

BAB I

PENDAHULUAN

1.1. Latar Belakang

Semakin berkembangnya teknologi menjadikan semakin banyak bentuk dokumen, mulai dari teks, citra, suara, maupun video. Proses penyimpanan maupun pencarian data pun sangatlah beragam. Khususnya, proses pencariannya dituntut secara cepat dan menghasilkan data yang akurat.

Citra merupakan salah satu bentuk data yang paling banyak digunakan oleh manusia. Oleh karena itu, diperlukan sebuah teknik pencarian yang prosesnya cepat dan akurasi tinggi.

Salah satu teknik yang digunakan untuk melakukan pencarian citra ialah Sistem Perolehan Citra berbasis Isi (SPCI) atau CBIR (*Content Based Image Retrieval*). Yaitu dengan mencari kemiripan citra yang menjadi *query* dalam pencarian kemudian dibandingkan kemiripan dengan citra yang ada pada *database* [1].

Hasil pencarian yang ditampilkan pada CBIR konvensional tidak dapat menemukan hasil yang optimal dan tidak sesuai dengan apa yang kita inginkan, dikarenakan pencariannya dilakukan berdasarkan nama *file* [2]. CBIR memiliki kelemahan yaitu meskipun hasil pencarian sesuai dengan *query* yang diberikan, tetapi tidak seperti yang diinginkan oleh *user*. Seperti CBIR yang selama ini kita kenal yaitu *Google Image Search* [3].

Untuk mengatasi permasalahan tersebut digunakan dua lapis pencarian citra, yaitu pertama dengan menggunakan *text retrieval* yang digunakan untuk melakukan pencarian berdasarkan *query* teks setelah itu dilanjutkan dengan memilih sebuah citra yang dijadikan *query* dan akan di-*filter* dari hasil pencarian teks untuk mencari citra yang mirip dengan *query* citra yang diberikan.

Selain untuk mendapatkan hasil citra yang diinginkan, tujuan kombinasi antara *text based* dengan *content based* adalah agar hasil pencarian lebih efisien tanpa mengurangi akurasi hasil *retrieval*. Misalnya, jika hanya menggunakan *query image* pada CBIR, proses pencarian dilakukan pada seluruh *database* citra. Akan tetapi, jika menggunakan metode yang diajukan, pencarian dilakukan pada hasil teratas yang ditentukan dari hasil filter pencarian menggunakan *text based*. Setelah menggunakan metode yang diajukan hasil akurasi

retrieval tetap tinggi meskipun tanpa harus menggunakan seluruh *database*.

Warna dan tekstur ialah dua *feature* tingkat rendah yang banyak digunakan untuk klasifikasi citra, *indexing*, dan *retrieval*. Warna biasanya direpresentasikan dengan menggunakan *histogram*, ialah pengukuran statistik yang diperoleh dari distribusi global warna pada sebuah citra.

Penelitian ini akan membahas implementasi metode *Integrated Color and Intensity Co-occurrence Matrix* (ICICM) yang digunakan untuk *image retrieval* dan untuk *text retrieval* menggunakan TF-IDF yang akan dikombinasikan pada 2 lapis temu kembali citra pada media sosial *twitter*.

ICICM merupakan metode yang merepresentasikan warna dan intensitas dari ketetanggaan piksel pada suatu citra menggunakan *Co-occurrence Matrix*[4]. CCM merupakan metode yang menghasilkan variasi warna pada suatu citra. CCM merepresentasikan matriks tiga dimensi, dimana pasangan p dan N_p (piksel dan tetangga piksel) diperoleh dari matriks dua dimensi pertama dan jarak spasial d diantara dua piksel pada tiga dimensi. CCM dihasilkan hanya dengan ruang bidang *Hue* dari ruang warna HSV (*Hue, Saturation, Intensity Value*).

Kelebihan metode ICICM dibandingkan dengan metode lainnya ialah meningkatnya akurasi *precision*. Karena metode ini melakukan integrasi variasi baik pada tingkat warna maupun intensitas setiap ketetanggaan piksel menggunakan ruang warna HSV [4]. Sedangkan metode lainnya, jumlah warna serta variasi intensitas setiap piksel dan tetangganya diperkirakan dengan menggunakan fungsi bobot. Kelebihan utama menggunakan ICICM ialah metode ini menghindari penggunaan bobot yang menggabungkan fitur warna dan tekstur .

Tujuan dari penelitian ialah untuk membandingkan hasil dari dua lapis temu kembali citra yang menggunakan metode ekstraksi fitur tekstur dan warna yaitu ICICM dikombinasikan dengan *term weighting* menggunakan TF-IDF dibandingkan dengan *text retrieval* dan *content based image retrieval*.

1.2. Rumusan Masalah

Rumusan masalah yang dibahas dalam penelitian ini adalah:

- Berapa perbandingan akurasi pencarian antara *text based retrieval* dengan kombinasi *text and content based retrieval*.
- Berapa perbandingan akurasi pencarian antara *content based retrieval* dengan kombinasi *text and content based retrieval*.

