Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3442–3450
Marseille, 20-25 June 2022

The Ellogon Web Annotation Tool: Annotating Moral Values and Arguments

Alexandros Fotios Ntogramatzis[2,1], **Anna Gradou**[1], **Georgios Petasis**[1], and **Marko Kokol**[3]

[1]Institute of Informatics and Telecommunications,
National Centre for Scientific Research (N.C.S.R.) "Demokritos"
GR-153 10, P.O.BOX 60228, Aghia Paraskevi, Athens, Greece
[2]Department of Informatics and Telecommunications,
National and Kapodistrian University of Athens
[3]Semantika Research, Semantika d.o.o., Zagrebška 40a, 2000 Maribor
ntogram@di.uoa.gr, agradou@iit.demokritos.gr, petasis@iit.demokritos.gr
marko.kokol@semantika.eu

Abstract

In this paper, we present the Ellogon Web Annotation Tool. It is a collaborative, web-based annotation tool built upon the Ellogon infrastructure offering an improved user experience and adaptability to various annotation scenarios by making good use of the latest design practices and web development frameworks. Being in development for many years, this paper describes its current architecture, along with the recent modifications that extend the existing functionalities and the new features that were added. The new version of the tool offers document analytics, annotation inspection and comparison features, a modern UI, and formatted text import (e.g. TEI XML documents, rendered with simple markup). We present two use cases that serve as two examples of different annotation scenarios to demonstrate the new functionalities. An appropriate (user-supplied, XML-based) annotation schema is used for each scenario. The first schema contains the relevant components for representing concepts, moral values, and ideas. The second includes all the necessary elements for annotating argumentative units in a document and their binary relations.

Keywords: annotation tools, inter-annotation reliability, collaborative annotation, web-based annotation

1. Introduction

The large amount of data that have been available over the last decades, has led to the development of special tools for extracting information that can be utilised in various tasks. Especially the need for the creation of high quality annotated corpora, has raised the necessity of developing user-friendly tools to facilitate the annotation process. There are two main annotation approaches. In the semi-automatic approach, the annotator's task is restricted to review and validate a pre-annotation process that was performed in an automatic way. On the contrary, in the manual method, the annotator is responsible for the whole annotation process. Concerning the annotation tool design, there are three categories: desktop applications, browser-based solutions and distributed/collaborative applications. Each one has its own different limitations, as the first one is limited to the corpora found in a single local storage, while the second one has to deal with browser limitations. However, during the last years, a diverse ensemble of tools that explore alternative approaches that leverage the advantages of previous categories has arise.

In this paper, we present a new version of the Ellogon Web-based annotation tool that was developed and used in the context of two research programs, VAST - Values Across Space and Time[1] and DebateLab[2]. The annotation tool covers the requirement of collaboration among partners without needing a special installation process and by providing real-time collaboration facilities for annotators. This tool incorporates some upgrades and extends some functionalities (e.g. uploading formatted text, handling multiple open documents) that are crucial for both the annotation process and the improvement of the user experience for the annotators. The new features can be summarised as follows:

- A completely redesigned user interface, based on current best practices and design trends.
- Ability to integrate machine learning models and external services to supply predictions for labels (pre-annotation), or perform continuous active learning.
- Annotation Analytics: useful statistics and insights that offer a clear overview of the annotated corpora, with multi-format export functionality for further analysis in external tools.
- Document Inspection: a detailed visualisation/overview of the annotated document, with detailed searching and filtering support.

[1]https://www.vast-project.eu/

[2]https://debatelab.ics.forth.gr/

- Annotation comparison: facilities for comparing annotations across different collections, documents and raters, allowing comparison among an unrestricted number of raters/annotators, and providing visual annotation sets differences, rater agreement tables, and calculating various metrics for inter-rater reliability/inter-annotator agreement.

- Collection/Document comparison: annotation comparison functionality can be applied on any selection of Collections/Documents available to each user.

The rest of the paper is organised as follows: Section 2 presents the state-of-the-art work regarding annotation tools, while section 3 presents the outline of the architecture and focuses on the new features of our web-based annotation tool. Section 4 presents two use cases from two different research projects, which differ in their aims and annotation needs. Finally, section 5 concludes this paper and suggests some future directions.

