

วงศ พงษ์พงศ์สรรค์ 2551: การจัดกลุ่มข้อมูลด้านความปลอดภัยของอาหารด้วยวิธีการ  
แบบสองขั้นตอน โดย SOM และขั้นตอนวิธีเชิงพันธุกรรม ปรินญาวิทยาศาสตร์  
มหาบัณฑิต (วิทยาการคอมพิวเตอร์) สาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการ  
คอมพิวเตอร์ ปรธานกรรมการที่ปรึกษา: รองศาสตราจารย์อนงค์นาฏ ศรีวิหค, Ph.D.  
75 หน้า

งานวิจัยนี้นำเสนอหลักการจัดกลุ่มข้อมูลด้านความปลอดภัยอาหารออกเป็นกลุ่ม ตาม  
ความคล้ายคลึงกันของข้อมูล ซึ่งพิจารณาจากค่าสำคัญของเอกสารที่กำหนดขึ้น โดยผู้เชี่ยวชาญ  
หรือเจ้าของเอกสารและผ่านการตัดคำและกรองคำหยุด จากนั้น ใช้ระบบน้ำหนัก Inverse  
Document Frequency (IDF) เพื่อปรับค่า น้ำหนักของคำสำคัญ เพื่อทดลองเปรียบเทียบระหว่าง  
จัดกลุ่มข้อมูลที่ไม่มีการปรับค่าน้ำหนัก (CN) กับการจัดกลุ่มข้อมูลที่มีการนำระบบน้ำหนักมาใช้  
(CW)

โดยการทดลองจะนำเสนอการใช้ขั้นตอนวิธีการจัดกลุ่มข้อมูลตามค่าสำคัญของเอกสาร  
แบบ 2 ขั้นตอน ระหว่างขั้นตอนวิธี Kohonen's Self-Organizing Maps (SOM) ซึ่งเป็นขั้นตอนวิธี  
ทางโครงข่ายประสาทเทียม (Neural Network) มาใช้ในการหาจำนวนกลุ่มที่เหมาะสมกับข้อมูล  
จากนั้น นำจำนวนกลุ่มที่ได้มาทำการจัดกลุ่มเปรียบเทียบกันระหว่างขั้นตอนวิธีเชิงพันธุกรรม  
(Genetic Algorithms: GA) กับขั้นตอนวิธีการจัดกลุ่มข้อมูลแบบเค-มีน (K-Means) โดย  
เปรียบเทียบประสิทธิภาพการจัดกลุ่มจากค่าความแปรปรวน (Variance) ความคล้ายคลึงของ  
เอกสารภายในกลุ่มโดยใช้ค่า RMSSTD และความแตกต่างระหว่างแต่ละกลุ่มโดยใช้ค่า (RS)  
เพื่อให้ทราบถึงขั้นตอนวิธีที่เหมาะสมที่ใช้ในการจัดกลุ่มข้อมูลด้านความปลอดภัยอาหาร

จากการวิจัยพบว่า การจัดกลุ่มที่ให้ผลการจัดกลุ่มที่ดีที่สุดคือ จำนวนกลุ่ม 8 กลุ่มและ  
ขั้นตอนวิธีที่ดีที่สุดคือ ขั้นตอนวิธีการจัดกลุ่มแบบ SOM และ Genetic Algorithms เมื่อ  
เปรียบเทียบกับผลการนำระบบน้ำหนักมาใช้พบว่าการใช้ระบบน้ำหนักจะให้ผลการจัดกลุ่มที่ดีกว่า

วงศ พงษ์พงศ์สรรค์  
ลายมือชื่อนิติ

ลายมือชื่อประธานกรรมการ

14 / 5 / 2551

Wongkot Pojpongsan 2008: Clustering of Food Safety Documents by Using Two Step Clustering: SOM and Genetic Algorithms. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Associate Professor Anongnart Srivihok, Ph.D. 75 pages.

This research proposes an approach to improve document clustering technique to segment Food Safety documents by the similarity of documents. The keywords that provided by the specialist were used to represent the documents. Keywords were extracted from documents by using Stemming and Stopping Word Removal technique and then Inverse Document Frequency (IDF) was used to calculate keywords' weight. The comparison of clustering documents by using weighted keywords (CW) and non-weighted keywords (CN) was conducted.

Two step clustering technique was improved in this study to cluster Food Safety documents. These techniques included (1) Self-Organizing Maps (SOM) and (2) Genetic Algorithms (GA) and (3) K-Means algorithm. SOM was used find the number of clusters. Then both GA and K-Means were used to cluster data and evaluate their performances. Evaluation methods included Variance analysis, Root-mean-square Standard Deviation (RMSSTD) and R-Square (RS). RMSSTD was used to find the cohesion of data or the difference between data in the same cluster while RS was used to find the differences between clusters.

Results revealed that the appropriate cluster numbers was equal to 8 while having the best of value variance analysis, RMSTD and RS. As well, Genetic Algorithms was the best performance algorithm in clustering. When we use document weighting technique. The results of clustering with weighted keyword Documents (CW) were better than clustering with non-weighted keywords Documents (CN).

Wongkot Pojpongsan

Student's signature

Anongnart Srivihok

Thesis Advisor's signature

14 / 5 / 2008