

การใช้เทคนิคการทำเหมืองข้อมูลกราฟด้วยวิธีการเรียนรู้แบบมีผู้สอนสำหรับจำแนก
ลักษณะเด่นของฟองน้ำในชั้นเดโมสปอนเจีย
Supervised Learning for *Demospongiae* Identification using Graph Mining
Technique

ณัฐธิมา สุรเดช^{1*}, วิลาวลย์ ยาทองคำ²
Natsima Suradet^{1*}, Wilawan Yathongkhum²

¹หลักสูตรวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
โรงเรียนบ้านแม่เทย จังหวัดลำพูน

¹Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Lanna,
Thailand

²Ban Mae Thoei School, Lamphun, Thailand

*Corresponding author. E-mail: natsima@hotmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการใช้เทคนิคการทำเหมืองข้อมูลกราฟ ด้วยวิธีการเรียนรู้แบบมีผู้สอน เพื่อจำแนกลักษณะเด่นของฟองน้ำในชั้นเดโมสปอนเจีย ซึ่งข้อมูลฟองน้ำที่ใช้เป็นฟองน้ำทะเลที่อาศัยอยู่ในมหาสมุทรแอตแลนติก และทะเลเมดิเตอร์เรเนียน จำนวน 7 วงศ์ ในงานวิจัยนี้ข้อมูลฟองน้ำจะถูกนำเสนอให้อยู่ในรูปแบบของข้อมูลแบบกราฟก่อนนำไปทำการวิเคราะห์เพื่อจำแนกลักษณะเด่น จากผลการวิเคราะห์พบว่าฟองน้ำในชั้นเดโมสปอนเจียแต่ละวงศ์มีลักษณะจำเพาะที่แตกต่างกันอย่างชัดเจน ซึ่งคุณลักษณะเหล่านั้นสามารถนำไปประยุกต์ใช้ในการพัฒนาต้นแบบระบบผู้เชี่ยวชาญในการจำแนกวงศ์ของฟองน้ำได้ เมื่อทำการทดสอบต้นแบบระบบผู้เชี่ยวชาญโดยวัดประสิทธิภาพจากความสามารถในการจำแนก พบว่าสามารถจำแนกวงศ์ของฟองน้ำได้ถูกต้อง คิดเป็นค่าความแม่นยำในการจำแนกเท่ากับ 76.47 เปอร์เซ็นต์ ซึ่งแสดงให้เห็นได้อย่างชัดเจนว่า เทคนิคการทำเหมืองข้อมูลกราฟด้วยวิธีการเรียนรู้แบบมีผู้สอนมีความเหมาะสมและสามารถนำมาใช้ในการจำแนกลักษณะเด่นของฟองน้ำในชั้นเดโมสปอนเจียได้จริง

คำสำคัญ: การทำเหมืองข้อมูลกราฟ วิธีการเรียนรู้แบบมีผู้สอน การจำแนกฟองน้ำชั้นเดโมสปอนเจีย

ABSTRACT

This research purposes a graph mining technique to identify the particular characteristic of *Demospongiae* by supervised learning method. The sponge dataset contained 7 families belonging to the *Demospongiae* class was collected from the Mediterranean and Atlantic

Received 18-01-2019
Revised 25-03-2019
Accepted 26-03-2019

oceans. The dataset of *Demospongiae* was performed on graph dataset and then was analyzed to identify the particular characteristic by using graph based supervised learning method. The learned substructures can identify the unique of a specific feature for each family and can use to develop the prototype of knowledge-based expert system for *Demospongiae* identification. The prototype was evaluated by measuring the efficiency from the ability to classify the sponge family. The experimental results showed the identification performance accuracy is 76.47%. This indicated that graph mining technique by supervised learning method is valid and practicable for *Demospongiae* identification.

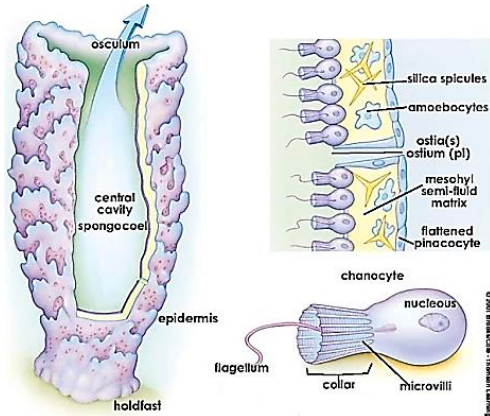
Keywords: Graph mining, Supervised learning, *Demospongiae* identification

1. บทนำ

ฟองน้ำ (Sponge) เป็นสิ่งมีชีวิตในไฟลัมพอริเฟอร่า (*Phylum Porifera*) ในปัจจุบันมีอยู่ประมาณ 7,000 สายพันธุ์ [1] ซึ่งฟองน้ำชั้นเดโมสปอนเจีย (*Demospongiae*) จัดได้ว่าเป็นชั้นที่มีความหลากหลายมากที่สุด ไฟลัมพอริเฟอร่า โดยคิดเป็นจำนวนมากถึง 90 % ของฟองน้ำทั้งหมด ฟองน้ำเป็นสัตว์ไม่มีกระดูกสันหลังที่โบราณที่สุด กำเนิดขึ้นในยุคแคมเบรียน (Cambrian) หรือเมื่อประมาณ 550 ล้านปีมาแล้ว ฟองน้ำจัดเป็นสิ่งมีชีวิตหลายเซลล์ แต่ยังไม่เป็นเนื้อเยื่อที่แท้จริง และไม่มีอวัยวะ [2] ด้วยเหตุนี้ฟองน้ำจึงจัดเป็นกลุ่มสัตว์ที่อยู่กึ่งกลางระหว่างโพรโทซัว (Protozoa) และอาณาจักรเมทาซัว (Metazoan) ฟองน้ำจัดอยู่ในไฟลัมพอริเฟอร่า ซึ่งสามารถจำแนกตามลักษณะของโครงสร้างออกเป็น 4 ชั้น (Class) คือ ชั้นคัลคาเรีย (*Calcarea*), ชั้นเฮกซาคตินา (*Hexactinellida*), ชั้นเดโมสปอนเจีย และ ชั้นสเกลอโรสปอนเจีย (*Sclerospongiae*) [3] ฟองน้ำส่วนใหญ่อาศัยอยู่ในทะเล และสามารถพบได้ตั้งแต่เขตน้ำขึ้นน้ำลงจนถึงพื้นทะเลลึก

ฟองน้ำเป็นสิ่งมีชีวิตหลายเซลล์ที่มีโครงสร้างซับซ้อนน้อยที่สุด โดยตัวเต็มวัยของฟองน้ำจะยึดเกาะ

