Final Report CLARIAH-core tasks Radboud University

In general, with respect to the development of software we would like to note that the way that the team operates is not entirely in sync with the original plan with milestones. The key difference is that most tasks have started before their planned start date. We are continuously working on our software stack, and have started tasks as soon as developers were in place (rather than starting them according to our plan in a non-overlapping order). In essence all our work is in 'on track'. All software is functional, documented, and continuously available (https://github.com/proycon/LaMachine). we welcome tests of recent versions of our software and are completely open to contributions, bug reports, wish lists and suggestions for improvements. All of our software is Open Source (GPL v.3).

This implies that on the one hand all tasks have started and can even be seen as finished to a certain extent, while on the other hand work on tasks that were planned to have ended still continues, due to our intention to continue implementing improvements as time allows. For our developers, this means a type of distributed attention over many different goals that has their preference over a more piecemeal division of time. On a regular basis, with at least weekly meetings between subsets of team members, changes in activity plans are discussed and agreed upon. Usually, short 'sprints' are defined on one of the software packages, where one or two developers focus more on one goal for a few weeks. Ad-hoc activities may also be caused by bug reports or feature requests from outside the team.

All NLP software mentioned below (Frog, Ucto, CLAM, FLAT, FoLiA libraries, optionally TICCL) is available under one meta-package called LaMachine. The code, including instructions on how to build virtual machines (e.g. Vagrant, VirtualBox), a Docker instance, or a virtual environment of LaMachine and all packaged software on the most common operating systems (Windows, macOS, Linux), can be found here: https://github.com/proycon/LaMachine

T10: Data Curation Service

A. Metadatacollection records

We have worked at the introduction of more elaborate and consistent metadata records for collections of (linguistic) data resources in the CLARIN/CLARIAH context. For this purpose we designed and implemented a CMDI profile. This profile is stored in CLARIN's Component Registry and can be viewed at https://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/1.1/profiles/clarin.eu:cr1:p_1493735943947/xml . We validated the profile in a first pilot in which we populated the profile for 45 Dutch language resources. Given the complexity of the profile and special purpose requirements we developed our own interface for creating, editing, listing, copying and exporting descriptions of metadata collection records (https://applejack.science.ru.nl/collbank/).

With the help of CLARIN we build our OAI-PMH bridge to release the resulting CMDI metadata records and make them available in CLARIN's Virtual Language Observatory (VLO). Via the interface and the harvesting bridge we have published the metadata records of 60 collections.

More information can be found in Van den Heuvel, Komen, Oostdijk (2018).

B. Data curation:

Data curation efforts have mainly concentrated on the regional dialect dictionaries for Brabant, Limburg and Gelderland (WBD, WLD, WGD). Many errors were corrected in the database records. From our own matching CLST realized search portals for these dialect dictionaries which were launched in 2016, and 2017, see https://e-wbd.nl/ and http://e-wld.nl/ . The search page e-wgd.nl is under construction and will be launched in 2019.

The LMF structure has been adopted to new insights (see Van den Heuvel, Sanders & Van der Sijs, 2016) permitting the inclusion of other regional dialect dictionaries (such as WALD and WOD), and local dialect dictionaries. A close collaboration with the INT has been established for WBD and WLD in order to integrate these dictionaries in the DSDD project of the University of Gent (see https://www.clarin.eu/sites/default/files/CLARIN2018_Poster-9_Paper-73_VanKeymeulen-Chambers-DeTier-deDoes-Depuydt-Schoonheim-Vandenberghe-Hellebaut.pdf). An important element is here that all entries at dialect word level obtained a unique ID.

Nelleke Oostdijk gave a presentation about corpus creation and preparation at the Provenance Workshop at DANS in September 2018: https://clariah.nl/files/presentations/provenance/00-CLARIAH-Provenance_Workshop.pdf

A task scheduled for 2019 is the linking of CGN and SoNaR metadata in OpenSoNaR+.

Publications:

Heuvel, H. van den, Oostdijk, N., Sanders, E., De Lint, V. (2015) Data curations by the Dutch Data Curation Service. Overview and future perspective. In: Odijk, J. (2015): Selected Papers from the CLARIN 2014 Conference, October

24-25, 2014, Soesterberg, The Netherlands. Linköping Electronic Conference Proceedings, 116, pp. 54-62. ISBN: 978-91-7685-954-4. http://www.ep.liu.se/ecp_article/index.en.aspx?issue=116;article=005

Heuvel. H. van den, Sanders, E.P., Sijs, N. van der (2016) Curation of Dutch Regional Dictionaries. In: Proceedings LREC 2016, Portorož, 23-28 May 2016. pp. 3249-3255.

Hout, R. van, Sijs, N. van der, Komen, E., Heuvel, H. van den (2018) A Fast and Flexible Webinterface for Dialect Research in the Low Countries. In Proceedings LREC 2018, Miyazaki, 7-10 May 2018, pp. 3617-3621.

Heuvel, H. van den, Komen, E., Oostdijk, N. (2018) Metadata Collection Records for Language Resources. In Proceedings LREC 2018, Miyazaki, 7-10 May 2018, pp. 1282-1288.

T70: RDM Guidelines

Our task to establish RDM guidelines has been approached from various angles. Henk van den Heuvel as research data officer of the Faculty of Arts at Radboud University was involved in establishing the Research Data Management policy for the faculty with a clear view on the type of research data that is relevant for CLARIAH being language and humanities related data. This has resulted in a number of relevant websites that are relevant for CLARIAH partners too:

- https://www.ru.nl/rdm
- https://www.ru.nl/rdm/vm/checklist-rdm/
- https://www.ru.nl/facultyofarts/research/research-data-management/
- https://www.radboudnet.nl/letteren/onderzoek/regelingen-werkprocessen/research-data-management/

Furthermore the GDPR which came into effect in May 2018 had an important impact on collecting, processing and sharing research data that includes person data. Person data are not only present in the metadata of many data collections but also at the heart of it (audio and video recordings). As a result of a questionnaire sent out to relevant CLARIAH partners in 2018 it appeared that institutes have not published binding guidelines and/or are not willing to share these (at this stage). Similarly, Radboud University published the webpages below on GDPR but with respect to research data an official policy has not been released yet.

- https://www.radboudnet.nl/privacy/english/protection-personal-data/personal-data/
- https://www.radboudnet.nl/privacy/english/use-personal-data/purpose-and-associated-purpose/
- https://www.radboudnet.nl/privacy/english/use-personal-data/faq/

Further, with respect to sensitive data Nelleke Oostdijk gave a presentation at the EUDAT workshop on sensitive data in Porto in January 2018.

CLARIN ERIC has installed a Legal Issues Committee (CLIC - https://www.clarin.eu/governance/legal-issues-committee) Henk van den Heuvel joined the special session of this workgroup at the Annual Conference in Pisa in October 2018, and the workshop in Vilnius in December 2018 (Hacking the GDPR).

Furthermore, he co-organised two CLARIN workshops of the DELAD taskforce. DELAD (=SHARED in Swedish) is an initiative to establish a digital archive of disordered speech and share this with interested researchers. DELAD seeks collaboration with CLARIN to this end. The first workshop was held in Cork in November 2017 and the second in Utrecht in January 2019. In the workshop in Utrecht CLIC members were invited to present their views on collecting processing and sharing corpora of disordered speech.

Finally, a new European infrastructure project named SSHOC started in 2019 in which Task 5.4 is concentrated on accessibility of sensitive data. Since CLST is actively involved here further results that are also relevant for CLARIAH may be expected in terms of guidelines, data and tools.

Relevant websites showing our activities:

- https://eudat.eu/sensitive-data-workshop
- https://www.clarin.eu/tags/clic
- https://www.clarin.eu/blog/report-clarin-delad-workshop
- https://www.clarin.eu/blog/clarin-workshop-delad-database-enterprise-language-and-speech-disorders
- https://www.sshopencloud.eu/

CLST being a CLARIN T-centre we realized Trusted CLARIN authentication arranged for OpenSoNaR+ & FAME search portals/pages. To have this facility is important for safeguarding data behind search portals now, and for later if CLST as a CLARIN data centre were to store sensitive data (which will be followed up in CLARIN-PLUS).

T21: CLAM

- Website: https://proycon.github.io/clam
- Source Repository: https://github.com/proycon/clam

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Ongoing development and maintenance: **accomplished**
- Updated Documentation: accomplished (completely renewed)
- Port backend to Python 3, web.py -> flask: **accomplished** (ahead of schedule prior to the official start)
- Support for users and service providers: **accomplished**
- OAuth reimplementation: **accomplished**
- Connecting CLAM services in workflow systems: to this end, support has been given in the WP3 VRE project, no separate initiatives have been started

T22: Frog

- Website: https://languagemachines.github.io/frog
- Source Repository: https://github.com/LanguageMachines/frog

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Ongoing development and maintenance: **accomplished**
- Frog Documentation is converted from a static PDf version of the documentation to an online website: (https://frognlp.readthedocs.io)
- Dependency parser in C++: **accomplished**
- Integration with Alpino (as in T-Scan): **not performed**, however, a separate Alpino-to-FoLiA converter is available and integrated in the CLAM-based Alpino webservice we host

Components of Frog have been refactored or reprogrammed (e.g. the dependency parser was rewritten in C++ from Python, speeding up parsing) and a long list of wishes and bugs were implemented and fixed. Work on Frog is continuous.

A major Frog update went into MBMA and MBLEM, the memory-based components responsible for morphological analysis and lemmatization, have been qualitatively improved. These tools, based on the CELEX and eLex lexical databases, generated errors due to inconsistencies in the original data. Morphological analysis is now also 'complete' in the sense that it produces fully nested morphological analyses.

We worked with Frog in a number of case studies. Frog was used as a test case in the Master course Text and Multimedia Mining from the Faculty of Arts, Radboud University. We ran experiments with Frog within case study with the KB with the 'ICT with Inductry 2019' Workshop. And Frog was investigated as part of a student internship where the student worked on enhancing Frog NER with more fine-grained categories using Wikipedia.

We performed an evaluation of the modules in Frog. The optimal evaluation of the Frog modules is performed by using a new and unseen test set. The developers of Frog are not suited to create the annotation for a new test set as they might be biased by the implementation decisions in Frog and will be inclined to follow those. However, finding independent test sets was not always possible, and we used 10-fold cross validation on training material to get a performance indication for the POS, lemmatizer and Chunker. The lemmatizer is trained on the CELEX word list. We evaluated all modules using tenfold-cross validaiton on the training material of the POS-tagger and chunker consisting of 11.133K tokens, 998.796 sentences. The POS tagger obtained an accuracy of 95.6 on the finegrained CGN-tags and 97.97% on the 12 main CGN-tags. The chunker acchieved an accuracy of 92.5 and the lemmatizer 95.85%. Do note that these results are biased over-optimistic results as part of this data was automatically labeled.

For the NER module we found several new unseen test sets: the CLIN26 shared task data set, a dataset of Dutch parlementary documents (Jonkers, 2016), a data set of historical documents from the KB and we had an intern student who created and annotated a data set of Wikipedia pages for NER labels. Overall, the NER module does not achieve very high results, between 25% and 60% F-scores. Names are often domain dependent, and also time dependent. As Frog is trained on an annotated data set of mostly news articles, many of the names are no longer relevant.

We presented the outcomes of the Frog evaluation as a poster in the CLIN 29 conference in Groningen, 30 March 2019. The Frog documentation still needs to be updated with these evaluation results.

T23: Frog Generator (Toad)

- Source Repository: https://github.com/LanguageMachines/toad

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Frog generator: training modules to infer a complete Frog from data (any language): **accomplished**; for named entity recognition, part-of-speech tagging, lemmatisation
- Documentation: accomplished

Froggen (for Frog Generator, a trainable version of Frog) has been created to allow for the generation of a functional Frog instantiation based on a lemmatized and part-of-speech tagged corpus in any language (these two levels are the current scope of Froggen; more levels will be added). Successful tests have been carried out with Old Greek and historical variants of Dutch.

T24: FLAT

- Source Repository: https://github.com/proycon/flat

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Ongoing development and maintenance: **accomplished**
- Updated documentation: accomplished
- Support for complex span annotation (dependency relations, syntax): **accomplished**
- Pagination: **accomplished**
- Visualization: **accomplished**
- Annotation of substrings and other text units (to allow the handling of pre-tokenized text): **not performed** (no use cases lowered the priority on this)
- Caching: **accomplished**
- Conversion from other formats (with FoLiA module in pandoc): **accomplished**, there is a mechanism for embedding convertors and this has been actively used in practise. There is, however, no automatic support for various other formats yet.
- OAuth Support for single-sign login: **not performed** (CLARIN does not yet offer the OAuth infrastructure that makes this possible, which prompted us to hold on to our own simple authentication system for the time being)
- Right-to-Left support (Arabic, Hebrew): **accomplished**

In June, FLAT was selected as the annotation tool of choice by the PARSEME COST Action (http://typo.uni-konstanz.de/parseme/), this led to some shifted priorities and implementation of various new features.

T55: Ucto

- Website: https://languagemachines.github.io/ucto
- Source Repository: https://github.com/LanguageMachines/ucto

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Ongoing development and maintenance: **accomplished**
- Updated documentation: We converted the static PDf version of the documentation to an online website and updated the content. The Ucto documentation was released on 30 Oct 2018 and is now available at https://ucto.readthedocs.io/.

We improved and validated the language-dependent part of Ucto and checked the rules and abbreviations for English, Italian, Portuguese, French and Turkish.

We evaluated the performed of Ucto on new unseen material. It turned out to be very difficult to find existing independently verified tokenized text, as for most corpora the tokenization is not manually checked. We did obtain a

tokenized sample of 2897 sentences from the Dutch DCOI data set that was manually corrected. We used a small sample of around 60 sentences from the NewReader corpus in English, Dutch and Italian as a second evaluation set for Ucto. As the tokenization in the NewsReader corpus was not manually verified, we asked a native speaker of Italian and an expert for English and Dutch to verify the tokenization. We ran Ucto on these testsets and compared the outcome to the manually corrected labels. The resulting F-scores were very high, ranging from 99.2 for English to 99.9 on the DCOI dataset. The remaining errors concerned splits on dashes and uncommon names with unexpected characters in the strings.

T71: FoLiA

- Website: https://proycon.github.io/folia
- Source Repository: https://github.com/proycon/folia

See the attached software release report for a full overview of all releases during CLARIAH core.

