

บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

ในบทที่ 2 นี้จะกล่าวถึงทฤษฎีแนวคิดและองค์ความรู้อื่นที่เกี่ยวข้องในการดำเนินการวิจัย ซึ่งจะแบ่งออกเป็น 2 ส่วน คือส่วนที่เป็นทฤษฎี กับส่วนที่เป็นงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังต่อไปนี้

2.1 ทฤษฎี

2.1.1 ความสัมพันธ์

Meronym (Cruse, Alan, *Meaning in Language: An Introduction to Semantics and Pragmatics*, Oxford University Press, Oxford, 2000) คือแนวคิดเกี่ยวกับความสัมพันธ์แบบมีความหมายใช้ในด้านภาษาศาสตร์โดยที่ meronym จะหมายถึงการเป็นส่วนประกอบของบางสิ่งบางอย่าง (Part-of) หรือการเป็นสมาชิกของบางสิ่งบางอย่าง (Member-of) ยกตัวอย่างเช่น ‘นิ้วมือ’ เป็น meronym ของ ‘มือ’ เนื่องจากนิ้วมือเป็นส่วนหนึ่งของมือ หรือ ‘ล้อ’ เป็น meronym ของ ‘รถจักรยาน’ เนื่องจากล้อเป็นองค์ประกอบของรถจักรยาน

ในบางครั้งการพูดคุยกัันหรือการถามตอบจะไม่มีวลี ‘เป็นส่วนหนึ่ง’ หรือ ‘เป็นองค์ประกอบ’ (‘part’) ในการสนทนาแต่จะมีแนวคิด ‘part’ (Morton E. Winston et al., 1987) เช่น

- 1) “Bicycles have wheels.” // “รถจักรยานมีล้อ”
- 2) “Bicycles are made of aluminum.” // “รถจักรยานทำด้วยอลูมิเนียม”

ซึ่งถ้าเปลี่ยนรูปประโยคให้มี ‘part’ จะได้

- 3) “Wheels are parts of bicycles.” // “ล้อเป็นส่วนประกอบของจักรยาน”
- 4) “Bicycles are partly aluminum.” // “จักรยานมีส่วนหนึ่งเป็นอลูมิเนียม”

จากตัวอย่างข้างต้นจะเห็นว่า ‘part’ ในประโยคแรกจะหมายถึงการเป็นส่วนประกอบ (component-integral object) และ ‘part’ ในประโยคที่ 2 จะหมายถึงการเป็นองค์ประกอบ (stuff-object) จึงได้มีการแบ่ง ‘Part-Whole relation’ เป็น 6 หมวด ดังต่อไปนี้

2.1.1.1 Component-Integral Object

ความสัมพันธ์หมวดนี้จะหมายถึงการนำวัตถุแต่ละชิ้นมาประกอบรวมกันเป็นวัตถุขนาดใหญ่ขึ้นมีคุณสมบัติมากขึ้นใช้งานได้หลากหลายขึ้นเช่น

- 1) “Wheels are parts of cars.” // “ล้อเป็นส่วนประกอบของรถยนต์”
- 2) “Chapters are parts of books.” // “บทเป็นส่วนประกอบของหนังสือ”
- 3) “The refrigerator is part of the kitchen.” // “ตู้เย็นเป็นส่วนประกอบของห้องครัว”
- 4) “Belgium is part of NATO.” // “ประเทศเบลเยียมเป็นส่วนหนึ่งขององค์การสนธิสัญญาป้องกันแอตแลนติกเหนือ”

- 1) “Phonology is part of linguistics.” // “สัทวิทยาเป็นสาขาของภาษาศาสตร์”

ส่วนประกอบแต่ละส่วนของตัวอย่างข้างต้นต่างก็เป็นวัตถุที่มีคุณสมบัติของตัวเองอยู่ แต่เมื่อนำมาประกอบกันจะได้เป็นกลุ่มใหม่, องค์กรใหม่, หรือวัตถุชิ้นใหม่ ซึ่งสามารถแยกวัตถุแต่ละชิ้นออกจากกันได้ ซึ่งเมื่อแยกออกจากกันแล้วก็ยังได้วัตถุชิ้นเก่าก่อนนำมาประกอบรวมกัน

2.1.1.2 Member-Collection

ความสัมพันธ์ประเภทนี้จะแตกต่างจากความสัมพันธ์แบบ Component-Integral Object ข้างต้นแม้จะเกิดจากการประกอบกันขึ้นมาก็ตาม เนื่องจากความสัมพันธ์แบบ Member-Collection ไม่ได้อาศัยคุณลักษณะของแต่ละองค์ประกอบเช่น

- 1) “A tree is part of forest.” // “ต้นไม้เป็นส่วนประกอบของป่า”
- 2) “A juror is part of a jury.” // “ตุลาการเป็นส่วนประกอบของคณะตุลาการ”
- 3) “This ship is part of a fleet.” // “เรือลำนี้เป็นสมาชิกของกองทัพเรือ”

แต่ความสัมพันธ์แบบ Member-Collection ก็แตกต่างจากคลาสเช่นกัน เพราะความสัมพันธ์ในคลาสไม่ใช่ความสัมพันธ์แบบ Meronymy เพราะใช้ ‘part’ แทนไม่ได้เช่น

- 1) “The Nile is a river.” // “ไนล์คือแม่น้ำ”
- 2) “Fido is a dog.” // “ฟีโดคือสุนัข”

จะเห็นว่าความสัมพันธ์ในคลาสจะเป็นแบบความคล้ายคลึงกันของแต่ละสมาชิก กล่าวคือ มีคุณสมบัติพื้นฐานเหมือนกัน ในขณะที่ความสัมพันธ์แบบ Member-Collection จะเป็นแบบการอยู่ร่วมกัน, อยู่ใกล้ชิดกัน หรือเป็นการติดต่อเชื่อมโยงกันทางสังคม ดังจะเห็นว่า ‘สมาชิกของป่าคือต้นไม้’ ต้นไม้ในที่นี้จะหมายถึงต้นไม้ที่อยู่ใกล้หรืออยู่ติดกับต้นไม้อื่นๆ ซึ่งอีกความหมายหนึ่งของความสัมพันธ์ประเภทนี้ก็คือการรวมกลุ่ม (groups)

2.1.1.3 Portion-Mass

ความสัมพันธ์ของบางส่วนของวัตถุหรือมิติทางกายภาพ จะแตกต่างจาก 2 ความสัมพันธ์ข้างต้นนั่นคือแต่ละองค์ประกอบที่ประกอบกันด้วยความสัมพันธ์ประเภทนี้จะเหมือนกันเช่น

