

• جعبة الكلمات Bag of Words

و هي طريقة جمع كل الكلمات الموجودة في النص , ثم تحديد هل الجملة الاولى او القطعة الاولى فيها هذه الكلمات ام لا

Term Document Matrix (TDM)			
Terms	Text 1	Text 2	Text 3
An	1	0	0
epic	1	0	0
adventure	1	0	0
in	1	1	0
time	1	1	0
space	1	1	1
and	1	1	1
life	1	0	0
the	0	1	0
cosmos	0	1	0
is	0	1	0
infinite	0	1	0
I	0	0	1
like	0	0	1
books	0	0	1
about	0	0	1
history	0	0	1

Document vector

Word vector

