

Levenshtein edit distance

Levenshtein edit distance เป็นขั้นตอนวิธีการวัดหาค่าความต่างกันของสายอักขระสองชุด ระหว่างชุดแรกที่เป็นต้นแบบ และ ชุดที่สองที่เป็นชุดเปรียบเทียบ โดยค่าความต่างกันจะวัดจากจำนวนของการที่จะต้องทำการตัดออก แทรก และแทนที่อักขระในชุดที่นำมาเปรียบเทียบจนกระทั่งมีลักษณะเหมือนชุดอักขระที่เป็นต้นแบบทุกประการ

วิธีการคือนำชุดอักขระ 2 ชุด มาเปรียบเทียบจำนวนความแตกต่างกัน โดยจะพิจารณาดังนี้

การแทรก เป็นการนำเอาอักขระตัวใด ๆ มา เพื่อให้ชุดอักขระชุดนั้นเหมือนกับอีกชุดอักขระหนึ่งในภายหลัง เช่น run → ruin จะแทรกตัว i ให้กับ run เพื่อให้ run กลายเป็น ruin เป็นต้น

การตัดออก เป็นการตัดอักขระออกครั้งละ 1 ตัว จากชุดอักขระตัวหนึ่ง เพื่อให้ชุดอักขระชุดนั้นเหมือนกับอีกชุดอักขระหนึ่งในภายหลัง เช่น dog → do จะตัดตัว g ออก เพื่อให้ dog กลายเป็น do เป็นต้น

การแทนที่ เป็นการนำอักขระของชุดอักขระหนึ่งไปแทนอักขระของอีกชุดอักขระหนึ่ง เพื่อให้ชุดอักขระชุดนั้นเหมือนกับอีกชุดอักขระหนึ่งในภายหลัง เช่น cat → rat จะแทนที่ r ด้วย c เพื่อให้ cat กลายเป็น rat หรือ อาจมองในทางกลับกันก็ได้ เป็นต้น

ทั้งนี้ให้ค่าใช้จ่ายในการเรียก การแทรก การตัดออก การแทนที่ 1 ครั้งคิดเป็น 1 หน่วย

ดังนั้น Edit Distance คือ ต้นทุนในการแก้ไข string ต้นฉบับเพื่อให้ได้ string ปลายทาง หน้าที่ของเราคือ หา Minimum Edit Distance คือทำอย่างไรให้ใช้ต้นทุนน้อยที่สุด

ข้อมูลนำเข้า

มี 2 บรรทัด บรรทัดแรกเป็นข้อความแรก บรรทัดที่สองเป็นข้อความที่สอง แต่ละข้อความยาวไม่เกิน 100 ตัวอักขระ

ข้อมูลส่งออก

มี 1 บรรทัดแทนค่าใช้จ่ายในการเปลี่ยนข้อความแรกเป็นข้อความที่สองที่น้อยที่สุด

ตัวอย่าง

ข้อมูลนำเข้า	ข้อมูลส่งออก
cat rat	1
method algorithm	8