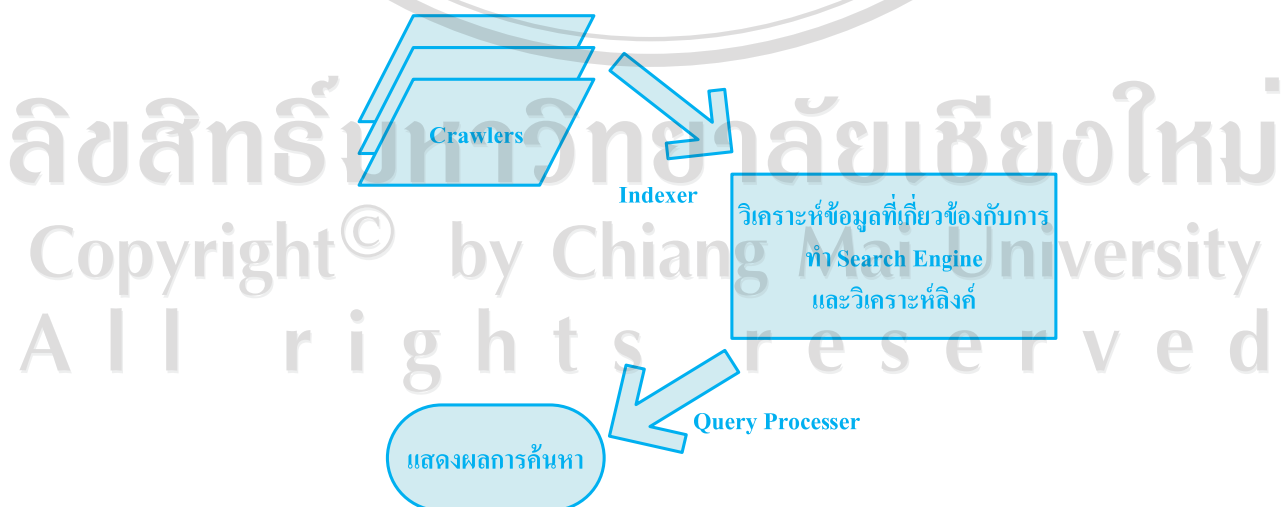


1.1 ที่มาของปัญหาที่นำไปสู่การค้นคว้าวิจัย และแนวทางการแก้ปัญหา

เสิร์ชเอนจิน (Search Engine) เป็นเครื่องมือสำหรับค้นหาเว็บเพจบนเครือข่ายอินเทอร์เน็ต ปัจจุบันมีเสิร์ชเอนจินที่ได้รับความนิยมใช้งานกันอยู่มากมาย เช่น กูเกิล (Google) ยาฮู (Yahoo!) หรือ บิง (Bing) เป็นต้น

ในปี 1998 นักวิจัยกลุ่มหนึ่งคิดค้นเครื่องมือสำหรับสำหรับค้นหาเอกสารบนอินเทอร์เน็ต โดยให้ชื่อว่า กูเกิล [1] และได้พัฒนาต่อมาเรื่อยๆ เนื่องจากประสิทธิภาพของการทำดัชนี (Index) ให้กับเว็บไซต์ ทำให้ผู้ใช้งานนิยมใช้งานกันมากที่สุดในปัจจุบัน ซึ่งการทำงานของกูเกิล แบ่งเป็น 3 ขั้นตอนหลักๆ ดังแสดงการทำงานของกูเกิลในรูปที่ 1.1 โดยจะเริ่มจากการค้นหาเว็บไซต์บนอินเทอร์เน็ต จากนั้นนำเว็บไซต์ทั้งหมดที่ค้นหาได้มาทำการวิเคราะห์ข้อมูล และสุดท้ายแสดงผลการสอบถาม (Query) ตามคำค้น (Keyword) ของผู้ใช้งาน

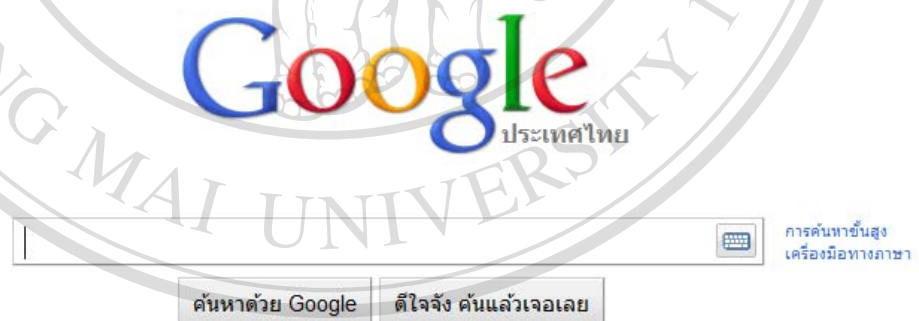


รูปที่ 1.1 แสดงการทำงานของกูเกิล

การค้นหาเว็บไซต์บนอินเทอร์เน็ต คุกกี้จะใช้เว็บคลอว์เลอร์ (Web Crawler) หรือเว็บสไปเดอร์ (Web Spider) ซึ่งเป็นโปรแกรมประยุกต์สำหรับค้นหาเว็บไซต์ทั้งหมดบนอินเทอร์เน็ต และนำกลับมาเก็บไว้ในสโตร์เซิร์ฟเวอร์ (Store Server)

สำหรับแต่ละเว็บเพจที่ได้จากเว็บคลอว์เลอร์ จะถูกนำมาแยกเอาข้อมูลที่จำเป็นสำหรับการทำงานของเสิร์ชเอนจินออกมาวิเคราะห์ด้วยกระบวนการต่างๆ ซึ่งกระบวนการหนึ่งที่เป็นพื้นฐานสำคัญของการพัฒนาเสิร์ชเอนจินคือการแยกลิงก์ (Link) ออกจากเว็บเพจเพื่อจัดลำดับเว็บเพจ (Webpage Ranking) ซึ่งคุกกี้เรียกขั้นตอนนี้ว่าการคำนวณค่าเพจเรงก์ (Page Rank) โดยจะพิจารณาจากลิงก์เข้า (Back Link) และลิงก์ออก (Forward Link) ของเว็บเพจ ซึ่งได้อธิบายรายละเอียดไว้ในบทที่ 2

การแสดงผลการสอบถามเป็นส่วนติดต่อกับผู้ใช้งาน โดยคุกกี้จะรับคำค้นจากผู้ใช้งานผ่านทางเว็บเบราว์เซอร์ (Web Browser) จากนั้นตัวประมวลผลการสอบถาม (Query Processor) จะทำการสอบถาม และแสดงผลกลับไปยังเว็บเบราว์เซอร์ของผู้ใช้งานที่ทำการค้นหา ดังแสดงส่วนต่อประสานกับผู้ใช้งาน (User Interface) ของคุกกี้ในรูปที่ 1.2

