



การออกแบบและพัฒนาเว็บแอปพลิเคชัน  
สำหรับการค้นหาข้อมูลกองทุนรวม

ณัฐพัฒน์ นนทชัย

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2555

**A Design and Development of Web Application for  
Mutual Fund Searching**

**Nathapat Nontachai**

เลขทะเบียน.....	0223899.....
วันลงทะเบียน.....	12 มี.ค. 2556.....
เลขเรียกหนังสือ.....	006.76.....
	๗๖31๕๗
	[๒๕๕๕]

**Thematic Paper Submitted in Partial Fulfillment of  
the Requirements for the Degree of  
Master of Science in Web Engineering  
Faculty of Information Technology, Dhurakij Pundit University**

**2012**



## ใบรับรองสารนิพนธ์

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิทยาศาสตรมหาบัณฑิต


หัวข้อสารนิพนธ์ การออกแบบและพัฒนาเว็บแอปพลิเคชันสำหรับการค้นหากองทุนรวม


เสนอโดย ณัฐพัฒน์ นนทชัย

สาขาวิชา วิศวกรรมเว็บ

อาจารย์ที่ปรึกษาสารนิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา


ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

  
.....ประธานกรรมการ  
(รองศาสตราจารย์ ดร.นุชรี เปรมชัยสวัสดิ์)

  
.....กรรมการและอาจารย์ที่ปรึกษาสารนิพนธ์  
(ผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา)

  
.....กรรมการ  
(อาจารย์ ดร.ธนภัทร์ อนุศาสน์อมรกุล)

คณะเทคโนโลยีสารสนเทศรับรองแล้ว

  
.....คณบดีคณะเทคโนโลยีสารสนเทศ  
(รองศาสตราจารย์ ดร.นุชรี เปรมชัยสวัสดิ์)

วันที่ 14 .....เดือน มิถุนายน .....พ.ศ. ....2555....

## กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือจาก ผศ.ดร. วรสิทธิ์ ชูชัยวัฒนา อาจารย์ที่ปรึกษาสารนิพนธ์ ที่ได้กรุณาสละเวลาให้คำปรึกษา ให้คำแนะนำ ให้วิชาความรู้ด้าน วิศวกรรมเว็บ (Web Engineering) ด้านเครื่องมือค้นหา (Search Engine) ตลอดจนการทำตลาด บนเว็บเบื้องต้น (E-Commerce) ตรวจสอบแก้ไขข้อบกพร่องเพื่อปรับปรุงแก้ไขให้สารนิพนธ์ฉบับนี้ เสร็จเรียบร้อยสมบูรณ์

ขอขอบคุณ ผศ.ดร. วราพร จิระพันธุ์ทอง และ คณาจารย์ในคณะเทคโนโลยีสารสนเทศ สาขาวิศวกรรมเว็บทุกท่านที่ได้กรุณาสละเวลาให้คำปรึกษา ให้คำแนะนำตลอดจนให้ความรู้ด้านการเขียนโปรแกรมบนเว็บ

ขอขอบคุณ ผศ.อัมพร สุคันธวิช อาจารย์ที่ปรึกษาจากคณะสังคมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ที่กรุณาอบรมสั่งสอน ให้คำแนะนำ และเป็นกำลังใจให้ผมเสมอมา

ขอขอบคุณบิดามารดาผู้ให้กำเนิด และคุณยายที่มีความมุ่งมั่น ตั้งใจ อบรมสั่งสอนและดูแลผมเป็นอย่างดี รวมถึงกรุณาให้ทุนทรัพย์เพื่อสนับสนุนการศึกษา ค่าใช้จ่ายอื่นๆ และทุนในการลงทุนทำธุรกิจเขียนเว็บอิสระ เพื่อหารายได้พิเศษด้วยตนเองอีกด้วย

ขอขอบคุณป้าหอม ศ.พญ.เพ็ญแข ถิ่นศิลา ที่ได้มอบชีวิตใหม่ให้แก่ผม ท่านเป็นผู้เปลี่ยนผมจากเด็กออกทิสติกคนหนึ่งที่มีความสามารถ และไม่เป็นที่ยอมรับในสังคมให้กลายเป็นคนที่สามารถใช้ชีวิตได้เหมือนคนปกติที่มีความรู้ความสามารถคนหนึ่ง และเป็นที่ยอมรับในสังคมจนถึงทุกวันนี้ ถ้าไม่มีบุคคลท่านนี้ ก็คงไม่มีผมอยู่จนถึงทุกวันนี้ รวมถึงอาจารย์ในโครงการพิเศษของโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ทุกท่าน

ผู้วิจัยขอกราบขอบพระคุณทุกท่านที่กล่าวมาด้วยความเคารพอย่างสูงมา ณ โอกาสนี้

นายณัฐวัฒน์ นนทชัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ฅ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ซ
สารบัญภาพ.....	ฌ
บทที่	
1 บทนำ.....	1
1.1 ที่มาและความสำคัญของงาน.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ประโยชน์และผลที่คาดว่าจะได้รับ.....	2
1.4 ขอบเขตการศึกษา.....	2
1.5 นิยามศัพท์.....	2
2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ความรู้เบื้องต้นเกี่ยวกับกองทุนรวม.....	4
2.2 เครื่องมือสืบค้นข้อมูลทางเว็บ.....	5
2.3 Web crawler.....	7
2.4 การจัดทำดัชนีของเครื่องมือค้นหา.....	23
2.5 คำค้นหาเว็บ.....	36
3 วิธีการดำเนินการและเครื่องมือ.....	39
3.1 แผนเวลาดำเนินการ.....	39
3.2 การออกแบบฟังก์ชัน (Functional Design).....	41
3.3 เครื่องมือที่ใช้ในการพัฒนา.....	46
3.4 วิธีการเก็บรวบรวมข้อมูล.....	46
3.5 การวิเคราะห์ข้อมูล.....	47

## สารบัญ (ต่อ)

บทที่	หน้า
4 ผลลัพธ์ของการพัฒนาระบบ.....	48
4.1 ความเร็วในการทำงานของในส่วนของ Crawler.....	48
4.2 ความเร็วในการค้นหาข้อมูล.....	52
4.3 ระบบ MFS สำหรับค้นหาข้อมูลกองทุนรวม.....	54
4.4 การเปรียบเทียบผลลัพธ์การค้นหาข้อมูลระหว่าง MFS และ Settrade.com.....	56
5 สรุปผลการทดลอง.....	59
5.1 ความเร็วในการทำงานของโปรแกรม Crawler.....	59
5.2 ความเร็วในการค้นหาข้อมูล.....	60
5.3 ข้อเสนอแนะ.....	61
บรรณานุกรม.....	62
ภาคผนวก.....	67
ประวัติผู้เขียน.....	98

## สารบัญตาราง

ตารางที่	หน้า
2.1 ครอบงำผกผัน.....	26
2.2 ครอบงำก้าวหน้า.....	28
4.1 เปรียบเทียบความเร็วในการทำงานของโปรแกรม Crawler.....	51
4.2 เปรียบเทียบความเร็วในการค้นหาข้อมูลแต่ละคำ.....	52
4.3 เปรียบเทียบความเร็วในการค้นหาข้อมูลแต่ละหน้า.....	52
4.4 เปรียบเทียบความเร็วในการค้นหาข้อมูลเมื่อใส่เครื่องหมาย Quote.....	53
4.5 เปรียบเทียบระหว่าง Mutual Fund Searching กับ Settrade.com .....	56

สารบัญภาพ

ภาพที่	หน้า
2.1 สถาปัตยกรรมขั้นสูงของ Web Crawler มาตรฐาน.....	5
2.2 สูตรคำนวณความสดใหม่ของข้อมูล.....	12
2.3 สูตรคำนวณอายุของข้อมูล.....	13
2.4 โครงสร้างขั้นสูงของ Web Crawler พื้นฐาน.....	17
3.1 แผนภูมิแกนที่ สำหรับวางแผนการดำเนินงาน.....	41
3.2 แผนภาพแสดงการทำงานของ Web Crawler (ภายใน).....	42
3.3 แผนภาพแสดงการทำงานของ Web Crawler (ภายนอก).....	43
3.4 แผนภาพแสดงการทำงานของ Search Application.....	44
3.5 แผนภาพแสดงการทำงานของแสดงผลลัพธ์การค้นคืน.....	45
3.6 แผนภาพแสดงการทำงานของการค้นคืนข้อมูลผ่านอุปกรณ์มือถือ.....	46
4.1 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2551.....	49
4.2 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2552.....	49
4.3 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2553.....	49
4.4 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2554.....	50
4.5 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2555.....	50
4.6 หน้าแรกก่อนเติมคำที่ต้องการค้นหา.....	54
4.7 หน้าแรกระหว่างกำลังเติมคำที่ต้องการค้นหา.....	55
4.8 หน้าผลลัพธ์หลังค้นหาข้อมูล.....	55
4.9 หน้าผลลัพธ์ระหว่างกำลังเติมคำที่ต้องการค้นหาเพิ่มเติม.....	56



หัวข้อสารนิพนธ์	การออกแบบและพัฒนาเว็บแอปพลิเคชันสำหรับการค้นหาข้อมูลกองทุนรวม
ชื่อผู้เขียน	ณัฐพัฒน์ นนทชัย
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.วรสิทธิ์ ชูชัยวัฒนา
สาขา	วิศวกรรมเว็บ
ปีการศึกษา	2554

### บทคัดย่อ

สารนิพนธ์นี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนาเว็บแอปพลิเคชันสำหรับการค้นหาข้อมูลกองทุนรวม โดยระบบดังกล่าวจะมีส่วนประกอบที่สำคัญ 2 ส่วน ได้แก่ ส่วนของ Crawler ซึ่งทำหน้าที่ในการรวบรวมข้อมูลจากแหล่งของข้อมูลกองทุนรวมหลากหลายที่ และส่วนของโปรแกรมการค้นหา ซึ่งทำหน้าที่เปรียบเทียบคำค้นกับข้อมูลกองทุนรวม และแสดงผลลัพธ์ของการสืบค้น

ผู้วิจัยได้ทำการวัดประสิทธิภาพของการทำงานของระบบ โดยการวัดความเร็วในการทำงานของส่วนของ Crawler และส่วนของโปรแกรมการค้นหา รวมทั้งได้เปรียบเทียบผลลัพธ์ของการค้นคืนของระบบที่ได้พัฒนาขึ้น กับระบบที่มีอยู่ในปัจจุบัน ซึ่งได้แก่ settrade.com พบว่าระบบที่ได้พัฒนาขึ้นมีความสามารถแตกต่างจาก settrade.com อยู่หลายประเด็น เช่น ความสามารถในการค้นหาชื่อเต็มของกองทุนรวม และการให้ข้อมูลเกี่ยวกับราคาซื้อขาย เป็นต้น

Thematic Paper Title	A design and development of web application for mutual fund searching
Author	Nathapat Nonthchai
Thesis Advisor	Asst.Prof.Dr. Worasit Choochaiwattana
Department	Web Engineering
Academic Year	2011

### **ABSTRACT**

This study aims at designing and developing a web application for mutual fund searching. The application has two main components, which are a crawler and a search function. The crawler is responsible for retrieve and download mutual fund information from several sources, while the search function is responsible for comparing query with mutual fund information and displaying search results.

Evaluations of efficiency of the developed system were conducted. The speed of the crawler and search function were measured. The search results from the developed system were compared with the result from settrade.com. The result showed that the capabilities of the developed system were better than settrade.com in some aspects such as the capability of search mutual information with a full name and the details information about selling and buying price of the mutual fund.

## บทที่ 1

### บทนำ

#### 1.1 ที่มาและความสำคัญของงาน

ที่มาของผลงานนี้เกิดจากกองทุนเป็นหนึ่งในเครื่องมือทางการเงินของนักลงทุนที่มีการใช้งานเป็นจำนวนมาก กองทุนรวมในประเทศไทยมีรายชื่อกองทุนรวมเป็นจำนวนมาก ทำให้ยากแก่การค้นหาข้อมูลของกลุ่มนักลงทุน ผู้วิจัยได้ศึกษาทั้งหมดมาบูรณาการเข้าด้วยกัน เพื่อสร้างระบบ Mutual Fund Search Engine เพื่อเป็นนวัตกรรมใหม่ที่เป็นประโยชน์แก่ทุกคนในโลกไซเบอร์ ประกอบกับกองทุนรวมที่มีอยู่ในประเทศไทยก็มีจำนวนเพิ่มมากขึ้นทุกวันๆ ราวกับดอกเห็ดแล้วทำไมในประเทศไทยจึงไม่มี Mutual Fund Search Engine เพื่อช่วยเพิ่มความสะดวกในการค้นหาข้อมูลเกี่ยวกับกองทุนรวมให้ใช้งานง่ายด้ายแบบในต่างประเทศบ้าง เช่น เว็บไซต์ MutualFundsNet ซึ่งเป็นเว็บไซต์ Search Engine ข้อมูลกองทุนรวม ที่ได้รับความนิยมของประเทศสหรัฐอเมริกา (<http://www.mutualfundsnet.com/>) อีกแหล่งก็คือเว็บไซต์ MSN money ของบริษัท Microsoft (<http://moneycentral.msn.com/investor/finder/predefffunds.aspx>) เป็นต้น แน่แน่นอนว่าเว็บไซต์นี้สามารถค้นหาข้อมูลกองทุนรวมจากทุก ๆ รัฐภายในประเทศสหรัฐอเมริกาได้จริงอย่างแน่นอนอีกด้วย เว็บไซต์นี้ยังทำหน้าที่เป็น Search Directory ในตัวเองอีกด้วย เพราะกองทุนรวมมีหลายกองและมีหลายประเภทมาก ๆ แต่ละกองจะมีบริษัทหลักทรัพย์จัดการกองทุนที่ไม่เหมือนกันอีกด้วย ขณะที่ Mutual Fund Search Directory ของกองทุนรวมในประเทศไทยมีอยู่แล้ว ซึ่งก็คือ เว็บไซต์ Thaimutualfund ของสมาคมบริษัทจัดการลงทุน (<http://www.thaimutualfund.com>) กับเว็บไซต์ Settrade ของตลาดหลักทรัพย์แห่งประเทศไทย (<http://www.settrade.com>) แต่ในปัจจุบันประเทศไทยยังขาดแคลนบุคลากรในการพัฒนาเว็บไซต์ประเภท Mutual Fund Search Engine สำหรับการค้นหาข้อมูลเกี่ยวกับข้อมูลกองทุนรวม โดยเฉพาะให้เป็นรูปแบบในต่างประเทศอยู่ และแน่นอนว่าระบบจำเป็นต้องสามารถใช้งานได้ง่ายด้ายแบบ Search Engine ชื่อใดจากหลาย ๆ สำนักอีกด้วย เพียงแค่พิมพ์และกด Enter เท่านั้นคำตอบก็ออกมาอย่างรวดเร็วทันทีที่ต้องการอีกด้วย

## 1.2 วัตถุประสงค์ของการศึกษา

เพื่อออกแบบ และพัฒนาเว็บแอปพลิเคชันสำหรับค้นหาข้อมูลกองทุนรวม

## 1.3 ประโยชน์และผลที่คาดว่าจะได้รับ

1. โครงการนี้มีส่วนช่วยเป็นเครื่องมือให้บุคคลที่เป็นนักลงทุนหรือคนทั่วไปก็ได้ สามารถค้นหาข้อมูลกองทุนรวมทุกค่ายที่ต้องการได้ง่ายดายยิ่งขึ้น
2. เป็นแหล่งรวบรวมข้อมูลกองทุนรวม สำหรับผู้สนใจลงทุนในกองทุนรวม

## 1.4 ขอบเขตการศึกษา

1. ศึกษาข้อมูลหน่วยลงทุนกองทุนรวมที่มีอยู่ในปัจจุบันและย้อนหลังโดยแยกเป็นประเภทต่างๆ จากทุก ๆ ค่ายหลักทรัพย์จัดการกองทุน เช่น กองทุนตราสารหนี้ กองทุนหุ้น กองทุนรวมที่สามารถลดหย่อนภาษีเงินได้ เป็นต้น โดยเลือกจำกัดการศึกษาเฉพาะภายในประเทศไทย ตั้งแต่ พ.ศ. 2538 – 2554 เพื่อเป็นต้นแบบของผลงาน
2. ศึกษาผลการดำเนินงานกองทุนรวมที่มีอยู่ในปัจจุบันและย้อนหลังโดยแยกเป็นประเภทต่างๆ จากทุก ๆ ค่ายหลักทรัพย์จัดการกองทุน โดยการวิเคราะห์ทางเทคนิค เพื่อจัดลำดับกองทุนรวมแต่ละประเภท

## 1.5 นิยามศัพท์

ในวิทยานิพนธ์ฉบับนี้ได้ใช้นิยามศัพท์ที่ใช้ในการวิจัยไว้ ดังนี้

**Crawler** หมายถึง เครื่องมือที่ทำงานเสมือน browser พิเศษ โดยการท่องโลกไซเบอร์เป้าหมาย ซึ่งโปรแกรมนี้มีส่วนช่วยจัดเก็บข้อมูลเป้าหมายบนโลกไซเบอร์ที่ต้องการค้นหาข้อมูลเก็บเข้าสู่ฐานข้อมูลไปพร้อมกันอย่างต่อเนื่องอีกด้วย โปรแกรมนี้เป็นโปรแกรมอยู่เบื้องหลังการทำงานของระบบโปรแกรม Search Engine และยังเป็นส่วนหนึ่งของระบบโปรแกรม Search Engine อีกด้วย

**Database** หมายถึง เครื่องมือที่ทำหน้าที่จัดเก็บรวบรวมข้อมูลทางสถิติ ซึ่งโปรแกรมนี้มีโปรแกรมนี้เป็นโปรแกรมอยู่เบื้องหลังการทำงานของระบบโปรแกรม Search Engine และยังเป็นส่วนหนึ่งของระบบโปรแกรม Search Engine อีกด้วย

**กองทุนรวม** หมายถึง หนังสือราชการ คือ เอกสารที่เป็นหลักฐานในราชการ ได้แก่ หนังสือที่มีไปมาระหว่างส่วนราชการ หนังสือที่ส่วนราชการมีไปถึงหน่วยงานอื่นใดซึ่งมิใช่ ส่วนราชการ หรือที่มีไปถึงบุคคลภายนอก หนังสือที่หน่วยงานอื่นใดซึ่งมิใช่ส่วนราชการหรือที่

บุคคลภายนอกมีมาถึงส่วนราชการ เอกสารที่ทางราชการจัดทำขึ้นเพื่อเป็นหลักฐานราชการ เอกสารที่ทางราชการจัดทำขึ้นตามกฎหมาย ระเบียบ หรือข้อบังคับ

**หน่วยลงทุน** คือ ตราสารทางการเงินที่ออกเสนอขายโดย บลจ. เพื่อระดมเงินทุนจากประชาชนเข้าสู่กองทุนรวมที่ได้จัดตั้งขึ้น และยังใช้เป็นหลักฐานแสดงความเป็นเจ้าของในเงินลงทุนนั้นๆ โดยปกติจะมีมูลค่าที่ตราไว้เท่ากับ 10 บาทต่อหน่วยลงทุน

## บทที่ 2

### วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

เอกสารเป็นทรัพยากรที่สำคัญอย่างหนึ่งขององค์กร ดังนั้นเอกสารที่เกิดจากการดำเนินงานของหน่วยงานจึงเป็นหลักฐานการดำเนินการกิจและกิจกรรมของหน่วยงานสามารถใช้เป็นเอกสารอ้างอิงเมื่อมีปัญหาทางกฎหมาย การนำเทคโนโลยีสารสนเทศมาประยุกต์ใช้ในการบริหารจัดการเอกสารอิเล็กทรอนิกส์จึงเป็นที่นิยมใช้กันอย่างแพร่หลายเพราะมีประโยชน์หลายด้าน ได้แก่ ลดปัญหาการใช้กระดาษ ประหยัดงบประมาณ ลดขั้นตอนการติดต่อสื่อสารภายในองค์กร และผู้ใช้งานสามารถค้นหาเรียกดูเอกสารได้อย่างรวดเร็ว เพิ่มประสิทธิภาพการทำงาน สามารถบริหารจัดการผ่านระบบเครือข่ายอินเทอร์เน็ตและอินทราเน็ตได้ง่าย

#### 2.1 ความรู้เบื้องต้นเกี่ยวกับกองทุนรวม

กองทุนรวม คือ เครื่องมือในการลงทุน (investment vehicle) สำหรับผู้ลงทุนรายย่อยที่ประสงค์จะนำเงินมาลงทุนในตลาดเงินตลาดทุนแต่ติดขัดด้วยอุปสรรคหลายประการที่ทำให้การลงทุนด้วยตนเองไม่สามารถได้ผลลัพธ์ตามเป้าหมายที่ต้องการ (ธัญวงศ์ กิริตวานิชย์ และภัสรา ชวาลกร, 2547) เช่น

2.1.1 มีทุนทรัพย์จำนวนจำกัดไม่สามารถกระจายการลงทุนในหลักทรัพย์ต่างประเทศได้มากพอเพื่อลดความเสี่ยงจากการลงทุนหรือ (ธัญวงศ์ กิริตวานิชย์ และภัสรา ชวาลกร, 2547)

2.2.2 ไม่มีประสบการณ์ความรู้ความชำนาญในการลงทุนหรือ (ธัญวงศ์ กิริตวานิชย์ และภัสรา ชวาลกร, 2547)

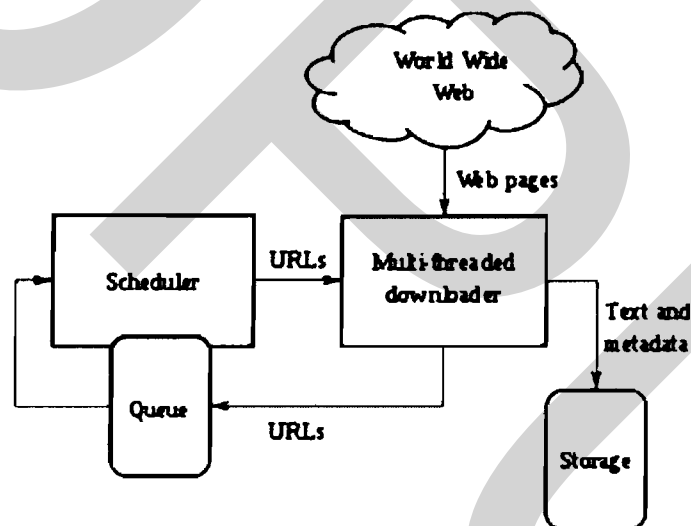
2.2.3 ไม่มีเวลาจะศึกษาค้นหาและติดตามข้อมูลเพื่อใช้ในการตัดสินใจการลงทุน (ธัญวงศ์ กิริตวานิชย์ และภัสรา ชวาลกร, 2547)

2.2.4 กองทุนรวมจึงเป็นเครื่องมือในการลงทุนที่มีประสิทธิภาพมีการจัดการลงทุนอย่างเป็นระบบโดยมีจุดมุ่งหมายให้การลงทุนได้รับผลตอบแทนที่ดีที่สุดภายใต้กรอบความเสี่ยงที่ผู้ลงทุนยอมรับได้ (ธัญวงศ์ กิริตวานิชย์ และภัสรา ชวาลกร:2547) รายละเอียดที่เหลือด้านความรู้เกี่ยวกับกองทุนรวมสามารถอ่านได้ที่ภาคผนวก ก.

## 2.2 เครื่องมือสืบค้นข้อมูลทางเว็บ

เครื่องมือสืบค้นข้อมูลทางเว็บได้รับการออกแบบมาเพื่อค้นหาข้อมูลบนเว็บไซต์ ไซด์เว็บ และเซิร์ฟเวอร์ FTP ผลการค้นหามักจะได้รับการนำเสนอในรูปแบบรายการผลการค้นหา ที่เรียกกันว่า SERPS หรือ “search engine results pages-หน้าผลการค้นหาจากเครื่องมือสืบค้นข้อมูล” จะประกอบไปด้วยหน้าเว็บ รูปภาพ ข้อมูล และไฟล์ในลักษณะอื่นๆ เครื่องมือสืบค้นข้อมูลบางตัว ยังดึงเอาข้อมูลที่มีในฐานข้อมูลหรือ directory เปิด (Open Directory) ไม่เหมือนกับ web directories ที่มีการดูแลรักษาโดยผู้ดูแลที่เป็นมนุษย์เท่านั้น เครื่องมือสืบค้นข้อมูลยังได้รับการดูแลรักษาข้อมูล แบบ real-time โดยการดำเนินการ Algorithm บน web crawler (Web search engine, Retrieved 2012, January 23)

### 2.2.1 เครื่องมือค้นหาเว็บไซต์ทำงานอย่างไร



ภาพที่ 2.1 สถาปัตยกรรมขั้นสูงของ Web Crawler มาตรฐาน

ที่มา: Wikipedia ([http://en.wikipedia.org/wiki/Web\\_search\\_engine](http://en.wikipedia.org/wiki/Web_search_engine))

เครื่องมือค้นหาทำงานเป็นลำดับดังนี้:

1. Web Crawling
2. Indexing
3. Searching

เครื่องมือค้นหาเว็บไซต์ทำงานโดยการเก็บข้อมูลเกี่ยวกับหน้าเว็บต่างๆ หลายหน้า ซึ่งได้มาจาก html นั้นเอง เพจเหล่านี้ถูกดึงมาโดย Web Crawler (บางครั้งก็รู้จักกันในนาม “แมงมุม”) ซึ่งเป็น Web Browser อัตโนมัติที่ติดตามการเชื่อมโยง (ลิงค์) ทุกอย่างในเว็บไซต์ การจะยกเว้นการเชื่อมโยงใด ๆ ทำได้โดยการใช้ robots.txt เนื้อหาในแต่ละหน้าสามารถนำมาวิเคราะห์เพื่อระบุว่าจะจัดทำเป็นดัชนี (index) อย่างไร (ตัวอย่างเช่น คำต่างๆ ถูกดึงออกมาจากชื่อเรื่อง หัวเรื่อง หรือส่วนพิเศษต่างๆ เรียกว่า meta tags) ข้อมูลเกี่ยวกับเว็บเพจนั้นถูกเก็บในฐานข้อมูลดัชนีเพื่อใช้ในอนาคต คำถามที่ตั้งสำหรับการค้นหานั้นอาจจะเป็นคำเดียวโดดๆ ก็ได้ วัตถุประสงค์ของดัชนี (index) ก็เพื่อให้ข้อมูลนั้นได้รับการค้นเจอเร็วที่สุดเท่าที่จะทำได้ เครื่องค้นหาข้อมูลบางตัว เช่น Google นั้นเก็บหน้าที่เป็นแหล่งข้อมูล (source) (เรียกว่า cache) ไว้ทั้งหมดหรือบางส่วน เช่นเดียวกับข้อมูลที่เกี่ยวข้องกับ web page ในขณะที่ตัวอื่นๆ เช่น AltaVista นั้นเก็บทุกคำของทุกหน้าที่มันค้นพบ หน้าเพจ cache นี้จะเก็บตัวอักษรที่ค้นหาตัวจริงไว้ตลอดเวลา เนื่องจากว่า มันเป็นตัวที่ถูกทำดัชนี ดังนั้น มันจึงเป็นประโยชน์มากเมื่อเนื้อหาของหน้าปัจจุบันได้รับการปรับเปลี่ยนให้เป็นปัจจุบัน และคำที่ใช้ค้นหานั้นหายไปแล้ว ปัญหานี้อาจจะได้รับการพิจารณาว่าเป็นปัญหาเล็กๆ สำหรับ linkrot และการจัดการของ Google ได้เพิ่มความสามารถในการใช้งาน โดยการทำให้ผู้ใช้พึงพอใจเพราะว่าคำที่ใช้ค้นหานั้นปรากฏอยู่ที่หน้าเว็บที่ปรากฏผลออกมา สิ่งนี้บรรลุหลักการประหลาดใจน้อยที่สุด เนื่องจากผู้ใช้อาจจะคาดหวังให้คำที่ใช้ค้นหานั้นปรากฏอยู่ในหน้าที่ปรากฏผลออกมาด้วย ความเกี่ยวข้องในการค้นหาที่เพิ่มมากขึ้นทำให้หน้าเว็บ cache เหล่านี้มีประโยชน์มาก นอกเหนือจากประโยชน์ที่หน้าเหล่านี้ มีข้อมูลที่ไม่สามารถหาได้จากที่อื่น (Web search engine, Retrieved 2012, January 23)

เมื่อผู้ใช้ใส่คำค้นหาลงในเครื่องมือค้นหา (โดยปกติแล้วจะใส่ คำสำคัญลงไป) เครื่องมือค้นหาจะตรวจสอบดัชนีของมันเอง และแสดงรายการหน้าเว็บที่มีความเกี่ยวข้องกับคำที่ค้นหามากที่สุดตามเกณฑ์ โดยปกติแล้วจะมีสรุปย่อสั้นๆ ที่บ่งบอกชื่อของเอกสาร และบางครั้งก็จะแสดงข้อความบางส่วนด้วย ดรรชนีนี้สร้างขึ้นจากข้อมูลที่เก็บไว้ และผ่านวิธีการนำข้อมูลไปทำดัชนี เป็นที่น่าเสียดายว่า ยังไม่มีเครื่องมือค้นหาสาธารณะใดที่อนุญาตให้ค้นหาเอกสารได้โดยวันที่ เครื่องมือค้นหาต่างๆ สนับสนุนการใช้งาน ตัวดำเนินการ Boolean และ หรือ และไม่ระบุคำค้นหาเพิ่มเติม ตัวดำเนินการ Boolean นั้นก็เพื่อการค้นหาคำตามตัวอักษร ที่อนุญาตให้ผู้ใช้สามารถเพิ่มการค้นหาให้ละเอียดยิ่งขึ้นและขยายผลการค้นหา เครื่องมือค้นหาจะค้นหาคำหรือวลีตรงตามตัวอักษรที่พิมพ์เข้าไป เครื่องมือค้นหาบางเครื่องมีเครื่องมือระดับสูงที่เรียกว่า การค้นหา ใกล้เคียง ที่อนุญาตให้ผู้ใช้ระบุระยะห่างระหว่างคำสำคัญต่างๆ ยังมีการค้นหาที่อยู่บนพื้นฐานของแนวคิด ที่การค้นหา มักจะต้องมีการใช้การวิเคราะห์เชิงสถิติ บนหน้าต่างๆ ที่มีคำหรือวลีที่คุณค้นหาอยู่



นอกจากนี้ยังมีการช่วยให้ผู้ใช้สามารถใช้คำพูดเหมือนที่ใช้ถามมนุษย์ด้วยกันในการพิมพ์ลงในช่องเพื่อการค้นหา เว็บไซต์ค้นหาเช่นนี้ คือ ask.com (Web search engine, Retrieved 2012, January23)

ประโยชน์ที่ได้จากเครื่องมือค้นหานี้ขึ้นอยู่กับความเกี่ยวข้องของชุดผลการค้นหาที่ได้รับ แม้ว่าจะมีหน้าเว็บเป็นล้านๆ ที่มีคำหรือวลีเฉพาะ บางหน้าหน้าก็อาจจะเกี่ยวข้องมากกว่าได้รับความนิยมนมากกว่า หรือมีอำนาจมากกว่าหน้าอื่นๆ เครื่องมือค้นหาส่วนใหญ่ใช้วิธีการจัดอันดับผลที่ได้เพื่อให้ผลที่ “ดีที่สุด” ขึ้นมาก่อน ส่วนคำถามที่ว่าเครื่องมือค้นหาตัดสินใจอย่างไรว่าหน้าไหนตรงกับคำค้นหามากที่สุด และลำดับใดที่ผลการค้นหาควรจะได้รับ การแสดงออกมานั้นเปลี่ยนแปลงไปตามเครื่องมือค้นหาแต่ละอัน วิธีการยังเปลี่ยนแปลงไปตามเวลาเนื่องจากการใช้อินเตอร์เน็ตนั้นเปลี่ยนแปลงไปและเทคนิคใหม่ๆ ได้รับการพัฒนาขึ้น มีเครื่องมือค้นหาสองแบบหลักๆ ที่เกิดขึ้น อย่างแรกคือระบบที่มีการให้คำจำกัดความและจัดอันดับคำสำคัญต่างๆ ไว้เรียบร้อยแล้วโดยมนุษย์ อีกแบบคือระบบที่สร้าง “ครรชนีส์กลับ-inverted index” โดยการวิเคราะห์ข้อความตามตำแหน่ง แบบที่สองนี้อาศัยคอมพิวเตอร์เป็นอย่างมากในการทำงานส่วนใหญ่ (Web search engine, Retrieved 2012, January23)

เครื่องมือค้นหาเว็บไซต์ส่วนใหญ่เป็นธุรกิจ ที่ได้รับการสนับสนุนโดยรายได้จากการโฆษณา ดังนั้น บางเครื่องมือจึงอนุญาตให้ผู้โฆษณา จ่ายเงินเพื่อให้รายการของพวกเขาอยู่ในอันดับที่สูงกว่าในผลการค้นหา เครื่องมือค้นหาที่ไม่รับเงินสำหรับผลการค้นหานี้ หายรายได้โดยการดำเนินการค้นหาการโฆษณาที่เกี่ยวข้อง ไปพร้อมกับผลการค้นหาตามปกติ เครื่องมือค้นหาจะได้รายได้ทุกครั้งที่มีการคลิกโฆษณา (Web search engine, Retrieved 2012, January23)

### 2.3 Web crawler

Web Crawler เป็นโปรแกรมคอมพิวเตอร์ ที่ค้นหา World Wide Web อย่างเป็นขั้นตอนและอัตโนมัติ หรือเป็นลำดับ คำที่ใช้เรียก Web Crawler อื่นๆ ได้แก่ ants, automatic indexers-ตัวสร้างครรชนีส์อัตโนมัติ, bots, Web Spiders, Web Robots หรือ ที่ใช้กันในชุมชน FOAF คือ Web Scutters (Web crawling, Retrieved 2012, January23)

กระบวนการนี้ เรียกว่า Web Crawling หรือ Spidering เว็บไซต์ต่างๆ มากมาย โดยเฉพาะอย่างยิ่ง เครื่องมือค้นหา ใช้การ spidering เป็นวิธีการปรับปรุงข้อมูลให้ทันสมัยอยู่เสมอ Web Crawlers มักจะถูกใช้ในการทำสำเนาหน้าเว็บที่เคยเข้าไปทั้งหมด เพื่อการดำเนินการโดยเครื่องมือค้นหาในอนาคตที่จะจัดระเบียบ จัดทำครรชนีส์หน้าเว็บที่ดาวน์โหลดมาแล้ว เพื่อให้สามารถค้นหาได้อย่างรวดเร็ว Crawlers ยังสามารถนำไปใช้เพื่อการดูแลรักษางานต่างๆ บนเว็บไซต์โดยอัตโนมัติ อาทิ การตรวจสอบลิงค์ (การเชื่อมโยง) หรือการยืนยันรหัส HTML

นอกจากนั้น Crawlers ยังสามารถนำไปใช้ในการรวบรวมข้อมูลเฉพาะประเภทต่างๆ จากหน้าเว็บ เช่นการค้นหาที่อยู่ e-mail (โดยปกติมักจะใช้ในการส่งจดหมายขยะ-spam) (Web crawling, Retrieved 2012, January23)

Web Crawler เป็น bot หรือเป็นตัวแทนซอฟต์แวร์ (software agent) ประเภทหนึ่ง โดยทั่วไปจะเริ่มด้วยรายการ URLs ที่จะไปค้นหา เรียกว่า seeds เมื่อ crawler เข้าเยี่ยมชม URLs ต่างๆ มันจะระบุ การเชื่อมโยงหลายมิติ (Hyperlinks) ทั้งหมดในหน้านั้นและเพิ่มการเชื่อมโยงเหล่านั้นลงไป ใน รายการ URLs ที่จะไปค้นหา เรียกว่า Crawl frontier ซึ่ง URLs ต่างๆจาก frontier นั้นจะถูกเข้าไปค้นหาแบบ recursive โดยจะเป็นไปตามชุดนโยบายที่กำหนด (Web crawling, Retrieved 2012, January23)

ปริมาณที่มากนั้นบ่งชี้ว่า crawler สามารถดาวน์โหลดเศษเสี้ยวหนึ่งของเว็บไซต์ในเวลาที่กำหนดเท่านั้น ดังนั้นมันจะต้องเรียงลำดับความสำคัญของรายการที่จะดาวน์โหลด อัตราการเปลี่ยนแปลงที่สูงนั้นสะท้อนว่าหน้าเว็บต่างๆ อาจจะมีการปรับปรุงให้ทันสมัยแล้วหรืออาจจะถูกลบไปแล้วก็ได้ (Web crawling, Retrieved 2012, January23)

จำนวน URLs ที่สามารถเข้าไปค้นหาโดย Crawler ที่ถูกสร้างขึ้นโดยซอฟต์แวร์ด้านเซิร์ฟเวอร์นั้นสร้างความยากลำบากให้กับ Web Crawlers ในการหลีกเลี่ยงการจับเก็บเนื้อหาที่ซ้ำซ้อนกัน มีชุดตัวแปร HTTP GET (บนพื้นฐานของ URL) มากมาย แต่มีเพียงแค่มือถือมือหนึ่งเท่านั้นที่จะให้เนื้อหาที่ไม่เหมือนใครออกมา ตัวอย่างเช่น Gallery รูปภาพออนไลน์ อาจจะเสนอตัวเลือก 3 ตัวเลือกให้กับผู้ใช้ ดังที่ระบุผ่านตัวแปร HTTP GET ใน URL หากมีวิธีการคัดแยกรูปภาพ 4 วิธี มีสามวิธีที่เป็นการดูรูป Thumbnail สองรูปแบบไฟล์ และอีกหนึ่งตัวเลือกเพื่อการหยุดการใช้เนื้อหาที่ผู้ใช้ป้อนเข้ามา เนื้อหาชุดเดียวกันจะสามารถเข้าถึงได้ด้วย 48 URLs ทุก URL อาจจะเชื่อมโยงมาที่เว็บไซต์ การจัดกลุ่มเชิงคณิตศาสตร์นี้ สร้างปัญหาให้กับ Crawler เนื่องจากพวกมันจะต้องจัดเรียงกลุ่มต่างๆ ที่มีมากมาย ที่มีการเปลี่ยนแปลงเพียงเล็กน้อย เพื่อให้ได้มาซึ่งเนื้อหาที่ไม่ซ้ำใคร (Web crawling, Retrieved 2012, January23)

ดังที่ Edwards *et al.* ได้กล่าวไว้ “เมื่อ Bandwidth สำหรับการดำเนินการ Crawl นั้นจำกัดและไม่ใช่ของฟรี มันจึงจำเป็นที่จะต้อง Crawl เว็บต่างๆ อย่างมีประสิทธิภาพ ไม่คำนึงถึงปริมาณที่สามารถเข้าถึงได้เป็นหลัก หากยังต้องการรักษาคุณภาพหรือความสดใหม่ให้คงอยู่” Crawler จะต้องคัดเลือกหน้าเว็บที่จะเข้าค้นหาในครั้งต่อไปอย่างระมัดระวัง (Web crawling, Retrieved 2012, January23)

พฤติกรรมของ Web Crawler เป็นผลมาจากการผสมผสานระหว่างนโยบายต่างๆ:

### 2.3.1 นโยบายการคัดเลือก

นโยบายการคัดเลือก ที่บ่งบอกว่าให้ดาวน์โหลดหน้าเว็บใด เนื่องจากขนาดของเว็บที่มีขอบเขตกว้างในปัจจุบัน แม้แต่เครื่องมือค้นหาขนาดใหญ่ยังครอบคลุมเพียงแค่ส่วนหนึ่งของเว็บไซต์ที่เปิดสู่สาธารณะ งานศึกษาปี ค.ศ.2005 แสดงให้เห็นว่าเครื่องมือค้นหาใหญ่ๆ จัดทำดัชนีได้ไม่เกิน 40-70% ของเว็บที่จัดทำดัชนีได้ งานศึกษาก่อนหน้าโดย Dr.Steve Lawrence และ Lee Giles แสดงให้เห็นว่า ไม่มีเครื่องมือค้นหาใดที่จัดทำดัชนีได้มากกว่า 16% ของเว็บไซต์ในปี ค.ศ.1999 เนื่องจาก Crawler ดาวน์โหลดเพียงแค่เศษเสี้ยวหนึ่งของหน้าเว็บต่างๆ จึงเป็นสิ่งที่ดีถ้าเศษเสี้ยวที่ดาวน์โหลดลงมานั้นมีหน้าเว็บที่เกี่ยวข้องมากที่สุด ไม่ใช่เพียงตัวอย่างสุ่มของเว็บไซต์ (Web crawling, Retrieved 2012, January23)

