

ABSTRAK

Kebutuhan manusia terhadap teknologi serta ketersediaan sarana prasarana yang memadai menjadi bukti bahwa saat ini teknologi menjadi bagian dari kebutuhan pokok manusia. Semakin bertambah banyaknya jurnal dan karya ilmiah, maka harus semakin selektif dalam memilih dan memilah meskipun sudah banyak penyedia jasa online dan portal jurnal. Penelitian tentang search engine dan plagiarism dan sistem rekomendasi sudah banyak dilakukan dengan berbagai metode yang dirasa tepat untuk meningkatkan performa sistem itu sendiri, makalah ini memiliki tujuan menghitung kemiripan antara satu artikel dengan artikel yang lain dengan mengimplementasikan n-gram dan tanimoto cosine. Jumlah artikel yang diuji sebanyak empat puluh tiga judul dan abstrak, ujicoba sebanyak lima puluh kali dengan kata kunci yang dipilih secara acak, dengan memecah setiap kalimat judul dan abstrak kedalam n karakter ($n=2$) termasuk spasi dan tanda baca, kemudian dihitung kemiripan dengan query atau kata kunci yang dipakai untuk ujicoba sistem. Pengujian dilakukan menggunakan beberapa variasi threshold. Setelah dilakukan pengamatan dari lima puluh kali pengujian, threshold 0,30 memiliki nilai akurasi = 0,86, presisi = 0,37 dan recall = 0,44.

Kata kunci: sistem rekomendasi, Tanimoto cosine, similarity, akurasi, presisi dan recall.

Abstract

Human needs of technology and the availability of adequate infrastructure are evidences that the technology currently becomes part of the human beings' basic necessities. Growing multitude of journals and scientific papers makes choosing and sorting become more selective though there have been many online journals service providers and portals. Research on search engines, plagiarism and recommendation system has been carried out with various methods to improve the performance of the system itself, this paper aims to calculate similarities between one article with other articles by implementing n-gram and tanimoto cosine. The number of articles tested were forty-three titles and abstracts, tests were carried fifty times with random selected keywords, by separating each sentences of the title and abstract into n characters ($n = 2$) including spaces and punctuation, then calculating similarity to the query or keywords used to test the system. Testing was done using several variation of the thresholds. After observing the fifty times-testings, the threshold value of 0.30, produced accuracy = 0.86, precision = 0.37 and recall = 0.44.

Keywords: *recommendation system, Tanimoto cosine, similarity, accuracy, precision and recall*