CLARIN-EL Web-based Annotation Tool

Ioannis Manousos Katakis, Georgios Petasis, Vangelis Karkaletsis
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
{gkatakis, petasis, vangelis}@iit.demokritos.gr

Abstract

This paper presents a new Web-based annotation tool, the "CLARIN-EL Web-based Annotation Tool". Based on an existing annotation infrastructure offered by the "Ellogon" language enginneering platform, this new tool transfers a large part of Ellogon's features and functionalities to a Web environment, by exploiting the capabilities of cloud computing. This new annotation tool is able to support a wide range of annotation tasks, through user provided annotation schemas in XML. The new annotation tool has already been employed in several annotation tasks, including the anotation of arguments, which is presented as a use case. The CLARIN-EL annotation tool is compared to existing solutions along several dimensions and features. Finally, future work includes the improvement of integration with the CLARIN-EL infrastructure, and the inclusion of features not currently supported, such as the annotation of aligned documents.

Keywords: Annotation tools, collaborative annotation, adaptable annotation schemas, Web-based annotation tools

1. Introduction

The large volumes of data that are being published on the Web every day, create the need for powerful systems able to discover information contained in them. Usually, such kind of systems exploit supervised machine-learning techniques, requiring some form of guidance through examples, which are quite often extracted from manually annotated corpora. Although the production of this kind of corpora is a demanding and time-consuming process, it can be significantly facilitated by the use of user-friendly and intuitive annotation tools.

In this paper, we present a new Web-based annotation tool which can cover a wide range of annotation tasks, through its support to end-user provided annotation schemas, defining the graphical elements, the tags, and the restrictions that can be used to manually annotate a corpus. The proposed annotation solution is based on the relevant desktop, cross-platform annotation tool and the annotation engine provided by the Ellogon language engineering platform[1] (Petasis et al., 2002), and attempts to transfer many of its features in a Web environment by exploiting the capabilities of cloud computing.

Ellogon's annotation engine is a framework that supports the creation of a wide range of annotation tools, either desktop applications such as (Fragkou et al., 2008) or collaborative tools like (Petasis, 2012a) and (Petasis, 2014) that distribute annotation through a centralised server. The range of annotation tasks that can be implemented include any annotation task that can be modelled through the selection of multiples text ranges, and group such ranges in annotations, along with user provided information. Under such a model, a wide range of annotation tasks can be accomplished, from morphological annotation (i.e. part-of-speech annotation), to phrase annotation (from syntactic annotation to named entities, sentiment, co-reference, etc.), and up to semantic annotation, where documents can be annotated with a semantic model (such as an OWL ontology) (Fragkou et al., 2008). The infrastructure presented in this paper aims to port this annotation engine in a Web environment, while retaining the ability to customise the generated annotation tools to a wide range of annotation tasks, through annotation schemas described in XML.

The rest of the paper is organised as follows: Section 2 presents relevant work in area of Web-based annotation tools, while section 3 presents the annotation tool in detail, including its architecture and novel aspects. Section 4 presents a comparison of the CLARIN-EL Web-based annotation tool with existing approaches, while section 5 presents a use case, which involves the annotation of a corpus with arguments. Finally, section 6 concludes this paper and presents some future directions.

2. Related Work

During the last decade, a large number of annotation tools has been developed. Each one of them, is built upon it's own logic and provides a different set of features, while some of them exploit previous experience acquired from their equivalent desktop versions. GATE Teamware[2] is an annotation solution which aims to facilitate the annotation process among teams, by leveraging its distributed architecture (Bontcheva et al., 2013). It offers a desktop application which enables users to add annotations, as well as a Web-based user interface from which the users are able to manage their projects and monitor their statistics.

Another popular annotation solution is BRAT[3], a Web-based tool for NLP-assisted text annotation. Its users are able to access and annotate their collections through their browsers, without the need of installing any additional software (Stenetorp et al., 2012). BRAT also offers collaboration features, meaning that two or more users have the ability to add and modify annotations in the same document, simultaneously. The changes take place in real time and everyone have access to the latest version of the document.