การจะทำเช่นนี้ได้ ต้องการการให้ความสำคัญกับการจัดลำดับความสำคัญของหน้าเว็บต่างๆ ความสำคัญของแต่ละหน้า เป็นฟังก์ชันของคุณภาพของตัวหน้าเว็บ ความนิยมในทำนองของการเชื่อมโยงหรือการเข้าชม และแม้แต่ URL ของมันเอง (อันสุดท้ายนี้เป็นกรณีของเครื่องมือค้นหาแนวตั้งที่ถูกจำกัดอยู่ใน โดเมนชั้นบนสุด (top-down domain) โดเมนเดียว หรือเครื่องมือค้นหาที่ถูกจำกัดอยู่ที่เว็บไซต์เว็บเดียว) การออกแบบนโยบายการคัดเลือกที่ดีนั้นเพิ่มความยากขึ้นมา เพราะว่ามันจะต้องทำงานร่วมกับข้อมูลเพียงส่วนเดียว เนื่องจากชุดหน้าเว็บสมบูรณ์นั้นไม่สามารถรู้ได้ ในระหว่างการ Crawling (Web crawling, Retrieved 2012, January23)

Cho et al. ได้ทำการศึกษา นโยบายต่างๆ สำหรับการจัดการตารางการ Crawling เป็นครั้งแรก ชุดข้อมูลของพวกเขา คือ หน้าเว็บ 180,000 หน้า ที่ crawl มาจากโดเมนของ stanford.edu ซึ่งมีการจำลองการ crawling ด้วยวิธีการที่แตกต่างกันไป Ordering Metrics ที่ได้รับการทดสอบนั้น อยู่ในรูปแบบของ Breadth-first, backlink-count และมีการคำนวณแบบเรียงลำดับหน้า (Pagerank) เป็นบางส่วน หนึ่งในข้อสรุปคือหาก crawler ต้องการดาวน์โหลดหน้าเว็บที่มีค่า Pagerank สูงก่อนเวลา ในช่วงที่มีการ Crawling นั้น การใช้วิธีการ Pagerank บางส่วนเป็นวิธีที่ดีกว่า ตามมาด้วย breadth-first และ backlink-count อย่างไรก็ตามผลที่ได้นี้เป็นผลสำหรับโดเมนเดียวเท่านั้น Cho ยังเขียนวิทยานิพนธ์ปริญญาเอกของเขาที่ Stanford เกี่ยวกับ Web Crawling อีกด้วย (Web crawling, Retrieved 2012, January23)

Najork and Wiener ได้ดำเนินการ crawl จริงๆ กับหน้าเว็บ 328 ล้านหน้าโดยใช้การจัดอันดับแบบ Breadth-first พวกเขาพบว่า การ crawl แบบ breadth-first นั้นจัดเก็บหน้าต่างๆ ที่มีค่า Pagerank สูงได้ตั้งแต่เนิ่นๆ ในการ crawl (แต่พวกเขาไม่ได้เปรียบเทียบวิธีการนี้กับวิธีการอื่นๆ)

คำอธิบายที่ให้โดยผู้เขียนสำหรับผลที่ได้นี้ก็คือ “หน้าที่มีความสำคัญมากที่สุด มีการเชื่อมโยง (ลิงค์) มากมายมาที่หน้าเหล่านี้โดยโฮสต์ต่างๆมากมาย และจะพบการเชื่อมโยงเหล่านี้ได้แต่เนิ่นๆ โดยไม่คำนึงว่าโฮสต์หรือหน้าเว็บใดที่การ Crawl เริ่มต้นขึ้น (Web crawling, Retrieved 2012, January23)

Abiteboul ได้ออกแบบวิธีการ crawling บนพื้นฐานของ algorithm ที่เรียกว่า OPIC (On-line Page Importance Computation- การคำนวณความสำคัญของหน้าเว็บออนไลน์) ใน OPIC นั้น หน้าเว็บแต่ละหน้าจะได้รับผลรวมอันดับของ “cash” ที่กระจายเท่าๆ กันไปตามหน้าเว็บต่างๆ ที่มีการบ่งชี้ไป คล้ายๆกับการคำนวณ Pagerank แต่เร็วกว่าและดำเนินการในขั้นตอนเดียว Crawler ที่ทำงานโดยใช้ OPIC ดาว์นโหลดหน้าเว็บต่างๆใน Crawling Frontier ที่มีปริมาณ 'cash' สูงก่อนเป็นอันดับแรก การทดลองนั้นดำเนินการกราฟสังเคราะห์ 100,000 หน้า ด้วยการแจกแจง ตามกฎเลขยกกำลัง ( power-law) ของ in-links อย่างไรก็ตาม ไม่มีการเปรียบเทียบกับวิธีการอื่นๆ และไม่มีการทดลองบนเว็บจริง (Web crawling, Retrieved 2012, January23)

Boldi et al. ได้ใช้การจำลองกับสับเซตของเว็บ 40 ล้านหน้า จากโดเมน .it และ 100 ล้านหน้าจาก WebBase crawl และทำการทดสอบ breadth-first กับ depth-first ทดสอบการจัดอันดับสุ่มและวิธีการรอบรู้ การเปรียบเทียบนั้นอยู่บนพื้นฐานประสิทธิภาพของการคำนวณค่า PageRank แท้จริงโดยประมาณ ในการ crawl บางส่วน เป็นที่น่าประหลาดใจว่าการค้นดูบางเว็บที่สะสม PageRank อย่างรวดเร็ว นั้น (ที่มีชื่อคือการค้นดูแบบ breadth-first และรอบรู้) ให้ผลการคาดการณ์ก้าวหน้าที่ต่ำมาก (Web crawling, Retrieved 2012, January23)

Baeza-Yates et al. ได้ใช้การจำลองกับสับเซต สองสับเซตของเว็บ 3 ล้านหน้า จากโดเมน .gr และ .cl ทดสอบวิธีการ crawling หลายวิธี พวกเขาได้แสดงให้เห็นว่าทั้งวิธี OPIC และวิธีที่ใช้ความยาวของการเข้าแถวต่อเว็บนั้นดีกว่า การ crawl แบบ breadth-first และยังเป็นวิธีที่มีประสิทธิภาพด้วยในการใช้ crawl ที่ผ่านมาในการแนะนำให้กับ crawl ปัจจุบัน หากทำได้ (Web crawling, Retrieved 2012, January23)

Daneshpajouh et al. ได้ออกแบบ algorithm บนพื้นฐานของชุมชน สำหรับการค้นหา seeds ที่ดี วิธีการนี้ จะทำการ crawl หน้าเว็บที่มี PageRank สูงจากหลายๆชุมชน ในรูปแบบที่ไม่ซ้ำกัน ในการเปรียบเทียบกับ crawl ที่เริ่มจาก seeds สุ่ม seed ที่ดีจะถูกดึงมาได้จากเว็บที่มีการ crawl ไปก่อนหน้า เมื่อมีการใช้วิธีการใหม่นี้ หากมีการใช้ seeds เหล่านี้ crawl แบบใหม่ก็มีโอกาสที่จะมีประสิทธิภาพสูง (Web crawling, Retrieved 2012, January23)

### 2.3.1.1 การCrawling แบบเน้นความสนใจ

ความสำคัญของหน้าเว็บสำหรับ Crawler อาจจะแสดงออกมาได้ในรูปแบบฟังก์ชันของความคล้ายคลึงกันของหน้า กับคำที่ใช้ในการค้นหา Web Crawler ที่พยายามจะดาวน์โหลดหน้าต่างๆที่มีความคล้ายคลึงกันนั้นเรียกว่า Crawler เน้นความสนใจ หรือ crawler เฉพาะเรื่อง หลักการของการ Crawling แบบเน้นความสนใจหรือเฉพาะเรื่องนี้ เริ่มมีการพูดถึงโดย Menczer และโดย Chakrabart et al. (Web crawling, Retrieved 2012, January23)

ปัญหาหลักของการ Crawling แบบเน้นความสนใจก็คือ ในบริบทของ Web Crawler นั้น เราต้องการที่จะสามารถพยากรณ์ความคล้ายคลึงกันของเนื้อหาข้อความในหน้าเว็บกับคำค้นหาได้ก่อนที่จะดาวน์โหลดหน้านั้นมาจริงๆ ผู้พยากรณ์ที่เป็นไปได้คือ anchor text ของการเชื่อมโยง (ลิงค์) ต่างๆ นี่เป็นแนวทางที่ใช้โดย Pinkerton ใน web crawler ตัวแรกๆ ของช่วงแรกๆ ของเว็บไซต์ Diligenti et al. เสนอให้มีการใช้เนื้อหาสมบูรณ์ของหน้าต่างๆที่ได้เคยค้นดูมาแล้ว ในการบ่งชี้ความคล้ายคลึงกันระหว่างคำค้นหากับหน้าต่างๆที่ยังไม่เคยค้นดู การทำงานของ Crawling แบบเน้นความสนใจนี้มักจะขึ้นอยู่กับจำนวนของการเชื่อมโยง (ลิงค์) ในหัวข้อเฉพาะที่กำลังถูกค้นหา และ Crawling แบบเน้นความสนใจมักจะอาศัยเครื่องมือค้นหาเว็บทั่วไปในการเริ่มต้น (Web crawling, Retrieved 2012, January23)

#### 1) การจำกัดการเชื่อมโยงที่มีการติดตามแล้ว

Crawler อาจจะต้องการหาหน้า HTML และหลักเสียงประเภท MIME อื่นๆ ในการที่จะขอทรัพยากร HTML เท่านั้น Crawler อาจจะต้องใช้คำขอ HTTP HEAD เพื่อระบุประเภท MIME ของทรัพยากรเว็บไซด์ก่อนที่จะขอทรัพยากรทั้งหมดด้วยคำขอ GET เพื่อหลีกเลี่ยงการขอ HEAD มากเกินไป Crawler อาจจะตรวจสอบ URL และขอทรัพยากรก็ต่อเมื่อ URL นั้นลงท้ายด้วยอักขระ เช่น .html, .htm, .asp, .aspx, .php, .jsp, .jspx หรือ เครื่องหมาย/ วิธีการนี้อาจจะทำให้ทรัพยากรเว็บ HTML มากมายตกหล่นไปโดยไม่ตั้งใจ (Web crawling, Retrieved 2012, January23)

Crawler บางตัวอาจจะหลีกเลี่ยงการขอทรัพยากรใดๆ ที่มีเครื่องหมาย ? อยู่ด้วย (เกิดขึ้นอย่าง dynamic- มีการทำให้เกิดขึ้น-ผู้แปล) เพื่อที่จะหลบกับดักแมงมุม (Spider Traps) ที่อาจทำให้ crawler ดาวน์โหลด URL จากเว็บไซด์มาอย่างไม่จำกัด วิธีการนี้ไม่น่าเชื่อถือหากเว็บไซด์ใช้การเขียน URL ใหม่เพื่อให้ URL ต่างๆง่ายขึ้น (Web crawling, Retrieved 2012, January23)

#### 2) การทำURL ให้เป็นมาตรฐาน

Crawler มักจะดำเนินการทำ URL ให้เป็นมาตรฐานอย่างใดอย่างหนึ่งเพื่อหลีกเลี่ยงการ Crawling ทรัพยากรที่เหมือนกันมากกว่าหนึ่งครั้ง ศัพท์คำว่า การทำ URL ให้เป็นมาตรฐาน ยังถูกเรียกว่า การเลือก URL ที่ดีที่สุด (URL Canonicalization) ได้อีกด้วย ซึ่งจะหมายถึงกระบวนการใน

การปรับเปลี่ยนและทำ URL ให้เป็นมาตรฐานอย่างสม่ำเสมอ มีประเภทการทำให้เป็นมาตรฐานหลายประเภท ที่อาจจะใช้ดำเนินการ นี้รวมถึงการแปลง URL ให้อยู่ในรูปตัวพิมพ์เล็ก การกำจัดส่วน “.” และ “..” ออก และการเพิ่มเครื่องหมาย / กับส่วนที่ไม่เป็นที่ว่าง เพื่อประโยชน์ในการติดตามค้นหา (Web crawling, Retrieved 2012, January 23)

### 3) การ Crawling แบบ Path-ascending (ขึ้นตามทาง-ผู้แปล)

Crawler บางตัวตั้งใจจะดาวน์โหลดทรัพยากรให้ได้มากที่สุดเท่าที่จะทำได้จากเว็บไซต์แต่ละแห่ง ดังนั้นจึงเกิด Path-ascending Crawler ขึ้น เพื่อให้เดินทางขึ้นตามทางทุกทางในแต่ละ URL ที่ Crawler นั้นประสงค์จะ crawl ตัวอย่างเช่น เมื่อได้ seed ที่มี URL เป็น <http://llama.org/hamster/monkey/page.html> Crawler จะพยายาม crawl คำว่า /hamster/monkey/, /hamster/, และ /. Cothey ค้นพบว่า Crawler แบบ path-ascending นี้มีประสิทธิภาพสูงในการค้นหาทรัพยากรที่อยู่โดดๆ หรือทรัพยากรที่โดยปกติแล้วจะหาไม่เจอโดยการเชื่อมโยงภายใน (inbound link) ในการ Crawling ตามปกติ (Web crawling, Retrieved 2012, January 23)

Crawler แบบ path-ascending หลายตัวยังเป็นที่รู้จักในนามซอฟต์แวร์เก็บเกี่ยวเว็บไซต์ (Web harvesting software) เพราะว่าพวกมันถูกใช้ในการ “เก็บเกี่ยว” หรือ เก็บรวบรวมเนื้อหาทั้งหมด – อาจจะเป็นอัลบั้มรูป- จากหน้าเว็บนั้นๆ หรือจากโฮสต์ (Web crawling, Retrieved 2012, January 23)

#### 2.3.1.2 นโยบายการกลับไปค้นดู

นโยบายการเข้าค้นดูอีกครั้ง ที่บ่งบอกว่า จะตรวจสอบดูการเปลี่ยนแปลงที่เกิดกับหน้าเว็บเมื่อไร เว็บไซต์นั้นเปลี่ยนแปลงไปได้ตลอดเวลา และการ Crawling เศษเสี้ยวหนึ่งของเว็บไซต์เหล่านั้นอาจใช้เวลาหลายสัปดาห์หรือหลายเดือน เมื่อ Web Crawler เสร็จสิ้นการ Crawl แล้ว ก็อาจจะไม่ทันต่อการเกิดขึ้นใหม่ของเว็บไซต์ การปรับข้อมูลให้ทันสมัยและการลบข้อมูลบางอย่างทิ้งไป (Web crawling, Retrieved 2012, January 23)

ในมุมมองของเครื่องมือค้นหา มีต้นทุนที่เข้ามาเกี่ยวข้อง เมื่อไม่สามารถตรวจพบเหตุการณ์ต่างๆ เหล่านั้นได้ ทำให้เกิดการสำเนาทรัพยากรที่ไม่เป็นปัจจุบัน ฟังก์ชันต้นทุนที่ใช้กันมากที่สุด คือ ความสดใหม่และอายุ (Web crawling, Retrieved 2012, January 23)

ความสดใหม่: คือหน่วยวัดทวิภาค ที่ระบุว่าสำเนาที่สร้างขึ้น ณ ตรงนั้น ถูกต้องหรือไม่ ความสดใหม่ของหน้า  $p$  ในที่เก็บ ณ เวลา  $t$  นั้นสามารถให้คำจำกัดความว่า

$$F_p(t) = \begin{cases} 1 & \text{หาก } p \text{ เท่ากับสำเนา ณ ตรงนั้น ณ เวลา } t \\ 0 & \text{นอกจากนั้น} \end{cases}$$

ภาพที่ 2.2 สูตรคำนวณความสดใหม่ของข้อมูล

ที่มา: Wikipedia ([http://en.wikipedia.org/wiki/Web\\_crawling](http://en.wikipedia.org/wiki/Web_crawling))

อายุ: เป็นหน่วยวัดที่ระบุว่าสำเนาที่สร้างขึ้น ณ ตรงนั้น ล้าสมัยไปมากเพียงใด อายุของหน้า  $p$  ในที่เก็บ ณ เวลา  $t$  นั้นสามารถให้คำจำกัดความว่า

$$A_p(t) = \begin{cases} 0 & \text{หาก } p \text{ ไม่ถูกเปลี่ยนแปลง ณ เวลา } t \\ t - & \text{เวลาที่ใช้ในการเปลี่ยนแปลง } p \end{cases}$$

ภาพที่ 2.3 สูตรคำนวณอายุของข้อมูล

ที่มา: Wikipedia ([http://en.wikipedia.org/wiki/Web\\_crawling](http://en.wikipedia.org/wiki/Web_crawling))

Coffman et al. ได้ศึกษาคำจำกัดความของเป้าหมายของ Web Crawler ที่มีความหมายเท่ากับความสดใหม่ แต่ใช้อีกคำหนึ่ง: พวกเขาเสนอว่า Crawler จะต้องลดเสียเวลาที่หน้าต่างๆ จะล้าสมัย ลงให้น้อยที่สุด พวกเขายังสังเกตอีกว่า ปัญหาของ Web Crawling สามารถนำมาแสดงในรูปแบบของ multiple-queue (การเข้าหลายแถว) Single-server polling system (ระบบการลงคะแนนเซิร์ฟเวอร์เดียว) ซึ่ง Web Crawler เป็นเซิร์ฟเวอร์และเว็บไซต์เป็นแถวหรือคิว การเปลี่ยนแปลงหน้าเว็บ คือ การเข้ามาของลูกค้า และ switch-over times (เวลาที่สลับไปมา) คือช่วงเวลาระหว่างการเข้าถึงหน้าเว็บไปจนถึงเว็บไซต์เว็บหนึ่ง ภายใต้รูปแบบนี้ ระยะเวลารอโดยเฉลี่ยสำหรับลูกค้าในระบบลงคะแนน นั้นจะเท่ากับอายุเฉลี่ยสำหรับ Web Crawler (Web crawling, Retrieved 2012, January 23)

เป้าหมายของ Crawler คือการรักษาความสดใหม่เฉลี่ยของหน้าต่างๆ ที่เก็บรวบรวมมา ให้สูงที่สุดเท่าที่จะทำได้ หรือทำให้อายุเฉลี่ยของหน้าต่างๆ นั้นต่ำที่สุดเท่าที่จะทำได้ เป้าหมายเหล่านี้ ไม่เหมือนกัน: ในกรณีแรก Crawler เพียงสนใจว่ามีหน้าเว็บที่ล้าสมัยกี่หน้า ในขณะที่ใน

กรณีที่สอง Crawler นั้นสนใจว่า สำเนาหน้าเว็บ ณ จุดนั้น มีอายุเท่าไรแล้ว Cho และ Garcia-Molina ได้ศึกษานโยบายการกลับไปค้นดูต่างๆ สองนโยบาย (Web crawling, Retrieved 2012, January 23)

นโยบายแบบแผนเดียวกัน (Uniform Policy): รวมเอาการกลับไปค้นดูหน้าเว็บทุกหน้าที่เก็บรวบรวมมาไว้และมีความถี่เดียวกัน ไม่คำนึงถึงอัตราการเปลี่ยนแปลงของหน้าเว็บเหล่านั้น

นโยบายสัดส่วน (Proportional Policy): รวมเอาการกลับไปค้นดูหน้าเว็บที่เปลี่ยนแปลงถี่กว่า มากกว่า ความถี่ของการเยี่ยมชมนั้น เป็นสัดส่วนแปรผันตรงกับความถี่การเปลี่ยนแปลง (โดยประมาณ)

(ในทั้งสองกรณี ลำดับการ crawling เข้าในหน้าเว็บเดิมอาจกระทำแบบสุ่มหรือแบบคำสั่งตายตัว)

Cho and Garcia-Molina ได้พิสูจน์ผลอันน่าประหลาดใจ ในกรณีความสดใหม่เฉลี่ยนั้น นโยบายแบบแผนเดียวกันให้ผลดีกว่า นโยบายสัดส่วน ทั้งใน Web Crawl จำลองและ Web Crawl จริง โดยสัญชาตญาณแล้วเหตุผลมีอยู่ว่า เนื่องจาก Web Crawler มีข้อจำกัดในเรื่องของจำนวนหน้าเว็บที่สามารถ Crawl ได้ในกรอบเวลาหนึ่ง 1.) พวกมันจะจัดสรรการ Crawl ใหม่ๆ ให้กับหน้าเว็บที่เปลี่ยนแปลงอย่างรวดเร็ว มากเกินไป ในขณะที่หน้าเว็บที่มีการเปลี่ยนแปลงในความถี่ที่น้อยกว่าถูกละเลย และ 2.) ความสดใหม่ของหน้าเว็บที่เปลี่ยนแปลงอย่างรวดเร็ว นั้นมีระยะเวลาที่สั้นกว่าหน้าเว็บที่มีการเปลี่ยนแปลงในความถี่น้อยกว่า พูดย่างๆ ก็คือ นโยบายสัดส่วนจัดสรรทรัพยากรให้การ Crawl หน้าเว็บที่มีการเปลี่ยนแปลงอย่างรวดเร็ว แต่ประสบกับความสดใหม่เฉลี่ยที่น้อยกว่า (Web crawling, Retrieved 2012, January 23)

ในการพัฒนาความสดใหม่ Crawler ควรจะมีบทลงโทษกับองค์ประกอบที่เปลี่ยนแปลงบ่อยเกินไป นโยบายการกลับไปค้นดูที่เหมาะสมที่สุดนั้น ไม่ใช่ทั้งนโยบายแบบแผนเดียวกันหรือนโยบายสัดส่วน วิธีการที่เหมาะสมที่สุดในการรักษาระดับความสดใหม่เฉลี่ยให้สูงนั้น คือการไม่สนใจหน้าเว็บที่เปลี่ยนแปลงเร็วเกินไป และวิธีการที่เหมาะสมที่สุดในการทำอายุเฉลี่ยให้ต่ำก็คือการใช้การเพิ่มขึ้นของความถี่การเข้าถึงฝ่ายเดียว (และ sub-linear) เมื่อเทียบกับอัตราการเปลี่ยนแปลงของหน้าเว็บแต่ละหน้า ในทั้งสองกรณี สิ่งที่เหมาะสมที่สุด จะเข้าใกล้กับนโยบายแบบแผนเดียวกันมากกว่า นโยบายสัดส่วน ดังที่ Coffman et al. สังเกต “ในการที่จะลดเวลาถ้าสมมุติโดยประมาณลงให้ต่ำที่สุด การเข้าถึงหน้าเว็บใดๆ ควรจะแจกแจงให้เท่าๆ กัน เท่าที่จะทำได้” โดยทั่วไปแล้วคงไม่สามารถหาสูตรที่ชัดเจนสำหรับนโยบายกลับไปค้นดูได้ แต่สามารถจะหาได้ในเชิงตัวเลข เนื่องจากสิ่งเหล่านี้ ขึ้นอยู่กับการแจกแจงการเปลี่ยนแปลงของหน้าเว็บ Cho และ Garcia-Molina แสดงให้เห็นว่าการแจกแจงแบบ exponential นั้นคือการอธิบายการเปลี่ยนแปลงของหน้า



ในขณะที่ Ipeiritise et al. แสดงให้เห็นว่าจะใช้เครื่องมือทางสถิติในการค้นหาตัวแปรที่ส่งผลกระทบต่อการแจจแจงได้อย่างไร สังเกตว่านโยบายการกลับไปค้นดูที่มีการพิจารณา ณ ที่นี้นั้น สมมติให้หน้าเว็บทั้งหมด มีคุณภาพเหมือนกันหมด (“ทุกๆ หน้าในเว็บ มีคุณค่าเท่ากัน”) ซึ่งในความเป็นจริงแล้วไม่ใช่อย่างนั้น ดังนั้นควรจะมีการนำข้อมูลเพิ่มเติมเกี่ยวกับคุณภาพของหน้าเว็บ รวมเข้าไปในการวิเคราะห์ด้วยเพื่อให้เกิดนโยบาย Crawling ที่ดียิ่งขึ้น (Web crawling, Retrieved 2012, January23)

### 2.3.1.3 นโยบายความสุภาพ

นโยบายความสุภาพ ที่บ่งบอกว่าจะหลีกเลี่ยงภาวะ Overload ของเว็บไซต์ต่างๆ ได้อย่างไร Crawler สามารถดึงข้อมูลได้เร็วกว่าและลึกกว่าผู้ค้นหาที่เป็นมนุษย์ ดังนั้น พวกมันจึงสามารถส่งผลกระทบต่อการทำงานของเว็บไซต์ เป็นที่รู้กันดีว่า หาก Crawler ใดๆ กำลังดำเนินการตามคำขอหลายๆ คำขอต่อวินาที และ/หรือดาวน์โหลดไฟล์ใหญ่ๆ หลายไฟล์ เซิร์ฟเวอร์ก็อาจจะต้องเผชิญกับความยุ่งยากในการดำเนินการตามคำขอจาก Crawler หลายๆ ตัว

ดังที่ Koster ได้สังเกต การใช้ Web Crawlers นั้นเป็นประโยชน์ในงานต่างๆ มากมาย แต่ก็มีข้อเสียต่อชุมชน (Web crawling, Retrieved 2012, January23) โดยทั่วไป ต้นทุนของการใช้ Web Crawlers นั้นมีดังนี้:

- 1) ทรัพยากรเครือข่าย เนื่องจาก Crawlers นั้นต้องการ bandwidth ที่มากพอสมควร และมีระดับการทำงานอย่างคู่ขนานสูง เป็นเวลานาน
- 2) เซิร์ฟเวอร์ Overload โดยเฉพาะอย่างยิ่งหากความถี่ของการเข้าถึงเซิร์ฟเวอร์นั้นๆ สูงเกินไป
- 3) Crawlers ที่ถูกเขียนขึ้นมาไม่ดี ซึ่งอาจทำให้เซิร์ฟเวอร์หรือเราเตอร์พัง หรืออาจจะดาวน์โหลดหน้าเว็บที่พวกมันไม่สามารถจะดูแลได้ และ
- 4) Crawlers ส่วนบุคคลที่ หากถูกใช้โดยผู้ใช้อย่างหลายคน อาจจะทำให้อุปกรณ์เครือข่ายและเซิร์ฟเวอร์เว็บไซต์ได้

ทางออกส่วนหนึ่งของปัญหาเหล่านี้ก็คือ การใช้โปรโตคอลแยก robots (Robots Exclusion Protocol) หรือเรียกอีกอย่างว่า robots.txt protocol สิ่งนี้เป็นมาตรฐานสำหรับผู้ดูแลในการระบุว่าส่วนใดของเซิร์ฟเวอร์เว็บของพวกเขาที่ไม่ควรจะถูกเข้าถึงได้โดย Crawlers มาตรฐานนี้ ไม่รวมการเสนอให้เข้าดูเป็นการภายในของเซิร์ฟเวอร์เดียวกัน แม้ว่าการเข้าถึงเป็นการภายในจะเป็นวิธีการที่มีประสิทธิภาพมากที่สุดในการหลีกเลี่ยง เซิร์ฟเวอร์ Overload เมื่อไม่นานมานี้ เครื่องมือค้นหาเชิงธุรกิจเช่น Ask Jeeves, MSN และ Yahoo นั้นสามารถใช้ตัวแปร “Crawl-delay” เพิ่มเติมได้

แล้ว ในไฟล์ robots.txt เพื่อระบุจำนวนวินาทีที่คำขอถูกทำให้เข้าไป (Web crawling, Retrieved 2012, January23)

ช่วงเวลาระหว่างการเชื่อมต่อที่เสนอมาอันแรกคือ 60 วินาที อย่างไรก็ตาม หากหน้าเว็บต่างๆ ถูกดาวน์โหลดจากเว็บไซด์ในอัตรานี้ ด้วยหน้าเว็บมากกว่า 100,000 หน้า แม้ว่าจะมีการเชื่อมต่อที่สมบูรณ์ มีค่าความหน่วงเวลาเป็นศูนย์ และมี Bandwidth ไม่จำกัด ยังจะต้องใช้เวลามากกว่าสองเดือนในการดาวน์โหลดเว็บไซด์เหล่านั้น นอกจากนี้ ยังมีเพียงเศษเสี้ยวของทรัพยากรจาก Web Server นั้นเท่านั้นที่จะถูกใช้ นี่ยังใช้ไม่ได้ (Web Search Engine, Retrieved 2012, January23)

Cho ใช้ช่วงเวลา 10 วินาที ในการเข้าถึง และ WIRE Crawler ใช้เวลา 15 วินาทีเป็นค่าปกติ Mercator Web Crawler ดำเนินการตามนโยบายความสุภาพที่ถูกปรับให้เหมาะสม: หากใช้เวลา  $t$  วินาทีในการดาวน์โหลดเอกสารหนึ่งจาก server ที่กำหนด crawler จะรอ  $10t$  วินาทีก่อนที่จะดาวน์โหลดหน้าต่อไป Dill et al. ใช้เวลา 1 วินาที (Web crawling, Retrieved 2012, January23)

สำหรับผู้ที่ใช้ Web Crawler เพื่อจุดประสงค์ในการค้นคว้า วิจัย จะต้องมีการวิเคราะห์ข้อดี-ข้อเสียอย่างละเอียดกว่านี้ และจะต้องมีการดำเนินการพิจารณาในเรื่องของจรรยาบรรณด้วยการตัดสินใจว่าจะ Crawl ที่ใดบ้างและจะ Crawl เร็วเท่าไร (Web crawling, Retrieved 2012, January23)

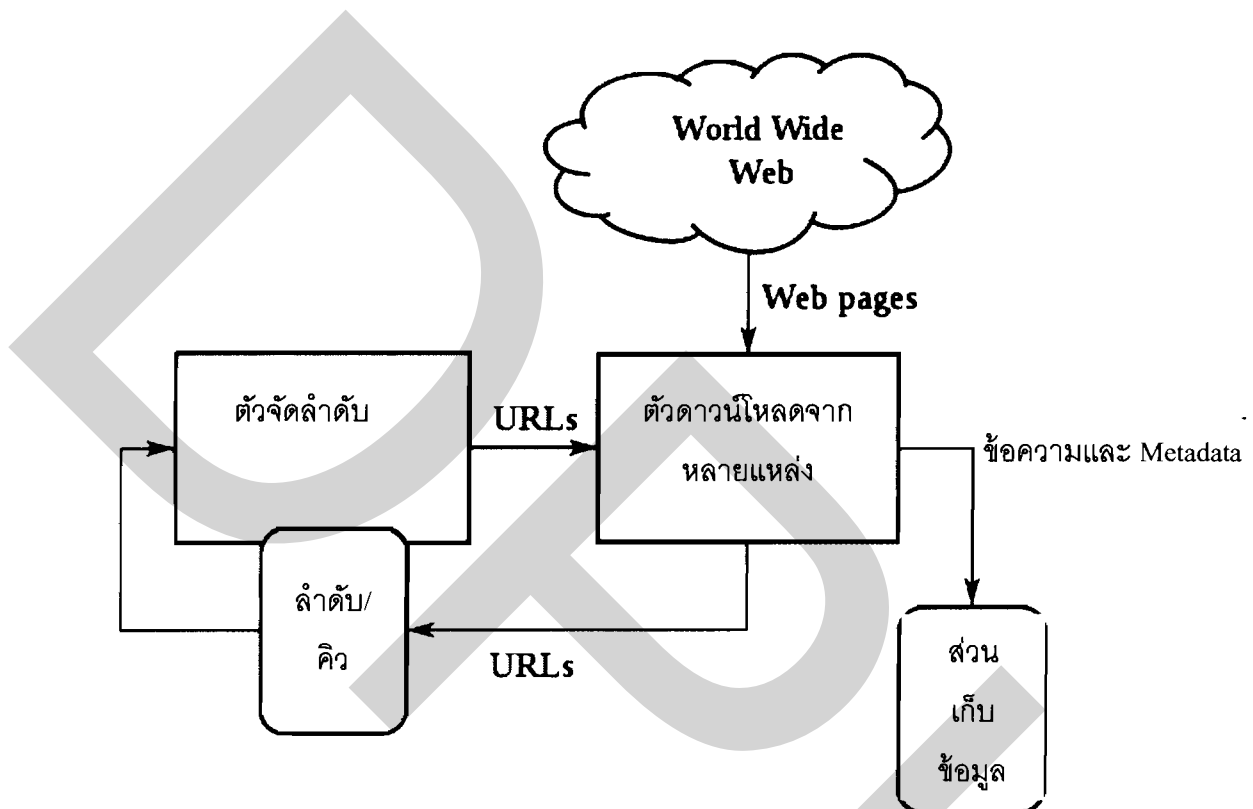
หลักฐานเรื่องราวจากบันทึกการเข้าถึง แสดงให้เห็นว่า ช่วงเวลาในการเข้าถึงจาก Crawlers ที่เป็นที่ยุติกันอยู่ระหว่าง 20 วินาทีและ 3-4 นาที เป็นที่น่าสังเกตว่า แม้ว่าจะสุภาพแล้ว และได้ดำเนินการตามมาตรการป้องกันต่างๆแล้วเพื่อหลีกเลี่ยงการ overload เซิร์ฟเวอร์ แต่ก็ยังได้รับคำร้องเรียนจากผู้ดูแลเว็บเซิร์ฟเวอร์ Brin and Page สังเกตว่า "...การใช้งาน Crawler ที่เชื่อมต่อกับเซิร์ฟเวอร์กว่าครึ่งล้าน (...) ทำให้มีอีเมลล์และโทรศัพท์เข้ามาว่าจำนวนหนึ่งเลยทีเดียว เนื่องจากมีคนออนไลน์อยู่มาก แน่แน่นอนว่ามีคนที่ไม่รู้ว่ crawler คืออะไร เพราะเป็นครั้งแรกที่พวกเขาเจอกับมัน (Web crawling, Retrieved 2012, January23)

#### 2.3.1.4 นโยบายดำเนินการกู่ขนาน

นโยบายการไม่ทำให้เกิดการขนาน ที่บ่งบอกว่าจะประสานงานกับ Web Crawler ต่างๆ ที่ส่งออกไปอย่างไร Crawler กู่ขนานคือ Crawler ที่ทำงานหลายๆ กระบวนการขนานกันไป เป้าหมายคือเพื่อให้ได้อัตราดาวน์โหลดสูงสุด ในขณะที่ลดค่าใช้จ่ายจากการทำงานกู่ขนานลงให้ต่ำที่สุด และหลีกเลี่ยงการดาวน์โหลดหน้าเดียวกันซ้ำๆ ในการหลีกเลี่ยงการดาวน์โหลดหน้าเดิมมากกว่าหนึ่งครั้งนั้น ระบบการ Crawling ต้องการนโยบายในการระบุ URLs ใหม่ๆ ที่พบเจอ

ระหว่างกระบวนการ Crawling เนื่องจากอาจจะเจอ URL เดียวกันได้อีก ในกระบวนการ Crawling อีกกระบวนการหนึ่ง (Web crawling, Retrieved 2012, January

### 2.3.2 โครงสร้าง



ภาพที่ 2.5 โครงสร้างขั้นสูงของ Web Crawler พื้นฐาน

ที่มา: Wikipedia ([http://en.wikipedia.org/wiki/Web\\_crawling](http://en.wikipedia.org/wiki/Web_crawling))

Crawler นั้นไม่เพียงแต่จะต้องมียุทธศาสตร์การ crawling ที่คิดงัดที่ได้กล่าวไว้ในก่อนหน้านี เท่านั้น แต่ยังต้องมีโครงสร้างที่เหมาะสมเป็นอย่างสูงอีกด้วย Shkapenyuk และ Suel ตั้งข้อสังเกตว่า (Web crawling, Retrieved 2012, January 23)

ในขณะที่การสร้าง crawler ที่ทำงานช้า ดาวน์โหลดไม่ก็หน้าต่อวินาที เป็นเวลาสั้นๆ นั้นค่อนข้างง่าย การสร้างระบบที่มีประสิทธิภาพสูงที่สามารถดาวน์โหลดหน้าเว็บเป็นร้อยล้านหน้า ในระยะเวลาหลายสัปดาห์นั้น มีความยากลำบากและเป็นความท้าทายต่อการออกแบบระบบ I/O และประสิทธิภาพของเครือข่าย และความสามารถในการบริหารจัดการได้ (Web crawling, Retrieved 2012, January 23)

Web Crawler นั้นเป็นหัวใจสำคัญของเครื่องมือค้นหา และรายละเอียดเกี่ยวกับ algorithm และ โครงสร้างนั้นถูกเก็บเป็นความลับทางธุรกิจ เมื่อการออกแบบ Crawler ได้รับการเปิดเผย มักจะมีรายละเอียดที่สำคัญขาดหายไปอยู่บ่อยๆ ซึ่งจะป้องกันไม่ให้คนอื่นลอกเลียนแบบได้ ยังมีความกังวลเกี่ยวกับ “Search Engine Spamming” เกิดขึ้นมา ซึ่งสิ่งนี้จะป้องกันไม่ให้เครื่องมือค้นหาใหญ่ๆ ลงอันดับ algorithm ของพวกเขา (Web crawling, Retrieved 2012, January23)

### 2.3.3 การระบุตัวตนCrawler

Web Crawler นั้นมักจะระบุตัวตนของตนเองต่อ Web Server โดยการใช้ส่วน User-agent ของคำขอ HTTP ผู้ดูแลเว็บไซต์มักจะตรวจสอบบันทึก web server และใช้ส่วน user agent เพื่อระบุว่า crawler ใดได้เข้าเยี่ยมชมเว็บไซต์ และเข้าชมบ่อยเพียงใด ส่วน user agent นั้นอาจจะใส่ URL ไว้ด้วยซึ่งผู้ดูแลเว็บไซต์อาจจะสามารถเข้าไปหาข้อมูลเพิ่มเติมเกี่ยวกับ Crawler ได้ Spambots และ Web Crawler ตัวร้ายอื่นๆ มักจะไม่ใส่ข้อมูลระบุตัวตนลงไปในส่วน user agent หรือพวกมันอาจจะปิดบังตัวตนที่แท้จริงอยู่ โดยใช้ browser หรือ crawler ที่เป็นที่รู้จักดีตัวอื่นมาบังหน้า (Web crawling, Retrieved 2012, January23)

การที่ Web Crawler ระบุตัวตนของตัวเองนั้นเป็นสิ่งสำคัญ เพื่อให้ผู้ดูแลเว็บไซต์สามารถติดต่อเจ้าของได้หากจำเป็น ในบางกรณี Crawler อาจจะถูกล็อกคักติดอยู่ใน Crawler Trap หรือพวกมันอาจจะกำลัง Overload เซิร์ฟเวอร์ด้วยคำขอต่างๆอยู่และเจ้าของมีความจำเป็นจะต้องหยุด Crawler การแสดงตัวตนนั้นยังมีประโยชน์สำหรับผู้ดูแลที่สนใจที่จะรู้ว่าพวกเขาจะคาดหวังให้หน้าเว็บต่างๆ ได้รับการจัดเข้าคิรอานโดยเครื่องมือค้นหาต่างๆ เมื่อใด (Web crawling, Retrieved 2012, January23)

### 2.3.4 ตัวอย่าง Crawler ทั่วไป

รายการต่อไปนี้เป็นรายการโครงสร้างของ Crawler ที่ได้รับการเผยแพร่แล้ว ซึ่งเป็นโครงสร้างสำหรับ Crawler ทั่วไป (ยกเว้น Crawler ที่เน้นเฉพาะเว็บไซต์) พร้อมคำอธิบายสั้นๆ ที่รวมถึงชื่อต่างๆ สำหรับส่วนประกอบต่างๆ และลักษณะเด่น:

2.3.4.1 Yahoo! Slurp เป็นชื่อของ Crawler ค้นหาของ Yahoo (Web crawling, Retrieved 2012, January23)

2.3.4.2 Bingbot เป็นชื่อของ Web Crawler ของ Bing แห่ง Microsoft ซึ่งมาแทน MSNbot (Web crawling, Retrieved 2012, January23)

2.3.4.3 FAST Crawler เป็น Crawler ที่กระจายไปทั่ว ถูกใช้งานโดยFast Search & Transfer, และคำอธิบายทั่วไปของโครงสร้าง (Web crawling, Retrieved 2012, January23)