2. Related Work

An Annotation tool is a special application that serves as an extended document viewer, which provides functionality for marking segments of textual documents or artefacts expressed in other modalities (e.g. audio, video) to perform various annotation tasks. The majority of these tools also provide a set of additional functionalities that can facilitate and automate the annotation process. We can divide annotation tools in three major categories depending on application design:

- Desktop applications: The user needs to install special software, and the annotation process takes place locally with the documents that are stored in a single machine.

- Browser-based applications: This category includes tools that operate as Web applications, typically running within a Web browser. There is no need for additional software installation, but these kind of tools usually have to deal with browser and other limitations.

- Distributed/Collaborative applications: These tools offer functionality for collaborative work with multiple annotators performing tasks in real-time and, in some cases, simultaneously, typically exploiting technologies from both aforementioned categories.

The above classification is not the unique way to divide the annotation applications, as a lot of them combine features from multiple categories and exploit the advantages of both desktop and browser-based designs with potentially collaborative features. The Ellogon language engineering platform [Petasis et al.2002] was among the first infrastructures to provide an adaptable annotation engine, which can be adapted to a wide variety of annotation tasks, through the provision of annotation schemes provided by the user, specified as declarative elements in XML. Starting as a desktop application, and later re-implemented as a Web application, the Ellogon Web Annotation Tool can support a wide range of annotation tasks through user-provided configurations. Use cases of the annotation tool developed on top of this platform are presented in this paper. The Ellogon Annotation Engine is available as both a desktop application that provides an engine with a wide range of features for linguistic annotation that runs on Windows, Linux and OSX, and as a Web application that can be installed either locally or accessed through a publicly available server[3]. It offers the ability to annotate segments of a corpus with schemes that vary in complexity - from keywords and labels to more complex schemes that highlight links among segments [Petasis and Tsoumari2012b]. It offers a graphical user interface with an annotator component that is adaptable in order to provide a layout that corresponds to the specifications of the selected annotation schema [Katakis et al.2016a, Petasis2014]. It has already been used for a variety of different annotation tasks such as annotation of part-of-speech tags and named entities [Petasis et al.2003], prosodic features [Spiliotopoulos et al.2005], semantic graphs [Fragkou et al.2008], document sections [Petasis et al.2008, Petasis and Tsoumari2012a], co-reference on aligned corpora [Tsoumari and Petasis2011], events [Petasis2012], and arguments [Petasis2014].
This same annotation engine has been used by NOMAD collaborative tool [Petasis2014] that is a desktop application that provides collaborative annotation through the use of a centralised server and instance messaging protocols. Furthermore, several features that were supported by the Ellogon annotation platform have been moved to a Web environment, thereby offering collaborative annotation in real-time in the CLARIN-EL Annotation Tool [Katakis et al.2016b].
In addition to Ellogon and Ellogon based annotation tools, there are a lot of existing solutions available to the NLP community. BRAT[4] is a real-time collaborative web-based text annotation tool for annotating text spans and relations [Stenetorp et al.2012]. Another popular solution is GATE Teamware[5], which is a web-based platform for collaborative annotation. It offers additional functionalities for project management like re-usable project templates and user roles, as well as features for better monitoring of the annotation processes like statistics and reports about annotators' activity. Thanks to these features, it fits for group annotation tasks [Bontcheva et al.2013]. Inforex[6] is a web

[3]The Ellogon Web Annotation is freely available at: `https://annotation.ellogon.org`

[4]http://brat.nlplab.org/

[5]https://gate.ac.uk/teamware/

[6]https://inforex-work.clarin-pl.eu/

system for creation and management of annotated text corpora. It offers an advanced system of access control that allows users to access the same document simultaneously. Users can also monitor the progress of their works via the control progress utilities of the system. Inforex also offers appropriate operations for calculating metrics for inter-annotation agreement. WebAnno [7] is a general purpose multi-user annotation tool that supports the creation of different user roles - annotator, curator and project manager, as well as special monitoring tools. It leverages the BRAT's visualisation system and supports cooperation with platforms and infrastructures [de Castilho et al.2014]. Label Studio [Tkachenko et al.2020 2022] is a Web-based, multimodal annotation tool, allowing customisation of annotation schemes and pre-annotation through machine-learning services integration.

Finally there are desktop annotation tools like the KIM Semantic Annotation Platform [Popov et al.2004], the SHOE Knowledge Annotator [Heflin et al.1999], Callisto [8], Wordfreak [9], MMAX28 [Müller and Strube2006], Knowtator [10] [Ogren2006], and AeroSWARM [Corcho2006]. There are also extensions (e.g. A.nnotate [11], Bounce [12], Diigo [13], iComment, MyStickies[14], AnnotateIt) that run inside Web browsers and offer the ability to annotate web pages. An fairly recent extensive review and comparison of several annotation tools for manual text annotation can be found in [Neves and Ševa2019].

