

Studi Penggunaan Data Exif Untuk Mengukur Pengaruhnya Terhadap Peningkatan Kinerja *Image Search Engine*

Nugroho Herucahyono (13504038)

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung

Jl. Ganesha 10, Bandung

E-mail: xinuc@xinuc.org

Abstrak - Data EXIF adalah data yang disimpan oleh kamera digital pada suatu *image* yang berisi informasi mengenai kondisi dan setting kamera digital pada waktu dilakukan pemotretan. Beberapa informasi yang disimpan pada data EXIF adalah tanggal dan waktu pengambilan gambar, ukuran, deskripsi, *copyright*, dan sebagainya. Data ini dapat digunakan untuk membantu pencarian gambar di Internet oleh *image search engine*.

Makalah ini membahas mengenai penggunaan data EXIF tersebut untuk meningkatkan kinerja *image search engine*. Perangkat lunak yang dibangun bernama ExifSearch, dikembangkan dengan *framework* Ruby on Rails dengan bahasa pemrograman Ruby. Pengujian dilakukan dengan menggunakan beberapa kasus uji yang telah diketahui relevansinya, sehingga efektivitas pencarian dapat dihitung menggunakan nilai *recall* dan *precision*.

Dari hasil pengujian, tidak didapatkan perubahan yang terpola pada nilai *precision* dan *recall*, sehingga belum dapat dipastikan pengaruh penggunaan data EXIF terhadap nilai *precision* dan *recall* suatu pencarian. Selain itu, penggunaan data EXIF cenderung mengurangi jumlah hasil pencarian. Keuntungan utama penggunaan data EXIF dalam pencarian yang dapat disimpulkan dari Tugas Akhir ini adalah memungkinkan user untuk membentuk *query* yang lebih spesifik.

Kata Kunci: Data EXIF, *information retrieval*, *image search engine*.

I. Pendahuluan

Internet telah berkembang menjadi sebuah “perpustakaan” data yang sangat besar. Di Internet, kita dapat menemukan berbagai data dan informasi yang kita perlukan. Dengan jumlah data yang sangat besar tersebut, kita akhirnya menghadapi suatu masalah baru, yaitu cara untuk menemukan data yang kita perlukan diantara milyaran data yang terdapat di Internet. Data yang terdapat di Internet memiliki berbagai macam bentuk, misalnya halaman html, xml, pdf, gambar / *image*, suara, dan lain – lain. Penggunaan mesin pencari (*search engine*) sangat membantu pengguna Internet untuk menemukan dokumen yang diinginkannya. Namun hingga saat ini, kemampuan *search engine* tersebut masih cukup terbatas, terutama jika berhubungan dengan data *non-text*, misalnya gambar, suara dan video.

Di sisi lain, perkembangan teknologi juga telah mendorong semakin banyaknya data yang disimpan dalam berbagai bentuk. Salah satu teknologi tersebut adalah kamera digital. Kamera digital merupakan teknologi yang berkembang sangat pesat dan hampir telah sepenuhnya menggantikan penggunaan kamera analog (menggunakan film). Kamera digital memungkinkan pengguna untuk mengambil gambar dengan biaya yang jauh lebih murah dibandingkan dengan menggunakan kamera analog. Hal ini akan semakin memperbanyak jumlah data dalam bentuk gambar yang tersebar di Internet.

Masalah yang dihadapi oleh *search engine* dalam menemukan data *non-text* terutama adalah karena dokumen yang tidak berupa *string*, sedangkan *query*

yang dimasukkan oleh pengguna adalah berupa *string*. Dalam makalah ini akan dibahas mengenai penggunaan data EXIF untuk membantu peningkatan kemampuan *search engine* tersebut.

Data EXIF (*Exchangeable Image File Format*) adalah suatu data yang disisipkan pada suatu *image* untuk menyimpan beberapa informasi mengenai *image* tersebut, terutama informasi mengenai *setting* kamera digital yang digunakan pada saat mengambil gambar. Pada awalnya data EXIF diusulkan oleh *Japan Electronics and Information Technology Industries Association* (JEITA) sebagai format data yang dapat dipertukarkan (*exchangeable*).

II. Analisis

Inti permasalahan dari sebuah *information retrieval system* adalah bagaimana menentukan apakah suatu dokumen relevan dengan *query* yang dimasukkan oleh user. Secara manual, manusia memiliki kemampuan intelektual untuk menentukan relevansi antara dokumen dengan *query* pada domain yang dikuasainya. Namun untuk pencarian dengan komputer, perlu dibuat suatu model sehingga relevansi dapat dihitung dan ditentukan secara kuantitas. Salah satu model yang cukup efektif dan banyak digunakan adalah model ruang vektor.

