**WIA1004 / WAES1102: FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE**

**TUTORIAL: NATURAL LANGUAGE PROCESSING (NLP)**

**APRIL 2016**

**A Corpus Development**

1. Describe the term Corpus (corpora) in a Computational Linguistic Field.
2. We will develop a Sentiment and Emotion Corpus, which can be used for other Linguistic application. Each group (with a maximum of 4 students) can choose any natural languages, including the under-resources languages and dialects. Your tasks are:
   1. Choose the language and give an overview about the language and its resources.
   2. Find at least 5 references about the language (and its digital system, if any)
   3. Using MS Excel, define a minimum of 200 words and terminologies that reflect your emotion. You may choose one of the 4 sentiment/emotion classifications (or you can also do all the four): happy, sad, hate and frighten.
   4. Label each word with a minimum of 4 tagging such as the English translation, part of speech (POS), antonym, synonym, example sentence (if any), the sms code, the iconic representation and any other relevance tagging.
   5. Create a language learning tool of your corpus: MS Power Point slides/ Website/ Wiki / look-up table/ manual/video / file etc. (refer to the example diagram).
3. Please present and submit / upload the language learning tool during the e-learning week (tutorial time).

Best Wishes !