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

- a. Mengetahui perbandingan akurasi pencarian antara *text based retrieval* dengan kombinasi *text and content based retrieval*.
- b. Mengetahui perbandingan akurasi pencarian antara *content based retrieval* dengan kombinasi *text and content based retrieval*.

1.4. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah untuk menyelesaikan masalah citra hasil pencarian dari CBIR konvensional yang seringkali tidak relevan sesuai dengan keinginan *user*. Dan dengan menggunakan dua lapis temu kembali citra diharapkan hasil *retrieval* menjadi lebih baik.

1.5. Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang dikumpulkan hanya data dengan *query* “wisata madura”.
2. Data yang digunakan adalah data *tweet* yang berasal dari media sosial *twitter*. Data diambil melalui *REST API Twitter* yang formatnya berupa *.json*. Kemudian dilakukan proses *parsing* untuk mengambil data *tweet* yang ada citranya, atau dengan kata lain *tweet* tersebut memiliki *key* “*images*”.
3. *Quantization factor* yang digunakan adalah 16 bit dengan sudut orientasi 0°.
4. Citra hasil *crawl* pada *database* di *resize* dengan *base* 180.

1.6. Metode Penelitian

Secara umum, penelitian dilakukan dengan beberapa tahapan, yaitu :

1. Studi Literatur

Pada tahap ini pengumpulan informasi yang dibutuhkan. Informasi didapat dari membaca, mempelajari literatur, makalah, jurnal ilmiah dan juga informasi yang disediakan oleh situs-situs *web* yang berkaitan dengan penelitian skripsi, diantaranya :

- a. Mempelajari *REST API twitter*.
- b. Mempelajari *parsing data twitter*.
- c. Mempelajari proses dasar pengolahan citra untuk CBIR.

- d. Mempelajari TF-IDF untuk pembobotan data *tweet* yang didapat dari *twitter*.
 - e. Mempelajari metode ICICM (*Integrated Color and Intensity Co-occurrence Matrix*).
 - f. Mempelajari pengukuran jarak / *similarity measurement* ICICM dengan menggunakan *Manhattan Distance*.
 - g. Mempelajari VSM (*Vector Space Model*) yang digunakan untuk mencari relevansi dari *document* yang menggunakan pembobotan dan direpresentasikan sebagai *vector*.
2. Pengumpulan Data
Data yang diperoleh berasal dari *REST API twitter* dan akan dikumpulkan sebanyak-banyaknya dengan *query* teks berupa “wisata madura”.
 3. Analisa dan Perancangan Sistem
Pada tahap ini diperoleh setelah melihat data yang didapatkan serta akan mengimplementasikan algoritma yang digunakan pada skripsi ini untuk mendapatkan sistem yang diinginkan.
 4. Implementasi Sistem
Pada skripsi ini sistem dibangun dengan menggunakan bahasa pemrograman *python*.
 5. Uji Coba Sistem
Sistem akan diuji coba dengan menggunakan data yang berasal dari media sosial *twitter*. Ada 3 skenario uji coba yang dilakukan yaitu dengan membandingkan hasil dari *text retrieval* dibandingkan dengan kombinasi *text & content based retrieval* serta *content based retrieval* dibandingkan dengan *text & content based retrieval*.
 6. Analisa dan Evaluasi
Setelah sistem selesai dibuat, sistem akan diuji akurasi dengan menggunakan skenario uji coba dan hasil *retrieval* nya akan di validasi oleh beberapa responden..
 7. Dokumentasi
Dilakukan penyusunan buku sebagai laporan skripsi yang berisi dokumentasi tahap awal sampai tahap akhir.

1.7. Sistematika Penulisan

Adapun sistematika penulisan skripsi ialah sebagai berikut.

1. BAB I PENDAHULUAN

Pada bab 1 ini berisikan tentang latar belakang dari sistem dua lapis

kembali yang mengkombinasikan *text retrieval* dengan *image retrieval*, persamaan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan skripsi.

2. BAB II STUDI PUSTAKA

Pada bab 2 ini berisikan tentang penjelasan dari teori yang mendasari penelitian ini. Teori yang akan dibahas pada bab ini yaitu CBIR, TF-IDF, ICICM, dan VSM (*Cosine Similarity*),

3. BAB III PERANCANGAN SISTEM

Pada bab 3 ini berisikan tentang seluruh perancangan pembuatan dari sistem di skripsi ini. Dari *software dependencies* dan *hardware* yang digunakan dalam penelitian skripsi ini. Serta *workflow* dari sistem.

4. BAB IV IMPLEMENTASI SISTEM

Pada bab 4 ini berisikan tentang penjelasan implementasi dari sistem yang sebelumnya sudah dibuat rancangannya pada bab 3. Dan sistem yang telah selesai akan diuji coba menggunakan skenario-skenario ujicoba yang sudah ditentukan untuk mengukur akurasi.

5. BAB V PENUTUP

Pada bab terakhir, bab 5 berisikan tentang kesimpulan dari penelitian skripsi ini yang sudah diuji coba dengan menggunakan skenario yang sudah ditentukan.

~Halaman Ini Sengaja Dikosongkan~