The SYNC3 Collaborative Annotation Tool

Georgios Petasis

Software and Knowledge Engineering Laboratory
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
petasis@iit.demokritos.gr

Abstract

The huge amount of the available information in the Web creates the need for effective information extraction systems that are able to produce metadata that satisfy user's information needs. The development of such systems, in the majority of cases, depends on the availability of an appropriately annotated corpus in order to learn or evaluate extraction models. The production of such corpora can be significantly facilitated by annotation tools, which provide user-friendly facilities and enable annotators to annotate documents according to a predefined annotation schema. However, the construction of annotation tools that operate in a distributed environment is a challenging task: the majority of these tools are implemented as Web applications, having to cope with the capabilities offered by browsers. This paper describes the SYNC3 collaborative annotation tool, which implements an alternative architecture: it remains a desktop application, fully exploiting the advantages of desktop applications, but provides collaborative annotation through the use of a centralised server for storing both the documents and their metadata, and instance messaging protocols for communicating events among all annotators. The annotation tool is implemented as a component of the Ellogon language engineering platform, exploiting its extensive annotation engine, its cross-platform abilities and its linguistic processing components, if such a need arises. Finally, the SYNC3 annotation tool is distributed with an open source license, as part of the Ellogon platform.

Keywords: annotation tools, collaborative annotation, adaptable annotation schemas

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 decade (Uren et al., 2006; Fragkou et al., 2008), 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 require 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, having to cope with the capabilities offered by browsers. Annotator tools that operate as Web services are easier to implement, as the corpora are kept on a single server and annotation is happening also on the server, triggered by actions that happen at the browser of each annotator. Despite the fact the almost all collaborative annotation tools follow this approach, there are a few disadvantages in comparison to desktop applications:

- The graphical user interface of Web applications is less capable than a desktop user interface. Usability features like assigning keyboard-shortcuts to buttons, or special actions to mouse buttons, are usually missing.
- It is very difficult to personalise the annotation tool for each annotator. Features like monitoring the actions of a specific annotator, and inducing a set of regular expressions to pre-annotate documents, are quite difficult to be implemented.
- The annotation tool cannot store files or run application on the computer of the annotator. This suggests that any pre-annotation can only occur at the server, and not on the clients.
- An internet connection to the server is constantly required. An annotator cannot annotate locally, uploading the annotation results at a latter time.

This paper describes the SYNC3[1] collaborative annotation tool, which implements an alternative architecture: it remains a desktop application, fully exploiting the advantages of desktop applications, but provides collaborative annotation through the use of a centralised server for storing both the documents and their metadata, while exploits instance messaging protocols for communicating events among all annotators. The annotation tool is implemented as a component of the Ellogon language engineering platform (Peta-

[1] http://www.sync3.eu

sis et al., 2002), exploiting its extensive annotation engine, its cross-platform abilities and its linguistic processing components, if such a need arises. Finally, the SYNC3 annotation tool is distributed with an open source license[2], as part of the Ellogon platform.

2. Related Work

A plethora of annotation tools has been made available to the NLP community during the last decade (Uren et al., 2006; Fragkou et al., 2008), targeting all related modalities (text, HTML, audio, video, etc.). Popular annotation tools like the ones included in GATE[3] (Cunningham et al., 2011), Ellogon (Petasis et al., 2002), the KIM Semantic Annotation Platform (Popov et al., 2004), the SHOE Knowledge Annotator (Heflin et al., 1999), Callisto[4], Wordfreak[5], MMAX2[6] (Müller and Strube, 2006), Knowtator[7] (Ogren, 2006), and AeroSWARM[8] (Corcho, 2006), allow the annotation of texts and HTML documents using either XML-based annotation schemas, or ontologies. Usually related with natural language engineering platforms/infrastructures, these tools are desktop applications that annotate corpora stored locally, on the same machine the annotation tool is used. On the other hand, there are several tools that allow the annotation of any Web page, such as A.nnotate[9], Bounce[10], Diigo[11], iComment[12], MyStickies[13], etc. Typically, these tools employ extensions that run inside a browser along with a centralised server (for storing the annotations), in order to allow the annotation of text and images in online material, such as Web pages. Usually, these annotations are free-form text fields, where users can type anything they wish. Not conforming to any annotation schema, these tools are not well suited to the same annotation tasks, as the tools aiming at linguistic annotation. However, they offer some interesting advantages, including ease of use by not requiring installation of complex applications, the accurate rendering of HTML documents, and of course the possibility of distributed/collaborative annotation. Distributed/collaborative annotation tools not only offer the ability to create annotated corpora by annotators that are not co-located, but also offer the possibility of appealing to larger crowds, like the "Phrase Detectives" system[14] (Chamberlain et al., 2008), where linguistic annotation is exposed as a online game.

