Software Release and Progress Report

Start date: 2019-10-15 End date: 2020-01-14

Short Summary

The past period, a major new CLAM version was released (v3.0), featuring a new and improved user interface and user experience in general. It contains numerous improvements and has been integrated in LaMachine and deployed in Nijmegen.

FLAT has once more been adopted as the annotation tool of choice for another round of the PARSEME project.

LaMachine has again had a few more interations of added software, improvements, and fixes. It now more explicitly targets certain Linux distribution/operating systems and indicates the level of support (gold/silver/bronze), to temper user expectation and keep things more maintainable in the long run.

Maarten has specified various potential workflows for the WP3 VRE project, as the VRE developers were struggling with the existing list that was very underspecified and incomplete. The list explores technical possibilities and gives suggestions, what is still lacking but is very important, is an actual user base! (I'm a bit worried about that actually)

The need arose for a webservice to do conversions, as an alternative to OpenConvert with which there were issues (from the VRE developers) and which has gotten out of date. Maarten wrote a new CLAM-based webservice called Piereling, leveraging existing solutions such as pandoc, the service is mostly focussed on conversions from and to FoLiA. The (limited!) TEI-to-folia converter developed as part of Nederlab is also integrated.

Maarten attended (virtually) the CLARIN switchboard workshop and is working on further integratations of Nijmegen-hosted CLAM-based services. He's in contact with the switchboard developers for closer collaboration.

Maarten spoke with Gertjan Filarski about CLARIAH-as-a-service and will help set up the workflow interest group with Jauco, and also participate in the Text Infrastructure group (and perhaps Annotation Infrastructure).

Fryske Akademy now runs their own LaMachine instance for deployment of Oersetter and no longer relies on Nijmegen's hosting.

The PyOpenSci community expressed an interest in the codemeta standard (3rd party) for software metadata and the codemetapy tool Maarten wrote for it (as also used by LaMachine).

Henk and Nelleke worked on a couple of issues:

- Design and implementation of the website for the new CLARIN Knowledge Centre for Atypical Communication Expertise: https://ace.ruhosting.nl/
- Three LREC papers were submitted, one addressed the potential of the ACE K-centre mentioned bove, another presented two case stuydies on the GDPR and Sensitive data, and a third one the CLLARIN Transcription Chain, portal and dissemination
- A second website was designed and implemented for the CLARIN taskforce for pathological language data resources: http://delad.net
- Nelleke started her report on compatibility of metadata for SoNaR and CGN and will finish the report in January 2020.

CLAM

Project/task/deliverable references: CLARIAH-PLUS WP3 T142; CLARIAH-CORE WP3 T21 [D2.8 (software); D2.9 (doc)]

clam v3.0.9

Bugfix release: * fixing information leakage between anonymous users: always operate on a COPY of a parameter

(Released by Maarten van Gompel on 2019-12-21) https://github.com/proycon/clam/releases/tag/v3.0.9

clam v3.0.8

Minor fix/update release: * output proper http headers in custom format in textstats example #87 * minor fix of misleading debug message #85 * make all the new extra metadata available in client/data API

(Released by Maarten van Gompel on 2019-12-13) https://github.com/proycon/clam/releases/tag/v3.0.8

clam v3.0.7

Minor bugfix release: * Better handling of unset parameters, raise exceptions rather than return default value False (except for booleans) * Fixed Action method

(Released by Maarten van Gompel on 2019-11-08) https://github.com/proycon/clam/releases/tag/v3.0.7

clam v3.0.6

Minor bugfix release: * Correctly show registration link in interface (if applicable) * Fixed a minor bug in error reporting

(Released by Maarten van Gompel on 2019-11-04) https://github.com/proycon/clam/releases/tag/v3.0.6

clam v3.0.5

Fix for the flat viewer

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.5

clam v3.0.4

Another post v3 release interface fix (first inputtemplate was missing)

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.4

clam v3.0.3

Another quick update after v3 release: fixes some styling issues that became apparent in production environments

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.3

clam v3.0.2

Minor update: * add ability to reference variables from the main service configuration in the external configuration file, not only environment variables.

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.2

clam v3.0.1

- Post-release bugfix: use https headers for jquery download

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.1

clam v3.0.0

Major new release featuring a new interface

Important Note: This release drops all Python 2.7 support, only Python 3.3 and above are supported now!

