

โปรแกรมเก็บข้อมูล Facebook fan page จากหน้าเว็บ

Facebook fan page scraper from web page

นายคมสัน นิภารัตน์ รหัส 5952100228

นายสุทิวัฒน์ แสงสุวรรณ รหัส 5952100317

โครงงานนี้เป็นส่วนหนึ่งของ โครงงานเทคโนโลยีสารสนเทศ 2 (IT 59306)

สาขาเทคโนโลยีสารสนเทศ

คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์

ปีการศึกษา 2561



โปรแกรมเก็บข้อมูล Facebook fan page จากหน้าเว็บ

Facebook fan page scraper from web page

นายคมสัน นิภารัตน์ รหัส 5952100228

นายสุทิวัฒน์ แสงสุวรรณ รหัส 5952100317

โครงงานนี้เป็นส่วนหนึ่งของ โครงงานเทคโนโลยีสารสนเทศ 2 (IT 59306)

สาขาเทคโนโลยีสารสนเทศ

คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์

ปีการศึกษา 2561

**คำนำ**

โครงงานเรื่องโปรแกรมเก็บข้อมูล Facebook fan page จากหน้าเว็บ จัดทำขึ้นเพื่อช่วยให้ผู้ที่ต้องการข้อมูลจาก Facebook fan page ที่มีข้อมูลจำนวนมาก สามารถนำข้อมูลที่ได้จากโครงงานไปใช้ในการทำประโยชน์ทั้งด้านวิเคราะห์ข้อมูลรวมถึงช่วยในการตัดสินใจในด้านต่างๆ ทั้งภาคการศึกษาและภาคธุรกิจ

คณะผู้จัดทำขอขอบพระคุณอาจารย์ในคณะวิศวกรรมศาสตร์และเทคโนโลยีรวมถึงผู้ที่เกี่ยวข้องทุกท่านที่ให้การสนับสนุนจนกระทั่งโครงงานฉบับนี้เสร็จสมบูรณ์

คณะผู้จัดทำหวังว่าโครงงานฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจในและผู้ที่ต้องการนำข้อมูลจาก Facebook fan page ไปใช้ในการวิเคราะห์ข้อมูลต่อไป

คณะผู้จัดทำ

มีนาคม 2562

**สารบัญ**

**หน้า**

[บทที่ 1 บทนำ 1](#_Toc8888884)

[1.1 ความเป็นมาและความสำคัญของโครงงาน 1](#_Toc8888885)

[1.2 วัตถุประสงค์ของโครงงาน 2](#_Toc8888886)

[1.6 เครื่องมือที่ใช้ในการดำเนินงาน 4](#_Toc8888887)

[บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 6](#_Toc8888888)

[2.1 ทฤษฏีที่เกี่ยวข้อง 6](#_Toc8888889)

[2.1.1 Web Mining 6](#_Toc8888890)

[2.1.2 An Architecture for Web Mining 6](#_Toc8888891)

[2.1.3 Web Scraping, Web Data Extraction 7](#_Toc8888892)

[2.2 เทคโนโลยีที่เกี่ยวข้อง 8](#_Toc8888893)

[2.2.1 Selenium 8](#_Toc8888894)

[2.2.3 JavaScript 9](#_Toc8888895)

[2.2.4 Sublime 9](#_Toc8888896)

[2.2.5 Visual Studio 9](#_Toc8888897)

[2.3 เอกสารและงานวิจัยที่เกี่ยวข้อง 10](#_Toc8888898)

[2.3.1 ขั้นตอนและวิธีการสลัดข้อมูลสินค้าบนเว็บเพจสำหรับเว็บครอเลอร์ที่ใช้ในโปรแกรมค้นหาสินค้า 10](#_Toc8888899)

[2.3.2 ระบบติดตามการคัดลอกเนื้อหาเว็บอัตโนมัติโดยใช้วิธีการเลือกข้อความสำคัญ 11](#_Toc8888900)

[บทที่ 3 วิธีการดำเนินการ 13](#_Toc8888901)

[3.1 การวิเคราะห์ออกแบบโครงงาน และลักษณะโครงสร้างของโครงงาน 13](#_Toc8888902)

[3.2 การวิเคราะห์ปัญหาที่อาจเกิดขึ้น และขั้นตอน 14](#_Toc8888903)

[3.3 กระบวนการในการตรวจหาข้อมูล, คำ, ประโยค หรือภาพ 15](#_Toc8888904)

[3.4 การทดลองโครงงาน 19](#_Toc8888905)

[3.5 ผลลัพธ์จากการตรวจหาข้อมูลตรงตามที่ต้องการและการจัดเก็บข้อมูล 20](#_Toc8888906)

[บทที่ 4 ผลการดำเนินงาน 21](#_Toc8888907)

[4.1 ส่วนของโปรแกรมขั้นตอนการ Login ของตัวโปรแกรม 21](#_Toc8888908)

[4.2 ผลการวิจัยในการดึงข้อมูลจาก Facebook Fan page 23](#_Toc8888909)

[บทที่ 5 สรุปผลและข้อเสนอแนะ 27](#_Toc8888910)

[5.1 สรุปผลการทดลอง 27](#_Toc8888911)

[5.2 ข้อเสนอแนะ 27](#_Toc8888912)

**สารบัญภาพ**

[รูปที่ 2.1: General Architecture for WEBMINER 7](#_Toc3765799)

[รูปที่ 2.2: Sample Entries from a Web Server Access Log 7](#_Toc3765800)

[รูปที่ 2.3: ตัวอย่างโหนดในภาษา HTML กับโหนดสินค้า 1 โหนด 1 สินค้า 10](#_Toc3765801)

[รูปที่ 2.4: แสดงการสกัดข้อความบนเพจภาพ 11](#_Toc3765802)

[รูปที่ 2.5: แสดงการใช้งานของ Syntax 12](#_Toc3765802)

[รูปที่ 3.1: แสดงการใช้งานของ Syntax 12](#_Toc3765802)

[รูปที่ 3.1**:** Architecture 14](#_Toc3765803)

[รูปที่ 3.2: sequence diagram 15](#_Toc3765804)

[รูปที่ 3.3: Facebook Fan page ต้นแบบ 15](#_Toc3765805)

[รูปที่ 3.4: Facebook Fan page โพส 16](#_Toc3765806)

[รูปที่ 3.5: Login Facebook Fan page 16](#_Toc3765807)

[รูปที่ 3.6: TAG HTML 16](#_Toc3765808)

[รูปที่ 3.7: TOKEN 16](#_Toc3765808)

[รูปที่ 3.8: Post Text 17](#_Toc3765809)

[รูปที่ 3.9: Post Picture 17](#_Toc3765810)

[รูปที่ 3.10: Getting Data Fb Fanpage 18](#_Toc3765811)

[รูปที่ 3.11: Flow Chart 19](#_Toc3765812)

[รูปที่ 3.12: DATA ที่ได้และการเปรียบเทียบความถูกต้อง 20](#_Toc3765813)

[รูปที่ 3.12: DATA ที่ได้และการเปรียบเทียบความถูกต้อง 20](#_Toc3765813)

[รูปที่ 4.1: หน้า Login เริ่มต้นการทำงาน 21](#_Toc3765813)

[รูปที่ 4.2: ตัวอย่างการกำหนดเว็บเพจที่ต้องการจะดึงข้อมูล 21](#_Toc3765813)

[รูปที่ 4.3: ตัวอย่างการใส่ Token ยืนยันตัวตน 22](#_Toc3765813)

[รูปที่ 4.4: การทำงานของโปรแกรมหลังจากตั้งค่าและทำการ Login 22](#_Toc3765813)

[รูปที่ 4.5: รูปแบบการแสดงค่าของการทำงานของโปรแกรม 22](#_Toc3765813)

[รูปที่ 4.6: การจัดเก็บข้อมูลเพจเป้าหมาย 23](#_Toc3765813)

[รูปที่ 4.7: ผลลัพธ์ของโปรแกรม 26](#_Toc3765813)

# บทที่ 1 บทนำ

## 1.1 ความเป็นมาและความสำคัญของโครงงาน

เนื่องจากปัจจุบันข้อมูลมีความสำคัญในด้านการประมวลผลและภาคธุรกิจโดยเฉพาะอย่างยิ่งถ้าเป็นข้อมูลขนาดใหญ่หรือข้อมูลปริมาณมากและมีการเปลี่ยนแปลงตลอดเวลา ในทางเทคนิคข้อมูลประเภทต่างๆ อาทิเช่น ข้อมูลส่วนบุคคล ข้อมูลการทำธุรกรรม ข้อมูลเว็บ เป็นข้อมูล Big Data การรวบรวมข้อมูลดังกล่าวจากการวิเคราะห์ การศึกษา และการวิจัย จะช่วยในการตัดสินใจในเรื่องต่างๆ ทำให้สามารถสร้างความได้เปรียบในเชิงภาคธุรกิจ เฟสบุ๊ค (Facebook) เป็นเว็บไซด์สังคมออนไลน์ (Social Media) ที่มีคนไทยเล่นเป็นจำนวนมาก ส่งผลให้มีข้อมูลของผู้ใช้ในโลกออนไลน์นั้นมีจำนวนมากเช่นกัน ข้อมูลที่ Facebook มีนั้น สำหรับบุคคลทั่วไปอาจเป็นเพียงข้อมูลทั่วไป แต่สำหรับธุรกิจที่ต้องการทำการค้าขายนั้นมองข้อมูลนี้เป็นสิ่งมีค่า โดยหากภาคธุรกิจทำการรวบรวมการสืบค้นต่างๆ เช่น การนำข้อมูลความสนใจในสิ่งต่างๆของผู้ใช้ใน Facebook มาวิเคราะห์ทางสถิติเพื่อให้ธุรกิจสามารถขายสินค้าและเสนอสินค้าได้ตรงตามความต้องการของผู้บริโภคที่กำลังหาสินค้าที่กำลังสนใจอยู่ได้ ข้อมูลในปี 2017 มีการประมาณเว็บไซต์ที่เพิ่มขึ้นขนาดกว่า 1,766 ล้านเว็บไซต์ เมื่อเทียบกับปี 2016 ซึ่งเพิ่มเป็น 69% และยังมีจำนวนผู้ใช้ Social Media มีจำนวนเพิ่มขึ้น 3,196 ล้านคน (ณรงค์ศักดิ์ ปลอดมีชัย, 2561) ดังนั้นการนำข้อมูลจาก Facebook ที่มีปริมาณมากจำเป็นต้องใช้โปรแกรมช่วยเหลือในการรวบรวมข้อมูล จากการศึกษาโปรแกรมเก็บรวมรวมข้อมูลจาก Facebook ที่มีอยู่ในปัจจุบัน ไม่มีความสามารถพอที่จะเก็บข้อมูลได้หลังจากการแสดงผลของ Facebook เปลี่ยนไป ทำให้โปรแกรมดังกล่าวไม่สามารถใช้งานได้ ส่งผลให้ขาดความต่อเนื่องในการใช้งานของโปรแกรม ซึ่งการปรับปรุงโปรแกรมจำเป็นต้องมีค่าใช้จ่ายในการปรับปรุง

การแก้ไขในปัจจุบันที่เป็นไปได้ในส่วนของโปรแกรมคือการปรับปรุงโปรแกรมดังกล่าวให้พัฒนาตามการเปลี่ยนแปลงของการแสดงผล เพื่อให้โปรแกรมสามารถใช้งานได้หรือการทำให้โปรแกรมสามารถเข้าใจถึงข้อมูลในส่วนต่างๆที่เปลี่ยนแปลงไปได้อย่างถูกต้องโดยไม่ต้องปรับปรุง

ผู้จัดทำจึงมีความสนใจที่จะพัฒนาโปรแกรมเก็บรวบรวมข้อมูลจาก Facebook โดยโปรแกรมสามารถรองรับการเปลี่ยนแปลงการแสดงผลของ Facebook ได้โดยไม่จำเป็นต้องปรับปรุงโปรแกรม โปรแกรมจะมุ่งเน้นที่การเก็บรวมรวมข้อมูลจาก Facebook fan page เนื่องจากมีข้อมูลของผู้บริโภคเฉพาะกลุ่มเป็นจำนวนมาก อาทิ Facebook fan page ร้านขายกระเป๋าจะมีข้อมูลของปริมาณลูกค้าที่ชื่นชอบกระเป๋าในรูปแบบต่างๆ เป็นต้น โดยโปรแกรมจะเรียนรู้รูปแบบการจัดเรียงของ Tag HTML จาก fan page ต้นแบบ และนำไปใช้กับ fan page เป้าหมายที่ต้องการเก็บรวบรวมข้อมูล

