ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย

THE IMAGE SEARCH ENGINE USING SEMANTIC WEB

ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย

อานนท์ ไกรเสวกวิสัย

การศึกษาเฉพาะบุคคลเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศและการจัดการ
มหาวิทยาลัยกรุงเทพ
พ.ศ.2552

© 2552

อานนท์ ไกรเสวกวิสัย

สงวนลิขสิทธิ์

บัณฑิตวิทยาลัย มหาวิทยาลัยกรุงเทพ

อนุมัติให้การศึกษาเฉพาะบุคคลนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต

เรื่อง ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย

ผู้วิจัย นาย อานนท์ ไกรเสวกวิสัย

ได้พิจารณาเห็นชอบโดย

อาจารย์ที่ปรึกษา [redacted]

(ดร. ธนกร หวังพิพัฒน์วงศ์)

ผู้แทนบัณฑิตวิทยาลัย [redacted]

(ดร. วุฒนิพงษ์ วราไกรสวัสดิ์)

กรรมการผู้ทรงคุณวุฒิ [redacted]

(ผู้ช่วยศาสตราจารย์ ดร. ภัทรชัย ลลิตโรจน์วงศ์)

[redacted]

(ดร. สุดารัตน์ ดิษยวรรธนะ จันทราวัฒนากูล)

คณบดีบัณฑิตวิทยาลัย

วันที่ 19 เดือน มิถุนายน พ.ศ. 2552

อานนท์ ไกรเสวกวิสัย. ปริญญาวิทยาศาสตรมหาบัณฑิต, มิถุนายน 2552, บัณฑิตวิทยาลัย มหาวิทยาลัยกรุงเทพ.
ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย (91 หน้า)
อาจารย์ที่ปรึกษา : ดร.ธนกร หวังพิพัฒนวงศ์

บทคัดย่อ

การสืบค้นรูปภาพบนเว็บส่วนใหญ่ใช้หลักการสืบค้นโดยใช้คำหลัก (Keyword) ซึ่งทำให้ได้ผลลัพธ์คือ เอกสารที่มีคำหลักที่ตรงกับคำที่ใช้สืบค้น โดยไม่สนใจความหมายของข้อมูล ส่งผลให้ผลลัพธ์จากการสืบค้นมีจำนวนมาก ผู้ใช้ต้องใช้เวลาคัดเลือกรูปภาพที่ตรงกับความต้องการ ในอนาคตการสืบค้นบนอินเตอร์เน็ตจะเปลี่ยนมาเป็นการสืบค้นเชิงความหมาย ซึ่งทำให้คอมพิวเตอร์สามารถเข้าใจความหมายของข้อมูลในเอกสารบนอินเตอร์เน็ตได้ ส่งผลทำให้ได้ผลลัพธ์จากการสืบค้นที่มีความแม่นยำมากขึ้น งานวิจัยนี้นำเสนอแนวความคิดการสืบค้นรูปภาพบนเว็บโดยใช้แนวความคิดของเว็บเชิงความหมาย งานวิจัยแบ่งงานออกเป็น 2 ขั้นตอน คือ การเตรียมข้อมูลสำหรับการสืบค้น ในขั้นตอนนี้จะมีการออกแบบโครงสร้างข้อมูลออนโทโลยีสำหรับบรรยายสถานที่ท่องเที่ยวในประเทศไทย แล้วทำการจัดเก็บคำบรรยายรูปภาพให้อยู่ในรูปแบบของภาษา RDF และในขั้นตอนที่สองเป็นการออกแบบขั้นตอนสืบค้นและพัฒนาแบบจำลองในการสืบค้น ใช้เครื่องมือคือ ความแม่นยำ (Precision) และการเรียกซ้ำ (Recall) ในการวัดประสิทธิภาพในการประเมินประสิทธิภาพของระบบ โดยระดับความแม่นยำอยู่ที่ 6.6 และระดับการเรียกซ้ำอยู่ที่ 7.1

Kraisawekwisai, A. M.S. (Information Technology and Management), June 2009, Graduate School, Bangkok University

The Image Search Engine Using Semantic Web (91 หน้า)

Independent Study Advisor : Thanakorn Wangpipatwong, Ph.D.

ABSTRACT

Current web image searching activities use keyword-based search engines. Traditional search engines return a large amount of irrelevant information and provide no semantic relationships among terms in other document. People spend more time to find out their required image. The future search engines on World Wide Web, the semantic search engine. They will be able to understand the meaning of the content on the web pages and carry out logical reasoning on them to perform complex search queries and return accurate result. In this paper propose a conceptual architecture for a semantic image search engine. This paper parts two steps. The first step to present Ontology models for representation of tourist attraction in Thailand which its properties are structured according to image type, attraction type and attraction sub-type, this model was able to represent in graph base on OWL then the image description has been stored as RDF Format. The second step presents semantic search algorithm for RDF Data. Finally, the precision and recall criterion is employed for efficient evaluation. The experimental results reported that the proposed method was able to output the precision 6.6 and recall 7.1

กิตติกรรมประกาศ

ในการทำโครงการวิจัยระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย (The Image Search Engine using Semantic Web) ภาควิชาเทคโนโลยีสารสนเทศและการจัดการ ได้รับความกรุณาอย่างดียิ่งจาก ดร.ธนกร หวังพิพัฒน์วงศ์ ที่ได้กรุณาให้คำปรึกษา คำแนะนำเกี่ยวกับการทำวิจัย รวมถึงช่วยแก้ไขรายงานการวิจัยให้มีความสมบูรณ์ ที่เกิดขึ้นด้วยความเอาใจใส่ ผู้ทำวิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ.ที่นี้

อานนท์ ไกรเสวกวิสัย

BANGKOK UNIVERSITY FOUNDED 1962

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ญ
สารบัญภาพ	ฏ
บทที่ 1 บทนำ	1
1.1 ชื่อโครงการ	1
1.2 ความเป็นมาและความสำคัญของปัญหา	1
1.3 วัตถุประสงค์ของการวิจัย	2
1.4 ขอบเขตของการวิจัย	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
1.6 แผนการดำเนินงาน	3
1.7 โครงสร้างของงานวิจัย	4
2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	6
2.1 วรรณกรรมที่เกี่ยวข้อง	6
2.1.1 วิธีการสืบค้นสารสนเทศบนเว็บ	6
2.1.2 แนวความคิดเว็บเชิงความหมาย	7
2.1.3 XML (eXtensible Markup Language)	9
2.1.4 RDF (Resource Description Language)	10
2.1.5 RDFS (RDF Schema)	13
2.1.6 RDQL (RDF Data Query Language)	14
2.1.7 ออนโทโลยี	15
2.1.8 OWL (Web Ontology Language)	16
2.1.9 Lexical Chains	20

สารบัญ (ต่อ)

		หน้า
	2.1.10 ฐานความรู้ธีซอรัส	21
	2.1.11 ฐานความรู้เวิร์ดเนต	21
	2.1.12 การวัดประสิทธิภาพของการสืบค้น	22
2.2	งานวิจัยที่เกี่ยวข้อง	25
	2.2.1 การจัดเก็บและสืบค้นรูปภาพเชิงความหมายบนฐานของ RDF	25
	2.2.2 การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย	31
3	ขั้นตอนการดำเนินงาน	36
3.1	การออกแบบจำลองการสืบค้นข้อมูลรูปภาพ	36
	3.1.1 Crawler (Robot Software)	37
	3.1.2 Text Operations	37
	3.1.3 การทำดัชนี (Indexing)	37
	3.1.4 การสืบค้น (Searching)	37
3.2	การออกแบบโครงสร้างข้อมูลออนโทโลยี	38
	3.2.1 กำหนดขอบเขตและแนวทางของออนโทโลยี	38
	3.2.2 การระบุเงื่อนไขในออนโทโลยี	38
	3.2.3 การกำหนดคลาสและลำดับชั้นของคลาส	38
3.3	การบรรยายออนโทโลยีด้วยภาษา OWL	43
	3.3.1 การกำหนด Namespace	43
	3.3.2 การอธิบายคลาสของออนโทโลยี	44
3.4	การเตรียมข้อมูลการสืบค้น	45
	3.4.1 การคัดกรองข้อมูล (Data Cleaning)	46
	3.4.2 การสร้าง Lexical Chains	46
	3.4.3 การอธิบายโครงสร้างข้อมูลรูปภาพด้วยภาษา OWL	48

สารบัญ (ต่อ)

	หน้า
3.5 การออกแบบการสืบค้นรูปภาพ	49
3.5.1 พิจารณาคำ, วลีสืบค้น	49
3.5.2 การสืบค้นข้อมูลจากฐานข้อมูล	50
3.6 เครื่องมือที่ใช้ในการดำเนินงาน	51
3.6.1 Protégé	51
3.6.2 เครื่องมือในการค้นหาข้อมูลในเอกสาร RDF	52
3.6.3 เครื่องมือที่ใช้ในการคิวรีข้อมูล	53
3.6.4 เครื่องมือในการแสดงผลข้อมูลการสืบค้นรูปภาพ	53
3.7 การทดสอบวัดประสิทธิภาพความแม่นยำในการสืบค้น	53
4 ผลการวิจัยและอภิปรายผล	55
4.1 ผลการวิจัย	55
4.2 การอภิปรายผล	55
4.2.1 การเก็บรวบรวมข้อมูลรูปภาพและคำบรรยายภาพ	55
4.2.2 การสืบค้นข้อมูลรูปภาพโดยใช้คำสำคัญ	59
4.2.3 ผลการทดลองสืบค้น	61
4.2.4 การเปรียบเทียบผลงานวิจัยกับงานวิจัยอื่น	64
5 สรุปผลการวิจัย	67
5.1 สิ่งที่ค้นพบในการศึกษาวิจัยและบทวิจารณ์	67
5.2 สิ่งที่ได้รับจากการศึกษาวิจัย	67
5.3 ข้อจำกัดและอุปสรรค์ในการศึกษาวิจัย	68
5.4 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต	68
5.5 สรุปผลการวิจัย	69
บรรณานุกรม	71
ภาคผนวก	74

สารบัญตาราง

หน้า

ตารางที่		
1.1	แผนการดำเนินงาน	4
2.1	ข้อมูลในโครงสร้างการอธิบาย RDF	12
2.2	ผลลัพธ์จาการค้นหาด้วย RDQL	14
2.3	แสดงจำนวนคำศัพท์ในฐานความรู้เวิร์ดเนต	22
3.1	รายละเอียดของคลาสในระดับ 2 ในออนโทโลยี	40
3.2	รายละเอียดของคลาสในระดับ 3 ในออนโทโลยี	41
3.3	รายละเอียดของคุณสมบัติของชนิดข้อมูลในออนโทโลยีสถานที่ท่องเที่ยว	43
3.4	การอธิบายคลาสด้วยภาษา OWL	44
4.1	คำและวลีสืบค้นที่ใช้การทดลอง	62
4.2	เกณฑ์การแปลความหมายข้อมูลและพิจารณา จากค่า Precision และ Recall	62

BANGKOK UNIVERSITY FOUNDED 1962

สารบัญภาพ

หน้า

ภาพที่

2.1 Semantic Web Layer cake 8

2.2 ตัวอย่างไวยากรณ์ภาษา XML 9

2.3 ส่วนประกอบของ Triple 11

2.4 แสดงโครงสร้างการอธิบาย RDF 12

2.5 ตัวอย่างของเอกสาร RDF 12

2.6 ตัวอย่างเอกสาร RDFS 13

2.7 ตัวอย่างไวยากรณ์ของ RDQL 14

2.8 ตัวอย่างการกำหนด Namespace 18

2.9 ตัวอย่าง Ontology Header 18

2.10 ตัวอย่างการกำหนดคลาสและคลาสย่อย 19

2.11 การกำหนดความสัมพันธ์ระหว่างทรัพยากร (Resource) 20

2.12 แสดงความสัมพันธ์ระหว่างกลุ่มข้อมูลในฐานข้อมูลเกี่ยวข้องกับคำสืบค้น 23

2.13 แสดงความสัมพันธ์ระหว่างกลุ่มของข้อมูลที่ใช้สืบค้นกับคำค้น 23

2.14 แสดงสมการการคำนวณในการหาค่า Recall 24

2.15 แสดงสมการการคำนวณในการหาค่าของ Precision 24

2.16 การอธิบายโครงสร้างรูปภาพภาษาธรรมชาติด้วยกราฟ 26

2.17 ตัวอย่าง Ontology คำศัพท์ธรรมชาติ 26

2.18 แสดงขั้นตอนการสืบค้น 28

2.19 การคำนวณน้ำหนักของคำสืบค้น 29

2.20 ลำดับขั้นตอนการสืบค้นข้อมูล 30

2.21 แสดงภาพรวมของออนโทโลยีระดับบนของออนโทโลยี 33

2.22 คลาส Medicine 34

2.23 คลาส Symptoms 35

สารบัญภาพ (ต่อ)

หน้า

ภาพที่

3.1	แบบจำลองการสืบค้นรูปภาพบนเว็บ	36
3.2	ภาพรวมของโครงสร้างข้อมูลออนโทโลยีสถานที่ท่องเที่ยวในประเทศไทย	39
3.3	ตัวอย่างการกำหนด Namespace ด้วยภาษา OWL	43
3.4	ตัวอย่างการอธิบายคลาสด้วย OWL	44
3.5	แสดงขั้นตอนการนำเข้าข้อมูล	45
3.6	ตัวอย่างรูปภาพเกาะพีพี	47
3.7	ตัวอย่างคำบรรยายรูปภาพเกาะพีพี	47
3.8	ตัวอย่างการอธิบายโครงสร้างรูปภาพด้วย RDF	48
3.9	อธิบายข้อมูลด้วยมาตรฐาน RDFS	49
3.10	ขั้นตอนการสืบค้นข้อมูล	50
3.11	ตัวอย่างการใช้งานโปรแกรม Protégé ในการสร้างออนโทโลยี	52
3.12	ตัวอย่างการใช้งานโปรแกรม Sesame	52
3.13	สมการในการวัดความแม่นยำ (Precision) และการเรียกซ้ำ (Recall)	54
4.1	แสดงหน้าจอโปรแกรม Web Crawler	56
4.2	แสดงการเก็บข้อมูลของโดยการกระจายแบบวิธี Breadth first search	57
4.3	ข้อมูลที่ได้จากการเก็บรวบรวมของ Web Crawler	58
4.4	แสดงผลลัพธ์ที่ได้หลังจากผ่านการทำความสะอาดข้อมูล	59
4.5	หน้าจอแสดงการสืบค้นข้อมูลรูปภาพ	59
4.6	แสดงผลลัพธ์จากการสืบค้นข้อมูล	60
4.7	แสดงรายละเอียดเพิ่มเติมเกี่ยวกับข้อมูลรูปภาพ	61
4.8	แสดงข้อจำกัดในการเก็บรวบรวมข้อมูล	64
ก-1	หน้าจอโปรแกรม Protégé	74
ก-2	การสร้างคลาสและซับคลาส	76

สารบัญภาพ (ต่อ)

หน้า

ก-3 ลำดับชั้นของคลาสที่สร้างตามออนโทโลยีที่ออกแบบไว้ 77

BANGKOK UNIVERSITY FOUNDED 1962

บทที่ 1
บทนำ

1.1 ชื่อโครงการ

ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย

The Image Search Engine Using Semantic Web

1.2 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน World Wide Web (WWW) หรืออินเตอร์เน็ต ได้รับความนิยมเป็นอย่างสูง ซึ่งเข้ามาเปลี่ยนแปลงช่องทางในการติดต่อสื่อสารของมนุษย์ และก่อให้เกิดการดำเนินธุรกิจในรูปแบบใหม่เกิดขึ้น

ในรายงานของ Internet World Stats (Internet World stats, internet, 2008) พบว่ามีจำนวนผู้ใช้อินเตอร์เน็ตสูงขึ้นถึง 6 พันล้านคน เมื่อเทียบกับสถิติจำนวนผู้ใช้อินเตอร์เน็ตในปี 2000 มีอัตราการเติมโตเพิ่มขึ้น 300.5% ด้วยจำนวนเอกสารบนเว็บและจำนวนผู้ใช้มีอัตราการเพิ่มขึ้นอย่างรวดเร็ว ปัญหาที่เกิดขึ้นตามมาคือ จำนวนเอกสารที่เสิร์จเอ็นจินที่มีอยู่ในปัจจุบัน เช่น Google, Yahoo และAlta Vista ค้นหามาได้ในแต่ละครั้งมีจำนวนมาก ผลลัพธ์จากการสืบค้นมีทั้งข้อมูลที่ต้องการและไม่ต้องการ ผู้ใช้งานต้องใช้เวลาในการคัดเลือกเอาเฉพาะข้อมูลที่ต้องการอีกครั้งหนึ่งจนกว่าจะพบข้อมูลที่ผู้ใช้ต้องการ ทั้งนี้ เนื่องจากเทคนิคในการสืบค้นของเสิร์จเอ็นจิน ส่วนใหญ่เหล่านี้ ตั้งอยู่บนพื้นฐานของการค้นหาคำ (ที่ผู้ใช้ต้องการสืบค้น) ที่สอดคล้องหรือเหมือนกันคำหลัก (Keyword-based Matching) ที่มีอยู่ในเอกสาร เพราะคำหลักใส่ไว้ในเอกสารโดยเจ้าของแต่ละเว็บไซด์ อาจบรรจุคำหลักที่เหมือนกันแต่มีความหมายที่แตกต่างกัน (Lee & Tsai, 2003)

เช่นเดียวกันกับการสืบค้นรูปภาพที่อยู่บนเอกสารเว็บ ที่มีจำนวนรูปภาพมากมายบนอินเตอร์เน็ต ในการสืบค้นรูปภาพโดยการใช้เสิร์จเอ็นจิน เช่น Google Image และ Yahoo image search จะใช้วิธีการค้นหาจากคำอธิบายรูปภาพ ผลลัพธ์ที่ได้มีทั้งรูปภาพที่ผู้ใช้ต้องการและไม่

ต้องการจำนวนมาก ยกตัวอย่างเช่น ผู้ใช้ต้องการสืบค้นรูปภาพของผลแอปเปิล โดยใช้คำหลักว่า “Apple” ผลลัพธ์ที่ได้จะมีทั้งรูปภาพของผลแอปเปิลที่เราต้องการ แต่ก็จะมีรูปภาพของ Apple Logo, Apple Ipod, Apple Iphone, apple pie, Fiona Apple และอีกมากมายรวมอยู่ในผลลัพธ์ของการสืบค้นด้วย (Jing, Wang, Yao, Deng, Zhang, & Ying Ma, 2006)

ดังนั้นจะเห็นได้ว่า การค้นหารูปภาพโดยการใช้คำหลัก (Keyword-based) จากคำอธิบายรูปภาพซึ่งเป็นวิธีการที่ยังมีข้อจำกัด คือการค้นหาด้วยคำหลักให้ผลลัพธ์ไม่ตรงกับความต้องการ เนื่องจากการสืบค้นด้วยคำหลักเป็นการเปรียบเทียบคำที่ผู้ใช้สืบค้นกับข้อมูลคำอธิบายของรูปภาพ และการแทนข้อมูลของรูปภาพด้วยคำหลักไม่สามารถแทนรายละเอียดของรูปภาพได้ทั้งหมดจึงทำให้ประสบปัญหาในการสืบค้น ผลลัพธ์ที่ได้ตรงกับความต้องการในระดับต่ำ (Hyvonen, Styrman, & Saarela, 2002)

จากปัญหาการสืบค้นดังกล่าว งานวิจัยนี้จึงให้ความสนใจในการเพิ่มประสิทธิภาพในการสืบค้นรูปภาพบนอินเตอร์เน็ตและให้ได้ผลลัพธ์จากการสืบค้นตรงกับความต้องการของผู้ใช้ เพื่อผลประโยชน์โดยตรงต่อผู้ใช้ในเรื่องของเวลาและได้ข้อมูลที่ประโยชน์ โดยการนำเอาแนวความคิดเว็บเชิงความหมาย (Semantic Web) ที่มีวัตถุประสงค์เพื่อให้คอมพิวเตอร์สามารถเข้าใจความหมายของเอกสารบนอินเตอร์เน็ตและสามารถเชื่อมโยงความสัมพันธ์ของเอกสารได้ จึงได้มีการนำพัฒนาเทคโนโลยี RDF (Resource Description Framework) เพื่อใช้ในการอธิบายรายละเอียดและความหมายของทรัพยากรต่างๆ บนอินเตอร์เน็ต เป็นมาตรฐานที่รับรองจากองค์กร W3C (World Wide Web Consortium)

นอกจากนี้ยังได้นำเอาโครงสร้างแนวความคิดของ Ontology เพื่อออกแบบการนำเข้าข้อมูลจากข้อความบรรยายภาพโดยใช้ภาษาธรรมชาติและทำการออกแบบการขั้นตอนสืบค้นข้อมูลเพื่อรองรับการสืบค้นเชิงความหมาย เป็นการจัดกลุ่มของรูปภาพที่เกี่ยวข้องกัน และเป็นการจัดการความสัมพันธ์ของกลุ่มของรูปภาพ เพื่อให้มีประสิทธิภาพในการค้นหารูปภาพที่แม่นยำและมีประสิทธิภาพมากขึ้น

1.3 วัตถุประสงค์ของการวิจัย

1.3.1 เพื่อศึกษาสถาปัตยกรรมระบบการสืบค้นข้อมูลรูปภาพบนอินเตอร์เน็ต

1.3.2 เพื่อพัฒนาตัวแบบของการอธิบายรายละเอียดของรูปภาพ

1.3.3 เพื่อออกแบบและพัฒนาคำอธิบายข้อมูลออนโทโลยีเพื่อแก้ปัญหาความขัดแย้งกันทางความหมายของข้อมูล

1.3.4 เพื่อออกแบบและพัฒนาแบบจำลองสำหรับระบบการสืบค้นรูปภาพที่สอดคล้องกับสถาปัตยกรรมที่ได้ออกไว้

1.4 ขอบเขตของการวิจัย

1.4.1 งานวิจัยนี้จะศึกษาถึงปัญหาและข้อจำกัดของการสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้คำอธิบายความหมาย

1.4.2 เป็นการสร้างระบบสืบค้นข้อมูลรูปภาพ ซึ่งจะค้นหาภายใต้โดเมนการท่องเที่ยวแห่งประเทศไทย

1.5 ประโยชน์ที่คาดว่าจะได้รับการวิจัย

1.5.1 เป็นการศึกษาแนวทางที่จะพัฒนาระบบสืบค้นข้อมูลรูปภาพเพื่อไปใช้งานจริงต่อไป

1.5.2 สามารถนำไปเป็นกรณีศึกษาและแนวทางในการพัฒนาการสืบค้นรูปภาพบนเว็บด้วยวิธีการอื่น

1.6 แผนการดำเนินงาน

1.6.1 ศึกษาและกำหนดขอบเขตเป้าหมาย

1.6.2 ศึกษาขั้นตอนการทำตัวแบบการสืบค้นรูปภาพเชิงความหมาย

1.6.3 ศึกษาเครื่องมือในการเก็บรวบรวมรูปภาพและคำบรรยายรูปภาพ

1.6.4 เตรียมข้อมูลรูปภาพและแปลงรูปแบบข้อมูลให้ถูกต้องพร้อมสำหรับการสืบค้น

1.6.5 การสร้างตัวแบบการสืบค้นรูปภาพเชิงความหมาย

1.6.6 ทดสอบและทำการประเมินผล

1.6.7 จัดทำคู่มือและเอกสารต่างๆ

ตารางที่ 1.1 แผนการดำเนินงาน

รายละเอียดการดำงาน	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
1. ศึกษาและกำหนดขอบเขตเป้าหมาย										
2. ศึกษาขั้นตอนการทำตัวแบบการสืบค้นเชิงความหมาย										
3. ศึกษาเครื่องมือในการรวบรวมข้อมูลรูปภาพ										
4. เตรียมข้อมูลและการแปลงรูปแบบข้อมูล										
5. การสร้างตัวแบบการสืบค้นรูปภาพเชิงความหมาย										
6. ทดสอบและประเมินผล										
7. จัดทำคู่มือและเอกสารต่างๆ										

1.7 โครงสร้างของงานวิจัย

ในงานวิจัยนี้จะแบ่งเนื้อหาออกในลักษณะเป็นบทซึ่งแต่ละบทประกอบด้วยเนื้อหาดังนี้

บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง จะอธิบายถึงความรู้พื้นฐานซึ่งใช้เป็นข้อมูลสำหรับงานวิจัย โดยจะกล่าวถึงมาตรฐานสำหรับการอธิบายข้อมูล รวมถึงงานวิจัยที่เกี่ยวข้องในการออกแบบโครงสร้างข้อมูลและการสืบค้นข้อมูล

บทที่ 3 การดำเนินงานวิจัย เป็นการนำเสนอเกี่ยวกับวิธีการและขั้นตอนการทำงานวิจัยซึ่งประกอบด้วยการนำเข้าข้อมูล การออกแบบโครงสร้างสำหรับการอธิบายรูปภาพและการออกแบบขั้นตอนการสืบค้นข้อมูลเพื่อแสดงให้เห็นการทดลองตามแนวความคิดที่ได้ตั้งขึ้น

บทที่ 4 การวิเคราะห์ผล เป็นการนำเสนอการสรุปผลจากการทดลองและแสดงผลการเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่นรวมถึงการอธิบายความรู้ที่ได้จากงานวิจัยนี้

บทที่ 5 สรุปผลและข้อเสนอแนะ เป็นการนำเสนอความรู้ที่ได้รับจากงานวิจัยนี้ เพื่อเป็นความรู้ที่จะนำไปออกแบบการสืบค้นข้อมูลเพื่อให้ได้ผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้และเป็นแนวทางในการพัฒนางานให้ได้ผลลัพธ์ดียิ่งขึ้นไปในอนาคต

BANGKOK UNIVERSITY FOUNDED 1962

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

บทนี้จะกล่าวถึงเนื้อหา 2 ส่วน คือ วรรณกรรมรวมถึงทฤษฎีที่เกี่ยวข้อง และงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้

2.1 วรรณกรรมที่เกี่ยวข้อง

2.1.1 วิธีการสืบค้นสารสนเทศบนเว็บ

วิธีการสืบค้นข้อมูลบนเว็บแบ่งออกได้เป็น 3 ประเภท ดังนี้

2.1.1.1 เสิร์จเอ็นจิน (Search Engine) เป็นระบบซอฟท์แวร์ที่มีซอฟต์แวร์หุ่นยนต์ (Robot Software) ซึ่งบางครั้งเรียกว่า Spider หรือ Crawler ทำหน้าที่ท่องไปในเว็บไซต์ต่างๆ ในอินเตอร์เน็ตเพื่อรวบรวมเอกสารบนเว็บ เพื่อนำมาสร้างเป็นฐานดัชนีสำหรับการสืบค้นเอกสารบนเว็บ เพื่อทำให้การสืบค้นกระทำได้อย่างรวดเร็ว ดังนั้นกระบวนการจัดเตรียมดัชนี จึงเป็นขั้นตอนสำคัญอย่างยิ่งสำหรับการสืบค้นข้อมูลบนเว็บ เนื่องจากการเตรียมข้อมูลหรือดัชนีสำหรับการสืบค้นเป็นกระบวนการที่จัดเตรียมโดยใช้ซอฟต์แวร์ ทำให้สามารถอัพเดตข้อมูลได้ตลอดเวลา นั่นคือสามารถส่งซอฟต์แวร์หุ่นยนต์ออกไปรวบรวมเอกสารบนเว็บได้ตลอดเวลา ตัวอย่างของเสิร์จเอ็นจินที่มีอยู่ในปัจจุบันและเป็นที่นิยมคือ Google อย่างไรก็ตาม วิธีการจัดทำดัชนีด้วยวิธีการนี้ ความถูกต้องของการสืบค้นจะขึ้นกับฐานข้อมูลดัชนี และวิธีการสืบค้น