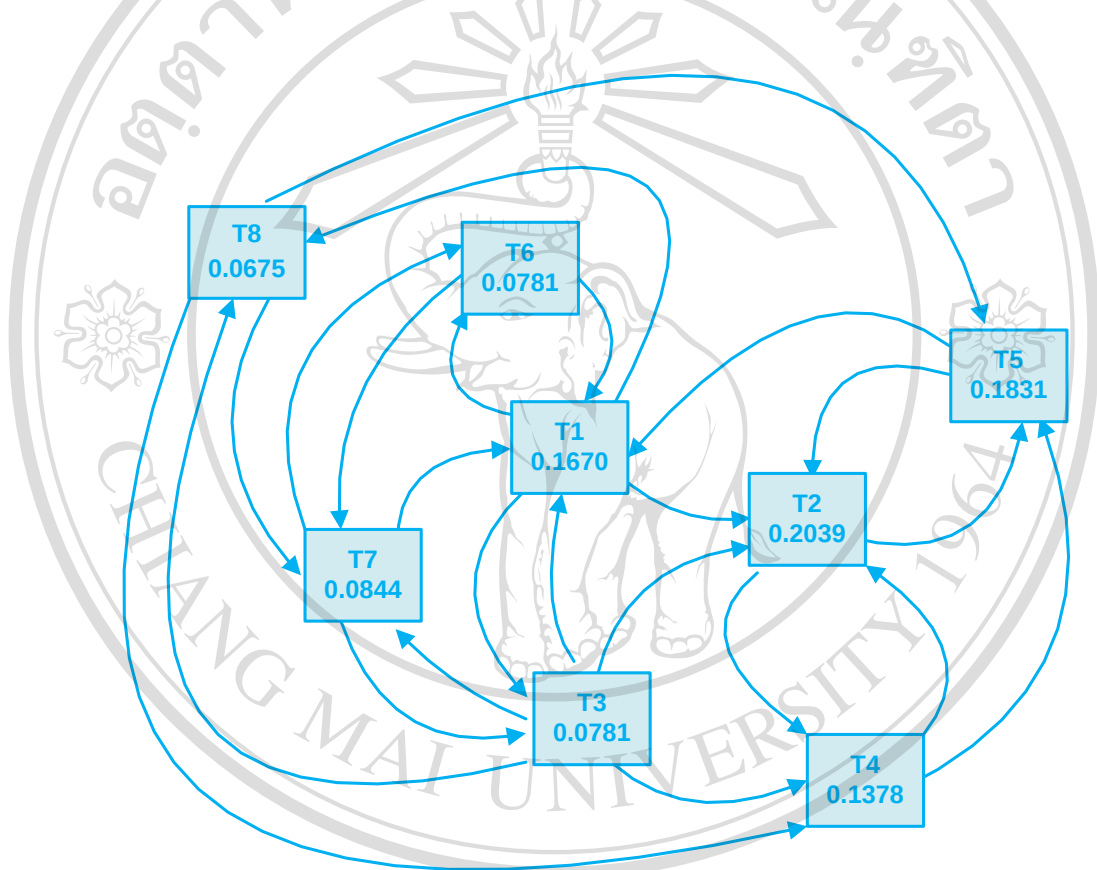


ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

รูปที่ 1.2 แสดงส่วนต่อประสานกับผู้ใช้งานของคุกกี้

(<http://www.google.com/>)

โครงสร้างของลิงค์เข้าและลิงค์ออกของเว็บเพจทั้งหมดที่ได้จากการวิเคราะห์ลิงค์ สามารถแทนได้ด้วยกราฟแบบมีทิศทาง (Directed Graph) โดยกำหนดให้แต่ละเว็บเพจแทนด้วยโหนด (Node) ของกราฟ และกำหนดให้แต่ละลิงค์ที่เชื่อมกันระหว่างเว็บเพจแทนด้วยเอดจ์ (Edge) ของกราฟ ซึ่งมีทิศทางจากเว็บเพจหนึ่งไปยังอีกเว็บเพจหนึ่ง เรียกกราฟของเว็บเพจที่มีความสัมพันธ์กันนี้ว่าเว็บลิงค์กราฟ (Web Link Graph) ซึ่งกราฟดังกล่าวสามารถนำไปคำนวณค่าเพจแรงค์ของแต่ละเว็บเพจในกราฟได้ ดังแสดงตัวอย่างของเว็บลิงค์กราฟและค่าเพจแรงค์ในรูปที่ 1.3

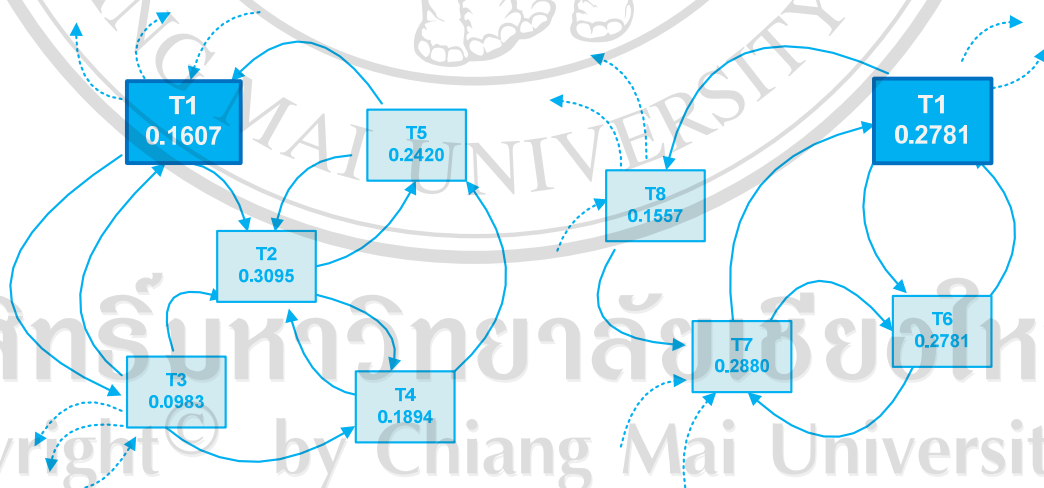


รูปที่ 1.3 แสดงตัวอย่างของเว็บลิงค์กราฟและค่าเพจแรงค์

เมื่อต้องการหาค่าเพจแรงค์ของแต่ละเว็บเพจในเว็บลิงค์กราฟ ทำได้ด้วยการคำนวณแบบขั้นตอนวิธีการทำซ้ำ (Iterative Algorithm) และมีผลลัพธ์ของสมการลู่เข้าหาค่าใดค่าหนึ่ง (Converge) ซึ่งจำนวนรอบของการปรับค่าเพจแรงค์เพื่อให้เข้าใกล้ค่าที่เหมาะสมที่สุดนั้น ขึ้นอยู่กับขนาดและโครงสร้างของเว็บลิงค์กราฟ

ภายใต้ข้อจำกัดในด้านทรัพยากรของคอมพิวเตอร์ ไม่ว่าจะเป็นข้อจำกัดของหน่วยความจำ (Memory) ฮาร์ดดิสก์ (Harddisk) หรือเวลาการประมวลผลข้อมูล (Computation Time) ล้วนแต่เป็นปัจจัยที่มีอิทธิพลในการจัดลำดับเว็บเพจทั้งสิ้น หากขนาดของเว็บลิงค์กราฟขยายใหญ่มากขึ้น จำเป็นต้องใช้ขนาดของฮาร์ดดิสก์และหน่วยความจำที่มากพอสำหรับจัดเก็บ และประมวลผลข้อมูล ในทำนองเดียวกันขนาดที่ใหญ่ขึ้นของเว็บลิงค์กราฟ จะส่งผลให้เวลาในการคำนวณเพจแรงค์เยอะขึ้นตามไปด้วย ดังนั้นนักวิจัยกลุ่มหนึ่งจึงหาวิธีแก้ปัญหาการขยายขนาดของเว็บลิงค์กราฟที่ส่งผลให้เกิดความไม่พอเพียงของทรัพยากร และประสิทธิภาพของเวลาในการคำนวณที่แย่ง โดยเสนอหลักการทำเสร็จเอ็นจินบนระบบกระจายศูนย์ (Distribute System) ขึ้น

อย่างไรก็ตาม การทำงานที่เป็นอิสระต่อกันของเว็บคลอเลอร์บนระบบกระจายศูนย์ อาจมีโอกาที่ข้อมูลแต่ละเพียร์ (Peer) มีการซ้อนทับกันสูง เนื่องจาก 1 เว็บเพจสามารถถูกเก็บไว้ในหลายๆ เพียร์ การคำนวณค่าเพจแรงค์เพียงแคในเพียร์ จะมีผลให้ค่าเพจแรงค์ของเว็บเพจเดียวกันที่ถูกเก็บไว้ต่างเพียร์ มีค่าเพจแรงค์ไม่เท่ากัน ดังแสดงตัวอย่างในรูปที่ 1.4 เว็บเพจ T1 ที่ถูกเก็บไว้ในเพียร์ A และ เว็บเพจ T1 ที่ถูกเก็บไว้ในเพียร์ B เป็นเว็บเพจเดียวกัน แต่มีค่าไม่เท่ากัน ซึ่งกรณีดังกล่าวจะส่งผลกระทบต่อความผิดพลาดของการจัดลำดับเว็บเพจได้ และใน [2, 4] นักวิจัยอีกกลุ่มหนึ่งเสนอหลักการสำหรับจัดลำดับเว็บเพจบนระบบกระจายศูนย์ขึ้น



