**Chapter 2**

**Related Works**

Though, the topic of text classification and identification is common for English language, related works regarding this topic is sparse in Bangla language. In English, the work on detection and retrieval of questions is done based on the contents generated from blogs, web portals, twitter and emails. A good number of work has been done for question detection from Community Question Answering(CQA). Deep learning for the classification of sentences is also highly practiced.

**2.1 Online Question Answering Services**

The question and query detection are mostly used and analyzed for developing online question answering Services. The most famous services are: Apple’s Siri, Microsoft’s Cortana, IBM Watson, Wolfram Alpha and Google search engine. They detect and extract answer in the following way [4]:

**2.1.1 IR-Based Approach**

The questions generated by the users are decomposed and searched in the web for finding the answer and related information. IBM Watson, Google search engine works in that way. This approach is hugely used for factoid questions.

**2.1.2 Knowledge-Based Approach**

In this approach, the answer is found by understanding the parts of the questions. Natural language processing is highly used in this approach.

**2.1.3 Hybrid Approach**

This is the combination of IR based approach and the knowledge-based approach. In this approach online databases and the web search is used to detect the question and finding the answers. Example: Apple’s Siri, Wolfram Alpha, True Knowledge Evi.

**2.2 Rule Based Approach of Question Identification**

Rule based approach is the most time-worn process of detecting questions. This approach is based on grammatical rules, taxonomy, data analysis. In [5], a rule-based approach is used to detect and analyze the questions asked in microblogging environment like twitter. The authors have made a taxonomy of questions from huge collections of questions taken from twitter. They have classified and defined and generalized the type of questions that can be found in twitter.

**2.3 Learning Based Approach of Question Identification**

Learning based approach is the automated approach of detecting questions. In [3], a learning-based approach based on lexical and syntactic features are implied to detect questions retrieved from Community Based Question Answering Services (CQA). Sequential patterns of the sentences are mined to detect the questions. The sentences are decomposed into a stream of tokens. In sequential pattern extraction, the Part-of-Speech (POS) tag of all the tokens of a sentence are used. Support Vector Machine (SVM) algorithm is used for the classification of the question and non-questions. In [7], another learning based approach is implied to identify questions from twitter. The prefix span algorithm is used for mining the frequent question patterns and SVM is also implied to differentiate the questions.

**2.4 Deep Learning for Question Answering System**

For designing question answering systems, deep learning is a very aspiring and efficient way. With the formation of Recurrent Neural Networks(RNN), it is possible to analyze longer text [9]. The authors here have designed GRU, sequence-to-sequence models and dynamic memory networks to imply the concept of deep learning.

**2.5 Convolutional Neural Network for Question Classification**

By forming Convolutional Neural Network of one layer of Convolution with the help of word2vec and static vectors, it is possible to design a state-of-the art language independent sentence classifier and question classifier [10], which performs better than most of the classifiers with the help of tuning the hyperparameters.

**2.6 Question Detection Using Machine Learning Techniques**

In [11], machine learning techniques SVM and Naïve Bayes are implied to detect Frequently Asked Question(FAQ). Information Gain (IG), ChiSquared Attribute Evaluation(Cfs), and

CfsSubset (Chi) methods are used for feature selections. Syntactic features, question words, semantic features, bag of words are used for forming the set of features.