

การเปรียบเทียบวิธีการสืบค้นข้อความภาษาไทยระหว่างเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมและเทคนิคการค้นหาดำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง

The Comparison of Data Query in Thai Text between Text Processing and Structured Query Language (SQL)

อิทธิศักดิ์ ศรีดำ¹

Idhisak Sridam¹

Received: 9 October 2019 ; Revised: 14 February 2020 ; Accepted: 13 May 2020

บทคัดย่อ

การสืบค้นข้อมูลที่เป็นข้อความภาษาไทยด้วยการเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้างพบปัญหาการนำส่งข้อมูลยังไม่เหมาะสมและไม่ตรงกับความต้องการ บทความนี้มีจุดมุ่งหมายในการเสนอการทดสอบการสืบค้นข้อมูลข้อความภาษาไทยระหว่างการประมวลผลข้อความตามมาตรฐานวัดความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมและการเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง

การศึกษานี้เป็นการผสมผสานวิธีการระหว่างการวิจัยเชิงคุณภาพและการวิจัยเชิงปริมาณด้วยการวิจัยเอกสาร และการวิจัยเชิงทดลอง โดยการพัฒนาโปรแกรมทดสอบการสืบค้นข้อมูลข้อความภาษาไทยด้วยการประมวลผลข้อความตามมาตรฐานวัดความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม และพัฒนาโปรแกรมทดสอบการเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง เพื่อเปรียบเทียบประสิทธิภาพ ของอัตราการค้นพบข้อความที่สืบค้น และเวลาที่ใช้ประมวลผลสืบค้นข้อมูล

ผลการศึกษา จากการทดสอบการสืบค้นข้อความภาษาไทย จำนวน 10 ครั้ง และหาค่าเฉลี่ยประสิทธิภาพตามเกณฑ์การประเมินประสิทธิภาพ 3 ด้าน ได้แก่ ค่าความระลึก ค่าความเที่ยงตรง และค่าเอฟเมเชอร์ พบว่า เทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม มีค่าเฉลี่ยความระลึก เท่ากับ 0.981 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ 0.986 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.983 ส่วนเทคนิคการค้นหาดำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง มีค่าเฉลี่ยความระลึก เท่ากับ 0.155 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ 0.295 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.203

คำสำคัญ: การประมวลผลข้อความ การหาความคล้ายคลึงเชิงความหมาย ขั้นตอนวิธีเชิงพันธุกรรม ภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง

Abstract

Most query systems use Structured Query Language (SQL) to Thai query data but the data is unmatched and dissimilar to the data required by the users. This article aims to propose the test of Thai text search by comparing the method of text processing to the semantic-based similarity measurements using a Genetic algorithm with SQL.

This research used both qualitative and quantitative research methodology which consisted of documentary and experimental research. The testing used Thai text in searching for the semantic-based similarity measurement.

The results revealed the efficiency average from searching Thai text 10 times according to the 3 efficiency evolution criteria, i.e., recall, precision, and F-Measure. The semantic-based similarity measurement technique using Genetic algorithm with the average of recall, precision, and F-Measure were 98.1%, 98.6%, and 98.3%, respectively.

¹ สาขาวิชาวิศวกรรมซอฟต์แวร์และระบบสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี สถาบันเทคโนโลยีปทุมวัน กรุงเทพมหานคร 10330
E-mail: idhisak@pit.ac.th

¹ Department of Software Engineering and Information System, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok, Thailand 10330 E-mail: idhisak@pit.ac.th

Moreover, the search technique using SQL had average of recall, precision, and F-Measure of 15.5%, 29.5%, and 20.3%, respectively.

Keywords: Text Processing, Similarity Measurement, Genetic Algorithm, Structured Query Language

บทนำ

The screenshot shows the Google Advanced Search page in Thai. The main heading is "การค้นหาขั้นสูง" (Advanced Search). Below it, there are several sections for refining search results:

- ค้นหาเว็บไซต์ที่มี...** (Search for websites that...):
 - ทุกคำเหล่านี้: (All these words) - Tip: พิมพ์คำสำคัญ: สามสี แรทเทอร์เจียร์
 - คำหรือวลีที่ตรงตามนี้: (Words or phrases that match) - Tip: ใส่คำที่ต้องการให้ตรงตามนี้ในเครื่องหมายอัญประกาศ: "เทอริเจอร์ท์อินท์นู"
 - คำใดๆ เหล่านี้: (Any of these words) - Tip: พิมพ์ OR ระหว่างทุกคำที่คุณต้องการ: ขนาดเล็ก OR ขนาดมาตรฐาน
 - ไม่มีคำเหล่านี้: (None of these words) - Tip: ใส่เครื่องหมายลบหน้าคำที่คุณไม่ต้องการ: -สัตว์พื้นทะเล, "-แจ็ค รัสเซลล์"
 - จำนวนตั้งแต่: (Number of pages from) - Tip: ใส่จุดสองจุดระหว่างตัวเลขและเพิ่มหน่วยของการวัด: 10..35 ปอนด์, \$300..\$500, 2010..2011
- แล้วจำกัดผลลัพธ์ของคุณโดย...** (Further limit your results by...):
 - ภาษา: (Language) - Tip: หาเว็บไซต์ในภาษาที่คุณเลือก
 - ภูมิภาค: (Region) - Tip: หาเว็บไซต์ที่เผยแพร่ในภูมิภาคใดภูมิภาคหนึ่ง
 - อัปเดตล่าสุด: (Most recent) - Tip: หาเว็บไซต์ที่มีการอัปเดตภายในเวลาที่คุณระบุ
 - ไซต์หรือโดเมน: (Site or domain) - Tip: ค้นหาไซต์หนึ่ง (เช่น wikipedia.org) หรือจำกัดผลการค้นหาของคุณเป็นโดเมน เช่น .edu, .org หรือ .gov
 - คำที่ปรากฏ: (Words that appear) - Tip: ค้นหาคำในหน้าเว็บทั้งหน้า ชื่อหน้าเว็บ หรือที่อยู่เว็บ หรือลิงก์ไปยังหน้าเว็บที่คุณต้องการ
 - ค้นหาปลอดภัย: (Safe search) - Tip: นอก ค้นหาปลอดภัย ว่าจะต้องการกรองเนื้อหาทางเพศที่โจ่งแจ้งหรือไม่
 - ประเภทไฟล์: (File type) - Tip: หาเว็บไซต์ในรูปแบบที่คุณต้องการ

Figure 1 Example of Google Advanced Search (Source: https://www.google.com/advanced_search, online)

การใช้คำหลัก (Keyword) ในการสืบค้นข้อมูลบนเว็บไซต์ที่ให้บริการข้อมูลเป็นที่นิยมของผู้ใช้งานเป็นอย่างมาก ทั้งข้อมูลที่เป็นข้อความ รูปภาพ ภาพเคลื่อนไหว เพลง ซอฟต์แวร์ แผนที่ ข้อมูลบุคคล กลุ่มข่าว และอื่นๆ ซึ่งแตกต่างกันไปแล้วแต่ผู้ให้บริการ เว็บไซต์ที่บริการข้อมูลหลักๆ จะเป็นเว็บไซต์ประเภทโปรแกรมสืบค้นข้อมูล (Search engine)¹

ผู้ให้บริการโปรแกรมสืบค้นข้อมูลเหล่านี้ได้พัฒนาระบบสืบค้นข้อมูลภายใต้เงื่อนไขต่างๆ ที่กำหนดขึ้น เพื่ออำนวยความสะดวก

ความสะดวกให้ผู้ใช้งานเข้าถึงข้อมูลได้โดยง่ายที่สุด ประเภทโปรแกรมสืบค้นข้อมูล แบ่งออกเป็น 3 ประเภท ได้แก่ การสืบค้นข้อมูลด้วยตัวสืบค้นอัตโนมัติมาเก็บไว้ที่ฐานข้อมูล (Database) การสืบค้นข้อมูลด้วยการใช้คนนำเข้าข้อมูลสู่ฐานข้อมูล (Directories search) และการสืบค้นข้อมูลที่อาศัยฐานข้อมูลของโปรแกรมสืบค้นข้อมูลอื่น (Meta search)²

