

การวัดความคล้ายกันของภาพด้วยการเชื่อมโยงกราฟแบบลำดับชั้นบนพื้นฐานของ

อนุกรมวิธาน WordNet

Similarity Images with Hierarchical Graph based on WordNet Thesaurus

นศพัชราณ ชินปัญญะ

วิทยาลัยนวัตกรรมการเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

ABSTRACT –Image retrieval is active research topic in image processing field. There are many research groups that have attempted to improve new technique for better results. However, the model cannot specific enough for representing the meaning of images. In this paper, we present a novel technique of the similarity images on hieratical graph based on WordNet. The model is divided into 3 steps that included (1) prepared the image data (2) related the relationship on images and (3) evaluated with similarity measure. The experimental results indicate that our proposed approach offers significant performance improvements in the interpretation of semantic image with the maximum of 85.7%

KEY WORDS – image retrieval, image processing, semantic image, hierarchical graph, similarity measure

บทคัดย่อ -- การค้นคืนภาพดิจิทัลเป็นหัวข้องานวิจัยที่ยังคงท้าทายอย่างมากในสาขาการประมวลผลภาพ มีนักวิจัยหลายกลุ่มพยายามคิดค้นและประยุกต์วิธีการต่างๆเพื่อให้ได้ผลลัพธ์ที่ได้จากการค้นคืนดียิ่งขึ้นตรงตามความต้องการมากขึ้น แต่อย่างไรวิธีการเหล่านั้นไม่สามารถที่จะนำมาใช้แทนความหมายของภาพได้อย่างแท้จริง ดังนั้นในงานวิจัยชิ้นนี้ได้นำเสนอแนวทางการวัดความคล้ายกันของภาพด้วยวิธีย้อนกลับบนพื้นฐานการเชื่อมโยงแบบลำดับชั้นบนอนุกรมวิธาน WordNet สำหรับการจัดกลุ่มความหมายภาพ แบ่งขั้นตอนการทำงานออกเป็น 3 ส่วนดังนี้ (1) ขั้นตอนการเตรียมข้อมูล (2) การเชื่อมโยงความสัมพันธ์ข้อมูลภาพ และ (3) การวัดความคล้ายกันของภาพ จากวิธีการที่นำเสนอนี้สามารถวัดความคล้ายกันได้ค่าความถูกต้องสูงถึง 85.7%

คำสำคัญ – การค้นคืนภาพ; การประมวลผลภาพ; ความหมายภาพ; กราฟลำดับชั้น; การวัดความหมายภาพ;

1. บทนำและงานวิจัยที่เกี่ยวข้อง

วิวัฒนาการของเครื่องมือและอุปกรณ์อิเล็กทรอนิกส์สำหรับการถ่ายภาพดิจิทัลได้พัฒนาอย่างรวดเร็ว จนทำให้ภาพถ่ายภาพดิจิทัลมีจำนวนมากมาย และแพร่กระจายอยู่ทั่วไปไม่ว่าจะอยู่ในสังคมออนไลน์อย่าง Facebook Google หรือ Instagram เป็นต้น ทำให้เกิดปัญหาที่ตามมาคือการจัดเก็บข้อมูลภาพที่เพิ่มขึ้น

การจัดเก็บอย่างไรให้มีระบบที่ดีสามารถสืบค้นข้อมูลภาพได้ง่ายและจำแนกข้อมูลภาพให้ตรงตามความหมายของภาพที่ต้องการของผู้ใช้มากที่สุด ทำให้งานวิจัยในปัจจุบันที่เกี่ยวข้องกับการค้นคืนรวมทั้งการจัดกลุ่มภาพให้ตรงกับความ ต้องการได้รับความสนใจจากนักวิจัยหลายกลุ่มเป็นงานด้านการประมวลผลภาพ (image processing) ด้านการค้นคืนสารสนเทศ (image retrieval) และการจำแนกประเภท



รูปที่ 1. ขั้นตอนการวัดความคล้ายกันของกลุ่มความหมายภาพ

ข้อมูลภาพ (image classification) เพื่อคัดเลือกภาพ ให้ตรงตามความต้องการของผู้ใช้งาน สำหรับงานวิจัยทางการประมวลผลภาพในการค้นคืนสารสนเทศ จะมีการค้นคืนตามคุณลักษณะพื้นฐานของภาพที่ถูกสกัดคุณลักษณะด้วยอัลกอริทึมต่างๆ เช่น สี (color) ลวดลาย (texture) รูปทรง (shape) เป็นต้น [1][2][3] งานวิจัยของ SIMPLiCity [4] เป็นการใช้อนุลักษณะระดับต่ำด้วยสี ลวดลาย และตำแหน่งของพื้นที่ของภาพ จากผลลัพธ์จะสังเกตว่าผลลัพธ์ของภาพเป็นภาพที่มีโทนสีคล้ายกันเป็นหลัก แต่มีลักษณะวัตถุที่แตกต่างกันอย่างเห็นได้ชัดเจนการประมวลผลภาพระดับต่ำนั้นค่อนข้างยากที่จะจัดให้หมวดหมู่เดียวกัน แต่มีกลุ่มนักวิจัยที่พยายามปรับปรุงอัลกอริทึม เพื่อทำการค้นคืนภาพที่มีลักษณะฟีเจอร์ที่ใกล้เคียงกับภาพที่ต้องการมาก การปรับปรุงเทคนิค ด้วยการนำวิธีการมาผสมผสานกันระหว่างคุณลักษณะเพื่อให้สามารถวิเคราะห์ในรูปแบบที่ซับซ้อน แต่อย่างไรก็ตามในการค้นคืนที่ใช้คุณลักษณะฟีเจอร์ระดับต่ำเพียงอย่างเดียว ทำให้ได้ผลลัพธ์ส่วนใหญ่ตรงกับคุณลักษณะของฟีเจอร์ที่สกัดมาแต่ไม่ได้ตรงกับความหมายที่เกิดภายในภาพที่ต้องการอย่างแท้จริง