[1] http://www.ellogon.org

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

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

WebAnno[4] follows the design philosophy of BRAT but it differentiates in multiple ways. It is a Web-based annotation solution which combines BRAT's visualisations with a fully-fledged back-end and delivers features like user and quality management, monitoring tools as well as an interface to crowd-sourcing (de Castilho et al., 2014). Moreover, offers a library of predefined schemas for various annotation tasks, and it supports different corpora formats enabling the cooperation with various existing platforms and infrastructures.

Inforex[5] is a Web-based system which facilitates the management and creation of annotated corpora. Its users are able to browse and edit the content of the annotated documents as well as to pre-process them. In addition, it integrates an advanced versioning system allowing users to revert every document of their collections in a previous state. Regarding the annotation process, it offers a number of predefined annotation schemas which can be customised according to the needs of each annotation task.

The Ellogon language engineering platform[6] (Petasis et al., 2002) provides an all-in-one desktop solution and an annotation engine, which allows the annotation of a wide range of information, ranging from information about words to complex annotation schemas involving links between aligned segments in bilingual texts (Petasis and Tsoumari, 2012b). In addition it supports collaborative/distributed annotation, where the annotation process can be shared among different annotators at different locations. Last but not least, it is an open source software which can be customised according to the requirements of each annotation task, exploiting a customisable engine for generating different layouts and user interfaces, driven by XML annotation schemas (Petasis, 2014). It has been applied to a wide range of tasks, ranging from 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 Tsoumari, 2012a), co-reference on aligned corpora (Tsoumari and Petasis, 2011), events (Petasis, 2012b), and arguments (Petasis, 2014).

The tool presented in this paper, is an attempt to move a number of features introduced by the Ellogon's annotation engine and tools in an intuitive and user-friendly Web environment, focusing on the improvement of the user experience during the annotation process.

3. The CLARIN-EL Annotation Tool

3.1. Architecture

In order to provide a rich and unobstructed user experience, we combined state-of-art technologies along with powerful services built upon the Ellogon language engineering platform, which expose various aspects of its annotation engine. The proposed annotation tool exploits the power of cloud computing in order to allow users to perform annotation tasks without the need of any kind of software. On top of that, it guarantees the consistency and synchronisation of user data across multiple devices. Even if the internet connectivity is lost, the user is able to continue the annotation task without losing the annotation progress. The architecture of the CLARIN-EL annotation tool is shown in Figure 1.

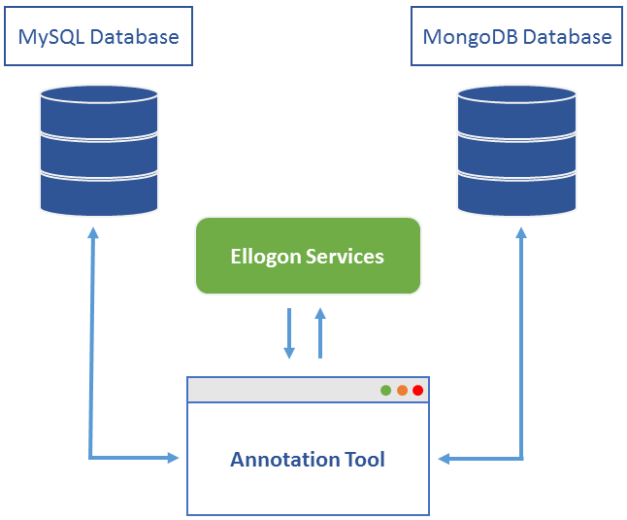

Figure 1: The architecture of CLARIN-EL annotation tool.

The annotation tool is written in JavaScript using the AnguarJS Framework[7], and the Laravel PHP Framework[8] as its backend. The core component of the CLARIN-EL annotation tool is a set of REST Web services developed on top of the Ellogon language engineering platform. These services are being used from the annotation tool in order to access the information of the stored annotation schemas, to derive a graphical user interface from the annotation schema, to validate user-uploaded schemas, and to interface the annotation tool with other existing infrastructures through suitable data import/export facilities. When a user selects an annotation schema, a template with all the included graphical components is served in HTML format. The generated HTML template corresponds to the user interface defined by the selected annotation schema. Afterwards, AngularJS transforms the HTML template into an interactive user interface and provides all the needed functionality for performing the annotation process.

The data of the Web application are stored in two different databases. A MySQL database is used to store the data of the users along with their collections of documents, while a MongoDB database is used to store the annotations which are being added by the users on documents. Considering the large amount of annotation-related data, we decided to store them in a NoSQL database in order to achieve better performance as well as more effective load balancing. Finally, the annotation tool supports recent versions of all major Web browsers[9].

[4] https://webanno.github.io/webanno
[5] http://nlp.pwr.wroc.pl/inforex
[6] http://www.ellogon.org
[7] https://angularjs.org/
[8] https://laravel.com/
[9] Recent versions of Google Chrome, Mozilla Firefox, Microsoft Edge, Opera and Apple Safari are supported.

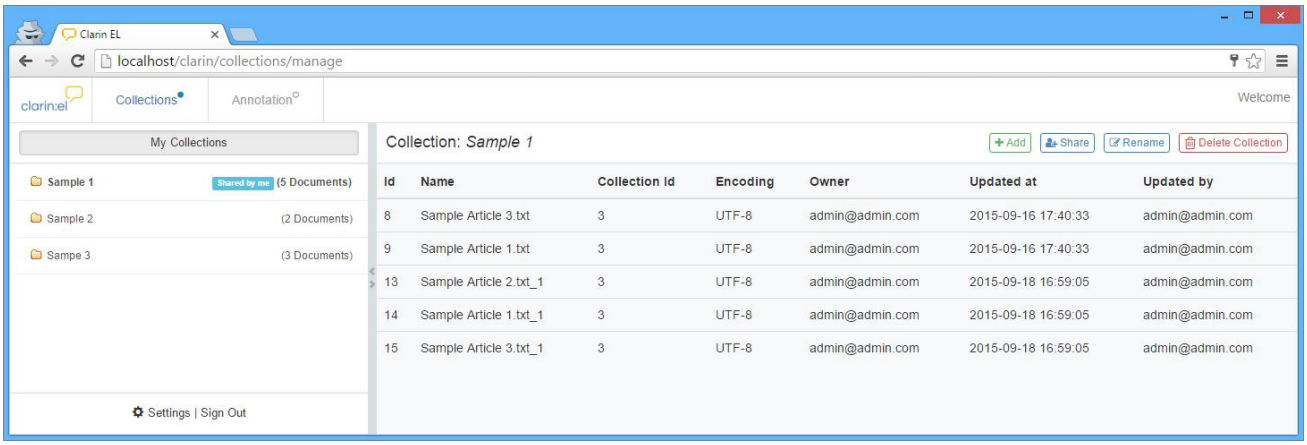

Figure 2: Collection management user interface.

3.2. CLARIN-EL Annotation Tool UI

The annotation tool is organised in a series of "pages", with each "page" concerning specific functionality. The current implementation offers two "pages", or sections: The first section concerns corpora management, which contains tools that enable users to manage their collections of documents, while the second section contains the elements that relate to the actual annotation process.

Figure 2 depicts the first section of the annotation tool. In this section, users are able to create and edit their own collections, from documents they upload through their Web browser. After a user uploads a new collection of documents, has the ability to rename it, modify the number of its documents, as well as to remove it entirely. In addition, the owner of a collection is able to share it with other users by supplying their email addresses. The invitation is being sent through email and includes a link for the invited user to confirm his or her acceptance. When a collection is shared among different users, it is eligible for collaborative annotation, meaning that multiple users can annotate the same documents simultaneously. If the owner of the collection wants to make it private again, can revoke the access of the invited users at any time, even if they annotate the selected collection at that time.

Users also have the option to export the annotations performed on a specific document in a format which is fully compatible with the Ellogon language engineering platform, along with all the formats supported by Ellogon for exporting data. The import of annotations performed from the Ellogon's annotation tool is also supported.

When the user has created or managed his/her corpora, he/she can switch to the "Annotation" page, in order to annotate documents from the available corpora. Once in the annotation page, the user is asked to select an annotation schema, from the pool of annotation schemas uploaded by the user, or public schemas uploaded by other users. Additionally, the user is asked to select the document which intends to annotate. The selection is performed by an interface similar to the one shown in Figure 3. Afterwards, the graphical user interface for performing the annotation process is being rendered, based on the guidelines embedded in the selected XML annotation schema. The set of available annotation schemas can be extended through user-provided annotated schemas in XML, conforming to Ellogon's annotation schemas guidelines[10].