2.1.1.2 ไดเร็กทอรี (Directory) แตกต่างจากเสิร์จเอ็นจินตรงที่ไดเร็กทอรีมีการแบ่งแยกหมวดหมู่ของเอกสารโดยคน เว็บไซต์ใดที่ต้องการมีรายชื่อในไดเร็กทอรีต้องติดต่อผู้ดูแลไดเร็กทอรี เพื่อให้ผู้ดูแลไดเร็กทอรีจำแนกแยกประเภทของเอกสารบนเว็บนั้นๆให้อยู่ในหมวดหมู่ที่เหมาะสม เนื่องจากไดเร็กทอรี ถูกจัดเป็นหมวดหมู่โดยมนุษย์ ซึ่งจะมีการอ่านและพิจารณาแยกหมวดหมู่ได้ถูกต้องกว่าคอมพิวเตอร์ การค้นหาเอกสารบนเว็บด้วยไดเร็กทอรีจึงให้ผลลัพธ์ที่ตรงกับความต้องการมากกว่าเสิร์จเอ็นจิน แต่ด้วยเหตุผลไดเร็กทอรีมีการจัดแบ่งหมวดหมู่โดยคน ทำให้จำนวนเอกสารของไดเร็กทอรีมีน้อยกว่าเอกสารที่มีอยู่ในเสิร์จเอ็นจิน ตัวอย่าง ไดเร็กทอรีที่มีอยู่ในปัจจุบันคือ Yahoo

2.1.1.3 ไฮบริดเสิร์จเอ็นจิน (Hybrid Search Engine) เป็นเสิร์จเอ็นจินที่เป็นการผสมผสานการทำงานของเสิร์จเอ็นจินและไดเร็คทอรี่เข้าด้วยกัน มีข้อดีของทั้งเสิร์จเอ็นจินและไดเร็คทอรีรวมอยู่ด้วยกัน

2.1.2 แนวความคิดเว็บเชิงความหมาย (Semantic Web)

แนวความคิดของเว็บเชิงความหมาย (Semantic Web) เป็นแนวความคิดของ Sir Tim Berners-Lee ซึ่งเป็นบุคคลที่คิดค้น WWW, URIs, HTTP และ HTML โดยได้แรงสนับสนุนจากองค์กร W3C (World Wide Web Consortium) โดยมีจุดมุ่งหมายเพื่อช่วยให้สามารถค้นหาข้อมูลบนอินเตอร์เน็ตได้อย่างมีประสิทธิภาพมากขึ้นและยังสามารถสร้างความสัมพันธ์ให้กับข้อมูลที่มีมาจากแหล่งข้อมูลที่ต่างกัน ซึ่งจะก่อให้เกิดเป็นฐานข้อมูลที่เชื่อมโยงกันทั่วโลก

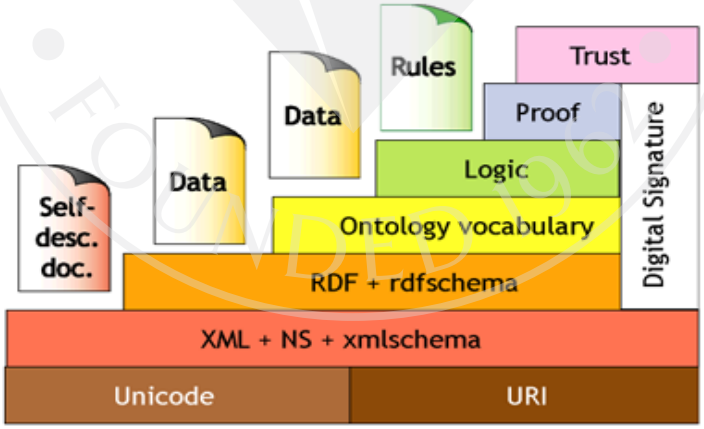
ปัจจุบันนี้ทรัพยากรที่อยู่บน WWW (World Wide Web) นั้นมีจำนวนมากมายมหาศาล ซึ่งก่อให้เกิดปัญหาที่สำคัญคือความแม่นยำในการสืบค้นด้วย Search Engine ลดลง กล่าวคือผลลัพธ์ที่ได้จากการสืบค้นด้วย Search Engine มีจำนวนมาก มีทั้งผลลัพธ์ที่ผู้ใช้ต้องการและผลลัพธ์ที่ผู้ใช้ไม่ต้องการ ส่งผลให้ผู้ใช้ต้องใช้เวลาเลือกเอกสารที่ตนเองต้องการ ที่เป็นเช่นนี้เนื่องจากเอกสาร HTML ไม่สามารถอธิบายความหมายสำหรับข้อมูลที่อยู่ภายใน HTML ได้และไม่สามารถเชื่อมโยงความสัมพันธ์ของข้อมูลได้

การที่จะทำให้แนวความคิดของเว็บเชิงความหมายเกิดขึ้นได้จริงนั้น จำเป็นต้องมีโครงสร้างของข้อมูลและหลักเกณฑ์ที่ดีสำหรับ Spider หรือ Crawling ที่จะการท่องไปตามเว็บต่างๆ เพื่อเก็บข้อมูลบนเว็บไซด์ สามารถเข้าใจความหมายของข้อมูลและเชื่อมโยงความสัมพันธ์ของข้อมูลได้ ดังนั้น XML และ RDF จึงเป็นเทคโนโลยีที่มีความสำคัญสำหรับแนวความคิดเว็บเชิงความหมาย โดยจะมีการใช้ XML ในการอธิบายโครงสร้างของข้อมูลและใช้ RDF ในการอธิบายความหมายของข้อมูล

อย่างไรก็ตามแนวความคิดเว็บเชิงความหมายยังอยู่ในระหว่างการศึกษาวิจัยและพัฒนา เนื่องจากยังมีความท้าทายหลายประการ อาทิ จะพัฒนาความร่วมมือระหว่างผู้พัฒนาเว็บไซด์ในการเตรียมข้อมูลให้เป็นไปตามมาตรฐานของ Semantic Web อย่างไร และจะจัดการกับปัญหาเกี่ยวกับการละเมิดลิขสิทธิ์ได้อย่างไร เมื่อข้อมูลทุกอย่างทุกเชื่อมโยงเข้าด้วยกัน

เทคโนโลยีเว็บเชิงความหมาย มีองค์ประกอบที่ถูกแบ่งออกเป็นระดับชั้น ดังภาพที่ 2-1 โดยข้อมูลเชิงความหมาย (Semantic-based Information) คือข้อมูลที่ถูกอธิบายด้วยการกำหนดความหมายไว้เป็นอย่างดี (Well-Defined Meaning) และข้อมูลเหล่านั้นสามารถอ้างอิงด้วย URI/URL และถูกอธิบายด้วยภาษาเชิงความหมาย (Semantics-Based language) เช่น XML, RDF และ RDFS โดยในการอธิบายข้อมูลเชิงความหมาย สามารถกำหนดคำ (Vocabulary) คุณสมบัติ (Property) และเงื่อนไข (Constraint) ต่างๆ สำหรับการอธิบายข้อมูลซึ่งถูกกำหนดไว้ในออนโทโลยี (Ontology) มาอธิบาย ในการตีความข้อมูลเชิงความหมายต้องใช้ตรรกะ (Logic) ต่าง ๆ มาพิจารณาในการตีความกระบวนการของการตีความได้ถูกจัดให้อยู่ในชั้นของการพิสูจน์ (Proof) ซึ่งใช้การอนุมาน (Inference) เป็นกลไกของการทำงาน และในชั้น Trust เป็นการสร้างความเชื่อถือได้ในการเผยแพร่และการนำข้อมูลไปใช้งาน อีกทั้งมีการกำหนดการรักษาความปลอดภัยให้กับข้อมูลที่มีการรับส่งระหว่างโปรแกรมประยุกต์ด้วยเทคโนโลยี Digital signature

ภาพที่ 2-1: Semantic Web Layer Cake



ที่มา: Tim Berners-Lee (Copyright 2000). Semantic Web on XML. สืบค้นวันที่ 20 สิงหาคม 2551 จาก http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html

2.1.3 XML (eXtensible Markup Language)

ภาษา XML (eXtensible Markup Language) เป็นส่วนหนึ่งของ SGML (the Standard Generalized Markup Language) โดยพัฒนาให้สามารถใช้งานได้ง่ายกว่าและมีความยืดหยุ่นมากกว่า โดย XML ถูกนำเสนอในปี ค.ศ.1996 ภายใต้การดูแลของ W3C (World Wide Web Consortium) ภาษา XML เป็น Meta-Language ที่ใช้แท็กในการสื่อความหมายของข้อมูล ซึ่งผู้พัฒนาสามารถกำหนดแท็กขึ้นใช้งานได้เองตามความต้องการ

เริ่มแรกภาษา XML ถูกพัฒนาขึ้นมาเพื่อแก้ไขปัญหาของภาษา HTML แต่ต่อมาได้เพิ่มขอบเขตความสามารถออกไป ได้แก่ สามารถแลกเปลี่ยนข้อมูลกันได้โดยไม่ขึ้นกับ Platform ใด Platform หนึ่ง ทำให้คอมพิวเตอร์สามารถติดต่อสื่อสารเพื่อแลกเปลี่ยนข้อมูลได้อย่างไร้ขีดจำกัด และต้องการให้ภาษา XML เป็นภาษามาตรฐานของเว็บที่ใช้กันทั่วโลกเพื่อให้ผู้พัฒนาที่เป็นรูปแบบเดียวกัน

ภาพที่ 2-2: ตัวอย่างไวยากรณ์ภาษา XML

```
<?xml version = “1.0” encoding = window-874” ?>
<!— filename = detailBook.xml -->
<INVENTORY>
        <BOOK ISBN=”974-94427-9-2”>
                <TITLE>A Semantic Web Primer</TITLE>
                <AUTHOR>Grioris Antoniou and Frank van Hermelen</AUTHOR>
                <PAGE>324</PAGE>
                <PRICE>500</PRICE>
        </BOOK>
<?INVENTORY>
```

2.1.4 RDF (Resource Description Language)

สิ่งที่สำคัญของแนวความคิดเว็บเชิงความหมาย (Semantic Web) คือการทำให้เครื่องคอมพิวเตอร์สามารถเข้าใจความหมายของข้อมูลได้

RDF เป็นมาตรฐานโครงสร้างการอธิบาย Metadata บนเว็บ ที่ถูกพัฒนาขึ้นมาโดยองค์กร W3C (World Wide Web Consortium) ซึ่งมีจุดมุ่งหมายที่จะพัฒนา RDF ให้เป็นส่วนสำคัญของ Semantic Web และเพื่อง่ายต่อการใช้งานในระบบที่ต่าง Platform กัน RDF จึงเขียนให้อยู่ในรูปแบบโครงสร้างของภาษา XML (World Wide Web Consortium, 2002)

ลาสสิร่าและสวิก (Lassila & Swick, 1999) ได้ทำการวิจัยโดยกล่าวว่า RDF สามารถนำไปใช้ประโยชน์ในหลากหลายด้าน ยกตัวอย่างเช่น การสืบค้นเอกสารบนเว็บ ด้วย Search Engine ได้อย่างมีประสิทธิภาพมากขึ้น, สามารถอธิบายข้อมูลและความสัมพันธ์ระหว่างข้อมูลโดยเฉพาะเว็บไซต์, เว็บเพจ โดย Intelligent Software Agent ซึ่งทำให้ง่ายต่อการแบ่งปันข้อมูลและการแลกเปลี่ยนข้อมูลซึ่งกันและกัน

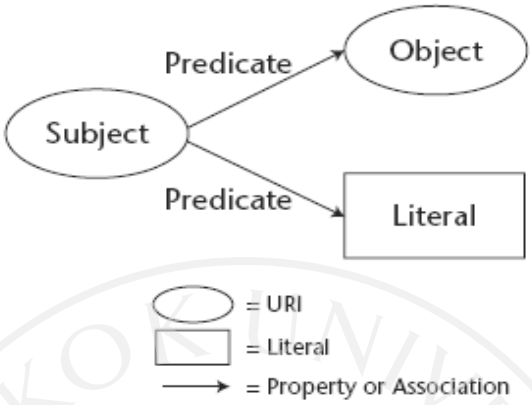
แนวความคิดพื้นฐานของ RDF ประกอบไปด้วย 3 ส่วนประกอบไปด้วย Resources, Properties, Statements ซึ่งแต่ละส่วนสามารถอธิบายแต่ละส่วนได้ดังนี้

2.1.4.1 Resources คือ สิ่งใดๆ ที่ต้องการจะอธิบาย โดยที่ทุกๆ Resource จะต้องถูกระบุด้วย URI (Universal Resource Identifier) ซึ่ง URI ที่ว่านี้อาจจะเป็น URL (Web Address) หรือสิ่งที่ระบุให้ Resource นั้นจะต้องมีคุณสมบัติที่ไม่ซ้ำกัน

2.1.4.2 Properties คือ การอธิบายคุณสมบัติหรือความสัมพันธ์ระหว่าง Resource ซึ่ง Properties จะต้องถูกระบุด้วย URI เช่นกัน

2.1.4.3 Statements คือ RDF จะมีการอธิบายข้อมูลในรูปแบบที่เรียกว่า “Triple” ประกอบด้วย Subject, Predicate และ Object

ภาพที่ 2-3: ส่วนประกอบของ Triple



ที่มา: Daconta, M., Obrst, L., & Smith, K. (2003). The Semantic Web. Indiana:Wiley Publishing

จากภาพที่ 2-3 แสดงส่วนประกอบของ Triple ประกอบไปด้วย

1) Subject คือ ทรัพยากรที่อธิบาย แทนที่ด้วยสัญลักษณ์วงรี ซึ่งแทนด้วย URI
2) Predicate คือ คุณสมบัติของทรัพยากรแทนด้วยลูกศรชี้จากส่วน Subject ชี้ไปยังส่วนของ Object หรือ Literal
3) Object คือ ค่าของคุณสมบัติ แทนด้วยสัญลักษณ์ที่เป็นวงรี ซึ่งแทนที่ด้วย URI แต่หากเป็นค่าของอักขระจะเรียกว่า Literal ซึ่งจะมีลักษณะเป็นสัญลักษณ์รูปสี่เหลี่ยม

ตัวอย่าง การแทนที่ข้อมูลในโครงสร้างการอธิบายข้อมูลของ RDF

John Lennon is the creator web page http://www.bu.ac.th/lennon/XML

http://www.bu.ac.th/lennon/XMLhas title Programming XML in Java

http://www.bu.ac.th/lennon/XMLhas date 2008-10-01

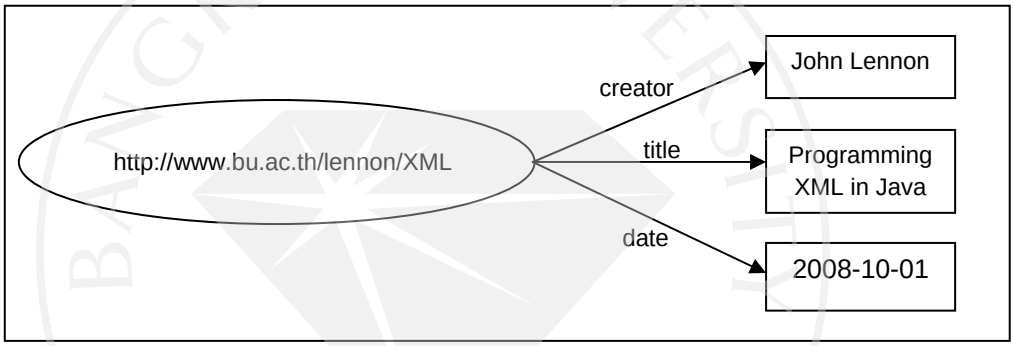
สามารถแทนข้อมูลในโครงสร้างการอธิบาย RDF ดังตารางที่ 2-1

ตารางที่ 2-1: ข้อมูลในโครงสร้างการอธิบาย RDF

Subject	Predicate	Object
http://www.bu.ac.th/lennon/XML	Creator	John Lennon
http://www.bu.ac.th/lennon/XML	Title	Programming XML in Java
http://www.bu.ac.th/lennon/XML	Date	2008-10-01

สามารถแทนที่ข้อมูลในโครงสร้างการอธิบาย RDF ได้ดังกราฟดังภาพที่ 2-4

ภาพที่ 2-4: แสดงโครงสร้างการอธิบาย RDF



จากกราฟสามารถแทนข้อมูลเป็น RDF Document ตาม Syntax ของ RDF ได้ดังภาพที่ 2-5

ภาพที่ 2-5: ตัวอย่างของเอกสาร RDF

```
<?xml version=”1.0” encoding =”UTF-8”?>
<rdf:RDF xmlns:pwterms =”http://www.bu.ac.th/terms#”
        xmlns:rdf=”http://www.w3c.org/1999/02/22-rdf-syntax-ns#”>
  <rdf:Description rdf:about=”http://www.bu.ac.th/lennon/XML
        <pwterms:creator>John Lennon</pwterms:creator>
        <pwterms:title>Programming XML in Java</pwterms:title>
        <pwterms:date>2008-10-01</pwterms:date>
  </rdf:Description>
</rdf:RDF>
```

2.1.5 RDFS (RDF Schema)

RDF Schema เป็นส่วนต่อขยายจาก RDF ใช้ในการอธิบายโครงสร้างของ Metadata ของทรัพยากรบนเว็บโดยกำหนด Vocabulary เพื่ออธิบายโครงสร้างของ Metadata ซึ่งประกอบไปด้วย Property และค่าของ Property โดยอธิบายให้อยู่ในรูปของ Class และ Property โดยสามารถสืบทอดเป็น SubClass และ SubProperty ได้ เมื่อเอกสาร RDF มี RDF Schema กำกับจะช่วยให้เครื่องอนุมานข้อมูลได้ (Inference Engine)

RDF Schema เป็นส่วนที่ขยายต่อมาจาก RDF โดยสามารถอธิบายกลุ่มของทรัพยากรบนเว็บที่มีความเกี่ยวข้องกันและสามารถเชื่อมต่อความสัมพันธ์ระหว่างทรัพยากรบนเว็บ ยกตัวอย่าง เช่น “Pupil” เป็นกลุ่มประเภทของ “Student” และ “Student” ก็เป็นคลาสย่อยของ “People” เป็นต้น (Swartz, internet, 2002)

ภาพที่ 2-6: ตัวอย่างเอกสาร RDFS

```
<?xml version="1.0"?>
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

<rdf:Description rdf:ID="Animal">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
</rdf:Description>

<rdf:Description rdf:ID="Dog">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
   <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>

<rdf:Description rdf:ID="Cat">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
   <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>

<rdf:Description rdf:ID="Donkey">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
   <rdfs:subClassOf rdf:resource="#Animal"/>
</rdf:Description>

<rdf:Description rdf:ID="PersianCat">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
   <rdfs:subClassOf rdf:resource="#Cat"/>
</rdf:Description>

</rdf:RDF>
```

จากภาพที่ 2-6 เป็นการแสดงความสัมพันธ์ระหว่างคลาสต่างๆ โดยบรรยายโดยใช้ RDF/RDFS ในตัวอย่างนี้ เป็นการแสดงโดยความสัมพันธ์โดยระบุว่า สุนัข (Dog), แมว (Cat), และ ลา (Donkey) เป็นซับคลาสของ สัตว์ (Animal) ส่วน แมวเปอร์เซีย (Persian Cat) มีชนิดเป็นแมว ดังนั้นเมื่ออธิบายโดยใช้ RDFS ไว้เครื่องคอมพิวเตอร์ก็สามารถเข้าใจได้ว่า แมวเปอร์เซียอยู่ในชนิดของสัตว์

2.1.6 RDQL (RDF Data Query Language)

RDQL (RDF Data Query Language) เป็นภาษาในการสืบค้นข้อมูลบนโครงสร้างของ RDF ซึ่งคล้ายกับการดำเนินการของภาษาในการสอบถามข้อมูลแบบโครงสร้าง SQL ของฐานข้อมูลเชิงสัมพันธ์ โดยเมื่อรับคำสั่งมาจากโปรแกรมประยุกต์แล้ว RDQL จะดำเนินการสืบค้นข้อมูลพร้อมกับส่งผลลัพธ์ที่ตรงกับเงื่อนไขกลับมายังผู้ใช้งาน (Seaborne, internet, 2004)

ภาพที่ 2-7: ตัวอย่างไวยากรณ์ของ RDQL

```
Select ?carBrand
Where(?x, http://www.owl-ontologies.com/carsale.owl#carBrand,?
carBrand)
```

ดังภาพที่ 2-7 ในคำสั่ง RDQL นี้เป็นการค้นหา วัตถุอ้างอิง (x) และยี่ห้อของรถยนต์ (Car Brand) ซึ่งได้ผลลัพธ์ดังแสดงในตารางที่ 2-2

ตารางที่ 2-2: ผลลัพธ์จากการค้นหาด้วย RDQL

X	CarBrand
http://www.owl-ontologies.com/carsale.owl#/Brand01	Honda
http://www.owl-ontologies.com/carsale.owl#/Brand02	Toyota

2.1.7 ออนโทโลยี (Ontology)

ออนโทโลยีเป็นคำที่ใช้กันอย่างแพร่หลายในสาขาที่เกี่ยวข้องกับปัญญาประดิษฐ์ (AI) และวิศวกรรมความรู้ มีผู้ที่ให้คำนิยามของออนโทโลยีไว้หลายแบบ คำนิยามที่มีผู้นิยมใช้กันอย่างแพร่หลายคือคำนิยามของ กลูเบอร์ (Gruber, 1993) กล่าวว่า Ontology เป็นรายละเอียดที่ชัดเจนแน่นอนของแนวความคิด (An ontology is an explicit specification of a conceptualization) ดังนั้นเมื่อเราต้องการแสดงหรือระบุแนวความคิดของวัตถุหรือสิ่งของนั้นๆ เราควรจะแสดงรายละเอียดที่ชัดเจนและแน่นอน และคำนิยามของ กรันริโน่ (Guarino, 1998) กล่าวว่า Ontology คือการเรียนรู้เกี่ยวกับชนิดของสิ่งใดๆ ที่มีอยู่จริงซึ่งอาจจะมีหลายรูปแบบ แต่ที่จำเป็นและเห็นกันอยู่ชัดเจน จะรวมถึงคำศัพท์ของคำหรือสิ่งของและรายละเอียดที่เฉพาะเจาะจงความหมายของคำหรือสิ่งของนั้นๆ

แต่อย่างไรก็ตามคำนิยามของกลูเบอร์และกรันริโน่เป็นคำนิยามที่กว้างจนเกินไป เพื่อให้เข้าใจได้ง่ายขึ้น เซิกทัวร์ (Seartout, 1996) ได้ให้ความหมายที่ชัดเจนขึ้นของออนโทโลยีโดยกล่าวว่าออนโทโลยีคือกลุ่มคำที่มีโครงสร้างแบบลำดับชั้นสำหรับอธิบายขอบเขตและเนื้อหาที่สนใจ และสามารถใช้เป็นโครงสร้างพื้นฐานของความสัมพันธ์สำหรับใช้ในฐานความรู้ได้ จึงกล่าวได้ว่า Ontology เป็นระบบคำศัพท์ที่มีความสัมพันธ์ในเชิงความหมาย เช่น ใช้เป็นโครงร่างพื้นฐานในการอธิบายความรู้เฉพาะด้าน เช่น การเกษตร คอมพิวเตอร์ ออนโทโลยีนอกจากนี้ยังสามารถนำไปประยุกต์ใช้ในการทำงานของระบบงานต่างๆ เช่น ช่วยในการทำงานของระบบสืบค้นข้อมูล โดยช่วยขยายคำค้น ทำให้ค้นหาและเข้าถึงข้อมูลที่ผู้ใช้งานต้องการจริงๆ ออนโทโลยี ยังมีส่วนช่วยในการทำงานของระบบประมวลผลสารสนเทศให้มีประสิทธิภาพมากขึ้น เช่น การจัดรวมสารสนเทศ (Information Integration) การจำแนกเอกสาร (Document Classification) การทำความสะอาดข้อมูลในฐานข้อมูล (Data Cleaning in database system) และการสกัดข้อสนเทศ (Information Extraction)

เฟรนเซล (Fensel, 2003) ได้แบ่งออนโทโลยีออกเป็น 2 ประเภทใหญ่ๆ คือ

2.1.7.1 Real World Semantic Ontology คือออนโทโลยีที่แสดงความสัมพันธ์ทั่วไปของแนวความคิดเป็นจริงเสมอในทุกสถานการณ์ ความสัมพันธ์ที่มีในออนโทโลยีชนิดนี้มักจะเป็นความสัมพันธ์แบบ synonym, antonym, hypernym และ meronym ตัวอย่างออนโทโลยีชนิดนี้ เช่น WordNet, EuroWordNet

2.1.7.2 Formal Semantic Ontology คือออนโทโลยีที่ใช้ในการแทนความรู้เฉพาะด้าน เฉพาะงาน (Task-Oriented Ontology) จะเป็นจริงเสมอสำหรับระบบงานที่นำไปใช้ ในระบบงานอื่นๆ อาจมองในมุมมองอื่นได้ และความสัมพันธ์ที่มีจะหลากหลายขึ้นอยู่กับระบบงาน เช่น UMLS (Unified Medical Language System) ซึ่งเป็น ออนโทโลยีเกี่ยวกับทางการแพทย์ KA2 ontology เป็นออนโทโลยีทางด้านการจัดการความรู้ (Knowledge management)

ออนโทโลยีสามารถนำมาประยุกต์ใช้ในการทำงานของระบบงานต่างๆ โดยเฉพาะระบบสืบค้นข้อมูล ออนโทโลยีมีส่วนช่วยในการขยายคำค้น ทำให้ค้นหาและเข้าถึงข้อมูลที่ตรงตามที่ผู้ใช้งานต้องการจริง ด้วยการพิจารณาความหมายของสิ่งต่างๆ ในโดเมน

