# Introduction

Sentence Segmentation is a basic text operation that segments the text into sentences. Commonly, the sentences are segmented on basis of “.” as sentence boundary indicator. But in reality this case is not always true. We have to deal with such sentences that are not ended on dot but some other tokens. Sometimes, we do not want to break sentence on a particular dot occurrence. For example, “My software version is 1.4.5.3 but still it is not working”. In this sentence, if we use simple dot as boundary indicator first sentence would break at “My software version is 1.” which is not valid.

Therefore, we need another way to handle different sentence indicators and some work with regular expressions is required to segment sentences efficiently and accurately.

# What will this software do?

This software is doing the sentence segmentation on basis of different rules, different sentence boundary indicators and different regular expressions. It takes a stream of text, parses it, tokenizes it and returns all the possible sentences in an output file.

# How to use the software DLL in your project?

First you need to get a zipped file that contains the DLL file and a simple text file (required for sentence segmentation). After getting the zipped you need to do the following steps in order to use the sentence segmentation library.

## Extract the files from zipped folder to ‘Debug’ folder of your project.

## Inside your project in solution explorer do right click on “References” and select “Add References”.

## Now go on ‘BROWSE’ tab and go to the debug folder of your project.

## Select “SentenceSegmentation.dll” file.

## Now add a namespace “using SentenceSegmentation”.

## Create an object of “Sent\_Seg”.

## Call the function “segment (string input\_file\_path, string output\_file\_path)”.

# Note:

Input file would contain a stream of text.

Output file will contain all the sentences from that text.