

จุดกำเนิดของ Google.com

นำเสนอเมื่อ : 14 ก.ค. 2551

เอามาให้อ่านกันเป็นความรู้ครับ จาก วิชาการ.คอม โดย ผศ.ดร.บุญญฤทธิ์
อุยยานนวาระสถาบันเทคโนโลยีนานาชาติสิรินธร (SIIT) มหาวิทยาลัยธรรมศาสตร์

เซอร์เก บริน1 (Sergey Brin) และ ลาร์รี่ เพจ (Larry Page)

จุดเริ่มต้นของเรื่องนี้อยู่ที่ มหาวิทยาลัยสแตนฟอร์ด (Stanford University) อีกหนึ่งสุดยอดแห่งมหาวิทยาลัยแดนไอที (จริงๆแล้ว Stanford ก็จัดว่าเป็นอันดับต้นๆของโลกในสาขาอื่นๆอีกมากมาย ด้วยเช่นกัน) มหาวิทยาลัยแห่งนี้ นอกจากจะเป็นต้นกำเนิดของสุดยอด 2 ไอเดียออนไลน์ระดับโลกอย่าง Yahoo! และ Google แล้ว มหาวิทยาลัยแห่งนี้ยังเป็นต้นกำเนิดของคอมพิวเตอร์ที่เราใช้กันอยู่ด้วยซ้ำไป ที่นี้เป็นที่ที่ คุณ John von Neuman (จอห์น วอน นูแมน) คิดและประดิษฐ์คอมพิวเตอร์ที่มีสถาปัตยกรรมที่เหมือนกับที่เราใช้อยู่นี้แหละ และเป็นคนแรกของโลก (แม้จะไม่ใช้คอมพิวเตอร์เครื่องแรกของโลกก็ตาม – สับสนมั๊ยครับ – คืออย่างนี้ครับ คอมพิวเตอร์เครื่องแรกของโลกชื่อว่า ENIAC ซึ่งเป็นเครื่องที่มีความซับซ้อนสูง ต่อมาคุณ von Neuman ซึ่งก็ช่วยงานในการสร้างคอมพิวเตอร์ ENIAC อยู่ด้วย แก่เสนอว่า เราน่าจะแยกส่วนของคอมพิวเตอร์ออกเป็น หน่วยประมวลผล หน่วยความจำ หน่วยรับเข้าและส่งออกข้อมูล ซึ่งเป็นสถาปัตยกรรมที่แตกต่างจากคอมพิวเตอร์เครื่องแรก และเป็นสถาปัตยกรรมที่เรายังใช้มาอยู่จนในปัจจุบัน หลายสถาบันก็เลยยกย่องให้ von Neuman เป็นบิดาของคอมพิวเตอร์เลยก็เดียว) แหม! พูดถึงมหาวิทยาลัยสแตนฟอร์ดชะยี้ดยาวเลย พักไว้แค่นี้ก่อน เตี่ยวหาใครที่กำลังเรียนแถมั้นมาบรรยายสรรพคุณของ Stanford ต่อ ตอนนี่เข้าเรื่องกุเกิ้ลดีกว่าครับ เตี่ยวกระดากจะหมดชะกอน

เรื่องก็เริ่มตอนปิดเทอมภาคฤดูร้อนปี 1995 ที่มหาวิทยาลัยแห่งนี้แหละครับ ตอนนั้น Sergey Brin (เซอร์เก บริน) 1 ใน 2 ของผู้ก่อตั้ง กูเกิ้ล เป็นแค่นักเรียนปริญญาเอกที่กำลังจะขึ้นปี 2 ของภาควิชา วิทยาศาสตร์คอมพิวเตอร์ (Computer Science) ธรรมดาๆ คนหนึ่ง ที่อาสาเข้ามาเป็นนักศึกษาช่วยงาน Open House

โดยปกติทุกๆปีในช่วง ก่อนเปิดเทอมนี้ มหาวิทยาลัยต่างๆเค้าจะมีการเปิดบ้านต้อนรับผู้มาเยือน เราเรียกว่า Open House (ที่เมืองไทยก็เห็นมีบ้างแล้วหลายมหาวิทยาลัย) คือว่าใครที่สนใจที่จะเรียนในมหาวิทยาลัยไหน คณะไหน ก็จะไปงาน Open House ของที่นั่น ที่จะมีคนมาคอยพาทัวร์ และแนะนำสถานที่ แนะนำคณะ แนะนำ Lab แนะนำครุอาจารย์ เป็นปกติเหมือนทุกๆปี แต่ปีนี้เองพระเอกคนที่สองของเรา คือคุณ Larry Page (ลาร์รี่ เพจ) ก็โผล่มาในงาน Open House ในปีนี้ หลังจากได้ติกรี วิศวกรรมศาสตรบัณฑิต จากมหาวิทยาลัยมิชิแกน (Michigan

University) มาหยกๆ

ทั้งสองเจอกันเพราะ Larry Page ไปอยู่ในกลุ่มทัวร์ ที่มี Sergey Brin เป็นหัวหน้ากลุ่มทัวร์พอดี

ดู ทำว่่างานนี้ไม่ใช่รักแรกพบครับ เพราะระหว่างทางที่เดินทัวร์มหาวิทยาลัย และเมือง San Francisco อยู่ ทั้งสองคนก็มีเรื่องให้ตกเตียงกันตลอดเวลา โดยเฉพาะเรื่องของการจัดฝั่งเมืองของ San Francisco (?!?!?)

