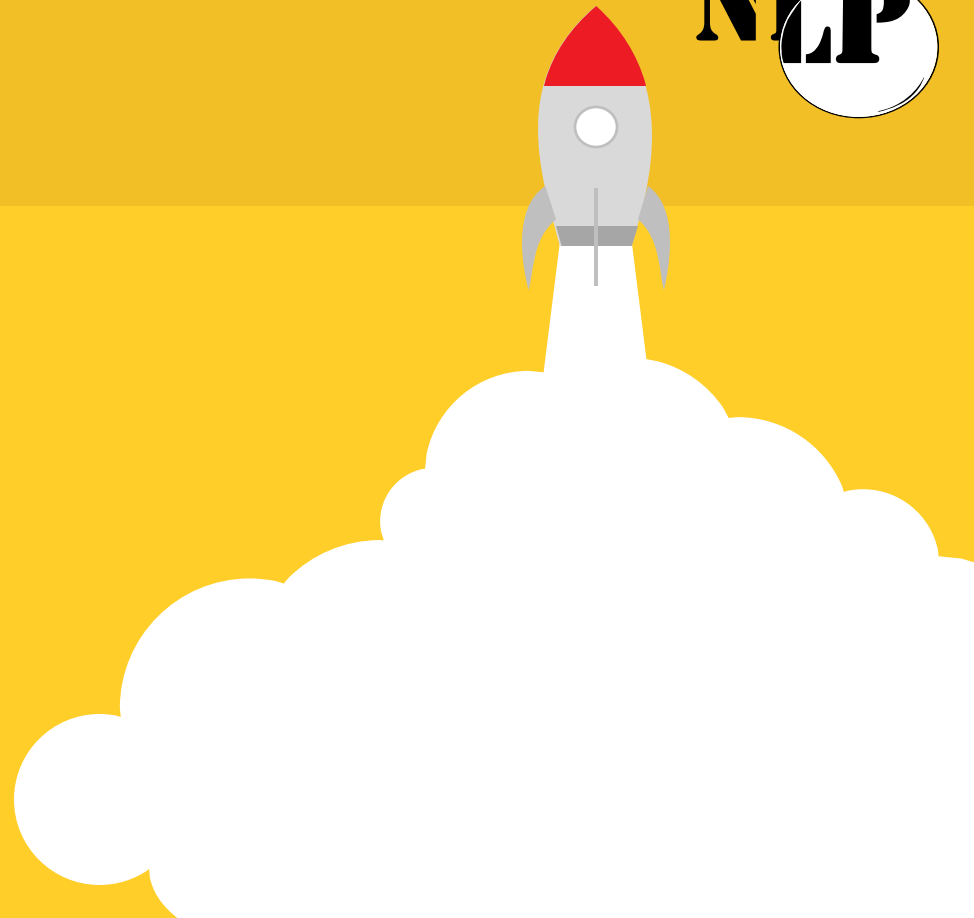


Five Years of Language Services

Zuzana Nevěřilová, 2023-12-08
RASLAN 2023, Červenohorské sedlo



Agenda



01

Language Services Overview

02

Usage Statistics

Fun facts found in the logs

03

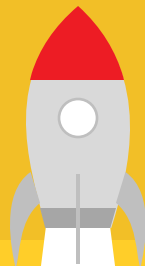
Language Services Misuse

Where users do not know how to use it right

04

Future Work

Improve or let go?



LanguageServices

NLP APIs <https://nlp.fi.muni.cz/languageservices/>

`service.py?call=<service_name>&key=value...`

Overview

service name	description	parameters	returns	example call	text length limit												
sholva Shallow ontology for Czech words	This service returns negative or positive membership of the word in the following classes: person, person-individual, event, substance.	<table><thead><tr><th colspan="2">text</th></tr><tr><th>type</th><td>string</td></tr><tr><th>desc</th><td>word</td></tr></thead></table>	text		type	string	desc	word	<table><thead><tr><th colspan="2">concepts</th></tr><tr><th>type</th><td>list</td></tr><tr><th>desc</th><td>The output format is a list of positive or negative memberships to some (or all) of the classes.</td></tr></thead></table>	concepts		type	list	desc	The output format is a list of positive or negative memberships to some (or all) of the classes.	service.py?call=sholva&lang=cs&output=json&text=šampaňské	1000
text																	
type	string																
desc	word																
concepts																	
type	list																
desc	The output format is a list of positive or negative memberships to some (or all) of the classes.																
topics Get topics of a text	This service returns a list of topics found in a short text. The topic candidates are noun phrases ordered by importance.	<table><thead><tr><th colspan="2">text</th></tr><tr><th>type</th><td>string</td></tr><tr><th>desc</th><td>text</td></tr></thead></table>	text		type	string	desc	text	<table><thead><tr><th colspan="2">response</th></tr><tr><th>type</th><td>list</td></tr><tr><th>desc</th><td>list of noun phrases ordered by score</td></tr></thead></table>	response		type	list	desc	list of noun phrases ordered by score	service.py?call=topics&lang=cs&output=json&text=Jednoho letního večera na návsi pod starou lípou hostinský Antonín Kučera vyvalil soudeček s pípou, nebylo to posvícení, nebyla to neděle, v naší obci mezi kopci plnily se korbele. Byl to ten slavný den, kdy k nám byl zaveden elektrický proud, střídavý, střídavý, silný elektrický proud, střídavý, střídavý, zkrátka elektrický proud.	1000
text																	
type	string																
desc	text																
response																	
type	list																
desc	list of noun phrases ordered by score																
phrases Extract (sub)phrases	This service processes an input sentence and returns all phrases	<table><thead><tr><th colspan="2">text</th></tr><tr><th>type</th><td>string</td></tr><tr><th>desc</th><td>text</td></tr></thead></table>	text		type	string	desc	text	<table><thead><tr><th colspan="2">phrases</th></tr><tr><th>type</th><td>list</td></tr><tr><th>desc</th><td>The list of phrases, i.e. lists of phrase</td></tr></thead></table>	phrases		type	list	desc	The list of phrases, i.e. lists of phrase	service.py?call=phrases&lang=cs&output=json&text=Konečně levný mobil - nástupce kompaktního modelu má být velkým lákadlem	1000
text																	
type	string																
desc	text																
phrases																	
type	list																
desc	The list of phrases, i.e. lists of phrase																

Unified Service Definition

```
{
  "name": "hello",
  "title": "Hello Service",
  "desc": "This service returns a greeting.
           It's practically useless since it is only an example.",
  "exec": "http://nlp.fi.muni.cz/projekty/languageservices/
           hello.py?name={text}&output={output}",
  "parameters":
    {"name":
      {"type": "string",
        "desc": "the person you want to greet or 'World' if not provided",
        "default": "World"}},
  "returns":
    {"response":
      {"type": "string", "desc": "a greeting", "default": "Hello World!"},
      "service":
        {"type": "string", "desc": "service name", "default": "hello"}},
  "example": "text=Dolly",
  "languages": ["en"],
  "limit": 40
}
```

Unified Service Definition

```
{  
  "name": "hello",  
  "title": "Hello Service",  
  "desc": "This service returns a greeting.  
          It's practically useless since it is only an example."  
}
```

hello

Hello Service

This service returns a greeting. It's practically useless since it is only an example.

name	
type	string
desc	the person you want to greet or 'World' if not provided

response	
type	string
desc	a greeting

service	
type	string
desc	service name

service.py?call=hello&lang=en&output=json&text=Dolly

40

```
  {"response":  
    {"type": "string", "desc": "a greeting", "default": "Hello World!"},  
    "service":  
    {"type": "string", "desc": "service name", "default": "hello"}},  
  "example": "text=Dolly",  
  "languages": ["en"],  
  "limit": 40  
}
```

Language Services Usage

Service name	Requests
declension	10,676
diacritics	12,312
get location	1,034
hello	422
majka	5,673
polite	1,649
tagger	38,759
topics	7,607
unknown	25
vocative	1,071
total	79,228

- **Diacritics** service peak usage in 2022-12
 - strelec (sniper) – Slovak IP addresses
- **Declension** peak usage in 2020-05 from Slovak IP addresses:
 - zmizelý Kim Čong-un (the missing Kim Jong Un)
 - dřevostavitelka (woodworker)
- **Declension** used by ssakhk.cz (high school in Hradec Králové)
- **Majka** peak usage by Czechitas project in 2019-12
- **Majka** peak usage (>3000) in 2022-12 from amazonaws.com:
 - diamanty (diamonds), sneholak (snowman)
- **Majka** peak usage in 2020-04 (388):
 - political agenda of the ANO political party
- **Tagger** service English usage: Goosey Gazette
- **Tagger** service peak usage in 2020-05 from Slovak IP



Language Services Usage

Service name	Requests
declension	10,676
diacritics	12,312
get location	1,034
hello	422
majka	5,673
polite	1,649
tagger	38,759
topics	7,607
unknown	25
vocative	1,071
total	79,228

- **Declension** used by ssakhk.cz (high school in Hradec Králové)
- **Majka** peak usage by Czechitas project in 2019-12
- **Majka** peak usage (>3000) in 2022-12 from amazonaws.com:
 - diamanty (diamonds), sneholak (snowman)
- **Majka** peak usage in 2020-04 (388):
 - political agenda of the ANO political party
- **Tagger** service English usage: Goosey Gazette
- **Tagger** service peak usage in 2020-05 from Slovak IP addresses:
 - Job position names
- **Topics** service peak usage in 2021-02 (>3000 requests)
 - Short texts by fans of Milion chvilek pro demokracii (Million moments for democracy)
 - Liberty, a better future, change



Services Misuse

Where users do not know...



Diacritics
Used for words
with diacritics



Diacritics
Used for single
words



Tagger
Used for single
words



Majka
Used for
keyword
extraction



Services Usage by IP Addresses

Who uses it for what

IP address	declension	topics	tagger	majka	polite	diacritics	vocative
Cloud							
Schools							
Slovak IP							
Czech IP							
None of the above							



Future Work

Informative API

Instead

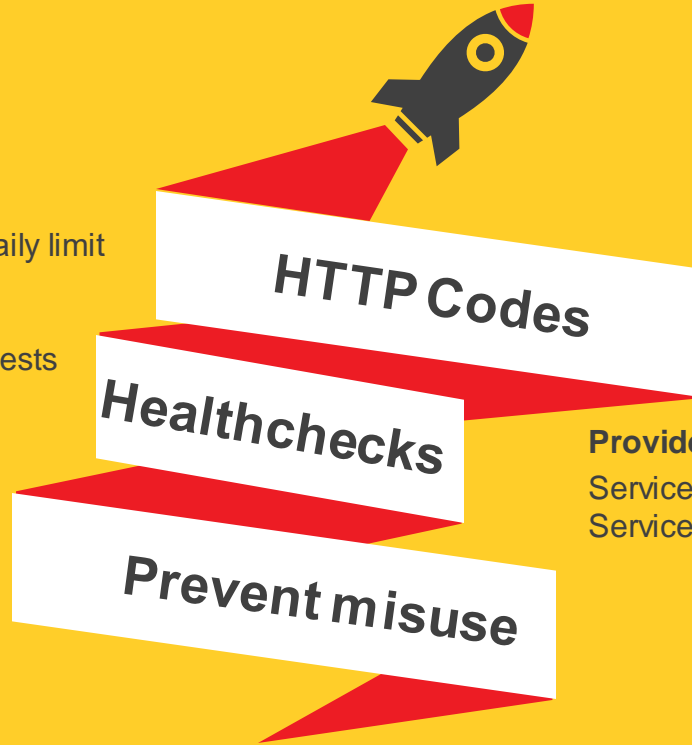
```
HTTP 200 {"message": "Daily limit exceeded"}
```

respond

```
HTTP 429 Too Many Requests
```

Monitor usage

Alert users when they repeatedly misuse a service.



HTTP Codes

Healthchecks

Prevent misuse

Provide info about application health

Service X is working.

Service Y is down.



And, of course,

New services!

Neural Tagger

netagger



```
"tokens":  
  [  
    { "lemma": "na",  
      "token": "Na",  
      "tag": "k7c6",  
      "conf": 0.9986273050308228 },  
    { "lemma": "okno",  
      "token": "okně",  
      "tag": "k1gNnSc6",  
      "conf": 0.8368463158607483 },  
    { "lemma": "sedět",  
      "token": "seděla",  
      "tag": "k5eAaImAgFnSrD",  
      "conf": 0.9982708394527435 },  
    { "lemma": "kočka",  
      "token": "kočka",  
      "tag": "k1gFnSc1",  
      "conf": 0.9978978931903839 },  
    { "lemma": "",  
      "token": ",",  
      "tag": "kIxX",  
      "conf": 0.9999971389770508 },  
    ...  
  ]
```

Named Entity Recognizer

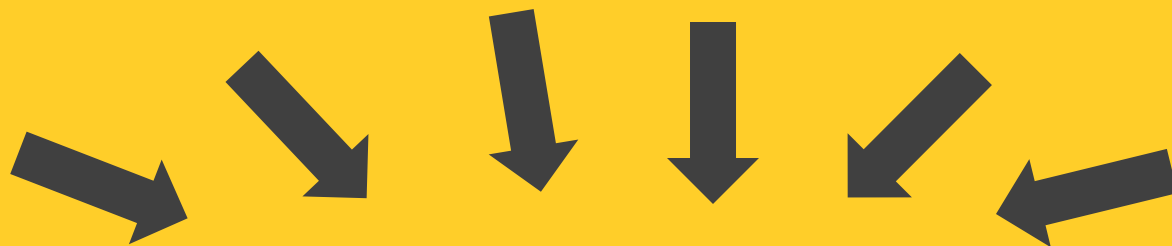
ner



RASLAN se letos poprvé koná na Červenohorském sedle. Od 8. prosince se tam sejdou lidi z NLP Centra.

```
{ "end": 6,  
  "entity_group": "I",  
  "score": 0.8467003703117371,  
  "word": " RASLAN",  
  "start": 0 },  
{ "end": 51,  
  "entity_group": "G",  
  "score": 0.8967188596725464,  
  "word": " Červenohorském sedle",  
  "start": 31 },  
{ "end": 68,  
  "entity_group": "T",  
  "score": 0.9001127481460571,  
  "word": " 8. prosince",  
  "start": 57 },  
{ "end": 100,  
  "entity_group": "I",  
  "score": 0.8720375299453735,  
  "word": " NLP Centra",  
  "start": 90 }
```

Your Greatest NLP App



<https://nlp.fi.muni.cz/languageservices/>