

การค้นคืนสารสนเทศออนไลน์โดยใช้จีเนติกอัลกอริทึม Online Information Retrieval using Genetic Algorithms

บังอร กลับบ้านเกาะ

สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ลาดกระบัง กรุงเทพฯ 10520

โทร (02) 7372551-4(EXT:802) โทรสาร 3269074 E-Mail:S0067034@kmitl.ac.th

เอื้อน ปิ่นเงิน

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ลาดกระบัง กรุงเทพฯ 10520

โทร (02) 3269969 E-Mail:kpouen@kmitl.ac.th

ABSTRACT -- This article presents an online information retrieval using genetic algorithms to increase information retrieval efficiency. Under vector space model, information retrieval is based on the similarity measurement between query and documents. Documents with high similarity to query are judge more relevant to the query and should be retrieved first. Under genetic algorithms, each query is represented by a chromosome. These chromosomes feed into genetic operator process: selection, crossover, and mutation until we get an optimize query chromosome for document retrieval.

KEY WORDS -- GENETIC ALGORITHMS, INFORMATION RETRIEVAL

บทคัดย่อ -- บทความนี้นำเสนอเกี่ยวกับวิธีการค้นคืนสารสนเทศออนไลน์โดยประยุกต์ใช้จีเนติกอัลกอริทึมเพื่อเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศ ภายใต้เวกเตอร์สเปซโมเดล (vector space model) การค้นคืนสารสนเทศได้นั้นขึ้นอยู่กับความคล้ายคลึง (similarity) ระหว่างเอกสารและควรี เอกสารใดที่มีความคล้ายคลึงกับควรีสูงย่อมแสดงว่าเอกสารนั้นมีความสัมพันธ์กับควรีมากกว่าและควรจะได้รับ การค้นคืนขึ้นมาก่อน ในขั้นตอนของจีเนติกอัลกอริทึมนั้น ควรีจะถูกแทนด้วยโครโมโซม ซึ่งโครโมโซมเหล่านี้จะถูกนำเข้าสู่กระบวนการจีเนติกโอเปอเรเตอร์ต่างๆ อันได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน จนกระทั่งได้โครโมโซมควรีที่เหมาะสมเพื่อนำไปค้นคืนสารสนเทศต่อไป

คำสำคัญ -- จีเนติกอัลกอริทึม, การค้นคืนสารสนเทศ

1. คำนำ

จีเนติกอัลกอริทึมคิดค้นขึ้นโดย John Holland ในปี.ศ. 1975 [1][2] เป็นการนำทฤษฎีวิวัฒนาการของสิ่งมีชีวิตมาประยุกต์ใช้ในงานปัญญาประดิษฐ์ เพื่อใช้สำหรับหาคำตอบที่ดีที่สุด (Optimization) ของปัญหาต่าง ๆ จากจำนวนคำตอบที่เป็นไปได้ทั้งหมดของการแก้ปัญหาานั้น

ปัจจุบันได้มีการประยุกต์ใช้จีเนติกอัลกอริทึม ในงานต่าง ๆ อย่างแพร่หลาย เช่น การใช้จีเนติกอัลกอริทึมในการแก้ปัญหาทางคณิตศาสตร์ การแก้ปัญหาการหาเส้นทางที่มีระยะทางสั้นที่สุด (Traveling Salesman Problem : TSP) [1] [2] การจัดตารางสอน (Timetable Scheduling Problem) [7] [12] การควบคุมหุ่นยนต์ (Robot Control) [8] [10] รวมทั้งการค้นคืนสารสนเทศด้วย [3] [4] [5]

