

Requirements for the Arabic Tex Tagger

The tagger will accept as input Unicode Arabic text and a set of tags associated with the Arabic text. Each tag will have a tag type, tag start position, and tag length. We assume that we also have a static data set that associates each tag type with a description.

The tagger accepts the Arabic text and calls a tagging library that will compute and return the set of tags. We assume this library exists and we model it with a static hard coded prototype for now.

The tagger can also load the tags from a tag file that is saved from a previous session as will be discussed later.

The tagger will show the Arabic text in a dominating part of its display area. The text should be visually annotated with legends, colors, and what have you to allow the visibility of the tags.

The display should also contain a tag legend association part that allows the user to understand and control the associations between legends and tags. The control part of this display allows the user to pick and chose what tags to display.

The tagger needs to allow the user to edit and correct tags. This process should be enabled by a context menu in the text display area. The user may enter new categories and subsequently new tags if he did not find a satisfactory category.

The text display area will have the text and the text will be highlighted, colored, and featured to reflect the settings of the corresponding tag. Since the tags can be plenty and can be overlapping as well, the features associated with each tag need to be studied carefully. Most interest and focus should be given to the text area that contains the cursor at a point in time. One suggestion is to dim most the features and only stress those in the neighborhood of the cursor.

The tagger will save the tags in a tag file that it can also read. You may change the fields and the format as you desire for performance and other reasons. The following is a suggest.

```
Tags {
  /* this is the time the tag file was generated*/
  TimeStamp = "Thu April 22, 2010 – 10:30:00 AM"

  /* details about the text file*/
  Source {
    name="/home/fadi/documents/nizarkabani.txt",
    timestamp="Wed April 21, 2010 – 9:39:00 AM",
    ... /* other details about the file*/
  }

  /* the static set of tags*/
  Set {
    proper_name {id=1, description="This is a proper name", legend=2},
    city {id=2, description="This is a city",legend=4},
```

```

    ...
    female_name {id=1010, source=user, description="This is a female proper name", legend=2},
    ...
}

Legends {
    bold_red {id = 2, foreground_color=red, font=bold, ...}
    ...
}

Tag 1 { type=proper_name, start=3, length=5},
Tag 2 { type=city, start=10, length=4},
...
}

```

Note that the female_name category has a source entry that says it was a user specified category.