# การออกแบบสถาปัตยกรรมของการเผยแพร่สารสนเทศด้านหลักทรัพย์ โดยใช้คำบรรยายลักษณะผู้ใช้สำหรับตลาดหลักทรัพย์แห่งประเทศไทย

นายกัน อุตะเดช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2546
ISBN 974-17-5143-5
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

# ARCHITECTURAL DESIGN OF USER-PROFILE BASED SECURITIES INFORMATION DISSEMINATION FOR STOCK EXCHANGE OF THAILAND

Mr.Gan Utadesh

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University

Academic Year 2003
ISBN 974-17-5143-5

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | การออกแบบสถาปัตยกรรมของการเผยแพร่สารสนเทศด้านหลักทรัพย์โดยใช้คำบรรยายลักษณะผู้ใช้สำหรับตลาดหลักทรัพย์แห่งประเทศไทย |
| โดย | นายกัน อุตะเดช |
| สาขาวิชา | วิทยาศาสตร์คอมพิวเตอร์ |
| อาจารย์ที่ปรึกษา | อาจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์ |

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(อาจารย์ ดร.ยรรยง เต็งอำนวย)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา)

..... กรรมการ
(อาจารย์ จารุมาตร ปิ่นทอง)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กัน อุตะเดช : การออกแบบสถาปัตยกรรมของการเผยแพร่สารสนเทศด้านหลักทรัพย์โดยใช้คำบรรยายลักษณะผู้ใช้สำหรับตลาดหลักทรัพย์แห่งประเทศไทย. (ARCHITECTURAL DESIGN OF USER-PROFILE BASED SECURITIES INFORMATION DISSEMINATION FOR STOCK EXCHANGE OF THAILAND) อ. ที่ปรึกษา: อ. ดร.ณัฐวุฒิ หนูไพโรจน์, 74 หน้า. ISBN 974-17-5143-5.

ปัจจุบัน ข้อมูลสารสนเทศทางด้านหลักทรัพย์ได้มีบทบาทเพิ่มมากขึ้นในการตัดสินใจของนักลงทุน และได้กลายเป็นสินค้าที่สำคัญประเภทหนึ่งในธุรกิจพาณิชย์อิเล็กทรอนิกส์ เนื่องจากความซับซ้อนของข้อมูลและความต้องการของนักลงทุนที่มีความหลากหลาย ทำให้การเผยแพร่สารสนเทศด้านหลักทรัพย์ มีความยุ่งยากในการจัดการ การให้บริการสารสนเทศในลักษณะดั้งเดิมยังไม่สามารถตอบสนองความต้องการของผู้รับบริการได้อย่างเต็มที่ ทั้งในด้านความสามารถและประสิทธิภาพ

ในวิทยานิพนธ์นี้ได้เสนอวิธีการเผยแพร่สารสนเทศด้านหลักทรัพย์ โดยการผสมผสานระหว่างการให้บริการโดยกำหนดให้ผู้รับบริการสมัครเป็นสมาชิกของผู้ให้บริการและ การเผยแพร่ข้อมูลตามคำบรรยายลักษณะผู้ใช้ส่วนบุคคล ซึ่งจะทำให้ผู้ให้บริการสามารถบริหารจัดการกับความต้องการของสมาชิกได้อย่างมีประสิทธิภาพ วิทยานิพนธ์นี้ได้ทำการออกแบบขั้นตอนวิธีในการกรองข้อมูลโดยใช้เทคนิคด้านการจัดทำดัชนี และได้ทำการทดลองเชิงประสิทธิภาพโดยใช้ข้อมูลจากตลาดหลักทรัพย์แห่งประเทศไทย ซึ่งเป็นจำลองสภาพข้อมูลเรียลไทม์บนสถาปัตยกรรมที่ออกแบบ จากการทดลองพบว่าระบบที่ออกแบบสามารถทำงานกับสถาพแวดล้อมที่มีอัตราการเกิดสารสนเทศสูง และสามารถรองรับปริมาณสารสนเทศจำนวนมากได้เป็นอย่างดี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา...วิศวกรรรมคอมพิวเตอร์..... ลายมือชื่อนิสิต.....
สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์..... ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา....2546.....

## 4470216221 : MAJOR COMPUTER SCIENCE

KEY WORD: INFORMATION DISSEMINATION / PUSH BASED MODEL / USER-PROFILE BASED DISSEMINATION / PUBLISH/SUBSCRIBE MODEL / CONTENT-BASED PUBLISH/SUBSCRIBE

GAN UTADESH : ARCHITECTURAL DESIGN OF USER-PROFILE BASED SECURITIES INFORMATION DISSEMINATION FOR STOCK EXCHANGE OF THAILAND. THESIS ADVISOR : NATAWUT NUPAIROJ,Ph.D., 74 pp.

ISBN 974-17-5143-5

Nowsaday, securities information plays a critical role in decision making of investors, and becomes important commodities in electronic commerce. Due to content complexity and wide ranges of investors' requirements, securities information dissemination becomes complicated and difficult to manage. Traditional dissemination model is inadequate to fulfill the needs of clients, both capability or performance requirements.

In this thesis, we have studied a new dissemination model, which combines subscription service and user-profile based information dissemination. Our model can manage customers' needs effectively. We have designed filtering algorithm base on indexing technique and conducted experiments by using data from the Stock Exchange of Thailand and real-time data simulation on our architecture. The experimental results indicate that our system can support real-time and high-volume data dissemination effectively.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department…Computer Engineering…. Student's Signature.....
Field of study ...Computer Science….. Advisor's Signature….....
Academic year ....2003….....

# กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณา และความช่วยเหลืออย่างดียิ่ง จากอาจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ รวมทั้งสนับสนุนการทำงานวิจัยของข้าพเจ้ามาด้วยดีตลอด

ขอขอบคุณ อาจารย์ ดร.ยรรยง เต็งอำนวย และ ผู้ช่วยศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา ที่ท่านอาจารย์ทั้งสองได้ให้ข้อคิดต่าง ๆ ที่เป็นประโยชน์ในวิทยานิพนธ์นี้ รวมถึง อาจารย์ จารุมาตร ปิ่นทอง ที่ท่านได้กรุณาเป็นกรรมการสอบวิทยานิพนธ์นี้

ขอขอบคุณเจ้าหน้าที่จากตลาดหลักทรัพย์แห่งประเทศไทยที่กรุณาให้ข้อมูลสำหรับทดลองในวิทยานิพนธ์นี้

ขอขอบคุณคณาจารย์ เพื่อน ๆ พี่ ๆ น้อง ๆ ในห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ และห้องปฏิบัติการวิศวกรรมซอฟต์แวร์ ที่ได้ให้คำแนะนำ และกำลังใจแก่ข้าพเจ้าตลอดเวลาที่ศึกษาในภาควิชาวิศวกรรมคอมพิวเตอร์แห่งนี้

ท้ายที่สุด ข้าพเจ้าใคร่ขอกราบขอบพระคุณบิดา มารดา และทุกคนในครอบครัว ซึ่งสนับสนุนและให้กำลังใจแก่ข้าพเจ้าเสมอมาจนสำเร็จการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฎ
สารบัญภาพ ..... ฎ

บทที่

1. บทนำ ..... 1
    1.1 ความเป็นมาและความสำคัญของปัญหา ..... 1
    1.2 วัตถุประสงค์ของการวิจัย ..... 4
    1.3 ขอบเขตของการวิจัย ..... 4
    1.4 ประโยชน์ที่คาดว่าจะได้รับ ..... 4
    1.5 วิธีดำเนินการวิจัย ..... 4
    1.6 ลำดับขั้นตอนในการเสนอผลการวิจัย ..... 5
2. เอกสารและงานวิจัยที่เกี่ยวข้อง ..... 6
    2.1 แนวคิดและทฤษฎี ..... 6
        2.1.1 ระบบการติดต่อแบบพับลิช/สับสไครบ์ ..... 6
            2.1.1.1 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์ผ่านทางแชนเนล ..... 7
            2.1.1.2 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่อง ..... 8
            2.1.1.3 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ..... 9
        2.1.2 การกรองสารสนเทศ ..... 10
            2.1.2.1 การกรองโดยพิจารณาจากเนื้อหา ..... 11
            2.1.2.2 การกรองโดยใช้ความร่วมมือ ..... 12
        2.1.3 คำบรรยายลักษณะผู้ใช้ ..... 13
            2.1.3.1 รูปแบบคำบรรยายลักษณะผู้ใช้ ..... 13
            2.1.3.2 หมวดข้อมูลของคำบรรยายลักษณะผู้ใช้ ..... 14

# สารบัญ(ต่อ)

หน้า

2.1.4 อัลกอริธึมของการจับคู่สำหรับการติดต่อแบบพลับลิช/สับสไคร์บโดยพิจารณาจากเนื้อหา ..... 16
2.1.4.1 อัลกอริธึมแบบนาอีฝ ..... 16
2.1.4.2 อัลกอริธึมแบบแฟร์เพรดิเคต ..... 17
2.1.4.3 อัลกอริธึมแบบนีออน ..... 19
2.1.4.4 อัลกอริธึมอื่น ๆ ..... 21
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง ..... 21
2.2.1 งานวิจัย “Matching Event in A Content-based Subscription System” ..... 21
2.2.2 งานวิจัย “Efficient Filtering of XML Document for Selective Dissemination of Information” ..... 22
2.2.3 งานวิจัย “Filtering Algorithm and Implementation for Very Fast Publish/Subscribe Systems” ..... 22
2.2.4 งานวิจัย “Predicate Matching and Subscription Matching in Publish/Subscribe Systems” ..... 23
2.2.5 งานวิจัย “Event Matching in Symmetric Subscription Systems” ..... 23
3. วิธีออกแบบการวิจัย ..... 24
3.1 ความต้องการโดยรวมของระบบ ..... 24
3.2 สถาปัตยกรรมของระบบซีรัพ ..... 26
3.2.1 คำบรรยายลักษณะผู้ใช้ ..... 28
3.2.1.1 หมวดข้อมูลส่วนบุคคล ..... 28
3.2.1.2 หมวดข้อมูลในการรวบรวมสารสนเทศ ..... 29
3.2.1.3 หมวดข้อมูลการส่งสารสนเทศไปให้สมาชิก ..... 29
3.2.1.4 สรุปรายละเอียดของโพรไฟล์ ..... 30
3.2.2 รูปแบบโพรไฟล์ ..... 31
3.2.3 ตัวกรองสารสนเทศ ..... 32
3.2.3.1 การจัดทำสกีมาของสารสนเทศที่มาจากตลาดหลักทรัพย์ ..... 33
3.2.3.2 การจัดทำดัชนีของคำบรรยายลักษณะผู้ใช้ ..... 35
3.2.3.3 ขั้นตอนการทำงานของแมทชิ่งมอดูล ..... 38

# สารบัญ(ต่อ)

หน้า

3.2.4 ขั้นตอนการทำงานของตัวเผยแพร่สารสนเทศ ..... 41
3.2.4.1 การเผยแพร่สารสนเทศแบบไม่เรียลไทม์ ..... 41
3.2.4.2 การเผยแพร่สารสนเทศแบบเรียลไทม์ ..... 42
4. รายละเอียดของระบบ ..... 44
4.1 แผนภาพคลาสของการวิจัย ..... 44
4.1.1 ระบบย่อยส่วนติดต่อผู้ใช้ ..... 44
4.1.1.1 แพ็กเกจหลัก ..... 45
4.1.1.2 แพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้ ..... 46
4.1.1.3 แพ็กเกจการจัดการกับสารสนเทศแบบเรียลไทม์ ..... 46
4.1.1.4 แพ็กเกจการจัดการกับสารสนเทศแบบไม่เรียลไทม์ ..... 47
4.1.2 ระบบย่อยตัวป้อนสารสนเทศแบบเรียลไทม์ ..... 48
4.1.3 ระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ ..... 49
4.1.4 ระบบย่อยส่วนติดต่อกับสมาชิก ..... 49
4.2 การจำลองระบบเพื่อใช้ทดสอบการวิจัย ..... 50
4.3 ฐานข้อมูลคำบรรยายลักษณะผู้ใช้ ..... 51
4.4 อัลกอริธึมที่ใช้ในการทดสอบการวิจัย ..... 58
4.4.1 อัลกอริธึมพื้นฐาน ..... 58
4.4.2 อัลกอริธึมของระบบซีรัฟ ..... 58
5. ผลการวิจัย ..... 61
5.1 สภาพแวดล้อมที่ใช้ในการทดสอบการวิจัย ..... 61
5.2 ข้อมูลที่ใช้ทดสอบการวิจัย ..... 61
5.3 การทดสอบการวิจัย ..... 62
5.4 วิธีประเมินผลการทดสอบการวิจัย ..... 62
5.4.1 การประเมินผลประสิทธิภาพของการเผยแพร่สารสนเทศแบบไม่เรียลไทม์ ..... 62
5.4.2 การประเมินผลประสิทธิภาพของการเผยแพร่สารสนเทศแบบเรียลไทม์จำลอง ..... 63
5.5 ผลการทดสอบการวิจัย ..... 64
5.6 ผลการวิเคราะห์ปัจจัย ..... 67
6. สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ ..... 69

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญ(ต่อ)

หน้า

6.1 สรุปผลการวิจัย ..... 69
6.2 ปัญหาและข้อจำกัดที่พบจากการวิจัย ..... 70
6.3 ข้อเสนอแนะในการพัฒนาต่อ ..... 70
รายการอ้างอิง ..... 72
ประวัติผู้เขียนวิทยานิพนธ์ ..... 74

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญตาราง

ตาราง หน้า

ตารางที่ 3.1 รายละเอียดและตัวอย่างคำบรรยายลักษณะผู้ใช้ ..... 30
ตารางที่ 3.2 การกำหนดลักษณะของสารสนเทศภายในรูปแบบโพรไฟล์ ..... 31
ตารางที่ 4.1 รายละเอียดตารางเก็บข้อมูลคำบรรยายลักษณะผู้ใช้..... 52
ตารางที่ 4.2 รายละเอียดตารางเก็บข้อมูลชื่อของสมาชิก..... 52
ตารางที่ 4.3 รายละเอียดตารางเก็บข้อมูลที่อยู่ของสมาชิก ..... 53
ตารางที่ 4.4 รายละเอียดตารางเก็บข้อมูลหมวดข้อมูลการรวบรวมสารสนเทศ..... 54
ตารางที่ 4.5 รายละเอียดตารางเก็บข้อมูลหมวดข้อมูลการส่งสารสนเทศ..... 54
ตารางที่ 4.6 รายละเอียดตารางเก็บข้อมูลประเภทหมวดของสารสนเทศ ..... 55
ตารางที่ 4.7 รายละเอียดตารางเก็บข้อมูลแอททริบิวท์ของสารสนเทศ..... 55
ตารางที่ 4.8 รายละเอียดตารางเก็บข้อมูลรหัสบริษัทและชื่อหลักทรัพย์..... 55
ตารางที่ 4.9 รายละเอียดตารางเก็บข้อมูลความสัมพันธ์ระหว่างหลักทรัพย์และบริษัท ..... 56
ตารางที่ 4.10 รายละเอียดตารางเก็บข้อมูลวิธีการส่งสารสนเทศ ..... 56
ตารางที่ 4.11 รายละเอียดตารางเก็บข้อมูลรูปแบบการส่งสารสนเทศ ..... 56
ตารางที่ 4.12 รายละเอียดตารางเก็บข้อมูลรูปแบบของสารสนเทศ..... 57
ตารางที่ 5.1 รายละเอียดตัวแปรภาระงานและค่าที่เป็นไปได้ ..... 61
ตารางที่ 5.2 ค่าที่เหมาะสมของปัจจัยที่มีผลต่อระบบ..... 68

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญภาพ

ภาพประกอบ หน้า

รูปที่ 1.1 เทคโนโลยีเชิงพูลล์ ..... 2
รูปที่ 1.2 เทคโนโลยีเชิงพุช ..... 3
รูปที่ 2.1 แบบจำลองระบบพับลิช/สับสไครบ์ ..... 7
รูปที่ 2.2 อัลกอริธึมแบบนาอีฝ ..... 16
รูปที่ 2.3 อัลกอริธึมแบบแฟร์เพรดิเคต ..... 18
รูปที่ 2.4 อัลกอริธึมแบบนีออน ..... 20
รูปที่ 3.1 แผนภาพยูสเคสของงานวิจัย ..... 26
รูปที่ 3.2 สถาปัตยกรรมของระบบ ..... 27
รูปที่ 3.3 ขั้นตอนการทำงานของตัวกรองสารสนเทศ ..... 33
รูปที่ 3.4 รายการสกีมาของสารสนเทศ ..... 34
รูปที่ 3.5 ตัวอย่างสารสนเทศที่เข้ามาสู่ระบบ ..... 34
รูปที่ 3.6 ตัวอย่างของสกีมาของสารสนเทศ ..... 34
รูปที่ 3.7 ตัวอย่างบัญชีสมาชิกของสมาชิก ..... 35
รูปที่ 3.8 คลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ ..... 36
รูปที่ 3.9 ตัวอย่างการจัดทำคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ ..... 36
รูปที่ 3.10 คลัสเตอร์ของสมาชิกตามเพรดิเคต ..... 37
รูปที่ 3.11 ตัวอย่างการจัดทำคลัสเตอร์ของสมาชิกตามเพรดิเคต ..... 38
รูปที่ 3.12 รายการเพรดิเคตพึงใจและสมาชิกที่ส่งเพรดิเคตพึงใจ ..... 39
รูปที่ 3.13 ขั้นตอนการทำงานของแมทชิ่งมอดูล ..... 40
รูปที่ 3.14 ขั้นตอนการเผยแพร่สารสนเทศแบบไม่เรียลไทม์ ..... 41
รูปที่ 3.15 ตัวอย่างเอกสารเอ็กซ์เอ็มแอลก่อนส่งให้กับสมาชิก ..... 42
รูปที่ 3.16 ขั้นตอนการเผยแพร่สารสนเทศแบบเรียลไทม์ ..... 43
รูปที่ 4.1 ความสัมพันธ์ระหว่างระบบย่อยต่าง ๆ ของระบบ ..... 44
รูปที่ 4.2 ความสัมพันธ์ระหว่างแพ็กเกจต่าง ๆ ของระบบย่อยส่วนติดต่อผู้ใช้ ..... 45
รูปที่ 4.3 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจหลัก ..... 45
รูปที่ 4.4 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้ ..... 46
รูปที่ 4.5 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการสารสนเทศแบบเรียลไทม์ ..... 47
รูปที่ 4.6 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการสารสนเทศแบบไม่เรียลไทม์ ..... 48

# สารบัญภาพ(ต่อ)

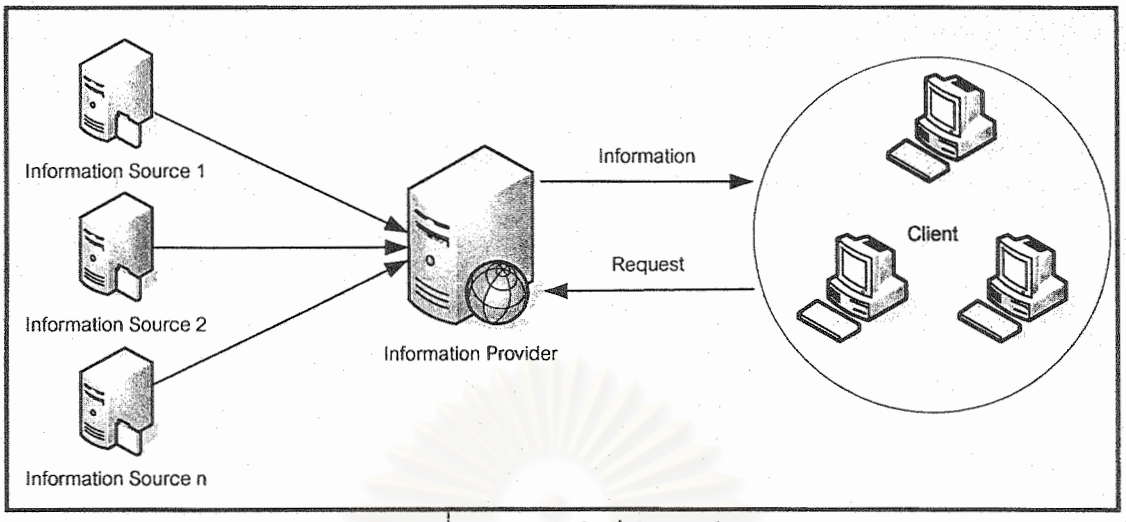
ภาพประกอบ หน้า
รูปที่ 4.7 ความสัมพันธ์ของคลาสในระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ ..... 49
รูปที่ 4.8 ความสัมพันธ์ของคลาสในระบบย่อยส่วนติดต่อกับสมาชิก ..... 50
รูปที่ 4.9 การจำลองการทำงานของระบบ ..... 51
รูปที่ 4.10 ตารางทีพีดีซี ..... 52
รูปที่ 4.11 ตารางทียูเอ็นเอ็ม ..... 52
รูปที่ 4.12 ตารางทียูเอดี ..... 53
รูปที่ 4.13 ตารางทีจีดีซี ..... 54
รูปที่ 4.14 ตารางทีดีดีซี ..... 54
รูปที่ 4.15 ตารางอาร์ซีเอที ..... 55
รูปที่ 4.16 ตารางอาร์เอทีที ..... 55
รูปที่ 4.17 ตารางอาร์ซีโอเอ็มพี ..... 55
รูปที่ 4.18 ตารางอาร์เอสอีซี ..... 56
รูปที่ 4.19 ตารางอาร์ดีอีเอ็ม ..... 56
รูปที่ 4.20 ตารางอาร์ดีทีเอ็ม ..... 56
รูปที่ 4.21 ตารางอาร์แอลเอโอ ..... 57
รูปที่ 4.22 ความสัมพันธ์ของตารางทั้งหมดภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ ..... 57
รูปที่ 4.23 อัลกอริธึมพื้นฐาน ..... 58
รูปที่ 4.24 อัลกอริธึมของระบบซีรัพ ..... 59
รูปที่ 5.1 การจับเวลาของสารสนเทศแบบไม่เรียลไทม์ ..... 62
รูปที่ 5.2 การคำนวณหาค่าเวลาเฉลี่ยที่ใช้ในการแมทชิ่งหนึ่งแอททริบิวท์ ..... 63
รูปที่ 5.3 การจับเวลาของสารสนเทศแบบเรียลไทม์จำลอง ..... 64
รูปที่ 5.4 เวลาเฉลี่ยที่ใช้ในการแมทช์หนึ่งแอททริบิวท์ ..... 64
รูปที่ 5.5 เปรียบเทียบจำนวนเพรดิเคตที่ถูกประเมิน ..... 65
รูปที่ 5.6 เวลาเฉลี่ยที่ใช้ในการแมทชิ่ง เมื่อเพิ่มจำนวนแมสเซจ ..... 66
รูปที่ 5.7 เวลาเฉลี่ยที่ใช้ในการแมทชิ่ง เมื่อเพิ่มจำนวนไคลเอนท์ ..... 67

# บทที่ 1
# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

สารสนเทศด้านหลักทรัพย์ มีความสำคัญต่อการตัดสินใจของนักลงทุน เนื่องจากเป็นสารสนเทศที่เกี่ยวข้องกับตลาดทุนและบริษัทจดทะเบียน ทำให้สารสนเทศด้านหลักทรัพย์ได้กลายเป็นสินค้าประเภทหนึ่งในธุรกิจพาณิชย์อิเล็กทรอนิกส์ และการได้รับสารสนเทศที่ถูกต้องตรงกับความต้องการ สะดวก รวดเร็ว เป็นสิ่งที่สมาชิกที่รับสารสนเทศต้องการ แต่เนื่องจากสารสนเทศด้านหลักทรัพย์มีลักษณะพิเศษ กล่าวคือเป็นข้อมูลที่มีอัตราการเกิดสูง และภายในสิ้นวันจะมีจำนวนสารสนเทศเกิดขึ้นเป็นจำนวนมาก เนื่องจากการเก็บข้อมูลของตลาดหลักทรัพย์จะมีโปรแกรมประยุกต์ที่ถูกออกแบบมาแยกกัน จึงทำให้มีการเก็บข้อมูลตามลักษณะการเกิดของข้อมูลเช่น ข้อมูลที่เกิดจากโปรแกรมประยุกต์ด้านการซื้อขายหลักทรัพย์ ข้อมูลที่เกิดจากโปรแกรมประยุกต์ด้านข้อมูลบริษัทจดทะเบียนเป็นต้น ซึ่งปัญหาปริมาณของข้อมูลและอัตราการเกิดข้อมูลสูงเป็นสาเหตุหนึ่งที่ทำให้เกิดความยากลำบากในการกรองข้อมูลปริมาณมากในแต่ละวัน จึงไม่สามารถที่จะส่งสารสนเทศเฉพาะส่วนที่สมาชิกต้องการไปให้กับสมาชิกได้ และสมาชิกที่ต้องการรับสารสนเทศแต่ละรายมีความต้องการข้อมูลในลักษณะที่แตกต่างกัน ทำให้ผู้เผยแพร่สารสนเทศมีความจำเป็นต้องเพิ่มรูปแบบวิธีการเผยแพร่สารสนเทศ ไปยังสมาชิกที่ต้องการรับสารสนเทศแต่ละราย ที่อาจมีความต้องการข้อมูลชนิดเดียวกันในลักษณะที่แตกต่างกัน

นอกจากนี้สถาปัตยกรรมของระบบเครือข่ายขนาดใหญ่ในปัจจุบัน มีรูปแบบของการติดต่อที่นิยมใช้งานคือ แพลตฟอร์มไคลเอนต์/เซิร์ฟเวอร์แบบซิงโครนัส (Synchronous client/server platform) หรือรูปแบบรีเควสท์/รีพลาย (Request/Reply model) เช่น เวิลด์ไวด์เว็บ คอร์บา (CORBA) เป็นต้น ซึ่งมีข้อดีคือความง่ายในการจัดการ แต่จะมีข้อเสียคือลักษณะของการติดต่อที่เป็นแบบซิงโครนัส (Synchronous) นั่นคือเมื่อไคลเอนต์ส่งคำร้องขอไปยังเซิร์ฟเวอร์แล้ว ไคลเอนต์จะต้องรอการตอบกลับมาจากเซิร์ฟเวอร์ก่อน จึงจะสามารถทำงานต่อได้ และนอกจากนี้ยังมีลักษณะของการผูกติดระหว่างคอมโพเนนท์ที่เกี่ยวข้อง โดยการทำงานของรูปแบบรีเควสท์/รีพลายจะเริ่มจากไคลเอนต์ ที่มีความต้องการสารสนเทศจะส่งคำร้องขอ (Request) สารสนเทศที่ต้องการไปยังเซิร์ฟเวอร์ โดยทางเซิร์ฟเวอร์จะทำการค้นหาสารสนเทศตามคำร้องขอของไคลเอนต์จากแหล่งสารสนเทศที่เซิร์ฟเวอร์รู้จัก และเมื่อได้มาซึ่งสารสนเทศแล้วจะทำการส่งต่อไปให้กับไคลเอนต์ ซึ่งเทคโนโลยีของการติดต่อลักษณะนี้จะเรียกว่า เทคโนโลยีแบบพูลล์ (Pull-Based Technology) ดังแสดงในรูปที่ 1.1

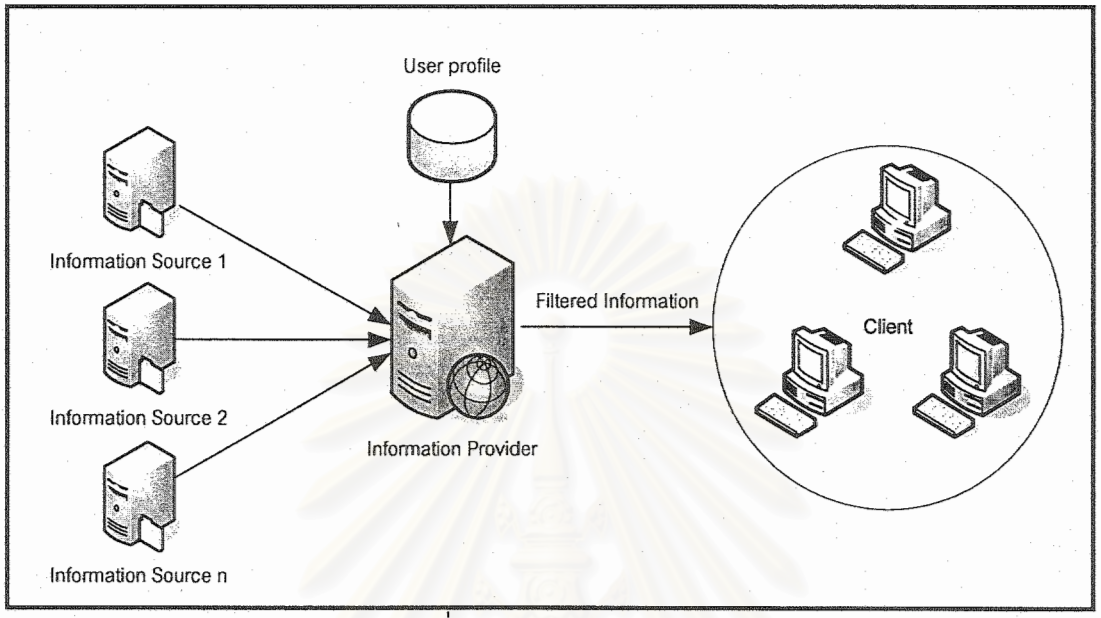



รูปที่ 1.1 เทคโนโลยีเชิงพูลล์

เมื่อพิจารณาลักษณะของสารสนเทศทางด้านหลักทรัพย์ กับเทคโนโลยีแบบพูลล์ พบว่าเทคโนโลยีแบบพูลล์ไม่เหมาะสม กับการเผยแพร่สารสนเทศทางด้านหลักทรัพย์ เนื่องจากไคลเอนต์จะต้องคอยวนส่งคำร้องขอหลักทรัพย์ที่สนใจจากเซิร์ฟเวอร์ที่ให้บริการ ทำให้เกิดการใช้งานทรัพยากรที่ไม่คุ้มค่า นอกจากนี้ในการส่งคำร้องขอมายังเซิร์ฟเวอร์แต่ละครั้ง ไคลเอนต์จะต้องรอการตอบกลับจากเซิร์ฟเวอร์ ในกรณีที่มีไคลเอนต์จำนวนมากส่งคำร้องขอไปยังเซิร์ฟเวอร์อาจทำให้เซิร์ฟเวอร์ไม่สามารถจัดการกับคำร้องขอจากไคลเอนต์แต่ละตัวได้ทัน นอกจากลักษณะของสารสนเทศด้านหลักทรัพย์ที่ไม่เหมาะสมกับเทคโนโลยีแบบพูลล์แล้ว ยังมีสารสนเทศแบบอื่นที่ไม่เหมาะกับการทำงานของเทคโนโลยีแบบพูลล์ โดยมักจะเป็นลักษณะของสารสนเทศที่ไม่สามารถทำนายอัตราการเกิดได้ เช่น สารสนเทศของการประมูลแบบออนไลน์ สารสนเทศของข่าวที่เกิดขึ้นในแต่ละวัน และสารสนเทศในการควบคุมการจราจร เป็นต้น

จากปัญหาความไม่เหมาะสมดังกล่าว ได้มีการเสนอแนวคิดในการกำหนดให้มีการระบุความต้องการ หรือลักษณะของสารสนเทศที่สนใจไว้ที่เซิร์ฟเวอร์ โดยเก็บไว้ในคำบรรยายลักษณะผู้ใช้ (User Profile) และเมื่อมีสารสนเทศใหม่เกิดขึ้นหรือมีการเปลี่ยนแปลงกับสารสนเทศเดิม ระบบจะทำการตรวจสอบสารสนเทศนั้นกับคำบรรยายลักษณะผู้ใช้ว่า เป็นสารสนเทศที่สมาชิกต้องการหรือไม่ ทำให้ไคลเอนต์ไม่ต้องคอยวนส่งคำร้องขอมายังเซิร์ฟเวอร์ โดยจะเรียกการทำงานในลักษณะนี้จะเรียกว่า รูปแบบพับลิช/สับสไครบ์ หรือ เทคโนโลยีแบบพุช (Push-Based Technology) ดังในรูปที่ 1.2 ซึ่งการติดต่อแบบพับลิช/สับสไครบ์จะเป็นการติดต่อแบบอะซิงโครนัส (Asynchronous) ที่แยกการผูกติดกันระหว่างคอมโพเนนท์ที่เกี่ยวข้อง ซึ่งรูปแบบการเผยแพร่ที่ใช้เทคโนโลยีแบบพุชจะเหมาะสมกับการเผยแพร่สารสนเทศด้านหลักทรัพย์ เนื่องจากเซิร์ฟเวอร์เป็นผู้

