

**PERBANDINGAN PERFORMANSI ANTARA ALGORITMA
COSINE SIMILARITY DENGAN *JACCARD SIMILARITY*
UNTUK PENCARIAN PADA PORTAL JURNAL**

TUGAS AKHIR

**Diajukan untuk Memenuhi Salah Satu Syarat Mencapai Gelar Strata Satu
Program Studi Informatika**



Disusun Oleh:

NOVI ESTU WARDANI

NIM. M0509049

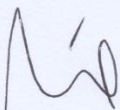
**PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
2016**

TUGAS AKHIR
PERBANDINGAN PERFORMANSI ANTARA ALGORITMA
***COSINE SIMILARITY* DENGAN *JACCARD SIMILARITY* UNTUK**
PENCARIAN PADA PORTAL JURNAL

Disusun Oleh :
NOVI ESTU WARDANI
M0509049

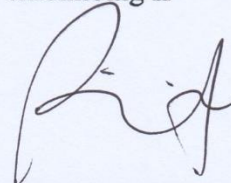
Telah disetujui untuk dipertahankan di hadapan dewan penguji
pada tanggal 18 Juli 2016

Pembimbing I



Ristu Saptono S.Si, M.T.
NIP. 197902 10 200212 1 001

Pembimbing II



Rini Anggrainingsih S.T, M.T.
NIP. 197809 09 200812 2 002

TUGAS AKHIR**PERBANDINGAN PERFORMANSI ANTARA ALGORITMA *COSINE SIMILARITY* DENGAN *JACCARD SIMILARITY* UNTUK PENCARIAN PADA PORTAL JURNAL**


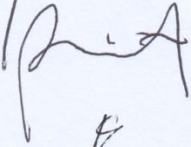
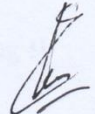
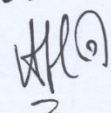
Disusun oleh :

NOVI ESTU WARDANI

Telah dipertahankan di hadapan Dewan Penguji

pada tanggal: 18 Juli 2016

Susunan Dewan Penguji

- | | | |
|--|-----------|--|
| 1. Ristu Saptono S.Si., M.T. | (Ketua) | () |
| NIP. 197902102002121001 | | |
| 2. Rini Anggrainingsih S.T., M.T. (Sekretaris) | | () |
| NIP . 197809092008122002 | | |
| 3. Esti Suryani, S.Si., M.Kom | (Anggota) | () |
| NIP. 197611292008122001 | | |
| 4. Afrizal Doewes S.Kom., M.Sc | (Anggota) | () |
| NIP. 198508312012121004 | | |

Disahkan Oleh

Kepala Program Studi Informatika


Drs. Bambang Harjito, M.App.Sc., Ph.D

NIP. 19621130 199103 1 002

MOTTO

Belajar dari kesalahan masa lalu, bekerja keras untuk masa kini, dan berharap hasil yang terbaik pada masa depan.



PERSEMBAHAN

Kupersembahkan karya ini untuk

Kedua orang tua tercinta atas dukungan dan tulus doanya

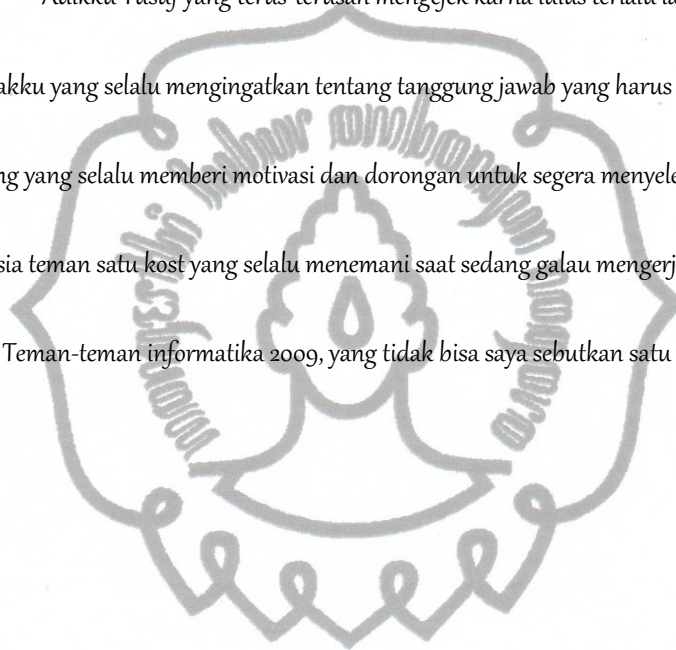
Adikku Yusuf yang terus-terusan mengejek karna lulus terlalu lama

Kakakku yang selalu mengingatkan tentang tanggung jawab yang harus diselesaikan

Mbah jambrong yang selalu memberi motivasi dan dorongan untuk segera menyelesaikan tugas akhir ini

Mbak nesia teman satu kost yang selalu menemani saat sedang galau mengerjakan tugas akhir

Teman-teman informatika 2009, yang tidak bisa saya sebutkan satu persatu



PERBANDINGAN PERFORMANSI ANTARA ALGORITMA *COSINE SIMILARITY* DENGAN *JACCARD SIMILARITY* UNTUK PENCARIAN PADA PORTAL JURNAL

NOVI ESTU WARDANI

Jurusan Informatika. Fakultas MIPA. Universitas Sebelas Maret

ABSTRAK

Pembuatan artikel ilmiah menjadi prasyarat kelulusan mahasiswa Universitas Sebelas Maret. Hal tersebut menyebabkan pertumbuhan jumlah artikel dalam portal jurnal UNS terus meningkat. Manajemen portal jurnal yang baik sangat dibutuhkan.

Fungsi pencarian merupakan fungsi yang penting dalam sebuah portal jurnal. Fungsi pencarian yang baik akan memudahkan pengguna dalam mencari jurnal yang diinginkan.

Penelitian ini menggunakan metode Cosine similarity dan Jaccard similarity untuk perhitungan kemiripan pada pencarian. Metode tersebut bertujuan untuk mengetahui precision maupun recall dari fungsi search. Data yang digunakan untuk pengujian adalah kata kunci yang diambil secara acak. Dari uji coba didapatkan hasil bahwa Cosine menghasilkan hasil pencarian yang lebih baik daripada Jaccard, nilai precision Cosine yang tertinggi mencapai 67.47%, dan nilai recall tertinggi yang dicapai adalah 99.91%, sedangkan nilai precision tertinggi Jaccard adalah 53.38% dan nilai recall tertingginya adalah 96.88%.

Kata kunci: *Jaccard similarity*, *Cosine similarity*, pencarian

PERFORMANCE COMPARATION BETWEEN OF COSINE SIMILARITY AND JACCARD SIMILARITY ALGORITHM FOR SEARCH IN PORTAL JOURNAL

NOVI ESTU WARDANI

Department of Informatic. Mathematic and Science Faculty
Sebelas Maret University

ABSTRACT

Article is one of graduation requirement for student of Sebelas Maret University, it case the growth of article amount in Sebelas Maret University journal portal is getting increassing. A good management in journal portal is really needed.

The function of search engine is an important function in a journal portal. A good function of finding will facilitate the users in looking for and searching the needed journal.

This study uses Cossine Similarity and Jaccard Similarity method to count the similarity on searching. The method is to know the precision as well as recall from search function. Data used in the study is keywords were selected randomly. From the try out, it was found that Cosine produces better search results than Jaccard, the highest precision cosine value reaches 67.47%, and the highest recall value is 99.91%, while the highest precision value of Jaccard is 53.38% and the highest recall value is 96.88%.

Keywords: *Jaccard similarity, Cosine similarity, search*

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Puji syukur penulis panjatkan kehadiran Allah SWT karena atas limpahan rahmat dan hidayahNya penulis dapat menyelesaikan penulisan skripsi. Shalawat serta salam senantiasa penulis haturkan kepada Baginda Rasulullah Muhammad SAW sebagai pembimbing umat manusia. Keberhasilan penyusunan skripsi ini tidak terlepas dengan bantuan dari banyak pihak, maka dari itu penulis ingin menyampaikan ucapan terima kasih yang sebesar- besarnya kepada :

1. Bapak Drs. Bambang Harjito, M.App.Sc., Ph.D. selaku Kepala program studi Informatika FMIPA UNS.
2. Bapak Ristu Saptono S.Si, M.T. selaku pembimbing I yang telah sabar memberikan banyak bimbingan, arahan, motivasi serta petunjuk kepada penulis dalam menyelesaikan skripsi ini.
3. Ibu Rini Anggrainingsih S.T, M.T. selaku pembimbing II yang telah sabar memberikan banyak bimbingan, arahan, motivasi serta petunjuk kepada penulis dalam menyelesaikan skripsi ini.
4. Bapak dan Ibu dosen Jurusan Informatika FMIPA UNS
5. Bapak dan Ibu Samsuri, orang tua terhebat yang tak hentinya memberikan semangat, serta dukungan moril maupun materil.
6. Teman-teman Informatika 2009 yang telah memberikan kebersamaan dan keteduhan selama menempuh studi di Jurusan Informatika UNS.

Semoga skripsi ini dapat memberikan manfaat sekaligus inspirasi kepada pembaca.

Wassalamu'alaikum Wr.Wb.

Surakarta,

Penulis

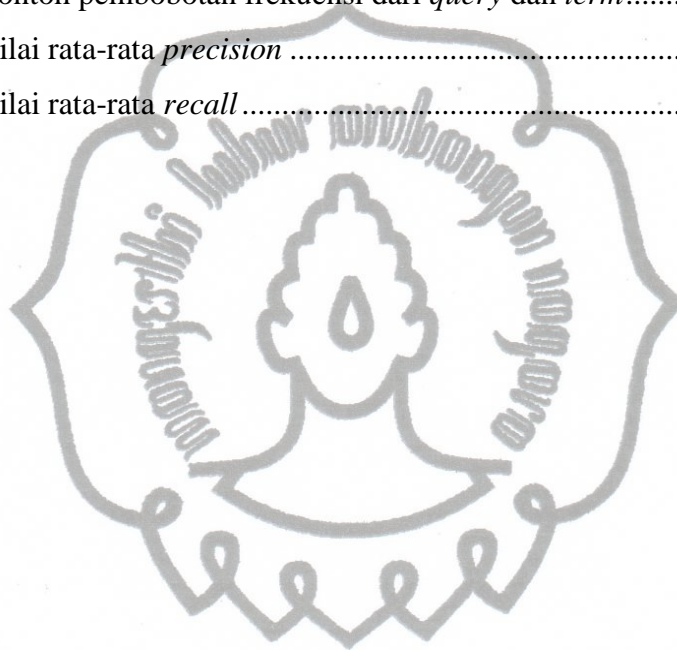
DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
MOTTO	iii
PERSEMBAHAN	iv
ABSTRAK.....	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I <u>PENDAHULUAN</u>	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II <u>TINJAUAN PUSTAKA</u>	5
2.1 Landasan Teori	5
2.1.1 Sistem Temu Kembali	5
2.1.2 <i>Text Mining</i>	6
2.1.3 Pembobotan TF/IDF	8
2.1.4 <i>Cosine similarity</i>	10
2.1.5 <i>Jaccard similarity</i>	11
2.1.6 <i>Precision dan Recall</i>	11
2.2 Penelitian Terkait	14
BAB III <u>METODOLOGI PENELITIAN</u>	18
3.1 Pengumpulan Data	18

3.2	Analisa penerapan metode.....	18
3.2.1	<i>Preprocessing</i> dokumen	19
3.2.2	Pembobotan dokumen	19
3.2.3	Perhitungan similaritas	19
3.3	Hasil dan evaluasi.....	19
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Deskripsi Data	20
4.2	Analisa dan penerapan metode.....	20
4.3	Implementasi <i>Vector Space Model</i>	22
4.4	Analisis Hasil Dan Pengujian.....	23
BAB V KESIMPULAN DAN SARAN.....		27
5.1	Kesimpulan.....	27
5.2	Saran.....	27
DAFTAR PUSTAKA		28
LAMPIRAN.....		30

DAFTAR TABEL

Tabel 2.1 Matrix Recall dan Precision	14
Tabel 2.2 Penelitian Terkait	17
Tabel 4.1 Daftar kata kunci yang digunakan untuk pengujian.....	21
Tabel 4.2 Batas similarity	22
Tabel 4.3 Contoh pembobotan frekuensi dari <i>query</i> dan <i>term</i>	23
Tabel 4.3 Nilai rata-rata <i>precision</i>	24
Tabel 4.4 Nilai rata-rata <i>recall</i>	25



DAFTAR GAMBAR

Gambar 2.1 Flowchart perhitungan Term Frequency (TF).....	8
Gambar 2.2 Flowchart Perhitungan TF-IDF.....	10
Gambar 2.3 Relasi antara <i>relevant</i> dan <i>retrieved</i> dokumen.	12
Gambar 2.4 Trade-off antara <i>precision</i> dan <i>recall</i>	14
Gambar 3.1 Diagram alir implementasi	18
Gambar 4.1 Grafik Perbandingan <i>Precision</i> Cosine dengan Jaccard	24
Gambar 4.2 Grafik Perbandingan <i>Recall</i> Cosine dengan Jaccard.....	25