The following was planned, we report on the status:

- Support for users, extending annotation types on request: **accomplished**
- Updated libfolia, the C++ library (used in Frog, T-Scan, TICCL): **accomplished**
- Updated Python library: **accomplished**
- Updated documentation: **accomplished**
- Deep Validation: **accomplished**
- Refactoring and syncing libfolia and pynlpl.formats.folia: **accomplished**
- Speech annotation: **accomplished**
- Stand-off format: **accomplished** (but removed again in 2019 with FoLiA v2 due to no users)
- Constructing FoLiA Set Definitions for FLAT: various sets have been created and the set definition format has been completely revised (now uses SKOS/RDF), converter from the old legacy format is provided
- Conversion module for FoLiA in pandoc: **partially accomplished**, not in pandoc but we decided on separate rst2folia tool (using docutils) that converts ReStructuredText to FoLiA.

LaMachine

- Website: https://proycon.github.io/lamachine
- Source Repository: https://github.com/proycon/lamachine

LaMachine was not initially planned in CLARIAH, but was conceived along the way. LaMachine is a unified software distribution for Natural Language Processing that distributes our entire software stack. This is also how we deliver our software to CLARIN centres and other users.

We integrate numerous open-source NLP tools, programming libraries, web-services, and web-applications in a single Virtual Research Environment that can be installed on a wide variety of machines.

The software included in LaMachine tends to be highly specialised and generally depends on a lot of other interdependent software. Installing all this software can be a daunting task, compiling it from scratch even more so. LaMachine attempts to make this process easier by offering pre-built recipes for a wide variety of systems, whether it is on your home computer or whether you are setting up a dedicated production environment, LaMachine will safe you a lot of work.

We address various audiences; the bulk of the software is geared towards data scientists who are not afraid of the command line and some programming. We give you the instruments and it is up to you to yield them. However, we also attempt to accommodate researchers that require more high-level interfaces by incorporating webservices and websites that expose some of the functionality to a larger audience.

Scientific Ouput: We submitted an extended abstract on LaMachine for the CLARIN conference, which was presented presented in Pisa in poster-form in Ocotober, 2018. We also worked on a full paper which was recently accepted for the proceedings of that conference.

Appendix: Software Release Report

From: 2015-06-01 (a bit before the official start, some possible overlap with CLARIN-NL) To: 2019-02-01 (some possible overlap with CLARIAH-PLUS)

CLAM

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

clam v2.4.4

Bugfix release: fix for forwarders in web interface #69

(Released by Maarten van Gompel on 2019-02-11) https://github.com/proycon/clam/releases/tag/v2.4.4

clam v2.4.3

- Allow CLAMClient to use HTTP Basic Auth instead of Digest
- check whether the proper authorization scheme is used

(Released by Maarten van Gompel on 2018-12-12) https://github.com/proycon/clam/releases/tag/v2.4.3

clam v2.4.2

- Bugfix: Boolean paramters (checkboxes) with default True did not get posted #73
- Allow HTTP digest authentication to be disabled (e.g. for allowing only HTTP basic authentication)

(Released by Maarten van Gompel on 2018-12-11) https://github.com/proycon/clam/releases/tag/v2.4.2

clam v2.4.1

Minor bugfix release:

- Fix in CLAMFile.read()

(Released by Maarten van Gompel on 2018-11-21) https://github.com/proycon/clam/releases/tag/v2.4.1

clam v2.4.0

- Ported documentation to sphinx and rewrote/restructured various sections (#72), the old documentation is now obsolete.
- added system details to the CLAM footer (so it's not just a CLAM footer)
- Added Forwarders (#69) and ForwardViewer
- Implemented CLAMData.get() (akin to dict.get()) #50
- parameter default fix (#71)

(Released by Maarten van Gompel on 2018-11-19) https://github.com/proycon/clam/releases/tag/v2.4.0

clam v2.3.6

- Further updates and fixes to clamnewproject to generate a new project with host-specific external configurations #70

(Released by Maarten van Gompel on 2018-11-06) https://github.com/proycon/clam/releases/tag/v2.3.6

clam v2.3.5

- Updated clamnewproject to generate projects with a setup.py #70

(Released by Maarten van Gompel on 2018-11-02) https://github.com/proycon/clam/releases/tag/v2.3.5

clam v2.3.4

Minor update:

- Added codemeta metadata
- Extra information on output templates on `info/` page (accommodating a request by @henkvdheuvel)

(Released by Maarten van Gompel on 2018-08-28) https://github.com/proycon/clam/releases/tag/v2.3.4

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

clam v2.3

Update with various enhancements and fixes.

- Investigate increased flexibility in cross-domain loading XSL of stylesheets #63
- Implement support for external configuration files (yaml) #67 (needed for LaMachine v2)
- Allow pre-setting global parameters #66 (needed for CLARIN LR switchboard)
- implement CUSTOMHTML_PROJECTFAILED #42
- [ui] center "show input files" button? #44
- Clean up info page (RESTful specification), doesn't filter duplicate output templates + minor issues #62

(Released by Maarten van Gompel on 2018-03-22) https://github.com/proycon/clam/releases/tag/v2.3

clam v2.2.5

- Minor update, added `JSONFormat` and renamed `UndefinedXMLFormat` to `XMLFormat` (with backwards compatibility link)

(Released by Maarten van Gompel on 2018-02-21) https://github.com/proycon/clam/releases/tag/v2.2.5

clam v2.2.4

- Minor bugfix releases, fixes issue #61

(Released by Maarten van Gompel on 2018-01-16) https://github.com/proycon/clam/releases/tag/v2.2.4

clam v2.2.3

Minor bugfix release (don't fall over dangling symlinks)

(Released by Maarten van Gompel on 2017-12-22) https://github.com/proycon/clam/releases/tag/v2.2.3

clam v2.2.2

- Minor bugfix release for FLATViewer

(Released by Maarten van Gompel on 2017-11-22) https://github.com/proycon/clam/releases/tag/v2.2.2

clam v2.2.1

- Fixed forwarded authentication (#55)

(Released by Maarten van Gompel on 2017-10-25) https://github.com/proycon/clam/releases/tag/v2.2.1

clam v2.2

- Implemented workaround for problems with HTTP 100 Continue (#54)
- Added multiple authentication support (Allowing basic & Digest at the same time), note that Basic is only available if an SSL connection is used, see the next time:
- Added `ASSUMESSL` setting, will be autodetected based on port if not set explicitly. Support for HTTP Basic Authentication is enabled automatically if set.
- Fixes for HTTP Basic Authentication

(Released by Maarten van Gompel on 2017-09-13) https://github.com/proycon/clam/releases/tag/v2.2

clam v2.1.14

Another important bugfix for Access-Control-Allow-Origin (Cross-Site Request Forgery) (#51)

(Released by Maarten van Gompel on 2017-06-27) https://github.com/proycon/clam/releases/tag/v2.1.14

clam v2.1.13

- Various fixes for configurable Access-Control-Allow-Origin (Cross-Site Request Forgery) (#51), was incomplete in v2.1.12

(Released by Maarten van Gompel on 2017-06-19) https://github.com/proycon/clam/releases/tag/v2.1.13

clam v2.1.12

- Added configurable Access-Control-Allow-Origin (Cross-Site Request Forgery) (#51)

(Released by Maarten van Gompel on 2017-05-15) https://github.com/proycon/clam/releases/tag/v2.1.12

clam v2.1.11

Minor update: Better feedback on error in FLAT viewer

(Released by Maarten van Gompel on 2017-05-05) https://github.com/proycon/clam/releases/tag/v2.1.11

clam v2.1.10

Minor bugfix release: fixes interface problem regarding input sources

(Released by Maarten van Gompel on 2017-04-05) https://github.com/proycon/clam/releases/tag/v2.1.10

clam v2.1.9

Minor bugfix release:

- Hide filename field in interface when filename is fixed anyway #23
- Implement more elegant solution for custom CSS (allow absolute paths now) #20

(Released by Maarten van Gompel on 2017-03-10) https://github.com/proycon/clam/releases/tag/v2.1.9

clam v2.1.8

Minor update release (adds `loadmetadata` parameter to `CLAMClient` API)

(Released by Maarten van Gompel on 2017-01-30) https://github.com/proycon/clam/releases/tag/v2.1.8

clam v2.1.7

- Cleanup by moving webservices for our NLP tools from CLAM itself to new separate repository `clamservices` (#47)
- Added a viewer to delegate FoLiA output to FLAT (#48)

(Released by Maarten van Gompel on 2017-01-05) https://github.com/proycon/clam/releases/tag/v2.1.7

clam v2.1.6

- Clean-up release
- Advertise full version string
- Entry point `clamservice/startclamservice` no longer re-invokes a Python interpreter
- Use of `exec()` replaced by importlib

(Released by Maarten van Gompel on 2016-11-13) https://github.com/proycon/clam/releases/tag/v2.1.6

clam v2.1.5

Minor bugfix release:

- Fixes problem with XSLTViewer in Python 2 (issue #46)

(Released by Maarten van Gompel on 2016-10-19) https://github.com/proycon/clam/releases/tag/v2.1.5

clam v2.1.4

Minor bugfix: directories in the temp directory did not get removed properly

(Released by Maarten van Gompel on 2016-04-12) https://github.com/proycon/clam/releases/tag/v2.1.4

clam v2.1.3

Bugfix release: template.sh was missing, causing clamnewproject to malfunction.

(Released by Maarten van Gompel on 2016-04-05) https://github.com/proycon/clam/releases/tag/v2.1.3

clam v2.1.2

Bugfix release: fixes custom validation functions (was partially defective since introduction in issue #39)

(Released by Maarten van Gompel on 2016-04-05) https://github.com/proycon/clam/releases/tag/v2.1.2

clam v2.1.1

Bugfix release; there was a problem with spaces in parameters in actions.

(Released by Maarten van Gompel on 2016-03-31) https://github.com/proycon/clam/releases/tag/v2.1.1

clam v2.1.0

This is a major update release of CLAM, containing various bug fixes as well as new functionality.

New features

- *OAuth2* and *external forwarded authentication* are functional again (issue #4)
- New *program* paradigm for use in e.g. Python wrapper scripts (issue #36), allows target-based programming through iteration over output files prior to their generation.
- Projects are indexed/cached and disk usage is tracked (issue #26)
- Added an interface to quickly delete multiple projects, see disk usage per project and total (issue #26)
- Fixes for the administrative interface
- Implement hooking up a custom value validation function for parameter validation (issue #39)
- Implemented temporary directories for both project and action paradigm, concurrency-safe (issue #37, issue #40)
- More elegant solution for custom formats (issue #21)
- Support for shell operators (pipes, redirects) in commands

Bug fixes

- Important fixes in CLAM Client API, files were not transferred properly (issue #33)
- Upload failures were not properly communicated to the interface, now they are.
- Proper SSL support for CLAM Client API, allow passing certificate (issue #31)
- Session nonces accumulated on disk, cleared now (issue #25)
- Some minor stylistic fixes (issue #23, issue #27, issue #34)
- Cleaner XML output (no unnecessary whitespace) for all REST responses.
- Fix for parsing unicode in non-unicode strings in service configuration in Python 2 (issue #24) All users of earlier CLAM versions are urged to upgrade to this release. The release is available through the Python Package Index, as well as Github.

(Released by Maarten van Gompel on 2016-03-30) https://github.com/proycon/clam/releases/tag/v2.1.0

clam v2.0.1

Stable release of CLAM (formerly v0.99 release candidate), using `flask` as backend.

(Released by Maarten van Gompel on 2016-02-09) https://github.com/proycon/clam/releases/tag/v2.0.1

clam v0.99.15

(Released by Maarten van Gompel on 2015-11-22) https://github.com/proycon/clam/releases/tag/v0.99.15

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

clamservices v1.5.1

Previous release missed some files

(Released by Maarten van Gompel on 2018-03-23) https://github.com/proycon/clamservices/releases/tag/v1.5.1

clamservices v1.5

New release for CLAM 2.3, with external configuration files. Breaks compatibility with LaMachine v1.

(Released by Maarten van Gompel on 2018-03-22) https://github.com/proycon/clamservices/releases/tag/v1.5

clamservices v1.4.4

Finalizing migration to new server in Nijmegen

(Released by Maarten van Gompel on 2017-12-04) https://github.com/proycon/clamservices/releases/tag/v1.4.4

clamservices v1.4.3

Fix for FLAT link on Nijmegen servers

(Released by Maarten van Gompel on 2017-10-31) https://github.com/proycon/clamservices/releases/tag/v1.4.3

clamservices v1.4.2

- Fix for ucto webservice in LaMachine vagrant/docker

(Released by Maarten van Gompel on 2017-09-14) https://github.com/proycon/clamservices/releases/tag/v1.4.2

clamservices v1.4.1

- Added support for new webserver in Nijmegen

(Released by Maarten van Gompel on 2017-08-25) https://github.com/proycon/clamservices/releases/tag/v1.4.1

clamservices v1.4

- Added missing language options for ucto (Swedish, Russian , Spanish, Portuguese, Turkish) (addresses issue clarin-eric/LRSwitchboard#7)
- Use iso-639-3 for ucto

(Released by Maarten van Gompel on 2017-07-11) https://github.com/proycon/clamservices/releases/tag/v1.4

clamservices v1.3

Added "Open in FLAT" viewer for Frog and ucto services

(Released by Maarten van Gompel on 2017-05-05) https://github.com/proycon/clamservices/releases/tag/v1.3

clamservices v1.2

- Fixes for frog configuration path

(Released by Maarten van Gompel on 2017-03-17) https://github.com/proycon/clamservices/releases/tag/v1.2

clamservices v1.1

Fixed issue with frog configuration path (#1)

(Released by Maarten van Gompel on 2017-03-09) https://github.com/proycon/clamservices/releases/tag/v1.1

clamservices v1.0

First release (moved from CLAM main repo)

(Released by Maarten van Gompel on 2017-01-05) https://github.com/proycon/clamservices/releases/tag/v1.0

FLAT

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

flat v0.7.14

Bugfix release:

- Fixes bug in adding certain span annotation elements (#138)

(Released by Maarten van Gompel on 2018-10-09) https://github.com/proycon/flat/releases/tag/v0.7.14

flat v0.7.13

Added support for URL prefix, i.e. hosting FLAT not at the root, needed by LaMachine #131

(Released by Maarten van Gompel on 2018-08-12) https://github.com/proycon/flat/releases/tag/v0.7.13

flat v0.7.12

- Added /config endpoint that just outputs all configurations in JSON (usually for #50)
- Documented public upload facility (#50)

(Released by Maarten van Gompel on 2018-03-13) https://github.com/proycon/flat/releases/tag/v0.7.12

flat v0.7.11

- Added Django 2.0 compatibility **Upgrade advisory:** If you run Django 2.0 and come from an earlier version, change MIDDLEWARE_CLASSES in your settings.py to MIDDLEWARE. Otherwise you will encounter an error 'WSGIRequest' object has no attribute 'user'

(Released by Maarten van Gompel on 2017-12-07) https://github.com/proycon/flat/releases/tag/v0.7.11

flat v0.7.10

Bugfix release: fixes issue correction annotation (#124)

(Released by Maarten van Gompel on 2017-11-01) https://github.com/proycon/flat/releases/tag/v0.7.10

flat v0.7.9

Bugfix release:

- Fix for #122: Correction class list box does not set current selection
- Reimplementing editing correction class (#120)
- Fix sort targets #119

(Released by Maarten van Gompel on 2017-10-18) https://github.com/proycon/flat/releases/tag/v0.7.9

flat v0.7.8

- Added option to exclude certain classes from a set definition (for the editor) #118
- Improvements for metadata editor
- Added autometadata configuration option, specifying metadata defaults

(Released by Maarten van Gompel on 2017-10-04) https://github.com/proycon/flat/releases/tag/v0.7.8

flat v0.7.7

Bugfix release: FLAT had a minor HTTP violation in communication with foliadocserve (and foliadocserve's backend cherrypy got stricter), fixes HTTP 400 responses from foliadocserve

(Released by Maarten van Gompel on 2017-10-02) https://github.com/proycon/flat/releases/tag/v0.7.7

flat v0.7.6

Minor update:

- Ship with migrations (#117)
- Added an option for automatically executing a preset search on page load

(Released by Maarten van Gompel on 2017-10-02) https://github.com/proycon/flat/releases/tag/v0.7.6

flat v0.7.5

Minor update:

- Allow disabling user registration (ALLOWREGISTRATION directive) (#114)
- Clarified documentation (#116)

(Released by Maarten van Gompel on 2017-06-16) https://github.com/proycon/flat/releases/tag/v0.7.5

flat v0.7.4.1

Minor bugfix release: exempt public anonymous interfaces from CSRF protection

(Released by Maarten van Gompel on 2017-05-08) https://github.com/proycon/flat/releases/tag/v0.7.4.1

flat v0.7.3

Minor bugfix release for public editor (#50)

(Released by Maarten van Gompel on 2017-05-05) https://github.com/proycon/flat/releases/tag/v0.7.3

flat v0.7.2

Bugfix release: fixes two regression issues (#107, #108) noticed immediately after v0.7.1 release, reverts part of #102

(Released by Maarten van Gompel on 2017-03-02) https://github.com/proycon/flat/releases/tag/v0.7.2

flat v0.7.1

- Major improvements in session handling, concurrency now works #106
- Implemented a public facility, allowing public anonymous uploads/views/edits (without login) #50
- Added numbers for structural items in perspective view (issue #86)
- Cleaned up "Add new annotation type" dialog #96
- Added syntax annotation test #95
- Stylistic improvements/fixes (thanks to @ErkanBasar) #101 #102
- Check version information from foliadocserve
- Various clarifications/fixes in documentation
- [bugfix] Global annotations disappear when an entity is labeled. #103
- [bugfix] FQL syntax error on adding features on their parent elements in one go #104
- [bugfix] Backend unloaded documents prematurely due to poor session handling #105

(Released by Maarten van Gompel on 2017-03-02) https://github.com/proycon/flat/releases/tag/v0.7.1

flat v0.7

Major update, builds on foliadocserve >= v0.5

- Refactor of data model, reduce memory consumption (#41)
- Edit more complex forms of span annotation: dependencies, co-references (#3)
- Editing features (#43)
- Selector mechanism that allows selecting structure elements that are not on the deepest level (#75)
- Use common FoLiA specification to create necessary structures (#76)
  - Use FoLiA onstrains for new annotations (#80)
  - Prevent duplicate annotations (#82)
- Updated and refactored test suite (#95)
- More clearly indicate text of selected span (#69)
- Support for phonetic content (#89)
- Support for visualisation of morphology (#13)
- Annotation menus cleaned up and properly sorted (#90)
- Syntax tree visualisation (#85)
- Added ability to predefine search queries in the configuration; adding them to the menu for quick access (#94)
- Fixes in handling of corrections (#79 #91)
- [bugfix] extra failsafes in global visualisation (#98)
- [bugfix] add new annotation dialog (#96)
- [bugfix] Modify both span and category of annotation at same time (#97)

(Released by Maarten van Gompel on 2016-12-29) https://github.com/proycon/flat/releases/tag/v0.7

flat v0.6.2

Update release, featuring several smaller fixes and improvements to search:

- [Search function] Changed highlight colour for search results (a red border and bold text), to not conflict with existing highlights.
- [Search function] Automatically resubmit searches when the page changes (highlight mode)
- [Search function] Search window's "perspective mode" now works, but requires carefully crafted FQL queries, `RETURN` statement must return something structural that FLAT can visualise. (issue #31)
- [Bugfix] Ensure colours of annotation focus are consistent (issue #81)
- [Bugfix] Don't stumble over empty comments boxes (issue #77)
- [Bugfix] Extra checks to prevent expired user sessions
- [Bugfix] Clear wait screen when going back to a page from the browser (issue #83)

(Released by Maarten van Gompel on 2016-10-29) https://github.com/proycon/flat/releases/tag/v0.6.2

flat v0.6.1

Bugfix release, fixes issue #73, related to the metadataindex.

(Released by Maarten van Gompel on 2016-10-14) https://github.com/proycon/flat/releases/tag/v0.6.1

flat v0.6.0

Major update release:

- New and updated documentation: *installation guide*, *administration guide* and *user guide*, hosted on http://flat.readthedocs.io/en/latest/ (issue #16, issue #61)
- Implemented file management features from the document index (issue #54, issue #20)
- New user permission model with support for groups and group namespaces. Permissions assignable from administrative interface (issue #63)
- Added ability to constrain metadata entry (issue #58)
- Added password reset facility (issue #21) and change password link in menu
- Administrative interface accessible through menu now
- Added facility to plug in converters for file upload (issue #62)
- Added `requiredeclaration` option for FLAT configuration, prevents loading wrong documents in the wrong configuration (issue #56)
- Added `autoselectspan` option to automatically press the select span button on opening the editor dialog (issue #71)
- Document index allows showing specific metadata fields (issue #55)
- Bugfix: Support usernames/namespaces with @ signs (issue #66)
- Bugfix: Global annotations are now rendered when a sub-part is updated (issue #27)
- Bugfix: Overlapping spans are properly visualised in global annotations (issue #70)
- Bugfix: Allow overlapping annotation of the same class (issue #67)
- Bugfix for comments or descriptions with a newline (issue #51)
- Dropped support for Django < 1.8

(Released by Maarten van Gompel on 2016-10-01) https://github.com/proycon/flat/releases/tag/v0.6.0

flat v0.5.3

Bugfix and documentation release:

- Fixes issue #64 (issue with editform new)
- Now with better documentation (issue #16, #61): installation guide, user guide, administration guide, hosted on http://flat.readthedocs.io/en/latest/

(Released by Maarten van Gompel on 2016-09-26) https://github.com/proycon/flat/releases/tag/v0.5.3

flat v0.5.2

Bugfix release:

- Fixes inability to add descriptions (issue #48)

- Fixes bug adding comments with new annotations (issue #47)

(Released by Maarten van Gompel on 2016-09-13) https://github.com/proycon/flat/releases/tag/v0.5.2

flat v0.5.1

- Adds compatibility with the latest django (1.10) (issue #46)

(Released by Maarten van Gompel on 2016-09-02) https://github.com/proycon/flat/releases/tag/v0.5.1

flat v0.5

- Styling improvements: some colours have changed (better contrast etc) and applied more consistently over the various forms and views.
- You can now assign comments and/or descriptions to any annotation (issue #38, comments are introduced in FoLiA v1.3), lays the foundation for further higher-order annotations
- Added confidence slider (issue #18), you can set a confidence value for any annotation
- Added a simple metadata editor (issue #39), supports FoLiA's limited native metadata format only.
- Added additional interface and integration tests for the new features
- The interface now uses the labels from the set definition, if defined (issue #37, introduced in FoLiA v1.3)

(Released by Maarten van Gompel on 2016-09-01) https://github.com/proycon/flat/releases/tag/v0.5

flat v0.4.2

- Fixes bug with revision save button (issue #32)

(Released by Maarten van Gompel on 2016-05-18) https://github.com/proycon/flat/releases/tag/v0.4.2

flat v0.4.1

- Adds right-to-left support (solves issue #2)

(Released by Maarten van Gompel on 2016-05-10) https://github.com/proycon/flat/releases/tag/v0.4.1

foliadocserve v0.6.8

Bugfix release:

- Fixed issue #14 (Metadata changes weren't saved properly if no other annotations were made)

(Released by Maarten van Gompel on 2018-09-19) https://github.com/proycon/foliadocserve/releases/tag/v0.6.8

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

foliadocserve v0.6.6

- Fixed stop and SIGUSR1 behaviour (#12)
- Log file now appends to existing file instea of overrwriting

(Released by Maarten van Gompel on 2018-02-05) https://github.com/proycon/foliadocserve/releases/tag/v0.6.6

foliadocserve v0.6.5

- Implemented extra protections in case document serialisation fails
  - Better log output on such failure
  - foliadocserve goes into a lockdown mode, prevent new document from being loaded
- Implemented extra protection in case contact with the background autoupdater thread is lost after a certain time limit (same lockdown mode)

(Released by Maarten van Gompel on 2018-02-04) https://github.com/proycon/foliadocserve/releases/tag/v0.6.5

foliadocserve v0.6.4

- Added support for git --shared (enabled by default for group) using `--gitshare` (#11)
- Added option to choose between different git modes using `--gitmode` (#11)
  - monolithic (ALL users share a single repository, NOT recommended because of scalability)
  - user (each user/namespace is its own git repository; this is the default)
  - nested (each subdirectory is its own git repository, maximum scalability)
- Implemented sane stop behaviour (#12)

(Released by Maarten van Gompel on 2017-10-23) https://github.com/proycon/foliadocserve/releases/tag/v0.6.4

foliadocserve v0.6.3

- Added support for FoLiA v1.5; strip all text redundancy by default (#10)

(Released by Maarten van Gompel on 2017-10-09) https://github.com/proycon/foliadocserve/releases/tag/v0.6.3

foliadocserve v0.6.2

Minor bugfix release: ensure all JSON properties use double quotes (invalid JSON otherwise), fixes proycon/flat#110

(Released by Maarten van Gompel on 2017-05-05) https://github.com/proycon/foliadocserve/releases/tag/v0.6.2

foliadocserve v0.6.1

Minor fix release:

- Increased max allowed upload document size to 1GB instead of 100MB (fixes proycon/flat#109)

(Released by Maarten van Gompel on 2017-03-20) https://github.com/proycon/foliadocserve/releases/tag/v0.6.1

foliadocserve v0.6.0

Update release, to be used with FLAT v0.7.1:

- Important fixes in session handling (#8), now properly support concurrency (proycon/flat#106)
- Fixes autounload problem (proycon/flat#105)
- Propagate page size to clients
- More verbose logging

(Released by Maarten van Gompel on 2017-03-02) https://github.com/proycon/foliadocserve/releases/tag/v0.6.0

foliadocserve v0.5.1

Bugfix release, fixes in foliadocserve version reporting

(Released by Maarten van Gompel on 2017-01-03) https://github.com/proycon/foliadocserve/releases/tag/v0.5.1

foliadocserve v0.5

Major update, compatible only with FLAT >= v0.7:

- New data model (#6)

(Released by Maarten van Gompel on 2016-12-29) https://github.com/proycon/foliadocserve/releases/tag/v0.5

foliadocserve v0.4.2

Minor bugfix release, fixes issue #4

(Released by Maarten van Gompel on 2016-10-25) https://github.com/proycon/foliadocserve/releases/tag/v0.4.2

foliadocserve v0.4.1

- Added file management facilities
- Changes in returning namespaces (now recursive over all)

(Released by Maarten van Gompel on 2016-10-01) https://github.com/proycon/foliadocserve/releases/tag/v0.4.1

foliadocserve v0.4

- FoLiA 1.3 support
- Various fixes
- Added tests for FLAT
- Support for metadata editing.

(Released by Maarten van Gompel on 2016-09-01) https://github.com/proycon/foliadocserve/releases/tag/v0.4

foliadocserve v0.3.6

Bugfix release: Fixes git revert functionality (issue proycon/flat#35)

(Released by Maarten van Gompel on 2016-05-25) https://github.com/proycon/foliadocserve/releases/tag/v0.3.6

foliadocserve v0.3.5

- Adds right-to-left support (issue proycon/flat#2)

(Released by Maarten van Gompel on 2016-05-10) https://github.com/proycon/foliadocserve/releases/tag/v0.3.5

foliadocserve v0.3.4

Initial release

(Released by Maarten van Gompel on 2016-04-02) https://github.com/proycon/foliadocserve/releases/tag/v0.3.4

FoLiA

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

folia 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/folia/releases/tag/v2.0.1

folia 2.0.0

This is a major new release of FoLiA, which includes some breaking changes such as renamed elements. Nevertheless, the FoLiA libraries retain backward compatibility and can read FoLiA v1 (and v0) documents and upgrade them. Points of general interest:

- Completely revised the FoLiA documentation, turned into more formal specification; automatically drawn from the official specification; with automatically validated examples. Now available as a webpage hosted on https://folia.readthedocs.io (PDF still available too) #43
  - The documentation includes some guidelines on good FoLiA practises (arose from #70 and others)
- Added proper support for provenance logging in FoLiA #46
- Renamed alignment annotation to relation annotation #59
- Ensured most examples are "sensible" #9
  - Extended tests using these examples, all examples are automatically tested now
- The FoLiA tools are now split from the central FoLiA repository into a separate project at https://github.com/proycon/foliatools #55
  - Cleaner output without stack traces from FoLiA validator #44

- Implemented the ability to add inline annotations on multi-word spans (group annotations) and solved related multi-word issues. These were previously reserved only for use with structural elements. #51
- Revised the structure annotation hierarchy (i.e. which structural elements are allowed under which parents) on certain points #42
- Implemented a hidden words annotation type, allowed a layer of implicit/empty/ghost words that can be referenced from span annotation. Needed e.g. for syntactic movement annotation. #58
- Allow encoding of soft word breaks / hyphenation #66 More technical points:
- Add support for provenance in FQL #60
- Annotation declaration overhaul and handle missing set attribute in declarations #54
- Explicitly forbid and prevent forward wrefs from span annotation #41
- Apply space attribute more generically to multiple structure elements #61
- Added a new property in the specification to detect tags that may be (or MUST be) used as Wrefs #63
- Added a new property to distinguish folia:id (IDREF) from xml:id (ID) #64
- Alias attribute does not propagate to RelaxNG schema yet #65 A new FoLiA library has been released (replacing the previous one in PyNLPl): https://github.com/proycon/foliapy/releases/tag/2.0.0 A new version of FoLiA tools has also been released: https://github.com/proycon/foliatools/releases/tag/2.0.0 You may also consult the FoLiA release plan (#68) for more information on upgrading and compatibility.

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/folia/releases/tag/2.0.0

folia v1.5.1.60

- [foliacorrect] Implemented `--acceptsuggestion`

(Released by Maarten van Gompel on 2018-01-15) https://github.com/proycon/folia/releases/tag/v1.5.1.60

folia v1.5.1.59

FoLiA v1.5.1

- Minor update: set comment printable to false explicitly FoLiA-Tools v1.5.1.59
- Fixes in `foliavalidator` for directory processing
- Prerelease of new `foliaeval` tool (still under construction)

(Released by Maarten van Gompel on 2017-11-28) https://github.com/proycon/folia/releases/tag/v1.5.1.59

folia v1.5.0.57

FoLiA v1.5

- Implemented text validation (#24); checks text redundancy and offsets
- Added facilities for metadata on sub-parts of a document (#30)
- Added support for aliases (short names) for set definitions (#31)
- More liberal acceptance of Linebreak and Whitespace
- Allow Paragraph and Part under ListItem
- TextContent and PhonContent were erroneously disallowed under Part **Important note:** Text validation is now the default for FoLiA v1.5+, this means that documents will be more strictly validated regarding their text content. Inconsistenties in text reduncancy or offsets results in invalid FoLiA. FoLiA-Tools v1.5.0.57
- Expanded `rst2folia` tool, better rst coverage and fixes
- `foliavalidator` adapted for text checking (default for v1.5+, off for older)

(Released by Maarten van Gompel on 2017-10-08) https://github.com/proycon/folia/releases/tag/v1.5.0.57

folia v1.4.3.56

FoLiA v1.4.3

- Added `textclass` attribute to make relation between different text classes and annotations explicit (#29) FoLiA-tools v1.4.3.56

- Added `foliaid` tool
- Fixes for `foliamerge` tool

(Released by Maarten van Gompel on 2017-08-16) https://github.com/proycon/folia/releases/tag/v1.4.3.56

folia v1.4.2.55

FoLiA v1.4.2 release:

- Allow more structural elements under
- Space attribute simplified #28 FoLiA-tools v.1.4.2.55:
- `foliavalidator`: Implemented experimental text validation (warnings only for now)
- Expanding `foliamerge` tool with support for merging as alternative

(Released by Maarten van Gompel on 2017-07-13) https://github.com/proycon/folia/releases/tag/v1.4.2.55

folia v1.4.1.54

FoLiA v1.4.1 release:

- Allow <w> directly under <listitem> (#26)
- Fixes for linebreak (<br/>) as text markup (allow xlink attribute) **FoLiA-tools v.1.4.1.54**:
- Updated foliavalidator to generate more extensive tracebacks

(Released by Maarten van Gompel on 2017-03-28) https://github.com/proycon/folia/releases/tag/v1.4.1.54

folia v1.4.0.53

FoLiA v1.4

- Migrated FoLiA Set Definitions to RDF using the SKOS model (#14)
- Deep validation is properly implemented now
  - Added an example documented for deep validation
  - Added an example document with deep morphology (Frog output)
- Various set definitions are provided, most notably those by Frog
  - CGN Part-of-Speech tagset covered now (#18)
- Constraining the number span roles (of a specific type) allowed under a span annotation
- Updated documentation

FoLiA-Tools v1.4.0.53

- Added `foliasetdefinition` tool, used to inspect and test set definitions, and to provide conversion from legacy XML sets to RDF SKOS sets
- `foliatextcontent` prints warnings instead of raising NotImplementErrors now
- Minor update to `foliaspec` (issue LanguageMachines/libfolia#8)

(Released by Maarten van Gompel on 2016-12-09) https://github.com/proycon/folia/releases/tag/v1.4.0.53

folia v1.3.2.52

FoLiA v1.3.2 release, minor update:

- Allow multiple foreign-data elements in metadata (issue #23)
- RelaxNG schema fix for foreign-data in metadata

(Released by Maarten van Gompel on 2016-10-11) https://github.com/proycon/folia/releases/tag/v1.3.2.52

folia v1.3.1.52

- FoLiA v1.3.1 release, minor update over v1.3.0
  - Allow features on CoreferenceChain
  - Allow multiple sense annotations of the same set

(Released by Maarten van Gompel on 2016-09-29) https://github.com/proycon/folia/releases/tag/v1.3.1.52

folia v1.3.0.52

FoLiA v1.3

- Added a 'comment' element (higher-order annotation) (issue #10)
- Added a label attribute on set definition (issue #11)
- Added a **predicate** span annotation element that groups semantic roles (issue #13)
- Added a new linguistic annotation type for various **observations** on a text (issue #15)
- Added a new linguistic annotation type for **sentiment analysis** (issue #16), this is now favoured over the older and more limited `subjectivity` token annotation element.
- Added a new linguistic annotation type for **attribution** (issue #17)

(Released by Maarten van Gompel on 2016-09-01) https://github.com/proycon/folia/releases/tag/v1.3.0.52

folia v1.2.0.51

Folia Tools:

- added `foliatree` tool, prints the structure of a document.

(Released by Maarten van Gompel on 2016-07-22) https://github.com/proycon/folia/releases/tag/v1.2.0.51

folia v1.2.0.50

Minor bugfix release for foliaspec, previous release was premature

(Released by Maarten van Gompel on 2016-05-25) https://github.com/proycon/folia/releases/tag/v1.2.0.50

folia v1.2.0.49

Minor bugfix release for foliaspec.

(Released by Maarten van Gompel on 2016-05-25) https://github.com/proycon/folia/releases/tag/v1.2.0.49

folia v1.2

FoLiA v1.2

- Support for in-document metadata in any foreign namespace (see section 2.11 of the manual)
- Support for foreign namespace annotation using the `foreign-data` element (see section 2.10.15 of the manual)
- Native FoLiA metadata does not formally predefine any fields (but various tools may)
- Extended linebreak element (`br`) with `newpage,pagenr` and `linenr` attributes. Element can now be used to signal page breaks. (see section 2.5.6 of the manual)
- Alignments (`alignment`) now allows pointing to external documents that are not FoLiA, using the `format` attribute in combination with the already existing `xlink:href` attribute. (see section 2.10.8 of the manaul)
  - `aref/@type` is now a free field, the type needs not correspond to a FoLiA element anymore. Necessary to allow linking to external resources.
- Similarly, references (`ref`) can now link to external resources using xlink attributes and the `format` attributes. (See section 2.5.12 of the manual)
- Xlink support has been extended to include attributes like `xlink:role` etc.. These are now allowed on all xlink-capable elements.
- Example document (`test/example.xml`) has been brought in line with some extra tests conducted by the `libfolia` library.
- A small arabic example document was added (`test/example.ar.xml`), intended to test right-to-left visualisation.

FoLiA Tools v1.2.0.48

- Improvements in `folia2html` (`folia2html.xsl`)
  - Support for right-to-left languages (based on explicit metadata field `direction: rtl`)
  - Explicit support for titles in dublin core format
  - Better propagation of metadata from FoLiA to HTML
  - Visualisation of some token annotation under morphemes
  - Render phonetics in annotation pop-up

(Released by Maarten van Gompel on 2016-05-24) https://github.com/proycon/folia/releases/tag/v1.2

folia v1.0.1.47

Bugfix release:

- Allow more optional attribs on annotation layers (was broken in v1.0.0 but frog NER and/or FLAT outputs it)

(Released by Maarten van Gompel on 2016-04-01) https://github.com/proycon/folia/releases/tag/v1.0.1.47

folia v1.0.0.47

FoLiA v1.0.0

- **New FoLiA specification** in `schemas/folia.yml`, to be used by external libraries to facilitate implementation and synchronization
- Minor fixes in documentation
- Library implementation status updated in documentation
- RelaxNG schema built from external specification (via `pynlpl.formats.folia` library)
- Better legacy support in RelaxNG schema (e.g `listitem` tag)

FoLiA-tools (rev 47)

- Added `foliaspec` tool that reads from external FoLiA specification and updates the library source code for `pynlpl` and `libfolia`

(Released by Maarten van Gompel on 2016-03-09) https://github.com/proycon/folia/releases/tag/v1.0.0.47

folia v0.12.2.46

- added ComplexAnnotation support in FoLiA schema

(Released by Maarten van Gompel on 2016-02-22) https://github.com/proycon/folia/releases/tag/v0.12.2.46

folia v0.12.2.45

Release

(Released by Maarten van Gompel on 2016-02-17) https://github.com/proycon/folia/releases/tag/v0.12.2.45

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

foliapy 2.0.1

Allow (limited!) serialisation of older FoLiA versions if `keepversion` is set.

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/foliapy/releases/tag/2.0.1

foliapy 2.0.0

Major new release; the former `pynlpl.formats.folia` module in PyNLPI has now been migrated to this new standalone package, which constitutes the new FoLiA library for Python. Included is the FoLiA library, FQL library and Set Definition library. This release implements the new FoLiA v2.0, consult the release notes there for a comprehensive list of changes. Library documentation is now hosted on https://foliapy.readthedocs.io/en/latest/. Upgrading from pynlpl is usually as easy as doing a `pip install folia` and replace within your software: `from pynlpl.formats import folia` with `import folia.main as folia`. The library is backward compatible.

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/foliapy/releases/tag/2.0.0

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

foliatools 2.0.1

Enable keepversion for certain tools, so they don't upgrade to FoLiA v2 if FoLiA v1 input is given, ensuring better backward compatibility.

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/foliatools/releases/tag/2.0.1

foliatools v2.0.0

This is the FoLiA v2 version of the FoLiA-tools, now split from the central FoLiA repository at https://github.com/proycon/folia and building on the new FoLiApy library. New documented resides at https://folia-tools.readthedocs.io/

- There is a new `foliaupgrade` tool to upgrade from FoLiA v1 (or v0) to v2 (although this is usually done implicitly by any of the other tools, or the FoLiAPy library itself, anyway)
- Refactored the entire foliavalidator
- Cleaner output without stack traces from FoLiA validator (proycon/folia#44)
- Added automatic injection of provenance information to various tools

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/foliatools/releases/tag/v2.0.0

foliautils v0.10

[Ko vd Sloot]

- fixed icu:namespace issues
- added FoLiA-abby, an ABBY to FoLiA convertor
- src/FoLiA-abby.cxx, src/FoLiA-page.cxx, src/FoLiA-pm.cxx:
    - Allow 'none' value for --prefix
- src/FoLiA-page.cxx, src/FoLiA-hocr.cxx: fixed Alignment info
- src/FoLiA-correct.cxx:
    - fixed a problem with correction of the last word of a trigram.
    - fix correction of paragraphs with only deeper text
    - The --rank option accepts more flavors of files
- src/FoLiA-stats.cxx:
    - added a --detokenize option
- several minor fixes, refactorings etc.
- updated tests

(Released by Ko van der Sloot on 2018-11-29) https://github.com/LanguageMachines/foliautils/releases/tag/v0.10

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

foliautils v0.8

- added -R option to FoLiA-collect
- FoLiA-collect now can work in parallel (-t option)
- modernized configuration, whit better Max OSX support (including OpenMP)
- all modules end with an exit code now.
- added more tests to 'make check'
- added output of Type-Token Ratio's (also in degrees)
- several bugfixes.
- code cleanup and refactoring, some speedup too

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/foliautils/releases/tag/v0.8

foliautils v0.7

[ko vd Sloot]

- updated and expanded tests
- fixed offset calculations in FoLiA-hocr, FoLiA-page.cxx and FoLiA-alto. We use unicode points now. (needed for folia v1.5 and above)
- Changed 'modes' in FoLiA-stats, to be a bit more comprehensible
- fixed problem with metadatatype when 'foreign-data' is present
- enhanced FoLiA-clean. Still not done...
- switched to dynamic OMP scheduling in most programs. (which process files with probably big differences in processing time)

- small bugfixes.
- general cleanup and refactoring

[Maarten van Gompel]

- Added and improved FoLiA-wordtranslate.cxx

(Released by Ko van der Sloot on 2017-10-24) https://github.com/LanguageMachines/foliautils/releases/tag/v0.7

foliautils v0.6

foliautils 0.6 04-04-2017 This is an intermediate release!! Work on some tools is developing rapidly. next releases won't take long. For now, backward compatibility is still maintained mostly.

[Ko van der Sloot]

- uses libfolia 1.7 now!
- FoLiA-correct now uses an other output file naming scheme (breaks backward compitablity)
- FoLiA-langcat now has a --tags parameter to select which nodes are searched
- FoLiA-stats:
  - a new --separator option is added
  - added a --max-ngram option.
  - added a --languages option for multiple languages
  - now we have a --aggregate option for multiple language statistics
  - fixed a bug in total counts
- added a first version of FoLiA-clean program. Cleans up tests/tags in FoLiA files.
- FoLiA-correct:
  - output statistics
  - verbosity option improved
- added and improved a lot of tests

(Released by Ko van der Sloot on 2017-04-04) https://github.com/LanguageMachines/foliautils/releases/tag/v0.6

foliautils v0.5

- based on libfolia 1.5 or higher
  - use recent ucto with textcat support
  - use ISO 639-3 language names
  - lot's of code refactoring
  - improved tests
  - bug fixes in FoLiA-correct unigram correction
  - extended and improved FoLiA-pm a lot
  - changed default values for '--lang' and '--class' in FoLiA-stats (issue #3)
  - FoLiA-alto can now work without a Didl too (issue #2)
  - numerous additions...

