In [2]:
```python
print(avanish)
```

```
   __      _     __    _  _   _   __  _
  /  |_    \ \   / ___ (_) __/ /_
 / /|  | | | / / _ `/ _ \/ / _ / __ \
/ __  | |/ / /_/ / / / / ( __ ) / / /
/_/  |_|__/\__,_/_/ /_/_/___/_/ /_/
```

In [3]:
```python
d1 = "I am Sam "
d2 = "Sam I am "
d3 = "I do not like green eggs and ham "
d4 = "I do not like them, Sam I am "
```

In [4]:
```python
def shingles_k1(text):
    return (text.split())
```

In [5]:
```python
def shingles_k2(text):
    l=[]
    words = text.split()
    pairs = [words[i]+' '+words[i+1] for i in range(len(words)-1)]
    l.append(pairs)
    return l
```

In [17]:
```python
def shingles_k3(text):
    l=[]
    words = text.split()

    pairs = [words[i]+' '+words[i+1]+' '+words[i+2] for i in range(len(words)-2)]
    l.append(pairs)
    return l
```

In [7]:
```python
union_doc = d1+d2+d3+d4
```

In [8]: ```python
shingles_k1(union_doc)
```

Out[8]: ```
['I',
 'am',
 'Sam',
 'Sam',
 'I',
 'am',
 'I',
 'do',
 'not',
 'like',
 'green',
 'eggs',
 'and',
 'ham',
 'I',
 'do',
 'not',
 'like',
 'them,',
 'Sam',
 'I',
 'am']
```

In [9]: `shingles_k2(union_doc)`

Out[9]:
```
[['I am',
  'am Sam',
  'Sam Sam',
  'Sam I',
  'I am',
  'am I',
  'I do',
  'do not',
  'not like',
  'like green',
  'green eggs',
  'eggs and',
  'and ham',
  'ham I',
  'I do',
  'do not',
  'not like',
  'like them,',
  'them, Sam',
  'Sam I',
  'I am']]
```

```
In [10]: shingles_k3(union_doc)
```

```
Out[10]: [['I am Sam',
          'am Sam Sam',
          'Sam Sam I',
          'Sam I am',
          'I am I',
          'am I do',
          'I do not',
          'do not like',
          'not like green',
          'like green eggs',
          'green eggs and',
          'eggs and ham',
          'and ham I',
          'ham I do',
          'I do not',
          'do not like',
          'not like them,',
          'like them, Sam',
          'them, Sam I',
          'Sam I am']]
```

```python
In [13]: def shingles_k4(text):
             l=[]
             words = text.split()

             pairs = [words[i]+' '+words[i+1]+' '+words[i+2]+' '+words[i+3] for i in range(len(words)-3)]
             l.append(pairs)
             return l
```

```
In [14]: d1_U_d2 = d1+d2
```

```
In [15]: shingles_k4(d1_U_d2)
```

```
Out[15]: [['I am Sam Sam', 'am Sam Sam I', 'Sam Sam I am']]
```

In [18]: `shingles_k3(d1_U_d2)`

Out[18]: `[['I am Sam', 'am Sam Sam', 'Sam Sam I', 'Sam I am']]`

In [ ]: