

BAB III LANDASAN TEORI

3.1 Metadata

Metadata adalah sebuah informasi yang secara struktur dapat menjelaskan, menemukan, atau menjadikan suatu informasi mudah untuk ditemukan kembali, digunakan, atau dikelola. Metadata sering disebut sebagai data tentang data atau informasi tentang informasi. Metadata ini mengandung informasi mengenai isi dari suatu data yang dipakai untuk keperluan manajemen file/data itu nantinya dalam suatu basis data.

Menurut (Subli, Sugiantoro and Prayudi, 2017) Metadata adalah informasi yang ditanam pada sebuah file yang isinya berupa penjelasan tentang file tersebut. Metadata ini mengandung informasi mengenai isi dari suatu data yang dipakai untuk keperluan manajemen file atau data itu nantinya dalam suatu basis data.

(Ii Sopiandi, 2017) pada penelitiannya menjelaskan bahwa metadata Metadata adalah informasi terstruktur yang mendeskripsikan, menjelaskan, menemukan atau setidaknya menjadikan suatu informasi mudah untuk ditemukan kembali, digunakan, atau dikelola. Metadata sering disebut sebagai data tentang data atau informasi tentang informasi. Metadata ini mengandung informasi mengenai isi dari suatu data yang dipakai untuk keperluan manajemen file/data itu nantinya dalam suatu basis data.

Metadata sangat diperlukan didalam proses indentifikasi digital yang berkaitan dengan sebuah file , dengan menggali informasi metadata pada sebuah file ini tentu akan memudahkan menarik benang merah pada sebuah kasus digital forensik. Serta dapat melihat rentetan suatu kejadian, kapan dirubah, apakah asli ataupun apakah file tersebut telah mengalami proses perubahan secara digital.

3.1.1 Jenis – Jenis Metadata

Berikut adalah jenis-jenis dalam metadata :

1. **Metadata Deskriptif:** menggambarkan sumber daya untuk tujuan seperti penemuan dan identifikasi. Ini dapat mencakup elemen-elemen seperti judul, abstrak, penulis, dan kata kunci.

2. **Metadata Struktural:** menunjukkan bagaimana objek majemuk disatukan, misalnya, bagaimana halaman diurutkan untuk membentuk bagian/bab .

16

3. **Metadata Administratif:** menyediakan informasi untuk membantu mengelola sumber daya, seperti kapan dan bagaimana sumber daya itu dibuat, jenis file dan informasi teknis lainnya, dan siapa yang dapat mengaksesnya.

4. **Metadata Pemeliharaan :** menawarkan informasi yang dapat memperkuat seluruh prosedur pemeliharaan objek/file digital tertentu. Informasi ini dapat mencakup detail penting yang diperlukan sistem untuk berkomunikasi atau berinteraksi dengan file tertentu.

5. **Metadata asal :** Dengan metadata asal, ini berarti paling relevan ketika sesuatu berubah atau sering diduplikasi. Sekarang, untuk objek fisik ini kurang penting, karena kita tidak sering menggandakannya. Namun, di dunia digital, ini terjadi sepanjang waktu.

Metadata dapat mendukung manajemen objek digital dengan menyediakan informasi yang dibutuhkan untuk mengekstrak konten digital secara spesifik , dengan seperti itu metadata mendukung navigasi di dalam bagian item, misalnya, dari satu halaman atau bagian ke halaman berikutnya, dan di antara berbagai versi objek, seperti berbagai resolusi gambar fotografi. Berikut tipe metadata dan contoh propertis menurut (Riley, 2017) :

Table 3.1 tipe metadata dan contoh propertis

Metadata Type	Example Properties
Descriptive metadata	Title Author Subject Genre Publication date
Technical metadata	File type File size Creation date/time Compression scheme

Preservation metadata	Checksum Preservation event
Rights metadata	Copyright status License terms Rights holder
Structural metadata	Sequence Place in hierarchy
Markup languages	Paragraph Heading List Name Date

3.1.2 Tipe Metadata

1. Machine readable cataloging (MARC)

MARC merupakan standar metadata *ANSI/NISO Z39.2 Information Interchange Format and ISO 2709 Information and documentation—Format for information exchange* yang dikembangkan pertama kali oleh Library of Congress. MARC adalah salah satu hasil dan juga sekaligus syarat penulisan katalog koleksi bahan pustaka perpustakaan. Format LC MARC sangat besar manfaatnya bagi penyebaran data katalogisasi bahan pustaka ke berbagai perpustakaan di Amerika Serikat. MARC merupakan bahasa metadata yang paling banyak digunakan di komunitas perpustakaan jauh sebelum XML dan RDF.