งานวิจัยอีกกลุ่มที่พยายามจะใช้เทคนิคของการเข้าใจความหมายของภาพแทน การสืบค้นแบบข้างต้น งานวิจัยในกลุ่มนี้พยายามที่จะมองข้อมูลบนภาพเป็นวัตถุ (object) [5] ที่มีความหมายและแทนวัตถุนั้นๆ ด้วยคำหลัก (keyword) บนภาพเรียกการแทนข้อมูลบนภาพแบบนี้ว่า Keyword-based image indexing [6] เป็นการให้ความหมายของวัตถุนั้นๆ เป็นชื่อวัตถุ หรือคำศัพท์ ที่สอดคล้องกัน เช่น “grass” , “plant” , “boat”, “sky” เป็นต้น หรือแสดงเนื้อหาภาพเป็นคำอธิบาย สิ่งต่างๆที่เกี่ยวข้องกับภาพ เช่น ชื่อเรื่อง, ผู้ถ่ายภาพ, รายละเอียด

ภาพ เป็นต้น ส่วนใหญ่ในการให้ความหมายหรือการเขียนคำอธิบายภาพจะเป็นลักษณะของอัตโนมัติบนภาพหรือใช้การคัดเลือกคำศัพท์ด้วยมนุษย์ ขึ้นกับระบบที่เลือกใช้งาน ในปัจจุบันระบบที่ใช้ในการให้ความหมายภาพ (annotation) จะสามารถให้ความหมายวัตถุนั้นภาพได้อย่างง่ายดายและรวดเร็ว ตัวอย่างระบบ เช่น LabelMe[7] Peekaboom[8] เป็นต้น การให้ความหมายของภาพด้วยคำศัพท์ต่างๆเพื่อทำให้การสืบค้นข้อมูลได้ผลดีขึ้นจากวิธีแรกที่ใช้เพียงคุณลักษณะต่ำเพียงอย่างเดียว แต่ยังคงมีปัญหาที่ตามมาจากผลของการค้นคืนเพื่อให้ถูกต้องและตรงตามความหมายมากยิ่งขึ้น การใช้หลายคำศัพท์ (multiple label) [9][10] เป็นอีกวิธีการหนึ่งที่ใช้ช่วยการค้นคืน แต่ก็ยังถูกจำกัดด้วยการค้นหาเพื่อให้ได้ภาพที่ตรงกับความหมายภาพที่เพิ่มขึ้น นักวิจัยบางกลุ่ม [11] พยายามใช้มิติของความสัมพันธ์ของคำศัพท์เพื่อหาแตกต่างกัน เช่น “mango” กับ “fruit” “mobile phone” กับ “electronic device” เป็นต้น การให้ความหมายเพียงผิวเผิน หรือความหมายที่ตรง กับวัตถุจริงหรือเพียงกลุ่มของวัตถุเท่านั้นจะมีผลต่อการค้นคืนภาพรวมทั้งมีผลต่อการแปลความหมายภาพด้วยเช่นกัน ดังนั้นในงานวิจัยนี้จึงได้นำเสนอวิธีการค้นคืนกลุ่มภาพและแทนข้อมูลความสัมพันธ์ภาพนั้นด้วยรูปแบบอนุกรมวิธาน WordNet [12][13] เป็นฐานข้อมูลพจนานุกรมภาษาอังกฤษที่มีโครงสร้างแบบลำดับชั้น และสร้างความสัมพันธ์ในเชิงความหมายเหมือนและความหมายต่าง รวมทั้งมีการจัดกลุ่มความหมายของคำตามลำดับชั้นในเชิงโครงสร้างต้นไม้ ซึ่งการจัดลำดับความสัมพันธ์ลักษณะนี้ถูกเรียกว่า Synset ตัวอย่างเช่นความสัมพันธ์ของความหมายของการแทนคำ ร่วมกับวิธีการ Kernel Canonical Correlation Analysis เพื่อสร้างความสัมพันธ์ของคุณลักษณะคำศัพท์ที่เกิดขึ้น M. Johnson and R. Cipolla

[10] ได้พยายามจัดกลุ่มภาพที่มีวัตถุเหมือนกันอยู่ในกลุ่มเดียวกันแทนที่จะเป็นการแยกทีละภาพ I. Simon และคณะ [14] ทำการจัดกลุ่มภาพในลักษณะเดียวกันโดยที่จะใช้ข้อมูลสถานที่ (location) ของภาพ และการจัดกลุ่มตามชื่อเรื่อง (topic) หรือภาพที่เกิดขึ้นเป็นภาพธรรมชาติ จัดกลุ่มตาม เหตุการณ์ของธรรมชาติ (natural event) แต่บางงานวิจัยจะมีการใช้วิธีการเรียนรู้ตามแม่แบบ (learning concept templates) [15] จัดกลุ่มภาพและเพิ่มส่วนของการตอบสนองกลับ (relevance feedback) เพื่อช่วยการการแก้ไขกลุ่มของภาพในการค้นคืนให้ได้ตรงตามกลุ่มเป้าหมายที่สร้างขึ้น Query-Frequency Pair [5] เป็นอีกวิธีการหนึ่งที่พยายามค้นคืนภาพด้วยคิวรี (query) ที่มีความซับซ้อนประกอบด้วยคำศัพท์ที่มีการบันทึกไว้ ผลลัพธ์จะได้ผลเมื่อการค้นคืนใช้รูปแบบของคำศัพท์ในแบบเดียวกัน แต่เมื่อมีความแตกต่างของการให้ความหมายคำศัพท์ผลที่ได้จะคิดไปเพื่อให้ผลลัพธ์ที่ได้มีความถูกต้องมากยิ่งขึ้น การค้นหภาพด้วยเทคนิคที่กล่าวมานี้จะได้ผลลัพธ์ที่ขึ้นกับคำศัพท์ที่ถูกให้ความหมายไว้บนภาพยังมีการแท็กข้อมูลบนภาพมากยิ่งขึ้นสามารถหาความเหมือนกันบนภาพมากขึ้นเท่านั้น แต่เมื่อการให้ความหมายของวัตถุบนภาพมาจากหลายที่ หลายแหล่งข้อมูล หรือหลายผู้บันทึกข้อมูลทำให้คำศัพท์บางคำมีความหมายที่เหมือนกันแต่ต่างคำกัน เช่น “sky” กับ “trees”, “car” กับ “vehicle” เป็นต้น หรือคำศัพท์บางคำที่ให้ความหมายได้ลึก “right” และ “ripe” เป็น Synset เดียวกัน เพราะมีความหมายใกล้เคียงกันเป็นต้น

