

โปรแกรมค้นหา (search engine): การสืบค้นและการประเมิน

พิมพ์ร่ำไพ เปรมสมิทธิ์*



ทรัพยากรสารสนเทศบนอินเทอร์เน็ตมีมากมาย บริการหนึ่งที่เป็นที่นิยมมากในขณะนี้ก็คือ เวิลด์ไวด์เว็บ ซึ่งคงจะเนื่องมาจากการที่เวิลด์ไวด์เว็บมีลักษณะการเชื่อมประสานที่ง่ายต่อการใช้ รวมทั้งการนำเสนอในรูปแบบสื่อประสมที่ดึงดูดใจทั้งผู้ใช้และผู้ให้บริการสารสนเทศ เวิลด์ไวด์เว็บช่วยให้ผู้ใช้สามารถเข้าถึงเอกสารอิเล็กทรอนิกส์บนอินเทอร์เน็ตที่อยู่บนเว็บเพจต่าง ๆ รวมทั้งการเชื่อมโยงไปยังเอกสารอื่น ๆ ได้ ดังนั้นเวิลด์ไวด์เว็บจะมีข้อมูลจำนวนมากและการสืบค้นข้อมูล จะทำได้ยากมาก หากไม่มีเครื่องมือใด ๆ มาช่วยในการค้น

วิธีการในการค้นหาข้อมูลในเว็บมีหลายวิธี ผู้ใช้อาจจะเข้าถึงเว็บเพจโดยตรงโดยใช้ที่อยู่เว็บที่ทราบ หรือใช้ตัวเชื่อมในระบบไฮเปอร์เท็กซ์ไปยังเว็บเพจอื่น ๆ ที่มีข้อมูลที่เกี่ยวข้อง แต่ที่สำคัญก็คือใช้โปรแกรมค้นหาที่ผู้ใช้สามารถระบุว่าจะค้นหาข้อมูลใดที่ต้องการ และแสดงผลข้อมูลที่เกี่ยวข้องตามที่ผู้ใช้ระบุ

โปรแกรมค้นหา เริ่มปรากฏในปี ค.ศ. 1994 (Chu and Rosenthal 1996: 1 ; Schwartz 1998: 1) เพื่อช่วยในการค้นหาข้อมูลที่มีในเว็บเพจที่มีเพิ่มมากขึ้น โปรแกรมค้นหาเป็นประเด็นที่ได้รับความสนใจในการศึกษาและวิจัยมาก โดยเฉพาะในเรื่องคุณลักษณะ ขีดความสามารถในการตอบสนองความต้องการของผู้ใช้ และความแตกต่างในการช่วยค้นหาข้อมูล

ประเภทของโปรแกรมค้นหา

โปรแกรมค้นหาเป็นชื่อเรียกโดยรวมของเครื่องมือที่ช่วยค้นหาเว็บเพจ โดยทั่วไปแล้ว โปรแกรมค้นหาที่มีหน้าที่หลัก 3 ประการ คือ (Gordon and Pathak 1999: 142)

1. รวบรวมกลุ่มของเว็บเพจออกมาเป็นสิ่งที่มวลที่ผู้ใช้จะสามารถค้นคืนสารสนเทศ
2. แสดงเว็บเพจในลักษณะที่แสดงให้เห็นถึงความพยายามในการที่จะเลือกและสกัดเนื้อหาสาระออกมา

* ผู้ช่วยศาสตราจารย์ประจำภาควิชาบรรณารักษศาสตร์ คณะอักษรศาสตร์

3. ให้ผู้ใช้สามารถตั้งคำถาม และใช้ขั้นตอนวิธี (algorithm) ในการค้นคืนที่จะช่วยให้ได้เว็บเพจที่ตรงตามต้องการมากที่สุด

ซึ่งหน้าที่ต่าง ๆ เหล่านี้ก็อาจแตกต่างกันไปในแต่ละโปรแกรมค้นหา

โปรแกรมค้นหา อาจแบ่งออกเป็น 2 ประเภท คือ

1. ประเภทจัดหมวดหมู่ / กลุ่มเรื่อง(directory)
2. ประเภทใช้โปรแกรมค้นหา (search engine) หรือโปรแกรมค้นหาโดยการสอบถาม (query-based engine)

1. ประเภทจัดหมวดหมู่ / กลุ่มเรื่อง

โปรแกรมค้นหาที่ใช้วิธีการนี้ที่มีชื่อเสียงมากก็คือ Yahoo ที่มีการจัดเรียงเว็บเพจเป็นหมวดหมู่ เรียงลำดับโดยมีหมวดหมู่ใหญ่ 14 หมวด และมีการแบ่งย่อยลงไปเรื่อย ๆ จนถึงรายการเว็บเพจในหัวข้อใดหัวข้อหนึ่งโดยเฉพาะ ในกรณีที่ผู้ใช้ต้องการค้นหาข้อมูลเกี่ยวกับโปรแกรมการศึกษาในสาขาบรรณารักษศาสตร์และสารนิเทศศาสตร์ อาจเริ่มค้นจาก Education แล้วเลือก Higher Education ตามด้วย Graduate Education จากนั้นเลือก By Subject จนถึง Library and Information Science หรือผู้ใช้อาจเลือกหมวดหมู่ Reference แล้วเลือก Libraries ต่อด้วย Library and Information Science จากนั้นไปสู่ Education แล้วคลิก College and Universities ไปถึง Departments and Programs จะเห็นได้ว่าวิธีการเข้าถึงแบบนี้จะมีลักษณะของการจัดหมวดหมู่ ลดหลั่นกันเป็นลำดับขั้น และผู้ใช้สามารถเลือกทาง (path) ได้ตามต้องการ ผู้เชี่ยวชาญเฉพาะสาขาหรือบรรณารักษศาสตร์จะเป็นผู้เลือกเว็บเพจ และระบุตำแหน่งในลำดับขั้น และยังมีการทำการปริทัศน์ (review) เว็บไซต์ต่าง ๆ ไปด้วย

ผู้จัดทำเว็บไซต์อาจมีการเปลี่ยนแปลงข้อมูลบนเว็บเพจได้โดยไม่มีผลกับรายการ/รายชื่อ เนื่องจากการค้นหาจะพยายามจับคู่ให้ตรงกับคำอธิบาย หรือข้อความที่ส่งเข้าไปในบัญชีรายชื่อ

2. ประเภทใช้โปรแกรมค้นหา

โปรแกรมค้นหาจะค้นหาข้อมูลบนเว็บโดยอัตโนมัติ แล้วจะทำดรรชนีจากเว็บเพจที่อ่านได้ และยังทำการตรวจสอบการเชื่อมโยงอีกด้วย องค์ประกอบที่สำคัญ 3 ประการของโปรแกรมค้นหา คือ

