Software Release and Progress Report

Start date: 2018-04-03 End date: 2018-06-21

CLAM

Official task and deliverable references: CLARIAH T21 [D2.8 (software); D2.9 (doc)]

clam v2.3.3

- Introducing `INTERNALURLPREFIX` configuration parameter as an attempted fix for issue #68

(Released by Maarten van Gompel on 2018-05-10) https://github.com/proycon/clam/releases/tag/v2.3.3

clam v2.3.2

- Fix for double slash problem in URLs

(Released by Maarten van Gompel on 2018-05-09) https://github.com/proycon/clam/releases/tag/v2.3.2

clam v2.3.1

- Fixed regression bug in v2.3 when using urlprefix

(Released by Maarten van Gompel on 2018-05-02) https://github.com/proycon/clam/releases/tag/v2.3.1

clamservices v1.6.2

Alpino bugfix release

(Released by Maarten van Gompel on 2018-05-04) https://github.com/proycon/clamservices/releases/tag/v1.6.2

clamservices v1.6.1

Bugfix release

(Released by Maarten van Gompel on 2018-05-02) https://github.com/proycon/clamservices/releases/tag/v1.6.1

clamservices v1.6

- Added alpino webservice (merged into clamservices from standalone repository)

(Released by Maarten van Gompel on 2018-05-02) https://github.com/proycon/clamservices/releases/tag/v1.6

FLAT

Official task and deliverable references: CLARIAH T24 [D2.10 (software); D2.11 (docs)]

foliadocserve v0.6.7

- Bugfix release for metadata changes not saving (#13), thanks to @OsmanMutlu

(Released by Maarten van Gompel on 2018-04-23) https://github.com/proycon/foliadocserve/releases/tag/v0.6.7

FoLiA

Official task and deliverable references: CLARIAH T71 [D2.5 (libs); D2.6 (docs); D2.7 (tools)]

foliautils v0.9.2

Bug fix release:

- append small prefixes to output filenames, to ALWAYS avoid names starting with a numeric value. 'FPM-' for FoLiA-pm. 'FP-' for FoLiA-page, 'FH-' for FoLiA-hocr Can bet set witth --prefix
- FoLiA-stats.cxx:
  - added --collect to usage() and 'man' page
- FoLiA-correct:
  - added --inputclass and --outputclass parameters (must be different)
  - Don't crash on empty text.

(Released by Ko van der Sloot on 2018-06-05) https://github.com/LanguageMachines/foliautils/releases/tag/v0.9.2

foliautils v0.9.1

Bug Fix release:

- the tests directory wasn't included in the release

(Released by Ko van der Sloot on 2018-05-17) https://github.com/LanguageMachines/foliautils/releases/tag/v0.9.1

foliautils v0.9

[Ko vd Sloot]

- FoLiA-stats.cxx:
  - added a --collect option, to create files with all n-grams together
  - clearer message in FoLiA-stats when no results were found
  - extract text from deeper nodes, if needed
  - fixed out-of-bounds problem
  - now fails when every input file fails
- FoLiA-txt:
  - now fails when every input file fails
- avoid xml:id's starting with a number. Add "id-" in front.
- added more tests

[Maarten van Gompel]

- added codemeta.json

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/foliautils/releases/tag/v0.9

libfolia v1.13

[Ko van der Sloot]

- disabled WordReference test. It was incomplete, and hard to do
- use icu:: namespace

[Maarten van Gompel]

- added codemeta.json
- fix spelling errors in error messages

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/libfolia/releases/tag/v1.13

pynlpl v1.2.7

- More insightful output on FoLiA text consistency errors

(Released by Maarten van Gompel on 2018-05-23) https://github.com/proycon/pynlpl/releases/tag/v1.2.7

pynlpl v1.2.6

- Fix for FoLiA library (#41)
- Partial fix in FrogClient library (#38)

(Released by Maarten van Gompel on 2018-04-30) https://github.com/proycon/pynlpl/releases/tag/v1.2.6

Frog

Official task and deliverable references: CLARIAH T22 [D2.1 (software); D2.2 (doc)]; T23 [D2.3 (froggen software), D2.4 (froggen docs)]

frog v0.15

[Ko vd Sloot]

