บทที่ 7

ระบบค้นคืนสารสนเทศบน WWW

วัตถุประสงค์

1. เพื่อให้นักศึกษาทราบถึง Traversing The Web
2. เพื่อให้นักศึกษาทราบถึง การทำงานของ Search Engine
3. เพื่อให้นักศึกษาทราบถึง Taxonomy For Search Tools And Service
4. เพื่อให้นักศึกษาทราบถึง วิธีการค้นหาข้อมูลด้วยsearch engine ที่ทำให้ค้นหาข้อมูลให้ง่ายขึ้น
5. เพื่อให้นักศึกษาทราบถึง Retrieval Effectiveness Assessment
6. เพื่อให้นักศึกษาทราบถึง การจัดลำดับของเพจ(PageRank)

สารบัญ

7.1 บทนำ

World Wide Web เกิดขึ้นในปี 1989 โดย Tim Berners-Lee แห่งห้องปฏิบัติการ CERN(Conseil European pour la Recherche Nucleaire) ประเทศสวิตเซอร์แลนด์ ซึ่ง ทิม เบอร์เนอร์ส ลี แห่งศูนย์วิจัย CERN ได้คิดค้นระบบไฮเปอร์เท็กซ์ขึ้นเว็บเบราวเซอร์ตัวแรกมีชื่อว่า World Wide Web แต่เว็บได้รับความนิยมอย่างจริงจังเมื่อศูนย์วิจัย NCSA ของมหาวิทยาลัยอิลลินอยส์เออร์ ฯ สหรัฐอเมริกา ได้คิดโปรแกรม MOSAIC (โมเสค) ซึ่งเป็นเว็บเบราว์เซอร์ระบบกราฟิก Web ได้มีการเติบโตอย่างรวดเร็วและมีการเชื่อมต่อแหล่งที่เป็นข้อมูลข่าวสาร เช่น Personal home pages ,Online digital Libraries , virtural Museums , Product and Service catalog , Government Information for public dissmension , research publication , Gopher , mail server , Usenet New , FTP และ ฯลฯ

ซึ่งจากการประมาณการในเวลานี้จะพบว่าเว็บมีจำนวนเป็นหลายร้อยล้าน page และจะทวีจำนวนเพิ่มขึ้นสองเท่าตัวทุก ๆ 4 เดือน ความสามารถในการค้นหาและค้นคืนข้อมูลข่าวสารจากเว็บให้เป็นไปอย่างมีประสิทธิภาพซึ่งจะต้องตะหนักถึงเทคโนโลยีที่ใช้เพื่อให้มีประสิทธิภาพที่เต็มกำลังและมีความเป็นไปได้โดยควรจะใช้เครื่อง Work station และระบบเทคโนโลยี Parallel Processing ที่มีประสิทธิภาพไม่ก่อให้เกิดปัญหาคอขวด (Bottlenneck)

Rank	Query Term
1	sex
2	(artifact)
3	(artifact)
4	porno
5	mp3
6	Halloween
7	sexo
8	chat
9	porn
10	YAHOO
11	KaZaA
12	xxx
13	Hentai
14	lyrics
15	hotmail

Rank	Query Term
16	crack
17	games
18	pussy
19	cracks
20	lolita
21	britney spears
22	ebay
23	sexe
24	Pamela Anderson
25	warez
26	divx
27	gay
28	harry potter
29	playboy
30	lolitas

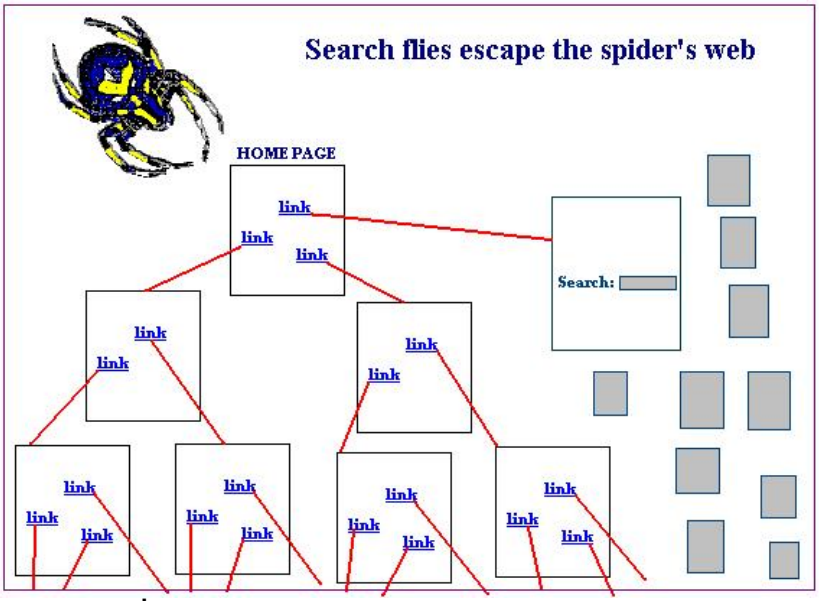
รูปที่ 7.1 : Queries on the Web (Year 2002)

รูปที่ 7.1 : Queries on the Web(Year 2002) แสดงถึง ลำดับของเทอมที่ผู้ใช้ร้องขอในปี ค.ศ.2002 มากที่สุดลดหลั่นตามลำดับ ระบบการค้นคืนในปัจจุบันมีประสิทธิภาพที่สำคัญแตกต่างกัน ปัจจุบันมี Search Tools ที่ใช้สำหรับการค้นคืนเอกสารเป็นจำนวนมากทั้งนี้จะขึ้นอยู่กับความสัมพันธ์ของข้อมูลตามที่ผู้ใช้ป้อนคำขอ (Query) และนอกจากนี้ความสัมพันธ์ของเอกสารที่ถูกค้นคืนมานั้นจะต้องเป็นเอกสารที่สำคัญและอยู่ใน TOP Rank ด้วย

สถาปัตยกรรมที่เกี่ยวกับโมเดลของการค้นคืนและวิธีการในการปฏิบัติการตาม Query ซึ่งจะต้องใช้ Search Tools ที่มีความเหมาะสมซึ่งการพัฒนา web Search Tools

7.2 Traversing The Web

สำหรับการค้นหาข้อมูลในเวปที่เต็มไปด้วยสารสนเทศจำนวนมากๆเหล่านี้ ดังนั้นจำเป็นต้องมีเครื่องมือช่วยในการสืบค้น จึงมีการพัฒนาสร้าง Search Engine เพื่อเป็นเครื่องมือสำคัญช่วยในการสืบค้นข้อมูล และจำเป็นอย่างยิ่งที่จะต้องพัฒนาประสิทธิภาพการทำงานของ Search Engine ให้สามารถสืบค้นเอกสารที่ผู้ใช้ต้องการในรูปของเว็บเพจได้อย่างครบถ้วนและถูกต้องมากที่สุด การสืบค้นข้อมูลแต่ละครั้ง Search Engine จะสืบค้นผ่านฐานข้อมูลที่ตนเองสร้างขึ้นที่รวบรวมจาก spiders ซึ่งจะทำการ Craw ไปตามลิงค์ต่างๆ เมื่อ Spider พบเว็ปเพจแล้วจะส่งต่อไปยังโปรแกรม indexing เพื่อกำหนดและแยกส่วนประกอบเพื่อทำเป็น คีย์เวิร์ด



รูปที่ 7.2 : แสดงการค้นหาแฟ้มจาก spider ของเวป

การทำงานของ Search Engine โดยทั่วไปประกอบด้วย 3 ส่วนหลัก คือ

1. โรบอต (Robots) หรือ สไปเดอร์ (Web Spider) หรือ ครอว์เลอร์ (Crawler) ทำหน้าที่ติดต่อไปยังเว็บไซต์ต่างๆ เพื่อสะสมไฟล์เอชทีเอ็มแอล (HTML) เก็บเป็นข้อมูลสำหรับสร้างดัชนีค้นหา ซึ่ง **Spidering Algorithm** สามารถกระทำได้ดังนี้

```
Initialize queue (Q) with initial set of known URL's
Until Q empty or page or time limit exhausted:
    Pop URL, L (Link), from front of Q
    If L is not to an HTML page (.gif, .jpeg, .ps, .pdf, .ppt…)
        continue loop
    If already visited L, continue loop
    Download page, P, for L
    If cannot download P (e.g. 404 error, robot excluded)
        continue loop
    Index P (e.g. add to inverted index or store cached copy).
    Parse P to obtain list of new links N
    Append N to the end of Q
```

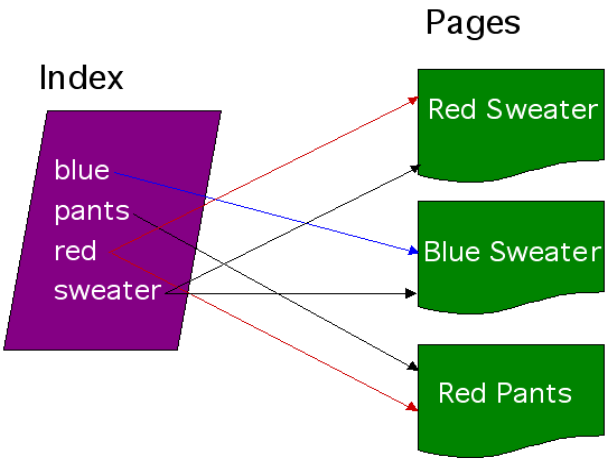
วิธีการทำงานนั้น web Robot จะทำการเข้าสำรวจเว็บไซต์ต่างๆ แล้วดึงข้อมูลเหล่านั้นมาอัพเดทใส่ในรายการฐานข้อมูล ส่วนมากมักจะเข้าไปอัพเดทข้อมูลเป็นรายเดือน โดยเริ่มต้นที่การตัดสินใจเลือก URL บนพื้นฐานของเว็บที่มีความ Popular มีการแบ่งแยก web space บนฐานข้อมูลของระบบ Internet name หรือ รหัสประเทศที่ได้

2. อินเด็กเซอร์ (Indexer) ทำหน้าที่สร้างดัชนีค้นหาจากไฟล์ HTMLที่โรบอตคัดลอกมาเอกสารเ HTML ใดๆ จะสามารถสืบค้นได้ก็ต่อเมื่อเอกสารนั้นผ่านการทำดัชนีแล้วเท่านั้น ดังรูปที่ 7.3 ซึ่งการทำ Indexing มีวัตถุประสงค์ 3 ข้อในการค้นคืนของสารสนเทศ

(1) ทำให้การหาที่อยู่ของเอกสารตามหัวเรื่องได้โดยง่าย

(2) นิยามสาขาของหัวเรื่องและความสัมพันธ์ของเอกสารอันหนึ่งกับอีกอันหนึ่ง

(3) ทำนายความเกี่ยวข้องของเอกสารที่กำหนดเข้ากับความต้องการของข่าวสารที่ระบุ

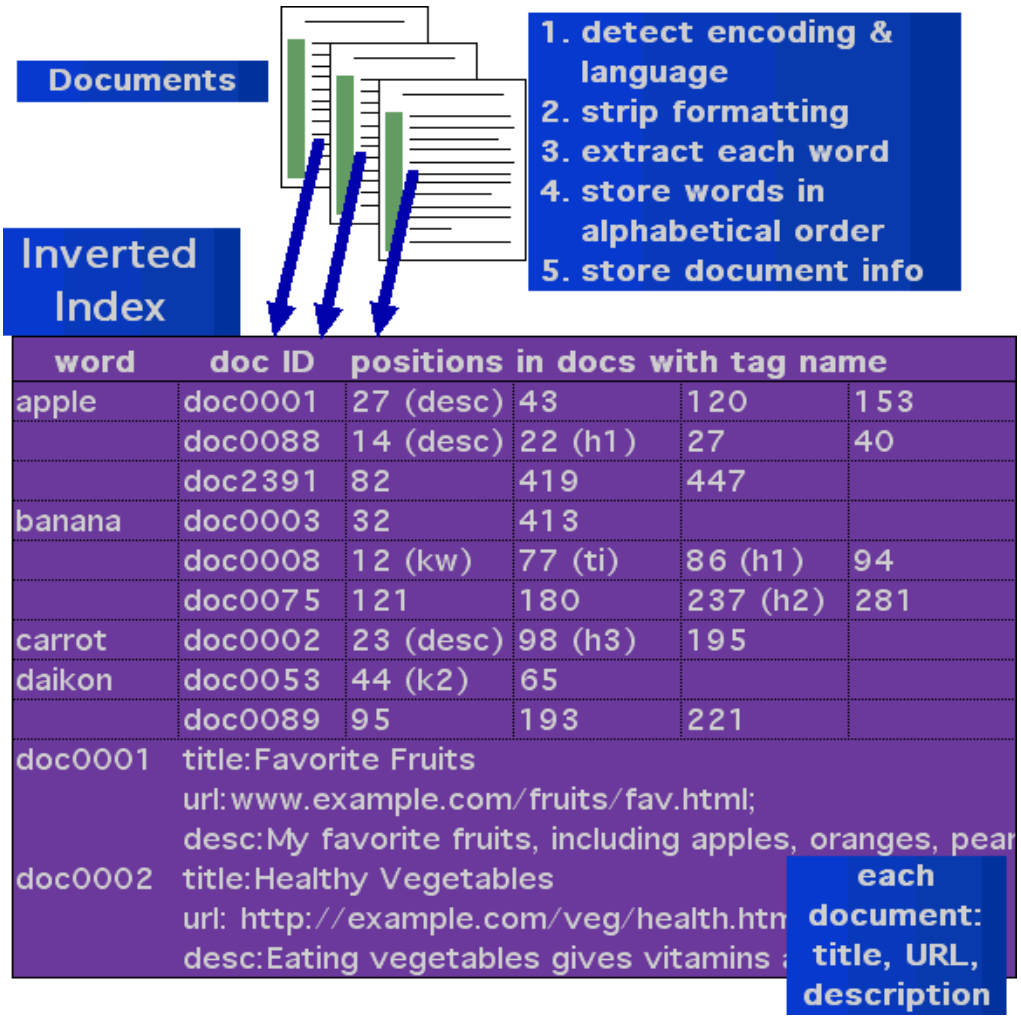


รูปที่ 7.3 : แสดงการสืบค้นจาก Index

การทำ Indexing ของเอกสารโดยทั่วไปจะเกี่ยวกับ Non Object Terms ซึ่งเป็นตัวสะท้อนกลับของข่าวสารอย่างชัดเจนของเอกสาร ซึ่งจะเป็นการออกแบบเกี่ยวกับการให้ค่าน้ำหนัก การนำเสนอ หรือการสะท้อนกลับในเนื้อหาของข่าวสาร

3. ระบบค้นหา ทำหน้าที่รับคิวรี (query) ที่ผู้ใช้ต้องการค้นหาผ่านทาง ซีจีไอ (Common GatewayInterface) โดยนำข้อสอบถามของผู้ใช้ที่ป้อนเข้ามาทำการค้นหาในดัชนีที่สร้างโดย Indexing เพื่อค้นหาเอกสาร HTML ที่ตรงกับความต้องการของผู้ใช้ นำผลลัพธ์แสดงต่อผู้ใช้ผ่านทางบราวเซอร์ ดังรูปที่ 7.4

ปัญหาหลักที่ Search Engine พบคือ Query จากผู้ใช้ ในกรณีที่ Query มีลักษณะทั่วไปอาจทำให้ได้รับเอกสารจากการค้นหามากมายมหาศาล ในขณะที่คิวรีที่เฉพาะเจาะจงอาจทำให้ไม่ได้รับเอกสารใดๆจากการค้นหาเลย เอกสารที่ได้รับจากคิวรีที่มีลักษะทั่วไป ส่วนใหญ่ไม่เกี่ยวข้องกับคิวรีนั้น และมีเอกสารหลายๆ ฉบับที่เกี่ยวข้องแต่ไม่ถูกค้นพบ เนื่องมาจากคิวรีนั้นไม่มีคีย์เวิร์ดที่บ่งชี้ไปถึงเอกสารนั้น Search Engine ทั่วไปจะสืบค้นเอกสารโดยพิจารณาจากการเรียงตัวอักษรของคำ (Pattern matching)โดยไม่คำนึงในแง่ของความหมาย (Semantic matching)



รูปที่ 7.4 : แสดงระบบค้นหา

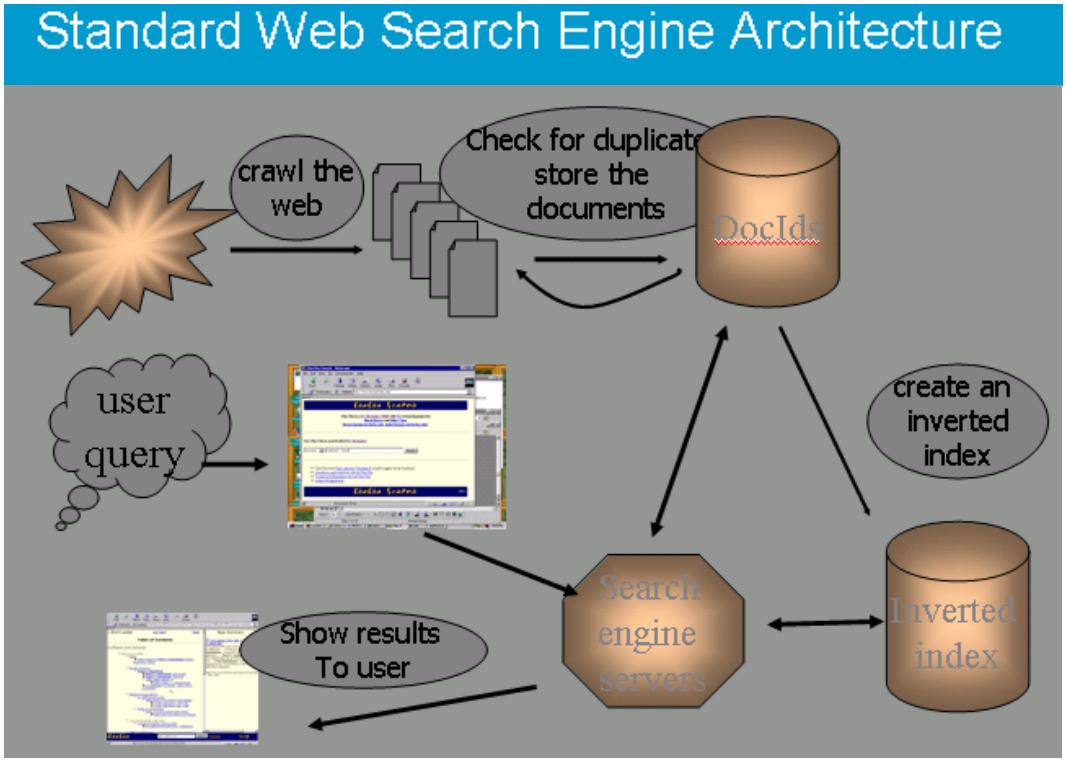
การค้นหาข้อมูล Search Engine สามารถกระทำได้หลายวิธีดังนี้

1. **ค้นหาจาก URL** เมื่อต้องการค้นหาข้อมูลโดยการใส่คำ (Keywords) หรือข้อความลงไป Search Engine จะเริ่มค้นหาคำที่ใส่ลงไปนั้นจากที่อยู่ของ URL ของเว็บไซต์ต่าง ๆ เช่น ต้องการค้นหาคำว่า “Travel” ระบบจะกระทำการค้นหาซึ่งได้ผลลัพธ์เป็นเว็บไซต์ที่มี URL เป็น www.Travel.com อยู่ในลำดับต้น ๆ แต่ก็อาจจะมีคำที่ค้นหาไม่พบเหมือนกัน

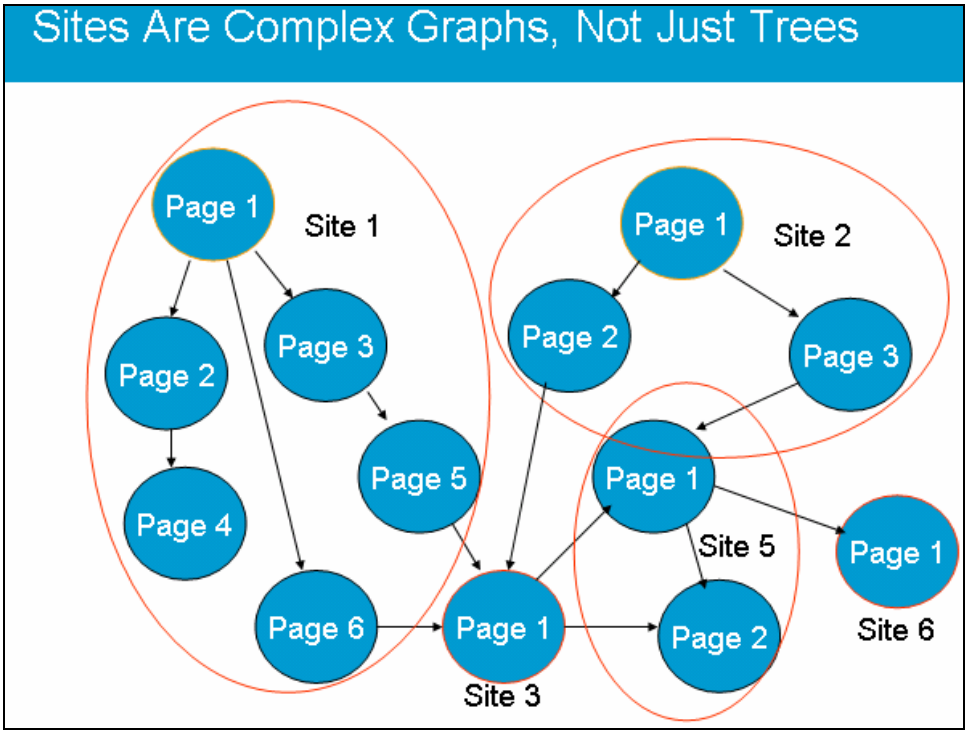
2. **ค้นหาจากคำที่อยู่ในไตเติ้ล (Title)** ของเว็บไซต์ที่เป็นประโยค หรือกลุ่มคำที่อยู่ทาง ด้านบนขอบด้านซ้ายของบราวเซอร์

3. การค้นหาจากส่วนที่อธิบายเว็บไซต์ สำหรับการแสดงคำอธิบายบอกลักษณะหรือเนื้อหาของ site ที่ทำ โดย Search Engine แต่ละตัวก็จะมีกระบวนการที่แตกต่างในการแสดงคำอธิบายของ site ส่วนหนึ่งที่จะนำมาจากคำสั่ง Tag ที่มีอยู่ในเอกสาร HTML ซึ่งเรียกว่า METE Description Tag โดยคำสั่งนี้จะเป็นตัวบอก หรืออธิบายว่า site นั้น ๆ ทำงานเกี่ยวกับอะไรบ้าง มีอะไรบ้าง ซึ่งตรงส่วนนี้จะไม่ปรากฏที่หน้าเว็บเพจแต่สามารถดูได้จาก Source Code ของเอกสาร HTML ซึ่งแหล่ง Search Engine จะใช้พวก Robots หรือ Spiders ในการสำรวจในไซต์ ต่าง ๆ มาแสดงผล

4. การค้นหาจากคำสั่ง Keyword การค้นหาจากคำสั่ง Keyword ก็เป็นการค้นหาจากคำหลักที่ระบุไว้ ในส่วนของเอกสาร HTML โดยตัว Search Engine ส่วนหนึ่ง จะนำมาจากคำสั่ง Tag ที่มีอยู่ในเอกสาร HTML ซึ่งเรียกว่า META Keyword Tag โดยคำสั่งนี้ จะบอกหรืออธิบายว่าในไซต์นั้น ๆ ทำงานเกี่ยวกับอะไร



รูปที่ 7.5 : แสดง สถาปัตยกรรมมาตรฐานของ Search Engine

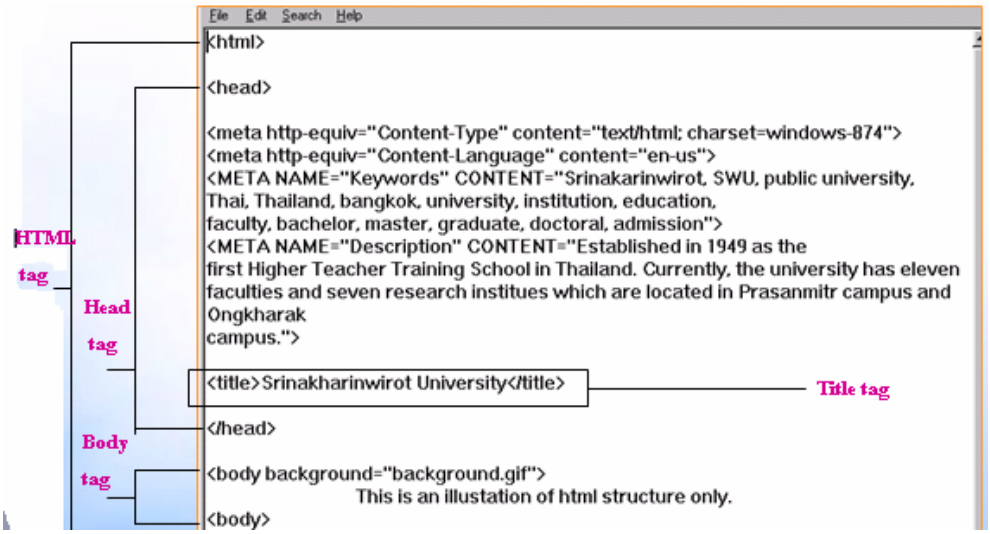


รูปที่ 7.6 : แสดงการเชื่อมโยงของ Site ต่าง ๆ เป็นโครงสร้างชนิด Graphs

7.3 A Taxonomy For Search Tools And Service

จากรูปที่ 7.6 แสดงให้เห็นถึงการเชื่อมโยงของ Site ต่างๆที่เป็นโครงสร้างชนิด Graphs ที่ซับซ้อน จึงมีการพัฒนาวิธีการในการค้นคืนสารสนเทศบนเว็ปขึ้นมาเพื่อเป็นเครื่องมือที่มีประสิทธิภาพในการสืบค้นข้อมูล รวมทั้งได้พัฒนาการบริการในการค้นหาให้แก่ผู้ใช้เพื่อให้การใช้ Web Search มีความง่ายขึ้นและตรงตามจุดประสงค์ของผู้ใช้ให้มากที่สุด

Search Tools ที่พัฒนาขึ้นมาเพื่อใช้งานนั้นสามารถแบ่งได้เป็น 3 กลุ่ม ซึ่งได้จำแนกตาม วิธีการในการสำรวจ Web(Method For Web Navigation) , เทคนิคในการสร้างดัชนี(Indexing Techniques) ,ภาษาสอบถาม(Query Language) , กลยุทธ์ในการจับคู่ระหว่างเอกสารและข้อสอบถาม(strategies for query – document matching) และวิธีการในการแสดงผลลัพธ์ที่ได้จากการสอบถาม(method for presenting the query output) ดังนี้



รูปที่ 7.7 : แสดงไฟล์ HTML

7.3.1 Indexing Search Engine

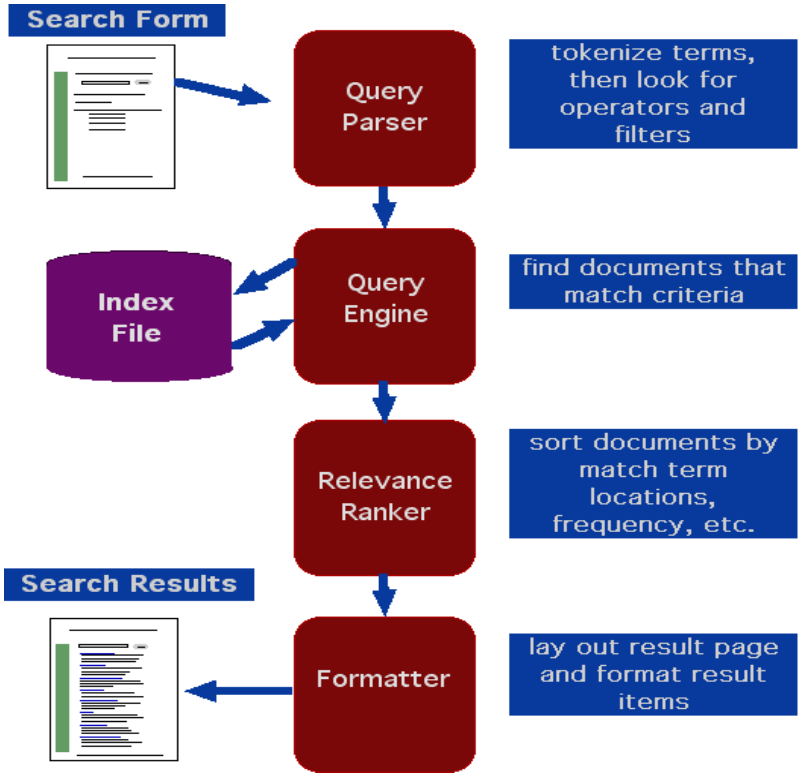
วิธีการทำงานของ Search Tools ชนิดนี้จะใช้ตัว Keyword Index เพื่อใช้ในการค้นหาจากข้อความในเว็บที่ได้สำรวจมาแล้ว ซึ่ง Search engine จะอ่านคำบนเว็บเพจอย่างน้อยที่สุด สองสามคำแรกที่ใช้ในเพจ ซึ่งรวมถึงคำที่ อยู่ใน Tag (รูปแบบของคำสั่งที่ใช้ในไฟล์ HTML)เช่น <title> </title> และหลังคำสั่ง alt (กรณีที่เป็นรูปภาพ) ซึ่ง search engine จะไม่นำเอาคำสั่ง Tag ของไฟล์ HTML เช่น คำสั่ง Java หรือ คำสั่งอื่น เช่น “and” , “the” , “by”, “for” เข้ามารวมเป็นข้อมูลที่ใช้ในการการค้นหาด้วย Web page ประเภทนี้จะยึดตำแหน่งและความบ่อยของการค้นเจอคำนั้นในข้อมูลมาจัดเรียงลำดับ page ก่อนหลัง ซึ่งวิธีการค้นหาข้อมูล โดย search Tools ชนิดนี้จะมีความรวดเร็วแต่ความละเอียดในการจัดแยกหมวดหมู่ของข้อมูลค่อนข้างน้อย เนื่องจากไม่ได้คำนึงถึงรายละเอียดของเนื้อหา ผลลัพธ์ที่ได้มากจนเกินไป ขาดการประเมินและกลั่นกรองเนื้อหาสาระใน Web pages ที่ไปเก็บรวบรวมมา แต่หากว่าต้องการแนวทางด้านกว้างของข้อมูลและความรวดเร็วในการค้นหา วิธีการนี้ก็ใช้ได้ผลดี แต่ก็ขึ้นอยู่กับข้อความที่ผู้ใช้ใช้สำหรับการค้นที่มีความซับซ้อนเพียงใดด้วย เว็บ Keyword Index ได้แก่