a.เว็บลิงค์กราฟในเพียร์ A

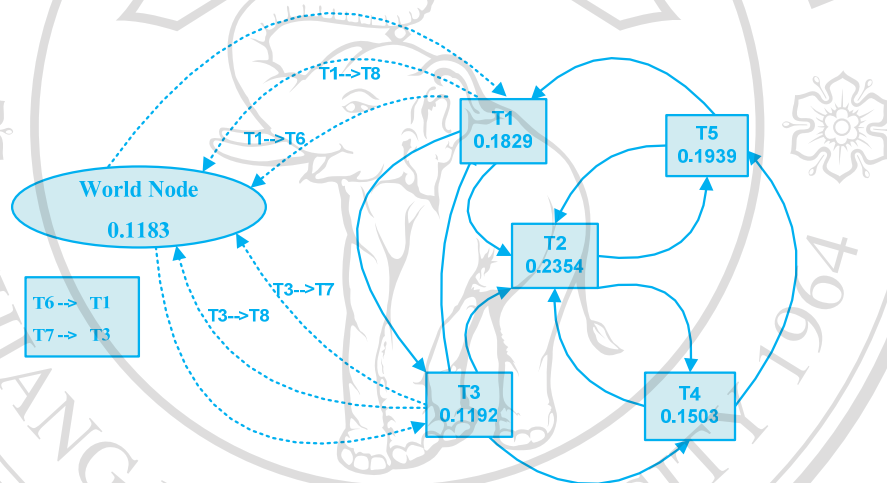
b.เว็บลิงค์กราฟในเพียร์ B

รูปที่ 1.4 แสดงตัวอย่างเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์และค่าเพจแรงค์

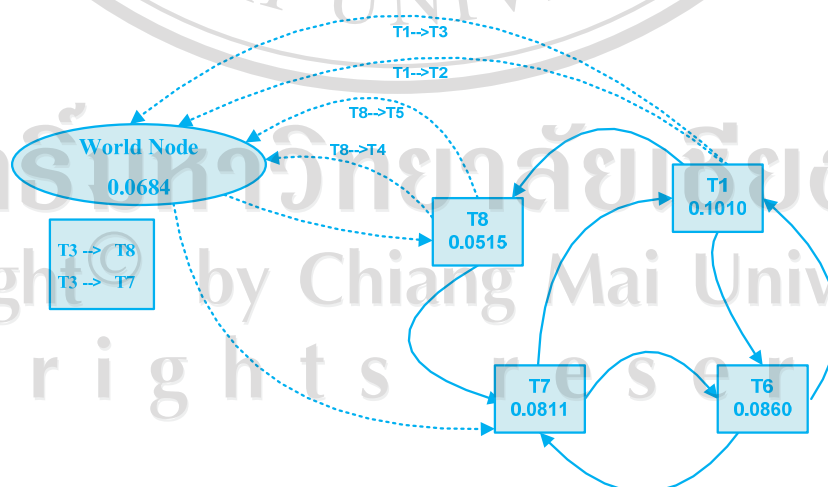
ในวิทยานิพนธ์เล่มนี้ ผู้วิจัยได้ศึกษางานวิจัย [3] ซึ่งเป็นงานวิจัยที่เสนอแนวคิดของการจัดลำดับเว็บเพจบนเครือข่ายเพียร์ทูเพียร์ (Peer to Peer Network) โดยพัฒนาขั้นตอนวิธี JXP สำหรับจัดลำดับเว็บเพจบนระบบกระจายศูนย์ ให้เสมือนว่ากำลังจัดลำดับเว็บเพจบนระบบแบบ

รวมศูนย์ (Centralized System) ซึ่งได้อธิบายการทำงานของขั้นตอนวิธี JXP ไว้ในบทที่ 2 หากเว็บลิงค์กราฟขยายขนาดใหญ่ขึ้น ขั้นตอนวิธี JXP จะมีข้อได้เปรียบในด้านประสิทธิภาพของเวลาการคำนวณค่าเพจแรงค์ และประสิทธิภาพของการใช้ทรัพยากรในคอมพิวเตอร์

ขั้นตอนวิธี JXP จะเรียกกราฟที่ถูกเก็บไว้ในแต่ละเฟียร์ว่าโลคอลเว็บลิงค์กราฟ (Local Web Link Graph) ดังรูป 3.4 ในขั้นตอนวิธี JXP จะเรียกเว็บลิงค์กราฟในรูปที่ 3.4a และ 3.4b ว่าโลคอลเว็บลิงค์กราฟในเฟียร์ A และโลคอลเว็บลิงค์กราฟในเฟียร์ B ตามลำดับ และโลคอลเว็บลิงค์กราฟในแต่ละเฟียร์จะเชื่อมต่อกับเวิลด์โหนด (World Node) ซึ่งเป็นโหนดสำหรับเก็บเว็บเพจภายนอกเฟียร์ (External Pages) ดังแสดงในรูปที่ 1.5



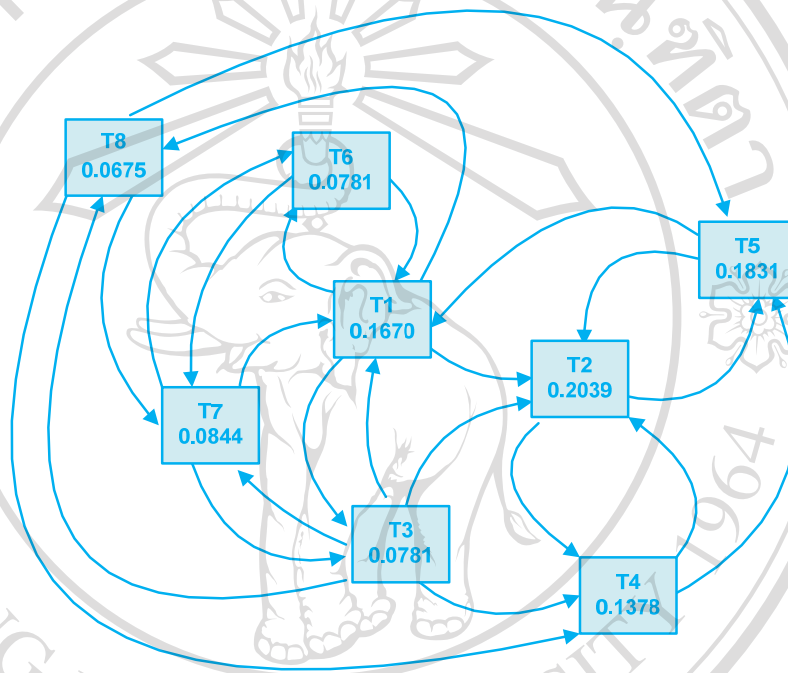
a. ตัวอย่างค่าน้ำหนักของเว็บเพจในโลคอลเว็บลิงค์กราฟในเฟียร์ A



b. ตัวอย่างค่าน้ำหนักของเว็บเพจใน โลคอลเว็บลิงค์กราฟในเฟียร์ B

รูปที่ 1.5 แสดงโลคอลเว็บลิงค์กราฟ ที่เชื่อมต่อกับเวิลด์โหนด

หากนำโหนดเว็บลิงค์กราฟที่เก็บไว้ในทุกเพียร์บนเครือข่าย มายูเนียน (Union) กัน จะได้เว็บลิงค์กราฟขนาดใหญ่เพียง 1 กราฟ เรียกเว็บลิงค์กราฟที่ได้นี้ว่าโกลบอลเว็บลิงค์กราฟ (Global Web Link Graph) ดังรูปที่ 1.6 แสดงโกลบอลเว็บลิงค์กราฟที่เกิดจากการยูเนียน โหนดเว็บลิงค์กราฟในเพียร์ A และเพียร์ B



รูปที่ 1.6 แสดงโกลบอลเว็บลิงค์กราฟ

สำหรับในขั้นตอนวิธี JXP จะพิจารณาค่าเพจแรงค์ในแต่ละโหนดเว็บลิงค์กราฟบนเครือข่ายให้เสมือนกับที่กำลังพิจารณาค่าเพจแรงค์บนโกลบอลเว็บลิงค์กราฟ ดังตัวอย่างเว็บลิงค์กราฟในรูปที่ 1.5a และ 1.5b ซึ่งเป็นเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์ A และ เพียร์ B ตามลำดับ จะถูกพิจารณาค่าเพจแรงค์ให้เสมือนกับที่กำลังพิจารณาค่าเพจแรงค์ของเว็บลิงค์กราฟในรูปที่ 1.6