กำหนดการส่งสารสนเทศที่เกิดขึ้นใหม่ หรือที่เปลี่ยนแปลงไปให้กับไคลเอนต์ โดยที่ไคลเอนต์ไม่ต้องมีการส่งคำร้องขอมายังเซิร์ฟเวอร์ ทำให้ลดปัญหาของการใช้งานทรัพยากรไม่คุ้มค่า และไคลเอนต์ไม่จำเป็นต้องรอการตอบกลับของเซิร์ฟเวอร์




รูปที่ 1.2 เทคโนโลยีเชิงพุช

ในปัจจุบันได้มีการนำเอารูปแบบการติดต่อสื่อสารแบบพับลิช/สับสไครบ์ มาใช้งานในการเผยแพร่สารสนเทศด้านหลักทรัพย์ แต่ไม่ได้มีการจัดทำรูปแบบของสารสนเทศตามความสนใจส่วนบุคคล (Personalization of Information) เนื่องจากความยากลำบากในการกรองข้อมูลปริมาณมากและมีอัตราการเกิดของสารสนเทศสูง จากปัญหาดังกล่าวนี้ทำให้สมาชิกที่รับสารสนเทศต้องรับสารสนเทศไปทั้งหมด ทำให้ได้สารสนเทศที่ไม่ตรงกับความต้องการ เกินความจำเป็น และต้องมีการทำการกรองข้อมูลด้วยตัวเอง ดังนั้นทางผู้วิจัยเห็นว่ามีความจำเป็นที่จะต้องศึกษาถึง แนวทางในการประยุกต์ใช้รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา เพื่อทำการเผยแพร่สารสนเทศด้านหลักทรัพย์ ที่มีการจัดทำรูปแบบของสารสนเทศตามความสนใจของสมาชิกโดยการทำการวิเคราะห์และออกแบบคำบรรยายลักษณะผู้ใช้ที่เหมาะสม ทำการหาขั้นตอนวิธีการกรองสารสนเทศเพื่อให้ตรงกับความต้องการของสมาชิก และปรับปรุงรูปแบบของสารสนเทศตามความต้องการของสมาชิก เพื่อให้ผู้เผยแพร่สารสนเทศสามารถลดภาระและงบประมาณในการจัดการกับสารสนเทศในลักษณะดังกล่าวและทำให้สมาชิกได้รับสารสนเทศในลักษณะที่ตนเองต้องการได้อย่างถูกต้องและครบถ้วน

## 1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและออกแบบสถาปัตยกรรมในการเผยแพร่สารสนเทศด้านหลักทรัพย์ ที่มีอัตราการเกิดของสารสนเทศสูง และมีปริมาณสารสนเทศมากโดยใช้คำบรรยายลักษณะผู้ใช้ในการกรองสารสนเทศที่จะส่งให้กับสมาชิก

## 1.3 ขอบเขตของการวิจัย

1.3.1 ทดสอบการเผยแพร่สารสนเทศโดยวิธีการจำลอง (Simulation)

1.3.2 พารามิเตอร์ที่คงตัว ได้แก่

1.3.2.1 จำนวนหมวดข้อมูลที่จัดเก็บของคำบรรยายลักษณะผู้ใช้

1.3.2.2 วิธีการจับคู่ระหว่างสารสนเทศกับคำบรรยายลักษณะผู้ใช้

1.3.3 พารามิเตอร์ที่ปรับเปลี่ยนค่าได้ ได้แก่

1.3.3.1 รูปแบบของสารสนเทศที่กำหนดในคำบรรยายลักษณะผู้ใช้

1.3.3.2 เนื้อหาของสารสนเทศที่กำหนดในคำบรรยายลักษณะผู้ใช้

1.3.4 การวัดผลจะวัดจาก

1.3.4.1 สมาชิกได้รับสารสนเทศในรูปแบบตรงกับที่กำหนดในคำบรรยายลักษณะผู้ใช้

1.3.4.2 สมาชิกได้รับสารสนเทศภายในเวลาที่กำหนดในคำบรรยายลักษณะผู้ใช้

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 ได้รูปแบบที่เหมาะสมในการทำการเผยแพร่สารสนเทศที่มีความแตกต่างกัน

1.4.2 สามารถจัดหาสารสนเทศที่ตรงกับความต้องการของสมาชิกได้มากขึ้น

## 1.5 วิธีดำเนินการวิจัย

1.5.1 ศึกษาความรู้เกี่ยวกับข้อมูลแต่ละประเภท

1.5.2 ศึกษาความรู้เกี่ยวกับการออกแบบโมเดลของคำบรรยายลักษณะผู้ใช้

1.5.3 ศึกษาความรู้เกี่ยวกับการเผยแพร่ข้อมูลในรูปแบบต่าง ๆ

1.5.4 วิเคราะห์และออกแบบส่วนของการกรองข้อมูล

1.5.5 วิเคราะห์และออกแบบส่วนของการเผยแพร่ข้อมูล

1.5.6 ทดสอบ เผยแพร่ และปรับปรุงงานวิจัย

1.5.7 สรุปผลการวิจัย และจัดทำรายงานวิทยานิพนธ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

## 1.6 ลำดับขั้นตอนในการเสนอผลการวิจัย

ในบทต่อไปของวิทยานิพนธ์นี้จะกล่าวถึงทฤษฎีที่นำมาประยุกต์ใช้ และงานวิจัยที่เกี่ยวข้อง ส่วนในบทที่ 3 จะกล่าวถึงแนวคิดที่ใช้ในการออกแบบงานวิจัยนี้ ในบทที่ 4 จะกล่าวถึงรายละเอียดของงานวิจัยและระบบจำลองที่ได้สร้างขึ้นเพื่อใช้ในการทดสอบ ในบทที่ 5 จะกล่าวถึงผลการทดสอบประสิทธิภาพของสถาปัตยกรรมที่ออกแบบ ผลการวิเคราะห์ข้อมูล และปัจจัยที่เกี่ยวข้อง และในบทสุดท้ายจะเป็นการสรุปผลการวิจัย อภิปรายผล และข้อแนะนำในการพัฒนาต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 2
# เอกสารและงานวิจัยที่เกี่ยวข้อง

## 2.1 แนวคิดและทฤษฎี

### 2.1.1 ระบบการติดต่อแบบพับลิช/สับสไครบ์ (Publish/Subscribe System) [1]

แนวคิดในการโต้ตอบแบบพับลิช/สับสไครบ์ (Publish/Subscribe interaction scheme) กำลังได้รับความสนใจเพิ่มขึ้นเนื่องจากระบบกระจายในปัจจุบันได้มีความเกี่ยวข้องกับเอ็นทิตีที่มีจำนวนมากขึ้น ทำให้เกิดความต้องการรูปแบบของการติดต่อสื่อสารที่มีลักษณะยืดหยุ่นและระบบที่มีลักษณะพลวัต รวมไปถึงลักษณะของแอพลิเกชันที่เป็นแบบไม่ผูกกัน (Decouple) โดยลักษณะแบบไม่ผูกกันระหว่างผู้เผยแพร่สารสนเทศ (Publisher) และสมาชิก (Subscriber) สามารถแบ่งออกได้เป็น 3 ด้าน

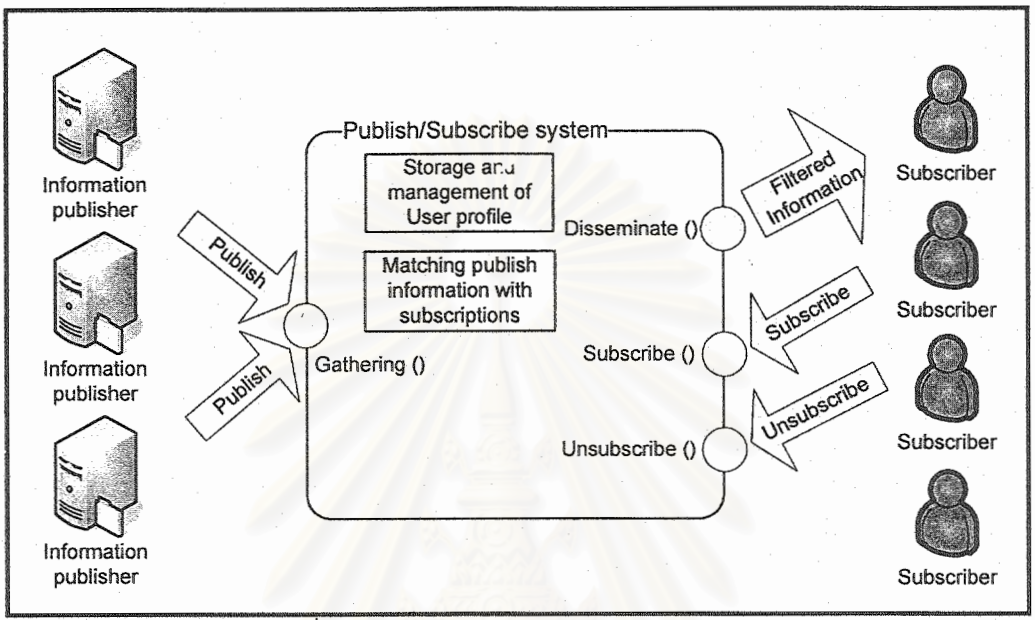
(1) ความไม่ผูกกันด้านพื้นที่ เนื่องจากทั้งผู้เผยแพร่สารสนเทศ และสมาชิกไม่มีการใช้พื้นที่ร่วมกัน

(2) ความไม่ผูกกันด้านเวลา เนื่องจากผู้เผยแพร่สารสนเทศ และสมาชิกไม่จำเป็นต้องเข้ามามีส่วนร่วมในเวลาเดียวกัน

(3) ความไม่ผูกกันด้านกระแสงาน เนื่องจากกระแสงานของทั้งผู้เผยแพร่สารสนเทศ และสมาชิกไม่ได้มีความเกี่ยวข้องกัน

การโต้ตอบแบบพับลิช/สับสไครบ์ เป็นการติดต่อที่ยอมให้สมาชิกระบุถึงความลักษณะของความสนใจในอีเวนต์ (Event) หรือลักษณะของข้อมูลที่ต้องการไว้กับผู้เผยแพร่สารสนเทศ เพื่อที่จะได้รับข้อมูลดังกล่าวเมื่อมีข้อมูลที่ตรงกับที่ระบุไว้ โดยข้อมูลความสนใจของสมาชิกจะถูกเก็บไว้ในรูปแบบของคำบรรยายลักษณะผู้ใช้ (User profile) ซึ่งแบบจำลองระบบพับลิช/สับสไครบ์ดังแสดงในรูปที่ 2.1 จะอาศัยหลักการของการบริการเผยแพร่สารสนเทศ ที่มีการจัดเก็บและจัดการกับบัญชีสมาชิก (Subscription) ที่เก็บอยู่ในคำบรรยายลักษณะผู้ใช้ และมีการแมทชิ่งสารสนเทศที่ถูกเผยแพร่กับบัญชีสมาชิก โดยระบบจะทำหน้าที่เป็นตัวกลางระหว่างผู้เผยแพร่สารสนเทศ (Information publisher) กับสมาชิก ในการทำงานสมาชิกจะทำการระบุความสนใจในสารสนเทศผ่านทางอินเตอร์เฟสที่ชื่อ Subscribe() ของระบบโดยจะไม่รู้จักกับผู้เผยแพร่สารสนเทศและผลของการระบุความสนใจจะถูกเก็บไว้ภายในระบบ ส่วนอินเตอร์เฟสที่ชื่อ Unsubscribe() จะเป็นการยกเลิกการเป็นสมาชิก

เมื่อผู้เผยแพร่สารสนเทศต้องการเผยแพร่สารสนเทศ จะเผยแพร่สารสนเทศให้กับระบบผ่านทางอินเตอร์เฟสที่เรียกว่า Gathering() ซึ่งจะทำให้ระบบทำการแมทชิ่งสารสนเทศกับบัญชีสมาชิกและส่งสารสนเทศที่ผ่านการกรองไปให้กับสมาชิก โดยการส่งผ่านทางอินเตอร์เฟสที่ชื่อ Disseminate()




รูปที่ 2.1 แบบจำลองระบบพับลิช/สับสไครบ์

ซึ่งวิธีการที่แตกต่างกันในการระบุถึงความสนใจในสารสนเทศของสมาชิก จะทำให้เกิดแบบแผนของการติดต่อสื่อสารที่แตกต่างกัน โดยแบบแผนหลักที่สำคัญในปัจจุบันได้แก่ รูปแบบการติดต่อแบบพับลิช/สับสไครบ์ผ่านทางแชนเนล (Channel-based publish/subscribe model) รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่อง หรือ คำสำคัญ (Topic-based publish/subscribe model) และรูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา (Content-based publish/subscribe model)

2.1.1.1 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์ผ่านทางแชนเนล

ระบบการโต้ตอบแบบพับลิช/สับสไครบ์ในสมัยแรก จะเป็นการใช้งานแชนเนลโดยสารสนเทศจะถูกเผยแพร่ตามแชนเนลที่ทางผู้เผยแพร่สารสนเทศระบุไว้ และสารสนเทศที่เผยแพร่เข้าไปในแชนเนลจะถูกส่งไปยังสมาชิกทุกคนที่สมัครเป็นสมาชิกของแชนเนลนี้ แต่การเผยแพร่สารสนเทศในลักษณะนี้จะมีข้อบกพร่องได้แก่

1) ความชัดเจนในการระบุความสนใจในสารสนเทศของสมาชิก หรือความสามารถในการกรองข้อมูลของแชนเนลจะถูกจำกัด เนื่องจากสารสนเทศจะถูกจัดหมวดหมู่ตามจำนวนของแชนเนล ซึ่งหากไม่มีแชนเนลที่ตรงกับความต้องการของสมาชิก อาจทำให้สมาชิกต้องเป็นสมาชิกของแชนเนลมากกว่าหนึ่งแชนเนล และสมาชิกอาจต้องมีการทำการกรองข้อมูลด้วยตนเองอีกครั้งหนึ่ง
2) การใช้งานแชนเนลมีลักษณะที่ไม่ยืดหยุ่น โดยในกรณีที่มีการเปลี่ยนแปลงสารสนเทศที่เผยแพร่เข้าไปในในแชนเนล ทำให้ทั้งผู้เผยแพร่และสมาชิกอาจต้องทำการเปลี่ยนแปลงด้วย โดยผู้เผยแพร่อาจต้องทำการเผยแพร่สารสนเทศนั้นไปในช่องแชนเนลอื่น ในขณะที่สมาชิกอาจต้องมีการเป็นสมาชิกของแชนเนลเพิ่มขึ้น และอาจต้องทำการกรองข้อมูลที่ได้รับมาอีกครั้ง

2.1.1.2 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่อง

สำหรับรูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่อง จะอาศัยหลักการที่ใช้หัวเรื่องเป็นตัวแทนคำสำคัญ (Keyword) เพื่อให้ผู้ที่เกี่ยวข้องทั้งผู้เผยแพร่สารสนเทศและสมาชิกสามารถระบุถึงความสนใจที่ตรงกันได้ โดยเป็นการใช้คำสำคัญในการแบ่งรูปแบบของข้อมูลที่ทำการเผยแพร่ ซึ่งปกติจะเป็นการระบุโดยใช้เครื่องหมายจุดเป็นตัวคั่น ยกตัวอย่างเช่น สมาชิกต้องการรับข้อมูลของบริษัท abc ที่จดทะเบียนอยู่ในตลาดหลักทรัพย์แห่งประเทศไทยอาจทำการสมัครเป็นสมาชิกของหัวเรื่อง market.quotes.SET.abc เป็นต้น และภายหลังได้มีการพัฒนาโดยมีการเพิ่มลำดับชั้นให้กับคำสำคัญ เพื่อให้สามารถแบ่งแยกประเภทและความสัมพันธ์ของคำสำคัญได้ดียิ่งขึ้นโดยจะทำให้สมาชิกสามารถระบุความสนใจได้ตรงกับความต้องการได้เพิ่มมากขึ้น ยกตัวอย่างเช่น มีการจัดทำหมวดหมู่ของหลักทรัพย์ตามกลุ่มอุตสาหกรรม และจัดหมวดหมู่ของหลักทรัพย์โดยประเภทหลักทรัพย์ ดังนั้นถ้าหากสมาชิกต้องการรับข้อมูลของหลักทรัพย์ abc ซึ่งอยู่ภายในกลุ่มอุตสาหกรรมธนาคาร จะต้องสมัครเป็นสมาชิกของหัวเรื่อง market.quotes.SET.Banking.abc หรือต้องการรับข้อมูลของกลุ่มอุตสาหกรรมธนาคารทั้งหมดอาจสมัครเป็นสมาชิกของหัวเรื่อง market.quotes.SET.Banking.* เป็นต้น

จากตัวอย่างข้างต้นจะพบว่า รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่องสามารถ ช่วยให้สมาชิกระบุถึงความต้องการทางด้านสารสนเทศได้ในระดับหนึ่งเท่านั้น กล่าวคือ สมาชิกยังไม่สามารถกำหนดความต้องการในระดับของแอททริบิวท์ที่ตนเองต้องการที่อยู่

นอกเหนือไปจากหัวเรื่องที่ได้กำหนดไว้ และการกำหนดหัวเรื่องในหลายมุมมองจะทำให้เกิดความยากลำบากในการจัดทำลำดับชั้นของหัวเรื่อง นอกจากนี้เมื่อมีการเปลี่ยนแปลงกับลำดับชั้นของหัวเรื่องที่ได้กำหนดไว้ทำให้สมาชิกที่เคยเป็นสมาชิกในหัวเรื่องนั้น อาจต้องมีการปรับปรุงหัวเรื่องที่ได้ลงทะเบียนไว้

2.1.1.3 รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา

จากข้อบกพร่องของ รูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากหัวเรื่องทำให้เกิดรูปแบบการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ซึ่งได้ทำการแก้ไขข้อบกพร่องนี้โดยการนำไปสู่แนวคิดของการเผยแพร่ข้อมูลโดยพิจารณาจากคุณสมบัติของข้อมูลที่ประกอบไปด้วยแอททริบิวต์ต่าง ๆ ดังนั้นสมาชิกสามารถที่จะระบุความสนใจโดยใช้คู่ลำดับของชื่อแอททริบิวต์และค่าของแอททริบิวต์ ของเนื้อหาสารสนเทศ และคู่ลำดับอาจเป็นการประกอบกันของคำเชื่อม (เช่น และ หรือ เป็นต้น) เพื่อสร้างรูปแบบของคำบรรยายลักษณะผู้ใช้ที่ซับซ้อน ซึ่งจะถูกแปลงให้เป็นเงื่อนไขในการกรองข้อมูลเพื่อส่งให้กับสมาชิก จากหลักการนี้ทำให้สามารถลดจำนวนสารสนเทศที่สมาชิกไม่ต้องการลง และสามารถส่งสารสนเทศได้เที่ยงตรงกับความต้องการของสมาชิกได้มากขึ้น

จากความสามารถดังกล่าว ทำให้การติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหาเป็นรูปแบบในการติดต่อสื่อสารที่เหมาะสมกับแอพลิเคชันที่มีอัตราการการเกิดและปริมาณของสารสนเทศสูง โดยเฉพาะในการทำงานบนอุปกรณ์พกพาที่มีข้อจำกัดทางด้านการประมวลผลและแบนด์วิดท์ของเครือข่าย ตัวอย่างเช่นบนเครือข่ายโทรศัพท์มือถือ การใช้งานการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ทำให้สามารถกำหนดระดับของราคาหลักทรัพย์ที่สนใจเมื่อราคาเกินจุดที่กำหนดและสามารถแจ้งเตือนสมาชิกได้ นอกจากนี้รูปแบบการติดต่อลักษณะนี้ทำให้ผู้เผยแพร่สารสนเทศและสมาชิกที่รับสารสนเทศถูกแยกออกจากกันอย่างสมบูรณ์ ซึ่งจะเห็นได้ว่าการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา มีลักษณะที่ยืดหยุ่นกว่าการติดต่อแบบพับลิช/สับสไครบ์แบบอื่น เนื่องจากการเพิ่มจำนวนของผู้เผยแพร่สารสนเทศและสมาชิกสามารถทำได้สะดวกมากขึ้น แต่อย่างไรก็ตามความสามารถในด้านการขยายก็ยังคงเป็นประเด็นที่มีความซับซ้อนที่จะต้องคำนึงถึงในการใช้งาน โดยความชัดเจนของการระบุถึงความสนใจจะเป็นสิ่งที่มีผลกระทบอย่างมากกับความสามารถในด้านการขยายของทุกรูปแบบของการติดต่อแบบพับลิช/สับสไครบ์

2.1.2 การกรองสารสนเทศ (Information Filtering) [2]

การกรองสารสนเทศ เป็นกระบวนงานที่เกี่ยวข้องกับการส่งสารสนเทศที่ตรงกับความต้องการไปยังบุคคลที่ต้องการ ซึ่งจะประกอบไปด้วยคุณสมบัติพื้นฐานดังต่อไปนี้

1) สารสนเทศจะถูกกรองจากกระแสของสารสนเทศซึ่ง การกรองอาจหมายถึงการค้นหา หรือนำสารสนเทศออกจากกระแสสารสนเทศ

2) กระบวนการกรองสารสนเทศ จะขึ้นอยู่กับคำอธิบายลักษณะของสารสนเทศที่สมาชิกต้องการ

3) ลักษณะของสารสนเทศที่จะทำการกรองมีลักษณะเป็น สารสนเทศไม่มีแบบแผน (Unstructured ) หรือ สารสนเทศกึ่งแบบแผน (Semi-structure data) เช่น ไฟล์อักขระ เป็นต้น

ปริมาณและอัตราการเกิดของกระแสสารสนเทศ จะขึ้นอยู่กับแต่ละแอพพลิเกชันโดเมน (Application domain) เช่น บริการกรองข้อมูลข่าว (News filtering service) อาจต้องทำการกรองหัวข้อข่าวจำนวนมาก เพื่อส่งให้กับสมาชิกจำนวนมาก ในขณะที่บริการกรองข้อมูลจดหมายอิเล็กทรอนิกส์ (E-mail filtering service) ต้องการเพียงแค่จำแนกประเภทของจดหมายอิเล็กทรอนิกส์ที่เข้ามาของสมาชิกหนึ่งคน ดังนั้นเมื่อต้องการที่จะสร้างระบบการกรองสารสนเทศ (Information filtering system) หลักที่ใช้ในการตัดสินใจในการออกแบบที่จะต้องพิจารณาคือ

1) วิธีการในการยึดเก็บ (Capture) ความต้องการสารสนเทศของสมาชิก

2) วิธีการในการนำเสนอความต้องการสารสนเทศของสมาชิก

3) วิธีการในการนำเสนอสารสนเทศที่ถูกเผยแพร่

4) หลักการที่ใช้ในการกรองสารสนเทศ กับ ความต้องการสารสนเทศของสมาชิก

สำหรับการยึดเก็บความต้องการสารสนเทศของสมาชิก จะทำในช่วงของการโต้ตอบระหว่างระบบ กับสมาชิก ซึ่งอาจมีลักษณะที่ชัดเจนโดยการใช้ส่วนต่อประสานกราฟิกกับผู้ใช้ (Graphical user interface) ในการเก็บข้อมูล หรืออาจใช้วิธีการสังเกตจากพฤติกรรมของสมาชิก

ซึ่งเทคนิคในการยึดเก็บความต้องการสารสนเทศของสมาชิกจะเรียกว่า ฟังก์ชันป้อนกลับ (Feedback function) และผลลัพธ์ของฟังก์ชันป้อนกลับอาจจะเก็บอยู่ในรูปแบบของคำบรรยายลักษณะผู้ใช้ (User profile) ซึ่งถือเป็นการแสดงรูปแบบภายใน (Internal representation) ของความต้องการสารสนเทศของสมาชิก และเป็นปัจจัยพื้นฐานของการกรองสารสนเทศ สำหรับทางด้านสารสนเทศที่เผยแพร่จะต้องมีการจัดทำการแสดงรูปแบบภายในเช่นกัน โดยจะเรียกการแสดงรูปแบบภายในของสารสนเทศว่า สกีมาของสารสนเทศ (Information schema)

ฟังก์ชันการกรอง เป็นส่วนงานของการจับคู่ของระบบกรองสารสนเทศ โดยอาศัยคำบรรยายลักษณะผู้ใช้ และ สกีมาของสารสนเทศ ซึ่งผลลัพธ์ที่ได้ควรจะเป็นสารสนเทศที่มีความเกี่ยวข้องกับสมาชิก หรือสารสนเทศที่ตรงกับความต้องการของสมาชิก

ในปัจจุบันการกรองสารสนเทศประกอบไปด้วยแนวคิด 2 แบบ ได้แก่

1) การกรองโดยพิจารณาจากเนื้อหา (Content based filtering) คือการที่สารสนเทศถูกกรองโดยพิจารณาจากลักษณะของสารสนเทศ

2) การกรองโดยใช้ความร่วมมือ (Collaborative filtering) คือการที่สารสนเทศถูกกรองโดยพิจารณาจากความคิดเห็นของสมาชิก และความเห็นของบุคคลอื่น

2.1.2.1 การกรองโดยพิจารณาจากเนื้อหา

ระบบกรองโดยพิจารณาจากเนื้อหา จะทำการกรองสารสนเทศโดยการพิจารณาจากลักษณะของสารสนเทศ จากการอธิบายลักษณะของสารสนเทศด้วยเนื้อหา ซึ่งการจัดทำสกีมาของสารสนเทศ อาจเป็นการนำเอาลักษณะบางส่วนของสารสนเทศที่เผยแพร่มาจัดทำเป็นสกีมาของสารสนเทศ ซึ่งรูปแบบที่มีลักษณะเด่นคือการใช้ลักษณะทางสถิติ เช่นสารสนเทศที่มีลักษณะเป็นแฟ้มข้อความ สามารถจัดทำสกีมาของสารสนเทศโดยใช้เวกเตอร์แสดงความถี่ของคำ ที่ปรากฏอยู่ในแฟ้มข้อความนั้น หรือในกรณีที่สารสนเทศมีลักษณะกึ่งแบบแผน อาจจัดทำสกีมาของสารสนเทศได้โดยการนำเอาแอททริบิวท์ของสารสนเทศ (Information attribute) มาเป็นสกีมาของสารสนเทศเพื่อใช้ในการกรองข้อมูลร่วมกับคำบรรยายลักษณะผู้ใช้ได้

ส่วนการจัดทำคำบรรยายลักษณะผู้ใช้นั้นจะขึ้นอยู่กับฟังก์ชันป้อนกลับ ซึ่งวิธีการที่เป็นที่นิยมจะมีอยู่สองวิธี โดยวิธีแรกเป็นแนวคิดที่แสดงถึงลักษณะสกีมาของสารสนเทศที่สมาชิกมี

ความสนใจ เช่นการเก็บเซตของกลุ่มคำสำคัญที่มีการถ่วงน้ำหนัก หรือการจัดเก็บแอททริบิวท์ที่อยู่ในเนื้อหาของสารสนเทศที่ถูกเผยแพร่ออกมา ส่วนวิธีที่สองเป็นการเก็บรายการของสกีมาของสารสนเทศ โดยรายการสกีมาของสารสนเทศแต่ละรายการจะมีตัวระบุว่าสกีมาของสารสนเทศนั้นมีความสัมพันธ์กับสมาชิกหรือไม่ ซึ่งสำหรับวิธีการแรกฟังก์ชันป้อนกลับจะได้ผลลัพธ์เป็นเซตของคำสำคัญ หรือเพรดิเคตที่สมาชิกระบุในการโต้ตอบกับระบบ ส่วนวิธีที่สองฟังก์ชันป้อนกลับจะได้ผลลัพธ์เป็นคะแนนที่สัมพันธ์กับเซตของสกีมาของสารสนเทศ ซึ่งสามารถใช้สร้างโมเดลของความสนใจของสมาชิกได้

สำหรับฟังก์ชันการกรองสำหรับการกรองสารสนเทศโดยพิจารณาจากเนื้อหา จะเป็นการพิจารณาความสัมพันธ์ระหว่าง สกีมาของสารสนเทศและคำบรรยายลักษณะผู้ใช้ เพื่อที่จะระบุว่าสารสนเทศนั้นตรงกับความต้องการของสมาชิกหรือไม่ ตัวอย่างเช่นถ้าหากคำบรรยายลักษณะผู้ใช้ประกอบไปด้วยเซตของกลุ่มคำสำคัญที่มีการถ่วงน้ำหนัก สามารถใช้เทคนิคทางด้านการค้นคืนสารสนเทศ (Information retrieval) ในการคำนวณหาความคล้ายคลึงของคะแนนระหว่างคำบรรยายลักษณะผู้ใช้กับสกีมาของสารสนเทศ ซึ่งถ้าหากคะแนนที่ได้มีค่าสูงเกินขีดแบ่ง (Threshold) ที่กำหนดแสดงว่าสารสนเทศนั้นมีแนวโน้มที่จะตรงกับความต้องการของสมาชิก หรือในกรณีที่คำบรรยายลักษณะผู้ใช้ประกอบไปด้วยเพรดิเคตของแอททริบิวท์ และค่าที่สอดคล้องกับแอททริบิวท์นั้น อาจใช้ตัวดำเนินการเปรียบเทียบกระทำระหว่างสารสนเทศที่กำลังพิจารณา กับเพรดิเคตที่อยู่ในคำบรรยายลักษณะผู้ใช้ เพื่อพิจารณาว่าสารสนเทศนั้นตรงกับความต้องการของสมาชิกหรือไม่

2.1.2.2 การกรองโดยใช้ความร่วมมือ

เป็นการพิจารณาว่าสารสนเทศตรงกับความต้องการของสมาชิกหรือไม่ โดยอาศัยความเห็นจากสมาชิกหรือค่าความนิยม (Rating) ซึ่งไม่มีการคำนึงถึงลักษณะของสารสนเทศ เมื่อเปรียบเทียบกับการกรองโดยพิจารณาจากเนื้อหาแล้ว การกรองโดยใช้ความร่วมมือจะอยู่บนข้อสันนิษฐานที่ว่าสมาชิกที่มีความเห็นเดียวกันในสารสนเทศหนึ่งในอดีต จะมีความเห็นเดียวกันสำหรับสารสนเทศนั้นในอนาคต ด้วยเหตุนี้จึงไม่จำเป็นต้องคำนึงถึงลักษณะของสารสนเทศ ทำให้ในการกรองข้อมูลโดยใช้ความร่วมมือ ไม่จำเป็นต้องจัดทำสกีมาของสารสนเทศ

ส่วนคำบรรยายลักษณะผู้ใช้สำหรับการกรองโดยใช้ความร่วมมือ อาจมีลักษณะที่ชัดเจนโดยเป็นเซตของค่าความนิยมในสารสนเทศนั้น หรืออาจเป็นการเก็บค่าการกระทำของสมาชิกต่อสารสนเทศนั้นเช่นการเก็บค่าว่าสมาชิกมีการเข้ามาดูสารสนเทศนี้หรือไม่

สำหรับฟังก์ชันการกรองสำหรับการกรองโดยใช้ความร่วมมือมีสองแบบ โดยแบบแรกเป็นการนำคำบรรยายลักษณะผู้ใช้ของสมาชิกไปแมทช์กับ คำบรรยายลักษณะผู้ใช้ของสมาชิกคนอื่นเพื่อทำการระบุถึงคำบรรยายลักษณะผู้ใช้ที่คล้ายคลึงกัน และสมาชิกจะทำการคำนวณเพื่อทำนายค่าความนิยมสำหรับสารสนเทศโดยการรวมค่านิยมของกลุ่มสมาชิกที่อยู่ใกล้ชิดกัน และถ่วงน้ำหนักด้วยค่าความคล้ายคลึงของสมาชิก ส่วนแบบที่สองเป็นการสร้างโมเดลทางด้านแมชชีนเลินนิ่ง (Machine learning) จากเซตของคำบรรยายลักษณะผู้ใช้ทั้งหมด ซึ่งโมเดลเหล่านี้จะถูกใช้ในการทำนายวิธีการที่สมาชิกจะให้ค่านิยมกับสารสนเทศใหม่

2.1.3 คำบรรยายลักษณะผู้ใช้ [3]

2.1.3.1 รูปแบบคำบรรยายลักษณะผู้ใช้ (User profile modeling)

ในการส่งสารสนเทศที่สอดคล้องกับความต้องการของสมาชิก สามารถแบ่งปัจจัยหลักที่เกี่ยวข้องได้ 2 ส่วนคือ ความต้องการของสมาชิก และ แหล่งที่มาของสารสนเทศ

ความต้องการของสมาชิก หมายถึง อะไรก็ตามที่สมาชิกกำลังค้นหา ซึ่งอาจมีความแตกต่างกันทั้งในด้านประเภท และเนื้อหาของสารสนเทศ นอกจากนี้ความต้องการของสมาชิกยังสามารถแบ่งแยกออกได้เป็น ความต้องการระยะสั้น ซึ่งอาจเรียกได้ว่าเป็นแบบเฉพาะกิจ และความต้องการระยะยาว ซึ่งเป็นความต้องการสารสนเทศซึ่งเป็นที่สนใจในช่วงเวลาหนึ่ง อาจกล่าวได้โดยสรุปว่าความต้องการสารสนเทศจะแตกต่างกันตามประเภท เนื้อหาและระยะเวลา (ระยะสั้นและระยะยาว) และยังแตกต่างกันตามความต้องการของสมาชิกอีกด้วย สิ่งเหล่านี้ล้วนต้องนำมาพิจารณาในระหว่างกระบวนการสร้างแบบจำลองของคำบรรยายลักษณะผู้ใช้