เนื่องจากระบบค้นหาบนเว็บไซต์มีกระบวนการเตรียมการข้อมูลมาเก็บไว้ในฐานข้อมูลเอาไว้ก่อนล่วงหน้า

สำหรับไว้ให้ผู้ใช้งานทำการสืบค้นข้อมูล จึงอนุมานได้ว่าผู้ใช้บริการโปรแกรมค้นหาจะต้องใช้พื้นที่ในการจัดเก็บข้อมูลที่มีขนาดใหญ่ การศึกษาพบว่าโปรแกรมค้นหาได้พัฒนาระบบการค้นหาขั้นสูงไว้สำหรับบริการผู้ใช้งาน เพื่อให้ผู้ใช้งานจะได้เจาะจงการสืบค้นข้อมูลที่ต้องการได้ชัดเจนขึ้น เช่น การค้นหาข้อมูลตามคำหรือวลีที่กำหนด การค้นหาข้อมูลยกเว้นตามคำหรือวลีที่กำหนด เป็นต้น³ แสดงตัวอย่างดัง Figure 1 อย่างไรก็ตามผู้ใช้งานส่วนใหญ่ไม่ได้ใช้ระบบการค้นหาขั้นสูงที่ผู้บริการเตรียมไว้ให้² อาจทำให้เกิดปัญหาที่ผู้สืบค้นข้อมูลจะไม่สามารถค้นพบข้อมูลที่ตรงหรือใกล้เคียงกับข้อมูลที่ต้องการได้ ซึ่งส่วนใหญ่โปรแกรมสืบค้นข้อมูลจะใช้การเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง (Structured query language: SQL)⁴

การบริหารจัดการข้อมูลขนาดใหญ่อย่างมีประสิทธิภาพโดยเน้นการสืบค้นข้อมูลให้ตรงตามความต้องการของผู้ใช้งาน และการประหยัดเวลาในการประมวลผลเป็นสิ่งสำคัญ⁴ โดยเฉพาะอย่างยิ่งการสืบค้นข้อมูลที่เป็นข้อความภาษาไทย ซึ่งการสืบค้นข้อมูลที่เป็นข้อความภาษาไทยด้วยการเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง พบว่า การนำส่งข้อมูลยังไม่เหมาะสม ยังไม่ได้ข้อมูลที่ตรง และยังไม่ใกล้เคียงกับข้อมูลที่ผู้ใช้งานต้องการ⁵

นอกจากนี้ การสืบค้นข้อมูลโดยใช้คำหลักโดยผู้ใช้งานยังถูกใช้ในระบบการบริหารลูกค้าสัมพันธ์ (Customer relationship management: CRM) สำหรับกรณีต่างๆ ปัจจุบันองค์ประกอบที่สำคัญของระบบ การบริหารลูกค้าสัมพันธ์ ได้แก่ การสืบค้นข้อมูลคำถามหรือคำตอบที่ลูกค้าถามบ่อย (Question and answer: Q&A) และระบบการสืบค้นข้อมูลที่เกี่ยวข้องกับสินค้าหรือบริการ⁶ เช่น ประเภทสินค้าหรือบริการ ราคาสินค้าหรือบริการ รายละเอียดสินค้าหรือบริการ การเปรียบเทียบสินค้าหรือบริการ เป็นต้น การแสดงข้อมูลผลิตภัณฑ์ ที่เหมาะสมและครบถ้วนสามารถเพิ่มแนวโน้มในการตัดสินใจซื้อของลูกค้าได้เป็นอย่างดี⁷

ดังนั้น การนำส่งข้อมูลที่เหมาะสมที่สุดจึงเป็นสิ่งจำเป็นที่สุดของระบบการสืบค้นข้อมูล การหาค่า ความเหมาะสมของข้อมูลที่ต้องการค้นหาด้วยขั้นตอนวิธีเชิงพันธุกรรม (Genetic algorithm: GA) น่าจะสามารถนำมาใช้ในการสืบค้นข้อความที่เหมาะสมได้เมื่อสมการหาค่าความเหมาะสมของขั้นตอนวิธีเชิงพันธุกรรมได้ถูกกำหนดไว้อย่างถูกต้อง และการนำเทคนิคการหาค่า ความคล้ายคลึงเชิงความหมาย (Semantic similarity) มาประยุกต์ใช้ในการกำหนดสมการหาค่าความเหมาะสมของขั้นตอนวิธีเชิงพันธุกรรมมีประโยชน์ในการจำแนกหมวดหมู่ หรือจัดกลุ่มประเภทข้อความ บทความนี้ได้นำเสนอแนวทางการสืบค้น

ข้อมูลที่เป็นข้อความภาษาไทยอย่างมีประสิทธิภาพในการนำส่งข้อมูลตามความต้องการของผู้ใช้งานด้วยเทคนิคการหา ความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม และการเปรียบเทียบวิธีการสืบค้นข้อความภาษาไทยระหว่างเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมและเทคนิค การค้นหาด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง

ขั้นตอนการวิจัย

1. ชุดข้อมูลข้อความมาจากการคัดลอกคำไทยที่พบบ่อยจำนวน 99 คำ จากฐานข้อมูลคลังเอกสารสาธารณะ (<http://www.openbase.in.th/>) ซึ่งคาดว่าเป็นจำนวนที่เหมาะสมสำหรับใช้ในการทดสอบ⁸ มาผสมเป็นข้อความจำนวน 99 ข้อความต่อคำไทย 1 คำ รวมเป็นข้อความทั้งสิ้น 9,801 ข้อความโดยสุ่มผสมคำอย่างง่ายสำหรับใช้ในการศึกษาครั้งนี้

2. กระบวนการศึกษา

การศึกษานี้เป็นการวิจัยแบบผสมวิธีมีลักษณะการผสมกันระหว่างการวิจัยเชิงคุณภาพและการวิจัยเชิงปริมาณ (Mixed method research)⁹ การวิจัยเชิงคุณภาพจะใช้การวิจัยเอกสาร (Documentary research) ด้วยการคัดเลือกเอกสารทุติยภูมิ¹⁰ ได้แก่ เอกสารที่เกี่ยวข้องและสอดคล้องกับวัตถุประสงค์ของการวิจัย ประกอบด้วย บทความวิจัย บทความวิชาการ เอกสารตำรา และงานวิจัยที่เกี่ยวข้อง¹⁰ ส่วนการวิจัยเชิงปริมาณได้ใช้การวิจัยเชิงทดลอง (Experimental research) โดยการพัฒนาโปรแกรมทดสอบการสืบค้นข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม และพัฒนาโปรแกรมทดสอบการเข้าถึงข้อมูลด้วยภาษา แบบสอบถาม ข้อมูลอย่างมีโครงสร้าง

3. การวิเคราะห์ข้อมูล

3.1 การวิจัยเชิงคุณภาพได้ใช้การวิเคราะห์ข้อมูลมีเกณฑ์คัดเลือกเอกสารที่เกี่ยวข้อง ได้แก่ 1) ความถูกต้องของข้อมูล เอกสารต้องมาจากแหล่ง ข้อมูลที่เชื่อถือได้ มีความถูกต้อง สมบูรณ์ สอดคล้องกับบริบทที่อยู่ในช่วงระยะเวลาที่ตีพิมพ์ 2) ข้อมูลต้องมี ความน่าเชื่อถือ เป็นเอกสารที่ไม่ใช่ข้อผิดพลาดและ การบิดเบือนข้อมูล 3) ข้อมูลมีความเป็นตัวแทน เป็นเอกสารที่สามารถแสดงรายละเอียดแทนเอกสารในกลุ่มเดียวกันได้¹¹ และ 4) ข้อมูลมีความหมายชัดเจน เป็นเอกสารที่มีความหมายที่เข้าใจได้ง่าย ชัดเจน และสอดคล้องกับวัตถุประสงค์ของการศึกษา¹²

3.2 การวิจัยเชิงปริมาณเป็นการประเมินผลเปรียบเทียบวิธีการสืบค้นข้อความภาษาไทยระหว่างเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิง

พันธุกรรมและเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง ซึ่งทดสอบ การสืบค้นข้อความภาษาไทยจำนวน 10 ครั้ง (เพื่อง่ายในการคำนวณเปรียบเทียบผล) มาหาค่าเฉลี่ยประสิทธิภาพตามเกณฑ์การประเมินประสิทธิภาพ 3 ด้าน ได้แก่ ค่าความระลึก (Recall) ค่าความเที่ยงตรง (Precision) และค่าเอฟเมเชอร์ (F-Measure) โดยมีสมการดังนี้

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{F-Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

โดยที่

TP คือ จำนวนข้อความภาษาไทยที่ถูกดึงออกมาถูกต้องออกมา

FN คือ จำนวนข้อความภาษาไทยที่ผิดพลาดแต่ไม่ถูกดึงออกมา

FP คือ จำนวนข้อความภาษาไทยที่ผิดพลาดที่ถูกดึงออกมา

4. ขอบเขตของการศึกษา

4.1 ขอบเขตเนื้อหาของข้อมูลที่ใช้ในการศึกษา ได้แก่ เอกสารที่เกี่ยวข้องกับการประมวลผลข้อความ เอกสารที่เกี่ยวข้องกับการหาค่าความคล้ายคลึงเชิงความหมาย และเอกสารที่เกี่ยวข้องกับขั้นตอนวิธีเชิงพันธุกรรม

4.2 ขอบเขตด้านเครื่องมือวิจัยเป็นเครื่องมือที่ใช้ในการออกแบบโปรแกรมทดสอบการสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาค่าความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม และออกแบบโปรแกรมทดสอบการเข้าถึงข้อมูลด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง คือ โปรแกรม MINITAB19 trial version โปรแกรมฐานข้อมูลที่ใช้คือ MySQL และเครื่องมือที่ใช้วิเคราะห์เชิงสถิติคือ โปรแกรม SPSS version 20

ผลการศึกษา

การนำส่งข้อมูลที่เหมาะสมเพื่อการสืบค้นข้อมูลจากฐานข้อมูล สามารถประยุกต์ใช้ทฤษฎีการหาค่าความเหมาะสมของคำตอบ (Solution optimization)¹³ การหาค่าความเหมาะสมของคำตอบที่ต้องการค้นหา มีหลากหลายวิธีการตามทฤษฎีปัญญาประดิษฐ์ (Artificial intelligence: AI) ได้แก่ การคำนวณเชิงวิวัฒนาการเป็นการใช้หลักการหาค่าที่เหมาะสมที่สุด

โดยเริ่มต้นจากกลุ่มของคำตอบกลุ่มหนึ่งที่ถูกสุ่มมาจากนั้นก็ทำการวิวัฒนาการผสมผสานเลือกเอากลุ่มตัวอย่างที่ดีที่สุดเพื่อความอยู่รอดในรุ่นถัดไป (การปรับการค้นหา) โดยใช้วิธีความฉลาดแบบกลุ่ม (Swarm intelligence) เป็นขั้นตอนวิธีเชิงวิวัฒนาการ ได้แก่ ขั้นตอนวิธีเชิงพันธุกรรม (Genetic algorithm: GA)¹⁴ ขั้นตอนวิธีเชิงพันธุกรรม เป็นวิธีสำหรับค้นหาคำตอบ (Solution) โดยการประมาณของปัญหาที่อาศัยวิธีการคัดเลือกตามธรรมชาติในการวิวัฒนาการโดยมีประชากรแทนกลุ่มของคำตอบที่อยู่ในลักษณะสายอักขระโครโมโซม ซึ่งมีรายละเอียดดังนี้¹⁵

1. การประเมินคุณภาพของโครโมโซม
2. เลือกโครโมโซมที่มีคุณภาพดี
3. ปฏิบัติการทางพันธุกรรมโดยการข้ามสายพันธุ์และการกลายพันธุ์
4. กำหนดโครโมโซมใหม่มาเป็นรุ่นถัดไป
5. ดำเนินการสร้างโครโมโซมใหม่ไปเรื่อยๆ จนพบคำตอบที่มีค่าความเหมาะสมที่สุด (นำกลับไปย้อนทำข้อ 1 ใหม่)

นอกจากนี้ การนำเทคนิคการหาค่าความคล้ายคลึงเชิงความหมาย (Semantic similarity) มาประยุกต์ใช้ในการค้นหาข้อความที่มีความหมายเดียวกันหรือคล้ายกันได้เป็นอย่างดีมีประสิทธิภาพในด้านการแก้ปัญหาด้านการนำส่งข้อมูลข้อความภาษาไทยไม่ได้ในกรณีที่เป็นข้อความที่มีวรรณยุกต์ต่างๆ อยู่ในข้อความ ซึ่งเป็นปัญหาของการนำส่งข้อมูลมาแสดงให้กับผู้ใช้งาน โดยการหาค่าความคล้ายคลึงเชิงความหมายจัดอยู่ในประเภทของการหาค่าความคล้ายคลึงกันของข้อความด้วยการวัดระยะห่าง (Text similarity measurement) ซึ่งเป็นงานย่อย (Subtask) ของการประมวลผลที่เกี่ยวข้องกับข้อมูลที่มีข้อความมาเกี่ยวข้องเป็นอย่างมาก เช่น การปรับปรุงประสิทธิภาพของงานค้นหา (Search task) การค้นคืนข่าวสาร (Information retrieval task)¹⁶ การค้นหาข้อความที่มีความหมายเดียวกันหรือคล้ายคลึงกัน โดยใช้คำหรือข้อความที่แตกต่างกัน ซึ่งมีประโยชน์ในการจำแนกหมวดหมู่ หรือจัดกลุ่มประเภทข้อความจากแหล่งต่างๆ เช่น การจัดกลุ่มของข่าวสารออกเป็นข่าวเศรษฐกิจ ข่าวการเมือง การจำแนกความคิดเห็นของลูกค้าจากสื่อสังคมออนไลน์ หรือการจำแนกข้อความที่แสดงความเห็นเชิงบวกออกจากข้อความที่แสดงความเห็นเชิงลบ หรือใช้ในงานสรุปย่อข้อความจากหลายเอกสาร เป็นต้น¹⁶

ผลการศึกษา พบว่า กระบวนการสืบค้นข้อมูลด้วยการประมวลผลข้อความภาษาไทยด้วยเทคนิคการหาค่าความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม ประกอบด้วย 13 กระบวนการหลัก แสดงดัง Figure 2

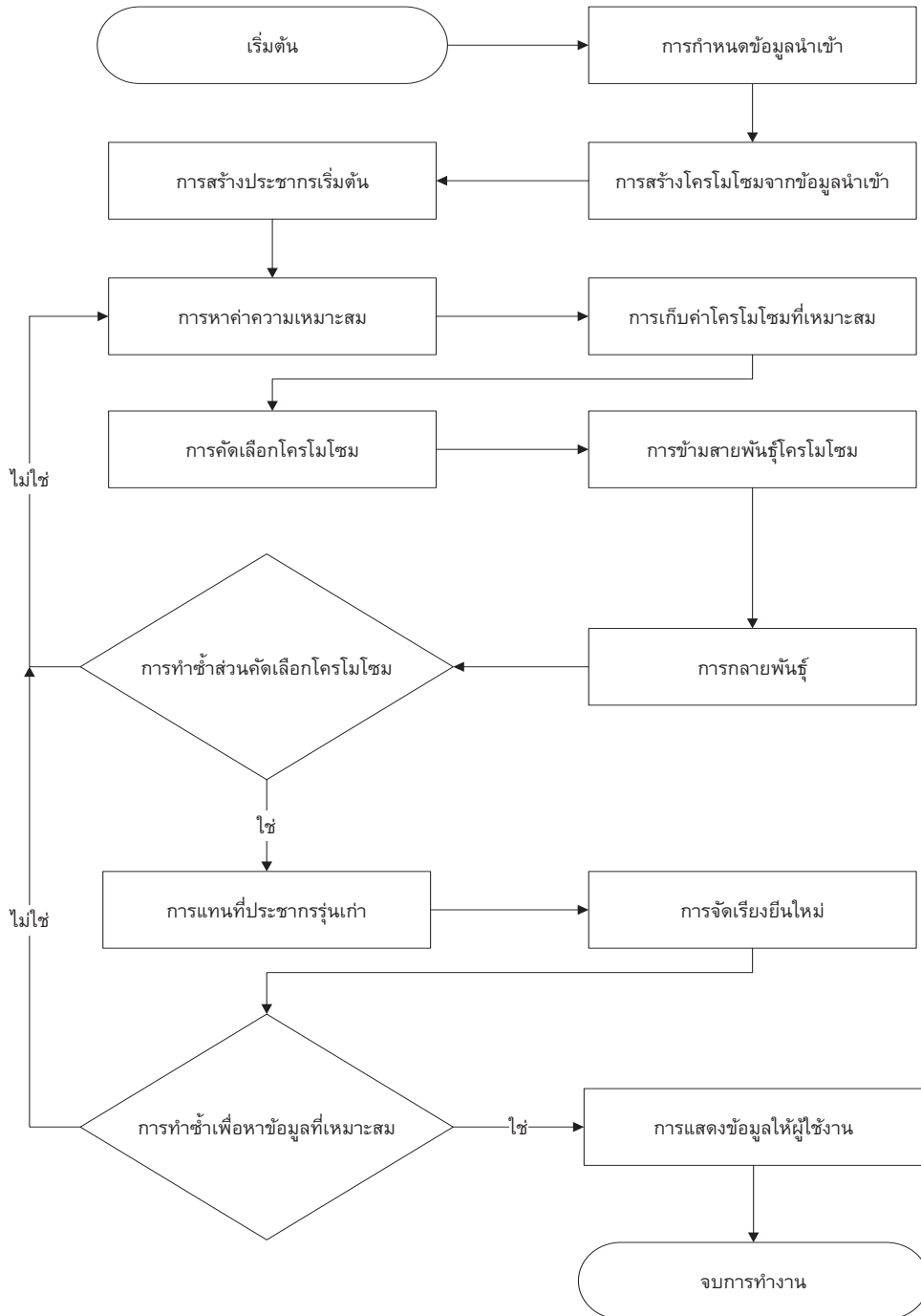


Figure 2 The concept of Thai text search employing text processing pursuant to the semantic-based similarity measurement using Genetic Algorithm

ขั้นตอนที่ 1 การกำหนดข้อมูลนำเข้า

นำข้อมูลคำหลัก และจำนวนรายการข้อมูลที่ต้องการให้แสดงต่อ 1 หน้าจอ เข้าสู่กระบวนการขั้นตอนวิธีเชิงพันธุกรรม โดยกำหนดให้ข้อมูลคำหลัก และจำนวนรายการข้อมูลที่ต้องการให้แสดงต่อ 1 หน้าจอเป็นข้อมูลนำเข้า ตัวอย่างเช่น จำนวนรายการข้อมูลของ URL ที่ต้องแสดงต่อ 1 หน้าจอของโปรแกรมค้นหา ซึ่งกำหนดให้มี 10 รายการต่อ 1 หน้าจอ¹⁷ ดังนั้น จำนวนรายการข้อมูลที่ต้องการให้แสดงต่อ 1 หน้าจอ

จะเท่ากับ 10 รายการค้นหาต่อการแสดง 1 หน้าจอ¹⁸ ดัง Figure 3 โดยกำหนดรายละเอียดของข้อมูลนำเข้า ดังนี้

1. ข้อมูลคำหลัก กำหนดให้เป็นกฎอแจไซรหัสโครโมโซม
2. จำนวนรายการข้อมูลที่ต้องการให้แสดงต่อ 1 หน้าจอ กำหนดให้เป็นข้อมูลจำนวนช่องโครโมโซม

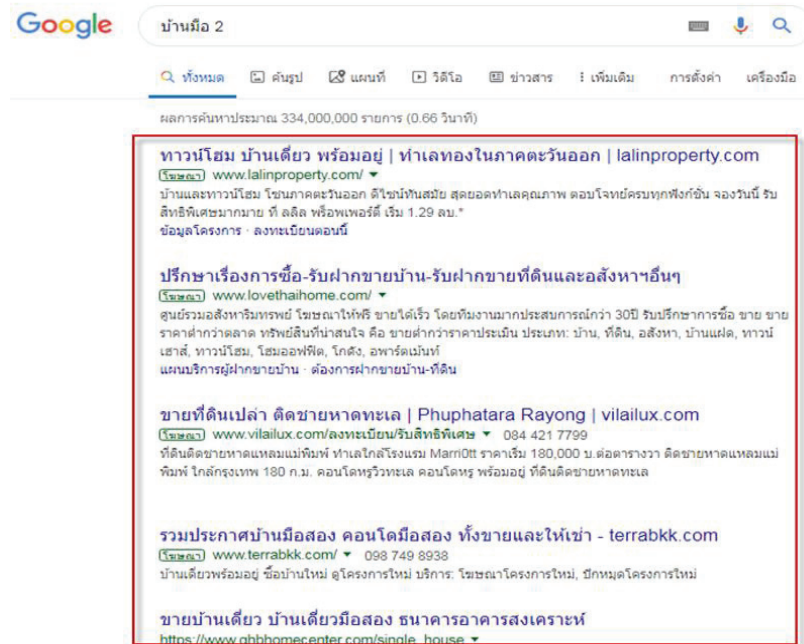


Figure 3 Example of displaying the particulars of information per page of Google Search
 (Source: <https://www.google.co.th/search?hl=th&source=#บ้านมือ2#>, online)

ขั้นตอนที่ 2 การสร้างโครโมโซมจากข้อมูลนำเข้า

สร้างโครโมโซมจากข้อมูลนำเข้าด้วยการเข้ารหัสโครโมโซม (Chromosome encoding) การเข้ารหัสสำหรับข้อมูลนำเข้าจะเป็นในลักษณะ การเป็นตัวแทนการเปลี่ยนแปลง (Permutation representation) เป็นรูปแบบของการเปลี่ยนแปลงของเซตค่ายีน¹⁹ โดยกำหนดให้ยีนของโครโมโซมมีจำนวน n ช่อง (จำนวน n เท่ากับข้อมูล

จำนวนยีนของโครโมโซมตามข้อมูลที่ได้ในขั้นตอนที่ 1) กำหนดให้หมายเลขของช่องเป็น 1 ถึง n แล้วโครโมโซมจะมีลักษณะเป็นสายของยีน โดยยีนจะไม่ซ้ำกัน สมมติถ้าค้นหาคำว่า "สิ่งประดิษฐ์" โครโมโซมจะแสดงลักษณะของยีนตามรหัสชุดคำพ้อง (Synonym set: Synset) จากการแปลงคำที่เกี่ยวข้องกับคำว่า "สิ่งประดิษฐ์" ดัง Figure 4

build.v.02	create.v.02	artifical.a.01	thing.n.01	recover.v.01	appoint.v.01	thing.n.03	artifical.v.02	thing.v.02	create.v.03
------------	-------------	----------------	------------	--------------	--------------	------------	----------------	------------	-------------

Figure 4 Simulation of the chromosome consisting of 10 genes

ขั้นตอนที่ 3 การสร้างประชากรเริ่มต้น

สร้างประชากรเริ่มต้น (Initial population) ด้วยการสร้างโครโมโซมแต่ละตัวโดยการสุ่มยีนไปสร้างเป็นโครโมโซมยีนที่ถูกสุ่มออกมาแล้วจะไม่ใส่กลับไป สุ่มจนครบตามจำนวนยีนสูงสุดที่กำหนด การสร้างโครโมโซม 1 ตัวและทำกระบวนการสุ่มซ้ำจนครบตามจำนวนโครโมโซมที่กำหนดไว้ซึ่งยีนได้มาจากมาตรวัดความคล้ายคลึงเชิงความหมายที่ได้มาจากการนำข้อมูลจากการนำกุญแจไขรหัสโครโมโซมมาค้นหาคำในฐานข้อมูลเครือข่ายคำภาษาไทย (Thai WordNet) โดยสร้างมาจากเครือข่ายคำของพรินสตัน (Princeton's wordnet: PWN) ที่เป็นแหล่งคำศัพท์ภาษาอังกฤษ โดยพิจารณาความคล้ายคลึงกันของเทอม (Term similarity) จาก

