



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Metadata merupakan sumber daya yang memiliki peran penting terhadap fungsi dan interoperabilitas berbagai informasi digital. Jika suatu informasi digital memiliki metadata yang baik, maka hal ini berdampak pada kemudahan informasi tersebut untuk ditemukan[1]. Namun, sulit untuk mengintegrasikan metadata ke dalam sistem informasi, baik cara membuat, memperluas, mengelola, maupun menggunakannya[2]. Selain itu, standar metadata yang beragam tentu mempersulit ketika ingin digunakan secara bersama-sama dalam suatu lingkup tertentu[2]. Pembentukan metadata secara otomatis merupakan salah satu solusi untuk menghasilkan metadata dengan kualitas yang baik sehingga mampu membantu penemuan informasi yang relevan. Otomatisasi tersebut juga mampu meningkatkan konsistensi serta efisiensi dalam pembentukan metadata dari pada proses manual[3].

Marko A.R[1] telah melakukan penelitian tentang pembentukan metadata secara otomatis pada dokumen *video*, *image*, dan *music*. Selain itu Darmawan Fatrionanda[3] juga melakukan penelitian yang sama tetapi dengan penerapan dokumen yang berbeda yaitu pada dokumen web. Kedua penelitian tersebut membuat *associative network* dengan metode *occurrence associative network* dan *cooccurrence associative network*. Hasil penelitian menyimpulkan bahwa *cooccurrence* memberikan hasil lebih baik dari pada *occurrence*. Setelah *associative network* terbentuk, dilakukan *sharing* metadata menggunakan salah satu teknik propagasi yaitu algoritma *particle swarm*. Namun, keduanya menganggap bahwa algoritma *particle swarm* memiliki tingkat komputasi tinggi sehingga memerlukan penelitian lebih lanjut dengan metode lain dalam propagasi metadata. Sedangkan kualitas rekomendasi metadata dapat dilihat dari nilai *precision* dan *recall*, di mana nilai *precision* dituntut agar

lebih tinggi dari pada nilai *recall*, tanpa menurunkan kualitas dari aspek produktivitas dalam menghasilkan metadata[1][14][15].

Heung-Nam Kim[11] telah mengembangkan pembentukan *associative network* dengan metode baru yang diberi nama *Associative Face Cooccurrence Network* (FCON). Metode ini diimplementasikan dalam sistem rekomendasi hubungan pertemanan di jejaring sosial berdasarkan *name tag* yang terdapat pada sekumpulan foto di album seorang pemilik akun jejaring sosial. Hasil penelitian tersebut menunjukkan bahwa nilai *precision* semakin menurun, sedangkan nilai *recall* semakin naik dalam setiap kali percobaan yang dilakukan. Namun, nilai rata-rata *precision* mencapai angka yang lebih tinggi dari pada nilai rata-rata *recall*. Hasil akhir dalam penelitian ini selaras dengan kualitas rekomendasi metadata yang baik.

Jane Greenberg[6], telah melakukan penelitian terhadap pengaruh proses *harvesting* dalam pembuatan metadata yang terotomatisasi. Hasil penelitian tersebut menunjukkan bahwa *harvesting* dari *tag* meta yang dibuat oleh manusia, memiliki dampak positif untuk menciptakan metadata yang optimal. Selain itu, Mangala Hirwade[19] juga menggunakan metode *harvesting* untuk mengatur keterkaitan sumber daya, berdasarkan metadata yang dimiliki oleh dokumen digital. Penelitiannya berhasil menciptakan *Metadata Harvesting Services* di India, di mana aplikasi tersebut mampu menciptakan keteraturan web serta mengontrol keterhubungan antar dokumen digital yang jumlahnya semakin meningkat dan meskipun muncul dalam format yang beragam. Selain kedua penelitian tersebut, Markus Hanse[30] juga melakukan penelitian tentang metode *harvesting* terhadap metadata yang dimiliki oleh berbagai *research paper* dari situs web konferensi ilmiah. Hasil penelitian tersebut mampu mencapai *accuracy*, *precision*, dan *recall* lebih dari 95%.

Daniel Parrott[20] telah melakukan penelitian guna mengoptimalkan algoritma *Particle Swarm Optimization* (PSO) yaitu dengan *Species-Based Particle Swarm Optimization*

(SPSO). Selain itu, Emad Elbeltagi[21] telah membandingkan algoritma Genetic, Memetic, PSO dan *Shuffled Frog Leaping* (SFL). Hasil penelitiannya mengungkapkan bahwa SFL memiliki waktu komputasi yang paling rendah. Pada algoritma SPSO dan SFL terdapat beberapa langkah dalam setiap iterasi yang dilakukan, salah satunya adalah melakukan *sorting* terhadap *fitness* secara *descending*[20][21].

