

LINDAT DSpace at CLARIN.SI

Cyprian Laskowski

Jožef Stefan Institute

9 November 2016

- CLARIN.SI is small
- The local LINDAT/Dspace repository is the core of CLARIN.SI
- Now a CLARIN B centre
- Close to 50 submissions, all with substantial data, mostly in HLT
- Mostly from Slovenia and for Slovene
- No tools yet
- Example submissions: SNABI (speech database), Emoji Sentiment (lexicon), SlovParl (transcriptions)

Find
Linguistic Data and NLP Tools
Citation Support (with Persistent IDs)

CLARIN.SI

Repository About Contact

Search

Advanced Search

Author	Subject	Language (ISO)
Erjavec, Tomaž (19)	lemmatisation (10)	Slovenian (31)
Ljubešić, Nikola (12)	tagging (9)	English (13)
Klubička, Filip (10)	parallel corpus (8)	Serbian (8)
Krek, Simon (9)	web corpus (8)	Bulgarian (5)
Dobrovoljc, Kaja (6)	historical language (5)	Croatian (5)
... View More	... View More	... View More

What's New

Corpus

CLARIN.SI Data & Tools

What can you do?

DEPOSIT

- Local LINDAT DSpace installation in 2014
- Local branding (based on common-theme)
- Two-year “test” phase (but without “fake” submissions)
- Periodic updates/upgrades
- Local overlay management
- Test server
- Localisation (Slovene translation)
- Hiding of unused features (e.g., Google Analytics)

- AAI
- OAI-PMH harvesting
- Overlays
- Common theme
- Shared submission
- Refbox
- Piwik
- Etc.

Features hoped for

- Simpler localisation
- Improved bibliography exports
- Integration with improved Kontext
- More harvesters (e.g, Thomson Reuters)
- Better statistics summaries
- Better admin documentation

- Stabilisation (minor local bug fixes, etc.)
- Better Kontext installation
- Improved monitoring
- Statistics analysis
- More complete localisation (and for DSpace?)
- Adding of CLARIN and custom licenses
- Move from sys to user support (documentation, webcasts, etc.)
- Providing web services and tools