As a result, unifying the two categories of annotation tools is an appealing research area. Among the first approaches that tried to support distributed/collaborative annotation is the AGTK toolkit (Maeda and Strassel, 2004), which utilises a relational database for storing and accessing corpora on a shared server, in order to offer a framework for development of collaborative annotation tools. A similar approach is GATE Teamware (Bontcheva et al., 2010): utilising also a shared server, it offers an annotation tool that can be embedded through Java inside a browser. One of the main advantages of GATE Teamware is its extensive support for "roles", by separating annotators into three groups (managers, editors, annotators), and arranging their actions into annotation workflows. The SYNC3 annotation tool shares architectural elements from both systems, as:

- Utilises a central server for storing corpora, similar to AGTK and Teamware. However, the server needs to store only a relational database, as in AGTK, avoiding the complex installation process of Teamware, or the need to use a commercial hosting platform like GateCloud[15].
- The SYNC3 annotation tool from the annotator perspective is a desktop application, distributed as a single executable, similarly to AGTK. However, the SYNC3 tool is adaptable to XML annotation schemas, similar to the Teamware tool.
- Supports management of corpora, allowing the creation/deletion of collections, and the addition/deletion of documents into them.

In addition, the SYNC3 annotation tool introduces several novel aspects. Unique features of the SYNC3 annotation tool with respect to the state of the art, include:

- Data integrity: All data held in the central server are also stored also locally, by every instance of the SYNC3 annotation tool. This ensures that data and operation can be immediately restored in case of a problem in the central server, or in case a new server acquires the role of the central repository. Multiple copies of the data ensure that the corpora will never be lost, even if something happens to the central server.
- Off-line annotation mode: An annotator can lock a document, annotate it without being connected to the central server, and upload modifications at a latter time.
- Robustness: Communication with the central server is not required during document annotation, consisting the annotation process immune to network connection temporal errors/drops.
- Personalisation of the tool to the needs of each annotator: the SYNC3 annotation tool allows the assignment of keyboard shortcuts/mouse shortcuts that are unique for each annotator, and stored locally by every tool instance.
- Data export: ability to export any corpus/collection/document in various formats, including XML, directly from the tool.

[2]The SYNC3 annotation tool is distributed under the LGPL version 3 license.

[3]http://gate.ac.uk/

[4]http://callisto.mitre.org/

[5]http://wordfreak.sourceforge.net/

[6]http://www.eml-research.de/english/research/nlp/download/mmax.php

[7]http://bionlp.sourceforge.net/Knowtator/index.shtml

[8]http://projects.semwebcentral.org/projects/aeroswarm/

[9]http://a.nnotate.com/

[10]http://www.bounceapp.com/

[11]http://www.diigo.com/

[12]http://www.icomment.com/

[13]http://www.mystickies.com/

[14]http://anawiki.essex.ac.uk/phrasedetectives/instructions.php

[15]https://gatecloud.net/

- Personalised automatic annotation support, through regular expression acquisition performed by monitoring the actions of each annotator.
- Easy to setup and administer, as the tool is distributed as a single executable file, and the central repository can be filled automatically by any instance of the tool, if it is new or empty.

In addition, the SYNC3 tool includes a complete distribution of the Ellogon language engineering platform, suggesting that all its processing components are available from inside the tool, and can be applied on a document/collection. Teamware offers similar functionality, but execution of linguistic processing components can happen only at the server, and not within the tool. Local execution of components within the SYNC3 tool allows access to the vast collection of Ellogon's components, including operators that transform imported documents, linguistic annotation viewers, annotation schema validators, inter-annotation agreement calculators, etc.

3. The SYNC3 Annotation Tool

In order to develop an annotation tool that would be easy to use and generic enough to support a wide range of annotation tasks, we identified four basic requirements for our system:

- The tool should be user-friendly, easy to be understand and operated by the annotators.
- The operation must be based on annotation schemas that define the annotation task and guide the annotators in their work. The annotation tool must adapt its user interface automatically according to the loaded annotation schema.
- The system should support collaborative/distributed annotation, where the annotation process can be shared among different annotators at different locations.
- The system should be tolerant to losses of internet connectivity, allowing the annotation to continue locally, if possible.