Page เล่าให้ฟังว่าตอนนั้นเค้าจำได้ว่า Sergey Brin เป็นคนที่มีความคิดที่ค่อนข้างจะติดยึด

เป็น คนที่ไม่ค่อยโศภาที่น่าจะอยู่ใกล้เท่าไร ถ้าคิดว่าตัวเองถูกละก็จะเถียงหัวชนฝา ซึ่งบังเอิญว่าตัวเอง (Page) ก็เป็นคนแบบนั้น ส่วน Sergey ก็บอกว่าจริงๆแล้ว Larry ก็ออกจะแปลกๆอยู่เหมือนกันและเถียงหัวชนฝา ไม่ค่อยยอมใคร (เอากันเข้าไป มีนาหละ เเถียงกันได้ทั้งวัน)

เอาเป็นว่า ทั้งคู่ถกเถียงกันเรื่องต่างๆทั้งวันที่เดินทัวร์ แม้จะไม่ลงรอยด้วยดี แต่ทั้งคู่ก็จำกันได้ดีก่อนจะแยกจากกันในตอนเย็น (แหม! พลืดยังกับหนังไทยเลยครับ พระเอกกับนางเอกเจอกันครั้งแรกจะต้องมีทะเลาะตบตี ต่างคนต่างบอกว่าเกลียด แต่ในใจคิดถึงอยู่)

อีก 2-3 เดือนถัดมา มหาวิทยาลัยก็เปิดเทอมครับ Page ก็เข้ามารายงานตัว และเลือก Prof. Terry Winograd ผู้เชี่ยวชาญด้านการโต้ตอบระหว่างคอมพิวเตอร์และมนุษย์ เป็นอาจารย์ที่ปรึกษา และก็เริ่มมองหาหัวข้อวิทยานิพนธ์

พ่อของ Larry Page (ขณะนั้นเป็นอาจารย์ด้าน Computer Science อยู่ที่มหาวิทยาลัยมิชิแกน) บอกว่า Thesis ปริญญาเอก จะเป็นเหมือนกรอบ ที่จะคอยกำหนดอนาคต ดานวิชาการของเราไปทั้งชีวิต ก่อนจะตัดสินใจเลือกทำให้ไตรตรองให้ดี ทำให้ Page ใช้เวลายาวนานในการเลือกหัวข้อวิทยานิพนธ์ หลังจากลองนึกๆดูสิบกว่าเรื่อง สุดท้ายก็มาลงที่เรื่อง World Wide Web นี้เอง

และแล้วจุดเริ่มต้นของไอเดียเล็กๆ ก็กลายเป็นไอเดียที่เปลี่ยนโลกอินเทอร์เน็ตทั้งใบ จุดกำเนิดของยักษ์ใหญ่ในวงการซอฟต์แวร์อีกตน ก็เริ่มขึ้นที่นี่ ...

อินเทอร์เน็ต ทฤษฎีกราฟ และ BackRub Project

Page เริ่มหัวข้อวิจัยเกี่ยวกับเว็บก็จริง แต่ไม่ได้เริ่มมองหาวิธีที่จะค้นหาข้อมูลบนเว็บ แต่สิ่งที่เค้ามองเห็นกลับเป็นมุมมองทางคณิตศาสตร์ของเว็บไซต์มากกว่า คือ Page มองแบบนี่ครับ...

ถ้าหากมองว่า 1 เซอร์ฟเวอร์ หรือ 1 เว็บไซต์ เช่น วิชาการ.คอม หรือ 1 คอมพิวเตอร์ เป็นเพียง จุด (Node, Vertex) จุดหนึ่งบนกราฟ (Graph) และ ลิงค (link) เช่น www.ipst.ac.th ที่ วิชาการ.คอม เชื่อมต่อไปยังเว็บไซต์อื่นๆ เหมือนกับเป็นทางเชื่อมต่อกัน หรือ ขอบ(Edge) ระหว่างจุดเหล่านั้น หรือ พูดยาวๆว่า Page มองเห็น อินเทอร์เน็ตเป็นกราฟนั่นเอง (แะๆ แบบนี้เรียกว่ามันอยู่ในสัญชาตญาณ มองอินเทอร์เน็ตเป็นกราฟ ทำได้ไงเนี่ย) ซึ่งบ้านเรา นิสิต นักศึกษาระดับปริญญาตรี ด้านทางสายวิทย์ มักจะได้เรียนเรื่อง กราฟ ประเภทนี้ในเรื่อง ทฤษฎีกราฟ (Graph Theory) แถวๆปีต้นๆ (ซึ่งน้องๆตัวเล็กๆอาจสับสนนิดหนึ่ง เพราะกราฟที่คุณเคยอาจจะหมายถึง กราฟที่เป็นตัวแทนของข้อมูล เช่น กราฟแท่ง กราฟเชิงเส้น ซะมากกว่า)

