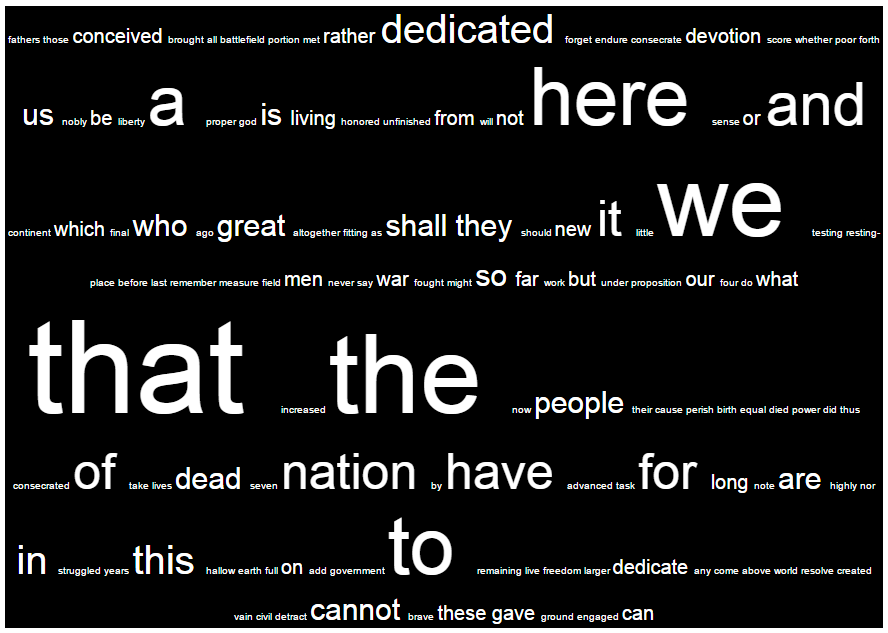
**Exercise: Make a Word Cloud**

A word (or Tag) cloud is a visual representation for text data typically used to depict keyword metadata (tags) on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or colour. This format is useful for quickly perceiving the most prominent terms and for locating a term to determine its relative prominence. We are going to use Abraham Lincoln's Gettysburg address of 1863 - this is famous for many things, including for being a short speech.

We are going to construct a simple web page. Each word will be enclosed in span tags and will have its individual font size set. A possible output might look like this:



As **this is not an exercise in HTML creation**, the text of the HTML page is provided below:

<!DOCTYPE html>

<html>

<head lang="en">

<meta charset="UTF-8">

<title>Tag Cloud Generator</title>

</head>

<body>

<div style="text-align: center; vertical-align: middle; font-family: arial; color: white; background-color:black; border:1px solid black">

<!--Your SPAN elements should be inserted here-->

</div>

</body>

</html>

The format of a span element is

<span style="font-size: XXpx"> WORD </span>

where XX is the size in pixels and WORD is the word being represented.

So for example, <span style="font-size: 20px"> our </span>

**Tasks**

You will need to do some simple analysis on the speech to complete this exercise including counting the number of occurrences of each word. You should then create the completed HTML page and write it an a .html file. local You can open this file to test it using a browser of your choice.

**Hints**

1. Import the string module. This gives string.whitespace, a string containing all of the white space characters and string.punctuation, a string containing all of the punctuation characters.
2. The first part will be to read the speech from a file and populate a dictionary using each word as a key, and value equals to the frequency of the word. Your dictionary may look something like:

{'fathers': 1, 'those': 1, 'conceived': 2, 'brought': 1, 'all': 1, 'battlefield': 1, …}

1. For each word in the dictionary of frequencies you’ll need to write a SPAN tag to your HTML file. The font size will vary depending on the frequency of the word – for example you can use count\*10 for a size, e.g. words that appear once will be size 10px, words that appear twice – 20px, etc.

**Advanced**

Modify your program to exclude common words from showing in the cloud. You can get a file of stopwords (like a, the, this, there, etc) from the internet.