

ชื่อเรื่อง การค้นคว้าแบบอิสระ เว็บไซต์ไปเคอร์แบบจำเพาะกลุ่มเป้าหมายโดยอาศัยหลัก
ทางเอสวีเอ็ม

ผู้เขียน นายเจษฎา เทโวชาติ

ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

อาจารย์ที่ปรึกษาการค้นคว้าแบบอิสระ ผู้ช่วยศาสตราจารย์ ดร. รัฐสิทธิ์ สุขะหุต

บทคัดย่อ

ปัจจุบันเว็บไซต์ไว้ค์เว็บ ได้รับความนิยมในการใช้งานเป็นอย่างมาก และปริมาณเนื้อหาข้อมูล ที่เพิ่มขึ้นอย่างรวดเร็วบนเว็บไซต์ต่างๆ ทำให้เกิดปัญหาในการจัดการกับ เว็บไซต์ที่เพิ่มขึ้นซึ่งรวม ไปถึงการคัดกรองเว็บไซต์ให้เป็นหมวดหมู่ ซึ่งจะช่วยให้สามารถสืบค้นเว็บไซต์ตามหมวดหมู่อย่าง มีประสิทธิภาพ

งานค้นคว้าอิสระฉบับนี้จึงได้นำเสนอ เว็บไซต์ไปเคอร์แบบจำเพาะกลุ่มเป้าหมายโดย อาศัย หลักทางของซัพพอร์ตเวกเตอร์แมชชีน ซึ่งเว็บไซต์ที่ใช้ในการค้นคว้าแบบอิสระ ได้มาจากการครอว์ เลอร์ของเว็บไซต์ไปเคอร์ โดยครอว์เลอร์โดเมน เว็บไซต์ไว้ค์เว็บคอตซีเอ็นเอ็นคอตคอม เป็นเอกสารเว็บ ที่เกี่ยวกับข่าว โดยนำเอกสาร สร้างแบบจำลองจำนวน 1,000 เอกสาร และเอกสารในการทดสอบ จำนวน 500 เอกสาร เอกสารกลุ่มข่าวที่นำมาจัดกลุ่มจำนวน 5 กลุ่มคือ 1.ข่าวธุรกิจ 2.ข่าวสุขภาพ 3.ข่าวกีฬา 4.ข่าวท่องเที่ยว 5.ข่าวคอมพิวเตอร์ การจัดกลุ่มเอกสารในงานค้นคว้าแบบอิสระนี้ อาศัย หลักการของซัพพอร์ตเวกเตอร์แมชชีนและ ได้มีการทดสอบประสิทธิภาพของการจัดกลุ่มเอกสาร ผลการทดสอบสามารถวัดประสิทธิภาพการจัดกลุ่มเอกสารของระบบคิดเป็นร้อยละ

72.46% จากการวัดค่าเอฟ ซึ่งจำแนกได้ 3 กลุ่มคือ 1.กลุ่มวัดประสิทธิภาพการจัดกลุ่มเอกสารที่ได้ ดีที่สุดคือกลุ่มข่าวกีฬา 2.กลุ่มวัดประสิทธิภาพการจัดกลุ่มเอกสารที่ได้ปานกลางคือกลุ่มข่าว สุขภาพและกลุ่มข่าวท่องเที่ยว 3.กลุ่มวัดประสิทธิภาพการจัดกลุ่มที่ได้ต่ำคือกลุ่มข่าวธุรกิจและกลุ่ม ข่าวคอมพิวเตอร์

Independent Study Title	Web Spider for Specific Domain Based on SVM
Author	Mr. Jedsada Tawokat
Degree	Master of Science (Computer Science)
Independent Study Advisor	Assistant Professor Dr. Rattasit Sukhahuta

ABSTRACT

Currently World Wide Web (WWW) has gained its popularity and the increment of data posted on websites has created not only web managing problem but also web classifying problem. With efficient classifying process, it will help enhance the search process on website.

This independent study propose Web spider for specific domain by using support vector machines. The reference website for this independent study came from crawler of Web spider with crawler the domain www.cnn.com is a website about news. There were 1,000 documents for create model and 500 documents are used in the test. There were 5 fields of news: 1.Business news 2.Health news 3.Sports news 4.Travel news and 5.Computer news. The document classification of this independent study applied support vector machine principle and the effectiveness of the classification document has been measured.

The test results by measuring the effectiveness of the classification process of systems with the experimental results are 72.46 percent after testing by F-measure, which identified three groups: 1.A group of classification that can best document is a group of sports news 2.Groups of classification that document the moderate are health news group and travel news group 3.The worst group of classification document are business news group and computer news group.