

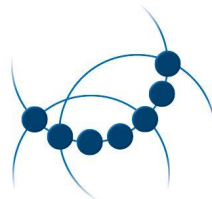
Website of CLASSLA

CLARIN Knowledge centre for
South Slavic languages

Nikola Ljubešić
Jožef Stefan Institute
Slovenia



CLARIN
K CENTRE

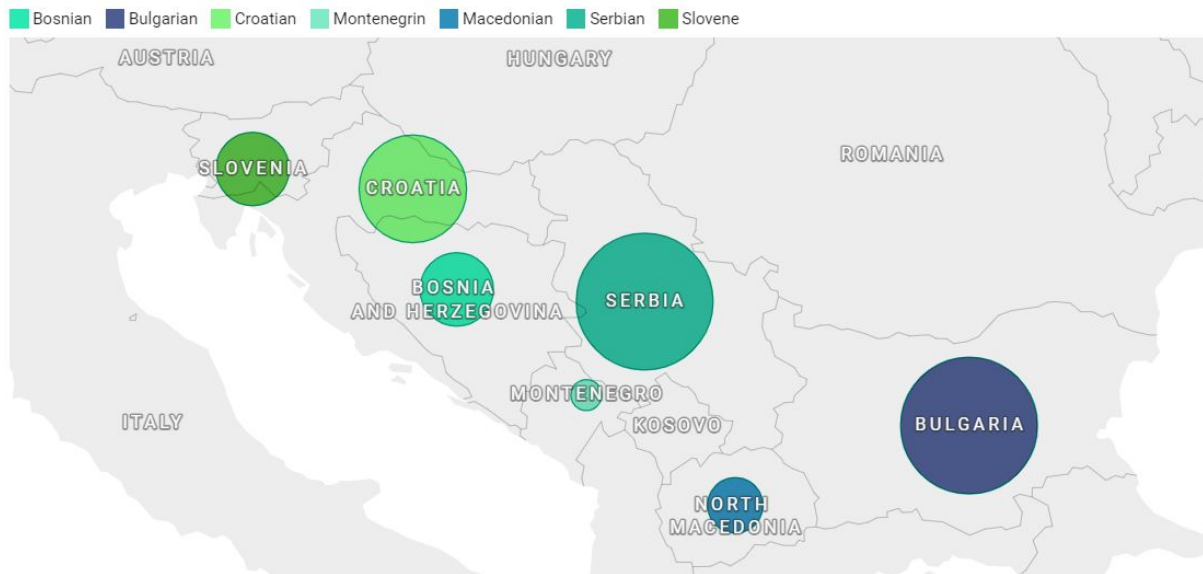


CLaDA**BG**

CLASSLA

Run by CLARIN.SI (Slovenia) and CLaDA-BG (Bulgaria)

Set up in March 2019, uses mostly the CLARIN.SI infrastructure





Slovenska raziskovalna infrastruktura za jezikovne vire in tehnologije
Common Language Resources and Technology Infrastructure,
Slovenia

CLASSLA: KNOWLEDGE CENTRE FOR SOUTH SLAVIC LANGUAGES

The CLARIN Knowledge Centre for South Slavic languages (CLASSLA) offers expertise on language resources and technologies for South Slavic languages. Its basic activities are (1) giving researchers, students, citizen scientists and other interested parties information on the available resources and technologies via its documentation, (2) supporting them in producing, modifying or publishing resources and technologies via its helpdesk and (3) organizing training activities.

The helpdesk of CLASSLA can be contacted via helpdesk.classla@clarin.si. The helpdesk offers additional clarifications regarding the CLASSLA documentation (detailed below) and support in using, modifying, producing, or publishing resources and technologies for South Slavic languages.

The Knowledge Centre currently offers frequently asked questions (FAQ) documentation for the

WHAT'S NEW:

Join CLASSLA's Mailing List

We are happy to announce that our Knowledge Centre for South Slavic Languages, [CLASSLA](#), now has a mailing list. You can subscribe to it [here](#) to be informed of new resources, technologies, events and projects for South Slavic languages. All relevant announcements will be also published [here](#).

