

193087

การวิจัยนี้มีวัตถุประสงค์เพื่อทำการจัดประเภทของเว็บไซต์โดยการเชื่อมโยงของเว็บในระดับเพจและบล็อกให้เป็นหมวดหมู่เพื่อช่วยด้านการสืบค้นข้อมูล โดยการนำเสนอการประยุกต์ใช้คุณลักษณะระดับเพจและบล็อกมาช่วยในการจัดประเภทของเว็บไซต์

ผู้วิจัยได้ใช้ประเภทของเว็บไซต์จำนวน 11 ประเภท จำนวน 265 เว็บไซต์ รวมทั้งสิ้น 2,650 เพจ โดยใช้อัลกอริทึม VIPS ในการแบ่งเพจเป็นบล็อก แล้วคำนวณค่าคุณลักษณะระดับเพจและบล็อก จากนั้นได้ใช้วิธีการเรียนรู้ของเครื่องแบบ See5 แสดงให้เห็นค่าความถูกต้องในการจัดประเภทของเว็บไซต์

ผลการทดลองที่ได้จากการวิจัยในครั้งนี้ แสดงค่าความผิดพลาดที่เกิดจากการจัดประเภทของเว็บไซต์ที่พิจารณาจากคุณลักษณะระดับเพจ คุณลักษณะระดับบล็อก คุณลักษณะระดับเพจและบล็อก ซึ่งให้ค่าผิดพลาดที่เรียงจากน้อยที่สุดไปหามากที่สุดคือ การจัดประเภทของเว็บไซต์โดยใช้คุณลักษณะระดับเพจเป็น 62.5 เปอร์เซ็นต์ การจัดประเภทของเว็บไซต์โดยใช้คุณลักษณะระดับเพจและระดับบล็อกเป็น 67.8 เปอร์เซ็นต์ และการจัดประเภทของเว็บไซต์โดยใช้คุณลักษณะระดับบล็อกเป็น 75.9 เปอร์เซ็นต์ ทั้งนี้แนวทางของการทดลองในงานวิจัยนี้อาจสามารถนำไปปรับปรุงให้ดีขึ้นโดยการให้ความสำคัญในการเลือกหน่วยตัวอย่าง และเพิ่มจำนวนเพจที่อยู่ในเว็บไซต์เพื่อนำมาซึ่งผลการทดลองที่ดีขึ้นต่อไป

193087

The purpose of this research is to classify web sites by connectivity at page and block level which helps information searching. This research applies features at page and block level for web classification

In carrying out this research, researcher uses 11 types of web sites, 265 web sites and 2,650 pages. VIPS algorithm is also used to divide page to block . Then page features and block features are computed. After that, See5 is used to classify and evaluate the accuracy of web site classification.

The result of this experiment has shown error of web site classification which considers by features in page level, block level and page&block level. The values of percentage error for using only page features, page and block features and only block features are 62.6%, 67.8% and 75.9% respectively. The experiment from this research can be improved by emphasizing on the selection of the sample units of web site and thoroughly considering to increase more pages in each web site to give more classification accuracy.