## Q1

Pseudocode:



## Q2

docA: "the sky is blue the sun is bright"

docB: "the sun in the sky is bright"

### (i)

2-shinggles





Jaccard similarity



### (ii)

Since the hash function is given:



The input matrix with bit vectors and corresponding hash value is:



Signature matrix can be derived from following steps:



Hence, the signature matrix is