หลังจากโปรแกรมเก็บรวบรวมข้อมูลจาก Facebook fan page ทำงาน โปรแกรมจะสามารถจัดเก็บข้อมูลจาก fan page เป้าหมายที่ผู้ใช้ต้องการในรูปแบบให้พร้อมใช้เป็นไฟล์ออฟไลน์ เพื่อสำหรับการประมวลผลเชิงสถิติต่อไป

## 1.2 วัตถุประสงค์ของโครงงาน

เพื่อพัฒนาโปรแกรมเก็บรวบรวมข้อมูลโดยอัตโนมัติจากโพสของ Facebook fan page ที่ผู้ใช้ต้องการแม้มีการเปลี่ยนแปลงการแสดงผลของหน้าเว็บเล็กน้อย ประกอบด้วยข้อความ รูปภาพ จำนวนการกดถูกใจ วัน เดือนที่โพสและจำนวนความคิดเห็น

**1.3 ขอบเขตของโครงงาน**

1.3.1 โปรแกรมสามารถเก็บรวบรวมข้อมูลโดยอัตโนมัติที่ประกอบด้วย ข้อความ รูปภาพ จำนวนการกดถูกใจ วัน เดือนที่โพสและจำนวนความคิดเห็น และบันทึกเป็นไฟล์ที่ประกอบด้วยข้อมูลดังกล่าวเป็นไฟล์นามสกุล csv โดยเริ่มเก็บรวบรวมจากวันที่โพสเรียงจากใหม่ไปเก่า

1.3.2 โปรแกรมสามารถรองรับการเปลี่ยนแปลงที่เกิดจากการพัฒนาส่วนติดต่อผู้ใช้ของ Facebook ได้ หากโพสในแต่ละโพสมีข้อมูลคุณลักษณะที่ยังคงอยู่ใน Node HTML เดียวกัน

1.3.3 โปรแกรมต้องการการป้อนข้อมูลจากผู้ใช้ที่ประกอบไปด้วย ชื่อผู้ใช้ รหัสผ่าน โทเคนของ Facebook fan page และ URL Facebook fan page ของเป้าหมายที่ต้องการเก็บรวบรวมข้อมูล

1.3.4 โปรแกรมสามารถเก็บรวบรวมข้อมูลได้อย่างถูกต้องในทุกๆ ส่วน เช่น ไฟล์ที่ถูกบันทึกแสดงจำนวนการแบ่งปัน มีข้อมูลที่ตรงกันกับโพสใน Facebook fan page นั้นๆ

1.3.5 โปรแกรมคำนึงถึงความถูกต้องของข้อมูลที่ได้เก็บข้อมูลมาเท่านั้น โดยไม่นำปัจจัยอื่นที่เกี่ยวข้องกับการวัดประสิทธิภาพ เช่น ความเร็วของการเก็บข้อมูล

1.3.6 ระยะเวลาดำเนินงาน 11 ธันวาคม พ.ศ. 2561 ถึง 16 พฤษภาคม พ.ศ. 2562

**1.4 ประโยชน์ที่ได้รับ**

1.4.1 สามารถใช้โปรแกรมเพื่อทำการรวบรวมข้อมูลโดยอัตโนมัติจาก Facebook fan page ที่ต้องการได้ โดยประกอบไปด้วย ข้อความ รูปภาพ จำนวนการกดถูกใจ วัน เดือนที่โพสและจำนวนความคิดเห็น

1.4.2 นักพัฒนาโปรแกรมไม่จำเป็นต้องพัฒนาโปรแกรมเพื่อรองรับการพัฒนาส่วนติดต่อผู้ใช้ของ Facebook หากเกิดการเปลี่ยนแปลงเล็กน้อย

1.4.3 เพื่ออำนวยความสะดวกให้กับผู้ที่ต้องการใช้ข้อมูลหรือนักวิเคราะห์ข้อมูลจาก Facebook Fan page โดยไม่มีค่าใช้จ่าย

1.4.4 เพื่อส่งเสริมชุมชนโอเพ่นซอร์ส (Open Source Community) โดยมีการเผยแพร่คำสั่งของโปรแกรม (Source Code) บนเว็บไซต์ https://github.com/qaz74107410/fb-page-scraper

**1.5 แผนดำเนินการของโครงงาน**

1.5.1 โครงงานและนำเสนออาจารย์ที่ปรึกษา

1.5.2 ศึกษาทฤษฏีและโปรแกรมที่ใช้ในการจัดทำโครงงาน

1.5.3 ออกแบบโครงสร้างของระบบการเก็บข้อมูล Facebook fan page จากหน้าเว็บ

1.5.4 สร้างโปรแกรมการเก็บข้อมูล Facebook fan page จากหน้าเว็บ

1.5.5 ทดสอบโปรแกรม สรุปผลจากข้อมูลที่เก็บรวบรวมได้จากโปรแกรม

1.5.6 การประเมินผลและปรับปรุงระบบ

1.5.7 จัดทำเอกสารโครงงาน

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| กิจกรรมการดำเนินงาน | ระยะเวลาในการดำเนินงาน (สัปดาห์) | | | | | | | | | | | | | | | | | | | | | | | | |
| ธันวาคม 2561 | | | | มกราคม 2562 | | | | กุมภาพันธ์ 2562 | | | | มีนาคม 2562 | | | | เมษายน 2562 | | | | พฤษภาคม 2562 | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| คิดค้นโครงงานและนำเสนออาจารย์ที่ปรึกษา |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ศึกษาทฤษฏีและโปรแกรมที่ใช้ในการจัดทำโครงงาน |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ออกแบบโครงสร้างของระบบของโปรแกรมเก็บรวบรวมข้อมูล |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| สร้างโปรแกรม Software การเก็บข้อมูล |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| การทดสอบโปรแกรม และสรุปผลการทดสอบ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| การประเมินผลและปรับปรุงระบบ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| จัดทำเอกสารโครงงาน |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

ตารางที่ 1.1 แสดงแผนการดำเนินงาน

## 1.6 เครื่องมือที่ใช้ในการดำเนินงาน

1.6.1 โปรแกรมคอมพิวเตอร์ (Software)

- คอมพิวเตอร์แบบพกพา

- Visual Studio

- Sublime

1.6.2 สิ่งแวดล้อมสำหรับการพัฒนา (Development Environment)

python

- selenium

- requests

- getpass

javascript

- jQuery

1.6.3 อุปกรณ์คอมพิวเตอร์ (Hardware)

Time of this report: 1/22/2019, 21:48:44

Machine name: LAPTOP-MNH7RRS4

Machine Id: {45FB1CB5-B0D2-4FFA-BC84-849A3246CD07}

Operating System: Windows 10 Home Single Language 64-bit (10.0, Build 17134) (17134.rs4\_release.180410-1804)

Language: English (Regional Setting: English)

System Manufacturer: ASUSTeK COMPUTER INC.

System Model: VivoBook S14 X411UF

BIOS: X411UF.302 (type: UEFI)

Processor: Intel(R) Core(TM) i3-8130U CPU @ 2.20GHz (4 CPUs), ~2.2GHz Memory: 8192MB RAM Available OS Memory: 8072MB RAM Page File: 15118MB used, 4216MB available Windows Dir: C:\WINDOWS DirectX Version: DirectX 12 DX

Setup Parameters: Not found

User DPI Setting: 96 DPI (100 percent)

System DPI Setting: 144 DPI (150 percent)

DWM DPI Scaling: Disabled Miracast: Available,

HDCP Microsoft Graphics Hybrid: Supported DxDiag Version: 10.00.17134.0001 64bit Unicode

**1.7 นิยามศัพท์เฉพาะ (Definitions)**

1.7.1 Browser (เบราเซอร์) คือ โปรแกรมดูเว็บไซต์ที่สามารถใช้ค้นหา ใช้ดูไฟล์ต่างๆ และสามารถเป็นเครื่องมือที่สามารถใช้งานได้หลากหลาย แต่ตอนนี้มีจำนวน Browser ให้เลือกใช้เพราะมีนักพัฒนามากมายเข้ามาเห็นความสำคัญ เช่น Internet Explorer, Firefox, Google Chrome, Opera, Safari

1.7.2Web Mining (เว็บไมนิ่ง) คือ กระบวนการทำงานโดยใช้ Data Mining (ดาต้าไมนิ่ง)และ algorithms (อัลกอริธึม) ในการดึงข้อมูลโดยตรงจากเว็บโดยแยกออกจากเว็บออนไลน์ เพื่อให้ความสนใจกับเนื้อหาเว็บไซต์

# บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิจัยข้อมูลรายละเอียดและองค์ประกอบการเก็บข้อมูลจากหน้าเว็บไซต์ซึ่งผู้จัดทำโครงงานได้ทำการทบทวนแนวคิด ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง ดังนี้

## 2.1 ทฤษฏีที่เกี่ยวข้อง

### 2.1.1 Web Mining

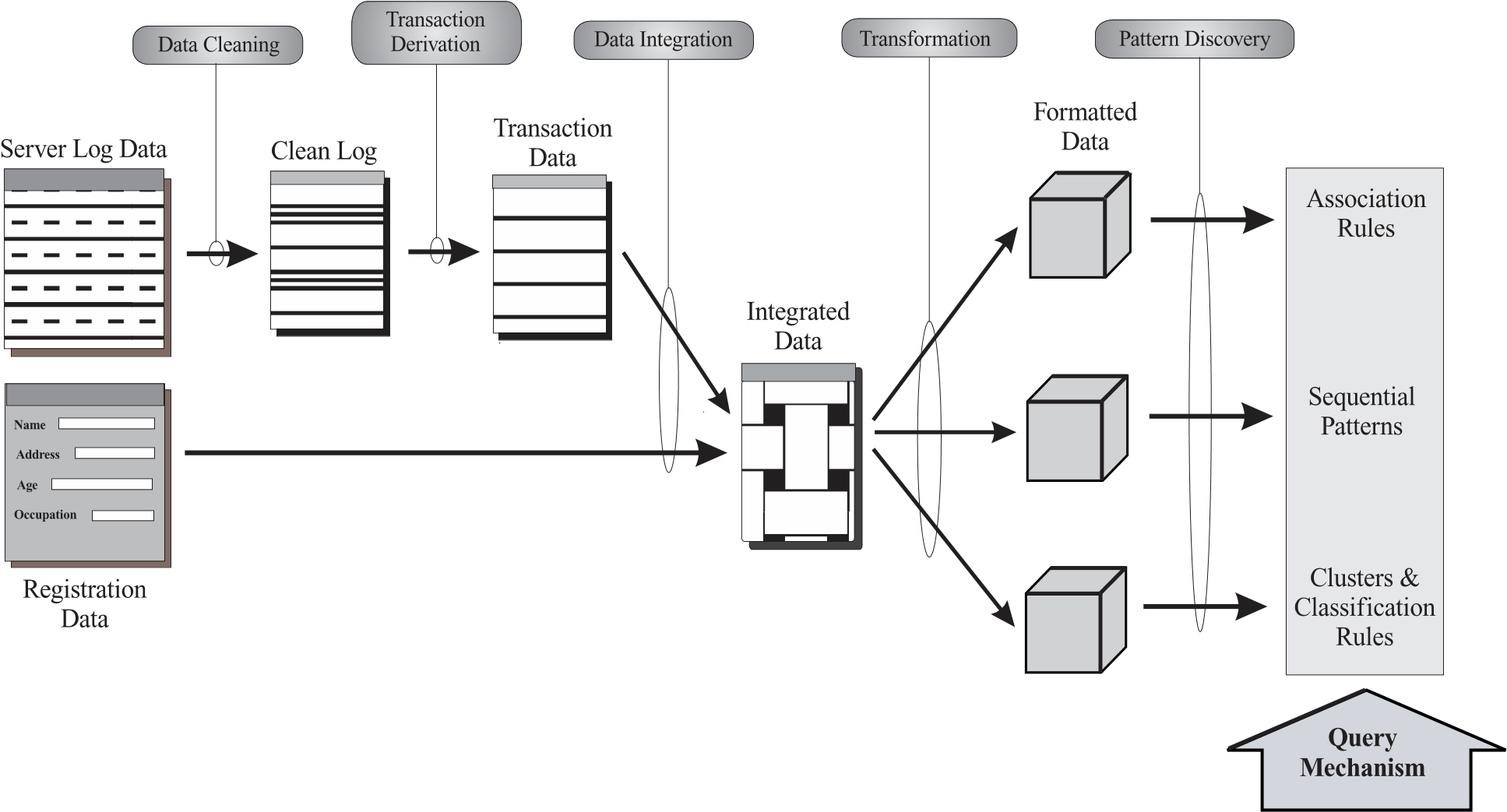
ธนวิชญ์ บุตรโคตร (2018) ได้ให้ข้อมูลและเทคนิควิธีการทำ Web Mining ดังนี้

การทำเหมืองข้อมูลคือ การใช้เทคนิคของการทำเหมืองข้อมูลเพื่อค้นหาและกรองข้อมูลสารสนเทศจากเอกสารบนเว็บและบริการบนเว็บโดยอัตโนมัติและนำข้อมูลมาทำสรุปและจัดเก็บข้อมูลในรูปแบบเอกสาร

การทำเหมืองเว็บคือ การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลกับเวิลด์ไวด์เว็บเรียกว่าการขุดเว็บเป็นจุดสนใจของโครงการวิจัยและงานวิจัยล่าสุดหลายฉบับ อย่างไรก็ตามไม่มีคำศัพท์ที่จัดตั้งอย่างเป็นทางการ แต่หากเปรียบเทียบกับการวิจัยหลายๆแห่งถูกใช้คำว่า Web Mining การทำเหมืองจากเนื้อหาบนเว็บไซต์เป็นกระบวนการของการค้นพบข้อมูลจากแหล่งต่างๆทั่วโลก รวมถึงเป็นกระบวนการของการขุดหารูปแบบการเรียกค้นของผู้ใช้และการเข้าถึง การกำหนด Web mining และนำเสนอภาพรวมของประเด็นการวิจัยเทคนิคและความพยายามในการพัฒนาที่หลากหลาย ระบบสำหรับการขุดการใช้งานเว็บและสรุปรายงานโดยการระบุปัญหาการวิจัย

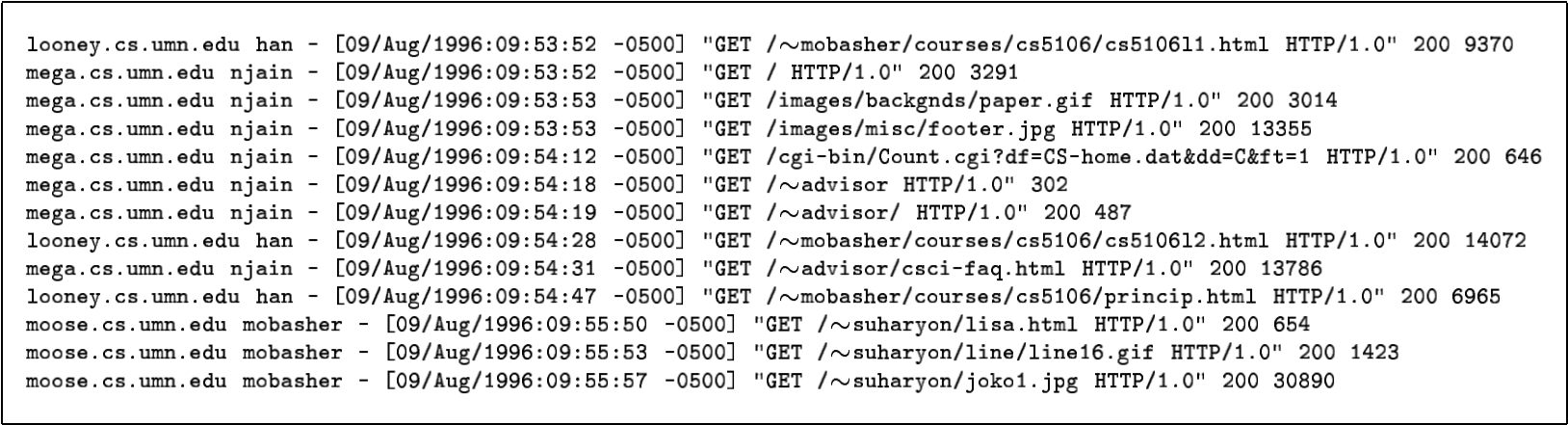
### 2.1.2 An Architecture for Web Mining

Bamshad Mobasher (1997) ได้ให้ข้อมูลเกี่ยวกับ An Architecture for Web Mining ว่า Web server ของ WWW จะแสดง Log ที่เกี่ยวกับ HTTP protocol และ Log นี้จะแสดง IP address, user id, access time, request method



รูปที่ 2.1: General Architecture for WEBMINER จาก Eduardo Mena's,

โดย Namit Jain 1997 สืบค้นจาก http://eolo.cps.unizar.es



รูปที่ 2.2: Sample Entries from a Web Server Access Log จาก Eduardo Mena's,

โดย Namit Jain 1997 สืบค้นจาก http://eolo.cps.unizar.es

โดยปกติของเว็บจะมีทั้งรูป เสียง ไฟล์ วิดีโอ ที่จะแสดงลงใน Log ไฟล์ เราสามารถใช้งานได้ แต่ต้องทำความสะอาดหรือการ Cleaning data เพื่อจะได้นำข้อมูลที่จำเป็นสำหรับเราไปใช้งาน

### 2.1.3 Web Scraping, Web Data Extraction

การเก็บข้อมูลจากหน้าเว็บไซต์ บางครั้งถูกเรียกว่า Web Scraping, Web Data Extraction, Screen Scraping หรืออื่นๆ คือ การได้มาซึ่งข้อมูลโดยวิธีการใดๆ ที่นอกเหนือจากการให้โปรแกรมติดต่อกับ API (หรือที่พบได้ชัดเจนคือผ่านการโต้ตอบโดยใช้เว็บบราวเซอร์) การเก็บข้อมูลจากหน้าเว็บไซต์ที่พบได้บ่อยครั้งคือ การให้โปรแกรมอัตโนมัติทำงานโดยส่งคำร้องขอหน้าเว็บไซต์ (ส่วนใหญ่เป็นรูปแบบ HTML หรือ ส่วนประกอบของเว็บที่เป็นไฟล์อื่นๆ) และวิเคราะห์ข้อมูลที่ต้องการออกมา

บางกรณีการเก็บข้อมูลจากหน้าเว็บไซต์มีความจำเป็นต่อการได้มาซึ่งข้อมูล ตัวอย่างเช่น เว็บไซต์ไม่มีการทำ API ส่งผลให้การได้มาซึ่งข้อมูลเป็นไปได้ยาก จึงต้องใช้เทคนิคดังกล่าวขั้นต้นเก็บข้อมูลจากหน้าเว็บไซต์

ในปัจจุบันการเก็บข้อมูลจากหน้าเว็บไซต์เป็นเทคนิคที่ถูกใช้อย่างแพร่หลายในการทำงานเชิงวิเคราะห์ข้อมูลและความปลอดภัย อีกทั้งข้อมูลสามารถถูกจัดเก็บได้ในรูปแบบของฐานข้อมูล ซึ่งผู้ที่มีข้อมูลสามารถนำไปใช้ประโยชน์อื่นๆต่อได้โดยไม่มีความจำเป็นต้องเข้าไปในหน้าเว็บไซต์

(เข้าถึงได้จาก: https://medium.com/@minatonamikazerak/python-ทำ-web-scraping-web-crawler-d2d00244ac94)

## 2.2 เทคโนโลยีที่เกี่ยวข้อง

### 2.2.1 Selenium

Selenium คือ ชุดคำสั่งของโปรแกรมที่พร้อมใช้งานเพื่อทดสอบการทำงานของ  
แอปพลิเคชั่นและเว็บไซต์ โดยสามารถทำการทดสอบกับภาษาต่างๆได้เช่น C#, Java, PHP, Python และยังมีเครื่องมือที่ใช้ร่วมกับ Selenium อีก 4 เครื่องมือ ได้แก่

2.2.1.1 Selenium Remote Control (RC)  
2.2.1.2 Selenium Integrated Development Environment (Selenium IDE)  
2.2.1.3 Selenium Grid  
2.2.1.4 Web Driver

Selenium Web Driver เป็นเครื่องมือที่ช่วยให้เราสามารถสร้างโปรแกรมในการทดสอบเว็บแอพพลิเคชั่นและมันได้รับการออกแบบให้เข้าใจง่ายและระบุข้อจำกัดเอาไว้โดยตัว  
Web Driver ทำงานผ่านบราว์เซอร์ทำให้มันมีประสิทธิภาพสูง (Beans Erik, 2551)

2.2.2 Python

เป็นภาษาการเขียนโปรแกรมเชิงวัตถุระดับสูงที่มีการตีความเชิงความหมายแบบไดนามิกส์ มีโครงสร้างข้อมูลระดับสูงประกอบกับการพิมพ์แบบไดนามิกส์ และการเชื่อมโยงแบบไดนามิกส์ ทำให้น่าสนใจมากสำหรับการพัฒนาแอพพลิเคชั่นอย่างรวดเร็ว เช่นเดียวกับการใช้เป็นภาษาสคริปต์หรือกาวเพื่อเชื่อมต่อส่วนประกอบที่มีอยู่ด้วยกัน ไวยากรณ์ของ Python ที่เรียบง่ายและเรียนรู้ง่ายเน้นการอ่านได้ง่ายและลดค่าใช้จ่ายในการบำรุงรักษาโปรแกรม Python สนับสนุนโมดูลและแพ็คเกจซึ่งสนับสนุนโมดูลแบบแยกส่วนและการใช้รหัสซ้ำ Python interpreter และไลบรารี่มาตรฐานมีให้บริการในรูปแบบซอร์สหรือไบนารีโดยไม่มีค่าใช้จ่ายสำหรับแพลตฟอร์มหลักทั้งหมด และสามารถแจกจ่ายได้อย่างอิสระ บ่อยครั้งที่โปรแกรมเมอร์ชอบใช้ Python เนื่องจากผลผลิตที่เพิ่มขึ้นนั้นมีให้ เนื่องจากไม่มีขั้นตอนการรวบรวมรอบการแก้ไขการทดสอบการดีบักจึงรวดเร็วอย่างไม่น่าเชื่อ การดีบักโปรแกรม Python นั้นง่าย ข้อผิดพลาดหรืออินพุตที่ไม่ดีจะไม่ทำให้เกิดการแบ่งเซ็กเมนต์ แต่เมื่อล่ามพบข้อผิดพลาดจะทำให้เกิดข้อยกเว้น เมื่อโปรแกรมไม่ตรวจจับข้อยกเว้นล่ามจะพิมพ์การติดตามสแต็ก ดีบักเกอร์ระดับซอร์สอนุญาติให้ตรวจสอบตัวแปรโลคอล และโกลบอลการประเมินผลของนิพจน์โดยพลการตั้งค่าเบรกพอยต์การก้าวผ่านโค้ดทีละบรรทัดและอื่น ๆ ตัวดีบั๊กเขียนด้วยภาษา Python เพื่อพิสูจน์ถึงพลังของ Python ในทางกลับกันมักจะเป็นวิธีที่เร็วที่สุดในการดีบักโปรแกรมคือการเพิ่มคำสั่งการพิมพ์บางอย่างลงในแหล่งที่มา: รอบการแก้ไขการทดสอบการแก้ปัญหาอย่างรวดเร็วทำให้วิธีการง่าย ๆ นี้มีประสิทธิภาพมาก (Guido van Rossum, 2533)

### 2.2.3 JavaScript

ภาษาสคริปต์เชิงวัตถุ หรือ JavaScript (จาวาสคริปต์) เป็นภาษาคอมพิวเตอร์ที่ใช้กันอย่างแพร่หลายในอินเทอร์เน็ต ดังนั้นจึงได้รับความนิยมเป็นอย่างสูง รวมถึงใช้งานอย่างกว้างขวางในอินเทอร์เน็ต เพื่อให้เว็บไซต์สามารถโต้ตอบกับผู้ใช้งานและทำให้ประสบการณ์ของผู้ใช้งานดียิ่งขึ้น โดยภาษา JavaScript จะมีมาตราฐานที่ชื่อ ECMA เพื่อให้บราวเซอร์แปลคำสั่ง ดังนั้น JavaScript จึงสามารถทำงานเฉพาะบนบราวเซอร์ที่สนับสนุนซึ่งปัจจุบันบราวเซอร์เกือบทั้งหมดเว้นแต่มาตราฐานใหม่ของ ECMA ซึ่งบางบราวเซอร์อาจไม่รองรับ JavaScript ถูกพัฒนาขึ้นโดย เน็ตสเคปคอมมิวนิเคชันส์ (Netscape Communications Corporation) โดยใช้ชื่อว่า Live Script ต่อมาได้มีการพัฒนา LiveScript ขึ้นมาใหม่เมื่อ ปี 2538 แล้วตั้งชื่อใหม่ว่า JavaScript Java Script สามารถทำให้ มีลูกเล่น ต่าง ๆ มากมาย และยังสามารถโต้ตอบกับผู้ใช้ได้อย่างทันที เช่น การจัดการการคลิกโต้ตอบกับผู้ใช้ การทำเว็บไซต์ให้ตอบสนองทั้งโทรศัพท์และคอมพิวเตอร์ เป็นต้น (Brendan Eich, 2538)

