

Raw text

Le ver vert va  
vers le verre vert.  
(6 mots qui  
restent)

One hot encoding

ver	1
vert	1
rouge	0
vers	1
verre	1
...	...

Bag of words

ver	1
vert	2
rouge	0
vers	1
verre	1
...	...

tf-idf

ver	$1/6 \times 2$
vert	$2/6 \times 1$
rouge	0
vers	$1/6 \times 0.1$
verre	$1/6 \times 1.5$
...	tf x idf

tf:term frequency

df:document frequency

idf: inverse document  
frequency  
 $:= -\ln(df)$