

Exercise leads to UTF-8 fun!

mark.yoshikawa@gmail.com

Thank you to Jose Portilla and Inanc Gumus...

It started with an exercise and since it wanted me to “5. Print a non-english sentence using a string literal” I decided to use Google’s translate web app...

(Note that Arabic, Urdu, and Kurdish in my experiment did not show up in the terminals, only as blank characters or some other non-language printing... so I got curious (Nested Note: In a Google Drive doc, apparently, those characters show up from a copy and paste from my Cygwin terminal. That should end my investigation but what's the fun of that. So pretend they were blank or some other characters that had no meaning in the intended language. **So I wanted to investigate what it was really printing out**))

```
$ cat main.go
// Copyright © 2018 Inanc Gumus
// Learn Go Programming Course
// License: https://creativecommons.org/licenses/by-nc-sa/4.0/
//
// For more tutorials : https://learnprogramming.com
// In-person training : https://www.linkedin.com/in/inancgumus/
// Follow me on twitter: https://twitter.com/inancgumus

package main

// -----
// EXERCISE: Print the literals
//
// 1. Print a few integer literals
//
// 2. Print a few float literals
//
// 3. Print true and false bool constants
//
// 4. Print your name using a string literal
//
// 5. Print a non-english sentence using a string literal
//
// -----
```

```

import (
    "fmt"
)

func main() {
    // Use fmt.Println()
    fmt.Println(42)
    fmt.Println(314159)
    fmt.Println(-256)
    fmt.Println(3.14159)
    fmt.Println(true)
    fmt.Println(false)
    fmt.Println("Mark Cyber33")
    fmt.Println("Let's try english and convert this sentence using Google's translate.")
    // Arabic not visible but runs and the cursor goes in the right direction when you move over
    the text i.e. right to left.
    fmt.Println("لنجرب اللغة الإنجليزية ونحوّل هذه الجملة باستخدام ترجمة Google.")
    // Spanish is good.
    fmt.Println("Probemos inglés y conviertamos esta oración usando el traductor de Google.")
    // Russian
    fmt.Println("Давайте попробуем по-английски и переведем это предложение с
    помощью переводчика Google.")
    // Chinese Traditional
    fmt.Println("讓我們試試英語並使用谷歌的翻譯來轉換這句話。")
    // Tamil
    fmt.Println("Google இன் மொழிபெயர்ப்பைப் பயன்படுத்தி இந்த வாக்கியத்தை
    ஆங்கிலத்தில் மாற்ற முயற்சிப்போம்.")
    // Urdu not so much but runs and the cursor goes in the right direction when you move over
    the text i.e. right to left.
    fmt.Println("آئیے انگریزی کی کوشش کریں اور گوگل کے ترجمہ کا استعمال کرتے ہوئے اس جملے کو تبدیل کریں۔")
    // Swahili
    fmt.Println("Hebu tujaribu Kiingereza na tubadili sentensi hii kwa kutumia tafsiri ya
    Google.")
    // Korean
    fmt.Println("Google 번역을 사용하여 영어를 시도하고 이 문장을 변환해 봅시다.")
    // Persian not correct visually but runs and the cursor goes in the right direction when you
    move over the text i.e. right to left.
    fmt.Println("بایا بید انگلیسی را امتحان کنیم و این جمله را با استفاده از ترجمه گوگل تبدیل کنیم")
    // Kurdish (Kurmanji)
    fmt.Println("Ka em îngilîzî biceribînin û vê hevokê bi wergera Google-ê veguherînin.")
}

```

DETECT LANGUAGE ARABIC PORTUGUESE ENGLISH ARABIC KURDISH (KURMANJI) GALICIAN

Let's try english and convert this sentence using Google's translate. × . لنجد اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.

linujarib allughat al'iinjiliziat wanuhul hadhih aljumlat biastikhdam tarjamat Google.

