

หัวข้อวิทยานิพนธ์	ตัวกรองอีเมลล์ขยะด้วยทฤษฎีเจเนติกอัลกอริทึม
นักศึกษา	นางสาวอุษารัตน์ แสนปากดี
รหัสประจำตัว	47060802
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2549
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ดร.สมศักดิ์ วลัยรัชต์

### บทคัดย่อ

ในปัจจุบัน มีการคิดค้นวิธีการกำจัดอีเมลล์ขยะขึ้นมาหลากหลายวิธี หนึ่งในวิธีที่นิยมคือการใช้เครื่องเรียนรู้ (Machine Learning) วิทยานิพนธ์นี้ได้นำเสนอ วิธีการกรองอีเมลล์ขยะด้วยทฤษฎีเจเนติกอัลกอริทึม (Genetic Algorithm) ซึ่งกระบวนการหลักๆ ของเจเนติกอัลกอริทึมที่นำมาใช้ประกอบด้วย การคัดเลือก (Selection) การครอสโอเวอร์ (Crossover) และการมิวเตชัน (Mutation) กระบวนการเหล่านี้ถูกนำมาใช้ในการสร้างแม่แบบของอีเมลล์ ซึ่งสืบทอดมาจากอีเมลล์ที่มีอยู่เดิม แม่แบบของอีเมลล์ที่ถูกสร้างขึ้นมาใหม่จะถูกคัดเลือกเฉพาะแม่แบบที่มีความเหมาะสมมากกว่าแม่แบบอื่นๆ เอาไว้ ตามหลักการของเจเนติกอัลกอริทึม (Fittest of Survival) ทั้งนี้ แม่แบบของอีเมลล์ที่ถูกสร้างขึ้นมาใหม่ทำให้ตัวกรองได้เรียนรู้รูปแบบของอีเมลล์ใหม่ๆ ส่งผลให้เกิดทางเลือกที่จะนำมาใช้ในการตัดสินใจว่าอีเมลล์เป็นอีเมลล์ประเภทใดเพิ่มมากขึ้น ซึ่งเมื่อนำตัวกรองอีเมลล์ขยะที่นำเสนอมาทดสอบประสิทธิภาพโดยเปรียบเทียบกับตัวกรองอีเมลล์ขยะโดยประยุกต์ใช้ทฤษฎีเบย์ เขียนพบว่ามีประสิทธิภาพที่ดีกว่า โดยมี ค่า Accuracy 87.05% ค่า Recall 88.50% และค่า Precision 86.35%

<b>Thesis Title</b>	Spam Mail Filtering Using Genetic Algorithm
<b>Student</b>	Miss Usarat Sanpakdee
<b>Student ID.</b>	47060802
<b>Degree</b>	Master of Engineering
<b>Program</b>	Computer Engineering
<b>Year</b>	2006
<b>Thesis Advisor</b>	Dr. Somsak Walairachth

### **ABSTRACT**

Nowadays, there are several methods to eliminate spam mail. One of popular method is eliminating spam mail by using machine learning. In this paper, we present a spam mail filtering using genetic algorithm. We applied Genetic operator, crossover and mutation to create varieties of mails templates which inherit from old mails. Therefore, it saves time for filter to prepared training set and not need large training set to learn like others method. New mails templates which have more fitness will be selected according to "fittest of survival" of genetic principle. New mail templates are the result of new learning and new choices which improve the spam mail filter efficiency and decided whether the incoming e-mail is spam. In this thesis, we compare the efficiency of filter with Bayesian spam filter. We found that the propose method has more efficient than Bayesian spam filter. With the accuracy 87.05%, recall is 88.50% and precision 86.35%.