



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 II

LANDASAN TEORI

2.1 Dublin Core (DC) sebagai Skema Metadata

Dublin Core (DC) merupakan salah satu skema metadata yang digunakan untuk *web resource description and discover*. Skema ini dihadirkan dengan tujuan untuk menyederhanakan kaidah dalam *Machine Readable Cataloguing* (MARC) yang juga merupakan skema metadata. MARC dirasa terlalu banyak memiliki unsur yang mengatur di dalamnya serta istilah-istilah yang digunakan tergolong sulit dimengerti oleh pihak yang tidak memiliki pengetahuan tentang ilmu pustaka[9][10].

DC dan MARC banyak digunakan di lingkungan perpustakaan. Namun, pada praktisnya penggunaan DC lebih banyak digunakan untuk program-program perpustakaan digital, di mana kesederhanaan unsur menjadi pertimbangan utamanya. Selain itu, unsur-unsur yang diadopsi oleh DC dapat diimplementasikan untuk kegiatan *sharing* metadata. Hal tersebut didukung dengan karakteristik DC yang memiliki kesamaan kategori kode, kesederhanaan deskripsi, semantik atau arti kata yang mudah dikenali secara umum, dan *expandable* atau memiliki potensi untuk dikembangkan lebih lanjut. Sedangkan elemen yang diterapkan oleh DC, secara keseluruhan ada lima belas, yaitu *contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject/keywords, title, dan type*[10]. Sedangkan salah satu contoh penerapan struktur DC dapat dilihat pada Gambar 2.1.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

```

<dc>
  <title>Tasamuh, bersama Prof. Dr. Alwi Shihab</title>
  <title>
    <alternative>Tasamuh, with Prof. Dr. Alwi Shihab</alternative>
  </title>
  <creator>Alwi Shihab, Prof. Dr.</creator>
  <publisher>JBPTITBPP</publisher>
  <subject>
    <keywords>eksklusivisme, inklusivisme, pluralisme</keywords>
  </subject>
  <description>Dalam setiap agama akan selalu ada kelompok-kelompok
    pemikiran dalam memandang agamanya maupun agama mitra dialognya ....
  </description>
  <contributor>NARATOR : Evi Yuliani; digitized by Indro</contributor>
  <date>1997-07-20</date>
  <date>
    <modified>2000-09-26 ::00</modified>
  </date>
  <type>dlearn</type>
  <type>
    <schema>general</schema>
  </type>
  <identifier>itb-dist-salma-1997-Alwi-tasa</identifier>
  <identifier>
    <hierarchy>/Distance_Learning/Religion/Salman_Mosque/
      Ceramah_Umum/</hierarchy>
  </identifier>
  <source>Studium Generale</source>
  <language>Bahasa Indonesia</language>
  <relation>
    <count>1</count>
  </relation>
  <relation>
    <no>1</no>
    <datemodified>2003-04-15 11:08:12</datemodified>
    <haspart>itb-dist-salma-1997-Alwi-tasa-part2.RM</haspart>
    <haspath>disk1/1/itb-dist-salma-1997-Alwi-tasa-part2.RM</haspath>
    <hasfilename>itb-dist-salma-1997-Alwi-tasa-part2.RM</hasfilename>
    <hasformat>application/vnd.rn-realmedia</hasformat>
    <hassize>7726181</hassize>
    <hasuri>/download.php?
      f=/disk1/1/itb-dist-salma-1997-Alwi-tasa-part2.RM</hasuri>
  </relation>
</dc>

```

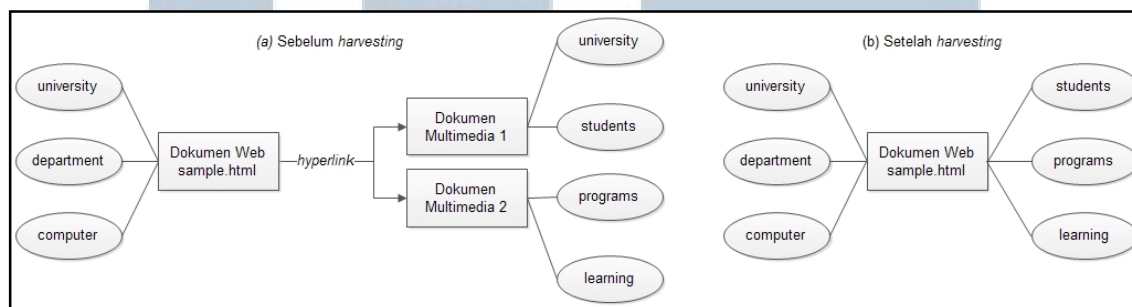
Gambar 2.1 Metadata dalam Format XML Standar Dublin Core[27]