อย่างไรก็ตามค่าเพจแรงค์ที่ถือว่ามีประสิทธิภาพและถูกต้องที่สุด คือค่าเพจแรงค์ที่ได้จากการพิจารณาบนโกลบอลเว็บลิงค์กราฟ ซึ่งในความเป็นจริงไม่มีความเป็นไปได้ที่จะได้ค่าดังกล่าวเนื่องจากสามารถพิจารณาได้แค่ในโหนดเว็บลิงค์กราฟเท่านั้น แต่ขั้นตอนวิธี JXP สามารถที่จะพิจารณาค่าเพจแรงค์จากโหนดเว็บลิงค์กราฟ ให้มีค่าใกล้เคียงกับค่าเพจแรงค์จากโกลบอลเว็บลิงค์กราฟได้

แม้ว่าขั้นตอนวิธี JXP จะทำงานได้อย่างมีประสิทธิภาพ แต่การเปลี่ยนแปลงโครงสร้างของเว็บลิงค์กราฟ ทำให้ขั้นตอนวิธี JXP ต้องเริ่มต้นคำนวณค่าเพจแรงค์ของทุกๆเว็บเพจในเว็บลิงค์กราฟใหม่ทุกครั้ง ซึ่งในบางกรณีไม่จำเป็นต้องคำนวณค่าเพจแรงค์ของเว็บเพจใหม่ทั้งหมด ดังแสดงในตารางที่ 1.1

L-WLG	P1	P2	P3	P4	P5	P6	P7	P8	P9
เริ่มต้น	0.037	0.037	0.399	0.037	0.399	0.037	0.053	-	-
เวลา T1	0.033	0.033	0.308	0.033	0.238	0.033	0.174	0.150	-
เวลา T2	0.029	0.029	0.246	0.029	0.189	0.029	0.131	0.211	0.106
เวลา T3	0.029	0.029	0.200	0.029	0.154	0.029	0.141	0.234	0.156

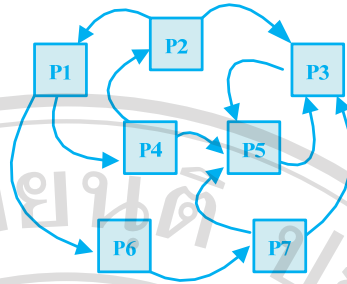
ตารางที่ 1.1 ค่าเพจแรงค์ของเว็บลิงค์กราฟเมื่อมีการเปลี่ยนแปลงของโครงสร้างของเว็บลิงค์กราฟ

จากตารางที่ 1.1 เป็นค่าเพจแรงค์ที่ได้จากเว็บลิงค์กราฟในรูปที่ 1.8 ซึ่งเริ่มต้นด้วยเว็บลิงค์กราฟที่มีเว็บเพจเพียง 7 โหนด ดังแสดงในรูปที่ 1.8a และเมื่อเวลาผ่านไปเป็น T1 มีเว็บเพจ P8 เพิ่มขึ้นในเว็บลิงค์กราฟ ดังแสดงในรูปที่ 1.8b จากนั้นเมื่อเวลาเปลี่ยนเป็น T2 มีเว็บเพจ P9 เพิ่มขึ้นในเว็บลิงค์กราฟ ดังแสดงในรูปที่ 1.8c ส่วนเมื่อเวลาเปลี่ยนเป็น T3 จำนวนเว็บเพจในเว็บลิงค์กราฟไม่เปลี่ยนแปลง เพียงแต่มีการเพิ่มขึ้นของเว็บลิงค์ ดังแสดงในรูป 1.8d

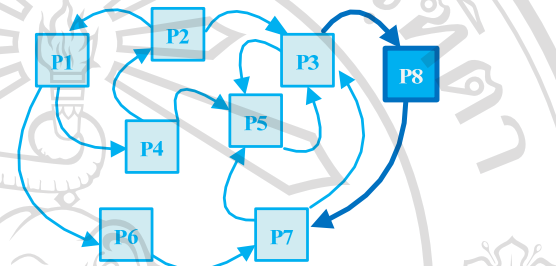
จากค่าของเพจแรงค์ของแต่ละเว็บเพจในตารางที่ 1.1 จะเห็นว่าแม้โครงสร้างของเว็บลิงค์กราฟจะเปลี่ยนแปลงไป แต่ค่าเพจแรงค์ของบางเว็บเพจในเว็บลิงค์กราฟเปลี่ยนแปลงไปเพียงเล็กน้อยเท่านั้น ดังตัวอย่างค่าเพจแรงค์จากเว็บเพจ P1, P2, P4 และ P6 ในตารางที่ 1.1 เมื่อเปรียบเทียบการเปลี่ยนแปลงของค่าเพจแรงค์ของเว็บเพจดังกล่าว กับการเปลี่ยนแปลงของค่าเพจแรงค์ของเว็บเพจอื่นๆ ที่เหลือในเว็บลิงค์กราฟแล้ว ถือว่าค่าเพจแรงค์เปลี่ยนแปลงไปเพียงเล็กน้อย ซึ่งไม่มีผลกระทบต่อการจัดลำดับของเว็บเพจมากนัก

ดังนั้นบางโครงสร้างของเว็บลิงค์กราฟ เมื่อมีการเปลี่ยนแปลงโครงสร้างใหม่ ไม่ว่าจะเป็นการเพิ่มขึ้นของเว็บเพจหรือการเพิ่มขึ้นของเว็บลิงค์ ไม่จำเป็นต้องเริ่มต้นคำนวณค่าเพจแรงค์ของทุกเว็บเพจใหม่คำนวณแค่บางส่วนก็ได้ ซึ่งการคำนวณเฉพาะบางส่วนนี้จะทำให้โครงสร้างเว็บลิงค์กราฟที่นำมาคำนวณค่าเพจแรงค์มีขนาดเล็กลงจึงใช้เวลาในการปรับค่าเพจแรงค์น้อยลงไปด้วย

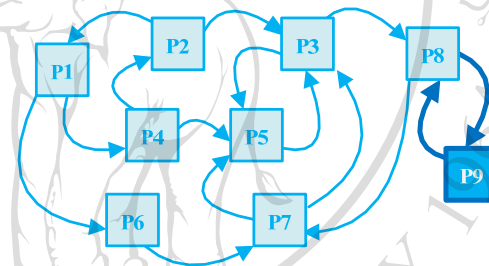
a. โดคอลเว็บลิงค์กราฟเริ่มต้น



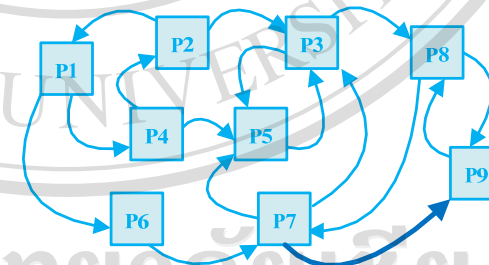
b. โดคอลเว็บลิงค์กราฟเมื่อเวลา T1



c. โดคอลเว็บลิงค์กราฟเมื่อเวลา T2



d. โดคอลเว็บลิงค์กราฟเมื่อเวลา T3



รูปที่ 1.7 แสดงการเปลี่ยนแปลงของโครงสร้างโดคอลเว็บลิงค์กราฟ

สำหรับในงานวิจัยฉบับนี้ ผู้วิจัยได้ศึกษาเพื่อแก้ปัญหาของการเริ่มต้นคำนวณค่าเพจแรงค์ใหม่หมดในทุกเว็บเพจของเว็บลิงค์กราฟเมื่อโครงสร้างของเว็บลิงค์กราฟเปลี่ยนแปลง เพื่อปรับปรุงประสิทธิภาพในเรื่องเวลาการคำนวณค่าเพจแรงค์ของขั้นตอนวิธี JXP โดยใช้หลักการของการจัดลำดับเว็บเพจแบบเพิ่มขึ้น (Incremental Web Ranking) เพื่อพิจารณาค่าเพจแรงค์เฉพาะ

บางส่วนของเว็บเพจในเว็บลิงก์กราฟ แทนการจัดลำดับเว็บเพจแบบธรรมดาในขั้นตอนวิธี JXP ดังแสดงส่วนของขั้นตอนวิธี JXP ที่ถูกปรับปรุงในรูปที่ 1.8

Algorithm1 Initial PageWeight Computation

```

1: input: local graph G and est. size of global graph N
2: n  $\leftarrow$  size(G)
3: Create world node W
4: PageWeight(plp  $\in$  G)  $\leftarrow \frac{1}{N}$ 
5: PageWeight(W)  $\leftarrow \frac{N-n}{N}$ 
6: add W to G
7: run PageRank power iteration on G
8: Update(PageWeights)

```

Algorithm2 the JXP Algorithm

```

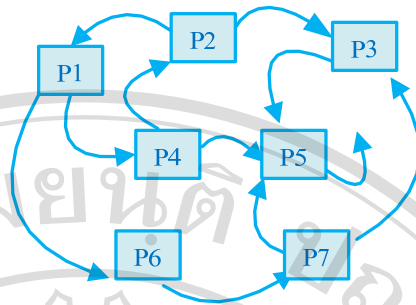
1: input: local graph  $G_A$ , world node  $W_A$ , score list  $L_A$ 
2: repeat
3:   Contact a random peer B in the network and exchange information
4:    $G_M \leftarrow \text{mergeGraphs}(G_A, G_B)$ 
5:    $W_M \leftarrow \text{mergeWorldNodes}(W_A, W_B)$ 
6:    $G'_M \leftarrow (G_M + W_M)$ 
7:    $L_M \leftarrow \text{combineLists}(L_A, L_B)$ 
8:    $PR \leftarrow \text{PageRank}(G'_M)$ 
9:    $L'_M \leftarrow \text{updateScoresList}(L_M, PR)$ 
10:  update( $L_A$ )
11:  update( $W_A$ )
12: Discard( $G_M, W_M, L_M, G_B, W_B, L_B$ )

```

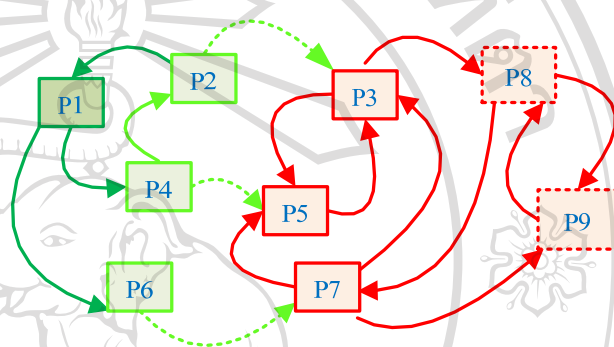
รูปที่ 1.8 แสดงส่วนของขั้นตอนวิธี JXP ที่ถูกปรับปรุง

โดยผู้วิจัยจึงได้แบ่งส่วนของกราฟออกเป็น 3 ส่วนคือ ส่วนแรกเป็นส่วนของเว็บเพจที่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงก์กราฟ ส่วนที่สองคือส่วนของเว็บเพจที่ไม่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงก์กราฟ และส่วนสุดท้าย คือส่วนของเว็บเพจในส่วนแรกที่มีเพียงลิงค์ออกไปยังเว็บเพจในส่วนที่สอง ดังตัวอย่างในรูปที่ 1.9 แสดงเว็บลิงก์กราฟที่ถูกแบ่งส่วน และจำคำนวณค่าเพจแรงค์เฉพาะส่วนแรกเท่านั้น

a. โดคูลเว็บลิงค์กราฟเริ่มต้น



b. โดคูลเว็บลิงค์กราฟเมื่อเวลา T1



รูปที่ 1.9 แสดงเว็บลิงค์กราฟที่ถูกแบ่งส่วน

ส่วนแรกของเว็บลิงค์กราฟ เป็นส่วนของกลุ่มของเว็บเพจที่มีการเปลี่ยนแปลงค่าเพจแรงค์เป็นอย่างมาก เมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ กล่าวคือกลุ่มของเว็บเพจทั้งหมดที่เป็นเดสเซนเด้นต์ (Descendent) ของเว็บเพจที่เพิ่มเข้ามาใหม่ ดังแสดงตัวอย่างใน รูปที่ 1.9b จะเห็นได้ว่าจากเว็บลิงค์กราฟเริ่มต้นในรูปที่ 1.9a เมื่อมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่จะมีโครงสร้างของเว็บลิงค์กราฟเปลี่ยนแปลงไปเป็นเว็บลิงค์กราฟในรูป 1.9b ซึ่งแสดงให้เห็นว่าทั้งเว็บเพจ P8 และ P9 ส่งผลกระทบบต่อค่าเพจแรงค์ของเว็บเพจ P3, P5 และ P7

ส่วนที่สอง เป็นส่วนของกลุ่มของเว็บเพจที่มีการเปลี่ยนแปลงของค่าเพจแรงค์เพียงเล็กน้อย เมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ ซึ่งส่วนดังกล่าวเป็นกลุ่มของเว็บเพจที่ไม่ใช่กลุ่มของเว็บเพจในส่วนแรกทั้งหมด ดังรูปที่ 1.9b เว็บเพจ P1, P2, P4 และ P6 เป็นกลุ่มของเว็บเพจในส่วนที่สอง

ส่วนที่สาม เป็นกลุ่มของเว็บเพจในส่วนที่สองที่มีลิงค์ออกไปยังกลุ่มของเว็บเพจในส่วนแรก ดังรูปที่ 1.9b เว็บเพจ P2, P4 และ P6 เป็นเว็บเพจในส่วนที่สาม เรียกกลุ่มของเว็บเพจในส่วนดังกล่าวว่าโหนดขอบ (Border Nodes) แม้ว่าโหนดขอบจะมีการเปลี่ยนแปลงของค่าเพจแรงค์เพียงเล็กน้อยเมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ แต่ถือว่ามีผลต่อการจัดลำดับเว็บเพจ

แบบเพิ่มขึ้นเป็นอย่างมาก เนื่องจากโหนดขอบจะให้ค่าเพจแรงค์จากกลุ่มของเว็บเพจในส่วนที่สอง ไปยังกลุ่มของเว็บเพจในส่วนแรก

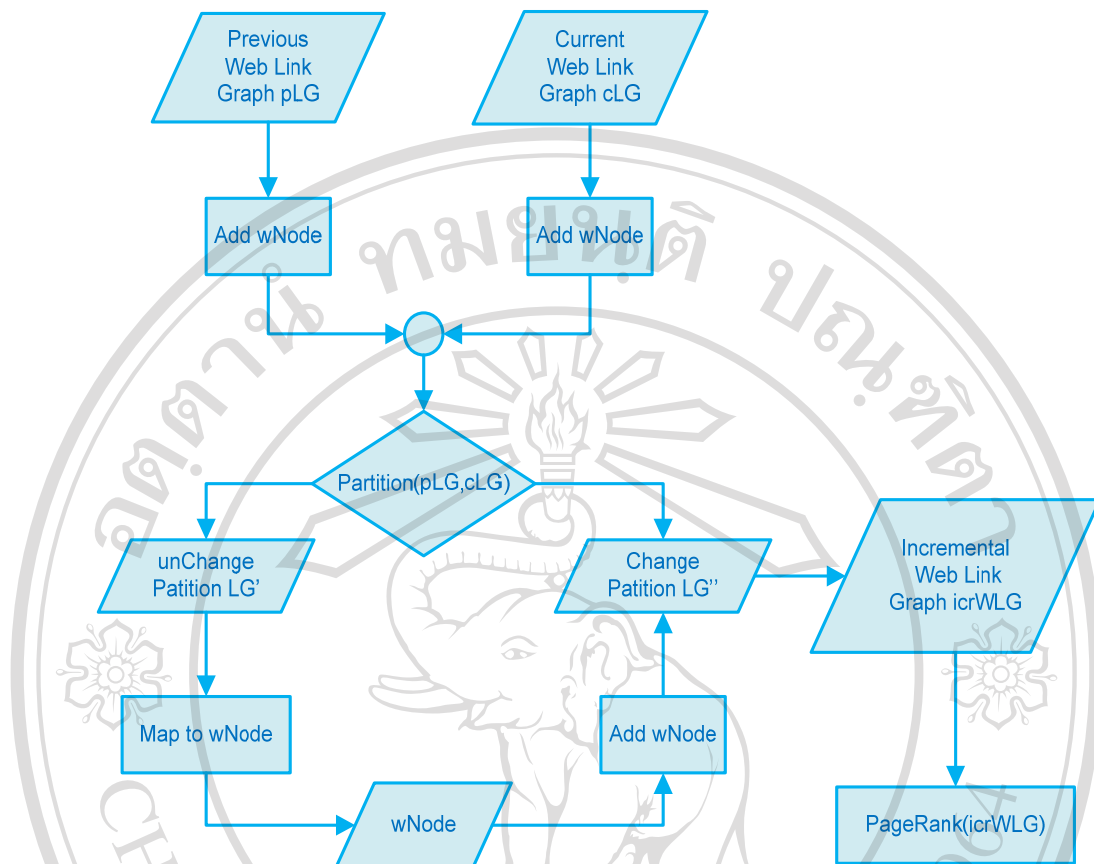
ซึ่งหลักการของการจัดลำดับแบบเพิ่มขึ้นจะคำนวณค่าเพจแรงค์เฉพาะในกลุ่มของเว็บเพจในส่วนแรกของเว็บลิงค์กราฟเท่านั้น เนื่องจากเป็นกลุ่มของเว็บเพจที่ได้รับผลกระทบจากการเพิ่มเข้ามาของเว็บเพจใหม่ ในเว็บลิงค์กราฟ ส่งผลให้ขนาดของเว็บลิงค์กราฟที่นำมาคำนวณค่าเพจแรงค์มีขนาดเล็กลง