- **AltaVista,** http://altavista.digital.com
- **AOL NetFind,** http://www.aol.com/netfine
- **Excite,** http://www.excite.com

- **HotBot,** http://www.hotbot.com
- **Infoseek,** http://www.infoseek.com
- **LookSmart,** http://www.looksmart.com
- **Lycos,** http://www.lycos.com
- **Northern Light,** http://www.nothernlight.com

ตัวอย่างที่ 7.1 การค้นหาโดยใช้ Indexing Search Engine

สมมุติว่าต้องการค้นหาคำว่า ”Pentium” ผลลัพธ์ที่ได้จากการทำงานคือ web ที่มีคำว่า “Pentium” ในคำสั่ง <titile> </title> โดย web page ที่มีคำว่า “Pentium” อยู่มากกว่า ขึ้นมาเป็นอันดับแรก ๆ ก่อน web page ที่มี”Pentium” น้อยกว่า เช่น web page ที่ประกอบด้วยคำว่า “Pentium” อยู่ 20 คำ ย่อมจะถูกนำขึ้นมาแสดงที่อยู่พร้อม URL ก่อน web page ที่ประกอบด้วยคำว่า “Pentium” น้อยกว่า



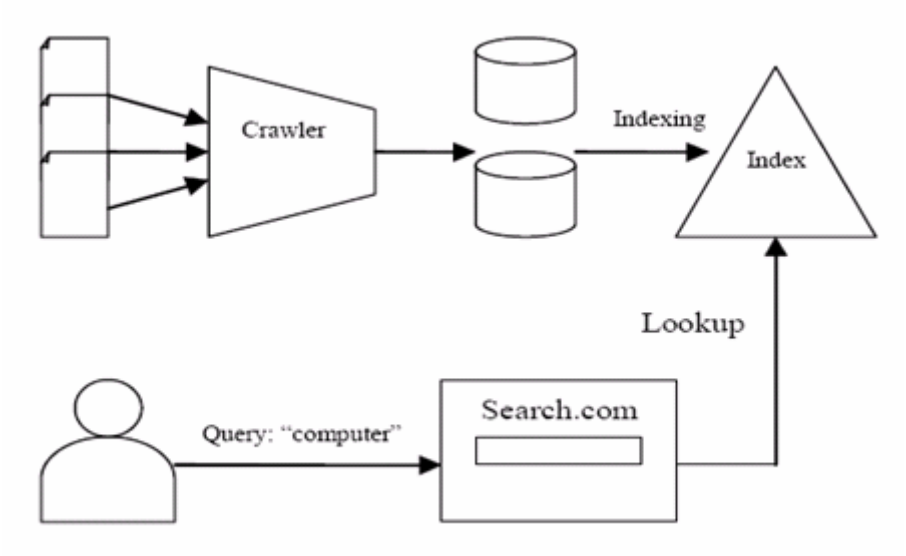
รูปที่ 7.8 : แสดงการค้นหาโดยใช้ Keyword Index

Alta Vista (www.altavista.com)

เป็น Search Engine ยอดนิยมอันดับต้น ๆ ที่ได้รับความเชื่อถือกันมานาน เพราะค้นหาได้รวดเร็วและมีความแม่นยำค่อนข้างดี โดยใช้คำหลัก (Key Words) จาก 25 ภาษา หากมีข้อมูลที่พบมากเกินไปก็จะมีวิธีการที่ฉลาดในการกลั่นกรองสิ่งที่ค้นหา Alta Vista ยังสามารถตรวจดูกลุ่มข่าวสารอีกนับพันบน Usenet สำหรับการส่งข่าวที่หลากหลาย Alta Vista มี Spider ที่เรียกว่า Scooter ในการ update index

Alta vista สามารถสนับสนุน Boolean , วลี , และตัวอักษรพิมพ์ใหญ่กับตัวอักษรพิมพ์เล็กแตกต่างกัน (Case – Sensitive) ผลของการค้นหาของความสัมพันธ์จะให้ค่า Score ที่สูงจากเอกสารที่คิวรี่ในสอง-สามคำแรก และเอกสารที่ทำการคิวรี่ของคำที่มีการค้นพบ ผลที่ได้จะประกอบด้วย Title , a shot Abstract , size , and date of the last modification ของแต่ละเอกสารที่ทำการค้นคืน

สิ่งที่ Alta Vista จะไม่ค้นหา หน้าลงทะเบียน,ข้อความในรูปภาพ และ Multimedia Files, XML , Java applets , Comment tags , Acrobat Files , Spammers ซึ่งอันดับมีผลต่อความเร็วในการ Load เนื้อหาและการจัดวาง มีความสำคัญมาก อิเช่น ส่วนบนสุดของหน้าเวบไซต์ รวมทั้ง หัวข้อ HTML สำคัญมาก แต่ละหน้าไม่ควรซ้ำกัน แต่สำหรับ Meta Tags ไม่มีความสำคัญ แต่ควรมี Altavista ใช้เวลาหลายเดือนในการ update index แต่ละครั้ง ความถี่ของการ update index จะขึ้นอยู่กับความสำคัญของข้อมูลและปัจจัยอื่นๆ



รูปที่ 7.9 : สถาปัตยกรรมของ Indexing Search Engine (keyword search)

Excite (www.excite.com)

Excite เคยเป็น search engine ที่เป็นที่นิยมของผู้ใช้ Excite จะมี Spider และ Indexer สำหรับการค้น ผลของการสืบค้นจะได้เป็นข้อมูลฉบับเต็ม (Full text) ที่ต้องการ ซึ่ง Spider จะทำการค้นคืนสำหรับ Web และ Usenet New ซึ่งใน Excite จะมีการรวบรวม Index ไว้ถึง 50 ล้าน URL Excite จะสามารถ Support การ search ชื่อบุคคล, Boolean operator และ Boolean Expression ผลของการค้นคืนจะให้ค่า Rank ซึ่งรวบรวมจาก keywords ที่ ปรากฏในส่วน title page หรือ keywords ที่ปรากฏซ้ำๆบ่อยๆ รวมทั้ง ความยาวของ URL ซึ่ง URL ที่ไม่ยาวมากจะถูก rank อยู่ในอันดับต้นๆ อีกทั้ง การโฆษณาให้กับ Excite ด้วย

Excite เป็น Search Engine ที่มีข้อเด่นคือ การใช้เทคนิค Concept Based ทำให้หาข้อมูลได้ตรงประเด็นมาก เช่น ถ้าต้องการค้นเรื่องเกี่ยวกับ ดอกบัว ผู้ใช้พิมพ์คำว่า Lotus เพื่อสืบค้นใน Search engine ตัวอื่นผลที่ปรากฏออกมานั้นคำว่า Lotus นี้อาจเป็นชื่อบริษัท ชื่อโปรแกรม และรถสปอร์ตยี่ห้อดัง แต่ถ้าค้นโดยใช้ Excite โดยเลือกแสดงผลในหัวข้อ Directory จะเห็นว่า Excite มีการจัดข้อมูลไว้เป็นหมวดหมู่จำแนกกันไปว่า Lotus ที่สนใจเป็นเรื่องเกี่ยวกับพืช หรือ รถยนต์ หรือโปรแกรมคอมพิวเตอร์ ทำให้สามารถทำการสืบค้นต่อไป ได้ตรงประเด็นขึ้น ในปัจจุบัน Excite ไม่รับการ submit URL แล้ว และไม่ได้ update index ด้วย spider ของตนเอง

Hotbot (www.hotbot.com)

เป็น Search Engine ที่จะไม่ค้นหาโดยการแบ่งหน้าจอ รวมทั้ง หน้าที่มีการใช้ File Cookies หรือ URL ที่มีเครื่องหมายพิเศษต่างๆ รวมทั้ง Spammers Hotbot จะสนใจถึงอันดับ ซึ่งอันดับของการแสดงผลนั้นจะมีผลต่อการโหลดถ้าเกิดว่า Server ช้าเกินไป นอกจากนี้ยังสนใจถึง เนื้อหาและการจัดวาง ซึ่งความยาวของเอกสาร และ ความถี่ของคำสำคัญที่พบ ส่งผลต่อการจัดอันดับเช่นกัน

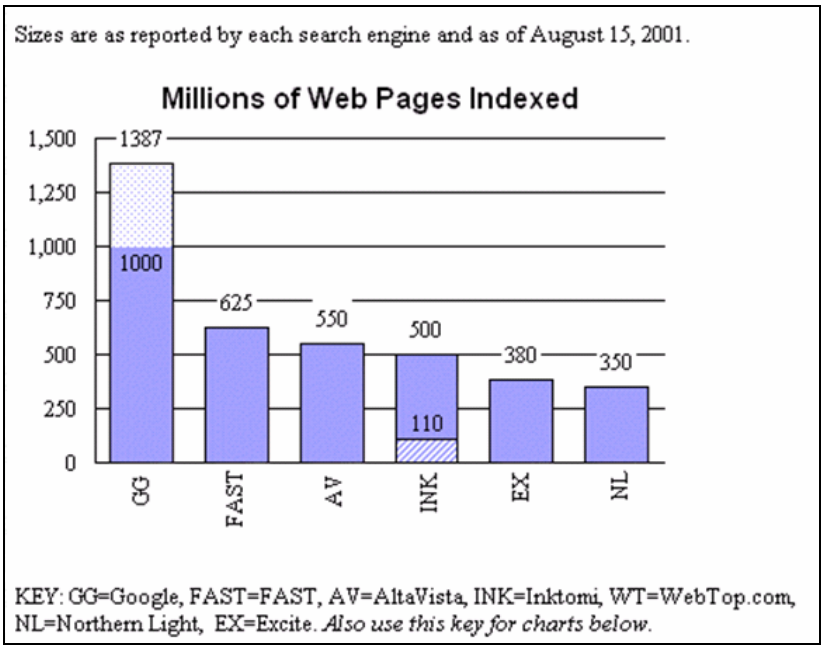
Hotbot ให้ความสำคัญมากสำหรับ หัวข้อ HTML , Meta Tags ซึ่งเป็นคำอธิบายไม่เกิน 150 ตัวอักษร และคำสำคัญ ไม่เกิน 75 ตัวอักษร รวมทั้งความถี่ของคำคัญ ปกติแล้วกำหนดไว้ที่ 3-7 % ของเนื้อหาทั้งหมด รวมทั้ง จำนวน Link ที่โยงเข้ามา มีการใช้ Inktomi เพื่อใช้สำหรับเป็นเงื่อนไขในการพิจารณา HotBot จะสามารถ Support เกี่ยวกับตัวอักษรพิมพ์ใหญ่กับตัวอักษรพิมพ์เล็กแตกต่างกัน (Case – Sensitive) , Boolean และ Advance เกี่ยวกับการระบุรายละเอียดของแต่ละทางเลือกที่มีความเฉพาะในรูปของ Media

Info Seek Guide (www.infoseek.com หรือ infoseek.go.com)

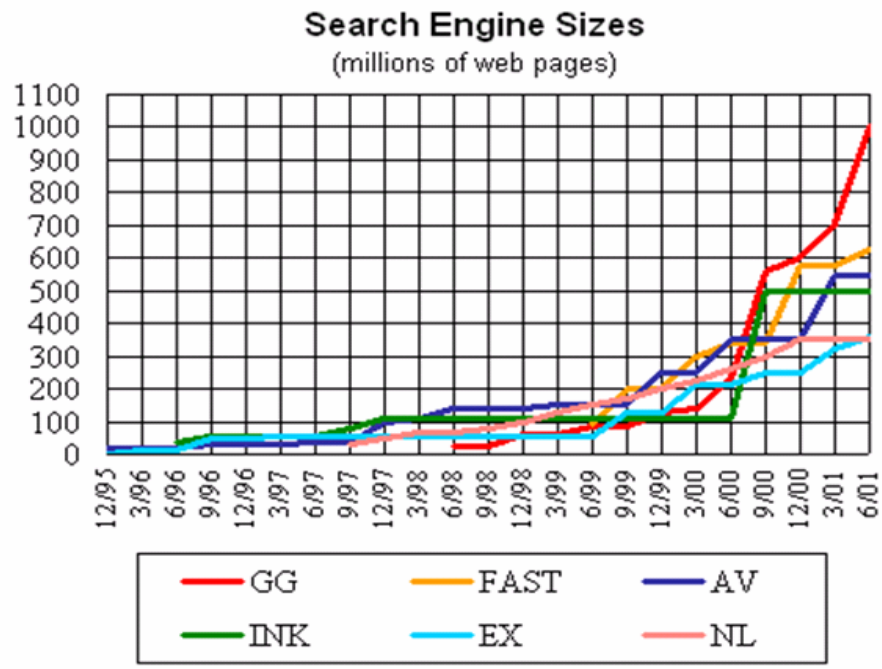
เป็น Search Engine ที่มีความ Popular เกี่ยวกับที่ใช้ Robot ในการค้นคืนเอกสารที่เกี่ยวกับ HTML , เอกสารที่เป็น PDF และเอกสารที่ข้อมูลฉบับเต็ม (Full text) Info seek จะทำการค้นหาใน Web , use net Group , และ Web FAQ ซึ่งจะใช้ตัว Index ในการกระจาย

