# 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. State of the art (11.03.2012)

Metacognition [6] refers to higher order thinking that involves active control over the thinking process involved in learning. It is a "thinking process" based on one's experience and knowledge about his/her own cognitive activities. This is why, the shortest definition used to "metacognition" is "thinking about thinking".

Basically, a student before actually learn details about a concept, he/her must be able to do so (that means he/she must know how to learn). This is why the teachers must not only transmit some pieces of information, but also must be aware of how to develop students' metacognitive abilities.

Metacognition can be refered to as a system [2] [4]. In the most general sense, a system is a configuration of parts connected and joined together by a web of relationships. Thereby, the most important components of metacognition are knowledge and strategy. In this case knowledge refers to the information one has and use in order to achieve his/her goal of learning another piece of information. In other words, metacognitive knowledge [3] is used in the work monitoring process in order to identify the main task that one must work on, to determine the progress his/her work etc. On the other hand, metacognitive strategies [3] are used to direct one's activities to his/her goal ("draw" the path of the learning process). So, the learning process requires resources allocation, task division and plans to achieve the resulting sub-tasks (the actual activities, but one must be aware also about the time, the intensity and the speed for each of them). Another component of the metacognition system is experience[1]. It refers to past experiences that have something to do with the current developed task. Many times, experience is not recognized as an independent component of the metacognition system, but rather a sub-component of the metacognitive knowledge.

Metacognitive knowledge can be declarative, procedural and conditional [5]. Declarative knowledge answers to the question "what" (e.g. a journal entry). Procedural knowledge answers to the question "how" (e.g. the required steps to write a journal entry); this type of knowledge underlies the metacognitive strategies. Conditional knowledge answers to the questions "when" and "why" (e.g. a journal entry must be written every time the users clicks on an image).

In order to evaluate what the student knows we must be able to compare his written text with the original one and see how many concepts are touched. First of all cue phrases have to be identified and the text should be split in the main ideas components. If the student's work is too poor such that there are no clue phrases the main ideas must be still extracted somehow.

After having the main ideas running a latent semantic analysis on each of them will transform the text into a highdimensional vector in the space of words. For better results, instead of using LSA it's best to

use a probabilistic model[7]. For a refined model we can use lexical chains of words for grouping words together before applying LSA or PLSA.

A similar treatment is done to the original text. In fact, the original text can be split into ideas from the beginning. Both alternatives have to be tested.

Then, for each idea of the original document we will compute the cosine similarity between the vector obtained from the student's work and the original document. This will provide an overview over how well this idea was understood. If this particular idea is missing in the student's work then the overview value will be 0.

The last step involves averaging all the overview values weighted by some predefined importances given to the ideas of the original text.

To evaluate the model a human expert will label several exemples and precision and recall will be computed.

## 6. to be continued

## 7. Bibliography


[1] http://studenttavern.com/2008/04/metacognitive-letters-evaluating-yourself/
[2] http://onlinelibrary.wiley.com/doi/10.1111/1467-8527.t01-1-00156/abstract
[3] http://www.hent.org/world/rss/files/metacognition.htm
[4] http://www.hent.org/world/rss/files/systems_think.htm
[5] http://www.education.com/reference/article/metacognitive-process-text-comprehension/
[6] http://catdir.loc.gov/catdir/samples/cam033/2002024499.pdf
[7] "*Probabilistic Latent Semantic Analysis*
", Thomas Hofmann
, UAI'99, Stockholm

[8] "*Measuring Metacognition and Self-Regulated*
*Learning in Educational Technologies*
", The 3rd Workshop on Metacognition and Self-Regulated
Learning in Educational Technologies
, http://www.andrew.cmu.edu/~iroll/workshops/its08/