- 1) สไปเดอร์ (spider) หรือเรียกอีกอย่างหนึ่งว่าโรบอต (Robot) หรือครอเลอร์ (crawler) สไปเดอร์จะเข้าไปอ่านข้อมูลในเว็บเพจ และตรวจสอบตัวเชื่อมโยงหรือลิงก์ (link) ไปยังเว็บเพจอื่น ๆ ในเว็บไซต์ และยังมีมีการกลับมาอ่านเพจในเว็บไซต์อย่างสม่ำเสมอ เช่น อาจจะทำทุกเดือนหรือทุก 2 เดือน โปรแกรมค้นหาบางโปรแกรมเช่น Alta Vista มีสไปเดอร์ที่ชื่อ "Scooter" ที่เก็บรวบรวม



ข้อมูลจากเว็บเพจจำนวนประมาณ 6 ล้านหน้าทุกวัน (Schwartz 1998) เพื่อสำรวจว่ามี การเปลี่ยนแปลงอะไรบ้าง สิ่งทีส่งไปเดออร์คั้นพบจะส่งไปยังอินเด็กเซอร์ (indexer) เพื่อทำดรรชนี (index)

2) ดรรชนี หรือในบางครั้งจะเรียกว่ารายการ (catalog) เป็นเสมือนสมุดเล่มใหญ่ที่มี ลำเนาของเว็บเพจทั้งหมดทีส่งไปเดออร์คั้นพบ ถ้าหากเว็บเพจมีความเปลี่ยนแปลงก็จะมีการปรับปรุง ข้อมูลให้ทันสมัย

เว็บเพจจะได้รับการจัดระบบเป็นกลุ่มทีช่วยให้ง่ายต่อการค้นหา เช่น ค้นหาบุคคล ค้นหา หน่วยงาน เป็นต้น และบางโปรแกรมก็จะได้ให้การเชื่อมโยงไปยังเว็บเพจต่าง ๆ และบางโปรแกรม ก็จะมีช่องกรอกคำค้นทีจะส่งไปยังโปรแกรมค้นหาอื่น ๆ ด้วย ในการแสดงผลการค้นหาจะทำเอา เว็บไซด์ทีเกี่ยวข้องมากที่สุดมาไว้ในรายการแรก และเว็บไซด์ทีมีจะเป็นเว็บไซด์ทีอยู่ในระบบ เท่านั้น ผู้สร้างเว็บไซด์จะเป็นผู้ทีแจ้งมายังผู้ทีดูแลระบบ ซึ่งจะจำแนกหมวดหมู่ให้ ในกรณีเช่นนี้ การสืบค้นอาจจะได้รับผลลัพธ์ทีตรงตามความต้องการมากขึ้น อย่างไรก็ตามการค้นแบบนี้ก็ยังมีข้อ จำกัดตรงทีไม่ได้ค้นข้อมูลบนเว็บทั้งหมด ส่วนมากแล้วการค้นหาแบบนี้จะตรวจตรารายการที กำหนดโดยผู้ทีดูแลและให้บริการเท่านั้น และมักจะทำการค้นหาเฉพาะทีโฮมเพจหรือหน้าแรกของ เว็บไซด์เท่านั้น

การทำดรรชนีของโปรแกรมค้นหาส่วนใหญ่จะเป็นลักษณะการทำดรรชนีเนื้อหาเต็ม ซึ่ง หมายความว่า คำทุกคำทีมีความหมายจะถือเป็นดรรชนี ยกเว้นคำทีไม่ใช่เป็นคำคั้น (stop word) เช่น a, an, the, and, but เป็นต้น อย่างไรก็ตามการทำดรรชนีของโปรแกรมคั้นแต่ละโปรแกรม อาจมีความแตกต่างกันออกไป บางโปรแกรมจะไม่ทำดรรชนีของข้อมูลทีอยู่ในเขตข้อมูลเมทา (meta field) เช่น เขตข้อมูลคำหลัก/คำสำคัญทีผู้แต่งกำหนด เป็นต้น บางโปรแกรมก็จะเลือกทำ ดรรชนีเฉพาะบางส่วน เช่น หัวเรื่องระดับสูงสุด (top-level headings) เท่านั้น นอกจากนั้นโดย ปกติจะไม่มีการทำดรรชนีของกราฟิกทีปรากฏบนเว็บเพจ

3) ซอฟต์แวร์ ซึ่งทำหน้าที่พยายามรับคำถามหรือคำคั้น และค้นหาในดรรชนี จากนั้นก็จะ นำผลทีได้มาจัดลำดับตามสิ่งที่เห็นว่าตรงตามความต้องการมากที่สุด

การใช้โปรแกรมค้นหา นั้น มีข้อเสียอยู่ทีผู้ใช้อาจได้ข้อมูลมากมายเกินไป และอาจไม่ใช่ ข้อมูลทีตรงตามความต้องการ ทำให้ผู้ใช้ต้องมากลั่นกรองและเลือกผลลัพธ์ทีได้จากการคั้นอีกครั้ง หนึ่ง

โปรแกรมค้นหาทีมีข้อเสียและเป็นทีนิยมในปัจจุบัน อาจไม่จัดอยู่ในประเภทที่ 1 หรือ ประเภทที่ 2 แต่จัดเป็นประเภทลูกผสม (hybrid search engine) ทีมีทั้งการจัดหมวดหมู่ทีผู้ เชี่ยวชาญมาทำ และมีการใช้ส่งไปเดออร์คั้นออกไปเก็บคำ และมีการทำดรรชนี เช่น Lycos ทีเริ่มด้วย การใช้ส่งไปเดออร์ และต่อมาในเดือนเมษายน 1999 มีการเปลี่ยนแปลงมาสู่แบบการจัดหมวดหมู่

เหมือน Yahoo รายชื่อหลักจะมาจาก Open Directory Project และผลลัพธ์รองจะมาจากการใช้สไปเดอร์ตรวจตราเว็บ