Pada model ruang vektor, kata – kata pada dokumen dan *query* (*index term*) dimodelkan sebagai suatu ruang vektor berdimensi n dengan n adalah jumlah term yang berbeda pada dokumen dan *query* tersebut. Relevansi dihitung dengan membandingkan perbedaan arah atau sudut yang terbentuk oleh vektor – vektor tersebut. Secara matematis, relevansi pada model ini dihitung dengan nilai \cos dari sudut yang terbentuk.

Model ruang vektor mengasumsikan bahwa baik dokumen maupun *query* berbentuk *string*, sehingga dapat dibandingkan dengan mudah. Model ruang vektor tidak dapat digunakan jika dokumen atau *query* memiliki bentuk selain *string*, misalnya gambar.

Untuk pencarian gambar, dapat digunakan beberapa model khusus yang diciptakan untuk mengenali bentuk dari gambar, dengan teknologi *image recognition*. Namun teknologi *image recognition* masih bersifat *experimental* dan dirasa kurang sesuai dengan kebutuhan *image search engine* dimana user lebih menyukai memasukkan *query* dalam bentuk *string*. Karena alasan tersebut, maka *image search engine* yang akan dibangun akan mencoba menerapkan teknik pencarian yang diterapkan pada

dokumen teks pada dokumen gambar, yaitu dengan model ruang vektor.

Dalam *image search engine* dimana dokumen berbentuk gambar, *term* yang digunakan dalam pengukuran relevansi diambil dari beberapa teks yang berhubungan dengan gambar tersebut, misalnya nama *file*, atribut ALT pada tag ``, judul halaman *web*, link menuju gambar tersebut, kata – kata di sekitar gambar dan sebagainya. Komponen teks yang paling akurat menunjukkan *content* dari suatu *image* adalah nama *file* dan atribut ALT. Pada makalah ini, komponen teks yang digunakan untuk mengetahui *content* dari suatu *image* adalah:

- a. Nama *file*.
- b. Atribut ALT pada tag ``.
- c. Metadata EXIF yang tersimpan di dalam *image*.

Dalam pencarian, suatu *term* yang digunakan mungkin lebih penting / signifikan dibandingkan dengan *term* yang lain. Pada *image search engine* yang akan dibangun, *term* yang digunakan berasal dari beberapa komponen teks gambar seperti yang telah disebutkan. Teks yang dihasilkan sederhana dan cukup pendek sehingga tidak memerlukan normalisasi serta penghitungan TF cukup menggunakan teknik *Raw TF*.

Agar penghitungan relevansi dokumen lebih cepat, maka perlu dilakukan proses *indexing*. Pada *image search engine* yang akan dibangun, proses *indexing* akan dilakukan terhadap suatu data deskripsi yang diekstrak dari *image* tersebut. Ketika suatu *image* diproses, maka komponen teks dari *image* tersebut diambil dan dijadikan “deskripsi” dari *image*. Deskripsi ini mewakili dokumen *image* tersebut dan dapat dipandang sebagai suatu dokumen teks biasa. *Term* yang digunakan dalam proses *indexing* diambil dari deskripsi ini.

Indexing dilakukan dengan membuat suatu *inverted index* yang menyimpan *term*, dokumen *image* yang mengandung *term* tersebut, dan jumlah kemunculan *term* pada dokumen (TF). Pada saat proses pencarian, relevansi dihitung dengan TF dari *index* ini, dan IDF dihitung dari jumlah dokumen dengan *term* tersebut.

Pada proses *searching*, relevansi dihitung dengan menggunakan *index* yang telah dibuat dengan model ruang vektor. Jika dokumen relevan dengan *query*, maka akan dimasukkan ke dalam hasil pencarian. Data EXIF (kecuali *image description*) akan digunakan dalam proses *filtering*. User dapat menentukan beberapa parameter, sehingga hasil

pencarian akan di *filter* menggunakan parameter tersebut sebelum ditampilkan kepada *user*.

Pada makalah ini, hanya akan digunakan beberapa *tag* saja yang dapat dimanfaatkan dalam proses *image searching*, yaitu:

1. *Date/Time*
Mencatat waktu pada saat gambar diambil.
2. *Exif Image Width*
Mencatat dimensi lebar *image* dalam *pixel*.
3. *Exif Image Height*
Mencatat dimensi tinggi *image* dalam *pixel*.
4. *Software*
Ditulis oleh *image manipulation program* pada saat menyimpan gambar akan mencatat *software terakhir* yang digunakan untuk memanipulasi *image*.
5. *Make*
Mencatat produsen / *manufacturer* kamera digital.
6. *Artist*
Tag ini secara *default* kosong. Dapat diisi dengan nama *artist* / *kameramen* yang mengambil gambar.
7. *Orientation*
Mencatat orientasi gambar.
8. *Image Description*
Tag ini secara *default* kosong. Dapat diisi dengan deskripsi *image* tersebut.
9. *Copyright*
Tag ini secara *default* kosong. Dapat diisi dengan data hak cipta *image* tersebut.

III. Perangkat Lunak

Perangkat lunak yang akan dibangun adalah sebuah *image search engine* berbasis *web* yang selanjutnya akan disebut dengan “ExifSearch”. ExifSearch dapat digunakan untuk mencari dokumen dalam bentuk *image JPEG* dengan menggunakan data EXIF untuk membantu pencarian. ExifSearch memiliki dua mode pencarian, yaitu:

- a. *Normal Search*
Pada mode ini, pencarian dilakukan dengan membandingkan *keyword query* yang dimasukkan *user* dengan data teks dari suatu *image*, yaitu nama *file*, atribut ALT dan keterangan *image description* dari data EXIF.
- b. *Advanced Search*
Pada mode ini, selain memasukkan *keyword*, *user* juga dapat menentukan berbagai parameter atribut suatu *image* yang dicarinya, misalnya waktu pengambilan dan ukuran gambar.

Parameter yang dimasukkan oleh *user* tersebut akan dibandingkan dengan data EXIF yang dimiliki oleh suatu *image*. Pencarian dilakukan dengan membandingkan *keyword* dengan nama *file*, atribut ALT, dan *image description* data EXIF, kemudian *image* yang sesuai akan difilter lagi dengan parameter yang dimasukkan *user* dengan bantuan data EXIF masing – masing *image*.

ExifSearch memiliki suatu *webcrawler* dimana *user* dapat memasukkan URL dari sebuah *web* yang kemudian akan diproses dengan diambil data *image* pada *web* tersebut. *Image* yang ditemukan kemudian diambil nama *file*, atribut ALT, dan data EXIF yang diperlukan untuk kemudian disimpan dalam *database*.

Ada beberapa kebutuhan fungsional dari perangkat lunak *image search engine* yang akan dikembangkan, yaitu:

1. Dapat melakukan *parsing* terhadap suatu halaman *web* dan menemukan *image* pada halaman *web* tersebut.
2. Dapat membaca / mengambil atribut teks dari *image* yang ditemukan, berupa nama *file*, atribut ALT dari *tag *, serta beberapa data EXIF yang telah disebutkan.
3. Dapat melakukan pencarian *image* sesuai dengan *query* yang dimasukkan *user*.
4. Memiliki opsi / pilihan pencarian, dengan menggunakan data EXIF dan tidak menggunakan data EXIF.
5. Dapat menampilkan hasil pencarian berupa *image* hasil pencarian (dalam ukuran lebih kecil) dan *link* menuju *image* tersebut.

ExifSearch merupakan aplikasi berbasis *web* yang dikembangkan dengan *framework* Ruby On Rails. Bahasa pemrograman yang digunakan adalah Ruby dengan menggunakan *interpreter* Jruby yang berjalan di atas Java Virtual Machine.

Aplikasi yang dibangun pada memiliki batasan sebagai berikut:

- a. *Indexing* yang dilakukan adalah *indexing* sederhana, tanpa adanya normalisasi, *stemming* (pengembalian ke bentuk dasar) dan penghilangan kata – kata tak berpengaruh (*stop word*)
- b. Relevansi suatu dokumen hanya dihitung menggunakan TF (*term frequency*) dan IDF (*inverse document frequency*), dimana TF dihitung menggunakan Raw TF (banyak

kemunculan suatu *term*) dan IDF dihitung dengan logaritma.

IV. Pengujian

Pada bagian ini akan dijelaskan pengujian perangkat lunak. Pengujian perangkat lunak yang dilakukan memiliki beberapa tujuan, yaitu:

1. Mengetahui apakah perangkat lunak yang diimplementasikan telah sesuai dengan kebutuhan – kebutuhan utama perangkat lunak yang dispesifikasikan pada tahap analisis.
2. Mengetahui apakah keluaran perangkat lunak berupa hasil pencarian telah sesuai dengan aturan yang diterapkan.
3. Mengetahui kualitas hasil pencarian serta dampak dari penggunaan data EXIF (*precision* dan *recall* sebelum dan sesudah penggunaan data EXIF).

