

# 1. Pendahuluan

## 1.1 Latar Belakang

Seiring dengan berkembangnya Teknologi Informasi dan ilmu pengetahuan, maka informasi yang terdapat di Internet pun semakin banyak dan luas. Dengan banyaknya informasi yang tersedia di Internet, informasi apapun yang dibutuhkan bisa dicari di Internet. Dengan bantuan *search engine*, informasi yang dibutuhkan akan didapat, hanya dengan cara mengetikkan suatu *keyword*. Namun seringkali informasi yang didapat tidak relevan dengan apa yang dicari.

Informasi yang didapat dari Internet biasanya berupa dokumen web yang memiliki elemen-elemen teks yang dapat diproses lebih lanjut untuk membantu penemuan dokumen yang relevan. Suatu dokumen web memiliki metadata berupa teks yang terstruktur, sedangkan di media lain metadata bisa berupa suara atau gambar. Dalam dokumen web, metadata ini membantu dalam menentukan dokumen web yang relevan. Meskipun memiliki nilai yang sangat penting, metadata pada umumnya diabaikan dan walaupun ada seringkali isinya tidak lengkap [4], karena menemukan, membuat atau memelihara metadata pada suatu halaman web secara manual ini terbilang sulit [3]. Hal ini tentu saja menghambat efektivitas pelayanan atau pemberian informasi yang relevan.

Pembentukan metadata secara otomatis dapat membantu penemuan informasi yang relevan. Selain itu bagi pembuat dokumen web, pembentukan metadata secara otomatis dalam suatu dokumen web ini menjadi lebih efisien dan lebih konsisten daripada proses manual (yang berorientasi pada manusia) [5]. Penelitian menunjukkan bahwa pembentukan metadata secara otomatis dapat menghasilkan metadata yang dapat diterima. Namun, pembentukan metadata ini tetap memerlukan pertimbangan intelektual manusia. Para peneliti telah menyimpulkan bahwa pembentukan metadata yang paling efektif adalah mempertimbangkan penilaian manusia maupun metode otomatis [6].

Teknik yang dilakukan dalam pembentukan metadata secara otomatis ini adalah dengan melakukan ekstraksi metadata dan isi dari dokumen. Teknik ekstraksi yang berdasarkan isi(*content*) ini kurang efisien untuk jenis dokumen multimedia (audio, video dan gambar) [16]. Selain itu, proses ekstraksi yang berdasar pada isi(*content*) ini ternyata menghabiskan biaya komputasi yang sangat besar [11].

Dalam suatu situs di Internet tentu ada dokumen yang saling berhubungan, contohnya dokumen-dokumen yang terhubung melalui *hyperlink* (syntax *href* dalam HTML) pada suatu halaman web. Contoh lain, dalam situs detik.com keterhubungan suatu artikel dapat dili hat dari data tentang penulis artikel tersebut. Dokumen-dokumen web yang saling berhubungan ini dapat berbagi(*shared*) metadata dari dokumen web yang memiliki metadata lengkap dapat berbagi ke dokumen web yang memiliki metadata kurang atau tidak lengkap. Teknik berbagi(*shared*) metadata ini merupakan teknik ekstraksi dengan pendekatan ekstraksi berdasar pada konten yang dapat membantu mengurai biaya komputasi [13].

Dalam Tugas Akhir ini dibangun aplikasi untuk membentuk metadata secara otomatis dengan menggunakan metode *associative network*. Metode *associative network* digunakan untuk menentukan keterhubungan antar dokumen. Terdapat dua algoritma yang digunakan untuk membuat *associative network*, yaitu *occurrence associative network* dan *coocurrence associative network*. Algoritma *occurrence associative network* melakukan pendekatan pada kutipan (biasanya berupa *hyperlink*) yang ada pada dokumen web, sedangkan *coocurrence associative network* melakukan pendekatan pada nilai properti-properti pada metadata seperti *author*, *keyword*, *identifier*, *description*, *publisher*, *format* dan tanggal *publish*. Setelah ditentukan dokumen yang memiliki keterhubungan, kemudian dilakukan propagasi metadata dengan algoritma *particle swarm* untuk proses *shared* (berbagi) metadata dari dokumen web yang memiliki metadata lengkap ke dokumen web yang metadatanya kurang atau tidak

lengkap[14]. Teknik propagasi metadata menggunakan algoritma *particle swarm* dipilih karena dalam proses komputasi algoritma *particle swarm* akan lebih unggul daripada algoritma sejenis seperti algoritma genetik, karena pada algoritma *particle swarm* tidak ada proses evolusi (mutasi dan *crossover*) [15]. Proses selanjutnya adalah mekanisme *filter* (penyaringan) terhadap energi dari metadata yang direkomendasikan, proses ini bersifat *optional*, karena untuk kebutuhan tertentu proses *filter* bisa tidak dilakukan.

### 1.2 Perumusan Masalah

Perumusan masalah yang diambil dalam penelitian Tugas Akhir ini adalah bagaimana mengatasi kesulitan untuk membuat, menemukan atau melengkapi metadata secara manual. Pembentukan metadata secara otomatis dengan metode *associative network* menggunakan algoritma *occurrence* dan *cooccurrence* dipilih untuk membantu permasalahan yang ada. Selain proses ekstraksi dilakukan juga proses propagasi menggunakan algoritma *particle swarm*. Hasil dari pembentukan metadata secara otomatis ini harus menginformasikan isi dari dokumen. Keakuratan dari hasil pembentukan metadata ini yang kemudian dievaluasi.

### 1.3 Batasan Masalah

Implementasi Tugas Akhir ini dibatasi oleh beberapa hal, sebagai berikut :

1. Data yang digunakan untuk membentuk metadata bentuk fisiknya adalah halaman web.
2. *Resource* metadata menggunakan data yang berasal dari lingkungan *Knowledge Discover Labotary* berupa file xml yang didapat dari <http://kdl.cs.umass.edu/proximity/proximity.html>, kemudian di-*export* ke mysql sebagai database. Jumlah data tersebut sebanyak 4135 data. Untuk file fisik dokumen html dapat di unduh di <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/>.
3. Properti-properti metadata yang digunakan dalam penerapan algoritma *coocurrence associative network* adalah *keyword*. *Keyword* dipilih karena merupakan properti yang sering digunakan. Properti pada dataset yang digunakan merupakan properti yang didefinisikan oleh suatu institusi yang bernama *Dublin Core*.

### 1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah

1. Membangun suatu aplikasi untuk membentuk metadata secara otomatis dengan metode *associative network* dengan algoritma *occurrence associative network* dan *cooccurrence associative network* untuk teknik ekstraksi. Kemudian mengimplementasikan teknik propagasi metadata dengan menerapkan algoritma *particle swarm* yang berfungsi menyebarkan properti metadata.
2. Evaluasi performansi algoritma *occurrence associative network* dan *cooccurrence associative network* yang dijadikan sumber pada teknik propagasi metadata (menyebarkan properti metadata) dengan algoritma *particle swarm* menggunakan nilai *precision*, *recall* dan *F-score*.

### 1.5 Metode Penyelesaian Masalah

#### a. Studi literatur

Studi Literatur dengan mempelajari literatur-literatur yang relevan dengan permasalahan yang meliputi : melakukan studi pustaka dan referensi mengenai *automatic metadata generation*, *associative network*, *occurance associative network*, *coocurrence associative network*, algoritma *particle swarm*, library Java untuk algoritma *particle swarm* (JSwarm), *Dublin Core metadata element set*.

b. Analisis Perancangan

Pada tahap ini dilakukan analisis, pengkajian dan pemahaman terhadap tahapan dan cara kerja *occurrence associative network*, *cooccurrence associative network* serta algoritma *particle swarm* untuk *metadata generation*. Dari dua algoritma ekstraksi (*occurrence* dan *cooccurrence*) dianalisis mana yang menghasilkan rekomendasi metadata yang lebih baik.

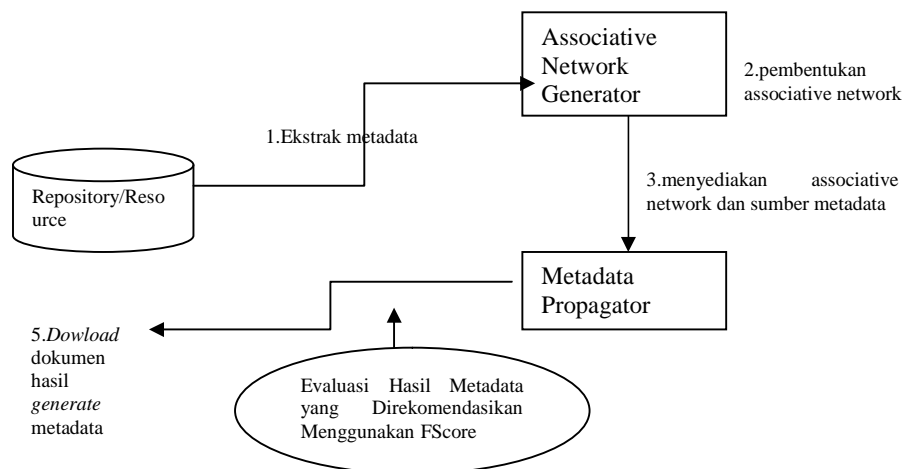
c. Implementasi

Melakukan pembangunan aplikasi untuk implementasi metode yang digunakan pada sistem pembentukan metadata secara otomatis. Aplikasi yang dibuat bersifat web-based. Aplikasi dapat meng-*insert* metadata yang direkomendasikan ke dokumen web yang tidak memiliki atau tidak lengkap metadata.

d. Testing dan Analisa Hasil

Pada tahap testing dan analisis hasil ini Implementasi algoritma akan dilakukan berdasarkan hasil analisis dan perancangan algoritma pada tahap sebelumnya. Pengujian algoritma akan dilakukan dengan menggunakan input berupa koleksi dokumen yang akan dilakukan proses pembuatan metadata secara otomatis kemudian dilakukan analisis hasil berdasar parameter yang telah ditentukan. Pengujian akan didasarkan pada 2 parameter, yaitu :

- a. Hasil dari pembentukan metadata secara otomatis ini dibandingkan dengan metadata asli yang ada di *resource*.
- b. Membentuk beberapa *associative network* menggunakan algoritma *occurrence* dan *cooccurrence* (properti metadata yang akan digunakan *author, date publish* dan *keyword*). Kemudian dari *associative network* yang tersebut dilakukan propagasi metadata dengan menyebarkan (*propagating*) properti metadata. Evaluasi dilakukan dengan menghitung *F-score*.



Gambar 1.1 Deskripsi Sistem