

PERBANDINGAN ALGORITMA *COSINE* DAN *DICE SIMILARITY* DALAM MENGHITUNG KEMIRIPAN DOKUMEN

M Dzul Romaini AL
16.04.111.00.175

Pembimbing 1 : Achmad Jauhari, S.T.,M.Kom. 19810109 200604 1 003
Pembimbing 2 : Sigit Susanto Putro, S.Kom., M.Kom. 19790313 200604 1 002

ABSTRAK

Sebuah Informasi adalah kebutuhan yang berguna untuk segala pihak sehingga berguna untuk kepentingan didunia akademis dan memberikan dampak yang besar, salah satunya yaitu internet. Situs pencarian seperti *Google*, *Bing*, dan *Yahoo* merupakan suatu tempat untuk mendapatkan informasi yang dibutuhkan tanpa bersusah payah. Dimulai dari berita terkini, penelitian, jurnal atau *paper* bisa didapatkan di era yang memiliki teknologi maju seperti sekarang, Namun tidak semua dampak baik bisa didapatkan di era yang penuh pengaruh teknologi informasi saat ini. Banyak oknum-oknum melakukan perilaku *negative* dari pemanfaatan informasi yang didapat seperti halnya melakukan kecurangan dengan cara mencuri penelitian orang lain atau bisa disebut *Plagiarism*. Peneliti melakukan sebuah penelitian tentang bagaimana mencari kemiripan berdasarkan pencocokan melalui *query* sehingga lebih mudah di temukan ketika sebuah dokumen melakukan *plagiarism*, pencarian kemiripan (*Similarity*) dilakukan dengan menggunakan metode *Cosine Similarity* dan *Dice Similarity*. Peneliti mengusulkan sebuah sistem pendeteksi kemiripan dokumen dengan membandingkan 2 metode yaitu *Cosine Similarity* dan *Dice Similarity* yang digunakan untuk mencari kemiripan dalam teks Bahasa Indonesia. Pada penelitian kali ini yang mana 2 metode tersebut menghasilkan perbandingan yang sangat berbeda dalam momen tertentu, dimana *Dice* Memiliki banyak

keunggulan dengan menghasilkan nilai rata-rata sebesar 19,356 pada abstrak dengan 1 data yang diambil secara utuh dan rata-rata sebesar 28,256 dalam abstrak yang diambil 1 paragraf pertama dengan 3 data yang dipilih secara *random*, memiliki time respons yang cukupcepat meskipun dalam situasi tertentu metode *Cosine* dapat mengungguli.

Kata Kunci : Dokumen, Deteksi kemiripan, *Similarity*, *Cosine Similarity*, *Dice Similarity*

COSINE AND DICE SIMILARITY ALGORITHM COMPARISON IN CALCULATE DOCUMENTS OF DOCUMENTS

M Dzul Romaini AL
16.04.111.00.175

Supervisor 1 : Achmad Jauhari, S.T.,M.Kom. 19810109 200604 1 003
Supervisor 2 : Sigit Susanto Putro, S.Kom., M.Kom. 19790313 200604 1 002

ABSTRACT

An information is a need that is useful for all parties so that it is useful for the interests of the academic world and has a big impact, one of which is the internet. Search sites like Google, Bing, and Yahoo are a place to get the information you need effortlessly. Starting from the latest news, research, journals or papers can be obtained in an era that has advanced technology as it is now, but not all good impacts can be obtained in an era that is full of the influence of information technology today. Many individuals carry out negative behavior from using the information obtained, such as committing fraud by stealing other people's research or it can be called plagiarism. Researchers conducted a study on how to look for similarities based on matching through queries so that it is easier to find when a document is plagiarized, similarity searches are carried out using the Cosine Similarity and Dice Similarity methods. The researcher proposes a document similarity detection system by comparing 2 methods, namely Cosine Similarity and Dice Similarity which are used to find similarities in Indonesian texts. In this study, where the 2 methods produce very different comparisons at certain moments, where Dice has many advantages by producing an average value of 19.356 in the abstract with 1 data taken in its entirety and an average of 28.256 in the abstract taken. The first 1 paragraph with 3 randomly selected data, has a fairly fast response time even though in certain situations the

Cosine method can outperform.

Keywords: Document, Similarity Detection, Similarity, Cosine Similarity, Dice Similarity