อย่างไรก็ตามด้วยหลักการของขั้นตอนวิธี IPR สามารถคำนวณค่าเพจแรงค์ในส่วนแรกได้ทันที เนื่องจากเว็บเพจทั้งหมดถูกเก็บรวมกันไว้ในที่เก็บเดียวกัน

แต่สำหรับขั้นตอนวิธี JXP เว็บลิงค์กราฟถูกแยกเก็บไว้ในแต่ละเพียร์ และมีการนำเวกต์โหนดเข้ามาพิจารณาด้วย ผู้วิจัยจึงได้นำแนวคิดของขั้นตอนวิธี IPR มาประยุกต์ใช้ โดยจะนำเว็บเพจในส่วนที่สองของแต่ละโหนดเว็บลิงค์กราฟไปรวมกับเวกต์โหนด ดังแสดงขั้นตอนวิธีอย่างง่ายในการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ โดยใช้เป็นแผนภาพประกอบการบรรยายในรูปที่ 1.10

จากรูปที่ 1.10 เมื่อมีการแบ่งส่วนของเว็บลิงค์กราฟ cLG เป็น 2 ส่วนคือ LG' ซึ่งเป็น ส่วนของเว็บเพจที่ไม่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ และ LG'' ซึ่งเป็น ส่วนของเว็บเพจที่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงค์กราฟ จากนั้นจะรวมส่วน LG' เข้ากับเวกต์โหนด และเชื่อมต่อเวกต์โหนดเข้ากับส่วน LG'' ได้เป็นเว็บลิงค์กราฟใหม่และคำนวณค่าเพจแรงค์ในเว็บลิงค์กราฟ icrWLG

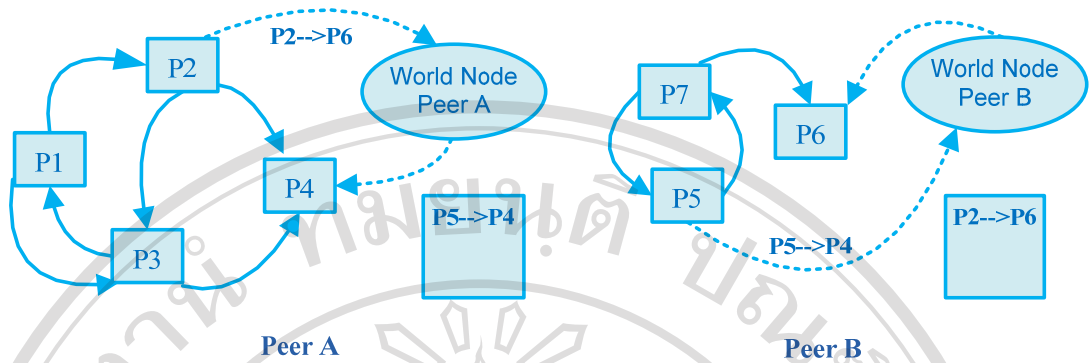
สำหรับการคำนวณค่าเพจแรงค์ในขั้นตอนวิธี JXP จะประกอบไปด้วยการคำนวณ 2 ส่วนด้วยกันคือ การคำนวณค่าเพจแรงค์ในโหนดเว็บลิงค์กราฟที่ถูกเก็บไว้ในแต่ละเพียร์ และการคำนวณค่าเพจแรงค์ในขณะรวมสองโหนดเว็บลิงค์กราฟ ให้เป็นกราฟใหม่ในกระบวนการจับคู่ระหว่างเพียร์ ดังนั้นจึงมีการใช้ขั้นตอนวิธีการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ 2 ครั้งด้วยกัน ในการทำงานของขั้นตอนวิธี JXP คือการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในโหนดเว็บลิงค์กราฟ และการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในกระบวนการจับคู่ระหว่างเพียร์



รูปที่ 1.10 แผนภาพการทำงานของขั้นตอนการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่

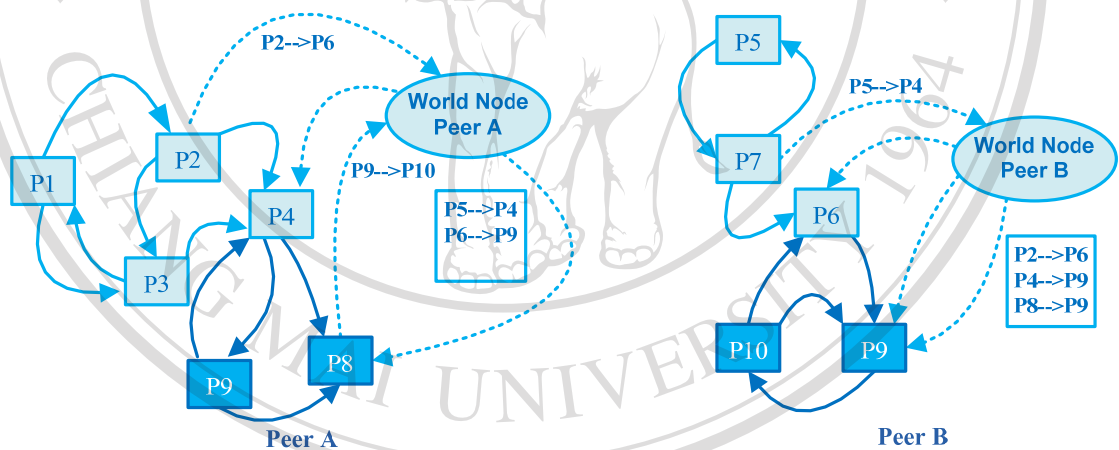
การคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในแต่ละโลกออลเว็บลิงค์กราฟ ทำงานในกระบวนการแรกของขั้นตอนวิธี JXP ซึ่งเป็นการคำนวณค่าเพจแรงค์เฉพาะในเฟิร์ส ซึ่งรายละเอียดของขั้นตอนวิธีจะอธิบายต่อไปในบทที่ 3

ตัวอย่างการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในแต่ละโลกออลเว็บลิงค์กราฟ กำหนดให้เครือข่ายเฟิร์มทูเฟิร์มมีเฟิร์สทั้งหมด 2 เฟิร์ส ซึ่งเก็บโลกออลเว็บลิงค์กราฟได้ดังรูปที่ 1.11



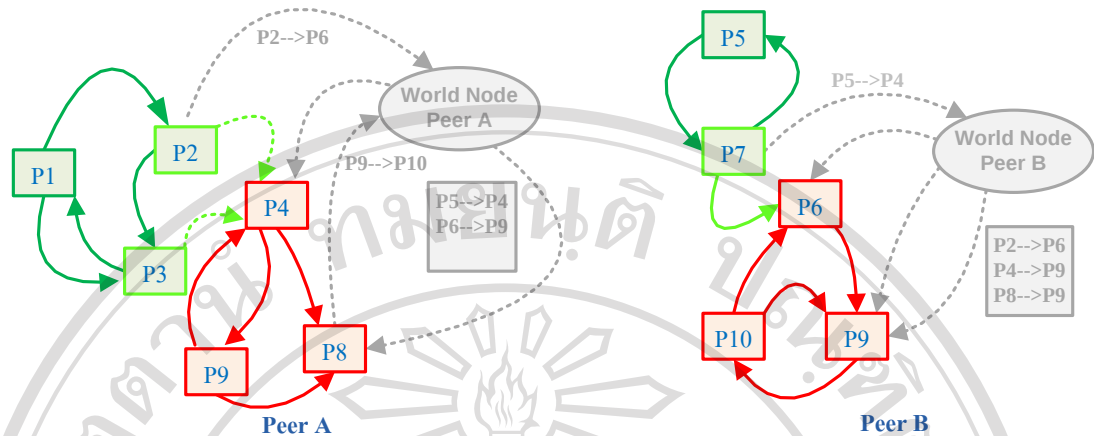
รูปที่ 1.11 โคลอเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์ A และเพียร์ B

