s):
Zeman, Daniel ; et al.

show everyone

This item contains 3 files (534.14 MB). Publicly Available

Browse

> All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

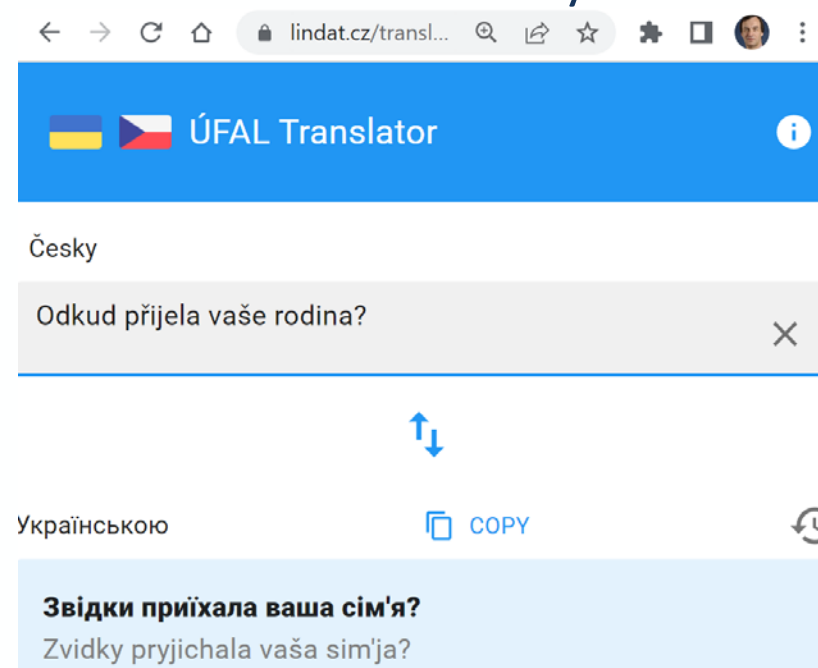
Help Desk

TOOLS AND SERVICES BY LINDAT/CLARIAH-CZ AND ÚFAL MFF UK

- Tools developed at Charles University and others available
 - From LINDAT/CLARIAH-CZ repository
 - Open Source
 - Trained models
 - Examples: MT, UDPipe, lexical tools etc.
- Services run by LINDAT/CLARIAH-CZ
 - Machine translation for (virtually) ~50 language pairs
 - Now includes Czech <-> Ukrainian MT
 - Language analysis (UDPipe) for 100+ languages
 - Access to lexical resources
 - Named entity recognition
 - ... and more

TOOLS AND SERVICES BY LINDAT/CLARIAH-CZ AND ÚFAL MFF UK

- Run for everybody
 - Terms of use: free for research, public, and for limited use for everybody
 - Commercial licenses available for Charles University-authored tools and model
- All tools run as services
 - E.g., Machine Translation
 - Example: Czech – Ukrainian
 - Android app coming soon
 - With ASR/TTS
 - spoken translation



TOOLS AND SERVICES BY LINDAT/CLARIAH-CZ AND ÚFAL MFF UK

- Many other language pairs
- English-cs, fr CUBBITT system (2021 WMT en-cs winner)
- API available for external programmatic use

LINDAT Translation

Translate Docs

The translation service is available for *personal and non-commercial use* (see [terms of use](#) for more details).

Models

en-fr (English->French (CUBBITT))

Source en **Target** fr

advanced

Input sentences

Before completing Hong Kong's mandatory three-week quarantine back in February, CNN features producer Zahra Jamshed sought advice from a number of seasoned hotel quarantiners, who suggested she put a routine in place from the get go. Although Jamshed was able to work from the hotel during her quarantine period, so her days had a regular structure to them, she decided to create a "to do" list with a number of allotted tasks so that each day would feel purposeful.

Translation

Avant de terminer en février la quarantaine obligatoire de trois semaines à Hong Kong, la productrice de CNN Zahra Jamshed a demandé conseil à un certain nombre de quarantenaires chevronnés des hôtels, qui lui ont suggéré de mettre en place une routine dès le départ. Bien que Jamshed ait pu travailler à l'hôtel pendant sa période de quarantaine, de sorte que ses journées avaient une structure régulière pour eux, elle a décidé de créer une liste de choses à faire avec un certain nombre de tâches assignées afin que chaque journée se sente utile.

Translate Choose file

ENJOY THE WORKSHOP!

hajic@ufal.mff.cuni.cz