- 1) “This slice is part of a pie.” // “พายชิ้นนี้เป็นส่วนหนึ่งของก้อนพาย”
- 2) “A yard is part of a mile.” // “หลาเป็นส่วนประกอบของไมล์”
- 3) “This hunk is part of my clay.” // “ดินก้อนนี้เป็นส่วนหนึ่งของดินของฉัน”

จากตัวอย่างพบว่า ทุกๆ ชิ้นส่วนของแต่ละองค์ประกอบจะต้องเหมือนกันเมื่อนำมาประกอบกันก็จะได้ชิ้นที่ใหญ่กว่าเดิม โดยลักษณะของPortion-Mass มักจะเป็นหน่วยการวัดพื้นฐานเช่น นิ้ว, เมตร, ปอนด์, ชั่วโมง โดยความสัมพันธ์แบบ Portion-Mass มักจะใช้ตัวดำเนินการทางคณิตศาสตร์ (บวก, ลบ, คูณ,หาร)

2.1.1.4 Stuff-Object

ความสัมพันธ์ประเภทนี้จะมีความสัมพันธ์ขององค์ประกอบที่เป็นส่วนหนึ่ง(‘is partly’)ของวัตถุ แต่ไม่เหมือนความสัมพันธ์แบบ Component-Integral Object ดังตัวอย่าง

- 1) “A martini is partly alcohol.” // “มาร์ตินีมีส่วนหนึ่งเป็นแอลกอฮอล์”
- 2) “The bike is partly steel.” // “จักรยานมีส่วนหนึ่งเป็นเหล็กกล้า”
- 3) “Water is partly hydrogen.” // “น้ำมีส่วนหนึ่งเป็นไฮโดรเจน”

จะเห็นว่าลักษณะของความสัมพันธ์ที่ใช้‘is partly’จะไม่สามารถแยกส่วนแอลกอฮอล์ออกจากมาร์ตินี หรือ แยกส่วนเหล็กกล้าออกจากจักรยาน หรือ แยกส่วนไฮโดรเจนออกจากน้ำ เหมือนกับการแยกวัตถุในความสัมพันธ์แบบ Component-Integral Object

‘is partly’จะหมายถึง‘ทำด้วย’,‘ผลิตด้วย’ หรือ ‘made of’ โดยที่ลักษณะของ ‘made of’ เป็นการสร้างวัตถุ, ผลิตวัตถุ เมื่อเราต้องการแยกองค์ประกอบของวัตถุจะทำได้ด้วยวิธีทางกายภาพเหมือนกับการแยกวัตถุในความสัมพันธ์แบบ Component-Integral Object

พบว่าลักษณะของ ‘Stuff-Object’ เมื่อพิจารณาด้วยตาเปล่าจะเป็นวัตถุที่เป็นเนื้อเดียวกัน เช่น เหล็กกล้า, เครื่องดื่มแอลกอฮอล์ เป็นต้น แต่เมื่อพิจารณาถึงระดับโมเลกุลหรืออะตอม วัตถุเกือบทุกชนิดก็จะเป็น‘Stuff-Object’ได้

ในบางครั้งการพิจารณาวัตถุที่ซับซ้อนว่าเป็นวัตถุชนิดไหน หรือมีความสัมพันธ์ขององค์ประกอบเป็นอะไร เช่น ‘จักรยาน’ จะเห็นว่าจักรยานมีองค์ประกอบหลายชนิดเมื่อแยกออกมาจากจักรยาน จักรยานนั้นก็ยิ่งเรียกว่าจักรยาน หรือ ‘สลัด’ เมื่อเอามะเขือเทศออกจากสลัด สลัดก็ยังคงเป็นสลัด ลักษณะของวัตถุแบบนี้จะไม่ใช่ ‘Stuff-Object’ แต่ถ้าวัตถุนั้นขาดองค์ประกอบบางอย่าง

1) “Alcohol is a constituent of wine.” // “แอลกอฮอล์เป็นส่วนผสมของไวน์”

2) “Tomato is an ingredient of salad.” // “มะเขือเทศเป็นส่วนผสมของสลัด”

จะเห็นว่าทั้ง แอลกอฮอล์ และ มะเขือเทศ ต่างก็เป็นเป็นส่วนผสมของวัตถุ โดย มะเขือเทศต้องผ่าน ‘การจัดเตรียม’ ก่อนนำไปปรุงเป็นสลัด แต่แอลกอฮอล์ไม่ต้องผ่านขั้นตอน ‘การจัดเตรียม’ ซึ่งใน ‘Part-Whole relation’ ทั้ง 6 หมวดนี้ ‘ingredient’ จะหมายถึง ‘component’ นั่นคือ แอลกอฮอล์เป็น Stuff-Object และมะเขือเทศเป็น Component-Integral Object

2.1.1.5 Feature-Activity

ความสัมพันธ์แบบนี้จะเป็นการใช้ ‘part’ ในคุณลักษณะของกิจกรรมและการดำเนินการ เช่น

1) “Paying is part of shopping.” // “การจ่ายเงินเป็นส่วนหนึ่งของการซื้อสินค้า”

2) “Ovulation is part of the menstrual cycle.” // “การตกไข่เป็นส่วนหนึ่งของรอบประจำเดือน”

3) “Bidding is part of playing bridge.” // “การบิ๊ดเป็นส่วนหนึ่งของเกมบริดจ์”

แต่ความสัมพันธ์แบบ Feature-Activity จะไม่สามารถอยู่ในรูป ‘X has Y’ ได้

2.1.1.6 Place-Area

ความสัมพันธ์แบบนี้จะเป็นความสัมพันธ์ของพื้นที่(area) กับสถานที่(place) เช่น

1) “The Everglades are part of Florida.” // “Everglades เป็นส่วนหนึ่งของรัฐฟลอริดา”

2) “An oasis is a part of a desert.” // “โอเอซิสเป็นส่วนหนึ่งของทะเลทราย”

3) “The baseline is a part of a tennis court.” // “เส้นหลังเป็นส่วนหนึ่งของสนามเทนนิส”

Place-Area จะมีลักษณะคล้าย Member-Collection ตรงที่ไม่ได้อาศัยคุณลักษณะของแต่ละองค์ประกอบเพื่อมาประกอบกันแต่จะเป็นตำแหน่งของสถานที่ (place) ในพื้นที่ (area) โดยที่ไม่สามารถแบ่งแยก สถานที่ (place) ออกจากพื้นที่ (area) ได้

2.1.2 อาณาจักรพืช(Kingdom Plantae)

พืชเป็นสิ่งมีชีวิตที่มีกำเนิดขึ้นมาแล้วไม่ต่ำกว่า 400 ล้านปี มีหลักฐานหลายอย่าง ที่ทำให้เชื่อว่า พืชมีวิวัฒนาการมาจากสาหร่ายสีเขียว กลุ่ม Charophytes โดยมีการปรับตัวจากสภาพที่เคยอยู่

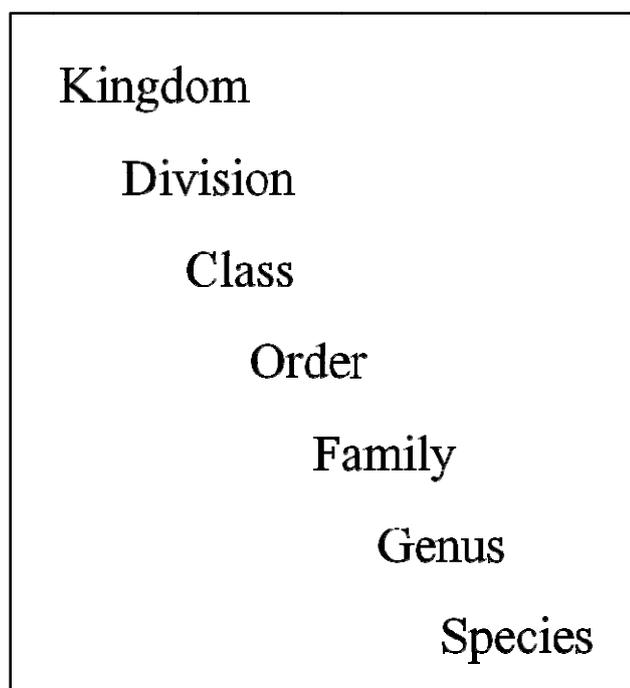
ในน้ำ ขึ้นมาอยู่บนบก ด้วยการสร้างคุณสมบัติต่างๆ ที่เหมาะสมขึ้นมา เช่น มีการสร้างคิวติน(cutin) ขึ้นมาปกคลุมผิว ของลำต้นและใบเรียกว่า คิวทิเคิล (cuticle) เพื่อป้องกันการสูญเสียน้ำ และการเกิด สโทมาตา (stomata) เพื่อทำหน้าที่ระบายน้ำ และแลกเปลี่ยนก๊าซ เป็นต้น

พืชมีโครงสร้างที่ประกอบขึ้นด้วยหลายเซลล์ที่มารวมกลุ่มกันเป็นเนื้อเยื่อที่ทำหน้าที่ เฉพาะ อย่างเซลล์ของพืชมีผนังเซลล์ที่มีสารประกอบ เซลลูโลส (cellulose) เป็นองค์ประกอบ ที่พบ เป็นส่วนใหญ่ พืชทุกชนิดที่คุณสมบัติที่สามารถสร้างอาหารได้เอง จากกระบวนการ สังเคราะห์ด้วย แสง โดยบทบาทของรงควัตถุ คลอโรฟิลล์(chlorophyll a & b) ที่อยู่ในคลอโรพลาสต์เป็นสำคัญ รงควัตถุหลักที่พบได้ในเซลล์พืช จะเหมือนกับที่พบในเซลล์ ของสาหร่ายสีเขียว ได้แก่ คลอโรฟิลล์ เอ คลอโรฟิลล์บี และแคโรทีนอยด์ นอกจากนี้ พืชยังสะสมอาหารในรูปของแป้ง (starch)

คุณสมบัติพื้นฐานที่แสดงว่าสิ่งมีชีวิตประเภทนั้นคือพืช มี 3 ประการ คือ

- 1) พืชต้องมีคลอโรฟิลล์ (chlorophyll)
- 2) พืชต้องมีผนังเซลล์ที่มีสารประกอบ เซลลูโลส (cellulose) เป็นองค์ประกอบ
- 3) พืชไม่สามารถเคลื่อนที่ได้เอง

การจำแนกสิ่งมีชีวิตในอาณาจักรพืช ทางชีววิทยาจะเรียกว่า อนุกรมวิธานของพืช (taxonomy) ซึ่งจะมีลำดับดังต่อไปนี้



ภาพที่ 2.1 แสดงอนุกรมวิธานพืช

ในทางชีววิทยา(แบบเรียนชีววิทยาของ สสวท.)ได้จำแนกสิ่งมีชีวิตในอาณาจักรพืช ออกเป็นกลุ่มใหญ่ที่เรียกว่า ดิวิชัน(division) จำนวน 9 ดิวิชัน ดังต่อไปนี้

2.1.2.1 ดิวิชันไบรโอไฟตา

เรียกโดยทั่วไปว่า ไบรโอไฟต์ (bryophyte) มีทั้งสิ้นประมาณ 16,000 ชนิด พืชในดิวิชันนี้มีขนาดเล็ก มีโครงสร้างง่าย ๆ ยังไม่มีราก ลำต้นและใบที่แท้จริง ชอบอาศัยอยู่ตามที่ชุ่มชื้น การสืบพันธุ์แบบอาศัยเพศยังต้องอาศัยน้ำสำหรับให้สเปิร์มที่มีแฟลกเจลลา (flagella) ว่ายไปผสมกับไข่ ต้นที่พบเห็นโดยทั่วไปคือแกมีโทไฟต์ (มีแกมีโทไฟต์เด่น) รูปร่างลักษณะมีทั้งที่เป็นแผ่นหรือแทลลัส (thallus) และคล้ายลำต้นและใบของพืชชั้นสูง (leafy form) มีไรซอยด์ (rhizoid) สำหรับยึดต้นให้ติดกับดินและช่วยดูดน้ำและแร่ธาตุ มีส่วนคล้ายใบ เรียก phylloid และส่วนคล้ายลำต้นเรียกว่า cauloid แกมีโทไฟต์ของไบรโอไฟต์มีสีเขียวเพราะมีคลอโรพิลล์สามารถสร้างอาหารได้เอง ทำให้อยู่ได้อย่างอิสระ เมื่อแกมีโทไฟต์เจริญเต็มที่สร้างเซลล์สืบพันธุ์คือสเปิร์มและไข่ต่อไป ภายหลังการปฏิสนธิของสเปิร์มและไข่จะได้ไซโกตซึ่งแบ่งตัวเจริญต่อไปเป็นเอ็มบริโอและสปอร์โรไฟต์ตามลำดับ สปอร์โรไฟต์ของไบรโอไฟต์มีรูปร่างลักษณะง่าย ๆ ไม่สามารถอยู่ได้อย่างอิสระจะต้องอาศัยอยู่บนแกมีโทไฟต์ตลอดชีวิต พืชในดิวิชันนี้สร้างสปอร์เพียงชนิดเดียว ตัวอย่างของพืชในกลุ่มนี้ได้แก่ ลิเวอร์เวิร์ต (liverwort), ฮอร์นเวิร์ต (hornwort) และ มอส (moss)

2.1.2.2 ดิวิชันไซโลไฟตา

พืชในดิวิชันนี้ที่พบได้ในประเทศไทย ได้แก่ Psilotum รู้จักกันในชื่อไทยว่า หวายทะนอยสปอโรไฟต์ของพืชนี้มีรูปร่างลักษณะง่าย ๆ คือมีแต่ลำต้นยังไม่มีรากและใบ ลำต้นมีลักษณะเป็นไม้เนื้ออ่อนขนาดสูงประมาณ 20 – 30 เซนติเมตร ขึ้นอยู่ตามพื้นดิน (terrestrial) หรือเกาะติดกับต้นไม้อื่น (epiphyte) ลำต้นแบ่งออกเป็น 2 ส่วน คือ ส่วนที่อยู่ใต้ดินเป็นลำต้นชนิดไรโซม (rhizome) มีสีน้ำตาล และมีไรซอยด์ทำหน้าที่ดูดน้ำและแร่ธาตุ ลำต้นส่วนที่อยู่เหนือพื้นดิน (acrial stem) มีสีเขียว มีลักษณะเป็นเหลี่ยม ลำต้นส่วนนี้ทำหน้าที่สังเคราะห์แสง ทั้งลำต้นใต้ดินและลำต้นเหนือพื้นดิน แตกกิ่งเป็น 2 แฉก (dichotomous branching) ที่ส่วนของลำต้นเหนือพื้นดินมีระยางค์เล็กๆ (appendage) ยื่นออกมาเห็นได้ทั่วไป สปอโรไฟต์ที่เจริญต้นที่จะสร้างอับสปอร์ที่มีรูปร่างเป็น 3 พู ที่ซอกของระยางค์บนลำต้นเหนือพื้นดิน อับสปอร์สร้างสปอร์ชนิดเดียว แกมีโทไฟต์มีขนาดเล็ก สีน้ำตาลไม่มีคลอโรพิลล์ รูปร่างเป็นแท่งทรงกระบอก แตกแขนงได้

2.1.2.3 ดิวิชันไลโลไฟตา

สปอโรไฟต์ของพืชดิวิชันนี้มีราก ลำต้น และใบครบทุกส่วน มีลักษณะเป็นไม้เนื้ออ่อนที่มีขนาดไม่ใหญ่มากนัก พวกที่เจริญอยู่บนพื้นดิน อาจมีลำต้นตั้งตรงหรือทอดนอน บางชนิดอาศัยเกาะบนต้นไม้อื่น ลำต้นแตกกิ่งเป็น 2 แฉก ใบมีขนาดเล็ก เป็นใบแบบไมโครพิลล์ (microphyll) คือ

Lycopodium รู้จักในชื่อไทยว่า ซ้องนางกลี สร้อยสุกรม สามร้อยยอด และหางสิงห์เป็นต้น ที่พบในปัจจุบันมีประมาณ 200 ชนิด ใบในขนาดเท่า ๆ กันเรียงตัวเป็นเกลียวโดยรอบลำต้น และกิ่ง เป็นพืชที่สร้างสปอร์ชนิดเดียว แกมีโทไฟต์มีขนาดเล็ก บางชนิดมีคลอโรฟิลล์เจริญอยู่บนพื้นดิน บางชนิดไม่มีคลอโรฟิลล์เจริญอยู่ใต้ดิน

2.1.2.4 คิวซันสปีโนไฟตา

คิวซันสปีโนไฟตา (Division Sphenophyta) พืชที่มีท่อลำเลียงในคิวซันนี้มีเพียง วงศ์เดียว คือ Equisetaceae แกมีโทไฟต์มีขนาดเล็ก เจริญอยู่ใต้ดิน สปอโรไฟต์มีขนาดใหญ่ อายุยืน มีซิกติกาลำต้นเป็นข้อปล้องชัดเจน ปล้องเป็นร่องและสัน ข้อมีใบแบบไมโครฟิลล์อยู่รอบข้อเรียงแบบ whorl เป็น homosporous plant โดยสปอโรแองเจียมเจริญอยู่บนโครงสร้างที่เรียกว่าสปอโรแองจิอพออร์ (sporangiophore) ตัวอย่างของพืชในกลุ่มนี้ได้แก่ หญ้าถอดปล้อง (equisetum)

2.1.2.5 คิวซันเทอโรไฟตา

พืชคิวซันนี้มีชื่อทั่วไปว่า เฟิร์น (fern) มีจำนวนมากที่สุดในบรรดาโรไฟต์ของเฟิร์นมีราก ลำต้นและใบเจริญดี เฟิร์นส่วนใหญ่มีลำต้นใต้ดิน ใบของเฟิร์นเรียกว่า ฟรอนด์ (frond) เป็นส่วนที่เห็นเด่นชัด มีขนาดใหญ่เป็นใบแบบเมกะฟิลล์ (megaphyll) มีรูปร่างลักษณะเป็นหลายแบบ มีทั้งที่เป็นใบเดี่ยว (simple leaf) และใบประกอบ (compound leaf) ใบอ่อนของเฟิร์นมีลักษณะพิเศษคือ จะม้วนเป็นวง (circinate venation) สปอโรไฟต์ที่เจริญเต็มที่จะสร้างอับสปอร์ ซึ่งมารวมกลุ่มอยู่ที่ด้านใต้ใบ แต่ละกลุ่มของอับสปอร์เรียกว่า ซอรัส (sorus) เฟิร์นส่วนใหญ่สร้างสปอร์ชนิดเดียว ยกเว้นเฟิร์นบางชนิดที่อยู่ในน้ำ และที่ชื้นแฉะ ได้แก่ จอกหูหนู แหนแดง และผักแว่นมีการสร้างสปอร์ 2 ชนิด

แกมีโทไฟต์ของเฟิร์นที่สร้างสปอร์ชนิดเดียว มีลักษณะเป็นแผ่นแบนบางสีเขียว (มีคลอโรฟิลล์) ด้านล่างมีไรซอยด์ ส่วนใหญ่มักมีรูปร่างคล้ายรูปหัวใจ (prothallus)

2.1.2.6 คิวซันโคนิเฟอโรไฟตา

เป็นจิมโนสเปิร์มที่มีจำนวนมากที่สุด มีหลายสกุลด้วยกัน ที่รู้จักกันดีคือ Pinus ได้แก่ สนสองใบ และสนสามใบ เป็นต้น สปอโรไฟต์ของ Pinus มีลักษณะเป็นไม้ยืนต้นขนาดค่อนข้างใหญ่ และแตกกิ่งก้านสาขาจำนวนมาก ใบมีขนาดเล็ก รูปร่างคล้ายเข็ม อยู่รวมกันเป็นกลุ่ม สปอโรไฟต์ที่เจริญเต็มที่จะสร้างโคนเพศผู้ที่มีขนาดเล็กและโคนเพศเมียที่มีขนาดใหญ่บนต้นเดียวกัน

2.1.2.7 คิวซันไซแคโดไฟตา

พืชคิวซันนี้มีอยู่ประมาณ 60 ชนิด ตัวอย่างที่รู้จักกันดีคือ พวงปรง (Cycas) สปอโรไฟต์มีลำต้นอวบ เตี้ย และมักไม่แตกแขนง มีใบเป็นใบประกอบแบบขนนกขนาดใหญ่ เกิดเป็นกระจุกที่บริเวณยอดของลำต้น ใบย่อยมีรูปร่างเรียวยาว และแข็ง สปอโรไฟต์ที่เจริญเต็มที่จะสร้างโคนเพศผู้และโคนเพศเมีย แยกตัวกัน

2.1.2.8 คิวซันกิงโกไฟตา

ปัจจุบันมีเพียงชนิดเดียวคือ Ginkgo biloba หรือแปะก๊วย เป็นพืชที่ขึ้นอยู่ในเขตอบอุ่น เช่น ในประเทศจีน สปอโรไฟต์มีลักษณะเป็นไม้ยืนต้นขนาดใหญ่ แตกกิ่งก้านสาขาเป็นจำนวนมาก ใบมีรูปร่างคล้ายพัด สปอโรไฟต์ที่เจริญเติบโตเต็มที่จะสร้างโคนเพศผู้และโคนเพศเมียแยกตัวกัน

2.1.2.9 คิวซันแอนโทไฟตา

พืชในกลุ่มนี้จะเป็นกลุ่มพืชชั้นสูง หรือพืชที่มีดอก ผล และสืบพันธุ์ด้วยเมล็ดแบ่งออกได้เป็น 2 คลาส คือ

1. คลาสไดคอตีเลโดเนส (Class Dicotyledones) ได้แก่ พืชใบเลี้ยงคู่ทั้งหมด มีอยู่ประมาณ 170,000 ชนิด ลักษณะทั่วไปคือ มีใบเลี้ยง 2 ใบ เส้นใบเป็นร่างแห รากเป็นระบบรากแก้ว และส่วนประกอบของดอก (เช่น กลีบเลี้ยง กลีบดอก) มีจำนวนเป็น 4-5 หรือทวีคูณของ 4-5

2. คลาสมอโนคอตีเลโดเนส (Class Monocotyledones) ได้แก่ พืชใบเลี้ยงเดี่ยวทั้งหมดมีอยู่ประมาณ 60,000 ชนิด ลักษณะทั่วไปคือ มีใบเลี้ยงใบเดียว ใบมีเส้นใบเรียงตัวแบบขนาน รากเป็นระบบรากฝอย ส่วนประกอบของดอกมีจำนวนเป็น 3 หรือทวีคูณของ 3

จากทั้ง 9 คิวซัน จะพบว่า มี 3 คิวซันที่พืชในกลุ่มนี้มีจำนวนน้อยมากจนใกล้สูญพันธุ์คือ คิวซันไซโลไฟตา คิวซันสไฟโนไฟตา และ คิวซันกิงโกไฟตา ซึ่งในการวิจัยนี้จะไม่นำมารวมด้วย

2.1.3 สารประกอบทุติยภูมิ (Secondary Metabolites)

ตัวตั้งแต่วัยเด็กกาลมนุษย์มีการใช้สารเคมีจากพืช ในขั้นต้นได้ใช้เพื่อการล่าและการทำลายชีวิต สารดังกล่าวได้แก่ ทูโบคูเรรีน (tubocurarine) ที่ส่วนใหญ่ได้จากพืชเถา Chondrodendron tomentosum วงศ์ Menispermaceae ในป่าแถบลุ่มน้ำอะเมซอน สารดังกล่าวหยุดการส่งผ่านของกระแสประสาทไปยังกล้ามเนื้อ เมื่อใช้อาบปลายลูกธนูจะทำให้สัตว์ที่ถูกยิงด้วยลูกธนูดังกล่าวเกิดเป็นอัมพาตเคลื่อนไหวไม่ได้ ในการผลิตชีวิตมีการใช้เฮมล็อก (Conium maculatum วงศ์ Umbelliferae) ซึ่งทำให้เกิดอัมพาตกล้ามเนื้อตามด้วยการชักและจบชีวิตลงด้วยอัมพาตของระบบทางเดินหายใจ เฮมล็อกมีพิษทั้งต้นเพราะมีอัลคาลอยด์มาก อัลคาลอยด์ที่สำคัญคือ โคนีนิน (conine)

สารประกอบทุติยภูมิ (Secondary Metabolites) คือกลุ่มของสารเคมีที่สร้างโดยพืช สัตว์ ราหรือแบคทีเรีย ที่ไม่มีความจำเป็นในขั้นวิกฤตต่อสิ่งมีชีวิตผู้ผลิต หากแต่ถูกสร้างโดยขบวนการทางชีวเคมีของผู้ผลิตเป็นสารจำเพาะต่อผู้ผลิตนั้นๆ เป็นสารที่ให้กลิ่น สี หรือสรรพคุณจำเพาะของพืชที่พบในอาหาร ยาและสารพิษต่างๆ จากพืชและมีการกระจายตัวอย่างจำกัด

การใช้สมุนไพรและตัวยาจากธรรมชาติมีประวัติการใช้งานอันยาวนาน การใช้พืชทั้งต้นหรือการเตรียมยาอย่างหายาๆ มีข้อดีหลายประการดังนี้

1. ความไม่แน่นอนของปริมาณสารที่ออกฤทธิ์จากแหล่งเพาะปลูกที่ต่างกัน หรือจากฤดูหนึ่งๆ ไปยังฤดูอื่น รวมถึงส่วนของพืชที่ต่างกันและลักษณะทางกายวิภาคที่ต่างกันด้วย ตัวอย่างได้แก่ เปล้าน้อยพืชที่ปลูกในจังหวัดปราจีนบุรีมีสารออกฤทธิ์น้อยกว่าพืชที่ปลูกในจังหวัดประจวบคีรีขันธ์

2. การสะสมของสารที่ไม่ต้องการซึ่งอาจมีส่วนก่อให้เกิดความเปลี่ยนแปลงต่อสารออกฤทธิ์เช่น มีฤทธิ์ต่างกันหรือมีฤทธิ์แรงขึ้น เช่น ลิวโรซีน (leucosine) เป็น indole alkaloid คล้ายอัลคาลอยด์จากต้นพวงพวยฝรั่ง สารสกัดหยาบอัลคาลอยด์ไม่มีฤทธิ์ชีวภาพในเซลล์มะเร็งเม็ดเลือดขาวแต่สารบริสุทธิ์มีพิษต่อเซลล์มะเร็งนี้ในหลอดทดลองอย่างชัดเจน

3. การสูญเสียสารออกฤทธิ์เนื่องจากความแตกต่างกันของวิธีเก็บเกี่ยวเก็บรักษาและการเตรียมวัตถุดิบเพื่อการใช้ในการรักษา ตัวอย่างได้แก่ต้นพิเวอร์พีจะรักษาสารสำคัญไว้ได้มากที่สุด ถ้าอบพืชแห้งทันทีที่ 60 องศาเซลเซียสในระยะเวลาอันสั้น ถ้าตากในร่มที่อุณหภูมิห้องต้องใช้เวลาดตาก 3-7 วันและจะได้ปริมาณพาร์ทีโนโลลด์ไม่เต็มที่

ข้อดีของการสกัดแยกสารผลิตภัณฑ์ธรรมชาติให้ได้สารออกฤทธิ์มีดังนี้

1. สารออกฤทธิ์ในรูปของสารบริสุทธิ์สามารถนำจ่ายได้ในปริมาณที่ถูกต้องและสามารถวัดให้เท่ากันได้

2. การรู้จักองค์ประกอบทางเคมีและโครงสร้างสารบริสุทธิ์สามารถนำไปสู่การพัฒนาวิธีทดสอบทางชีวภาพหาสารอื่นที่มีฤทธิ์เดียวกันจากสิ่งมีชีวิตในกลุ่มเดียวกันหรือกลุ่มที่ใกล้เคียงกันได้

3. การทราบ โครงสร้างที่แน่นอนของสารออกฤทธิ์อาจเป็นการชี้แนะให้เกิดการสังเคราะห์สารดังกล่าวหรือการเปลี่ยนแปลงองค์ประกอบโมเลกุลเล็กน้อยให้ได้สารที่มีฤทธิ์มากขึ้นหรือมีพิษต่ำลง ตัวอย่างได้แก่ สารโพโดฟิลโลทอกซิน (podophyllotoxin) ซึ่งเป็นลิแกนด์ในยางของ Podophyllumpeltatum มีฤทธิ์ใช้รักษามะเร็งได้แต่มีผลข้างเคียงมาก ขณะที่ยาสังเคราะห์อีโทโปไซด์ (etoposide) และเทนิโปไซด์ (teniposide) มีสูตรโครงสร้างหลักเหมือนกันต่างกันที่อนุพันธ์มีฤทธิ์แรงกว่าและมีผลข้างเคียงน้อยกว่าสารที่ได้จากธรรมชาติมาก

4. สามารถศึกษาถึงเส้นทางชีวสังเคราะห์ของสารนั้นๆ ทำให้รู้จักสารตั้งต้นของมัน อาจมีผลทำให้ใช้ส่วนของพืชที่มีสารสำคัญน้อยลงโดยใช้สารอื่นในเส้นทางชีวสังเคราะห์ทดแทน หรือนำสารในเส้นทางเมตาบอลิซึมมาเป็นสารตั้งต้นในการสังเคราะห์ด้วยที่ต้องการ ตัวอย่าง ได้แก่ ยาพาชีทาเซลหรือแท็กซอล เป็นต้น

ซึ่งในงานวิจัยนี้ได้แบ่งประเภทของสารประกอบทุติยภูมิออกเป็น 5 กลุ่มใหญ่ตาม โครงสร้างหลักได้แก่

1. อัลคาลอยด์ (Alkaloids)
2. เทอร์พีนอยด์ (Terpenoid)
3. ไกลโคไซด์ (Glycoside)
4. ฟีนอล (Phenols)
5. โพลีคีไทด์ (Polyketides)

2.1.4 ตัวจำแนกประเภทเนอ็ฟเบย์ (Naïve Bayes Classifier)

ตัวจัดประเภทเนอ็ฟเบย์ (Naïve Bayes classifier, NB) (Mitchell 1997) หรือ ตัวเรียนรู้ NB เป็นวิธีการเรียนรู้ที่นิยมใช้กันมาก และเป็นการเรียนรู้ที่อยู่บนพื้นฐานของความน่าจะเป็น (Probability) กับข้อมูลที่สังเกต (Observed Data) ตามที่ Mitchell T.M., (1997) ได้กล่าวว่าตัวจัดประเภท NB สามารถประยุกต์ใช้กับงานเรียนรู้ที่ซึ่งแต่ละตัวอย่าง x (Instance x) ได้ถูกอธิบายโดยการเชื่อมโยงค่าแอททริบิวท์ (Attribute Values) ต่างๆ และที่ซึ่งฟังก์ชันเป้าหมาย (Target Function, $f(x)$) สามารถแสดงค่าคลาส (Class Value, v) จาก คลาสไฟไนท์เซต (Class Finite Set, V) ดังนั้นเซตของตัวอย่างการเรียนรู้ของฟังก์ชันเป้าหมายได้ถูกกำหนดไว้ให้ และเมื่อมีตัวอย่างใหม่เกิดขึ้นก็สามารถอธิบายได้ คือบอกค่าคลาสได้ด้วยทูปเพิล (Tuple) ของค่าแอททริบิวท์ $\langle a_1, a_2, \dots, a_n \rangle$ นั่นคือตัวเรียนรู้ทำนายค่าเป้าหมายหรือการจัดแบ่งประเภทสำหรับตัวอย่างใหม่ที่เข้ามา

แนวทางเบย์ที่จะจัดประเภทให้กับตัวอย่างใหม่ที่เข้ามานั้นเป็นการกำหนดค่าเป้าหมายที่มีโอกาสเป็นไปได้มากที่สุด หรือที่เรียกว่า v_{maximum} a posterior (v_{MAP}) เมื่อกำหนดค่าแอททริบิวท์ต่างๆ ให้ $\langle a_1, a_2, \dots, a_n \rangle$ ที่ใช้อธิบายตัวอย่าง ดังแสดงในสมการ(2) และ (3)

$$v_{\text{MAP}} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, \dots, a_n) \quad (1)$$

$$v_{\text{MAP}} = \underset{v_j \in V}{\operatorname{argmax}} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (2)$$

ตัวจัดประเภท NB ดำเนินงานบนพื้นฐานของข้อสมมติฐานแบบง่าย ๆ ที่มีเงื่อนไขว่าค่าแอททริบิวต์แต่ละแอททริบิวต์ จะต้องเป็นอิสระต่อกันเมื่อกำหนดค่าเป้าหมายไว้ให้ กล่าวคือข้อสมมติฐานเป็นการกำหนดค่าเป้าหมายของตัวอย่าง (คือคลาสของตัวอย่าง) ฉะนั้นความน่าจะเป็นของการสังเกตการเชื่อมโยงกันของ a_1, a_2, \dots, a_n คือผลคูณของค่าความน่าจะเป็นของแอททริบิวต์ต่างๆ ดังนั้นตัวจัดประเภท NB (v_{NB}) สามารถแสดงได้ดังต่อไปนี้

$$V_{\text{NB}} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j) \quad (3)$$

สำหรับงานวิจัยนี้ เราได้ประยุกต์ใช้ตัวจัดประเภท NB ที่เป็นสมการ (4) สำหรับการเรียนรู้แยกประเภทของแนวความคิดของสารเคมีและแนวความคิดของพืชในแต่ละประโยคดังต่อไปนี้

StuffRelationForNaturalProduct

$$\begin{aligned} &= \underset{\text{class} \in \text{Class}}{\operatorname{argmax}} P(\text{class} | c_1, c_2, \dots, c_{\text{max1}}, s_1, s_2, \dots, s_{\text{max2}}, a_1, a_2, \dots, a_{\text{max3}}) \\ &= \underset{\text{class} \in \text{Class}}{\operatorname{argmax}} P(\text{class}) \prod_{\text{num}=1}^{\text{max1}} P(c_{\text{num}}) \prod_{\text{num}=1}^{\text{max2}} P(s_{\text{num}}) \prod_{\text{num}=1}^{\text{max3}} P(a_{\text{num}}) \end{aligned} \quad (4)$$

เมื่อ $\text{Class} = \{\text{"yes"}, \text{"no"}\}$

r = ขนาดกรอบหน้าต่างของคำระหว่างแนวคิดสารเคมีกับแนวคิดของพืช (ดูหัวข้อ 4.1)

โดย r เป็นเซตของตัวเลข $\{1, 2, 3, 4, 5\}$ ซึ่งเป็นค่า max3

$r = \text{set of } \{1, 2, 3, 4, 5\}$

สำหรับงานวิจัยนี้ $a_1, a_2, \dots, a_n \in A$, $A =$ เซตของคำที่มีความถี่สูงซึ่งอยู่ระหว่างแนวคิดของสารเคมี(c_i)กับแนวคิดของพืช(s_j)

$c_1, c_2, \dots, c_i \in C$, $C =$ เซตของแนวคิดของสารเคมี

$s_1, s_2, \dots, s_j \in S$, $S =$ เซตของแนวคิดของพืช

$i = \{1, 2, \dots, m\}$

$j = \{1, 2, \dots, k\}$

จากการศึกษาข้อมูล $\max_1 = 16$ $\max_2 = 14$ $\max_3 = r$ หรือ n (n คือจำนวนคำที่อยู่ระหว่างคำที่เป็นแนวคิดของสารเคมี และคำที่เป็นแนวคิดของพืช) โดย r มีขนาด 3,4, และ 5 และ n มีค่ามากที่สุดคือ 10

เมื่อตัวแปร “Class” เป็นไฟไนท์เซต (Finite Set) ของประเภทความสัมพันธ์แบบสตัพฟ์สำหรับ แนวคิดของสารเคมีและความคิดของพืช (StuffRelationForNaturalProduct) โดยใช้ฟีเจอร์ (Feature) ต่างๆที่ประกอบด้วย 3 กลุ่มดังนี้ ฟีเจอร์กลุ่มแรกคือคำหรือสมาชิก (Element) ของเซต C ซึ่งเป็น เซตของแนวคิดของสารเคมี $\langle c_1, c_2, \dots, c_r \rangle$ กลุ่มที่ 2 คือคำหรือสมาชิก ของเซต S ซึ่งเป็น เซตของแนวคิดของพืชที่เป็น Natural Source ในระดับ Genus $\langle s_1, s_2, \dots, s_j \rangle$ และกลุ่มที่ 3 ซึ่งแบ่งออกเป็น 2 แบบคือ แบบที่เป็นคำหรือสมาชิกต่างๆใน เซต A ซึ่งเป็นเซตของคำที่อยู่ระหว่างแนวคิดของสารเคมีและแนวคิดของพืช โดยจะใช้คำทั้งหมดเหล่านี้ $\langle a_1, a_2, \dots, a_r \rangle$ ที่มีความถี่มากที่สุด ภายใต้ขนาดกรอบหน้าต่าง r ของคำต่างๆระหว่างแนวคิดสารเคมีกับแนวคิดของพืช แบบที่ 2 เป็นคำทั้งหมด (n) ที่อยู่ระหว่างแนวคิดของสารเคมีและแนวคิดของพืช และเป็นคำใน เซต A $\langle a_1, a_2, \dots, a_n \rangle$

2.2 งานวิจัยที่เกี่ยวข้อง

R. Girju ใช้เทคนิคการเรียนรู้ของเครื่องชนิด ID3(C4.5) โดยใช้เงื่อนไขในรูปแบบโครงสร้างทางไวยากรณ์ (Lexico-syntactic pattern) เพื่อหาความสัมพันธ์แบบ Part-Whole จากบทความ LA Times ของคลังข้อความ TREC-9 โดยแบ่งความสัมพันธ์แบบ Part-Whole เป็น 3 ประเภทตาม WordNet ได้แก่ Member-of (เช่น UK IS-MEMBER-OF NATO), Stuff-of (เช่น carbon IS-STUFF-OF coal) และ Part-of (เช่น leg IS-PART-OF table) จาก 10,000 ประโยคตัวอย่างพบว่า ได้ความถูกต้องเป็น ความแม่นยำ (precision) 83% และ การเรียกคืน (recall) 98%

P. Pantel และ M. Pennacchiotti เสนอ Espresso algorithm ซึ่งเป็นอัลกอริทึมการเรียนรู้แบบมีการสอนแบบอ่อนๆ (weakly-supervised algorithm) เพื่อใช้สกัดความสัมพันธ์แบบ Is-a (Is-a relation) และความสัมพันธ์แบบ Part-of (Part-of relation) Espresso algorithm สามารถรองรับข้อมูลปริมาณมหาศาลที่ดาวน์โหลดมาจากเว็บด้วยการใช้รูปแบบทั่วไปที่นำเชื่อถือในการเก็บข้อมูลซึ่งโดยทั่วไปผลที่ได้จากรูปแบบที่นำเชื่อนี้จะให้ precision ที่สูงแต่ recall จะต่ำมากเช่น “X consists of Y” ในกรณีของความสัมพันธ์แบบ Part-of (Part-of relation)

โดยที่ P. Pantel และ M. Pennacchiotti ได้ทำการทดลองโดยใช้ข้อมูล 2 กลุ่ม คือ คลังข้อความ CHEM และ คลังข้อความ TREC-9 ซึ่งพบว่าจากคลังข้อความ CHEM ได้ความถูกต้องของความสัมพันธ์แบบ Part-of (Part-of relation) เป็น precision 51% และ relative recall เป็น 46 ในขณะที่

และเป็นเครื่องมือในการสกัดความสัมพันธ์จากข้อความใน ส่วนของบทคัดย่อของคลังข้อความ MEDLINE ซึ่งเป็นคลังข้อความที่มีเนื้อหาเกี่ยวกับชีวการแพทย์ โดย ReLex เป็นโมเดลที่ใช้ตรวจจับเหตุการณ์ร่วม (co-occurrences) ของนิพจน์ระบุนามภายใน ประโยคหรือบทคัดย่อและใช้ เซตของกฎอย่างง่าย, ใช้การเครื่องมือในการกำหนดประเภทของคำ (part-of-speech-tagging), noun-phrase-chunking และการขึ้นต่อกันเพื่อหาความสัมพันธ์ของนิพจน์ ระบุนามของยีนกับนิพจน์ระบุนามของโปรตีนซึ่งจะใช้กฎทางภาษาศาสตร์ที่พบมากใน ภาษาอังกฤษดังนี้

- (1) effector-relation-effectee (' α activates β ')
- (2) relation-of-effectee-by-effector ('Activation of α by β ')
- (3) relation-between-effector-and-effectee ('Interaction between α and β ').

โดยได้ความถูกต้อง precision เป็น 80 % และ recall 80 %

G. I. Brown นำเสนอระบบการสกัดความสัมพันธ์โดยใช้ตัวจำแนก support vector machine (SVM) โดยใช้คลังข้อความที่เป็นบทวิจารณ์จาก J.D. Power และ Associates Sentiment 3 ประเภท ดังนี้ 1) ผู้เชี่ยวชาญเป็นผู้เขียน 2) บล็อกเกอร์เป็นผู้เขียน และ 3) ผู้ใช้ทั่วไป โดยคุณลักษณะ (features) ที่ใช้ใน SVM ประกอบไปด้วย 'คำ' ที่เกี่ยวข้องกับค่านามที่ปรากฏในวลี, ประเภทของ นิพจน์ระบุนาม และคลาส token ซึ่งงานนี้ได้เน้นการสกัดระหว่างความแตกต่างของประเภทบท วิจารณ์ โดยได้ความถูกต้องของการสกัดแบบ Part-of เป็น precision 46% โดยเฉลี่ย และ recall เป็น 33% โดยเฉลี่ย

จะเห็นว่าการวิจัยเหล่านี้ดำเนินการบนพื้นฐานที่มีการกำหนดชนิดของคำ (part of speech) และโดยส่วนใหญ่จะสนใจเฉพาะค่านามอย่างเดียว แต่สำหรับงานวิจัยนี้ไม่สามารถ กำหนดชนิดของคำได้อย่างอัตโนมัติ เนื่องจากความซับซ้อนของคำที่เป็นชื่อสารเคมี ต่อไปนี้จะ แสดงให้เห็นถึงความซับซ้อนของชื่อสารเคมี เนื่องจากว่าชื่อเรียกของสารเคมีมีหลากหลายชื่อ เช่น ชื่อทางการค้า, ชื่อสามัญ, ชื่อ IUPAC, ชื่อย่อ, ชื่อที่ใช้สูตรอย่างย่อ ดังตารางที่ 2.3 จะเห็นว่าชื่อของ สารเคมีจะมีสัญลักษณ์ต่างๆ เช่น นขลิจิต "(", "[", "{", "ยัติภังค์ "-", อะพอสโทรฟี่ "'", จุลภาค ";", มหัพภาค "." รวมถึงช่องว่าง เป็นต้น บางครั้งก็จะมีอักษรภาษาอังกฤษที่เป็นตัวโดดและเป็น ตัวพิมพ์ใหญ่

ตารางที่ 2.1 แสดงตัวอย่างของชื่อของสารเคมี

ชื่อของสารเคมี
1-(3-hydroxyphenyl)-3-(4-hydroxy-2,5-dimethoxyphenyl)propane
5-bromo-3-(3'-hydroxy-3'-methylpent-4'-enylidene)-2,4,4-trimethylcyclohexanone
isoliquiritigenin
methyl carnosate
4R-hydroxy-18-normanoyl oxide
N-pentacosanoyl-4,5-dihydroxy-tryptamine
alanine
2'-hydroxy-6,4',6'',4'''-tetramethoxy-[7-O-7'']-bisisoflavone
(7S,8R,1'S,5'S,6'R)-2',8'-3',6'-dihydroxy-5'-methoxy-3,4-methylenedioxy-4'-oxo-8.1',7.5'-neolignan

การที่จะใช้เครื่องมือในการระบุชนิดของคำ โดยเฉพาะส่วนที่เป็นชื่อสารเคมีอย่างอัตโนมัติ นั้นทำได้ เนื่องจากลักษณะต่างๆ ที่ปรากฏอยู่ในชื่อของสารเคมี ดังนั้น งานวิจัยนี้จึงเสนอการสกัดความสัมพันธ์แบบสตัทพ์ โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ที่เป็นเนอิวเบย์ (Naïve Bayes) ด้วยการเปรียบเทียบคำต่างๆ ในเอกสารงานวิจัยทางวิทยาศาสตร์กับฐานข้อมูล NCI-PubChem และ NCI-taxonomy