การสืบค้น

หน้าที่สำคัญของโปรแกรมค้นหาคือ ช่วยผู้ใช้ให้สามารถค้นและได้รับสารสนเทศตามที่ต้องการ ในบางครั้งโปรแกรมค้นหาอาจให้ความช่วยเหลือในการค้นได้ไม่มากนัก ผู้ใช้อาจได้ผลลัพธ์การค้นที่น้อยหรือพบรายการที่ไม่ตรงตามความต้องการ อย่างไรก็ตาม โปรแกรมค้นหาจะใช้เทคนิควิธีการในการค้นคืนสารสนเทศ ซึ่งรวมถึงขีดความสามารถในการสืบค้นโดยใช้ตรรกบูลีน (Boolean logic) การตัดคำ (Wildcard & truncation) การค้นโดยใช้รากคำ (stemming) การค้นโดยใช้วลี (phrase searching) เป็นต้น หรือใช้เทคนิควิธีการขั้นสูง (advanced search) ที่ช่วยในการปรับกลวิธีการสืบค้น เช่น เลือกจำกัดการสืบค้นจากผลลัพธ์ที่ได้เท่านั้น เป็นต้น

โดยทั่วไปเมื่อต้องการสืบค้น ผู้ใช้จะใส่คำค้นลงในช่องกรอกคำค้น และอาจใช้เทคนิควิธีการต่าง ๆ ในโปรแกรมค้นหาจะมีตัวดำเนินการโดยปริยาย (default) อยู่บ้าง เช่น การค้นหาโดยวลีโดยอัตโนมัติ (automatic phrase searching) การตัดคำ การค้นทุกรูปแบบของคำ เช่น study, studied, studying เป็นต้น และโปรแกรมค้นหาส่วนมากจะพยายามช่วยเหลือผู้ใช้เพิ่มมากขึ้นด้วยสิ่งที่เรียกว่าการ สืบค้นขั้นสูง หรือการสืบค้นแบบมีพลัง (power search)

ในตอนนี้จะกล่าวถึงการค้นโดยใช้วิธีการต่าง ๆ ที่ปรากฏในโปรแกรมค้นหาซึ่งต่างโปรแกรมก็อาจจะใช้วิธีการแตกต่างกันไป ผู้ที่สนใจอาจเข้าไปศึกษาจากบริการโปรแกรมค้นหาแต่ละโปรแกรม หรืออาจดูจากบริการที่ศึกษาค้นคว้าเกี่ยวกับโปรแกรมค้นหาโดยเฉพาะ เช่น เข้าถึง <http://www.searchenginewatch.com> หรือ <http://www.searchengineshowdown.com> ก็จะได้รับข้อมูลเกี่ยวกับคำสั่งและวิธีการค้น

การสืบค้นโดยใช้ตรรกบูลีน

โปรแกรมค้นหาแทบทุกโปรแกรมจะเปิดโอกาสให้ผู้ใช้ใช้ตัวดำเนินการบูลีน (Boolean operators) คือ AND OR และ NOT

“AND” เป็นตัวดำเนินการที่ใช้ค้นหาเพจที่มีแนวคิดหรือคำค้นสองคำอยู่ด้วยกันเท่านั้น เช่น ต้องการค้นเรื่องการใช้อินเทอร์เน็ตในประเทศไทย คำค้น คือ Internet และ Thailand ในกรณีนี้จะใช้ตัวดำเนินการ “AND” เพื่อให้ได้เพจที่กล่าวถึงเรื่องนี้เท่านั้น

ตัวดำเนินการ “AND” จะมีในโปรแกรมค้นหาแทบทุกโปรแกรม ยกเว้น All The Web Google LookSmart และ Yahoo เนื่องจากถือเป็นตัวดำเนินการโดยปริยาย ในโปรแกรมแล้ว



“OR” เป็นตัวดำเนินการที่ใช้รวมแนวคิดการสืบค้นสองแนวคิด หรือคำค้นสองคำเข้าด้วยกัน เพื่อให้ได้ผลออกมารวมกัน เช่น ต้องการเพจที่เกี่ยวกับมลภาวะทางเสียง และมลภาวะทางอากาศ จะใช้ข้อความว่า noise pollution OR air pollution ผลลัพธ์ที่ได้ก็จะมีเรื่องราวเกี่ยวกับมลภาวะทั้ง 2 ประเภท

ตัวดำเนินการ “OR” จะมีในโปรแกรมค้นหาแทบทุกโปรแกรม ยกเว้น LookSmart และ Yahoo โดยเป็นตัวดำเนินการโดยปริยายในโปรแกรกดังกล่าว

“NOT” เป็นตัวดำเนินการที่ใช้สำหรับตัดแนวคิดที่ไม่เกี่ยวข้องออกไป โดยเลือกเฉพาะเพจที่มีแนวคิดเดียว เช่น noise pollution NOT air pollution หมายความว่าต้องการเฉพาะมลภาวะทางเสียงเท่านั้น ไม่ต้องการมลภาวะทางเสียงและทางอากาศ หรือมลภาวะทางอากาศแต่อย่างเดียว ผลลัพธ์จะออกมาเฉพาะเพจที่มีเรื่องมลภาวะทางเสียงเท่านั้น

ตัวดำเนินการ “NOT” จะไม่มีให้ใช้ใน Google LookSmart และ Yahoo แต่จะใช้เป็น “AND NOT” ใน Alta Vista AOL NetFind Inktomi(MSN Search) Netscape และ Snap

ในการสืบค้นที่ใช้ตัวดำเนินการบูลีน วิธีการค้นที่โปรแกรมค้นหา ได้พยายามทำให้เหมือนกับการค้นออนไลน์ทั่วไป ก็คือวิธีการ “Nesting” หรือการใช้วงเล็บ เช่น

preservation AND (books or tapes)

ข้อความนี้ผู้ค้นต้องการเรื่องเกี่ยวกับการสงวนรักษาทั้งหนังสือและเทป

สำหรับวิธีการนี้ใช้ได้กับเกือบทุกโปรแกรม ยกเว้น All The Web Google LookSmart และ Yahoo

นอกจากนั้นยังมีการค้นแบบระบุตำแหน่ง (proximity searching) โดยทั่วไปจะใช้คำสั่ง NEAR ที่จะระบุว่าคำค้นจะอยู่ห่างกันเท่าไร เช่น yellow NEAR river

การระบุตำแหน่งจะแตกต่างกันไปตามโปรแกรมค้นหา เช่น Alta Vista จะกำหนดว่าห่างกันภายใน 10 คำ Lycos ห่างกัน 25 คำ ใน AOL Search ผู้ใช้สามารถกำหนดใหม่ได้ว่าจะให้ห่างกันกี่คำ เช่น yellow NEAR/5 river

การค้นโดยใช้คำสั่งเชิงคณิตศาสตร์ (Math Command)

ผู้ใช้อาจใช้คำสั่งเชิงคณิตศาสตร์ (math command) โดยใช้เครื่องหมายบวก (+) และเครื่องหมายลบ (-)

