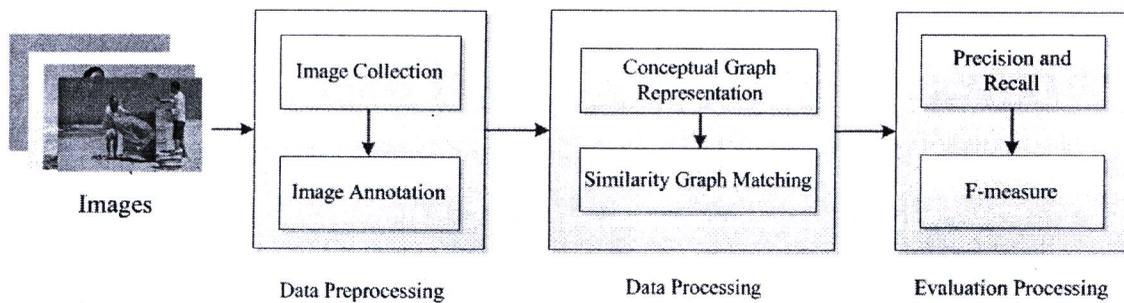


บทที่ 3

ขั้นตอนวิธีการดำเนินงานวิจัย

ในการศึกษางานวิจัยในครั้งนี้ได้มีการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยออกเป็น 3 ส่วนหลักดังนี้ ขั้นตอนการเตรียมข้อมูล (data preprocessing) ขั้นตอนการประมวลผล (data processing) และ ขั้นตอนการวัดประสิทธิภาพ (evaluation) โดยขั้นตอนทั้งหมดถูกแสดงไว้ในภาพที่ 3.1



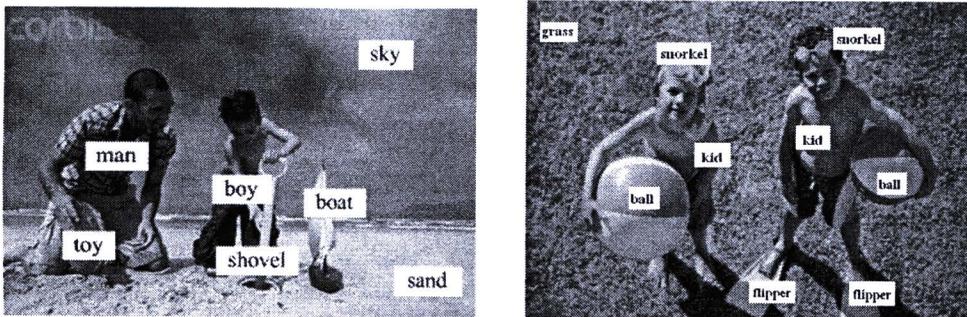
ภาพที่ 3.1 ขั้นตอนการจำแนกกลุ่มความหมายภาพ

3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการเตรียมข้อมูล (data preprocessing) โดยการทำการแยกและคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ สำหรับภาพบางภาพจะไม่นำเข้ามาทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวม ภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะถูกคัดเลือกภาพนั้นออกไป และทำการรวบรวมข้อมูลภาพที่ต้องการที่มาจากหลายฐานข้อมูลจุดประสงค์ก็เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ดังนั้นกระบวนการทั้งหมดนี้จะประกอบด้วย 2 กระบวนการดังนี้

3.1.1 ฐานข้อมูลรูปภาพ

แหล่งข้อมูลภาพมีหลายแหล่งข้อมูลที่ได้รับการยอมรับและสามารถนำมาใช้เป็นฐานข้อมูลรูปภาพได้ เช่น Fotosearch stock² The Corbis Stock³ The Corel Corporation⁴ เป็นต้น เป็นแหล่งข้อมูลภาพที่หลากหลาย อาจจะมีภาพที่ไม่เหมาะสมกับการทดลองที่นำเสนอ เนื่องจาก ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดเจนคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บังชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ภาพบางภาพอาจจะไม่สามารถแปลความหมาย หรือภาพมีความหมายกำกวมจนทำให้ไม่สามารถหาความหมายภาพได้ ทำให้ต้องมีการคัดเลือกภาพออกไป ไม่นำมาใช้ในการทดลอง ดังนั้นในการหาแหล่งข้อมูลของการนำภาพเข้ามาใช้จึงจำเป็นต้องสมบูรณ์ที่สุด ภาพจะต้องความเหมาะสมกับงานที่จะนำมาใช้เพื่อตอบสนองกับความต้องการของการทดลองมากที่สุด



ภาพที่ 3.2 ภาพตัวอย่างที่ถูกแท็กด้วยคำหลัก⁵

สำหรับการทดลองโดยทั่วไปในสาขาคอมพิวเตอร์วิชัน (computer vision) ในส่วนของ การประมวลผลภาพระดับสูง รูปภาพที่นำเข้ามาทดลองจะเป็นภาพที่วัตถุถูกแท็ก หรือกระทำการ ให้ความหมายมาล่วงหน้าก่อน จะเรียกภาพจำพวกนี้ว่า annotated images [Carbonetto P., 2004][Jia Li, 2003][Winn J., 2005] ดังแสดงในภาพที่ 3.2 แสดงรูปภาพถูกแท็กด้วยคำหลัก ดังนั้นในการทดลองจะต้องมีฐานข้อมูลภาพที่สมบูรณ์เพียงพอที่จะสามารถนำส่วนของคำหลักที่ถูกแท็กมาใช้งานได้โดยไม่มีผลข้างเคียงต่อกระบวนการที่นำเสนอ ส่วนใหญ่แหล่งข้อมูลภาพจะทำการคัดเลือกบริเวณ (region) ที่เหมาะสมสำหรับการแท็กเป็นคำหลักหรือคำสำคัญ (keyword)

² Fotosearch Stock: <http://www.fotosearch.com>

³ The corbis Stock: <http://pro.corbis.com>

⁴ The Corel Corporation: <http://www.corel.com/>

⁵ รูปภาพ อ้างอิง <http://www.corbisimages.com/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

เพื่อใช้สำหรับการสืบค้นข้อมูล ดังนั้นบางแหล่งข้อมูลจะใช้การแบ่งหมวดหมู่วัตถุ (object categories) เข้ามาช่วยในการจัดกลุ่มของคำที่จะนำมาทดแทนส่วน บริเวณนั้น ๆ [Everingham M., 2006] [Winn J., 2005] ซึ่งคำที่จะใช้จะมาจากพจนานุกรม (dictionary) จากตารางที่ 3.1 [A. Hanbury, 2007] แสดงตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลัก ในการทดลองสำหรับการประมวลผลแบบการให้ความหมายภาพ (annotation)

