วิทยานิพนธ์นี้ได้นำเสนอการจัดกลุ่มเอกสารภาษาไทยสำหรับผลการสืบค้นเว็บ โดยในการจัดกลุ่มได้ใช้ขั้นตอนวิธีนอนเน็กกาทีฟเมทริกซ์แฟกทอไรเซชั่น และขั้นตอนวิธีออนไลน์ สเฟียริเคิลเค-มีนส์ ในการเพิ่มประสิทธิภาพความถูกต้อง การทดลองได้ใช้ข้อมูลคำอธิบายเว็บภาษาไทย จากสารบบเว็บหรรษา สารบบเว็บน่ารัก และสารบบเว็บสนุก จำนวน 6 ชุดเอกสาร 2 รูปแบบ จากหมวดสารบบเว็บและผลจากการสืบค้นคำสำคัญจากหมวดสารบบเว็บ ผ่านกระบวนการเตรียมเอกสาร ตัดคำโดยใช้พจนานุกรมด้วยวิธีการเทียบคำที่ยาวที่สุดและแทนเอกสารการเวคเตอร์โมเดล ผลการทดลองแสดงให้เห็นว่าผลการจัดกลุ่มด้วยขั้นตอนวิธีนอนเน็กกาทีฟเมทริกซ์แฟกทอไรเซชั่นให้ผลในระดับดี และผลการจัดกลุ่มขั้นตอนวิธีออนไลน์สเฟียริเคิลเค-มีนส์ มีความถูกต้องในโดยเฉลี่ยรวมเพิ่มขึ้นจากขั้นตอนวิธีนอนเน็กกาทีฟเมทริกซ์แฟกทอไรเซชั่น จากการทดลองยังพบว่าการใช้ภาษาไทยมีความยุ่งยากในการประมวลผลเอกสารข้อความคำอธิบายเว็บภาษาไทยจำนวนมากใช้คำทับศัพท์ภาษาต่างประเทศ คำชื่อเฉพาะ ใช้คำที่เขียนไม่ถูกต้อง ซึ่งมีผลต่อการจัดกลุ่มเอกสารภาษาไทยสำหรับผลการสืบค้นเว็บ

ABSTRACT

218303

This research aimed to investigate Clustering for web search results in Thai documents using Non-negative Matrix Factorization(NMF) and Online Spherical K-Means Algorithm(OSKM) for increase of accuracy in performance of clustering, using 2 types of 6 Thai web-snippet documents from Hunsa, Sanook, Narak webdirectories. The method for treating data was by single word searching. Data preprocessing was done by using the dictionary approach for word segmenting and the longest string matching. All of the segmented words of each document were transformed into Vector Space Model. The experimental results showed that the average performance NMF is at a high level and the performance of OSKM is better than that of NMF. The experiment revealed difficulty in Thai document processing due to some words being transliterated from a foreign languages and some words derived from common nouns and used incorrectly.