2.3.4.4 Googlebot นั้นได้รับการอธิบายอย่างละเอียด แต่แหล่งอ้างอิงนั้นมีถึงแค่โครงสร้างช่วงแรกๆ ซึ่งอยู่บนพื้นฐานของ C++ และ Python Crawler นั้นได้รับการบูรณาการเข้ากับกระบวนการจัดทำดัชนี เนื่องจากมีการใช้การวิเคราะห์ข้อความกับการจัดทำดัชนีข้อความทั้งหมดและใช้การวิเคราะห์ข้อความในการดึงเอา URL ออกมาด้วย มีเซิร์ฟเวอร์ URL ที่ส่งรายการ URL ที่จะต้องถูกดึงมาโดยกระบวนการ Crawling ต่างๆ ระหว่างกระบวนการวิเคราะห์ข้อความ URL ที่พบ ถูกส่งต่อไปยังเซิร์ฟเวอร์ URL ที่ทำการตรวจสอบว่า URL นั้นได้ผ่านตามาก่อนหรือไม่ หากไม่ URL จะถูกเพิ่มเข้าไปในคิวของเซิร์ฟเวอร์ URL (Web crawling, Retrieved 2012, January23)

2.3.4.5 PolyBot เป็น Crawler กระจายไปทั่ว เขียนขึ้นโดยใช้ C++ และ Python ซึ่งมีส่วนประกอบคือ “ตัวบริหารการ Crawl- Crawl Manager” “ตัวดาวน์โหลด” หนึ่งตัวหรือมากกว่า และ “DNS Resolvers” หนึ่งตัวหรือมากกว่า URL ที่ถูกรวบรวมมาจะถูกเพิ่มเข้าไปในคิวบนดิสก์ และจะถูกดำเนินการต่อไป เพื่อค้นหา URL ที่เคยพบเห็นแล้วในโหมด Batch นโยบายความสุภาพพิจารณาโดเมนทั้งในระดับที่สามและระดับที่สอง (เช่น: www.example.com และ www2.example.com เป็นโดเมนระดับที่สาม) เนื่องจากโดเมนระดับที่สามนั้นมักจะมีโฮสต์เป็นเซิร์ฟเวอร์เว็บตัวเดียวกัน (Web crawling, Retrieved 2012, January23)

2.3.4.6 RBSE เป็น Web Crawler ที่ได้รับการเผยแพร่ตัวแรก มันอยู่บนพื้นฐานของสองโปรแกรม โปรแกรมแรก “spider” นั้นจะรักษาคิว สัมพันธ์กับฐานข้อมูล และ โปรแกรมที่สอง “mite” เป็น browser ของ www.ASCII ที่ได้รับการปรับปรุงแล้วที่ดาวน์โหลดหน้าเว็บต่างๆ จากเว็บไซต์ (Web crawling, Retrieved 2012, January23)

2.3.4.7 WebCrawler ใช้งานเพื่อสร้างดัชนีข้อความทั้งหมดที่เผยแพร่ต่อสาธารณะ ดรรชนีแรก มันอยู่บนพื้นฐานของ lib-WWW เพื่อใช้ดาวน์โหลดหน้าเว็บและโปรแกรมอื่นเพื่อวิเคราะห์และสั่งการ URLs สำหรับการสำรวจ breadth-first ของ Web Graph มันยังรวมเอา Crawler แบบ real-time ที่จะติดตามลิงค์ (การเชื่อมโยง) ต่างๆ บนพื้นฐานของความคล้ายคลึงกันของข้อความ anchor (anchor text) กับคำถามที่ต้องการค้นหา (Web crawling, Retrieved 2012, January23)

2.3.4.8 World Wide Web Worm เป็น Crawler ที่ถูกใช้ในการสร้างดรรชนีอย่างง่ายของชื่อเอกสาร และ URLs สามารถค้นหาดรรชนีได้โดยการใช้คำสั่ง grepUnix (Web crawling, Retrieved 2012, January23)

2.3.4.9 WebFountain เป็น Crawler กระจายไปทั่ว และเป็น Crawler แบบหน่วย (Modular Crawler) คล้ายๆกับ Mercator แต่ถูกเขียนขึ้นโดยใช้ C++ มีตัว “ควบคุม” ที่ประสานงานกับหน่วย “มด” ต่างๆ หลังจากที่ได้ดาวน์โหลดหน้าต่างๆซ้ำๆกันแล้ว อัตราการเปลี่ยนแปลงจะถูกอนุมานสำหรับหน้าแต่ละหน้า และจะใช้วิธีการ non-linear programming ในการแก้ระบบสมการเพื่อให้ค่าความสดใหม่สูงที่สุด ผู้เขียนแนะนำให้ใช้คำสั่ง Crawling ชนิดนี้ในขั้นตอนแรกๆ ของการ Crawl จากนั้นจึงเปลี่ยนไปใช้ คำสั่ง Crawling ที่เป็นมาตรฐาน ซึ่งหน้าต่างๆ จะได้รับการเข้ากันคู่ด้วยอัตราความถี่เท่าๆกัน (Web crawling, Retrieved 2012, January23)

2.3.4.10 WebRACE เป็นหน่วย Crawling และ Caching ที่ใช้งานใน Java และใช้งานเป็นส่วนหนึ่งของระบบ Generic ที่เรียกว่า eRACEระบบนี้จะรับคำขอจากผู้ใช้ในการดาวน์โหลดหน้าเว็บ ดังนั้น Crawler จะทำหน้าที่ส่วนหนึ่งเป็นเซิร์ฟเวอร์ Smart Proxy ระบบยังต้องรับคำขอสำหรับการ “subscriptions” กับหน้าเว็บต่างๆที่จะต้องถูกติดตาม: เมื่อหน้าเว็บมีการเปลี่ยนแปลงหน้าเว็บเหล่านั้นจะต้องถูกดาวน์โหลดโดย Crawler และตัวรับการ subscription จะต้องได้รับแจ้ง ลักษณะที่โดดเด่นที่สุดของ WebRACEคือ ในขณะที่ Crawler ส่วนใหญ่เริ่มต้นจากชุด URLs “Seed” WebRACEจะได้รับ URLs เริ่มต้นใหม่ๆ อย่างต่อเนื่อง เพื่อใช้ในการเริ่มการ Crawl (Web crawling, Retrieved 2012, January23)

### 2.3.5 ตัวอย่าง Open-source crawlers

นอกเหนือจากโครงสร้าง Crawler เฉพาะ ในรายการข้างต้นแล้ว ยังมีโครงสร้าง Crawler ทั่วไป ที่เผยแพร่โดย Cho and Chakrabarti อีก (Web crawling, Retrieved 2012, January23)

2.3.5.1 Aspeek เป็น Crawler เป็นตัวจัดทำดัชนี และเป็นเครื่องมือค้นหาที่เขียนขึ้นใน C++ และจดลิขสิทธิ์ภายใต้ GPL (Web crawling, Retrieved 2012, January23)

2.3.5.2 DataparkSearch เป็น Crawler และเครื่องมือค้นหาที่ออกมาภายใต้ GNU General Public License (ลิขสิทธิ์สาธารณะทั่วไป) (Web crawling, Retrieved 2012, January23)

2.3.5.3 GNU Wget เป็น Crawler ที่ทำงานด้วยสายคำสั่ง เขียนขึ้นใน C และออกสู่สาธารณะภายใต้ GPL มักใช้ในการ mirror เว็บและ FTP sites (Web crawling, Retrieved 2012, January23)

2.3.5.4 GNU Wget เป็น Crawler ที่ทำงานด้วยสายคำสั่ง เขียนขึ้นใน C และออกสู่สาธารณะภายใต้ GPL มักใช้ในการ mirror เว็บและ FTP sites (Web crawling, Retrieved 2012, January23)

2.3.5.5 GRUB เป็น Crawler ค้นหาที่กระจายทั่วไปและเป็น open source ซึ่ง Wikia Searchใช้ในการค้นหาเว็บไซต์ (Web crawling, Retrieved 2012, January23)

2.3.5.6 Heritrix เป็น Crawler ที่มีคุณภาพในการเก็บข้อมูลของ Internet Archive ถูกออกแบบมาเพื่อให้เก็บข้อมูลบางส่วนของเว็บไซต์ต่างๆ ซึ่งมีจำนวนมาก เขียนขึ้นใน Java (Web crawling, Retrieved 2012, January23)

2.3.5.7 ht://Dig นั้นรวมเอา Web Crawler ไว้ในเครื่องมือการจัดทำดัชนี includes a Web crawler in its indexing engine. (Web crawling, Retrieved 2012, January23)

2.3.5.8 HTTrack ใช้ Web Crawler ในการสร้าง mirror ของเว็บไซต์สำหรับการเข้ามาดูแบบ off-line เขียนขึ้นใน C และออกสู่สาธารณะภายใต้ GPL (Web crawling, Retrieved 2012, January23)

2.3.5.9 ICDL Crawler เป็น Web Crawler ข้ามแพลตฟอร์ม ที่เขียนขึ้นใน C++ และตั้งใจที่จะ Crawl เว็บไซต์ต่างๆบนพื้นฐานของ Web-site Parse Templates โดยใช้ทรัพยากร CPU ที่ว่างของคอมพิวเตอร์เท่านั้น (Web crawling, Retrieved 2012, January23)

2.3.5.10 mnoGoSearch เป็น Crawler เป็นตัวจัดทำดัชนี และเป็นเครื่องมือค้นหาที่เขียนขึ้นใน C และจดลิขสิทธิ์ภายใต้ GPL (เครื่อง Linux เท่านั้น) (Web crawling, Retrieved 2012, January23)

2.3.5.11 Nutch เป็น Crawler เขียนขึ้นใน Java และออกภายใต้ลิขสิทธิ์ Apache สามารถนำไปใช้ร่วมกับแพ็คเกจการจัดทำดัชนีข้อความ Lucene ได้ (Web crawling, Retrieved 2012, January23)

2.3.5.12 Open Search Server เป็นเครื่องมือค้นหาและซอฟต์แวร์ Web Crawler ที่ออกภายใต้ GPL (Web crawling, Retrieved 2012, January23)

2.3.5.13 Pavuk เป็นเครื่องมือสายคำสั่งสำหรับ Web Mirror พร้อมด้วยตัวเลือก Crawler X11 GUI และออกมาภายใต้ GPL มันมีลักษณะขั้นสูงหลายอย่าง เมื่อเทียบกับ wget และ httrack เช่น มีกฎการกรองและการสร้างไฟล์ที่แสดงออกมาอย่างสม่ำเสมอ เป็นต้น (Web crawling, Retrieved 2012, January23)

2.3.5.14 PHP-Crawler เป็น Crawler บนพื้นฐานของ PHP และ MySQL อย่างง่าย ที่ออกภายใต้ BSD ติดตั้งง่าย และเป็นที่ยอมรับสำหรับเว็บไซต์ที่ขับเคลื่อนโดย MySQL ตัวเล็กๆ ที่อยู่บน Host ร่วม(shared hosting) (Web crawling, Retrieved 2012, January23)

2.3.5.15 tkWWW Robot เป็น Crawler ที่อยู่บนพื้นฐานของ Web Browser tkWWW (จดลิขสิทธิ์ภายใต้ GPL) (Web crawling, Retrieved 2012, January23)

2.3.5.16 YaCy เป็นเครื่องมือค้นหากระจายไปทั่วซึ่งฟรี สร้างขึ้นบนหลักการเครือข่าย peer-to-peer (จดลิขสิทธิ์ภายใต้ GPL) (Web crawling, Retrieved 2012, January23)

2.3.5.17 Seeks เป็นเครื่องมือค้นหากระจายไปทั่วซึ่งฟรี (จดลิขสิทธิ์ภายใต้ Affero General Public License) (Web crawling, Retrieved 2012, January23)

#### 2.3.6 การCrawling เว็บไซต์เชิงลึก

หน้าเว็บจำนวนมากนั้นอยู่ในกลุ่ม เว็บไซต์ลึกหรือเว็บไซต์ที่มองไม่เห็น หน้าเว็บเหล่านี้มักจะเข้าถึงได้โดยการส่งคำค้นหาไปที่ฐานข้อมูล และ crawler ธรรมดาไม่น่าจะสามารถที่จะหาหน้าเหล่านี้ได้หากไม่มีการเชื่อมโยง (ลิงค์) ที่ชี้ไปที่พวกมัน โพรโตคอล Sitemap ของ Google และ mod oai นั้นตั้งใจที่จะอนุญาตให้มีการค้นหาทรัพยากรเว็บเชิงลึกเหล่านี้ (Web crawling, Retrieved 2012, January23)

การ Crawling เว็บเชิงลึกนั้นเพิ่มจำนวนการเชื่อมโยงเว็บต่างๆที่จะต้อง Crawl ขึ้นเป็นทวีคูณ บาง Crawler เพียงแค่นำเอา URL ที่หน้าตาแบบนี้ <a href="URL" ไปเท่านั้น ในบางกรณี เช่น Googlebot การ Web Crawling นั้นจะกระทำกับข้อความทั้งหมดที่มีอยู่ภายในเนื้อหา hypertext, tags หรือ text (ข้อความ) (Web crawling, Retrieved 2012, January23)

#### 2.3.7 การนำการCrawling Web 2.0 ไปใช้

2.3.7.1 Sheeraj Shah แสดงให้เห็นข้อมูลเชิงลึกของการนำการ Crawling Web 2.0 ที่ขับเคลื่อนโดย Ajax ไปใช้งาน (Web crawling, Retrieved 2012, January23)

2.3.7.2 ผู้อ่านที่สนใจอาจจะต้องการอ่าน AJAXSearch: Crawling, Indexing and Searching Web 2.0 Applications (Web crawling, Retrieved 2012, January23)

2.3.7.3 Making AJAX Applications Crawable, จาก Google Code เอกสารนี้ให้คำจำกัดความข้อขัดแย้งระหว่าง web servers และ crawlers เครื่องมือค้นหา ที่อนุญาตให้เนื้อหาที่มีการสร้างอย่างต่อเนื่องนั้นมองเห็นได้โดย Crawler โดย Google ก็สนับสนุนข้อขัดแย้งนี้อยู่ในปัจจุบัน (Web crawling, Retrieved 2012, January23)

### 2.4 การจัดทำดัชนีของเครื่องมือค้นหา

การจัดทำดัชนีของเครื่องมือค้นหา นั้น รวบรวม วิเคราะห์ และเก็บข้อมูลเพื่อช่วยให้เกิดการดึงข้อมูลอย่างรวดเร็วและแม่นยำ การออกแบบดัชนีนั้น รวมเอาหลักการจากหลายๆ ศาสตร์เข้าด้วยกัน ทั้งภาษาศาสตร์ จิตวิทยาการรู้การเข้าใจ คณิตศาสตร์ สุนทรศาสตร์ ฟิสิกส์ และวิทยาศาสตร์คอมพิวเตอร์ ชื่อที่เวียนเปลี่ยนกันไปสำหรับกระบวนการ ในบริบทของเครื่องมือค้นหาที่ได้รับการออกแบบมาเพื่อค้นหาหน้าเว็บบนอินเทอร์เน็ตนั้น คือ การจัดทำดัชนีเว็บ (Web Indexing) (Search engine indexing, Retrieved 2012, January23)



เครื่องมือที่เป็นที่นิยมนั้นเน้นไปที่การจัดทำดัชนีข้อความทั้งหมด ของเอกสารออนไลน์ และใช้ภาษาธรรมชาติ ประเภทของสื่อ อาทิ วิดีโอและเสียงและกราฟิกนั้นก็สามารถค้นหาได้เช่นกัน (Search engine indexing, Retrieved 2012, January23)

Meta Search Engine (เครื่องมือค้นหา Meta) นั้นจะใช้งานดัชนีของการบริการอื่นๆ ซ้ำอีก และจะไม่เก็บดัชนีท้องถิ่น (local index) ไว้ ในขณะที่เครื่องมือค้นหาบนพื้นฐานของ cache นั้นจะเก็บข้อมูลดัชนีไว้ตลอด พร้อมกับ corpus ด้วยการบริการข้อความบางส่วน จะจำกัดความลึกที่จัดทำดัชนี เพื่อลดขนาดดัชนีลง ไม่เหมือนกับดัชนีข้อความทั้งหมด การบริการที่มีขนาดใหญ่ขึ้น มักจะดำเนินการในห้วงเวลาที่ระบุไว้ก่อนหน้าแล้ว เนื่องจากระยะเวลาที่ต้องการและต้นทุนในการดำเนินการ ในขณะที่เครื่องมือค้นหาที่อยู่บนพื้นฐานของ agent (ตัวแทน) จะจัดทำดัชนีแบบทันที (real-time) (Search engine indexing, Retrieved 2012, January23)

#### 2.4.1 การจัดทำดัชนี

วัตถุประสงค์ของการจัดเก็บดัชนีก็เพื่อให้ความเร็วและการทำงานมีประสิทธิภาพสูงสุด ในการค้นหาเอกสารที่เกี่ยวข้องกับคำค้นหา หากไม่มีดัชนี เครื่องมือค้นหาจะต้องค้นหาเอกสารทุกเอกสารใน คลังข้อมูล ซึ่งกินเวลาและพลังงานคอมพิวเตอร์มาก ตัวอย่างเช่น ดรรชนีของเอกสาร 10,000 เอกสารนั้นสามารถค้นหาได้ภายในเวลาไม่กี่วินาที แต่การสแกนหาคำทุกคำ ในเอกสารใหญ่ๆ 10,000 เอกสาร อาจใช้เวลานานหลายชั่วโมง ต้นทุนที่เก็บข้อมูลคอมพิวเตอร์เพิ่มเติมเพื่อเก็บรักษาดัชนี เช่นเดียวกับเวลาที่เพิ่มมากขึ้นเพื่อให้มีการปรับข้อมูล (อัปเดต) ข้อมูลให้เป็นปัจจุบัน สิ่งเหล่านี้ เป็นสิ่งที่จำเป็นจะต้องแลกกับเวลาที่น้อยลงในระหว่างการดึงข้อมูล (Search engine indexing, Retrieved 2012, January23)

##### 2.4.1.1 ปัจจัยในการออกแบบดัชนี

ปัจจัยหลักๆ ในการออกแบบโครงสร้างเครื่องมือค้นหานั้นมีดังนี้:

###### 1) ปัจจัยผสมผสาน

ข้อมูลเข้าไปอยู่ในดัชนีได้อย่างไร หรือ คำต่างๆ หรือ ลักษณะของหัวข้อถูกเพิ่มเข้าไปในดัชนีในระหว่างการเชื่อมโยงข้ามไปมาระหว่างคลังข้อความ และดัชนีที่มีหลายดัชนีนั้นสามารถทำงานอย่างอิสระ (asynchronously) ได้หรือไม่ ตัวจัดทำดัชนีจะต้องตรวจสอบเสียก่อนว่ามันกำลังทำข้อมูลเนื้อหาทำให้ทันสมัยหรือกำลังเพิ่มข้อมูลใหม่เข้าไป การเชื่อมโยงข้ามไปมา (traversal) นั้นมักจะสัมพันธ์กับนโยบายการรวบรวมข้อมูล การผสมผสานดัชนีเครื่องมือค้นหานั้นมีหลักการคล้ายๆกับคำสั่งการผสมผสาน SQL และอัลกอริทึมการผสมผสานอื่นๆ (Search engine indexing, Retrieved 2012, January23)

## 2) วิธีการเก็บรักษาข้อมูล

วิธีการเก็บรักษาข้อมูลครรรชนี นั้นคือ ข้อมูลควรจะเป็นข้อมูลที่ถูกลบอัดหรือถูกกรองขนาดของครรรชนี (Search engine indexing, Retrieved 2012, January23)

ครรรชนีต้องการหน่วยความจำคอมพิวเตอร์มากเท่าไรในการสนับสนุนความเร็วในการค้นหา (Search engine indexing, Retrieved 2012, January23)

ความสามารถในการค้นหาหนึ่งคำในครรรชนีผกผัน ความเร็วในการค้นหาข้อมูลในโครงสร้างข้อมูล เปรียบเทียบกับความเร็วที่สิ่งนั้นจะได้รับการปรับปรุงให้ทันสมัย หรือถูกลบออก เหล่านี้เป็นหัวใจสำคัญของวิทยาศาสตร์คอมพิวเตอร์ (Search engine indexing, Retrieved 2012, January23)

## 3) การดูแลรักษา

ครรรชนีจะได้รับการดูแลรักษาอย่างไร (Search engine indexing, Retrieved 2012, January23)

## 4) ความทนทานต่อการทำงานที่ผิดพลาด

ความสำคัญของการบริการที่จำเป็นจะต้องเชื่อถือได้ ปัญหาต่างๆ นั้นมีตั้งแต่การจัดการกับครรรชนีที่มีปัญหา การระบุว่าข้อมูลไม่ดี (Bad Data) สามารถจัดการได้เดี๋ยวนั้นหรือไม่ การจัดการกับฮาร์ดแวร์ไม่ดี การทำพาร์ทิชัน และรายการต่างๆ เช่น การทำพาร์ทิชันแบบ hash-based หรือ composite เช่นเดียวกับกระบวนการ replication (Search engine indexing, Retrieved 2012, January23)

### 2.4.1.2 โครงสร้างข้อมูลครรรชนี

โครงสร้างเครื่องมือค้นหานั้นมีความหลากหลายในวิธีการในการดำเนินการจัดทำครรรชนีและวิธีการในการเก็บรักษาครรรชนีให้เป็นไปตามปัจจัยการออกแบบทั้งหลาย (Search engine indexing, Retrieved 2012, January23) ประเภทของครรรชนีมีดังนี้

#### 1) Suffix Tree

โครงสร้างเปรียบเทียบได้กับต้นไม้ สนับสนุนการค้นหาแบบ linear time (เวลาเส้นตรง) สร้างขึ้นโดยการเก็บส่วนเสริมท้าย (suffix) ของคำแต่ละคำ Suffix Tree นั้นเป็นประเภทหนึ่งของ Trie โดย Trie นั้นสนับสนุนการ hashing ที่ยืดหยุ่นได้ ซึ่งเป็นสิ่งสำคัญในการจัดทำครรรชนีของเครื่องมือค้นหา โครงสร้างนี้ใช้สำหรับการค้นหาแบบแผนในลำดับ DNA และการรวมกลุ่มกันข้อเสียเปรียบใหญ่ๆ ก็คือการเก็บคำใน tree อาจจะต้องการพื้นที่มากกว่าการเก็บคำหลายๆ การนำเสนออีกแบบหนึ่ง อยู่ในรูป suffix array (การจัดเรียง ส่วนเสริมท้าย) ซึ่งว่ากันว่า ต้องการ

หน่วยความจำเสมือน (virtual memory) ที่น้อยกว่า และสนับสนุนการบีบอัดข้อมูลเช่น อัลกอริทึม BWT (Search engine indexing, Retrieved 2012, January23)

## 2) วรรณียศผผัน

เก็บรายการการปรากฏขึ้นของเกณฑ์การค้นหาย่อยแต่ละเกณฑ์ โดยมาก มักจะอยู่ในรูปแบบตาราง hash หรือ binary tree (Search engine indexing, Retrieved 2012, January23)

## วรรณียศอ้างอิง

เก็บการอ้างอิงหรือการเชื่อมโยงหลายมิติ ระหว่างเอกสารต่างๆ เพื่อสนับสนุนการวิเคราะห์การอ้างอิง โดยจะเป็นไปตาม Bibilometrics (Search engine indexing, Retrieved 2012, January23)

## 3) วรรณียศ Ngram

เก็บลำดับของความยาวของข้อมูลเพื่อสนับสนุนการดึงข้อมูลหรือการค้นหาข้อความ (Search engine indexing, Retrieved 2012, January23)

## 4) Document-term matrix

ใช้ในการวิเคราะห์แบบ latent semantic เก็บการปรากฏขึ้นของคำต่างๆ ในเอกสารในรูปแบบของตารางวิเคราะห์สองมิติ (two-dimensional sparse matrix) (Search engine indexing, Retrieved 2012, January23)

### 2.4.1.3 ความท้าทายในการทำงานกลุ่มงาน

ความท้าทายหลักในการออกแบบเครื่องมือค้นหาคือการบริหารจัดการกระบวนการคอมพิวเตอร์ที่ต่อเนื่องกัน อาจจะมีโอกาสมากมายที่จะเกิดภาวะการแข่งขันและข้อผิดพลาดซึ่งสอดคล้องกัน ตัวอย่างเช่น เอกสารใหม่ได้ถูกเพิ่มเข้าไปในคลังและวรรณียศจะต้องได้รับการปรับปรุงให้ทันสมัย แต่ในขณะเดียวกันวรรณียศนั้นก็ยังต้องตอบสนองต่อคำค้นหาต่างๆ ที่ถูกป้อนเข้ามา นี่เป็นการชนกันของงานที่ทำงานแข่งกันสองงาน ลองคิดว่า ผู้เขียนเป็นผู้ผลิตข้อมูล และ web crawler เป็นลูกค้าของข้อมูลนี้ ที่ค้นหาข้อความ และนำมาเก็บไว้ใน cache (หรือคลัง-corpus) วรรณียศก้าหน้าคือลูกค้าของข้อมูลที่ผลิตขึ้น โดยคลัง และวรรณียศผผันคือลูกค้าของข้อมูลที่ผลิตขึ้นจากวรรณียศก้าหน้า สถานการณ์เช่นนี้มักจะถูกเรียกว่า โมเดลผู้ผลิต-ลูกค้า (producer-consumer model) ตัวจัดทำวรรณียศเป็นผู้ผลิตข้อมูลที่สามารถค้นหาได้ และผู้ใช้ก็คือลูกค้าที่ต้องการจะค้นหา ความท้าทายนั้นเพิ่มขึ้นเมื่อทำงานกับหน่วยเก็บข้อมูลและกระบวนการที่กระจายไปทั่วในความพยายามที่จะประมวลผลด้วยข้อมูลที่ได้รับการจัดทำวรรณียศที่มีอยู่มากกว่า โครงสร้างของเครื่องมือค้นหาจะต้องนำ การประมวลผลแบบกระจาย มาใช้ ซึ่งในการประมวลผลแบบนี้ นั้น เครื่องมือค้นหาจะประกอบไปด้วยเครื่องมือทำงานหลายเครื่องทำงานด้วยกันอย่างพร้อมเพียง

กัน นี้จะช่วยเพิ่มความเป็นไปได้ของการเกิดปัญหาความไม่สอดคล้องกัน และทำให้มีความยากลำบากมากขึ้นในการดูแลรักษาโครงสร้างที่ทำงานพร้อมกัน กระจาย และคู่ขนานกัน (Search engine indexing, Retrieved 2012, January23)

#### 2.4.1.4 ธรรมชาติผกผันต่างๆ

เครื่องมือค้นหาหลายเครื่องมือได้นำเอาธรรมชาติผกผันเข้ามาใช้ด้วย เมื่อทำการประเมินคำค้นหา ให้สามารถระบุตำแหน่งของเอกสารที่มีคำต่างๆ ที่ต้องการค้นอยู่และจัดอันดับเอกสารเหล่านี้ตามเกณฑ์ความเกี่ยวข้อง เนื่องจากธรรมชาติผกผันนี้จัดเก็บรายการเอกสารต่างๆ ที่มีคำแต่ละคำอยู่ เครื่องมือค้นหาสามารถใช้การเข้าถึงโดยตรง (direct access) ในการค้นหาเอกสารที่เกี่ยวข้องกับคำแต่ละคำที่ค้นหา เพื่อให้สามารถดึงเอกสารที่เกี่ยวข้องออกมาได้อย่างรวดเร็ว ต่อไปนี้คือตัวอย่างประกอบง่ายๆ ของธรรมชาติผกผัน (Search engine indexing, Retrieved 2012, January23)

ตารางที่ 2.1 ธรรมชาติผกผัน

คำ	เอกสาร
The	เอกสาร 1, เอกสาร3, เอกสาร 4, เอกสาร 5
Cow	เอกสาร 2, เอกสาร 3, เอกสาร 4
Says	เอกสาร 5
Moo	เอกสาร 7

ที่มา : Wikipedia ([http://en.wikipedia.org/wiki/Index\\_\(search\\_engine\)](http://en.wikipedia.org/wiki/Index_(search_engine)))

ธรรมชาตินี้สามารถระบุได้เพียงคำว่า คำคำนั้นปรากฏอยู่ในเอกสารนั้นๆ หรือไม่ เนื่องจากว่ามันไม่ได้เก็บข้อมูลเกี่ยวกับความถี่และตำแหน่งของคำเอาไว้ จึงถือกันว่าเป็นธรรมชาติ Boolean ธรรมชาติรูปแบบนี้จะระบุว่าเอกสารใดตรงกับคำค้นหาแต่จะไม่เรียงลำดับเอกสารที่ตรงกันให้ ในการออกแบบบางชิ้น ธรรมชาติจะรวมเอาข้อมูลเพิ่มเติมเช่นความถี่ของคำแต่ละคำในเอกสารแต่ละชิ้น หรือตำแหน่งของคำในเอกสารแต่ละชิ้น ข้อมูลตำแหน่งช่วยให้อัลกอริทึมการค้นหาสามารถระบุความใกล้เคียงของคำ เพื่อสนับสนุนการค้นหาวิธีต่างๆ ความถี่นั้นสามารถใช้เพื่อช่วยในการเรียงลำดับความเกี่ยวข้องของเอกสารกับคำค้นหา หัวข้อต่างๆเหล่านั้นเป็นหัวใจสำคัญในงานวิจัยในเรื่องของการดึงข้อมูล (information retrieval) (Search engine indexing, Retrieved 2012, January23)

ดัชนีผกผันเป็นเมตริกซ์ (ตาราง) วิเคราะห์ เนื่องจากคำทุกคำนั้นไม่ได้ปรากฏอยู่ในเอกสารแต่ละชิ้น เพื่อลดหน่วยความจำของคอมพิวเตอร์ที่ต้องใช้ ดรรชนีจะถูกเก็บโดยวิธีที่แตกต่างจากการจัดเรียงสองมิติ ดรรชนีนี้จะคล้ายกับ term document matrices ที่ถูกใช้งาน โดยการวิเคราะห์แบบ latent semantic ดรรชนีผกผันนั้นสามารถพิจารณาได้ว่าเป็นรูปแบบหนึ่งของตาราง hash ในบางกรณี ดรรชนีนั้นเป็นรูปแบบหนึ่งของ binary tree ซึ่งต้องการหน่วยความจำเพิ่มเติมแต่อาจย่นระยะเวลาการค้นหาลงได้ ในดรรชนีที่มีขนาดใหญ่กว่านี้ โครงสร้างนั้นจะอยู่ในรูปแบบของตารางการแจกแจง hash (distributed hash table) (Search engine indexing, Retrieved 2012, January23)

#### 2.4.1.5 การรวมดรรชนี

ดรรชนีผกผันนั้นจะถูกเพิ่มผ่านการผสมผสานหรือสร้างใหม่ การสร้างใหม่นั้นใกล้เคียงกับการผสมผสานแต่ในขั้นแรกจะลบเนื้อหาของดรรชนีผกผันก่อน โครงสร้างอาจได้รับการออกแบบเพื่อสนับสนุนการจัดทำดรรชนีเพิ่มเติมขึ้นเรื่อยๆ ส่วนการผสมผสานนั้นจะมีการระบุเอกสารที่จะมีการเพิ่มเติมหรือมีการปรับปรุงให้ทันสมัย และจะมีการวิเคราะห์เอกสารเหล่านั้นให้ได้ผลออกมาเป็นคำ เพื่อความแม่นยำทางเทคนิค การผสมผสานจะรวมเอาเอกสารที่เพิ่งจะผ่านการจัดทำดรรชนีเข้าไปด้วย โดยปกติแล้วก็จะอยู่ในหน่วยความจำเสมือน และมีcache ของดรรชนีอยู่ในฮาร์ดไดรฟ์คอมพิวเตอร์ หนึ่งหรือหลายเครื่อง (Search engine indexing, Retrieved 2012, January23)

หลังจากการวิเคราะห์ ตัวจัดทำดรรชนีจะเพิ่มเอกสารที่อ้างถึงแล้วเข้าไปในรายการเอกสารสำหรับคำต่างๆที่เหมาะสม ในเครื่องมือค้นหาที่ใหญ่กว่า กระบวนการค้นหาคำแต่ละคนในดรรชนีผกผัน (เพื่อรายงานว่ามีการปรากฏของคำนั้นในเอกสาร) อาจจะกินเวลามากเกินไป และทำให้กระบวนการนี้ มักจะถูกแบ่งออกเป็นสองส่วน ส่วนการพัฒนาของดรรชนีก้าวหน้าและกระบวนการที่แบ่งแยกเนื้อหาของดรรชนีก้าวหน้าออกไปเป็นดรรชนีผกผัน ดรรชนีผกผันได้ชื่อนี้มาเพราะว่าเป็นตัวผกผันของดรรชนีก้าวหน้า (Search engine indexing, Retrieved 2012, January23)

#### 2.4.1.6 วรรณีก้าวหน้า

วรรณีก้าวหน้าจัดเก็บรายการคำต่างๆสำหรับเอกสารแต่ละชิ้น รูปแบบต่อไปนี้เป็นรูปแบบอย่างง่ายของวรรณีก้าวหน้า:

ตารางที่ 2.2 วรรณีก้าวหน้า

เอกสาร	คำ
Document 1	the,cow,says,moo
Document 2	the,cat,and,the,hut
Document 3	the,dish,ran,away,with,the,spoon

ที่มา : Wikipedia ([http://en.wikipedia.org/wiki/Index\\_\(search\\_engine\)](http://en.wikipedia.org/wiki/Index_(search_engine)))

หลักการเบื้องหลังการพัฒนาวรรณีก้าวหน้าก็คือ ในขณะที่เอกสารต่างๆกำลังผ่านการวิเคราะห์ น่าจะเป็นการดีกว่าหากมีการจัดเก็บคำต่างๆสำหรับเอกสารแต่ละชิ้นในทันที การวาดโครงร่าง (Delineation) จะช่วยในเรื่องของกระบวนการทำงานของระบบอย่างอิสระ ซึ่งจะเลี่ยงปัญหาข้อขัดข้องของการปรับปรุงข้อมูลวรรณีก้าวหน้าให้ทันสมัยได้บ้าง วรรณีก้าวหน้านี้ถูกจัดระเบียบให้เปลี่ยนแปลงตัวเองไปเป็นวรรณีก้าวหน้า วรรณีก้าวหน้าจริงๆแล้วก็คือรายการของกลุ่มที่มีเอกสารและคำ เปรียบเทียบโดยยึดเอกสาร การเปลี่ยนจากวรรณีก้าวหน้าเป็นวรรณีก้าวหน้านี้เป็นเพียงแค่การจัดกลุ่มโดยใช้คำเท่านั้น ดังนั้น วรรณีก้าวหน้าจึงเป็นวรรณีก้าวหน้าที่มีการจัดกลุ่มคำใหม่ (Search engine indexing, Retrieved 2012, January23)

#### 2.4.1.7 การบีบอัด

การสร้างหรือการดูแลรักษาดัชนีของเครื่องมือค้นหาขนาดใหญ่ นั้นสะท้อนให้เห็นถึงความท้าทายใหญ่หลวงในเรื่องของการเก็บรักษาและการประมวลผลข้อมูล เครื่องมือค้นหามากมายใช้ประโยชน์จากรูปแบบหนึ่งของการบีบอัดเพื่อลดขนาดของดัชนีบนดิสก์ลง (Search engine indexing, Retrieved 2012, January23) ลองพิจารณาสถานการณ์ต่อไปนี้ ซึ่งเกี่ยวกับเครื่องมือค้นหาในข้อความเต็ม (full text) ในอินเทอร์เน็ต

1) จากการประมาณการณพบว่า มีหน้าเว็บที่แตกต่างกันถึง 2,000,000,000 หน้า ในปี ค.ศ. 2000 (Search engine indexing, Retrieved 2012, January23)

2) สมมติว่าแต่ละหน้าเว็บมีคำอยู่ 250 คำ (ตั้งอยู่บนสมมติฐานว่าหน้าเว็บนั้น เหมือนกับหน้าของหนังสือนิยาย) (Search engine indexing, Retrieved 2012, January23)

3) ใช้เนื้อที่ 8 บิต (หรือ 1 ไบต์) ในการเก็บอักขระหนึ่งตัว บางรหัสอักขระก็ใช้ 2 ไบต์ต่อหนึ่งอักขระ (Search engine indexing, Retrieved 2012, January23)

4) จำนวนอักขระเฉลี่ยในแต่ละคำบนหน้าเว็บ อาจจะประมาณการณืได้ว่าอยู่ที่ 5 ตัว (Search engine indexing, Retrieved 2012, January23)

5) คอมพิวเตอร์ส่วนบุคคลนั้น โดยเฉลี่ยแล้วจะมาพร้อมกับเนื้อที่ความจำ 100-250 กิกะไบต์ (Search engine indexing, Retrieved 2012, January23)

ในสถานการณ์เช่นนี้ ธรรมชาติที่ไม่มีการบีบอัด (สมมติว่าเป็นธรรมชาติที่ไม่มีการผสมผสาน และเป็นธรรมชาติอย่างง่าย) สำหรับหน้าเว็บ 2 พันล้านหน้าจะต้องจัดเก็บคำทั้งหมด 5 แสนล้านคำ ในอัตรา 1 ไบต์ต่อหนึ่งอักขระ หรือ 5 ไบต์ต่อคำ นี้จะต้องใช้เนื้อที่ 2500 กิกะไบต์ในการจัดเก็บ ซึ่งมากกว่าเนื้อที่ในฮาร์ดดิสก์ของเครื่องคอมพิวเตอร์ส่วนบุคคลเฉลี่ยถึง 25 เครื่องรวมกัน ความต้องการเนื้อที่จัดเก็บนี้อาจจะเพิ่มมากขึ้นสำหรับโครงสร้างหน่วยความจำกระจายที่ทนทานต่อข้อผิดพลาด ธรรมชาตินั้นสามารถลดขนาดลงไปได้ถึงเศษเสี้ยวของขนาดปกติ ขึ้นอยู่กับวิธีการบีบอัดที่เลือกใช้ สิ่งที่ต้องแลกมาก็คือเวลา และพลังงานในการดำเนินการ ที่จะต้องใช้ในการบีบอัด และขยายออก (Search engine indexing, Retrieved 2012, January23)

เป็นที่น่าสังเกตว่า การออกแบบเครื่องมือค้นหาขนาดใหญ่นั้นจะรวมเอาต้นทุนในการจัดเก็บเช่นเดียวกับต้นทุนค่าไฟฟ้าในการให้พลังงานที่จัดเก็บเข้าไปด้วย การบีบอัดจึงเป็นเรื่องของต้นทุน (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2 การวิเคราะห์เอกสาร

การวิเคราะห์เอกสารจะแยกส่วนประกอบ (คำ) ของเอกสารหรือสื่อในรูปแบบอื่นๆ ออกเพื่อการใส่ธรรมชาติก้าวหน้าและธรรมชาติผกผันเข้าไป คำต่างๆ ที่พบจะถูกเรียกว่า token และในบริบทของการจัดทำดัชนีเครื่องมือค้นหาและการประมวลผลภาษาธรรมชาติ การวิเคราะห์นั้นจะถูกเรียกกันโดยทั่วไปว่า tokenization บางครั้งก็ยังเรียกว่า การเพิ่มความชัดเจนให้กับขอบเขตของคำ การตัดป้าย การแบ่งข้อความออกเป็นส่วนๆ การวิเคราะห์เนื้อหา การวิเคราะห์ข้อความ การค้นหาข้อความ การสร้างความสอดคล้องกัน การแบ่งคำพูดออกเป็นส่วนๆ การ lexingหรือการวิเคราะห์คำ ศัพท์เช่น “การจัดทำดัชนี” “การวิเคราะห์” “tokenization” นั้นถูกใช้แทนกันได้ หากใช้เป็นคำพูดอย่างไม่เป็นทางการในองค์กร (Search engine indexing, Retrieved 2012, January23)

การประมวลผลภาษาธรรมชาตินั้น ในปี ค.ศ.2006 นั้นเป็นศาสตร์แห่งการวิจัยอย่างต่อเนื่องและการปรับปรุงเทคโนโลยี Tokenization นั้นเสนอให้เห็นความท้าทายมากมายในการดึงข้อมูลที่จำเป็นออกมาจากเอกสารเพื่อการค้นหาคุณลักษณะสนับสนุน การ Tokenization สำหรับการจัดทำดัชนีนั้น จะรวมเอาเทคโนโลยีที่หลากหลายเข้าไปด้วย ซึ่งการนำไปใช้นั้นมักจะถูกเก็บเป็นความลับขององค์กร (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2.1 ความท้าทายในการประมวลผลภาษาธรรมชาติ

##### 1) ความคลุมเครือของขอบเขตของคำ

ผู้ที่ใช้ภาษาอังกฤษเป็นภาษาแม่ก็น่าจะคิดว่าการ tokenization เป็นงานที่ง่าย แต่นี่จะไม่ง่ายเลยในกรณีของการออกแบบตัวจัดทำดัชนีพหุภาษา ในรูปแบบดิจิทัล ข้อความภาษาอื่นๆ เช่นภาษาจีน ญี่ปุ่น หรืออาร์บิกนั้นสะท้อนให้เห็นความท้าทายที่เพิ่มมากขึ้นเนื่องจาก คำต่างๆ ไม่ได้ถูกค้นโดยช่องว่างสีขาว เป้าหมายในการ tokenization คือการระบุคำต่างๆ ที่ผู้ใช้จะหา ระบบเหตุผลเฉพาะภาษานั้นจะถูกใช้เพื่อระบุขอบเขตของคำอย่างเหมาะสม ซึ่งมักเป็นเหตุผลในการออกแบบตัววิเคราะห์สำหรับทุกภาษาที่มีการสนับสนุน (หรือสำหรับกลุ่มของภาษาที่มีตัวบอขอบเขตและวากยสัมพันธ์ คล้ายๆ กัน) (Search engine indexing, Retrieved 2012, January23)

เพื่อให้ความช่วยเหลือด้วยการจัดอันดับเอกสารที่ตรงกันอย่างเหมาะสมนั้น เครื่องมือค้นหาหลายตัวก็รวบรวมข้อมูลเพิ่มเติมเกี่ยวกับคำแต่ละคำเช่น ภาษา หรือ หมวดหมู่ของคำ (หน้าที่ของคำในประโยค) วิธีการเหล่านี้ขึ้นอยู่กับภาษาด้วย เนื่องจากมีวากยสัมพันธ์ที่แตกต่างกันไปในแต่ละภาษา เอกสารต่างๆ มักจะไม่ได้ระบุภาษาของเอกสารหรือบ่งบอกอย่างชัดเจนและถูกต้องนัก ในการ Tokenizing เอกสารนั้น เครื่องมือค้นหาบางเครื่องพยายามที่จะระบุภาษาของเอกสารโดยอัตโนมัติ (Search engine indexing, Retrieved 2012, January23)

##### 2) รูปแบบของไฟล์ (format) ที่หลากหลาย

เพื่อให้สามารถระบุว่าส่วนใดของเอกสารที่มีอักขระอยู่ จะต้องมีการจัดการรูปแบบของไฟล์อย่างถูกต้อง เครื่องมือค้นหาที่สนับสนุนรูปแบบของไฟล์ที่หลากหลาย จะต้องสามารถเปิดและเข้าถึงเอกสารได้อย่างถูกต้องและสามารถดำเนินการ tokenize อักขระต่างๆของเอกสารนั้น ความผิดพลาดที่เกิดขึ้นกับที่เก็บข้อมูล (Search engine indexing, Retrieved 2012, January23)

คุณภาพของข้อมูลภาษาธรรมชาตินั้นอาจจะไม่สมบูรณ์เสมอไป มีเอกสารจำนวนหนึ่งโดยเฉพาะอย่างยิ่งบนอินเทอร์เน็ต ที่ไม่เป็นไปตามระเบียบการจัดการไฟล์ที่เหมาะสม อักขระทวิภาค อาจจะถูกเข้ารหัสในหลายๆ ส่วนของเอกสาร หากไม่มีการระบุอักขระเหล่านี้ และหากไม่มีการดำเนินการอย่างเหมาะสม คุณภาพของดัชนีหรือการทำงานของตัวจัดทำดัชนี อาจจะไม่ดี (Search engine indexing, Retrieved 2012, January23)



#### 2.4.2.2 Tokenization

ไม่เหมือนกับมนุษย์ที่มีความรอบรู้ คอมพิวเตอร์ไม่เข้าใจโครงสร้างเอกสาร ภาษามนุษย์ และไม่สามารถรับรู้คำและประโยคต่างๆ ได้โดยอัตโนมัติ สำหรับคอมพิวเตอร์ เอกสารก็คือลำดับของไบนารีเท่านั้น คอมพิวเตอร์ไม่ “รู้” ว่าที่วางนั้นแยกคำต่างๆ ออกจากกัน มนุษย์ต่างหากที่จะต้องตั้งคำ ตั้งโปรแกรมให้คอมพิวเตอร์ระบุว่าอะไรเป็นคำเดี่ยว หรือคำโคด ซึ่งจะเรียกว่า token โปรแกรมดังกล่าวมักจะเรียกกันทั่วไปว่า tokenizer หรือ parser หรือ lexer เครื่องมือค้นหาหลายตัว เช่นเดียวกับซอฟต์แวร์ประมวลผลภาษาธรรมชาติอื่นๆ จะผนึกเอาโปรแกรมเฉพาะทางสำหรับการวิเคราะห์ (parsing) เช่น YACC หรือ Lex เข้าไปด้วย (Search engine indexing, Retrieved 2012, January23)

ในระหว่างการ tokenization นั้น parser จะระบุลำดับอักขระที่บ่งบอกถึงคำและองค์ประกอบอื่นๆ เช่น เครื่องหมายวรรคตอน ที่จะแสดงด้วยรหัสตัวเลข บางรหัสก็เป็นอักขระที่ไม่มีบนเครื่องพิมพ์ Parser ยังสามารถระบุเอกลักษณ์เช่น ที่อยู่อีเมล (e-mail address) หมายเลขโทรศัพท์ และ URLs ได้ เมื่อระบุ token แต่ละตัวแล้ว คุณสมบัติหลายข้อก็จะถูกเก็บไว้ เช่น ตำแหน่งของ token (อักขระบน อักขระล่าง ผสม เหมาะสม) ภาษา หรือการเข้ารหัส หมวดหมู่ของคำ (หน้าที่ของคำในประโยคเช่น เป็นคำนาม หรือคำกริยา) ตำแหน่ง หมายเลขประโยค ตำแหน่งของประโยค ความยาว และหมายเลขบรรทัด (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2.3 การระบุภาษา

หากเครื่องมือค้นหาสนับสนุนภาษาหลายภาษา ขั้นตอนเริ่มต้นร่วมกันระหว่างการ tokenization คือการระบุภาษาของแต่ละเอกสาร ขั้นตอนต่อมาหลายขั้นตอนนั้นจะขึ้นอยู่กับแต่ละภาษา (เช่นการตัดปลายที่มากเกินไป (stemming) และหน้าที่ของคำในประโยค (part of speech)) การระบุภาษานั้นเป็นกระบวนการที่โปรแกรมคอมพิวเตอร์พยายามจะระบุหรือจัดหมวดหมู่ภาษาของเอกสาร โดยอัตโนมัติ ชื่ออื่นๆ สำหรับการระบุภาษานั้นได้แก่ การจัดหมวดหมู่ภาษา การวิเคราะห์ภาษา การระบุภาษา และการตัดปลายภาษา การระบุภาษาโดยอัตโนมัติเป็นหัวข้อที่กำลังวิจัยกันอยู่ในการประมวลผลภาษาธรรมชาติ การค้นหาว่าคำต่างๆ นั้นอยู่ในภาษาใด อาจจะต้องใช้ตารางการระบุภาษา (language recognition chart) (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2.4 การวิเคราะห์รูปแบบ (format)

หากเครื่องมือค้นหาสนับสนุนรูปแบบเอกสารหลายรูปแบบ เอกสารก็จะต้องมีการเตรียมพร้อมสำหรับการ tokenization ความท้าทายก็คือ รูปแบบเอกสารต่างๆ มีข้อมูลรูปแบบที่เพิ่มขึ้นมาจากเนื้อหาข้อความ ตัวอย่างเช่น เอกสาร HTML ก็จะมีป้าย HTML ดิด ซึ่งจะบ่งบอกชี้เฉพาะถึงข้อมูลรูปแบบ เช่น การเริ่มต้นบรรทัดใหม่ การเน้นคำ และขนาดอักษรหรือรูปแบบของอักษร

หากเครื่องมือค้นหาไม่สนใจข้อแตกต่างระหว่างเนื้อหาและข้อมูล “ส่วนที่เพิ่มเข้ามา” ข้อมูลที่เพิ่มเติมเข้ามาจะถูกรวมเข้าไปอยู่ในครรชนนี้ด้วย ทำให้ผลการค้นหาออกมาไม่ดี การวิเคราะห์รูปแบบคือการระบุและจัดการกับเนื้อหาเชิงรูปแบบที่ฝังอยู่ในเอกสารซึ่งควบคุมแนวทางที่เอกสารนั้นแสดงออกมาบนจอคอมพิวเตอร์หรือแปลออกมาโดยโปรแกรมซอฟต์แวร์ การวิเคราะห์รูปแบบนี้ยังถูกเรียกว่า การวิเคราะห์โครงสร้าง การ parsing รูปแบบ การดึงป้ายที่ติดออก การดึงรูปแบบออก การจัดข้อความให้เป็นมาตรฐาน การทำความสะอาดข้อความและการเตรียมข้อความ ความท้าทายของการวิเคราะห์รูปแบบนั้นถูกทำให้ยุ่งยากมากขึ้น โดยความซับซ้อนของรูปแบบไฟล์ที่หลากหลาย รูปแบบไฟล์บางประเภทนั้นมีข้อมูลที่เปิดเผยออกมาไม่มาก ในขณะที่ไฟล์ประเภทอื่นๆ นั้นมีข้อมูลให้ค้นหาได้ทั่วไป (Search engine indexing, Retrieved 2012, January23)

1) รูปแบบไฟล์ที่พบกันทั่วไป มีข้อมูลเยอะ และได้รับการสนับสนุนจากเครื่องมือค้นหาหลายตัวนั้น (Search engine indexing, Retrieved 2012, January23) มีดังนี้

- HTML
- ไฟล์ข้อความ ASCII (เอกสารข้อความที่ไม่มีรูปแบบเฉพาะเพื่อให้คอมพิวเตอร์สามารถอ่านได้)

- Adobe's Portable Document Format (PDF)
- PostScript (PS)
- LaTeX
- UseNetnetnews server formats
- XMLและสิ่งที่ต่อยอดมาเช่น RSS
- SGML
- Multimediameta data เช่นรูปแบบ ID3
- Microsoft Word
- Microsoft Excel
- Microsoft Powerpoint
- IBM Lotus Notes

2) รูปแบบไฟล์ที่บีบอัด

ตัวเลือกในการจัดการกับรูปแบบต่างๆ นั้นได้แก่การใช้เครื่องมือวิเคราะห์ที่มีให้หาซื้อได้ทั่วไป ที่ออกโดยองค์กรที่ได้พัฒนา ดูแล หรือเป็นเจ้าของรูปแบบนั้น และการเขียน parser (ตัววิเคราะห์) ขึ้นมาเอง (Search engine indexing, Retrieved 2012, January23)

เครื่องมือค้นหาบางตัวนั้นสนับสนุนการตรวจสอบไฟล์ที่ถูกเก็บอยู่ในรูปแบบที่บีบอัดหรือเข้ารหัส เมื่อต้องทำงานกับรูปแบบที่บีบอัดนั้น ตัวจัดทำดัชนีจะต้องขยายเอกสารออกมา ก่อน ขั้นตอนนี้อาจจะทำให้เกิดไฟล์หนึ่งไฟล์หรือมากกว่า ซึ่งแต่ละไฟล์ที่ออกมาจะต้องได้รับการจัดทำดัชนีแยกต่างหาก (Search engine indexing, Retrieved 2012, January23) รูปแบบไฟล์ที่บีบอัด ที่มีการสนับสนุนอยู่ทั่วไปได้แก่:

- ZIP - Zip archive file
- RAR - RoshalARchive File
- CAB - Microsoft Windows Cabinet File
- Gzip – ไฟล์ที่ถูกบีบอัดด้วย gzip
- BZIP – ไฟล์ที่ถูกบีบอัดด้วย bzip2
- Tape ARchive (TAR), Unix archive file, ไม่มีการบีบอัด (ด้วยตัวมันเอง)
- TAR.Z, TAR.GZ or TAR.BZ2 - Unix archive ถูกบีบอัดด้วย GZIP หรือ

BZIP2

### 3) ตัวอย่างของการใช้รูปแบบเอกสารไปในทางที่ไม่ดี

การวิเคราะห์รูปแบบอาจมีวิธีการพัฒนาคุณภาพ เพื่อหลีกเลี่ยงการรวม “ข้อมูลขยะ” เข้าไปในดัชนีด้วย เนื้อหาที่สามารถจัดการข้อมูลรูปแบบเพื่อให้รวมเอาเนื้อหาเพิ่มเติมเข้าไปด้วย (Search engine indexing, Retrieved 2012, January23) ตัวอย่างของการใช้รูปแบบเอกสารไปในทางที่ไม่ดี ในการ spamdexing นั้นได้แก่

- การรวมเอาคำเป็นร้อยๆ เป็นพันๆ คำ เข้าไปในส่วนที่ซ่อนจากจอกคอมพิวเตอร์ แต่มองเห็นได้ด้วยตัวจัดทำดัชนี โดยการใช้การจัดรูปแบบ (เช่น ป้าย “div” ซ่อนไว้ใน HTML ซึ่งอาจจะรวมเอาการใช้ CSS หรือ Javascript ในการทำการนี้ด้วย) (Search engine indexing, Retrieved 2012, January23)

- การตั้งคำศัพท์ของคำในตำแหน่งด้านหน้า ให้เหมือนกับสีพื้นหลัง ทำให้คำต่างๆ ถูกซ่อน ผู้ที่อ่านจากหน้าจอกคอมพิวเตอร์มองไม่เห็น แต่ตัวจัดทำดัชนีสามารถมองเห็นได้ (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2.5 การระบุหมวด/ส่วน/ภาค (section)

หากดัชนีเครื่องมือค้นหาเนื้อหาเป็นถ้าเป็นเนื้อหาปกติที่มีคุณภาพของดัชนีและคุณภาพการค้นหาอาจจะลดลงอันเนื่องมาจากเนื้อหาแบบผสมและความใกล้ชิดคำที่ไม่เหมาะสม สองปัญหาหลักที่จะตั้งข้อสังเกต (Web Search Engine, Retrieved 2012, January23)

เครื่องมือค้นหาบางตัวรวมเอาการระบุภาคส่วนเข้าไปด้วยซึ่งก็คือ การระบุภาคส่วนที่สำคัญของเอกสารก่อนที่จะดำเนินการ tokenization เอกสารทุกชิ้นในคลังนั้น ไม่ใช่ทุกชิ้นที่สามารถอ่านได้เหมือนกับหนังสือที่มีการเขียนมาอย่างดี มีการแบ่งเป็นบทและหน้าต่างๆ เอกสารหลายๆ ชิ้นบนเว็บเช่น จดหมายข่าว และรายงานองค์กรนั้น มีเนื้อหาที่จัดวางไม่ถูกต้องอยู่ และมีส่วนเพิ่มเติมที่ไม่มีเนื้อหาหลักๆ (ที่เกี่ยวข้องโดยตรงกับเนื้อหาของเอกสาร) เช่น บทความนี้มีเมนูด้านข้างที่เชื่อมต่อไปยังหน้าเว็บอื่นๆ รูปแบบไฟล์บางประเภทเช่น HTML หรือ PDF นั้นอนุญาตให้เนื้อหาถูกแสดงออกมาในรูปของคอลัมน์ได้ แม้ว่าเนื้อหาจะถูกแสดงออกมา หรือตั้งใจให้ออกมาในพื้นที่แบบต่างๆ แต่เนื้อหาเพิ่มเติมก็อาจเก็บข้อมูลนี้ ตามลำดับ คำต่างๆ ที่แสดงออกมาตามลำดับในเนื้อหาในเนื้อหาจากแหล่งที่มาจะถูกจัดสรรนี้ตามลำดับ แม้ว่าประโยคและย่อหน้าเหล่านี้จะถูกจัดให้อยู่ในส่วนที่แตกต่างกันบนจอคอมพิวเตอร์ หากเครื่องมือค้นหาจัดทำดัชนีของเนื้อหาที่เสมือนว่ามันเป็นเนื้อหาปกติ คุณภาพของดัชนีและคุณภาพของการค้นหาอาจจะแย่ลง (Search engine indexing, Retrieved 2012, January23)

การวิเคราะห์ส่วนอาจจะต้องทำให้เครื่องมือค้นหา ใช้งานการระบบตรรกะการ render ของเอกสารแต่ละชิ้น โดยเฉพาะอย่างยิ่ง เนื้อหาคร่าวๆ ของเอกสารจริง และจากนั้นจึงค่อยจัดทำดัชนีจากเนื้อหาคร่าวๆ นั้นแทน ตัวอย่างเช่น เนื้อหาบางเนื้อหาบนอินเทอร์เน็ตนั้นได้รับการ render โดยใช้ Javascript หากเครื่องมือค้นหาไม่ render หน้าและประเมิน Javascript ภายในหนึ่งหน้า เครื่องมือค้นหาจะไม่สามารถ “เห็น” เนื้อหาในทำนองเดียวกันและจะจัดทำดัชนีเอกสารอย่างผิดวิธี เนื่องจากเครื่องมือค้นหาบางตัวไม่สนใจในเรื่องการ render ผู้ออกแบบหน้าเว็บหลายๆ คนจึงหลีกเลี่ยงที่จะแสดงเนื้อหาผ่าน Javascript หรือใช้ป้าย Noscript เพื่อให้มั่นใจว่าหน้าเว็บได้รับการจัดดัชนี อย่างเหมาะสม ในขณะเดียวกัน ความจริงอันนี้สามารถนำไปใช้เพื่อให้ตัวจัดทำดัชนี “มองเห็น” เนื้อหาที่แตกต่างจากสิ่งที่ผู้อ่านมองเห็น (Search engine indexing, Retrieved 2012, January23)

เนื่องจากเนื้อหาที่ผสมปนเปและความใกล้เคียงของคำที่ไม่เหมาะสม มีปัญหาหลักๆ สองปัญหาที่พบ นั่นก็คือ:

- 1) เนื้อหาที่อยู่คนละส่วนนั้นถูกจัดเสมือนว่าเกี่ยวข้องกัน ในดัชนี แม้ว่าในความเป็นจริงแล้ว จะไม่เกี่ยวข้องกันเลย (Search engine indexing, Retrieved 2012, January23)
- 2) เนื้อหาจัดระบบ “ด้านข้าง” นั้นถูกรวมให้อยู่ในดัชนีด้วย แต่เนื้อหาด้านข้างนั้นไม่มีส่วนที่เกี่ยวข้องกับเนื้อหาของเอกสาร และทำให้ดัชนีนั้นเต็มไปด้วยตัวแทนเอกสารที่แยก (Search engine indexing, Retrieved 2012, January23)

#### 2.4.2.6 การจัดทำดัชนี Meta Tag

เอกสารเฉพาะมักจะมีข้อมูล Meta ฝังอยู่ เช่น ผู้เขียน คำสำคัญ คำอธิบาย และภาษา สำหรับหน้า HTML ตัว meta tag จะมีคำสำคัญที่สามารถนำไปใส่ไว้ในครรชนีได้ เทคโนโลยี เครื่องมือค้นหาอินเทอร์เน็ตในสมัยก่อนจะจัดทำครรชนีเฉพาะคำสำคัญที่อยู่ใน meta tags สำหรับ ครรชนีก้าวหน้า เอกสารตัวเต็มจะไม่ถูกวิเคราะห์ ในสมัยนั้น การจัดทำครรชนีข้อความเต็มยังไม่ แพร่หลาย ยังไม่มีแม้แต่ฮาร์ดแวร์ที่จะสามารถสนับสนุนเทคโนโลยีเช่นนั้นได้ การออกแบบภาษา เพิ่มเติม HTML นั้นเริ่มแรกเลยก็รวมการสนับสนุน meta tags เพื่อให้ง่ายและเหมาะสมต่อการจัดทำ ครรชนี โดยไม่ต้องใช้การ tokenization (Search engine indexing, Retrieved 2012, January23)

เมื่ออินเทอร์เน็ตเติบโตขึ้นในช่วงทศวรรษที่ 1990s องค์กรแบบเดิม (brick-and-mortar corporations) ต่างก็หันตัวเอง “ออนไลน์” กันหลายองค์กร และสร้างเว็บไซต์องค์กรขึ้นมา คำสำคัญ ที่ใช้ในการอธิบายหน้าเว็บ (ซึ่งหน้าเว็บหลายหน้าในสมัยนั้นก็คือนำหน้าเว็บขององค์กรที่คล้ายกับใบ โฆษณาสินค้า) เปลี่ยนแปลงจากคำอธิบายธรรมดา เป็น คำสำคัญที่มุ่งเน้นไปที่การตลาด ออกแบบ มาเพื่อส่งเสริมการขาย โดยการนำหน้าเว็บขึ้นไปอยู่ในอันดับต้นๆ ของการค้นหาสำหรับคำค้นหา บางคำ การที่คำสำคัญเหล่านี้ถูกกำหนดมาเป็นการเฉพาะเพื่อประโยชน์ส่วนบุคคลนั้นทำให้เกิด spamdexing ซึ่งกลายมาเป็นตัวกระตุ้นให้เครื่องมือค้นหาหลายตัวหันมาใช้เทคโนโลยีการจัดทำ ครรชนีแบบข้อความเต็ม ในช่วงปี 1990s ผู้ออกแบบเครื่องมือค้นหาและบริษัทต่างๆ ทำได้เพียงใส่ “คำสำคัญด้านการตลาด” หลายๆ คำลงในเนื้อหาของหน้าเว็บก่อนที่จะค่อยๆ ปลอมข้อมูลที่น่าสนใจ และมีประโยชน์ออกมา เนื่องจากการขัดผลประโยชน์กับเป้าหมายทางธุรกิจของการออกแบบ เว็บไซต์ที่เน้นผู้ใช้ ซึ่ง “ไม่ค่อยมีการเปลี่ยนแปลง” สมการความคุ้มค่าชั่วอายุลูกค้าได้ถูกเปลี่ยนไป เพื่อการรวมเอาเนื้อหาที่มีประโยชน์มากขึ้นเข้าไปในเว็บไซต์ด้วยความหวังที่จะรักษาผู้เข้ามาเยี่ยมชม เอาไว้ ในทำนองเดียวกัน การจัดทำครรชนีข้อความเต็ม เป็นสิ่งที่เน้นประโยชน์ และเพิ่มคุณภาพ ของผลการค้นหาของเครื่องมือค้นหา เนื่องจากมันเป็นการก้าวออกมาจาก การควบคุมการจัดวางผล การค้นหาของเครื่องมือค้นหาเพื่อประโยชน์ส่วนบุคคล ซึ่งก็ทำให้เกิดการพัฒนาการวิจัยเทคโนโลยี การจัดทำครรชนีข้อความเต็ม (Search engine indexing, Retrieved 2012, January23)

ในการค้นหาบนเครื่องคอมพิวเตอร์นั้น ผลการค้นหาหากมี meta tags เพื่อเปิดทาง ให้ผู้เขียนตั้งคำถามได้ว่า จะให้เครื่องมือค้นหาจัดทำครรชนีเนื้อหาจากไฟล์หลายๆ ไฟล์ได้ โดย เนื้อหาเหล่านั้นไม่จำเป็นต้องมาจากเนื้อหาของไฟล์ การค้นหาบนเครื่องคอมพิวเตอร์นั้นอยู่ ภายใต้การควบคุมของผู้ใช้มากกว่า ในขณะที่เครื่องมือค้นหาในอินเทอร์เน็ตนั้นจะต้องเน้นไปที่การ จัดทำครรชนีข้อความเต็มมากกว่า (Search engine indexing, Retrieved 2012, January23)

## 2.5 คำค้นหาเว็บ

คำค้นหาเว็บ เป็นคำค้นหาที่ผู้ใช้กรอกเข้าไปในเครื่องมือค้นหาเว็บ เพื่อให้บรรลุความต้องการข้อมูลของเขา/เธอ คำค้นหาเว็บนั้นมีเอกลักษณ์ตรงที่คำเหล่านี้ไม่มีโครงสร้างและมักจะไม่ใช่คำเต็ม คำเหล่านี้แตกต่างกันไปตามมาตรฐานภาษาของคำค้นหาซึ่งควบคุมโดยกฎวากยสัมพันธ์เครื่องคิด (Web search query, Retrieved 2012, January23)

### 2.5.1 ประเภท

มีหมวดหมู่กว้างๆ 4 หมวด ที่ครอบคลุมคำค้นหาเว็บเกือบทั้งหมด เครื่องมือค้นหา มักจะสนับสนุนคำค้นหาแบบที่สี่ ซึ่งมีการใช้ไม่ค่อยบ่อยนัก (Web search query, Retrieved 2012, January23)

2.5.1.1 คำค้นหาข้อมูล เป็นคำค้นหาที่ครอบคลุมหัวข้อกว้างๆ (เช่น *Colorado* หรือ *trucks*) อาจะปรากฏผลการค้นหาหลายพันผลที่เกี่ยวข้อง (Web search query, Retrieved 2012, January23)

2.5.1.2 คำค้นหาทางไปสู่วีป เป็นคำค้นหาที่หาเว็บไซต์ หรือเว็บเพจหนึ่งๆ (เช่น *youtube* หรือ *delta air lines*) (Web search query, Retrieved 2012, January23)

2.5.1.3 คำค้นหาทางธุรกรรม เป็นคำค้นหาที่สะท้อนความต้องการของผู้ใช้ในการดำเนินการอย่างใดอย่างหนึ่ง เช่น ซื้อรถหรือดาวน์โหลด screen saver (Web search query, Retrieved 2012, January23)

2.5.1.4 คำค้นหาเพื่อการเชื่อมโยง เป็นคำค้นหาที่รายงานการเชื่อมโยงของ web graph ที่ได้รับการจัดทำดัชนีแล้ว(เช่นการเชื่อมโยง (ลิงค์) ใดที่เชื่อมไปที่ URL นี้ และมีหน้าก่หน้าที่ได้รับการจัดทำดัชนี จากชื่อโดเมนอันนี้) (Web search query, Retrieved 2012, January23)

### 2.5.2 ลักษณะต่างๆ

เครื่องมือค้นหาเว็บที่เป็นธุรกิจที่สุด จะไม่เปิดเผยบันทึกการค้นหา ดังนั้นจึงเป็นการยากที่จะได้มาซึ่งข้อมูลเกี่ยวกับว่าผู้ใช้ได้ค้นหาอะไรบ้างบนเว็บ อย่างไรก็ตาม งานศึกษาในปี ค.ศ.2001 นั้นได้วิเคราะห์คำค้นหาจากเครื่องมือค้นหา Excite และแสดงให้เห็นลักษณะที่น่าสนใจบางประการของการค้นหาบนเว็บ (Web search query, Retrieved 2012, January23)

2.5.2.1 ความยาวของคำค้นหาโดยเฉลี่ยอยู่ที่ 2.4 คำ (Web search query, Retrieved 2012, January23)

2.5.2.2 ประมาณครึ่งหนึ่งของผู้ใช้ ใส่คำค้นหาเดี่ยวๆ ในขณะที่เกือบหนึ่งในสามของผู้ใช้ใส่คำค้นหาที่เป็นเอกลักษณ์ สามคำค้น หรือมากกว่า (Web search query, Retrieved 2012, January23)

2.5.2.3 เกือบครึ่งหนึ่งของผู้ใช้ดูผลการค้นหาเพียงแค่หนึ่งหรือสองหน้าแรกเท่านั้น (10 ผลการค้นหาต่อหน้า) (Web search query, Retrieved 2012, January23)

2.5.2.4 น้อยกว่า 5% ของผู้ใช้ใช้งานการค้นหาขั้นสูง (สำหรับผู้ประกอบการ Boolean เช่น และ, หรือ, ไม่ (AND, OR, NOT)) (Web search query, Retrieved 2012, January23)

2.5.2.5 คำสั่งอันดับแรกที่ใช้มากที่สุดคือ (ว่าง), *and*, *of*, และ *sex*. (Web search query, Retrieved 2012, January23)

การศึกษานี้ที่ศึกษาคำค้นหาของ Excite อันเดียนั้นพบว่า 19% ของคำค้นหานั้น มีคำที่บ่งบอกถึงภูมิศาสตร์ (เช่นชื่อสถานที่ รหัสไปรษณีย์ ลักษณะทางภูมิศาสตร์ ฯลฯ) (Web search query, Retrieved 2012, January23)

ในงานศึกษานี้ที่ศึกษาคำค้นหาของ Yahoo ในปี ค.ศ.2005 พบว่า 33% ของคำค้นหาต่างๆ ที่มาจากผู้ใช้คนเดียวกันนั้นเป็นคำค้นหาซ้ำ และ 87% ผู้ใช้จะคลิกที่ผลการค้นหาเดิม นี่แสดงให้เห็นว่าผู้ใช้หลายคนใช้คำค้นหาซ้ำเดิมในการเข้าดูข้อมูลเดิม หรือค้นหาข้อมูลเดิมอีกครั้ง การวิเคราะห์นี้ได้รับการยืนยันโดยเรื่องราวในบล็อกของเครื่องมือค้นหา Bing ที่ระบุว่าราว 30% ของคำค้นหาเป็นคำค้นหาทางไปสู่วีป (Web search query, Retrieved 2012, January23)

นอกจากนี้ การวิจัยเพิ่มเติมได้แสดงให้เห็นว่าตารางแจกแจงความถี่คำค้นหานั้นเป็นไปตามกฎเลขยกกำลัง (power law) หรือ กราฟเส้นโค้งแจกแจงทางยาว นั่นก็คือ มีค่าเป็นสัดส่วนไม่มากที่อยู่ภายในที่คำค้นหาขนาดใหญ่ (เช่น มากกว่า 100 ล้านคำค้นหา) ถูกใช้บ่อยที่สุดในขณะที่คำที่เหลือ ถูกใช้งานไม่บ่อยนัก ตัวอย่างนี้ ของหลักการ Pareto (หรือกฎ 80-20) ทำให้เครื่องมือค้นหา นำวิธีการเพื่อความเหมาะสมที่สุด เช่นการจัดทำดัชนี หรือการทำพาร์ทิชันฐานข้อมูล การ caching หรือการ ดึงล่วงหน้า (pre-fetching) มาใช้ (Web search query, Retrieved 2012, January23)

แต่ในงานศึกษาเมื่อไม่นานมานี้ ในปี ค.ศ.2011 มีการค้นพบว่าความยาวของคำค้นหาได้เพิ่มมากขึ้นอย่างสม่ำเสมอตามกาลเวลา และความยาวเฉลี่ยของคำค้นหาที่ไม่ใช่ภาษาอังกฤษก็เพิ่มขึ้นมากกว่าคำค้นหาภาษาอังกฤษ (Web search query, Retrieved 2012, January23)

### 2.5.3 คำค้นหาที่มีโครงสร้าง

ในเครื่องมือค้นหาที่สนับสนุนการประกอบการ Boolean และ การใช้วงเล็บนั้น วิธีการที่ใช้กันมาแต่เดิม โดยบรรณารักษ์นั้นอาจนำมาใช้ได้ ผู้ใช้ที่กำลังหาเอกสารที่ครอบคลุมหัวเรื่องหลายๆ หัวเรื่องหรือ มุมมองนั้นอาจจะต้องการที่จะอธิบายแต่ละหัวเรื่องหรือมุมมองด้วยการแยกคำบ่งบอกลักษณะออกจากกันเช่น ยานพาหนะ หรือ รถ หรือ รถยนต์ คำค้นหาที่บ่งบอกมุมมอง คือ การรวมเอามุมมองต่างๆ เช่น (electronic OR computerized OR DRE) AND (voting OR elections OR election OR balloting OR cllectoral) จะมีความเป็นไปได้สูงที่จะพบเอกสารเกี่ยวกับ electronic

voting แม้ว่าเอกสารเหล่านั้นจะละคำว่า “electronic” และ “voting” หรือ ละไว้ทั้งสองคำ (Web search query, Retrieved 2012, January 23)



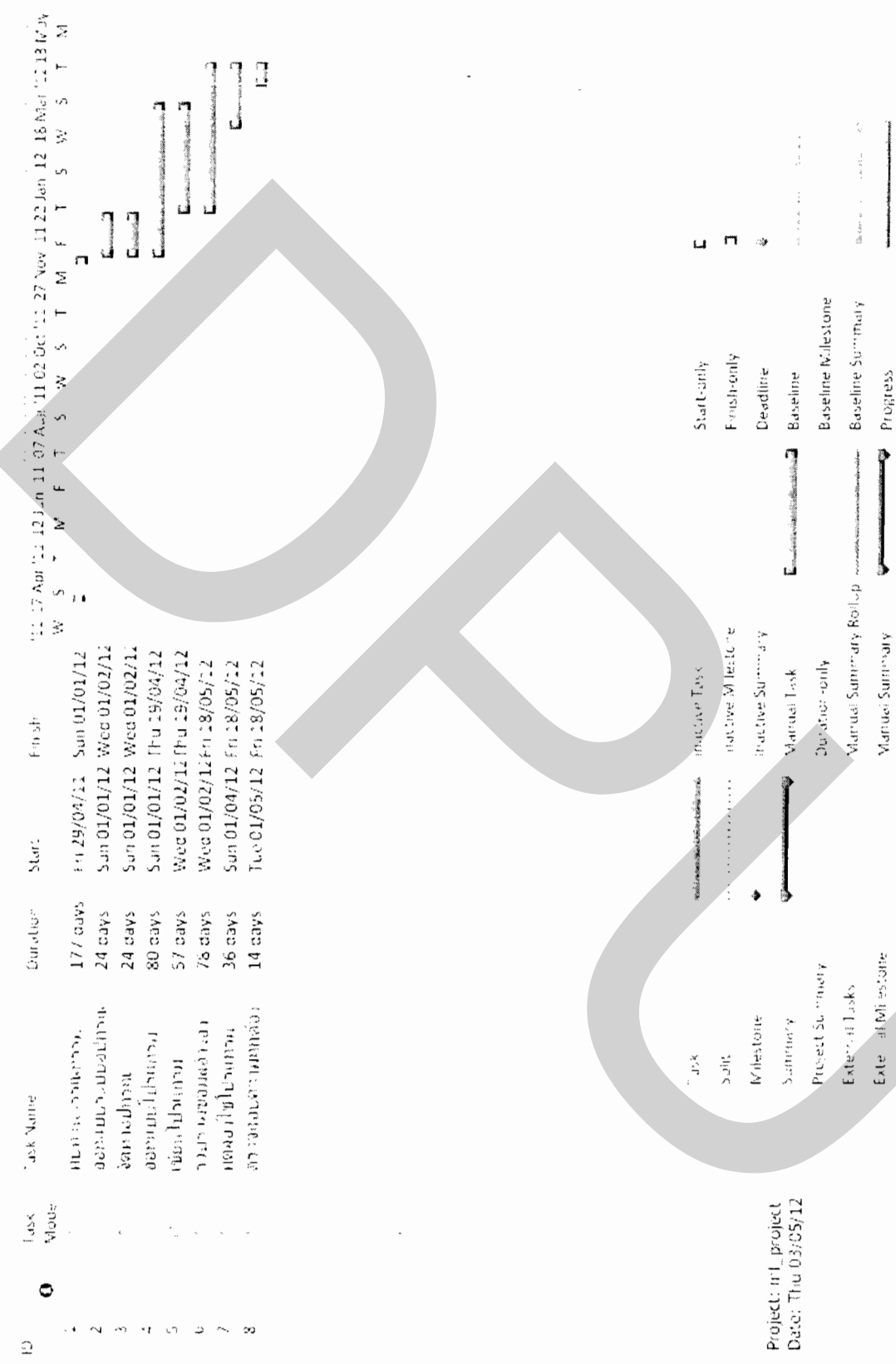


## บทที่ 3

### วิธีการดำเนินการและเครื่องมือ

#### 3.1 แผนเวลาดำเนินการ

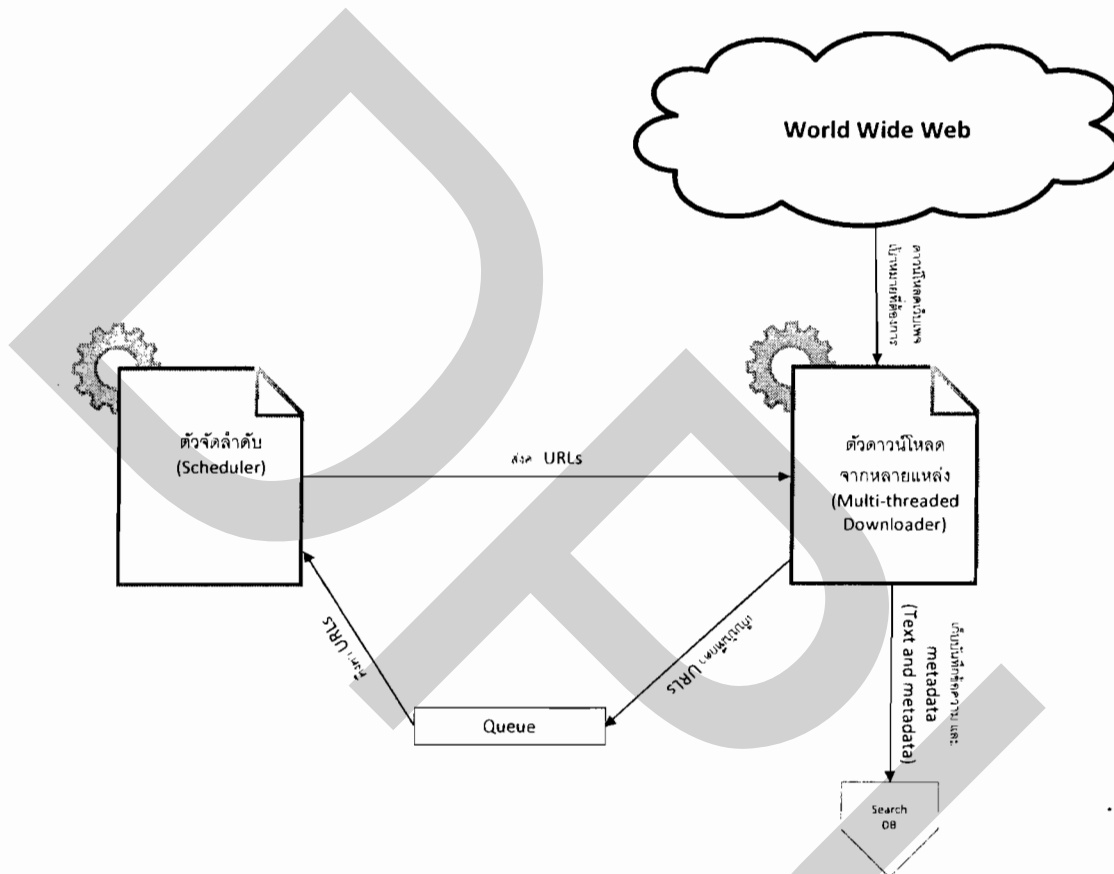
การวางแผนการดำเนินงาน เป็นขั้นตอนที่สำคัญขั้นตอนหนึ่ง เพราะนอกจากจะช่วยในเรื่องของการจัดทำตารางเวลาที่ใช้ในการวิจัยแล้ว ยังช่วยให้สามารถดำเนินการวิจัยได้ตามแผนที่วางไว้อีกด้วย ผู้วิจัยได้วางแผนการดำเนินงานโดยใช้แผนภูมิแกนต์(Gantt Chart) ซึ่งพัฒนาโดย Henry L. Gantt แผนภูมิแกนต์เป็นแผนภูมิแท่งที่ใช้ควบคุมโครงการ เป็นเครื่องมือที่จัดทำขึ้นเพื่อแสดงเวลาเริ่มต้นและสิ้นสุดของกิจกรรมต่างๆที่ได้กำหนดไว้ในโครงการ โดยได้ผลของการวางแผนการดำเนินงาน ตาม0