### 2.2.4 Sublime

Sublime เป็นโปรแกรมชนิดหนึ่งที่ใช้ในการเขียนโค้ดสำหรับนักพัฒนาหรือบุคคลเริ่มต้นที่จะเขียน ตัวโปรแกรมนี้สนับสนุนให้สมารถเขียนได้หลากหลายภาษา ตัวอย่างเช่นการภาษา C, C++, Java, JavaScript, C#, CSS, and Python และยังอีกหลากหลายภาษา เป็น(ส่วนเสริมการทำงาน) SASS, LESS, jQuery (Jon Skinner, 2551)

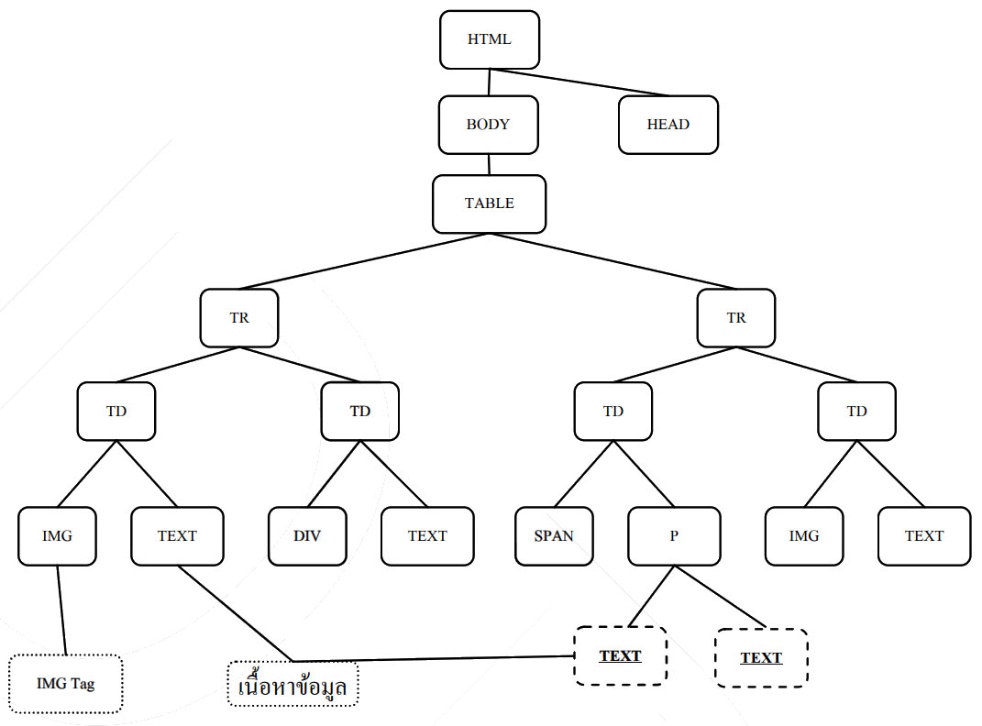
### 2.2.5 Visual Studio

คุณสมบัติใหม่เช่นการตรวจสอบความถูกต้องทางสถาปัตยกรรมสดช่วยให้พัฒนา DevOps ก่อนหน้านี้ในกระบวนการพัฒนา นอกจากนี้การปรับปรุงคุณสมบัติยอดนิยม เช่นการนำรหัส IntelliSense การปรับโครงสร้างและการแก้ไขรหัสช่วยให้คุณประหยัดเวลาและความพยายามโดยไม่คำนึงถึงภาษาหรือแพลตฟอร์ม Visual Studio สนับสนุนฟีเจอร์ภาษาการเขียนโปรแกรมล่าสุด ไม่ว่าจะทำงานกับ C #, Visual Basic, C ++, Typescript, F # หรือแม้แต่ภาษาของบุคคลที่สามเช่น JavaScript และ Python **(Luxsana Vathin, 2551)**

## 2.3 เอกสารและงานวิจัยที่เกี่ยวข้อง

### 2.3.1 ขั้นตอนและวิธีการสลัดข้อมูลสินค้าบนเว็บเพจสำหรับเว็บครอเลอร์ที่ใช้ในโปรแกรมค้นหาสินค้า

ในงานวิจัยของ กลยุทธ บพิตร ปี 2555 ศึกษาเกี่ยวกับเรื่องพฤติกรรมการซื้อสินค้าออนไลน์ของผู้บริโภค โดยผู้วิจัยพบว่าผู้บริโภคค้นหาข้อมูลของสินค้าไม่ครอบคุมและยังคงไม่ตรงกับความต้องการของผู้ใช้งาน ดังนั้นผู้วิจัยจึงพัฒนาโปรแกรมในการค้นหาสินค้าโดยสกัดข้อมูลสินค้าสำหรับเว็บครอเลอร์ เพื่อแยกแยะข้อมูลสินค้าต่างๆ เช่น รูปภาพ ราคา และอื่นๆ สำหรับเว็บขายสินค้าในประเทศไทย

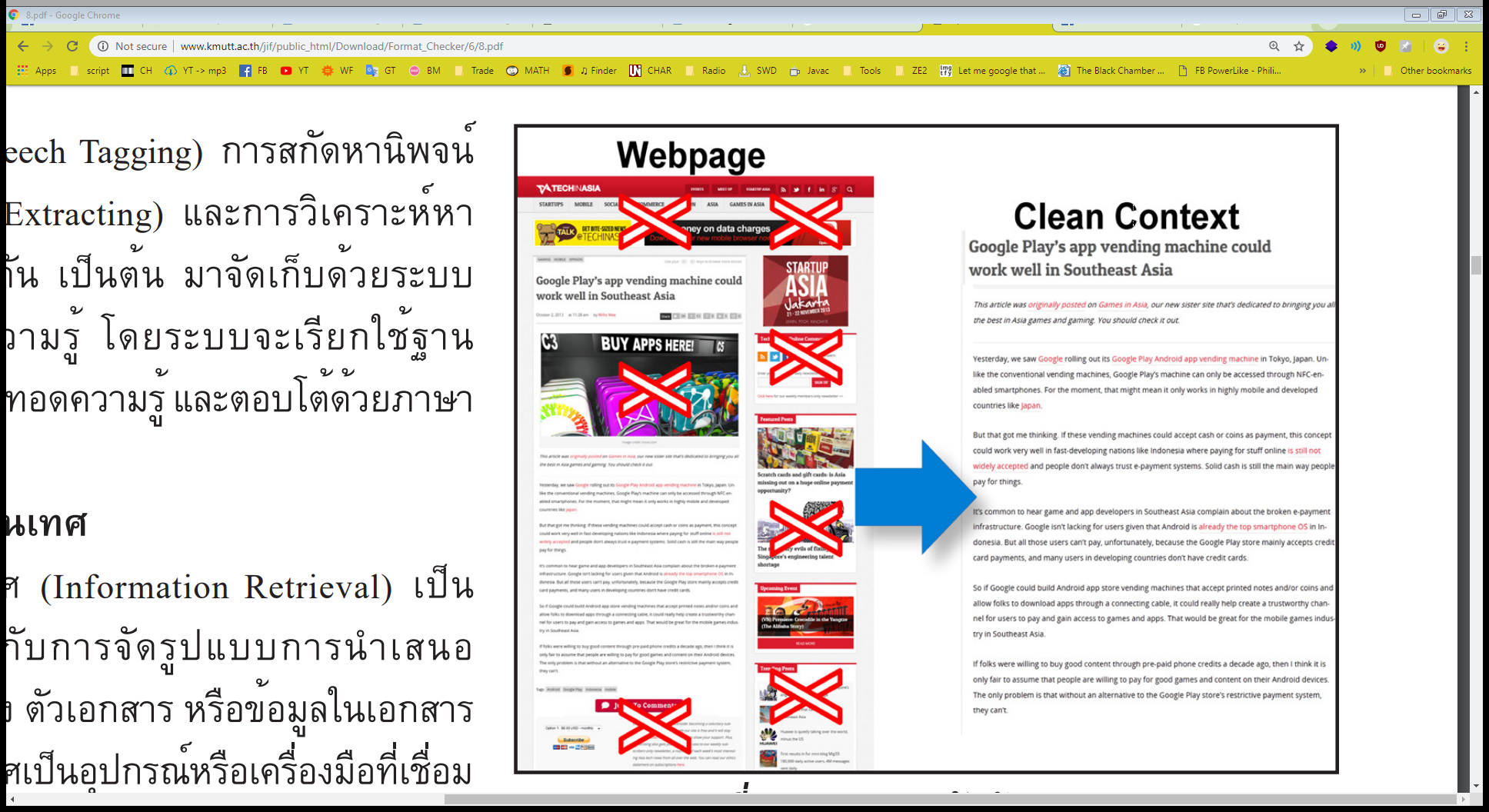


รูปที่ 2.3: ตัวอย่างโหนดในภาษา HTML กับโหนดสินค้า 1 โหนด 1 สินค้า จาก แหล่งข้อมูลปัญญา DPU, โดย กลยุทธ บพิตร 2555 สืบค้นจาก http://libdoc.dpu.ac.th/thesis

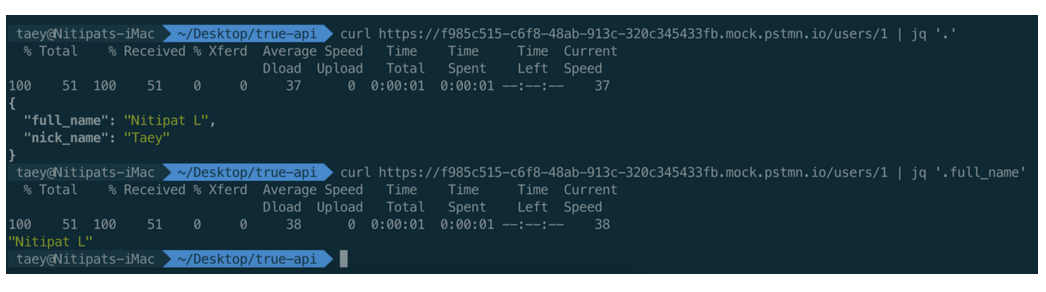
หลักการของโปรแกรมในการค้นหาข้อมูลของสินค้าคือ ทำการสกัดข้อมูลจากโครงสร้าง HTML ที่มีลักษณะคล้ายโครงสร้างต้นไม้ตามลำดับ โดยโปรแกรมจะค้นหาข้อมูลสินค้าที่มีข้อความตรงกัน รวมถึงโหนดที่อยู่ใกล้เคียงกันของสินค้า รวมกับการวิเคราะห์หาข้อมูลของสินค้าที่อยู่ใกล้เคียงกับข้อความที่ใช้ค้นหา โดยพิจารณาจากโหนดที่อยู่ภายใต้โหนดเดียวกัน

### 2.3.2 ระบบติดตามการคัดลอกเนื้อหาเว็บอัตโนมัติโดยใช้วิธีการเลือกข้อความสำคัญ

งานวิจัยของศิริพร อ่วมมีเพียร และ สันติพงษ์ ไทยประยูร ปี 2559 ได้ศึกษาเกี่ยวกับการป้องกันบทความถูกการคัดลอกเนื้อหาไม่ว่าจะเจตนาหรือไม่เจตนาก็ตาม โดยผู้เขียนบทความไม่ยินดี ส่งผลให้เกิดผลเสียต้องผู้เขียนบทความ แม้ว่าในปัจจุบันสามารถนำข้อความบางส่วนไปค้นหาผ่านเครื่องมือสืบค้น แต่ต้องใช้เวลานานและต้องใช้ความละเอียดของการสืบค้น ดังนั้นผู้วิจัยจึงพัฒนาโปรแกรมสำหรับค้นหาข้อความสำคัญจากเนื้อหาในเว็บไซต์อัตโนมัติโดยใช้เทคนิคต่างๆ ประกอบด้วยการประมวลผลภาษาธรรมชาติ การค้นคืนสารสนเทศ การสกัดข้อความบนเว็บ การเลื่อนกรอบข้อความ การตรวจสอบความคล้ายกันของเอกสาร และการให้ค่าน้ำหนักคำ



รูปที่ 2.4: แสดงการสกัดข้อความบนเพจภาพ จาก KMUTT Research, โดยศิริพร อ่วมมีเพียร และ สันติพงษ์ ไทยประยูร 2559 สืบค้นจาก http://www.kmutt.ac.th



รูปที่ 2.5:แสดงการใช้งานของ Syntax จาก A Medium Corp., โดย Nitipat Lowichakornthikun 2562 สืบค้นจาก <https://medium.com/i-gear-geek>