2) ออนโทโลยีต่างจากฐานข้อมูล ในด้านของการเก็บรายละเอียดต่างๆ เช่น บทบาท (Role) ความสัมพันธ์ (Relation) ในขณะที่ฐานข้อมูลเป็นเพียงแหล่งรวบรวมข้อมูล หากใช้ออนโทโลยีช่วยในการค้นหาข้อมูล จะช่วยให้ได้ข้อมูลข่าวสารที่ถูกต้องยิ่งขึ้น

3) ระบบงานที่ต้องการผนวกความรู้ (Knowledge Integration) เพื่อสร้างความเข้าใจเบื้องต้นของความรู้ระหว่างโดเมน

4) ระบบที่มีความต้องการนำข้อมูลกลับมาใช้อีก (Reuse) เช่นระบบงานหนึ่งอาจจะกำหนดออนโทโลยีเพื่อใช้งาน และมีการอนุญาตให้ระบบงานอื่น สามารถนำไปใช้งานได้อีก โดยที่ระบบที่นำไปใช้งานภายหลัง อาจกำหนดข้อมูลเชิงความหมายเพิ่มเติมได้โดยไม่ขัดแย้งกับข้อมูลเชิงความหมายของออนโทโลยีที่ถูกนำข้อมูลกลับมาใช้อีก

2.1.8 OWL (Web Ontology Language)

ถึงแม้ว่า RDF Schema เป็นภาษาที่สามารถอธิบายออนโทโลยี แต่เนื่องจากภาษา RDF/RDFS ยังมีข้อจำกัดที่สามารถสรุปได้ดังนี้

Local scope ของ properties เนื่องจากคำสั่ง rdf:range ใช้ในการกำหนด Range ของ Property ให้กับทุกคลาส เช่น Property “การกิน” มีโดเมนคือ ‘สัตว์’ และมี Range คือ ‘เนื้อสัตว์’ ดังนั้นคลาสที่เป็นคลาสย่อยของ ‘สัตว์’ จะต้องมีข้อกำหนดของ Property ตามนี้ ไม่สามารถกำหนดได้คลาส ‘วัว’ ซึ่งเป็นคลาสย่อยของสัตว์มี Range ของ Property ‘การกิน’ เป็น ‘พืช’

Disjoint ของคลาส ตัวอย่างเช่น คลาส ‘ผู้ชาย’ และ ‘ผู้หญิง’ เป็น Disjoint class กัน หมายถึงไม่มี instance ใดที่เป็นสมาชิกของทั้งคลาส “ผู้ชาย” และ “ผู้หญิง”

Boolean combination ของคลาส ในบางครั้งเราต้องการที่จะสร้างคลาสใหม่จากคลาสเดิม โดยการนำคลาสเดิมมา union, intersection หรือ complement กัน เช่น คลาส ‘คน’ เกิดจากการ union กัน ของคลาส ‘ผู้ชาย’ และ ‘ผู้หญิง’ ซึ่งไม่สามารถกำหนดได้ใน RDFS

Cardinality restriction การสร้างข้อกำหนดของคุณสมบัติให้มีค่าเป็นตัวเลข เช่น คลาส Person จะมีแม่ (Mother) ได้ 1 คนเท่านั้น ซึ่งไม่สามารถกำหนดได้ใน RDFS

ดังนั้นทางองค์กร W3C (World Wide Web Consortium) จึงได้ทำการพัฒนา OWL (Web Ontology Language) เพื่อเป็นส่วนขยายต่อจากภาษา RDF/RDFS ภาษา OWL จัดเป็นองค์ประกอบหนึ่งในงานเว็บเชิงความหมาย (Semantic Web) ที่ใช้ในการบรรยายข้อมูลเชิงความหมาย สามารถกำหนดโครงสร้างข้อมูลในลักษณะลำดับชั้น และอธิบายข้อมูล (Metadata) ที่มีความสัมพันธ์ในระบบฐานข้อมูลได้ รวมทั้งสามารถรองรับการบรรยายข้อมูลเชิงตรรกะ ชนิดข้อมูล และตัวบ่งปริมาณได้ ทำให้ข้อมูลที่ถูกแทนที่นั้นมีความหมายมากยิ่งขึ้น ลักษณะการบรรยายจะอยู่ในรูปของคลาส คุณสมบัติของคลาส และความสัมพันธ์ของคลาส เพื่ออธิบายเอ็นทิตี้ (Entity) และความสัมพันธ์ต่างๆ ที่เกิดขึ้น ซึ่งทางองค์กร W3C ได้แบ่งภาษา OWL ออกเป็น 3 กลุ่มคือ

- OWL Lite เหมาะสำหรับการเก็บข้อมูลความสัมพันธ์แบบระดับชั้น (Hierarchy) และมีข้อกำหนดอย่างง่าย
- OWL DL(Description Logic) เป็นภาษาที่รองรับคำสั่งทุกคำสั่งของภาษา OWL แต่จะต้องเขียนภายใต้ข้อกำหนด เช่น คลาสสามารถเป็นซับคลาสของคลาสอื่นๆ ได้มากว่าหนึ่งคลาส
- OWL Full เป็นภาษาที่มีความซับซ้อนมากที่สุด โดยมีลักษณะพิเศษ เช่น Syntax อิสระจาก RDF, คลาสสามารถเป็น Instance ได้

ทั้ง OWL DL และ OWL Full ต่างก็สนับสนุนเซตของภาษา OWL แต่มีข้อจำกัดของคุณลักษณะบางอย่างที่แตกต่างกันบนพื้นฐานของ RDFS โดย OWL Full จะมีการผสมผสานระหว่าง OWL และ RDFS โดยไม่มีการบังคับในส่วนการแบ่งคลาส การกำหนดคุณสมบัติ และค่าของข้อมูล ส่วน OWL DL จะมีข้อบังคับในการใช้ RDF การกำหนดคลาส การกำหนดคุณสมบัติ และค่าของข้อมูล เป็นต้น

โครงสร้างของภาษา OWL (Ontology Web Language)

1) Namespace

การกำหนด Namespace จะประกาศไว้ที่ส่วนเริ่มต้นของเอกสาร เพื่อเป็นการกำหนดกลุ่มในการอ้างอิงข้อมูลเอกสาร OWL ที่ถูกสร้างจะขึ้นอยู่กับโครงสร้างที่ถูกนิยามด้วย RDF/RDFS และชนิดข้อมูลของ XML Schema การเขียน Namespace จะประกาศไว้ภายใต้คำสั่ง rdf:RDF syntax ตัวอย่างในภาพที่ 2-8 แสดงการกำหนด Namespace

ภาพที่ 2-8: ตัวอย่างการกำหนด Namespace

```
xmlns=”http://www.owl-ontologies.com/car.owl#”
xmlns:rdf=”http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:rdfs=”http://www.w3.org/2000/01/rdf-schema#”
xmlns:owl=”http://www.w3.org/2002/07/owl#”
```

2) Ontology Headers

แสดงการอธิบายรายละเอียดเบื้องต้นของออนโทโลยี ว่าเป็นโครงสร้างข้อมูลเกี่ยวกับอะไร ภายใต้ Element <owl:Ontology rdf:about=””> ประกอบด้วยคำสั่ง <owl:versionInfo> ใช้แสดงรุ่นของข้อมูลที่สร้าง คำสั่ง <rdf:comment> ใช้แสดงการอ้างอิงเกี่ยวกับแหล่งข้อมูลว่ามาจากที่ใด ดังอย่างการแทนค่าใช้งานแสดงดังภาพที่ 2-9

ภาพที่ 2-9: ตัวอย่าง Ontology Header

```
<owl:Ontology rdf:about=”MyCarOntology”/>
<owl:Class rdf:ID=”Car”/>
```

3) การกำหนดคลาส

ในการอธิบายคลาสในออนโทโลยี จะมีคลาสเริ่มต้นคือ owl:Class โดยกำหนดให้ owl:Class เป็นคลาสใหญ่ที่สามารถครอบคลุม ทุกคลาสข้อมูลได้ ดังนั้นไม่ว่าผู้ใช้งานกลุ่มใดสร้างคลาสขึ้นมา จะเสมือนว่าเป็นสมาชิกอยู่ภายใต้คลาส owl:class

กำหนดคลาส “Island” เป็นคลาสย่อย (Subclass) ของคลาส “Natural Place” ความสัมพันธ์แบบคลาสย่อยนี้ทำให้เกิดคลาสทั่วไป (Generic class) และคลาสจำเพาะเจาะจง (Specific class) กล่าวคือ Natural Place เป็นคลาสที่แสดงความหมายของสถานที่ท่องเที่ยวเชิงธรรมชาติโดยทั่วไป ในขณะที่ Island คือคลาสที่แสดงความหมายที่จำเพาะเจาะจงว่าเป็นสถานที่ท่องเที่ยวที่เป็นเกาะ ซึ่งจากความสัมพันธ์ดังกล่าวนี้สามารถเขียนอธิบายด้วยภาษา OWL ได้ดังภาพที่ 2.10

ภาพที่ 2.10: ตัวอย่างการกำหนดคลาสและคลาสย่อย

```
<owl:Class rdf:ID=”NaturalPlace”/>
<owl:Class rdf:ID=”Island”/>
        <rdfs:subClassOf
rdf:resource=”#NaturalPlace”/>
```

4) การกำหนดคุณสมบัติ (Property)

การกำหนดคุณสมบัติของคลาสใน OWL สามารถกำหนดได้ 2 ประเภทคือ

- การกำหนดคุณสมบัติด้วย owl:DatatypeProperty เพื่อกำหนดการอธิบายคุณสมบัติของคลาสที่เป็นค่าชนิดพื้นฐาน เช่น การอธิบายข้อมูลราคาสินค้า

- การกำหนดคุณสมบัติด้วย owl:ObjectProperty เพื่อกำหนดการอธิบายข้อมูลซึ่งต้องการอธิบายคุณสมบัติของคลาส ซึ่งเป็น Resource หรือการกำหนดการเชื่อมโยงความสัมพันธ์ระหว่างคลาส 2 คลาส แสดง

5.) การกำหนดความสัมพันธ์ระหว่าง Resource

ทรัพยากร (Resource) หมายถึง คลาสที่มีการอ้างอิงได้ด้วยการกำหนด URL ซึ่งการกำหนดความสัมพันธ์ระหว่างทรัพยากรเป็นการกำหนดเงื่อนไข Domain และ Range ให้กับการอธิบายคุณสมบัติที่เชื่อมโยงระหว่าง Resource ดังตัวอย่างในภาพที่ 2-11 เป็นการอธิบายข้อมูลราคาด้วยคุณสมบัติ hasPrice ซึ่งสามารถกำหนดโดเมนคือ “Car” และ Range คือชนิดพื้นฐานซึ่งเป็นตัวเลขจำนวนเต็ม (Integer)

ภาพที่ 2-11: การกำหนดความสัมพันธ์ระหว่างทรัพยากร (Resource)

```
<owl:DatatypeProperty rdf:ID=”hasPrice”>
<rdfs:domain rdf:resource=”#Car”/>
<rdfs:range rdf:resource=”http://www.w3.org/2001/XMLSchema#int”/>
</owl:DatatypeProperty>
```

2.1.9 Lexical Chains

Lexical Chains เป็นวิธีการจัดหมวดหมู่ของคำที่มีความหมายอยู่ในกลุ่มเดียวกันจากประโยค บทความหรือวารสาร (Morris, & Hirst, 1991) เพื่อใช้ในการจับใจความสำคัญของบทความและหาโครงสร้างของบทความการสร้าง Lexical Chains นั้นมีอยู่ 2 ขั้นตอนหลักคือ ขั้นตอนการหาคำสำคัญ (Candidate Word) และขั้นตอนการสร้าง Lexical Chains

ขั้นตอนการหาคำสำคัญเป็นการเลือกคำสำคัญจากบทความซึ่งจะมีวิธีการเลือกคำสำคัญโดยการเลือกเอาเฉพาะคำที่สื่อความหมายได้ซึ่งได้แก่ คำนาม (Noun) ต่างๆ และจะไม่ในใจคำประเภทอื่นๆ เช่น คำสันธาน (Pronouns) คำสรรพนาม (Preposition) คำศัพท์ (Adjective) คำกริยาวิเศษ (Adverb) คำกริยาต่างๆ และคำศัพท์ที่ใช้บ่อยๆ ในประโยค แต่สื่อความหมายไม่ได้

ขั้นตอนการสร้าง Lexical Chains เป็นการนำคำสำคัญที่ได้จากขั้นตอนการเลือกคำสำคัญมาจัดกลุ่มโดยอาศัยฐานข้อมูล Thesaurus และฐานข้อมูล Word Net ที่มีการแยกประเภทและจัดกลุ่มคำศัพท์เป็นหมวดหมู่

2.1.10 ฐานความรู้ธีซอรัส (Thesaurus)

ธีซอรัส (นฤมล ปราชญโยธิน, ทวีศักดิ์ กออนันตกูล, เปรมิน จินดาวิมลเลิศ, 2536) เป็นคลังคำ แต่เป็นคลังคำของศัพท์ควบคุม คือ จะกำหนดหรือแนะให้ใช้ศัพท์คำใดคำหนึ่ง เป็นตัวแทนของกลุ่มคำที่มีรูปต่างกัน หรือเขียนต่างกันแต่มีความหมายเดียวกัน และจะประมวลศัพท์ที่อยู่ในกลุ่มเดียวกันไว้ด้วยกัน โดยมีการแสดงโครงสร้างความสัมพันธ์ของคำในกลุ่ม และมีสัญลักษณ์แสดงความสัมพันธ์กำกับไว้ภายใต้รูปแบบที่ได้มาตรฐาน โดยสัญลักษณ์ที่ใช้แทนความสัมพันธ์จะมีรูปแบบเฉพาะตัวมากกว่าบัญชีคำศัพท์ประเภทอื่น ๆ

วัตถุประสงค์ของธิซอรัส คือ จัดทำขึ้นเพื่อเป็นเครื่องมือที่มีประสิทธิภาพในการจัดเก็บและค้นคืนสารสนเทศด้วยวิธีการเข้าถึงสารสนเทศผ่านทางเนื้อหา (Subject access) คำว่าประสิทธิภาพในที่นี้ได้แก่ความสามารถในการสืบค้นข้อมูลเรื่องใดเรื่องหนึ่งได้อย่างครบถ้วน โดยไม่ต้องทราบคำศัพท์ที่เกี่ยวข้องทั้งหมดและความสามารถในการช่วยเหลือให้ผู้จัดทำดรรชนีและผู้สืบค้นใช้ศัพท์ที่เป็นมาตรฐานเดียวกัน ทำให้การกำหนดคำแทนสาระของเอกสารทีความเที่ยงตรง

2.1.11 ฐานความรู้เวิร์ดเนต

ฐานความรู้เวิร์ดเนตหรือฐานความรู้อภิธานเวิร์ดเนตเป็นฐานข้อมูลคำศัพท์ภาษาอังกฤษโดยการแบ่งการเก็บข้อมูลคำศัพท์ตามชนิดของคำ เช่น คำนาม คำกริยา คำวิเศษณ์ เป็นต้น การจัดโครงสร้างคำศัพท์ในฐานข้อมูลเวิร์ดเนตจะจัดตามความหมายของคำ (Word meaning) มากกว่ารูป

ของคำ (Word Form) โดยที่คำหนึ่งอาจมีได้หลายความหมายและคำหลายคำอาจมีความเดียวกัน (Chirstiane, 1998)

ตารางที่ 2-3: แสดงจำนวนคำศัพท์ในฐานความรู้เวิร์ดเนต

Part of speech	Unique String	Synsets	Total word Sense pair
Noun (n)	114,648	79,689	146,273
Verb (v)	11,306	13,508	24,691
Adjective (a,s)	21,793	18,563	31,016
Adverb (r)	4,660	3,664	5,808
Total	152,407	115,424	207,788

จากตารางที่ 2-3 แสดงจำนวนข้อมูลทั้งหมดและประเภทคำต่างๆ ซึ่งแบ่งได้เป็น 3 กลุ่ม ดังนี้

กลุ่มที่ 1: Unique String คือ กลุ่มของคำศัพท์ที่ถูกจัดหรือคัดออกมาให้เหลือเพียงหนึ่งคำศัพท์จากคำศัพท์ที่ซ้ำกัน

กลุ่มที่ 2: Synsets คือ กลุ่มของคำศัพท์ทั้งหมดที่มีความหมายใกล้เคียงหรือเหมือนกันกับกลุ่มของ Unique String

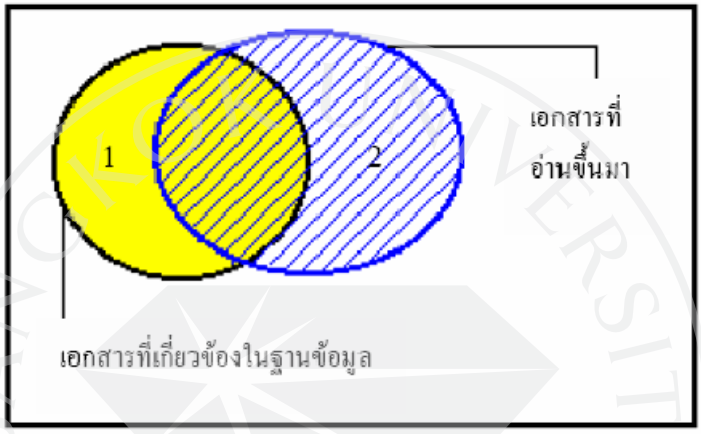
กลุ่มที่ 3: Total word Sense pair กลุ่มของคำศัพท์ทั้งหมดที่มีความหมายใกล้เคียงกันหรือเหมือนกันกับคำศัพท์ที่ไม่ได้ถูกจัดกลุ่มให้เหลือเพียงหนึ่งคำจากคำที่ซ้ำกัน

2.1.12 การวัดประสิทธิภาพของการสืบค้น

Precision and Recall (Richard, 2000 และ Cyril and Eric, 2005) คือวิธีการหนึ่งสำหรับการประเมินประสิทธิภาพการสืบค้นเอกสารซึ่งแสดงให้เห็นตามภาพที่ 2-12 ซึ่งสามารถอธิบายได้ดังนี้

1. กลุ่มของเอกสารที่เก็บไว้ในฐานความรู้ซึ่งมีความสัมพันธ์กับข้อมูลของการสืบค้น (1)
2. กลุ่มของเอกสารที่ถูกเลือกขึ้นมาเป็นผลลัพธ์ (2)

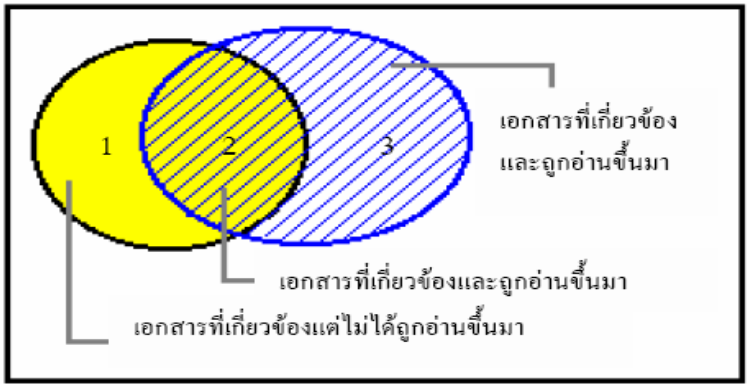
ภาพที่ 2-12: แสดงความสัมพันธ์ระหว่างกลุ่มข้อมูลในฐานข้อมูลเกี่ยวข้องกับคำสืบค้น



จากภาพที่ 2-13 สามารถอธิบายได้ดังนี้

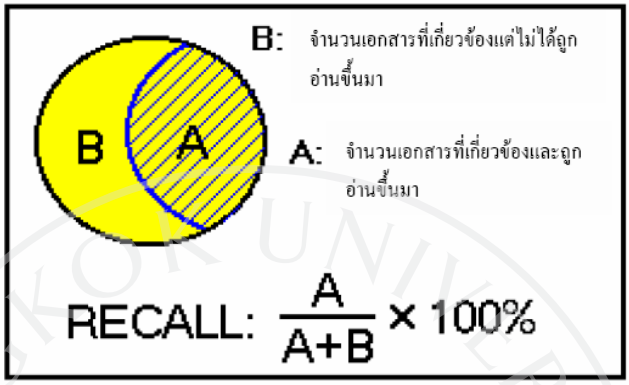
1. กลุ่มของเอกสารที่เกี่ยวข้องกับข้อมูลสืบค้นแต่ไม่ได้เลือกขึ้นมา (1)
2. กลุ่มของเอกสารที่เกี่ยวข้องกับข้อมูลสืบค้นและถูกเลือกขึ้นมา (2)
3. กลุ่มของเอกสารที่ไม่เกี่ยวข้องกับข้อมูลสืบค้นแต่ถูกเลือกขึ้นมา (3)

ภาพที่ 2-13: แสดงความสัมพันธ์ระหว่างกลุ่มของข้อมูลที่ใช้สืบค้นกับคำค้น



ภาพที่ 2-14: แสดงสมการการคำนวณในการหาค่า Recall

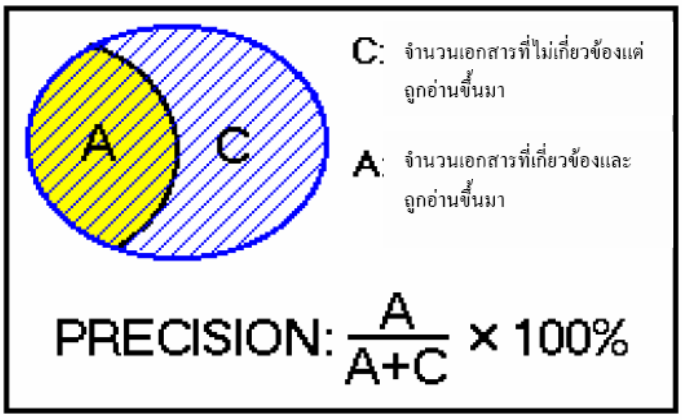
$$\text{RECALL} = \frac{A}{A+B} \times 100\ \%$$



ภาพที่ 2-14 คือการหาค่าของ Recall เมื่อ B คือจำนวนของเอกสารที่เกี่ยวข้องหรือสัมพันธ์กับคำสืบค้นแต่ไม่ได้ถูกเลือกมาเป็นผลลัพธ์ และ A คือจำนวนเอกสารที่เกี่ยวข้องหรือสัมพันธ์กับคำสืบค้นและถูกเลือกมาเป็นผลลัพธ์ ซึ่งสามารถคำนวณได้จากสมการดังนี้

ภาพที่ 2-15: แสดงสมการการคำนวณในการหาค่าของ Precision

$$\text{Precision} = \frac{A}{A+C} \times 100\ \%$$



ภาพที่ 2-15 คือการหาค่าของ Precision เมื่อ C คือจำนวนของเอกสารที่ไม่เกี่ยวข้องหรือไม่สัมพันธ์กับคำสืบค้นแต่ถูกเลือกมาเป็นผลลัพธ์ และ A คือจำนวนของเอกสารที่เกี่ยวข้องหรือสัมพันธ์กับคำสืบค้นและถูกเลือกมาเป็นผลลัพธ์ ซึ่งสามารถคำนวณได้จากสมการดังนี้

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การจัดเก็บและสืบค้นรูปภาพเชิงความหมายบนฐานของ RDF (นพคุณ บุญสิน, 2547)

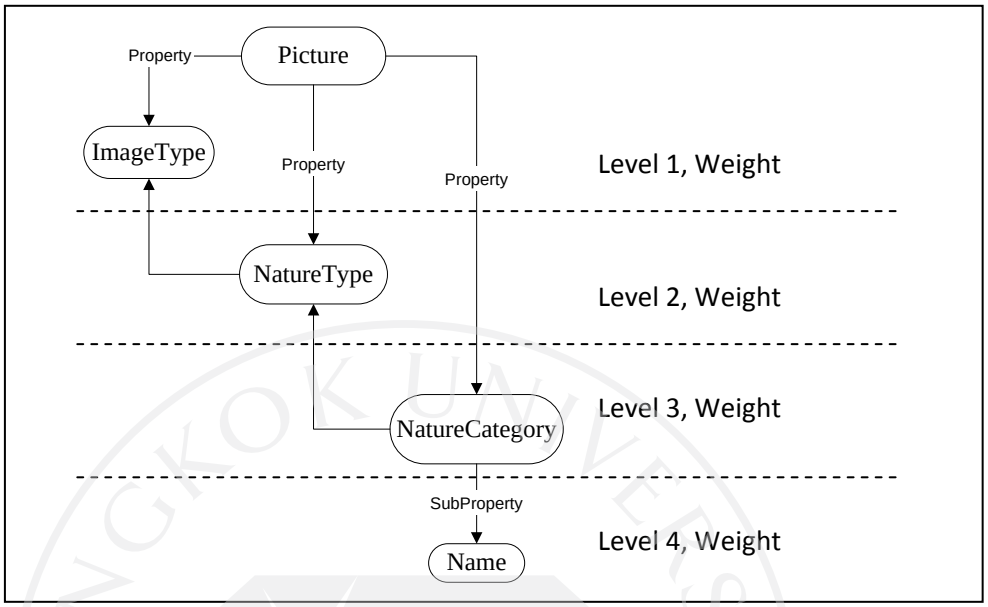
งานวิจัยนี้จะนำเสนอตัวแบบการอธิบายข้อมูลรูปภาพธรรมชาติ โดยตัวแบบจะเป็นโครงสร้างข้อมูลสำหรับอธิบายคุณลักษณะของรูปภาพ ซึ่งประกอบด้วย ประเภทของรูปภาพ หมวดของธรรมชาติและหมวดย่อยของธรรมชาติ ตัวแบบดังกล่าวสามารถอธิบายด้วยกราฟบนฐานของ RDF

ในงานวิจัยนี้มีการใช้ RDF ในการอธิบายรายละเอียดของรูปภาพ โดยมีขั้นตอนที่สำคัญดังต่อไปนี้

2.2.1.1 การออกแบบโครงสร้างข้อมูล

เพื่ออธิบายข้อมูลของรูปภาพ โดยแบ่งออกเป็น 4 ระดับเชิงความหมาย ในระดับชั้นที่ 1 คุณสมบัติประเภทของรูปภาพ (Image Type) เป็นคุณสมบัติที่ออกแบบเพื่ออธิบายประเภทของรูปภาพ เช่น ภาพธรรมชาติ ภาพคน และภาพกีฬา เป็นต้น ระดับชั้นที่ 2 คุณสมบัติหมวดของธรรมชาติ (Nature Type) คุณสมบัติที่ออกแบบเพื่ออธิบายหมวดของธรรมชาติที่ปรากฏอยู่ในรูป เช่น ภัยพิบัติ (Disaster) ภูมิประเทศ (Geography) เป็นต้น ระดับชั้นที่ 3 คุณสมบัติหมวดย่อยธรรมชาติ (Nature Category) เป็นคุณสมบัติที่ออกแบบเพื่ออธิบายหมวดย่อยของธรรมชาติในรูปภาพ เช่น หมวดย่อยของภัยพิบัติประกอบด้วย น้ำท่วม ไฟไหม้และแผนดินไหว เป็นต้น ระดับชั้นที่ 4 คุณสมบัติชื่อ (Name) เป็นคุณสมบัติที่ออกแบบเพื่ออธิบายชื่อเฉพาะของหมวดย่อยธรรมชาติในรูป เช่น สึนามิ (Tsunami) อันดามัน (Andaman) เป็นต้น