รูปที่ 3.1 แผนภูมิแกนต์ สำหรับวางแผนการดำเนินงาน

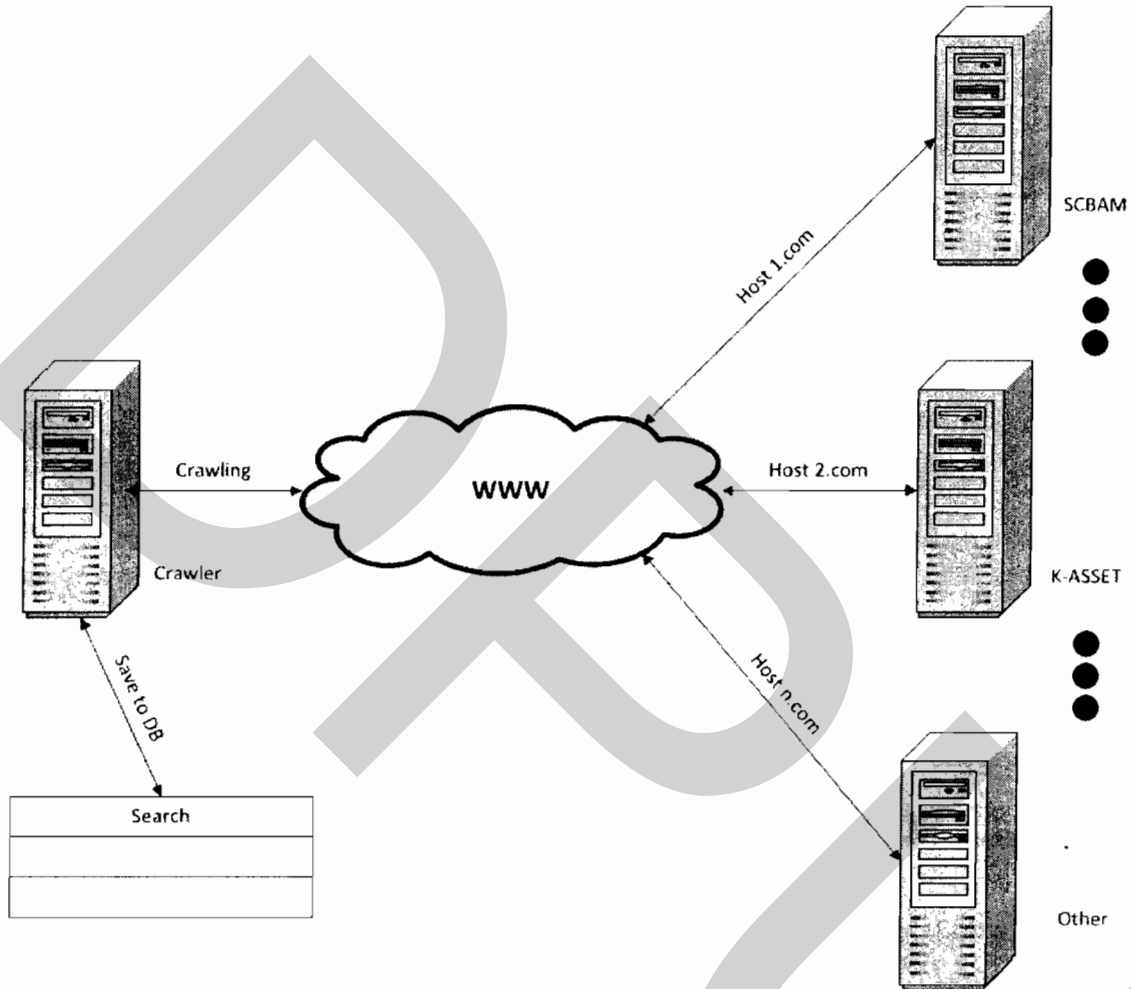
### 3.2 การออกแบบฟังก์ชัน (Functional Design)

#### 3.2.1 การออกแบบฟังก์ชันในส่วนของการทำงานของ Web Crawler (ภายใน)



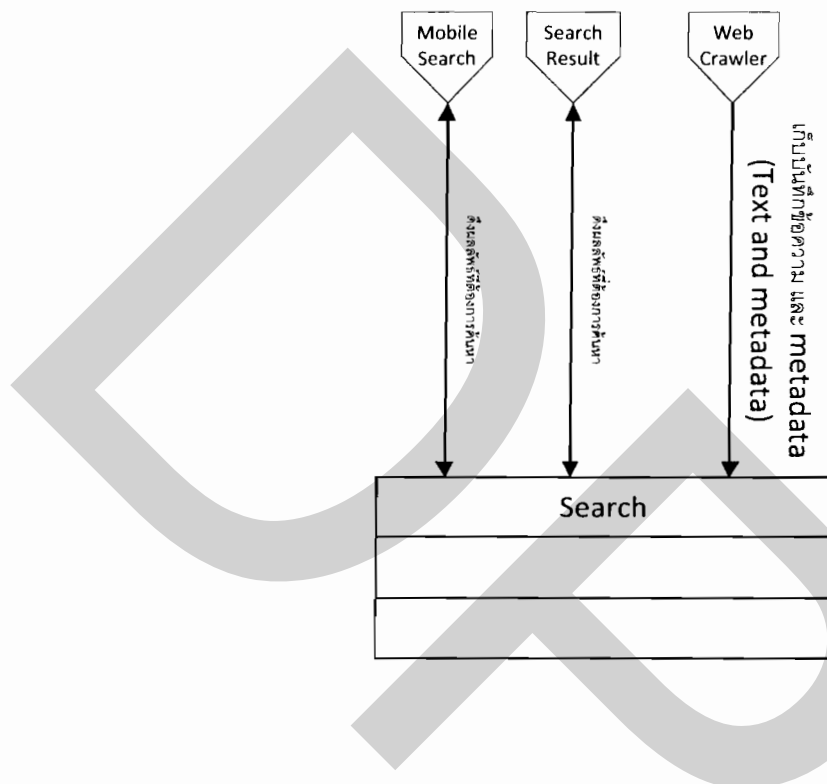
รูปที่ 3.2 แผนภาพแสดงการทำงานของ Web Crawler (ภายใน)

### 3.2.2 การออกแบบฟังก์ชันในส่วนการทำงานของ Web Crawler (ภายนอก)



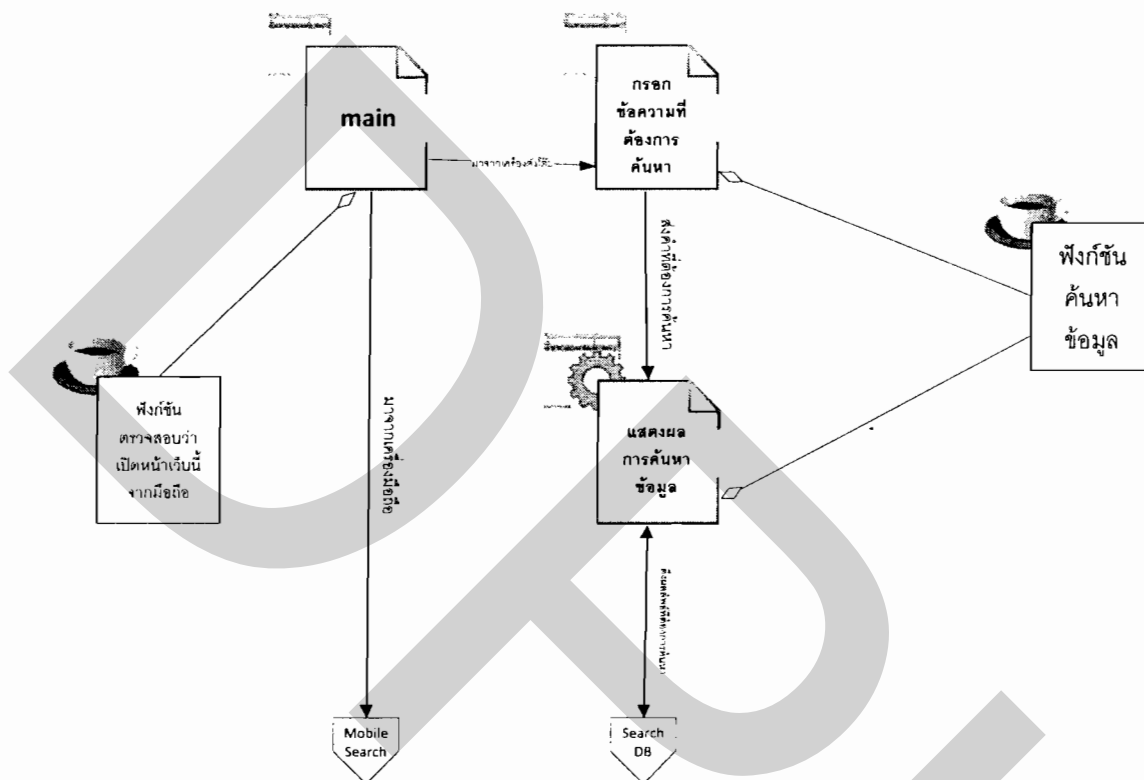
รูปที่ 3.3 แผนภาพแสดงการทำงานของ Web Crawler (ภายนอก)

### 3.2.3 การออกแบบฟังก์ชันในส่วนของการทำงานของ Search Application



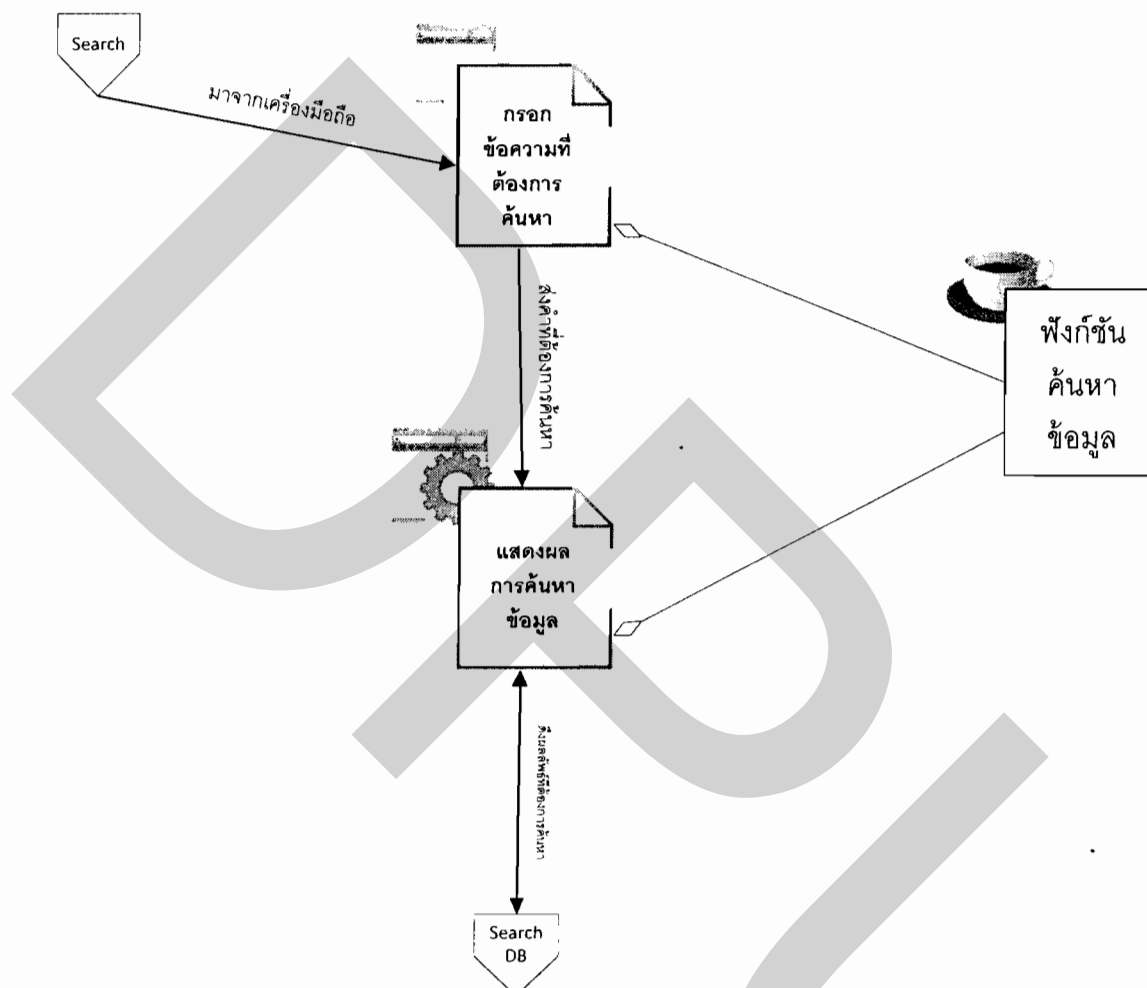
รูปที่ 3.4 แผนภาพแสดงการทำงานของ Search Application

3.2.4 การออกแบบฟังก์ชันในส่วนของการทำงานของการแสดงผลการค้นหา



รูปที่ 3.5 แผนภาพแสดงการทำงานของการทำงานของการแสดงผลการค้นหา

### 3.2.5 การออกแบบฟังก์ชันในส่วนของการทำงานของการค้นคืนข้อมูลผ่านอุปกรณ์มือถือ



รูปที่ 3.6 แผนภาพแสดงการทำงานของการทำงานของการค้นคืนข้อมูลผ่านอุปกรณ์มือถือ

### 3.3 เครื่องมือที่ใช้ในการพัฒนา

#### 3.3.1 Hardware

เครื่อง Apple (Intel - Platform) ระบบ Macintosh ความสามารถของเครื่องขั้นต่ำ

3.3.1.1 CPU : 2.66 GHz Intel Core 2 Duo

3.3.1.2 RAM : 4 GB 1067 MHz DDR3

3.3.1.3 HDD : 500 GB Serial ATA 2 Drive

#### 3.3.2 Software

โปรแกรมที่ใช้งานประกอบด้วย

3.3.2.1 Adobe Dreamweaver

3.3.2.2 Adobe Photoshop

3.3.2.3 BBEdit

3.3.2.4 Mac OS X Server Snow Leopard Ver. 10.66 หรือเวอร์ชันสูงกว่านั้นก็ได้

3.3.2.5 Terminal

3.3.2.6 MAMP Ver.2.00 หรือเวอร์ชันสูงกว่านั้นก็ได้

3.3.2.7 CronniX

3.3.2.8 DynDns Updater

3.3.2.9 Safari

### 3.4 วิธีการเก็บรวบรวมข้อมูล

3.4.1 ใช้คำสั่งของภาษา PHP ชื่อ fopen() เป็นเบื้องหลังในการช่วยเก็บรวบรวมข้อมูล

3.4.2 ใช้ MySQL Database เป็นตัวจัดเก็บและจัดการข้อมูลเป็นจำนวนมากทั้งหมด

3.4.3 ใช้ XML เป็นตัวเสริมในการช่วยเพิ่มความเร็วในการค้นหาข้อมูล และแลกเปลี่ยนข้อมูล ซึ่งมีที่มาจากกรองเฉพาะข้อมูลสำคัญจาก MySQL Database อีกทีหนึ่ง

### 3.5 การวิเคราะห์ข้อมูล

3.5.1 ใช้วิธีการวิเคราะห์ข้อมูลทั้งหมดที่มาจากกรองข้อมูลหลังจากส่งคำสั่ง fopen() ในภาษา PHP โดยการเปรียบเทียบข้อมูลทั้งหมด ใช้หลักการวิจัยเชิงปริมาณ โดยใช้เครื่องคอมพิวเตอร์ในการจับเวลา และคิดคำนวณเวลาดึงข้อมูลเป็นหลัก



3.5.2 ใช้วิธีการวิเคราะห์ข้อมูลทั้งหมดที่มาจาก MySQL Database โดยการเปรียบเทียบข้อมูลทั้งหมด ใช้หลักการวิจัยเชิงปริมาณ โดยใช้เครื่องคอมพิวเตอร์ในการจับเวลา ค้นหาข้อมูล คำนวณข้อมูล ประมวลผลข้อมูลเป็นหลัก

3.5.3 ใช้วิธีการวิเคราะห์ข้อมูลทั้งหมดที่มาจาก XML โดยการเปรียบเทียบข้อมูลทั้งหมด ใช้หลักการวิจัยเชิงปริมาณ โดยใช้เครื่องคอมพิวเตอร์ในการจับเวลา ค้นหาข้อมูล และแลกเปลี่ยนข้อมูล ที่มาจากการกรองเฉพาะข้อมูลสำคัญจาก MySQL Database อีกทีหนึ่ง

## บทที่ 4

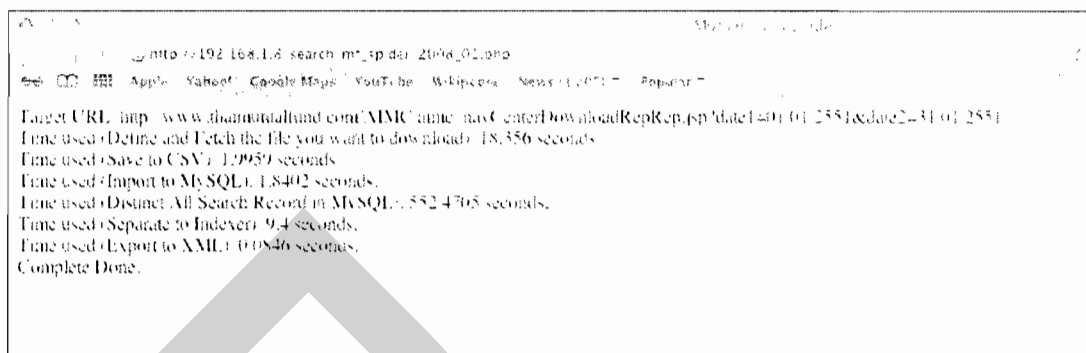
### ผลลัพธ์ของการพัฒนาระบบ

ผู้วิจัยได้ทดลองเพื่อวัดประสิทธิภาพในการทำงานของระบบที่ได้พัฒนาขึ้น และทำการเปรียบเทียบผลลัพธ์ที่ได้จากการค้นคืนกับระบบอื่น ๆ ที่มีอยู่ในปัจจุบัน เนื้อหาในบทนี้จึงถูกแบ่งออกเป็นหัวข้อย่อย ๆ ดังต่อไปนี้

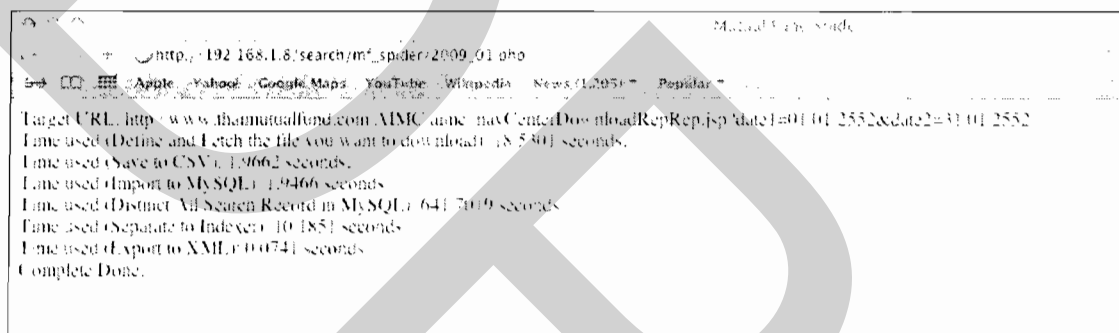
- 4.1 ความเร็วในการทำงานของในส่วนของ Crawler
- 4.2 ความเร็วในการค้นหาข้อมูล
- 4.3 ระบบ MFS สำหรับค้นหาข้อมูลกองทุนรวม
- 4.4 การเปรียบเทียบผลลัพธ์การค้นคืนข้อมูลระหว่าง MFS และ Settrade.com

#### 4.1 ความเร็วในการทำงานของโปรแกรม Crawler

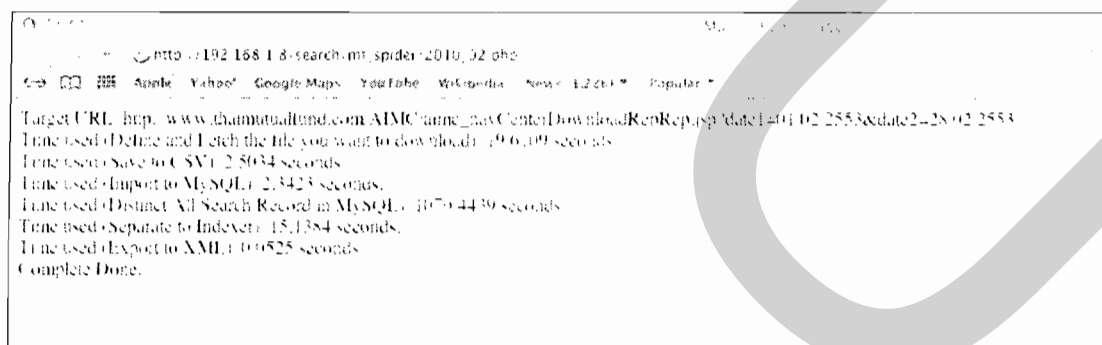
ในการวัดความเร็วในการทำงานของ Crawler ผู้วิจัยทำการกำหนดสภาพแวดล้อมในการทำงาน ให้เหมือนกับการค้นหาข้อมูลบนระบบจริง โดยผู้วิจัยทำการเชื่อมโยงเครื่องแม่ข่ายที่มีการติดตั้งระบบ MFS อยู่เข้ากับเครือข่ายอินเทอร์เน็ต ผ่านระบบ dynamic dns ของ dyndns.org ซึ่งเป็นการจำลองให้เครื่องแม่ข่ายที่ใช้ในการพัฒนาระบบ เสมือนกับเป็นเครื่องแม่ข่ายเครื่องหนึ่งบนอินเทอร์เน็ต โดยในการวัดความเร็วของการทำงานของ Crawler นั้น ผู้วิจัยใช้เครื่องคอมพิวเตอร์อีกเครื่องทำหน้าที่ เป็นเครื่องลูกข่าย เพื่อเข้าถึงระบบ MFS ผลลัพธ์ของการวัดความเร็วแสดงอยู่ในภาพที่ 4.1 – 4.5 และสรุปไว้ในตารางที่ 4.1



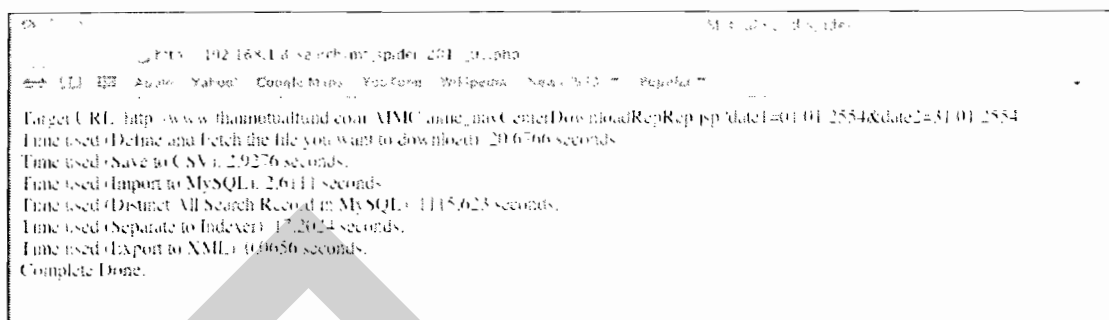
ภาพที่ 4.1 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2551



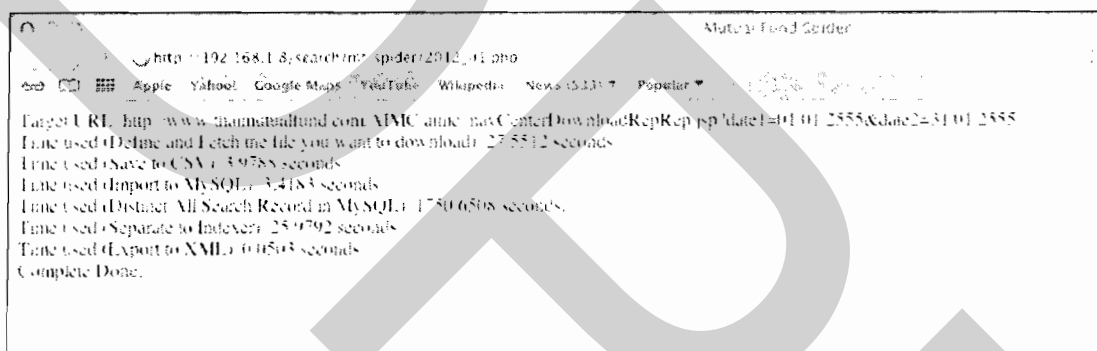
ภาพที่ 4.2 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2552



ภาพที่ 4.3 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2553



ภาพที่ 4.4 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2554



ภาพที่ 4.5 ความเร็วในการทำงานของ Web Crawler เมื่อดึงข้อมูลใน ม.ค. 2555

ตารางที่ 4.1 เปรียบเทียบความเร็วในการทำงานของโปรแกรม Crawler

Activity/Year	Jan 2551	Jan 2552	Jan 2553	Jan 2554	Jan 2555
Define and Fetch the file	18.3560	18.5301	19.6109	20.6766	27.5512
Save to CSV	1.9959	1.9620	2.5034	2.9276	3.9788
Import to MySQL	1.8402	1.9466	2.3423	2.6111	3.4183
Distinct All Search Record in MySQL	552.4705	641.7019	1070.4439	1115.6230	1750.6508
Separate to Indexer	9.4000	10.1851	15.1384	17.2024	25.9792
Export to XML	0.0846	0.0741	0.0525	0.0656	0.0503
<b>Total Second</b>	<b>584.1472</b>	<b>674.3998</b>	<b>1110.0914</b>	<b>1159.1063</b>	<b>1811.6286</b>
<b>Imported Record</b>	<b>10,524</b>	<b>10,653</b>	<b>12,851</b>	<b>15,393</b>	<b>20,143</b>
<b>Performance</b>	<b>0.0555</b>	<b>0.0633</b>	<b>0.0864</b>	<b>0.0753</b>	<b>0.0899</b>

#### 4.2 ความเร็วในการค้นหาข้อมูล

ในกรณีศึกษาความเร็วในการค้นหาข้อมูล โดยแบ่งและจำแนกรูปแบบผลการทดลอง ออกเป็น 3 รูปแบบ

##### 4.2.1 ความเร็วในการค้นหาข้อมูลแต่ละคำ

ตารางที่ 4.2 เปรียบเทียบความเร็วในการค้นหาข้อมูลแต่ละคำ

	ค้นหาที่กว้าง	ค้นหาที่เฉพาะเจาะจง
คำค้นที่ใช้ชื่อ	K-ASSET	K-MONEY
กลุ่มคำค้นเป็น	ชื่อของ บลจ. คำหนึ่ง	ชื่อย่อของกองทุนรวมคำหนึ่ง
จำนวนหน้าทั้งหมด	13,031 หน้า	109 หน้า
ตำแหน่งหน้าปัจจุบันที่	1	1
ใช้เวลาค้นหา	6.3871 วินาที	5.7739 วินาที
มีจำนวน	130,303 เรคอร์ด	1,083 เรคอร์ด
ค่าประสิทธิภาพ	$4.9017 \times 10^{-5}$	$5.3314 \times 10^{-3}$

#### 4.2.2 ความเร็วในการค้นคืนข้อมูลแต่ละหน้า

ตารางที่ 4.3 เปรียบเทียบความเร็วในการค้นคืนข้อมูลแต่ละหน้า

	ค้นหาที่กว้าง	ค้นหาที่เฉพาะเจาะจง
คำค้นที่ใช้ชื่อ	SCBAM	SCBAM
กลุ่มคำค้นเป็น	ชื่อของ บลจ. ค่ายหนึ่ง	ชื่อของ บลจ. ค่ายหนึ่ง
จำนวนหน้าทั้งหมด	13,470 หน้า	13,470 หน้า
ตำแหน่งหน้าปัจจุบันที่	1	2
ใช้เวลาค้นหา	6.3294 วินาที	6.0395 วินาที
มีจำนวน	134,693 เรคอร์ด	134,693 เรคอร์ด
ค่าประสิทธิภาพ	$4.6991 \times 10^{-5}$	$4.4839 \times 10^{-5}$

#### 4.2.3 ความเร็วในการค้นหาข้อมูลเมื่อใส่เครื่องหมาย Quote

ตารางที่ 4.4 เปรียบเทียบความเร็วในการค้นหาข้อมูลเมื่อใส่เครื่องหมาย Quote

	ค้นหาที่กว้าง	ค้นหาที่เฉพาะเจาะจง
คำค้นที่ใช้ชื่อ	EQ	'EQ'
กลุ่มคำค้นเป็น	ชื่อประเภทกองทุนรวม	ชื่อประเภทกองทุนรวม
จำนวนหน้าทั้งหมด	71,979 หน้า	52,496 หน้า
ตำแหน่งหน้าปัจจุบันที่	1	1
ใช้เวลาค้นหา	6.5723 วินาที	4.0619 วินาที
มีจำนวน	130,303 เรคอร์ด	1,083 เรคอร์ด
ค่าประสิทธิภาพ	$9.1309 \times 10^{-6}$	$7.7375 \times 10^{-6}$

### 4.3 ระบบ MFS สำหรับค้นหาข้อมูลกองทุนรวม



ภาพที่ 4.6 หน้าแรกก่อนเติมคำที่ต้องการค้นหา





MFS	
SCBSE	SCB SEC. INDEX FUND กองทรัสต์ไทยพาณิชย์ อีเอ็ม บี แคช ฟันด์
SCBSI	SCB SAVING FUND IN OMV OPEN FUND กองทรัสต์ไทยพาณิชย์ อีเอ็ม บี แคช ฟันด์
<b>SEARCH RESULT</b>	
Keyword: SCBAM	
Asset Class: SCBAM	
Total: 7545 Results Page 1 of 74 Pages Found: 3208 records	
Date:	3/04/2009
Type:	CI
Category:	EQ
Asset:	SCBAM
AssetID:	SCBAM
FundID:	ไทยพาณิชย์อำนวยการ
FundID:	THAI CAPITAL FUND
SymbolFund:	TCF
FundNetAsset:	893,291,860.94
NAV:	1.067600
DividCat:	
DivDate:	
Out:	
Fee:	
Date:	3/04/2009
Type:	CI
Category:	EQ
Asset:	SCBAM
AssetID:	SCBAM
FundID:	กองทรัสต์ไทยพาณิชย์ อีเอ็ม บี แคช ฟันด์
FundID:	SCB SEC INDEX FUND

ภาพที่ 4.9 หน้าผลลัพธ์ระหว่างกำลังเติมคำที่ต้องการค้นหาเพิ่มเติม

#### 4.4 การเปรียบเทียบผลลัพธ์การค้นคืนข้อมูลระหว่าง MFS และ Settrade.com

ในงานวิจัยนี้สามารถเปรียบเทียบระหว่าง Mutual Fund Searching กับ Settrade.com สามารถสรุปได้ดังตารางต่อไปนี้

ตารางที่ 4.5 เปรียบเทียบระหว่าง Mutual Fund Searching กับ Settrade.com

Mutual Fund Searching	Settrade.com
<ol style="list-style-type: none"> <li>1. สามารถค้นหาชื่อย่อของกองทุนรวมได้</li> <li>2. สามารถค้นหาชื่อเต็มภาษาไทยของกองทุนรวมได้</li> <li>3. สามารถค้นหาชื่อเต็มภาษาอังกฤษของกองทุนรวมได้</li> <li>4. สามารถบอกระยะเวลาการค้นหาข้อมูล</li> <li>5. สามารถบอกข้อมูลต่างๆที่เกี่ยวข้องกับกองทุนรวม เช่น ประเภทของกองทุนรวม, ค่าหลักทรัพย์จัดการกองทุนรวมของกองทุนรวมได้โดยตรง</li> <li>6. สามารถบอกเวลาล่าสุดของข้อมูลซื้อขายกองทุนได้</li> <li>7. สามารถบอกราคาซื้อขายในส่วนของกองทุนรวมได้</li> <li>8. กรณีค้นหากองทุนรวมสามารถค้นหาได้โดยตรงได้เลย</li> <li>9. สามารถใช้เครื่องหมาย Quote ร่วมในการค้นหาข้อมูล เพื่อค้นหาอย่างละเอียดได้</li> <li>10. ไม่สามารถใช้เครื่องหมาย AND, OR เพื่อเป็น Boolean Mode ในการค้นหาได้</li> </ol>	<ol style="list-style-type: none"> <li>1. สามารถค้นหาชื่อย่อของกองทุนรวมได้</li> <li>2. ไม่สามารถค้นหาชื่อเต็มภาษาไทยของกองทุนรวมได้</li> <li>3. ไม่สามารถค้นหาชื่อเต็มภาษาอังกฤษของกองทุนรวมได้</li> <li>4. ไม่สามารถบอกระยะเวลาการค้นหาข้อมูลได้</li> <li>5. สามารถบอกข้อมูลต่างๆที่เกี่ยวข้องกับกองทุนรวม เช่น ประเภทของกองทุนรวม, ค่าหลักทรัพย์จัดการกองทุนรวมของกองทุนรวมโดยผ่านเว็บของ Thaimutualfund.com อีกทีหนึ่ง</li> <li>6. สามารถบอกเวลาล่าสุดของข้อมูลซื้อขายได้</li> <li>7. ไม่สามารถบอกราคาซื้อขายในส่วนของกองทุนรวมได้</li> <li>8. กรณีค้นหากองทุนรวมด้วระบบจะเปิดหน้าเพจของ Thaimutualfund.com อีกทีหนึ่ง</li> </ol>

## ตารางที่ 4.5 (ต่อ)

Mutual Fund Searching	Settrade.com
<p>11. ไม่สามารถค้นหาชื่อย่อหลักทรัพย์อื่นๆ ที่ไม่เกี่ยวข้องกับกองทุนรวมได้เลย เนื่องจากระบบนี้ถูกออกแบบให้ค้นหาแต่เฉพาะกองทุนรวมโดยเฉพาะ ดังนั้นจึงไม่สามารถค้นหาหลักทรัพย์อื่นๆ ที่ไม่เกี่ยวข้องกับกองทุนรวมได้เลย</p>	<p>9. ไม่สามารถใช้เครื่องหมาย Quote ร่วมในการค้นหาข้อมูลเพื่อค้นหาอย่างละเอียดได้</p> <p>10. ไม่สามารถใช้เครื่องหมาย AND, OR เพื่อเป็น Boolean Mode ในการค้นหาได้</p> <p>11. สามารถค้นหาชื่อย่อหลักทรัพย์อื่นๆ ที่ไม่เกี่ยวข้องกับกองทุนรวมได้</p>

## บทที่ 5

### สรุปผลการทดลอง

#### 5.1 ความเร็วในการทำงานของโปรแกรม Crawler

จากผลการทดลอง ผู้วิจัยดึงข้อมูลตั้งแต่ปี 2550 จนถึงปี 2555 พบว่าในปี 2550 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 23.9417 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลงฐานข้อมูลใช้เวลาประมาณ 1.6008 วินาที นำเข้าฐานข้อมูลใช้เวลาประมาณ 1.875 วินาที ตัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 430.1954 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 7.84 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0449 วินาที ในปี 2551 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 18.356 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลงฐานข้อมูลใช้เวลาประมาณ 1.9959 วินาที นำเข้าฐานข้อมูล ใช้เวลาประมาณ 1.8402 วินาที ตัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 552.4705 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 9.4 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0846 วินาที ในปี 2552 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 18.5301 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลงฐานข้อมูลใช้เวลาประมาณ 1.962 วินาที นำเข้าฐานข้อมูลใช้เวลาประมาณ 1.9466 วินาที ตัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 641.7019 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 10.1851 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0741 วินาที ในปี 2553 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 19.6109 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลงฐานข้อมูลใช้เวลาประมาณ 2.5034 วินาที นำเข้าฐานข้อมูลใช้เวลาประมาณ 2.3423 วินาที ตัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 1070.4439 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 15.1384 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0525 วินาที ในปี 2554 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 20.6766 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลงฐานข้อมูลใช้เวลาประมาณ 2.9276 วินาที นำเข้าฐานข้อมูลใช้เวลาประมาณ 2.6111 วินาที ตัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 1115.623 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 17.2024 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0656 วินาที ในปี 2555 ใช้ความเร็วการดึงข้อมูลจากแหล่งอื่นใช้เวลาประมาณ 27.5512 วินาที เก็บข้อมูลบันทึกเป็น CSV เพื่อเป็นไฟล์ชั่วคราวก่อนลง

ฐานข้อมูลใช้เวลาประมาณ 3.9788 วินาที นำเข้าฐานข้อมูลใช้เวลาประมาณ 3.4183 วินาที คัดเรคอร์ดที่ซ้ำใช้เวลาประมาณ 1750.6508 วินาที แยกเป็นกลุ่มคำที่ค้นหาใช้เวลาประมาณ 25.9792 วินาที เขียน XML เพื่อทำระบบ autocomplete ใช้เวลาประมาณ 0.0503 วินาที สามารถสรุปได้ว่ายังใช้โปรแกรม Crawler ในการเก็บข้อมูลที่มีความล่าช้าของข้อมูลเท่าไร ยิ่งใช้เวลานานมากขึ้นเท่านั้น อาจเป็นไปได้ว่าแต่ละปีมีข้อมูลที่เพิ่มมากขึ้น ยิ่งข้อมูลมีขนาดใหญ่ การเก็บข้อมูลก็จะช้าลงอย่างเห็นได้ชัด

## 5.2 ความเร็วในการค้นหาข้อมูล

### 5.2.1 ความเร็วในการค้นหาข้อมูลแต่ละคำ

จากผลการทดลอง ผู้วิจัยได้ลองค้นหาคำที่กว้าง ในตัวอย่างผู้วิจัยใช้คำว่า KASSET ซึ่งเป็นชื่อของบจก. คำหนึ่ง โดยแสดงผลลัพธ์การค้นหา จำนวน 130,303 เรคอร์ด จำนวนหน้าทั้งหมด 13,031 หน้า ตำแหน่งหน้าปัจจุบันที่ 1 และใช้เวลาค้นหา ประมาณ 6.3871 วินาที และผู้วิจัยได้ลองค้นหาคำที่เฉพาะเจาะจง ในตัวอย่างผู้วิจัยใช้คำว่า K-MONEY ซึ่งเป็นชื่อย่อของกองทุนรวม คำหนึ่ง โดยแสดงผลลัพธ์การค้นหา จำนวน 1,083 เรคอร์ด จำนวนหน้าทั้งหมด 109 หน้า ตำแหน่งหน้าปัจจุบันที่ 1 และใช้เวลาค้นหาประมาณ 5.7739 วินาที สรุปได้ว่ายังใช้คำที่เฉพาะเจาะจงมากเท่าไร ยังมีผลลัพธ์ในการค้นหาที่แคบลง และค้นหาได้รวดเร็วมากขึ้นเท่านั้น

### 5.2.2 ความเร็วในการค้นหาข้อมูลแต่ละหน้า

จากผลการทดลอง ผู้วิจัยได้ลองค้นหาคำ ในตัวอย่างผู้วิจัยใช้คำว่า SCBAM ซึ่งเป็นชื่อของบจก. คำหนึ่ง โดยแสดงผลลัพธ์การค้นหา จำนวน 134,693 เรคอร์ด จำนวนหน้าทั้งหมด 13,470 หน้า ตำแหน่งหน้าปัจจุบันที่ 1 และใช้เวลาค้นหา ประมาณ 6.3294 วินาที ผู้วิจัยได้ลองเปลี่ยนไปหน้าถัดไป จึงทำให้ช่อง URL มีการเพิ่มเติมรายละเอียดเล็กน้อยจากเครื่องมือค้นหาเพื่อใช้เป็นคำอ้างอิงในแต่ละหน้า อีกทั้งยังทำให้การเครื่องมือบอกตำแหน่งหน้าปัจจุบันมีการเปลี่ยนแปลงจากตำแหน่งเดิมเล็กน้อย โดยแสดงผลลัพธ์การค้นหาเท่าเดิม จำนวนหน้าทั้งหมดเท่าเดิม ตำแหน่งหน้าปัจจุบันที่ 2 และใช้เวลาค้นหาประมาณ 6.0395 วินาที สรุปได้ว่าแทบไม่มีผลต่อความเร็วในการค้นคืน เมื่อเปลี่ยนข้อมูลเป็นหน้าถัดไป

### 5.2.3 ความเร็วในการค้นหาข้อมูลเมื่อใส่เครื่องหมาย Quote

จากผลการทดลอง ผู้วิจัยได้ลองค้นหาคำ ในตัวอย่างผู้วิจัยใช้คำว่า EQ ซึ่งเป็นชื่อประเภทกองทุนรวม โดยแสดงผลลัพธ์การค้นหา จำนวน 719,786 เรคอร์ด จำนวนหน้าทั้งหมด 71,979 หน้า ตำแหน่งหน้าปัจจุบันที่ 1 และใช้เวลาค้นหา ประมาณ 6.5723 วินาที ผู้วิจัยได้ลองใส่เครื่องหมาย Quote ด้านหน้ากับด้านหลังคำเดิมที่ต้องการค้นหา ทำให้กลายเป็น 'Quote' ในทันที

อีกทั้งยังพบว่าแสดงผลการค้นหา จำนวน 524,960 เรคอร์ด จำนวนหน้าทั้งหมด 52,496 หน้า ตำแหน่งหน้าปัจจุบันที่ 1 และใช้เวลาค้นหา ประมาณ 4.0619 วินาที สรุปได้ว่าการใส่เครื่องหมาย Quote ด้านหน้ากับด้านหลังคำเคมที่ต้องการค้นหา ทำให้มีผลลัพธ์ในการค้นหาที่แคบลง และมีความเร็วในการค้นหาคำเพิ่มขึ้นอีกด้วย

### 5.3 ข้อเสนอแนะ

ผู้วิจัยมองว่าเป็นโปรแกรมที่ไม่สมบูรณ์พอที่จะเรียกว่า Search Engine เนื่องจากเป็นการค้นคืนจากฐานข้อมูลโดยไม่ผ่านการทำดัชนีคำค้น (Indexer) แบบเต็มรูปแบบ อีกทั้งโปรแกรมก็ทำงานช้าอยู่





ป  
ร  
ร  
ณ  
า  
น  
ุ  
ก  
ร  
ม

บรรณานุกรม

## บรรณานุกรม

### ภาษาไทย

#### หนังสือ

- ดร. ธนัยวงศ์ กิรติวานิชย์ และภัศรา ชาวลกร. (2547). **รู้วิเคราะห์ เจาะเรื่อง กองทุนรวม**.กรุงเทพฯ: ตลาดหลักทรัพย์แห่งประเทศไทย.
- ผศ.ดร. คະชา ชาญศิลป์. (2553). **คู่มือเรียน Web Programming ด้วย PHP, MySQL และ AJAX**. กรุงเทพฯ: โปรวิชั่น.
- กาญจนา ตันวิสุทธิ. (2551). **เก่ง AJAX+PHP ให้ครบสูตร**.กรุงเทพฯ: วิตดี กรุ๊ป.
- นิรุช อำนวยศิลป์. (2548). **PHP How-To and Web-Based Application Techniques เทคนิค PHP เพื่อการต่อยอด**.กรุงเทพฯ: ดวงกมลสมัย.
- นิรุช อำนวยศิลป์. (2548). **PHP How-To and Web-Based Application Techniques เทคนิค PHP เพื่อการต่อยอด 2**.กรุงเทพฯ: ดวงกมลสมัย.
- นิรุช อำนวยศิลป์. (2548). **PHP How-To and Web-Based Application Techniques เทคนิค PHP เพื่อการต่อยอด 3**.กรุงเทพฯ: ดวงกมลสมัย.
- สาธิต ชัยวิวัฒน์ตระกูล. (2547). **เติมเทคนิค MySQL ให้เต็มประสิทธิภาพ**.กรุงเทพฯ: วิตดี กรุ๊ป.

#### บทความ

- สิทธิโชค ปัญญาฤกษ์ชัย และ ศิพาลี นุชิตประสิทธิ์ชัย. (2552). "ระบบการค้นคืนสารสนเทศโดยใช้เทคนิค N-Gram." วารสารการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ(เล่ม 5, หน้า 307-312).กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

#### วิทยานิพนธ์

จิรัชยา นครชัย. (2553). ระบบจัดการเอกสารอิเล็กทรอนิกส์. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยี  
มหานคร.

### สารสนเทศจากสื่ออิเล็กทรอนิกส์

สมาคมบริษัทจัดการลงทุน. (2549). ผลการดำเนินงานของรายกองทุนรวม. สืบค้นเมื่อ 28

กุมภาพันธ์ 2554, จาก <http://www.aimc.or.th>

ตลาดหลักทรัพย์แห่งประเทศไทย. (2555). ข้อมูลการซื้อขายหลักทรัพย์. สืบค้นเมื่อ 1 มิถุนายน

2555, จาก <http://www.settrade.com>

ไทยครีเอท. (2553). PHP Upload and Convert/Import CSV To MySQL. สืบค้นเมื่อ 12 เมษายน

2554, จาก <http://www.thaicreate.com/php/forum/051489.html>

ไทยครีเอท. (2553). PHP Upload and Convert/Import CSV To MySQL. สืบค้นเมื่อ 28 กุมภาพันธ์

2554, จาก <http://www.thaicreate.com/php/php-upload-convert-import-csv-to-mysql.html>

ไทยครีเอท. (2553). PHP Convert and Import CSV to MySQL Database. สืบค้นเมื่อ 28

กุมภาพันธ์ 2554, จาก <http://www.thaicreate.com/php/php-convert-csv-to-mysql.html>

สมาคมบริษัทจัดการลงทุน และ ตลาดหลักทรัพย์แห่งประเทศไทย. (2545). เรื่องนำร่องของกองทุน

รวม. สืบค้นเมื่อ 28 กุมภาพันธ์ 2554, จาก <http://www.thaimutualfund.com>

นกลีเก. (2554). Webbot programming. สืบค้นเมื่อ 28 กุมภาพันธ์ 2554, จาก

[http://www.noklek.com/wiki2/index.php/Webbot\\_programming](http://www.noklek.com/wiki2/index.php/Webbot_programming)

นกลีเก. (2554). Webbot, Spiders and Screen Scrapers. สืบค้นเมื่อ 4 เมษายน 2554, จาก

[http://www.noklek.com/wiki2/index.php/Webbot,\\_Spiders\\_and\\_Screen\\_Scrapers](http://www.noklek.com/wiki2/index.php/Webbot,_Spiders_and_Screen_Scrapers)

ภาษาต่างประเทศ

**BOOKS**

Deitel,P.J. and Deitel,H.M..(2009). **Internet & World Wide Web How to Program.**(4th ed.).

NJ: Pearson Education International.

Croft,BruceW., Metzler,Donald and Strohman,Trevor. (2010).**Search Engines Information**

**Retrieval in Practice.**NJ: Pearson Education International.

Schrenk,Michael. (2007). **Webbots, Spiders, and Screen Scrapers.**CA: No Starch Press.

**ARTICLES**

Transier, Frederik and Sanders,Peter. (2010). "Engineering Basic Algorithms of an In-Memory Text Search Engine."Proceeding of **ACM Symposium on User Interface Software and Technology, 2010.**

Bar-Yossef, Ziv and Gurevich,Maxim. (2011). "Efficient Search Engine Measurement." Proceeding of **ACM Symposium on User Interface Software and Technology, 2011.**

Meng, Weiyi., Yu Clement and Liu King-Lup. (2002). "Building Efficient and Effective Metasearch Engine."Proceeding of **ACM Symposium on User Interface Software and Technology, 2002.**

Kobayashi, Mei and Takeda,Koichi. (2000). "Information Retrieval on the web."Proceeding of **ACM Symposium on User Interface Software and Technology, 2000.**

**ELECTRONIC SOURCES**

Schrenk, Michael. (2007). Webbot, Spiders and Screen Scrapers. Retrieved February 28, 2011, from <http://www.schrenk.com/nostarch/webbots>

Schrenk, Michael. (2007). Webbot, Spiders and Screen Scrapers. Retrieved February 28, 2011, from [http://www.schrenk.com/nostarch/webbots/DSP\\_download.php](http://www.schrenk.com/nostarch/webbots/DSP_download.php)

Wikipedia. (2012). Web search engine. Retrieved January 23, 2012, from

**[http://en.wikipedia.org/wiki/Web\\_search\\_engine](http://en.wikipedia.org/wiki/Web_search_engine)**

Wikipedia. (2012). Web crawling. Retrieved January 23, 2012, from

**[http://en.wikipedia.org/wiki/Web\\_crawling](http://en.wikipedia.org/wiki/Web_crawling)**

Wikipedia. (2012). Search engine indexing. Retrieved January 23, 2012, from

**[http://en.wikipedia.org/wiki/Index\\_\(search\\_engine\)](http://en.wikipedia.org/wiki/Index_(search_engine))**

Wikipedia. (2012). Web search query. Retrieved January 23, 2012, from

**[http://en.wikipedia.org/wiki/Web\\_search\\_query](http://en.wikipedia.org/wiki/Web_search_query)**



ภาคผนวก





**ภาคผนวก ก**  
**ความรู้เบื้องต้นเกี่ยวกับกองทุนรวม**

## ความรู้เบื้องต้นเกี่ยวกับกองทุนรวม

### 1. ความรู้เบื้องต้นเกี่ยวกับกองทุนรวม

#### 1.1 ความเป็นมา

บริษัทหลักทรัพย์จัดการกองทุนรวมแห่งแรกของประเทศไทยได้จัดตั้งขึ้นในปีพ.ศ. 2518 โดยความร่วมมือกันระหว่างรัฐบาลไทยและบรรษัทการเงินระหว่างประเทศ ( International Finance Corporation – IFC ) ซึ่งเป็นหน่วยงานหนึ่งของธนาคาร โลกและได้มีการจัดตั้งและจัดการกองทุนรวมโครงการแรกในปีพ.ศ. 2520 ภายใต้ชื่อ “โครงการกองทุนสินภิญโญ” ด้วยขนาดกองทุน 100 ล้านบาทและมีอายุโครงการ 10 ปี ในปีพ.ศ. 2535 พระราชบัญญัติหลักทรัพย์และตลาดหลักทรัพย์พ.ศ. 2535 มีผลบังคับใช้จึงได้มีการจัดตั้งสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ขึ้นเพื่อทำหน้าที่กำกับดูแลและส่งเสริมพัฒนาธุรกิจหลักทรัพย์ธุรกิจจัดการลงทุนและตลาดทุนของประเทศโดยในวันที่ 19 มีนาคมพ.ศ. 2535 กระทรวงการคลังได้พิจารณาให้ใบอนุญาตการจัดการกองทุนรวมแก่บริษัทหลักทรัพย์จัดการกองทุนรวมที่จัดตั้งขึ้นใหม่เพิ่มอีก 7 บริษัทรวมกับบริษัทเดิมอีก 1 บริษัทเป็น 8 บริษัทมีผลให้ธุรกิจการจัดการกองทุนรวมขยายตัวขึ้นอย่างรวดเร็วและกลายเป็นส่วนสำคัญในการพัฒนาตลาดทุนของประเทศไทย ปลายปีพ.ศ. 2538 สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ได้เปิดให้ยื่นคำขอรับใบอนุญาตประกอบธุรกิจหลักทรัพย์ประเภทการจัดการกองทุนรวมเพิ่มเติมซึ่งมีกลุ่มผู้ผ่านการพิจารณาของสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์จำนวน 7 กลุ่ม ( ไม่นับรวมผู้จัดตั้งบริษัทหลักทรัพย์จัดการกองทุนรวมเพื่อผู้ลงทุนต่างด้าวจำกัด ) โดยต่อมาในปีพ.ศ. 2539 – 2540 กลุ่มผู้ผ่านการพิจารณาดังกล่าวได้ทยอยจัดตั้งบริษัทขึ้นเพื่อรับใบอนุญาตประกอบธุรกิจหลักทรัพย์ประเภทการจัดการกองทุนรวมจากกระทรวงการคลังซึ่งมีกลุ่มที่สามารถดำเนินการจัดตั้งบริษัทได้เพียง 6 บริษัทรวมกับบริษัทหลักทรัพย์จัดการกองทุนรวมที่จัดตั้งแล้วก่อนหน้านี้เป็น 14 บริษัท ปัจจุบันณวันที่ 15 กุมภาพันธ์ 2551 มีบริษัทหลักทรัพย์จัดการกองทุนรวมทั้งสิ้น 23 บริษัทคือ

1.1.1 บริษัทหลักทรัพย์จัดการกองทุน กรุงศรี จำกัด เลขที่ 898 อาคารเพลลินจิตทาวเวอร์ ชั้น 1-2 โซนเอ และชั้น 12 ถนนเพลลิตจิต แขวงลุมพินี เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2657-5757 โทรสาร 0-2657-5777 <http://www.ayfunds.com>

1.1.2 บริษัทหลักทรัพย์จัดการกองทุน กรุงไทย จำกัด (มหาชน) เลขที่ 11 อาคารคิวเฮ้าส์ สาทร ชั้น G, M และ 10 ถนนสาทรใต้ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120 โทร 0-2670-4900 โทรสาร 0-2679-1820 <http://www.ktam.co.th>



1.1.3 บริษัทหลักทรัพย์จัดการกองทุน กสิกรไทย จำกัด เลขที่ 400/22 อาคารธนาคาร กสิกรไทย ชั้น 6 ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพฯ 10400 โทร 0-2673-3999 โทรสาร 0-2673-3900 <http://www.kasikornasset.com>

1.1.4 บริษัทหลักทรัพย์จัดการกองทุน ซีไอเอ็มบี-พริ้นซิเพิล จำกัด เลขที่ 44 ชั้น 16 ถนน หลังสวน แขวงลุมพินี เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2686-9500 โทรสาร 0-2657-3167 <http://www.cimb-principal.co.th>

1.1.5 บริษัทหลักทรัพย์จัดการกองทุน ทหารไทย จำกัด เลขที่ 990 อาคารอับดุลราฮิมเพลส ชั้น 32 ถนนพระราม 4 แขวงสีลม เขตบางรัก กรุงเทพฯ 10500 โทร 0-2636-1800 โทรสาร 0-2636-1820 <http://www.tmbam.com>

1.1.6 บริษัทหลักทรัพย์จัดการกองทุนรวม ทองคำ แอสเซท จำกัด 100/43 อาคารสาทร นครทาวเวอร์ ชั้น 23 ถนนสาทรเหนือ แขวงสีลม เขตบางรัก กรุงเทพมหานคร 10500 โทร 0-2685-0555 โทรสาร 0-2636-7960 <http://www.thongcomeasset.com>

1.1.7 บริษัทหลักทรัพย์จัดการกองทุน ทิสโก้ จำกัด เลขที่ 48/16 -- 17 อาคารทิสโก้ ทาวเวอร์ ชั้น 9 ถนนสาทรเหนือ แขวงสีลม เขตบางรัก กรุงเทพฯ 10500 โทร 0-2633-6000 โทรสาร 0-2633-7300 <http://www.tiscoasset.com>

1.1.8 บริษัทหลักทรัพย์จัดการกองทุน ธนชาติ จำกัด เลขที่ 898 อาคารเพลินิจิตทาวเวอร์ ชั้น 15,18 ถนนเพลินิจิต แขวงลุมพินี เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2263-0800 โทรสาร 0-2263-0811-4 <http://www.thanachartfund.com>

1.1.9 บริษัทหลักทรัพย์จัดการกองทุน ฟินันซ่า จำกัด เลขที่ 48/21, 48/24 อาคารทิสโก้ ทาวเวอร์ ชั้น 12 เอ ถนนสาทรเหนือ แขวงสีลม เขตบางรัก กรุงเทพฯ 10500 โทร 0-2352-4000 โทรสาร 0-2352-4098-9 <http://www.finansa-asset.com>

1.1.10 บริษัทหลักทรัพย์จัดการกองทุน ยูโอบี (ไทย) จำกัด เลขที่ 191 ชั้น 11 ถนนสาทรใต้ แขวงยานนาวา เขตสาทร กรุงเทพฯ 10120 โทร. 0-2676-7100 โทรสาร 0-2676-7880-7 <http://www.uobam.co.th>

1.1.11 บริษัทหลักทรัพย์จัดการกองทุน อเบอร์ดีน จำกัด เลขที่ 179 อาคารบางกอกซิตี ทาวเวอร์ ชั้น 28 ถนนสาทรใต้ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120 โทร 0-2352-3333 โทรสาร 0-2352-3388 <http://www.aberdeen-asset.co.th>

1.1.12 บริษัทหลักทรัพย์จัดการกองทุน เกียรตินาคิน จำกัด เลขที่ 942/135 อาคารชาญ อิศระทาวเวอร์ ชั้น 4 ถนนพระราม4 แขวงสุริวงส์ เขตบางรัก กรุงเทพฯ 10500 โทร. 0-2624-8500 โทรสาร 0-2624-8599 <http://www.kk-fund.com>

1.1.13 บริษัทหลักทรัพย์จัดการกองทุน เอ็มเอฟซี จำกัด ( มหาชน ) เลขที่ 199 ชั้น G และ ชั้น 21-23 อาคารคอลลัมน์ทาวเวอร์ ถนนรัชดาภิเษก เขตคลองเตย กรุงเทพฯ 10110 โทร 0-2649-2000 โทรสาร 0-2649-2100, 0-2649-2111 <http://www.mfcfund.com>

1.1.14 บริษัทหลักทรัพย์จัดการกองทุนแมนูไลฟ์ (ประเทศไทย) จำกัด เลขที่ 364/30 อาคารแมนูไลฟ์เพลส ชั้น 6 ถนนศรีอยุธยา แขวงถนนพญาไท เขตราชเทวี กรุงเทพฯ 10400 โทรศัพท์ 0-2246-7650 โทรสาร 0-2642-6341 <http://www.manulife-asset.co.th>

1.1.15 บริษัทหลักทรัพย์จัดการกองทุนรวม แลนด์ แอนด์ เฮ้าส์ จำกัด เลขที่ 11 อาคารลิฟต์ เฮ้าส์ สแควร์ ชั้น 14 ถนนสาทรใต้ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120 โทร 0-2286-3484 โทรสาร 0-2286-3585 <http://www.lhfund.co.th>

1.1.16 บริษัทหลักทรัพย์จัดการกองทุน แอสเซท พลัส จำกัด เลขที่ 175 อาคารสาทรซิตี้ ทาวเวอร์ ชั้น 17 ถนนสาทรใต้ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120 โทร. 0-2672-1000 โทรสาร 0-2286-4472 <http://www.assetfund.co.th>

1.1.17 บริษัทหลักทรัพย์จัดการกองทุน ไทยพาณิชย์ จำกัด เลขที่ 19 ไทยพาณิชย์ ปาร์ค พลาซ่า อาคาร 3 ชั้น 21-22 ถนนรัชดาภิเษก แขวงจตุจักร เขตจตุจักร กรุงเทพฯ 10900 โทร. 0-2949-1500 โทรสาร 0-2949-1501 <http://www.scbam.com>

1.1.18 บริษัทหลักทรัพย์จัดการกองทุน ไอเอ็นจี (ประเทศไทย) จำกัด เลขที่ 130-132 อาคารสินทรทาวเวอร์ 3 ชั้น 15 และ 17 ถนนวิฑู แขวงลุมพินี เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2688-7777 โทรสาร 0-2688-7700 <http://www.ingfunds.co.th>

1.1.19 บริษัทหลักทรัพย์จัดการกองทุนรวม กิมเอ็ง จำกัด เลขที่ 999/9 อาคารสำนักงาน ดิ ออฟฟิศเสสแอทเซ็นทรัล วิลด์ ชั้น 20 ถนนพระรามที่ 1 แขวงปทุมวัน เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2264-5111 โทรสาร 0-2264-5132 <http://www.kcat.co.th>

1.1.20 บริษัทหลักทรัพย์จัดการกองทุนรวม ซีมิโก้ จำกัด เลขที่ 287 อาคารลิเบอร์ตีส์แควร์ ชั้น 8 ถนนสีลม แขวงสีลม เขตบางรัก กรุงเทพฯ 10500 โทร 0-2624-6300 โทรสาร 0-2624-6330 <http://www.seamicoasset.com>

1.1.21 บริษัทหลักทรัพย์จัดการกองทุนรวม บัวหลวง จำกัด เลขที่ 175 อาคารสาทรซิตี้ทาวเวอร์ ชั้น 7 ชั้น 21 และ ชั้น 26 ถนนสาทรใต้ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120 โทร 0-2674-6488 โทรสาร 0-2679-5995-6 <http://www.bblam.co.th>

1.1.22 บริษัทหลักทรัพย์จัดการกองทุนรวม ฟิลลิป จำกัด เลขที่ 849 อาคารรววัฒน์ ชั้น 22 ถนนสีลม แขวงสีลม เขตบางรัก กรุงเทพฯ 10500 โทร. 0-2635-3033 โทรสาร 0-2635-3040 <http://www.phillipasset.co.th>

1.1.23 บริษัทหลักทรัพย์จัดการกองทุนรวม วรณ จำกัด เลขที่ 989 อาคารสยามทาวเวอร์ ชั้น 24 ( สยามดิศพ์เวร์ ) ถนนพระราม 1 แขวงปทุมวัน เขตปทุมวัน กรุงเทพฯ 10330 โทร 0-2659-8888 โทรสาร 0-2659-8860-1 <http://www.one-asset.com>

1.2 กองทุนรวมคือเครื่องมือในการลงทุน (investment vehicle) สำหรับผู้ลงทุนรายย่อยที่ประสงค์จะนำเงินมาลงทุนในตลาดเงินตลาดทุนแต่ติดขัดด้วยอุปสรรคหลายประการที่ทำให้การลงทุนด้วยตนเองไม่สามารถได้ผลลัพธ์ตามเป้าหมายที่ต้องการเช่น

1.2.1 มีทุนทรัพย์จำนวนจำกัดไม่สามารถกระจายการลงทุนในหลักทรัพย์ต่างประเทศได้มากพอเพื่อลดความเสี่ยงจากการลงทุนหรือ

1.2.2 ไม่มีประสบการณ์ความรู้ความชำนาญในการลงทุนหรือ

1.2.3 ไม่มีเวลาจะศึกษาค้นหาและติดตามข้อมูลเพื่อใช้ในการตัดสินใจการลงทุน

1.2.4 กองทุนรวมจึงเป็นเครื่องมือในการลงทุนที่มีประสิทธิภาพมีการจัดการลงทุนอย่างเป็นระบบโดยมีจุดมุ่งหมายให้การลงทุนได้รับผลตอบแทนที่ดีที่สุดภายใต้กรอบความเสี่ยงที่ผู้ลงทุนยอมรับได้

1.3 ผู้เกี่ยวข้องกักองทุนรวม ภายในโครงสร้างของกองทุนรวมถูกกำหนดขึ้นเพื่อผลประโยชน์ของผู้ถือหุ้นลงทุนโดยประกอบด้วยผู้ที่ได้รับมอบหมายให้ทำหน้าที่ต่างๆและผู้กำกับดูแลทั้งที่เป็นองค์กรของภาคเอกชนและภาครัฐได้แก่

1.3.1 บริษัทจัดการบริษัทจัดการต้องเป็นบริษัทหลักทรัพย์ที่ได้รับใบอนุญาตจัดการลงทุนจากกระทรวงการคลังเท่านั้นบริษัทจัดการเป็นผู้กำหนดโครงการกองทุนรวมนโยบายการลงทุนและวัตถุประสงค์เพื่อนำเสนอขออนุมัติจากสำนักงานก.ล.ต. และต้องบริหารจัดการลงทุนตามวัตถุประสงค์และนโยบายการลงทุนนั้นโดยเคร่งครัดทั้งนี้บริษัทจัดการจะแจ้งนโยบายการลงทุนและวัตถุประสงค์ในการลงทุนให้ผู้ลงทุนทราบในหนังสือชี้ชวนเสนอขายหน่วยลงทุนที่แจกจ่ายให้แก่ผู้ลงทุนและผู้สนใจลงทุนได้ศึกษาก่อนที่จะลงทุน

1.3.2 ผู้ดูแลผลประโยชน์ผู้ดูแลผลประโยชน์เป็นสถาบันการเงินที่มีคุณสมบัติตามที่สำนักงานก.ล.ต. กำหนดและต้องไม่มีส่วนเกี่ยวข้องทั้งทางตรงและทางอ้อมกับบริษัทจัดการผู้ดูแลผลประโยชน์จะเป็นตัวแทนของผู้ถือหุ้นลงทุนทำหน้าที่รักษาผลประโยชน์ทั้งหมดของผู้ถือหุ้นลงทุนเช่น

1.3.2.1 ดูแลให้บริษัทจัดการจัดการกองทุนให้เป็นไปตามวัตถุประสงค์และนโยบายการลงทุนของโครงการลงทุนที่ได้รับอนุมัติจากสำนักงานก.ล.ต. และที่ได้จัดแจ้งไว้ในหนังสือชี้ชวน

1.3.2.2 ทำหน้าที่ชำระราคาซื้อและรับชำระราคาจากการขายทรัพย์สิน

1.3.2.3 เก็บรักษาทรัพย์สินทั้งหมดของกองทุนรวม

1.3.2.4 สอบทานความถูกต้องของมูลค่าทรัพย์สินของกองทุนรวม

1.3.2.5 ดำเนินคดีฟ้องร้องแทนผู้ถือหน่วยลงทุนหากบริษัทจัดการปฏิบัติหน้าที่

โดยมิชอบ

1.3.3 ตัวแทนสนับสนุนการขายหน่วยลงทุนปัจจุบันบุคคลที่จะทำหน้าที่เสนอขายหน่วยลงทุนของกองทุนรวมได้ต้องเป็นบุคคลที่ได้รับอนุญาตจากสำนักงานก.ล.ด. เท่านั้นตัวแทนสนับสนุนขายหน่วยลงทุนต้องมีคุณสมบัติตามที่กฎหมายกำหนดและผ่านการทดสอบความรู้ในหลักสูตรการเป็นตัวแทนขายจากสถาบันที่สำนักงานก.ล.ด. เห็นชอบขึ้นทะเบียนรายชื่อกับสำนักงานก.ล.ด. ต้องปฏิบัติและทำหน้าที่ในการขายตามกรอบที่กฎหมายกำหนดเพื่อป้องกันการขายและการโฆษณาชวนเชื่อให้ผู้ลงทุนเข้าใจผิดในสาระสำคัญตัวแทนสนับสนุนขายหน่วยลงทุนหรือที่เรียกกันอย่างเป็นทางการว่าผู้ทำหน้าที่ขายหรือรับซื้อคืนหน่วยลงทุนกองทุนรวมมีสองระดับได้แก่

1.3.3.1 ตัวแทนสนับสนุนการขาย-ระดับหนึ่ง [ Investment Planner (IP)] หมายถึงบุคคลที่สามารถให้คำแนะนำในการลงทุนในหน่วยลงทุนของกองทุนรวมแก่ผู้ถือหน่วยลงทุนและผู้สนใจลงทุนทั่วไปทั้งที่เป็นคำแนะนำทั่วไปและคำแนะนำเฉพาะเจาะจง

1.3.3.2 ตัวแทนสนับสนุนการขาย-ระดับสอง [ Fundamental Guide (FG) ] หมายถึงบุคคลที่สามารถให้คำแนะนำในการลงทุนในหน่วยลงทุนของกองทุนรวมแก่ผู้ถือหน่วยลงทุนและผู้สนใจลงทุนทั่วไปเฉพาะที่เป็นคำแนะนำทั่วไปเท่านั้น

1.3.4 นายทะเบียนหน่วยลงทุนนายทะเบียนหน่วยลงทุนเป็นสถาบันการเงินที่ได้รับอนุญาตจากสำนักงานก.ล.ด. ให้มีหน้าที่ดูแลทะเบียนรายชื่อผู้ถือหน่วยลงทุนตลอดจนสิทธิประโยชน์ของผู้ถือหน่วยลงทุนเช่นการจ่ายเงินปันผลและสิทธิประโยชน์อื่นๆบริษัทจัดการอาจทำหน้าที่เป็นนายทะเบียนหน่วยลงทุนสำหรับกองทุนรวมภายใต้การจัดการของตนก็ได้

1.3.5 ผู้สอบบัญชีรับอนุญาตผู้สอบบัญชีรับอนุญาตเป็นบุคคลที่ได้รับอนุญาตให้เป็นผู้สอบบัญชีและมีชื่อขึ้นทะเบียนไว้กับสำนักงานก.ล.ด. ต้องไม่มีส่วนเกี่ยวข้องทั้งทางตรงและทางอ้อมกับบริษัทจัดการมีหน้าที่ตรวจสอบบัญชีทรัพย์สินของกองทุนรวมตรวจสอบและให้ความเห็นชอบงบการเงินของกองทุนให้ถูกต้องตามมาตรฐานบัญชี

1.3.6 สมาคมบริษัทจัดการลงทุนสมาคมบริษัทจัดการลงทุนจัดตั้งขึ้นภายใต้พระราชบัญญัติหลักทรัพย์และตลาดหลักทรัพย์.ศ. 2535 เป็นสมาคมที่เกี่ยวข้องกับธุรกิจหลักทรัพย์จดทะเบียนสมาคมกับสำนักงานก.ล.ด. มีบริษัทจัดการที่ประกอบธุรกิจหลักทรัพย์ประเภทการจัดการกองทุนรวมการจัดการกองทุนส่วนบุคคลและการจัดการกองทุนส่วนบุคคลที่

เป็นกองทุนสำรองเลี้ยงชีพเป็นสมาชิกสมาคมมีหน้าที่กำหนดจรรยาบรรณและวางมาตรฐานในการปฏิบัติให้บริษัทสมาชิกยึดถือและปฏิบัติเป็นมาตรฐานเดียวกันทุกรายกำหนดบทลงโทษเมื่อบริษัทสมาชิกฝ่าฝืนและไม่ปฏิบัติตาม

1.3.7 สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์สำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์เป็นองค์กรของภาครัฐทำหน้าที่กำกับดูแลธุรกิจหลักทรัพย์รวมถึงการจัดการลงทุนออกระเบียบข้อบังคับประกาศหรือข้อกำหนดตามความในกฎหมายว่าด้วยหลักทรัพย์และตลาดหลักทรัพย์

1.4 ด้วยกองทุนรวมเป็นเสมือนหนึ่งเครื่องมือในการลงทุนของผู้ลงทุนดังนั้นจึงต้องมีความหลากหลายเพื่อให้มีความเหมาะสมกับแต่ละลักษณะของผู้ลงทุนโดยทั่วไปกองทุนรวมสามารถแบ่งออกได้ดังนี้

1.4.1 แบ่งตามประเภทของการขายคืนหน่วยลงทุนสามารถแบ่งได้ 2 ประเภท

1.4.1.1 กองทุนปิด (Closed-End fund) กองทุนรวมที่มีหน่วยลงทุนคงที่ไม่เพิ่มขึ้นและไม่ลดลงและเปิดให้มีการจองซื้อเพียงครั้งเดียวเมื่อจัดตั้ง โครงการมีกำหนดอายุโครงการแน่นอนและบริษัทจัดการไม่รับซื้อคืนหน่วยลงทุนก่อนครบกำหนดอายุโครงการผู้ถือหน่วยลงทุนไม่สามารถไถ่ถอนหน่วยลงทุนก่อนครบกำหนดอายุโครงการได้โดยส่วนใหญ่แล้วอายุโครงการของกองทุนรวมในประเทศไทยจะมีกำหนด 3 ปี 5 ปีหรือ 10 ปีและเพื่อเพิ่มสภาพคล่องให้แก่ผู้ถือหน่วยลงทุนบริษัทจัดการอาจนำหน่วยลงทุนของกองทุนปิดไปจดทะเบียนซื้อขายในตลาดรอง (ตลาดหลักทรัพย์แห่งประเทศไทย) หรือจัดให้มีตัวแทนจัดการซื้อขาย (Market maker)

1.4.1.2 กองทุนเปิด (Open-End fund) กองทุนรวมที่สามารถเพิ่มหรือลดหน่วยลงทุนได้ไม่มีกำหนดอายุโครงการและบริษัทจัดการรับซื้อคืนหน่วยลงทุนตามกำหนดเวลาที่ระบุไว้ในหนังสือชี้ชวนเช่นทุกวันทุกสัปดาห์ทุกสองสัปดาห์ทุกเดือนทุกไตรมาสหรือทุกหกเดือนกองทุนเปิดจึงเป็นที่นิยมมากกว่ากองทุนปิดเพราะมีสภาพคล่องมากกว่า

ตารางที่ 6.1 เปรียบเทียบระหว่างกองทุนปิดกับกองทุนเปิด

ตารางเปรียบเทียบกองทุนปิดและกองทุนเปิด		
	กองทุนปิด	กองทุนเปิด
จำนวนหน่วยลงทุน	กำหนดแน่นอนไม่เพิ่มไม่ลด	สามารถเพิ่มหรือลดลงได้
อายุโครงการ	มีกำหนดแน่นอน	ไม่มีกำหนด (evergreen)
การซื้อขายหน่วยลงทุน	เปิดให้จองซื้อครั้งเดียวเมื่อเริ่มโครงการหากประสงค์ซื้อเพิ่มในภายหลังต้องเข้าซื้อในตลาดรอง (กรณีบริษัทจัดการนำหน่วยลงทุนเข้าจดทะเบียนซื้อขาย) หรือแสดงความจำนงกับตัวแทนขาย (market maker) ที่บริษัทจัดการแต่งตั้ง	สามารถซื้อเพิ่มจำนวนหน่วยกับบริษัทจัดการโดยตรงหรือติดต่อผ่านตัวแทนสนับสนุนการขายที่บริษัทจัดการแต่งตั้งทั้งที่เป็นบุคคลธรรมดาและนิติบุคคล เช่น ธนาคารหรือบริษัทหลักทรัพย์ซึ่งจะทำหน้าที่ส่งคำสั่งซื้อขายมายังบริษัทจัดการ
การขายคืนหน่วยลงทุน	บริษัทจัดการไม่รับซื้อคืนหน่วยลงทุนจนกว่าจะครบอายุโครงการหากผู้ลงทุนมีความจำเป็นต้องใช้เงินต้องขายหน่วยลงทุนที่ถือไว้ในตลาดรองในราคาตลาดให้แก่ผู้ประสงค์ซื้อ	บริษัทจัดการรับซื้อคืนหน่วยลงทุนตามกำหนดเวลาที่ระบุไว้ในหนังสือชี้ชวน (รายวันรายสัปดาห์รายเดือน) ในราคาเท่ากับมูลค่าทรัพย์สินสุทธิหักด้วยค่าธรรมเนียม(ถ้ามี)
การจดทะเบียนซื้อขาย	นิยมจดทะเบียนซื้อขายหน่วยลงทุนในตลาดรองเช่น ตลาดหลักทรัพย์แห่งประเทศไทย	ไม่นิยมจดทะเบียนซื้อขายหน่วยลงทุนในตลาดรองเพราะสามารถซื้อขายผ่านตัวแทนสนับสนุนการขายได้อยู่แล้ว

ที่มา : ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

#### 1.4.2 แบ่งตามนโยบายการลงทุน 10 แบบมาตรฐานของสำนักงานก.ล.ต. มีดังนี้

1.4.2.1 กองทุนรวมตราสารแห่งทุน (Equity fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งตราสารทุน โดยเฉลี่ยในรอบปีบัญชีไม่น้อยกว่าร้อยละ 65 ของมูลค่าทรัพย์สินสุทธิของกองทุนรวมบริษัทจัดการต้องรายงานค่าเฉลี่ยการถือครองตราสารทุนให้สำนักงานก.ล.ต. ทราบทุกกรอบระยะเวลาสามเดือนหกเดือนเก้าเดือนและสิบสองเดือนของรอบบัญชีกองทุนหากค่าเฉลี่ยการถือครองตราสารทุนไม่ถึงร้อยละ 65 ในรอบระยะเวลาใดให้บริษัทจัดการแสดงผลโดยชัดเจนเพื่อที่สำนักงานก.ล.ต. จะได้นำไปเปิดเผยให้แก่ผู้ลงทุนและผู้สนใจลงทุนทราบต่อไป โดยทั่วไปแล้วกองทุนรวมตราสารแห่งทุนมีความเสี่ยงสูงกว่ากองทุนรวมที่มีนโยบายลงทุนในตราสารประเภทอื่นจึงเหมาะสำหรับผู้ลงทุนที่ยอมรับความเสี่ยงได้สูงและควรลงทุนเพื่อหวังผลที่ดีกว่าในระยะยาว

1.4.2.2 กองทุนรวมตราสารแห่งหนี้ (General fixed income fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้เฉพาะเงินฝากหรือหลักทรัพย์หรือทรัพย์สินอื่นหรือการขาดดอกผลโดยวิธีอื่นตามที่สำนักงานก.ล.ต. กำหนดหรือให้ความเห็นชอบให้กองทุนประเภทดังกล่าวลงทุนได้ห้ามมิให้กองทุนรวมตราสารแห่งหนี้ลงทุนหรือมีไว้ซึ่งตราสารทุนหรือตราสารกึ่งหนี้กึ่งทุน (หุ้นกู้แปลงสภาพ) ยกเว้นแต่สำนักงานก.ล.ต. จะพิจารณาอนุญาตเมื่อมีผู้ให้คำรับรองที่น่าเชื่อถือได้ว่าจะเป็นผู้รับซื้อตราสารทุนหลังการแปลงสภาพนั้นออกไปจากกองทุน โดยทั่วไปแล้วกองทุนรวมตราสารแห่งหนี้มีความเสี่ยงน้อยกว่ากองทุนรวมที่มีนโยบายลงทุนในตราสารทุนจึงเหมาะสำหรับผู้ลงทุนที่ยอมรับความเสี่ยงได้น้อยกว่า

1.4.2.3 กองทุนรวมตราสารแห่งหนี้ระยะยาว (Long-term fixed income fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้เฉพาะเงินฝากหรือหลักทรัพย์หรือทรัพย์สินอื่นหรือการขาดดอกผลโดยวิธีอื่นตามที่สำนักงานก.ล.ต. กำหนดหรือให้ความเห็นชอบให้กองทุนประเภทดังกล่าวลงทุนได้โดยกองทุนมีวัตถุประสงค์ที่จะดำรงพอร์ตโฟลิโอดูเรชัน (portfolio duration) ในขณะที่ขณะหนึ่งของกองทุนรวมนั้นมากกว่าหนึ่งปีขึ้นไปพอร์ตโฟลิโอดูเรชัน (portfolio duration) หมายถึงอายุถัวเฉลี่ยแบบถ่วงน้ำหนักของกระแสเงินที่ได้รับจากทรัพย์สินของกองทุนรวมพอร์ตโฟลิโอดูเรชันมากกว่าหนึ่งปีมีความหมายโดยทั่วไปว่าทรัพย์สินที่กองทุนลงทุนและมีไว้มีอายุเฉลี่ยมากกว่าหนึ่งปีเหมาะสำหรับผู้ลงทุนที่ยอมรับความเสี่ยงต่ำและสามารถลงทุนระยะยาวได้

1.4.2.4 กองทุนรวมตราสารแห่งหนี้ระยะสั้น (Short-term fixed income fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้เฉพาะเงินฝากหรือหลักทรัพย์หรือทรัพย์สินอื่นหรือการขาดดอกผลโดยวิธีอื่นตามที่สำนักงานก.ล.ต. กำหนดหรือให้ความเห็นชอบให้กองทุนประเภท

ดังกล่าวลงทุนได้โดยกองทุนมีวัตถุประสงค์ที่จะดำรงพอร์ตโฟลิโอคู่เรชัน (portfolio duration) ในขณะใดขณะหนึ่งของกองทุนรวมนั้นไม่เกินหนึ่งปีพอร์ตโฟลิโอคู่เรชันต่ำกว่าหนึ่งปีมีความหมายโดยทั่วไปว่าทรัพย์สินที่กองทุนลงทุนและมีไว้มีอายุเฉลี่ยน้อยกว่าหนึ่งปีเหมาะสำหรับผู้ลงทุนที่ต้องการลงทุนระยะสั้นและต้องการความเสี่ยงต่ำ

1.4.2.5 กองทุนรวมผสม (Balanced fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งหลักทรัพย์หรือทรัพย์สินอื่นหรือการหาผลตอบแทนโดยวิธีอื่นตามที่สำนักงาน ก.ล.ต. กำหนดหรือให้ความเห็นชอบให้กองทุนประเภทดังกล่าวลงทุนได้โดยมีวัตถุประสงค์ที่จะดำรงอัตราส่วนการลงทุนในหรือมีไว้ซึ่งตราสารทุนในขณะใดขณะหนึ่งไม่เกินร้อยละ 65 และไม่ต่ำกว่าร้อยละ 35 ของมูลค่าทรัพย์สินสุทธิของกองทุนรวมกองทุนผสมลงทุนในตราสารได้ทุกประเภทผู้จัดการกองทุนสามารถแสวงหาโอกาสลงทุนที่ดีกว่าได้ทั้งในตลาดตราสารทุนและตลาดตราสารหนี้แต่เป็นการจัดสรรเงินลงทุนประเภทสมดุลเพราะมีข้อกำหนดเกี่ยวกับ ceiling และ floor ในการลงทุนในตราสารทุนกองทุนผสมเหมาะสำหรับผู้ลงทุนที่ยอมรับความเสี่ยงได้ปานกลาง

1.4.2.6 กองทุนรวมผสมแบบยืดหยุ่น (Flexible portfolio fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งหลักทรัพย์หรือทรัพย์สินอื่นหรือการหาผลตอบแทนโดยวิธีอื่นตามที่สำนักงาน ก.ล.ต. กำหนดหรือให้ความเห็นชอบให้กองทุนประเภทดังกล่าวลงทุนได้ทั้งนี้การลงทุนในหรือมีไว้ซึ่งหลักทรัพย์หรือทรัพย์สินอื่นหรือการหาผลตอบแทนโดยวิธีอื่นดังกล่าวขึ้นกับการตัดสินใจลงทุนของผู้จัดการกองทุนรวมตามความเหมาะสมและสภาพการณ์ในแต่ละขณะกองทุนรวมผสมแบบยืดหยุ่นสามารถลงทุนในตราสารทุกประเภทเช่นเดียวกับกับกองทุนรวมผสมแต่ไม่มีข้อกำหนดเกี่ยวกับ ceiling และ floor ในการลงทุนในตราสารทุนแต่อย่างใดการจัดสรรเงินลงทุนของกองทุนรวมผสมแบบยืดหยุ่นระหว่างตลาดตราสารทุนและตลาดตราสารหนี้ขึ้นอยู่กับดุลพินิจของผู้จัดการกองทุนกองทุนรวมผสมแบบยืดหยุ่นเหมาะสำหรับผู้ลงทุนที่ยอมรับความเสี่ยงได้ปานกลาง

