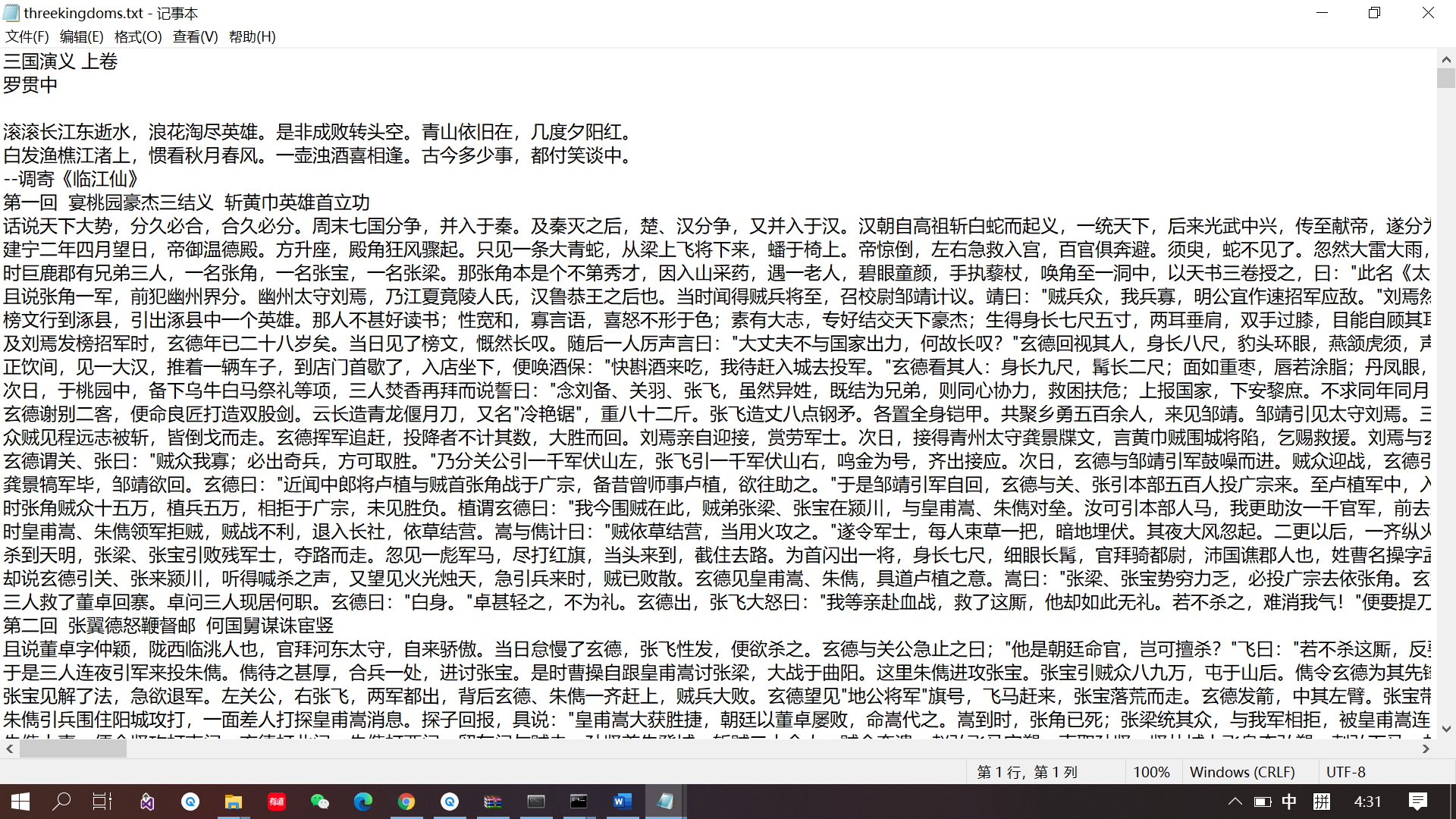
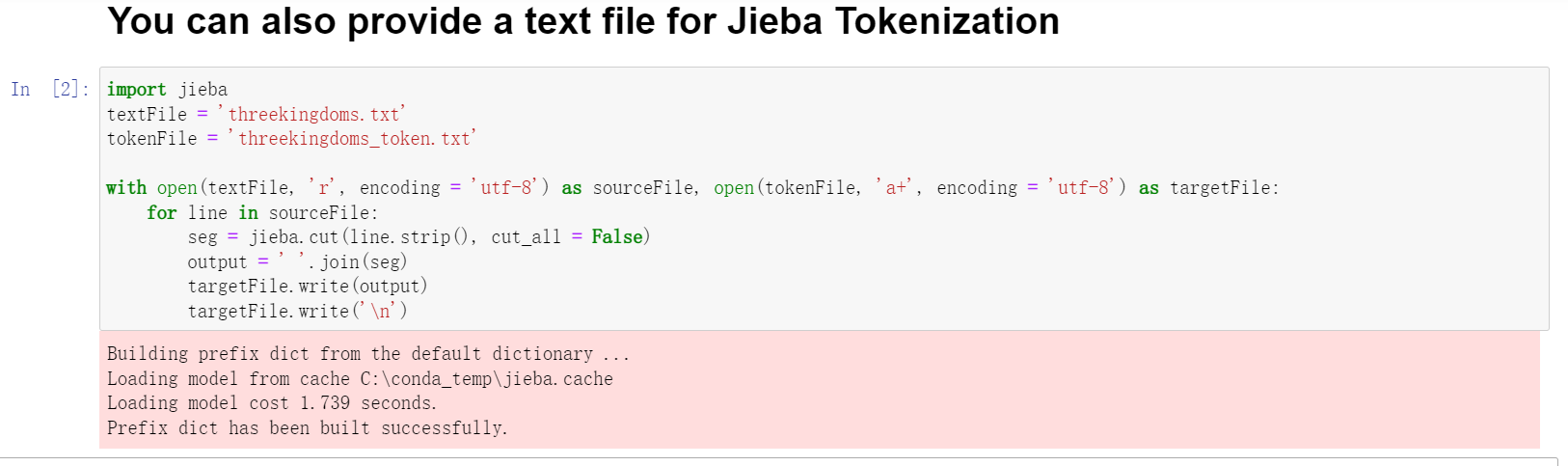
**Bonus**