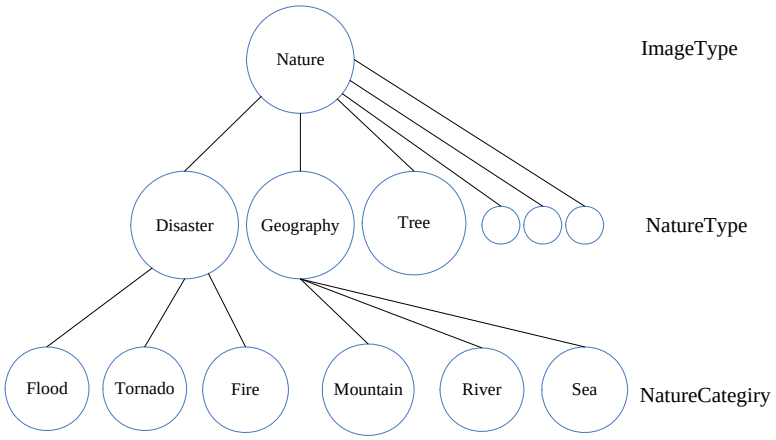
ภาพที่ 2-16: การอธิบายโครงสร้างรูปภาพภาษาธรรมชาติด้วยกราฟ



2.2.1.2 นิยามออนโทโลยี

คำศัพท์เกี่ยวกับธรรมชาติ ในการจัดเก็บและสืบค้นเชิงความหมายจำเป็นจะต้องนิยามออนโทโลยีคำศัพท์เกี่ยวกับธรรมชาติที่มีความหมายเหมือนกันเพื่อใช้ในการจัดเก็บและสืบค้นรูปภาพ โดยจะใช้ RDF Schema ในการอธิบายข้อมูลของออนโทโลยี

ภาพที่ 2-17: Ontology คำศัพท์ธรรมชาติ



2.2.1.3 การนำเข้าข้อมูล

เป็นการจัดเตรียมข้อมูลก่อนดำเนินการสืบค้น ซึ่งถือเป็นขั้นตอนที่มีความ สำคัญ งานวิจัยนี้จึงได้ออกแบบการนำเข้าข้อมูล โดยมีขั้นตอนดังนี้

1. การสร้าง Lexical Chains จากคำอธิบายรูปภาพ ซึ่งนำเข้าโดยผู้นำเสนอภาพ จะเป็นภาษาธรรมชาติในการอธิบายข้อมูลภาพการสร้าง Lexical Chains ประกอบด้วย 2 ขั้นตอนหลักคือ

1.1 การหาคำสำคัญ จากตัวอย่างคำอธิบายรูปภาพ “While parts of the west coast of Thailand were effect by the 2004 disaster tsunami, most tourist areas had reopened within weeks or even days of the event. Aside from some isolated sports, tourism infrastructure is now intact and business are operating as usual. On badly damaged Koh Phuket.” จะได้คำสำคัญคือ {parts, coast, Thailand, disaster, Tsunami, tourism, areas spots, infrastructure, intact, businesses, Koh, Phuket}

1.2 การสร้าง Lexical Chains จากคำสำคัญดังกล่าวจะนำไปสร้าง Lexical Chains โดยดำเนินการตามลำดับขั้นตอนของการทำ Lexical Chains

2. สิ่งที่ได้จากขั้นตอนที่ 1 จะนำข้อมูลดังกล่าวไปอธิบายตามโครงสร้างที่ได้ออกแบบไว้ ซึ่งในขั้นตอนนี้มีการใช้ Ontology จากคำศัพท์ธรรมชาติเข้ามาร่วมด้วยคล้ายกับการสร้าง Lexical Chains

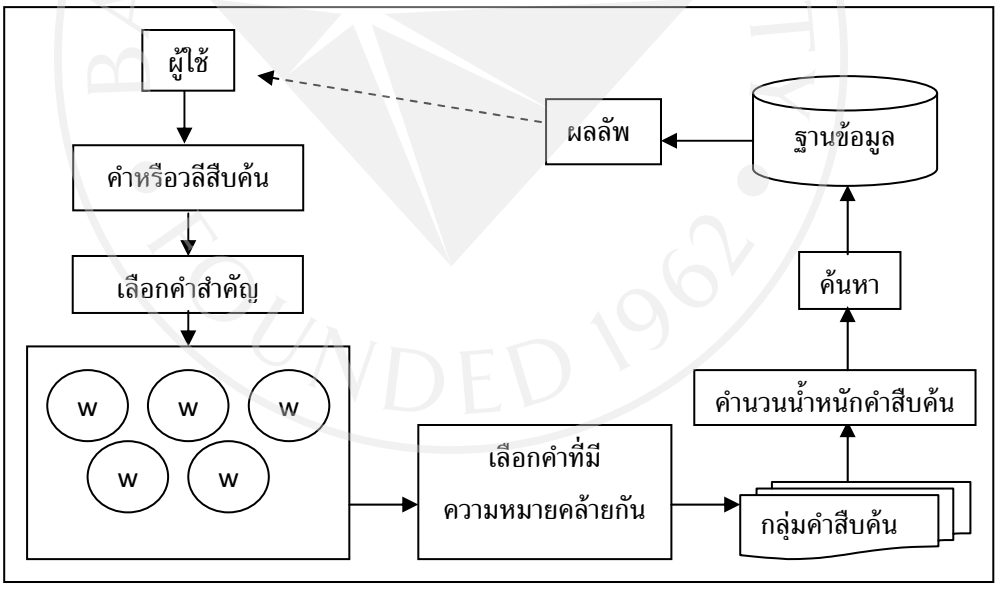
3. นำข้อมูลไปอธิบายด้วยภาษา RDF

4. จัดเก็บข้อมูลในฐานข้อมูลในรูปแบบของเอกสาร RDF ที่เชื่อมโยงกับรูปภาพ

2.2.1.4 การออกแบบการสืบค้นรูปภาพ

การสืบค้นข้อมูลโดยการใช้คำสำคัญ (Keyword) และคำสืบค้นด้วยภาษาธรรมชาติ (Natural Language) ล้วนแต่มีประสิทธิภาพต่ำ เนื่องมาจากข้อมูลที่สืบค้นมีโครงสร้างที่ไม่แน่นอน จึงทำให้การสืบค้นข้อมูลได้ผลลัพธ์ไม่ตรงกับความต้องการ การออกแบบขั้นตอนการสืบค้นข้อมูลเป็นขั้นตอนสำคัญในการสืบค้นข้อมูล โดยการออกแบบขั้นตอนจะดำเนินตามโครงสร้างข้อมูลที่ได้กำหนดขึ้น เริ่มจากนำคำหรือวลีสืบค้นมาเลือกสำคัญ ต่อจากนั้นจะนำคำสำคัญดังกล่าวมาค้นหาคำที่มีความหมายคล้ายกัน (Synonym) จาก Ontology คำศัพท์ธรรมชาติ หลังจากนั้นจะทำการคำนวณน้ำหนักของกลุ่มคำสืบค้น และนำกลุ่มคำสำคัญไปค้นหาในฐานข้อมูล ภาพที่ 2-18 แสดงขั้นตอนการสืบค้นข้อมูล

ภาพที่ 2-18: แสดงขั้นตอนการสืบค้น

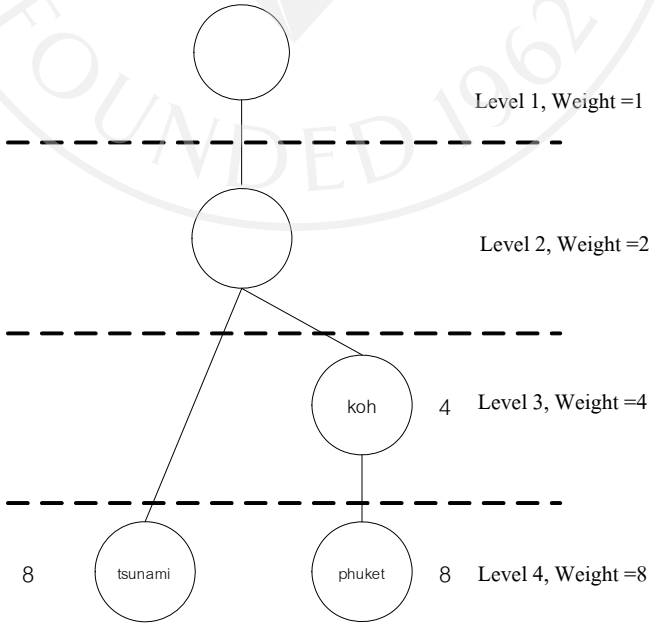


1) การเลือกคำสำคัญจากคำหรือวลีสืบค้น จากคำหรือวลีสืบค้นที่ผู้ใช้ ส่งเข้าสู่ระบบเพื่อสืบค้นข้อมูล คำหรือวลีดังกล่าวเป็นภาษาธรรมชาติ (Natural Language) ซึ่งการค้นหาด้วยรูปแบบดังกล่าวทำให้ผลลัพธ์จากการสืบค้นไม่ตรงกับความต้องการ ดังนั้นจึงจำเป็นต้องมีการจัดรูปแบบโดยการการพิจารณาเลือกเฉพาะคำสำคัญที่สื่อความหมายได้มาเป็นคำสำคัญในการสืบค้นโดยอาศัยฐานข้อมูล Ontology คำศัพท์ธรรมชาติ ตัวอย่าง วลีสืบค้น “Picture of disaster about tidalwave

tsunami in koh phuket.” จากวลีดังกล่าวจะเลือกคำสำคัญได้ดังนี้ {disaster, tidalwave tsunami koh phuket} นำคำสำคัญดังกล่าวเลือกคำศัพท์ที่มีความหมายเหมือนกัน (Synonym) ความสำคัญของคำที่มีความหมายเหมือนกัน เนื่องจากการสืบค้นด้วยคำสำคัญ เป็นการเปรียบเทียบคำสำคัญกับข้อมูลในฐานข้อมูล ถ้าการเปรียบเทียบไม่พบข้อมูลแสดงว่าไม่มีข้อมูลที่ตรงกับความต้องการของผู้ใช้ สำหรับการเลือกคำที่มีความหมายเหมือนกันเพื่อเป็นกลุ่มคำในการสืบค้นข้อมูล มีจุดประสงค์คือ ทำการสืบค้นข้อมูลด้วยคำที่มีความหมายเหมือนกันหรือใกล้เคียงกันจะทำให้การสืบค้นมีประสิทธิภาพและเชิงความหมายด้วย ภาพที่ 3.18 เป็นการนิยามคำศัพท์ที่มีความความหมายเหมือนกันโดยภาษา RDF Schema

2) การคำนวณน้ำหนักของคำสืบค้น จากกลุ่มคำสืบค้นที่ได้จากขั้นตอนที่ 3.6.2 ในขั้นตอนนี้เป็นการคำนวณน้ำหนักของคำสืบค้นเพื่อเป็นคะแนนในเปรียบเทียบและสืบค้นรูปภาพ โดยการคำนวณน้ำหนักของคำสืบค้น จะอาศัยฐานข้อมูล Ontology ของคำศัพท์ธรรมชาติและน้ำหนักของโครงสร้างที่ได้นิยามขึ้นในขั้นตอนที่ 3.2 ซึ่งมีการให้น้ำหนักตามลำดับดังนี้ ระดับที่ 1 2 3 และ 4 มีน้ำหนักคือ 1 2 4 และ 8 ภาพที่ 2-19 เป็นตัวอย่างการคำนวณน้ำหนักของคำสืบค้น โดยในคำสืบค้นประกอบด้วย คำในระดับที่ 3 (1 คำ) และคำระดับที่ 4 (2 คำ) ผลรวมของน้ำหนักเท่ากับ 20

ภาพที่ 2-19: การคำนวณน้ำหนักของคำสืบค้น



3) การสืบค้นข้อมูลจากฐานข้อมูล

เมื่อทำการจัดรูปแบบข้อมูลนำเข้าและคำสืบค้นเรียบร้อยแล้ว ขั้นตอนต่อมาคือการออกแบบการสืบค้นข้อมูลจะดำเนินการตามโครงสร้างข้อมูลเพื่อให้ได้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ภาพที่ 2-20 แสดงลำดับขั้นตอนการการสืบค้นข้อมูล

ภาพที่ 2-20: ลำดับขั้นตอนการสืบค้นข้อมูล

```
MF(Query[Q1,Q2,…,Qn], Picture[P1,P2,…,Pm],QW)
{
 i=0, j=0;
PW = 0;
    For i from 0 to n
        For j from 0 to m
        Switch (ontology_type (Qi)) ;แยกประเภทเปรียบเทียบตามคุณสมบัติ
        {
            Case ImageType ; บวกค่าเพิ่มตามน้ำหนักของแต่ละคุณสมบัติ
                If (Qi = Pj) {PW=PW+1 }
            Case NatureType
                If (Qi = Pj) {PW=PW+2 ; }
            Case NatureCategory
                If (Qi = Pj) {PW=PW+4 ; }
            Case Name
                If (Qi = Pj) {PW=PW+8; }
             }
             If (PW >= ½(QW)) ; เปรียบเทียบน้ำหนักของรูปภาพ กับ น้ำหนักของคำสืบค้น
             {   Return “Picture [name]” ;ชื่อของรูป หรือที่อยู่ของรูป}
}
```

ลำดับขั้นตอนการสืบค้นข้อมูลตามภาพที่ 2-20 มีลำดับขั้นตอนการสืบค้นข้อมูลดังนี้

3.1 นำข้อมูลเข้าสู่ขั้นตอนการสืบค้น (Matching Function = MF) ข้อมูลประกอบ ด้วยกลุ่มคำสืบค้น (Query $[Q_1,Q_2,\ldots,Q_n]$) โดย $Q_1,Q_2,\ldots,Q_n$ คือกลุ่มคำสืบค้น ข้อมูลของรูปภาพ (Picture $[P_1,P_2,\ldots,P_m]$) โดย $P_1,P_2,\ldots,P_m$ คือคุณสมบัติของรูปภาพและน้ำหนักของคำสืบค้น (Query Weight = QW)

3.2 แยกประเภทของกลุ่มคำสืบค้นโดยใช้ Ontology คำศัพท์ธรรมชาติ

3.3 คำนวณน้ำหนักของรูปภาพ (Picture Weight =PW) ตามคุณสมบัติที่ค้นพบโดยแต่ละคุณสมบัติจะมีน้ำหนักดังนี้ ระดับที่ 1 2 3 และ 4 มีน้ำหนักคือ 1 2 4 และ 8

3.4 นำค่าน้ำหนักของรูปภาพที่ได้ มาทำการเปรียบเทียบกับน้ำหนักของกลุ่มคำสืบค้น ถ้าน้ำหนักของภาพที่ค้นพบ มากกว่าร้อยละ 50 ของน้ำหนักกลุ่มคำสืบค้น แสดงภาพภาพดังกล่าว มีความถูกต้องและตรงกับความต้องการ การเปรียบเทียบคะแนนของน้ำหนักของการสืบค้นและน้ำหนักของรูปภาพ ถ้าคะแนนของรูปภาพมีน้ำหนักน้อยกว่าร้อยละ 50 ของคะแนนน้ำหนักของคำสืบค้นจะทำให้ได้ผลลัพธ์ไม่ตรงกับความต้องการ ถ้าคะแนนรวมของน้ำหนักมากกว่าร้อยละ 50 โดยเฉลี่ยจะได้ผลลัพธ์ที่ตรงกับความต้องการ

2.2.2 การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย (สิริรัตน์ ประกฤติกรชัย, 2550)

เป็นงานวิจัยที่มีวัตถุประสงค์เพื่อพัฒนาต้นแบบระบบออนโทโลยีของพืชสมุนไพรไทย เพื่อสร้างและเก็บข้อมูลออนโทโลยีพืชสมุนไพรที่ออกแบบไว้ ซึ่งครอบคลุมพืชสมุนไพรไทยที่ใช้เป็นส่วนประกอบของยาสามัญประจำบ้านแผนโบราณ 27 ชนิด ตามประกาศกระทรวงสาธารสุข เรื่องยาสามัญฉบับที่ 2 และเพื่อให้สามารถทดสอบและใช้งานจากผู้ใช้ทั่วไปได้ ข้อมูลพืชสมุนไรถูกถ่ายโอนออกให้อยู่ในรูป OWL จากนั้นถูกใช้เป็นฐานความรู้ในการสร้างเว็บแอพพลิเคชัน ซึ่งเว็บแอพพลิเคชันที่สร้างขึ้นสามารถใช้ค้นหาทั้งข้อมูลและแสดงความสัมพันธ์ระหว่างของพืชสมุนไพรไทยกับยาสามัญประจำบ้านแผนไทยได้ จากการประเมินประสิทธิภาพของระบบโดยผู้เชี่ยวชาญ ที่ระดับนัยสำคัญ 0.05 พบว่าระบบมีประสิทธิภาพโดยรวมในด้านความสัมพันธ์ การแบ่งหมวดหมู่ของ

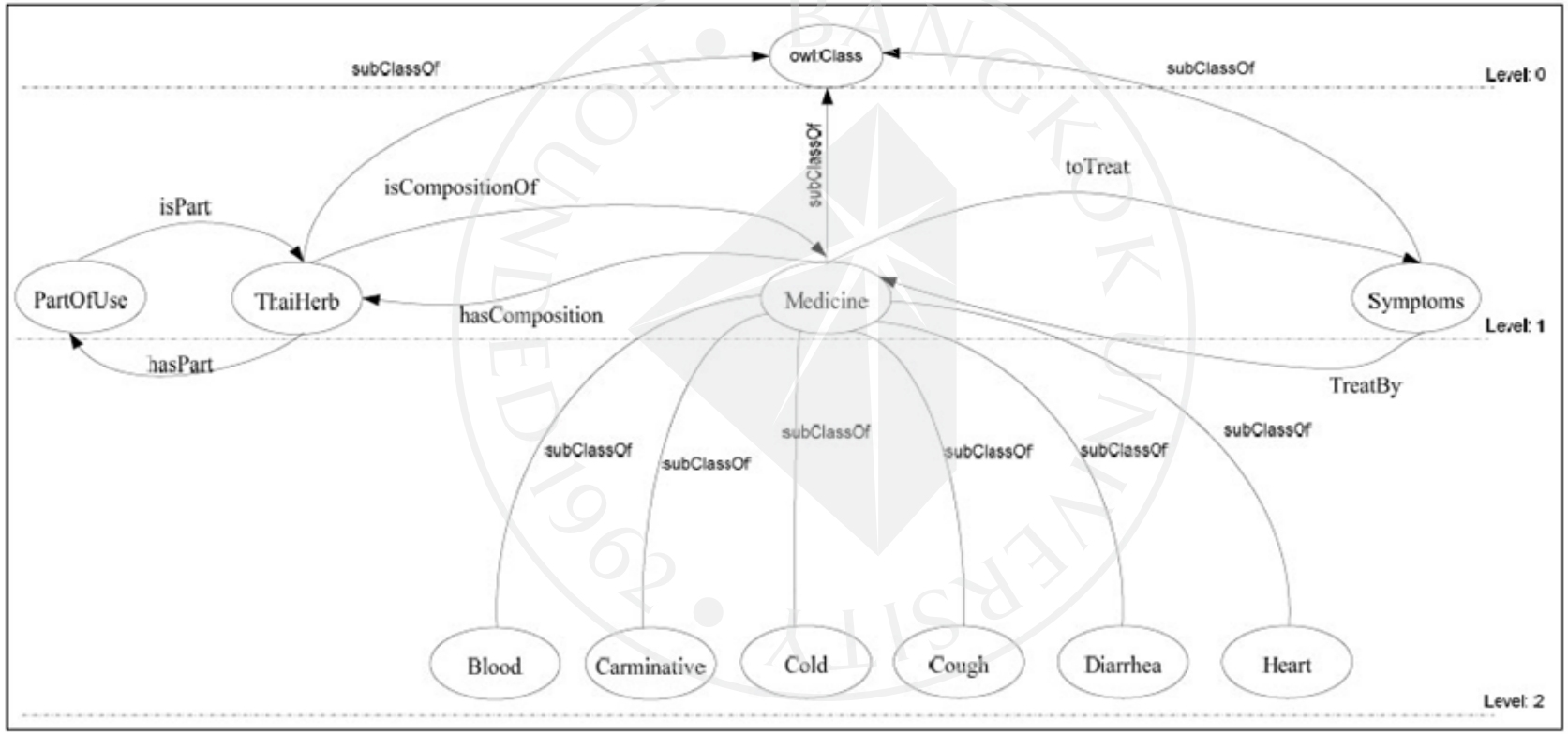
ระบบออนโทโลยี และด้านการใช้งานกับด้านการแสดงผลของเว็บแอพพลิเคชันอยู่ในระดับดี (จากมาตรวัดการประมาณค่าของ Recall ค่าเฉลี่ยของการประเมินเท่ากับ 4.36 จาก 5)

ในการพัฒนาต้นแบบฐานข้อมูลความรู้พืชสมุนไพรไทยที่เป็นส่วนประกอบของยาสามัญประจำบ้านแผนโบราณ จากการศึกษาค้นคว้าจากตำราแพทย์แผนโบราณทั่วไป สาขาเภสัชกรรม และสอบถามจากผู้เชี่ยวชาญด้านสมุนไพรไทย ทำให้สามารถกำหนดขอบเขต และออกแบบความสัมพันธ์ของข้อมูลโดยใช้หลักการของออนโทโลยีมาพัฒนาระบบ ในการจัดการอธิบายถึงความสัมพันธ์ระหว่างข้อมูลในระบบ และสามารถพิจารณาการอนุมานเพื่อค้นหาข้อมูลที่มีความเกี่ยวข้องกัน โดยพิจารณาได้จากความสำคัญของข้อมูลเชิงความหมาย จากการวิเคราะห์ความต้องการในการใช้งานออนโทโลยี สามารถสรุปได้ดังนี้

1. ออนโทโลยีสามารถอธิบายแนวความคิดพืชสมุนไพรไทย ยาสามัญประจำบ้านแผนโบราณ และกลุ่มอาการ
2. ออนโทโลยีสามารถสร้างความสัมพันธ์ระหว่างข้อมูลพืชสมุนไพรไทย ยาสามัญประจำบ้านแผนโบราณ และกลุ่มอาการ

ในภาพที่ 2.21 แสดงภาพรวมของออนโทโลยีระดับบนของออนโทโลยีการพัฒนาฐานข้อมูลองค์ความรู้พืชสมุนไพรไทยที่ได้ออกแบบสำหรับสารนิพนธ์นี้

2.21 ภาพรวมของออนโทโลยีระดับบนของออนโทโลยีการพัฒนาต้นแบบฐานข้อมูลองค์ความรู้พืชสมุนไพรไทยที่เป็นส่วนประกอบของยาสามัญประจำบ้านแผนโบราณ



คลาสพืชสมุนไพรไทย (Thai Herb Class)

เป็นคลาสชนิดของพืชสมุนไพรไทย โดยมี 2 ซับคลาส (Subclass) คือ

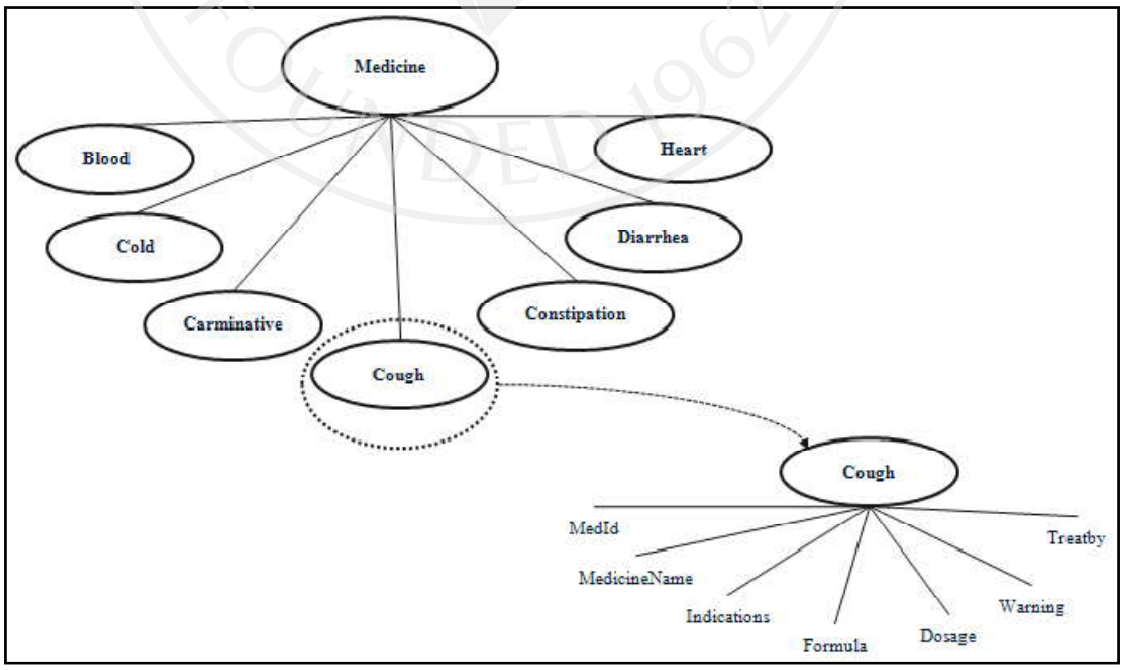
1) คลาส ThaiHerb มีซับคลาสตามชื่อวงศ์ของพืชสมุนไพร มีทั้งหมด 46 คลาส โดยแต่ละคลาสมีคุณสมบัติของชนิดข้อมูล (Datatype Property) 14 Properties

2) คลาส PartOfUse เป็นคลาสแสดงส่วนของพืชที่มีสรรพคุณทางยา เป็นคลาสที่อยู่ในระดับที่ 1 และมีคุณสมบัติของชนิดข้อมูล (Datatype Property) 4 Properties

คลาสยาสามัญประจำบ้านแผนโบราณ (Medicine)

ประกอบด้วยซับคลาสย่อย กลุ่มอาการ 7 ซับคลาส และมีคุณสมบัติชนิดข้อมูล 7 Properties รายละเอียดตามภาพที่ 2-22

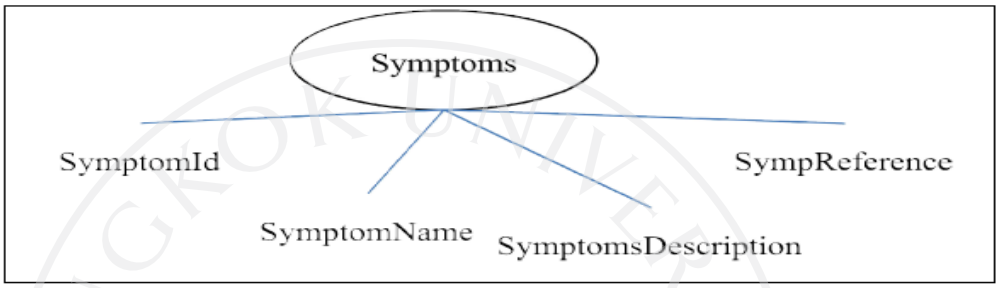
ภาพที่ 2-22: คลาส Medicine



คลาสกลุ่มอาการ (Symptoms)