เครื่องหมาย + ใช้เมื่อต้องการรวมคำค้นให้ได้ผลลัพธ์ตามต้องการ เช่น

+ chuan + taksin

+ education + reform

คำสั่งที่ใช้เครื่องหมายบวก (+) จะมีในทุกโปรแกรมยกเว้น LookSmart

เครื่องหมายลบ (-) ใช้เมื่อต้องการเฉพาะเรื่องแรกที่ไม่มีข้อมูลของเรื่องก็ตามหลังเครื่องหมายมาปน เช่น

pollution – water

หมายความว่า ต้องการเรื่องเกี่ยวกับมลภาวะ ที่ไม่เกี่ยวข้องกับน้ำ

เครื่องหมายลบ (-) จะมีในทุกโปรแกรมยกเว้น LookSmart เช่นเดียวกัน

นอกจากนั้นผู้ใช้อย่างสามารถใช้เครื่องหมาย “ ” ในการทำการค้นวลี ในบางครั้งผู้ใช้อาจเห็นว่าการใช้เครื่องหมายบวกร ก็เท่ากับเป็นการเพิ่มคำค้นหรือแนวคิดอยู่แล้ว แต่ในการค้นคำค้นจะไม่อยู่ติดกัน เช่น ตัวอย่างที่ใช้คำค้น education และ reform อาจได้เพจที่พูดเรื่องการศึกษา (education) และมีเรื่องการปฏิรูป (reform) ด้านอื่น ๆ ได้ แต่ถ้าใช้ “education reform” ผลลัพธ์ก็จะเป็นเพจที่มีเรื่องราวการปฏิรูปการศึกษาอย่างแน่นอน

การสืบค้นแบบนี้เรียกอีกอย่างหนึ่งว่าการค้นโดยวลี (phrase searching) ซึ่งมีในทุกโปรแกรมยกเว้น LookSmart Direct Hit ส่วนโปรแกรม เช่น Google มีลักษณะที่เป็นกึ่งอัตโนมัติเมื่อผู้ใช้ใส่คำว่า education reform ถ้าทำการค้นและพบทั้งวลีเช่นนี้ โปรแกรมจะระบุทันทีว่าเป็น Phrase Match ซึ่งหมายความว่าในเพจมีวลีนี้ หรือมีตัวเชื่อมโยงไปยังเพจที่มีวลีนี้

การใช้คำสั่งสืบค้นอื่น ๆ

ในการค้นหาสารนิเทศบนเว็บ โดยใช้โปรแกรมค้นหา อาจใช้คำสั่งสืบค้นที่เฉพาะเจาะจง อาทิ

การสืบค้นโดยชื่อเรื่อง การสืบค้นโดยชื่อเรื่องในที่นี้คือ ชื่อเรื่องบนเว็บเพจ เมื่อเวลาค้นผู้ใส่จะระบุข้อความสืบค้นดังนี้

title : Use your search engine

สำหรับโปรแกรมค้นหาบางโปรแกรมจะใช้เมนู เช่น Lycos ในขณะที่ Yahoo ผู้ค้นจะใช้ t : meta engine ในการค้นหาเว็บเพจที่มีชื่อเรื่องตามต้องการ โปรแกรมค้นหา เช่น Excite LookSmart Netscape HotBot จะไม่มีวิธีการค้นนี้

การสืบค้นโดยเว็บไซต์

ในกรณีที่ต้องการค้นเว็บเพจทุกหน้าที่อยู่ในเว็บไซต์ก็สามารถใช้คำสั่งดังนี้

host : searchenginewatch.com สำหรับโปรแกรมค้นหา เช่น Alta Vista เป็นต้น

domain : searchenginewatch.com สำหรับโปรแกรมค้นหา เช่น Inktomi HotBot เป็นต้น

site : searchenginewatch.com สำหรับโปรแกรมค้น เช่น Google เป็นต้น

คำสั่งสืบค้นนี้ไม่มีในโปรแกรมค้นหา เช่น AOL Search HotBot Yahoo เป็นต้น



ในบางกรณีผู้ใช้อาจต้องการค้นเรื่องเฉพาะที่มีในไซต์นี้ เช่น “political reform” + host : edu หมายความว่าเลือกเฉพาะไซต์ประเภทที่เป็นสถาบันการศึกษาเท่านั้น

การสืบค้นโดย URL

การสืบค้นวิธีนี้จะมีลักษณะคล้ายกับการสืบค้นโดยไซต์ รูปแบบก็คือ

url : chula.ac.th ใช้ในโปรแกรมค้นหา เช่น Alta Vista Northern Light เป็นต้น

url.all : chula.ac.th ใช้ในโปรแกรมค้นหาเช่น All the Web Lycos เป็นต้น

allinurl: chula.ac.th ใน โปรแกรมค้นหา เช่นGoogle เป็นต้น

originurl: chula.ac.th สำหรับ โปรแกรมค้นหา เช่นInktomi เป็นต้น

u : chula.ac.th สำหรับโปรแกรมค้นหา เช่น Yahoo เป็นต้น

การสืบค้นวิธีนี้แสดงให้เห็นว่าเว็บเพจใดในฐานข้อมูลของโปรแกรมค้นหา มีการเชื่อมโยงกับ URL ใดบ้าง ทำให้ผู้ใช้สามารถระบุเว็บไซต์ที่เกี่ยวข้องได้

การสืบค้นโดยตัวเชื่อมโยง (Link Search)

โปรแกรมค้นหาบางโปรแกรมมีความสามารถที่จะค้นทุกเพจที่เชื่อมโยงกับเพจหรือโดเมนใดโดยเฉพาะ คำสั่งคือ

link : dlib.org ใช้ในโปรแกรมค้นหา เช่น Alta Vista Google เป็นต้น

link domain : pcworld.com ใช้ในโปรแกรมค้นหา เช่น HotBot เป็นต้น

การสืบค้นแบบตัดคำ (wildcard & truncation)

ผู้ใช้สามารถใช้เครื่องหมายดอกจัน (*) ในการทำการสืบค้นแบบนี้ เช่น

sing* จะถือว่าค้นคำว่า singing และ sings ด้วย นอกจากนี้ผู้ค้นสามารถใส่เครื่องหมาย ? ใน AOL Search ได้

วิธีการนี้ใช้ได้กับโปรแกรมค้นหา เช่น Alta Vista HotBot Yahoo เป็นต้น ส่วนบางโปรแกรม เช่น Google Lycos All the Web จะไม่มีวิธีการนี้

การสืบค้นโดยใช้รากคำ (stemming)

โปรแกรมค้นหาบางโปรแกรมจะมีเทคนิคการค้นโดยตัดคำทำไว้โดยอัตโนมัติ เช่น Lycos โปรแกรมบางโปรแกรมจะจัดให้มีเทคนิคที่จะค้นตามความหลากหลายของไวยากรณ์ได้ เช่น เมื่อค้นด้วยคำว่า “sought” จะได้ seek และ seeking ด้วย

ในการค้นผู้ใช้อาจต้องพิจารณาลักษณะเฉพาะ เช่น การใช้ตัวใหญ่ ตัวเล็ก ในภาษาอังกฤษ โปรแกรมค้นหาบางโปรแกรมสามารถระบุความแตกต่างของการใช้ตัวใหญ่ ตัวเล็ก ได้ ซึ่งจะสามารถช่วยในการค้นศัพท์ที่มีความหมายต่างกัน หากใช้ตัวใหญ่และตัวเล็ก เช่น คำว่า AIDS และ aids โดยทั่วไปหากผู้ใช้พิมพ์เป็นตัวเล็กทั้งหมด โปรแกรมค้นหาก็ค้นทั้งตัวใหญ่และตัวเล็ก ในขณะที่หากใช้ตัวใหญ่หมดก็จะให้ผลลัพธ์เฉพาะที่ตรงตามที่พิมพ์เข้าไป เช่น

house จะค้นทั้ง HOUSE House และ house

HOUSE จะค้น HOUSE เท่านั้น

หรือ House จะค้น House เท่านั้น

การค้นหาสารนิเทศบนเว็บด้วยเทคนิควิธีการต่าง ๆ จะช่วยให้ผู้ค้นสามารถปรับแต่งแก้ไข เพื่อให้ได้สารนิเทศตรงตามความต้องการที่สุด การใช้ตัวดำเนินการบูลีน จะช่วยจำกัดการสืบค้น ให้ได้ผลตามความต้องการ และสามารถใช้ร่วมกันเพื่อการค้นที่ค่อนข้างซับซ้อน หรือในบางกรณี ผู้ใช้อาจใช้การค้นด้วยวลี อาจเหมาะสมกว่าการใช้คำค้นเพียงคำเดียวที่อาจทำให้ได้ผลมากมายจนเกินกว่าที่จะตรวจตราหรือคัดเลือกได้ ในกรณีที่ต้องการผลลัพธ์ที่ครอบคลุมกว้างขวาง ขึ้นก็ให้ใช้วิธีการค้นแบบ wildcard โดยการใส่ดอกจันไว้ท้ายคำ เช่น

librar* จะได้ข้อมูลเกี่ยวกับ libraries library librarian librarians มาด้วย เป็นต้น

โปรแกรมค้นหาจะมีแบบฟอร์มที่ช่วยให้มีการปรับแต่งการค้นหา โดยใช้เทคนิควิธีการ ขั้นสูง ซึ่งจะทำให้ผู้ใช้สามารถทำการค้นโดยไม่ต้องใช้ภาษาสืบค้นขั้นสูง

ผลการค้น (search results)

โดยทั่วไปเมื่อทำการค้นหาด้วยโปรแกรมค้นหา ผลที่ได้รับมักจะเป็นผลที่จัดลำดับการนำเสนอตามความเกี่ยวข้อง หมายความว่า ผลลัพธ์ในลำดับต้นก็จะมีเกี่ยวข้องหรือให้คำตอบตรงตามที่ต้องการมากกว่าผลลัพธ์ในลำดับท้าย วิธีการทำงานของโปรแกรมค้นหาที่จัดลำดับผล ส่วนมากจะเป็นการดูว่าคำค้นนั้นจะมีปรากฏบ่อยมากน้อยเพียงไรในเว็บไซต์นั้น เช่น ค้นคำว่า tourism เว็บไซต์ที่มีคำนี้ 25 คำ ก็อยู่ในลำดับต้นกว่าเว็บไซต์ที่มีเพียง 10 คำ

นอกจากนั้นการเสนอผลที่ค้นได้ตามลำดับนั้นอาจพิจารณาจำนวนความถี่ของคำค้นที่ปรากฏในฐานข้อมูลต่าง ๆ รวมทั้งการที่คำค้นอยู่ที่ไหนในเอกสารบนเว็บ เช่น ถ้าอยู่ในชื่อเรื่องหรือ Meta Description Tag ก็จะมีน้ำหนักมากกว่าคำที่ปรากฏในส่วนอื่นของเอกสาร เป็นต้น หรือในกรณีคำค้นนั้นอยู่ติดกันหรือเรียงตามลำดับกันอย่างถูกต้อง

เมื่อได้ผลการค้นแล้ว โปรแกรมค้นจะแสดงผลทีละ 10 รายการ เช่น Alta Vista หรือ All The Web ในขณะที่บางโปรแกรม เช่น MSN Search อาจเสนอทีละ 15 รายการ Yahoo ทีละ 20 รายการได้เช่นกัน

รูปแบบการแสดงผล ซึ่งอาจจะมีแบบสั้นหรือแบบยาว ก็มักจะประกอบด้วยข้อมูลต่อไปนี้ (Schwartz 1998 : 976)

1. ชื่อเรื่อง (ชื่อเว็บเพจ)
2. คะแนน/ความถี่ของความตรงต่อความต้องการ
3. สรุปรูป ซึ่งอาจอยู่ในรูปสาระสังเขป หรือเป็นโครงเรื่อง ที่ดึงมาจากข้อความที่อยู่ตอนต้นหรือคำที่มี



4. ขนาดของแฟ้มข้อมูล
5. วันเดือนปีของแฟ้มข้อมูล
6. วันเดือนปีที่นำข้อมูลเข้าในฐานข้อมูล
7. URL
8. ภาษา
9. ข้อความแสดงประเภท (ถ้าในโปรแกรมค้นหามีการจำแนกประเภทข้อมูล)
10. คำค้นที่ปรากฏ

โปรแกรมค้นหามักจะมีการเสนอทางเลือกในการแสดงผล เช่น Alta Vista จะมีทางเลือกแบบปกติ แบบสั้นและแบบที่เป็นข้อความเท่านั้น หรือ HotBot จะให้ทางเลือกการแสดงผลแบบเต็ม แบบย่อ และแบบที่แสดงเฉพาะ URL เท่านั้น

การแสดงผลในลักษณะอื่น ๆ แต่ละโปรแกรมค้นหาเสนอให้กับผู้ใช้มีอาทิ การจัดเรียงตามลำดับวันเดือนปี จะมีโปรแกรม เช่น Northern Light แต่การจัดเรียงแบบนี้มีปัญหาตรงที่ในเว็บไซต์บางแห่งจะระบุวันเดือนปีผิด หรือบางที่ไม่ระบุเลย

