

การวิเคราะห์ลิงก์เว็บไซต์ของคณะวิศวกรรมศาสตร์ มทร.พระนคร

RMUTP Faculty of Engineering Website Link Analysis

นิลमित นิลาศ¹

¹สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
1381 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพมหานคร E-mail: nilamit.n@rmutp.ac.th

บทคัดย่อ

จากพัฒนาการของเว็บไซต์ มีความจำเป็นในการศึกษา รายละเอียดของการเชื่อมโยงลิงก์ของเว็บเพื่อให้ได้ข้อมูลมาประกอบในการพัฒนาเว็บไซต์จากเหตุที่ว่ากรณีที่โปรแกรมเมอร์ที่พัฒนาเว็บไซต์ มีหลายคนและมีการสร้างข้อมูลเพิ่มเติมหรือลบข้อมูลบางส่วนออกไปจากการปรับปรุงเว็บ ทำให้มีความจำเป็นในการศึกษาถึงลิงก์ของเว็บไซต์ที่ใช้งานของหน่วยงาน เพื่อหาว่าลิงก์ที่อยู่บนหน้าเว็บเพจรวมถึงลิงค์ด้านออกจากเว็บไซต์ของคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนครหรือภายในเซิร์ฟเวอร์หรือต่างเซิร์ฟเวอร์มีจำนวนเท่าใด และมีการเปลี่ยนแปลงอย่างไรในระหว่างช่วงการทำวิจัยในระยะ 1 ไตรมาสรวมถึงหน้าเว็บเพจที่เชื่อมโยงอย่างครบถ้วนและหน้าเว็บเพจที่ขาดหายไปจากเส้นทางการเชื่อมโยง (Path) ของไคลเรคตอรีที่เก็บไฟล์ของหน้าเพจ ซึ่งพบว่ามีลิงค์เข้า และออกเป็นจำนวน 211 ลิงค์ที่มีข้อผิดพลาด โดยมีรหัส 504 มากสุดประมาณ 96% โดยมี HTML เพิ่มขึ้น 147 มี Script Links ลดลง 2

คำสำคัญ: ลิงค์ วิเคราะห์เว็บ เว็บไซต์เมตริก วิเคราะห์ลิงค์

Abstract

For feature website development necessary to study in to the link details. The website data collection essential for website improvement. The website programmers have special roles to add, to delete, and to change, etc. from the step forward to represent content from the web server to several clients. The website codes consist of many programming languages. The adjustment website is important to adapt the content to suitable on time, especially, inside links and outside links of RMUTP Faculty of Engineering. The comparison results from started several data collections in a started week then 3 months later collected data again on average to change as the HTML files:+147, mail to Links:-2, graphic links:-1, style sheet links:-1, and script links:-2.

The RMUTP engineering website shown 96% of the 504 code on the 211 link errors.

Keywords: link, website analysis, Webometrics, link analysis

1. บทนำ

สภาวะการใช้งานของเว็บไซต์มีมากมายหลายลักษณะแต่โดยพื้นฐานของเว็บไซต์มีองค์ประกอบคือการตอบสนองต่อผู้ใช้งานในด้านข้อมูลที่เกิดจากความต้องการของผู้ใช้ส่วนใหญ่ผ่านจากหน้าเว็บ Browser [1-2] โดยเนื้อในทั่วไปส่วนใหญ่เป็นข้อความเนื้อหาสนทนาด้วยภาษาไทยและภาษาอังกฤษ ส่วนภาษาที่นอกเหนือจากนี้มีปริมาณน้อยสำหรับเว็บไซต์ของคนไทย โดยเฉพาะเว็บไซต์ของมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร คณะวิศวกรรมศาสตร์ประกอบด้วยข้อมูลหลัก ๆ เป็นเนื้อความในสองภาษาคือภาษาไทยและภาษาอังกฤษ

