



[Home](#)

[Repository](#)

[PML TreeQuery](#)

[Tools & Services](#)

[CLARIN](#)

[Contact](#)

# LINDAT/CLARIN Tools & Services

Michal Josífko

**ÚFAL Seminar**

Sedlec-Prčice, September 15 -16, 2014

## Treex::Web - Highly Modular NLP Framework Online

**Authors:**

Martin Popel, Zdeněk Žabokrtský, Jan Ptáček, Petr Pajas

**Description:**

Treex (formerly TectoMT) is a highly modular NLP software system implemented in Perl programming language under Linux.

It is primarily aimed at Machine Translation, making use of the ideas and technology created during the Prague Dependency Treebank project. At the same time, it is also hoped to significantly facilitate and accelerate development of software solutions of many other NLP tasks, especially due to re-usability of the numerous integrated processing modules (called blocks), which are equipped with uniform object-oriented interfaces.

**URI to cite/link:**

<http://hdl.handle.net/11858/00-097C-0000-0023-44AF-C>

[PROJECT HOME](#)



Treex::Web

[Česílko](#)

[PML-Tree Query](#)

[PDT-Vallex](#)

[EngVallex](#)

[Morph](#)

[MorphoDITA](#)

[NameTag](#)

[IRLB](#)

[EllixirFM](#)

[Dialogy.Org](#)

[Korektor](#)

## Cesilko - Machine Translation System for Closely Related Languages

**Authors:**

Jan Hajič, Vladislav Kuboň, Petr Homola

**Description:**

The system Česílko was designed as a tool enabling the fast and efficient translation from one source language into many target languages, which are mutually related. The system receives as its input a high quality human translation of the original into Czech (from any language). It translates the Czech input into a number of languages related to Czech. The system contains at the moment 5 language pairs, 4 of them only as experiments, namely Czech into Polish, Lithuanian, Macedonian and Lower Sorbian. Unfortunately, the system cannot be tested on arbitrary texts for these language pairs due to a small size of all dictionaries. The only working language pair (and at the same time also exploitable outside

# So, what is it?

- Single entry point for all ÚFAL tools & services
- Show-case
  - web presentation, demo
- Commitment
  - availability, REST(like) API, Single Sign On (if needed)

# Is it good for something? **Yleees!**

- **I**nstitute level presentation
- **E**asy to understand
- **E**asy to orientate oneself in
- **E**asy to find ([http://ufal.ms.mff.cuni.cz/~josifko/best\\_project\\_you\\_will\\_never\\_ever\\_find/version\\_6/?info](http://ufal.ms.mff.cuni.cz/~josifko/best_project_you_will_never_ever_find/version_6/?info))
- **S**imple to share

# MorphoDiTa

[Basic Info](#)[Demo](#)[REST API Documentation](#)

MorphoDiTa: Morphological Dictionary and Tagger is an open-source tool for morphological analysis of natural language texts. It performs morphological analysis, morphological generation, tagging and tokenization and is distributed as a standalone tool or a library, along with trained linguistic models. In the Czech language, MorphoDiTa achieves state-of-the-art results with a throughput around 10-200K words per second. MorphoDiTa 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.

Copyright 2014 by Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University In Prague, Czech Republic.

**Authors**

Milan Straka, Jana Straková

**Homepage**<http://ufal.mff.cuni.cz/morphodita>**Development repository**<http://github.com/ufal/morphodita>**Status**

version 1.1.0

**OS**

Linux, Windows

**License of the library**[LGPL](#)**License of the models**[CC BY-NC-SA](#)**Contact**[straka@ufal.mff.cuni.cz](mailto:straka@ufal.mff.cuni.cz) CLARIN CENTRE B**Partners, Coordination, Funding**

Dept. of Cybernetics, Univ. of West Bohemia  
Institute of Formal and Applied Linguistics (Prague)  
Institute of Czech Language (Prague)  
NLP Centre, Masaryk University (Brno)  
Ministry of Education, Sports and Youth of the Czech Republic