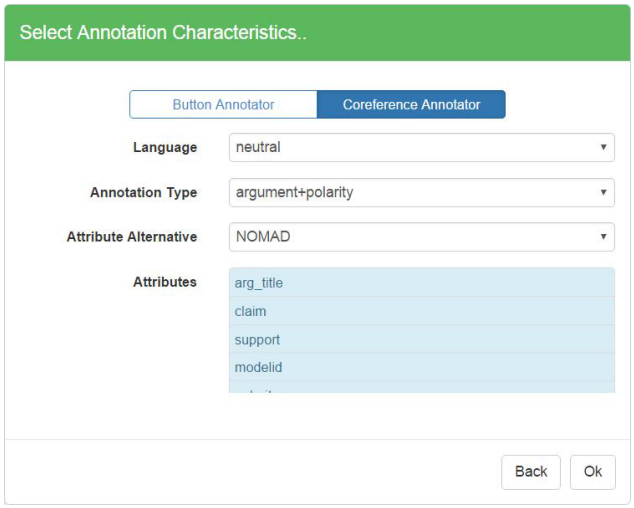

Figure 3: Annotation schema selection.

The graphical interface of the annotation page is divided into main three parts, as shown in Figure 4. The component that visualises the selected document along with the annotations that have been added to it, is placed on the left side of the screen. Users can select one of the pre-existing annotations in order to edit it, or get more information about it from the panel on the right. In addition, they are able to select a new text segment by using the available mouse shortcuts (e.g. middle mouse button click wraps whole words at once).

On the right side of the annotation tool, the user has the ability to switch between two tabs. The first contains the user interface (UI) components defined by the selected annotation schema, while the second contains information

[10]The CLARIN-EL Web-annotation tool shares the same annotation schemas as the Ellogon's standalone annotation tool - a desktop application - which can be found at: `http://www.ellogon.org/index.php/annotation-tool`.

