แบบจำลองการสืบค้นแบบรูปการออกแบบ

RETRIEVING MODEL FOR DESIGN PATTERNS

วีณาวดี ม่วงอ้น
WEENAWADEE MUNGON

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี
พ.ศ. 2549

ISBN 974-15-2403-9

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

แบบจำลองการสืบค้นแบบรูปการออกแบบ

RETRIEVING MODEL FOR DESIGN PATTERNS

วีณาวดี ม่วงอ้น

WEENAWADEE MUNGON

วน.
ว 815 ว
2549

เลขหมู่.....
เลขทะเบียน.....63452
วัน,เดือน,ปี.....28 ส.ค. 2549

.b.....11637420
.i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2549

ISBN 974-15-2409-9

RETRIEVING MODEL FOR DESIGN PATTERNS

WEENAWADEE MUANGON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN COMPUTER SCIENCE

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2006

ISBN 974-15-2409-9

COPYRIGHT 2006

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อวิทยานิพนธ์	แบบจำลองการสืบค้นรูปแบบการออกแบบ
นักศึกษา	นางสาววีณาวดี ม่วงอ้น
รหัสประจำตัว	45064610
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2549
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ.ดร.ศรัณย์ อินทโกสุม

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ พัฒนาแบบจำลองการสืบค้นแบบรูปการออกแบบ ซึ่งสามารถสืบค้นได้ตามลักษณะโดเมนของปัญหา ทำให้มีความยืดหยุ่นกว่าการสืบค้นโดยการใช้ ชื่อ หรือ หมวดหมู่ ตามที่มีอยู่ในปัจจุบัน ซึ่งจะช่วยให้ผู้พัฒนาระบบสามารถนำแบบรูปการออกแบบที่เหมาะสมไปใช้งานในโครงการต่าง ๆ ได้โดยสะดวกยิ่งขึ้น งานวิจัยนี้ได้นำเสนอแบบจำลองการสืบค้นแบบรูปการออกแบบที่ประกอบด้วย การวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น และการออกแบบวิธีคำนวณค่าถ่วงน้ำหนักให้เหมาะสมกับดัชนี โดยได้ใช้แบบจำลองเวกเตอร์สเปซในการคำนวณหาค่าความคล้ายคลึงระหว่างคำถามและเอกสาร จากผลการทดลองพบว่าแบบจำลองที่เสนอในงานวิจัยนี้ จะให้ความแม่นยำเฉลี่ยในการค้นแบบรูปการออกแบบอยู่ที่ประมาณร้อยละ 70

Thesis Title	Retrieving Model for Design Patterns
Student	Ms.Weenawadee Muangon
Student ID.	45064610
Degree	Master of Science
Programme	Computer Science
Year	2006
Thesis Advisor	Asst. Prof. Dr. Sarun Intakosum

ABSTRACT

The purpose of this research is to develop a retrieving model for design patterns, based on problem domain context. This model is more flexible than existing tools that use only patterns name or category. As a result, developers can easily apply suitable design patterns to their projects. This model is composed of two major parts, the analysis of design pattern documents to create search index, and the calculation of index weight. Vector space model is used for calculation of similarity between queries and documents. The result of this research shows that precision of the proposed model, in retrieving the correct design patterns, is about 70 percents in average.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จได้อย่างดี ด้วยคำแนะนำ คำปรึกษา ความรู้ ความดูแลเอาใจใส่อย่างดี จาก ผศ.ดร. ศรัณย์ อินทโกสุม ซึ่งเป็นอาจารย์ผู้ควบคุมงานวิจัย ซึ่งท่านได้สละเวลาให้กับข้าพเจ้าอย่างเต็มที่ในการแก้ปัญหาต่างๆ อันเป็นประโยชน์ต่องานวิจัยนี้เป็นอย่างดี ข้าพเจ้ารู้สึกซาบซึ้งเป็นอย่างยิ่ง จึงใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ รศ.ดร. วีระ บุญจริง และ ผศ.ดร.จีรพร ศรีสวัสดิ์ สำหรับ ความรู้ คำแนะนำ และคำปรึกษาต่างๆ ในการทำวิทยานิพนธ์ด้วยดีตลอดมา

ขอกราบขอบพระคุณ ดร.เฉลิมศักดิ์ เลิศวงศ์เสถียร ซึ่งให้ความกรุณามาเป็นตัวแทนกรรมการจากบุคคลภายนอกทำให้ได้รับคำปรึกษาและคำแนะนำต่างๆในการสอบวิทยานิพนธ์

ขอขอบพระคุณนายรุ่งโรจน์ วชิรจารุ ที่มีส่วนร่วมในการให้คำปรึกษา และคำแนะนำ ในการทำวิทยานิพนธ์ เป็นอย่างดี

ขอขอบคุณนายกฤษดา เพชรทับทิม เพื่อนที่น่ารักที่คอยให้กำลังใจ และช่วยเหลือให้คำปรึกษาในเรื่องต่างๆเป็นอย่างดีตลอดมา

ขอขอบคุณ สาว กิ้ว เพื่อนที่แสนดี ที่คอยดูแล ช่วยเหลือในขณะศึกษาปริญญาโทและทำวิทยานิพนธ์ และขอบคุณน้องตี๋ น้องแก่น น้องอั๋น และเพื่อนๆ พี่ๆ น้องๆ นักศึกษาปริญญาโททุกคน ที่คอยให้การช่วยเหลือและให้คำปรึกษาด้านต่างๆ และยังเป็นกำลังใจต่อผู้วิจัยอย่างใกล้ชิดตลอดมา

ขอขอบคุณ นายอนันต์ชัย ม่วงอ้น พี่ชายที่แสนดี ที่คอยให้กำลังใจ และให้คำปรึกษาในเรื่องต่างๆ จนทำให้ข้าพเจ้ามีกำลังใจทำวิทยานิพนธ์สำเร็จด้วยดี

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่คอยดูแล สั่งสอน ให้กำลังใจ และสนับสนุนในการเรียนเป็นอย่างดีตลอดมา ทำให้ข้าพเจ้าสามารถทำวิทยานิพนธ์เล่มนี้สำเร็จลงด้วยดี

คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแด่ผู้มีพระคุณทุกท่าน

วีณาวดี ม่วงอ้น

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป IX

บทที่ 1 บทนำ 1

1.1 ความสำคัญและที่มาของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ 2

1.3 สมมุติฐานของการศึกษา 2

1.4 ขอบเขตการศึกษา 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 ส่วนประกอบของวิทยานิพนธ์ 3

บทที่ 2 แบบรูปการออกแบบ 5

2.1 ความรู้ทั่วไปเกี่ยวกับแบบรูปการออกแบบ 5

2.1.1 หมวดหมู่ของแบบรูปการออกแบบ 7

2.1.1.1 หมวดหมู่การสร้าง 7

2.1.1.2 หมวดหมู่โครงสร้าง 8

2.1.1.3 หมวดหมู่พฤติกรรม 9

2.1.2 แม่แบบการอธิบายแบบรูปการออกแบบ 11

2.2 เครื่องมือสืบค้นแบบรูปการออกแบบ 13

2.3 งานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงสร้างการอธิบายแบบรูปการออกแบบ 15

บทที่ 3 การสืบค้นระบบสารสนเทศ 22

3.1 ความรู้ทั่วไปเกี่ยวกับการสืบค้นระบบสารสนเทศ 22

3.1.1 การเตรียมฐานข้อมูล 23

3.1.1.1 ขั้นตอนการเตรียมเอกสาร 23

สารบัญ (ต่อ)

หน้า

5.3.1 วิธีการทดลอง 54
5.3.2 ผลการทดลอง 55
5.3.3 วิเคราะห์แบบจำลองการสืบค้นแบบรูปการออกแบบ 56

บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ 60
5.1 สรุป 60
5.2 ข้อเสนอแนะ 60

เอกสารอ้างอิง 62

ภาคผนวก ก แสดงคำถามที่ใช้ในการทดลองการสืบค้นแบบรูปการออกแบบ 65
ภาคผนวก ข แสดงดัชนี และค่าถ่วงน้ำหนักของดัชนี 69

ประวัติผู้เขียน 81

สารบัญตาราง

ตารางที่ หน้า

2.1 แสดงหมวดหมู่ของแบบรูปการออกแบบ 6
2.2 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้าง 16
2.3 แสดงโครงสร้างใหม่ของแอบสแทรกต์ แฟกทอรี 17
2.4 แสดงโครงสร้างใหม่ของซิงเกิลตอน 17
2.5 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่โครงสร้าง 17
2.6 แสดงโครงสร้างใหม่ของอะแดปเตอร์ 18
2.7 โครงสร้างวากยสัมพันธ์ของหมวดหมู่พฤติกรรม 18
2.8 แสดงโครงสร้างใหม่ของออบเซิฟเวอร์ 19
3.1 แสดงจำนวนความถี่ของคำที่ปรากฏในแต่ละเอกสารและคำถาม 28
3.2 แสดงความถี่ผกผันของดัชนีตามข้อมูลในรูปที่ 3.2 ก) 29
3.3 แสดงตัวอย่างค่าถ่วงน้ำหนักของดัชนีที่ใช้ในการคำนวณค่าการป้อนกลับของผู้ใช้ 32
4.1 แสดงตัวอย่างชื่อรองของแบบรูปการออกแบบ 40
4.2 แสดงตัวอย่างฟังก์ชันของหมวดหมู่แบบรูปการออกแบบ 41
4.3 แสดงตัวอย่างฟังก์ชันของแบบรูปการออกแบบ 41
4.4 แสดงตัวอย่างรูปแบบวัตถุของแบบรูปการออกแบบ 42
4.5 แสดงตัวอย่างการเก็บดัชนีเดี่ยว 43
4.6 แสดงตัวอย่างการเก็บดัชนีวลี 44
4.7 แสดงค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของตัวแทนแบบรูปการออกแบบ 47
4.8 แสดงรายละเอียดความสามารถในการบ่งชี้ของแต่ละดัชนีในตัวอย่างที่ 4.1 50
4.9 แสดงค่าถ่วงน้ำหนักดัชนีแบบเดี่ยวของเอกสารแอบสแทรกต์ แฟกทอรี 51
5.1 แสดงความสามารถในการบ่งชี้ของตัวแทนแบบรูปการออกแบบ 55
5.2 แสดงค่าเฉลี่ยความแม่นยำในค่าความระลึก 11 จุด ของผลการทดลอง 56
5.3 แสดงตัวอย่างการสืบค้นคำถาม “I want design pattern that can create class for convert interface” 57
5.4 แสดงตัวอย่างการสืบค้นคำถาม “I want design pattern that can class to notify dependent or relate object” 57
5.5 แสดงค่าถ่วงน้ำหนักของดัชนีในคำถามตารางที่ 5.4 58
5.6 แสดงผลลัพธ์การสืบค้นหลังการปรับค่าถ่วงน้ำหนักของคำถามในตารางที่ 5.4 59

สารบัญรูป

รูปที่ หน้า

2.1 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบแอบสแทรกต์ แฟกทอรี 7

2.2 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบอะแดปเตอร์ 8

2.3 แสดงลักษณะการทำงานของแบบรูปการออกแบบออบเซิฟเวอร์ 10

2.4 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบออบเซิฟเวอร์ 10

2.5 แสดงเว็บไซต์ที่ให้บริการสืบค้นแบบรูปการออกแบบ 13

2.6 แสดงการสืบค้นแบบรูปการออกแบบของ Frank และคณะ 14

2.7 แสดงส่วนการสืบค้นรูปแบบการออกแบบของเครื่องมือเอ็มวีเคส 15

2.8 แสดงการประยุกต์แบบจำลองเมทาพีซีเอ็มเอล 20

2.9 แสดงตัวอย่างความสัมพันธ์ภาคแสดงและภาคประธานของอะแดปเตอร์ 21

3.1 แสดงกระบวนการสืบค้นสารสนเทศ 22

3.2 แสดงการแทนของดัชนี และ คำสำคัญที่ใช้ในการสืบค้นเป็นเวกเตอร์ 27

3.3 แสดงกราฟค่าความแม่นยำเทียบกับค่าความระลึก 35

4.1 แสดงโครงสร้างการอธิบายแบบรูปการออกแบบ 40

4.2 แสดงลำดับความสามารถในการบ่งชี้โครงสร้างข้อมูลแบบรูปการออกแบบ 45

5.1 แสดงผลการทดลองการใช้แบบจำลองการสืบค้นแบบรูปการออกแบบ 55

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

การออกแบบเป็นขั้นตอนสำคัญขั้นตอนหนึ่งในงานพัฒนาซอฟต์แวร์ ซึ่งถ้าซอฟต์แวร์ใดมีการออกแบบที่ดี การพัฒนาก็จะสามารถทำได้ง่าย รวดเร็ว และมีประสิทธิภาพ เทคนิคหนึ่งที่ใช้ในการออกแบบซอฟต์แวร์ในปัจจุบันนี้ คือ แบบรูปการออกแบบ (Design Pattern) เทคนิคนี้เกิดขึ้นจากผู้มีประสบการณ์ในการออกแบบซอฟต์แวร์ ทำการคิดค้นแบบรูป (Pattern) เพื่อใช้สำหรับแก้ไขปัญหาในแต่ละด้าน (Problem in Context) ของการออกแบบซอฟต์แวร์ แบบรูปการออกแบบมีความยืดหยุ่นในการทำงานสูง และคุณสมบัติในการนำกลับมาใช้ใหม่ (Reusable) จึงทำให้การออกแบบและพัฒนาซอฟต์แวร์สามารถทำได้รวดเร็วมากขึ้น อีกทั้งเทคนิคนี้ยังช่วยให้การติดต่อสื่อสารกันระหว่างผู้ออกแบบและผู้พัฒนาซอฟต์แวร์ สามารถทำได้ง่ายมากยิ่งขึ้น

อย่างไรก็ตามผู้ที่นำแบบรูปการออกแบบมาใช้ในการออกแบบซอฟต์แวร์ จำเป็นที่จะต้องมีความรู้เกี่ยวกับแบบรูปการออกแบบเป็นอย่างดี นั่นคือจะต้องทราบว่าปัญหาการออกแบบนี้จะต้องใช้แบบรูปการออกแบบชื่ออะไรในการแก้ไขปัญหา จากนั้นจึงไปสืบค้นเอกสารอ้างอิงดูรายละเอียดของแบบรูปที่ต้องการ ซึ่งอาจอยู่ในรูปแบบของ หนังสือ เว็บไซต์ หรือเครื่องมืออิเล็กทรอนิกส์ที่ช่วยในการสืบค้นแบบรูปการออกแบบ เช่น เครื่องมือเอ็มวีเคส (MVCASE TOOL) [7] วิธีการสืบค้นแบบรูปการออกแบบเหล่านี้อาจไม่มีประโยชน์มากนักสำหรับผู้พัฒนาซอฟต์แวร์ที่ยังไม่มีประสบการณ์ในการออกแบบ หรือผู้ที่ไม่มีความรู้ในเรื่องแบบรูปการออกแบบ เนื่องจากไม่สามารถวิเคราะห์ได้ว่าปัญหาการออกแบบที่ประสบอยู่นี้ จะต้องใช้แบบรูปการออกแบบใดมาแก้ไขจึงจะเหมาะสม ดังนั้นวิธีการหนึ่งที่มีประโยชน์มากกว่า คือ การให้ผู้ใช้ได้สืบค้นแบบรูปการออกแบบตามลักษณะปัญหาที่ผู้ใช้ต้องการแก้ไข

จากปัญหาดังกล่าวจึงมีแนวความคิดที่จะทำวิจัยเพื่อออกแบบแบบจำลองการสืบค้นแบบรูปการออกแบบที่สามารถสืบค้นแบบรูปการออกแบบได้ตามลักษณะโดเมนปัญหาการออกแบบที่ผู้พัฒนาซอฟต์แวร์ประสบ โดยสืบค้นตามคำสำคัญที่ผู้ใช้ป้อนเกี่ยวกับปัญหาแล้วทำการจัดลำดับให้ว่าผลลัพธ์ใดคล้ายคลึงกับคำถามที่สุด แบบจำลองนี้จะช่วยให้ผู้พัฒนาซอฟต์แวร์สามารถเลือกใช้แบบรูปการออกแบบได้ตรงตามปัญหา และเป็นแนวทางหนึ่งที่สามารถลดระยะเวลาในการเรียนรู้แบบรูปการออกแบบได้อีกด้วย

1.2 ความมุ่งหมาย และวัตถุประสงค์

งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบแบบจำลองการสืบค้นแบบรูปการออกแบบ ให้มีความสามารถในการสืบค้นแบบรูปการออกแบบได้ครอบคลุมตามโดเมนปัญหาการออกแบบที่ผู้ใช้ประสบ

1.3 สมมติฐานการศึกษา

แบบจำลองการสืบค้นแบบรูปการออกแบบจากงานวิจัยนี้ จะช่วยให้ผู้ใช้งานแบบรูปการออกแบบสามารถสืบค้นแบบรูปการออกแบบได้ตรงกับโดเมนปัญหาการออกแบบที่ประสบ

1.4 ขอบเขตการศึกษา

1.4.1 งานวิจัยนี้ออกแบบแบบจำลองเพื่อสืบค้นแบบรูปการออกแบบ โดยวิเคราะห์ตัวแทนเอกสารเพื่อสร้างดัชนีคำค้น และออกแบบวิธีคำนวณค่าถ่วงน้ำหนักให้เหมาะสมกับดัชนี

1.4.2 งานวิจัยนี้ใช้แบบจำลองเวกเตอร์สเปซ ในการคำนวณค่าความคล้ายคลึงระหว่างเอกสารและคำถาม

1.4.3 การทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบนี้ จะทดลองโดยการพัฒนาดัชนีจากเทคนิคการวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น และคำนวณค่าถ่วงน้ำหนักของดัชนีตามแบบจำลองการสืบค้นแบบรูปการออกแบบ จากนั้นทำการพัฒนาโปรแกรมการสืบค้นซึ่งมีวิธีการคำนวณหาค่าความคล้ายคลึงตามหลักการของเวกเตอร์สเปซ

1.4.4 ลักษณะคำถามในการสืบค้นจะอยู่ในรูปของตัวอักษร และผลลัพธ์ของการสืบค้นจะเป็นชื่อของแบบรูปการออกแบบ

1.4.5 ดัชนีที่ใช้ในการทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบนี้จะไม่ได้นำคุณสมบัติรากศัพท์ (Stemming) มาใช้ในการสืบค้นด้วย

1.4.6 การทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบนี้ ผู้ทำวิจัยเป็นผู้เชี่ยวชาญที่ทำการวิเคราะห์ว่าแบบรูปการออกแบบที่ทำการสืบค้นมานั้นเหมาะสมกับคำถามหรือไม่ เพื่อนำไปใช้ในการประเมินผลลัพธ์การสืบค้น

1.4.7 การประเมินผลลัพธ์การสืบค้นของแบบจำลองการสืบค้นแบบรูปการออกแบบนี้ ใช้หลักการประเมินของกระบวนการสืบค้นสารสนเทศคือ วิธีหาค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยความแม่นยำ (Average Precision)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้แนวทางในการวิเคราะห์โครงสร้างข้อมูลของแบบรูปการออกแบบเพื่อสร้างเป็นดัชนีบ่งชี้ถึงแบบรูปการออกแบบ

1.5.2 ได้แบบจำลองสำหรับนำไปพัฒนาเป็นเครื่องมือสืบค้นแบบรูปการออกแบบได้

1.5.3 เครื่องมือสืบค้นแบบรูปการออกแบบที่ได้จากแบบจำลองนี้ สามารถลดความยุ่งยากในขั้นตอนการวิเคราะห์ปัญหาการออกแบบของผู้ออกแบบซอฟต์แวร์ เนื่องจากผู้ใช้สามารถสืบค้นแบบรูปการออกแบบเชิงวัตถุได้ตามโดเมนปัญหาที่ผู้ใช้ประสบ จึงทำให้สามารถลดระยะเวลาในการออกแบบซอฟต์แวร์ได้

1.5.4 ทำให้ผู้พัฒนาซอฟต์แวร์มีความสนใจ และมีความต้องการใช้แบบรูปการออกแบบมากขึ้น เนื่องจากสามารถสืบค้นแบบรูปการออกแบบมาใช้งานได้ตรงตามความต้องการ

1.6 ส่วนประกอบของวิทยานิพนธ์

รายละเอียดส่วนอื่นๆที่จะนำเสนอในวิทยานิพนธ์ฉบับนี้ สามารถแบ่งการอธิบายได้เป็นบทต่างๆ ดังนี้

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับแบบรูปการออกแบบ ซึ่งแบ่งการนำเสนอออกเป็น 2 ส่วน โดยในส่วนแรกจะนำเสนอข้อมูลพื้นฐานที่เกี่ยวข้องกับแบบรูปการออกแบบ ส่วนที่สองจะนำเสนองานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงสร้างการอธิบายแบบรูปการออกแบบให้มีความสามารถในการสืบค้นมากยิ่งขึ้น

บทที่ 3 กล่าวถึงทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการสืบค้นระบบสารสนเทศ ซึ่งแบ่งการนำเสนอออกเป็น 2 ส่วน โดยส่วนแรกจะนำเสนอทฤษฎีพื้นฐานในการสืบค้นระบบสารสนเทศ และส่วนที่สองจะนำเสนองานวิจัยที่เกี่ยวข้องกับการออกแบบการคำนวณค่าถ่วงน้ำหนักในระบบสืบค้นสารสนเทศ

บทที่ 4 กล่าวถึงแนวทางการออกแบบและพัฒนาแบบจำลองการสืบค้นแบบรูปการออกแบบ ซึ่งแบ่งการนำเสนอออกเป็น 2 ส่วน คือ ส่วนแรกจะกล่าวถึงแบบจำลองการวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น ส่วนที่สองเป็นการออกแบบวิธีการคำนวณค่าถ่วงน้ำหนักให้เหมาะสมกับดัชนีที่ออกแบบ

บทที่ 5 กล่าวถึงการประเมินผลลัพธ์ของแบบจำลองการสืบค้นแบบรูปการออกแบบ โดยแบ่งออกเป็น 3 ส่วน คือ ส่วนแรกจะกล่าวถึงวิธีการหาค่าความคล้ายคลึงระหว่างเอกสารและคำถาม ส่วนที่สองจะกล่าวถึงวิธีการประเมินผลลัพธ์ในการสืบค้นของระบบ และส่วนสุดท้ายกล่าวถึงการ

ทดลอง ซึ่งประกอบด้วย วิธีการทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลองที่ได้จากการสืบค้นแบบรูปการออกแบบ

บทที่ 6 กล่าวสรุปผลการวิจัยทั้งหมดรวมทั้งข้อเสนอแนะเพื่อการทำวิจัยต่อในอนาคต

บทที่ 2
แบบรูปการออกแบบ

ในบทนี้จะ กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับแบบรูปการออกแบบ ซึ่งแบ่งการนำเสนอออกเป็น 2 ส่วน โดยในส่วนแรกจะนำเสนอข้อมูลพื้นฐานที่เกี่ยวข้องกับแบบรูปการออกแบบ ส่วนที่สองนำเสนองานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงสร้างการอธิบายแบบรูปการออกแบบให้มีความสามารถในการสืบค้นมากยิ่งขึ้น

2.1 ความรู้ทั่วไปเกี่ยวกับแบบรูปการออกแบบ (Design Pattern)

ในทางวิศวกรรมซอฟต์แวร์แบบรูปการออกแบบคือ ผลเฉลยของปัญหาทั่วๆไปที่เกิดขึ้นในการออกแบบซอฟต์แวร์ และเป็นเทคนิคที่ช่วยให้ได้การออกแบบที่เหมาะสมที่สุดสำหรับซอฟต์แวร์ แบบรูปการออกแบบนี้เกิดขึ้นจากการเปรียบเทียบหลักการสร้างบ้านของสถาปนิก ที่มองว่าการออกแบบในแต่ละส่วนของบ้าน เป็นแบบรูป (Pattern) โดยเมื่อจะทำการสร้างบ้านสถาปนิกก็จะเลือกเอาแบบรูปของแต่ละส่วนมารวมกันเป็นพิมพ์เขียว (Blueprint) ของบ้านทั้งหลังเพื่อใช้ในการสื่อสารกับเจ้าของบ้าน วิศวกร และทีมงาน เมื่อนำแนวความคิดนี้มาประยุกต์เข้ากับการออกแบบซอฟต์แวร์แล้ว แบบรูปการออกแบบ ก็คือเทคนิคที่เป็นแบบรูป โดยเมื่อทำการออกแบบซอฟต์แวร์นักออกแบบซอฟต์แวร์ก็จะเลือกนำเอาแบบรูปที่เหมาะสมกับการแก้ไขปัญหามาประยุกต์เข้ากับซอฟต์แวร์ จึงทำให้การพัฒนาซอฟต์แวร์สามารถทำได้ง่าย และรวดเร็วมากยิ่งขึ้น อีกทั้งแบบรูปการออกแบบนี้จะทำให้การติดต่อสื่อสารกันระหว่างผู้ออกแบบซอฟต์แวร์ และผู้พัฒนาซอฟต์แวร์ สามารถทำได้ง่ายมากยิ่งขึ้น

ในปัจจุบันนี้แบบรูปการออกแบบได้ถูกคิดค้นขึ้นมากมาย โดยงานวิจัยที่เกี่ยวข้องกับเทคนิคแบบรูปการออกแบบนี้ได้ถูกรวบรวมไว้เป็นส่วนหนึ่งของหนังสือที่ทำการรวบรวมงานวิจัยเกี่ยวกับแบบรูป [6][21] หรือ หนังสือนำเสนอการอธิบายแบบรูปการออกแบบโดยตรง เพื่อความสะดวกสำหรับผู้ที่ต้องการศึกษาแบบรูปการออกแบบ เช่น หนังสือ Design Pattern: Element of Reusable Object Oriented Software [10] หนังสือเล่มนี้ได้ทำการรวบรวมแบบรูปการออกแบบโดยกลุ่มผู้มีประสบการณ์ทางด้านการออกแบบซอฟต์แวร์ 4 คน คือ Erich Gamma, Richard Helm, Ralph John Vlissides หรือรู้จักกันดีในชื่อกลุ่ม จีโอเอฟ (GOF: Gang of Four) โดยได้นำเสนอแบบรูปการออกแบบไว้ 23 แบบรูป ซึ่งแบ่งตามลักษณะการทำงานเป็น 3 หมวดหมู่ คือ หมวดหมู่การสร้าง หมวดหมู่โครงสร้าง และหมวดหมู่พฤติกรรม โดยหนังสือเล่มนี้เป็นที่รู้จักกันดีในกลุ่มนักออกแบบซอฟต์แวร์เนื่องจากเป็นหนังสือที่ทำการรวบรวมแบบรูปการออกแบบที่สามารถแก้ไข

ปัญหาการออกแบบซอฟต์แวร์ที่พบบ่อยครั้งในการออกแบบซอฟต์แวร์ นอกเหนือจากการหนังสือแล้วแบบรูปการออกแบบยังสามารถหาได้จากเว็บไซต์ [25][26][27][28] ซึ่งส่วนหนึ่งของเว็บไซต์เหล่านี้ได้อธิบายถึงแบบรูปการออกแบบ เช่น เว็บไซต์วิกิพีเดีย (Wikipedia) [27] เป็นเว็บไซต์หนึ่งที่สามารถค้นคว้าหาความรู้ที่เกี่ยวข้องกับแบบรูปการออกแบบประเภทต่างๆได้ ซึ่งในปัจจุบันเว็บไซต์นี้มีข้อมูลแบบรูปการออกแบบอยู่ 48 แบบรูป โดยในบทนี้จะยกตัวอย่างแบบรูปการออกแบบส่วนหนึ่งที่เป็นที่รู้จักกันดีในการออกแบบซอฟต์แวร์นั่นคือ แบบรูปการออกแบบของจีโอเอฟ และแบบรูปการออกแบบจากเว็บไซต์วิกิพีเดีย ซึ่งแสดงได้ดังตารางที่ 2.1

ตาราง 2.1 แสดงหมวดหมู่ของแบบรูปการออกแบบ

หมวดหมู่แบบรูปการออกแบบ		
หมวดหมู่การสร้าง	**หมวดหมู่โครงสร้าง**	**หมวดหมู่พฤติกรรม**
Abstract Factory	Adapter	Interpreter
Factory Method	Bridge	Chain of Responsibility
Builder	Composite	Command
Singleton	Decorator	Iterator
Prototype	Facade	Mediator
Anonymous subroutine object	Flyweight	Momento
Lazy initialization	Proxy	Observer
	Container	State
	Pipe and filter	Strategy
	DataBus	Visitor
	RoleObject	HierarchicalVisitor
		Event listener

