การวิเคราะห์ลิงค์เว็บไซต์ของคณะวิศวกรรมศาสตร์ มทร.พระนคร

RMUTP Faculty of Engineering Website Link Analysis

นิลมิต นิลาศ

งานวิจัยนี้ได้รับทุนอุดหนุนจากงบประมาณรายได้ ประจำปีงบประมาณ พ.ศ 2561

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

การวิเคราะห์ลิงค์เว็บไซต์ของคณะวิศวกรรมศาสตร์ มทร.พระนคร

นิลมิต นิลาศ

งานวิจัยนี้ได้รับทุนอุดหนุนจากงบประมาณรายได้ ประจำปีงบประมาณ พ.ศ 2561
คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการเชื่อมโยงของหน้าเว็บไซต์คณะวิศวกรรมศาสตร์ทั้งการเชื่อมโยงภายในและภายนอกและการเกิดข้อผิดพลาดของการเข้าถึงลิงค์ในหน้าเว็บแต่ละสาขาวิชารวมทั้งสิ้น 9 หน้าเว็บของแต่ละสาขา จากเหตุผลที่นักพัฒนาเว็บมุ่งมั่นในการพัฒนามีการสร้างเพิ่มข้อมูลเข้าไปและบางส่วนมีการปรับปรุงข้อมูลออก โดยใช้ช่วงระยะเวลาเก็บข้อมูลประมาณ 4 เดือนระหว่างกุมภาพันธ์ 2561 ถึงพฤษภาคม 2561 โดยการเข้าถึงทุกเส้นทาง ได้ผลลัพธ์คือมีลิงค์ที่เกิดรหัสข้อผิดพลาด 404 (Not Found) มากสุด 405 (Method Not Allowed) และ 504 (Gateway Timeout) ตามลำดับรวมประมาณ 78 เปอร์เซ็นต์ของการแสดงหมายเลขรหัสความผิดพลาด โดยรวมแล้วมีลิงค์จากเว็บไซต์คณะวิศวกรรมศาสตร์ทั้งออกและเข้าประมาณ 7,500 ลิงค์

คำสำคัญ: วิเคราะห์ลิงค์ การเชื่อมโยงเว็บ สถานะ HTTP

Abstract

This research aims to study the website link analysis of Faculty of Engineering RMUTP both of the internal link or the external link via the HTTP standard response code of all 9 under graduate departments web pages. From reasons of the web developers create and modify some part of content such as to include more the pieces of information or to exclude its. The website data collection period started from February 2018 to May 2018 via several directory paths or link-paths and the author found 404 (Not Found) most often than 405 (Method Not Allowed) and 504 (Gateway Timeout) in order, on average 78 percent. Totally, the links into and out of the engineering faculty website approximately 7,500 connections.

Keywords: Web link, Link analysis, HTTP Status

สารบัญ

เรื่อง	หน้า
บทคัดย่อ	ก
สารบัญ	ข
สารบัญรูป	ค
สารบัญตาราง	ง
กิตติกรรมระกาศ	จ
บทที่ 1 บทนำ	
1.1 ความสำคัญและที่มาของปัญหาที่ทำการวิจัย	1
1.2 วัตถุประสงค์ของโครงการวิจัย	1
1.3 ขอบเขตของโครงการวิจัย	1
1.4 ประโยชน์ที่คาดว่าจะได้รับ	1
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	
2.1 ทฤษฎีและกรอบแนวความคิด	2
2.2 กระบวนการเข้าถึงข้อมูลและการวิภัช	3
2.3 การทบทวนวรรณกรรมที่เกี่ยวข้อง	4
บทที่ 3 การดำเนินการวิจัย	
3.1 การศึกษาการเชื่อมโยง	7
3.2 การติดตั้งโปรแกรมช่วยในการทำงาน	7
3.3 การเตรียมการทดสอบเครื่องมือ	10
บทที่ 4 ผลการวิจัย	
4.1 ผลการเก็บข้อมูล	12
4.2 การประมวลผลจำนวนรหัสความผิดพลาดของลิงค์	16
บทที่ 5 สรุป และ ข้อเสนอแนะ	
5.1 ผลการวิเคราะห์ข้อมูล	17
5.2 ข้อเสนอแนะ	17
เอกสารอ้างอิง	18
ภาคผนวก	19

สารบัญรูป

รูปที่		หน้า
2.1	การเชื่อมโยงและตอบสนองของ HTTP ในการใช้งานข้อมูลบนเครื่อง Server	2
2.2	ช่วงเวลาในการเกิดเหตุการณ์ของการใช้งานระหว่างผู้ใช้เว็บ	3
2.3	ตัวอย่างผังโครงสร้างของ Sever บนระบบ DMZ	4
2.4	ผังแบบ Bow-tie ที่บ่งถึงความเชื่อมโยงระหว่างอินพุตแกน และเอาต์พุต	4
2.5	ผังหน้าของเว็บไซต์คณะวิศวกรรมศาสตร์ด้านขวามือคือโค๊ดโปรแกรมที่แสดงหน้าเว็บ	5
2.6	ตัวอย่าง HTML ส่วนหนึ่งของหน้าเว็บไซต์คณะวิศวกรรมศาสตร์	5
2.7	ตัวอย่างลิงค์ของรูปพื้นหลัง	6
3.1	สัญลักษณ์ของโครมเบราว์เซอร์	7
3.2	เบราว์เซอร์ Chrome กับการสืบค้นข้อมูล	8
3.3	สัญลักษณ์ไออี	8
3.4	เบราว์เซอร์ ie กับการสืบค้นข้อมูล	8
3.5	สัญลักษณ์ไฟล์ฟอกซ์	9
3.6	การใช้ Firefox สืบค้นข้อมูลจาก Google.com	9
3.7	การใช้โปรแกรม Python ดึงข้อมูลจากหน้าเว็บไซต์คณะวิศวกรรมศาสตร์ มทร.พระนคร	9
3.8	การเรียกใช้งาน R บน MS-Windows 10	10
3.9	ตัวอย่างการเรียกดูข้อมูลจากหน้าเว็บ	10
3.10	การใช้ Search Engine ในการหาความสัมพันธ์การเชื่อมโยง	11
4.1	รายการลิงค์ของสาขาวิชาต่าง ๆ	12
4.2	กรณีไม่พบเซิร์ฟเวอร์หรือลิงค์มีปัญหาอย่างสิ้นเชิง	12

สารบัญตาราง

ตารางที่ หน้า

4.1 จำนวนลิงค์ที่เกิดความผิดพลาดในการสืบค้นบนเว็บแต่ละสาขาวิชา 17

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร วช. สถาบันวิจัยและพัฒนา คณะวิศวกรรมศาสตร์ ที่ได้สนับสนุนทุนวิจัยและสถานที่ในการดำเนินงานไว้ ณ ที่นี้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหาที่ทำการวิจัย

จากการพัฒนาเว็บไซต์ของแต่ละหน่วยงานนั้นบนมาตรฐาน Web 2.0 มีสิ่งที่สามารถนำมาเป็นข้อมูลทางวิศวกรรมในการวิเคราะห์เพื่อหาข้อมูลและปรับปรุงเว็บไซต์ในการให้บริการแก่บุคคลากรทางการศึกษา ร่วมถึงบุคคลทั่วไปเช่น ผู้ปกครอง นักวิจัย ผู้บริหารของหน่วยงาน ในการเข้าถึงข้อมูลได้ง่ายสะดวกการทำงานขององค์กรทั่วไปหน้าของเว็บไซต์เป็นเสมือนหน้าบ้านเช่นเดียวกัน หน้าตาของเว็บไซต์ของคณะวิศวกรรมศาสตร์ก็เปรียบเสมือนหน้าตาของหน่วยงานดังนั้นจึงมีความจำเป็นอย่างยิ่งที่ต้องศึกษาลงรายละเอียด เพราะการสืบค้นบนเครื่องมือสืบค้นต่าง ๆ ทำให้โลกภายนอกมองเห็นข้อมูลของหน่วยงานที่ต้องการนำเสนออีกทั้งส่วนของการเชื่อมโยงทั้งภายนอกรวมถึงภายในตัว Server หรือต่างเครื่อง Server หรือ Host แต่อยู่ในกลุ่มเดียวกัน กรณีที่เส้นทางในการเชื่อมโยงมีปัญหาทำให้ผู้ใช้งานไม่สามารถเข้าถึงข้อมูลได้ ก็จะทำให้เกิดการขาดข้อมูลที่ผู้ใช้งานต้องการดังนั้นการศึกษาถึงลิงค์เชื่อมโยงของเว็บไซต์คณะวิศวกรรมศาสตร์จึงมีความสำคัญและจำเป็นอย่างยิ่งเพื่อพัฒนาการตอบสนองต่อผู้ใช้งาน รวมถึงกรณีอาจาเกิดการที่มีลิงค์เชื่อมโยงแต่ไม่มีไฟล์ข้อมูลอยู่จริงก็ทำให้ผู้ใช้งานไม่สามารถเข้าถึงข้อมูลของหน่วยงานได้ อีกทั้งการวิจัยนี้ต้องอาศัยเทคนิคในการเข้าถึงข้อมูลเพื่อคัดกรองข้อมูลและตรวจสอบความถูกต้องจากเครื่องมือต่าง ๆ ให้ข้อมูลที่แม่นยำด้วยเทคนิคด้านวิศวกรรมการสืบค้นข้อมูลที่เหมาะสม