2. งานวิจัยที่เกี่ยวข้อง

จินตคณิตอัลกอริทึมเป็นกระบวนการแก้ปัญหาเอนกประสงค์ที่สามารถนำไปประยุกต์ใช้กับปัญหาประเภทต่าง ๆ ได้ สำหรับงานวิจัยที่เกี่ยวข้องเท่าที่ศึกษาและค้นคว้ามีดังนี้ คือ “การเรียนรู้ของแขนหุ่นยนต์โดยการโปรแกรมพันธุกรรม” โดยจุมพล พลวิชัย [8], “การปรับปรุงประสิทธิภาพของโปรแกรมหุ่นยนต์ซึ่งก่อกำเนิดโดยการโปรแกรมเชิงพันธุศาสตร์” โดย ธวัชชัย เอี่ยมมนัสกุล [10], “การลดทอนความพิชิตพยายามเชิงคำนวณของวิธีการเรียนรู้แบบการโปรแกรมเชิงพันธุศาสตร์” โดยชัชวรินทร์ เจริญภาพกรณ์ [9], “การจัดตารางสอนของโรงเรียนแบบอัตโนมัติโดยจินตคณิตอัลกอริทึม” โดยกาญจณี วงศ์วิภาพร [7] ในส่วนของงานประยุกต์จินตคณิตอัลกอริทึมกับการค้นคืนสารสนเทศ เช่น Jing-Jye Yang, Robert R. Korfhage และ Edie Rasmussen ได้ประยุกต์ใช้จินตคณิตอัลกอริทึมในการปรับปรุงควิรี (Query Improvement in Information Retrieval using Genetic Algorithms) [4] โดยมีวัตถุประสงค์ที่จะปรับปรุงควิรี ให้สามารถค้นคืนข้อมูลได้ตรงตามความต้องการของผู้ใช้มากขึ้น ข้อจำกัดของงานวิจัยนี้คือใช้การปรับปรุงเฉพาะน้ำหนัก (term weight) ของควิรี ไม่มีการเพิ่มเติมคำใหม่ ๆ เข้ามาในกระบวนการของ จินตคณิต จะเน้นการค้นหาคำที่ค้ำอยู่เฉพาะที่คำที่มีอยู่ในควิรีเดิม ในขณะที่ Maria J. Martin-Bustista, Henrik L. Larsen, Jacob Nicolaisen และ Torben Svendsen ได้ประยุกต์ใช้จินตคณิตอัลกอริทึมร่วมกับฮินส์แบบฟัซซีในงานวิจัยเรื่อง An Approach to An Adaptive Information Retrieval Agent using Genetic Algorithms with Fuzzy Set Genes [5] โดยมีวัตถุประสงค์ที่จะสร้างเอเจนต์เพื่อการกลั่นกรองส่วนบุคคล (personal filter agent) ระหว่างผู้ใช้กับกลไกค้นหาข้อมูลบนอินเทอร์เน็ต (internet search engine) ข้อจำกัดของงานวิจัยนี้คือการใช้ฮินส์แบบฟัซซีทำให้ยากต่อการกำหนดรูปแบบ โครโมโซม ได้ถูกต้องยากแก่การใช้งานจริง

จะเห็นว่าแนวโน้มการใช้จินตคณิตอัลกอริทึมในการค้นคืนสารสนเทศมีมากขึ้นและผลที่ได้ก็เป็นที่น่าพึงพอใจแต่ทั้งนี้ยังต้องมีการวิจัยพัฒนาต่อไป เพื่อเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศให้ดียิ่งขึ้น

3. หลักการของจินตคณิตอัลกอริทึม

3.1 องค์ประกอบของจินตคณิตอัลกอริทึม

จินตคณิตอัลกอริทึมมีองค์ประกอบที่สำคัญ 5 องค์ประกอบ คือ

- 1) รูปแบบโครโมโซมที่ใช้ในการนำเสนอทางเลือกที่สามารถจะเป็นได้ของแต่ละปัญหา

- 2) วิธีสร้างประชากรต้นกำเนิด (initial population) ของทางเลือกที่สามารถจะเป็นไปได้
- 3) ฟังก์ชันสำหรับประเมินค่าความเหมาะสม (fitness) เพื่อให้คะแนนแต่ละทางเลือก
- 4) จินตคณิตโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ในการปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ ได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน
- 5) ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้สำหรับ จินตคณิตอัลกอริทึม เช่น ขนาดของประชากร, ความน่าจะเป็นของการใช้จินตคณิตโอเปอเรเตอร์ และจำนวนรุ่นเป็นต้น
- 6) ฟังก์ชันสำหรับประเมินค่าความเหมาะสม (fitness) เพื่อให้คะแนนแต่ละทางเลือก
- 7) จินตคณิตโอเปอเรเตอร์ (Genetic Operator) ซึ่งใช้ในการปรับเปลี่ยนองค์ประกอบของข้อมูลตลอดกระบวนการ ได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน
- 8) ค่าพารามิเตอร์ต่างๆ ซึ่งต้องใช้สำหรับ จินตคณิตอัลกอริทึม เช่น ขนาดของประชากร, ความน่าจะเป็นของการใช้จินตคณิตโอเปอเรเตอร์ และจำนวนรุ่นเป็นต้น

3.2 การทำงานของจินตคณิตอัลกอริทึม

ขั้นตอนแรกของจินตคณิตอัลกอริทึมคือการกำหนดฟังก์ชันความเหมาะสมรวมทั้งรูปแบบโครโมโซมเสียก่อน จากนั้นจึงเริ่มสร้างประชากรต้นกำเนิดตามรูปแบบโครโมโซมที่ได้กำหนดไว้ เมื่อได้ประชากรต้นกำเนิดแล้วก็ทำการวัดค่าความเหมาะสม (fitness) ของแต่ละโครโมโซม เพื่อคัดเลือกเข้าสู่กระบวนการจินตคณิตโอเปอเรเตอร์ โดยทำการคัดเลือกเอาเฉพาะโครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจชุดหนึ่งเก็บไว้โครโมโซมที่คัดเลือกไว้นั้นจะถูกนำมาทำการครอสโอเวอร์และมิวเตชันได้เป็นโครโมโซมชุดใหม่ ซึ่งเราจะนำโครโมโซมชุดใหม่นี้มาวัดค่าความเหมาะสมเพื่อทำการคัดเลือกและดำเนินการต่อไปจนสิ้นสุดตามเงื่อนไขที่ได้กำหนดไว้ ก็จะได้โครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจ หรือได้คำตอบของปัญหาที่ต้องการดังแสดงในรูปที่ 1

4. การค้นคืนสารสนเทศออนไลน์โดยใช้จินตคณิตอัลกอริทึม

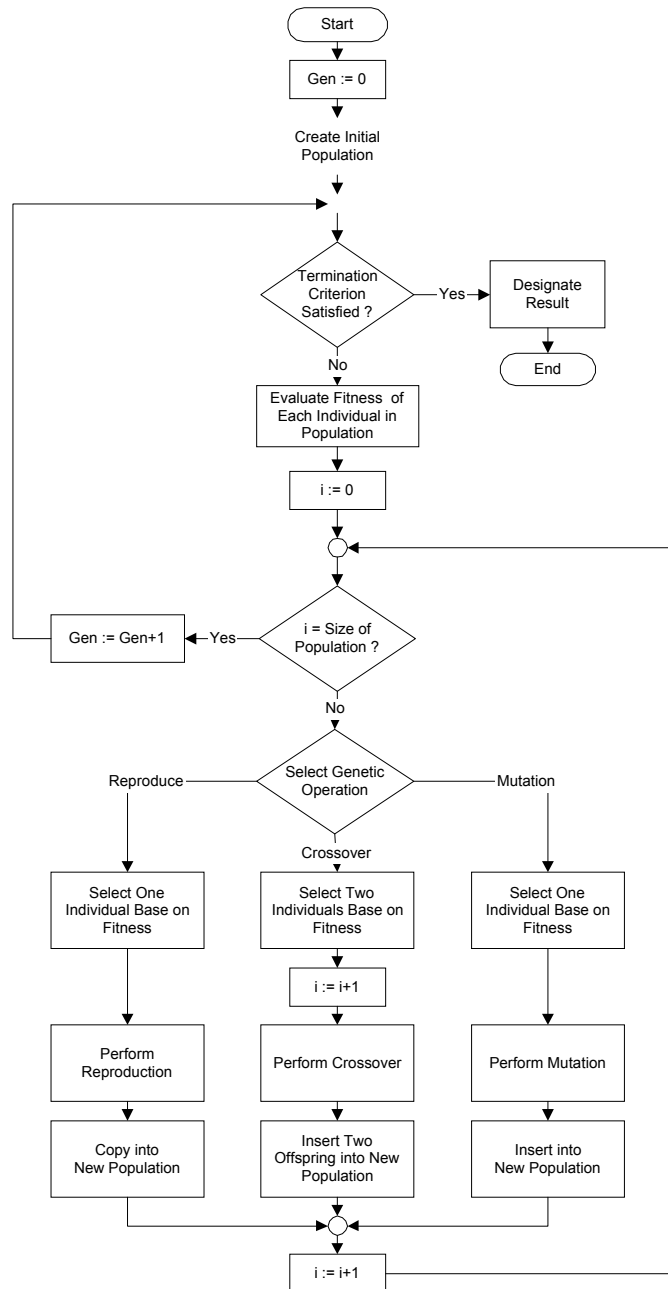
4.1 รูปแบบการนำเสนอ

ระบบค้นคืนสารสนเทศออนไลน์โดยใช้จินตคณิตอัลกอริทึมนี้ จัดทำขึ้นภายใต้เวกเตอร์สเปซโมเดล โดยแต่ละเอกสารแทนด้วยเวกเตอร์ของคำสำคัญและควิรีแทนด้วยเวกเตอร์ของคำที่ใช้ในควิรี(query terms) การใช้เวกเตอร์ควิรีในการค้นคืนทำได้โดยการจับคู่ระหว่างเอกสารกับควิรีแล้วทำการคำนวณหาค่าความคล้ายคลึง โดยถ้าหากปรากฏค่า ณ

ตำแหน่งนั้นในเอกสารหรือคิวรีจะให้ค่าเป็น “1” หากไม่ปรากฏให้ค่าเป็น “0” ค่าที่คำนวณได้เป็นการแสดงว่าเอกสารนั้น ๆ ตรงกับคิวรีเพียงใด ซึ่งค่าความคล้ายคลึงที่วัดได้นี้จะถูกนำไปใช้ในขั้นตอนการคัดเลือกของกระบวนการจenetik ตัวอย่างเช่น ให้ Doc เป็นเวกเตอร์ของเอกสาร และ Query เป็นเวกเตอร์ของคิวรี

$$Doc = (term_1, term_2, \dots, term_n)$$

$$Query = (qterm_1, qterm_2, \dots, qterm_m)$$



รูปที่ 1 การทำงานของจenetikอัลกอริทึม [12]

เมื่อผู้ใช้ได้คิวรีเข้าสู่ระบบสามารถค้นคืนชุดของเอกสาร 5 ฉบับประกอบด้วยคำสำคัญดังนี้

Doc₁ ประกอบด้วย Relational Database, Query, Data Retrieval, Computer

Networks, DBMS
Doc₂ ประกอบด้วย Artificial Intelligence, Internet, Indexing, Natural Language Processing

- Doc₃ ประกอบด้วย Databases, Expert System, Information Retrieval System, Multimedia
- Doc₄ ประกอบด้วย Fuzzy Logic, Neural Network, Computer Networks
- Doc₅ ประกอบด้วย Object-Oriented, DBMS, Query, Indexing

สามารถจัดเรียงเป็นชุดของคำสำคัญ คือ Artificial Intelligence, Computer Networks, Data Retrieval, Databases, DBMS, Expert System, Fuzzy Logic, Indexing, Information Retrieval System, Internet, Multimedia, Natural Language Processing, Neural Network, Object-Oriented, Query, Relational Databases

นำเสนอในรูปแบบโครโมโซม ได้ดังนี้คือ

- Doc₁ = 0110100000000011
- Doc₂ = 1000000101010000
- Doc₃ = 0001010010100000
- Doc₄ = 0100001000001000
- Doc₅ = 0000100100000110

โครโมโซมชุดแรกที่ได้มานี้จะเรียกว่าประชากรต้นกำเนิด ซึ่งจะนำไปผ่านกระบวนการจินตนาการไป ความยาวของโครโมโซมเหล่านี้จะขึ้นอยู่กับจำนวนคำสำคัญของชุดเอกสารทั้งหมดที่ตรงตามคิวรี จากตัวอย่างแต่ละโครโมโซมมีความยาวเท่ากับ 16 บิต

4.2 การวัดค่าความเหมาะสม

ฟิตเนสฟังก์ชันหรือฟังก์ชันวัดความเหมาะสม คือ ฟังก์ชันที่ใช้ในการประเมินว่าแต่ละทางเลือก (solution) นั้น มีความเหมาะสม สามารถใช้แก้ปัญหาได้ดีเพียงใด สำหรับปัญหาของการค้นหาสารสนเทศก็คือ ทำอย่างไรจึงจะสามารถค้นหาเอกสารที่ตรงตามความต้องการของผู้ใช้ จากการที่เลือกใช้เทคนิคการสุ่มโครโมโซม การที่จะกำหนดว่าเอกสารใดตรงตามคิวรีหรือคล้ายคลึงกับคิวรีนั้น สามารถทำได้โดยใช้ฟังก์ชันในการวัดความคล้ายคลึงซึ่งก็มีด้วยกันหลายรูปแบบ ในที่นี้เลือกใช้ฟังก์ชันดังตารางที่ 1 [6] เพื่อนำมาใช้เป็นฟิตเนสฟังก์ชัน

