```
In [15]: import string
         punc=" '!""$%&\'()*-;?[\\]^_`{|}~'"
         punc
```

```
Out[15]: " '!$%&'()*-;?[\\]^_`{|}~'"
```

```
In [18]: txt_file=open('origin football.txt',encoding="utf-8")
         #read whole file to a string
         data = txt_file.read()

         #close file
         txt_file.close()

         print(data)
```

...

remove punctutuation

```
In [19]: txtData=""
         for char in data:
             if (char not in punc  or char ==" " and char.isascii()):
                 txtData=txtData+char
```

```
In [20]: print(txtData)
```

...

Type *Markdown* and LaTeX: α^2

```
In [ ]:
```

```
In [ ]:
```

```
In [21]: import re
```

Remove numbers

```
In [22]: pattern = r'[0-9]'
         new_string = re.sub(pattern, '', txtData)
         print(new_string)
```

...

remove mails
```

```
In [23]: out = ' '.join([item for item in new_string.split() if '.com' not in item])
         print(out)
```

...

remove arabic words

```
In [24]: out=re.sub('[\u0600-\u06ff]+', '',out)
```

```
In [25]: print(out)
```

...

```
In [1]: !jupyter nbconvert 'temperature'.ipynb --to html
```

...

```
In [ ]:
```

```
In [ ]:
```