1.2 วัตถุประสงค์ของโครงการวิจัย

1) เพื่อศึกษาลิงค์บนเว็บไซต์ของคณะวิศวกรรมศาสตร์

1.3 ขอบเขตของโครงการวิจัย

1) ตรวจสอบลิงค์ของเว็บไซต์คณะวิศวกรรมศาสตร์

2) หาลิงค์เชื่อมโยงภายนอกของเว็บไซต์คณะวิศวกรรมศาสตร์

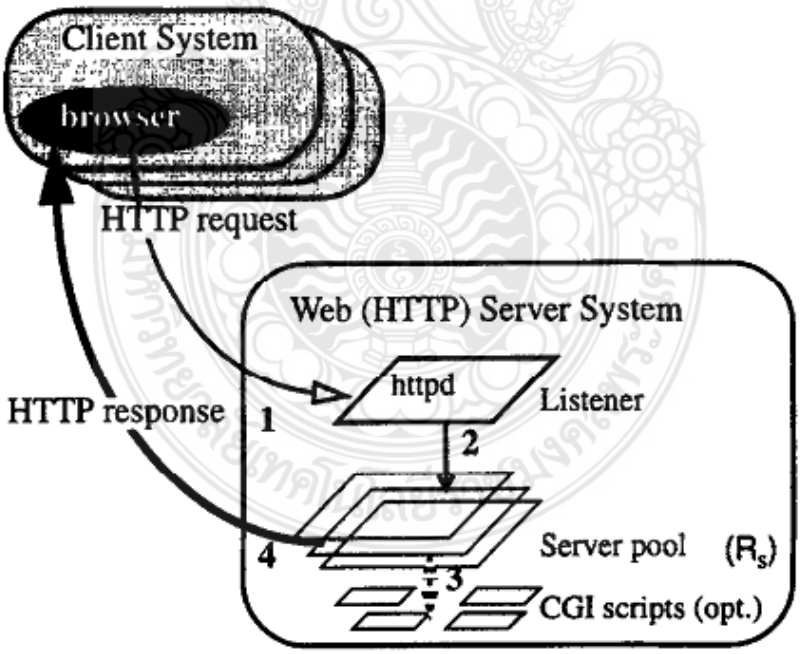
1.4 ประโยชน์ที่คาดว่าได้รับ

ได้ข้อมูลของลิงค์ต่าง ๆ ในเว็บไซต์ของคณะวิศวกรรมศาสตร์เพื่อเป็นข้อมูลให้ผู้บริหารจัดการรวมถึงผู้ดูแลเว็บไซต์ทราบถึงสถานะของเว็บไซต์ของหน่วยงานใช้ประโยชน์ในการปรับปรุงพัฒนาเว็บไซต์ต่อไป

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีและกรอบแนวความคิด

จากการใช้อินเตอร์เน็ตจากการสืบค้นบน Gopher ที่ต้องดูข้อมูลจากหัวข้อรายการเพื่อเข้าสู่ฐานข้อมูลในแต่ละเรื่องพัฒนามาถึงการแสดงผลเป็นตัวอักษรและกราฟิกที่อยู่บนหน้าเว็บไซต์ด้วย Tag เพื่อกำกับข้อความหรือภาพที่อยู่บนหน้าเว็บไซต์ ปัจจุบันได้พัฒนาสู่เว็บไซต์ที่มีความฉลาดแต่โดยปกติในท้องตลาดทั่วไปการใช้งานเว็บไซต์ต้องการนำเสนอข้อมูลหรือใช้เพื่อรับข้อมูลจากผู้ใช้ทั้งสองด้านบนมาตรฐาน Web 2.0 ของหน่วยงาน W3C[1]จากรูปที่ 2.1 เห็นได้ว่าเป็นการใช้เบราว์เซอร์ (Browser) ร้องขอข้อมูลผ่านโปรโตคอล HTTP จาก Web Server System ที่เรียกว่า HTTP request ซึ่งทางเซิอร์ฟเวอร์จำต้องทำการสนองตอบการร้องขอ โดยลงเข้าลึกไปสู่พื้นที่เก็บข้อมูลและนำกลับขึ้นมาสำหรับข้อมูลที่เป็นสาธารณะทั่วไปที่บุคคลภายนอกสามารถเข้าถึงได้หรือสำหรับข้อมูลสวงนเฉพาะในกรณีได้รับอนุญาตให้เข้าถึงข้อมูลได้ เพื่อส่งกลับไปยังผู้ร้องขอหรือเรียกใช้งานดังนั้นถ้าไม่สามารถเข้าถึงข้อมูลที่อยู่ในเบื้องลึกได้การนำเสนอข้อมูลไปยังผู้ใช้ก็ไม่ปรากฏข้อมูลดังกล่าว ซึ่งสาเหตุมีได้หลายประการ เช่นข้อมูลมีอยู่แต่เส้นทาง (Path) เข้าถึงไม่ถูกต้องหรือข้อมูลไม่มีเพราะมีการปรับปรุงข้อมูลแต่ลิงค์บางอันไม่ได้ถูกยกเลิกหรือลบทิ้ง ข้อมูลมีแต่ชื่อของข้อมูลที่ถือว่าเป็น Object นั้น ๆ แต่ไม่ปรากฏ Object ที่ตรงเพราะมีการแยกประเภทข้อมูลผิดพลาดหรือไม่ถูกต้องเป็นต้น



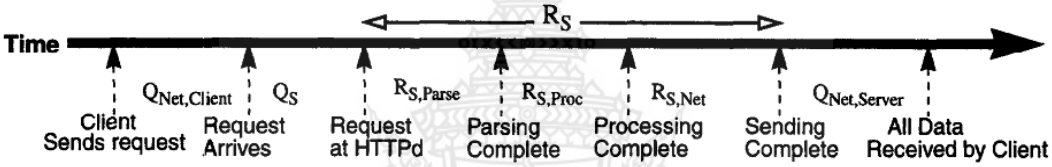
รูปที่ 2.1 การเชื่อมโยงและตอบสนองของ HTTP ในการใช้งานข้อมูลบนเครื่อง Server[3]

การเข้าถึงเหล่านี้ผ่านการทำงานของโปรแกรมกลุ่มสคริปต์ที่ใช้เพื่อเชื่อมทางเข้าออก Common Gateway Interface (CGI Scripts) ที่ตอบสนองต่อการร้องขอได้รวดเร็วทำให้เกิดกระบวนการ HTTP response ขึ้นจาก

การร้องขอผ่านเบราว์เซอร์ (Browser) ที่เครื่องผู้ใช้งานขั้นปลาย (End-user) หรือระบบ Client System ที่อยู่บนเครื่องลูกข่ายกรณีที่ผู้ใช้มีการคลิกเปิดพร้อมกันหลาย ๆ เบราว์เซอร์บนเครื่องเดียวกัน ก็ทำให้เกิดสภาวะดังกล่าววนซ้ำไปมาหลาย ๆ การร้องขอบนเครื่องคอมพิวเตอร์ การเกิดการร้องขอปกติหนึ่งครั้งก็มีการตอบสนองจากเว็บเซิร์ฟเวอร์จากนั้นระบบก็นิ่งเรียกรอการร้องขอถัดไปสภาวะนี้ว่าเป็น Stateless การใช้งานสามารถเกิดบนโทรศัพท์มือถือ แท็บเล็ตหรือที่ปัจจุบันเรียกอุปกรณ์สมาร์ต (Smart devices)