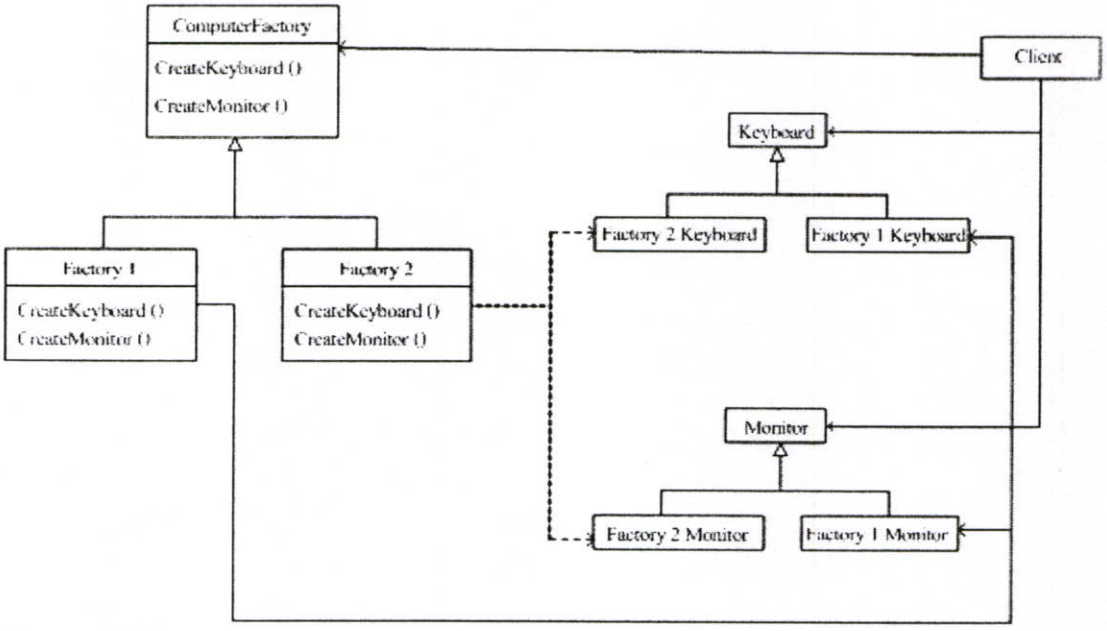
ตารางที่ 2.1 แสดงแบบรูปการออกแบบที่อยู่ในแต่ละหมวดหมู่ ซึ่งหมวดหมู่นี้จะยึดตามหมวดหมู่ของจีโอเอฟและแบบรูปการออกแบบนี้นำมาจากหนังสือจีโอเอฟ และเว็บไซต์วิกิพีเดีย ซึ่งหัวข้อที่ 2.1.1 ได้อธิบายการทำงานของแต่ละหมวดหมู่ และตัวอย่างการทำงานของแบบรูปการออกแบบ และหัวข้อที่ 2.1.2 อธิบายโครงสร้างที่ใช้ในการอธิบายแบบรูปการออกแบบซึ่งในแต่ละแบบรูปการออกแบบอาจใช้วิธีอธิบายการทำงานไม่เหมือนกัน

2.1.1 หมวดหมู่แบบรูปการออกแบบ มีรายละเอียดดังต่อไปนี้

2.1.1.1 หมวดหมู่การสร้าง คือ แบบรูปการออกแบบที่จัดการเกี่ยวกับกลไกการสร้าง เพื่อเพิ่มความยืดหยุ่นในการสร้างออบเจ็กต์ให้กับระบบซอฟต์แวร์ โดยหมวดหมู่นี้จีโอเอฟนำเสนอแบบรูปการออกแบบไว้ 5 แบบรูป ดังนี้ แอบสแทรกต์แฟกทอรี (Abstract Factory), แฟกทอรีเมท็อด (Factory Method), โปรโตไทป์ (Prototype), ซิงเกิลตอน (Singleton) และบีลเดอร์ (Builder) นอกจากนี้ยังมีแบบรูปการออกแบบอื่นซึ่งจัดอยู่ในหมวดหมู่นี้ที่มีการนำเสนอไว้ในเว็บไซต์วิกีพีเดีย เช่น อโนนิมัสซับรูทีนออบเจ็กต์ (Anonymous subroutine object) และเลซี อินนิเชียลไลเซชัน (Lazy initialization) เป็นต้น ตัวอย่างที่ 2.1 แสดงตัวอย่างแบบรูปการออกแบบ แอบสแทรกต์แฟกทอรี

ตัวอย่าง 2.1 แบบรูปการออกแบบแอบสแทรกต์แฟกทอรี

แบบรูปการออกแบบแอบสแทรกต์แฟกทอรี เป็นแบบรูปการออกแบบหนึ่งในหมวดหมู่การสร้างที่ออกแบบกระบวนการสร้างสำหรับสร้างออบเจ็กต์ที่มีความสัมพันธ์กัน ซึ่งแบบรูปการออกแบบนี้จะใช้อินเทอร์เฟซในการประกาศเมท็อดที่ทำการสร้างออบเจ็กต์ที่สัมพันธ์กัน ซึ่งอินเทอร์เฟซจะเป็นตัวกลางในการติดต่อระหว่างลูกข่าย (Client) และส่วนของการสร้างออบเจ็กต์ โดยเมื่อมีการเปลี่ยนแปลงรหัสต้นฉบับ (Source Code) ในส่วนของการสร้างออบเจ็กต์ อินเทอร์เฟซจะช่วยให้ส่วนที่เป็นรหัสต้นฉบับของลูกข่ายไม่จำเป็นที่จะต้องเปลี่ยนแปลงตามไปด้วย แผนผังองค์ประกอบแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี แสดงตัวอย่างได้ดังรูปที่ 2.1



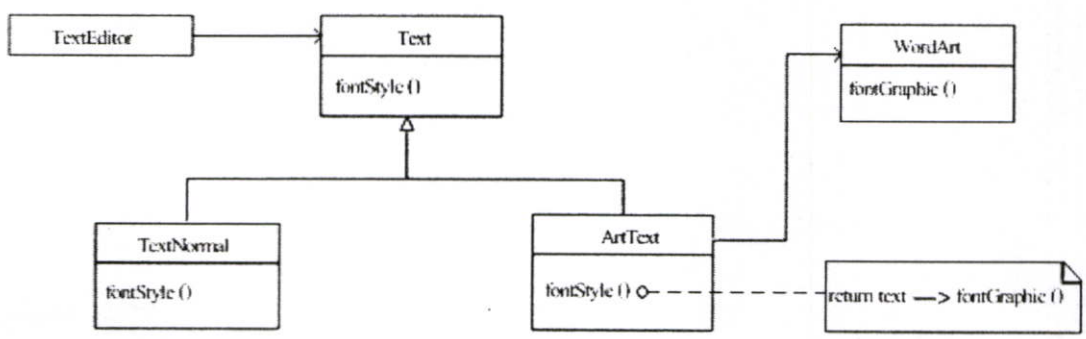
รูปที่ 2.1 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี

จากรูปที่ 2.1 แสดงแผนผังการทำงานในระบบการสร้างคอมพิวเตอร์ของ 2 โรงงาน ที่ใช้แบบรูปการออกแบบแอบสแทรกต์แฟกทอรีในส่วนของการสร้างออบเจ็กต์ที่สัมพันธ์กัน โดยคอมพิวเตอร์จะมีส่วนประกอบที่สัมพันธ์กันคือ จอภาพ (Monitor) และแป้นพิมพ์ (Keyboard) โดยส่วนประกอบเหล่านี้จะถูกพัฒนาแตกต่างกันขึ้นอยู่กับการพัฒนาของแต่ละโรงงาน แอบสแทรกต์แฟกทอรีจะใช้อินเทอร์เฟซเป็นตัวกลางระหว่างลูกข่าย และโรงงาน ซึ่งจะช่วยให้ลูกข่ายไม่จำเป็นต้องรู้รายละเอียดในส่วนของของโรงงาน เช่น ในกรณีที่มีการเพิ่มโรงงานเข้าไปในระบบ ลูกข่ายก็ไม่จำเป็นที่จะต้องทราบรายละเอียดในการเพิ่มโรงงาน เนื่องจากลูกข่ายจะทำการติดต่อโรงงานผ่านอินเทอร์เฟซซึ่งไม่มีการเปลี่ยนแปลง

2.1.1.2 หมวดหมู่โครงสร้าง คือแบบรูปการออกแบบที่ทำการออกแบบการประกอบกันของคลาส และออบเจ็กต์ เพื่อให้ระบบสามารถมีโครงสร้างขนาดใหญ่ขึ้นได้ โดยหมวดหมู่นี้จีโอเอฟนำเสนอแบบรูปการออกแบบไว้ 7 แบบรูป ดังนี้ คอมโพสิท (Composite), บริดจ์ (Bridge), เดคอเรเตอร์ (Decorator), ฟาซาด (Facade), ฟลายเวท (Flyweight), พรอคซี่ (Proxy) และอะแดปเตอร์ (Adapter) นอกจากนี้ยังมีแบบรูปการออกแบบอื่นซึ่งสามารถจัดอยู่ในหมวดหมู่นี้ได้ที่มีการนำเสนอไว้ในเว็บไซต์วิกีพีเดีย เช่น คอนเทรนเนอร์ (Container), ดาต้าบัส (DataBus) โรลออบเจ็กต์ (Role object) และไปป์แอนด์ฟิวเตอร์ (Pipe and filter) เป็นต้น ตัวอย่างที่ 2.2 แสดงตัวอย่างแบบรูปการออกแบบอะแดปเตอร์

ตัวอย่าง 2.2 แบบรูปการออกแบบอะแดปเตอร์

แบบรูปการออกแบบอะแดปเตอร์เป็นแบบรูปหนึ่งในหมวดหมู่โครงสร้าง โดยมีลักษณะการทำงานเช่นเดียวกับ อะแดปเตอร์ที่เป็นตัวแปลงกระแสไฟซึ่งทำการแปลงกระแสไฟจากปลั๊กไฟให้อยู่ในรูปแบบที่ผู้ใช้ต้องการใช้งาน ในระบบซอฟต์แวร์แบบรูปการออกแบบอะแดปเตอร์จะทำการแปลงอินเทอร์เฟซของคลาสที่ไม่สามารถทำงานด้วยกันได้ ให้สามารถทำงานด้วยกันได้ โดยแผนผังแสดงองค์ประกอบการใช้แบบรูปการออกแบบอะแดปเตอร์ แสดงตัวอย่างได้ดังรูปที่ 2.2



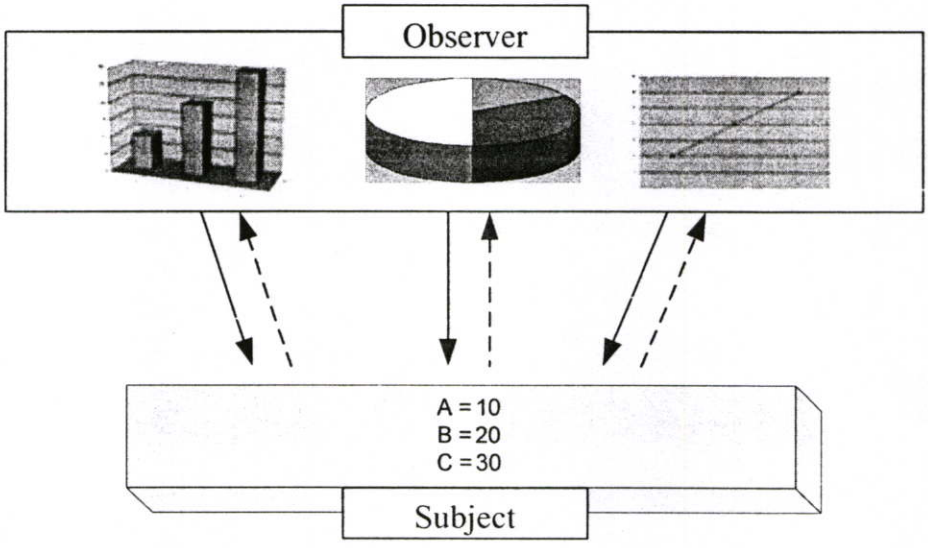
รูปที่ 2.2 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบอะแดปเตอร์

จากรูปที่ 2.2 แสดงการใช้แบบรูปการออกแบบอะแดปเตอร์ในระบบบรรณาธิการตัวอักษร (TextEditor) ระบบนี้จะจัดการเกี่ยวกับตัวอักษร เช่น การจัดการรูปแบบของอักษร (Font style) โดยระบบนี้มีคลาส TextNormal ที่ทำการพัฒนาการจัดการเกี่ยวกับตัวอักษร ซึ่งคลาสนี้จะถ่ายทอด (Inheritance) มาจากอินเทอร์เฟซ Text เมื่อระบบต้องการเพิ่มในส่วนของกราฟฟิกเข้าไปในอักษร จึงได้นำเมท็อด fontGraphic จากคลาส WordArt มาใช้ แต่เนื่องจากอินเทอร์เฟซของทั้งสองระบบนี้ต่างกันซึ่งไม่สามารถใช้งานร่วมกันได้โดยตรง ดังนั้นระบบจึงสร้างคลาสอะแดปเตอร์ ArtText เพื่อเป็นตัวกลางในการเรียกใช้เมท็อด fontGraphic ของคลาส WordArt ซึ่งเมท็อด fontGraphic นี้จะถูกเรียกใช้โดยเมท็อด fontStyle () ของคลาส ArtText

2.1.1.3 หมวดหมู่พฤติกรรม คือ แบบรูปการออกแบบที่ออกแบบการติดต่อสื่อสารระหว่างออบเจ็กต์ โดยออกแบบให้ออบเจ็กต์มีความยืดหยุ่นในการติดต่อมากขึ้น หมวดหมู่นี้จีโอเอฟนำเสนอแบบรูปการออกแบบไว้ 11 แบบรูป ดังนี้ คอมมานด์ (Command), อินเทอร์พริเตอร์ (Interpreter), อิตเทอร์เรเตอร์ (Iterator), มิดิเอเตอร์ (Mediator), มิเมนโต (Memento), ออบเซิฟเวอร์ (Observer), สเตท (State), สเตรทีจี (Strategy), เทมเพลตเมท็อด (Template Method), วิสิทเตอร์ (Visitor) และเชนออฟเรสปอนสิบิลิตี้ (Chain of Responsibility) รวมทั้งในเว็บไซต์วิกีพีเดียได้นำเสนอแบบรูปการออกแบบอื่นซึ่งจัดอยู่ในหมวดหมู่นี้ เช่น อีเวนต์ ลิสเซิลเนอร์ (Event Listener) และ ไฮลาชิคอลวิสิทเตอร์ (Hierarchical Visitor) เป็นต้น ตัวอย่างที่ 2.3 แสดงตัวอย่างแบบรูปการออกแบบ ออบเซิฟเวอร์

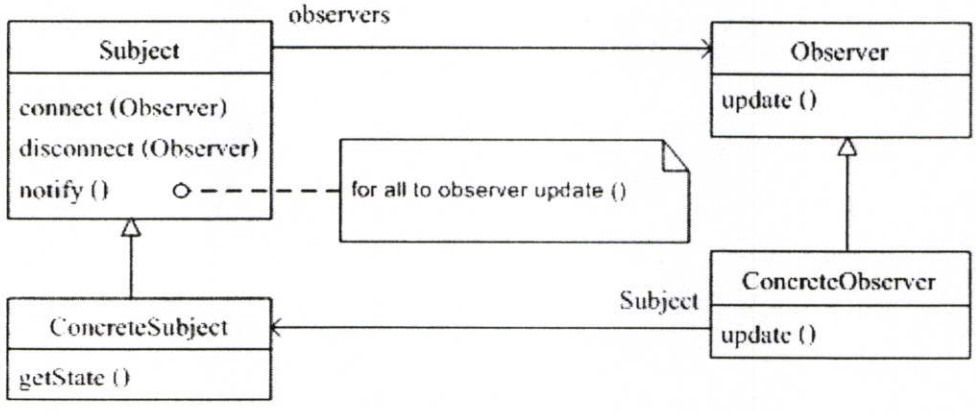
ตัวอย่าง 2.3 แบบรูปการออกแบบออบเซิฟเวอร์

แบบรูปออบเซิฟเวอร์ เป็นแบบรูปหนึ่งในหมวดหมู่พฤติกรรม ที่ทำการจัดการออบเจ็กต์ที่มีการติดต่อสื่อสารกันแบบ หนึ่งต่อกลุ่ม (One-to-Many) โดยความสัมพันธ์ระหว่างออบเจ็กต์เป็นแบบขึ้นต่อกัน (Dependent) เมื่อออบเจ็กต์ตัวหนึ่งมีการเปลี่ยนแปลงสถานะ ทุกๆออบเจ็กต์ที่ขึ้นต่อกันจะมีการปรับสถานะโดยอัตโนมัติ แสดงได้ดังรูปที่ 2.3



รูปที่ 2.3 แสดงลักษณะการทำงานของแบบรูปการออกแบบออบเซิฟเวอร์

รูปที่ 2.3 แสดงลักษณะการทำงานของแบบรูปการออกแบบออบเซิฟเวอร์ ในระบบการแสดงผล ที่มีแสดงผลแตกต่างกันออกไปตามข้อมูลที่กำหนด เช่น แสดงเป็น กราฟแท่ง กราฟวงกลม และกราฟเส้น เมื่อข้อมูลมีการเปลี่ยนแปลง กราฟการแสดงผลทั้ง 3 รูปแบบนี้ก็ต้องเปลี่ยนแปลงตามข้อมูลด้วย โดยแผนผังโครงสร้างความสัมพันธ์ของแบบรูปการออกแบบออบเซิฟเวอร์ แสดงได้รูปที่ 2.4



รูปที่ 2.4 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบออบเซิฟเวอร์

รูปที่ 2.4 แสดงแผนผังความสัมพันธ์ของแบบรูปการออกแบบออบเซิฟเวอร์ ซึ่งอธิบายกระบวนการทำงานได้ดังนี้ ออบเซิฟเวอร์ (Observer) คือคลาสที่ทำการลงทะเบียนกับคลาสซับเจ็กต์ (Subject) เพื่อประกาศการเป็นสมาชิกระหว่างซับเจ็กต์และออบเซิฟเวอร์ ซึ่งเมื่อซับเจ็กต์มีการ

เปลี่ยนแปลงข้อมูล ซับเจ็กต์จะส่งสัญญานโดยใช้เมท็อด notify () ไปยังออบเซิฟเวอร์ที่เป็นสมาชิกของซับเจ็กต์เพื่อบอกว่าซับเจ็กต์มีการเปลี่ยนแปลงข้อมูล หลังจากนั้นออบเซิฟเวอร์ก็จะใช้เมท็อด update() ทำการอัพเดตข้อมูลตามซับเจ็กต์ และซับเจ็กต์จะใช้เมท็อด getState() ทำการตรวจเช็คสถานะของข้อมูลออบเซิฟเวอร์

เมื่อนำแบบรูปการออกแบบออบเซิฟเวอร์ไปเทียบกับการทำงานในรูปที่ 2.3 จะพบว่า ซับเจ็กต์ คือ ข้อมูล และออบเซิฟเวอร์ คือ กราฟทั้งสามประเภท โดยเมื่อข้อมูลมีการเปลี่ยนแปลงกราฟทั้งสามประเภทก็จะได้รับสัญญานเพื่อทำการเปลี่ยนแปลงกราฟให้ตรงกับข้อมูลที่เปลี่ยน

2.1.2 แม่แบบการอธิบายแบบรูปการออกแบบ (Design Pattern Template)

แม่แบบการอธิบายแบบรูปการออกแบบ คือ โครงสร้างที่ใช้ในการอธิบายการทำงานของแบบรูปการออกแบบ ซึ่งอาจประกอบด้วยตัวอย่างปัญหาการออกแบบซอฟต์แวร์ที่ประสบและอธิบายวิธีการแก้ไขปัญหาของแบบรูปการออกแบบ โดยจะอธิบายวิธีการแก้ไขปัญหาในรูปแบบความสัมพันธ์ระหว่างคลาสและออบเจ็กต์ ซึ่งแม่แบบการอธิบายแบบรูปการออกแบบที่นิยมใช้ในปัจจุบันนี้ คือ แม่แบบของ อเล็กซานเดอร์ [1] และแม่แบบของจีโอเอฟ [10] ทั้งสองแม่แบบนี้จะมีโครงสร้างการอธิบายที่คล้ายคลึงกันนั่นคืออธิบายแบบรูปการออกแบบโดยยกตัวอย่างสถานการณ์เพื่ออธิบายการเกิดปัญหาการออกแบบและวิธีการแก้ไขปัญหาของแบบรูปการออกแบบ รายละเอียดแม่แบบของอเล็กซานเดอร์แสดงได้ดังตัวอย่างที่ 2.4 และจีโอเอฟแสดงได้ดังตัวอย่างที่ 2.5

ตัวอย่างที่ 2.4 แสดงการอธิบายรายละเอียดแม่แบบการอธิบายของอเล็กซานเดอร์

- ชื่อ (Name) คือ ส่วนที่เป็นชื่อของแบบรูปการออกแบบ
- ปัญหา (Problem) คือ ส่วนที่อธิบายถึงสถานะของปัญหาการออกแบบที่แบบรูปการออกแบบพยายามจะแก้ไข
- บริบท (Context) คือ ส่วนที่ยกตัวอย่างสถานการณ์ที่ทำให้เกิดปัญหาการออกแบบและอธิบายความสามารถของแบบรูปการออกแบบในการนำไปประยุกต์ใช้
- ข้อบังคับ (Force) คือ แสดงบทเรื่อง (Scenario) ในการใช้แบบรูปการออกแบบ
- ผลเฉลย (Solution) คือ ส่วนที่อธิบายวิธีการแก้ไขปัญหา โดยอธิบายว่าแบบรูปการออกแบบนั้นแก้ไขปัญหาอย่างไร
- ตัวอย่าง (Example) คือ ตัวอย่างการประยุกต์ใช้แบบรูปการออกแบบ ซึ่งจะช่วยให้ผู้ใช้แบบรูปการออกแบบสามารถเข้าใจแบบรูปการออกแบบได้ง่ายขึ้น
- บริบทผลลัพธ์ (Resulting Context) คือ ส่วนที่อธิบายสถานะของระบบหลังจากที่ได้ใช้แบบรูปการออกแบบแก้ไขปัญหาแล้ว
- เหตุผล (Rationale) คือ อธิบายขั้นตอนหรือ กฎ ของแบบรูปการออกแบบ

- แบบรูปการออกแบบที่เกี่ยวข้องกัน (Related Patterns) คือ อธิบายความสัมพันธ์ระหว่างตัวแบบรูปการออกแบบเอง และกับแบบรูปการออกแบบอื่น
- ตัวอย่างการใช้งานจริง (Known Uses) คือ ตัวอย่างการใช้งานแบบรูปการออกแบบในระบบจริง

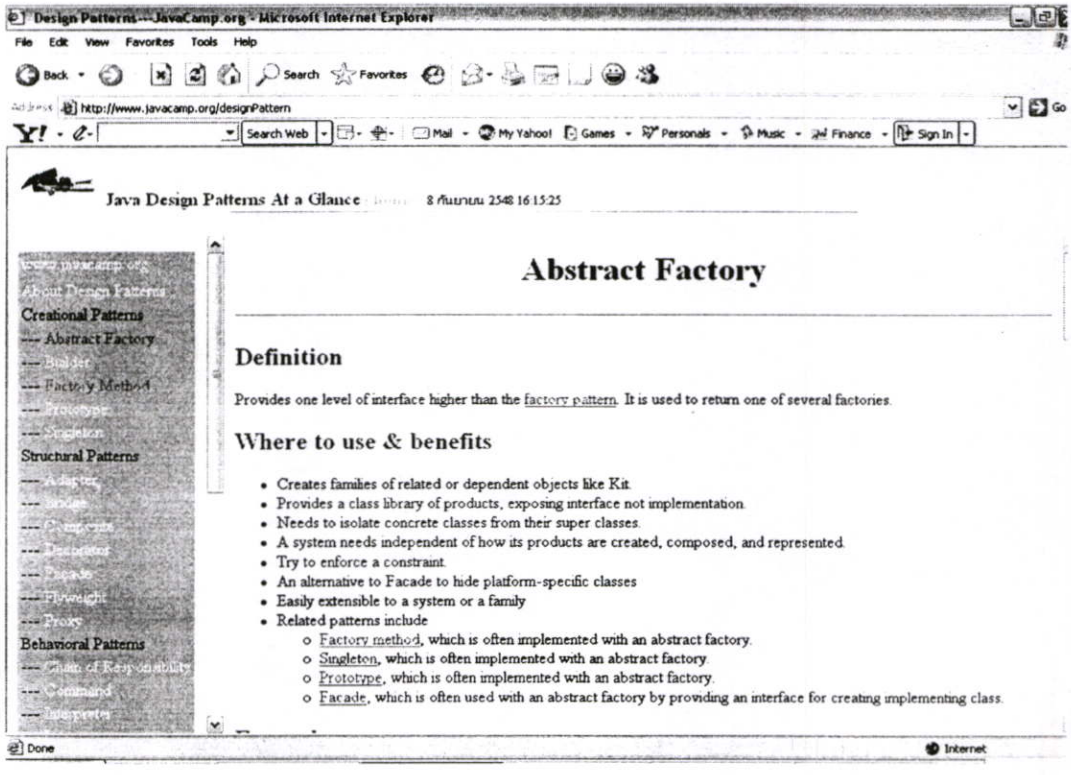
ตัวอย่างที่ 2.5 แสดงการอธิบายรายละเอียดแม่แบบการอธิบายของจีโอเอฟ

- ชื่อ และ หมวดหมู่ของแบบรูป (Pattern Name and Classification) ทุกๆแบบรูปการออกแบบจะต้องมีชื่อ และแต่ละชื่อนั้นจะต้องเป็นหนึ่งเดียว (unique) เพื่อประสิทธิภาพในการอ้างถึง และในแต่ละแบบรูปการออกแบบนั้นจะมีหมวดหมู่เพื่ออธิบายถึงการทำงานโดยทั่วไปของแบบรูปการออกแบบ
- จุดมุ่งหมาย (Intent) คือ ส่วนที่อธิบายถึงเป้าหมายหลักในการทำงานของแบบรูปการออกแบบ
- ชื่อรอง (Also Known As) คือ ชื่ออื่นของแบบรูปการออกแบบที่เป็นที่รู้จัก
- แรงจูงใจ (Motivation) คือ ส่วนที่ยกตัวอย่างสถานการณ์ปัญหาการออกแบบที่แบบรูปการออกแบบสามารถแก้ไขได้
- การประยุกต์ใช้ (Applicability) คือ สถานการณ์ที่แบบรูปการออกแบบสามารถนำไปประยุกต์ใช้ได้
- โครงสร้าง (Structure) คือ ส่วนการแทนแบบรูปการออกแบบด้วยกราฟิกในลักษณะแผนผังคลาส โดยเป็นแผนผังการติดต่อสื่อสารกันระหว่าง คลาส และออบเจ็กต์
- ผู้ร่วมทำงาน (Participants) คือ ส่วนที่อธิบายถึงคลาสและออบเจ็กต์ที่ทำงานในแบบรูปการออกแบบ
- การร่วมทำงาน (Collaboration) คือ ส่วนที่อธิบายว่า คลาส และออบเจ็กต์ทำงานร่วมกันอย่างไร
- ผลที่เกิดขึ้นภายหลัง (Consequence) คือ ส่วนที่บอกถึงข้อดี ข้อเสีย และผลลัพธ์ของการใช้แบบรูปการออกแบบนี้
- การพัฒนา (Implementation) คือ ส่วนที่อธิบายเทคนิคการพัฒนาแบบรูปการออกแบบในรูปแบบโปรแกรม
- ตัวอย่างโปรแกรม (Sample Code) คือ ส่วนที่ยกตัวอย่างการนำหลักการของแบบรูปการออกแบบไปใช้ในการเขียนโปรแกรม
- ตัวอย่างการใช้งานจริง (Known Uses) คือ ส่วนที่ยกตัวอย่างแบบรูปการออกแบบที่ใช้ในการออกแบบระบบจริง

- แบบรูปการออกแบบที่เกี่ยวข้องกัน (Related Patterns) คือ แบบรูปการออกแบบที่เกี่ยวข้องกันในการทำงาน

2.2 เครื่องมือสืบค้นแบบรูปการออกแบบ

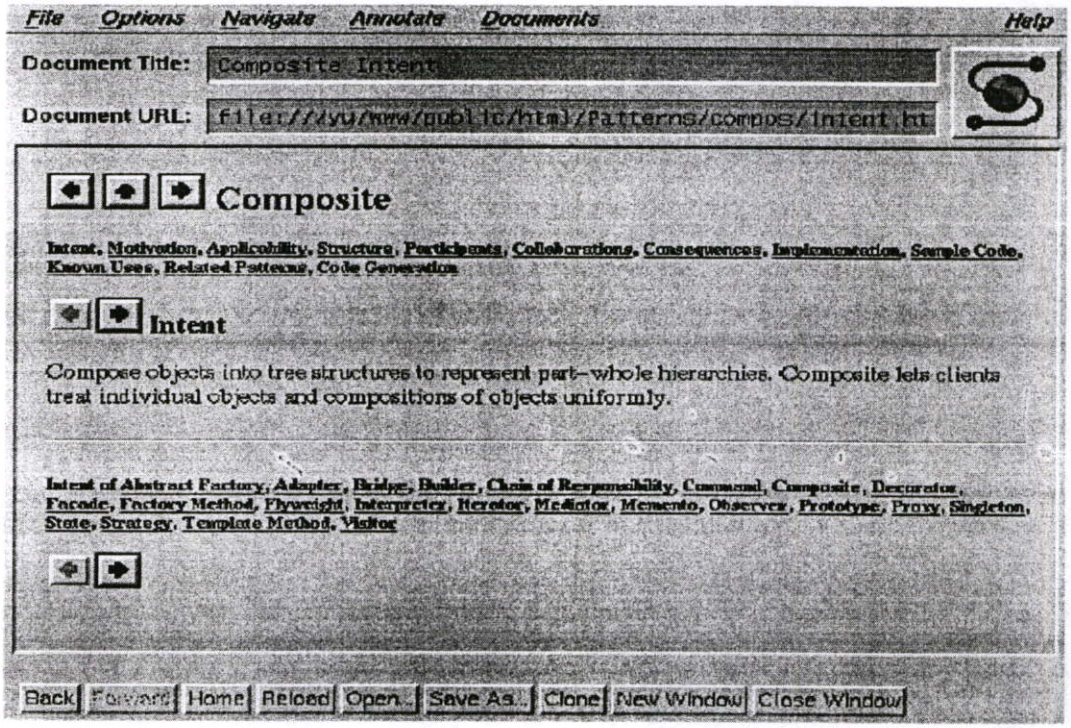
การสืบค้นแบบรูปการออกแบบโดยทั่วไปแล้วจะทำการสืบค้นจากหนังสือที่อธิบายแบบรูปการออกแบบ [6][10][21] แต่การสืบค้นจากหนังสือนี้อาจไม่สะดวกมากนักสำหรับผู้ออกแบบซอฟต์แวร์ เนื่องจากผู้ออกแบบซอฟต์แวร์จะต้องทำการเปิดหนังสือเพื่อหาคำอธิบายแบบรูปการออกแบบจึงเป็นวิธีที่ใช้เวลานาน วิธีที่สามารถสืบค้นได้รวดเร็วกว่าคือการสืบค้นจาก เว็บไซต์ หรือเครื่องมืออิเล็กทรอนิกส์ที่บริการการสืบค้นแบบรูปการออกแบบ รูปที่ 2.5 แสดงตัวอย่างเว็บไซต์ javacamp ที่ให้บริการการสืบค้นแบบรูปการออกแบบ โดยใช้ชื่อของแบบรูปการออกแบบเป็นตัวกลางในการเชื่อมโยงหลายมิติ (Hyperlink) ซึ่งเป็นวิธีที่เว็บไซต์ส่วนใหญ่ใช้ เช่น [25][27][28]



