

การจำแนกหัวเรื่องด้วยวิธีการแบ่งเว็บเพจเป็นส่วนย่อย
และซอฟต์แวร์เว็บเบราว์เซอร์แมชชีนเพื่อปรับปรุงโปรแกรมอ่านหน้าจอ

HClassifier: Using Web Page Segmentation and SVM based

Web Heading Extraction for Improving Screen Readers

คณาฤดี วงศ์เกษม , รัชฎา คงคะจันทร์

บทคัดย่อ

หัวเรื่องซึ่งถูกกำหนดรูปแบบการแสดงผลด้วยป้าย ายระบุหัวเรื่องช่วยให้ผู้พิการทางสายตา เข้าถึง ข้อมูลของเว็บเพจผ่านทางโปรแกรมอ่านหน้าจอได้รวดเร็วขึ้น แต่เว็บไซต์จำนวนมากใช้รูปแบบการ แสดงผลในการสร้างหัวเรื่อง เช่น กำหนดขนาดตัวอักษรให้มีขนาดใหญ่ แทนการใช้ป้ายระบุหัวเรื่องทำให้ โปรแกรมอ่านหน้าจอไม่สามารถระบุได้ว่าส่วนใดเป็นหัวเรื่อง ดังนั้นวิทยานิพนธ์นี้ได้นำเสนอการจำแนกหัว เรื่องของเว็บเพจจากรูปแบบการแสดงผลด้วยวิธีการแบ่งเว็บเพจเป็นส่วนย่อยควบคู่กับการใช้ซอฟต์แวร์ เว็บเบราว์เซอร์แมชชีน โดยทำการทดลองกับเว็บเพจชนิดต่างๆ จำนวน 100 เว็บเพจ ผลการทดลองพบว่าวิธีการ ที่นำเสนอให้ผลการจำแนกหัวเรื่องด้วยค่าความแม่นยำร้อยละ 86 ค่าระลึกร้อยละ 81 และ ค่า f-measure ร้อยละ 83

Abstract

Headings being properly marked up with heading tags can help visually impaired users to rapidly get information of web pages through screen readers. Unfortunately, heading of web pages in the real world are created by using visual styles i.e. font size and font color instead of heading tags which screen readers cannot identify them correctly. To address this problem, we propose web heading classification that is based on machine learning over visual styles by using web page segmentation and support vector machine. To evaluate the proposed, we tested our approach with 100 web pages. The results show that our method yields 86%, 81% and 83% of precision, recall and fmeasure, respectively.