# Meta-cognition evaluation
## project journal

### 1. Short project description (as given by the project description document)

Starting from a set of student verbalizations, our aim is to measure the similarity between these metacognitions and the initial read texts for better estimating the one's comprehension level. The following criteria will be taken into consideration:

- lexical similarity based on lexical chains
- semantic similarity (LSA, LDA)
- cue phrases

### 2. Team and repository

Our team is formed of only two members:

- Maruseac Mihai
- Neață Sofia

All resources used for this project including this journal will be stored in a Git repository over GitHub at https://github.com/mihaimaruseac/nlp. Although the repository is public, this cannot be a problem for this project.

### 3. Motivation for choosing this project

The main reason why we have decided to choose this project is because sometimes each of us had to do a review of a book, of an article, etc and wanted to see how much of the review is similar with the original text and how much is new text. Also, this can be used to test the coverage of the original text given by the summary.

For example, we can use this tool to test how much a blog article reviewing a book uses phrases from the original text and how much coverage of the book is given by the article. We don't want to give spoilers of the book, for example.

Lastly, this can be used to test how much a student understood from the lecture by analyzing his summary of the lecture. If the results are not satisfactory the student knows that he needs to pay more attention next time and reread / rewatch the lecture if possible.

### 4. Preview of tools used (as of 26.02.2012)

As of now, since the state of the art is not analysed, we can only give a preview of the tool which will be used in the project.

Basically, we will be using either WEKA or RapidMiner for implementing Machine Learning algorithms which will be needed in the process of developing the project. For kernel machines we can

use either libSVM or Shogun, depending on the complexity of the task where we will need them, if any.

As for the Natural Language Processing tools which we will use we analysed both OpenNLP from Apache or MontyLingua. However, we will use GATE for two reasons: it supports Romanian (in case we will port the project for this language) and it has plugins to interract with WEKA, libSVM and other tools.

## 5. to be continued