# *INF 141/CS 121: Assignment 3*
# *Milestone #1*
# *Implement Indexer and Build an Index*

**Prepared by:**
Nathanial Benjamin
Theodore “Buddy” Matula
Vincent Valentine

***The Number of Documents in the Index……*** **49956**

***The Number of Unique Words in the Index……*** **174,555**

***A Sample Index (a text representation of the key/value pairs the index's dictionaries:***

Dict1 = {URL: [Words] }

Dict2 = {Word: [URLs] }

Dict3 = {URL:ID#}

Dict4 = {URL: [Title] }

Dict5 = {URL:# Words}

Dict6 = {URL:
{Word:TF-IDF} }

***The Total Size (in KB) of the Index on Disk……*** **513,900 KB**

***The Time Taken to Create the Index……*** **8 minutes**

Dict1 is a dictionary with every document's URL as the key and a list of the unique words in that document as the values.
Dict2 is the opposite: every unique word throughout all the documents are the keys and the value is a list of URLs that the word appears in.
Dict3 has the URLs of the documents as keys and the document's ID number as the value.
Dict4 has the Document's URL as the key and a list of the tokens in the document's title as the value.
Dict5 has the URLs of the documents as the key and the total number of words in the document as the value.
Dict6 has the URLs of the documents as keys and for the value has a mapping of each word in that URL to its TF-IDF. All of the dictionaries were saved in their own file in JSON format.

Specs of desktop computer used to create index:
Processor: Intel Core i7-2600k @ 3.4 GHz
RAM: 8GB
System Type: 64-bit