# บทที่ 3 วิธีการดำเนินการ

การจัดทำโครงงานชิ้นนี้มีวัตถุประสงค์เพื่อสร้างและพัฒนาระบบการค้นหาและการเก็บข้อมูลผ่าน Facebook fan page โดยผู้ทำโครงงานได้ศึกษาเอกสารและงานวิจัยหรือโครงงานที่เกี่ยวข้องเพื่อหาแนวทางการทำโครงงานออกมาให้สำเร็จ

3.1 การวิเคราะห์ออกแบบโครงงาน และลักษณะโครงสร้างของโครงงาน

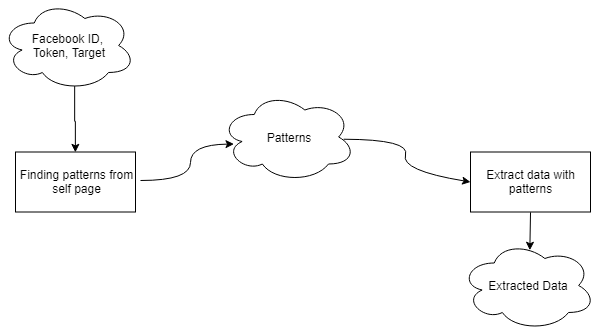
3.2 การวิเคราะห์ปัญหาที่อาจเกิดขึ้น และขั้นตอน

3.3 กระบวนการ ในการตรวจหาข้อมูล, คำ, ประโยค หรือภาพ

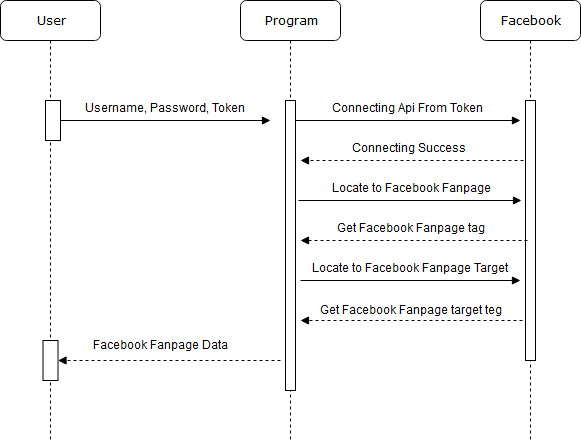
3.4 การทดลองโครงงาน

3.5 ผลลัพธ์จากการตรวจหาข้อมูลตรงตามที่ต้องการและการจัดเก็บข้อมูล

## 3.1 การวิเคราะห์ออกแบบโครงงาน และลักษณะโครงสร้างของโครงงาน

การออกแบบระบบที่ใช้ในการทำโครงงานแบ่งตามขั้นตอนการออกแบบดังนี้

รูปที่ 3.1: Architecture



รูปที่ 3.2: sequence diagram

3.1.1 วิเคราะห์โครงสร้างของเว็บไซต์ Facebook fan page โดยการตรวจสอบการแสดงผลของหน้าเว็บไซต์ รวมถึงการทำงาน ความแต่งต่างระหว่าง Browser และ Platform

3.1.2 ออกแบบขั้นตอนการดึงข้อมูล และวิธีการเชื่อมต่อเพื่อรับส่งข้อมูลกับเว็บไซด์ Facebook

3.1.3 ออกแบบรูปแบบการจัดเก็บผลที่ได้จากการดึงข้อมูล และการประเมินความถูกต้องของผลลัพธ์ของโครงงาน

## 3.2 การวิเคราะห์ปัญหาที่อาจเกิดขึ้น และขั้นตอน

ปัญหาที่อาจเกิดขึ้นจากการดึงข้อมูลจากการแสดงผลของหน้าเว็บโดยใช้ python และ JavaScript ดังนี้

3.2.1 การแสดงผลของเว็บไซต์ของ Facebook มีการเปลี่ยนแปลงในรูปแบบวันต่อวัน (Dynamic Website) ส่งผลให้วิธีการตรวจหาข้อมูลในแต่ละส่วนของเว็บไซต์ต้องเปลี่ยนแปลงตาม

3.2.2 เว็บไซต์เฟสบุ๊คอาจมีการจำกัดจำนวนการร้องขอข้อมูล ซึ่งอาจส่งผลให้โครงงานมีการทำงานผิดพลาดหรือล่าช้า

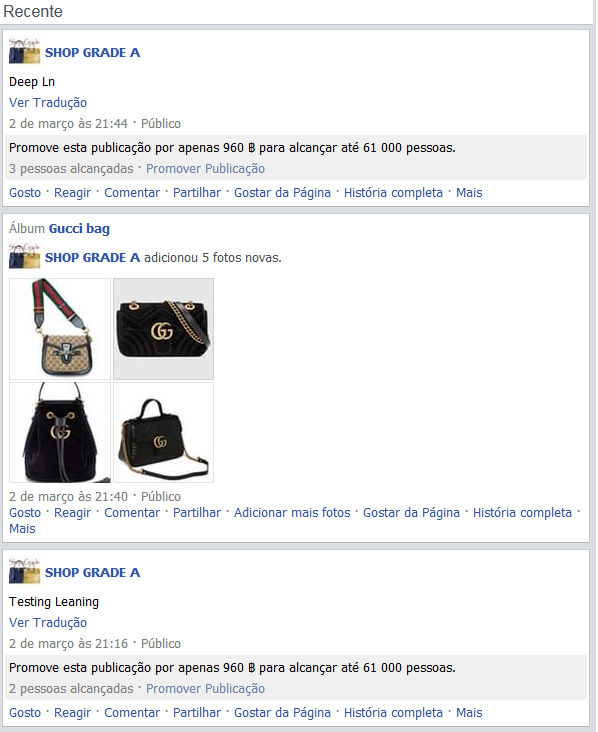
## 3.3 กระบวนการในการตรวจหาข้อมูล, คำ, ประโยค หรือภาพ

3.3.1 เรียนรู้รูปแบบของการแสดงผลข้อมูลบนหน้าเว็บไซต์

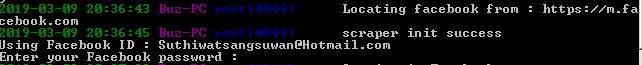
3.3.1.1 การเตรียมข้อมูลต้นแบบ และการตั้งค่าของเว็บไซต์



รูปที่ 3.3: Facebook Fan page ต้นแบบ



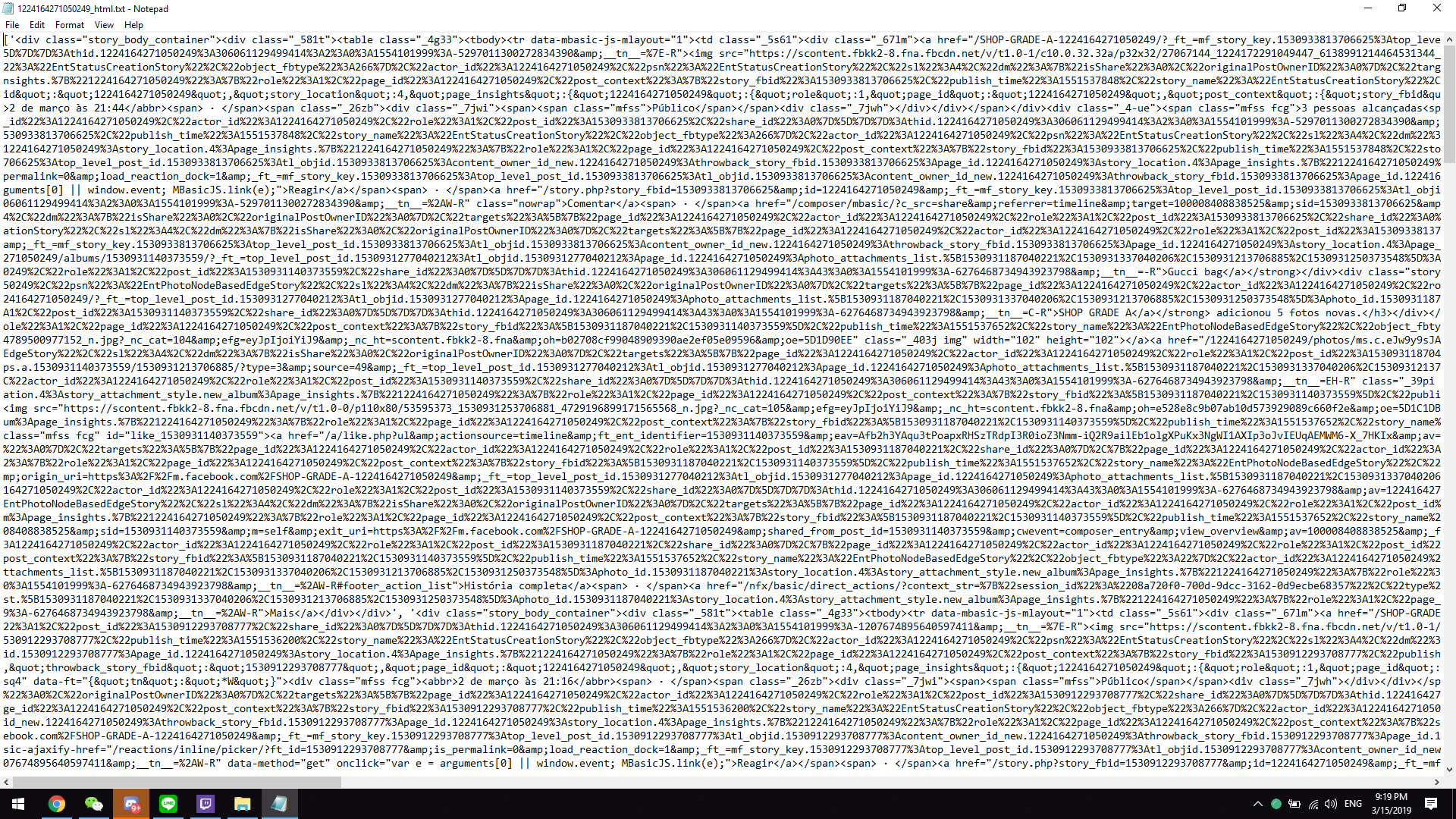
รูปที่ 3.4: Facebook Fan page โพส



รูปที่ 3.5: Login Facebook Fan page

จากรูปที่ 3.3 ถึง 3.5 เป็นข้อมูลต้นแบบเพื่อใช้ในการตรวจสอบความถูกต้องของข้อมูลผลลัพธ์ และใช้หารูปแบบของการแสดงผลของเว็บไซต์เฟสบุ๊ค รวมถึงการตั้งค่าภาษาเพื่อง่ายต่อการค้นหาปุ่มหรือข้อความต่างๆ ในการค้นหา

3.3.1.2 การตรวจสอบแท็ก HTML ในแต่ละโพสต์จากการแสดงผล



รูปที่ 3.6: TAG HTML

จากรูปที่ 3.6 เป็นข้อมูลแท็กในแต่ละโพส ที่ได้จากการแสดงผลของเว็บไซต์ Facebook fan page ของข้อมูลต้นแบบ โดยได้จากการหา parent ของแท็กชื่อเพจกระทั่งแท็กนั้นๆเกิดการซ้อนทับกันกับข้อความขอบนอกในแต่ละโพส

3.3.1.3 การดึงข้อมูลเว็บไซต์จาก token



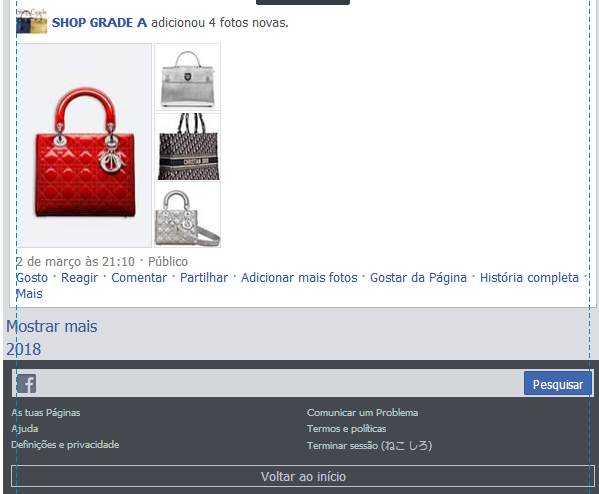
รูปที่ 3.7: TOKEN

จากรูปที่ 3.7 เป็น token ของ Facebook fan page เพื่อใช้ในการดึงข้อมูลจาก feed เพื่อนำไปเปรียบเทียบกับแท็ก โดยเทียบถึงความลึก ชนิดของแท็ก ที่มีข้อความนั้นๆ และนำมาบันทึกเป็นรูปแบบของโพสชนิดนั้นๆ

