การค้นคืนภาพดิจิทัลด้วยการวัดความคล้ายกันบนกราฟแบบมีทิศทางหลายลำดับ

SIMILARITY MEASURE WITH DIRECTED UNIVERSAL HIERARCHICAL GRAPH FOR DIGITAL IMAGE RETRIEVAL

นัศพ์ชาณัณ ชินปัญช์ธนะ
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์
110/1-4 ถ.ประชาชื่น เขตหลักสี่ กรุงเทพฯ 10210

Nutchanun Chinpanthana
College of Innovative Technology and Engineering, Dhurakij Pundit University
110/1-4 Prachachuen rd. Laksi, Bangkok, 10210, Thailand.

บทคัดย่อ

การค้นคืนความหมายภาพยังเป็นหัวข้อที่มีความท้าทายในสาขาการประมวลผลภาพ มีนักวิจัยหลายกลุ่มพยายามปรับวิธีการเพื่อแก้ไขปัญหาของการแทนความหมายภาพด้วยรูปแบบที่แตกต่างกัน ดังนั้นจึงนำเสนอการแทนด้วยโครงสร้างของกราฟแบบมีทิศทาง มีส่วนช่วยให้การค้นหาภาพมีความหมายที่ใกล้เคียงกันมากที่สุด ดังนั้นขั้นตอนสุดท้ายของการค้นหาความหมายภาพ ด้วยการเปรียบเทียบความคล้ายข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการเปรียบเทียบกันทั้งหมด 4 วิธี คือ การวัดความคล้ายด้วยอนุกรมวิธาน การวัดความคล้ายด้วยคิวรี่คำศัพท์ การวัดความคล้ายด้วยคิวรี่ย้อนกลับ และนำเสนอวิธีการวัดความคล้ายกันของภาพด้วยโครงสร้างกราฟตามความสัมพันธ์ลำดับขั้น เปรียบเทียบความหมายลงใน 5 กลุ่มเหตุการณ์ประกอบด้วย ภาพพักผ่อนภายนอก (Outdoor leisure), ภาพพิธีการ (Ceremony), ภาพทำงานในสำนักงาน (Office working), ภาพเล่นกีฬา (Sport game) และ ภาพครอบครัว (Family time) จากการทดลองพบว่าได้ค่าความถูกต้องสูงถึง 81.2% ด้วยการวัดความคล้ายด้วยคิวรี่ย้อนกลับ

คำสำคัญ: การประมวลผลภาพ, กราฟลำดับ, การวัดความคล้าย

ABSTRACT

Digital image retrieval is an active problem in image processing field. Researchers have attempted to solve problem for a new model. Therefore, we are using a direct universal

hierarchical graph concept to represent the image meaning. We compared with 4 similarity measures including WordNet similarity, Keyword-based query similarity, Relevance feedback based on query similarity, and Semantic graph similarity with relational hierarchy structure. The group of images are organized in 5 categories: outdoor leisure, ceremony, office working, sport game and family time. The experimental results indicate that our purposed approach offers significant performance improvements in the interpretation of semantic images with maximum of 81.2% accuracy.

KEYWORDS: Image processing, Hierarchical graph, Similarity measure

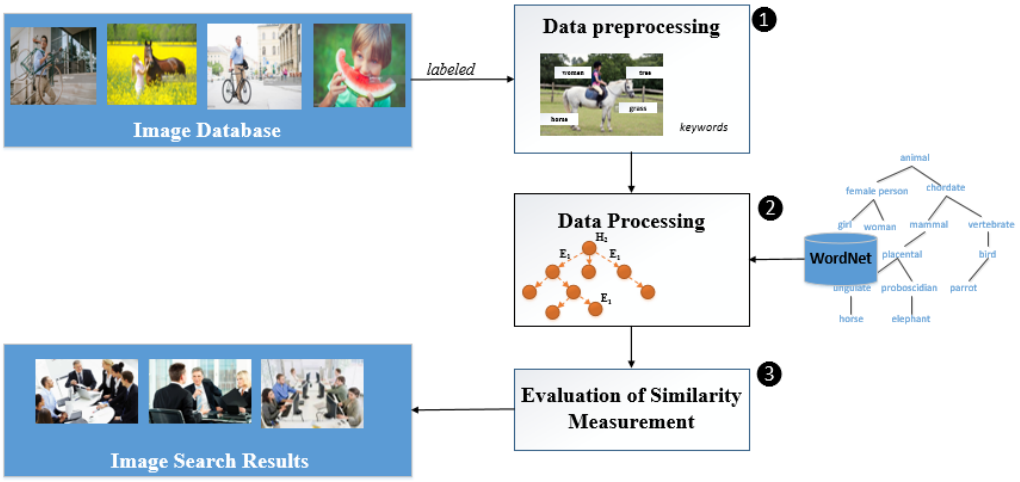
1. บทนำ

การประมวลผลภาพวิวัฒนาการของเครื่องมือและอุปกรณ์อิเล็คทรอนิคสำหรับการถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็วจนทำให้ภาพถ่ายภาพดิจิทัลมีจำนวนมากมาย และแทรกอยู่ทั่วไปไม่ว่าจะอยู่ใน สังคมออนไลน์อย่างเฟสบุค (Facebook), กูเกิล (Google) หรือ อินสตาแกรม (Instagram) เป็นต้น ทำให้เกิดปัญหาที่ตามมาก็คือการจัดเก็บข้อมูลภาพที่เพิ่มขึ้น การจัดเก็บอย่างไรให้มีระบบที่ดีสามารถสืบค้นข้อมูลภาพได้อย่างง่าย และจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความต้องการได้รับความสนใจจากนักวิจัยหลายกลุ่ม เป็นงานด้านการประมวลผลภาพ (Image processing) ด้านการค้นคืนสารสนเทศ (Image retrieval) และการจำแนกประเภทข้อมูลภาพ (Image classification) เพื่อคัดเลือกภาพ ให้ตรงตามความต้องการของผู้ใช้งาน สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ เช่น สี (Color) ลวดลาย (Texture) รูปทรง (Shape) เป็นต้น [1] งานวิจัยของ Xin Lu [2] เป็นการใช้คุณลักษณะระดับต่ำด้วยรูปทรง สี ลวดลาย และลักษณะของอารมณ์ภายในภาพเพื่อค้นหาข้อมูลภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างเห็นได้ชัดเจน การประมวลผลภาพระดับต่ำนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน แต่มีกลุ่มนักวิจัยที่พยายามปรับปรุงอัลกอริทึม เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมาก การปรับปรุงเทคนิค ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ใกล้เคียงกับภาพที่ต้องการให้มากขึ้นในรูปแบบที่ซับซ้อน แต่อย่างไรก็ตามในการค้นคืนที่ใช้คุณลักษณะฟีเจอร์ระดับต่ำเพียงอย่างเดียว ทำให้ได้ผลลัพธ์ส่วนใหญ่ตรงกับคุณลักษณะของฟีเจอร์ที่สกัดมาแต่ไม่ได้ตรงกับความหมายที่เกิดภายในภาพที่ต้องการอย่างแท้จริง

งานวิจัยอีกกลุ่มที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) [3] ที่มีความหมายและแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บนภาพ เรียกการแทนข้อมูลบนภาพแบบนี้ว่า Keyword-based image indexing [4] เป็นการให้ความหมายของวัตถุบนภาพเป็น ชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกัน เช่น “grass” , “plant” , “boat”, “sky” เป็นต้น หรือแสดงเนื้อหาภาพเป็นคำอธิบาย สิ่งต่างๆ ที่เกี่ยวข้องกับภาพ เช่น ชื่อเรื่อง, ผู้ถ่ายภาพ, รายละเอียดภาพ เป็นต้น ส่วนใหญ่ในการให้ความหมายหรือการเขียนคำอธิบายภาพจะเป็นลักษณะของอัตโนมัติบนภาพหรือใช้การคัดเลือกคำศัพท์ด้วยมนุษย์ ขึ้นกับระบบที่เลือกใช้งาน ในปัจจุบันระบบที่ใช้ในการให้ความหมายภาพ จะสามารถให้ความหมายวัตถุบนภาพได้อย่างง่ายดายและรวดเร็ว การให้ความหมายของภาพด้วยคำศัพท์ต่างๆเพื่อทำให้การสืบค้นข้อมูลได้ผลดีขึ้นจากวิธีแรกที่ใช้เพียงคุณลักษณะต่ำเพียงอย่างเดียว แต่ยังคงมีปัญหาที่ตามมาจากผลของการค้นคืนเพื่อให้ถูกต้องและตรงตามความหมายมากยิ่งขึ้น การใช้หลายคำศัพท์ (Multiple label) เป็นอีกวิธีการหนึ่งที่ใช้ช่วยการค้นคืนแต่ก็ยังถูกจำกัดด้วยการค้นหาเพื่อให้ได้ภาพที่ตรงกับความหมายภาพที่เพิ่มขึ้น นักวิจัยบางกลุ่ม [5, 6] พยายามใช้มิติของความสัมพันธ์ของคำศัพท์ เพื่อหาความสัมพันธ์ของความหมายของการแทนคำ ร่วมกับวิธีการเรียนรู้ด้วย Neural Network เพื่อสร้างความสัมพันธ์ของคุณลักษณะคำศัพท์ที่เกิดขึ้น นักวิจัยบางกลุ่มจัดภาพที่มีวัตถุเหมือนกันให้อยู่ในกลุ่มเดียวกันด้วยวิธี Query-Frequency Pair [4] เป็นอีกวิธีการหนึ่งที่พยายามค้นคืนภาพด้วย คิวรี่ที่ซับซ้อนที่ประกอบด้วยคำศัพท์ที่มีการบันทึกไว้ ผลลัพธ์ที่ได้นั้นจะได้ผลที่ดีเมื่อการค้นคืนใช้รูปแบบของคำศัพท์ในแบบเดียวกัน แต่เมื่อมีความแตกต่างของการให้ความหมายคำศัพท์ผลที่ได้จะผิดเพี้ยนไป

เพื่อให้ผลลัพธ์ที่ได้มีความถูกต้องมากยิ่งขึ้น การค้นหาภาพด้วยเทคนิคที่กล่าวมานี้ จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกให้ความหมายไว้บนภาพ ยิ่งมีการแท็กข้อมูลบนภาพมากยิ่งสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่เมื่อการให้ความหมายของวัตถุบนภาพมาจากหลายที่ หลายแหล่งข้อมูล หรือหลายผู้บันทึกข้อมูลทำให้คำศัพท์บางคำมีความหมายที่เหมือนกันแต่ต่างคำกัน เช่น “sky” กับ “trees”, “car” กับ“vehicle” เป็นต้น หรือคำศัพท์บางคำที่ให้ความหมายได้ลึกแตกต่างกัน เช่น “mango” กับ “fruit”, “mobile phone” กับ “electronic device” เป็นต้น การให้ความหมายเพียงผิวเผิน หรือความหมายที่ตรงกับวัตถุจริงหรือเพียงกลุ่มของวัตถุเท่านั้น จะมีผลต่อการค้นคืนภาพรวมทั้งมีผลต่อการแปลความหมายภาพด้วยเช่นกัน ดังนั้นในงานวิจัยนี้จึงนำเสนอวิธีการจัดการกลุ่มคำศัพท์และแทนข้อมูลความสัมพันธ์ภาพนั้นด้วยรูปแบบออนโทโลจี (Ontology) [7] ซึ่งออนโทโลจีมีโครงสร้างแบบดัชนีหลายลำดับ (Multilevel index) สามารถนำมาแทนกลุ่มความหมายคำศัพท์ได้ และใช้วิธีการเปรียบเทียบความเหมือนกันของความหมายภาพด้วยการหาความเหมือนของภาพด้วย รูปแบบการวัดความเหมือนของโครงสร้างกราฟ (Similarity Graph Matching) และ การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering Algorithm)

ดังนั้นจึงนำเสนอการแทนด้วยโครงสร้างของกราฟ มีส่วนช่วยให้การค้นหาภาพมีความหมายที่ใกล้เคียงกันมากที่สุด ดังนั้นขั้นตอนสุดท้ายของการค้นหาความหมายภาพ ด้วยการเปรียบเทียบความคล้ายข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการเปรียบเทียบกันทั้งหมด 4 วิธี การวัดความคล้ายด้วยอนุกรมวิธาน (WordNet Similarity: WS) การวัดความคล้ายด้วยคิวรี่คำศัพท์ (Keyword-based Query Similarity: KQS) การวัดความคล้ายด้วยคิวรี่ย้อนกลับ (Relevance Feedback based on Query Similarity: RFQS) และนำเสนอวิธีการวัดความคล้ายกันของภาพด้วยโครงสร้างกราฟตามความสัมพันธ์ลำดับขั้น (Semantic Graph Similarity with Relational Hierarchy Structure: RHS) เปรียบเทียบความหมายลงใน 5 กลุ่มเหตุการณ์ประกอบด้วย ภาพพักผ่อนภายนอก (Outdoor leisure), ภาพพิธีการ (Ceremony), ภาพทำงานในสำนักงาน (Office working), ภาพเล่นกีฬา (Sport game) และ ภาพครอบครัว (Family time)



รูปที่ 1 ขั้นตอนการค้นคืนภาพดิจิทัล

2. ขั้นตอนการค้นคืนภาพดิจิทัล

การเชื่อมโยงความสัมพันธ์ของข้อมูลภายในภาพดิจิทัลเพื่อค้นคืนภาพที่ต้องการ เริ่มจากการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสม เพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอนวิธีการดำเนินงานวิจัยโดยทั่วไปจะแบ่งออกเป็น 3 ส่วนหลักดังนี้ (1) ขั้นตอนการเตรียมข้อมูล (Data preprocessing) เป็นการทำงานในส่วนของการนำข้อมูลเข้าด้วยเครื่องมือ (Image annotation tool) และการแทนวัตถุลงในกราฟ (Object representation into graph) (2) ขั้นตอนการประมวลผล (Data processing) ทำการนำข้อมูลที่ได้ทั้งหมดมาทำงานบนกราฟแบบลำดับขั้น

(3) ขั้นตอนการวัดประสิทธิภาพความคล้ายกันของภาพ (Evaluation of similarity measurement) เป็นการเปรียบเทียบการทำงานของวิธีการที่นำเสนอ ดังแสดงในรูปที่ 1

ขั้นตอนการเตรียมข้อมูล เริ่มจากคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมายภาพนั้นได้อย่างสมบูรณ์ ภาพบางภาพมีลักษณะผิดปกติ (Outlier) หรือคุณลักษณะวัตถุ (Object characteristic) ไม่ชัดคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถบ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (Close up) ทำให้ต้องมีการคัดเลือกภาพออกไปไม่นำมาใช้ในการทดลอง ได้มีการคัดเลือกฐานข้อมูลภาพแอ็คชันสำหรับการทดลองจาก PASCAL action [8] โดยจัดให้อยู่ในหมวดหมู่และทำการแท็กข้อมูลภาพสำหรับการทดลองใหม่ เพื่อให้อยู่ในกลุ่มและคำหลักที่สามารถใช้ทดลองได้ ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะไม่ถูกนำมาพิจารณา

3. การแทนที่ข้อมูลกราฟแบบมีทิศทางหลายลำดับ

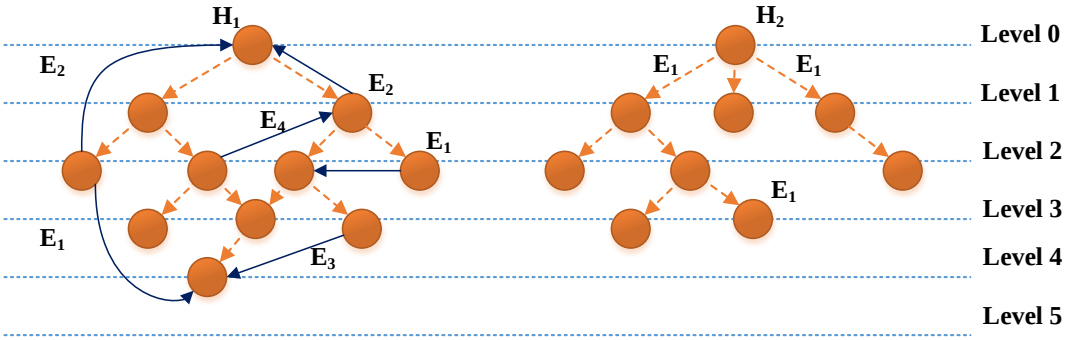
การแทนที่ข้อมูลภายในภาพด้วยคำหลักกับวัตถุที่เด่นชัด โดยจะทำการคัดเลือกเฉพาะคำหลักที่เป็นคำนาม (Noun) เช่น “woman”, “kid”, “bike” เป็นต้น โครงสร้างความสัมพันธ์ของคำหลักถูกอ้างอิงจากความสัมพันธ์ของ WordNet [9] ตามหมวดหมู่ที่มีการแบ่งประเภทของคำนามไว้

3.1 ความสัมพันธ์ระหว่างวัตถุด้วยแนวคิดกราฟ

หลังจากผ่านขั้นตอนของการเตรียมข้อมูลภาพ ฐานข้อมูลคำหลักและฐานข้อมูลภาพได้ถูกคัดเลือกเข้ามา จะทำการจัดเก็บข้อมูลคำหลักที่ได้ลงในกราฟ ข้อมูลวัตถุภายในภาพที่ถูกจัดเก็บลงในเมตริกซ์จะถูกสร้างความสัมพันธ์ระหว่างวัตถุด้วยแนวคิดกราฟ (Conceptual graph) จากคุณสมบัติของคลาสกราฟสามารถเขียนเป็นทฤษฎีกราฟได้ดังนี้

กำหนดให้ $V \neq \phi$ เป็นเซตของเวคเตอร์ $H = (\mathrm{V}, \mathrm{E}), |V| < \infty$ เป็นกราฟจำกัดที่ระบุทิศทาง เมื่อ$E \subseteq \binom{V}{2}, V$ แทนจุดหรือโหนดและ E คือเซตของเส้นเชื่อมความสัมพันธ์ระหว่างจุด

กำหนดให้ $V \neq \phi$ เป็นเซตของเวคเตอร์ และเรียก $H = (\mathrm{V}, \mathrm{E}), E \subseteq V \times V, |V| < \infty$ เป็นกราฟจำกัดที่มีการระบุทิศทาง



รูปที่ 2 ตัวอย่างการแทนความสัมพันธ์ตามแนวคิดกราฟ

กำหนดให้ $H = (\mathrm{V}, \mathrm{E})$ เป็นจำนวนจำกัดสำหรับกราฟมีทิศทาง และ $G = (\widehat{\mathrm{V}}, \widehat{\mathrm{E}})$ เป็นกราฟย่อยของ H หมายถึง $\widehat{\mathrm{V}} \subseteq V$ และ $\widehat{\mathrm{E}} \subseteq E$ กำหนดให้กราฟย่อย G ของ H ด้วย $G \subset H$ ในกรณีของ $\widehat{\mathrm{E}} = E \cap (\widehat{\mathrm{V}} \times \widehat{\mathrm{V}})$ เมื่อเรียก G เป็นกราฟย่อยของ H,

กำหนดให้ $H = (V, E)$ เป็นกราฟจำกัดแบบมีทิศทาง โดยกำหนดให้มีคุณลักษณะต่าง ๆ ดังนี้ $\xi_+(v) \coloneqq \{\tilde{v} \in V\{v\} | (v, \tilde{v}) \in E\}, \xi_-(v) \coloneqq \{\tilde{v} \in V\{v\} | (\tilde{v}, v) \in E\}$, แล้ว $\varsigma_{out}(v) \coloneqq |\xi_+(v)|, \varsigma_{in}(v) \coloneqq |\xi_-(v)|$.

กำหนดให้โหนดรากของต้นไม้เป็น $T = (V_T, E_T)$ เป็นกราฟที่มีเพียงหนึ่งโหนด $r \in V_T$ เมื่อ $\varsigma_{in}(r) = 0$ ทุกๆ โหนดใน T จะไม่ซ้ำกันและสามารถเข้าถึงจาก r ที่ถูกแทนเป็นโหนดราก ทฤษฎีของโครงสร้างต้นไม้ทั่วไป สามารถเขียน $T = (V, E_1)$ แทนรากต้นไม้ จุดกำเนิดสามารถเขียนได้ดังนี้

$$V \coloneqq \{v_{0,1}, v_{0,2}, \dots, v_{0,|V_0|}, v_{1,1}, v_{1,2}, \dots, v_{1,|V_1|} v_{2,1}, v_{2,2}, \dots, v_{d,1}, v_{d,2}, \dots, v_{d,|V_d|}\}$$

โดยที่กำหนดให้ $|V| < \infty$. $|\kappa|$ แทนค่าเป็นคาร์ดินอลลิตี้ (Cardinality) ของระดับ เซต $\boldsymbol{\kappa}$ และ $\boldsymbol{d}$ แทนความลึกของ T $|\kappa| = d + 1$ และทำการแมพ $L: V \to \kappa$ ถูกเรียกว่าฟังกชันหลายระดับ (Multi-level function) โดยกำหนดให้กับทุกๆจุดในอิลิเมนต์ของระดับในเซต $\boldsymbol{\kappa}$ เมื่อกำหนดให้ $v_{i,j}$ เป็นจุดของตำแหน่งที่ j บนระดับ i เมื่อ $0 \le i \le d, 1 \le j \le |V_i|$. $|V_i|$ กำหนดให้จำนวนจุดบนระดับ i เซตของเส้นเชื่อมระหว่างจุด $E_{GT} \coloneqq E_1 \cup E_2 \cup E_3 \cup E_4$ ของกลุ่มต้นไม้ที่ถูกกำหนดไว้แล้ว รูปแบบของ E_1 เป็นเซตของเส้นเชื่อมที่อยู่ภายใต้รากของต้นไม้ T, E_2 เป็นเส้นเชื่อมความสัมพันธ์ของลำดับขั้นของจุดก่อนหน้า (Up-edge) โดยที่รูปแบบความสัมพันธ์ดังนี้

$$\begin{aligned} E_2 \coloneqq \{(v_{i+s,\eta^{i+s}}, v_{i,\eta^i}) \Big| & v_{i+s,\eta^{i+s}}, v_{i,\eta^i} \in V, L(v_{i,\eta^i}) = L(v_{i+s,\eta^{i+s}}) - s, \\ & 1 \le s \le d \wedge \exists! (v_{i,\eta^i}, v_{i+1,\eta^{i+1}}), \dots, (v_{i+s-1,\eta^{i+s-1}}, v_{i+s,\eta^{i+s}})), \\ & 1 \le \eta^i \le |V_i|, \dots, 1 \le \eta^{i+s-1} \le |V_{i+s-1}|, 1 \le \eta^{i+s} \le |V_{i+s}|\}, \end{aligned}$$

E_3 เป็นเส้นเชื่อมที่แสดงความสัมพันธ์ของโครงสร้างบนต้นไม้ ทุกจุดที่อยู่ถัดไป (Down-edge) โดยที่รูปแบบความสัมพันธ์ดังนี้

$$E_3 := \{(v_{i,\eta^i}, v_{i+s,\eta^{i+s}}) \Big| v_{i,\eta^i}, v_{i+s,\eta^{i+s}} \in V, L(v_{i+s,\eta^{i+s}}) = L(v_{i,\eta^i}) + s, \\ 1 \le s \le d \wedge \exists! (v_{i,\eta^i}, v_{i+1,\eta^{i+1}}), \dots, (v_{i+s-1,\eta^{i+s-1}}, v_{i+s,\eta^{i+s}})), \\ 1 \le \eta^i \le |V_i|, \dots, 1 \le \eta^{i+s-1} \le |V_{i+s-1}|, 1 \le \eta^{i+s} \le |V_{i+s}|\},$$

E_4 เป็นเส้นเชื่อมที่แสดงความสัมพันธ์ระหว่างจุดของโครงสร้างลำดับ โดยกำหนดให้เซต $E_4 := E_4^{i\to i} \cup E_4^{i+s\to i} \cup E_4^{i\to i+s}$ เป็นความสัมพันธ์ที่กำหนดการรวมกันระหว่างเซตย่อยของ E_4 เมื่อ

$$E_4^{i\to i} := \{(v_{i,\tilde{\eta}^i}, v_{i,\hat{\eta}^i}) \Big| v_{i,\tilde{\eta}^i}, v_{i,\hat{\eta}^i} \in V, 0 \le i \le d, L(v_{i,\tilde{\eta}^i}) = L(v_{i,\hat{\eta}^i}) \wedge (\tilde{\eta}^i < \hat{\eta}^i \vee \tilde{\eta}^i > \hat{\eta}^i)\},$$

$$E_4^{i+s\to i} := \{(v_{i+s,\eta^{i+s}}, v_{i,\eta}) \Big| v_{i+s,\eta^{i+s}}, v_{i,\eta^i} \in V, (v_{i+s,\eta^{i+s}}, v_{i,\eta^i}) \notin E_2, L(v_{i,\eta^i}) = L(v_{i+s,\eta^{i+s}}) - s, 1 \le s \le d\},$$

$$E_4^{i\to i+s} := \{(v_{i,\eta^i}, v_{i+s,\eta^{i+s}}) \Big| v_{i,\eta^i}, v_{i+s,\eta^{i+s}} \in V, (v_{i+s,\eta^i}, v_{i+s,\eta^{i+s}}) \notin E_1, E_3, L(v_{i+s,\eta^{i+s}}) = L(v_{i,\eta^i}) + s, 1 \le s \le d\}.$$

ถ้า E_2, E_3หรือ E_4 มีข้อมูลแล้ว $H = (V, E_{GT})$ แทนเป็นลักษณะของต้นไม้ทั่วไป (Finite general tree) จากรูปที่ 2 แสดงความสัมพันธ์ของเส้นเชื่อมบนต้นไม้ทั่วไป H_1 และโครงสร้างของต้นไม้ ทั่วไป H_2 ที่สัมพันธ์กับระดับต่างๆ

3.2 กราฟลำดับขั้นแบบมีทิศทาง (Directed Universal Hierarchical Graphs)

ในส่วนนี้ได้กล่าวถึง กราฟลำดับขั้นแบบมีทิศทาง (Directed Universal Hierarchical Graphs) เป็นคลาสของกราฟทั่วไปในลักษณะหลายลำดับขั้นในรูปแบบโครงสร้างต้นไม้ทั่วไป เมื่อกำหนดให้โหนดรากอยู่ในระดับ 0 ซึ่งกราฟจะมีความซับซ้อนกว่าโครงสร้างต้นไม้ จะมีการกำหนดโครงสร้างพื้นฐานของกราฟแบบมีทิศทางไว้ดังนี้กำหนดเป็น

$$V := \{v_{0,1}, v_{0,2}, \dots, v_{0,|V_0|}, v_{1,1}, v_{1,2}, \dots, v_{1,|V_1|} v_{2,1}, v_{2,2}, \dots, v_{d,1}, v_{d,2}, \dots, v_{d,|V_d|}\}$$

โดยที่กำหนดให้ $|V| < \infty$. และกำหนดให้ $L: V \to \kappa$ เป็นฟังก์ชันหลายระดับ จะถูกกำหนดให้ทุกจุดของอิลิเมนต์ในระดับเซต κ และ $d = |\kappa| - 1$ กำหนดให้ $v_{i,j}$ เป็นจุดของตำแหน่งที่ j บนระดับ i เมื่อ $0 \le i \le d, 1 \le j \le |V_i|$. $|V_i|$ กำหนดให้จำนวนจุดบนระดับ i เซตของเส้นเชื่อม $E_{DUHG} := E_1 \cup E_2 \cup E_3$ ถูกกำหนดไว้แล้วดังนี้

E_1 เส้นเชื่อมลง (Down-edges) เป็นเส้นเชื่อมที่มีการเปลี่ยนแปลง 1 ระดับ

$$E_1 := \{(v_{i,\eta^i}, v_{i+s,\eta_j^{i+s}}) \mid v_{i,\eta^i}, v_{i+s,\eta_j^{i+s}} \in V, 0 \le i \le d, 1 \le j \le |V_i|, L(v_{i+s,\eta_j^{i+s}}) = L(v_{i,\eta^i}) + s, 1 \le s \le d\}.$$

E_2 เส้นเชื่อมขึ้น (Up-edges) เป็นการเปลี่ยนแปลงอย่างน้อย 1 ระดับ

$$E_2 := \{(v_{i+s,\eta_j^{i+s}}, v_{i,\eta^i}) \mid v_{i+s,\eta_j^{i+s}}, v_{i,\eta^i} \in V, 0 \le i \le d, 1 \le j \le |V_i|, L(v_{i,\eta^i}) = L(v_{i+s,\eta_j^{i+s}}) - s, 1 \le s \le d\},$$

E_3 เส้นเชื่อมระหว่างกัน (Across-edges) ไม่มีการเปลี่ยนแปลงระดับ

$$E_3 := \{(v_{i,\tilde{\eta}^i}, v_{i,\hat{\eta}^i}) \mid v_{i,\tilde{\eta}^i}, v_{i,\hat{\eta}^i} \in V, 0 \le i \le d, L\left(v_{i,\tilde{\eta}^i}\right) = L(v_{i,\hat{\eta}^i}) \wedge (\tilde{\eta}^i < \hat{\eta}^i \vee \tilde{\eta}^i > \hat{\eta}^i)\},$$

แล้ว $H = (V, E_{DUHG})$ ถูกเรียกว่ากราฟลำดับขั้นแบบมีทิศทาง

4. การวัดความคล้ายของกราฟแบบหลายลำดับขั้น

การเปรียบเทียบความคล้ายกันของกราฟ (Graph Similarity) เป็นการนำข้อมูลที่มีความสัมพันธ์ในรูปแบบกราฟทั้งหมดที่เก็บรวบรวม ผ่านกระบวนการขั้นตอนการเปรียบเทียบความคล้ายกันของกราฟแยกแยะข้อมูลลงในแต่ละกลุ่มที่จัดไว้ โดยในแต่ละกลุ่มของข้อมูลนั้นจะมีคุณลักษณะเด่นของแต่ละกลุ่มที่แตกต่างกัน ขึ้นกับข้อมูลที่เก็บรวบรวมมาได้ การเชื่อมโยงความสัมพันธ์ข้อมูลภาพ ด้วยกราฟลำดับขั้น ข้อมูลภาพได้จากการแท็กของผู้ใช้ออนไลน์ ที่มีผู้ใช้ที่แตกต่างกันทำให้เกิดความแตกต่างของคำศัพท์บ้างเล็กน้อย กับการแท็กวัตถุประเภทเดียวกัน แต่มีการแท็กคำที่ต่างกัน (เช่น คำศัพท์ “stone” หรือ “rock” มีความหมายใกล้เคียงกัน) เมื่อมีการค้นคืนด้วยคำศัพท์จะถูกสร้างความแตกต่างของภาพ แต่ในความเป็นจริงแล้วมีความหมายที่ใกล้เคียงกัน ดังนั้นจึงทำการลดความซ้ำซ้อนหรือกำกวมของคำศัพท์ที่เกิดขึ้นด้วยการใช้รูปแบบความสัมพันธ์ของคำศัพท์ WordNet [9] ดังนั้นเมื่อคำศัพท์สองคำหรือมากกว่ามีความหมายเหมือนกัน (synonym) จะถูกแทนค่าวัดความเหมือนได้ด้วยความหมายเดียวกัน ในรูปแบบความสัมพันธ์ Synset ตามอนุกรมวิธาน (Taxonomy) ของ WordNet และขั้นตอนสุดท้ายขั้นตอนการประมวลผล (Data processing) ดังนั้นวิธีการเปรียบเทียบความคล้ายข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการเปรียบเทียบกันทั้งหมด 4 วิธี

4.1 การวัดความคล้ายด้วยอนุกรมวิธาน

การวัดความคล้ายด้วยอนุกรมวิธาน (WordNet Similarity: WS) ลำดับแรกจะมีการกำหนดความเหมือนกันของคำศัพท์ที่อยู่บนพื้นฐานของอนุกรมวิธาน WordNet ที่มีความสัมพันธ์ของคำศัพท์แบบลำดับขั้น โดยที่จะมีการเลือกใช้ในส่วนเฉพาะโครงสร้างของ WS ที่มีสัดส่วนของเฉพาะข้อมูลคำศัพท์ที่มีในฐานข้อมูลภาพเท่านั้นทำให้มีการเปรียบเทียบที่จำกัดลง และการวัดความคล้ายที่ถูกนำมาใช้เป็นรูปแบบของความสัมพันธ์ระหว่างคำ (Interword relationships) ที่มีการสอดคล้องกับ Synset ดังนั้นกำหนดให้ การวัดความคล้ายของ s_1 และ s_2 เป็น Synset และมีความลึกของโหนดบรรพบุรุษของ Synset เป็น s_a แล้วสามารถเขียนสมการแสดงการวัดความคล้ายกันของความลึกระหว่างของ s_1 และ s_2 ได้ดังแสดงในสมการ

$$Sim\_WS(s_1, s_2) = depth(s_a)/depth_{max}$$

4.2 การวัดความคล้ายด้วยคิวรี่คำศัพท์

การวัดความคล้ายด้วยคิวรี่คำศัพท์ (Keyword-based Query Similarity: KQS) เป็นการวัดความคล้ายแบบคำต่อคำที่มีการกำหนดค่าถ่วงน้ำหนักภายในคำศัพท์ที่ถูกแท็กเพื่อใช้เป็นเครื่องมือวัดรูปแบบหนึ่งระหว่างคิวรี่และภาพ โดยที่เมื่อกำหนดให้ $\rho = \{< k_1, w_{k_1} >, < k_2, w_{k_2} >, \dots, < k_m, w_{k_m} >\}$ ให้ k_i แทนคำศัพท์หรือ Synset ที่ i ที่มีการแท็กบนภาพด้วยค่าถ่วงน้ำหนักมีค่าเป็น w_{k_i} ที่ i คิวรี่ของการวัดความคล้ายสามารถเขียนได้ $Q = \{< q_1, w_{q_1} >, < q_2, w_{q_2} >, \dots, < q_m, w_{q_m} >\}$ ทุกคำศัพท์เมื่อผู้ใช้งานมีการค้นหาจะมีค่าถ่วงน้ำหนักเริ่มต้นคงเดิม แต่เมื่อมีการโต้ตอบของผู้ใช้งานเกิดขึ้นจะมีการปรับค่าถ่วงน้ำหนักไป ซึ่งสมการของการปรับเปลี่ยนค่าถ่วงน้ำหนักสามารถเขียนได้ดังนี้

$$Sim\_KQS(Q, \rho) = \sum_{i=1}^{n} \max_{j=1\dots m} [sim(q_i, k_j) \cdot w_{q_i} \cdot w_{k_j}] / n$$

เมื่อกำหนดให้ $sim\_KQS(q_i, k_j)$ เป็นการวัดความคล้ายระหว่าง q_i และ k_j เป็นการค้นหาคำศัพท์ของแต่ละคิวรี่และมีการคำนวณค่าเฉลี่ยที่ได้จากค่าสูงสุดของการวัดค่าความคล้ายระหว่างคิวรี่และภาพค่าความคล้ายกันของค่าน้ำหนักของคำศัพท์ w_{q_i} และ w_{k_j} ค่ามากจะแสดงถึงความเหมือนกันของคำศัพท์

4.3 การวัดความคล้ายด้วยคิวรี่ย้อนกลับ

การวัดความคล้ายด้วยคิวรี่ย้อนกลับ (Relevance Feedback based on Query Similarity: RFQS) เป็นการสร้างตำแหน่งและที่ตั้งที่มีความสอดคล้องกันบนโครงสร้างความสัมพันธ์ก่อนหน้า

เพื่อนำมาใช้ในการวัดความคล้ายด้วยคิวรี่คำศัพท์เพื่อค้นหาภาพที่คล้ายกัน ผลที่ได้จากการค้นคืนด้วยคิวรี่ จะได้ผลย้อนกลับ (Relevance feedback) ที่เป็นทั้งภาพคล้ายกันและภาพต่างกันตามระดับลักษณะของภาพ ดังนั้นในการนำผลที่ได้จากการค้นคืนด้วยคิวรี่หลายครั้งมารวมกันเพื่อให้ได้ผลลัพธ์ที่ใกล้เคียงมากที่สุด เมื่อทุกคำศัพท์ที่ถูกแท็กบนตัวอย่างที่ถูกต้องจะเป็นคิวรี่บวก (positive query: Q_p) ด้วยค่าน้ำหนักที่เกิดขึ้นจากแต่ละการแท็ก และQ_n เป็นคิวรี่ที่ถูกสร้างจากการย้อนกลับแบบลบ (negative query: Q_n) ดังนั้นการสร้างค่าความคล้ายสุดท้ายของ S_RFQS_i เมื่อ i เป็นลำดับภาพบนฐานข้อมูลสามารถเขียนเป็นสมการได้ดังนี้

$$S\_\mathrm{RFQS}_i = sim(Q,\rho_i) + \delta\left(1 + sim(Q_p,\rho_i)\right)\sum_{k\epsilon Np} S_{ik} + \beta(1+1+sim(Q_n,\rho_i))\sum_{k\epsilon Nn} S_{ik}$$

เมื่อกำหนดให้ δ และ β เป็นค่าคงที่ และ S_{ik} เป็นค่าความคล้ายระหว่างภาพ

4.4 การวัดความคล้ายด้วยโครงสร้างความสัมพันธ์ลำดับขั้น

การวัดความคล้ายกันของภาพด้วยโครงสร้างความสัมพันธ์ลำดับขั้น (Semantic similarity with relational hierarchy structure: RHS) ดังนั้นเป็นขั้นตอนสุดท้ายของกระบวนการจำแนกความหมายภาพ โดยจะมีการแยกเปรียบเทียบสองส่วนคือ ส่วนที่เกิดจากวัตถุภายในภาพซึ่งจะเปรียบเทียบความต่างของกลุ่มวัตถุและเปรียบเทียบความแตกต่างของภาพซึ่งถูกแทนด้วยกราฟ ดังนั้นจึงได้ทำการแบ่ง ลักษณะของการเปรียบเทียบออกเป็น 2 ส่วนดังนี้

4.4.1 การเปรียบเทียบความเหมือนของวัตถุ

กำหนดให้ $\xi_1,\xi_1,\dots,\xi_n$ เป็นจำนวนเต็มและให้ O_i,O_j เป็นวัตถุ ดังนั้นเขียนเป็นสมการเพื่อแสดงความสัมพันธ์ของวัตถุดังนี้ $\Upsilon\left(O_i,O_j,\xi_1,\xi_1,\dots,\xi_n\right) := \frac{\beta(\xi_1),\beta(\xi_1),\dots,\beta(\xi_n)}{n}$ เมื่อ $\beta(\xi_i) \le 1, 1 \le i \le n$, เมื่อ $\Upsilon(\xi_1,\xi_1,\dots,\xi_n)$ คือการวัดความเหมือนของวัตถุ และได้มีการกำหนดเงื่อนไขเพิ่มเติมของการวัดค่าความเหมือน เมื่อ $\beta(\xi_i) \le 1, 1 \le i \le n$, และ $\Upsilon(\xi_1,\xi_1,\dots,\xi_n)$ เป็น ภายใต้สมมุติฐานเมื่อ $O_i = O_j, \beta(\xi_i) = 0, 1 \le i \le n$ นอกเหนือจากนี้ยังมีการวัดความเหมือนของวัตถุจากจำนวนดีกรีเข้าและออกในระดับ i จากโหนดดังนี้

$$\Upsilon_i^{out} = \Upsilon_i^{out}(\xi_{|i,1|}^{out},\xi_{|i,2|}^{out},\dots,\xi_{|i,n|}^{out}) \in [0,1]$$

และ

$$\Upsilon_i^{in} = \Upsilon_i^{in}(\xi_{|i,1|}^{in},\xi_{|i,2|}^{in},\dots,\xi_{|i,n|}^{in}) \in [0,1]$$

4.4.2 การเปรียบเทียบความเหมือนของภาพ

กำหนดให้ H_1 และ H_2 เป็นกราฟลำดับขั้นที่แทนรูปที่ 2 ตามลำดับและ $v \coloneqq \max(d^{H_1}, d^{H_2})$ โดยที่ $d^{H_i} = |\Im^{H_i}| - 1$ ดังนั้นความเหมือนของกราฟสามารถเขียนเป็นสมการได้ดังนี้

$$s(H_1, H_2) \coloneqq \frac{\prod_{i=0}^{v} \overline{\Upsilon}_i}{\sum_{i=0}^{v} \overline{\Upsilon}_i / v + 1}$$

$$s(H_1, H_2) = s(H_2, H_1),$$

$$s(H_1, H_1) = 1, 0 < s(H_1, H_2) \leq s(H_1, H_1)$$

เมื่อกำหนดให้

$$\overline{\Upsilon}_i = \overline{\Upsilon}\left(\xi^{out}_{|i,1|}, \xi^{out}_{|i,2|}, \dots, \xi^{out}_{|i,n|}, \xi^{in}_{|i,1|}, \xi^{in}_{|i,2|}, \dots, \xi^{in}_{|i,n|}\right);$$

$$\alpha.\Upsilon^{out}_i\left(\xi^{out}_{|i,1|}, \xi^{out}_{|i,2|}, \dots, \xi^{out}_{|i,n|}\right) + (1-\alpha).\Upsilon^{in}_i\left(\xi^{in}_{|i,1|}, \xi^{in}_{|i,2|}, \dots, \xi^{in}_{|i,n|}\right), \alpha \in [0,1]$$

เมื่อมีการกำหนด $\bar{s}(H_1, H_2) \coloneqq \sum_{i=0}^{v} \overline{\Upsilon}_i / (v+1)$ แล้วจากสมการข้างต้นที่กล่าวถึงความสมมาตรกันของกราฟนั้นสามารถเขียนได้ว่า $\bar{s}(H_1, H_2) = \bar{s}(H_2, H_1)$ เมื่อ $\overline{\Upsilon}_i \leq \alpha.1 + (1-\alpha).1 = 1, \bar{s}(H_1, H_2) = \frac{1+1+\cdots+1_{(v+1-times)}}{v+1} = 1.$

5. ผลการทดลอง

ข้อมูลภาพสำหรับการทดลองได้ใช้ฐานข้อมูลที่เก็บรวบรวมและมาจากแอพพลิเคชัน LabelMe [10] และภาพภายนอก (Outdoor) มาจากฐานข้อมูล Corbis คัดเลือกคำศัพท์ภาษาอังกฤษมาทั้งหมด 200 คำหลัก โดยคำศัพท์ที่นำมาใช้จะเป็นคำหลักที่มาจากวัตถุทั่วไปที่มักพบในกลุ่มภาพตามหมวดหมู่ที่สอดคล้องกับภาพที่คัดเลือกมาเป็นหลัก คำหลักที่ถูกสร้างขึ้นนั้นได้พยายามหลีกเลี่ยงคำหลักที่มีความหมายกำกวมและคำหลักที่เป็นทั้งคำเหมือน (Synonym) เช่น “abbey”, “church”, “cathedral” เป็นต้น ดังนั้นในการทดลองนี้จะมีการแท็กคำหลักตามที่กำหนดไว้บนภาพที่ถูกคัดเลือกมาตามความหมายพื้นฐานจากการสุ่มและตัดสินใจจากการจำแนกความหมายภาพด้วยคนเป็นหลัก (Human scenes classification)

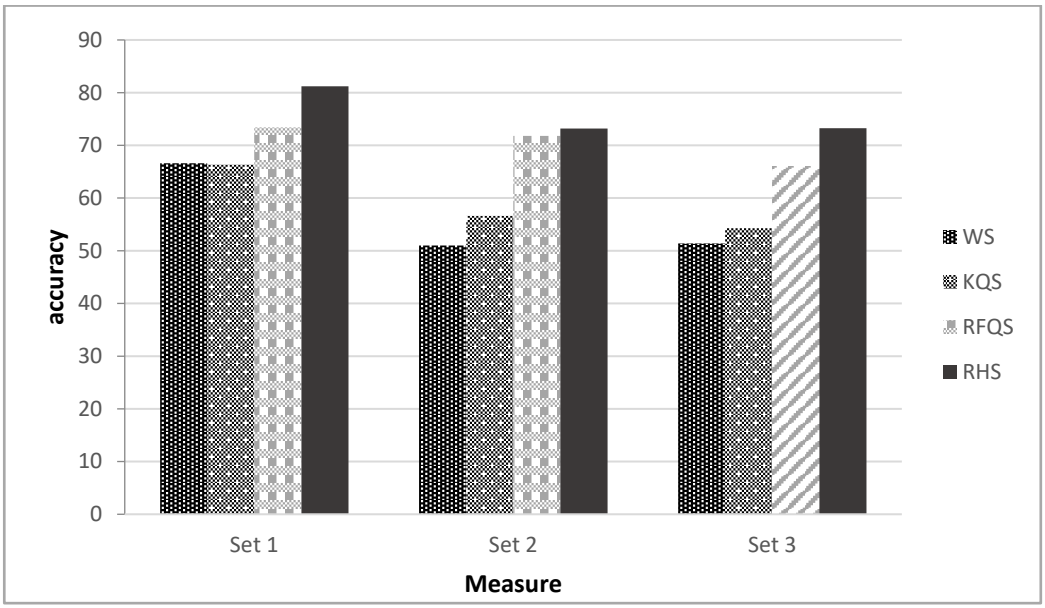
การทดลองได้เปรียบเทียบตามลักษณะของการแทนด้วยโครงสร้างของกราฟที่มีโครงสร้างเบื้องต้นจาก WordNet [9] ซึ่งจะมีส่วนช่วยให้การค้นหาภาพมีความหมายที่ใกล้เคียงกันมากที่สุด ดังนั้นขั้นตอนสุดท้ายของการค้นหาความหมายภาพ ด้วยการเปรียบเทียบความคล้ายข้อมูล ซึ่งในงานวิจัยนี้ได้ใช้วิธีการเปรียบเทียบกันทั้งหมด 4 วิธี การวัดความคล้ายด้วยอนุกรมวิธาน (WordNet Similarity: WS) การวัดความคล้ายด้วยคิวรี่คำศัพท์ (Keyword-based Query Similarity: KQS)

การวัดความคล้ายด้วยคิวรี่ย้อนกลับ (Relevance Feedback based on Query Similarity: RFQS) และนำเสนอวิธีการวัดความคล้ายกันของภาพด้วยโครงสร้างกราฟตามความสัมพันธ์ลำดับขั้น (Semantic Graph Similarity with Relational Hierarchy Structure: RHS) เปรียบเทียบความหมายลงใน 5 กลุ่มเหตุการณ์ประกอบด้วย ภาพพักผ่อนภายนอก (Outdoor leisure), ภาพพิธีการ (Ceremony), ภาพทำงานในสำนักงาน (Office working), ภาพเล่นกีฬา (Sport game) และ ภาพครอบครัว (Family time) การทดลองจะแบ่งชุดข้อมูลภาพเพื่อใช้ทำการทดลองออกเป็น 3 ชุด ด้วยการสุ่มและจัดคำหลักที่มีความสอดคล้องกันตามชุดข้อมูล แบ่งชุดการทดลองออกเป็น 3 ชุด ชุดที่ 1 ใช้ภาพสุ่ม 500 ภาพ คำศัพท์เกี่ยวข้อง 80 คำหลัก ชุดที่ 2 ใช้ภาพสุ่ม 750ภาพ คำศัพท์เกี่ยวข้อง 120 คำหลัก ชุดที่ 3 ใช้ภาพสุ่ม 1,200 ภาพ คำศัพท์เกี่ยวข้อง 180 คำหลัก จากตารางที่ 1

ตารางที่ 1 ผลลัพธ์การเปรียบเทียบความคล้ายกันของความหมายภาพ

Measur	Class	Performance (%)								
		Set 1			Set 2			Set 3		
		Prec.	Recall	F_1	Prec.	Recall	F_1	Prec.	Recall	F_1
WS	Outdoor leisure	69.9	65.0	67.4	54.6	53.0	53.8	55.3	52.0	53.6
	ceremony	69.1	67.0	68.0	50.0	51.0	50.5	54.0	54.0	54.0
	Office indoor	61.5	64.0	62.7	50.9	55.0	52.9	45.3	48.0	46.6
	sport game	67.3	68.0	67.7	47.6	49.0	48.3	50.5	51.5	51.0
	family time	65.7	69.0	67.3	52.2	47.0	49.5	52.5	51.5	52.0
	Accuracy		66.6			51			51.4	
KQS	Outdoor leisure	66.7	64.0	65.3	58.1	54.0	56.0	53.3	57.0	55.1
	ceremony	67.3	68.0	67.7	58.7	54.0	56.3	53.2	58.0	55.5
	Office indoor	66.0	64.0	65.0	54.2	52.0	53.1	54.2	52.0	53.1
	sport game	66.7	68.0	67.3	56.4	62.0	59.0	56.3	53.5	54.8
	family time	65.1	67.6	66.3	56.0	61.0	58.4	54.8	51.0	52.8
	Accuracy		66.3			56.6			54.3	
RFQS	Outdoor leisure	71.2	74.0	72.5	68.6	72.0	70.2	67.0	65.0	66.0
	ceremony	74.5	73.0	73.7	72.1	75.0	73.5	68.3	69.0	68.7
	Office indoor	81.0	68.0	73.9	77.2	71.0	74.0	66.0	64.0	65.0
	sport game	72.6	77.0	74.8	70.6	72.0	71.3	65.4	68.0	66.7
	family time	69.4	75.0	72.1	71.1	69.0	70.1	63.4	64.0	63.7
	Accuracy		73.4			71.8			66.0	
RHS	Outdoor leisure	83.0	78.0	80.4	70.6	72.0	71.3	74.5	76.0	75.2
	ceremony	81.7	76.0	78.8	72.3	73.0	72.6	75.5	76.2	75.9
	Office indoor	80.4	86.0	83.1	81.4	70.0	75.3	72.6	69.0	70.8
	sport game	81.6	84.0	82.8	72.6	77.0	74.8	73.7	73.0	73.4
	family time	79.6	82.0	80.8	70.5	74.0	72.2	69.9	72.0	70.9
	Accuracy		81.2			73.2			73.3	

การเปรียบเทียบโดยใช้ช่วงความลึกของโครงสร้างกราฟถึงคำหลักที่ถูกแท็กไว้บนภาพ จากตารางที่ 1 แสดงให้เห็นว่ากลุ่มภาพ set 1 ที่ถูกเปรียบเทียบความคล้ายของภาพ ด้วย WS จะมีค่าความถูกต้องสูงถึง 66.6% จะเห็นว่าสามารถบอกความหมายของภาพได้ดีกว่า KQS ถึง 0.03% แต่อย่างไรก็ตามเมื่อเปรียบเทียบกับ set 2 และ 3 และ KQS กลับมีความค่าความถูกต้องที่สูงกว่า 4.6% และ 2.9% ตามลำดับ RFQS สามารถเปรียบเทียบความคล้ายของภาพใน set 1 ได้อย่างดีค่าความถูกต้องสูงถึง 73.4% แต่เมื่อ set 2 และ 3 กลับมีค่าความถูกต้องที่ลดลงเมื่อมีจำนวนภาพและจำนวนของคำศัพท์ที่มากขึ้นเช่นเดียวกันกับวิธีการ RHS แต่อย่างไรก็ตามค่าความถูกต้องใน set 1 มีค่าความถูกต้องที่สูงถึง 81.2% และเมื่อใช้กลุ่มข้อมูลของ set 2 และ 3 กลับมีค่าความถูกต้องที่ต่ำลง 73.2% และ 73.3% แต่มีความแตกต่างไม่มากทั้งๆที่จำนวนของภาพใน set 2 และ set 3 มีจำนวนภาพเพิ่มขึ้นเท่าตัวทั้งนี้ขึ้นกับจำนวนของลำดับความสัมพันธ์ที่เกิดขึ้นบนกราฟมีความซับซ้อนแต่ยังสามารถเปรียบเทียบเพื่อค้นหาความหมายของภาพได้อย่างสมบูรณ์จนทำให้มีค่าใกล้เคียงกัน ดังแสดงกราฟผลลัพธ์ในรูปที่ 3



รูปที่ 3 การเปรียบเทียบความคล้ายกันด้วยค่าความถูกต้องระหว่างกลุ่มภาพ

จากการเปรียบเทียบวิธี KQS จะเห็นว่าสามารถเปรียบเทียบความหมายภาพใน set 1 ลงในกลุ่ม ceremony มีค่า Prec. 67.3% Recall 68% และ sport game มีค่า Prec. 66.7% Recall 68% แต่ใน set 1 ที่มีการเปรียบเทียบด้วย RFQS ลงในกลุ่ม Office indoor มีค่า Prec. 81% Recall 68% สำหรับการเปรียบเทียบความหมายใน set 3 ด้วย RHS จะเห็นว่าสามารถวัดความคล้ายของ

ภาพได้สูงถึง 73.3% และลงในกลุ่ม ceremony ได้ค่า F1 ถึง 75.9% เช่นเดียวกันกับวิธีอื่น WS, KQS และ RFQS จะสามารถเปรียบเทียบความคล้ายของภาพในกลุ่ม Outdoor leisure ได้ค่า F1 สูงที่สุด 75.2% ใน set 3 ด้วยวิธีการ RHS แต่ในทางกลับกัน กลุ่มที่มีการเปรียบเทียบได้ค่า F1 ที่น้อยนั้นส่วนใหญ่เป็นกลุ่ม family time ด้วยวิธี RHS จะได้เพียง 70.9% ดังแสดงภาพตัวอย่างผลลัพธ์ในรูปที่ 4

ก. ภาพกลุ่ม outdoor leisure

ข. ภาพกลุ่ม ceremony

ค. ภาพกลุ่ม office indoor

ง. ภาพกลุ่ม sport game

จ. ภาพกลุ่ม family time

รูปที่ 4 ตัวอย่างผลลัพธ์ของการเปรียบเทียบความคล้ายกันของความหมายภาพ

6. สรุปผลการทดลอง

การเปรียบเทียบความคล้ายกันของความหมายภาพเพื่อให้ได้ความหมายของภาพที่อยู่ในหมวดหมู่เดียวกัน โดยปกติทั่วไปนั้นการใช้อัลกอริทึมที่มาสกัดข้อมูลภาพนั้นมักจะใช้สกัดเพียงข้อมูลที่เกิดขึ้นภายในภาพ แล้วนำมาประมวลผลเพื่อใช้ในการสืบค้นข้อมูลภาพ แต่ปัจจุบันได้มีการนำคำหลักที่ได้จากการให้ความหมายของการแท็กวัตถุบนภาพมาหาความสัมพันธ์ภายใน โดยพยายามหาความสัมพันธ์ที่คล้ายกันของวัตถุในหมวดหมู่เดียวกัน ในงานวิจัยนี้ก็เช่นเดียวกันแต่จะเฉพาะเจาะจงลงในกลุ่มของภาพส่วนบุคคลเฉพาะเหตุการณ์กิจกรรม โดยใช้แนวคิดกราฟที่แทนข้อมูลภาพในภาพทั้งหมด จากการทดลองแสดงให้เห็นว่ากลุ่มภาพ set 1 ที่ถูกเปรียบเทียบความคล้ายของภาพ ด้วย WS จะมีค่าความถูกต้องสูงถึง 66.6% RFQS สามารถเปรียบเทียบความคล้ายของภาพใน set 1 ได้อย่างดีค่าความถูกสูงถึง 73.4% แต่เมื่อ set 2 และ 3 กลับมีค่าความถูกต้องที่ลดลงเมื่อมีจำนวนภาพที่มากขึ้นที่ แต่อย่างไรก็ตามค่าความถูกต้องใน set 1 มีค่าความถูกต้องที่สูงถึง 81.2% และเช่นเดียวกันเมื่อใช้กลุ่มข้อมูลของ set 2 และ 3 กลับมีค่าความถูกต้องที่ต่ำลง 73.2% และ 73.3% ลักษณะเหตุผลของค่าความถูกต้องที่ไม่มีการเปลี่ยนแปลงอาจจะเกิดจากความหลากหลายของคำศัพท์ทำให้การเปรียบเทียบตามเส้นการเชื่อมโยงนั้น ยังมีความสัมพันธ์ที่น้อยเกินไป แต่อย่างไรก็ตามค่าความถูกต้องที่ได้จากการจัดกลุ่มด้วย RFQS ได้มากกว่าวิธีการอื่นๆ อัลกอริทึมที่นำเสนอสามารถนำมาประยุกต์ใช้ในการประมวลผลการค้นคืนฐานข้อมูลภาพที่มีขนาดใหญ่ได้ เพื่อทำให้ผลลัพธ์ที่ได้นั้นมีความครบถ้วนสมบูรณ์ และตรงตามความหมายมากยิ่งขึ้น

References

[1] Guang-Hai Liu, etc. (2015). “Content-based image retrieval using computational visual attention model”. **Pattern Recognition.** Vol. 48 (No. 8): Pages 2554-2566.

[2] Xin Lu, Poonam Suryanarayan, Reginald B. Adams, Jr., Jia Li, Michelle G. Newman and James Z. Wang. (2012). “On Shape and the Computability of Emotions”. **Proceedings of the ACM Multimedia Conference.** : Pages 229-238.

[3] Galleguillos C., Belongie S. (2010). “Context Based Object Categorization: A Critical Survey”. **Computer Vision and Image Understanding (CVIU).** Vol. 114 (Issue 6): Pages 712-722.

[4] Tie Hua Zhou, Ling Wang, and Keun Ho Ryu. (2015). “Supporting Keyword Search for Image Retrieval with Integration of Probabilistic Annotation”. **Sustainability.** Vol. 7 (Issue 5): Pages 6303-6320.

[5] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. (2016). “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. **IEEE Transactions on Pattern Analysis and Machine Intelligence.** Vol. 39 (No. 6): Pages1137-1149.

[6] Sun Ting and Geng Guohua. (2016). “Image Retrieval Method for Deep Neural Network”. **International Journal of Signal Processing, Image Processing and Pattern Recognition.** Vol. 9 (No. 7): Pages 33-42.

[7] J. Euzenat, and P. Shvaiko. **Ontology Matching.** 2nd edition. Springer.

[8] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. (2010). “The pascal visual object classes (Voc) challenge”. **International Journal of Computer Vision.** Vol. 88 (No. 2): Pages 303–338.

[9] Patwardhan and Pedersen. (2006). “Using WordNet Based Context Vectors to Estimate the Semantic Relatedness of Concepts”. **Proceedings of the EACL 2006 Workshop Making Sense of Sense - Bringing Computational Linguistics and Psycholinguistics Together.** : Pages 1-8.

[10] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy and William T. Freeman. (2008). “LabelMe: A database and Web-Based Tool for Image Annotation”. **International Journal Computer Vision**. Vol. 77: Pages 157-173.

ประวัติผู้เขียนบทความ

นัศพ์ชาณัณ ชินปัญช์ธนะ อาจารย์ประจำวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์ งานวิจัยทางด้านการประมวลผลภาพดิจิทัลทางด้านความหมายภาพ งานประมวลผลภาพทางด้านโรงงานอุตสาหกรรมการผลิต การจำแนกภาพดิจิทัล การรู้จำภาพดิจิทัล การประมวลผลภาพวิดีโอแบบเรียลไทม์ โทรศัพท์ 0-2954-7300 อีเมล์ nutchanun.cha@dpu.ac.th