

”KOMBINASI ANTARA *TERM WEIGHTING* DENGAN *INTEGRATED COLOR AND INTENSITY CO- OCCURENCE MATRIX* UNTUK DUA LAPIS TEMU KEMBALI CITRA PADA MEDIA SOSIAL *TWITTER*”

Reza Pahlevi Bahrudin
(14.04.111.00056)

Pembimbing I : Dr. Indah Agustien Siradjuddin, S.Kom, M.Kom.
Pembimbing II : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRAK

Content Based Image Retrieval (CBIR) ialah salah satu teknik yang digunakan untuk melakukan pencarian citra berbasis isi. Meskipun hasil pencarian dari CBIR relevan dengan *query* yang diberikan oleh *user*, tetapi terkadang hasilnya tidak seperti yang diinginkan oleh *user*. Sehingga untuk mengatasi permasalahan tersebut dalam penelitian ini akan digunakan dua lapis sistem temu kembali yang mengkombinasikan *text retrieval* dengan model *Vector Space Model* (VSM) dan pembobotan menggunakan *Term Frequency - Inverse - Document Frequency* (TF-IDF) dikombinasikan dengan CBIR yang menggunakan *Integrated Color and Intensity Co-Occurrence Matrix* (ICICM) sebagai metode *feature extraction* untuk citra. *Feature* yang digunakan pada ICICM ialah tekstur dan warna yang merupakan 2 *feature* tingkat rendah yang sering digunakan untuk klasifikasi citra dan *retrieval*. Tujuan dari penelitian ini ialah mengkombinasikan dua lapis sistem temu kembali yaitu *text* dan *content based* untuk mendapatkan hasil *retrieval* yang lebih baik. Data yang digunakan dalam penelitian ini berasal dari media sosial *Twitter* dengan *query* teks berupa “wisata madura”. Hasil yang didapat dari penelitian ini adalah dengan menggunakan kombinasi *text based* dengan *content based* dapat meningkatkan *precision* dan *recall* hasil *retrieval* meskipun tanpa melakukan pencarian dalam seluruh *dataset*. Hasil yang didapatkan pada penelitian ini adalah dengan menggunakan kombinasi antara *text based* dengan *content based* hasilnya lebih baik dibandingkan dengan menggunakan *text based* saja yaitu dengan kenaikan sebesar 0,0576 untuk *precision* dan dengan *recall* yang sama. Sedangkan dengan menggunakan kombinasi antara *text based* dengan

content based hasilnya lebih baik dibandingkan dengan menggunakan *content based* saja yaitu dengan kenaikan sebesar 0,0238 untuk *precision* dan dengan *recall* sebesar 0,2000.

Kata kunci : *Content Based Image Retrieval, Integrated Color and Intensity Co-Occurrence Matrix, Vector Space Model, Term Frequency-Inverse - Document Frequency, Twitter*, dua lapis sistem temu kembali

DAFTAR ISI

HALAMAN PENGESAHAN.....	v
HALAMAN PERNYATAAN ORISINALITAS.....	vii
HALAMAN PERSEMBAHAN.....	ix
ABSTRAK.....	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL.....	xxiii
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	3
1.5. Batasan Masalah.....	3
1.6. Metode Penelitian.....	3
1.7. Sistematika Penulisan.....	4
BAB II.....	7
STUDI PUSTAKA.....	7
2.1. Penelitian Sebelumnya.....	7
2.2. <i>Dataset tweet</i> dari media sosial <i>twitter</i>	12
2.3. <i>Content Based Image Retrieval</i>	13
2.4. <i>Term Weighting</i> menggunakan <i>TF-IDF</i>	14
2.5. <i>Cosine Similarity</i>	17
2.6. <i>Integrated Color and Intensity Co-occurrence Matrix (ICICM)</i>	17
2.7. <i>City Block Similarity (Manhattan Distance)</i>	23
2.8. <i>Precision and Recall</i>	24
BAB III.....	25
PERANCANGAN SISTEM.....	25
3.1. Analisa sistem.....	25
3.2. Analisa kebutuhan.....	25
3.3. Desain Rancangan Sistem.....	26
3.4. Arsitektur sistem.....	28
BAB IV.....	33
IMPLEMENTASI SISTEM.....	33
4.1. Lingkungan Uji Coba.....	33
4.2. <i>Dataset tweet</i>	33