ประกอบไปด้วยคุณสมบัติของชนิดข้อมูล จำนวน 3 Properties รายละเอียดตามภาพที่ 2-23

ภาพที่ 2-23: คลาส Symptoms



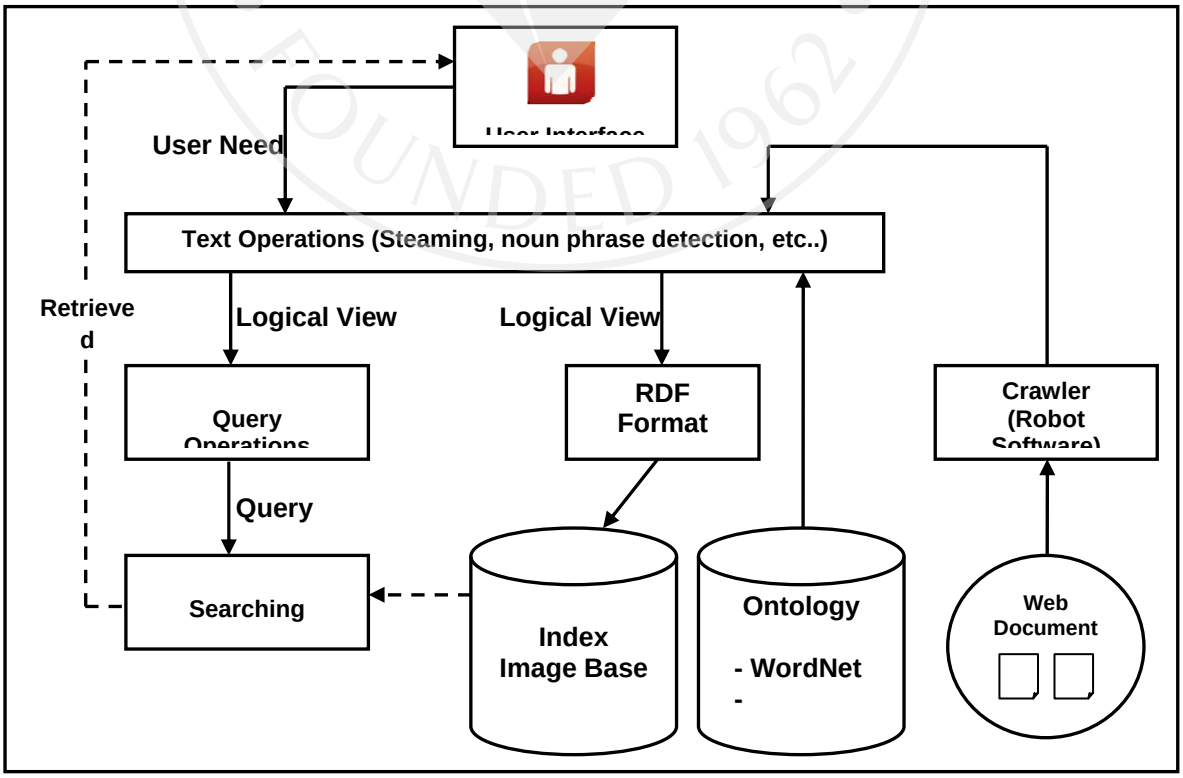
บทที่ 3
วิธีการดำเนินงาน

การดำเนินงานของงานวิจัยในบทนี้ เป็นการนำเสนอเกี่ยวกับวิธีการและขั้นตอนการทำงานวิจัยซึ่งประกอบด้วยการนำเข้าข้อมูล การออกแบบโครงสร้างสำหรับการอธิบายรายละเอียดของรูปภาพด้วยออนโทโลยีและการออกแบบขั้นตอนการสืบค้นข้อมูลรูปภาพ

3.1 การออกแบบจำลองการสืบค้นข้อมูลรูปภาพ

แบบจำลองการสืบค้นข้อมูลรูปภาพ ออกแบบโดยตั้งอยู่บนพื้นฐานของแบบจำลองการสืบค้นสารสนเทศ (Information Retrieval Model) ของริคาร์โด้ (Ricardo, 1999) ซึ่งมีองค์ประกอบดังภาพ 3-1

ภาพที่ 3-1: แบบจำลองการสืบค้นรูปภาพบนเว็บ



องค์ประกอบของแบบจำลองการสืบค้นรูปภาพบนเว็บประกอบด้วย

3.1.1 Crawler (Robot Software)

Crawler คือ ซอฟต์แวร์ที่ทำหน้าที่ในการเก็บรวบรวมเอกสารอิเล็กทรอนิกส์บนเว็บทั้งหมด (เช่น แฟ้มเอกสารประเภท HTML, PHP, PDF, DOC และอื่นๆ) บนเว็บ ซึ่งจะเรียกสั้นๆ ว่า เอกสารบนเว็บ (Web Documents)

3.1.2 Text Operations

เป็นขั้นตอนการเตรียมข้อมูลเบื้องต้น เป็นกระบวนการนำเอกสารที่รวบรวมได้เข้าสู่กระบวนการเตรียมข้อความก่อนการทำดัชนี เช่น การแยกคำ การตัดคำ การเลือกกลุ่มคำที่เป็นคำหลัก (Keyword) กระบวนการนี้ จะใช้ฐานความรู้ออนโทโลยีเป็นเครื่องมือ โดยใช้ฐานความรู้เวิร์ดเนตสำหรับภาษาอังกฤษในการเลือกคำที่สื่อความหมาย และฐานความรู้ธีซอรัสในการจัดกลุ่มคำที่มีความหมายเหมือนกัน

3.1.3 การทำดัชนี (Indexing)

เป็นวิธีการในการจัดทำดัชนีของคำที่ได้จากการสกัดคำหลักจากเอกสาร เพื่อนำมาบรรยายให้อยู่ในรูปแบบของภาษา RDF (Resource Description Framework) ที่สามารถอธิบายรายละเอียดของทรัพยากรต่างๆ บนอินเตอร์เน็ต ที่คอมพิวเตอร์สามารถเข้าใจความหมายของข้อมูลได้ ตามโครงสร้างของออนโทโลยีที่ได้ทำการออกแบบเอาไว้ เอกสาร RDF จะถูกเก็บรวบรวมไว้เป็นฐานความรู้เพื่อจัดเตรียมไว้สำหรับการสืบค้น

3.1.4 การสืบค้น (Searching)

เมื่อผู้ใช้กรอกข้อมูลสืบค้น (Query) ที่ต้องการ ข้อมูลสืบค้นจะถูกเข้าสู่กระบวนการของการเตรียมข้อมูลเบื้องต้น เพื่อทำการแบ่งคำที่เป็นไปได้ทั้งหมดก่อน และเมื่อแบ่งกลุ่มคำเสร็จแล้วจะเข้าสู่กระบวนการของ Query Operations ซึ่งเป็นวิธีการกำหนดค่าน้ำหนักของคำเมื่อเทียบกับข้อมูลสืบค้นที่มีอยู่ในแต่ละเอกสาร จากนั้นจะเข้าสู่กระบวนการสืบค้น

แบบจำลองดังภาพที่ 3-1 ประกอบด้วยกระบวนการสำหรับการสืบค้นรูปภาพบนเว็บที่สำคัญ 2 กระบวนการ คือ (1) การสืบค้นรูปภาพบนเว็บ (2) การจัดเตรียมข้อมูลสำหรับการสืบค้น

3.2 การออกแบบโครงสร้างข้อมูลออนโทโลยี

ในการพัฒนาต้นแบบการสืบค้นข้อมูลรูปภาพสถานที่ท่องเที่ยวในประเทศไทยบนอินเตอร์เน็ต ต้องมีการออกแบบออนโทโลยี เพื่อนำข้อมูลบรรยายรูปภาพมาบรรยายตามออนโทโลยีที่ออกแบบไว้ ในการออกแบบออนโทโลยีจะเป็นเป็นออนโทโลยีประเภท Formal Semantic Ontology ซึ่งเป็นออนโทโลยีที่ออกแบบมาเพื่องานอย่างใดอย่างหนึ่ง ขั้นตอนการออกแบบออนโทโลยีมีดังนี้

3.2.1 กำหนดขอบเขตและแนวทางของออนโทโลยี (Scope Determination)

เริ่มแรกของการพัฒนาออนโทโลยี ต้องเริ่มทำการกำหนดขอบเขตของออนโทโลยี ในงานวิจัยนี้มีขอบเขตเพื่อนำออนโทโลยีมาใช้ในการบรรยายสถานที่ท่องเที่ยวในประเทศไทย

3.2.2 การระบุเงื่อนไขในออนโทโลยี (Term Enumeration)

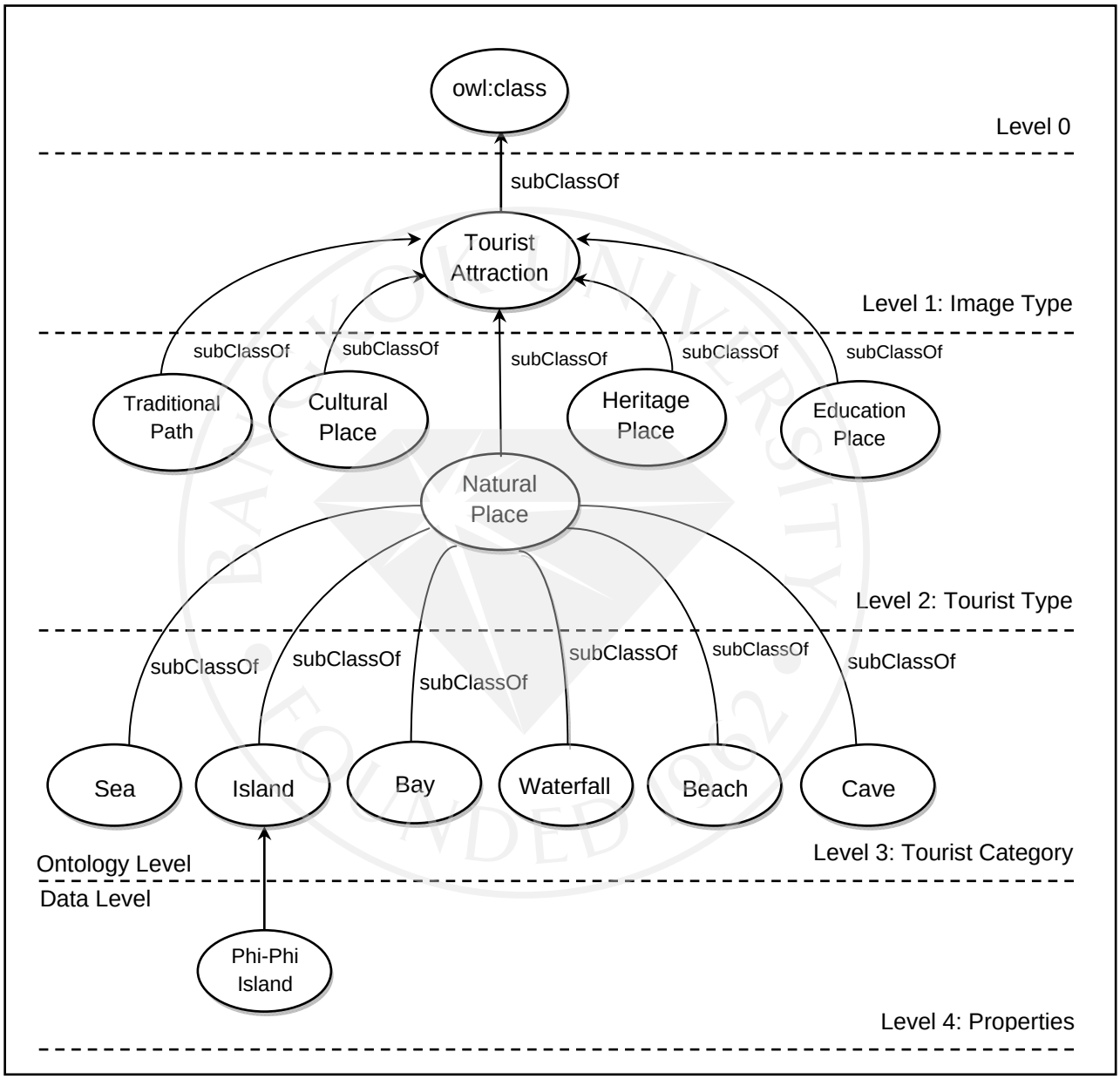
เพื่อกำจัดขอบเขตของรายการหรือเทอมทั้งหมด เพื่ออธิบายขอบเขตการทำงาน เช่น ทะเล น้ำตก ชายหาด ถ้ำ เกาะ เป็นต้น

3.3.3 การกำหนดคลาสและลำดับชั้นของคลาส (Class Definition)

คลาสหมายถึงแนวคิดที่อยู่ในโดเมนซึ่งประกอบด้วยส่วนต่างๆ ในการพัฒนาออนโทโลยี ตัวอย่างเช่น ถ้าคลาสสถานที่ท่องเที่ยวเชิงธรรมชาติเป็นซูเปอร์คลาส (Super Class) ของคลาสทะเล (Sea) แล้ว Instance ทั้งหมดที่อยู่ในคลาสทะเล ก็จะเป็น Instance ของคลาสสถานที่ท่องเที่ยวเชิงธรรมชาติด้วย ส่วนการกำหนดลำดับของคลาสจะใช้วิธีการแบบบนลงล่าง (Top-Down)

ในภาพที่ 3.2 แสดงภาพรวมของออนโทโลยีการพัฒนาฐานข้อมูลองค์ความรู้สถานที่ท่องเที่ยวในประเทศไทยที่ได้ออกแบบสำหรับงานวิจัยนี้

ภาพที่ 3-2: ภาพรวมของออนโทโลยีของโครงสร้างข้อมูลออนโทโลยีสถานที่ท่องเที่ยวในประเทศไทย



จากภาพที่ 3-2 เป็นตัวอย่างการแยกประเภทและหมวดหมู่ของออนโทโลยีสถานที่ท่องเที่ยว โดยจะแบ่งออกเป็น 4 ระดับ ประกอบด้วยระดับประเภทของรูปภาพ (Image Type) ระดับประเภทสถานที่ท่องเที่ยว (Tourist Type) ระดับประเภทหมวดย่อยของสถานที่ท่องเที่ยว (Tourist Category) และระดับล่างสุดเป็นคุณสมบัติของสถานที่ท่องเที่ยว

ประเภทรูปภาพ (Image Type)

เป็นส่วนที่ใช้ในการอธิบายประเภทของรูปภาพ เช่น รูปภาพการแข่งขันกีฬา รูปภาพธรรมชาติและรูปภาพคน เป็นต้น โดยในงานวิจัยนี้ให้ความสนใจในส่วนของรูปภาพสถานที่ท่องเที่ยวในประเทศไทย เพียงอย่างเดียวเท่านั้น

ประเภทสถานที่ท่องเที่ยว (Tourist Type)

เป็นส่วนที่อธิบายประเภทของสถานที่ท่องเที่ยว รูปภาพในหมวดนี้ รายละเอียดของคลาสต่างๆ ที่กำหนดในออนโทโลยีสถานที่ท่องเที่ยวในประเทศไทย ในระดับที่ 2 ประกอบด้วย 6 คลาส ดังที่จะแสดงในตาราง 3-1 ในการแบ่งประเภทของสถานที่ท่องเที่ยว ได้มีการออกแบบตามการแบ่งประเภทตามการท่องเที่ยวแห่งประเทศไทย

ตาราง 3-1: รายละเอียดของคลาสในระดับ 2 ในออนโทโลยี

คลาสที่	คลาส	รายละเอียด
1	Natural Place	คลาสแทนข้อมูลสถานที่ท่องเที่ยวเชิงธรรมชาติ
2	Cultural Place	คลาสแทนข้อมูลสถานที่ท่องเที่ยวเชิงวัฒนธรรม
3	Heritage Place	คลาสแทนข้อมูลสถานที่ท่องเที่ยวเชิงประวัติศาสตร์
4	Education Place	คลาสแทนข้อมูลสถานที่ท่องเที่ยวเชิงศึกษา
5	Traditional Path	คลาสแทนข้อมูลสถานที่ท่องเที่ยวตามวิถีชีวิต
6	Recreational Place	คลาสแทนข้อมูลสถานที่ท่องเที่ยวเชิงสันทนาการ

ประเภทหมวดย่อยสถานที่ท่องเที่ยว (Tourist Category)

เป็นส่วนที่อธิบายหมวดย่อยของประเภทสถานที่ท่องเที่ยวแต่ละประเภท รายละเอียดของคลาสต่างๆ ที่กำหนดในออนโทโลยีสถานที่ท่องเที่ยวในประเทศไทย ในระดับที่ 3 ประกอบด้วย 28 คลาส แสดงในตารางที่ 3-2

ตารางที่ 3-2: รายละเอียดของคลาสในระดับ 3 ในออนโทโลยี

ลำดับ	คลาส	รายละเอียด	หมวดย่อยของ
1	Mountain	คลาสแทนข้อมูลรูปภาพภูเขา	Natural Place
2	Waterfall	คลาสแทนข้อมูลรูปภาพน้ำตก	Natural Place
3	Cave	คลาสแทนข้อมูลรูปภาพถ้ำ	Natural Place
4	Sea	คลาสแทนข้อมูลรูปภาพทะเล	Natural Place
5	Beach	คลาสแทนข้อมูลรูปภาพหาด	Natural Place
6	Bay	คลาสแทนข้อมูลรูปภาพอ่าว	Natural Place
7	Island	คลาสแทนข้อมูลรูปภาพเกาะ	Natural Place
8	Dam	คลาสแทนข้อมูลรูปภาพเขื่อน	Natural Place
9	Flower Field	คลาสแทนข้อมูลรูปภาพสวนดอกไม้	Natural Place
10	Hot Spring	คลาสแทนข้อมูลรูปภาพบ่อน้ำพุร้อน	Natural Place
11	National Park	คลาสแทนข้อมูลรูปภาพอุทยานแห่งชาติ	Natural Place
12	Archaeological Site	คลาสแทนข้อมูลรูปภาพ	Heritage Place
13	Monument	คลาสแทนข้อมูลรูปภาพอนุสาวรีย์	Heritage Place
14	Royal Palace	คลาสแทนข้อมูลรูปภาพพระราชวัง	Heritage Place
15	Historical Museum	คลาสแทนข้อมูลรูปภาพพิพิธภัณฑ์ประวัติศาสตร์	Heritage Place
16	Temple	คลาสแทนข้อมูลรูปภาพวัด	Culture Place
17	Art Gallery	คลาสแทนข้อมูลรูปภาพหอศิลปะ	Culture Place

(ตารางมีต่อ)

ตารางที่ 3-2 (ต่อ): รายละเอียดของคลาสในระดับ 3 ในออนโทโลยี

ลำดับ	คลาส	รายละเอียด	หมวดย่อยของ
18	Cultural Museum	คลาสแทนข้อมูลรูปภาพพิพิธภัณฑ์ด้านวัฒนธรรม	Culture Place
19	Royal Project	คลาสแทนข้อมูลรูปภาพโครงการหลวง	Educational
20	Botanical Garden	คลาสแทนข้อมูลรูปภาพสวนพฤกษศาสตร์	Educational
21	Science Museum	คลาสแทนข้อมูลรูปภาพพิพิธภัณฑ์วิทยาศาสตร์	Educational
22	Village	คลาสแทนข้อมูลรูปภาพหมู่บ้าน	Traditional Path
23	Market	คลาสแทนข้อมูลรูปภาพตลาด	Traditional Path
24	Floating Market	คลาสแทนข้อมูลรูปภาพตลาดน้ำ	Traditional Path
25	Theater	คลาสแทนข้อมูลรูปภาพโรงละคร	Recreational Place
27	Zoo	คลาสแทนข้อมูลรูปภาพสวนสัตว์	Recreational Place
27	Animal Farm	คลาสแทนข้อมูลรูปภาพฟาร์มเลี้ยงสัตว์	Recreational Place
28	Amusement Park	คลาสแทนข้อมูลรูปภาพสวนสนุก	Recreational Place

รายละเอียดของการกำหนดคุณสมบัติของชนิดข้อมูล (Datatype Property) ทั้งหมด 6 พร็อพเพอร์ตี้ แสดงในตารางที่ 3-3

ตารางที่ 3-3: รายละเอียดของคุณสมบัติของชนิดข้อมูลในออนโทโลยีสถานที่ท่องเที่ยว

ลำดับที่	Property	Data Type	รายละเอียด
1	ImageID	Integer	เป็นการระบุค่าคงที่เป็นหมายเลขของรูปภาพแต่ละรูป

(ตารางมีต่อ)

ตารางที่ 3-3 (ต่อ): รายละเอียดของคุณสมบัติของชนิดข้อมูลในออนโทโลยีสถานที่ท่องเที่ยวในประเทศไทย

ลำดับที่	Property	Data Type	รายละเอียด
2	ImageURL	String	เป็นการระบุค่าคงที่ แสดง URL ของรูปภาพ
3	TouristType	String	เป็นการระบุค่าคงที่ แสดงประเภทสถานที่ท่องเที่ยว
4	TouristCategory	String	เป็นการระบุค่าคงที่ แสดงหมวดย่อยสถานที่ท่องเที่ยว
5	Name	String	เป็นการระบุค่าคงที่แสดงชื่อของสถานที่
6	Location	String	เป็นการระบุค่าคงที่แสดงสถานที่ตั้ง

3.3 การบรรยายออนโทโลยีด้วยภาษา OWL (Ontology Web Language)

หลังจากการออกแบบโครงสร้างข้อมูลออนโทโลยีเพื่อบรรยายถึงสถานที่ท่องเที่ยวในประเทศไทย จะต้องนำโครงสร้างข้อมูลออนโทโลยีที่ออกแบบไว้มาบรรยายด้วยภาษาสำหรับการบรรยายออนโทโลยี ในงานวิจัยนี้ได้เลือกใช้ภาษา OWL (Ontology Web Language) ในการบรรยายออนโทโลยี

3.3.1 การกำหนด Namespace

เพื่อใช้ในการอ้างอิงเอกสาร การกำหนด Namespace ที่เขียนด้วยภาษา OWL ดังที่แสดงในภาพที่ 3-3

ภาพที่ 3-3: ตัวอย่างการกำหนด Namespace ด้วยภาษา OWL

```
<?xml version="1.0"?>
<rdf:RDF
        xmlns="http://www.owl-ontologies.com/Ontology1204346619.owl#"
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:owl="http://www.w3.org/2002/07/owl#"
        xml:base="http://www.owl-ontologies.com/Touris.owl">
```

3.3.2 การอธิบายคลาสของออนโทโลยี

การอธิบายคลาส ใช้อ้างอิงในลักษณะลำดับชั้นในออนโทโลยี ซึ่งประกอบด้วยคำสั่งดังที่ปรากฏในตารางที่ 3-4

ตารางที่ 3-4: การอธิบายคลาสในภาษา OWL

ภาษา OWL	คำอธิบาย
owl:class	ใช้บรรยายคลาสซึ่งใช้อ้างอิงไปยังฐานข้อมูล
owl:subClassOf	ใช้บรรยายคลาสย่อยซึ่งใช้อ้างอิงไปยังคุณสมบัติต่าง ๆ ที่ใช้สร้างตาราง
owl:onProperty	ใช้บรรยายคุณสมบัติคลาสซึ่งใช้อ้างอิงไปยังฟิลด์ข้อมูลในตาราง
owl:allValueFrom	ใช้บรรยายประเภทข้อมูลซึ่งใช้อ้างอิงประเภทฟิลด์ที่ถูกสร้าง

ตัวอย่างการอธิบายคลาสน้ำตก (Waterfall) ซึ่งเป็นคลาสย่อยของสถานที่ท่องเที่ยวเชิงธรรมชาติ (Natural Place) โดยอธิบายด้วยคำสั่ง owl:subClassOf ส่วนคำสั่ง owl:disjointWith เป็นการระบุถึง Instance ใดที่เป็นสมาชิกของคลาส Waterfall แล้วจะไม่สามารถเป็น Instance ของคลาสอื่นในระดับชั้นเดียวกันได้อีก ดังภาพที่ 3-4

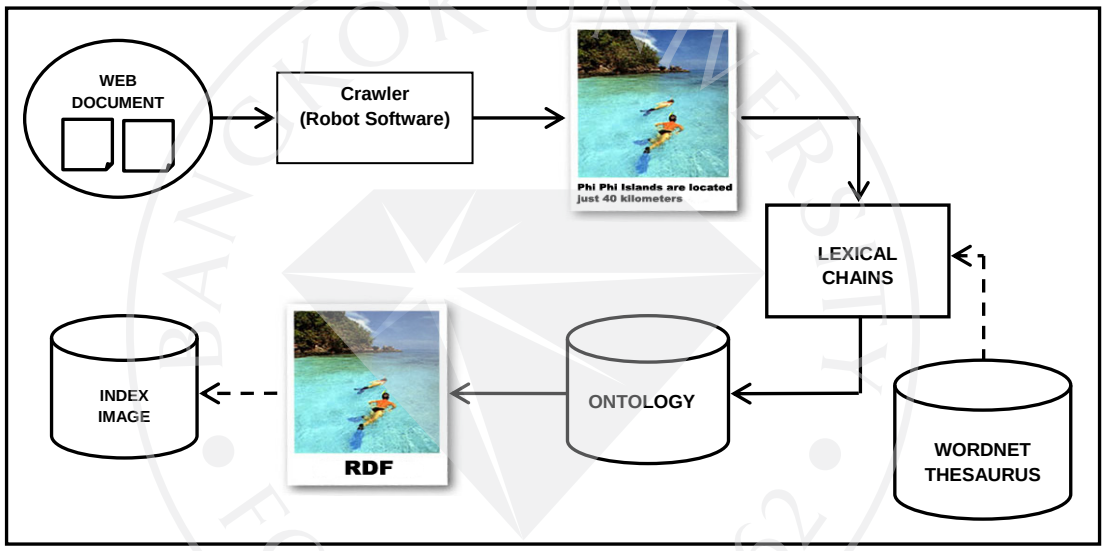
ภาพที่ 3-4: ตัวอย่างการอธิบายคลาสด้วย OWL

```
<owl:Class rdf:ID="Waterfall">
   < owl:subClassOf rdf:resource="#NaturalPlace"/>
   <owl:disjointWith rdf:resource="#Sea"/>
   <owl:disjointWith rdf:resource="#Island"/>
   <owl:disjointWith rdf:resource="#Bay"/>
   <owl:disjointWith rdf:resource="#Beach"/>
   <owl:disjointWith rdf:resource="#Cave"/>
</owl:Class>
```