3.3.1.2 การจัดชนิดของโพสจากข้อมูลต้นแบบ



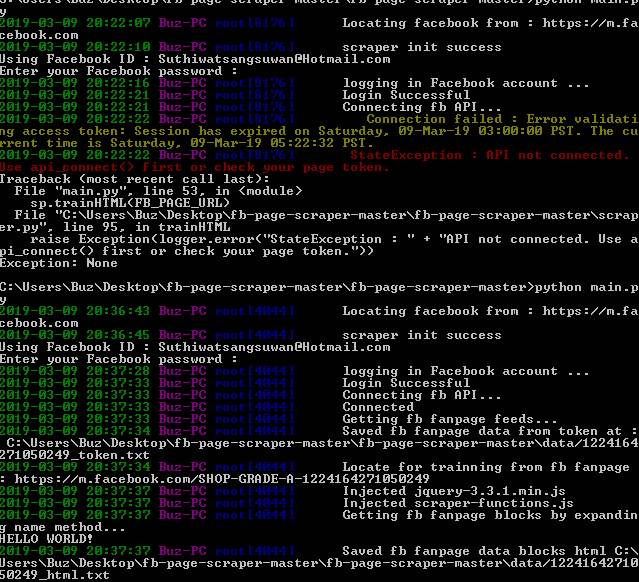
รูปที่ 3.8: Post Text



รูปที่ 3.9: Post Picture

จากรูปที่ 3.8 และ 3.9 เป็นภาพรวมวิธีการหารูปแบบรวมเช่น การทำให้โปรแกรมเรียนรู้จากการโพสเป็นภาพอย่างเดียวหรือ เรียนรู้จากการโพสเป็นข้อความอย่างเดียวเพื่อนำไปใช้กับเว็บไซต์ Facebook fan page ที่ต้องการจริง

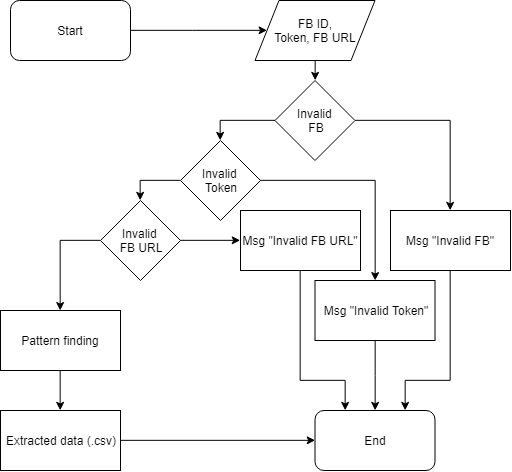
3.3.2 ดึงข้อมูลจากเว็บไซต์ Facebook fan page จากรูปแบบที่ได้จากต้นแบบ



รูปที่ 5.10: Getting Data Fb Fanpage

จากรูปที่ 3.10 เป็นภาพที่กำลังเชื่อมต่อเข้ากับ API ของทาง Facebook และกำลังนำข้อมูลที่ต้องการออกมาเป็นไฟล์เพื่อที่จะทำการตรวจแท็กของภาพหรือข้อความที่เราต้องการ

## 3.4 การทดลองโครงงาน



รูปที่ 3.11: Flow Chart

3.4.1 การทดลองโครงงานโดยการรันโปรแกรมผ่าน CMD เริ่มตั้งแต่การใส่รหัส User Account และ Password เพื่อทำการ Login เข้าระบบหลังจากนั้น โปรแกรมจะทำการเชื่อมต่อเข้ากับ API Facebook เพื่อทำการนำข้อมูลออกมา

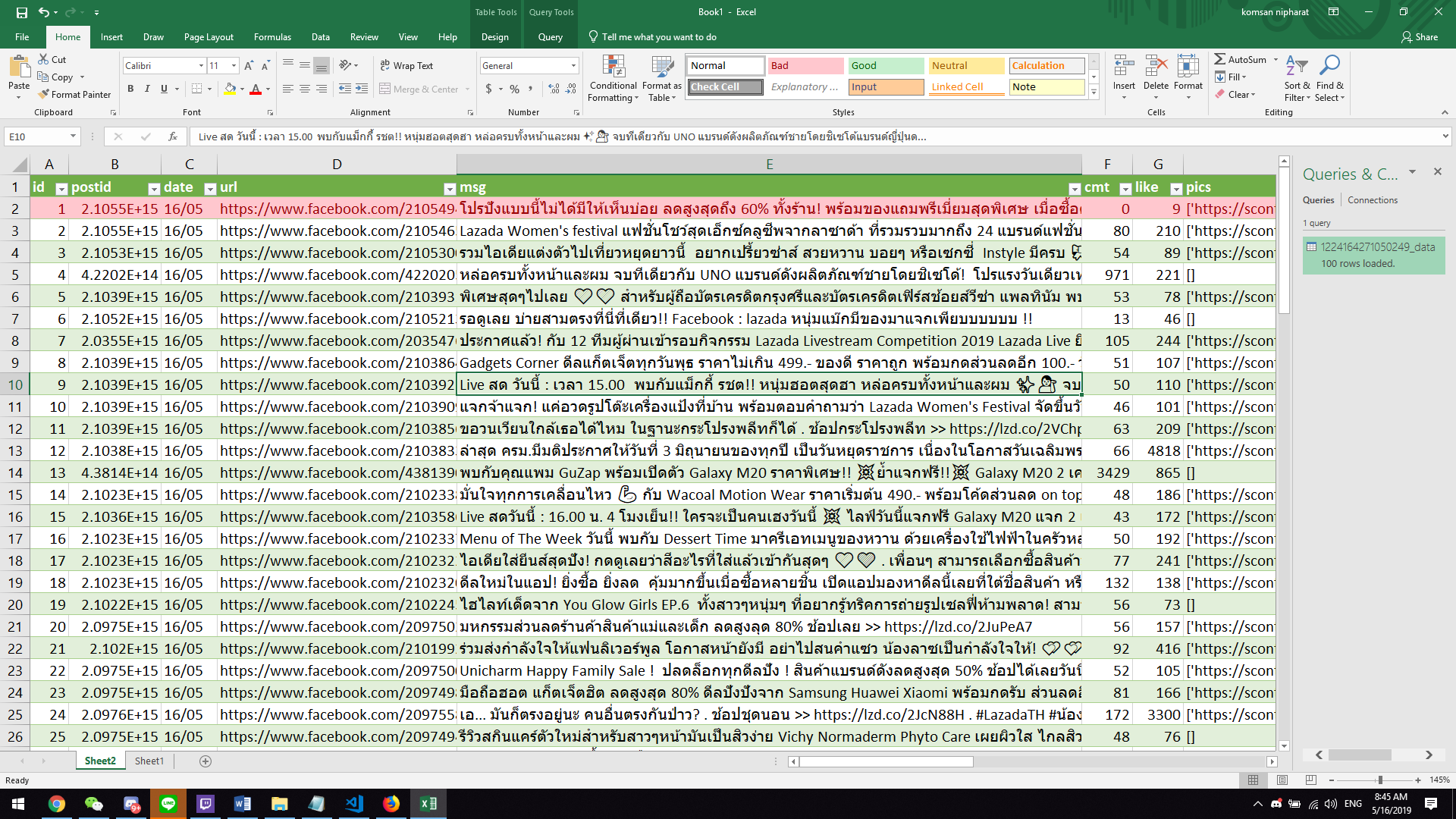
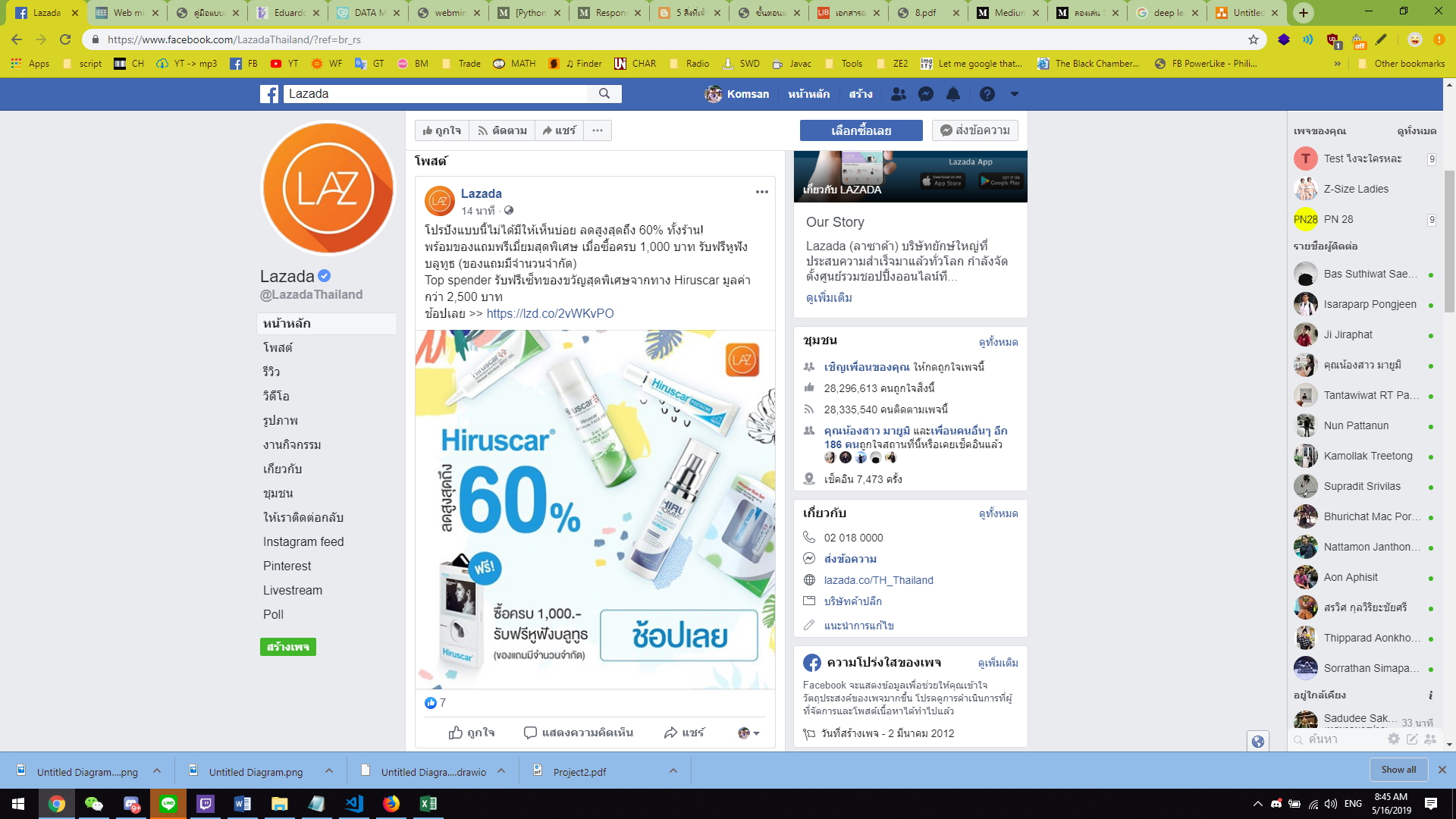
3.4.2 โปรแกรมได้ทำการนำข้อมูล TAG HTML ของโครงสร้างทาง Facebook fan page ต้นแบบออกมาในแต่ละแท็กข้อมูลเพื่อบ่งบอกวันเวลา, โพสต์รูป, โพสต์ข้อความ

3.4.3 โปรแกรมนำข้อมูล Facebook fan page ของต้นแบบที่ได้จาก API และ TAG HTML มาเทียบกันเพื่อหารูปแบบของการแสดงผล

3.4.4 โปรแกรมทำการเก็บข้อมูลของ Facebook fan page เป้าหมาย โดยใช้รูปแบบที่ได้วิเคราะห์จากข้อ

3.4.5 บันทึกลงในไฟล์ ข้อมูลในไฟล์จะประกอบด้วยโพสต์ต่างๆที่ตรงกันกับ Facebook fan page เป้าหมาย

## 3.5 ผลลัพธ์จากการตรวจหาข้อมูลตรงตามที่ต้องการและการจัดเก็บข้อมูล



รูปที่ 3.11: DATA ที่ได้และการเปรียบเทียบความถูกต้อง

จากรูปที่ 3.11 เป็นภาพข้อมูลตัวอย่างที่ถูกบันทึกหลังจากการเก็บข้อมูลจาก Facebook fan page เป้าหมาย ประกอบด้วย ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ(ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น วันและเดือน ที่โพสต์

# บทที่ 4 ผลการดำเนินงาน

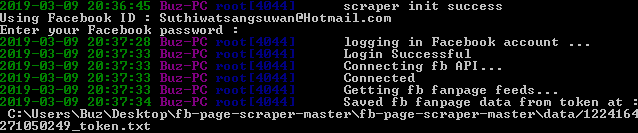
จากการรวบรวมข้อมูลและวิเคราะห์ปัญหา คณะผู้จัดทำได้ทำการพัฒนาโปรแกรมเกี่ยวกับการดึงข้อมูลทาง Facebook fan page อัตโนมัติ โดยโปรแกรมสามารถรองรับการเปลี่ยนแปลงการแสดงผลของ Facebook ได้โดยไม่จำเป็นต้องปรับปรุงโปรแกรม ซึ่งมีส่วนของการพัฒนาและดำเนินงานดังนี้

4.1 ส่วนของโปรแกรมขั้นตอนการ Login ของตัวโปรแกรม

4.2 ผลการวิจัยในการดึงข้อมูลจาก Facebook Fan page

## 4.1 ส่วนของโปรแกรมขั้นตอนการ Login ของตัวโปรแกรม

4.1.1 รูปแบบการ Login ของโปรแกรม



รูปที่ 4.1: หน้า Login เริ่มต้นการทำงาน

จากภาพที่ 4.1 การใช้งานในหน้านี้คือการทำ Login เข้าสู่ระบบเพื่อที่จะทำการเก็บข้อมูล โดยป้อน ID (identification) จะต้องเป็นรหัสที่มีอยู่แล้วที่ใช้ในการ Login Facebook Account เนื่องจากหน้าเว็บไซต์ Facebook fan page ไม่สามารถเข้าถึงได้หากไม่มีการ Login

4.1.2 การกำหนดเว็บเพจที่ต้องการจะดึงข้อมูล

C:\Users\Buz\Desktop\Pic\New folder\link page to find data.PNG

รูปที่ 4.2: ตัวอย่างการกำหนดเว็บเพจที่ต้องการจะดึงข้อมูล

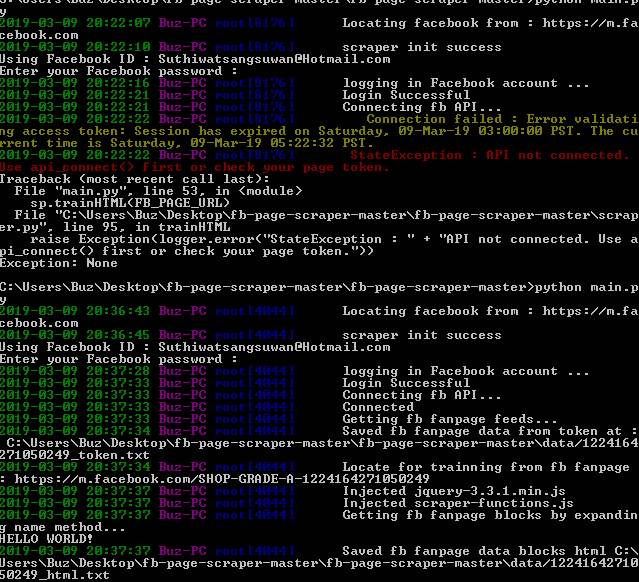
จากภาพที่ 4.2 การใช้งานในหน้านี้สามารถใส่เป้าหมายหรือเพจที่เราต้องการที่จะนำข้อมูลออกมาและสามารถเปลี่ยนแปลงเพจได้ตามที่ผู้ใช้งานต้องกา

4.1.3 รูปแบบการรับ Token ยืนยันตัวตนเข้าสู่การระบบการเก็บข้อมูล



รูปที่ 4.3: ตัวอย่างการใส่ Token ยืนยันตัวตน

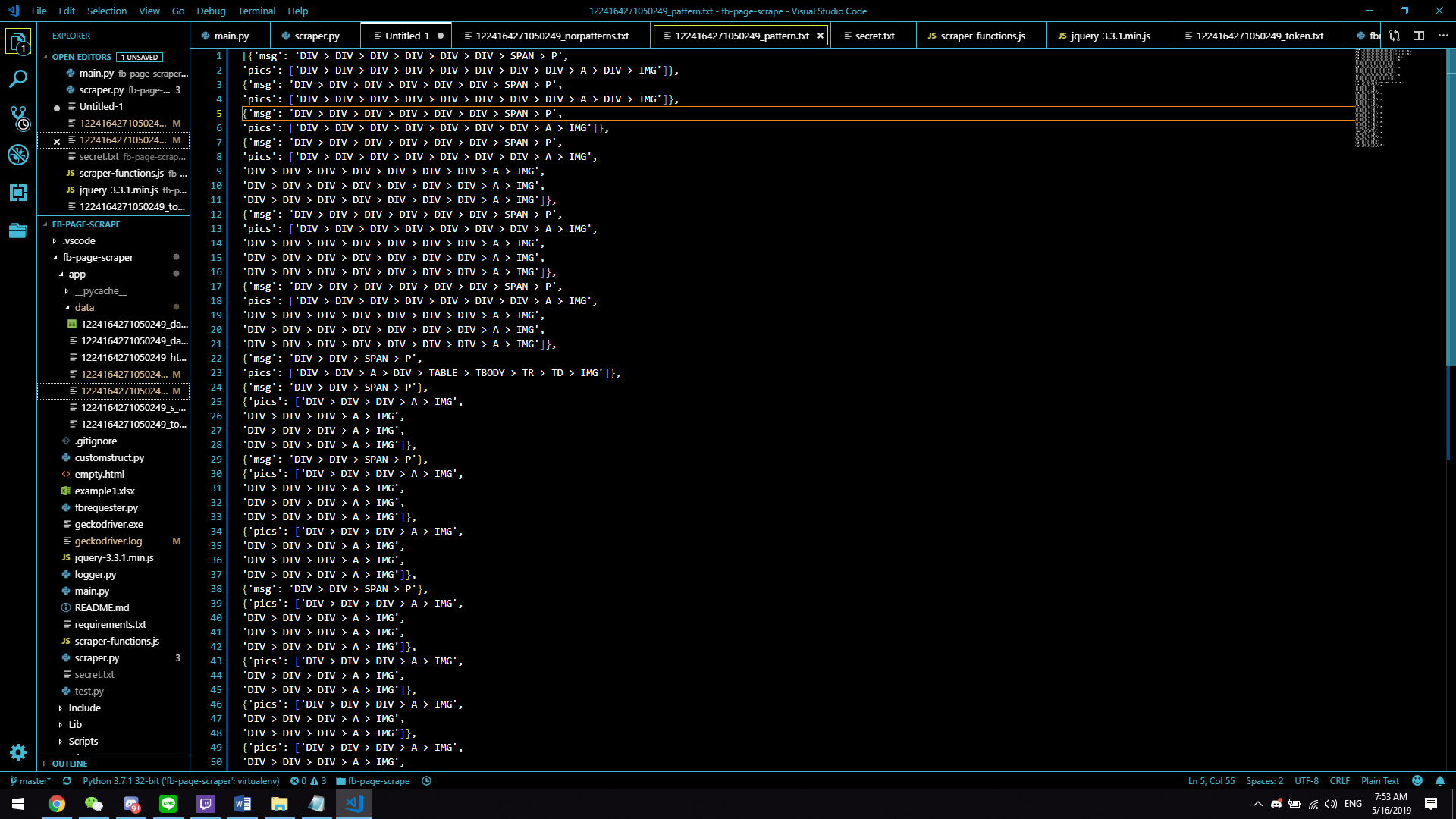
จากภาพที่ 4.3 เป็นการใส่ Token(การยืนยันตัวตน)เพื่อทำการเข้าระบบอย่างรวดเร็วโดย Token นั้นจะได้ก็ต่อเมื่อขอจากทาง Facebook เรียบร้อยแล้วเท่านั้น และเมื่อขอแล้วสามารถใส่ลงไปในโปรแกรมที่เราจัดเตรียมให้

4.1.4 รูปแบบการทำงานของโปรแกรมเมื่อทำการตั้งค่าต่างๆเสร็จสิ้น

รูปที่ 4.4: การทำงานของโปรแกรมหลังจากตั้งค่าและทำการ Login

จากภาพที่ 4.4 โปรแกรมได้เริ่มการทำงานหลังจากมีการตั้งค่าของตัวโปรแกรมให้ทำการดึงข้อมูลจากเพจต้นแบบที่ได้ทำการเลือกไว้และทำการเก็บข้อมูลหน้าเว็บไซต์ลงสู่เครื่องของผู้ใช้งาน

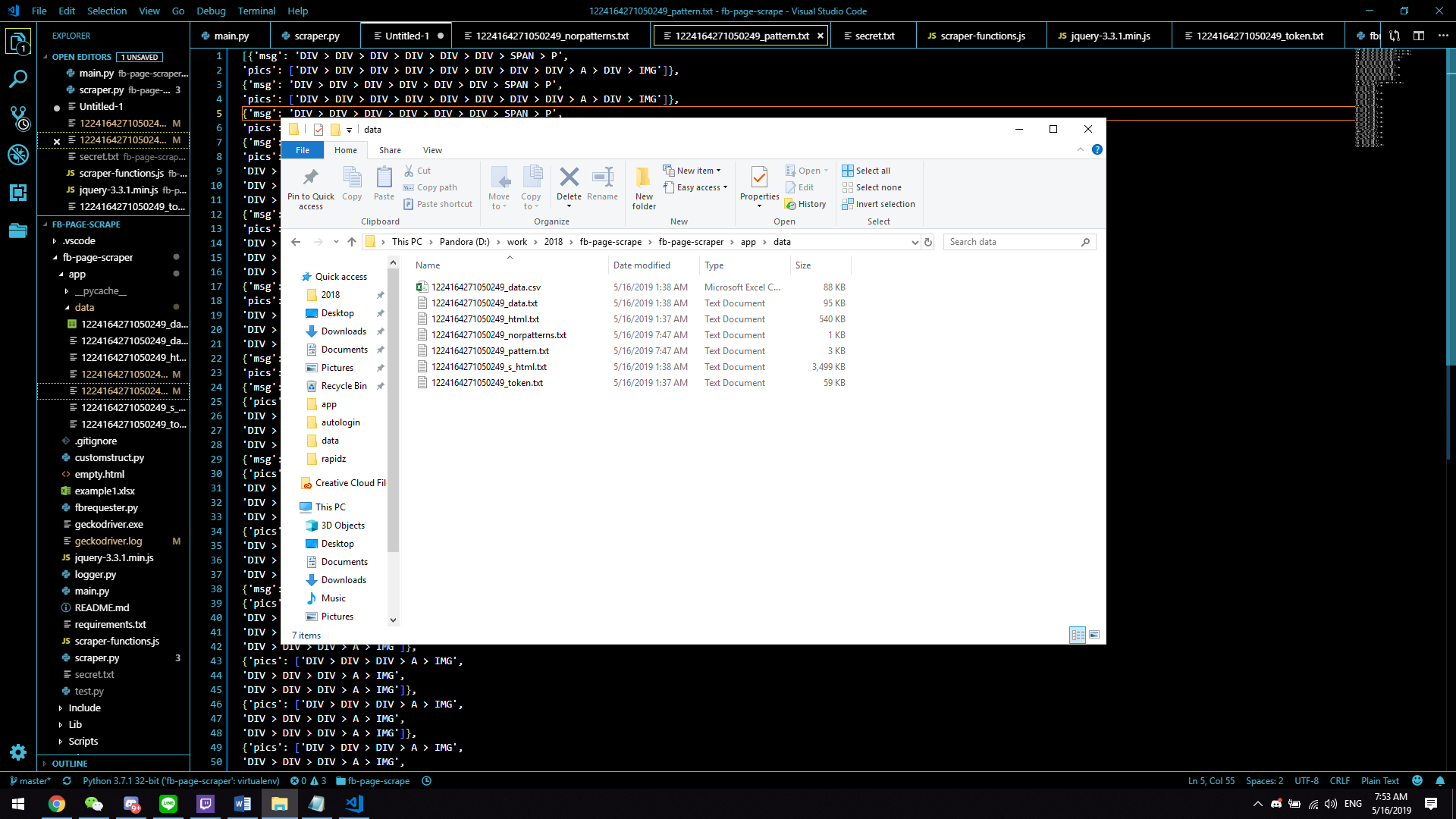
4.1.5 รูปแบบของการแสดงหลังจากการทำงานของโปรแกรม



รูปที่ 4.5: รูปแบบการแสดงค่าของการทำงานของโปรแกรม

จากภาพที่ 4.5 โปรแกรมได้ทำงานและได้แสดงค่าของ ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ(ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น วันและเดือน ที่โพสต์โดยทำการวิเคราะห์รูปแบบจากเพจต้นแบบหลายๆโพสต์ ซึ่งทำได้โดยการเทียบข้อมูลที่ได้จากการใช้ Token ของเพจต้นแบบและ HTML ไฟล์ที่เกิดจากการแสดงผลบนหน้าเว็บ