4.3. Pembuatan <i>Program</i>	36
4.4. Skenario Percobaan.....	50
4.5. Perhitungan <i>Precision</i> dan <i>Recall</i>	52
4.6. Hasil Uji Coba.....	105
4.7. Analisa Hasil Uji Coba.....	120
BAB V.....	125
PENUTUP.....	125
5.1. Kesimpulan.....	125
5.2. Saran.....	125
Daftar Pustaka.....	127

DAFTAR GAMBAR

Gambar 2. 1 Struktur .json dari sebuah <i>tweet</i>	12
Gambar 2. 2 <i>overview of CBIR system</i> [14].....	14
Gambar 2. 3 Contoh <i>text preprocessing</i> pada sebuah dokumen <i>tweet</i>	16
Gambar 2. 4 <i>Logical view of ICICM matrix</i> [4].....	20
Gambar 2. 5 Algoritma untuk <i>update matrix ICICM</i> [4].....	21
Gambar 2. 6 Hasil konversi citra RGB ke HSV.....	21
Gambar 2. 7 Ilustrasi Matrix awal Intensity.....	22
Gambar 2. 8 Ilustrasi Matrix Intensity setelah dikuantisasi menjadi 4 bit.....	22
Gambar 2. 9 Ilustrasi <i>matrix ICICMII</i> dengan <i>angle</i> 0°.....	22
Gambar 2. 10 Ilustrasi titik a dan b.....	23
Gambar 3. 1 Diagram IPO sistem.....	26
Gambar 3. 2 <i>Flowchart</i> rancangan sistem.....	28
Gambar 3. 3 <i>Flowchart text retrieval</i> menggunakan <i>TF-IDF</i> dan <i>Cosine Similarity</i>	31
Gambar 3. 4 <i>Flowchart image retrieval</i> menggunakan <i>ICICM</i> dan <i>Manhattan Distance</i>	32
Gambar 4. 1 Beberapa contoh citra pada <i>dataset</i>	34
Gambar 4. 2 <i>Database sqlite</i> untuk menyimpan <i>data tweet</i>	35
Gambar 4. 3 Tampilan CBIR.....	49
Gambar 4. 4 Tampilan <i>Text Retrieval</i>	49
Gambar 4. 5 <i>Text Retrieval</i> untuk kombinasi <i>Text and Image Retrieval</i>	50
Gambar 4. 6 <i>Image Retrieval</i> dari k teratas hasil <i>Text Retrieval</i>	50
Gambar 4. 7 Tampilan Setting.....	50
Gambar 4. 8 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 <i>query</i> teks 1 responden 1.....	54
Gambar 4. 9 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 <i>query</i> teks 2 responden 1.....	55
Gambar 4. 10 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 <i>query</i> teks 3 responden 1.....	56
Gambar 4. 11 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 <i>query</i> teks 5 responden 1.....	57
Gambar 4. 12 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 1 responden 1.....	58
Gambar 4. 13 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 2 responden 1.....	59
Gambar 4. 14 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 3 responden 1.....	60

Gambar 4. 15 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 query citra 5 responden 1.....	61
Gambar 4. 16 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 1 responden 1.....	62
Gambar 4. 17 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 2 responden 1.....	63
Gambar 4. 18 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 3 responden 1.....	64
Gambar 4. 19 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 5 responden 1.....	65
Gambar 4. 20 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 query teks 1 responden 2.....	66
Gambar 4. 21 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 query teks 2 responden 2.....	67
Gambar 4. 22 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 query teks 3 responden 2.....	68
Gambar 4. 23 Grafik <i>precision & recall</i> untuk 9 citra teratas skenario 1 query teks 4 responden 2.....	68
Gambar 4. 24 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 query teks 5 responden 2.....	69
Gambar 4. 25 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 query citra 1 responden 2.....	70
Gambar 4. 26 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 query citra 2 responden 2.....	71
Gambar 4. 27 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 query citra 3 responden 2.....	72
Gambar 4. 28 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 1 responden 2.....	73
Gambar 4. 29 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 2 responden 2.....	74
Gambar 4. 30 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 3 responden 2.....	75
Gambar 4. 31 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 4 responden 2.....	76
Gambar 4. 32 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi query teks dan citra 5 responden 2.....	77
Gambar 4. 33 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1 query teks 1 responden 3.....	78
Gambar 4. 34 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	

