

In [1]:

```
#Answer 3

import math
import re
from collections import Counter

f1 = open('doc1.txt', 'r')
f2 = open('doc2.txt', 'r')
doc1 = f1.read()
doc2 = f2.read()
#print(doc1)
#print(doc2)

WORD = re.compile(r"\w+")

def get_cosine(vec1, vec2):
    intersection = set(vec1.keys()) & set(vec2.keys())
    print("Intersection set=", intersection)
    numerator = sum([vec1[x] * vec2[x] for x in intersection])
    print("Numerator=", numerator)

    sum1 = sum([vec1[x] ** 2 for x in list(vec1.keys())])
    sum2 = sum([vec2[x] ** 2 for x in list(vec2.keys())])
    denominator = math.sqrt(sum1) * math.sqrt(sum2)
    print("Denominator=", denominator)

    if not denominator:
        return 0.0
    else:
        return float(numerator) / denominator

def text_to_vector(text):
    words = WORD.findall(text)
    return Counter(words)

vector1 = text_to_vector(doc1)
vector2 = text_to_vector(doc2)

cosine = get_cosine(vector1, vector2)

print("Cosine distance:", cosine)
print("Cosine angle:", 1-cosine)

f1.close()
f2.close()
```

```
Intersection set= {'with', 'a', 'of', 'is', 'basic', 'MATLAB'}
Numerator= 9
Denominator= 31.432467291003416
Cosine distance: 0.28632814333910006
Cosine angle: 0.7136718566608999
```