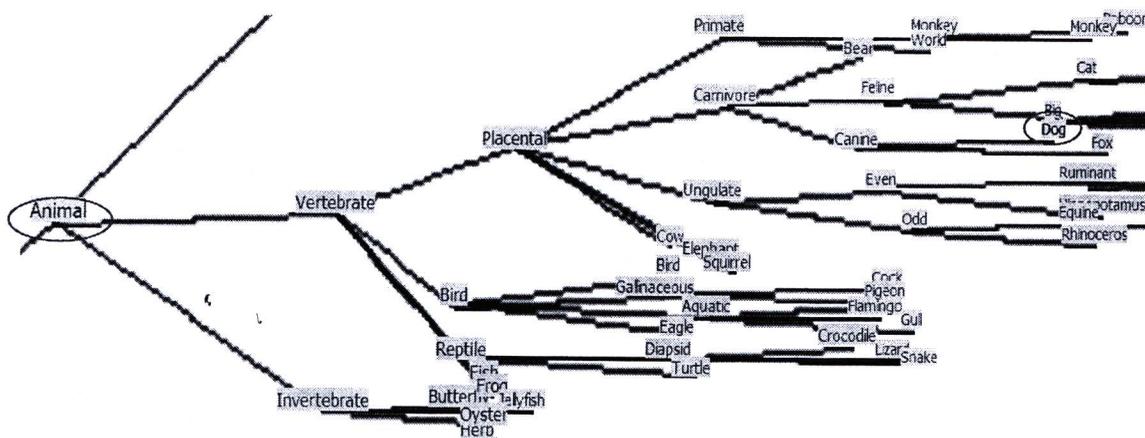
แหล่งข้อมูลภาพ	จำนวนคำหลัก
EU LAVA Project [Peronnin F., 2006]	10
Chen and Wang [Chen Y., 2004]	20
Microsoft Research Cambridge Database [Winn J., 2005]	35
Carbonetto et al., [Carbonetto P., 2004]	55
PASCAL VOC Challenge 2005 Database [Everingham M., 2006]	101
Fei Fei et al., [L. Fei-Fei, 2004]	101
Li and Wang [Jia Li, 2003]	433
Barnard et al., [Barnard K., 2003]	323
University of Washington Ground Truth Image Database ⁶	392

ตารางที่ 3.1 ตัวอย่างงานวิจัยที่ใช้ฐานข้อมูลคำหลักในการให้ความหมายภาพ

WordNet [Miller G.A., 1990] [Zinger S., 2005] เป็นงานวิจัยที่ได้รับความนิยมมากมีนักวิจัยหลายกลุ่มที่อ้างอิงการใช้ฐานข้อมูลของคำหลักที่แท็กบนภาพ [Adrian Popescu 2008][Javier Álvez, 2008] โดย WordNet จะใช้คำนาม (noun) ถูกเลือกจากพจนานุกรม (Dictionary) เป็นคำหลักและนำคำหลักเหล่านั้นมาจัดกลุ่ม (class) ตามลำดับชั้น (hierarchy) ถึง 73,733 กลุ่มซึ่งจะมีคำนามที่ถูกจัดกลุ่มอยู่ภายในกิ่งก้าน (leave) ของลำดับชั้นมากมายถึง 60,000 กิ่ง และมีคำนามที่ใช้ถึง 116,364 คำ ดังแสดงโครงสร้างคำหลักของ WordNet ในภาพที่ 3.3 แสดงกิ่งก้านลำดับชั้นของคำว่า “dog” จะอยู่ในกิ่งลำดับชั้นหรือรากฐานมาจากคำหลัก “animal” และจากภาพที่ 3.4 แสดงถึงความสัมพันธ์ของคำหลักกับโครงสร้างใน WordNet ของคำว่า “dog”

⁶ <http://www.cs.washington.edu/research/imagetdatabase/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

เพื่อให้ส่วนอื่นไปเรียกใช้ โดย WordNet จะถูกใช้คู่กับโปรแกรมเพื่อทำการสืบค้นข้อมูล และยังสามารถกำหนดความสัมพันธ์ของคำด้วยโครงสร้างของ WordNet เองและสามารถสร้างความหมายเหมือนกัน (synonymy) ได้เช่นกัน เป็นข้อดีของการนำ WordNet เข้ามาใช้ตั้งจะกล่าวในหัวข้อถัดไป



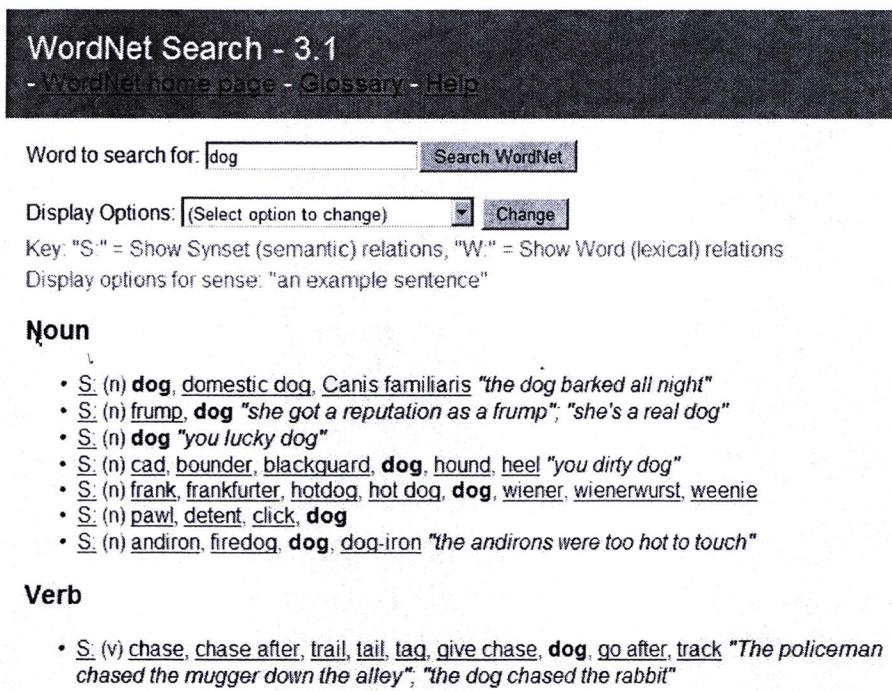
ภาพที่ 3.3 ส่วนย่อยของโครงสร้างลำดับชั้นของคำหลัก“dog” บน WordNet⁷

