In [6]:

```
from sklearn.feature_extraction.text import CountVectorizer
vect = CountVectorizer(binary = True)
```

In [7]:

```
corpus=["I have a german shepard","German shepard is from german","germans love gossiping"]
```

In [8]:

```
vect.fit(corpus)
```

Out[8]:

```
CountVectorizer(binary=True)
```

In [9]:

```
vocab=vect.vocabulary_
```

In [10]:

```
for key in sorted(vocab.keys()):
    print("{}:{}" .format(key,vocab[key]))
```

```
from:0
german:1
germans:2
gossiping:3
have:4
is:5
love:6
shepard:7
```

In [11]:

```
print(vect.transform(["Germany has german shepard"]).toarray())
```

```
[[0 1 0 0 0 0 0 1]]
```

In [16]:

```
from sklearn.metrics.pairwise import cosine_similarity
similarity=cosine_similarity(vect.transform(["German has German shepard"]).toarray(),vect.transform(["Geraman has capital"]).to
```

In [17]:

```
print(similarity)
```

```
[[0.]]
```

In [ ]:

```

```