- ucto_tokenizer_mod: removed call of (useless) ucto:setSentenceDetection(true)
- fix to close the server when a socket fails
- when frogging a file, and the docID is NOT specified, use the filename as the docID (filtering out non-NCName characters)
- fix building the documentation from TeX files
- a lot of small code improvements

[Maarten van Gompel]

- added codemeta.json
- Fixed python-frog example in documentation (closes #48)

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/frog/releases/tag/v0.15

frogdata v0.16

- cleanup and additions to the NER gazeteer files

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/frogdata/releases/tag/v0.16

LaMachine

Official task and deliverable references: CLARIAH: LaMachine v2 plan in scope of CLARIAH WP3 VRE

LaMachine v2.2.7

- `lamachine-update` fix for Docker #94
- improved git ident fix when stashing #88
- added note on disabling Hyper-V on Windows #93
- Better managed version number Due to a bug in `lamachine-update` that affects docker users only, docker users should update to this release by obtaining the new image from Docker Hub.

(Released by Maarten van Gompel on 2018-06-20) https://github.com/proycon/LaMachine/releases/tag/v2.2.7

LaMachine v2.2.6

- Nextflow permission fix
- nginx: allow upload of large files (up to 1GB)
- Apache and puppet examples (thanks to @lbiemans)
- Explicitly force english UTF-8 locale in lamachine-start-webserver script
- Set default HTTP port to 8080 instead of 80 if root is disabled
- Redirect uwsgi output to log + more info in lamachine-start-webserver
- Attempted fix for git ident problem on stash #88 #94
- Rehash $PATH in bootstrap script

(Released by Maarten van Gompel on 2018-06-14) https://github.com/proycon/LaMachine/releases/tag/v2.2.6

LaMachine v2.2.3

- Updated flavour descriptions to cause less confusion
- Set Labirinto portal_registry_url to be relative by default
- Added U_USING_ICU_NAMESPACE for old ticcltools to make it work with newer icu
- added --force parameter to bootstrap (#88), and you can now do `lamachine-update force=2` to explicitly remove all sources prior to updating
- Attempted fix for git stash problems due to missing identity by setting git environment variables; introduced new "maintainer_name" and "maintainer_mail" properties. #88
- Valkuil metadata registration
- Ensure nextflow permissions are sane also in multi-user settngs
- Attempted fix for global installations without sudo (not really well tested yet)

(Released by Maarten van Gompel on 2018-06-05) https://github.com/proycon/LaMachine/releases/tag/v2.2.3

LaMachine v2.2.0

- **Added Labirinto portal** (https://github.com/proycon/labirinto), adds a web-based portal of available software to LaMachine (provided you included webserver support)
- Added `nodejs-core`
- Integrated FLAT #49
- Stick to django 2.0 if user is on python 3.4 (2.1 does not support 3.4)
- Fixed nginx webservertype issue #86
- Various metadata fixes and enhancements
- Added option for including extra metadata of tools not in LaMachine
- Split nextflow into separate role rather than as part of PICCL
- linking uctodata on macOS is now obsolete (fbkarsdorp/homebrew-lamachine#9)
- git stash fixes
- force english utf-8 locale for lamachine-update (prevents "ascii codec can not decode" errors on some systems)

(Released by Maarten van Gompel on 2018-05-22) https://github.com/proycon/LaMachine/releases/tag/v2.2.0

LaMachine v2.1.7

- Various improvements to metadata registry
- Better webservice integration, various fixes
  - alpino & tscan (#47)
  - PICCL (#75)
  - Nginx local compilation fixes (#83)
  - Webserver start/stop script fixes (also when no systemd present)
- also force reinstallation of python packages when `force=1`
- Propagate force parameter to homebrew
- Miscellaneous fixes (#82)

- First functional version of kaldi integration (#32)

(Released by Maarten van Gompel on 2018-05-09) https://github.com/proycon/LaMachine/releases/tag/v2.1.7

LaMachine v2.1.3

- Introduction of new software metadata scheme, using codemeta (#74)
- Fix for extra activation scripts
- Fix for alpino (#61)
- Extra checks for conflicts with anaconda
- Do not download spacy data if it already exists
- Add (locally compiled) nginx in local env variant
- Various other fixes

(Released by Maarten van Gompel on 2018-04-23) https://github.com/proycon/LaMachine/releases/tag/v2.1.3

LaMachine v2.1

- Bootstrap refactoring; use proper host names in host configuration files, no `lamachine-` prefix needed.
- Cleanup after bootstrap (#64)
- LaMachine can now be bootstrapped by an internal or external *controller*
- Added the *remote* flavour
- Various fixes for webservices
- Fixed PICCL (v0.5)
- Added `services` configuration option to explicitly constrain desired services (defaults to: `[ all ]`)
- Added `--targetdir` and `--services` flags to bootstrap (#73)
- `lamachine-update` now allows a `--noninteractive` flag
- Initial add of software by CLTL, VU Amsterdam (#70)
- Introduced `unix_group` and `web_group` configuration options for explicit group ownership
- Added instructions about security to the README
- Expand icu 61 workaround to all platforms #69
- uwsgi fixes for CentOS 7
- Fixes for Linux Mint (hopefully, to be confirmed)

(Released by Maarten van Gompel on 2018-04-06) https://github.com/proycon/LaMachine/releases/tag/v2.1

codemetapy v0.1.4

- Bugfix release

(Released by Maarten van Gompel on 2018-05-19) https://github.com/proycon/codemetapy/releases/tag/v0.1.4

codemetapy v0.1.3

- Added a `resolve()` function that resolves nodes that only have an `@id` when such a node was previously introduced (not used internally yet)

(Released by Maarten van Gompel on 2018-05-10) https://github.com/proycon/codemetapy/releases/tag/v0.1.3

codemetapy v0.1.2

- making registry jsonld complaint
- added schema:audience property
- Work on entrypoints, defining extra context for entrypoints (codemeta/codemeta#183)
- lowercase all identifiers

(Released by Maarten van Gompel on 2018-05-02) https://github.com/proycon/codemetapy/releases/tag/v0.1.2

codemetapy v0.1.1

- Minor fix: omit empty fields, use lower case identifiers in registry

(Released by Maarten van Gompel on 2018-04-23) https://github.com/proycon/codemetapy/releases/tag/v0.1.1

labirinto v0.2.3

- Fixed collapse bug in interface
- Fix for small screen (=mobile) detection

(Released by Maarten van Gompel on 2018-06-13) https://github.com/proycon/labirinto/releases/tag/v0.2.3

labirinto v0.2.2

Minor update: now supports being served at a different base url/prefix (set BASE in `prod.env.js`)

(Released by Maarten van Gompel on 2018-06-07) https://github.com/proycon/labirinto/releases/tag/v0.2.2

labirinto v0.2.1

Minor bugfix (nothing changed except the version number, npm demands three components)

(Released by Maarten van Gompel on 2018-06-07) https://github.com/proycon/labirinto/releases/tag/v0.2.1

labirinto v0.2

- Internal refactoring to make things more modular (split things into neat components)
- Added some less-interactive endpoints that filter by interface type (`/cli` , `/lib`, `/web`, `/webservices`)
- Complete restyling, more radboud-like colours
- Headers and footer should be omitted when included in an iframe
- Better display on mobile platforms

(Released by Maarten van Gompel on 2018-06-07) https://github.com/proycon/labirinto/releases/tag/v0.2

labirinto v0.1.3

- Support for obtaining metadata registry from relative URLs
- Show error feedback when obtaining metadata registry fails
- Allow Vue development tools in production mode for now

(Released by Maarten van Gompel on 2018-06-01) https://github.com/proycon/labirinto/releases/tag/v0.1.3

labirinto v0.1.2

- Added metadata

(Released by Maarten van Gompel on 2018-05-19) https://github.com/proycon/labirinto/releases/tag/v0.1.2

labirinto v0.1.1

- First update

(Released by Maarten van Gompel on 2018-05-19) https://github.com/proycon/labirinto/releases/tag/v0.1.1

labirinto v0.1.0

First release

(Released by Maarten van Gompel on 2018-05-17) https://github.com/proycon/labirinto/releases/tag/v0.1.0

Miscellaneous

Official task and deliverable references: Dependants of others

ticcutils v0.19

[Ko van der Sloot]

- Bumped library version to 5.0
- SocketBasics:
  - invalidate sockets, when something looks not OK
- Unicode:
  - added Unicode (UTF8) aware versions of uppercase() and lowercase() added a UnicodeString splitter, analogue to the 'old' TiCC::split() functions
  - added icu:: namespace (required for ICU61 and up)
  - fixed a problem with Transcriptor rules not handled correctly.
- Small edits to fix some compiler warnings.
- fixed potential memory leaks in Trie.h and Tar.cxx
- CommandLine:
  - implemented gnu-like extensions in command-line parsing. (intermixing mass options with normal options)
- added tests for the new features/functions.

[Maarten van Gompel]

- added codemeta.json file

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.19

PICCL & TICCL

Official task and deliverable references: CLARIAH T26 [D1.3 (PICCL software); T1.4 (PICCL docs)]

PICCL v0.6.2

Minor bugfix release (debug mode was broken)

(Released by Maarten van Gompel on 2018-06-06) https://github.com/LanguageMachines/PICCL/releases/tag/v0.6.2

PICCL v0.6.1

- more verbose output from clam wrapper
- added debug option
- force en_US.UTF-8 locale in CLAM wrapper (solves LanguageMachines/ticcltools#18)
- added a test

(Released by Maarten van Gompel on 2018-06-06) https://github.com/LanguageMachines/PICCL/releases/tag/v0.6.1

PICCL v0.6

This is an important bugfix release with some new features as well:

- **New features**:
  - Added TICCL-chainer
  - Propagate alphabet file to resolver (TICCL-LDcalc)
- **Fixes**:
  - Fixed and refactored integrations tests, travis works again #16
  - Some fixes for running ticcl.nf for folia files with different extension #32 (but not extensively tested yet)
  - frog.nf couldn't find frog xml output #29
  - Use new inputclass/outputclass parameters for FoLiA-correct #34
  - ocr.nf could not find FoLiA-hocr output files #30 This release depends upon the new releases (released today) of **ticcltools** (v0.6) and **foliautils** (v0.9.2).

(Released by Maarten van Gompel on 2018-06-05) https://github.com/LanguageMachines/PICCL/releases/tag/v0.6

PICCL v0.5.3

- Bugfix release: frog.nf cannot find frog xml output #29

(Released by Maarten van Gompel on 2018-05-23) https://github.com/LanguageMachines/PICCL/releases/tag/v0.5.3

PICCL v0.5.2

- Added metadata
- Bugfix: Frog did not honour --skip option #28

(Released by Maarten van Gompel on 2018-05-19) https://github.com/LanguageMachines/PICCL/releases/tag/v0.5.2

PICCL v0.5.1

- Implemented frog selection options #25

(Released by Maarten van Gompel on 2018-04-13) https://github.com/LanguageMachines/PICCL/releases/tag/v0.5.1

PICCL v0.5

- Use CLAM 2.3 and the new external configuration file capability
- More detailed output from nextflow in webservice error.log
- Some improvements in handling quotes/spaces, but not complete yet (#24), usage of spaces in input filenames is still not supported!
- Fix in tokeniser invocation from webservice
- Compatibility with LaMachine v2 (this may break LaMachine v1 compatibility), installation instructions updated accordingly

(Released by Maarten van Gompel on 2018-04-05) https://github.com/LanguageMachines/PICCL/releases/tag/v0.5

ticcltools v0.6

Intermediate release, with a lot of new code to handle N-grams Also a lot of refactoring is done, for more clear and maintainable code. This is work in progress still.

- TICCL-unk:
  - more extensive acronym detection
  - fixed artifreq problems in 'clean' punctuated words
  - added filters for 'unwanted' characters

- added a ligature filter to convert evil ligatures
- normalize all hyphens to a 'normal' one (-)
- use a better definition of punctuation (unicode character class is not good enough to decide)
- TICCL-lexstat:
  - the 'separator' symbol should get freq=0, so it isnt counted
  - the clip value is added to the output filename
- TICCL-indexer:
  - indexer and indexerNT now produce the same output, using different strategies when a --foci files is used.
- TICCL-LDcalc: major overhaul for n-grams
  - added a ngram point column to the output (so NOT backward compatible!)
  - produce a '.short' list for short word corrections
  - produce a '.ambi' file with a list of n-grams related to short words
  - prune a lot of ngrams from the output
- TICCL-rank:

- output is sorted now
- honor the ngram-points from the new LDcalc. (so NOT backward compatible!)

- TICCL-chain: new module to chain ranked files
- TICCL-lexclean: -added a -x option for 'inverse' alphabet
- TICCL-anahash:
  - added a --list option to produce a list of words and anagram values
- added metadata file: codemeta.json

(Released by Ko van der Sloot on 2018-06-05) https://github.com/LanguageMachines/ticcltools/releases/tag/v0.6

Timbl & Mbt

Official task and deliverable references: Pre-CLARIN

mbt v3.3.2

[Ko van der Sloot]

Maintenance release:

- updated configure/build/test
- typos in man pages corrected

[Maarten van Gompel]

- Added codemeta.json

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/mbt/releases/tag/v3.3.2

mbtserver v0.12

Maintenance release:

- configure/build/test stuff
- typo in man page
- MbtServer: --debug is also a valid option

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/mbtserver/releases/tag/v0.12

python-timbl v2018.04.23

Yet another round of Mac OS X compilation fixes.. (homebrew's boost-python3 changed something)

(Released by Maarten van Gompel on 2018-04-23) https://github.com/proycon/python-timbl/releases/tag/v2018.04.
23

timbl v6.4.12

[Ko van der Sloot]

Bugfix release:

- updated usage(). Info on -G 2 option was wrong.
- changed an error message to be more clear.
- fixed building of the TeX documentation

[Maarten van Gompel]

- Added codemeta.json metadata

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.12

timblserver v1.12

Maintenance release:

- updated and improved configure/build
- typo's corrected
- some code cleanup
- better error messages.

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/timblserver/releases/tag/v1.12

Ucto

Official task and deliverable references: CLARIAH T55 [D1.1 (software), D1.2 (docs)]

python-ucto v0.4.7

- Fixed issue #4 and compatibility release for ucto v0.13

(Released by Maarten van Gompel on 2018-05-16) https://github.com/proycon/python-ucto/releases/tag/v0.4.7

ucto v0.13.2

Bug fix release:

- uctodata is mandatory. So don't install default rules anymore

(Released by Ko van der Sloot on 2018-05-17) https://github.com/LanguageMachines/ucto/releases/tag/v0.13.2

ucto v0.13.1

Bug fix release:

- configure now finds out the location of the uctodata files. should make it work on Mac systems too

(Released by Ko van der Sloot on 2018-05-17) https://github.com/LanguageMachines/ucto/releases/tag/v0.13.1

ucto v0.13

[Ko van der Sloot]

- improved configure/build/test
- added a --split option
- fixed -P option
- removed -S option (never used, and only half implemented)
- added a --add-tokens option, to add special tokens for the default language
- generally use the icu:: namespace
- added more tests
- fixed uninitialized variable.
- added code to use an alternative search-path for uctodata

[Maarten van Gompel]

- added codemeta.json

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/ucto/releases/tag/v0.13

uctodata v0.7.1

Bug fix release:

- install some datafiles originally provided by 'ucto'

(Released by Ko van der Sloot on 2018-05-17) https://github.com/LanguageMachines/uctodata/releases/tag/v0.7.1

uctodata v0.7

[Ko vd Sloot]

- tokconfig-nld-historical: typo in rule
- updated all languages with new ABBREVIATION and NUMBER-ORDINAL rules: = accommodate ABBREVIATIONS within brackets. = avoid needless backtracking in NUMBER-ORDINAL

[Maarten van Gompel]

- Apparent bug in Italian config

(Released by Ko van der Sloot on 2018-05-16) https://github.com/LanguageMachines/uctodata/releases/tag/v0.7

uctodata v0.6

several fixes for problems addressed in https://github.com/LanguageMachines/ucto/issues/46 Notes:

- the suffix problems were already addressed in 0.5
- the colon problem is not addressed. Do we need REVERSE-SMILEY?

(Released by Ko van der Sloot on 2018-04-04) https://github.com/LanguageMachines/uctodata/releases/tag/v0.6