3.4 การเตรียมข้อมูลสำหรับการสืบค้น

การสืบค้นรูปภาพซึ่งไม่มีการจัดเตรียมข้อมูลก่อนการดำเนินการสืบค้น ผลลัพธ์ของการสืบค้นจะมีประสิทธิภาพและความแม่นยำในระดับต่ำ ดังนั้นการจัดเตรียมข้อมูลก่อนดำเนินการสืบค้นจึงมีความสำคัญต่อประสิทธิภาพของการสืบค้นรูปภาพ งานวิจัยนี้จึงได้มีการออกแบบขั้นตอนการนำเข้าข้อมูลรูปภาพ เพื่อจัดรูปแบบข้อมูลรูปภาพก่อนการสืบค้นรูปภาพ ดังภาพที่ 3-5

ภาพที่ 3-5: แสดงขั้นตอนการนำเข้าข้อมูล



เป็นขั้นตอนการประมวลผลเพื่อเตรียมข้อมูลรูปภาพให้อยู่ในรูปแบบที่เหมาะสมสำหรับการสืบค้น เป็นการประมวลผลแบบ Batch Processing นั่นคือ ระบบจะส่งซอฟต์แวร์หุ่นยนต์ (Robot Software/Crawler) เพื่อไปอ่านเอกสารที่มีอยู่บนเว็บ (ในขอบเขตของโดเมนที่กำหนด) เพื่อมาประมวลผลให้ได้ข้อมูลที่ต้องการสำหรับสืบค้น จากนั้นนำข้อมูลที่ซอฟต์แวร์หุ่นยนต์เก็บรวบรวมมาได้ มาทำการคัดกรองข้อมูล (Data Cleaning) และสร้างดัชนีให้กับรูปภาพแต่ละภาพแล้วเก็บเป็นฐานความรู้ในรูปแบบที่กำหนด

เป้าหมายของขั้นตอนการเตรียมข้อมูลสำหรับการสืบค้น การได้กลุ่มคำหลักจากคำบรรยายของแต่ละภาพ เพื่อทำหน้าที่เป็นตัวแทนเชิงความหมายของแต่ละรูปภาพ เพื่อเป็นข้อมูลสำหรับการเปรียบเทียบระดับความเหมือน กับข้อมูลการสืบค้น (Query) ของผู้ใช้ ซึ่งงานในขั้นตอนนี้ประกอบด้วยงานเป็นลำดับดังนี้ คือ การสร้าง Lexical Chains จากคำบรรยายภาพ การบรรยายรูปภาพด้วยภาษา RDF และการทำดัชนี

3.4.1 การคัดกรองข้อมูล (Data Cleaning)

เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป ให้เหลือไว้เฉพาะข้อมูลที่เราต้องการนำไปใช้เท่านั้น ข้อมูลที่ซอฟต์แวร์หุ่นยนต์เก็บรวบรวมจากภายในเว็บไซต์ที่กำหนด จะถูกแปลงให้อยู่ในรูปของฐานข้อมูล สำหรับกระบวนการในการคัดข้อมูลมีดังต่อไปนี้

3.4.1.1 ทำการลบคอลัมน์ที่ไม่ต้องการนำไปใช้ออกไป

3.4.1.2 ทำการกำจัดเร็คคอร์ดที่ซ้ำกันออกไป โดยในที่นี้คีย์หลักของเรา (Primary Key) คือ URL ของรูปภาพหรือที่อยู่เฉพาะของรูปภาพภายในเว็บไซต์ ดังนั้นผลลัพธ์ที่ได้ URL ของรูปภาพจะไม่มีการซ้ำกันเลย

3.4.1.3 ทำการกำจัดข้อมูลรูปภาพ ที่มีการลิงค์หรือเชื่อมโยงไปยังหน้าเว็บไซต์ทั้งภายในโดเมนและภายนอกโดเมน เนื่องจากรูปภาพเหล่านี้ จะมีลักษณะการทำงานเป็นปุ่ม (Button) และเป็นรูปภาพประเภทแบนเนอร์โฆษณาประชาสัมพันธ์ต่างๆ

3.4.2 การสร้าง Lexical Chains

มอร์ริชและเฮิร์ต (Morris, & Hirst,1991) ได้ทำการออกแนวคิดของ Cohesion Chain หรือเรียกในอีกชื่อหนึ่งว่า Lexical Chains ในการสร้าง Lexical Chains เป็นการนำคำบรรยายรูปภาพมาทำการสร้างความสัมพันธ์ของคำที่สัมพันธ์กันในระดับความหมายนำมาไว้ในสายของลิสต์เดียวกัน ซึ่งวิธีดังกล่าวนี้ใช้หลักการเดียวกันการใช้สำคัญ (Keyword) ของเอกสารนำมาเป็นแทนเอกสาร แต่ได้เพิ่มการพิจารณาระดับคำ (word cohesion) เข้ามาร่วม ขั้นตอนการสร้าง Lexical Chain มีดังนี้

3.4.2.1 การนำเข้าข้อความบรรยายรูปภาพ ที่ได้จากหน้าเว็บเพจ เพื่อนำมาสร้าง Lexical Chains ของคำบรรยายภาพ

ภาพที่ 3-6: ตัวอย่างรูปภาพเกาะพีพี

ภาพที่ 3-7: ตัวอย่างคำบรรยายรูปภาพเกาะพีพี

> Regarded by travelers from all walks of life as being among the most beautiful tropical islands of the world, Phi-Phi Islands have become one of the most popular tourist attractions in Thailand. a total area of 390 square kilometers, were recently established as a national park Phi Phi Islands are located just 40 kilometers away from Krabi

3.4.2.2 ในขั้นตอนนี้จะมีการนำฐานความรู้เวิร์ดเนต (WordNet) เป็นฐานข้อมูลคำศัพท์ภาษาอังกฤษโดยการแบ่งการเก็บข้อมูลคำศัพท์ตามชนิดของคำ เช่น คำนาม คำกริยา คำวิเศษณ์ เป็นต้น มาใช้ในการเลือกเฉพาะคำนามที่อยู่ในคำบรรยายภาพ เพราะคำที่สามารถสื่อความหมายได้ โดยไม่สนใจคำประเภทต่อไปนี้

- คำสรรพนาม (Pronoun)
- คำบุพบท (Preposition)
- คำคุณศัพท์ (Adjective)
- คำกริยาวิเศษ (Adverb)
- คำกริยาต่างๆ (Verb)

จากภาพที่ 3-7 ที่เป็นคำบรรยายภาพเมื่อนำมาคัดเลือกคำเฉพาะคำนาม ผลลัพธ์ที่ได้จากการคัดเลือกคำนามจากคำบรรยายภาพด้วยฐานข้อมูลเวิร์ดเนต

3.4.2.3 นำผลลัพธ์ที่ได้จากเลือกคำนามจากขั้นตอนที่ 2 ทำการสร้าง Lexical Chains โดยใช้ฐานความรู้ Thesaurus ซึ่งเป็นฐานข้อมูลคำศัพท์ที่มีการแยกประเภทคำศัพท์ออกเป็นหมวดหมู่โดยจัดคำศัพท์ที่มีความหมายเหมือนกันไว้ด้วยกัน เพื่อสร้างความสัมพันธ์ของคำที่มีความหมายในระดับเดียวกัน นำมาไว้ในสายลิสต์เดียวกัน

3.4.3 การอธิบายโครงสร้างข้อมูลรูปภาพด้วยภาษา OWL

เป็นกระบวนการนำเอา Lexical Chains ที่ได้จากกระบวนการสกัดคำที่เกี่ยวข้องกับสถานที่ท่องเที่ยวในประเทศไทย มาจัดเก็บในรูปแบบของภาษา OWL ที่เป็นภาษาที่ใช้ในการอธิบายทรัพยากรบนอินเตอร์เน็ต ตามออนโทโลยีที่ได้ทำการออกแบบเอาไว้ เพื่อใช้ในการสืบค้นข้อมูลรูปภาพ

ภาพที่ 3-8: ตัวอย่างการอธิบายโครงสร้างรูปภาพด้วย RDF

```
<ImageID rdf:datatype="http://www.w3.org/2001/XMLSchema#int">10000008</
ImageID>
<TouristType rdf:datatype="http://www.w3.org/2001/XMLSchema#string">#Natural
Place </ TouristType >
<TouristCategory rdf:datatype="http://www.w3.org/2001/XMLSchema#string"># Island
</ TouristCategory >
<Name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Samed
Island</Name>
<Location>rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Rayong
</Location>
```

จากภาพที่ 3-8 เป็นการอธิบายข้อมูลรูปภาพด้วยภาษา RDF (Resource Description Framework) โดยประกอบด้วย rdf:Description เป็นการระบุถึงสิ่งที่ต้องการบรรยาย ในตัวอย่างนี้คือ URL ของรูปภาพที่ต้องการบรรยาย “http://www.tourist.com/D0001.jpg” โดยที่รูปภาพเป็นรูปภาพประเภท (Image Type) เป็นรูปภาพสถานที่ท่องเที่ยว มีประเภทสถานที่ท่องเที่ยว (Tourist

Type) เป็นสถานที่ท่องเที่ยวเชิงธรรมชาติ ที่เป็นเกาะ (Island) โดยมีคุณสมบัติเฉพาะคือชื่อ “เกาะพีพี” และที่ตั้งอยู่ที่ “จังหวัดกระบี่”

3.5 ขั้นตอนการสืบค้นข้อมูลรูปภาพ

ขั้นตอนการสืบค้นข้อมูลเริ่มจาก การพิจารณาคำ, วลีสืบค้นที่ผู้ใช้นำเข้า เลือกคำสำคัญพร้อมทั้งคำที่มีความหมายเหมือนกัน (Synonym) หลังจากนั้น นำกลุ่มคำสำคัญไปค้นหาในฐานข้อมูล

3.5.1 พิจารณาคำ, วลีสืบค้น

ที่ผู้ใช้ทำการสืบค้นโดยเลือกเฉพาะคำสำคัญ (Candidate word) จากคำหรือวลีสืบค้นซึ่งเกี่ยวข้องกับการอธิบายรูปภาพสถานที่ท่องเที่ยว และทำการแบ่งกลุ่มตามประเภทของคำศัพท์เกี่ยวกับการแบ่งสถานที่ท่องเที่ยวที่นิยามไว้ดังภาพที่ 3-9 ตัวอย่าง “Picture of beach in koh phuket.” จากวลีดังกล่าววิเคราะห์คำสำคัญได้ดังนี้ {beach, koh, phuket} นำคำสำคัญดังกล่าวมาเลือกคำศัพท์ที่มีความหมายเหมือนกัน จากฐานข้อมูลซึ่งใช้มาตรฐานของ RDFS อธิบายข้อมูล ตัวอย่างเช่น

ภาพที่ **3-9** อธิบายข้อมูลด้วยมาตรฐาน **RDFS**

```
<rdf:Description rdf:nodeID="Island"
        vCard:FN="Island">
        <rdfs:subClassOf rdf:nodeID="Nature"/>
        <rdfs:seeAlso rdf:nodeID="Koh"/>
        <rdfs:seeAlso rdf:nodeID="isle"/>
        <rdfs:seeAlso rdf:nodeID="archipelago"/>
</rdf:Description>
```

จากโครงสร้างดังกล่าวเป็นการอธิบายข้อมูลของภัยพิบัติ(Disaster) ซึ่งคำศัพท์ดังกล่าวจะมีความหมายเหมือนกับคำว่า Catastrophe และ Calamity โดยคุณสมบัติ seeAlso ของ RDFS เป็นตัวช่วยในการอธิบายลักษณะที่เหมือนกันของข้อมูล

จากคำสำคัญที่เลือกพิจารณา มีดังนี้

{Beach, shore, coast}

{koh, isle, island, archipelago}

{phuket}

3.5.2 การสืบค้นข้อมูลจากฐานข้อมูล นำกลุ่มคำสืบค้นจากขั้นตอน 3.5.1 มาสืบค้นข้อมูลในฐานข้อมูลโดยมีขั้นตอนดังนี้

ภาพที่ 3-10 ขั้นตอนการสืบค้นข้อมูล

```
REPEAT
READ next กลุ่มคำสืบค้น(i)
CASE TouristType (หมวดประเภทสถานที่ท่องเที่ยว)
        Result1(i)= Search กลุ่มคำสืบค้น(i) จากฐานข้อมูล
CASE TouristCategory (หมวดย่อยสถานที่ท่องเที่ยว)
        Result2(i)= Search กลุ่มคำสืบค้น(i) จากฐานข้อมูล

CASE Name (ชื่อสถานที่ท่องเที่ยว)
        Result3(i)= Search กลุ่มคำสืบค้น(i) จากฐานข้อมูล
IF Result1(i) AND Result1(i+1) AND…. Result2(i) AND Result2(i+1)
            Result3(i) AND Result3(i+1)..not NULL
        THEN
            BestPic = Result1(i) □ Result1(i+1) □ ….
            Result2(i) □ Result2(i+1) □ ….
            Result3(i) □ Result3(i+1) □ ….
        ELSE
            NormalPic =  Result2(i) + Result2(i+1) + … or  Result3(i) + Result3(i+1) + ….
            WorsePic = Result1(i) + Result1(i+1) + ….
END IF
END REPEAT
```

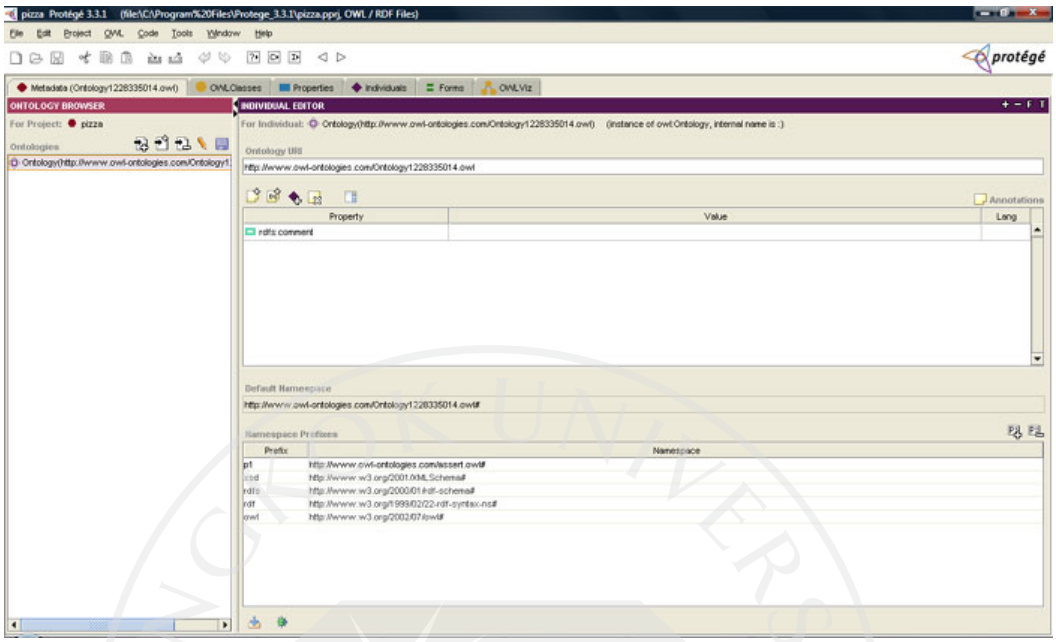
จากขั้นตอนการสืบค้นดังกล่าว เริ่มจาก การนำกลุ่มคำสืบค้นมาแยกประเภทตามคุณสมบัติของตัวแบบ(หมวดประเภทสถานที่ท่องเที่ยวและหมวดย่อยของสถานที่ท่องเที่ยว) หลังจากนั้น นำคำสืบค้นไปค้นหาในฐานข้อมูล โดยผลลัพธ์ จากภาพที่ 3-10 แบ่งออกเป็น 3 ส่วน คือผลลัพธ์ของแต่ละคุณสมบัติ (Result1,Result2,Result3) และกลุ่มคำสืบค้น (Result1(i),- Result(i+1)) พิจารณาผลลัพธ์ดังนี้ ถ้าผลลัพธ์ที่ค้นพบได้จากผลลัพธ์ทุกกลุ่ม (Result1(i)..,Result2(i), Result3(i)…) แสดงว่ามีระดับความแม่นยำมาก ถ้าผลลัพธ์ที่ค้นพบได้จากผลลัพธ์ของบางกลุ่ม ให้ผลลัพธ์ที่ได้จากการค้นหาในระดับหมวดย่อยของสถานที่ท่องเที่ยว มีระดับความแม่นยำปานกลาง และผลลัพธ์ได้ที่จากระดับหมวดประเภทสถานที่ท่องเที่ยว มีระดับความแม่นยำน้อย ผลลัพธ์จากการสืบค้นข้อมูลเป็นผลลัพธ์ซึ่งมาจากคำค้นที่มีความหมายอยู่ในกลุ่มเดียวกัน เพราะฉะนั้นผลลัพธ์ที่ได้จะเป็นผลลัพธ์เชิงความหมาย

3.6 เครื่องมือที่ใช้ในการดำเนินงาน

3.6.1 Protégé

Protégé เป็นซอฟต์แวร์ประเภทโอเพ่นซอร์ส ใช้งานสำหรับการสร้างออนโทโลยีและฐานความรู้ รองรับการสร้างออนโทโลยีด้วยภาษา RDF/RDFS และ OWL พัฒนา ขึ้นโดยมหาวิทยาลัยสแตนฟอร์ด ประเทศสหรัฐอเมริกา มีส่วนการติดต่อ ผู้ใช้งานเป็นแบบกราฟิก (Graphical User Interface : GUI) รองรับการทำงานแบบหลายผู้ใช้ จัดเก็บออนโทโลยีในรูปแบบแฟ้มข้อมูลและฐานข้อมูลเชิงสัมพันธ์ มีเครื่องมือสำหรับสร้างโดเมนของออนโทโลยี และรูปแบบข้อมูลที่สะดวกในการป้อนข้อมูลโดยยอมให้ผู้ใช้ทำงานพร้อมกันบนคลาส หรืออินสแตนซ์ใหม่ และช่วยแก้ปัญหาที่เกิดขึ้นในการพัฒนาส่วนของวิธีการโดยหลาย ๆ โปรแกรมประยุกต์สามารถใช้งานโดเมน เพื่อแก้ปัญหาที่ต่างกัน และวิธีการนั้นสามารถนำไปประยุกต์ใช้กับออนโทโลยีที่ต่างกันได้ (Protégé Project, internet, 2007)

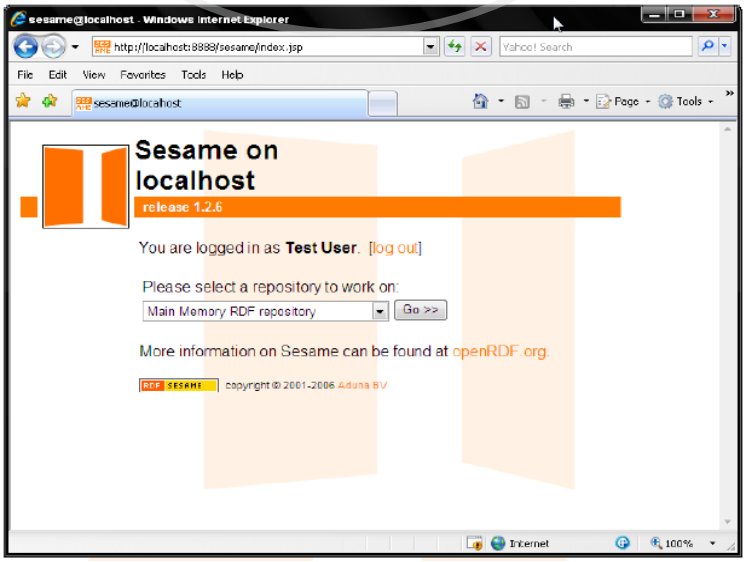
ภาพที่ 3-11: ตัวอย่างการใช้งานโปรแกรม Protégé ในการสร้างออนโทโลยี



3.6.2 เครื่องมือในการค้นหาข้อมูลในเอกสาร RDF

Sesame เป็นโอเพ่นซอร์สที่พัฒนาบนภาษาจาวาเพื่อจัดเก็บและสืบค้นสำหรับเอกสาร RDF และ RDF Schema สามารถนำ Library ไปพัฒนาต่อเป็นเว็บแอปพลิเคชั่นได้

ภาพที่ 3-12: ตัวอย่างการใช้งานโปรแกรม Sesame



3.6.3 เครื่องมือที่ใช้ในการคิวรีข้อมูล

Jena Framework เป็นโอเพ่นซอร์สสำหรับรองรับการพัฒนาแอพพลิเคชั่นเว็บเชิงความหมาย โดยใช้ภาษาจาวาในการพัฒนา โดยสามารถที่จะนำข้อมูลไปจัดเก็บในรูปแบบของภาษา RDF (Resource Description Framework) (Jena Framework, internet, 2008) Jena มีลักษณะเหมือนกับ Sesame แต่จะแตกต่างกันตรงที่ Jena นั่นรองรับได้ทั้งภาษา OWL, RDF, RDFS และ SPARQLส่วน Sesame นั่นรองรับเฉพาะ RDF และ RDFS เท่านั้น

3.6.4 เครื่องมือในการแสดงผลข้อมูลการสืบค้นรูปภาพ

ในขั้นตอนนี้จะเป็นการพัฒนาเว็บแอพพลิเคชันสำหรับการสืบค้นข้อมูลรูปภาพ ซึ่งได้อาศัย JSP (Java Server Pages) ในการพัฒนา เป็นเทคโนโลยีที่มีลักษณะคล้าย ASP (Microsoft Active Server Pages) แต่จะมีข้อแตกต่างคือ JSP พัฒนาโดยใช้ภาษาจาวา ซึ่งเป็นภาษาเชิงวัตถุ ซึ่งช่วยทำให้ง่ายต่อการพัฒนาในโครงการขนาดใหญ่ ตลอดจนสามารถนำส่วนประกอบต่าง ๆ กลับมาใช้ได้อีก

3.7 การทดสอบวัดประสิทธิภาพความแม่นยำในการสืบค้น

ประสิทธิภาพของแบบจำลองในการสืบค้นรูปภาพเชิงความหมายบนเว็บไซด์นี้ วัดจากระดับความแม่นยำ (Precision) และการเรียกซ้ำ (Recall) ของการสืบค้น ดังสมการดังภาพที่ 3-13 เพื่อทดสอบประสิทธิภาพโครงสร้างของข้อมูลของออนโทโลยีสำหรับบรรยายสถานที่ท่องเที่ยวในประเทศที่ได้ออกแบบเอาไว้ และขั้นตอนของการสืบค้น

ภาพที่ 3-13: สมการในการวัดความแม่นยำ (Precision) และการเรียกซ้ำ (Recall)

$$\text{Precision} = \frac{A}{A + B} \times 100\ \%$$

$$\text{Recall} = \frac{A}{A + C} \times 100\ \%$$

เมื่อ

A	คือ จำนวนรูปภาพที่สามารถสืบค้นได้และถูกต้อง
B	คือ จำนวนรูปภาพที่สามารถสืบค้นได้แต่ไม่ถูกต้อง
C	คือ จำนวนรูปภาพที่ต้องการ แต่ไม่ถูกอ่านขึ้นมา

ในการทดสอบประสิทธิภาพได้จัดเตรียมรูปภาพสถานที่ท่องเที่ยวในประเทศจำนวน 200 รูปที่ถูกบรรยายและจัดเก็บอยู่ในรูปแบบของ RDF Format และทำการสืบค้นโดยใช้คำหลักทั้งหมด 30 คำ และทำการวัดประสิทธิภาพโดยใช้สมการวัดความแม่นยำ (Precision) และการเรียกซ้ำ (Recall) ของการสืบค้นแต่ละคำหลัก เพื่อหาเฉลี่ยของการสืบค้นจำนวน 30 ครั้ง

บทที่ 4
ผลการวิจัยและอภิปรายผล

4.1 ผลการวิจัย

จากวิธีการดำเนินงานวิจัย ซึ่งนำคำบรรยายรูปภาพสถานที่ท่องเที่ยวในประเทศไทย มาบรรยายตามโครงสร้างของออนโทโลยี เพื่อแทนความหมายของข้อมูลรูปภาพได้อย่างมีความหมายตามแนวความคิดของเว็บเชิงความหมาย

ผลลัพธ์จากการสืบค้นด้วยวิธีการพิจารณาความเหมือนของข้อมูลสืบค้น (Query) สามารถสรุปได้ว่า ข้อมูลรูปภาพที่ค้นพบ เป็นข้อมูลที่มีความหมายตรงกับข้อมูลสืบค้นหรือมีความหมายเหมือนกับข้อมูลสืบค้น และในกรณีที่ในหน้าเว็บไซต์ประกอบด้วยรูปภาพ 1 ภาพ ผลลัพธ์ที่ได้จะตรงกับความหมายที่แท้จริงของรูปภาพมากกว่า กรณีที่ในหน้าเว็บไซต์มีรูปภาพตั้งแต่ 2 ภาพขึ้นไป เนื่องจากในการขั้นตอนการเก็บรวบรวมข้อมูลยังไม่สามารถระบุได้ว่าคำบรรยายภาพในส่วนใดเป็นคำบรรยายของแต่ละภาพ

ในขณะที่ผลการทดลองสืบค้น โดยใช้วิธีการของ Precision และ Recall ซึ่งเป็นวิธีการหนึ่งสำหรับการประเมินประสิทธิภาพการสืบค้นเอกสาร พบว่าผลการประเมินของทั้ง Precision และ Recall อยู่ในระดับปานกลาง

