

Sharing data in repositories

Class #14, May 16 2023

Barbora Hladká hladka@ufal.mff.cuni.cz

Data

= information in digital form
for computer processing

- text, audio, video,
image, software

Metadata

= data about data

Data

Mozilla Firefox

MAZON - Mozilla Firefox

https://quest.ms.mff.cuni.cz/teitok-dev/teitok/teaching/mazon/index.php?action=file&id=ces/A/

TEITOK

Jiří Mirovský
Available Corpora

MAZON

Home
COL Search
PMLTQ Search
Facsimile Search
Named Entities

DEV Home

user: JIMI

Admin

Na Letné 25. srpna 1919.
Velevážený pane **kollego**,

Váš milý lístek je nyní nesnadno, než z mladších kolegů rodiny. Jeden vša přihlásí se ihned. Keplerova ul. 219 najdeme počátku p. kapitán snad si na několik dní uč. laskavou zprávou Gotfriedovi, kupci v poněvadž je tu bratr ze Slovenska a tím můj pravidelný život je porušen. Tím také, prosím, omluvte laskavě

kollego
kollega
NOUN
UD POS tag
National POS
tag
NNMSS—A—
Animacy=Anim,
Case=Voc, Gender=Masc,
Number=Sing,
Polarity=Pos
vocative

ndělí.) Je
te některého
ova a bez
rdi) a
IV,
voval,
šho. Kdyby
řejal bych
om o
ímu

Na Letné 25. srpna 1919.
Velevážený pane kollego,
Váš milý lístek je nyní nesnadno, než z mladších kolegů rodiny. Jeden vša přihlásí se ihned. Keplerova ul. 219 najdeme počátku p. kapitán snad si na několik dní uč. laskavou zprávou Gotfriedovi, kupci v poněvadž je tu bratr ze Slovenska a tím můj pravidelný život je porušen. Tím také, prosím, omluvte laskavě

Metadata

Cote	Date	Type de document	Nb de f.	Langue	Auteur du document	Lieu
AMA.8.7.1	10.03.1908	Lettre manuscrite	1	allemand	Hartmann, Erich	Bautzen
AMA.8.12.34	24.11.1913	Lettre dactylographiée	1	allemand	Böhme, Erich	Berlin
AMA.8.13.40	26.04.1914	Carte manuscrite	1	allemand	Irmer, Hermann	Harkov
AMA.8.13.27	01.04.1914	Lettre manuscrite	1	allemand	Leskien, August	Leipzig
AMA.8.13.31	06.04.1914	Lettre manuscrite	1	anglais	Minns, Ellis H.	Cambridge
AMA.8.12.12	04.04.1913	Lettre dactylographiée	1	anglais	Miller, Arthur William Kaye	Londres
AMA.8.12.25	29.07.1913	Lettre manuscrite et dactylographiée	1	anglais	Miller, Arthur William Kaye	Londres
AMA.8.15.4	06.05.1919	Lettre manuscrite	2	bulgare	Stoilov, Nikola	indéterminé
AMA.8.13.37	16.04.1914	Lettre manuscrite	1	français	Demidov, Elim Pavlovich	Athènes

Data repository

= a digital infrastructure to share data

- i.e. to preserve data and help others to find them
- for many subjects, check e.g.,
 - <https://fairsharing.org/>
 - <https://www.re3data.org/>

LINDAT repository :: <https://lindat.cz>

initially LINGuistic DATa only, non-linguistic content now as well

Catalogue

LINDAT/CLARIAH-CZ is a Czech centre for data providing certified storage and natural language processing services

Who can download?
Who can deposit?
Publication schedule

CLARIN CENTRE B OPEN ACCESS

Cite Deposit

Linguistic (language) data

= resources for Linguistics and Natural Language Processing

corpora

Corpus LINDAT / CLARIAH-CZ

SYN v9: large corpus of written Czech

(Charles University, Faculty of Arts, Institute of the Czech National Corpus / 2021-12-05)

Author(s):
Křen, Michal ; et al.

▶ show everyone

 This item contains 1 file (21.87 GB).

Academic Use 

[url](#)

Linguistic (language) data

= resources for Linguistics and Natural Language Processing

annotated corpora


Corpus LINDAT / CLARIAH-CZ

Czech Named Entity Corpus 1.1

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2014-01-09)

Author(s):
Ševčíková, Magda ; Žabokrtský, Zdeněk ; Straková, Jana and Straka, Milan

This item contains 1 file (10.48 MB).

