

LAB 8 : Text Clustering & Topic Modeling

วัตถุประสงค์

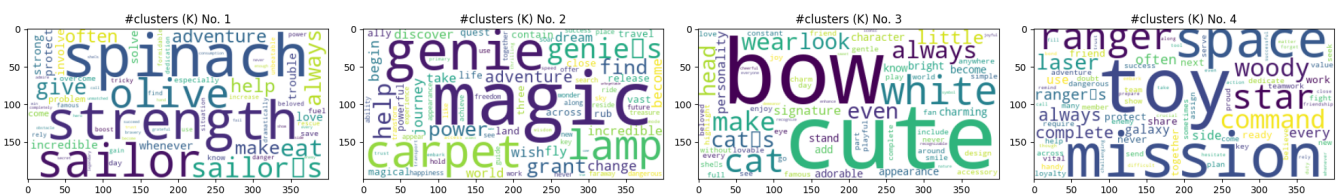
นศ. สามารถเขียนโปรแกรมด้วยภาษา Python เพื่อ Text Clustering & Topic Modeling

* ไม่นอนุญาตให้ส่งงานหลังกำหนด และนศ.ต้องตรวจสอบสถานะการส่งงาน หลัง submit งานทุกครั้ง

คำสั่ง

กำหนดให้ใช้ไฟล์ข้อมูลชื่อ UnknownTopics1.csv สำหรับนักศึกษาที่มีรหัสลงท้ายด้วยเลขคี่ และไฟล์ข้อมูลชื่อ UnknownTopics2.csv สำหรับนักศึกษาที่มีรหัสลงท้ายด้วยเลขคู่

- กำหนดให้ใช้ไฟล์ตัวอย่าง ที่ชื่อว่า Lab8_1_TextClustering.ipynb มาปรับแก้ และให้ทำการหาจำนวน Cluster ที่เหมาะสมสำหรับไฟล์ข้อมูล จากนั้นให้ทำการ Capture ผลลัพธ์การแสดงผล Word Cloud ที่แสดงการจัดกลุ่มข้อมูลตามจำนวน Cluster ที่ท่านเห็นว่าแสดงการจัดกลุ่มข้อมูลที่ดีและเหมาะสมที่สุด มาแสดงในช่องว่างที่เว้นไว้



- กำหนดให้ใช้ไฟล์ตัวอย่าง ที่ชื่อว่า Lab8_2_Topic_modeling.ipynb มาปรับแก้ และให้ทำการหาจำนวน Topic ที่เหมาะสมสำหรับไฟล์ข้อมูล จากนั้นให้ทำการ Capture ผลลัพธ์การแสดงผล Intertopic Distance Map ที่แสดงด้วย pyLDAvis โดยให้นำผลลัพธ์มาแสดงในทุกๆ หัวข้อตามจำนวน Topic ที่ท่านเห็นว่าแสดงการจัดกลุ่มข้อมูลที่ดีและเหมาะสมที่สุด มาแสดงในช่องว่างที่เว้นไว้