**Repository**

[Main page](#)  
[Contact](#)

**More**

[Clarín](#)  
[META-Net](#)

 LINDAT CLARIN MŠMT  
MINISTRY OF EDUCATION,  
YOUTH AND SPORTS

THE LINDAT/CLARIN PROJECT (LM2010013) IS FULLY SUPPORTED BY THE MINISTRY OF EDUCATION, SPORTS AND YOUTH OF THE CZECH REPUBLIC UNDER THE PROGRAMME LM OF "LARGE INFRASTRUCTURES".

Přibližný počet výsledků: 8 860 000 (0,26 s)

## Zobrazeny výsledky pro dotaz **fal services**

Místo toho hledat **ufal services**

### FAL | Services

[www.falservices.cz/](http://www.falservices.cz/) ▾

Services. ... Vítejte na stránce **FAL Services!** Projekty ... Spolupráce.  
Kariéra · Kontakt · Facebook. © 2013 All Rights Reserved - **FAL Servi**

#### Kontakt

Jméno a příjmení: (povinné). Email: (povinné). Předmět: Vaše zpráva:.

#### O nás

Filozofii firmy FAL S  
poskytovat služby z

#### Kariéra

K úspěchu a rozvoji firmy přispívá především kvalitní tým. Proto ...

[Další výsledky z webu falservices.cz »](#)

### FAL Services, s.r.o. (Horní Suchá) • Firmy.cz

[www.firmy.cz/detail/2572931-fal-services-horni-sucha.html](http://www.firmy.cz/detail/2572931-fal-services-horni-sucha.html) ▾

Aktuálně ověřené informace, kontakty a hodnocení **FAL Services, s.r.o.**  
Zabýváme se tvorbou internetových obchodů, základních, středních a

### FAL Services s.r.o. - Obchodní rejstřík firem - Kurz

[rejstrik-firem.kurz.cz/28566084/fal-services-sro/](http://rejstrik-firem.kurz.cz/28566084/fal-services-sro/) ▾

**FAL Services** s.r.o., Na Františku, Horní Suchá, výpis z obchodního r

Přibližný počet výsledků: 511 000 000 (0,23 s)

Tip: Hledejte pouze výsledky v **češtině**. Jazyk vyhledávání můžete nastavit na stránce [Nastavení](#)

### PDT-Vallex - LINDAT-Clarín - Univerzita Karlova

[lindat.mff.cuni.cz/services/PDT-Vallex/](http://lindat.mff.cuni.cz/services/PDT-Vallex/) ▾ [Přeložit tuto stránku](#)

Please see also the PDT-Vallex main research page for additional resources. These webpages and its contents are licensed to you under the CC-BY-NC-ND ...

### LINDAT/CLARIN repository and services

<https://lindat.mff.cuni.cz/privacypolicy.html> ▾ [Přeložit tuto stránku](#)

Repository and **services** of LINDAT/CLARIN project at Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in ...

### Cesilko - LINDAT-Clarín - Univerzita Karlova

[lindat.mff.cuni.cz/services/cesilko/](http://lindat.mff.cuni.cz/services/cesilko/) ▾ [Přeložit tuto stránku](#)

Homepage, <http://lindat.mff.cuni.cz/services/cesilko/>. Tutorials, Demo. Status, version 1.0; released. OS, Linux. License, Attribution-NonCommercial-NoDerivs 3.0 ...

### REST API Documentation - LINDAT-Clarín

[lindat.mff.cuni.cz/services/cesilko/api-docs.php](http://lindat.mff.cuni.cz/services/cesilko/api-docs.php) ▾ [Přeložit tuto stránku](#)

The translate is the main operation/**service** offered by the REST API. This **service** translates the given text in Czech language into Slovak. The possible ...

