

สุเชษ อมาตยกุล, พันเอก 2552: อัลกอริทึมการตรวจจับเว็บอนาจารด้วยลิงก์ฟาร์ม
ปริญญาวิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์) สาขาวิทยาการคอมพิวเตอร์
ภาควิชาวิทยาการคอมพิวเตอร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก: ผู้ช่วยศาสตราจารย์
สุพุมล กิตติสิน, Ph.D. 62 หน้า

การศึกษานี้มีวัตถุประสงค์เพื่อหาแนวทางลดภาระของผู้เชี่ยวชาญในการรวบรวม
รายชื่อเว็บไซต์อนาจาร โดยได้นำเสนออัลกอริทึมสำหรับตรวจจับเว็บไซต์อนาจารแบบ
กึ่งอัตโนมัติ เรียกว่า Link Farm Based Pornographic Web Detection Algorithm (LFPD) LFPD
ทำงานโดยใช้ผู้เชี่ยวชาญตรวจหาเว็บไซต์อนาจารเพื่อเป็นข้อมูลเริ่มต้น จากนั้นทำงานต่อไป
โดยการตรวจสอบโครงสร้างเส้นเชื่อมระหว่างเว็บไซต์ เพื่อหาเว็บไซต์ที่มีเส้นเชื่อมชี้ไปและกลับ
ร่วมกับกลุ่มของเว็บไซต์อนาจารที่ตรวจพบก่อนหน้านี้ ซึ่งสามารถแบ่งการทำงานของ LFPD ได้
เป็นสามขั้น ได้แก่ การกำหนดกลุ่มเว็บไซต์อนาจารเริ่มต้นโดยผู้เชี่ยวชาญ การรวบรวมข้อมูลเว็บ
และการตรวจจับเว็บไซต์อนาจาร

ผลการทดลองพบว่าจากเว็บไซต์อนาจารเริ่มต้นที่ระบุโดยผู้เชี่ยวชาญจำนวน 8 เว็บไซต์
เมื่อเราให้ LFPD ทำงานจำนวน 3 รอบ สามารถตรวจจับเว็บไซต์อนาจารได้ตั้งแต่ 18,725 - 44,123
เว็บไซต์ ขึ้นอยู่กับค่า T ซึ่งอยู่ในช่วงตั้งแต่ 1 ถึง 4 โดยเมื่อ T มีค่ามากขึ้นจำนวนเว็บไซต์ที่
ตรวจจับได้ก็จะลดลงตามลำดับ เมื่อพิจารณาขีดความสามารถของ LFPD พบว่าในรอบที่หนึ่งและ
สองมีผลการตรวจจับที่ความแม่นยำสูงกว่าร้อยละ 90 ในทุกๆ ค่าของ T อย่างไรก็ตามจำนวน
เว็บไซต์อนาจารสูงสุดที่ตรวจจับได้ด้วยความแม่นยำสูงกว่าร้อยละ 90 เป็นผลการทำงานในรอบที่
สามเมื่อใช้ T 4 โดยตรวจจับเว็บไซต์อนาจารได้จำนวน 18,725 เว็บไซต์ นอกจากนี้เรายังพบว่า
เว็บไซต์ทั่วไปที่มีใช้เว็บไซต์อนาจารแต่มีความสัมพันธ์ระหว่างกัน โดยมีการสร้างเส้นเชื่อมชี้ไป
และกลับร่วมกับเว็บไซต์อนาจารโดยเจตนา

Suez Amatyakul, Colonel 2009: Link Farm Based Pornographic Web Detection Algorithm. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Assistant Professor Sukumal Kitisin, Ph.D. 62 pages.

The purpose of this research is to find a better solution and more effective way to detect and identify pornographic websites. This research proposes a semi-automatic technique called Link Farm Based Pornographic Web Detection Algorithm (LEPD) algorithm as a mean for examining and identifying pornographic websites. The algorithm will assist experts to search for a set of pornographic websites based on an initial seed set. The LEPD will then examine the structures of the websites in the seed set to find any links that will lead to connections with other pornographic websites previously identified. LEPD works in three steps; first, experts identify a group of pornographic websites as a seed set; second, the algorithm automatically collects pornographic websites' information and structures; third, the algorithm examines and identifies whether the collected URLs are pornographic websites.

The result from this research founded that from the 8 websites initially identified by expertise. When we run LEPD 3 times, we were able to detect pornographic websites from 18,725 to 44,123 websites. The number of websites detected is depended on value of T which is in the range between 1 to 4. As the value of T increases the number of websites detected decreases. When we consider the capability of LEPD it founded that in the first and second run the accuracy rate of detection was higher than 90 percent in every mean of T. However, the highest number (18,725) of pornographic websites that LFPD was able to detect with accuracy rate of 90 percent or higher was from the third run where T equal to 4. Furthermore, we founded that there are other websites that are not pornographic websites which intentionally link back and forth with pornographic websites; for example, commercial and match-making services.