Terdapat beberapa kasus uji yang digunakan dalam pengujian perangkat lunak ini, yaitu:

1. Kasus uji untuk pengujian fungsional:
 - a. Kasus uji untuk melihat fungsionalitas *crawler*. *Database* dalam keadaan kosong, disediakan beberapa halaman *web* yang saling terhubung. URL dari halaman tersebut dimasukkan ke dalam *crawler* untuk memulai proses *crawling*.
 - b. Kasus uji untuk melihat fungsi pembobotan dalam pencarian. Pada kasus uji ini, *database* sudah terisi dengan data yang diketahui bobot masing – masing *image*. Dilakukan pencarian dengan suatu *keyword* tertentu kemudian dilihat apakah hasil pencarian terurut berdasarkan bobot *image*.
 - c. Kasus uji untuk melihat fungsi pencarian dan untuk membandingkan antar pencarian normal dengan pencarian lanjut. Pada kasus uji ini, *database* terisi data dengan jumlah cukup besar. Dilakukan pencarian dengan mode pencarian normal dan pencarian lanjut, kemudian dilihat apakah hasil pencarian berbeda.
2. Kasus uji untuk pengujian pencarian:

Kasus uji pada bagian ini ditujukan untuk mengukur dan membandingkan efektifitas pencarian antara mode pencarian normal dan pencarian lanjut. Untuk semua kasus yang digunakan, telah diketahui relevansi / hasil pencarian yang seharusnya didapatkan. Karena *query* yang diberikan untuk kedua mode pencarian tersebut berbeda, maka dokumen yang relevan untuk masing – masing mode pencarian juga berbeda. Efektifitas yang diukur dengan

nilai *recall* dan *precision* juga akan menggunakan jumlah dokumen relevan yang berbeda. Pengujian ini membutuhkan data dengan jumlah yang cukup besar dan acak. Pada kasus uji ini akan digunakan data 2000 *image* yang diambil secara acak dari Internet yang meliputi berbagai topik yang berbeda.

- a. Kasus uji pencarian pertama, menggunakan *keyword* “*bird*” untuk melakukan pencarian terhadap *image* yang berhubungan dengan burung. Parameter pencarian yang akan dimasukkan pada mode pencarian lanjut adalah ukuran lebar *image* yang kurang dari 2000 *pixels*.
- b. Kasus uji pencarian kedua, menggunakan *keyword* “*car*” untuk mencari *image* yang berhubungan dengan mobil. Parameter pencarian yang akan dimasukkan pada mode pencarian lanjut adalah tanggal pengambilan *image* setelah 1 Januari 2008.
- c. Kasus uji pencarian ketiga, menggunakan *keyword* “*cat*” untuk mencari *image* yang berhubungan dengan kucing. Parameter pencarian yang akan dimasukkan pada mode pencarian lanjut adalah orientasi *image portrait*.
- d. Kasus uji pencarian keempat, menggunakan *keyword* “*asmirandah*” untuk mencari *image* yang berhubungan dengan Asmirandah. Parameter pencarian yang akan dimasukkan pada mode pencarian lanjut adalah *image* tanpa *copyright* (*field copyright* kosong).

Hasil pengujian Fungsional

- a. Pada kasus uji pertama, *crawler* berhasil menemukan *image* yang berada pada halaman – halaman *web* yang disediakan, serta mampu mengambil berbagai data yang diperlukan untuk *indexing*.
- b. Pada kasus uji kedua, pencarian dilakukan dengan *keyword* “*asmirandah*”, dimana ada 16 dokumen *image* yang relevan dengan *keyword* tersebut. Masing – masing *image* memiliki jumlah kata “*asmirandah*” yang berbeda pada deskripsinya. Hasil dari pencarian ditampilkan terurut sesuai bobot *image* yang telah diketahui.
- c. Pada kasus uji ketiga, pencarian dengan *keyword* “*asmirandah*” dilakukan dengan mode normal dan lanjut, dimana pada mode lanjut dimasukkan parameter hasil pencarian berupa tanggal pengambilan gambar : sebelum tanggal 1 Januari 2007.