แหล่งที่มาของสารสนเทศ หรือต้นทางสารสนเทศ ได้แก่ ผู้เผยแพร่สารสนเทศเชิงดิจิทัล โดยเป็นผู้เผยแพร่สารสนเทศที่เป็นที่สนใจของสมาชิก ที่มีความแตกต่างกันในลักษณะของสารสนเทศที่จัดหาให้ บริการที่จัดหาให้ และกลุ่มของสมาชิกที่ให้บริการ ซึ่งเป้าหมายหลักของผู้เผยแพร่สารสนเทศคือการจัดหาสารสนเทศที่ถูกต้องให้กับสมาชิกอย่างตรงเวลา และด้วยวิธีการที่

เหมาะสม ในการทำงานสามารถแบ่งการทำงานออกเป็น 2 ขั้นตอนคือ การรวบรวมสารสนเทศที่เกี่ยวข้องกับความต้องการ และ การส่งออกไปให้กับสมาชิก

2.1.3.2 หมวดข้อมูลของคำบรรยายลักษณะผู้ใช้

จากที่กล่าวมาข้างต้นคำบรรยายลักษณะผู้ใช้ จะต้องบอกข้อมูลเกี่ยวกับสมาชิกอย่างน้อย 2 เรื่องได้แก่ สารสนเทศที่สมาชิกต้องการ และวิธีการในการส่งไปให้กับสมาชิก

ซึ่งสามารถจำแนกได้ 5 หมวดหมู่ ได้แก่

- หมวดข้อมูลส่วนบุคคล
- หมวดข้อมูลเกี่ยวกับการรวบรวมข้อมูล
- หมวดข้อมูลเกี่ยวกับการส่งข้อมูล
- หมวดข้อมูลเกี่ยวกับการปฏิบัติการของผู้ใช้
- หมวดข้อมูลด้านความปลอดภัย

1) หมวดข้อมูลส่วนบุคคล

เป็นการจัดเก็บข้อมูลส่วนตัวของสมาชิก เช่น ชื่อบัญชีสมาชิก รหัสผ่าน ชื่อ-สกุลของสมาชิก และรายละเอียดส่วนตัวอื่น ๆ

2) หมวดข้อมูลเกี่ยวกับการรวบรวมข้อมูล

เป็นการรวบรวมข้อมูลเกี่ยวกับค่าตั้งพึงใจของผู้ใช้ระบบ (User preference) และข้อจำกัดเกี่ยวกับเอกสารที่สมาชิกต้องการ ซึ่งสามารถแบ่งได้เป็น 3 หมวดย่อย ได้แก่

2.1) หมวดย่อยเกี่ยวกับเนื้อหาของเอกสาร เป็นการระบุถึงลักษณะของเอกสารที่จะทำการส่งไปให้กับสมาชิก โดยพิจารณาจากค่าตั้งพึงใจของลักษณะของเอกสารที่สัมพันธ์กับเนื้อหาสาระที่สมาชิกต้องการ

2.2) หมวดย่อยเกี่ยวกับโครงสร้างของเอกสาร เป็นการระบุถึงลักษณะของเอกสารที่สนใจในเชิงของโครงสร้างของเอกสาร เช่นรูปแบบของเอกสารและประเภทของเอกสาร เป็นต้น

2.3) หมวดย่อยเกี่ยวกับแหล่งที่มาของเอกสาร เป็นการระบุถึงแหล่งข้อมูลที่ใช้ในการรวบรวมเอกสาร

3) หมวดข้อมูลเกี่ยวกับการส่งข้อมูล

เป็นการระบุถึงค่าตั้งพึงใจเกี่ยวกับวิธีการส่งข้อมูลซึ่งสามารถแบ่งออกได้เป็น 2 หมวดย่อยซึ่งได้แก่

3.1) หมวดย่อยเกี่ยวกับวิธีการส่งข้อมูล เป็นการระบุถึงวิธีการส่งข้อมูลที่จะใช้ในการส่งสารสนเทศที่สมาชิกต้องการ

3.2) หมวดย่อยเกี่ยวกับเวลาในการส่งข้อมูล เป็นการระบุถึงช่วงเวลาในการส่งข้อมูล ทั้งในกรณีที่เป็นคาบของเวลาและไม่เป็นคาบของเวลา

4) หมวดข้อมูลเกี่ยวกับการปฏิบัติการของผู้ใช้

การบริการส่วนบุคคลควรจะตอบสนองต่อความต้องการของสมาชิกอย่างสูง โดยเฉพาะทางด้านความต้องการสารสนเทศในระยะยาว ซึ่งเกี่ยวพันการปฏิสัมพันธ์อย่างซ้ำไปซ้ำมากับสมาชิก เนื่องจากไม่สามารถสันนิษฐานได้ว่าความต้องการของสมาชิกจะคงที่ โดยการเปลี่ยนแปลงของความต้องการของสมาชิก สามารถเป็นการเปลี่ยนแปลงอย่างเล็กน้อยที่สัมพันธ์กับลำดับความสำคัญของสารสนเทศที่สนใจ ไปจนถึงการเลิกสนใจในบางโดเมนและไปสนใจในโดเมนอื่นแทน โดยทั่วไปแล้วระบบต้องสามารถตรวจหาหรือต้องยอมให้สมาชิกระบุถึงความเปลี่ยนแปลงของความต้องการและควรตอบสนองโดยการปรับเข้าหาการเปลี่ยนแปลงที่ถูกระบุไว้ ระบบจะต้องสามารถในการค้นหาโดเมนใหม่ ๆ และคาดการณ์ลักษณะภายหน้าของสารสนเทศที่อาจเป็นที่สนใจของสมาชิก ดังนั้นภายใต้หมวดข้อมูลเกี่ยวกับการปฏิบัติการของสมาชิก จะต้องรวบรวมกลุ่มของการกระทำของสมาชิก กับระบบซึ่งข้อมูลทางปฏิบัติการอาจเป็นประโยชน์ในการใช้สำหรับการทำให้การระบุข้อมูลของสมาชิกได้ละเอียดมากขึ้น

5) หมวดข้อมูลด้านความปลอดภัย

เป็นการรวบรวมค่าตั้งพึงใจของสมาชิก และเงื่อนไขที่เกี่ยวกับข้อมูลที่แสดงถึงคำบรรยายลักษณะของผู้ใช้ เรื่องใดบ้างที่สามารถเข้าถึงได้ ซึ่งค่าตั้งพึงใจเหล่านี้โดยปกติแล้วสมาชิกอาจตั้งค่าข้อกำหนดภาวะส่วนตัวที่แตกต่างกันสำหรับแต่ละบริการที่สมาชิกมีการเข้าถึง

2.1.4 อัลกอริธึมของการจับคู่สำหรับการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา

สำหรับอัลกอริธึมของการจับคู่ใด ๆ มักจะมีกระบวนการก่อนประมวลผล (Preprocess) ที่ทำหน้าที่ในการแปลงกลุ่มของบัญชีสมาชิก (Subscription) ที่อยู่ภายในคำบรรยายลักษณะผู้ใช้ที่ต้องการจะแมทช์ให้อยู่ในรูปแบบภายในที่จะถูกใช้งานโดยอัลกอริธึม เพื่อพิจารณาว่าบัญชีสมาชิกใดที่แมทช์กับสารสนเทศที่เข้ามา

สิ่งที่ต้องคำนึงถึงในการออกแบบอัลกอริธึมของการจับคู่ ของการติดต่อแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ได้แก่ จำนวนเพรดิเคตที่จะต้องทำการประเมิน จำนวนประเภทของแอททริบิวท์ของสารสนเทศ และอัตราการเกิดของสารสนเทศ

2.1.4.1 อัลกอริธึมแบบนาอีฝ (Naïve algorithm) [4]

เป็นอัลกอริธึมที่ง่ายที่สุดในการแก้ปัญหาของการหาบัญชีสมาชิกที่แมทช์กับสารสนเทศ โดยจะทำการทดสอบกับทุกเพรดิเคตของทุกบัญชีสมาชิกแบบหนึ่งต่อหนึ่ง กับสารสนเทศที่เผยแพร่ ดังแสดงในรูปที่ 2.2

```
1:// S is the set of subscriptions, d is the document to match
2:// s is the subscription, p is the predicate
3:// M is the set of matched subscriptions
4: M = {}
5: for every s in S do
6:     m = true
7:     for every p in s do
8:         if not Match(p,d) then
9:             m = false
10:            break
11:        endif
12:    endfor
13:    if m then
14:        add s to M
15:    endif
16:endfor
```

รูปที่ 2.2 อัลกอริธึมแบบนาอีฝ

ฟังก์ชั่น *Match*(*p,d*) ในบรรทัดที่ 8 เป็นการตรวจสอบเพรดิเคตกับสารสนเทศที่เข้ามา ซึ่งถ้าเพรดิเคตสามารถจับคู่กับสารสนเทศจะส่งผลลัพธ์กลับมาเป็นจริง และจะส่งผลลัพธ์เป็นเท็จในกรณีกลับกัน โดยเวลาที่ใช้ในการจัดการกับสารสนเทศ *d* คือ

$$O\left(\sum_{s \in S}(\#\text{predicate of } s)C_p\right)$$

เมื่อ $s$ แทนบัญชีสมาชิกของสมาชิก

$S$ แทนเซตของบัญชีสมาชิกของสมาชิก

$C_P$ แทนเวลาที่ใช้ในการแมทช์เพรดิเคต

จากค่าผลการปฏิบัติการของอัลกอริธึมแบบนาอีฝ จะมีค่าลดลงเมื่อจำนวนของสมาชิกเพิ่มมากขึ้น ซึ่งปกติแล้วอัลกอริธึมในการจับคู่จะถูกใช้งานในสภาพแวดล้อมที่มีสมาชิกจำนวนมาก ซึ่งถ้าหากเวลาตอบสนองเป็นองค์ประกอบที่สำคัญ และอัตราการเกิดของสารสนเทศมีสูง อัลกอริธึมแบบนาอีฝ จะไม่เป็นทางเลือกที่เหมาะสม

จากการศึกษาพบว่าปัญหาหลักของอัลกอริธึมแบบนาอีฝ คือการที่มีเพรดิเคตซ้ำซ้อนและต้องทำการประเมินกับทุก ๆ เพรดิเคตที่มีอยู่ ดังนั้นแนวคิดที่จะแก้ไขข้อบกพร่องนี้เริ่มขึ้นโดยให้ความสำคัญกับเพรดิเคตมากกว่าบัญชีสมาชิก และอัลกอริธึมของการจับคู่จะต้องหลีกเลี่ยงการประเมินเพรดิเคตซ้ำ โดยการแยกบัญชีสมาชิกออกเป็นเพรดิเคตและอาจใช้หลักการนิรนัย (Deduction) ในการลดจำนวนเพรดิเคตที่จะต้องทำการประเมินเช่นถ้าเพรดิเคต (ราคาหุ้น,<, 10) แมทช์กับสารสนเทศแล้วจะทำให้เพรดิเคต (ราคาหุ้น,<, 20) แมทช์ด้วย ซึ่งจะทำให้เซตของเพรดิเคตที่จะต้องประเมินถูกจัดการอย่างมีประสิทธิภาพ

2.1.4.2 อัลกอริธึมแบบแฟร์เพรดิเคต (Fair predicate algorithm) [4]

ในอัลกอริธึมนี้จะมีการจัดเพรดิเคตให้เป็นคลัสเตอร์ (Cluster) ระหว่างการทำกระบวนการก่อนประมวลผล โดยการใช้ตัวดำเนินการเปรียบเทียบ (Comparison operator) และแอททริบิวท์ในการจัดทำคลัสเตอร์ และใช้ตารางเพรดทูสับ (*pred_to_subs*) ในการเชื่อมความสัมพันธ์ระหว่างเพรดิเคตกับบัญชีสมาชิกที่เป็นเจ้าของเพรดิเคตนั้น ทำให้มีการจัดเก็บเพรดิเคตเพียงที่เดียว ดังแสดงในรูปที่ 2.3

การทำงานของอัลกอริธึมแบบแฟร์เพรดิเคต จะแบ่งออกเป็นสองขั้นตอนโดยในขั้นตอนแรก (บรรทัดที่ 5-7) จะเป็นการประเมินเพรดิเคตกับสารสนเทศที่เข้ามาเพื่อค้นหาจำนวนเพรดิเคต

ที่แมทช์กับสารสนเทศ และเก็บค่าไว้ในตัวแปร *satisfied_preds* โดยฟังก์ชัน *eqpredmatch*(*d*) *lesspredmatch*(*d*) และ *greaterpredmatch*(*d*) ใช้สำหรับค้นหาเพรดิเคต แบบเท่ากับ น้อยกว่า และมากกว่า ที่แมทช์กับสารสนเทศ *d* ตามลำดับ ส่วนขั้นตอนที่สอง (บรรทัดที่ 9-18) เป็นการคำนวณหาเซตของบัญชีสมาชิกที่แมทช์กับสารสนเทศที่เข้ามา จากเซตของเพรดิเคตที่แมทช์กับสารสนเทศที่ได้จากขั้นตอนแรก และจะทำการนับจำนวนของเพรดิเคตที่ตรงกับสารสนเทศของบัญชีสมาชิกแต่ละอัน (บรรทัดที่ 9-12) แล้วจึงทำการเปรียบเทียบกับจำนวนของเพรดิเคตของบัญชีสมาชิกนั้น ถ้าจำนวนที่นับและจำนวนเพรดิเคตของบัญชีสมาชิกเท่ากันแสดงว่าบัญชีสมาชิกนั้นแมทช์กับสารสนเทศที่เข้ามา

```
1:// d is the document to match
2:// M is set of matched subscriptions
3:M = {}
4:// Step 1 – compute the predicates satisfied by d
5:satisfied_preds = eqpredmatch(d)
6:satisfied_preds = satisfied_preds + lesspredmatch(d)
7:satisfied_preds = satisfied_preds + greaterpredmatch(d)
8:// Step 2 – counts the number of satisfied predicates for each subscription
9:for every p in satisfied_pred do
10:     for every s in pred_to_subs[p] do
11:          hitcount[s] = hitcount[s] + 1
12:     endfor
13:endfor
14:for every s in S do
15:     if hitcount[s] = #subscriptions[i] pred then
16:          add s to M
17:     endif
18:endfor
```

รูปที่ 2.3 อัลกอริธึมแบบแฟร์เพรดิเคต

ฟังก์ชัน *eqpredmatch*(*d*) ใช้ในการค้นหาเพรดิเคตที่มีค่าเดียวกับค่าแอททริบิวท์ภายในคลัสเตอร์แบบเท่ากัน (Equality cluster) ที่สัมพันธ์กับค่าแอททริบิวท์ของสารสนเทศสำหรับแต่ละแอททริบิวท์ของสารสนเทศ ซึ่งถ้าคลัสเตอร์ถูกเก็บแบบเรียงลำดับและใช้ อัลกอริธึมการค้นหาแบบไบนารี (Binary searching algorithm) จะมีค่าความซับซ้อนเชิงเวลา (Time complexity) เป็น O ($\sum_{a \in d.atts}$(log (#$D_i$)+1) แต่ในกรณีที่ใช้ตารางแฮช จะมีค่าเป็น O (#*e.atts*) ซึ่ง *e.atts* แทนแอททริบิวท์ของสารสนเทศ

ฟังก์ชัน *lesspredmatch*(*d*) ใช้ในการค้นหาเพรดิเคตที่มีค่ามากที่สุด ที่มีค่าน้อยกว่าหรือเท่ากับค่าแอททริบิวท์ที่อยู่ในคลัสเตอร์แบบน้อยกว่า (Less-than cluster) ที่มีความสัมพันธ์กับแอททริบิวท์ของสารสนเทศ สำหรับแต่ละแอททริบิวท์ของสารสนเทศ ซึ่งเมื่อพบเพรดิเคตที่แมทช์กับสารสนเทศ จะทำให้เราทราบว่าเพรดิเคตลำดับก่อนหน้านั้นทั้งหมดจะตรงกับสารสนเทศที่เข้ามาและค่าความซับซ้อนเชิงเวลาจะมีค่าเป็น O ($\sum_{a \in d.atts}(\log(\#D_i)+\#D_i)$ และสำหรับฟังก์ชัน *greaterpredmatch*(*d*) จะมีลักษณะคล้ายคลึงกับฟังก์ชัน *lesspredmatch* โดยมีค่าความซับซ้อนเชิงเวลาที่เท่ากัน

อัลกอริธึมแบบแฟร์เพรดิเคต ได้แก้ไขปัญหาของการมีเพรดิเคตที่ซ้ำซ้อน และการทำการประเมินทุกเพรดิเคต แต่ข้อจำกัดของอัลกอริธึมนี้คือกระบวนการที่จำเป็นต้องคำนวณหาบัญชีสมาชิกที่แมทช์กับสารสนเทศจากเพรดิเคตพึงใจ (Satisfied predicate) ซึ่งค่าความซับซ้อนเชิงเวลาของอัลกอริธึมนี้คือ

$$O\ (C_{=}+C_{\leq}+C_{\geq}+C_{add}P_{sat}+C_{comp}(\#S))$$

เมื่อ $C_{=}$ $C_{\leq}$ และ$C_{\geq}$ แทนเวลาในการคำนวณหาเพรดิเคตพึงใจในกรณีเท่ากับ น้อยกว่าและมากกว่า ตามลำดับ

$C_{add}$ แทนเวลาที่ใช้ในการรวมผลลัพธ์

$P_{sat}$ แทนจำนวนของเพรดิเคตที่ตรงกับสารสนเทศ

$C_{comp}$ แทนเวลาที่ใช้ในการเปรียบเทียบ และ $S$ เป็นเซตของบัญชีสมาชิก

2.1.4.3 อัลกอริธึมแบบนีออน (Neon algorithm) [5]

เป็นผลิตภัณฑ์ของนีออนรูลส์ (NEONRules) ในอัลกอริธึมนี้มีการใช้แนวคิดที่ต่างออกไปในการแก้ไขปัญหาแมทชิ่ง โดยภายในอัลกอริธึมนี้จะไม่ได้ทำการแยกตัวประกอบของเพรดิเคต ทำให้เพรดิเคตสามารถถูกเก็บแบบซ้ำซ้อนตามบัญชีสมาชิก และอัลกอริธึมนี้ได้มีการพิจารณาถึงความแตกต่างของประเภทสารสนเทศ ซึ่งประเภทสารสนเทศแต่ละอันจะสอดคล้องกับสกีมาของสารสนเทศที่แตกต่างกัน บัญชีสมาชิกจะมีความสัมพันธ์กับประเภทสารสนเทศและจะถูกประเมิน ถ้าประเภทสารสนเทศของบัญชีสมาชิกสอดคล้องกับประเภทสารสนเทศของสารสนเทศที่จะถูกประมวลผล โดยบัญชีสมาชิกจะถูกระบุด้วยตัวเลขที่ไม่ซ้ำกัน และในช่วงเฟสของกระบวนการก่อนการประมวลผลบัญชีสมาชิกจะถูกจัดให้อยู่ในรูปแบบภายในดังนี้

ตาราง *predicates* จะทำหน้าที่เก็บเพรดิเคตของแต่ละบัญชีสมาชิก ซึ่งจะถูกระบุว่าเป็นของบัญชีสมาชิกใดโดยการใช้ตัวระบุบัญชีสมาชิก นอกจากนี้ยังเก็บประเภทของสารสนเทศแอททริบิวท์ของสารสนเทศ ตัวดำเนินการเปรียบเทียบที่อยู่ในเพรดิเคต และค่าแวลูย์ที่ใช้ในการเปรียบเทียบ

ตาราง *operations* ใช้สำหรับการระบุถึงการมีอยู่ของเพรดิเคต ด้วยคุณลักษณะพิเศษเฉพาะ ซึ่งตารางนี้ยอมให้มีการประเมินของเพรดิเคตที่อยู่ในตาราง *predicates* ที่อ้างถึงสารสนเทศที่กำลังประมวลผล โดยไม่สนใจเพรดิเคตตัวอื่น

ตาราง *n_predicates* เก็บรักษาจำนวนของเพรดิเคตสำหรับบัญชีสมาชิกแต่ละอัน

```
1:// d is the document to process. Operation and predicates are data
2:// structure filled in during the processing of the subscriptions.
3: M = {}
4: for every i in operation such as i.doc_type = d.type do
5:     value = value of d corresponding to attribute i.attribute
6:     if i.comparison = "EQ" then
7:         for every pred in predicates such as pred.doc_type = i.doc_type and
8:                 pred.attr = i.attr and pred.comparison = i.comparison and
9:                 pred.value = value do
10:            if some_{j in and} pred.sub_id = j.sub_id then
11:                j.count = j.count + 1
12:            else
13:                add (pred.sub_id, 1) to and
14:            endif
15:        endfor
16:        … // similar to less and greater than comparison
17:    endif
18: endfor
19: for every sub in and do
20:    if sub.count = n_predicates[sub.sub_id] then
21:        M = M ∪ sub.sub_id
22:    endif
23: endfor
```

รูปที่ 2.4 อัลกอริธึมแบบนีออน

อัลกอริธึมนี้แสดงในรูปที่ 2.4 โดยเริ่มต้นจากการพิจารณาอิลิเมนท์ภายในตาราง *operations* ทุกอิลิเมนท์ ที่แสดงว่าเพรดิเคตมีแนวโน้มจะแมทช์กับสารสนเทศที่เข้ามา (บรรทัดที่ 4) นั่นคืออิลิเมนท์ที่มีค่าของฟิลด์ *doc_type* เท่ากับค่าฟิลด์ *doc_type* ของสารสนเทศที่เข้ามา

สำหรับแต่ละอิลิเมนท์ที่ได้จะได้ค่าแวลูย์ ของสารสนเทศที่สอดคล้องกับแอททริบิวท์ของสารสนเทศที่ถูกระบุโดยอิลิเมนท์ที่กำลังพิจารณา (บรรทัดที่ 5) และเพรดิเคตที่ถูกแสดงด้วยอิลิเมนท์นี้จะถูกประมวลผลเพื่อหาเพรดิเคตพึงใจ ซึ่งการประมวลผลจะขึ้นอยู่กับตัวดำเนินการเปรียบเทียบที่อยู่ในเพรดิเคต สำหรับการเปรียบเทียบแบบเท่ากัน (Equality comparison) (บรรทัดที่ 6-15) โดยเพรดิเคตพึงใจจะเป็นเพรดิเคตที่ตรงกันกับอิลิเมนท์ของตาราง *predicates* ที่มีฟิลด์ *doc_type attribute* และ *comparison* เท่ากับค่าฟิลด์ทั้งสามของสารสนเทศที่เข้ามาตามลำดับ และมีฟิลด์แวลูย์เท่ากับค่าแวลูย์ของสารสนเทศที่เข้ามา และสำหรับแต่ละเพรดิเคตที่แมทช์กับสารสนเทศ อิลิเมนท์ของเพรดิเคตและโครงสร้างข้อมูล *and* จะถูกปรับปรุงให้ทันสมัย ซึ่งสำหรับตัวดำเนินการเปรียบเทียบประเภทอื่นจะถูกจัดการด้วยวิธีการเดียวกันกับการเปรียบเทียบแบบเท่ากัน และหลังจากที่ทุกอิลิเมนท์ในตาราง *operations* ถูกประมวลผลเรียบร้อยแล้ว จะทำการหาบัญชีสมาชิกที่แมทช์ (บรรทัดที่ 19-23) โดยบัญชีสมาชิกจะแมทช์กับสารสนเทศที่เข้ามา ถ้าจำนวนของเพรดิเคตพึงใจของบัญชีสมาชิกมีค่าเท่ากับจำนวนเพรดิเคตของบัญชีสมาชิก

#### 2.1.4.4 อัลกอริธึมอื่น ๆ

นอกจากอัลกอริธึมทั้งสามแบบที่ได้กล่าวไปแล้ว ยังมีอัลกอริธึมอื่น [6-8] สำหรับการแก้ปัญหาทางด้านการแมทชิ่งระหว่างบัญชีสมาชิกและสารสนเทศ โดยจะมีการใช้งานแผนภูมิต้นไม้เป็นตัวแทนของบัญชีสมาชิก โดยที่ลีฟโหนด (Leaf node) จะเก็บค่าบัญชีสมาชิกที่ตรงกับสารสนเทศตามค่าของแอททริบิวท์ที่ระบุไว้ในโหนดกิ่ง (Non-leaf node) และเส้นเชื่อม (Edge) จะแสดงผลลัพธ์ที่สอดคล้องผลการดำเนินการกับค่าแอททริบิวท์ของโหนดกิ่ง ซึ่งในบางอัลกอริธึมจะมีการใช้งาน *-edge เพื่อแสดงว่าสามารถท่องไปถึงบัญชีสมาชิกโดยผ่านเส้นเชื่อมนี้โดยไม่สนใจผลลัพธ์ของการดำเนินการที่เกี่ยวข้องกับโหนดกิ่งของเส้นเชื่อมนั้น

## 2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

### 2.2.1 งานวิจัย “Matching Event in A Content-based Subscription System” [8]

ในงานวิจัย [8] ได้นำเสนอวิธีการแก้ปัญหาของการจับคู่ระหว่างแอททริบิวท์ของเพรดิเคต และแอททริบิวท์ของอีเวนต์ (หรือสารสนเทศ) โดยการใช้แผนภูมิต้นไม้ในการหาบัญชีสมาชิกที่แมทช์กับแต่ละอีเวนต์ ซึ่งวิธีการที่นำเสนอได้พยายามให้มีค่าความซับซ้อนเชิงเวลาเป็นลักษณะสับลิเนียร์ (Sub linear) กับจำนวนของบัญชีสมาชิก และมีค่าความซับซ้อนเชิงพื้นที่เป็นแบบ

ลิเนียร์ (Linear) และการพิสูจน์ในขั้นต้นมีค่าความคาดหมายทางด้านเวลาในการแมทช์กับอีเวนต์แบบสุ่มไม่เกิน $O(N^{1-\lambda})$ เมื่อ $N$ คือจำนวนบัญชีสมาชิก และ $\lambda$ คือนิพจน์รูปแบบปิด (Closed-form expression) ที่ขึ้นอยู่กับจำนวน และชนิดของแอททริบิวท์

2.2.2 งานวิจัย “Efficient Filtering of XML Document for Selective Dissemination of Information” [9]

ในงานวิจัย [9] ได้นำเสนอสถาปัตยกรรมที่มีชื่อว่า เอ็กซ์ฟิลเตอร์ (XFilter) ซึ่งเป็นการจับคู่เอกสารเอ็กซ์เอ็มแอล กับคำบรรยายลักษณะผู้ใช้จำนวนมากอย่างมีประสิทธิภาพ โดยการใช้โครงสร้างทางดัชนี (Index structure) ในการหาโพรไฟล์ที่ตรงกับเอกสารเอ็กซ์เอ็มแอล ส่วนภาษาที่ใช้ในการอธิบายคำบรรยายลักษณะผู้ใช้ ได้มีการใช้ภาษาเอ็กซ์พาธ (XPath) ซึ่งถูกออกแบบสำหรับการใช้งานของเอ็กซ์เอสแอลที (XSLT: XSL Transformation) และภาษาเอ็กซ์พอยเตอร์ (XPointer language)

2.2.3 งานวิจัย “Filtering Algorithm and Implementation for Very Fast Publish/Subscribe Systems” [10]

ในงานวิจัย [10] ได้มีการเสนอระบบโต้ตอบแบบพับลิช/สับสไครบ์ที่รองรับการทำงานกับบัญชีสมาชิกจำนวนมาก และอีเวนต์ที่เข้ามามีปริมาณสูง โดยผู้วิจัยได้นำเสนออัลกอริธึมสำหรับการแมทชิ่งบนหน่วยความจำหลักที่มีประสิทธิภาพ ซึ่งมีลักษณะเด่นได้แก่ ประการแรก อัลกอริธึมมีการสร้างโครงสร้างข้อมูล ที่ปรับปรุงให้เข้ากับความซับซ้อนของภาษาที่ใช้อธิบายคำบรรยายลักษณะผู้ใช้ ประการที่สอง อัลกอริธึมมีลักษณะเป็นตัวประมวลผลแคชเชิงสำนึก (Processor cache-conscious) และมีการใช้เทคนิคที่หลีกเลี่ยงการพลาดแคช (Cache miss) โดยการใช้คำสั่งของตัวประมวลผลที่ชื่อว่า พรีเฟทช์ (Prefetch) ประการที่สาม อัลกอริธึมมีการใช้งานการจัดทำคลัสเตอร์บนสกีมาของอีเวนต์ (Event schema) ที่ใช้แนวคิดหลัก 2 ประการคือ (1) การจัดกลุ่มบัญชีสมาชิกตามจำนวนเพรดิเคตและการเชื่อมแบบทั่วไป (Common conjunction) ของเพรดิเคตแบบเท่ากัน ทำให้สามารถประเมินบัญชีสมาชิก (บางส่วน) ได้จากการเปรียบเทียบเพียงครั้งเดียว (2) มีการใช้งานดัชนีแฮชชิ่งแบบหลายแอททริบิวท์ (Multi-attribute hashing index) ทำให้แอททริบิวท์ของบัญชีสมาชิกสามารถประเมินได้จากการเปรียบเทียบเพียงครั้งเดียว และประการสุดท้าย ได้มีการเตรียม cost-base algorithm ที่ให้รายละเอียดของบัญชีสมาชิก และ

สถิติของข้อมูลที่เข้ามา เพื่อให้สามารถปรับเพิ่มประสิทธิภาพของการทำคลัสเตอร์ให้เข้ากับความเปลี่ยนแปลงของบัญชีสมาชิก และรูปแบบของสารสนเทศ

2.2.4 งานวิจัย “Predicate Matching and Subscription Matching in Publish/Subscribe Systems” [11]

ในงานวิจัย [11] ได้มีการตั้งข้อสังเกตว่าค่าประเภทโดเมน (Domain type) ของเพรดิเคตทั้งหมดจะมีค่าเป็นจำนวนตัวเลขที่แน่นอน จากข้อสังเกตนี้ผู้วิจัยได้นำเสนอ แนวคิดการค้นหาโดยใช้ตาราง (Table-based look-up scheme) สำหรับการประเมินเพรดิเคตที่ทำให้ค้นหาทุกเพรดิเคตที่แมทช์สำหรับแต่ละประเภทด้วยการค้นภายในตารางเพียงครั้งเดียว นอกจากนี้ยังได้มีการเสนออัลกอริธึมที่ใช้ในการแมทชิ่งในลักษณะของฐานข้อมูล (DBMS-based matching algorithm) อีก 2 อัลกอริธึม และทำการเปรียบเทียบอันที่ดีกว่ากับแมทชิ่งอัลกอริธึมสำหรับพับลิช/สับสไครบ์ที่มีโดยเฉพาะ ซึ่งทางผู้วิจัยได้ชี้ให้เห็นอย่างชัดเจนว่าอัลกอริธึมทั้งสองแบบนี้จะไม่เหมาะสมกับการทำงานกับสถานการณ์ที่มีภาระงานมาก และมีการประมวลผลอีเวนต์ในปริมาณมาก

2.2.5 งานวิจัย “Event Matching in Symmetric Subscription Systems” [12]

ในงานวิจัยได้กล่าวถึงความสำคัญของอัลกอริธึมของการแมทชิ่งอีเวนต์ (Event matching algorithm) ที่มีต่อ ผลการปฏิบัติงานและความสามารถในการขยาย ของระบบการโต้ตอบแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา โดยได้เสนอเทคนิคของการแมทชิ่งอีเวนต์ และ เทคนิคของการแก้ปัญหาแมทชิ่งอีเวนต์แบบการจัดการฐานข้อมูล (DBMS solutions) ซึ่งระบบการโต้ตอบแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ยอมให้สมาชิกที่รับสารสนเทศกำหนดเกณฑ์การกรองสารสนเทศ เมื่อลงทะเบียนความสนใจที่จะได้รับการแจ้งอีเวนต์ที่แมทช์กับความต้องการ ซึ่งผู้เผยแพร่สารสนเทศไม่ได้มีการจัดการในลักษณะที่ยืดหยุ่นแบบเดียวกัน ยิ่งไปกว่านั้นระบบการโต้ตอบแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหา ต้องการทราบค่าของแต่ละแอททริบิวท์ที่อยู่ภายในสกีมาของอีเวนต์ (Event schema) ซึ่งจะรู้ได้ก็ต่อเมื่ออยู่ในช่วงเวลาของการเผยแพร่สารสนเทศและ ผู้เผยแพร่สารสนเทศมียังความต้องการที่เลือกประเภทของสมาชิกที่รับสารสนเทศ ที่สามารถจะรับการเผยแพร่ข้อมูลนั้น และความสามารถในการปรับปรุงความเป็นส่วนตัวในการเผยแพร่สารสนเทศ (Personalize publication) ในระดับของแอททริบิวท์ ซึ่งงานวิจัยนี้ได้แนะนำระบบที่จัดการกับประเด็นข้างต้น