การพิจารณาเว็บนั้นเสมือนกับกราฟที่มีการโยงหากันมีทั้งเข้าหาและมีทั้งออกมาจากจุดหรือ โหนด โดยให้ $W_p[i, j]:[5]$ แทนด้วยเว็บเพจที่มีลิงค์จากหน้าหรือเพจ i ไปยังหน้าเพจ j ถ้ามีการเชื่อมโยงถึงกันให้เป็น '1' ถ้าไม่มีการเชื่อมโยงถึงกันแทนด้วย '0' ทำให้เราสามารถหาจำนวนการเชื่อมโยงได้ โลกภายนอกที่เชื่อมโยงถึงมีทั้งภายในเว็บไซต์ที่พิจารณาออกสู่โลกภายนอกและทั้งมีส่วนของลิงค์ที่เชื่อมโยงภายในระหว่างหน้าเว็บเพจที่เชื่อมโยงลิงค์ไปมาทำให้เกิดจำนวนลิงค์มากขึ้นตามอายุที่เปิดบริการส่วนหนึ่ง ตามการจัดการในลิงค์ต่าง ๆ ของโปรแกรมเมอร์อีกส่วนหนึ่งและจากการเลิกใช้งานลิงค์นั้นไม่ว่าโดยการสิ้นสุดของข้อมูลในการให้บริการหรือการหมดอายุการแสดงผลจากเงื่อนไขของระบบเองส่วนหนึ่งทำให้เกิดมีภาวะของลิงค์ที่ไม่มีเชื่อมโยงได้จริงเกิดขึ้น และอีกส่วนคือลิงค์ยังถูกต้องแต่ไม่มีข้อมูลปลายทางที่มีอยู่จริงทำให้เกิดข้อผิดพลาดขณะที่ผู้ใช้งานเลือกคลิกจากหน้าเว็บเพจ การเกิดภาวะการนี้ทำให้เป็นภาระของเซิร์ฟเวอร์[4-6] โดยเปล่าประโยชน์แต่เป็นการเพิ่มภาระงานให้กับเซิร์ฟเวอร์เนื่องจากเมื่อมีการร้องขอ Request จากหน้าเพจเช่น GET ก็ทำให้เกิดการสนองตอบจากเซิร์ฟเวอร์นั้นคือทำให้เซิร์ฟเวอร์ใช้เวลาในการเข้าถึงข้อมูลตามลิงค์ที่ปรากฏอยู่แต่ผลลัพธ์คือไม่มีข้อมูลที่ต้องการ สูญเสียทรัพยากรของเครื่องทั้งระบบไฟฟ้าที่ใช้ อายุการทำงานของ Hard disk เวลาของการบริการของ CPU เครื่องเซิร์ฟเวอร์

บทความวิจัย

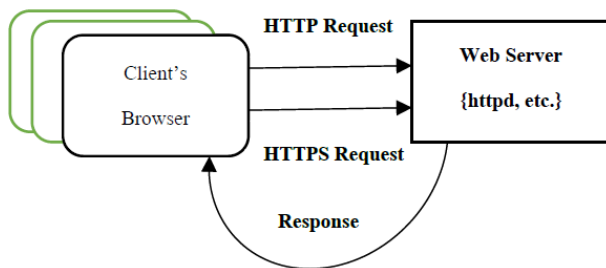
การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 3
Proceedings of the 3rd RMUTP Conference of Engineering and Technology

และทั้งทำให้ผู้ใช้งานที่คลิกเสียเวลาในการรอคอยพร้อมทั้งเกิดความรู้สึกที่ไม่ดีต่อการบริการของเซิร์ฟเวอร์ดังกล่าว จึงเป็นข้อพิจารณาในการทำงานวิจัยในเรื่องของการวิเคราะห์หลักการในการเชื่อมต่อข้อมูลของเว็บไซต์ของคณะวิศวกรรมศาสตร์ โดยใช้เครื่องมือสืบค้น Google.com รวมถึงการใช้ เครื่องมือ ใน Microsoft Office [7-8] Python[9] R Programming[10]

