

Abstrak

Perkembangan dokumen teks sangat cepat di internet, perpustakaan digital, dan artikel-artikel. Informasi di internet sangat bermanfaat bagi pengguna, khususnya artikel berita dalam bentuk dokumen teks. Begitu banyak artikel berita yang ada di internet sehingga sangat menyulitkan pengguna internet untuk mendapatkan artikel berita yang diinginkan. Untuk itu diperlukan kategorisasi artikel berita berdasarkan informasi yang terkandung di dalamnya. Sehingga artikel berita tersebut bisa di kategorikan pada topik tertentu.

Klasterisasi dokumen/artikel merupakan salah satu metode yang dilakukan untuk menggali informasi yang terkandung dalam dokumen/artikel tersebut. Klasterisasi dilakukan untuk membuat klaster agar dokumen yang berhubungan atau dokumen yang informasinya mirip satu sama lain bisa berada dalam klaster yang sama. Sebuah dokumen mempunyai sifat dimensi tinggi dan volume data yang besar. Untuk itu diperlukan metode yang bisa menangani dimensi yang tinggi dan volume data yang besar.

Pada umumnya algoritma klasterisasi hanya fokus pada klasterisasi satu arah, misalnya membuat klaster dokumen berdasarkan distribusi kata, atau membuat klaster kata/word berdasarkan distribusi dokumen. Oleh sebab itu, dikembangkan metode *Co-Clustering* yang membuat klaster secara simultan pada kedua dimensi tabel. Hal ini dilakukan untuk mengurangi dimensi secara efektif dan efisien.

Kata kunci: *co-clustering, information theory, mutual information*