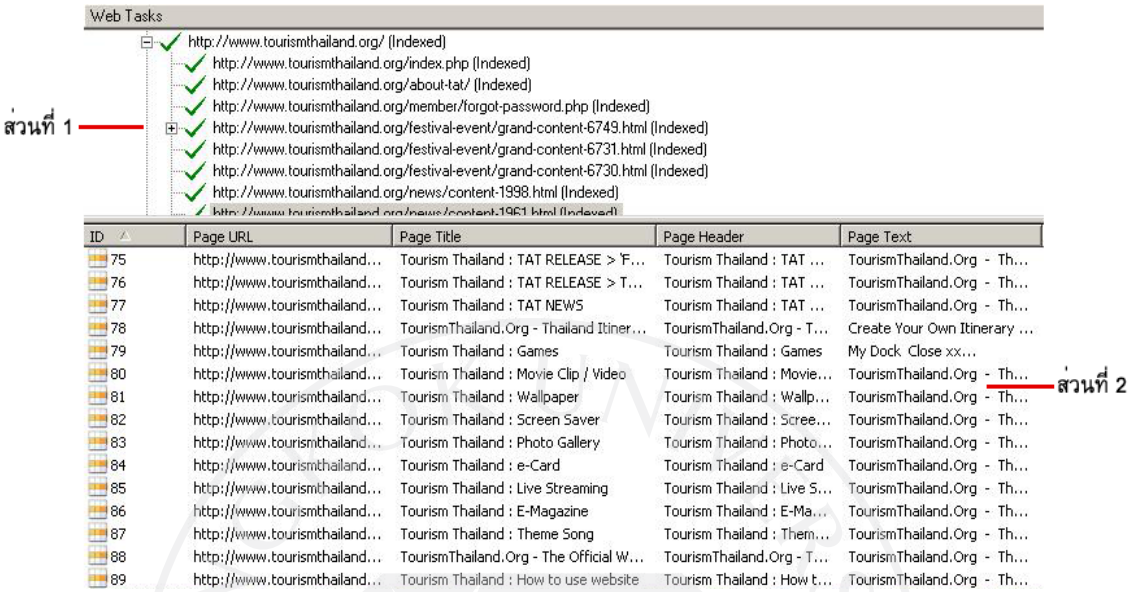
4.2 การอภิปรายผล

4.2.1 การเก็บรวบรวมข้อมูลรูปภาพและคำบรรยายภาพ

Web Crawler คือซอฟต์แวร์ที่ทำหน้าที่ในการเก็บรวบรวมเอกสารอิเล็กทรอนิกส์บนเว็บทั้งหมด ในงานวิจัยนี้จะใช้ Web Crawler ในการเก็บรวบรวมข้อมูลรูปภาพสถานที่ท่องเที่ยวในประเทศไทย พร้อมคำบรรยายรูปภาพ ซึ่งวิธีการทำงานของโปรแกรมคือป้อน URL เริ่มต้นก่อนคือ http://www.tourismthailand.org จากนั้นโปรแกรมจะทำการเปิดหน้าเว็บขึ้นมาและทำการดึงข้อมูลต่างๆ ที่มาแสดงผลดังภาพที่ 4-1 ซึ่งสามารถอธิบายได้ดังนี้

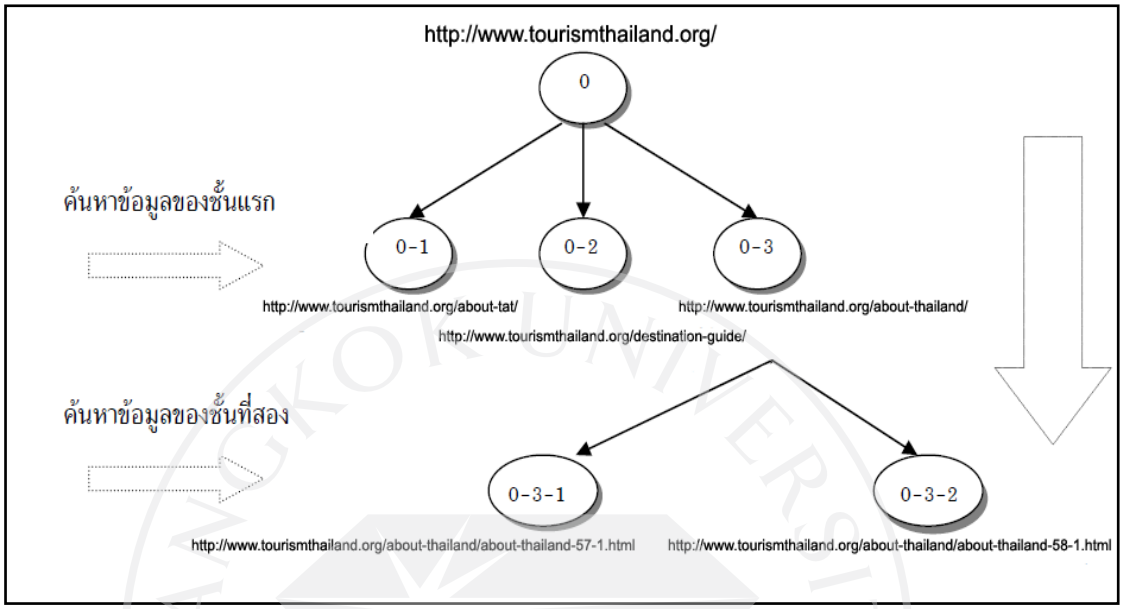
ส่วนที่ 1 แสดงข้อมูลในส่วนของการแสดงการทำงานของ Web Crawler และเส้นทางในการเก็บรวบรวมข้อมูล ส่วนที่ 2 แสดงข้อมูลในส่วนของ Hyperlink, Tag Title, Tag Body และ Tag ต่างๆ ที่อยู่ภายในหน้าเว็บ และเมื่อได้ข้อมูลครบถ้วนภายในเว็บทั้งหมดแล้วจะนำข้อมูลที่เก็บรวบรวมได้บันทึกลงในฐานข้อมูล

ภาพที่ 4-1: แสดงหน้าจอโปรแกรม Web Crawler



จากภาพที่ 4-2 เป็นภาพแสดงเส้นทางการเดินทางของตัว Crawler ที่ใช้ท่องไปในแต่ละเอกสารบนอินเตอร์เน็ต เพื่อค้นหาข้อมูลในแต่ละแท็ก (Tag) แต่ละลิงค์ (Link) โดยเริ่มต้นที่เว็บ http://www.tourismthailand.org จากนั้นจะมาที่โหนด http://www.tourismthailand.org/about-tat/ และโหนด http://www.tourismthailand.org/about-thailand/ ตามลำดับ เมื่อเดินทางไปยังแต่ละโหนดของชั้นแรกเสร็จ โปรแกรมจะท่องไปยังชั้นที่ 2 โดยไปในแต่ละโหนดจนครบไล่จากฝั่งทางซ้ายไปขวาจนครบทุกโหนด

ภาพที่ 4-2: แสดงการเก็บข้อมูลของโดยการกระจายแบบวิธี Breadth first search



จากภาพที่ 4-3 เป็นภาพที่แสดงถึงข้อมูลที่ Web Crawler เก็บรวบรวมมาได้และแปลงให้อยู่ในรูปของฐานข้อมูล เพื่อนำมาทำทำความสะอาดข้อมูล (Data Cleaning) เพื่อให้เหลือเฉพาะข้อมูลที่เราต้องการนำไปใช้เท่านั้น ข้อมูลที่เก็บรวบรวมได้มีดังนี้

- Page URL เป็นตำแหน่งหรือที่อยู่เฉพาะของแต่ละเว็บเพจภายในเว็บไซต์
- Page Title เป็นข้อมูลที่อยู่ภายในแท็ก Title เป็นส่วนที่แสดงชื่อของเอกสาร
- Page Header เป็นข้อมูลที่อยู่ภายในแท็ก Head เป็นส่วนที่แสดงชื่อเรื่องของแต่ละหน้าภายในเว็บไซต์
- Page Body เป็นข้อมูลที่อยู่ภายในแท็ก Body เป็นส่วนที่แสดงเนื้อหาของหน้าเว็บเพจนั้นๆ
- Meta Tag เป็นข้อมูลที่อยู่ภายในแท็ก Meta เป็นส่วนไว้สำหรับใส่คำอธิบายข้อมูล หรือข้อมูลที่บอกรายละเอียดของข้อมูล (Meta data)
- Img URL เป็นตำแหน่งหรือที่อยู่เฉพาะของรูปภาพภายในเว็บไซต์
- Img Link เป็นลิงค์

ภาพที่ 4-3: ข้อมูลที่ได้จากการเก็บรวบรวมของ Web Crawler

Page URL	Page Title	Page Header	Page Body
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Doi Inthanon National Park
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Phra That Si Chom Thong Te
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Phra That Si Chom Thong Te
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Phra That Si Chom Thong Te
http://www.tourismthailand.org/attraction,	Tourism Thailand : Destina	TourismThailand.Org	Chiang Dao Cave Tham Chia
http://www.tourismthailand.org/attraction,	TourismThailand.Org - The	TourismThailand.Org	Chiang Dao Cave Tham Chia
http://www.tourismthailand.org/attraction,	TourismThailand.Org - The	TourismThailand.Org	Chiang Dao Cave Tham Chia
http://www.tourismthailand.org/attraction,	Tourism Thailand : Multim	TourismThailand.Org	Chiang Dao Cave Tham Chia

จำนวนเอกสารที่เก็บรวบรวมเอกสารบนเว็บมีทั้งหมด 14,599 เอกสาร จากนั้นจะนำเอกสารที่เก็บรวบรวมทำผ่านกระบวนการคัดกรองข้อมูล (Data Cleaning) เป็นคัดกรองให้เหลือเฉพาะข้อมูลที่ต้องการเท่านั้น ก่อนที่จะนำข้อมูลดังกล่าวมาสกัดค่าคำสำคัญ แล้วนำไปบรรยายตามโครงสร้างของออนโทโลยีที่ได้มีการออกแบบเอาไว้ สำหรับขั้นตอนกระบวนการในการคัดกรองข้อมูล ดังที่ระบุไว้ในบทที่ 3 ผลลัพธ์ที่ได้หลังจากการทำความสะอาดข้อมูลคือจำนวนข้อมูล 1,641 ภาพ

ภาพที่ 4-4 เป็นการแสดงถึงผลลัพธ์ของข้อมูลที่ได้ หลังจากผ่านกระบวนการทำความสะอาดข้อมูล ผลลัพธ์ที่ได้ประกอบด้วย Page URL, IMG URL และ Page Body ซึ่งส่วนคำบรรยายภาพในส่วนของ Page Body เพื่อนำไปเตรียมข้อมูลสำหรับการสืบค้น เพื่อได้กลุ่มคำหลักจากคำบรรยายของแต่ละภาพ เพื่อทำหน้าที่เป็นตัวแทนเชิงความหมายของแต่ละรูปภาพ เพื่อเป็นข้อมูลสำหรับการเปรียบเทียบระดับความเหมือน กับข้อมูลการสืบค้น (Query) ของผู้ใช้

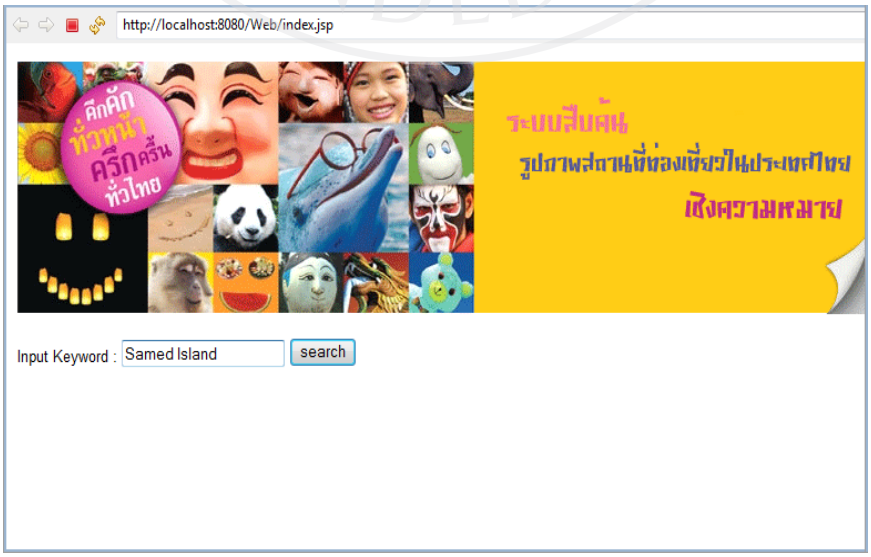
ภาพที่ 4-4: แสดงผลลัพธ์ที่ได้หลังจากผ่านการทำความสะอาดข้อมูล

Page URL	Img URL	Page Body
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	A favorite get-away island for Bangkokians, Ko San
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	A favorite get-away island for Bangkokians, Ko San
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	A favorite get-away island for Bangkokians, Ko San
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Also dubbed "Wat Lum Pradu" which offers intere:
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Suan Sri Muang is a public park and health center. I
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Hat Laem Charoen This beach is located 5 kilomete
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Hat Mae Ramphueng This 12-kilometre stretch of v
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Wat Saranat Thammaram Near the turning point at
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Five kilometers further on from Laem Mae Phim oı
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Laem Mae Phim This cape can be reach either from
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Ko Man Nai This small island located approximatel
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Suphattraland Thirty kilometres from town is Suph
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Hat Phayun and Hat Phala are located approximate
http://www.tourismthailand.org/attraction/rayoı	http://www.tourismthailand.org/ajaxer	Namtok Klong Pla Kang Located some 3 kilometers
http://www.tourismthailand.org/attraction/roiet	http://www.tourismthailand.org/ajaxer	Sim Wat Chakkrawan Phum Phinit or Wat Nong Mu
http://www.tourismthailand.org/attraction/roiet	http://www.tourismthailand.org/ajaxer	It is situated in Wat Burapha Ku Ka Sing, Tambon Kı
http://www.tourismthailand.org/attraction/sako	http://www.tourismthailand.org/ajaxer	Phra That Choeng Chum is an historical site that ha
http://www.tourismthailand.org/attraction/sako	http://www.tourismthailand.org/ajaxer	Nong Han is a large freshwater lake with an area o
http://www.tourismthailand.org/attraction/samu	http://www.tourismthailand.org/ajaxer	Pom Phra Chulachomklao Naval History Park at Chu

4.2.2 การสืบค้นข้อมูลรูปภาพโดยใช้คำสำคัญ

ผู้ใช้งานกำหนดคำสำคัญของชื่อสถานที่ท่องเที่ยวในประเทศไทยที่ต้องการค้นหา จากนั้นกดปุ่ม Search เพื่อทำการรูปภาพสถานที่ท่องเที่ยว ที่มีข้อมูลตรงกับคำสำคัญที่ผู้ใช้กำหนด ภาพที่ 4-5 แสดงหน้าจอในขณะที่ผู้ใช้ใส่คำสำคัญคำว่า “Samed Island”

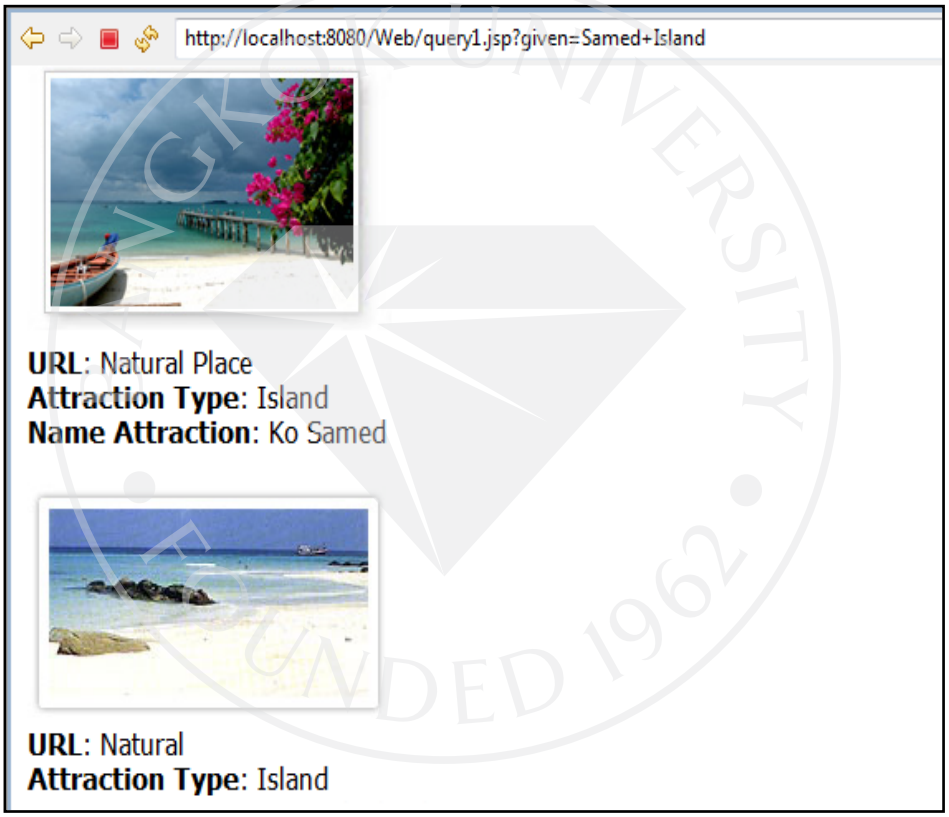
ภาพ 4-5: หน้าจอแสดงการสืบค้นข้อมูลรูปภาพ



หลังจากที่ผู้ใช้งานกดปุ่ม Search ถ้าระบบมีข้อมูลตรงกับคำสำคัญที่ผู้ใช้กำหนดผลลัพธ์ที่ได้จะถูกแบ่งออกเป็นสองส่วนด้วยกัน คือ

4.2.2.1 แสดงผลลัพธ์จากการคิวรีในส่วนของข้อมูลรูปภาพสถานที่ท่องเที่ยวในประเทศไทย โดยจะแสดงรูปภาพ, ชื่อสถานที่ท่องเที่ยว, หมวดย่อยของสถานที่ท่องเที่ยวและที่ตั้งของสถานที่ท่องเที่ยว ดังที่แสดงดังภาพที่ 4-6

ภาพที่ 4-6: แสดงผลลัพธ์จากการสืบค้นข้อมูล



4.2.2.2 เมื่อผู้ใช้งานต้องการทราบรายละเอียดเพิ่มเติมเกี่ยวกับรูปภาพ สามารถกดเรียกโดยทำการคลิกไปที่รูปภาพ จะแสดงผลลัพธ์ดังภาพที่ 4-7

ภาพที่ 4-7: แสดงรายละเอียดเพิ่มเติมเกี่ยวกับข้อมูลรูปภาพ

4.2.3 ผลการทดลองสืบค้น

หลังจากพัฒนาตัวแบบฐานข้อมูลองค์ความรู้สถานที่ท่องเที่ยวในประเทศไทย เพื่อสำหรับการสืบค้นข้อมูลรูปภาพสถานที่ท่องเที่ยวในประเทศไทย โดยมีจัดเก็บข้อมูลตามรูปแบบออนโทโลยีที่ได้มีการออกแบบเอาไว้ และนำมาจัดเก็บให้อยู่ในรูปแบบของเอกสาร RDF (Resource Description Framework)

ในการทดลองผลการสืบค้นจะมีการทดลองจำนวน 2 ครั้ง ในแต่ละครั้งจะมีการสุ่มตัวอย่างข้อมูลรูปภาพจำนวน 200 รูป จากนั้นทดลองค้นหาโดยใช้คำสำคัญในการสืบค้นข้อมูลจำนวน 20 คำสำคัญที่มีความหมายแตกต่างกัน และใช้วิธีการประเมินประสิทธิภาพการสืบค้นด้วยวิธีการของ Precision และ Recall ข้อมูลสืบค้นที่ใช้ในการทดลองผลการสืบค้น ดังที่แสดงในตารางที่ 4-1

ตารางที่ 4-1: คำและวลีสืบค้นที่ใช้การทดลอง

1. Monument	11. Natural Place
2. Island	12. Traditioanal Path
3. Koh chang	13. Cultural Place and Heritage Place
4. Phu chee fah	14. Koh Samed Rayong
5. Beach and Sea	15. National Park
6. Doi Inthanon	16. Waterfall and Mountain
7. Bay in Thailand	17. Palace
8. The highest Mountain	18. Waterfall in Chiangmai
9. Wat in north of Thailand	19. Koh phi phi
10. Cave	20. Long beach in samui

ในตารางที่ 4-2 เป็นเกณฑ์การแปลความหมายของข้อมูลที่ได้จากการประเมินโดยการใช้วิธีการวัดแบบประสิทธิภาพของการสืบค้นข้อมูล Precision และ Recall

ตาราง 4-2: เกณฑ์การแปลความหมายข้อมูลและพิจารณา จากค่า Precision และ Recall

ระดับเกณฑ์	คำหมาย
0.85-1.00	ประสิทธิภาพของระบบอยู่ในระดับดีมาก
0.75-0.84	ประสิทธิภาพของระบบอยู่ในระดับดี
0.55-0.74	ประสิทธิภาพของระบบอยู่ในระดับปานกลาง
0.35-0.54	ประสิทธิภาพของระบบอยู่ในระดับน้อย
0.00-0.34	ประสิทธิภาพของระบบอยู่ในระดับใช้ไม่ได้

การทดลองครั้งที่ 1 จากการทดลอง 20 ข้อมูลสืบค้นที่แตกต่างกัน 5 คำสืบค้นที่ไม่มีผลลัพธ์ ส่วนอีก 15 คำสืบค้นมีผลลัพธ์ ผลการวัดประสิทธิภาพในการสืบค้นด้วยวิธี Precision เท่ากับ 5.8 และวิธี Recall เท่ากับ 6.9

การทดลองครั้งที่ 2 จากการทดลอง 20 ข้อมูลสืบค้นที่แตกต่างกัน 4 คำสืบค้นที่ไม่มีผลลัพธ์ ส่วนอีก 16 คำสืบค้นมีผลลัพธ์ ผลการวัดประสิทธิภาพในการสืบค้นด้วยวิธี Precision เท่ากับ 6.6 และวิธี Recall เท่ากับ 7.1

สรุปผลการทดลอง ผลลัพธ์ที่ได้จากการทดลองทั้ง 2 ครั้งประสิทธิภาพของระบบอยู่ที่ระดับปานกลาง ตามเกณฑ์การแปลความหมายของข้อมูลที่ได้จากการประเมิน ระบบเข้าใจความหมายของคำที่มีความหมายเหมือนกัน ทำให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้และผลลัพธ์ที่ได้ก็มีความครบถ้วนมากขึ้น เพราะการบรรยายภาพของแต่ละคนอาจจะใช้คำที่แตกต่างกัน แต่ในเชิงความหมายมีความหมายเหมือนกัน แต่ในกรณีของคำสืบค้น “Long beach in samui” หรือ “Wat in north of Thailand” ซึ่งเป็นคำที่มีความหมายหลากหลายรวมกันอยู่ในข้อมูลสืบค้น ผลลัพธ์ที่ได้จากการสืบค้นยังมีประสิทธิภาพที่ไม่ค่อยดีเท่าไรนัก จำเป็นต้องมีการปรับเปลี่ยนอัลกอริทึ่มในการสืบค้น เพื่อให้สามารถสืบค้นข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น

ข้อจำกัดที่สังเกตได้ของผลลัพธ์ที่ได้จากการสืบค้น ดังตัวอย่างการสืบค้นในภาพที่ 4-8 แสดงถึงภาพที่มีชื่อว่า “Koh Samet” มีประเภทของหมวดย่อยของสถานที่ท่องเที่ยวเป็นประเภทของเกาะ เมื่อเปรียบเทียบจากข้อมูลจริงในหน้าของเว็บไซต์จะพบว่ารูปภาพผลลัพธ์ดังกล่าว เป็นภาพของ หาดทรายแก้ว (Hai Sai Kaeo) เป็นสถานที่หนึ่งบนเกาะเสม็ด ข้อมูลในการบรรยายภาพควรที่บรรยายชื่อของสถานที่คือ “Hai Sai Kaeo” และมีประเภทของหมวดย่อยของสถานที่ท่องเที่ยวเป็นหาดทราย (Beach) แต่เนื่องจากข้อจำกัดในการเก็บรวบรวมข้อมูลในกรณีที่ในหน้าเว็บไซต์มีรูปภาพมากกว่า 1 ภาพ ตัว Web Crawler ไม่สามารถแยกแยะได้ว่าคำบรรยายภาพในส่วนใดเป็นของรูปภาพไหน ส่งผลให้รูปภาพทั้งหมดในหน้าเว็บไซต์เดียวกันมีคำบรรยายเหมือนกันทั้งหมด

ภาพที่ 4-8: แสดงข้อจำกัดในการเก็บรวบรวมข้อมูล

4.2.4 การเปรียบเทียบผลงานวิจัยกับงานวิจัยอื่น

งานวิจัยที่เกี่ยวข้องและนำมาเปรียบเทียบคืองานวิจัยที่มีชื่อว่า “Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques” เป็นงานวิจัยเกี่ยวกับการสืบค้นข้อมูลโดยใช้เทคนิค Case-based Learning ในการสืบค้นรูปภาพ ในระบบมีการออกแบบการจัดข้อมูลอธิบายรูปภาพให้อยู่ในรูปแบบของ RDF Triple และนำข้อมูลที่ได้ไปเก็บไว้ในฐานข้อมูลของ Case-based เพื่อเป็นข้อมูลสำหรับการสืบค้น สำหรับการสืบค้นจะทำการจัดรูปแบบของคำสืบค้นให้อยู่ในรูปแบบของ RDF Triple เช่นกันและนำข้อมูลดังกล่าวไปสืบค้นในฐานข้อมูล ในการสืบค้นมีการออกแบบขั้นตอนการสืบค้นจากโครงสร้างที่ได้ออกแบบคือทำการแบ่งคำสืบค้นออกเป็นส่วนย่อยตาม RDF Triple จะเรียกว่า Query Code Sequence (QCS) ที่ส่งเข้ามาสืบค้นและนำข้อมูลดังกล่าวไปเปรียบเทียบกับข้อมูลในฐานข้อมูลเพื่อให้ได้ผลลัพธ์ที่ตรงกับคำสืบค้นมากที่สุด โดยข้อมูลในฐานข้อมูลก็จะถูกแบ่งย่อยออกเช่นกันจะเรียกว่า Case Code Sequence (CCS)

ส่วนในงานวิจัยนี้ได้มีการนำเอาโครงสร้างแนวความคิดของ Ontology เพื่อออกแบบการนำเข้าข้อมูลจากข้อความบรรยายภาพโดยใช้ภาษา OWL (Web Ontology Language) และทำการออกแบบการขั้นตอนสืบค้นข้อมูลเพื่อรองรับการสืบค้นเชิงความหมาย เป็นการจัดกลุ่มของรูปภาพที่เกี่ยวข้องกันเพื่อให้มีประสิทธิภาพในการค้นหารูปภาพที่แม่นยำและมีประสิทธิภาพมากขึ้น และนำข้อมูลทั้งหมดจัดเก็บให้อยู่ในรูปของเอกสาร RDF

การทดสอบสืบค้นเพื่อวัดประสิทธิภาพความถูกต้องของผลลัพธ์ของทั้ง 2 งานวิจัยที่นำเปรียบเทียบกันครั้งนี้ ตั้งอยู่บนสภาพแวดล้อมที่มีความแตกต่างกัน ในด้านของความแตกต่างทางด้านจำนวนข้อมูล และตัวอย่างข้อมูลที่ใช้ในการทดสอบ

ในงานวิจัย Case-based Learning ได้มีการทดลองสืบค้นข้อมูลเพื่อวัดประสิทธิภาพความถูกต้องของผลลัพธ์ โดยในมีข้อมูลภาพในการทดลองจำนวน 49 รูป ทำการสืบค้นด้วยคำสืบค้น 30 คำที่แตกต่างกัน ผลปรากฏว่า มี 14 คำสืบค้นที่ไม่มีผลลัพธ์จากการสืบค้น ส่วนอีก 16 คำสืบค้นมีผลลัพธ์ จากการทดสอบพบว่าประสิทธิภาพในการสืบค้นในงานวิจัยโดยเฉลี่ยแล้วอยู่ที่ 0.55

