

Abstraksi

Clustering dokumen merupakan cara yang tepat untuk mempermudah search engine dalam melakukan query terhadap kumpulan dokumen yang besar. Dokumen-dokumen yang memiliki kesamaan akan dikelompokkan sehingga membentuk topik-topik atau subtopik yang berbeda. Algoritma-algoritma *clustering* dokumen yang sering dipelajari adalah algoritma-algoritma *batch clustering*, di mana keseluruhan dokumen diperlukan sejak awal dan *clustering* dilakukan dengan banyak iterasi terhadap dokumen tersebut. Namun, dengan adanya publikasi *online* di web yang semakin berkembang seperti sekarang ini, terjadi ledakan jumlah informasi yang bertambah setiap harinya. Metode *batch clustering* dianggap tidak efisien untuk kasus semacam ini. Agar proses *clustering* dapat dilakukan segera setelah dokumen masuk, maka *clustering* perlu dilakukan secara *incremental*.

Terdapat beberapa algoritma *incremental clustering* yang populer. Salah satunya adalah algoritma Cobweb yang diimplementasikan pada Tugas Akhir ini. Tugas Akhir ini menggunakan Cobweb untuk mengelompokkan *retrieved documents* hasil pencarian *search engine*. Cobweb melakukan *clustering* data dengan membangun *classification tree* di mana tiap node dari *tree* tersebut menggambarkan cluster yang berisi objek-objek data. Dalam membangun *tree*, Cobweb menggunakan *category utility* (CU) untuk mengevaluasi *tree* dan mendapatkan pengelompokan data yang paling tepat. Dari pengujian yang dilakukan pada Tugas Akhir ini, hasil akhir menunjukkan bahwa *clustering* menggunakan algoritma Cobweb yang diterapkan pada *retrieved documents* memberikan solusi dengan kualitas yang baik, karena meskipun pada pohon klaster pasti terjadi *overlapping*, tetap terbukti memiliki sifat kohesif. Kohesif adalah keadaan di mana persamaan antardokumen dalam klaster yang sama lebih besar daripada persamaan antardokumen pada klaster yang berbeda.

Kata kunci : dokumen, *incremental clustering*, Cobweb, *search engine*, *retrieved documents*, *classification tree*, *category utility*, persamaan, kohesif