รูปที่ 2.5 แสดงเว็บไซต์ที่ให้บริการสืบค้นแบบรูปการออกแบบ

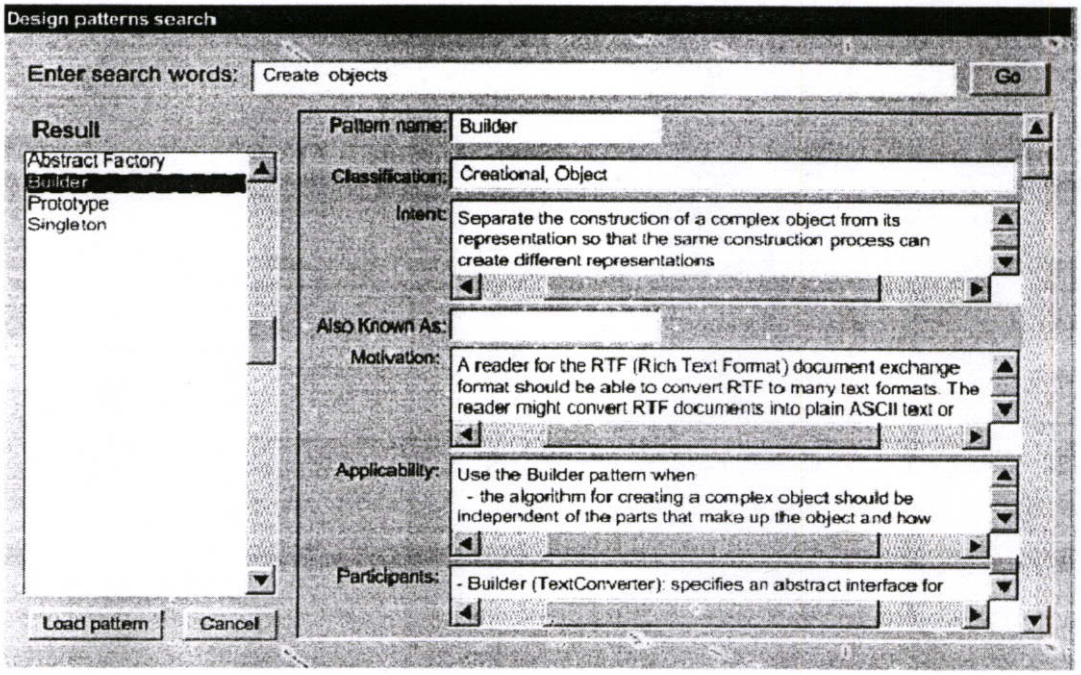
อีกแนวทางหนึ่งที่ใช้ในการสืบค้นแบบรูปการออกแบบ คือ การใช้เครื่องมืออิเล็กทรอนิกส์ ซึ่งมีงานวิจัยที่เกี่ยวข้อง ดังนี้

Frank และคณะ [9] เสนองานวิจัยที่อธิบายถึงสถาปัตยกรรม และการพัฒนาเครื่องมือการสร้างรหัสต้นฉบับ (Source code) ของแบบรูปการออกแบบโดยอัตโนมัติ โดยส่วนหนึ่งของเครื่องมือ คือ ส่วนสืบค้นคำอธิบายจุดมุ่งหมายของแบบรูปการออกแบบ ซึ่งมีวิธีการสืบค้น 2 รูปแบบ คือ แบบเรียงลำดับ (Sequence) และแบบสุ่ม (Randomly) การสืบค้นแบบเรียงลำดับ คือ การกดปุ่มลูกศรเพื่อหาแบบรูปการออกแบบที่ต้องการ ส่วนการสืบค้นแบบสุ่ม คือ การเลือกจากชื่อของแบบรูปการออกแบบที่เป็นตัวเชื่อมโยงหลายมิติ ซึ่งแสดงได้ดังรูปที่ 2.6 [9]



รูปที่ 2.6 แสดงการสืบค้นแบบรูปการออกแบบของ Frank และคณะ

งานวิจัย [7] เสนอเครื่องมือเอ็มวีเคส ซึ่งเป็นเครื่องมือที่ช่วยให้การทำงานร่วมกับแบบรูปการออกแบบทำได้ง่ายขึ้น เช่น การสร้างแบบรูปการออกแบบ การเก็บแบบรูปการออกแบบ และส่วนหนึ่งในการทำงานของเครื่องมือนี้คือการสืบค้นแบบรูปการออกแบบ การสืบค้นของเครื่องมือนี้ใช้ ชื่อ และหมวดหมู่ ของแบบรูปการออกแบบเป็นคำหลักในการสืบค้นแสดง ซึ่งส่วนการสืบค้นแบบรูปการออกแบบแสดงได้ดังรูปที่ 2.7 [7]



รูปที่ 2.7 แสดงส่วนการสืบค้นแบบรูปการออกแบบของเครื่องมือเอ็มวีเคส

การสืบค้นแบบรูปการออกแบบวิธีดังกล่าวมานี้ผู้ใช้จะสามารถทำผ่านชื่อ และหมวดหมู่ของแบบรูปการออกแบบเท่านั้น ซึ่งนักออกแบบซอฟต์แวร์จะต้องมีความรู้เป็นอย่างดีจึงจะสามารถเลือกแบบรูปการออกแบบได้เหมาะสมกับความต้องการ วิธีการหนึ่งที่ช่วยให้การสืบค้นแบบรูปการออกแบบมีความยืดหยุ่นมากขึ้น คือ การสืบค้นโดยนำคำสำคัญอื่นที่สามารถช่วยบ่งชี้ถึงแบบรูปการออกแบบได้ นอกเหนือจากการใช้ชื่อ และหมวดหมู่ โดยคำสำคัญที่สามารถบ่งชี้ได้นี้อาจได้มาจากคำอธิบายแบบรูปการออกแบบ ซึ่งในปัจจุบันมีงานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงสร้างคำอธิบายแบบรูปการออกแบบ เพื่อให้มีความสามารถบ่งชี้ได้ดียิ่งขึ้น ซึ่งรายละเอียดของงานวิจัยจะนำเสนอในหัวข้อที่ 2.3

2.3 งานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงสร้างคำอธิบายแบบรูปการออกแบบ

ในการสร้างแบบจำลองการสืบค้นที่สามารถสืบค้นได้ตามโดเมนของปัญหาการออกแบบขั้นตอนสำคัญขั้นตอนหนึ่ง คือการวิเคราะห์คำที่นำมาเป็นดัชนีในการบ่งชี้เอกสาร งานวิจัยที่จัดรูปแบบคำอธิบายแบบรูปการออกแบบให้มีความสามารถในการบ่งชี้ได้อย่างมีประสิทธิภาพ มีดังต่อไปนี้

Andras และ Mattias [2] ได้เสนอวิธีการจัดรูปประโยคใหม่ให้กับจุดมุ่งหมายในแม่แบบการอธิบายของจีโอเอฟ เพื่อให้มีความสามารถในการบ่งชี้ถึงแบบรูปการออกแบบมากยิ่งขึ้น วิธีการ

คือ ออกแบบโครงสร้างวากยสัมพันธ์ (Syntactic Structure) ในแต่ละหมวดหมู่ของแบบรูปการออกแบบ และจัดแบ่งประโยคที่เป็นจุดมุ่งหมายใหม่ ตามความหมาย (Semantic) ของประโยคโดยการจัดแบ่งประโยคนี้จะยึดหลักโครงสร้างวากยสัมพันธ์ที่ออกแบบ ซึ่งโครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้าง หมวดหมู่โครงสร้าง และหมวดหมู่พฤติกรรม แสดงได้ดังนี้

1) โครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้าง แสดงได้ดังตารางที่ 2.2

ตาราง 2.2 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้าง

Negation	**Create**	Entity	Action	Force	Supplementary information

จากตารางที่ 2.2 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้าง ซึ่งโครงสร้างวากยกสัมพันธ์นี้จะประกอบไปด้วย

- Negationคือ ส่วนที่ใช้ในการอธิบายแบบรูปการออกแบบที่มีจุดมุ่งหมายตรงข้ามกับแบบรูปการออกแบบส่วนใหญ่ที่อยู่ในหมวดหมู่เดียวกัน ดังตัวอย่างที่ 2.7
- Create คือ ส่วนที่บ่งบอกการทำงานหลักของแบบรูปการออกแบบ
- Entity คือ ส่วนที่เป็นเอนทิตี (Entity)ที่ทำงานในแบบรูปการออกแบบ
- Action คือ ส่วนที่เป็นการกระทำ (Action) ของแบบรูปการออกแบบ
- Force คือ ข้อบังคับในการทำงานของแบบรูปการออกแบบ
- Supplementary information คือ ข้อมูลเสริมในการทำงานของแบบรูปการออกแบบ

โดยตัวอย่างที่ 2.6 การดึงความหมายจากประโยคจุดมุ่งหมายของแบบรูปการออกแบบแอบสแทรกต์แฟกทอรีลงโครงสร้างวากยสัมพันธ์ของหมวดหมู่การสร้างซึ่งการจัดโครงสร้างวากยสัมพันธ์ของแบบรูปการออกแบบซิงเกิลตอนนั้นจะมีประโยคที่ดัดแปลงไม่ตรงกับของจีโอเอฟนัก อย่างไรก็ตามความหมายของประโยคยังคงสื่อความหมายไม่ต่างจากเดิมซึ่งแสดงได้ดังตัวอย่างที่ 2.7

ตัวอย่างที่ 2.6

ประโยคจุดมุ่งหมายของแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี คือ

Provide an interface for creating families of related or dependent object without specifying their concrete classes. [10]

จุดมุ่งหมายของแอบสแทรกต์แฟกทอรี ข้างต้น สามารถดึงความหมายของประโยคมาจัดวางตามโครงสร้างวากยสัมพันธ์ ได้ดังตารางที่ 2.3 [2]

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ตาราง 2.3 แสดงโครงสร้างใหม่ของแอบสแทรกต์แฟกทอรี

Create	Entity	Action	Force
Create	Families of related or dependent object	By provide an interface for doing so	Without specifying their concrete class

ซึ่งได้คำอธิบายใหม่ของแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี คือ Create families of related or dependent objects by providing an interface for doing so, without specifying their concrete class.

ตัวอย่างที่ 2.7

ประโยคจุดมุ่งหมายของแบบรูปการออกแบบซิงเกิลตอน คือ

Ensure a class only has one instance, and provide a global point of access to it [10]

การอธิบายโดยตรงในประโยคจุดมุ่งหมายของซิงเกิลตอนนี้อาจไม่ชัดเจนมากนัก ซึ่งถ้าใช้การอธิบายแบบตรงกันข้ามคือ นำเอา Negation มาใช้ในการอธิบายจะทำให้สามารถเข้าใจประโยคจุดมุ่งหมายได้ดียิ่งขึ้น ซึ่งสามารถเขียนเป็นโครงสร้างวากยสัมพันธ์ ได้ดังตารางที่ 2.4 [2]

ตาราง 2.4 แสดงโครงสร้างใหม่ของซิงเกิลตอน

Negation	**Create**	Entity	Action
Do not	Create	More than one instance of a class	Provide a global point of access to it

คำอธิบายใหม่ของแบบรูปการออกแบบซิงเกิลตอน คือ Do not create more than one instance of a class. Provide a global point of access to it.

2) โครงสร้างวากยสัมพันธ์ของหมวดหมู่โครงสร้าง แสดงได้ดังตารางที่ 2.5

ตาราง 2.5 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่โครงสร้าง

Action to Apply	Entity	Outcome	Result

จากตารางที่ 2.5 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่โครงสร้าง ซึ่งโครงสร้างวากยสัมพันธ์นี้จะประกอบไปด้วย

63452

- Action to Apply คือ การกระทำที่ทำกับเอนทิตี
- Entity คือ ส่วนที่เป็นเอนทิตี (Entity) ที่ทำงานในแบบรูปการออกแบบ
- Outcome คือ ส่วนที่เป็นผลลัพธ์ที่เกิดขึ้นจากการกระทำของแบบรูปการออกแบบ
- Result คือ ผลลัพธ์ที่ได้จากการทำงานของแบบรูปการออกแบบ

โดยตัวอย่างที่ 2.8 การดึงความหมายจากประโยคจุดมุ่งหมายของแบบรูปการออกแบบอะแดปเตอร์ลงโครงสร้างวากยสัมพันธ์ของหมวดหมู่โครงสร้าง

ตัวอย่างที่ 2.8

ประโยคจุดมุ่งหมายของแบบรูปการออกแบบอะแดปเตอร์ คือ

Convert the interface of a class into another interface clients expect. Adaptee lets classes work together that couldn't otherwise because of incompatible interfaces. [10]

จุดมุ่งหมายของอะแดปเตอร์ข้างต้น สามารถดึงความหมายของประโยคมาจัดวางตามโครงสร้างวากยสัมพันธ์ ได้ดังตารางที่ 2.6 [2]

ตาราง 2.6 แสดงโครงสร้างใหม่ของอะแดปเตอร์

Action to Apply	Entity	Outcome	Result
Convert	The interface of class	Into another interface client expect	Adapter let classes work together that couldn't otherwise because of incompatible interfaces.

คำอธิบายใหม่ของแบบรูปการออกแบบอะแดปเตอร์ คือ Convert the interface of a class into another interface clients expect. Adapter lets classes work together that couldn't otherwise because of incompatible interface

3) โครงสร้างวากยสัมพันธ์ของหมวดหมู่พฤติกรรม แสดงได้ดังตารางที่ 2.7

ตารางที่ 2.7 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่พฤติกรรม

Objective	Course of action / proceedings	Result

จากตารางที่ 2.7 แสดงโครงสร้างวากยสัมพันธ์ของหมวดหมู่พฤติกรรม ซึ่งโครงสร้างนี้จะประกอบไปด้วย

- Objective คือ จุดประสงค์การทำงานของแบบรูปการออกแบบ
- Course of action / proceedings คือ ลำดับการทำงานของแบบรูปการออกแบบ
- Result คือ ผลลัพธ์ที่ได้จากการทำงานของแบบรูปการออกแบบ

โดยตัวอย่างที่ 2.9 การดึงความหมายจากประโยคจุดมุ่งหมายของแบบรูปการออกแบบออบเซิฟเวอร์ ลงโครงสร้างวากยสัมพันธ์ของหมวดหมู่พฤติกรรม

ตัวอย่างที่ 2.9

ประโยคจุดมุ่งหมายของแบบรูปการออกแบบออบเซิฟเวอร์ คือ

Define a one-to-many dependency between objects so that when one object changes state, all its dependents are notified and updated automatically. [10]

จุดมุ่งหมายของออบเซิฟเวอร์ข้างต้น สามารถดึงความหมายของประโยคมาจัดวางตามโครงสร้างวากยสัมพันธ์ ได้ดังตารางที่ 2.8 [2]

ตาราง 2.8 แสดงโครงสร้างใหม่ของออบเซิฟเวอร์

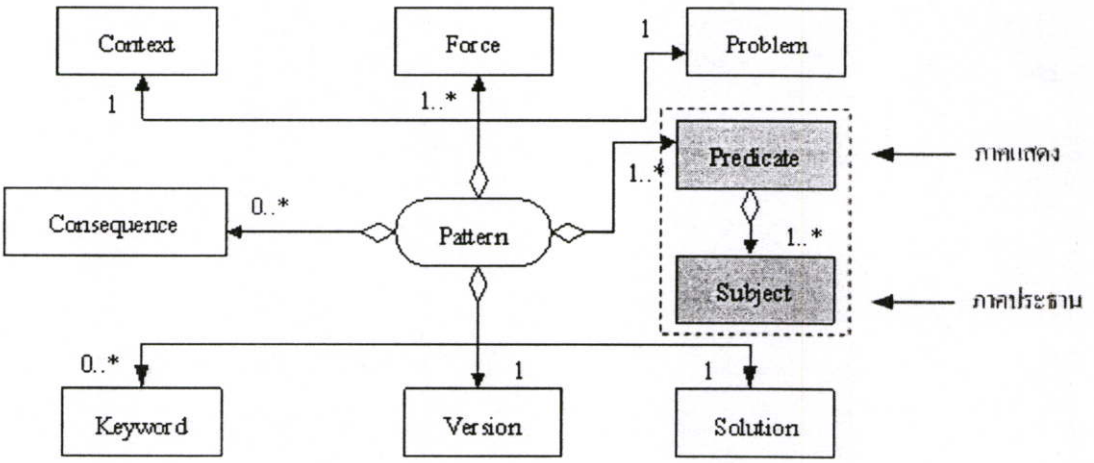
Objective	Course of action / proceedings
All dependencies are notified and updated automatically when one object changes sate	By defining a one-to-many relationship between objects.

คำอธิบายใหม่ของแบบรูปการออกแบบออบเซิฟเวอร์ คือ All dependencies are notified and updated automatically when one object changes state, by defining a one-to-many relationship between objects.

ซึ่งการจัดโครงสร้างการอธิบายแบบรูปการออกแบบใหม่นี้จะช่วยลดความคลุมเครือในคำอธิบายแบบรูปการออกแบบ และทำให้ผู้ใช้สามารถเข้าใจการทำงานของแต่ละแบบรูปได้ง่ายและรวดเร็วยิ่งขึ้น

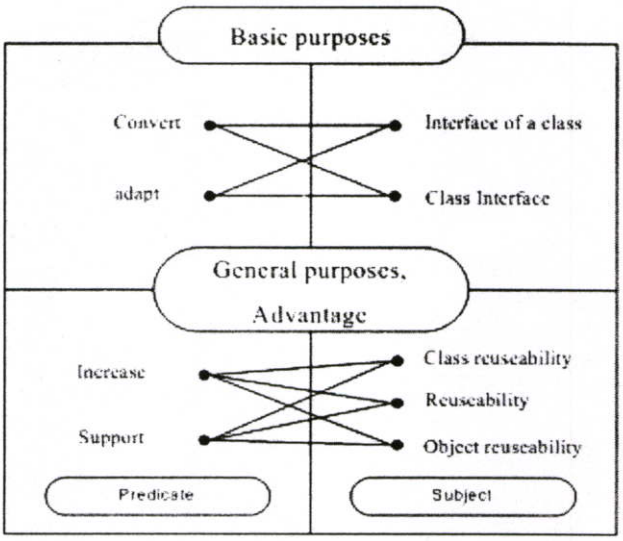
Objectventure [16] ได้นำเสนอพีซีเอ็มแอล (PCML : Pattern and Component Markup Languages) ซึ่งเป็นภาษาที่ใช้ในการอธิบายแบบรูปการออกแบบ ที่ใช้ภาษาเอ็กซ์เอ็มแอล (XML) เป็นภาษากลางในการพัฒนา พีซีเอ็มแอลจะอธิบายแบบรูปการออกแบบเป็นแบบจำลองเมทา (meta model) ซึ่งช่วยให้การอธิบายแบบรูปการออกแบบเป็นมาตรฐาน และทำให้สามารถเข้าใจการทำงานของแบบรูปการออกแบบมากขึ้น

Daniel และ Gyorgy [8] นำเสนอวิธีการประยุกต์แบบจำลองเมทาพีซีเอ็มแอล ให้มีความสามารถในการสืบค้นแบบรูปการออกแบบได้เจาะจงมากยิ่งขึ้น โดยการเพิ่มในส่วนของภาคแสดง (Predicate) และ ภาคประธาน (Subject) เข้ากับแบบจำลองเมทาพีซีเอ็มแอล ซึ่งจะทำให้แบบจำลองการอธิบายแบบรูปการออกแบบที่ได้นั้นสามารถอธิบายปัญหาและจุดมุ่งหมายของแบบรูปการออกแบบได้เจาะจงได้มากยิ่งขึ้น แสดงได้ดังรูปที่ 2.8



รูปที่ 2.8 แสดงการประยุกต์แบบจำลองเมทาพีซีเอ็มแอล

รูปที่ 2.8 แสดงการประยุกต์แบบจำลองเมทาพีซีเอ็มแอลโดยการเพิ่มในส่วนของภาคแสดงและภาคประธาน โดยภาคแสดง คือ คำกริยาหลัก (Key verb) ที่สื่อความหมายถึงการทำงานของแบบรูปการออกแบบ และภาคประธานคือ คำที่เป็นประธานในการทำงานของภาคแสดง คำเหล่านี้ได้มาจากการสรุป (Summary) คำอธิบายปัญหา และจุดมุ่งหมายของแบบรูปการออกแบบ โดยภาคแสดงนั้นจะมีความสัมพันธ์ต่อแบบรูปการออกแบบ คือ 1...* นั่นหมายความว่าแบบรูปการออกแบบหนึ่งสามารถมีคำที่เป็นภาคแสดงได้หลายคำ และภาคประธานสัมพันธ์กับภาคแสดงแบบ 1...* นั่นหมายความว่า ภาคประธานสามารถสัมพันธ์กับคำที่เป็นภาคแสดงได้หลายคำ เช่นกัน โดยตัวอย่างการอธิบายแบบรูปการออกแบบอะแดปเตอร์โดยใช้วิธีของ Daniel และ Gyorgy แสดงได้ดังรูปที่ 2.9 [8]



รูปที่ 2.9 แสดงตัวอย่างความสัมพันธ์ภาคแสดงและภาคประธานของอะแดปเตอร์

รูปที่ 2.9 แสดงการการอธิบายแบบรูปการออกแบบอะแดปเตอร์ โดย Daniel และ Gyorgy ได้สรุปจุดประสงค์การทำงานของอะแดปเตอร์ออกเป็น 2 ส่วน คือ จุดประสงค์พื้นฐาน และจุดประสงค์ทั่วๆไปของประโยชน์ในการใช้แบบรูปการออกแบบอะแดปเตอร์ ซึ่งในแต่ละส่วนจะได้มาจากการสรุปจุดมุ่งหมายการทำงานของแบบรูปการออกแบบ โดยจุดมุ่งหมายนี้จะถูกสรุปเป็นจุดประสงค์พื้นฐาน คือ มีคำกริยา “Convert” และ “Adapt” และมีประธาน คือ “Class interface” และ “Interface of a class” เมื่อจัดความสัมพันธ์ของประธาน และกริยาที่มีในจุดประสงค์นี้จะได้ว่าแบบรูปการออกแบบอะแดปเตอร์จะทำการ “Convert class interface” หรือ “Convert interface of a class” เป็นต้น

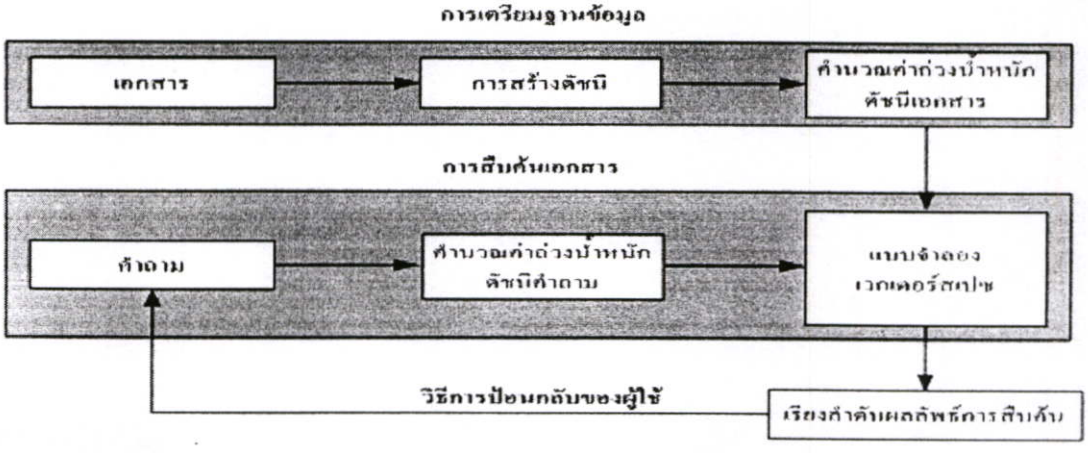
การเพิ่มส่วนของภาคแสดงและภาคประธานในแบบจำลองเมทาพีซีเอ็มแอลนี้ทำให้ความสามารถในการสืบค้นแบบรูปการออกแบบมากขึ้น เนื่องจากดัชนีที่ใช้ในการสืบค้นมีความเจาะจงถึงแบบรูปการออกแบบมากยิ่งขึ้น อย่างไรก็ตามแบบจำลองเมทาทำการเก็บข้อมูลทั่วไปของการอธิบายแบบรูปการออกแบบ จึงเป็นการยากที่จะสามารถระบุคำที่จะใช้เป็นดัชนีคำค้นแบบรูปการออกแบบนั้น ดังนั้นถ้าสามารถจัดคำอธิบายแบบรูปการออกแบบนี้โดยใช้แนวความคิดของงานวิจัย [2] ซึ่งทำการจัดแบ่งคำอธิบายออกเป็นส่วนๆ ตามความหมายของประโยค ก็จะทำให้สามารถระบุคำที่จะใช้เป็นดัชนีคำค้นแบบรูปการออกแบบได้เจาะจงมากยิ่งขึ้น

บทที่ 3
การสืบค้นระบบสารสนเทศ

ในบทนี้จะกล่าวถึงทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการสืบค้นระบบสารสนเทศ ซึ่งแบ่งการนำเสนอออกเป็น 2 ส่วน โดยส่วนแรกจะนำเสนอทฤษฎีพื้นฐานในกระบวนการสืบค้นระบบสารสนเทศ และวิธีการประเมินระบบสืบค้นสารสนเทศ ส่วนที่สองจะนำเสนองานวิจัยที่เกี่ยวข้องกับการออกแบบการคำนวณค่าถ่วงน้ำหนักและจัดการโครงสร้างเอกสารเพื่อการสืบค้นในระบบสารสนเทศ

3.1 ความรู้ทั่วไปเกี่ยวกับการสืบค้นระบบสารสนเทศ (Information Retrieval)

การสืบค้นระบบสารสนเทศ คือ กระบวนการที่เกี่ยวข้องกับการจัดเก็บข้อมูล การเข้าถึงข้อมูลและรูปแบบการแทนข้อมูล จุดประสงค์หลักของกระบวนการสืบค้นสารสนเทศ คือ การสืบค้นเอกสารให้ตรงตามความต้องการของผู้ใช้มากที่สุด โดยผู้ใช้เป็นผู้ป้อนคำถาม (Query) แล้วกระบวนการสืบค้นสารสนเทศจะทำการคำนวณค่าความคล้ายคลึง (Similarity) ระหว่างคำถามและเอกสาร ผลลัพธ์ที่ได้จะเป็นรายการเอกสารที่เกี่ยวข้องกับคำถามซึ่งเรียงตามลำดับค่าความคล้ายคลึง หลังจากนั้นนำผลลัพธ์ที่ได้เข้าสู่วิธีการป้อนกลับของผู้ใช้ (Relevance Feedback) เพื่อทำการปรับปรุงคำถามให้สามารถสืบค้นได้ผลลัพธ์ที่มีความแม่นยำมากขึ้น ซึ่งกระบวนการสืบค้นสารสนเทศแสดงได้ดังรูปที่ 3.1 โดยประกอบด้วย 3 ส่วน คือ ส่วนการเตรียมฐานข้อมูล การสืบค้นข้อมูล และ การป้อนกลับของผู้ใช้ ซึ่งรายละเอียดในแต่ละกระบวนการอธิบายดังหัวข้อ 3.1.1, 3.1.2 3.1.3 และรายละเอียดวิธีประเมินระบบสืบค้นสารสนเทศ อธิบายดังหัวข้อที่ 3.1.4 [3][13]



รูปที่ 3.1 แสดงกระบวนการสืบค้นสารสนเทศ

3.1.1 การเตรียมฐานข้อมูล

ในส่วนของการเตรียมฐานข้อมูล เป็นกระบวนการจัดการกับเอกสารเพื่อให้พร้อมสำหรับการสืบค้น ซึ่งกระบวนการนี้ประกอบด้วย ขั้นตอนการเตรียมเอกสาร ขั้นตอนการสร้างดัชนี และขั้นตอนการคำนวณค่าถ่วงน้ำหนักของดัชนี ซึ่งรายละเอียดได้อธิบายไว้ดังหัวข้อที่ 3.1.1.1, 3.1.1.2 และ 3.1.1.3

3.1.1.1 ขั้นตอนการเตรียมเอกสาร (Document Processing)

ขั้นตอนนี้ทำการเตรียมเอกสารเพื่อให้เหมาะสมในการเป็นดัชนี โดยประกอบด้วยการวิเคราะห์คำศัพท์ การหาสต็อปเวิร์ด และการหารากศัพท์ ซึ่งมีรายละเอียดดังนี้

1) การวิเคราะห์คำศัพท์ (Lexical Analysis) คือ ขั้นตอนที่เป็นการกำจัดส่วนไม่จำเป็นในประโยค เช่น เว้นวรรค ย่อหน้า เครื่องหมายขีดกลาง (-) และเครื่องหมายจุด (.) เป็นต้น ซึ่งทั้งหมดนี้มีจะมีผลต่อประสิทธิภาพการสืบค้น

2) การหาสต็อปเวิร์ด (Stop words) คือ ขั้นตอนในการหากลุ่มคำที่ไม่ต้องการนำมาเป็นดัชนีในระบบสืบค้นสารสนเทศ เช่น a, also, and, as, be และ but เป็นต้น คำเหล่านี้ไม่มีประโยชน์ในการสืบค้น จึงควรกำจัดออกเพื่อช่วยลดขนาดของแฟ้มดัชนี (Invert Index File) ข้อเสีย คือ จะไม่สามารถค้นหาคำซึ่งเป็นสต็อปเวิร์ดได้

3) การหารากศัพท์ (Stemming) คือ การรวมคำที่มีรากศัพท์เดียวกันไว้ด้วยกัน ซึ่งคำที่มีรากศัพท์เดียวกันอาจแตกต่างกันที่ คำเติมหน้า (Prefix) หรือ คำต่อท้าย (Suffix) ในรูปแบบของรากศัพท์ทำให้สามารถมีคำสำคัญได้แค่หนึ่งคำ เช่น คำว่า Computer เป็นรากศัพท์ของ Computed, Computers และComputing เป็นต้น การใช้ขั้นตอนวิธีการหารากศัพท์นี้จะช่วยลดขนาดของดัชนี

3.1.1.2 ขั้นตอนการสร้างดัชนี (Indexing)

หลังจากที่เอกสารผ่านขั้นตอนการเตรียมเอกสารแล้ว ขั้นตอนต่อไป คือ การเลือกคำสำคัญ (Selection index) เพื่อนำมาเป็นดัชนีในการบ่งชี้เอกสาร โดยคำสำคัญเหล่านี้อาจนำมาเป็นดัชนีเดี่ยว (Single Index) หรือ เป็นดัชนีวลี (Phrase Index)

1) **ดัชนีเดี่ยว** คือ ดัชนีซึ่งเป็นคำเดียว เช่น Computer และInformation เป็นต้น

2) **ดัชนีวลี** คือ ดัชนีที่นำดัชนีเดี่ยวมาประกอบกัน เช่น Information Retrieval และ Design Pattern เป็นต้น

การเลือกคำสำคัญเพื่อนำมาสร้างเป็นดัชนีนี้เป็นขั้นตอนที่สำคัญในกระบวนการสืบค้นสารสนเทศ เนื่องจากเป็นขั้นตอนที่มีผลกระทบต่อประสิทธิภาพการสืบค้นโดยตรง การเลือกคำสำคัญสามารถทำได้หลักๆ 2 วิธี คือการเลือกโดยให้ผู้เชี่ยวชาญเป็นคนกำหนด (Manually Index) และ การเลือกคำสำคัญแบบอัตโนมัติ (Automatic Index) โดย [13] กล่าวว่าการเลือกคำสำคัญจากผู้เชี่ยวชาญนี้มีข้อดี คือ สามารถควบคุมคำที่ต้องการนำมาเป็นคำสำคัญได้ ทำให้มีความแม่นยำในการสืบค้นสูง ข้อเสีย คือ วิธีการนี้จะใช้เวลาในการทำงานสูง ส่วนวิธีเลือกคำสำคัญแบบอัตโนมัติจะเป็นการเลือกคำสำคัญโดยใช้เครื่องมือ จึงไม่สามารถควบคุมคำที่นำมาเป็นคำสำคัญได้ ส่วนระยะเวลาที่ใช้ในการเลือกคำสำคัญจะแน่นอนและรวดเร็วกว่า อย่างไรก็ตามวิธีการเลือกคำสำคัญจากเชี่ยวชาญจะให้ประสิทธิภาพการสืบค้นที่ดีกว่าแบบอัตโนมัติ

3.1.1.3 ขั้นตอนการคำนวณค่าถ่วงน้ำหนักของดัชนี (Term Weighting)

การคำนวณหาค่าถ่วงน้ำหนักของดัชนี คือ การคำนวณค่าถ่วงน้ำหนักของดัชนีแต่ละคำในแต่ละเอกสาร ค่าถ่วงน้ำหนักของดัชนีนี้จะเป็นค่าที่สามารถบ่งชี้ถึงความสำคัญของดัชนีในแต่ละเอกสารได้ว่าดัชนีนั้นมีความสำคัญต่อเอกสารมากน้อยเพียงใด ซึ่งตัวอย่างการคำนวณค่าถ่วงน้ำหนักของดัชนี เช่น การคำนวณค่าถ่วงน้ำหนักโดยคำนวณจากผลคูณของจำนวนความถี่ของคำ (Term Frequency – tf) กับค่าความถี่ผกผันของเอกสาร (Inverse Document Frequency – idf) โดยการคำนวณค่าความถี่ผกผันนี้จะใช้หลักความเป็นจริงที่ว่าดัชนีใดปรากฏในเอกสารมากฉบับจะต้องมีความสำคัญน้อยกว่าคำที่ปรากฏในเอกสารน้อยฉบับ เนื่องจากดัชนีที่ปรากฏอยู่น้อยฉบับจะมีคุณสมบัติการบ่งชี้มากกว่า โดยมีสมการที่ใช้ในการคำนวณน้ำหนักความถี่ผกผันของเอกสารดังสมการที่ 3.1 [13]

$$idf_k = \log\left(\frac{N}{n_k}\right) \tag{3.1}$$

เมื่อ idf_k แทน ค่าความถี่ผกผันของเอกสารของดัชนีตัวที่ k เมื่อ $k \in \{1, 2, 3, ..., M\}$

M แทน จำนวนดัชนีทั้งหมด

N แทน จำนวนเอกสารทั้งหมด

n_k แทน จำนวนเอกสารที่เจอดัชนี ตัวที่ k

วิธีการคำนวณค่าถ่วงน้ำหนักของดัชนีสามารถแสดงได้ดังสมการที่ 3.2

$$w_{ik} = tf_{ik} * idf_k \tag{3.2}$$
$$w_{ik} = tf_{ik} * \log\left(\frac{N}{n_k}\right)$$

เมื่อ w_{ik} แทน น้ำหนักของดัชนี ตัวที่ k ในเอกสารลำดับที่ i

tf_{ik} แทน ความถี่ของดัชนี ตัวที่ k ในเอกสารลำดับที่ i

$k \in \{1, 2, 3, ..., M\}$ และ $i \in \{1, 2, 3, ..., N\}$

เมื่อทำการเตรียมฐานข้อมูลเรียบร้อยแล้วขั้นตอนต่อไป คือ การสืบค้นเอกสาร ขั้นตอนนี้จะนำดัชนีที่ได้จากขั้นตอนการเตรียมฐานข้อมูลมาทำการแทนเข้ากับแบบจำลองการสืบค้นเพื่อทำการคำนวณหาผลลัพธ์การสืบค้น โดยแบบจำลองการสืบค้นนี้จะทำการแทนข้อมูลของเอกสารในรูปแบบต่างๆ ซึ่งมีรายละเอียดดังหัวข้อที่ 3.1.2

3.1.2การสืบค้นเอกสาร

ในระบบสืบค้นสารสนเทศมีแบบจำลองที่ทำการสืบค้นหลายรูปแบบ [13] โดยแบบจำลองที่เป็นต้นแบบในการสืบค้น และยังคงนิยมอยู่ในปัจจุบัน คือแบบจำลองคลาสสิค (Classic Model) ซึ่งประกอบไปด้วย แบบจำลองบูลีน (Boolean Model) แบบจำลองความน่าจะเป็น (Probabilistic Model) และ แบบจำลองเวกเตอร์สเปซ (Vector Space Model) โดยมีรายละเอียดของแต่ละแบบจำลอง ดังต่อไปนี้

3.1.2.1 แบบจำลองบูลีน เป็นแบบจำลองที่สืบค้นโดยใช้แนวความคิดเชิงตรรกะหรือพีชคณิตบูลีน มาสืบค้นร่วมกับคำสำคัญที่เป็นคำถาม ซึ่งตรรกะหรือพีชคณิตบูลีนที่ใช้ในการสืบค้น คือ และ (AND) หรือ (OR) ไม่ (NOT) โดยแต่ละคำมีความหมาย ดังนี้

- AND คือการกำหนดว่าต้องมีคำสำคัญทั้งสองคำปรากฏอยู่ในเอกสาร
- OR คือการกำหนดว่ามีอย่างน้อยหนึ่งคำปรากฏอยู่ในเอกสาร
- NOT คือการกำหนดว่าคำที่ระบุต้องไม่ปรากฏอยู่ในเอกสาร

แบบจำลองบูลีนนี้จะใช้ค่า 0 และ 1 ในการเป็นตัวแทนการปรากฏของดัชนีในเอกสาร ซึ่ง 0 คือ ไม่มีดัชนีปรากฏในเอกสาร และ 1 คือมีดัชนีปรากฏในเอกสาร ดังนั้นลักษณะการสืบค้นจึงมีคำตอบคือ ใช่ กับ ไม่ใช่ เท่านั้น ซึ่งจะไม่มีการจัดลำดับเอกสารที่สามารถสืบค้นมาได้และไม่มีการเปรียบเทียบความคล้ายคลึงกันระหว่างเอกสารและคำถาม

3.1.2.2 แบบจำลองความน่าจะเป็น เป็นแบบจำลองที่ใช้ทฤษฎีความน่าจะเป็นมาช่วยในการสืบค้นหาผลลัพธ์ ซึ่งแบบจำลองความน่าจะเป็นนี้จะทำการสมมติค่าเริ่มต้น โดยทั่วไปจะใช้ค่า 0.5 จากนั้นเมื่อทำการสืบค้นเอกสารที่ได้จากการสืบค้นจะถูกสมมติให้ว่าจะมีความเหมาะสม (Relevance) กับคำถามด้วยค่าความน่าจะเป็นสูงกว่าค่าเริ่มต้นที่ตั้งไว้ ซึ่งค่าความน่าจะเป็นที่เอกสารได้จากการสืบค้นต้องเหมาะสมกับคำถาม และค่าความน่าจะเป็นไม่จำเป็นต้องรวมกันเป็น 1 ซึ่งการสืบค้นโดยใช้แบบจำลองความน่าจะเป็นนี้จะใช้จำนวนความถี่ของคำในการคำนวณความน่าจะเป็นของแต่ละคำ และความถี่ของคำจะสามารถนำมาใช้คาดการณ์ความน่าจะเป็นของเอกสารที่มีกลุ่มของคำสำคัญอยู่ได้

จากการศึกษาแบบจำลองข้างต้นจะพบว่าแบบจำลองบูลีนนั้นเป็นแบบจำลองที่เข้าใจง่าย และมีความแม่นยำในการสืบค้น แต่ไม่สามารถหาความคล้ายคลึงและไม่มีการจัดลำดับเอกสารที่สืบค้นมาได้ ส่วนแบบจำลองความน่าจะเป็นนั้น เป็นแบบจำลองที่ยากต่อการพัฒนาและใช้จำนวนความถี่ของคำในการคำนวณความน่าจะเป็นของแต่ละคำ จึงไม่เหมาะสำหรับระบบสืบค้นสารสนเทศที่ใช้โครงสร้างของเอกสารเป็นตัวกำหนดค่าถ่วงน้ำหนักของเอกสาร ซึ่งแบบจำลองที่เหมาะสำหรับการสืบค้นเอกสารในงานวิจัยนี้คือแบบจำลองเวกเตอร์สเปซเนื่องจากเป็นแบบจำลองที่มีการใช้คณิตศาสตร์เรียบง่ายในการคิดมีการพิจารณาค่าน้ำหนักของคำที่ละเอียด และสามารถจัดลำดับผลการสืบค้นของเอกสารได้ โดยมีรายละเอียดดังต่อไปนี้

3.1.2.3 แบบจำลองเวกเตอร์สเปซ เป็นแบบจำลองทางคณิตศาสตร์ที่ใช้หลักการของเวกเตอร์ แทนเอกสารและคำถามของผู้ใช้ โดยแกนแต่ละแกนของเวกเตอร์ คือ ดัชนีที่ปรากฏอยู่ในเอกสาร และใช้หลักการค่าถ่วงน้ำหนักดัชนีเข้ามาช่วยในการกำหนดขนาดของเวกเตอร์ แบบจำลองเวกเตอร์สเปซนี้จะประกอบไปด้วยเวกเตอร์เอกสาร (Document Vector) และเวกเตอร์ของคำถาม (Query Vector) ซึ่งแต่ละเวกเตอร์จะประกอบไปด้วยขนาดและทิศทาง ถ้าเวกเตอร์เอกสารใดชี้ไปในทิศทางใกล้เคียงกับเวกเตอร์คำถาม นั่นแสดงว่าเอกสารนั้นมีเนื้อหาใกล้เคียงกับคำถามและควรที่จะเป็นคำตอบสำหรับคำถามนั้น วิธีการเปรียบเทียบเวกเตอร์ระหว่างเวกเตอร์เอกสารกับเวกเตอร์คำถาม สามารถทำได้โดยการคำนวณค่าความคล้ายคลึง (Similarity) ระหว่างเอกสารและคำถามโดยทำการคูณภายใน (Dot Product) ระหว่างเวกเตอร์ ผลลัพธ์ที่ได้เป็นค่าตัวเลขของแต่ละเอกสาร ค่าตัวเลขนี้บ่งบอกถึงว่าเอกสารกับคำถามมีเวกเตอร์ที่ชี้ไปในทิศทางเดียวกันมากน้อยแค่ไหน หรือเอกสารมีเนื้อหาที่ตรงกับคำถามมากน้อยเท่าใด การหาค่าความคล้ายคลึง แสดงได้ดังสมการที่ 3.3

$$Sim(q,d_i) = \frac{\sum_{k=1}^{N} w_{q,k} * w_{i,k}}{\sqrt{\sum_{k=1}^{N} (w_{q,k})^2 * \sum_{k=1}^{N} (w_{i,k})^2}} \qquad (3.3)$$

เมื่อ $w_{i,k}$ แทน ค่าถ่วงน้ำหนักของดัชนีที่ k ในเอกสารลำดับที่ i

$w_{q,k}$ แทน ค่าถ่วงน้ำหนักของดัชนีที่ k ในคำถาม

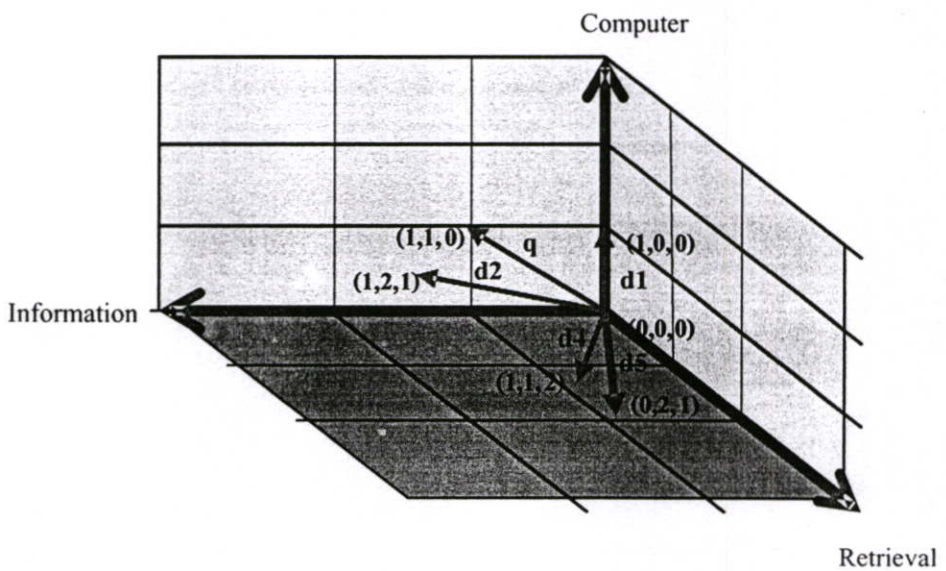
$Sim(q,d_i)$ แทน ค่าความคล้ายคลึงระหว่างคำถามและเอกสารลำดับที่ i

$i \in \{1, 2, 3, ..., N\}$ เมื่อ N แทน จำนวนเอกสารทั้งหมด

ตัวอย่าง 3.1 การใช้เวกเตอร์ในการแทนเอกสารและคำถาม สามารถแสดงได้ดังรูปที่ 3.2 ก) และ ข) โดย ก) แสดงตัวอย่างเอกสาร (d1, d2, d3, d4) และคำถาม (q) ที่ใช้ในการสืบค้น และ ข) แสดงการแทนดัชนีของเอกสารและคำถามเป็นเวกเตอร์

d1	Computer system
d2	Computer Information retrieve information
d3	Retrieve document by information retrieval
d4	Information Retrieval is retrieve data of computer
q	Computer Information

ก) ตัวอย่างเอกสารและคำถาม



ข) เวกเตอร์ของดัชนีในเอกสารและ คำถาม

รูปที่ 3.2 แสดงการแทนของดัชนี และคำสำคัญที่ใช้ในการสืบค้นเป็นเวกเตอร์

จากตัวอย่างข้อมูลข้างต้นมีคำสำคัญที่เป็นดัชนีอยู่ 3 คำ คือ "Computer" "Information" และ "Retrieval" ดังนั้นแกนของเวกเตอร์แบ่งออกเป็น 3 แกนตามดัชนี โดยแต่ละเวกเตอร์ที่แสดงใน ข) เป็นเวกเตอร์ของเอกสาร d1, d2, d3, d4 และ เวกเตอร์คำถาม q ซึ่งมีขนาดตามค่าถ่วงน้ำหนักของดัชนี และทิศทางตามดัชนีที่ปรากฏในเอกสารและคำถาม

ตัวอย่าง 3.2 แสดงวิธีการหาค่าความถี่ผกผันของดัชนีโดยกำหนดให้มีเอกสารทั้งหมด $N = 23$ และ ใช้ตัวอย่างข้อมูลคำถามและเอกสารจากรูปที่ 3.2 ก)
การคำนวณค่าถ่วงน้ำหนักของดัชนี ของ ดัชนี "Computer" ในเอกสาร d2 แสดงได้ดังนี้

$$idf_k = \log\left(\frac{N}{n_k}\right)$$

$$idf_1 = \log\left(\frac{23}{3}\right)$$

$$= 0.4771$$

จากข้อมูลในรูปที่ 3.2ก) สามารถหา ค่าความถี่ผกผัน และค่าความถี่ของดัชนีในแต่ละเอกสาร ได้ดังตารางที่ 3.1

ตารางที่ 3.1 แสดงจำนวนความถี่ของคำที่ปรากฏในแต่ละเอกสารและคำถาม

ลำดับ	ดัชนี	ค่าความถี่ผกผันของดัชนี (idf)	จำนวนความถี่ของคำที่ปรากฏในแต่ละเอกสารและคำถาม (tf)				
			d1	d2	d3	d4	q
1	Computer	0.4771	1	1	0	1	1
2	Information	0.4771	0	2	1	1	1
3	Retrieval	0.4771	0	1	2	2	0

ตัวอย่าง 3.3 แสดงวิธีการหาค่าถ่วงน้ำหนักของดัชนี โดยใช้ข้อมูลค่าความถี่ผกผันและจำนวนความถี่ของคำที่ปรากฏในแต่ละเอกสารจากตารางที่ 3.1 และจากข้อมูลในตารางที่ 3.1 นี้สามารถคำนวณค่าถ่วงน้ำหนักได้ดังตารางที่ 3.2
การคำนวณค่าถ่วงน้ำหนักของดัชนี ของ ดัชนี "Computer" ในเอกสาร d2 แสดงได้ดังนี้

$$w_{12} = tf_{12} * idf_1$$

$$w_{12} = 1 * 0.4771$$

$$= 0.4771$$

ตัวอย่าง 3.4 แสดงขั้นตอนการคำนวณค่าความคล้ายคลึงระหว่างเอกสารและคำถามโดยใช้ค่าถ่วงน้ำหนักดัชนีที่ได้จากตารางที่ 3.2

ตารางที่ 3.2 แสดงค่าความถี่ผกผันของดัชนีตามข้อมูลในรูปที่ 3.2 ก)

ลำดับ	ดัชนี	ค่าถ่วงน้ำหนักของดัชนีในเอกสารและคำถาม				
		d1	d2	d3	d4	q
1	Computer	0.4771	0.4771	0	0.4771	0.4771
2	Information	0	0.9542	0.4771	0.4771	0.4771
3	Retrieval	0	0.4771	0.9542	0.9542	0

จากข้อมูลในตารางที่ 3.2 สามารถแสดงการคำนวณค่าความคล้ายคลึงระหว่างเอกสารและคำถามได้ดังนี้

1) ขั้นแรกคำนวณค่าความยาวเวกเตอร์ของเอกสารและคำถาม จากตารางที่ 3.1

โดยนิยามให้ $|d_i|$ แทน ความยาวเวกเตอร์ของเอกสาร

$|q|$ แทน ความยาวเวกเตอร์ของคำถาม

$$|d_i| = \sqrt{\sum_{k=1}^{N} (w_{i,k})^2}$$

$$|d_1| = \sqrt{0.4771^2} = \sqrt{0.2276} = 0.4770$$
$$|d_2| = \sqrt{0.4771^2 + 0.9542^2 + 0.4771^2} = \sqrt{1.3656} = 1.16858$$
$$|d_3| = \sqrt{0.4771^2 + 0.9542^2} = \sqrt{1.1380} = 1.0667$$
$$|d_4| = \sqrt{0.4771^2 + 0.4771^2 + 0.9542^2} = \sqrt{1.3656} = 1.16858$$

$$|q| = \sqrt{\sum_{k=1}^{N} (w_{q,k})^2}$$

$$|q| = \sqrt{0.4771^2 + 0.4771^2} = \sqrt{0.4552} = 0.6746$$

2) คำนวณ ผลคูณภายในของทุกคำที่เจอในเอกสารและคำถาม

$$q \bullet d_1 = 0.4771 * 0.4771 = 0.2276$$

$$q \bullet d_2 = 0.4771 * 0.4771 + 0.4771 * 0.9542 = 0.6828$$

$$q \bullet d_3 = 0.4771 * 0.4771 = 0.2276$$

$$q \bullet d_4 = 0.4771 * 0.4771 + 0.4771 * 0.4771 = 0.4552$$

3) คำนวณค่าความคล้ายคลึงของเอกสารและคำถามของผู้ใช้โดยใช้สมการที่ 3.3

$$Sim(q, d_1) = \frac{0.2276}{0.4770 * 0.6746} = 0.7090$$

$$Sim(q, d_2) = \frac{0.6828}{1.16858 * 0.6746} = 0.8661$$

$$Sim(q, d_3) = \frac{0.2276}{1.0667 * 0.6746} = 0.3163$$

$$Sim(q, d_4) = \frac{0.4552}{1.16858 * 0.6746} = 0.5574$$

จากค่าความคล้ายคลึงที่ได้สามารถเรียงลำดับเอกสารที่มีความคล้ายคลึงกับคำถามของผู้ใช้ได้ดังนี้ d_2, d_1, d_4 และ d_2 ตามลำดับ

3.1.3 วิธีการป้อนกลับของผู้ใช้

จากการที่ผลลัพธ์การสืบค้นโดยคำถามบางคำถามอาจทำให้สืบค้นได้เอกสารที่ไม่เกี่ยวข้องมาด้วย ซึ่งสาเหตุเกิดจากคำสำคัญที่ใช้ในการสืบค้นเอกสารอาจจะปรากฏอยู่ในเอกสารอื่นด้วย วิธีการลดข้อผิดพลาดนี้สามารถทำได้โดยใช้วิธีการป้อนกลับของผู้ใช้ (Relevance Feedback) [12] ซึ่งวิธีการนี้เป็นเทคนิคที่ทำการปรับปรุงคำถามให้สามารถชี้เฉพาะเอกสารได้เจาะจงมากขึ้น

วิธีการปรับปรุงคำถามของผู้ใช้ มีดังต่อไปนี้

1) เพิ่ม หรือ ลด ดัชนีคำค้น เพื่อเพิ่มความแม่นยำในการสืบค้น

2) เพิ่มโครงสร้างในกระบวนการถาม เพื่อให้สามารถชี้เฉพาะเอกสารได้เจาะจงยิ่งขึ้น เช่น เพิ่มบูลีนให้กับโครงสร้างการถาม เช่น AND และ OR เป็นต้น

3) ทำการปรับค่าถ่วงน้ำหนักของดัชนีให้เหมาะสม โดยการปรับค่าถ่วงน้ำหนักของดัชนีมีหลักการ คือ ใช้ข้อมูลจากคำถามเดิมที่ผู้ใช้ป้อนมาทำการปรับปรุงค่าถ่วงน้ำหนัก โดยถ้าดัชนีมีความสามารถในการบ่งชี้เอกสารสูงก็จะถูกกำหนดค่าถ่วงน้ำหนักให้เป็นบวก และถ้าดัชนีไม่มีความสามารถในการบ่งชี้เอกสารค่าถ่วงน้ำหนักก็จะถูกลด จนกระทั่งเป็น 0

วิธีการปรับปรุงคำถามนี้จะทำการปรับปรุงคำถามเดิมให้สามารถชี้เฉพาะเจาะจงคำตอบได้ดียิ่งขึ้น โดยวิธีนี้สามารถที่จะเลือกใช้ได้ตามความเหมาะสมกับระบบสืบค้นสารสนเทศ วิธีที่หนึ่งคือการเลือกคำที่เหมาะสมกับการบ่งชี้เอกสารเพิ่มเข้าไปในแฟ้ม และทำการคัดเอาดัชนีที่ไม่

สามารถสืบค้นได้ออกจากแฟ้มดัชนี เพื่อเพิ่มความแม่นยำในการสืบค้น วิธีที่สองทำการปรับปรุงวิธีการถามซึ่งวิธีการนี้ไม่ได้ปรับเปลี่ยนดัชนีหรือค่าถ่วงน้ำหนัก แต่จะทำการเพิ่มโครงสร้างให้กับคำถามเพื่อให้สามารถเจาะจงถึงเอกสารได้มากยิ่งขึ้น และวิธีที่สามเป็นวิธีการที่ทำการปรับค่าถ่วงน้ำหนักของคำถามเพื่อให้มีความแม่นยำในการสืบค้นมากขึ้น ซึ่งสมการในการใช้ปรับปรุงค่าถ่วงน้ำหนักของคำถามมีอยู่หลายสมการ [4] โดยสมการที่ 3.4 แสดงสมการของ Rocchio ซึ่งเป็นสมการต้นแบบที่ใช้เพื่อปรับค่าถ่วงน้ำหนักของคำถามในแบบจำลองเวกเตอร์สเปซ ซึ่งสมการนี้มีหลักการคือ จะทำการเพิ่มค่าถ่วงน้ำหนักให้กับดัชนีในคำถามที่สามารถสืบค้นแล้วได้เอกสารที่เหมาะสม และจะทำการลดค่าถ่วงน้ำหนักให้กับดัชนีที่สืบค้นแล้วได้เอกสารที่ไม่เหมาะสม

$$Q' = \alpha Q_0 + \frac{\beta}{|D_u|}\sum_{d_r \in D_u} d_r - \frac{\gamma}{|D_v|}\sum_{d_i \in D_v} d_i \qquad (3.4)$$

$Q_0 = (q_1, q_2, ..., q_i)$ แทน ชุดค่าถ่วงน้ำหนักของคำถามเดิมที่ผู้ใช้ป้อน

$Q' = (q_1', q_2', ...q_i')$ แทน ชุดค่าถ่วงน้ำหนักของคำถามที่ทำการปรับปรุงค่าถ่วงน้ำหนัก

q_i แทน ค่าถ่วงน้ำหนักของคำสำคัญในคำถามเดิมที่ผู้ใช้ป้อน

q_i' แทน ค่าถ่วงน้ำหนักของคำสำคัญในคำถามที่ทำการปรับปรุงค่าถ่วงน้ำหนัก

d_r แทน เอกสารที่สืบค้นแล้วเหมาะสม (Relevance Document)

d_i แทน เอกสารที่สืบค้นแล้วไม่เหมาะสม (Non-Relevance Document)

α , β และ γ แทน พารามิเตอร์แสดงความสำคัญของ คำถามเดิม เอกสารที่เหมาะสม และเอกสารที่ไม่เหมาะสม ตามลำดับ

D_u แทน เอกสารที่สืบค้นแล้วเหมาะสมทั้งหมด

D_v แทน เอกสารที่สืบค้นแล้วไม่เหมาะสมทั้งหมด

สมการนี้ใช้ α , β และ γ เป็นพารามิเตอร์ในการปรับค่าถ่วงน้ำหนักของคำถาม โดย α แทนความสำคัญของคำถามเดิม ซึ่งส่วนมากจะแทนด้วย 1 เพื่อคำถาม Q_0 มีความสำคัญเท่าเดิม และ ค่า β ควรที่จะสูงกว่าค่า γ เนื่องจาก β เป็นพารามิเตอร์ที่ทำการเพิ่มค่าถ่วงน้ำหนักให้กับดัชนีในคำถามที่สามารถสืบค้นแล้วได้เอกสารที่เหมาะสม และ γ เป็นพารามิเตอร์ที่ทำการลดค่าถ่วงน้ำหนักให้กับดัชนีที่ทำการสืบค้นแล้วได้เอกสารที่ไม่เหมาะสม โดยบางระบบจะกำหนดความสัมพันธ์ของ (α, β, γ) เป็น (1,1,0) นั่นหมายความว่าไม่สนใจเอกสารที่สืบค้นแล้วไม่เหมาะสม จะทำการเพิ่มค่าถ่วงน้ำหนักให้กับดัชนีที่สืบค้นแล้วได้เอกสารที่เหมาะสมเท่านั้น หรือบางระบบจะใช้เป็น $(1, \frac{1}{2}, \frac{1}{4})$ นั่นคือ ทำการปรับค่าถ่วงน้ำหนักให้กับดัชนีที่สืบค้นแล้วได้เอกสาร