จากเว็บลิงค์กราฟในรูปที่ 1.11 เมื่อมีเว็บเพจ P8, P9 และ P10 เพิ่มเข้ามาใหม่ ทำให้โครงสร้างของเว็บลิงค์กราฟในแต่ละเพียร์เปลี่ยนแปลงไป ซึ่งแสดงโครงสร้างของเว็บลิงค์กราฟที่เปลี่ยนแปลงไปของแต่ละเพียร์ในรูปที่ 1.12



รูปที่ 1.12 โครงสร้างของโคลอเว็บลิงค์กราฟที่เปลี่ยนแปลงในเพียร์ A และเพียร์ B

จากรูปที่ 1.12 จะเห็นได้ว่าที่โคลอเว็บลิงค์กราฟในเพียร์ A จะมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่ และที่โคลอเว็บลิงค์กราฟในเพียร์ B จะมีเว็บเพจ P9 และ P10 เพิ่มเข้ามาใหม่ ซึ่งการเพิ่มขึ้นของโหนดดังกล่าว สามารถแบ่งโคลอเว็บลิงค์กราฟในแต่ละเพียร์ได้ดังรูปที่ 1.13

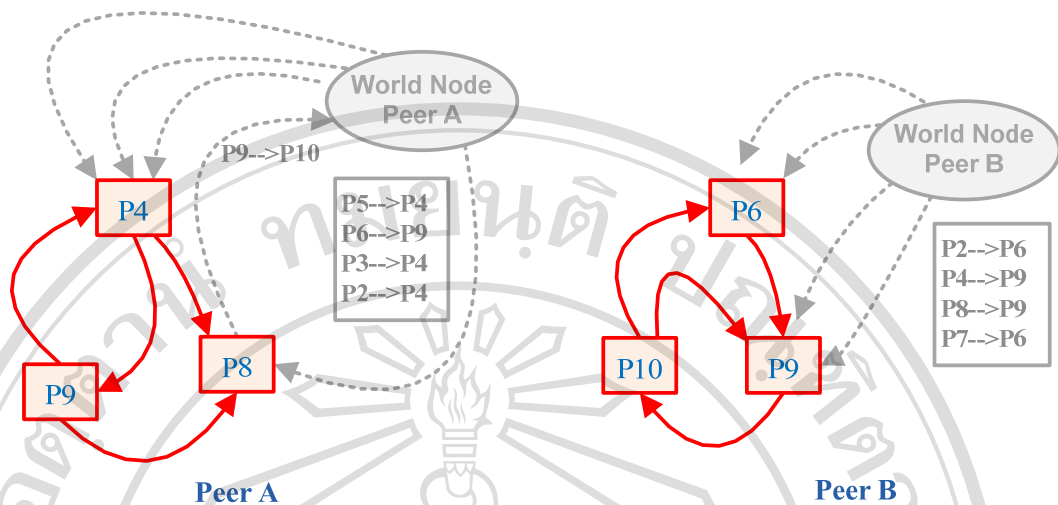


รูปที่ 1.13 การแบ่งส่วนของโลกคอลเว็บลิงค์กราฟในเพียร์ A และเพียร์ B

การแบ่งส่วนที่เพียร์ A จากรูปที่ 1.13 เมื่อมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่จะส่งผลกระทบต่อค่าเพจแรงค์ของโหนดเดสเซ็นด์นทั้งหมด กล่าวคือเว็บเพจ P4, P8 และ P9 ลักษณะดังกล่าวจะทำให้เว็บลิงค์กราฟในเพียร์ A ถูกแบ่งเป็น 2 ส่วนอย่างชัดเจน ดังรูปที่ 1.13 คือกลุ่มของเว็บเพจที่เป็นเดสเซ็นด์นของเว็บเพจที่เพิ่มเข้ามาใหม่ และกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เพียงเล็กน้อยเมื่อมีเว็บเพจใหม่เพิ่มขึ้นมา ซึ่งได้แก่เว็บเพจ P1, P2 และ P3 ซึ่งเว็บเพจทั้งสองกลุ่มจะเชื่อมต่อกันด้วยโหนดขอบ P2 และ P3

ในทำนองเดียวกันการแบ่งส่วนที่เพียร์ B จากรูปที่ 1.13 เมื่อมีเว็บเพจ P9 และ P10 เพิ่มเข้ามาใหม่ จะส่งผลกระทบต่อค่าเพจแรงค์ของโหนดเดสเซ็นด์นกล่าวคือเว็บเพจ P6, P9 และ P10 ซึ่งเว็บลิงค์กราฟจะถูกแบ่งเป็น 2 ส่วนเช่นเดียวกับเพียร์ A คือ กลุ่มของเว็บเพจที่เป็นเดสเซ็นด์นของเว็บเพจที่เพิ่มเข้ามาใหม่ และกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เพียงเล็กน้อย เมื่อมีเว็บเพจใหม่เพิ่มขึ้นมา ซึ่งได้แก่เว็บเพจ P5 และ P7 โดยมีการเชื่อมต่อกันที่โหนดขอบ P7

อย่างไรก็ตาม จะเห็นได้ว่าแต่ละเพียร์การแบ่งส่วนของโลกคอลเว็บลิงค์กราฟจะไม่พิจารณาเว็บลิงค์ที่เชื่อมต่อไปยังเว็ลด์โหนด และส่วนของกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เมื่อมีเว็บเพจใหม่เพิ่มเข้ามา จะถูกนำไปรวมกับเว็ลด์โหนด ดังแสดงตัวอย่างในรูปที่ 1.14 ซึ่งเป็นเว็บลิงค์กราฟที่เกิดจากการพิจารณาเฉพาะเดสเซ็นด์นของกลุ่มของเว็บเพจที่เพิ่มเข้ามาใหม่ ซึ่งจะนำโลกคอลเว็บลิงค์กราฟที่ได้นี้ไปคำนวณค่าเพจแรงค์ตามวิธีการของขั้นตอนวิธี JXP ที่จะกล่าวถึงต่อไปในบทที่ 2



รูปที่ 1.14 แสดงโลกดวลเว็บลิงค์กราฟในเพียร์ A และเพียร์ B ที่นำมาคำนวณค่าเพจแรงค์

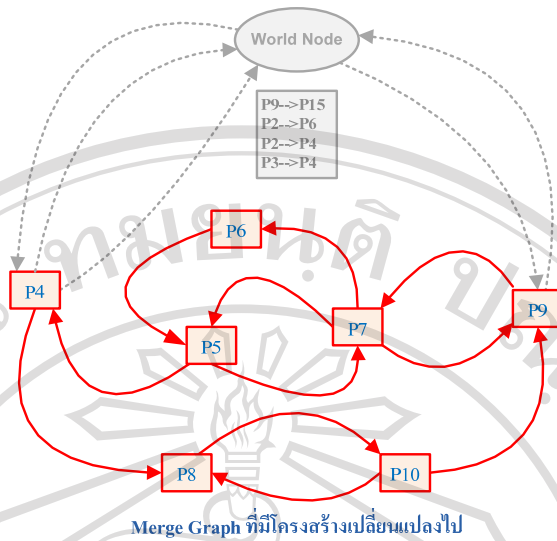
หลังจากคำนวณค่าเพจแรงค์แบบเฉพาะส่วนที่เพิ่มขึ้นมาใหม่ จนครบในทุกๆเพียร์ในเครือข่ายเพียร์ทูเพียร์แล้ว จากนั้นจะเป็นการคำนวณค่าเพจแรงค์ในขณะรวมสองโลกดวลเว็บลิงค์กราฟให้เป็นกราฟใหม่ในกระบวนการจับคู่ระหว่างเพียร์ ซึ่งจะอธิบายรายละเอียดของขั้นตอนวิธีต่อไปในบทที่ 3

