

Problem1 Word Tokenization

โจทย์ต้องการส่งข้อความหาแฟนสาวคนสวยชาวต่างประเทศของเขา เขาจึงมีความจำเป็นต้องพิมพ์ภาษาอังกฤษ แต่ปุ่ม space ดันกดไม่ได้ ทำให้โจทย์ต้องพิมพ์ตัวอักษรติดกัน โจทย์ต้องการโปรแกรมสำหรับการตัดคำเพื่อแก้ปัญหานี้

โจทย์ข้อนี้ต้องการเขียนโปรแกรมตัดคำเพื่อช่วยโจโดยแบ่งคำแบบ longest matching ซึ่งจะตัดคำตามคำที่ยาวที่สุดที่พบใน dictionary ตามลำดับ

Longest matching

1. เริ่มต้นไล่ดูที่อักขระตัวแรกจนถึงอักขระตัวสุดท้าย หาชุดอักขระที่ตรงกับคำที่อยู่ใน dictionary
2. ให้เลือกตัดตามคำที่ยาวที่สุดที่พบใน dictionary ออกจาก string ตั้งต้น
3. ใช้ string ที่เหลือ วนกลับไปทำข้อ 1 จนกว่าจะแบ่งไม่ได้แล้ว

**** หากไม่พบใน dictionary ให้ตัดรวมอักขระทั้งหมดที่แบ่งไม่ได้เป็นคำเดียวกัน ****

ตัวอย่างวิธีการตัดคำ

ข้อความที่ยังไม่ได้แบ่งคำ : “tappleisafod”

dictionary : app, apple, a, at

output : “t apple is a food”

เริ่มดูตั้งแต่อักขระตัวแรกจาก t ไปจนถึงตัวสุดท้ายคือ d ซึ่งไม่พบคำใน dictionary ที่ขึ้นต้นด้วย t จึงแบ่ง t

ออกมาจากข้อความตั้งต้น จากนั้นจึงเริ่มที่อักขระตัวถัดไปจาก a จนถึง d ตัวสุดท้าย จะได้ชุดของอักขระ a, app, apple

ที่ตรงกับคำที่อยู่ใน dictionary ให้เลือกแบ่งเป็นคำที่ยาวที่สุดก็จะได้เป็น apple แบ่งออกมาจากข้อความตั้งต้น

แล้วทำต่อโดยเริ่มที่อักขระตัวถัดไปคือ i ซึ่งไม่พบคำใน dictionary ที่ขึ้นต้นด้วย i จึงแบ่งออกไปเป็น i ถัดมาเริ่มที่ s

ก็ไม่พบคำที่อยู่ใน dictionary ที่ขึ้นต้นด้วย s เช่นกัน ก็ให้รวม i กับ s เข้าด้วยกันกลายเป็น is ตัว a ถัดมามีอยู่ใน dictionary และ

คำที่ยาวที่สุดแบ่งได้คือ a จึงแบ่ง a ออกมา สุดท้าย food อยู่ใน dictionary เลยกกลายเป็นคำเดียว

ข้อมูลนำเข้า

ข้อมูลนำเข้าจะรับจากแป้นพิมพ์ที่ละบรรทัดเริ่มต้นด้วยข้อความที่ยังไม่ได้แบ่งคำ ต่อมาเป็นจำนวนคำใน dictionary และตามด้วยคำต่างๆตามจำนวนที่ระบุ เช่น

tappleisacat # ข้อความที่ยังไม่ได้แบ่งคำ

4 # จำนวนคำใน dictionary

app

apple

a

****หมายเหตุ** ข้อความนำเข้าเป็นตัวพิมพ์เล็กทั้งหมด

ข้อมูลส่งออก

ข้อความผลลัพธ์หลังการแบ่งคำที่มี space คั่นระหว่างคำ เช่น “apple is a cat”

input (จากแป้นพิมพ์)	output (ทางจอภาพ)
Thisisadog 8 these this is a and his sad dog	this is a dog
thisisscotchandhisdog 8 these this is a and his sad dog	this is scotch and his dog
thisiscocodog 2 dog scotch	thisiscoco dog