การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

A METADATA SEARCH APPROACH TO KEYWORD SEARCHING IN DATABASES

จารุณี แซ่หลี
JARUNEE ZALEE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2549

ISBN 974-15-2304-1

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

A METADATA SEARCH APPROACH TO KEYWORD SEARCHING IN DATABASES

จารุณี แซ่หลี

JARUNEE ZALEE

เลขหมู่.....
เลขทะเบียน.....61687
วัน,เดือน,ปี.....19 ก.ค. 2549

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2549

ISBN 974-15-2304-1

A METADATA SEARCH APPROACH TO KEYWORD SEARCHING IN DATABASES

JARUNEE ZALEE

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 974-15-2304-1

COPYRIGHT 2006

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อวิทยานิพนธ์	การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล
นักศึกษา	นางสาวจารุณี แซ่หลี
รหัสประจำตัว	45064614
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2549
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.วีระ บุญจริง

บทคัดย่อ

การค้นหาข้อมูลบนอินเทอร์เน็ตโดยใช้คำสำคัญเป็นที่นิยมอย่างแพร่หลาย ในขณะที่การค้นหาข้อมูลบนฐานข้อมูลผู้ใช้จำเป็นต้องมีความรู้ด้านโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม การนำเทคนิคการค้นหาคำสำคัญบนอินเทอร์เน็ตมาใช้กับฐานข้อมูลก็ไม่สามารถทำได้โดยตรง เนื่องจากข้อมูลที่เก็บบนอินเทอร์เน็ตและฐานข้อมูลอยู่ในรูปแบบที่ต่างกัน นอกจากนี้ระบบการค้นหาคำสำคัญในฐานข้อมูลส่วนใหญ่กำหนดให้คำสำคัญเป็นคำที่ตรงกับค่าข้อมูล งานวิจัยนี้จึงได้เสนอวิธีการค้นหาคำสำคัญในฐานข้อมูล ซึ่งผู้ใช้สามารถใช้คำสำคัญหรือกลุ่มคำสำคัญที่ตรงกับค่าข้อมูล ข้อมูลฐานข้อมูล และคำพ้องความหมายในการค้นหาได้ โดยอาศัยแบบจำลองข้อมูลที่ประกอบด้วยข้อมูล 3 ระดับ คือ โครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส เรียกว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล เมื่อทำการเปรียบเทียบผลลัพธ์ที่ดีที่สุดที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ พบว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลให้ผลลัพธ์ที่ถูกต้องแม่นยำกว่าระบบที่มีอยู่เดิม

Thesis Title	A Metadata Search Approach to Keyword Searching in Databases
Student	Ms. Jarunee Zalee
Student ID.	45064614
Degree	Master of Science
Programme	Computer Science
Year	2006
Thesis Advisor	Assoc.Prof.Dr. Veera Boonjing

ABSTRACT

Search engines on the Web have popularized the keyword-based search paradigm, while searching in databases users need to know a database schema and a query language. Keyword search techniques on the Web cannot directly be applied to databases because the data on the Internet and database are in different forms. Moreover, existing systems for keyword searching in databases limit type of keywords to database values. Thus, this research aims to propose a method for keyword searching in databases which allows users to search either with database values, metadata or synonyms. A metadata search approach is used to achieve this method. A data model consists of database schema, synonyms and cases. The most effective results from each approach show that the metadata search approach brings about the more accurate result than other approaches.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี ข้าพเจ้าขอกราบขอบพระคุณ รศ.ดร. วีระ บุญจริง อาจารย์ผู้ควบคุมวิทยานิพนธ์ ที่กรุณาให้คำปรึกษา คำแนะนำ แนวคิด ความเอาใจใส่ ตลอดจนการแก้ปัญหาต่างๆ อันเป็นประโยชน์ต่อวิทยานิพนธ์เป็นอย่างดี

ขอกราบขอบพระคุณ ผศ.ดร. ศรัณย์ อินทโกสุม และ ผศ.ดร. จีรพร ศรีสวัสดิ์ สำหรับคำแนะนำ และคำปรึกษาต่างๆ และยังให้ความกรุณามาเป็นตัวแทนกรรมการจากภาควิชา

ขอกราบขอบพระคุณ ดร. เฉลิมศักดิ์ เลิศวงศ์เสถียร ซึ่งให้ความกรุณามาเป็นตัวแทนกรรมการจากบุคคลภายนอก ทำให้ได้รับคำปรึกษาและคำแนะนำต่างๆ ในการสอบวิทยานิพนธ์

ขอกราบขอบพระคุณคณาจารย์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกๆ ท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ ให้กับข้าพเจ้าในระดับปริญญาโท

ขอกราบขอบพระคุณคณาจารย์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานีทุกๆ ท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ ในระดับปริญญาตรี ซึ่งช่วยสร้างฐานความรู้ที่มั่นคงให้กับข้าพเจ้า

ขอกราบขอบพระคุณ โครงการพัฒนาอาจารย์สาขาขาดแคลนฯ ที่ให้การสนับสนุนทุนการศึกษาในระดับปริญญาโทมาโดยตลอด

ขอกราบขอบพระคุณ ทุนสนับสนุนการทำวิทยานิพนธ์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ให้การสนับสนุนทุนการศึกษาในการทำวิทยานิพนธ์

ขอขอบคุณ สาว ออย และเพื่อนๆ พี่ๆ น้องๆ ในระดับปริญญาโททุกคน ที่ให้คำแนะนำและความช่วยเหลือต่างๆ ตลอดจนเป็นกำลังใจในการทำวิทยานิพนธ์เป็นอย่างดีเสมอมา

ขอขอบคุณ หยิ่ม เย้ง ฝง ที่คอยเป็นแรงผลักดันในการทำงาน และให้คำปรึกษาด้านการเงินเสมอมา และขอขอบคุณ เหน่ง โอ แบม หญิง นงนุช และเพื่อนๆ ทุกคน ที่คอยให้กำลังใจและเป็นอีกหนึ่งแรงผลักดันในการทำวิทยานิพนธ์เป็นอย่างดี

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณ บิดา มารดา ที่ให้การสนับสนุนด้านการเรียนเป็นอย่างดีมาโดยตลอด อีกทั้งยังเป็นกำลังใจอันยิ่งใหญ่ในการทำวิทยานิพนธ์ฉบับนี้ให้สำเร็จลุล่วงด้วยดี

สำหรับคุณค่าและประโยชน์อันพึงมีมาจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบแด่ผู้มีพระคุณทุกท่าน

จารุณี แซ่หลี

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป X

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 สมมติฐานของการศึกษา 2

1.4 ขอบเขตของการศึกษา 2

1.5 ส่วนประกอบของวิทยานิพนธ์ 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 การค้นหาคำสำคัญในฐานข้อมูล 4

2.1.1 สถาปัตยกรรมของระบบ 5

2.1.2 แบบจำลองข้อมูลและดัชนีผกผัน 5

2.1.3 ภาษาสอบถาม 6

2.1.4 ผลลัพธ์จากการสอบถาม 6

2.1.5 การจัดลำดับผลลัพธ์ 6

2.2 ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีตรงกับค่าข้อมูล 7

2.2.1 ดาตาสปอต 7

2.2.2 ดิสโคเวอร์ 9

2.2.3 ดีบีเอกซ์โพลเรอ 13

2.2.4 อีเคเอสโอ 15

2.3 ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล 17

สารบัญ(ต่อ)

หน้า

2.3.1 ดีบีเซอร์ฟเฟอร์ 17
2.3.2 บีเอเอ็นเคเอส 19
2.3.3 มราจียาติ 22
2.4 การวิเคราะห์ระบบการค้นหาคำสำคัญในฐานข้อมูล 26

บทที่ 3 การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล 28
3.1 แบบจำลองข้อมูล 28
3.1.1 โครงสร้างฐานข้อมูล 29
3.1.2 เคส 30
3.1.3 คำพ้องความหมาย 31
3.2 สถาปัตยกรรมของระบบ 31
3.2.1 หน่วยจัดการแบบจำลองข้อมูล 32
3.2.2 ตัวสร้างดัชนี 33
3.2.3 การวิเคราะห์คำจากการสอบถาม 34
3.2.4 กระบวนการหาผลลัพธ์ 35
3.2.5 การสร้างคำสั่งเอสคิวแอล 36
3.3 อัลกอริทึมที่ใช้ในการหาผลลัพธ์ 37
3.4 การจัดการเคส 39
3.4.1 กระบวนการในการสร้างเคส 39
3.4.2 กระบวนการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส 39
3.4.3 กระบวนการรียูสเคส 40

บทที่ 4 การประเมินผล 41
4.1 วิธีการประเมินผลของระบบ 41
4.2 ตัวอย่างฐานข้อมูลที่ใช้ในระบบ 42
4.3 ตัวอย่างการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล 44
4.4 ตัวอย่างการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส 60
4.5 สรุปการประเมินผล 64

สารบัญ(ต่อ)

หน้า

บทที่ 5 สรุปและข้อเสนอแนะ 73
5.1 สรุป 73
5.2 ข้อเสนอแนะ 74

เอกสารอ้างอิง 76

ภาคผนวก ก เมตาดาตาสำหรับฐานข้อมูลดีบีแอลพี 79
ภาคผนวก ข ตัวอย่างค่าข้อมูลจากฐานข้อมูลดีบีแอลพี 82
ภาคผนวก ค ตัวอย่างคำพ้องความหมายในฐานข้อมูลดีบีแอลพี 101
ภาคผนวก ง ตัวอย่างเคส 104

ประวัติผู้เขียน 106

สารบัญตาราง

ตารางที่ หน้า

2.1 แสดงวิธีการเก็บดัชนีผกผันในระบบดีบีเอ็กซ์โพลเรอ 13

2.2 แสดงตัวอย่างข้อมูลในเอนทิตี Member 23

2.3 แสดงตัวอย่างข้อมูลในเอนทิตี Activity 23

3.1 แสดงตัวอย่างดัชนีผกผัน 34

3.2 แสดงตัวอย่างการเชื่อมเงื่อนไขของเงื่อนไขการเลือก 37

4.1 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Get address of Jason Rennie" 44

4.2 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo" 47

4.3 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology" 49

4.4 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me references of The VOCAL Test Methodology" 51

4.5 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "What are papers cite to The VOCAL Test Methodology" 54

4.6 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "What were papers published in year 1999" 58

4.7 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Get address of Jason Rennie" 61

4.8 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Who write The VOCAL Test Methodology" 62

4.9 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Publication of David Zuckerman and Russell Impagliazzo" 63

4.10 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Get address of Jason Rennie" 65

4.11 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Get address of Jason Rennie" 65

4.12 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo" 66

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.13 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo" 66

4.14 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology" 67

4.15 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology" 67

4.16 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me references of The VOCAL Test Methodology" 68

4.17 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Give me refcrences of The VOCAL Test Methodology" 68

4.18 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "What are papers cite to The VOCAL Test Methodology" 69

4.19 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "What are papers cite to The VOCAL Test Methodology" 70

4.20 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "What were papers published in year 1999" 71

4.21 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "What were papers published in year 1999" 71

4.22 แสดงการเปรียบเทียบความสามารถในการค้นหาคำสำคัญตามประเภทของคำดัชนี 72

ก.1 แสดงข้อมูลจาก Entity 80

ก.2 แสดงข้อมูลจาก Attribute 80

ก.3 แสดงข้อมูลจาก Equivalent 81

ก.4 แสดงข้อมูลจาก Integrity 81

ข.1 แสดงตัวอย่างข้อมูลจากเอนทิตี Publication 83

ข.2 แสดงตัวอย่างข้อมูลจากเอนทิตี Author 95

ข.3 แสดงตัวอย่างข้อมูลจากเอนทิตี Writes 97

ข.4 แสดงตัวอย่างข้อมูลจากเอนทิตี Citation 100

ค.1 แสดงคำพ้องความหมายของฐานข้อมูลดีบีแอลพี 102

ง.1 แสดงตัวอย่างเคส 105

สารบัญรูป

รูปที่ หน้า

2.1 สถาปัตยกรรมของดาตาสปอต 7
2.2 ตัวอย่างฐานข้อมูลและกราฟไฮเปอร์เบส 8
2.3 สถาปัตยกรรมของดิสโคเวอร์ 10
2.4 ตัวอย่างข้อมูลจากฐานข้อมูล TPC-H ที่ใช้ในดิสโคเวอร์ 12
2.5 จอยทรี 14
2.6 สถาปัตยกรรมของระบบอีเคเอสโอ 15
2.7 ตัวอย่างข้อมูลจากฐานข้อมูล TPC-H ที่ใช้ในอีเคเอสโอ 16
2.8 ตัวอย่างเอกสารเสมือนในรูปของภาษาเอกซ์เอ็มแอลที่ได้จากฐานข้อมูลดีบีแอลพี 18
2.9 ตัวอย่างทรีที่เป็นผลลัพธ์จากการการสอบถามด้วย “soumen byron” 20
2.10 ตัวอย่างโครงสร้างฐานข้อมูลของบีเอเอ็นเคเอส 21
2.11 สถาปัตยกรรมของของระบบมราจียาติ 22
2.12 ทรีที่ได้จากการสอบถามด้วย “John” 24
2.13 ทรีที่ได้จากการสอบถามด้วย “John sport” 24
2.14 ทรีที่ได้จากการสอบถามด้วย “John activity” 25
3.1 แบบจำลองข้อมูล 28
3.2 ตัวอย่างโครงสร้างฐานข้อมูล 29
3.3 ตัวอย่างกราฟโครงสร้างฐานข้อมูล 29
3.4 แผนภาพแสดงความสัมพันธ์ระหว่างข้อมูลฐานข้อมูล 30
3.5 สถาปัตยกรรมของระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล 32
3.6 หน่วยจัดการแบบจำลองข้อมูล 33
4.1 โครงสร้างฐานข้อมูลดีบีแอลพี 43
4.2 กราฟโครงสร้างฐานข้อมูลดีบีแอลพีในแบบจำลองข้อมูล 43
4.3 เส้นทางพื้นฐานของ QI_1 = {ADDRESS| (A, A003), Jason Rennie | (V, A002)} 45
4.4 เส้นทางพื้นฐานของ QI_2 = {ADDRESS| (A, P008), Jason Rennie | (V, A002)} 45
4.5 กราฟที่เป็นไปได้ FG_1 ของ QI_1 45
4.6 กราฟที่เป็นไปได้ FG_2 ของ QI_2 46
4.7 เส้นทางพื้นฐานของ QI_1 = {David Zuckerman| (V, A002), Russell Impagliazzo| (V, A002)} 48

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.8 กราฟที่เป็นไปได้ FG_1 ของ QI_1 48

4.9 เส้นทางพื้นฐานของ QI_1 = {AUTHOR| (E, AUTHOR), NAME| (A, A002), WRITES| (E, WRITES), The VOCAL Test Methodology| (V, P004)} 50

4.10 กราฟที่เป็นไปได้ FG_1 ของ QI_1 50

4.11 เส้นทางพื้นฐานของ QI_1 = {CITATION| (E, CITATION), The VOCAL Test Methodology| (V, P004)} 52

4.12 กราฟที่เป็นไปได้ FG_{11} ของ QI_1 52

4.13 กราฟที่เป็นไปได้ FG_{12} ของ QI_1 53

4.14 เส้นทางพื้นฐานของ QI_1 = {PUBLICATION| (E, PUBLICATION), CITING| (A, C001), The VOCAL Test Methodology| (V, P004)} 55

4.15 เส้นทางพื้นฐานของ QI_2 = {PUBLICATION| (E, PUBLICATION), CITED| (A, C002), The VOCAL Test Methodology| (V, P004)} 55

4.16 กราฟที่เป็นไปได้ FG_{11} และ FG_{12} ของ QI_1 56

4.17 กราฟที่เป็นไปได้ FG_{21} และ FG_{22} ของ QI_2 56

4.18 เส้นทางพื้นฐานของ QI_1 = {PUBLICATION| (E, PUBLICATION), YEAR| (A, P007), 1999| (V, P007)} 59

4.19 กราฟที่เป็นไปได้ FG_1 ของ QI_1 59

4.20 กราฟคำตอบที่ได้จากการรีดูสเคส C_3 63

4.21 กราฟคำตอบที่ได้จากการรีดูสเคส C_2 64

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เทคนิคการค้นหาคำสำคัญเป็นที่นิยมอย่างแพร่หลายในการสืบค้นข้อมูลจากอินเทอร์เน็ต ผู้ใช้สามารถค้นหาข้อมูลจากเอกสารบนอินเทอร์เน็ตได้โดยระบุคำสำคัญหรือกลุ่มคำสำคัญที่เกี่ยวข้องกับข้อมูลที่ต้องการค้นหา ในขณะที่การค้นข้อมูลจากฐานข้อมูลผู้ใช้อาจค้นหาได้โดยการใช้ภาษาสอบถามเชิงโครงสร้างหรือภาษาเอสคิวแอล (Structured Query Language หรือ SQL) ซึ่งผู้ใช้จำเป็นต้องมีความรู้เกี่ยวกับโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม หรืออาจค้นหาได้จากแบบฟอร์มที่ผู้ให้บริการระบบฐานข้อมูลนั้นจัดเตรียมไว้ (Form-Based Interface) [4] การนำเทคนิคการค้นหาคำสำคัญมาใช้กับฐานข้อมูลจึงน่าจะช่วยให้ผู้ใช้ค้นหาข้อมูลได้สะดวกขึ้น แต่เนื่องจากข้อมูลบนอินเทอร์เน็ตและข้อมูลบนฐานข้อมูลเก็บอยู่ในรูปแบบที่ต่างกัน กล่าวคือ บนอินเทอร์เน็ตข้อมูลอยู่ในรูปของเอกสาร ในขณะที่ฐานข้อมูลเก็บข้อมูลแบบมีโครงสร้าง คือ อยู่ในรูปของหลายๆ เอนทิตี (Entity) และมีคีย์เชื่อมโยงความสัมพันธ์ระหว่างข้อมูล (Primary Key-Foreign Key) จึงไม่สามารถนำเทคนิคการค้นหาคำสำคัญบนอินเทอร์เน็ตมาใช้กับการค้นหาข้อมูลจากฐานข้อมูลได้โดยตรง

การค้นหาคำสำคัญในฐานข้อมูล เป็นวิธีหนึ่งที่ผู้ใช้สามารถค้นหาข้อมูลจากฐานข้อมูลได้โดยที่ผู้ใช้ไม่จำเป็นต้องมีความรู้ด้านโครงสร้างฐานข้อมูลและภาษาเอสคิวแอล ผู้ใช้สามารถทำการค้นหาได้โดยระบุคำสำคัญที่เกี่ยวกับข้อมูลที่ต้องการค้นหา คำสำคัญดังกล่าวอาจปรากฏในค่าข้อมูลหรือข้อมูลฐานข้อมูล (ชื่อเอนทิตี หรือชื่อแอทริบิว) ผลลัพธ์ที่ได้จากการค้นหา คือ ระเบียนข้อมูลที่ประกอบด้วยคำสำคัญ และ/หรือชื่อเอนทิตีหรือชื่อแอทริบิวตรงกับคำสำคัญตามที่ผู้ใช้ต้องการ โดยระเบียนเหล่านั้นมีความสัมพันธ์กันทางโครงสร้างฐานข้อมูล

ปัจจุบันมีงานวิจัยที่เกี่ยวกับการค้นหาคำสำคัญในฐานข้อมูล โดยมีการสร้างแบบจำลองที่ใช้ในการจัดเก็บโครงสร้างฐานข้อมูลรวมทั้งค่าข้อมูลต่างๆ ที่เปรียบได้กับเอกสารบนอินเทอร์เน็ต และนำดัชนีผกผัน (Inverted Index) มาใช้ในการระบุตำแหน่งของคำสำคัญในแบบจำลอง แล้วนำตำแหน่งที่ได้มาค้นหาผลลัพธ์ตามอัลกอริทึมที่แตกต่างกันออกไป ซึ่งสามารถแบ่งระบบการค้นหาคำสำคัญในฐานข้อมูลตามประเภทของคำดัชนีได้ 2 ประเภท คือ ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูล [3][5][7][11] และระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล [1][10][13] ซึ่งระบบที่ใช้

คำดัชนีตรงกับค่าข้อมูลไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ ในขณะที่ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูลบางระบบไม่สามารถค้นหาด้วยชื่อเอนทิตีได้ [13] และบางระบบไม่สามารถค้นหาคำสำคัญที่มาจากเอนทิตีมากกว่า 2 เอนทิตีได้ [10] นอกจากนี้ ระบบส่วนใหญ่ไม่อนุญาตให้ผู้ใช้ค้นหาคำสำคัญด้วยคำพ้องความหมายของค่าข้อมูล หรือข้อมูลฐานข้อมูลได้

งานวิจัยนี้จึงได้เสนอวิธีการค้นหาคำสำคัญในฐานข้อมูล ซึ่งผู้ใช้สามารถใช้คำสำคัญหรือกลุ่มคำสำคัญที่ตรงกับค่าข้อมูล ข้อมูลฐานข้อมูล (ชื่อเอนทิตี และชื่อแอทริบิว) และคำพ้องความหมายในการค้นหา คำสำคัญเหล่านี้จะถูกแปลงให้อยู่ในรูปของภาษาเอสคิวแอลโดยใช้แบบจำลองข้อมูลที่ประกอบด้วยข้อมูล 3 ประเภท คือ โครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส เรียกว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

งานวิจัยนี้เป็นการพัฒนาวิธีการค้นหาคำสำคัญในฐานข้อมูล โดยใช้การค้นข้อมูลฐานข้อมูลซึ่งประกอบด้วยข้อมูล 3 ประเภท คือ โครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส ทำให้ผู้ใช้สามารถค้นหาข้อมูลจากฐานข้อมูลได้โดยใช้คำสำคัญที่ตรงกับค่าข้อมูล ข้อมูลฐานข้อมูล (ชื่อเอนทิตี และชื่อแอทริบิว) และคำพ้องความหมาย

1.3 สมมติฐานของการศึกษา

การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลให้ผลลัพธ์ที่ถูกต้องแม่นยำกว่าการค้นหาคำสำคัญที่ตรงกับค่าข้อมูลเพียงอย่างเดียว

1.4 ขอบเขตของการศึกษา

ออกแบบแบบจำลองข้อมูลและดัชนีผกผัน ซึ่งแบบจำลองข้อมูลใช้ในการเก็บโครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส รวมทั้งพัฒนาวิธีการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

1.5 ส่วนประกอบของวิทยานิพนธ์

ส่วนที่เหลือของวิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 4 บท ดังต่อไปนี้

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง ซึ่งแบ่งเนื้อหาออกเป็น 3 ส่วน คือ ส่วนที่ 1 กล่าวถึงข้อมูลพื้นฐานของการค้นหาคำสำคัญในฐานข้อมูล ส่วนที่ 2 กล่าวถึงวิธีการค้นหาคำสำคัญจากฐานข้อมูลในวิธีต่างๆ ซึ่งแบ่งเป็นระบบที่ใช้คำสำคัญที่ตรงกับค่าข้อมูล และระบบที่ใช้คำสำคัญที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล และสุดท้าย ส่วนที่ 3 เป็นการสรุปผลการวิเคราะห์วิธีการค้นหาคำสำคัญในฐานข้อมูล

บทที่ 3 กล่าวถึงการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล ซึ่งแบ่งเนื้อหาเป็น 4 ส่วนหลักๆ คือ ส่วนแรกกล่าวถึงแบบจำลองข้อมูล ส่วนที่ 2 กล่าวถึงสถาปัตยกรรมของระบบ ส่วนที่ 3 กล่าวถึงอัลกอริทึมที่ใช้ในการค้นหาผลลัพธ์ และส่วนที่ 4 กล่าวถึง การจัดการเคส

บทที่ 4 เป็นการประเมินผล ซึ่งแบ่งเนื้อหาเป็น 5 ส่วน คือ ส่วนแรก กล่าวถึงวิธีการประเมินผลของระบบ ส่วนที่ 2 กล่าวถึงตัวอย่างฐานข้อมูลที่ใช้ในการประเมินผล ส่วนที่ 3 กล่าวถึงตัวอย่างการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล ส่วนที่ 4 กล่าวถึงตัวอย่างการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส และส่วนที่ 5 กล่าวถึงการสรุปผลการประเมินผล

บทที่ 5 เป็นบทสรุปผลการวิจัยและข้อเสนอแนะ

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้เป็นการนำเสนอวิธีการค้นหาคำสำคัญในฐานข้อมูลวิธีต่างๆ โดยที่ผู้ใช้ไม่จำเป็นต้องมีความรู้เกี่ยวกับโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม ซึ่งในส่วนที่ 2.1 กล่าวถึงข้อมูลพื้นฐานของการค้นหาคำสำคัญในฐานข้อมูล ส่วนที่ 2.2 กล่าวถึง ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลระบบต่างๆ ส่วนที่ 2.3 กล่าวถึง ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล และสุดท้าย ส่วนที่ 2.4 เป็นการวิเคราะห์วิธีการค้นหาคำสำคัญในฐานข้อมูลจากระบบต่างๆ

2.1 การค้นหาคำสำคัญในฐานข้อมูล

ระบบการค้นหาคำสำคัญในฐานข้อมูล เป็นระบบที่ช่วยให้ผู้ใช้สามารถค้นหาข้อมูลจากฐานข้อมูลได้สะดวกยิ่งขึ้น โดยที่ผู้ใช้ไม่จำเป็นต้องมีความรู้เกี่ยวกับโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม เช่น ภาษาเอสคิวแอล ซึ่งคล้ายกับการค้นหาข้อมูลบนอินเทอร์เน็ต

การค้นหาคำสำคัญ (Keyword Searching) คือ การที่ผู้ใช้ระบุคำที่มีนัยสำคัญ (Significant Word) หรือคำที่สามารถค้นหาได้จากดัชนี เพื่อค้นหาข้อมูลที่ต้องการ โดยคำดัชนีต้องเป็นคำที่มีนัยสำคัญที่ปรากฏอยู่ในชื่อเรื่อง เนื้อเรื่อง หรือบทความ [8]

ในที่นี้กำหนดให้ คำสำคัญ คือ คำที่ผู้ใช้ระบุในการสอบถามและมีนัยสำคัญ หรือคำที่ปรากฏในดัชนีผกผัน ส่วนคำดัชนี คือ คำหรือดัชนี ในแฟ้มดัชนีผกผัน

การค้นหาคำสำคัญในฐานข้อมูล คือ การค้นหาระเบียนข้อมูลทุกระเบียนจากฐานข้อมูลที่ประกอบด้วยคำสำคัญ และ/หรือระเบียนข้อมูลที่มีชื่อเอนทิตีหรือชื่อแอทริบิวตรงกับคำสำคัญตามที่ผู้ใช้ต้องการ โดยที่ระเบียนเหล่านั้นต้องมีความสัมพันธ์กันตามโครงสร้างฐานข้อมูล คำสำคัญในที่นี้เป็นคำที่ปรากฏอยู่ในข้อมูล 2 ประเภท คือ ค่าข้อมูลที่ได้จากฐานข้อมูล และข้อมูลฐานข้อมูล เช่น ชื่อเอนทิตี และชื่อแอทริบิว เป็นต้น หากคำสำคัญปรากฏในค่าข้อมูล คำสำคัญนั้นจะใช้ในการระบุระเบียนข้อมูล และหากคำสำคัญปรากฏในข้อมูลฐานข้อมูล คำสำคัญนั้นจะใช้ในการระบุเอนทิตีหรือแอทริบิว

ระบบการค้นหาคำสำคัญในฐานข้อมูลในปัจจุบันสามารถแบ่งตามประเภทของคำดัชนีได้ 2 ประเภท คือ ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูล และระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล

1. ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูล ประกอบด้วย ระบบดาตาสปอต (DTL's DataSpot) [5] ระบบดิสโคเวอร์ (Discover) [7] ระบบดีบีเอกซ์โพลเรอ (DBXplorer) [3] และระบบอีเคเอสโอ (EKSO) [11]

2. ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูลแบ่งเป็น 2 วิธี ได้แก่

2.1 คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อแอทริบิว ประกอบด้วย ระบบดีบีเซอร์ฟเฟอร์ (DBSurfer) [13]

2.2 คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี/ชื่อแอทริบิว ประกอบด้วย ระบบบีเอเอ็นเคเอส (BANKS) [1][2] และระบบมราจียาติ (Mragyati) [10]

หลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูลมีดังต่อไปนี้

2.1.1 สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบการค้นหาคำสำคัญในฐานข้อมูลแบ่งออกเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูลเริ่มต้น และส่วนการสอบถาม ในส่วนการเตรียมข้อมูลเริ่มต้น เป็นการเตรียมข้อมูลจากฐานข้อมูลเพื่อใช้ในขั้นตอนการค้นหา เรียกว่า แบบจำลองข้อมูล (Data Model) [12] ส่วนที่สอง คือ ส่วนการสอบถาม เป็นขั้นตอนการค้นหาผลลัพธ์ โดยเริ่มจากการส่งผ่านคิวรีไปค้นหาคำสำคัญ แล้วนำคำสำคัญที่ได้ไปพิจารณาหาผลลัพธ์ตามขั้นตอนของระบบ

2.1.2 แบบจำลองข้อมูลและดัชนีผกผัน

แบบจำลองข้อมูลส่วนใหญ่แบ่งเป็น 2 ประเภท คือ แบบจำลองในรูปของกราฟ [1][3][5][7] และแบบจำลองในรูปของเอกสารเสมือน (Virtual Document) [11] [13]

1. แบบจำลองในรูปของกราฟ แบ่งเป็น 2 ประเภท คือ กราฟโครงสร้างฐานข้อมูล และกราฟข้อมูล กราฟทั้งสองประเภทอาจเป็นกราฟแบบมีทิศทางหรือกราฟไม่มีทิศทาง ขึ้นอยู่กับขั้นตอนในการค้นหาผลลัพธ์ของผู้ออกแบบระบบ

1.1 กราฟโครงสร้างฐานข้อมูล กำหนดให้โหนดใช้แทนเอนทิตี และเส้นเชื่อมใช้แทนความสัมพันธ์ระหว่างเอนทิตี

1.2 กราฟข้อมูล กำหนดให้โหนดใช้แทนระเบียนข้อมูล และเส้นเชื่อมใช้แทนความสัมพันธ์ระหว่างระเบียน

แม้ว่ากราฟข้อมูลจะให้ประสิทธิภาพในการค้นหาดีกว่ากราฟโครงสร้างฐานข้อมูล แต่กราฟข้อมูลจำเป็นต้องใช้หน่วยความจำในการเก็บข้อมูลมากกว่า ในกรณีที่ฐานข้อมูลมีขนาดใหญ่กราฟข้อมูลต้องใช้หน่วยความจำมากยิ่งขึ้น ในขณะที่กราฟโครงสร้างฐานข้อมูลยังคงใช้หน่วยความจำเท่าเดิม

2. แบบจำลองในรูปของเอกสารเสมือน คือ แบบจำลองที่ได้จากการนำค่าข้อมูลจากฐานข้อมูลมาแปลงให้อยู่ในรูปของเอกสารเสมือนหรือเว็บเพจ โดยในแต่ละเอกสารเสมือนอาจประกอบด้วยข้อมูลจากระเบียนข้อมูลเดียวกัน หรือข้อมูลจากหลายระเบียนเชื่อมต่อกันตามความสัมพันธ์ของโครงสร้างฐานข้อมูล

ส่วนดัชนีผกผัน คือ คำดัชนีต่างๆ ที่มีรายการอ้างอิงไปยังตำแหน่งของคำดัชนีนั้นในแบบจำลองข้อมูล

2.1.3 ภาษาสอบถาม

ภาษาสอบถามเป็นการกำหนดรูปแบบของภาษาที่ใช้ในการสอบถาม ระบบการค้นหาคำสำคัญในฐานข้อมูลส่วนใหญ่กำหนดรูปแบบให้ผู้ใช้ระบุคำสำคัญเป็นการสอบถามแบบคำเดี่ยว (Single Word Query) การสอบถามแบบวลี (Phrase Query) การสอบถามแบบบูล (Boolean Query) [8] การสอบถามแบบพร็อกซิมิตี (Proximity Query) [6] หรือการสอบถามแบบกำหนดรูปแบบ (Pattern Matching Query)

2.1.4 ผลลัพธ์จากการสอบถาม

ผลลัพธ์ที่ได้จากการสอบถาม ประกอบด้วยระเบียนข้อมูลจากเอนทิตีใดเอนทิตีหนึ่ง หรือการจอยกันของระเบียนข้อมูลจากหลายๆ เอนทิตีที่มีความสัมพันธ์กันตามโครงสร้างฐานข้อมูล โดยที่ระเบียนข้อมูลเหล่านั้นอาจประกอบด้วยคำสำคัญที่ได้จากการสอบถามทั้งหมดหรือไม่ก็ได้

2.1.5 การจัดลำดับผลลัพธ์

เนื่องจากคำสำคัญที่ได้จากการสอบถามอาจตรงกับแบบจำลองข้อมูลในหลายตำแหน่ง ดังนั้นผลลัพธ์จึงอาจมีได้หลายผลลัพธ์ การจัดลำดับความสำคัญของผลลัพธ์เพื่อแสดงผลแก่ผู้ใช้จึงเป็นสิ่งจำเป็น วิธีการจัดลำดับความสำคัญของผลลัพธ์สามารถทำได้โดยการให้คะแนนกับผลลัพธ์ หลักเกณฑ์ในการให้คะแนนสามารถพิจารณาได้จากปัจจัย 3 ปัจจัย คือ

1. คะแนน IR (IR Score) ของแอทริบิว สามารถคำนวณได้จากจำนวนของคำสำคัญที่ปรากฏในแอทริบิวนั้น หรืออาจใช้วิธีการคำนวณหาน้ำหนักของระบบสืบค้นสารสนเทศ (Information Retrieval Weighting Method) อื่นๆ เช่น การให้ค่า tf-idf เป็นต้น

2. โครงสร้างของกราฟคำตอบ คะแนนของผลลัพธ์ขึ้นอยู่กับขนาดของกราฟคำตอบ ตัวอย่างเช่น จำนวนของโหนดหรือเส้นเชื่อมในกราฟคำตอบ

3. ความหมายของเส้นเชื่อม คะแนนของโหนดขึ้นอยู่กับเส้นเชื่อมระหว่างโหนดๆ หนึ่งกับโหนดใดๆ

การให้คะแนนโดยการพิจารณาจากปัจจัยดังกล่าวขึ้นอยู่กับนิยามของผลลัพธ์ในระบบนั้นๆ ถ้าผลลัพธ์ประกอบด้วยโหนดหลายๆ โหนด การให้คะแนนให้คะแนนตามโครงสร้างของกราฟคำตอบ แต่ถ้าผลลัพธ์มีเพียงโหนดเดียว โครงสร้างของกราฟคำตอบก็ไม่จำเป็นต้องนำมาพิจารณา

2.2 ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีตรงกับค่าข้อมูล

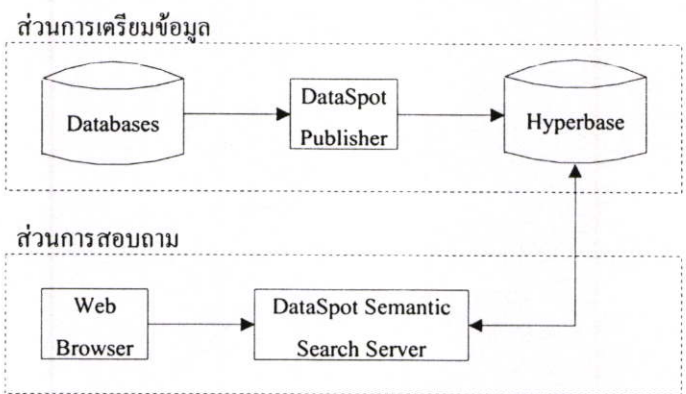
ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีตรงกับค่าข้อมูล ได้แก่ ดาตาสปอต ดิสโคเวอร์ ดีบีเอกซ์โพลเรอ และอีเคเอสโอ ซึ่งมีรายละเอียดดังต่อไปนี้

2.2.1 ดาตาสปอต

ดาตาสปอต [5] เป็นระบบการค้นหาข้อมูลจากฐานข้อมูลด้วยการระบุข้อความที่เรียบง่าย โดยอาศัยแบบจำลองข้อมูลที่เรียกว่า ไฮเปอร์เบส (Hyperbase) ผลลัพธ์ที่ได้ คือ กราฟย่อยของไฮเปอร์เบส ซึ่งคำสำคัญปรากฏในโหนดใบของกราฟย่อยนั้น

1. สถาปัตยกรรมของระบบ

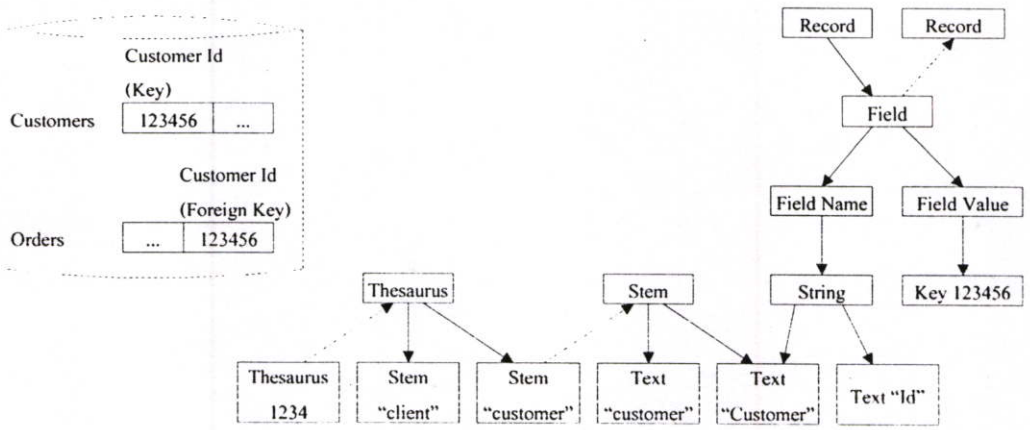
สถาปัตยกรรมของดาตาสปอตเป็นดังรูปที่ 2.1 ซึ่งแบ่งเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูล และส่วนการสอบถาม ในส่วนการเตรียมข้อมูล ระบบใช้แบบจำลองข้อมูลที่เรียกว่าไฮเปอร์เบส ในส่วนการสอบถาม ระบบทำการค้นหาคำสำคัญจากไฮเปอร์เบสโดยตรง



รูปที่ 2.1 สถาปัตยกรรมของดาตาสปอต

2. แบบจำลองข้อมูลและดัชนีผกผัน

แบบจำลองข้อมูลของดาตาสปอต คือ กราฟโครงสร้างฐานข้อมูลรวมกับกราฟข้อมูลแบบมีทิศทาง เรียกว่า ไฮเปอร์เบส ซึ่งประกอบด้วยโหนดซึ่งใช้แทนเอนทิตี แอทริบิว และค่าข้อมูล นอกจากนี้ยังประกอบด้วยโหนดซึ่งใช้แทนรากศัพท์และคำพ้องความหมายของโหนดค่าข้อมูลด้วย ในส่วนของดัชนีผกผัน ดาตาสปอตไม่จำเป็นต้องใช้ดัชนีผกผันในการค้นหาคำสำคัญ แต่สามารถค้นหาจากไฮเปอร์เบสได้โดยตรง ตัวอย่างฐานข้อมูลและไฮเปอร์เบสเป็นดังรูปที่ 2.2



รูปที่ 2.2 ตัวอย่างฐานข้อมูลและกราฟไฮเปอร์เบส

3. ขั้นตอนวิธีในการหาผลลัพธ์

ระบบทำการส่งผ่านคิวรีไปค้นหาคำสำคัญและตำแหน่งของโหนดใบที่คำสำคัญนั้นปรากฏอยู่จากไฮเปอร์เบส ผลลัพธ์ที่ได้คือ กราฟย่อยที่เชื่อมโยงโหนดดังกล่าวเข้าด้วยกัน และคำสำคัญทุกคำต้องปรากฏในโหนดของกราฟย่อยนั้น ผลลัพธ์ที่ได้เรียงลำดับความสำคัญตามขนาดของกราฟย่อย

4. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

4.1 สถาปัตยกรรมของระบบ ระบบแบ่งการทำงานเป็น 2 ส่วนหลัก คือ การเตรียมข้อมูลและการสอบถาม

4.2 แบบจำลองข้อมูลและดัชนีผกผัน ระบบใช้กราฟโครงสร้างฐานข้อมูลร่วมกับกราฟข้อมูลเป็นแบบจำลองข้อมูล รวมทั้งนำรากศัพท์ และคำพ้องความหมายมาเป็นส่วนหนึ่งของแบบจำลอง ทำให้ผู้ใช้สามารถค้นหาคำสำคัญจากรากศัพท์ และคำพ้องความหมายได้ แต่ไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ และเนื่องจากระบบทำการคัดลอกค่าข้อมูลไปยังไฮเปอร์เบส และไม่มีการนำดัชนีผกผันมาใช้ ระบบจึงทำการค้นหาคำสำคัญจากไฮเปอร์เบส

โดยตรง นั่นคือ การค้นหาตำแหน่งของคำสำคัญต้องค้นหาจากกราฟแบบจำลองข้อมูล ซึ่งระบบต้องทำงานกับกราฟขนาดใหญ่

4.3 ภาษาสอบถาม ระบบอนุญาตให้ผู้ใช้ใช้ภาษาข้อความอิสระแบบง่ายๆ หรือการสอบถามแบบวลีในการสอบถาม

4.4 ผลลัพธ์จากการสอบถาม คือ กราฟย่อยของไฮเปอร์เบส กราฟย่อยดังกล่าวประกอบด้วยโหนดต่างๆ โดยที่โหนดใบต้องมีคำสำคัญทั้งหมดปรากฏอยู่

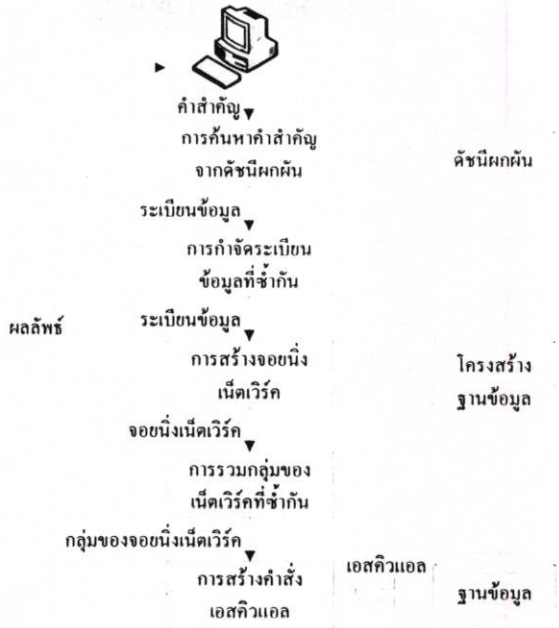
4.5 การจัดลำดับผลลัพธ์ ระบบจัดลำดับความสำคัญของผลลัพธ์ตามโครงสร้างของกราฟคำตอบ หรือตามขนาดของกราฟ ซึ่งกราฟที่มีขนาดเล็กที่สุดเป็นกราฟคำตอบที่มีนัยสำคัญมากที่สุด การจัดลำดับความสำคัญดังกล่าวเพิ่มความความสะดวกให้กับผู้ใช้ในการเลือกผลลัพธ์

2.2.2 ดิสโคเวอร์

ดิสโคเวอร์ [7] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูลเชิงสัมพันธ์ โดยใช้กราฟโครงสร้างฐานข้อมูลในการค้นหาผลลัพธ์ ผลลัพธ์ที่ได้จากการค้นหา คือ ระเบียนข้อมูลที่สัมพันธ์กัน ระเบียนข้อมูลเหล่านี้ประกอบด้วยคำสำคัญที่ได้จากการสอบถามทั้งหมดหรือไม่ก็ได้ ในกรณีที่ผลลัพธ์มีหลายผลลัพธ์ ระบบทำการจัดลำดับความสำคัญของผลลัพธ์ตามจำนวนการจอยกันของระเบียนข้อมูล และคะแนน IR ของแอทริบิว

1. สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบดิสโคเวอร์เน้นไปที่ส่วนการสอบถาม กล่าวคือ ระบบใช้กราฟโครงสร้างข้อมูลเป็นแบบจำลองข้อมูล และในส่วนการสอบถามประกอบด้วยขั้นตอนการทำงานหลัก 5 ขั้นตอน ดังรูปที่ 2.3 ซึ่งขั้นตอนการทำงานดังกล่าว ได้แก่ การค้นหาคำสำคัญจากดัชนี การกำจัดระเบียนข้อมูลที่ซ้ำกัน การสร้างจอยนิ่งเน็ตเวิร์ค การรวมกลุ่มของเน็ตเวิร์คที่ซ้ำกัน และการสร้างคำสั่งเอสคิวแอล



รูปที่ 2.3 สถาปัตยกรรมของดิสโคเวอร์

2. แบบจำลองข้อมูลและดัชนีผกผัน

ระบบใช้แบบจำลองในรูปของกราฟโครงสร้างฐานข้อมูล สมมติให้ฐานข้อมูลหนึ่งประกอบด้วยเอนทิตีจำนวน n เอนทิตี ($E_1, E_2, \ldots, E_n$) แต่ละเอนทิตี E_i มีจำนวนแอทริบิวเท่ากับ m_i แอทริบิว ($a^i_1, a^i_2, \ldots, a^i_{m_i}$) แบบจำลองข้อมูลของดิสโคเวอร์ถูกเก็บอยู่ในรูปของกราฟแบบมีทิศทาง โดยกำหนดให้โหนด E_i ใช้แทนเอนทิตี E_i และเส้นเชื่อม $E_i \rightarrow E_j$ ใช้แทนความสัมพันธ์ระหว่างเอนทิตีจากคีย์หลักไปยังคีย์นอก ซึ่งเชื่อมระหว่างเซตของแอทริบิว ($a^i_{b_1}, a^i_{b_2}, \ldots, a^i_{b_l}$) ของ E_i ไปยังเซตของแอทริบิว ($a^j_{b_1}, a^j_{b_2}, \ldots, a^j_{b_l}$) ของ E_j เมื่อ $a^i_{b_k} \equiv a^j_{b_k}$ โดยที่ k = 1, 2, …, l

ระบบใช้ดัชนีผกผันในการเก็บคำดัชนี แต่ละคำดัชนีมีการอ้างอิงไปยังแบบจำลองข้อมูล โดยกำหนดให้คำดัชนีคือค่าข้อมูลประเภทอักขระ

3. ขั้นตอนวิธีในการหาผลลัพธ์

ในส่วนของขั้นตอนวิธีในการหาผลลัพธ์ จากตัวอย่างข้อมูลในรูปที่ 2.4 เมื่อทำการค้นหาด้วยคำสำคัญ K = {Smith, Miller} ดิสโคเวอร์มีขั้นตอนการทำงานดังต่อไปนี้

ขั้นตอนที่ 1 จากดัชนีผกผัน เมื่อได้ตำแหน่งของคำสำคัญในแบบจำลองข้อมูลแล้วระบบจะทำการดึงระเบียนข้อมูลที่คำสำคัญนั้นปรากฏอยู่ จากตัวอย่างการค้นหา K ได้ระเบียนข้อมูลดังนี้

Smith ได้ระเบียน O_1

Miller ได้ระเบียน O_2 และ O_3

ขั้นตอนที่ 2 พิจารณาหาระเบียน R_i^K ที่เป็นไปได้ทั้งหมด เพื่อกำจัดระเบียนที่อาจซ้ำกันในขั้นตอนที่ 1

ขั้นตอนที่ 3 นำกลุ่มของระเบียนที่ได้มาจอย (Join) กันตามเงื่อนไขที่ได้จากแบบจำลองข้อมูล เรียกว่า จอยนิ่งเน็ตเวิร์ค (Joining Network) จากการสอบถาม K ได้จอยนิ่งเน็ตเวิร์ค J_1 และ J_2 ดังนี้

$$J_1 = \text{ORDERS}^{\text{Smith}} \bowtie \text{CUSTOMER}^{\{\}} \bowtie \text{ORDERS}^{\text{Miller}}$$

$$J_2 = \text{ORDERS}^{\text{Smith}} \bowtie \text{CUSTOMER}^{\{\}} \bowtie \text{NATION}^{\{\}} \bowtie \text{CUSTOMER}^{\{\}} \bowtie \text{ORDERS}^{\text{Miller}}$$

ขั้นตอนที่ 4 พิจารณาส่วนของจอยนิ่งเน็ตเวิร์คที่เหมือนกันในแต่ละเน็ตเวิร์ค เพื่อลดจำนวนการจอยกันของข้อมูลและสามารถใช้ส่วนของจอยนิ่งเน็ตเวิร์คนั้นร่วมกันได้ ผลลัพธ์ที่ได้เรียงลำดับตามจำนวนการจอยกันในแต่ละเน็ตเวิร์ค โดยผลลัพธ์ที่มีจำนวนการจอยกันของค่าข้อมูลน้อยกว่าจะถือว่ามีความสำคัญมากกว่า จากตัวอย่าง ส่วนของจอยนิ่งเน็ตเวิร์คที่เหมือนกัน คือ T_1

$$T_1 \leftarrow J_1 = \text{ORDERS}^{\text{Smith}} \bowtie \text{CUSTOMER}^{\{\}}$$

$$C_1 \leftarrow T_1 \bowtie \text{ORDERS}^{\text{Miller}}$$

$$C_2 \leftarrow T_1 \bowtie \text{NATION}^{\{\}} \bowtie \text{CUSTOMER}^{\{\}} \bowtie \text{ORDERS}^{\text{Miller}}$$

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

ORDERS

	ORDKEY	CUSTKEY	ORDSTATUS	TOTALPRICE	ORDDATE	CLERK
o1	100105	12312	complete	$5,000	5/2/2001	John Smith
o2	100111	12312	incomplete	$3,000	5/1/2001	Mike Miller
o3	100125	10001	incomplete	$7,000	5/1/2001	Mike Miller
o4	100110	10002	complete	$8,000	4/2/2001	Keith Brown

CUSTOMER

	CUSTKEY	NAME	ADDRESS	NATKEY
c1	12312	Brad Lou	3811 State Drive, Los Angeles	01
c2	10001	George Wales	43655 Ave, New York	01
c3	10013	John Roberts	3321 Broadways St, San Francisco	01

NATION

	NATKEY	NAME	REGIONKEY
n1	01	USA	N America

รูปที่ 2.4 ตัวอย่างข้อมูลจากฐานข้อมูล TPC-H

4. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

4.1 สถาปัตยกรรมของระบบ เน้นในส่วนการสอบถาม ประสิทธิภาพของระบบขึ้นอยู่กับจำนวนของคำสำคัญ การวิเคราะห์หาจอยนิ่งเน็ตเวิร์ค จำนวนของจอยนิ่งเน็ตเวิร์คที่ได้ และการพิจารณาส่วนของจอยนิ่งเน็ตเวิร์คที่เหมือนกันในแต่ละเน็ตเวิร์คเพื่อสร้างส่วนของการจอยที่สามารถใช้ร่วมกันได้

4.2 แบบจำลองข้อมูลและดัชนีผกผัน ระบบใช้กราฟโครงสร้างฐานข้อมูลเป็นแบบจำลองข้อมูล และใช้ดัชนีผกผันในการชี้ตำแหน่งของคำดัชนีไปยังแบบจำลองข้อมูล โดยกำหนดให้คำดัชนีคือค่าข้อมูลเท่านั้น ทำให้ผู้ใช้ไม่สามารถค้นหาคำพ้องความหมายของคำดัชนี และไม่สามารถค้นหาด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้

4.3 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว ไม่สามารถใช้ข้อความอิสระในการสอบถามได้ นอกจากนี้คำสำคัญที่ใช้ค้นหาอาจตรงกับค่าข้อมูล หรือส่วนหนึ่งของค่าข้อมูล (Sub-String Matching)

4.4 ผลลัพธ์จากการสอบถาม คือ ระเบียนข้อมูลที่สัมพันธ์กัน ระเบียนข้อมูลเหล่านี้ประกอบด้วยคำสำคัญที่ได้จากการสอบถามทั้งหมดหรือไม่ก็ได้

4.5 การจัดลำดับผลลัพธ์ ในกรณีที่ผลลัพธ์มีหลายผลลัพธ์ ระบบมีการจัดลำดับความสำคัญของผลลัพธ์โดยพิจารณาจากจำนวนการจอยกันของระเบียนข้อมูล หรือตามโครงสร้างของกราฟคำตอบ โดยพิจารณาร่วมกับการให้คะแนน IR ของแอทริบิว เพื่อให้ผู้ใช้เลือกผลลัพธ์ที่ดีที่สุด

2.2.3 ดีบีเอกซ์โพลเรอ

ดีบีเอกซ์โพลเรอ [3] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูลอีกระบบหนึ่ง การหาผลลัพธ์ทำได้โดยการค้นหาเซตของเอนทิตี และแอทริบิวหรือระเบียนที่ประกอบด้วยคำสำคัญ เอนทิตีทั้งหมดที่ได้เชื่อมต่อกันตามโครงสร้างฐานข้อมูล ผลลัพธ์ที่ได้ คือ กราฟย่อยที่ประกอบด้วยคำสำคัญทั้งหมดจากการสอบถาม จากนั้นจึงแปลงกราฟย่อยเป็นภาษาเอสคิวแอล ในกรณีที่ผลลัพธ์มีหลายผลลัพธ์ ระบบทำการจัดลำดับความสำคัญของผลลัพธ์ตามกราฟย่อยที่มีระยะทางที่สั้นที่สุด นั่นคือ ผลลัพธ์ที่มีความสัมพันธ์ระหว่างข้อมูลใกล้ชิดกันมากที่สุดให้ถือว่าเป็นผลลัพธ์ที่ดีที่สุด

1. สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบดีบีเอกซ์โพลเรอแบ่งออกเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูลที่เรียกว่า พับลิช (Publish Component) และส่วนการสอบถาม เรียกว่า เสิร์ช (Search Component) ในส่วนของพับลิชประกอบด้วยอินเทอร์เฟส (Interface) สำหรับการเลือกฐานข้อมูล การเลือกเอนทิตีหรือแอทริบิว และการปรับปรุง หรือลบ หรือจัดการแบบจำลองข้อมูล ในส่วนของการสอบถามประกอบด้วยอินเทอร์เฟสสำหรับการค้นหาคำสำคัญจากแบบจำลองข้อมูล และการระบุเอนทิตีแอทริบิว หรือระเบียนสำหรับขั้นตอนการหาผลลัพธ์

2. แบบจำลองข้อมูลและดัชนีผกผัน

ระบบดีบีเอกซ์โพลเรอทำการจัดเก็บแบบจำลองข้อมูลในรูปของกราฟโครงสร้างฐานข้อมูลแบบไม่มีทิศทาง โดยกำหนดให้โหนดใช้แทนเอนทิตี และเส้นเชื่อมใช้แทนความสัมพันธ์ระหว่างเอนทิตี ส่วนการจัดเก็บดัชนีผกผัน ระบบทำการเก็บคำดัชนี 2 ระดับ คือ ระดับคอลัมน์ (Column Level Granularity) และระดับเซลล์ (Cell Level Granularity) ในการเก็บคำดัชนีระดับคอลัมน์ คำดัชนีถูกระบุตำแหน่งไปยังคอลัมน์หรือแอทริบิวที่คำดัชนีนั้นปรากฏอยู่ ส่วนการเก็บคำดัชนีในระดับเซลล์ คำดัชนีถูกระบุตำแหน่งไปยังเซลล์ที่คำดัชนีนั้นปรากฏ ดังตารางที่ 2.1

