

Software Release and Progress Report

Start date: 2019-03-18 End date: 2019-06-25

Short Summary

At the end of last period, FoLiA v2 was released, which signified a major milestone promised for CLARIAH-PLUS that adds provenance support, rectifies some misnomers from the past, and adheres to a stricter paradigm. This period has seen development on various tools and libraries, making them compatible with the latest version. We released a FoLiA v2 compliant version of libfolia, and of Frog, ucto, and other software relying on libfolia. We also released a new version of FLAT, which aside from all the FoLiA v2 improvements, now also integrates the new CONLL-U to FoLiA converter, allowing uploads in this popular format. Moreover, the ReStructuredText to FoLiA converter is also integrated, providing a good entry point for uploading structured text with markup information.

Certain components that used to be part of PICCL have been split up into separate projects (without harming the functionality of PICCL), allowing for better modularity and maintainability. This concerns the Nederlab pipeline for linguistic enrichment of historical dutch, and aNtiLoPe, workflows that wrap around Frog and ucto.

Catch-up work has been done for Nederlab, which will eventually also be relevant for CLARIAH-PLUS considering the follow-up task we were granted. The catch-up work resulted in two new tools that are of more general interest:

- a new TEI to FoLiA converter as part of the folia tools, and is derived from prior work by Jesse de Does (INT).
- wikiente, a named entity recognition tool and linker, which is a simple but
- fully FoLiA-capable wrapper around DBPedia Spotlight.

Further activities carried out in this period:

- Metadata mapping SoNaR and CGN (N. Oostdijk); Report to be written
- Application for Knowledge Centre for Atypical Communication submitted to CLARIN ERIC in collaboration with MPI's TLA and approved in June
- H. van den Heuvel participated in ParlaFormat (23-24 May), a workshop on TEI-guidelines for parliamentary speech: <https://www.clarin.eu/event/2019/parlaformat-workshop>
- H. van den Heuvel has been invited for a presentation about secure management of sensitive data at <https://www.tapas-etn-eu.org/events/te3>
- H. van den Heuvel was involved in preparations for DH2019 workshop on Transcription Chain: <https://www.eventbrite.co.uk/e/dh2019-workshop-the-case-of-interview-data-tickets-59161515686>
- I. Hendrickx worked on experiments & bug testing of frog modules, mainly morphological analysis
- Official creation of the Stichting Open Spraaktechnologie (10 May 2019)

Done in April:

- e-WGD updated with text and further curated dictionary files
- Papers for CLARIN AC submitted:
 1. GDPR & data use cases sensitive data
 2. Transcription Chain
- ASR webservice for English set up (https://webservices-1st.science.ru.nl/eng_ASR)
- G2P webservice for English created (<https://webservices-1st.science.ru.nl/g2pservice>)

CLAM

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

clam v2.4.5

- Added support for CUSTOM_VIEWERS configuration directive
- silently accept None as positional parameter in various constructors to omit a certain value
- Fixes on info page

(Released by Maarten van Gompel on 2019-06-07) <https://github.com/proycon/clam/releases/tag/v2.4.5>

FLAT

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

flat v0.8.0

- Updated for FoLiA v2, now uses the new foliapy library instead of predecessor pynlpl.formats.folia
 - Implemented provenance support #140
 - Annotator information now shows multiple levels (nested processors)
 - Older FoLiA documents are automatically updated by the new FLAT
 - Various major fixes for adding syntactic annotation (#138)
 - Fixes in visualisation of syntax trees (#138)
 - Implemented hidden word support (#141)
 - Adding support for appending/prepending hidden words in the annotation editor #145
 - Added a `script/` dir instead of having scripts in `style/`, which was a misnomer
 - Implemented visualisation of relations #84 (a)
 - Disable edit forms for corrections and alternatives in the default full configuration
 - Show warning when loading set definition fails #147
 - No write permission check needed for public uploads #146
 - Adding default converters for ReStructuredText (#74) and CONLL-U (#144), through foliatools
 - Stylistic improvements
 - Improved the top menu, various UI style and usability improvements
 - Expanded footer, added documentation links
- Important notes:**
- This release requires some changes to your `settings.py` if you use a custom one! Please inspect the diffs.
 - This release produces only FoLiA v2 documents (it can still read older documents but will automatically upgrade those)

(Released by Maarten van Gompel on 2019-04-30) <https://github.com/proycon/flat/releases/tag/v0.8.0>

foliadocserve v0.7.0

- Updated for FoLiA v2, now uses the foliapy library instead of predecessor pynlpl.formats.folia
 - Implemented provenance support
 - Older FoLiA documents are automatically updated by the new foliadocserve
 - Implemented hidden word support (proycon/flat#141)
 - Allow set definition failures and propagate failed set definitions to output (proycon/flat#147)
- Important note:** This release produces only FoLiA v2 documents (it can still read older documents but will automatically upgrade those)

(Released by Maarten van Gompel on 2019-04-30) <https://github.com/proycon/foliadocserve/releases/tag/v0.7.0>

FoLiA

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

folia v2.1.0

- Set definitions can now define constraints on the combination of classes/subsets used. #50
- Added a `t-ref` element as a text markup counterpart to `ref`
- Added flexibility for structure elements (more nesting options)

(Released by Maarten van Gompel on 2019-05-22) <https://github.com/proycon/folia/releases/tag/v2.1.0>

folia v2.0.3

- Added extra examples and fixes some other
- Added set definition for Penn Treebank and Spacy Named Entities (thanks to @ErkanBasar)
- Added ISO-639-3 set definition (and a link for backward compatibility) (LanguageMachines/ucto#67)

- span roles do not carry independent processor information
- reducing complexity: no attributes except ID allowed on annotation layers anymore

(Released by Maarten van Gompel on 2019-04-26) <https://github.com/proycon/foia/releases/tag/v2.0.3>

foia v2.0.2

- added attributes src and format to processor element
- added some examples
- added simple default set for phonetics
- various documentation updates
- various fixes
- annotator/processor/annotatortype attribute on suggestion element is deprecated (only confidence and n are still allowed)
- [specification] added a hidden property, used by the Hiddenword class

(Released by Maarten van Gompel on 2019-04-16) <https://github.com/proycon/foia/releases/tag/v2.0.2>

foia v2.0.1

- Minor fix for hyphenation annotation
- README update and some documentation fixes

(Released by Maarten van Gompel on 2019-03-20) <https://github.com/proycon/foia/releases/tag/v2.0.1>

foliapy v2.1.4

- Documentation updates
- Add a fix to the library to handle an automatically fix certain invalid FoLiA where an explicit processor attribute is missing on an annotation and no default can be assigned due to there being multiple possible processors/annotators (need to be explicitly enabled)

(Released by Maarten van Gompel on 2019-06-12) <https://github.com/proycon/foiapy/releases/tag/v2.1.4>

foliapy v2.1.3

- migrated annotator2processor from foliatools/foiaupgrade to the library; now handle 'annotator' keyword argument in a backward-compatible manner on element instantiation (fix for proycon/foia#73)
- fix serialisation of annotator attribute (proycon/foia#73)

(Released by Maarten van Gompel on 2019-06-06) <https://github.com/proycon/foiapy/releases/tag/v2.1.3>

foliapy v2.1.2

Bugfix release:

- Fix in set definition parsing

(Released by Maarten van Gompel on 2019-05-29) <https://github.com/proycon/foiapy/releases/tag/v2.1.2>

foliapy v2.1.1

- prevent creating invalid XML NCName for processor IDs

(Released by Maarten van Gompel on 2019-05-29) <https://github.com/proycon/foiapy/releases/tag/v2.1.1>

foliapy v2.1.0

- Updated for FoLiA v2.1.0
 - Implemented TextMarkupReference (t-ref)
 - Wider range of nesting of structure elements
 - Implemented a constraint parser for deep validation against set definition that specify constraint (proycon/foia#50)

- Various fixes (e.g. for aliases, no longer lose annotator on text declarations, parsing of linebreaks with IDs, fix in test framework)

(Released by Maarten van Gompel on 2019-05-22) <https://github.com/proycon/foliapy/releases/tag/v2.1.0>

foliapy v2.0.8

- Adapted for FoLiA v2.0.3
- Added `Document.erase()` method to erase annotation types (#3)
- fixes for FoLiA v1 compatibility and conversion
- More flexible handling in case of missing (404) set definitions (proycon/flat#147)
- Implemented JSON serialiation for relations
- Various FQL fixes
- Various other fixes

(Released by Maarten van Gompel on 2019-04-26) <https://github.com/proycon/foliapy/releases/tag/v2.0.8>

foliapy v2.0.7

Bugfix release, implements FoLiA v2.0.2, contains various fixes. Upgrade recommended.

- various FQL fixes, adapted to new PROCESSOR statement (required for FLAT/foliadocserve)
- fix in `next()` method for querying multiple classes at once
- expanding Provenance data (processors) with a `src` and `format` attribute
- handle `format` attribute on declarations (implementation was missing)
- implemented JSON serialisation of provenance data
- Use `AbstractWord` interface for `Morpheme` and `Phoneme` as intended
- Added `Processor.insert()` and `Provenance.insert()`
- Ignore annotator and annotortype assignments if a processor is already set

(Released by Maarten van Gompel on 2019-04-16) <https://github.com/proycon/foliapy/releases/tag/v2.0.7>

foliapy v2.0.6

Minor update (minor refactoring)

(Released by Maarten van Gompel on 2019-04-01) <https://github.com/proycon/foliapy/releases/tag/v2.0.6>

foliapy v2.0.5

Another round of significant fixes and refactoring in the wake of the v2 release, mostly to do with backward compatibility. One notable change: this release changes the default value for `set` in a lot of functions. `Set` is now `False` when it means *any set* (useful in querying functions), it is `None` to specifically mean *no set/setless* and of course a string with ideally (but not necessarily) a URL in all other cases.

(Released by Maarten van Gompel on 2019-04-01) <https://github.com/proycon/foliapy/releases/tag/v2.0.5>

foliapy v2.0.4

Various more fixes and refactoring in the wake of the FoLiA v2 release:

- Fixed autodeclaration of setless elements

(Released by Maarten van Gompel on 2019-03-23) <https://github.com/proycon/foliapy/releases/tag/v2.0.4>

foliapy v2.0.3

Small fix for backward compatibility in serialisation of FoLiA v1

(Released by Maarten van Gompel on 2019-03-22) <https://github.com/proycon/foliapy/releases/tag/v2.0.3>

foliapy v2.0.2

Updated documentation and minor fixes for FoLiA 2.0.1

(Released by Maarten van Gompel on 2019-03-20) <https://github.com/proycon/foliapy/releases/tag/v2.0.2>

foliatools v2.1.7

- [foliamerge] fixing breakage with new library (reported in proycon/flat#148)
- [foliaupgrade] exit code was wrong

(Released by Maarten van Gompel on 2019-06-17) <https://github.com/proycon/foliatools/releases/tag/v2.1.7>

foliatools v2.1.6

Previous release was broken, fixed now.

(Released by Maarten van Gompel on 2019-06-12) <https://github.com/proycon/foliatools/releases/tag/v2.1.6>

foliatools v2.1.5

- added language detection tool (folialangid) based on lang.py
- --fixunassignedprocessor didn't work properly yet

(Released by Maarten van Gompel on 2019-06-12) <https://github.com/proycon/foliatools/releases/tag/v2.1.5>

foliatools v2.1.4

- [foliaupgrade] Added --fixunassignedprocessor flag to fix a known issue with some FoLiA documents

(Released by Maarten van Gompel on 2019-06-12) <https://github.com/proycon/foliatools/releases/tag/v2.1.4>

foliatools v2.1.3

- [tei2folia] prevent creating a text content element with only one (or more) non-breaking spaces (was not handled by normalize-space() as expected)
- [foliaupgrade] Migrated part of the folia upgrader to the foliapy library itself

(Released by Maarten van Gompel on 2019-06-06) <https://github.com/proycon/foliatools/releases/tag/v2.1.3>

foliatools v2.1.1

Minor fixes for conllu2folia and tei2folia

(Released by Maarten van Gompel on 2019-05-27) <https://github.com/proycon/foliatools/releases/tag/v2.1.1>

foliatools v2.1.0

- Implemented a tei2folia tool, implementing a *subset* of TEI (tested on the DBNL collection by the KB). Loosely based on prior work by @JesseDeDoes.