จากตารางที่ 1 กำหนดให้ $X=(x_1, x_2, x_3, \dots, x_n)$,
 $|X|$ = จำนวนคำที่ปรากฏใน X ,
 $|X \cap Y|$ = จำนวนคำที่ปรากฏทั้งใน X และ Y

ฟิตเนสฟังก์ชันมี 2 รูปแบบคือแบบถ่วงน้ำหนักและแบบไบนารี แต่ในการวิจัยครั้งนี้จะใช้แบบไบนารี

ผลลัพธ์ที่ได้จากการคำนวณของฟิตเนสฟังก์ชันจะมีค่าอยู่ระหว่าง 0.0-1.0 โดยที่ 1.0 หมายถึงเอกสารและคิวรีนั้นเหมือนกัน ค่าที่เข้าใกล้ 1.0 แสดงว่าเอกสารและคิวรีนั้นมีความสัมพันธ์กันมาก ส่วนค่าที่เข้าใกล้ 0.0 แสดงว่าเอกสารและคิวรีนั้นมีความสัมพันธ์กันน้อย ค่าที่ได้จากการคำนวณนี้เรียกว่าค่าความเหมาะสม (fitness)

ตารางที่ 1 ฟิตเนสฟังก์ชัน

Similarity Measure	Binary Term Vectors	Weighted Term Vectors
Dice coefficient	$2 \frac{ X \cap Y }{ X + Y }$	$\frac{2 \sum_{i=1}^l x_i \cdot y_i}{\sum_{i=1}^l x_i^2 + \sum_{i=1}^l y_i^2}$
Cosine coefficient	$\frac{ X \cap Y }{ X ^{1/2} \cdot Y ^{1/2}}$	$\frac{\sum_{i=1}^l x_i \cdot y_i}{\sqrt{\sum_{i=1}^l x_i^2 \cdot \sum_{i=1}^l y_i^2}}$
Jaccard Coefficient	$\frac{ X \cap Y }{ X + Y - X \cap Y }$	$\frac{\sum_{i=1}^l x_i \cdot y_i}{\sum_{i=1}^l x_i^2 + \sum_{i=1}^l y_i^2 - \sum_{i=1}^l x_i \cdot y_i}$

4.3 การคัดเลือก

หลังจากได้ค่าความเหมาะสมของแต่ละโครโมโซมแล้ว ขั้นตอนต่อมาคือการคัดเลือกสายพันธุ์ (Selection) การคัดเลือกสายพันธุ์เป็นไปตามหลักการอยู่รอดของสิ่งที่เหมาะสมที่สุด (Survival of the fittest) โดยโครโมโซมที่มีค่าความเหมาะสมเป็นที่น่าพอใจก็จะได้รับการคัดเลือกไว้ ส่วนโครโมโซมที่มีค่าความเหมาะสมต่ำ ก็จะไม่ได้รับการคัดเลือกน้อยหรือไม่ได้รับการคัดเลือกเลย

4.4 การครอสโอเวอร์

การครอสโอเวอร์ (Crossover) คือ การนำโครโมโซม 2 โครโมโซมมาทำการตามขั้นตอนต่างๆ ซึ่งจะให้ค่าโครโมโซมใหม่ที่จะนำไปใช้ในการคัดเลือกครั้งถัดไป หรือเป็นการนำโครโมโซมสองโครโมโซมมาผสมกันเพื่อให้ได้ค่าโครโมโซมใหม่ขึ้นมาตัวเอง ในขั้นตอนนี้จะพยายามสร้างทางเลือกใหม่และปรับปรุงทางเลือกให้ดีขึ้นโดยการครอสโอเวอร์ ซึ่งจินตนาการที่มักจะพยายามสร้างทางเลือกที่ดีขึ้นโดยการรวมลักษณะที่ดีของแต่ละโครโมโซมเข้าด้วยกัน โครโมโซมที่มีค่าความเหมาะสมสูงกว่า มักจะถูกเลือกมาครอสโอเวอร์บ่อยครั้งกว่าส่งผลให้มีโอกาสในการรอดไปยังรุ่นต่อ ๆ ไปมีมากกว่า

การครอสโอเวอร์สามารถทำได้หลายวิธีเช่น การครอสโอเวอร์หนึ่งตำแหน่ง (one point crossover) การครอสโอเวอร์สองตำแหน่ง (two point crossover) หรือการครอสโอเวอร์หลายตำแหน่ง (multiple point crossover)

