Assignment on Text Analytics Part I

Creating Word Cloud

Presented by Yubraj Das

Importing packages

In [76]:
```
import os
from wordcloud import WordCloud, STOPWORDS,ImageColorGenerator
from PIL import Image
import numpy as np
import matplotlib.pyplot as plt
from stop_words import get_stop_words
from PIL import Image
import matplotlib.pyplot as plt
```

Open the image & convert it into array format

In [77]:
```
original = np.array(Image.open('C:\\data\\Text Analytics\\hunter3.png'))
```

Findind stop words to remove from the text

In [78]:
```
stop_words = get_stop_words('english')
stop_words.append('my_new_stop_word')
stop_words.remove('my_new_stop_word')
```

Preprocessing of the image for wordcloud

In [79]:
```
wc = WordCloud(background_color='white',
               mask=original,
               collocations=False,
               width=600,
               height=300,
               contour_width=3,
               contour_color='black',
               stopwords=stop_words)
```

Import & genertaing WordCloud

In [80]:
```
text = open('C:\data\Text Analytics\Tiger.txt').read()
```

In [81]:
```
wc.generate(text_file)
```

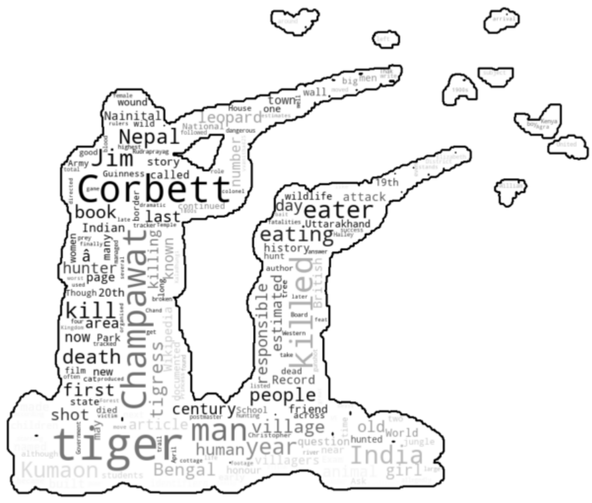
Out[81]:
```
<wordcloud.wordcloud.WordCloud at 0x228fdba6820>
```

In [82]:
```
image_colors = ImageColorGenerator(original)
wc.recolor(color_func=image_colors)
```

Out[82]:
```
<wordcloud.wordcloud.WordCloud at 0x228fdba6820>
```

Plotting WordCloud

In [83]:
```
plt.figure(figsize=(20, 10))
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
wc.to_file('wordcloud.png')
plt.show()
```



In [84]:
```
plt.imshow(wc, interpolation='bilinear')
plt.axis("off")
plt.figure()
plt.imshow(original, cmap=plt.cm.gray, interpolation='bilinear')
plt.axis("off")
plt.show()
```

In [ ]:
```

```