ในการค้นบางครั้งจะได้รับผลการค้นมาจากเว็บไซต์เดียว โปรแกรมค้นหาบางโปรแกรมจะมีคำสั่งให้ดึงผลลัพธ์เฉพาะ 1 หน้า ต่อ 1 เว็บไซต์ สำหรับผลการค้นที่อยู่ในลำดับต้น ผู้ค้นจะได้เว็บไซต์ที่หลากหลาย

ในการแสดงผลของแต่ละโปรแกรมค้นหาอาจมีรายละเอียดบางประการที่ต่างกัน เช่น ขนาดความยาวของชื่อเรื่อง เป็นการกำหนดจำนวนอักขระ (characters) ที่จะแสดงผลจากชื่อเรื่องซึ่งมักประมาณ 70-80 ตัว หรือในกรณีที่ไม่มีชื่อเรื่องโปรแกรมค้นหาบางโปรแกรมจะใช้คำว่าไม่มีชื่อเรื่อง (No Title, Untitled) บางโปรแกรมจะบอก URL หรือบางโปรแกรมก็จะใช้ข้อความในบรรทัดแรกบนเว็บเพจ

โปรแกรมค้นหาต่างก็พยายามในการนำเสนอผลที่ตรงตามความต้องการของผู้ใช้และอยู่ในรูปแบบที่จะเป็นประโยชน์ การให้ข้อมูลสรุปเกี่ยวกับเพจ ก็อาจมีส่วนช่วยให้ผู้ใช้สามารถถกค้นกรองผลลัพธ์ได้ แต่ขีดความสามารถของโปรแกรมในการทำสรุป (summary) หรือสาระสังเขป (abstract) อัปเดตโนมันั้นแตกต่างกัน และสิ่งที่ยังดูเป็นข้อด้อยก็คือ การจัดลำดับความสำคัญของผลลัพธ์ รวมทั้งการที่ยังมีปัญหาในเรื่องการแสดงผลที่ซ้ำซ้อน และบางครั้งไม่สามารถนำไปใช้ได้

การประเมิน

ปัจจุบันผู้ใช้เว็บจะพบการปริทัศน์ (review) โปรแกรมค้นหาอยู่มากมายทั้งที่อยู่ในรูปบทความวารสารที่ดีพิมพ์ เช่น ในวารสาร PC Magazine หรือ Internet World รวมทั้งการประเมินบนเว็บเอง งานวิจัยที่เกี่ยวข้องกับโปรแกรมค้นหา ก็มักจะทำการประเมินการค้นโดยใช้โปรแกรมค้นหาซึ่งมักจะใช้วิธีการเช่นเดียวกับการประเมินการค้นคืนสารสนเทศ

ลักษณะการประเมินประสิทธิภาพการปฏิบัติงานของโปรแกรมค้นหาที่ทำโดยหน่วยงานทางธุรกิจ หรืออุตสาหกรรมคอมพิวเตอร์ มักจะเป็นการเปรียบเทียบโปรแกรมค้นหาในด้านความเร็ว ความง่ายในการใช้ การออกแบบ การเชื่อมประสานกับผู้ใช้ ลักษณะวิธีการค้นรวมทั้งขีดความสามารถในการค้น การแสดงผลการค้น

ในการทดลองค้น การประเมินมักจะมุ่งเน้นอยู่ที่ความถูกต้องตรงตามความต้องการของผลการค้น เช่น การเสนอผลการทดลองของ Haskin (1998) ซึ่งทดสอบโปรแกรมค้นหา 6 โปรแกรม คือ Alta Vista Excite HotBot InfoSeek Lycos Web Crawler โดยพิจารณาจากขีดความสามารถในการค้นประเภทต่าง ๆ คือ การค้นโดยง่าย (simple search) และการค้นขั้นสูง (advanced search) การเชื่อมประสานกับผู้ใช้ว่ายุ่งยากเพียงไร การแสดงผลบนจอ รูปแบบการแสดงผล

Searchenginewatch (2002) ได้แสดงการจัดลำดับและผลการปฏิบัติงานของโปรแกรมค้นหา มีการนำเสนอผลการจัดลำดับของหน่วยงาน เช่น SatMarket ที่ให้ข้อมูลเมื่อเดือนเมษายน 2002 ว่า Google เป็นไซต์สืบค้นลำดับที่ 1

การศึกษาเกี่ยวกับโปรแกรมค้นหาในขณะนี้จะให้ความสนใจอย่างมากกับการประเมิน โดยเฉพาะในเรื่องประสิทธิภาพการค้น Gordon and Pathak (1999 : 146-147) เสนอลักษณะ 7 ประการที่จะทำในการประเมินเป็นไปอย่างถูกต้องแม่นยำและให้เนื้อหาสาระ และมีเป้าหมายที่จะวัดโปรแกรม ค้นหาเหล่านี้โดยการค้นจริง ลักษณะดังกล่าวได้แก่

1. การค้นควรจะเกิดมาจากความต้องการสารสนเทศจริงของผู้ใช้
2. ถ้าทำการทดลองค้นในหัวข้อที่ผู้อื่นระบุไว้ ก็ต้องพยายามค้นให้ครอบคลุมให้มากที่สุดตามบริบทของความต้องการของผู้ใช้นั้น
3. ควรทำการค้นให้มากเพียงพอที่จะสามารถนำมาประเมินประสิทธิภาพของโปรแกรมค้นหาได้
4. ควรทดลองกับโปรแกรมค้นหาที่สำคัญ ๆ ให้มากที่สุด
5. การวิเคราะห์ประสิทธิภาพของโปรแกรมค้นหา แต่ละโปรแกรมควรจะใช้ลักษณะพิเศษของแต่ละโปรแกรม ซึ่งหมายความว่าไม่จำเป็นต้องใช้กลยุทธ์การสืบค้นเดียวกันในแต่ละโปรแกรม มิฉะนั้นลักษณะพิเศษที่ดีที่สุดอาจไม่ได้นำมาใช้



6. ผู้ที่ต้องการสารสนเทศจะต้องตัดสินใจความต้องการของผลการค้น

7. การทดลองที่ดี ซึ่งหมายความว่าผลการทดลองที่ทำตามการออกแบบอย่างดี มีการใช้เกณฑ์ที่เป็นที่ยอมรับในการค้นคืนสารสนเทศ และใช้เทคนิคทางสถิติเป็นเรื่องจำเป็น

ในการประเมินโปรแกรมค้นหา อาจไม่มีการดำเนินการตามลักษณะที่ควรจะเป็นครบทั้ง 7 ประการ แต่สิ่งที่การประเมินจะใช้ก็คือ การตัดสินใจความต้องการของผลการค้น

โดยทั่วไปโปรแกรมค้นหาจะเสนอผลการค้นโดยการเรียงลำดับตามความต้องการ เกณฑ์ในการจัดลำดับความต้องการมักจะเป็นจำนวนคำที่พบ ตำแหน่งของคำในเอกสาร และความถี่ของคำที่ปรากฏในเอกสารและในทั้งฐานข้อมูล รวมทั้งความยาวของเอกสาร กลยุทธ์การจัดลำดับผลการค้นมักจะแตกต่างกันไปตามโปรแกรมค้นหา และมักถือเป็นความลับด้วย บางโปรแกรมเน้นความถี่ของคำที่ปรากฏ บางโปรแกรมพิจารณาจำนวนคำกับความยาวของเอกสาร โดยที่เอกสารที่สั้นกว่าจะได้รับการจัดลำดับที่ดีกว่าเอกสารยาว หากมีจำนวนคำค้นที่พบในเอกสารเท่ากัน คำที่พบในชื่อเรื่องและ metatag จะมีน้ำหนักมากกว่าคำที่พบในเนื้อหา

การจัดลำดับความสำคัญของการเสนอผล มีความสำคัญกับความพึงพอใจของผู้ใช้โปรแกรมค้นหา รวมทั้งความสำเร็จของการสืบค้น ส่วนมากแล้วผู้ใช้จะตัดสินใจถูกต้องตรงตามความต้องการจากจำนวนคำค้นที่ปรากฏในเอกสาร ดังนั้นเมื่อผู้ใช้อ่านผลการค้นและพบคำค้นของตน ก็มักจะเห็นว่าเอกสารนั้น ๆ ตรงตามความต้องการ

ในการทำการประเมิน ผู้ประเมินมักจะใช้ข้อความสืบค้นจำนวนหนึ่งแล้วทดลองค้นจริง Chu และ Rosenthal (1996) ใช้คำถามที่ผู้ใช้ถามบรรณารักษ์บริการตอบคำถามที่ Long Island University จำนวน 10 คำถาม และเน้นเทคนิคการวิจัยการค้นคืนสารสนเทศในการประเมิน ได้แก่ เวลาสนองตอบ (response time) ซึ่งหมายถึงเวลาที่ใช้ในการรอผลการค้น ตั้งแต่เริ่มใช้คำสั่งสืบค้นจนถึงผลการค้นชุดแรกปรากฏบนจอ และอัตราความถูกต้องตรงตามความต้องการ (precision) โดยเลือกอ่านระเบียบเว็บ (web records) ที่ดาวน์โหลดจำนวนไม่เกิน 10 ระเบียบ ในแต่ละข้อความสืบค้น และคำนวณอีกว่าความถูกต้องตรงตามความต้องการโดยรวมของการสืบค้น 10 ครั้ง สำหรับ 10 คำถาม ในแต่ละโปรแกรมค้นหา ส่วน Ding และ Marchionini (1996) ใช้คำถามสืบค้น 5 คำถาม และให้คะแนนผลการสืบค้น คะแนนสูงสุด 5 คะแนน ให้กับการค้นได้เอกสารเนื้อหาเต็ม หรือข้อมูลที่ครอบคลุมเกี่ยวกับเรื่องที่ค้นให้ 4 คะแนน สำหรับบทความหรือเพจที่เกี่ยวข้องกับเรื่องให้ 3 คะแนน สำหรับเอกสารที่ชี้แนะไปยังเอกสารที่อยู่ในระดับ 5 หรือ 4 ให้ 2 คะแนน สำหรับโฮมเพจทั่ว ๆ ไปให้ 1 คะแนน สำหรับเอกสารที่มีการกล่าวถึงเรื่องที่ค้นบ้าง และ 0 คะแนน สำหรับเอกสารที่ไม่มีเนื้อหาที่ต้องการ จากนั้นก็นำคะแนนไปคำนวณอัตราความถูกต้องตรงตามความต้องการ

งานวิจัยของ Courtois และ Berry (1999) ใช้วิธีการทดสอบโดยใช้ข้อความสืบค้น 12 ข้อความ เพื่อดูความสามารถของโปรแกรมค้นหาในการจัดลำดับผล และดาวน์โหลดรายการที่พบ 100 รายการ แล้วพิจารณาว่าเอกสารนั้นมีค่าทุกคำในข้อความสืบค้นมีวลีที่เป็นข้อความสืบค้นตามลำดับ เช่น latent semantic indexing อย่างน้อย 1 วลีหรือไม่ และคำค้นจะปรากฏในชื่อเรื่อง หัวเรื่อง หรือ metatag อย่างน้อย 1 ครั้งหรือไม่ ผลการทดลองพบว่า Excite ได้รับคะแนนดีที่สุดในการสืบค้นแบบทุกคำ ส่วน AltaVista จะปฏิบัติได้ดีในการค้นแบบระบุตำแหน่ง

Bar-Ilan (2002) ระบุว่าเกณฑ์การประเมินแบบเดิม ไม่ว่าจะเป็นเรื่องความครอบคลุม อัตราเรียกค้น อัตราความถูกต้องตรงตามความต้องการ เวลาสนองตอบ ความพยายามของผู้ใช้ และรูปแบบผลลัพธ์ ไม่เพียงพอในการวัดประสิทธิภาพการปฏิบัติงานของโปรแกรมค้นหาในสภาพแวดล้อมที่เป็นพลวัตเช่นปัจจุบัน จึงเสนอเกณฑ์ต่อไปนี้

- อัตราถูกต้องตรงตามความต้องการเชิงเทคนิค (technical precision) คือ ร้อยละของเอกสารที่ตรงตามความต้องการ คิดจากเอกสารที่ค้นได้ทั้งหมด
- อัตราความเกี่ยวข้องเชิงเทคนิค (technical relevancy) คือเอกสารที่ค้นได้ด้วยคำถามที่ตั้ง เช่น คำค้น หรือ วลีที่ใช้ค้นจะต้องปรากฏในเอกสาร
- ความครอบคลุมต่อรอบการค้น (relative coverage per search round) เป็นจำนวน URL ที่ตรง/เกี่ยวข้องที่โปรแกรมค้นได้หารด้วยจำนวน URL ที่ค้นได้ในรอบนั้น
- จำนวน URL ที่ใหม่ ที่พบในการค้นในโปรแกรมใดโปรแกรมหนึ่ง โดยยังไม่เคยพบในรอบการค้นก่อนหน้า ส่วน จำนวน URL ใหม่โดยรวม คือ รายการที่ไม่เคยค้นได้มาก่อนในทุก ๆ โปรแกรมที่ค้น