อยู่กับที่ (Sessile) และกินอาหารโดยการกรอง ผนังลำตัวของฟองน้ำประกอบด้วยรูพรุนขนาดเล็กทั่วตัว มีทางให้น้ำผ่านเข้า เรียกว่า ออสเทีย (Ostia) ซึ่งเป็นการนำอาหารเข้าสู่ลำตัว ส่วนช่องขนาดใหญ่เป็นทางให้น้ำออกเรียกว่า ออสคูลัม (Osculum) ฟองน้ำเป็นสัตว์เพียงกลุ่มเดียวที่มีการจัดเรียงตัวของเซลล์ล้อมรอบระบบท่อ (Cannal system) ช่องทางผ่านของน้ำและโพรทอยาใน (Chamber) ทำให้น้ำไหลผ่านเข้าออกจากตัวฟองน้ำได้อย่างต่อเนื่อง ผนังลำตัวประกอบด้วยเนื้อเยื่อสองชั้นคือ เนื้อเยื่อชั้นนอกเป็นเซลล์รูปร่างแบนเรียงกันคล้ายแผ่นกระเบื้อง ประกอบด้วยเซลล์เป็นบล็อก มีแฟลกเจลลัมช่วยโบกพัดให้น้ำเคลื่อนผ่านลำตัว และทำหน้าที่กินสิ่งมีชีวิตเล็ก ๆ ที่ปนมากับน้ำ เรียกเซลล์เหล่านี้ว่า เซลล์บล็อกคอ (Choanocyte) ระหว่างเนื้อเยื่อสองชั้น จะมีชั้นกลางที่มีลักษณะคล้ายวุ้นเรียกว่า มีเซนไคม์ (Mesenchyme) ในชั้นนี้มีเซลล์พิเศษที่สามารถเคลื่อนที่ได้ เรียกว่า อะมีโบไซต์ (Amoebocyte) ซึ่งเป็นเซลล์ที่มีหน้าที่หลายอย่าง เช่น ย่อยอาหาร สร้างเซลล์สืบพันธุ์ และสร้างโครงสร้างค้ำจุนร่างกาย เป็นต้น ภาพที่ 1 แสดงโครงสร้างภายในของฟองน้ำ



ภาพที่ 1 โครงสร้างภายในของฟองน้ำ [4]

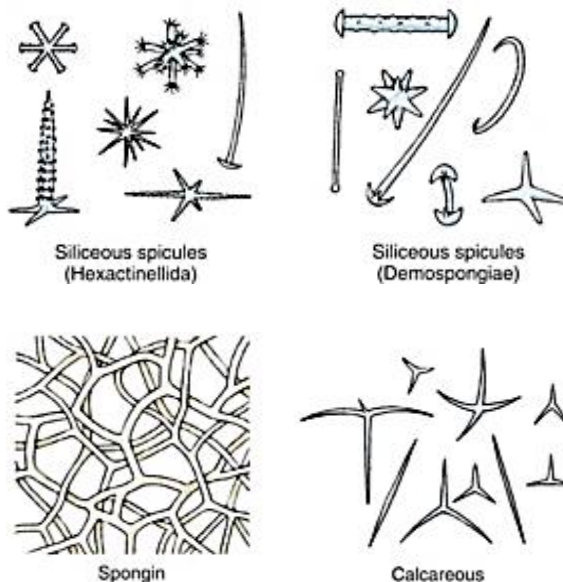
ฟองน้ำเป็นสัตว์ไม่มีกระดูกสันหลัง แต่มีโครงร่างอย่างอื่นทำหน้าที่แทนกระดูกค้ำจุนให้คงรูปอยู่ได้ เรียกว่า โครงร่าง (Skeleton) มี 2 ประเภท คือ

- เส้นใยสปอนจิน (Spongin fiber) เป็นเส้นใยโปรตีนที่มีองค์ประกอบซับซ้อนสร้างจากสปันโกไซต์ (Spongocyte) เป็นโครงร่างที่ทำให้ฟองน้ำมีความเหนียวและอ่อนนุ่มยืดหยุ่นตัวได้ดี

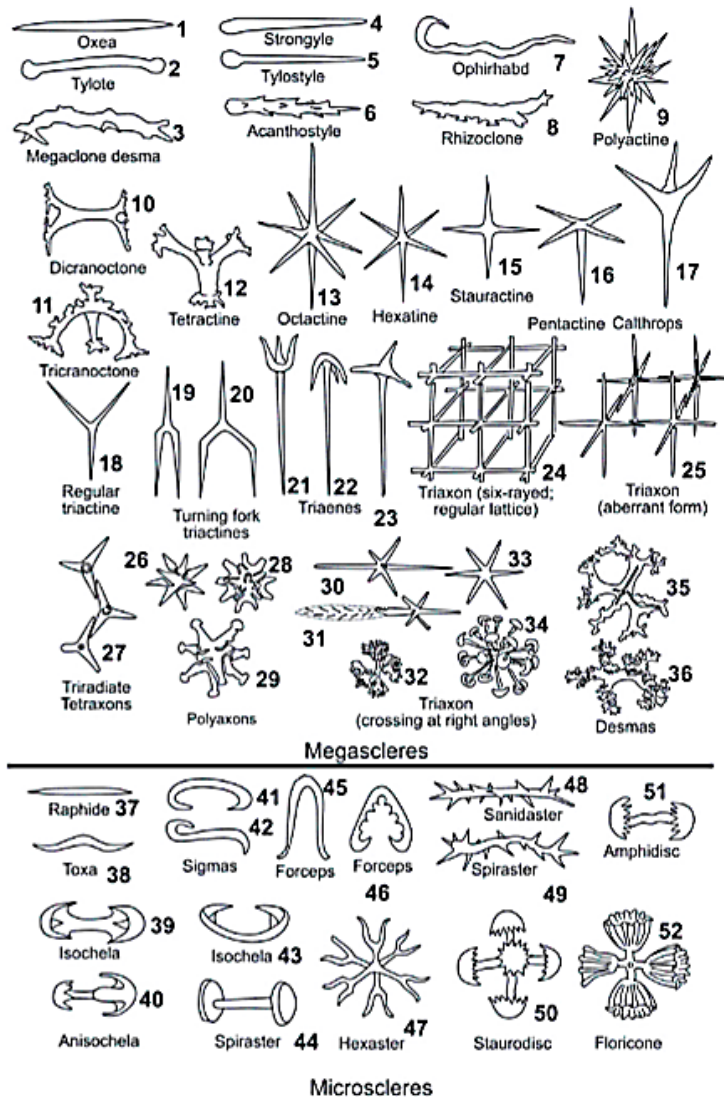
- ขวาก (Spicule) เป็นโครงร่างแข็งแบ่งออกเป็น 2 กลุ่มตามองค์ประกอบ คือ ขวากหินปูน

