

Ellogon Casual Annotation Infrastructure

Georgios Petasis, Leonidas Tsekouras

Institute of Informatics and Telecommunications,
National Centre for Scientific Research (N.C.S.R.) “Demokritos”,
P.O. BOX 60228, Aghia Paraskevi, GR-153 10, Athens, Greece.
{petasis, ltsekouras}@iit.demokritos.gr

Abstract

This paper presents a new annotation paradigm, casual annotation, along with a proposed architecture and a reference implementation, the Ellogon Casual Annotation Tool, which implements this paradigm and architecture. The novel aspects of the proposed paradigm originate from the vision to tightly integrate annotation with the casual, everyday activities of users. Annotating in a less “controlled” environment, and removing the bottleneck of selecting content and importing it to annotation infrastructures, casual annotation provides the ability to vastly increase the content that can be annotated and ease the annotation process through automatic pre-training. The proposed paradigm, architecture and reference implementation has been evaluated for more than two years on an annotation task related to sentiment analysis. Evaluation results suggest that, at least for this annotation task, there is a huge improvement in productivity after casual annotation adoption, in comparison to the more traditional annotation paradigms followed in the early stages of the annotation task.

Keywords: Authoring Tools, Corpus (Creation, Annotation, etc.), LR Infrastructures and Architectures, Casual Annotation, Annotation Tools, Collaborative Annotation, Web-based annotation tools

1. Introduction

The development and maintenance of annotated corpora can be significantly facilitated through the use of annotation tools, as annotation tools can control most aspects of the annotation process, from the presentation of the relevant information to the annotators to the validation of annotated information according to a predefined schema. A plethora of annotation tools has been presented during the last two decades (Uren et al., 2006; Fragkou et al., 2008a; Katakis et al., 2016), covering a wide range of annotation tasks and offering various levels of support. Annotation solutions can be divided into manual and semi-automatic methods: manual solutions provide the required infrastructure (i.e. storage management, graphical user interface, etc.) for annotators to annotate a corpus with a completely manual approach, where all information must be manually entered by the annotators. Semi-automatic solutions on the other hand, try to pre-annotate corpora, reducing the role of annotators into validation of existing pre-annotation. However, several of the existing annotation tools are desktop applications, allowing the annotation of corpora found on a single computer. A more recent category of annotation solutions, are *distributed* or *collaborative* annotation tools, where several annotators (not necessarily co-located) can annotate the same corpus, and in some cases even the same document. However, the construction of annotation tools that operate in a distributed environment is a challenging task, while the majority of these tools are implemented as Web applications (such as Ellogon’s Collaborative Annotation Tool¹ (Katakis et al., 2016), WebAnno² (Yimam et al., 2013) or BRAT³ (Stenetorp et al., 2012a)), having to cope with the capabilities offered by browsers.

This paper describes a new annotation tool, the Ellogon⁴

(Petasis et al., 2002a) *casual* annotation tool, which proposes a new annotation paradigm (“casual” annotation) and implements an alternative architecture employing an HTTP proxy server.

2. Casual Annotation

Typically the annotation of textual corpora is a tightly controlled process. It is not uncommon for the annotation process to involve several groups of annotators, often with various roles. Quite frequently the annotation task is facilitated with either custom-made or more generic annotation tools, such as the Ellogon Annotation Infrastructure (Katakis et al., 2016; Petasis et al., 2002a), UIMA⁵, BRAT (Stenetorp et al., 2012a)), WebAnno (Yimam et al., 2013), etc. All these tools typically involve three main steps: 1) importing of documents (usually converted to plain text) into the annotation tool infrastructure; 2) annotating the imported textual documents (either manually, semi-automatic or automatic) following the annotation guidelines and procedures; and 3) exporting the annotated texts from the annotation infrastructure according to a predefined format. While this process has been well studied and has been used to produce a vast number of high-quality annotated corpora, there are some disadvantages, such as: 1) formatting is usually lost, as well as contextual features like images or dialog utterances (i.e. in comments or tweets), leading to interpretation of textual fragments in isolation; 2) data must be imported into annotation infrastructure restricting annotation to pre-selected resources, a task that must be performed without support from the annotation infrastructure; and 3) not all evaluation scenarios can be supported, especially those requiring application to larger corpora than the ones used for training.

