MAE ver. 0.9
User Manual

Amber Stubbs
astubbs@cs.brandeis.edu

September 29, 2011

Contents

1   Introduction

MAE (Multi-purpose Annotation Environment) was created for annotation tasks requiring extents, links, and/or non-consuming tags. Input is in the form of DTD-like files that describe the task name, tags, and attributes. A sample task definition is included with the MAE distribution, and instructions for creating your own tasks are included in this file. Mae was written in Java on an Ubuntu Linux system, and has been tested on Windows XP, Windows 7, and Mac OSX. It uses the SQLiteJDBC Java driver (more information is available here: `http://www.zentus.com/sqlitejdbc/`).

2   License

Multi-purpose Annotation Environment (MAE) and Multi-document Adjudication Interface (MAI) by Amber Stubbs at Brandeis University is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License `http://creativecommons.org/licenses/by-nc-sa/3.0/`.

The funding for both these projects was provided by NIH grant NIHR21LM009633-02 with Principle Investigator James Pustejovsky.

3   Installation and Running

MAE requires Java to run. On Windows and Unix system it should be used with the most recent version of Java 6 (it must have at least update 14 in order to run properly on Windows and Unix), though it will also run on Java 5, so it can also be run on Macs. To run MAE on any operating system, open a terminal and navigate to the directory where the .jar file exists, then run the command:

java -jar Mae_v0.9.jar

On Windows, it is also possible to open the program by double-clicking on the Mae_v0.9.jar file, however doing so will not allow all error messages to be displayed, so it is recommended to use the terminal. On all systems, you should see the following window:

4 Loading Tasks and Files

4.1 Loading a Task

In order to use MAE, you must load both a task definition file (.dtd) and a file to be annotated (.txt or .xml). The task definition must be loaded first, by selecting "Load DTD" from the File menu. Once the dtd is loaded the lower section of the MAE window will be filled with tabs representing the different tag types

4.2 Loading a file

Once the task is loaded it is now possible to annotate files. To load a file for annotation, use the "Load File" option in the File menu. You may load a file containing only text, or if there is only a portion of a file that you want to be available for annotation, you can enclose that portion `<TEXT>` and `</TEXT>` tags. Please note that these tags must be on lines by themselves, like this:

```
<TEXT>
this is the text that will be annotated.

More text to be annotated.
</TEXT>
```

Please note that if you wish to use the `<TEXT>` tags to limit what text is annotated, any text in the file that does not appear between those tags will

4

not be included in the file that is output by MAE.

You may also, of course, load a file that has been previously annotated in MAE or in any other program, as long as it adheres to the following format:

```
<TaskName>
<TEXT><![CDATA[
this is the text that will be annotated.

More text to be annotated.
]]></TEXT>
<TAGS>
<tag1... />
<tag2... />
</TAGS>
</TaskName>
```

When the input file is loaded, the text for annotation will appear in the top MAE window. The file being loaded must use UTF-8 encoding.

4.3   Annotating Entities

Entity annotation is done with the mouse. Simply highlight the word to be annotated with the left mouse button. With the word highlighted, right-click with the mouse to open a menu of possible entity tags. Select the desired tag from the menu, and the information about the tag will be added to the appropriate tab at the bottom of the screen. The id, start, end, and text features are automatically generated by MAE, and are best left unmodified. The other features can be changed by the annotator as needed.