2.2 Harvesting

Harvesting digunakan untuk menghubungkan sumber daya yang berbeda jenis[19]. Proses tersebut terjadi ketika metadata dikumpulkan dari seluruh dokumen digital (seperti: *.doc, dan *.pdf), di mana dokumen tersebut memiliki hubungan *occurrence* dengan dokumen web yang menjadi tempat berkumpulnya metadata tersebut[6]. Arti dari hubungan *occurrence* adalah hubungan langsung diantara dua dokumen yang terlihat dengan adanya

kutipan, umumnya seperti *hyperlink*. Sedangkan hubungan *cooccurrence* merupakan hubungan langsung antar dokumen, seperti kesamaan metadata[3].

Proses *harvesting* bergantung pada metadata yang dihasilkan oleh manusia atau oleh *full/semi-automatic* yang didukung perangkat lunak[6]. Pada penelitian ini, *harvesting* dilakukan sebelum pembentukan *associative network* terjadi. Sedangkan ilustrasi proses *harvesting* dapat dilihat pada Gambar 2.2.



Gambar 2.2 Contoh Proses Harvesting

2.3 Associative Face Cooccurrence Network

Associative Face Cooccurrence Network (FCON) merupakan metode yang dipublikasikan pada tahun 2010, sehingga sulit untuk mencari sumber literatur yang berkaitan dengan metode tersebut. Salah satu jenis metode pembentukan *associative network* ini dikembangkan pertama kali dengan tujuan untuk memberikan rekomendasi hubungan pertemanan dalam jejaring sosial. Metode ini membangun *associative network* dengan memanfaatkan foto pribadi, di mana setiap orang yang berada di dalam foto merupakan *node* atau *vertex* yang saling berhubungan satu sama lain[11]. Namun, dalam penelitian ini, yang berperan sebagai foto adalah dokumen web. Sedangkan setiap orang yang berada di dalam foto adalah setiap *keyword* yang berada di dalam dokumen web.

FCON merupakan jenis *graph* yang bersifat *undirected*, memiliki bobot, dan tidak ada perulangan ke diri sendiri. Secara formal, FCON direpresentasikan sebagai $N = (V, E)$, E adalah himpunan semua elemen dua himpunan bagian dari $V: E \subseteq V \times V$. Kemudian, $e(v_i, v_j)$ adalah *edge* di antara dua *vertex* v_i dan v_j . Bobot pada setiap *edge* merepresentasikan kekuatan *associative*, yang diperoleh dari frekuensi *cooccurrence* dari metadata di dalam dokumen web[11].

Pembangunan FCON dimulai dengan melakukan iterasi kepada setiap foto yang terdapat pada sebuah album. Kemudian pada tiap foto dilakukan iterasi pula terhadap *name tag* yang terdapat pada foto tersebut. Sebagai contoh, terdapat delapan foto yang dilengkapi *name tag* di album milik Brandon, seperti yang ditunjukkan pada Tabel 2.1[11].

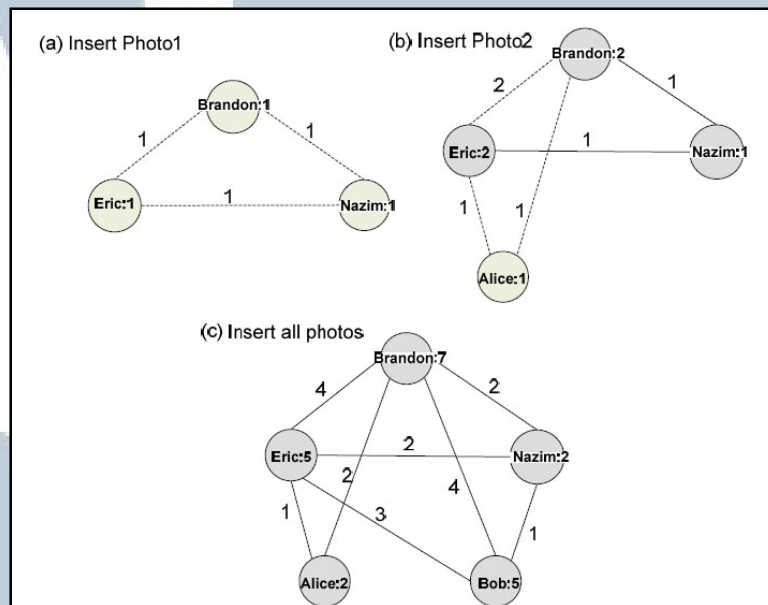
Tabel 2.1 Contoh Album Foto Brandon[11]

ID Foto	Wajah pada Foto
Foto1	Brandon, Eric, Nazim
Foto2	Alice, Brandon, Eric
Foto3	Bob, Brandon, Eric
Foto4	Alice, Brandon
Foto5	Bob, Eric
Foto6	Bob, Brandon
Foto7	Bob, Brandon, Eric, Nazim
Foto8	Bob, Brandon

Pertama, jika tidak ada sebuah FCON untuk Brandon, maka FCON baru harus dibuat terlebih dahulu. Untuk foto dengan id Foto1, *name tag* Brandon, Eric, Nazim, ditransformasikan ke dalam FCON, di mana Brandon dihubungkan dengan Eric, Eric dihubungkan dengan Nazim, dan Nazim dihubungkan dengan Brandon (*fully connected graph*). Kemudian, *attribute* setiap *edge* dan *vertex* yang baru terbentuk diinisialisasi dengan

nilai 1. Gambar 2.3-(a) mengilustrasikan FCON setelah menambahkan *name tag* pada Foto1[11].

Pada foto dengan id Foto2, karena *name tag* Brandon dan Eric telah terdapat pada FCON, jadi bobot edge $e(\text{Brandon}, \text{Eric})$ ditambah sebanyak 1, yaitu menjadi 2. Sedangkan *name tag* Alice dibuatkan *vertex* baru yang dihubungkan dengan *vertex* Brandon dan Eric, yang masing-masing *attribute* diinisialisasi dengan nilai 1, seperti yang diilustrasikan pada Gambar 2.3-(b). Setelah semua foto dimasukkan ke dalam FCON, hasil akhir dari pembentukan FCON pribadi milik Brandon dapat dilihat pada Gambar 2.3-(c)[11].



Gambar 2.3 Contoh Pembangunan FCON untuk Brandon[11]

2.4 Quicksort

Dokumen web yang dianggap miskin adalah dokumen web yang memiliki metadata tidak lengkap atau dengan kata lain suatu metadata tidak mampu merepresentasikan isi dokumen secara menyeluruh sehingga informasi yang ada di dalamnya sulit untuk ditemukan[3]. Pada setiap dokumen web yang tergolong miskin, dilakukan iterasi terhadap

elemen *keywords* untuk mencari *vertex* yang memiliki hubungan *cooccurrence* dengan setiap elemennya. Selanjutnya *vertex* sebanyak n dengan kekuatan *associative* (bobot *edge*) terbaik diambil sebagai rekomendasi metadata, selama kekuatan *associative* setiap *vertex* memenuhi kriteria. Oleh karena itu, metode *sorting* dilakukan untuk mempermudah pemilihan *vertex* berdasarkan nilai kekuatan *associative*[11]. Sedangkan algoritma *sorting* yang dipilih dalam penelitian ini adalah *quicksort*.

Quicksort adalah algoritma pengurutan sekuensial yang banyak diyakini sebagai yang tercepat. Algoritma ini merupakan algoritma rekursif yang menggunakan metode “*Divide and Conquer*” untuk menyortir semua *key*[22]. *Quicksort* bekerja dengan memilih sebuah elemen e dari *array*. Kemudian dengan cara menukar elemen secara berpasangan, *array* asli dipartisi menjadi dua bagian *subarray* yaitu kiri dan kanan, di mana pada akhir putaran *subarray* kiri $\leq e$ dan *subarray* kanan $\geq e$ [23]. Sedangkan kompleksitas algoritma pada *quicksort* adalah $O(n \log n)$ [17].

Langkah untuk menentukan posisi yang sebenarnya bagi elemen paling kiri ($\text{data}[\text{kiri}]$) adalah sebagai berikut[17]:

- a. Cari dari arah kiri ke kanan data yang nilainya tidak lebih kecil dari pada $\text{data}[\text{kiri}]$, misalkan data tersebut $\text{data}[\text{j}]$,
- b. Cari dari arah kanan ke kiri data yang nilainya tidak lebih besar dari pada $\text{data}[\text{kiri}]$, misalkan data tersebut $\text{data}[\text{k}]$,
- c. Apabila $\text{j} < \text{k}$ maka tukar $\text{data}[\text{j}]$ dan $\text{data}[\text{k}]$, ulangi kembali langkah pertama sampai dengan ketiga,
- d. Tukar $\text{data}[\text{kiri}]$ dan $\text{data}[\text{k}]$, $\text{data}[\text{k}]$ telah berada pada posisinya yang tepat, dan
- e. Ulangi langkah pertama sampai dengan keempat untuk data partisi kiri ($\text{data}[\text{kiri}]$ sampai dengan $\text{data}[\text{k}-1]$) dan data partisi kanan ($\text{data}[\text{k}+1]$ sampai dengan $\text{data}[\text{kanan}]$), proses pengulangan ini berlangsung selama $\text{kiri} < \text{kanan}$.

