Phase 1 Report

1. Understanding of the problem statement
   The problem statement required us to generate Wikipedia articles in Indian languages for 2 domains which include schools and hospitals of India. This is done in order to achieve a larger aim of enriching the Indian Wikipedia.

2. Tasks completed
   a. Hindi, Telugu, Marathi school articles
   b. Hindi, Marathi initial hospital articles
   c. Basic data collected and cleaned for various hospitals around the country.

3. Errors/scope of improvement
   a. Not enough data to generate large articles
   b. More variation can be introduced into the articles
   c. The transliteration is not very good
   d. Only the article text is being generated without any other components of the Wikipedia article

4. Progress with respect to the proposed timeline in our scope document
   a. We proposed to completely port the code for school articles to other languages as our midpoint deliverable which has been successfully completed
   b. Our other proposal included having an end-to-end system for writing articles for the school domain in Hindi and Marathi which has been achieved as well
   c. Our initial scope document also included the collection of data for generating articles for hospitals which has been completed and we are still looking for more data to improve the quality of the generated articles.

5. The novelty of our solution
   a. We have proposed a solution that is language independent. The user has to create a template in his own preferred language, use this as an input file. The bot automatically creates articles with the data and templates.
   b. It happens so that no translator or transliterator is the best, therefore we propose a solution to use multiple translators according to their performance.
   c. Keeping the template and code separate allowed us to use our solution with more scaled data i.e we can add or remove features in our dataset at any time regardless of changing the code each time.

Future Scope:

1. More data
   One of our main objectives going into the next phase of the project would be to include more data in terms of the number of fields and the types of sentences that are generated

2. UI improvements
   We plan on having a proper command-line based UI for interacting with the script.

3. Publishing the content
   If needed we plan to publish the generated articles on Wikipedia/Wikimedia.