4.3.1   Attribute Information

Once a tag has been created and appears in the table at the bottom of the screen it is possible to set values for the different attributes contained in the tag. Depending on the task, the attribute values can either be entered as free text or will be selected from a list, as shown:
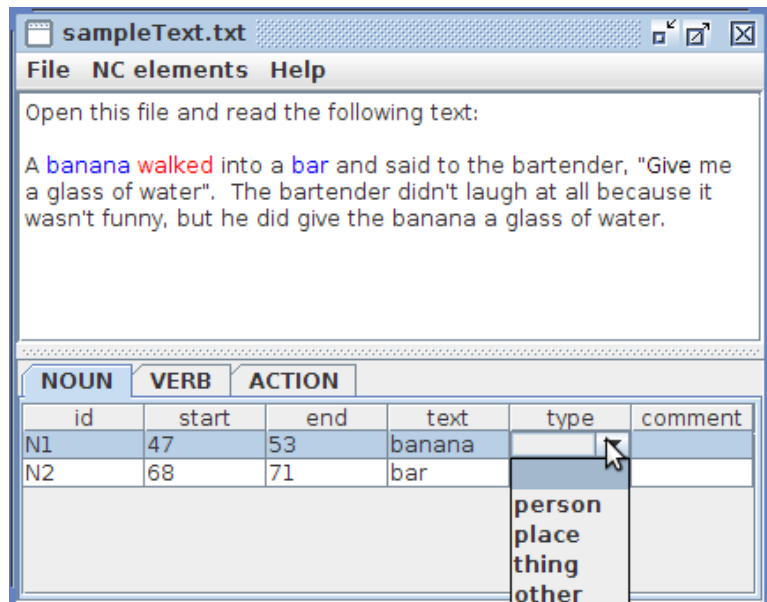
4.3.2 Non-consuming tags

It is possible in MAE to have annotation schemes which have entity tags that do not require extents to be associated with them. This can be useful for entities that are implied, but not explicitly mentioned. Non-consuming tags can be added to the annotation by selecting the tag type from the "non-consuming tags" menu at the top of the MAE window. These will be given ID tags, but the start and end elements will be set to "-1".

All non-consuming tags will be listed as options for linking to and from in the link creation window.

4.4 Annotating Links

To create links between two entities, simply make sure that the text window is active (left-clicking anywhere on the text will suffice for that), hold down the ctrl key (or the command key, if you are on a Mac) and left click each of the entities that will be included in the link, in the order that you want them to be in for the link. Once this is done, a window will pop up that will allow you to select where the link starts and where it ends, and the type of link that you want to create between them. When you click on the "Create link" button at the bottom of the pop-up window, the link will be added to

the appropriate table, where you can fill in the rest of the information about it.

4.5   Deleting tags

There are two ways to delete entity tags: highlight all or part of the tagged text in the window, and right-click on it. The pop-up menu will list all the tags at that location, with the option to remove each of them.

The other way to remove an extent tag also works for links and non-consuming tags: just select the row of the tag that you want to remove in the table, and right-click. You will have the option to remove all the highlighted rows in the active tab.

When an extent tag is removed, all the link tags that is is a part of are removed as well to maintain consistency. The reverse is not true for link tags– removing a link tag will have no effect on extent tags.

Please note that there is no undo function.

5   Saving files

To save the information you have annotated, select "Save as XML" from the File menu. This will create a file that contains the text and the tags that you have created, with the tags in stand-off format. Once you have created this file, you will be able to load it back into Mae to continue or edit existing annotations.

There is another option in the File menu called "Create RTF". This option *will not* save all the information about the tags and annotation you created, but it is a handy way to look at what each word has been annotated as.

6   Defining your own Task

Creating an annotation task for MAE is fairly straightforward. The format of the input is similar to DTDs (Document Type Definitions) used for XML. There are three main parts of task creation: the task name, the tag names, and the tag attributes. DTDs are simply text files with a .dtd extension–you can create them in any text editing program.

The specifics of how these lines should be formatted might not work well in the .pdf, so please refer to the included sample DTD (samples/sampleTask.dtd) for help if necessary.

6.1   Task Name

The task name is defined with the !ENTITY tag. If you wanted to create a task called "myTask", then create the !ENTITY line with "name" and the name of the task in quotes. The line to do so would look like this:

```
<!ENTITY name "myTask">
```

This simply provides a name to be used in the output files.

6.2   Elements (aka Tags)

Elements (defined by !ELEMENT tags) are used to define the names of the tags being used in your annotation task. MAE recognizes two types of tags: extent tags (tags used to label actual words in the document) and link tags (tags that identify a relationship between two extent tags).

To define an extent tag for your task, the line in your DTD will look like this: `<!ELEMENT TagName ( #PCDATA ) >` while a link tag will look like this: `<!ELEMENT LinkName EMPTY >`

The "(#PCDATA)" indicates that the tag will have extents, while the "EMPTY" indicates that the tag will be used for linking.

You cannot have two tags with the same name, even if they are of different types. **Note:** Currently, there are a limited number of font colors that are being assigned to tags–the colors will repeat, which may make annotation confusing if you have more than 11 tags.

6.3   Attributes

Attributes (defined by the !ATTLIST tags) contain the information associated with each tag. Some attributes are pre-defined by MAE–extent tags will always have start, end, text, and id attributes, even if they are not defined in the DTD. Link tags will always have to, from, toText, fromText, and id attributes.

Attributes must include the name of the element that they refer to, followed by the name of the attribute and the type of the attribute, like so:

```
<!ATTLIST TagName attribute1 ( YES | NO ) #IMPLIED >

<!ATTLIST TagName attribute2 CDATA #IMPLIED >
```

In the following subsections I will go over the details of the attribute descriptions.

6.3.1   id attributes

If no id attribute is created, then MAE will assume by default that the prefix for the ids for that tag will be the first letter of that tag name. So, for example, a tag called Verb will have the ids V1, V2, V3, etc. Id values are automatically assigned by the program to help prevent two tags having the same ID. If you want to specify your own prefix, add prefix="MC" to your element attribute, like so:

```
<!ATTLIST TagName id ID prefix="MC" #REQUIRED >
```

In a future version of MAE, the #REQUIRED and #IMPLIED values will have an impact on how the files are output to XML (whether an exception will be thrown if a required attribute is left blank), but at the moment they don't actually do anything, except when used for the 'start' attribute.

6.3.2   start attribute

As previously mentioned, all extent tags have an attribute called 'start', which denotes where the tag begins in the text. It is possible in MAE to create a "non-consuming" extent tag—an extent tag that doesn't actually cover any text, but instead exists as a placeholder. By default, MAE will not allow a tag to be non-consuming, but by putting a line for the start attribute with the value #IMPLIED, the tag that the start attribute belongs to will be added to the list of non-consuming tags. For example, this line:

```
<!ATTLIST Tag1 start #IMPLIED >
```

Will make it so that the tag Tag1 is allowed to be non-consuming.

If you do not want to allow a tag to be non-consuming, it is not necessary to mention the start attribute in the DTD at all.

6.3.3   Attribute types

It is possible in MAE to have a set of options for an attribute value, rather than asking the annotators to fill in their own values each time. If you want

to have a list of values, create the attribute and include a list of options in parenthesis, separated by ' | ', like so:

```
<!ATTLIST TagName attribute1 ( YES | NO ) #IMPLIED >
```

If, on the other hand, you want the annotator to be able to enter their own values in to the field for that attribute, assign the #CDATA value:

```
<!ATTLIST TagName attribute2 CDATA #IMPLIED >
```

Again, the #IMPLIED and #REQUIRED values for these attributes don't currently have an impact on how MAE operates, and it is fine to not include them.

6.3.4 Default Attribute Values

Starting in version 0.9, MAE allows you to set default values for any attribute by placing the desired value in quotes at the end of the attribute definition, like so:

```
<!ATTLIST TagName attribute1 ( YES | NO ) #IMPLIED "YES">
<!ATTLIST TagName attribute2 CDATA #IMPLIED "default">
```

Please not that if a list of options is defined in a attribute but the default value does not appear in the list, MAE will not provide that default value when creating a new tag.

7 Frequently Asked Questions

What does it mean when text is underlined?
Text in the annotation window is underlined when there is more than one extent tag at that location. I experimented with mixing the colors of the two (or more) overlapping tags, but the colors were muddy and difficult to distinguish. Underlining seemed like a more recognizable visual clue.

Why doesn't MAE do XYZ?/Hey, this doesn't work!
MAE is a work in progress. If you find any bugs or have any suggestions for improvement, please contact the creator, Amber Stubbs at astubbs@cs.brandeis.edu.

Can MAE do multi-level annotations?
MAE is much more suited for single-layer annotation tasks—if you want to create an annotation involving multiple layers, you should look at a more complex annotation tool, such as the UIMA framework.