INTRODUCTION

Language is an effective medium of communication. It explicitly represents the ideas and expressions of humans mind. More than 5000 languages and thousands of dialects exist in the world which reflects the linguistic diversity. It is difficult for an individual to know and understand all the languages of the world. Hence, the methodology of machine translation was adopted to communicate or translate the messages from one natural language to another by using computer machine.

In the 1951 census, the first census after India attained independence, 845 languages (dialects) were identified, of which 60 were spoken by at least 100,000 people each. The Indian constitution identifies 22 languages, of which six languages (Hindi, Telugu, Tamil, Bengali, Marathi and Gujarati) are spoken widely.

India has 22 officially recognized languages with hundreds of dialects and each state have interest in their regional/state languages; State governments work in their respective regional languages and the union government’s official documents and reports are in bilingual form (Hindi/English), so there is a great demand for translation of documents. In India, there are certain Machine Translation systems were already developed for most of the commonly used natural languages to translate from English to Indian languages, Indian languages to English and from regional to regional languages considering different approaches, like rule-based direct word-to-word approach and data-driven approach.

With this perspective, this proposal is made for “Development of a Hindi to Marathi Machine Translation System and its Evaluation”, will implement and evaluate the system with appropriate parameters and then made available to the target users. As of author’s knowledge, no such machine translation system for Hindi to Marathi is available.
Hindi and Marathi are mutually comprehensible languages spoken by more than 700 million people in India & abroad. Hindi and Marathi are written in the Devanagari script from left to right. Over the time, with the influence of Sanskrit in Hindi and Marathi, the vocabularies of the two languages have also become different though they still share more lot of common words.

Hindi is an Indo-Aryan language with about 577 million speakers all over the world and is ranked as 5th most widely spoken language in the world. It is one of the official language of India and is the main language used in the northern states of Rajasthan, Delhi, Haryana, Uttarakhand, Uttar Pradesh, Madhya Pradesh, Chhattisgarh, Himachal Pradesh, Jharkhand and Bihar, and is spoken in much of north and central India alongside other languages such as Punjabi, Gujarati, Marathi or Bengali. In other parts of India, as well as in Nepal, Bangladesh and Pakistan, Hindi is understood.

Hindi first started to be used in writing during the 4th century AD. It was originally written with the Brahmi script but since the 11th century AD it has been written with the Devanagari alphabet. Hindi is mostly written in a script called Nagari or Devanagari script. Hindi is normally spoken using a combination of 52 sounds, 10 vowels, 40 consonants, nasalisation and a kind of aspiration. There are thirty three basic consonants and seven other consonants that are formed with a dot diacritic on basic consonants form. In addition to this, Devanagari script has eleven vowel characters, ten vowel symbols, two symbols for nasalized sound and four consonant conjuncts.

Marathi is an Indo-Aryan language spoken by about 90 million people mainly by the native people of Maharashtra and neighbouring states. Marathi is also spoken in Israel and Mauritius. Marathi is the 4th most spoken language in India and the 15th most spoken language in
the world. Marathi is the oldest of the regional literature in Indo-Aryan languages, dating back from about 1000 AD.

Marathi first appeared in writing during the 11th century in the form of inscriptions on stones and copper plates. Long before this, Marathi must have been spoken by people living in the region. The earliest reference to spoken Marathi is found in the 8th century poem “kuvalayamA1A” of Udyotansuri. Marathi was the court language during the reign of the Yadava kings of Devgiri. From the 13th century until the mid 20th century, it was written with the Modi alphabet. This script looked more like today’s dravidian scripts. Since 1950 it has been written with the Devanagari alphabet. Marathi belongs to the group of Indo-Aryan languages which are a part of the larger group of Indo-European languages, all of which can be traced back to a common root.

Among the Indo-Aryan languages, Marathi is the southern-most language. All of the Indo-Aryan languages originated from Sanskrit. Three Prakrit languages, simpler in structure, emerged from Sanskrit. These were Saurensi, Magadhi and Maharashtri. Marathi is said to be a descendent of Maharashtri which was the Prakrit spoken by people residing in the region of Maharashtra. Marathi script consists of sixteen vowels and thirty-six consonents making a total of fifty-two alphabets. It has total 57 alphabets but some are no longer used.