

CSE484/654 – INTRODUCTION TO NATURAL LANGUAGE PROCESSING

ASSIGNMENT #2

Merve GÜRLER

1901042685

- 1) Downloaded the Turkish Wikipedia Dump from link that was given in assingment PDF.
- 2) To separate each word into syllables, used the syllabilization program that was used at previous assignment. But changed the code a little bit to write Turkish characters correctly. (In first assignment Turkish characters were written in wrong format, it is fixed now.)

There is a screenshot to prove how the syllabilization program works:

```
| <do id= . url . wi ki pe dia di ? org /wi ki cu rid= tit le=" Cen giz Han" >
Cen giz Han

Cen giz Han ("Ceng his , Khan" "Cing gis , Ha an" ya da do ğum a diy la Te mu çin (an la mı : , de mir ci) Mo ğol ca: ya d
a kl Ka ra , Hı tay , Ma ve ra ünne ün ne hir , Ha rezm Ho ra san ve I ran' da ki Har zem şah lar i le Kaf kas ya' da , Gür cü ler Dest -i

Boz kır ge le ne ğin den ge len on lu teş ki la tı kul la na rak Me ri tok ra tik (li yâ ka ta bağ lı) bir or du mey da na ge ti ren Cen
a ra la rın da bir ay rım ya pıl mak sı zın say gı gös te ril me si ve ver gi den mu af tu tül ma la rı nı ka nun laş tır mış tır cen giz , Ha
ş le ye rek in san lık ta ri hi nin gör dü ğü bi ti şik sı nır la ra sa hip en bü yük im pa ra tor luk ha li ne . gel di

Cen giz Han 'ın se ce re si ya rı mi to lo jik bir se kil de sis per de si . ar ka sın da dır O nun a ta la rı nın ve ken di si nin
lu da ha . ol muş tur Cen giz Han ve o nun men sup ol du ğu Bör çü ğün ler bu ço cuk lar dan "Bo don car" ad lı en kü çük o
"Han" un va nı nı . al mış tır Cen giz Han 'ın ba ba sı Ye sü gey Ba ha dır o nun . to ru nu dur

Ka bul Han ve o nun ha le fi Am ba kay Han za ma nın da , Mo ğol lar Cin' de ki Jin Im pa ra tor lu ğu i le mü ca de le e de cek
Jin , Ha ne da nı Mo ğol la rı he zi me te . uğ rat tı Mo ğol lar bir sü re kar ma şa i çin de . da ğıl dı lar Se fa let i çin de yü z
çir me si i çin yar dım . et ti Tu ğ rul ve Ye sü gey an da i le kar de ş i lik ye mi ni et tik ten , son ra da ha son ra la rı M

Ri va yet le re ve ef sa ne ye gö re Ye sü gey Ba ha dır bir gün O non neh ri k ı y ı la rın da şa hi ni i le av la nır ken ge lin le r
di ye rek . ev len miş ti

Ev li lik ten bir sü re son ra , Ye sü gey Ta tar lar ü ze ri ne yap tı ğı bir a kin dan ge ri , dön dü ğün de Hö e lin ko ca sı nı
a şık ke mi ği şek lin de bir kan pı h tı sı i le . doğ du . Ye gü sey Ba ha dır düş man ol du ğu Ta tar ka bi le le rin den bi ri nin

A rap ve I ran lı ta rih ç i le re gö re Te mu çin 'in do ğum ta ri hi O cak . 115 An cak Çin kay nak la rın ca , Te mu çin
bu ye rin O non ya kın la rın da Dü lün- Bol dak a diy la bi li nen böl ge ol du ğu . yaz mak ta dır Bu a dın an la mı "Da lak Te

Ye gü sey Ba ha dır ve Hö e lin 'in Te mu çin' den baş ka Ha sar ve Ha ç i ad la rı nı ta sı yan i ki o ğ lu i le Te mu lun a dın d
ka dar e vin re i si ne hiz met . e de cek ti Fa kat Ye sü gey Ba ha dır e vi ne dö ner ken yol da kar şı laş tı ğı ve ko nuk ol du
Ye su gey 'in ka bi le si Te mu çin 'in an ne si Hö e lin 'in ya nı bir ka dı nın em ri al tın da kal mak is te me ye rek göç . et ti
la rı o de re ce i ler le miş ti ki "an da" ya nı kan kar de ş i ol mak i çin ye min . et ti ler Te mu çin ya şın da i ken
ır kü pek gi bi boy nu na bir tah ta dan ya pıl mış tas ma ta ka rak ka bi le le re teş hir . e dil di Fa kat bir gün bek ç i si
u zak tan Te mu çin 'i a ra yan ta kip ç i le rin bu ta ra fa gel di ği ni So han sı ra on la ra ma nı ol mak i çin he
t tu lar da ha son ra da Te mu çin 'i ko yun yü nü le do lu bir a ra ba nın i çin de . sak la dı lar Er te si gün , Tay ciutl ut l
rek so nun da a i le si ne ge ri . dön dü , Te mu çin i le ri de Cen giz Han ol du ğun da ken di si ne yar dım e den le ri hiç
ha di se nin ü ze rin den bir yıl . geç miş ti A i le sa de ce bir sü rü ye ve do kuz ta ne a ta . sa hip ti Bir gün T
ı nı sü rü sün de ki zin de at lar dan bi ri ye de ği ş tir me si i çin is rar . et ti Te mu çin a tı nı de ği ş ti rip o ra dan ay rı lı
şe lim sa de ce han ğı le ri ni al mak is te di ği ni söy le ye ter . de di Ba ba sı var lık lı bi ri o lan Bug hur ç i ha yır di ye

Te mu çin ya şın day ken ni şan lan dı ğı Bör te i le ba ba sı nın öl me den ön ce ka rar laş tır dı ğı gi bi ev li lik dü ğü nü yap m
he nü ye ni ev len di ği e ş i Bör te Mer kit ler ta ra fın dan . Ka çı rıl dı Ba tak lık la rı ya ra rak düş ma nı a ğır mağ lu bi ye te
```

To create ngrams, I used the “**nlTK**” library of python.

Firstly, I create unigram table.

```
##### calculating 1-Grams #####

with open('train.txt', encoding='utf-8') as f:
    lines = f.read()

f = open("1gram.txt", 'w', encoding = "utf-8")

n = 1
bigrams = ngrams(lines.split(), n)

for grams in bigrams:
    f.write(grams.__str__() + '\n')

with open('1gram_prob_train.txt', encoding='utf-8') as f:
    grams = f.read()

total_word_num = grams.count('\n') + 1

f = open("1gram_prob_train.txt", 'w', encoding = "utf-8")

j = 0
probability_ar = {}

for i in grams.split('\n'):
    prob = calculate_probabilities(i, grams,total_word_num,n)
    probability_ar[j] = prob
    f.write("gram: %s\tProbability: %Lf\n" % (i, prob))
    j += 1

calculate_perplexity_of_sentences(grams, probability_ar, "1gram_perplexities_train.txt")
```

It reads train data than creates unigrams, and writes unigrams to a temporary file which will be delete end of the training. Then, it reads from temporary unigram file and for each unigram it calculates probabilities. And writes a “1gram_prob_train.txt” file which is the unigram table. After that, it calculates perplexity of each sentence in train data.

```
##### To test model #####  
  
with open('test.txt', encoding='utf-8') as f:  
    test = f.read()  
  
n = 1  
bigrams_test = ngrams(test.split(), n)  
  
f = open("1gram_test.txt", 'w', encoding = "utf-8")  
  
for grams in bigrams_test:  
    f.write(grams.__str__() + '\n')  
  
with open('1gram_test.txt', encoding='utf-8') as f:  
    grams_test = f.read()  
  
with open('1gram.txt', encoding='utf-8') as f:  
    grams = f.read()  
  
f = open("1gram_prob_test.txt", 'w', encoding = "utf-8")  
  
k=0  
probability_ar_2 = {}  
  
for i in grams_test.split('\n'):  
    prob = calculate_probabilities(i, grams,total_word_num,n)  
    probability_ar_2[k] = prob  
    f.write("gram: %s\tProbability: %lf\n" % (i, prob))  
    k += 1  
  
calculate_perplexity_of_sentences(grams_test, probability_ar_2, "1gram_perplexities_test.txt")  
  
ind = 0  
ind = find_max_probability(grams, probability_ar)  
  
create_paragraphs(grams, probability_ar, ind, "1gram_paragraphs.txt",n)
```

In the second part, I tested the train data with test data. All steps are same with training part except at the end I created random paragraph.

All of the above steps are same for bigrams, trigrams, 4-grams and 5-grams.

Unfortunately, I could use only a little piece of the Turkish Wikipedia dump. Because my computer does not allow me to perform all of the Wikipedia dump.

For this reason, my output paragraphs, that are created randomly from ngrams, are not the best ones. Also, I could not choose the 5 best letters first, because my train data was too small due I did not have 5 best letters for each ngram.

This is the main flow of my language model.

I explained how did I calculate probability, ngram tables, perplexity, GT smoothing and randomly created paragraph below.

To calculate probabilities of each gram:

```
def calculate_probabilities(gram, grams, total_word_num, n):
    probability = 0

    for i in range(0, gram.count(',') + 1):
        gram = gram.replace(' ', ' ')
        gram = gram.replace('(', ' ')
        grams = grams.replace(' ', ' ')
        grams = grams.replace('(', ' ')
        string = gram.split(',')
        for j in range(0, n - 1):
            if n > 1:
                pay = string[j] + ',' + string[j + 1]
                if grams.count(string[j + 1]) != 0 and grams.count(pay) != 0:
                    if i == 0:
                        probability = (grams.count(pay) / grams.count(string[j + 1]))
                    else:
                        probability = (grams.count(pay) / grams.count(string[j + 1])) * probability
                else:
                    if i == 0:
                        probability = calculate_probability_GT_smoothing(grams, total_word_num)
                    else:
                        probability *= calculate_probability_GT_smoothing(grams, total_word_num)
            else:
                pay = string[j]
                if grams.count(string[j]) != 0 and grams.count(pay) != 0:
                    if i == 0:
                        probability = (grams.count(pay) / grams.count(string[j]))
                    else:
                        probability = (grams.count(pay) / grams.count(string[j])) * probability
                else:
                    probability *= calculate_probability_GT_smoothing(grams, total_word_num)
    return probability
```

“nltk” library’s ngram function returns ngrams as (‘some’, ‘thing’), for that reason I removed the paranthesis and split gram by ‘,’ to achieve syllables. This function calculates probability of one gram than returns the probability of the current gram. It checks if it is a unigram or not, if it is a unigram **pay** variable must have one syllable. If ngram is not a unigram, **pay** variable will be the consecutive syllables. In inner for loop, chain rule applied. If one of the pay or string[j+1] variable is equal to zero, I call the calculate_probability_GT_smoothing function.

```
def calculate_probability_GT_smoothing(grams, total_word_num):
    occurred_ones = 0
    grams = grams.replace(' ', ' ')
    grams = grams.replace('(', ' ')

    for i in grams.split('\n'):
        i = i.replace(' ', ' ')
        i = i.replace('(', ' ')
        string = i.split(',')
        if (grams.count(string[0]) == 1):
            occurred_ones += 1
    if occurred_ones > 0:
        probability = occurred_ones / total_word_num
    else:
        probability = 1 / total_word_num
    print(str(probability))
    return probability
```

calculate_probability_GT_smoothing function counts the number of words that are occurred one times in dataset, then returns this count.