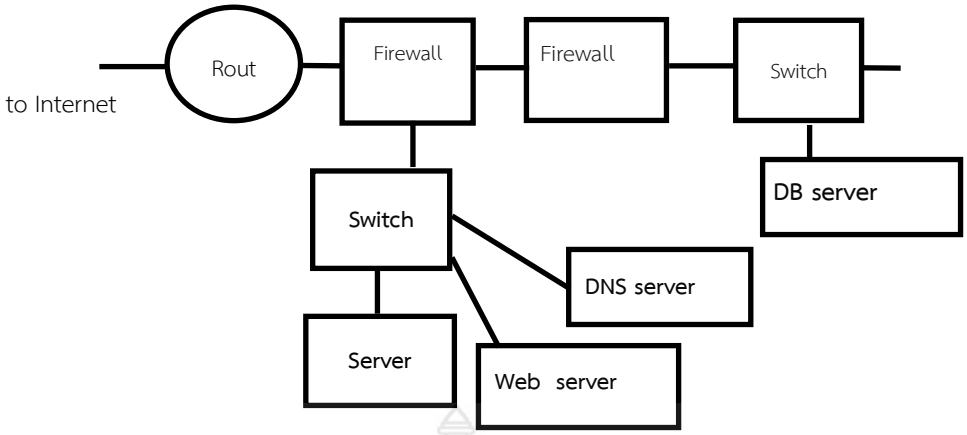
2.2 กระบวนการเข้าถึงข้อมูลและการวิภัช

โดยธรรมชาติเช่นตัวอย่างการสนทนาฝั่งคู่สนทนาต้องรับฟังและทำความเข้าใจในประโยคที่คู่สนทนาพูดเช่นเดียวกับการสื่อสารของเครื่องเซิร์ฟเวอร์กับในรูที่ 2.2 เป็นลำดับของการร้องขอซึ่งเป็นช่วงแต่ละเวลาที่ทำให้เกิดการวิภัชและการนำเสนอข้อมูลไปยังหน้าเว็บ เห็นได้ว่าแต่ละส่วนต้องมีการใช้เวลาในการประมวลผลซึ่งส่งผลกระทำ



รูปที่ 2.2 ช่วงเวลาในการเกิดเหตุการณ์ของการใช้งานระหว่างผู้ใช้เว็บ[3]

การเชื่อมโยงของ Web server นั้นขอยกตัวอย่างโครงสร้างที่นิยมใช้ออกแบบเพื่อใช้งานคือการทำ DMZ (De-militarized Zone) ดังในรูปที่ 2.3 ดังนั้นในการร้องขอจากผู้ใช้หรือการทำจากภายนอกของระบบสู่ภายในต้องมีการผ่านไฟร์วอลล์ (Firewall) ข้อมูลจากตัวเว็บเซิร์ฟเวอร์ (Web server) ถูกส่งต่อผ่านไฟร์วอลล์ออกไปสู่ภายนอก ดังนั้นในการค้นหาข้อมูลจากเว็บไซต์จึงต้องผ่านไฟร์วอลล์เข้าไปและดึงข้อมูลบางส่วนออกมา สิ่งที่เชื่อมโยงหรือลิงค์จากภายในสู่ภายนอกและการเชื่อมโยงกันภายในเว็บเดียวกันหรือต่างเซิร์ฟเวอร์กันที่อยู่ภายใต้ DMZ หรือนอก DMZ หรือนอกไฟร์วอลล์ เพื่อนำข้อมูลของการลิงค์ถึงกันมาแสดง เพื่อตรวจสอบจำนวนลิงค์ที่มี จำนวนลิงค์ที่ชี้ไปหาเพื่อใช้ในเวลาค้นหาผ่านเครื่องมือสืบค้นที่เรียกว่า Search engine



รูปที่ 2.3 ตัวอย่างผังโครงสร้างของ Sever บนระบบ DMZ

เครื่องมือในการรวบรวมข้อมูลจาก Server ของคณะวิศวกรรมศาสตร์

Chrome Browser

MS IE Browser

Firefox Browser

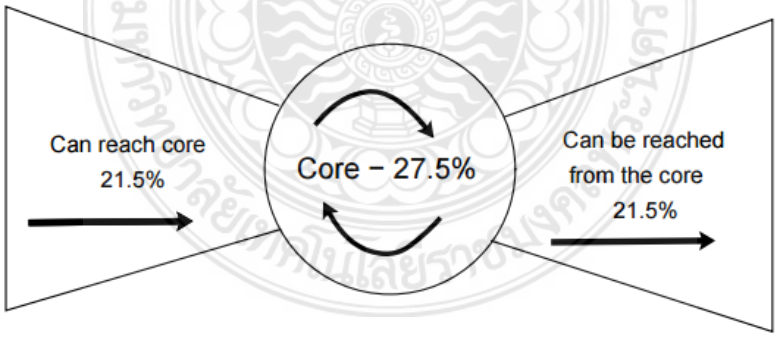
Python

R Program

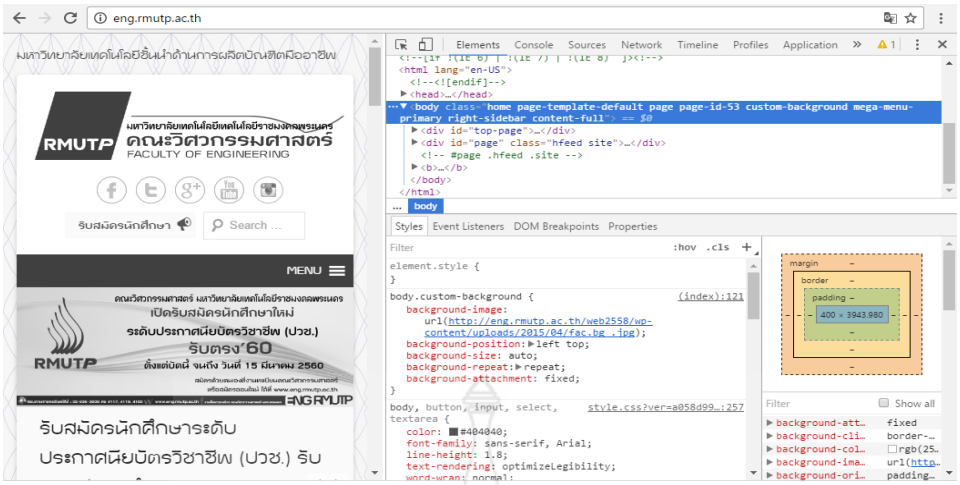
MS Excel

2.3 การทบทวนวรรณกรรมที่เกี่ยวข้อง

การใช้งานเว็บไซต์นั้นในระบบ Internet มีปริมาณจำนวนมหาศาล แต่โดยหลักการพื้นฐานมาจากการพิจารณา 3 องค์ประกอบหลักคือ Input Core และส่วน Output ที่เรียกว่าแบบ Bow-tie[6]



รูปที่ 2.4 ผังแบบ Bow-tie ที่บ่งถึงความเชื่อมโยงระหว่างอินพุตแกน และเอาต์พุต[6]



รูปที่ 2.5 ผังหน้าของเว็บไซต์คณะวิศวกรรมศาสตร์ด้านขวามือคือโค๊ดโปรแกรมที่แสดงหน้าเว็บ

รูปที่ 2.6 ตัวอย่าง HTML ส่วนหนึ่งของหน้าเว็บไซต์คณะวิศวกรรมศาสตร์

จากเฟรมในรูปที่ 2.5 และบางส่วนของโค๊ดในรูปที่ 2.6 แสดงให้เห็นถึงส่วนของเฟรมบนหน้าเว็บไซต์การจัดองค์ประกอบต่าง ๆ และการมีสคริปต์ในหน้าเว็บ ส่วนในรูปที่ 7 เป็นตัวอย่างการทำให้เกิดภาพพื้นหลังของหน้าเว็บซึ่งเป็นการอ้างชื่อไฟล์รูปภาพด้วยลิงค์ ซึ่งจะมีลักษณะแบบนี้ในหลาย ๆ หน้าเว็บเพราะผู้ออกแบบต้องการให้หน้าเว็บไซต์มีรูปภาพทำให้ง่ายต่อการใช้งานและการจดจำดังนั้นกระบวนการในการเก็บรวบรวมข้อมูลจึงสำคัญและต้องมีการแยกแยะในส่วนต่าง ๆ ออก

```
body.custom-background {
  background-image: url(http://eng.rmutp.ac.th/web2558/wp-content/uploads/2015/04/fac.bg_.jpg);
  background-position: left top;
  background-size: auto;
  background-repeat: repeat;
  background-attachment: fixed;
}
```

รูปที่ 2.7 ตัวอย่างลิงค์ของรูปพื้นหลัง

เพื่อนำข้อมูลที่ต้องการพิจารณาหรือตรวจสอบออกมาแสดงประกอบการพิจารณาในการปรับปรุงหน้าเว็บรวมถึงรายละเอียดในแต่ละระดับของหน้าเว็บ

บทที่ 3

การดำเนินการวิจัย

3.1 ศึกษาการเชื่อมโยง

จากการศึกษาถึงโครงข่ายของเว็บไซต์ทำให้ทราบว่าโครงข่ายของเว็บไซต์มีการเชื่อมโยงเป็นลักษณะที่ซับซ้อนบนพื้นฐานของการเชื่อมต่อกันแบบ Mesh แต่โดยสภาพทั่วไปมีการพิจารณาถึง โดยต้องเข้าใจการใช้งานของข้อมูลผ่านโปโตคอลที่สำคัญของการใช้งานเว็บไซต์ แต่อาจมีการโปรโตคอลอื่น ๆ ร่วมในการทำงาน การใช้งานแต่ละโปรโตคอลจำเป็นต้องมีพอร์ตรองรับ เช่น HTTP ใช้งานพอร์ต 80 HTTPs ใช้งานพอร์ต 443 เป็นต้น

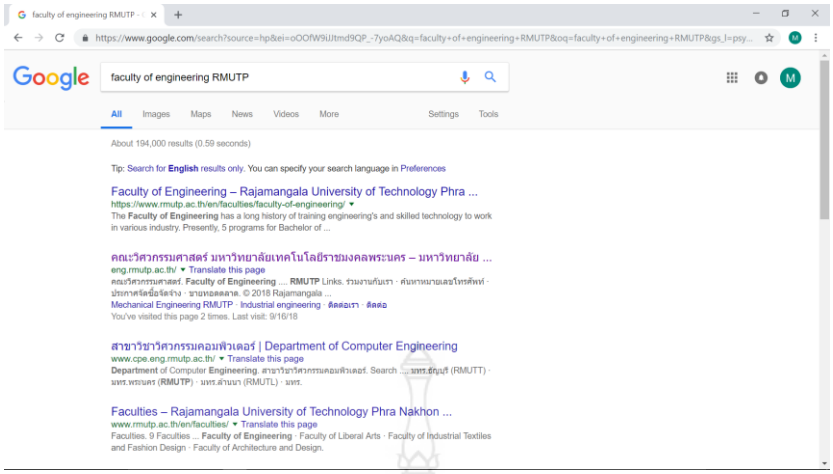
การติดต่อสื่อสารของแต่ละภาระงานของเลเยอร์ในการสื่อสารไม่ว่าในส่วนของ OSI Reference Model ที่มีจำนวนชั้นทั้งสิ้น 7 เลเยอร์อันเป็นบรรทัดฐานในการพิจารณาแต่ละส่วนของระบบเครือข่ายตามชั้นเลเยอร์เพื่อให้แยกการทำงานชัดเจน หรือ TCP/IP Layers ที่มีทั้งสิ้น 4 เลเยอร์เพื่อสะดวกในการพิจารณาเฉพาะข้อที่นำรวมกันให้สะดวกต่อ ในส่วนของเว็บซึ่งเกี่ยวข้องโดยตรงกับ W3C นั้นในการเข้าถึงก็มีทั้งโดยตรงเพราะข้อมูลเปิดเผยเป็นสาธารณะกับข้อมูลเฉพาะผู้เกี่ยวข้อง ซึ่งการศึกษาวิจัยนี้ต้องพิจารณาร่วมเพราะการเข้าถึงข้อมูลเฉพาะนั้นทำให้เกิดผลลัพธ์คือไม่สามารถเข้าถึงข้อมูลได้นั้นคือสาเหตุของการเกิดรหัสหมายเลขความผิดพลาดที่เกิดได้ ดังนั้นการศึกษานี้จึงต้องมีข้อควรระวังหลายจุด ที่อาจทำให้ได้ข้อมูลไม่ตรงหรือทับซ้อนหรือแยกแยะยากกว่าปกติ แต่ข้อดีของโปรแกรมที่นำมาประยุกต์ใช้งานมีศักยภาพสูงในการทำงานในด้านนี้โดยตรง

3.2 การติดตั้งโปรแกรมช่วยในการทำงาน

โปรแกรมที่จำเป็นในการใช้งานมีดังต่อไปนี้

3.2.1 Chrome เบราว์เซอร์ เป็นที่นิยมใช้งานและมีการทำงานที่รวดเร็วสะดวกในการใช้งานสามารถดาวน์โหลดฟรีเพื่อติดตั้งในเครื่องคอมพิวเตอร์และมีการปรับปรุงคุณสมบัติตลอดเวลาทำให้มีจุดอ่อนน้อยน่าใช้งาน จากรูปที่ 3.2 เห็นได้ว่าใช้เวลาในการสืบค้นข้อมูลเพียง 0.59 วินาที เป็นที่นิยมใช้งานในการท่องอินเทอร์เน็ต

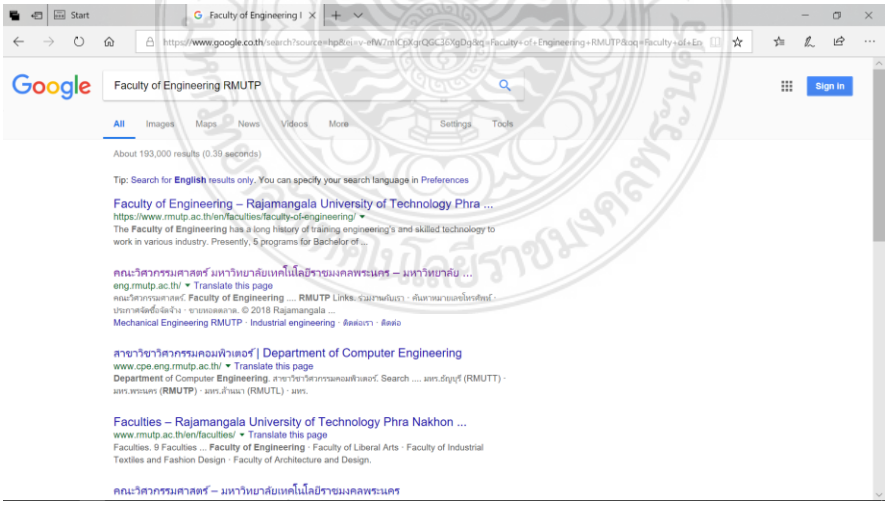
รูปที่ 3.1 สัญลักษณ์ของโครมเบราว์เซอร์



รูปที่ 3.2 เบราว์เซอร์ Chrome กับการสืบค้นข้อมูล

3.2.2 ie เบราว์เซอร์ ของค่าย Microsoft ที่สะดวกในการใช้งานถ้าใช้ระบบปฏิบัติการ MS-Windows ซึ่งการออกแบบเว็บไซต์ของแต่ละหน่วยงานก็ตอบสนองต่อเบราว์เซอร์บางรุ่นบางยี่ห้อได้ดี บางรุ่นบางยี่ห้ออาจไม่สะดวกเพราะการปรับปรุงเบราว์เซอร์มีตลอดเวลาของทุกค่าย ie เบราว์เซอร์ดังในรูปที่ 3.4 เห็นได้ว่ามีความเร็วในการสืบค้นเพื่อนำเสนอที่ 0.39 วินาที

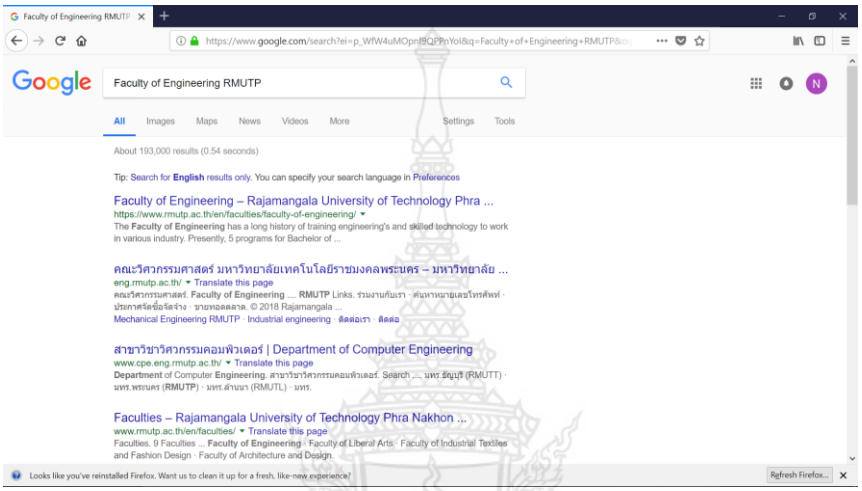
รูปที่ 3.3 สัญลักษณ์ไออี



รูปที่ 3.4 เบราว์เซอร์ ie กับการสืบค้นข้อมูล

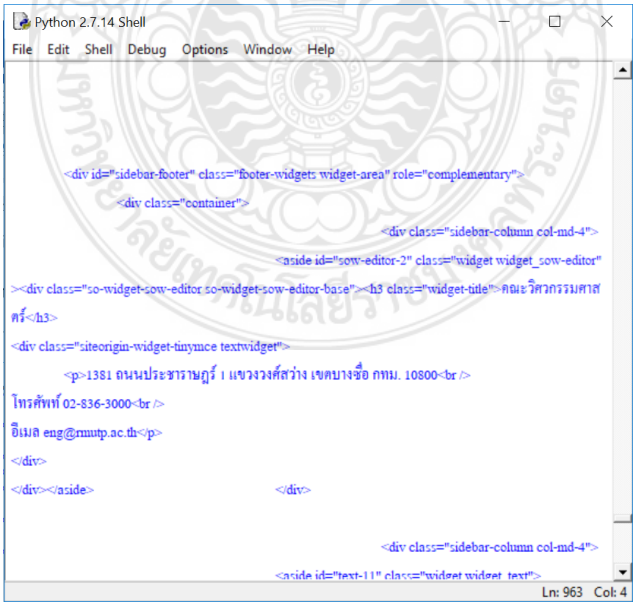
3.2.3 Firefox เบราว์เซอร์ อีกค่ายที่เป็นที่นิยมเพราะมี Add-on ที่หลากหลายสะดวกต่อผู้ใช้งาน

รูปที่ 3.5 สัญลักษณ์ไฟล์ฟอกซ์



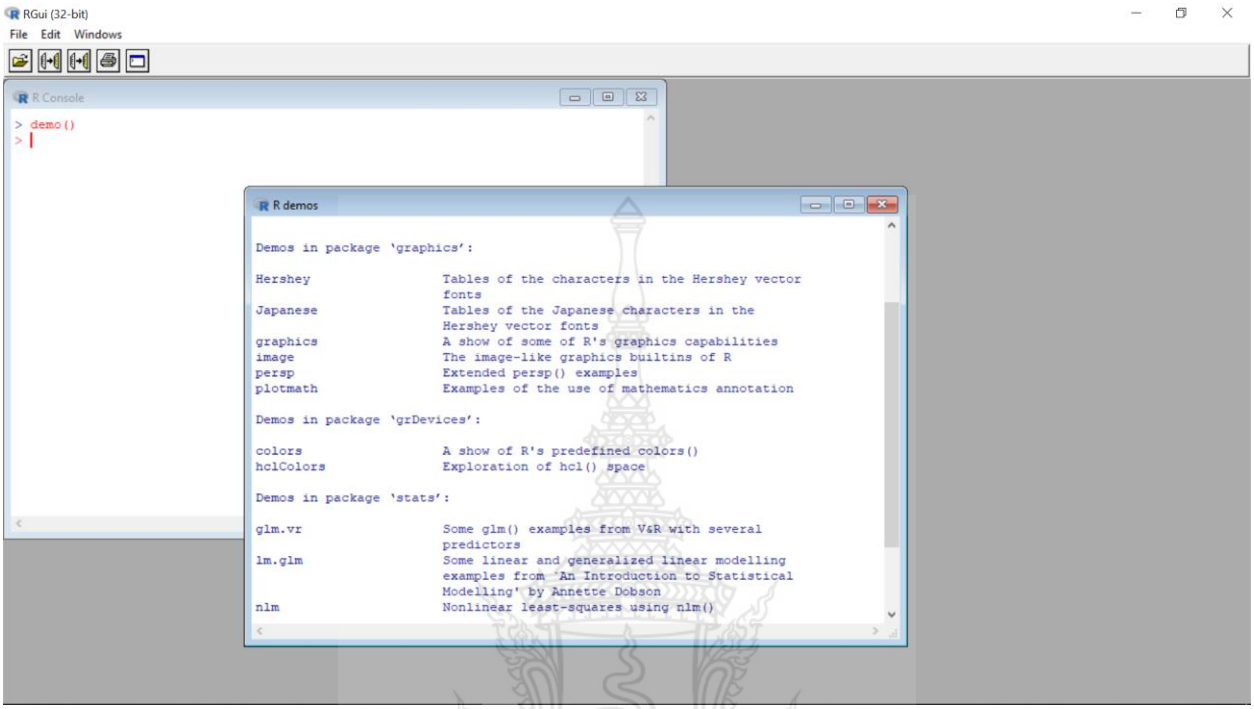
รูปที่ 3.6 การใช้ Firefox สืบค้นข้อมูลจาก Google.com

3.2.4 Python เป็นโปรแกรมที่มีศักยภาพในการทำงานเกี่ยวกับอินเทอร์เน็ตได้เป็นอย่างดี ภาษา Python เป็นภาษาสคริปต์ สะดวกในการใช้งาน และสามารถดึงข้อมูลจากหน้าเว็บได้เป็นอย่างดีจากรูปที่ 3.7



รูปที่ 3.7 การใช้โปรแกรม Python ดึงข้อมูลจากหน้าเว็บไซต์คณะวิศวกรรมศาสตร์ มทร.พระนคร

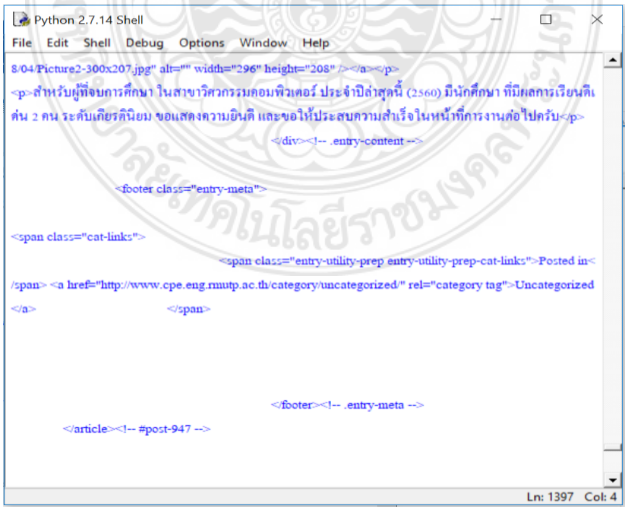
3.2..6 R เป็นภาษาโปรแกรมมิ่งที่มีการใช้งานอย่างกว้างขวางเนื่องเพราะฟรี ทั้งยังมีการพัฒนาอย่างต่อเนื่องตลอด



รูปที่ 3.8 การเรียกใช้งาน R บน MS-Windows 10

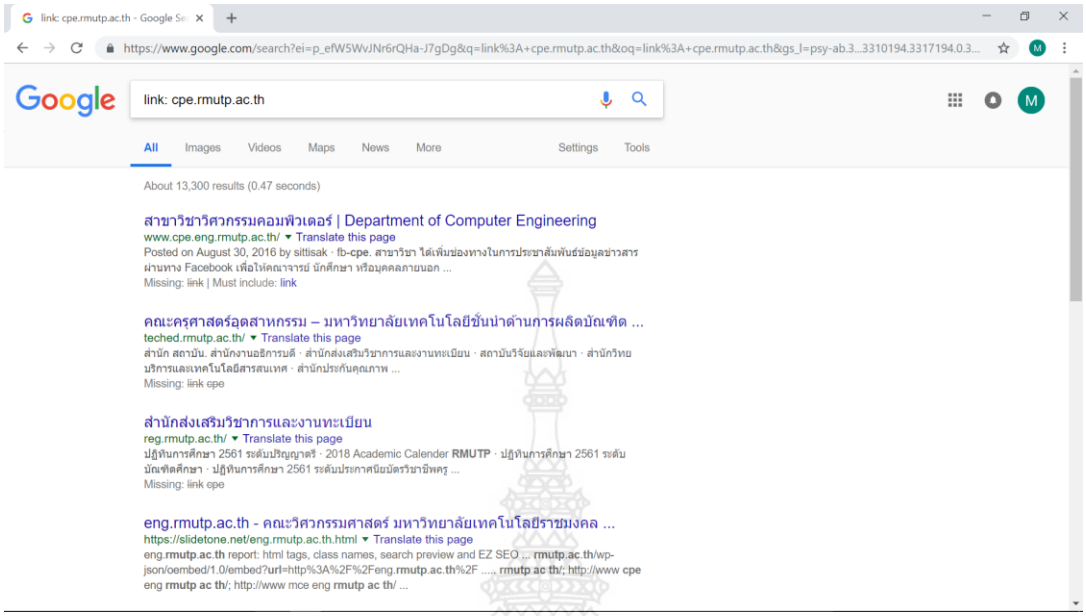
3.3 การเตรียมการทดสอบเครื่องมือ

3.3.1 การทดสอบ Python เพื่อเรียกดูข้อมูลจากเว็บไซต์โดยตรง



รูปที่ 3.9 ตัวอย่างการเรียกดูข้อมูลจากหน้าเว็บ

3.3.2 การทดสอบเบราว์เซอร์เพื่อตรวจสอบ



รูปที่ 3.10 การใช้ Search Engine ในการหาความสัมพันธ์การเชื่อมโยง

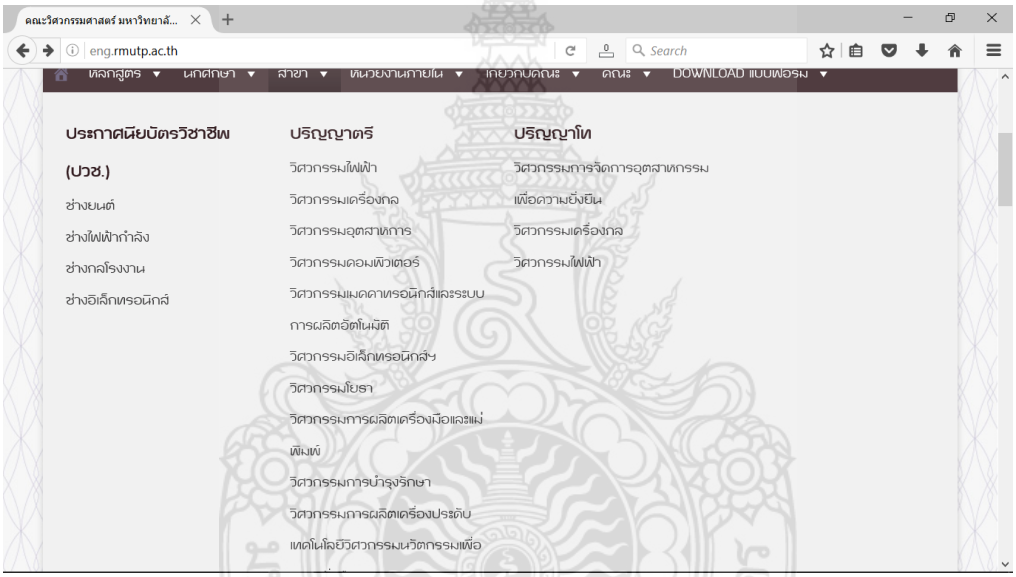
บทที่ 4

ผลการวิจัย

4.1 ผลการเก็บข้อมูล

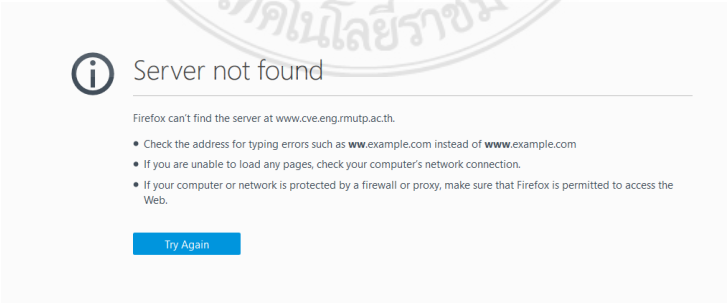
ในการดำเนินการวิจัยสี่เดือนระหว่างช่วงของแต่ละเดือนจากเดือนกุมภาพันธ์ 2561 ถึง พฤษภาคม 2561 ติดต่อกันแล้วนำมาหาค่าเฉลี่ยด้วยโปรแกรม R และ Python โดยพิจารณายึดรหัสหมายเลขความผิดพลาดของการเข้าถึงบนหน้าเว็บซึ่งแสดงในภาคผนวก สรุปได้ข้อมูลที่สำคัญดังนี้

4.1.1 หน้าเว็บไซต์ของคณะวิศวกรรมศาสตร์มีรายการของลิงค์ระดับปริญญาตรีทั้งสิ้น 11 ลิงค์



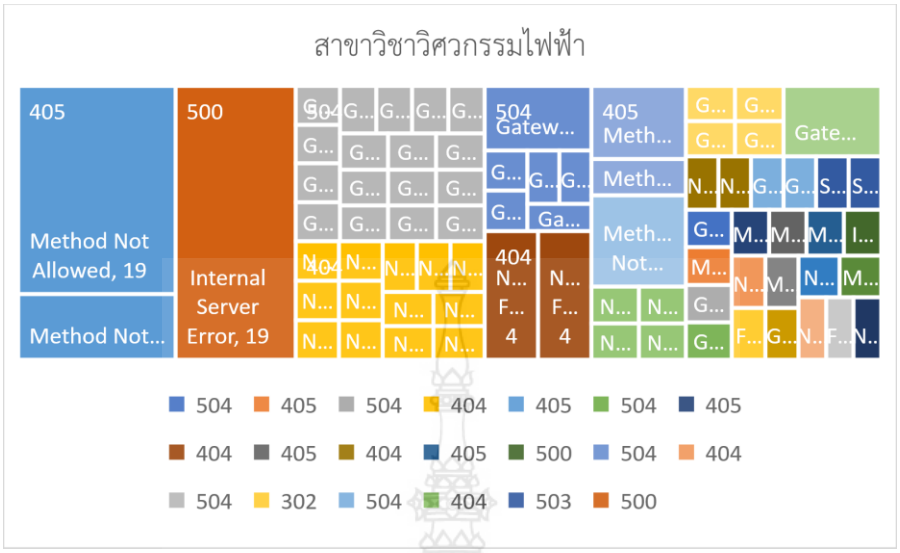
รูปที่ 4.1 รายการลิงค์ของสาขาวิชาต่าง ๆ

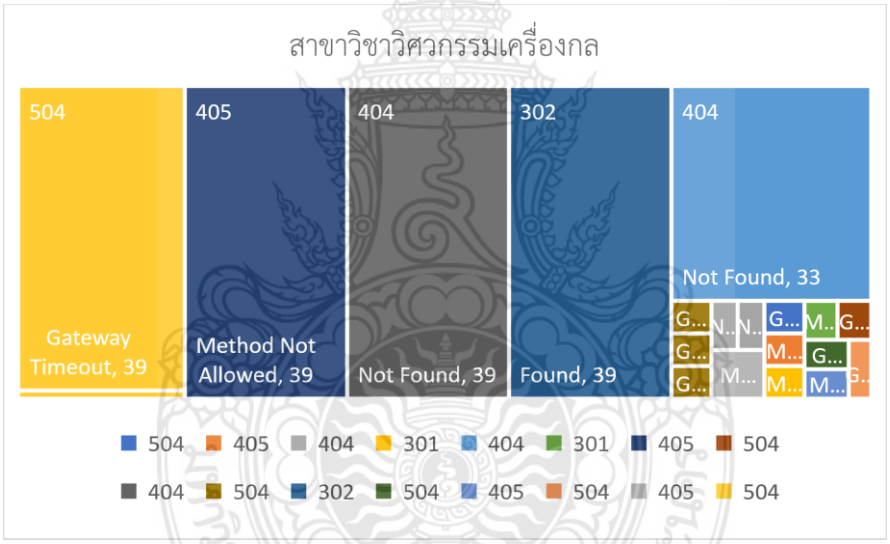
โดยมี 2 ลิงค์ที่ไม่สามารถเข้าถึงได้เลยดังรูปที่ 4.2

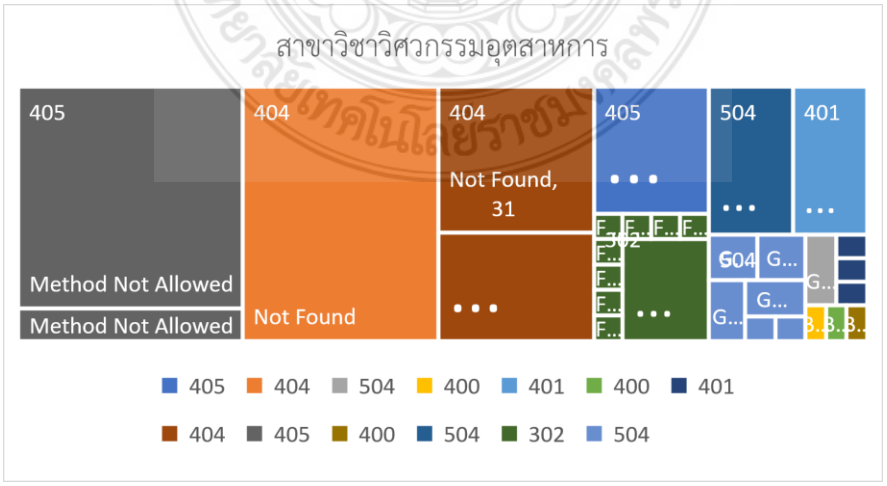


รูปที่ 4.2 กรณีไม่พบเซิร์ฟเวอร์หรือลิงค์มีปัญหาอย่างสิ้นเชิง

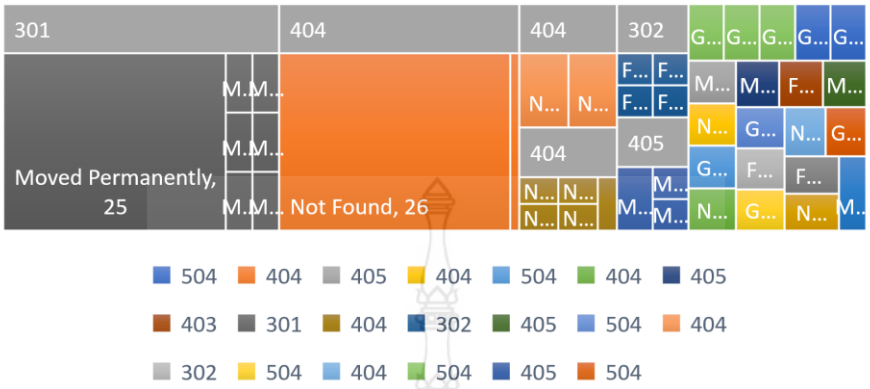
4.1.2 ความผิดพลาดของลิงค์สาขาวิชาของคณะวิศวกรรมศาสตร์ มทร.พระนคร



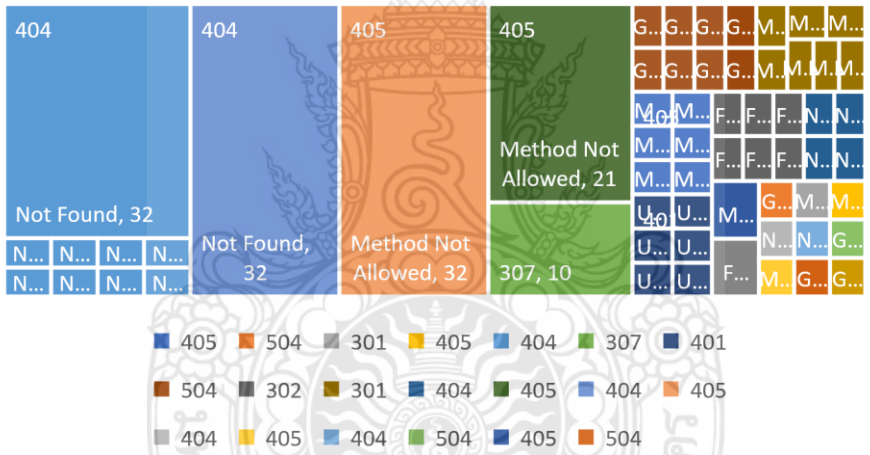




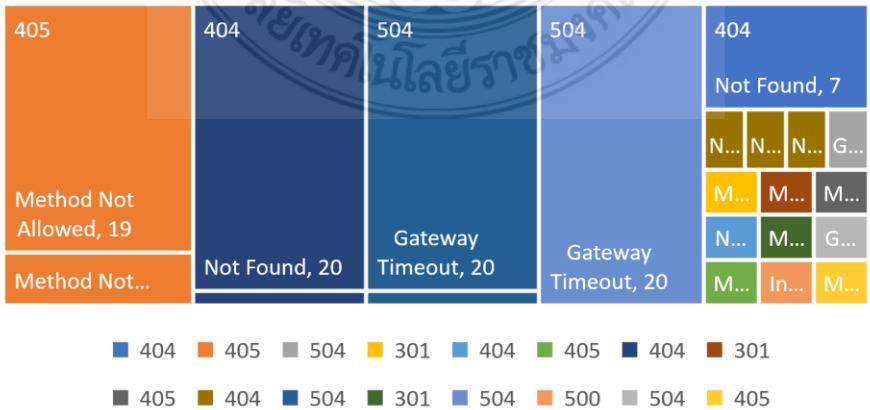
สาขาวิชาวิศวกรรมคอมพิวเตอร์

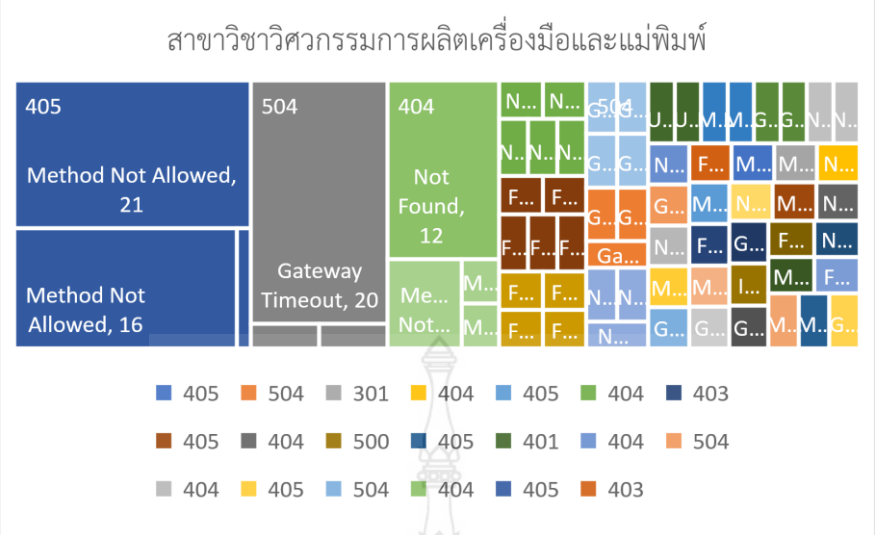


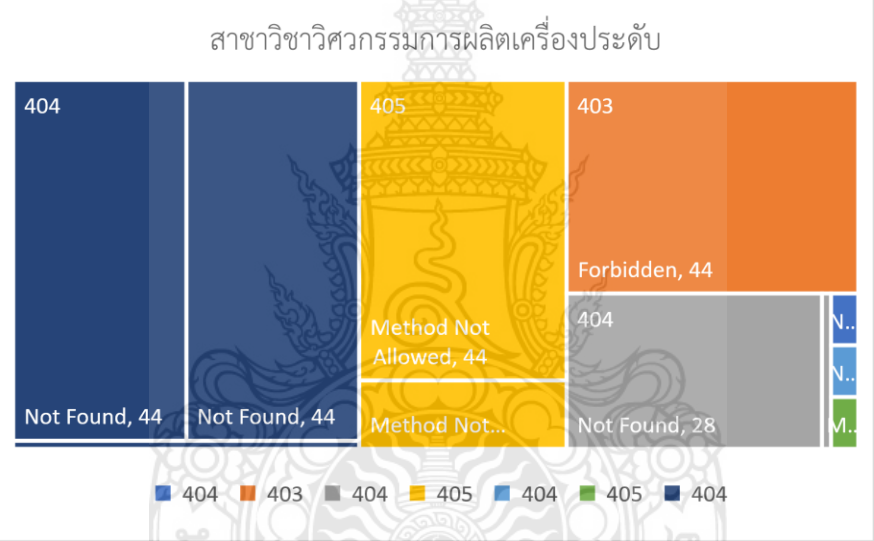
สาขาวิชาวิศวกรรมแมคคาทรอนิกส์และระบบการผลิตอัตโนมัติ

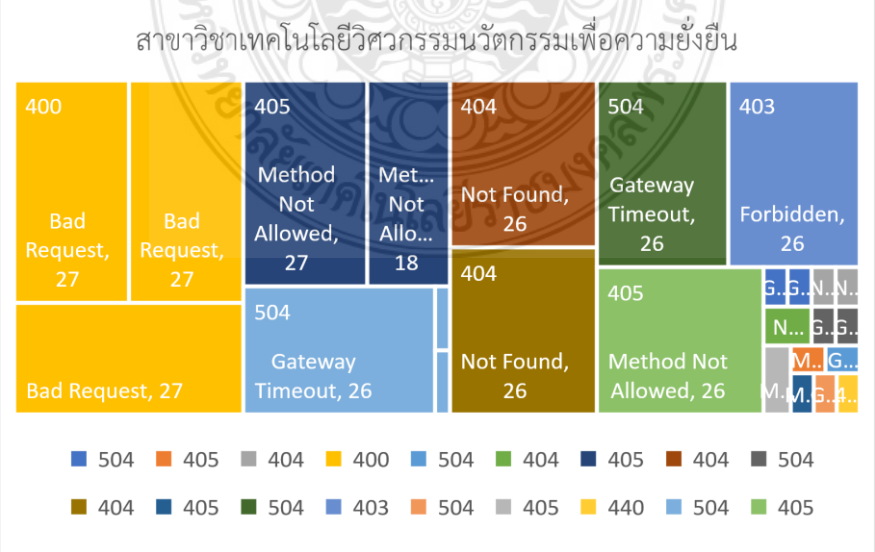


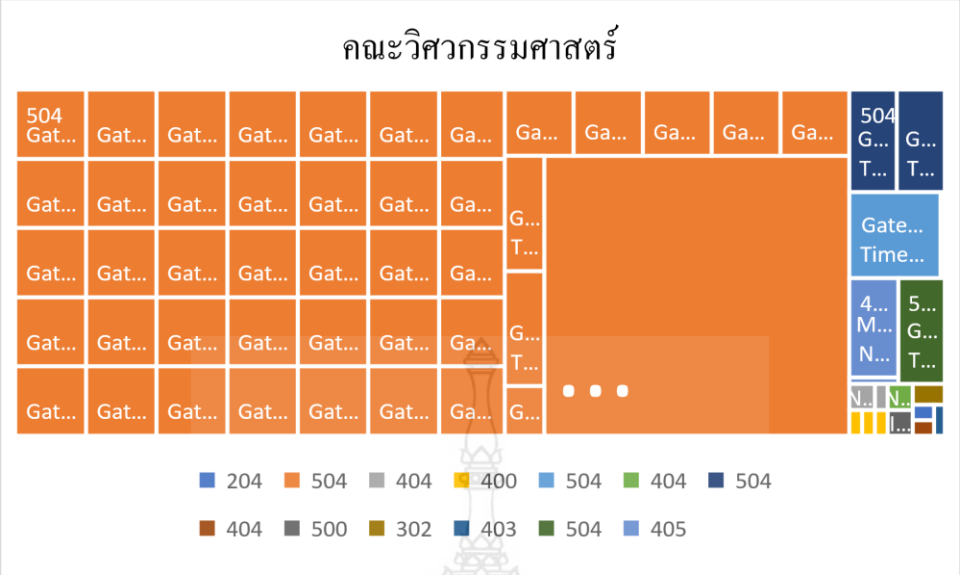
สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์ฯ




สาขาวิชาวิศวกรรมการผลิตเครื่องมือและแม่พิมพ์
405
Method Not Allowed, 21
Method Not Allowed, 16
504
Gateway Timeout, 20
404
Not Found, 12
405 504 301 404 405 404 403
405 404 500 405 401 404 504
404 405 504 404 405 403


สาขาวิชาวิศวกรรมการผลิตเครื่องประดับ
404
Not Found, 44
Not Found, 44
405
Method Not Allowed, 44
Method Not...
403
Forbidden, 44
404
Not Found, 28
404 403 404 405 404 405 404


สาขาวิชาเทคโนโลยีวิศวกรรมนวัตกรรมเพื่อความยั่งยืน
400
Bad Request, 27
Bad Request, 27
Bad Request, 27
405
Method Not Allowed, 27
Met... Not Allo... 18
504
Gateway Timeout, 26
404
Not Found, 26
404
Not Found, 26
504
Gateway Timeout, 26
403
Forbidden, 26
405
Method Not Allowed, 26
504 405 404 400 504 404 405 404 504
404 405 504 403 504 405 440 504 405



4.2 การประมวลผลจำนวนรหัสความผิดพลาดของลิงค์

การประมวลผลด้วยโปรแกรม R และ Python เพื่อหาข้อมูลด้วยวิธีการเข้าถึงข้อมูลในแต่ละลิงค์ โดยพบว่าเฉพาะหน้าเว็บของคณะวิศวกรรมศาสตร์มีความผิดพลาดเกิดขึ้นทั้งหมด 987 ลิงค์ซึ่งแตกต่างกันออกไปส่วนใหญ่เป็นรหัสความผิดพลาดหมายเลข 504 (Gateway Timeout) มากสุดเฉลี่ยมากกว่า 85 เปอร์เซ็นต์รองจากนั้นเป็นรหัสความผิดพลาด 400 (Bad Request) 404 (Not Found) และเป็นส่วนน้อยมากเป็น 403 (Forbidden) 204 (No Content) ตามลำดับ ส่วนลิงค์ของหน้าเว็บแต่ละสาขาวิชาฯ ได้ข้อสรุปของข้อมูลที่มีค่ารหัสความผิดพลาดมากกว่าสิบลิงค์ขึ้นไปมีดังนี้

ตารางที่ 4.1 จำนวนลิงค์ที่เกิดความผิดพลาดในการสืบค้นบนเว็บของแต่ละสาขาวิชา

รหัสหมายเลขความผิดพลาด	จำนวน	รหัสหมายเลขความผิดพลาด	จำนวน
301	46	403	77
302	82	404	498
307	10	405	422
400	81	504	230
401	27	รวมทั้งสิ้น 1,473 ลิงค์	

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 ผลวิเคราะห์ข้อมูล

จากการสืบค้นบนระบบอินเทอร์เน็ตของเว็บไซต์คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร พบว่ามีลิงค์ที่เกิดความผิดพลาดในการเข้าถึงจากหลายสาเหตุ แต่ที่มีจำนวนมากนับร้อยลิงค์ขึ้นไปคิดเป็นประมาณ 78 เปอร์เซ็นต์ของสภาวะที่เกิดขณะทำการเก็บข้อมูล พอแยกโดยสังเขปตามลำดับได้ดังนี้

404 Not Found (มากสุดเป็นอันดับต้น)

405 Method Not Allowed

504 Gateway Timeout

จากการข้อมูลที่ได้มีปริมาณคิดเป็น 78 เปอร์เซ็นต์ของทั้งหมดของแต่ละหน้าเว็บของแต่ละสาขาวิชานำมารวมกัน นอกจากนี้การหาการเชื่อมโยงภายในออกสู่ภายนอกและจากนอกมาของเว็บไซต์คณะวิศวกรรมศาสตร์รวมมีค่าประมาณ 7,500 เว็บ

5.2 ข้อเสนอแนะ

เนื่องจากการปรับปรุงเว็บไซต์มีความจำเป็นต้องกระทำตลอดเวลา ดังนั้นการเกิดข้อผิดพลาดของลิงค์ของข้อมูลหรือลิงค์การเชื่อมโยงภายในเซิร์ฟเวอร์และภายนอกจึงมีความไม่แน่นอนแปรเปลี่ยนได้ตลอดเวลา เพียงแต่การดูแลรักษาของผู้ปฏิบัติงานต้องกระทำอย่างระมัดระวังและสม่ำเสมอ เนื่องจากข้อมูลที่สอบถามผู้ดูแลเว็บของแต่ละสาขาวิชาฯมีภาระงานที่หลากหลายมาก ทำให้ภาระการทำหน้าที่และความสม่ำเสมอไม่ครอบคลุม เนื่องจากข้อมูลที่ได้บ่งชี้ให้เห็นว่ามีจำนวนของลิงค์ที่มีข้อผิดพลาดประมาณ 1,500 ลิงค์โดยรวมทำให้ต้องมีการตรวจสอบเนื้อหาและลิงค์ต่าง ๆ ในหน้าเว็บไซต์อย่างสม่ำเสมอ การปฏิบัติงานควรมีเอกสารกำกับเพื่อให้เกิดการทวนสอบได้

เอกสารอ้างอิง

[1] WCAG 2.0, W3C, https://www.w3.org/TR/WCAG20/.

[2] W3C Web Content Accessibility Guidelines 2.0 https://www.w3.org/WAI/WCAG2-Conformance.

[3] Virgilio A.F. Almeida, Daniel Menasce, Capacity Planning: Essential Tool for Managing Web Services, IEEE IT Pro August 2002.

[4] Mark Levene and Alexandra Poulovassilis, *Web Dynamics*, Springer Verlag 2004.

[5] Jon M. Kleinberg, Authoritative sources in a hyper-linked environment, Proceeding ACM-SIAM Symposium on Discrete Algorithms, May 1998.

[6] Network Working Group, *Hypertext Transfer Protocol, RFC2616*, Standard Track, June 1999.

[7] Ed Bott, Microsoft Office Inside Out: 2013, Microsoft Press Inc., 2013.

[8] Scott A. Helmers, Microsoft Visio 2013 Step by Step, Microsoft Press Inc., 2013.

[9] M. O. Faruque Sarker, *Python Network Programming Cookbook*.

[10] Hadley Wickham, Garrett Grolemund, R for Data Science, O'Reilly, December, 2016.

[11] นิลมิต นิลาศ *การวิเคราะห์ลิงค์เว็บไซต์ของคณะวิศวกรรมศาสตร์ มทร.พระนคร* การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 3 พ.ศ.2561.