3.1.2 การแท็กคำหลักบนภาพ

สิ่งที่สำคัญถัดจากการเลือกใช้ฐานข้อมูลสำหรับกระบวนการของการประมวลผลภาพระดับสูง (high-level image processing) คือ การให้คำอธิบายภาพ (annotated images) โดยใช้ฐานข้อมูลคำหลักที่คัดเลือกมา การแท็กคำหลักบนภาพ โดยทั่วไป ภาพ (image) จะประกอบด้วย พื้นหลัง (background) และ พื้นหน้า (foreground) กล่าวคือภาพหนึ่งภาพจะประกอบด้วยวัตถุ (objects) หลายวัตถุ ทำให้ทุกวัตถุควรจะถูกแท็กด้วยคำหลักที่เหมาะสมจากฐานข้อมูล ดังนั้นในการแบ่งแยกวัตถุ (segmentation) เป็นหัวข้อที่มีการวิจัยอย่างต่อเนื่องในการประมวลผลภาพระดับต่ำ (low-level image processing) [Qian Huang, 1995][Vailaya A., 2001] ด้วยคุณลักษณะของภาพหลายคุณลักษณะ (features) เพื่อทำการแบ่งแยกให้ได้วัตถุที่สมบูรณ์แบบ แต่อย่างไรก็ตามการแบ่งแยกวัตถุที่มีรูปร่างลักษณะคล้ายคลึงกันหรือแตกต่างกันแต่มีความหมายเดียวกัน ยังคงเป็นเรื่องที่ค่อนข้างยากสำหรับการประมวลผลภาพระดับต่ำรวมถึงกระบวนการรู้จำรูปแบบวัตถุ (pattern recognition) เพื่อบอกความหมายของวัตถุหรือชื่อของวัตถุ แต่อย่างไรก็ตามในการวิเคราะห์ความหมายของภาพในงานวิจัยนี้ได้เฉพาะเจาะจงในส่วน

⁷ <http://people.csail.mit.edu/torralba/research/LabelMe/wordnet/test.html> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

การประมวลผลระดับสูง จึงได้ข้ามในส่วนของการแบ่งแยกวัตถุ และการรู้จำรูปแบบวัตถุ ดังนั้นในส่วนที่กล่าวต่อไปสำหรับการประมวลผลภาพระดับสูงคือการให้คำอธิบายภาพด้วยการแท็ก (tag) [Ismail Haritaoglu, 1998] [Tele Tan, 2002] [Vasileios Mezaris, 2003] [R. Zhao, 2002] ข้อมูลภาพให้อยู่ในรูปของชื่อวัตถุ หรือคำหลักจากฐานข้อมูล



ภาพที่ 3.4 ความสัมพันธ์ของโครงสร้างคำหลักบน WordNet⁸

มีหลายกลุ่มงานวิจัยที่คิดค้นกระบวนการให้ความหมาย หรือแท็กวัตถุบนภาพด้วยวิธีการแตกต่างกัน จากกลุ่มเครื่องมือที่ให้ความหมายภาพที่นิยม [Jeroen Steggink, 2011] สามารถจำแนกวิธีการให้ความหมายตามเครื่องมือได้ 3 รูปแบบ ดังนี้

1. แบบ desktop PC [Yao, B.,2007][Petridis, K.,2006][Hollink, L.,2004] เป็นรูปของการให้ความหมายบนโปรแกรมจะสามารถทำงานได้ด้วยการติดตั้งลงบนเครื่องคอมพิวเตอร์ส่วนตัว (personal computer) ข้อดีของโปรแกรมที่ถูกติดตั้งลงบนเครื่องคอมพิวเตอร์ส่วนตัวคือสามารถเพิ่มขีดความสามารถของการประมวลผลในส่วนขอโปรแกรมได้มากกว่าวิธีอื่น ๆ ดังนั้นการโปรแกรมจะมีความสามารถได้มากกว่าวิธีอื่น ๆ มีความยืดหยุ่นกว่า แต่อย่างไรก็ตามข้อจำกัดของวิธีการนี้ คือ เนื่องจากการให้ความหมายบนวัตถุนั้นสามารถทำได้ด้วยตนเองทำให้ ฐานข้อมูลภาพถูกจำกัด ขอบเขตของคำหลักถูกจำกัด หรืออาจไม่ครอบคลุมมาก

⁸ WordNet: <http://wordnetweb.princeton.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

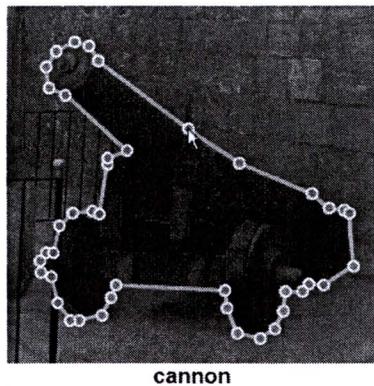
พอการให้ความหมายภาพอาจจะถูกให้ความหมายด้วยตนเองหรือเพียงกลุ่มบุคคลใดเพียงกลุ่มเดียว ทำให้การทดลองไม่ครอบคลุมเท่าที่ควร ตัวอย่างโปรแกรม เช่น Image Parsing [Yao, B., 2007], M-OntoMat-Annotizer [Petridis, K.,2006], Photostuff [Halaschek-Wiener,2005], Spatial Annotation [Hollink, L.,2004] ดังแสดงในตารางที่ 3.2