ที่เหมาะสมโดยการเพิ่มค่าถ่วงน้ำหนักครั้งละ $\frac{1}{2}$ และดัชนีใดได้เอกสารที่ไม่เหมาะสมก็จะลดครั้งละ $\frac{1}{4}$ ทั้งนี้ค่าพารามิเตอร์จะเป็นเท่าไรขึ้นอยู่กับความต้องการปรับค่าถ่วงน้ำหนักของระบบ ตัวอย่างที่ 3.5 แสดงการปรับค่าถ่วงน้ำหนักของดัชนีของคำถามโดยใช้สมการของ Rocchio

ตัวอย่าง 3.5 แสดงตัวอย่างการปรับค่าถ่วงน้ำหนักของดัชนีของให้เหมาะสมโดยใช้สมการของ Rocchio

จากสมการที่ 3.4 กำหนดให้ค่าพารามิเตอร์ $\alpha = 1, \beta = \frac{1}{2}, \gamma = \frac{1}{4}$ และมีค่าถ่วงน้ำหนักของดัชนี "Computer" "Information" "Retrieval" "Design" "Pattern" ในคำถาม Q และเอกสาร d1, d2 เป็นดังตารางที่ 3.3 โดยกำหนดให้เอกสาร d1 เป็นเอกสารที่สืบค้นแล้วเหมาะสมกับคำถาม และเอกสาร d2 เป็นเอกสารที่สืบค้นแล้วไม่เหมาะสม

ตารางที่ 3.3 แสดงตัวอย่างค่าถ่วงน้ำหนักของดัชนีที่ใช้ในการคำนวณค่าการป้อนกลับของผู้ใช้

	Computer	Information	Retrieval	Design	Pattern
Q	4	0	2	0	1
d1	2	1	2	0	0
d2	1	0	0	0	2

จากตารางที่ 3.3 ค่าถ่วงน้ำหนักในแต่ละดัชนีของชุดคำถามเดิมที่ผู้ใช้ป้อน คือ

$$Q_o = (q_1, q_2, ..., q_i)$$

$$Q_0 = (4,0,2,0,1)$$

ค่าถ่วงน้ำหนักในแต่ละดัชนีของเอกสารที่สืบค้นแล้วเหมาะสม คือ

$$\sum_{d_r \in D_n} d_r = (2,1,2,0,0)$$

ค่าถ่วงน้ำหนักในแต่ละดัชนีของเอกสารที่สืบค้นแล้วไม่เหมาะสม คือ

$$\sum_{d_i \in D_v} d_i = (1,0,0,0,2)$$

จากข้อมูลข้างต้นสามารถคำนวณค่าถ่วงน้ำหนักของคำถามใหม่ได้ดังต่อไปนี้

$$Q' = Q_0 + \frac{1}{2}\left(\sum_{d_r \in D_n} d_r\right) - \frac{1}{4}\left(\sum_{d_i \in D_v} d_i\right)$$

$$Q' = (4,0,2,0,1) + \frac{1}{2}(2,1,2,0,0) - \frac{1}{4}(1,0,0,0,2)$$

$$= (4\frac{3}{4}, \frac{1}{2}, 4, 0, \frac{1}{2})$$

จากการคำนวณข้างต้นสามารถสรุปได้ว่าค่าถ่วงน้ำหนักในคำถามใหม่นั้นสามารถถูกปรับได้ดังต่อไปนี้

1. ค่าถ่วงน้ำหนักของ "Computer" ถูกปรับเพิ่มขึ้นจาก 4 เป็น $4\frac{3}{4}$ และ "Retrieval" ถูกปรับเพิ่มขึ้นจาก 2 เป็น 4 นั้นหมายความว่าในคำถาม Q ดัชนี "Computer" และ"Retrieval" มีความสามารถในการบ่งชี้ถึงเอกสารที่เหมาะสมได้ดังนั้นค่าถ่วงน้ำหนักจึงถูกปรับเพิ่มขึ้น

2. ค่าถ่วงน้ำหนักของ "Pattern" ถูกปรับลดลงจาก 1 เป็น $\frac{1}{2}$ นั้นหมายความว่าในคำถาม Q ดัชนี "Pattern" นั้นไม่สามารถบ่งชี้ถึงเอกสารที่เหมาะสมได้ ดังนั้นค่าถ่วงน้ำหนักจึงถูกปรับลดลง

3. และสุดท้ายทำการเพิ่มดัชนี "Information" เข้าไป นั้นหมายความว่าในคำถาม Q นั้น "Information" มีความสามารถในการบ่งชี้ถึงเอกสารที่เหมาะสมได้ ดังนั้นจึงเพิ่มดัชนี"Information" เข้าไปในคำถาม

ซึ่งการปรับค่าถ่วงน้ำหนักในคำถามที่ใช้ในการสืบค้นนี้จะทำให้ได้ผลลัพธ์การสืบค้นที่ดีขึ้น อย่างไรก็ตามค่าถ่วงน้ำหนักที่ปรับได้นี้อาจไม่สามารถสืบค้นเอกสารที่เหมาะสมได้ในการปรับเพียงครั้งเดียว จึงต้องทำการวนซ้ำ (Iteration) การปรับค่าถ่วงน้ำหนัก เพื่อให้ได้ค่าถ่วงน้ำหนักที่เหมาะสมสำหรับคำถามนั้นๆ

3.1.4 วิธีการประเมินระบบสืบค้นสารสนเทศ

วิธีการประเมินประสิทธิผลของระบบสืบค้นสารสนเทศ ทำได้โดยการวัดค่าความแม่นยำ ค่าความระลึก และค่าเฉลี่ยความแม่นยำ [13] ซึ่งมีรายละเอียด ดังหัวข้อที่ 3.1.4.1, 3.1.4.2 และ 3.1.4.3 และตัวอย่างกราฟค่าความแม่นยำเทียบกับค่าความระลึก แสดงได้ดังรูปที่ 3.3

3.1.4.1 ค่าความแม่นยำ หมายถึง ค่าความถูกต้องของผลลัพธ์ที่ระบบการสืบค้นสารสนเทศสามารถสืบค้นมาได้ ซึ่งค่าความแม่นยำนี้หาได้จากอัตราส่วนของการสืบค้นเอกสารที่ถูกต้องจากจำนวนเอกสารทั้งหมดที่ทำการสืบค้นมาได้ ดังสมการที่ 3.5

$$\text{ค่าความแม่นยำ } (P) = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารทั้งหมดที่สืบค้นออกมาได้}} \qquad (3.5)$$

3.1.4.2 ค่าความระลึก หมายถึง ค่าที่แสดง (Rank) ว่าระบบการสืบค้นสารสนเทศสามารถดึงข้อมูลที่เกี่ยวข้องกับคำถามได้ม เกน้อยเพียงไร โดยค่าความระลึกหาได้จากอัตราส่วนของการสืบค้นเอกสารที่ถูกต้องจากจำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล ดังสมการที่ 3.6

$$\text{ค่าความระลึก } (R) = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล}} \tag{3.6}$$

3.1.4.3 ค่าเฉลี่ยความแม่นยำ หมายถึง ค่าความแม่นยำในการสืบค้นของระบบสืบค้นสารสนเทศที่ได้จากการเฉลี่ยค่าความแม่นยำกับค่าความระลึก 11 จุด ซึ่งสามารถหาได้จากสมการที่ 3.7 [12]

$$\overline{P}(r) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q} \tag{3.7}$$

เมื่อ $\overline{P}(r)$ แทน ค่าเฉลี่ยความแม่นยำ ณ ค่าระลึกลำดับที่ r

$P_i(r)$ แทน ค่าความแม่นยำ ณ ค่าระลึกลำดับที่ r สำหรับคำถามที่ i

N_q แทน จำนวนคำถามทั้งหมดที่ใช้ในการหาค่าเฉลี่ยความแม่นยำ

3.1.4.4 ความสัมพันธ์ระหว่าง ค่าความแม่นยำ และค่าความระลึก

พิจารณาความสัมพันธ์ระหว่างค่าความแม่นยำ และค่าความระลึกนี้จะพบว่าค่าทั้งสองมีความสัมพันธ์ในลักษณะแปรผกผันกันดังแสดงในรูปที่ 3.3 นั่นคือ ถ้าระบบมีค่าความแม่นยำสูงแล้ว ระบบจะให้ค่าความระลึกต่ำ ในทางกลับกันถ้าระบบมีค่าความแม่นยำต่ำแล้ว ระบบจะให้ค่าความระลึกสูง ตัวอย่างความสัมพันธ์ระหว่างค่าความแม่นยำและค่าความระลึกแสดงได้ดังตัวอย่างที่ 3.6 ดังนี้

ตัวอย่างที่ 3.6 แสดงตัวอย่างความสัมพันธ์ระหว่างค่าความแม่นยำและค่าความระลึก

เมื่อระบบมีการป้อนคำถาม q ให้กับระบบแล้วระบบสืบค้นคำตอบออกมาโดยมีการเรียงลำดับ ดังต่อไปนี้

1. d_4 ✓	4. d_{123}	7. d_1	10. d_{135} ✓	13. d_{20}
2. d_9	5. d_5	8. d_{47}	11. d_{111}	14. d_{59}
3. d_{256} ✓	6. d_{13} ✓	9. d_{68}	12. d_{43}	15. d_7 ✓

สัญลักษณ์ ✓ หลังเอกสารแสดงถึงเอกสารที่สืบค้นมาแล้วเหมาะสมกับคำถาม โดยสมมติให้ในฐานข้อมูลมีเอกสารที่เกี่ยวข้องกับคำถาม 10 เอกสาร คือ d_4 , d_7 , d_{11} , d_{13} , d_{16} , d_{73} , d_{99} , d_{135} , d_{178} , d_{256}

จากข้อมูลข้างต้นสามารถคำนวณค่าความแม่นยำ และค่าความระลึกได้ ดังนี้

1. เอกสาร d_4 เป็นเอกสารที่อยู่ในกลุ่มของเอกสารที่เกี่ยวข้องกับคำถาม และสามารถสืบค้นมาได้ในลำดับที่ 1 ดังนั้น ณ จุดนี้ระบบจะให้

ค่าความระลึก = 1/10 = 0.1 และค่าความแม่นยำ = 1/1 = 1

2. เอกสาร d_{256} เป็นเอกสารที่อยู่ในกลุ่มของเอกสารที่เกี่ยวข้องกับคำถาม และสามารถสืบค้นมาได้ในลำดับที่ 2 ดังนั้น ณ จุดนี้ระบบจะให้

ค่าความระลึก = 2/10 = 0.2 และค่าความแม่นยำ = 2/3 = 0.66

3. เอกสาร d_{13} เป็นเอกสารที่อยู่ในกลุ่มของเอกสารที่เกี่ยวข้องกับคำถาม และสามารถสืบค้นมาได้ในลำดับที่ 3 ดังนั้น ณ จุดนี้ระบบจะให้

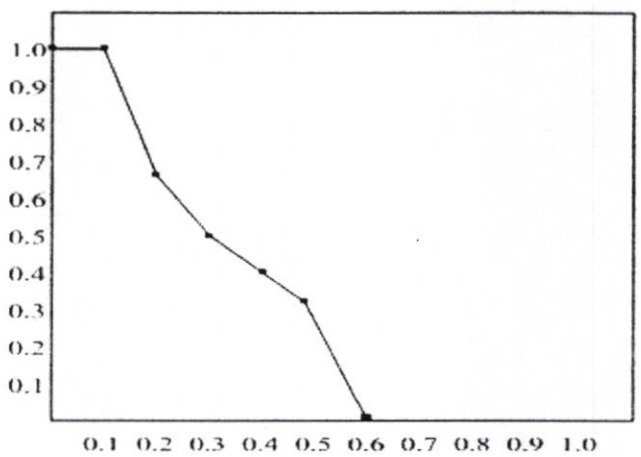
ค่าความระลึก = 3/10 = 0.3 และค่าความแม่นยำ = 3/6 = 0.5

4. เอกสาร d_{135} เป็นเอกสารที่อยู่ในกลุ่มของเอกสารที่เกี่ยวข้องกับคำถาม และสามารถสืบค้นมาได้ในลำดับที่ 4 ดังนั้น ณ จุดนี้ระบบจะให้

ค่าความระลึก = 4/10 = 0.4 และค่าความแม่นยำ = 4/10 = 0.4

5. เอกสาร d_7 เป็นเอกสารที่อยู่ในกลุ่มของเอกสารที่เกี่ยวข้องกับคำถาม และสามารถสืบค้นมาได้ในลำดับที่ 5 ดังนั้น ณ จุดนี้ระบบจะให้

ค่าความระลึก = 5/10 = 0.5 และค่าความแม่นยำ = 5/15 = 0.33



รูปที่ 3.3 แสดงกราฟความสัมพันธ์ระหว่างค่าความแม่นยำ และค่าความระลึก

รูปที่ 3.3 แสดงความสัมพันธ์ระหว่างค่าความแม่นยำ และค่าความระลึก โดยกราฟนี้เกิดจากความสัมพันธ์ของที่ค่าความแม่นยำ และค่าความระลึกที่ได้จากตัวอย่างที่ 3.6 จะเห็นได้ว่า ค่าความแม่นยำในระดับค่าความระลึกมากกว่า 0.5 นั้นจะมีค่าเป็น 0 เนื่องจากไม่มีเอกสารใดที่เกี่ยวข้องกับคำถามได้ถูกสืบค้นขึ้นมาอีกแล้ว โดยค่าความระลึกที่ใช้เทียบกับค่าความแม่นยำนี้จะมี 11 ค่า ซึ่งเรียกว่าค่าความระลึกมาตรฐาน (Standard Recall) ซึ่งประกอบด้วยค่า 0, 0.1, 0.2, 0.3, 0.4,..., 1.0 โดยในตัวอย่างที่ 3.6 นี้จะแสดงการหาค่าความแม่นยำ และค่าความระลึกของคำถาม 1 คำถาม แต่โดยทั่วไปแล้วการประเมินประสิทธิผลของระบบสืบค้นสารสนเทศนั้นจะวัดจากคำถามหลายๆคำถามที่แตกต่างกัน ทำให้ได้กราฟแสดงค่าความแม่นยำ และค่าความระลึกที่แตกต่างกันหลายๆกราฟ ซึ่งการทดสอบโดยใช้คำถามหลายๆคำถามนี้จะถูกวัดประสิทธิผลต่อไปโดยใช้ค่าเฉลี่ยความแม่นยำ ซึ่งกล่าวไว้ในหัวข้อที่ 3.1.4.3

3.2 งานวิจัยที่เกี่ยวข้องกับการคำนวณค่าถ่วงน้ำหนักและการจัดโครงสร้างเอกสารเพื่อการสืบค้นในระบบสืบค้นสารสนเทศ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับวิธีการคำนวณค่าถ่วงน้ำหนักของดัชนีในระบบการสืบค้นสารสนเทศ สามารถสรุปการคำนวณค่าถ่วงน้ำหนักของดัชนีได้เป็น 2 กรณี คือ กรณีที่ให้ค่าถ่วงน้ำหนักโดยไม่สนใจโครงสร้างของเอกสาร และกรณีที่ให้ค่าถ่วงน้ำหนักตามโครงสร้างของเอกสาร โดยในกรณีที่ไม่สนใจโครงสร้างของเอกสารนั้นจะใช้จำนวนความถี่ในการปรากฏของคำในการส่งผลต่อความสามารถในการบ่งชี้ ซึ่งรายละเอียดอธิบายไว้ในหัวข้อที่ 3.2.1 และในอีกกรณีหนึ่ง คือ การให้ค่าถ่วงน้ำหนักตามโครงสร้างการอธิบายของเอกสารในกรณีนี้จะไม่คำนึงถึงความถี่ของคำในเอกสารแต่จะให้ค่าถ่วงน้ำหนักตามการปรากฏของคำในโครงสร้างของเอกสาร เช่น คำที่เป็นชื่อจะมีค่าถ่วงน้ำหนักมากกว่าคำที่เป็นส่วนอธิบายทั่วๆไป เนื่องจากชื่อมีความสามารถในการบ่งชี้มากกว่า เป็นต้น รายละเอียดอธิบายไว้ในหัวข้อที่ 3.2.2 ดังต่อไปนี้

3.2.1 ความถี่ของคำ วิธีการนี้จะใช้ค่าความถี่ของคำเป็นหลัก โดยนับความถี่ของคำที่ปรากฏในเอกสาร ซึ่งโดยทั่วไปคำที่ปรากฏอยู่ในเอกสารจะถูกพิจารณาว่าเป็นคำสำคัญถ้าคำนั้นปรากฏหลายครั้งในเอกสาร แต่ปรากฏน้อยครั้งในจำนวนเอกสารทั้งหมด หลักการนี้ถูกนำมาใช้ในสูตรการหาค่า TF/IDF ซึ่งรายละเอียดได้กล่าวไว้ในหัวข้อที่ 3.1.1.3 ตัวอย่างงานวิจัยระบบสืบค้นสารสนเทศที่ใช้ค่าน้ำหนักของดัชนีโดยใช้ค่าความถี่ของดัชนีเป็นหลัก มีดังนี้

Wei Zhao และคณะ [22] เป็นงานวิจัยหนึ่งซึ่งให้ค่าถ่วงน้ำหนักของดัชนีโดยยึดตามค่าความถี่ของดัชนี โดยงานวิจัยนี้นำเสนอวิธีการหากลุ่มของฟังก์ชันที่ สอดคล้องในแต่ละความต้องการ (Requirement) โดยแนวคิดหลักของงานวิจัยนี้ คือ การรวมเอาเทคนิคการสืบค้นระบบสารสนเทศ กับการวิเคราะห์โครงสร้างคงที่ของรหัสต้นฉบับไว้ด้วยกัน โดยใช้ความต้องการที่เป็น

ฟังก์ชัน (Functional Requirements) เป็นคำถามในการสืบค้น และเป้าหมายในการสืบค้น คือ ข้อมูลที่ระบุ (Identifier Information) เช่น ชื่อของฟังก์ชัน, ชื่อตัวแปร, ชื่อพารามิเตอร์ เป็นต้นโดยส่วนการสืบค้นของงานวิจัยนี้จะสืบค้นโดยใช้แบบจำลองเวกเตอร์สเปซ และใช้ค่าความถี่ของดัชนีเป็นตัวหลักในการคำนวณค่าถ่วงน้ำหนัก

จากงานวิจัยข้างต้นเป็นต้นอย่างระบบการสืบค้นสารสนเทศหนึ่งที่ใช้ความถี่ของคำเป็นหลักในการคำนวณค่าถ่วงน้ำหนักของดัชนี ซึ่งวิธีการให้ค่าถ่วงน้ำหนักดัชนีอีกวิธีหนึ่ง คือการให้ค่าถ่วงน้ำหนักตามความสำคัญของโครงสร้างของเอกสาร ซึ่งจะทำให้สามารถบ่งชี้ได้ว่าส่วนใดของเอกสารสำคัญมากน้อยกว่ากัน ซึ่งรายละเอียดดังหัวข้อที่ 3.2.2 ดังนี้

3.2.2 โครงสร้างของเอกสาร วิธีการนี้จะให้ค่าน้ำหนักของดัชนีตามโครงสร้างหลักของเอกสารโดยโครงสร้างนี้จะขึ้นอยู่กับการตีความเอกสารจากฐานความรู้ที่มีอยู่โดยจะทำการพิจารณาเอกสาร และทำการแปลงเนื้อหาของเอกสารให้มีลักษณะเป็นโครงสร้างที่สามารถแทนความรู้ของเอกสารได้ [14][15][17][19] ซึ่งการกระทำเช่นนี้จะทำให้สามารถสรุปใจความสำคัญของเอกสารได้ ในส่วนของวิธีแปลงเนื้อหาเอกสารให้เป็นโครงสร้างนี้อาจใช้คอนเซปต์ชวลกราฟ (Conceptual Graph) [17] มาช่วยสร้างโครงสร้างของเนื้อหา ซึ่งแต่ละส่วนของโครงสร้างอาจได้มาจาก หัวเรื่อง (Topic) ชื่อเรื่อง (Title) หัวข้อ (Heading) หรือคำที่เป็นตัวแทนการอธิบายเอกสาร ปกติแล้ววิธีการให้ค่าน้ำหนักของดัชนีที่จัดเรียงแบบโครงสร้างนี้จะมีลักษณะเป็นลำดับ ซึ่งการให้ค่าน้ำหนักเป็นลำดับชั้นนี้จะให้ค่าน้ำหนักในแต่ละส่วนของโครงสร้างเอกสารไม่เท่ากัน โดยทั่วไปแล้วจะให้ค่าถ่วงน้ำหนักสูงในส่วนของโครงสร้างเอกสารที่มีความสามารถในการบ่งชี้สูง และจะให้ค่าถ่วงน้ำหนักลดหลั่นกันตามความสามารถในการบ่งชี้ งานวิจัยที่เกี่ยวข้องกับแนวความคิดการวิเคราะห์โครงสร้างของเอกสาร มีดังนี้

Noriko Kando [15] นำเสนอบทความที่ทำการรวบรวมงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์เอกสารเป็นโครงสร้างลำดับชั้นก่อนทำการสืบค้น นอกจากนี้ยังทำการวิเคราะห์งานวิจัยและทำการสรุปว่าการจัดเอกสารให้เป็นโครงสร้างก่อนการสืบค้นนั้นจะทำให้ผลการสืบค้นที่แม่นยำมากขึ้น

Philippe และ Laurence [17] ชี้ให้เห็นถึงประโยชน์ในการจัดเอกสารให้มีลักษณะเป็นโครงสร้างสำหรับการเก็บ และการแทนข้อมูล เพื่อการสืบค้น นอกจากนี้ยังแสดงการแทนเอกสารด้วยฐานความรู้ ของเอกสารโดยใช้คอนเซปต์ชวล

Song mao และคณะ [19] นำเสนอบทความการวิเคราะห์ขั้นตอนวิธี ในการสร้างโครงสร้างเอกสารโดยได้ทำการรวบรวมงานวิจัยที่ทำการนำเสนอขั้นตอนวิธีในการแทนเอกสารเป็นโครงสร้างทั้งทางกายภาพ และทางตรรกวิทยา อีกทั้งทำการประเมินขั้นตอนวิธีในการวิเคราะห์โครงสร้างเอกสาร

Magnus Rosell [14] นำเสนอบทความที่ทำการสรุปงานวิจัยที่ออกแบบขั้นตอนวิธีในการสรุปเอกสาร โดยใช้ฐานความรู้ ซึ่งเป็นวิธีที่ทำการตีความจากเอกสารและจัดโครงสร้างให้เอกสารในลักษณะต่างๆ เช่น กราฟ

งานวิจัยที่เกี่ยวข้องกับระบบสืบค้นสารสนเทศ ที่ให้ค่าน้ำหนักของดัชนีตามโครงสร้างเอกสาร มีดังต่อไปนี้

Yohei Seki [24] เสนองานวิจัยที่ทำการสรุปข่าวในหนังสือพิมพ์ ซึ่งงานวิจัยนี้ได้ใช้กลวิธีในการคำนวณค่าถ่วงน้ำหนักของประโยคเพื่อใช้สืบค้นบทสรุปของข่าวในหนังสือพิมพ์ โดยการจัดลำดับความสำคัญของโครงสร้างหนังสือพิมพ์ในแต่ละส่วนให้ไม่เท่ากัน โดยการกำหนดให้ค่าถ่วงน้ำหนักในส่วนที่สนใจมากกว่าส่วนอื่น เช่น ให้ค่าถ่วงน้ำหนักในประโยคที่มีวลีหัวข้อข่าวปรากฏอยู่มากกว่าประโยคทั่วไป หรือให้ค่าถ่วงน้ำหนักมากขึ้นในประโยคสุดท้ายของประโยค เนื่องจากเป็นประโยคการสรุปของข่าว การออกแบบการคำนวณค่าถ่วงน้ำหนักโดยแบ่งให้แต่ละส่วนของโครงสร้างข่าวไม่เท่ากันนี้จะทำให้สามารถแยกประโยคซึ่งบ่งชี้ถึงใจความสำคัญของข่าวได้

Tao Qin และคณะ [20] ได้เสนอระบบสืบค้นยูสเคส (Uses Case) จากรหัสต้นฉบับโดยใช้ขั้นตอนวิธี บีอาร์ซีจี (BRCG : Branch Reserving Call Graph) ในการแทนโครงสร้างของรหัสต้นฉบับในรูปแบบของกราฟ ซึ่งออกแบบการแทนโครงสร้างของรหัสต้นฉบับเป็น 2 ส่วน คือ ประโยคแยกทาง (branch statement) และประโยคการเรียกฟังก์ชัน (function call statement) โดยส่วนของประโยคแยกทาง นั้นจะทำการแทนคำสั่งในการเขียนโปรแกรมต่างๆ เช่น คำสั่ง if และ While เป็นต้น ประโยคการเรียกฟังก์ชันเป็นส่วนที่แทนฟังก์ชันที่ใช้ในโปรแกรม ซึ่งงานวิจัยดังกล่าวออกแบบค่าถ่วงน้ำหนักในแต่ละส่วนที่กล่าวมานั้นไม่เท่ากัน โดยให้ค่าถ่วงน้ำหนักในส่วนของประโยคแยกทาง มีค่าสูงกว่าส่วนประโยคการเรียกฟังก์ชัน เนื่องจากมีโอกาสในการทำงานมากกว่า การออกแบบให้ค่าถ่วงน้ำหนักในแต่ละส่วนไม่เท่ากันนี้สามารถบ่งบอกความสำคัญของคำในแต่ละส่วนของโครงสร้างรหัสต้นฉบับได้

บทที่ 4
แบบจำลองการสืบค้นแบบรูปการออกแบบ

ในบทที่ 4 นี้จะกล่าวถึงแบบจำลองการสืบค้นแบบรูปการออกแบบ โดยจะเน้นในส่วนของขั้นตอนการเตรียมฐานข้อมูลของแบบรูปการออกแบบจีโอเอฟ ซึ่งขั้นตอนนี้จะประกอบไปด้วยการวิเคราะห์แบบรูปการออกแบบเพื่อสร้างเป็นดัชนีคำค้น และ วิธีการคำนวณค่าถ่วงน้ำหนักของดัชนี โดยมีรายละเอียด ดังนี้

4.1 การวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น

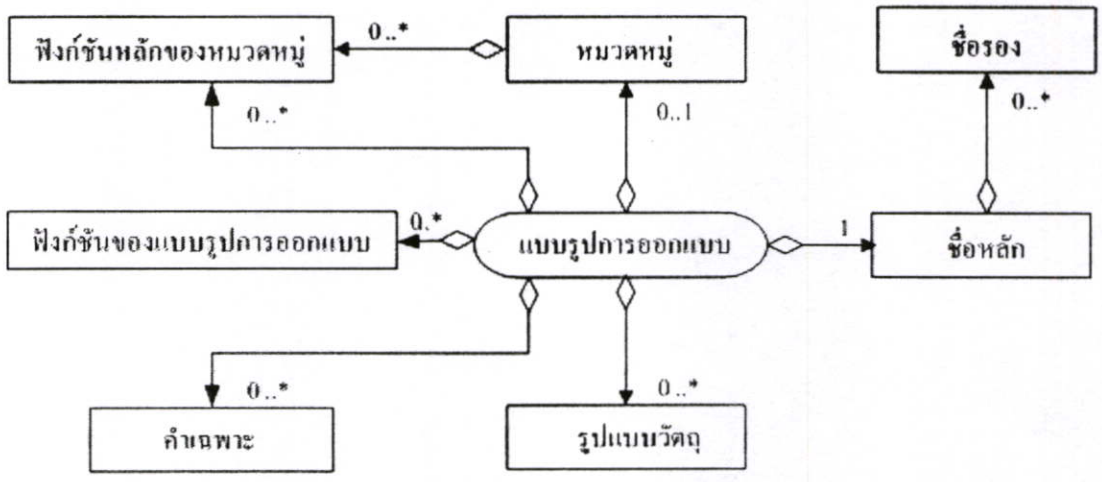
การวิเคราะห์โครงสร้างของเอกสารเพื่อนำมาสร้างเป็นดัชนีคำค้น เป็นขั้นตอนแรกในกระบวนการสืบค้นสารสนเทศ โดยขั้นตอนนี้ทำการวิเคราะห์ข้อมูลในเอกสารเพื่อหาตัวแทนในการบ่งชี้เอกสาร งานวิจัยนี้ทำการวิเคราะห์ตัวแทนที่สามารถบ่งชี้ถึงวิธีการแก้ปัญหาของแบบรูปการออกแบบ เพื่อให้สามารถสืบค้นได้ตามลักษณะปัญหาการออกแบบที่ผู้ออกแบบซอฟต์แวร์ประสบ โดยแบ่งตัวแทนในการบ่งชี้แบบรูปการออกแบบที่นำมาสร้างเป็นดัชนีคำค้นออกเป็น 3 ส่วน ดังต่อไปนี้

ส่วนที่หนึ่ง คือ ชื่อของแบบรูปการออกแบบ ส่วนนี้จะสามารถบ่งชี้ได้ถึงแบบรูปการออกแบบได้โดยตรง โดยจะประกอบไปด้วยชื่อหลัก และชื่อรอง

ส่วนที่สอง คือ หมวดหมู่ของแบบรูปการออกแบบ ส่วนนี้จะสามารถบ่งชี้ถึงการทำงานโดยรวมของแบบรูปการออกแบบ โดยประกอบไปด้วยคำที่สามารถบ่งชี้ถึงหมวดหมู่โดยตรง และคำที่บ่งชี้ถึงการทำงานของหมวดหมู่

ส่วนที่สาม คือ ส่วนที่ได้จากการสรุปการทำงานของแบบรูปการออกแบบ โดยงานวิจัยการพัฒนาโครงสร้างคำอธิบายแบบรูปการออกแบบ [2] ได้กล่าวว่าในแบบรูปการออกแบบของจีโอเอฟ ส่วนคำอธิบายจุดมุ่งหมาย เป็นส่วนที่เหมาะสมที่สุดในการบ่งชี้ถึงการทำงานและวิธีการแก้ปัญหา ของแบบรูปการออกแบบ ดังนั้นงานวิจัยนี้จึงได้สรุปจุดมุ่งหมายของจีโอเอฟ โดยแบ่งเป็น ฟังก์ชันก์ของแบบรูปการออกแบบ รูปแบบวัตถุ และคำเฉพาะ

โดยทั้ง 3 ส่วนที่กล่าวมาข้างต้นนี้สามารถอธิบายเป็นโครงสร้างข้อมูลได้ดังรูปที่ 4.1



รูปที่ 4.1 แสดงโครงสร้างการอธิบายแบบรูปการออกแบบ

โครงสร้างข้อมูลของแบบรูปการออกแบบดังกล่าวสามารถนำมาใช้เป็นดัชนีคำค้น โดยมีรายละเอียดในแต่ละส่วน ดังต่อไปนี้

1) ชื่อหลัก คือ ส่วนที่เป็นชื่อของแบบรูปการออกแบบ เช่น คำดัชนี "Abstract Factory" เป็นชื่อของแบบรูปการออกแบบ แอบสแทรกต์แฟกทอรี จึงเป็นคำที่สามารถบ่งชี้ถึงแอบสแทรกต์แฟกทอรีได้โดยตรง

2) ชื่อรอง คือ ชื่ออีกชื่อหนึ่งของแบบรูปการออกแบบซึ่งเป็นที่รู้จักกันดี เช่น Kit เป็นชื่อรองของแบบรูปการออกแบบ แอบสแทรกต์แฟกทอรี โดยชื่อเหล่านี้จะสื่อถึงลักษณะการทำงานของแบบรูปการออกแบบตัวนั้นด้วย ตัวอย่างชื่อรองของแบบรูปการออกแบบแสดงได้ดังตารางที่ 4.1

ตาราง 4.1 แสดงตัวอย่างชื่อรองของแบบรูปการออกแบบ

แบบรูปการออกแบบ	ชื่อรองของแบบรูปการออกแบบ
Abstract Factory	Kit
Adapter	Wrapper
Observer	Dependent, Publish-Subscribe

3) หมวดหมู่ คือ ส่วนที่เป็นหมวดหมู่ของแบบรูปการออกแบบ โดยหมวดหมู่จะทำการแบ่งแบบรูปการออกแบบตามลักษณะการทำงาน เช่น หมวดหมู่การสร้าง หมวดหมู่โครงสร้าง และหมวดหมู่พฤติกรรม เป็นต้น

4) ฟังก์ชันหลักของหมวดหมู่ คือ ส่วนที่อธิบายถึงฟังก์ชันหลักในการทำงานของหมวดหมู่แบบรูปการออกแบบ ซึ่งคำเหล่านี้อาจหาได้จากคำอธิบายโดยรวมที่แบบรูปการออกแบบในหมวดหมู่เดียวกันกระทำเหมือนกัน เช่น คำดัชนี “Create” และ“Instantiate” เป็นคำที่อธิบายถึงฟังก์ชันหลักของหมวดหมู่การสร้าง เป็นต้น ตัวอย่างฟังก์ชันหลักของหมวดหมู่แบบรูปการออกแบบแสดงได้ดังตารางที่ 4.2

ตาราง 4.2 แสดงตัวอย่างฟังก์ชันของหมวดหมู่แบบรูปการออกแบบ

หมวดหมู่แบบรูปการออกแบบ	ส่วนดำเนินงานหลักของหมวดหมู่
Creational	Create, Instantiate
Structural	Compose, Structure
Behavioral	Communication, Behavior

5) ฟังก์ชันของแบบรูปการออกแบบ คือ ส่วนที่เป็นฟังก์ชันของแบบรูปการออกแบบเป็นส่วนที่อธิบายการกระทำของแบบรูปการออกแบบ ซึ่งคำเหล่านี้อาจหาได้จากส่วนของการกระทำในงานวิจัย [2] เช่น คำดัชนี “Convert” และ“Adapt” เป็นฟังก์ชันของแบบรูปการออกแบบอะแดปเตอร์ หรือ “Update” “Notify” เป็นการดำเนินงานของแบบรูปออบเซิร์ฟเวอร์ เป็นต้น ตัวอย่างฟังก์ชันของแบบรูปการออกแบบแสดงได้ตารางที่ 4.3

ตาราง 4.3 แสดงตัวอย่างฟังก์ชันของแบบรูปการออกแบบ

แบบรูปการออกแบบ	ฟังก์ชันของแบบรูปการออกแบบ
Abstract Factory	Create, without specify concrete class
Adapter	Convert, Adapt
Observer	Update, Notify

6) รูปแบบวัตถุ คือ ส่วนที่เป็นรูปแบบของวัตถุที่ใช้ในการทำงานของแบบรูปการออกแบบ ซึ่งคำในส่วนนี้อาจหาได้จากส่วนที่เป็นเอนทิตีในงานวิจัย [2] เช่น แบบรูปการออกแบบแอบสแทรกต์แฟกทอรี เป็นแบบรูปที่ทำการสร้างออบเจ็กต์ และวัตถุที่แบบรูปการออกแบบสร้างคือ ออบเจ็กต์ที่สัมพันธ์กัน ดังนั้นดัชนีรูปแบบวัตถุของแอบสแทรกต์แฟกทอรี คือ ดัชนีวลี “Relate object” “Family object” และ“Dependent object” เป็นต้น ตัวอย่างรูปแบบวัตถุของแบบรูปการออกแบบแสดงได้ดังตารางที่ 4.4

ตาราง 4.4 แสดงตัวอย่างรูปแบบวัตถุของแบบรูปการออกแบบ

แบบรูปการออกแบบ	รูปแบบวัตถุ
Abstract Factory	Relate object, Dependent object , Family object
Adapter	Interface of class
Observer	One to many dependent object

7) คำเฉพาะ คือ คำที่สามารถบ่งชี้ได้ถึงลักษณะเฉพาะของการอธิบายแบบรูปการออกแบบ เช่น ลักษณะเฉพาะของการทำงาน คือ คำที่แสดงการทำงานที่โดดเด่นของแบบรูปการออกแบบนั้น

ข้อกำหนดความสัมพันธ์ในแต่ละส่วนของโครงสร้างข้อมูลแบบรูปการออกแบบนี้ เป็นส่วนสำคัญส่วนหนึ่งในการวิเคราะห์ลำดับความสามารถในการบ่งชี้ของดัชนี (รายละเอียดในหัวข้อที่ 4.3) โดยมีรายละเอียดความสัมพันธ์ ดังต่อไปนี้

1) ชื่อของแบบรูปการออกแบบ สามารถมีได้เพียงชื่อเดียวเท่านั้น

2) ชื่อรองของแบบรูปการออกแบบจะสามารถมีได้หลายชื่อ เนื่องจากโดยทั่วไปแบบรูปการออกแบบจะมีชื่อเรียกอื่นที่เป็นที่รู้จักกันดีนอกเหนือจากชื่อหลักได้หลายชื่อ

3) หมวดหมู่ของแบบรูปการออกแบบสามารถมีหรือไม่มีก็ได้ แต่ถ้ามีการกำหนดหมวดหมู่แบบรูปการออกแบบจะสามารถมีได้มากสุดเพียงหมวดหมู่เดียวเท่านั้น เนื่องจากการกำหนดหมวดหมู่นั้นจะเป็นการแบ่งแยกถึงวิธีการทำงานของแบบรูปการออกแบบซึ่งแต่ละหมวดหมู่จะมีการทำงานที่ต่างกัน

4) ในส่วนของฟังก์ชันหลักของหมวดหมู่ ฟังก์ชันของแบบรูปการออกแบบ รูปแบบวัตถุและคำเฉพาะที่ใช้ในการเป็นตัวแทนของแบบรูปการออกแบบ สามารถมีคำที่ใช้อธิบายในแต่ละส่วนดังที่กล่าวมาได้หลายคำ เช่น ส่วนฟังก์ชันของออบเซิร์ฟเวอร์ อาจถูกแทนด้วยคำดัชนี “Update” และ“Notify” หรือรูปแบบวัตถุของแอบสแทรกต์แฟกทอรี ถูกแทนด้วยคำดัชนีวลี “Relate object” “family object” และ“Dependent object”

4.2 แบบจำลองการสร้างดัชนีการสืบค้น

ดัชนี เป็นส่วนที่ใช้ในการสืบค้นเอกสารในกระบวนการสืบค้นสารสนเทศ โดยรูปแบบของดัชนีเป็นตัวแปรสำคัญตัวหนึ่งในการบ่งบอกประสิทธิภาพของระบบสืบค้นสารสนเทศ เนื่องจากการใช้ดัชนีในแต่ละรูปแบบนั้นจะให้ผลลัพธ์ของการสืบค้นต่างกัน โดยส่วนใหญ่แล้วดัชนีวลีจะให้ผลลัพธ์ในการสืบค้นที่ดีกว่า ในส่วนของดัชนีการสืบค้นนี้ จะนำคำมาจากขั้นตอนการวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น (รายละเอียดในหัวข้อที่ 4.1) งานวิจัยนี้ใช้รูปแบบดัชนีการ

สืบค้น 2 ประเภท คือ ดัชนีเดี่ยวและดัชนีวลี ซึ่งรายละเอียดของแต่ละประเภทสามารถอธิบายได้ดังต่อไปนี้

4.2.1 ดัชนีเดี่ยว

ดัชนีเดี่ยว หรือ คำ (Word) คือ ดัชนีที่เป็นคำๆเดียว เช่น “Create”, “Convert” และ“Adapt” โดยดัชนีแบบเดี่ยวนี้จะทำการเก็บร่วมกับคำอธิบายแบบรูปการออกแบบเพื่อใช้ในการคำนวณค่าถ่วงน้ำหนักของดัชนีต่อไป (รายละเอียดในหัวข้อที่ 4.3) เช่น “Create” คือฟังก์ชันหลักของหมวดหมู่ หรือ “Convert” คือ ฟังก์ชันของแบบรูปการออกแบบ

การเก็บดัชนีเดี่ยวนี้มีเงื่อนไข คือ ไม่ทำการเก็บดัชนีที่เป็นคำอธิบายเชิงวัตถุ เช่น “Object” “Instance”, “Subclass” และ “Class” เนื่องจากคำอธิบายแบบรูปการออกแบบของจีโอเอฟนั้นเป็นเรื่องของการออกแบบเชิงวัตถุ ดังนั้นคำอธิบายเชิงวัตถุเหล่านี้จะพบได้ในทุกเอกสารของแบบรูปการออกแบบ จึงเป็นคำที่ไม่สามารถบ่งชี้ถึงเอกสารได้ อย่างไรก็ตามเมื่อนำคำอธิบายเชิงวัตถุนี้ประกอบเข้ากับดัชนีเดี่ยวตัวอื่นเพื่อใช้เป็นดัชนีวลี (รายละเอียดในหัวข้อที่ 4.2.2) คำเหล่านี้ก็จะมีคุณสมบัติในการบ่งชี้ถึงแบบรูปการออกแบบได้ดีมากยิ่งขึ้น ดังนั้น ในการเก็บดัชนีวลีจึงเก็บคำที่เป็นคำอธิบายเชิงวัตถุเข้าไปด้วยเพื่อเพิ่มประสิทธิภาพในการบ่งชี้แบบรูปการออกแบบ ตัวอย่างดัชนีเดี่ยวแสดงได้ดังตารางที่ 4.5

ตาราง 4.5 แสดงตัวอย่างการเก็บดัชนีเดี่ยว

ชื่อแบบรูปการออกแบบ	ดัชนีเดี่ยว	คำอธิบายแบบรูปการออกแบบ
Abstract Factory	Abstract	ชื่อหลัก
Abstract Factory	Create	ฟังก์ชันหลักของหมวดหมู่
Abstract Factory	Relate	รูปแบบวัตถุ
Factory Method	Factory	ชื่อหลัก
Composite	Compose	ฟังก์ชันหลักของหมวดหมู่
Composite	Structure	ฟังก์ชันหลักของหมวดหมู่
Composite	Tree	คำเฉพาะ
Singleton	Single	คำเฉพาะ
…	…	….

4.2.2 ดัชนีวลี

ดัชนีวลี คือ ดัชนีที่เกิดจากการประกอบของคำที่อยู่ติดกันอย่างน้อยสองคำขึ้นไป โดยดัชนีวลีเกิดจากการนำดัชนีแบบเดี่ยวมาประกอบกัน อาทิ “Create family” และ“tree structure” ซึ่งในดัชนีวลีจะสามารถใช้คำอธิบายเชิงวัตถุประกอบเข้ากับดัชนีเดี่ยวเพื่อสร้างเป็นดัชนีวลีได้ เช่น “notify object”, “family object”, “create object”, “create family object”, “structure object” และ “tree object” ซึ่งดัชนีวลีนี้อาจหาได้จากการใช้หลักการจับคู่ภาคแสดงและภาคประธานในงานวิจัยของ Daniel และ Gyorgy [8]

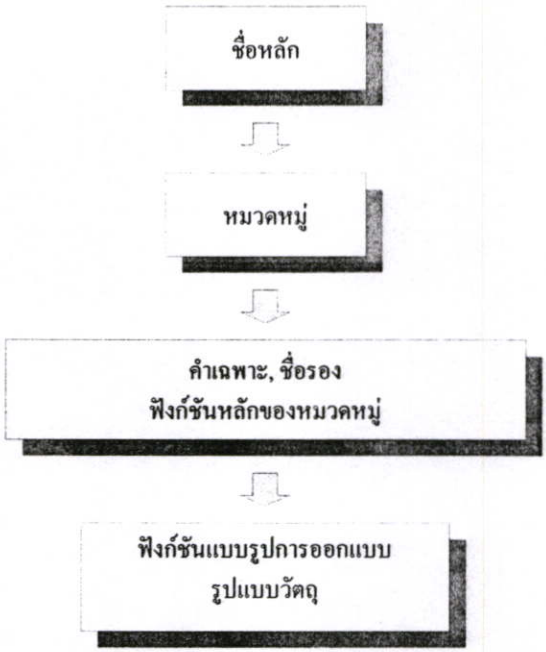
การใช้ดัชนีวลีนี้จะทำให้มีความแม่นยำการสืบค้นมากขึ้น โดยดัชนีวลีอาจมีมากกว่า 2 คำก็ได้ อย่างไรก็ตามจากงานวิจัยของ Wesley และ Chu [23] ได้ทำการทดลองพบว่าการใช้ดัชนีวลีเพียง 2 ถึง 3 คำ ก็สามารถสืบค้นเอกสารได้อย่างแม่นยำแล้ว แม้ว่าดัชนีวลีมีจำนวนคำมากขึ้นกว่านี้ก็ทำให้ความแม่นยำในการสืบค้นเพิ่มขึ้นไม่มาก ดังนั้นงานวิจัยนี้จึงใช้ดัชนีวลีไม่เกิน 3 คำ ตาราง 4.6 แสดงตัวอย่างการเก็บดัชนีวลี

ตาราง 4.6 แสดงตัวอย่างการเก็บดัชนีวลี

ชื่อแบบรูปการออกแบบ	ดัชนีวลี
Abstract Factory	Abstract Factory
Abstract Factory	Create Family Object
Abstract Factory	Family Object
Abstract Factory	Dependent Object
Abstract Factory	Relate Object
Factory Method	Factory Method
Template Method	Template Method
Adapter	Part Whole
Adapter	Tree Structure
Adapter	Convert Interface
...	...

4.3 แบบจำลองการคำนวณค่าถ่วงน้ำหนัก

การคำนวณน้ำหนักของเอกสารเป็นส่วนสำคัญอีกส่วนหนึ่งในการบ่งชี้ว่าระบบสืบค้นสารสนเทศนั้นมีประสิทธิภาพในการสืบค้นเพียงใด เนื่องจากการกำหนดค่าถ่วงน้ำหนักให้กับดัชนีก็เปรียบเสมือนกับการกำหนดระดับความสามารถในการบ่งชี้เอกสารให้กับดัชนีตัวนั้น ดังนั้นถ้าระบบสืบค้นใดออกแบบวิธีการคำนวณค่าถ่วงน้ำหนักได้เหมาะสมกับโครงสร้างข้อมูลของเอกสารก็จะทำให้ความสามารถในการบ่งชี้เอกสารมีประสิทธิภาพการสืบค้นสูง แบบจำลองนี้ได้ออกแบบวิธีการคำนวณค่าถ่วงน้ำหนักของดัชนีให้เหมาะสมตามโครงสร้างข้อมูลของแบบรูปการออกแบบที่ได้กล่าวไว้ในหัวข้อที่ 4.1 โดยทำการจัดลำดับความสามารถในการบ่งชี้แบบรูปการออกแบบในแต่ละส่วนของโครงสร้างข้อมูลนั้นไม่เท่ากัน แสดงได้ดังรูปที่ 4.2



รูปที่ 4.2 แสดงลำดับความสามารถในการบ่งชี้โครงสร้างข้อมูลแบบรูปการออกแบบ

จากโครงสร้างข้อมูลแบบรูปการออกแบบในหัวข้อที่ 4.1 จะสามารถลำดับความสามารถในการบ่งชี้ได้ดัง รูปที่ 4.2 ซึ่งมีรายละเอียดในแต่ละส่วน ดังต่อไปนี้

ความสามารถในการบ่งชี้ลำดับที่ 1 คือ ชื่อหลักของแบบรูปการออกแบบ

ความสามารถในการบ่งชี้ลำดับที่ 2 คือ หมวดหมู่ของแบบรูปการออกแบบ

ความสามารถในการบ่งชี้ลำดับที่ 3 คือ ชื่อรอง คำเฉพาะ และ ฟังก์ชันหลักของหมวดหมู่

ความสามารถในการบ่งชี้ลำดับที่ 4 คือ ฟังก์ชันของแบบรูปการออกแบบ และ รูปแบบวัตถุ

ในการจัดลำดับความสามารถในการบ่งชี้ของโครงสร้างแบบรูปการออกแบบนี้สามารถวิเคราะห์ได้จากความสัมพันธ์ในแต่ละส่วนของโครงสร้างข้อมูลแบบรูปการออกแบบ(รายละเอียดดังรูปที่ 4.1) โดยความสามารถในการบ่งชี้ลำดับที่ 1 คือ ชื่อหลัก เนื่องจากมีค่าความสัมพันธ์เป็น 1 ลำดับที่ 2 คือ หมวดหมู่ เนื่องจากมีค่าความสัมพันธ์เป็น 0..1 นั่นหมายความว่าอาจไม่มีคำสำคัญในการบ่งชี้แบบรูปการออกแบบก็ได้ ในลำดับที่ 3 และ 4 นั้นมีค่าความสัมพันธ์เป็นเท่ากัน คือ 0..* นั่นหมายความว่าคำในส่วนนี้สามารถมีคำที่ใช้อธิบายแบบรูปการออกแบบได้หลายคำ ดังนั้นในงานวิจัยนี้ จึงนำเอาจำนวนคำสำคัญที่ใช้ในการบ่งชี้ถึงแบบรูปการออกแบบมาร่วมพิจารณาลำดับความสามารถในการบ่งชี้ โดยคำในลำดับที่ 3 นั้นเป็นคำที่สามารถบ่งชี้ได้ถึงแบบรูปการออกแบบได้ในคำเดียว เช่น “Create” บ่งชี้ฟังก์ชันหลักของหมวดหมู่การสร้าง ดังนั้นคำว่า “Create” นี้จึงสามารถบ่งชี้แบบรูปการออกแบบที่อยู่ในหมวดหมู่การสร้างได้ทั้งหมด หรือ “kit” คือ ชื่อรองของแอบสแทรกซ์แฟกทอรี จึงสามารถบ่งชี้ถึงแอบสแทรกต์แฟกทอรีได้โดยตรง เป็นต้น และคำในลำดับที่ 4 เป็นคำทั่วไปที่ใช้ในการอธิบายถึงแบบรูปการออกแบบ ซึ่งคำที่อยู่ในส่วนนี้อาจจะปรากฏอยู่ในการอธิบายแบบรูปการออกแบบตัวอื่นด้วยเหมือนกัน ดังนั้นจึงต้องใช้คำมากกว่า 1 คำร่วมในการบ่งชี้แบบรูปการออกแบบ เช่น “Single” คือ รูปแบบวัตถุของทั้งซิงเกิลตอน และฟาซาด โดยถ้าจะให้สามารถบ่งชี้ถึงฟาซาดได้นั้นจะต้องใช้คำมากกว่าหนึ่งคำ เช่น “Single“ “Simplify” เป็นต้น

ในการคำนวณค่าถ่วงน้ำหนักของดัชนีการสืบค้นแบบรูปการออกแบบนี้ได้ใช้การจัดลำดับความสามารถในการบ่งชี้แบบรูปการออกแบบดังกล่าวข้างต้น เป็นตัวร่วมในการบ่งชี้ว่าดัชนีตัวนั้นมีความสามารถในการบ่งชี้แบบรูปการออกแบบมากน้อยเพียงใด โดยกำหนดค่าคงที่ให้กับลำดับความสามารถในการบ่งชี้ ซึ่งมีหลักการว่าส่วนใดมีความสามารถในการบ่งชี้สูงก็จะมีค่าคงที่มาก และค่าคงที่นี้จะลดลงตามลำดับความสามารถในการบ่งชี้แบบรูปการออกแบบ ซึ่งการกำหนดค่าความสามารถในการบ่งชี้ให้กับดัชนีตามลำดับดังกล่าวจะนำความสามารถในการบ่งชี้ของคำมาใช้ได้อย่างเหมาะสม อย่างไรก็ตามผลลัพธ์การสืบค้นไม่ได้ขึ้นอยู่กับค่าคงที่ดังกล่าวเพียงอย่างเดียว แต่ต้องพิจารณาร่วมกับจำนวนคำสำคัญที่ใช้ในการสืบค้นด้วย โดยเบื้องต้นได้ยกตัวอย่างค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ไว้ตามตารางที่ 4.7 ซึ่งค่าเหล่านี้เป็นค่าที่ได้จากการทดลองซึ่งรายละเอียดอยู่ในบทที่ 5

ตาราง 4.7 แสดงค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของตัวแทนแบบรูปการออกแบบ

คำอธิบายแบบรูปการออกแบบ	ค่าคงที่แสดงลำดับความสามารถการบ่งชี้
ชื่อหลัก	0.5
หมวดหมู่	0.3
ชื่อรอง คำเฉพาะ และส่วนฟังก์ชันหลักของหมวดหมู่	0.2
ฟังก์ชันของแบบรูปการออกแบบ และรูปแบบวัตถุ	0.1

เมื่อได้ลำดับความสามารถในการบ่งชี้ในแต่ละดัชนีแล้ว ขั้นตอนต่อไปในการคำนวณค่าถ่วงน้ำหนัก คือ นำค่าคงที่แสดงลำดับความสามารถในการบ่งชี้มาคำนวณร่วมกับค่าความถี่ผกผันของดัชนี จากหัวข้อที่ 4.2 การสืบค้นเอกสารงานวิจัยนี้ได้ใช้ดัชนี 2 ประเภทด้วยกัน คือ ดัชนีแบบเดี่ยว และดัชนีแบบวลี ดังนั้นการคำนวณค่าถ่วงน้ำหนักของดัชนีจึงแบ่งเป็น 2 ประเภทเช่นกัน คือ การคำนวณค่าถ่วงน้ำหนักในดัชนีแบบเดี่ยว และการคำนวณค่าถ่วงน้ำหนักในดัชนีแบบวลี ซึ่งรายละเอียดสามารถอธิบายได้ดังหัวข้อ 4.3.1 และ 4.3.2

4.3.1 การคำนวณค่าถ่วงน้ำหนักในดัชนีเดี่ยว

การคำนวณค่าถ่วงน้ำหนักในดัชนีเดี่ยว สามารถทำได้โดยการบวกค่าคงที่แสดงลำดับความสามารถในการบ่งชี้เข้ากับค่าความถี่ผกผันของดัชนี โดยจะใช้ค่าคงที่ ที่ได้แสดงไว้ในตารางที่ 4.7 การคำนวณค่าถ่วงน้ำหนักของดัชนีเดี่ยวนี้จะแบ่งออกเป็น 4 กรณี ซึ่งจะกล่าวถึงในสมการที่ 4.1-4.4 โดยสมการทั้งหมดจะใช้ ตัวแปรในการคำนวณ ดังต่อไปนี้

t_i แทน ดัชนีลำดับที่ i โดยนิยามให้ $i \in \{1,2,3,...,M\}$ เมื่อ M คือ จำนวนดัชนีทั้งหมด

d_k แทน เอกสารลำดับที่ k โดยนิยามให้ $k \in \{1,2,3,...,N\}$ เมื่อ N คือ จำนวนเอกสารทั้งหมด

$idf(t_i)$ แทน ค่าความถี่ผกผันของดัชนีลำดับที่ i

$weight(t_i, d_k)$ แทน ค่าถ่วงน้ำหนักของดัชนีลำดับที่ i ในเอกสารลำดับที่ k

PC_Name แทน ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของชื่อ

$PC_Category$ แทน ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของหมวดหมู่

$PC_Keyword$ แทน ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของ ชื่อรอง คำเฉพาะ และส่วนฟังก์ชันหลักของหมวดหมู่

$PC_Function$ แทน ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของฟังก์ชันของแบบรูปการออกแบบ และรูปแบบวัตถุ

ก) กรณีที่ 1 ดัชนี คือชื่อแบบรูปการออกแบบ การค่าถ่วงน้ำหนักของดัชนีแสดงได้ดังสมการที่ 4.1

$$weight(t_i, d_k) = idf(t_i) + PC_Name \tag{4.1}$$

ข) กรณีที่ 2 ดัชนี คือ หมวดหมู่ของแบบรูปการออกแบบ การค่าถ่วงน้ำหนักของดัชนีแสดงได้ดังสมการที่ 4.2

$$weight(t_i, d_k) = idf(t_i) + PC_Category \tag{4.2}$$

ค) กรณีที่ 3 ดัชนี คือ ชื่อรอง คำเฉพาะ หรือฟังก์ชันหลักของหมวดหมู่ของแบบรูปการออกแบบ การค่าถ่วงน้ำหนักของดัชนี แสดงได้ดังสมการที่ 4.3

$$weight(t_i, d_k) = idf(t_i) + PC_Keyword \tag{4.3}$$

ง) กรณีที่ 4 ดัชนี คือ คือ ฟังก์ชันของแบบรูปการออกแบบ หรือรูปแบบวัตถุของแบบรูปการออกแบบ การค่าถ่วงน้ำหนัก แสดงได้ดังสมการที่ 4.4

$$weight(t_i, d_k) = idf(t_i) + PC_Function \tag{4.4}$$

4.3.2 การคำนวณค่าถ่วงน้ำหนักในดัชนีวลี

การคำนวณค่าถ่วงน้ำหนักดัชนีวลีมีหลักการ คือ นำค่าถ่วงน้ำหนักของดัชนีเดี่ยวมาบวกกัน เช่น เมื่อดัชนีวลีเกิดจากการนำดัชนีเดี่ยวมาประกอบกัน นั่นคือ ดัชนีเดี่ยวตัวที่ i และดัชนีเดี่ยวตัวที่ j ดังนั้นค่าถ่วงน้ำหนักสามารถหาได้โดย นำค่าถ่วงน้ำหนักของดัชนีเดี่ยวตัวที่ i มาบวกกับค่าถ่วงน้ำหนักของดัชนีเดี่ยวตัวที่ j ซึ่งการสร้างดัชนีวลีในงานวิจัยนี้มีได้ 2 ถึง 3 คำ นั่นคือ ดัชนีเดี่ยวตัวที่ i ดัชนีเดี่ยวตัวที่ j ดัชนีเดี่ยวตัวที่ l มาประกอบกับ แสดงได้ดังสมการที่ 4.5

$$phrase_weight(t_i + t_j + [t_l], d_k) = weight(t_i, d_k) + weight(t_j, d_k) + [weight(t_l, d_k)]$$

เมื่อ $i \neq j \neq l$ (4.5)

ในดัชนีวลีนี้สามารถใช้คำอธิบายเชิงวัตถุประกอบกับดัชนีเดี่ยวเพื่อสร้างเป็นดัชนีวลีได้ โดยงานวิจัยนี้ให้ความสำคัญกับดัชนีที่เป็นตัวแทนคำอธิบายเชิงวัตถุในระดับกลางเนื่องจากดัชนีอธิบายเชิงวัตถุนี้เป็นดัชนีที่เลือกใช้เฉพาะกับดัชนีวลีเท่านั้น ดังนั้นถ้ากำหนดค่าถ่วงน้ำหนักให้กับดัชนีอธิบายเชิงวัตถุน้อยไปก็จะไม่สามารถช่วยบ่งชี้ถึงเอกสารได้ หรือถ้ากำหนดให้มากเกินไปก็จะทำให้มีผลกระทบกับการสืบค้นของดัชนีอื่น โดยค่าถ่วงน้ำหนักสามารถคำนวณได้จากสมการ ที่ 4.6

$$phrase_weight(t_i + [t_j] + t_object, d_k) = weight(t_i, d_k) + [weight(t_j, d_k)] + PC_Object$$

เมื่อ $i \neq j$ และ t_object สามารถอยู่ตำแหน่งใดในดัชนีวลีก็ได้ (4.6)

ค่า PC_Object นี้ หาได้จากการนำค่า $Max\,idf$ หารด้วย 2 เพื่อหาค่ากลางในการบ่งชี้เอกสาร เนื่องจากค่า idf แสดงถึงความสำคัญของดัชนีต่อทุกเอกสาร (รายละเอียดดังหัวข้อที่ 3.1.1.3) ดังนั้นค่า $Max\,idf$ ซึ่งเป็นค่าสูงสุดของค่า idf ซึ่งหาได้จากดัชนีที่ปรากฏในเอกสารเพียงฉบับเดียวเท่านั้น เมื่อทำการหาร 2 จึงได้เป็นค่ากลาง แสดงได้ดังสมการที่ 4.7 และตัวอย่างการคำนวณค่า PC_Object แสดงได้ดังตัวอย่างที่ 4.2

$$PC_Object = \frac{Max\ idf}{2} \tag{4.7}$$

โดยที่

$t_i, t_j\ t_l$ แทน ดัชนีลำดับที่ i, j และ l ตามลำดับโดยนิยามให้ $i, j, l \in \{1,2,3,...,M\}$ เมื่อ

