A survey on annotation tools for the biomedical literature

Mariana Neves and Ulf Leser

Submitted: 7th September 2012; Received (in revised form): 5th November 2012

Abstract

New approaches to biomedical text mining crucially depend on the existence of comprehensive annotated corpora. Such corpora, commonly called gold standards, are important for learning patterns or models during the training phase, for evaluating and comparing the performance of algorithms and also for better understanding the information sought for by means of examples. Gold standards depend on human understanding and manual annotation of natural language text. This process is very time-consuming and expensive because it requires high intellectual effort from domain experts. Accordingly, the lack of gold standards is considered as one of the main bottlenecks for developing novel text mining methods. This situation led the development of tools that support humans in annotating texts. Such tools should be intuitive to use, should support a range of different input formats, should include visualization of annotated texts and should generate an easy-to-parse output format. Today, a range of tools which implement some of these functionalities are available. In this survey, we present a comprehensive survey of tools for supporting annotation of biomedical texts. Altogether, we considered almost 30 tools, 13 of which were selected for an in-depth comparison. The comparison was performed using predefined criteria and was accompanied by hands-on experiences whenever possible. Our survey shows that current tools can support many of the tasks in biomedical text annotation in a satisfying manner, but also that no tool can be considered as a true comprehensive solution.

Keywords: annotation tools; curation tools; gold standard corpora; text mining

INTRODUCTION

In text mining, the development of new applications crucially depends on the existence of pre-annotated texts, called gold standard corpora (GSC). GSC are important for evaluating the performance of tools, for learning patterns or models in supervised information extraction and for an unbiased comparison of tools [1, 2]. This is particularly true in knowledge-rich domains such as the life sciences where texts deal with millions of entities whose naming often does not follow any regularities or conventions [3] and where relationships between entities may differ in highly subtle ways [4]. These properties imply that approaches purely built on background knowledge, such as dictionaries for entity names or language patterns specific for the intended relationships, often lead to insufficient recall [5].

A GSC is a set of documents where all mentions of facts of interest have been manually annotated by a human expert. GSC essentially provide examples for algorithms which learn commonalities between the annotated facts and which then try to generalize these commonalities into rules also finding novel facts [6]. This approach has proven to be highly successful in many biomedical text mining tasks [7–9], leading to an increasing interest in biomedical GSC. Examples at the entity level are corpora for detecting gene names [10], species names [11] or names of chemical entities [12]; at the relationship level, examples are corpora for studying protein–protein interactions (PPIs) [13],...
gene regulatory relationships [14], complex events [15] or drug–drug interactions [16]. A comprehensive collection of corpora for studying relationships between biomedical entities can be found at http://corpora.informatik.hu-berlin.de.

The manual annotation of text documents is a time-consuming and error-prone task, as it requires a high intellectual effort from domain experts, such as biologists or pharmacists [17]. Annotation of comprehensive corpora frequently involves a team of annotators, often with slightly different opinions about the facts under study, leading to inconsistent annotations. Comprehensive annotation guidelines are therefore important, i.e. documents which convey a precise definition of the facts to be annotated. The development of such guidelines is also a complex undertaking that often proceeds in an incremental fashion, which in turn makes necessary the re-annotation of texts [17]. This situation makes the creation of large GSC a project that often turns out to be much more expensive and much more involved than expected at the start. However, the creation of GSC is a task that can be well supported by computational tools. Supportive tasks include graphical interfaces for highlighting and tagging texts, management of text collections, assignment of texts to annotators, definition of annotation schemes (classes of entities and relationships that may be annotated) or export of the result into various formats. If available, existing text mining tools can be integrated to provide pre-tagged input texts.

This situation led to the development of a number of tools that support experts in annotating texts. These tools range from simple word processors with auto-completion functionalities (e.g. Notepad++) over standalone tools with graphical user interfaces [18–20] to web-based solutions suitable for large distributed annotation projects [21–24]. Tools vary widely in their functionalities. For instance, some tools allow arbitrary annotation schemes [20] while others support only a fixed scheme [25, 26]; some accept multiple input and/or create multiple output formats [27] while others are bound to a single format [24]; some focus on linguistic text properties [19] while others excel in domain-specific mentions [21].

Despite the importance of such tools for text mining, there exist very few previous surveys on them. The study by Uren et al. [28] focuses on the Semantic Web and does not include most of the tools covered in this work. Dipper et al. [29] provide a survey of five XML-based annotation tools focusing on linguistic annotations. To the best of our knowledge, a survey specifically focusing on the needs of the biomedical domain does not exist yet.