การครอสโอเวอร์แบบหนึ่งตำแหน่ง โดย ครอสโอเวอร์ ณ ตำแหน่งที่ 8

101111110011101

100110011110000

ผลที่ได้หลังการครอสโอเวอร์ คือ

101111111110000

100110010011101

4.5 การมิวเตชัน

การมิวเตชัน (Mutation) เป็นลักษณะของการผ่าเหล่าคือการนำโครโมโซมเก่ามาสุ่มแก้ไขบางส่วนของโครโมโซม เช่น บิทบางบิทให้เปลี่ยนไป ทำให้ได้โครโมโซมใหม่ที่มีสายพันธุต่างจากเดิม ซึ่งมีโอกาสที่จะเป็นโครโมโซมที่ดีขึ้นหรือเลวลงก็ได้ หากโครโมโซมที่ได้ใหม่นี้เป็นโครโมโซมที่เลวลง โครโมโซมที่ได้นี้ก็จะถูกคัดออกไปในขั้นตอนการคัดเลือกเอง วัตถุประสงค์ของการมิวเตชันคือเพื่อป้องกันการสูญหายของข้อมูลและเพื่อความหลากหลายของข้อมูล ตัวอย่างของการทำมิวเตชัน เช่น สุ่มเลือกเปลี่ยนโครโมโซมตำแหน่งที่ 10

101111110011101

ผลที่ได้คือ 101111110111101

จีเนติกอัลกอริทึม จะทำเป็นวัฏจักรหมุนเวียนอยู่เช่นนี้จนกระทั่งถึงจุดหนึ่งตามเงื่อนไข โดยอาจสิ้นสุดเมื่อถึงรุ่น (generation) ตามที่กำหนดหรือสิ้นสุดเมื่อพบค่าตอบที่ดีที่สุดแล้วหรือถึงเรซสโลต์ (threshold) ตามที่ได้กำหนดไว้ล่วงหน้าแล้วนั้น

4.6 ขั้นตอนการทำงานของระบบ

1. ผู้ใช้ใส่คิวรีเข้าสู่ระบบ
2. นำคิวรีไปค้นหาจากรายการคำสั่งสำคัญ ทั้งหมดของระบบที่มีอยู่
3. นำชุดของเอกสารที่ตรงตามคิวรีมาแปลงเป็นโครโมโซม จะได้เป็นประชากรต้นกำเนิด

4. นำชุดโครโมโซม (ประชากรต้นกำเนิด) ที่ได้นี้เข้าสู่กระบวนการจีเนติกโอเปอเรเตอร์ อันได้แก่ การคัดเลือก การครอสโอเวอร์ และการมิวเตชัน ตามที่ได้อธิบายไว้แล้ว
ดำเนินการตามข้อ 4 จนกระทั่งถึงรุ่นที่กำหนด จะได้โครโมโซมคิวรีที่เหมาะสม (optimize query chromosome) เพื่อค้นคืนสารสนเทศต่อไป
5. แปลงโครโมโซมคิวรีที่เหมาะสมนี้เป็นคิวรีเพื่อค้นคืนสารสนเทศจากฐานข้อมูล

5. การทดลอง

5.1 วิธีการทดลอง

การทดลองเป็นการทดลองค้นคืนจากคิวรี 21 คิวรี ซึ่งใช้ฟิตเนสฟังก์ชันหรือฟังก์ชันวัดความเหมาะสมที่แตกต่างกัน 3 ฟังก์ชันดังที่ได้กล่าวมาแล้ว คือ Jaccard Coefficient, Cosine Coefficient และ Dice Coefficient โดยที่แต่ละฟิตเนสฟังก์ชันนั้นจะทดสอบโดยใช้ค่าความน่าจะเป็นในการครอสโอเวอร์ 0.8 และ ค่าความน่าจะเป็นในการมิวเตชัน 0.01, 0.10 และ 0.30 จำนวนรุ่นสูงสุด (Max Generation) เท่ากับ 30 รุ่น เพื่อทดสอบเปรียบเทียบประสิทธิภาพ โดยประสิทธิภาพของการค้นคืนสารสนเทศวัดได้จากค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

ค่าความระลึก (R) เป็นอัตราส่วนของการค้นพบเอกสารที่ถูกต้องจากจำนวนเอกสารที่ถูกต้องทั้งหมด ดังสมการที่ 1 [3][6]

$$R = \frac{\text{จำนวนเอกสารที่ถูกต้องที่ค้นคืนได้}}{\text{จำนวนเอกสารที่ถูกต้องทั้งหมดในฐานข้อมูล}} \quad (1)$$