2. เว็บไซต์เทคนิค

2.1 โครงสร้างการเข้าถึงเว็บไซต์

การเข้าถึงเว็บไซต์นั้นมีหลักการจากอินเทอร์เน็ตที่เกิดจากผู้ใช้งานคลิกผ่านลิงค์ที่มีอยู่บนหน้าเว็บเพจ เมื่อเว็บเพจมีการร้องขอจากผู้ใช้งานไปยังเซิร์ฟเวอร์เกิดการส่งคำสั่งเช่นคำสั่ง Get ถ้าด้านเซิร์ฟเวอร์เมื่อรับมาแล้วและมีรูปแบบของคำสั่งร้องขอถูกต้องก็ทำการส่งข้อมูลหรือหน้าเว็บเพจที่ต้องการผ่านระบบเครือข่ายอินเทอร์เน็ตไปให้ยังหน้าเพจบน Browser ของผู้ร้องขอ การส่งคำสั่งร้องขอนี้ถ้าผ่านจากผู้ใช้งานไม่ถึงฝั่งเซิร์ฟเวอร์ก็ไม่เกิดการส่งข้อมูลกลับมาทำให้เกิดข้อผิดพลาดขึ้นบนหน้าเพจ Error เช่นที่พบบ่อยกรณีไม่มีหน้าเว็บเพจนั้นคือรหัส 404 แต่ถ้า



รูปที่ 1 ผังการร้องขอระหว่างผู้ใช้งานและเว็บเซิร์ฟเวอร์

หน้าเว็บเพจมีข้อมูลครบ และที่ด้านเว็บเซิร์ฟเวอร์มีเส้นทางเข้าถึงข้อมูลตรงกับทางด้านผู้ใช้งานก็ส่งข้อมูล Response กลับไปยังผู้ใช้งานบน Browser ก็มีข้อมูลนำเสนอให้ผู้ใช้งาน กรณีที่มี Add-on ลงไว้พร้อมใช้งานสำหรับภาพหรือมีเดียไฟล์บน Browser ก็สามารถนำเสนอได้อย่างสมบูรณ์

2.2 การทำงานของหน้าเว็บ

การทำงานบนระบบเครือข่ายอินเทอร์เน็ตนั้นมีความยาวนานในระยะเวลาการส่งและการตอบสนองกลับจากผู้ใช้งานไปยังเซิร์ฟเวอร์ไม่แน่นอนทั้งนี้ขึ้นกับปัจจัยหลายประการ เช่นในแต่ละช่วงการเคลื่อนที่ของข้อมูลออกสู่โลกภายนอก แต่ช่วงเวลามีความหนาแน่นในการใช้งานที่แตกต่างกันไปตามสภาพการใช้งานจริง ดังนั้นผู้ใช้งานเมื่อส่งข้อมูลไปยังปลายทางอาจมีช่วงระยะเวลาตอบสนองที่แตกต่างกันออกไป

แต่โดยระบบแล้วการหาเส้นทางที่สั้นที่สุด ดีที่สุดเป็นหน้าที่ของอุปกรณ์หาเส้นทาง

ถ้าพิจารณาโดยจับกลุ่มของปริมาณข้อมูลระยะทางและแบนด์วิดท์ที่ให้ความสำคัญในเรื่องความเร็วในการรับส่งข้อมูลและการตอบสนองของโนดปลายทางหรือเซิร์ฟเวอร์แล้วนั้น โดยระยะแต่ละช่วงของเส้นแปรตามข้อมูลดังกล่าวที่พิจารณา เห็นได้ว่าทุกส่วนล้วนอยู่บนฐานของเวลา และเป็นตัวชี้ถึงการตอบสนองต่อผู้ใช้งาน แต่การเชื่อมโยงต่าง ๆ ของเว็บไซต์อยู่บนหน้าเว็บเพจ โดยการแทรกด้วย Tag การลิงค์ก็อีกทั้งจากเว็บไซต์ภายนอกที่ลิงค์ย้อนกลับมาหาจึงต้องนำมาพิจารณา จากตัวอย่างของ Tag ในไฟล์ HTML

```
<!DOCTYPE html>
<html> ... </html>
<body>
<h1> to <h6> tags
<p> tag
<b> ... </b> tags เป็นต้น
```

โดยหลักแล้วประกอบด้วยส่วนที่เรียกว่า “child” “parent” “sibling”

ในส่วนของโปรแกรมบนหน้าเว็บเพจจากตัวอย่างรูปที่ 2 มี Tag ที่เป็น องค์ประกอบหลักที่ควรพิจารณาอยู่หลายส่วนดังนี้

```
... html lang="en-US" == $0
<!--<![endif]-->
<head>...</head>
<body class="home page-template-default page page-id-53 custom-background mega-menu-primary right-sidebar content-full">
  <div id="top-page">...</div>
  <div id="page" class="hfeed site">...</div>
  <!-- #page .hfeed .site -->
  <b>...</b>
</body>
</html>
```

รูปที่ 2 โครงสร้าง HTML หน้าเว็บคณะวิศวกรรมศาสตร์ มทร.พระนคร

2.3 ทฤษฎีที่เกี่ยวข้อง

วิธีการในการพิจารณาสภาพแวดล้อม Hyperlinked ของ Jon Kleinberg [5] นับเป็นวิธีการที่นิยมใช้งานวิธีหนึ่งแม้ว่าทำให้ต้องเป็นภาระในการประมวลผลเนื่องจากต้องทำกับสมาชิกสองชุดเพื่อคำนวณย้อนกับไปยัง Parent pages นั้นมีความแม่นยำที่ดีสำหรับการลิงค์ Hyperlink-Induced Topic Search หรือ HITS เป็นที่นิยมใช้งานในปัจจุบัน และยังมีหลายเทคนิควิธีการในการคำนวณที่สามารถให้ผลลัพธ์ที่ใกล้เคียงความเป็นจริง แต่กรณีศึกษาที่ทำอยู่ปรับใช้แบบ Visit page และทำการ View เพื่อหาข้อมูลที่ต้องการในเรื่องของลิงค์ โดยเฉพาะการเชื่อมต่อที่ทำให้ทราบถึงข้อมูลเบื้องต้นของหน้าเว็บเพจที่ต้องการ ส่วน

บทความวิจัย

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 3
Proceedings of the 3rd RMUTP Conference of Engineering and Technology

การหาหน้าหนักนั้นก็เป็นอีกเทคนิคที่ดี ดังนั้นการทดสอบโปรแกรมด้วยการรันเทคนิคต่าง ๆ มีความสำคัญในการพัฒนาวิธีการที่เหมาะสมกับสถานการณ์ที่แตกต่างกันไป

2.4 ข้อมูลเบื้องต้นจากการสืบค้น

การสืบค้นได้กระทำในช่วงกลางเดือนพฤศจิกายน 2560 โดยได้มีการเก็บข้อมูลจากหน้าเว็บเพจของคณะวิศวกรรมศาสตร์มหาวิทยาลัยเทคโนโลยีราชมงคลพระนครจำนวน 9 ครั้งและหาค่าเฉลี่ยคิดเป็นเปอร์เซ็นต์จากสัดส่วนข้อมูลที่ได้และเก็บอีกช่วงในปลายเดือนกุมภาพันธ์ 2561 ซึ่งอยู่ในรอบไม่เกิน 1 ไตรมาส ได้ข้อมูลที่นำเสนอได้แสดงในตารางที่ 1 ในช่องจำนวนที่ได้จากการค้นหาโดยช่องทางย่อยซ้ำมือเป็นช่วงแรกในการค้นหาข้อมูลส่วนช่องค้นหาซ้ำมือเป็นข้อมูลในช่วงประมาณ 1 ไตรมาสของการค้นหาข้อมูลโดยสรุป

พิจารณาจากข้อมูลที่ได้ พบว่าจำนวนหน้าเพจของเอกสาร HTML มากกว่าสองพันส่วนการลิงค์หรือเชื่อมโยงนั้นมีเปอร์เซ็นต์มากคือจาก Scripts ไม่ว่าจะจากทางส่วนของเซิร์ฟเวอร์และที่รองรับจาก Clients หรือผู้ใช้งาน โดยมีรหัสผิดพลาดตามมาตรฐาน HTTP/1.1 RFC2616[6]

จากข้อมูลที่ได้โดยเฉลี่ยของการสืบค้นในสัปดาห์แรกที่ค้นหาข้อมูลหลายรอบแล้วนำมาเฉลี่ยมีจำนวนลิงค์มายังหน้าเว็บประมาณ 924 และมีลิงค์ไปจากหน้าเว็บเพจประมาณ 925 มีผิดพลาด 504 (Gateway timeout) 405 (Method not allowed) 404 (Not found) และเกิดรหัสผิดพลาด 301 (Moved permanently) รวมประมาณ 211 ลิงค์ส่วนใหญ่เป็นรหัสผิดพลาด 504 ประมาณ 96% อีก 2.4% เป็นรหัสผิดพลาด 404

ตารางที่ 1 การเก็บข้อมูลจากระบบอินเทอร์เน็ต

ผลที่พบ	จำนวนเฉลี่ยจากการค้นหา		เปลี่ยน
HTML	2135	2282	+147
Mail to Links	5	3	-2
Graphic Links	29	28	-1
Style sheet Links	28	27	-1
Text Links	23	23	0
XML Links	25	24	-1
Script Links (Clients+Severs)	163	161	-2
Generic Links	3	2	-1

2.5 การเปรียบเทียบลิงค์ระหว่างสาขาวิชา

จากค้น ข้อมูลจากเว็บไซต์ eng.rmup.ac.th โดยแยกการค้นหาค้นหาแต่ละส่วนของสาขาวิชาแต่ละสาขาวิชารวบรวมได้ทั้งสิ้น 9 สาขาวิชา ปัจจุบันมีการเปิดสอนในหลักสูตรระดับปริญญาตรี ปริญญาโทในบางสาขาวิชาฯ รวมถึงระดับ ปวช. ในบางสาขาวิชาฯ ซึ่งตรวจสอบแล้วอยู่ภายใต้สาขาวิชานั้น ๆ เช่นสาขาวิชาวิศวกรรมเครื่องกล มีการให้การศึกษาระดับปริญญาโทและระดับ ปวช. โดยภาพรวมแล้วที่อยู่ภายในหน้าเว็บไซต์สาขาวิชาวิศวกรรมเครื่องกล หรือสาขาวิชาวิศวกรรมไฟฟ้าก็เช่นกัน ส่วนสาขาวิชาวิศวกรรมคอมพิวเตอร์มีเฉพาะระดับปริญญาตรีเพียงอย่างเดียว แต่ละสาขาวิชาต้องรับผิดชอบในการดูแลเว็บเพจของสาขา อาจารย์แต่ละสาขามีภาระกิจมางานวิจัยนี้ช่วยนำเสนอข้อมูลในเบื้องต้น ดังนั้น โดยรวมในงานวิจัยนี้ใช้การพิจารณาเฉพาะส่วนข้อมูลจากหน้าเว็บไซต์ของสาขาวิชาในระดับปริญญาตรียึดหลักจากเมนูลิงค์บนหน้าเว็บไซต์จากบนลงล่างประกอบด้วยสาขาวิชาดังต่อไปนี้

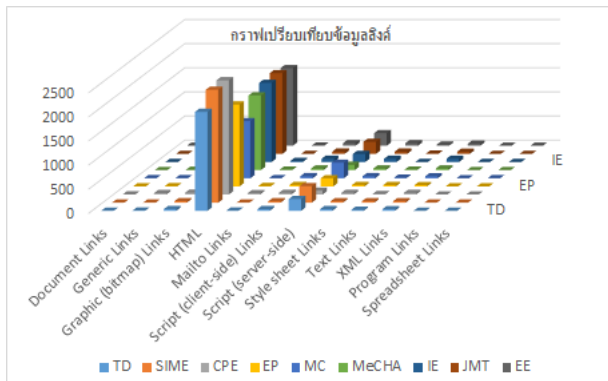
- 1) สาขาวิชาวิศวกรรมไฟฟ้า (EP: www.eerg.eng.rmup.ac.th)
- 2) สาขาวิชาวิศวกรรมเครื่องกล (MC:me.eng.rmup.ac.th)
- 3) สาขาวิชาวิศวกรรมอุตสาหกรรม (IE:www.rie.eng.rmup.ac.th)
- 4) สาขาวิชาวิศวกรรมคอมพิวเตอร์ (CPE:www.cpe.eng.rmup.ac.th)
- 5) สาขาวิชาวิศวกรรมแมคคาทรอนิกส์ (MeCHA:www.mce.eng.rmup.ac.th)
- 6) สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม (EE:www.ee.eng.rmup.ac.th)
- 7) สาขาวิชาวิศวกรรมโยธา (CE:www.cve.eng.rmup.ac.th)
- 8) สาขาวิชาวิศวกรรมการผลิตเครื่องมือและแม่พิมพ์ (TD:tde.eng.rmup.ac.th)
- 9) สาขาวิชาวิศวกรรมอัญมณีและเครื่องประดับ (JMT: jmt.eng.rmup.ac.th)
- 10) สาขาวิศวกรรมเพื่อความยั่งยืน (SIME:sime.eng.rmup.ac.th)

จากการสืบค้นบนเซิร์ฟเวอร์พบว่าส่วนของบางสาขาวิชาไม่มีหน้าเว็บไซต์ในระบบจึงขอยกไว้ไม่นำมาพิจารณา

ลิงค์ของเซิร์ฟเวอร์ eng.rmup.ac.th จากการค้น หาบนเครื่องมือค้นหา Google เฉลี่ยมีประมาณ 7,209 เห็นได้ว่าแต่ละสาขาวิชา มีลิงค์เชื่อมโยงที่มีลักษณะใกล้เคียงกับหน้าเว็บหลักของคณะวิศวกรรมศาสตร์ มทร.พระนคร ส่วนใหญ่มากกว่าร้อยละ 90 เป็น HTML link รองลงมาเป็นลิงค์ในสคริปต์ของทั้ง Client-Server

บทความวิจัย

การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 3
Proceedings of the 3rd RMUTP Conference of Engineering and Technology



รูปที่ 3 เปรียบเทียบข้อมูลของแต่ละสาขาวิชา

3. สรุป

การใช้หน้าเว็บไซต์ให้เป็นประโยชน์ในการประชาสัมพันธ์ เพื่อให้ผู้ใช้งานเข้าถึงหน้าเว็บได้ในปริมาณที่ ดังนั้นการศึกษานี้เป็นประโยชน์ด้านข้อมูลของเว็บไซต์ของหน่วยงานสำหรับนักพัฒนาเว็บไซด์และโปรแกรมเมอร์เพื่อการจัดการและการเฟิร์มแวร์ทำให้เว็บไซต์มีการสนองตอบต่อผู้ใช้งาน ส่วนการพิจารณาถึงลิงค์ต่าง ๆ ที่มีบนหน้าเว็บเพจที่ต้องลดการเกิดภาวะลิงค์ (Link) และเส้นทาง (Path) ผิดพลาดรหัส 504 405 404 301 หรือรหัสผิดพลาดอื่น ๆ ตาม RFC2616 เพื่อทำให้ผู้ใช้เมื่อใช้งานเว็บไซต์มีความรู้สึกที่ดีสำหรับการสนองตอบของเซิร์ฟเวอร์ในการคลิกเพื่อเข้าถึงข้อมูลที่เตรียมไว้ไปถึงเทคนิคในการเพิ่มลิงค์สำหรับผู้พัฒนาเว็บไซต์เพื่อลดการเกิดข้อผิดพลาดของเส้นทาง

4. กิตติกรรมประกาศ

ผู้เขียนขอขอบคุณทางคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร สวพ. สวส. ที่สนับสนุนเครื่องมือและทุนวิจัยประจำปี 2561 และที่ขาดไม่ได้ Google Search Engine

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 Poulouvasilis, *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.