

ToC Assignment [20 คะแนน]

โจทย์ : ให้แต่ละกลุ่มสร้าง Web application ที่ทำหน้าที่เป็น Web crawler โดยใช้ Regular expression library ของ Python (import re) ในการรวบรวมรายชื่อของหมวดหมู่ที่นักศึกษาสนใจ จำนวน 1 หมวด (แต่ละกลุ่มห้ามซ้ำหมวดกัน) เช่น นักแสดงไทย, นักร้องเกาหลี, อาหารนานาชาติ เป็นต้น โดยจำนวนชื่อในหมวดที่เลือก ต้องไม่ต่ำกว่า 200 ชื่อ สามารถ Crawl มาจากหลายแหล่งได้ เช่น Wikipedia, IMDb เป็นต้น และต้องใช้ regular expression ในการ extract รายชื่อ

การจัดกลุ่ม ให้แต่ละกลุ่มจับกลุ่มเองประมาณ 10-11 คนต่อกลุ่ม โดยกรอกข้อมูลในลิงค์

https://docs.google.com/spreadsheets/d/1RQH20YCWMj7UCK-LHIGfNHpfQjb-fxDNZl8dsvRNdMw/edit?usp=share_link

Output คือ แสดงรายชื่อของสิ่งที่อยู่หมวดที่แต่ละกลุ่มเลือก เช่น นักแสดงไทย ให้แสดงเฉพาะชื่อของนักแสดงไทยทั้งหมดที่หน้าเว็บแอป และ สามารถ export รายชื่อทั้งหมดที่ได้เป็นไฟล์ csv ในรูปแบบหนึ่งชื่อต่อหนึ่งบรรทัดได้ (ควรมีความลึกของการ Crawl มากกว่า 1 ระดับขึ้นไป)

กำหนดส่ง วันอังคารที่ 8 ตุลาคม 2567 ก่อนเวลา 9.00 น. (ลิงค์ส่ง Assignment อยู่ในไฟล์ Course syllabus)

สิ่งที่ต้องส่ง

1. ลิงค์ไปยังคลิปบน YouTube พิธีแต่งงานไม่เกิน 10 นาที โดยเน้นที่วิธีการ **crawl** และ การประยุกต์ใช้ **regular expression**
2. ลิงค์ไปยังหน้าเว็บแอปฯ โดยที่หน้าเว็บจะต้องมีลิงค์ไปยัง **Source code**

เกณฑ์การให้คะแนน

1. แหล่งการเข้าถึงข้อมูลของหมวดหมู่ที่เลือกควรมีความลึกของการ **Crawl** มากกว่า 1 ระดับ
2. ความซับซ้อนของ **Regular Expression** ควรเหมาะสมกับปัญหา และไม่ควรง่ายจนเกินไป
3. ความถูกต้องของข้อมูลที่ดึงมาได้
4. การออกแบบ Web Application