The architecture of the SYNC3 annotation tool is shown in figure 1. The central component is an SQL database server, where all tools are registering themselves upon start-up and termination. The database server is used to store collections of documents, either annotated or not, along with any other information required by the annotation tools. Each annotation tool communicates with the database server through SQL queries, and the supported databases are MySQL[16], PostgreSQL[17], and Microsoft SQL Server[18]. A single instance of the SYNC3 annotation tool can be run in any number of computers. Each instance registers itself with the database server, using the credentials of the user running the application (taken from the operating system), and

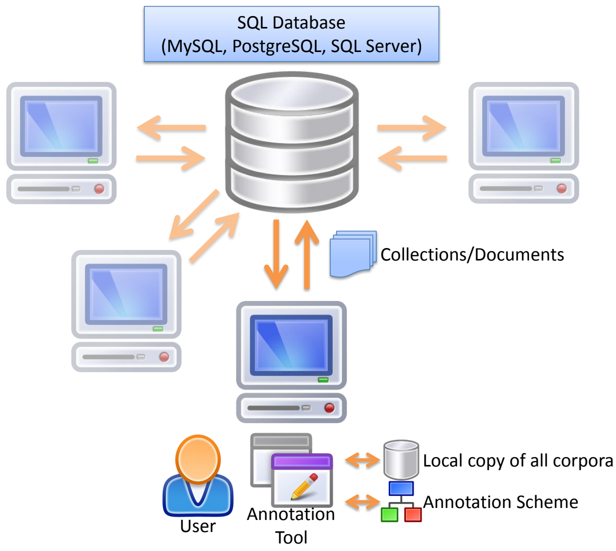

Figure 1: The architecture of the SYNC3 annotation tool.

synchronises its local copy with the database server. Each annotation tool stores locally all information kept in the server, ensuring that all data can be restored even if the database server gets replaced with a new server, and at the same time providing the ability to annotate off-line.

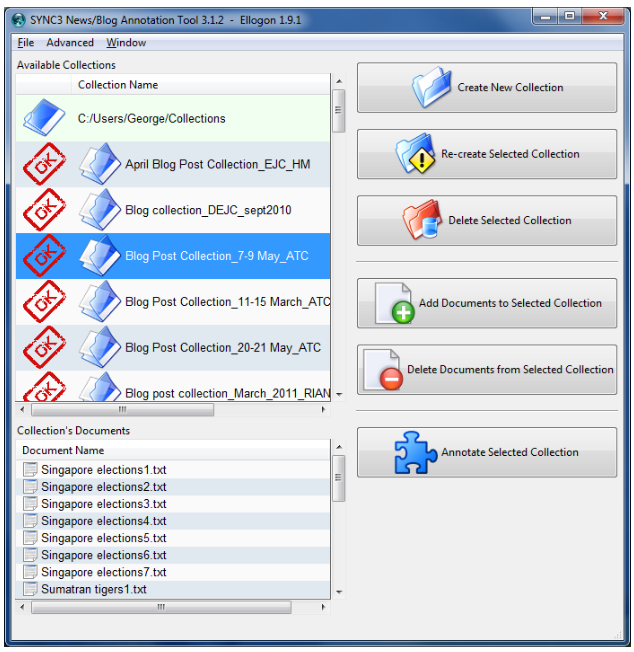

Figure 2: The main window the SYNC3 annotation tool.

The main window of the annotation tool is shown in figure 2, while the window for annotating documents (with the SYNC3 annotation schema loaded) is shown in figure 3. The main window is organised around collection and document management, where any annotator can create/delete/modify collections by adding or removing documents. The annotation window adapts itself to the selected annotation schema, allowing the user to select segments and

[16]http://www.mysql.com/
[17]http://www.postgresql.org/
[18]http://www.microsoft.com/sqlserver/en/us/default.aspx

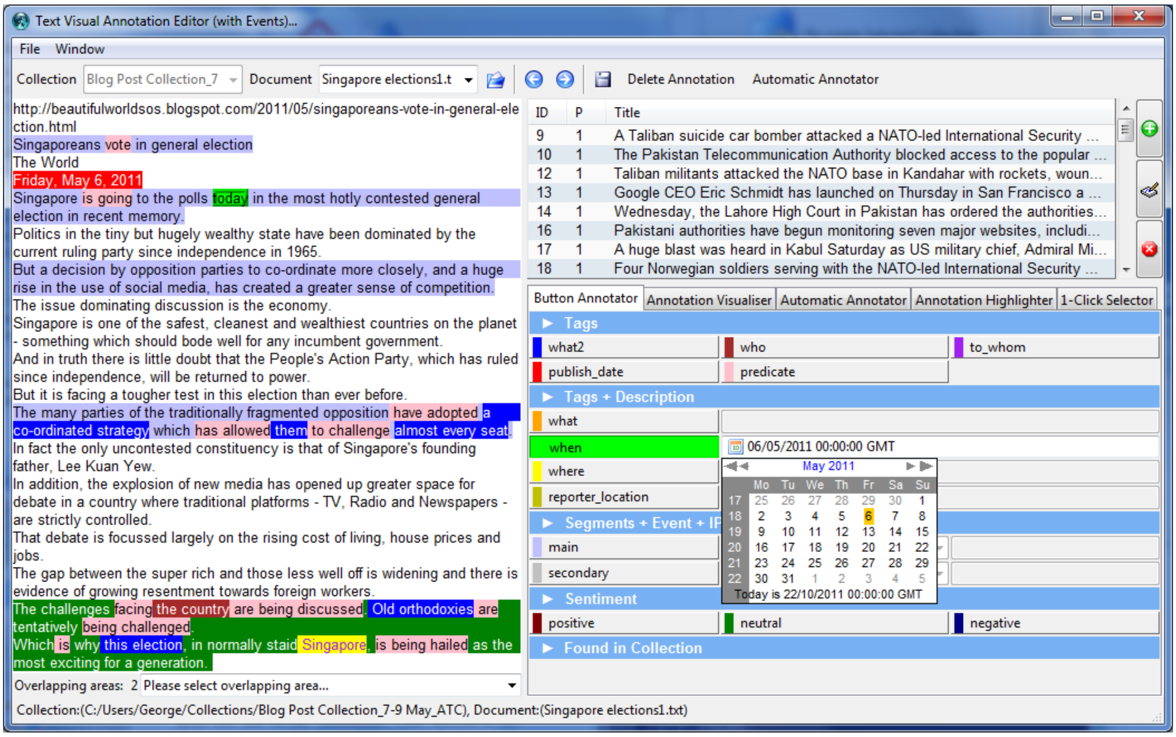

Figure 3: The annotation window of the SYNC3 annotation tool.

annotate them with a set of attributes, according to the annotation schema. The user has complete control over the way segments are selected (i.e. by configuring mouse buttons or key combinations to select whole words), and annotated (i.e. by configuring key combinations for any category). In addition, the tool monitors the annotation performed by the user, and tries to extract regular expressions from already annotated items. The user is able to revise these expressions (if desired) and apply them to automatically annotate the rest of the document, or other documents.

The SYNC3 tool supports both distributed and collaborative annotation. The distributed mode is the most frequent and easy to use mode: each annotator locks a whole document for editing, by simply opening this document in its tool, preventing any other annotator to open the same document for annotation. The collaborative mode is more complex, and requires a different configuration of the tool. Withe the help of the open instance messaging protocol Jubber[19], also used by Google Talk, actions performed by annotators are shared among all instances of the tool, effectively sharing annotations among all users that annotate the same document. However, no conflict resolution is performed: if two users annotate the same text segment, both annotations are kept into the system, no matter if they are overlapping or contradicting. Finally, the annotation tool currently runs under the Windows (XP, Vista, 7) and Linux (32 and 64 bit) operating systems.

[19] The Jubber instance messaging protocol: http://www.jabber.org/

4. Reusing Ellogon's Annotation Engine

The Ellogon language engineering platform (Petasis et al., 2002) offers an extensive annotation engine, allowing the construction of a wide range of annotation tools for both plain text and HTML documents. This annotation engine provides a wide range of features, including:

- cross-platform graphical user interface (supporting Windows, Linux and OS X),
- use of standard formats (including stand-off annotation in XML),
- support for user centered design and user friendly interface,
- support of customized annotation schemata,
- support for annotating rendered HTML pages,
- support for performing automatic annotation, and
- comparison facilities, to identify mismatches among independent annotations of the same document, or calculate inter-annotation agreement.

Despite the fact that these features are not unique among the available annotation tools (i.e. most of these features are also supported by tools offered by Callisto, Wordfreak, GATE, MMAX2, Knowtator, and AeroSWARM), reusing an annotation engine allows for rapid and robust development of a new annotation tool, through the re-use of tested components.

