— LaMaPUn Blue Note* —
Current Web State of Annotation Tools

Alex Dumitru, Vlad Merticariu
Computer Science, Jacobs University
http://kwarc.info

May 2, 2013

1

Abstract

Contents

*Inspired by the "blue book" in Alan Bundy's group at the University of Edinburgh, LaMaPUn blue notes, are documents used for fixing and discussing ϵ-baked ideas in projects by the LaMaPUn group (see http://trac.kwarc.info/lamapun/). Unless specified otherwise, they are for project-internal discussions only. Please only distribute outside the LaMaPUn group after consultation with the author.

1 Introduction

The purpose of this report is to present a few of the current annotation tools available on the web, in the perspective of extracting the most important features and assembling a new system, suitable for annotating mathematical text.

Text annotation is the practice of adding a note to a text, which may include highlights or underlining, comments, footnotes or links. In most of the cases, annotations can be thought of as text-metadata because they are usually added post hoc and provide information about the text without fundamentally altering it. A web annotation is an online annotation associated with a web resource. The annotation of web-based data by user communities is a widely used mean to augment and add value to the resources and there are numerous examples of different types of annotation systems across the web, some of which we are going to analyze in this report. Different types of web-based projects will require different approaches to annotations, allowing us to distinguish between 2 main categories:

1. Dynamic annotations: implemented by systems which allow the annotation of the text itself. In such a system, the anchor of an annotation is a piece of digitized text.
2. Static annotations: implemented by systems which allow the annotation of a specific region of content. In this case, the anchor of an annotation is the fixed, in-page, position of the annotated region.

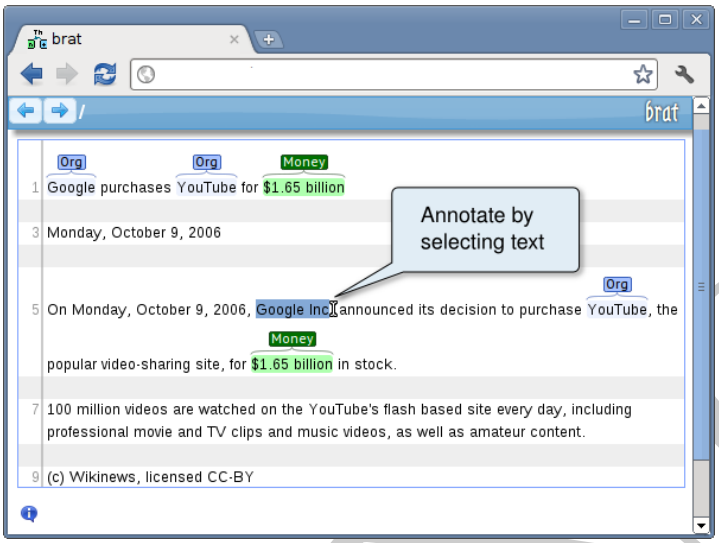
Digitized, mathematical text lays in the center of our research direction during this project and, for this reason, we are going to focus mainly on the first category. However, in order to have a complete overview of the available features, the last section of this report presents one example from the second category.

An important aspect of our desired system is that, even though it is very close to a dynamic annotation system, we are looking for a more stable and semantically rich format, such as OMDoc, to anchor the annotations. In the following sections we are presenting, one by one, the investigated systems.

2 brat Annotation Tool

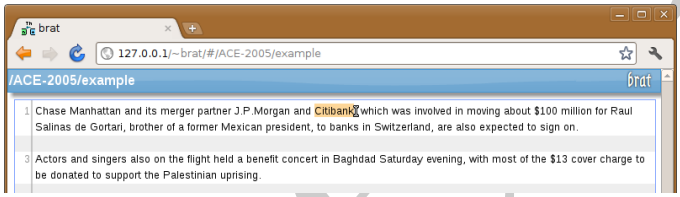
brat[1] is a web based tool for text annotation. Is is designed in particular for structured annotation, where the notes are not freeform text but have a fixed form that can be automatically processed and interpreted by a computer and implicitly belongs to the category of dynamic annotation tools.

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

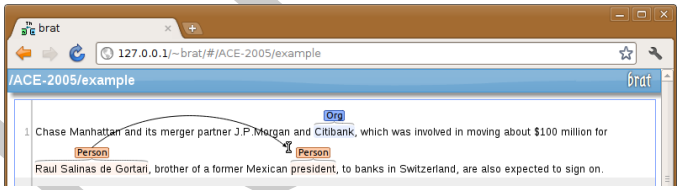


The most important features that we identified were:

1. The text must be preprocessed into a particular, fixed format before being annotated.
2. There are 2 types of annotation supported:
   - text span annotations: simple annotation of a piece of text:



   - relation annotations: two separated pieces of text are annotated and a connection between them is created:



3. Extensive search functionality for annotations:

4. Export interface: annotations can be exported in an internal format, which can then be converted in other formats such as PDF or HTML. Furthermore, visualizations can be exported as well, into SVG and bitmap formats (PNG).

5. Each annotation is accessible by URL: every brat annotation can be uniquely addressed within the brat server. Together with the URL of the server, this form of addressing provides a globally unique address for every brat annotation.

6. Brat provides a validation system for its annotations: the admin can define validation grammar rules, forcing the user to adhere to a specific annotation format.

7. Annotations are anchored based on a word-counter: this makes the annotated text unchangeable and constitutes one of our main concerns regarding the system.

3 Yawas Annotation Tool

Yawas[2] is an annotation system designed as an extension for Firefox and Google Chrome.

[2] http://www.keeness.net/yawas/index.htm

It is different than the other systems that we analyzed in that the annotation doesn't belong to the text but to the user. After installing the extension, the use can navigate to any page and can highlight any piece of text. The annotation is saved in the user's google account and is displayed every time the user accesses the page.

Feature-wise the system is much weaker than the previously analyzed brat (and much older), however it contains a potentially useful idea: user-specific annotations which can later be made accessible only to specific users or groups of users.

4 Annotatie Systeem Annotation Tool

Annotatie[3] is an annotation tool for printed documents and belongs to the static annotation category.

The anchoring of the annotations is done by page positioning: this seems inflexible at a first glance, however, it constitutes a good alternative when the text in the page is not digitized. The feature that caught our attention is the extensive comment sectioned derived from the annotations:

- Each annotation represents a new comment thread.
- In each comment thread other users can further discuss on both the contents of the annotated text and the annotation itself.
- All the annotations in a page are displayed in the right side of the page, as collapsed threads.

We believe that allowing discussion based on one user's annotation is an important feature in such a system, as the context introduced by annotations is one which emphases user collaboration.

5 Conclusion

We have presented the three most interesting annotation tools, feature-wise, available on the web, in the hope of identifying the features needed for an annotation system suitable for mathematical text documents.

We believe that brat is the most advanced existing system and a future implementation should follow the direction that brat started. However, Yawas and Annotatie can add to the system two main features: user-owned annotations and annotation-based comments.

An open question remains the in-text anchoring of the annotations. After deciding the supporting format of the text, deciding how to create the link between the annotated text and the annotation itself will constitute a crucial step in the implementation.

[3] http://www.annotatiesysteem.nl/