- Brand new and completely restyled user interface. (#82)
    - Improved user experience, including on mobile devices.
    - Fixed various minor UI nags from the past (such as #43, #53)
    - Added support for a cover image.
    - Allow a link to a registration page.

- – Added support for custom css.
  - Added a 'porch' in the interface, a landing page to the webservice that requires no authentication.
  - Actions that are marked to allow anonymous usage will show up right on the porch page in the interface.
  - Reworked the format validation mechanism, now with support for added constraints on the input. (#84)
  - You can now add add viewer to an action, making them produce styled output rather than a raw download.
  - More system metadata.
  - CUSTOM_FORMATS needs not to be redefined client-side but is automatically read from the XML response. (#81)
  - Viewers can now be associated with actions as well.
  - Access to RESTful API specification from the interface.
  - Added extra webservice metadata, which also shows in the interface.
  - Improved the clamnewproject script to create a cleaner project structure.
  - Various minor fixes (such as #83, 78)

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clam/releases/tag/v3.0.0

clamservices v2.1.0

Added SpaCy service (requires spacy and spacy2folia)

(Released by Maarten van Gompel on 2020-01-14) https://github.com/proycon/clamservices/releases/tag/v2.1.0

clamservices v2.0.2

Minor fixes

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clamservices/releases/tag/v2.0.2

clamservices v2.0.1

Corrected some urls to use HTTPS instead of HTTP

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clamservices/releases/tag/v2.0.1

clamservices v2.0.0

Added extra metadata and interface options for use with the new CLAM v3.

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/clamservices/releases/tag/v2.0.0

FLAT

Project/task/deliverable references: CLARIAH-PLUS WP3 T062 & T063; CLARIAH-CORE WP3 T24 [D2.10 (software); D2.11 (docs)]

flat v0.9.1

Minor bugfix release: * session configuration should not override explicitly set configuration for public documents

(Released by Maarten van Gompel on 2019-11-07) https://github.com/proycon/flat/releases/tag/v0.9.1

FoLiA

Project/task/deliverable references: CLARIAH-PLUS WP3 T108; CLARIAH-CORE WP3 T71 [D2.5 (libs); D2.6 (docs); D2.7 (tools)]

foliapy v2.2.5

Minor fix: * no hard fail on missing version, but assume an old version instead #17

(Released by Maarten van Gompel on 2020-01-13) https://github.com/proycon/foliapy/releases/tag/v2.2.5

foliapy v2.2.4

Minor fix in setup.py so it installs even when not on a proper utf-8 locale. (#16)

(Released by Maarten van Gompel on 2020-01-03) https://github.com/proycon/foliapy/releases/tag/v2.2.4

foliapy v2.2.3

- Fix for items() and added elements() (proycon/foliatools#7)

(Released by Maarten van Gompel on 2019-12-21) https://github.com/proycon/foliapy/releases/tag/v2.2.3

foliatools v2.2.7

- Fixed excessively slow foliaupgrade #7

(Released by Maarten van Gompel on 2019-12-21) https://github.com/proycon/foliatools/releases/tag/v2.2.7

foliatools v2.2.6

Significant fixes and improvements for tei2folia (was previously only tested on DBNL collection, now on several others too).

(Released by Maarten van Gompel on 2019-10-20) https://github.com/proycon/foliatools/releases/tag/v2.2.6

foliatools v2.2.5

- Added txt2folia tool
- [conllu2folia] added –outputfile parameter

(Released by Maarten van Gompel on 2019-10-19) https://github.com/proycon/foliatools/releases/tag/v2.2.5

libfolia v2.3.2

Bug fix release * fix for https://github.com/LanguageMachines/foliautils/issues/37 * fix for https://github.com/LanguageMachines/foliautils/issues/38 * fixes in Correction handling. * fixed a Multi-Threading problem with the static reverse_old map

(Released by Ko van der Sloot on 2020-01-13) https://github.com/LanguageMachines/libfolia/releases/tag/v2.3.2

libfolia v2.3.1

- Bug fix release for gcc 9.1 It stumbles upon some inline functions
- replaced call to unsafe 'tmpnam()' by 'TiCC::tempname()'

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/libfolia/releases/tag/v2.3.1

Frog

Project/task/deliverable references: CLARIAH-PLUS WP3 T139; CLARIAH-CORE WP3 T22 [D2.1 (software); D2.2 (doc)]; T23 [D2.3 (froggen software), D2.4 (froggen docs)]

frog v0.19.1

- fixed an overseen incompatibility problem with the new libfolia. (https://github.com/proycon/tscan/issues/13)
- removed dependency on MbtServer
- Some documentation updates
- improved using Alpino, using unique filenames now.

