1. Using the tokens generated by the above program, give the incidence matrix for the set of documents given with this assignment. Let document 0 be the first column, document 1 be the second column, and so on.



1. Show the bit vector that would be the result of each of the following Boolean queries.

2.1 ((fear OR earth) AND NOT it)

<0,1,0,0,0,0,0,0,0,0> || <0> && !<0,0,0,0,0,0,1,0,0,0> =>

<0,1,0,0,0,0,0,0,0,0>

2.2 NOT(Harry OR last) AND NOT(the OR can)

~<1001000111> and ~<0010110000> =>

<0110111000> and <1101001111> => <0100001000>

1. Using the tokens generated by the above program, give the inverted index for the given set of documents.
2. Show the results of using the above index for processing the following query:
   1. the Earth

<0,4,8,9> //Earth and the are both contained in 4

* 1. Earth last

<0,4,8,9> //last and Earth located in 0,8, and 9

1. Give the bi-word index for the given set of documents.



1. Show the results of using the above bi-word index for processing the following phrase query:

Note: the words must appear in this exact order in the document.

* 1. “the Earth”

<4>

* 1. “Earth last”

<> //no document matches such a query