2. แบบออนไลน์ (online) [Russell, B.C.,2008][Volkmer, T.,2005] [flickr] รูปแบบนี้จะสามารถให้ความหมายวัตถุบนภาพได้ทางออนไลน์ผ่านทางเว็บไซต์ เป็นข้อดีสำหรับรูปแบบนี้ เพราะทุกคนสามารถเข้าถึงได้อย่างง่าย ข้อมูลภาพและกลุ่มข้อมูลคำหลักจะถูกจัดเก็บลงบนเซิร์ฟเวอร์ เดียวกัน ฐานข้อมูลภาพหลากหลายและกลุ่มคนที่ให้ความหมายมีหลายกลุ่ม แต่อย่างไรก็ตามการที่มีฐานข้อมูล และกลุ่มคำศัพท์ที่เปิดกว้างมากเกินไปนี้อาจจะเป็นข้อเสียสำหรับการทดลองที่มีขีดจำกัดเช่นเดียวกัน ส่วนกลุ่มบุคคลที่เข้ามา จะไม่สามารถควบคุมได้ การให้ความหมายภาพให้เป็นไปตามที่กำหนดของส่วนโปรแกรมที่ทำการทดลอง ตัวอย่างโปรแกรม เช่น Flickr [flickr] LabelMe [Russell, B.C.,2008] IBM EVA [Volkmer, T.,2005] ดังแสดงในตารางที่ 3.1

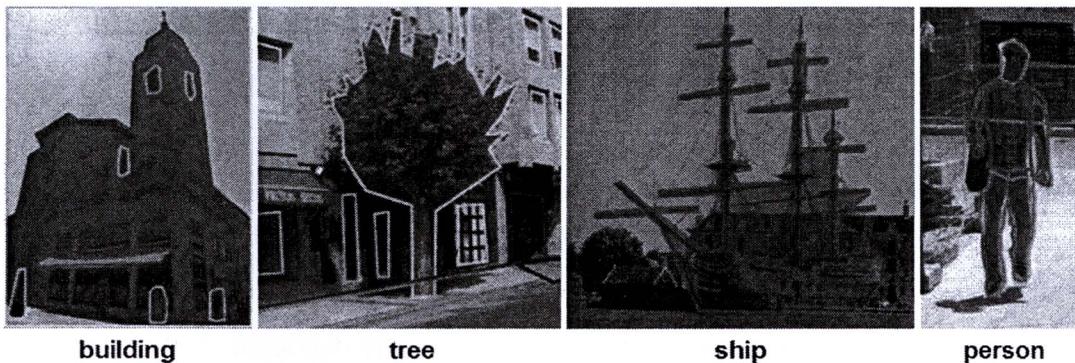
รูปแบบเครื่องมือ	เครื่องมือคำอธิบายภาพ	รูปแบบการเลือกวัตถุ			
		global	bounding box	polygon	free hand
Desktop PC	Image Parsing	✓		✓	
	M-OntoMat-Annotizer		✓		✓
	Photostuff		✓		
	Spatial Annotation	✓			
Online	Flickr	✓	✓		
	IBM EVA	✓			
	LabelMe			✓	✓
Online game	ESP game		✓		
	Peekaboom				✓
	Squigl				✓

ตารางที่ 3.2 แหล่งฐานข้อมูลของคำหลัก

จากเครื่องมือให้ความหมายภาพ (image annotation tool) ในตารางที่ 3.2 ที่แบ่งเป็น 3 รูปแบบ แต่ละรูปแบบมีความสามารถทำงานได้แตกต่างกันมีทั้งข้อดีและข้อจำกัด ดังนั้นในงานวิจัยนี้ได้เลือกเครื่องมือแบบออนไลน์ โดยใช้โปรแกรม LabelMe⁹ [B. C. Russell,2008][A. Torralba,2010] เป็นเครื่องมือที่ได้รับการยอมรับอย่างกว้างขวาง สำหรับงานวิจัยทางด้าน Computer Vision โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย (Web-based annotation tools) เริ่มตั้งแต่ปี 2005 ปัจจุบันมีวัตถุภาพที่ถูกให้ความหมายรวมทั้งสิ้น 400,000 วัตถุ [Von Ahn and L. Dabbish, 2004.] [B. C. Russell, 2008] [A. Sorokin and D. Forsyth, 2008] [M. Spain and P. Perona, 2007] [D. G. Stork, 1999] ผู้ใช้สามารถเข้าถึงโปรแกรมผ่านทางเครือข่ายออนไลน์ได้ สามารถแท็กวัตถุผ่านทางออนไลน์ได้ดังแสดงในภาพที่ 3.5 แสดงหน้าเว็บไซต์ของโปรแกรม LabelMe โปรแกรมสามารถทำงานร่วมกันได้หลายแพลตฟอร์ม ซึ่งเป็นวัตถุประสงค์หลักของคณะผู้จัดทำโปรแกรมนี้ ทำให้ผู้ใช้งานที่เข้ามาให้ความหมายภาพมาได้มากมาย และมีพื้นฐานของการให้ความหมายที่แตกต่างกันตามความสามารถ ของแต่ละบุคคล



cannon

ภาพที่ 3.6 การเลือกสกัดส่วนของวัตถุบนภาพ¹⁰

building

tree

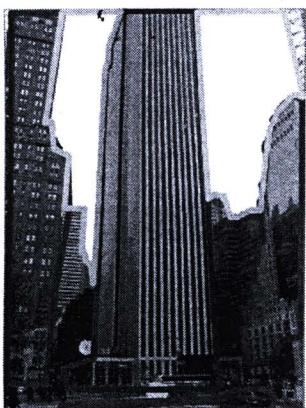
ship

person

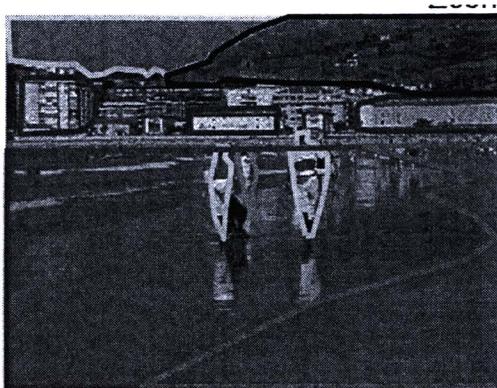
ภาพที่ 3.7 ตัวอย่างวัตถุที่ถูกแท็กด้วยโปรแกรม LabelMe¹⁰

