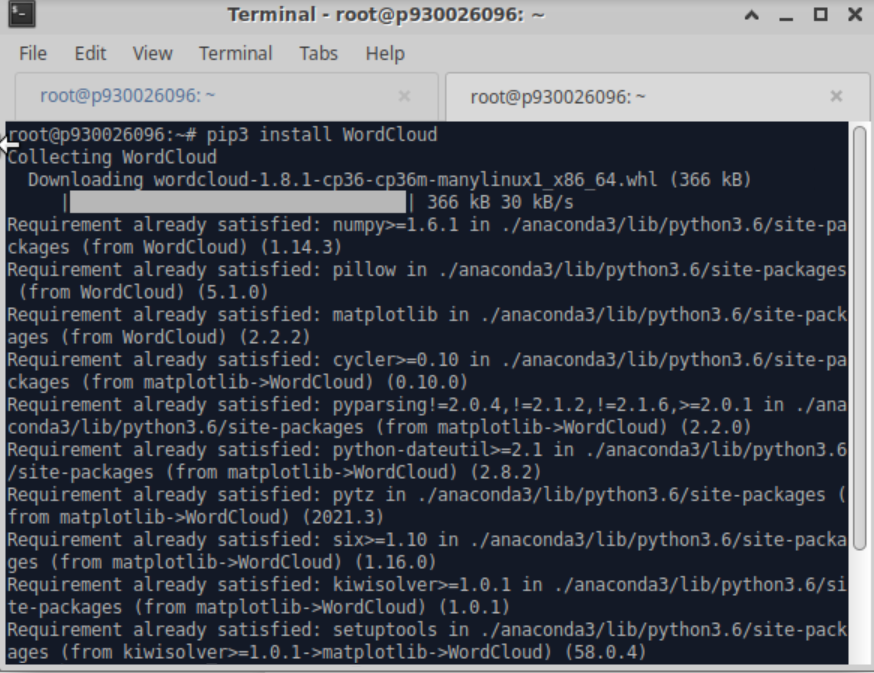
**Basic**



**Install the WordCloud library in python.**



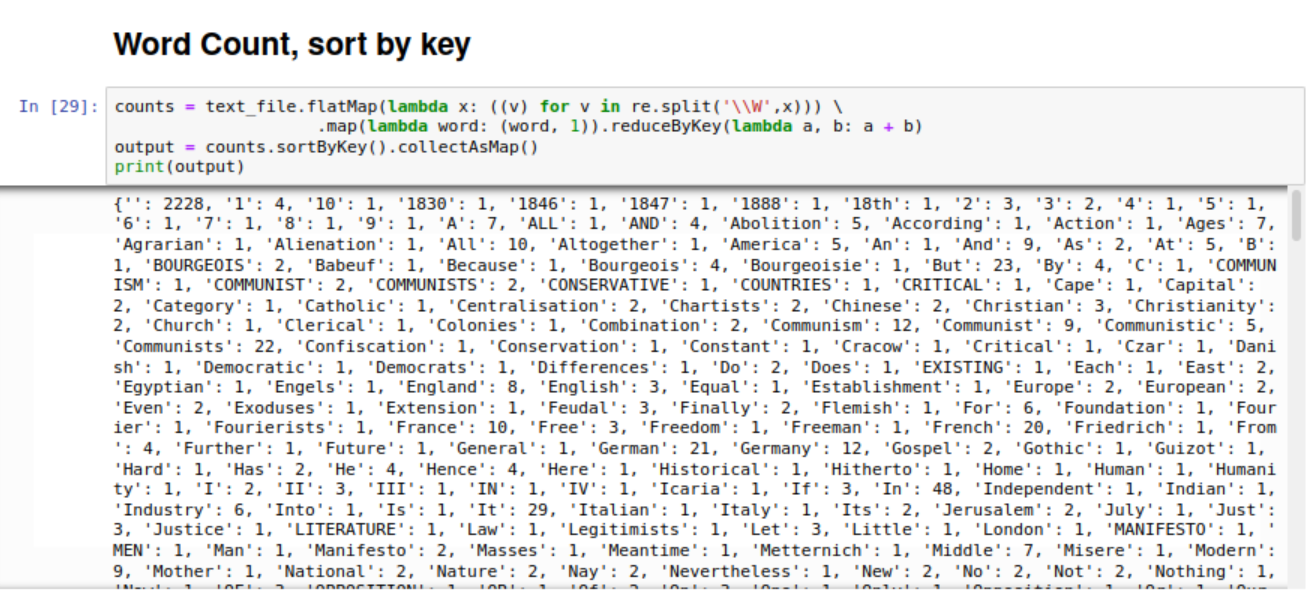
**Run the first lab WordCloud.**



**Import libraries.**



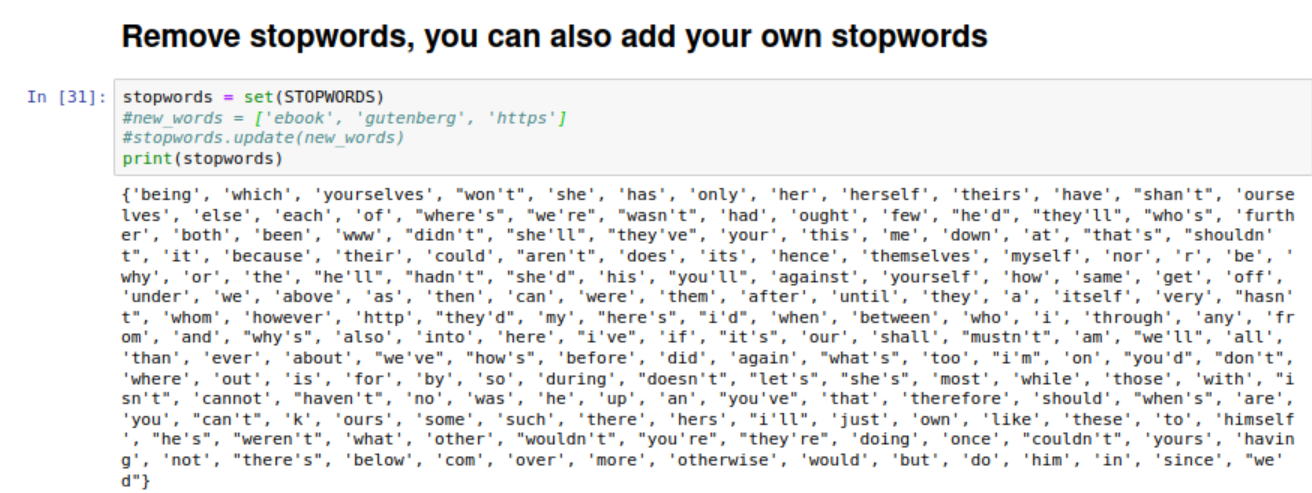
**Create a spark context and load text file.**



**Count the word and sorted by key.**



**Get the list of words**



**Remove stopwords and you can update the stopwords by update function.**



**Generate and show the word cloud.**



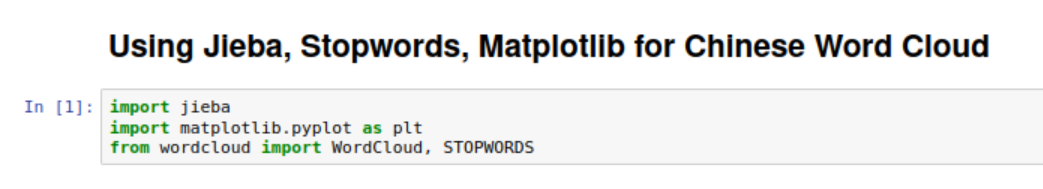
**You can include a picture for the word cloud (change the background picture size from default rectangle to picture shape).**



**The word cloud is transformed into heart shape.**



**Run the second lab WordCloudCN.**



**Import libraries.**



**Load text file and stop words file, using Jieba for cutting.**



**Generate and show word cloud, and you can use a Chinese font.**



**The default shape word cloud for Chinese text.**



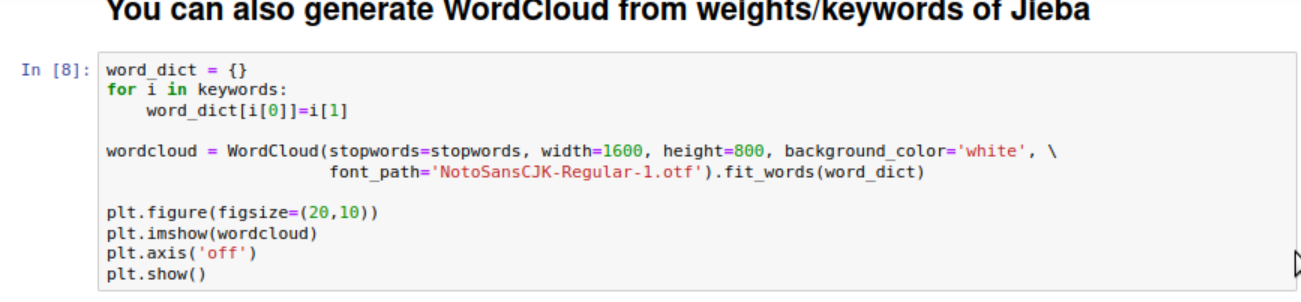
**Include a picture mask for the result.**



**The word cloud with a heart shape as the background.**



**Word Ranking from Jieba.**



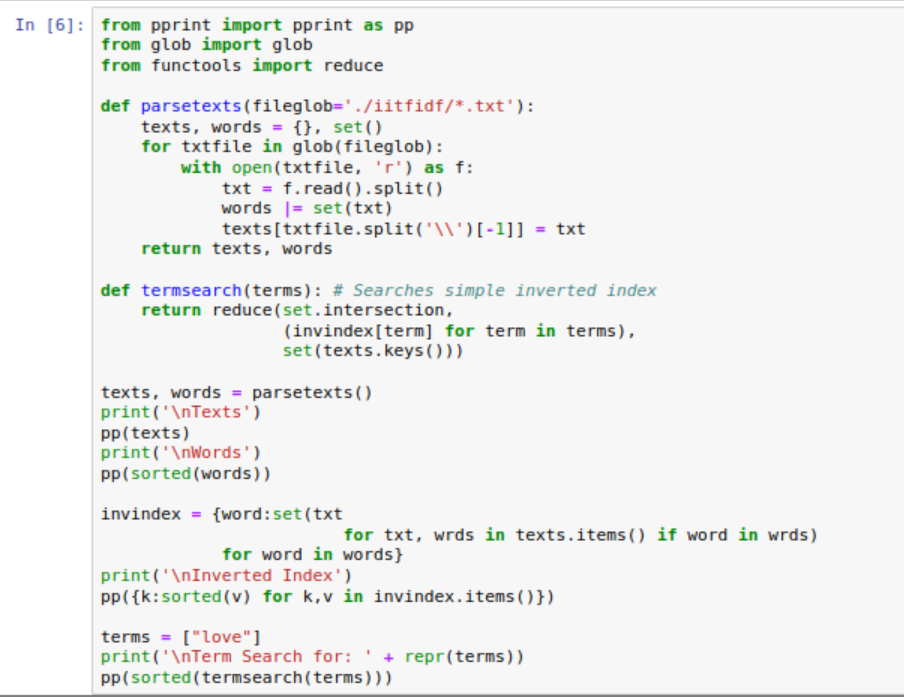
**Generate word cloud from weights/keywords of Jieba.**



**Show the word cloud with weights.**



**Run the third Lab SimpleInvIndex.**

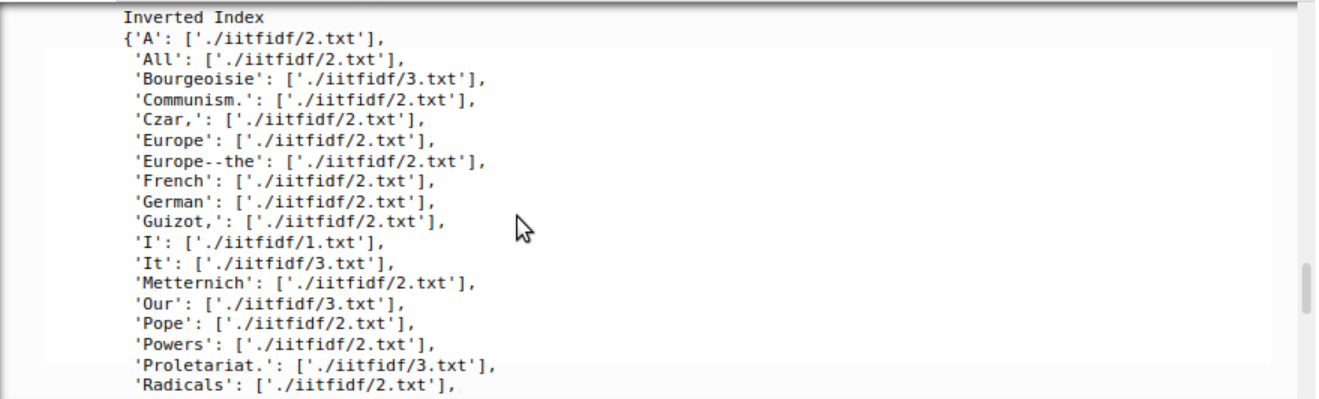


**Read all the from a local directory and split each file. Get the filename and split words**

**And define a function that can perform word searching.**



**The result of word splitting.**



**The result of inverse indexing.**



**Run the forth lab SparkInvIndex.**



**Using Spark to perform inverse indexing, the result contains the word, the document included this word and term frequency.**