Then I calculated the perplexity of each sentence.

```
def calculate_perplexity_of_sentences(grams, probabilities, name, total_word_num):
    count = 0
    perplexity = 0
    flag = 0

    f = open(name, 'w', encoding = "utf-8")
    for i in grams.split('\n'):
        i = i.replace(' ', '')
        i = i.replace('(', ' ')
        string = i.split(',')
        dot = '.'
        if string[0] == dot:
            if flag == 0:
                perplexity = probabilities[count]
            else:
                perplexity *= probabilities[count]
            f.write("%s %s %s" % (string[1], str(perplexity), '\n'))
            if perplexity != 0:
                perplexity = 1 / perplexity
                perplexity = perplexity ** (1 / total_word_num)
            else:
                print(calculate_probability_GT_smoothing(grams, total_word_num))
                perplexity = calculate_probability_GT_smoothing(grams, total_word_num)
                perplexity = 1 / perplexity
                perplexity = perplexity ** (1 / total_word_num)
            perplexity = 0
            flag = 0
            count += 1
        else:
            f.write(string[0] + ' ')
            if flag == 0:
                perplexity = probabilities[count]
            else:
                perplexity *= probabilities[count]
            flag = 1
            count += 1
    return perplexity
```

This function reads grams while a dot occurs. While there is no dot it adds grams to an empty string to achieve a full sentence then multiplies old perplexity with new grams perplexity, when a dot occurs it writes the string and calculates the final perplexity by $(1/PP(w_1, w_2, \dots, w_n))^{1/N}$ formula and writes to file.

To create random sentences:

```
def create_paragraphs(grams, probabilities, ind, name, n):
    count = 0
    string1 = {}
    string2 = ""

    f = open(name, 'w', encoding = "utf-8")
    for i in grams.split('\n'):
        if count < len(probabilities) and probabilities[count] == probabilities[ind]:
            f.write(i)
            i = i.replace(' ', '')
            i = i.replace('(', ' ')
            string = i.split(',')
            if bool(string):
                string1[n-1] = string[n-1]
                for k in range(0, 10):
                    rand_gram = choose_rand_gram(string1[n-1], grams, probabilities, n)
                    rand_gram = rand_gram.replace(' ', '')
                    rand_gram = rand_gram.replace('(', ' ')
                    string1 = rand_gram.split(',')
                    string2 = ""
                if n > 1:
                    for l in range(1, n):
                        if l < n-1:
                            string2 += string1[l] + ' '
                        else:
                            string2 += string1[l]
                    f.write(str(string2))
                else:
                    f.write(str(string1[0]))
                break
            count += 1
    f.close()
```

Firstly, using `find_max_probability` function, I take the index of the gram that has the highest probability. In `create_paragraph` function, I found

the gram that has the highest probability, and write it to file. Then, by using choose_rand_gram function, I had a gram then I wrote it to file.

```
def choose_rand_gram(last_syllable,grams,probabilities,n):
    all_prob = {}
    second_choice = {}
    string = {}
    temp = ""
    j = 0
    k = 0

    for i in grams.split('\n'):
        i = i.replace(' ','')
        i = i.replace('(',' ')
        string = i.split(',')
        if k < (grams.count('\n') - 1):
            second_choice = i
        if k < (grams.count('\n') - 1) and last_syllable == string[0]:
            temp = ""
            for t in range(0, n):
                if t < n-1:
                    temp += string[t] + ','
                else:
                    temp += string[t]
            all_prob[j] = temp
            j += 1
            k += 1
    if bool(all_prob):
        return random.choice(all_prob)
    else:
        return second_choice
```

This function takes last syllable of previous ngram than collects all ngrams that has this syllable as first element to a dictionary. Than returns from this dictionary **randomly**.

```
def find_max_probability(grams, probabilities,last_ind):
    count = 0
    max_prob = 0
    max_prob_ind = 0

    for i in grams.split('\n'):
        if probabilities[count] > max_prob and count != last_ind:
            max_prob = probabilities[count]
            max_prob_ind = count
            last_ind = count
        count += 1
    return max_prob_ind
```

This function checks the current probability is the highest one or not and is it equal to previous gram or not and returns the index of the ngram that has the highest probability.

N-Gram Tables:

1-Gram:

1-Gram Table For Train Data

| | |
|-------------------|-----------------------|
| gram: ('<do',) | Probability: 0.000000 |
| gram: ('id=',) | Probability: 0.000000 |
| gram: ('.',) | Probability: 0.000000 |
| gram: ('url',) | Probability: 0.000000 |
| gram: ('.',) | Probability: 0.000000 |
| gram: ('wi',) | Probability: 0.000000 |
| gram: ('ki',) | Probability: 0.000000 |
| gram: ('pe',) | Probability: 0.000000 |
| gram: ('dia',) | Probability: 0.000000 |
| gram: ('di',) | Probability: 0.000000 |
| gram: ('?',) | Probability: 0.000000 |
| gram: ('org',) | Probability: 0.000000 |
| gram: ('/wi',) | Probability: 0.000000 |
| gram: ('ki',) | Probability: 0.000000 |
| gram: ('cu',) | Probability: 0.000000 |
| gram: ('rid=',) | Probability: 0.000000 |
| gram: ('tit',) | Probability: 0.000000 |
| gram: ('le=',) | Probability: 0.000000 |
| gram: ('Cen',) | Probability: 0.000000 |
| gram: ('giz',) | Probability: 0.000000 |
| gram: ('Han"',) | Probability: 0.000000 |
| gram: ('>',) | Probability: 0.000000 |
| gram: ('Cen',) | Probability: 0.000000 |
| gram: ('giz',) | Probability: 0.000000 |
| gram: ('Han',) | Probability: 0.000000 |
| gram: ('Cen',) | Probability: 0.000000 |
| gram: ('giz',) | Probability: 0.000000 |
| gram: ('Han',) | Probability: 0.000000 |
| gram: ('("Ceng',) | Probability: 0.000000 |
| gram: ('his',) | Probability: 0.000000 |
| gram: ('',) | Probability: 0.000000 |
| gram: ('Khan"',) | Probability: 0.000000 |

1-Gram Table For Test Data

| | |
|----------------|-----------------------|
| gram: ('An',) | Probability: 0.000000 |
| gram: ('tik',) | Probability: 0.000000 |
| gram: ('ya',) | Probability: 0.000000 |
| gram: ('zar',) | Probability: 0.000000 |
| gram: ('lar',) | Probability: 0.000000 |
| gram: ('ya',) | Probability: 0.000000 |
| gram: ('p1',) | Probability: 0.000000 |
| gram: ('nin',) | Probability: 0.000000 |
| gram: ('mi',) | Probability: 0.000000 |
| gram: ('ma',) | Probability: 0.000000 |
| gram: ('r1',) | Probability: 0.000000 |
| gram: ('nin',) | Probability: 0.000000 |
| gram: ('ol',) | Probability: 0.000000 |
| gram: ('du',) | Probability: 0.000000 |
| gram: ('gu',) | Probability: 0.000000 |
| gram: ('nu',) | Probability: 0.000000 |
| gram: ('.',) | Probability: 0.000000 |
| gram: ('kay',) | Probability: 0.000000 |
| gram: ('det',) | Probability: 0.000000 |
| gram: ('mek',) | Probability: 0.000000 |
| gram: ('te',) | Probability: 0.000000 |
| gram: ('dir',) | Probability: 0.000000 |
| gram: ('Ay',) | Probability: 0.000000 |
| gram: ('r1',) | Probability: 0.000000 |
| gram: ('ca',) | Probability: 0.000000 |
| gram: ('"un",) | Probability: 0.000000 |
| gram: ('a',) | Probability: 0.000000 |
| gram: ('d1',) | Probability: 0.000000 |
| gram: ('da',) | Probability: 0.000000 |
| gram: ('.',) | Probability: 0.000000 |
| gram: ('geç',) | Probability: 0.000000 |
| gram: ('mek',) | Probability: 0.000000 |

2-Gram:

2-Gram Table For Train Data

| | |
|--------------------------|-----------------------|
| gram: ('<do', 'id=') | Probability: 0.500000 |
| gram: ('id=', '.') | Probability: 0.004950 |
| gram: ('.', 'url') | Probability: 0.500000 |
| gram: ('url', '.') | Probability: 0.004950 |
| gram: ('.', 'wi') | Probability: 0.500000 |
| gram: ('wi', 'ki') | Probability: 0.012500 |
| gram: ('ki', 'pe') | Probability: 0.250000 |
| gram: ('pe', 'dia') | Probability: 0.250000 |
| gram: ('dia', 'di') | Probability: 0.019608 |
| gram: ('di', '?') | Probability: 0.500000 |
| gram: ('?', 'org') | Probability: 0.500000 |
| gram: ('org', '/wi') | Probability: 0.500000 |
| gram: ('/wi', 'ki') | Probability: 0.012500 |
| gram: ('ki', 'cu') | Probability: 0.071429 |
| gram: ('cu', 'rid=') | Probability: 0.500000 |
| gram: ('rid=', 'tit') | Probability: 0.250000 |
| gram: ('tit', 'le="') | Probability: 0.500000 |
| gram: ('le=', 'Cen') | Probability: 0.026316 |
| gram: ('Cen', 'giz') | Probability: 0.475000 |
| gram: ('giz', 'Han"') | Probability: 0.500000 |
| gram: ('Han"', '>') | Probability: 0.500000 |
| gram: ('>', 'Cen') | Probability: 0.026316 |
| gram: ('Cen', 'giz') | Probability: 0.475000 |
| gram: ('giz', 'Han') | Probability: 0.357143 |
| gram: ('Han', 'Cen') | Probability: 0.026316 |
| gram: ('Cen', 'giz') | Probability: 0.475000 |
| gram: ('giz', 'Han') | Probability: 0.357143 |
| gram: ('Han', '("Ceng") | Probability: 0.500000 |
| gram: ('("Ceng", 'his') | Probability: 0.500000 |
| gram: ('his', ',') | Probability: 0.000000 |
| gram: (',', 'Khan"') | Probability: 0.000071 |
| gram: ('Khan"', '"Cing') | Probability: 0.500000 |

