

#Ex6smoothing

```
from collections import defaultdict
```

```
def calculate_ngram_probabilities(corpus):
```

```
    ngrams = defaultdict(int)
```

```
    context = defaultdict(int)
```

```
    for sentence in corpus:
```

```
        words = sentence.split()
```

```
        for i in range(len(words)-2):
```

```
            trigram = tuple(words[i:i+3])
```

```
            ngrams[trigram] += 1
```

```
            context[trigram[:2]] += 1
```

```
        print(ngrams)
```

```
        print("-----")
```

```
        print(context)
```

```
        probabilities = defaultdict(float)
```

```
    for trigram, count in ngrams.items():
```

```
        context_count = context[trigram[:2]]
```

```
        probabilities[trigram] = (count+1)/(context_count + len(ngrams))
```

```
    return probabilities
```

```
corpus=[
```

```
    "I love to playgame",
```

```
    "Python is a programming language",
```

```
    "coding is dead",
```

```
    "I hate coding in Python",
```

```
    "I am vetri venthan"
```

```
]
```

```
trigram_probabilities = calculate_ngram_probabilities(corpus)
```

```
for trigram, probability in trigram_probabilities.items():
```

```
    print(f"Trigram: {trigram}, Probability: {probability:.4f}")
```

Output:

```
student@ab1-cse106: ~/Documents/vetri
student@ab1-cse106:~/Documents/vetri$ python3 Ex6smoothing.py
defaultdict(<class 'int'>, {(('I', 'love', 'to'): 1, ('love', 'to', 'playgame'): 1})
-----
defaultdict(<class 'int'>, {(('I', 'love'): 1, ('love', 'to'): 1})
defaultdict(<class 'int'>, {(('I', 'love', 'to'): 1, ('love', 'to', 'playgame'): 1, ('Python', 'is', 'a'): 1, ('is', 'a', 'programming'): 1, ('a', 'pro
gramming', 'language'): 1})
-----
defaultdict(<class 'int'>, {(('I', 'love'): 1, ('love', 'to'): 1, ('Python', 'is'): 1, ('is', 'a'): 1, ('a', 'programming'): 1})
defaultdict(<class 'int'>, {(('I', 'love', 'to'): 1, ('love', 'to', 'playgame'): 1, ('Python', 'is', 'a'): 1, ('is', 'a', 'programming'): 1, ('a', 'pro
gramming', 'language'): 1, ('coding', 'is', 'dead'): 1})
-----
defaultdict(<class 'int'>, {(('I', 'love'): 1, ('love', 'to'): 1, ('Python', 'is'): 1, ('is', 'a'): 1, ('a', 'programming'): 1, ('coding', 'is'): 1})
defaultdict(<class 'int'>, {(('I', 'love', 'to'): 1, ('love', 'to', 'playgame'): 1, ('Python', 'is', 'a'): 1, ('is', 'a', 'programming'): 1, ('a', 'pro
gramming', 'language'): 1, ('coding', 'is', 'dead'): 1, ('I', 'hate', 'coding'): 1, ('hate', 'coding', 'in'): 1, ('coding', 'in', 'Python'): 1})
-----
defaultdict(<class 'int'>, {(('I', 'love'): 1, ('love', 'to'): 1, ('Python', 'is'): 1, ('is', 'a'): 1, ('a', 'programming'): 1, ('coding', 'is'): 1, ('
I', 'hate'): 1, ('hate', 'coding'): 1, ('coding', 'in'): 1})
defaultdict(<class 'int'>, {(('I', 'love', 'to'): 1, ('love', 'to', 'playgame'): 1, ('Python', 'is', 'a'): 1, ('is', 'a', 'programming'): 1, ('a', 'pro
gramming', 'language'): 1, ('coding', 'is', 'dead'): 1, ('I', 'hate', 'coding'): 1, ('hate', 'coding', 'in'): 1, ('coding', 'in', 'Python'): 1, ('I',
'am', 'vetri'): 1, ('am', 'vetri', 'venthan'): 1})
-----
defaultdict(<class 'int'>, {(('I', 'love'): 1, ('love', 'to'): 1, ('Python', 'is'): 1, ('is', 'a'): 1, ('a', 'programming'): 1, ('coding', 'is'): 1, ('
I', 'hate'): 1, ('hate', 'coding'): 1, ('coding', 'in'): 1, ('I', 'am'): 1, ('am', 'vetri'): 1})
Trigram: ('I', 'love', 'to'), Probability: 0.1667
student@ab1-cse106:~/Documents/vetri$
```