One important distinction to make is that between annotation and curation. By annotation, we mean the complete tagging of a given text with respect to some intended facts [17], usually for the construction of a gold standard. Sometimes, also facts that might be irrelevant for the intended scope should be annotated. For instance, if a GSC should be created for PPIs, also mentions of proteins not taking part in an interaction should be tagged, as this provides important negative information for training and evaluation. By curation, in turn, we mean the analysis of a given document with respect to some information being sought for. For instance, curated PPI databases scan documents to extract information about protein interactions [30]. Curating a document does not always imply tagging at the text level but sometimes only at the document level. Furthermore, for instance, many literal PPIs might be ignored because they do not obey quality standards of the curators (e.g. hypothetical PPI or negated PPI). Curation can be supported by text mining [31] and other tools [32, 33], but the requirements for the later are quite different.

This survey focuses on annotating texts; however, in the ‘Discussion’ section, we will also briefly discuss how well the evaluated tools are suitable for curation. Altogether, we considered almost 30 tools, 13 of which were selected for an in-depth analysis. We excluded tools which are not freely available or that focus only on linguistic annotations. Thus, we only considered tools which had been applied at least once in the biomedical domain. The 13 tools meeting these constraints were analyzed in detail with respect to 35 predefined criteria, including issues of documentation and extensibility, supported formats and platforms, implemented functionality and popularity. Furthermore, we locally installed and tested most of the 13 tools to be able to report on difficulty of installation and the general (naturally somewhat subjective) look-and-feel. We also discuss which tools are more suitable for specific situations, such as which of them support the annotation of full texts as well as relationships besides only entities.

MATERIALS AND METHODS
We screened PubMed and Google Scholar, checked corpora publications and used Google search engine to create a list of tools potentially useful for
annotating texts. All appropriate tools were evaluated due to the selection criteria presented in 'Selection of tools' section. Tools matching these requirements were assessed in details using the criteria described in 'Criteria for selected tools' section.

**Selection of tools**

There is a variety of tools available for the annotation of textual documents. In this survey, we focus on tools for manual annotation of semantic facts in biomedical documents and, optionally, for supporting database curation [34]. Therefore, a tool should allow manual annotation of texts. Manual annotation may be at any level, e.g. document, snippet, sentence or mention. Tools should support annotation of named entities and, optionally, the annotation of relationships. We put no constraints regarding supported formats. Finally, tools should be freely available, which, among others, excluded for instance [25].

We focused on general purpose tools for the manual annotation of biomedical documents, which is testified by at least one report on successful application of the tool for creating a gold standard or curating biomedical data. This, for instance, excluded tools with a sole linguistic purpose or which have been used for named-entity annotation in other domains, such as Ellogon [35], PalinkA [36], SALTO [37], Slate [38] or UAM Corpus [39]. We also excluded tools which have been used for linguistic annotation in biomedical documents (but not for biomedical facts). For instance, in [40], biomedical documents belonging to the Genia corpus [10] have been annotated on discourse relations using the PDTB tool. Regarding the tools which have been used for biomedical corpus annotation, we only excluded CADIXE (http://caderige.imag.fr/Cadixe/) used in [41, 42], which it is only available by demand, because we did not receive an answer from the tool's developers.

Tools which focus on annotations of Web documents have not been considered either, such as Domeo [43]. General purpose curation tools which do not allow the manual annotation of texts, such as Textpresso [33] and PaperBrowser [44], are also not the focus of this survey. Finally, we did not consider further all tools that were developed for a single specific question without apparent abilities to be adapted to other tasks, such as [32, 45], GOAnnotator [46] and PreBIND [47].

We tried to install all tools fulfilling these requirements. Also tools for which we did not succeed in this task were considered further, skipping the hands-on experiences. Still, downloading functionality must be working. We give our opinion regarding the easiness of installation in the detailed evaluation of each tool in Section 2 in Supplementary Data.

**Criteria for selected tools**

We identified 13 tools meeting our requirements. For those, we perform an in-depth evaluation based on published features and hands-on experiences. Tools were evaluated with respect to a list of 35 criteria grouped into the following categories (cf. Table 1): basics, publication, system properties, data and functionalities.

- Basics criteria assess the quality of the documentation and the existence of support (e.g. mailing lists and forum).
- Publication criteria encompass year of the last publication, number of citations (according to Google Scholar) and abundance of papers describing application of the tool to biomedical texts.
- System properties include availability of code, supported operating systems, general architecture (stand-alone, web-based, plug-in) and license.
- The data section addresses formats supported by the tool for input documents, input of pre-annotations, output for the annotations and annotation scheme.
- Finally, the functionalities group contains features to support the annotation process, like the smallest unit of annotation (character or token), built-in biomedical named-entity extraction, support for fast annotation (shortcut keys), pre-annotations, ontologies or inter-annotator agreement.

**RESULTS**

We screened the scientific literature for descriptions of tools that support the annotation of biomedical corpora. Tools were selected for an in-depth analysis if they are freely available, support a sufficient general class of annotation tasks and have been previously applied to biomedical texts. Thirteen tools met these criteria: @Note, Argo, Bionotate, Brat, Callisto, DjangoLogic, GATE Teamware, Knowtator, MMAX2, MyMiner, Semantator, WordFreak and XConc Suite. Wherever possible and feasible, we also installed the tools locally and experimented with their functionality and user
interface. Installation succeeded with one exception: GATE Teamware.

In this chapter, we present each of these tools in detail and highlight their strengths and pitfalls. The main technical features of each tool are summarized in Table 2 while the functional ones are shown in Table 3. Tools are presented in alphabetical order. A more detailed description of each tool can be found in Section 2 in Supplementary Data.

@Note, University of Minho (Portugal) and University of Vigo (Spain)
@Note [48] is a workbench supporting article retrieval, journal crawling, PDF-to-text conversion, pre-processing and annotation of documents. The tool has been used for annotating microbial cellular responses [49] and stringent response for Escherichia coli [50]. It is not available as open source, although plug-ins may be added to it through the AIBench Framework. It only allows the annotation of named entities and a limited set of dictionaries is included to perform manual or automatic annotation. It is pre-configured for the annotation of 14 biological entity classes and can be extended to support arbitrary entity types. Documents collections are constructed through queries to Medline and full texts can be annotated only if the corresponding PDF is found in the Web.

Argo, National Centre for Text Mining (NaCTeM), University of Manchester (UK)
Argo [24] is a web-based curation tool that has been used for annotating interactions between marine drugs and enzymes [24]. It is similar to

<table>
<thead>
<tr>
<th>Table 1: Criteria used for the evaluation of annotation tools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category</strong></td>
</tr>
<tr>
<td>Basics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Publication</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>System properties</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Data</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Functionalities</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
U-Compare [51] in terms that users can build their own annotation workflow using UIMA components. Argo is curation oriented but allows manual annotation of text, although full texts might not be visualized properly. The tool only supports annotation of named entities and the annotation schema is limited to those entities included in the UIMA-type systems. Therefore, its use might not be intuitive for those not familiar with UIMA. Analysis engines are available to pre-process documents and automatic extraction is performed using ABNER, GENIA tagger and OscarMER, for instance.

Bionotate, University of Granada (Spain) and Harvard Medical School (USA)

Bionotate [52] is a web-based tool aimed at annotating snippets in a collaborative way. Authors report on a pilot case study for the annotation of PPIs and gene disease [52] and Bionotate has also been used as part of the Genome-Environment-Trait Evidence (GET-Evidence) system [53]. It is an open source tool and features a very simple installation. The annotation scheme is composed of predefined entity types and so-called questions encoding relationships. The existence of a relationship is expressed by the answer to the question, which is limited to a single one per snippet. Therefore, Bionotate might not be suitable for the annotation of an unlimited number of entities or relationships in a text. The latest version also allows integration of named entity recognition services [54] and assignment of ontology terms to curated entities [55] (not yet available).