ก่อนการเปลี่ยนแปลงโครงสร้างของเว็บลิงค์กราฟในแต่ละเพียร์ เมื่อนำเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์มารวมกันจะได้เว็บลิงค์กราฟใหม่ดังตัวอย่างที่แสดงในรูป 1.15 เมื่อมีเว็บเพจเพิ่มขึ้นมาใหม่ในแต่ละเพียร์ จะทำให้โครงสร้างเว็บลิงค์กราฟในแต่ละเพียร์เปลี่ยนจากเดิม เมื่อมีการจับคู่ระหว่างเพียร์จะทำให้เว็บลิงค์กราฟที่เกิดจากการรวมเพียร์ในรูป 1.15a กลายเป็นเว็บลิงค์กราฟที่เกิดจากการรวมเพียร์ในรูป 1.15b

จากรูป 1.15b จะเห็นว่าเว็บเพจ P8, P9 และ P10 เพิ่มขึ้นมาใหม่ ซึ่งเว็บเพจดังกล่าว เป็นเว็บเพจที่เพิ่มขึ้นมาใหม่ในแต่ละโลกดวลเว็บลิงค์กราฟ ซึ่งเว็บเพจที่เพิ่มขึ้นมาใหม่เหล่านี้ จะเป็นส่วนที่เพิ่มขึ้นมาใหม่ของเว็บลิงค์กราฟที่เกิดจากการรวมสองโลกดวลเว็บลิงค์กราฟเข้าด้วยกันด้วย

ของค่าเพนแรงค์เมื่อมีโหนดใหม่เพิ่มเข้ามา ดังแสดงตัวอย่างของการแบ่งเว็บลิงค์กราฟที่เกิดจากการรวมกันของโหนดกราฟใน 2 เพียร์ดังรูปที่ 1.16

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved



รูปที่ 1.17 แสดง Merge Graph ที่นำมาคำนวณค่าเพจแรงค์

ดังที่กล่าวมาจะพบว่า ทั้งการพิจารณาค่าเพจแรงค์เฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ในโลคอลเว็บลิงค์กราฟของแต่ละเพียร์ และการพิจารณาค่าเพจแรงค์เฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟที่เกิดจากการรวมโลคอลเว็บลิงค์กราฟในเพียร์ เว็บลิงค์กราฟที่ถูกพิจารณา จะมีโครงสร้างที่เล็กลงเมื่อเทียบกับเว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ ดังตารางที่ 1.2 แสดงตัวอย่างการเปรียบเทียบขนาดและเวลาการคำนวณค่าเพจแรงค์ ระหว่างโลคอลเว็บลิงค์กราฟที่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ และไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่

เพียร์	เว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่		เว็บลิงค์กราฟที่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่	
	ขนาด (โหนด)	เวลา (รอบ)	ขนาด (โหนด)	เวลา (รอบ)
Peer 1	1960	17	1897	17
Peer 2	1957	5	1450	5
Peer 3	1957	5	875	4
Peer 4	1957	4	965	3
Peer 5	1947	7	42	1

ตารางที่ 1.2 การเปรียบเทียบขนาดและเวลาในการคำนวณค่าเพจแรงค์ของโลคอลเว็บลิงค์กราฟที่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ และไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่

จากตารางที่ 1.2 จะเห็นว่าเว็บลิงค์กราฟที่พิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่ จะมีจำนวนเว็บเพจที่นำมาคำนวณค่าเพจแรงค่าน้อยกว่าเว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่ แต่จะสังเกตเห็นว่า โคลอเว็บลิงค์กราฟที่เพียร์ 3, เพียร์ 4 และเพียร์ 5 มีจำนวนของเว็บเพจที่จะนำมาพิจารณาค่าเพจแรงค่าน้อยกว่าครึ่งของขนาดโคลอเว็บลิงค์กราฟ ส่งผลให้เวลาที่ใช้คำนวณค่าเพจแรงค่าน้อยกว่าแบบเห็นได้ชัดเจน

1.2 สรุปสาระสำคัญจากเอกสารที่เกี่ยวข้อง

Brin, S. และคณะ [1, 3] ได้เสนอเครื่องมือการค้นหาเอกสารบนเครือข่ายอินเทอร์เน็ต ซึ่งให้ชื่อว่ากูเกิลและเสนอวิธีการคำนวณเพจแรงค้ำ เพื่อจัดลำดับเว็บเพจบนระบบแบบรวมศูนย์

Parreira, J. X. และคณะ [2, 4] ได้เสนอขั้นตอนวิธี JXP ซึ่งเป็นขั้นตอนการจัดลำดับเว็บเพจบนเครือข่ายแบบเพียร์ทูเพียร์

Desikan, P และคณะ [5] ได้เสนอขั้นตอนวิธี IPR ซึ่งเป็นขั้นตอนวิธีการจัดลำดับเว็บเพจแบบเพิ่มขึ้นบนระบบแบบรวมศูนย์

สุมาลี และคณะ [11] ได้เสนอขั้นตอนวิธี ICR ซึ่งเป็นขั้นตอนวิธีการจัดลำดับเว็บเพจแบบเพิ่มขึ้นบนเครือข่ายแบบเพียร์ทูเพียร์

1.3 หลักการและเหตุผล

ในการจัดลำดับเว็บเพจบนเครือข่ายแบบเพียร์ทูเพียร์ สามารถแก้ไขปัญหาการขยายขนาดของเว็บลิงค์กราฟได้อย่างมีประสิทธิภาพ แต่การเพิ่มขึ้นมาใหม่ของเว็บเพจในเว็บลิงค์กราฟจำเป็นต้องเริ่มต้นคำนวณน้ำหนักของแต่ละเว็บเพจใหม่ทั้งหมด วิทยานิพนธ์นี้จึงเสนอแนวทางการจัดลำดับเว็บเพจแบบเพิ่มขึ้นบนเครือข่ายแบบเพียร์ทูเพียร์ขึ้น เพื่อเพิ่มประสิทธิภาพในการจัดลำดับเว็บเพจบนเครือข่ายแบบเพียร์ทูเพียร์

1.4 วัตถุประสงค์ของการศึกษา

พัฒนาขั้นตอนวิธีแบบเพิ่มขึ้นสำหรับการจัดลำดับเว็บเพจบนเครือข่ายเพียร์ทูเพียร์อย่างมีประสิทธิภาพ ซึ่งจะช่วยลดความสิ้นเปลืองของเวลาในการคำนวณ และช่วยลดความสิ้นเปลืองของการสื่อสารภายในเครือข่าย โดยยังคงไว้ซึ่งความแม่นยำในการจัดลำดับ

1.5 ขอบเขตการทำวิจัย

1.6.1 ขั้นตอนวิธี ICR ที่จะพัฒนาขึ้นจะปรับปรุงจากขั้นตอนวิธี JXP ในส่วนของการเริ่มต้นคำนวณน้ำหนักของเว็บเพจใหม่ทั้งหมด โดยนำแนวคิดจากขั้นตอนวิธี IPR มาประยุกต์ใช้ในการแก้ปัญหา

1.6.2 ทดลองโดยใช้ข้อมูลจาก Amazon.com สำหรับสร้างโกลบอลเว็บลิงก์กราฟและโลคอลเว็บลิงก์กราฟ

1.6.3 ทดลองภายใต้สมมุติฐานว่า โกลบอลเว็บลิงก์กราฟมีขนาดที่แน่นอน โดยแบ่งโกลบอลเว็บลิงก์กราฟไปเก็บไว้ในแต่ละเพียร์สำหรับสร้างโลคอลเว็บลิงก์กราฟเพื่อใช้ในการทดลอง

1.6.4 ขั้นตอนวิธีที่จะพัฒนาขึ้นจะไม่ครอบคลุมกรณีที่มีการลดลงของโครงสร้างในเว็บลิงก์กราฟ ทั้งในด้านการลดลงของเว็บเพจ และเว็บลิงค์

1.6.5 ใช้จำนวนข้อความ (Message) ที่ส่งระหว่างเพียร์ เป็นตัวชี้วัดค่าความสิ้นเปลืองด้านการสื่อสารภายในเครือข่าย (Communication Cost)

1.6.6 ใช้จำนวนรอบ (Rounds) ในการปรับค่าเพจแรงค์ให้เข้าใกล้ค่าที่เหมาะสม เป็นตัวชี้วัดค่าความสิ้นเปลืองด้านเวลาการคำนวณ (Computational Cost)