2-Gram Table For Test Data

| | |
|----------------------|-----------------------|
| gram: ('An', 'tik') | Probability: 0.000208 |
| gram: ('tik', 'ya') | Probability: 0.000208 |
| gram: ('ya', 'zar') | Probability: 0.000208 |
| gram: ('zar', 'lar') | Probability: 0.000208 |
| gram: ('lar', 'ya') | Probability: 0.000208 |
| gram: ('ya', 'p1') | Probability: 0.000208 |
| gram: ('p1', 'n1n') | Probability: 0.000208 |
| gram: ('n1n', 'mi') | Probability: 0.000208 |
| gram: ('mi', 'ma') | Probability: 0.000208 |
| gram: ('ma', 'r1') | Probability: 0.000208 |
| gram: ('r1', 'n1n') | Probability: 0.047619 |
| gram: ('n1n', 'ol') | Probability: 0.000208 |
| gram: ('ol', 'du') | Probability: 0.190476 |
| gram: ('du', 'ġu') | Probability: 0.352941 |
| gram: ('ġu', 'nu') | Probability: 0.035714 |
| gram: ('nu', '.') | Probability: 0.004950 |
| gram: ('.', 'kay') | Probability: 0.187500 |
| gram: ('kay', 'det') | Probability: 0.000208 |
| gram: ('det', 'mek') | Probability: 0.000208 |
| gram: ('mek', 'te') | Probability: 0.068966 |
| gram: ('te', 'dir') | Probability: 0.200000 |
| gram: ('dir', 'Ay') | Probability: 0.000208 |
| gram: ('Ay', 'r1') | Probability: 0.000208 |
| gram: ('r1', 'ca') | Probability: 0.000208 |
| gram: ('ca', '"un') | Probability: 0.000208 |
| gram: ('"un', 'a') | Probability: 0.000208 |
| gram: ('a', 'd1') | Probability: 0.041667 |
| gram: ('d1', 'da') | Probability: 0.000208 |
| gram: ('da', '.') | Probability: 0.014851 |
| gram: ('.', 'geç') | Probability: 0.250000 |
| gram: ('geç', 'mek') | Probability: 0.000208 |
| gram: ('mek', 'te') | Probability: 0.068966 |

3-Gram:

3-Gram Table For Train Data

| | |
|---------------------------------|-----------------------|
| gram: ('<do', 'id=', '.') | Probability: 0.000022 |
| gram: ('id=', '.', 'url') | Probability: 0.002934 |
| gram: ('.', 'url', '.') | Probability: 0.000029 |
| gram: ('url', '.', 'wi') | Probability: 0.002934 |
| gram: ('.', 'wi', 'ki') | Probability: 0.000185 |
| gram: ('wi', 'ki', 'pe') | Probability: 0.001852 |
| gram: ('ki', 'pe', 'dia') | Probability: 0.037037 |
| gram: ('pe', 'dia', 'di') | Probability: 0.000228 |
| gram: ('dia', 'di', '?') | Probability: 0.011619 |
| gram: ('di', '?', 'org') | Probability: 0.296296 |
| gram: ('?', 'org', '/wi') | Probability: 0.296296 |
| gram: ('org', '/wi', 'ki') | Probability: 0.000185 |
| gram: ('/wi', 'ki', 'cu') | Probability: 0.000151 |
| gram: ('ki', 'cu', 'rid=') | Probability: 0.042328 |
| gram: ('cu', 'rid=', 'tit') | Probability: 0.074074 |
| gram: ('rid=', 'tit', 'le="') | Probability: 0.148148 |
| gram: ('tit', 'le=', 'Cen') | Probability: 0.000821 |
| gram: ('le=', 'Cen', 'giz') | Probability: 0.014074 |
| gram: ('Cen', 'giz', 'Han"') | Probability: 0.281481 |
| gram: ('giz', 'Han"', '>') | Probability: 0.296296 |
| gram: ('Han"', '>', 'Cen') | Probability: 0.000821 |
| gram: ('>', 'Cen', 'giz') | Probability: 0.014074 |
| gram: ('Cen', 'giz', 'Han') | Probability: 0.143613 |
| gram: ('giz', 'Han', 'Cen') | Probability: 0.000586 |
| gram: ('Han', 'Cen', 'giz') | Probability: 0.014074 |
| gram: ('Cen', 'giz', 'Han') | Probability: 0.143613 |
| gram: ('giz', 'Han', '("Ceng') | Probability: 0.211640 |
| gram: ('Han', '("Ceng', 'his') | Probability: 0.296296 |
| gram: ('("Ceng', 'his', ',') | Probability: 0.000000 |
| gram: ('his', ', ', 'Khan"') | Probability: 0.000000 |
| gram: (' ', 'Khan"', '"Cing') | Probability: 0.000021 |
| gram: ('Khan"', '"Cing', 'gis') | Probability: 0.296296 |

3-Gram Table For Test Data

| | |
|-----------------------------|-----------------------|
| gram: ('An', 'tik', 'ya') | Probability: 0.000000 |
| gram: ('tik', 'ya', 'zar') | Probability: 0.000000 |
| gram: ('ya', 'zar', 'lar') | Probability: 0.000000 |
| gram: ('zar', 'lar', 'ya') | Probability: 0.000000 |
| gram: ('lar', 'ya', 'p1') | Probability: 0.000000 |
| gram: ('ya', 'p1', 'n1n') | Probability: 0.000000 |
| gram: ('p1', 'n1n', 'mi') | Probability: 0.000000 |
| gram: ('n1n', 'mi', 'ma') | Probability: 0.000000 |
| gram: ('mi', 'ma', 'r1') | Probability: 0.000000 |
| gram: ('ma', 'r1', 'n1n') | Probability: 0.000001 |
| gram: ('r1', 'n1n', 'ol') | Probability: 0.000000 |
| gram: ('n1n', 'ol', 'du') | Probability: 0.000013 |
| gram: ('ol', 'du', 'gu') | Probability: 0.056242 |
| gram: ('du', 'gu', 'nu') | Probability: 0.001067 |
| gram: ('gu', 'nu', '.') | Probability: 0.000002 |
| gram: ('nu', '.', 'kay') | Probability: 0.000413 |
| gram: ('.', 'kay', 'det') | Probability: 0.000000 |
| gram: ('kay', 'det', 'mek') | Probability: 0.000000 |
| gram: ('det', 'mek', 'te') | Probability: 0.000002 |
| gram: ('mek', 'te', 'dir') | Probability: 0.006539 |
| gram: ('te', 'dir', 'Ay') | Probability: 0.000000 |
| gram: ('dir', 'Ay', 'r1') | Probability: 0.000000 |
| gram: ('Ay', 'r1', 'ca') | Probability: 0.000000 |
| gram: ('r1', 'ca', '"un') | Probability: 0.000000 |
| gram: ('ca', '"un', 'a') | Probability: 0.000000 |
| gram: ('"un', 'a', 'd1') | Probability: 0.000001 |
| gram: ('a', 'd1', 'da') | Probability: 0.000000 |
| gram: ('d1', 'da', '.') | Probability: 0.000000 |
| gram: ('da', '.', 'geç') | Probability: 0.002200 |
| gram: ('.', 'geç', 'mek') | Probability: 0.000000 |
| gram: ('geç', 'mek', 'te') | Probability: 0.000002 |
| gram: ('mek', 'te', 'dir') | Probability: 0.006539 |

4-Gram:

4-Gram Table For Train Data

| | |
|---------------------------------------|-----------------------|
| gram: ('<do', 'id=', '.', 'url') | Probability: 0.000003 |
| gram: ('id=', '.', 'url', '.') | Probability: 0.000000 |
| gram: ('.', 'url', '.', 'wi') | Probability: 0.000013 |
| gram: ('url', '.', 'wi', 'ki') | Probability: 0.000000 |
| gram: ('.', 'wi', 'ki', 'pe') | Probability: 0.000010 |
| gram: ('wi', 'ki', 'pe', 'dia') | Probability: 0.000003 |
| gram: ('ki', 'pe', 'dia', 'di') | Probability: 0.000001 |
| gram: ('pe', 'dia', 'di', '?') | Probability: 0.000051 |
| gram: ('dia', 'di', '?', 'org') | Probability: 0.000205 |
| gram: ('di', '?', 'org', '/wi') | Probability: 0.133484 |
| gram: ('?', 'org', '/wi', 'ki') | Probability: 0.000002 |
| gram: ('org', '/wi', 'ki', 'cu') | Probability: 0.000000 |
| gram: ('/wi', 'ki', 'cu', 'rid=') | Probability: 0.000002 |
| gram: ('ki', 'cu', 'rid=', 'tit') | Probability: 0.000341 |
| gram: ('cu', 'rid=', 'tit', 'le="') | Probability: 0.033371 |
| gram: ('rid=', 'tit', 'le=', 'Cen') | Probability: 0.000005 |
| gram: ('tit', 'le=', 'Cen', 'giz') | Probability: 0.000317 |
| gram: ('le=', 'Cen', 'giz', 'Han"') | Probability: 0.000334 |
| gram: ('Cen', 'giz', 'Han"', '>') | Probability: 0.120469 |
| gram: ('giz', 'Han"', '>', 'Cen') | Probability: 0.000019 |
| gram: ('Han"', '>', 'Cen', 'giz') | Probability: 0.000317 |
| gram: ('>', 'Cen', 'giz', 'Han') | Probability: 0.000122 |
| gram: ('Cen', 'giz', 'Han', 'Cen') | Probability: 0.000009 |
| gram: ('giz', 'Han', 'Cen', 'giz') | Probability: 0.000162 |
| gram: ('Han', 'Cen', 'giz', 'Han') | Probability: 0.000122 |
| gram: ('Cen', 'giz', 'Han', '("Ceng') | Probability: 0.061464 |
| gram: ('giz', 'Han', '("Ceng', 'his') | Probability: 0.068104 |
| gram: ('Han', '("Ceng', 'his', ',') | Probability: 0.000000 |
| gram: ('("Ceng', 'his', ',, 'Khan"') | Probability: 0.000000 |
| gram: ('his', ',, 'Khan"', '"Cing') | Probability: 0.000000 |
| gram: ('', 'Khan"', '"Cing', 'gis') | Probability: 0.000000 |
| gram: ('Khan"', '"Cing', 'gis', 'Ha') | Probability: 0.000389 |

