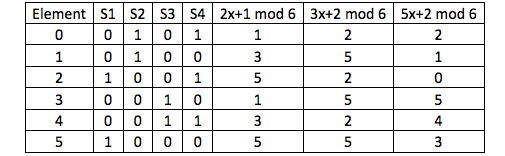
**Exercise 3.2.1**

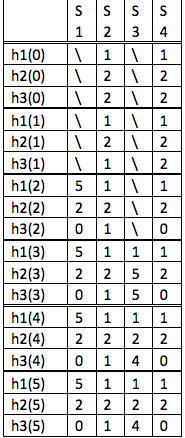
The set of the first 10 3-shingles is {“The”, “he “, “e m”, “ mo”, “mos”, “ost”, “st “, “t e”, “ ef”, “eff” }. Or {“The most effective”, “most effective way”, “effective way to”, “way to represent”, “to represent documents”, “represent documents as”, “documents as sets”, “as sets for”, “sets for purpose”, “for purpose of”}

**Exercise 3.3.3**

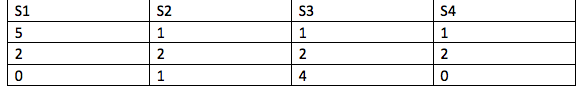
a)



Minhash Signature Matrix



The final minhash signature matrix is:



(b) Only function h3 is a true permutation

(c)The estimated Jaccard similarities are not close to the true ones at all.