

ePETaLS: Online Annotation Tool for Emotional Text Labeling

Ekaterina Volkova, Arsenij Mstislavskij

May 19, 2012

Text annotation is one of the most popular methods of linguistic data collection. The quality of the resulting corpus is one of the major concerns for the researchers, especially when the annotation process is performed by participants who have not received any specific task-related training. One important factor that can help to ensure high resulting quality is a user-friendly annotation environment.

In this talk we present a new annotation system ePETaLS [1] that can help researchers to collect texts annotated for various emotions. Pre-formatted texts can be uploaded onto the system and their annotation can be assigned to a participant, whose task is to mark each phrase in the text with a specific emotion or leave it neutral. For each phrase, the annotator is also asked to assign the emotional force and mark the word on which the emotional emphasis falls. Before submission the annotation is checked and the user is informed of any missing values. This step helps to ensure higher quality of the resulting texts. The time spent on each annotation is also logged which helps to detect outliers who spend extremely little or too much time on their annotation tasks. The resulting annotation is saved in the XML format and is ready for data extraction.

Before an annotation procedure can begin, each text is automatically split into small annotation units. These units correspond to short phrases that people would usually pronounce without pausing when they read the text out loud. Each sentence in the text can contain one and more of such units, a typical unit length is three to seven word tokens. This component of ePETaLS is based on supervised machine learning system TiMBL [2] and uses WebLicht [3] for linguistic data extraction, e.g. lemmas, POS, dependency relation, etc. The machine learning algorithm uses a small corpus of texts that were split into phrases by naïve participants.

The annotation system is at present used for collecting a corpus of fairy tales in English written down by Andrew Lang [4]. Each text is annotated for ten to thirteen emotions. The final goal of the project is to create an automatic sentiment analysis system for emotional virtual character animation.

[1] e-petals.kyb.tuebingen.de/uploads

[2] <http://ilk.uvt.nl/timbl/>

[3] <http://weblicht.sfs.uni-tuebingen.de/weblicht.shtml>

[4] <http://www.gutenberg.org/browse/authors/l#a79>