# บทที่ 3
# วิธีออกแบบการวิจัย

ในบทนี้จะกล่าวถึง สถาปัตยกรรมของการเผยแพร่สารสนเทศด้านหลักทรัพย์โดยใช้คำบรรยายลักษณะผู้ใช้สำหรับตลาดหลักทรัพย์ (Security Information Repository and User Profile based Dissemination: SIRUP) โดยจะกล่าวถึงแนวคิดที่ใช้ออกแบบการวิจัย โครงสร้างข้อมูลที่ใช้ในการวิจัย รวมไปถึงขั้นตอนวิธีการทำงานของส่วนที่สำคัญของงานวิจัย

## 3.1 ความต้องการโดยรวมของระบบ

ในการส่งสารสนเทศจากตลาดหลักทรัพย์จะแบ่งสารสนเทศออกเป็น 2 ประเภทคือ สารสนเทศแบบเรียลไทม์และสารสนเทศแบบไม่เรียลไทม์ โดยสารสนเทศแบบเรียลไทม์จะถูกส่งในขณะที่ตลาดหลักทรัพย์ฯ เปิดทำการและสารสนเทศแบบไม่เรียลไทม์จะถูกส่งหลังจากที่ตลาดหลักทรัพย์ฯ ปิดทำการแล้ว ซึ่งสมาชิกที่ต้องการรับสารสนเทศจะต้องทำการสมัครสมาชิกด้วยการระบุลักษณะของสารสนเทศที่ต้องการจะรับ โดยการจัดทำเป็นเพรดิเคตแล้วทำการบันทึกไว้ในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ หลังจากนั้นเจ้าหน้าที่จะเป็นผู้จัดทำรูปแบบโพรไฟล์ และเป็นผู้จัดการกับสารสนเทศแบบไม่เรียลไทม์ ได้แก่การนำเข้าสารสนเทศแบบไม่เรียลไทม์ให้กับตัวกรองสารสนเทศ การกรองสารสนเทศก่อนส่งให้สมาชิกซึ่งการกรองสารสนเทศทำได้โดยการตรวจสอบเพรดิเคตที่อยู่ภายในคำบรรยายลักษณะผู้ใช้กับสารสนเทศที่ได้รับจากตลาดหลักทรัพย์ฯ โดยในกรณีของการรับสารสนเทศแบบไม่เรียลไทม์สมาชิกสามารถเลือกรูปแบบโพรไฟล์ของสารสนเทศที่ต้องการรับได้

จากผลการวิเคราะห์ดังกล่าวสามารถสรุปเพื่อใช้สร้างแผนภาพยูสเคส ได้ดังนี้

1) แอคเตอร์ (Actor) ของระบบมีทั้งหมด 3 แอคเตอร์ ได้แก่ ตลาดหลักทรัพย์ซึ่งทำหน้าที่เป็นผู้ให้ข้อมูล เจ้าหน้าที่ และสมาชิก

2) ยูสเคสที่เกี่ยวข้องกับแอคเตอร์ตลาดหลักทรัพย์ที่จำเป็นต้องมีคือ ยูสเคสการรับข้อมูลจากตลาดหลักทรัพย์ ซึ่งแบ่งออกเป็น 2 ยูสเคสย่อยคือ การรับข้อมูลแบบเรียลไทม์ และการรับข้อมูลแบบไม่เรียลไทม์

3) เนื่องจากข้อมูลที่ส่งมาจากตลาดหลักทรัพย์มี 2 ประเภทจึงต้องมี ยูสเคสการกรองข้อมูลที่แยกออกเป็น 2 ยูสเคสย่อยคือ การจับคู่ข้อมูลแบบไม่เรียลไทม์กับ

คำบรรยายลักษณะผู้ใช้ และการจับคู่ข้อมูลแบบเรียลไทม์กับคำบรรยายลักษณะผู้ใช้

4) ยูสเคสที่เกี่ยวข้องกับแอคเตอร์สมาชิกคือ ยูสเคสการสมัครสมาชิก ซึ่งแบ่งออกเป็น 3 ยูสเคสย่อยคือ การสร้างคำบรรยายลักษณะผู้ใช้ การสร้างเพรดิเคต และการบันทึกข้อมูลคำบรรยายลักษณะผู้ใช้ และยูสเคสการส่งข้อมูลเพื่อทำการรับข้อมูลจากระบบ ซึ่งยูสเคสการสมัครสมาชิกจะต้องเรียกใช้ (Uses) ยูสเคสการสร้างคำบรรยายลักษณะผู้ใช้ และยูสเคสการสร้างคำบรรยายลักษณะผู้ใช้ก็จะต้องเรียกใช้ยูสเคสการสร้างเพรดิเคต

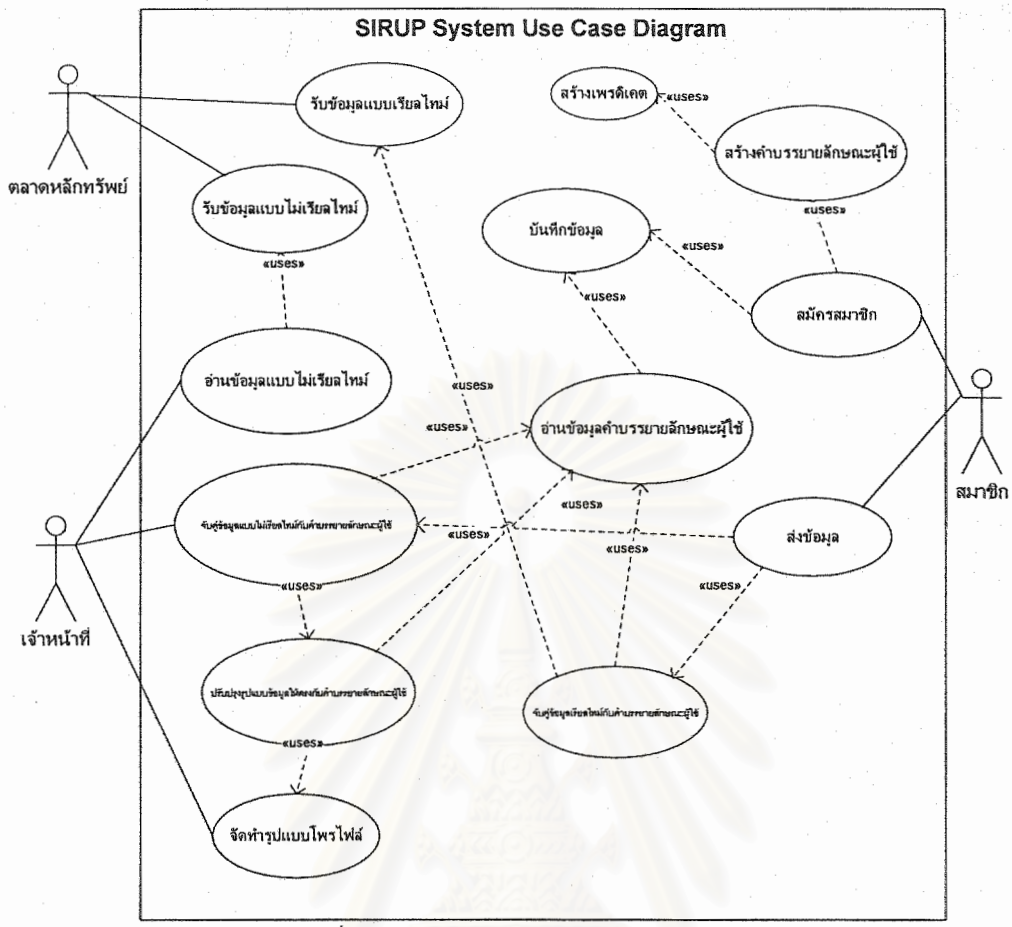
5) การสมัครสมาชิกจะต้องมีการจัดเก็บข้อมูล จึงควรมียูสเคสบันทึกข้อมูลเพิ่มขึ้น และยูสเคสการสมัครสมาชิกจะต้องเรียกใช้ยูสเคสบันทึกข้อมูลนี้

6) ยูสเคสที่เกี่ยวข้องกับแอคเตอร์เจ้าหน้าที่คือ ยูสเคสการอ่านข้อมูลแบบไม่เรียลไทม์ โดยจะเรียกใช้ยูสเคสการรับข้อมูลแบบไม่เรียลไทม์ และยูสเคสการจัดทำรูปแบบโพรไฟล์

7) สำหรับการกรองข้อมูล จะต้องมีการอ่านข้อมูลจากคำบรรยายลักษณะผู้ใช้ ดังนั้นจึงควรมี ยูสเคสอ่านข้อมูลคำบรรยายลักษณะผู้ใช้เพิ่มขึ้นและยูสเคสการกรองข้อมูล (ทั้ง 2 ยูสเคส) จะต้องเรียกใช้ยูสเคสอ่านข้อมูลคำบรรยายลักษณะผู้ใช้นี้ และยูสเคสอ่านข้อมูลคำบรรยายลักษณะผู้ใช้จะต้องเรียกใช้ยูสเคสบันทึกข้อมูล

8) เนื่องจากกรณีของข้อมูลแบบไม่เรียลไทม์ สมาชิกสามารถเลือกรูปแบบโพรไฟล์ของข้อมูลที่จะรับได้ จึงควรมียูสเคสปรับปรุงรูปแบบข้อมูลให้ตรงกับรูปแบบโพรไฟล์ในคำบรรยายลักษณะผู้ใช้ โดยจะมีการเรียกใช้ยูสเคสการจัดทำรูปแบบโพรไฟล์

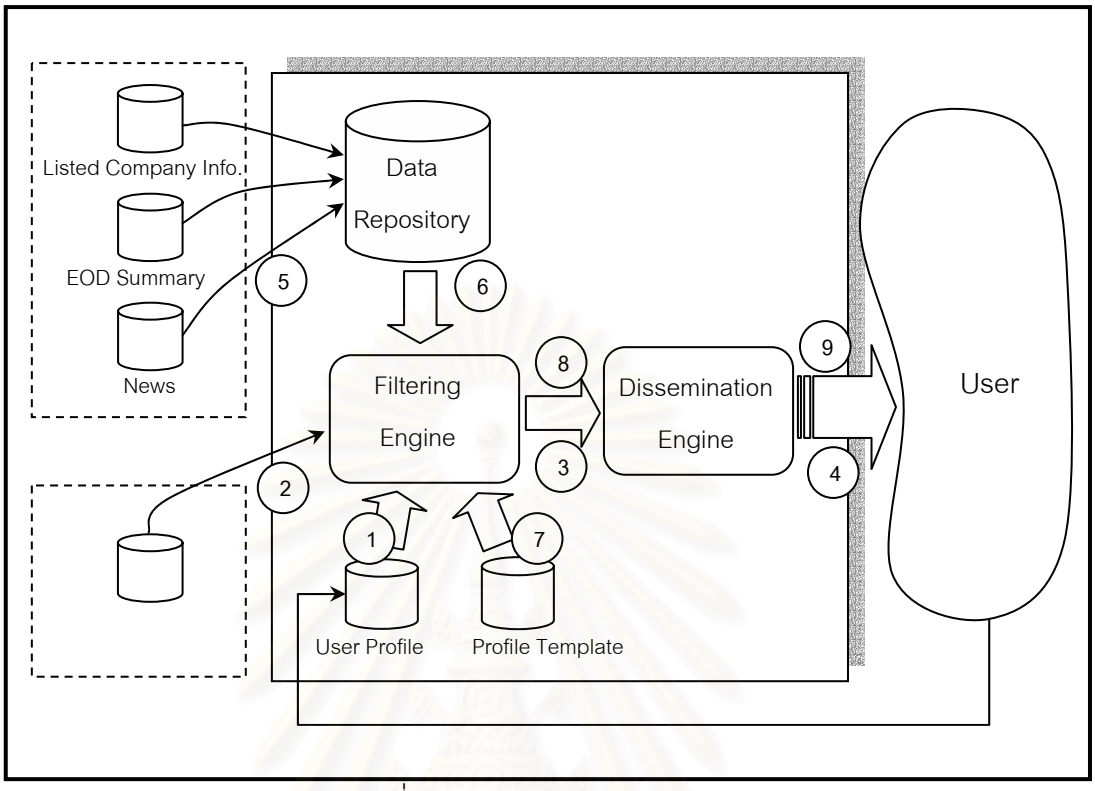
จากการวิเคราะห์จะทำให้เราได้แผนภาพยูสเคส ดังแสดงในรูปที่ 3.1




รูปที่ 3.1 แผนภาพยูสเคสของงานวิจัย

## 3.2 สถาปัตยกรรมของระบบซีรัฟ

สถาปัตยกรรมของงานวิจัยประกอบไปด้วยส่วนต่าง ๆ ได้แก่ ที่เก็บสารสนเทศแบบไม่เรียลไทม์ (Data repository) ตัวกรองสารสนเทศ (Filtering engine) ตัวเผยแพร่สารสนเทศ (Dissemination engine) คำบรรยายลักษณะผู้ใช้ (User profile) และรูปแบบโพรไฟล์ (Profile template) ซึ่งตัวอย่างของการทำงานของระบบเช่น นักลงทุนที่ต้องการรับสารสนเทศที่เกี่ยวข้องกับหลักทรัพย์ของบริษัท ก จะต้องทำการสมัครเป็นสมาชิกโดยทำการสร้างเพรดิเคตที่ระบุถึงบริษัท ก และวิธีการในการส่งและรูปแบบของการส่งสารสนเทศเก็บไว้ในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ และเมื่อระบบได้รับสารสนเทศที่มาจากตลาดหลักทรัพย์ฯ จะทำการพิจารณาสารสนเทศที่ได้รับ กับเพรดิเคตที่สมาชิกระบุไว้ ถ้าพบว่าสารสนเทศนั้นมีเนื้อหาเกี่ยวกับบริษัท ก และสอดคล้องกับเพรดิเคตของสมาชิก ระบบจะส่งสารสนเทศนั้นไปให้กับสมาชิกตามวิธีการและรูปแบบในการส่งสารสนเทศที่สมาชิกระบุไว้ในคำบรรยายลักษณะผู้ใช้




รูปที่ 3.2 สถาปัตยกรรมของระบบ

จากรูปที่ 3.2 แสดงขั้นตอนการทำงานของสถาปัตยกรรม ดังนี้

1) ระบบทำการอ่านข้อมูลคำบรรยายลักษณะผู้ใช้ เข้าไปเก็บไว้ในหน่วยความจำ โดยมีการจัดทำดัชนีแยกตามประเภทของแอททริบิวท์ ที่ปรากฏอยู่ในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ และเริ่มต้นการทำงานของเซิร์ฟเวอร์ขาเข้า(เพื่อรับสารสนเทศแบบเรียลไทม์จากตลาดหลักทรัพย์ฯ) และเริ่มต้นการทำงานของเซิร์ฟเวอร์ขาออก (เพื่อส่งสารสนเทศแบบเรียลไทม์ที่ผ่านการกรองแล้วให้แก่สมาชิก)
2) เมื่อตลาดหลักทรัพย์ฯ เปิดทำการและเริ่มมีการซื้อขาย ระบบจะได้รับสารสนเทศแบบเรียลไทม์จากทางเซิร์ฟเวอร์ขาเข้า โดยสารสนเทศแบบเรียลไทม์จะถูกส่งให้กับตัวกรองสารสนเทศ เพื่อทำการกรองสารสนเทศแบบเรียลไทม์กับคำบรรยายลักษณะผู้ใช้ที่อยู่ในหน่วยความจำ
3) ตัวกรองสารสนเทศส่งสารสนเทศที่ผ่านการกรองแล้วให้กับตัวเผยแพร่สารสนเทศ
4) ตัวเผยแพร่สารสนเทศทำการส่งสารสนเทศออกไปให้กับสมาชิก ที่ต้องการสารสนเทศนั้นผ่านทางเซิร์ฟเวอร์ขาออก

5) หลังจากที่ตลาดหลักทรัพย์ฯ ปิดทำการ เมื่อระบบได้รับสารสนเทศแบบไม่เรียลไทม์จากตลาดหลักทรัพย์ฯ จะทำการเก็บสารสนเทศแบบไม่เรียลไทม์ไว้ในที่เก็บสารสนเทศแบบไม่เรียลไทม์
6) ระบบอ่านสารสนเทศแบบไม่เรียลไทม์ที่ได้รับให้กับตัวกรองสารสนเทศ
7) หลังจากผ่านการกรองแล้ว ระบบจะทำการปรับปรุงรูปแบบสารสนเทศตามที่สมาชิกระบุไว้ในคำบรรยายลักษณะผู้ใช้โดยใช้รูปแบบโพรไฟล์
8) ตัวกรองสารสนเทศส่งสารสนเทศที่ผ่านการกรอง และปรับปรุงรูปแบบแล้วไปให้กับตัวเผยแพร่สารสนเทศ
9) ตัวเผยแพร่สารสนเทศทำการส่งสารสนเทศให้กับสมาชิกตามรูปแบบและเวลาที่กำหนดไว้ในคำบรรยายลักษณะผู้ใช้

โดยส่วนที่สำคัญของสถาปัตยกรรมมีรายละเอียดดังนี้

### 3.2.1 คำบรรยายลักษณะผู้ใช้ (User Profile)

จากทฤษฎีในบทที่ 2 สามารถออกแบบรูปแบบของคำบรรยายลักษณะผู้ใช้ในงานวิจัยได้ดังนี้

คำบรรยายลักษณะ (Profile) จะสามารถระบุได้โดยใช้ตัวระบุคำบรรยายลักษณะ (ProfID) ซึ่งไม่ซ้ำกัน สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$\text{Profile} = \text{ProfID} \rightarrow \text{UserProfile}$$

โดยคำบรรยายลักษณะผู้ใช้จะประกอบไปด้วย 3 หมวดได้แก่ หมวดข้อมูลส่วนบุคคล (Personal Data: PersData) หมวดข้อมูลในการรวบรวมข้อมูล (Gathering Data: GathData) และหมวดข้อมูลการส่งข้อมูล (Delivery Data: DeliData) สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$\text{UserProfile} = (\text{PersData} \times \text{GathData} \times \text{DeliData})$$

#### 3.2.1.1 หมวดข้อมูลส่วนบุคคล

หมวดข้อมูลส่วนบุคคล จะประกอบไปด้วยข้อมูลเกี่ยวกับการระบุรายละเอียดส่วนบุคคลของสมาชิก สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$\text{PersData} = (\text{Account} \times \text{Name} \times \text{Address})$$

3.2.1.2 หมวดข้อมูลในการรวบรวมสารสนเทศ

หมวดข้อมูลในการรวบรวมข้อมูล เป็นส่วนที่บอกถึงความต้องการประเภทสารสนเทศของสมาชิกโดยการระบุเป็นเพรดิเคต ซึ่งภายในเพรดิเคตจะประกอบไปด้วยหมวดของสารสนเทศ ชื่อแอททริบิวท์ ตัวดำเนินการเปรียบเทียบ และค่าของแอททริบิวท์ โดยในส่วนของหมวดของสารสนเทศจะมีการแบ่งประเภทเพื่อให้สมาชิกเลือก ได้แก่ข้อมูลแบบเรียลไทม์ (Real Time Data: REAL) ข้อมูลข่าว (News: NEWS) ข้อมูลสรุปประจำวัน (End of Day Summary: EOD) และข้อมูลบริษัทจดทะเบียน (Listed Company Information: COMP) และในกรณีที่สมาชิกมีการระบุเพรดิเคตมากกว่าหนึ่งเพรดิเคต ระบบจะทำการเชื่อมเพรดิเคตแต่ละเพรดิเคตด้วยคำว่า “หรือ” ซึ่งสามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$GathData = PredicateID \rightarrow Predicate$$

$$Predicate = (Category \times Attribute \times Condition \times Value)$$

$$Category = (REAL + NEWS + EOD + COMP)$$

3.2.1.3 หมวดข้อมูลการส่งสารสนเทศไปให้สมาชิก

หมวดข้อมูลการส่งสารสนเทศไปให้สมาชิก เป็นส่วนที่บอกถึงวิธีการ (DeliMode) และเวลาในการส่งข้อมูล (DeliTime) ให้กับสมาชิกโดยขึ้นอยู่กับคำบรรยายลักษณะผู้ใช้ ดังนั้นสามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$DeliData = ProfID \rightarrow (DeliMode \times DeliTime)$$

วิธีการในการส่งข้อมูล (Delivery Mode: DeliMode) เป็นส่วนที่ระบุถึงวิธีการที่ใช้ในการส่งสารสนเทศ ลักษณะของสารสนเทศที่ส่งให้กับสมาชิก และที่อยู่ปลายทางในการส่งสารสนเทศโดยขึ้นอยู่กับวิธีการส่ง ซึ่งเขียนเป็นสัญลักษณ์ได้ดังนี้

$$DeliMode = (DeliMean \times Layout \times Destination)$$

วิธีการในการส่งสารสนเทศ (Delivery Mean: DeliMean) เป็นการระบุวิธีการส่งสารสนเทศไปให้กับสมาชิกเช่น วงจรเช่า (Lease line) จดหมายอิเล็กทรอนิกส์ (E-mail) เป็นต้นซึ่งสามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

$$DeliMean = (LLine + EMail)$$

ลักษณะของสารสนเทศ (Layout) เป็นการระบุลักษณะของสารสนเทศที่สมาชิกต้องการ ได้แก่ สารสนเทศแบบสมบูรณ์ สารสนเทศแบบธรรมดา และสารสนเทศแบบขั้นต่ำ สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

Layout = (ComDoc + TypDoc + MinDoc)

ที่อยู่ปลายทางในการส่ง (Destination) เป็นการระบุถึงปลายทางในการส่งสารสนเทศไปให้สมาชิกตามวิธีการส่งเช่น ถ้าวิธีการส่งเป็นอีเมล์จะจัดเก็บที่อยู่อีเมล์ของสมาชิก เป็นต้น

เวลาในการส่งสารสนเทศ (Delivery Time: DeliTime) เป็นส่วนที่ระบุถึงรูปแบบเวลาในการส่งสารสนเทศ โดยจะประกอบไปด้วยการส่งประเภทเป็นคาบเวลาแบบคงที่ (Fixed time interval) เช่น ส่งสารสนเทศทุกวันเวลา 9.00 น. หรือ การส่งแบบทันที (As soon as possible: ASAP) สามารถเขียนเป็นสัญลักษณ์ได้ดังนี้

DeliTime = (TimeInterval + ASAP)

3.2.1.4 สรุปรายละเอียดของโพรไฟล์

ตารางที่ 3.1 รายละเอียดและตัวอย่างคำบรรยายลักษณะผู้ใช้

| หมวดข้อมูล | รายละเอียด | ตัวอย่าง |
|---|---|---|
| PersData | Account = (UserName+Passwd) | Username = ‘g44scr’<br>Passwd = ‘44icugd’ |
| | Name = (FirstName+LasrtName) | FirstName = ‘สมชาย’<br>LastName = ‘รักไทย’ |
| | Address = (Address No +<br>Street+<br>City +<br>Province +<br>Country +<br>PostalCode) | AddressNo = ‘40/2 ซ.อิ่มบุญ’<br>Street = ‘ถ.พญาไท’<br>City = ‘เขตปทุมวัน’<br>Province = ‘กรุงเทพมหานคร’<br>Country = ‘ประเทศไทย’<br>PostalCode = ‘10330’ |
| GathData | Category | Category = ‘REAL’ |
| | Attribute | Attribute = ‘SecurityName’ |
| | Condition | Condition = ‘=’ |
| | Value | Value = ‘abc’ |

ตารางที่ 3.1(ต่อ) รายละเอียดและตัวอย่างคำบรรยายลักษณะผู้ใช้

| หมวดข้อมูล | รายละเอียด | ตัวอย่าง |
|---|---|---|
| DeliData | DeliMean | DeliMean = 'Email' |
| | Layout | Layout = 'ComDoc' |
| | Destination | Destination = 'abc@chula.com' |
| | DeliTime | DeliTime = 'ASAP' |

3.2.2 รูปแบบโพรไฟล์ (Profile Template)

เป็นตัวกำหนดลักษณะของสารสนเทศที่สมาชิกต้องการได้รับ ซึ่งจะถูกใช้งานสำหรับสารสนเทศแบบไม่เรียลไทม์ เนื่องจากสารสนเทศแบบไม่เรียลไทม์มีปริมาณมากจึงต้องมีการระบุถึงลักษณะของสารสนเทศที่จะรับ เพื่อให้สมาชิกได้รับสารสนเทศในลักษณะที่ตรงกับความต้องการมากที่สุด โดยมีการออกแบบไว้ 3 ลักษณะ ได้แก่ ลักษณะของสารสนเทศแบบสมบูรณ์ (COMDOC) ลักษณะของสารสนเทศแบบปกติ (TYPDOC) และลักษณะของสารสนเทศแบบขั้นต่ำ (MINDOC) ซึ่งการกำหนดลักษณะของสารสนเทศแสดงในตารางที่ 3.2

ตารางที่ 3.2 การกำหนดลักษณะของสารสนเทศภายในรูปแบบโพรไฟล์

| | ชื่อฟิลด์ | ความหมาย | COM DOC | TYP DOC | MIN DOC |
|---|---|---|---|---|---|
| 1 | Security Name | ชื่อย่อหลักทรัพย์ | ☑ | ☑ | ☑ |
| 2 | Security Id | รหัสหลักทรัพย์ | ☑ | ☑ | ☒ |
| 3 | Company Id | รหัสบริษัท | ☑ | ☒ | ☒ |
| 4 | Security Type | ประเภทหลักทรัพย์ | ☑ | ☑ | ☑ |
| 5 | Local Security Name | ชื่อย่อหลักทรัพย์ของหลักทรัพย์แบบโลคอล | ☑ | ☑ | ☒ |
| 6 | Local Security Id | รหัสหลักทรัพย์ของหลักทรัพย์แบบโลคอล | ☑ | ☒ | ☒ |
| 7 | Isin No. | หมายเลขกำกับหลักทรัพย์ | ☑ | ☑ | ☑ |

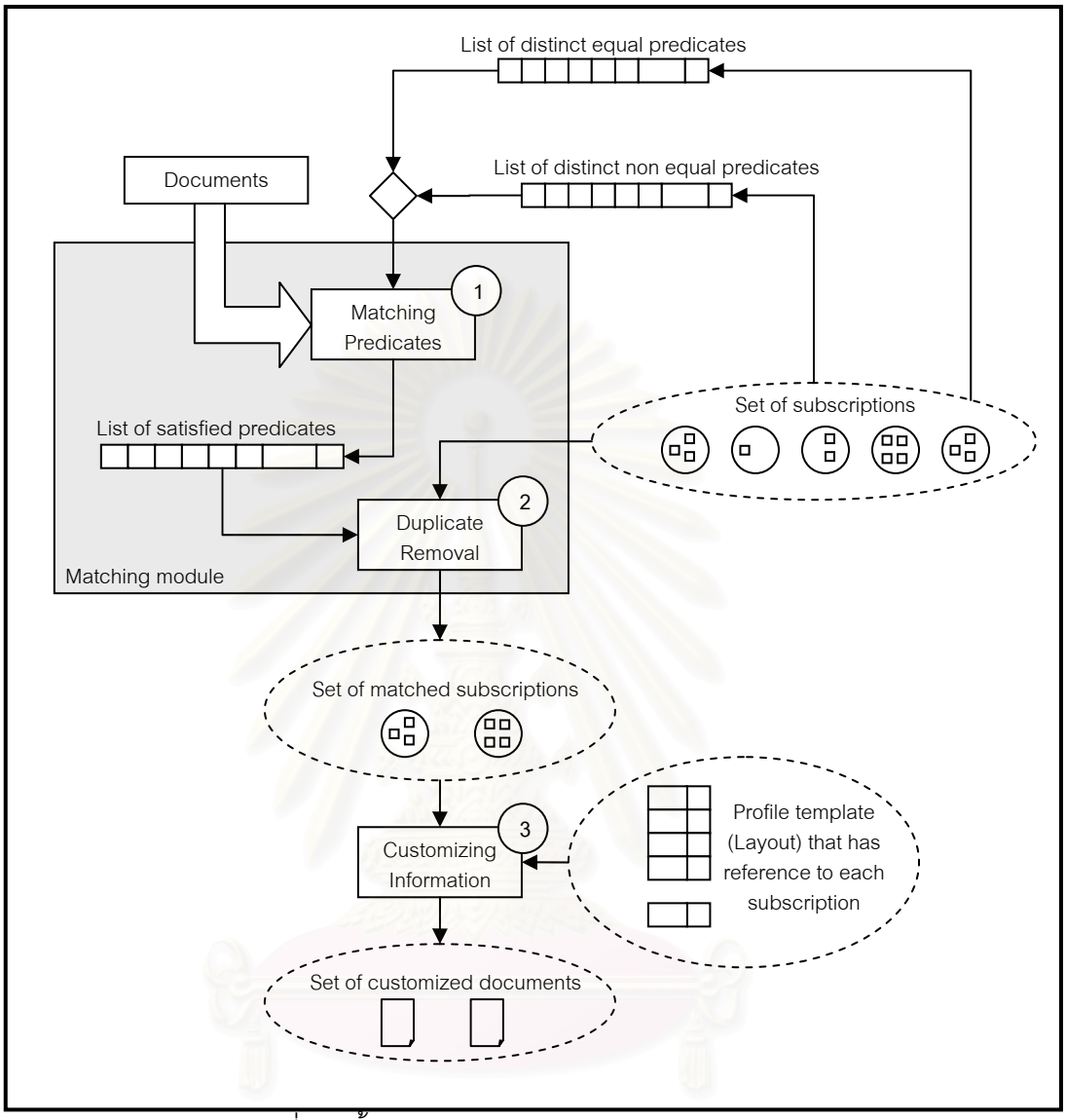
☑ แสดงถึงมีการส่งข้อมูลฟิลด์นั้นให้กับสมาชิก

☒ แสดงถึงไม่มีการส่งข้อมูลฟิลด์นั้นให้กับสมาชิก

3.2.3 ตัวกรองสารสนเทศ (Filtering Engine)

สำหรับตัวกรองสารสนเทศจะประกอบไปด้วยส่วนต่าง ๆ ที่สำคัญได้แก่ การจัดทำสกีมาของสารสนเทศที่มาจากตลาดหลักทรัพย์ การจัดทำดัชนีของคำบรรยายลักษณะผู้ใช้ และแมทชิ่งมอดูล ซึ่งในส่วนขั้นตอนการทำงานของตัวกรองสารสนเทศ ผู้วิจัยได้ออกแบบขั้นตอนโดยการทำงานหลักจะประกอบไปด้วยขั้นตอน 3 ขั้นตอนได้แก่ การจับคู่ระหว่างสารสนเทศกับเพรดิเคต (Matching predicate) การกำจัดการส่งสารสนเทศซ้ำ (Duplicate removal) การแปลงสารสนเทศ (Customizing Information) ดังแสดงในรูปที่ 3.3 ซึ่งมีขั้นตอนการทำงานดังนี้

1) จับคู่สารสนเทศเผยแพร่กับเพรดิเคตที่อยู่ในคำบรรยายลักษณะผู้ใช้ เพื่อสร้างรายการเพรดิเคตพึงใจ ที่สอดคล้องกับสารสนเทศที่กำลังพิจารณา ตัวอย่างเช่น สารสนเทศที่เผยแพร่เป็นสารสนเทศที่เกี่ยวกับหลักทรัพย์ กขค ซึ่งเป็นของบริษัท กขค จำกัด (มหาชน) ระบบจะทำการหาเพรดิเคตทั้งหมดที่สอดคล้องกับสารสนเทศของสารสนเทศดังกล่าว จากเพรดิเคตที่อยู่ในคำบรรยายลักษณะผู้ใช้มาสร้างเป็นรายการเพรดิเคตพึงใจ
2) จากรายการเพรดิเคตพึงใจที่ได้ อาจมีเพรดิเคตบางตัวที่ทำให้เกิดความซ้ำซ้อนในการส่งสารสนเทศให้กับสมาชิก ตัวอย่างเช่น สารสนเทศที่เข้ามาเป็นสารสนเทศที่เกี่ยวข้องกับหลักทรัพย์ กขค ซึ่งอยู่ในกลุ่มอุตสาหกรรมธนาคารและสถาบันการเงิน ถ้าหากมีสมาชิกส่งเพรดิเคตว่าขอรับข้อมูลของหลักทรัพย์ กขค และข้อมูลของกลุ่มอุตสาหกรรมธนาคารและสถาบันการเงิน จะทำให้สารสนเทศนั้นถูกส่งให้กับสมาชิก 2 ครั้งจึงจำเป็นที่จะต้องมีการกำจัดการส่งสารสนเทศซ้ำ ซึ่งหลังจากขั้นตอนนี้จะได้ผลลัพธ์เป็นเซตของบัญชีสมาชิกที่แมทช์กับสารสนเทศ
3) แปลงสารสนเทศให้ตรงกับ รูปแบบของสารสนเทศที่สมาชิกระบุไว้ในคำบรรยายลักษณะผู้ใช้ โดยการอ่านข้อมูลรูปแบบโพรไฟล์ของแต่ละบัญชีสมาชิกที่อยู่ในเซตของบัญชีสมาชิกที่แมทช์กับสารสนเทศ ตัวอย่างเช่น สมาชิกระบุข้อมูลรูปแบบโพรไฟล์ที่ต้องการจะรับเป็น สารสนเทศแบบสมบูรณ์ ระบบจะสร้างโครงสร้างของเอกสารแบบสมบูรณ์ จากสารสนเทศที่เผยแพร่ และทำการส่งโครงสร้างดังกล่าวให้กับ ตัวเผยแพร่สารสนเทศ

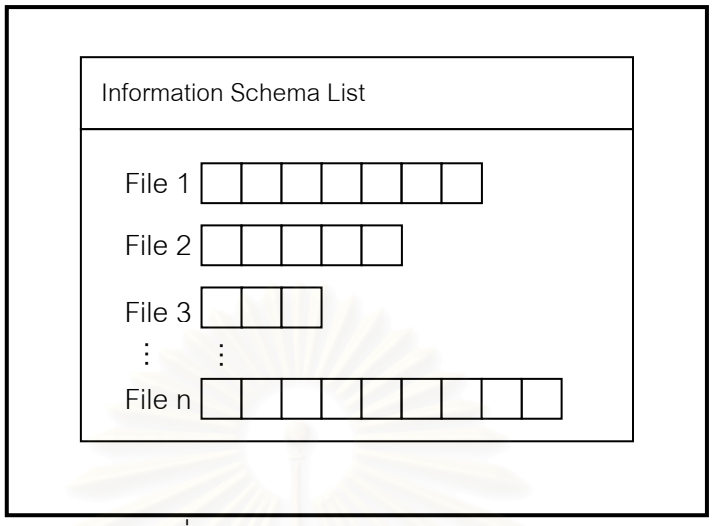



รูปที่ 3.3 ขั้นตอนการทำงานของตัวกรองสารสนเทศ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3.2.3.1 การจัดทำสกีมาของสารสนเทศที่มาจากตลาดหลักทรัพย์

เมื่อระบบได้รับสารสนเทศแบบไม่เรียลไทม์จากตลาดหลักทรัพย์ จะมีการจัดทำสกีมาของแต่ละไฟล์ที่เข้ามาโดยการเก็บประเภทของแอททริบิวท์ โดยได้แก่ความหมายของแต่ละฟิลด์ที่อยู่ในไฟล์ที่ได้รับ เพื่อใช้เป็นดัชนีในการเปรียบเทียบกับข้อมูลคำบรรยายลักษณะผู้ใช้ ซึ่งจัดทำเป็นรายการสกีมาของสารสนเทศในแต่ละไฟล์ ดังแสดงในรูปที่ 3.4

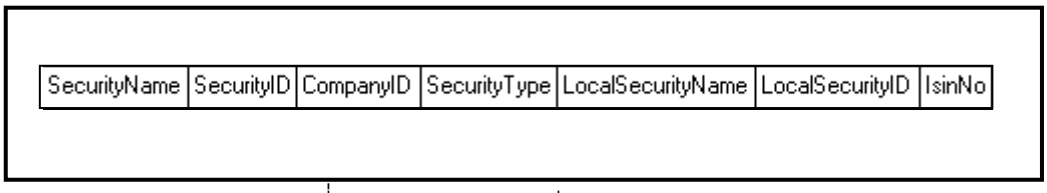



รูปที่ 3.4 รายการสกีมาของสารสนเทศ

ตัวอย่างเช่น กำหนดให้สารสนเทศแบบไม่เรียลไทม์ที่เผยแพร่มีลักษณะดังรูปที่ 3.5 เมื่อจัดทำสกีมาของสารสนเทศแล้วจะได้สกีมาของสารสนเทศดังรูปที่ 3.6

| SecurityName | SecurityID | Company | SecurityType | LocalSecurityName | LocalSecurityID | IsinNo |
|---|---|---|---|---|---|---|
| CPF-W1-R | 3710 | 101 | R | CPF-W1 | 2014 | TH010105C7R5 |
| KGI-V3 | 3711 | 171 | F | KGI-W3 | 1666 | TH017105C618 |
| KGI-W3-R | 3712 | 171 | R | KGI-W3 | 1666 | TH017105C6R2 |
| KK-V4 | 3707 | 121 | F | KK-W4 | 2093 | TH012105L119 |
| MSC-W1-R | 3713 | 569 | R | MSC-W1 | 2881 | TH056905FBR4 |
| NOBLE-V1 | 3714 | 564 | F | NOBLE-W1 | 2454 | TH056405C812 |
| NOBLE-W1-R | 3715 | 564 | R | NOBLE-W1 | 2454 | TH056405C8R3 |
| ROYNET-R | 3720 | 663 | R | ROYNET | 2572 | TH0663010R13 |
| SF-W1-R | 3716 | 699 | R | SF-W1 | 3392 | TH069905FCR7 |
| SSEC-W1-R | 3717 | 683 | R | SSEC-W1 | 2891 | TH068305FCR1 |
| TEAM-R | 3721 | 640 | R | TEAM | 2492 | TH0640010R11 |
| TTA-V2 | 3705 | 535 | F | TTA-W2 | 1889 | TH053505EB16 |
| ZMICO-V1 | 3718 | 472 | F | ZMICO-W1 | 1895 | TH047205C511 |

รูปที่ 3.5 ตัวอย่างสารสนเทศที่เข้ามาสู่ระบบ




รูปที่ 3.6 ตัวอย่างของสกีมาของสารสนเทศ

3.2.3.2 การจัดทำดัชนีของคำบรรยายลักษณะผู้ใช้

ในการจัดทำดัชนีของคำบรรยายลักษณะผู้ใช้จะมีการสร้างคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ และคลัสเตอร์ของสมาชิกตามเพรดิเคต เพื่อใช้ในการจับคู่กับสารสนเทศที่เผยแพร่ เช่นกำหนดให้บัญชีสมาชิกที่อยู่ในคำบรรยายลักษณะผู้ใช้มีลักษณะดังรูปที่ 3.7

ตัวอย่างเพรดิเคตของผู้ใช้

| UserID | Category | AttributeCD | Condition | Value |
|---|---|---|---|---|
| USER1 | COMP | SecurityName | = | KK-V4 |
| USER1 | COMP | SecurityName | = | KGI-V3 |
| USER2 | EOD | AccountFormID | = | 2 |
| USER4 | REAL | SecurityName | = | KDH |
| USER4 | REAL | SecurityName | = | SHIN |
| USER4 | COMP | SecurityName | = | ABICO |
| USER4 | EOD | ParValue | = | 34.46 |
| USER6 | NEWS | ParValue | < | 34.46 |
| USER6 | COMP | SecurityName | = | BROOK |
| USER8 | EOD | SecurityType | = | V |
| USER8 | COMP | ShareholderType | = | 0 |
| USER8 | EOD | NewsType | = | 53 |
| USER8 | NEWS | SecurityType | = | 0 |

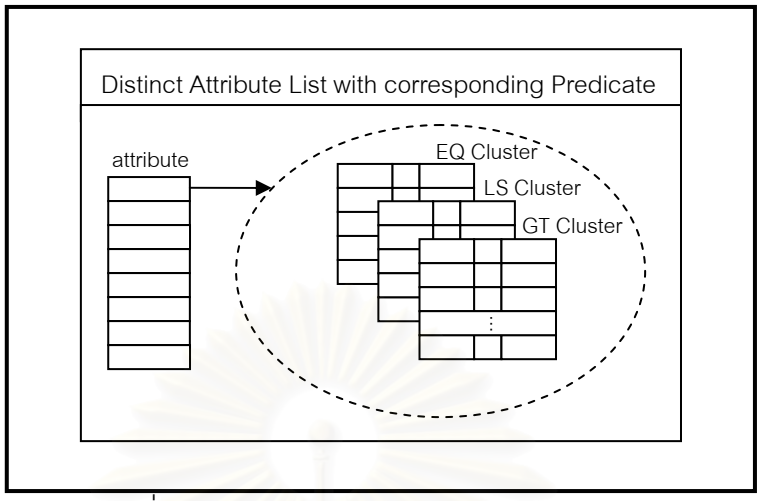
ตัวอย่างลักษณะของสารสนเทศที่ผู้ใช้ต้องการและวิธีการจัดส่งสารสนเทศไปให้กับผู้ใช้

| UserID | rdemCD | rlaoCD | rdtmCD | Destination | deliveryTime |
|---|---|---|---|---|---|
| USER1 | Email | COMDOC | ASAP | g44gud@cp.eng.chula.ac.th | |
| USER2 | Email | TYPDOC | INTERVAL | g44gud@cp.eng.chula.ac.th | 19:45 |
| USER4 | Email | TYPDOC | INTERVAL | g44gud@cp.eng.chula.ac.th | 23:00 |
| USER6 | Email | MINDOC | INTERVAL | g44gud@cp.eng.chula.ac.th | 22:15 |
| USER8 | Email | MINDOC | INTERVAL | g44gud@cp.eng.chula.ac.th | 10:00 |

รูปที่ 3.7 ตัวอย่างบัญชีสมาชิกของสมาชิก

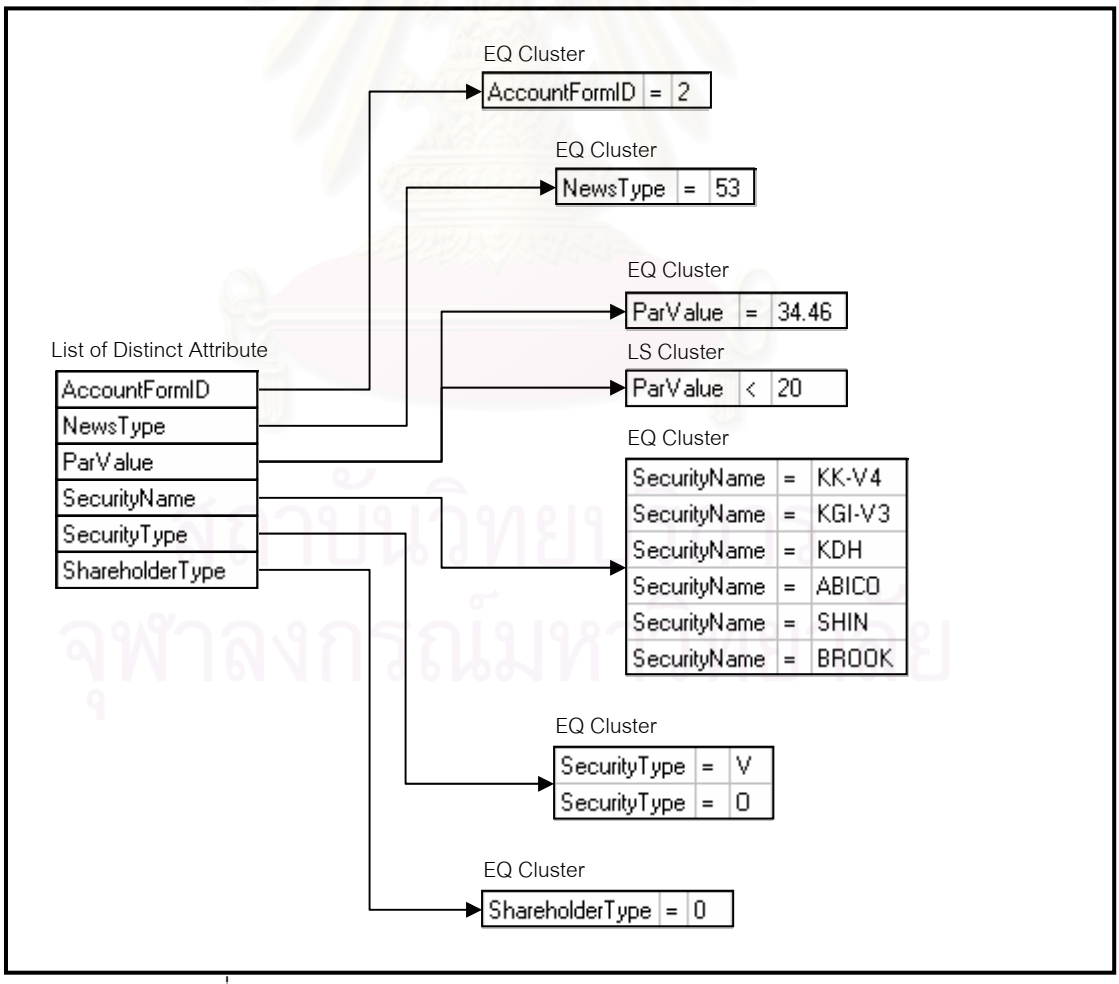
1) การจัดทำคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์

ในการสร้างคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ เป็นการสร้างคลัสเตอร์ของเพรดิเคตของแต่ละแอททริบิวท์ที่อยู่ภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ โดยในแต่ละแอททริบิวท์จะประกอบด้วยคลัสเตอร์ของเพรดิเคตที่สนใจในแอททริบิวท์นั้น 3 คลัสเตอร์ ได้แก่ คลัสเตอร์ของเพรดิเคตแบบเท่ากัน คลัสเตอร์ของเพรดิเคตแบบน้อยกว่า และคลัสเตอร์ของเพรดิเคตแบบมากกว่า โดยจะเก็บเพรดิเคตแบบเท่ากัน น้อยกว่า และมากกว่า ตามลำดับ ดังแสดงในรูปที่ 3.8




รูปที่ 3.8 คลัสเตอร์ของเพรดิเคตตามแอททริบิวท์

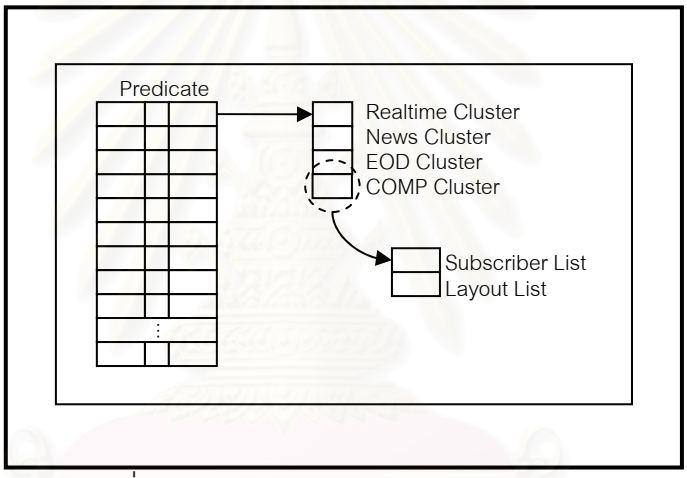
เมื่อนำบัญชีสมาชิกในรูปที่ 3.7 มาสร้างคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ จะได้คลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ ดังรูปที่ 3.9




รูปที่ 3.9 ตัวอย่างการจัดทำคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์

2) การจัดทำคลัสเตอร์ของสมาชิกตามเพรดิเคต

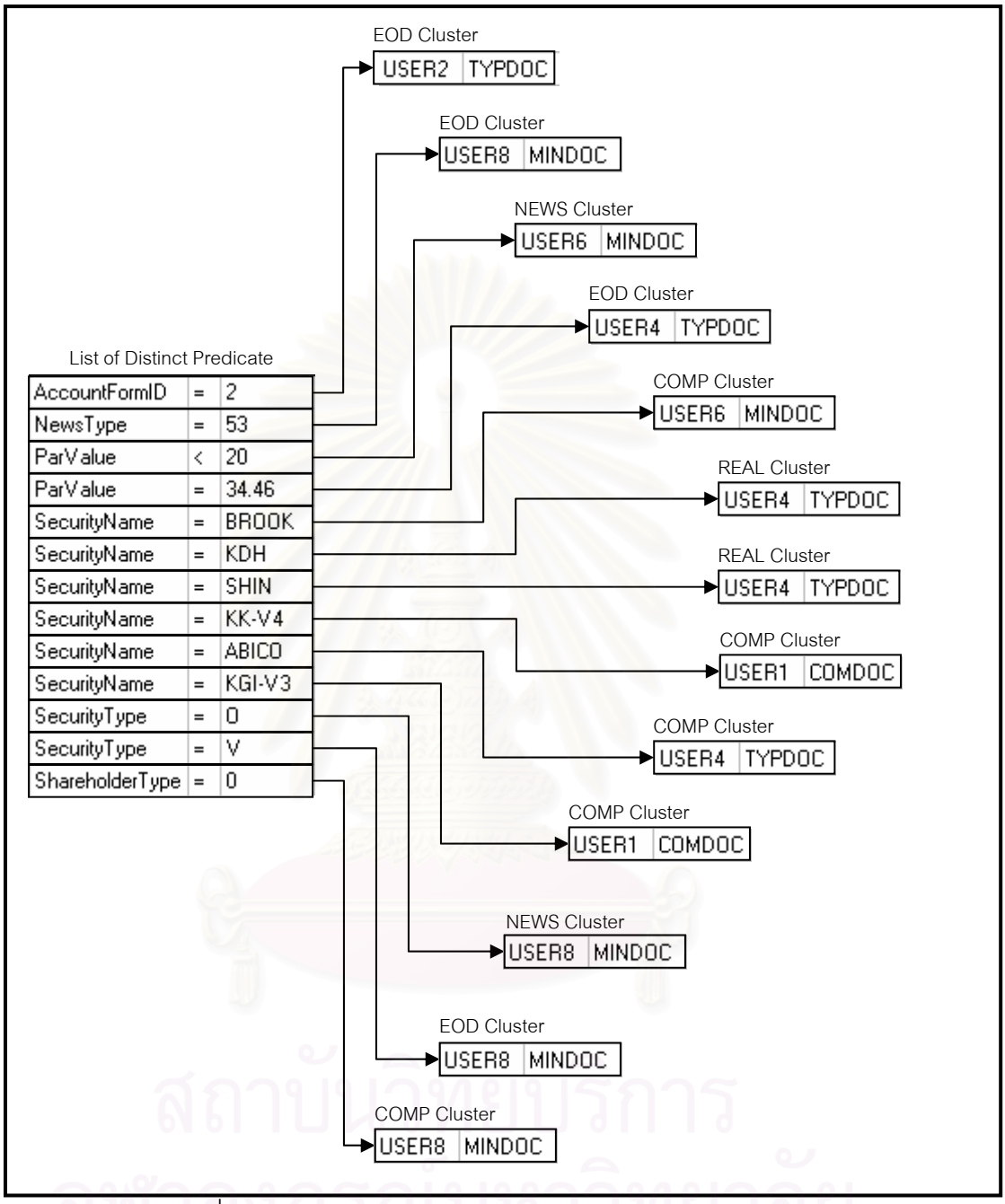
ในการสร้างคลัสเตอร์ของสมาชิกตามเพรดิเคต เป็นการสร้างคลัสเตอร์ของสมาชิกที่สนใจในเพรดิเคตเดียวกัน โดยในแต่ละเพรดิเคตจะประกอบไปด้วยคลัสเตอร์ 4 คลัสเตอร์ตามประเภทของหมวดข้อมูลสารสนเทศ (Information category) ซึ่งได้แก่ คลัสเตอร์ของหมวดข้อมูลเรียลไทม์ คลัสเตอร์ของหมวดข้อมูลข่าว คลัสเตอร์ของหมวดข้อมูลสรุปประจำวัน และคลัสเตอร์ของหมวดข้อมูลบริษัทจดทะเบียน ตามลำดับ ซึ่งภายในแต่ละคลัสเตอร์จะเก็บชื่อสมาชิกและรูปแบบของสารสนเทศที่ต้องการจะรับ ดังแสดงในรูปที่ 3.10




รูปที่ 3.10 คลัสเตอร์ของสมาชิกตามเพรดิเคต

และเมื่อนำบัญชีสมาชิกในรูปที่ 3.7 มาสร้างคลัสเตอร์ของสมาชิกตามเพรดิเคตจะได้คลัสเตอร์ของสมาชิกตามเพรดิเคต ดังรูปที่ 3.11

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย




รูปที่ 3.11 ตัวอย่างการจัดทำคลัสเตอร์ของสมาชิกตามเพรดิเคต

3.2.3.3 ขั้นตอนการทำงานของแมทชิ่งมอดูล

สำหรับการทำงานของแมทชิ่งมอดูล จะประกอบด้วยขั้นตอนทั้งหมด 4 ขั้นตอน ได้แก่

1) พิจารณาแต่ละแอททริบิวท์ ที่อยู่ภายในรายการแอททริบิวท์ภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ กับสกีมาของสารสนเทศที่เข้ามา โดยตรวจสอบว่าสกีมา

ของสารสนเทศมีค่าแอททริบิวท์ที่กำลังพิจารณาอยู่หรือไม่ ถ้าหากไม่มีจะเลื่อนไปพิจารณาแอททริบิวท์ตัวถัดไป

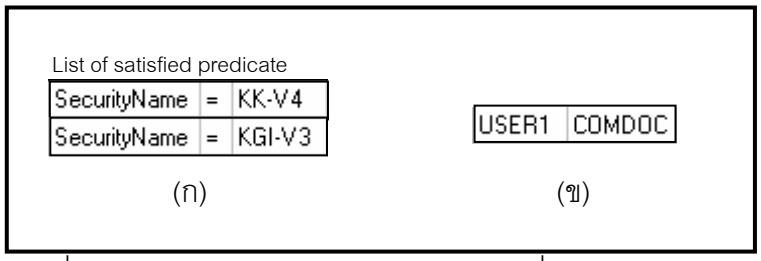
2) ถ้าสกีมาของสารสนเทศมีแอททริบิวท์ที่กำลังพิจารณา จะทำการตรวจสอบสารสนเทศนั้นกับคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ ทั้ง 3 คลัสเตอร์ เพื่อจัดทำรายการเพรดิเคตพึงใจ (List of satisfied predicate) หลังจากได้รายการเพรดิเคตพึงใจแล้ว จะนำเอารายการเพรดิเคตพึงใจที่ได้ไปหาบัญชีสมาชิกที่แมทช์กับ สารสนเทศที่เข้ามาจากคลัสเตอร์ของสมาชิกตามเพรดิเคตเพื่อหาสมาชิกที่ส่งเพรดิเคตนั้น

3) นำรหัสสมาชิกที่ได้จากขั้นตอนที่สองไปทำการตรวจสอบกับ โครงสร้างข้อมูล GenXMLUser เพื่อป้องกันการส่งสารสนเทศซ้ำซ้อนให้กับสมาชิก โดยทำการตรวจสอบว่ามีการทำเครื่องหมาย (Mark) ในโครงสร้างข้อมูล GenXMLUser ของสมาชิกนั้นหรือไม่ถ้ามีการทำเครื่องหมายว่าเคยแมทช์ไปแล้ว จะไม่ทำการสร้างรายการของสารสนเทศนั้นให้กับสมาชิกซ้ำอีก ซึ่งโครงสร้างข้อมูล GenXMLUser จะถูกล้างทุกครั้งเมื่อมีสารสนเทศใหม่เข้ามา

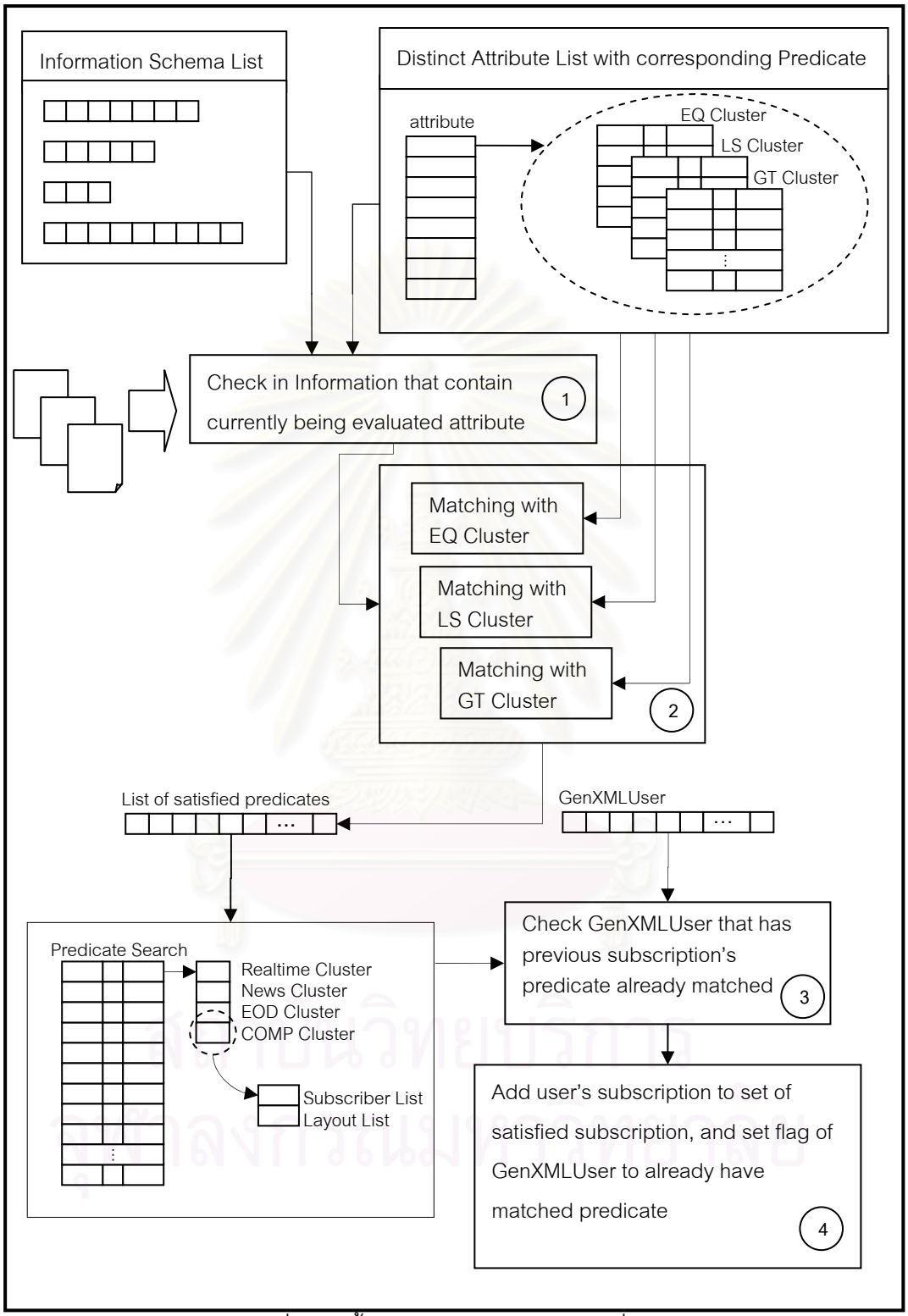
4) ในกรณีที่ยังไม่มีการทำเครื่องหมายในโครงสร้างข้อมูล GenXMLUser จะทำการเพิ่มรหัสสมาชิกนั้นเข้าไปในเซตของบัญชีสมาชิกที่แมทช์ (Set of matched subscriptions) และทำเครื่องหมายในโครงสร้างข้อมูล GenXMLUser ว่าเคยมีเพรดิเคตที่แมทช์ไปแล้ว

ซึ่งขั้นตอนการทำงานของแมทชิ่งมอดูล แสดงในรูปที่ 3.13

และเมื่อนำสกีมาของสารสนเทศในรูปที่ 3.6 มาตรวจสอบกับคลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ในรูปที่ 3.9 ตามขั้นตอนที่สองจะได้รายการของเพรดิเคตพึงใจ ดังรูปที่ 3.12(ก) และเมื่อนำรายการเพรดิเคตพึงใจไปหาสมาชิกที่ส่งเพรดิเคตนั้นจากคลัสเตอร์ของสมาชิกตามเพรดิเคต จะได้ดังรูปที่ 3.12(ข)




รูปที่ 3.12 รายการเพรดิเคตพึงใจและสมาชิกที่ส่งเพรดิเคตพึงใจ



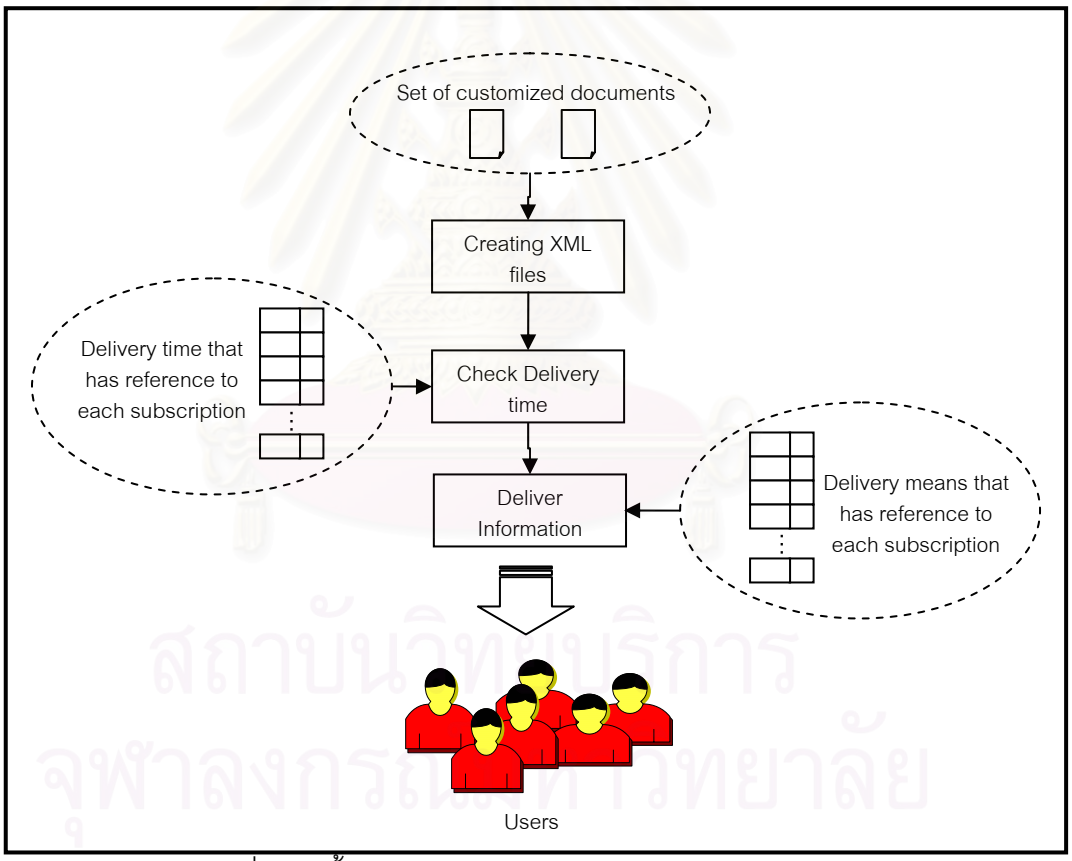

รูปที่ 3.13 ขั้นตอนการทำงานของแมทชิ่งมอดูล

3.2.4 ขั้นตอนการทำงานของตัวเผยแพร่สารสนเทศ

ในส่วนของขั้นตอนการทำงานของตัวเผยแพร่สารสนเทศจะแบ่งการทำงานออกเป็น 2 ส่วนได้แก่ การเผยแพร่สารสนเทศแบบไม่เรียลไทม์ และการเผยแพร่สารสนเทศแบบเรียลไทม์

3.2.4.1 การเผยแพร่สารสนเทศแบบไม่เรียลไทม์

สำหรับการเผยแพร่สารสนเทศแบบไม่เรียลไทม์ หลังจากได้สารสนเทศที่ผ่านการกรองแล้วตัวเผยแพร่สารสนเทศจะทำการสร้างเอกสารเอ็กซ์เอ็มแอล จากเซตของสารสนเทศที่ได้จากตัวกรองสารสนเทศ




รูปที่ 3.14 ขั้นตอนการเผยแพร่สารสนเทศแบบไม่เรียลไทม์

หลังจากนั้นตัวเผยแพร่สารสนเทศ จะตรวจสอบรูปแบบของการส่งสารสนเทศที่อยู่ภายในบัญชีสมาชิก โดยสำหรับสมาชิกที่ต้องการสารสนเทศแบบทันที (As soon as possible: ASAP) ตัวเผยแพร่สารสนเทศจะทำการส่งเอกสารเอ็กซ์เอ็มแอลให้กับสมาชิกทันที เมื่อสร้างเอกสารเอ็กซ์เอ็มแอลเสร็จตามวิธีการจัดส่งสารสนเทศที่อยู่ในบัญชีสมาชิก ส่วนสมาชิกที่ต้องการ