นอกจากนี้ยังมีการพิจารณา จำนวน URL ที่เกี่ยวข้อง แต่โปรแกรมค้นหาทำตกหล่น หรือ สูญหายไป โดยอาจถูกค้นพบในรอบการค้นต่อไป

ผู้วิจัยใช้การเลือกหัวข้อสืบค้นที่พิจารณาแล้วว่ามีความคงที่ (stable) โดยโปรแกรมค้นหาแต่ละตัวจะตีความเหมือนกัน เลือกคำคำเดียว ที่ไม่มี รากคำ หรือ คำขยาย และ มีความเป็น “นานาชาติ” เพื่อไม่ให้จำกัดเฉพาะเพจที่เป็นภาษาอังกฤษ ค้นจากโปรแกรมค้นหา 6 โปรแกรม คือ Alta Vista Excite Fast (All The Web) HotBot Google และ Northern Light

การประเมินโปรแกรมค้นหาเป็นความพยายามในการที่จะช่วยในการปรับปรุงวิธีการสืบค้น และการเสนอผลของโปรแกรมค้นหา เช่น ในขณะนี้โปรแกรมค้นหา ก็เพิ่มวิธีการจัดอันดับผลการสืบค้นโดยพิจารณาจากจำนวนการเชื่อมโยงที่มีในแต่ละเว็บเพจ อย่างไรก็ตาม สิ่งที่ต้องคำนึงถึงนอกเหนือไปจากการประเมินเพื่อช่วยปรับปรุงโปรแกรมค้นหา ก็คือ การให้ความสนใจกับ



ความพยายามช่วยเหลือผู้ค้นหาสารนิเทศบนเครือข่ายด้วยวิธีอื่น ๆ ไม่ว่าจะเป็นการมีแหล่งสารนิเทศที่ให้บริการเพิ่มคุณค่าในเรื่องของการเข้าถึงสารนิเทศบนอินเทอร์เน็ต หรือการมีเครื่องมือที่ช่วยให้ผู้ใช้สามารถค้นหา (browse) สารนิเทศบนอินเทอร์เน็ตได้ หรือการมีคอลเล็กชันที่ช่วยในการค้นหาสารนิเทศ เช่น โครงการ Access to Network Resource ซึ่งเป็นส่วนหนึ่งของโครงการ eLib (Electronic Libraries Programme) ของสหราชอาณาจักรที่พัฒนาเกตเวย์ สำหรับสาขาต่าง ๆ แต่ละเกตเวย์จะมีแหล่งสารนิเทศบนเครือข่ายที่มีคุณภาพสูงที่ได้เลือกสรรมาแล้วเป็นอย่างดี และมีระเบียบเมตาเดตา (metadata) ที่ระบุข้อมูลที่เกี่ยวข้องถึงคำสำคัญ ศัพท์สัมพันธ์ การจัดหมู่ต่าง ๆ เพื่อช่วยการค้นหาและการค้นหาผ่านเกตเวย์ส่วนใหญ่จะใช้ซอฟต์แวร์และมาตรฐานเดียวกัน ทำให้ผู้ใช้ได้เห็นสิ่งที่เหมือน ๆ กัน และได้เครื่องมือในการค้นหาสารนิเทศที่มีคุณภาพเช่นเดียวกัน

การใช้โปรแกรมค้นหาเชิงพาณิชย์ ก็ยังคงเป็นวิธีการสำคัญในการค้นหาสารนิเทศที่มีมากมายมหาศาลบนเว็บ แต่หากโปรแกรมค้นหาใดที่อ้างว่าได้มีการทำดัชนีทุกสิ่งทุกอย่างบนเว็บแล้วนั้นก็คงเป็นสิ่งที่ไม่เป็นความจริง โปรแกรมค้นหาแต่ละโปรแกรมมีจุดเด่นที่อาจแตกต่างกันออกไป ถึงแม้จะมีขีดความสามารถสูงในการค้นหาสารนิเทศ เพื่อตอบสนองผู้ใช้ก็ตาม ผู้ใช้ก็ควรที่จะได้ติดตามการปริทัศน์และการประเมินโปรแกรมค้นหาต่าง ๆ เพื่อช่วยประกอบการตัดสินใจในการสืบค้นสารนิเทศ พัฒนาการของบริการค้นหาสารนิเทศบนเว็บคงดำเนินไปอย่างไม่หยุดยั้ง และคงจะมีบทบาทสำคัญในการใช้ทรัพยากรบนเครือข่ายเพิ่มมากขึ้นเช่นเดียวกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

- Barr-Ilan, Judit. "Methods for Measuring Search Engine Performance Over Time."
Journal of the American Society for Information Science and Technology. 53,4
(February 2002): 308-319
- Chu, Heting and Rosenthal, Marilyn. " Search Engines for the World Wide Web: A
Comparative Study and Evaluation Methodology." In Steve Hardis, ed. **Global
Complexity: Information, Chaos, and Control: Proceedings of the 59th ASIS
Annual Meeting** vol.33, p. 127-135. Medford : Information Today, 1996. Also
[Online] Available: [http://www.asis.org/annual-96/Electronic Proceedings/chu.html](http://www.asis.org/annual-96/Electronic%20Proceedings/chu.html)
- Curtois, Martin P.; and Berry, Michael W. " Results Ranking in Web Search Engines"
Online 3,3 (May/June 1999): 39-46.
- Ding, Wei; and Marchionini, G. "A Comparative Study of Web Search Service
Performance. Steve Hardis, ed. **Global Complexity: Information, Chaos, and
Control : Proceedings of the 59th ASIS Annual Meeting** vol.33, p. 136-142,
Medford : Information Today, 1996. Also [Online] Available:
[http://www.asis.org/annual-96/ Electronic Proceedings/ding.html](http://www.asis.org/annual-96/Electronic%20Proceedings/ding.html)
- Gordon, Michael; and Pathak, Praveen. " Finding Information on the World Wide Web:
the Retrieval Effectiveness of Search engines." **Information Processing and
management** 35 (1999) : 141-180.
- Haskin, David. " The Right Search Engine : IW Labs Test." [Online]. Available : [http://
www.iw.com/ 1997/09/report.html](http://www.iw.com/1997/09/report.html),1998.
- Schwartz, Candy. " Web Search Engines." **Journal of the American Society for
Information Science**. 49,11 (1998): 973-982.
- Searchenginewatch. "Ratings, Reviews and Tests." [Online]. Available:
<http://www.Searchenginewatch.com/reports/index.html>, 2002.