Publicly Available 

[url](#)

Linguistic (language) data

= resources for Linguistics and Natural Language Processing

treebanks

Corpus
LINDAT / CLARIAH-CZ


Universal Dependencies 2.9

(Universal Dependencies Consortium / 2021-11-15)


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

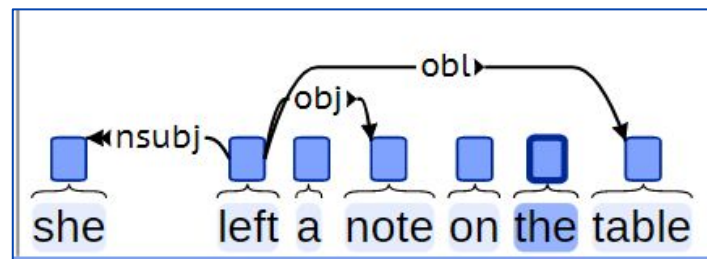
▶ show everyone

This item contains 3 files (534.14 MB).



Publicly Available





see Lec. #6

Linguistic (language) data

= resources for Linguistics and Natural Language Processing

lexicons


LexicalConceptualResource LINDAT / CLARIAH-CZ





[Bosworth-Toller's Anglo-Saxon Dictionary online](#)

(Charles University, Faculty of Arts, Department of English Language and ELT Methodology / 2021-04-09)

Author(s):
Tichý, Ondřej ; et al.

► show everyone

 This item contains 3 files (689.16 MB).

 Publicly Available   

Linguistic (language) tools

- Natural Language Processing tools
 - Machine translation, UDPipe, NameTag (see Lec. #7), ...
- Lexicons
 - Morphology - MorphoDiTa (basic morphological analysis and synthesis)
 - Syntax - Valency lexicons for Czech and English
 - Semantics - SynSemClass (synonym verbal lexicon alias event-type ontology)
- Search services
 - many corpora, all lexicons – KonText, PML-TQ, TEITOK (see Lec. #7)

Linguistic (language) tools

- Can be downloaded
- Users install them at their computing environment (PC, notebook, OS, ...)
- Users run them on their data locally (or any other computing facility, cloud, etc.)

Language tools vs. applications

- Web applications accessed through web browsers
- User opens service's web application page, enters data or uploads file
- Run link at <https://lindat.cz/services>
 - e.g., [NameTag](#)

For twenty-two years, **Bohumil Veselý** filmed important personalities, both well-known and lesser-known, on the courtyard balcony of his flat at 38 **Školská Street** in **Prague's New Town**.

Language tools vs. services :: <https://lindat.cz/services>

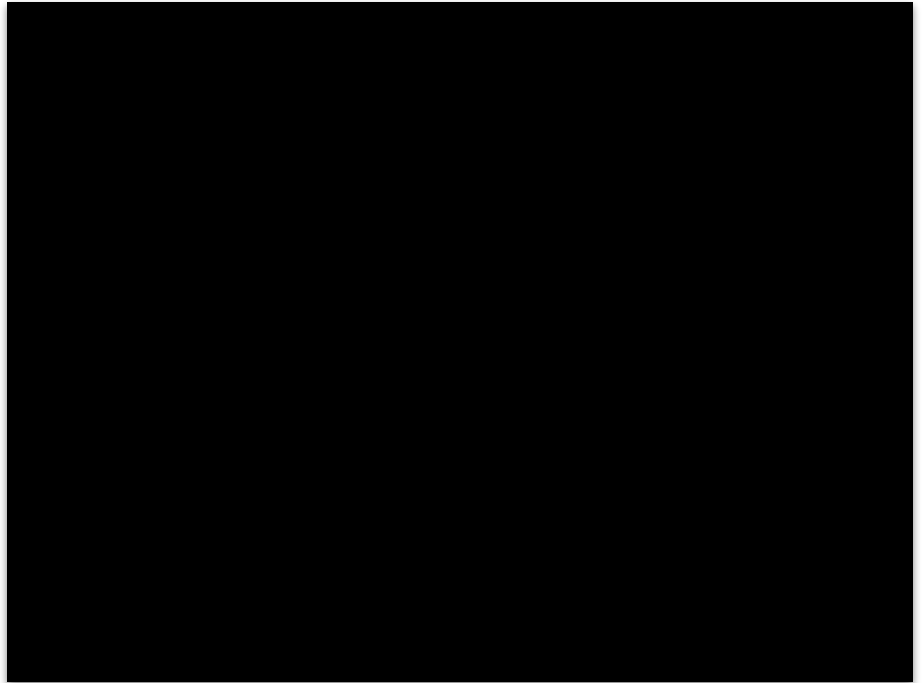
- Tools running @LINDAT in the background, 24/7
- Each service has its application
- Users can use it remotely, by calling scripts, or from their own code
- Means: Application Programming Interface (API)
 - = a software that allows two applications to talk to each other (send request, return response), e.g., each time you check the weather on your phone, you're using an API

