



HACETTEPE ÜNİVERSİTESİ-BİLİŞİM ENSTİTÜSÜ

VBM682-DOĞAL DİL İŞLEME

PROF. DR. İLYAS ÇİÇEKLİ

ÖDEV-1

TUĞÇE USTA

N22231481

BÖLÜM-1

1. a ve b karakterlerinde oluşan ve ilk ve son harfleri farklı olan karakter dizelerinin kümesi.

- $(?: (^a(a|b)^*b\$)| (^b(a|b)^*a\$))$

Sadece küçük harflerden oluşan kelimeler kümesi.

- $(^ [a-z]) ([a-z])^* \$$

Ev kelimesin çekilmiş hallerinin bir alt kümesi olan {ev,eve,evi,evde,evden,evler,evlere, evleri,evlerde,evlerden,evim,evime,evimi,evimde,evimden,evlerim,evlerime,evlerimi, evlerimde,evlerimden} kümesi.

- $^(ev)(ler|im|lerim)?(i|e|de|den)?\$$

2. Kalem ve akal kelimeleri arasındaki minimum edit distance tablosu

m	5	4	5	4	3
e	4	3	4	3	2
l	3	2	3	2	1
a	2	1	2	1	2
k	1	2	1	2	3
#	0	1	2	3	4
	#	a	k	a	l

3. a-) Unigram sayıları:

sen	dün	eve	geldin	gittin	<s>	</s>
3	1	4	2	2	4	4

Unigram olasılıkları:

$$P(\text{sen}) = 3/20 \text{ (0.15)} \quad P(\text{eve}) = 4/20 \text{ (0.2)}$$

$$P(\text{dün}) = 1/20 \text{ (0.05)} \quad P(\text{geldin}) = 2/20 \text{ (0.1)}$$

$$P(\text{gittin}) = 2/20 \text{ (0.1)} \quad P(\text{<s>}) = 4/20 \text{ (0.2)}$$

$$P(\text{</s>}) = 4/20 \text{ (0.2)}$$

Bigram sayıları:

	<s>	sen	dün	eve	geldin	gittin	</s>
<s>	0	2	0	2	0	0	0
sen	0	0	1	1	0	1	0
dün	0	0	0	1	0	0	0
eve	0	1	0	0	2	1	0
geldin	0	0	0	0	0	0	2
gittin	0	0	0	0	0	0	2
</s>	0	0	0	0	2	2	0

Bigram olasılıkları:

$P(\text{sen} \text{<s>}) = 2/4$ (0.5)	$P(\text{eve} \text{<s>}) = 2/4$ (0.5)
$P(\text{dün} \text{sen}) = 1/3$ (0.3)	$P(\text{eve} \text{sen}) = 1/3$ (0.3)
$P(\text{gittin} \text{sen}) = 1/3$ (0.3)	$P(\text{eve} \text{dün}) = 1/1$ (1)
$P(\text{sen} \text{eve}) = 1/4$ (0.2)	$P(\text{geldin} \text{eve}) = 2/4$ (0.5)
$P(\text{gittin} \text{eve}) = 1/4$ (0.2)	$P(\text{</s>} \text{geldin}) = 2/2$ (1)
$P(\text{</s>} \text{gittin}) = 2/2$ (1)	$P(\text{geldin} \text{</s>}) = 2/4$ (0.5)
$P(\text{gittin} \text{</s>}) = 2/4$ (0.5)	

b-)add-k smoothing yöntemi ile bigram olasılıkları (k=0,5)

P()	<s>	sen	dün	eve	geldin	gittin	</s>
<s>	0,066	0,333	0,066	0,333	0,066	0,066	0,066
sen	0,076	0,076	0,466	0,466	0,076	0,466	0,076
dün	0,111	0,111	0,111	0,333	0,111	0,111	0,111
eve	0,066	0,2	0,066	0,066	0,333	0,2	0,066
geldin	0,09	0,09	0,09	0,09	0,09	0,09	0,454
gittin	0,09	0,09	0,09	0,09	0,09	0,09	0,454
</s>	0,066	0,066	0,066	0,066	0,333	0,333	0,066

c-) <s>Eve dün sen geldin</s>

$P(\text{eve} | \text{<s>}) = 0,333$

$P(\text{dün} | \text{eve}) = 0,066$

$P(\text{sen} | \text{dün}) = 0,111$

$P(\text{geldin} | \text{sen}) = 0,076$

$P(\text{</s>} | \text{geldin}) = 0,454$

Sonuç = 0,000084