This paper presents the latest re-implementation of the Ellogon Web Annotation Tool, a modern web-based upgrade of the CLARIN-EL tool offering a user-friendly solution for collaborative annotation and powerful curation and management features, like machine-learning assisted pre-annotation, annotation analytics, document and annotation comparison functionalities, and inter-annotator reliability metrics. Its unique features include the annotation of long documents (e.g. annotation of whole theatrical plays, like ancient Greek tragedies and comedies) and the ability to add relations between annotation segments, which are visualised as arrows on top of the text, without limiting display to text fragments, allowing the user to explore a document in its entirety.

3. The Ellogon Web Annotation Tool

3.1. Architecture

The new web-based annotation tool presented in this paper is based on the CLARIN-EL annotation tool and leverages a set of REST Web services built upon the Ellogon language engineering platform. Therefore, the structure of its architecture presents several similarities with CLARIN-EL annotation tool. The new version is implemented using the most recent state of the art frameworks for front-end and back-end web development to allow users to perform annotation tasks without having to install additional software on their computers. It should be noted that this annotation tool is not a simple re-implementation of the CLARIN-EL annotation tool, merely using different web development technologies. Some of the existing functionalities have been significantly enhanced, and several new features have been added. Additionally, it provides a more modern and attractive interface that improves the user experience significantly. More details on new functionality and features are provided in section 3.3 .

The User Interface (UI) is implemented in Typescript using the Angular[15] framework. A set of REST web services is implemented on Python Django[16] Framework for user management and collection and document handling (e.g. collection creation/share/export, document upload), also supporting pre-annotation through Python components (that can leverage deep learning support through various frameworks), or external REST services. In order to perform a set of actions such as loading the graphical user interface that derives from an annotation schema, several powerful services were developed on top of the Ellogon language engineering platform, whose annotation engine is still responsible for implementing the visual representation as a UI of user-provided annotation schemes. The UI component receives declarations that are converted into UI components with interactive elements (e.g. buttons, combo-boxes, check-boxes, text fields etc.). The elements of these components are used for the annotation creation. As the annotation data is primarily shallow (no deep links/relations), unstructured data, a NoSQL MongoDB[17] was chosen to allow for easy horizontal scalability and performance scaling, even for a large number of annotations and documents database. [Kalayda2021] The rest of the data are stored in an traditional SQL relational database. The Ellogon Web Annotation Tool is publicly available as open-source on Github[18], and freely accessible from [19].

3.2. New features of the Ellogon Web Annotation Tool

The Ellogon Web Annotation Tool maintains the same advantages and novel aspects of the Ellogon-based family of tools. Furthermore, some of the existing functionalities of CLARIN-EL tool have been extended, and some new features have been added.

[7] https://webanno.github.io/webanno/
[8] https://mitre.github.io/callisto/
[9] http://wordfreak.sourceforge.net/
[10] http://knowtator.sourceforge.net/
[11] http://a.nnotate.com/
[12] https://bounceapp.com
[13] https://www.diigo.com/
[14] https://www.mystickies.com/
[15] https://angular.io/
[16] https://www.djangoproject.com/
[17] https://www.mongodb.com/
[18] https://github.com/iit-Demokritos/clarin-el-annotation-tool
[19] https://annotation.ellogon.org/

Firstly, we focus on the existing functionalities that have been extended in order to improve the user experience:

- Upload formatted text: Apart from plain text documents, users can upload TEI XML documents, where the available metadata is used as formatting information.This extension facilitates users annotating their documents, especially long documents, such as a theatrical play.

- Handling multiple opened documents (documents with unsaved changes): Users can have multiple documents with unsaved changes. Before opening a new document, the tool will inform the annotator about the current state of his/her opened documents. If the user is the document owner, he/she can discard the unsaved changes, save them or skip taking action. If the opened document is owned by another user, the annotator cannot discard the unsaved changes. Users can also close opened documents via the collection management panel. This modification improves the user experience in collaborative annotation and firmly ensures the robustness and annotation data integrity in case of internet connection or browser problems.