- **NameTag:** `curl -F 'data=@sentences.txt' -F 'output=vertical' http://lindat.mff.cuni.cz/services/nametag/api/recognize`
 - `data` (points to `@sentences.txt`)
 - `output format` (points to `output=vertical`)
 - `service` (points to `recognize`)

Non-linguistic content :: Video

Bohumil Veselý's Gallery ([url](#))

Tennis player Karel Koželuh teaching
at his tennis school in Horní Liboc ([url](#))



Non-linguistic content :: Img

Corpus


LINDAT / CLARIAH-CZ


Danish Fungi 2020

(IEEE/CVF / 2022-01-01)

Author(s):
Picek, Lukáš ; et al.

► show everyone



 This item contains 2 files (138.95 GB).

Publicly Available

[url](#)

Two ways how to get into LINDAT repository

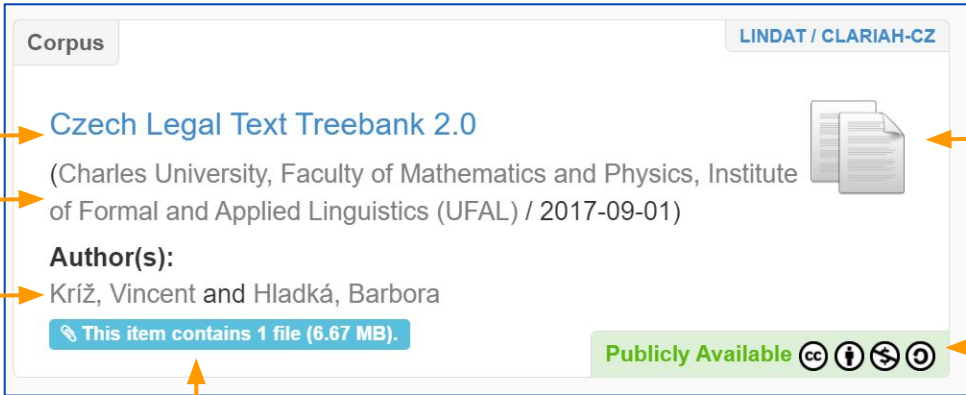
1. <http://lindat.cz>



The screenshot shows the LINDAT website's search interface. At the top, there are navigation links: **Catalogue**, Corpora, Treebanks, ČDK, and Bibliography. Below these is a search bar with the placeholder text "search by type of data" and a red "Search" button. Underneath the search bar, there is a hint: "e.g. [corpus](#) or [lexicon](#) or [editor](#)".

2. <http://lindat.cz/repository>

Repository items :: Basic view



The screenshot shows a repository item view for 'Czech Legal Text Treebank 2.0'. The item is part of the 'Corpus' collection and is associated with 'LINDAT / CLARIAH-CZ'. The title is 'Czech Legal Text Treebank 2.0'. The publisher is '(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2017-09-01)'. The author(s) are 'Kříž, Vincent and Hladká, Barbora'. The item contains 1 file (6.67 MB). The item is publicly available under a Creative Commons license (CC BY-NC-ND).

Annotations with arrows point to the following elements:

- title**: Czech Legal Text Treebank 2.0
- publisher**: (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2017-09-01)
- author(s)**: Kříž, Vincent and Hladká, Barbora
- type**: Document icon
- licenses**: Publicly Available CC BY-NC-ND icons
- how many files to download**: This item contains 1 file (6.67 MB).

Repository items :: Detailed view

Czech Legal Text Treebank 2.0

Please use the following text to cite this item or export to a predefined format: [BIBTEX](#) [CMBX](#)

Kříž, Vincent and Hladká, Barbora, 2017, Czech Legal Text Treebank 2.0, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (UFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-2498>

This resource is also integrated in following services: [PML-TO](#) [KonText](#) [Share](#) [f](#) [t](#)

LINDAT / CLARIAH-CZ

Authors Kříž, Vincent and Hladká, Barbora

Item identifier <http://hdl.handle.net/11234/1-2498>

Project URL <http://ufal.mff.cuni.cz/clit2.0>

Date issued 2017-09-01

Type corpus, text

Size 1121 sentences

Language(s) Czech

Description The Czech Legal Text Treebank 2.0 (CLTT 2.0) annotates the same texts as the CLTT 1.0. These texts come from the legal domain and they are manually syntactically annotated. The CLTT 2.0 annotation on the syntactic layer is more elaborate than in the CLTT 1.0 from various aspects. In addition, new annotation layers were added to the data: (i) the layer of accounting entities, and (ii) the layer of semantic entity relations.

