

**Lab Report 6**

**Submitted by: Submitted to:**

Pratyush Nepal Birodh Rijal

0519 (Artificial Intelligence Lecturer)

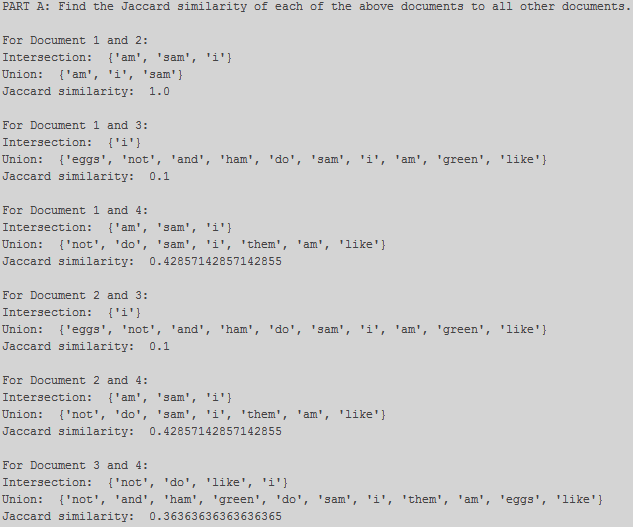
2019 ‘A’

Find the Jaccard similarity of each of the above documents to all other documents.

**Answer:**

The Jaccard similarity is defined as,

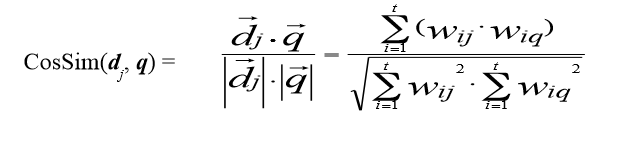
Where, A and B are two documents, denote the total number of unique words on both of them and denotes the number of common words both the documents have.



Calculate the Cosine similarity of the above documents.

**Answer:**

Cosine similarity is calculated as,



Where is a document vector which is calculated by the weights of all the words in both the documents with respect to document j. It is computed as

*wij = tfij idfi = tfij* log2 (*N/ dfi*)

where,

*tfij =fij / maxi*{*fij*}

*fij* is the frequency of ith word in jth document.

*df i* = document frequency of term *i*

