

211041

งานวิจัยในการตรวจจับลิงก์ฟาร์มโดยทั่วไปมีแนวคิดในการหาอัลกอริทึมในการตรวจจับให้มีความถูกต้องเพียงอย่างเดียวโดยไม่ได้คำนึงถึงโครงสร้างลิงก์ฟาร์ม ดังนั้นจึงมีการพัฒนาไวยากรณ์กราฟมาใช้อธิบายตัวแบบของลิงก์ฟาร์ม และพัฒนาไวยากรณ์กราฟสำหรับตรวจจับลิงก์ฟาร์ม ซึ่งในกระบวนการตรวจจับลิงก์ฟาร์มนั้นมีการนับจำนวนของการใช้โปรดักชันจากข้อมูลสอน และมีอัลกอริทึมตรวจจับลิงก์ฟาร์มซึ่งใช้กฎตรรกศาสตร์ในการจำแนกความเป็นสแปมโฮสจากระดับตรวจจับทั้งหมด 20 ระดับ ผลการทดลองพบว่าเมื่อเปรียบเทียบกับงานวิจัยที่เกี่ยวข้องประสิทธิภาพในการตรวจจับลิงก์ฟาร์มนั้นได้ผลที่ดี ดังนั้นการตรวจจับลิงก์ฟาร์มโดยใช้ไวยากรณ์กราฟที่นำเสนอใช้นั้นเมื่อใช้จำนวนการใช้โปรดักชันทั้งในเว็บกราฟระดับเว็บเพจและโฮสพิจารณาาร่วมกันสามารถนำมาใช้ในการตรวจจับลิงก์ฟาร์มได้เป็นอย่างดี

211041

There are many link farm detection techniques proposed in the literature review. These techniques only involve designing algorithms for detection with high precision without considering the structure of link farm. In our work, we introduce a new graph grammar model for expressing the structure of a link farm and a graph grammar for the link farm detection. Supervised graph grammar induction is modified to fit the training data with the number of applying production rules. Link farm detection algorithm is proposed and it uses logical rule to classify target hosts with 20 steps of detection. Compared with the related works, graph grammar in the experiments can effectively recognize link farms from web spam dataset. The comparison between the frequency of usage of some productions of spam and those of normal hosts indicates that graph grammar seem to be a good mechanism for detecting link farm.