Tour de CLARIN: The CLARIN Knowledge Centre for South



Slovenska raziskovalna in
tehnologije
Common Language Resource
Slovenia

- FAQ for Slovene
- FAQ for Croatian
- FAQ for Serbian
- FAQ for Bulgarian
- FAQ for Macedonian
- Web services
- Workshops
- Success stories

kovne vire in
ogy Infrastructure,

CLASSLA: KNOWLEDGE CENTRE FOR SOUTH SLAVIC LANGUAGES

The CLARIN Knowledge Centre for South Slavic languages (CLASSLA) offers expertise on language resources and technologies for South Slavic languages. Its basic activities are (1) giving researchers, students, citizen scientists and other interested parties information on the available resources and technologies via its documentation, (2) supporting them in producing, modifying or publishing resources and technologies via its helpdesk and (3) organizing training activities.

The helpdesk of CLASSLA can be contacted via helpdesk.classla@clarin.si. The helpdesk offers additional clarifications regarding the CLASSLA documentation (detailed below) and support in using, modifying, producing, or publishing resources and technologies for South Slavic languages.

The Knowledge Centre currently offers frequently asked questions (FAQ) documentation for the

WHAT'S NEW:

Join CLASSLA's Mailing List

We are happy to announce that our Knowledge Centre for South Slavic Languages, [CLASSLA](#), now has a mailing list. You can subscribe to it [here](#) to be informed of new resources, technologies, events and projects for South Slavic languages. All relevant announcements will be also published [here](#).

Tour de CLARIN: The CLARIN Knowledge Centre for South

1. Online Slovene language resources

Q1.1: Where can I find Slovene dictionaries?

Q1.2: How can I analyse Slovene corpora online?

Q1.3: Which Slovene corpora can I analyse online?

Q1.4: What linguistic annotation schemas are used in Slovene corpora?

Q1.5: Where can I download Slovene resources?

2. Tools to annotate Slovene texts

Q2.1: How can I perform basic linguistic processing of my Slovene texts?

Q2.2: How can I standardize my texts prior to further processing?

Q2.3: How can I annotate my texts for named entities?

Q2.4: How can I syntactically parse my texts?

3. Datasets to train Slovene annotation tools

Q3.1: Where can I get word embeddings or pre-trained language models for Slovene?

Q3.2: What data is available for training a text normaliser for Slovene?

Q3.3: What data is available for training a part-of-speech tagger for Slovene?

Q3.4: What data is available for training a lemmatiser for Slovene?

Q3.5: What data is available for training a named entity recogniser for Slovene?

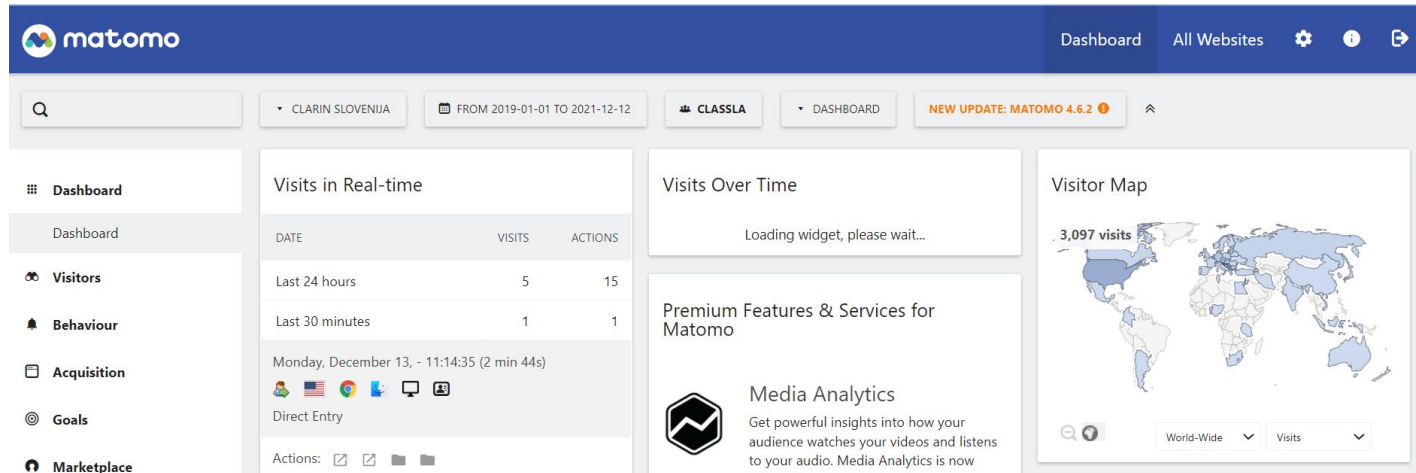
Q3.6: What data is available for training a syntactic parser for Slovene?