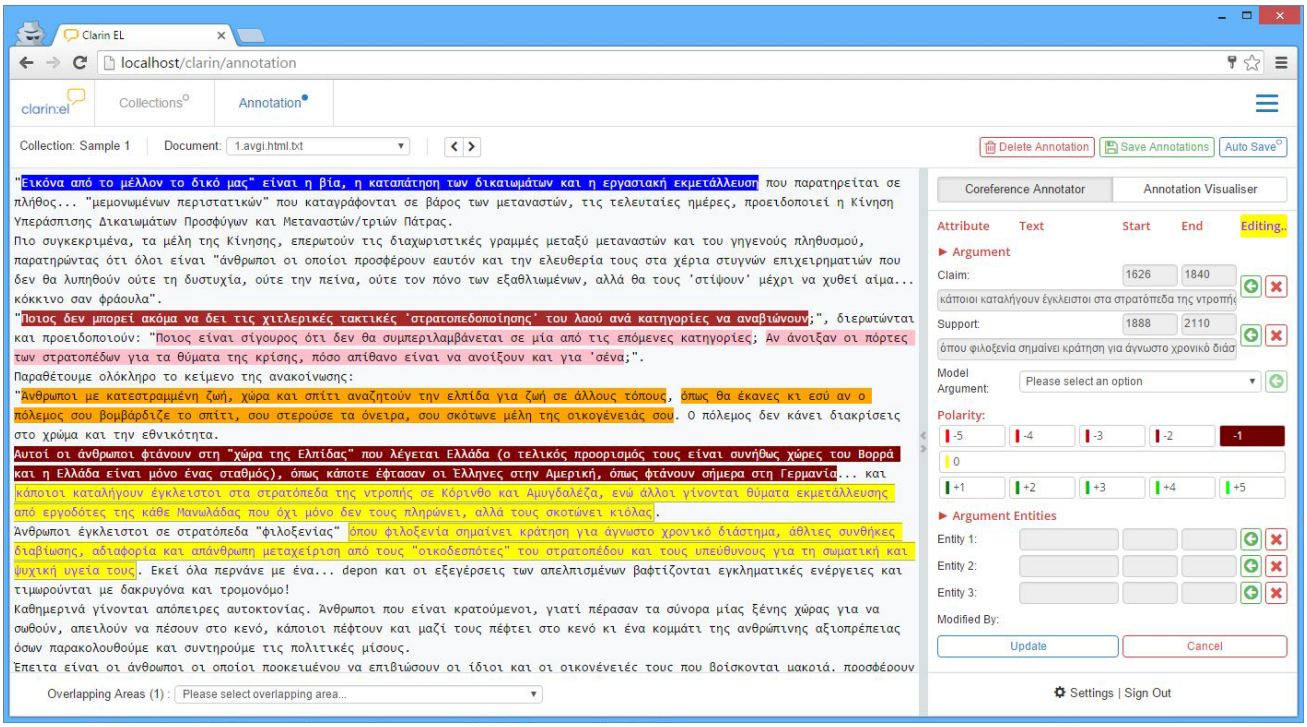

Figure 4: CLARIN-EL Annotation tool UI for annotating a document.

about the annotation that has been selected, if any. When a user has selected a text segment, the user is able to annotate it with an attribute by clicking on the desired component. In case that a user has selected an existing annotation and clicks on a different attribute, the annotation is being updated with the new value.

On the upper side of the annotation tool, there is a tool-bar which allows users to navigate between the different documents of the selected collection, save the current progress, as well as to delete an existing annotation, if any selected. In addition, the users are able to enable the auto-save feature, which monitors when the document changes and saves automatically the current progress.

3.3. Novel aspects of CLARIN-EL Annotation Tool

Being based on the annotation infrastructure provided by Ellogon and its associated annotation tools, the CLARIN-EL annotation tool tries to maintain some of requirements and novel aspects of Ellogon family of tools, such as data integrity, off-line annotation, robustness, etc. (Petasis, 2014). In order to ensure robustness and integrity of the annotation data, in cases such as internet connectivity failures or Web browser crashes, the annotation tool is able to recover the annotation status just before the connection failure or the crash: When a user begins an annotation session on a document, the current annotated statues of the document is also maintained on the server automatically, even between saves of the document (on a separate, "shadow" cache of the document). As a result, the server always maintains the last saved state of the document, but also the current state, shared among all annotators currently annotating the document. If a sudden failure occurs and the user needs to re-open the annotation tool, a dialog box with three options appears: The user is able to choose between continuing the annotation process, saving the current progress, or reverting the document to its last saved state.

The CLARIN-EL annotation tool also provides support for collaborative annotation. Users are able to share collections among each other and annotate the same document simultaneously. With the help Server-sent events (SSE), a technology standardised as part of HTML5[11] by the World Wide Web Consortium (W3C)[12], actions performed on each document are shared among all the users that annotate the specific document. If an annotator creates, modifies or deletes an annotation, the change is immediately propagated to all annotators, and it is visible in real-time to all users that annotate the same document.

Last but not least, the CLARIN-EL annotation tool is configurable through annotation schemas in XML, and users are able to upload their own annotation schemas. The annotation schemas must be in XML format and must follow the requirements of the Ellogon language engineering platform. With this feature, users can create custom annotation schemas, and thus different layouts and UIs, which meet the requirements of their annotation task. Moreover, the tool supports the export of the custom annotation schemas in XML, so as to be used as examples of how existing annotation schemas have been configured. For the various graphical elements available through the annotation schemas, please refer to section 5.1 of (Petasis, 2014).

[11] http://www.w3.org/TR/html5/
[12] http://www.w3.org/

4. Comparison with existing solutions

During the development of the proposed annotation tool, multiple existing annotation solutions have been examined. Each one of them, offers a different set of features and intends to cover a broad range of user defined requirements. Although the number of available annotation tools seems to constantly increase, an all-in-one tool has not been developed yet. In that manner, users have to choose the one that fits best their requirements and meets the needs of their specific annotation task.

Following the evolution of Web 2.0, most annotation tools are built upon a client/server architecture. Users tend to prefer plug and play solutions which require minimum effort to operate and are accessible through their favourite Web browsers. For this reason, the majority of annotation tools that have been developed during the last years are Web-based (BRAT, Inforex, WebAnno, Clarin-EL etc.). While all of these tools successfully meet the aforementioned requirement, such kind of approaches are usually limited in terms of functionality due to the limitations and inconsistencies of Web browsers. On the other hand, desktop annotation tools such as Ellogon or hybrid approaches (Desktop-Web), such as GATE Teamware, are more customisable and deliver a wider range of features to the end users.