Brat, University of Tokyo (Japan) and University of Manchester (UK)

Brat [21] is a web-based tool for text annotation which has been used for a variety of projects, including cancer biology, epigenetics and post-translational modifications (EPI) and the BioNLP 2011 Shared Tasks [21]. It has also been employed for annotating stem cell-related documents [56] and anatomical terms [57]. Documents are displayed graphically by sentences in a very appealing manner; interactive annotation of entities and relationships is intuitive and well supported by appropriate shortcut keys. Pre-annotations can be loaded easily using a plain text stand-off format and tools accessible as web services can be integrated for automatic text pre-processing and annotation. At the downside, Brat has difficulties with the annotation of full texts. To work with the tool, we had to split all documents into parts composed at most of 50 sentences [56].
### Table 3: Comparison of all tools according to selected functional criteria

<table>
<thead>
<tr>
<th>Tool</th>
<th>Speed</th>
<th>Full</th>
<th>Relat.</th>
<th>Pre-pros.</th>
<th>NER</th>
<th>Pre-ann.</th>
<th>Ontol.</th>
<th>IAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>@Note</td>
<td>Fine</td>
<td>Poor</td>
<td>–</td>
<td>Sentences, tokens, stopwords, POS tags</td>
<td>Dictionary-based (14 types)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Argo</td>
<td>Fine</td>
<td>Poor</td>
<td>–</td>
<td>Sentences, tokens</td>
<td>ABNER, GENIA, OscarMER, species</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Bionotate</td>
<td>Good</td>
<td>Poor</td>
<td>Poor (questions)</td>
<td>Sentences</td>
<td>–</td>
<td>–</td>
<td>Good (XML)</td>
<td>Poor (no. equal answers)</td>
</tr>
<tr>
<td>Brat</td>
<td>Very good</td>
<td>Poor (connectors)</td>
<td>Very good</td>
<td>Web-services</td>
<td>Good (TXT)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Callisto</td>
<td>Fine</td>
<td>Good</td>
<td>Fine (attributes)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Djangology</td>
<td>Good</td>
<td>Good</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Good (P/R/F, side-by-side comparison)</td>
</tr>
<tr>
<td>GATE</td>
<td>Good</td>
<td>Good</td>
<td>Fine (attributes)</td>
<td>GATE components</td>
<td>GATE components</td>
<td>–</td>
<td>–</td>
<td>Good (P/R/F, kappa, comparison, consensus)</td>
</tr>
<tr>
<td>Knowtator</td>
<td>Fine</td>
<td>Good</td>
<td>Fine (slot filling)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MMAX2</td>
<td>Fine</td>
<td>Fine</td>
<td>Fine (connectors)</td>
<td>Tokenization</td>
<td>Dictionary-based, ABNER, Linnaeus</td>
<td>Poor (XML)</td>
<td>Protégé</td>
<td>Good metrics and consensus</td>
</tr>
<tr>
<td>MyMiner</td>
<td>Good</td>
<td>Poor</td>
<td>Poor (matrix check-box)</td>
<td>Sentences</td>
<td>BioPortal, cTAKES</td>
<td>Poor (XML)</td>
<td>OWL</td>
<td>–</td>
</tr>
<tr>
<td>Semantator</td>
<td>Good</td>
<td>Good</td>
<td>Fine (slot filling)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>WordFreak</td>
<td>Fine</td>
<td>Fine</td>
<td>–</td>
<td>Sentences, POS tags, parsing</td>
<td>BioPortal, cTAKES</td>
<td>Poor (XML)</td>
<td>Fine (XML)</td>
<td>–</td>
</tr>
<tr>
<td>XCconc</td>
<td>Fine</td>
<td>Poor</td>
<td>Poor (connectors)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Good (XML)</td>
<td>OWL</td>
</tr>
</tbody>
</table>

Callisto, MITRE Corporation (USA)

Callisto [58] is a stand-alone Java text annotation framework and customized versions of it have been used for annotating a variety of corpora, including ITI TXM [59], temporal clinical data [60] and also clinical data in [61]. The annotation scheme is configured by developing a custom plug-in based on a DTD file, therefore, users should be familiar with this format. We failed in annotating relationships properly probably due to errors in our DTD, and the system lacks examples of DTDs supporting relationships. Callisto is rather popular in the biomedical domain and supports the annotation of both named entities and relationships.

Djangology, DePaul University (USA) and Northwestern University (USA)

Djangology [22] is a web application for collaborative and distributed annotation of documents. It was originally developed for annotation of named entities in medical studies of trauma, shock and sepsis conditions. The tool is based on the Django web framework and annotations are saved in a database (PostgreSQL, MySQL, Oracle or SQLite are supported). Annotations and documents can be easily loaded into the system or through direct connection to the database. Annotation of relationships is not supported. It allows annotation by multiple annotator, which are also easily managed through the web interface. Djangology is capable of computing inter-annotator agreement and features a helpful side-by-side comparison of documents annotated by distinct annotators.

