**Indian Institute of Information Technology (IIIT) Chittoor, Sricity Questionnaire for BTP-Progress Evaluation**

**Wednesdays, 3:30 – 5pm**

**Spring 2016**

**1. Describe what you have done in the last 3 weeks.**

We have read articles related to speech acts, inter-rater agreement, a paper ( A Semi supervised dialog act tagging for telugu) and have collected data ( of about 1000 conversations) and annotated it under different speech acts. We have also implemented a java code which reads the text file and gives the word frequency and generates feature vectors of the given sentences (arff file). We are also implementing a java code to calculate the Cohen's Kappa coefficient.

**2. How many times did you meet / talk with your faculty / guide?**

We have a discussions with our faculty mentor atleast thrice a week.

**3. How many papers / articles / technical materials have you read in the last 3 weeks?**

* **Paper:** A Semi supervised dialog act tagging for telugu.
* **Articles:** Machine learning algorithms( Naive Bayes), inter-rater agreement ( Calculation of Cohen Kappa coefficient), Boot-strapping technique to train data, Implication Analysis, Speech acts (dialogue acts) and their classification.

**4. Provide a brief summary of your learning?**

We have learnt what speech acts are and how to classify the sentences based on these speech acts. Once the same set of sentences are classified by 2 different users we calculate the inter-rater agreement between them. Once the data is annotated we extract the feature vector matrix and train weka with the data set using boot strapping technique and machine learning algorithms like Naive bayes.

**5. What development / programming / practical activity did you do in the last 3 weeks.**

* We have collected a corpus of 1000 conversations from facebook, Twitter and Quora.
* Manually annotated 65 sentences according to the 43 speech acts and started implementing the program to calculate the kappa co-efficent.
* We have implemented a code in java using eclipse where a text file with sentences is given as an input and it gives the word frequency( takes the top 20%) and generates a Feature vector file (arff file) and train weka with this data set.
* We are impleming the code in java to calculate the Cohen's Kappa co-fficient (from Confusion matrix).

**6. How close/far are you from the milestone set by your Guide?**

We are half way through the code of the confusion matrix (calculation of Cohen's Kappa coefficient) and started writing the code to build the classifier.

**7. What specific challenges are you facing/you faced in the last 3 weeks?**

There was a little problem in getting familiarized with and using weka. We have learnt it now and are using it.

**8. Propose your plan for the next 3 weeks; as agreed with your supervisor. It would be verified in the next round** (Q1).

* To build a classifier which automatically predicts the class (speech act) to which a sentence belongs to and further improve the classifier to get accurate results.
* Implication Analysis – collecting the data set which would have sentences and the different implications on the sentence made by different users.