Creating a word cloud

> A word cloud represents word usage in a document by resizing individual words in said document proportionally to how frequently they are used, and then jumbling them into some vaguely artistic arrangement.
>
> — Word clouds considered harmful

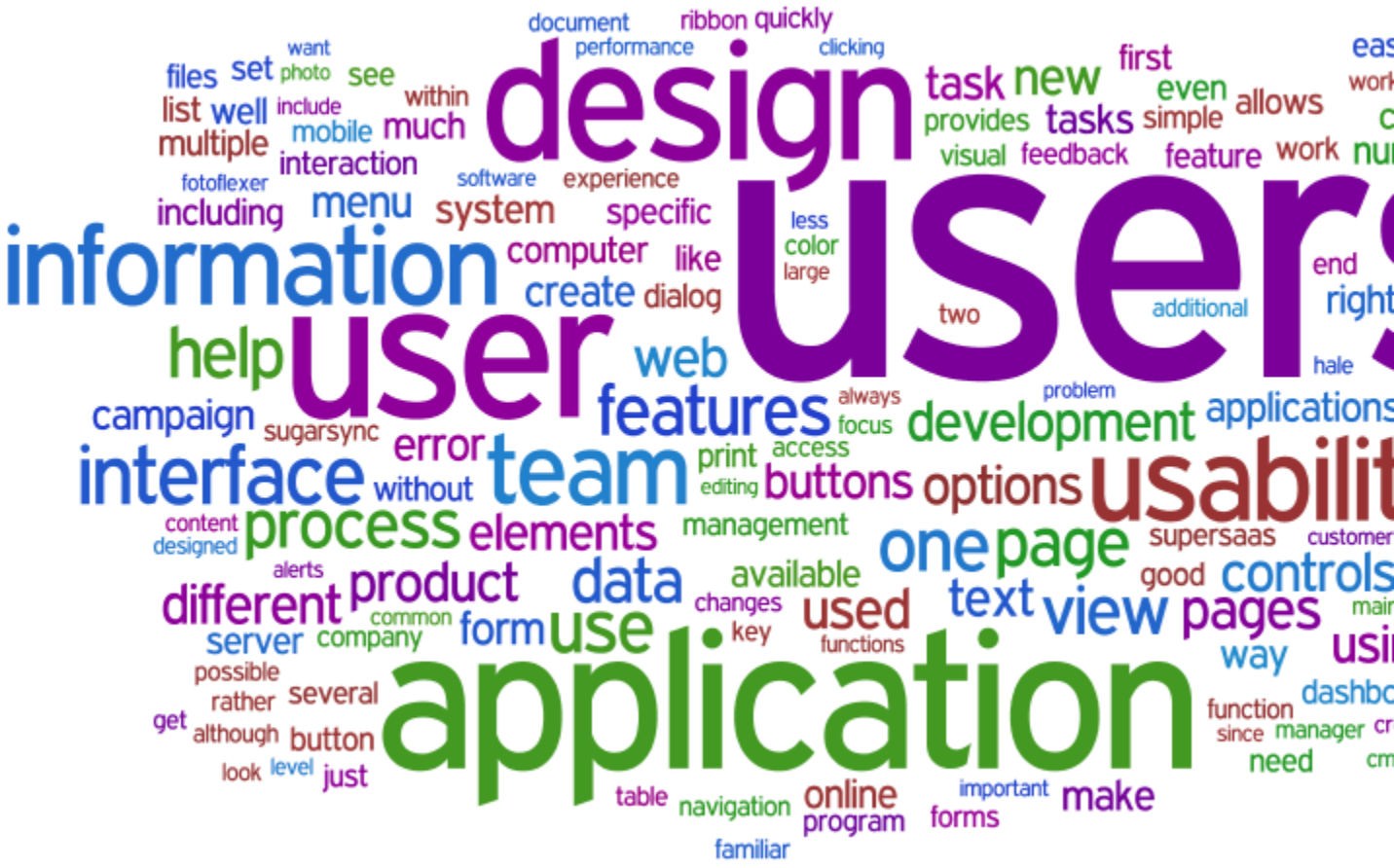

Figure 1: example of a word cloud

Essentially, a word cloud uses the size of a word to indicate the frequency of that word in a piece of text. It is meant to show word frequency not as a detailed data analysis for the content being studied. It is not intended for deeper data analysis

We'll use Python's Natural Language Toolkit to generate a list of the 200 most used Words in Moby Dick. Make sure you follow the instructions on chapter 1 of Language Processing and Python to get the data set up.

The script uses content from Natural Language Processing with Python to

generate the frequency distribution and the English stop words included with the Natural Language Toolkit to eliminate 1 letter words and punctuation signs to make the cloud more relevant to the text.

```
#!/usr/bin/env python3

# Imports json module
import json
# Loads the books we downloaded
from nltk.book import *
# Import stopwords list for English
from nltk.corpus import stopwords

# Set the stopwords words to English
stop = set(stopwords.words('english'))
'english'# Creates a frequency distribution for
fdist1 = FreqDist(text1)

# Creates a list of the 200 most common words on Moby Dick
mostCommon = fdist1.most_common(200)

# Print out most common
#print(mostCommon)

#filteredList = [w for w[0] in mostCommon if w not in stopwords]

# Write o utput to filen('cloud.json', "w") as f:
    f.write(json.dumps(mostCommon, indent=2))
'cloud.json'
```

Once we have the data, saved as `cloud.json`, we can do something like the following:

```
// Starting values were:
// width: 2140 - margin.right - margin.left
// height : 1640 - margin.top - margin.bottom
var margin = {top: 20, right: 120, bottom: 20, left: 120},
```

```
    width = 1070 - margin.right - margin.left,
    height = 820 - margin.top - margin.bottom;

d3.select()
```