GATE Teamware, University of Sheffield (UK)

GATE Teamware is a web-based collaborative annotation and curation environment [23] based on the GATE infrastructure [18]. The system has been used for annotating documents related to fungal enzymes [62], lignocellulose research [63] and for the OT (OrganismTagger) corpus [64]. It allows GATE-based routines for pre-processing and pre-annotation. Annotation schemes are configured using ontologies (relationships are modeled as properties in the ontology). The tool supports multiple projects with multiple members by the flexible definition and assignment of roles to users. It also offers monitoring of progress and generation of statistics in real time. The complexity of the tool probably renders it unsuitable for small projects.

Knowtator, Mayo Clinic (USA)

Knowtator [20] is a general purpose annotation tool which runs in Protégé Frames (http://protege.stanford.edu/overview/protege-frames.html). It is certainly one of the most popular tools and was used by a number of biomedical annotation projects, such as the CRAFT corpus [65], annotation of electronic patient records in Swedish [66], clinical entities and relationships in the CLEF Corpus [67], semantic analysis of PubMed queries [68, 69], concept analysis [70], radiology reports [71], temporal expressions in medical narratives [72] and drug–drug interactions [73]. Annotation scheme are defined as ontologies and may include hierarchical relationships. Relationships are defined using slot filling, including co-references and relationships constrained to certain entity types. In our hands-on experiments, importing pre-annotations did not work satisfactorily, specially if some manual annotations had already been included in the document. Knowtator includes a consensus module to perform inter-annotator agreement and to create a consensus gold standard corpus.

MMAX2, EML Research (Germany)

MMAX2 [19] is a general purpose annotation tool supporting creation, browsing, visualization and querying of annotations. It has been integrated into the JANE active learning framework [74] and as such, it was used for annotating corpora related to gene regulation [14] and hematopoietic stem cell transplantation [75]. Input documents, annotation schemes and stand-off annotations are represented in XML format and it allows annotation of entities and relationships, as well as computation of inter-annotation agreement. We found usage of the tool rather unintuitive and had difficulties configuring and using MMAX2 for annotation of relationships.

MyMiner, Monash University (Australia), Spanish National Cancer Research Center (Spain), Université de la Méditerranée (France), Indian Institute of Technology (India)

MyMiner [76] is a web-based tool supporting a range of tasks in text curation, such as document labeling, annotation of entities and binary relationships and entity normalization. It participated in the BioCreative III IAT [77] and has been used for curation of PPIs [78] and for assisting manual document classification in the BioCreative III ACT [79]. Pre-annotation can be performed using ABNER and Linneaus, and the
system can be configured for manual annotations of other entity types. Relationships are annotated by means of a matrix with the distinct mentions found in text. While it is suitable for data curation, it is problematic for annotation, as sometimes two entities do interact in some part of a document but not in another. Additionally, a simple inter-annotator agreement functionality is provided for showing differences between annotations.

Semantator, Mayo Clinic (USA) and Lehigh University (USA)
Semantator [80] is a Protege OWL (http://protege.stanford.edu/overview/protege-owl.html) plug-in which allows manual and semi-automatic annotation of texts and was used for annotation of temporal data in clinical narratives [81]. It is similar to Knowtator (cf. above) in that the annotation schema is defined by a hierarchical ontology, but Semantator has the advantage of supporting OWL ontologies. Although the documentation claims annotation of co-references and of relationships, we were not able to perform this functionality during our hands-on experiments. For pre-annotation, users can choose among tools from the BioPortal and the Mayo Clinics Clinical Text Analysis and Knowledge Extraction System.

WordFreak, University of Pennsylvania (USA)
WordFreak [27] is a Java-based open source annotation tool built around a highly modular plug-in architecture which can be used to integrate tools for pre-annotation. It is one of the most popular tools and has been used in several projects, such as semantic role labeling [82], for the GREC corpus [83, 84], for SNPs in [85], for histone modifications in [86] and for semantic frames [87]. Annotation schemes have to be specified as plug-ins, which means that it might require a certain amount of Java programming. It is not clear whether annotation of complex relationships, such as biological events, can be handled by the tool. We installed and tested WordFreak successfully, but refrained from testing its abilities for adaptation.

XConc Suite, University of Tokyo (Japan) and University of Manchester (UK)
XConc Suite [10] is a general purpose annotation tool provided as an Eclipse plug-in. It has been used for annotating terms in the GENIA corpus [15], meta-knowledge also in the GENIA corpus [88], curating pathway [89] and for the HANAPIN corpus [90]. The tool allows importing ontologies in OWL format and it is equipped with an Ontology viewer. It allows the annotation of named entities as well as complex relationships. However, we found the annotation of complex events rather cumbersome as events are shown below a sentence, instead of being displayed in the text itself, which might also limit its suitability for annotating full texts.

DISCUSSION
We compared 13 tools for annotating biomedical texts with respect to a set of 35 criteria. Our evaluation showed that there is no single tool which supports all use cases with equal robustness. Instead, all tools have their pros and cons, making them more or less suitable depending on the task at hand. In the following, we highlight strengths of different tools by discussing their suitability for common use cases in biomedical annotation. An overview of the features supported by each tool is shown in Figure 1.

Annotation versus curation
Most of the tools described in this survey were developed for annotation purposes. Exceptions are @Note, Argo and Bionotate which were originally intended to support curation. Therefore, they contain functionality that typically is not in focus during annotation, such as document classification (cf. [34] for examples of biocuration workflows). Of the other tools, only MyMiner provides a component for document categorization which can be helpful for document triage [34]. Selection of texts is otherwise left to the users. To enable annotation, all tools present large portions of the texts to be curated and let the user select and then annotate tokens. An exception is Bionotate which only shows text snippets (cf. criterion C27). Therefore, it makes annotation of longer texts a bit painful since they need to be broken into smaller pieces, which in turn hinders a quick consideration of the context of a fact.

Supported annotations
Three of the 13 tools only support annotation of named entities, namely @Note, Argo and Django. This is probably a severe limitation for many projects, as annotation of relationships has become a de facto standard in biomedicine [14, 13,
91]. The way how annotation of relationships is supported varies considerably between the tools (cf. criteria C24). Bionotate poses questions to the user, while MyMiner generates a matrix of possible relationships, which implies that annotation at the mention level is not possible. Callisto, GATE Teamware, Knowtator, Semantator and WordFreak use slot filling which we found very hard to use in large-scale annotation as no visualization inside the text is provided. Only Brat, MMAX2 and XConc Suite show visual links between the entities directly on the text, although its use in MMAX2 is not straightforward and its visualization in XConc is confusing.

Pre-annotation

Clearly, manual annotation can benefit substantially from pre-annotation. If a comprehensive and high-quality pre-annotation is available, the task of the annotator changes from careful reading of texts to confirmation or rejection of system-suggested annotation. However, one has to keep in mind that such an approach is only useful if pre-annotations are of extremely high quality, which still is wishful thinking for all but a few entity types and way beyond the state of the art for relationships. @Note, Argo, MyMiner and Semantator come along with a named-entity module for the recognition of certain biological entity classes (cf. criterion C31). Also the GATE framework contains a set of NER components which can be integrated into GATE Teamware. Finally, Brat allows integration of the output of automatic text annotation and pre-processing tools through web services. Some of the other tools, i.e. Bionotate, Brat, Knowtator, WordFreak and XConc Suite, allow importing pre-annotated texts, but leave it to the users to compute such pre-annotations with tools of their choice (cf. criteria C14, C17, C32). This can be seen as an advantage—for the experienced user who wants to use her/his favorite tools—or as a disadvantage—for the less experienced user who is not capable of performing independent pre-annotation. In our tests,
we could add pre-annotations without problems for Bionotate, Brat and Djangology. But it might as well be feasible using the XML format supported by XConc Suite. In contrary, manipulating the input file failed in the case of Callisto, although others reported that it has been used for validation of pre-annotation [60]. Knowtator is provided with an import functionality but warns user of some risks when using it; actually, we experienced severe problems while trying it.

Simple versus comprehensive
Biomedical annotation may be carried out by a range of different users. Some users are proficient in computer usage and experienced in working with natural language texts. Thus, they often seek a tool with rich functionality and usually accept the necessity to invest a certain amount of work for adapting or configuring the tool to specific needs. Other users are less experienced and prefer tools which are simple to use and provide all functionality out-of-the-box. Table 3 and Figure 1 indicate that GATE, Knowtator and Semantator are the most comprehensive ones in terms of functionality. They support all standard formats (cf. criteria C14, C15, C16 and C17), are able to integrate ontologies to define annotation vocabulary (cf. criterion C25), allow annotation of entities and relationships (cf. criterion C24) and also adequately support projects with many annotators (cf. criteria C33 and C34). Besides, also having comprehensive annotation guidelines is important; however, supporting their creation is beyond the scope of typical annotation tools. Regarding simplicity, we recommend Brat, Bionotate, Djangology, Knowtator and MyMiner. These tools are simple to install (cf. criterion C12), can be configured easily (cf. criterion C1) and provide good documentation (cf. criterion C2) or self-explaining examples. Still, defining a proper annotation schema was not simple with any of the tools.