Send feedback

Seems simple enough... just copy and put into my fmt.Println statements... but noticed my terminal did not show correctly. I knew it had to be in something like the driver for my Cygwin terminal so tried my Windows terminal and actually worse. Still I did not doubt Go :-).

```
mark_8LAPTOP-3D73FF3E /cygdrive/c/Users/mark_/_Data/Google/Interview Prep/Golang/learngo/06-variables/01-basic-data-types/exercises/01-print-the-literals/TryIt
$ ./main.go
42
314159
-256
3.14159
true
false
Mark Cyber33
Let's try english and convert this sentence using Google's translate.
Probemos inglés y conviertamos esta oración usando el traductor de Google.
Попробуйте попробуем по-английски и переведем это предложение с помощью переводчика Google.
Google 를 사용해보세요 문장을 시도하고 이 문장을 번역해 봅시다.
Ka em inglezî biceribînin û vê hevokê bi wergera Google-ê veguherînin.
mark_8LAPTOP-3D73FF3E /cygdrive/c/Users/mark_/_Data/Google/Interview Prep/Golang/learngo/06-variables/01-basic-data-types/exercises/01-print-the-literals/TryIt
$
```

\$./main.go

42
314159
-256
3.14159
true
false
Mark Cyber33

Let's try english and convert this sentence using Google's translate.

. لنجد اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.

Probemos inglés y conviertamos esta oración usando el traductor de Google.

Давайте попробуем по-английски и переведем это предложение с помощью переводчика Google.

讓我們試試英語並使用谷歌的翻譯來轉換這句話。

Google ଇଣ୍ଟ ମେମାଟିପେଯର୍‌ପପେପ୍ ପାଯନ୍‌ପଟୁତ୍ତି ଇନ୍ତ ବାକକିଯତ୍ତରେ ଆନ୍ଦିଲତ୍ତରିଲ୍ ମାର୍ଗ ମୁଯର୍‌ଶିପ୍‌ପୋମ୍.

آئیں انگریزی کی کوشش کریں اور گوگل کے ترجمہ کا استعمال کرتے ہوئے اس جملے کو تبدیل کریں۔

Hebu tujaribu Kiingereza na tubadili sentensi hii kwa kutumia tafsiri ya Google.

Google 번역을 사용하여 영어를 시도하고 이 문장을 번역해 봅시다.

بیابید انگلیسی را امتحان کنیم و این جمله را با استفاده از ترجمه گوگل تبدیل کنیم.

Ka em îngilîzî biceribînin û vê hevokê bi wergera Google-ê veguherînin.

\$

My first solution was to use od -xc... and decided to focus the code on just the English and Arabic version:

```
....  
func main() {  
    // Use fmt.Println()  
    fmt.Println("Let's try english and convert this sentence using Google's translate.")  
    fmt.Println("\n\n\n\n")  
    // Arabic not visible but runs and the cursor goes in the right direction when you move over  
    the text i.e. right to left.  
    fmt.Println("لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.")  
}
```

```
$ /home/mark/_go/bin/go run englishAndArabic.go | od -xc  
0000000 654c 2774 2073 7274 2079 6e65 6c67 7369  
    L e t ' s   t r y   e n g l i s  
0000020 2068 6e61 2064 6f63 766e 7265 2074 6874  
    h   a n d   c o n v e r t   t h  
0000040 7369 7320 6e65 6574 636e 2065 7375 6e69  
    i s   s e n t e n c e   u s s i n  
0000060 2067 6f47 676f 656c 7327 7420 6172 736e  
    g   G o o g l e ' s   t r a n s  
0000100 616c 6574 0a2e 0a0a 0a0a d90a d984 d886  
    I a t e . \n \n \n \n \n 331 204 331 206 330  
0000120 d8ac d8b1 20a8 a7d8 84d9 84d9 bad8 a9d8  
    254 330 261 330 250 330 247 331 204 331 204 330 272 330 251  
0000140 d820 d9a7 d884 d9a5 d886 d9ac d984 d88a  
    330 247 331 204 330 245 331 206 330 254 331 204 331 212 330  
0000160 d9b2 d88a 20a9 88d9 86d9 add8 88d9 84d9  
    262 331 212 330 251 331 210 331 206 330 255 331 210 331 204  
0000200 d920 d887 d9b0 2087 a7d8 84d9 acd8 85d9  
    331 207 330 260 331 207 330 247 331 204 330 254 331 205  
0000220 84d9 a9d8 d820 d8a8 d8a7 d8b3 d8aa d8ae  
    331 204 330 251 330 250 330 247 330 263 330 252 330 256 330  
0000240 d8af d9a7 2085 aad8 b1d8 acd8 85d9 a9d8  
    257 330 247 331 205 330 252 330 261 330 254 331 205 330 251  
0000260 4720 6f6f 6c67 2e65 000a  
    G o o g l e . \n  
0000271
```

\$

I looked at the [Arabic UTF-8 table](#) to see if I'm close to what I suspect and recognize the "d8" and "d9" hex codes, but **maybe a better way is to have it print out hex from Go's Printf...**

```
func main() {
    // Use fmt.Println()
    fmt.Println("Let's try english and convert this sentence using Google's translate.")
    fmt.Println("\n\n\n\n")
    // Arabic not visible but runs and the cursor goes in the right direction when you move over
    // the text i.e. right to left.
    fmt.Println("لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.")
    aStr := "لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google."
    fmt.Printf("%x\n", aStr)
}
```

\$ /home/mark/_go/bin/go run englishAndArabic.go

Let's try english and convert this sentence using Google's Translate.

لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.

d9 84 d9 86 d8 ac d8 b1 d8 a8 20 d8 a7 d9 84 d9 84 d8 ba d8 a9 20 d8 a7 d9 84 d8 a5 d9 86
d8 ac d9 84 d9 8a d8 b2 d9 8a d8 a9 20 d9 88 d9 86 d8 ad d9 88 d9 84 20 d9 87 d8 b0 d9 87
20 d8 a7 d9 84 d8 ac d9 85 d9
84 d8 a9 20 d8 a8 d8 a7 d8 b3 d8 aa d8 ae d8 af d8 a7 d9 85 20 d8 aa d8 b1 d8 ac d9 85 d8 a9
20 47 6f 6f 67 6c 65 2e

For me that is much better since the grouping seems to focus on just the string than in od... but
let me confirm that the method to copy the Arabic text is true UTF-8...

```
func main() {
    // Use fmt.Println()
    fmt.Println("Let's try english and convert this sentence using Google's translate.")
    fmt.Println("\n\n\n\n")
    // Arabic not visible but runs and the cursor goes in the right direction when you move over
    // the text i.e. right to left.
    fmt.Println("لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google.")
    // aStr := "لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google."
    // This is to confirm I used the copy 'button' in the Google Translate app and not just the
    // mouse - I'm trusting Google at this point the copy is the true UTF-8 Arabic text
    aStr := "لتجرب اللغة الإنجليزية ونحو هذه الجملة باستخدام ترجمة Google."
    fmt.Printf("%x\n", aStr)
}
```

```
$ /home/mark/_go/bin/go run englishAndArabic.go  
Let's try english and convert this sentence using Google's translate.
```

لنجرب اللغة الإنجليزية ونحوّل هذه الجملة باستخدام ترجمة Google.
d9 84 d9 86 d8 ac d8 b1 d8 a8 20 d8 a7 d9 84 d9 84 d8 ba d8 a9 20 d8 a7 d9 84 d8 a5 d9 86
d8 ac d9 84 d9 8a d8 b2 d9 8a d8 a9 20 d9 88 d9 86 d8 ad d9 88 d9 84 20 d9 87 d8 b0 d9 87
20 d8 a7 d9 84 d8 ac d9 85 d9
84 d8 a9 20 d8 a8 d8 a7 d8 b3 d8 aa d8 ae d8 af d8 a7 d9 85 20 d8 aa d8 b1 d8 ac d9 85 d8 a9
20 47 6f 6f 67 6c 65 2e

I'm no expert in Arabic or English for that matter but something tells me that I should question my results...

لنجرب اللغة الإنجليزية ونحوّل هذه الجملة باستخدام ترجمة Google.

Ok look from right to left and magnify the characters...

نجر

Now from left to right after the Google (or before in Arabic)...

مة ج

Check against the UTF-8 Arabic table:

U+0644	ل	d9 84	ARABIC LETTER LAM
--------	---	-------	-------------------

U+0646	ن	d9 86	ARABIC LETTER NOON
--------	---	-------	--------------------

U+060C	,	d8 8c	ARABIC COMMA
--------	---	-------	--------------

Not sure if things are matching up, at least visually!

Ok just thought of one more thing to do, how about if I just put the Arabic text into a text file and then cat it and see if it appears here in Google docs...

```
$ vi englishAndArabic.go
```

```
$ cat > Arabic.txt
↳ Google.
```

```
$ cat Arabic.txt
↳ Google.
```

Hmmm did not work!

Well how about writing it to a file from Go...

```
$ cp -p englishAndArabic.go writeArabicToFile.go
```

```
$ vi writeArabicToFile.go
```

```
$ cat writeArabicToFile.go
package main
```

```
import (
    "fmt"
    "log"
    "os"
)

func main() {
    // From: https://zetcode.com/golang/writefile/

    f, err := os.Create("Arabic.txt")

    if err != nil {
        log.Fatal(err)
    }

    defer f.Close()

    aStr := "Google." لنجرب اللغة الانجليزية ونحوں هذه الجملة باستخدام ترجمة"
    _, err2 := f.WriteString(aStr)

    if err2 != nil {
        log.Fatal(err2)
    }
}
```

```
    }  
  
    fmt.Println("done")  
}  
  
$ /home/mark/_go/bin/go run writeArabicToFile.go
```

```
done
```

```
$ cat Arabic.txt  
لنجرب اللغة الإنجليزية ونحوه هذه الجملة باستخدام ترجمة Google.
```

Well the results look the same...

```
$ od -x Arabic.txt  
0000000 84d9 86d9 acd8 b1d8 a8d8 d820 d9a7 d984  
0000020 d884 d8ba 20a9 a7d8 84d9 a5d8 86d9 acd8  
0000040 84d9 8ad9 b2d8 8ad9 a9d8 d920 d988 d886  
0000060 d9ad d988 2084 87d9 b0d8 87d9 d820 d9a7  
0000100 d884 d9ac d985 d884 20a9 a8d8 a7d8 b3d8  
0000120 aad8 aed8 afd8 a7d8 85d9 d820 d8aa d8b1  
0000140 d9ac d885 20a9 6f47 676f 656c 002e  
0000155
```

So I went to big endian and the order looks correct and what I thought was boundary shifting has to do with spaces (hex 20) in the Arabic sentence which requires two bytes and seems to shift the boundary in the od output that groups 16 bits.

```
$ od -x --endian=big Arabic.txt  
0000000 d984 d986 d8ac d8b1 d8a8 20d8 a7d9 84d9  
0000020 84d8 bad8 a920 d8a7 d984 d8a5 d986 d8ac  
0000040 d984 d98a d8b2 d98a d8a9 20d9 88d9 86d8  
0000060 add9 88d9 8420 d987 d8b0 d987 20d8 a7d9  
0000100 84d8 acd9 85d9 84d8 a920 d8a8 d8a7 d8b3  
0000120 d8aa d8ae d8af d8a7 d985 20d8 aad8 b1d8  
0000140 acd9 85d8 a920 476f 6f67 6c65 2e00  
0000155
```

```
$
```

That's it for now as I have to do all the exercises/projects I'm behind on...