### LINDAT/CLARIN Research Infrastructure

[www.lindat.cz/](http://www.lindat.cz/) ▾ [Přeložit tuto stránku](#)

Search for language resources in the LINDAT/CLARIN repository. Search. **SERVICES**. LINDAT/CLARIN **services**. Take a look at the available tools and **services**.

### Automatic Docker Service Announcement with Registrar ...

[progrium.com/.../automatic-docker-service-announc...](http://progrium.com/.../automatic-docker-service-announc...) ▾ [Přeložit tuto stránku](#)

před 2 dny - Progrium // Jeff Lindsav ... **Service** discovery requires your **services** to

# Current status

- 12 on-line services
- Treex::Web, Česílko, PML-Tree Query, PDT-Vallex, EngVallex, Morph, MorphoDiTa, NameTag, IRLB  
ElixirFM, Dialogy.Org, Korektor
- Česílko integrated into Weblicht SOA environment
- New services are welcome!

# **We want you! Our today's offer:**

- Environment for web presentations or proxying (quest)
- Infrastructure for running services
- Authentication package for Single Sign On (AAI)
- “Corporate design” package (headers, footers, CSS)
- Know-how & help with AAI, REST API or anything else







# AAI


- Authentication and Authorization Infrastructure
- Based on Shibboleth (SAML2)
- GUI: open-source project Disco Juice
- Stable and user-friendly interface
- Easily incorporable into existing projects
- Users from many (not only) European institutions

## Sign in to LINDAT/CLARIN Repository



Select your Identity Provider (e.g., Home Institution)

-  Univerzita Karlova v Praze  
 Czech Republic 4 km
- CESNET, a. l. e.   
 Czech Republic Nearby
- The National Library of the Czech Republic  Národní knihovna České republiky  
 Czech Republic Nearby
- National Library of Technology   
 Czech Republic Nearby
- Institute of Organic Chemistry and Biochemistry AS CR   
 Czech Republic Nearby
- Czech Technical University in Prague, Faculty of Electrical Engineering   
 Czech Republic Nearby
- Academy of Sciences Library   
 Czech Republic Nearby
- Czech University of Life Sciences Prague   
 Czech Republic Nearby

 or search for a provider, such as Example University

► Please help, I cannot find my provider



Locate me and show nearby providers

Show providers in  show all countries



# REST (like) API

- Usually very easy to implement
- HTTP GET/POST/PUT/DELETE in, JSON/XML out
- Provides great added value for users (no need to install and set up tools to accomplish required tasks)
- Usually very easy to use
- Only needs proper documentation

# Treex::Web API Documentation

## /auth

[Show/Hide](#) | [List Operations](#) | [Expand Operations](#) | [Raw](#)

GET	/auth	Check for whether the session exists
POST	/auth	Provides user login
DELETE	/auth	Provides user logout

## /scenarios

[Show/Hide](#) | [List Operations](#) | [Expand Operations](#) | [Raw](#)

GET	/scenarios	Get list of all scenarios
POST	/scenarios	Create new scenario for user
GET	/scenarios/languages	Get list of all languages for which there are scenarios available
GET	/scenarios/{scenarioId}	Get scenario by Id
PUT	/scenarios/{scenarioId}	Update scenario info
DELETE	/scenarios/{scenarioId}	Deletes scenario

## /treex

[Show/Hide](#) | [List Operations](#) | [Expand Operations](#) | [Raw](#)

GET	/treex/languages	Returns all available language options
-----	------------------	--

## /input

[Show/Hide](#) | [List Operations](#) | [Expand Operations](#) | [Raw](#)

POST	/input/url	Downloads html content from given url
GET	/input/samples	Returns list of all available sample files

# Latest service LINDAT KonText

- Customized installation of ÚČNK corpus manager
- Single GUI for available corpora (including parallel and speech corpora)
- Searchable using Federated Content Search
- No dependency trees, only CQL queries
- Your corpora are welcome!



User: Michal Jozifko (logout)



Query Subcorpora Save Concordance Filter Frequency Collocations View options Help Custom...

corpus HindEnCorp 0.5 - English- POS tags only 3,739,891 positions

Hits: 409 | I.p.m. : 109.36 (related to the whole HindEnCorp 0.5 - English- POS tags only) | ARF : 221 | Result is shuffled

1 / 11

HindEnCorp 0.5 - English- POS tags only		HindEnCorp 0.5 - Hindi - POS tags only	
launchpad	Possible values are : ' never ' - do not allow move with drag & drop of folders in folder tree , ' always ' - allow move with drag & drop of folders in folder tree without asking , or ' ask ' - ( or any other value ) will ask user .	launchpad	संभावित मान हैं : ' कभी नहीं ' खींच & drop के साथ खिसकाने नहीं दें ; फ़ोल्डर पेड़ में फ़ोल्डर को छोड़ें , ' हमेशा ' खींच & drop के साथ खिसकाएँ ; बिना फ़ोर्डे फ़ोल्डर पेड़ में फ़ोल्डर छोड़ दें , या ' पूछें ' ( या कोई अन्य मान ) उपयोगकर्ता से पूछने के लिए .
tides	And although the tree has no religious significance-villagers even use the peel of the trunk to treat skin ailments-It does have great sentimental value .	tides	और इस पेड़े का कोई धार्मिक महत्व भले न हो-ग्रामीण लग इसके तने की छाल का प्रयोग लम्बे लोगों के उपचार में करते हैं-लेकिन इसका खासा भावनात्मक महत्व है .
Indic2012	There are numerous beautiful trees here , hence this place has been named " Sunderban " ( Beautiful forest )	Indic2012	यहाँ बड़ी तादाद में सुंदरी पेड़ मिलते हैं जिनके नाम पर ही इन वनों का नाम सुंदरवन पड़ा है ।
Intercorp	He was at home with the silence of the desert , and he was content just to look at the trees .	Intercorp	वह रेगिस्तान की खामोशी में खो गया था और पेड़ों की कतार को देखकर ही संतुष्ट था ।
tides	Figures such as elephants , swans , bulls , cows , the papiha bird , parrots the mynah bird , chakor-chakon , peacocks and peahens , trees and shrubs like the Banana tree , the Jamxtn tree , the Bunyan tree , the Saru tree , the Sambal tree , clouds , lightning , and clay pots are common to most of these paintings .	tides	इन चित्रों में हाथी , हंस , बैल , गाय , पपीहा , तोता , मैना , चकोरचकोरी , मोर मोरनी , पेड़जड़ पशुधे , केला , जानुन , बट , सरु , सेंबल , लता , मेघ , घटाएँ बिजली , सूर्यार्हचन्द्र आदि का प्रयोग अतन्त्र-कल्प-यंत लोकप्रिय है .
launchpad	Width , in pixels , of the tree view grid lines	launchpad	तरु दृश्य गिरेड लाइन की चौड़ाई , पिक्सल में
Indic2012	Tagore established Shanti Niketan in the company of nature with trees , gardens and a library .	Indic2012	प्रकृति के सान्निध्य में पेड़ों बगीचों और एक लाइब्रेरी के साथ टैगोर ने शांतिनिकेतन की स्थापना की ।
tides	What seems to be meant here is that there was a tree of the species called vata at the point of confluence of the Ganga and the Jamuna at Prayaga which was known as the playaga tree , or the tree at Prayaga .	tides	इससे आशय यह रहा होगा कि प्रयाग में गंगा और जमुना के संगम पर वट जाती का कोई वृक्ष था जिसे प्रयाग-वृक्ष या प्रयाग का वृक्ष कहते होंगे .
quote-sent	Character is like a tree and reputation like a shadow . The shadow is what we think of it ; the tree is the real thing .	quote-sent	चरित्र एक वृक्ष है और मान एक छाया । हम हमेशा छाया की सोचते हैं ; लेकिन असलियत तो वृक्ष ही है ।
