

Harshavardhan Reddy Mallannagari

11603389

NLP - HW2

$k=0$  ~~alphabet~~ alphabets as  $[A-Z, a-z]$

Initial	Corpus	Vocabulary
2	B, a, k, e, s, -	<del>a</del> - 26
2	b, e, t, t, y, -	a/A 6
2	L, O, U, -	b/B 15
2	b, o, u, g, h, t, -	d/D 1
2	s, o, m, e, -	e/E 20
2	b, i, t, t, e, s, -	g/G 2
2	b, u, t, t, e, s, -	h/H 4
1	b, u, t, -	i/I 3
1	i, t, -	k/K 3
1	m, a, d, e, -	l/L 2
2	h, e, f, -	m/M 4
2	b, a, t, t, e, s, -	o/O 8
1	s, a, -	r/R 12
2	b, e, t, t, e, r, -	s/S 3
1	t, o, -	t/T 25
1	m, a, k, e, -	u/U 7

Iteration - 0

Initial values

k = 1

occurrences - 12

## Resation 1

combine  $e_{18} \rightarrow e_{18}$

### Vocabulary

- 26

a 6

b 15

d 1

**e 8**

g 2

h 4

i 3

k 3

l 2

m 4

o 8

**x 0**

s 3

t 25

u 7

**ey 12**

2 B, a, k, e, s, -  
2 b, r, e, i, t, t, i, y, -  
2 d, i, o, u, -  
2 b, r, o, u, g, h, t, -  
2 s, i, o, m, i, e, -  
2 b, i, i, t, t, i, e, s, -  
2 b, r, u, t, t, i, e, s, -  
1 b, r, u, t, -  
1 i, t, -  
1 m, a, i, d, j, e, -  
2 h, e, s, -  
2 b, r, a, t, t, i, e, s, -  
2 s, i, o, -  
2 b, r, e, s, t, i, t, e, s, -  
1 t, r, o, -  
1 m, a, k, e, -

k=2

occurrences

(12)

ex, - → ex-

combine ~~ex, -~~

Vocabulary

Count

Iteration-2

Corpus

2	b, a, k, ex -
2	b, r, e, t, i, t, y, -
2	d, i, q, u, i, d -
2	b, i, q, u, g, i, h, t, -
2	s, i, o, m, i, e, -
2	b, i, t, i, t, i, e, s -
2	b, u, r, d, i, t, i, e, s -
1	b, u, i, t, -
1	i, t, -
1	m, a, d, i, e, -
2	h, i, e, s -
2	b, a, i, t, t, i, e, s -
1	s, i, o, -
2	b, i, e, t, t, i, e, s -
1	t, p, -
1	m, a, k, i, e, -

-	14
---	----

a	6
b	15
d	1
e	28
g	2
h	4
i	3
k	3
l	2
m	4
o	8
s	3
t	25
u	7

ex -	12
ex	0

k=3

occurrences (10)

### Iteration - 3

combine tit  $\rightarrow$  tt  
vocabulary.

count

Corpus

2	b, a, k, e, -
2	b, e, t, t, y, -
2	d, o, u, -
2	b, r, o, u, g, h, t, -
2	s, o, m, e, -
2	b, i, t, t, e, -
2	b, u, t, t, e, -
1	b, u, t, -
1	i, t, -
1	m, a, d, e
2	<del>h, e, s</del> h, e, s -
2	b, a, t, t, e, s -
1	s, o, -
2	b, e, t, t, e, s -
1	t, o, -
1	m, a, k, e

-	14
a	6
b	15
d	1
e	8
g	2
h	4
i	3
k	3
l	2
m	4
o	8
s	3
t	5
u	7
e, s -	12
tt	10

$k=4$

occurrences (2)

Iteration 4  
Count Corpus

Combine tt, es- → tter-  
Vocabulary

2	b, a, k, es-
2	b, e, tt, y, -
2	d, o, u, -
2	b, o, u, g, h, t, -
2	s, o, m, e-
2	b, i, tter-
2	b, u, tter-
1	b, u, t, -
1	i, t, -
1	m, a, i, d, e, -
2	h, i, es-
2	b, a, tter-
1	s, o, -
2	b, e, tter-
1	t, o, -
1	m, a, k, e, -

-	14
a	6
b	15
d	1
e	8
g	2
h	4
i	3
k	3
l	2
m	4
o	8
s	3
t	5
u	7

es-	4
tt	2
tter-	8

k=5

occurrences (4)

Iteration 5  
corpus

count

combine bie → be  
vocabulary

2	b, a, k, e, s, -
2	<del>b, a, k, e, s, -</del> b, e, t, t, i, y, -
2	d, i, o, u, -, -
2	b, i, o, u, g, h, t, -, -
2	s, i, o, m, e, -, -
2	b, i, t, t, e, s, -
2	b, i, u, t, t, e, s, -
1	b, u, t, -, -
1	i, t, -, -
1	m, a, d, e, -, -
2	h, i, e, s, -, -
2	b, a, t, t, e, s, -
1	s, i, o, -, -
2	b, e, t, t, e, s, -
1	t, i, o, -, -
1	m, a, k, e, -, -

-	14
a	6
b	11
d	1
e	4
g	2
h	4
i	3
k	3
l	2
m	4
o	8
s	3
t	5
u	7
es, -	4
tt	2
ttes, -	8
be	4

$k=6$

occurrences (4)

<u>Count</u>	<u>Iteration - 6</u> <u>Corpus</u>
2	b, a, k, e, s -
2	b, e, t, t, i, n, g, -
2	d, o, u, -
2	b, o, u, g, h, t, -
2	s, i, o, m, e, -
2	b, i, t, t, e, s -
2	b, u, t, t, e, s -
1	b, u, t, -
1	p, t, -
1	m, a, d, e, -
2	h, e, s -
2	b, a, t, t, e, s -
1	s, p, -
2	b, e, t, t, e, s -
1	t, i, o, -
1	m, a, i, k, e, -

combine o, u → ou  
Vocabulary

-	14
a	6
b	11
d	1
e	4
g	2
h	4
i	3
k	3
l	2
m	4
<b>o</b>	<b>4</b>
p, s	3
t	5
<b>u</b>	<b>3</b>
e, s -	4
t, t	2
t, t, e, s -	8
b, e	4
<b>ou</b>	<b>4</b>



$k=7$

Occurrences (4)

Iteration - 7

Count

cospus

2	ba, kies-
2	be, tt, y, -
2	d, ou, -
2	b, ou, g, hit, -
2	s, o, m, e, -
2	b, i, ttes-
2	b, u, ttes-
1	b, u, t, -
1	i, t, -
1	m, a, d, e, -
2	hies-
2	ba, ttes-
1	s, o, -
2	be, ttes-
1	t, o, -
1	m, a, k, e, -

Combine b, a  $\rightarrow$  ba  
Vocabulary

- 14	
a	2
b	7

d 1

e 4

g 2

h 4

i 3

k 3

l 2

m 4

o 4

s 3

t 5

u 3

es- 4

tt 2

ttes- 8

be 4

ou 4

ba	4
----	---

Considering  $a/A \rightarrow a$  ie,  $[A-Z, a-z]$   
in Alphabets  
case insensitive.

Iteration - 0

All the Initial values that are present in the given text.

Iteration - 1

In this iteration, combining  $e/x \rightarrow ex$   
count - 12  $[(bakes, 2), (bitter, 2), (butter, 2),$   
 $(better, 2), (batter, 2), (hex, 2)]$

Iteration - 2

In this iteration combining  $co/x \rightarrow ex-$   
count - 12

Iteration - 3

In the third iteration, we are combining  
 $t/t \rightarrow tt$

which uses 20 t's but the total  
count is 10

betty - 2

bitter - 2

better - 2

butter - 2

batter - 2

#### Iteration 4

In this iteration, we are considering the vocabulary  $tl, es \rightarrow ttes-$  which has a total count of 8

#### Iteration 5

Here in this combining  $b, e$   
count - 4 (betty, 2) (better, 2)

#### Iteration 6

For the 6th iteration considering  $o, u$  vocabulary.  
combine  $o, u \rightarrow ou$   
count - 4 (lou, 2) (bought, 2)

#### Iteration 7

For the 7th iteration considering  $b, a$  vocabulary  
for combining  $b, a \rightarrow ba$   
count - 4  
(baker - 2  
batter - 4