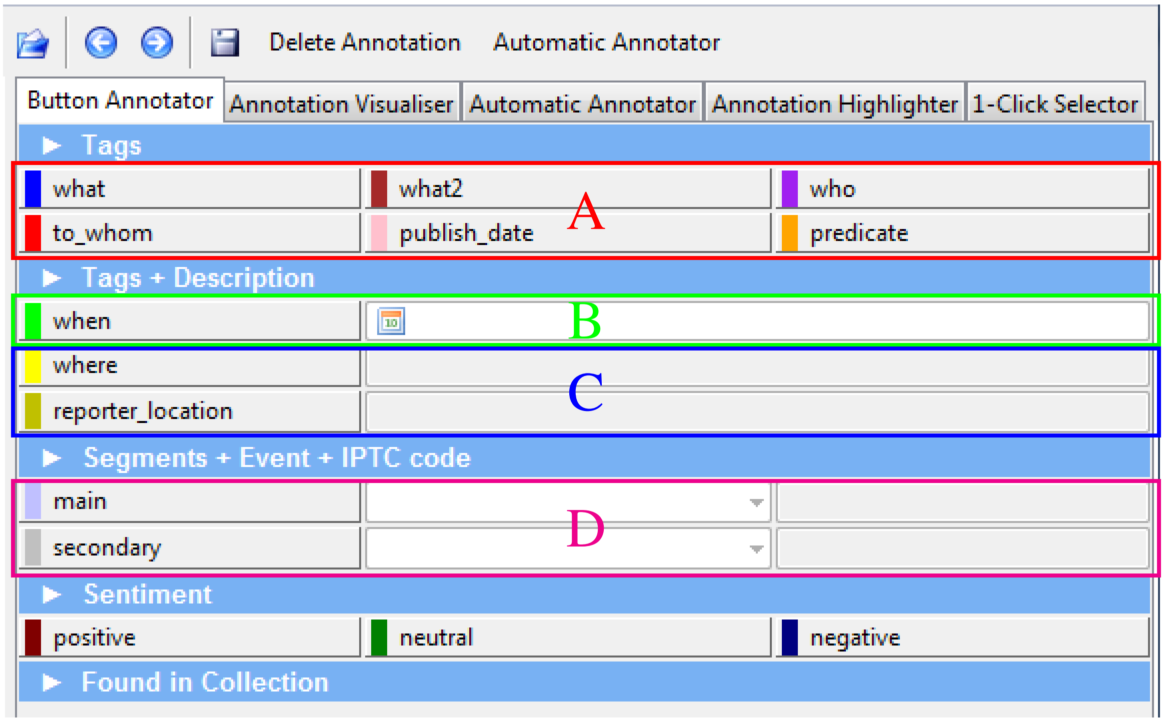

Figure 4: Various types of annotation input.

4.1. Annotation Schemes

The annotation engine of the Ellogon language engineering platform is configurable through XML files, that define annotation schemas. The tool reads the annotation schema from an XML file, and presents to the annotator a suitable GUI for annotating text segments. The SYNC3 tool follows a different approach than other tools, such as GATE Teamware: Instead of implementing floating windows which show only a small fragment of the annotation schema, the SYNC3 tool shows the whole schema, so as not to impose to the user the need to perform excessive mouse usage. In addition, the colours are not related to annotation groups (as in GATE Teamware), but on category/attribute values. The XML annotation schema language provides a variety of types that can be annotated. The most important types, along with their visual representation in the GUI, are shown in the following list:

- A **category** (figure 4-A) can be used to assign a specific category to selected segments. It is usually represented by a button widget. Typical usage of this schema type is to annotate POS tags, named-entity types, polarity, etc.

- A **date** (figure 4-B) can be used to assign a specific category to selected segments and in addition associate a date. It is usually represented by a button widget along with a date picker, to select the associated date. The date can be formatted according to the format specified in the annotation schema. Typical usage of this type is when dates in text must be associated/grounded with a normalised date, such as marking the text segment “yesterday”, and ground its date to “21 May 2012” in the linguistic annotation.

- A **category** with a **description** (figure 4-C). This input type can be used to assign a category and an arbitrary description (comment) to a selected segment. It is usually represented by a button widget and an entry widget, allowing the entry of arbitrary text. This type can be used when a category may be associated with a note/comment from the annotator.

- A **category** with a **detail** and possibly a **description** (figure 4-D). This input type can associate a category to a segment, along with a “detail”, a sub-category of predefined values, along with a description (arbitrary text). It is usually represented by a button widget and a combo-box widget, allowing the selection of a value among a set of predefined values. In case an optional description has been specified in the annotation schema, it is represented by an entry widget, allowing the entry of arbitrary text. Typical usage of this type is when a category has too many values to be represented as buttons, and the values should be selected from a drop-down list. For example, the SYNC3 annotation schema (shown in figure 3) includes a widget for defining “events” in its top-right corner: the annotator may identify the events contained in a news item, and define them, with a small description. Each defined event gets a unique id. In addition to the event definition, the annotator can mark the segment from where the event was extracted (figure 4-D), and select from the drop-down list the id of the defined event, so as to associate a segment to an event definition, which may be the same across many documents.

In addition, there are some types of annotation input that relate to grouping several segments and other information in a single annotation, to facilitate annotation of co-reference or other types of relations:

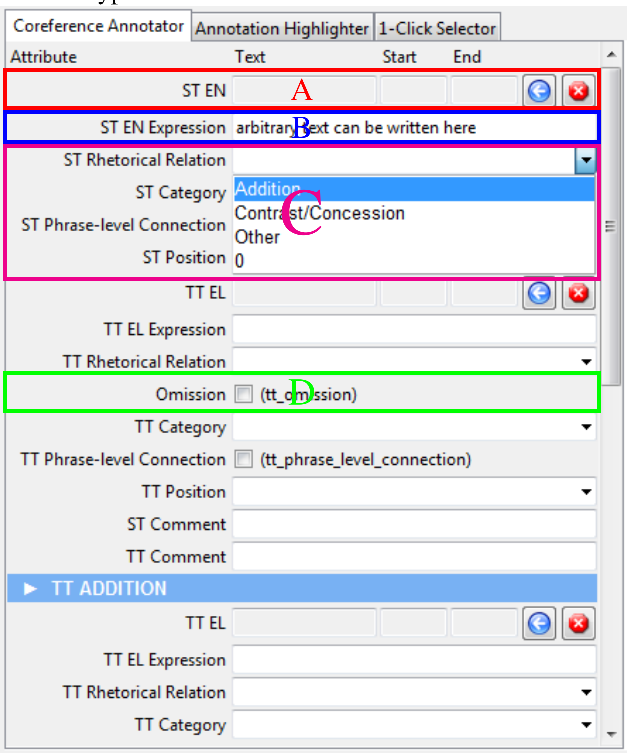

Figure 5: Types of annotation input within groups.

- A **span** or **segment** (figure 5-A). This input type is represented by a textual label (specified by the annotation schema), the text of the segment, its offsets, a button to fill in the segment from the current selection, and a button to clear the segment. It should be noted that the annotator is not required to type anything. For example, if the annotation schema defines a "source" attribute, the annotator is expected to either select an already annotated segment, or select a text segment with the mouse, and press the button with the blue arrow icon, to fill the "source" property.

- A **description** (figure 5-B), which the user can fill with arbitrary text. Represented by a textual label and an entry widget, where arbitrary text can be entered.

- A **category** (figure 5-C), selectable from a set of predefined categories by the annotation schema. Represented by a textual label and a combo-box widget, allowing the user to select a category from a set of predefined categories.

- A **boolean value** (figure 5-D), denoting the presence or absence of an attribute. Represented by a textual label and a check-box widget.

Finally, the annotation inputs can be separated in groups having a label, through the annotation schema, as shown in figure 4. Some more annotation input types related to template element filling, can be found in Fragkou et al. (2008), while an annotation schema for linguistic analysis of connectives on bi-lingual aligned corpora, can be found in Tsoumari and Petasis (2011).

4.2. The SYNC3 Annotation Scheme

The objective of the SYNC3 project is to take media monitoring and tagging to another level by comparing the latest news from traditional media sources and the blogosphere, enabling users to track their evolution, and to share favourite stories. The SYNC3 system seeks to crawl traditional news sources, and cluster news items in order to extract "events", where an "event is a particular thing that happens at a specific time and place". Having structured traditional media sources, SYNC3 tries to also structure the blogosphere, by locating blog posts, relevant to these "events". More specifically, SYNC3 automatically builds a news thematology, based on a statistical modelling approach that derives fine clusters of news articles, the so-called "news events". These events are classified into a hierarchy of news topics and themes, based on the IPTC taxonomy[20], and can be further labelled and linked with each other, according to detected temporal, geographical, and causal relations. Subsequently, the system adapts the statistical news event models to the blogosphere domain, allowing the system to automatically find blog posts that comment on these events. Finally, SYNC3 aims to determine the blog post author sentiment towards these events, through sentiment analysis.

An annotation corpus has been constructed, for evaluating various subsystems of SYNC3, using the SYNC3 annotation tool, essentially a minimal Ellogon distribution, with a SYNC3-related annotation schema, and the distributed controller, packaged as a single executable for the Windows and Linux platforms. The annotation teams were located in Maastricht, the Netherlands, Athens, Greece, and Moscow, Russia. The central server was located initially in Athens, Greece, and re-located to Maastricht, the Netherlands, by simply transferring the database. Two corpora have been created, with the first one containing the "gold" corpus that will be used for evaluation, and with the second corpus containing a small subset of the first corpus, annotated for measuring inter-annotation agreement.

The corpus was partly created by the annotators: 1000 news items from various news agencies were selected: the news items were read by the annotators and events were identified and annotated. The annotation engine of Ellogon allows not only to specify custom annotation schemas, but also to build an annotation tool by combining ready-to use components, such as the dynamic category definer (shown in top-right corner of figure 3), used in order to define the "events". Once an event has been identified, it must be located on the news item, by annotating the segments that contain the event, and identify various actors, that answer the "what", "who", "to whom", "when" and "where" questions. The annotation schema devised is shown in detail in figure 4. The upper group, labeled "Tags" and marked with "A", tries to capture exactly these categories, by offering buttons for what happened ("what" and "what2"), who did the event ("who"), and to whom the event was attributed to ("to_whom"). The "when" tag is associated with a date selector, where the annotator must select the cor-

[20]Metadata Taxonomies for the News Industry. International Press Telecommunications Council (IPTC). http://www.iptc.org

rect date, if the date cannot be automatically extracted correctly from the selected text segment. The "where" and "reporter_location" are associated with free text entries, where the annotation may type additional information, if required by the annotation guidelines. Regarding events, the annotators were requested to mark only two events in each news item, separating them into the primary event the news item is about ("main"), and the 2^{nd} most important event in the news item, marked as "secondary". Once a segment has been annotated as containing either a primary or a secondary event, the event id must be selected, from the drop-down list containing all the events defined in the system for all documents. Then the annotators were asked to search the Web, in order to locate blog posts relevant to the "main" event of the news item, import them to the corpus, and annotate the blog post with the same information. In addition, blog posts must be annotated with polarity information, by marking segments that express polarity towards the event with the "positive", "neutral" or "negative" categories.

Once the annotation of a document has been completed, the annotator marked the document as finished in the tool, while once all documents of a collection has been marked as finished, the collection was also marked as finished automatically. Once a collection has been marked as finished, the manager of the annotation process applied a set of validity tests (figure 6) from within the SYNC3 annotation tool. If the collection passed validation tests, it was marked as validated and shown with the "OK" stamp icon within the tool, as shown in figure 2.

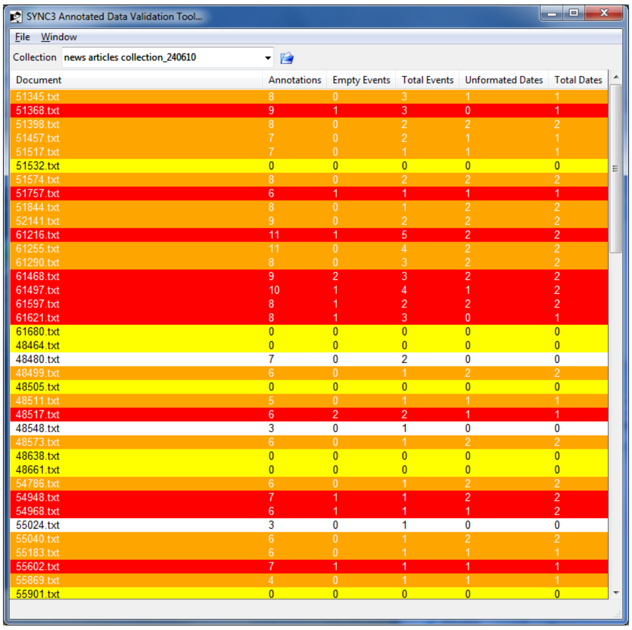

Figure 6: Applying validation checks within the SYNC3 annotation tool.