tides	Leading from the open yards , there should preferably be a common spacious grazing ground on either side with a few shady trees .	tides	इन खुले अहातों से लगा खूब खूबा एक एक मैदान भी दोनों तरफ होना चाहिए . उसमें कुछ छायादार वृक्ष भी होने चाहिए .
launchpad	Enable Tree Lines	launchpad	तरु पवित्त सकिरय करें
tides	The manna mentioned in the Bible , is the dried and solidified honeydew secreted by the bug Trabutina , living on the tamarisk tree .	tides	बाइबल में जिन मन्ना का उल्लेख आया है वह झाड़ू के वृक्ष पर रहने वाले दूरैयूटिना बग द्वारा स्रावित मधुबिंदु है जिसे सुखाकर टोस बनाया गया है .
tides	The crowing of a crow on house top is supposed to be a prior intimation of the arrival of a house guest , but if he sits on a withered tree and cows , it is supposed to foreshadow tragedy .	tides	यदि कौए का परात : ही मुडेर पार बोलना किसी अतिथि के आगमन का संकेत है तो किसी टूट पर कार्य-कार्य करना किसी अशुभ समाचार का .
agro-exact	If the new leaves at the apex of a mango tree are getting burnt mainly because of a fungus , then spray 0.5 % Bordo mixture . Spray the solution of 25 gm M-45 in 10 litre water , once in 15 days .	agro-exact	यदि आम के पेड़ के शीर्ष पर स्थित नवीन पत्तियों मुख्य रूप से कवको के कारण जलने लगती हैं , तब 0.5 % बोर्डोमिश्रण का छिड़काव करें तथा 10 लीटर पानी में 25 ग्राम M-45 को घोलकर उसका पन्द्रह दिनों में एक बार छिड़काव करें ।
Intercorp	Where have I seen this before , he wondered , where is this pool with reed-grown banks , shut in by level meadows , and with an avenue of poplar trees lining the road that loses itself in the white shadow of a birch copse ?	Intercorp	यह सब मैंने पहले कहाँ देखा है ? तालाब , जिसके तटों पर लम्बे - लम्बे बाँस लगे हैं , समतल चरागाहों से घिरा हुआ ; सड़क पर झुमते हुए विनार के पेड़ों की पंक्ति जो कहीं दूर जाकर सनोवर - झाड़ी की सफेद छाया में गायब हो जाती है ।
tides	Many insects , which ordinarily deposit eggs on leaves of trees during spring and summer , take in autumn the additional precaution of also fixing the leaves to the twig with a hard secretion , so that the leaf with the eggs may not fall to the ground in the general autumnal leaf-fall .	tides	ऐसे अनेक कीट जो साधारणतया बसंत और गर्मी के दौरान पेड़ों की पत्तियों पर अंडे देते हैं वे एक वृद्ध स्राव द्वारा पत्तियों को टहनी से चिपकाने के अतिरिक्त सावधानी बरतते हैं ताकि सामान्य पतझड़-ऋतु में जब पत्तियाँ झड़ती हैं तो अंडे भी उनके साथ जमीन पर न गिर जाएं .
wiklner2013inflected	List of Indian timber trees	wiklner2013inflected	भारत के इमारती लकड़ी वाले वृक्ष

बट प्रोथ के अनुसार इस मोरिथम में कई शाखाएँ थीं और लोग संतान प्रणति की इच्छा से प्रकृति प्रजा करने

# Coming soon... **with your help**

- Web interface both in English and Czech
- More services
- More data
- More REST (like) APIs
- ~~More users~~ More **active and satisfied** users!

# Open Issues (coffee break topics)

- Keeping up with the design updates
- Keeping up with software development
- (Global) user authorizations
- Legal issues regarding usage of data and services
- Limited resources (API usage)
- API sustainability (versioning, availability, ...)



**Thank you for  
your attention**

<http://lindat.mff.cuni.cz/services>