

BLG505 Doğal Dil İşleme Ödev 1 Raporu

Sümeyye ÖZTÜRK

504191569

Kasım 18, 2020

1) XFST Aracının Kullanımı

Yüklemiş olduğumuz xfst aracını “./xfst” komutunu kullanarak aşağıdaki resimde gösterildiği gibi terminalimiz üzerinden çalıştırıyoruz.

```
bin — xfst — 80x24
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]:
```

Figure 1.1: XFST Çalıştırma Komutu

2) XFST Aracında Türkçe Karakterleri Kullanmak İçin Ayarları Yapılandırma

Fsmbook [1] sitesinde gösterildiği gibi Utf-8 harf kodu ile çalışmak için:

“set char-encoding utf-8” komutunu kullanarak bu işlemi gerçekleştiririz.

```
bin — xfst — 80x24
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]: set char-encoding utf-8
variable char-encoding = UTF-8
xfst[0]:
```

Figure 2.1: UTF-8 Harf Kodu Çalıştırma Komutu

3) XFST İfadelerini Tanımak İçin Yapılan Çalışmalar

XFST: Intro Notes [2] sitesinde bulunan alıştırmlar terminal üzerinden denenerek yapılacak olduğumuz çalışma için önden hazırlık yapılmıştır. Aşağıdaki resimlerde de yapılan bazı örnekler gösterilmiştir.

```
bin — xfst — 80x24
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]: regex [d o g | c a t];
928 bytes. 6 states, 6 arcs, 2 paths.
xfst[1]: up
apply up> dog
dog
apply up> cat
cat
apply up> zebra
???
apply up> fir
???
apply up> END;
xfst[1]: down
apply down> dog
dog
apply down>
```

Figure 3: Alıştırma 1

```
bin — xfst — 80x24
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]: read regex [d o g | c a t];
928 bytes. 6 states, 6 arcs, 2 paths.
xfst[1]: print net
Sigma: a c d g o t
Size: 6.
Flags: deterministic, pruned, minimized, epsilon_free, loop_free
Arity: 1
s0: c -> s1, d -> s2.
s1: a -> s3.
s2: o -> s4.
s3: t -> fs5.
s4: g -> fs5.
fs5: (no arcs)
xfst[1]: print words
cat
dog
xfst[1]:
```

Figure 4: Alıştırma 2

```
bin — xfst — 80x24
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]: regex {dog}:{köpek};
1.4 Kb. 6 states, 5 arcs, 1 path.
xfst[1]: down dog
köpek
xfst[1]: up köpek
dog
xfst[1]:
```

Figure 5: Alıştırma 3

4) Script Örneği

İngilizce Sayılar için verilmiş “NumbersToNumerals” scriptini çalıştırıp üzerinde denemeler gerçekleştirdikten sonra yapılacak olan ödev için gerekli ön çalışma bitmiş oldu.

```
bin — xfst — 80x25
~/bin — xfst

xfst[0]: source NumbersToNumerals.script
Opening input file 'NumbersToNumerals.script'
November 18, 2020 11:31:46 GMT
Defined 'OneToNine': 2.3 Kb. 25 states, 32 arcs, 9 paths.
Defined 'TeenTen': 1.8 Kb. 19 states, 23 arcs, 6 paths.
Defined 'Teens': 2.8 Kb. 34 states, 42 arcs, 10 paths.
Defined 'TenStem': 2.0 Kb. 21 states, 27 arcs, 8 paths.
Defined 'Tens': 3.8 Kb. 49 states, 64 arcs, 80 paths.
Defined 'OneToNinetyNine': 6.0 Kb. 81 states, 113 arcs, 99 paths.
Print random numbers
30
79
4
44
42
Print random numerals
forty-seven
eighty-eight
six
sixty-one
seventy-six
Map numeral to number
Input: seventy-nine
79
Closing file NumbersToNumerals.script...
```

Figure 6: Alıştırma 4

```
xfst[1]: push OneToNine;
xfst[2]: apply down
apply down> 1
one
apply down> 2
two
apply down> 9
nine
apply down> END;
xfst[2]: apply up
apply up> one
1
apply up> nine
9
apply up> END;
xfst[2]:
```

Figure 7: Alıştırma 5

5) Türkçe Sayı Dönüşümleri İçin XFST Çalışması

Bu bölümde ödevde istenmiş olan Türkçe sayıların rakam ve metin halleri arasında dönüşümü gerçekleştiren bir xfst script'i yazılmıştır. Yazılan bu Script 1'den 9999'a kadar olan sayılar için gerekli dönüşümü gerçekleştirebiliyor.

Yapılan çalışmada aşağıdaki ilk iki resimde görüldüğü gibi önce oluşturduğumuz ifadelerin tanımlanması gerçekleştirilip ardından da hem yazılı hem de rakam halleri için rastgele farklı 10 adet sayı üretilerek gösterilmiştir.

```
~/bin — xfst
sumeyyeozturk@Sumeyye-MacBook-Pro bin % ./xfst
Copyright © Palo Alto Research Center 2001-2020
PARC Finite-State Tool, version 2.15.7 (libcfsm-2.25.11) (svn 34269)

Type "help" to list all commands available or "help help" for further help.

xfst[0]: source NumbersToNumeralsTurkish.script
Opening input file 'NumbersToNumeralsTurkish.script'
November 18, 2020 11:53:12 GMT
Defined 'TwoToNine': 3.6 Kb. 22 states, 28 arcs, 8 paths.
Defined 'OneToNine': 3.7 Kb. 24 states, 31 arcs, 9 paths.
Defined 'TenToNinety': 4.2 Kb. 32 states, 39 arcs, 9 paths.
Defined 'Tens': 5.9 Kb. 56 states, 72 arcs, 90 paths.
Defined 'OneHundred': 6.4 Kb. 64 states, 82 arcs, 100 paths.
Defined 'OtherHundreds': 7.8 Kb. 87 states, 112 arcs, 800 paths.
Defined 'Hundreds': 8.0 Kb. 89 states, 117 arcs, 900 paths.
Defined 'OneThousand': 6.6 Kb. 67 states, 86 arcs, 101 paths.
Defined 'OtherThousands': 9.9 Kb. 119 states, 154 arcs, 7200 paths.
Defined 'Thousands': 10.3 Kb. 126 states, 165 arcs, 7301 paths.
Defined 'OneToThousands': 10.8 Kb. 127 states, 182 arcs, 8300 paths.
```

```
Print random numbers
997
8958
125
2780
6597
2365
2017
4653
701
3043
Print random numerals
beş-bin-on-altı
dokuz-bin-kırk-beş
dokuz-yüz
iki-bin-dört-yüz-doksan-beş
yedi-bin-dokuz-yüz-yirmi-beş
iki-bin-sekiz-yüz-otuz-altı
altı-bin-üç-yüz-doksan-dört
üç-bin-dokuz-yüz-iki
yedi-yüz-seksen-sekiz
yedi-bin-dokuz-yüz-dokuz
Closing file NumbersToNumeralsTurkish.script...
xfst[1]:
```

Figure 8: Çalışma Script'i

Ardından En son oluşturduğumuz sonlu durum ifadesini “push OneToThousand” komutunu kullanarak xfst aracına veriyoruz.

Daha sonra “apply down” ve “apply up” komutlarını kullanarak rastgele sayılar üzerinde aşağıdaki resimde de görüldüğü gibi dönüşümler gerçekleştiriyoruz.

```
[xfst[1]: push OneToThousands
[xfst[2]: apply down
[apply down> 56
elli-altı
[apply down> 489
dört-yüz-seksen-dokuz
[apply down> 7634
yedi-bin-altı-yüz-otuz-dört
[apply down> END;
[xfst[2]: apply up
[apply up> yirmi-iki
22
[apply up> beş-yüz-yetmiş-iki
572
[apply up> üç-bin-üç-yüz-doksan-dokuz
3399
[apply up> END;
```

Figure 9: Çalışma Deneme Çıktıları

Kaynakça

- [1] K. R. Beesley ve L. Karttunen, «Finite State Morphology Homepage,» 9 Ağustos 2010. [Çevrimiçi]. Available: <https://web.stanford.edu/~laurik/fsmbook/faq/utf8.html>. [Erişildi: 9 Kasım 2020].
- [2] J. M. Gawron, «XFST: Intro Notes,» 2014. [Çevrimiçi]. Available: https://gawron.sdsu.edu/compling/course_core/xfst_intro/intro_notes.html#getting-started. [Erişildi: 9 Kasım 2020].