4-Gram Table For Test Data

| | |
|-----------------------------------|-----------------------|
| gram: ('An', 'tik', 'ya', 'zar') | Probability: 0.000000 |
| gram: ('tik', 'ya', 'zar', 'lar') | Probability: 0.000000 |
| gram: ('ya', 'zar', 'lar', 'ya') | Probability: 0.000000 |
| gram: ('zar', 'lar', 'ya', 'p1') | Probability: 0.000000 |
| gram: ('lar', 'ya', 'p1', 'n1n') | Probability: 0.000000 |
| gram: ('ya', 'p1', 'n1n', 'mi') | Probability: 0.000000 |
| gram: ('p1', 'n1n', 'mi', 'ma') | Probability: 0.000000 |
| gram: ('n1n', 'mi', 'ma', 'r1') | Probability: 0.000000 |
| gram: ('mi', 'ma', 'r1', 'n1n') | Probability: 0.000000 |
| gram: ('ma', 'r1', 'n1n', 'ol') | Probability: 0.000000 |
| gram: ('r1', 'n1n', 'ol', 'du') | Probability: 0.000000 |
| gram: ('n1n', 'ol', 'du', 'gu') | Probability: 0.000000 |
| gram: ('ol', 'du', 'gu', 'nu') | Probability: 0.000000 |
| gram: ('du', 'gu', 'nu', '.') | Probability: 0.000000 |
| gram: ('gu', 'nu', '.', 'kay') | Probability: 0.000000 |
| gram: ('nu', '.', 'kay', 'det') | Probability: 0.000000 |
| gram: ('.', 'kay', 'det', 'mek') | Probability: 0.000000 |
| gram: ('kay', 'det', 'mek', 'te') | Probability: 0.000000 |
| gram: ('det', 'mek', 'te', 'dir') | Probability: 0.000000 |
| gram: ('mek', 'te', 'dir', 'Ay') | Probability: 0.000000 |
| gram: ('te', 'dir', 'Ay', 'r1') | Probability: 0.000000 |
| gram: ('dir', 'Ay', 'r1', 'ca') | Probability: 0.000000 |
| gram: ('Ay', 'r1', 'ca', '"un') | Probability: 0.000000 |
| gram: ('r1', 'ca', '"un', 'a') | Probability: 0.000000 |
| gram: ('ca', '"un', 'a', 'd1') | Probability: 0.000000 |
| gram: ('"un', 'a', 'd1', 'da') | Probability: 0.000000 |
| gram: ('a', 'd1', 'da', '.') | Probability: 0.000000 |
| gram: ('d1', 'da', '.', 'geç') | Probability: 0.000000 |
| gram: ('da', '.', 'geç', 'mek') | Probability: 0.000000 |
| gram: ('.', 'geç', 'mek', 'te') | Probability: 0.000000 |
| gram: ('geç', 'mek', 'te', 'dir') | Probability: 0.000000 |
| gram: ('mek', 'te', 'dir', ',') | Probability: 0.000000 |

5-Gram:

5-Gram Table For Train Data

| | |
|---|-----------------------|
| gram: ('.', 'wi', 'ki', 'pe', 'dia') | Probability: 0.000011 |
| gram: ('wi', 'ki', 'pe', 'dia', 'di') | Probability: 0.000000 |
| gram: ('ki', 'pe', 'dia', 'di', '?') | Probability: 0.000025 |
| gram: ('pe', 'dia', 'di', '?', 'org') | Probability: 0.000025 |
| gram: ('dia', 'di', '?', 'org', '/wi') | Probability: 0.000025 |
| gram: ('di', '?', 'org', '/wi', 'ki') | Probability: 0.000001 |
| gram: ('?', 'org', '/wi', 'ki', 'cu') | Probability: 0.000000 |
| gram: ('org', '/wi', 'ki', 'cu', 'rid=') | Probability: 0.000000 |
| gram: ('/wi', 'ki', 'cu', 'rid=', 'tit') | Probability: 0.000000 |
| gram: ('ki', 'cu', 'rid=', 'tit', 'le=') | Probability: 0.002036 |
| gram: ('cu', 'rid=', 'tit', 'le=', 'Cen') | Probability: 0.000001 |
| gram: ('rid=', 'tit', 'le=', 'Cen', 'giz') | Probability: 0.000009 |
| gram: ('tit', 'le=', 'Cen', 'giz', 'Han') | Probability: 0.000009 |
| gram: ('le=', 'Cen', 'giz', 'Han', '>') | Probability: 0.000009 |
| gram: ('Cen', 'giz', 'Han', '>', 'Cen') | Probability: 0.000001 |
| gram: ('giz', 'Han', '>', 'Cen', 'giz') | Probability: 0.000009 |
| gram: ('Han', '>', 'Cen', 'giz', 'Han') | Probability: 0.000002 |
| gram: ('>', 'Cen', 'giz', 'Han', 'Cen') | Probability: 0.000000 |
| gram: ('Cen', 'giz', 'Han', 'Cen', 'giz') | Probability: 0.000002 |
| gram: ('giz', 'Han', 'Cen', 'giz', 'Han') | Probability: 0.000001 |
| gram: ('Han', 'Cen', 'giz', 'Han', 'Ceng') | Probability: 0.000003 |
| gram: ('Cen', 'giz', 'Han', 'Ceng', 'his') | Probability: 0.014834 |
| gram: ('giz', 'Han', 'Ceng', 'his', ',') | Probability: 0.000000 |
| gram: ('Han', 'Ceng', 'his', ', ', 'Khan') | Probability: 0.000000 |
| gram: ('Ceng', 'his', ', ', 'Khan', 'Cing') | Probability: 0.000000 |
| gram: ('his', ', ', 'Khan', 'Cing', 'gis') | Probability: 0.000000 |
| gram: (' ', 'Khan', 'Cing', 'gis', 'Ha') | Probability: 0.000000 |
| gram: ('Khan', 'Cing', 'gis', 'Ha', 'an') | Probability: 0.000440 |

5-Gram Table For Test Data

| | |
|--|-----------------------|
| gram: ('An', 'tik', 'ya', 'zar', 'lar') | Probability: 0.000000 |
| gram: ('tik', 'ya', 'zar', 'lar', 'ya') | Probability: 0.000000 |
| gram: ('ya', 'zar', 'lar', 'ya', 'p1') | Probability: 0.000000 |
| gram: ('zar', 'lar', 'ya', 'p1', 'n1n') | Probability: 0.000000 |
| gram: ('lar', 'ya', 'p1', 'n1n', 'mi') | Probability: 0.000000 |
| gram: ('ya', 'p1', 'n1n', 'mi', 'ma') | Probability: 0.000000 |
| gram: ('p1', 'n1n', 'mi', 'ma', 'r1') | Probability: 0.000000 |
| gram: ('n1n', 'mi', 'ma', 'r1', 'n1n') | Probability: 0.000000 |
| gram: ('mi', 'ma', 'r1', 'n1n', 'ol') | Probability: 0.000000 |
| gram: ('ma', 'r1', 'n1n', 'ol', 'du') | Probability: 0.000000 |
| gram: ('r1', 'n1n', 'ol', 'du', 'gu') | Probability: 0.000000 |
| gram: ('n1n', 'ol', 'du', 'gu', 'nu') | Probability: 0.000000 |
| gram: ('ol', 'du', 'gu', 'nu', '.') | Probability: 0.000000 |
| gram: ('du', 'gu', 'nu', '.', 'kay') | Probability: 0.000000 |
| gram: ('gu', 'nu', '.', 'kay', 'det') | Probability: 0.000000 |
| gram: ('nu', '.', 'kay', 'det', 'mek') | Probability: 0.000000 |
| gram: ('.', 'kay', 'det', 'mek', 'te') | Probability: 0.000000 |
| gram: ('kay', 'det', 'mek', 'te', 'dir') | Probability: 0.000000 |
| gram: ('det', 'mek', 'te', 'dir', 'Ay') | Probability: 0.000000 |
| gram: ('mek', 'te', 'dir', 'Ay', 'r1') | Probability: 0.000000 |
| gram: ('te', 'dir', 'Ay', 'r1', 'ca') | Probability: 0.000000 |
| gram: ('dir', 'Ay', 'r1', 'ca', 'un') | Probability: 0.000000 |
| gram: ('Ay', 'r1', 'ca', 'un', 'a') | Probability: 0.000000 |
| gram: ('r1', 'ca', 'un', 'a', 'd1') | Probability: 0.000000 |
| gram: ('ca', 'un', 'a', 'd1', 'da') | Probability: 0.000000 |
| gram: ('un', 'a', 'd1', 'da', '.') | Probability: 0.000000 |
| gram: ('a', 'd1', 'da', '.', 'geç') | Probability: 0.000000 |
| gram: ('d1', 'da', '.', 'geç', 'mek') | Probability: 0.000000 |
| gram: ('da', '.', 'geç', 'mek', 'te') | Probability: 0.000000 |
| gram: ('.', 'geç', 'mek', 'te', 'dir') | Probability: 0.000000 |
| gram: ('geç', 'mek', 'te', 'dir', ',') | Probability: 0.000000 |
| gram: ('mek', 'te', 'dir', ', ', 'Vit') | Probability: 0.000000 |

Created paragraphs:

1-Gram:

|(' <do',) 're' 'iç' 'ra' 'miş' 'sak' 'ki' 'l1' 'ris' 'Ke' 'tir'

2-Gram:

1 ('Han', ('Ceng')) 'his' '' 'ol' 'mak' 'i' 'şa' 'yan' ""ın" 'ka'

3-Gram:

1 ('Han', ('Ceng', 'his')) ' ' ' ' 'ol' 'du' 'ğ u' 'nu' ' ' 'et' 'miş' 'te' 'ka' 'nıt' 'la' 'r1' ' ' 'dir' 'ye' 'sü' 'gey'

4-Gram:

('di', '?', 'org', '/wi') 'ki' 'cu' 'rid=' 'tit' 'le="" 'Cen' 'giz' 'Han' 'ın' 'şe' 'ce' 're' 'kar' 'ma' 'şa' 'i' 'çin' 'de' 'si' ' ' 'ar' 'dın' 'dan' ' ' 'Gür' 'cü' 'ler' 'Deşt' '-i'

5-Gram:

('s1', 'zın', 'say', 'g1', |gös') 'te' 'ril' 'me' 'si' 'kim' 'li' 'ği' 'ça' 'l1ş' 't1' 'Mo' 'ğol' 'Her' 'kü' 'lü"" 'un' 'va' 'n1' 'n1' ' ' 'e' 'dil' 'di' 'Cen' 'giz' 'Han' 'Han' 'so' 'yu' 'il' 'gi' 'len' 'dir' 'me' 'ye' ' ' 'ça' 'ğır' 'm1ş' 've' 'Mo'

When the number of n is increasing the sentence is become more understandable and meaningfull.

Perplexities:

1-Gram

Train Data

```
| <do' 'id= 0
'url' 0
'w1' 'ki' 'pe' 'dia' 'di' '?' 'org' '/wi' 'ki' 'cu' 'rid=' 'tit' 'le="" 'Cen' 'giz' 'Han' "> 'Cen' 'giz' 'Han' 'Cen' 'giz' 'Han' ' 'Ceng' 'his' ' 'Khan"" ""Çing' 'gis' 'Ha' 'an' 'ya' 'd
0
'ö' 'A' 'ğuş' 'tos' ' 'Mo' 'ğol' ' 'ko' 'mu' 'tan' 'hü' 'küm' 'dar' 've' 'Mo' 'ğol' 'ım' 'pa' 'ra' 'tor' 'lu' "ğ u" 'nun' 0
'ku' 'ru' 'cu' 'su' 'dur' 'Cen' 'giz' ' 'Han' 0
'Yüz' 'y1' 'lın' 'ba' 'şın' 'da' 'Or' 'ta' 'As' "ya" 'da' 'ki' 'tüm' 'gö' 'çe' 'be' 'boz' 'kırn' 'ka' 'vim' 'le' 'ri' 'ni' 'bir' 'leş' 'ti' 're' 'rek' 'bir' 'u' 'lus' 'ha' 'li' 'ne' 'ge' '
'top' 'la' 'd1' 'Dün' 'ya' 'ta' 'ri' 'hi' 'nin' 'en' 'bü' 'yük' 'as' 'ke' 'ri' 'de' 'ha' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'la' 'rak' 'ka' 'bul' 'e' 'di' 'len' 'Cen' 'giz' ' 'Han' 'hü' 'küm'
'kur' 'du' 'Boz' 'kır' 'ge' 'le' 'ne' 'ğın' 'den' 'ge' 'len' 'on' 'lu' 'teş' 'ki' 'la' 't1' 'kul' 'la' 'na' 'rak' 'Me' 'ri' 'tok' 'ra' 'tik' 'li' 'yâ' 'ka' 'ta' 'bağ' 'l1' 'bir' 'or' 'du'
'tu' 'tar' 'Se' 'fer' 'le' 'ri' ' 'so' 'nu' 'cun' 'da' 'pek' 'çok' 'şe' 'hir' 'tah' 'rip' 'ol' 'muş' 've' 'mil' 'yon' 'lar' 'ca' 'in' 'san' 'da' 'kat' 'le' 'dil' 'miş' 'ti' 'an' 'cak' ""Ce
ş' 've' 'Mo' 'ğol' 'ca' 'i' 'çin' 'Uy' 'gur' 'al' 'fa' 'be' 'si' 'ni' 'u' 'yar' 'la' 'ta' 'rak' 'bu' 'nu' 'ço' 'cuk' 'la' 'r1' 'na' 'da' 'öğ' 'ret' 'me' 'si' 'ni' 0
'is' 'te' 'miş' 'tir' 'y1' 'lın' 'da' 'ku' 'zey' "Çin"" 'de' 'ki' 'Tan' 'gut' 'lar' 'ü' 'ze' 'ri' 'ne' 'çık' 't1' 'ğ1' 'se' 'fer' 'es' 'na' 'sın' 'da' 'ra' 'hat' 'sız' 'la' 'na' 'rak' 'ha'
'kay' 'bet' 'ti' 'Kur' 'muş' 'ol' 'du' 'ğ u' 'im' 'pa' 'ra' 'tor' 'luk' 'gü' 'nü' 'müz' 'de' 'Rus' 'ya' 'ha' 'riç' 'tüm' 'ül' 'ke' 'ler' 'den' 'da' 'ha' 'ge' 'niş' 'top' 'rak' 'lar' 'ü' 'ze
'gel' 'di' 'Cen' 'giz' 'Han' 'ın' 'şe' 'ce' 're' 'si' 'ya' 'r1' 'mi' 'to' 'lo' 'jik' 'bir' 'şe' 'kil' 'de' 'sis' 'per' 'de' 'si' 0
'ar' 'ka' 'sın' 'da' 'dır' 'o' 'nun' 'a' 'ta' 'la' 'r1' 'nin' 've' 'ken' 'di' 'si' 'nin' 'do' 'ğuş' ' 'ef' 'sa' 'ne' 'si' 'Mo' 'ğol' 'm1' 'to' 'lo' 'ji' 'si' 'nin' 'ö' 'nem' 'li' 0
'bel' 'ge' 'le' 'rin' 'den' 'dir' 'Ya' 'z1' 'l1ş' 'ta' 'ri' 'hi' 'i' 'ti' 'ba' 'r1' 'i' 'le' 'Cen' 'giz' 'Han' 'ın' 'ya' 'şa' 'd1' 'ğ1' 'dö' 'ne' 'me' 'en' 'ya' 'k1n' 'o' 'la' 'n1' 'şa' '
'e' 'dil' 'miş' 'tir' 'Mo' 'ğol' 'la' 'rın' 'Giz' 'li' 'Ta' 'ri' 'hin' 'de' 'yer' 'a' 'lan' 'ef' 'sa' 'ne' 'ye' 'gö' 're' 'A' 'lan' 'go' 'ya' 'dul' 'kal' 'dık' 'tan' 'son' 'ra' 'ev' 'len'
'ol' 'muş' 'tur' 'Cen' 'giz' 'Han' 've' 'o' 'nun' 'men' 'sup' 'ol' 'du' 'ğ u' 'Bör' 'ç1' 'ğın' 'ler' 'bu' 'ne' 'ço' 'cuk' 'lar' 'dan' ""Bo' 'don' 'car"" 'ad' 'l1' 'en' 'kü' 'çük' 'o' 'la' 'n1'
'gel' 'mek' 'te' 'dir' 'A' 'lan' 'go' 'ya' 'ef' 'sa' 'ne' 'si' 'yal' 'nız' 'ca' 'Cen' 'giz' 'Han' 'ı1' 'de' 'ğil' 'o' 'nun' 'la' 'bir' 'lik' 'te' "Ni' 'run"" 'ya' 'ni' '1' 'ş1' 'ğın' 'ço'
'e' 'dil' 'miş' 'tir' 'Ni' 'run' 'boy' 'la' 'rın' 'dan' 'ön' 'ce' 'lik' 'le' 'Cen' 'giz' 'Han' 'ın' 'bo' 'yu' 'Bor' 'ç1' 'ğın' 'ler' 'ar' 'dın' 'dan' ' 'Tay' 'ciütl' 'ut' 'lar' ' 'Bar' 'l
'sa' 'y1' 'la' 'bi' 'lir' 'y1' 'lın' 'da' 'Mo' 'ğol' 'ka' 'bi' 'le' 'le' 'rin' 'den' 'Bör' 'ç1' 'ğın' 'le' 're' 'men' 'sup' ' 'Ka' 'bul' 'bü' 'tün' 'Mo' 'ğol' 'la' 'rın' 'ilk' 'li' 'de' 'r
'al' 'miş' 'tır' 'Cen' 'giz' 'Han' 'ın' 'ba' 'ba' 's1' 'Ye' 'sü' 'gey' 'Ba' 'ha' 'dır' 'o' 'nun' 0
'to' 'ru' 'nu' 'dur' 'Ka' 'bul' 'Han' 've' 'o' 'nun' 'ha' 'le' 'fi' 'Am' 'ba' 'kay' 'Han' 'za' 'ma' 'nın' 'da' ' 'Mo' 'ğol' 'lar' "Çin"" 'de' 'ki' 'jin' 'ım' 'pa' 'ra' 'tor' 'lu' 'ğ u' 'i'
'et' 'ti' 'ler' 'Am' 'ba' 'kay' 'hiç' 'a' 'l1' 'şıl' 'ma' 'dık' 'bir' ' 'şe' 'kil' 'de' ""tah' 'ta' 'e' 'şek' 'şek' 'l1"" 'de' 'nen' 'bir' 'du' 'ru' 'ma' 'so' 'ku' 'la' 'rak' 'çar' 'm1' 'h
'e' 'dil' 'di' 'Cen' 'giz' 'Han' 'ın' 'bü' 'yük' 'a' 'ma' 'ca' 's1' ' 'Ku' 'tua' 'me' 'tu' 'bu' 'ha' 'ka' 're' 'te' 'çin' 'ü' 'ze' 'ri' 'ne' 've' 'Ta' 'tar' 'la' 'ra' 'bir' 'di' 'zi' 'sal' 'd1
'ka' 'zan' 'd1' ' 'Fa' 'kat' 'y1' 'lın' 'da' 'de' 'tay' 'la' 'r1' 'bi' 'lin' 'me' 'yen' 'bir' 'di' 'zi' 'o' 'lay' ' 'so' 'nun' 'da' 'ku' 'zey' 'çin' 'ın' 'ha' 'ki' 'm1' 'jin' ' 'Ha' 'ne
'uğ' 'rat' 't1' 'Mo' 'ğol' 'lar' 'bir' 'sü' 're' 'kar' 'ma' 'şa' 'i' 'çin' 'de' 0
'da' 'ğıl' 'd1' 'lar' 'Se' 'fa' 'let' 'i' 'çin' 'de' 'yü' 'zen' 'bu' 'kar' 'ma' 'şık' 'hal' 'de' 'ki' 'Mo' 'ğol' 'lar' 'ın' 'i' 'çe' 'ri' 'sin' 'de' 'ki' 'ö' 'nem' 'siz' 'li' 'der' 'ler'
'ça' 'l1ş' 't1' 'Mo' 'ğol' 'la' 'rın' 'ba' 't1' 'kom' 'şu' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'lan' 'Türk' 'boy' 'la' 'rın' 'dan' 'Ke' 'rait' 'it' 'ler' 0
'i' 'di' 'Ke' 'rait' 'it' 'ler' 'yıl' 'dan' 'be' 'ri' 'Nas' 'tu' 'ri' 0
'H1' 'ris' 'ti' 'yan' ' 'd1' 'H1' 'ris' 'ti' 'yan' 'Ke' 'rait' 'it' 'le' 'rin' 'o' 'za' 'man' 'ki' 'li' 'de' 'ri' 'Tuğ' 'rul' 0
```

Test Data

'An' 'tik' 'ya' 'zar' 'lar' 'ya' 'pı' 'nın' 'mi' 'ma' 'rı' 'nın' 'ol' 'du' 'ğu' 'nu' 0
'kay' 'det' 'mek' 'te' 'dir' 'Ay' 'rı' 'ca' '""un' 'a' 'dı' 'da' 0
'geç' 'mek' 'te' 'dir' ' ' 'Vit' 'ru' 'vius' 'us' 'MÖ' 0
'yüz' 'yı' 'lın' 'en' 'ö' 'nem' 'li' 'dört' 'hey' 'kel' 'tı' 'ra' 'şı' 'nın' 'bu' 'ya' 'pı' 'da' 'ça' 'lıış' 'tı' 'ğı' 'nı' 0
'kay' 'det' 'miş' 'tır' 'Do' 'ğu' 'da' ' ' 'Sko' 'pas' 'ba' 'tı' 'da' ' ' 'Le' 'ok' 'ha' 'res' 'ku' 'zey' 'de' ' ' 'Bryak' 'sis' 'gü' 'ney' 'de' 'Ti' 'mot' 'heos' 'os' 0
'ça' 'lıış' 'miş' 'tır' ' ' 'Bryak' 'sis' 'Kar' 'ya' 'lı' 'bir' 0
'sa' 'nat' 'cı' 'dır' 'Di' 'ğer' 'sa' 'nat' 'cı' 'lar' 'Yu' 'na' 'nis' 'tan' ' ' 'dan' 0
'ge' 'ti' 'ril' 'miş' 'tır' 'Dört' 'at' 'lı' 'a' 'ra' 'ba' 'yı' 'Mi' 'mar' 'Pythe' 'os' '""un' 'yap' 'tı' 'ğı' 0
'söy' 'len' 'mek' 'te' 'dir'