Web traffic analysis

Matomo - instance installed at the Jožef Stefan Institute

Wordpress has similar plugins

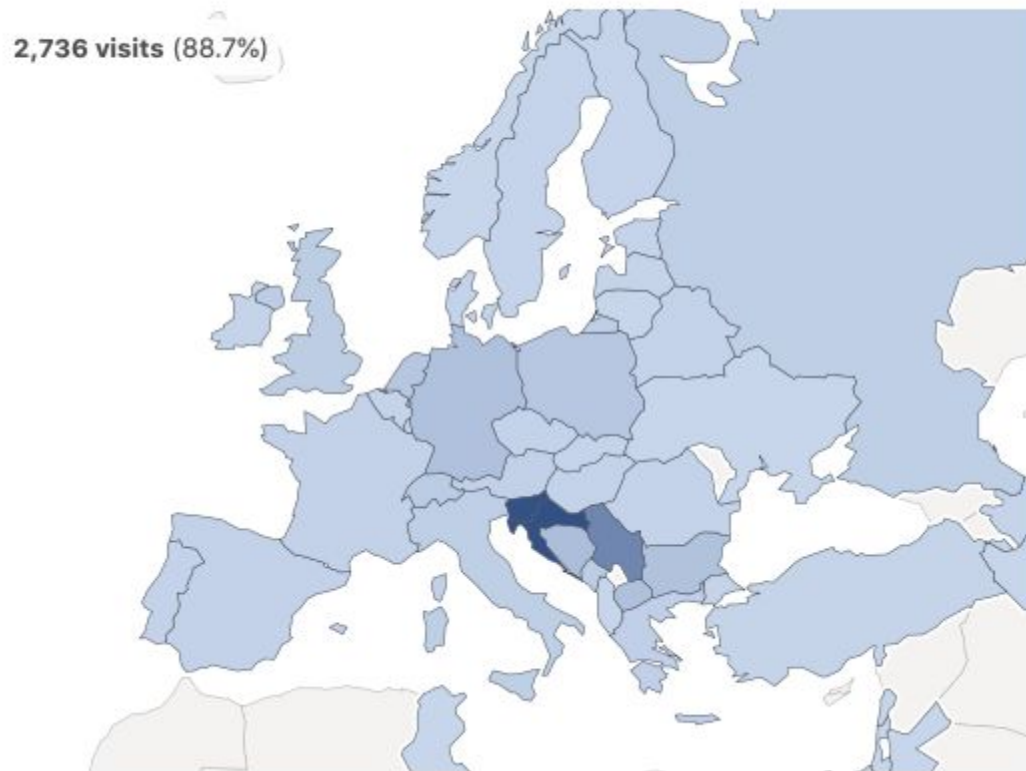
We consider this highly useful tool to understand and tailor content('s visibility)