¹⁰ LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554

รูปแบบของการให้ความหมายภาพ หรือแท็กข้อมูลบนโปรแกรม LabelMe จากภาพที่ 3.5 ประกอบด้วยเครื่องมือช่วยการแท็ก (ด้านบน) รูปภาพสามารถคัดเลือกจากฐานข้อมูลภายในโปรแกรม หรือ โหลดรูปภาพที่ต้องการจากเครื่องคอมพิวเตอร์เข้ามาในโปรแกรม LabelMe ได้ วิธีการแท็กวัตถุ แสดงในภาพที่ 3.6 เพื่อทำการแท็กคำหลักบนวัตถุ ด้วยการใช้เมาส์ลากตามสัดส่วนของวัตถุแบบ freehand บนรูปภาพ เมื่อลากตามสัดส่วนของวัตถุเสร็จสิ้นโปรแกรมจะให้ใส่คำหลักเพื่อให้ความหมายของวัตถุ ข้อมูลคำหลักจะถูกจัดเก็บลงบนฐานข้อมูลพร้อมกับรูปภาพ และ ในภาพที่ 3.7 แสดงตัวอย่างของภาพที่ถูกแท็ก จากภาพที่ 3.5 ข้อมูลคำหลักที่ถูกแท็กแล้วจะแสดงไว้ทางขวามือ จะได้ข้อมูลคำหลักของวัตถุบนภาพประกอบด้วย grass snorkel snorkel kid kid ball ball flipper และ flipper



คำหลัก: car traffic light car taxi building sky



คำหลัก: person woman sea person woman sky sea sand



คำหลัก: bus person wheel building



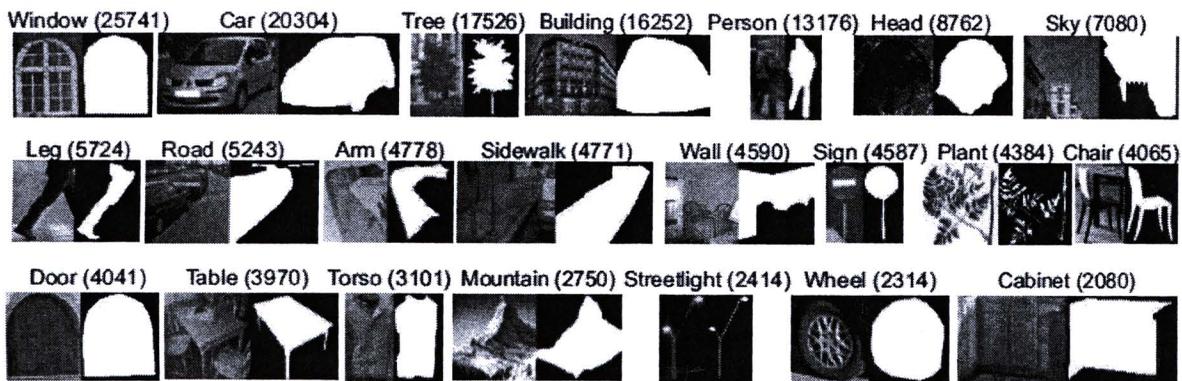
คำหลัก: boat sea water sky mountain wheel building

ภาพที่ 3.8 ตัวอย่างภาพที่ถูกแท็กคำหลักบนภาพ ด้วยโปรแกรม LabelMe¹¹

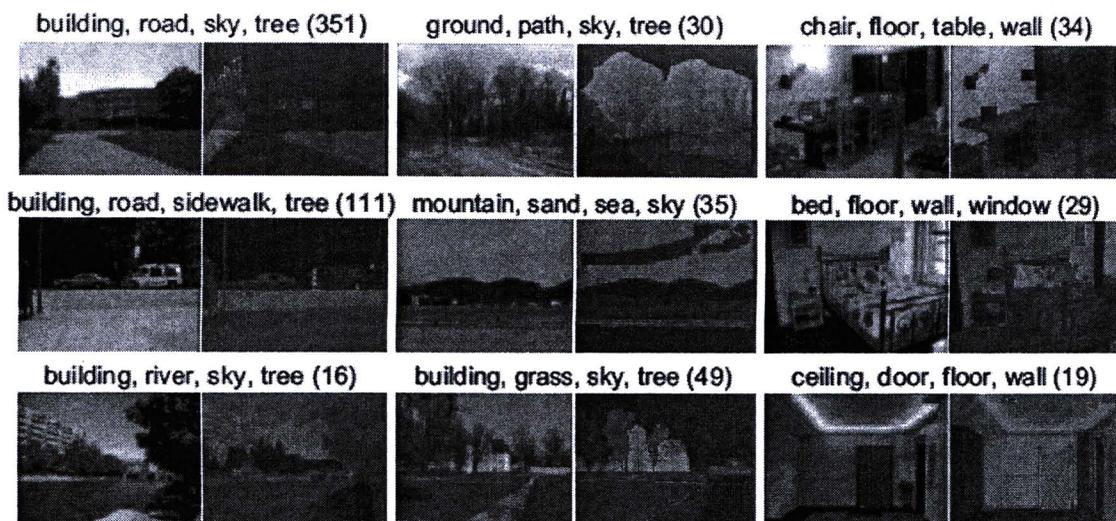
¹¹ LabelMe: <http://labelme.csail.mit.edu/> ค้นคืนเมื่อวันที่ 9 กันยายน 2554



หลังจากที่ผู้ใช้งานแท็กภาพในโปรแกรม LabelMe หรือทำการค้นหารูปภาพที่ถูกแท็ก โปรแกรมจะมีการแสดงภาพที่ถูกแท็ก ดังแสดงในภาพที่ 3.8 แสดงภาพที่ถูกแท็ก เส้นแสดงขอบเขตของคำหลักแต่ละคำที่แท็กไว้ ดังนั้นการใช้งานข้อมูลในโปรแกรม สามารถดาวน์โหลดมาใช้ร่วมกันได้ และข้อมูลคำหลักที่ถูกนำมาใช้มีมากมายหลายฐานดังแสดงในตารางที่ 3.1 ดังนั้นข้อมูลสำหรับการทดลอง จะมีการใช้ฐานข้อมูลที่เก็บรวบรวมมาจาก LabelMe [B. C. Russell, 2008] โดยจะทำการให้ความหมายในลักษณะของรูปทรงแบบ polygon ที่วาดลงบนวัตถุในภาพ ดังแสดงตัวอย่างในภาพที่ 3.9 [A. Torralba, 2010] จากโปรแกรม LabelMe เป็นตัวอย่างคำหลักที่พบบ่อยหรือเกิดขึ้นบ่อยในการให้ความหมายของวัตถุบนภาพ ตัวเลขด้านข้างแทนจำนวนของการให้ความหมาย สำหรับในภาพที่ 3.10 [A. Torralba, 2010] แสดงตัวอย่างของข้อมูลคำหลักที่ถูกแท็กบนภาพในโปรแกรม LabelMe เช่นกัน



ภาพที่ 3.9 ตัวอย่างของคำหลักที่พบบ่อยในการให้ความหมายวัตถุ



ภาพที่ 3.10 ตัวอย่างของคำหลักในการให้ความหมายวัตถุบนภาพ

เพื่อเพิ่มประสิทธิภาพของคำหลักที่มีอยู่จำนวนมากนั้นโปรแกรม LableMe ได้นำความสามารถของ WordNet ที่มีโครงสร้างคล้ายต้นไม้ในการเชื่อมโยงความสัมพันธ์ของคำหลักเป็นการจัดกลุ่มของคำตามกิ่งก้านสาขามาช่วยทำให้ลดความซ้ำซ้อนของคำที่ถูกแท็กลงบนภาพเพื่อจัดข้อมูลให้คงที่เพื่อพร้อมสำหรับการประมวลผลต่อไป [B. C. Russell,2008][A. Torralba,2010]

3.2 ขั้นตอนการประมวลผล

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลคำหลักและฐานข้อมูลภาพได้ถูกคัดเลือก ดังนั้นข้อมูลทั้งหมดที่ได้จะมีการนำภาพเข้ามาเพื่อทำการประมวลผลตามทฤษฎีที่นำเสนอ โดยแบ่งออกเป็น 2 ส่วนดังนี้

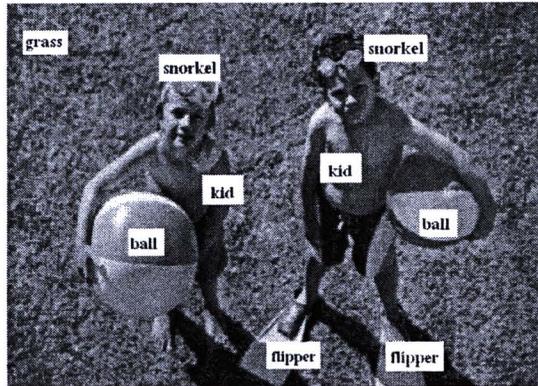
3.2.1 การแทนข้อมูลภาพด้วยแนวคิดกราฟ

ข้อมูลรูปภาพที่ได้มาจะถูกแทนที่ด้วยแนวคิดกราฟ (Conceptual Graph Representation) กำหนดให้รูปภาพใดใด สามารถเขียนความสัมพันธ์ของแนวคิดกราฟได้ดังนี้ $CG \equiv \langle V, E \rangle$ เมื่อ CG เป็นสเปเชียลเอ็นทิตี (spatial entities) ที่แทนด้วยเซตของ V คือจุด (vertex) หรือโหนด (node) เมื่อ $i \in \{1 \dots N\}$ และความสัมพันธ์ของวัตถุภายในภาพใด เกิดจากโหนดสองโหนดที่เชื่อมต่อกันด้วย E เมื่อ $E \subset V \times V$, เมื่อ E คือความสัมพันธ์ระหว่างโหนดสองโหนด (edge) ความสัมพันธ์รูปแบบนี้ถูกเรียกว่า binary spatial relationship ดังนั้นสามารถแทนความสัมพันธ์ของสองจุดได้ด้วย $e_{ab} \equiv \langle v_a, v_b \rangle \in E$ เมื่อกำหนดให้ $v_a, v_b \in V$ จากตัวอย่างภาพที่ 3.11 ก. แสดงภาพตัวอย่างจากโปรแกรม LabelMe ถูกให้ความหมายวัตถุตามคำหลักดังนี้ kid kid grass ball ball snorkel snorkel flipper ดังแสดงในภาพที่ 3.11 ข. และสามารถแทนด้วยความสัมพันธ์ของวัตถุ ด้วยแนวคิดกราฟโหนดแต่ละโหนดบนกราฟถูกแทนด้วยคำหลัก ดังนั้น grass เป็น root node ที่มีความสัมพันธ์ร่วมกันกับ kid ทั้งสองและ โหนดของ kid จะเกิดความสัมพันธ์ต่อไปยังส่วนต่างๆ เช่น ball snorkel และ flipper ดังนั้นการแสดงถึงความสัมพันธ์จะถูกเชื่อมเข้าด้วยกันเป็นค่าของ E

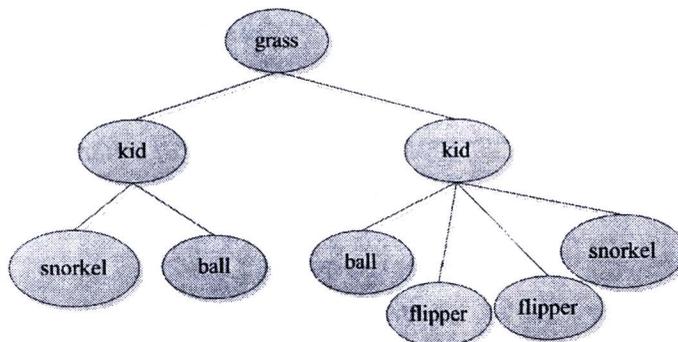


คำหลัก: kid kid grass ball ball snorkel snorkel flipper

ก. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe



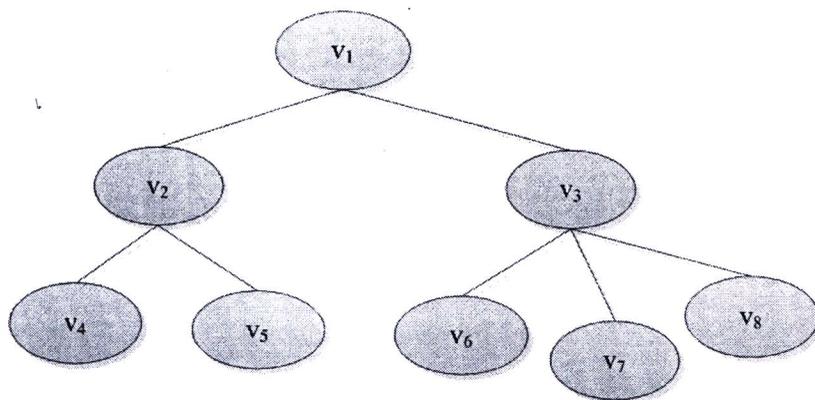
ข. วัตถุที่ถูกแท็กบนภาพจากโปรแกรม LabelMe