สารสนเทศแบบคาบเวลา (Time interval) ตัวเผยแพร่สารสนเทศจะคอยตรวจสอบเวลาระหว่างเวลาของระบบ กับเวลาที่สมาชิกต้องการจะรับสารสนเทศที่อยู่ภายในบัญชีสมาชิก ซึ่งถ้าหากถึงเวลาที่สมาชิกกำหนดระบบจึงจะส่งเอกสารเอ็กซ์เอ็มแอล ตามวิธีการจัดส่งที่ระบุภายในบัญชีสมาชิกให้กับสมาชิก โดยขั้นตอนของการส่งสารสนเทศแบบไม่เรียลไทม์ แสดงในรูปที่ 3.14 และตัวอย่างเอกสารเอ็กซ์เอ็มแอลที่สร้างจากการแมทชิ่งระหว่างสารสนเทศในรูปที่ 3.5 กับบัญชีสมาชิกของสมาชิกในรูปที่ 3.7 แสดงในรูปที่ 3.15

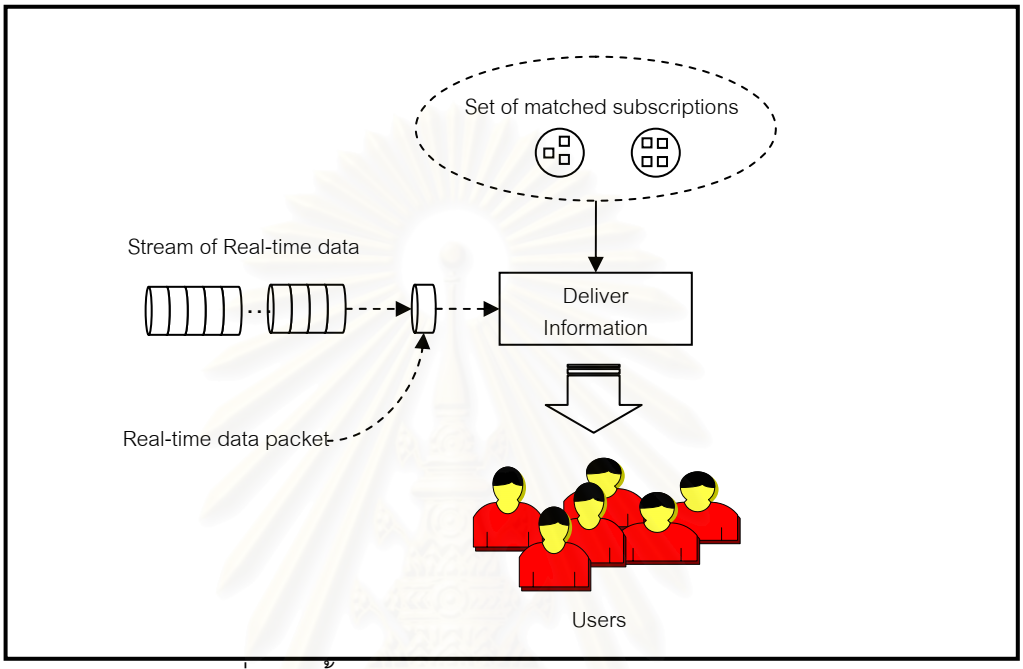
```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!DOCTYPE DeliveryData (View Source for full doctype...)>
- <DeliveryData>
  - <Category name="Listed Company Information">
    - <CompInfo type="Security Detail Information">
        <SecurityName>KGI-V3</SecurityName>
        <SecurityID>3711</SecurityID>
        <CompanyID>171</CompanyID>
        <SecurityType>F</SecurityType>
        <LocalSecurityName>KGI-W3</LocalSecurityName>
        <LocalSecurityID>1666</LocalSecurityID>
        <IsinNo>TH017105C618</IsinNo>
      </CompInfo>
    - <CompInfo type="Security Detail Information">
        <SecurityName>KK-V4</SecurityName>
        <SecurityID>3707</SecurityID>
        <CompanyID>121</CompanyID>
        <SecurityType>F</SecurityType>
        <LocalSecurityName>KK-W4</LocalSecurityName>
        <LocalSecurityID>2093</LocalSecurityID>
        <IsinNo>TH012105L119</IsinNo>
      </CompInfo>
    </Category>
  </DeliveryData>
```

รูปที่ 3.15 ตัวอย่างเอกสารเอ็กซ์เอ็มแอลก่อนส่งให้กับสมาชิก

3.2.4.2 การเผยแพร่สารสนเทศแบบเรียลไทม์

สำหรับการเผยแพร่สารสนเทศแบบเรียลไทม์ เนื่องจากสารสนเทศแบบเรียลไทม์ มีลักษณะของสารสนเทศที่วิกฤตเชิงเวลา (Time critical) นั่นคือเป็นลักษณะของสารสนเทศที่ถ้าหากสารสนเทศนั้นไปถึงสมาชิกช้ากว่าเวลาที่สมาชิกกำหนด จะเป็นสารสนเทศที่สมาชิกไม่ต้องการ เนื่องจากมีอัตราการเกิดของสารสนเทศแบบเรียลไทม์สูง และสารสนเทศควรจะไปถึงสมาชิกโดยเร็วที่สุด ดังนั้นถ้าหากมีการปรับปรุงสารสนเทศหลังจากที่ผ่านการกรองข้อมูลแล้ว

อาจจะทำให้สารสนเทศส่งไปถึงสมาชิกช้าเกินไป ดังนั้นทางผู้วิจัยจึงหลีกเลี่ยงการปรับปรุงรูปแบบสารสนเทศของสารสนเทศแบบเรียลไทม์ โดยสารสนเทศแบบเรียลไทม์ที่ผ่านการกรองแล้วจะถูกส่งไปให้กับสมาชิกที่ระบุทันที ดังรูปที่ 3.16




รูปที่ 3.16 ขั้นตอนการเผยแพร่สารสนเทศแบบเรียลไทม์

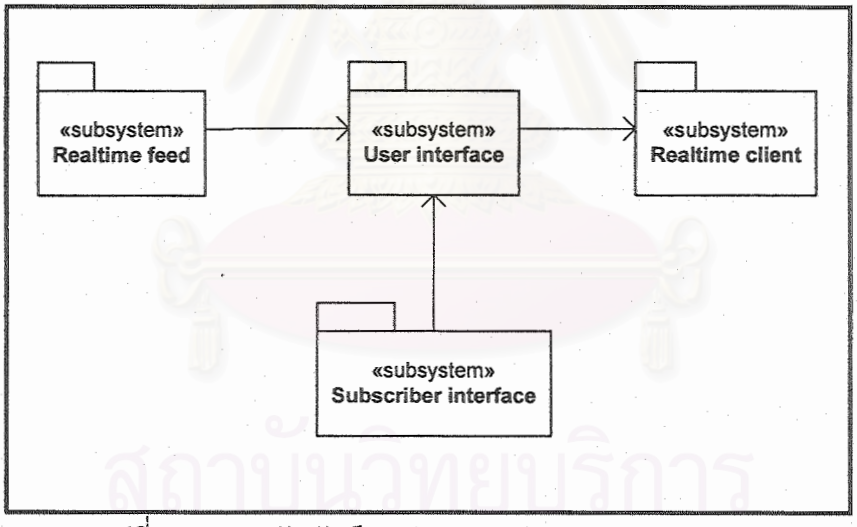
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 4
# รายละเอียดของระบบ

ในบทนี้จะกล่าวถึงรายละเอียดของระบบที่ใช้ในงานวิจัยตามที่ได้ออกแบบไว้ในบทที่ 3 ได้แก่ แผนภาพคลาสของระบบ แผนภาพแสดงความสัมพันธ์ระหว่างเอนทิตีของฐานข้อมูลของคำบรรยายลักษณะผู้ใช้ และอัลกอริธึมที่ใช้ในการทดสอบการวิจัย โดยจะทำการทดลองเปรียบเทียบอัลกอริธึมที่ออกแบบไว้ในบทที่ 3 กับอัลกอริธึมพื้นฐาน เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริธึมที่ออกแบบ และการจำลองระบบเพื่อทดสอบงานวิจัย

## 4.1 แผนภาพคลาสของการวิจัย

แผนภาพคลาสของระบบสามารถแบ่งออกเป็น 4 ระบบย่อย (Subsystem) ได้แก่ ระบบย่อยส่วนติดต่อผู้ใช้ (User interface subsystem) ระบบย่อยตัวป้อนสารสนเทศแบบเรียลไทม์ (Real-time feed subsystem) ระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ (Real-time client subsystem) และระบบย่อยส่วนติดต่อสมาชิก (Subscriber interface subsystem) ซึ่งสามารถแสดงความสัมพันธ์ระหว่างระบบย่อยต่าง ๆ ของระบบได้ ดังรูปที่ 4.1

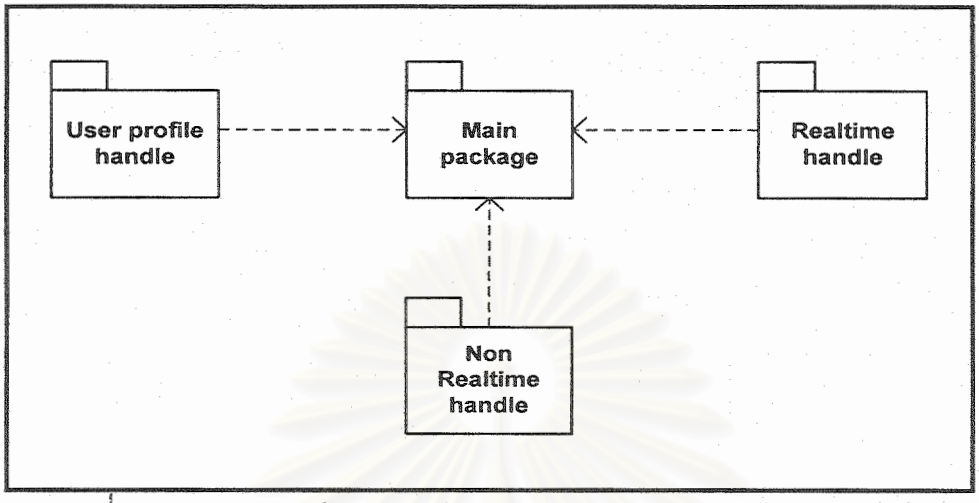



รูปที่ 4.1 ความสัมพันธ์ระหว่างระบบย่อยต่าง ๆ ของระบบ

### 4.1.1 ระบบย่อยส่วนติดต่อผู้ใช้

ระบบย่อยส่วนติดต่อผู้ใช้ เป็นส่วนที่ทำหน้าที่ติดต่อกับผู้ใช้งานระบบ ซึ่งเป็นระบบย่อยหลักของระบบ ประกอบไปด้วยคลาสทั้งหมด 74 คลาส โดยแบ่งออกเป็น 4 แพ็กเกจ (Package) ได้แก่ แพ็กเกจหลัก (Main package) แพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้ (User profile

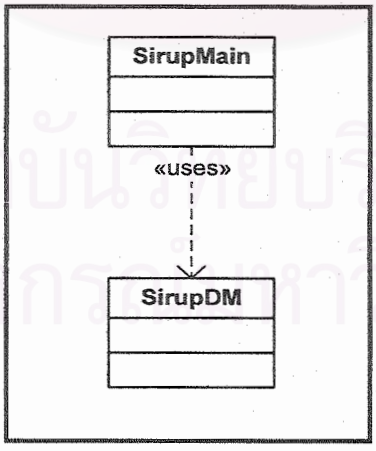
handle) แพ็กเกจการจัดการกับสารสนเทศแบบเรียลไทม์ (Realtime handle) และแพ็กเกจการจัดการกับสารสนเทศแบบไม่เรียลไทม์ (Non realtime handle) ดังรูปที่ 4.2




รูปที่ 4.2 ความสัมพันธ์ระหว่างแพ็กเกจต่าง ๆ ของระบบย่อยส่วนติดต่อผู้ใช้

4.1.1.1 แพ็กเกจหลัก

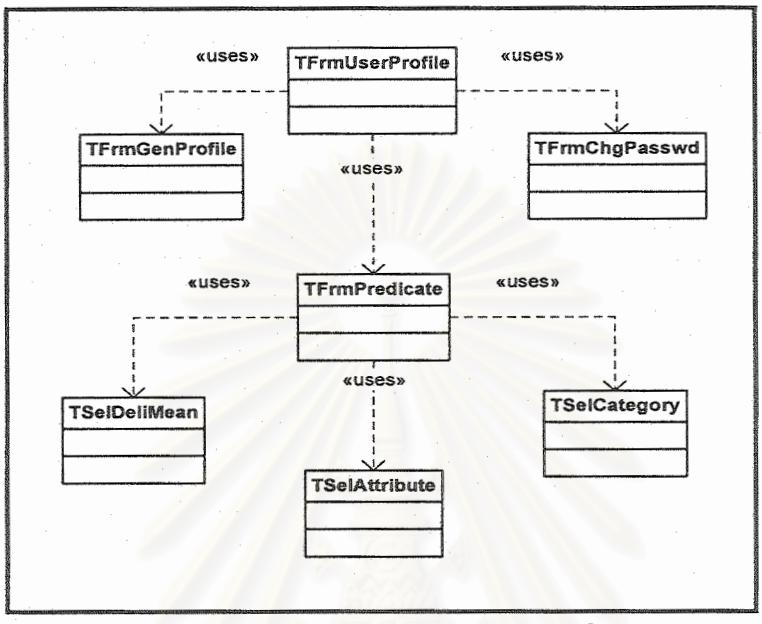
แพ็กเกจหลัก เป็นชุดของคลาสที่ทำหน้าที่หลักในระบบ ซึ่งประกอบไปด้วย คลาสซีรัพเมน (SirupMain) เป็นคลาสที่ทำหน้าที่จัดการในส่วนการติดต่อกระแสงานทั้งหมดของระบบ และเป็นส่วนติดต่อกับผู้ใช้งานระบบ และคลาสซีรัพดีเอ็ม (SirupDM) เป็นคลาสที่จัดการกับส่วนการติดต่อฐานข้อมูลของคำบรรยายลักษณะผู้ใช้ ซึ่งความสัมพันธ์ระหว่างคลาสในแพ็กเกจหลัก แสดงในรูปที่ 4.3




รูปที่ 4.3 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจหลัก

4.1.1.2 แพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้

แพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้ทำหน้าที่จัดการเกี่ยวกับการเพิ่ม แก้ไข และลบคำบรรยายลักษณะผู้ใช้ออกจากระบบ ซึ่งความสัมพันธ์ของคลาสแสดงในรูปที่ 4.4



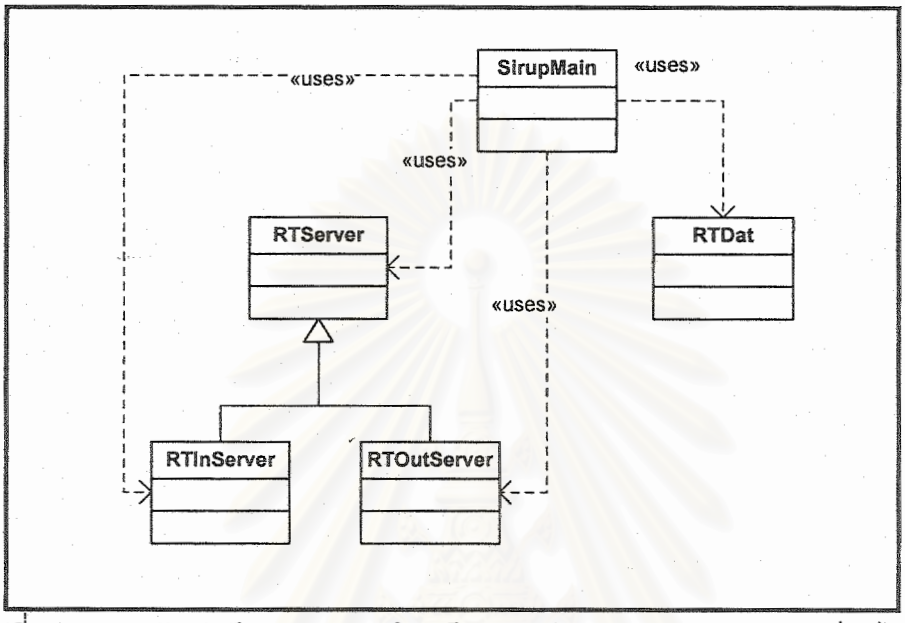

รูปที่ 4.4 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้

ซึ่งแพ็กเกจการจัดการคำบรรยายลักษณะผู้ใช้ ประกอบไปด้วยคลาส 7 คลาสได้แก่ คลาสจัดการคำบรรยายลักษณะผู้ใช้ (TFrmUserProfile) คลาสสำหรับสร้างคำบรรยายลักษณะผู้ใช้โดยอัตโนมัติ (TFrmGenProfile) คลาสสำหรับการแก้ไขรหัสผ่านของสมาชิก (TFrmChgPasswd) คลาสจัดการกับเพรดิเคต (TFrmPredicate) คลาสสำหรับการเลือกหมวดของสารสนเทศ (TSelCategory) คลาสสำหรับการเลือกวิธีการส่งสารสนเทศ (TSelDeliMean) และคลาสสำหรับการเลือกแอททริบิวท์ของสารสนเทศ (TSelAttribute)

4.1.1.3 แพ็กเกจการจัดการกับสารสนเทศแบบเรียลไทม์

แพ็กเกจการจัดการกับสารสนเทศแบบเรียลไทม์ ทำหน้าที่จัดการเกี่ยวกับการรับสารสนเทศแบบเรียลไทม์เพื่อส่งต่อให้กับคลาสซีรับเมน ที่อยู่ในแพ็กเกจหลัก และส่งสารสนเทศแบบเรียลไทม์ที่ผ่านการกรองแล้วให้กับระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ โดยจะมีคลาสทั้งหมด 4 คลาส ได้แก่ คลาสที่จัดการรูปแบบสารสนเทศแบบเรียลไทม์ (RTDat) ซึ่งจะถูกเรียกใช้ในคลาสซีรับเมน คลาสตัวบริการสารสนเทศแบบเรียลไทม์ (RTServer) คลาสตัวบริการ

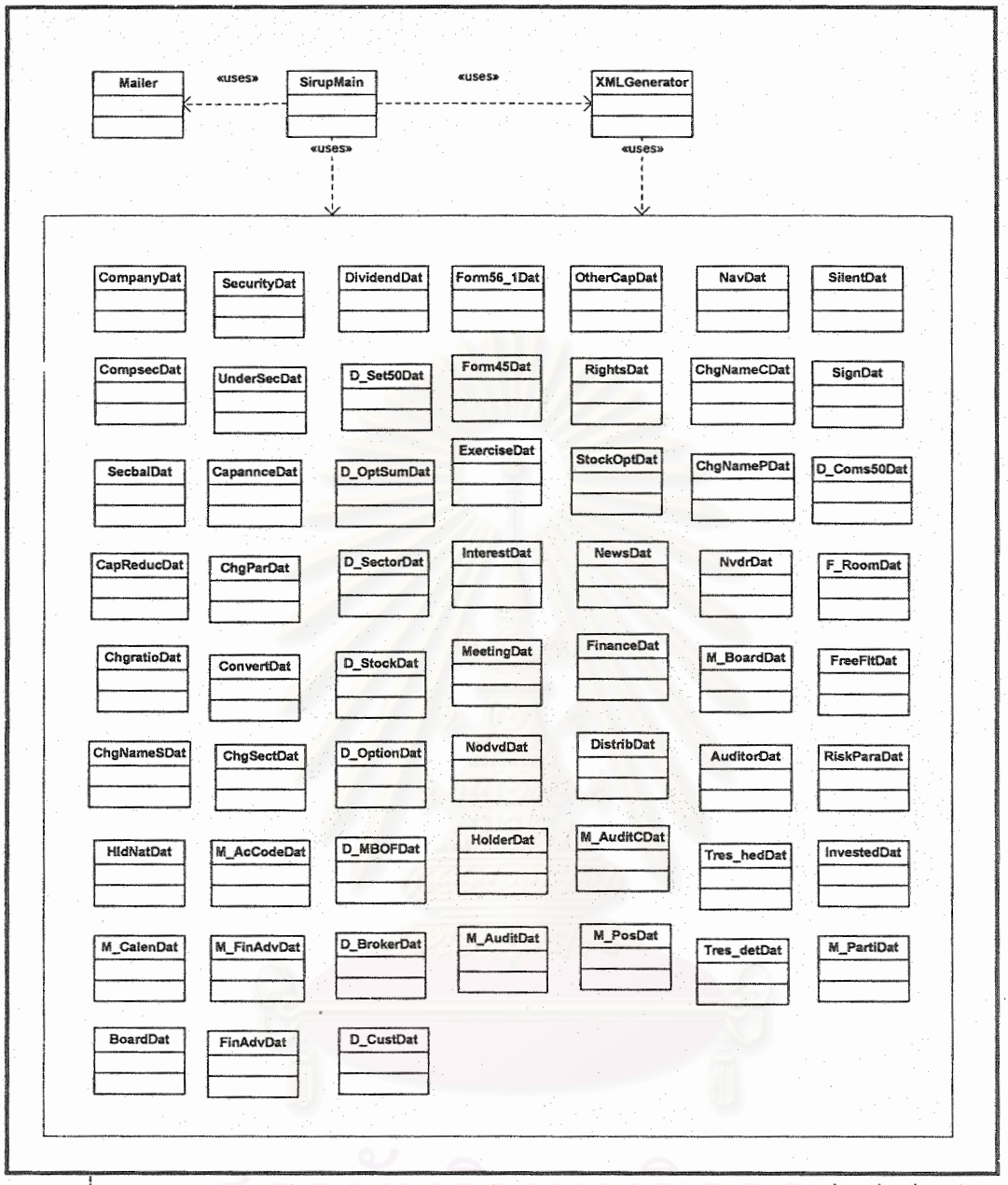
สารสนเทศขาเข้า (RTInServer) และคลาสตัวบริการสารสนเทศขาออก (RTOutServer) ซึ่งคลาสตัวบริการสารสนเทศขาเข้า และคลาสตัวบริการสารสนเทศขาออกจะสืบทอดคุณสมบัติ (Inherited) มาจากคลาสตัวบริการสารสนเทศแบบเรียลไทม์ ซึ่งความสัมพันธ์ของคลาสในแพ็กเกจนี้ แสดงในรูปที่ 4.5




รูปที่ 4.5 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการสารสนเทศแบบเรียลไทม์

4.1.1.4 แพ็กเกจการจัดการกับสารสนเทศแบบไม่เรียลไทม์

แพ็กเกจการจัดการสารสนเทศแบบไม่เรียลไทม์ ทำหน้าที่จัดการเกี่ยวกับการรับสารสนเทศ กรองสารสนเทศแบบไม่เรียลไทม์ การจัดทำการแสดงรูปแบบภายในของสารสนเทศแบบไม่เรียลไทม์ การปรับปรุงสารสนเทศให้ตรงกับรูปแบบโพรไฟล์ที่สมาชิกเลือก และการส่งสารสนเทศแบบไม่เรียลไทม์ให้กับสมาชิกโดยประกอบไปด้วยคลาสทั้งหมด 61 คลาส ดังแสดงในรูปที่ 4.6 (ยกเว้นคลาสซีรัพเมน) โดยคลาสตัวส่งจดหมายอิเล็กทรอนิกส์ (Mailer) จะทำหน้าที่ในการส่งเมล์ให้กับสมาชิก และคลาสสร้างเอกสารเอ็กซ์เอ็มแอล (XMLGenerator) จะทำหน้าที่สร้างเอกสารเอ็กซ์เอ็มแอล ส่วนคลาสที่เหลือจะเป็นการจัดการกับสารสนเทศแบบไม่เรียลไทม์ที่ส่งมาจากตลาดหลักทรัพย์ฯ



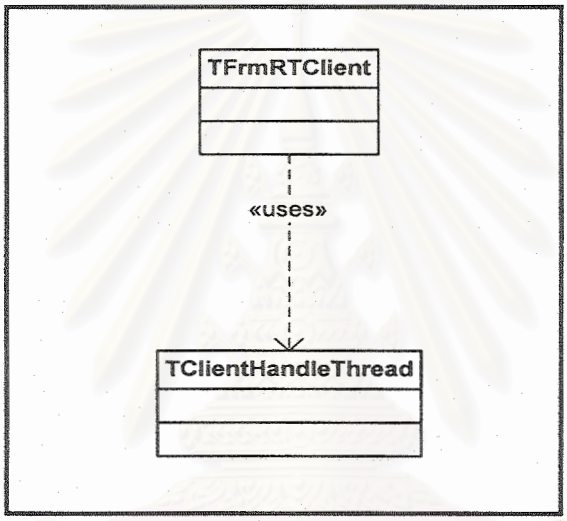

รูปที่ 4.6 ความสัมพันธ์ระหว่างคลาสในแพ็กเกจการจัดการสารสนเทศแบบไม่เรียลไทม์

### 4.1.2 ระบบย่อยตัวป้อนสารสนเทศแบบเรียลไทม์

ระบบย่อยตัวป้อนสารสนเทศแบบเรียลไทม์ เป็นส่วนที่ทำหน้าที่ป้อนสารสนเทศแบบเรียลไทม์ให้กับระบบย่อยส่วนติดต่อผู้ใช้ โดยประกอบด้วยคลาสตัวป้อนสารสนเทศแบบเรียลไทม์ (TFrmRTGen) ที่ทำหน้าที่ติดต่อกับคลาสตัวบริการขาเข้า ที่อยู่ภายในแพ็กเกจการจัดการสารสนเทศแบบเรียลไทม์

### 4.1.3 ระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์

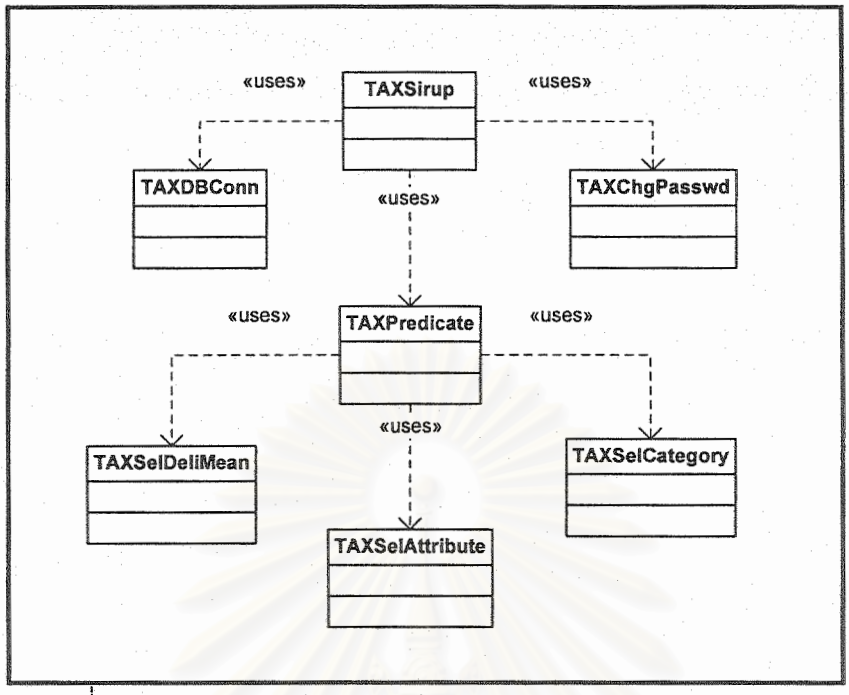
ระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ เป็นส่วนที่ทำหน้าที่รับสารสนเทศแบบเรียลไทม์ที่ผ่านการกรองแล้ว ซึ่งจะส่งออกมาจากระบบย่อยส่วนติดต่อผู้ใช้ โดยประกอบไปด้วยคลาสตัวรับบริการสารสนเทศแบบเรียลไทม์ (TFrmRTClient) ที่ทำหน้าที่เป็นตัวติดต่อกับ คลาสตัวบริการขาออก ซึ่งอยู่ภายในแพ็กเกจการจัดการสารสนเทศแบบเรียลไทม์ และคลาสที่ทำหน้าที่จัดการกับเธรดรับบริการ (TClientHandleThread) ซึ่งจะเป็นตัวรับช่วงการควบคุมการทำงานกับฝั่งตัวบริการขาออก เมื่อคลาสตัวรับบริการสารสนเทศแบบเรียลไทม์สามารถทำการติดต่อสำเร็จ โดยความสัมพันธ์ของคลาสภายในระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์ แสดงในรูปที่ 4.7




รูปที่ 4.7 ความสัมพันธ์ของคลาสในระบบย่อยตัวรับสารสนเทศแบบเรียลไทม์

### 4.1.4 ระบบย่อยส่วนติดต่อกับสมาชิก

ระบบย่อยส่วนติดต่อกับสมาชิก เป็นส่วนที่ทำหน้าที่ในการติดต่อกับสมาชิก ในกรณีที่สมาชิกต้องการเปลี่ยนแปลงข้อมูลคำบรรยายลักษณะผู้ใช้ โดยทำงานเป็นแอ็กทิฟเอ็กซ์ (ActiveX) และประกอบไปด้วยคลาสทั้งหมด 7 คลาส ได้แก่ คลาสส่วนติดต่อกับสมาชิก (TAXSirup) คลาสสำหรับติดต่อกับฐานข้อมูลคำบรรยายลักษณะผู้ใช้ (TAXDBConn) คลาสสำหรับการแก้ไขรหัสผ่านของสมาชิก (TAXChgPasswd) คลาสจัดการกับเพรดิเคต (TAXPredicate) คลาสสำหรับการเลือกหมวดของสารสนเทศ (TAXSelCategory) คลาสสำหรับการเลือกวิธีการส่งสารสนเทศ (TAXSelDeliMean) และคลาสสำหรับการเลือกแอททริบิวท์ของสารสนเทศ (TAXSelAttribute) ซึ่งความสัมพันธ์ของคลาสแสดงในรูปที่ 4.8



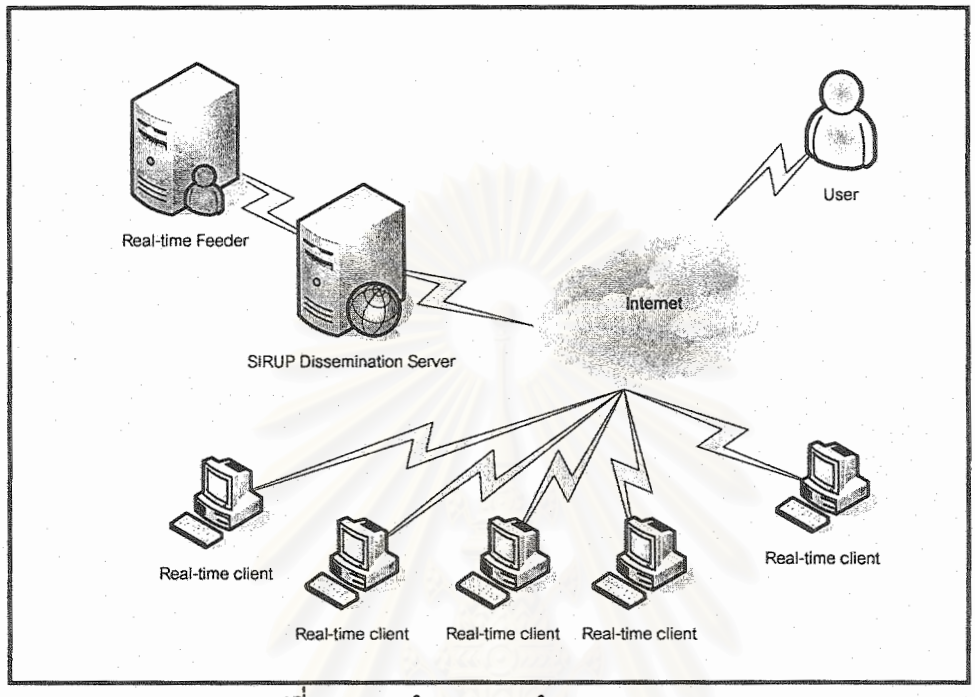

รูปที่ 4.8 ความสัมพันธ์ของคลาสในระบบย่อยส่วนติดต่อกับสมาชิก

## 4.2 การจำลองระบบเพื่อใช้ทดสอบการวิจัย

ระบบจำลองที่ใช้ทดสอบการวิจัย จะประกอบด้วยโปรแกรมประยุกต์ 4 โปรแกรมตามระบบย่อยของงานวิจัยที่ได้กล่าวไว้ในเรื่องแผนภาพคลาสของงานวิจัย ได้แก่ โปรแกรมประยุกต์ที่ใช้ในการติดต่อกับสมาชิก โปรแกรมประยุกต์สำหรับการป้อนสารสนเทศแบบเรียลไทม์ โปรแกรมประยุกต์สำหรับการกรองและเผยแพร่สารสนเทศ และโปรแกรมประยุกต์สำหรับการรับสารสนเทศแบบเรียลไทม์ จากรูปที่ 4.9 แสดงการจำลองการทำงานของระบบ ดังนี้