1.4.2.7 กองทุนรวมหน่วยลงทุน (Fund of funds) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งหน่วยลงทุนและใบสำคัญแสดงสิทธิที่จะซื้อหน่วยลงทุนของกองทุนรวมโดยเฉลี่ยในรอบปีบัญชีไม่น้อยกว่าร้อยละ 65 ของมูลค่าทรัพย์สินสุทธิของกองทุนรวมเนื่องจากกองทุนรวมมีข้อดีหลายประการที่สำคัญคือมีการกระจายการลงทุนความเสี่ยงจึงลดลงทั้งยังมีต้นทุนเฉลี่ยต่ำกว่ากองทุนรวมหน่วยลงทุนจึงรับเอาข้อได้เปรียบดังกล่าวมานอกจากนั้นแล้วกองทุนรวมหน่วยลงทุนยังกระจายการลงทุนไปในหลายกองทุนรวมภายใต้การจัดการของหลายผู้จัดการกองทุนและหลายบริษัทจัดการจึงเป็นการกระจายความเสี่ยงที่กว้างขวางกว่าข้อเสียของกองทุนรวมหน่วยลงทุนอยู่ที่มีค่าธรรมเนียมในการจัดการและค่าใช้จ่ายอื่นๆซ้ำซ้อน



1.4.2.8 กองทุนรวมใบสำคัญแสดงสิทธิ (Warrant fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งใบสำคัญแสดงสิทธิที่จะซื้อหุ้นใบสำคัญแสดงสิทธิที่จะซื้อหุ้นกู้ใบสำคัญแสดงสิทธิที่จะซื้อหน่วยลงทุนและใบสำคัญแสดงสิทธิที่จะซื้อหุ้นเพิ่มทุน โดยเฉลี่ยในรอบปีบัญชีไม่น้อยกว่าร้อยละ 65 ของมูลค่าทรัพย์สินสุทธิของกองทุนรวมการลงทุนในใบสำคัญแสดงสิทธิที่จะซื้อหุ้นมีความเสี่ยงสูงกองทุนประเภทนี้จึงมีความเสี่ยงสูงมาก

1.4.2.9 กองทุนรวมกลุ่มธุรกิจ (Sector fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งตราสารทุนของบริษัทที่มีธุรกิจหลักประเภทเดียวกันตามที่ตลาดหลักทรัพย์แห่งประเทศไทยกำหนด โดยเฉลี่ยในรอบปีบัญชีไม่น้อยกว่าร้อยละ 65 ของมูลค่าทรัพย์สินสุทธิของกองทุนรวมกองทุนรวมกลุ่มธุรกิจมีการลงทุนกระจุกตัวจึงมีความเสี่ยงสูงกว่ากองทุนรวมตราสารแห่งทุนโดยทั่วไป

1.4.2.10 กองทุนรวมตลาดเงิน (Money market fund) กองทุนรวมที่มีนโยบายการลงทุนในหรือมีไว้ซึ่งตราสารหนี้ที่มีคุณภาพและมีกำหนดชำระเงินต้นเมื่อทวงถามหรือมีอายุคงเหลือไม่เกิน 1 ปี กองทุนรวมตลาดเงินมีนโยบายการลงทุนที่คล้ายคลึงกับกองทุนรวมตราสารแห่งหนี้ระยะสั้นมีความเสี่ยงต่ำสุดเหมาะสำหรับการลงทุนระยะสั้นของผู้ลงทุนที่ไม่ต้องการความเสี่ยง

#### 1.5 กองทุนรวมประเภทพิเศษแบ่งออกเป็น 4 ประเภท ได้แก่

1.5.1 กองทุนรวมมีประกัน (Guarantee fund) กองทุนรวมมีประกันคือกองทุนรวมที่บริษัทจัดการจัดให้มีสถาบันการเงินเป็นผู้ประกันต่อผู้ถือหน่วยลงทุนว่าจะจ่ายเงินลงทุนหรือเงินลงทุนและผลตอบแทนตามจำนวนเงินที่ประกันไว้ (อาจจะเป็นส่วนหรือทั้งหมด) ให้แก่ผู้ถือหน่วยลงทุนเมื่อถือหน่วยลงทุนครบอายุตามระยะเวลาประกันที่กำหนดวัตถุประสงค์ของการจัดให้มีกองทุนรวมมีประกันก็เพื่อที่จะทำให้ผู้ลงทุนมีความมั่นใจว่าเงินลงทุนของตนจะไม่สูญนโยบายการลงทุนของกองทุนรวมมีประกันการลงทุนอาจเป็นแบบใดแบบหนึ่งในมาตรฐาน 10 แบบตามที่ได้กล่าวมาแล้วข้างต้นผู้ลงทุนต้องพิจารณาให้ดีก่อนการลงทุนสถาบันการเงินที่เป็นผู้ประกันของกองทุนรวมมีประกันต้องมีคุณสมบัติอย่างใดอย่างหนึ่งต่อไปนี้

1.5.1.1 เป็นธนาคารหรือสถาบันการเงินอื่นที่มีกฎหมายเฉพาะจัดตั้งขึ้นหรือธนาคารพาณิชย์ตามกฎหมายว่าด้วยการธนาคารพาณิชย์หรือบริษัทเงินทุนตามกฎหมายว่าด้วยการประกอบธุรกิจเงินทุนธุรกิจหลักทรัพย์และธุรกิจเครดิตฟองซิเอร์ทั้งนี้ผู้ประกันดังกล่าวจะต้องสามารถดำรงเงินกองทุนและกันเงินสำรองได้ตามหลักเกณฑ์ของกฎหมายที่ควบคุมการประกอบธุรกิจนั้น

1.5.1.2 เป็นธนาคารต่างประเทศที่ได้รับการจัดอันดับความน่าเชื่อถือในสี่อันดับแรกจากสถาบันการจัดอันดับความน่าเชื่อถือที่ได้รับการยอมรับจากสำนักงาน ก.ล.ต.

1.5.2 กองทุนรวมเพื่อการเลี้ยงชีพ (Retirement Mutual Fund : RMF) กองทุนรวมเพื่อการเลี้ยงชีพคือกองทุนรวมที่มีวัตถุประสงค์ในการส่งเสริมการออมและการลงทุนของบุคคลเพื่อเตรียมความพร้อมไว้สำหรับการเกษียณอายุที่มีคุณภาพผู้ลงทุนในกองทุนรวมเพื่อการเลี้ยงชีพจะได้รับสิทธิประโยชน์ทางภาษีที่มากกว่าการลงทุนในกองทุนรวมทั่วไปเพราะเงินลงทุนในกองทุนรวมเพื่อการเลี้ยงชีพได้รับการยกเว้นภาษีเงินได้ไม่เกินปีละ 300,000 บาท ทั้งนี้ให้นับรวมเงินลงทุนในกองทุนสำรองเลี้ยงชีพหรือกองทุนบำเหน็จบำนาญข้าราชการแล้วแต่กรณีผู้ลงทุนจะได้รับประโยชน์จากการประหยัดภาษีเงินได้ทันทีตั้งแต่ปีแรกที่เริ่มลงทุนการลงทุนในกองทุนรวมเพื่อการเลี้ยงชีพมีดังนี้

#### 1.5.2.1 กองทุนรวมเพื่อการเลี้ยงชีพเหมาะสำหรับใคร?

- 1) ผู้ประกอบอาชีพอิสระซึ่งแต่เดิมมาจากโอกาสสะสมเงินลงทุนแบบปลอดภาษีเพราะไม่มีระบบบำเหน็จบำนาญรองรับ
- 2) ลูกจ้างที่นายจ้างยังไม่พร้อมที่จะจัดให้มีกองทุนสำรองเลี้ยงชีพทำให้ลูกจ้างไม่สามารถสะสมเงินลงทุนเพื่อวัยเกษียณได้
- 3) ลูกจ้างหรือข้าราชการที่อยู่ในระบบบำเหน็จบำนาญอยู่แล้วและประสงค์ที่จะลงทุนมากกว่าเดิมเพื่อใช้สิทธิประโยชน์ทางภาษีให้เต็มวงเงิน 300,000 บาทตามที่รัฐบาลให้การสนับสนุนและส่งเสริม

#### 1.5.2.2 เงื่อนไขการลงทุนเพื่อให้ได้รับสิทธิประโยชน์ทางภาษีได้แก่

- 1) เงินลงทุนมาจากเงินได้พึงประเมินตามมาตรา 40 แห่งประมวลรัษฎากร
- 2) ผู้ลงทุนที่มีเงินได้ต้องลงทุนแบบผูกพันคือลงทุนอย่างสม่ำเสมอและต่อเนื่องแต่เมื่อมีความจำเป็นอาจระงับการลงทุนได้ไม่เกินหนึ่งปีติดต่อกันแต่หากผู้ลงทุนไม่มีเงินได้ในปีใดหรือหลายปีติดต่อกันผู้ลงทุนสามารถว่างเว้นการลงทุนได้โดยไม่ถือว่าเป็นการผิดเงื่อนไขการลงทุนและเมื่อมีเงินได้ก็ให้ลงทุนต่อไปโดยนับอายุการลงทุนตั้งแต่ปีแรกที่ลงทุน
- 3) เงินลงทุนขั้นต่ำต้องไม่น้อยกว่าร้อยละสามของรายได้หรือไม่น้อยกว่าห้าพันบาทต่อปีอย่างใดอย่างหนึ่งที่มีจำนวนเงินต่ำกว่าในการคำนวณรวมเงินลงทุนขั้นต่ำให้รวมเงินลงทุนในทุกๆกองทุนที่ลงทุนในปีนั้นๆ
- 4) เงินลงทุนขั้นสูงต้องไม่เกินร้อยละสิบห้าของเงินได้แต่ต้องไม่เกินสามแสนบาทต่อปีในการคำนวณรวมเงินลงทุนขั้นสูงให้รวมเงินลงทุนในทุกๆกองทุนที่ลงทุนในปีนั้นๆ

5) กองทุนไม่จ่ายเงินปันผล

6) ห้ามนำหน่วยลงทุนของกองทุนไปจำหน่ายจ่ายโอนหรือนำไปเป็นประกัน

7) หากขายคืนหน่วยลงทุนก่อนกำหนดที่ผู้ลงทุนจะมีอายุครบห้าสิบห้าปี บริบูรณ์และถือหน่วยลงทุนมาน้อยกว่าห้าปีผู้ลงทุนต้องนำเงินสิทธิประโยชน์ทางภาษีที่ได้รับในช่วงห้าปีปฏิทินล่าสุดไปคืนกรมสรรพากรและนำเงินกำไรส่วนเกินทุนที่เกิดจากการขายคืนหน่วยลงทุนนั้นไปคำนวณรวมเป็นรายได้เพื่อเสียภาษีเงินได้ในปีที่มีการขายคืนหน่วยลงทุนนั้น

#### 1.5.2.3 นโยบายการลงทุนของกองทุนรวมเพื่อการเลี้ยงชีพ

การลงทุนอาจเป็นแบบใดแบบหนึ่งในมาตรฐาน 10 แบบของสำนักงานก.ล.ด. ตามที่ได้กล่าวมาแล้วข้างต้นซึ่งมีความเสี่ยงและผลตอบแทนในระดับที่แตกต่างกันผู้ลงทุนต้องพิจารณาให้ถี่ถ้วนก่อนการลงทุนทั้งนี้ขึ้นอยู่กับอายุของผู้ลงทุนการยอมรับความเสี่ยงและการคาดหวังผลตอบแทน

#### 1.5.3 กองทุนรวมที่ลงทุนในต่างประเทศ (Foreign Investment Fund : FIF)

1.5.3.1 กองทุนรวมที่ลงทุนในต่างประเทศคือกองทุนรวมที่มีวัตถุประสงค์เพื่อนำเงินที่ได้จากการจำหน่ายหน่วยลงทุนในประเทศไปลงทุนในต่างประเทศโดยธนาคารแห่งประเทศไทยได้พิจารณาอนุญาตให้มีการนำเงินไปลงทุนในต่างประเทศได้ในวงเงินจำกัดในแต่ละปีกองทุนรวมที่ลงทุนในต่างประเทศจึงนับเป็นช่องทางเพียงช่องทางเดียวที่ผู้ลงทุนไทยจะสามารถกระจายเงินลงทุนของตนให้กว้างขวางขึ้นและเป็นการลดความเสี่ยงในการลงทุนบริษัทจัดการที่สามารถจัดตั้งกองทุนรวมที่ลงทุนในต่างประเทศต้องมีคุณสมบัติตามที่สำนักงานก.ล.ด. เห็นชอบบริษัทจัดการอาจให้ผู้จัดการกองทุนในต่างประเทศทำหน้าที่เป็นผู้จัดการกองทุนรวมได้เนื่องจากการลงทุนในต่างประเทศต้องอาศัยผู้ที่มีความเชี่ยวชาญในเรื่องดังกล่าวในการจัดการลงทุนนโยบายการลงทุนของกองทุนรวมที่ลงทุนในต่างประเทศอาจเป็นแบบใดแบบหนึ่งในมาตรฐาน 10 แบบของสำนักงานก.ล.ด. ตามที่ได้กล่าวมาแล้วข้างต้นซึ่งมีความเสี่ยงและผลตอบแทนในระดับที่แตกต่างกันผู้ลงทุนต้องพิจารณาให้ถี่ถ้วนก่อนการลงทุนทั้งนี้ขึ้นอยู่กับอายุของผู้ลงทุนการยอมรับความเสี่ยงและการคาดหวังผลตอบแทนกองทุนรวมที่ลงทุนในต่างประเทศต้องลงทุนในหรือมีไว้ซึ่งหลักทรัพย์หรือทรัพย์สินอื่นหรือการขาดอกผลโดยวิธีอื่นตามที่สำนักงานก.ล.ด. กำหนดกล่าวคือ

1) ต้องนำเงินไปลงทุนในต่างประเทศทั้งจำนวนวันแต่เป็นกรณีเงินฝากในประเทศเพื่อสำรองเงินไว้สำหรับการดำเนินงานของกองทุนหรือการลงทุนรักษาสภาพคล่องของกองทุนเป็นต้น

2) ต้องลงทุนในประเทศที่มีหน่วยงานกำกับดูแลด้านหลักทรัพย์และตลาดซื้อขายหลักทรัพย์ที่เป็นสมาชิกสามัญของ International Organization of Securities Commissions (IOSCO) หรือในประเทศที่มีตลาดซื้อขายหลักทรัพย์เป็นสมาชิกของ Federation International des Bourses de Valeurs (FIBV)

3) ผู้ออกหลักทรัพย์หรือตราสารการเงินอื่นใดและผู้รับฝากเงินต้องเป็นบุคคลที่อยู่ภายใต้บังคับกฎหมายของประเทศที่มีหน่วยงานกำกับดูแลด้านหลักทรัพย์และตลาดซื้อขายหลักทรัพย์เป็นสมาชิกสามัญของ IOSCO หรือของประเทศที่มีตลาดซื้อขายหลักทรัพย์เป็นสมาชิกของ FIBV

4) การลงทุนในหรือมีไว้ซึ่งหุ้นหุ้นนั้นต้องมีการซื้อขายใน Organized markets ของประเทศนั้นๆ (ตลาดซื้อขายหลักทรัพย์หรือศูนย์ซื้อขายหลักทรัพย์ที่อยู่ภายใต้การกำกับดูแลของหน่วยงานที่กำกับดูแลด้านหลักทรัพย์และตลาดซื้อขายหลักทรัพย์ที่เป็นสมาชิกสามัญของ IOSCO หรือในตลาดซื้อขายหลักทรัพย์เป็นสมาชิกของ FIBV)

1.5.3.2 บริษัทจัดการต้องยื่นคำขอต่อตลาดหลักทรัพย์แห่งประเทศไทยเพื่อขอให้พิจารณารับหน่วยลงทุนของกองทุนรวมที่ลงทุนในต่างประเทศเป็นหลักทรัพย์จดทะเบียนภายใน 15 วันนับตั้งแต่วันถัดจากวันจดทะเบียนกองทุนหากตลาดหลักทรัพย์ตั้งไม่รับหน่วยลงทุนของกองทุนรวมที่ลงทุนในต่างประเทศบริษัทจัดการต้องดำเนินการเพื่อยกเลิกกองทุน

#### 1.5.4 กองทุนรวมหุ้นระยะยาว (Long Term Equity Fund : LTF)

1.5.4.1 กองทุนรวมหุ้นระยะยาวเป็นกองทุนรวมที่เน้นลงทุนในหุ้นโดยทางการสนับสนุนให้จัดตั้งขึ้นเพื่อเพิ่มสัดส่วนผู้ลงทุนสถาบัน (ซึ่งก็คือกองทุนรวม) ที่จะลงทุนระยะยาวในตลาดหลักทรัพย์แห่งประเทศไทยการเพิ่มผู้ลงทุนสถาบันดังกล่าวจะช่วยให้ตลาดทุนไทยมีเสถียรภาพมากขึ้นทั้งนี้ผู้ที่ลงทุนในกองทุนรวมหุ้นระยะยาวที่เป็นบุคคลธรรมดาจะได้รับสิทธิประโยชน์ทางภาษีเพื่อเป็นแรงจูงใจในการลงทุนกองทุนรวมหุ้นระยะยาวเหมาะสำหรับใคร ? กองทุนรวมหุ้นระยะยาวเหมาะสำหรับคนทุกกลุ่มที่ต้องการลงทุนในหุ้นระยะยาวแต่อาจไม่มีความชำนาญเกี่ยวกับการลงทุนในหุ้นหรือไม่มีความใจลงลงทุนผ่านกองทุนรวมทั้งนี้ผู้ลงทุนจะต้องเข้าใจและยอมรับความเสี่ยงจากการลงทุนและเงื่อนไขเกี่ยวกับระยะเวลาในการลงทุนได้เงื่อนไขการลงทุนเพื่อให้ได้รับสิทธิประโยชน์ทางภาษีได้แก่

- 1) เงินลงทุนมาจากเงินได้พึงประเมินตามมาตรา 40 แห่งประมวลรัษฎากร
- 2) เมื่อผู้ลงทุนซื้อกองทุนรวมหุ้นระยะยาวแล้วต้องถือหน่วยลงทุนไว้ไม่

น้อยกว่า 5 ปีปฏิทิน

3) สามารถลงทุนได้สูงสุด 15% ของเงินได้ในแต่ละปีแต่ทั้งนี้เฉพาะส่วนที่ไม่เกิน 300,000 บาท

4) หากมีการขายคืนหน่วยลงทุนก่อนครบกำหนด 5 ปีปฏิทินถือว่าผิดเงื่อนไขการลงทุนจะต้องคืนเงินภาษีที่ได้รับยกเว้นไปพร้อมเงินเพิ่มในอัตราร้อยละ 1.5 ต่อเดือน โดยนับตั้งแต่เดือนเมษายนของปีที่ผู้ลงทุนยื่นขอยกเว้นภาษีจนถึงเดือนที่มีการยื่นคืนเงินภาษี นอกจากนี้ต้องจ่ายภาษีของกำไรส่วนเกินทุน (capital gain) โดยถูกหักภาษีที่จ่าย 3% ของเงินกำไรที่ได้รับและยังต้องนำกำไรที่ได้รับจากการขายคืนหน่วยลงทุนไปรวมเป็นเงินได้เพื่อเสียภาษีตอนปลายปีอีกด้วย

1.5.4.2 นโยบายการลงทุนของกองทุนรวมหุ้นระยะยาวกองทุนมีนโยบายการลงทุนแบบเดียวคือลงทุนในหุ้นที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยไม่น้อยกว่าร้อยละ 65 ของมูลค่าทรัพย์สินสุทธิของกองทุนโดยอาจเน้นลงทุนในหุ้นกลุ่ม SET 50 หุ้นตามกลุ่มอุตสาหกรรมหรือลงทุนในหุ้นตามที่บริษัทจัดการเห็นควรก็ได้ขึ้นอยู่กับรายละเอียดนโยบายการลงทุนของกองทุนรวมหุ้นระยะยาวแต่ละกอง

1.6 หนังสือชี้ชวนเสนอขายหน่วยลงทุนคือเอกสารสำคัญที่บริษัทจัดการจะต้องจัดทำขึ้นและต้องเผยแพร่ให้ผู้ลงทุนทราบหรือแจกจ่ายให้แก่ผู้ลงทุน (เมื่อถูกร้องขอ) ทุกครั้งที่มีการเสนอขายหน่วยลงทุนหนังสือชี้ชวนเสนอขายหน่วยลงทุนของกองทุนรวมซึ่งบริษัทจัดการจัดทำขึ้นให้มี 2 ส่วนคือ

1.6.1 ส่วนสรุปข้อมูลสำคัญที่ผู้ลงทุนควรทราบ โดยในส่วนสรุปข้อมูลสำคัญที่ผู้ลงทุนควรทราบต้องมีรายการอย่างน้อยดังต่อไปนี้

#### 1.6.1.1 ลักษณะที่สำคัญของกองทุนรวม (key feature)

- 1) ชื่อประเภทและอายุของโครงการ
- 2) จำนวนเงินทุนของโครงการมูลค่าที่ตราไว้จำนวนและราคาของหน่วยลงทุนที่เสนอขาย
- 3) นโยบายและวัตถุประสงค์การลงทุน
- 4) นโยบายการจ่ายเงินปันผล
- 5) วันที่เสนอขายหน่วยลงทุน
- 6) สถานที่ติดต่อซื้อขายหน่วยลงทุน
- 7) ชื่อที่อยู่และหมายเลขโทรศัพท์ของผู้ดูแลผลประโยชน์นายทะเบียนหน่วยลงทุนและผู้สอบบัญชี

8) กำหนดเวลาการขายและรับซื้อคืนหน่วยลงทุนภายหลังการเสนอขายครั้งแรกและระยะเวลาการชำระเงินค่าขายคืนหน่วยลงทุน (เฉพาะกองทุนเปิด)

9) หลักเกณฑ์และกำหนดเวลาการรับซื้อคืนหน่วยลงทุนอัตโนมัติ (ถ้ามี) (เฉพาะกองทุนปิด)

10) ข้อมูลอื่นๆเพิ่มเติมในกรณีที่เป็นกองทุนรวมอสังหาริมทรัพย์และกองทุนรวมมีประกันเป็นต้น

1.6.1.2 คำเตือนและข้อแนะนำเกี่ยวกับการลงทุนในหน่วยลงทุนเช่น "การลงทุนในหน่วยลงทุนมิใช่การฝากเงินและมีความเสี่ยงของการลงทุนผู้ลงทุนอาจได้รับเงินลงทุนคืนมากกว่าหรือน้อยกว่าเงินลงทุนเริ่มแรกก็ได้" (ยกเว้นกรณีกองทุนรวมมีประกัน) "ผู้ลงทุนควรศึกษาข้อมูลในหนังสือชี้ชวนให้เข้าใจก่อนซื้อหน่วยลงทุนและเก็บไว้เป็นข้อมูลเพื่อใช้อ้างอิงในอนาคต หากต้องการทราบข้อมูลเพิ่มเติมสามารถขอหนังสือชี้ชวนส่วนข้อมูลโครงการได้ที่บริษัทจัดการหรือผู้ขายหน่วยลงทุน" เป็นต้น

1.6.1.3 ความเสี่ยงในการลงทุนของกองทุนรวมนั้นให้มีรายละเอียดอย่างน้อยดังนี้

- 1) ความเสี่ยงจากการดำเนินงานของผู้ออกตราสาร (credit risk)
- 2) ความเสี่ยงจากความผันผวนของราคาตราสาร (market risk)
- 3) ความเสี่ยงจากการขาดสภาพคล่องของตราสาร (liquidity risk)
- 4) ความเสี่ยงในความสามารถในการชำระหนี้ในอนาคตของผู้ประกัน

(เฉพาะกองทุนรวมมีประกัน)

5) แนวทางการบริหารเพื่อลดความเสี่ยง (ถ้ามี)

6) การเปรียบเทียบความเสี่ยงและผลตอบแทนกับกองทุนรวมประเภท

อื่น (ถ้ามี)

1.6.1.4 ตารางแสดงค่าธรรมเนียมและค่าใช้จ่ายทั้งหมดที่เรียกเก็บจากผู้ซื้อหรือผู้ถือหน่วยลงทุนรวมหรือกองทุนรวม

1.6.1.5 วันเดือนปีที่รวบรวมข้อมูลไว้ในหนังสือชี้ชวน

1.6.2 ส่วนข้อมูลโครงการ โดยในส่วนข้อมูลโครงการต้องมีรายการตามที่มีปรากฏในส่วนสรุปข้อมูลสำคัญและอื่นๆดังต่อไปนี้

1.6.2.1 ประเภทชื่อจำนวนและมูลค่าตามราคาตลาดหลักทรัพย์หรือทรัพย์สินหรือการหาผลตอบแทนโดยวิธีอื่นที่กองทุนรวมลงทุนไว้

1.6.2.2 ความเห็นของผู้ดูแลผลประโยชน์เกี่ยวกับการจัดการกองทุนรวมของบริษัทจัดการ

1.6.2.3 ผลการดำเนินงานของกองทุนรวมโดยใช้วิธีวัดผลการดำเนินงานมาตรฐานที่สมาคมกำหนดหากใช้วิธีวัดผลการดำเนินงานอื่นใดให้แสดงผลการดำเนินงานที่ใช้วิธีวัดผลการดำเนินงานตามมาตรฐานที่สมาคมกำหนดควบคู่กันไปด้วยเป็นต้นบริษัทจัดการต้องจัดทำหนังสือชี้ชวนของกองทุนเปิดใหม่ในรอบปีบัญชีซึ่งต้องแสดงข้อมูลที่เป็นปัจจุบันณวันสิ้นปีบัญชียุทธศาสตร์นี้ หนังสือชี้ชวนแบ่งออกเป็น 2 ส่วนเช่นเดียวกับกับหนังสือชี้ชวนที่จัดทำขึ้นสำหรับการเสนอขายหน่วยลงทุนครั้งแรกและให้เพิ่มเติมวันที่จดทะเบียนกองทุนรวมในส่วนลักษณะที่สำคัญของกองทุนรวมด้วย

1.7 ตราสารที่ลงทุนแบ่งได้เป็น 2 กลุ่มตราสารที่ลงทุนมีดังนี้

1.7.1 ตราสารทุน (Equity Instruments) หมายถึงตราสารที่บริษัทออกให้แก่ผู้ถือเพื่อแสดงสิทธิของความเป็นเจ้าของในกิจการนั้นประเภทของตราสารทุนได้แก่

1.7.1.1 หุ้นสามัญ (Common Stocks หรือ Ordinary Shares) คือตราสารสิทธิที่แสดงความเป็นเจ้าของกิจการและเมื่อกิจการมีกำไรจากการดำเนินงานผู้ถือหุ้นสามัญจะได้รับเงินปันผลในอัตราที่จัดสรรโดยที่ประชุมใหญ่ผู้ถือหุ้นโดยคำนวณตามสัดส่วนของจำนวนหุ้นที่ถือครองทั้งนี้เงินปันผลอาจมากหรือน้อยขึ้นอยู่กับผลกำไรจากการดำเนินงานประจำปีของกิจการ

1.7.1.2 หุ้นบุริมสิทธิ (Preferred Stocks) คือตราสารสิทธิที่แสดงความเป็นเจ้าของกิจการที่มีการจัดบุริมสิทธิไว้อย่างชัดเจนไม่สามารถยกเลิกได้เมื่อกิจการมีกำไรจากการดำเนินงานผู้ถือหุ้นบุริมสิทธิจะได้รับเงินปันผลในอัตราคงที่ตามที่จัดบุริมสิทธิไว้ว่าจะมากหรือน้อยกว่าผู้ถือหุ้นสามัญก็ได้แต่หากกิจการนั้นต้องเลิกดำเนินการและมีการชำระบัญชีโดยการขายทรัพย์สินผู้ถือหุ้นบุริมสิทธิจะได้รับเงินคืนทุนก่อนผู้ถือหุ้นสามัญ

1.7.1.3 ใบสำคัญแสดงสิทธิในหุ้น (Stock Warrants) คือตราสารสิทธิที่กิจการออกให้แก่ผู้ลงทุนเพื่อให้สิทธิในการซื้อหุ้นออกใหม่ในราคาจำนวนและภายในระยะเวลาที่กำหนดผู้ลงทุนจะมีสิทธิในความเป็นเจ้าของกิจการก็ต่อเมื่อได้ใช้สิทธิในการซื้อหุ้นของกิจการนั้นแล้วเท่านั้น

1.7.1.4 หน่วยลงทุนในกองทุนรวมตราสารทุนคือตราสารสิทธิในการเป็นเจ้าของหน่วยลงทุนของกองทุนรวมที่มีนโยบายลงทุนเน้นการลงทุนในตราสารทุนผู้ลงทุนจะมีสิทธิในความเป็นเจ้าของกิจการที่กองทุนรวมนั้นลงทุนไว้ตามสิทธิที่เฉลี่ยระหว่างผู้ถือหุ้นหน่วยลงทุนทั้งหมดในกองทุนรวมนั้นนั่นเอง

1.7.1.5 ตราสารแสดงสิทธิในอนาคตที่มีหลักทรัพย์อ้างอิงเป็นหุ้น (Stock Options & Futures) คือสัญญาที่ผู้ลงทุนสองฝ่ายตกลงกันเพื่อซื้อหรือขายหุ้นในราคาจำนวนและภายในระยะเวลาที่กำหนด

1.7.2 ตราสารหนี้ (Debt Instruments) หมายถึงตราสารแสดงความเป็นหนี้หรือสัญญาเงินกู้ที่บริษัทออกให้แก่ผู้ลงทุนทั่วไปโดยสัญญาว่าจะใช้เงินตามกำหนดและจ่ายดอกเบี้ยตามกำหนดกล่าวอีกนัยหนึ่งคือเป็นตราสารสิทธิที่แสดงความเป็น "เจ้าหนี้ของกิจการ" โดยทั่วไปแล้วการลงทุนในตราสารหนี้จะมีความเสี่ยงน้อยกว่าการลงทุนในตราสารทุนประเภทของตราสารหนี้ได้แก่

1.7.2.1 ตราสารหนี้ภาครัฐ ได้แก่ พันธบัตรรัฐบาล (government bond) , พันธบัตรรัฐวิสาหกิจ (state-owned enterprise bond) , พันธบัตรธนาคารแห่งประเทศไทย, พันธบัตรกองทุนเพื่อการฟื้นฟูและพัฒนาระบบสถาบันการเงิน, ตั๋วเงินคลัง (treasury bill) เป็นต้น ตราสารหนี้ภาครัฐมีความเสี่ยงต่ำสุดในด้านความสามารถในการชำระหนี้แต่ตราสารหนี้ภาครัฐก็จะมีอัตราผลตอบแทนไม่สูงนักส่วนใหญ่จะมีอายุการลงทุนยาวเพื่อมิให้เป็นภาระของรัฐในด้านการบริหารและการจัดการหนี้ยกเว้นแต่กรณีของตั๋วเงินคลังซึ่งรัฐบาลออกเพื่อใช้ในการกู้ยืมเงินระยะสั้น (ไม่เกิน 180 วัน) หรือเพื่อดูดซับเงินสภาพคล่องส่วนเกินในตลาดเงินเพื่อรักษาระดับอัตราดอกเบี้ยเท่านั้น

1.7.2.2 ตราสารหนี้ภาคเอกชนได้แก่

1) หนี้กู้ (Debenture) มีลักษณะและคุณสมบัติตามสถานะของการเป็นเจ้าหนี้

- หนี้กู้มีประกัน (secured debt) มีการค้ำประกันหนี้โดยบุคคลที่สาม (ส่วนใหญ่ได้แก่บริษัทแม่หรือสถาบันการเงิน) หรือมีการวางหลักทรัพย์ไว้เป็นประกันการชำระหนี้ผู้ลงทุนทรงสิทธิของความเป็นเจ้าหนี้เหนือกว่าเจ้าหนี้รายอื่น

- หนี้กู้ไม่มีประกัน (non-secured debt) ปลอดภัยการค้ำประกันและปลอดภัยหลักทรัพย์ที่วางไว้เป็นประกันการชำระหนี้ผู้ลงทุนทรงสิทธิของความเป็นเจ้าหนี้ด้อยกว่าหนี้กู้มีประกัน

- หนี้กู้ไม่ด้อยสิทธิ (senior debt) ผู้ลงทุนทรงสิทธิของความเป็นเจ้าหนี้เท่าเทียมกับเจ้าหนี้รายอื่นแต่ด้อยกว่าหนี้กู้มีประกัน

- หนี้กู้ด้อยสิทธิ (subordinated debt) ผู้ลงทุนทรงสิทธิของความเป็นเจ้าหนี้เป็นรองเจ้าหนี้รายอื่นที่ไม่ด้อยสิทธิ นั่นคือได้รับชำระหนี้คืนหลังสุด

2) ตั๋วแลกเงิน (Bill of exchange) คือตราสารการเงินระยะสั้นที่บุคคลรายหนึ่งสั่งให้บุคคลอีกรายหนึ่งจ่ายเงินตามจำนวนที่ระบุไว้ในตั๋วแลกเงินนั้นให้แก่บุคคลอีกรายหนึ่งในวันที่กำหนดบนหน้าตั๋วแลกเงินนั้นตั๋วแลกเงินสามารถซื้อขายเปลี่ยนมือได้ในตลาดเงิน (money market) ส่วนใหญ่จะมีธนาคารหรือสถาบันการเงินค้ำประกันหรือรับรองหรือรับอาวัลหรือสลักหลังอย่างไม่มีเงื่อนไข



3) ตัวสัญญาใช้เงิน (Promissory note) คือตราสารการเงินระยะสั้นที่ออกโดยบุคคลรายหนึ่งสัญญากับบุคคลอีกรายหนึ่งว่าจะใช้เงินจำนวนที่ระบุบนหน้าตัวสัญญาใช้เงินพร้อมด้วยดอกเบี้ยให้ในวันที่กำหนดตัวสัญญาใช้เงิน โดยส่วนใหญ่จะแลกเปลี่ยนมือไม่ได้ (non-negotiable) และแสดงข้อความไว้บนหน้าตัวแต่ถ้าไม่มีการแสดงไว้ดังกล่าวและตัวสัญญาใช้เงินนั้นมีธนาคารหรือสถาบันการเงินค้ำประกันหรือรับรองหรือรับอวัลตัวสัญญาใช้เงินนั้นก็สามารถนำมาซื้อขายในตลาดเงินได้

4) บัตรเงินฝากแลกเปลี่ยนมือได้ (Negotiable Certificate of Deposit) คือตราสารแสดงการฝากเงินกับธนาคารสามารถแลกเปลี่ยนมือได้ในตลาดรองหากพิจารณาจากความสามารถในการชำระหนี้ตราสารหนี้ภาคเอกชนจะมีความเสี่ยงมากกว่าตราสารหนี้ของภาครัฐ แต่ตราสารหนี้ภาคเอกชนจะมีอัตราผลตอบแทนสูงกว่าและมีอายุการลงทุนให้เลือกมากทั้งระยะสั้นปานกลางและระยะยาว

#### 1.8 ผลตอบแทนจากการลงทุนใน 2 กลุ่มตราสารที่ลงทุนมีดังนี้

##### 1.8.1 ผลตอบแทนจากการลงทุนในตราสารทุนได้แก่

1.8.1.1 เงินปันผล (Dividend) คือเงินส่วนแบ่งกำไรจากการดำเนินงานรายปีของกิจการพิจารณาจัดสรร โดยที่ประชุมใหญ่ผู้ถือหุ้นที่จัดให้มีขึ้นภายหลังจากการรับรองงบการเงินของกิจการซึ่งส่วนใหญ่จะประมาณเดือนเมษายนของทุกปีกฎหมายกำหนดให้กิจการต้องจัดให้มีประชุมใหญ่ผู้ถือหุ้นปีละหนึ่งครั้งภายในสี่เดือนนับจากวันสิ้นสุดปีบัญชีและไม่เกินรอบ 12 เดือนนับจากวันที่ประชุมครั้งล่าสุด

1.8.1.2 กำไรส่วนเกินทุน (Capital Gain) คือเงินได้ที่เกิดขึ้นจากผลต่างของราคาขายหลักทรัพย์ที่สูงกว่าราคาทุน

##### 1.8.2 ผลตอบแทนการลงทุนในตราสารหนี้ได้แก่

###### 1.8.2.1 ดอกเบี้ยรับ (Interest Received)

ผู้ลงทุนจะได้รับดอกเบี้ยเป็นประจำตามจำนวนเงินที่คำนวณจากอัตราดอกเบี้ยที่ตราไว้ (coupon rate) บนตราสารหนี้และตามกำหนดเวลาที่ระบุไว้ตัวอย่างเช่นพันธบัตรรัฐบาลมีอัตราดอกเบี้ยที่ตราไว้ร้อยละ 7 จ่ายดอกเบี้ยทุกหกเดือนหมายความว่าผู้ลงทุนจะได้รับดอกเบี้ยเป็นจำนวนเงิน 3.50 บาทในเดือนมิถุนายนและธันวาคมของทุกปีเป็นต้นและงวดสุดท้ายจะได้รับดอกเบี้ยพร้อมการชำระคืนเงินต้นเต็มจำนวน

###### 1.8.2.2 ส่วนลดรับ (Discount Earned)

1) ในกรณีของตราสารหนี้ประเภท zero coupon bond ผู้ลงทุนสามารถซื้อตราสารหนี้ในราคาซื้อลดหรือในมูลค่าที่ต่ำกว่าจำนวนเงินหน้าตัว (face value) ที่ระบุไว้ว่าจะใช้

คืนในวันกำหนดชำระตัวอย่างเช่นตราสารหนี้ประเภท Zero coupon bond ที่ไม่มีอัตราดอกเบี้ยที่ตราไว้จำนวนเงินที่จะชำระคืน 1,000 บาทกำหนดชำระคืนในปีที่ 5 นับจากวันลงทุนจำหน่ายให้แก่ผู้ลงทุนในอัตราผลตอบแทนจนถึงวันครบกำหนด (Yield to maturity) ที่ 7% ต่อปีราคาซื้อของตราสารหนี้ประเภท Zero coupon bond จะคำนวณได้เท่ากับ 708.91 บาท

2) ส่วนลรับจะเท่ากับ  $1,000 - 708.91 = 291.09$  บาท

3) ส่วนลรับจึงเท่ากับส่วนต่างระหว่างราคาซื้อและราคารับชำระคืนเมื่อครบกำหนดนั่นเอง

### 1.8.2.3 กำไรส่วนเกินทุน (Capital Gain)

ส่วนใหญ่จะเกิดขึ้นเมื่ออัตราดอกเบี้ยในตลาดเงินลดลงและมีผลให้อัตราผลตอบแทนในการลงทุน (current yield) ที่มีผู้ประสงค์ซื้อจะลดลงด้วยราคาซื้อขายของตราสารหนี้ที่มี coupon rate ที่ตราไว้ในอัตราสูงกว่า current yield จะขยับตัวสูงขึ้นและเป็นที่มาของกำไรส่วนเกินทุน

1.9 ความเสี่ยงในการลงทุนหมายถึงการที่ผู้ลงทุนได้รับผลตอบแทนจริงจากการลงทุนเบี่ยงเบนไปจากผลตอบแทนที่คาดหวังไว้ว่าจะได้รับเมื่อแรกเริ่มลงทุนไม่ว่าผลตอบแทนที่ได้รับจริงจะมากกว่าหรือน้อยกว่าที่คาดหวังถือว่าเป็นความเสี่ยงทั้งสิ้นดังนั้นจึงไม่มีการลงทุนประเภทใดที่ไม่มีความเสี่ยง

### 1.9.1 ความเสี่ยงใน 2 กลุ่มตราสารที่ลงทุนมีดังนี้

#### 1.9.1.1 ความเสี่ยงจากการลงทุนในตราสารทุน

1) ผู้ลงทุนสามารถลดความเสี่ยงในการลงทุนในตราสารทุนได้ในระดับหนึ่งโดยการกระจายการลงทุนในหลักทรัพย์ของหลายอุตสาหกรรมและหลายภาคธุรกิจทั้งยังควรคัดเลือกตราสารที่จะลงทุนอย่างพิถีพิถัน โดยพิจารณาจากปัจจัยทางพื้นฐานของบริษัทผู้ออกตราสารทุนนั้นๆเป็นสำคัญเช่น

- กิจกรรมที่มีฐานะทางการเงินมั่นคงมีผลกำไรในการประกอบการอย่างต่อเนื่องและสม่ำเสมอทุกปีและถ้าเป็นกิจการที่จัดตั้งขึ้นใหม่ก็ต้องมีศักยภาพในการดำเนินธุรกิจสูง

- กิจกรรมที่มีแนวโน้มในการเจริญเติบโตที่ชัดเจนมีส่วนแบ่งตลาดสูงมีความสามารถในการแข่งขันสูง

- กิจกรรมที่มีคณะผู้บริหารที่มีคุณภาพและไม่มีประวัติการปฏิบัติงานที่ค้างพร้อยหรือส่อในทางไม่สุจริต

- กิจกรรมที่มีข้อมูลประกอบธุรกิจที่ชัดเจนโปร่งใสเป็นประโยชน์ในการตัดสินใจของผู้ลงทุน

- กิจกรรมที่มีงบแสดงฐานะทางการเงินที่ครบถ้วนตามมาตรฐานบัญชีที่

มีคุณภาพ

- 2) ความเสี่ยงของตราสารทุนสามารถแยกออกเป็น 3 ประเภทที่สำคัญได้

ดังนี้คือ

- Company Risk หรือ Credit Risk

ความเสี่ยงที่เกี่ยวกับบริษัทนั้นอันเกิดจากสาเหตุทั้งภายในและภายนอกที่ผู้ลงทุนไม่สามารถคาดการณ์หรือรับรู้มาก่อนได้ เช่น การเปลี่ยนแปลงคณะกรรมการหรือผู้บริหารสำคัญซึ่งอาจมีผลให้มีการเปลี่ยนแปลงนโยบายในการดำเนินงานของกิจการ, ความผิดพลาดในการบริหารจนทำให้เกิดความเสียหายและความสูญเสียต่อผู้ถือหุ้น, การแข่งขันในเชิงการตลาดของกลุ่มที่มีผลกระทบต่อปริมาณการขายและรายได้ของกิจการ, ความนิยมของผู้บริโภคที่เปลี่ยนไปและมีผลกระทบต่อปริมาณการขายและรายได้ของกิจการ, ความเสียหายจากภัยพิบัติทำให้ขบวนการผลิตและการจัดการหยุดชะงักและกระทบต่อกระแสรายได้ของกิจการ, การถูกฟ้องร้องดำเนินคดีอันนำมาซึ่งความเสื่อมเสียชื่อเสียงและค่านิยม (goodwill), การล้มละลายความสูญเสียทั้งหมดของผู้ถือหุ้น

- Sector Risk หรือ Industry Risk

ความเสี่ยงที่เกิดจากคุณสมบัติเฉพาะของภาคธุรกิจหรืออุตสาหกรรมนั้นๆเอง เช่น ธุรกิจหรืออุตสาหกรรมที่มีการผลิตหรือเสนอขายสินค้าหรือบริการเฉพาะเจาะจงเพียงไม่กี่ประเภท เช่น บริษัทสายการบิน บริษัททำเหมืองแร่ บริษัทคอมพิวเตอร์และ softwares เป็นต้น ธุรกิจหรืออุตสาหกรรมประเภทนี้ตลอดจนราคาหุ้นของบริษัทเหล่านี้สามารถพุ่งสูงขึ้นและสามารถตกต่ำได้ในชั่วพริบตาหากมีเหตุการณ์อันไม่คาดฝันเกิดขึ้น

- Market Risk

ความเสี่ยงที่เกิดจากการเปลี่ยนแปลงของสภาพการณ์ลงทุนในตลาดหุ้นที่มีผลต่อการเปลี่ยนแปลงของราคาหุ้น โดยที่ปัจจัยทางพื้นฐานของหุ้นนั้นไม่ได้เปลี่ยนแปลงไปเลยกรณีเช่นนี้มักจะเกิดขึ้นในขณะที่ตลาดหุ้นอยู่ในสภาพร้อนแรง (bull market) หรือชบเซา (bear market) อันมีผลมาจากกระแสความรู้สึกโดยรวมของผู้ลงทุน (market sentiment) ในขณะนั้นๆ สภาพเศรษฐกิจการเมืองและการเงินทั้งภายในและภายนอกประเทศจะมีส่วนกำหนด market sentiment ของผู้ลงทุนด้วย

#### 1.9.1.2 ความเสี่ยงจากการลงทุนในตราสารหนี้

- 1) โดยทั่วไปแล้วการลงทุนในตราสารหนี้จะมีความเสี่ยงน้อยกว่าการลงทุนในตราสารทุนแต่ในสถานการณ์ที่นโยบายการเงินของภาครัฐไม่ชัดเจนอัตราดอกเบี้ยขยับตัว

ขึ้นลงมากตลาดตราสารหนี้อาจมีความผันผวนและมีความเสี่ยงสูงการลงทุนในตราสารหนี้ในสถานการณ์ดังกล่าวก็อาจมีความเสี่ยงสูงกว่าการลงทุนในตราสารทุนก็เป็นได้

## 2) ความเสี่ยงของตราสารหนี้มีอยู่ 10 ประเภทด้วยกัน

### - Interest Rate Risk หรือ Market Risk

ความเสี่ยงที่เกิดจากอัตราดอกเบี้ยในตลาดเงินผันผวนตัวอย่างเช่นเมื่ออัตราดอกเบี้ยในตลาดเงินขยับตัวสูงขึ้นหรือมีท่าทีว่าจะขยับตัวสูงขึ้นอัตราดอกเบี้ยที่ตราไว้ (coupon rate) ของตราสารหนี้ที่ออกใหม่ก็จะสูงขึ้นด้วยตราสารหนี้ที่ออกมาก่อนหน้าและมีการซื้อขายในตลาดรองก็จะมี การซื้อขายในระดับราคาที่ลดลงเพื่อค้ำให้อัตราผลตอบแทน (Yield) ขยับสูงขึ้นไปอยู่ในระดับที่ เทียบเคียงกันกับอัตราดอกเบี้ยปัจจุบันฉะนั้นยิ่งตราสารหนี้ที่มีอายุยาวเพียงใดหรือมีอัตราดอกเบี้ยที่ ตราไว้ (coupon rate) ต่ำเพียงใดตราสารหนี้นั้นก็มีความเสี่ยงต่ออัตราดอกเบี้ยที่อาจจะปรับตัว สูงขึ้นในอนาคตมากขึ้นเท่านั้น

### - Credit Risk หรือ Default Risk

ความเสี่ยงที่เกิดจากการที่ผู้ออกตราสารหนี้ไม่สามารถจ่ายชำระดอกเบี้ยหรือชำระคืน เงินต้นได้เต็มตามจำนวนเงินหรือตามเวลาที่กำหนดไว้ในบรรดาตราสารหนี้ทั้งหมดของภาครัฐและ ภาคเอกชนตัวเงินคลังของรัฐบาลจะไม่มี credit risk เลยในขณะที่ตัวเงินหรือหุ้นกู้หรือตราสารหนี้ ประเภทอื่นๆที่ออกโดยภาคเอกชนจะมี credit risk มากบ้างน้อยบ้างในระดับที่แตกต่างกันไป ทั้งนี้ ขึ้นอยู่กับความสามารถในการชำระหนี้เงินต้นและดอกเบี้ยของผู้ออกตราสารหนี้เหล่านั้นๆ

การพิจารณาความสามารถในการชำระหนี้ของผู้ออกตราสารหนี้อาจดูจากผลการจัด อันดับความน่าเชื่อถือ (credit rating) ตราสารหนี้เหล่านั้นๆที่จัดทำโดยบริษัทจัดอันดับความน่าเชื่อถือ (credit rating agency) ซึ่งเป็นองค์กรอิสระไม่มีความสัมพันธ์กับผู้ออกตราสารหนี้ในการจัดอันดับ ความน่าเชื่อถือนั้นอาจมีการจัดอันดับทั้งด้านคุณภาพของผู้ออกตราสารหนี้ (ความสามารถในการ จ่ายชำระหนี้) และด้านคุณภาพของตัวตราสารหนี้เอง (ความเหมาะสมของเงื่อนไขในการชำระหนี้)

### - Purchasing Power Risk หรือ Inflation Risk

ความเสี่ยงต่อการมีอำนาจซื้อที่ลดลงในอนาคตภาวะเงินเฟ้อมีอิทธิพลต่ออำนาจซื้อของ ผู้ลงทุนในขณะที่ภาวะเงินเฟ้อเป็นสิ่งที่คู่กันไปกับการเจริญเติบโตทางเศรษฐกิจผู้ลงทุนในตราสาร หนี้ที่มีอายุยาวเกินกว่า 10 ปีจึงมีความเสี่ยงต่อภาวะเงินเฟ้อและอำนาจซื้อที่ลดลงในอนาคตเพราะ ดอกเบี้ยที่ได้รับจากการลงทุนในตราสารหนี้ที่มีอายุยาวจะมีจำนวนเงินที่ได้รับในแต่ละงวดเท่าเดิม ตลอดอายุของตราสารหนี้เหล่านั้นในขณะที่ภาวะเงินเฟ้อทำให้ราคาสິงของแพงขึ้น

## อัตราดอกเบี้ยรับ – อัตราเงินเฟ้อ = ผลตอบแทนที่แท้จริง ...[1]

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

ถ้าภาวะเงินเฟ้อสูงขึ้นผลตอบแทนที่แท้จริง (real return) ของการลงทุนในตราสารหนี้จะลดลงและทำให้อำนาจซื้อของผู้ลงทุนลดลงด้วย

หากภาวะเงินเฟ้อในปีใดสูงเกินกว่าดอกเบี้ยรับของตราสารหนี้ที่ผลตอบแทนที่แท้จริงที่ผู้ลงทุนจะได้รับอาจถึงขั้นติดลบก็ได้กล่าวได้ว่ารายรับจากดอกเบี้ยของตราสารหนี้ที่ลงทุนไว้ไม่เพียงพอกับรายจ่ายจะนั้นผู้ลงทุนที่นำเงินออมจำนวนมากๆไปลงทุนในตราสารหนี้อายุยาวที่มีดอกเบี้ยในอัตราคงที่ (fixed coupon rate) อาจประสบกับความเสียหายประเภท purchasing power risk เมื่อเงินเฟ้อพุ่งสูงขึ้นรายจ่ายเพิ่มขึ้นตามภาวะเงินเฟ้อแต่มีรายได้จากดอกเบี้ยรับคงที่โดยส่วนใหญ่แล้วผู้ลงทุนมิได้คำนึงถึงความเสียหายประเภทนี้มากนัก

ผู้ลงทุนจึงควรกระจายเงินที่จะลงทุนไปในตราสารหนี้ที่มีอายุ (maturity) หลากหลายเพื่อเป็นการกระจายความเสี่ยงประเภท purchasing power risk ไปในตัว

### - Reinvestment Risk

ความเสี่ยงที่เกิดขึ้นจากการที่ผู้ลงทุนนำเอาดอกเบี้ยรายงวดที่ได้รับจากตราสารหนี้ไปลงทุนต่อในตราสารที่ให้อัตราผลตอบแทนที่ลดลงจากเดิมในขณะที่อัตราดอกเบี้ยในตลาดเงินลดต่ำลงถึงแม้ว่าตราสารหนี้ที่ลงทุนไว้จะมีอัตราดอกเบี้ยที่ตราไว้คงที่ก็ตามแต่กระแสเงินสดที่ผู้ลงทุนได้รับจากดอกเบี้ยเป็นงวดๆจะถูกนำไปลงทุนใหม่อีกครั้งในตราสารอื่นที่มีอัตราดอกเบี้ยลดต่ำลงกล่าวง่ายก็คืออัตราดอกเบี้ยบนดอกเบี้ยรับลดลงนั่นเองจึงทำให้อัตราผลตอบแทนโดยรวมในการลงทุนในตราสารหนี้ของผู้ลงทุนนั้นลดลง

### - Rollover Risk

ความเสี่ยงที่เกิดในกรณีที่ตราสารหนี้ที่ลงทุนไว้เดิมครบกำหนดอายุในขณะที่อัตราดอกเบี้ยในตลาดเงินลดลงและผู้ลงทุนต้องนำเงินคืนที่ได้รับชำระคืนจากตราสารหนี้ที่ลงทุนใหม่ในตราสารหนี้ใหม่ที่มีอัตราดอกเบี้ยต่ำลงในกรณีเช่นนี้อัตราผลตอบแทนจากการลงทุนรอบใหม่จะลดลงเป็นความเสี่ยงที่มีลักษณะเหมือนกับ Reinvestment Risk แต่จะเกิดกับเงินต้นที่อ่อนตัวและอัตราดอกเบี้ยอยู่ในช่วงขาลงตราสารหนี้ระยะสั้นจะมี rollover risk สูงสุดแต่ในช่วงที่ตลาดเงิน

ตั้งตัวและดอกเบี้ยอยู่ในช่วงขาขึ้นตราสารหนี้ระยะสั้นก็จะมี rollover risk ต่ำและสามารถปรับตัวได้เร็วกว่าตราสารหนี้ประเภทอื่น

#### - Call Risk

ความเสี่ยงที่เกิดจากการที่ผู้ออกตราสารหนี้ชำระคืนหนี้ก่อนครบกำหนดความเสี่ยงประเภทนี้จะเกิดขึ้นเฉพาะกับตราสารหนี้ที่มีการระบุเงื่อนไข call option ไว้ล่วงหน้าว่าผู้ออกตราสารหนี้มีสิทธิที่จะจ่ายชำระคืนหนี้ได้ก่อนกำหนดส่วนใหญ่จะเกิดขึ้นเมื่ออัตราดอกเบี้ยลดลงและผู้ออกตราสารหนี้สามารถหาแหล่งเงินทุนอื่นที่มีต้นทุนต่ำกว่ามาจ่ายคืน (refinancing) ผู้ลงทุนในตราสารหนี้ประเภทนี้จะถูกบังคับให้รับคืนเงินและต้องนำเงินที่ได้รับคืนนั้นไปลงทุนใหม่ในตราสารอื่นที่มีอัตราผลตอบแทนที่ต่ำกว่าเดิมที่เคยได้รับ

#### - Prepayment Risk

ความเสี่ยงที่เกิดจากการที่ลูกหนี้จ่ายชำระคืนหนี้ก่อนครบกำหนดอายุเช่นเดียวกับกับ call risk แต่ความเสี่ยงประเภท prepayment risk นี้มักจะเกิดขึ้นกับตราสารหนี้ประเภทที่มีบัญชีลูกหนี้พร้อมหลักทรัพย์ค้ำประกันเป็นประกันการชำระคืนของตราสารหนี้ (mortgage-backed securities) หากลูกหนี้ตามสัญญาจ่ายชำระคืนเงินต้นและดอกเบี้ยก่อนกำหนดและถอนหลักทรัพย์ค้ำประกันตราสารหนี้ก็จะต้องรับคืนเงินก่อนกำหนดและหากอยู่ในระหว่างอัตราดอกเบี้ยขาลงผู้ลงทุนในตราสารหนี้ก็จะเสียเปรียบเช่นเดียวกับกรณีของ reinvestment risk

#### - Currency Risk หรือ Exchange Risk

ความเสี่ยงที่เกิดขึ้นจากอัตราแลกเปลี่ยนเงินตราต่างประเทศเปลี่ยนแปลงจะเกิดขึ้นกับกรณีของการลงทุนข้ามประเทศตัวอย่างเช่น

นาย A นำเงิน 1 ดอลลาร์สหรัฐเข้ามาลงทุนในตลาดทุนของประเทศไทยในช่วงก่อนเงินบาทลอยตัวในปี 2540 ซึ่งมีอัตราแลกเปลี่ยน  $US\$ 1 = 25$  บาท

นาย A ลงทุนด้วยเงิน 25 บาทเป็นระยะเวลา 5 ปีได้กำไรทุกปีในอัตราร้อยละ 10 ต่อปี ดังนั้นนาย A มีเงินลงทุนพร้อมดอกเบี้ยทบต้นมีมูลค่าสะสมรวมเท่ากับ 40.27 บาท

ต่อมานาย A ต้องการที่จะถอนเงินที่ลงทุนไว้และส่งกลับไปยังประเทศของตน

นาย A ต้องนำเงินลงทุนรวมทั้งผลกำไรที่เกิดขึ้น (40.27 บาท) ไปแลกกลับเป็นเงินสกุลดอลลาร์สหรัฐ

ในปัจจุบันอัตราแลกเปลี่ยน  $US\$ 1 = 42$  บาท

นาย A พบว่าผลกำไรที่เกิดขึ้นในระหว่างที่ลงทุน 5 ปีไม่คุ้มกับการขาดทุนค่าของเงิน

สิ่งที่เกิดขึ้นกับนาย A คือ currency risk หรือ exchange risk ที่เกิดขึ้นกับนักลงทุนต่างประเทศในตลาดทุนไทยนอกเหนือไปจากการขาดทุนในการลงทุนทั่วไปฉะนั้นเมื่อสถานการณ์

เศรษฐกิจการเงินและการเมืองภายในประเทศไม่เอื้ออำนวยจะพบว่าไม่มีเงินลงทุนจากต่างประเทศไหลเข้าทั้งที่ลงทุนอยู่แล้วก็เริ่มไหลออกไปยังตลาดทุนอื่นที่มีความเสี่ยงน้อยกว่าด้วย

- Liquidity Risk

ความเสี่ยงที่เกิดขึ้นจากการขาดสภาพคล่องในการซื้อขายเปลี่ยนมือของตราสารหนี้ นั้นๆมักจะเกิดขึ้นกับกรณีที่ผู้ลงทุนลงทุนในตราสารหนี้ที่ไม่เป็นที่นิยมและมีปริมาณซื้อขายในตลาดรองน้อยมากหากผู้ลงทุนประสงค์จะขายตราสารหนี้ขึ้นเพื่อแลกเปลี่ยนเป็นเงินสดในกรณีจำเป็นผู้ลงทุนอาจจะต้องยอมลดราคาขายลงต่ำกว่าราคาตลาดโดยทั่วไปเพื่อตั้งให้อัตราผลตอบแทนสูงขึ้นและดึงดูดความสนใจให้ผู้ซื้อเข้ามาซื้อราคาซื้อขายดังกล่าวไม่ใช่ราคาที่ยุติธรรมต่อผู้ขายอย่างแน่นอนหรืออาจจะถึงขั้นที่เป็นราคาขาดทุนก็เป็นได้ฉะนั้นการลงทุนในตราสารหนี้ที่ไม่มีสภาพคล่องก็เป็นความเสี่ยงอย่างหนึ่งที่สำคัญ

- Event Risk

ความเสี่ยงที่เกิดขึ้นจากการเปลี่ยนแปลงที่มีนัยสำคัญต่อนโยบายในการดำเนินงานของผู้ออกตราสารหนี้เช่นการเปลี่ยนแปลงผู้ถือหุ้นรายใหญ่หลังจากการถือครองงำกิจการและอาจมีคุณพินิจที่ไม่เป็นคุณแก่เจ้าหนี้ตราสารหนี้เดิมเช่นการประกาศเพิ่มอัตราส่วนการจ่ายเงินปันผลต่อกำไร (payout ratio) หรือประกาศเพิ่มทุนครั้งใหญ่ซึ่งหมายถึงกระแสเงินสดจำนวนมากที่จะต้องจ่ายเป็นเงินปันผลประจำปีและมีผลทำให้ฐานทุนของบริษัทลดความเข้มแข็งลงตัวอย่างของตราสารหนี้และความเสี่ยงในตราสารหนี้แต่ละประเภท

ตารางที่ 6.2 ประเภทของตราสารหนี้กับความเสี่ยงที่มี

ประเภทของตราสารหนี้	ความเสี่ยงที่มี
ตัวเงินคลัง	rollover risk ในระหว่างดอกเบี้ยขาลง reinvestment risk ในระหว่างดอกเบี้ยขาลง

พันธบัตร รัฐบาลระยะ ยาว	interest rate risk ในระหว่างดอกเบี้ยขาขึ้น reinvestment risk ในระหว่างดอกเบี้ยขาลง purchasing power risk ในระหว่างดอกเบี้ยขาขึ้น
หุ้นกู้บริษัท	credit risk จะมากน้อยขึ้นอยู่กับฐานะทางการเงิน liquidity risk ถ้าเป็น issue เล็กและมีผู้ลงทุนน้อย interest rate risk ยกเว้นประเภทอัตราดอกเบี้ยลอยตัว (floating rate) event risk ในกรณีถูกครอบงำกิจการหรือถูกควบรวมกิจการ call risk หรือ prepayment risk หากไม่มีข้อห้ามในสัญญาซื้อผูกพัน

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

1.9.2 วิธีลดความเสี่ยงจากการลงทุนที่วิธีหนึ่งก็คือการกระจายการลงทุนอย่างเหมาะสม  
เช่น

1.9.2.1 การกระจายการลงทุนในอุตสาหกรรมหรือธุรกิจหลายประเภท

1.9.2.2 เลือกลงทุนในหลายหลักทรัพย์ (ควรมากกว่า 15 หลักทรัพย์) ที่มีความ  
แตกต่างกัน

1.9.2.3 หลีกเลี่ยงอุตสาหกรรมหรือธุรกิจที่อาจจะมีผลประกอบการต่ำกว่า  
(underperform) อุตสาหกรรมหรือธุรกิจอื่น



1.9.2.4 หลีกเลียงหลักทรัพย์ที่อาจจะมีผลประกอบการหรือสร้างผลตอบแทนต่ำกว่า (underperform) หลักทรัพย์อื่น

1.9.2.5 เน้นการลงทุนระยะยาวควรกำหนดระยะเวลาในการลงทุนที่แน่นอนไม่คืนตระหนกหากมีความผันผวนในระยะสั้น

1.9.2.6 ไม่ลงทุนแบบเก็งกำไรระยะสั้นๆ

1.9.2.7 ไม่ซื้อขายบ่อยเพราะการซื้อขายในแต่ละครั้งมีค่าใช้จ่ายเกิดขึ้นเสมอ

1.10 ภาษีที่เกี่ยวข้องกับการลงทุนในตราสารทุน

ตารางที่ 6.3 ประเภทเงินได้จากการลงทุนในตราสารทุน

กรณีผู้ลงทุนเป็นบุคคลธรรมดา		
ประเภทเงินได้	ผู้ลงทุนไทย	ผู้ลงทุนต่างประเทศ
เงินปันผล	ผู้ลงทุนมีสิทธิที่จะเลือกหักภาษีที่หากอยู่ในประเทศไทย 180 วันขึ้นไปเสียจ่ายในอัตรา 10% หรือนำไปรวมภาษีเช่นเดียวกับผู้มีสัญชาติไทยแต่หากจำนวนเป็นรายได้เพื่อเสียภาษีปลายปีอยู่ในประเทศไทยน้อยกว่า 180 วันถูก	

	และได้รับเครดิตภาษีเงินปันผล	หักภาษีที่จ่ายในอัตรา 10%
กำไรส่วนเกินทุน	ได้รับยกเว้น	ได้รับยกเว้น
กรณีผู้ลงทุนเป็นนิติบุคคล		
ประเภทเงินได้	ผู้ลงทุนไทย	ผู้ลงทุนต่างประเทศ
เงินปันผล	<p>กรณีทั่วไป : ถูกหักภาษีที่จ่าย 10% และให้นำครึ่งหนึ่งของเงินปันผลที่ได้รับมาคำนวณเป็นรายได้เพื่อเสียภาษี สามารถนำภาษีที่ถูกหักมาจ่ายมาเครดิตออกจากภาษีที่ต้องเสียปลายปี</p> <p>กรณีบริษัทจดทะเบียน : ได้รับการยกเว้นหากถือหุ้นไว้ไม่น้อยกว่าสามเดือนก่อนและหลังวันที่ได้รับเงินปันผลนั้น</p> <p>กรณีบริษัทแม่ถือหุ้นในบริษัทลูกเกินกว่า 25% : ได้รับการยกเว้นหากถือหุ้นไว้ไม่น้อยกว่าสามเดือนก่อนและหลังวันที่ได้รับเงินปันผลนั้น</p>	ถูกหักภาษีที่จ่ายในอัตรา 10%
กำไรส่วนเกินทุน	รวมคำนวณเป็นรายได้รายได้หลังหักค่าใช้จ่ายจะเสียภาษีเงินได้นิติบุคคลในอัตรา 30%	ถูกหักภาษีที่จ่ายในอัตรา 15%

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

#### 1.11 ภาษีที่เกี่ยวข้องกับการลงทุนในตราสารหนี้

ตารางที่ 6.4 ประเภทเงินได้จากการลงทุนในตราสารหนี้

กรณีผู้ลงทุนเป็นบุคคลธรรมดา		
ประเภทเงินได้	ผู้ลงทุนไทย	ผู้ลงทุนต่างประเทศ
ดอกเบี้ย	หักภาษีที่จ่ายในอัตรา 15% และมี	หักภาษีที่จ่ายในอัตรา 15% หากจำหน่าย

	สิทธิเลือกที่จะไม่นำไปรวมคำนวณ	ในต่างประเทศจะได้รับการยกเว้น
	เป็นรายได้เพื่อเสียภาษีสิ้นปี	
ส่วนลดรับ	ส่วนลดรับหักภาษีที่จ่ายในอัตรา 15% เฉพาะผู้รับที่เป็นผู้ทรงคนแรก และมีสิทธิเลือกที่จะไม่นำไปรวมคำนวณเป็นรายได้เพื่อเสียภาษีสิ้นปี	หักภาษีที่จ่ายในอัตรา 15% เฉพาะผู้รับที่เป็นผู้ทรงคนแรก
ส่วนลดรับ	หักภาษีที่จ่ายในอัตรา 15% เฉพาะผู้รับที่เป็นผู้ทรงคนแรก	
กำไรส่วนเกินทุน	หักภาษีที่จ่ายในอัตรา 15% และมีสิทธิเลือกที่จะไม่นำไปรวมคำนวณเป็นรายได้เพื่อเสียภาษีสิ้นปี	หักภาษีที่จ่ายในอัตรา 15% ยกเว้นเฉพาะ Zero coupon bond ที่ผู้ทรงคนแรกได้
กำไรส่วนเกินทุน	Zero coupon bond ที่ผู้ทรงคนแรกได้	หักภาษีที่จ่ายในอัตรา 15% แล้ว
กำไรส่วนเกินทุน	ถูกหักภาษีที่จ่ายในอัตรา 15% แล้ว	
กรณีผู้ลงทุนเป็นนิติบุคคล		
ประเภทเงินได้	ผู้ลงทุนไทย	ผู้ลงทุนต่างประเทศ
ดอกเบี้ย	หักภาษีที่จ่ายในอัตรา 1% และต้องนำไปคำนวณเป็นรายได้เพื่อเสียภาษีสิ้นปี	หักภาษีที่จ่ายในอัตรา 15%
ส่วนลดรับ	หักภาษีที่จ่ายในอัตรา 1%	หักภาษีที่จ่ายในอัตรา 15%
กำไรส่วนเกินทุน	ไม่ถูกหักภาษีที่จ่าย	หักภาษีที่จ่ายในอัตรา 15%

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

#### 1.12 ผลตอบแทนและความเสี่ยงจากการลงทุนในกองทุนรวม

กองทุนรวมไม่รับประกันผลตอบแทนที่ผู้ลงทุนจะได้รับ (ยกเว้นกรณีของกองทุนรวมมีประกัน) ผู้ลงทุนอาจจะได้รับผลตอบแทนหรือไม่ก็ได้ หรืออาจจะขาดทุนจากการขายหน่วยลงทุนก็ได้

เมื่อการลงทุนของกองทุนรวมมีกำไรผู้ถือหุ้นหน่วยลงทุนจะได้รับส่วนแบ่งกำไรในรูปแบบของเงินปันผล (Dividend) (ในกรณีที่กองทุนรวมนั้นมีนโยบายที่จะจ่ายเงินปันผล) และมูลค่าทรัพย์สินสุทธิต่อหน่วยที่เพิ่มขึ้น

ผู้ลงทุนก็จะได้รับกำไรส่วนเกินมูลค่าหน่วยลงทุน (Capital Gain) เมื่อผู้ลงทุนนั้นขายคืนหน่วยลงทุน

ส่วนความเสี่ยงของกองทุนรวมแต่ละประเภทนั้นมีความเสี่ยงในทำนองเดียวกันกับตราสารหรือหลักทรัพย์ที่กองทุนรวมนั้นๆ เน้นลงทุนเช่นกองทุนรวมตราสารทุนก็จะมีความเสี่ยงแบบเดียวกันความเสี่ยงของตราสารทุนกองทุนรวมตราสารหนี้ก็จะมีความเสี่ยงแบบเดียวกับตราสารหนี้ เป็นต้น

### 1.13 ภาษีที่เกี่ยวข้องกับการลงทุนในกองทุนรวม

กองทุนรวมมีลักษณะเป็นนิติบุคคลแยกต่างหากจากบริษัทจัดการผลประโยชน์ต่างๆ ที่กองทุนรวมได้รับจะได้รับยกเว้นไม่ต้องเสียภาษีเช่นกองทุนรวมลงทุนในตราสารทุนได้รับเงินปันผลและหรือกำไรส่วนเกินทุนไม่ต้องเสียภาษีกองทุนรวมลงทุนในตราสารหนี้ได้รับดอกเบี้ยส่วนลดรับและหรือกำไรส่วนเกินทุนก็ไม่ต้องเสียภาษีเช่นกัน เป็นต้น ส่วนผู้ถือหน่วยลงทุนในกองทุนรวมนั้นเมื่อได้รับเงินปันผลหรือกำไรส่วนเกินมูลค่าหน่วยลงทุนจะมีภาระภาษีสำหรับบุคคลธรรมดาและนิติบุคคลในลักษณะเช่นเดียวกับภาระภาษีจากการลงทุนในตราสารทุนดังกล่าวไว้แล้วข้างต้นทุกประการเว้นแต่กรณีผู้ลงทุนซึ่งเป็นบุคคลธรรมดาและเลือกที่จะนำเงินปันผลที่ได้รับจากกองทุนรวมไปรวมคำนวณเป็นเงินได้เพื่อเสียภาษีผู้ลงทุนจะไม่สามารถเครดิตภาษีเงินปันผลได้

### 1.14 การคำนวณราคาและการกำหนดมูลค่าทรัพย์สินของกองทุนรวม

กองทุนรวมต้องคำนวณมูลค่าทรัพย์สินสุทธิของหลักทรัพย์และทรัพย์สินที่กองทุนรวมลงทุนไว้ทุกวันทำการการคำนวณมูลค่าทรัพย์สินสุทธิและมูลค่าต่อหน่วยจะกระทำตามขั้นตอนดังนี้

คำนวณมูลค่าทรัพย์สินสุทธิโดยการคำนวณมูลค่าหลักทรัพย์และทรัพย์สินที่กองทุนลงทุนตามหลักการมูลค่ายุติธรรม (mark to market) กล่าวคือต้องคำนวณมูลค่าจากราคาปิดหรือราคาเสนอซื้อครั้งสุดท้าย (ถ้าหลักทรัพย์นั้นไม่มีการซื้อขายในวันนั้น) บวกกับเงินสดและรายได้ค้างรับทั้งหมดที่มีหักด้วยหนี้สิน (ที่ยังไม่ได้ชำระราคา)

[มูลค่าทรัพย์สินสุทธิ (NAV) = มูลค่าทรัพย์สินตามราคาตลาด + รายได้ค้างรับ + เงินสด-หนี้สิน]

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

คำนวณมูลค่าต่อหน่วยโดยการนำมูลค่าทรัพย์สินสุทธิมาหารด้วยจำนวนหน่วยลงทุนที่ออกจำหน่ายแล้วทั้งหมดของกองทุนนั้น

(มูลค่าทรัพย์สินตามราคาตลาด + รายได้ค้างรับ + เงินสด - หนี้สิน) = มูลค่าต่อหน่วย ...[2]

จำนวนหน่วยลงทุน (NAV ต่อหน่วย)

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

กองทุนรวมต้องประกาศมูลค่าทรัพย์สินมูลค่าทรัพย์สินสุทธิและมูลค่าต่อหน่วยให้ผู้ลงทุนทราบทุกวันสุดท้ายของสัปดาห์ในกรณีของกองทุนปิดและทุกวันทำการที่มีการซื้อขายหน่วยลงทุนในกรณีของกองทุนเปิดการประกาศจะกระทำในหน้าหนังสือพิมพ์อย่างน้อยหนึ่งฉบับซึ่งส่วนใหญ่ก็จะเป็นหนังสือพิมพ์ข่าวธุรกิจและที่ทำการของบริษัทตัวแทนจำหน่ายเพื่อเผยแพร่ให้นักลงทุนทราบ

1.15 การวัดผลและการเปรียบเทียบการดำเนินงานของกองทุนรวม การประเมินผลการดำเนินงานของกองทุนรวมโดยการเปรียบเทียบการเปลี่ยนแปลงเป็นอัตราร้อยละ (percentage change) ของมูลค่าทรัพย์สินสุทธิ (ซึ่งก็คือมูลค่าของเงินลงทุนในครั้งแรกของกองทุนรวมนั้นบวก (หรือลบ) ด้วยผลตอบแทน (หรือผลขาดทุน) สะสมที่ได้จากการลงทุนนั้น) กับ

1.15.1 ดัชนีมาตรฐาน (benchmark) ที่มีกลุ่มหลักทรัพย์ที่คล้ายคลึงกับกลุ่มหลักทรัพย์ที่กองทุนรวมนั้นลงทุนดัชนีมาตรฐานจะถูกกำหนดโดยสมาคมบริษัทจัดการลงทุนเช่น

1.15.1.1 กองทุนรวมตราสารทุนเปรียบเทียบผลการดำเนินงานกับดัชนีราคาหุ้นตลาดหลักทรัพย์ (SET Index)

1.15.1.2 กองทุนรวมตราสารหนี้เปรียบเทียบผลการดำเนินงานกับดัชนี ThaiBDC Bond Index และอัตราดอกเบี้ยเงินฝากของธนาคารพาณิชย์เป็นต้น

1.15.2 กองทุนรวมที่มีนโยบายการลงทุนเหมือนกันในการแสดงข้อมูลเปรียบเทียบข้อมูลบริษัทจัดการจะต้องแสดงอัตราส่วนการเปลี่ยนแปลงของผลการดำเนินงานนั้นย้อนหลังจากวันที่คำนวณปัจจุบันจนถึงอดีตเช่นสามเดือนหกเดือนสิบสองเดือนสามปีและตั้งแต่วันแรกที่จัดตั้งกองทุนนั้นอัตราผลตอบแทนการลงทุนเป็นปัจจัยหลักในการวัดผลดำเนินงานของกองทุนหนึ่งเพื่อเทียบกับผลตอบแทนของกองทุนอื่นกองทุนต่างๆจึงถูกจัดกลุ่มตามวัตถุประสงค์และนโยบายการลงทุนที่คล้ายคลึงกันและนำอัตราผลตอบแทนการลงทุนมาเปรียบเทียบกันกองทุนที่มีผลการดำเนินงานดีก็จะสามารถนำมาใช้ในการจัดอันดับเพื่อผลทางการตลาดได้

(มูลค่าทรัพย์สินตามราคาตลาด + รายได้ค้างรับ + เงินสด - หนี้สิน) = มูลค่าต่อหน่วย ...[2]

จำนวนหน่วยลงทุน(NAV ต่อหน่วย)

ที่มา: ตลาดหลักทรัพย์แห่งประเทศไทย (2554)

กองทุนรวมต้องประกาศมูลค่าทรัพย์สินมูลค่าทรัพย์สินสุทธิและมูลค่าต่อหน่วยให้ผู้ลงทุนทราบทุกวันสุดท้ายของสัปดาห์ในกรณีของกองทุนปิดและทุกวันทำการที่มีการซื้อขายหน่วยลงทุนในกรณีของกองทุนเปิดการประกาศจะกระทำในหน้าหนังสือพิมพ์อย่างน้อยหนึ่งฉบับซึ่งส่วนใหญ่ก็จะเป็นหนังสือพิมพ์ข่าวธุรกิจและที่ทำการของบริษัทตัวแทนจำหน่ายเพื่อเผยแพร่ให้นักลงทุนทราบ

1.15 การวัดผลและการเปรียบเทียบการดำเนินงานของกองทุนรวม การประเมินผลการดำเนินงานของกองทุนรวมโดยการเปรียบเทียบการเปลี่ยนแปลงเป็นอัตราร้อยละ (percentage change) ของมูลค่าทรัพย์สินสุทธิ (ซึ่งก็คือมูลค่าของเงินลงทุนในครั้งแรกของกองทุนรวมนั้นบวก (หรือลบ) ด้วยผลตอบแทน (หรือผลขาดทุน) สะสมที่ได้จากการลงทุนนั้น) กับ

1.15.1 ดัชนีมาตรฐาน (benchmark) ที่มีกลุ่มหลักทรัพย์ที่คล้ายคลึงกับกลุ่มหลักทรัพย์ที่กองทุนรวมนั้นลงทุนดัชนีมาตรฐานจะถูกกำหนดโดยสมาคมบริษัทจัดการลงทุนเช่น

1.15.1.1 กองทุนรวมตราสารทุนเปรียบเทียบผลการดำเนินงานกับดัชนีราคาหุ้นตลาดหลักทรัพย์ (SET Index)

1.15.1.2 กองทุนรวมตราสารหนี้เปรียบเทียบผลการดำเนินงานกับดัชนี ThaiBDC Bond Index และอัตราดอกเบี้ยเงินฝากของธนาคารพาณิชย์เป็นต้น

1.15.2 กองทุนรวมที่มีนโยบายการลงทุนเหมือนกันในการแสดงข้อมูลเปรียบเทียบข้อมูลบริษัทจัดการจะต้องแสดงอัตราส่วนการเปลี่ยนแปลงของผลการดำเนินงานนั้นย้อนหลังจากวันที่คำนวณปัจจุบันจนถึงอดีตเช่นสามเดือนหกเดือนสิบสองเดือนสามปีและตั้งแต่วันแรกที่จัดตั้งกองทุนนั้นอัตราผลตอบแทนการลงทุนเป็นปัจจัยหลักในการวัดผลดำเนินงานของกองทุนหนึ่งเพื่อเทียบกับผลตอบแทนของกองทุนอื่นกองทุนต่างๆจึงถูกจัดกลุ่มตามวัตถุประสงค์และนโยบายการลงทุนที่คล้ายคลึงกันและนำอัตราผลตอบแทนการลงทุนมาเปรียบเทียบกันกองทุนที่มีผลการดำเนินงานดีก็จะสามารถนำมาใช้ในการจัดอันดับเพื่อผลทางการตลาดได้

## ประวัติผู้เขียน

ชื่อ-นามสกุล

นายณัฐพัฒน์ นนทชัย

ประวัติการศึกษา

สำเร็จการศึกษาระดับปริญญาตรีสาขาวิชาสังคมวิทยา  
และมานุษยวิทยา มหาวิทยาลัยเกษตรศาสตร์ ปี  
การศึกษา 2551

ตำแหน่งและสถานที่ทำงานปัจจุบัน

- เจ้าของเว็บไซต์ x321.dyndns.org
- เว็บมาสเตอร์อิสระ
- เว็บโปรแกรมเมอร์อิสระ
- เว็บไซต์ไนเนอร์อิสระ
- นักลงทุนอิสระ

ประสบการณ์ ผลงานทางวิชาการ รางวัล  
หรือทุนการศึกษา

- เลขทำงานเป็นผู้ออกแบบเว็บและผู้เขียนโปรแกรม  
ต่าง ๆ บนเว็บของสมาคมส่งเสริมถิ่นฐานคน  
เมื่อเดือนมีนาคม ถึง เดือนพฤษภาคม ในปี พ.ศ.  
2553
- เลขทำงานเป็นผู้ช่วยอาจารย์ในตำแหน่งผู้ดูแลเว็บ  
ผู้ออกแบบเว็บและผู้เขียนโปรแกรมต่าง ๆ บนเว็บ  
ให้เว็บของภาควิชาสังคมวิทยาและมานุษยวิทยา  
คณะสังคมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ เมื่อ  
เดือนกรกฎาคม ในปี พ.ศ. 2548 ถึง เดือน  
พฤษภาคม ในปี พ.ศ. 2553
- เลขทำงานเป็นพิธีกรและวิทยากรให้กับทาง  
โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์  
ศูนย์วิจัยและพัฒนาการศึกษา ในงาน Autistic  
Awareness Day ณ พระราชวังสวนกุหลาบ เมื่อเดือน  
สิงหาคม ในปี พ.ศ. 2547