เพื่อให้ได้จำนวนคำสืบค้นเท่ากันกับงานวิจัยที่ใช้เทคนิค Case-based Learning เพื่อที่จะได้เปรียบเทียบได้ตรงกัน จึงได้มีการทดสอบประสิทธิภาพของระบบใหม่อีกครั้ง โดยครั้งนี้มีการทดสอบโดยการใช้คำสืบค้น 30 คำที่แตกต่างกัน กับจำนวนรูปภาพทั้งหมด 200 ภาพ ผลปรากฏว่ามี 6 คำสืบค้นที่ไม่มีผลลัพธ์ ส่วนอีก 24 คำสืบค้นมีผลลัพธ์จากการสืบค้น จากผลการทดสอบผมว่าประสิทธิภาพในการสืบค้นในงานวิจัยโดยเฉลี่ยอยู่ที่ 0.64 จะเห็นได้ว่าประสิทธิภาพของงานวิจัยที่มีการออกแบบโครงสร้างข้อมูล รวมทั้งมีการนิยามออนโทโลยีและการระบุคำที่มีความหมายเหมือนกัน (Synonym) ช่วยให้การจัดรูปแบบข้อมูลและการสืบค้นข้อมูล มีประสิทธิภาพดีกว่าการใช้ Case-Based Learning อยู่ 9%

อีกหนึ่งงานวิจัยที่เกี่ยวข้องและนำมาเปรียบเทียบคืองานวิจัยที่ชื่อ “การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย” เป็นงานวิจัยพัฒนาต้นแบบระบบออนโทโลยีของพืชสมุนไพรไทย ในงานนี้ได้สร้างโมเดลความรู้ โดยนิยามแนวคิดเกี่ยวกับพืชสมุนไพรในรูปคลาสและพรอพเพอร์ตี้ของออนโทโลยี แล้วใช้โปรแกรมโปรทีเจ เพื่อสร้างและเก็บข้อมูลออนโทโลยีพืชสมุนไพรที่ออกแบบไว้ ซึ่งครอบคลุมพืชสมุนไพรไทยที่ใช้เป็นส่วนประกอบของยาสามัญประจำบ้านแผนโบราณ 27 ชนิด ตามประกาศกระทรวงสาธารณสุข

ในการพัฒนาต้นแบบฐานข้อมูลความรู้พืชสมุนไพรไทยที่เป็นส่วนประกอบของยาสามัญประจำบ้านแผนโบราณ จากการศึกษาค้นคว้าจากตำราแพทย์แผนโบราณทั่วไป สาขาเภสัชกรรม และสอบถามจากผู้เชี่ยวชาญด้านสมุนไพรไทย ทำให้สามารถกำหนดขอบเขต และออกแบบความสัมพันธ์ของข้อมูลโดยใช้หลักการของออนโทโลยีมาพัฒนาระบบ ในการจัดการอธิบายถึงความสัมพันธ์ระหว่างข้อมูลในระบบ และสามารถพิจารณาการอนุมานเพื่อค้นหาข้อมูลที่มีความเกี่ยวข้องกัน โดยพิจารณาได้จากความสำคัญของข้อมูลเชิงความหมาย จากการวิเคราะห์ความต้องการในการใช้งานออนโทโลยี สามารถสรุปได้ดังนี้

1. ออนโทโลยีสามารถอธิบายแนวความคิดพืชสมุนไพรไทย ยาสามัญประจำบ้านแผนโบราณ และกลุ่มอาการ

2. ออนโทโลยีสามารถสร้างความสัมพันธ์ระหว่างข้อมูลพืชสมุนไพรไทย ยาสามัญประจำบ้านแผนโบราณ และกลุ่มอาการ

ในการเปรียบเทียบจะเปรียบเทียบในเชิงของการพัฒนาต้นแบบออนโทโลยี ที่ทั้งสองงานวิจัยมีการนำเทคโนโลยี OWL (Web Ontology Language) มาใช้ในการพัฒนาต้นแบบออนโทโลยีเหมือนกัน ในลักษณะที่เหมือนกันก็คือมีการจัดกลุ่มประเภทกลุ่มของข้อมูลที่มีความเกี่ยวข้องกัน แต่ความแตกต่างก็คืองานวิจัยที่นำมาเปรียบเทียบมีการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูล ทำให้ผลลัพธ์ในการสืบค้นสามารถค้นหาข้อมูลที่มีความเกี่ยวข้องกันได้ แต่ในงานวิจัยนี้ไม่ได้มีการนำหลักการเว็บเชิงความหมาย ในการเชื่อมโยงความสัมพันธ์ของข้อมูลมาใช้

บทที่ 5
สรุปผลการวิจัย

5.1 สิ่งที่ค้นพบในการศึกษาวิจัยและการอภิปราย

จากงานวิจัยชิ้นนี้ ทำให้ได้เข้าใจเกี่ยวกับแนวความคิดเว็บเชิงความหมาย (Semantic Web) ถึงแม้ว่าแนวความคิดนี้ยังอยู่ในขั้นตอนของการศึกษาและวิจัย แต่จากงานวิจัยทั้งของประเทศไทยและของต่างประเทศ ที่ได้มีการนำเอาแนวความคิดนี้ไปประยุกต์ใช้ในงานด้านต่างๆ ทำให้สามารถเข้าใจถึงประโยชน์ของแนวความคิดนี้ได้ชัดเจนมากยิ่งขึ้น

นอกจากนี้ยังได้ทำให้เข้าใจถึงกระบวนการในการสร้างระบบสืบค้นข้อมูล (Search Engine) ตั้งแต่การเก็บรวบรวมข้อมูลโดยใช้ซอฟต์แวร์หุ่นยนต์ ขั้นตอนในการเตรียมฐานความรู้ให้อยู่ในรูปแบบที่เหมาะสม เพื่อรองรับการสืบค้นข้อมูล และอัลกอริทึ่มที่ใช้ในการสืบค้นข้อมูล

จากการศึกษาจากแหล่งข้อมูลหลายแหล่งซึ่งให้รายละเอียดของเรื่องเว็บเชิงความหมาย ทฤษฎีนี้จะเข้ามาช่วยยกระดับให้กับเทคโนโลยีของเว็บขึ้นไปอีกระดับหนึ่ง ทำให้เครื่องสามารถเข้าใจถึงความหมายของตัวข้อมูล และสามารถเชื่อมโยงข้อมูลที่มีความเกี่ยวข้องกันได้อย่างมีประสิทธิภา หรือจะกล่าวก็คือเป็นเทคโนโลยีที่ทำให้คอมพิวเตอร์มีความฉลาดมากขึ้น ในแง่ของการสืบค้นข้อมูลเว็บเชิงความหมายจะให้ผลลัพธ์ของการสืบค้นที่ตรงกับความต้องการของผู้ใช้ได้ดีกว่าการสืบค้นด้วยวิธีของ Keyword-Based

แต่อย่างไรก็ตามแนวความคิดเว็บเชิงความหมายยังอยู่ระหว่างการศึกษาและวิจัย และได้มีการพัฒนาเครื่องมือที่มาสนับสนุนแนวความคิดนี้ เช่น ภาษา OWL และ DAML + OIL อนาคตคงจะเห็นเครื่องมือที่ออกมาสนับสนุนการพัฒนาระบบที่ใช้แนวความคิดเว็บเชิงความหมายมากขึ้น

5.2 สิ่งที่ได้รับจากการศึกษาวิจัย

ในงานวิจัยชิ้นนี้ได้สร้างตัวแบบออนโทโลยีสำหรับการบรรยายสถานที่ท่องเที่ยวในประเทศไทย ตัวแบบที่ได้มีการออกแบบมานี้สามารถที่จะนำมาไปพัฒนาให้สามารถบรรยายได้ครอบคลุมสถานที่ท่องเที่ยวทั่วโลก รวมไปถึงการเชื่อมโยงข้อมูลที่เกี่ยวข้องกับรูปภาพที่ทำการ

สืบค้นออกมา ยกตัวอย่างเช่น ลิงค์ของโรงแรมที่พัก, รายละเอียดการเดินทาง และความคิดเห็นเกี่ยวกับสถานที่ท่องเที่ยวนั้นๆ เป็นต้น

งานวิจัยที่นำเอาแนวความคิดเว็บเชิงความหมายมาใช้กับระบบสืบค้นข้อมูล ได้มีการออกแบบการสืบค้นให้มีการใช้คำที่มีความหมายเหมือนกัน (Synonym) ทำให้คอมพิวเตอร์สามารถเข้าใจถึงกลุ่มคำที่มีความหมายเหมือนกัน มาใช้ในการสืบค้นดังนั้นทำให้ผลลัพธ์ที่ได้เป็นผลลัพธ์เชิงความหมายซึ่งมีลักษณะดีกว่าการสืบค้นด้วยคำสำคัญ (Keyword-Based) เนื่องจากในการเขียนคำบรรยายในเนื้อหาบทเว็บไซต์ ผู้เขียนแต่ละคนอาจมีการใช้คำที่แตกต่างกัน แต่มีความหมายเหมือนกัน ในกรณีระบบสืบค้นที่ใช้หลักการสืบค้นด้วยคำสำคัญจะดึงเฉพาะเอกสารที่มีคำที่ตรงกับคำสืบค้นเท่านั้น แต่ในงานวิจัยนี้นอกจากจะดึงเอกสารที่ตรงกับสืบค้นได้แล้ว ยังสามารถที่จะดึงเอกสารที่มีคำที่มีความหมายเหมือนกับคำสืบค้นขึ้นมาแสดงผลงด้วย ทำให้ได้ผลลัพธ์ที่ถูกต้องและครบถ้วนมากกว่า

5.3 ข้อจำกัดและอุปสรรค์ในการศึกษาวิจัย

1. ในขั้นตอนการออกแบบออนโทโลยีนั้นค่อนข้างยาก เนื่องจากการออกแบบออนโทโลยีสามารถออกแบบได้ในหลาย ๆ มุมมองขึ้นกับวัตถุประสงค์หลัก เพราะฉะนั้นก่อนที่จะลงมือออกแบบออนโทโลยี ต้องทำการศึกษาวิเคราะห์ความต้องการให้ชัดเจน จึงจะทำให้การนำเอาออนโทโลยีมาประยุกต์ใช้เกิดประโยชน์และมีประสิทธิภาพมากที่สุด

2. เนื่องจากปัจจุบันแนวความคิดเว็บเชิงความหมาย (Semantic Web) ยังอยู่ระหว่างการศึกษาและวิจัย ทำให้ยังขาดเอกสารที่เกี่ยวข้องกับการพัฒนาระบบโดยการนำเอาแนวความคิดเว็บเชิงความหมายมาประยุกต์ใช้

5.4 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต

1. ประสิทธิภาพของซอฟต์แวร์หุ่นยนต์ (Spider Robot) และขั้นตอนวิธีในการเตรียมฐานความรู้ การพัฒนาซอฟต์แวร์หุ่นยนต์ที่มีประสิทธิภาพ ทำให้การรวบรวมข้อมูลจากเว็บได้รวดเร็ว และสามารถเก็บรายละเอียดได้มากขึ้น มีผลโดยตรงต่อประสิทธิภาพของผลการสืบค้นข้อมูลที่จะได้ข้อมูลที่ถูกต้องและแม่นยำมากขึ้น

2. งานวิจัยนี้ขอบเขตของงานคือการค้นหาได้เฉพาะรูปภาพที่มีคำอธิบายข้อมูลด้วยภาษาอังกฤษ งานวิจัยที่สามารถทำต่อไปได้คือการดำเนินการบนภาษาไทย โดยจะต้องพัฒนาเวิร์ดเนตของภาษาไทยเพื่อให้สามารถแบ่งกลุ่มคำในภาษาไทย เพื่อให้สามารถจำแนกคำที่มีความหมายเหมือนกัน ตรงข้ามกัน รวมถึงกลุ่มคำกริยา และกลุ่มคำอื่นๆ เช่นเดียวกับเวิร์ดเนตของภาษาอังกฤษ

3. งานวิจัยที่สามารถทำต่อไปได้ในอนาคตคือการออกแบบโครงสร้างข้อมูลเพิ่มเติมให้สามารถอธิบายรายละเอียดของรูปภาพในคุณสมบัติต่างๆ ยกตัวอย่างเช่น ชนิดของรูปภาพ (JPG, GIF และ PNG), ขนาดของรูปภาพ, ความสัมพันธ์กับรูปภาพอื่น เป็นต้น

5.5 สรุปผลการวิจัย

จากการดำเนินงานวิจัย ซึ่งประกอบด้วย 2 ส่วนหลักๆ คือการเตรียมข้อมูลสำหรับการสืบค้น และวิธีการสืบค้นข้อมูลรูปภาพ โดยงานการเตรียมฐานความรู้สำหรับการสืบค้น เป็นการรวบรวมเอกสารบนเว็บ (Web Document) จากเว็บ (Web—World Wide Web) โดยในงานวิจัยนี้ได้กำหนดขอบเขตของเว็บไว้เฉพาะโดมเมนของการท่องเที่ยวแห่งประเทศไทย ในรูปแบบภาษาอังกฤษ (http://www.tourismthailand.org/) เพื่อทำการรวบรวมข้อมูลรูปภาพและคำบรรยายภาพ มีการใช้ Web Crawler ทำหน้าที่ในการเก็บรวบรวมข้อมูล และสามารถเก็บข้อมูลมาได้ทั้งหมด 14,599 เอกสาร ก่อนที่จะนำข้อมูลที่เก็บรวบรวมได้ไปทำความสะอาดข้อมูลให้เหลือเฉพาะข้อมูลที่ต้องการเท่านั้น

การเตรียมฐานความรู้ เป็นกระบวนการอธิบายรายละเอียดของรูปภาพตามโครงสร้างที่ได้ทำการออกแบบไว้ เพื่อให้อยู่ในรูปแบบที่เหมาะสมกับการสืบค้น งานวิจัยครั้งนี้ได้ทำการรวบรวมรูปภาพบนเว็บมาทั้งหมดจำนวน 1,641 ภาพ

การสืบค้นข้อมูล แบ่งออกเป็น 2 ขั้นตอนใหญ่ๆ คือ ขั้นตอนการเตรียมข้อมูลสืบค้น (Query) และขั้นตอนการสืบค้น ในการเตรียมข้อมูลสืบค้น เป็นขั้นตอนการแบ่งกลุ่มคำหลักที่มีอยู่ในข้อมูลสืบค้นออกเป็นกลุ่มๆ

ขั้นตอนการสืบค้น เป็นการนำข้อมูลสืบค้น ไปเปรียบเทียบระดับความเหมือนระหว่างข้อมูลสืบค้นและข้อมูลของคำบรรยายรูปภาพที่เตรียมไว้แล้วในฐานความรู้จากการทดลองเพื่อวัดประสิทธิภาพของการสืบค้นโดยเปรียบเทียบกับวิธีการโดยใช้เครื่องมือคือ ความแม่นยำ (Precision)

และการเรียกซ้ำ (Recall) ในการทดสอบมีการทดสอบจำนวน 2 ครั้ง โดยการซุ่มข้อมูลรูปภาพจำนวน 200 ภาพ ทำการทดสอบกับคำสืบค้นจำนวน 20 คำสืบค้นที่มีความแตกต่างกัน ผลการทดลองสามารถสรุปได้ว่า ครั้งที่ดีที่สุด มีระดับความแม่นยำ (Precision) เฉลี่ย 6.6 ในขณะที่ระดับการเรียกคืน (Recall) มีประสิทธิภาพเฉลี่ย 7.1

ถึงแม้ว่าจะได้ผลลัพธ์จากการทดสอบประสิทธิภาพของระบบจะที่อยู่ในระดับที่ระดับปานกลาง แต่ก็ยังมีข้อจำกัดในกรณีที่ในหน้าเว็บไซด์นั้นประกอบด้วยภาพจำนวนหลายภาพ ในงานวิจัยนี้ยังไม่สามารถแยกได้ว่าภาพแต่ละภาพมีความหมายเหมือนกันหรือแตกต่างกันอย่างไร

BANGKOK UNIVERSITY FOUNDED 1962

บรรณานุกรม

นฤมล ปราชญโยธิน, ทวีศักดิ์ กออนันตกูล, เปรมิน จินดาวิมลเลิศ. (2536). ธิซอรัสกับระบบสารสนเทศ. กรุงเทพฯ: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.

นพคุณ บุญสิน. (2547). การจัดเก็บและสืบค้นรูปภาพเชิงความหมายบนฐานของ RDF. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น.

สิริรัตน์ ประกฤติกรชัย. (2550). การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย. สารนิพนธ์ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า พระนครเหนือ.

Christiane F. (1998). Wordnet: An Electronic Lexical Database. MA: The MIT Press.

Cyril, G. & Eric G. (2005). A probabilistic interpretation of precision, Recall and F-score, with Implication for Evaluation. Proceeding of the European Colloquium on IR Research (ECIR'05), LLNCS 3408 (Springer), 345-359.

Feng Jing, Changhu Wang, Yuhuan Yao, Kefeng Deng, Lei Zhang, & Wei-Ying Ma. (2006). A web image search engine clustering of search result. Proceedings of the 14th annual ACM international conference on Multimedia, 377-384.

Fensel, D. (2003). Ontologies: Silverbullet for Knowledge Management and Electronic Commerce. (2rd ed.). New York:Springer.

Gruber, T. R. (1993). A translation approach to portable ontology specification. Knowledge Acquisition, 5(2), 199-220.

Guarino, G. (1998). Formal Ontology in Information Systems. Guarino N. 1st International Conference on Formal Ontology in Information Systems FOIS'98

Hyvönen, E., Styrman, A., & Saarela, S. (2002). Ontology-based image retrieval. Proceedings of XML Finland 2002 Conference, 15-27.

Internet World stats (Copyright 2008). "Internet Usage and World Population Statistics." สืบค้นวันที่ 29 ตุลาคม 2551 จาก http://www.internetworldstats.com/stats.htm.

Jena Framework (Copyright 2008). "Jena Framework." สืบค้นวันที่ 20 พฤศจิกายน 2551 จาก http://en.wikipedia.org/wiki/Jena_(framework).

Lassila, O., & Swick, R. (Copyright 1999). "Resource Description Framework (RDF) Model and Syntax Specification W3C." สืบค้นเมื่อ 25 สิงหาคม 2551 จาก http://www.w3.org/TR/RECrdf-syntax/.

Lee, W. P. and Tsai, T. C. (2003). An interactive agent-based system for concept-based web search. Journal of expert systems with application, 24, 365-373.

Morris, J., & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computation Linguistics, 17, 21-48.

Protégé Project (Copyright 2007). "What is protégé?." สืบค้นวันที่ 19 พฤศจิกายน 2551 จาก http://protege.stanford.edu/index.html/.

Ricardo, Y. (1999). Modern information retrieval. London: Addison Wesley.

Richard J. (Copyright 2000). "Measuring Information Retrieval. London: Addison Wesley." สืบค้นวันที่ 31 สิงหาคม 2551 จาก http://www.hsl.creighton.edu/HSL/searching/Recall-Precision.html.

Seaborne, A. (Copyright 2004). "A programmer's introduction to RDQL." สืบค้นวันที่ 25 สิงหาคม 2551 จาก http://www.hpl.hp.com/semweb/doc/tutorial/RDQL/.

Swartz, A. (Copyright 2002). "The Semantic Web in Breadth." สืบค้นเมื่อ 25 สิงหาคม 2551 จาก http://logicerror.com/semanticWeb-long.

Tim Berners-Lee (Copyright 2000). "Semantic Web on XML." สืบค้นเมื่อ 20 สิงหาคม 2551 จาก http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html.

Von-Wun, S. Chen-Yu, L. Chung-Cheng, L. Shu Lei, C. & Ching-chih, C. (2003). Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-based Learning Techniques. Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries, 61-72.

World Wide Web Consortium (Copyright 2002). "Frequently Asked Questions About RDF." สืบค้นเมื่อ 25 สิงหาคม 2551 จาก http://www.w3.org/RDF/.

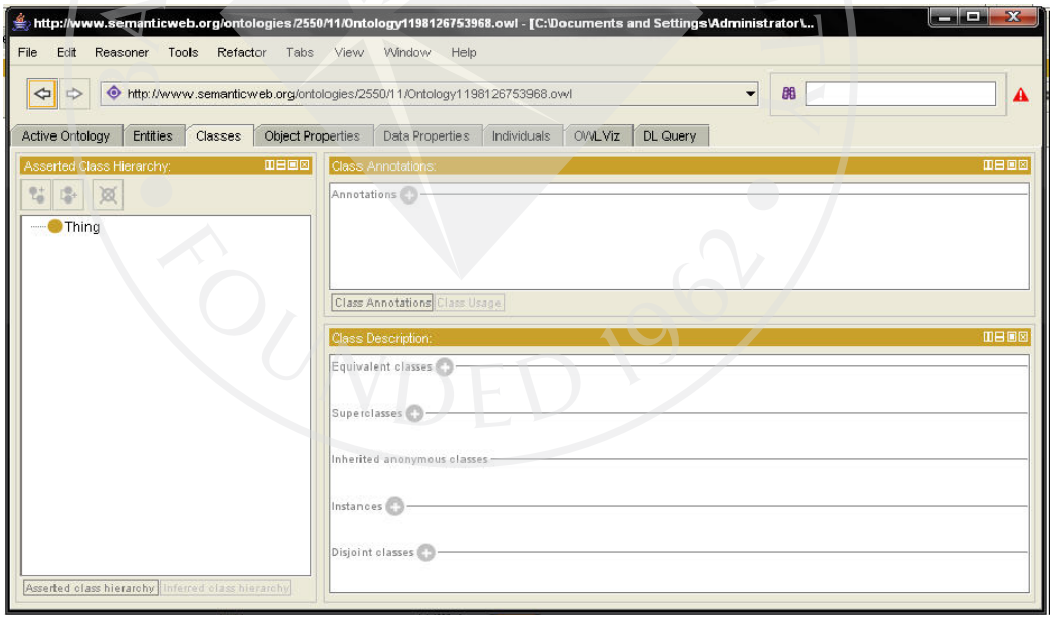
ภาคผนวก

การสร้างออนโทโลยีโดยใช้โปรแกรม Protégé

1. โปรแกรม Protégé

Protégé เป็นซอฟต์แวร์ประเภทโอเพ่นซอร์ส ใช้งานสำหรับการสร้างออนโทโลยีและฐานความรู้ รองรับการสร้างออนโทโลยีด้วยภาษา RDF/RDFS และ OWL พัฒนา ขึ้นโดยมหาวิทยาลัยสแตนฟอร์ด ประเทศสหรัฐอเมริกา มีส่วนการติดต่อ ผู้ใช้งานเป็นแบบกราฟิก (Graphical User Interface : GUI)

ภาพที่ ก-1: หน้าจอโปรแกรม Protégé



2. การสร้าง OWL ชื่อ Tourism.owl

ออนโทโลยีสถานที่ท่องเที่ยวในประเทศที่ได้มีการออกแบบไว้ในบทที่ 3 จะนำการสร้าง Ontology ที่เป็นเอกสาร OWL โดยมีรายละเอียดของคลาสดังต่อไปนี้

Class ประกอบด้วย

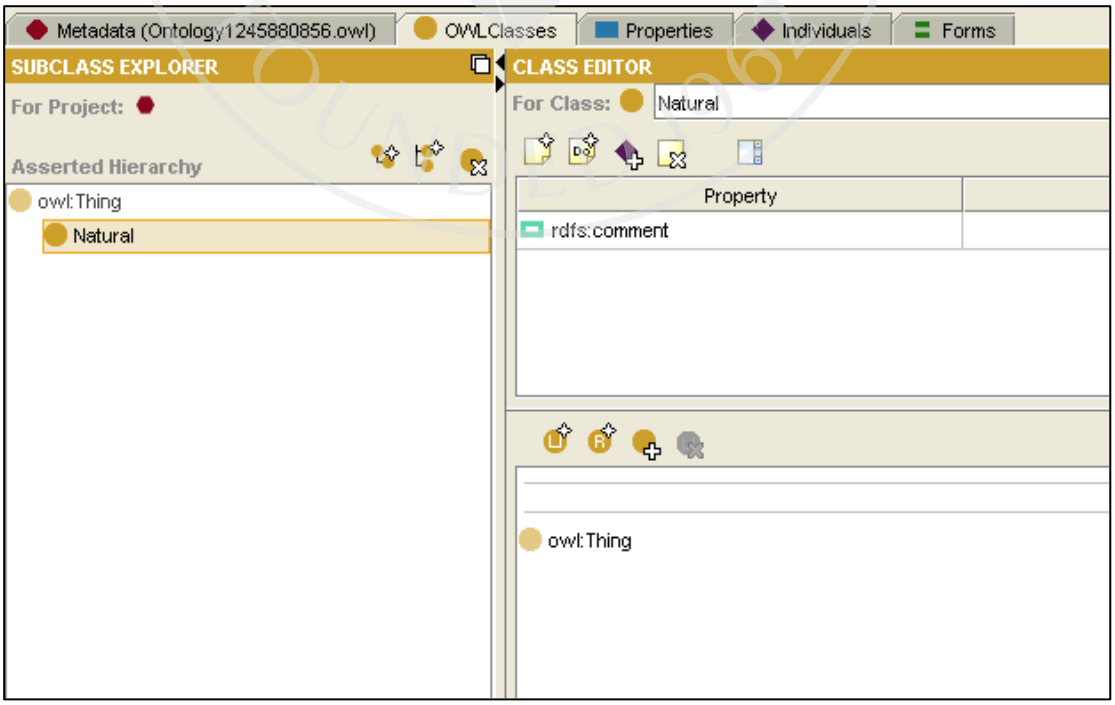
- Natural Place
    - Mountain
    - Waterfall
    - Cave
    - Sea
    - Beach
    - Bay
    - Island
    - Dam
    - Flower Field
    - Hot Spring
    - National Park
- Cultural Place
    - Temple
    - Art Gallery
    - Cultural Museum
- Heritage Place
    - Archaeological Site
    - Monument
    - Royal Palace
    - Historical Museum
- Education Place
    - Royal Project
    - Botanical Garden
    - Science Museum
- Traditional Path

- Village
- Market
- Floating Market

- Recreational Place
  - Theater
  - Zoo
  - Animal Farm
  - Amusement Park

เมื่อเปิด OWL Project ขึ้นมา ที่แท็บคลาส จะปรากฏคลาสเริ่มต้นที่ชื่อ owl:Thing ที่ Hierarchy Tree view และเมื่อมีการเพิ่มคลาสเข้าไป ทุกคลาสจะเป็นซับคลาส ของ owl:Thing ดังภาพที่ 3 มีการเพิ่ม Class Natural เข้าไป โดยการกดปุ่ม สร้างซับคลาส และใส่ชื่อเข้าไป

ภาพที่ ก-2: การสร้างคลาสและซับคลาส



หลังจากนั้นสร้างคลาสและซับคลาส ตามที่ได้มีการออกแบบเอาไว้ ผลลัพธ์ที่ได้แสดงดังภาพที่ ก-3

ภาพที่ ก-3: ลำดับชั้นของคลาสที่สร้างตามออนโทโลยีที่ออกแบบไว้