ถ้า ไม่ใช่ในวิชาคณิตศาสตร์ ก็จะเป็นวิชา Algorithm โดยเฉพาะพวกที่เรียน วิทยาการคอมพิวเตอร์ ก็น่าจะผ่านหูผ่านตากันมาแล้วทุกคน (ถ้าไม่ใช่ให้คืนอาจารย์ไปหมดแล้วซะก่อน) (มีอาจารย์และนักคณิตศาสตร์ในเมืองไทยหลายคน ที่เชี่ยวชาญเรื่อง Graph Theory เช่น ดร.จรียา อุษยะเสถียร ภาควิชาคณิตศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่จบปริญญาเอกเรื่องนี้มาโดยเฉพาะ เดี่ยวจะลองเกี่ยวมาช่วยเขียน เรื่อง Graph Theory อีกซักบทความ)

แต่คราวนี้ลองใช้จินตนาการดูกันหน่อยนะครับ ว่าแน่นอนเว็บไซต์หนึ่งเว็บ ก็ลิงค์ไปยังหลายร้อย หลายพันเว็บ และมีเว็บไซต์หลายๆเว็บ ที่ลิงคมายังเว็บไซต์หนึ่งๆ และปัจจุบันเรามีกันเป็น พันๆ ลานเว็บ เพราะฉะนั้น กราฟ ที่เราใช้แทนอินเทอร์เน็ต ก็จะเป็นกราฟขนาดมหึมา และมีความซับซ้อน มีเส้นโยงกันไปโยงกันมาอย่างยุ่งเหยิง

ซึ่งตรงนี้แหละที่ Page มองแล้วเห็นว่ามันช่างน่าตื่นเต้น น่าสนใจ น่าติดตาม เสียเหลือเกิน Page เคยบอกว่า Internet คือ กราฟที่ใหญ่ที่สุดที่มนุษย์เคยสร้างขึ้น และมันก็ยิ่งจะเติบโต ใหญ่ขึ้น ใหญ่ขึ้น ทุกๆวัน ด้วยอัตราเร็วในการเติบโตสูงมาก โอ! มันช่างเป็นเรื่องที่น่าสนใจทำวิทยานิพนธ์เหลือเกิน (ถ้าเป็นคนสามัญชนคนไทยธรรมดา ก็อาจจะบอกว่า โอ! มันซับซ้อนเหลือเกิน ไม่มีทางทำไต่หระก ทำไปเดียวไม่จบ หนีดึกกว่า) ซึ่ง Prof. Winograd อาจารย์ที่ปรึกษาของเค้าก็เห็นด้วย และเห็นว่าน่าจะศึกษาเรื่องของโครงสร้างของกราฟของเว็บ เป็นการเริ่มต้นวิทยานิพนธ์

Page ทำการศึกษาด้วยตัวเองอยู่ไม่นาน เค้าก็เจอปัญหาแรกเข้าให้...

โอเคตรงที่เราหาเรื่อง graph theory กันนิดหนึ่ง (เอาเป็นว่าผมพยายามมองเล็บภาษาอังกฤษของคำไทยไว้ด้วยครับ จะได้ช่วยให้คนที่คุ้นเคยกับคำอังกฤษในวิชาแล้ว ได้เห็นภาพง่ายขึ้นนะ) คืออย่างนี้ ในกราฟปกติ ขอบของกราฟ (Edge) จะเป็นตัวบ่งถึงความสัมพันธ์กันระหว่าง จุด (Vertex) ซึ่งโดยปกติแล้ว เราจะรู้และนับจำนวนได้ว่า จากจุดจุดหนึ่ง มีขอบ หรือ เส้นลากไปยังจุดอื่นๆ อีกกี่จุด และมีกี่จุดที่ลากมาหาตัวเอง แต่หน้าเว็บเพจกลับไม่เป็นแบบนี้ซะทีเดียว เพราะที่หน้าเว็บเพจหนึ่งๆ (สมมติว่าเว็บ 1 หน้าเป็น 1 จุดในกราฟ) เราดูครับว่า จากจุดที่เราอยู่ปัจจุบัน มันลิงค์ไปยังหน้าไหนบ้าง คือ เราดูว่ามีจำนวนขอบที่วิ่งออกไป (Out Degree) จากตัวเองก็ขอบก็เส้น และไปที่ไหนบ้าง แต่ที่เราไม่รู้ก็คือว่า มีเว็บเพจใดบ้าง ที่ลิงคมาหาหน้าที่เราสนใจ

โอ้ย สับสนไหม้ยครับ ผมยกตัวอย่างเปรียบเทียบแบบนี้ก็แล้วกัน ถ้าผมถามคุณว่าคุณรู้จักคนกี่คน คุณอาจจะต้องนั่งไล่นับนิ้วไปเรื่อยๆ แต่คุณก็บอกได้ว่าคุณรู้จักใครบ้าง และคล้ายๆกัน ผมถามว่า "คุณรู้มั่วว่ามีใครในโลกนี้รู้จักคุณบ้าง?" (เปรียบกับ "รู้มั่วว่ามีเว็บไหนลิงคมาที่เราบ้าง") คำตอบคือ ไม่รู้ จะไปรู้อย่างไรว่าใครรู้จักเราบ้าง

โอเค เริ่มเห็นภาพนะครับ ลองนึกตามแบบนี้ครับ ว่าหน้าเว็บที่คุณอ่านอยู่ตอนนี้เนี่ย ลิงคมาจากหน้าไหน URL อะไรบ้าง ถ้าคุณไม่กด Back มีข้อมูลตรงไหนบอกมั่วครับ ? หรือแมแต่คุณจะไม่กด Back คุณก็รู้แต่ลิงคเดียวที่ลิงคมาหาหน้านี้ แต่ที่จริง อาจจะมีหน้าเว็บอื่นๆอีกเป็นร้อยๆ ที่มีลิงคมาหาหน้านี้ ที่เราไม่รู้ คำถามคือ แล้วเราจะรู้อย่างไร ?

ซึ่ง ตรงนี้แหละยาก เพราะอินเทอร์เน็ตไม่ได้ให้ข้อมูลนี้มา และตรงนี้เองที่ Page คิดว่า มันน่าจะดี ถ้าหากว่าเรา (หรืออย่างน้อย มีวิธีการที่จะทำไหว) ว่าใครลิงคหาใครบ้าง หรือ มีใครลิงคมาหาหน้าบ้างก็คน Page ก็เลยเลือกเอาปัญหานี้ มาทำเป็นวิทยานิพนธ์ปริญญาเอก และตั้งชื่อเล่นโปรเจ็คของเค้าว่า "BackRub Project" (โครงการ "ถูหลัง" -แห่ม! ผมแปลตรงตัวไปหน่อยหรือปล่าวเนี่ย - สงสัยว่า Page คงอยากรูว่า ตอนนี้อ่านนี่ ใครถูหลังให้เค้าบ้าง - เอา ! วาไปโน่น)

โอเค งั้นกลับมาที่คำถามเดิม ถ้าเราอยากจะทำว่ามีใครรู้จักเราบ้างก็คน เราจะทำอย่างไร จริงๆคำตอบนี้ง่ายมากครับ คุณก็ถามคนทุกคนทั่วทั้งโลกนี้เลยสิครับว่าเค้ารู้จักใครบ้าง ด้วยวิธีนี้ พอคุณถามครบทุกคนทั้งโลก คุณก็จะรู้ว่าทั้งโลกนี้มีคนรู้จักคุณกี่คนใช่ป่าวครับ แหม! คิดได้ไง ง่ายจัง

คล้ายๆกัน เพื่อจะรู้ว่าใครลิงคมาที่หน้าเว็บนี้บ้าง Page ก็เริ่มจากการไล่ไปที่ละหน้าเว็บแล้วดูว่าหน้าไหนลิงคไปที่ไหนบ้าง (เหมือนว่าหน้าไหนรู้จักใครบ้าง) แล้วเก็บลิงคทั้งหมดในหน้านั้นมาเข้าคิวไว้ เพื่อจะได้ไล่ถามไปเรื่อยๆ (ตรงนี้แหละครับที่เรียกว่า Crawler - หน้าถัดไปจะอธิบายอีกที) Page คิดว่าแห่มไล่ไปเรื่อยๆแบบนี้ ซักอาทิตย์หนึ่งก็เก็บเว็บหมดจากทั่วโลกแล้ว ... นาน.... ก็เป็นจุดเริ่มต้นของ Backrub โปรเจ็คนั่นเอง

จาก BackRub Project ที่ค่อยๆ เติบโตมา ด้วยน้ำมือของนักศึกษา 2 คน ที่ใช้ห้องนอนที่หอพักนักศึกษา ทำเป็น ห้อง Server และ ห้องเขียนโปรแกรม ได้กลายเป็น Google Project ไปๆมาๆ Search Engine ขนาดจิว ที่ดูตลกพิการ Network ของมหาวิทยาลัย สแตนฟอร์ด ที่เดี๋ยวนี้อาจ Network ที่เร็วเป็นอันดับต้นๆของโลก