ค่าความแม่นยำ (P) เป็นอัตราส่วนของการค้นพบเอกสารที่ถูกต้องจากจำนวนเอกสารทั้งหมดที่ทำการค้นคืนมาได้ ดังสมการที่ 2 [3][6]

$$P = \frac{\text{จำนวนเอกสารที่ถูกต้องที่ค้นคืนได้}}{\text{จำนวนเอกสารทั้งหมดที่ค้นคืนออกมาได้}} \quad (2)$$

ฐานข้อมูลสำหรับงานวิจัยนี้ เป็นฐานข้อมูลโครงการนักศึกษา คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จำนวน 343 โครงการงาน

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพ

ผลงานวิจัย	ค่าความแม่นยำ	ค่าความระลึกลับ
วิธีการของ Kraft	0.842	0.664
วิธีการที่นำเสนอ	0.746	0.971

5. ผลการค้นคืนสารสนเทศโดยวิธีการที่นำเสนอเมื่อเปรียบเทียบกับวิธีการโดยทั่วไปพบว่า วิธีการที่นำเสนอให้ค่าความระลึกลับสูงกว่าวิธีการทั่วไป เนื่องจากวิธีการที่นำเสนอนอกจากจะสามารถค้นคืนเอกสารที่ค้นคืนได้ตามวิธีเปรียบเทียบแบบตรงกัน (Exact Match) แล้วระบบยังสามารถค้นคืนสารสนเทศที่เกี่ยวข้องแต่ไม่ได้ระบุโดยตรงจากคีย์อีกด้วย และเมื่อเปรียบเทียบกับผลงานวิจัยที่ใกล้เคียงคือผลงานของ Kraft [4] ซึ่งเป็นการปรับปรุงคีย์เพื่อใช้ในการค้นคืนสารสนเทศเช่นเดียวกันนั้น ผลปรากฏดังตารางที่ 3 คือวิธีการที่นำเสนอมีค่าความแม่นยำต่ำกว่า แต่ให้ค่าความระลึกลับสูงกว่าวิธีการของ Kraft และเมื่อเปรียบเทียบประสิทธิภาพ (E-Measure) ตามวิธีการของ van Rijsbergen [11] โดยรวมค่าความระลึกลับและค่าความแม่นยำออกมาเป็นค่าเดียว ดังสมการที่ 3

$$E = 1 - \frac{(1+b^2)PR}{b^2P+R} \quad (3)$$

โดยที่ P คือค่าความแม่นยำ R คือค่าความระลึกลับและ b เป็นการวัดความสำคัญเชิงสัมพัทธ์ของค่าความระลึกลับและความแม่นยำของผู้ใช้ ให้ $b=1$ คือความสำคัญระหว่างค่าความระลึกลับและค่าความแม่นยำเท่ากัน ผลปรากฏว่าวิธีการที่นำเสนอมีค่า E เท่ากับ 0.156 และวิธีของ Kraft มีค่า E เท่ากับ 0.258 ซึ่งค่า E ที่ยิ่งเข้าใกล้ 0 แสดงว่ายังมีประสิทธิภาพดี ดังนั้นสรุปได้ว่าวิธีการที่นำเสนอมีประสิทธิภาพดีกว่า

6. สรุป

จากผลการทดลองในเบื้องต้นจะเห็นว่าค่าความแม่นยำและค่าความระลึกลับจะมีลักษณะเชิงผกผันกัน การจะเลือกใช้พารามิเตอร์ใดขึ้นกับความเหมาะสมว่าต้องการใช้ค้นคืนสารสนเทศเพื่ออะไร กรณีที่ต้องการเอกสารที่มีค่าความแม่นยำสูงก็ควรจะใช้ค่าความน่าจะเป็นในการครอสโอเวอร์สูงและค่าความน่าจะเป็นในการมิวเตชันต่ำ ในขณะที่หากต้องการเอกสารที่เกี่ยวข้องมาก (ค่าความระลึกลับสูง) ก็อาจใช้ค่าความน่าจะเป็นในการมิวเตชันสูงและค่าความน่าจะเป็นในการครอสโอเวอร์ต่ำลง จากผลการทดลองในเบื้องต้นนี้จะเห็นว่าเราสามารถใช้อัลกอริทึมกับการค้นคืนสารสนเทศได้ นอกจากนี้ วิธีการที่นำเสนอเป็นการใช้ค่าสำคัญเป็นหลักในการค้นหา ซึ่งค่าสำคัญเหล่านั้นจะ

เป็นตัวชี้ไปยังเอกสารอีกทีหนึ่ง ดังนั้นวิธีการที่นำเสนอจึงสามารถนำไปประยุกต์ใช้สำหรับค้นคืนสารสนเทศภาษาใด ๆ ก็ได้

สำหรับแนวทางในการวิจัยต่อไปคือทำการทดลองกับฐานข้อมูลที่มีขนาดใหญ่ขึ้น พร้อมทั้งนำเสนอเอกสารจากการค้นคืนตามลำดับของค่าความเหมาะสมที่วัดได้จากฟิตเนสฟังก์ชันซึ่งจะบ่งบอกถึงลำดับความต้องการของผู้ใช้

เอกสารอ้างอิง

- [1] David, L. **Handbook of genetic algorithms**. New York : Van Nostrand Reinhold. 1991.
- [2] Goldberg, D.E. **Genetic Algorithms: in Search, Optimization, and Machine Learning**. New York : Addison-Wesley Publishing Co. Inc. 1989.
- [3] Korfhage, R.R. **Information Storage and Retrieval**. New York : Wiley Computer Publishing. 1997.
- [4] Kraft, D.H. et. al. "The Use of Genetic Programming to Build Queries for Information Retrieval." in **Proceedings of the First IEEE Conference on Evolutionary Computation**. New York : IEEE Press. PP. 468-473.
- [5] Martin-Bautista, M.J. et. al. "An Approach to An Adaptive Information Retrieval Agent using Genetic Algorithms with Fuzzy Set Genes." In **Proceeding of the Sixth International Conference on Fuzzy Systems**. New York : IEEE Press. 1997. PP. 1227-1232.
- [6] Salton, G. **Automatic text processing : the transformation, analysis, and retrieval of information by computer**. New York : Addison-Wesley Publishing Co. Inc. 1989.
- [7] กาญจน์ วงศ์วิภาพร. "การจัดตารางสอนของโรงเรียนแบบอัตโนมัติโดยจินตคณิต อัลกอริทึม." วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.
- [8] จุมพล พลวิชัย. "การเรียนรู้ของแขนหุ่นยนต์โดยใช้การโปรแกรมพันธุกรรม." วิทยานิพนธ์ วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.
- [9] ชัยวัฒน์ เจริญภาพกรณ์. "การลดทอนความเพียรพยายามเชิงคำนวณของวิธีการเรียนรู้แบบการโปรแกรมเชิงพันธุศาสตร์"

วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2540.

- [10] ชวัชชัย เอี่ยมมนัสสกุล. “การปรับปรุงประสิทธิภาพของโปรแกรมหุ่นยนต์ซึ่งก่อกำเนิดโดยการโปรแกรมเชิงพันธุศาสตร์.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2540.
- [11] นิพนธ์ เจริญกิจการ. “การจัดเก็บและค้นคืนสารสนเทศ ฉบับปรับปรุงครั้งที่ 1.” [Online]. Available : <http://web.it.kmutt.ac.th/nipon/yllabus-temp.html>. 2542.
- [12] สุรเดช ตรีวิทยากรานต์. “การจัดตารางเวลาสอบหัวข้อวิจัยโดยอาศัยเทคนิคจินตึก อัลกอริทึม.” วิทยานิพนธ์วิทยาศาสตร์ มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, มหาวิทยาลัยพระจอมเกล้าธนบุรี. 2540.



Bangorn Klabbankoh received bachelor degree in education technology from King Mongkut’s University of Technology Thonburi (KMUTT) in 1997, bachelor degree in Business Administration (Marketing) from Ramkhamhaeng University in 1999 and master degree in information technology from King Mongkut’s Institute of Technology

Ladkrabang (KITL) in 2000. Her research interests are Genetic Algorithm, Information Retrieval and Expert System.



ผู้ช่วยศาสตราจารย์ เอื้อน ปิ่นเงิน

สำเร็จการศึกษา ระดับปริญญาตรีจากมหาวิทยาลัยศรีนครินทรวิโรฒสาขาคณิตศาสตร์ ปริญญาโทจากจุฬาลงกรณ์มหาวิทยาลัย และ Oregon State University สาขาคอมพิวเตอร์ และ

ได้สำเร็จปริญญาเอกทางคอมพิวเตอร์จาก University of Nebraska ประเทศสหรัฐอเมริกา ปัจจุบันเป็นอาจารย์ประจำภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง