

ABSTRAKSI

Kebanyakan mesin pencari yang ada saat ini pada umumnya menampilkan dokumen hasil pencarian sesuai dengan urutan dokumen (*document ranking*) tanpa mengelompokkan atau mengkategorikan dokumen sesuai dengan kemiripan dokumen. Dengan jumlah dokumen yang cukup besar akan memberikan dampak negatif bagi pengguna, yaitu dibutuhkan waktu yang relatif lama untuk memilah-milah dokumen yang sesuai dengan kebutuhan pengguna. Untuk mempermudah pengguna dalam mencari informasi pada kumpulan dokumen yang cukup besar, salah satu solusinya adalah dengan cara mengelompokkan dokumen hasil pencarian sesuai dengan *keyword* yang diinputkan oleh pengguna. Dengan adanya pengelompokan dokumen hasil pencarian ini, maka pengguna tidak perlu membuka halaman terlalu banyak karena dokumen hasil pencarian telah dikelompokkan berdasarkan kemiripan dokumen-dokumen tersebut.

Salah satu algoritma *partitional* yang dapat mengelompokkan dokumen yang belum berlabel adalah *Expectation-Maximization*, yaitu algoritma yang berfungsi untuk menemukan nilai estimasi *Maximum Likelihood* dari parameter dalam sebuah model probabilistik [2]. Ciri-ciri dari algoritma ini adalah dapat mengelompokkan dokumen yang belum berlabel atau *unlabeled data* dan juga hasil pengelompokannya akan selalu *convergence*. Dari hasil percobaan didapatkan kesimpulan bahwa algoritma EM dapat mengelompokkan dokumen hasil pencarian, hal ini dapat membantu pengguna untuk mencari dokumen yang diharapkan. Akurasi tertinggi mencapai 70% dan terendah 32.58%. Penambahan algoritma *stemming* Arifin Setiono mampu meningkatkan performansi algoritma EM hingga 10%.

Kata Kunci: *Clustering, Expectation-Maximization, Unlabeled, Stemming.*