

ABSTRAK

IMPLEMENTASI COSINE SIMILARITY UNTUK MENENTUKAN GENRE BUKU BERKONTEN ISLAMI DI PT MIZAN MEDIA UTAMA

Oleh

Moh Iqbal
1211705110

Email : Moh.iqbal@student.uinsgd.ac.id

Banyaknya dokumen judul buku yang terkumpul dalam bentuk *soft file* yang tidak terklasifikasi dengan baik mengakibatkan proses pencarian kembali menjadi sulit. Untuk mengakses informasi yang dibutuhkan menjadi kurang cepat dan tepat apabila keseluruhan dokumen disimpan dalam satu *folder database*. Maka dari itu diperlukan suatu sistem yang dapat mengklasifikasikan dokumen secara otomatis ke dalam *folder* berbeda pada *database* agar lebih mudah dalam mengelola dokumen yang ada. Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode *cosine similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah *vector* dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran. Metode pengembangan sistem yang digunakan dalam penelitian ini adalah model *prototype*, Data latih yang digunakan dalam penelitian ini berjumlah 50 dokumen skripsi dengan beberapa kategori yang berbeda. Hasil penelitian menunjukkan bahwa persentase tingkat ketepatan klasifikasi sistem adalah sebesar 75,92%.

Kata kunci : klasifikasi dokumen, cosine similarity, TF-IDF, vector space



uin

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
B A N D U N G

ABSTRACT

THE IMPLEMENTATION OF AN ALGORITHM BOYER-MOORE ON THE APPLICATION ENCYCLOPEDIA TO SUFISM THE ANDROID BASED

By

Moh Iqbal
1211705110

Email : Moh.iqbal@student.uinsgd.ac.id

The number of document titles collected in soft files that are not well classified means that the search process is difficult again. To access the required information becomes less fast and precise if the entire document is stored in a database folder. Therefore we need a system that can classify documents automatically into different folders on the database so that it is easier to manage existing documents. The TF-IDF method is a way to give weight to the relationship of a word (term) to a document. The cosine similarity method is a method for calculating the similarity between two objects which are expressed in two vectors using keywords (keywords) from a document as a measure. The system development method used in this study is a prototype model, the training data used in this study amounted to 50 thesis documents with several different categories. The results showed that the percentage level of accuracy of the system classification was 75.92%.

Keywords: document classification, cosine similarity, TF-IDF, vector space

The logo of Universitas Islam Negeri Sunan Gunung Djati Bandung, featuring the lowercase letters 'uin' in a stylized, rounded font.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG