

```

#-----#
#Autoren: Emir Arnautovic & Marco Costa
#HTW Chur / Studienarbeit 2
#TET-TZ08 / TET-VZ08
#-----#
# Begrüssung
cat(" #####", "\n", "# Sprachenerkennung
V.1.0 FS09 #, "\n", "#####", "\n\n")
#-----#

```

```

#Offset=2; Z.B. wenn a<-5% ist, dann beim suchen kann diesen Wert von 3-7% wandeln
offset<-5

```

```

#Sonderzeichen werden nicht beruecksichtigt
buchstaben <- matrix(0,26,2)

```

```

buchstaben[1,1] <- "a"
buchstaben[1,2] <- 0
buchstaben[2,1] <- "b"
buchstaben[2,2] <- 0
buchstaben[3,1] <- "c"
buchstaben[3,2] <- 0
buchstaben[4,1] <- "d"
buchstaben[4,2] <- 0
buchstaben[5,1] <- "e"
buchstaben[5,2] <- 0
buchstaben[6,1] <- "f"
buchstaben[6,2] <- 0
buchstaben[7,1] <- "g"
buchstaben[7,2] <- 0
buchstaben[8,1] <- "h"
buchstaben[8,2] <- 0
buchstaben[9,1] <- "i"
buchstaben[9,2] <- 0
buchstaben[10,1] <- "j"
buchstaben[10,2] <- 0
buchstaben[11,1] <- "k"
buchstaben[11,2] <- 0
buchstaben[12,1] <- "l"
buchstaben[12,2] <- 0
buchstaben[13,1] <- "m"
buchstaben[13,2] <- 0
buchstaben[14,1] <- "n"
buchstaben[14,2] <- 0
buchstaben[15,1] <- "o"
buchstaben[15,2] <- 0
buchstaben[16,1] <- "p"
buchstaben[16,2] <- 0
buchstaben[17,1] <- "q"
buchstaben[17,2] <- 0
buchstaben[18,1] <- "r"
buchstaben[18,2] <- 0
buchstaben[19,2] <- "s"
buchstaben[19,2] <- 0

```

```
buchstaben[20,2] <- "t"  
buchstaben[20,2] <- 0  
buchstaben[21,2] <- "u"  
buchstaben[21,2] <- 0  
buchstaben[22,2] <- "v"  
buchstaben[22,2] <- 0  
buchstaben[23,2] <- "w"  
buchstaben[23,2] <- 0  
buchstaben[24,2] <- "x"  
buchstaben[24,2] <- 0  
buchstaben[25,2] <- "y"  
buchstaben[25,2] <- 0  
buchstaben[26,2] <- "z"  
buchstaben[26,2] <- 0
```

```
#Buchstabenhaeufigkeitstabelle - Deutsch / Angabe in %  
#Quelle - http://en.wikipedia.org/wiki/Letter\_frequencies  
#Erste Parameter ist die Sprache, zweite Parameter ist die id des Buchstabe
```

```
languages <- matrix(0,3,26)
```

```
languages[1,1] <- 6.51  
languages[1,2] <- 1.89  
languages[1,3] <- 3.06  
languages[1,4] <- 5.08  
languages[1,5] <- 17.4  
languages[1,6] <- 1.66  
languages[1,7] <- 3.01  
languages[1,8] <- 4.76  
languages[1,9] <- 7.55  
languages[1,10] <- 0.27  
languages[1,11] <- 1.21  
languages[1,12] <- 3.44  
languages[1,13] <- 2.53  
languages[1,14] <- 9.78  
languages[1,15] <- 2.51  
languages[1,16] <- 0.79  
languages[1,17] <- 0.02  
languages[1,18] <- 7  
languages[1,19] <- 7.27  
languages[1,20] <- 6.15  
languages[1,21] <- 4.35  
languages[1,22] <- 0.67  
languages[1,23] <- 1.86  
languages[1,24] <- 0.03  
languages[1,25] <- 0.04  
languages[1,26] <- 1.13
```

```
#Buchstabenhaeufigkeitstabelle - Italienisch / Angabe in %  
#Quelle - http://en.wikipedia.org/wiki/Letter\_frequencies  
#Erste Parameter ist die Sprache, zweite Parameter ist die id des Buchstabe
```

```
languages[2,1] <- 11.74
languages[2,2] <- 0.92
languages[2,3] <- 4.5
languages[2,4] <- 3.73
languages[2,5] <- 11.79
languages[2,6] <- 0.95
languages[2,7] <- 1.64
languages[2,8] <- 1.54
languages[2,9] <- 11.28
languages[2,10] <- 0
languages[2,11] <- 0
languages[2,12] <- 6.51
languages[2,13] <- 2.51
languages[2,14] <- 6.88
languages[2,15] <- 9.83
languages[2,16] <- 3.05
languages[2,17] <- 0.51
languages[2,18] <- 6.73
languages[2,19] <- 4.98
languages[2,20] <- 5.62
languages[2,21] <- 3.01
languages[2,22] <- 2.1
languages[2,23] <- 0
languages[2,24] <- 0
languages[2,25] <- 0
languages[2,26] <- 0.49
```

#Buchstabenhaeufigkeitstabelle - Bosnisch / Angabe in %

#Erste Parameter ist die Sprache, zweite Parameter ist die id des Buchstabe

```
languages[3,1] <- 11.0753
languages[3,2] <- 1.4647
languages[3,3] <- 2.6334
languages[3,4] <- 3.8584
languages[3,5] <- 9.2974
languages[3,6] <- 0.2811
languages[3,7] <- 1.5668
languages[3,8] <- 0.5737
languages[3,9] <- 10.0268
languages[3,10] <- 6.0788
languages[3,11] <- 3.2428
languages[3,12] <- 2.7299
languages[3,13] <- 3.4733
languages[3,14] <- 6.7714
languages[3,15] <- 9.1003
languages[3,16] <- 2.6774
languages[3,17] <- 0.0139
languages[3,18] <- 4.3387
languages[3,19] <- 5.8055
languages[3,20] <- 4.6139
languages[3,21] <- 4.2278
languages[3,22] <- 3.512
languages[3,23] <- 0.0166
```

```
languages[3,24] <- 0.0066
languages[3,25] <- 0.01
languages[3,26] <- 2.6024
```

```
#-----#}
```

```
#Datei einlesen
```

```
cat("Bitte wählen Sie ein Textdatei...", "\n\n", sep = "")
```

```
id<-readline()
```

```
  if(id == 1)
```

```
    textfile <- file(file.choose(), "r")
```

```
    text <- readLines(textfile, warn = FALSE)
```

```
textSplitted<-substring(text,1:nchar(text),1:nchar(text))
```

```
i<-1
```

```
#-----#
```

```
#Die Buchstaben im Text werden gezählt
```

```
while(i <= length(textSplitted)){
```

```
  a<-1
```

```
  while(a <= length(buchstaben)/2){
```

```
    if(gregexpr(buchstaben[a,1], textSplitted[i])==1){
```

```
      buchstaben[a,2]<-as.numeric(buchstaben[a,2])+1
```

```
      break
```

```
    }
```

```
    a <- a+1
```

```
  }
```

```
  i <- i+1
```

```
}
```

```
chars<-length(textSplitted)
```

```
a<-0
```

```
while(a <= length(buchstaben)/2){
```

```
  buchstaben[a,2]<-((as.numeric(buchstaben[a,2])*100)/chars)
```

```
  a <- a+1
```

```
}
```

```
#-----#
```

```
#Hier wird entscheidet was für eine Sprache den Text hat.
```

```
i<-1
```

```
ok<-0
```

```
while(i <= length(languages)){
```

```
  a<-1
```

```
  while(a <= length(buchstaben)/26){
```

```
    buchstabeWert<-as.numeric(buchstaben[a][2])
```

```
    if(buchstabeWert >= as.numeric(languages[i,a])+offset && buchstabeWert <=
```

```
as.numeric(languages[i,a])-offset ){
```

```
      ok<-ok+1
```

```
    }
    else {
        ok<-0
        break
    }
    if(ok==26){break}
    a <- a+1
}
if(ok==26){break}
i <- i+1
}

if(ok==1){
    if(i==1) {lang="Deutsch"}
    if(i==2) {lang="Italienisch"}
    if(i==3) {lang="Bosnisch"}
    cat(lang)
}
#-----#
```