In addition, there is a number of tasks that is not well supported by the typical annotation procedure, especially in

¹<https://www.ellogon.org/clarin/welcome>

²<https://webanno.github.io/webanno/>

³<https://brat.nlplab.org/index.html>

⁴<https://www.ellogon.org>

⁵<https://uima.apache.org/>

the context of curating resources, like morphological, sentiment or other types of lexicons, or during the task of selecting suitable textual resources that should be manually annotated. When seeking coverage of a resource or when selecting candidates for further annotation, a closer integration between browsing for resources and using an annotation infrastructure is required. This need has driven the integration of annotation infrastructures into Web browsers, through initiatives like “AnnotateIt”⁶ and “Hypothesis”⁷, where Web browser plugins or JavaScript “scriptlets” are used to “add” annotation support to Web pages, after they have been rendered within a browser.

Although these browser-based annotation infrastructures enable the annotation of arbitrary Web resources, a) they are browser and browser version dependent as they rely on plugins; and b) they offer limited to no support in pre-processing the displayed content. The annotation infrastructure presented in this paper goes one step further: instead of relying on browser-specific properties, the infrastructure is placed before the browser, between the browser and the Web, acting as a Web proxy. This placement has some advantages, as a) the annotation infrastructure has increased content modification power over content, being able to even block parts of the content and modify HTTP request headers; b) any form of pre-processing can be applied to the content; and c) there is no need for plug-ins or actions applied by the user through “scriptlets”, enhancing compatibility with any modern Web browser.

The combination of server-side pre-processing with the absence of needed user actions for activation of the infrastructure, enables the annotation infrastructure to be continuously available, even outside the context of a “controlled” annotation process. This continuous availability is the main motivation behind casual annotation: casual annotation can be defined as the evaluation and annotation of a manually curated resource while performing every-day tasks, like browsing, reading news or interacting with social media. For example, casual annotation can be considered the evaluation of a sentence splitter while an annotator is casually reading news, comments in news articles and tweets. In such a case the annotator can easily identify errors and annotate them as such, with the annotations stored in a centralised database, similar to “AnnotateIt” and “Hypothesis”. Similarly, in the case of annotating sentiment, the annotator can see automatically annotated news items, comments, tweets or Facebook posts by applying the sentiment analyser, and annotate errors or omissions, while doing casual tasks like reading the news, or visiting their Twitter/Facebook timelines. Casual annotation has the potential to increase annotation productivity, mainly for two reasons: a) the annotation happens while the annotator is engaged in casual, everyday type activities, which do not feel as “controlled” as typical annotation tasks but are rather more relaxed; and b) annotation is not restricted to pre-selected artifacts. Instead the annotation can be applied to a wider range and more diverse content.

3. Architecture

The architecture of the Ellogon Casual Annotation Tool is centered around the “SQUID Web Proxy Cache” proxy server. Squid⁸ is a caching proxy for the Web supporting HTTP, HTTPS, FTP, and more. As a caching server, it reduces bandwidth and improves response times by caching and reusing frequently-requested web pages. Squid provides the ability to include plugins (written in the C/C++ languages) through the “eCAP”⁹ software interface, allowing the application of arbitrary operations on a Web resource, before the Web resource is served to a Web browser that utilises the proxy. Through the eCAP plugin for the Tcl language “ecap-tcl”¹⁰, the Ellogon language infrastructure and its annotation engine can be integrated within Squid, as shown in Figure 1. The “ecap-tcl” plugin includes sample code for the implementation of an HTML processor.

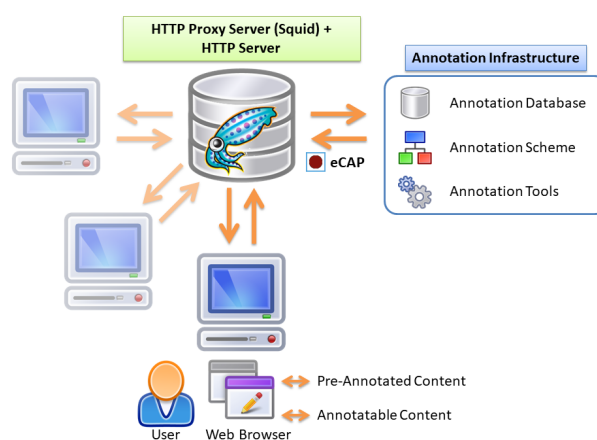


Figure 1: The architecture of Ellogon Casual Annotation Tool.

When the user requests a Web page through a browser, the browser sends the request to the Squid proxy server. The Squid server retrieves the Web page from its original location, and through eCAP asks the Annotation (Ellogon) infrastructure for a modified version of the page. The Annotation infrastructure may decide to modify the resource, according to its configuration. Since all HTTP headers are available during the modification request, the Annotation infrastructure has many parameters in order to select whether to modify a resource and what types of modifications are needed, including the URL of the resource, its type, its encoding, etc. The architecture displayed in Figure 1 is generic in the sense that is not tied to a specific Annotation infrastructure. However, the current implementation of the aforementioned infrastructure has been tested only with Ellogon. The configuration that has been extensively tested (for around four years) modifies all resources of type HTML, and some resources of type JSON, when their origin is within a set of servers (i.e. Twitter or Face-

⁶AnnotateIt: <https://annotateit.org/>
⁷Hypothesis.io: <https://web.hypothes.is/>

⁸SQUID: <http://www.squid-cache.org/>
⁹eCAP: <https://www.e-cap.org/>
¹⁰eCAP Tcl: <https://github.com/petasis/ecap-tcl>

book). Processing of JSON resources is supported, as it may be required in some cases (e.g. for Twitter), especially with Web pages formed through XHR requests. However, the interception of JSON messages typically require a site-specific implementation¹¹. Figure 2 shows an example of how a pre-annotated Web page is shown within a browser (Firefox). The annotation task shown in Figure 2 is re-

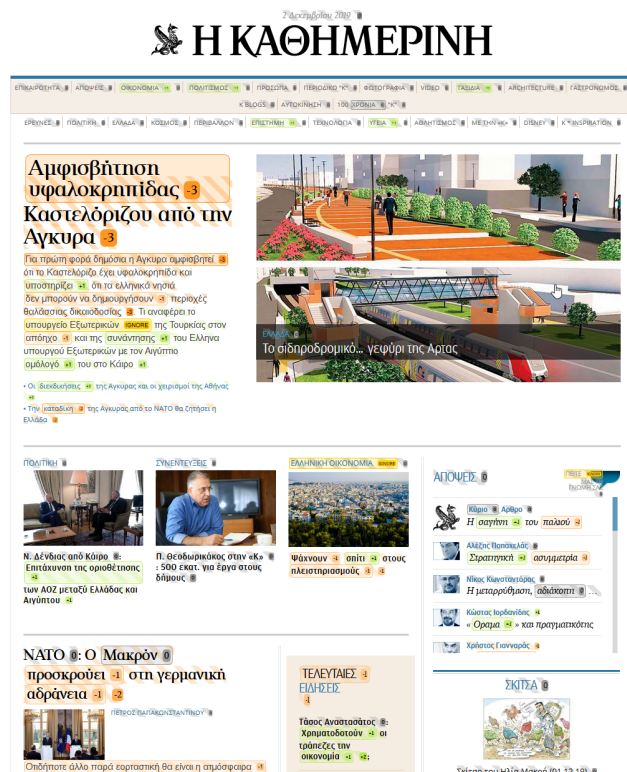


Figure 2: An example of Ellogon Casual Annotation Tool usage.

lated to sentiment analysis and the curation of a relevant lexicon for words and phrases. As shown in Figure 2, the whole Web page is annotated with existing lexicon listings (even parts loaded dynamically with JavaScript and AJAX requests), along with the identification of clauses and sentences (denoted by the slanted box areas). The annotator is able to edit existing annotations through a JavaScript User Interface based on “AnnotateIt”, activated when clicking on an annotated textual segment (as a floating window). In addition, the annotator can annotate new segments, simply by selecting them, and then clicking on a small indicator that appears on top of the selection. It is interesting to note that no browser plugins/scriptlets are required, since all the needed modifications (like loading of the JavaScript annotation infrastructure in the Web page and pre-annotating the page contents) occur within the Squid proxy server, before the content is received by the browser. As a result, the annotator is able to fully browse sites by following links: all pages will appear pre-annotated and able to be annotated. There are various ways to specify which sites will be directed through the annotator proxy server, with the most

¹¹The current implementation supports JSON from Twitter and DISQUS.

common ones being a browser extension that toggles the usage of a proxy server (like “Proxy Toggle” for Firefox), or through a “Proxy Auto-Configuration File”¹².

4. Use case: Annotating Sentiment

The paradigm of casual annotation and the Ellogon Casual Annotation tool has been applied in the context of an annotation task that has been running for several years (2010 – today). The annotation task relates to sentiment analysis, where the annotators need to annotate both words and larger fragments, such as phrases or sentences. During the first years (from 2010 to roughly 2016) the annotation task has followed the traditional paradigm, where relevant content has been collected using various criteria, including diversity of thematic domains, words and phrases not already covered, variance over document types (longer documents to micro-blogging posts), inclusion of several news sources and user generated content, etc. The collected content was subsequently imported into an annotation infrastructure, where words, segments, phrases, sentences or whole documents were annotated mainly for opinion polarity. At pre-determined intervals (i.e once a month), the annotated data were used to train an automatic annotation pipeline based on machine learning, which could be applied on new un-annotated content. Figure 3 shows the total number of annotations performed by a single annotator on various points in time. The points marked with circles (in blue) monitor how the total number of annotations increased over time. As can be seen from the figure, annotations are increased roughly linearly over time, from around 3,000 annotations in 2013 to about 14,000 annotations in 2017.

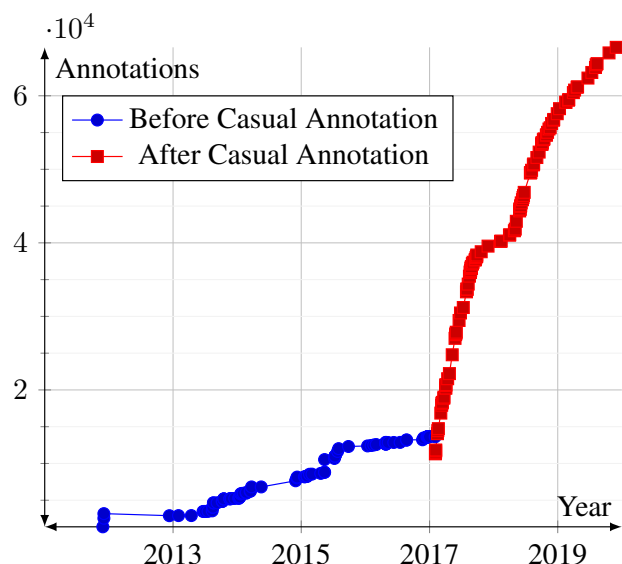


Figure 3: Number of Annotations before and after adopting Casual Annotation.

At the beginning of 2017, the annotation task switched to casual annotation, with all annotators switching to the Ellogon Casual Annotation Tool. How the total number of annotations has increased over time (for the same annotator)

¹²PAC: https://en.wikipedia.org/wiki/Proxy_auto-config

is also shown in Figure 3 (denoted by red squares). As can be seen, the slope of the increase has drastically changed, and the annotator was able to annotate much more content over time, increasing annotations from around 14,000 in 2017 to more than 65,000 at the end of 2019¹³. There are several factors that have contributed to this productivity increase, including a) the task of selecting content for annotation has been completely eliminated, saving a lot of time and reducing annotation effort; b) content is constantly pre-annotated, allowing the annotator to quickly see what is annotated, what has been erroneously annotated and must be corrected, and what needs to be annotated; c) the annotation process is significantly less “controlled”, usually to the point that the annotator does not feel like annotating, but rather performing casual activities like reading news or interacting with social media, resulting in making them spend more time on the task. At least for the case of annotation for sentiment, where errors and omissions are relatively easy to identify and annotate.

At the same time, quality of annotation has maintained the same levels, mainly due to the automatic detection of contradictions among annotators and their presentation to the annotators for resolution, before a commit can be made for each annotator. Unfortunately, the number of conflicts has not been recorded during this annotation use case, but typically lies in the range of 50–100 conflicts per year for each annotator (in the context of 4000–6000 new entries per year by each annotator). In addition, the rapid increase in available annotations and the amount of new annotations on a daily basis, allowed us to shorten the cycle of producing an updated automated annotator, from training a new annotator on a monthly basis to training a new annotator on a daily basis.

Note that in such an annotation tasks, where we aim for the maximum coverage of words and phrases over multiple document types and thematic domains, the elimination of the content selection step does not affect the quality or relevance of the annotations. In other tasks, where careful selection of documents and control over the annotation process are important, casual annotation may not be the best annotation paradigm to use, although it could still complement such a process.

5. Related Work

During the last two decades, a large number of annotation tools has been presented. Each one of them, is built upon its own logic and provides a different set of features, while some of them exploit previous experience acquired from their equivalent desktop versions. GATE Teamware¹⁴ 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 BRAT¹⁵, 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., 2012b). 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 has access to the latest version of the document. WebAnno¹⁶ 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, it 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.

Inforex¹⁷ 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 to 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 platform¹⁸ (Petasis et al., 2002b) 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 through Ellogon’s Collaborative Annotation Tool¹⁹ (Katakis et al., 2016), 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; Katakis et al., 2016). 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., 2008b), document sections (Petasis et al., 2008; Petasis and Tsoumari, 2012a), co-reference on aligned corpora (Tsoumari and Petasis, 2011), events (Petasis, 2012), arguments (Petasis, 2014; Katakis et al., 2016), and sentiment. In addition, Ellogon’s infrastructure is the only infrastructure that has implemented the casual annotation paradigm. Distributed collaborative annotation similar to “AnnotateIt” and “Hypothesis” (relying on “WebAnnotations” W3C

¹³The slight deceleration that can be observed at the beginning of 2018, was caused by a change in Twitter APIs, causing a temporary inability to annotate Twitter content.

¹⁴<https://gate.ac.uk/teamware>

¹⁵<https://brat.nlplab.org/>

¹⁶<https://webanno.github.io/webanno>

¹⁷<http://nlp.pwr.wroc.pl/inforex>

¹⁸<https://www.ellogon.org>

¹⁹<https://www.ellogon.org/clarin/welcome>

standards) have been employed in use cases such as disinformation and “fake news” (Rehm et al., 2018), annotating PDF documents (Shindo et al., 2018), annotating complex linguistic phenomena through annotation graphs (Forbes et al., 2018) and hierarchies (Helfrich et al., 2018), or curating resources like morphological lexicons and annotating morphosyntactic annotation of inflectional languages (Alosaimy and Atwell, 2018). Most of the presented tools and approaches are largely open source, i.e. Doccano (Nakayama et al., 2018), although there are also commercial approaches like TagTog²⁰.

Comparing all aforementioned solutions to the tool presented in this paper, the Ellogon Casual Annotation Tool suggests a novel annotation paradigm (casual annotation), going beyond what is currently available in Web-based, collaborative annotation tools, alleviating problems like browser compatibility, content pre-annotation and embedding the annotation infrastructure within the casual browsing activities of the users.

6. Conclusions and Future Work

This paper presents a new annotation paradigm, casual annotation, along with a proposed architecture based on an HTTP proxy, and an annotation tool, the Ellogon Casual Annotation Tool, which implements this paradigm and architecture. The novel aspects of the proposed paradigm originate from the vision to tightly integrate annotation with the casual, everyday activities of the users. Annotating in a less “controlled” environment, and removing the bottleneck of selecting content and importing it to annotation infrastructures, casual annotation provides the ability to vastly increase the content that can be annotated and ease the annotation process through automatic pre-training. The proposed paradigm, architecture and reference implementation has been evaluated for more than two years on an annotation task related to sentiment analysis. Evaluation results suggest that at least for this annotation task there is a huge improvement in productivity after casual annotation adoption, in comparison to the more traditional annotation paradigms followed in the early stages of the annotation task.

As future work, we aim to evaluate the casual annotation paradigm and provide support for more use-cases, identifying more annotation tasks that can be potentially benefited by the new paradigm. In addition, the reference annotation infrastructure must be enhanced in the direction of supporting more annotation schemas, and possibly enhance integration with recent approaches, such as “Hypothesis”.

Acknowledgments

We acknowledge support of this work by the project “APOLLONIS: Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation” (MIS 5002738) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

References

- Alosaimy, A. and Atwell, E. (2018). Web-based Annotation Tool for Inflectional Language Resources. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).
- 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.
- 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.
- Forbes, A., Lee, K., Hahn-Powell, G., Valenzuela-Escarcega, M. A., and Surdeanu, M. (2018). Text Annotation Graphs: Annotating Complex Natural Language Phenomena. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).
- Fragkou, P., Petasis, G., Theodorakos, A., Karkaletsis, V., and Spyropoulos, C. (2008a). 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/>.
- Fragkou, P., Petasis, G., Theodorakos, A., Karkaletsis, V., and Spyropoulos, C. D. (2008b). 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.
- Helfrich, P., Rieb, E., Abrami, G., Lücking, A., and Mehler, A. (2018). TreeAnnotator: Versatile Visual Annotation of Hierarchical Text Relations. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).
- Katakis, I. M., Petasis, G., and Karkaletsis, V. (2016). 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).
- Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., and Liang, X. (2018). doccano: Text annotation tool for human. Software available from <https://github.com/chakki-works/doccano>.
- Petasis, G. and Tsoumari, M. (2012a). A New Annotation Tool for Aligned Bilingual Corpora. In Petr Sojka, et al.,

²⁰<https://www.tagtog.net/>

- 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. (2002a). 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., Androutsopoulos, I., and Spyropoulos, C. D. (2002b). 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. (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).
- Rehm, G., Moreno-Schneider, J., and Bourgonje, P. (2018). Automatic and Manual Web Annotations in an Infrastructure to handle Fake News and other Online Media Phenomena. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).
- Shindo, H., Munesada, Y., and Matsumoto, Y. (2018). PDFAnno: a Web-based Linguistic Annotation Tool for PDF Documents. In Nicoletta Calzolari (Conference chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. 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. (2012a). 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, EACL '12*, pages 102–107, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012b). 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 (AEP2), in 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011)*, pages 43–52, September 15.
- Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., and Ciravegna, F. (2006). Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semant.*, 4:14–28, January.
- Yimam, S. M., Gurevych, I., de Castilho, R. E., and Biemann, C. (2013). Webanno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (System Demonstrations) (ACL 2013)*, pages 1–6, Stroudsburg, PA, USA, August. Association for Computational Linguistics.