

Say your corpus is \rightarrow Dog bites man. Man bites dog.
Dog eats meat. Man eats food.

- 1) Clean text eg lower all words, remove punctuations,
- 2) Remove stopwords
- 3) Apply stemming / lemmetization
- 4) Vocabulary set = {dog, bites, man, eats, meat, food}
(word tokenization)

WORD REPRESENTATION

word vector of 'dog' $\rightarrow [1 \ 0 \ 0 \ 0 \ 0 \ 0]$
of dimension 1×6

word vector of 'bites' $\rightarrow [0 \ 1 \ 0 \ 0 \ 0 \ 0]$
of dimension 1×6

word vector of 'man' $\rightarrow [0 \ 0 \ 1 \ 0 \ 0 \ 0]$
of dimension 1×6

word vector of 'eats' $\rightarrow [0 \ 0 \ 0 \ 1 \ 0 \ 0]$
of dimension 1×6

word vector of 'meat' $\rightarrow [0 \ 0 \ 0 \ 0 \ 1 \ 0]$
of dimension 1×6

word vector of 'food' $\rightarrow [0 \ 0 \ 0 \ 0 \ 0 \ 1]$
of dimension 1×6

SENTENCE REPRESENTATION

dog bites man $\rightarrow [1 \ 0 \ 0 \ 0 \ 0 \ 0] + [0 \ 1 \ 0 \ 0 \ 0 \ 0] + [0 \ 0 \ 1 \ 0 \ 0 \ 0]$
 $= [1 \ 1 \ 1 \ 0 \ 0 \ 0]$

man bites dog $\rightarrow [0 \ 0 \ 1 \ 0 \ 0 \ 0] + [0 \ 1 \ 0 \ 0 \ 0 \ 0] + [1 \ 0 \ 0 \ 0 \ 0 \ 0]$
 $= [1 \ 1 \ 1 \ 0 \ 0 \ 0]$

dog eats meat $\rightarrow [1 \ 0 \ 0 \ 0 \ 0 \ 0] + [0 \ 0 \ 0 \ 1 \ 0 \ 0] + [0 \ 0 \ 0 \ 0 \ 1 \ 0]$
 $= [1 \ 0 \ 0 \ 1 \ 1 \ 0]$

Man eats food $\rightarrow [0 \ 0 \ 1 \ 0 \ 0 \ 0] + [0 \ 0 \ 0 \ 1 \ 0 \ 0] + [0 \ 0 \ 0 \ 0 \ 0 \ 1]$
 $= [0 \ 0 \ 1 \ 1 \ 0 \ 1]$

method 1

dog bites man $\rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

man bites dog $\rightarrow \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

dog eats meat $\rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Man eats food $\rightarrow \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

method 2

Paragraph / Corpus Representation

Dog bites man. Man bites dog.
Dog eats meat. Man eats food. $\rightarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$

Method 1 - stacking

Dog bites man. Man bites dog.
Dog eats meat. Man eats food. $\rightarrow \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$

Method 2 - adding