4.1.6 การจัดเก็บข้อมูลเพจเป้าหมาย



รูปที่ 4.6: การจัดเก็บข้อมูลเพจเป้าหมาย

จากภาพที่ 4.6 หลังจากโปรแกรมสามารถวิเคราะห์รูปแบบของการแสดงผลได้แล้ว โปรแกรมจะทำการเก็บรวบรวมข้อมูลจากเพจเป้าหมายโดยอัตโนมัติ โดยวิธีการเก็บข้อมูลจะใช้รูปแบบที่วิเคราะห์ออกมาก่อนหน้า จากนั้นจะทำการบันทึกไฟล์ข้อมูลที่ได้ออกมาเป็นนามสกุล csv

## 4.2 ผลการวิจัยในการดึงข้อมูลจาก Facebook Fan page

ซึ่งในผลการวิจัยในส่วนของการดึงข้อมูลมีการจัดเก็บตัวผลลัพธ์จากการตรวจหาข้อมูลตรงตามที่ต้องการและการจัดเก็บข้อมูลโดยจะมีตารางแสดงให้ดูประกอบไปด้วย ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ(ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น วันและเดือน ที่โพสต์

|  |  |
| --- | --- |
| Project Name: | Facebook fan page scraper from web page |
| Module Name: | Login |
| Created by: | Team |
| Date of Creation: | 10/2/2561 |
| Date of Review: | 16/3/2561 |

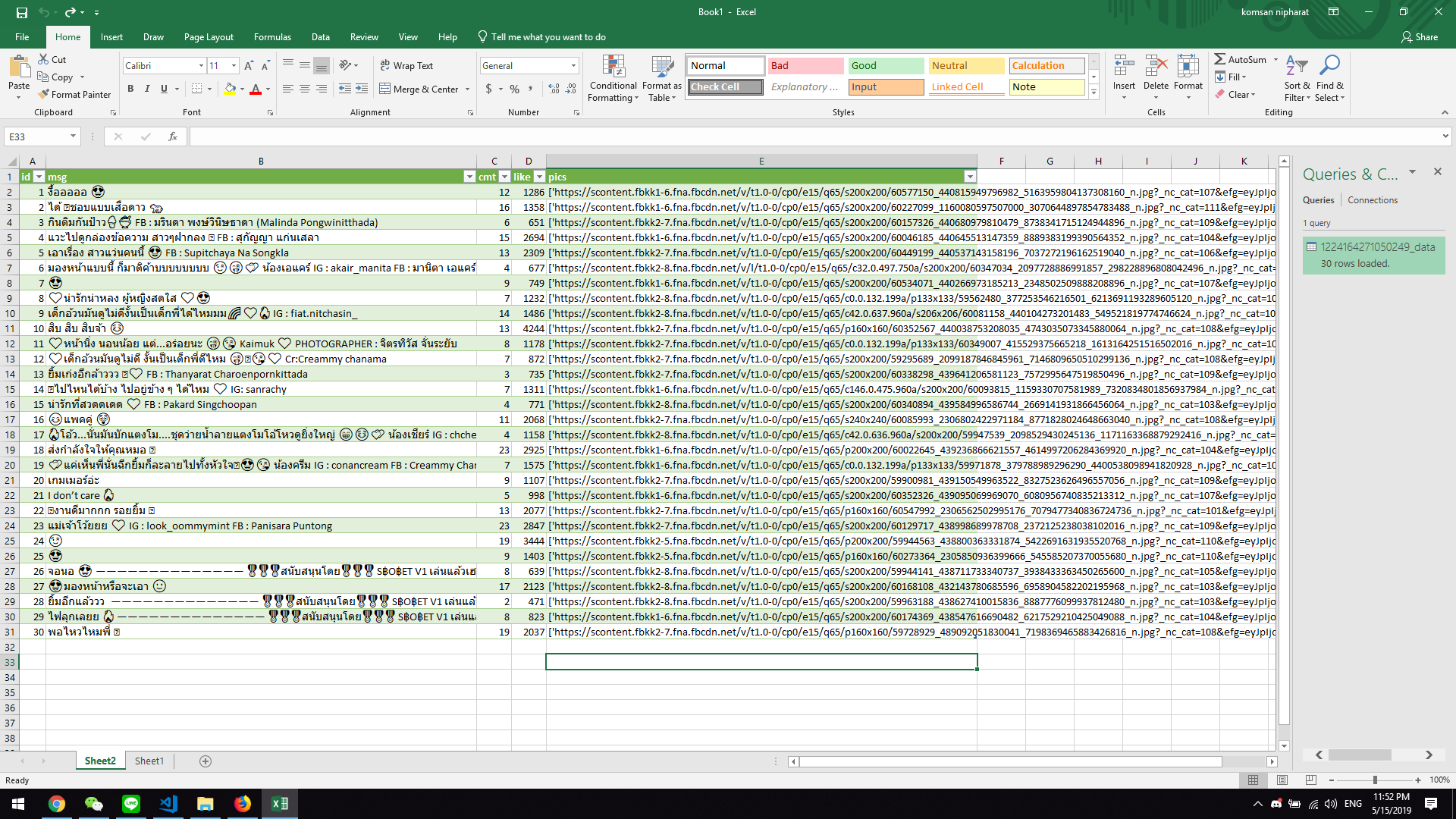
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TEST CASE ID | TESTSCEARIO | TEST CASE | PRECONDITION | TEST STEPS | TEST DATA | EXPECTED RESULT |
| TC\_LOGIN\_001 | Verify the login of token Facebook | Enter valid User name and valid Password | 1.Need a valid Token to do login | 1.Paste Token | Valid Token | Login Successful. |
| 2.Enter User Name | Valid Name |
| 3.Enter Password | Valid Password |
| 4.Push Button Enter |  |
| TC\_LOGIN\_002 | Verify the login of token Facebook | Enter valid User name and valid Password | 1.Need a valid Token to do login | 1.Paste Token | Valid Token | A message “Login can’t Success Password Incorrect.” |
| 2.Enter User Name | Valid Username |
| 3.Enter Invalid Password | Invalid Password |
| 4.Push Button Enter |  |
| TC\_LOGIN\_003 | Verify the login of token Facebook | Enter valid User name and valid Password | 1.Need a valid Token to do login | 1.Paste Invalid Token | Invalid Token | A message “API not connected check your page token.” |
| 2.Enter User Name | Valid Name |
| 3.Enter Password | Valid Password |
| 4.Push Button Enter |  |
| TC\_LOGIN\_004 | Verify the login of token Facebook | Enter valid User name and valid Password | 1.Need a valid Token to do login | 1.Paste Token | Valid Name | A message “Login can’t Success Username Incorrect |
| 2.Enter Invalid User Name | Invalid Username |
| 3.Enter Password | Valid password |
| 4.Push Button Enter |  |
| TC\_LOGIN\_005 | Verify the login of token Facebook | Enter valid User name and valid Password | 1.Need a valid Token to do login | 1.Paste Invalid Token | Invalid Token | A message” Login can’t Success Api not connect Username and Password Incorrect.” |
| 2.Enter Invalid User Name | Invalid Username |
| 3.Enter Invalid Password | Invalid Password |
| 4.Push Button Enter |  |

4.2.1 ตารางแสดง Test Case

|  |  |
| --- | --- |
| Project Name: | Facebook fan page scraper from web page |
| Module Name: | extract Result |
| Created by: | Team |
| Date of Creation: | 10/2/2561 |
| Date of Review: | 16/3/2561 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TEST CASE ID | TESTSCEARIO | TEST CASE | PRECONDITION | TEST STEPS | TEST DATA | EXPECTED RESULT |
| TC\_HD\_001 | Check extract result | Check headline form Result | 1.Need a preform program | 1. Open csv file | Valid to open csv | Extract data are correct |
| 2. Open Facebook Fan page Which you have chosen | Valid Fan page |
| 3. Check post, photo, comment to be the same data | Valid Data |
| TC\_HD\_002 | Check extract result | Check headline form Result | 1. Need a preform program | 1. Open csv file | Valid to open csv | A message “can’t Success Fan page Incorrect.” |
| 2. Open Invalid Facebook Fan page Which you have chosen | Invalid Fan page |
| TC\_HD\_003 | Check extract result | Check headline form Result | 1. Need a preform program | 1. Open csv file | Valid to open csv | Extract data shown as blank field |
| 2. Open Facebook Fan page Which you have chosen | Valid Fan page |
| 3. Check post, photo, comment to be the same data | Invalid Data |

4.2.2 ตารางแสดงผลการจัดเก็บข้อมูล



รูปที่ 4.7 ผลลัพธ์ของโปรแกรม

จากรูปภาพที่ 4.7 จะแสดงให้เห็นถึงองค์ประกอบไปด้วย ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ (ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น วันและเดือน ที่โพสต์

# บทที่ 5 สรุปผลและข้อเสนอแนะ

ในการทดลองการใช้งานโปรแกรมดึงข้อมูลจาก Facebook Fan Page สามารถสรุปผลการทดลองและข้อเสนอแนะดังนี้

## 5.1 สรุปผลการทดลอง

การทดลองใช้งานโปรแกรมดึงข้อมูลจาก Facebook fan page โดยตัวโปรแกรมจะช่วยดึงข้อมูลมาจาก Facebook fan page โดยการทดลองใช้นั้นทำให้เราได้ข้อมูลที่จะระบุถึงชื่อไฟล์ทั้ง ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ(ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น และวันที่โพสต์ และตัวโปรแกรมสามารถใช้งานได้แต่จำเป็นต้องสามารถเปิดโปรแกรม python ได้ ผลลัพธ์ที่ได้จากการทดลองใช้ระบบสามารถสรุปได้ว่าการทดลองทำงานได้ดังนี้

5.1.1 ระบบสามารถให้ผู้ใช้งานสามารถกำหนดเพจที่ต้องการดึงข้อมูลออกมาได้  
 5.1.2 ระบบสามารถแสดง ไอดี ข้อความของโพสต์ ลิ้งค์รูปภาพ(ถ้ามี) จำนวนการกดถูกใจ จำนวนความคิดเห็น วันและเดือนที่โพสต์

## 5.2 ข้อเสนอแนะ

การเพิ่มความถูกต้องของโปรแกรมสามารถเพิ่มให้ดียิ่งขึ้นได้ โดยสามารถทำได้ดังนี้

5.2.1 เพิ่มข้อมูลต้นแบบ

การเพิ่มความถูกต้องของโปรแกรมสามารถเพิ่มความถูกต้องได้ หากแต่ข้อมูลเพจต้นฉบับจำเป็นต้องมีข้อมูลเพิ่มขึ้น และมีรูปแบบการแสดงผลที่หลากหลาย เพื่อการวิเคราะห์ของโปรแกรมสามารถทำงานได้ดีขึ้น

5.2.2 ใช้ระบบปัญญาประดิษฐ์ซอฟแวร์คอมพิวเตอร์ที่่เลียนแบบการทำงานของระบบโครงข่ายประสาท (Deep learning)

การวิเคราะห์ข้อมูลจากเพจต้นแบบสามารถทำได้มีประสิทธิ์ภาพยิ่งขึ้น หากมีการนำปัญญาประดิษฐ์มาใช้ร่วม ทั้งนี้ข้อมูลต้นแบบจำเป็นจะต้องมีคุณลักษณะที่ดี กล่าวคือมีจำนวนมากและถูกกำหนดชนิดของโพส (Label) นั้นๆไว้

เอกสารอ้างอิง

Beans, Erik. (2551). **Selenium WebDriver**. Retrieved from: <https://www.seleniumhq.org/projects/webdriver/>

กลยุทธ บพิตร. (2555). **ขั้นตอนและวิธีการสกัดข้อมูลสินค้าบนเว็บเพจสำหรับเว็บครอเลอร์ที่ใช้ในโปรแกรมค้นหาสินค้า.** หลักสูตรวิทยาศาสตรมหาบัณทิตสาขาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณทิตย์

ศิริพร อ่วมมีเพียร และ สันติพงษ์ ไทยประยูร. (2559). **ระบบติดตามการคัดลอกเนื้อหาเว็บอัตโนมัติโดยใช้วิธีการเลือกข้อความสำคัญ**.

Guido van Rossum. (2533). **Python**. Retrieved from:

<https://www.python.org/>

Brendan Eich. (2538). **JavaScript**. Retrieved form:

<https://javascript.info/>

Jon Skinner. (2551). **Sublime**. Retrieved from:

<http://www.sublimetext.com/>

**Luxsana Vathin. (2551).** **Visual Studio**. Retrieved from:

<https://sites.google.com/site/luxsanavathin/gui/xngkh-prakxb-phun-than>