In order to facilitate and enhance the annotation process, available functionality was extended with more features. The novel aspects of the tool make good use of the observations and the needs of experienced annotators and meet the needs of annotators' groups to a better extent. The following new features facilitate the collaboration among multiple annotators and offer an inspection and validation process for the annotated corpora. Following we present in more detail the most crucial features:

- Document Analytics: The user can select a document and a specific annotation schema. Then, there is the possibility to see an overview of the annotation data in various data visualisation options (e.g. table, charts etc). There is also the possibility of exporting the data in various formats (like CSV/EXCEL/JSON files) or image formats for the case of the charts.

- Document Inspection: In order to offer the possibility of a document-centered overview that can include multiple annotation schemes, we added a functionality that offers a general document inspection. The user interface of the document inspection component is split horizontally into three parts. The two largest parts are: a table with the annotations of the selected document and a viewer in which the content of the document is displayed. By clicking on the table entry, the selected annotation is visualised on top of the content. By double clicking on the table entry, the user can see a drop down panel with details about the selected annotation, and have a complete overview of the specific case. Finally, the tool offers a query builder, fully customisable, and designed in a user-friendly way. This query builder serves as a powerful filter that can combine multiple conditions that concern the annotation schemes or other information of the text.

- Annotation Comparison: New functionality was added to facilitate the comparison of the annotations that appear within one single document. On top of the screen, there is a toolbar with which the user can select a document for inspection. The rest of the screen is a component with two tabs. In the first, an interface that is divided vertically into multiple parts. The user decide how many parts are necessary in order to perform the desired task by adding or removing them. Each part has a query builder that serves as a powerful filter that can combine multiple conditions in a way similar to what was described above. The rest of the screen has a table with the annotation information and a document viewer that visualises the selected annotation on top of the text. Furthermore, by double clicking on the annotation entry, the user can see additional information that are related to the current selection. There is also a tab where the user can see the rating and agreement tables for the selected document. An example of a rating table is given in figure 1. Finally, on the third tab, the annotator can see the calculated values of several inter-rater reliability metrics (e.g. Fleiss Kappa, Cohen Kappa, Krippendorff's alpha).

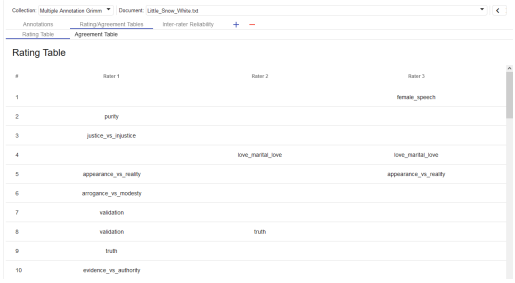

Figure 1: Rating table of a task involving three annotators.

- Collection/Document Comparison: In order to provide the ability to perform comparisons among different documents, we designed and implemented a new feature that offers an overview of multiple documents and their annotations on the same screen. More specifically, the UI of this component shares the same layout and functionality that was described above. Additionally, the annotation tables are sorted in a row depending on the part of the document in which they appear,

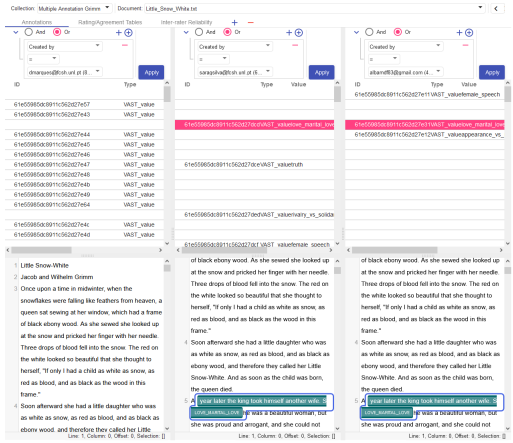

Figure 2: Comparing annotations created by different annotators for the same content.

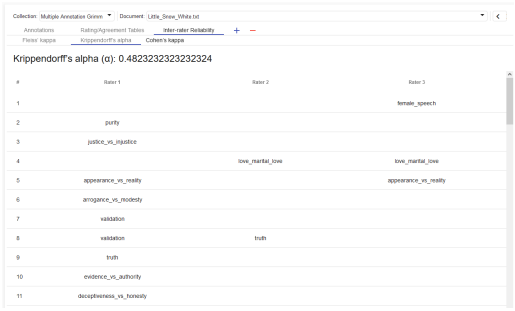

Figure 3: Calculation of inter-rater reliability metrics.

forming a visual "diff" on annotations. The sorting concerns the annotations of all the documents. If a span is not included in the annotations of a specific document, the row remains empty in order to provide a visual clue for the user that performs the comparison.