1) สมาชิกที่ต้องการรับสารสนเทศจากระบบ ทำการสมัครสมาชิกและสร้างคำบรรยายลักษณะผู้ใช้เก็บไว้ โดยใช้โปรแกรมประยุกต์สำหรับการติดต่อกับสมาชิกผ่านทางระบบอินเตอร์เน็ท
2) ตลาดหลักทรัพย์ฯ ส่งสารสนเทศแบบเรียลไทม์และสารสนเทศแบบไม่เรียลไทม์ให้กับระบบ โดยสำหรับสารสนเทศแบบเรียลไทม์จะส่งโดยใช้โปรแกรมประยุกต์สำหรับการป้อนสารสนเทศแบบเรียลไทม์ ส่วนสารสนเทศแบบไม่เรียลไทม์อาจส่งทางจดหมายอิเล็กทรอนิกส์ หรือการใช้งานตัวบริการเอฟทีพี (FTP server)
3) ระบบทำการกรองสารสนเทศโดยใช้ขั้นตอนการทำงานที่ออกแบบในบทที่ 3 พร้อมทั้งจัดทำรูปแบบของสารสนเทศ และทำการเผยแพร่สารสนเทศให้สมาชิก

4) สมาชิกใช้โปรแกรมประยุกต์สำหรับการรับสารสนเทศแบบเรียลไทม์เพื่อรับข้อมูลแบบเรียลไทม์ที่ระบบทำการเผยแพร่ โดยผ่านทางโพรโตคอลทีซีพีไอพี (TCP/IP protocol)




รูปที่ 4.9 การจำลองการทำงานของระบบ

## 4.3 ฐานข้อมูลคำบรรยายลักษณะผู้ใช้

จากการออกแบบคำบรรยายลักษณะผู้ใช้ในบทที่ 3 สามารถสร้างฐานข้อมูลของคำบรรยายลักษณะผู้ใช้โดยจะแบ่งประเภทของตารางเป็น 2 ประเภท ดังนี้

1) ตารางหลักของฐานข้อมูล ได้แก่ ตารางทีพีดีซีทำหน้าที่เก็บข้อมูลคำบรรยายลักษณะผู้ใช้ ตารางทียูเอ็นเอ็มทำหน้าที่เก็บข้อมูลรายละเอียดชื่อสมาชิก ตารางทียูเอดีทำหน้าที่เก็บข้อมูลรายละเอียดที่อยู่สมาชิก ตารางทีจีดีซีทำหน้าที่เก็บข้อมูลหมวดข้อมูลการรวบรวมข้อมูล ตารางทีดีดีซีทำหน้าที่เก็บข้อมูลหมวดข้อมูลการส่งข้อมูล

2) ตารางอ้างอิงของฐานข้อมูลได้แก่ ตารางอาร์ดีอีเอ็มทำหน้าที่เก็บข้อมูลวิธีการส่งสารสนเทศ ตารางอาร์ดีทีเอ็มทำหน้าที่เก็บข้อมูลรูปแบบการส่งสารสนเทศ ตารางอาร์แอลเอโอทำหน้าที่เก็บข้อมูลรูปแบบของสารสนเทศ ตารางอาร์ซีเอทีทำหน้าที่เก็บข้อมูลประเภทหมวดของสารสนเทศ ตารางอาร์เอทีทีทำหน้าที่เก็บ

ข้อมูลแอททริบิวท์ของสารสนเทศ ตารางอาร์ซีโอเอ็มพีทำหน้าที่เก็บข้อมูลรหัสบริษัทและชื่อหลักทรัพย์ และตารางอาร์เอสอีซีทำหน้าที่เก็บข้อมูลความสัมพันธ์ระหว่างชื่อหลักทรัพย์และบริษัท

| tpdc | | |
|---|---|---|
| PK | ProfileID | INTEGER |
| | **UserID** | **VARCHAR(10)** |
| | **Passwd** | **VARCHAR(100)** |

รูปที่ 4.10 ตารางทีพีดีซี

ตารางที่ 4.1 รายละเอียดตารางเก็บข้อมูลคำบรรยายลักษณะผู้ใช้

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| ProfileID | ตัวระบุคำบรรยายลักษณะผู้ใช้ | กำหนดโดยอัตโนมัติ ไม่ซ้ำกัน |
| UserID | รหัสสมาชิก | กำหนดโดยสมาชิก ไม่ซ้ำกัน |
| Passwd | รหัสผ่านของสมาชิก | กำหนดโดยสมาชิก |

| tunm | | |
|---|---|---|
| **PK,FK1** | **ProfileID** | **INTEGER** |
| | NamePrefix | VARCHAR(30) |
| | **FirstName** | **VARCHAR(100)** |
| | MiddleName | VARCHAR(10) |
| | **LastName** | **VARCHAR(200)** |
| | NameSuffix | VARCHAR(30) |

รูปที่ 4.11 ตารางทียูเอ็นเอ็ม

ตารางที่ 4.2 รายละเอียดตารางเก็บข้อมูลชื่อของสมาชิก

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| ProfileID | ตัวระบุคำบรรยายลักษณะผู้ใช้ | กำหนดจากตารางทีพีดีซี |
| NamePrefix | คำนำหน้าชื่อ | กำหนดโดยสมาชิก |
| FirstName | ชื่อต้น | กำหนดโดยสมาชิก |
| MiddleName | ชื่อกลาง | กำหนดโดยสมาชิก |
| LastName | ชื่อสกุล | กำหนดโดยสมาชิก |
| NameSuffix | คำต่อท้ายชื่อ | กำหนดโดยสมาชิก |

| tuad | | |
|---|---|---|
| **PK,FK1** | **ProfileID** | **INTEGER** |
| | Address | VARCHAR(100) |
| | Street | VARCHAR(50) |
| | City | VARCHAR(50) |
| | StateProv | VARCHAR(80) |
| | Province | VARCHAR(80) |
| | PostalCode | VARCHAR(8) |
| | Country | VARCHAR(50) |
| | Telephone | VARCHAR(50) |
| | Fax | VARCHAR(50) |
| | Mobile | VARCHAR(50) |
| | Pager | VARCHAR(50) |
| | Email | VARCHAR(100) |
| | Homepage | VARCHAR(50) |

รูปที่ 4.12 ตารางทียูเอดี

ตารางที่ 4.3 รายละเอียดตารางเก็บข้อมูลที่อยู่ของสมาชิก

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| ProfileID | ตัวระบุคำบรรยายลักษณะผู้ใช้ | กำหนดจากตารางทีพีดีซี |
| Address | เลขที่+ชื่ออาคาร+ชื่อหมู่บ้าน+ชื่อซอย | กำหนดโดยสมาชิก |
| Street | ชื่อถนน | กำหนดโดยสมาชิก |
| City | ชื่อแขวง หรือชื่อตำบล | กำหนดโดยสมาชิก |
| StateProv | ชื่อเขต หรือชื่ออำเภอ | กำหนดโดยสมาชิก |
| Province | ชื่อจังหวัด | กำหนดโดยสมาชิก |
| PostalCode | รหัสไปรษณีย์ | กำหนดโดยสมาชิก |
| Country | ชื่อประเทศ | กำหนดโดยสมาชิก |
| Telephone | หมายเลขโทรศัพท์ | กำหนดโดยสมาชิก |
| Fax | หมายเลขโทรสาร | กำหนดโดยสมาชิก |
| Mobile | หมายเลขโทรศัพท์มือถือ | กำหนดโดยสมาชิก |
| Pager | หมายเลขเพจเจอร์ | กำหนดโดยสมาชิก |
| Email | ที่อยู่จดหมายอิเล็กทรอนิกส์ | กำหนดโดยสมาชิก |
| Homepage | โฮมเพจ | กำหนดโดยสมาชิก |

| tgdc | | |
|---|---|---|
| PK | PredicateID | INTEGER |
| FK1 | ProfileID | INTEGER |
| FK2 | Category | VARCHAR(10) |
| FK3 | AttributeCD | VARCHAR(50) |
| | Condition | CHAR(2) |
| FK4,FK5 | Value | VARCHAR(30) |

รูปที่ 4.13 ตารางทีจีดีซี

ตารางที่ 4.4 รายละเอียดตารางเก็บข้อมูลหมวดข้อมูลการรวบรวมสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| PredicateID | ตัวระบุเพรดิเคต | กำหนดโดยอัตโนมัติ ไม่ซ้ำกัน |
| ProfileID | ตัวระบุคำบรรยายลักษณะ | กำหนดจากตารางทีพีดีซี |
| Category | หมวดหมู่ข้อมูลที่จะรับ | กำหนดจากตารางอาร์ซีเอที |
| AttributeCD | รหัสแอททริบิวท์ | กำหนดจากตารางอาร์เอเอที |
| Condition | ตัวดำเนินการเปรียบเทียบ | กำหนดโดยสมาชิก |
| Value | ค่าที่ใช้เปรียบเทียบ | กำหนดโดยสมาชิก |

| tddc | | |
|---|---|---|
| PK,FK5 | ProfileID | INTEGER |
| FK2 | rdemCD | VARCHAR(10) |
| FK4 | rlaoCD | VARCHAR(10) |
| FK3 | rdtmCD | VARCHAR(10) |
| | Destination | VARCHAR(255) |
| | DeliveryTime | VARCHAR(10) |

รูปที่ 4.14 ตารางทีดีดีซี

ตารางที่ 4.5 รายละเอียดตารางเก็บข้อมูลหมวดข้อมูลการส่งสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| ProfileID | ตัวระบุคำบรรยายลักษณะ | กำหนดจากตารางทีพีดีซี |
| rdemCD | รหัสวิธีการส่งสารสนเทศ | กำหนดจากตารางอาร์ดีอีเอ็ม |
| rlaoCD | รหัสรูปแบบของสารสนเทศ | กำหนดจากตารางอาร์แอลเอโอ |
| rdtmCD | รหัสรูปแบบการส่งสารสนเทศ | กำหนดจากตารางอาร์ดีทีเอ็ม |
| Destination | ที่อยู่ปลายทาง | กำหนดโดยสมาชิก |
| DeliveryTime | เวลาที่ส่งสารสนเทศ | กำหนดโดยสมาชิก |

| rcat | | |
|---|---|---|
| PK | CateCD | VARCHAR(10) |
| | Description | VARCHAR(255) |

รูปที่ 4.15 ตารางอาร์ซีเอที

ตารางที่ 4.6 รายละเอียดตารางเก็บข้อมูลประเภทหมวดของสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| CateCD | รหัสหมวดของสารสนเทศ | กำหนดโดยสมาชิก |
| Description | คำอธิบายหมวดของสารสนเทศ | กำหนดโดยสมาชิก |

| ratt | | |
|---|---|---|
| PK | AttributeCD | VARCHAR(10) |
| | Description | VARCHAR(255) |

รูปที่ 4.16 ตารางอาร์เอทีที

ตารางที่ 4.7 รายละเอียดตารางเก็บข้อมูลแอททริบิวท์ของสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| AttributeCD | รหัสแอททริบิวท์ | กำหนดโดยสมาชิก |
| Description | คำอธิบายแอททริบิวท์ | กำหนดโดยสมาชิก |

| rcomp | | |
|---|---|---|
| PK | CompanyId | INTEGER |
| | SecurityName | VARCHAR(12) |

รูปที่ 4.17 ตารางอาร์ซีโอเอ็มพี

ตารางที่ 4.8 รายละเอียดตารางเก็บข้อมูลรหัสบริษัทและชื่อหลักทรัพย์

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| CompanyID | รหัสบริษัท | กำหนดโดยสมาชิก |
| SecurityName | ชื่อหลักทรัพย์ | กำหนดโดยสมาชิก |

| rsec | | |
|---|---|---|
| PK | Security | VARCHAR(10) |
| | Company | VARCHAR(255) |

รูปที่ 4.18 ตารางอาร์เอสอีซี

ตารางที่ 4.9 รายละเอียดตารางเก็บข้อมูลความสัมพันธ์ระหว่างหลักทรัพย์และบริษัท

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| Security | ชื่อหลักทรัพย์ | กำหนดโดยสมาชิก |
| Company | ชื่อบริษัท | กำหนดโดยสมาชิก |

| rdem | | |
|---|---|---|
| PK | rdemCD | VARCHAR(10) |
| | description | VARCHAR(255) |

รูปที่ 4.19 ตารางอาร์ดีอีเอ็ม

ตารางที่ 4.10 รายละเอียดตารางเก็บข้อมูลวิธีการส่งสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| rdemCD | รหัสวิธีการส่งสารสนเทศ | กำหนดโดยสมาชิก |
| Description | คำอธิบายวิธีการส่งสารสนเทศ | กำหนดโดยสมาชิก |

| rdtm | | |
|---|---|---|
| PK | rdtmCD | VARCHAR(10) |
| | description | VARCHAR(255) |

รูปที่ 4.20 ตารางอาร์ดีทีเอ็ม

ตารางที่ 4.11 รายละเอียดตารางเก็บข้อมูลรูปแบบการส่งสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| rdtmCD | รหัสรูปแบบการส่งสารสนเทศ | กำหนดโดยสมาชิก |
| Description | คำอธิบายรูปแบบการส่งสารสนเทศ | กำหนดโดยสมาชิก |

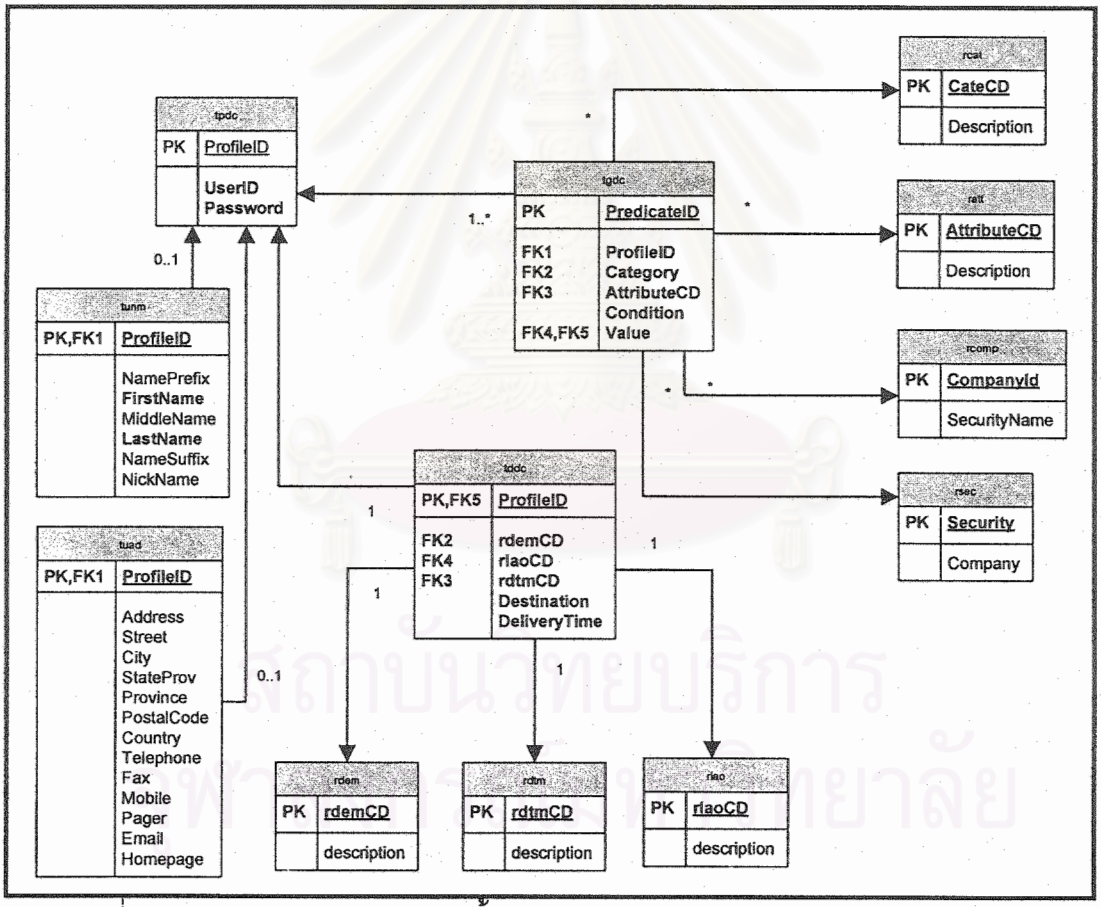
| rlao | | |
|---|---|---|
| PK | rlaoCD | VARCHAR(10) |
| | description | VARCHAR(255) |

รูปที่ 4.21 ตารางอาร์แอลเอโอ

ตารางที่ 4.12 รายละเอียดตารางเก็บข้อมูลรูปแบบของสารสนเทศ

| สดมภ์ | คำอธิบาย | หมายเหตุ |
|---|---|---|
| rlaoCD | รหัสรูปแบบของสารสนเทศ | กำหนดโดยสมาชิก |
| Description | คำอธิบายรูปแบบของสารสนเทศ | กำหนดโดยสมาชิก |

จากตารางที่เกี่ยวข้องทั้งหมดสามารถเขียนรูปของความสัมพันธ์ระหว่างตารางทั้งหมดภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ได้ ดังแสดงในรูปที่ 4.22




รูปที่ 4.22 ความสัมพันธ์ของตารางทั้งหมดภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้

## 4.4 อัลกอริธึมที่ใช้ในการทดสอบการวิจัย

### 4.4.1 อัลกอริธึมพื้นฐาน

สำหรับอัลกอริธึมพื้นฐานที่ใช้ในการทดสอบงานวิจัย จะมีลักษณะคล้ายคลึงกับอัลกอริธึมแบบนาอีฝ โดยจะทำการประเมินทุกเพรดิเคตของทุกบัญชีสมาชิกแบบหนึ่งต่อหนึ่ง กับสารสนเทศที่เข้ามา แต่จะแตกต่างจากอัลกอริธึมแบบนาอีฝ โดยจะมีการทดสอบเพิ่มเติมว่าแอททริบิวท์ของเพรดิเคตที่กำลังพิจารณามีอยู่ในรายการสกีมาของสารสนเทศหรือไม่ ซึ่งจะเป็นการลดจำนวนการวนซ้ำของแอททริบิวท์ที่ไม่เกี่ยวข้องออกไป ดังแสดงในรูปที่ 4.23

```
1:// r is the record in file to process. FileSchema is the schema of file that currently process
2:// Distinct_AttrList is the data structure that keep all distinct attribute in user profile
3:// p is the predicate. M is the set of matched predicates
4: M = {}
5: for every r do
6:     for every j in Distinct_AttrList do
7:         if FileSchema has Distinct_AttrList[j] then
8:             for every p in each subscription do
9:                 if (p.attribute = r.attribute) and (p.condition = r.condition) and
10:                  (p.value = r.value) then
11:                     insert p to M
12:                 endif
13:             endfor
14:         endif
15:     endfor
16: endfor
```

รูปที่ 4.23 อัลกอริธึมพื้นฐาน

### 4.4.2 อัลกอริธึมของระบบซีรัพ

สำหรับอัลกอริธึมที่ของระบบซีรัพ ดังแสดงในรูปที่ 4.24 จะใช้หลักการของการจัดทำคลัสเตอร์ของเพรดิเคต ในขั้นตอนของกระบวนการก่อนประมวลผล โดยการใช้ตัวดำเนินการเปรียบเทียบและแอททริบิวท์ในการจัดทำคลัสเตอร์เช่นเดียวกับในอัลกอริธึมแบบแฟร์เพรดิเคต โดยคลัสเตอร์แต่ละอันจะมีรายการเชื่อมโยงกับรายการของแอททริบิวท์ที่อยู่ภายในคำบรรยายลักษณะผู้ใช้ที่มีค่าแอททริบิวท์สอดคล้องกับแอททริบิวท์ของคลัสเตอร์นั้น

การทำงานของอัลกอริธึมจะเริ่มจาก การตรวจสอบว่าแอททริบิวท์ของเพรดิเคตที่กำลังพิจารณามีอยู่ในสกีมาของสารสนเทศหรือไม่ ซึ่งในการทำงานส่วนนี้จะมีลักษณะเดียวกันกับ

อัลกอริธึมแบบนีออน โดยถ้าหากมีแอททริบิวท์ที่กำลังพิจารณาอยู่ภายในสกีมาของสารสนเทศ จะทำการแมทช์แอททริบิวท์นั้นกับ คลัสเตอร์ของเพรดิเคตตามแอททริบิวท์ โดยฟังก์ชัน *EqualityMatch*(*p*,*r*) *LessMatch*(*p*,*r*) และ *GreaterMatch*(*p*,*r*) จะใช้ในการแมทช์ระหว่างเพรดิเคตกับเรคคอร์ดของสารสนเทศแบบเท่ากับ น้อยกว่า และมากกว่าตามลำดับ

```
1:// r is the record in file to process. FileSchema is the schema of file that currently process
2:// Distinct_AttrList is the data structure that keep all distinct attribute in user profile
3:// Distinct_PredList is the data structure that keep all distinct predicate in user profile
4:// Satisfied_Pred is the data structure that keep satisfied predicate
5:// p is the predicate. M is the set of matched predicates
6: M = {}
7: Satisfied_Pred = {}
8: for every r do
9:     for every j in Distinct_AttrList do
10:        if FileSchema has Distinct_AttrList[j] then
11:            for every p in Distinct_AttrList[j].EQ_Cluster then
12:                Satisfied_Pred = Satisfied_Pred ∪ EqualityMatch(p,r)
13:            endfor
14:            for every p in Distinct_AttrList[j].LS_Cluster then
15:                Satisfied_Pred = Satisfied_Pred ∪ LessMatch(p,r)
16:            endfor
17:            for every p in Distinct_AttrList[j].GT_Cluster then
18:                Satisfied_Pred = Satisfied_Pred ∪ GreaterMatch(p,r)
19:            endfor
20:        endif
21:    endfor
22: endfor
23: for every p in Satisfied_Pred do
24:    if p in Distinct_PredList then
25:        Get List of User that submit p
26:        for every user do
27:            if User[j] not have matched predicate then
28:                M = M ∪ User[j]
29:            endif
30:        endfor
31:    endif
32: endfor
```

รูปที่ 4.24 อัลกอริธึมของระบบซีรัพ

ฟังก์ชัน *EqualityMatch* ใช้ในการค้นหาเพรดิเคตที่มีค่าเดียวกับค่าแอททริบิวท์ ที่อยู่ภายในคลัสเตอร์แบบเท่ากัน โดยการใช้งานอัลกอริธึมในการค้นหาแบบไบนารีทำให้ได้ค่าความซับซ้อนเชิงเวลาเป็น O ($\sum_{a \in d.atts}(\log (\#D_i)+1)$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ฟังก์ชัน *LessMatch* และ *GreaterMatch* เป็นการค้นหาเพรดิเคตที่ตรงกับเรคคอร์ดของสารสนเทศแบบน้อยกว่าและมากกกว่าตามลำดับ โดยในการค้นหาถ้าหากพบเพรดิเคตที่แมทช์กับเรคคอร์ดของสารสนเทศแล้ว จะใช้หลักการนิรนัยในการลดจำนวนเพรดิเคตที่จะต้องทำการประเมินลง

หลังจากที่ได้รายการเพรดิเคตพึงใจ จะทำการอ่านข้อมูลของสมาชิกที่ระบุเพรดิเคตตรงกับเพรดิเคตที่อยู่ในรายการเพรดิเคตพึงใจ แล้วทำการตรวจสอบสมาชิกว่าเคยมีเพรดิเคตที่แมทช์ไปก่อนหน้านี้หรือไม่ ถ้าไม่มีจะทำการเพิ่มสมาชิกเข้าไปในเซตของบัญชีสมาชิกที่แมทช์กับสารสนเทศที่เข้ามา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 5
# ผลการวิจัย

ในบทนี้จะกล่าวถึงการทดสอบการวิจัย รวมถึงการเปรียบเทียบผลการทดลองที่ได้ และสรุปถึงปัจจัยที่มีผลกระทบต่อประสิทธิภาพการทำงานของระบบ และค่าที่เหมาะสมสำหรับปัจจัยแต่ละตัว

## 5.1 สภาพแวดล้อมที่ใช้ในการทดสอบการวิจัย

ในการวิจัยใช้คอมพิวเตอร์แบบพีซี Pentium III 1 กิกะเฮิรตซ์ หน่วยความจำ 2 กิกะไบต์ ฮาร์ดดิส 20 กิกะไบต์ และใช้ระบบปฏิบัติการไมโครซอฟท์วินโดวส์เอ็กซ์พี โดยมีตัวแปรภาระงานและค่าที่ใช้ในการทดลองแสดงในตารางที่ 5.1

ตารางที่ 5.1 รายละเอียดตัวแปรภาระงานและค่าที่เป็นไปได้

| คำอธิบาย | ค่าที่เป็นไปได้ |
|---|---|
| จำนวนข้อมูลแบบไม่เรียลไทม์ | 247,020 เรคคอร์ด |
| จำนวนข้อมูลแบบเรียลไทม์จำลอง | 70,000 แมสเซจ |
| จำนวนบัญชีสมาชิก | 10,000 ถึง 40,000 |
| จำนวนแอททริบิวท์ต่อการเผยแพร่ | 4 ถึง 51 |
| จำนวนแอททริบิวท์ที่ใช้กรองข้อมูล | 29 |
| จำนวนของเพรดิเคตต่อบัญชีสมาชิก | 3 |
| จำนวนเพรดิเคตแบบเท่ากันต่อบัญชีสมาชิก | 1-3 |
| จำนวนเพรดิเคตแบบไม่เท่ากันต่อบัญชีสมาชิก | 1-3 |

## 5.2 ข้อมูลที่ใช้ทดสอบการวิจัย

ข้อมูลที่ใช้ในการทดสอบ แบ่งออกเป็นสองประเภทได้แก่ ข้อมูลแบบเรียลไทม์ และข้อมูลแบบไม่เรียลไทม์ โดยสำหรับข้อมูลแบบเรียลไทม์เป็นข้อมูลจำลองที่สร้างจากข้อมูลจริงของตลาดหลักทรัพย์แห่งประเทศไทย และข้อมูลแบบไม่เรียลไทม์จะเป็นข้อมูลจริงจากตลาดหลักทรัพย์แห่งประเทศไทย โดยเป็นข้อมูลในช่วงวันที่ 3 มิถุนายน พ.ศ. 2546 วันที่ 28 กรกฎาคม พ.ศ. 2546 ข้อมูลช่วงวันที่ 18-20 และ 26 ของเดือนสิงหาคม พ.ศ. 2546 และข้อมูลช่วงเดือนกันยายน พ.ศ. 2546 ในวันที่ตลาดหลักทรัพย์แห่งประเทศไทยเปิดทำการ

## 5.3 การทดสอบการวิจัย

สำหรับการทดสอบการวิจัยได้ทำการทดสอบด้านต่าง ๆ ดังนี้

5.3.1 ทดสอบประสิทธิภาพโดยการเพิ่มปริมาณของจำนวนบัญชีสมาชิก ที่อยู่ภายในคำบรรยายลักษณะผู้ใช้ เพื่อทดสอบว่าเมื่อจำนวนของคำบรรยายลักษณะผู้ใช้เพิ่มมากขึ้นมีผลกระทบกับประสิทธิภาพของระบบอย่างไร

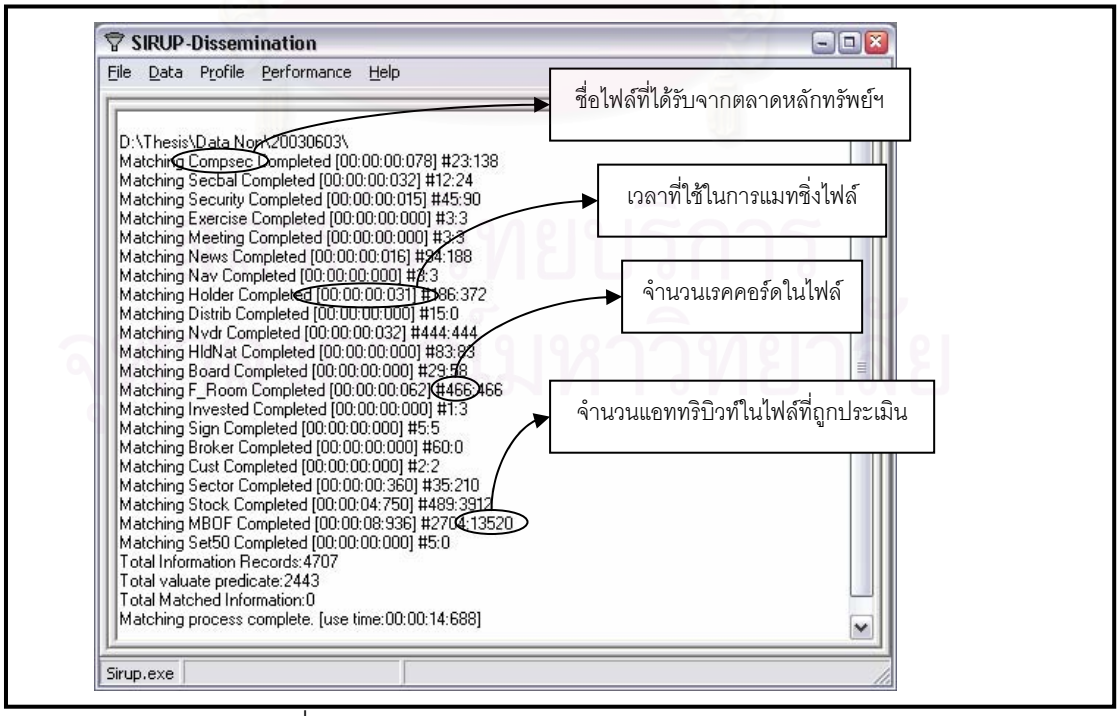
5.3.2 ทดสอบประสิทธิภาพโดยการเพิ่มจำนวนสารสนเทศ เพื่อทดสอบว่าอัตราการเพิ่มขึ้นของปริมาณสารสนเทศ มีผลกระทบกับประสิทธิภาพของระบบอย่างไร

5.3.3 ทดสอบประสิทธิภาพโดยการเพิ่มจำนวนไคลเอนท์ (Client) เพื่อทดสอบว่าเมื่อมีจำนวนผู้รับบริการเพิ่มมากขึ้น มีผลกระทบกับประสิทธิภาพของระบบอย่างไร

## 5.4 วิธีประเมินผลการทดสอบการวิจัย

5.4.1 การประเมินผลประสิทธิภาพของการเผยแพร่สารสนเทศแบบไม่เรียลไทม์

สำหรับการประเมินผลประสิทธิภาพของการแมทชิ่งสารสนเทศแบบไม่เรียลไทม์ จะทำการจับเวลาจากการแมทชิ่งจำนวนเรคคอร์ดที่อยู่ในไฟล์ที่ได้รับจากตลาดหลักทรัพย์ดังรูปที่ 5.1




รูปที่ 5.1 การจับเวลาของสารสนเทศแบบไม่เรียลไทม์

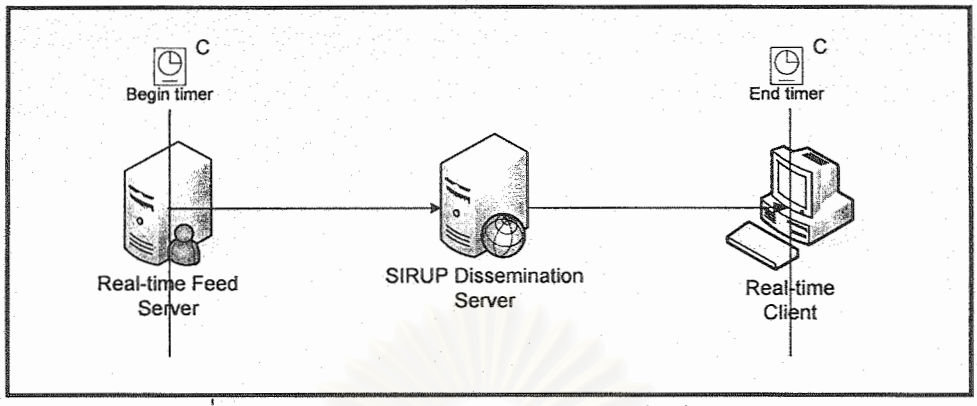
หลังจากนั้นนำผลที่ได้ไปคำนวณหาค่าเวลาเฉลี่ยที่ใช้ในการแมทชิ่งหนึ่งแอททริบิวท์ดังรูปที่ 5.2

| 20030603 | Data files from SET | Hour | Min | Sec | MS | Total Rec | No. of Eval Attr | Total MS (Minx60000) +(Secx1000) +MS | MS/Rec Total MS / Total Rec | Attr/Rec No. of Eval Attr / Total Rec | MS/Attr (MS/Rec) / (Attr/Rec) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Compsec | 0 | 0 | 0 | 360 | 23 | 138 | 360 | 15.652174 | 6 | 2.60869565 |
| | Secbal | 0 | 0 | 0 | 156 | 12 | 24 | 156 | 13 | 2 | 6.5 |
| | Security | 0 | 0 | 0 | 0 | 45 | 90 | 0 | 0 | 2 | 0 |
| | Exercise | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 1 | 0 |
| | Meeting | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 1 | 0 |
| | News | 0 | 0 | 0 | 16 | 94 | 188 | 16 | 0.1702128 | 2 | 0.08510638 |
| | Nav | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 1 | 0 |
| | Holder | 0 | 0 | 0 | 79 | 186 | 372 | 79 | 0.4247312 | 2 | 0.21236559 |
| | Distrib | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 |
| | Nvdr | 0 | 0 | 0 | 62 | 444 | 444 | 62 | 0.1396396 | 1 | 0.13963964 |
| | HldNat | 0 | 0 | 0 | 16 | 83 | 83 | 16 | 0.1927711 | 1 | 0.19277108 |
| | Board | 0 | 0 | 0 | 15 | 29 | 58 | 15 | 0.5172414 | 2 | 0.25862069 |
| | F_Room | 0 | 0 | 0 | 110 | 466 | 466 | 110 | 0.2360515 | 1 | 0.2360515 |
| | Invested | 0 | 0 | 0 | 0 | 1 | 3 | 0 | 0 | 3 | 0 |
| | Sign | 0 | 0 | 0 | 0 | 5 | 5 | 0 | 0 | 1 | 0 |
| | Broker | 0 | 0 | 0 | 0 | 60 | 0 | 0 | 0 | 0 | 0 |
| | Cust | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 0 |
| | Sector | 0 | 0 | 0 | 609 | 35 | 210 | 609 | 17.4 | 6 | 2.9 |
| | Stock | 0 | 0 | 16 | 641 | 489 | 3912 | 16641 | 34.030675 | 8 | 4.25383436 |
| | MBOF | 0 | 0 | 24 | 347 | 2704 | 13520 | 24347 | 9.004068 | 5 | 1.80081361 |
| | Set50 | 0 | 0 | 0 | 16 | 5 | 0 | 16 | 3.2 | 0 | 0 |
| Total Information Records | | 4707 | | | | | | | | | |
| | | | | | | | | | Average time | | 1.74435441 |

รูปที่ 5.2 การคำนวณหาค่าเวลาเฉลี่ยที่ใช้ในการแมทชิ่งหนึ่งแอททริบิวท์

5.4.2 การประเมินผลประสิทธิภาพของการเผยแพร่สารสนเทศแบบเรียลไทม์จำลอง

สำหรับการวัดประสิทธิภาพของการเผยแพร่สารสนเทศแบบเรียลไทม์จำลอง จะทำการจับเวลาหลังจากที่โปรแกรมด้านการป้อนสารสนเทศแบบเรียลไทม์ ทำการส่งข้อมูลเป็นจุดเริ่มต้นและจับเวลาอีกครั้ง เมื่อโปรแกรมด้านการรับสารสนเทศแบบเรียลไทม์ได้รับข้อมูลเป็นจุดสิ้นสุด หลังจากนั้นจึงนำเวลาที่ได้มาทำการเปรียบเทียบกันว่า ในการส่งสารสนเทศแบบเรียลไทม์จำลองแต่ละแมสเซจใช้เวลาเท่าไร หลังจากนั้นจึงทำการหาค่าเฉลี่ยของแมสเซจที่สมาชิกแต่ละคนได้รับแล้วนำค่าเฉลี่ยที่ได้ไปทำการหาค่าเฉลี่ยรวมต่อสมาชิกอีกครั้ง แล้วนำค่าเฉลี่ยรวมต่อสมาชิกแต่ละคนไปทำการหาค่าเฉลี่ยรวมของสมาชิกทุกคน เพื่อวัดประสิทธิภาพของระบบจากเวลาที่ใช้ในการเผยแพร่สารสนเทศแบบเรียลไทม์จำลอง ซึ่งเวลาที่ใช้ของการเผยแพร่สารสนเทศแบบเรียลไทม์จำลองจะประกอบด้วย เวลาที่ใช้ในการอ่านสกีมาของสารสนเทศแบบเรียลไทม์จำลองเข้าไปในหน่วยความจำ หรือเวลาของกระบวนการก่อนการประมวลผล (Pre-processing time) เวลาที่ใช้ในการแมทชิ่งบัญชีสมาชิกกับสารสนเทศ (Matching time) และเวลาที่ใช้ในการเผยแพร่สารสนเทศไปให้กับสมาชิก (Dissemination time) โดยในการทดลองจะลดจำนวนของเวลาที่ใช้ในการเผยแพร่สารสนเทศให้เหลือน้อยที่สุด เพื่อวัดประสิทธิภาพของอัลกอริธึมที่ออกแบบได้ใกล้เคียงกับเวลาที่ใช้จริงมากที่สุด ดังรูปที่ 5.3



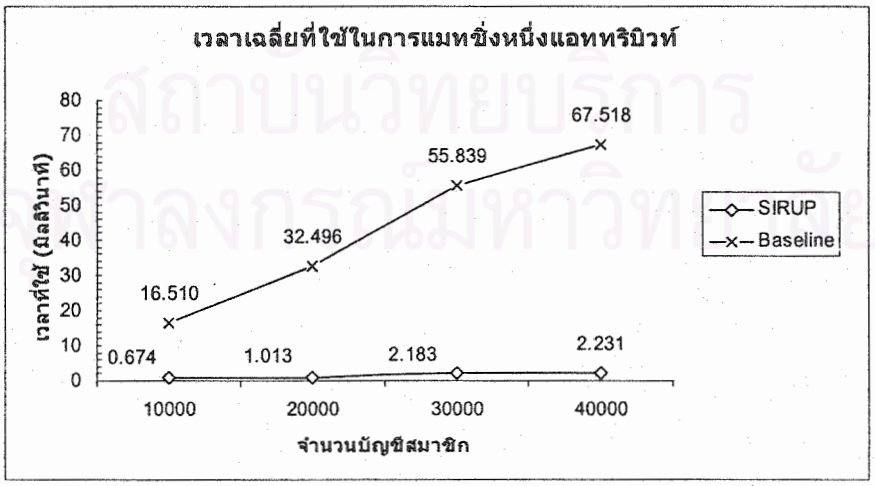

รูปที่ 5.3 การจับเวลาของสารสนเทศแบบเรียลไทม์จำลอง

## 5.5 ผลการทดสอบการวิจัย

### 5.5.1 การทดสอบประสิทธิภาพโดยการเพิ่มปริมาณของจำนวนบัญชีสมาชิก

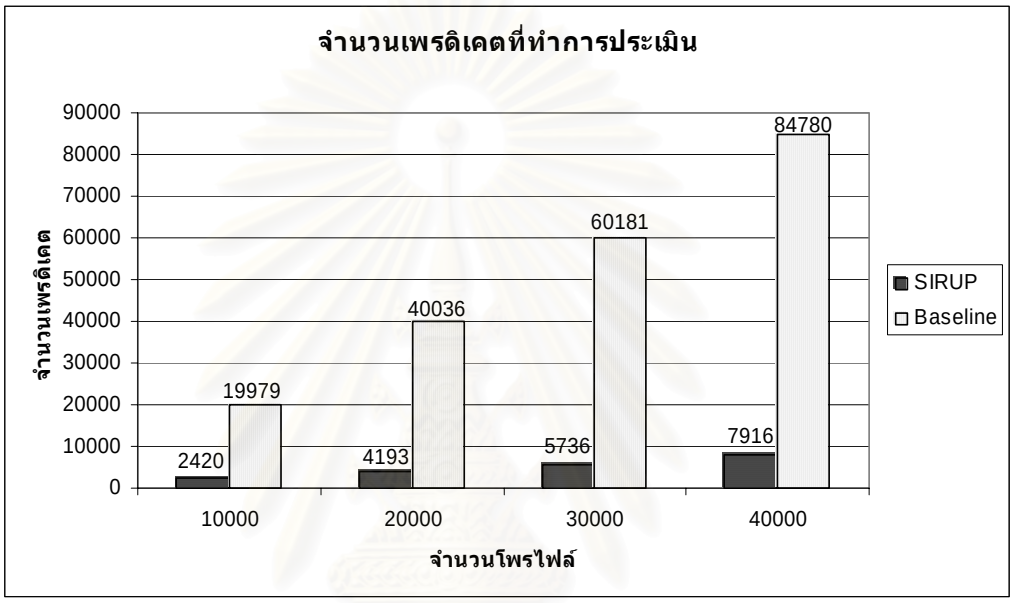
การทดลองนี้ ทำการทดสอบกับสารสนเทศแบบไม่เรียลไทม์ทั้งหมดจำนวน 247,020 รายการจากไฟล์อักขระทั้งหมด 37 ไฟล์ กับจำนวนบัญชีสมาชิกที่มีจำนวน 10,000 20,000 30,000 และ 40,000 บัญชีสมาชิก ตามลำดับ โดยในการทดลองจะทำการประเมินประสิทธิภาพของอัลกอริธึมที่ออกแบบ โดยการเปรียบเทียบกับอัลกอริธึมพื้นฐาน

จากการทดสอบโดยใช้อัลกอริธึมพื้นฐาน พบว่าเมื่อเพิ่มปริมาณจำนวนบัญชีสมาชิกของคำบรรยายลักษณะผู้ใช้ มีผลทำให้เวลาที่ใช้เพิ่มมากขึ้น สำหรับการทดสอบกับอัลกอริธึมที่ได้ออกแบบ พบว่าจำนวนบัญชีสมาชิกที่เพิ่มขึ้นมีผลกระทบต่อประสิทธิภาพการทำงานของระบบน้อยมาก ดังรูปที่ 5.4




รูปที่ 5.4 เวลาเฉลี่ยที่ใช้ในการแมทช์หนึ่งแอททริบิวท์

จากรูปที่ 5.4 เนื่องจากมีอัลกอริธึมที่ออกแบบมีการประเมินเพรดิเคตแบบไม่ซ้ำกันเพียงครั้งเดียว ดังนั้นเมื่อเปรียบเทียบจำนวนของเพรดิเคตที่ทำการประเมินแล้วพบว่า อัลกอริธึมที่ออกแบบมีการประเมินเพรดิเคตในจำนวนที่น้อยกว่า เนื่องจากความซ้ำซ้อนของเพรดิเคตที่อยู่ภายในคำบรรยายลักษณะผู้ใช้ ดังรูปที่ 5.5 ทำให้การเพิ่มปริมาณของบัญชีสมาชิกมีผลกระทบต่อประสิทธิภาพของอัลกอริธึมที่ออกแบบไม่มากนัก



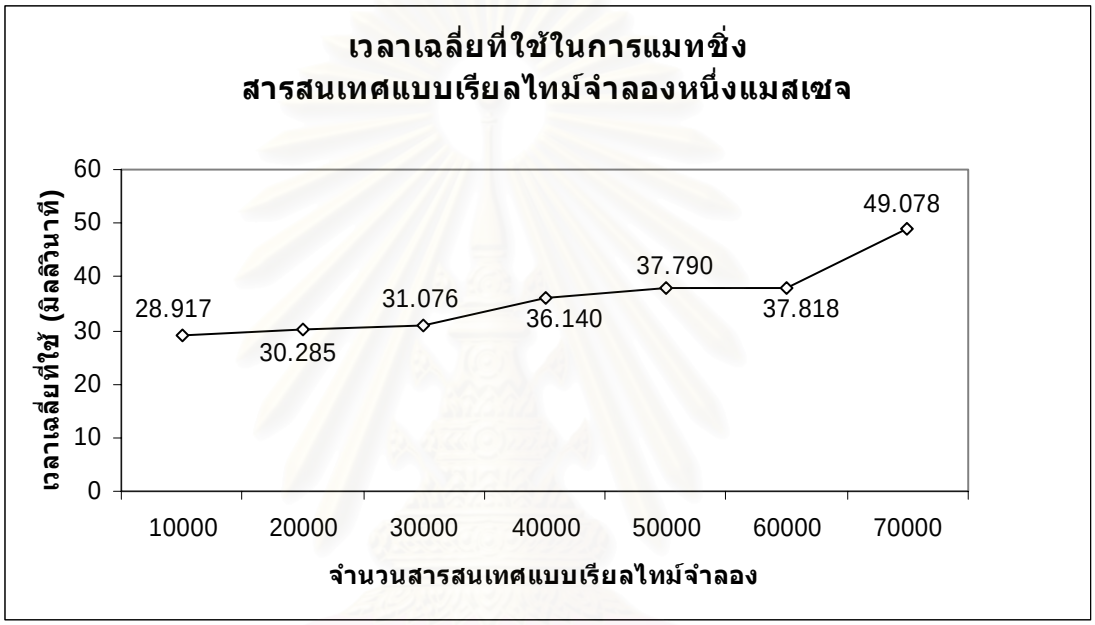

รูปที่ 5.5 เปรียบเทียบจำนวนเพรดิเคตที่ถูกประเมิน

5.5.2 ทดสอบประสิทธิภาพโดยการเพิ่มจำนวนสารสนเทศแบบเรียลไทม์จำลอง

การทดลองนี้กำหนดให้มีไคลเอนต์ที่เชื่อมต่ออยู่กับระบบจำนวน 180 ไคลเอนต์ และมีจำนวนบัญชีสมาชิกภายในฐานข้อมูลคำบรรยายลักษณะผู้ใช้ 10,000 บัญชีสมาชิก โดยมีการส่งสารสนเทศแบบเรียลไทม์จำลองจากระบบไปยังไคลเอนต์จำนวน 10,000 20,000 30,000 40,000 50,000 60,000 และ 70,000 แมสเซจตามลำดับ ซึ่งจะทำการหาค่าเฉลี่ยของเวลาที่ใช้ในการแมทชิ่งเพรดิเคตกับสารสนเทศ โดยทำการจับเวลาเมื่อมีการส่งสารสนเทศมาจากโปรแกรมประยุกต์สำหรับการป้อนสารสนเทศแบบเรียลไทม์มาให้ระบบ เมื่อระบบทำการแมทช์กับเพรดิเคตที่อยู่ภายในคำบรรยายลักษณะผู้ใช้ และส่งสารสนเทศต่อไปให้กับสมาชิก แล้วจึงทำการจับเวลาเมื่อสมาชิกได้รับสารสนเทศ หลังจากนั้นจะนำเวลาทั้งสองมาเปรียบเทียบหาค่าเวลาที่ใช้ในการเผยแพร่สารสนเทศแบบเรียลไทม์จำลอง และผลลัพธ์ที่ได้จะเป็นผลรวมของ เวลาในการอ่านสกีมาของสารสนเทศแบบเรียลไทม์จำลอง เวลาในการหาเซตของบัญชีสมาชิกที่แมทช์กับสารสนเทศ และเวลาที่ใช้ในการเผยแพร่สารสนเทศ ทั้งนี้การทำการทดลองตั้งอยู่บนสมมติฐานที่ว่าค่า

ความหน่วงของเครือข่าย (Network delay) มีค่าเท่ากับศูนย์ เพื่อที่จะทำการวัดประสิทธิภาพของอัลกอริธึมที่ออกแบบโดยไม่คำนึงถึงลักษณะของเครือข่าย

จากการทดลองพบว่าปริมาณสารสนเทศที่เพิ่มขึ้น มีผลกระทบต่อระบบเพียงเล็กน้อย เนื่องจากอัตราส่วนของสารสนเทศที่เพิ่มขึ้นกับจำนวนเวลาที่ใช้เพิ่มขึ้นมีไม่มากนัก ดังรูปที่ 5.6

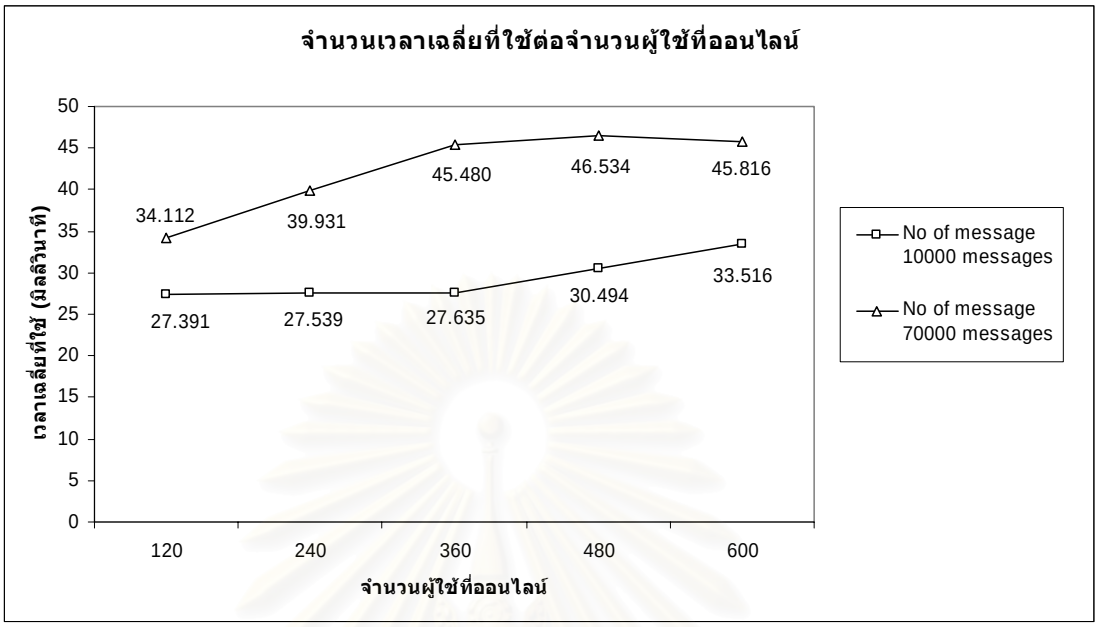



รูปที่ 5.6 เวลาเฉลี่ยที่ใช้ในการแมทชิ่ง เมื่อเพิ่มจำนวนแมสเซจ

5.5.3 ทดสอบประสิทธิภาพโดยการเพิ่มจำนวนไคลเอนท์

การทดลองนี้จะทดลองโดยมีพารามิเตอร์ที่คงที่ได้แก่จำนวนบัญชีสมาชิก โดยให้มีค่าเป็น 10,000 บัญชีสมาชิก และทำการทดสอบกับปริมาณของเรียลไทม์แมสเซจที่จำนวน 10,000 แมสเซจ และ 70,000 แมสเซจ เพื่อทดสอบประสิทธิภาพของอัลกอริธึม

จากการทดลองโดยเพิ่มจำนวนไคลเอนท์ที่เชื่อมต่อกับระบบพร้อม ๆ กัน เมื่อจำนวนไคลเอนท์เพิ่มขึ้น ระบบต้องจัดการกับรายการของไคลเอนท์มากขึ้น ดังนั้นเวลาที่ใช้ในการแมทชิ่งจึงเพิ่มขึ้น แต่อยู่ในลักษณะเชิงเส้นที่มีความชันต่ำ จึงทำให้จำนวนไคลเอนท์ที่เพิ่มขึ้นมีผลกระทบกับระบบไม่มากนัก ดังรูปที่ 5.7




รูปที่ 5.7 เวลาเฉลี่ยที่ใช้ในการแมทชิ่ง เมื่อเพิ่มจำนวนไคลเอนท์

## 5.6 ผลการวิเคราะห์ปัจจัย

จากการทดลองพบว่าปัจจัยที่มีผลต่อระบบได้แก่

### 5.6.1 ความซ้ำซ้อนของเพรดิเคต

เมื่อความซ้ำซ้อนของเพรดิเคตมีค่ามาก จะทำให้จำนวนของเพรดิเคตที่จะต้องทำการประเมินลดลง ทำให้ใช้เวลาในการแมทช์มีค่าลดลง และถึงแม้จะมีการเพิ่มขึ้นของจำนวนเพรดิเคตที่ไม่ซ้ำกัน แต่อัลกอริธึมที่ออกแบบจะมีแนวโน้มการใช้เวลาในการแมทช์ลดลง

### 5.6.2 ปริมาณของเพรดิเคตแบบเท่ากัน

เนื่องจากอัลกอริธึมมีการใช้งานการค้นหาแบบไบนารี จะทำให้เวลาที่ใช้ในการค้นหาว่าแอททริบิวท์ที่พิจารณาแมทช์กับเพรดิเคตหรือไม่ มีค่าความซับซ้อนเชิงเวลาอยู่ในรูปของฟังก์ชันลอการิธึม ทำให้ผลการทำงานของอัลกอริธึมมีประสิทธิภาพที่ดีขึ้น และการกำหนดเพรดิเคตในลักษณะของเพรดิเคตแบบเท่ากัน จะทำให้ระบบสามารถกรองข้อมูลได้ตรงตามความต้องการของสมาชิกมากขึ้น

5.6.3 ปริมาณของเพรดิเคตแบบไม่เท่ากัน

สำหรับปริมาณของเพรดิเคตแบบไม่เท่ากัน มีผลกระทบกับระบบคือ ในกรณีที่เพรดิเคตแบบไม่เท่ากันไม่สามารถแมทช์กับสารสนเทศที่เข้ามาได้ ระบบจะเสียเวลาในการประเมินเพรดิเคตแบบไม่เท่ากันทั้งหมด แต่ถ้าหากสามารถระบุได้ว่าเพรดิเคตแบบไม่เท่ากันแมทช์กับสารสนเทศที่พิจารณาได้เร็ว จะทำให้ระบบลดจำนวนของเพรดิเคตแบบไม่เท่ากันที่ต้องทำการประเมินลง โดยใช้หลักการนิรนัย

5.6.4 อัตราการเกิดของสารสนเทศและปริมาณของสารสนเทศ

สำหรับปัจจัยทางด้านอัตราการเกิดของสารสนเทศ จะมีผลกระทบกับระบบคือในกรณีสารสนเทศมีอัตราการเกิดจำนวนมากในเวลาหนึ่งวินาที ระบบอาจไม่สามารถแมทช์สารสนเทศที่เข้ามากับเพรดิเคตได้ทันเวลาทำให้ไม่สามารถส่งสารสนเทศที่ตรงกับความต้องการของสมาชิกได้ทันเวลา เช่นเดียวกับปัจจัยทางด้านปริมาณของสารสนเทศ เมื่อปริมาณของสารสนเทศมีจำนวนมาก จะมีผลกระทบต่อระบบคือระบบจะต้องใช้เวลาในการแมทชิ่งมากขึ้น อาจทำให้ไม่สามารถส่งสารสนเทศให้กับสมาชิกได้ทันเวลา

จากปัจจัยที่ได้กล่าวมาแล้ว พบว่าในการพัฒนาระบบกรองสารสนเทศสำหรับปริมาณข้อมูลจำนวนมาก และมีอัตราการเกิดสูง มักจะมีแอททริบิวท์ที่ถูกใช้เป็นเงื่อนไขในการกรองข้อมูลบ่อยครั้ง ทำให้ความซ้ำซ้อนของเพรดิเคตมีสูง และปริมาณของเพรดิเคตแบบเท่ากันจะมีจำนวนมากกว่า ปริมาณของเพรดิเคตแบบไม่เท่ากัน โดยค่าที่เหมาะสมของแต่ละปัจจัยสามารถสรุปได้ดังตารางที่ 5.2

ตารางที่ 5.2 ค่าที่เหมาะสมของปัจจัยที่มีผลต่อระบบ

| ปัจจัย | ลักษณะ / ความต้องการ |
|---|---|
| ความซ้ำซ้อนของเพรดิเคต | มาก |
| ปริมาณของเพรดิเคตแบบเท่ากัน | มาก |
| ปริมาณของเพรดิเคตแบบไม่เท่ากัน | ปานกลาง |
| อัตราการเกิดและปริมาณของสารสนเทศ | มาก |

# บทที่ 6
# สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากการทดสอบในบทที่ 5 พบว่าการพัฒนาระบบการเผยแพร่สารสนเทศทางด้านหลักทรัพย์โดยใช้คำบรรยายลักษณะผู้ใช้ในการกรองสารสนเทศนั้น จะต้องสามารถให้สมาชิกระบุถึงความสนใจในสารสนเทศที่ต้องการจะรับได้อย่างชัดเจนให้มากที่สุด เนื่องจากการระบุความต้องการที่ชัดเจนจะทำให้ระบบสามารถกรองสารสนเทศได้อย่างมีประสิทธิภาพ และใช้เวลาลดลง นอกจากนี้ การโต้ตอบแบบพับลิช/สับสไครบ์โดยพิจารณาจากเนื้อหาเป็นรูปแบบการติดต่อที่เหมาะสมทั้งนี้เนื่องจากลักษณะของการทำงานที่มีลักษณะผูกกันเชิงอิสระ (Loosely coupled) และการทำงานที่เป็นแบบอะซิงโครนัส ซึ่งจะเหมาะกับสารสนเทศเทศด้านหลักทรัพย์ เนื่องจากอัตราการเกิดของสารสนเทศมีสูง และสารสนเทศมีการเปลี่ยนแปลงบ่อย จึงเหมาะสมที่จะมีการเก็บความต้องการของสารสนเทศของสมาชิกในระยะยาว (Long-term interest) นั่นคือเป็นความต้องการที่ไม่มีการเปลี่ยนแปลงบ่อย เพื่อหลีกเลี่ยงการที่จะต้องคอยส่งคำร้องขอไปยังเซิร์ฟเวอร์ซึ่งเป็นการใช้งานทรัพยากรทางด้านเครือข่ายที่ไม่จำเป็น

## 6.1 สรุปผลการวิจัย

จากการออกแบบงานวิจัยและผลการทดลองพบว่า

6.1.1 การแยกองค์ประกอบของบัญชีสมาชิกออกเป็นเพรดิเคต ช่วยเพิ่มประสิทธิภาพในการแมทชิ่งระหว่าง แอททริบิวท์ของสารสนเทศกับแอททริบิวท์ของเพรดิเคต เนื่องจากความสนใจของสมาชิกที่รับสารสนเทศทางด้านหลักทรัพย์ มักจะมีความคล้ายคลึงกัน ทำให้ในแต่ละบัญชีสมาชิกอาจมีเพรดิเคตที่เหมือนกันอยู่ ดังนั้นเมื่อทำการแยกองค์ประกอบออกมาเป็นเพรดิเคต ทำให้สามารถทำการแมทช์ได้อย่างมีประสิทธิภาพมากขึ้น

6.1.2 การจัดทำสกีมาของสารสนเทศและดัชนีของคำบรรยายลักษณะผู้ใช้ มีส่วนช่วยให้การแมทช์ใช้เวลาลดลง ถึงแม้ว่าอาจต้องเสียเวลาให้กับกระบวนการก่อนการประมวลผลเพิ่มมากขึ้น แต่จะทำให้เวลาการทำงานโดยรวมของระบบลดลง

6.1.3 ความชัดเจนของความสามารถในการระบุความต้องการของสมาชิก เป็นส่วนสำคัญสำหรับความยืดหยุ่นของระบบ และการกรองสารสนเทศโดยพิจารณาจากเนื้อหาเป็นการกรองข้อมูลที่สามารถให้สมาชิกระบุความชัดเจนของความต้องการ

ได้มากที่สุด เนื่องจากยอมให้มีการกำหนดลักษณะของความต้องการจากเนื้อหาของข้อมูลที่ส่งออกมา

## 6.2 ปัญหาและข้อจำกัดที่พบจากการวิจัย

6.2.1 โปรแกรมเดลไฟ (Delphi) ที่ใช้ในการพัฒนามีการใช้งานบอร์แลนดาต้าเบสเอนจิน (Borland Database Engine: BDE) ในการติดต่อกับฐานข้อมูล ซึ่งถ้าหากเครื่องไคลเอนต์มีความต้องการที่จะเข้าถึงฐานข้อมูล จะต้องมีการติดตั้งบอร์แลนดาต้าเบสเอนจินที่เครื่องไคลเอนต์ด้วย

6.2.2 ฐานข้อมูลมายเอสคิวแอล (Mysql) มีข้อจำกัด ทางด้านการประมวลผลด้วยรายการเปลี่ยนแปลง (Transaction processing) ซึ่งยังไม่สามารถจัดการกับฐานข้อมูลในลักษณะของรายการเปลี่ยนแปลงได้

6.2.3 สารสนเทศแบบเรียลไทม์เป็นสารสนเทศที่จำลองขึ้น เนื่องจากนโยบายของสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ (กลต.) จะไม่ทำการเผยแพร่สารสนเทศแบบเรียลไทม์ให้กับบุคคลทั่วไป ดังนั้นทางผู้วิจัยจึงได้ทำการจำลองสารสนเทศแบบเรียลไทม์ โดยศึกษาจากเอกสารทางด้านโพรโตคอลที่ตลาดหลักทรัพย์ส่งสารสนเทศให้กับผู้เกี่ยวข้อง

6.2.4 สารสนเทศแบบไม่เรียลไทม์ที่ได้จากตลาดหลักทรัพย์บางไฟล์ มีค่าแอททริบิวท์บางค่าเป็นค่าว่าง ทำให้ไม่สามารถใช้แอททริบิวท์นั้นเพื่อใช้ในการกรองสารสนเทศได้

## 6.3 ข้อเสนอแนะในการพัฒนาต่อ

6.3.1 ในการวิจัยไม่ได้คำนึงถึงปัญหาทางด้านความปลอดภัยในการส่งสารสนเทศ ซึ่งเป็นประเด็นที่มีความสำคัญ โดยจะต้องมีการพิจารณาถึงหลักการยืนยันของความปลอดภัยของสารสนเทศและ สิทธิส่วนบุคคลของสมาชิก

6.3.2 การเพิ่มความซับซ้อนของคำบรรยายลักษณะผู้ใช้ เพื่อให้สมาชิกสามารถระบุถึงความต้องการได้ชัดเจนยิ่งขึ้น โดยการเพิ่มคำเชื่อมระหว่างเพรดิเคต (เช่น และ หรือ เป็นต้น) เพื่อทำให้ระบบสามารถกรองข้อมูลได้ตรงตามความต้องการของสมาชิกมากที่สุด

6.3.3 ในการทำการวิจัยทางผู้วิจัยไม่ได้คำนึงถึงความหน่วงของเครือข่าย ดังนั้นควรมีการจำลองสถานการณ์ที่มีการคำนึงถึงค่าดังกล่าว เพราะอาจส่งผลต่อประสิทธิภาพของระบบที่ออกแบบได้

6.3.4 ในการทำการวิจัยผู้วิจัยได้ทำการเชื่อมเพรดิเคตด้วยคำเชื่อมหรือเท่านั้น ดังนั้นเพื่อให้สามารถระบุถึงความต้องการของสารสนเทศได้ละเอียดมากขึ้น ควรมีการให้สมาชิกระบุคำเชื่อมของเพรดิเคตได้ด้วยตนเอง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# รายการอ้างอิง


1. Eugster, P.T., et al. The Many Faces of Publish/Subscribe. ACM Computing Surveys (CSUR) 35 (June, 2003).
2. Coster, R. The Architecture and Implementation of a system for Collaborative and Content-based Filtering. Technical Report, Department of Computer and System Sciences, Stockholm University, November 2002.
3. Amato, G. and Straccia, U. User Profile Modeling and Applications to Digital Libraries. Proceedings of the 3rd European Conference on Digital Libraries (ECDL). Paris, France, September, 1999.
4. Fabret, F., et al. Efficient Matching for Content-based Publish/Subscribe System. Technical Report, INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE(INRIA), 2000.
5. New Era of Networks Inc. NEONRules. (Online). Available from: http://www.neonsoft.com/whitepapers/MQSIRules.html
6. Gough, K.J. and Smith, G. Effcient recognition of events in distributed systems. Proceedings of ACSC-18. 1995.
7. Hanson, E.N., et al. A Predicate matching algorithm for database rule systems. SIGMOD'90 (1990): 271-280.
8. Aguilera, M.K., et al. Matching Event in A Content-based Subscription System. 18th ACM Symposium on Principles of Distributed Computing (PODC'99) (1999).
9. Altinel, M. and Franklin, M.J. Efficient Filtering of XML Document for Selective Dissemination of Information. Proceedings of the 26th International Conference on Very Large Database (VLDB) Conference, Cairo, Egypt. 2000.
10. Fabret, F., et al. Filtering Algorithm and Implementation for Very Fast Publish/Subscribe Systems. ACM SIGMOD 2001 (May, 2001).
11. Ashayer, G., Leung, H.K.Y., and Jacobsen, H.-A. Predicate Matching and Subscription Matching in Publish/Subscribe Systems. Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops (ICDCSW'02). 2002.

12. Rjaibi, W., Dittrich, K.R., and Jaepel, D. Event Matching in Sysmetric Subscription Systems. Proceedings of 2002 conference of the Center for Advanced Studies on Collaborative research. 2002.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

# ประวัติผู้เขียนวิทยานิพนธ์

นาย กัน อุตะเดช เกิดเมื่อวันที่ 18 กันยายน พ.ศ. 2519 สำเร็จการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ประยุกต์ ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ เมื่อปีการศึกษา 2541 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2544

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย