

SISTEM Pencarian Teks Al-Qur'an Terjemahan Berbahasa Indonesia dengan Metode *Dice Similarity* dan Nazief dan Adriani

Nadila Hidayanti
(160411100182)

Pembimbing I : Achmad Jauhari, S.T., M.Kom

Pembimbing II : Ika Oktavia Suzanti, S.Kom., M.Cs

ABSTRAK

Information Retrieval (IR) merupakan pengambilan informasi yang tersimpan didalam dokumen yang berisikan konteks dari dokumen tersebut. Adanya *IR* memiliki pengaruh yang pesat dalam kemajuan teknologi seperti perpustakaan digital, buku-buku yang diubah menjadi digital. Salah satunya aplikasi Al-qur'an dan tafsir dalam berbagai bahasa yang bertujuan untuk memudahkan publik untuk memahami dari setiap ayat dalam bahasa Indonesia. Akan tetapi, masih terdapat kendala dalam melakukan pencarian dengan topik yang diinginkan misalnya hasil pencarian yang muncul kurang sesuai dengan yang diinginkan. Penelitian ini menggunakan metode dengan tahapan *pre-processing*, *stemming* Nazief dan Adriani, pembobotan *TF-IDF* dan perhitungan *Dice Similarity*. Proses perhitungan tanpa *stemming* dengan metode *dice similarity* didapatkan nilai *recall* 77,20%, *precision* 76,77% dan *f-measure* 68,28%, sedangkan pada proses menggunakan *stemming* nazief dan adriani didapat nilai *recall* 89,76%, *precision* 78,37%, dan *f-measure* 73,35%. Dari hasil pengujian penggunaan *stemming* sangat berpengaruh dalam mendapatkan hasil yang relevan.

Kata Kunci : *Information Retrieval*, tafsir Al-qur'an, Nazief dan Adriani, *TF-IDF*, *Dice Similarity*.

KATA PENGANTAR

Bismillahirrahmanirrahim, Alhamdulillah segala puji syukur kehadiran Allah SWT atas berkah, rahmat dan hidayah-Nya yang senantiasa dilimpahkan kepada penulis sehingga bisa menyelesaikan skripsi dengan judul **“Sistem Pencarian Teks Al-qur’an Terjemahan Berbahasa Indonesia Dengan Metode Dice Similarity dan Nazief dan Adriani”** sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Teknik dan Jurusan Teknik Informatika. Dengan terselesainya laporan skripsi ini, penulis tak lupa menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah memudahkan dalam menyelesaikan kuliah hingga skripsi selesai.
2. Kedua orang tua yang telah memberikan semangat, kasih sayang dan doa yang tiada henti. Semoga kedua orang tua selalu dalam lindungan Allah SWT.
3. Achmad Jauhari, S.T., M.kom selaku dosen pembimbing I dan Ika Oktavia Suzanti, S.Kom., M.Cs selaku dosen pembimbing II. Terima kasih telah banyak meluangkan waktu untuk membimbing dan memberikan arahan kepada penulis selama penelitian skripsi.
4. Bapak dan Ibu Dosen Fakultas Teknik, khususnya di Program Studi S1 Teknik Informatika Universitas Trunojoyo Madura yang telah membagikan ilmunya kepada penulis selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini tidak luput dari kesalahan karena keterbatasan pengalaman. Oleh karena itu penulis menyatakan permohonan maaf atas segala kekurangan yang ada. Semoga skripsi ini dapat bermanfaat bagi paca pembaca.

Bangkalan, 13 Juli 2020

Penulis

DAFTAR ISI

SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN ORISINALITAS	iv
HALAMAN PERSEMBAHAN	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR KODE PROGRAM	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.2.1 Permasalahan (<i>Research Problem</i>)	2
1.2.2 Metode Usulan (<i>Proposed Method</i>)	2
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penelitian	3
BAB II KAJIAN PUSTAKA	5
2.1 <i>Information Retrieval</i>	5
2.2 <i>Pre-Processing</i>	6
2.2.1 Tokenizing	6

2.2.2 Filtering / Stop Words.....	6
2.2.3 Stemming.....	6
2.2.4 Term Frequency Inverse Document Frequency (TF-IDF)	9
2.3 Similarity.....	10
2.3.1 Cosine Similarity.....	10
2.3.2 Dice Similarity.....	10
2.4 Terjemahan Al-qur'an As-Sa'di.....	11
2.5 Penelitian Sebelumnya	12
BAB III METODE USULAN	17
3.1 Metode.....	17
3.1.1 Stemming Nazief dan Adriani	17
3.1.2 Flowchart TF-IDF	19
3.1.3 Flowchart Dice Similarity.....	20
3.2 Perancangan Sistem.....	22
3.2.1 Arsitektur Sistem.....	22
3.2.2 Flowchart Pre-Processing	23
3.2.3 IPO.....	24
3.2.4 Use Case	25
3.2.5 Conceptual Data Model (CDM).....	26
3.2.6 Physical Data Model (PDM).....	26
3.3 Tahapan Penelitian	27
3.3.1 Metode Penelitian	27
3.4 Skenario Pengujian.....	28
3.4.1 Pengujian Ketepatan Software.....	28
3.4.2 Perhitungan Recall, Precision, F-Measure	29
BAB IV HASIL DAN PEMBAHASAN.....	30
4.1 Lingkungan Uji Coba	30