3.3. Comparison to other annotation tools

Before developing the updated annotation tool we performed an analysis of available tools based on a number of requirements: whether the tools were release under open source licenses, what technology stack they used, the primary type of the application (web or desktop), whether they support collaborative annotation and whether this is supported in real-time, the level of role management supported, support for progress monitoring, inclusion of annotation statistics, the ability to include pre-annotators using machine learning (ML) or other approaches, the ability to calculate inter-annotator agreement, the ability to compare annotations (including the ability for the tool to mark the differences), whether they support the annotation of longer texts as a single task, and whether labels can be extended in real-time by annotators. The comparison is presented in Table 1.

4. Use Cases

The new, upgraded version of the Ellogon Web Annotation Tool has been used in two research projects for annotating a plethora of different documents and for creating annotated corpora for different research purposes. Following, we present one typical use case for each research project. First, we briefly describe the project, the aim and the purposes of the research, the annotation schema that was used, and finally we give an example that demonstrates the way the new tool satisfies the needs of the project and facilitates the collaboration among the partners.

4.1. Use case: VAST Project

4.1.1. Project Description

VAST (Values Across Space & Time) aims to study the transformation of moral values, such as freedom, democracy, equality, tolerance, etc., across space and time. Values are the trails of our common legacy, our collective memory, the way we think about ourselves and the others. Values are historically dynamic, they travel through material culture (artefacts, books, scientific instruments etc.), they are appropriated in different places and times by different people, and they re-emerge in new cultural forms.

The project aims to analyse narratives expressed in natural language and trace the emerging values. The analysed narratives were decided to include three different genres: art, science and folklore. More specifically, three different pilots were designed and each one covers a different selection of documents: ancient Greek drama, 17^{th} Century Scientific Revolution texts and fairy tales. This specific use case involves the analysis and annotation of a corpus of 54 documents, including tragedies of Aeschylus, Sophocles, and Euripides, Aristophanes's comedies, documents like Kepler's "Dissertatio" and Copernicus's "De revolutionibus" and selected tales by the Grimm brothers.

4.1.2. Annotating values

The VAST project aims to analyse the aforementioned corpus with respect to values. Therefore, the designed annotation schema includes a set of "keywords" (annotation labels) that represent concepts, moral values and ideas that are being tracked down the texts. These keywords are organised under three categories: Key/Main Concepts/ Values/Ideas, Expanded/Various Content and Opposite Concepts/Conceptual Couples/Bi-polarities. Also, the tool offers the opportunity to the users to enrich the existing annotation schema by adding custom values via a user-friendly interface. In addition to the annotation information, document metadata/attributes that provide details about the document were added: Document Title, Document Synopsis, Relations/Conditions/Bi-polarities, Ideas/Values/Messages, Emblematic/Archetypical Timeless Values, Entities/persons, Interpretation/Content, Relevant values in other (similar or not) texts and Comments.

	BRAT	Clarin-EL	Ellogon Annotation Platform	GATE Teamware	Label Studio	WebAnno
Open Source	Yes	Yes	Yes	Yes	Yes (Enterprise Features may not be included)	Yes
Tech Stack	Python, HTML, JavaScript	PHP, HTML, JavaScript	Python, Angular, HTML	Java	Python, HTML, Javascript	Java
Application Type	Web	Web	Web	Desktop, Web	Web	Web
Collaborative Annotation	Yes	Yes	Yes	Yes	Yes	Yes
Real-time Collaborative Annotation	Yes	Yes	Yes	No	No	No
Role Management	Basic	Basic	Basic	Advanced	Advanced in Enterprise (Paid)	Advanced
Progress Monitoring	No	No	No	Yes	Yes	Yes
Annotation Statistics	No	No	Yes	Yes	Yes	Yes
Automatic Annotation	Yes	No	Yes	Yes	Yes	Yes
Inter-annotator Agreement	Plugin	No	Yes	No	Enterprise (Paid)	No
Annotation Comparison	Partial	No	Yes	No	Enterprise (Paid)	Partial
Long Text Annotation	No	Yes	Yes	No	No	No
Real-time Schema Extension	No	Yes	Yes	No	No	No

Table 1: Feature comparison of existing annotation solutions.