(Calcareous spicule) มีหินปูนเป็นองค์ประกอบ และขวากแก้ว (Siliceous spicule) มีซิลิกา (Silica) เป็นองค์ประกอบ ในภาพที่ 2 เป็นลักษณะโครงร่างค้ำจุนของฟองน้ำแต่ละประเภท

ฟองน้ำในชั้นเดโมสปอนเจีย มีขวากที่เป็นสารประกอบซิลิกา 2 ขนาด คือ ขวากขนาดใหญ่ (ยาวมากกว่า 100 ไมครอน) เรียกว่า Megasclere และขวากขนาดเล็ก (ยาวน้อยกว่า 100 ไมครอน) เรียกว่า Microsclere โดยขวากขนาดใหญ่จะสานตัวเป็นโครงร่างค้ำจุนหลักของฟองน้ำเชื่อมด้วยเส้นใยสปอนจิน และมีขวากขนาดเล็กแทรกอยู่ระหว่างโครงร่างนั้น เพื่อค้ำจุนเนื้อฟองน้ำที่อ่อนนุ่ม [3] ซึ่งลักษณะรูปร่างของขวากมีหลายแบบและมีชื่อเรียกแตกต่างกันออกไป ลักษณะรูปร่างของขวากถูกใช้เป็นส่วนหนึ่งในการจัดหมวดหมู่ ในภาพที่ 3 แสดงลักษณะรูปร่างของขวากแบบต่าง ๆ



ภาพที่ 2 ลักษณะโครงร่างของฟองน้ำ [5]



ภาพที่ 3 ลักษณะรูปร่างของขากแบบต่างๆ [6]

การทำเหมืองข้อมูล (Data mining) สามารถสืบค้นความรู้ที่เป็นประโยชน์บนฐานข้อมูลขนาดใหญ่ โดยทำการวิเคราะห์ข้อมูลเพื่อหาข้อสรุปหรือความรู้ที่สำคัญ ซึ่งผลลัพธ์ที่ได้มาเป็นคุณสมบัติหรือลักษณะเฉพาะของข้อมูลที่อยู่ในฐานข้อมูลนั้น การทำเหมืองข้อมูลกราฟ (Graph mining) เป็นการทำเหมืองข้อมูลรูปแบบหนึ่ง ซึ่งทำการวิเคราะห์ข้อมูลที่อยู่ในรูปแบบของกราฟ เพื่อค้นหาข้อสรุปหรือความรู้ โดย

ผลลัพธ์ที่ได้จะแสดงโครงสร้างและความสัมพันธ์ต่าง ๆ ในชุดข้อมูล

เทคนิคการค้นหากราฟย่อยที่ปรากฏบ่อย (Frequent subgraphs) เป็นเทคนิคหนึ่งในการทำเหมืองกราฟ กระบวนการค้นหาความถี่ของกราฟย่อยประกอบไปด้วย การหาความถี่ทั้งหมดของกราฟย่อย และตรวจสอบความถี่ของแต่ละกราฟย่อยว่ามีค่าความถี่น้อยกว่า Minimum support หรือไม่ [7]

SUBDUE เป็นระบบการเรียนรู้เชิงสัมพันธ์ด้วยกราฟซึ่งถูกนำมาประยุกต์ใช้กับงานในหลายด้าน เช่น โครงสร้างการทำเหมืองเว็บ การวิเคราะห์เครือข่ายสังคม และงานทางด้านชีวสารสนเทศ เป็นต้น โดยนำไปใช้ในการค้นหากราฟย่อยที่ปรากฏอยู่ในฐานข้อมูลกราฟ [8]

SUBDUE เป็นระบบการเรียนรู้ที่ใช้หลัก Minimum description length (MDL) heuristic ในการหากราฟย่อย ซึ่งค่าของกราฟย่อยหาได้จากสมการที่ (1)

$$value(S, G) = \frac{DL(G)}{(DL(S) + DL(G|S))} \quad (1)$$

เมื่อ S คือ กราฟย่อยที่ต้องการการค้นหา, G คือ กราฟนำเข้า, DL(G) คือ จำนวนบิตที่ต้องการในการเข้ารหัสกราฟ, DL(S) คือ จำนวนบิตที่ต้องการในการเข้ารหัสกราฟย่อยที่ค้นหา และ DL(G|S) คือ จำนวนบิตที่ต้องการในการเข้ารหัสกราฟ G หลังจากโครงสร้างย่อย S ถูกบีบอัด

```
Subdue(Graph, BeamWidth, MaxBest, MaxSubSize, Limit)
  ParentList = {}
  ChildList = {}
  BestList = {}
  ProcessedSubs = 0
  Create a substructure from each unique vertex label and
  its single-vertex instances; insert the resulting
  substructures in ParentList
  while ProcessedSubs <= Limit and ParentList is not empty do
    while ParentList is not empty do
      Parent = RemoveHead(ParentList)
      Extend each instance of Parent in all possible ways
      Group the extended instances into Child substructures
      foreach Child do
        if SizeOf(Child) <= MaxSubSize then
          Evaluate the Child
          Insert Child in ChildList in order by value
          if Length(ChildList) > BeamWidth then
            Destroy the substructure at the end of ChildList
      ProcessedSubs = ProcessedSubs + 1
      Insert Parent in BestList in order by value
      if Length(BestList) > MaxBest then
        Destroy the substructure at the end of BestList
      Switch ParentList and ChildList
  return BestList
```

ภาพที่ 4 ขั้นตอนการค้นหาของ SUBDUE

ขั้นตอนวิธีการค้นหาของ SUBDUE ใช้ Beam search เป็นหลักในการค้นหากราฟย่อย ในภาพที่ 4 เป็นขั้นตอนวิธีที่ใช้ในการค้นหากราฟย่อย โดยขั้นแรก

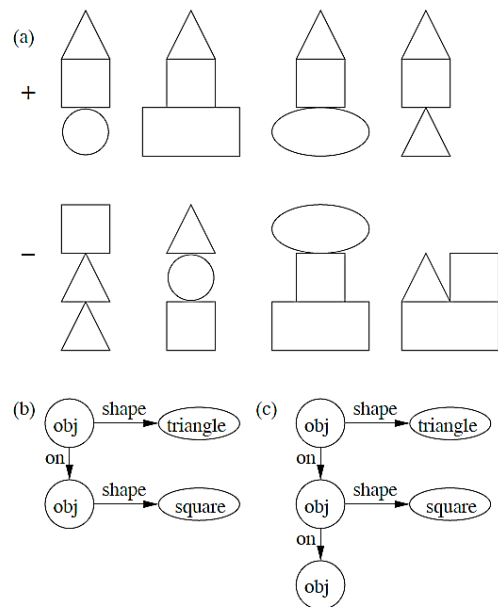
ของขั้นตอนวิธีเริ่มจากการสร้างลิสต์ที่มีชื่อว่า ParentList (ประกอบไปด้วยโครงสร้างย่อยที่จะทำการขยาย), ChildList (ประกอบไปด้วยโครงสร้างย่อย