Metode *sorting* dapat dilakukan dengan berbagai jenis algoritma, seperti *bubble sort*, *selection sort*, *insertion sort*, *quicksort*, *mergesort*, *heapsort*, *shellsort*, *radix sort*, dan *external sorting*[17]. Namun, pada umumnya *quicksort* adalah metode *sorting* yang paling cepat[23]. Selain itu, Philippas[22] menyebutkan bahwa *quicksort* tidak hanya algoritma yang paling cepat untuk melakukan *sorting* terhadap data yang ukurannya besar, tetapi juga memiliki kinerja *cache* yang baik serta sederhana untuk diimplementasikan ke dalam aplikasi.

Salah satu kelebihan dari *web base application* adalah kemampuannya yang dapat diakses oleh berbagai *platform* seperti windows, linux, dan MacOS[26]. Selain itu, aplikasi tidak menggunakan sumber daya pada *client* terlalu besar karena beban dipangku oleh *server*[24]. Kelebihan yang lain adalah kemudahan dalam melakukan perawatan aplikasi karena lokasinya yang terpusat pada *server*. Hal tersebut juga berdampak pada biaya perawatan yang dapat diminimalisir[24].

1.2 Batasan Masalah

Batasan masalah dalam penelitian ini:

- a. Objek yang digunakan untuk membentuk FCON bukan sekumpulan foto yang dilengkapi dengan *name tag*, melainkan sekumpulan dokumen web yang dilengkapi dengan *keywords*.

- b. *Input* aplikasi adalah dokumen XML yang diperoleh dari *tools* proximity, sedangkan dokumen fisik HTML dapat diunduh di <http://www-2/cs.cmu.edu/~WebKB/>,
- c. Skema metadata yang digunakan adalah Dublin Core karena memiliki deskripsi yang sederhana, arti kata yang mudah dikenali, bersifat *expandable*, dan banyak digunakan oleh dokumen web. Sedangkan properti metadata yang dibuat secara otomatis adalah properti *keywords*,
- d. Jumlah maksimal metadata yang dianggap miskin, jumlah maksimal metadata yang direkomendasikan oleh aplikasi, dan minimum nilai dari kualitas *link* antar metadata, ditentukan oleh user saat sebelum memulai *training*,
- e. Tipe dokumen web dan dokumen multimedia yang dapat direkomendasikan oleh aplikasi ini bergantung kepada *file input* XML. Namun, *file input* XML pada penelitian ini hanya berfokus kepada dokumen web bertipe *.html,
- f. Proses validasi dilakukan secara manual oleh manusia sebagai pengguna aplikasi, dan
- g. Proses penyebaran metadata dilakukan di dalam lingkup *database*.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini:

- a. Apakah proses yang dilakukan sebelum membangun *associative network* guna memberi pengaruh positif terhadap *variable* uji?
- b. Bagaimanakah melakukan pembentukan metadata secara otomatis dengan mengimplementasikan metode FCON?
- c. Bagaimana cara yang dapat dilakukan untuk menyebarkan metadata?

- d. Bagaimanakah nilai *precision* dan *recall* yang dihasilkan dalam penelitian ini apabila dibandingkan dengan nilai *precision* dan *recall* yang telah dilakukan oleh Darmawan Fatriananda[3]?

1.4 Tujuan Penelitian

Tujuan penelitian ini:

- a. Mengimplementasikan metode *harvesting* sebelum dilakukan pembuatan *associative network* sehingga mampu memberi pengaruh positif terhadap *variable* uji,
- b. Mengembangkan aplikasi berbasis web guna membentuk metadata secara otomatis dengan mengimplementasikan metode FCON,
- c. Mengimplementasikan metode *sorting* yaitu dengan algoritma *quicksort* guna melakukan propagasi metadata, dan
- d. Membandingkan nilai *precision* dan *recall* yang dihasilkan dalam penelitian ini dengan hasil penelitian yang dilakukan Darmawan Fatriananda[3].

1.5 Manfaat Penelitian

Penelitian ini bermanfaat untuk menghasilkan metadata pada dokumen web secara otomatis dengan kualitas yang baik, yaitu nilai *precision* lebih tinggi dari pada nilai *recall*.

