LANGUAGE RESEARCH INFRASTRUCTURE IN CZECHIA

LINDAT/CLARIN

PAVEL STRAŇÁK
LINDAT/CLARIN
LIN DAT/CLARIN
LINguistic
LINDAT/CLARIN
LINGuistic
DATa ... very broadly
LINDAT/CLARIN
LINGuistic DATa
/
CLARIN
LINDAT/CLARIN
LINGuistic
DATa
/
Common
LAnGuage
Research and technology
INfrastructure
LINDAT/CLARIN

- Czech national project; node of CLARIN ERIC
- Operational since 2014
- Users:
  - Researchers in SSH and Computational Linguistics
- Technology:
  - Repository (resources), Services, Applications
- Knowledge, Support and Training
LANGUAGE TECHNOLOGY

• Natural Language Processing - NLP
  • Analysis, synthesis of spoken and written language
  • Machine Translation, Information Extraction, ...
  • Search in texts, audio, video, images

• State-of-the-art technology in NLP
  • “Statistical” methods:
    • Machine learning incl. neural networks
    • Need for (large) Language Resources – Texts, multimodal
      • Repositories, identification, replication of experiments, standards
USERS

• Everyone
  • communicates in and works with natural language!

• ... immediate users of the infrastructure:
  • Language Technology researchers
    • Universities, Research organisations
    • Need lots of data, easy to get, clean open licensing

• “Content” users:
  • Linguists, historians, teachers, psychologists, sociologists, ...
    • Need identifiable data, preprocessed, searchable, easy-to-use
      services and applications

• General public:
  • proper language use – IRLG; holocaust documentation: CVH Malach
Data Repository

PRESCRIBE AND FIND LANGUAGE DATA AND NLP TOOLS

Author
- Hajič, Jan (47)
- Žabokrtský, Zdeněk (32)
- Straka, Milan (29)
- Zeman, Daniel (29)
- Bojar, Ondřej (28)

Subject
- Germanistik (47)
- machine translation (39)
- corpus (34)
- treebank (30)
- morphology (26)

Language (ISO)
- English (222)
- Czech (192)
- German (159)
- Dutch (92)
- Spanish (83)

What's New

CorpusExplorer

Author(s): Rüdiger, Jan Oliver

Description:
Software for corpus-freelists and text/data mining enthusiasts. The Corpus Explorer combines over 45 interactive tools for various language data processing tasks.
DATA REPOSITORY

- ~ 500 registered users
- submitters & users signing licenses (not everything can be Open Access)
- 200+ Data Records
- > 1000 Metadata Records
- 80 languages
- 100 TB+ Data in Repository (+ 1PB of UCS Shoah Foundation Archive)
DATA REPOSITORY

- ~ 500 registered users
- submitters & users signing licenses (not everything can be Open Access)
- 200+ Data Records
- > 1000 Metadata Records
- 80 languages
- 100 TB+ Data in Repository (+ 1PB of UCS Shoah Foundation Archive)
DATA REPOSITORY

- Safe preservation (upload and don’t worry)
- Discovery & Reuse
- Direct data citation (works in Google Scholar)
- Licensing (Open Access, but also more options)
- Versioning
- Language data and tools
- Worldwide (for everyone), easy to use
How to Deposit

Only authenticated users can deposit items. If you cannot find your home organisation in the Login dialog list of organisations then register at clarin.eu and authenticate using "clarin.eu website account". In case you cannot use any authentication method above or if you encounter a problem, do not hesitate to contact our Help Desk and we can create a local account for you.

Step 1: Login

To start a new submission you have to login first. Click Login under My Account in the right menu panel.

Step 2: Starting a new submission

Now you have a new menu item 'Submissions' under My Account. Click on Submissions to go to the Submissions screen.
Vědecké články o Prague dependency treebank 3.0

Prague dependency treebank 3.0 - Bejček - Počet citací tohoto článku: 47
Prague Dependency Treebank - Hajič - Počet citací tohoto článku: 385
The Prague dependency treebank - Böhmová - Počet citací tohoto článku: 423

Prague Dependency Treebank 3.0 | ÚFAL
https://ufal.mff.cuni.cz/pdt3.0 ▼ Přeložit tuto stránku
Introduction. The Prague Dependency Treebank 3.0 (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDIT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed and improved in various aspects. Moreover ...

The Prague Dependency Treebank 2.0.
https://ufal.mff.cuni.cz/pdt2.0/ ▼ Přeložit tuto stránku
The Prague Dependency Treebank 2.0 (PDT 2.0) contains a large amount of Czech texts with complex and interlinked morphological (2 million words), syntactic (1.5 MW) and complex semantic annotation ... Please note that new versions of this corpus have been published: PDT 3.0 (2013), PDIT 1.0 (2012), PDT 2.5 (2012).

Prague Dependency Treebank 3.0 (PDT 3.0)
https://lindat.mff.cuni.cz/repository/xmlui/bitstream/.../PDT30_index_lindat.html...?
Prague Dependency Treebank 3.0 (PDT 3.0). Overview. The Prague Dependency Treebank 3.0 (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDIT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed ...
Please use the following text to cite this item or export to a predefined format:

Masaryk University, NLP Centre, 2011, enTenTen, LINDAT/CLARIN 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.

This resource is also integrated in following services:

- KonText

Item identifier: http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8

Date issued: 2011-12-16

Type: corpus

Language(s): English

Description: Very large English web corpus enTenTen, comprising 3,268,798,527 tokens.

Publisher: Masaryk University, NLP Centre

Acknowledgement: Lexical Computing Ltd.

Subject(s): English large corpus

Collection(s): LINDAT / CLARIN Data & Tools
Choose a License

Choose a license from the options below:

- Start again

What do you want to deposit?
- Software
- Data

Search for a license...

Public Domain Mark (PD)
The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.

Publicly Available (PD)

Public Domain Dedication (CC Zero)
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available (CC 0)
LICENSING

AS OPEN AS POSSIBLE (NOT MORE)

Publisher  Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague

Acknowledgement  Ministerstvo školství, mládeže a tělovýchovy

Project code: LM2011023

Project name: Český národní korpus

Subject(s)  representative corpus, written language

Collection(s)  LINDAT / CLARIN Data & Tools

Show full item record

Files in this item

Download instructions for command line

This item is Academic Use and licensed under:
Czech National Corpus (Shuffled Corpus Data)

Name  syn2015.gz
Size  1.48 GB
Format  application/x-gzip
Description  corpus
MD5  e0242cc77e999794af6cfaf57f843c12

Download file
PREFER LATEST, PRESERVE ALL

Project name: Internet jako jazykový korpus
Ministerstvo školství, mládeže a tělovýchovy České republiky
Project code: LN00A063

Project name: Centrum komputační lingvistiky
Ministerstvo školství, mládeže a tělovýchovy České republiky
Project code: MSM 0021620838

Project name: Moderní metody, struktury a systémy informatiky

Subject(s)
- MorphoDiTa
- Czech
- morphological analysis
- morphological generation
- PoS tagging

Collection(s)
- LINDAT / CLARIN Data & Tools

This item is replaced by a newer submission:
http://hdl.handle.net/11234/1-1836
Please refer to the submission above for the latest available data. If you nevertheless need the original data, please click here.
PREFER LATEST, PRESERVE ALL

Collection(s)  LINDAT / CLARIN Data & Tools

Other versions
- List all versions
  - Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphiDiTa 161115
  - Czech Models (MorfFlex CZ 160310 + PDT 3.0) for MorphiDiTa 160310
  - Czech Models (MorfFlex CZ + PDT) for MorphiDiTa

Show full item record

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  czech-morfflex-pdt-161115.zip
Size  69.18 MB
Format  application/zip
Description  Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphiDiTa 161115
MD5  adde38cd363219759e19165b06bba4ce
REPOSITORY SOFTWARE

- CLARIN-DSpace

- DSpace + licensing, versioning and more

- LINDAT’s project converting to community

- 13 deployments 10 countries
SERVICES

• Web services
  • REST, stable services only
    • open source software, high-standards requirements

• Applications and user services
  • Wrappers around services (web-based applications)
  • Internet Language Reference Book (at Institute for Czech Language)

• “Physical Service”: Centre for Visual History Malach
  • Shoah collection access point
  • Reusable audio search technology
Web Services and Applications

20 SERVICE TYPES (WITH 100+ CORPORA/LEXICONS)

NameTag

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

Straka, Milan and Straková, Jana, 2014, NameTag, LINDAT/CLARIN 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-0000-0023-43CE-E.

Authors:
Milan Straka, Jana Straková

Description:
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance (Straková et. al. 2013). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.
Please use the following text to cite this item or export to a predefined format:

Straka, Milan and Straková, Jana, 2014, NameTag, LINDAT/CLARIN 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-0000-0023-43CE-E.

Authors:
Milan Straka, Jana Straková

Description:
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance (Straková et. al. 2013). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.
Please use the following text to cite this item or export to a predefined format:

Straka, Milan and Straková, Jana, 2014, *NameTag*, LINDAT/CLARIN 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-0000-0023-43CE-E](http://hdl.handle.net/11858/00-097C-0000-0023-43CE-E).

**Authors:**
Milan Straka, Jana Straková

**Description:**
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance (*Straková et. al. 2013*). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.
Please use the following text to cite this item or export to a predefined format:

Straka, Milan and Straková, Jana, 2014, NameTag, LINDAT/CLARIN 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-0000-0023-43CE-E.

Authors:
Milan Straka, Jana Straková

Description:
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance (Straková et. al. 2013). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.

The Internet Language Reference Book

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

Šmerk, Pavel; Pravdová, Markéta; Beneš, Martin; et al., 2009, The Internet Language Reference Book, LINDAT/CLARIN 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-0000-0023-8BD2-2.
Straka, Milan and Straková, Jana, 2014, *NameTag*, LINDAT/CLARIN 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-0000-0023-43CE-E](http://hdl.handle.net/11858/00-097C-0000-0023-43CE-E).

**Authors:**
Milan Straka, Jana Straková

**Description:**
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance ([Straková et. al. 2013](https://example.com)). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.

**The Internet Language Reference Book**

Šmerk, Pavel; Pravdová, Markéta; Beneš, Martin; et al., 2009, *The Internet Language Reference Book*, LINDAT/CLARIN 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-0000-0023-8BD2-2](http://hdl.handle.net/11858/00-097C-0000-0023-8BD2-2).
Please use the following text to cite this item or export to a predefined format:

Straka, Milan and Straková, Jana, 2014, *NameTag*, LINDAT/CLARIN 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-0000-0023-43CE-E](http://hdl.handle.net/11858/00-097C-0000-0023-43CE-E).

**Authors:**
Milan Straka, Jana Straková

**Description:**
*NameTag* is an open-source tool for named entity recognition (NER). *NameTag* identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. *NameTag* is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, *NameTag* achieves state-of-the-art performance ([Straková et. al. 2013](https://example.com)). *NameTag* is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.
NameTag

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

Straka, Milan and Straková, Jana, 2014, NameTag, LINDAT/CLARIN 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-0000-0023-43CE-E.

Authors:
Milan Straka, Jana Straková

Description:
NameTag is an open-source tool for named entity recognition (NER). NameTag identifies proper names in text and classifies them into predefined categories, such as names of persons, locations, organizations, etc. NameTag is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, NameTag achieves state-of-the-art performance (Straková et. al. 2013). NameTag is a free software under LGPL license and the linguistic models are free for non-commercial use and distributed under CC BY-NC-SA license, although for some models the original data used to create the model may impose additional licensing conditions.

The Internet Language Reference Book

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

Šmerk, Pavel; Pravdová, Markéta; Beneš, Martin; et al., 2009, The Internet Language Reference Book, LINDAT/CLARIN 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-0000-0023-8BD2-2.
UDPipe

UDPipe is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files. UDPipe is language-agnostic and can be trained given an input in CoNLL-U format. Trained models are provided for nearly all UD treebanks. UDPipe is available as a binary for Linux/Windows/OS X, as a library for C++, Python, Perl, Java, and as a web service. Third-party R CRAN package also exists.

UDPipe is a free software distributed under the Mozilla Public License 2.0 and the linguistic models are free for non-commercial use and distributed under the CC BY-NC-SA, although for some models the original data used to create the model may impose additional licensing conditions. UDPipe is versioned using Semantic Versioning.

Copyright 2017 by Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic.

Description of the available methods is available in the API Documentation and the models are described in the UDPipe User's Manual.

**Service**

The service is freely available for testing. Respect the CC BY-NC-SA licence of the models – explicit written permission of the authors is required for any commercial exploitation of the system. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the systems at UFAL. All comments and reactions are welcome.

**Model:**
- UD 2.0 (description)
- CoNLL17 Baseline UD 2.0 (description)
- UD 1.2 (description)

**Actions:**
- Tag and Lemmatize
- Parse

**Advanced Options**
UDPipe is a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files. UDPipe is language-agnostic and can be trained given an input in CoNLL-U format. Trained models are provided for nearly all UD treebanks. UDPipe is available as a binary for Linux/Windows/OS X, as a library for C++, Python, Perl, Java as a web service. Third-party R CRAN package also exists.

UDPipe is a free software distributed under the Mozilla Public License 2.0 and the linguistic models are free for non-commercial use and distributed under the CC BY-NC-SA, although for some models the original data used to create the model may impose additional licensing conditions. UDPipe is versioned using Semantic Versioning.

Copyright 2017 by Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic.

Description of the available methods is available in the API Documentation and the models are described in the UDPipe User's Manual.

The service is freely available for testing. Respect the CC BY-NC-SA licence of the models – explicit written permission of the authors is required for any commercial exploitation of the system. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the systems at UFAL. All comments and reactions are welcome.

**Model:**
- UD 2.0 (description)
- CoNLL17 Baseline UD 2.0 (description)
- UD 1.2 (description)

**Actions:**
- Tag and Lemmatize
- Parse
In the first place I hope you will live twenty-three years longer. Mr. Tom Lefroy's birthday was yesterday, so that you are very near of an age.

After this necessary preamble I shall proceed to inform you that we had an exceeding good ball last night, and that I was very much disappointed at not seeing Charles Fowle of the party, as I had previously heard of his being invited. In addition to our set at the Harwoods' ball, we had the Grants, St. Johns, Lady Rivers, her three daughters and a
In the first place I hope you will live twenty-three years longer. Mr. Tom Lefroy's birthday was yesterday, so that you are very near of an age.

After this necessary preamble I shall proceed to inform you that we had an exceeding good ball last night, and that I was very much disappointed at not seeing Charles Fowle of the party, as I had previously heard of his being invited. In addition to our set at the Harwoods' ball, we had the Grants, St. Johns, Lady Rivers, her three daughters and a
In the first place I hope you will live twenty-three years longer. Mr. Tom Lefroy's birthday was yesterday, so that you are very near of an age.

After this necessary preamble I shall proceed to inform you that we had an exceeding good ball last night, and that I was very much disappointed at not seeing Charles Fowle of the party, as I had previously heard of his being invited. In addition to our set at the Harwoods' ball, we had the Grants, St. Johns, Lady Rivers, her three daughters and a
UDPipe web service is available on [http(s)://lindat.mff.cuni.cz/services/udpipe/api/](http(s)://lindat.mff.cuni.cz/services/udpipe/api/).

The web service is freely available. Respect the CC BY-NC-SA licence of the models – explicit written permission of the authors is required for any commercial exploitation of the system. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the systems at UFAL. All comments and reactions are welcome.

**API Reference**

The UDPipe REST API can be accessed directly or via any other web programming tools that support standard HTTP request methods and JSON for output handling.

<table>
<thead>
<tr>
<th>Service Request</th>
<th>Description</th>
<th>HTTP Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>models</td>
<td>return list of models and supported methods</td>
<td>GET/POST</td>
</tr>
<tr>
<td>process</td>
<td>process supplied data</td>
<td>GET/POST</td>
</tr>
</tbody>
</table>

**Method models**

Return the list of models available in the UDPipe REST API, and for each model enumerate components supported by this models (model can contain a tokenizer, tagger and a parser). The default model (used when user supplies no model to a method call) is also returned – this is guaranteed to be the latest Czech model.

**Browser Example**

```
http://lindat.mff.cuni.cz/services/udpipe/api/models
try this
```

**Example JSON Response**

```
{
    "models": {
        "czech-ud-1.2-160523": ["tokenizer", "tagger", "parser"],
        "english-ud-1.2-160523": ["tokenizer", "tagger", "parser"]
    },
    "default_model": "czech-ud-1.2-160523"
}
```

**Method process**


The response is in JSON format of the following structure:

```json
{
  "model": "Model used",
  "acknowledgements": ["URL with acknowledgements", ...],
  "result": "processed_output"
}
```

The `processed_output` is the output of the UDPipe in the requested output format.

**Browser Examples**


---

## Model Selection

There are several possibilities how to select required model using the `model` option:

- If `model` option is not specified, the default model (returned by `models` method) is used – this is guaranteed to be the latest Czech model.
- The `model` option can specify one of the models returned by the `models` method.
- The `model` option may be only several first words of model name. In this case, the latest most suitable model is used.
- The `model` can be ISO 639-1 or ISO 639-2 code of a language. If available, newest model for the requested language is used.

Note that the last two possibilities allow using `czech`, `cs`, `ces`, `cze`, `english`, `en` or `eng` as models.

---

## Accessing API using Curl

The described API can be comfortably used by `curl`. Several examples follow:

### Passing Input on Command Line (if UTF-8 locale is being used)

```bash
curl --data 'tokenizer=&tagger=&parser=&data=Děti pojedou k babičce. Už se téší.' http://lindat.mff.cuni.cz/services/udpipe/api/process
```

### Using Files as Input (files must be in UTF-8 encoding)

```bash
curl -F data=@input_file.txt -F tokenizer= -F tagger= -F parser= http://lindat.mff.cuni.cz/services/udpipe/api/process
```

### Specifying Model Parameters

```bash
curl -F data=@input_file.txt -F model=english -F tokenizer= -F tagger= -F parser= http://lindat.mff.cuni.cz/services/udpipe/api/process
```

### Converting JSON Result to Plain Text

```bash
curl -F data=@input_file.txt -F model=english -F tokenizer= -F tagger= -F parser= http://lindat.mff.cuni.cz/services/udpipe/api/process | PYTHONIOENCODING=utf-8 python -c "import sys,json; sys.stdout.write(json.load(sys.stdin)['result'])"
```
• Preservation and dissemination of language data

• Creation of linguistic datasets

• Creation of language processing tools

• Language processing services

• Search interfaces for language datasets: corpora, dictionaries, audio-visual data

• Support in utilising all of this for your research


http://hdl.handle.net/
11858/00-097C-0000-0023-8BD2-2
(http://prirucka.ujc.cas.cz/)

http://malach-centrum.cz
Thank you!

http://lindat.cz