M แทน จำนวนดัชนีทั้งหมด

d_k แทน เอกสารลำดับที่ k โดยนิยามให้ $k \in \{1,2,3,...,N\}$ เมื่อ N คือ จำนวนเอกสารทั้งหมด

$Max\ idf$ แทน ค่าสูงสุดของค่าความถี่ผกผัน

$weight(t_i, d_k)$ แทน ค่าถ่วงน้ำหนักดัชนีเดี่ยวลำดับที่ i ในเอกสารลำดับที่ k

$weight(t_j, d_k)$ แทน ค่าถ่วงน้ำหนักดัชนีเดี่ยวลำดับที่ j ในเอกสารลำดับที่ k

$weight(t_l, d_k)$ แทน ค่าถ่วงน้ำหนักดัชนีเดี่ยวลำดับที่ l ในเอกสารลำดับที่ k

$phrase_weight(t_i + t_j + [t_l], d_k)$ แทน ค่าถ่วงน้ำหนักดัชนีวลี ในดัชนีลำดับที่ i ที่ประกอบกับดัชนีลำดับที่ j ซึ่งเป็นค่าถ่วงน้ำหนักของดัชนีวลี 2 คำ การประกอบดัชนีนี้อาจจะเพิ่มดัชนี

ลำดับที่ l เพื่อเป็นการคำนวณค่าถ่วงน้ำหนักของดัชนี 3 คำได้ โดยดัชนีเหล่านี้ต้องอยู่ในเอกสารลำดับที่ k

$phrase_weight(t_i + [t_j] + t_object, d_k)$ แทน ค่าถ่วงน้ำหนักดัชนีวลีในดัชนีลำดับที่ i ประกอบกับดัชนีที่เป็นคำอธิบายเชิงวัตถุ t_object เพื่อใช้เป็นค่าถ่วงน้ำหนักของดัชนีวลี 2 คำ หรือการประกอบนี้อาจจะเพิ่มดัชนีลำดับที่ j เพื่อใช้เป็นค่าถ่วงน้ำหนักของดัชนี 3 คำได้ โดยดัชนีเหล่านี้ต้องอยู่ในเอกสารลำดับที่ k

t_object แทน ดัชนีที่เป็นคำอธิบายเชิงวัตถุ

PC_Object แทน ค่าถ่วงน้ำหนักของดัชนีที่เป็นคำอธิบายเชิงวัตถุ

การคำนวณค่าถ่วงน้ำหนักของดัชนีเดี่ยว และดัชนีวลี สามารถแสดงตัวอย่างการคำนวณได้ดังต่อไปนี้

ตัวอย่าง 4.1 แสดงการคำนวณค่าถ่วงน้ำหนักดัชนีแบบเดี่ยวของเอกสาร แอบสแทรกต์แฟกทอรี

โดยสมมติให้เอกสาร d_1 คือ แอบสแทรกต์แฟกทอรี มีดัชนีแบบเดี่ยว คือ "Abstract", "Factory", "Creational", "Create" และ"Relate" โดยดัชนีเหล่านี้สามารถคำนวณหาค่าถ่วงน้ำหนักได้ดังขั้นตอนต่อไปนี้

ขั้นตอนแรก คือ แบ่งแยกดัชนีตามลำดับความสามารถในการบ่งชี้ ตารางที่ 4.8 แสดงการกำหนดความสามารถในการบ่งชี้ของดัชนีในเอกสารแอบสแทรกต์แฟกทอรี

ตารางที่ 4.8 แสดงรายละเอียดความสามารถในการบ่งชี้ของแต่ละดัชนีในตัวอย่างที่ 4.1

ลำดับ	ดัชนี	คำอธิบายแบบรูปการออกแบบ	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของดัชนี
1	Abstract	ชื่อหลัก	0.5
2	Factory	ชื่อหลัก	0.5
3	Creational	หมวดหมู่	0.3
4	Create	ฟังก์ชันหลักของหมวดหมู่	0.2
5	Relate	รูปแบบวัตถุ	0.1
6	Family	รูปแบบวัตถุ	0.1

ขั้นตอนที่ 2 คือ นำค่าความถี่ผกผันของดัชนีแต่ละตัวมาบวกเข้ากับ ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของดัชนี ที่แสดงในตารางที่ 4.7โดยค่าความถี่ผกผันของดัชนีหาได้จากสมการ ดังต่อไปนี้

$$idf(t_i) = \log\left(\frac{N}{n_k}\right)$$

เมื่อ idf_k แทน ค่าความถี่ผกผันของเอกสาร

n_k แทน จำนวนเอกสารที่เจอดัชนี $k \in \{1,2,3,...,N\}$ เมื่อ N คือ จำนวนเอกสารทั้งหมด

ตัวอย่างการหาค่าความถี่ผกผันของดัชนี "Creational" แสดงได้ดังนี้ เมื่อ $N = 23, n_k = 5$

$$idf(t_3) = \log\left(\frac{23}{5}\right)$$

$$= 0.6627$$

เนื่องจากดัชนี "Creational" มีความสามารถในการบ่งชี้ คือ เป็นหมวดหมู่ของแบบรูปการออกแบบ ดังนั้น ค่าถ่วงน้ำหนักของดัชนี มีค่าดังต่อไปนี้

$$weight(t_3, d_1) = idf(t_3) + PC_Category$$

$$= 0.6627 + 0.3$$

$$= 0.9627$$

โดยค่าถ่วงน้ำหนักดัชนีเดี่ยวของแอบสแทรกต์แฟกทอรี สามารถแสดงได้ดังตารางที่ 4.9

ตารางที่ 4.9 แสดงค่าถ่วงน้ำหนักดัชนีแบบเดี่ยวของเอกสารแอบสแทรกต์แฟกทอรี

ดัชนี	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้ของดัชนี	ค่าความถี่ผกผันของดัชนี	ค่าถ่วงน้ำหนักของดัชนี
Abstract	0.5	1.3617	1.8617
Factory	0.5	1.0606	1.5606
Creational	0.3	0.6627	0.9627
Create	0.2	0.6627	0.8627
Relate	0.1	1.3617	1.4617

ตัวอย่าง 4.2 แสดงการคำนวณค่าถ่วงน้ำหนักของดัชนีที่เป็นคำอธิบายเชิงวัตถุโดยใช้ข้อมูลเดียวกับตัวอย่างที่ 4.1

$$PC_Object = \frac{Max\ idf}{2}$$
$$= \frac{\log\left(\frac{23}{1}\right)}{2} = 0.6808$$

ตัวอย่าง 4.3 แสดงการคำนวณค่าถ่วงน้ำหนักดัชนีวลีของเอกสาร Abstract Factory โดยใช้ข้อมูลค่าถ่วงน้ำหนักดัชนีเดี่ยว ดังตารางที่ 4.9 และใช้ค่าถ่วงน้ำหนักของดัชนีที่เป็นคำอธิบายเชิงวัตถุในตัวอย่างที่ 4.2

ให้เอกสาร d_1 ซึ่งคือแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี ซึ่งมีดัชนีวลี คือ "Abstract Factory", "Relate Object" และ"Family Object" โดยดัชนีเหล่านี้สามารถหาค่าถ่วงน้ำหนักได้ดังนี้

นำค่าถ่วงน้ำหนักของดัชนีเดี่ยวแต่ละตัวมาคำนวณหาค่าถ่วงน้ำหนักดัชนีวลี โดยค่าถ่วงน้ำหนักของดัชนีวลี "Abstract Factory" มีค่าดังต่อไปนี้

$$phrase_weight(t_1 + t_2, d_1) = weight(t_1, d_1) + weight(t_2, d_1)$$

ค่าถ่วงน้ำหนักของดัชนีวลี "Relate Object" มีค่าดังต่อไปนี้

$$phrase_weight(t_4 + t_object, d_1) = weight(t_4, d_1) + PC_Object$$
$$= 1.4617 + 0.6808$$
$$= 2.1425$$

บทที่ 5
การประเมินแบบจำลองการสืบค้นแบบรูปการออกแบบ

บทที่ 5 กล่าวถึงการทดลองและการประเมินผลลัพธ์ของแบบจำลองการสืบค้นแบบรูปการออกแบบ โดยแบ่งออกเป็น 3 คือ ส่วนแรกจะกล่าวถึงวิธีการหาค่าความคล้ายคลึงระหว่างเอกสารและคำถาม ส่วนที่สองจะกล่าวถึงวิธีการประเมินผลลัพธ์ในการสืบค้น และส่วนสุดท้ายกล่าวถึงการทดลอง ซึ่งประกอบด้วย วิธีการทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลองที่ได้จากการสืบค้นแบบรูปการออกแบบ

5.1 วิธีการหาความคล้ายคลึงระหว่างคำถามและเอกสาร

การหาความคล้ายคลึงระหว่างคำถามและเอกสารในแบบจำลองการสืบค้นแบบรูปการออกแบบนี้ได้เลือกใช้แบบจำลองเวกเตอร์สเปซ ในการคำนวณหาค่าความคล้ายคลึง โดยในกระบวนการเวกเตอร์สเปซนี้แต่ละเอกสารจะถูกแทนด้วยเวกเตอร์ของดัชนี และคำถามจะถูกแทนด้วยเวกเตอร์ของคำที่ใช้ในการ โดยเวกเตอร์คำถามจะถูกจับคู่กับเวกเตอร์เอกสาร และทำการคำนวณหาค่าความคล้ายคลึงกันระหว่างคำถามและดัชนีของเอกสาร (รายละเอียดในหัวข้อที่ 3.1.2)

5.2 การประเมินผลการสืบค้น

วิธีการวัดประสิทธิผลของแบบจำลองการสืบค้นในงานวิจัยนี้ ทำได้โดยการวัดค่าความแม่นยำ ค่าความระลึก และค่าเฉลี่ยความแม่นยำ ซึ่งแสดงได้ดังสมการที่ 5.1 5.2 และ 5.3 (รายละเอียดในหัวข้อที่ 3.1.4)

$$\text{ค่าความแม่นยำ } (P) = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารทั้งหมดที่สืบค้นออกมาได้}} \quad (5.1)$$

$$\text{ค่าความระลึก } (R) = \frac{\text{จำนวนเอกสารที่ถูกต้องที่สืบค้นได้}}{\text{จำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล}} \quad (5.2)$$

$$\text{ค่าเฉลี่ยความแม่นยำ } (\overline{P}(r)) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q} \quad (5.3)$$

เมื่อ $\overline{P}(r)$ แทน ค่าเฉลี่ยความแม่นยำ ณ ค่าระลึกลำดับที่ r

$P_i(r)$ แทน ค่าความแม่นยำ ณ ค่าระลึกลำดับที่ r สำหรับคำถามที่ i

N_q แทน จำนวนคำถามทั้งหมดที่ใช้ในการหาค่าเฉลี่ยความแม่นยำ

5.3 การทดลอง

5.3.1 วิธีการทดลอง

ขั้นตอนการทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบในงานวิจัยนี้เริ่มจากการนำเอกสารมาทำดัชนี ซึ่งผู้วิจัยได้ทำการพัฒนาดัชนีตามแบบจำลองที่ได้อธิบายไว้ในบทที่ 4 จากนั้นจึงนำชุดคำถาม ไปสืบค้นโดยการจับคู่การสืบค้นระหว่างคำถามกับเอกสารแล้วทำการหาค่าความคล้ายคลึง (รายละเอียดในหัวข้อที่ 5.1) ผลการสืบค้นที่ได้นั้นจะถูกนำไปวิเคราะห์โดยผู้ชำนาญแบบรูปการออกแบบเพื่อวัดประสิทธิผลในการสืบค้น โดยประสิทธิผลการสืบค้นวัดได้จากค่าความแม่นยำ และค่าความระลึก (รายละเอียดในหัวข้อที่ 5.2)

การทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบนี้เริ่มต้นจะทำการสุ่มเลือกตัวเลขที่แสดงความสามารถในการบ่งชี้ของแต่ละตัวแทนแบบรูปการออกแบบ ซึ่งตัวเลขนี้จะมากน้อยลดหลั่นกันโดยมีหลักการว่าให้ช่วงตัวเลขที่แสดงความสามารถในการบ่งชี้ของชื่อและหมวดหมู่มีมากกว่าช่วงตัวเลขที่แสดงความสามารถในการบ่งชี้ระหว่างหมวดหมู่, ชื่อรอง คำเฉพาะ ส่วนฟังก์ชันหลักของหมวดหมู่, ฟังก์ชันของแบบรูปการออกแบบ รูปแบบวัตถุ เนื่องจากชื่อสามารถบ่งชี้ถึงแบบรูปการออกแบบได้มากที่สุดและแต่ละแบบรูปการออกแบบมีเพียงชื่อเดียวเท่านั้น จากนั้นจึงทดลองสืบค้นตามขั้นตอนที่ได้อธิบายไว้ข้างต้น และทำการปรับตัวเลขค่าคงที่จนกระทั่งได้ค่าคงที่ที่สามารถให้ประสิทธิผลการสืบค้น 70 เปอร์เซ็นต์

ข้อมูลที่ใช้ในการทดลอง มีดังต่อไปนี้

1) คำถามที่ใช้ในการทดลองทั้งสิ้น 105 คำถาม รายละเอียดคำถามที่ใช้ในการทดลองสามารถดูได้ที่ภาคผนวก ก โดยคำถามที่ใช้ในการทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบในงานวิจัยนี้นำมาจาก

ก) นักศึกษาที่มีความรู้เบื้องต้นเกี่ยวกับแบบรูปการออกแบบ 10 คน

ข) คำอธิบายแบบรูปการออกแบบรูปแบบต่างๆในอินเตอร์เน็ต

2) เอกสารที่ใช้ทดลองแบบจำลองการสืบค้นแบบรูปการออกแบบในงานวิจัยนี้มี 23 เอกสาร ซึ่งทั้ง 23 เอกสารเป็นแบบรูปการออกแบบที่อธิบายไว้ใน [10]

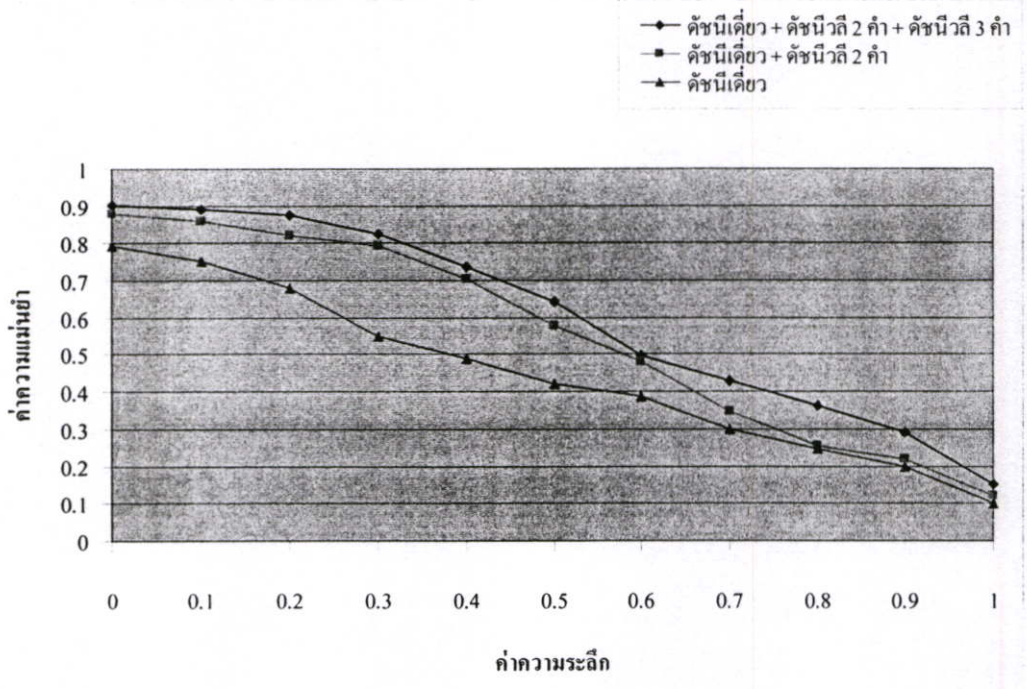
3) ดัชนีที่ใช้ในการสืบค้นทั้งสิ้น 290 ตัว โดยแบ่งเป็น ดัชนีเดี่ยว 210 ตัว, ดัชนีวลี 2 คำ 50 ตัว และ ดัชนีวลี 3 คำ 30 ตัว โดยรายละเอียดดัชนีกับ ค่าถ่วงน้ำหนักของดัชนีที่ใช้ในการทดลองสามารถดูได้ที่ภาคผนวก ข

5.3.2 ผลการทดลอง

จากการทดลองโดยใช้ค่าคงที่ความสามารถในการบ่งชี้ ดังตารางที่ 5.1จะให้ผลการทดลองเป็นดังรูปที่ 5.1

ตารางที่ 5.1 แสดงค่าความสามารถในการบ่งชี้ของตัวแทนแบบรูปการออกแบบ

คำอธิบายแบบรูปการออกแบบ	ค่าคงที่แสดงลำดับความสามารถการบ่งชี้
ชื่อหลัก	0.5
หมวดหมู่	0.3
ชื่อรอง คำเฉพาะ และส่วนฟังก์ชันหลักของหมวดหมู่	0.2
ฟังก์ชันของแบบรูปการออกแบบ และรูปแบบวัตถุ	0.1



รูปที่ 5.1 แสดงผลการทดลองการใช้แบบจำลองการสืบค้นแบบรูปการออกแบบ

รูปที่ 5.1 แสดงผลการทดลองการใช้แบบจำลองการสืบค้นแบบรูปการออกแบบ จากกราฟจะสรุปได้ว่าค่าความแม่นยำเมื่อเทียบกับค่าความระลึกจะมีแนวโน้มในลักษณะเชิงผกผันกัน นั่นคือผลการสืบค้นมีความแม่นยำสูงในค่าความระลึกต่ำ หมายความว่า แบบจำลองการสืบค้นนี้ สามารถสืบค้นเอกสารที่ต้องการได้ในลำดับต้นๆ และจากการคำนวณค่าเฉลี่ยความแม่นยำที่ช่วงค่าความระลึก 11 จุดแสดงได้ดังตารางที่ 5.2 กล่าวคือ การใช้ดัชนีเดี่ยวและดัชนีวลี 2 คำร่วมกันสืบค้นแบบรูปการออกแบบนั้นให้ผลดีขึ้นกว่าการใช้ดัชนีเดี่ยว 13% และการใช้ ดัชนีเดี่ยว ดัชนีวลี 2 คำ และดัชนี 3 คำร่วมกันนั้นทำให้มีความสามารถในการสืบค้นเพิ่มขึ้น 16% เมื่อเทียบกับการใช้ดัชนีเดี่ยว นั่นหมายความว่า การใช้ดัชนีวลีช่วยให้ความสามารถในการสืบค้นแบบรูปการออกแบบสามารถทำได้แม่นยำมากยิ่งขึ้น

ตาราง 5.2 แสดงค่าเฉลี่ยความแม่นยำในค่าความระลึก 11 จุด

รูปแบบดัชนีที่ใช้ในการสืบค้น	ค่าเฉลี่ยความแม่นยำ	ร้อยละ
ดัชนีเดี่ยว	0.548182	54.8 %
ดัชนีเดี่ยว + ดัชนีวลี 2 คำ	0.671429	67 .1%
ดัชนีเดี่ยว + ดัชนีวลี 2 คำ + ดัชนีวลี 3 คำ	0.705120	70.5 %

5.3.3 วิเคราะห์การใช้แบบจำลองสืบค้นแบบรูปการออกแบบ

จากผลการทดลองจะเห็นได้ว่า แบบจำลองสามารถสืบค้นแบบรูปการออกแบบได้ตรงกับคำถามที่ผู้ใช้ป้อนร้อยละ 70 อย่างไรก็ตามมีลักษณะคำถามบางคำถามที่อาจเป็นผลทำให้ค่าความแม่นยำการสืบค้นแบบรูปการออกแบบลดลง ดังนี้

1) คำถามบางคำถามอาจทำให้การสืบค้นได้เอกสารที่ไม่เกี่ยวข้องมาด้วย มีสาเหตุเนื่องมาจากคำสำคัญที่ใช้ในการสืบค้นเอกสารอาจจะปรากฏอยู่ในเอกสารอื่นด้วย ดังตารางที่ 5.3 แสดงตัวอย่างคำถาม “I want design pattern that can create class for convert interface “ซึ่งผลลัพธ์ที่ได้จากการสืบค้นควรเป็นแบบรูปการออกแบบอะแดปเตอร์ แต่เนื่องจากคำถามมีคำสำคัญ “Create” ซึ่งเป็นคำที่แสดงถึงฟังก์ชันหลักของหมวดหมู่การสร้าง ดังนั้นแบบรูปการออกแบบที่อยู่ในหมวดหมู่การสร้างจึงถูกสืบค้นขึ้นมาด้วย

ตาราง 5.3 แสดงตัวอย่างการสืบค้นคำถาม "I want design pattern that can create class for convert interface"

ลำดับที่ค้นพบ	คำสำคัญที่ค้นพบ	แบบรูปการออกแบบ	ความเหมาะสมของเอกสารที่สืบค้น
1	"convert interface"	Adapter	เหมาะสม
2	"create"	Abstract Factory	ไม่เหมาะสม
3	"create"	Factory Method	ไม่เหมาะสม
4	"create"	Prototype	ไม่เหมาะสม
5	"create"	Builder	ไม่เหมาะสม
6	"create"	Singleton	ไม่เหมาะสม

2) คำถามบางคำถามอาจได้เอกสารที่ไม่เกี่ยวข้องขึ้นมาเป็นลำดับต้นกว่าเอกสารที่เกี่ยวข้อง มีสาเหตุเนื่องมาจากผู้ป้อนคำถามใส่คำสำคัญของเอกสารอื่นมากกว่าเอกสารที่ต้องการ ดังตารางที่ 5.4 แสดงตัวอย่างคำถาม "I want design pattern that notify dependent or relate object "ซึ่งผลลัพธ์ที่ได้จากการสืบค้นควรเป็นแบบรูปการออกแบบออบเซิฟเวอร์ แต่เนื่องจากคำถามมีคำสำคัญ "Relate object" ซึ่งดัชนีวลีที่ใช้ในการบ่งชี้ถึงแบบรูปการออกแบบแอบสแทรกต์แฟกทอรี ดังนั้นจึงมีความสามารถในในการบ่งชี้สูงกว่าดัชนีเดี่ยวจึงได้แบบรูปการออกแบบแอบสแทรกต์แฟกทอรีขึ้นมาเป็นอันดับแรก

ตาราง 5.4 แสดงตัวอย่างการสืบค้นคำถาม "I want design pattern that notify dependent or relate object "

ลำดับที่ค้นพบ	คำสำคัญที่ค้นพบ	แบบรูปการออกแบบ	ความเหมาะสมของเอกสารที่สืบค้น
1	"dependent" "relate object"	Abstract Factory	ไม่เหมาะสม
2	"notify" "dependent"	Observer	เหมาะสม

จากตัวอย่างการสืบค้นแบบรูปการออกแบบข้างต้นสรุปได้ว่าข้อผิดพลาดในการสืบค้นอาจเกิดขึ้นจากคำถามที่ป้อนประกอบด้วยคำสำคัญของเอกสารที่ต้องการและเอกสารอื่นๆที่ไม่ต้องการ จึงทำให้ผลการสืบค้นได้เอกสารที่ไม่ต้องการขึ้นมาด้วย หรืออีกรณีหนึ่ง ข้อผิดพลาดในการสืบค้นนั้นเกิดขึ้นจากค่าถ่วงน้ำหนักของดัชนี ในกรณีนี้วิธีการลดข้อผิดพลาดอาจทำโดยการใช้วิธีการ

ป้อนกลับของผู้ใช้ (รายละเอียดในหัวข้อที่ 3.1.3) มาทำปรับค่าถ่วงน้ำหนักของคำถามให้เหมาะสมสำหรับสืบค้น แสดงได้ดังตัวอย่างที่ 5.1

ตัวอย่างที่ 5.1 แสดงตัวอย่างการปรับค่าถ่วงน้ำหนักของคำถาม ในตัวอย่างการสืบค้นของตารางที่ 5.4 โดยใช้วิธีของ Rocchio ซึ่งค่าถ่วงน้ำหนักของดัชนีในเอกสาร และคำถาม ที่ใช้ในการคำนวณค่าความคล้ายคลึงแสดงได้ดังตารางที่ 5.5

ตารางที่ 5.5 ค่าถ่วงน้ำหนักของดัชนี "Dependent", "Notify" และ "Relate object" ในคำถาม Q : "I want design pattern that notify dependent or relate object " และเอกสารแอบสแทรกต์แฟกทอรีกับออบเซิฟเวอร์

	Dependent	**Notify**	**Relate object**
Q	1.060698	1.261728	2.823455
Abstract Factory	1.160698	0	2.823455
Observer	1.160698	1.361728	0

จากตารางที่ 5.4 สามารถทำการปรับค่าถ่วงน้ำหนักของดัชนีให้กับคำถาม โดยการใช้วิธีป้อนกลับของ Rocchio ได้ดังนี้กำหนดให้ค่าพารามิเตอร์ $\alpha = 1, \beta = \frac{1}{2}, \gamma = \frac{1}{4}$

$$Q' = \alpha Q_0 + \frac{\beta}{|D_u|}\sum_{d_r \in D_u} d_r - \frac{\gamma}{|D_v|}\sum_{d_i \in D_v} d_i$$

$$Q' = (1.060698,1.261728,2.823455) + \frac{1}{2}(1.160698,1.361728,0)$$

$$-\frac{1}{4}(1.160698,0,2.823455)$$

$$= (1.516047,1.945368,2.11759125)$$

ผลลัพธ์การปรับค่าถ่วงน้ำหนักของคำถามสามารถสรุปได้ดังนี้

1. ค่าถ่วงน้ำหนักของ Dependent ถูกปรับเพิ่มขึ้นจาก 1.060698 เป็น 1.516047 หมายความว่าในคำถาม Q ดัชนี Dependent มีความสามารถในการบ่งชี้ถึงเอกสารที่เหมาะสมกับคำถามได้ ดังนั้นค่าถ่วงน้ำหนักจึงถูกปรับเพิ่มขึ้น

2. ค่าถ่วงน้ำหนักของ Notify ถูกปรับเพิ่มจาก 1.261728 เป็น 1.945368 นั้นหมายความว่าในคำถาม Q ดัชนี Notify มีความสามารถในการบ่งชี้ถึงเอกสารที่เหมาะสมกับคำถามได้ ดังนั้นค่าถ่วงน้ำหนักจึงถูกปรับเพิ่มขึ้น

3. ค่าถ่วงน้ำหนักของ Relate object ถูกปรับลดลงจาก 2.8234551เป็น 2.11759125 นั้นหมายความว่าในคำถาม Q ดัชนี Relate object ไม่มีความสามารถในการบ่งชี้ถึงเอกสารที่เหมาะสมกับคำถามได้ ดังนั้นค่าถ่วงน้ำหนักจึงถูกปรับลดลง

ในการปรับค่าถ่วงน้ำหนักของคำถามนี้จะทำให้ได้ผลลัพธ์ในการสืบค้นดีขึ้น อย่างไรก็ตามการปรับค่าถ่วงน้ำหนักนี้อาจต้องทำการวนซ้ำเพื่อให้ได้ค่าถ่วงน้ำหนักที่เหมาะสม ซึ่งคำถามในตารางที่ 5.4 นี้ จะต้องทำการปรับน้ำหนักวนซ้ำ 3 รอบ เพื่อให้ได้ผลลัพธ์ที่ดีขึ้น ดังนี้
การปรับค่าถ่วงน้ำหนักของคำถาม Q ในรอบที่ 2 แสดงได้ดังนี้

$$\begin{aligned} Q' &= (1.516047,1.945368,2.11759125) + \frac{1}{2}(1.160698,1.361728,0) \\ &\quad -\frac{1}{4}(1.160698,0,2.823455) \\ &= (1.80622155,2.62657,1.4117275) \end{aligned}$$

การปรับค่าถ่วงน้ำหนักของคำถาม Q ในรอบที่ 3 แสดงได้ดังนี้

$$\begin{aligned} Q' &= (1.80622155,2.62657,1.4117275) + \frac{1}{2}(1.160698,1.361728,0) \\ &\quad -\frac{1}{4}(1.160698,0,2.823455) \\ &= (2.09639605,3.31021,0.70586375) \end{aligned}$$

โดยเมื่อทำการปรับค่าถ่วงน้ำหนักในคำถาม Q วนซ้ำ 3 รอบแล้วจะได้ผลลัพธ์การสืบค้นที่ดีขึ้นแสดงได้ดังแสดงได้ในตารางที่ 5.6

ตารางที่ 5.6 แสดงผลลัพธ์การสืบค้นหลังการปรับค่าถ่วงน้ำหนักของคำถามในตารางที่ 5.4

ลำดับที่ค้นพบ	แบบรูปการออกแบบ	ความเหมาะสมของเอกสารที่สืบค้น
1	Observer	เหมาะสม
2	Abstract Factory	ไม่เหมาะสม

บทที่ 6
สรุปและข้อเสนอแนะ

6.1 สรุป

แบบรูปการออกแบบเป็นเทคนิคหนึ่งที่ช่วยให้การออกแบบซอฟต์แวร์เชิงวัตถุ สามารถทำได้อย่างรวดเร็ว และมีประสิทธิผล โดยช่วยแก้ไขปัญหาทั่วไปที่เกิดขึ้นในการออกแบบซอฟต์แวร์เชิงวัตถุ อย่างไรก็ตามเป็นการยากสำหรับนักออกแบบที่ยังไม่มีประสบการณ์การออกแบบมากนักจะสามารถเลือกใช้แบบรูปการออกแบบได้อย่างเหมาะสมกับปัญหาที่ประสบ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อ พัฒนาแบบจำลองการสืบค้นแบบรูปการออกแบบ ซึ่งสามารถสืบค้นได้ตามลักษณะโดเมนของปัญหา โดยสามารถสืบค้นตามคำสำคัญที่ผู้ใช้ป้อนเกี่ยวกับปัญหาแล้วทำการจัดลำดับให้ว่าผลลัพธ์ใดคล้ายคลึงกับคำถามที่สุด โดยแบบจำลองการสืบค้นแบบรูปการออกแบบนี้จะประกอบไปด้วย การวิเคราะห์เอกสารเพื่อสร้างเป็นดัชนีคำค้น และการออกแบบวิธีคำนวณค่าถ่วงน้ำหนักให้เหมาะสมกับดัชนีที่ใช้สืบค้น โดยได้ใช้แบบจำลองเวกเตอร์สเปซในการคำนวณหาค่าความคล้ายคลึงระหว่างคำถามและเอกสาร

การทดสอบประสิทธิผลแบบจำลองการสืบค้นนี้ใช้แบบรูปการออกแบบของจีโอเอฟเป็นเอกสารในการทดสอบแบบจำลอง และคำถามในการทดสอบนี้นำมาจากนักศึกษาที่มีความรู้เบื้องต้นเกี่ยวกับแบบรูปการออกแบบจำนวน 10 คน และคำอธิบายแบบรูปการออกแบบต่างๆในอินเตอร์เน็ต โดยผู้วิจัยเป็นผู้เชี่ยวชาญในการวิเคราะห์ความเหมาะสมและไม่เหมาะสมของผลลัพธ์การสืบค้น ซึ่งจากการทดลองสามารถสรุปได้ว่าแบบจำลองการสืบค้นแบบรูปการออกแบบนี้สามารถสืบค้นได้ผลลัพธ์ที่ครอบคลุมกับคำสำคัญของโดเมนปัญหาที่ผู้ใช้ป้อนและผลลัพธ์ที่มีความคล้ายคลึงกับคำถามมากที่สุดนั้นสามารถสืบค้นออกมาได้ในลำดับต้นๆ จากการทดลองสืบค้นโดยคำถาม 105 คำถามจะได้ค่าเฉลี่ยความแม่นยำของแบบจำลองการสืบค้นแบบรูปการออกแบบนี้ร้อยละ 70

6.2 ข้อเสนอแนะ

งานวิจัยเบื้องต้นนี้จะทำการพิจารณาน้ำหนักของดัชนีที่ใช้ในการสืบค้น แล้วคำนวณหาค่าความคล้ายคลึงระหว่างเอกสารและคำถาม จากนั้นจึงจัดลำดับผลลัพธ์ที่ได้ โดยให้เอกสารที่คล้ายคลึงกับคำถามมากที่สุดอยู่ในลำดับต้นๆ ซึ่งการพิจารณาน้ำหนักของดัชนีเพียงส่วนเดียวนี้อาจจะให้ผลลัพธ์ที่ไม่ตรงกับคำถามมากนัก แนวทางที่ทำให้แบบจำลองมีประสิทธิผลในการสืบค้น

มากขึ้น คือ การเพิ่มการวิเคราะห์คำถามที่ผู้ใช้ป้อนให้มีความถูกต้องมากยิ่งขึ้น เช่น การวิเคราะห์ไวยากรณ์ (Syntactic Analysis) ของประโยค หรือการวิเคราะห์ความหมาย (Semantic Analysis)

เอกสารอ้างอิง

[1] Aleksandra, T. "What is a Pattern ?." **IEEE Computer Society Press. 1996.**

[2] Andreas, G., Mattias, E. **"Formalizing the Intent of Design Patterns." An Approach Towards a Solution to the Indexing Problem Technical report 1999-006.** Uppsala University. 1999

[3] Baeza-Yates, R. and Berthier Ribiero-Neto(1999). **Modern Information Retrieval.** Addison Wesley.

[4] C. Lundquist, D. A. Grossman and O. Frieder. "Improving relevance feedback in the vector space model." **Proceedings of the sixth international conference on Information and knowledge management (CIKM '97). Las Vegas.** 1997. pp 16-23.

[5] Chris, B. **The Importance of Proper Weighting Methods.** [Online]. Available : http://www.cs.mu.oz.au/acl/H/H93/H93-1070.pdf

[6] Coplien, J. O. Schmidt, D. C. **Pattern Languages of Program Design**. Addison-Wesley. Reading. MA. 1995

[7] Daniel, L., Alexandre, A., Eduardo, S., Antonio, F. "MVCASE Tool-Working with Design Patterns." **The Third Latin American Conference On Pattern Languages Of Programming**. Porto de Galinhas – PE. 2003

[8] Daniel, P., Gyorgy, C. "Design Pattern Matching." **Periodica Polytechnica Ser. el. Eng. Vol. 47. No. 3-4.** 2003. pp. 205-212

[9] Frank Budinsky, Marilyn Finnie, Patsy Yu. "Automatic Code Generation from Design Patterns" **IBM Systems Journal**, 35, 2, 151-171.

[10] Gamma, E., Helm, R., Vlissides, J. **Design Pattern: Elements of Reusable Object Oriented Software**. Addison-Wesley Publishing Company, Addison-Wesley Professional Computing Series, New York, NY, 1977.

[11] Gerard Salton, A. Wong, and C. S. Yang. "A vector space model for information retrieval." **Communications of the ACM**. 1975. pp. 613-620.

[12] Gerard Salton and C. Buckley. "Improving retrieval performance by relevance feedback." **Journal of the American Society for Information Science**. 1990. pp 288-297.

เอกสารอ้างอิง (ต่อ)

[13] Gerard Salton, M.j. McGill, **Introduction to Modern Information Retrieval**. McGraw-Hill, New York, 1983

[14] Maguns Rosell. "Summary of "Knowledge-Based Text Summarization: Salience and Generalization Operators for Knowledge Base Abstraction" by Udo Hahn and Ulrich Reimer". **Department of Numerical Analysis and Computer Science**

[15] Kando, N. "Text-level structure of research papers: Implications for text-based information processing systems." **Proceedings of the British Computer Society Annual Colloquium of Information Retrieval Research**. April 1997, 68-81.

[16] Object Venture Inc. **Pattern and Component Markup Language**, 2002, [Online]. Available : http://www.objectventure.com/pcml.html.

[17] Philippe M., Laurence A. "Conceptual Structures and Structured Documents." **In Proceedings of ICCS 1996, 4th International Conference on Conceptual Structures**. pp. 145-159, Sydney. Australia. August 19-22. 1996..

[18] Smeaton, A. F. and Kelledy, F. "User-Chosen Phrases in Interactive Query Formulation for Information Retrieval." **In Proceedings of the 20th BCS-IRSG Colloquium, Grenoble,France, Springer-Verlag Workshops in Computing (1998).** 1998.

[19] Song M., Azriel R.,Tapas K. "Document Structure Analysis Algorithms: A Literature Surver." **Proceedings of SPIE Volume 5010 Document Recognition and Retrieval X. January 2003**. pp. 197-207

[20] Tao Qin, Lu Zhang, Zhiying Zhou, Dan Hao, and Jiasu Sun. "Discovering Use Cases from Source Code using the Branch-Reserving Call Graph." **In Proceedings of the Tenth Asia-Pacific Software Engineering Conference (APSEC'03).** 2003. pp 60-67

[21] Vlissides, J. M. Coplien, J. O. Kerth, N.L. **Pattern Languages of Program Design**. Vol. 2. Addison-Wesley. Reading. MA 1996

[22] Wei Zhao, Lu Zhang, Yin Liu, Jing Luo, Jiasu Sun. "Understanding How the Requirements Are Implements in Source Code." **In Proceedings of the Tenth Asia-Pacific Software Engineering Conference (APSEC' 03).** 2003 pp. 68

เอกสารอ้างอิง (ต่อ)

[23] Wesley W., Chu, F. Meng. "Database Query Formation from Natural Language using Semantic Modeling and Statistical Keyword Meaning Disambiguation." **Technical Report 990003. UCLA. USA.** 1999

[24] Yohei Seki. "Sentence Extraction by tf/idf and Position Weighting from Newspaper." **Proceedings of the Third NTCIR Workshop on Research in Information Retrieval, Automatic Text Summarization and Question Answering. NII. Japan. October 2002.** 2002. pp 71-77

[25] May 22 2006 [Online]. Available : http://c2.com/ppr/index.html

[26] May 22 2006 [Online]. Available : http://hillside.net/patterns/.

[27] May 22 2006 [Online]. Available : http://en.wikipedia.org/wiki/Design_Patterns

[28] May 22 2006 [Online]. Available : http://www.javacamp.org/designpattern/

ภาคผนวก

ภาคผนวก ก

แสดงคำถามที่ใช้ในการทดลองการสืบค้นแบบรูปการออกแบบ

ตาราง ก แสดงคำถามที่ใช้ในการทดลองการสืบค้นแบบรูปการออกแบบ

ลำดับคำถาม	คำถาม
1	Creates an instance of several families of classes
2	Allow client to create family object without specify their concrete class
3	Interface to create and return one of several family of relate…
4	Design a common interface for create family of similarly or relate object
5	Separate object construction from representation
6	Subclass decide which concrete class to create
7	Provide a simple decision making class that returns one of several ….
8	A fully initialized instance to be copy or clone
9	Starts with an initialized and instantiated class and copy or clone …
10	A class of which only a single instance can exist
11	Ensure one and only object to create
12	Design pattern that can change internal state
13	To make sure that there is precise one object in a class and provide access..
14	I want to create only one object
15	Wrap object and provide a different interface to it
16	Class interface match with another class
17	Class to convert the interface of one component into another interface…
18	Ccreate middleware for communication with another object
19	An abstract class must use for create object
20	Separates an object's interface from its implementation
21	A tree structure of simple and composite object
22	Client treat collection of object and individual object uniform
23	Make complex and simple kinds of an object share behavior
24	Add responsibility to object dynamically
25	Wrap an object to provide new behavior
26	A class can take an object and wrap new functions around it
27	A single class that represents an entire subsystem
28	Single class represent an entire subsystem
29	Separates an object's interface from its implementation

ตาราง ก (ต่อ)

ลำดับคำถาม	คำถาม
30	Add responsibility to object dynamically
31	Class or object provides a single point of entry for services of a subsystem
32	Hide a complex system of objects behind a single object
33	Provide unify interface to set of interface in subsystem
34	Define higher level interface that make subsystem easier to use
35	Simplify the interface of a set of class
36	A fine grain instance use for efficient share
37	Pattern for share object
38	Share of objects to save space
39	An object represent another object
40	Wrap an object to control access to it
41	A way of passing a request between a chain of object
42	Allows an even further decoupling between class
43	Separates the construction of a complex object from its representation
44	Encapsulate a command request as an object
45	Encapsulate a request as an object
46	Simple way to separate execution of a command from the interface …
47	A way to include language element in a program
48	Sequentially access the elements of a collection
49	Provide a way to traverse a collection of object without expose …
50	Defines simplify communication between class
51	Communication between classes simplify by using another class…
52	Capture and restore an object's internal state
53	A way of notify change to a number of classes
54	Allow object to be notify when state change
55	Way a number of classes can be notified of a change
56	Alter an object's behavior when its state change
57	Encapsulate state base behavior and use delegation to switch between …
58	When an object changes behavior responsibility with time place the state…

ตาราง ก (ต่อ)

ลำดับคำถาม	คำถาม
59	Encapsulates an algorithm inside a class
60	Encapsulate interchangeable behavior and use delegation to decide ...
61	Defer the exact steps of an algorithm to a subclass
62	Subclass decide how to implement step in an algorithm
63	How a compose object gets created
64	Structure and composition of an object
65	Compose object like hierarchy
66	Create class for convert interface
67	Match interface with another
68	Create family of relate object
69	Class of object that is instantiate
70	Subclass of object that is instantiate
71	Responsibility of an object without subclass
72	Single class to a control all subclass
73	Family of product object
74	Pattern that can create class for convert interface
75	Grammar and interpretation of a language
76	Separate process of create for crate complex object
77	Construction process can create different representation
78	Pattern that notify dependent or relate object
79	Class defer instantiation to subclass
80	Provide interface for create by without specify concrete class
81	Create only one object and provide global point of access
82	Convert interface to another
83	Let class work together
84	Copy object by prototype
85	Specify the kind of object to create using this prototypical instance
86	Provide unify interface for control subsystem
87	Tree structure to represent part-whole hierarchies

ตาราง ก (ต่อ)

ลำดับคำถาม	คำถาม
88	Compose tree object uniformly
89	Access element of an aggregate object sequentially
90	Object to alter its behavior when its internal state change
91	Skeleton of algorithm in one operation
92	Redefine certain steps of an algorithm without changing the algorithm
93	instantiate only one object for system
94	When object change it send message to all the dependent object
95	Send message to another object in chain
96	Left is an object that additional function
97	Dynamically attach additional responsibility
98	Notify object to change state
99	Compose tree object into tree structure
100	Incompatible interface work together
101	Encapsulate group of individual factory
102	Compose of one or more similar object
103	Way to notify change to a number of class
104	Interface for create dependent object
105	Left is an object that additional behavior

ภาคผนวก ข

แสดงดัชนี และค่าถ่วงน้ำหนักของดัชนี

ตาราง ข.1 แสดงดัชนี และค่าถ่วงน้ำหนักของดัชนี

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
1	Abstract	Abstract Factory	0.4	1.761728
2	Factory	Abstract Factory	0.4	1.460698
3	Abstraction	Abstract Factory	0.2	1.561728
4	Create	Abstract Factory	0.3	0.962758
5	Dependent	Abstract Factory	0.1	1.160698
6	Family	Abstract Factory	0.1	1.160698
7	Instantiate	Abstract Factory	0.3	0.962758
8	Relate	Abstract Factory	0.1	1.461728
9	Specific	Abstract Factory	0.1	1.160698
10	Kit	Abstract Factory	0.2	1.561728
11	Make	Abstract Factory	0.1	1.461728
12	Creational	Abstract Factory	0.3	0.962758
13	Encapsulate	Abstract Factory	0.1	0.984607
14	Group	Abstract Factory	0.1	1.461728
15	Builder	Builder	0.4	1.761728
16	Complex	Builder	0.1	1.461728
17	Construct	Builder	0.3	1.661728
18	Create	Builder	0.3	0.962758
19	Different	Builder	0.1	1.461728
20	Instantiate	Builder	0.3	0.962758
21	Representation	Builder	0.1	1.160698
22	Separate	Builder	0.1	1.461728
23	Creational	Builder	0.3	0.962758
24	Create	Factory Method	0.3	0.962758
25	Decide	Factory Method	0.1	1.461728
27	Factory	Factory Method	0.4	1.460698

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
28	Instantiate	Factory Method	0.3	0.962758
29	Method	Factory Method	0.4	1.460698
30	Visual	Factory Method	0.2	1.561728
31	Constructor	Factory Method	0.2	1.561728
32	Creational	Factory Method	0.3	0.962758
33	Access	Singleton	0.1	1.060698
34	Create	Singleton	0.3	0.962758
35	Ensure	Singleton	0.1	1.361728
36	Global	Singleton	0.1	1.361728
37	Instantiate	Singleton	0.3	0.962758
38	One	Singleton	0.1	1.060698
39	Singleton	Singleton	0.4	1.761728
40	Creational	Singleton	0.3	0.962758
41	Single	Singleton	0.1	1.561728
42	Unify	Singleton	0.1	1.060698
43	Copy	Prototype	0.1	1.461728
44	Create	Prototype	0.3	0.962758
45	Instantiate	Prototype	0.3	0.962758
46	Prototype	Prototype	0.4	1.761728
47	Prototypical	Prototype	0.2	1.561728
48	Specific	Prototype	0.1	1.160698
49	Creational	Prototype	0.3	0.962758
50	Clone	Prototype	0.2	1.561728
51	Cloneable	Prototype	0.2	1.561728
52	Adapter	Adapter	0.4	1.761728
53	Compose	Adapter	0.2	0.716629
54	Convert	Adapter	0.2	1.561728

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
55	Incompatible	Adapter	0.1	1.561728
56	Structural	Adapter	0.3	1.184607
57	Wrapper	Adapter	0.2	1.260698
58	Middleware	Adapter	0.2	1.561728
59	Match	Adapter	0.1	1.461728
60	Adapt	Adapter	0.1	1.461728
61	Wrapper	Adapter	0.2	1.561728
62	Bridge	Bridge	0.4	1.761728
63	Decouple	Bridge	0.1	1.461728
64	Independent	Bridge	0.1	0.984607
65	Structural	Bridge	0.3	1.184607
66	Handle	Bridge	0.2	1.561728
67	Compose	Bridge	0.2	0.71129
68	Body	Bridge	0.2	1.561728
69	Composition	Composite	0.3	1.661728
70	Uniform	Composite	0.2	1.561728
71	Tree	Composite	0.2	1.561728
72	Structure	Composite	0.3	1.661728
73	Hierarchy	Composite	0.2	1.561728
74	Composite	Composite	0.4	1.761728
75	Compose	Composite	0.2	0.716629
76	Structural	Composite	0.3	1.184607
77	Whole	Composite	0.1	1.461728
78	Part	Composite	0.1	1.461728
79	Addition	Decorator	0.1	1.461728
80	Alternative	Decorator	0.1	1.461728
81	Attach	Decorator	0.1	1.461728

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
82	Compose	Decorator	0.2	0.716629
83	Wrapper	Decorator	0.2	1.260698
84	Decorator	Decorator	0.4	1.761728
85	Dynamically	Decorator	0.1	1.461728
86	Extend	Decorator	0.1	1.461728
87	Flexible	Decorator	0.1	1.461728
88	Functional	Decorator	0.1	1.461728
89	Structural	Decorator	0.3	1.184607
90	Responsibility	Decorator	0.1	1.16070
91	Simplify	Facade	0.1	1.160698
92	Facade	Facade	0.4	1.761728
93	Compose	Facade	0.2	0.716629
94	Higher	Facade	0.2	1.561728
95	Level	Facade	0.1	1.461728
96	Subsystem	Facade	0.1	1.461728
97	Unify	Facade	0.2	1.260698
98	Structural	Facade	0.3	1.184607
99	Single	Facade	0.1	1.561730
100	High	Facade	0.2	1.561728
101	Access	Iterator	0.1	1.160698
102	Aggregate	Iterator	0.1	1.461728
103	Element	Iterator	0.1	1.461728
104	Expose	Iterator	0.1	1.461728
105	Iterator	Iterator	0.4	1.761728
106	Representation	Iterator	0.1	1.160698
107	Sequence	Iterator	0.1	1.461728
108	Underly	Iterator	0.1	1.461728

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
109	Behavioral	Iterator	0.3	0.620335
110	Behavior	Iterator	0.2	0.520335
111	Traverse	Iterator	0.2	1.561728
112	Cursor	Iterator	0.2	1.561728
113	Behavioral	Iterator	0.3	0.620335
114	Couple	Mediator	0.1	1.461728
115	Encapsulate	Mediator	0.1	0.984607
116	Explicitly	Mediator	0.1	1.461728
117	Independent	Mediator	0.1	0.984607
118	Interact	Mediator	0.1	1.461728
119	Interaction	Mediator	0.1	1.461728
120	Loose	Mediator	0.1	1.461728
121	Mediator	Mediator	0.4	1.761728
122	Promote	Mediator	0.1	1.461728
123	Behavioral	Mediator	0.3	0.620335
124	Behavior	Mediator	0.2	0.520335
125	Capture	Memento	0.1	1.461728
126	Encapsulation	Memento	0.1	1.461728
127	Token	Memento	0.2	1.561728
128	Externalize	Memento	0.1	1.461728
129	Internal	Memento	0.1	1.160698
130	Memento	Memento	0.4	1.761728
131	Restore	Memento	0.1	1.461728
132	State	Memento	0.1	0.984607
133	Violating	Memento	0.1	1.461728
134	Without	Memento	0.1	0.984607
135	Behavioral	Memento	0.3	0.620335

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
136	Change	Observer	0.1	0.984607
137	Dependency	Observer	0.1	1.461728
138	Dependent	Observer	0.1	1.160698
139	Notify	Observer	0.1	1.461728
140	Observer	Observer	0.4	1.761728
141	One	Observer	0.1	1.160698
142	Many	Observer	0.1	1.160698
143	State	Observer	0.1	0.984607
144	Update	Observer	0.1	1.461728
145	Observation	Observer	0.2	1.561728
146	Behavioral	Observer	0.3	0.620335
147	Behavior	Observer	0.2	1.561728
148	Publish	Observer	0.2	1.561728
149	Subscribe	Observer	0.2	1.561728
150	Notify	Observer	0.1	1.361728
151	Behavior	State	0.2	0.783577
152	Change	State	0.1	0.984606
153	Internal	State	0.1	1.160698
154	State	State	0.4	1.284607
155	Behavioral	State	0.3	0.620335
156	Algorithm	State	0.2	1.561728
157	Encapsulate	State	0.1	0.984607
158	Family	Strategy	0.1	1.160698
159	Independent	Strategy	0.1	0.984607
160	Interchangeable	Strategy	0.1	1.461728
161	Strategy	Strategy	0.4	1.761728
162	Behavioral	Strategy	0.3	0.620335

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
163	Behavior	Strategy	0.2	0.520335
164	Policy	Strategy	0.2	1.561728
165	Algorithm	Strategy	0.1	1.561727
166	Flyweight	Flyweight	0.4	1.761727
167	Structure	Flyweight	0.2	1.184606
168	Share	Flyweight	0.1	1.461728
169	Fine	Flyweight	0.1	1.461728
170	Gain	Flyweight	0.1	1.461728
171	Compose	Flyweight	0.2	0.716629
172	Structure	Proxy	0.2	1.184606
173	Structural	Proxy	0.3	1.284606
174	Proxy	Proxy	0.4	1.761727
175	Control	Proxy	0.1	1.160698
176	Compose	Proxy	0.2	0.716629
177	Access	Proxy	0.1	1.160698
178	Represent	Proxy	0.1	1.160698
179	Surrogate	Proxy	0.2	1.561728
180	Placeholder	Proxy	0.2	1.561728
181	Behavioral	Chain of Responsibility	0.3	0.620335
182	Chain	Chain of Responsibility	0.4	1.761727
183	Responsibility	Chain of Responsibility	0.4	1.360707
184	Behavior	Chain of Responsibility	0.2	1.461728
185	Command	Command	0.4	1.761727
186	Behavioral	Command	0.3	0.620335
187	Parameterize	Command	0.1	1.461728
188	Undoable	Command	0.1	1.461728
189	Action	Command	0.2	1.561728

ตาราง ข.1 (ต่อ)

ลำดับที่	ดัชนีเดี่ยว	เอกสาร	ค่าคงที่แสดงลำดับความสามารถในการบ่งชี้	ค่าถ่วงน้ำหนักดัชนี
190	Transaction	Command	0.2	1.561728
191	Behavior	Command	0.2	0.520335
192	Interpreter	Interpreter	0.4	1.761727
193	Behavioral	Interpreter	0.3	0.620335
194	Behavior	Interpreter	0.2	0.520335
195	Grammar	Interpreter	0.1	1.461728
196	Language	Interpreter	0.1	1.461728
197	Interpret	Interpreter	0.2	1.561728
198	Behavioral	Template Method	0.3	0.620335
199	Behavior	Template Method	0.2	0.520335
200	Method	Template Method	0.4	1.260698
201	Template	Template Method	0.4	1.761727
202	Behavioral	Template Method	0.3	0.620335
203	Skeleton	Template Method	0.2	1.561728
204	Redefine	Template Method	0.1	1.461728
205	Step	Template Method	0.1	1.461728
206	Visitor	Visitor	0.4	1.761727
207	Behavioral	Visitor	0.3	0.620335
208	Behavior	Visitor	0.2	0.520335
209	New	Visitor	0.1	1.461728
210	Operation	Visitor	0.1	1.461728

ตาราง ข.2 แสดงดัชนี และค่าถ่วงน้ำหนักของดัชนีวลี 2 คำ

ลำดับที่	ดัชนีวลี 2 คำ	เอกสาร	ค่าถ่วงน้ำหนักดัชนีวลี 2 คำ
1	Relate object	Abstract Factory	2.823455
2	Dependent object	Abstract Factory	2.522425
3	Group object	Abstract Factory	2.823455
4	Family object	Abstract Factory	2.522425
5	Create object	Abstract Factory	2.324485
6	Factory Method	Factory Method	3.121396
7	Visual Constructor	Factory Method	3.123456
8	Create object	Factory Method	2.324485
9	Convert interface	Adapter	2.923507
10	Incompatible interface	Adapter	2.923507
11	Adapt interface	Adapter	2.823455
12	Match interface	Adapter	2.823455
13	Compose object	Composite	2.078356
14	Tree structure	Composite	3.223456
15	Compose structure	Composite	3.223456
16	Tree object	Composite	2.923455
17	Whole part	Composite	2.923455
18	Hierarchy object	Composite	2.923455
19	Unify interface	Facade	2.622425
20	Single Class	Facade	2.923455
21	Higher level	Facade	3.023456
22	Share object	Flyweight	2.823455
23	Fine gain	Flyweight	2.923456
24	Interpret language	Interpreter	3.223455
25	Grammar language	Interpreter	2.923456
26	Interpret grammar	Interpreter	3.223455
27	Traverse object	Iterator	2.923455
28	Access object	Iterator	2.522425
29	Notify object	Observer	2.823455

ตาราง ข.2 (ต่อ)

ลำดับที่	ดัชนีวลี 2 คำ	เอกสาร	ค่าถ่วงน้ำหนักดัชนีวลี 2 คำ
30	Update object	Observer	2.823455
31	Publish subscribe	Observer	3.123456
32	Prototype object	Prototype	3.123507
33	Clone object	Prototype	2.923455
34	Copy object	Prototype	2.823455
35	Create object	Prototype	2.324485
36	Control access	Proxy	2.321396
37	Create object	Singleton	2.324485
38	Single object	Singleton	2.923455
39	One object	Singleton	2.422425
40	Global access	Singleton	2.522426
41	One instance	Singleton	2.422425
42	Single instance	Singleton	2.923455
43	Construct object	Builder	3.023455
44	State object	State	2.646334
45	State change	State	2.269213
46	Internal state	State	2.445305
47	Template Method	Template Method	3.121396
48	Step subclass	Template Method	2.823455
49	Subclass redefine	Template Method	2.823455
50	New operation	Visitor	1.461728

ตาราง ข.3 แสดงดัชนี และค่าถ่วงน้ำหนักของดัชนีวลี 3 คำ

ลำดับที่	ดัชนีวลี 3 คำ	เอกสาร	ค่าถ่วงน้ำหนักดัชนีวลี 3 คำ
1	Create family object	Abstract Factory	3.485183
2	Create relate object	Abstract Factory	3.786213
3	Create dependent object	Abstract Factory	3.485183
4	Create group object	Abstract Factory	3.786213
5	Interface create object	Abstract Factory	3.686212
6	Compose Incompatible interface	Adapter	3.640136
7	Convert incompatible interface	Adapter	4.485235
8	Separate construct object	Builder	4.485183
9	Compose tree structure	Composite	3.940085
10	Compose tree object	Composite	3.640084
11	Whole part hierarchy	Composite	4.485183
12	Compose hierarchy object	Composite	3.640084
13	Compose whole part	Composite	3.640084
14	Tree structure hierarchy	Composite	4.785184
15	Interface compose object	Facade	3.440083
16	Simplify compose object	Facade	3.239054
17	Interpret grammar language	Interpreter	4.985182
18	Sequence access object	Iterator	3.984153
19	Notify update object	Observer	4.285183
20	Create copy object	Prototype	3.786213
21	Create clone object	Prototype	3.886213
22	Create prototype object	Prototype	4.086265
23	Control access object	Proxy	3.683123
24	Create one object	Singleton	3.385183
25	Create single instance	Singleton	3.886213
26	Create single object	Singleton	3.886213
27	Unify access instance	Singleton	3.483123
28	Unify access object	Singleton	3.483123

ตาราง ข.3 (ต่อ)

ลำดับที่	ดัชนีวลี 3 คำ	เอกสาร	ค่าถ่วงน้ำหนักดัชนีวลี 3 คำ
29	instantiate one object	Singleton	3.385183
30	Change internal state	State	3.429911

ประวัติผู้เขียน

ชื่อ-สกุล	นางสาววีณาวดี ม่วงอ้น
วัน เดือน ปี เกิด	20 เมษายน 2523
ที่อยู่	100/2 หมู่ 4 ตำบลบางขุนกอง อำเภอบางกรวย จังหวัดนนทบุรี 11130
ประวัติการศึกษา	2544 วิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์ประยุกต์ คณะวิทยาศาสตร์ เทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ประวัติการทำงาน

มิถุนายน-กันยายน 2548 อาจารย์พิเศษ มหาวิทยาลัยกรุงเทพ วิทยาเขตรังสิต

ทุนการศึกษา

2546-2547	ทุนพัฒนาอาจารย์สาขาขาดแคลน สาขาวิทยาการคอมพิวเตอร์ สังกัดมหาวิทยาลัยศิลปากร
พฤศจิกายน 2548- มีนาคม 2549	ทุนฝึกอบรมด้านเทคโนโลยีสารสนเทศโครงการความร่วมมือ ไทย-อินเดีย ด้านเทคโนโลยีสารสนเทศ โดยกระทรวงวิทยาศาสตร์ และเทคโนโลยี
2549	ทุนสนับสนุนการทำวิทยานิพนธ์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพมหานคร