

การจัดกลุ่มผลการสืบค้นด้วยเทคนิค Suffix Tree Clustering (STC) เป็นเทคนิคการจัดกลุ่มผลการสืบค้นที่ทำงานเร็วและมีความต่อเนื่อง (Fast Incremental Clustering) ซึ่งเหมาะสำหรับการจัดกลุ่มผลการสืบค้น ที่ต้องการเปลี่ยนแปลงของผลการจัดกลุ่มตามผลการสืบค้นที่มีการเปลี่ยนแปลงตลอดเวลา (Dynamic or on-the-fly Clustering) แต่อย่างไรก็ตามเทคนิค Suffix Tree Clustering (STC) ก็ยังไม่เพียงพอสำหรับผลการจัดกลุ่มและป้ายชื่อกลุ่ม (Cluster Label) ที่ถูกตัดขาดจากการทำงานร่วมกับเทคนิค n-gram รายงานฉบับนี้ เราได้นำเสนอแนวทางการจัดกลุ่มผลการสืบค้นด้วย Suffix Tree Clustering (STC) แนวใหม่ ด้วยการรวมกลุ่มพื้นฐานจากการเชื่อมป้ายชื่อกลุ่มเอกสารจากการพิจารณาการมีอยู่ร่วมกันของคำที่ปรากฏในป้ายชื่อกลุ่มของคู่กลุ่มเอกสาร และพิจารณาความคล้ายคลึง (similarity) ของกลุ่มเอกสารเพื่อเชื่อมกลุ่มจากเงื่อนไข “ สมาชิกของกลุ่มย่อยมากกว่าหนึ่งเอกสารเป็นสมาชิกของกลุ่มใหญ่ในลักษณะที่ตำแหน่งของคำเชื่อมต่อกัน ” ซึ่งเป็นเทคนิคที่จะช่วยเพิ่มจำนวนเอกสารที่ตรงกับความต้องการ (Relevance Web Document) และลดจำนวนเอกสารที่ทับซ้อนกัน (Overlap Web Document) ของกลุ่มเอกสารได้มากกว่าการทำงานในรูปแบบเดิมของการจัดกลุ่มผลการสืบค้นด้วยเทคนิค Suffix Tree Clustering (STC)

Suffix Tree Clustering (STC) is incremental and linear time (in the document collection size) algorithm for dynamic or on-the-fly web search results clustering. However, STC is inadequate since they generate interrupted cluster label due to using n-gram technique. In this paper, we propose a new Suffix Tree Clustering algorithm that combining base cluster by cluster label combination to become new cluster and cluster label, consider the combined cluster label from word which joint appear within the cluster label of pair cluster is subset of member document within cluster and consider to adjoin of word location in document using word together. It's provides more relevance and less overlap of web document clusters than conventional Suffix Tree Clustering algorithms.