

**TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**



DİNAMİK PROGRAMLAMA

16011705 - YUNUS EMRE DEMİR

Algoritma Analizi Ödevi - III

Aralık, 2020

Problem 2: Sorgulanan bir cümlede yanlış yazılmış kelimeler varsa bu kelimelerin yerine doğru kelimeler öneren bir sistem tasarlanacaktır.

```
-----
Bir cumle giriniz : my name is Yunus
'my' sozlukte var
'name' sozlukte var
'is' sozlukte var

'yunus' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...
Distance:1 olan kelime yok distance:2 olana var mi bakiliyor...
Distance'si <=2 olan bir kelime bulunamadi.'

New Sentence : my name is yunus
```

Yukarıdaki örnekte “my”, “name” ve “is” sözcüklerinin 1. sözlükte (smallDicrtionary.txt) bulunduğunu, dolayısıyla doğru yazıldıkları anlaşılmaktadır. “Yunus” sözcüğü ise 1. Sözlükte aranmıştır. Bulunamayınca 2. Sözlüğe (hatalı kelime tablosunda) bakıldığı, orada da bulunamayınca 1. Sözlükteki yakın anlamlılarına bakılmıştır. Yakın anlamlılarda da uzaklığı 2’den küçük sözcük bulunamadığı için bir öneri sunulmamıştır. Direkt olarak kabul edilmiştir.

```
-----
Bir cumle giriniz : it is a taxi
'it' sozlukte var
'is' sozlukte var
'a' sozlukte var

'taxi' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...
Distance:1 olan kelime yok distance:2 olana var mi bakiliyor...
Distance'si <=2 olan bir kelime bulunamadi.'

New Sentence : it is a taxi
```

Yukarıdaki örnekte de ilk 3 kelime sözlükte bulunduğu için doğru kabul edilmiştir. “taxi” sözcüğü ise 1. Ve 2. sözlükte bulunamamıştır. Yine yakın anlamlılar sözlüğünde de benzer kelimeye rastlanmamıştır. Bu sebeple doğrudan alınmıştır.

```
-----  
Bir cumle giriniz : the function successfully working  
'the' sozlukte var  
'function' sozlukte var  
'successfully' sozlukte var  
  
'working' 1. sozlukte yok! Hash Table 2'ye bakiliyor...  
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...  
Distance:1 olan kelime yok distance:2 olana var mi bakiliyor...  
warning --distance: 2  
  
'working' is't in the dict. Did you mean : 'warning' or ? :warning  
  
New Sentence : the function successfully warning  
  
-----  
Bir cumle giriniz :  
  
New Sentence :  
-----  
Bir cumle giriniz : the function successfully working  
'the' sozlukte var  
'function' sozlukte var  
'successfully' sozlukte var  
  
'working' 1. sozlukte yok! Hash Table 2'ye bakiliyor...  
  
New Sentence : the function successfully warning
```

Yukarıdaki örnekte “working” kelimesi 1. Ve 2. Sözlükte bulunamamıştır. Yakın anlamlılarına bakıldığında 1 uzaklıklı kelime bulunamamış ve 2 uzaklıklı kelimelere bakıldığında “warning” kelimesi bulunmuş ve kullanıcıya önerilmiştir. Kullanıcı bu kelimeyi doğru kelime olarak seçmiştir. Bu kelime seçildikten sonra 2. Sözlüğe (yanlış ve doğru hali ile) eklenmektedir.

Kullanıcıdan 2. defa yeni cümle istendiğinde, kullanıcı ilk cümlelerin aynısını girmiştir. Bu sefer “working” kelimesi için 2. Sözlüğe bakıldığında doğru yazımı bulunmuştur. Kullanıcıya 2. Sözlükten düzeltilmiş olarak geri gönderilmiştir.

```
Bir cumle giriniz : this dictionaryy is totally wrong
'this' sozlukte var

'dictionaryy' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...
dictionary --distance: 1

'dictionaryy' is't in the dict. Did you mean : 'dictionary' or ? :dictionary
'is' sozlukte var

'totally' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...
Distance:1 olan kelime yok distance:2 olana var mi bakiliyor...
Distance'si <=2 olan bir kelime bulunamadi.'
'wrong' sozlukte var

New Sentence : this dictionary is totally wrong
```

“dictionaryy” yanlış yazılmıştır. Edit distance ile doğrusu önerilmiştir. Kullanıcı direktifi doğrultusunda doğrusu 2. Sözlüğe eklenmiştir. “totally” kelimesi hiçbir sözlükte bulunamayıp, yakın anlamlı sözcükte önerilememiştir. Olduğu gibi alınmıştır.

```
-----
Bir cumle giriniz : this dictionaryy is totally wrong
'this' sozlukte var

'dictionaryy' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
'is' sozlukte var

'totally' 1. sozlukte yok! Hash Table 2'ye bakiliyor...
HashTable 2'de de yok! Benzer kelimeler arastiriliyor...
Distance:1 olan kelime yok distance:2 olana var mi bakiliyor...
Distance'si <=2 olan bir kelime bulunamadi.'
'wrong' sozlukte var

New Sentence : this dictionary is totally wrong
```

Tekrar aynı cümle gönderildiğinde “dictionaryy” kelimesi için 2. sözlükte doğrusu bulunmuştur. Doğrusu 2. sözlükten getirilmiştir.

Bonus : Verilen A[n] ve B[m] stringler arasındaki edit distance hesabını sonuna kadar yapmayıp, A ve B stringleri arasındaki mesafe k değerinden(bu ödev için mesafesi 2'den büyük olan sözlük kelimeleri kullanılmayacağı için k=2'dir) büyük olacağı kesin olduğu adımda işlemi bitirenler ek 20 puanalacaktır. Verilen A[n] ve B[m] stringler ve k değeri için değişim matrisinde ana diagonalde [0,0]'dan en fazla kaç adım ilerlenmesi yeterlidir? Kısaca açıklayınız.

For two sequences a and b with lengths i and j respectively, the Levenshtein distance is defined as:

$$ld(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} ld(i-1, j) + 1 \\ ld(i, j-1) + 1 \\ ld(i-1, j-1) + k \end{cases} & \text{otherwise.} \end{cases}$$

where $k = 0$ if $(a_i = b_j)$, 1 otherwise.

Edit distancenin formülü gereği bir eleman oluşturulurken:

- Ya $M(i-1, j-1)$ deki değeri aynen alır yada değeri +1 artmak zorundadır. (nereden geldiğinin bir önemi yok)
- Matris oluşturulurken, önce 1. Satır sonra 2. Satır sonra 3. ... diye ilerler.
- Bu bağlamda 4. Satırı oluştururken 3. satırdaki tüm değerler 2 den büyükse 4. Satırdaki tüm değerlerinde 2 den büyük olacağını garanti etmiş olur.
 - Çünkü 4. Satırdaki herhangi bir değer alacağı en küçük değer diagonalin aynısı olmak zorundadır. Eğer diagonal 2 den büyükse ilgili indiste 2'den büyük olmak zorundadır.
- Dolayısıyla Matris oluşturulurken her satır için, satırda 3'ten küçük sayı var mı diye kontrol edilir. Eğer yoksa matrisi oluşturmaya gerek yoktur. Çünkü edit distance en iyi ihtimalle 3 çıkacak demektir.

Matris									
0	1	2	3	4	5	6	7	8	9
1	1	2	3	3	4	5	6	7	8
2	2	2	3	4	3	4	5	6	7
3	3	3	3	4	4	3	4	5	6
4	4	4	4	4	5	4	3	4	5
5	5	5	5	5	4	5	4	3	4
6	6	6	6	6	5	5	5	4	3

usemre
yunusemre

3. satırdan sonrasına
bakmasına gerek yok.

Matris									
0	1	2	3	4	5	6	7	8	9
1	1	2	3	3	4	5	6	7	8
2	2	2	3	4	3	4	5	6	7
3	3	3	3	4	4	3	4	5	6
4	4	4	4	4	5	4	3	4	5
5	5	5	5	5	4	5	4	3	4
6	6	6	6	6	5	5	5	4	3

y semre
yunusemre

3. Satırdan sonra
Bakmasına gerek yok

Matris									
0	1	2	3	4	5	6	7	8	9
1	0	1	2	3	4	5	6	7	8
2	1	0	1	2	3	4	5	6	7
3	2	1	0	1	2	3	4	5	6
4	3	2	1	0	1	2	3	4	5
5	4	3	2	1	0	1	2	3	4
6	5	4	3	2	1	1	2	3	3

yunuse
yunusemre

Bu durum ise worstcase
durumdur.

Ana dioganelden kaç adım ilerlemelidir? Sorusunun cevabı tamamen girdiler ile alakalıdır. Eğer farklılık kelimenin başındaysa 3-4 adım ilerlemeyle farkına varır. (Bestcase durumu) Eğer değişim kelime sonunda ise matrisi tamamlamak zorundadır. (worstcase durumu) Çünkü oraya gelene kadar kelime birebir aynıdır.