A common feature among all the aforementioned annotation solutions, is the ability to configure the annotation schemas, according to the requirements of each annotation task. More specifically, users are able to either reuse one of the preconfigured annotation schemas that every tool comes with, or create a new schema that will satisfy their needs. Most of the tools presented in this paper support three basic annotation types. The first one refers to a label/tag (Part of Speech tag, Name Entity tag, etc.) which is being assigned to a segment of the annotated document. The second type of annotations, describes a relation between two or more annotated segments while the third one defines additional attributes, which can be added to a specific annotation. A potential difference of the reviewed approaches lies within the flexibility that is offered to the users with respect to the level of customisation of the annotation schemas.

The annotation tasks usually involve a significant number of people who are working on the same collections of documents. When a corpus is being annotated by multiple annotators, a more robust and accurate result is potentially being generated. In these terms, the existence of collaborative features inside the annotation tools is a necessity. The collaborative text annotation is quite a complex task, requiring a sophisticated approach in order to address the issues that arise from the simultaneous editing of the documents. A flexible, yet powerful implementation is needed in order to handle the multiple interactions between the different actors of an annotation task, as well as an efficient, shared storage system. Each of the annotation solutions compared in this section follow a different approach in order to fulfill this requirement.

In order to make the collaborative text annotation more efficient, GATE Teamware and WebAnno provide support for multiple user roles. GATE Teamware, distinguishes three user roles: “Administrators”, “Project Managers” and “Annotators”. Based on the role of each user, a different user interface is provided, while different access levels are assigned. WebAnno follows a quite similar approach by defining the roles of “Project Manager”, “Annotator” and “Curator”. The main difference between these annotation tools, is that WebAnno separates the role “Curator” as the user who reviews, merges the annotations of the different users and produces the final result. On the other hand, BRAT, Clarin-EL and Inforex do not provide a clear separation of user roles. The users are distinguished in two categories. The first type of users can share a collection of documents with other users and annotate documents, while the second type of users have only annotation rights. Finally, Ellogon does not provide support for different user roles.

In addition to the collaborative text annotation, some annotation tools such as BRAT and Clarin-EL offer real-time updates of the document’s annotations. This feature allows multiple users that have access on the same document to interact with each other (create, edit or delete annotations) and view the changes at the same time. In contrast to this approach, annotation solutions such as GATE Teamware and WebAnno create a separate view of the annotated document for each annotator and only the curator of the annotation task is able to review the results of the annotation process.

Some of the annotation tools (GATE Teamware, Inforex, WebAnno), which are compared in this section, provide monitoring tools that enable project managers to track the progress of the annotation tasks and get general information and statistics about specific annotation tasks. Inforex, allows project managers to define a custom set of predefined flags, which can be used to describe the work state of each document contained in a specific corpus. These flags are declared during the initial configuration of the annotation tasks. In addition to this customisable progress tracking system, it offers a view with statistics for each corpus. Although GATE Teamware and WebAnno does not allow this level of customisation as regards the progress tracking system, they provide a more advanced monitoring interface in comparison to Inforex. In more detail, WebAnno provides a dedicated monitoring interface which depicts the progress of the annotation projects, shows inter-annotator agreement statistics and allows curators to redistribute the workload in different annotators of an annotation team. Similarly, GATE Teamware provides an advanced monitoring interface that allow project managers to watch metrics similar to WebAnno, as well as statistics for each annotator (which documents have been annotated from a specific user, how much time spent on each document, etc.). Furthermore, it gives the ability to project managers to check whether a corpus is currently assigned to a project, or lock a corpus from further editing. On the other hand, Ellogon does not offer a monitoring tool for tracking the progress of a specific annotation task, but it provides statistics such as inter-annotator agreement metrics, word count, annotation counts, etc. Unfortunately, BRAT and Clarin-EL have not integrated such features (progress tracking, annotation statistics etc.) yet.