(Released by Ko van der Sloot on 2017-01-17) https://github.com/LanguageMachines/foliautils/releases/tag/v0.5

foliautils v0.4

A new program *FoLiA-pm* is added it converts Political Mashup files into FoLiA. Needs libfolia v1.2

(Released by Ko van der Sloot on 2016-05-24) https://github.com/LanguageMachines/foliautils/releases/tag/v0.4

foliautils v0.3

New release. Now based on libfolia 1.0

(Released by Ko van der Sloot on 2016-03-10) https://github.com/LanguageMachines/foliautils/releases/tag/v0.3

foliautils v0.2

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/foliautils/releases/tag/v0.2

libfolia v1.15

- added (still experimental) code for a FoLiA Builder, Processor and TextProcessor class. Use with care. The API may change unannounced!
- a foliadiff script (using folialint) is installed now
- several refactorings, to make the code more clear.
- the 'ref' attribute was not serialized for TextContent
- several smaller small bug fixes
- the .so version is bumped to 9 because of a lot of API/ABI changes

(Released by Ko van der Sloot on 2018-11-29) https://github.com/LanguageMachines/libfolia/releases/tag/v1.15

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

libfolia v1.12

- configuration cleanup. MacOSX is better supported now.
- folialint now supports --fixtext (handle with care!)
- library version bumped to 8.0, due to changes in the API
- regenerated FoLiA properties (to FoLiA version 1.5.1)
- several small bug fixes

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/libfolia/releases/tag/v1.12

libfolia v1.11

Bug fix release:

- handling of <comment> tags within <t> nodes
- better handling of <wref> tags. Forbid forward references

(Released by Ko van der Sloot on 2017-12-04) https://github.com/LanguageMachines/libfolia/releases/tag/v1.11

libfolia v1.10.1

Minor fix

- bumped the .so version to 7.0

(Released by Ko van der Sloot on 2017-11-06) https://github.com/LanguageMachines/libfolia/releases/tag/v1.10.1

libfolia v1.10

Major Release, implementing FoLiA spec 1.5

- added text checking for all 1.5 documents and up
- added offset and ref checking for Text in all 1.5 documents and up
- 'empty' text inside TextContent, PhonContent and Textmarkup is significant
- better version checking
- text checking can be dis/enabled using FOLIA_TEXT_CHECK environment variable
- added submetadata mechanism
- implemented aliases for annotation setnames
- added an xmlstring() serializer for Document

- bug fixes:
  - in LineBreak serializing
  - XmlComment is textless.
  - miscellaneous small fixes

(Released by Ko van der Sloot on 2017-10-18) https://github.com/LanguageMachines/libfolia/releases/tag/v1.10

libfolia v1.9

Bug fix release

- accept ICU 50 too (was 52) to make CentOS happy
- XmlComment INSIDE <t> lead to crashes. fixed.
- code changes in code that is only executed for documents in folia 1.5 format (that shouldn't exist in the wild)

(Released by Ko van der Sloot on 2017-08-30) https://github.com/LanguageMachines/libfolia/releases/tag/v1.9

libfolia v1.8

Implements FoLiA spec 1.4.3 API changed. Bumped library version to 6.2.0

- added textclass attribute
- added experimental textchecking code. only working for FoLiA documents according to spec 1.5. NOT RELEASED YET! Work in Progress
- fix in generate_id. AUTO_GENERATE_ID property was ignored.
- numerous small bug fixes

(Released by Ko van der Sloot on 2017-08-16) https://github.com/LanguageMachines/libfolia/releases/tag/v1.8

libfolia v1.7

API changed so bumped library version to 6.1.0

[Ko van der Sloot]

- textcontent() and phoncontent() return const pointers, and also work for TexContent and PhonContent elements now
- some refactoring, as suggested by CPPCHECK
- typos
- added dangerous functions to manipulate the class of a TextContent
- added reference counting on annotations. This allows to remove unneeded declarations.
- small bug fixes:
  - str() should never throw.
  - avoid memory leak

[maarten van Gompel]

- fixes in folia_properties for FoLiA spec 1.4.1

(Released by Ko van der Sloot on 2017-04-04) https://github.com/LanguageMachines/libfolia/releases/tag/v1.7

libfolia v1.6

This release implements FoLiA spec 1.4

- ABI breakage. .so name bumped to 6.0.0 reason:
  - new properties added
  - implementation of generateId() is changed
- enhancements to folialint. Saving a document with --strip also implies canonical output (--kanon)
- some bug fixes

(Released by Ko van der Sloot on 2017-01-05) https://github.com/LanguageMachines/libfolia/releases/tag/v1.6

libfolia v1.5

- Bumped the .so name. Should have been done in 1.4!
- addition: text() mebmer for document-
- minor bug fixes:
  - isNCname test now conforms to XML definition
  - improved am error messag in Document
  - check empty attributes in Feature() construction

(Released by Ko van der Sloot on 2016-11-15) https://github.com/LanguageMachines/libfolia/releases/tag/v1.5

libfolia v1.4

This version implements FoLiA spec: 1.3.2

- multiple ForeignData nodes
- added more Feature nodes, like Polarity, Strenght
- Source, Target, Relation, Predicate, Sentiment Statement, Observation Annotations and Layers.
- Comment node
- better version checking.(and a bit relaxed too) some bug fixes and code improvement.
- str() works more as expected
- fixup ref 'id' vs. 'xml:id'
- improved sanity check to better test errors in the specs.
- added a language getter and setter.

(Released by Ko van der Sloot on 2016-10-18) https://github.com/LanguageMachines/libfolia/releases/tag/v1.4

libfolia v1.3.1

Very minor release update to facilitate debian packaging (+ updated README)

(Released by Maarten van Gompel on 2016-07-29) https://github.com/LanguageMachines/libfolia/releases/tag/v1.3.1

libfolia v1.3

Maintenance release:

- long options --help and --version added
- fix in LineBreak: text() generates a newline

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/libfolia/releases/tag/v1.3

libfolia v1.2

This release adds new features from the FoLiA 1.2 specification. This includes:

- ForeignData nodes
- Foreign metadata node
- less restrictions on `aref/@type` and `ref/@type`
- fully implemented `xlink:type="simple"` and `xlink:type="locator"` Also minor bug fixes and code improvements are included

(Released by Ko van der Sloot on 2016-05-24) https://github.com/LanguageMachines/libfolia/releases/tag/v1.2

libfolia v1.0.1

Bugfix release. Fixes linker and symbol lookup failure on Mac OS X (clang). Closes issue #3

(Released by Maarten van Gompel on 2016-03-14) https://github.com/LanguageMachines/libfolia/releases/tag/v1.0.1

libfolia v1.0

Major release: Implemented on the common properties shared by both the Python and C++ versions. From now on, both implementations will support exact the same tags, attributes and syntax.

(Released by Ko van der Sloot on 2016-03-10) https://github.com/LanguageMachines/libfolia/releases/tag/v1.0

libfolia v0.13

First release of libfolia from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/libfolia/releases/tag/v0.13

NAFFoLiAPy v0.1.1

A release of the current status-quo with some latest fixes, still much to do

(Released by Maarten van Gompel on 2018-06-29) https://github.com/cltl/NAFFoLiAPy/releases/tag/v0.1.1

pynlpl v1.2.9

The FoLiA library is being migrated out of PyNLPl to a new standalone project: https://github.com/proycon/foliapy (`pip install folia`), which is backward compatible with the one in pynlpl. Retaining the old library for a transition period, but implemented warnings and notices.

(Released by Maarten van Gompel on 2019-03-13) https://github.com/proycon/pynlpl/releases/tag/v1.2.9

pynlpl v1.2.8

- Enable proper confusion matrix in case of a dissimilarity between goals and observations #43

(Released by Maarten van Gompel on 2018-11-12) https://github.com/proycon/pynlpl/releases/tag/v1.2.8

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

pynlpl v1.2.5

- Minor folia fixes

(Released by Maarten van Gompel on 2017-11-28) https://github.com/proycon/pynlpl/releases/tag/v1.2.5

pynlpl v1.2.4

Minor bugfix release

(Released by Maarten van Gompel on 2017-10-27) https://github.com/proycon/pynlpl/releases/tag/v1.2.4

pynlpl v1.2.2

- FQL update, added [DELETE...] RESTORE ORIGINAL modifier

(Released by Maarten van Gompel on 2017-10-17) https://github.com/proycon/pynlpl/releases/tag/v1.2.2

pynlpl v1.2.1

- Fixed problem with text validation and text delimiters of sentences (#34)

(Released by Maarten van Gompel on 2017-10-10) https://github.com/proycon/pynlpl/releases/tag/v1.2.1

pynlpl v1.2

- Updated FoLiA library for FoLiA v1.5
  - Implemented text validation (proycon/folia#24)
  - Added set aliases (#33, proycon/folia#31)
  - Implemented submetadata (proycon/folia#30)
  - Refactoring of metadata implementation, added `ExternalMetadata` class.
  - Fixes in parsing old inline IMDI

(Released by Maarten van Gompel on 2017-10-08) https://github.com/proycon/pynlpl/releases/tag/v1.2

pynlpl v1.1.11

- Support for FoLiA v1.4.3, implements `textclass` attribute

(Released by Maarten van Gompel on 2017-08-16) https://github.com/proycon/pynlpl/releases/tag/v1.1.11

pynlpl v1.1.10

Minor update to tie folia tests to correct FoLiA release (was missing in v1.1.9)

(Released by Maarten van Gompel on 2017-07-13) https://github.com/proycon/pynlpl/releases/tag/v1.1.10

pynlpl v1.1.9

General:

- Restructured package (into a dedicated `pynlpl` subdir rather than root dir), removed patchy stuff from `setup.py`
- Updated installation instructions
- Added `pynlpl.algorithms.possiblesplits()` FoLiA library:
- Updated for FoLiA v1.4.2
- Experimental text validation support
- always serialize XML with FoLiA version of library rather than FoLiA version from document; can't guarantee there are no changes
- `text()` method recurses in a semantically saner fashion

(Released by Maarten van Gompel on 2017-07-13) https://github.com/proycon/pynlpl/releases/tag/v1.1.9

pynlpl v1.1.8

Minor fix in syncing pynlpl-folia tests for premature v1.1.7 release

(Released by Maarten van Gompel on 2017-03-28) https://github.com/proycon/pynlpl/releases/tag/v1.1.8

pynlpl v1.1.7

Minor update:

- Fixes for Linebreak as text markup in FoLiA library (#32)
- Better support for multiple inheritance and ACCEPTED_DATA in FoLiA Library
- Added ParseError exception in FoLiA Library, more detailed error feedback when loading invalid folia
- Updated to FoLiA v1.4.1

(Released by Maarten van Gompel on 2017-03-28) https://github.com/proycon/pynlpl/releases/tag/v1.1.7

pynlpl v1.1.6

Minor bugfix release, fixes issue in folia library with underscores in NCName #31

(Released by Maarten van Gompel on 2017-03-20) https://github.com/proycon/pynlpl/releases/tag/v1.1.6

pynlpl v1.1.5

> Minor update: output more details on FoLiA RelaxNG validation failure

(Released by Maarten van Gompel on 2017-02-27) https://github.com/proycon/pynlpl/releases/tag/v1.1.5

pynlpl v1.1.4

> Bugfix release, fixes issue in folia library in declaration of new sets (#29)

(Released by Maarten van Gompel on 2017-02-21) https://github.com/proycon/pynlpl/releases/tag/v1.1.4

pynlpl v1.1.3

> Minor update:
>
> - added class to perform evaluation of ordinal classification (by @fkunneman)

(Released by Maarten van Gompel on 2017-01-12) https://github.com/proycon/pynlpl/releases/tag/v1.1.3

pynlpl v1.1.2

> Minor bugfix release, fixes issue in FQL

(Released by Maarten van Gompel on 2016-12-16) https://github.com/proycon/pynlpl/releases/tag/v1.1.2

pynlpl v1.1.1

> Minor bugfix release, fixes `previous()`/`next()` method in FoLiA library (issue #26)

(Released by Maarten van Gompel on 2016-12-12) https://github.com/proycon/pynlpl/releases/tag/v1.1.1

pynlpl v1.1.0

> - Documentation fixes (also thanks to @irushchyshyn)
> - Updated FoLiA library for FoLiA v1.4
>   - Support for RDF-based FoLiA Set Definitions, split into a separate submodule now
>   * Support for conversion of legacy XML sets to RDF
>   - Support for deep validation (#19)
>   - Numerous FQL fixes and enhancements: support for features, span roles, nested span annotations, SPAN statement

(Released by Maarten van Gompel on 2016-12-09) https://github.com/proycon/pynlpl/releases/tag/v1.1.0

pynlpl v1.0.9

> Minor update release to fix tests

(Released by Maarten van Gompel on 2016-10-18) https://github.com/proycon/pynlpl/releases/tag/v1.0.9

pynlpl v1.0.8

> - FoLiA v1.3.2 release (minor update regarding metadata, issue proycon/folia#23)

(Released by Maarten van Gompel on 2016-10-11) https://github.com/proycon/pynlpl/releases/tag/v1.0.8

pynlpl v1.0.7

> - FoLiA v1.3.1 support (minor update)
> - handle escaped newlines (\n) in FQL (relates to issue proycon/flat#51)
> - minor FQL fixes
> - Got rid of numpy dependency

(Released by Maarten van Gompel on 2016-09-29) https://github.com/proycon/pynlpl/releases/tag/v1.0.7

pynlpl v1.0.5

- Minor update release, removed old examples.
- This version has been uploaded to Debian as well as `python-pynlpl` and `python3-pynlpl`

(Released by Maarten van Gompel on 2016-09-07) https://github.com/proycon/pynlpl/releases/tag/v1.0.5

pynlpl v1.0.4

Minor update: Ship test suite (v1.0.2 and v1.0.3 were wrong)

(Released by Maarten van Gompel on 2016-09-05) https://github.com/proycon/pynlpl/releases/tag/v1.0.4

pynlpl v1.0.3

Minor update: Ship test suite (v1.0.2 was wrong)

(Released by Maarten van Gompel on 2016-09-05) https://github.com/proycon/pynlpl/releases/tag/v1.0.3

pynlpl v1.0.2

Minor update: Ship tests in python package

(Released by Maarten van Gompel on 2016-09-05) https://github.com/proycon/pynlpl/releases/tag/v1.0.2

pynlpl v1.0.1

Minor fix release: FoLiA tests now reproducible for older versions

(Released by Maarten van Gompel on 2016-09-05) https://github.com/proycon/pynlpl/releases/tag/v1.0.1

pynlpl v1.0

FoLiA Library

- Update to FoLiA v1.3, consult the FoLiA 1.3 release notes

FQL Library

- Fix (issue #17)

(Released by Maarten van Gompel on 2016-09-01) https://github.com/proycon/pynlpl/releases/tag/v1.0

pynlpl v0.9.4

- Updated API Documentation, mostly in the FoLiA library

(Released by Maarten van Gompel on 2016-08-07) https://github.com/proycon/pynlpl/releases/tag/v0.9.4

pynlpl v0.9.3

- Bugfix release for FoLiA library, fixes issue #13

(Released by Maarten van Gompel on 2016-06-22) https://github.com/proycon/pynlpl/releases/tag/v0.9.3

pynlpl v0.9.2

Minor bugfix release for FoLiA library: v0.9.1 was released prematurely: `t-style` is a primary element. Fixed

(Released by Maarten van Gompel on 2016-05-25) https://github.com/proycon/pynlpl/releases/tag/v0.9.2

pynlpl v0.9.1

> Bugfix release for FoLiA library. Fixes issue proycon/flat#33

(Released by Maarten van Gompel on 2016-05-25) https://github.com/proycon/pynlpl/releases/tag/v0.9.1

pynlpl v0.9.0

FoLiA library

- Support for FoLiA v1.2, consult FoLiA v1.2 release notes
- RelaxNG generation supports new `foreign-data` element and correctly ignores foreign namespace attributes.

FQL Library

- Fixes for `SUGGESTION SPLIT/MERGE`

(Released by Maarten van Gompel on 2016-05-24) https://github.com/proycon/pynlpl/releases/tag/v0.9.0

pynlpl v0.8.2

- Update to FoLiA v1.0.1
- generate_id is more flexible
- API documentation regenerated

(Released by Maarten van Gompel on 2016-04-01) https://github.com/proycon/pynlpl/releases/tag/v0.8.2

pynlpl v0.8.0

FoLiA library

- FoLiA v1.0.0 released
- FoLiA library is now largely constructed on the base of an external FoLiA specification (issue #9), property values have been revised significantly
- Added `AUTO_GENERATE_ID` property to indicate preference that an ID should be generated on an item
- Contains fixes in RelaxNG generation: `speech` element was missing, legacy `listitem` was not supported.

Frog client

- Fix for Python 3 (pull request #10)

(Released by Maarten van Gompel on 2016-03-09) https://github.com/proycon/pynlpl/releases/tag/v0.8.0

pynlpl v0.7.8

- Added ComplexAlignment support in FoLiA library

(Released by Maarten van Gompel on 2016-02-22) https://github.com/proycon/pynlpl/releases/tag/v0.7.8

pynlpl v0.7.7.1

(Released by Maarten van Gompel on 2016-02-15) https://github.com/proycon/pynlpl/releases/tag/v0.7.7.1

pynlpl v0.7.7

> First formal release on github

(Released by Maarten van Gompel on 2016-02-10) https://github.com/proycon/pynlpl/releases/tag/v0.7.7

Frog

Project/task/deliverable references: CLARIAH-CORE WP3 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

frog v0.14

- use TiCC::UniFilter now
- use TiCC::diacritics_filter now
- configuration modernized. OSX build supported too
- XML (FoLiA) files are autodetected
- some more logging and time stamps added
- added code to NER module to override original tags (e.g. from gazeteer)

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/frog/releases/tag/v0.14

frog v0.13.10

Bug fix release to fix a compilation problem on Max OSX

(Released by Ko van der Sloot on 2018-01-29) https://github.com/LanguageMachines/frog/releases/tag/v0.13.10

frog v0.13.9

Bug fix release, to get all our releases into balance. (Toad release requires 0.13.9)

(Released by Ko van der Sloot on 2017-11-07) https://github.com/LanguageMachines/frog/releases/tag/v0.13.9

frog v0.13.8

- added -t / --textredundancy option, which is passed to ucto
- set textclass attributes on entities (folia 1.5 feature)
- better textclass handling in general
- multiple types of entities (setnames) are stored in different layers
- some small provisions for 'multi word' words added. mblem may use them other modules just ignore them (seeing a multiword as multi words)
- added --inpuclass and --outputclass options. (prefer over textclass)
- added a --retry option, to redo complete directories, skipping what is done.
- added a --nostdout option to suppress the tabbed output to stdout
- refactoring and small fixes

(Released by Ko van der Sloot on 2017-10-26) https://github.com/LanguageMachines/frog/releases/tag/v0.13.8

frog v0.13.7

- Data files are now in `share/` rather than `etc/` (requires frogdata >= v0.13)

(Released by Maarten van Gompel on 2017-01-23) https://github.com/LanguageMachines/frog/releases/tag/v0.13.7

frog v0.13.6

- rework done on compounding in MBMA. (still work in progress)
- lots of improvement in MBMA rule handling. (but still work in progress)
  - support for 'glue' rules added.
  - support for 'hidden' morphemes added.
  - proper CELEX tags are outputted now in the XML
  - some structure labels have better names now
- removed exit() calls from library modules (issue #17)
- added languages option which is handled over to ucto too.
  - detect multiple languages
  - handle a selected language an ignore the rest

(Released by Ko van der Sloot on 2017-01-05) https://github.com/LanguageMachines/frog/releases/tag/v0.13.6

frog v0.13.5

- Added safeguards against faulty data
- Added manpage for `ner` tool (issue #8)
- Added some more compounding rules
- Read and display frogdata version

(Released by Maarten van Gompel on 2016-09-13) https://github.com/LanguageMachines/frog/releases/tag/v0.13.5

frog v0.13.4

- added long options --help and --version
- interactive use is limited to TTY's only, so pipes from std in work
- added a --language='name' option. it tries to read the configuration from a subdirectory with 'name' in the configdir The default is 'nl'
- tokenizer timing is fixed at last
- be robust against a missing clex tag
- better warning when OpenMP is not present
- adaptation in mbma
- added 2 convenience functions to FragAPI: get_full_morph_analysis() and get_compound_analysis()
- CompoundType is now in it;s own namespace
- some code refactoring, as usual

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/frog/releases/tag/v0.13.4

frog v0.13.3

New release. Now based on libfolia 1.0

(Released by Ko van der Sloot on 2016-03-10) https://github.com/LanguageMachines/frog/releases/tag/v0.13.3

frog v0.13.0

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/frog/releases/tag/v0.13.0

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

frogdata v0.15

- new enhanced NER data, including gazeteer data
- new enhanced IOB data
- numerous fixes in MBMA rules
- some updates in MBLEM

(Released by Ko van der Sloot on 2017-10-26) https://github.com/LanguageMachines/frogdata/releases/tag/v0.15

frogdata v0.13

- Data files are now in `share/` instead of `etc/` (incompatible with frog < 0.13.7)

(Released by Maarten van Gompel on 2017-01-23) https://github.com/LanguageMachines/frogdata/releases/tag/v0.13

frogdata v0.12.1

Minor bugfix release to facilitate debian packaging

(Released by Maarten van Gompel on 2017-01-06) https://github.com/LanguageMachines/frogdata/releases/tag/v0.12.1

frogdata v0.12

0.12 [Ko van der Sloot] 11-07-2016

- generally use ISO 639-3 for language codes
- quite a big update in the MBMA rules

(Released by Ko van der Sloot on 2017-01-05) https://github.com/LanguageMachines/frogdata/releases/tag/v0.12

frogdata v0.11

Minor update:

- Version number added to configuration itself
- Fixed a typo in two rules

(Released by Maarten van Gompel on 2016-09-13) https://github.com/LanguageMachines/frogdata/releases/tag/v0.11

frogdata v0.10

- The data is now stored in language dependend subdirectories (for now only /nl) for backward compatability, the dutch /nl data is copied to the root configdir too

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/frogdata/releases/tag/v0.10

frogdata v0.9

First FrogData release from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/frogdata/releases/tag/v0.9

python-frog v0.3.7

workaround for libicu 61 compatibility

(Released by Maarten van Gompel on 2018-03-29) https://github.com/proycon/python-frog/releases/tag/v0.3.7

python-frog v0.3.6

- More fixes for Mac OS X compilation

(Released by Maarten van Gompel on 2018-03-07) https://github.com/proycon/python-frog/releases/tag/v0.3.6

python-frog v0.3.5

- Fix for Mac OS X compilation (use libc++ explicitly)

(Released by Maarten van Gompel on 2018-03-03) https://github.com/proycon/python-frog/releases/tag/v0.3.5

python-frog v0.3.4

- Compatibility with Frog v0.13.9

(Released by Maarten van Gompel on 2017-11-07) https://github.com/proycon/python-frog/releases/tag/v0.3.4

python-frog v0.3.3

- Fixed compatibility with Frog 0.13.6 and above (language aware configuration files)

(Released by Maarten van Gompel on 2017-06-01) https://github.com/proycon/python-frog/releases/tag/v0.3.3

python-frog v0.3.2

Release against Frog v0.13.4 (won't compile against older)

(Released by Maarten van Gompel on 2016-07-11) https://github.com/proycon/python-frog/releases/tag/v0.3.2

python-frog v0.3.1

Setup fix, prior release was broken in PyPI. For Frog v0.13.3 (won't compile against v0.13.4 or higher!)

(Released by Maarten van Gompel on 2016-03-23) https://github.com/proycon/python-frog/releases/tag/v0.3.1

python-frog v0.3.0

Release for FoLiA v1.0, compiles against new libfolia and frog

(Released by Maarten van Gompel on 2016-03-10) https://github.com/proycon/python-frog/releases/tag/v0.3.0

python-frog v0.2.5.1

Stable release

(Released by Maarten van Gompel on 2016-02-10) https://github.com/proycon/python-frog/releases/tag/v0.2.5.1

toad v0.4

Major release:

- Added modules te create new-style datafiles with enrichments for NER and CHUNKER Needs a recent frog too.
- lots of code changes and fixes
- make more use of recent additions to ticcutils.

(Released by Ko van der Sloot on 2017-11-07) https://github.com/LanguageMachines/toad/releases/tag/v0.4

toad v0.3

This is the first release of toad. Toad is a collection of tools to help constructing and maintaining data files for frog, the most important being **froggen** It is work in progress, and not crystallized yet.

(Released by Ko van der Sloot on 2016-10-24) https://github.com/LanguageMachines/toad/releases/tag/v0.3

LaMachine

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

codemetapy v0.2.1.1

(Minor rerelease without changes just to trigger a DOI on Zenodo)

(Released by Maarten van Gompel on 2019-01-16) https://github.com/proycon/codemetapy/releases/tag/v0.2.1.1

codemetapy v0.2.1

- better failure and exit code if identifier was not found in registry

(Released by Maarten van Gompel on 2018-10-08) https://github.com/proycon/codemetapy/releases/tag/v0.2.1

codemetapy v0.2.0

- Added some simple support for converting debian package metadata from apt show to codemeta (#1)

(Released by Maarten van Gompel on 2018-09-17) https://github.com/proycon/codemetapy/releases/tag/v0.2.0

codemetapy v0.1.6

Minor bugfix release

(Released by Maarten van Gompel on 2018-08-31) https://github.com/proycon/codemetapy/releases/tag/v0.1.6

codemetapy v0.1.5

Minor update release:

- Added `--with-orcid` parameter to generate placeholders for ORCIDs in author details (#2)

(Released by Maarten van Gompel on 2018-08-30) https://github.com/proycon/codemetapy/releases/tag/v0.1.5

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 v0.2.4.1

(minor rerelease just to trigger DOI update on Zenodo)

(Released by Maarten van Gompel on 2019-01-16) https://github.com/proycon/labirinto/releases/tag/v0.2.4.1

labirinto v0.2.4

- Added `REWRITE_HOST` configuration parameter that automatically rewrites the specified hosts (if occurring in entryPoints.urlTemplate) to the one determined at run-time, also introduced a `{host}` variable that does the same explicitly (proycon/LaMachine#97)
- Adding support for CLARIN specific attributes
- Added tooltips
- Show `dateCreated` property in interface

(Released by Maarten van Gompel on 2018-06-24) https://github.com/proycon/labirinto/releases/tag/v0.2.4

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

LaMachine v2.5.0

Adds support for FoLiA 2.0, see also the release plan in proycon/folia#68

- Added new FoLiApy
- FoLiA-tools is now split from the central FoLiA repository and LaMachine is adapted accordingly. The central repository containing examples and specifications is still supplied with LaMachine, but usually not needed for most end-users.

(Released by Maarten van Gompel on 2019-03-14) https://github.com/proycon/LaMachine/releases/tag/v2.5.0

LaMachine v2.4.11

- Added FLAIR to pytorch package
- Added phonetisaurus and g2pservice (limited availability), includes openfst independent of kaldi

(Released by Maarten van Gompel on 2019-03-05) https://github.com/proycon/LaMachine/releases/tag/v2.4.11

LaMachine v2.4.10

- implemented initial support for LXC/LXD containers #134
- fix for macOS installation of uwsgi
- added Go
- added a -h/--help flag to lamachine-update
- added --only option for lamachine-update to more selectively update

(Released by Maarten van Gompel on 2019-02-06) https://github.com/proycon/LaMachine/releases/tag/v2.4.10

LaMachine v2.4.9

- Added GLEM webservice
- listen on all IPs for Jupyter Lab (fixes #129)
- fixes for continuous integration on travis
- fixed for installation on Ubuntu 18.04

(Released by Maarten van Gompel on 2019-01-16) https://github.com/proycon/LaMachine/releases/tag/v2.4.9

LaMachine v2.4.8

- Fixes lamachine update error in docker/VM #127

(Released by Maarten van Gompel on 2018-12-19) https://github.com/proycon/LaMachine/releases/tag/v2.4.8

LaMachine v2.4.7

- Set default path of docker container to home dir and create convenience symlinks to $DATA_PATH and $LAMACHINE_PATH (arose from #125)
- explicitly install python3-setuptools (debian/ubuntu)
- Travis-CI integration tests are finally on ubuntu 16.04 instead of 14.04
- Use pip instead of `python setup.py install` whenever possible, with fallbacks
- Added GLEM to LaMachine
- Improved sanity check and error feedback for possibly outdated/malfunctioning virtualenv python
- Improved Ansible log output (colors and less cruft)

(Released by Maarten van Gompel on 2018-12-08) https://github.com/proycon/LaMachine/releases/tag/v2.4.7

LaMachine v2.4.6

Minor bugfix release:

- Fix: PICCL didn't get registered suddenly?

(Released by Maarten van Gompel on 2018-11-21) https://github.com/proycon/LaMachine/releases/tag/v2.4.6

LaMachine v2.4.5

- Added Oersetter Frisian-Dutch MT (#121)
- Added Moses (#110)
- Metadata registration fixes

(Released by Maarten van Gompel on 2018-10-17) https://github.com/proycon/LaMachine/releases/tag/v2.4.5

LaMachine v2.4.4

Bugfix release:

- Added compatibility for BSD sed (on mac), not just GNU sed

(Released by Maarten van Gompel on 2018-10-12) https://github.com/proycon/LaMachine/releases/tag/v2.4.4

LaMachine v2.4.3

- Fixes in bootstrap for prebuilt VM

(Released by Maarten van Gompel on 2018-10-12) https://github.com/proycon/LaMachine/releases/tag/v2.4.3

LaMachine v2.4.2

- Added an `extra_disksize` variable for VM (`--disksize` parameter for bootstrap), configures an extra virtual disk to put LaMachine on in order to allow installation of large software that would not fit in the default disk image.
- Allow allow Tesseract German Fraktur installation to fail (not present on all distros, removed from ubuntu 18.04) should solve #117
- Do not require python-core for kaldi
- Various small fixes
- Adding fy_nl_ASR webservice #116 (not fully functional yet)
- Attempt to get kaldi working on Arch Linux still failed for now
- Added LaMachine poster presentation for CLARIN conference in Pisa

(Released by Maarten van Gompel on 2018-10-08) https://github.com/proycon/LaMachine/releases/tag/v2.4.2

LaMachine v2.4.1

Minor update release:

- Register metadata for debian packages (proycon/codemetapy#1)
- Register versions for nextflow and alpino

(Released by Maarten van Gompel on 2018-09-19) https://github.com/proycon/LaMachine/releases/tag/v2.4.1

LaMachine v2.4.0

- Integrated Jupyter Labs as a web-based notebook/IDE/terminal environment (#108)
- Included Python Course for the Humanities by @fbkarsdorp (and myself)
- Added `lamachine-add` script as a way to add software, i.e. automatically edit the installation manifest #111
- Added `lamachine-passwd` script to set passwords
- Improved documentation (README), included screenshots
- Fix in delegation to FLAT from CLAM webservices (viewer)
- Added webserver support for macOS virtualenv flavour, but still limited (uwsgi doesn't work yet, so webservices don't work) #114
  - **Upgrade notice**: If you come from an earlier installation you will need to set `webserver: true` in your LaMachine configuration for this to take effect
- Added Stanford CoreNLP (#60)
- Added wopr for macOS virtualenv flavour (i.e. added to homebrew now)
- Added extra utilities for data science (extra-utils), provides: jq, ack, graphviz, gnuplot, glances, csvkit, xmlstarlet and a lot more

- Added conll-u parse module to python-core

(Released by Maarten van Gompel on 2018-09-06) https://github.com/proycon/LaMachine/releases/tag/v2.4.0

LaMachine v2.3.0

- Always upgrade pip, as we want to rely on the newer `--upgrade-strategy only-when-needed` flag, preventing unnecessary updates of dependencies if not explicitly requested.
- An explicit `PYTHONPATH` is now set and used for global LaMachine installations (affects all VM and Docker installations). Users are recommended to start anew with a fresh LaMachine VM or Docker container if possible, as the regular `lamachine-update` will leave artefacts.
- Fixed CentOS 7 regression problem caused by v2.2.12
- FLAT integration completed #49
- Implemented (limited) custom version support (#51)
  - Custom versioning on macOS is not supported (and probably won't be supported in the future either unless there is a huge demand)
  - Custom versioning necessarily excludes packages supplied by the linux distribution's package manager. To ensure getting a LaMachine environment close to the original, ensure you build it on the same distribution.
  - Versioning of java/maven software is not supported yet (but will be in the future)
  - Some software is excluded as either upstream doesn't provide the necessary release history, or it needs more effort on my part to support that release history as it deviates from the default, examples are: Nextflow, Alpino

(Released by Maarten van Gompel on 2018-08-13) https://github.com/proycon/LaMachine/releases/tag/v2.3.0

LaMachine v2.2.12

- Added Stanford CoreNLP #60
  - Added support for pulling software from Maven Central
  - Set a default java classpath
- Fixed locale problems, set locale for docker and vagrant, always prevent C or POSIX locales #107
- Added PyTorch
- Always update pip (a proper recent pip that supports the upgrade policy parameter is needed)
- Detect if the virtualenv python diverges from the global one, and suggest reinstallation of LaMachine in that case #105
- Some Travis-CI fixes

(Released by Maarten van Gompel on 2018-08-09) https://github.com/proycon/LaMachine/releases/tag/v2.2.12

LaMachine v2.2.11

Minor fixes:

- Set $MANPATH in activation script (should solve #104)
- fixed hunspell-de package #102

(Released by Maarten van Gompel on 2018-07-15) https://github.com/proycon/LaMachine/releases/tag/v2.2.11

LaMachine v2.2.10

- added networkx to default python packages to install
- Added NAFFoLiAPy to LaMachine after first release
- Adding zip/unzip #100
- Disabling allow-unauthenticated parameter to apt module (solves #101)
- Bootstrap had no hashbang
- Adding pcre libraries to global dependencies, required in case of manual nginx compilation
- Updated kaldi metadata #32
- Added ubuntu 18.04 venv test
- Added hunspell tool and dictionaries if gecco is chosen #102

(Released by Maarten van Gompel on 2018-07-12) https://github.com/proycon/LaMachine/releases/tag/v2.2.10

LaMachine v2.2.8

Bugfix release:

- uwsgi template somehow failed to produce a newline, causing bug #95, this should fix it
- Set REWRITE_HOST for labirinto (v0.2.4) to allow rewriting hostname with the one detected at run time, allows accessing underlying services when not resolvable
- fixed bug in previous fix for #88 #94, hopefully now fixed for good
- documentation: add hostname to recommended docker command after bootstrap #97

(Released by Maarten van Gompel on 2018-06-24) https://github.com/proycon/LaMachine/releases/tag/v2.2.8

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)

42

- 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

LaMachine v2.0.0

First LaMachine v2 release

Implements phases 2 and most of 3 of the original plan (https://github.com/proycon/LaMachine/blob/master/lamachine_
This release implements (https://github.com/proycon/LaMachine/milestone/1):

- Provide a single point-of-entry bootstrap script #40

- Change base distribution (debian 9) and test other distributions #41
- Rewrite entire provisioning infrastructure through Ansible #42
- Integrate full LanguageMachines C++ software stack #43
- Integrate Python based software (Radboud) #44
- Integrate Python bindings (Radboud) #45
- Integrate Nextflow and PICCL #46
- Integrate FLAT #49 [not completely finalized]
- Write contributor guidelines #55
- Make a new LaMachine website (and updated README) #52
  - The new website is at https//proycon.github.io/LaMachine
- Set up a rich test infrastructure #54
- Set up webserver and integrate CLAM webservices #48
- Include simple analytics #50
- Include gecco + valkuil enhancement #53 [valkuil: not completely finalized]
- Integrate Alpino and T-scan enhancement #47 [tscan: not completely finalized]

(Released by Maarten van Gompel on 2018-03-29) https://github.com/proycon/LaMachine/releases/tag/v2.0.0

Miscellaneous

Project/task/deliverable references: Dependants of others

ticcutils v0.20

- PrettyPrint: added printing of pairs
- several small bug fixes
- added more tests to 'make check'
- fixed icu::namespace issues

(Released by Ko van der Sloot on 2018-11-27) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.20

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

ticcutils v0.18

- added more unicode helper classes and functions:
  - Unicode Filter class
  - diacriticsfilter()
  - Unicode Normalizer.

- small fix in Configuration Class
- Bumped library version to 4.0
- more tests added to make check.
- overhauled autoconfigure stuff.
  - MacOSX is now better supported.
  - libzlib, libbz2 and libtar ar no longer optional

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.18

ticcutils v0.17

New release due to mixup in the previous release

(Released by Ko van der Sloot on 2017-11-07) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.17

ticcutils v0.16.1

Minor fix:

- bumped the .so version to 3.0

(Released by Ko van der Sloot on 2017-11-06) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.16.1

ticcutils v0.16

- update autoconfig stuff
- added more and enhanced split() functions, including JAVA like variants
- refactoring and cleanup
- improved assertEqual()

(Released by Ko van der Sloot on 2017-10-26) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.16

ticcutils v0.15

small bug fixes:

- added newline in logging
- check on 0 pointer in SocketBasics
- added more safeguards against abuse of CommandLine and StringOps
- added more tests tot runtest

(Released by Ko van der Sloot on 2017-04-04) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.15

ticcutils v0.14

New release:

- better configure scheme. including test for zlib1g-dev
- fixed tests

(Released by Ko van der Sloot on 2016-10-18) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.14

ticcutils v0.13.2

Minor update release:

- fixed pthread checking in configure
- added getatt() member to Configuration class, to get in line with setatt() and clearatt(). lookUp() still supported
- added code to use TCP_KEEPALIVE. Not used and little tested.

(Released by Ko van der Sloot on 2016-10-18) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.13.2

ticcutils v0.13.1

Very minor release update to facilitate debian packaging (adds commit ff76ca6f33748befebe3a372c6c85e3ccdb69c28)

(Released by Maarten van Gompel on 2016-07-26) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.13.1

ticcutils v0.13

Bug fix release:

- LogStream: fixed init and LogLevel stuff. Added tests
- Configuration: add output operators
- ServerBase: removed unsupported options. (some Timbl remains removed)

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.13

ticcutils v0.12

Bug fix release.

- fixed compiler warnings when libtar-dev is missing
- in Config code:
    - trim spaces in entries
    - don't add more then 1 configDir option
    - Config::dump() is fixed
    - added a clearatt() function
- added and updated tests.
- XMLtools:
    - added a function to poll ALL Defines namespaces on a node

(Released by Ko van der Sloot on 2016-06-07) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.12

ticcutils v0.11

First release based on GIT repository.

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/ticcutils/releases/tag/v0.11

PICCL & TICCL

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

PICCL v0.7.6

- Another fix for plain text input and no ocr AND no ticcl scenario (addressed in #43)
- Clean up in the wrapper script (it's becoming too convoluted)

(Released by Maarten van Gompel on 2019-03-05) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.6

PICCL v0.7.5.1

Minor correction for missing version update only, no functional changes

(Released by Maarten van Gompel on 2019-02-28) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.5.1

PICCL v0.7.5

- Fix for plaintext input and no ocr scenario (previous fix was still wrong)

(Released by Maarten van Gompel on 2019-02-27) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.5

PICCL v0.7.4

Added Autosearch forwarder

(Released by Maarten van Gompel on 2019-02-11) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.4

PICCL v0.7.3.1

(Minor rerelease without changes to trigger a DOI update on Zenodo)

(Released by Maarten van Gompel on 2019-01-16) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.3.1

PICCL v0.7.3

At least for some PDFs, the PDF to image file convertor in PICCL, i.e. PDFimage, created spurious image files. These sometimes resulted in 'pages' of garbage. Also, when we started building this pipeline, PDFimages did not yet convert straight into tiff-format. So we also used 'convert'. Both have now been replaced by pdftoppm, which seems to produce exactly the same amount of output tiff-files as regular PDF viewers report.

(Released by martinreynaert on 2018-12-12) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.3

PICCL v0.7.2

- fixed --help flag for ocr and ticcl (was broken)
- fix for text input when skipping TICCL
- Use 300dpi instead of the default 72dpi when converting bitmaps to TIF, should reduce garbage output #45
- Minor logging improvements: output tesseract version to standard output (#45) + feedback on why frog is enabled

(Released by Maarten van Gompel on 2018-12-11) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.2

PICCL v0.7.1

- Implemented FLAT viewers (allows to delegate FoLiA output to FLAT for visualisation) #42

(Released by Maarten van Gompel on 2018-11-21) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.1

PICCL v0.7.0

- added a document CONTRIBUTE.md with contributor guidelines and technical details
- added and expanded comments to aid @martinreynaert in understanding the Nextflow pipelines
- Restructured the webservice profiles (CLAM):
    - Publish relevant output of intermediate stages for the end-user, not just a single final end-result.
    - Less duplication
    - Some small fixes
    - Removed obsolete/implicit tokeniser option for Frog
- Fixes in the wrapper script
    - Fixes for text input
    - Fix: Output did not show up for download when only OCR is enabled #40
- Updated the startserver* scripts for the piccl webservice, made them more LaMachine-aware
- Prevent accidentally feeding Nextflow's trace.txt log as input
- Report input files to stdout for some pipelines (ticcl,frog, tokenize)
- Fix in nederlab pipeline, allow untokenised folia input and add --tok option to force tokenisation
- README fix. #41 See also https://github.com/proycon/clam/issues/69

(Released by Maarten van Gompel on 2018-11-19) https://github.com/LanguageMachines/PICCL/releases/tag/v0.7.0

PICCL v0.6.3

- Better input validation #37
- Split list of enrichment steps in webinterface #33
- Minor fixes and increased debug in webservice script

(Released by Maarten van Gompel on 2018-07-12) https://github.com/LanguageMachines/PICCL/releases/tag/v0.6.3

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

PICCL v0.4.4

First careful release of the current state of PICCL so we can at least differentiate production from develpment (releases are mandatory for LaMachine v2 inclusion). Some issues are still awaiting further testing from @martinreynaert , so the stability of the ticcl pipeline is still uncertain. The DBNL pipeline for Nederlab is functional and this version corresponds with the delivered corpus enriched documents.

(Released by Maarten van Gompel on 2018-03-09) https://github.com/LanguageMachines/PICCL/releases/tag/v0.4.4

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

ticcltools v0.5

- updated configuration. also for Mac OSX
- use of more ticcutils stuff: diacriticsfilter
- added a TICCL-mergelex program
- the OMP_THREAD_LIMIT environment variable was ignored sometimes
- TICCL-unk:
    - fixed a problem in artifreq handling
    - changed acronym detection (work in progress)
    - added -o option TICCL-lexstat:
    - added TTR output
    - added -o option TICCL-indexer
    - now also handles --foci file. with some speed-up
    - added a -t option TICCL-LDcalc:
    - be less picky on a few wrong lines in the data
- added some tests
- when libroaring is installed we built roaring versions of some modules (experimental)

- updated man pages

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/ticcltools/releases/tag/v0.5

ticcltools v0.4

- first official release.
  – added functions to test on Word2Vec datafiles
  – refactoring and modernizing stuff all around

(Released by Ko van der Sloot on 2017-04-04) https://github.com/LanguageMachines/ticcltools/releases/tag/v0.4

Software Quality Guidelines

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

software-quality-guidelines v1.0

First release after initial feedback round. Release is ready to be put in practice, from which we hope to obtain more feedback. Solves or addresses the following issues:

- issue #1
- issue #2
- issue #3
- issue #4
- issue #5
- issue #6
- issue #9
- issue #10
- issue #11
- issue #12
- issue #14

(Released by Maarten van Gompel on 2016-08-02) https://github.com/CLARIAH/software-quality-guidelines/releases/tag/v1.0

Timbl & Mbt

Project/task/deliverable references: Pre-CLARIN

mbt v3.4

- added an option '--tabbed' to set the word/tag separator to TABS
- better option parsing
- added some timers for debugging

(Released by Ko van der Sloot on 2018-11-28) https://github.com/LanguageMachines/mbt/releases/tag/v3.4

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

mbt v3.3.1

Bug fix release: A 3.2.17 tarball was released with the tag v3.3 which is confusing. So we release it again as 3.3.1 with the correct tarball.

(Released by Ko van der Sloot on 2017-11-08) https://github.com/LanguageMachines/mbt/releases/tag/v3.3.1

mbt v3.3

Maintenance release

- fixed a bug due tot uninitialized memory
- code refactoring
- added OPENMP safeguards
- fixed statistics for Enriched format
- fixed issue #3 (sending manifest to cerr not cout)

(Released by Ko van der Sloot on 2017-11-07) https://github.com/LanguageMachines/mbt/releases/tag/v3.3

mbt v3.2.16

Maintenance release

- some code refactoring
- fixed a problem with the -eNL option

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/mbt/releases/tag/v3.2.16

mbt v3.2.15

First release of MBT from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/mbt/releases/tag/v3.2.15

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

mbtserver v0.11

Bug fix:

- also work with more recent ticcutils

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/mbtserver/releases/tag/v0.11

mbtserver v0.10

First release of MbtServer from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/mbtserver/releases/tag/v0.10

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

python-timbl v2018.03.07

- More Mac OS X compilation fixes.. (using homebrew's boost-python3)

(Released by Maarten van Gompel on 2018-03-07) https://github.com/proycon/python-timbl/releases/tag/v2018.03.07

python-timbl v2018.03.03

- Another small fix for compilation on Mac OS X

(Released by Maarten van Gompel on 2018-03-03) https://github.com/proycon/python-timbl/releases/tag/v2018.03.03

python-timbl v2018.02.27

- More robust setup, actively looking for boost libraries and others and including homebrew support
  for Mac OS X

(Released by Maarten van Gompel on 2018-02-27) https://github.com/proycon/python-timbl/releases/tag/v2018.02.27

python-timbl v2017.11.09

Compatibility release for timbl 6.4.10

(Released by Maarten van Gompel on 2017-11-09) https://github.com/proycon/python-timbl/releases/tag/v2017.11.09

python-timbl v2017.04.04

Compatibility release for timbl v6.4.9

(Released by Maarten van Gompel on 2017-04-04) https://github.com/proycon/python-timbl/releases/tag/v2017.04.04

python-timbl v2016.06.02

Minor bugfix release, fixes compilation problem (issue proycon/LaMachine#14) on some platforms

(Released by Maarten van Gompel on 2016-06-02) https://github.com/proycon/python-timbl/releases/tag/v2016.06.02

python-timbl v2016.03.13

Now works on Mac OS X as well (provided boost is properly installed with --with-python3/--with-python)

(Released by Maarten van Gompel on 2016-03-13) https://github.com/proycon/python-timbl/releases/tag/v2016.03.13

python-timbl v2015.09.05

Stable release

(Released by Maarten van Gompel on 2016-02-10) https://github.com/proycon/python-timbl/releases/tag/v2015.09.05

timbl v6.4.13

- added a '--limit' option to use only the most significant features.

(Released by Ko van der Sloot on 2018-11-28) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.13

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

timbl v6.4.11

Bugfix release:

- Fixed a major bug in similarity metric calculations. (Cosine and Dot product)

(Released by Ko van der Sloot on 2018-01-09) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.11

timbl v6.4.10

- allow for spaces in TABBED input (they are significant)
- corrected some typos in messages and man page
- minor code refactoring

(Released by Ko van der Sloot on 2017-11-09) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.10

timbl v6.4.9

[Ko van der Sloot]

Maintenance release:

- removed unused/non-functional functions from the API
- code refactoring. Mostly based on CPPCHECK static analyzer.
- small bugs: -e options didn't always what you expected
- added missing files in docs

[Maarten van Gompel]

- updated README.md

(Released by Ko van der Sloot on 2017-04-04) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.9

timbl v6.4.8

Maintenance release:

- code refactoring and improvement
- relying more on ticcutils
- fixed exit codes
- accept long options: --version and --help
- fix out-of-range problem in Sparse Format

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.8

timbl v6.4.7

First release of Timbl from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/timbl/releases/tag/v6.4.7

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

timblserver v1.11

Maintenance release:

- accept long options --version and --help
- improved a warning

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/timblserver/releases/tag/v1.11

timblserver v1.10

First release of TimblServer from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/timblserver/releases/tag/v1.10

Ucto

Project/task/deliverable references: CLARIAH-CORE WP3 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

python-ucto v0.4.5

workaround for libicu 61 compatibility

(Released by Maarten van Gompel on 2018-03-29) https://github.com/proycon/python-ucto/releases/tag/v0.4.5

python-ucto v0.4.4

- More fixes for Mac OS X compilation (with homebrew)

(Released by Maarten van Gompel on 2018-03-07) https://github.com/proycon/python-ucto/releases/tag/v0.4.4

python-ucto v0.4.3

- Fix for Mac OS X compilation

(Released by Maarten van Gompel on 2018-03-03) https://github.com/proycon/python-ucto/releases/tag/v0.4.3

python-ucto v0.4.0

- No need for absolute path to configuration file anymore
- Fix for Python 2.7 compatibility
- Configurable include/library paths on setup
- Better documentation

(Released by Maarten van Gompel on 2016-09-25) https://github.com/proycon/python-ucto/releases/tag/v0.4.0

python-ucto v0.3.0

Release FoLiA v1.0, compiles against new libfolia and ucto

(Released by Maarten van Gompel on 2016-03-10) https://github.com/proycon/python-ucto/releases/tag/v0.3.0

python-ucto v0.2.4

Stable release

(Released by Maarten van Gompel on 2016-02-10) https://github.com/proycon/python-ucto/releases/tag/v0.2.4

ucto v0.14.1

- fixed textcat installation problems om Debian and OpenBSD (https://github.com/LanguageMachines/ucto/issues/59)
- typo in the man page fixed

(Released by Ko van der Sloot on 2018-12-10) https://github.com/LanguageMachines/ucto/releases/tag/v0.14.1

ucto v0.14

[Ko van der Sloot]

- updated usage() and removed -S option (never used)
- make sure the right textclass is assigned to <w> nodes in FoLiA
- minor code fixes/refactorings
- added more tests
- updated man.1 page

[Maarten van Gompel]

- updated README.md

[Iris Hendrickx]

- Updated and extended the manual

(Released by Ko van der Sloot on 2018-11-29) https://github.com/LanguageMachines/ucto/releases/tag/v0.14

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

ucto v0.12

- now use the UniFilter Unicode Filter from ticcutils
- now use the UnicodeNormalizer from ticcutils
- improved configuration. Support for Mac OSX added

(Released by Ko van der Sloot on 2018-02-19) https://github.com/LanguageMachines/ucto/releases/tag/v0.12

ucto v0.11

Bug fix release:

- problems with text inside <cell> elements

(Released by Ko van der Sloot on 2017-12-04) https://github.com/LanguageMachines/ucto/releases/tag/v0.11

ucto v0.10

New release due to outdated files in the previous release.

(Released by Ko van der Sloot on 2017-11-07) https://github.com/LanguageMachines/ucto/releases/tag/v0.10

ucto v0.9.9

Minor fix:

- bumped the .so version to 3.0.0

(Released by Ko van der Sloot on 2017-11-06) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.9

ucto v0.9.8

Bug-fix release.

- fixed utterance handling in FoLiA input. Don't try sentence detection!

(Released by Ko van der Sloot on 2017-10-23) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.8

ucto v0.9.7

- added textredundancy option, default is 'minimal'
- small adaptations to work with FoLiA 1.5 specs
  - set textclass on words when outputclass != inputclass
  - DON'T filter special characters when inputclass == outputclass
- -F (folia input) is automatically set for .xml files
- more robust against texts with embedded tabs, etc.
- more and better tests added
- better logging and error messaging
- improved language handling. TODO: Language detection in FoLiA
- bug fixes:
  - correctly handle xml-comment inside a
  - better id generation when parent has no id
  - better reaction on overly long 'words'

(Released by Ko van der Sloot on 2017-10-18) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.7

ucto v0.9.6

- Moving data files from `etc/` to `share/`, as they are more data files than configuration files that should be edited. Requires uctodata >= 0.4. Should solve debian packaging issues (#18)
- Minor updates to the manual (#2)
- Some refactoring/code cleanup, temper expectations regarding ucto's date-tagging abilities (#16, thanks also to @sanmai-NL)

(Released by Maarten van Gompel on 2017-01-23) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.6

ucto v0.9.5

> Bug fix release:
>
> - updated tokconfig-generic, which is removed from the uctodata package
> - configure no longer insists on the presence of uctodata, it merely warns when missing.

(Released by Ko van der Sloot on 2017-01-06) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.5

ucto v0.9.4

> Major update
>
> - Language support
>   - added support for multiple languages
>   - auto detection of languages using textcat
> - some refactoring
>   - no more call to exit()
>   - Better logging and Warning messages
>   - some folia output improvements
> - bug fixes
>   - in passthru,
>   - issue #11

(Released by Ko van der Sloot on 2017-01-05) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.4

ucto v0.9.3

> New release, implementing recursive rule application. Check for uctodata version >= 0.2 (but works reasonable with older version)

(Released by Ko van der Sloot on 2016-09-28) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.3

ucto v0.9.2

> Minor update release to facilitate debian packaging, readme has been updated as well.

(Released by Maarten van Gompel on 2016-07-30) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.2

ucto v0.9.1

> Bug fix release:
>
> - fixed autoconfig issue

(Released by Ko van der Sloot on 2016-07-12) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.1

ucto v0.9.0

> Major update
>
> - now use uctodata for language specific information ucto itself only supports a generic tokenizer
> - interactive use now uses readline library
> - accept long options --help and --version
> - UTF16BE now works
> - better support for crooked Windows files in general
> - added a --normalize option to map tokens in a certain TokenClass to it's generic name

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/ucto/releases/tag/v0.9.0

ucto v0.8.6

> Fix in the ABRREVIATION rules to avoid spurious linebreaks

(Released by Ko van der Sloot on 2016-04-25) https://github.com/LanguageMachines/ucto/releases/tag/v0.8.6

ucto v0.8.5

Bug fix release. Fixes issue #5

(Released by Ko van der Sloot on 2016-04-25) https://github.com/LanguageMachines/ucto/releases/tag/v0.8.5

ucto v0.8.4

new release based on libfolia1.0

(Released by Ko van der Sloot on 2016-03-10) https://github.com/LanguageMachines/ucto/releases/tag/v0.8.4

ucto v0.8.2

First Ucto release from GIT

(Released by Ko van der Sloot on 2016-01-18) https://github.com/LanguageMachines/ucto/releases/tag/v0.8.2

uctodata v0.8

[Ko van der Sloot]

- separated .abr files from there main files for all Languages
- updated italian data (thanks to @texttheater)

[Iris Hendricks]

- updated abbrev files for Portuguese Turkish and French based on https://en.wiktionary.org/wiki/Category:Portuguese_abbr and https://en.wiktionary.org/wiki/Category:Turkish_initialisms.
- added full list of French abbreviations.
- added 'aub' to Dutch list

(Released by Ko van der Sloot on 2018-11-29) https://github.com/LanguageMachines/uctodata/releases/tag/v0.8

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

uctodata v0.5

- adding slightly adapted tokenizer configuration for historical dutch (a bit more conservative on splitting). INL/nederlab-linguistic-enrichment/#7

(Released by Ko van der Sloot on 2017-10-18) https://github.com/LanguageMachines/uctodata/releases/tag/v0.5

uctodata v0.4

- Data files are moved to `share/` instead of `etc/` (incompatible with ucto < 0.9.6)
- `DATE` pattern expanded (LanguageMachines/ucto#16)

(Released by Maarten van Gompel on 2017-01-23) https://github.com/LanguageMachines/uctodata/releases/tag/v0.4

uctodata v0.3.1

Bug fix release:

- fixed install problem in debian packaging using DESTDIR
- cleaned all rules from 'empty' entries (which lead to harmless warnings)

(Released by Ko van der Sloot on 2017-01-06) https://github.com/LanguageMachines/uctodata/releases/tag/v0.3.1

uctodata v0.3

major change

- new direcory structure an filenames based on ISO 693-3 language codes.

(Released by Ko van der Sloot on 2017-01-05) https://github.com/LanguageMachines/uctodata/releases/tag/v0.3

uctodata v0.2

This release is targeted at the newest ucto 0.9.3 which supports recursive application of rules. It contains a series of updates and improvements for several languages too. In general it will also work with older ucto versions, but with slightly different results.

(Released by Ko van der Sloot on 2016-09-28) https://github.com/LanguageMachines/uctodata/releases/tag/v0.2

uctodata v0.1.1

Minor release update to facilitate debian packaging (no changes to the ucto configurations whatsoever)

(Released by Maarten van Gompel on 2016-07-30) https://github.com/LanguageMachines/uctodata/releases/tag/v0.1.1

uctodata v0.1

First version of a separate data package for ucto

(Released by Ko van der Sloot on 2016-07-11) https://github.com/LanguageMachines/uctodata/releases/tag/v0.1