

Juliana McCausland

Ling 473

Project 3

Results:

คู่แข่ง ชัน ต่าง ก็ คุ่ม เขิง กัน

เขา เียงบ ไปครู่ หนึ่ง แล้ว พุด ขึ้น

เธอ หัน มา คู้ย ทราย ขึ้น มา ใหม่

ยิน ดี ที่ ได้ รู้ จัก คุณ

ฉัน เห็น เขา ตาม เธอ แจ ไปทุก ที่ เป็น เมา ตาม ตัว ตลอด เลย

นก เล็ก มี หน้า ออก สี แดง

...

...

Description:

This project was both challenging and simple. The program opens the utf8 file containing the Thai text using the codecs tool, creates an output file, and iterates through the characters of the file while making transitions to the appropriate states (using the enumerate function so that each character for each line is assigned a number value). States 7 through 9 use a dictionary to access previous characters and create white space where necessary.

States 0 through 6 were fairly straightforward, although they required many transitions, which caused a bit of confusion. The states in which syllable breaks were detected (7 through 9) required more work to figure out.

My code for states 7 through 9 is likely more complicated than it needs to be. I struggled with keeping track of numbers and transitions. I also learned a little more about dictionaries (I always manage to fit dictionaries into my programs, but I hadn't realized their full potential until now). The most challenging part of this project was working with unfamiliar characters, as it required greater attention to detail.

I also struggled with opening the file and writing the output. After some research, I discovered the codecs tool, but I am still unsure of how exactly it works.