Last but not least, annotations platforms such as BRAT, Ellogon, GATE Teamware, Inforex and WebAnno have inte-

	BRAT	Clarin-EL	Ellogon	GATE Teamware	Inforex	WebAnno
Open Source	Yes	Yes	Yes	Yes	Yes	Yes
Core Lanuage(s)	Python, HTML, JavaScript	PHP, HTML, JavaScript	C, C++,Tcl	Java	Python, HTML, Javascript	Java
Application Type	Web	Web	Desktop	Desktop, Web	Web	Web
Collaborative Annotation	Yes	Yes	Yes	Yes	Yes	Yes
Real-time Collaborative Annotation	Yes	Yes	No	No	No	No
Role Management	Basic	Basic	No	Advanced	Basic	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

Table 1: Feature comparison of existing annotation solutions.

grated automatic annotation features. This kind of features may reduce significantly the cost and the time required for the creation of a manually annotated corpus. Although this approach can be very helpful, sometimes it may have disadvantages, if the service in use for the automatic annotation is trained on a different domain from the one that the annotation task refers to. BRAT enables users to either import the results of external automatic annotation systems into the annotation tool, or invoke external automatic annotation services through the interface of BRAT. The automatic annotation feature which is integrated in Inforex, allow users to run external modules for named entity recognition. After the completion of the pre-annotation phase, the curator of the project has the ability to review the inserted annotations and decide which of them are correct or need editing. WebAnno uses its built in machine learning capabilities to suggest possible annotations. The annotator can either accept or reject the suggestions of the system, which at the same time learns from the annotations of the user. The flexible architecture of Ellogon and GATE Teamware allow uses to create pipelines with custom machine learning services, which can be applied on the documents of a corpus in order to partially pre-annotate them. Similar to Inforex, a human annotator has to validate the generated result in Ellogon and GATE Teamware. A summary of the comparison of the aforementioned existing solutions can be found in Table 1.

5. Use case: Annotating arguments

Argumentation is a branch of philosophy that studies the act or process of forming reasons and of drawing conclusions in the context of a discussion, dialogue, or conversation. Being an important element of human communication, its use is very frequent in texts, as a means to convey meaning to the reader. As a result, argumentation has attracted significant research focus from many disciplines, ranging from philosophy to artificial intelligence. Central to argumentation is the notion of argument, which according to (Besnard and Hunter, 2008) is "a set of assumptions (i.e. information from which conclusions can be drawn), together with a conclusion that can be obtained by one or more reasoning steps (i.e. steps of deduction)". The conclusion of the argument is often called the claim, or equivalently the consequent or the conclusion of the argument, while the assumptions are called the support, or equivalently the premises of the argument, which provide the reason (or equivalently the justification) for the claim of the argument. The process of extracting conclusions/claims along with their supporting premises, both of which compose an argument, is known as argument mining (Goudas et al., 2015; Goudas et al., 2014) and constitutes an emerging research field. More information about argument mining can be found in the series of relevant workshops, such as (Green et al., 2014; Cardie, 2015).

In this context, the CLARIN-EL annotation tool was employed in order to create a small manually annotated corpus (around 300 news items pre-selected so as to contain arguments) with the help of two annotators. A large number of news articles (more than 5000) were downloaded from the Greek newspaper "Avgi", which were manually filtered to remove the ones that do not contain arguments. From this set, 300 documents were selected for annotation, aiming to cover as many thematic domains as possible. The size of the selected articles varies between one to ten paragraphs. The corpus as a whole consists of the following annotation units: 11640 sentences (approximately 38 per document), and 309136 words (approximately 1023 words per document). Each sentence has on average 26 tokens (including punctuation). Some of the documents were examined in order to create the annotation guidelines, which were distributed to the annotators. The annotation schema described in (Petasis, 2014) was re-used, which includes two argument components, a "claim" and one or more "premises", which support or attack the claim stated by the article's

author. The following example (example 1) illustrates an argument which consists of a claim "The sea level is rising" and a premise "greenhouse gases are causing the atmosphere to warm at a rapid rate" which supports the aforementioned statement. It should be noted that argument elements can be represented by phrases, segments smaller than sentences.

Example 1. The sea level is rising because greenhouse gases are causing the atmosphere to warm at a rapid rate.

Such kind of argumentative structures can be easily identified because of the cue words and cue phrases that they include. Cue words are defined as the connective expressions that link spans of discourse, and signal semantic relations in a text (e.g. "because", "in order to", "but", etc.). Arguments can also be expressed without the existence of cue words. Example 2 shows a more complex argumentative structure.

Example 2. Good nutrition can help prevent disease and promote health. Consumption of important fruits and vegetables ensures lower level of mortality and reduces various degenerative diseases.

In the above example the author justifies the benefits of good nutrition by the fact that the consumption of important fruits and vegetables increases the expected life span and reduces the risk of chronic diseases. These types of arguments are more difficult to be identified by the annotators, as sentences may not be successive in the text of the article. This suggests that an annotation tool that displays the document as a whole must be used, instead for example, annotation tools that specialise in annotating at the sentence level.

The annotation schema, beyond the "claim" and "support" components, also supports the annotation of polarity, and up to three entities that are involved in the argument.

The corpus was uploaded to the CLARIN-EL annotation tool as two separate corpora, along with the annotation schema. Each annotator had access only to his private version of the corpus, so as each annotator to annotate the corpus in isolation. Once each annotator has finished annotating all 300 documents, the two versions of the corpus were exported as Ellogon collections, and transfered to Ellogon platform for calculating the agreement. It should be noted that the current release of the CLARIN-EL annotation tool does not support monitoring of the annotation progress, inter-annotation agreement calculation, and merging/resolution of annotated information.

6. Conclusions and Future Work

In this paper presents a new Web-based annotation tool which attempts to eliminate the need of installing additional software during the annotation process. The presented annotation tool leverages a number of services built upon the Ellogon language engineering platform in order to exploit its annotation engine, and provides a robust annotation solution for the users, accessible from anywhere through a Web browser. Finally, the annotation tool has been developed in the context of CLARIN-EL project, and is distributed under an open license (LGPL), while it is publicly accessible from `http://clarin.ellogon.org/`, with sources available on Github[13].

As future work, we aim to provide support for even more annotation schemas, supporting for example annotation of bi-lingual, aligned documents, thus offering support for more features already offered from Ellogon's annotation engine. In addition, the integration with the CLARIN-EL infrastructure and processing services must be enhanced, so as to better integrate the annotation tool with the rest of the CLARIN infrastructure. Finally, we intend to add features that will allow users to have a more personalised user experience inside the environment of the annotation tool, mainly through the addition of roles and the ability to monitor their progress withing a collaborative annotation task.

Acknowledgements

The research leading to these results has received funding from the European Regional Development Fund of the European Union and from national funds in the context of the research project "Construction of repository and interface with internet linguistic services" (MIS 441451), of the action "CLARIN ATTICA Support and development of greek partners for the participation in the Consortia of the European Research Infrastructures in the context of the ESFRI/2006" under the Regional Operational Programme "Attica" (Priority Axis 3 "Improving competitiveness, innovation and digital convergence").

References

Besnard, P. and Hunter, A. (2008). *Elements of Argumentation*. MIT Press.

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.

Claire Cardie, editor. (2015). *Proceedings of the 2nd Workshop on Argumentation Mining*. Association for Computational Linguistics, Denver, CO, June.

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. D. (2008). BOEMIE ontology-based text annotation tool. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.

Goudas, T., Louizos, C., Petasis, G., and Karkaletsis, V. (2014). Argument extraction from news, blogs, and social media. In *A. Likas and K. Blekas and D. Kalles (Eds.): SETN 2014, LNCS 8445*, pages 287–299. Springer International Publishing Switzerland.

Goudas, T., Louizos, C., Petasis, G., and Karkaletsis, V. (2015). Argument extraction from news, blogs, and the

[13]The sources of the CLARIN-EL annotation tool can be found at: `https://github.com/iit-Demokritos/clarin-el-annotation-tool`.

social web. *International Journal on Artificial Intelligence Tools*, 24(05):1540024.

Nancy Green, et al., editors. (2014). *Proceedings of the First Workshop on Argumentation Mining*. Association for Computational Linguistics, Baltimore, Maryland, June.

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 Third International Conference on Language Resources and Evaluation, LREC 2002, May 29-31, 2002, Las Palmas, Canary Islands, Spain*. 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. (2012a). The SYNC3 collaborative annotation tool. In Nicoletta Calzolari, et al., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012*, pages 363–370. European Language Resources Association (ELRA).

Petasis, G. (2012b). 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, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26-31, 2014.*, pages 1930–1937. European Language Resources Association (ELRA).

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.

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.