

Programming Exercise 2: โครงสร้างข้อมูล: ต้นไม้

(โอลิมปิกวิชาการ ค่ายสอง ม. ศิลปากร 20 มีนาคม 2556, โดย ดร.ภิญโญ แท้ประสาทสิทธิ์)

Problem 2: ค้นหาคำยอดนิยม

Find Top-Hit Words

(ส่งคำตอบมาในไฟล์ชื่อ top_hit_word.*)

เราสามารถประยุกต์ใช้ทฤษฎีในการวิเคราะห์ข้อความ (text analysis) ทั้งในเอกสารทั่วไปและเอกสารบนเว็บได้ และหนึ่งในการประยุกต์ใช้ที่สำคัญก็คือการหาคำที่มีการใช้บ่อยที่สุดในเอกสาร เราต้องการทราบคำที่ใช้บ่อยที่สุด (คำยอดนิยม) เพราะมันสามารถบอกเราได้ว่าเอกสารนั้นมีเนื้อหาเกี่ยวกับอะไร

Task

ให้ขยายและดัดแปลงโครงสร้างข้อมูลและฟังก์ชันของทรีที่มีอยู่เดิมให้สามารถนับจำนวนคำที่ซ้ำกันได้ โดยรูปแบบของอินพุตจะเป็นคำสั่งในลักษณะเดิม แต่จะมีเฉพาะการใส่คำเพิ่มเข้าไปเท่านั้น ไม่มีคำสั่ง F Q และ P (ดูตัวอย่างประกอบหน้าถัดไป)

ยกตัวอย่าง เช่น

I Cat

I Cat

ในแบบฝึกหัดอันเดิมการใส่ Cat เข้าไปครั้งที่สองจะไม่เกิดการเปลี่ยนแปลงขึ้น แต่ในแบบฝึกหัดนี้เราจะต้องบันทึกไว้ด้วยว่าคำศัพท์แต่ละคำถูกใส่เข้ามาก็ครั้ง

เมื่อจัดการอินพุตเสร็จแล้ว ให้ทำการประมวลผลตรวจสอบว่าคำที่ซ้ำกันบ่อยที่สุดซ้ำกันกี่ครั้ง มีกี่ตัวที่ซ้ำกันบ่อยที่สุด (เพราะคำยอดนิยมอาจมีมากกว่าหนึ่งตัวได้) และให้รายงานด้วยว่ามีคำยอดนิยมอะไรบ้าง ในกรณีที่คำยอดนิยมมีมากกว่าหนึ่งตัวให้แสดงออกมาให้หมดทุกคำ คำละบรรทัด ดังตัวอย่างหน้าถัดไป

ตัวอย่าง input และ output (input_5.txt)

Input	Output
22	4
I iPod	2
I iPad	iPad
I iTune	iPhone
I iPhone	
I Apple	
I XBox	
I Wii	
I PlayStation	
I Microsoft	
I Zony	
I Nokia	
I Nintendo	
I Zony	
I Nintendo	
I iPhone	
I iPad	
I iPhone	
I iPad	
I iPhone	
I iPad	
I XBox	
I Wii	