All annotation has been performed by journalists, using the SYNC3 annotation tool for all tasks, annotating a corpus of about 2500 news items and blog posts. Once the annotation has been finished, analysis has been performed outside the annotation tool, with facilities offered by the Ellogon platform, in order to calculate inter-annotation agreement, and

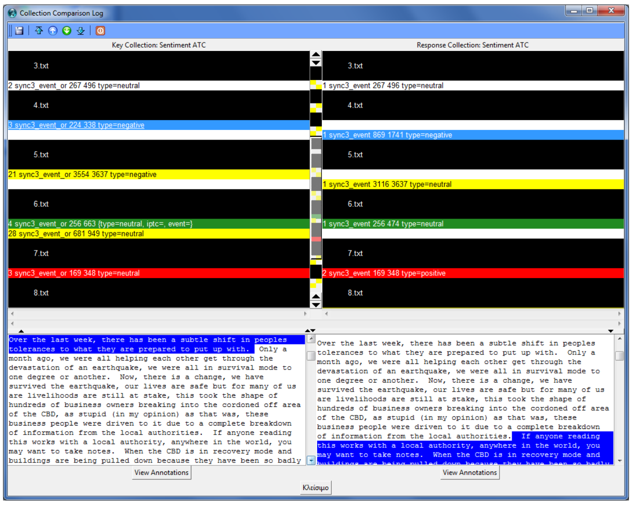

Figure 7: Comparing documents in the Ellogon language engineering platform.

examine cases of disagreement (figure 7). Measured inter-annotation agreement was good regarding the definition of events and event localisation/grounding on the documents (achieving a score of 87.87%), but was lower regarding polarity annotation, achieving an accuracy of 50% on the document level, with Cohen's kappa being equal to 0.255.

5. Conclusions and Future Work

In this paper we presented a new distributed/collaborative annotation tool, which tries to combine distributed/collaborative annotation with desktop applications, following a different approach from Web based distributed annotation tools. The presented annotation tool is implemented as an extension (plug-in) of the Ellogon language engineering platform, exploiting facilities like graphical user interface elements and its extensive annotation engine. The annotation tool has been used in the context of the SYNC3 research project, in order to annotate news items and blog posts with various types of information, including events dynamically extracted from the news, and polarity towards these events in blog posts. As future work, we aim to enhance the ability of annotating a single document by more than one annotator, especially towards conflict resolution, as the currently provided conflict resolution facilities are quite limited.

Acknowledgments

The author would like to acknowledge partial support of this work from the European Community Seventh Framework Programme, as part of the FP7 – 231854 SYNC3 project.

6. References

Kalina Bontcheva, Hamish Cunningham, Ian Roberts, and Valentin Tablan. 2010. Web-based collaborative corpus annotation: Requirements and a framework implementation. In René Witte, Hamish Cunningham, Jon Patrick, Elena Beisswanger, Ekaterina Buyko, Udo Hahn, Karin Verspoor, and Anni R. Coden, editors, *New Challenges* *for NLP Frameworks (NLPFrameworks 2010), in Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, Valletta, Malta, May 22. ELRA.

Jon Chamberlain, Massimo Poesio, and Udo Kruschwitz. 2008. Phrase detectives - a web-based collaborative annotation game. In *In Proceedings of I-Semantics*.

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

Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, Adam Funk, Angus Roberts, Danica Damljanovic, Thomas Heitz, Mark A. Greenwood, Horacio Saggion, Johann Petrak, Yaoyong Li, and Wim Peters. 2011. *Text Processing with GATE (Version 6)*.

Pavlina Fragkou, Georgios Petasis, Aris Theodorakos, Vangelis Karkaletsis, and Constantine Spyropoulos. 2008. Boemie ontology-based text annotation tool. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, 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/.

Jeff Heflin, James Hendler, and Sean Luke. 1999. Shoe: A knowledge representation language for internet applications. Technical report.

Kazuaki Maeda and Stephanie Strassel. 2004. Annotation tools for large-scale corpus development: Using AGTK at the Linguistic Data Consortium. *Proceedings of the 4^{th} International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal*.

Christoph Müller and Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2. In Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.

Philip V. Ogren. 2006. Knowtator: A protégé plug-in for annotated corpus construction. In Robert C. Moore, Jeff A. Bilmes, Jennifer Chu-Carroll, and Mark Sanderson, editors, *HLT-NAACL*. The Association for Computational Linguistics.

Georgios Petasis, Vangelis Karkaletsis, Georgios Paliouras, Ion Androutsopoulos, and Constantine D. Spyropoulos. 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.

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

Mara Tsoumari and Georgios Petasis. 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.

Victoria Uren, Philipp Cimiano, José Iria, Siegfried Handschuh, Maria Vargas-Vera, Enrico Motta, and Fabio Ciravegna. 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semant.*, 4:14–28, January.