(Released by Ko van der Sloot on 2019-11-15) https://github.com/LanguageMachines/frog/releases/tag/v0.19.1

frog v0.19

- added code to use a locally installed Alpino parser
- added code to use a remote Alpino Server
- added code to use (remote) timblservers and mbtservers for alle modules using JSON calls. Stil experimental.
- several code refactoring and small fixes:
  - memory leaks
  - using NER files in non-standard locations
  - bug fixes for some corner cases.
- frog.*.debug files are cleaned up after 1 day.

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/frog/releases/tag/v0.19

python-frog v0.4.2

- Fixed weird double freememory error (#17), some caution is in order whether this doesn't cause a memory leak though.

(Released by Maarten van Gompel on 2019-10-30) https://github.com/proycon/python-frog/releases/tag/v0.4.2

toad v0.6

- updated to work with newest mbt/mbtserver

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/toad/releases/tag/v0.6

LaMachine

Project/task/deliverable references: CLARIAH-PLUS WP3 T098; CLARIAH CORE WP3: LaMachine v2 plan in scope of CLARIAH WP3 VRE

codemetapy v0.3.1

Minor bugfix release: do not reset entrypoints when chaining

(Released by Maarten van Gompel on 2019-11-20) https://github.com/proycon/codemetapy/releases/tag/v0.3.1

codemetapy v0.3.0

This release makes some changes to the way codemetapy works: * Instead of parsing pip output, the tool now uses importlib.metadata to query for metadata. As metadata is read after installation, this work regardless of how the metadata was initially specified (setup.yp, setup.cfg or pyproject.toml) New features: * Added an output file parameter (-O) * Added an integration hook for setuptools, allowing users to add a codemeta command to setup.py Fixes: * Prevent duplicates in authors and other fields

(Released by Maarten van Gompel on 2019-11-15) https://github.com/proycon/codemetapy/releases/tag/v0.3.0

labirinto v0.2.5

- Updated a whole bunch of dependencies: webpack 4 and more

(Released by Maarten van Gompel on 2019-11-20) https://github.com/proycon/labirinto/releases/tag/v0.2.5

LaMachine v2.14

- Added julia #165
- Added spaCy webservice (a CLAM webservice with spacy2folia for FoLiA output support)
- Docker: expose more ports by default (8080, 8888 and 9999 are now exposed by default in the Dockerfile)
- Adding kaldi-nl resource (#106) and oral history (+webservice) (#140) [requires private access]
- Minor updates to contributing guidelines **v2.13**:
- Fix for alpino, it is linked against extra libraries which are in a non-default location

- Continue even if shared data path can't be created (due to permission issues sometimes)
- Various fixes for improved jupyter lab integration
  – auto-starting it
  – store jupyter notebooks in www_data_path and allow this to be stored on the shared data volume (relevant for docker/VM)
- Implemented new www_data_path option and a shared_www_data boolean that by defaults sets the data path to be on the shared volume (for VM/container)
  – explicitly ask during boostrap whether to put web data on the shared volume
- **LaMachine now specifically provides support status indications for different Linux distributions/operating systems**, we distinguish four categories (gold, silver, bronze and deprecated). Support indications are given both in the bootstrap strict as in the documentation.
- CentOS 8 support (silver support status), CentOS 7 support is being deprecated.
- Added clam_include directive to include a base CLAM configuration from the configuration of each clam service
- adding various "maintainted by lamachine" comments to configuration files, warning the user not to edit them.
- Fixes and updated instructions for for LXD containers
- Fixes in lamachine-start-webserver and cleanup of lamachine-start-webserver output

(Released by Maarten van Gompel on 2020-01-14) https://github.com/proycon/LaMachine/releases/tag/v2.14