In figure 4, we present a typical example of text annotation in the context of the VAST project use case. On the right panel, the UI is adjusted to the used annotation schema. Also, the user can add custom labels and enrich the annotation schema by using the label creation button at the bottom of the screen. On the left panel, the text of "Cinderella" tales is displayed, along with some annotations. The user can read the text in its entirety and click on the colored segments to highlight them, in order to see annotation data and edit its details. An important feature of this application is the navigation through overlapping annotations (if there are any) by using the combo-box that exist at the bottom of the UI.

4.2. Use case: DebateLab Project

4.2.1. Project Description

During the last years, the web has been changed radically because of the rapid development of social media, blogs and other sites that offer the opportunity of expressing user ratings, opinions, and comments about various topics. Nowadays, the web is not only a simple resource of information and knowledge but it is also enriched with a vast collection of views and arguments. The machines cannot interpret these opinions and arguments because they are published in a disorganised and unstructured way. Therefore, it is not feasible to find logical correlations between them in order to relate them to particular topics. In this project,

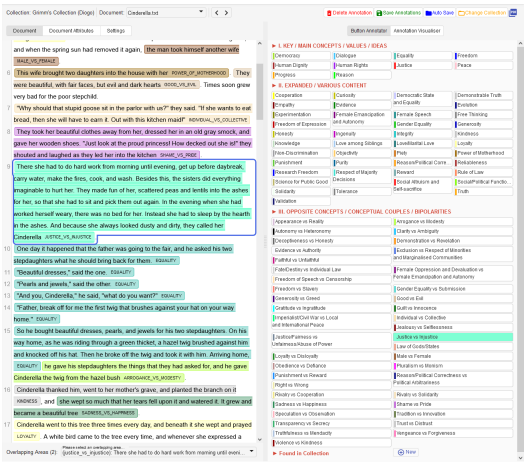

Figure 4: Annotating a tale with values.

DebateLab leverages state-of-the art approaches from fields, such as Computational Argumentation, Natural Language Processing, Information Retrieval, Machine Learning, and the Semantic Web in order to develop a suite of tools and services for e-journalism and the appropriate infrastructures for representing, mining and reasoning with online arguments.

4.2.2. Annotating and Interlinking Arguments

Argumentation is an indispensable part of human communication and its use is very frequent in articles (and of course also in other types of documents). Thus, it has evolved into an interdisciplinary field of scientific research. The DebateLab project aims to deal with arguments, extracted in nearly real-time from existing sources, such as news feeds. An important aspect of this project, is the establishment of models for argument mining, to support the automated, machine-learning based, methods for identifying arguments in written texts.

In this project, a large set of articles from a variety of different web sources (e.g. online newspapers, user generated content) has been collected, via an automated crawling process. The These articles have been pre-processed to extract the content of interest. Then, articles that are argumentative have been identified, and have been classified in predefined categories (i.e., views, opinions, journalistic surveys, analyses, blogs, letters, etc.). All the gathered argumentative articles and resources have been manually analysed in order to identify argumentative units (e.g. claims and premises), to classify them into predefined types according to certain specifications and to identify relations among them in order to reconstruct the related arguments. For this analysis, a new annotation schema has been defined in the Ellogon Web Annotation Tool.

An argument typically consists of several statements. In its simplest form it includes one claim that is supported by at least one premise. According to the modelling represented in the annotated schema, there are three types of Argumentative Discourse Units (ADUs): “major claim”, “claim” and “premise”. In an article, it is assumed that there is the major claim, that is often considered a thesis statement and represents the viewpoint of the author about the article topic. Usually, the major claim is present in the introduction, the conclusion of an article or in both. In the introduction it has the characteristics of a general assertion or an opinion with respect to the topic, whereas in the conclusion the major claim summarises the argumentation according to the author’s stance.

A claim is a direct support (or refutation) of the author’s viewpoint. In other words, it is a direct reason given in support (or attack in the case of a counter argument) of the major claim. Generally, a claim is usually supported with one or several reasons/premises. In the introduction or the conclusion, a claim appears as a direct reason of the major claim.

A premise is a reason given for supporting or attacking an argumentative discourse unit. Therefore, it can be regarded as an explanation or a refutation for persuading a reader for the truth or the falsity of a claim. A premise is always connected to another ADU (claim or premise). Several premises that are linked together in order to support a claim form a reasoning chain.

