

INTISARI

APLIKASI PENGELOMPOKAN HASIL PENCARIAN DOKUMEN BERITA MENGGUNAKAN ALGORITMA *BISECTING K-MEANS*

ERRIC ALFAJRI

11/316655/PA/13790

Penelusuran dokumen berita merupakan aktivitas umum yang dilakukan oleh pengguna internet. Dalam mendapatkan informasi, pengguna dapat menggunakan *search engine* yang dapat mengembalikan hasil pencarian berdasarkan sebuah kata kunci. Pencarian informasi dapat menjadi sebuah permasalahan karena terdapat beberapa subtopik dari kata kunci dicampur bersama-sama dalam satu daftar panjang. Fitur *clustering* dapat diimplementasikan terhadap hasil pencarian untuk mempersempit pilihan pencarian melalui teknik penyajian per-*cluster*.

Pada penelitian ini, telah dibuat aplikasi *search engine* dokumen berita. Aplikasi dilengkapi dengan fitur pengelompokan hasil pencarian. Terdapat beberapa pilihan sumber situs berita yang dijadikan sumber pencarian dokumen. Google *Custom Search* dimanfaatkan untuk mendapatkan dokumen sesuai kriteria pencarian, setelah itu diimplementasikan *clustering* terhadap hasil pencarian menggunakan algoritma *bisecting k-means* dengan *library* yang disediakan oleh *carrot2*.

Hasil dari penelitian ini adalah sebuah aplikasi *search engine* dokumen berita secara *online* yang dapat menyajikan hasil pencarian berdasarkan kelompok-kelompok. Setiap kelompok diwakili oleh label kelompok. Evaluasi dilakukan terhadap hasil *clustering*, didapatkan *precision* 0,57 dan *cluster label quality* 0,78.

Kata kunci: *clustering*, *bisecting k-means*, *carrot2*, Google *Custom Search*.

ABSTRACT

APPLICATION OF NEWS DOCUMENT SEARCH RESULTS CLUSTERING USING BISECTING K-MEANS ALGOTIRHM

ERRIC ALFAJRI

11/316655/PA/13790

Search for news document is common activity of internet users. To reach the information, user can use internet search engines, which can return search results depends on the keyword. Finding the information needed on the search results can be a problem, because several of subtopic are mix together in a long list of search results. Clustering feature can be implemented in search engine to present the search results in several clusters.

In this research, a search engine application of news document is build. A search result clustering feature is added to the search engines. There are news sites which are used as sources for searching news documents. Google Custom Search is used for searching documents, then clustering is implemented on the search results using bisecting k-means algorithm with a library supplied by carrot2.

This research produces a search engine application of news documents which can present search results based on clusters. Each cluster is represented by a label cluster. The evaluation was done on the clustering results, found that the precision is 0.57 and the cluster label quality is 0,78.

Keywords: clustering, bisecting k-means, carrot2, Google Custom Search.