-



จอมใต้เว็บ Crawler และ PageRank

Page เริ่มที่จะคิดว่า เราจะทำไปถึงจะรู้ว่า ลิงค์ใดบ้างที่ลิงค์มายังเว็บหนึ่งๆ หลังจากที่เราศึกษาเกี่ยวกับเรื่องนี้ได้ไม่นาน(ไม่กี่เดือน) Page ก็พบว่า จริงๆแล้ว เรื่องของการลิงค์ไปลิงค์มาของเอกสารนี้ มีกันมานานแล้วในวงการวิชาการ ก็คือเรื่องของ ผลงานวิชาการ นั่นเอง คือโดยปกติแล้ว หากนักวิชาการท่านใด คิดทฤษฎีอะไรออกมาใหม่ๆ หรือค้นพบอะไรใหม่ หรือต้องการจะแก้ไขสิ่งที่มีอยู่แล้ว ก็จะทำการตีพิมพ์ผลงานของตนเองในวารสารวิชาการ (Journal) โดยจะต้องอ้างอิงถึงที่มาของความรู้ หรือ ผลงานที่มีมาก่อนของผู้อื่น หรือที่ใกล้เคียง ก็เพื่อให้องค์ความรู้ใหม่ที่ตีพิมพ์ มีรากฐานจากองค์ความรู้ ที่ได้รับการพิสูจน์แล้ว (ตีพิมพ์แล้ว) นั่นเอง ดังนั้น ผลงานวิชาการ ไหนที่ได้รับการอ้างอิง (Citation) บ่อยๆ จาก นักวิชาการคนอื่นๆ แสดงว่า ผลงานวิชาการชิ้นนั้นได้รับการยอมรับอย่างจริง ในวงการวิชาการเรามีตัววัดกันอยู่แล้ว ผลงานหนึ่งๆ มีการถูกอ้างอิงมากน้อยเพียงใด เราเรียกดัชนีตัวนี้ว่า Citation index ซึ่งการอ้างอิงด้านวิชาการถือเป็นเรื่องใหญ่ ใหญ่ขนาดไหนครับ ก็ใหญ่พอที่จะมีวิชาที่วาดวยเรื่องนี้โดยเฉพาะเลย คือวิชา bibliometrics (ผมเองก็ไม่เคยเรียนครับ แต่เขาใจว่าน่าจะเป็นทางกลุ่มนักศึกษา วารสาร หรือ บรรณารักษ์ หรือ สารสนเทศ - เดียวจะคนมาหาว่าที่ไหนสอนบ้างในเมืองไทย)

ตอนที่ Tim Berners-Lee (ตอนนี้ได้รับการแต่งตั้ง เป็น Sir Tim Berners-Lee เรียกร้อยแล้ว) วิศวกรอิสระของ CERN คิดค้น World Wide Web ขึ้นเป็นครั้งแรกของโลก (ไว้วันหลังค่อยเขียนเรื่อง จุดกำเนิดของ WWW อีกทีแล้วกัน) คุณ Tim แก่ก็คิดว่า เราน่าจะมีวิธีที่ละลิงค์ผลงานวิชาการของนักวิชาการเขาด้วยกันเลย ไม่ต้องมานั่งกำหนดรูปแบบที่ยุงยาก (เหมือนที่เห็นในกรอบด้านบน) คุณ Tim ก็เลย คิดเรื่องของ Hypertext ขึ้นมา แต่สิ่งที่ Page กำลังทำเป็นการ Reverse Engineer ของ WWW เพราะเขาต้องการค้นหาถึงที่มา ต้นตอของเอกสารที่ลิงค์กันนั่นเอง ด้วยความรู้นี้ งานของเขาก็ง่ายขึ้นเยอะครับ (แม่ที่เหลือกก็จะมีที่สุดที่ตามมา) ที่เหลือกก็คือว่าเขาจะต้องหาให้ได้ว่า ใคร อ้างอิงจาก ใคร โดยอัตโนมัติ พุดงายๆ ว่าเขาต้องวาด กราฟของอินเทอร์เน็ต ขึ้นมา บนเครื่องคอมพิวเตอร์ของเขาเอง แนนอนว่า กราฟที่เขาจะสร้างขึ้น จะมีความซับซ้อนสูง และการคำนวณจำนวนลิงค์ ที่เชื่อมหากันก็ทำได้ยาก เนื่องจากกราฟมีการเจริญเติบโตเรื่อยๆ เพราะความซับซ้อนของข้อมูลสูง ดังนั้นสูตรการคำนวณเพื่อให้คะแนนแต่ละหน้า ก็จะต้องมีความซับซ้อนด้วย ตรงนี้เอง ที่ทำให้ Brin กระโดดเข้ามาในโปรเจกต์นี้

ด้วยพื้นเพ back ground เดิมของ Brin ที่เป็นนักคณิตศาสตร์ ที่จัดว่าเข้าขั้นเป็นอัจฉริยะคนหนึ่ง มีเชื้อสายเป็นคนรัสเซีย เกิดที่รัสเซีย มีพ่อเป็นนักวิทยาศาสตร์ชาวรัสเซีย ที่ทำงานที่องค์การ NASA และเป็นอาจารย์สอนคณิตศาสตร์ ที่มหาวิทยาลัยแมริแลนด์ (University of Maryland) โดยครอบครัวเขาอพยพ มาอยู่ที่อเมริกา ตอนที่ Brin อายุแค่ 6 ขวบ Brin เรียนจบ ม.ปลาย 1 ปีก่อนเข้าวามัน และหลังจากจบปริญญาตรี ที่แมริแลนด์ Brin ก็มาต่อเอกที่ Stanford ตัว Brin เองก็ต้องมองหาโปรเจกต์ปริญญาเอก ด้วยเช่นกัน แต่เขาเลือกไปเลือกมาเกือบ 2 ปีแล้ว ก็ยังหาหัวข้อลงตัวไม่ได้ จนได้เขามคลุกคลีกับโปรเจกต์ของ Page ก็เกิดความสนใจ ที่จะเข้ามาทำในสวนคณิตศาสตร์ ของโปรเจกต์นี้ และอีกสาเหตุก็คือเขาชอบ Page (ฮันแน! กะแล้ว เหมือนหนังไทยไม่มีผิด)