The new annotation schema developed for this use case, allows the user to annotate segments of texts as “major claim’, “claim” or “premise’. After annotating argument components, users can link them with argumentative relations. There are 6 types of relations. Each type is represented by a pair of combo-boxes (one for the source argument component and another one for the target argument component). These types can be grouped into two categories simple and complex relations. In the simple relation group, there are two types:

- Support Relations: A support relation between two argument components indicates that the source component is a reason or a justification of the target relation. Therefore, only premises are valid components for source combo-box and claims are valid components for target combo-box. (The annotation schema specification language allows the inclusion of this kind of restrictions).

- Attack Relations: An attack relation between two argument components indicates that the source component is a refutation or a rebuttal of the target relation. Therefore, only premises are valid components for source combo-box and claims are valid components for target combo-box.

In the relations of the second group, the simple relations (attack, support) are elaborated and combined with the stance of the author. For the source combo-box, only claims are valid argument components. For the target combo-box, only major claims are valid argument components:

- Support / For/Pro Relations
- Attack / Against Relations
- Support / Against Relations
- Attack / For/Pro Relations

In figure 5, we present a typical example of text annotation in the context of the DebateLab project. Apart from displaying the annotations on the text, there is also a visualisation of the relations between argumentative units. Each relation is represented as a labeled arrow. The annotator can toggle the visibility of the relations according to his/her needs, or select a "router" that manages how arrows are routed over annotated segments.

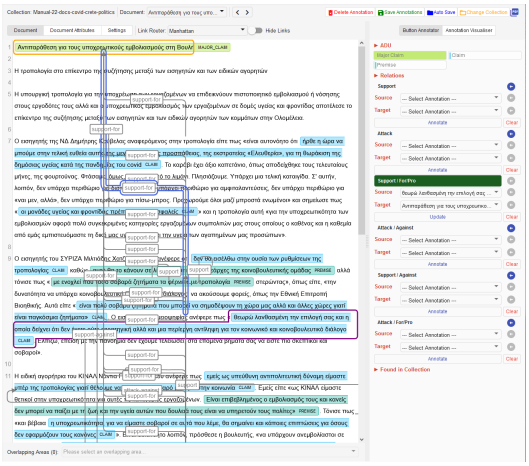

Figure 5: Annotating and interlinking arguments of an article.

5. Conclusions and Future Work

In this paper, we presented an extended version of CLARIN-EL Web based annotation tool. The new tool is named as the "Ellogon Web Annotation Tool" and it has a redesigned, user-friendly UI that offers various additional functionalities for advanced annotation tasks and support for complex schemes. It allows users to upload documents in several formats, such as documents in TEI XML, with the ability to extract formatting information from available metadata. Also, it improves the user experience in real-time collaborative annotation by proving a complete system of handling open documents with unsaved annotations in progress. Additionally, it provides a range of new features for inspecting and comparing annotated documents, and calculate inter-rater reliability metrics. Moreover, it offers the opportunity of exporting statistics of document's annotations in various formats. Finally, this version of the Web based annotation tool has been used and tested in the context of the "VAST" and "DebateLab" research projects, with a use case briefly described for each project.

As future work, we aim to include even more inter-rater reliability metrics. Moreover, we intent to implement an appropriate UI for managing annotation schemes. We are examining the possibility to use python packages for performing pre-annotation tasks and finally, we intend to implement an OpenAPI REST framework for managing collections, documents, annotations and other data, offering the ability to automate pre-annotation even by external tools.

6. Acknowledgments

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme, in the context of VAST project, under grant agreement No 101004949. In addition, this research was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to support Faculty Members and Researchers and the procurement of high-cost research equipment" grant (Project Number: 4195). This paper reflects only the view of the authors and the European Commission is not responsible for any use that may be made of the information it contains.

7. Bibliographical References

Bontcheva, K., Cunningham, H., Roberts, I., Roberts, A., Tablan, V., Aswani, N., and Gorrell, G. (2013). Gate teamware: a web-based, collaborative text annotation framework. *Language Resources and Evaluation*, 47(4):1007–1029.

Corcho, O. (2006). Ontology based document annotation: trends and open research problems. *Int. J. Metadata Semant. Ontologies*, 1:47–57, January.

de Castilho, R. E., Biemann, C., Gurevych, I., and Yimam, S. M. (2014). Webanno: a flexible, web-based annotation tool for clarin. *Proceedings of the CLARIN Annual Conference (CAC) 2014*, October.

