48309322 : สาขาวิชาเทคโนโลยีสารสนเทศ
คำสำคัญ : การคัดกรองเว็บไซต์ที่ไม่เหมาะสม / ซัพพอร์ตเวกเตอร์แมชชีน / การวิเคราะห์องค์ประกอบหลัก

ชาญพัฒณ์ ภินันท์รัชต์ธร : การพัฒนาระบบคัดกรองเว็บไซต์ที่ไม่เหมาะสมในระดับมัธยมศึกษา. อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผศ.ดร.ปานใจ ธารทัศนวงศ์. 154 หน้า.

วิทยานิพนธ์นี้ได้พัฒนาขึ้นเพื่อการคัดกรองเว็บไซต์ที่ไม่เหมาะสมภายใต้ระบบปฏิบัติการลีนุกซ์ โดยใช้โปรแกรม Squid ซึ่งมี ACL (Access Control List) เป็นตัวควบคุมการใช้งานเว็บไซต์ และได้นำเอาอัลกอริทึมการวิเคราะห์องค์ประกอบหลัก PCA (Principal Component Analysis) และอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน SVM (Support Vector Machine) มาใช้ในการคัดกรองเว็บไซต์ เพื่อช่วยควบคุมการใช้งานอินเทอร์เน็ตภายในโรงเรียนให้นักเรียนได้รับข้อมูลที่เหมาะสม

หลักการสร้างระบบคัดกรองเว็บไซต์ที่ไม่เหมาะสมนั้น ได้ใช้โครงสร้างของเว็บไซต์ตาม TAG ต่าง ๆ ดังนี้ META, IMG, A HREF, SCRIPT, TITLE และ BODY มาเป็นองค์ประกอบในการสร้างระบบคัดกรอง โดยนำเว็บไซต์ไม่เหมาะสมมาหาความสัมพันธ์ และใช้ PCA มาพิจารณาค่าความแปรปรวนในแต่ละองค์ประกอบเพื่อสร้างตัวแบบ จากนั้นทำการทดสอบโดยนำเว็บไซต์ความรุนแรง ยาเสพติด 200 เว็บไซต์ และเว็บไซต์ลามกอนาจาร 200 เว็บไซต์ มาทดสอบกับระบบ ผลของการศึกษาตัวแบบพบว่าระบบที่สร้างขึ้นสามารถแบ่งแยกเว็บไซต์ไม่เหมาะสมออกจากเว็บไซต์ปกติได้ โดยมีความถูกต้องสำหรับกลุ่มที่มีความรุนแรง ยาเสพติด 89.5 % และกลุ่มลามกอนาจาร 94.5% และทำการเปรียบเทียบวิธีการคัดกรองเว็บไซต์ที่ไม่เหมาะสมกับอัลกอริทึม SVM โดยใช้ข้อมูลทดสอบชุดเดียวกัน พบว่าวิธี SVM มีความถูกต้องในการคัดกรองสำหรับกลุ่มเว็บไซต์ความรุนแรง ยาเสพติด 89 % และกลุ่มเว็บไซต์ลามกอนาจาร 91% ในขณะที่ วิธี PCA สามารถแบ่งกลุ่มข้อมูลได้ดีกว่าวิธี SVM และจากการวิเคราะห์องค์ประกอบต่าง ๆ ของเว็บไซต์ทำให้รู้ว่าองค์ประกอบที่มีความสำคัญมากระหว่างเว็บไซต์ปกติและเว็บไซต์ไม่เหมาะสมนั้นคือองค์ประกอบของคำที่อยู่ภายใน BODY TAG อย่างไรก็ตามเทคนิคนี้ยังไม่ครอบคลุมกลุ่มคำที่มีความกำกวมและกรณีที่เว็บไซต์นั้นๆไม่สามารถตรวจสอบ HTML Code ได้

ภาควิชาคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ปีการศึกษา 2552
ลายมือชื่อนักศึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ .....

48309322 : MAJOR : INFORMATION TECHNOLOGY
KEY WORD : WEB FILLTER / SUPPORT VECTOR MACHINE / PRINCIPAL COMPONENT ANALYSIS
CHANPAT PINUNRATCHATHORN : SYSTEM DEVELOPING OF WEB CONTENT FILTERING IN SECONDARY SCHOOL. THESIS ADVISOR : ASST.PROF. PANJAI TANTATSANAWONG, Ph.D. 154 pp.

This thesis has developed to filter inappropriate websites under the Linux operating system using Squid proxy software with ACL (Access Control List), which controlled of the web site accessing. The research applied two algorithms to analyze including: Principal Component Analysis (PCA) and Support Vector Machine (SVM). These algorithms are used for filtering websites to help control Internet access in schools and allowed students to receive the appropriate information.

Principal of the inappropriate websites filtering system used the TAG elements structure of the site as the following META, IMG, A HREF, SCRIPT, TITLE and BODY to create the filtering system. The inappropriate website brought to model the relationship and used PCA to determine the value of transformation for each component to create the models. Then model was tested by the site of 200 violence and drug websites, 200 pornographic web sites. The results of model evaluation showed that systems can filter inappropriate websites from normal websites, with accuracy for a group of violent and drug 89.5% and pornographic group 94.5%. Researcher also compared to the other filtering inappropriate algorithms called SVM using the same set of test data. The results showed that SVM method has the accuracy of filtering websites as the following: violent and drug 89%, pornographic 91%, while the PCA algorithms can segment data better than SVM. From the analysis showed that the elements in BODY TAG of various sites are very important to classify normal and inappropriate websites. However, these techniques do not apply to group words that are ambiguous and websites which can not check the HTML Code.

Department of Computing Graduate School, Silpakorn University Academic Year 2009
Student's signature .....
Thesis Advisor's signature .....