ความหมายที่คำนวณได้จากทรัพยากรคำศัพท์ที่มีโครงสร้างแบบอนุกรมวิธาน (Taxonomy) ของเครือข่ายคำ (Wordnet) กล่าวคือ เป็นข้อมูลคลังคำศัพท์ที่มีการจัดหมวดหมู่ของคำไว้เป็นกลุ่มไวพจน์ (Synonym set) อธิบายความหมายรายละเอียดของคำจำกัดความ และความสำคัญทางความหมายระหว่างกลุ่มไวพจน์^{16, 20}

โครงสร้างเครือข่ายคำเริ่มต้นจะเป็นทรัพยากรคำศัพท์ที่สำคัญในการประมวลผลภาษาอังกฤษที่ใช้เป็นจุดเริ่มต้น ส่วนคำของเครือข่ายคำในภาษาอื่นๆ มักถูกออกแบบให้จับคู่กับเครือข่ายคำภาษาอังกฤษเท่าที่ทำได้ โครงสร้างของเครือข่ายจะแทนคำเป็นชุดคำพ้อง (Synonym set: Synset) ซึ่ง Synset จะเป็นของหนึ่งความหมายโดยมีรูปแบบดังนี้¹⁶

"<word>.<pos>.<number>"
 <word> คือ รากศัพท์ของคำ
 <pos> แสดงบทบาทของคำในประโยคมีบทบาทสำคัญ ได้แก่ Noun (n), Verb (v), Adjective (a or s), Adverb (r)
 <number> เป็นเลขเรียงลำดับเริ่มตั้งแต่ 01 เป็นต้นไป โดยในภาษาอังกฤษจะเรียงตามความถี่ของคำที่เกิดขึ้น ตัวอย่างเช่น คำว่า "success" จะมี 4 ลักษณะดังนี้¹⁶
 Synset('success.n.01'),
 Synset('success.n.02'),
 Synset('success.n.03'),
 Synset('achiever.n.01')
 โดยแต่ละแบบจะมีนิยามและตัวอย่างการใช้งานที่แตกต่างกัน เช่น¹⁶

Success.n.01 หมายถึง "an event that accomplishes its intended purpose" ใช้กล่าวถึงตัวความสำเร็จ เช่น "let's call heads a success and tails a failure"

Success.n.02 หมายถึง "an attainment that is successful" ใช้กล่าวถึงการประสบความสำเร็จ เช่น "his new play was a great success"

สำหรับการเข้าถึงคำศัพท์ภาษาอังกฤษ เข้าถึงได้โดยผ่านเครือข่ายคำภาษาไทย (Thai wordnet) ซึ่งเครือข่ายคำภาษาไทยจะมีกลุ่มไวพจน์ภาษาอังกฤษกำกับไว้^{16, 21} ด้วยการสร้างคลาสคำสังโปรแกรมย่อยผ่านคุณสมบัติเครือข่ายคำภาษาไทยที่ใช้วิธีการแปลงคำภาษาไทยมาเป็นภาษาอังกฤษในขั้นตอนนี้ ดังนั้น สรุปได้ว่า ยินจะมีรูปแบบการเก็บข้อมูลในลักษณะดังนี้

"<word>.<pos>.<number>"

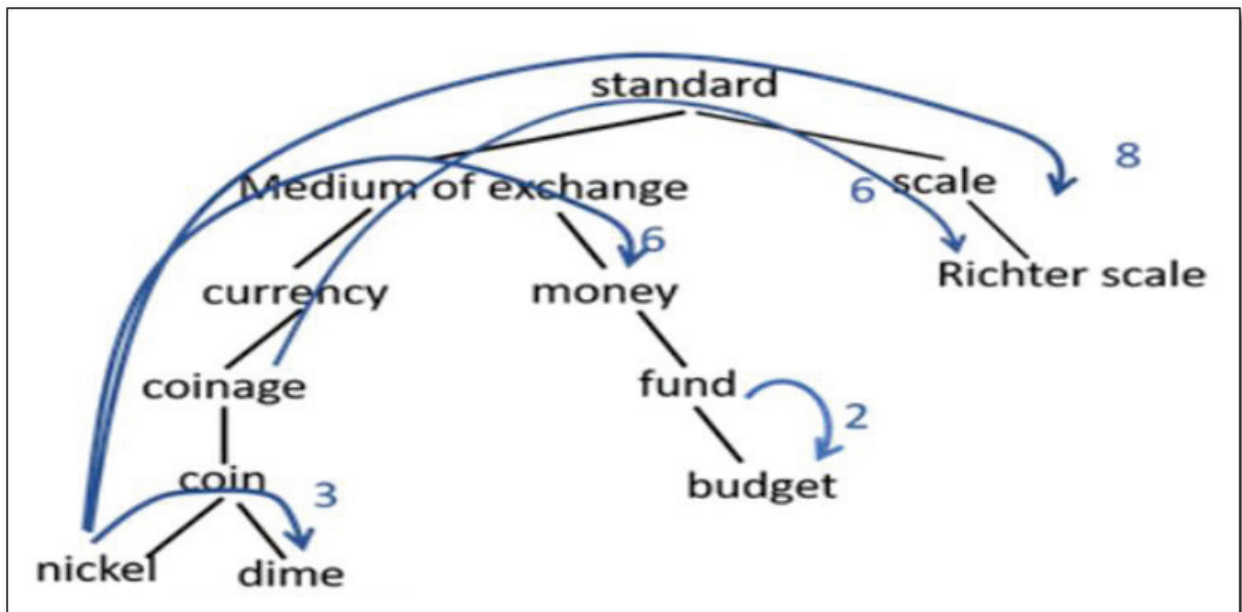


Figure 5 Example of the calculation of Path Distance Similarity⁸

ขั้นตอนที่ 4 การหาค่าความเหมาะสม

หาค่าความเหมาะสม เป็นการหาค่าความเหมาะสมให้แก่โครโมโซมตามฟังก์ชันค่าความเหมาะสม (Fitness function) จากสมการมาตรวัดความคล้ายคลึงเชิงลำดับชั้น ได้แก่ สมการ Path distance similarity^{22, 23} เป็นกลุ่มมาตรวัดที่คำนวณจากความสัมพันธ์ดังนี้

"is-ataxonomy (Hypernym, Hypnoym)" (4)

โดยที่

is-ataxonomy คือ กลุ่มคำศัพท์ที่มีความสัมพันธ์กันในลักษณะความหมายคล้ายคลึงกัน

Hypernym คือ คำศัพท์ที่มีความสัมพันธ์กันในลักษณะความหมายคล้ายคลึงกันในระดับหมวดหลัก (Superclass)

Hypnoym คือ คำศัพท์ที่มีความสัมพันธ์กันในลักษณะความหมายคล้ายคลึงกันในระดับหมวดย่อย (Subclass)

โดยจะคำนวณจากจำนวนโหนด (Node) ในเส้นทางที่สั้นที่สุดที่เชื่อมเทอมทั้งสองในความสัมพันธ์แบบ is-a taxonomy (Hypernym, Hypnym) และคืนค่าคะแนนตั้งแต่ 0 ถึง 1 โดยอย่างยิ่งใกล้ 1 มากเท่าไร ก็คือ มีความคล้ายคลึงมากเท่านั้นแสดงดัง Figure 5 และมีสมการต่อไปนี้¹⁶

$$pathsim = \frac{1}{pathlen(c1, c2)} \quad (5)$$

โดยที่

pathsim คือ ค่าความคล้ายคลึงเชิงความหมาย

pathlen คือ จำนวนโหนดความสัมพันธ์ระหว่างคำศัพท์ *c1* และ *c2*

c1 คือ ข้อมูลกฎเกณฑ์รหัสโครโมโซม

c2 คือ กลุ่มคำในยีน

จาก Figure 5 สามารถคำนวณค่า Path Distance Similarity ดังตัวอย่างดังนี้

$$pathsim(\text{fund, budget})=1/2=0.50$$

$$pathsim(\text{nickel, dime})=1/3=0.33$$

$$pathsim(\text{nickel, money})=1/6=0.16$$

$$pathsim(\text{nickel, Richter scale})=1/8=0.13$$

$$pathsim(\text{coinage, Richter scale})=1/6=0.16$$

ดังนั้น ฟังก์ชันค่าความเหมาะสมมีสมการดังนี้

$$Max(pathsim) \quad (6)$$

จากสมการที่ 3 ค่าความเหมาะสม คือ ค่าความคล้ายคลึงเชิงความหมายที่มีค่าสูงที่สุด

ขั้นตอนที่ 5 การเก็บค่าโครโมโซมที่เหมาะสม

เก็บค่าโครโมโซมที่เหมาะสม โดยการเก็บโครโมโซมที่มีค่าความเหมาะสมสูงที่สุดจำนวนร้อยละ 20 ของโครโมโซมทั้งหมด^{24, 25}

ขั้นตอนที่ 6 การคัดเลือกโครโมโซม

การคัดเลือกโครโมโซมแบบ Roulette wheel selection เป็นการใช้ค่าความเหมาะสม (Fitness) ของโครโมโซมมาพิจารณา โดยโครโมโซมที่มีค่าความเหมาะสมมากก็จะมีโอกาสถูกเลือกขึ้นมาด้วยความน่าจะเป็นที่มากกว่าโครโมโซมที่มีค่าความเหมาะสมน้อย²⁶

ขั้นตอนที่ 7 การข้ามสายพันธุ์โครโมโซม

การข้ามสายพันธุ์โครโมโซม เป็นการข้ามสายพันธุ์แบบ Partially mapped crossover (PMX) เป็นการสุ่มตำแหน่งในโครโมโซมมา 2 ตำแหน่ง จากทั้ง 2 โครโมโซม จะได้ส่วนสำหรับจับคู่ (Mapping section) ที่เป็นส่วนที่อยู่ระหว่างจุดตัด 2 จุด ที่สุ่มมา และยีนที่เหลือจะถูกสลับกันโดยพิจารณาเฉพาะส่วนที่จับคู่^{27, 28}จะได้โครโมโซมใหม่ จำนวน 2 โครโมโซม เลือกโครโมโซมแรกมาเพียง 1 ตัว สำหรับความหลากหลายของคำตอบที่ทำให้ประชากรรุ่นใหม่สืบทอดจากประชากรรุ่นก่อนหน้าให้มากที่สุด

ขั้นตอนที่ 8 การกลายพันธุ์

กลายพันธุ์ คือ ทำการกลายพันธุ์โครโมโซมแบบ Swap mutation โดยมีขั้นตอนเริ่มจากการสุ่มตำแหน่งยีนในโครโมโซมมา 2 ตำแหน่ง และสลับค่าระหว่าง 2 ตำแหน่งที่ถูกเลือกมา²⁹

ขั้นตอนที่ 9 การทำซ้ำส่วนคัดเลือกโครโมโซม

ทำซ้ำส่วนคัดเลือกโครโมโซมในขั้นตอนที่ 5 ถึงขั้นตอนที่ 7 จนได้โครโมโซมที่เป็นประชากรรุ่นใหม่ครบตามที่กำหนดในจำนวนโครโมโซม คือ อีกร้อยละ 80 ของจำนวนโครโมโซม^{27, 28}

ขั้นตอนที่ 10 การแทนที่ประชากรรุ่นเก่า

แทนที่ประชากรรุ่นเก่าด้วยประชากรรุ่นใหม่ แบ่งเป็นร้อยละ 20 จากโครโมโซมที่ได้จากขั้นตอนที่ 5 และร้อยละ 80 ที่เหลือจากโครโมโซมที่ได้จากขั้นตอนที่ 9²⁹

ขั้นตอนที่ 11 การจัดเรียงยีนใหม่

จัดเรียงยีนใหม่หลังจากได้โครโมโซมที่เป็นประชากรรุ่นใหม่ครบแล้ว ทำการจัดเรียงยีนสำหรับจำนวนกลุ่มคำที่จะแสดง เริ่มจากกลุ่มคำแรก แล้วจัดเรียงใหม่ เมื่อได้ค่าความเหมาะสมที่มากกว่าเดิม จะแทนโครโมโซมที่มีค่าเดิมด้วยโครโมโซมที่มีค่าใหม่ ทำการจัดเรียงยีนสำหรับจำนวนกลุ่มคำที่จะแสดงในโครโมโซมถัดไปจนถึงโครโมโซมสุดท้ายด้วยวิธีเดียวกันจนได้ค่าสูงที่สุดของร้อยละ 20 ของโครโมโซมทั้งหมด^{27, 28, 29}

ขั้นตอนที่ 12 การทำซ้ำเพื่อหาข้อมูลที่เหมาะสม

ทำซ้ำเพื่อหาข้อมูลที่เหมาะสม เป็นการซ้ำตั้งแต่ขั้นตอนที่ 4 ถึง ขั้นตอนที่ 12^{27, 28, 29}

ขั้นตอนที่ 13 การแสดงข้อมูลให้ผู้ใช้งาน

แสดงข้อมูลให้ผู้ใช้งาน เมื่อได้ยีนในโครโมโซมที่เหมาะสมที่สุดแล้ว ให้นำกลุ่มคำที่อยู่ในยีนมาสืบค้นข้อมูลที่มีอยู่ในฐานข้อมูลมาแสดงในระบบสืบค้นข้อมูลเพื่อแสดงข้อมูลให้กับผู้ใช้งาน

การทดลองใช้โปรแกรมเปรียบเทียบวิธีการสืบค้นข้อความภาษาไทยระหว่างเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมและเทคนิคการ

ค้นหาด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง ตัวอย่างโปรแกรมแสดงได้ดัง Figure 6 และ Figure 7 โดยมีรายละเอียดสรุปผลการประเมินเปรียบเทียบ ดัง Table 1

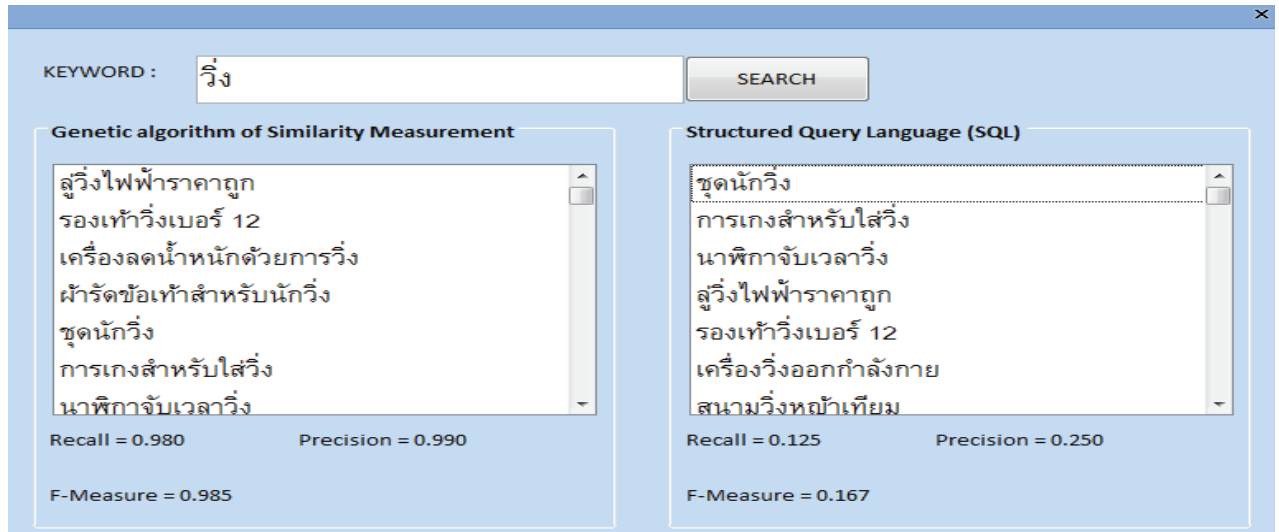


Figure 6 Example of the testing programs for the comparison between Thai text search in term of semantic-based similarity measurement technique using Genetic Algorithm and the search technique using Structured Query Language (SQL), single word search

