

First Unofficial Draft of the TANGO Tool Manual

The Toolkit for graphical annotation of TimeML is used to insert TimeML links in such a way that is intuitive to the annotator. It is used in conjunction with another annotation tool such as Callisto or Alembic. Once a document has been fully tagged with `<EVENT>`, `<TIMEEX3>`, and `<SIGNAL>` tags, the TANGO tool is used to represent the anchoring and ordering relationships that TLINKs contain, the subordinating relationships that SLINKs signify, and the aspectual information that ALINKs supply. The TANGO tool offers several useful features to facilitate the annotation of TimeML links.

1 Functionality

The TANGO user interface is split into three parts. The top of the window (the text window) contains the text of the document as a reference for the annotator. In the future, this window may be replaced by Callisto, thus making it possible to add non-link tags to the document. Currently, the text window is static, although text that has previously been captured with a TimeML tag is color coded to represent this information. All of these colors can be modified by the user, but, in this document, green text represents events, blue text represents time expressions, and purple text represents signals.

The left side of the TANGO window (the pending window) contains a list of all of the tags that have not been moved to the graphical annotation palette (the final section of the window). When a document that has no previously annotated links in it is first opened in TANGO, this list will contain all of the events that exist in the document. It may also contain some signal tags. It will not contain any time expressions because they are automatically placed in the palette window.

The final portion of the TANGO interface is the graphical annotation palette (the palette window). It is here that the annotator draws relationships between time expressions, between event instances, or between a time expression and an event instance. When a document is first opened in TANGO, providing it does not already contain links, this window will be empty except for an automatically generated timeline at the top of the window. The top line in this window represents the ordering of those time expressions with comparable values. They are ordered from left to right with the leftmost expression being the earliest value. The document creation time is marked in red with two stars to distinguish it from the rest of the time expressions. Any time expressions that could not fit into this ordering are listed below the initial line. The annotator can discover the value for any time expression in the palette window by scrolling the mouse over the time expression. This results in a tool tip that provides the complete value, if possible, or any other helpful information provided in the specification of the given time expression.

The primary use of the TANGO toolkit is to facilitate the manual insertion of TimeML links. Before

a link that involves an event instance can be drawn, the instance must be created by moving the event from the pending list to the palette. This is done by selecting the event from the pending list and clicking the '=>' button. This automatically generates an instance of the event that can now participate in a TimeML link. The user can move the event instance around the palette by selecting the “Events” button and dragging the instance wherever he or she wants it. The user can also add several event instances at once to the palette by holding the shift or control key while selecting the events from the pending list. Finally, double clicking on an event instance in the palette window has the effect of generating a new instance of that event. However, the multiple instances of that event are not distinguished from each other in the palette. Scrolling over the event instance in the palette window results in a tool tip that provides the class of the event. This information is often useful for the annotator when he or she is manually generating links.

Links can be inserted by selecting the type of link from the tool bar and then simply drawing a line between the two TimeML entities that share the relationship. Once the annotator releases the mouse button, a window will pop up asking for the specification of the link. This gives the annotator the opportunity to specify the relationship type between the two entities (FIGURE). If the link requires a signal ID, the signal must first be moved from the pending list to the palette. Once this is done, the signal will appear in the Signal ID pop-up window. Once a signal has been used in a link, it is automatically removed from the palette.

Each kind of link in TimeML is color-coded. Orange lines represent TLINKs, red lines represent SLINKs, and aqua lines represent ALINKs. Again, all of these colors can be changed by the user (by selecting View Prefs). Scrolling over a link in the palette window results in a tool tip that explains what the link is. If the annotator wants to change some of the information in a given link, he or she can either delete the faulty link and insert a new one or double click on the current link to bring up the specification window for that particular link.

Most of the links in a given document are TLINKs. Many event instances are anchored to specific times using this tag while temporal orderings between event instances and time expressions are also represented with TLINKs. The SmartLink tool facilitates the creation of these kinds of links. To use this feature, the annotator selects the SmartLink button and then holds the control key while clicking on the item in the palette window that will be included in all of the TLINKs generated by SmartLink. The user can then drag another item in the palette window to the left or right of the selected item. If the second item is dropped to the left, then a BEFORE TLINK is automatically produced. If it is dropped to the right of the selected item, an AFTER TLINK is created. If the second item is dropped below the selected item, SmartLink will pop up the TLINK specification window so that the user can specify exactly what relationship he or she means by this action.

Once a document has been annotated with links in the Tango tool, the user has several choices for how to view the graphical annotation. Many of these features are also useful throughout the annotation process to remove clutter from the palette while the annotator is working. Buttons for each of these features can be found in the top, center row in the tool bar. The Show/Hide Grid button allows the user to toggle a grid on or off for the palette window. The Hide/Show buttons for each of the link varieties allows the user to clear the screen of a particular kind of link. For instance, a fully annotated document can have so many TLINKs that it is hard to see the SLINKs or ALINKs. Hiding the TLINKs allows the annotator to analyze only these links without disturbing the placement of any TLINKs. These buttons can also be used to analyze only those links that are related to a particular item or set of items in the palette window. To make use of this feature, the

user begins by hiding all three kinds of links. Making sure that the Select button in the toolbar is highlighted, the annotator can then select an item in the palette window. Only those links that contain that particular item are shown. To select a set of items when using this feature, the user holds down the control key while making his or her selections.

2 Auto-Layout Feature

In addition to the manual link insertion capabilities of the Tango tool, the software is also useful for link analysis of documents that have already been fully annotated with TimeML links using another program or by hand. When such a document is opened in Tango, all of the event instances that participate in links are listed on the left side of the palette window. The timeline at the top of this window is ordered as if the document had been opened in Tango with no links. The annotator has several options for manipulating the layout of the palette window to assist with his or her analysis of the document. By hiding certain link types, the user can more easily see which event instances participate in what links. The user can also manually move each event instance to another part of the palette window to organize the annotation. However, this can be a tedious process, especially with a large, fully annotated document. The auto-layout feature in Tango is a powerful tool for mechanically organizing the palette window for the annotator.

The 'Arrange' button calls the Auto-Layout routine. It organizes the event instances within the palette window in four ways:

1. The left to right ordering of BEFORE links and their inverses is respected.
2. IDENTITY, SIMULTANEOUS, and INCLUDES links are represented by placing event instances in the same column as the associated time expression.
3. Event instances are placed horizontally with respect to the time expressions with which they are associated.
4. Time expressions that are represented in the same column are sorted in such a way as to reduce link crossing and to make it clear when links are not present.

The resulting palette window after the Auto-layout routine has been run resembles the kind of organization many annotators wish to enforce manually.

The auto-layout algorithm goes through three stages. First, the event tags are assigned to ranks (columns) based on the BEFORE, AFTER, and similar links. This is done by a topological sort. The algorithm is as follows:

```

procedure initRanks:
  For each tag assign tag.degree to be the number of incoming
  AFTER links and outgoing BEFORE links. Initialize tag.
  rank to 0.

```

```

Insert all tags with degree 0 into queue Q.

While Q is not empty do:
  dequeue tag t
  foreach link (t BEFORE t2) or (t2 AFTER t) do:
    t2.rank := t.rank + 1
    t2.degree := t2.degree - 1
    if t2.degree = 0 then Q.enqueue(t2)
  end
end
end

```

Next, the event instances are anchored to the time expressions. For all INCLUDES, HOLDS, SIMULTANEOUS and related links to time expressions, event instances are placed directly under the time expression. For BEFORE links from an event instance to a time expression we ensure that the event is to the left of the time expression. Similarly, AFTER links ensure that the event is to the right of the time expression. Event instances linked to other event instances that have moved are then adjusted recursively in the same manner.

Lastly, within the columns of the graph, tags are sorted to improve the appearance of the graph. Currently, auto-layout attempts to place events that are more directly related to time expressions closer to the top of the panel. Other things that could be done but are not yet implemented include placing identical or simultaneous events closer together in the column or arranging the columns to reduce link crossings between multiple columns.

3 Scoring Feature

The goal of the scorer is to compare the annotation of links from the current document against the ones from a reference document. The document opened in the Tango GUI is the current document, the reference document is chosen by the user with a file chooser dialog box when he or she presses the score button. One limitation of the scorer is that the two documents must have the same annotation for the EVENT and TIMEX3 tags. This is because the link tags are the only ones added by Tango and it does not make sense to score links when the underlying EVENT and TIMEX3 tags are not the same.

When the reference document is chosen, the scorer uses it as a baseline to compute precision, recall, and the F-measure of each of the link tags. For each link of the reference document, the algorithm takes the two endpoints of the link, which are TIMEX3 IDs or IDs of an instance of an event, and tries to find a link in the current document with the same endpoints. When this is the case, the link is counted as correct and the algorithm proceeds to check if the relation type is the same for the two links. After all the links (TLINK, SLINK and ALINK) are processed, a new window is created showing the results as follows. In order to achieve accurate scoring, at least one of the documents should be closed. This is the easiest way to capture inverse link relationships such as “BEGINS” and “BEGUN_BY”.