2.5 Evaluasi Menggunakan Precision dan Recall

Hasil uji coba dievaluasi menggunakan nilai *precision* dan *recall* dari setiap *resource* metadata. *Precision* merupakan jumlah nilai properti metadata yang diterima satu *resource* yang relevan terhadap nilai properti yang direkomendasikan secara keseluruhan (Persamaan 1)[3][15].

$$Pr(\mu) = \frac{|meta(n_i, \mu) \cap rec(n_i, \mu)|}{|rec(n_i, \mu)|} \quad \text{Persamaan (1)}$$

Sedangkan *recall* (Persamaan 2) didefinisikan sebagai proporsi dari nilai properti relevan yang diterima, dengan kemungkinan jumlah total nilai properti yang relevan. Sebagai contoh, jika sebelumnya *resource* n_i memiliki nilai properti *keywords* = {swarm} dan nilai properti yang direkomendasikan pun sama yaitu *keywords* = {swarm}, maka nilai *recall* adalah 100%. Di sisi lain, jika sebelumnya *resource* n_i memiliki nilai properti *keywords* = {swarm, network} dan nilai properti *keywords* yang direkomendasikan hanya {swarm}, maka nilai *recall* adalah 50%. Sedangkan dalam kedua kasus tersebut *precision* memiliki nilai sebesar 100%[1][15].

$$Re(\mu) = \frac{|meta(n_i, \mu) \cap rec(n_i, \mu)|}{|meta(n_i, \mu)|} \quad \text{Persamaan (2)}$$

Precision dan *recall* cenderung memiliki keterkaitan yang berbanding terbalik, $Pr \approx \frac{1}{Re}$. Hubungan tersebut dapat dimengerti dengan baik ketika sedang memeriksa kasus yang dinilai cukup ekstrim. Apabila pada setiap kemungkinan nilai properti menyediakan *resources* ($|rec(n_i, \mu)| \rightarrow \infty$), serta pada awalnya *resources* tersebut dinilai hanya memiliki satu nilai properti ($|meta(n_i, \mu)| = 1$), maka *recall* menjadi 100% sementara *precision* mendekati 0%.

Pada relasi ekstrim yang berlawanan, apabila *resource* sebelumnya memiliki setiap kemungkinan nilai properti pada original metadata ($|\text{meta}(\eta_i, \mu)| \rightarrow \infty$) dan direkomendasikan hanya satu nilai properti ($|\text{rec}(\eta_i, \mu)| = 1$), maka *precision* menjadi 100%, tetapi *recall* mendekati 0% [1].

2.6 Data Uji Banding

Pengujian dilakukan terhadap metode FCON dengan mengukur variable *recall*, dan *precision*. Hasil data uji coba dibandingkan dengan hasil penelitian (Tabel 2.2) yang telah dilakukan oleh Darmawan Fatriananda[3]. Penelitian tersebut menggunakan 2487 node/dokumen data training dan 500 node data testing, di mana hasil penelitian setelah dilakukan training diperoleh rata-rata *precision* sebesar 0.6108, dan rata-rata *recall* sebesar 0.79025. Sedangkan dokumen yang dapat di-generate metadatanya sebanyak 50%.

Tabel 2.2 Ringkasan Hasil Pengujian Associative Network[3]

Jenis <i>Associative Network</i>	Penyaringan Energi	Generate Metadata	Recall		Precision		F-score	
			Avg	Max	Avg	Max	Avg	Max
<i>Occurrence</i>	Tidak	46%	0.57068	1	0.32823	1	0.3255	1
<i>Occurrence</i>	Ya	29%	0.3927431	1	0.45523	1	0.34856	1
<i>Cooccurrence</i>	Tidak	50%	0.79025	1	0.6108	1	0.64885	1
<i>Cooccurrence</i>	Ya	47%	0.5571131	1	0.76154	1	0.60784	1

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A