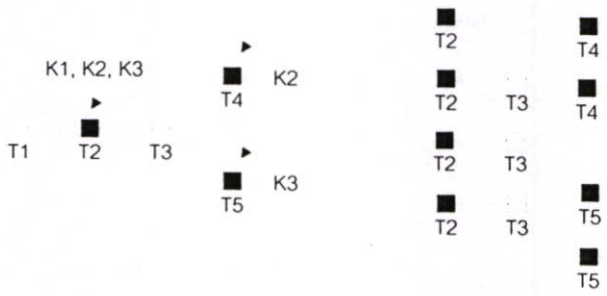
ตารางที่ 2.1 แสดงวิธีการเก็บดัชนีผกผันในระบบดีบีเอ็กซ์โพลเรอ

ระดับการเก็บดัชนี	วิธีการอ้างถึง
ระดับคอลัมน์	table.column
ระดับเซลล์	table.column.rowid

3. ขั้นตอนวิธีในการหาผลลัพธ์

ระบบนำตำแหน่งของคำสำคัญที่ได้จากดัชนีไปตรวจสอบกับกราฟโครงสร้างฐานข้อมูลว่าอยู่ในโหนดใดของกราฟ แล้วจึงทำการตัดโหนดที่ไม่ปรากฏคำสำคัญและโหนดที่ไม่ได้เชื่อม

ระหว่างคำสำคัญทิ้ง กราฟที่เหลืออยู่จะประกอบด้วยเอนทิตีที่คำสำคัญนั้นปรากฏอยู่ทั้งหมด โดยไม่มีคำสำคัญใดปรากฏซ้ำกันมากกว่า 1 โหนด ดังตัวอย่างจากรูปที่ 2.5 สมมติให้ K1, K2, และ K3 เป็นคำสำคัญที่ถูกพบในโหนดดังรูปทางซ้าย ส่วนรูปทางขวาถือเป็นกราฟคำตอบทั้งหมดที่ได้จากคำสำคัญ



รูปที่ 2.5 จอยทรี

4. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

4.1 สถาปัตยกรรมของระบบแบ่งเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูลที่เรียกว่า พับลิช และส่วนการสอบถาม เรียกว่า เสิร์ช

4.2 แบบจำลองข้อมูลและดัชนีผกผัน แบบจำลองข้อมูลของดีบีเอกซ์โพลเรออยู่ในรูปของกราฟโครงสร้างฐานข้อมูล โดยใช้ดัชนีผกผัน 2 ระดับ คือ ระดับคอลัมน์ และระดับเซลล์ ซึ่งจากการทดสอบดัชนีทั้งสองระดับใน [3] พบว่าดัชนีระดับคอลัมน์ให้ประสิทธิภาพในการทำงานดีกว่าดัชนีระดับเซลล์ นอกจากนี้คำดัชนีที่ใช้ในระบบ คือ ค่าข้อมูลจากฐานข้อมูล ทำให้ผู้ใช้ไม่สามารถค้นหาคำพ้องความหมายของคำดัชนี และไม่สามารถค้นหาด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้

4.3 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว ไม่สามารถใช้ข้อความอิสระในการสอบถามได้ นอกจากนี้คำสำคัญที่ใช้ค้นหาอาจตรงกับค่าข้อมูล หรือส่วนหนึ่งของค่าข้อมูล

4.4 ผลลัพธ์จากการสอบถาม คือ กราฟย่อยที่ประกอบด้วยคำสำคัญทั้งหมดจากการสอบถาม

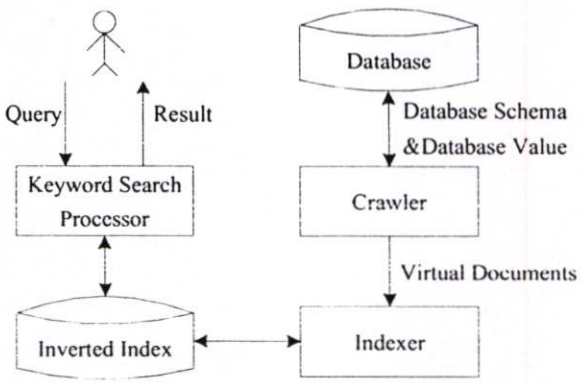
4.5 การจัดลำดับผลลัพธ์ ระบบดีบีเอกซ์โพลเรอใช้โครงสร้างของกราฟคำตอบในการพิจารณาจัดลำดับความสำคัญของผลลัพธ์

2.2.4 อีเคเอสโอ

อีเคเอสโอ [11] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูลเชิงสัมพันธ์ โดยใช้แบบจำลองข้อมูลในรูปของเอกสารเสมือน และใช้ดัชนีผกผันที่มีคำดัชนีตรงกับค่าข้อมูล ซึ่งรายละเอียดของระบบมีดังนี้

1. สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบเป็นดังรูปที่ 2.6 ซึ่งประกอบด้วยส่วนสำคัญ 3 ส่วน ส่วนแรก คือ ครอว์เลอร์ (Crawler) มีหน้าที่ในการดึงข้อมูลจากฐานข้อมูลมาสร้างเอกสารเสมือน ส่วนที่สอง คือ ตัวสร้างดัชนี (Indexer) มีหน้าที่สร้างดัชนีผกผันจากเอกสารเสมือน และสุดท้าย คือ กระบวนการค้นหาคำสำคัญ (Keyword Search Processor)



รูปที่ 2.6 สถาปัตยกรรมของระบบอีเคเอสโอ

2. แบบจำลองข้อมูลและดัชนีผกผัน

แบบจำลองข้อมูลของระบบอยู่ในรูปของเอกสารเสมือน จากรูปที่ 2.6 เอกสารเสมือนถูกสร้างโดยครอว์เลอร์ วิธีการสร้างเอกสารเสมือนแบ่งเป็น 2 ขั้นตอน คือ การสร้างเทกซ์ออบเจกต์ และการสร้างเอกสารเสมือน

2.1 การสร้างเทกซ์ออบเจกต์

เทกซ์ออบเจกต์ประกอบด้วยเซตของระเบียนข้อมูลที่สัมพันธ์กับรูททูเพิล (Root Tuple) [9] ซึ่งถูกกำหนดโดยอัตโนมัติหรือกำหนดโดยผู้ดูแลระบบ จากตัวอย่างข้อมูลในรูปที่ 2.7 เทกซ์ออบเจกต์ของรูททูเพิล P1 ประกอบด้วยเซตของระเบียนข้อมูล {P1, PS1, L1, S1, O1, C1}

2.2 การสร้างเอกสารเสมือน

เอกสารเสมือนประกอบด้วยค่าข้อมูลจากแอทริบิวที่มีชนิดข้อมูลเป็นอักขระ โดยที่แอทริบิวเหล่านั้นมาจากระเบียนข้อมูลในเทกซ์ออบเจกต์ จากตัวอย่างเทกซ์ออบเจกต์ดังกล่าว เอกสารเสมือนที่ได้ คือ Discus fish Acme Wisconsin Bob Madison

Part

	PARTKEY	NAME	...
P1	10	Discus fish	
P2	11	Apple PowerBook	
P3	12	Snow shevel	

Supplier

	SUPPKEY	NAME	ADDRESS	...
S1	1000	Acme	Wisconsin	
S2	1001	Vidgets	California	

Partsupp

	PARTKEY	SUPPKEY	...
PS1	10	1000	
PS2	11	1001	
PS3	12	1000	
PS4	12	1001	

Lineitem

	ORDERKEY	PARTKEY	SUPPKEY	LINENUMBER	...
L1	10001	10	1000	2	
L2	10002	11	1001	1	
L3	10003	12	1000	3	

Orders

	ORDERKEY	CUSTKEY	...
O1	10001	100	
O2	10002	100	
O3	10003	101	

Customer

	CUSTKEY	NAME	ADDRESS	...
C1	100	Bob	Madison	
C2	101	Alis	San Francisco	

รูปที่ 2.7 ตัวอย่างข้อมูลจากฐานข้อมูล TPC-H

ในส่วนของดัชนีผกผัน จากรูปที่ 2.6 ตัวสร้างดัชนีทำการสร้างดัชนีผกผันจากเอกสารเสมือนดังกล่าว โดยใช้ดีบีทูเอ็นเอสอี (DB2 NSE: DB2 Net Search Extender)

3. ขั้นตอนวิธีในการหาผลลัพธ์

ในกระบวนการค้นหาคำสำคัญ โปรแกรมค้นหาเอ็นเอสอี (NSE Search Engine) ทำการค้นหาคำสำคัญจากดัชนีผกผัน และคืนค่าผลลัพธ์ซึ่งเป็นรหัสในเทกซ์ออบเจกต์ ซึ่งผู้ใช้สามารถเลือกได้ว่าต้องการค้นดูข้อมูลในฐานข้อมูลจากรหัสเหล่านี้ หรือต้องการสอบถามใหม่

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

4. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

4.1 สถาปัตยกรรมของระบบแบ่งเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูล ซึ่งประกอบด้วย ครอว์เลอร์และตัวสร้างดัชนี และส่วนการสอบถาม

4.2 แบบจำลองข้อมูลและดัชนีผกผัน แบบจำลองข้อมูลของระบบอยู่ในรูปของเอกสารเสมือน โดยในแต่ละเอกสารเสมือนประกอบด้วยค่าข้อมูลและโครงสร้างฐานข้อมูล ทำให้การค้นหามีประสิทธิภาพยิ่งขึ้น ในส่วนของดัชนีผกผันซึ่งสร้างจากเอกสารเสมือนดังกล่าว ประกอบด้วยคำดัชนีที่มาจากข้อมูลประเภทอักขระ ระบบจึงไม่สามารถค้นหาคำพ้องความหมายของคำดัชนี ไม่สามารถจัดการกับคำดัชนีที่เป็นจำนวนหรือวันที่ได้ และไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้

4.3 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว

4.4 ผลลัพธ์จากการสอบถาม คือ ระเบียนข้อมูลที่ปรากฏคำสำคัญทั้งหมด

2.3 ระบบการค้นหาคำสำคัญในฐานข้อมูลที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล

ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูลแบ่งเป็น คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อแอทริบิว คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี และคำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี/ชื่อแอทริบิว ซึ่งระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลหรือชื่อแอทริบิว ได้แก่ ดีบีเซอร์ฟเฟอร์ ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี ได้แก่ บีเอเอ็นเคเอส และระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี/ชื่อแอทริบิว ได้แก่ และมราจียาติ

2.3.1 ดีบีเซอร์ฟเฟอร์

ดีบีเซอร์ฟเฟอร์ [13] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูล ที่ใช้แบบจำลองข้อมูลในรูปของกราฟโครงสร้างฐานข้อมูล และดัชนีผกผันที่สร้างจากเอกสารเสมือน ซึ่งคำดัชนีเป็นคำที่ตรงกับค่าข้อมูลหรือชื่อแอทริบิว ผลลัพธ์ที่ได้จากการสอบถามอยู่ในรูปของเทรล (Trail) [14] ซึ่งไม่จำเป็นต้องแปลงให้อยู่ในรูปของภาษาเอสคิวแอล

1. แบบจำลองข้อมูลและดัชนีผกผัน

ดัชนีผกผันของระบบสร้างจากเอกสารเสมือน หรือเว็บเพจ โดยการดึงข้อมูลในแต่ละระเบียนมาสร้างเป็นเอกสารเสมือนซึ่งอยู่ในรูปของภาษาเอกซ์เอ็มแอล (XML: Extensible Markup Language) ดังตัวอย่างจากรูปที่ 2.8 ซึ่งเป็นเอกสารเสมือนที่ได้จากการดึงข้อมูลจากฐานข้อมูลดีบี

แอลพี (DBLP) [14] ระบบทำการเก็บดัชนีผกผันที่ได้จากเนื้อความ (Textual Content) ของเอกสารเสมือนนี้ รวมทั้งค่า tf.idf ของคำดัชนี

```
1. <PUBLICATION>
2.    <row>
3.       <JOURNAL> Advances in Computers </JOURNAL>
4.       <KEY> journals/ac/Dam66 </KEY>
5.       <PAGES> 239-290 </PAGES>
6.       <TITLE> Computer Driven Displays and Their Use in Man/Machine Interaction. </TITLE>
7.       <TYPE> article </TYPE>
8.       <URL> http://dblp.uni-trier.de/db/journals/ac/ac7.html#Dam66 </URL>
9.       <VOLUME> 7 </VOLUME>
10.      <YEAR> 1966 </YEAR>
11.   </row>
12. </PUBLICATION>
```

รูปที่ 2.8 ตัวอย่างเอกสารเสมือนในรูปของภาษาเอกซ์เอ็มแอลที่ได้จากฐานข้อมูลดีบีแอลพี

ส่วนแบบจำลองข้อมูลของระบบ คือ เอกสารเสมือนที่เปรียบได้กับกราฟข้อมูล และกราฟโครงสร้างฐานข้อมูลที่เรียกว่า ลิงค์กราฟ (Link Graph) ซึ่งพิจารณาจากเงื่อนไขบังคับของคีย์นอก (Foreign Key Constraint) โดยมีลิงค์เชื่อมระหว่างเอนทิตีและแอทริบิว ระหว่างระเบียนข้อมูลกับเอกสารเสมือน ระหว่างเอกสารเสมือนกับเอกสารเสมือน และระหว่างเอกสารเสมือนหรือยูอาร์แอล (URL) ของเอกสารเสมือน

2. ขั้นตอนวิธีในการหาผลลัพธ์

ขั้นตอนวิธีในการหาผลลัพธ์มี 4 ขั้นตอน คือ ขั้นตอนแรก ระบบทำการคำนวณคะแนนของแต่ละโหนดหรือเอกสารเสมือนที่ประกอบด้วยคำสำคัญอย่างน้อยหนึ่งคำ ขั้นตอนที่สอง คือ การสร้างเทรลโดยใช้เบสเทรลอัลกอริทึม (Best Trail Algorithm) [14] ขั้นตอนที่สาม คือ การลบเทรลที่ซ้ำกัน ขั้นตอนสุดท้าย คือ การคำนวณคะแนนของเทรล เพื่อจัดลำดับความสำคัญของเทรลหรือผลลัพธ์ และแสดงผลต่อผู้ใช้

3. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

3.1 แบบจำลองข้อมูลและดัชนีผกผัน ระบบดีบีเซอร์ฟเฟอร์ใช้แบบจำลองข้อมูลในรูปของเอกสารเสมือน โดยเอกสาร 1 เอกสารใช้แทนข้อมูล 1 ระเบียน และอาศัยกราฟโครงสร้างฐานข้อมูลในการเชื่อมโยงเอกสารเหล่านั้นเข้าด้วยกัน ซึ่งเทียบได้กับการเก็บแบบจำลองในรูปของกราฟข้อมูล การหาผลลัพธ์ของระบบจึงต้องทำงานกับกราฟจำนวนมาก และต้องใช้หน่วยความจำใน

การเก็บเอกสารเหล่านั้นมากขึ้น ส่วนข้อดีของการเก็บแบบจำลองดังกล่าว คือ สามารถเข้าถึงค่าข้อมูลได้โดยผ่านทางเว็บเพจโดยไม่จำเป็นต้องใช้ภาษาเอสคิวแอลในการติดต่อฐานข้อมูล

ในส่วนของดัชนีผกผันซึ่งสร้างจากเอกสารเสมือนดังกล่าว ประกอบด้วยคำดัชนีที่มาจากข้อมูลประเภทอักขระเท่านั้น ระบบจึงไม่สามารถค้นหาคำพ้องความหมายของคำดัชนี ไม่สามารถจัดการกับคำดัชนีที่เป็นจำนวนหรือวันที่ได้ และไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้

3.2 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว นอกจากนี้ระบบอนุญาตให้ผู้ใช้ทำการสอบถามแบบกำหนดรูปแบบเพื่อให้ใช้คำสำคัญที่ตรงกับชื่อแอทริบิวได้ โดยใช้เครื่องหมาย "=" ในการกำหนดแอทริบิว รูปแบบของภาษาสอบถาม คือ x = y หมายถึง แอทริบิว x มีค่าเท่ากับ y ตัวอย่างเช่น การสอบถาม "simon" หมายถึงข้อมูลจากทุกระเบียนที่ปรากฏคำว่า "simon" ส่วนการสอบถาม "name = simon" หมายถึงระเบียนข้อมูลที่แอทริบิว "name" มีค่าเป็น "simon" เป็นต้น

3.3 ผลลัพธ์จากการสอบถาม คือ เซตของเอกสารเสมือนที่ประกอบด้วยคำสำคัญทั้งหมด เอกสารเสมือนดังกล่าวถูกแสดงผลในรูปของเทรล

3.4 การจัดลำดับผลลัพธ์ ระบบจัดลำดับความสำคัญของผลลัพธ์โดยพิจารณาจากจำนวนของคำสำคัญในแต่ละเทรล จำนวนครั้งของคำสำคัญที่ปรากกฏในเทรล และคะแนน IR ของเทรล

2.3.2 บีเอเอ็นเคเอส

ระบบบีเอเอ็นเคเอส [1][2] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้กราฟข้อมูลที่กำหนดให้โหนดแทนระเบียนข้อมูล และเส้นเชื่อมแทนความสัมพันธ์ระหว่างระเบียน ผลลัพธ์จากการสอบถามเป็นกราฟย่อยที่ประกอบด้วยคำสำคัญทั้งหมด การจัดลำดับความสำคัญของผลลัพธ์ขึ้นอยู่กับผลรวมของน้ำหนักโหนด (Node Score) และน้ำหนักของเส้นเชื่อม (Edge Weight)

1. แบบจำลองข้อมูลและดัชนีผกผัน

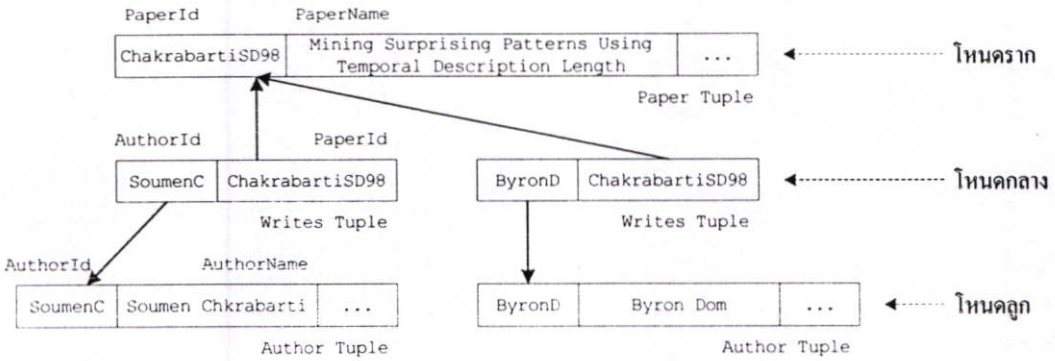
บีเอเอ็นเคเอสมีการเก็บแบบจำลองข้อมูลเป็นกราฟข้อมูลแบบมีทิศทาง โดยการให้โหนดแทนระเบียนข้อมูล และเส้นเชื่อมแทนความสัมพันธ์ระหว่างคีย์หลัก-คีย์นอกของระเบียนข้อมูลนั้น และเก็บดัชนีผกผันเพื่ออ้างไปยังโหนดที่คำสำคัญนั้นปรากฏอยู่

2. ขั้นตอนวิธีในการหาผลลัพธ์

ในการหาผลลัพธ์ บีเอเอ็นเคเอสจะทำการค้นหาเซตของโหนด (S_i) ทั้งหมดที่มีคำสำคัญ (t_i) ตามที่ผู้ใช้ระบุจากดัชนีผกผัน ผลลัพธ์ที่ได้จากการสอบถาม คือ ทรีแบบมีทิศทางที่ประกอบด้วยโหนดราก (Root Node) โหนดใบ (Leaf Node) และ/หรือโหนดกลาง (Intermediate Node) โหนดรากเป็นโหนดที่อยู่ตรงกลางของทรีที่สามารถเชื่อมต่อไปยังทุกๆ โหนดใบได้ โหนดใบเป็นโหนดซึ่งมีคำสำคัญปรากฏอยู่ ส่วนโหนดกลางเป็นโหนดซึ่งเชื่อมระหว่างโหนดรากและ

โหนดใบ ในแต่ละทรีต้องประกอบด้วยอย่างน้อย 1 โหนดที่มาจากแต่ละ S_i โดยที่ทรีนั้นอาจมีโหนดที่ไม่จำเป็นต้องเป็นสมาชิกของ S_i ใดๆ ก็ได้เพื่อเป็นโหนดที่เชื่อมระหว่างโหนดใบกับโหนดราก

ตัวอย่างทรีที่เป็นผลลัพธ์จากการการสอบถามด้วย "soumen byron" โดยมีโหนดราก โหนดกลาง และโหนดลูก เป็นดังรูปที่ 2.9



รูปที่ 2.9 ตัวอย่างทรีที่เป็นผลลัพธ์จากการการสอบถามด้วย "soumen byron"

ในกรณีที่ผลลัพธ์มีหลายผลลัพธ์ จะมีเพียงผลลัพธ์ส่วนหนึ่งที่ถูกนำเสนอแก่ผู้ใช้ ผลลัพธ์ดังกล่าว คือ ทรีที่มีน้ำหนักของทรีที่น้อยที่สุดจำนวน n ลำดับ ทรีแต่ละทรีจะถูกนำมาคำนวณน้ำหนัก โดยที่

คะแนนของทรี = ค่า proximity + ค่า prestige

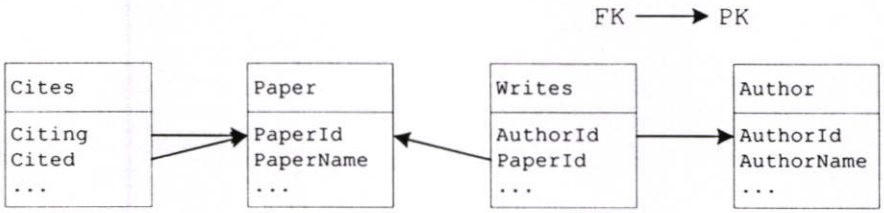
ค่า proximity ของแต่ละทรีได้มาจากน้ำหนักของกิ่งในทรีนั้น ซึ่งสามารถหาได้ ดังต่อไปนี้

proximity = edge score

= log(1/$\sum$ น้ำหนักของกิ่ง)

ในกรณีที่กิ่งนั้นเป็นฟอร์เวิร์ดเอด (Forward Edge) หรือกิ่งที่ชี้ไปยังทิศทางของโหนดราก [3] น้ำหนักของกิ่งจะขึ้นอยู่กับโครงสร้างฐานข้อมูล เช่น จากรูปที่ 2.10 น้ำหนักของกิ่งระหว่างตาราง Cites และ Paper มีค่ามากกว่าน้ำหนักของกิ่งระหว่างตาราง Writes และ Paper ดังนั้นจากรูปที่ 2.9 น้ำหนักของฟอร์เวิร์ดเอดทั้งสองจึงเท่ากัน

ในกรณีที่กิ่งนั้นเป็นแบคเวิร์ดเอด (Backward Edge) หรือกิ่งที่ชี้ไปยังทิศตรงข้ามกับโหนดราก น้ำหนักของกิ่งเท่ากับจำนวนอินดีกรี (Indegree) ของกิ่งที่ชี้มายังโหนดนั้น เช่น จากรูปที่ 2.9 น้ำหนักของกิ่งทั้งสองเท่ากับ 1 เนื่องจากโหนดลูกทั้งสองนั้นมีจำนวนอินดีกรีเท่ากัน



รูปที่ 2.10 ตัวอย่างโครงสร้างฐานข้อมูลของบีเอเอ็นเคเอส

ค่า prestige หาได้จากน้ำหนักของโหนดในทรี โดยพิจารณาเฉพาะโหนดรากและโหนดใบเท่านั้นเพื่อลดผลกระทบจากโหนดกลางที่อาจมีจำนวนมาก ซึ่งสามารถหาได้ดังนี้

prestige = node score

= น้ำหนักของโหนดราก + $\sum$ น้ำหนักของโหนดใบ

โดยที่

น้ำหนักโหนด = log(1+ อินดีกรี)

จากรูปที่ 2.9 โหนดรากมีจำนวนอินดีกรีเท่ากับ 2 โหนดลูกทางซ้ายและโหนดลูกทางขวามีจำนวนอินดีกรีเท่ากัน คือ เท่ากับ 1

3. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

3.1 แบบจำลองข้อมูลและดัชนีผกผัน เนื่องจากระบบใช้แบบจำลองข้อมูลที่เป็นกราฟข้อมูลทำให้ระบบต้องจัดการกับกราฟจำนวนมาก การทำงานของระบบจึงค่อนข้างช้า และในกรณีที่ฐานข้อมูลมีขนาดใหญ่ ระบบต้องใช้หน่วยความจำในการเก็บข้อมูลมากยิ่งขึ้น ในส่วนของดัชนีผกผัน คำดัชนีที่เก็บ คือ ค่าข้อมูล และชื่อเอนทิตี คำสำคัญที่ใช้ค้นหาจึงเป็นคำที่มีอยู่ในค่าข้อมูลหรือชื่อเอนทิตีเท่านั้น ไม่สามารถค้นหาคำพ้องความหมายของคำดัชนีได้ นอกจากนี้ แม้ระบบจะสามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ แต่ยังคงไม่สามารถค้นหาจากชื่อแอทริบิวได้

3.2 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว ไม่สามารถใช้ข้อความอิสระในการสอบถามได้

3.3 ผลลัพธ์จากการสอบถาม คือ กราฟย่อยที่ประกอบด้วยระเบียนข้อมูลต่างๆ โดยที่ระเบียนเหล่านั้นต้องประกอบด้วยคำสำคัญทั้งหมดที่ได้จากการสอบถาม และมีความสัมพันธ์กันตามโครงสร้างฐานข้อมูล

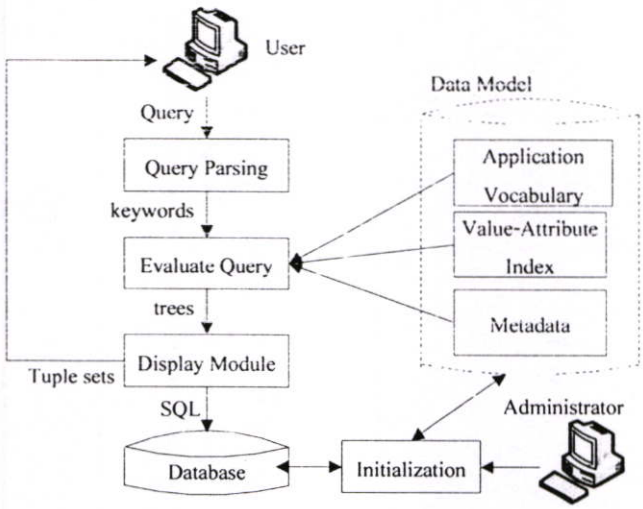
3.4 การจัดลำดับผลลัพธ์ของระบบบีเอเอ็นเคเอสพิจารณาจากโครงสร้างของกราฟคำตอบและความหมายของเส้นเชื่อม

2.3.3 มราจียาติ

มราจียาติ [10] เป็นระบบการค้นหาคำสำคัญในฐานข้อมูลที่ต่างจากวิธีการดังกล่าวข้างต้นทั้งในเรื่องของแบบจำลองข้อมูล ขั้นตอนในการหาผลลัพธ์ และการจัดลำดับความสำคัญของผลลัพธ์ ซึ่งมีรายละเอียดดังต่อไปนี้

1. สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบมราจียาติเป็นดังรูปที่ 2.11 ซึ่งประกอบด้วยส่วนสำคัญ 2 ส่วน คือ ส่วนการเตรียมข้อมูลเริ่มต้น และส่วนการสอบถาม ในส่วนการเตรียมข้อมูลเริ่มต้น ผู้ดูแลระบบเป็นผู้จัดการแบบจำลองข้อมูล และในส่วนการสอบถามประกอบด้วยการทำงาน 3 ขั้นตอน คือ การวิเคราะห์การสอบถาม การหาผลลัพธ์ และการแสดงผลลัพธ์



รูปที่ 2.11 สถาปัตยกรรมของระบบมราจียาติ

ในส่วนการวิเคราะห์การสอบถาม ระบบเริ่มค้นหาคำสำคัญจากคำแอบพลิเคชัน และดัชนีค่าข้อมูล-แอทริบิว ตามลำดับ หากไม่พบคำสำคัญจากแบบจำลองข้อมูลดังกล่าว ระบบจึงค้นหาคำสำคัญจากข้อมูลฐานข้อมูลต่อไป

2. แบบจำลองข้อมูลและดัชนีผกผัน

มราจียาติเก็บแบบจำลองข้อมูลในรูปของตารางโครงสร้างฐานข้อมูล โดยแบ่งข้อมูลเป็น 3 ส่วน คือ

- คำแอบพลิเคชัน คือ คำที่มีความหมายสื่อถึงชื่อเอนทิตี ชื่อแอทริบิว และค่าข้อมูลที่อยู่ในรูปของรหัส หรือคำย่อ เช่น ค่าข้อมูล 'F' แทนด้วย 'female' หรือ ค่าข้อมูล 'CS' แทนด้วย 'computer science' เป็นต้น
- ดัชนีค่าข้อมูล-แอทริบิว คือ ดัชนีที่ใช้เก็บค่าข้อมูล และตำแหน่งของค่าข้อมูลในฐานข้อมูล ซึ่งจะอยู่ในรูปของ <ค่าข้อมูล, แอทริบิว, เอนทิตี>
- ข้อมูลฐานข้อมูล ประกอบด้วยตาราง 4 ตาราง คือ รายการเอนทิตี รายการแอทริบิว รายการคีย์หลัก และรายการคีย์นอก

3. ขั้นตอนวิธีในการหาผลลัพธ์

ข้อแตกต่างระหว่างระบบมราจียาติและวิธีที่ผ่านมา คือ ผู้ใช้สามารถระบุคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ ซึ่งระบบจะใช้คำสำคัญนี้ในการเจาะจงระเบียนผลลัพธ์เป็นแอทริบิวใดแอทริบิวหนึ่งหรือเอนทิตีใดเอนทิตีหนึ่งได้ ตัวอย่างการสอบถามของผู้ใช้จากตัวอย่างข้อมูลในตารางที่ 2.2 และ 2.3

Q1 : "John" หมายถึง ผู้ใช้ต้องการข้อมูลเกี่ยวกับ John

Q2 : "John Sport" หมายถึง ผู้ใช้ต้องการข้อมูล Sport ของ John

Q3 : "John Activity" หมายถึง ผู้ใช้ต้องการข้อมูล Activity ของ John

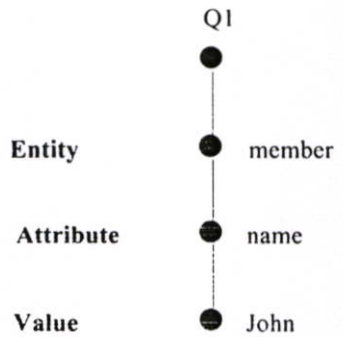
ตารางที่ 2.2 แสดงตัวอย่างข้อมูลในเอนทิตี Member

Name	City	Age
John	New York	45
...	...	...

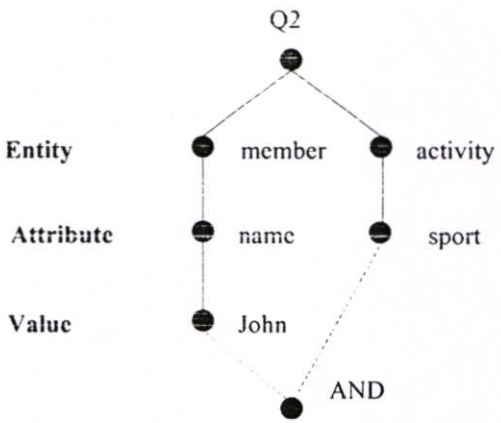
ตารางที่ 2.3 แสดงตัวอย่างข้อมูลในเอนทิตี Activity

Name	Sport
John	Running
John	Swimming
...	...

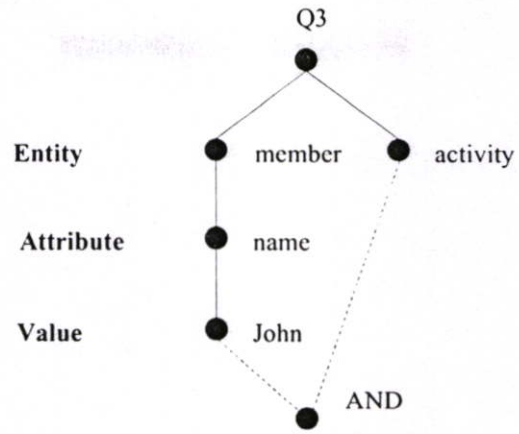
วิธีการหาผลลัพธ์ของมราจียาติแตกต่างจากวิธีการที่ผ่านมาเช่นกัน โดยระบบนำตำแหน่งที่ได้จากดัชนีมาสร้างเป็นทรีที่สามารถเป็นไปได้ทั้งหมด ทรีที่ได้มีความสูง 4 ระดับ คือ ระดับรากของทรี ระดับเอนทิตี ระดับแอทริบิว และระดับค่าข้อมูล โดยการสร้างทรีจะเริ่มจากระดับล่างสุด คือ ระดับค่าข้อมูล ระดับแอทริบิว ระดับเอนทิตี และระดับรากของทรี ตามลำดับ แล้วนำทรีที่ได้มาแปลงให้อยู่ในรูปของภาษาเอสคิวแอล ตัวอย่างทรีที่ได้จากการสอบถามด้วย Q1 Q2 และ Q3 เป็นดังรูปที่ 2.12 2.13 และ 2.14 ตามลำดับ



รูปที่ 2.12 ทรีที่ได้จากการสอบถามด้วย "John"



รูปที่ 2.13 ทรีที่ได้จากการสอบถามด้วย "John sport"



รูปที่ 2.14 ทรีที่ได้จากการสอบถามด้วย "John activity"

4. การวิเคราะห์

การวิเคราะห์ระบบตามหลักการทั่วไปของการค้นหาคำสำคัญในฐานข้อมูล มีดังต่อไปนี้

4.1 สถาปัตยกรรมของระบบ แบ่งเป็น 2 ส่วน คือ ส่วนการเตรียมข้อมูล และส่วนการสอบถาม ในส่วนการสอบถาม วิธีการหาผลลัพธ์โดยการสร้างทรีดังกล่าวทำให้ระบบไม่จำเป็นต้องทำงานกับกราฟจำนวนมาก แต่ทำให้ไม่สามารถค้นหาผลลัพธ์ที่ประกอบด้วยเอนทิตีมากกว่า 2 เอนทิตีได้

4.2 แบบจำลองข้อมูลและดัชนีผกผัน ระบบมราจียาติเก็บดัชนีผกผันร่วมกับแบบจำลองข้อมูล กล่าวคือ แบบจำลองข้อมูลของระบบประกอบด้วยข้อมูล 3 ส่วน คือ คำแอบพลิเคชัน ดัชนีค่าข้อมูล-แอทริบิว และข้อมูลฐานข้อมูล ซึ่งข้อมูลทั้ง 3 ส่วนนี้เก็บอยู่ในรูปของตารางที่แยกจากกัน การค้นหาคำสำคัญจึงต้องทำการค้นหาทีละส่วน ข้อดีของการเก็บแบบจำลองดังกล่าว คือ ผู้ใช้สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูล และคำแอบพลิเคชันดังกล่าวได้ ส่วนข้อเสีย คือ ในกรณีที่คำสำคัญตรงกับค่าข้อมูล และข้อมูลฐานข้อมูล ระบบทำการสร้างทรีเฉพาะคำสำคัญที่ตรงกับค่าข้อมูลเท่านั้น ดังนั้นผลลัพธ์ที่ได้ คือ ทรีที่ประกอบด้วยคำสำคัญที่ตรงกับค่าข้อมูลเท่านั้น ส่วนทรีที่ประกอบด้วยคำสำคัญตรงกับข้อมูลฐานข้อมูลจะไม่ถูกสร้าง

4.3 ภาษาสอบถาม ระบบกำหนดให้การสอบถามเป็นการระบุเซตของคำสำคัญ หรือเซตของคำสำคัญแบบเดี่ยว

4.4 ผลลัพธ์จากการสอบถาม คือ ทรีที่ประกอบด้วยคำสำคัญ ซึ่งจำนวนเอนทิตีของทรีต้องไม่เกิน 2 เอนทิตี ทำให้การสอบถามบางการสอบถามไม่สามารถหาผลลัพธ์ได้

4.5 การจัดลำดับผลลัพธ์ของมราจียาติทำได้ 2 วิธี คือ การจัดลำดับผลลัพธ์ตามการจัดเรียงของแอทริบิวใดแอทริบิวหนึ่ง ซึ่งผู้ใช้เป็นผู้กำหนด และการจัดลำดับผลลัพธ์ตามจำนวนของอิน

ดีกรีของโหนดราก ซึ่งการจัดลำดับผลลัพธ์ทั้ง 2 วิธีดังกล่าวไม่ได้จัดเรียงตามความสำคัญของผลลัพธ์

2.4 การวิเคราะห์ระบบการค้นหาคำสำคัญในฐานข้อมูล

โดยทั่วไป การค้นหาคำสำคัญในฐานข้อมูลนำดัชนีผกผันมาใช้ในการเก็บคำสำคัญ ส่วนแบบจำลองข้อมูลถูกเก็บในรูปของกราฟโครงสร้างฐานข้อมูล หรือกราฟข้อมูล การหาผลลัพธ์จึงอาศัยความสัมพันธ์ระหว่างโหนดในกราฟ ซึ่งในการสอบถามแต่ละครั้งอาจมีจำนวนผลลัพธ์มากกว่า 1 ผลลัพธ์ ดังนั้นในการพิจารณาความสำคัญของผลลัพธ์แต่ละผลลัพธ์ จึงอาศัยแนวความคิดความใกล้กันของโหนดในกราฟ หรือระยะทางที่สั้นที่สุด โดยถือว่ากราฟที่มีระยะทางในการเชื่อมต่อกันของโหนดสั้นที่สุดคือผลลัพธ์ที่ดีที่สุด

ข้อจำกัดของวิธีการค้นหาคำสำคัญจากฐานข้อมูลมีดังต่อไปนี้

1. คำสำคัญที่ใช้ค้นหาต้องเป็นคำที่มีอยู่ในค่าข้อมูลหรือเป็นส่วนหนึ่งของค่าข้อมูลเท่านั้น ไม่สามารถค้นหาคำพ้องความหมายของคำสำคัญได้ เช่น ในฐานข้อมูลมีการเก็บค่าข้อมูล “CA” แทนคำว่า “California” เมื่อผู้ใช้ค้นหาด้วยคำว่า “California” ระบบไม่สามารถค้นหาคำสำคัญดังกล่าวนี้ได้ ดังนั้นการเพิ่มความสามารถในการค้นหาด้วยคำพ้องคำสำคัญช่วยให้ผู้ใช้ค้นหาข้อมูลได้แม่นยำยิ่งขึ้น

2. ระบบการค้นหาคำสำคัญในฐานข้อมูลส่วนใหญ่ไม่สามารถใช้คำสำคัญที่ตรงกับข้อมูลของฐานข้อมูลเพื่อช่วยในการสอบถามได้ เช่น การสอบถามด้วย “simon” ในตัวอย่างจากหัวข้อ 2.3.1 ระบบต้องพิจารณาระบียนข้อมูลทุกระเบียนที่มีคำว่า “simon” แต่ถ้าผู้ใช้เจาะจงการสอบถามด้วย “name simon” ระบบจะพิจารณาเฉพาะระเบียนข้อมูลในแอทริบิว Name เท่านั้น การสอบถามด้วยข้อมูลฐานข้อมูลดังกล่าวจึงช่วยในการจำกัดขอบเขตของค่าข้อมูล และสามารถจำกัดผลลัพธ์โดยการระบุเอนทิตีหรือแอทริบิวได้ ทำให้การค้นหาคำสำคัญในฐานข้อมูลให้ผลลัพธ์ที่แม่นยำยิ่งขึ้น

3. ในบางระบบ อนุญาตให้ผู้ใช้สอบถามด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ แต่ยังคงมีการจำกัดรูปแบบของการสอบถาม เช่น การใช้เครื่องหมาย “=” ใน [13] ดังนั้นการสอบถามด้วยข้อความอิสระแต่ยังคงสอบถามด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ จึงน่าจะช่วยเพิ่มความสะดวกให้กับผู้ใช้มากยิ่งขึ้น

4. ระบบการค้นหาคำสำคัญส่วนใหญ่กำหนดให้ใช้คำดัชนีที่มีชนิดข้อมูลเป็นอักขระเท่านั้น ไม่สามารถใช้คำดัชนีที่เป็นจำนวนหรือวันที่ได้ เช่น จากตัวอย่างข้อมูลในรูปที่ 2.7 คำดัชนีคือ คำที่มาจากแอทริบิว NAME และ ADDRESS เท่านั้น ไม่สามารถนำคำจากแอทริบิวอื่นที่มีชนิดข้อมูลเป็นจำนวนมาสร้างเป็นคำดัชนีได้

5. ในกรณีที่ผู้ใช้สอบถามข้อมูลโดยใช้คำสำคัญชุดเดียวกัน ระบบส่วนใหญ่ต้องเริ่มกระบวนการค้นหาใหม่ การจัดการเรื่องเคสเบสเลินนิง (Case-Based Learning) และเคสเบสรีซันนิง (Case-Based Reasoning) จะช่วยยืนยันความถูกต้องของผลลัพธ์และเพิ่มประสิทธิภาพในการสอบถามได้

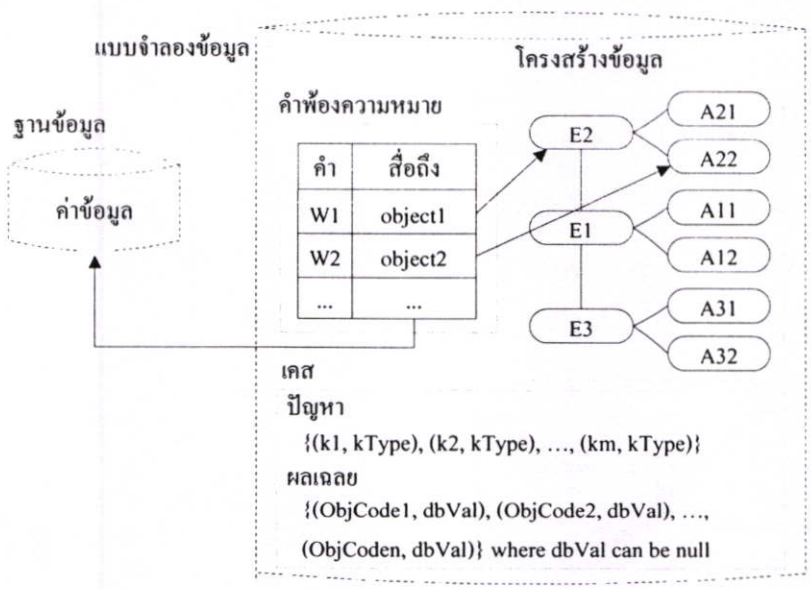
บทที่ 3
การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

บทนี้เป็นการนำเสนอระบบการค้นหาคำสำคัญในฐานข้อมูล ซึ่งผู้ใช้สามารถค้นหาด้วยคำสำคัญที่ปรากฏในค่าข้อมูล หรือข้อมูลฐานข้อมูล (ชื่อเอนทิตี หรือชื่อแอทริบิว) และสามารถค้นหาจากคำพ้องความหมายได้ ผลลัพธ์ที่ได้จากการค้นหา คือ ระเบียนข้อมูลที่ประกอบด้วยคำสำคัญและ/หรือชื่อเอนทิตีหรือชื่อแอทริบิวตรงกับคำสำคัญตามที่ผู้ใช้ต้องการ โดยระเบียนเหล่านั้นมีความสัมพันธ์กันตามโครงสร้างฐานข้อมูล

ในส่วนที่ 3.1 เป็นการนำเสนอแบบจำลองข้อมูล ส่วนที่ 3.2 นำเสนอสถาปัตยกรรมของระบบ ส่วนที่ 3.3 กล่าวถึงอัลกอริทึมที่ใช้ในการค้นหาผลลัพธ์ และส่วนสุดท้าย กล่าวถึงการจัดการเคส

3.1 แบบจำลองข้อมูล

แบบจำลองข้อมูลจัดเก็บข้อมูลที่จำเป็นสำหรับการค้นหาผลลัพธ์ แบ่งออกเป็น 3 ส่วน คือ โครงสร้างฐานข้อมูล เคส และคำพ้องความหมาย ดังรูปที่ 3.1 ซึ่งมีรายละเอียด ดังต่อไปนี้



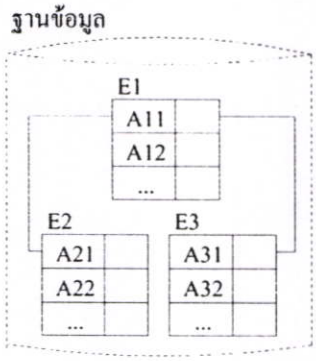
รูปที่ 3.1 แบบจำลองข้อมูล

3.1.1 โครงสร้างฐานข้อมูล

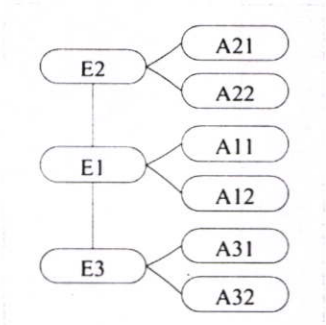
โครงสร้างฐานข้อมูลถูกจัดเก็บในรูปของกราฟแบบไม่มีทิศทางที่กำหนดให้โหนดใช้แทนเอนทิตีและแอทริบิว ส่วนเส้นเชื่อมใช้แทนความสัมพันธ์ระหว่างเอนทิตีกับเอนทิตี และความสัมพันธ์ระหว่างเอนทิตีกับแอทริบิว ซึ่งเส้นเชื่อมระหว่างเอนทิตีกับเอนทิตีอาศัยความสัมพันธ์ระหว่างคีย์หลักและคีย์นอก กราฟโครงสร้างข้อมูลมีนิยาม ดังนี้

นิยามที่ 1 กราฟโครงสร้างข้อมูล SG (Schema Graph) คือ กราฟ <V, E> โดยที่ V คือเซตของจุด 2 ประเภท คือ เอนทิตีและแอทริบิว ส่วน E คือ เซตของข้อบังคับในการเชื่อมต่อระหว่างจุด ซึ่งได้แก่ การเชื่อมต่อระหว่างแอทริบิวกับเอนทิตี (E^{ac}) และเอนทิตีกับเอนทิตี (E^{cc})

ตัวอย่างกราฟโครงสร้างฐานข้อมูล สมมติให้ฐานข้อมูลประกอบด้วยเอนทิตี 3 เอนทิตี และมีความสัมพันธ์ระหว่างคีย์หลักและคีย์นอกเป็นดังรูปที่ 3.2 กราฟโครงสร้างฐานข้อมูลเป็นดังรูปที่ 3.3

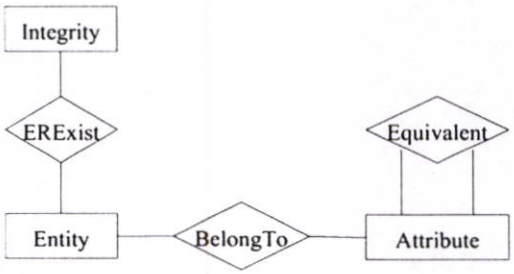


รูปที่ 3.2 ตัวอย่างโครงสร้างฐานข้อมูล



รูปที่ 3.3 ตัวอย่างกราฟโครงสร้างฐานข้อมูล

การจัดเก็บกราฟโครงสร้างข้อมูลดังตัวอย่างในรูปที่ 3.3 สามารถแสดงความสัมพันธ์ระหว่างข้อมูลฐานข้อมูลได้ดังรูปที่ 3.4 ซึ่งประกอบด้วยเอนทิตี Entity สำหรับเก็บข้อมูลเอนทิตีต่างๆ ในฐานข้อมูล เอนทิตี Attribute สำหรับเก็บข้อมูลแอทริบิว เอนทิตี Equivalent สำหรับเก็บแอทริบิวที่เท่าเทียมกัน และเอนทิตี Integrity เก็บข้อมูลความสัมพันธ์ระหว่างเอนทิตี



รูปที่ 3.4 แผนภาพแสดงความสัมพันธ์ระหว่างข้อมูลฐานข้อมูล

3.1.2 เคส

ระบบนำเคสในเคส-เบสรีซันนิง [4] มาใช้ในการพิจารณาหาผลลัพธ์ที่เหมาะสม โดยกำหนดให้เคสประกอบด้วยปัญหา (Problem) และผลเฉลย (Solution) ที่ได้จากการสอบถามหนึ่งๆ ในส่วนของปัญหาคือเซตของคำสำคัญที่ได้จากการสอบถาม และผลเฉลยสร้างจากกราฟคำตอบ โดยที่

$$Case = (Problem, Solution)$$

ซึ่ง Problem คือ เซตของจุดที่ค้นพบในกราฟคำตอบ และ Solution คือ กราฟคำตอบ โดยที่

$$Problem = \{(v_1, kType), (v_2, kType), \ldots, (v_m, kType)\}$$

เมื่อ v_i แทน จุดที่ค้นพบคำสำคัญ i โดยที่ $v_i \in$ กราฟคำตอบ

kType แทน ประเภทของคำสำคัญ โดย kType $\in$ {A: Attribute, E: Entity, V: Value}

m แทน จำนวนคำสำคัญที่ได้จากการสอบถาม

$$Solution = (V^{ac}, E^{cc})$$

เมื่อ V^{ac} คือ เซตของคู่ลำดับระหว่างจุดแอทริบิวหรือจุดเอนทิตี และจุดค่าข้อมูล กล่าวคือ $V^{ac} = \{(ObjCode_1, dbVal_1), (ObjCode_2, dbVal_2), \ldots, (ObjCode_n, dbVal_n)\}$ โดยที่ n แทน จำนวนจุดแอทริบิวและจุดเอนทิตีที่ไม่ปรากฏจุดแอทริบิวของผลลัพธ์ในเคส และ $ObjCode_i$ แทนจุดแอทริบิว หรือจุดเอนทิตีของกราฟคำตอบ

E^{cc} คือ เซตของเส้นเชื่อมระหว่างเอนทิตีกับเอนทิตีในกราฟคำตอบ

3.1.3 คำพ้องความหมาย

คำพ้องความหมาย คือ คำย่อ คำหรือวลีที่ใช้แทนหรือมีความหมายสื่อถึงเอนทิตี แอทริบิว หรือค่าข้อมูล เช่น คำพ้องความหมาย "California" ใช้แทนค่าข้อมูล "CA" คำพ้องความหมาย "Writer" ใช้แทนเอนทิตี "Author" เป็นต้น คำพ้องความหมายมีรูปแบบการสื่อถึงเอนทิตี แอทริบิว หรือค่าข้อมูล ดังต่อไปนี้

คำพ้องความหมาย s สื่อถึง (sType, Object)

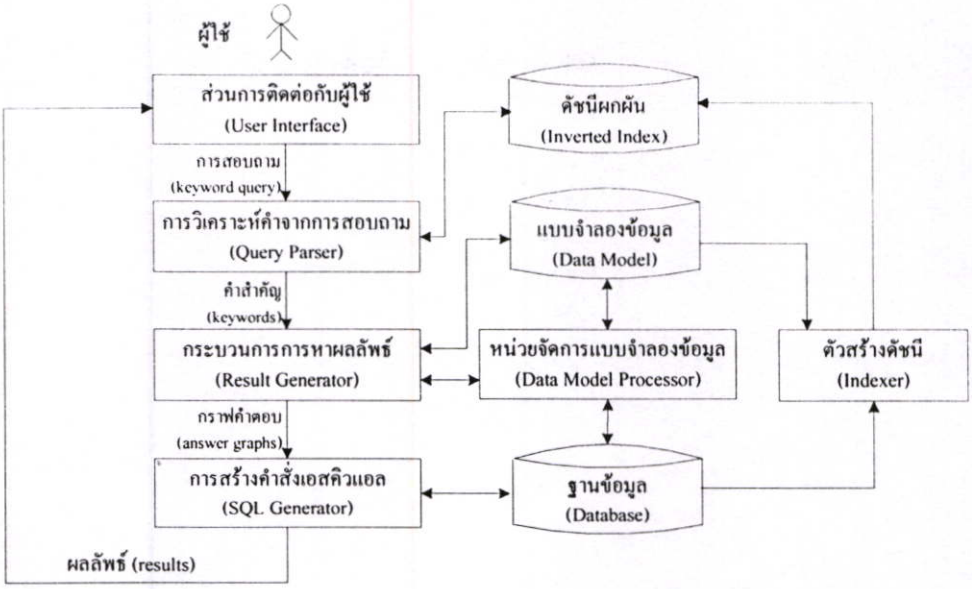
เมื่อ sType คือ ประเภทของคำพ้องความหมาย โดย sType $\in$ {A: Attribute, E: Entity, V: Value}

Object แทน คำที่คำพ้องความหมายสื่อถึง

นิยามที่ 2 เซตของคำพ้องความหมาย S_{i^t} คือ เซตของคำที่มีความหมายสื่อถึงคำดัชนี i^t เมื่อ t คือ ประเภทของคำดัชนี โดยที่ t = A E หรือ V ซึ่งใช้แทนแอทริบิว เอนทิตี และค่าข้อมูลตามลำดับ กล่าวคือ สมาชิกของ S_{i^t} คือ $\{s_1, s_2, ..., s_j\}$ เมื่อ j คือ จำนวนคำพ้องความหมายที่สื่อถึง i^t และแต่ละ s_j สื่อถึง i^t

3.2 สถาปัตยกรรมของระบบ

สถาปัตยกรรมของระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล เป็นดังรูปที่ 3.5 ส่วนประกอบที่สำคัญของระบบประกอบด้วยส่วนสำคัญ 5 ส่วน คือ หน่วยจัดการแบบจำลองข้อมูล (Data Model Processor) ตัวสร้างดัชนี (Indexer) ตัววิเคราะห์คำจากการสอบถาม (Query Parser) กระบวนการหาผลลัพธ์ (Result Generator) และตัวสร้างคำสั่งเอสคิวแอล (SQL Generator) ในแต่ละส่วนมีรายละเอียดดังต่อไปนี้



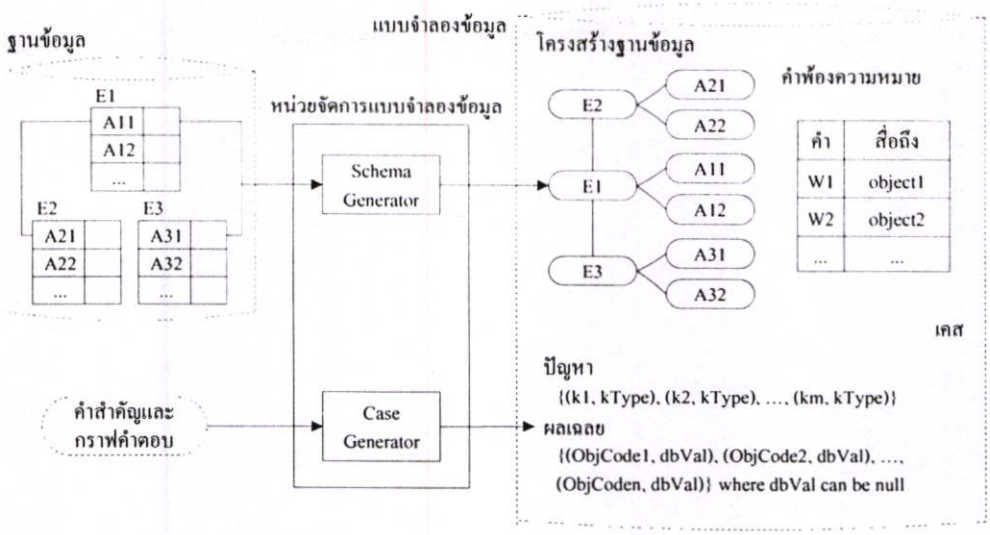
รูปที่ 3.5 สถาปัตยกรรมของระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

3.2.1 หน่วยจัดการแบบจำลองข้อมูล

หน่วยจัดการแบบจำลองข้อมูลมีหน้าที่จัดการข้อมูล 2 ส่วน คือ ข้อมูลโครงสร้างฐานข้อมูล และข้อมูลเกี่ยวกับเคส ดังรูปที่ 3.6

1. ตัวสร้างโครงสร้างฐานข้อมูล (Schema Generator) มีหน้าที่ทำการวิเคราะห์ฐานข้อมูล เพื่อนำข้อมูลฐานข้อมูลซึ่งได้แก่ ชื่อเอนทิตี ชื่อแอทริบิว ความสัมพันธ์ระหว่างเอนทิตีกับเอนทิตี และความสัมพันธ์ระหว่างเอนทิตีกับแอทริบิวสร้างเป็นแบบจำลองโครงสร้างฐานข้อมูลที่อยู่ในรูปของกราฟแบบไม่มีทิศทาง โดยกำหนดให้โหนดใช้แทนเอนทิตีและแอทริบิว ส่วนด้านใช้แทนความสัมพันธ์ระหว่างเอนทิตีกับเอนทิตี และความสัมพันธ์ระหว่างเอนทิตีกับแอทริบิว

2. ตัวสร้างเคส (Case Generator) มีหน้าที่สร้างเคสจากการสอบถามหนึ่งๆ ซึ่งประกอบด้วยปัญหาและผลเฉลย ในส่วนของปัญหาสร้างจากจุดที่ค้นพบของคำสำคัญที่ได้จากกราฟคำตอบ และผลเฉลยสร้างจากจุดแอทริบิวและจุดเอนทิตีที่ไม่ปรากฏจุดแอทริบิวในกราฟคำตอบ



รูปที่ 3.6 หน่วยจัดการแบบจำลองข้อมูล

3.2.2 ตัวสร้างดัชนี

ตัวสร้างดัชนีทำการสร้างดัชนีผกผันจากแบบจำลองข้อมูลและค่าข้อมูลจากฐานข้อมูล ในดัชนีผกผัน ประกอบด้วย คำดัชนี และรายการอ้างอิง ดังตัวอย่างจากตารางที่ 3.1 ซึ่งมีรายละเอียดดังต่อไปนี้

1. คำดัชนี คือ คำหรือวลี 4 ประเภท คือ ค่าข้อมูลจากฐานข้อมูล ข้อมูลจากแบบจำลองข้อมูล และคำที่ไม่มีนัยสำคัญ

1.1 ค่าข้อมูลในฐานข้อมูล คือ ค่าข้อมูลทั้งหมดจากฐานข้อมูล ทั้งข้อมูลประเภทอักขระ จำนวน และวันที่ โดยที่ค่าข้อมูลจากแอทริบิวเดียวกันจะไม่ซ้ำกัน ซึ่งตัวสร้างดัชนีจะทำการสร้างคำดัชนีประเภทนี้อัตโนมัติ

1.2 ข้อมูลจากแบบจำลองข้อมูล ได้แก่ ข้อมูลจากโครงสร้างฐานข้อมูล ข้อมูลจากเคส และข้อมูลจากคำพ้องความหมาย คำดัชนีที่มาจากโครงสร้างฐานข้อมูลประกอบด้วยชื่อเอนทิตีและชื่อแอทริบิว คำดัชนีที่มาจากเคสประกอบด้วยคำสำคัญจากปัญหาในแต่ละเคส ซึ่งตัวสร้างดัชนีจะทำการสร้างคำดัชนีจากแบบจำลองข้อมูลโดยอัตโนมัติ

1.3 คำที่ไม่มีนัยสำคัญ คือ คำที่ไม่ปรากฏอยู่ในคำดัชนีทั้ง 3 ประเภทดังกล่าว ซึ่งถือว่าเป็นคำที่ไม่มีนัยสำคัญในการค้นหา เนื่องจากผู้ใช้สามารถใช้ข้อความอิสระในการสอบถามได้ จึงอาจมีคำที่ไม่มีนัยสำคัญปรากฏอยู่ เช่น “give me detail of smith” คำที่ไม่มีนัยสำคัญ คือ “give” “me” “detail” และ “of” คำที่ไม่มีนัยสำคัญนี้ช่วยเพิ่มประสิทธิภาพในกระบวนการวิเคราะห์คำจากการสอบถามได้ ซึ่งคำดัชนีประเภทนี้ผู้ดูแลระบบจะเป็นผู้กำหนด

2. รายการอ้างอิง คือ การอ้างถึงตำแหน่งที่อยู่ของคำดัชนี สามารถอ้างอิงได้ดังนี้

รายการอ้างอิง = {(ประเภทของคำดัชนี, รหัสอ็อบเจกต์)}

โดยที่ ประเภทของคำดัชนี (Index Type) ประกอบด้วย A C E N S และ V

A (Attribute) หมายถึง คำดัชนีที่อ้างไปยังแอทริบิว

C (Case) หมายถึง คำดัชนีที่ปรากฏอยู่ในปัญหาของเคส

E (Entity) หมายถึง คำดัชนีที่อ้างไปยังเอนทิตี

N (Not use) หมายถึง คำดัชนีที่ไม่มีนัยสำคัญ

S (Synonym) หมายถึง คำดัชนีที่เป็นคำพ้องความหมาย

V (Value) หมายถึง คำดัชนีที่อ้างไปยังค่าข้อมูลในฐานข้อมูล

รหัสอ็อบเจกต์ (Object Code) ถูกกำหนดตามประเภทของคำดัชนี ดังนี้

A : กำหนดให้ใช้รหัสของแอทริบิว

C : กำหนดให้ใช้รหัสของเคส

E : กำหนดให้ใช้รหัสของเอนทิตี

N : กำหนดให้เป็นค่าว่าง

S : กำหนดให้ใช้รหัสของคำพ้องความหมาย

V : กำหนดให้ใช้รหัสของแอทริบิว

ตารางที่ 3.1 แสดงตัวอย่างดัชนีผกผัน

คำดัชนี	รายการอ้างอิง
w_1	{($iType_{11}$, $objCode_{11}$), ($iType_{12}$, $objCode_{12}$), …, ($iType_{1n}$, $objCode_{1n}$)}
w_2	{($iType_{21}$, $objCode_{21}$), ($iType_{22}$, $objCode_{22}$), …, ($iType_{2n}$, $objCode_{2n}$)}
…	
w_m	{($iType_{m1}$, $objCode_{m1}$), ($iType_{m2}$, $objCode_{m2}$), …, ($iType_{mn}$, $objCode_{mn}$)}

3.2.3 การวิเคราะห์คำจากการสอบถาม

การวิเคราะห์คำจากการสอบถามเป็นกระบวนการวิเคราะห์หาคำสำคัญและตำแหน่งของคำสำคัญในแบบจำลองข้อมูล โดยถือว่าคำหรือวลีจากการสอบถามที่สามารถจับคู่กับคำดัชนีในดัชนีผกผันได้คือคำสำคัญ ดังนิยามต่อไปนี้

นิยามที่ 3 กำหนดให้การสอบถาม Q (Query) เป็นการสอบถามด้วยข้อความอิสระ ที่มีการแบ่งคำด้วยอักขระว่าง โดยที่

$Q = w_1, w_2, \ldots, w_m$ เมื่อ w คือ คำจากการสอบถาม และ m คือ จำนวนคำจากการสอบถาม

นิยามที่ 4 คำสำคัญ K (Keyword) ของ Q คือ ส่วนของ Q ที่สามารถจับคู่กับคำดัชนีจากดัชนีผกผันได้ โดยที่

$K = \{ k_1, k_2, \ldots, k_n \}$ เมื่อ n คือ จำนวนคำสำคัญที่ค้นพบ โดยที่ $n \leq m$

นิยามที่ 5 แต่ละ k_i เมื่อ i = 1, 2, ..., n ประกอบด้วยรายการอ้างอิง L_{k_i} (Reference List) ดังที่กล่าวไว้ในส่วนที่ 3.2.2

3.2.4 กระบวนการหาผลลัพธ์

การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลอาศัยแบบจำลองข้อมูลดังที่กล่าวไว้ในส่วนที่ 3.1 เพื่อความเข้าใจในการอ้างถึงคำศัพท์ต่างๆ ในกระบวนการหาผลลัพธ์จึงได้นิยามคำศัพท์ไว้ ดังต่อไปนี้

นิยามที่ 6 กราฟเสมือน G (Virtual Graph) คือ กราฟโครงสร้างฐานข้อมูล SG ที่มีการเชื่อมต่อจุดแอทริบิวด้วยจุดค่าข้อมูล กล่าวคือ กราฟ G คือกราฟ <V, E> โดยที่ V คือเซตของจุด 3 ประเภท คือ เอนทิตี แอทริบิว และค่าข้อมูล ส่วน E คือ เซตของข้อบังคับในการเชื่อมต่อระหว่างจุด 3 ประเภท ซึ่งได้แก่ การเชื่อมต่อระหว่างเอนทิตีกับเอนทิตี (E^{ee}) ระหว่างแอทริบิวกับเอนทิตี (E^{ae}) และระหว่างแอทริบิวกับค่าข้อมูล (E^{va})

นิยามที่ 7 เซตของจุดที่ค้นพบ (Searchable Vertex Set) ของคำสำคัญ k_i หรือ V_{k_i} คือ เซตของจุดในกราฟ G ที่มีค่าสอดคล้องกับ k_i หรือมีคำพ้องความหมายสอดคล้องกับ k_i กล่าวคือ สมาชิกของ V_{k_i} คือ จุดที่ค้นพบ $v^t_{k_i}$ เมื่อ t = E A หรือ V โดยที่ t คือ ประเภทของจุดซึ่งแทนด้วยเอนทิตี แอทริบิว และค่าข้อมูลตามลำดับ

นิยามที่ 8 กำหนดให้จำนวนคำสำคัญเท่ากับ n เมื่อ $n \neq 0$ รูปจำลองการสอบถาม (Query Image) คือ เซตของจุดที่ค้นพบ v^t_i เมื่อ i = 1, 2, ..., n และ $v^t_i \in v^t_{k_i}$ ตามลำดับ

นิยามที่ 9 เส้นทางพื้นฐาน P_i (Basic Path) ของ v_i คือ เซตของจุดที่น้อยที่สุดในกราฟ G ที่ประกอบด้วย v_i และจุดเอนทิตี โดยที่จุดเอนทิตีนั้นสามารถเชื่อมต่อกับ v_i นั้นได้

นิยามที่ 10 เซตของจุดระหว่างกลาง V^C (Intermediate Vertex) คือ เซตของจุดที่เชื่อมระหว่างเส้นทางพื้นฐาน P_i กับ P_j โดยที่ $i \neq j$ กล่าวคือ $v^C \in G$ และ $v^C \notin P$ และมี E^{∞} เชื่อมระหว่าง v^C กับ v^E โดยที่ $v^E \in P$

นิยามที่ 11 กราฟที่เป็นไปได้ FG (Feasible Graph) ของรูปจำลองการสอบถามที่มีจำนวนคำสำคัญที่ค้นพบเท่ากับ n เมื่อ $n \neq 0$ คือ เซตของจุดในเส้นทางพื้นฐาน $P_1 \cup P_2 \cup K \cup P_n$ และจุดระหว่างกลาง โดยที่ P_i คือ เส้นทางพื้นฐานที่ถูกกำหนดโดยจุดที่ค้นพบ v_i ของรูปจำลองการสอบถาม

นิยามที่ 12 กราฟคำตอบ (Answer Graph) คือ กราฟที่ได้จากกราฟ FG โดยที่ $FG \subset G$ ซึ่ง FG ประกอบด้วย <V, E> เมื่อ V คือ เซตของจุดเอนทิตี จุดแอทริบิว และจุดค่าข้อมูล และ E คือ เซตของเส้นเชื่อมต่อระหว่างจุดดังกล่าว

ในกระบวนการหาผลลัพธ์แบ่งเป็น 2 กรณี คือ การหาผลลัพธ์โดยพิจารณาจากเคส และการหาผลลัพธ์จากการค้นข้อมูลฐานข้อมูล ในการหาผลลัพธ์โดยพิจารณาจากเคสระบบทำการเปรียบเทียบเซตของคำสำคัญที่ได้กับปัญหาของเคส หากเซตของคำสำคัญสอดคล้องกับปัญหาของเคส ผลลัพธ์ที่ได้คือผลเฉลยของเคส หรือการประยุกต์ใช้ผลเฉลยของเคสนั่นเอง ในกรณีที่เป็นการหาผลลัพธ์จากการค้นข้อมูลฐานข้อมูล ระบบทำการค้นหาเซตของจุดที่ค้นพบ (นิยามที่ 7) จากกราฟ G (นิยามที่ 6) และกำหนดรูปจำลองการสอบถาม (นิยามที่ 8) ทั้งหมด จากนั้นจึงพิจารณาหากราฟที่เป็นไปได้ (นิยามที่ 11) จากรูปจำลองการสอบถามโดยอาศัยเส้นทางพื้นฐาน (นิยามที่ 9) ซึ่งกราฟคำตอบ (นิยามที่ 12) คือ กราฟที่ได้จากกราฟที่เป็นไปได้จำนวน r กราฟ โดยเรียงลำดับความสำคัญของกราฟตามต้นทุนของกราฟ กราฟคำตอบที่ได้ถูกแปลงให้อยู่ในรูปภาษาเอสคิวแอลและนำเสนอต่อผู้ใช้ต่อไป ซึ่งหากผู้ใช้ยืนยันคำตอบที่ถูกต้องของผลลัพธ์และต้องการเก็บผลลัพธ์นั้นเป็นเคส ระบบจะนำเซตของคำสำคัญและกราฟคำตอบที่ได้เก็บให้อยู่ในรูปของเคสดังกล่าวข้างต้น

3.2.5 การสร้างคำสั่งเอสคิวแอล

จากกราฟคำตอบสามารถสร้างคำสั่งเอสคิวแอลได้ด้วยการพิจารณาความสัมพันธ์ระหว่างข้อมูลในกราฟ และเงื่อนไขที่ได้จากการสอบถาม ซึ่งได้แก่ รายการเอนทิตี (Entity List) รายการ

แอทริบิว (Attribute List) เงื่อนไขการจอย (Join Condition) และเงื่อนไขการเลือก (Selection Condition)

1. รายการเอนทิตี คือ เอนทิตีที่ปรากฏอยู่ในกราฟคำตอบทั้งหมด เป็นรายการที่บ่งบอกถึงระเบียนผลลัพธ์มาจากเอนทิตีใด

2. รายการแอทริบิว คือ แอทริบิวทุกแอทริบิวที่ปรากฏในกราฟคำตอบ และแอทริบิวจากเอนทิตีที่ไม่ปรากฏจุดแอทริบิวในกราฟคำตอบ

3. เงื่อนไขการจอย คือ การกำหนดความสัมพันธ์ระหว่างเอนทิตีโดยอาศัยความสัมพันธ์ของคีย์หลักและคีย์นอกของรายการเอนทิตี

4. เงื่อนไขการเลือก คือ การกำหนดระเบียนข้อมูล ในกรณีที่เงื่อนไขการเลือกมีหลายเงื่อนไข ซึ่งเป็นเงื่อนไขที่มีแอทริบิวต่างกัน ให้เชื่อมเงื่อนไขนั้นด้วยตัวดำเนินการ “AND” แต่ถ้าเป็นเงื่อนไขที่มีแอทริบิวเดียวกัน ให้เชื่อมเงื่อนไขนั้นด้วยตัวดำเนินการ “OR” ดังตัวอย่างจากตารางที่ 3.2

ตารางที่ 3.2 แสดงตัวอย่างการเชื่อมเงื่อนไขของเงื่อนไขการเลือก

เงื่อนไขการเลือก	การเชื่อมเงื่อนไข
{PUBLICATION.TYPE = ‘misc’, PUBLICATION.YEAR = 1999}	WHERE PUBLICATION.TYPE = ‘misc’ AND PUBLICATION.YEAR = 1999
{ PUBLICATION.TYPE = ‘misc’, PUBLICATION.TYPE = ‘article’}	WHERE PUBLICATION.TYPE = ‘misc’ OR PUBLICATION.TYPE = ‘article’

3.3 อัลกอริทึมที่ใช้ในการหาผลลัพธ์

กำหนดให้การสอบถาม Q = $w_1, w_2, \ldots, w_m$ เมื่อ w คือ คำจากการสอบถาม และ m คือ จำนวนคำจากการสอบถาม อัลกอริทึมที่ใช้ในการหาผลลัพธ์มีดังต่อไปนี้

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

1.1 หาเซตของคำสำคัญ K จาก Q โดยการค้นหาจากดัชนีผกผัน โดยที่คำสำคัญอาจเป็นคำ หรือวลีจาก Q ที่มีค่าตรงกับคำดัชนีหรือส่วนหนึ่งของคำดัชนี ดังนั้น

$K = \{ k_1, k_2, \ldots, k_n \}$ เมื่อ n คือ จำนวนคำสำคัญที่ค้นพบ โดยที่ $n \leq m$

1.2 หาเซตของจุดที่ค้นพบ V_{k_i} ของคำสำคัญ k_i เมื่อ i = 1, 2, ..., n ถ้า $k_i \in$ Synonym แล้ว $V_{k_i} \in G$ ดังนั้นเซตทั้งหมดของจุดที่ค้นพบ คือ $V_K = \{ V_{k_1}, V_{k_2}, ..., V_{k_n} \}$

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

ทำการกำหนดรูปจำลองการสอบถาม QI โดย QI_j คือ เซตของจุดที่ค้นพบในแต่ละ k_i กล่าวคือ $QI_j = \{ v_{k_1,j} \in V_{k_1}, v_{k_2,j} \in V_{k_2}, K, v_{k_i,j} \in V_{k_i} \}$ เมื่อ i = 1, 2, ..., n และ j = 1, 2, ..., m โดยที่ n คือ จำนวนคำสำคัญ และ m คือ จำนวนรูปจำลองการสอบถาม ซึ่ง $m = | V_{k_1} | | V_{k_2} | ... | V_{k_i} |$

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

หาเส้นทางพื้นฐาน $P_{i,j}$ ของ $v_{i,j}$ ในแต่ละรูปจำลองการสอบถาม เมื่อ i = 1, 2, ..., n และ j = 1, 2, ..., m ดังเงื่อนไขต่อไปนี้

3.1 ถ้า $v_{i,j} = v^D_{i,j}$ แล้ว $P_{i,j} = \{(v^E, v^A, v^D_{i,j}) | \ v^E, v^A \in G$ และมี $E^{v^D_{i,j} v^A}$, $E^{v^A v^E} \in G\}$

3.2 ถ้า $v_{i,j} = v^A_{i,j}$ แล้ว $P_{i,j} = \{(v^E, v^A) | \ v^E \in G$ และมี $E^{v^A_{i,j} v^E} \in G\}$

3.3 ถ้า $v_{i,j} = v^E_{i,j}$ แล้ว $P_{i,j} = \{ v^E_{i,j} \}$

ขั้นตอนที่ 4 หากราฟคำตอบ

4.1 พิจารณาหากราฟที่เป็นไปได้ FG_j จากแต่ละ QI_j ซึ่ง FG_j เป็นกราฟย่อยของกราฟ G โดยที่ FG_j ประกอบด้วยเซตของจุดใน $P_{i,j}$ เมื่อ i = 1, 2, ..., n และ v^C ซึ่ง $v^C \in G$ และ v^C เป็นจุดระหว่างกลาง

4.2 พิจารณากราฟคำตอบ โดยการเรียงลำดับกราฟคำตอบจากต้นทุนของกราฟ กล่าวคือ cost(FG_j) = (number of vertices of FG_j) – 1 กราฟที่มีต้นทุนต่ำที่สุดคือกราฟคำตอบที่มีนัยสำคัญมากที่สุด

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

SELECT [Attribute List]

FROM [Entity List]

WHERE [Join Condition] [AND] [Selection Condition]

3.4 การจัดการเคส

การจัดการเคส ประกอบด้วย กระบวนการในการสร้างเคส กระบวนการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส และกระบวนการรียูสเคส

3.4.1 กระบวนการในการสร้างเคส

เมื่อผู้ใช้ยืนยันความถูกต้องของผลลัพธ์ และต้องการเก็บเคสนั้นเพื่อการค้นหาครั้งต่อไป ระบบ หรือตัวสร้างเคสจะทำการสร้างเคสจากขั้นตอน ดังต่อไปนี้

ขั้นตอนที่ 1 กำหนดปัญหา

ให้ Problem = $\{(v_{k_1}, k_1\text{Type}), (v_{k_2}, k_2\text{Type}), \ldots, (v_{k_n}, k_n\text{Type})\}$ เมื่อ v_{k_i} คือ เซตของจุดที่ค้นพบคำสำคัญ k_i และ $v_{k_i} \in FG$ โดยที่ n คือจำนวนคำสำคัญ และ $k_i\text{Type} \in \{A, E, V\}$

ขั้นตอนที่ 2 กำหนดผลเฉลย

2.1 กำหนดจุดของกราฟคำตอบให้เป็นจุดผลเฉลย โดยที่

สำหรับแต่ละ v_i^t โดยที่ $v_i^t \in FG$ และ $t \in \{A, E\}$

ถ้า v_i^t โดยที่ $t = A$ แล้ว $V^{ac} \mathrel{+}= \{(v_i^A, v_j^D)\}$ เมื่อ $v_i^D \in FG$ และ v_i^D มีเส้นเชื่อมไปยัง v_i^A หรือ $V^{ac} \mathrel{+}= \{(v_i^A, -)\}$

มิฉะนั้น ถ้า v_i^t โดยที่ $t = E$ และ v_i^t ไม่ปรากฏ v_i^A กำหนดให้ $V^{ac} \mathrel{+}= \{(v_i^E, -)\}$

2.2 กำหนดเส้นเชื่อมระหว่างเอนทิตีกับเอนทิตีให้เป็นผลเฉลย โดยที่ สำหรับแต่ละ e_i^{cc} โดยที่ $e_i^{cc} \in FG$ กำหนดให้ $E^{cc} \mathrel{+}= \{e_i^{cc}\}$

3.4.2 กระบวนการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส

วิธีการพิจารณาเคสที่เหมาะสมอาศัยค่าความคล้ายคลึงระหว่างคำสำคัญที่ได้จากการสอบถาม และปัญหาของเคส โดยใช้ค่าเฉลี่ยน้ำหนัก [4] ซึ่งมีวิธีการ ดังต่อไปนี้

ขั้นตอนที่ 1 พิจารณาหาเคสที่อาจเป็นไปได้

พิจารณารายการอ้างอิงของแต่ละ k_i เมื่อ $i = 1, 2, \ldots, n$ เมื่อ n คือจำนวนคำสำคัญ ถ้า $iType_{k_i} =$ C แล้ว เซตของเคสที่อาจเป็นไปได้สามารถพิจารณาได้จาก

$\{ objCode_1 \cap objCode_2 \cap K \cap objCode_m \}$ โดยที่ เซตดังกล่าวไม่เท่ากับเซตว่าง เมื่อ $objCode_i$ แทนรายการอ้างอิง และ m แทนจำนวนรายการอ้างอิงที่มากที่สุดที่สามารถอินเตอร์เซกชันกันได้ ซึ่ง $m \leq n$

ขั้นตอนที่ 2 พิจารณาหาเคสที่เหมาะสม

2.1 หาค่าความคล้ายคลึงระหว่างการสอบถามกับเคส โดยพิจารณาจากเซตของจุดที่ค้นพบ V_k กับปัญหาของแต่ละ C_i โดยที่

$$\text{ค่าความคล้ายคลึง}\ (V_k, C_i) = \frac{\text{จำนวนจุดเทียบเท่า}}{\text{จำนวนจุดพื้นฐาน}}$$

เมื่อ จำนวนจุดเทียบเท่า คือ จำนวนจุดจาก V_k ที่ตรงกับจุดจาก C_i

จำนวนจุดพื้นฐาน คือ จำนวนจุดที่ได้จาก $V_k \cup C_i$

2.2 เคสที่ให้ค่าความคล้ายคลึงมากที่สุด คือ เคสที่เหมาะสม

ค่าความคล้ายคลึงมีค่าตั้งแต่ 0-1 ถ้าค่าความคล้ายคลึงมีค่าเท่ากับ 1 แสดงว่าผลเฉลยของเคส คือ ผลลัพธ์ที่ได้จากการค้นหา ถ้าค่าความคล้ายคลึงมีค่ามากกว่าหรือเท่ากับ θ แสดงว่าผลเฉลยของเคสสามารถนำมารียูสได้ โดยค่า θ คือ ค่าการตัดสินใจที่ระบบยอมรับได้สำหรับการนำเคสมารียูส

3.4.3 กระบวนการรียูสเคส

การรียูสเคส คือ การปรับปรุงผลเฉลยของเคสเพื่อนำมาเป็นส่วนหนึ่งกราฟคำตอบ ซึ่งสามารถแบ่งเป็น 2 กรณี คือ กรณีแรก คือ การรียูสเคสเมื่อจุดแอทริบิวจากเป็นเซตย่อยของผลเฉลยจากเคส และกรณีที่สอง คือ การรียูสเคสเมื่อผลเฉลยจากเคสเป็นเซตย่อยของจุดที่ค้นพบ การรียูสเคสทั้งสองกรณีมีขั้นตอน ดังต่อไปนี้

1. การรียูสเคสเมื่อจุดที่ค้นพบเป็นเซตย่อยจุดจากเคส สามารถทำได้โดยกำหนดจุดผลเฉลยที่ได้จาก $V_k \cap C_i$ และทำการหากราฟที่เป็นไปได้จากผลเฉลยภายในเคส
2. การรียูสเคสเมื่อผลเฉลยจากเคสเป็นเซตย่อยของจุดที่ค้นพบ สามารถทำได้โดยหาเส้นทางพื้นฐานของจุด $V_k - (V_k \cap C_i)$ แล้วนำเส้นทางพื้นที่ได้มาเชื่อมกับผลเฉลยของเคส C_i

บทที่ 4
การประเมินผล

บทนี้เป็นการยกตัวอย่างการสอบถาม โดยใช้แบบจำลองข้อมูลและวิธีการค้นหาตามที่ได้ออกแบบไว้ เพื่อประเมินผลแบบจำลองข้อมูล และวิธีการค้นหาคำสำคัญดังกล่าว โดยในส่วนที่ 4.1 กล่าวถึง วิธีการประเมินผลของระบบ ส่วนที่ 4.2 กล่าวถึงตัวอย่างฐานข้อมูลที่ใช้ในการยกตัวอย่างการสอบถาม ส่วนที่ 4.3 กล่าวถึงตัวอย่างการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล โดยการสอบถามด้วยข้อความอิสระในรูปแบบต่างๆ ส่วนที่ 4.4 กล่าวถึงตัวอย่างการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส และสุดท้าย ส่วนที่ 4.5 สรุปการประเมินผล

4.1 วิธีการประเมินผลของระบบ

การประเมินผลของระบบพิจารณาจากคำสำคัญที่ใช้ในการสอบถาม แบ่งเป็น 3 ประเภท คือ การสอบถามด้วยคำสำคัญที่ตรงกับค่าข้อมูล การสอบถามด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูล และการสอบถามด้วยคำสำคัญที่ตรงกับคำพ้องความหมาย

เพื่อประเมินผลการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลว่าระบบสามารถค้นหาคำสำคัญที่ตรงกับค่าข้อมูล ทั้งข้อมูลประเภทอักขระ จำนวน และวันที่ได้ วิธีการประเมินผลทำได้โดยการยกตัวอย่างการสอบถามโดยใช้ข้อความที่ประกอบด้วยคำสำคัญที่ตรงกับค่าข้อมูลประเภทดังกล่าว เช่นเดียวกับการประเมินผลการสอบถามด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลและคำพ้องความหมาย โดยยกตัวอย่างการสอบถามโดยใช้ข้อความที่ประกอบด้วยคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลและคำพ้องความหมาย แล้วนำผลลัพธ์ที่ได้เปรียบเทียบกับผลลัพธ์ที่ได้จากระบบการค้นหาคำสำคัญระบบต่างๆ โดยแบ่งระบบการค้นหาคำสำคัญเป็น 2 กลุ่ม ตามประเภทของคำดัชนีที่ใช้ในระบบ ดังที่กล่าวไว้ในบทที่ 2 คือ คำดัชนีที่ตรงกับค่าข้อมูล และคำดัชนีที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล

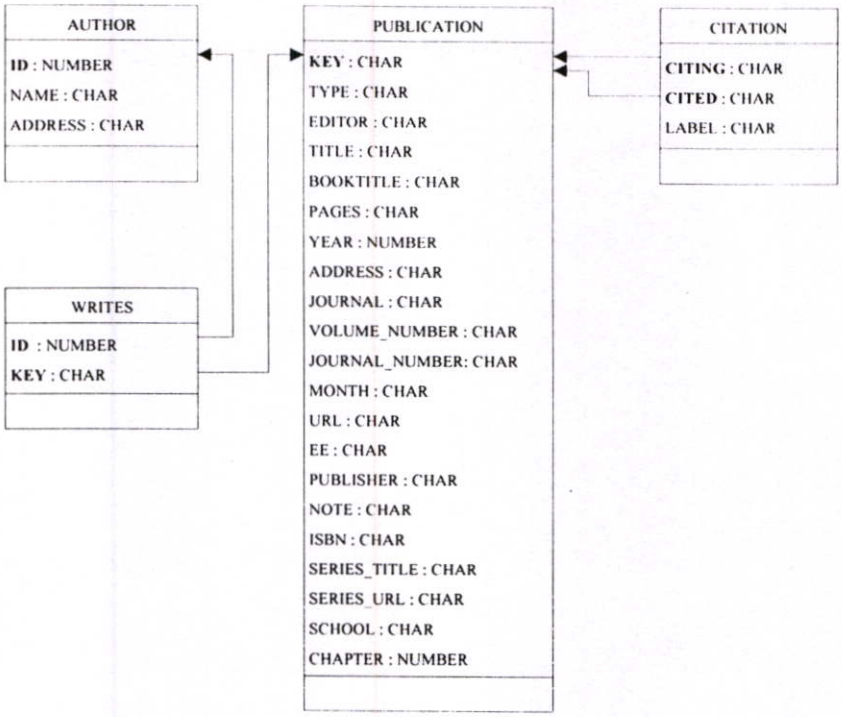
1. ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล ในที่นี้ประกอบด้วย ดิสโคเวอร์ ดีบีเอกซ์โพลเรอ และอีเคเอสโอ
2. ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล ในที่นี้ประกอบด้วย บีเอ เอ็นเคเอส และมราจียาติ

เพื่อให้ผลลัพธ์ที่ได้จากการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ อยู่ในรูปแบบเดียวกัน จึงกำหนดให้ผลลัพธ์ดังกล่าวอยู่ในรูปของภาษาเอสคิวแอล โดยผลลัพธ์ที่ได้จากระบบดิสโคเวอร์ ระบบดีบีเอกซ์โพลเรอ ระบบอีเคเอสโอ และระบบมราจียาติได้จากการพิจารณาหาผลลัพธ์ตามขั้นตอนของระบบดังกล่าว ส่วนผลลัพธ์ที่ได้จากระบบบีเอเอ็นเคเอสได้จากโปรแกรมต้นแบบ [15] ร่วมกับการพิจารณาตามขั้นตอนวิธีของระบบเพื่อแปลงผลลัพธ์จากโปรแกรมต้นแบบให้อยู่ในรูปภาษาเอสคิวแอล ในกรณีที่แต่ละการสอบถามประกอบด้วยผลลัพธ์หลายผลลัพธ์ ผู้ประเมินจะทำการเลือกผลลัพธ์ที่ดีที่สุด โดยพิจารณาจากวิธีการจัดลำดับความสำคัญของผลลัพธ์ของแต่ละระบบ

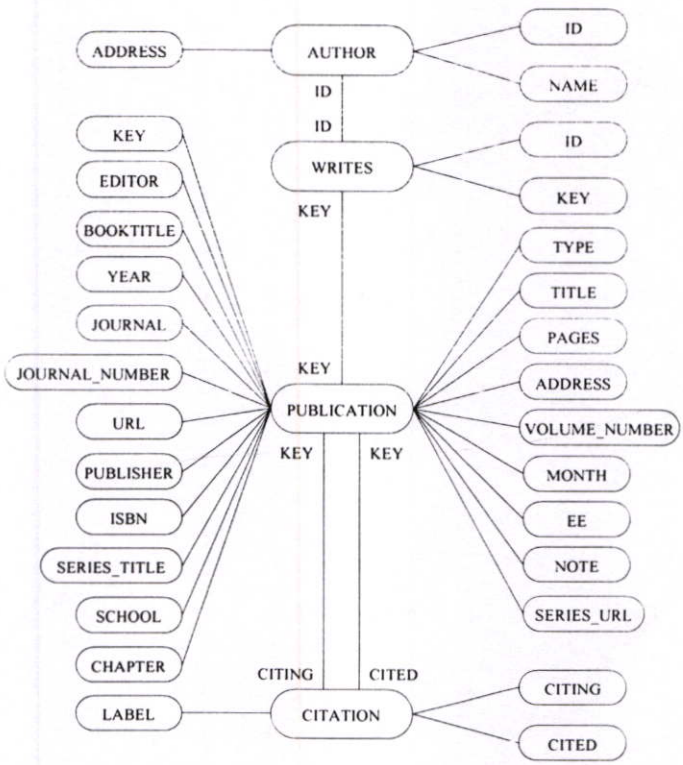
4.2 ตัวอย่างฐานข้อมูลที่ใช้ในระบบ

ระบบใช้ตัวอย่างข้อมูลจากฐานข้อมูลดีบีแอลพี (DBLP: Digital Bibliography & Library Project) ซึ่งเป็นฐานข้อมูลที่ใช้สำหรับเก็บข้อมูลบรรณานุกรมของบทความทางวิทยาการคอมพิวเตอร์ โครงสร้างฐานข้อมูลดีบีแอลพีเป็นดังรูปที่ 4.1 ซึ่งประกอบด้วยเอนทิตี 4 เอนทิตี คือ Publication Citation Author และ Writes เอนทิตี Publication เก็บข้อมูลรายละเอียดของวารสารและบทความต่างๆ เอนทิตี Citation เก็บข้อมูลการอ้างอิงระหว่างบทความต่างๆ เอนทิตี Author เก็บข้อมูลรายชื่อผู้เขียน และเอนทิตี Writes เก็บข้อมูลการเขียนบทความของผู้เขียน

จากโครงสร้างฐานข้อมูลในรูปที่ 4.1 กราฟโครงสร้างฐานข้อมูลในแบบจำลองข้อมูลเป็นดังรูปที่ 4.2 ซึ่งประกอบด้วยจุดเอนทิตี จุดแอทริบิว และเส้นเชื่อมแสดงความสัมพันธ์ระหว่างเอนทิตีกับเอนทิตี และแอทริบิวกับเอนทิตี



รูปที่ 4.1 โครงสร้างฐานข้อมูลดีบีแอลพี



รูปที่ 4.2 กราฟโครงสร้างฐานข้อมูลดีบีแอลพีในแบบจำลองข้อมูล

4.3 ตัวอย่างการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล

ตัวอย่างการสอบถามข้อมูลจากฐานข้อมูลดีบีแอลพี และตัวอย่างการสร้างเคสจากผลลัพธ์ที่ได้ โดยมีตัวอย่างค่าข้อมูลดังภาคผนวก ข และคำพ้องความหมายในภาคผนวก ค เป็นดังต่อไปนี้

ตัวอย่างที่ 1 เมื่อสอบถามด้วย "Get address of Jason Rennie"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.1 เมื่อจำนวนคำสำคัญ n = 2

ตารางที่ 4.1 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Get address of Jason Rennie"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
address	{ ADDRESS| (A, A003), ADDRESS| (A, P008)}
Jason Rennie	{ Jason Rennie | (V, A002)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

จำนวนของรูปจำลองการสอบถาม คือ 2

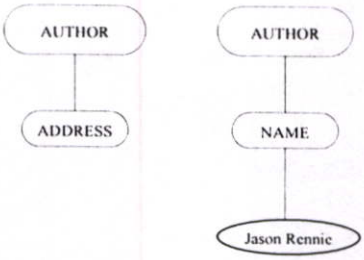
รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {ADDRESS| (A, A003), Jason Rennie | (V, A002)}

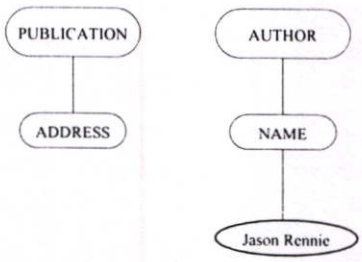
QI_2 = {ADDRESS| (A, P008), Jason Rennie | (V, A002)}

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.3 และเส้นทางพื้นฐานของ QI_2 เป็นดังรูปที่ 4.4



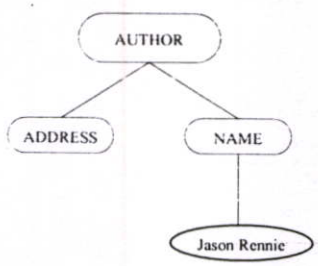
รูปที่ 4.3 เส้นทางพื้นฐานของ QI_1 = {ADDRESS| (A, A003), Jason Rennie | (V, A002)}



รูปที่ 4.4 เส้นทางพื้นฐานของ QI_2 = {ADDRESS| (A, P008), Jason Rennie | (V, A002)}

ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_1 ของ QI_1 เป็นดังรูปที่ 4.5 ซึ่งมีต้นทุนของกราฟเท่ากับ 3 และกราฟที่เป็นไปได้ FG_2 ของ QI_2 เป็นดังรูปที่ 4.6 และมีต้นทุนของกราฟเท่ากับ 5 ดังนั้นกราฟคำตอบ คือ FG_1 และ FG_2 โดยถือว่ากราฟคำตอบ FG_1 เป็นกราฟคำตอบที่มีนัยสำคัญมากกว่า FG_2



รูปที่ 4.5 กราฟที่เป็นไปได้ FG_1 ของ QI_1

รูปที่ 4.6 กราฟที่เป็นไปได้ FG_2 ของ QI_2

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_1 ซึ่งประกอบด้วย

รายการเอนทิตี คือ AUTHOR

รายการแอทริบิว คือ AUTHOR.ADDRESS, AUTHOR.NAME

เงื่อนไขการจอย คือ -

เงื่อนไขการเลือก คือ AUTHOR.NAME = 'Jason Rennie'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT AUTHOR.ADDRESS, AUTHOR.NAME
FROM AUTHOR
WHERE (AUTHOR.NAME = 'Jason Rennie')
```

จากกราฟที่เป็นไปได้ FG_2 ซึ่งประกอบด้วย

รายการเอนทิตี คือ PUBLICATION, WRITES, AUTHOR

รายการแอทริบิว คือ PUBLICATION.ADDRESS, WRITES.*,
AUTHOR.NAME

เงื่อนไขการจอย คือ PUBLICATION.KEY = WRITES.KEY and
WRITES.ID = AUTHOR.ID

เงื่อนไขการเลือก คือ AUTHOR.NAME = 'Jason Rennie'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT PUBLICATION.ADDRESS, WRITES.*, AUTHOR.NAME
FROM PUBLICATION, WRITES, AUTHOR
WHERE (PUBLICATION.KEY = WRITES.KEY and
        WRITES.ID = AUTHOR.ID) and
```

(AUTHOR.NAME = 'Jason Rennie')

ขั้นตอนที่ 6 การสร้างเคส

สมมติให้ผู้ใช้เลือกกราฟคำตอบในรูปที่ 4.5 เป็นผลลัพธ์ที่เหมาะสม เคสที่ได้จากกราฟคำตอบดังกล่าวเป็นดังนี้

Problem = {(ADDRESS| A003, A), (Jason Rennie| A002, V)}

Solution = $(V^{ae}_{Sol}, E^{ee}_{Sol})$

V^{ae}_{Sol} = {(A003, -), (A002, Jason Rennie)}

E^{ee}_{Sol} = {}

<u>ตัวอย่างที่ 2</u> เมื่อสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.2 เมื่อจำนวนคำสำคัญ n = 2

ตารางที่ 4.2 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
David Zuckerman	{David Zuckerman| (V, A002)}
Russell Impagliazzo	{Russell Impagliazzo| (V, A002)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

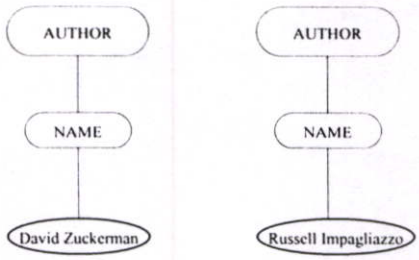
จำนวนของรูปจำลองการสอบถาม คือ 1

รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {David Zuckerman| (V, A002), Russell Impagliazzo| (V, A002)}

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

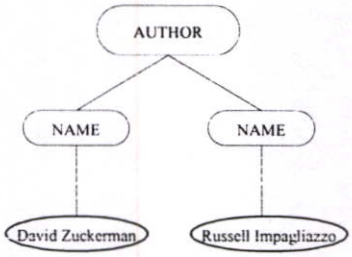
เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.7



รูปที่ 4.7 เส้นทางพื้นฐานของ QI_1 = {David Zuckerman| (V, A002), Russell Impagliazzo| (V, A002)}

ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_1 ของ QI_1 เป็นดังรูปที่ 4.8 ซึ่งมีต้นทุนของกราฟเท่ากับ 4



รูปที่ 4.8 กราฟที่เป็นไปได้ FG_1 ของ QI_1

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_1 ซึ่งประกอบด้วย

รายการเอนทิตี คือ AUTHOR

รายการแอทริบิว คือ AUTHOR.ID, AUTHOR.ADDRESS, AUTHOR.NAME

เงื่อนไขการจอย คือ -

เงื่อนไขการเลือก คือ AUTHOR.NAME = 'David Zuckerman' or AUTHOR.NAME = 'Russell Impagliazzo'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT AUTHOR.ID, AUTHOR.ADDRESS, AUTHOR.NAME
FROM AUTHOR
WHERE (AUTHOR.NAME = 'David Zuckerman' or
```

AUTHOR.NAME = 'Russell Impagliazzo')

ขั้นตอนที่ 6 การสร้างเคส

จากกราฟคำตอบในรูปที่ 4.8 สามารถสร้างเคสได้ดังนี้

Problem = {(David Zuckerman| A002, V), (Russell Impagliazzo| A002, V)}

Solution = $(V^{ae}_{Sol}, E^{ee}_{Sol})$

V^{ae}_{Sol} = {(A002, David Zuckerman), (A002, Russell Impagliazzo)}

E^{ee}_{Sol} = {}

<u>ตัวอย่างที่ 3</u> เมื่อสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.3 เมื่อจำนวนคำสำคัญ n = 4

ตารางที่ 4.3 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
author	{AUTHOR| (E, AUTHOR)}
name	{NAME| (A, A002)}
write	{WRITES| (E, WRITES)}
The VOCAL Test Methodology	{The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

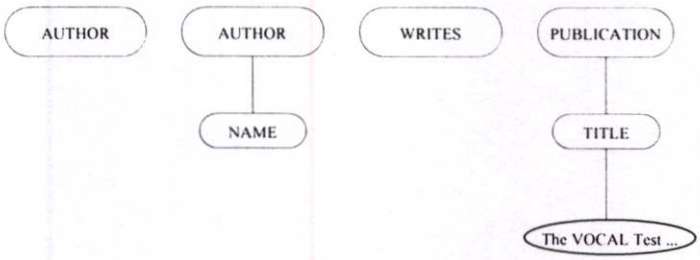
จำนวนของรูปจำลองการสอบถาม คือ 1

รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {AUTHOR| (E, AUTHOR), NAME| (A, A002), WRITES| (E, WRITES), The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

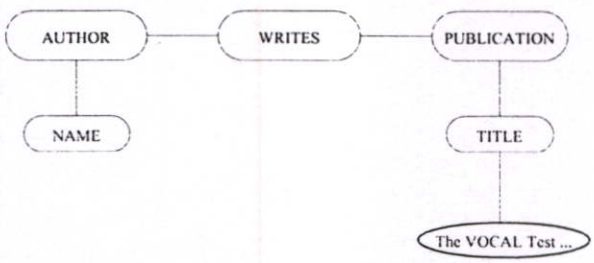
เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.9



รูปที่ 4.9 เส้นทางพื้นฐานของ QI_1 = {AUTHOR| (E, AUTHOR), NAME| (A, A002), WRITES| (E, WRITES), The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_1 ของ QI_1 เป็นดังรูปที่ 4.10 ซึ่งมีต้นทุนของกราฟเท่ากับ 5



รูปที่ 4.10 กราฟที่เป็นไปได้ FG_1 ของ QI_1

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_1 ซึ่งประกอบด้วย

รายการเอนทิตี คือ AUTHOR, WRITES, PUBLICATION

รายการแอทริบิว คือ AUTHOR.NAME, WRITES.*,
PUBLICATION.TITLE

เงื่อนไขการจอย คือ AUTHOR.ID = WRITES.ID and
WRITES.KEY = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT AUTHOR.NAME, WRITES.*, PUBLICATION.TITLE
FROM AUTHOR, WRITES, PUBLICATION
WHERE (AUTHOR.ID = WRITES.ID and
        WRITES.KEY = PUBLICATION.KEY) and
        (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

ขั้นตอนที่ 6 การสร้างเคส

จากกราฟคำตอบในรูปที่ 4.8 สามารถสร้างเคสได้ดังนี้

Problem = {(AUTHOR| AUTHOR, E), (NAME| A002, A), (WRITES| WRITES, E), (The VOCAL Test Methodology| P004, V)}

Solution $= (V^{ae}_{Sol}, E^{ee}_{Sol})$

V^{ae}_{Sol} = {(A002, -), (WRITES, -), (P004, The VOCAL Test Methodology)}

E^{ee}_{Sol} = {AWID, PWKey}

<u>ตัวอย่างที่ 4</u> เมื่อสอบถามด้วย "Give me references of The VOCAL Test Methodology"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.4 เมื่อจำนวนคำสำคัญ n = 2

ตารางที่ 4.4 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Give me references of The VOCAL Test Methodology"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
references	{CITATION| (E, CITATION)}
The VOCAL Test Methodology	{The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

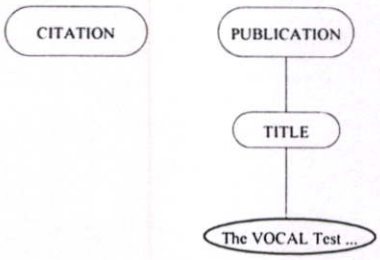
จำนวนของรูปจำลองการสอบถาม คือ 1

รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {CITATION| (E, CITATION), The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

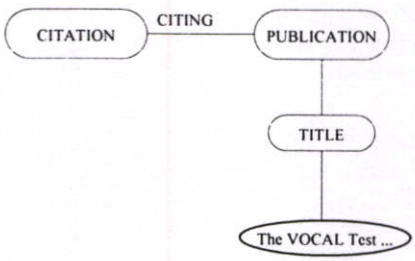
เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.11



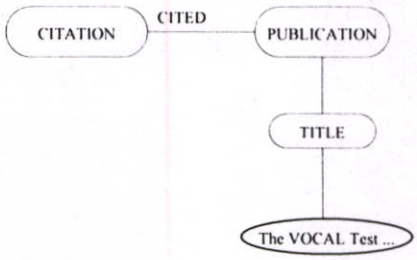
รูปที่ 4.11 เส้นทางพื้นฐานของ QI_1 = {CITATION| (E, CITATION), The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_{11} ของ QI_1 เป็นดังรูปที่ 4.12 ซึ่งมีต้นทุนของกราฟเท่ากับ 3 และ FG_{12} เป็นดังรูปที่ 4.13 ซึ่งมีต้นทุนของกราฟเท่ากับ 3 เช่นกัน ดังนั้นกราฟคำตอบ คือ FG_{11} และ FG_{12}



รูปที่ 4.12 กราฟที่เป็นไปได้ FG_{11} ของ QI_1



รูปที่ 4.13 กราฟที่เป็นไปได้ FG_{12} ของ QI_1

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_{11} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.*, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITING = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.*, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITING = PUBLICATION.KEY) and
      (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

จากกราฟที่เป็นไปได้ FG_{12} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.*, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITED = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.*, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITED = PUBLICATION.KEY) and
```

(PUBLICATION.TITLE = 'The VOCAL Test Methodology')

ขั้นตอนที่ 6 การสร้างเคส

สมมติให้ผู้ใช้เลือกกราฟคำตอบในรูปที่ 4.12 เป็นผลลัพธ์ที่เหมาะสม เคสที่ได้จากกราฟคำตอบดังกล่าวเป็นดังนี้

Problem = {(CITATION| CITATION, E), (The VOCAL Test Methodology| P004, V)}

$Solution = (V^{ae}_{Sol}, E^{ce}_{Sol})$

V^{ae}_{Sol} = {(CITATION, -), (P004, The VOCAL Test Methodology)}

E^{ce}_{Sol} = {PCCiting}

<u>ตัวอย่างที่ 5</u> เมื่อสอบถามด้วย "What are papers cite to The VOCAL Test Methodology"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.5 เมื่อจำนวนคำสำคัญ n = 3

ตารางที่ 4.5 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "What are papers cite to The VOCAL Test Methodology"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
papers	{PUBLICATION| (E, PUBLICATION)}
cite	{CITING| (A, C001), CITED| (A, C002)}
The VOCAL Test Methodology	{The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

จำนวนของรูปจำลองการสอบถาม คือ 2

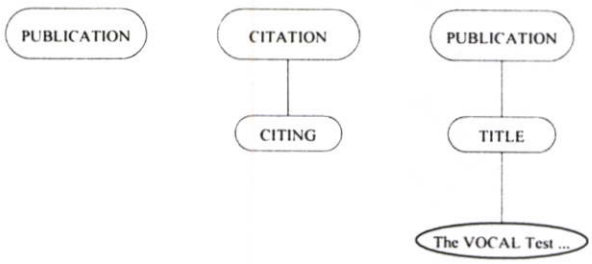
รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {PUBLICATION| (E, PUBLICATION), CITING| (A, C001), The VOCAL Test Methodology| (V, P004)}

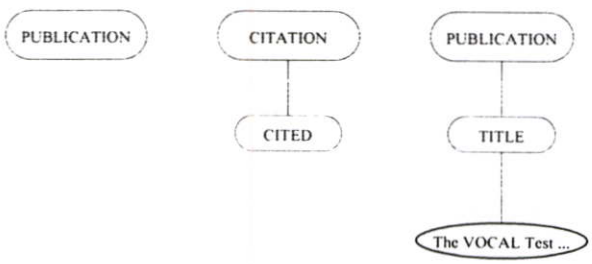
QI_2 = {PUBLICATION| (E, PUBLICATION), CITED| (A, C002),
The VOCAL Test Methodology| (V, P004)}

ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.14 และเส้นทางพื้นฐานของ QI_2 เป็นดังรูปที่ 4.15



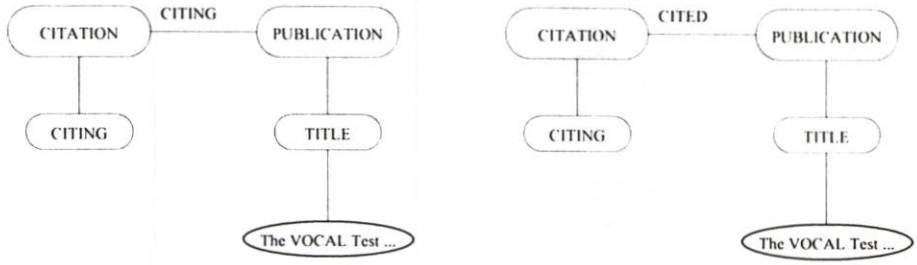
รูปที่ 4.14 เส้นทางพื้นฐานของ QI_1 = {PUBLICATION| (E, PUBLICATION), CITING| (A, C001), The VOCAL Test Methodology| (V, P004)}



รูปที่ 4.15 เส้นทางพื้นฐานของ QI_2 = {PUBLICATION| (E, PUBLICATION), CITED| (A, C002), The VOCAL Test Methodology| (V, P004)}

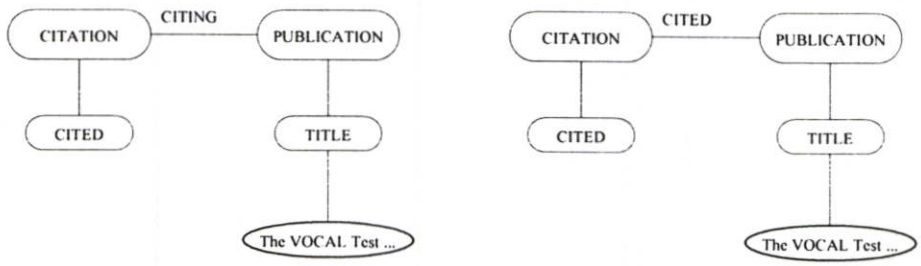
ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_{11} และ FG_{12} ของ QI_1 เป็นดังรูปที่ 4.16 ซึ่งทั้งสองมีต้นทุนของกราฟเท่ากับ 4 และ FG_{21} และ FG_{22} ของ QI_2 เป็นดังรูปที่ 4.17 ซึ่งทั้งสองมีต้นทุนของกราฟเท่ากับ 4 เช่นกัน ดังนั้นกราฟคำตอบ คือ FG_{11} FG_{12} FG_{21} และ FG_{22}



FG11 FG12

รูปที่ 4.16 กราฟที่เป็นไปได้ FG_{11} และ FG_{12} ของ QI_1



FG21 FG22

รูปที่ 4.17 กราฟที่เป็นไปได้ FG_{21} และ FG_{22} ของ QI_2

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_{11} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.CITING, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITING = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.CITING, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITING = PUBLICATION.KEY) and
      (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

จากกราฟที่เป็นไปได้ FG_{12} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.CITING, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITED = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.CITING, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITED = PUBLICATION.KEY) and
      (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

จากกราฟที่เป็นไปได้ FG_{21} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.CITED, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITING = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.CITED, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITING = PUBLICATION.KEY) and
      (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

จากกราฟที่เป็นไปได้ FG_{22} ซึ่งประกอบด้วย

รายการเอนทิตี คือ CITATION, PUBLICATION

รายการแอทริบิว คือ CITATION.CITED, PUBLICATION.TITLE

เงื่อนไขการจอย คือ CITATION.CITED = PUBLICATION.KEY

เงื่อนไขการเลือก คือ PUBLICATION.TITLE = 'The VOCAL Test Methodology'

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT CITATION.CITED, PUBLICATION.TITLE
FROM CITATION, PUBLICATION
WHERE (CITATION.CITED = PUBLICATION.KEY) and
        (PUBLICATION.TITLE = 'The VOCAL Test Methodology')
```

ขั้นตอนที่ 6 การสร้างเคส

สมมติให้ผู้ใช้เลือกกราฟคำตอบ FG21 จากรูปที่ 4.17 เป็นผลลัพธ์ที่เหมาะสม เคสที่ได้จากกราฟคำตอบดังกล่าวเป็นดังนี้

Problem = {(PUBLICATION| PUBLICATION, E), (CITING| C001, A),
(The VOCAL Test Methodology| P004, V)}

Solution = (V^{ae}_{Sol} , E^{ee}_{Sol})

V^{ae}_{Sol} = {(C002, -), (P004, The VOCAL Test Methodology)}

E^{ee}_{Sol} = {PCCiting}

ตัวอย่างที่ 6 เมื่อสอบถามด้วย "What were papers published in year 1999"

ขั้นตอนที่ 1 วิเคราะห์คำจากการสอบถาม

สามารถค้นหาคำสำคัญและเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.6 เมื่อจำนวนคำสำคัญ n = 3

ตารางที่ 4.6 แสดงคำสำคัญที่ค้นพบและเซตของจุดที่ค้นพบจากการสอบถามด้วย "What were papers published in year 1999"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
papers	{PUBLICATION| (E, PUBLICATION)}
year	{YEAR| (A, P007)}
1999	{1999| (V, P007)}

ขั้นตอนที่ 2 กำหนดรูปจำลองการสอบถาม

จำนวนของรูปจำลองการสอบถาม คือ 1

รูปจำลองการสอบถามที่ได้ คือ

QI_1 = {PUBLICATION| (E, PUBLICATION), YEAR| (A, P007), 1999| (V, P007)}

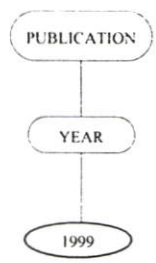
ขั้นตอนที่ 3 หาเส้นทางพื้นฐาน

เส้นทางพื้นฐานของ QI_1 เป็นดังรูปที่ 4.18

รูปที่ 4.18 เส้นทางพื้นฐานของ QI_1 = {PUBLICATION| (E, PUBLICATION), YEAR| (A, P007), 1999| (V, P007)}

ขั้นตอนที่ 4 หากราฟคำตอบ

กราฟที่เป็นไปได้ FG_1 ของ QI_1 เป็นดังรูปที่ 4.19 ซึ่งมีต้นทุนของกราฟเท่ากับ 2



รูปที่ 4.19 กราฟที่เป็นไปได้ FG_1 ของ QI_1

ขั้นตอนที่ 5 สร้างคำสั่งเอสคิวแอล

จากกราฟที่เป็นไปได้ FG_1 ซึ่งประกอบด้วย

รายการเอนทิตี คือ PUBLICATION

รายการแอทริบิว คือ PUBLICATION.*

เงื่อนไขการจอย คือ -

เงื่อนไขการเลือก คือ PUBLICATION.YEAR = 1999

จากข้อมูลดังกล่าว สามารถเขียนให้อยู่ในรูปภาษาเอสคิวแอลได้ ดังนี้

```
SELECT PUBLICATION.*
FROM PUBLICATION
WHERE (PUBLICATION.YEAR = 1999)
```

ขั้นตอนที่ 6 การสร้างเคส

จากกราฟคำตอบในรูปที่ 4.19 สามารถสร้างเคสที่ได้ดังนี้

Problem = {(PUBLICATION| PUBLICATION, E), (YEAR| P007, A), (1999| P007, V)}

Solution = $(V^{ae}_{Sol}, E^{ee}_{Sol})$

V^{ae}_{Sol} = {(P007, 1999)}

E^{ee}_{Sol} = {}

4.4 ตัวอย่างการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส

จากตัวอย่างเคสในภาคผนวก ง การพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส โดยกำหนดให้ค่า θ เท่ากับ 0.5 เป็นดังต่อไปนี้

ตัวอย่างที่ 7 เมื่อสอบถามด้วย "address of Jason Rennie"

ขั้นตอนที่ 1 พิจารณาหาเคสที่อาจเป็นไปได้ โดยพิจารณาหาเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.7

ตารางที่ 4.7 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Get address of Jason Rennie"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
address	{ ADDRESS| (A, A003), ADDRESS| (A, P008), ADDRESS| C, C00001)}
Jason Rennie	{ Jason Rennie | (V, A002), Jason Rennie | (C, C00001)}

เคสที่อาจเป็นไปได้ คือ C00001

ขั้นตอนที่ 2 พิจารณาหาเคสที่เหมาะสม

V_k = {(ADDRESS| A003, A), (Jason Rennie| A002, V)}

C_1 = {(ADDRESS| A003, A), (Jason Rennie| A002, V)}

Edge(C_1) = {}

Sim (V_k , C_1) = 1

ขั้นตอนที่ 3 การรียูสเคส

เนื่องจาก ค่า Sim (V_k , C_1) = 1 ดังนั้น กราฟคำตอบจากการสอบถาม คือ ผลเฉลยของ C_1

<u>ตัวอย่างที่ 8</u> เมื่อสอบถามด้วย "Who write The VOCAL Test Methodology"

ขั้นตอนที่ 1 พิจารณาหาเคสที่อาจเป็นไปได้ โดยพิจารณาหาเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.8

ตารางที่ 4.8 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Who write The VOCAL Test Methodology"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
write	{WRITES| (E, WRITES), (WRITES| C, C00003), (WRITES| C, C00008)}
The VOCAL Test Methodology	{The VOCAL Test Methodology| (V, P004), The VOCAL Test Methodology| (C, C00003), The VOCAL Test Methodology| (C, C00004), The VOCAL Test Methodology| (C, C00005)}

เคสที่อาจเป็นไปได้ คือ C00003

ขั้นตอนที่ 2 พิจารณาหาเคสที่เหมาะสม

V_k = {(WRITES| WRITES, E), (The VOCAL Test Methodology | P004, V)}

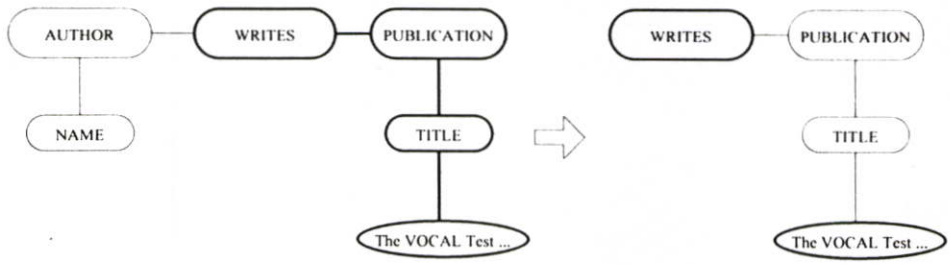
C_3 = {(AUTHOR| AUTHOR, E), (NAME| A002, A), (WRITES| WRITES, E), (The VOCAL Test Methodology| P004, V)}

Edge(C_3) = {AWID, PWKey}

Sim (V_k , C_3) = 0.5

ขั้นตอนที่ 3 การรียูสเคส

กราฟคำตอบ คือ การรียูสผลเฉลยของ C_3 ดังรูปที่ 4.20 ซึ่งประกอบด้วย เซตของจุดในกราฟคำตอบ คือ {(WRITES| WRITES, E), (The VOCAL Test Methodology | P004, V)} และเซตของเส้นเชื่อมของ C_3 คือ {PWKey}



รูปที่ 4.20 กราฟคำตอบที่ได้จากการรีดิวสเคส C_3

<u>ตัวอย่างที่ 9</u> เมื่อสอบถามด้วย "Publication of David Zuckerman and Russell Impagliazzo"

ขั้นตอนที่ 1 พิจารณาหาเคสที่อาจเป็นไปได้ โดยพิจารณาหาเซตของจุดที่ค้นพบ ได้ดังตารางที่ 4.9

ตารางที่ 4.9 แสดงคำสำคัญและเซตของจุดที่ค้นพบจากการสอบถามด้วย "Publication of David Zuckerman and Russell Impagliazzo"

คำสำคัญที่ค้นพบ	เซตของจุดที่ค้นพบ
Publication	{PUBLICATION| (E,PUBLICATION), (PUBLICATION | C, C00005), (PUBLICATION | C, C00006), (PUBLICATION | C, C00008)}
David Zuckerman	{David Zuckerman| (V, A002), David Zuckerman| (C, C00002)}
Russell Impagliazzo	{Russell Impagliazzo| (V, A002), Russell Impagliazzo| (C, C00002)}

เคสที่อาจเป็นไปได้ คือ C00002

ขั้นตอนที่ 2 พิจารณาหาเคสที่เหมาะสม

V_k = {(PUBLICATION| PUBLICATION, E), (David Zuckerman| A002, V),
(Russell Impagliazzo| A002, V)}

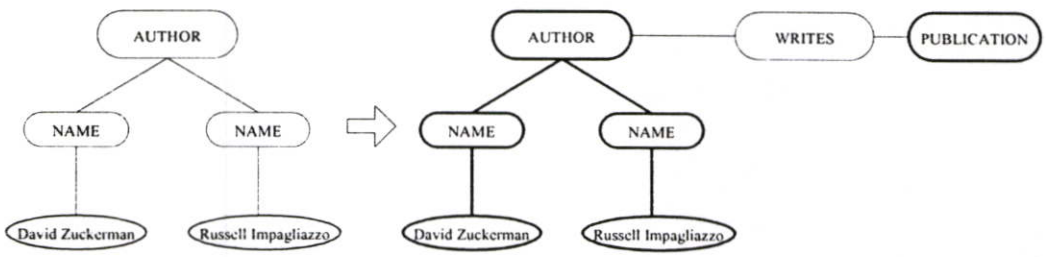
C_2 = {(David Zuckerman| A002, V), (Russell Impagliazzo| A002, V)}

Edge(C_2) = { }

Sim (V_k , C_2) = 0.67

ขั้นตอนที่ 3 การรียูสเคส

กราฟคำตอบ คือ การรียูสผลเฉลยของ C_2 ดังรูปที่ 4.21 ซึ่งประกอบด้วย เซตของจุดในกราฟคำตอบ คือ {(David Zuckerman| A002, V), (Russell Impagliazzo| A002, V)} ∪ {(PUBLICATION| PUBLICATION, E)} และจุด {(WRITES| WRITES, E)} ที่เชื่อมระหว่าง C_2 กับจุด {(PUBLICATION| PUBLICATION, E)}



Case C2 Answer Graph

รูปที่ 4.20 กราฟคำตอบที่ได้จากการรียูสเคส C_2

4.5 สรุปการประเมินผล

จากตัวอย่างการสอบถามตัวอย่างที่ 1 – 6 สามารถเปรียบเทียบผลลัพธ์จากการสอบถามด้วยระบบการค้นหาคำสำคัญที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล (Value Index) และระบบการค้นหาคำสำคัญที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล (Value-Matadata Index) ซึ่งระบบการค้นหาคำสำคัญที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล ได้แก่ ดิสโคเวอร์ ดีบีเอกซ์โพลเรอ และอีเคเอสโอ และระบบการค้นหาคำสำคัญที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูลซึ่งแบ่งเป็น ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตี (Value-Entity Index) ได้แก่ บีเอเอ็นเคเอส และระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อแอทริบิว/ชื่อเอนทิตี (Value-Entity-Attribute Index) ได้แก่ มราจียาติ

การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 1 สอบถามด้วย "Get address of Jason Rennie" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.10

ตารางที่ 4.10 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Get address of Jason Rennie"

คำสำคัญ	ประเภทของคำสำคัญ
address	ข้อมูลฐานข้อมูล (ชื่อแอทริบิว)
Jason Rennie	ค่าข้อมูล

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.11 โดยที่

1. ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล และระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูลหรือชื่อเอนทิตีสามารถใช้เฉพาะคำ "Jason Rennie" ในการค้นหาเท่านั้น

2. ระบบมราจียาติที่สามารถค้นหาคำสำคัญที่ตรงกับค่าข้อมูล ชื่อแอทริบิว หรือชื่อเอนทิตี เมื่อสอบถามด้วยการสอบถามดังกล่าว ผลลัพธ์ที่เป็นแอทริบิว ADDRESS และแอทริบิว NAME อยู่ภายในเอนทิตีเดียวกัน ระบบจึงไม่สามารถกำหนดรายการแอทริบิวได้

3. การค้นข้อมูลฐานข้อมูล สามารถกำหนดรายการการเลือก ซึ่งได้แก่ แอทริบิว ADDRESS ตามความต้องการของผู้ใช้ได้

ตารางที่ 4.11 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Get address of Jason Rennie"

วิธีการ	ผลลัพธ์
Value Index	Select * From AUTHOR Where name = '*Jason Rennie*'
Value-Entity Index	Select * From AUTHOR Where name = '*Jason Rennie*'
Value-Entity-Attribute Index	Select * From AUTHOR Where name = '*Jason Rennie*'
Metadata Search	Select *ADDRESS*, NAME From AUTHOR Where AUTHOR.NAME = '*Jason Rennie*'

การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 2 สอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.12

ตารางที่ 4.12 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo"

คำสำคัญ	ประเภทของคำสำคัญ
David Zuckerman	ค่าข้อมูล
Russell Impagliazzo	ค่าข้อมูล

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.13 พบว่า ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล และระบบบีเอเอ็นเคเอส ให้ผลลัพธ์ที่ต่างจากมราจียาติ และการค้นข้อมูลฐานข้อมูล เนื่องจากระบบอาศัยความสัมพันธ์ระหว่างระเบียนข้อมูลในการหาผลลัพธ์ จึงไม่สามารถใช้ใช้ตัวดำเนินการ OR ในการค้นหาได้ ในขณะที่มราจียาติ และการค้นข้อมูลฐานข้อมูลใช้โครงสร้างข้อมูลในการหาผลลัพธ์ จึงสามารถใช้ตัวดำเนินการ OR ในการกำหนดเงื่อนไขการเลือกได้

ตารางที่ 4.13 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วย "Give me detail of David Zuckerman and Russell Impagliazzo"

วิธีการ	ผลลัพธ์
Value Index	Select * From AUTHOR, WRITES, PUBLICATION Where AUTHOR.ID = WRITES.ID And WRITES.KEY = PUBLICATION And AUTHOR.NAME = '*David Zuckerman*' And AUTHOR.NAME = '*Russell Impagliazzo*'
Value-Entity Index	Select * From AUTHOR, WRITES, PUBLICATION Where AUTHOR.ID = WRITES.ID And WRITES.KEY = PUBLICATION And AUTHOR.NAME = '*David Zuckerman*' And AUTHOR.NAME = '*Russell Impagliazzo*'
Value-Entity-Attribute Index	Select * From AUTHOR Where name = '*David Zuckerman*' Or name = '*Russell Impagliazzo*'
Metadata Search	Select * From AUTHOR Where name = '*David Zuckerman*' Or name = '*Russell Impagliazzo*'

การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 3 สอบถามด้วย "Give me an author name who write The VOCAL Test Methodology" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.14

ตารางที่ 4.14 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me an author name who write The VOCAL Test Methodology"

คำสำคัญ	ประเภทของคำสำคัญ
author	ข้อมูลฐานข้อมูล (ชื่อเอนทิตี)
name	ข้อมูลฐานข้อมูล (ชื่อแอทริบิว)
write	ข้อมูลฐานข้อมูล (ชื่อเอนทิตี)
The VOCAL Test Methodology	ค่าข้อมูล

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.15 พบว่า

1. ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล สามารถใช้เฉพาะคำสำคัญ "The VOCAL Test Methodology" ในการค้นหาเท่านั้น

2. ระบบบีเอเอ็นเคเอสสามารถใช้ชื่อเอนทิตีในการค้นหาได้ ผลลัพธ์จึงประกอบด้วยระเบียนข้อมูลที่มาจากเอนทิตีและระเบียนข้อมูลจากค่าข้อมูลดังกล่าว

3. เนื่องจากผลลัพธ์ประกอบด้วยเอนทิตีมากกว่า 2 เอนทิตี มราจียาติจึงไม่สามารถค้นหาผลลัพธ์ได้

4. การค้นข้อมูลฐานข้อมูล สามารถกำหนดรายการการเลือก ซึ่งได้แก่ แอทริบิว AUTHOR.NAME ตามความต้องการของผู้ใช้ได้

ตารางที่ 4.15 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถาม "Give me an author name who write The VOCAL Test Methodology"

วิธีการ	ผลลัพธ์
Value Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Value-Entity Index	Select * from *AUTHOR*, *WRITES*, PUBLICATION Where (AUTHOR.ID = WRITES.ID And WRITES.KEY = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')

ตารางที่ 4.15 (ต่อ)

วิธีการ	ผลลัพธ์
Value-Entity-Attribute Index	-
Metadata Search	Select *AUTHOR.NAME*, *WRITES.**, PUBLICATION.TITLE From AUTHOR, WRITES, PUBLICATION Where (AUTHOR.ID = WRITES.ID And WRITES.KEY = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')

การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 4 สอบถามด้วย "Give me references of The VOCAL Test Methodology" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.16

ตารางที่ 4.16 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "Give me references of The VOCAL Test Methodology"

คำสำคัญ	ประเภทของคำสำคัญ
references	คำพ้องความหมาย (คำพ้องชื่อเอนทิตี)
The VOCAL Test Methodology	ค่าข้อมูล

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.17 พบว่า ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล ระบบบีเอเอ็นเคเอส และระบบมราจียาติ สามารถใช้เฉพาะคำสำคัญ "The VOCAL Test Methodology" ในการค้นหาเท่านั้น ในขณะที่การค้นข้อมูลฐานข้อมูล สามารถใช้คำพ้องความหมายในการค้นหาผลลัพธ์ได้

ตารางที่ 4.17 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถาม "Give me references of The VOCAL Test Methodology"

วิธีการ	ผลลัพธ์
Value Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Value-Entity Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'

ตารางที่ 4.17 (ต่อ)

วิธีการ	ผลลัพธ์
Value-Entity-Attribute Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Metadata Search	Select *CITATION*.*, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITING = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')
	Select *CITATION*.*, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITED = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')

การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 5 สอบถามด้วย "What are papers cite to The VOCAL Test Methodology" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.18

ตารางที่ 4.18 แสดงคำสำคัญที่ได้จากการสอบถามด้วย "What are papers cite to The VOCAL Test Methodology"

คำสำคัญ	ประเภทของคำสำคัญ
papers	คำพ้องความหมาย (คำพ้องชื่อเอนทิตี)
cite	คำพ้องความหมาย (คำพ้องชื่อแอทริบิว)
The VOCAL Test Methodology	ค่าข้อมูล

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.19 ซึ่งพบว่า ระบบที่ใช้คำดัชนีที่ตรงกับค่าข้อมูล ระบบบีเอเอ็นเคเอส และระบบมราจียาติสามารถใช้เฉพาะคำสำคัญ "The VOCAL Test Methodology" ในการค้นหาเท่านั้น ในขณะที่การค้นข้อมูลฐานข้อมูล สามารถใช้คำพ้องความหมายในการค้นหาผลลัพธ์ได้

ตารางที่ 4.19 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถาม "What are papers cite to The VOCAL Test Methodology"

วิธีการ	ผลลัพธ์
Value Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Value-Entity Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Value-Entity-Attribute Index	Select * from PUBLICATION Where TITLE = '*The VOCAL Test Methodology*'
Metadata Search	Select *CITATION.CITING*, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITING = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')
	Select *CITATION.CITING*, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITED = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')
	Select *CITATION.CITED*, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITING = PUBLICATION.KEY) And (PUBLICATION.TITLE = '*The VOCAL Test Methodology*')
	Select CITATION.CITED, PUBLICATION.TITLE From CITATION, PUBLICATION Where (CITATION.CITED = PUBLICATION.KEY) And (PUBLICATION.TITLE = 'The VOCAL Test Methodology')

<u>การเปรียบเทียบผลลัพธ์จากตัวอย่างที่ 6</u> สอบถามด้วย "What were papers published in year 1999" ประกอบด้วยคำสำคัญและประเภทของคำสำคัญดังตารางที่ 4.20

ตารางที่ 4.20 แสดงคำสำคัญที่ได้จากการสอบถามด้วย"What were papers published in year 1999"

คำสำคัญ	ประเภทของคำสำคัญ
papers	คำพ้องความหมาย (คำพ้องชื่อเอนทิตี)
year	ข้อมูลฐานข้อมูล (ชื่อแอทริบิว)
1999	ค่าข้อมูล (ชนิดจำนวน)

ผลลัพธ์ที่ได้จากการค้นหาด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ เป็นดังตารางที่ 4.21 พบว่า ระบบบีเอเอ็นเคเอส ไม่สามารถใช้คำสำคัญที่ตรงกับค่าข้อมูลชนิดจำนวนในการค้นหาได้ ในขณะที่มราจียาติ และการค้นข้อมูลฐานข้อมูลสามารถค้นหาจากคำสำคัญดังกล่าวได้

ตารางที่ 4.21 แสดงการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถาม "What were papers published in year 1999"

วิธีการ	ผลลัพธ์
Value Index	Select * from PUBLICATION Where YEAR = *1999*
Value-Entity Index	-
Value-Entity-Attribute Index	Select * from PUBLICATION Where *YEAR = 1999*
Metadata Search	Select PUBLICATION.* From PUBLICATION Where *PUBLICATION.YEAR = 1999*

<u>การประเมินผล</u>

เมื่อทำการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ แล้ว พบว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลให้ผลลัพธ์ที่ถูกต้องแม่นยำกว่าระบบการค้นหาคำสำคัญวิธีอื่นๆ ดังตารางที่ 4.22 ที่แสดงให้เห็นว่า

1. ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลเพียงอย่างเดียวไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูล และคำพ้องความหมายได้

2. ระบบที่ใช้คำสำคัญที่ตรงกับค่าข้อมูลและชื่อเอนทิตีไม่สามารถค้นหาคำสำคัญที่ตรงกับค่าข้อมูลประเภทจำนวนหรือวันที่ได้ ไม่สามารถค้นหาคำสำคัญที่ตรงกับชื่อแอทริบิว และคำพ้องความหมายได้

3. ระบบที่ใช้คำสำคัญที่ตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล แม้จะสามารถค้นหาด้วยคำสำคัญที่ตรงกับชื่อเอนทิตีและชื่อแอทริบิวได้ แต่ระบบดังกล่าวไม่สามารถค้นหาคำสำคัญที่ผลลัพธ์ประกอบด้วยเอนทิตีมากกว่า 2 เอนทิตีได้ และไม่สามารถค้นหาด้วยการเจาะจงแอทริบิวที่ขึ้นอยู่กับเอนทิตีเดียวกับคำสำคัญที่ตรงกับค่าข้อมูล ดังตัวอย่างการค้นหาคำสำคัญในตัวอย่างที่ 4.1 ได้ นอกจากนี้ ระบบไม่สามารถค้นหาคำสำคัญที่ตรงกับคำพ้องความหมายได้

4. ระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลสามารถค้นหาคำสำคัญที่ตรงกับค่าข้อมูล (ประเภทอักขระและจำนวน) ข้อมูลฐานข้อมูล (ชื่อเอนทิตีและชื่อแอทริบิว) และคำพ้องความหมายได้

ตารางที่ 4.22 แสดงการเปรียบเทียบความสามารถในการค้นหาคำสำคัญตามประเภทของคำดัชนี

คำดัชนี / วิธีการ	ค่าข้อมูล		ข้อมูลฐานข้อมูล		คำพ้องความหมาย
	อักขระ	จำนวน/วันที่	ชื่อเอนทิตี	ชื่อแอทริบิว	
Value Index	✓	✓	✗	✗	✗
Value-Entity Index	✓	✗	✓	✗	✗
Value-Entity-Attribute Index	✓	✓	✓*	✓*	✗
Metadata Search	✓	✓	✓	✓	✓

บทที่ 5
สรุปและข้อเสนอแนะ

5.1 สรุป

การค้นหาข้อมูลในฐานข้อมูลส่วนใหญ่ต้องใช้ภาษาเอสคิวแอล ซึ่งผู้ใช้จำเป็นต้องมีความรู้ด้านโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม ระบบการค้นหาคำสำคัญในฐานข้อมูลช่วยให้ผู้ใช้สามารถค้นหาข้อมูลจากฐานข้อมูลได้โดยระบุคำสำคัญเช่นเดียวกับการค้นหาข้อมูลบนอินเทอร์เน็ต แต่ระบบการค้นหาคำสำคัญส่วนใหญ่ต้องระบุคำสำคัญที่ตรงกับค่าข้อมูลที่มีชนิดข้อมูลเป็นอักขระ ไม่สามารถค้นหาจากคำพ้องความหมาย หรือคำที่ตรงกับข้อมูลฐานข้อมูลได้ นอกจากนี้ เมื่อผู้ใช้ค้นหาด้วยคำสำคัญกลุ่มเดิม ระบบการค้นหาคำสำคัญในฐานข้อมูลส่วนใหญ่ต้องเริ่มกระบวนการค้นหาใหม่

งานวิจัยนี้จึงได้พัฒนาวิธีการค้นหาคำสำคัญในฐานข้อมูล ซึ่งผู้ใช้สามารถใช้คำสำคัญหรือกลุ่มคำสำคัญดังกล่าวในการค้นหา ซึ่งจากระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูล พบว่า

1. แบบจำลองข้อมูลและดัชนีผกผัน

งานวิจัยนี้ได้ออกแบบแบบจำลองข้อมูลสำหรับการเก็บข้อมูล 3 ประเภท คือ โครงสร้างข้อมูล คำพ้องความหมาย และเคส

1.1 โครงสร้างข้อมูลเก็บอยู่ในรูปของกราฟโครงสร้างข้อมูล ซึ่งไม่จำเป็นต้องคัดลอกค่าข้อมูลจากฐานข้อมูลมายังแบบจำลอง

1.2 คำพ้องความหมาย ช่วยให้ผู้ใช้สามารถค้นหาด้วยคำที่มีความหมายเหมือนค่าข้อมูล หรือข้อมูลฐานข้อมูลได้ เช่น ในฐานข้อมูลประกอบด้วยค่าข้อมูลเดือน ซึ่งเก็บ "JUN" ซึ่งผู้ใช้สามารถค้นหาได้ด้วยชื่อเดือน คือ "June"

1.3 เคส ช่วยในการหาผลลัพธ์ของการสอบถามปัจจุบัน โดยการค้นคืนเคส หรือรียูสเคส ซึ่งในกรณีที่เป็นการสอบถามด้วยคำสำคัญกลุ่มเดิม ผู้ใช้จะได้รับคำตอบที่เหมาะสมตามความต้องการของผู้ใช้

1.4 ดัชนีผกผัน เป็นการนำเทคนิคการค้นคืนสารสนเทศ (Information Retrieval) มาประยุกต์ใช้กับการค้นหาคำสำคัญในฐานข้อมูล ซึ่งช่วยให้ระบบไม่จำเป็นต้องค้นหาคำสำคัญจากฐานข้อมูล หรือแบบจำลองข้อมูลโดยตรง

2. ภาษาสอบถาม

ระบบอนุญาตให้ผู้ใช้สามารถใช้ข้อความอิสระในการสอบถาม โดยไม่จำเป็นต้องระบุเฉพาะคำสำคัญ

3. ผลลัพธ์จากการสอบถาม

ผลลัพธ์จากการสอบถามประกอบด้วย ระเบียนข้อมูลที่ประกอบด้วยคำสำคัญ และ/หรือ ระเบียนข้อมูลที่มีชื่อเอนทิตี หรือชื่อแอทริบิวตรงกับคำสำคัญตามที่ผู้ใช้ต้องการ ซึ่งผู้ใช้สามารถยืนยันความถูกต้องของผลลัพธ์ และเก็บข้อมูลการสอบถามนั้นเป็นเคส เพื่อความถูกต้องของผลลัพธ์ในการสอบถามครั้งต่อไป

นอกจากนี้เมื่อทำการเปรียบเทียบผลลัพธ์ที่ได้จากการสอบถามด้วยระบบการค้นหาคำสำคัญในฐานข้อมูลระบบต่างๆ แล้ว พบว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลให้ผลลัพธ์ที่ถูกต้องแม่นยำกว่าระบบที่ใช้การค้นหาคำสำคัญที่ตรงกับค่าข้อมูลเพียงอย่างเดียว โดยสามารถกำหนดผลลัพธ์ได้จากการเลือกระเบียนข้อมูล การเลือกแอทริบิว และการเลือกเอนทิตี ในขณะที่ระบบที่ใช้การค้นหาคำสำคัญที่ตรงกับค่าข้อมูลเพียงอย่างเดียวสามารถกำหนดผลลัพธ์ได้จากการเลือกระเบียนข้อมูลเท่านั้น

4. การจัดลำดับความสำคัญของผลลัพธ์

ระบบจัดลำดับความสำคัญของผลลัพธ์ตามต้นทุนของกราฟคำตอบ นั่นคือ กราฟคำตอบใดมีขนาดของกราฟเล็กที่สุดจะถือว่ากราฟคำตอบนั้น คือ ผลลัพธ์ที่มีนัยสำคัญมากที่สุด

5.2 ข้อเสนอแนะ

งานวิจัยนี้นำเสนอแบบจำลองข้อมูลและวิธีการค้นหาคำสำคัญในฐานข้อมูล ซึ่งสามารถนำไปสร้างเป็นโปรแกรมต้นแบบเพื่อทดสอบประสิทธิภาพของระบบ และสามารถนำไปประยุกต์ใช้กับการค้นหาคำสำคัญจากฐานข้อมูลต่างๆ ได้

ในส่วนของเคส-เบสรีซันนิงสามารถพัฒนาวิธีการปรับปรุงเคส เพื่อการรียูส และการรีไวส์ (Revise) แล้วนำไปสร้างร่วมกับโปรแกรมต้นแบบ เพื่อพัฒนาเคส-เบสให้เกิดการเรียนรู้มากขึ้น และทดสอบประสิทธิภาพการทำงานของระบบการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นข้อมูลฐานข้อมูลระหว่างการใช้เคส-เบสรีซันนิง และไม่ใช้เคส-เบสรีซันนิง

นอกจากนี้อาจสามารถพัฒนาระบบให้รองรับการสอบถามด้วยภาษาธรรมชาติ (NLP: Natural Language Processing) เช่น การสอบถามด้วย "What are papers cite to The VOCAL Test Methodology" ในตัวอย่างการสอบถามที่ 5 จากบทที่ 4 จะเห็นว่าผลลัพธ์ที่ได้จากการสอบถามมีจำนวน 4 ผลลัพธ์ ซึ่งถ้าระบบสามารถวิเคราะห์ไวยากรณ์ของภาษาได้ ก็จะช่วยจำกัดขอบเขตของผลลัพธ์ให้น้อยลง

เอกสารอ้างอิง

[1] Aditya, B. et. al. 2003. **BANKS: Browsing and Keyword Searching in Relational Databases.** [Online]. Available: http://www.vldb.org/conf/2002/S33P11.pdf.

[2] Aditya, B. et. al. 2003. **Keyword Searching and Browsing in Databases using BANKS.** [Online]. Available: http://www.cse.iitb.ac.in/banks/Public/Publications.

[3] Agrawal, S. et. al. 2002. **DBXplorer: A System for Keyword-Based Search over Relational Databases.** [Online]. Available: http://csdl.computer.org/comp/proceedings /icde/2002/1531/00/15310005abs.htm.

[4] Bergmann, R. **Introduction to Case-Based Reasoning.** [Online]. Available: http://www.dfki.uni-kl.de/~aabecker/Mosbach/Bergmann-CBR-Survey.pdf.

[5] Dar, S. et. al. 1998. **DTL's DataSpot: Database Exploration Using Plain Language.** [Online]. Available: http://citeseer.nj.nec.com/context/1431382/0.

[6] Goldman, R. et. al. **Proximity Search in Databases. [Online].** Available: http://www-db.stanford.edu/lore/pubs/proximity-full.pdf

[7] Hristidis, V. and Papakonstantinou, Y. 2002. **DISCOVER: Keyword Search in Relational Databases.** [Online]. Available: http://feast.ucsd.edu/people/vagelis/ DISCOVER.pdf.

[8] Johnson K. 2002. **Database Search Trips and Tricks.** [Online]. Available: www.csuchico.edu/lins/handouts/ Database_Search_Tips_Feb_03.pdf.

[9] Kolodner, J. 1993. **Case-based reasoning.** San Mateo, CA : Morgan Kaufmann.

[10] Sarda, N. L. and Jain, A. 2001. **Mragyati: Asystem for Keyword-based Searching in Databases.** [Online]. Available: http://arxiv.org/abs/cs.DB/0110052.

[11] Su, Q. and Widom J. **Indexing Relational Database Content Offline for Efficient Keyword-Based Search.** [Online]. Available: http://www-db.standford.edu/~qi/ ekso_ideas05.pdf.

[12] Wang, S. and Zhang K. L. 2005. **Searching Databases with Keywords.** [Online]. Available: http://www.springerlink.com.

เอกสารอ้างอิง (ต่อ)

[13] Wheeldon, R. et. al. **DbSurfer: A Search and Navigation Tool for Relational Databases.** [Online]. Available: http://www.rswheeldon.com/pubs/dbsurfer-short.pdf.

[14] Wheeldon, R. et. al. 2003. **Search and Navigation in Relational Databases.** [Online]. Available: http://arxiv.org/abs/cs.DB/0307073.

[15] http://www.cse.iitb.ac.in/banks/demo/dev-stable/servlet/SearchForm

ภาคผนวก

ภาคผนวก ก

เมตาดาตาสำหรับฐานข้อมูลดีบีแอลพี

ตารางที่ ก.1 แสดงข้อมูลจาก Entity

EntName	EntDescript
AUTHOR	Details of each individual author
CITATION	Links publications with those which reference them
PUBLICATION	Details of the journal, article and book entries
WRITES	Links between publication and author

ตารางที่ ก.2 แสดงข้อมูลจาก Attribute

AttCode	AttName	AttType	AttDescript	EntName
A001	ID	Number		AUTHOR
A002	NAME	Char		AUTHOR
A003	ADDRESS	Char		AUTHOR
C001	CITING	Char		CITATION
C002	CITED	Char		CITATION
C003	LABEL	Char		CITATION
P001	KEY	Char		PUBLICATION
P002	TYPE	Char		PUBLICATION
P003	EDITOR	Char		PUBLICATION
P004	TITLE	Char		PUBLICATION
P005	BOOKTITLE	Char		PUBLICATION
P006	PAGES	Char		PUBLICATION
P007	YEAR	Number		PUBLICATION
P008	ADDRESS	Char		PUBLICATION
P009	JOURNAL	Char		PUBLICATION
P010	VOLUME	Number		PUBLICATION
P011	JOURNAL_NUMBER	Char		PUBLICATION
P012	MONTH	Char		PUBLICATION
P013	URL	Char		PUBLICATION
P014	EE	Char		PUBLICATION
P015	PUBLISHER	Char		PUBLICATION
P016	NOTE	Char		PUBLICATION

ตารางที่ ก.2 (ต่อ)

AttCode	AttName	AttType	AttDescript	EntName
P017	ISBN	Char		PUBLICATION
P018	SERIES_TITLE	Char		PUBLICATION
P019	SERIES_URL	Char		PUBLICATION
P020	SCHOOL	Char		PUBLICATION
P021	CHAPTER	Number		PUBLICATION
W001	ID	Number		WRITES
W002	KEY	Char		WRITES

ตารางที่ ก.3 แสดงข้อมูลจาก Equivalent

EqiCode	PAtt	FAtt
AWID	A001	W001
PCCiting	P001	C001
PCCited	P001	C002
PWKey	P001	W001

ตารางที่ ก.4 แสดงข้อมูลจาก Integrity

IntName	IntType	Master	Slave
WriteAuthor	FK	WRITES	AUTHOR
WritePub	FK	WRITES	PUBLISHER
CitePub	FK	CITATION	PUBLISHER

ภาคผนวก ข

ตัวอย่างค่าข้อมูลจากฐานข้อมูลดีบีแอลพี

ตารางที่ ข.1 แสดงตัวอย่างข้อมูลจากเอนทิตี Publication

KEY	TYPE	EDITOR	TITLE
chiang00statistical	misc		Statistical Parsing With an Automatically-Extracted Tree Adjoining Grammar
agarwal99geometric	misc		Geometric Range Searching and Its Relatives
rennie99using	inproceedings	Ivan Bratko and Saso Dzeroski	Using Reinforcement Learning to Spider the Web Efficiently
sommerville97viewpoints	article		Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering
lamport95byzantine	incollection		The Byzantine Generals Problem
shao01adaptive	misc		Adaptive Scheduling of Master/Worker Applications on Distributed Computational Resources
butz01algorithmic	article		An Algorithmic Description of XCS
dolbec-component	misc		A Component Based Software Reliability Model
turk96visual	misc		Visual Interaction With Lifelike Characters
impagliazzo89how	inproceedings		How to Recycle Random Bits
path-measuring	misc		Measuring Bottleneck Bandwidth Of Targeted Path
radoslavov01topologyinformed	inproceedings		Topology-Informed Internet Replica Placement

ตารางที่ ข.1 (ต่อ)

KEY	TYPE	EDITOR	TITLE
boppana90approximating	inproceedings	J. R. Gilbert and R. Karlsson	Approximating Maximum Independent Sets by Excluding Subgraphs
bradshaw97introduction	incollection	Jeffrey M. Bradshaw	An Introduction to Software Agents
deb00fast	inproceedings	Marc Schoenauer and Kalyanmoy Deb and G Rudolph and Xin Yao and Evelyne Lutton and J. J. Merelo and Hans-Paul Schwefel	A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II
feigeminimal	inproceedings		A Minimal Model for Secure Computation
fletcher97nonlinear	techreport		Nonlinear Programming without a penalty function
fletcher97nonlinear	misc		Building Awareness in Global Software Engineering: Using Issues as Context
chang01oblivious	article		Oblivious Polynomial Evaluation and Oblivious Neural Learning
pemberton-vocal	misc		The VOCAL Test Methodology
voas95software	article		Software Testability: The New Verification
sommerville97viewpoints	article		Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering
impagliazzo89how	inproceedings		How to Recycle Random Bits

ตารางที่ ข.1 (ต่อ)

KEY	TYPE	EDITOR	TITLE
hastad93construction	misc		Construction of Pseudorandom Generator from any One-Way Function
lindell00privacy	article		Privacy Preserving Data Mining
gray97data	article	Usama Fayyad and Heikki Mannila and Gregory Piatetsky-Shapiro	Data Cube: {A} Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals
clarkson89applications	article		Applications of Random Sampling in Computational Geometry, {II}
agarwal95davenportschinzel	techreport		Davenport--Schinzel Sequences and Their Geometric Applications
menczer97arachnid	inproceedings		ARACHNID: Adaptive Retrieval Agents Choosing Heuristic Neighborhoods for Information Discovery
hofmann98statistical	techreport		Statistical Models for Co-occurrence Data

ตารางที่ ข.1 (ต่อ)

KEY	BOOKTITLE	PAGES	YEAR	ADDRESS	JOURNAL
rennie99using	Proceedings of {ICML}-99, 16th International Conference on Machine Learning	335--343	1999	Bled, SL	
sommerville97viewpoints		101-130	1997		Annals of Software Engineering
lamport95byzantine	Advances in Ultra-Dependable Distributed Systems, N. Suri, C. J. Walter, and M. M. Hugue (Eds.), {IEEE} Computer Society Press		1995		
shao01adaptive			2001		
butz01algorithmic		253--??	2001		Lecture Notes in Computer Science
dolbec-component					
turk96visual			1996		
impagliazzo89how	{IEEE} Symposium on Foundations of Computer Science	248-253	1989		
path-measuring					
radoslavov01topologyinformed	Proceedings of WCW'01: Web Caching and Content Distribution Workshop, Boston, MA		2001		

ตารางที่ ข.1 (ต่อ)

KEY	BOOKTITLE	PAGES	YEAR	ADDRESS	JOURNAL
boppana90approximating	{SWAT} 90 2nd Scandinavian Workshop on Algorithm Theory	13--25	1990		
bradshaw97introduction	Software Agents	3--46	1997		
deb00fast	Proceedings of the Parallel Problem Solving from Nature {VI} Conference	849--858	2000	Paris, France	
feigeminimal		554--563			
fletcher97nonlinear			1997	Dundee DD1 4HN, UK	
fletcher97nonlinear					
chang01oblivious		369+	2001		Lecture Notes in Computer Science
pemberton-vocal					
voas95software		17--28	1995		IEEE Software
sommerville97viewpoints		101-130	1997		Annals of Software Engineering
impagliazzo89how	{IEEE} Symposium on Foundations of Computer Science	248-253	1989		

ตารางที่ ข.1 (ต่อ)

KEY	BOOKTITLE	PAGES	YEAR	ADDRESS	JOURNAL
hastad93construction			1993		
lindell00privacy		36--??	2000		Lecture Notes in Computer Science
gray97data		29--53	1997		J. Data Mining and Knowledge Discovery
clarkson89applications		387--421	1989		Discrete and Computational Geometry
agarwal95davenportschinzel			1995		
menczer97arachnid	Machine Learning: Proceedings of the Fourteenth International Conference	227--235	1997		
hofmann98statistical		21	1998		

ตารางที่ ข.1 (ต่อ)

KEY	VOLUME	JOURNAL_NUMBER	MONTH	URL	EE
chiang00statistical				citeseer.ist.psu.edu/chiang00statistical.html	
agarwal99geometric				citeseer.ist.psu.edu/article/agarwal99geometric.html	
rennie99using				citeseer.ist.psu.edu/article/rennie99using.html	
sommerville97viewpoints	3			citeseer.ist.psu.edu/sommerville97viewpoints.html	
lamport95byzantine				citeseer.ist.psu.edu/lamport82byzantine.html	
shao01adaptive				citeseer.ist.psu.edu/shao01adaptive.html	
butz01algorithmic	1996			citeseer.ist.psu.edu/butz01algorithmic.html	
dolbec-component				citeseer.ist.psu.edu/dolbec95component.html	
turk96visual				citeseer.ist.psu.edu/turk96visual.html	
impagliazzo89how				citeseer.ist.psu.edu/impagliazzo89how.html	
path-measuring				citeseer.ist.psu.edu/harfoush01measuring.html	
radoslavov01topologyinformed			jun	citeseer.ist.psu.edu/radoslavov01topologyinformed.html	
boppana90approximating	447			citeseer.ist.psu.edu/boppana92approximating.html	
bradshaw97introduction				citeseer.ist.psu.edu/bradshaw97introduction.html	
deb00fast				citeseer.ist.psu.edu/deb00fast.html	
feigeminimal				citeseer.ist.psu.edu/feige94minimal.html	

ตารางที่ ข.1 (ต่อ)

KEY	VOLUME	JOURNAL_NUMBER	MONTH	URL	EE
fletcher97nonlinear				citeseer.ist.psu.edu/fletcher00nonlinear.html	
fletcher97nonlinear				citeseer.ist.psu.edu/568949.html	
chang01oblivious	2248			citeseer.ist.psu.edu/article/chang01oblivious.html	
pemberton-vocal				citeseer.ist.psu.edu/pemberton98vocal.html	
voas95software	12	3	May	citeseer.ist.psu.edu/voas95software.html	
sommerville97viewpoints	3			citeseer.ist.psu.edu/sommerville97viewpoints.html	
impagliazzo89how				citeseer.ist.psu.edu/impagliazzo89how.html	
hastad93construction				citeseer.ist.psu.edu/hastad93construction.html	
lindell00privacy	1880			citeseer.ist.psu.edu/lindell00privacy.html	
gray97data	1	1		citeseer.ist.psu.edu/gray97data.html	
clarkson89applications	4	1		citeseer.ist.psu.edu/clarkson95applications.html	
agarwal95davenportschinzel		Technical report DUKE--TR--1995--21		citeseer.ist.psu.edu/340081.html	
menczer97arachnid				citeseer.ist.psu.edu/menczer97arachnid.html	
hofmann98statistical		AIM-1625		citeseer.ist.psu.edu/hofmann98statistical.html	

ตารางที่ ข.1 (ต่อ)

KEY	PUBLISHER	NOTE	ISBN
chiang00statistical			
agarwal99geometric			
rennie99using	Morgan Kaufmann Publishers, San Francisco, US		
sommerville97viewpoints			
lamport95byzantine			
shao01adaptive			
butz01algorithmic			
dolbec-component			
turk96visual			
impagliazzo89how			
path-measuring			
radoslavov01topologyinformed			
boppana90approximating			
bradshaw97introduction	AAAI Press / The MIT Press		
deb00fast	Springer. Lecture Notes in Computer Science No. 1917		

ตารางที่ ข.1 (ต่อ)

KEY	PUBLISHER	NOTE	ISBN
feigeminimal			
fletcher97nonlinear			
fletcher97nonlinear			
chang01oblivious			
pemberton-vocal			
voas95software			
sommerville97viewpoints			
impagliazzo89how			
hastad93construction			
lindell00privacy			
gray97data clarkson89applications agarwal95davenportschinzel menczer97arachnid hofmann98statistical	Kluwer Academic Publishers		

ตารางที่ ข.1 (ต่อ)

KEY	SERIES_TITLE	SERIES_URL	SCHOOL	CHAPTER
chiang00statistical				
agarwal99geometric				
rennie99using				
sommerville97viewpoints				
lamport95byzantine				
shao01adaptive				
butz01algorithmic				
dolbec-component				
turk96visual				
impagliazzo89how				
path-measuring				
radoslavov01topologyinformed				
boppana90approximating				
bradshaw97introduction				
deb00fast				
feigeminimal				

ตารางที่ ข.1 (ต่อ)

KEY	SERIES_TITLE	SERIES_URL	SCHOOL	CHAPTER
fletcher97nonlinear				
fletcher97nonlinear				
chang01oblivious				
pemberton-vocal				
voas95software				
sommerville97viewpoints				
impagliazzo89how				
hastad93construction				
lindell00privacy				
gray97data				
clarkson89applications				
agarwal95davenportschinzel				
menczer97arachnid				
hofmann98statistical				

ตารางที่ ข.2 แสดงตัวอย่างข้อมูลจากเอนทิตี Author

ID	NAME	ADDRESS
0001	Allen H. Dutoit	Menlo Park, CA
0002	Amrit Pratap	Oakland, CA
0003	Andrew McCallum	Berkeley, CA
0004	Azer Bestavros	San Jose, CA
0005	Bernd Bruegge	Oakland, CA
0006	Chi-Jen Lu	Beijing, China
0007	David Chiang	Berkeley, CA
0008	David Zuckerman	Palo Alto, CA
0009	Deborah Estrin	Covelo, CA
0010	Duncan Pemberton	San Francisco, CA
0011	Gary Shao	Nashville, TN
0012	Ian Sommerville	Corvallis, OR
0013	Jason Rennie	Walnut Creek, CA
0014	Jean Dolbec	Ann Arbor, MI
0015	Jeff Erickson	Gary, IN
0016	Jeffrey M. Bradshaw	Oakland, CA
0017	Joe Kilian	Paris, France
0018	John Byers	Oakland, CA
0019	Kalyanmoy Deb	Boston, MA
0020	Khaled Harfoush	Washington, DC
0021	Leslie Lamport	Chicago, IL
0022	Magnús M. Halldórsson	Dallas, TX
0023	Marshall Pease	Munchen, Germany
0024	Martin V. Butz	San Francisco, CA
0025	Matthew Turk	Landon, England
0026	Moni Naor	New York, NY
0027	Oliver Creighton	Liverpool, England
0028	Pankaj K. Agarwal	Berkeley, CA
0029	Pavlin Radoslavov	New York, NY

ตารางที่ ข.2 (ต่อ)

ID	NAME	ADDRESS
0030	Pete Sawyer	Seattle, WA
0031	Rafael Kobylinski	Tustin, CA
0032	Ramesh Govindan	Las Gatos, CA
0033	Ravi Boppana	Fremont, CA
0034	Robert Shostak	Portland, OR
0035	Roger Fletcher	Remulade, WA
0036	Russell Impagliazzo	Nashville, TN
0037	Samir Agrawal	Dallas, TX
0038	Stewart W. Wilson	Covelo, CA
0039	Sven Leyffer	Seattle, WA
0040	T Meyarivan	Corvallis, OR
0041	Terry Shepard	Dallas, TX
0042	Uri Feige	San Jose, CA
0043	Yan-Cheng Chang	Chicago, IL
0044	Jeffrey M. Voas	Malet St, London
0045	Keith W. Miller	San Diego, CA
0046	Johan Hastad	Seattle, WA
0047	Leonid A. Levin	New York, USA
0048	Michael Luby	Bournemouth, UK
0049	Yehuda Lindell	Lawrence, KS
0050	Benny Pinkas	Bombay, India
0051	Jim Gray	Dallas, TX
0052	Surajit Chaudhuri	Washington, DC
0053	Adam Bosworth	Oakland, CA
0054	Andrew Layman	Mumbai, India
0055	Don Reichart	Varanasi, India
0056	Murali Venkatrao	Portland, OR
0057	Frank Pellow	Berkeley, CA
0058	Hamid Pirahesh	Bournemouth, UK

ตารางที่ ข.2 (ต่อ)

ID	NAME	ADDRESS
0059	Kenneth L. Clarkson	Regina, Canada
0060	Peter W. Shor	Warsaw, Poland
0061	Micha Sharir	Tokyo, Japan
0062	Filippo Menczer	Warsaw, Poland
0063	Thomas Hofmann	Remulade, WA
0064	Jan Puzicha	Boston, MA

ตารางที่ ข.3 แสดงตัวอย่างข้อมูลจากเอนทิตี Writes

ID (AUTHOR)	KEY(PUBLICATION)
0007	chiang00statistical
0028	agarwal99geometric
0015	agarwal99geometric
0013	rennie99using
0003	rennie99using
0012	sommerville97viewpoints
0030	sommerville97viewpoints
0021	lamport95byzantine
0034	lamport95byzantine
0023	lamport95byzantine
0011	shao01adaptive
0024	butz01algorithmic
0038	butz01algorithmic
0014	dolbec-component
0041	dolbec-component
0025	turk96visual
0036	impagliazzo89how
0008	impagliazzo89how
0020	path-measuring
0004	path-measuring

ตารางที่ ข.3 (ต่อ)

ID (AUTHOR)	KEY(PUBLICATION)
0018	path-measuring
0029	radoslavov01topologyinformed
0032	radoslavov01topologyinformed
0009	radoslavov01topologyinformed
0033	boppana90approximating
0022	boppana90approximating
0016	bradshaw97introduction
0019	deb00fast
0037	deb00fast
0002	deb00fast
0040	deb00fast
0042	feigeminimal
0017	feigeminimal
0026	feigeminimal
0035	fletcher97nonlinear
0039	fletcher97nonlinear
0031	fletcher97nonlinear
0027	fletcher97nonlinear
0001	fletcher97nonlinear
0003	fletcher97nonlinear
0043	chang01oblivious
0006	chang01oblivious
0010	pemberton-vocal
0044	voas95software
0045	voas95software
0012	sommerville97viewpoints
0030	sommerville97viewpoints
0008	impagliazzo89how
0036	impagliazzo89how

ตารางที่ ข.3 (ต่อ)

ID (AUTHOR)	KEY(PUBLICATION)
0036	hastad93construction
0046	hastad93construction
0047	hastad93construction
0048	hastad93construction
0049	lindell00privacy
0050	lindell00privacy
0051	gray97data
0052	gray97data
0053	gray97data
0054	gray97data
0055	gray97data
0056	gray97data
0057	gray97data
0058	gray97data
0059	clarkson89applications
0060	clarkson89applications
0061	agarwal95davenportschinzel
0028	agarwal95davenportschinzel
0062	menczer97arachnid
0063	hofmann98statistical
0064	hofmann98statistical

ตารางที่ ข.4 แสดงตัวอย่างข้อมูลจากเอนทิตี Citation

CITING	CITED	LABEL
pemberton-vocal	voas95software	
pemberton-vocal	sommerville97viewpoints	
chang01oblivious	impagliazzo89how	
chang01oblivious	hastad93construction	
chang01oblivious	lindell00privacy	
agarwal99geometric	gray97data	
agarwal99geometric	clarkson89applications	
agarwal99geometric	agarwal95davenportschinzel	
rennie99using	menczer97arachnid	
rennie99using	hofmann98statistical	

ภาคผนวก ค

ตัวอย่างคำพ้องความหมายในฐานข้อมูลดีบีแอลพี

ตารางที่ ค.1 แสดงคำพ้องความหมายของฐานข้อมูลดีบีแอลพี

Synonym	SType	MeanTo
book	A	BOOKTITLE|P005
California	V	CA|A003
caption	A	TITLE|P004
cite	A	CITING|C001
cite	A	CITED|C002
code	A	ID|A001
code	A	ID|W001
England	V	UK|P008
heading	A	TITLE|P004
Illinois	V	IL|A003
Indiana	V	IN|A003
journal	E	PUBLICATION| PUBLICATION
journals	E	PUBLICATION| PUBLICATION
June	V	jun|P011
Kansas	V	KS|A003
Massachusetts	V	MA|A003
Michigan	V	MI|A003
miscelloneous	V	misc|P002
NYC	V	New York|A003
Oregon	V	OR|A003
paper	E	PUBLICATION| PUBLICATION
papers	E	PUBLICATION| PUBLICATION
reference	E	CITATION| CITATION
tech report	V	techreport|P002
Tennessee	V	TN|A003
Texas	V	TX|A003
tome	E	PUBLICATION| PUBLICATION
treatise	E	PUBLICATION| PUBLICATION

ตารางที่ ข.1 (ต่อ)

Synonym	SType	MeanTo
United Kingdom	V	UK|P008
WA	V	Washington|A003
write	E	WRITES|WRITES
writer	E	AUTHOR|AUTHOR

ภาคผนวก ง

ตัวอย่างเคส

ตารางที่ ง.1 แสดงตัวอย่างเคส

CaseCode	Problem	Solution	
		Attribute-Value	Edge
C00001	{(ADDRESS| A003, A), (Jason Rennie| A002, V)}	{(A003, -), (A002, Jason Rennie)}	{}
C00002	{(David Zuckerman| A002, V), (Russell Impagliazzo| A002, V)}	{(A002, David Zuckerman), (A002, Russell Impagliazzo)}	{}
C00003	{(AUTHOR| AUTHOR, E), (NAME| A002, A), (WRITES| WRITES, E), (The VOCAL Test Methodology| P004, V)}	{(A002, -), (WRITES, -), (P004, The VOCAL Test Methodology)}	{AWID, PWKey}
C00004	{(CITATION| CITATION, E), (The VOCAL Test Methodology| P004, V)}	{(CITATION, -), (P004, The VOCAL Test Methodology) }	{PCCiting}
C00005	{(PUBLICATION| PUBLICATION, E), (CITING| C001, A), The VOCAL Test Methodology| P004, V)}	{(C002, -), (P004, The VOCAL Test Methodology)}	{PCCiting}
C00006	{(PUBLICATION| PUBLICATION, E), (YEAR| P007, A), (1999| P007, V)}	{(P007, 1999)}	{}
C00007	{(Chi-Jen Lu| A002, A)}	{(A002, Chi-Jen Lu)}	{}
C00008	{(PUBLICATION| PUBLICATION, E), (Chi-Jen Lu| A002, A), (WRITES| WRITES, E)}	{(PUBLICATION,-), (WRITES, -), (A002, Chi-Jen Lu)}	{PWKey, AWID}

ประวัติผู้เขียน

ชื่อ-สกุล	นางสาวจารุณี แซ่หลี
วัน เดือน ปี เกิด	20 มิถุนายน 2523
ที่อยู่	7/12 หมู่ 1 ตำบลสะบ้าย้อย อำเภอสะบ้าย้อย
	จังหวัดสงขลา 90210
ประวัติการศึกษา	2544 วิทยาศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์ประยุกต์
	คณะวิทยาศาสตร์และเทคโนโลยี
	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

ประวัติการทำงาน

มิถุนายน-กันยายน 2548	อาจารย์พิเศษ มหาวิทยาลัยกรุงเทพ วิทยาเขตรังสิต
2547	ผู้ประสานงาน สถาบันศูนย์ศึกษาต่อเนื่องจุฬาฯ สาขาซีคอนสแคว์
มีนาคม-เมษายน 2543	ฝึกงานที่บริษัท Forward Management Services Co., Ltd. (FMS)
	293 ถ. เทียมร่วมมิตร เขต ห้วยขวาง กรุงเทพฯ

ทุนการศึกษา

2549	ทุนสนับสนุนการทำวิทยานิพนธ์
	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
	กรุงเทพมหานคร
2546-2547	ทุนพัฒนาอาจารย์สาขาขาดแคลน สาขาวิทยาการคอมพิวเตอร์
	สังกัดมหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี จังหวัดปัตตานี