ที่ถูกขยาย) และ BestList (ประกอบไปด้วยโครงสร้างย่อยที่มีค่าสูงสุด) ให้เป็นค่าว่าง และให้ ProcessedSubs มีค่าเท่ากับ 0 ขั้นตอนถัดไปทำการสร้างโครงสร้างย่อยจากแต่ละป้ายชื่อของจุดยอดที่ไม่ซ้ำกันทั้งหมด โดยเอาไปใส่ไว้ใน ParentList หลังจากนั้นภายในลูป while แต่ละโครงสร้างย่อยจะถูกเอาออกจาก ParentList และขยายทุกทางที่เป็นไปได้โดยการเพิ่มเส้นเชื่อมและจุดยอด หรือเพิ่มเส้นเชื่อมระหว่างจุดยอดสองจุด เมื่อ ParentList มีค่าว่าง จะมีการสลับระหว่าง ParentList กับ ChildList [9] ข้อมูลนำเข้าของ SUBDUE เป็นกราฟที่แทนวัตถุและลักษณะของวัตถุ โดยกำหนดให้ป้ายชื่อของจุดยอดแทนชื่อของวัตถุหรือลักษณะของวัตถุและป้ายชื่อของเส้นเชื่อมแทนความสัมพันธ์ระหว่างวัตถุกับวัตถุหรือความสัมพันธ์ระหว่างวัตถุกับลักษณะ

การเรียนรู้แบบมีผู้สอน (Supervised learning) ของ SUBDUE มุ่งเน้นไปที่ชุดข้อมูลที่มีเพียง 2 กลุ่ม (Class) เท่านั้น การเรียนรู้แบบมีผู้สอนจะเป็นการค้นหาโครงสร้างกราฟย่อยที่เกิดขึ้นในกราฟเชิงบวก แต่มักไม่อยู่ในกราฟเชิงลบ ข้อมูลนำเข้าจะอยู่ในรูปแบบของกราฟที่มีการระบุว่าเป็นกราฟเชิงบวก หรือกราฟเชิงลบ ซึ่งกราฟในแต่ละประเภทอาจจะมีเพียงกราฟเดียวหรือหลายกราฟก็ได้ โดย G^+ แทนกราฟเชิงบวก และ G^- แทนกราฟเชิงลบ ดังภาพที่ 5(a) วิธีการเรียนรู้แบบมีผู้สอน คือการหากราฟย่อยที่มักปรากฏในกราฟเชิงบวก แต่ไม่อยู่ในกราฟเชิงลบ วิธีการนี้เป็นการค้นหากราฟย่อยที่มีขนาดเล็กซึ่งหาได้จากสมการที่ (2)

$$\frac{|\{g \in G^+ | S \not\subseteq g\}| + |\{g \in G^- | S \subseteq g\}|}{|G^+| + |G^-|} \quad (2)$$

เมื่อ $S \subseteq g$ หมายถึง S เป็นกราฟถอดแบบ (Isomorphic) กับกราฟย่อย g จากสมการพจน์แรกเป็นผลลบที่ผิดพลาด (False negatives) และพจน์ที่สองเป็นผลบวกที่ผิดพลาด (False positives) วิธีการนี้มีเป้าหมายเพื่อนำไปสู่การค้นหากราฟย่อยที่มีขนาดเล็กซึ่งสามารถแยกแยะได้ดี ดังตัวอย่างกราฟในภาพที่ 5(b)



ภาพที่ 5 ตัวอย่างการเรียนรู้แบบมีผู้สอน (a) กราฟเชิงบวกและกราฟเชิงลบ (b, c) กราฟที่เป็นไปได้

เราสามารถอคติ (Bias) ในการค้นหาคุณลักษณะโดยใช้ค่าทางสารสนเทศเพื่อหากราฟย่อยเชิงบวกที่ไม่มีในกราฟเชิงลบ ถ้า $DL(G)$ คือ จำนวนบิตที่แทนความยาวของคุณลักษณะ (Description length) ของกราฟ G และ $DL(G|S)$ คือ ความยาวของคุณลักษณะ (เป็นบิต) ของกราฟ G ที่ถูกย่อโดยกราฟย่อย S เราสามารถหา S ได้จากสมการที่ (3)

$$DL(G^+ | S) + DL(S) + DL(G^-) - DL(G^- | S) \quad (3)$$

วิธีนี้จะนำไปสู่การค้นหากราฟย่อยที่มีขนาดใหญ่เพื่อแสดงลักษณะของกราฟเชิงบวก โดยลักษณะต่างๆ เหล่านี้มักไม่มีในกราฟเชิงลบ ตัวอย่างกราฟในภาพที่ 5(c)

กระบวนการนี้สามารถทำซ้ำเพื่อเรียนรู้สมมติฐาน หากใช้เครื่องมือวัดความผิดพลาดแล้วกราฟเชิงบวกใด ๆ ที่มีการเรียนรู้กราฟย่อยจะถูกลบออกจากการทำซ้ำในรอบถัดไป ถ้าใช้ค่าทางสารสนเทศ ตัวอย่างของการเรียนรู้กราฟย่อยในกราฟเชิงบวกและกราฟเชิงลบจะถูกย่อเป็นจุดยอด (Vertex) เดียว [10]

Yada et al. [11] ได้นำเทคนิคการทำเหมืองข้อมูลกราฟไปประยุกต์ใช้กับการวิเคราะห์พฤติกรรมของผู้บริโภคในด้านการตลาด โดยนำข้อมูลประวัติการซื้อสินค้าของผู้บริโภคแปลงเป็นโครงสร้างกราฟแล้วใช้เทคนิคการค้นหารูปแบบกราฟที่เกิดขึ้นบ่อยมาวิเคราะห์ความนิยมในการบริโภคอาหารเพื่อสรุปพบว่า การใช้ข้อมูลที่มีโครงสร้างของกราฟเพื่อแสดงพฤติกรรมผู้บริโภคทำให้สามารถนำเสนอข้อมูลที่มีคุณสมบัติตามลำดับเวลาและในความสัมพันธ์ที่มีอยู่ระหว่างกลุ่มผลิตภัณฑ์ต่างๆ ได้อย่างมีประสิทธิภาพ

You et al. นำเสนอวิธีการค้นหารูปแบบที่มีความหมายทางชีวภาพเกี่ยวกับวิถีเมแทบอลิก (Metabolic pathway) โดยใช้ SUBDUE ซึ่งเป็นระบบการเรียนรู้ด้วยกราฟเชิงสัมพันธ์ ในการทดลองใช้อัลกอริทึมการเรียนรู้เชิงสัมพันธ์บนกราฟทั้งในสถานการณ์ที่อยู่ภายใต้การเรียนรู้แบบมีผู้สอนและการเรียนรู้แบบไม่มีผู้สอนพบว่าโครงสร้างกราฟที่เรียนรู้โดย SUBDUE นั้นมีความเข้าใจความหมายทางชีวภาพของวิถีเมแทบอลิก โดยในการเรียนรู้แบบมีผู้สอนสามารถระบุคุณสมบัติที่เป็นเอกลักษณ์ของวิถีเมแทบอลิกแต่ละประเภทได้อย่างชัดเจน ซึ่งความแม่นยำเฉลี่ยของการทดลองเป็น 71.76 เปอร์เซ็นต์

ในส่วนของการเรียนรู้แบบไม่มีผู้สอนผลลัพธ์ที่ได้จากการเรียนรู้โดย SUBDUE จะอยู่ในรูปแบบการจัดกลุ่มแบบลำดับชั้น (Hierarchical cluster) ที่แสดงคุณสมบัติทั่วไปของวิถีเมแทบอลิกในแต่ละประเภท [12]

ในปัจจุบันการศึกษาวิจัยเกี่ยวกับฟองน้ำมีอยู่เป็นจำนวนมาก โดยส่วนใหญ่เป็นการศึกษาวิจัยด้านชีววิทยาของฟองน้ำ เช่น ลักษณะทางสัณฐานวิทยา การจำแนกสายพันธุ์ หรือการค้นพบสายพันธุ์ใหม่โดยการวิเคราะห์วิวัฒนาการโมเลกุลของฟองน้ำ เป็นต้น แต่การประมวลผลด้วยคอมพิวเตอร์ของข้อมูลเกี่ยวกับฟองน้ำยังมีอยู่เป็นจำนวนน้อยมากเมื่อเทียบกับจำนวนของข้อมูลดิบที่มีอยู่อย่างมหาศาล ดังนั้นทางผู้วิจัยจึงมีแนวคิดที่จะนำเทคนิคการทำเหมืองข้อมูลกราฟมาวิเคราะห์ข้อมูลลักษณะของฟองน้ำชั้นเดโมสปอนเจียเพื่อหาลักษณะเฉพาะ หรือจุดเด่นของฟองน้ำในแต่ละวงศ์ ซึ่งข้อมูลจำเพาะเหล่านั้นสามารถนำไปใช้ประโยชน์ในการประมวลผลด้วยคอมพิวเตอร์ด้านปัญญาประดิษฐ์ได้ เช่น ระบบผู้เชี่ยวชาญเกี่ยวกับการจำแนกฟองน้ำ เป็นต้น

ในบทความนี้แบ่งเนื้อหาออกเป็นส่วนต่างๆ ดังนี้ ส่วนที่ 2 กล่าวถึงวัสดุ อุปกรณ์ และวิธีการในการดำเนินการวิจัย ในส่วนที่ 3 เป็นรายละเอียดเกี่ยวกับผลการทดลองและอภิปรายผล สำหรับส่วนที่ 4 เป็นบทสรุปของการวิจัยนี้

2. วัสดุ อุปกรณ์ และวิธีการทดลอง

2.1 ชุดข้อมูลที่ใช้ในการทดลอง

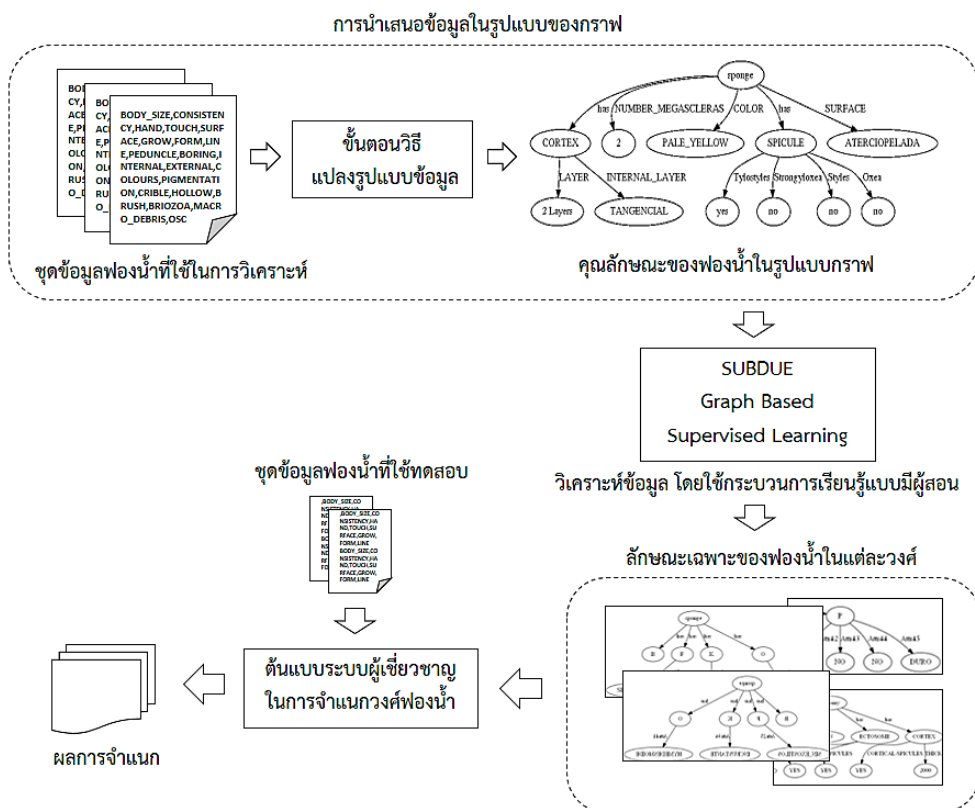
ชุดข้อมูลที่ใช้ในการทดลองนี้เป็นข้อมูลฟองน้ำทะเลที่อาศัยอยู่ในมหาสมุทรแอตแลนติก และทะเลเมดิเตอร์เรเนียนซึ่งอยู่ในชั้นเดโมสปอนเจีย โดยประกอบไปด้วยข้อมูลฟองน้ำทั้งหมด 76 ตัว จำแนกได้

7 วงศ์ (Class) จาก UCI Machine Learning Repository [13] ฟองน้ำแต่ละตัวประกอบไปด้วย 45 คุณลักษณะ เช่น ลักษณะผิวฟองน้ำ (Cortex) สีของฟองน้ำ ลักษณะของขากขนาดใหญ่ และลักษณะของขากขนาดเล็ก เป็นต้น

2.2 วิธีการดำเนินการวิจัย

ในการวิจัยนี้มีขั้นตอนในการดำเนินการวิจัยดังแสดงในภาพที่ 6 โดยขั้นตอนแรกเป็นการนำเสนอชุด

ข้อมูลที่ใช้ในการทดลองให้อยู่ในรูปแบบของกราฟ จากนั้นนำคุณลักษณะของฟองน้ำในรูปแบบกราฟไปวิเคราะห์ข้อมูล โดยใช้กระบวนการเรียนรู้แบบมีผู้สอน และนำลักษณะเฉพาะของฟองน้ำในแต่ละวงศ์ที่ได้จากการวิเคราะห์ มาพัฒนาต้นแบบผู้เชี่ยวชาญในการจำแนกวงศ์ของฟองน้ำในชั้นแต่โมสพอนเจีย

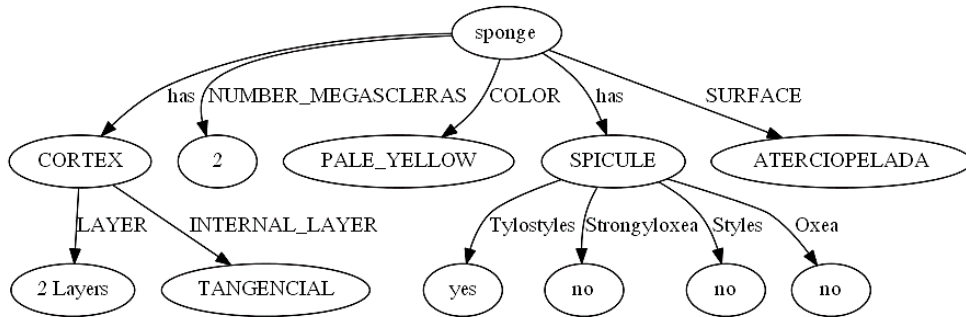


ภาพที่ 6 ขั้นตอนในการดำเนินการวิจัย

2.2.1 การนำเสนอข้อมูลในรูปแบบของกราฟ

ข้อมูลฟองน้ำแต่ละตัวจะถูกนำเสนอในรูปแบบของกราฟ ดังตัวอย่างในภาพที่ 7 ซึ่งเป็นการแทนคุณลักษณะของฟองน้ำทั้ง 45 คุณลักษณะ โดยจุดยอดแต่ละจุดในกราฟจะแสดงค่าของข้อมูลซึ่งเป็น

คุณลักษณะที่ปรากฏ หรือเป็นข้อความซึ่งบอกรายละเอียดเกี่ยวกับส่วนประกอบของฟองน้ำ เช่น ผิวฟองน้ำ (Cortex) ขาก (Spicule) เป็นต้น ในส่วนของเส้นเชื่อมจะแสดงความสัมพันธ์ระหว่างจุดยอดที่มีเส้นเชื่อมกัน โดยมีชื่อของคุณลักษณะกำกับอยู่บนเส้น



ภาพที่ 7 การนำเสนอคุณลักษณะทั้ง 11 คุณลักษณะของฟองน้ำในรูปแบบกราฟ

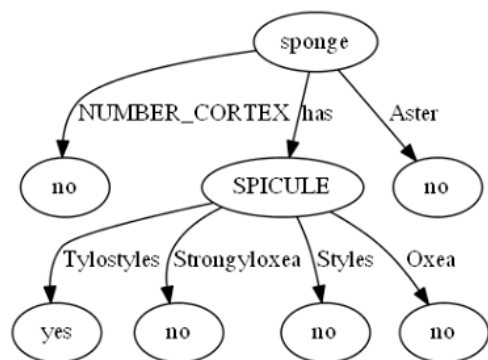
2.2.2 การวิเคราะห์ข้อมูลโดยใช้

กระบวนการในการเรียนรู้แบบมีผู้สอน

การทดลองนี้ผู้วิจัยใช้ SUBDUE ในการวิเคราะห์ข้อมูลโดยใช้กระบวนการในการเรียนรู้แบบมีผู้สอน ซึ่งชุดข้อมูลที่ใช้ในการทดลองได้จำแนกตามวงศ์ของฟองน้ำในชั้นเดโมสปอนเจียทั้งหมด 7 วงศ์ดังต่อไปนี้ คือ *Tethyidae*, *Clionidae*, *Timeidae*, *Suberitidae*, *Stylocordylidae*, *Polymastiidae* และ *Spirastrellidae* ในการทดลองจะนำชุดข้อมูลมาทำการเรียนรู้ที่ละวงศ์ โดยชุดข้อมูลฟองน้ำที่อยู่ในวงศ์ที่ต้องการเรียนรู้จะใส่สัญลักษณ์ XP ส่วนชุดข้อมูลฟองน้ำที่อยู่ในวงศ์อื่นจะใส่สัญลักษณ์เป็น NP ซึ่งกระบวนการเรียนรู้แบบมีผู้สอนดังกล่าว เป็นการค้นหาลักษณะเฉพาะของฟองน้ำในแต่ละวงศ์ โดยที่คุณลักษณะเหล่านั้นมักจะปรากฏในฟองน้ำวงศ์อื่น

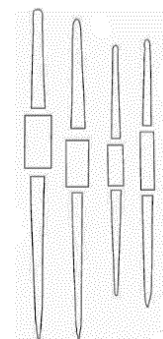
3. ผลการทดลองและอภิปรายผล

ตัวอย่างผลการวิเคราะห์ข้อมูลกราฟของฟองน้ำวงศ์ *Suberitidae* ในชั้นเดโมสปอนเจีย โดยใช้เทคนิคการทำเหมืองข้อมูลกราฟ ด้วยกระบวนการเรียนรู้แบบมีผู้สอน แสดงดังภาพที่ 8 เมื่อทำการอธิบายผลที่ได้จากการวิเคราะห์ข้อมูลกราฟพบว่าฟองน้ำในแต่ละวงศ์มีลักษณะเฉพาะดังต่อไปนี้



ภาพที่ 8 ผลที่ได้จากการวิเคราะห์ข้อมูลกราฟของฟองน้ำในวงศ์ *Suberitidae*

ฟองน้ำที่อยู่ในวงศ์ *Tethyidae* ส่วนใหญ่มีรูปร่างของขากขนาดใหญ่เป็นแบบ *Strongyloxea* ซึ่งมีลักษณะเรียวยาว โดยปลายด้านหนึ่งแหลมและอีกด้านหนึ่งไม่แหลม ดังภาพที่ 9



ภาพที่ 9 ลักษณะของขากแบบ *Strongyloxea* [14]

ฟองน้ำที่อยู่ในวงศ์ *Clionidae* มีลักษณะของ
ขากขนาดใหญ่เป็นทรงกระบอกเรียบและปลายทั้ง
สองด้านเรียวแหลม ดังภาพที่ 10



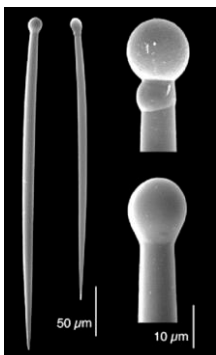
ภาพที่ 10 ลักษณะของขากแบบ Oxea [15]

ลักษณะเด่นของฟองน้ำในวงศ์ *Timeidae*
ประกอบด้วยขากขนาดใหญ่แบบ Styles มีลักษณะ
ดังภาพที่ 11 โดยปลายด้านหนึ่งจะขยายใหญ่ขึ้น ส่วน
ปลายอีกด้านหนึ่งเรียวแหลม และในส่วนของผิว
ฟองน้ำ (Cortex) ไม่พบการสานเรียงตัวกันของขาก
ขนาดใหญ่ (Megasclere) ในชั้น Ectosome



ภาพที่ 11 ลักษณะของขากแบบ Styles [16]

ฟองน้ำที่อยู่ในวงศ์ *Suberitidae* มีขากขนาด
ใหญ่เป็นรูปร่างแบบ Tylostyles ซึ่งส่วนหัวจะมี
ลักษณะคล้ายทรงกลม และส่วนปลายอีกด้านหนึ่งจะ
เรียวแหลม ดังแสดงในภาพที่ 12 ในส่วนของขาก
ขนาดเล็กที่มีรัศมีกระจายออกจากจุดศูนย์กลาง
โดยรอบคล้ายรูปดาว หรือเรียกว่า Aster นั้นไม่พบใน
ฟองน้ำวงศ์นี้และไม่มีส่วนที่เป็นผิวของฟองน้ำ

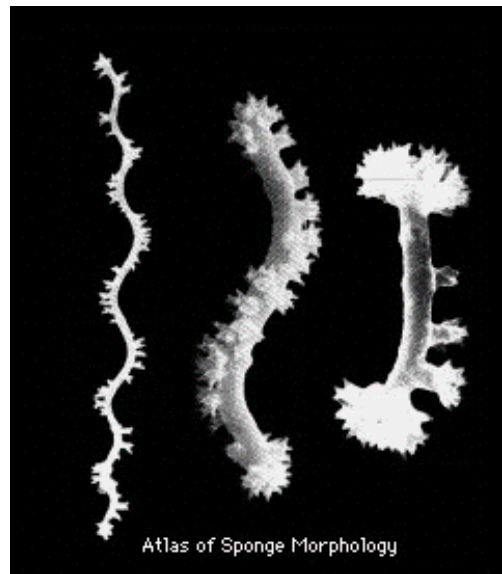


ภาพที่ 12 ลักษณะของขากแบบ Tylostyles [17]

ฟองน้ำที่อยู่ในวงศ์ *Stylodordylidae*
ประกอบด้วยขากขนาดใหญ่แบบ Styles ซึ่งมี
ลักษณะเดียวกับฟองน้ำในวงศ์ *Timeidae*

ฟองน้ำที่อยู่ในวงศ์ *Polymastiidae* จะมีส่วน
ที่เป็นผิวของฟองน้ำ ในส่วนของการยึดเกาะ จะ
เกาะติดกับวัตถุที่มีลักษณะแข็ง สำหรับขากขนาด
ใหญ่บริเวณชั้น Ectosome จะมีรูปร่างแบบ
Tylostyles และไม่พบขากขนาดเล็กที่มีลักษณะแบบ
Aster ในฟองน้ำวงศ์นี้

ฟองน้ำที่อยู่ในวงศ์ *Spirastrellidae* ไม่พบ
ส่วนที่เป็นผิวของฟองน้ำ ในส่วนของขากขนาดเล็กมี
รูปร่างแบบ Spiraster ซึ่งมีลักษณะเป็นแท่งคล้าย
เกลียวและมีหนามอยู่โดยรอบ ดังภาพที่ 13



ภาพที่ 13 ลักษณะของขากแบบ Spiraster [18]

จากลักษณะเด่นของฟองน้ำในแต่ละวงศ์ที่ได้
จากการวิเคราะห์ด้วยเทคนิคการทำเหมืองข้อมูลกราฟ
ผู้วิจัยได้นำมาสร้างเป็นต้นแบบระบบผู้เชี่ยวชาญที่
สามารถจำแนกวงศ์ของฟองน้ำชั้นเดโมสปอนเจีย โดย
การประมวลผลด้วยคอมพิวเตอร์ ซึ่งข้อมูลที่ใช้ในการ

ทดสอบต้นแบบระบบผู้เชี่ยวชาญนั้นเป็นข้อมูลของ
ฟองน้ำที่พบในมหาสมุทรแอตแลนติก และทะเลเมดิ-
เตอร์เรเนียน ซึ่งอยู่ในชั้นเดโมสปอนเจียเช่นเดียวกัน
โดยข้อมูลนี้ได้มาจากฐานข้อมูล UCI Machine
Learning Repository [19] ผู้วิจัยได้สุ่มตัวอย่าง
ฟองน้ำทั้ง 7 วงศ์ มาจำนวน 51 ตัวอย่าง พบว่า
ต้นแบบระบบผู้เชี่ยวชาญสามารถจำแนกวงศ์ของ
ฟองน้ำได้ถูกต้อง 39 ตัวอย่าง ซึ่งคิดเป็นค่า
ความแม่นยำในการจำแนกเท่ากับ 76.47% โดยมีผล
การทดสอบแสดงในตารางที่ 1 ซึ่งประสิทธิภาพในการ
จำแนกวงศ์ของฟองน้ำ วัดได้จากค่าความแม่นยำดัง
สมการที่ (4)

$$\text{ความแม่นยำ} = \frac{TP}{TP + FP} \quad (4)$$

โดยที่ TP คือจำนวนวงศ์ที่ต้นแบบระบบ
ผู้เชี่ยวชาญทำการจำแนกได้อย่างถูกต้อง และ FP คือ
จำนวนวงศ์ที่ต้นแบบระบบผู้เชี่ยวชาญมีการจำแนก
ผิดพลาด

ตารางที่ 1 ผลการจำแนกวงศ์ของฟองน้ำจากการ
ทดสอบต้นแบบระบบผู้เชี่ยวชาญ

วงศ์ของฟองน้ำ	TP	FP
<i>Tethyidae</i>	7	0
<i>Clionidae</i>	4	6
<i>Timeidae</i>	3	5
<i>Suberitidae</i>	12	0
<i>Stylocordylidae</i>	2	0
<i>Polymastiidae</i>	9	0
<i>Spirastrellidae</i>	2	1

4. สรุปผลการทดลอง

งานวิจัยนี้ได้นำเสนอการใช้เทคนิคการทำ
เหมืองข้อมูลกราฟด้วยวิธีการเรียนรู้แบบมีผู้สอน
สำหรับจำแนกลักษณะเด่นของฟองน้ำในแต่ละวงศ์ซึ่ง
อยู่ในชั้นเดโมสปอนเจีย จากผลการทดลองดังกล่าว
แสดงให้เห็นว่า เทคนิคการทำเหมืองข้อมูลกราฟ
สามารถนำมาวิเคราะห์ลักษณะเด่นของฟองน้ำในแต่
ละวงศ์ได้ เมื่อนำลักษณะจำเพาะเหล่านั้นไปพัฒนา
ต้นแบบระบบผู้เชี่ยวชาญเพื่อจำแนกวงศ์ของฟองน้ำ
โดยการประมวลผลด้วยคอมพิวเตอร์ พบว่าระบบมีค่า
ความแม่นยำในการจำแนกเท่ากับ 76.47% จาก
ผลการวิจัยแสดงให้เห็นว่า เทคนิคการทำเหมืองข้อมูล
กราฟด้วยวิธีการเรียนรู้แบบมีผู้สอนสามารถนำมาใช้
ในการจำแนกลักษณะเด่นของฟองน้ำในชั้นเดโม-
สปอนเจียได้จริง

เมื่อพิจารณาผลการจำแนกวงศ์ของฟองน้ำจาก
ต้นแบบระบบผู้เชี่ยวชาญ พบว่าฟองน้ำวงศ์
Clionidae และ *Timeidae* มีผลการจำแนกผิดพลาด
มากกว่าวงศ์อื่นๆ และลักษณะเด่นที่ได้จาก
การวิเคราะห์ข้อมูลกราฟของฟองน้ำ จะเห็นได้
อย่างชัดเจนว่าฟองน้ำวงศ์ *Timeidae* มีลักษณะเด่น
บางอย่างคล้ายกับฟองน้ำในวงศ์ *Stylocordylidae*
และลักษณะเด่นที่พบของฟองน้ำในวงศ์ *Clionidae*
นั้นแม้ว่าจะมีความจำเพาะ แต่ก็มียาจำนวนคุณลักษณะที่
โดดเด่นเป็นจำนวนน้อย

เนื่องจากในการวิจัยนี้ใช้ข้อมูลในการวิเคราะห์
ลักษณะเด่นของฟองน้ำเพียงแค่ 76 ข้อมูลเท่านั้น หาก
นำเทคนิคการทำเหมืองข้อมูลกราฟโดยกระบวนการ
เรียนรู้แบบมีผู้สอนไปใช้ในการวิเคราะห์ลักษณะเด่น
ของฟองน้ำในแต่ละวงศ์ โดยใช้ชุดข้อมูลในการ
วิเคราะห์ที่มีจำนวนมากขึ้น ลักษณะเด่นที่ได้จาก
การวิเคราะห์ก็必将มีความจำเพาะมากยิ่งขึ้น ซึ่งอาจ

ส่งผลให้สามารถพัฒนาระบบผู้เชี่ยวชาญในการจำแนกวงศ์ฟองน้ำในชั้นเดโมสปอนเจียที่มีประสิทธิภาพมากยิ่งขึ้นได้

5. อ้างอิง

- [1] Hooper JNA, Van Soest RWM. Systema Porifera: A guide to the classification of sponges. New York: Kluwer Academic/Plenum Publishers; 2002.
- [2] บพิท จารุพันธุ์, นันทพร จารุพันธุ์. สัตว์ไม่มีกระดูกสันหลัง I โพรโทซัวถึงทาร์ดิกราตา. พิมพ์ครั้งที่ 1. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์; 2545.
- [3] Hartman WD. Porifera. Synopsis and classification of living organisms 1982; 1:640-666.
- [4] Starr C, Evers CA, Starr L. Biology: Concepts and Applications. 10th Ed. Boston: Cengage Learning, Inc; 2016.
- [5] Niedzlek-Feaver M. Sponges. [Internet]. North Carolina: NC State; 2016 [cited 2018 Nov 20]. Available from: http://projects.ncsu.edu/project/bio402_315/lec3porifera/Porifera/Untitled31-summ.html
- [6] Jain S. Fundamentals of Invertebrate Palaeontology: Macrofossils. [Internet]. India: Springer India; 2017 [cited 2018 Nov 20]. doi: 10.1007/978-81-322-3658-0.
- [7] Cook DJ, Holder LB, editors. Mining Graph Data. John Wiley & Sons, Inc; 2007.
- [8] Cook DJ, Holder LB. Substructure discovery using minimum description length and background knowledge. J Artif Intell Res. 1993; 1:231-255.
- [9] Cook DJ, Holder LB. Graph-based data mining. IEEE Intell Syst. 2000; 15:32-41.
- [10] Cook DJ, Holder LB. Graph-based relational learning: current and future directions. ACM SIGKDD Explorations Newsletter. 2003; 5:90-3.
- [11] Yada K, Motoda H, Washio T, et al. Consumer Behavior Analysis by Graph Mining Technique. In: Negoita MGH, Howlett RJ, Jain LC, editors. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems; 2004 Sep 20-25; Wellington, New Zealand. 2004. P. 800-6.
- [12] You CH, Holder LB, Cook DJ. Application of Graph-based Data Mining to Metabolic Pathways. Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06); 2006 Dec 18-22; Hong Kong, China. IEEE; 2006. P. 169 - 173.
- [13] Uriz L, Domingo M. UCI Machine Learning Repository: Sponge Data Set. [Internet]. Irvine, CA: University of California, School of Information and Computer Science. [cited 2018 Oct 7]. Available from: <http://archive.ics.uci.edu/ml/datasets/sponge>

- [14] The World Porifera Database [Internet]. Rob van Soest. [cited 2018 Nov 20]. Available from: <http://www.marinespecies.org/porifera>
- [15] Moreno AG., et al. Practices of Zoology: Study and diversity of the Poriferous. [Internet]. 2011 [cited 2018 Nov 20]. Available from: <http://www.ucm.es/data/cont/docs/568-2013-12-16-01-Poriferos.pdf>
- [16] Boury-Esnault N, Rutzler K. Thesaurus of sponge morphology, Smithsonian Contributions to Zoology. [Internet]. 1997 [cited 2018 Nov 20]. Available from: <http://doi.org/10.5479/si.00810282.596>
- [17] Rutzler K., et al. Diversity of sponges (Porifera) from cryptic habitats on the Belize barrier reef near Carrie Bow Cay, Zootaxa. 2014. doi: 10.11646/zootaxa.3805.1.1.
- [18] De Vos L, Donadey C, Vacelet J, et al. Atlas of sponge morphology, Smithsonian Institution Press; 1991.
- [19] Armengol E, Plaza E, Domingo M, et al. UCI Machine Learning Repository: Demospongiae Data Set. [Internet]. [cited 2018 Nov 2]. Available from: <https://archive.ics.uci.edu/ml/datasets/Demospongiae>