Fragkou, P., Petasis, G., Theodorakos, A., Karkaletsis, V., and Spyropoulos, C. (2008). Boemie ontology-based text annotation tool. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

Heflin, J., Hendler, J., and Luke, S. (1999). Shoe: A knowledge representation language for internet applications. Technical report.

Kalayda, A. V. (2021). Promising directions for the development of modern databases. *Journal of Physics: Conference Series*, 2131(2):022087, dec.

Katakis, I. M., Petasis, G., and Karkaletsis, V. (2016a). CLARIN-EL web-based annotation tool. In Nicoletta Calzolari, et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož,*

Slovenia, May 23-28, 2016. European Language Resources Association (ELRA).

Katakis, I. M., Petasis, G., and Karkaletsis, V. (2016b). CLARIN-EL web-based annotation tool. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4505–4512, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Müller, C. and Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. In Sabine Braun, et al., editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.

Neves, M. and Ševa, J. (2019). An extensive review of tools for manual annotation of documents. *Briefings in Bioinformatics*, 22(1):146–163, 12.

Ogren, P. V. (2006). Knowtator: A protégé plug-in for annotated corpus construction. In Robert C. Moore, et al., editors, *HLT-NAACL*. The Association for Computational Linguistics.

Petasis, G. and Tsoumari, M. (2012a). A New Annotation Tool for Aligned Bilingual Corpora. In Petr Sojka, et al., editors, *Text, Speech and Dialogue*, volume 7499 of *Lecture Notes in Computer Science*, pages 95–104. Springer Berlin Heidelberg, Brno, Czech Republic, September 3–7.

Petasis, G. and Tsoumari, M. (2012b). A new annotation tool for aligned bilingual corpora. In *Text, Speech and Dialogue*, pages 95–104. Springer.

Petasis, G., Karkaletsis, V., Paliouras, G., Androutsopoulos, I., and Spyropoulos, C. D. (2002). Ellogon: A New Text Engineering Platform. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 72–78, Las Palmas, Canary Islands, Spain, May 29–31. European Language Resources Association.

Petasis, G., Karkaletsis, V., Paliouras, G., and Spyropoulos, C. D. (2003). Using the Ellogon Natural Language Engineering Infrastructure. In *Proceedings of the Workshop on Balkan Language Resources and Tools, 1st Balkan Conference in Informatics (BCI 2003)*, Thessaloniki, Greece, November 21. http://labs-repos.iit.demokritos.gr/skel/bci03_workshop/.

Petasis, G., Fragkou, P., Theodorakos, A., Karkaletsis, V., and Spyropoulos, C. D. (2008). Segmenting HTML pages using visual and semantic information. In *Proceedings of the 4th Web as a Corpus Workshop (WAC-4), 6th Language Resources and Evaluation Conference (LREC 2008)*, pages 18–24, Marrakech, Morocco, June 1. Proceedings: The 4th Web as Corpus: Can we do better than Google? `http://www.lrec-conf.org/proceedings/lrec2008/workshops/W19_Proceedings.pdf`.

Petasis, G. (2012). The SYNC3 Collaborative Annotation Tool. In *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, pages 363–370, Istanbul, Turkey, May. European Language Resources Association.

Petasis, G. (2014). Annotating arguments: The nomad collaborative annotation tool. In Nicoletta Calzolari, et al., editors, *LREC*, pages 1930–1937. European Language Resources Association (ELRA).

Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., and Kirilov, A. (2004). KIM - a semantic platform for information extraction and retrieval. *Journal of Natural Language Engineering*, 10(3-4):375–392.

Spiliotopoulos, D., Petasis, G., and Kouroupetroglou, G. (2005). Prosodically Enriched Text Annotation for High Quality Speech Synthesis. In *Proceedings of the 10th International Conference on Speech and Computer (SPECOM-2005)*, pages 313–316, Patras, Greece, October 17–19.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107. Association for Computational Linguistics.

Tkachenko, M., Malyuk, M., Holmanyuk, A., and Liubimov, N. (2020-2022). Label Studio: Data labeling software. Open source software available from https://github.com/heartexlabs/label-studio.

Tsoumari, M. and Petasis, G. (2011). Coreference Annotator - A new annotation tool for aligned bilingual corpora. In *Proceedings of the Second Workshop on Annotation and Exploitation of Parallel Corpora (AEPC 2), in 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011)*, pages 43–52, September 15.