4.2 Implementasi Sistem	30
4.3 Pelaksanaan Skenario Pengujian	34
4.3.1 Pelaksanaan Pengujian Ketepatan Software.....	34
4.3.2 Pelaksanaan Pengujian <i>Recall, Precision, F-Measure</i>	39
4.3.3 Pelaksanaan Pengujian Waktu.....	43
4.4 Analisa Uji Coba	44
BAB V PENUTUP	47
5.1 Kesimpulan.....	47
5.2 Saran.....	47
DAFTAR PUSTAKA.....	48

DAFTAR GAMBAR

Gambar 3.1 Flowchart Stemming Nazief dan Adriani	18
Gambar 3.2 Flowchart TF-IDF.....	20
Gambar 3.3 Flowchart Dice Similarity	21
Gambar 3.4 Arsitektur Sistem	22
Gambar 3.5 Flowchart Tokenizing.....	23
Gambar 3.6 Flowchart Filtering	24
Gambar 3.7 Diagram IPO.....	25
Gambar 3.8 Use Case Sistem	25
Gambar 3.9 Rancangan Conceptual Data Model	26
Gambar 3.10 Rancangan Physical Data Model.....	27
Gambar 4.1 Uji Menggunakan Sistem tanpa Stemming Dice Similarity.....	34
Gambar 4.2 Uji Menggunakan Sistem tanpa Stemming Cosine Similarity	35
Gambar 4.3 Uji Menggunakan Sistem dengan Stemming Dice Similarity.....	37
Gambar 4.4 Uji Menggunakan Sistem dengan Stemming Cosine Similarity	37

DAFTAR TABEL

Tabel 2.1 Aturan Stemming Nazief dan Adriani.....	8
Tabel 2.2 Penelitian Sebelumnya	14
Tabel 4.1 Lingkungan Uji Coba	30
Tabel 4.2 TF-IDF Query tanpa Stemming	35
Tabel 4.3 Perhitungan dxq pada Query Mendapatkan	35
Tabel 4.4 Perhitungan dxq pada Query Pahala	35
Tabel 4.5 Hasil Perhitungan dxq tanpa Stemming	36
Tabel 4.6 Nilai TF-IDF ² Dokumen tanpa Stemming	36
Tabel 4.7 Hasil Perhitungan Similarity tanpa Stemming	36
Tabel 4.8 TF-IDF Query dengan Stemming.....	38
Tabel 4.9 Perhitungan dxq pada Query Bangkit	38
Tabel 4.10 Perhitungan dxq pada Query Semangat	38
Tabel 4.11 Hasil Perhitungan dxq dengan Stemming	38
Tabel 4.12 Nilai TF-IDF ² Dokumen dengan Stemming	39
Tabel 4.13 Perhitungan Similarity dengan Stemming.....	39
Tabel 4.14 Pengujian query tanpa Stemming Recall, Precision,F-Measure pada Dice Similarity	40
Tabel 4.15 Uji query dengan Stemming Recall, Precision,F-Measure pada Dice Similarity	41
Tabel 4.16 Pengujian query tanpa Stemming Recall, Precision,F-Measure pada Cosine Similarity	41
Tabel 4.17 Uji query dengan Stemming Recall, Precision,F-Measure pada Cosine Similarity	42
Tabel 4.18 Pengujian Waktu Dice Similarity Sudah di Eksekusi	44
Tabel 4.19 Pengujian Waktu Cosine Similarity Sudah di Eksekusi.....	44
Tabel 4.20 Pengujian Waktu Dice Similarity Belum di Eksekusi.....	44
Tabel 4.21 Pengujian Waktu Cosine Similarity Belum di Eksekusi	44
Tabel 4.22 Perbandingan nilai Recall, Precision, F-Measure setiap metode dengan tanpa stemming	45

Tabel 4.23 Perbandingan nilai Recall, Precision, F-Measure setiap metode dengan stemming	45
--	----

DAFTAR KODE PROGRAM

Kode Program 4. 1 Algoritma Stemming Nazief dan Adriani	32
Kode Program 4. 2 Algoritma Dice Similarity.....	33