Info seek จะสามารถ Support เกี่ยวกับตัวอักษรพิมพ์ใหญ่กับตัวอักษรพิมพ์เล็กแตกต่างกัน (Case – Sensitive) ,สัญลักษณ์ (Symbol) , วลี (Phrase) และเกี่ยวกับชื่อบุคคลนอกจากนี้สามารถ search เกี่ยวกับรูปภาพ โดยการใส่ชื่อหัวข้อเข้าไป

Info Seek จะทำการจัดอับดับ Output ของการประมวลผลโดยใช้ RSV โดยจะให้ค่าน้ำหนักของเอกสารตามที่บรรจุคิวรี่ของคำของเอกสารเริ่มต้น และการกลับคืนของ Short Summary Relevancy Score , และ Document Size



รูปที่ 7.10 : แสดงขนาดของ Search Engine ต่าง ๆในปี ค.ศ.2001



รูปที่ 7.11 : แสดงการใช้งานSearch Engine ระหว่างปี ค.ศ.1995 ถึง ค.ศ.2001

7.3.2 Subject Directories

Search Tools ชนิดนี้ใช้ Subject Directories web ซึ่งจะแสดงหัวข้อของสิ่งที่ค้นหา ซึ่งข้อดีของ Search Tools ชนิดนี้ คือ Site แต่ละ Site จะถูกแบ่งกลุ่มตามเนื้อหาสาระก่อนที่จะถูกนำมาใช้เป็นข้อมูลในการค้นหาต่อไป และ การค้นหาของค่อนข้างจะตรงประเด็นกับผลลัพธ์ที่ได้ ดังรูปที่ 7.12

Web ประเภทนี้ เหมือนกับแคตตาล็อกสินค้า โดยมันจะแสดงหัวข้อของสิ่งที่ต้องการค้นหา แต่ละหัวข้อจะแสดงถึง URL, รายละเอียดเกี่ยวกับ URL นั้น ๆ ซึ่งได้มาจากการวิเคราะห์ดูเนื้อหา ของแต่ละ web page ว่าเกี่ยวกับอะไร ซึ่งอยู่ที่ต้องใช้คนพิจารณาในการเลือกหัวข้อ แยกหัวข้อ

ข้อดี

1. Search engine ประเภทนี้ site แต่ละ site จะถูกแบ่งกลุ่มตามเนื้อหาสาระของมัน ก่อนที่จะถูกนำมาใช้เป็นข้อมูลในการค้นหาต่อไป

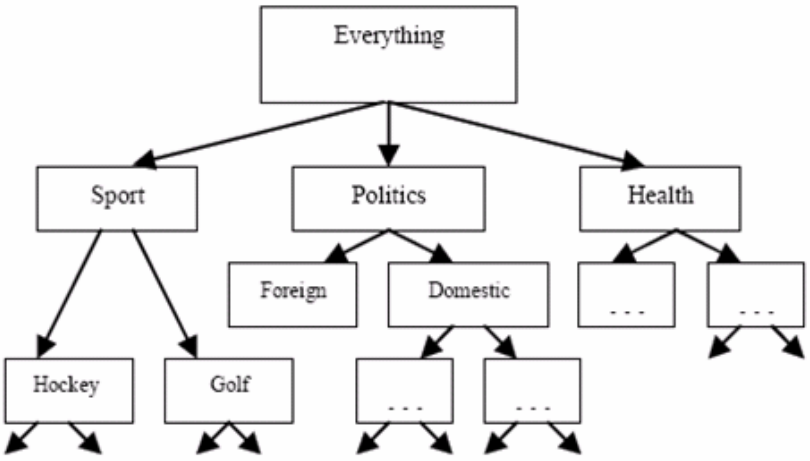
2. การค้นหาของผู้ใช้ ค่อนข้างจะตรงประเด็นกับผลลัพธ์ที่ได้จาก search engine ชนิดนี้ เพราะการใช้คนในการจัดหัวข้อต่าง ๆ แล้วมาพิจารณาอย่างละเอียดถี่ถ้วน เกี่ยวกับรายละเอียดของ site ต่าง ๆ จะได้ใจความที่สำคัญออกมา ที่จะใช้ในการจัดทำหัวข้อนั้น ย่อมดีกว่าใช้ คอมพิวเตอร์ในการพิจารณา และยังสามารถจัดหัวข้อได้ตรงกับเนื้อความอีกด้วย

ข้อเสีย

1. ในด้านการแยกหัวข้อจากความคิดของแต่ละคนอาจไม่ตรงกัน เวลาค้นหาก็ใช้หัวข้อที่แตกต่างไป ทำให้ไม่ได้ข้อมูลที่ต้องการได้ ผลลัพธ์ที่ได้อาจไม่ตรงตามความต้องการนัก เพราะการตีความที่แตกต่างกัน

ตัวอย่างเว็บ ชนิด Subject Directories

- **Galaxy,** http://www.einet.net
- **Magellan,** http://www.mckinley.com
- **NetGuide,** http://www.netguide.com
- **WebCrawler,** http://www.webcrawler.com
- **Yahoo,** http://www.yahoo.com



รูปที่ 7.12 : แสดง Topic hierarchy taxonomies ของ Subject Directory

Yahoo (www.yahoo.com)

เป็นเว็บไซต์จักกันดีที่สุดในการค้นหาออนไลน์ มีบริการข่าวสาร การเงิน และจดหมายอิเล็กทรอนิกส์ฟรี รูปแบบการจัดหน้า Site จัดไว้เป็นหมวดหมู่ตั้งแต่ศิลปะ และเรื่องราวของมนุษย์ ไปจนถึงคอมพิวเตอร์และธุรกิจ Yahoo ยังได้ทำ Site พิเศษสำหรับบางประเทศเช่น เยอรมัน และฝรั่งเศส และสำหรับบางเมืองเช่น บอสตัน และชิคาโก

สิ่งที่ Yahoo จะไม่ค้นหาคือ Spammers สำหรับเนื้อหาและการจัดวาง จะพิจารณาโดยกองบรรณาธิการของ Web เพื่อให้อยู่ในหมวดหมู่ที่เหมาะสม หัวข้อ HTML อาจเป็น ชื่อเรื่องที่ผู้ใช้ป้อน อีกทั้ง Meta Tags ซึ่งเป็นคำอธิบายและคำสำคัญมีบทบาทมากเพื่อนำมาเป็นเงื่อนไขในการพิจารณา โดยผลของการปฏิบัติงานนั้นจะได้คำอธิบายและคำสำคัญกะทัดรัดและถูกต้อง และอยู่ในหมวดหมู่ที่เหมาะสม

7.3.3 Meta Search

Search Service ชนิดนี้จะใช้ Meta search Engines ซึ่งลักษณะเด่นของเว็บประเภทนี้ได้แก่ ความหลากหลายของข้อมูล เช่น ผลลัพธ์ที่ได้จะมาจาก Search engine ชนิดต่าง ๆ ซึ่งพยายามที่จะขจัดปัญหาข้อจำกัดในเรื่องของขอบเขต โดยการเสนอการ query โดยใช้หลาย ๆ มาตรฐานการ search engines ให้เป็นหนึ่งเดียว จุดเด่นของการค้นหาด้วยวิธีการนี้ คือ สามารถเชื่อมโยงไปยัง Search Engine ประเภทอื่นๆ และแสดงออกมาโดยจะต้องมีความสอดคล้องกับ user interface ด้วย และยังมีความหลากหลายของข้อมูล เช่น ผลลัพธ์ที่ได้มาจาก search engine ชนิดต่างๆเราสามารถเชื่อมต่อไปยัง Search Engine ชนิดต่างๆได้ ตัวอย่างเช่น web MetaCrawler ที่จะมีการสร้างตัวเชื่อมไปยังเว็บ search engine ประเภทต่าง ๆ ไว้ท้ายข้อมูลที่ query ออกมา แต่การค้นหาด้วยวิธีนี้มีจุดด้อย คือ วิธีการนี้จะไม่ให้ความสำคัญกับขนาดเล็กใหญ่ของตัวอักษร และมักจะผ่านเลยคำประเภท Natural Language (ภาษาพูด) เช่น Dog pile, Inference Find, Met Crawler

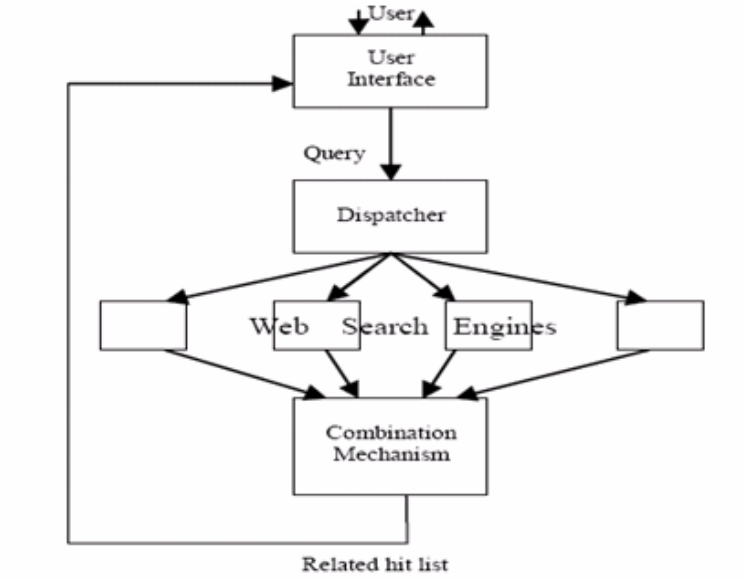
NEC Research Institute(NECI) พัฒนา Mata-search engine เพื่อปรับปรุงผลและความแม่นยำโดยการ download และการวิเคราะห์เอกสารแต่ละเอกสารและการแสดงผลที่แสดง term query ช่วยให้ผู้ใช้ส่วนมากสามารถตัดสินใจได้อย่างรวดเร็ว ถ้าเอกสารมีความสัมพันธ์กับการ download แต่ละหน้า เทคนิคอย่างหนึ่งที่มันสามารถทำให้ได้ผลอย่างมาก คือ การทำให้ลักษณะทั่วๆไป สัมพันธ์กับขนาดของ drivers และการจัดระเบียบที่ไม่ดีของ

Anastasios Tombros ได้ทำการวิจัยเพื่อพิสูจน์ถึงข้อสรุปในข้อดีของการ query ชนิดนี้ การศึกษาของเขาที่ระบุว่า user จะรู้สึกดีกับการ query ที่เร็วและตรงตามที่ต้องการมากกว่าความถูกต้อง และผู้ใช้ที่ทำงานแบบสรุปอย่างเร็วหรือการ query ที่ไม่ดีกับเอกสารแบบสรุป การ query ข้อมูลแบบสรุปจะต้องลดความต้องการเอกสารของผู้ใช้แบบเต็ม

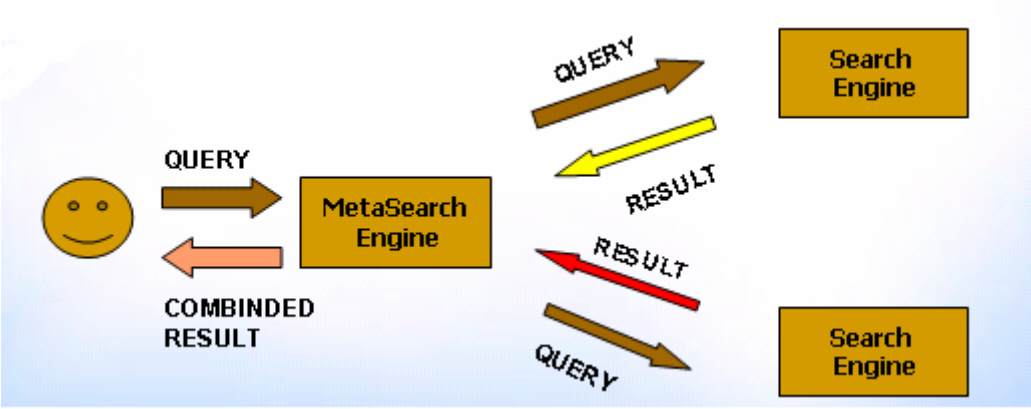
ตัวอย่าง Web ชนิด Meta-search Engines ได้แก่

- **All4One Search Machine,** http://www.all4one.com
- **Dogpile,** http://www.dogpile.com
- **Highway 61,** http://www.highway61.com
- **HuskySearch,** http://huskysearch.cs.washington.edu
- **Inference Fine,** http:www.inference.com/ifind
- **Mamma,** http://www.mamma.com
- **MetaCrawler,** http://www.metacrawler.com
- **MetaFine,** http://www.metafind.com
- **OneSeek.Com,** http://www.oneseek.com
- **Profusion,** http://profusion.ittc.ekans.edu
- **SavySearch,** http://savvy.cs.colostate.edu:2000

จากรูปที่ 7.13 แสดงถึงสถาปัตยกรรมของ Meta-Search Engine และรูปที่ 7.14 แสดงถึงการทำงานของ Meta-Search Engine ที่มีการนำคำถามหรือข้อสอบถามจากผู้ใช้ (Query) นำไปสอบถาม Search Engine ตัวอื่น แล้วนำผลลัพธ์มารวมกันเพื่อนำเสนอในรูปแบบที่เหมาะสม



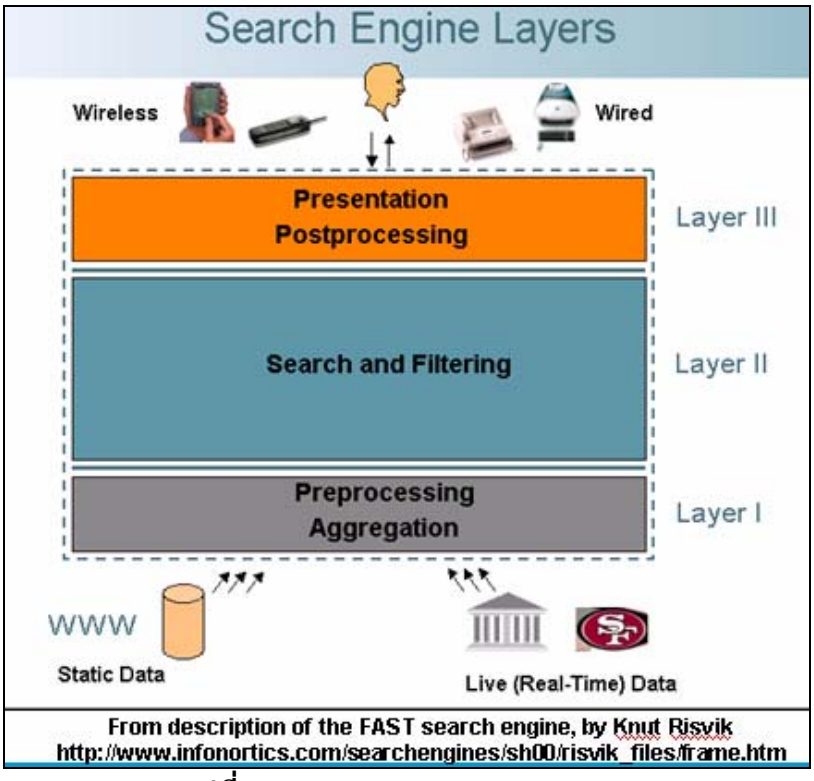
รูปที่ 7.13 :แสดงสถาปัตยกรรมของ Meta-Search Engine