LaMachine v2.12

- Vagrant base box updated to Debian 10 (buster)
- Changing the way metadata is collected, adapting to codemetapy 0.3, making it easier to regenerate metadata, and removing some complexity
- Added piereling (and installing poppler and pandoc by default)
- Allowing quicker updates with the lamachine-update –only parameter no longer process all dependencies if it is set
- Remove labirinto sources before update because it often fails to upgrade nicely
- Allow LaMachine to build against a custom Python 3, by setting python_bin and pip_bin (building the custom Python is not included in LaMachine) #159
- Fixes for Arch Linux #161
- adding hdf5 to base packages (required by keras)
- adding new translitteration service (t2t)
- Various smaller fixes

(Released by Maarten van Gompel on 2019-11-27) https://github.com/proycon/LaMachine/releases/tag/v2.12

LaMachine v2.10.1

Various fixes: * lamachine-list: get $LM_PREFIX dynamically * lamachine-list update * metadata fixes for glem and g2pservice * adding valkuil frontend installation and configuration (not done) * allow setting ignore_errors in package for lamachine-package-install, as it does not propagate to tasks within ignore_role * disabling cask installation for homebrew, they say it comes by default

(Released by Maarten van Gompel on 2019-10-17) https://github.com/proycon/LaMachine/releases/tag/v2.10.1

Miscellaneous

Project/task/deliverable references: Dependencies/wrappers and or unforeseen tools (related to CLARIAH projects)

piereling v0.2

Minor fixes and added metadata

(Released by Maarten van Gompel on 2019-11-08) https://github.com/proycon/piereling/releases/tag/v0.2

piereling v0.1

First experimental release. Some convertors (like hocr and alto) are still to be implemented.

(Released by Maarten van Gompel on 2019-10-21) https://github.com/proycon/piereling/releases/tag/v0.1

ticcutils v0.23

- re-implementation of ServerBase.cxx As this breaks the API/ABI, the library version is bumped.
- added 2 functions to FileUtils:
  - erase() to erase a file or directory
  - tempname() to create a temporary file. (to replace the dreaded tmpnam() function)
- fixed small bug in CommandLine parsing

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.23

Nederlab

Project/task/deliverable references: Nederlab

(no releases this period)

PICCL & TICCL

Project/task/deliverable references: CLARIAH-PLUS WP3 ???; CLARIAH-CORE WP3 T26 [D1.3 (PICCL software); T1.4 (PICCL docs)]

(no releases this period)

Software Quality Guidelines

Project/task/deliverable references: CLARIAH-CORE WP2 Task 54.100

(no releases this period)

Timbl & Mbt

Project/task/deliverable references: Pre-CLARIN

mbt v3.5

- added JSON support Still experimental!
- several code improvements; including 1 to avoid inappropriate exit codes

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/mbt/releases/tag/v3.5

mbtserver v0.13

- added JSON support. The JSON syntax might change in the future, so handle with care.
- lots of refactoring.

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/mbtserver/releases/tag/v0.13

python-timbl v2019.12.09

- Minor installation expansion: added lib64 to default search path for boost-python

(Released by Maarten van Gompel on 2019-12-08) https://github.com/proycon/python-timbl/releases/tag/v2019.12.09

timbl v6.4.14

- added JSON support. Still EXPERIMENTAL! the JSON syntax might change in the future. So handle with care.
- confidence score calculation is now a real TimblExperiment member
- removed Boost dependency.

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.14

timblserver v1.13

- Added JSON support. The JSON syntax might change in the future, so handle with care.
- lots of refactoring. Depends on recent ticcutils and timbl

(Released by Ko van der Sloot on 2019-10-21) https://github.com/LanguageMachines/timblserver/releases/tag/v1.13

Ucto

Project/task/deliverable references: CLARIAH-CORE WP3 T55 [D1.1 (software), D1.2 (docs)]

ucto v0.20

Bug fix release. solving: * https://github.com/LanguageMachines/frog/issues/84 * https://github.com/LanguageMachines/fro * https://github.com/LanguageMachines/ucto/issues/76 * https://github.com/LanguageMachines/ucto/issues/74

(Released by Ko van der Sloot on 2019-11-27) https://github.com/LanguageMachines/ucto/releases/tag/v0.20