การ สร้างกราฟของอินเทอร์เน็ตโดยอัตโนมัติ เป็นจุดเริ่มให้ Page เขียนโปรแกรมเล็กๆ ประเภท Crawler ขึ้นมาตัวหนึ่ง ในห้องนอน ตอนที่ Page เริ่มเขียน crawler นี้ จำนวนหน้าเว็บทั่วโลกก็มีอยู่ประมาณ 10 ล้านหน้าเห็นจะได้ แต่จำนวนลิงค์ที่เชื่อมกันอยู่นั้นนับไม่ถ้วน โดยหวังจะให้เจ้า Crawler ไต่ไปเก็บข้อมูลมาสร้างเป็นกราฟโดยอัตโนมัติ ใน ตอนนั้น เขาอาจจะยังไม่รู้อีก ว่าโปรแกรมเล็กๆที่เขาเริ่มเขียนในห้องนอน จะเป็นสิ่งที่ประสบความสำเร็จที่สุดในโลก ต่อจาก Internet

หลาย คนคงอาจจะยังไม่คุ้นกับคำว่า Crawler ผมขอยายความให้อีกนิด Crawler เป็นโปรแกรมเล็กๆโปรแกรมหนึ่ง ที่ทำหน้าที่ไปดึงเว็บเพจต่างๆมา ซึ่งปกติแล้วข้อมูลแสดงหน้าเว็บไซต์ที่เราเห็นนี่ เป็นแค่ text file หรือ ข้อมูลตัวอักษรธรรมดาๆนี่เอง (ลองกดที่เมนู view->source ดูนะครับ นั่นแหละคือข้อมูลของหน้าเว็บที่แท้จริง) พอโปรแกรมประเภท บราวเซอร์ เช่น IE หรือ Firefox ได้รับข้อมูลพวกนี้แล้วมันก็จะทำการแปล และแสดงให้เป็นสิ่งที่คุณเห็นบนหน้าจอตอนนี้ก็) (อีกที)

หลังจากที่โปรแกรมประเภท crawler ได้รับข้อมูลมาแล้ว มันก็จะทำการแยกข้อมูล และลิงค์(ที่จะไปหน้าอื่น) ออกมา

สมมุติ ว่าหน้าที่คุณอ่านอยู่ตอนนี้ มีลิงค์ออกจากมันไป อีกประมาณ 30 ลิงค์ เจ้าตัว crawler ก็จะทำการจัดการเอาลิงค์เหล่านี้มาเขาคิวเรียงกันไว้ แล้วก็ไล่ไปทีละลิงค์ตามคิว แล้วก็ไปดึงข้อมูลเหล่านั้นมา แล้วแยกลิงค์แบบเดิมอีก แล้วลิงค์ที่ไต่จากหน้าถัดไปนี่ก็จะเอามาเขาคิว เรียงต่อกันไปเรื่อยๆ เพื่อจะทำการไปดึงข้อมูลมาในเวลาถัดๆไป เพราะฉะนั้นมันก็เลยให้ความรู้สึกคล้ายๆกับว่า เจ้า crawler มันค่อยๆคืบคลานออกจากจุดเริ่มต้นไปที่ละน้อย ทีละน้อย

และแล้ว ในเดือนมีนาคม 1996 (เพียงแค่นี้ไม่ถึงปีจากที่เค้าเริ่มศึกษา) Page ก็ปล่อยเจ้า crawler ตัวแรกให้เริ่มทำงาน โดยไต่จากหน้าเว็บเพจของตัวเอง ที่อยู่บนเว็บไซต์มหาวิทยาลัยสแตนฟอร์ด เจ้า crawler เวอร์ชันแรกของ Page ไต่ไปตามเว็บเพื่อเก็บแค้ ชื่อเว็บ และ ข้อมูลใน header เท่านั้นเอง แต่ที่ถือว่าเป็นจุดเริ่มต้นอย่างเป็นทางการของ Google (ที่ในปัจจุบันกลายเป็น ซุปเปอร์อภิมหาอมตะนิรันดร์กาล crawler ไปแล้ว เพราะมันไต่ไปเก็บข้อมูลทุกอย่าง ของทุกหน้าเว็บ) เพราะตอนนั้นขึ้นเก็บทุกอย่าง ทรัพยากรของระบบ เช่น memory หรือ ฮาร์ดดิสต์ ที่จะต้องใช้ในการจัดการกับข้อมูลเหล่านี้ ก็คงต้องมีขนาดใหญ่มที่มาก และมันก็มากเกินไปจนจะเป็นโปรเจกต์ของเด็กนักเรียนคนหนึ่ง

โปรแกรมเล็กๆ ที่ถูกปล่อยออกจากห้องนอนที่หอพักนักศึกษาคนหนึ่ง ได้กลายเป็นปรากฏการณ์ ที่ใหญ่ที่สุด รองจากการคิดค้นอินเทอร์เน็ต

จาก โปรเจ็ค BackRub Project สู่ โปรเจ็ค Google

Page และ Brin ได้ร่วมกันคิดหาสูตร หรือ วิธีคำนวณว่า จะให้คะแนนแต่ละหน้าเว็บเพจอย่างไรดี

เว็บ เพจหนึ่งหน้า ถ้ามองให้ดีมันก็เป็นเหมือนผลงานวิชาการชิ้นหนึ่ง ภายในผลงานวิชาการนี้ ก็จะมีการอ้างอิงผลงานคนอื่น (หรือลิงค์ไปหาคนอื่นนั่นเอง) (แต่บางทีเราอาจจะไม่ เป็นแบบนี้ เพราะบางทีเรา ภาเห็นว่าเรื่องไหนดี ก็ใช้วิธี cut-paste เนื้อหานั้นไปเลย โดยไม่ได้อ้างอิงที่มา นอกจากจะละเมิดลิขสิทธิ์แล้ว ยังถือว่าเป็นการเสียมารยาทอย่างรุนแรงด้วยครับ แม้วางจะนำไปใช้ในทางไม่ถูกรายได้ ที่อาจจะไม่ผิดลิขสิทธิ์ในบางกรณี แต่ก็น่าจะมีมารยาทในการอ้างอิงที่มาของเนื้อหาด้วย ซวยๆกันผลิตภัณฑ์ใหม่ในไปในทางถูกต้องมากขึ้นนะครับ) เพราะฉะนั้นหากหน้าเว็บใดมีคนอ้างอิงถึง (ลิงค์) ถึงหน้านั้นเยอะ ก็น่าจะแสดงว่าหน้านั้นมีข้อมูลที่ดี หรือ น่าเชื่อถือ ซึ่งหากมีการให้คะแนนแต่ละหน้าเว็บ Page คิดว่าการอ้างอิงจากเว็บอื่นก็จะมีส่วนสำคัญต่อคะแนน

ภายหลังเค้าตั้งชื่อเล่นระบบคิดคะแนนของเค้าว่า Pagerank (ซึ่งล้อค่า 2 ความหมาย คำแรกคำว่า Page หมายถึงได้ทั้งหน้าเว็บ หรือ หมายถึงชื่อของเค้าเอง) ซึ่งก็คือเลียนการคิดคะแนนมาจากการอ้างอิงกันของผลงานวิชาการ เพราะเค้ารู้ว่า การอ้างอิงกันของเอกสารทำอะไร มีมีการให้คะแนน Citation Index อย่างไร และเค้าก็ได้เพิ่มเรื่องของกราฟและลดคะแนนพิเศษด้วย หากวลิงค์ที่ลิงคมาหาหน้าใดหน้าหนึ่ง เป็นลิงค์ที่มีคะแนนสูง เป็นเว็บที่มีความน่าเชื่อถือ เว็บที่ถูกลิงค์ ก็จะได้รับคะแนนสูงด้วย และหากนำเว็บมาเรียงลำดับกัน เว็บที่มีคะแนนสูงกว่า ก็จะอยู่ลำดับต้นๆ ส่วนเว็บที่มีคะแนนต่ำก็จะอยู่ท้ายๆ

สมมุติว่าเว็บ วิชาการ.คอม มีลิงค์ไปยังเว็บของ สสวท. ในขณะที่เดียวกัน เว็บของเด็กชาย ก. จากจังหวัดสงขลา ก็ลิงคมายังเว็บของ สสวท เช่นกัน แต่มีคนลิงคมาหา เว็บ วิชาการ.คอม อยู่ก่อนแล้ว สมมุติว่าเป็น 1000 ลิงค์ ในขณะที่มีคนลิงคไปหาเว็บของ เด็กชาย ก. เพียง 10 ลิงค์ ทำให้ วิชาการ.คอม มีคะแนน สูงกว่าเว็บของ เด็กชาย ก. ดังนั้น เว็บของ สสวท. ก็จะให้คะแนน จาก เว็บของ วิชาการ.คอม มากกว่า จากเว็บของ เด็กชาย ก. ด้วยนั่นเอง (คือไม่ได้นับเฉพาะจำนวนของลิงค์)

สังเกตว่า เว็บสีแดงด้านบนจะมีขนาดใหญ่กว่าเว็บสีเขียวด้านล่าง แม้จะมีลิงค์เข้ามาแค่ 1 ลิงค์ จากสี่สี่หมื่น เท่านั้น แต่เพราะเว็บสีส้มมีความสำคัญสูง เพราะได้รับการลิงคมาจากหลายที่ เว็บสีแดงด้านบนก็เลย มีความน่าเชื่อถือด้วย (ดังนั้นถ้าเว็บไหน ได้รับการลิงคจากเว็บที่เป็นที่ยอมรับ ก็จะได้รับคะแนนยอมรับด้วย)

ตอนที่ทั้งสองคนคิดระบบให้คะแนนนี้ขึ้น ทั้งสองคนไม่ได้คิดถึงเรื่องของการค้นหาข้อมูลบนเว็บเลย ที่คิดอยู่ในหัวก็มีแต่เรื่องที่ว่า จะค้นหาให้ได้ว่าใครลิงคมาที่เว็บเพจไหนบ้าง (Backlinks) โดยที่ทั้งสอง ทำโปรเจ็ค Backrub นี้ มาจนถึงขั้นที่ รับ ลิงค์ (URL) มาหนึ่งลิงค์ มันก็จะให้ ผลลัพธ์ มาเป็น ลิงค์ที่ลิงคมาหา (backlinks) ทั้งหมดที่ลิงคมาหาหน้าที่กำหนด โดยเรียงลำดับตามความสำคัญ (เช่นถ้าใส่ เว็บ สสวท ไป ก็จะได้ทั้งเว็บ วิชาการ.คอม และ เว็บเด็กชาย ก. และเว็บอื่นๆ ที่ลิงคมาหา สสวท เป็น ผลการค้นหา แต่เว็บ วิชาการ.คอมจะอยู่ด้านบนกว่า เว็บของ เด็กชาย ก. เป็นต้น)

ซึ่ง เค้าก็พบว่าจริงแล้ว มันสามารถประยุกต์ ไปทำเป็น Search Engine น่าจะได้ หลังจากเล่นไปเล่นมากับ กราฟของเว็บ ที่ทั้งคู่สร้างขึ้นมา พบว่า การค้นหาข้อมูลทำได้รวดเร็ว และ ถูกต้อง อยางไรก็ตามเชื่อ ชนิดที่ทั้งสองคนยังงงว่าทำไมโอเคเดียวที่จะใช้เรื่องนี้ทำ Search Engine ถึงไม่ผุดมาในหัวตั้งแต่ตอน เพราะมันชัดเจนเหลือเกิน

หลังจากทดลองปรับแต่ง Backrub ให้กลายเป็น โปรแกรม Search engine ทำการค้นหาข้อมูล บนหน้าเว็บไหนใดหน้าหนึ่ง (ผม ละเชื่อเค้าเลย สองคนนี่ ปกติแล้วเรื่อง การค้นหาข้อมูลใน document นี้ไม่ใช่เรื่องง่าย เพราะไม่ได้คนกันแบบ เค้าค่าเทียบค่า แต่ต้องมีการแปลงรูปแบบของ ข้อมูลให้เป็นเมตริกซ์ ทั้ง SVM กับ LSI (Latent Semantic Indexing - เป็นวิธีการที่กำลังมาแรง) ก็สุดหิน ที่อยู่ในวิชา Information Retrieval - แต่สองคนนี่เล่นเรียนรูเรื่องเหล่านี้ชนิดที่ประมาณว่า ชิวชามคีน) เค้าพบว่าผลการค้นหา ดีกว่า search engine ที่มีอยู่ในตอนนั้น เช่น AltaVista หรือ Excite มาก โดย Search Engine ที่มีอยู่นั้นมักจะให้ผลการค้นหา ที่ไม่ค่อยจะเกี่ยวข้องกับเรื่องที่ต้องการหาจริงๆ เพราะอาศัยเพียงแค่การจับคู่คำ ไม่คำนึงถึงสัญญาณ หรือข้อมูลอื่นๆที่ใกล้เคียง search engine ใหม่ของพวกเค้า ไม่เพียงว่าผลการค้นหาจะดี แต่ ระบบการให้คะแนนนี้จะดีขึ้นเรื่อยๆ เมื่อเว็บใดขึ้นเรื่อยๆ กราฟของเว็บใดขึ้นเรื่อยๆ ผลการค้นหา ก็จะดีขึ้นเรื่อยๆ เพราะจะมีตัวให้คะแนนซึ่งกันและกัน เยอะขึ้น ตรงนี้เองที่ ทั้ง Brin และ Page รู้ว่าเค้า สะดุดชุมทรัพย์ มหิมาเขาให้แล้ว...

และ แล้ว โปรเจ็ค search engine โดยนักศึกษา 2 คน ... ก็เริ่มเป็นรูปร่างที่ชัดเจน , เค้าเริ่มขอบริจาคเครื่องคอมพิวเตอร์เก่าๆที่มหาวิทยาลัยไม่ได้ใช้แล้วมาต่อ กันเป็น server เค้าเริ่มจะหาเงินมาซื้อ harddisk เพิ่มเติม เพราะตอนนี้ข้อมูลที่ crawler เก็บมามันได้ขึ้นเรื่อยๆ

ทั้งสองคนจึงตั้งชื่อ search engine ตัวใหม่ของเค้าว่า Google ที่เพี้ยนมาจากคำว่าgoogol (ที่แปลว่า เลข หนึ่ง ตามด้วย ศูนย์ 100 ตัว - สังเกตสัญลักษณ์ของ Goooooooooooooooooooooogle) และ

Server ของ Google เมื่อปี 1998 สมัยที่ยังอยู่ที่ Stanford เป็นเครื่องที่ได้จากการบริจาคทุกเครื่อง (ปัจจุบัน Search Engine ของ Google ทำงานอยู่บน Linux Server Farm จำนวนประมาณ 250,000 เครื่อง)

Google เวอร์ชันแรก ถูกปล่อยออกสู่สาธารณะชน ใ้บนูเว็บของมหาวิทยาลัย Stanford เป็นครั้งแรก ตอนเดือนสิงหาคม 1996 - เพียง 1 ปี หลังจากที่ทั้งคู่เจอกันครั้งแรก - เพียงไม่กี่เดือนนับจากเริ่มศึกษาปัญหา - โอ จอร์จ ทำได้ยังไงเนี่ย ! ! ! ! อัจฉริยะจริงๆ ! ! !

ที่มา <http://webboard.hitech.sanook.com/forum/?topic=2422729>