

Penerapan Algoritma TF-IDF Untuk Pencarian Karya Ilmiah

ABDUL AZIS MA'ARIF

*Program Studi Teknik Informatika - S1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111201005666@mhs.dinus.ac.id

ABSTRAK

Pemilahan sebuah karya penulisan ilmiah dapat dilakukan dengan mudah oleh manusia, tetapi pemilahan dokumen dilakukan secara otomatis dengan komputer akan membawa permasalahan tersendiri. Begitu pula dengan mengukur tingkat kemiripan suatu dokumen dengan dokumennya lainnya, manusia dapat dengan mudah mengukur apakah suatu dokumen memiliki tingkat kemiripan/similaritas dengan dokumen lainnya. Kata kunci yang digunakan didapat dari proses ekstraksi dokumen pada proses pemilahan kategori dokumen. Agar hasil pengukuran tingkat similaritas dokumen dengan kata kunci mendapatkan hasil yang optimal maka digunakan algoritma text mining dimana dalam prosesnya digunakan algoritma TF-IDF (Term Frequency "Inversed Document Frequency) dari model IR (information retrieval) sebagai pengukur tingkat similaritas antara dokumen dengan keyword yang didapat dari ekstraksi teks pada dokumen. Tujuan penelitian ini adalah untuk menerapkan algoritma TF-IDF yang dapat digunakan untuk mencari karya ilmiah sebagai pengukur tingkat similaritas antara dokumen dengan keyword yang didapat dari ekstraksi teks pada dokumen sehingga mendapatkan data yang terurut dari yang kemiripannya (tingkat similaritas) paling tinggi sehingga pencarian karya ilmiah menjadi lebih efisien sebagai informasi yang relevan

Kata Kunci : TF-IDF, Karya Ilmiah, Kata Kunci

Application of TF-IDF algorithm to search Scientific Work

ABDUL AZIS MA'ARIF

*Program Studi Teknik Informatika - S1, Fakultas Ilmu
Komputer, Universitas Dian Nuswantoro Semarang*

URL : <http://dinus.ac.id/>

Email : 111201005666@mhs.dinus.ac.id

ABSTRACT

Sorting a scientific paper can be done easily by humans, but sorting of documents is done automatically by the computer will bring its own problems. Similarly, by measuring the level of similarity of a document with other documents, humans can easily measure whether a document has the level of similarity/similaritas with other documents. Keyword that is used in the extraction process document in the process of sorting the categories document. In order for the results of the measurement of level of similaritas documents with keywords to get optimal results then used algorithms for text mining algorithm used in the process whereby the TF-IDF (Term Frequency " Inversed Document Frequency) of the model IR (information retrieval) as a measure of the level of similaritas between documents with keywords obtained from the extraction of the text in the document. The purpose of this study was to apply the algorithm TF-IDF that can be used to find the scientific papers as the measuring level similaritas between documents with keywords obtained from the extraction of the text in the document so that it gets sorted data from that similarity (level similaritas) most high so the search papers become more efficient as the relevant information

Keyword : Tf-IDF, Scientific Papers, Keyword