ค. แทนความสัมพันธ์ของวัตถุที่ถูกแท็กด้วยแนวคิดกราฟ

ภาพที่ 3.11 ตัวอย่างความสัมพันธ์ของวัตถุด้วยแนวคิดกราฟ

จากภาพที่ 3.11 ค. แสดงความสัมพันธ์ ที่เกิดขึ้นดังนั้นตามแนวคิดกราฟคือโหนดทั้งหมด ประกอบด้วยคำหลักดังนี้ grass kid kid ball ball snorkel snorkel flipper และจุดเชื่อมต่อ ระหว่าง grass กับ kid ถูกแทนด้วย $e_{grass,kid} \equiv \langle v_{grass}, v_{kid} \rangle \in E$ ดังนั้นสามารถเขียนรูปแบบ ใหม่เพื่อ่ายในการอ่านดังแสดงในภาพที่ 3.12 สามารถเขียนเป็นสมการได้ดังนี้ $CG \equiv \langle V, E \rangle$ กำหนดให้ โหนดต่างๆบนภาพด้วย $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$, และแทนความสัมพันธ์ ระหว่างโหนดต่างๆด้วยสมการ เมื่อกำหนดให้ $e_{1,2} \equiv \langle v_1, v_2 \rangle$ ดังนั้นสามารถเขียน edge ทั้งหมด ในกราฟได้ดังนี้ $E = \{e_{1,2}, e_{1,3}, e_{2,4}, e_{2,5}, e_{3,6}, e_{3,6}, e_{3,6}, e_{3,7}, e_{3,8}\}$



ภาพที่ 3.12 ความสัมพันธ์ของวัตถุแทนด้วยแนวคิดกราฟ

3.2.2 การกำหนดค่าน้ำหนักข้อมูล

การกำหนดค่าน้ำหนักสำหรับข้อมูลคำหลักบนภาพ โดยคิดค่าน้ำหนักจากความสัมพันธ์ ข้อมูลคำหลักที่เกิดขึ้นทั้งหมดระหว่างคำหลักภายในภาพโดยจะเก็บเป็นค่า IC (Information Content) [Resnik et al.,1995] แทนรายละเอียดข้อมูลดังนั้นสามารถนับจำนวนคำหลักที่เกิดขึ้น เป็นค่าของความน่าจะเป็นของแต่ละคำหลักบนภาพได้ด้วยสมการดังนี้

$$P(v_i) = \frac{freq(v_i)}{\sum_i freq(v_i)}$$

เมื่อกำหนดให้ $P(v_i)$ แทนความน่าจะเป็นของคำหลัก v_i และ $freq(v_i)$ แทนความถี่ของคำหลัก v_i ที่เกิดขึ้น

$$freq(v_i) = \sum_{n \in word(v)} count(n).$$

เมื่อกำหนดให้ $word(v)$ แทนกลุ่มของคำหลักที่อยู่ในกลุ่มคำหลัก v ดังนั้นเมื่อมีการคำนวณใน แต่ละโหนดสามารถเขียนเป็นสมการของ IC ใน v_i ได้ดังนี้

$$IC(v_i) = \log^{-1} P(v_i)$$

เมื่อคำหลักมีความสัมพันธ์มากกว่าหนึ่งกลุ่มใน WordNet สามารถเขียนสมการหาความสัมพันธ์ของกลุ่มใน WordNet สามารถที่จะเขียนสมการหาความสัมพันธ์ที่เกิดขึ้นสูงสุดได้ของกลุ่มได้จากการหาความเหมือนกันของข้อมูลมากที่สุดของคำหลักดังนี้

$$\text{sim}(v_1, v_2) = \max_{v_1, v_2} [\text{sim}(v_1, v_2)]$$

เมื่อกำหนดให้ v_1 และ v_2 เป็นคำหลักที่เกิดความสัมพันธ์ต่อกันในกลุ่มของ v_1 และ v_2 ใน WordNet ตามลำดับ

3.2.3 การหาความสัมพันธ์ของสายเชื่อมโยงข้อมูล

ภายในภาพหนึ่ง ๆ มีวัตถุที่ถูกแทนเป็นโหนดได้หลายโหนดแต่ละโหนดถูกเชื่อมโยงด้วย edge ระหว่างกันเป็นทอด ๆ สามารถแสดงลำดับความสัมพันธ์ระหว่างโหนดได้ดังแสดงในภาพที่ 3.12 v_2 เป็นโหนดที่มีความสัมพันธ์กับโหนดที่อยู่ด้านบนหรือโหนดพ่อแม่ (parent node) คือ v_1 และ v_2 มีโหนดที่อยู่ด้านล่างหรือโหนดลูก (children node) คือ v_3 และ v_4 ดังนั้นจะเห็นว่าความสัมพันธ์ที่เกิดขึ้นของ v_2 นั้นเกิดขึ้นร่วมกันหลายโหนด ดังนั้นจะหาความสัมพันธ์ที่เกิดขึ้นของโหนดสายการเชื่อมโยง(link) โดยจะใช้การคำนวณในลักษณะของความน่าจะเป็นจากความสัมพันธ์ที่เกิดขึ้นเป็นค่าของ link strength (LS) เป็นลอการิทึมแบบผกผันเพื่อหาค่า LS ของโหนด v_i ได้ดังนี้

$$LS(v_i) = -\log(P(v_i | \text{parent}(v_i)))$$

$$LS(v_i) = IC(v_i) - IC(\text{parent}(v_i))$$

โดยที่ $\text{parent}(v_i)$ แทนโหนดที่อยู่เหนือโหนด v_i เมื่อ $i \in m$, m คือจำนวนของโหนดที่เกิดขึ้นทั้งหมดในภาพ และ $IC(v_i)$ แทน Information Content ของโหนด v_i ดังนั้นสามารถเขียนสมการของความน่าจะเป็นที่จะเกิดขึ้นได้ดังนี้

$$P(v_i / \text{parent}(v_i)) = \frac{P(c_i \cap \text{parent}(v_i))}{P(\text{parent}(v_i))}$$

$$P(v_i / \text{parent}(v_i)) = \frac{P(v_i)}{P(\text{parent}(v_i))}$$

3.2.4 การหาความสัมพันธ์แพทเทิร์นระหว่างโหนด

สำหรับการหาความสัมพันธ์แพทเทิร์นระหว่างโหนด (link pattern: LP) จะเป็นการสื่อถึงความหมายของภาพที่สำคัญ เพราะในภาพจะมีโหนดหรือวัตถุที่มีการเชื่อมโยงต่อกันใน

รูปแบบของแพทเทิร์น ดังนั้นจากข้างต้นมีการกำหนดโหนดต่างๆบนภาพด้วย $V = \{v_1, v_2, \dots, v_n\}$, เมื่อ n แทนจำนวนโหนดบนภาพ v_1 แทนโหนดที่ 1 และแทนความสัมพันธ์ระหว่างโหนดต่างๆด้วย $e_{1,2}$ คือ edge ที่เชื่อมโยงระหว่างโหนด v_1 และโหนด v_2 เมื่อกำหนดให้ $e_{1,2} \equiv \langle v_1, v_2 \rangle$ ดังนั้น $\delta_i = \{e_{1,2}, e_{2,3}, \dots, e_{m,n}\}$ เมื่อ $\delta_i \in \delta$ สามารถเขียนสมการของการ LP ได้ดังนี้

$$LP(\delta_i) = \frac{freq(\delta_i)}{\sum_i freq(\delta_i)}$$

เมื่อกำหนดให้ $freq(\delta_i)$ แทนความถี่ที่เกิดขึ้นจากแพทเทิร์นของ δ_i เมื่อ $i \in n$, n คือจำนวนของแพทเทิร์นที่เกิดขึ้นทั้งหมดในภาพ

3.2.5 การวัดความคล้ายกันของกราฟแบบจับคู่

การหาความคล้ายกันของภาพต้องมีการเปรียบเทียบในแต่ละส่วนของกราฟตามค่านำหนักของแต่ละกิ่งก้านที่เกิดขึ้นบนกราฟ โดยกำหนดให้ G_1 และ G_2 แทนกราฟที่จะทำการเปรียบเทียบการเปรียบเทียบความคล้ายกันมีทั้งหมด 2 ขั้นตอน

– การเปรียบเทียบความคล้ายด้วยข้อมูล

การเปรียบเทียบความคล้ายด้วยข้อมูลภายในกราฟ G_1 และ G_2 แทนกราฟที่ 1 และ 2 โดยที่มีข้อมูลภายในเป็น $v_i \in G_1$ และ $v_j \in G_2$ โดยใช้วิธีการเปรียบเทียบข้อมูลภายในของแต่ละกราฟด้วยสมการดังนี้

$$sim_content(G_1, G_2) = content_match(G_1, G_2)$$

โดยที่อัลกอริทึมของ $content_match$ มีดังนี้

Algorithm 1: Content_Matching.

1. input: Semantic Conceptual Graph G_1 ,
Representative Graph G_2 ;
2. output: the content matching coefficient.
3. BEGIN
4. Initialize $v_i \in G_1, v_j \in G_2$ where v_i, v_j are leaves,
 $cm = 0, total_cm = 0$.
5. While $G_1(v_i)$ is not NULL
6. if $find_content(G_1(v_i) == G_2(v_j))$

```

7.          $cm = cm + LS(G_2(v_j)).$ 
8.          $increment(total\_cm, i).$ 
9.     else if  $find\_content(G_1(parent(v_i)) == G_2(parent(v_j))$ 
10.         $cm = cm + LS(G_2(parent(v_j))).$ 
11.         $increment(total\_cm, i).$ 
12.        else increment  $i$  .
13.    end if
14. end if
15. end while
16. RETURN  $\left(\frac{cm}{total\_cm}\right).$ 

```

– การเปรียบเทียบความคล้ายด้วยความสัมพันธ์ของข้อมูลกราฟ

การเปรียบเทียบความคล้ายด้วยความสัมพันธ์ของข้อมูลภายในกราฟ จะใช้วิธีการเปรียบเทียบข้อมูลภายในของแต่ละกราฟด้วยสมการดังนี้

$$sim_relation(G_1, G_2) = relationship_match(G_1, G_2)$$

โดยที่อัลกอริทึมของ Relationship_match มีดังนี้

Algorithm 2: Relationship_Matching.

```

1. input: Conceptual Graph  $G_1$ ,
        Representative Graph  $G_2$ ;
2. output: the relationship matching coefficient.
3. BEGIN
4. Initialize  $\delta_i \in G_1, \delta_j \in G_2, rm = 0, total\_rm = 0.$ 
5. While  $G_1(\delta_i)$  is not NULL
6.     if  $find\_relationship(G_1(\delta_i) == G_2(\delta_j))$ 
7.          $rm = rm + LP(G_1(R\delta_j)).$ 
8.          $increment(total\_rm, i).$ 
9.     end if
10. end while

```

$$11. \text{ RETURN } \left(\frac{rm}{total_rm} \right).$$

3.3 ขั้นตอนการวัดประสิทธิภาพ

จากภาพที่ 3.1 ขั้นตอนการวัดประสิทธิภาพ (evaluation) เป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ จะนำเอาผลการจำแนกกลุ่มของภาพที่จัดได้มาทำการวัดประสิทธิภาพของสิ่งที่นำเสนอข้างต้น โดยจะตรวจสอบกลุ่มของภาพที่จัดได้ว่ามีค่าเป็นอย่างไร เมื่อเทียบกับกลุ่มของภาพที่ถูกต้อง ซึ่งต้องมีการวัดค่าความระลึก (recall) และค่าความแม่นยำ (precision) จะเป็นค่าที่แสดงว่า การค้นคืนข้อมูลได้ตรงกับความต้องการเพียงใด ส่วนค่าความระลึกจะเป็นค่าที่แสดงถึงความครอบคลุมในการจัดกลุ่มภาพ หลังจากนั้นจะนำค่ามาคำนวณในรูปแบบของค่าความถูกต้อง (accuracy) และ F-measure ต่อไปและนำค่าทั้งหมดมาแปลผลและประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสม หรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ในรูปแบบที่สามารถเข้าใจได้ง่าย เช่น ตาราง หรือกราฟซึ่งในงานวิจัยนี้ได้แสดงในรูปแบบของตารางเชื่อมโยงปฏิสัมพันธ์ความสับสน (confusion matrix) สามารถอ่านเพิ่มเติมได้ในบทที่ 2