2-Gram

Train Data

'<do' 'id=' 'url' 0.0012376237623762376
'url' 'wi' 0.0024752475247524753
'wi' 'ki' 'pe' 'dia' 'di' '?' 'org' /'wi' 'ki' 'cu' 'rid=' 'tit' 'le=' ' ' 'Cen' 'giz' 'Han' ' ' '>' 'Cen' 'giz' 'Han' 'Cen' 'giz' 'Han' ' ' 'Ceng' 'his' ' ' 'Khan' ' ' ' 'Cing' 'gis' 'Ha' 'an' ' 'ya' 'd
'ö' 0.11111111111111111
'ö' 'A' 'ğus' 'tos' ' ' 'Mo' 'ğöl' ' ' 'ko' 'mu' 'tan' 'hü' 'küm' 'dar' 've' 'Mo' 'ğöl' 'İM' 'pa' 'ra' 'tor' 'lu' " 'ğu" 'nun' 'ku' 1.4818834921742642e-41
'ku' 'ru' 'cu' 'su' 'dur' 'Cen' 'giz' ' ' 'Han' 'Yüz' 2.7834374118267235e-19
'Yüz' 'yı' 'lın' 'ba' 'şın' 'da' 'Or' 'ta' 'As' 'ya' " 'da' 'ki' 'tüm' 'gö' 'ce' 'be' 'boz' 'kır' 'ka' 'vim' 'le' 'ri' 'ni' 'bir' 'leş' 'ti' 're' 'rek' 'bir' 'u' 'lus' 'ha' 'li' 'ne' 'ge' ' ' 'top' 'la' 'dı' 'Dün' 'ya' 'ta' 'rı' 'hi' 'nin' 'en' 'bü' 'yük' 'as' 'ke' 'ri' 'de' 'ha' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'la' 'rak' 'ka' 'bul' 'e' 'di' 'len' 'Cen' 'giz' ' 'Han' 'hü' 'küm' 'kur' 'du' 'Boz' 'kır' 'ge' 'le' 'ne' 'ğın' 'den' 'ge' 'len' 'on' 'lu' 'teş' 'ki' 'la' 'tı' 'kul' 'la' 'na' 'rak' 'Me' 'ri' 'tok' 'ra' 'tik' ' ' 'li' 'yâ' 'ka' 'ta' 'bağ' 'lı' 'bir' 'or' 'du' 'tu' 'tar' 'Se' 'fer' 'le' 'ri' ' 'so' 'nu' 'cun' 'da' 'pek' 'çok' 'şe' 'hir' 'tah' 'rip' 'ol' 'muş' 've' 'mil' 'yon' 'lar' 'ca' 'ın' 'san' 'da' 'kat' 'le' 'dil' 'miş' 'ti' 'an' 'cak' '""Ce
s' 've' 'Mo' 'ğöl' 'ca' 'i' 'çin' 'Uy' 'gur' 'al' 'fa' 'be' 'si' 'ni' 'u' 'yar' 'la' 'ta' 'rak' 'bu' 'nu' 'ço' 'cuk' 'la' 'rı' 'na' 'da' 'öğ' 'ret' 'me' 'si' 'ni' 'is' 5.639957517959265e-30
'is' 'te' 'miş' 'tır' 'yı' 'lın' 'da' 'ku' 'zey' 'Cin' ' 'de' 'ki' 'Tan' 'gut' 'lar' 'ü' 'ze' 'ri' 'ne' 'çık' 'tı' 'ğı' 'se' 'fer' 'es' 'na' 'sın' 'da' 'ra' 'hat' 'sız' 'la' 'na' 'rak' 'ha' 'kay' 'bet' 'ti' 'Kur' 'muş' 'ol' 'du' 'ğu' 'im' 'pa' 'ra' 'tor' 'luk' 'gü' 'nü' 'müz' 'de' 'Rus' 'ya' 'ha' 'riç' 'tüm' 'ül' 'ke' 'ler' 'den' 'da' 'ha' 'ge' 'niş' 'top' 'rak' 'lar' 'ü' 'ze
'gel' 'di' 'Cen' 'giz' 'Han' ' 'ın' 'şe' 'ce' 're' 'si' 'ya' 'rı' 'mi' 'to' 'lo' 'jik' 'bir' 'şe' 'kil' 'de' 'sis' 'per' 'de' 'si' 'ar' 1.906180761372464e-30
'ar' 'ka' 'sın' 'da' 'dır' 'O' 'nun' 'a' 'ta' 'la' 'rı' 'nın' 've' 'ken' 'di' 'si' 'nin' 'do' 'ğuş' ' 'ef' 'sa' 'ne' 'si' 'Mo' 'ğöl' 'mi' 'to' 'lo' 'ji' 'si' 'nin' 'ö' 'nem' 'li' 'bel' 4.2
'bel' 'ge' 'le' 'rin' 'den' 'dir' 'Ya' 'zı' 'lıış' 'ta' 'rı' 'hi' 'i' 'ti' 'ba' 'rı' 'i' 'le' 'Cen' 'giz' 'Han' ' 'ın' 'ya' 'şa' 'dı' 'ğı' 'dö' 'ne' 'me' 'en' 'ya' 'kın' 'o' 'la' 'nı' 'Sa' ' 'e' 'dil' 'miş' 'tır' 'Mo' 'ğöl' 'la' 'rın' 'Giz' 'li' 'Ta' 'rı' 'hin' 'de' 'yer' 'a' 'lan' 'ef' 'sa' 'ne' 'ye' 'gö' 're' 'A' 'lan' 'go' 'ya' 'dul' 'kal' 'dik' 'tan' 'son' 'ra' 'ev' 'len' 'ol' 'muş' 'tur' 'Cen' 'giz' 'Han' 've' 'o' 'nun' 'men' 'sup' 'ol' 'du' 'ğu' 'Bör' 'çi' 'gin' 'ler' 'bu' 'co' 'cuk' 'lar' 'dan' '""Bo' 'don' 'car' ""ad' 'lı' 'en' 'kü' 'çük' 'o' 'la' 'nı' ' 'gel' 'mek' 'te' 'dir' 'A' 'lan' 'go' 'ya' 'ef' 'sa' 'ne' 'si' 'yal' 'nız' 'ca' 'Cen' 'giz' 'Han' 'ı' 'de' 'ğil' 'o' 'nun' 'la' 'bir' 'lik' 'te' '""Ni' 'run' ""ya' 'ni' 'ı' 'şı' 'ğın' 'co' 'e' 'dil' 'miş' 'tır' 'Ni' 'run' 'boy' 'la' 'rın' 'dan' 'ön' 'ce' 'lik' 'le' 'Cen' 'giz' 'Han' ' 'ın' 'bo' 'yu' 'Bor' 'ci' 'gin' 'ler' 'ar' 'dın' 'dan' ' ' 'Tay' 'ciutl' 'ut' 'lar' ' 'Bar' 'l
'sa' 'yı' 'la' 'bi' 'lir' 'yı' 'lın' 'da' 'Mo' 'ğöl' 'ka' 'bi' 'le' 'le' 'rin' 'den' 'Bör' 'çi' 'gin' 'le' 're' 'men' 'sup' 'Ka' 'bul' 'bü' 'tün' 'Mo' 'ğöl' 'la' 'rın' 'ilk' 'li' 'de' 'r
'al' 'miş' 'tır' 'Cen' 'giz' 'Han' ' 'ın' 'ba' 'ba' 'sı' 'Ye' 'sü' 'gey' 'Ba' 'ha' 'dır' 'o' 'nun' 'to' 1.9153848029038692e-18
'to' 'ru' 'nu' 'dur' 'Ka' 'bul' 'Han' 've' 'o' 'nun' 'ha' 'le' 'fi' 'Am' 'ba' 'kay' 'Han' 'za' 'ma' 'nın' 'da' ' ' 'Mo' 'ğöl' 'lar' 'Cin' ' 'de' 'ki' 'Jin' 'İM' 'pa' 'ra' 'tor' 'lu' 'ğu' 'i' 'et' 'ti' 'ler' 'Am' 'ba' 'kay' 'hiç' 'a' 'lı' 'şıl' 'ma' 'dik' 'bir' ' 'şe' 'kil' 'de' '""tah' 'ta' 'e' 'şek' 'şek' 'lı' ""de' 'nen' 'bir' 'du' 'ru' 'ma' 'so' 'ku' 'la' 'rak' 'çar' 'mı' 'h
'e' 'dil' 'di' 'Cen' 'giz' 'Han' ' 'ın' 'bü' 'yük' 'a' 'ma' 'ca' 'sı' ' 'ku' 'tua' 'tu' 'bu' 'ha' 'ka' 're' 'te' 'Cin' 'ü' 'ze' 'ri' 'ne' 've' 'Ta' 'tar' 'la' 'ra' 'bir' 'di' 'zi' 'sal' 'd
'ka' 'zan' 'dı' ' 'Fa' 'kat' ' 'yı' 'lın' 'da' 'de' 'tay' 'la' 'rı' 'bi' 'lin' 'me' 'yen' 'bir' 'di' 'zi' 'o' 'lay' ' 'so' 'nun' 'da' 'Ku' 'zey' 'Cin' ' 'in' 'ha' 'ki' 'mi' 'Jin' ' 'Ha' 'ne
'uğ' 'rat' 'tı' 'Mo' 'ğöl' 'lar' 'bir' 'sü' 're' 'kar' 'ma' 'sa' 'ı' 'çin' 'de' 'da' 7.593183304638943e-24
'da' 'ğıl' 'dı' 'lar' 'Se' 'fa' 'let' 'i' 'çin' 'de' 'yü' 'zen' 'bu' 'kar' 'ma' 'şık' 'hal' 'de' 'ki' 'Mo' 'ğöl' 'lar' ' 'ın' 'i' 'çe' 'ri' 'sin' 'de' 'ki' 'ö' 'nem' 'siz' 'li' 'den' 'ler' 'ca' 'lıış' 'tı' 'Mo' 'ğöl' 'la' 'rın' 'ba' 'tı' 'kom' 'şu' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'lan' 'Türk' 'boy' 'la' 'rın' 'dan' 'Ke' 'rait' 'it' 'ler' 'i' 2.9258186196875556e-31
'i' 'di' 'Ke' 'rait' 'it' 'ler' 'yıl' 'dan' 'be' 'ri' 'Nas' 'tu' 'ri' 'Hı' 1.7006864410097422e-17
'Hı' 'ris' 'tı' 'yan' 'dı' 'Hı' 'ris' 'ti' 'yan' 'Ke' 'rait' 'it' 'le' 'rin' 'o' 'za' 'man' 'ki' 'li' 'de' 'ri' 'Tuğ' 'rul' 'i' 1.429502453347625e-28
'i' 'di' 'Tuğ' 'rul' 'yı' 'lan' 'da' 'iç' 'me' 'se' 'le' 'ler' 'so' 'nu' 'cu' 'tah' 'tı' 'nı' 'kay' 2.268347973376551e-29
'kay' 'bet' 'miş' 'ti' 'Mo' 'ğöl' 'la' 'rın' 'li' 'de' 'ri' ' 'Ye' 'sü' 'gey' 'Tuğ' 'ru' 'la' 'ka' 'bi' 'le' 'si' 'nin' 'ön' 'der' 'li' 'ğı' 'ni' 'ye' 'ni' 'den' 'e' 'le' 'ge' 'çin' 'me' ' 'et' 'ti' 'Tuğ' 'rul' 've' 'Ye' 'sü' 'gey' 'an' 'da' 'i' 'le' 'kar' 'de' 'şi' 'lik' 'ye' 'mi' 'ni' 'et' 'tik' 'ten' ' 'son' 'ra' 'da' 'ha' 'son' 'ra' 'la' 'rı' 'Mo' 'ğöl' 'la' 'rın' 'ye' ' 'kur' 'du' 'lar' 'Ri' 'va' 'yet' 'le' 're' 've' 'ef' 'sa' 'ne' 'ye' 'gö' 're' 'Ye' 'sü' 'gey' 'Ba' 'ha' 'dır' 'bir' 'gün' 'O' 'non' 'neh' 'ri' 'kı' 'yı' 'la' 'rın' 'da' 'şa' 'hi' 'ni' 'i' 'rast' 'la' 'dı' ' 'Ye' 'sü' 'gey' 'a' 'ra' 'ban' 'da' 'o' 'tu' 'ran' 'kı' 'za' 'ilk' 'gö' 'rüş' 'te' 'a' 'şık' 'ol' 2.7577105399266325e-35
'ol' 'muş' 'tu' 'Ye' 'sü' 'gey' 'i' 'ki' 'kar' 'de' 'şi' 'ni' 'ya' 'nı' 'na' 'a' 'la' 'rak' 've' 'bir' 'lik' 'te' 'a' 'ğır' 'a' 'ğır' 'gı' 'den' 'dü' 'ğün' 'a' 'ra' 'ba' 'sı' 'na' 'ye' 9.1
'ye' 'tiş' 'ti' 'ler' 'Üç' 'kar' 'deş' 'Ong' 'gi' 'rat' 'bo' 'yu' 'nun' 'Ol' 'ku' 'nat' 'ka' 'bi' 'le' 'sin' 'den' 'o' 'lan' 'Hö' 'e' 'lin' 'a' 'dın' 'da' 'ki' 'ye' 'ni' 'ev' 'li' 'ge' 'li
'ya' 'ka' 'la' 'dı' 'lar' 'Hö' 'e' 'lin' 'baş' 'ke' 'seç' 'me' 'şan' 'sı' 'ol' 'ma' 'dı' 'ğı' 'i' 'çin' 'Ye' 'sü' 'gey' 'Ba' 'ha' 'dır' 'ı' 'ye' 'ni' 'ko' 'ca' 'sı' 'o' 'la' 'rak' 'ka' 1.
'ka' 'bul' 'len' 'di' 'Ye' 'sü' 'gey' ' 'o' 'nun' 'la' 'do' 'ğur' 'du' 'ğu' 'oğ' 'lan' 'bir' 'kah' 'ra' 'man' 'o' 'la' 'cak' 'di' 'ye' 'rek' 'ev' 5.1767303949298744e-42