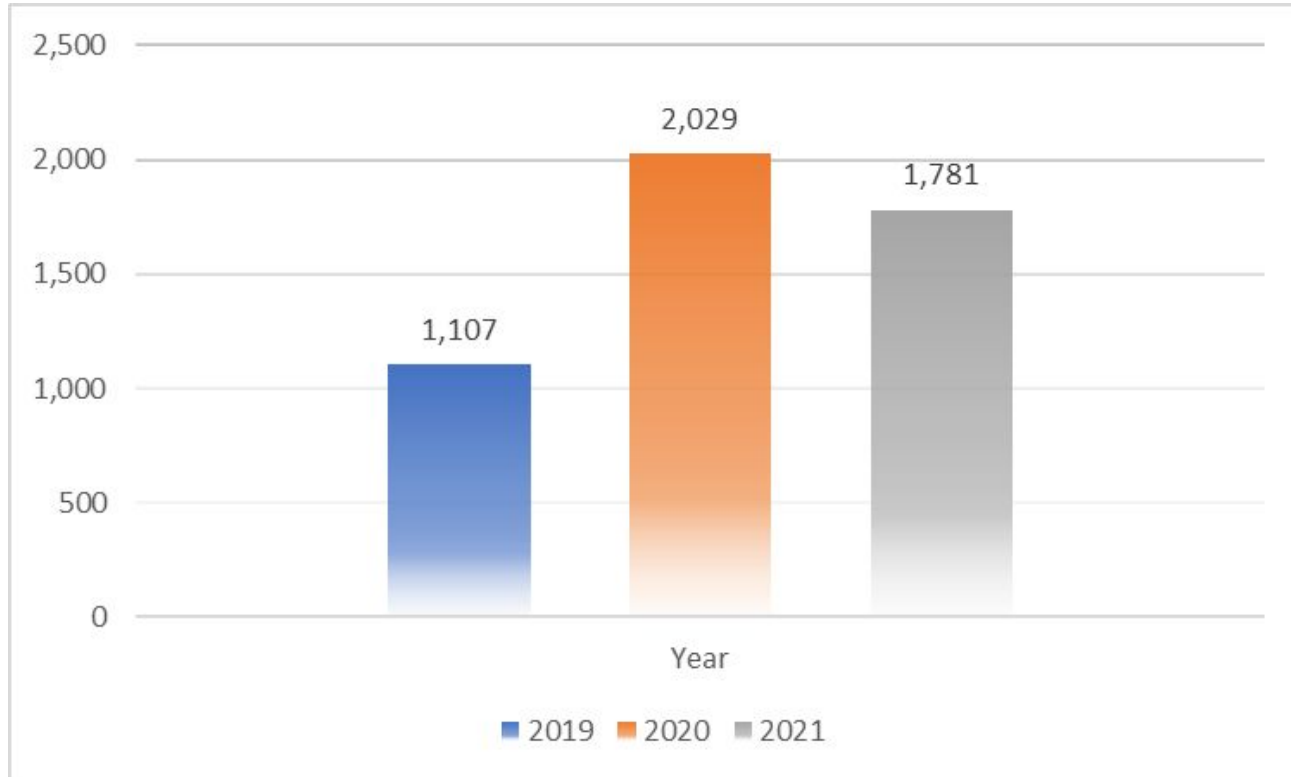
Page views

[-] k-centre	4,951
[+] web-services-documentation	1,890
[↗] /index	1,261
[+] faq4slovene	337
[+] faq4serbian	362
[+] faq4croatian	330
[+] workshops	270
[+] faq4macedonian	266
[+] faq4bulgarian	198
[+] success-stories	30
[+] k-center	1,002

Geographic distribution

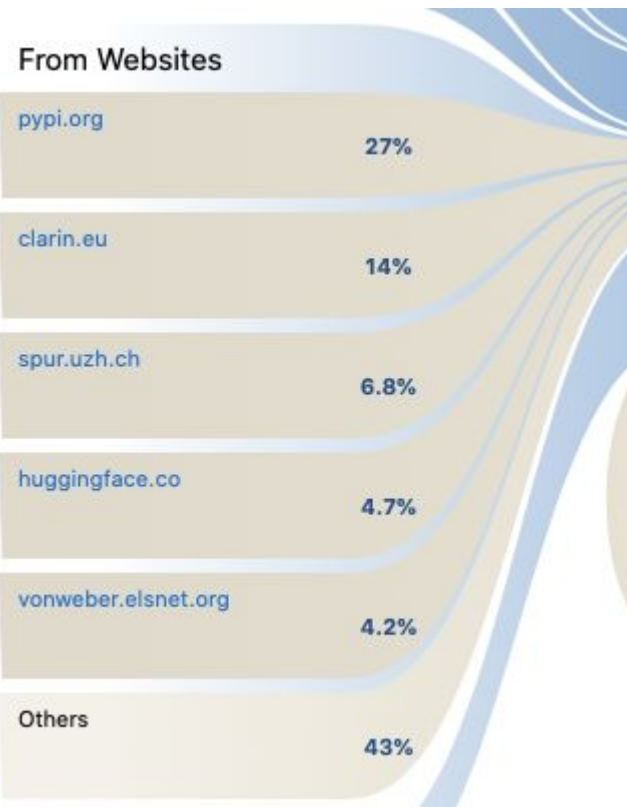


Time distribution



Entry points

CHANNEL TYPE	▼ VISITS	ACTIONS	ACTIONS PER VISIT	AVG. TIME ON WEBSITE
Direct Entry	1,403	6,291	4.5	5 min 11s
⊕ Search Eng...	1,008	4,126	4.1	5 min 32s
⊕ Websites	624	3,323	5.3	7 min 23s
⊕ Social Net...	24	118	4.9	11 min 13s



Takeaways

- Be practical in the web page setup – can be part of a CLARIN centre website
- Consider web traffic analysis solutions early on
- NLP is hot - technologies might bring most visitors to your website
- Use existing web strongholds to promote your website
- We still consider FAQ pages to be very useful, not only for answering specific questions, but also widening horizons

Thanks!