

Natural Language Processing (NLP) Techniques to Measure the Syllabus Coverage with the Final Exam Paper

A. G. K. S. Dharmapriya¹, J. R. K. C. Jayakody

Computer and information systems, Wayamba University of Sri Lanka, Kuliyaipitiya, Sri Lanka

Exam questions usually play a pivotal role in the education and it is the main assessment technique to evaluate the Intended learning Outcome (ILO). The main Intention of the subject syllabus is to cover the ILO. If the syllabus is not covered effectively through the exam questions, it is a challenging task to identify whether students have acquired and enhance the necessary skills and the knowledge which are specified in the given ILO. Therefore the preparation of the exam papers and its evaluation plays an important role in education to improve the performance of the students. Having appropriate level of exam questions which covers the entire syllabus is a time consuming, tedious and challenging work for the instructors. Therefore this research work was done with the view of setting up effective question paper to measure the depth of the syllabus coverage. Further, Identification of the students' knowledge level was an output. Natural Language Processing (NLP) techniques such as tokenization, stop word removal, non-alpha numeric word removals and tagging were used to process the contents and questions. NLP with NLTK and cosine similarity with TF-IDF (term frequency inverse document frequency), TF-IDF variations and semantic similarity algorithms were developed to generate a unique set of rules to identify the best syllabus coverage contents for exam questions. Rule based set of logics were developed to classify the exam questions under different syllabus topic. Based on the experiment output, evaluators and instructors can redesign their exam paper. 72 exam questions were used as the initial dataset. Final evaluation was done based on the total value generated from TF-IDF, TF-IDF variations, TF-IDF with cosine similarity and semantic similarity.

Keywords: Natural language processing, Cosine similarity with TF-IDF, TF-IDF variations, Semantic similarit

¹ Corresponding author. Tel.: 071-724-8167
E-mail address: krishanisudharshi@gmail.com