Test Data

'An' 'tik' 'ya' 'zar' 'lar' 'ya' 'pı' 'nın' 'mi' 'ma' 'rı' 'nın' 'ol' 'du' 'ğu' 'nu' 'kay' 3.4131655449097456e-48
'kay' 'det' 'mek' 'te' 'dir' 'Ay' 'rı' 'ca' '""un' 'a' 'dı' 'da' 'geç' 7.584983601684808e-36
'geç' 'mek' 'te' 'dir' ' ' 'Vit' 'ru' 'vius' 'us' 'MÖ' 'yüz' 1.8254431158509846e-38
'yüz' 'yı' 'lın' 'en' 'ö' 'nem' 'li' 'dört' 'hey' 'kel' 'tı' 'ra' 'şı' 'nin' 'bu' 'ya' 'pı' 'da' 'ça' 'lıış' 'tı' 'ğı' 'nı' 'kay' 4.846019473434205e-63
'kay' 'det' 'miş' 'tır' 'Do' 'ğu' 'da' ' ' 'Sko' 'pas' 'ba' 'tı' 'da' ' ' 'Le' 'ok' 'ha' 'res' 'ku' 'zey' 'de' ' ' 'Bryak' 'sis' 'gü' 'ney' 'de' 'Ti' 'mot' 'heos' 'os' 'ça' 4.368561785965644e-11
'ça' 'lıış' 'miş' 'tır' ' ' 'Bryak' 'sis' 'Kar' 'ya' 'lı' 'bir' 'sa' 8.705069348824607e-38
'sa' 'nat' 'cı' 'dır' 'Di' 'ğer' 'sa' 'nat' 'cı' 'lar' 'Yu' 'na' 'nis' 'tan' ' 'dan' 'ge' 3.0018389024984876e-58
'ge' 'ti' 'ril' 'miş' 'tır' 'Dört' 'at' 'lı' 'a' 'ra' 'ba' 'yı' 'Mi' 'mar' 'Pythe' 'os' '""un' 'yap' 'tı' 'ğı' 'söy' 3.4017741042112624e-59
'söy' 'len' 'mek' 'te' 'dir' 'Kar' 'ya' 'sat' 'ra' 'bı' ' 'Ma' 'u' 'so' 'los' 'ken' 'di' 'yö' 'ne' 'ti' 'mi' 'za' 'ma' 'nın' 'da' 'muh' 'te' 'me' 'len' 'MÖ' '35' 'ya' 'pı' 'ya' 'baş' 6.907
'baş' 'la' 'miş' 'tır' 'O' 'nun' 'ö' 'lü' 'mün' 'den' 'son' 'ra' ' 'M' ' 'ka' 'rı' 'sı' 'ay' 'nı' 'za' 'man' 'da' 'kız' 'kar' 'de' 'şi' 'Ar' 'te' 'meisia' 'i' 'sia' 'si' 'a' 'nı' 'tın' 'ya' 'et' 'miş' 'ler' 'dir' ' 'Muh' 'te' 'me' 'len' 'in' 'şaat' 'at' 'MÖ' '34' 'Pik' 'so' 'da' 'ros' ' 'la' 'A' 'da' 'a' 'ra' 'sın' 'da' 'ki' 'sat' 'rap' 'lık' 'mü' 'ca' 'de' 'le' 'si' 'sı' 'ra'

Test Data

| 'An' 'tik' 'ya' 'zar' 'lar' 'ya' 'pı' 'nın' 'mi' 'ma' 'rı' 'nın' 'ol' 'du' 'ğu' 'nu' 'kay' 3.4118499891e-312
'kay' 'det' 'mek' 'te' 'dir' 'Ay' 'rı' 'ca' '""un" 'a' 'dı' 'da' 'geç' 4.447855269787033e-228
'geç' 'mek' 'te' 'dir' ' ' 'Vit' 'ru' 'vius' 'us' 'Mö' 'yüz' 3.208200249355322e-229
'yüz' 'yı' 'lın' 'en' 'ö' 'nem' 'li' 'dört' 'hey' 'kel' 'tı' 'ra' 'şı' 'nın' 'bu' 'ya' 'pı' 'da' 'ça' 'lış' 'tı' 'ğı' 'nı' 'kay' 0.0
'kay' 'det' 'miş' 'tir' 'Do' 'ğu' 'da' ' ' 'Sko' 'pas' 'ba' 'tı' 'da' ' ' 'Le' 'ok' 'ha' 'res' 'ku' 'zey' 'de' ' ' 'Bryak' 'sis' 'gü' 'ney' 'de' 'Ti' 'mot' 'heos' 'os' 'ça' 0.0
'ça' 'lış' 'mış' 'tır' ' ' 'Bryak' 'sis' 'Kar' 'ya' 'lı' 'bir' 'sa' 3.5754405905582624e-236
'sa' 'nat' 'çı' 'dır' 'Di' 'ğer' 'sa' 'nat' 'çı' 'lar' 'Yu' 'na' 'nis' 'tan' ' ' 'dan' 'ge' 0.0
'ge' 'ti' 'ril' 'miş' 'tir' 'Dört' 'at' 'lı' 'a' 'ra' 'ba' 'yı' 'Mi' 'mar' 'Pythe' 'os' '""un" 'yap' 'tı' 'ğı' 'söy' 0.0

5-Gram

Train Data

| '<do' 'id=' 'url' 2.8786920572363746e-36
'url' 'wi' 3.681383851178938e-15
'wi' 'ki' 'pe' 'dia' 'di' '?' 'org' '/wi' 'ki' 'cu' 'rid=' 'tit' 'le=' ' ' 'Cen' 'giz' 'Han' ' ' '>' 'Cen' 'giz' 'Han' 'Cen' 'giz' 'Han' ' ' 'Ceng' 'his' ' ' 'Khan' ' ' ' 'Cing' 'gis' 'Ha' 'an' ' ' 'ya' 'd
'ö' 2.504354108454811e-06
'ö' 'A' 'ğus' 'tos' ' ' 'Mo' 'ğol' ' ' 'ko' 'mu' 'tan' 'hü' 'kü'm' 'dar' 've' 'Mo' 'ğol' 'İM' 'pa' 'ra' 'tor' 'lu' "ğu" ' ' 'nun' 'ku' 7.638025081501351e-302
'ku' 'ru' 'cu' 'su' 'dur' 'Cen' 'giz' ' ' 'Han' 'Yüz' 3.637346515762314e-139
'Yüz' 'yı' 'lın' 'ba' 'şın' 'da' 'or' 'ta' 'As' 'ya' ' ' 'da' 'ki' 'tüm' 'gö' 'çe' 'be' 'boz' 'kır' 'ka' 'vim' 'le' 'ri' 'ni' 'bir' 'leş' 'ti' 're' 'rek' 'bir' 'u' 'lus' 'ha' 'li' 'ne' 'ge' ' ' 'top' 'la' 'dı' 'Dün' 'ya' 'ta' 'rı' 'hi' 'nin' 'en' 'bü' 'yük' 'as' 'ke' 'ri' 'de' 'ha' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'la' 'rak' 'ka' 'bul' 'e' 'di' 'len' 'Cen' 'giz' ' ' 'Han' 'hü' 'kü'm' ' ' 'kur' 'du' 'Boz' 'kır' 'ge' 'le' 'ne' 'ğın' 'den' 'ge' 'len' 'on' 'lu' 'teş' 'ki' 'la' 'tı' 'kul' 'la' 'na' 'rak' 'Me' 'ri' 'tok' 'ra' 'tik' ' ' 'li' 'yâ' 'ka' 'ta' 'bag' 'lı' 'bir' 'or' 'du' ' ' 'tu' 'tar' 'Se' 'fer' 'le' 'ri' ' ' 'so' 'nu' 'cun' 'da' 'pek' 'çok' 'şe' 'hir' 'tah' 'rip' 'ol' 'muş' 've' 'mil' 'yon' 'lar' 'ca' 'in' 'san' 'da' 'kat' 'le' 'dil' 'miş' 'ti' 'an' 'cak' ' ' 'Ce
ş' 've' 'Mo' 'ğol' 'ca' 'i' 'çin' 'Uy' 'gur' 'al' 'fa' 'be' 'si' 'ni' 'u' 'yar' 'la' 'ta' 'rak' 'bu' 'nu' 'ço' 'cuk' 'la' 'rı' 'na' 'da' 'ög' 'ret' 'me' 'si' 'ni' 'is' 0.0
'is' 'te' 'miş' 'tir' 'yı' 'lın' 'da' 'ku' 'zey' "Cin" ' ' 'de' 'ki' 'Tan' 'gut' 'lar' 'ü' 'ze' 'ri' 'ne' 'çık' 'tı' 'ğı' 'se' 'fer' 'es' 'na' 'sın' 'da' 'ra' 'hat' 'sız' 'la' 'na' 'rak' 'ha' ' ' 'kay' 'bet' 'ti' 'Kur' 'muş' 'ol' 'du' 'ğu' 'im' 'pa' 'ra' 'tor' 'luk' 'gü' 'nü' 'müz' 'de' 'Rus' 'ya' 'ha' 'riç' 'tüm' 'ül' 'ke' 'ler' 'den' 'da' 'ha' 'ge' 'niş' 'top' 'rak' 'lar' 'ü' 'ze
'gel' 'di' 'Cen' 'giz' 'Han' ' ' 'ın' 'şe' 'ce' 're' 'si' 'ya' 'rı' 'mi' 'to' 'lo' 'jik' 'bir' 'şe' 'kil' 'de' 'sis' 'per' 'de' 'si' 'ar' 8.486573644772507e-251
'ar' 'ka' 'sın' 'da' 'dır' 'O' 'nun' 'a' 'ta' 'rı' 'nın' 've' 'ken' 'di' 'si' 'nin' 'do' 'ğuş' ' ' 'ef' 'sa' 'ne' 'si' 'Mo' 'ğol' 'mi' 'to' 'lo' 'ji' 'si' 'nin' 'ö' 'nem' 'li' 'bel' 0.0
'bel' 'ge' 'le' 'rin' 'den' 'dir' 'Ya' 'zı' 'lış' 'ta' 'rı' 'hi' 'i' 'ti' 'ba' 'rı' 'i' 'le' 'Cen' 'giz' 'Han' ' ' 'ın' 'ya' 'şa' 'dı' 'ğı' 'dö' 'ne' 'me' 'en' 'ya' 'kın' 'o' 'la' 'nı' 'sa' ' ' 'e' 'dil' 'miş' 'tir' 'Mo' 'ğol' 'la' 'rın' 'Giz' 'li' 'Ta' 'rı' 'hin' 'de' 'yer' 'a' 'lan' 'ef' 'sa' 'ne' 'ye' 'gö' 're' 'A' 'lan' 'go' 'ya' 'dul' 'kal' 'dik' 'tan' 'son' 'ra' 'ev' 'len' ' ' 'ol' 'muş' 'tur' 'Cen' 'giz' 'Han' 've' 'o' 'nun' 'men' 'sup' 'ol' 'du' 'ğu' 'Bör' 'çi' 'ğın' 'ler' 'bu' 'ço' 'cuk' 'lar' 'dan' ' ' 'Bo' 'don' 'car' ' ' 'ad' 'lı' 'en' 'kü' 'çük' 'o' 'la' 'nı' ' ' 'gel' 'mek' 'te' 'dir' 'A' 'lan' 'go' 'ya' 'ef' 'sa' 'ne' 'si' 'yal' 'nız' 'ca' 'Cen' 'giz' 'Han' ' ' 'ı' 'de' 'ğil' 'o' 'nun' 'la' 'bir' 'lik' 'te' ' ' 'Ni' 'run' ' ' 'ya' 'ni' 'ı' 'şı' 'ğın' 'ço' ' ' 'e' 'dil' 'miş' 'tir' 'Ni' 'run' 'boy' 'la' 'rın' 'dan' 'ön' 'ce' 'lik' 'le' 'Cen' 'giz' 'Han' ' ' 'ın' 'bo' 'yu' 'Bor' 'ci' 'ğın' 'ler' 'ar' 'dın' 'dan' ' ' 'Tay' 'ciutl' 'ut' 'lar' ' ' 'Bar' 'l
'sa' 'yı' 'la' 'bi' 'lir' 'yı' 'lın' 'da' 'Mo' 'ğol' 'ka' 'bi' 'le' 'le' 'rın' 'den' 'Bör' 'çi' 'ğın' 'le' 're' 'men' 'sup' ' ' 'Ka' 'bul' 'bü' 'tün' 'Mo' 'ğol' 'la' 'rın' 'ilk' 'li' 'de' 'r
'al' 'miş' 'tır' 'Cen' 'giz' 'Han' ' ' 'ın' 'ba' 'ba' 'sı' 'Ye' 'sü' 'gey' 'Ba' 'ha' 'dır' 'o' 'nun' 'to' 4.3446588739087573e-162
'to' 'ru' 'nu' 'dur' 'Ka' 'bul' 'Han' 've' 'o' 'nun' 'ha' 'le' 'fi' 'Am' 'ba' 'kay' 'Han' 'za' 'ma' 'nın' 'da' ' ' 'Mo' 'ğol' 'lar' "Cin" ' ' 'de' 'ki' 'Jin' 'İM' 'pa' 'ra' 'tor' 'lu' 'ğu' 'i' ' ' 'et' 'ti' 'ler' 'Am' 'ba' 'kay' 'hiç' 'a' 'lı' 'sıl' 'ma' 'dik' 'bir' ' ' 'şe' 'kil' 'de' ' ' 'tah' 'ta' 'e' 'şek' 'şek' 'li' ' ' 'de' 'nen' 'bir' 'du' 'ru' 'ma' 'so' 'ku' 'la' 'rak' 'çar' 'mı' 'h
'e' 'dil' 'di' 'Cen' 'giz' 'Han' ' ' 'ın' 'bü' 'yük' 'a' 'ma' 'ca' 'sı' ' ' 'Ku' 'tua' 'tu' 'bu' 'ha' 'ka' 're' 'te' 'Cin' 'ü' 'ze' 'ri' 'ne' 've' 'Ta' 'tar' 'la' 'ra' 'bir' 'di' 'zi' 'sal' 'dı
'ka' 'zan' 'dı' ' ' 'Fa' 'kat' ' ' 'yı' 'lın' 'da' 'de' 'tay' 'la' 'rı' 'bi' 'lin' 'me' 'yen' 'bir' 'di' 'zi' 'o' 'lay' ' ' 'so' 'nun' 'da' 'ku' 'zey' 'Cin' ' ' 'in' 'ha' 'ki' 'mi' 'Jin' ' ' 'Ha' 'ne
'uğ' 'rat' 'tı' 'Mo' 'ğol' 'lar' 'bir' 'sü' 're' 'kar' 'ma' 'şa' 'i' 'çin' 'de' 'da' 3.018071061541877e-192
'da' 'ğıl' 'dı' 'lar' 'Se' 'fa' 'let' 'i' 'çin' 'de' 'yü' 'zen' 'bu' 'kar' 'ma' 'şık' 'hal' 'de' 'ki' 'Mo' 'ğol' 'lar' ' ' 'ın' 'i' 'çe' 'ri' 'sin' 'de' 'ki' 'ö' 'nem' 'siz' 'li' 'der' 'ler'
'ça' 'lış' 'tı' 'Mo' 'ğol' 'la' 'rın' 'ba' 'tı' 'kom' 'şu' 'la' 'rın' 'dan' 'bi' 'ri' 'o' 'lan' 'Türk' 'boy' 'la' 'rın' 'dan' 'Ke' 'rait' 'it' 'ler' 'i' 5.529920094539071e-244
'i' 'di' 'Ke' 'rait' 'it' 'ler' 'yı' 'dan' 'be' 'ri' 'Was' 'tu' 'ri' 'Hı' 8.659724029203064e-141
'Hı' 'ris' 'ti' 'yan' 'dı' 'Hı' 'ris' 'ti' 'yan' 'Ke' 'rait' 'it' 'le' 'rin' 'o' 'za' 'man' 'ki' 'li' 'de' 'ri' 'Tuğ' 'rul' 'i' 1.960357303158267e-264
'i' 'di' ' ' 'Tuğ' 'rul' 'yı' 'lın' 'da' 'iç' 'me' 'se' 'le' 'ler' 'so' 'nu' 'cu' 'tah' 'tı' 'nı' 'kay' 6.348542629061118e-222
'kay' 'bet' 'miş' 'ti' 'Mo' 'ğol' 'la' 'rın' 'li' 'de' 'ri' ' ' 'Ye' 'sü' 'gey' 'Tuğ' 'ru' 'la' 'ka' 'bi' 'le' 'si' 'nin' 'ön' 'der' 'li' 'ğı' 'ni' 'ye' 'ni' 'den' 'e' 'le' 'ge' 'çir' 'me' ' ' 'et' 'ti' 'Tuğ' 'rul' 've' 'Ye' 'sü' 'gey' 'an' 'da' 'i' 'le' 'kar' 'de' 'şi' 'lik' 'ye' 'mi' 'ni' 'et' 'tik' 'ten' ' ' 'son' 'ra' 'da' 'ha' 'son' 'ra' 'la' 'rı' 'Mo' 'ğol' 'la' 'rın' 'ye' ' '

Test Data

'An' 'tik' 'ya' 'zar' 'lar' 'ya' 'pı' 'nın' 'mi' 'ma' 'rı' 'nın' 'ol' 'du' 'ğu' 'nu' 'kay' 0.0
'kay' 'det' 'mek' 'te' 'dir' 'Ay' 'rı' 'ca' '""un" 'a' 'dı' 'da' 'geç' 0.0
'geç' 'mek' 'te' 'dir' ' ' 'Vit' 'ru' 'vius' 'us' 'Mö' 'yüz' 0.0
'yüz' 'yı' 'lın' 'en' 'ö' 'nem' 'li' 'dört' 'hey' 'kel' 'tı' 'ra' 'şı' 'nın' 'bu' 'ya' 'pı' 'da' 'ça' 'lış' 'tı' 'ğı' 'nı' 'kay' 0.0
'kay' 'det' 'miş' 'tir' 'Do' 'ğu' 'da' ' ' 'Sko' 'pas' 'ba' 'tı' 'da' ' ' 'Le' 'ok' 'ha' 'res' 'ku' 'zey' 'de' ' ' 'Bryak' 'sis' 'gü' 'ney' 'de' 'Ti' 'mot' 'heos' 'os' 'ça' 0.0
'ça' 'lış' 'mış' 'tır' ' ' 'Bryak' 'sis' 'Kar' 'ya' 'lı' 'bir' 'sa' 0.0
'sa' 'nat' 'çı' 'dır' 'Di' 'ğer' 'sa' 'nat' 'çı' 'lar' 'Yu' 'na' 'nis' 'tan' ' ' 'dan' 'ge' 0.0
'ge' 'ti' 'ril' 'miş' 'tir' 'Dört' 'at' 'lı' 'a' 'ra' 'ba' 'yı' 'Mi' 'mar' 'Pythe' 'os' '""un" 'yap' 'tı' 'ğı' 'söy' 0.0