|  |
| --- |
| **Sentence Segmentation** |

This project emphasizes on the concept of regular expression and basic python techniques. In this project, we try on segmenting the given text into sentences with utmost accuracy. While doing so, we take care of the special occasions or conditions such as Mr. Mrs. U.N.O and so on.

I wrote this program in Python 2.7. In this program, we first tackle the problem with brute approach. I passed the prefixes, suffixes, websites pattern at the very beginning of the code. After that, I applied those in getting the needed sentences.

Applied techniques: Replacing “.” With <use> so that it does not segment the sentence on special conditions like A.B.C., A.B., MR., MRS., .com, 123.45, .88, ….., i.e., e.g., and so on.

It covers prefixes, suffixes, digits, acronyms, starters. Given more time, I would have dealt more on punctuations and numbers pattern itself.

Doing so, applying the model, I obtained 2121 sentences from the corpus provided.