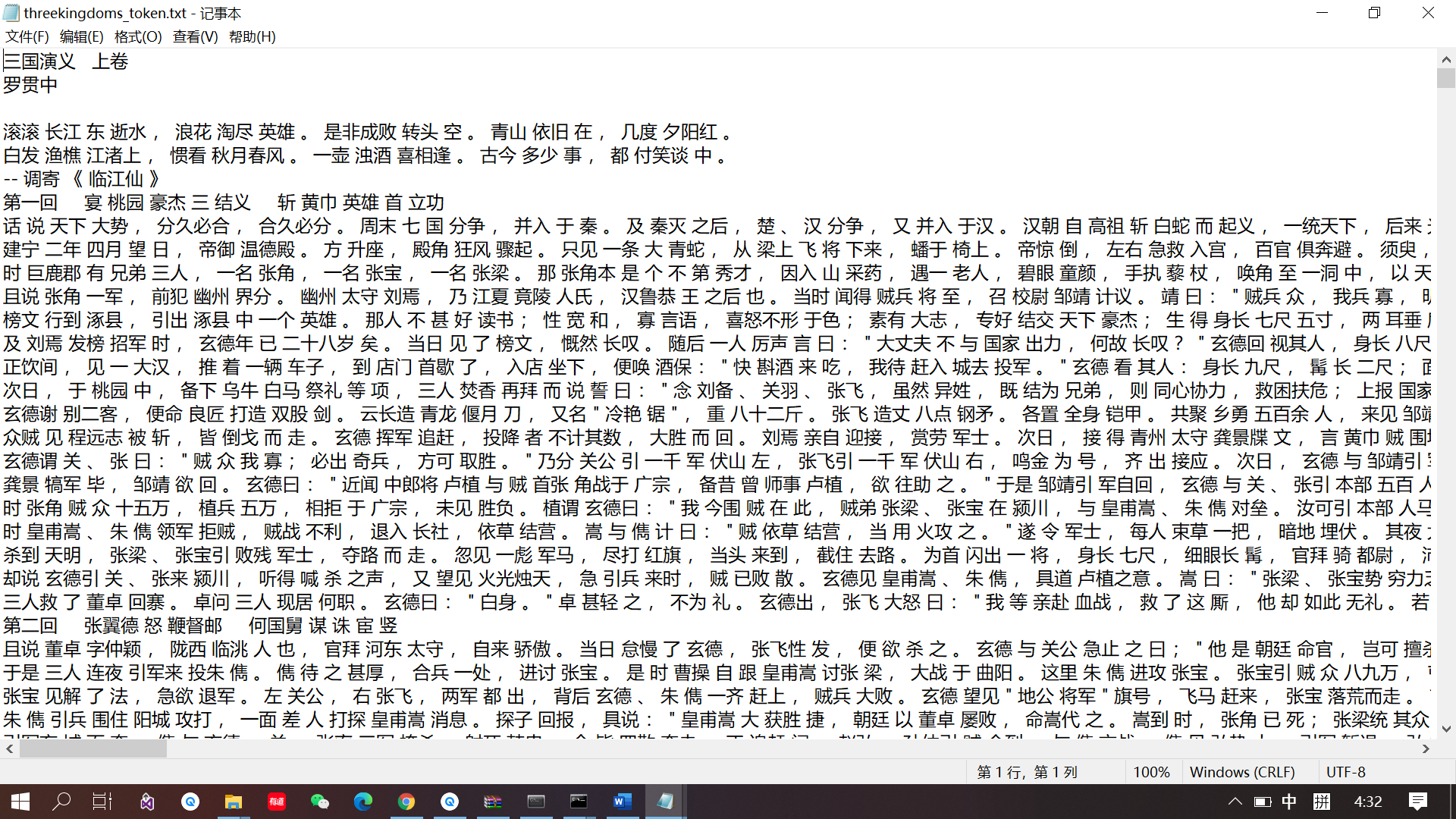
The content of the Chinese text file.



First do tokenization to the text.



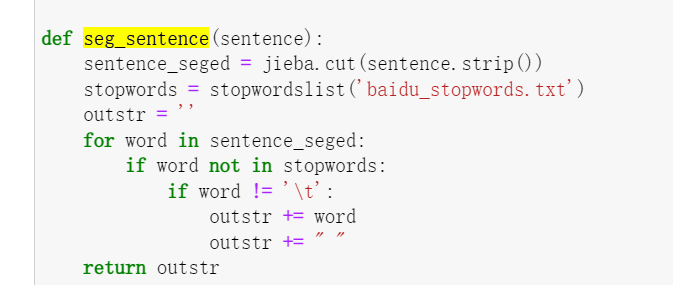
Read the text file as txt. And apply jieba.cut to txt. Save the tokenization in another file.

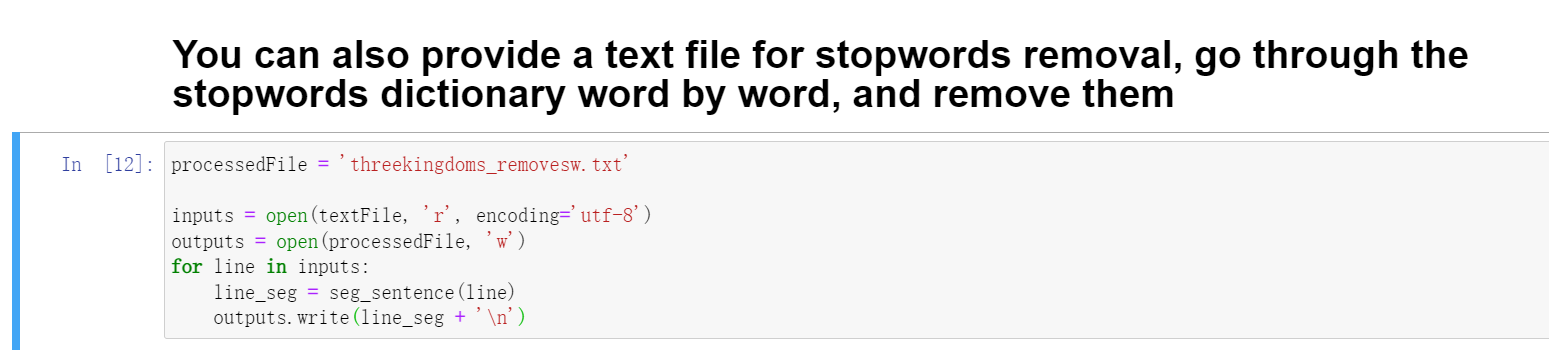


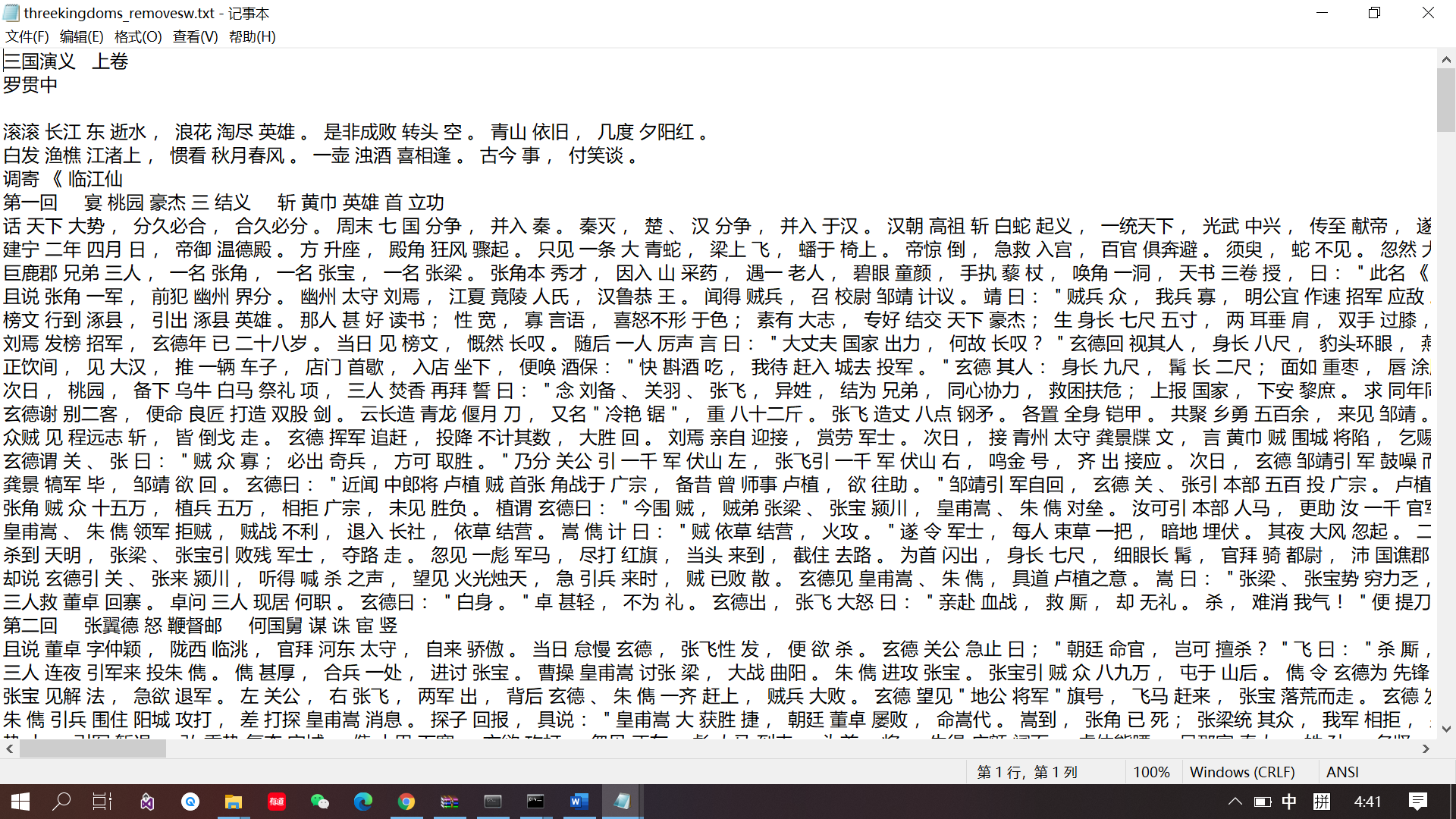
Open the tokenization file. It gives all the tokenization of the original text file.



Second, remove stop words in the text file.





Use seg\_sentence function to remove stop words in baidu\_stopwords.txt, and write the output to a file.