(Released by Maarten van Gompel on 2019-05-22) <https://github.com/proycon/foliatools/releases/tag/v2.1.0>

foliatools v2.0.7

- added a new foliaerase tool that can remove certain annotation types
- foliaupgrade: various fixes
- rst2folia: adding a wrapper for FLAT integration
- conllu2folia: adding a wrapper for FLAT integration

(Released by Maarten van Gompel on 2019-04-26) <https://github.com/proycon/foliatools/releases/tag/v2.0.7>

foliatools v2.0.6

- foliaupgrade: refactoring and updates
- foliaupgrade: set src and format attributes on pre-upgraded version
- foliavalidator: add proof of validation (not cryptographic) to provenance chain (only when --output is chosen)
- folia2html: updated for FoLiA v2

(Released by Maarten van Gompel on 2019-04-16) <https://github.com/proycon/foliatools/releases/tag/v2.0.6>

foliatools v2.0.5

- foliacat: More provenance data for foliacat
- foliacat: Compatibility for serialising older FoLiA versions
- Added conllu2folia tool (#1)
- implemented foliaupgrade, convert old-style annotator data to provenance

(Released by Maarten van Gompel on 2019-04-01) <https://github.com/proycon/foliatools/releases/tag/v2.0.5>

foliatools v2.0.3

Minor refactoring of foliacat and better version reporting

(Released by Maarten van Gompel on 2019-03-22) <https://github.com/proycon/foliatools/releases/tag/v2.0.3>

foliatools v2.0.2

- Fix for foliacat when dealing with older versions

(Released by Maarten van Gompel on 2019-03-19) <https://github.com/proycon/foliatools/releases/tag/v2.0.2>

foliautils v0.12

Released for FoLiA 2.0

(Released by Ko van der Sloot on 2019-05-29) <https://github.com/LanguageMachines/foliautils/releases/tag/v0.12>

foliautils v0.11

- Updated and added some tests
- started moving common code to a separate file and build a library (libfoliautils)
 - hemp detection is one of them
- FoLiA-stats:
 - added possibility to read a list of directories + file-names to process into separate output directories. (could be generalized to other programs)
 - better hemp detection
- FoLiA-correct:
 - use same hemp detection as FoLiA-stats
- FoLiA-abby:
 - support more flavors
- FoLiA-clean:
 - avoid removing the last remaining text on nodes
 - cleaning of tokenization now works

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/foliautils/releases/tag/v0.11>

libfolia v2.1

Bug fixes and enhancements:

- provenance:
- added 'generate_id' attribute with 'auto()' and 'next()' values
- some code improvements

- bugs:
- When using the FoLiA-engine, we have to save the ORIGINAL annotationdefaults, and use these when parsing.

(Released by Ko van der Sloot on 2019-06-19) <https://github.com/LanguageMachines/libfolia/releases/tag/v2.1>

libfolia v2.0

This is a major release.

- Supports the new FoLiA 2.0 features:
 - provenance support
 - more stricter checking on annotation declarations
 - added the new TextMarkupReference class
 - supports Hidden Words.
 - All structure elements can have the 'space' attribute
 - support for groupannotations
 - many more.
- API and ABI breaches:
 - library version bumped to version 10
 - many functions are renamed
 - the text() functions have an ENUM parameter now to select for STRICT, RETAIN or HIDDEN
- bug fixes
 - support for xlink: improved
 - there as a rare mixup between nodes and nodes in the folia::Engine
 - all nodes get a 't' attribute now on serializing.
 - reading Extrenal FoLiA could get in an endless loop
- code refactoring and cleanup

(Released by Ko van der Sloot on 2019-05-29) <https://github.com/LanguageMachines/libfolia/releases/tag/v2.0>

libfolia v1.16

Stabilizing release for folia1.5. The next release will support the new FoLiA 2.0 Changes:

- renamed folia::Processor to folia::Engine
- extended and improved Engine code a lot
- avoid spurious newline on Document output
- Will read and ignore some FoLiA 2.0 additions
- numerous small additions and fixes
- make sure that the XmlParser uses the HUGE model everywhere

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/libfolia/releases/tag/v1.16>

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 v0.18.1

Bug fix release. "tabbed" output contained 1 tab to much when --skip=a was specified

(Released by Ko van der Sloot on 2019-06-19) <https://github.com/LanguageMachines/frog/releases/tag/v0.18.1>

frog v0.18

Bug fixes and enhancements:

- provenance uses new 'generate_id' option in libfolia:processor
- solved problems when frogging partly tokenized FoLiA
- solved problems when processing with --skip=t

- small improvement in compound detection (still more to do...)

(Released by Ko van der Sloot on 2019-06-19) <https://github.com/LanguageMachines/frog/releases/tag/v0.18>

frog v0.17

- some bug fixes
 - trust the tokenizer to get the default language
 - don't stumble upon empty sentences introduced by a non-breaking-space
 - provenance data is added for all the modules

(Released by Ko van der Sloot on 2019-05-29) <https://github.com/LanguageMachines/frog/releases/tag/v0.17>

frog v0.16

This is the last release using pre FoLiA 2.0 It includes a total rework of the Frog Internals, aiming at better maintainability and hoping for a speedup and a smaller memory footprint. This work will continue in the upcoming release for FoLiA 2.0 Major Changes:

- total rework. Not using a FoLiA document as the internal datastructure anymore but a FrogData structure.
- use folia::engine for all FoLiA processing
- -Q option is NOT supported anymore. It was unreliable anyway
- builds on the newest ucto versions only
- fix for <https://github.com/proycon/LaMachine/issues/135> <https://github.com/LanguageMachines/frog/issues/66>
- handles some corner cases in FoLiA better
- lots of code cleanup
- numerous small fixes (e.g. in NER and MBMA results)
- improved working of --languages option
- avoid invalid FoLiA: <https://github.com/LanguageMachines/frog/issues/60>
- fixed memory leaks
- better handling of weird FoLiA

[Maarten van Gompel]

- added skeleton for new Frog documentation

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/frog/releases/tag/v0.16>

frogdata v0.17

[Ko van der Sloot]

- small fix in mblem data, and retrained
- updated NER gazetteer files
- added cgn constraints file

[Maarten van Gompel]

- config/nld/ners.known, config/nld/plaatsen-nld.ner: added dutch cities/villages, may overlap with geonames but this is more specific
- config/nld/ners.known, config/nld/straten-nld.ner: Added dutch streets gazetteer data (extracted from OpenStreetMap)
- Added middle dutch model (INL/nederlab-linguistic-enrichment#15) in /dum/ subdir

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/frogdata/releases/tag/v0.17>

python-frog v0.4.0

- Compatibility with Frog v0.16
- Fixed for adaptations in FrogAPI (#11) and FoLiAPy library

(Released by Maarten van Gompel on 2019-05-24) <https://github.com/proycon/python-frog/releases/tag/v0.4.0>

toad v0.5

- some bug fixes
- added --eos and --postags options to froggen.
- cleaned up ICU namespaces
- froggen prefers higher frequency POS tags over lower, when ambiguous
- added a 'bootstrap' function to 'nergen' to enrich a data file with gazetteer info BEFORE training.

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/toad/releases/tag/v0.5>

LaMachine

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

LaMachine v2.8.0

- Added aNtiLoPe and nederlab-pipeline, (both split off from PICCL), they are nabled by default now as part of the languagemachines-pipelines package
- Adding DBPedia spotlight #143
- Added langid.py
- Added fasttext
- Added WikiEnte
- Fix for spacy #148
- Fixed proper validation in case disk size wasn't a number
- opening port 9999 by default in vagrant for user-defined purposes v2.7.0 (forgot to release formally here):
- New FoLiA v2 based releases for Ucto, frog, libfolia, FoLiA-utils
- [CentOS7] compile SQLite from scratch because system version is too old (django complains, affects FLAT)
- attempted fix for redhat (CentOS)
- fix: flair requires python >3.6, do not install on older distributions
- ansible 2.8 compatibility; added note for ansible upgrade error #144

(Released by Maarten van Gompel on 2019-06-14) <https://github.com/proycon/LaMachine/releases/tag/v2.8.0>

LaMachine v2.6.0

- Added spacy2folia
- Added Freeling #123
- Added eng_ASR #139 (still to be properly tested)
- Adapted for FLAT v0.8.0 and foliadocserve v0.7.0
- Fix in npm upgrade
- This version also holds the final camera ready paper on LaMachine as published in the CLARIN 2018 proceedings

(Released by Maarten van Gompel on 2019-04-30) <https://github.com/proycon/LaMachine/releases/tag/v2.6.0>

Miscellaneous

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

ticcutils v0.21

NOTE: this is hopefully the last release that needs boost::regex The next releases will depend on competent C++11 compilers

- small fixes in bootstrapping and building phase
- neater output from CL_Options::toString()
- improved Timer clas. Added + operator
- added more self-tests
- zipper: Simplified code and better error messages

(Released by Ko van der Sloot on 2019-05-14) <https://github.com/LanguageMachines/ticcutils/releases/tag/v0.21>

wikiente v0.1.0

First release

(Released by Maarten van Gompel on 2019-06-12) <https://github.com/proycon/wikiente/releases/tag/v0.1.0>

Nederlab

Project/task/deliverable references: Nederlab

nederlab-pipeline v0.8.0

This is a first release after the split from PICCL. Introduces a renewed Nederlab pipeline consisting of the following components (most can be enabled/disabled when needed):

- Conversion from TEI to FoLiA through `tei2folia` (part of Folia tools, this replaces the old scripts by Erik Tjong Kim Sang.
- FoLiA v2 compatible, stricter and explicit validation and explicit provenance information.
- Tokenisation through `ucto`
- Language detection using `folialangid` (part of Folia tools)
- Linguistic enrichment through Frog (part-of-speech tagging, lemmatisation, named entity recognition)
- Named Entity Linking (after another recognition stage) is handled by `wikiente`, which communicates with the DBpedia Spotlight service, and replaces the older `foliaentity` tool by Erwin Komen.

(Released by Maarten van Gompel on 2019-06-14) <https://github.com/proycon/nederlab-pipeline/releases/tag/v0.8.0>

PICCL & TICCL

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

PICCL v0.8.0

- **Several workflows that used to be part of PICCL have been split off into separate projects now**, this concerns:
 - The nederlab pipeline for enrichment of historical dutch, they are now in <https://github.com/proycon/nederlab-pipeline>
 - The frog, ucto and folia validation pipelines <https://github.com/proycon/aNtiLoPe>, PICCL depends on this new aNtiLoPe project now
 - This is an organisational change in favour of more modularity, clarity and better maintainability, it does not affect the functionality or installation of PICCL!
- Allow unsetting `flaturl` in external yaml configuration to disable flat viewers (`proycon/clam#75`)
- Propagate existing input `textclass` option to PICCL and assume a default of 'current' (rather than 'OCR') if OCR is skipped (`#48`) and change TICCL inputclass default to 'current' instead of 'OCR' when dealing with FoLiA input
- Delete zero byte input files prior to FoLiA-correct (artefact of earlier patchy error ignore strategy) `#49`

(Released by Maarten van Gompel on 2019-06-14) <https://github.com/LanguageMachines/PICCL/releases/tag/v0.8.0>

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

(no releases this period)

Ucto

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

python-ucto v0.5.0

- Compatibility release for ucto v0.15

(Released by Maarten van Gompel on 2019-05-24) <https://github.com/proycon/python-ucto/releases/tag/v0.5.0>

ucto v0.17

Bug-fix release:

- solved problems when tokenizing (partly-)tokenized FoLiA (but this is a very complicated situation. Might need more work)
- solved problems with --passthru on FoLiA
- avoid empty lines in FoLiA output
- use the new generate_id attribute for provenance/processors
- added more tests **KNOW PROBLEM:** On TravisCI/MacOSX some tests fail for unclear reasons.

(Released by Ko van der Sloot on 2019-06-19) <https://github.com/LanguageMachines/ucto/releases/tag/v0.17>

ucto v0.16

Major release supporting FoLiA 2.0

- bug fixes for:
 - empty sentences in FoLiA introduced by NonBreakingSpace
 - provide provenance data

(Released by Ko van der Sloot on 2019-05-29) <https://github.com/LanguageMachines/ucto/releases/tag/v0.16>

ucto v0.15

Stabilizing release for pre FoLiA 2.0

- uses new folia::engine to process FoLiA
- lots of refactoring and cleanup
- some small bug fixes
- added tests for corner cases in FoLiA
- improved TextCat handling and debugging

(Released by Ko van der Sloot on 2019-05-15) <https://github.com/LanguageMachines/ucto/releases/tag/v0.15>