Popularity and support
One way of assessing tools is their recent popularity. This may be estimated from citation counts or from concrete reports on recent usage for annotation of biomedical corpora. MMAX2 (87), Knowtator (62), WordFreak (35) and Callisto (16) lead in terms of citation counts [counts are according to Google Scholar (as of August 2012)] (cf. criterion C5). When taking into account their usage for the biomedical domain, we found nine reports for Knowtator, seven for WordFreak, four for XConc, three for Callisto, Brat and GATE and only two for MMAX2 (cf. criterion C6). However, tools which have been published only recently, like Brat and MyMiner, certainly have a disadvantage in both these criteria. Another important factor is the quality of documentation and the liveliness of a tool, reflected in an ongoing development and active user support by mailing lists or forums (cf. criteria C2 and C3). Some of the tools, especially Callisto, Knowtator and WordFreak, seem not to be maintained anymore and their last versions date from more than a couple of years ago. Callisto’s website even became unavailable recently. Also questions posted to the forums of Knowtator, MMAX2 and WordFreak have not been answered by the developers within the last months. In contrast, Brat, GATE, Knowtator, MyMiner, Semantator and XConc provide very good documentation and particularly Brat and GATE have an active developer base.

Support for full texts
The use of full texts is a crucial issue in biomedical research, as studies have proven that the differences between results obtained only from abstracts compared to full texts are significant [92, 93]. All standalone tools, as well as Djangology and WordFreak, support visualization and annotation of full texts, although sometimes the text loses its format when imported into a XML or in plain text formats (cf. criterion C27). We found XConc Suite to be less suitable for full texts as it displays relationships below each sentence, which is confusing when working with long documents. For Bionotate or Brat, texts must be broken into smaller pieces to be handled adequately. In our experience with Brat for the CellFinder corpus [56], we had to split all documents into parts composed at most of 50 sentences. Full texts are also not correctly displayed in Argo and MyMiner, whose text area seems to be limited to a fixed size. Finally, @Note and MyMiner claim to support PDF documents, functionality which has not been assessed in this survey (cf. criterion C14).

Software architecture and availability of code
An important difference between tools is their architecture and the platforms they support. Actually, all tools we considered are generally platform independent (cf. criterion C8). However, as some run as
CONCLUSIONS

In this work, we surveyed 13 tools suitable for manual annotation of biomedical texts. These tools exhibit a large and diverse class of features and differ substantially in many aspects, such as the basic architecture (web-based or stand-alone), the types of facts that can be annotated (only fixed entities, configurable entities, relationships), their support for larger annotation projects (inter-annotator agreement), maturity of the software and support from the developers and handling of pre-annotations. We presented a brief description of each tool, evaluated them according to a comprehensive set of pre-defined criteria, reported on hands-on experiences for most of them and discussed them according to common situations in biomedical annotations.

Our study shows that there is no perfect tool which would be recommendable in any situation. Nonetheless, we believe that as long as the requirements of a project are not too special, researchers aiming at annotating a biomedical corpus should be able to find a tool which complies with their needs. However, non-standard requirements in all cases require adaptation at the code level. Overall, we hope that this survey raises the awareness of what is already out there and helps to diminish the resources going into costly developments of functionality already available for free.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bib.oxfordjournals.org/.

Key Points

- There exist a multitude of tools available for annotating biomedical corpora as an important resource in text mining.
- None of these tools satisfies all wishes and needs, but comprehensive and easy-to-use solutions exist for many use cases.
- Most of the tools are released as open source which may compensate deficiencies in functionality and facilitates integration with other systems or frameworks.
- Only few tools allow the use of ontologies, specially standard formats such as OBO and OWL.
- The number of tools offered as web application increases. These completely relieve from software installation and maintenance, but store annotations on remote servers.

Acknowledgements

We are thankful for the valuable suggestions, comments and corrections provided by Philippe Thomas and Michael Weidlich.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) [LE 1428/3-1].

References