Hasil Pengujian Pencarian

- a. Kasus uji pertama menggunakan *keyword* “bird”. Total semua *image* yang relevan dengan *keyword* tersebut yang ada di *database* adalah 9. Dari pencarian normal dihasilkan 11 *image* (2 *image* mengandung kata *bird* pada deskripsi, sedang *image* tersebut sebenarnya bukan tentang burung). Sedang pada *query* lanjut dengan lebar *image* kurang dari 2000 *pixels*, ada 4 *image* yang relevan. Hasil pencarian memunculkan 3 *image* (1 *image* tidak memiliki data EXIF).
- b. Kasus uji kedua menggunakan *keyword* “car”. Total semua *image* yang relevan dengan *keyword* tersebut yang ada di *database* adalah 54. Dari pencarian normal dihasilkan 45 *image* dengan 41 *image* relevan. Sedang pada *query* pencarian lanjut dengan tanggal pengambilan setelah 1 Januari 2008, menghasilkan 37 *image* dengan 35 *image* relevan. Total *image* relevan di *database* untuk *query* ini adalah 38.
- c. Kasus uji ketiga menggunakan *keyword* “cat”. Total semua *image* yang relevan dengan *keyword* tersebut yang ada di *database* adalah 30. Dari pencarian normal dihasilkan 22 *image* dengan 18 *image* relevan. Sedang pada *query* pencarian lanjut dengan orientasi *image portrait*, menghasilkan 5 *image* dengan 4 *image* relevan. Total *image* relevan di *database* untuk *query* ini adalah 4.
- d. Kasus uji ketiga menggunakan *keyword* “asmirandah”. Total semua *image* yang relevan dengan *keyword* tersebut yang ada di *database* adalah 4. Dari pencarian normal dihasilkan 4 *image* dengan 4 *image* relevan. Sedang pada *query* pencarian lanjut dengan tanpa *copyright*, menghasilkan 1 *image* dengan 1 *image* relevan. Total *image* relevan di *database* untuk *query* ini adalah 1.

Tabel -1 Hasil Pengujian Pencarian

Kasus	Mode	Ditemukan	Image Relevan	Recall	Precision
1	Normal	11	9	1	0,81
	Lanjut	3	3	0,75	1
2	Normal	45	41	0,75	0,91
	Lanjut	37	35	0,92	0,94
3	Normal	22	18	0,6	0,81
	Lanjut	5	4	1	0,8
4	Normal	4	4	1	1
	Lanjut	1	1	1	1

V. Kesimpulan

Kesimpulan yang dapat diambil dari makalah ini antara lain:

1. Data EXIF dapat diterapkan pada *image search engine* dengan metode pencarian yang menggunakan proses *filtering*. Proses *filtering* bertujuan menyaring hasil pencarian sehingga sesuai dengan parameter pencarian yang dimasukkan oleh *user*. Pada proses *filtering*, parameter yang dimasukkan oleh *user* akan dibandingkan dengan data EXIF yang dimiliki *image* tersebut.
2. Penggunaan data EXIF tidak memiliki pengaruh yang terpola terhadap perubahan nilai *precision* dan *recall*, sehingga belum dapat disimpulkan bahwa penggunaan data EXIF akan meningkatkan kinerja *image search engine* yang diukur menggunakan *precision* dan *recall*.
3. Keuntungan utama dari penggunaan data EXIF pada *image search engine* adalah memungkinkan *user* untuk membentuk *query* pencarian yang lebih spesifik.
4. Penggunaan data EXIF pada pencarian cenderung menurunkan jumlah *image* hasil pencarian. Hal ini disebabkan *query* yang digunakan oleh *user* akan lebih spesifik, serta banyaknya *image* yang ada di Internet yang tidak memiliki data EXIF.

VI. Daftar Pustaka

- [MUN01] Munson, Ethan V; Tsymbalenko, Yelena. *To Search for Images on the Web, Look at the Text, Then Look at the Images*. Department of EECS University of Wisconsin-Milwaukee, Milwaukee, 2001.
- [MAN08] Manning, Christopher D; Raghavan, Prabhakar; Schütze, Hinrich. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, 2008.
- [RIJ95] Rijsbergen, C. J. Van. *Information Retrieval*. Department of Computing Science University of Glasgow, Glasgow, 1995.
- [SCL95] Sclaroff, Stan. *World Wide Web Image Search Engines*. Image and Video Computing Group Computer Science Department Boston University, Boston, 1995.
- [FRA96] Frankel, Charles; Swain, Michael J.; Athitsos, Vassilis. *WebSeer: An Image Search Engine for the World Wide Web*. Computer Science Department, The University of Chicago, Chicago, 1996.