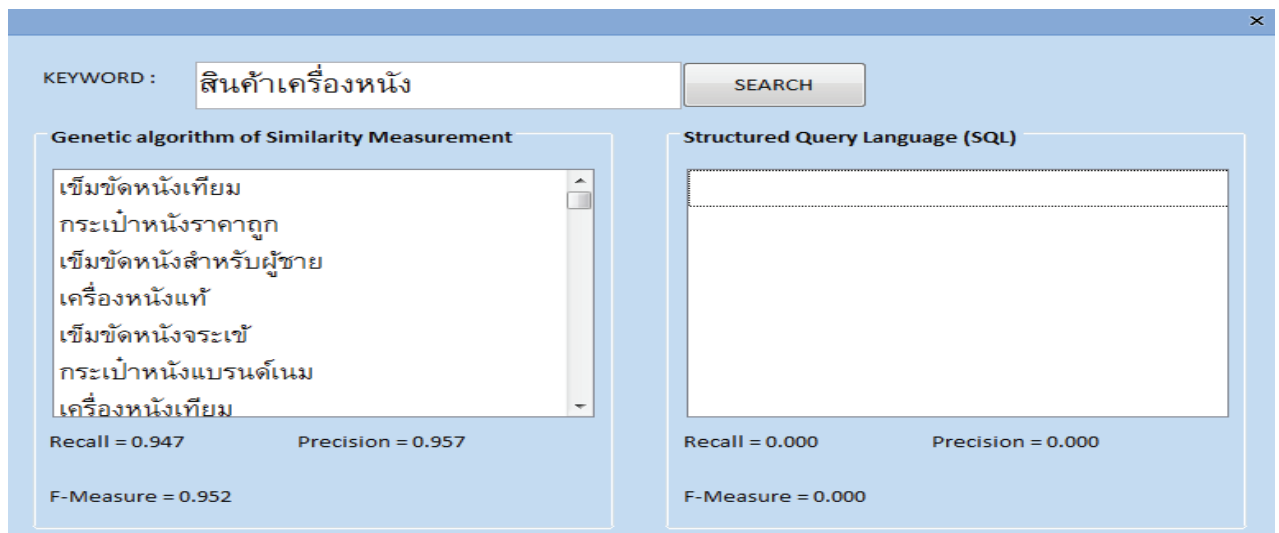


Figure 7 Example of the testing programs for the comparison between Thai text search in term of semantic-based similarity measurement technique using Genetic Algorithm and the search technique using Structured Query Language (SQL), sentence search

Table 1 Evaluation Result

คุณลักษณะ	ค่าเฉลี่ยความระลึก	ค่าเฉลี่ยความเที่ยงตรง	ค่าเฉลี่ยเอฟเมเชอร์
เทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม	0.981	0.986	0.983
เทคนิคการค้นหาด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง	0.155	0.295	0.203

ผลการทดสอบการสืบค้นข้อความภาษาไทย จำนวน 10 ครั้ง และหาค่าเฉลี่ยประสิทธิภาพตามเกณฑ์การประเมิน ประสิทธิภาพ 3 ด้าน ได้แก่ ค่าความระลึก ค่าความเที่ยงตรง และค่าเอฟเมเชอร์ พบว่า เทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม มีค่าเฉลี่ยความระลึก เท่ากับ 0.981 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ 0.986 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.983 ส่วนเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง มีค่าเฉลี่ยความระลึก เท่ากับ 0.155 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ 0.295 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.203

จากผลการศึกษา พบว่า การสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมมีประสิทธิภาพสูงกว่าเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง โดยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม ประกอบด้วย 13 ขั้นตอน ได้แก่ 1) การกำหนดข้อมูลนำเข้า 2) การสร้างโครโมโซมจากข้อมูลนำเข้า 3) การสร้างประชากรเริ่มต้น 4) การหาค่าความเหมาะสม 5) การเก็บค่าโครโมโซมที่เหมาะสม 6) การคัดเลือกโครโมโซม 7) การข้ามสายพันธุ์โครโมโซม 8) การกลายพันธุ์ 9) การทำซ้ำส่วนคัดเลือกโครโมโซม 10) การแทนที่ประชากรรุ่นเก่า 11) การจัดเรียงยีนใหม่ 12) การทำซ้ำเพื่อหาข้อมูลที่เหมาะสม และ 13) การแสดงผลข้อมูลให้ผู้ใช้งาน

วิจารณ์และสรุปผล

ผลการศึกษา พบว่า แนวคิดการสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม ประกอบด้วย 13 ขั้นตอน ได้แก่ 1) การกำหนดข้อมูลนำเข้า 2) การสร้างโครโมโซมจากข้อมูลนำเข้า 3) การสร้างประชากรเริ่มต้น 4) การหาค่าความเหมาะสม 5) การเก็บค่าโครโมโซมที่เหมาะสม 6) การคัดเลือกโครโมโซม 7) การข้ามสายพันธุ์โครโมโซม 8) การกลายพันธุ์ 9) การทำซ้ำส่วนคัดเลือกโครโมโซม 10) การแทนที่ประชากรรุ่นเก่า 11) การจัดเรียงยีนใหม่ 12) การทำซ้ำเพื่อหาข้อมูลที่เหมาะสม และ 13) การแสดงผลข้อมูลให้ผู้ใช้งาน ส่วนการสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมมีประสิทธิภาพสูงกว่าการเข้าถึงข้อมูลด้วยเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง โดยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม มีค่าเฉลี่ยความระลึก เท่ากับ 0.981 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ 0.986 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.983 ส่วนเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้าง มีค่าเฉลี่ยความระลึก เท่ากับ 0.155 ค่าเฉลี่ยความเที่ยงตรง เท่ากับ

0.295 และค่าเฉลี่ยเอฟเมเชอร์ เท่ากับ 0.203 ซึ่งให้เห็นว่าการเข้าถึงข้อมูลด้วยเทคนิคการค้นหาคำด้วยภาษาแบบสอบถามข้อมูลอย่างมีโครงสร้างจะค้นหาคำข้อมูลได้เฉพาะคำที่เป็นคำโดดและต้องเป็นคำที่มีความหมายที่ตรงตัวเท่านั้น (Figure 6) ส่วนการสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมมีประสิทธิภาพการค้นหาคำที่สูงทั้งคำค้นหาที่เป็นคำโดดและคำที่เป็นรูปประโยค โดยค้นหาได้ทั้งคำที่มีความหมายโดยตรงและคล้ายคลึงได้ (Figure 7) ซึ่งผลการศึกษาสอดคล้องกับการศึกษาของ Netisopakul (2019)⁸ ที่ศึกษาเกี่ยวกับการทดสอบมาตรฐานวัดความคล้ายคลึงเชิงความหมายสำหรับคำเหมือนและคำตรงข้ามภาษาไทย ซึ่งการหาความคล้ายคลึงเชิงความหมายสามารถประยุกต์ใช้วัดความคล้ายคลึงในคำภาษาไทยได้อย่างมีประสิทธิภาพสูง และผลการศึกษาได้รับประโยชน์ดังนี้

1. ได้แนวความคิดการสืบค้นข้อมูลข้อความภาษาไทยด้วยการประมวลผลข้อความตามมาตรฐานวัดความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม
2. นักพัฒนาโปรแกรมสามารถนำผลการวิจัยไปใช้สำหรับอ้างอิงในการพัฒนาการสืบค้นข้อมูลให้มีประสิทธิภาพสูงขึ้นได้
3. นักวิจัยที่เกี่ยวข้องกับสาขาการจัดการสารสนเทศ สามารถนำผลการวิจัยไปใช้สำหรับอ้างอิงเป็นแนวทางในการวิจัยในอนาคตได้

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. นักพัฒนาโปรแกรมที่มีความต้องการพัฒนาระบบสืบค้นข้อมูลข้อความภาษาไทยควรนำแนวคิดการสืบค้นข้อมูลข้อความภาษาไทยด้วยการประมวลผลข้อความตามมาตรฐานวัดความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมไปประยุกต์ใช้ในการพัฒนาระบบสืบค้นข้อมูลข้อความภาษาไทย
2. การนำแนวคิดการสืบค้นข้อมูลข้อความภาษาไทยด้วยการประมวลผลข้อความตามมาตรฐานวัดความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรมไปใช้งานควรนำไปทดสอบหาค่าประสิทธิภาพก่อนนำไปใช้งาน เนื่องจากในขั้นตอนที่ 3 การสร้างประชากรเริ่มต้น ควรที่จะสร้างคลาสคำสั่งโปรแกรมย่อยผ่านคุณสมบัติเครือข่ายคำภาษาไทยในการเข้าถึงคำศัพท์ภาษาอังกฤษซึ่งเครือข่ายคำภาษาไทยจะมีกลุ่มไวยากรณ์ภาษาอังกฤษกำกับไว้
3. การกำหนดยีนของโครโมโซมอาศัยคุณสมบัติเครือข่ายคำภาษาไทยในการเข้าถึงคำศัพท์ภาษาอังกฤษ

ซึ่งจำนวนคำศัพท์ที่มีสูงสุดขึ้นอยู่กับเครือข่ายคำภาษาไทย ดังนั้น ควรมีการปรับปรุงรุ่น ของเครือข่ายคำภาษาไทยอย่างสม่ำเสมอเพื่อความทันสมัยของข้อมูล

ข้อเสนอแนะในการทำวิจัยต่อไป

1. การศึกษาครั้งนี้ ได้ใช้การหาความคล้ายคลึงเชิงความหมายแบบ Path distance similarity เป็นพื้นฐานของสมการหาค่าความเหมาะสมเท่านั้น การศึกษาครั้งต่อไป ควรใช้การหาความคล้ายคลึงเชิงความหมายแบบอื่นๆ ได้แก่ Leacock chodorow similarity, Wu-palmer Similarity, Resnik similarity, Jiang-conrath, Similarity และ Lin similarity มาใช้ในการศึกษาในอนาคต

2. การศึกษาครั้งต่อไปควรศึกษาเกี่ยวกับประสิทธิภาพของการหาความคล้ายคลึงเชิงความหมายแบบต่างๆ มาเปรียบเทียบหามาตรวัดความคล้ายคลึงเชิงความหมายที่มีประสิทธิภาพสูงสุดในการนำมาใช้เป็นพื้นฐานของสมการหาค่าความเหมาะสมของการสืบค้นข้อมูลข้อความภาษาไทยด้วยเทคนิคการหาความคล้ายคลึงเชิงความหมายด้วยขั้นตอนวิธีเชิงพันธุกรรม

เอกสารอ้างอิง

- Sapsomboon B. Medical biology information search. Available from: URL: <https://www.si.mahidol.ac.th/simi/pdf/pubmed> Assessed May 16, 2019.
- Laothongdee D. Search engine algorithm study and web promotion analysis. (Maters Thesis, Thammasat University). 2003.
- Google. Advanced Search. Available from: URL: https://google.com/advanced_search May 16, 2019.
- Boonrawd P, Srisuktaksin P. Retrieving Big Data Using NoSQL and Semantic Web Technology. The Journal of KMUTNB 2015 ; 5(2): 255-264.
- Korler P. SQL Issues. Available from: URL: <http://stackoverflow.com/quest/SQL/Like-not-found> May 16, 2019.
- Svetthitikun Y, Supasetsiri P, Pakdeeronachit S. The analysis of factors for e-commerce affects the purchasing behavior of consumers' fashion apparel. The Journal of Social Communication Innovation 2017 ; 5(1): 21-31.
- Vongsirisin K. The product value affecting the customer's purchase decision on V CARE U Natural & Premium Dog Food in Bangkok. Maters Thesis, Bangkok University). 2016. (in thai).
- Netisopakul P. Text similarity measurement for synonyms term and antonyms term of Thai. Available from: URL: http://www.it.kmitl.ac.th/index.php/main_journal/article May 10, 2019. (in thai).
- Pattrawadee M. Research Design for Mixed Method Research. Journal of the Association of Researchers 2016 ; 2(2): 19-31.
- Wellington J. Education research: contemporary issues and practical approaches. London: Continuum ; 2000.
- Scott, J. Social research and documentary sources. Sage Benchmarks in Social Research Methods, Documentary Research Volume 1. Thousand Oaks, California: SAGE Publication. 2006 ; 3-40.
- Rigotti, B. How do I get my website on the first page of Google. Available from: URL: <https://cristersmedia.com/how-do-i-get-my-website-on-the-first-page-of-google/> May 16, 2019.
- Mathworks. Find global minima for highly nonlinear problems. Available from: URL: <https://mathworks.com/genetic-algorithm.html> May 16, 2019.
- Salah FAAS, Mohamad SA, El-Horbaty ESM. Swarm Intelligent Algorithms for solving load balancing in cloud computing. Egyptian Computer Science Journal 2019 ; 43(1): 45-57.
- Whitley D. Next generation genetic algorithms: a user's guide and tutorial. In: Gendreau M, Potvin JY. (eds) Handbook of metaheuristics. International Series in Operations Research & Management Science 2019 ; 272: Springer, Cham.
- Netisopakul P. text similarity measurement for synonyms term and antonyms term of Thai. Available from: URL: http://www.it.kmitl.ac.th/~journal/index.php/main_journal/article/download/100/49 May 16, 2019. (in thai).
- Rigotti B. How do I get my website on the first page of Google. Available from: URL: <https://cristersmedia.com/how-do-i-get-my-website-on-the-first-page-of-google/> May 16, 2019.
- Google. Data search. Available from: URL: <https://www.google.co/search> May 16, 2019.

19. Eiben K, Agoston E, Smith JE. *Introduction to Evolutionary Computing*, Springer. Goldberg, David E. (1989), *Genetic Algorithm in Search Optimization and Machine Learning*, New York: Addison Wesley. 2003.
20. Sagum RA, Ramos AD, Llanes MT. FICOBU: Filipino WordNet construction using decision tree and language modeling. *International Journal of Machine Learning and Computing* 2019 ; 9(1): 103-107.
21. Thoongsup SR. Thai WordNet construction. *Proceedings of the 7th workshop on Asian language resources Association for computational linguistics* 2009: 139-144.
22. Yang D, Powers DM. Measuring semantic similarity in the taxonomy of WordNet. *Australian Computer Society* 2005 ; 38(1): 315-322.
23. Cai Y, Zhang Q, Che WA. Hybrid approach for measuring semantic similarity based on IC-weighted path distance in WordNet. *Journal of Intelligent Information Systems* 2018 ; 51(1): 33-47.
24. Fazlia W, Zagerb AAK, Rozaidaa G. Using improved firefly algorithm based on genetic algorithm crossover operator for solving optimization problems. *Journal of Intelligent & Fuzzy Systems* 2019 ; 36(2): 1547-1562.
25. Haoweia Z, Junweia X, Jiaanga G, Zhaojianb Z, Binfengc ZA. Hybrid adaptively genetic algorithm for task scheduling problem in the phased array radar *European Journal of Operational Research* 2019 ; 272 (3): 868-878.
26. Krishna GJ, Ravi V. Feature subset selection using adaptive differential evolution: an application to banking. *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data* 2019: 157-163.
27. Juneja SS, Saraswat P, Singh K, Sharma J, Majumdar R, Chowdhary S. Travelling salesman problem optimization using genetic algorithm. *Amity International Conference on Artificial Intelligence (AICAI)* 2019.
28. Mishra A, Sainul IA, Bhuyan, S, Deb S, Sen D, Deb AK. Development of a flexible assembly system using industrial robot with machine vision guidance and dexterous multi-finger gripper. *Precision Product-Process Design and Optimization* 2018 ; 31-71.
29. Yifei T, Meng Z, Jingwei L, Dongbo L, Yulin W. Research on intelligent welding robot path optimization based on GA and PSO algorithms. *IEEE Access* 2019 ; 6: 65397-65404.