รูปที่ 7.14 : แสดงถึงการทำงานของ Meta-Search Engine

ปัญหาของ Meta-Search Engine นั้นเป็นเรื่องของผลลัพธ์ที่ได้จากการค้นหามักจะซ้ำกัน เนื่องจากตัว Search Engine ต้องไปดึงข้อมูลจากหลายๆที่ และไม่มีการจัดลำดับ Ranking

สำหรับผู้ใช้ที่ต้องการค้นหาข้อมูลทั่ว ๆ ไปที่ไม่เน้นความสำคัญ ของเรื่องต่าง ๆ เหล่านี้แล้ว ให้ใช้เว็บประเภทนี้ เป็น search engines จะเหมาะสมที่สุด แต่ถ้าต้องการข้อมูลที่ถูกต้อง และมีการกลั่นกรองอย่างพิถีพิถันแล้ว Meta-search Engines นั้นไม่ใช่ทางเลือกที่ดี



รูปที่ 7.15 : แสดง Search Engine Layer

7.4 วิธีการค้นหาข้อมูลด้วยsearch engine ที่ทำให้ค้นหาข้อมูลให้ง่ายขึ้น

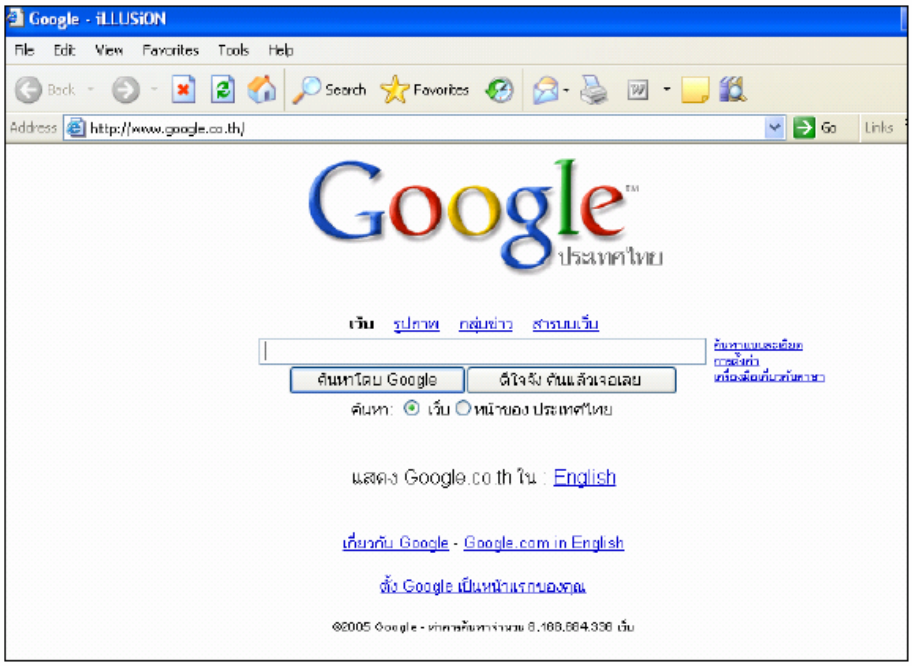
มีวิธีการหลายวิธีและหลายเทคนิคในการค้นหาข้อมูลที่ให้ประสิทธิภาพและประสิทธิผลที่ดี เพียงแต่เวลาเรา Search ข้อมูล เราต้องรู้จักใช้ ให้ถูกต้อง จะได้ข้อมูลอย่างรวดเร็วและตรงตามที่ต้องการมากที่สุด อาจจะกระทำโดย

1. การบีบประเด็นให้แคบลง เนื่องจากจำนวนข้อมูลที่มีมาก เพิ่มขึ้นจำนวนไม่น้อย ทำให้การค้นหาที่ได้ก็มีมาก เราควรทำการค้นหาโดยทั่วๆ ไปก่อน เกี่ยวกับเรื่องที่ต้องการค้นหา ต่อจากนั้นค่อยระบุ หรือใส่ option เพิ่มเติมเกี่ยวกับสิ่งที่ต้องการลงไป เพื่อบีบประเด็นให้ครอบคลุมมากยิ่งขึ้น

2. ใช้ advanced search เพื่อช่วยในการค้นหาข้อมูล ในsearch engine ต่าง ๆ ในปัจจุบันจะมีการแทรก option ต่าง ๆ เข้าไปเช่น “Advanced Option” หรือ “Power Search”

7.4.1 การค้นหาข้อมูลจากเว็บไซต์ www.Google.com

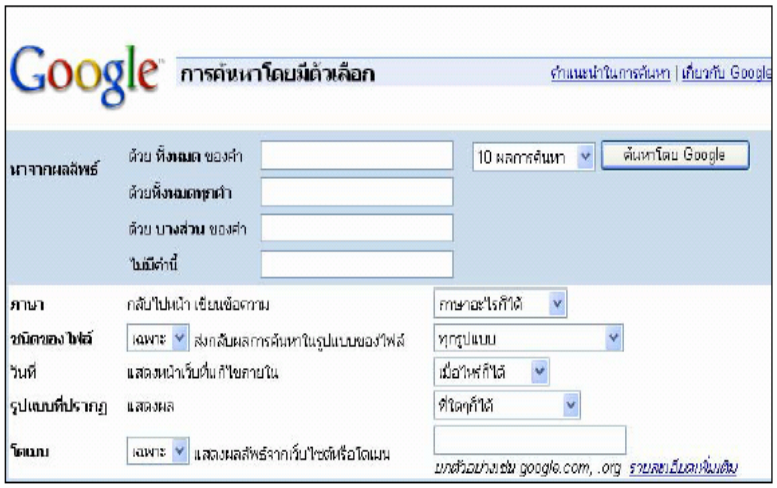
Google เป็นเว็บไซต์ฐานข้อมูลที่ใหญ่มากที่สุดของโลก ในอดีตเป็นบริษัทที่ดำเนินการด้านฐานข้อมูลเพื่อให้บริการแก่เว็บไซต์ค้นหาอื่น ๆ ปัจจุบันได้เปิดเว็บไซต์ค้นหาเอง ด้วยฐานข้อมูลมากกว่า สามพันล้านเว็บไซต์และเพิ่มขึ้นเรื่อย ๆ ทุกวัน ที่เหนือกว่าบริการรายอื่น ๆ คือ เป็นเว็บไซต์ที่สนับสนุนภาษาต่าง ๆ มากกว่า 80 ภาษาทั่วโลก (รวมทั้งภาษาไทย) และมีเครื่องserver ให้บริการในส่วนต่าง ๆ ของโลกมากถึง 36 ประเภท (รวมทั้งในประเทศไทย)



รูปที่ 7.16 : แสดง Search Engine: Google

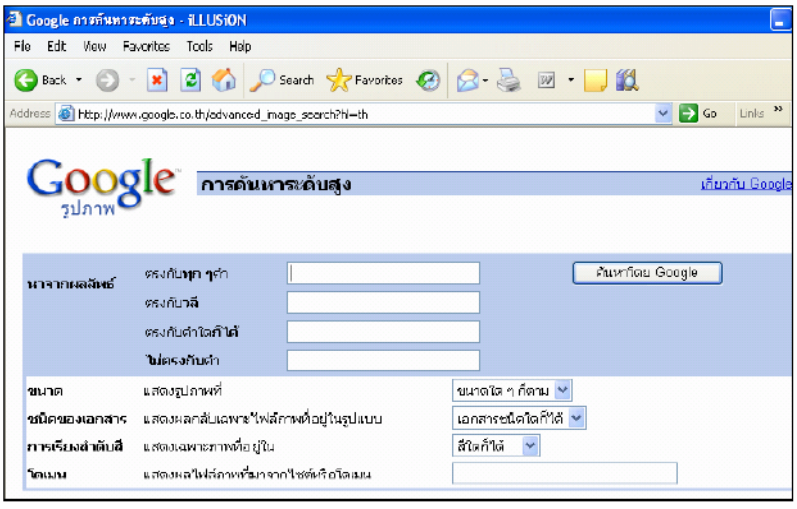
เมื่อเราเปิด Browser `และพิมพ์ URL : www.google.com ลงไป ด้วยระบบตรวจสอบภาษาของเว็บไซต์ Google เมื่อพบว่าเราใช้บราวเซอร์บนวินโดว์ภาษาไทยจะสวิทซ์เป้าหมายมายัง www.google.co.th โดยอัตโนมัติดังรูปที่ 7.16 การบริการค้นหาของ Google แยกฐานข้อมูลออกเป็น 4 หมวด ได้แก่ เว็บ , รูปภาพ , กลุ่มข่าว และสารบบเว็บดังมีรายละเอียดดังนี้

1. เว็บ : เป็นการค้นหาข้อมูลจากเว็บไซต์ต่าง ๆ ทั่วโลก การค้นหาแบบเจาะลึกเกี่ยวกับเว็บ สามารถระบุรายละเอียดต่าง ๆ ได้ เพื่อให้สามารถค้นหาได้ในวงจำกัด เช่น การกำหนดคำหลักที่ต้องการ คำที่คล้ายคลึง และคำที่ไม่ต้องการให้ปรากฏอยู่ด้วย ,กำหนดเฉพาะภาษา, ชนิดของไฟล์ (html ,word)) ช่วงระยะเวลาที่เอกสารนั้นสร้างขึ้น หรือ จากโดเมนเว็บไซต์ชื่ออะไร เป็นต้น



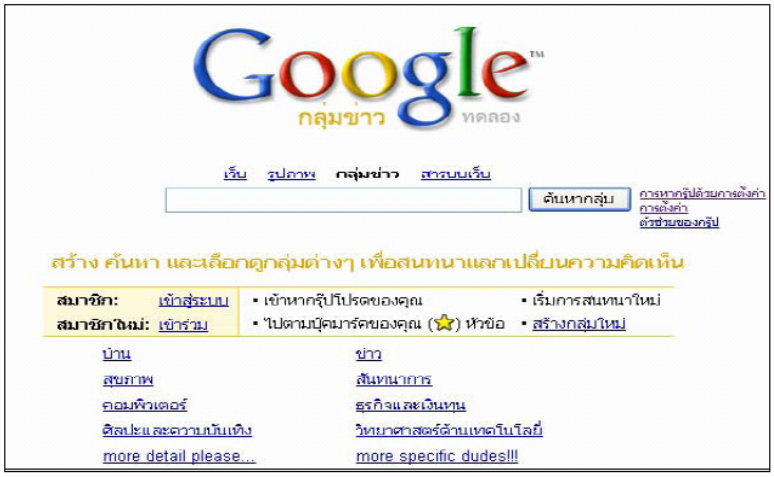
รูปที่ 7.17 : แสดงการค้นหาข้อมูลจากเว็บไซต์ต่าง ๆทั่วโลก

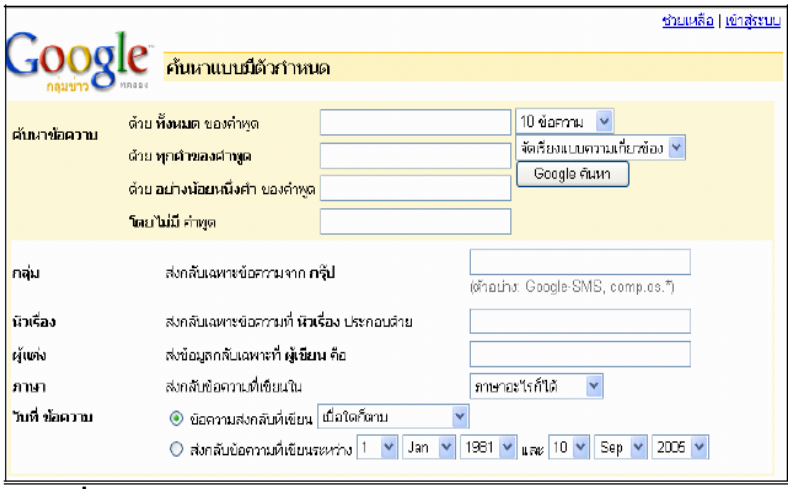
2. รูปภาพ : เป็นการค้นหารูปภาพหลากหลาย Format จากเว็บไซต์ต่าง ๆ ทั่วโลก การค้นหาภาพเพื่อให้สามารถค้นหาได้รวดเร็วควรใช้การค้นหาภาพระดับสูงเพราะสามารถระบุชื่อหรือบางส่วนของชื่อ ชนิดรูปภาพเป็นไฟล์ฟอร์แมตใด (JPG, GIF, PNG) ชนิดของสี (Black/White, Grayscale, Color) ชื่อของโดเมนที่คาดว่าจ่าจะมีภาพนั้น ๆ



รูปที่ 7.18 : แสดงการค้นหารูปภาพจาก Google.com

3. กลุ่มข่าว : เป็นการค้นหาเรื่องราวที่น่าสนใจจากกลุ่มข่าวต่างๆ เว็บไซต์ที่เป็นกลุ่มข่าวนั้นโดยปกติจะมีการแบ่งแยกเป็นหมวดหมู่ที่น่าสนใจอยู่แล้ว ผู้ใช้สามารถคลิกที่ชื่อกลุ่มข่าวที่สนใจได้ทันที แต่บางทีผู้ใช้อาจจะไม่แน่ใจว่าเรื่องที่สนใจนั้นอยู่ในกลุ่มข่าวใด ผู้ใช้สามารถใช้การค้นหาแบบพิเศษเป็น การค้นหากลุ่มข่าวแบบระบุรายละเอียด โดยระบุข้อความที่ต้องการค้นหาจากกลุ่มข่าวด้วยคำ หรือบางส่วนของข้อความ เช่นเดียวกับการค้นหาเว็บเพจ แต่สามารถคัดเลือกเอาเฉพาะคำที่ปรากฏในกลุ่มข่าว ผู้เขียน หมายเลขข้อความ ภาษาที่ใช้ รวมทั้งช่วงระยะเวลาตามที่ต้องการได้ด้วย





รูปที่ 7.19 : แสดงการค้นหากลุ่มข่าวแบบระบุรายละเอียด

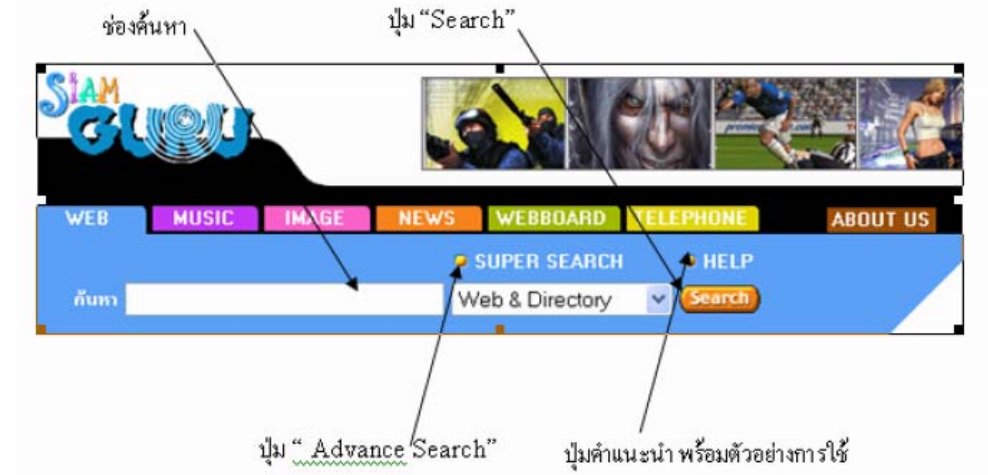
4 .สารบบเว็บ : การค้นหาข้อมูลจากเว็บไซต์ที่แยกออกเป็นหมวดหมู่

7.4.2 การค้นหาข้อมูลด้วย Basic Search จากเว็บไซต์ www.siamguru.com

Basic Search คือเครื่องมือในการค้นหาเว็บไซต์ ทำหน้าที่ในการให้บริการค้นหาข้อมูล (Search Engine) โดยเน้นเรื่องความสามารถในการค้นหาข้อมูลภาษาไทยบนอินเตอร์เน็ต มีความสามารถเทียบเท่า search engine ชื่อดังจากต่างประเทศ โดยการค้นหาจะเป็นแบบค้นหาข้อมูลจากทุกคำของข้อมูลจริง (Full Text Search) ทั้งภาษาไทยและภาษาอังกฤษจากเว็บเพจจำนวนหลายล้านหน้า มีการเก็บรวบรวมข้อมูลเว็บเพจที่เกี่ยวข้องกับประเทศไทยมาจัดทำดัชนี (index) โดยอัตโนมัติ ผสมกับการจัดแยกหมวดหมู่อย่างชัดเจน เพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูลได้ง่ายและรวดเร็วมากที่สุด

เว็บไซต์ www.siamguru.com แบ่งการค้นหาเป็น 4 รูปแบบคือ

1. ค้นหาเว็บไซต์ (Basic Search) เหมาะสำหรับผู้ที่ใช้อินเทอร์เน็ตทั่วๆ ไปที่ยังไม่มีความรู้มากนัก
2. ค้นหาเว็บไซต์แบบซูเปอร์เสิร์ช (Search) เป็นบริการสืบค้นข้อมูลแบบมีเงื่อนไขสำหรับการค้นหาที่มีการเฉพาะเจาะจงมากยิ่งขึ้น
3. ค้นหาเพลง (Music Search) บริการค้นหาเพลง , เนื้อร้อง จากเว็บเพจต่างๆ โดยสามารถค้นหาได้จากชื่อเพลง ชื่อนักร้อง ชื่ออัลบั้ม หรือ คำร้องจากท่อนใดท่อนหนึ่งก็ได้
4. ค้นหารูปภาพ (Image Search) บริการค้นหา ภาพถ่าย ภาพกราฟฟิก ภาพการ์ตูน เป็นต้น



รูปที่ 7.20 : แสดงการใช้ Basic Search จาก WWW.siamguru.com

การค้นหาด้วย Super Search

Super Search เป็นเครื่องมือค้นหาข้อมูลในอินเทอร์เน็ตอีกประเภทหนึ่ง เหมาะสำหรับผู้ที่คุ้นเคยกับ Basic Search อยู่แล้ว แต่ต้องการค้นหาข้อมูลให้ได้ผลลัพธ์ตรงความต้องการมากขึ้นกว่าที่จะสามารถทำได้ใน Basic Search ด้วยวิธีการสร้างเงื่อนไขการค้นหาขึ้น ซึ่งจะได้ผลลัพธ์ที่น่าพอใจกว่าใน Basic Search ในขณะเดียวกันการค้นหาแบบ Super Searchก็จะมีความซับซ้อนในการใช้งานด้วยเช่นกัน

รูปที่ 7.21 : แสดงการค้นหาด้วย Super Search

จากรูปที่ 7.21 จะเห็นได้ว่าประกอบด้วยส่วนประกอบ 4 ส่วนคือ

ส่วนที่ 1. ข้อความแบบมีเงื่อนไข เป็นช่องสำหรับกำหนดข้อความที่เป็นเงื่อนไขในการค้นหา

ตัวอย่างที่ 7.1 :

พิมพ์ **ไทย and จีน** ลงในช่องข้อความแบบมีเงื่อนไข จะหมายถึง ค้นหาคำว่า ไทย และ จีน โดยผลลัพธ์จากการค้นหา จะปรากฏคำว่า "ไทย" และ "จีน" อยู่ในหน้าเว็บเพจเดียวกัน

ตัวอย่างที่ 7.2 :

พิมพ์ **กีฬา or ดนตรี** ลงในช่องข้อความแบบมีเงื่อนไข Super Search จะค้นหาข้อมูลที่ปรากฏคำว่า "กีฬา" หรือ "ดนตรี" ในหน้าเว็บเพจ

ตัวอย่างที่ 7.3 :

พิมพ์ **กีฬา not ฟุตบอล** จะหมายถึง การค้นหาเว็บเพจที่ปรากฏคำว่า "กีฬา" แต่ต้องไม่ปรากฏคำว่า "ฟุตบอล

ตัวอย่างที่ 7.4 :

พิมพ์ **วัด near อยุธยา** หมายถึง การค้นหาเว็บเพจที่มีทั้งคำว่า วัด และ อยุธยา อยู่ในหน้า เว็บเพจเดียวกัน และคำทั้งสองน่าจะปรากฏอยู่ใกล้เคียงกัน

การค้นหาโดยใช้เงื่อนไข "NEAR" หมายถึง เป็นการระบุให้ผลลัพธ์ของการค้นหาต้องปรากฏทั้ง A และ B และทั้งสองคำนี้จะต้องปรากฏอยู่ใกล้ๆกัน รูปแบบการค้นหาแบบนี้จะคล้ายกับการใช้เงื่อนไข "AND" แต่ต่างกันเพียง คำทั้งสองจะต้องปรากฏอยู่ห่างกันไม่เกิน 10 คำ ซึ่งเราจะเห็นว่าการใช้เงื่อนไข NEAR จะมีประสิทธิภาพที่ดีกว่าการใช้เงื่อนไข "AND" ในกรณีที่คำทั้งสองมีความเกี่ยวข้องกัน โดยคาดหวังว่าคำทั้งสองน่าจะปรากฏอยู่ใกล้เคียงกัน

ตัวอย่างที่ 7.5 :

พิมพ์ **(การเมือง or เศรษฐกิจ) near รัฐสภา** หมายถึง การสั่งให้ค้นหาหน้าเอกสารเว็บเพจที่ปรากฏคำว่า "การเมือง" หรือ "เศรษฐกิจ" และ จะต้องปรากฏอยู่ใกล้เคียงกับคำว่า "รัฐสภา" ด้วย

การค้นหาโดยใช้เครื่องหมายวงเล็บ "( )" การใช้เครื่องหมายวงเล็บคร่อมข้อความที่เป็นเงื่อนไข หมายถึง การเจาะจงให้ประมวลผลข้อความที่อยู่ภายในวงเล็บ

ส่วนที่2.เสียงคล้าย เป็นช่องระบุว่าต้องการคำที่ออกเสียงคล้ายคลึงกันได้

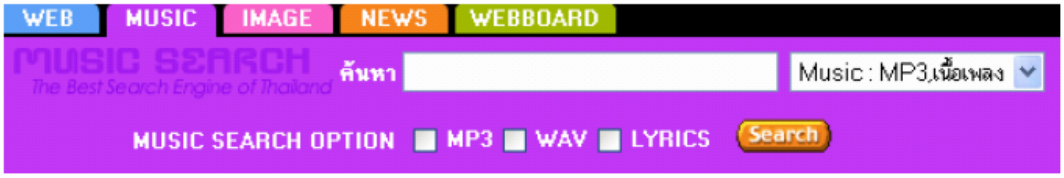
ส่วนที่3.คำแนะนำพร้อมตัวอย่างการใช้งาน เป็นข้อความที่อยู่ภายใต้ช่องค้นหาเพื่อแนะนำการใช้งาน Search Engine อย่างง่าย พร้อมตัวอย่างการใช้งาน

ส่วนที่4. ปุ่ม "Search" ปุ่มสำหรับสั่งให้ทำการค้นหา

การค้นหาเพลงด้วย Music Search

Music Search เป็นเครื่องมือในการค้นหาเนื้อร้อง, เพลง mp3 ทั้ง เพลงไทย, เพลงสากล, เพลงญี่ปุ่น และ เพลงประกอบภาพยนตร์ **(Sound track)** โดย **Music Search** ของ **SiamGURU** แบ่งรูปแบบการค้นหาเป็น 2 รูปแแบบคือ

1. **ค้นหาแบบธรรมดา** เป็นการค้นหาที่ง่ายต่อการใช้งาน เพราะ เราเพียงแต่ทราบข้อมูลเกี่ยวกับเพลงเพียงเล็กน้อย เราก็สามารถค้นหาเนื้อร้อง หรือ เพลง mp3 เหล่านั้นได้



รูปที่ 7.22 : แสดงการค้นหาเพลงแบบธรรมดา

การค้นหาแบบธรรมดา ประกอบด้วย

- **ช่องค้นหาแบบธรรมดา** เป็นช่องที่คุณไว้สำหรับพิมพ์ข้อมูลทุกอย่างที่เกี่ยวกับเพลง เช่น ชื่อ ศิลปินเพลง, ชื่อเพลง, ชื่ออัลบั้ม, ชื่อภาพยนตร์ หรือ ชื่อโฆษณา หรือ ชื่อละคร ที่ใช้เพลงนี้ประกอบ เช่น เพลงประกอบภาพยนตร์เรื่อง **Matrix** เราพิมพ์ **Matrix** ลงในช่องค้นหา
- **คำแนะนำพร้อมตัวอย่างการใช้งาน (Example)** เป็นข้อความที่อยู่ภายใต้ช่องค้นหาแบบธรรมดา เพื่อแนะนำการใช้งาน Music Search อย่างง่าย
- **ปุ่ม "Search" (Search Button)** ปุ่มสำหรับสั่งให้เริ่มทำการค้นหา

ในการค้นหาแบบธรรมดานั้นผู้ใช้อาจได้สารสนเทศไม่ตรงกับความต้องการ หากเราทราบข้อมูลเกี่ยวกับเพลงนั้นอย่างแน่ชัด เช่น ทราบว่าเป็นอัลบั้มใด ,เพลงชื่ออะไร ,ทราบชื่อศิลปิน อย่างแน่ชัด แนะนำให้ใช้งานในส่วนของ Advanced Search จะดีกว่า เพราะจะได้ข้อมูลที่ตรงกับความต้องการของผู้ใช้งานมากที่สุด

2. **ค้นหาแบบเจาะจง (Advanced Search)** เป็นการค้นหาเนื้อเพลง หรือ เพลง mp3 ในกรณีที่ เราทราบข้อมูล เกี่ยวกับศิลปิน (Artist), ชื่ออัลบั้ม(Album), ชื่อเพลง (Song) หรือ เนื้อร้อง (Lyrics) โดยการค้นหาแบบนี้จะทำให้เราได้ผลลัพธ์ที่รวดเร็ว และ ตรงความต้องการของผู้ใช้งานมากยิ่งขึ้น

รูปที่ 7.23 : แสดงการค้นหาเพลงแบบเจาะจง

ตามรูปที่ 7.23 สำหรับการค้นหาแบบเจาะจง จะประกอบไปด้วย

• **ศิลปิน** เป็นช่องสำหรับกรอกข้อมูลศิลปินเพลงที่เราทราบลงไป เช่น "Backstreet Boys" ,"Barry Manilow" ,"aztec camera"," เสาวลักษณ์ ลีละบุตร"

• **อัลบั้ม** เป็นช่องสำหรับกรอกข้อมูลอัลบั้มของเพลง เช่น "Baby One More Time" , "OST from The Nutty Professor" (เพลงประกอบภาพยนตร์

• **ชื่อเพลง** เป็นช่องสำหรับกรอกข้อมูลที่เป็นชื่อเพลง เช่น "I Wanna Be With You", "Ano Tsuku Kotoba","Ano Tsuku Kotoba", "น้องไก่"

• **เนื้อร้อง** เป็นช่องสำหรับกรอกข้อมูลที่เป็นเนื้อเพลง เพียงพิมพ์เนื้อเพลงบางท่อนลงไป เช่น "แค่กระจกใสบางๆ ที่กั้นตรงกลางเท่านั้น ได้แค่เพียงเห็นหน้ากัน", "Every little thing I do Never seems enough for you You" เป็นต้น

• **ปุ่ม "Search" (Search Button)** ปุ่มสำหรับสั่งให้เริ่มทำการค้นหา

สิ่งที่ต้องระวังในการค้นหานั่นคือ ชื่อศิลปิน (Artist) , ชื่ออัลบั้ม (Album), ชื่อเพลง (Song), เนื้อร้อง (Lyrics) ที่มีช่องว่างระหว่างคำ (space) จะต้องระวังในเรื่องการใส่เครื่องหมายฟันหนู " " คร่อมข้อความเหล่านั้นด้วย เช่น "Yesterday once more", "เจตริน วรรธนะสิน", "เจ็บจนเริ่มชา อยากจะลาก็คงไม่ว่ากัน ถามคำเดียว เพราะฉันทำผิดไป หรือ เพราะเธอมีใครที่ต้องการ" และพยายามระบุข้อมูลเกี่ยวกับเพลงที่เราทราบให้มากที่สุด เพราะยิ่งเราบอกข้อมูลเกี่ยวกับเพลงได้มากเท่าใด จะทำให้เราสามารถค้นหาเนื้อเพลง เพลง mp3 ได้ง่าย และตรงความต้องการมากขึ้นเท่านั้น เช่นเราต้องการค้นหาเพลงของ Westlife ชื่อเพลง Fool Again ให้เราใส่ Westlife ในช่องศิลปิน และ "Fool Again" ในช่องเพลง เป็นต้น และ หากต้องการเนื้อร้อง หรือ เพลง mp3 ทั้งหมดของเพลงประกอบภาพยนตร์ หรือ ละคร ให้พิมพ์ชื่อภาพยนตร์ หรือ ละคร ในช่องอัลบั้ม เช่น ต้องการหา เพลงประกอบละครเรื่องเก้าอี้ขาวในห้องแดง ทั้งหมด ให้พิมพ์ เก้าอี้ขาวในห้องแดง ในช่องอัลบั้ม

การค้นหาภาพด้วย Images Search

Images Search เป็นเครื่องมือค้นหาข้อมูลในอินเทอร์เน็ตอีกประเภทหนึ่ง เหมาะสำหรับผู้ที่คุ้นเคยกับ Basic Search อยู่แล้ว แต่ต้องการค้นหาข้อมูล ให้ได้ผลลัพธ์ตรงความต้องการมากขึ้นกว่า ที่จะสามารถทำได้ใน Basic Search ด้วยวิธีการสร้างเงื่อนไขการค้นหาขึ้น ซึ่งจะได้ผลลัพธ์ที่น่าพอใจกว่าใน Basic Search ในขณะเดียวกันการค้นหาแบบ Super Search ก็จะมีความซับซ้อนในการใช้งานด้วยเช่นกัน

รูปที่ 7.24 : แสดงการค้นหาภาพด้วย Image Search

จากรูปที่ 7.23 : แสดงการค้นหาเพลงแบบเจาะจง ประกอบไปด้วย

- **ช่องค้นหารูปภาพ** เป็นช่องสำหรับใส่คำที่ต้องการค้นหา เช่น ชื่อรูปภาพ หรือบางส่วนของชื่อ คำที่เกี่ยวข้องกับรูปภาพนั้น
- **ตัวอย่างการใช้งานอย่างง่าย ๆ** เป็นข้อความที่อยู่ภายใต้ช่องค้นหารูปภาพ เพื่อแนะนำการใช้งาน Image Search อย่างง่าย พร้อมตัวอย่างการใช้งานอย่างสั้นๆ
- **ปุ่ม "Search"** ปุ่มสำหรับสั่งให้ทำการค้นหา

เทคนิคการค้นหารูปภาพแบบต่าง ๆ เราสามารถค้นหารูปภาพที่ต้องการโดยใช้เงื่อนไข เพื่อเป็นการเจาะจงผลลัพธ์ โดย

การใช้เครื่องหมายบวก "+" หมายถึง การระบุเงื่อนไข "และ" นั่นคือ หากเราต้องการให้ได้ภาพตรงความต้องการเรามากที่สุด เราจะต้องใส่เครื่องหมายบวกนำหน้าคำหลักที่เราใช้ในการค้นหา โดยเครื่องหมายบวกจะต้องเขียนติดกับคำหลัก เช่น เราพิมพ์ +นิโคล +เทริโอ หมายถึง ค้นหารูปภาพที่เป็นนิโคล เทริโอ เท่านั้น

การใช้เครื่องหมายลบ "-" หมายถึง การระบุเงื่อนไข "ต้องไม่ใช่" นั่นคือ เราจะใส่เครื่องหมายลบติดกับคำหลักที่เราไม่ต้องการให้ค้นหารูปภาพนั้นออกมา เช่น นิโคล -เทริโอ หมายถึง ค้นหารูปภาพ นิโคล แต่ไม่ใช่ "นิโคล เทริโอ" ซึ่งผลลัพธ์อาจจะออกมาเป็น นิโคล คนอื่นๆ เช่น นิโคล คิดแมน

นอกจากนี้ถ้าต้องการค้นหาให้ได้ผลลัพธ์ทีละมากๆ กระทำได้ โดยการเว้นช่องว่างระหว่างคำ เช่น นิโคล มอส ทาทา ก็จะได้ผลลัพธ์รูปภาพที่มีนิโคล หรือ มอส หรือ ทาทา จะเห็นได้ว่า Advance Search เป็นการช่วยในการ Search ข้อมูล เพราะสามารถช่วยในการบีบประเด็นหัวข้อให้แคบลง ซึ่งทำให้ได้รายชื่อเว็บไซต์ที่ตรงกับความต้องการในการ Search มากขึ้น

สรุปได้ว่า ในการใช้ Search engine จะให้เกิดประโยชน์อย่างเต็มที่ ผู้ใช้จำเป็นจักต้องศึกษา ถึงวิธีการใช้ Search แต่ละวิธี ไม่ว่าจะเป็น Keyword Index , Subject Directories, Metadata Search engine นอกจากนี้ยังต้องศึกษาถึง การใช้ tool วิธีการใช้ Advance Search เพื่อกำหนดขอบเขตการ Search ข้อมูลให้แคบลง จักได้ Page ที่ถูกต้องและตรงตามความต้องการมากที่สุด

7.5 Retrieval Effectiveness Assessment

การประเมินประสิทธิผลของการค้นคืนสารสนเทศนั้น ที่นิยมใช้กันอยู่นั้นมี Search Tools และ Search Service ที่ใช้อยู่ 2 Query ด้วยกันโดยเปรียบเทียบการค้นคืนของคำให้ได้จำนวนของเอกสารที่ทำการค้นคืน ได้แก่ Latex Software และ Multi agent systems Architecture ดังรายละเอียดต่อไปนี้

7.5.1 Latex Software

เป็น Software ที่เป็นการรวมชุดคำสั่งและเป็น Software เสรีที่ผู้ใช้สามารถ Download ตัวโปรแกรมและรหัสต้นฉบับได้โดยไม่ต้องเสียค่าใช้จ่ายใด ๆ ซึ่งสามารถ Download ได้ที่ http://ww.tug.org และสามารถใช้ได้กับคอมพิวเตอร์ทุกแพลตฟอร์ม (Platform) ไม่ว่าจะเป็น Windows , Mac , Unix ฯลฯ ซึ่งไฟล์ของ Latex จะเป็นแบบลักษณะ WYSIWYG (*What You See Is What You Get*) คือเป็นไฟล์ที่อ่านได้ (Human readable) ซึ่งต่างจาก Word processor เพราะว่าถ้าไม่มีโปรแกรมโปรแกรมเวิร์ดโปรเซสเซอร์ก็จะไม่สามารถดูเนื้อหาได้

การเขียนสูตรคณิตศาสตร์ในโปรแกรม Latex สามารถเขียนได้ง่าย พร้อมทั้งสามารถแปลงไฟล์เป็น PDF (Portable Document Format) การใช้ภาษาไทยกับ Software Latex ยังไม่ค่อยเป็นที่นิยมใช้เนื่องจากส่วนใหญ่มักจะใช้โปรแกรมเวิร์ดโปรเซสเซอร์เมื่อเทียบกับ Software Latex ที่ใช้ยากว่า

คำเชื่อมบางครั้งถูกนำมาใช้กับลักษณะของการปรากฏของคำใน Indexing คำสองคำที่ถูกเชื่อมถ้ามันปรากฏอยู่ด้วยกันที่มีความสัมพันธ์เชิงความหมายบางอย่าง คำใน Indexing จะปรากฏอยู่ในบทบาทหนึ่งบทบาทหรือมากกว่า 1 บทบาท เป็นการแสดงหน้าที่หรือการใช้

จากตารางที่ 1 : แสดงผลเปรียบเทียบของผลลัพธ์ของ Search Engine ต่างๆที่ได้จากการร้องขอสารสนเทศโดยใช้ latex software

- ในคอลัมน์ที่ 1 เป็นการแสดง Search Tools ที่ใช้
- คอลัมน์ที่สองเป็นการใช้ Disjunctive Query คือ จะมีการเลือกเฉพาะ Query ที่มีการใช้ตรรกะในลักษณะของการเชื่อมด้วย " (OR) หรือ "
- คอลัมน์ที่สามเป็นการใช้ Conjunctive Query คือ จะมีการเลือกเฉพาะ Query ที่มีการใช้ตรรกะในลักษณะของการเชื่อมด้วย " (AND) และ "

- คอลัมน์สุดท้ายจะเป็นการแสดงถึงจำนวนของเอกสารทั้งหมดที่ทำการค้นคืนโดยการอธิบายแบบการใช้ Phrase Query ในการ Retrieval ข้อูมล คือ จะมีการเลือกเฉพาะ Query ที่เป็น วลีเท่านั้น
- ส่วน N / A (Not Available) ในตาราง หมายถึงไม่มีข้อมูล

Table 1. Comparison of number of results for "latex software" query.

Search tool/ service	Disjunctive query	Conjunctive query	Phrase query
AltaVista	200,000	30,000	100
Excite	134,669	29,287	29,287
HotBot	3,696,449	61,830	17,630
InfoSeek Guide	3,111,835	427	100
Lycos	29,881	26	N/A
OpenText	481,846	2,541	6
WebCrawler	158,751	864	6
WWW Worm	4,999	2	N/A
Galaxy	6,351	20	N/A
Magellan	17,658	17,658	N/A
Yahoo	373 categories 18,344 sites	1 category 3 sites	N/A 101 sites
IBM InfoMarket	100	N/A	N/A
MetaCrawler	29	32	34

เครื่องมือที่ใช้ในการสืบค้น (Search Tools) เช่น Search Engine ของ Info seek ซึ่งจากตารางจะพบว่า info seek มีจำนวนเอกสารที่ทำการค้นคืนมาได้มากกว่า 3 ล้านเอกสาร ส่วน WWWW ทำการค้นคืนได้น้อยที่สุดเป็นจำนวน 4,999 เอกสาร

โดยทั่วไปจำนวนเอกสารที่ทำการค้นคืน จาก Disjunction ไปยัง Conjunction และไปยัง Pharse Query จะมีจำนวนที่ลดลงเรื่อย ๆ หรือในกรณีของ Excite จะมีจำนวนของเอกสารที่ทำการค้นคืนมาได้นั้นของConjunctive และ Phares Query มีจำนวนเท่ากัน

7.5.2 Multi agent systems Architecture

Table 2. Results for "multiagent system architecture" query.

Search tool/ service	Disjunctive query	Conjunctive query	Phrase query
AltaVista	40,000,000	700	6
Excite	514,321	3	561
HotBot	18821	383	0
InfoSeek Guide	10,960,353	79	38
Lycos	53,525	0	N/A
OpenText	604,487	33	0
WebCrawler	150,619	0	0
WWW Worm	2,000	0	N/A
Galaxy	797	0	N/A
Magellan	58,080	84	N/A
Yahoo	615 categories	0	0
	23,560 sites	5,890 sites	6 sites
IBM InfoMarket	100	N/A	N/A
MetaCrawler	38	29	6

จากตารางที่ 2 เป็นการนำเสนอโดยการใช้ Multiagent systems Architecture ยกเว้นสำหรับ WWWW และ Galaxy เพราะว่าเป็น Literature ซึ่ง Conjunction และ phrase Query จะให้ผลที่ดีกว่าแบบ Disjunction Query ซึ่ง Query ส่วนใหญ่ยากในการประเมินและทำให้เห็น Query ในหลายด้านและผลของการใช้ Top Rank ของเอกสารสามารถแสดงคุณสมบัติของการใช้ Multiagent systems Architecture

จากวิธีการทั้งสองจะพบว่า การทดสอบ query , ความเกี่ยวข้องกันของเอกสารและความไม่เกี่ยวข้องกันของเอกสาร ในการจัดอันดับของ query ซึ่งการค้นคืนของ User ไม่สามารถทดสอบได้จาก 2-3 เอกสารเท่านั้นที่จะให้ค่า Top Rank อย่างไรก็ตามจำนวนของเอกสารในการแสดงผล Rank query ในจำนวนหนึ่งพัน

7.6 Improving Retrieval Effectiveness

การพัฒนาการค้นคืนให้มีประสิทธิภาพนั้นสิ่งที่สำคัญคือการออกแบบและการพัฒนาการใช้ Web search Tools โดยการใช้ Query ที่ focusedหรือเฉพาะเจาะจง นอกจากนี้ยังขึ้นอยู่กับความเร็วของการทำงานและขนาดของ database อีกด้วย สามารถสรุปถึงวิธีการในการเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศได้หลายวิธีการดังนี้

1. Relevance Feedback Techniques

การใช้เทคนิคการป้อนกลับความเกี่ยวข้องของสารสนเทศ โดยจะมีการประเมินของตัวอย่างเอกสารที่ถูกค้นคืนออกมา และการประเมินนี้จะถูกนำไปใช้เพื่อแก้ไขกระบวนการค้นคืน ซึ่งเป็นสิ่งที่ผู้ใช้กำหนดหลังจากทีมีการ Retrieval มาแล้วโดยมีการพิจารณาว่าข้อมูลที่ได้ถูก Retrieval ขึ้นมานั้นมีความตรงใจมากน้อยแค่ไหน โดยมีการนำผลลัพท์อันแรกมาทำการ Revise ใหม่แล้วทำการ search ใหม่อีกครั้งเป็นการกระทำซ้ำจนกระทั่งได้ผลลัพธ์เป็นที่น่าพอใจโดยทั่วไป Relevance Feedback ไม่ค่อยนิยมใช้ เพราะว่าUser ไม่ต้องการให้มา Feedback เอกสารใหม่เพราะว่าต้องมาพิจารณาเอกสารใหม่เสียเวลาของ User และบางครั้งการให้ Feed back ได้ยากเพราะว่าบางครั้งเกิดจากความรู้สึก

2. Modifying The query representation

การแก้ไขตัวแทนข้อสอบถาม ในกรณีที่ผลลัพธ์ที่ได้ไม่ตรงความต้องการ เราสามารถข้อสอบถามใหม่โดยวิธีการที่เรียกว่า

2.1 Query Expression โดยมีการ Add Key Word ใหม่เพื่อให้ Query นั้นไปหาเอกสารอีกครั้งหนึ่งซึ่งน่าจะตรงกว่าครั้งแรก

2.2 Term Reweighing มีการปรับ Weight ของคำใน Key Word และลด Weight ของเทอมบางเทอมในเอกสารที่ไม่สำคัญ อาจใช้วิธีการที่เรียกว่า **Query Splitting** เป็นการแบ่งเอกสารออกเป็น Cluster โดยจะมีการแยก Query ออกเป็น Sub query ซึ่งแต่ละ Sub query จะหมายถึง 1 Cluster ในกลุ่มของ Positive Feedback ซึ่งน้ำหนักของแต่ละ term ใน Sub query สามารถปรับจัดหรือขยาย ไปยังสองวิธีการก่อนหน้า

3. Pseudo Feedback จะเป็นการพยายามปรับปรุงจาก Relevance Feed back ที่ User ให้การปรับ Weight โดยการเอาผลลัพธ์ที่ได้จากการ Retrieval ครั้งแรกมา สร้าง Query ใหม่โดยการเทียบเคียงจากเอกสารที่สัมพันธ์กันซึ่งฐานข้อมูล Pseudo Feedback ใช้ในการเพิ่มประสิทธิภาพของฐานข้อมูล TREC

4. **Modifying The Document representation** มีการปรับจัด Document Vector ในลักษณะของ Relevance Feedback โดยมีการอ้างถึง User – Oriented Cluster ซึ่งเป็นการปรับปรุงโดยการปรับค่าน้ำหนักของการค้นคืนและเอกสารที่เกี่ยวข้อง

5. Agent – Base Filtering and Routing

7.7 การคำนวณ PageRank

วิธีการในการวิเคราะห์ลิงค์ที่ใช้โดย Google (Brin & Page, 1998) นั้น ซึ่งลำดับของเพจนั้นขึ้นกับ authority ซึ่งประยุกต์ไปยังเวปที่เป็นเพื่อนบ้านภายใน local neighborhood of pages) ที่ล้อมรอบผลลัพธ์ของข้อสอบถาม(query) เริ่มต้นของ Google นั้น จะจัดลำดับของเพจโดยใช้การเชื่อมโยงโครงสร้างไปยังเวป ค้นหาผลลัพธ์ของการค้นหาการจัดลำดับของคีย์เวิร์ดในโครงสร้างฐานข้อมูล โดยจะมีการจัดลำดับโดยการรวมกับลำดับของเพจของบุคคล (Individual PageRanks) ในแง่ของเหตุการณ์ของเทอมบนเพจ(Occurrences of a term on a page) , ตำแหน่งของคียเวิร์ด(Position of the keywords) หรือ ขนาดรูปแบบของเทอม(Font size of the terms)

แนวคิดในการจัดลำดับเพจนั้นมีดังนี้

1. เริ่มจากใช้โครงสร้างของการเชื่อมโยง(link structure) ของเวป ไปคำนวณหา quality ranking (PageRank) ของแต่ละเวปเพจ
2. ใช้วิธี Citation counting a metric สำหรับวัดคุณภาพของเพจหรือเอกสาร
3. ในการหาค่าของ PageRank จะไม่อ้างเหตุผลประกอบที่ผิด(PageRank a more sophisticated citation counting method), ไม่โอนเอียงในการจัดการ(not prone to manipulation).
4. แต่ละเพจจะมีค่าของ PageRank เพียงค่าเดียวและเป็นอิสระกับคีย์เวิรด์ของข้อสอบถาม(Each page has unique PageRank, independent of keyword query)

5. PageRank จะไม่แสดงความคล้ายคลึงกันของหน้าไปยังข้อสอบถาม(PageRank does NOT express relevance of page to query)
6. คำนวณค่าที่เกิดขึ้นโดยสัญชาติญาณ(Calculation Intuition) :โดยค่าของ PageRank ของ เพจ P เมื่อมีเพจที่มี PageRank ที่มีขนาดใหญ่กว่าชี้มาที่จุด P (PageRank of page P increases when pages with large PageRanks point to P.)
7. ลำดับของเพจมีโอกาสเท่ากันในการกระจายไปยังเพจต่างๆในการเชื่อมโยงไปข้างหน้า (The rank of a page is evenly distributed among its forward links.)

PageRank(PR) คือจำนวนในการเยี่ยมไปยังแต่ละเพจ

สูตรในการคำนวณเป็นดังนี้

PR(A) = (1-d) + d*(PR(T1)/C(T1)+…+ PR(Tn)/C(Tn))

โดย

PR(A) คือ PageRank หน้า A

PR(Ti) คือ PageRank ของหน้า Ti ที่เชื่อมโยงกับหน้า A

C(Ti) คือ จำนวนของการเชื่อมโยงออกจากเพจ Ti)
(the number of outbound links on page Ti)

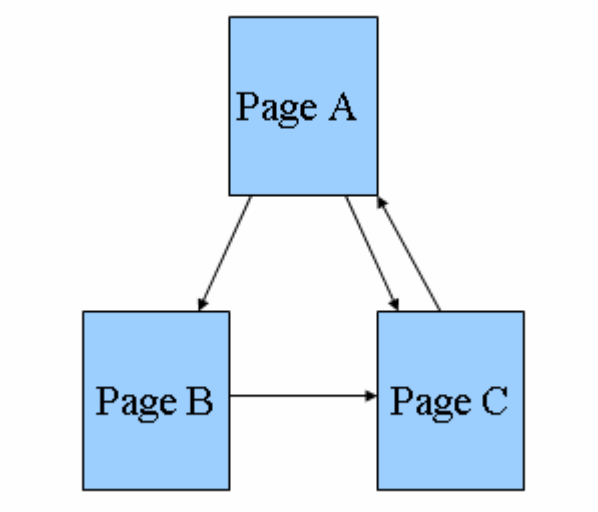
T1, …, Tn: เป็นหน้าที่มีจุดเชื่อมโยงไปยังหน้า A(pages pointing to page A)

d คือ ค่าหน่วงที่มีค่าอยู่ระหว่าง 0 ถึง 1 ค่ามาตรฐานจะกำหนดที่ 0.85 โดยจะใช้สำหรับนับ PageRank ที่จม

(a damping factor which can be set between 0 and 1. normally this is set to 0.85)

Note: d counts for PageRank sinks

ตัวอย่างที่ 7.6 สมมุติว่า Web site หนึ่งมีการเชื่อมโยงเพจดังรูปต่อไปนี้



สมมุติว่า ค่าของ d มีค่าเท่ากับ 0.5

แทนสูตรในการคำนวณจาก

PR(A) = (1-d) + d*(PR(T1)/C(T1)+…+ PR(Tn)/C(Tn))

ดังนั้น

PR(A) = 0.5 + 0.5 PR(C)

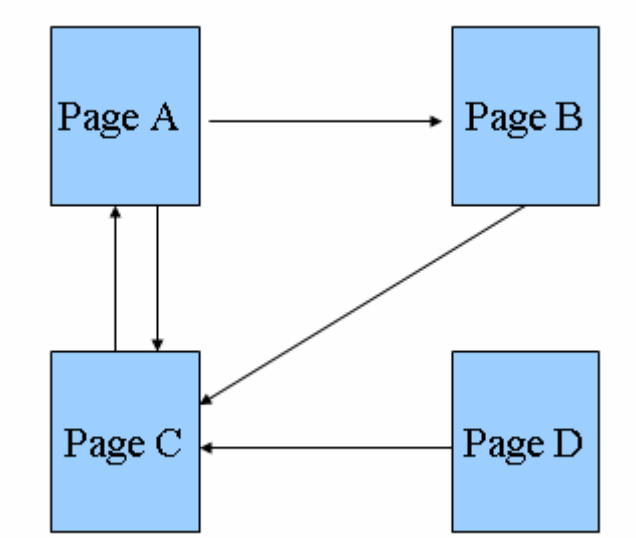
PR(B) = 0.5 + 0.5 (PR(A) / 2)

PR(C) = 0.5 + 0.5 (PR(A) / 2 + PR(B))

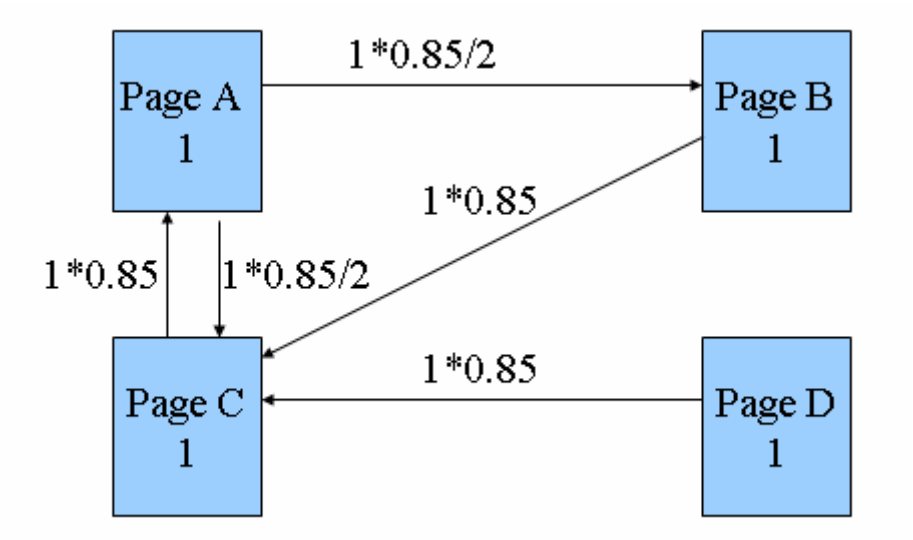
Iteration	PR(A)	PR(B)	PR(C)
0	1	1	1
1	1	0.75	1.125
2	1.0625	0.765625	1.1484375
3	1.07421875	0.76855469	1.15283203
4	1.07641602	0.76910400	1.15365601
5	1.07682800	0.76920700	1.15381050
6	1.07690525	0.76922631	1.15383947
7	1.07691973	0.76922993	1.15384490
8	1.07692245	0.76923061	1.15384592

9	1.07692296	0.76923074	1.15384611
10	1.07692305	0.76923076	1.15384615
11	1.07692307	0.76923077	1.15384615
12	1.07692308	0.76923077	1.15384615

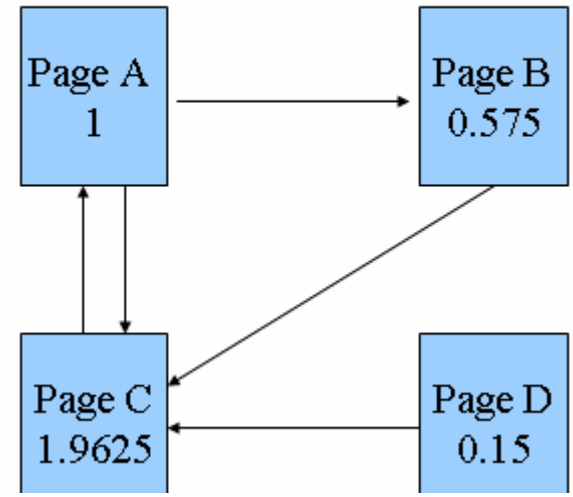
ตัวอย่างที่ 7.7 Web site หนึ่งมีการเชื่อมโยงเพจดังรูปต่อไปนี้



รอบที่ 1 สมมุติว่า d = 0.85



รอบที่ 2



- **Each page has not passed on 0.15, so we get:**

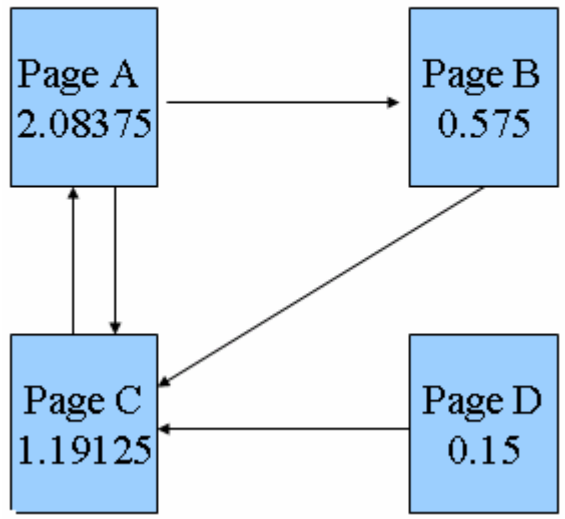
Page A: 0.85 (from Page C) + 0.15 (not transferred) = 1

Page B: 0.425 (from Page A) + 0.15 (not transferred) = 0.575

Page C: 0.85 (from Page D) + 0.85 (from Page B) + 0.425 (from Page A) + 0.15 (not transferred) = 1.9625

Page D: receives none, but has not transferred 0.15 = 0.15

รอบที่ 3



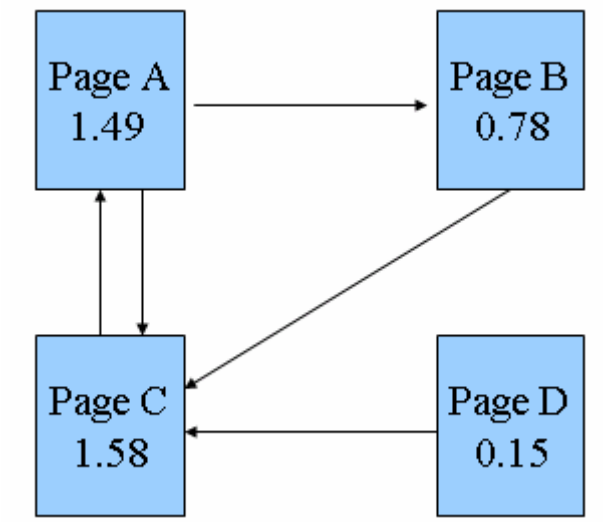
Page A: 1.9625*0.85 (from Page C) + 0.15 (not transferred) = 1.818125

Page B: 1*0.85/2 (from Page A) + 0.15 (not transferred) = 0.575

Page C: 0.15*0.85 (from Page D) + 0.575*0.85(from Page B) + 1*0.85/2 (from Page A) +0.15 (not transferred) = 1.19125

Page D: receives none, but has not transferred 0.15 = 0.15

รอบสุดท้าย



ผลจากการคำนวณทำให้ทราบว่า Page C มีค่าของ PageRank มากที่สุด และ Page A มีค่ารองลงมา ซึ่งเราสรุปได้ว่า Page C มีความสำคัญมากที่สุดในกราฟเพจนี้ และการที่มีการคำนวณหลายรอบนั้นเพื่อชักจูงให้มาบรรจบกันของ PageRank นั่นเอง

แบบฝึกหัด

1. จงอธิบายถึง Traversing The Web มาพอเข้าใจ
2. จงอธิบายถึง การทำงานของ Search Engine มาพอเข้าใจ
3. จงอธิบายถึง Taxonomy For Search Tools And Service มาพอเข้าใจ
4. จงอธิบายถึง วิธีการค้นหาข้อมูลด้วย search engine ที่ทำให้ค้นหาข้อมูลให้ง่ายขึ้น
5. จงอธิบายถึง Retrieval Effectiveness Assessment มาพอเข้าใจ
6. จงอธิบายถึงการจัดลำดับของเพจ(PageRank) มาพอเข้าใจ

บรรณานุกรม

Avi Rappoport ," **How Search Engines Work: A Technology Overview ",** Search Tools Consulting, UC Berkeley SIMS class 202 , September 16, 2004 , www.searchtools.com ,consult1@searchtools.com

Prof. Ray Larson & Prof. Marc Davis , **"Information Organization and Retrieval :** Web Search Issues and Algorithms **"**,UC Berkeley SIMS , http://www.sims.berkeley.edu/ academics/ courses/is202/f04/

Sirak Kaewjamnong, **Web Spidering and Web Link Analysis**

Steve Lawrence And C.Lee Giles ,"**Context and Page Analysis For Improve Web Search**" ,http://clgiles.ist.psu.edu/papers/IEEE.IC.search-web.pdf

http://www.elib-online.com/computers/internet_search2.html

Improving Web searching with user preferences Web Search Your Way

www.google.co.th

www.siamguru.com