Publisher Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL)

Acknowledgement Univerzita Karlova (mimo GAUK)
Project code: SVV 260 453
Project name: Specifický vysokoškolský výzkum
Ministerstvo Školství, mládeže a tělovýchovy České republiky
Project code: LM2015071
Project name: LINDAT/CLARIN: Institut pro analýzu, zpracování a distribuci lingvistických dat

Subject(s) [treebank](#) [Prague dependencies](#) [named entities](#) [semantic relations](#)

Collection(s) LINDAT / CLARIAH-CZ Data & Tools

Other versions [List all versions](#)

Show full item record

Repository items :: Files to download

Files in this item



Download instructions for command line

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)



Name cltt_2.0.zip
Size 6.67 MB
Format application/zip
Description CLTT 2.0
MDS5 c0aef10f0c49ff159db68e59d37825e4







Download file Preview

File Preview

data		
sentences		
pml		
document_02_004.w		201 kB
document_02_006.a		771 kB
document_01_004.m		479 kB
document_01_003.w		143 kB
document_01_005.a		766 kB
document_01_001.a		522 kB

Repository items :: Data types

 Corpus	 Lexical conceptual	 Language description	 Technology / Tool / Service
<p>i Type of the resource: "Corpus" refers to text, speech and multimodal corpora. "Lexical Conceptual Resource" includes lexica, ontologies, dictionaries, word lists etc. "language Description" covers language models and grammars. "Technology / Tool / Service" is used for tools, systems, system components etc.</p>			


Corpus LINDAT / CLARIAH-CZ

Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2020)

Author(s):
Hajič, Jan ; et al.
▶ show everyone

This item contains 1 file (2.6 GB).

Publicly Available 


LexicalConceptualResource LINDAT / CLARIAH-CZ

EngVallex - English Valency Lexicon 2.0

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2021-02-25)

Author(s):
Cinková, Silvie ; Fučíková, Eva ; Šindlerová, Jana and Hajič, Jan

This item contains 1 file (9.52 MB).

Publicly Available 


LanguageDescription LINDAT / CLARIAH-CZ

NameTag 2 Models (2021-09-16)

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2020-09-16)

Author(s):
Straková, Jana and Straka, Milan

This item contains 7 files (609.97 MB).

Publicly Available 

ToolService LINDAT / CLARIAH-CZ

MorphoDiTa: Morphological Dictionary and Tagger

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2014-02-14)

User activities

- Searching the repository
- Downloading data
- Storing data

Searching the repository :: <http://lindat.cz>

Catalogue Corpora Treebanks ČDK Bibliography

e.g. [corpus](#) or [lexicon](#) or [editor](#)

Search results and faceted search

treebank

Advanced Search

Limit your search

- Author
- Subject
- Rights
- Language (ISO)
- Type
- Contain Files
- Community



Showing 1 through 10 out of 113 results

1 2 3 > 12

Corpus LINDAT / CLARIAH-CZ

Czech Legal Text Treebank 2.0
 (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2017-09-01)



Author(s):
 Križ, Vincent and Hladká, Barbora

Corpus LINDAT / CLARIAH-CZ

Prague Czech-English Dependency Treebank 2.0
 Coref
 (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2016-03-30)

Author(s):
 Nedoluzhko, Anna ; Novák, Michal ; Cinková, Silvie ; Mikulová, Marie and Mirovský, Jiří

Limit your search

Author

Subject

Rights

Language (ISO)

Czech (73)

English (51)

French (36)

German (35)

Italian (34)

Hungarian (33)

Polish (33)

Swedish (33)

Russian (32)

Spanish (32)

Bulgarian (31)

Modern Greek (1453-) (31)

Dutch (30)

Finnish (30)

Basque (29)

Hebrew (29)

Portuguese (29)

Romanian (29)

Slovak (29)

Slovenian (29)

... View More

Type

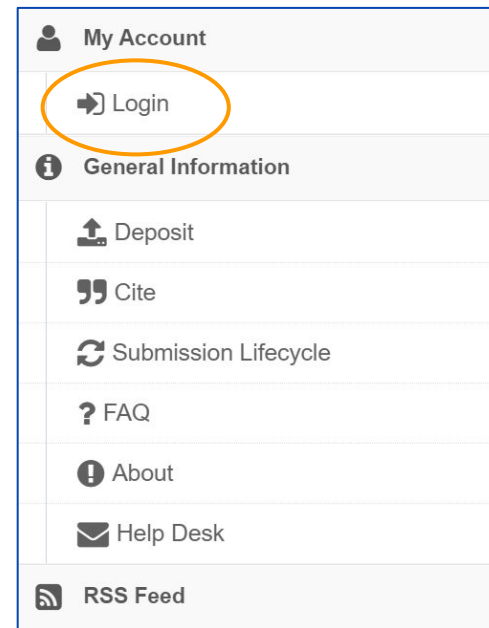
Contain Files

User activities

- Searching the repository
- Downloading data
- Storing data

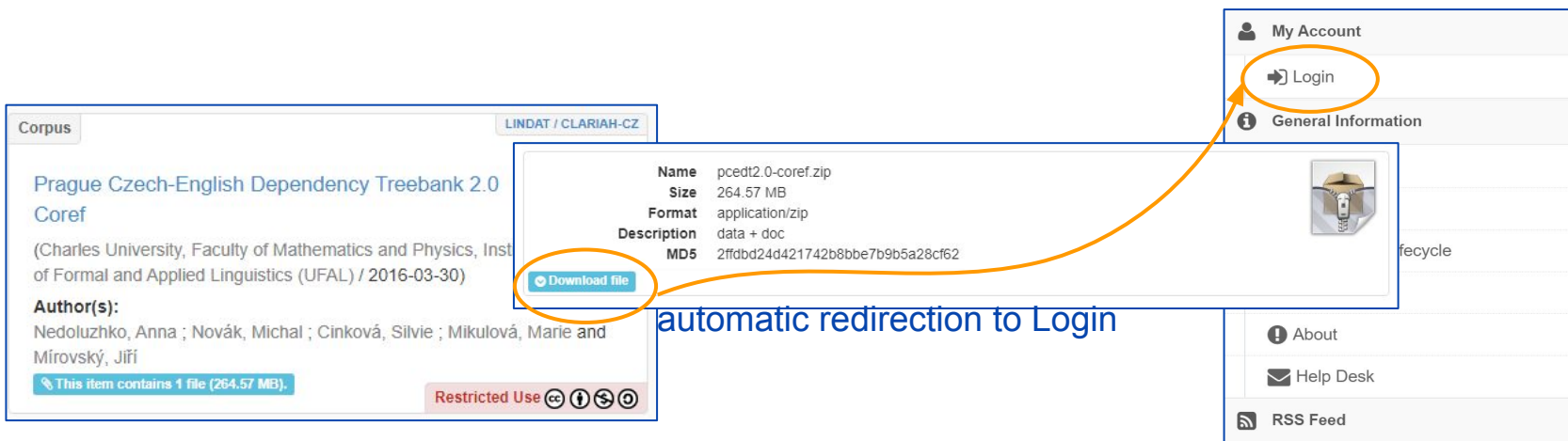
Downloading data

- Login is not required when downloading data the license of which allows free sharing, i.e. Creative Commons licenses and Open Access licenses.



Downloading data

- Login is required when downloading data and tools with the required which require a license agreement to be signed.



Corpus LINDAT / CLARIAH-CZ

Prague Czech-English Dependency Treebank 2.0
Coref

(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2016-03-30)

Author(s):
Nedoluzhko, Anna ; Novák, Michal ; Cinková, Silvie ; Mikulová, Marie and Mírovský, Jiří

This item contains 1 file (264.57 MB).

Restricted Use

Name	pcedt2.0-coref.zip
Size	264.57 MB
Format	application/zip
Description	data + doc
MD5	2ffdbd24d421742b8bbe7b9b5a28cf62

Download file

My Account

Login

General Information

About

Help Desk

RSS Feed

automatic redirection to Login

Login

- Find an academic institution where you have an account
- If it does not exist (= you're a “homeless” researcher), create a CLARIN account at [url](#)

Sign in to **LINDAT/CLARIAH-CZ Repository**

Login via Your home institution (e.g. university)

- Univerzita Karlova v Praze
Czech Republic 251 km
- FH Burgenland
Austria
- University of Kentucky
- Brookhaven National Laboratory - SDCC.BNL.GOV
- Iowa State University
- Mengo Hospital
- Okta, Inc
- Pädagogische Hochschule Oberösterreich
Austria
- Institut Mines Telecom Business School and Telecom SudParis (v4)
- Northern State University
- Energimuseet
- Lauder Business School

or search for a provider, such as Example University

User activities

- Searching the repository
- Downloading data
- Storing data

Storing data :: Upload and don't worry

= uploading datasets = creating new items in the repository

Catalogue

Who can download?
Who can deposit?
Publication schedule

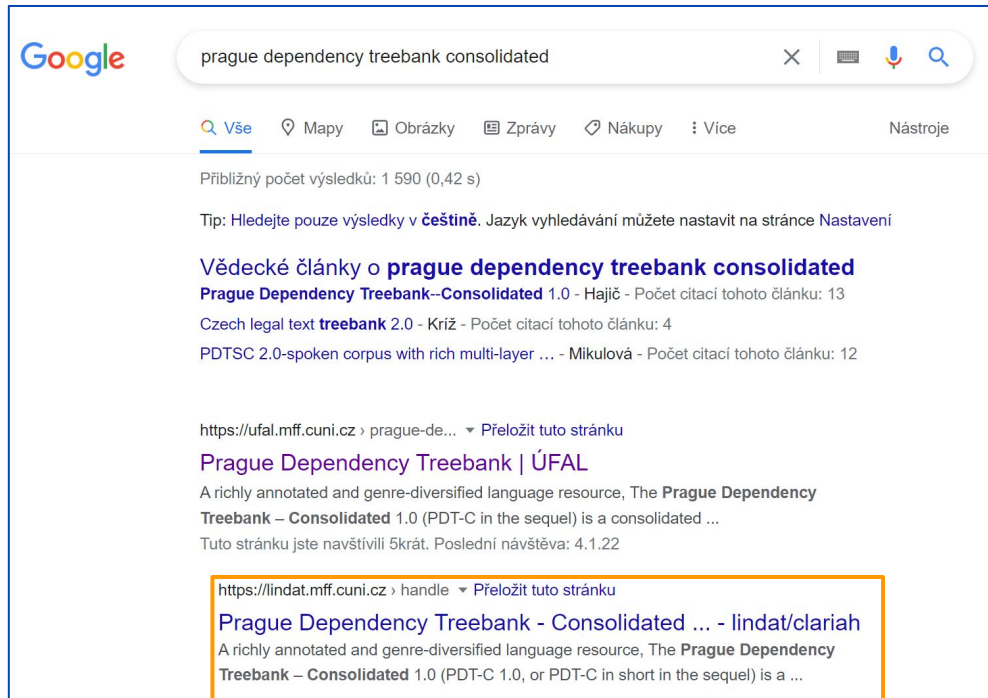
LINDAT/CLARIAH-CZ is a Czech centre for data providing certified storage and natural language processing services

CLARIN CENTRE B OPEN ACCESS

Cite Deposit

Why to create a new item

Your work is visible



Google

prague dependency treebank consolidated

Vše Mapy Obrázky Zprávy Nákupy Více Nástroje

Přibližný počet výsledků: 1 590 (0,42 s)

Tip: Hledejte pouze výsledky v češtině. Jazyk vyhledávání můžete nastavit na stránce Nastavení

Vědecké články o **prague dependency treebank consolidated**

Prague Dependency Treebank--Consolidated 1.0 - Hajič - Počet citací tohoto článku: 13

Czech legal text **treebank** 2.0 - Križ - Počet citací tohoto článku: 4

PDTSC 2.0-spoken corpus with rich multi-layer ... - Mikulová - Počet citací tohoto článku: 12

<https://ufal.mff.cuni.cz> > prague-de... > Přeložit tuto stránku

Prague Dependency Treebank | ÚFAL

A richly annotated and genre-diversified language resource, The **Prague Dependency Treebank – Consolidated** 1.0 (PDT-C in the sequel) is a consolidated ...

Tuto stránku jste navštívili 5krát. Poslední návštěva: 4.1.22

<https://lindat.mff.cuni.cz> > handle > Přeložit tuto stránku

Prague Dependency Treebank - Consolidated ... - lindat/clariah

A richly annotated and genre-diversified language resource, The **Prague Dependency Treebank – Consolidated** 1.0 (PDT-C 1.0, or PDT-C in short in the sequel) is a ...

Why to create a new item

Your work is visible :: Direct data citation, integration with Google scholar

enTenTen 📄

“ Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

Masaryk University, NLP Centre, 2011, *enTenTen*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8>. 📄

bibtex

```
@misc{11858/00-097C-0000-0001-CCDF-8,
title = {{enTenTen}},
url = {http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8},
note = {{LINDAT}/{{CLARIAH}}-{{CZ}} digital library at the Institute of Formal and Applied Linguistics
({{\U}}FAL)}, Faculty of Mathematics and Physics, Charles University},
copyright = {{NLP} Centre Web Corpus License},
year = {2011} }
```

📄

Why to create a new item

Your work is permanently visible

- Persistent identifier PID (= stable reference)
- PID makes a URL that always leads to the metadata
 - it always works, even when moving data
 - included in the citation

Czech Legal Text Treebank 2.0 🔍









“ Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

Križ, Vincent and Hladká, Barbora, 2017, *Czech Legal Text Treebank 2.0*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-2498>. 📄

🧩 This resource is also integrated in following services: Share: [f](#) [t](#)

PML-TQ
KonText

LINDAT / CLARIAH-CZ

 Authors	Križ, Vincent and Hladká, Barbora	
 Item identifier	http://hdl.handle.net/11234/1-2498	
 Project URL	http://ufal.mff.cuni.cz/cltt2.0	
 Date issued	2017-09-01	
 Type	corpus, text	
 Size	1121 sentences	
 Language(s)	Czech	

Why to create a new item

Your work can make other people happy

- Licencing
- LINDAT team supports open data sharing while respecting the licenses of your data
 - Open Access in 5 minutes ([url](#))
 - free and open online access to academic information, such as publications and data

Why to create a new item

Your work is safe

- on/off-site backup

Storing data

= uploading datasets = creating new items in the repository, [help](#)

1. Login
2. Submissions, Start
3. Select the data type
4. Fill in the metadata
= describe a new item
5. Upload the files
6. Select a license
7. Fill in a note for the editors
8. Check the added information
9. Submit the item to the editors for review

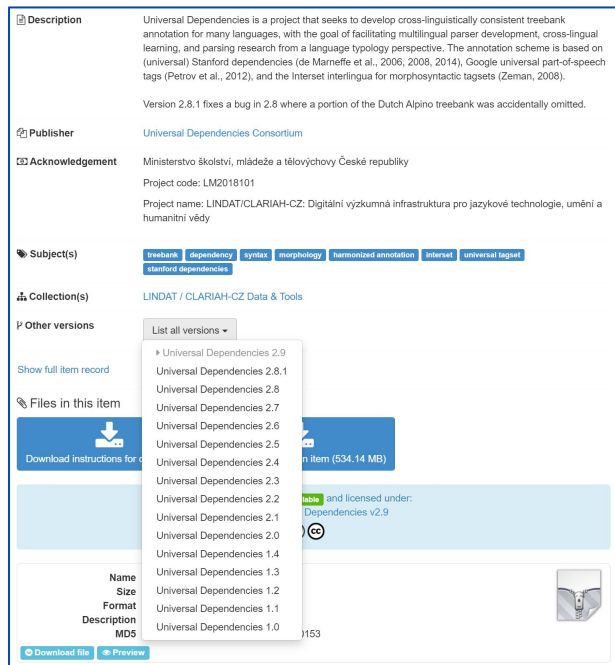
Submissions

Start a new submission

The submission process includes describing the item and uploading the file(s) comprising it. Each community or collection may set its own submission policy.

Storing data :: Versions

- Prefer latest, preserve all
- Changes → new version, new PID, new citation



Description

Universal Dependencies is a project that seeks to develop cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. The annotation scheme is based on (universal) Stanford dependencies (de Marneffe et al., 2006, 2008, 2014), Google universal part-of-speech tags (Petrov et al., 2012), and the Intersect interlingua for morphosyntactic tagsets (Zeman, 2008).

Version 2.8.1 fixes a bug in 2.8 where a portion of the Dutch Alpino treebank was accidentally omitted.

Publisher [Universal Dependencies Consortium](#)

Acknowledgement Ministerstvo školství, mládeže a tělovýchovy České republiky
Project code: LM2018101
Project name: LINDAT/CLARIAH-CZ: Digitální výzkumná infrastruktura pro jazykové technologie, umění a humanitní vědy

Subject(s) [treebank](#) [dependency](#) [syntax](#) [morphology](#) [harmonized annotation](#) [intersect](#) [universal tagset](#)
[standard dependencies](#)

Collection(s) LINDAT / CLARIAH-CZ Data & Tools

Other versions

List all versions ▾

- ▶ Universal Dependencies 2.9
- ▶ Universal Dependencies 2.8.1
- ▶ Universal Dependencies 2.8
- ▶ Universal Dependencies 2.7
- ▶ Universal Dependencies 2.6
- ▶ Universal Dependencies 2.5
- ▶ Universal Dependencies 2.4
- ▶ Universal Dependencies 2.3
- ▶ Universal Dependencies 2.2
- ▶ Universal Dependencies 2.1
- ▶ Universal Dependencies 2.0
- ▶ Universal Dependencies 1.4
- ▶ Universal Dependencies 1.3
- ▶ Universal Dependencies 1.2
- ▶ Universal Dependencies 1.1
- ▶ Universal Dependencies 1.0

[Download instructions for](#)

[Download file](#) [Preview](#)

Name Universal Dependencies 2.9
Size 534.14 MB
Format MDS
Description

Storing data :: Connection to services

- Greater user comfort

Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)

Please use the following text to cite this item or export to a predefined format:

BIBTEX CMDI

Hajič, Jan; et al., 2020, *Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-3185>.

This resource is also integrated in following services:

TEITOK

Share:

TEITOK

Login
Available Corpora
TEITOK@LINDAT

Powered by <TEI+TOK>
Maarten Janssen, 2014

Corpora

Below is the list of corpora in the TEITOK/Kontext hybrid set-up, hosted at the [UFAL](http://ufal.institute) institute. To get a the [repository](#). For corpora that have previous in TEITOK, you can click on the version number to see a

	Acronym	Latest	Token size	Corpus Type	Corpus Status	Corpus Language(s)
info	PDT-C	1.0	4M	Treebank	stable	Czech
info	ParCzech	3.0	27M	Parliamentary corpus	stable	Czech
info	Skript 2015		400k	Learner Corpus	live	Czech
info	Universal Dependencies	2.7	26M	Treebank	stable	Many

<http://lindat.mff.cuni.cz/services/teitok/pdct10/index.php>

Sum up :: FAIR data principles https://en.wikipedia.org/wiki/FAIR_data

- Findable
The first step in (re)using data is to find them, incl. metadata
- Accessible
Once we find the data we need to know how to access them
- Interoperable
The data can be exchanged and used across different applications and systems
- Reusable
The data are well documented and curated

LINDAT Helpdesk

- Frequently Asked Questions <https://lindat.cz/faq-repository>



Harvesting metadata

- extract metadata from various sources
- [OAI-PMH](#) = a protocol for metadata harvesting that specifies how repositories can expose their metadata for other to harvest

CLARIN Virtual Language Observatory

- [CLARIN](#) is a digital infrastructure offering data, tools and services to support research based on language resources. It operates through the national centers all over Europe (LINDAT is one of them)
- Virtual Language Observatory (VLO) <https://vlo.clarin.eu>
 - online overview of the data that is available at a variety of computing centres

Virtual Language Observatory Search Contributors Help CLARIN

VLO / Faceted search / Search results

LINDAT

Showing 1 to 10 of 526 results for LINDAT Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

- Language
- Collection
- Resource type
- Format

Temporal Coverage

Availability

Search options

<< < 1 2 3 4 5 6 7 8 9 10 > >>

LINDAT Translation service 1 [h]
 (Part of LINDAT / CLARIAH-CZ Data & Tools)
 Source code of the LINDAT Translation service frontend. The service provides a UI and a simple rest api that accesses machine translation models served by tensorflow serving. The most recent version of the code is available at https://github.com/ufal/lindat_translation.
 Landing page for this record

KonText Web Demo 1 [h] 1 [h]
 (Part of LINDAT / CLARIAH-CZ Data & Tools)
 An interactive web demo for querying selected ÚFAL and LINDAT corpora. LINDAT/CLARIN KonText is a fork of ÚČNK KonText (<https://github.com/czcorpus/kontext>, maintained by Tomáš Machálek) that contains some modifications and additional features. Kontext, in turn, is a fork of the Bonito 2.68 python web interface to the ...
 Czech English
 Landing page for this record

Korektor 1 [h] 3 [h]
 (Part of LINDAT / CLARIAH-CZ Data & Tools)
 Statistical spell- and (occasional) grammar-checker. There are three versions: a unix command line utility and an OS X SpellServer with a System Service, that integrates with native OS X GUI applications, and a web service run by Lindat-Clarín, that can be used either through a web form in a browser, or by web applicat...
 Czech
 Landing page for this record