2. ขั้นตอนการประมวลผลภาพ

การเชื่อมโยงความสัมพันธ์ของข้อมูลเพื่อค้นคืนความหมายภาพเริ่มจากการเก็บรวบรวมข้อมูลภาพและคัดเลือกภาพที่เหมาะสมเพื่อเตรียมเป็นข้อมูลภาพเบื้องต้น ดังนั้นข้อมูลภาพที่เตรียมพร้อมจะสามารถเข้าสู่กระบวนการประมวลผลภาพ โดยในงานวิจัยจะมีการแบ่งขั้นตอน วิธีการดำเนินงานวิจัยแบ่งออกเป็น 3 ส่วนหลักโดยขั้นตอนทั้งหมดจะสามารถอธิบายได้ดังรูปที่ 1 ดังนี้

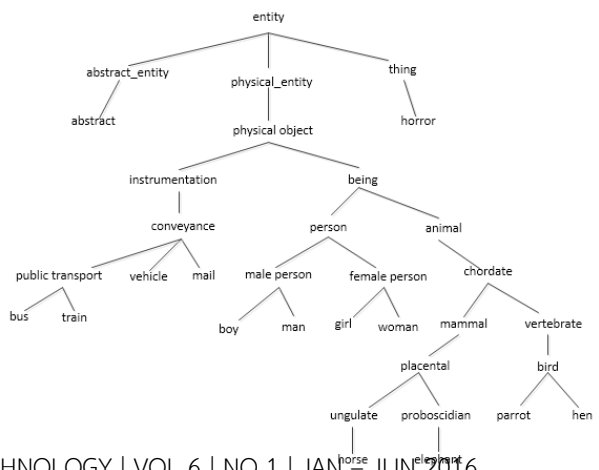
2.1 การเตรียมข้อมูล

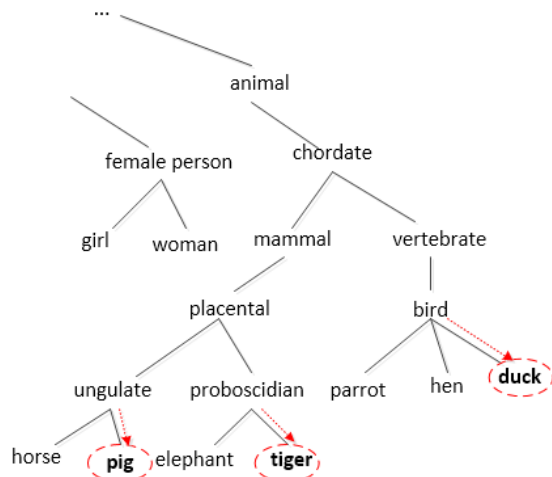
ขั้นตอนการเตรียมข้อมูล (data preprocessing) เริ่มจากคัดเลือกข้อมูลภาพดิจิทัลที่มีวัตถุบนภาพที่เด่นชัด มีวัตถุภาพพื้นหลัง และภาพที่คัดเลือกเข้ามานั้นสามารถให้มนุษย์แปลความหมาย

ภาพนั้นได้อย่างสมบูรณ์ ภาพบางภาพมีลักษณะผิดปกติ (outlier) หรือ คุณลักษณะวัตถุ (object characteristic) ไม่ชัดเจนคลุมเครือ มีขนาดวัตถุขนาดเล็กเกินไปไม่สามารถ บ่งชี้ชื่อวัตถุได้ ภาพถ่ายระยะใกล้ (close up) ทำให้ต้องมีการคัดเลือกภาพออกไปไม่นำมาใช้ในการทดลอง สำหรับภาพบางภาพจะไม่นำเข้ามาทำการทดลองนั้นจะเป็นภาพที่มีความหมายกำกวมภาพไม่มีความหมาย แปลความหมายไม่ได้ หรือภาพที่มนุษย์แปลได้หลายความหมาย ภาพที่มีการโฟกัสระยะใกล้ ข้อมูลภาพที่มีความซ้ำซ้อน หรือไม่สอดคล้องกันจะไม่ถูกนำมาพิจารณา การแท็กข้อมูลภายในภาพเป็นการทำงานในส่วนของการนำข้อมูลเข้าด้วยเครื่องมือ (image annotation tool) ใช้ฐานข้อมูลภาพจาก LabelMe [7]

2.2 การเชื่อมโยงความสัมพันธ์ข้อมูลภาพ

การเชื่อมโยงความสัมพันธ์ข้อมูลภาพ ด้วยกราฟลำดับชั้นข้อมูลภาพได้จากการแท็กของผู้ใช้ออนไลน์ ที่มีผู้ใช้ที่แตกต่างกันทำให้เกิดความแตกต่างของคำศัพท์บ้างเล็กน้อย กับการแท็กวัตถุประเภทเดียวกัน แต่มีการแท็กคำที่ต่างกัน (เช่น คำศัพท์ “stone” หรือ “rock” มีความหมายใกล้เคียงกัน) เมื่อมีการค้นคืนด้วยคำศัพท์จะถูกสร้างความแตกต่างของภาพ แต่ในความเป็นจริงแล้วมีความหมายที่ใกล้เคียงกัน ดังนั้นจึงทำการลดความซ้ำซ้อนหรือกำกวมของคำศัพท์ที่เกิดขึ้นด้วยการใช้รูปแบบความสัมพันธ์ของคำศัพท์ WordNet [12] ดังนั้นเมื่อคำศัพท์สองคำหรือมากกว่ามีความหมายเหมือนกัน (synonym) จะถูกแทนกล่าวด้วยความเหมือนได้ด้วยคำความหมายเดียวกัน ในรูปแบบความสัมพันธ์ Synset ตามอนุกรมวิธาน (taxonomy) ของ WordNet และขั้นตอนสุดท้ายขั้นตอนการประมวลผล (data processing) ด้วยการวัดความคล้ายแบบกราฟลำดับชั้น กล่าวในลำดับถัดไป





รูปที่ 3. ตัวอย่างการแทรกคำศัพท์ลงใน HSG

3. แบบจำลองความสัมพันธ์โครงสร้างกราฟลำดับขั้น

แบบจำลองความสัมพันธ์โครงสร้างกราฟลำดับขั้นบนพื้นฐานของอนุกรมวิธาน WordNet และเรียกกราฟลำดับขั้นที่มีความสัมพันธ์นี้ว่า (Hierarchical Synset Graph :HSG) กำหนดให้ HSG เป็นแนวคิดโครงสร้างลำดับขั้นที่ประกอบด้วย โหนดเริ่มต้นดังนี้ event, physical object, thing, instrument, action, living form, object, place, event, phenomenon, group, possession and condition และทำการเพิ่มโหนดต่างๆต่อจาก โหนดรากตามคุณสมบัติของ Offspring นั้นๆ โดยที่เมื่อ HSG มีคำศัพท์ใหม่ที่เกิดขึ้นจากการแก้ไขบนภาพจะมีการตรวจสอบตามคุณสมบัติจากโครงสร้างของ WordNet เพื่อให้ให้อยู่ในตำแหน่งที่เหมาะสมบน HSG ดังแสดงในรูปที่ 2 ส่วนอัลกอริทึมแสดงรายละเอียดด้านล่าง การแทรกด้วยอัลกอริทึมด้านบนจะเป็นการแทรกข้อมูลบนโครงสร้างของ WordNet เมื่อกำหนดให้ คำศัพท์หลักแสดงในกรอบสี่เหลี่ยม เมื่อมีคำศัพท์ใหม่ที่เข้ามาบน HSG ที่โหนดราก physical object ดังนี้ duck, pig และ tiger ผลลัพธ์ที่ได้จากการแทรกข้อมูลดังแสดงในรูปที่ 3 จะเห็นว่าคำศัพท์ที่แทรกเข้าไปตามลำดับของโครงสร้าง WordNet ซึ่ง HSG ที่ถูกนำมาใช้นั้นจะสร้างความสัมพันธ์เพียงส่วนหนึ่งของโครงสร้าง WordNet เท่านั้น ไม่สามารถแทนความสัมพันธ์ของ WordNet ทั้งหมดได้ ความสัมพันธ์ที่เกิดขึ้นจะเป็นความสัมพันธ์ของโครงสร้างกลุ่มความหมายที่

เหมือนกันและ กลุ่มเดียวกันที่มีการตกทอดของความสัมพันธ์ที่ต่อเนื่องกันเท่านั้น โดยแสดงความสัมพันธ์ไว้ดังนี้

Algorithm : Insertion Synset Keyword on HSG

กำหนดให้ S_n เป็นคำศัพท์, W เป็นกราฟ

Step 1: ค้นหา ความสัมพันธ์ของคำศัพท์ Synset เป็น S_n ใน W บน WordNet.

Step 2: เริ่มต้น S_n และทำการเดินจากล่างไปบนตามโครงสร้างบน WordNet จนถึงโหนดบรรพบุรุษแรกของ S_n และทำการออกจาก HSG และเปลี่ยน Synset เป็น S_a

Step 3: กำหนดให้ $\{S_1, S_2, \dots, S_m\}$ เป็นโหนดลูกของ S_a ใน HSG เมื่อให้ m แทนจำนวนของโหนดลูกใน S_a

For $i = 1$ **to** m

ค้นหาโหนดบรรพบุรุษในระดับที่ต่ำที่สุดที่มี Synset

S_{co_a} ของทั้ง S_i และ S_n บนโครงสร้างของ WordNet

If $S_{co_a} \neq S_a$ **ไปยัง** Step 5.

End For

Step 4: เพิ่มโหนด S_n ลงใน HSG เป็นโหนดลูกของ S_a

ออกจากอัลกอริทึม

Step 5: เพิ่มโหนด S_n ลงใน HSG เป็นโหนดลูกของ S_a

ให้ S_i เป็นโหนดลูกของ S_a แล้วทำการเพิ่มโหนด S_{co_a}

If $S_n \neq S_{aco_a}$ แล้วทำการเพิ่ม S_n ลงใน HSG เป็น

โหนดลูกของ S_{aco_a}

ออกจากอัลกอริทึม

4. การวัดความคล้ายกันของกลุ่มความหมายภาพ

โดยทั่วไปวิธีการวัดความคล้ายกันของภาพมักจะเปรียบเทียบความคล้ายกันด้วยคุณสมบัติของข้อมูลที่เหมือนกัน หรือใกล้เคียงกันตามคุณลักษณะที่มีการบันทึกไว้ เช่น ตัวเลข หรือ คำศัพท์ เท่านั้น เช่นเดียวกันกับการวัดความคล้ายของภาพที่เก็บข้อมูลคำศัพท์ที่ถูกแก้ไขบนภาพแบบคำต่อคำ (Word by Word Similarity: WWS) โดยจะทำการเปรียบเทียบเฉพาะคำที่ปรากฏบนภาพที่มีรูปแบบเหมือนกันเท่านั้น ดังนั้นถ้าจำนวนคำศัพท์ที่แก้ไขบนภาพตรงกับคำศัพท์ของความถี่การใช้งานมากเท่าไรก็แสดงว่ามีความคล้ายกันของภาพมากเท่านั้น แต่อย่างไรก็ตามเป็นวิธีการหาภาพที่เหมือนกันเบื้องต้นได้แต่อาจจะเกิดความผิดพลาดได้ถ้ามีการแก้ไขคำศัพท์ที่มีความหมายเหมือนกันเกิดขึ้น ดังนั้นจึงได้มีการวัดความคล้ายกันของภาพตามความเหมือนกันของลำดับขั้นที่ถูกอ้างอิงด้วย WordNet จะแบ่งเป็น 2 ขั้นตอนดังนี้

4.1 การเตรียมข้อมูล

การวัดความคล้ายด้วยกราฟลำดับชั้น (Hierarchical Graph Similarity: HGS) ลำดับแรกจะมีการกำหนดความเหมือนกันของคำศัพท์ที่อยู่บนพื้นฐานของอนุกรมวิธาน WordNet ที่มีความสัมพันธ์ของคำศัพท์แบบลำดับชั้น โดยที่จะมีการเลือกใช้ในส่วนเฉพาะโครงสร้างของ HSG ที่มีสัดส่วนของเฉพาะข้อมูลคำศัพท์ที่มีในฐานข้อมูลภาพเท่านั้น ทำให้มีการเปรียบเทียบที่จำกัดลง 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 ได้ดังแสดงในสมการดังนี้

$$Sim(s_1, s_2) = depth(s_a) / depth_{max} \quad (1)$$

4.2 การเตรียมข้อมูล

การวัดความคล้ายด้วยคำศัพท์ (Keyword-based Query Similarity: KQS) จะเป็นการวัดความคล้ายแบบคำต่อคำที่มีการกำหนดค่าถ่วงน้ำหนักภายในคำศัพท์ที่ถูกแท็กเพื่อใช้เป็นเครื่องมือวัดรูปแบบหนึ่งระหว่างคำศัพท์และภาพ เมื่อกำหนดให้ $\rho = \{ \langle k_1, w_{k_1} \rangle, \langle k_2, w_{k_2} \rangle, \dots, \langle k_m, w_{k_m} \rangle \}$ เมื่อกำหนดให้ k_i แทนคำศัพท์หรือ Synset ที่ i ที่มีการแท็กบนภาพด้วยค่าถ่วงน้ำหนักมีค่าเป็น w_{k_i} ที่ i คำศัพท์ของการวัดความคล้ายสามารถเขียนได้ $Q = \{ \langle q_1, w_{q_1} \rangle, \langle q_2, w_{q_2} \rangle, \dots, \langle q_m, w_{q_m} \rangle \}$ ทุกคำศัพท์เมื่อผู้ใช้งานมีการค้นหาจะมีค่าถ่วงน้ำหนักเริ่มต้นคงเดิม แต่เมื่อมีการโต้ตอบของผู้ใช้งานเกิดขึ้นจะมีการปรับค่าถ่วงน้ำหนักไป ซึ่งสมการของการปรับเปลี่ยนค่าถ่วงน้ำหนักสามารถเขียนได้ดังนี้

$$SimF(Q, \rho) = \sum_{i=1}^n \max_{j=1 \dots m} [sim(q_i, k_j) \cdot w_{q_i} \cdot w_{k_j}] / n \quad (2)$$

เมื่อกำหนดให้ $sim(q_i, k_j)$ เป็นการวัดความคล้ายระหว่าง q_i และ k_j ใช้การคำนวณจากสมการที่ (1) จากสมการที่ (2) จะเป็นการค้นหาคำศัพท์ของแต่ละคำศัพท์และมีการคำนวณค่าเฉลี่ยที่ได้จากค่าสูงสุดของการวัดค่าความคล้ายระหว่างคำศัพท์

และภาพค่าความคล้ายกันของค่าน้ำหนักของคำศัพท์ w_{q_i} และ w_{k_j} ค่ามากจะแสดงถึงความเหมือนกันของคำศัพท์

4.2 การวัดความคล้ายด้วยคำศัพท์ย้อนกลับ

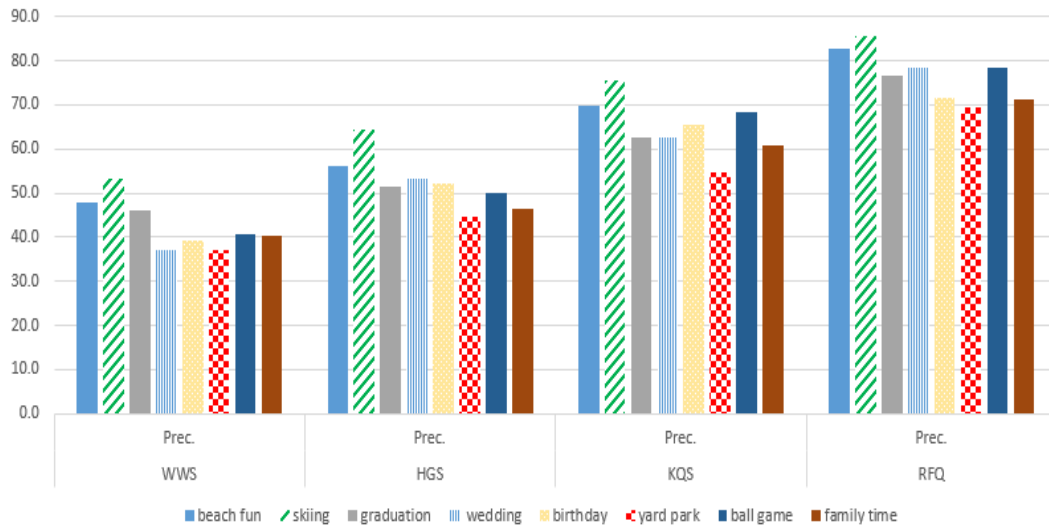
การวัดความคล้ายด้วยคำศัพท์ (Keyword-based Query Similarity :KQS) จากการสร้างแบบจำลองความสัมพันธ์โครงสร้างกราฟลำดับชั้น HSG เมื่อมีการเพิ่มคำศัพท์ลงบน HSG จะมีการเพิ่มโหนดตามแนวคิดของ Synset บน WordNet ลงบนกราฟที่กำหนดขึ้น และจะสร้างตำแหน่งและที่ตั้งที่มีความสอดคล้องกันบนโครงสร้าง HSG ดังนั้นในการใช้คำศัพท์ในการวัดความคล้ายด้วยคำศัพท์เพื่อค้นหาภาพที่คล้ายกัน ดังสมการที่ (2) ผลที่ได้จากการค้นคืนด้วยคำศัพท์ จะได้ผลย้อนกลับ (relevance feedback) ที่เป็นทั้งภาพคล้ายกันและภาพต่างกันตามระดับลักษณะของภาพ ดังนั้นในการนำผลที่ได้จากการค้นคืนด้วยคำศัพท์หลายครั้งมารวมกันเพื่อให้ได้ผลลัพธ์ที่ใกล้เคียงมากที่สุด เรียกวิธีการนี้ว่าการวัดความคล้ายด้วยคำศัพท์ย้อนกลับ (Relevance Feedback based on Query Expansion: RFQ) เมื่อทุกคำศัพท์ที่ถูกแท็กบนตัวอย่างที่ถูกต้องจะเป็นคำศัพท์บวก (positive query: Q_p) ด้วยค่าน้ำหนักที่เกิดขึ้นจากแต่ละการแท็ก และ Q_n เป็นคำศัพท์ที่ถูกสร้างจากการย้อนกลับแบบลบ (negative query) ดังนั้นการสร้างค่าความคล้ายสุดท้ายของ S_i เมื่อ i เป็นลำดับภาพบนฐานข้อมูลสามารถเขียนเป็นสมการได้ดังนี้

$$S_i = sim(Q, \rho_i) + \delta (1 + sim(Q_p, \rho_i)) \sum_{k \in N_p} S_{ik} + \beta (1 + 1 + sim(Q_n, \rho_i)) \sum_{k \in N_n} S_{ik} \quad (3)$$

เมื่อกำหนดให้ δ และ β เป็นค่าคงที่ และ S_{ik} เป็นค่าความคล้ายระหว่างภาพที่ i กับคำศัพท์ตัวอย่างย้อนกลับที่ k , N_p และ N_n เป็นจำนวนรวมของคำศัพท์ย้อนกลับทั้งบวกและลบ

4.4 การวัดประสิทธิภาพ

การวัดประสิทธิภาพ (Evaluation) เป็นขั้นตอนสุดท้ายเพื่อทำการตรวจสอบวิธีการที่ทำการทดลองมาข้างต้นว่ามีประสิทธิภาพมากหรือน้อยเพียงใดเมื่อนำมาใช้งานจริง จะเป็นขั้นตอนที่สำคัญเพราะการนำวิธีการที่นำเสนอไปข้างต้นมาใช้งานได้นั้นจะต้องสอดคล้องกับความต้องการ จึงต้องมีการทดสอบศักยภาพการนำไปใช้ สถาปัตยกรรมที่ใช้วัดความสำเร็จหลังการนำไปใช้หากนำไปใช้แล้วไม่ประสบ



รูปที่ 4. กราฟแสดงการเปรียบเทียบผลการค้นคืนตามกลุ่มความหมายภาพด้วยค่าความแม่นยำ(%)

ผลสำเร็จต้องย้อนกลับ ไปเริ่มกระบวนการแรกใหม่ จึงต้องมีการประเมินผลก่อนการใช้งาน ในการประเมินนั้นกระทำได้ โดยการวัดประสิทธิภาพของการจัดกลุ่มภาพมักจะถูกพิจารณาเป็นค่าของความถูกต้องของแต่ละกลุ่มข้อมูลซึ่งจะประกอบด้วย การวัดค่าความแม่นยำ ค่าความระลึก และ ค่าความถูกต้อง

ค่าความแม่นยำ (False positive rate / Precision: Pr) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพทั้งหมดที่ทำการค้นหามาได้

$$Pr = \frac{\varepsilon}{\partial} \quad (4)$$

เมื่อกำหนดให้ ε แทนจำนวนข้อมูลที่เกี่ยวข้องและถูกดึงออกมาได้อย่างถูกต้อง และ ∂ แทนจำนวนข้อมูลที่ถูกดึงออกมาทั้งหมด ค่าความระลึก (True positive rate / Recall: Re) เป็นอัตราส่วนของการค้นพบภาพที่ถูกต้องจากจำนวนภาพที่ถูกต้องทั้งหมด

$$Re = \frac{\varepsilon}{\delta} \quad (5)$$

เมื่อ δ แทนจำนวนข้อมูลที่เกี่ยวข้องทั้งหมด

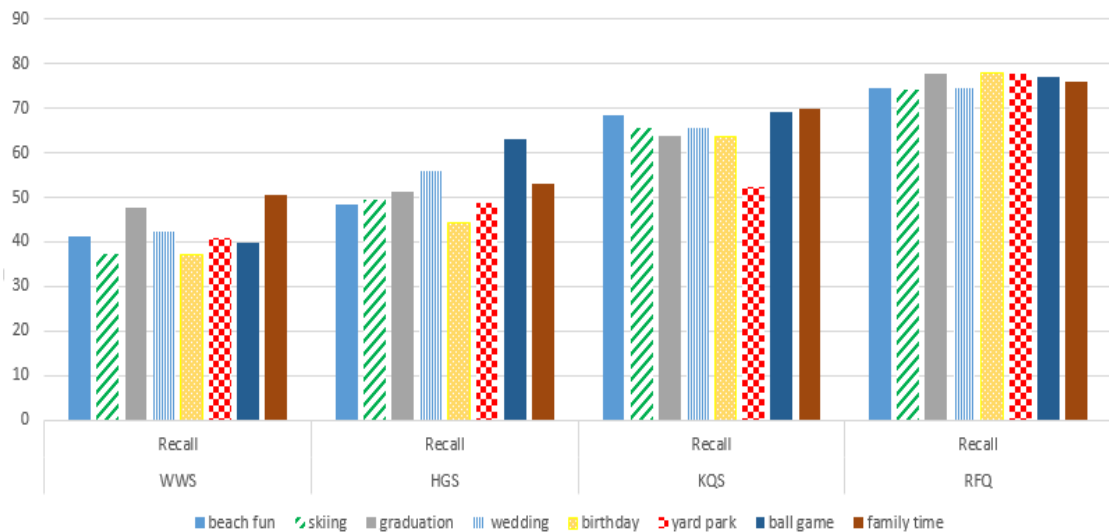
5. การวัดความคล้ายกันของกลุ่มความหมายภาพ

โดยทั่วไปสำหรับการทดลองได้ใช้ฐานข้อมูลที่เกี่ยวข้องและมาจากแอปพลิเคชัน LabelMe [7] โดยแอปพลิเคชันนี้สามารถทำงานได้อย่างเต็มรูปแบบบนเว็บในลักษณะของเครื่องมือให้ความหมาย และได้มีการคัดเลือกภาพสำหรับการทดลองให้ครอบคลุมกับการทดลองโดยเป็นภาพส่วนบุคคลที่อยู่ในหมวดหมู่ของภาพภายใน (indoor) และภาพภายนอก

(outdoor) มาจากฐานข้อมูล LabelMe [7] และคำศัพท์ที่ได้อ้างอิงจากฐานข้อมูลอนุกรมวิธาน WordNet [12][13] คัดเลือกคำศัพท์มาทั้งหมด 95 คำหลักที่เป็นภาษาอังกฤษ โดยคำหลักที่นำมาใช้จะเป็นคำหลักที่มาจากวัตถุทั่วไปที่มักพบในกลุ่มความหมายภาพเกิดเพื่อให้สอดคล้องกับแอปพลิเคชันที่ใช้งาน

การวัดความหมายภาพจะใช้ความสัมพันธ์ที่ได้จาก HSG ของตำแหน่งคำศัพท์บนกราฟ นำมาใช้วัดความคล้ายกันของความหมายที่ได้ตามอนุกรมวิธาน WordNet เพื่อทำให้ได้ผลลัพธ์ของความหมายภาพอย่างแท้จริง โดยทำการเปรียบเทียบความคล้ายกันของความหมายภาพด้วย การวัดความคล้ายกันด้วยคำศัพท์ต่อคำศัพท์ (WWS), การวัดความคล้ายด้วยกราฟลำดับชั้น (HGS), การวัดความคล้ายด้วยคิวรี (KQS) และการวัดความคล้ายกันด้วยคิวรีย้อนกลับ (RFQ) ลงใน 9 กลุ่มเหตุการณ์ ประกอบด้วย beach fun, skiing, graduation, wedding, birthday party, Christmas, yard park, ball game และ family time แสดงภาพตัวอย่างข้อมูลในรูปแบบที่ 5

จากการทดลองการวัดความหมายภาพสามารถสรุปได้ว่า การวัดความคล้ายกันด้วยวิธี WWS เป็นการเปรียบเทียบแบบคำต่อคำที่เกิดขึ้นบนภาพจะเห็นว่ามีค่าความถูกต้องรวมของทุกกลุ่มความหมายเพียง 46.8% ซึ่งมีค่าที่ต่ำที่สุดของการวัดความคล้ายของภาพและ 58%, 73% และ 85.7% สำหรับการวัดความคล้ายด้วยการใช้อนุกรมวิธาน WordNet HSG, KQS และ RFQ



รูปที่ 5. กราฟแสดงการเปรียบเทียบผลการค้นคืนตามกลุ่มความหมายด้วยค่าความระลึก (%)

ตามลำดับ เมื่อทำการเปรียบเทียบการวัดความคล้ายกันตามกลุ่มความหมายภาพแล้วจะเห็นว่า ค่าความแม่นยำในกลุ่มความหมายภาพ skiing มีค่าเปลี่ยนแปลงของค่าความแม่นยำที่เพิ่มขึ้นอย่างต่อเนื่องดังนี้ 53.2%, 64.4%, 75.5% และ 85.6% สำหรับค่าความระลึกเป็น 37.2%, 49.6%, 65.5% และ 74.1% ของการวัดแบบ WWS, HGS, KQS และ RFQ ตามลำดับ จะเห็นว่าการวัดแบบสุดท้าย RFQ ที่มีการใช้คิวรีย้อนกลับสามารถได้ค่าความแม่นยำสูงถึง 85.6% และค่าความระลึก 74.1% แต่เมื่อเปรียบเทียบกับกลุ่มความหมายของ family time จะเห็นว่าการวัดความคล้ายด้วย WWS และ HGS ได้ค่าความแม่นยำเป็น 40.3%, 46.5% ตามลำดับ จะเห็นว่าเพิ่มขึ้นเพียง 6.2% เท่านั้น 50.5% 53.1% และเมื่อใช้การวัดความคล้ายแบบ KQS และ RFQ จะได้ค่าความแม่นยำเป็น 60.7% และ 71.1% ตามลำดับ ซึ่งเพิ่มขึ้น 10.4% ความแตกต่างกันของกลุ่มความหมายภาพอาจเป็นส่วนหนึ่งที่ทำให้การวัดความคล้ายกันของภาพนั้นได้ผลที่คลาดเคลื่อนได้ จากการทดลองครั้งนี้สามารถสรุปได้ว่า การที่นำทฤษฎีของโครงสร้างกราฟแบบลำดับขั้น เพื่อใช้ในการเปรียบเทียบด้วยอนุกรมวิธาน WordNet และการใช้รูปแบบการถ่วงน้ำหนักของคิวรีย้อนกลับมีส่วนช่วยให้เพิ่มค่าความถูกต้องของผลลัพธ์จากเดิมได้สูงถึง 85.7%

6. สรุปผลการทดลอง

จากงานที่ได้นำเสนอการค้นคืนความหมายภาพด้วยคิวรีย้อนกลับบนพื้นฐานการเชื่อมโยงแบบลำดับขั้นบนอนุกรมวิธาน WordNet นั้นจะเห็นว่า ความสัมพันธ์ที่ได้จาก HSG ของตำแหน่งคำศัพท์บนกราฟ นำมาใช้วัดความคล้ายกันของกลุ่มความหมายภาพจะ

เห็นว่าเมื่อเปรียบเทียบการวัดความคล้ายด้วย การวัดความคล้ายด้วยคิวรี (KQS) ใน 9 กลุ่มเหตุการณ์ได้ผลลัพธ์ของค่าความถูกต้องได้สูงถึง 85.7%

สิ่งควรจะมีการปรับปรุงเพิ่มเติมเพื่อให้การค้นคืนความหมายของภาพได้ดียิ่งขึ้น ก็คือการพิจารณารายละเอียดของวัตถุของภาพ การเพิ่มความหลากหลายของกลุ่มคำศัพท์ควรจะมีการจัดกลุ่มของวัตถุที่เข้ามามีการทดลองให้มีความหมายที่รัดกุมมากขึ้น และเมื่อภาพที่มีส่วนของสภาพแวดล้อมที่คล้ายคลึงกันมาก จะถูกจัดอยู่ในกลุ่มเดียวกันทั้งที่ วัตถุภายในภาพมีความแตกต่างกันของรายละเอียดภายในซึ่งบางครั้งการจัดคำหลักจะมีผลต่อการแปลความหมายโดยตรง แต่อย่างไรก็ตามการให้ความสัมพันธ์ของวัตถุภายในภาพนั้นสามารถช่วยทำการแปลความหมายที่ได้มีความสมบูรณ์ขึ้นอย่างเห็น

เอกสารอ้างอิง

- [1] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R., "Content-Based Image Retrieval at the End of the Early Years", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22, 12.
- [2] Ying Liua, Dengsheng Zhanga, , Guojun Lua, , Wei-Ying Mab, A survey of content-based image retrieval with high-level semantics, Pattern Recognition, Volume 40, Issue 1, January 2007, Pages 262–282.
- [3] Kevin Lin, Huei-Fang Yang, Kuan-Hsien Liu, Jen-Hao Hsiao, Chu-Song Chen Rapid Clothing Retrieval via Deep

Learning of Binary Codes and Hierarchical Search ACM International Conference on Multimedia Retrieval (ICMR), 2015.

[4] Jia Li, James Z. Wang. “Automatic linguistic indexing of pictures by a statistical modeling approach”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25, 9, pp. 1075-1088.

[5] Galleguillos C., Belongie S, Context Based Object Categorization: A Critical Survey. Computer Vision and Image Understanding (CVIU), Vol. 114, 2010, pp. 712-722.

[6] Tie Hua Zhou, Ling Wang , and Keun Ho Ryu, Supporting Keyword Search for Image Retrieval with Integration of Probabilistic Annotation , 7,Sustainability 2015, 6303-6320.

[7] Russell, B.C., Torralba, A., Murphy, K.P., and Freeman, W.T.. “LabelMe: a database and web-based tool for image annotation,” International Journal Computer Vision, vol.77, 2008.

[8] L. Von Ahn, Liu, and M. Blum, “Peekaboom: a game for locating objects in images”, In: Proceedings SIGCHI conference on Human Factors in Computing Systems, 2006, pp. 55–64.

[9] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M.I. Jordan, “Matching words and pictures,” J. Mach. Learn. Res., vol.3, pp. 1107–1135, 2003.

[10] M. Johnson and R. Cipolla, “Improved image annotation and labeling through multi-label boosting,” in Brit. Machine Vision Conf., 2005.

[11] Venkatesh N. Murthy, Subhransu Maji, R. Manmatha, Automatic Image Annotation using Deep Learning Representations, Proceeding ICMR '15 Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, 2015, pp. 603-606.

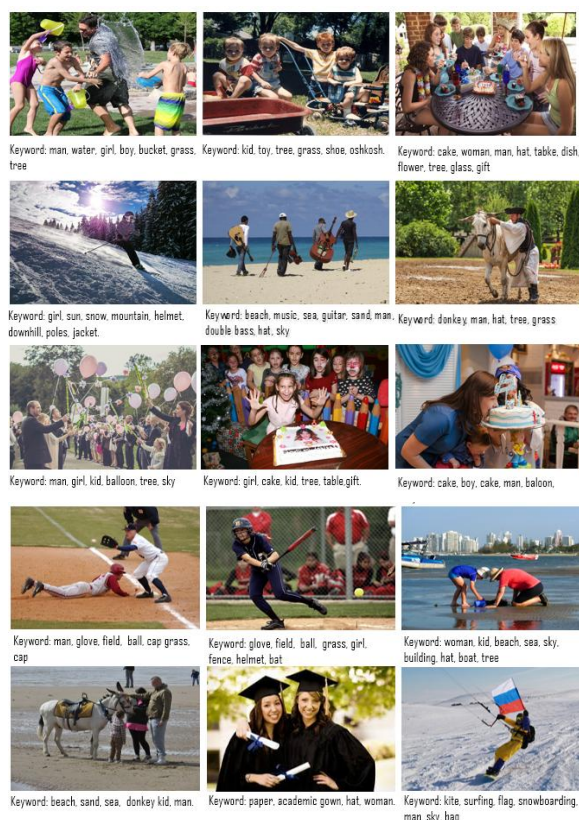
[12] Patwardhan and Pedersen, “Using WordNet Based Context Vectors to Estimate the Semantic Relatedness of Concepts,” Proceedings of the EACL 2006 Workshop Making Sense of Sense - Bringing Computational Linguistics and

Psycholinguistics Together, pp. 1-8, April 4, 2006, Trento, Italy.

[13] Miller, George A. “WordNet: An on-line lexical database,” International Journal of Lexicography, Vol. 3, 1990, pp. 235–312.

[14] I. Simon, N. Snavely, and S. Seitz, “Scene summarization for online image collections,” in IEEE Int. Conf. Computer Vision, 2007, pp. 1–8.

[15] Y. Wu, J.-Y. Bouguet, A. Nefian, and I. Kozintsev, “Learning concept templates from web images to query personal image databases,” in IEEE Int. Conf. Multimedia and Expo, 2007, pp. 1986–1989.



รูปที่ 5. ตัวอย่างการใส่คำหลักบนภาพตามในแต่ละกลุ่มความหมาย