<i>query</i> teks 2 responden 3.....	79
Gambar 4. 35 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 3 responden 3.....	80
Gambar 4. 36 Grafik <i>precision & recall</i> untuk 9 citra teratas skenario 1	
<i>query</i> teks 4 responden 3.....	81
Gambar 4. 37 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 5 responden 3.....	82
Gambar 4. 38 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 1 responden 3.....	83
Gambar 4. 39 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 2 responden 3.....	84
Gambar 4. 40 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 3 responden 3.....	85
Gambar 4. 41 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 4 responden 3.....	85
Gambar 4. 42 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 5 responden 3.....	86
Gambar 4. 43 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3	
kombinasi <i>query</i> teks dan citra 1 responden 3.....	87
Gambar 4. 44 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3	
kombinasi <i>query</i> teks dan citra 2 responden 3.....	88
Gambar 4. 45 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3	
kombinasi <i>query</i> teks dan citra 3 responden 3.....	89
Gambar 4. 46 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3	
kombinasi <i>query</i> teks dan citra 4 responden 3.....	90
Gambar 4. 47 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3	
kombinasi <i>query</i> teks dan citra 5 responden 3.....	91
Gambar 4. 48 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 1 responden 4.....	92
Gambar 4. 49 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 2 responden 4.....	93
Gambar 4. 50 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 3 responden 4.....	94
Gambar 4. 51 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 1	
<i>query</i> teks 5 responden 4.....	95
Gambar 4. 52 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 1 responden 4.....	96
Gambar 4. 53 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2	
<i>query</i> citra 2 responden 4.....	97

Gambar 4. 54 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 3 responden 4.....	98
Gambar 4. 55 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 4 responden 4.....	99
Gambar 4. 56 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 2 <i>query</i> citra 5 responden 4.....	100
Gambar 4. 57 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi <i>query</i> teks dan citra 1 responden 4.....	101
Gambar 4. 58 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi <i>query</i> teks dan citra 2 responden 4.....	102
Gambar 4. 59 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi <i>query</i> teks dan citra 3 responden 4.....	103
Gambar 4. 60 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi <i>query</i> teks dan citra 4 responden 4.....	104
Gambar 4. 61 Grafik <i>precision & recall</i> untuk 20 citra teratas skenario 3 kombinasi <i>query</i> teks dan citra 5 responden 4.....	105
Gambar 4. 62 Beberapa citra hasil <i>retrieval</i> dengan <i>query</i> “Pantai indah”	108
Gambar 4. 63 Beberapa citra <i>retrieval</i> dengan <i>query</i> “Bukit Jaddih Madura”.....	108
Gambar 4. 64 Beberapa citra hasil <i>retrieval</i> dengan <i>query</i> “Api Madura”	108
Gambar 4. 65 Beberapa citra hasil <i>retrieval</i> dengan <i>query</i> “Kerapan Sapi”	109
Gambar 4. 66 Beberapa citra hasil <i>retrieval</i> dengan <i>query</i> “Jembatan Suramadu”.....	109
Gambar 4. 67 Beberapa citra <i>retrieval</i> dengan <i>query</i> citra 1.....	113
Gambar 4. 68 Beberapa citra <i>retrieval</i> dengan <i>query</i> citra 2.....	114
Gambar 4. 69 Beberapa citra <i>retrieval</i> dengan <i>query</i> citra 3.....	114
Gambar 4. 70 Beberapa citra <i>retrieval</i> dengan <i>query</i> citra 4.....	114
Gambar 4. 71 Beberapa citra <i>retrieval</i> dengan <i>query</i> citra 5.....	114
Gambar 4. 72 Contoh beberapa citra <i>retrieval</i> hasil uji coba dengan <i>query</i> 1.....	119
Gambar 4. 73 Contoh beberapa citra <i>retrieval</i> hasil uji coba dengan <i>query</i> 2.....	120
Gambar 4. 74 Contoh beberapa citra <i>retrieval</i> hasil uji coba dengan <i>query</i> 3.....	120
Gambar 4. 75 Contoh beberapa citra <i>retrieval</i> hasil uji coba dengan <i>query</i> 4.....	120
Gambar 4. 76 Contoh beberapa citra <i>retrieval</i> hasil uji coba dengan <i>query</i> 5.....	120
Gambar 4. 77 Grafik perbandingan rata-rata performansi <i>retrieval</i>	

seluruhnya..... 122