2. Dublin Core

Dublin Core Metadata Element Set (DCMES) berkembang dari pertemuan tahun 1995 di Dublin, Ohio, yang difokuskan pada metadata untuk informasi elektronik jaringan. Dublin Core dihadirkan karena ada beberapa pihak yang merasa kurang sesuai untuk menggunakan bentuk MARC sehingga diadakan suatu kesepakatan menyusun sebuah metadata baru yang lebih mudah dan fleksibel serta mempunyai kemampuan untuk dikembangkan dibanding MARC. Data Dublin Core tersusun atas 15 element dasar yaitu :

- Title : judul dari sumber informasi

- Creator : pencipta sumber informasi
- Subjet : pokok bahasan sumber informasi, biasanya dinyatakan dalam bentuk kata kunci atau nomor klasifikasi
- Description : keterangan suatu isi dari sumber informasi, misalnya berupa abstrak, daftar isi atau uraian
- Publisher : orang atau badan yang mempublikasikan sumber informasi
- Contributor : orang atau badan yang ikut menciptakan sumber informaso
- Date : tanggal penciptaan sumber informasi
- Type : jenis sumber informasi, laporan, peta dan sebagainya
- Format : bentuk fisik sumber informasi, format, ukuran, durasi, sumber informasi
- Identifier : nomor atau serangkaian angka dan huruf yang mengidentifikasi sumber informasi. Contoh URL, alamat situs
- Source : rujukan ke sumber asal suatu sumber informasi
- Language : bahasa yang intelektual yang digunakan sumber informasi
- Relation : hubungan antara satu sumber informasi dengan sumber informasi lainnya
- Coverage : cakupan isi ditinjau dari segi geografis atau periode waktu
- Rights : pemilik hak cipta sumber informasi

3. MODS

MODS adalah singkatan dari Metadata Object Description Schema. merupakan salah satu standar metadata yang dikembangkan oleh Library of Congress Network Development bekerjasama dengan MARC standar office, skema ini dikembangkan sebagai respon terhadap keluhan bahwa skema Dublin Core terlampaui sederhana untuk lingkungan perpustakaan, sedangkan format MARC 21 terlalu kompleks dan kurang bersahabat bagi pengguna di luar sistem perpustakaan.

4. EXchangeable Image File Format (Exif)

Struktur tag untuk metadata tertanam dalam file gambar digital. Format file TIFF dan JPEG mendukung Exif yang disematkan, tetapi tidak termasuk format file JPEG2000, PNG, dan GIF. Spesifikasi Exif mencakup elemen metadata seperti dimensi piksel, tanggal dan waktu diambil, pengaturan ISO, aperture, white balance, dan informasi tentang lensa yang digunakan.

3.2 Skema metadata

skema metadata mengandung 3 komponen yaitu semantik, isi dan sintaksis, sebagaimana uraian berikut:

1. Semantik (semantics) adalah kesepakatan mengenai makna bagaimana mengartikan suatu unsur sebuah skema metadata.
2. Isi (content) adalah isi dari unsur yang sudah diberi makna kesepakatan.
3. Sintaksis (syntax) dalam skema metadata dapat berarti sebagai machine readable (dapat dibaca mesin) atau dengan kata lain bahasa pemrograman. Sehingga semantic dan content yang telah dibuat dapat dibaca oleh mesin

3.3 Standar Metadata

Standar-standar yang digunakan dalam skema metadata, di antaranya :

1. CDWA (Categories for Descriptions of Works of Art): skema untuk deskripsi karya seni (http://www.getty.edu/research/publications/electronic_publications/cdwa/)
2. DCMES (Dublin Core Metadata Element Set): skema umum untuk deskripsi beraneka ragam sumber digital
3. EXIF (EXchangeable Image File Format) skema untuk metadata tertanam dalam file gambar digital. (http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf)
3. EAD (Encoded Archival Description): skema untuk menciptakan sarana temu kembali bahan kearsipan (archival finding aids) dalam bentuk elektronik.
4. GEM (Gateway to Educational Materials): skema untuk bahan pendidikan dan pengajaran

5. MPEG (Moving Pictures Experts Group) MPEG-7 dan MPEG-21: standar untuk rekaman audio dan video dalam bentuk digital
6. ONIX (Online Information Exchange), untuk data bibliografi lingkungan penerbit dan pedagang buku
7. TEI (Text Encoding Initiative): panduan untuk encoding teks dalam bentuk elektronik menggunakan SGML dan XML, khususnya untuk kalangan peneliti teks bidang humaniora.
8. VRA (Visual Resources Association) Core: skema untuk deskripsi karya visual dan representasinya. (<http://www.loc.gov/standards/vracore/>)
9. METS (Metadata Encoding and Transmission Standard): skema metadata untuk obyek digital kompleks yang tersimpan dalam koleksi perpustakaan
10. MARC (Machine Readable Cataloguing): skema yang digunakan di lingkungan perpustakaan sejak tahun 1960-an untuk membuat cantuman bibliografi elektronik standar. (http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf)
11. MODS (Metadata Object Description Standard): skema untuk deskripsi rinci sumber-sumber elektronik. (<http://www.loc.gov/standards/mods/>)

3.4 Pengertian Korelasi File

Menurut (Subli, Sugiantoro and Prayudi, 2017) korelasi dapat diartikan sebagai hubungan. Namun ketika dikembangkan lebih jauh, korelasi tidak hanya dapat dipahami sebatas pengertian tersebut. Korelasi merupakan salah satu teknik analisis dalam statistik yang digunakan untuk mencari hubungan antara dua variabel yang bersifat kuantitatif. Hubungan dua variabel tersebut dapat terjadi karena adanya hubungan sebab akibat atau dapat pula terjadi karena kebetulan saja. Secara sederhana korelasi ini bisa diartikan sebagai hubungan. Dua variabel dikatakan berkorelasi apabila perubahan pada variabel yang satu akan diikuti perubahan pada variabel yang lain secara teratur dengan arah yang sama (korelasi positif) atau berlawanan (korelasi negatif).

3.5 Fungsi Metadata

Berberapa fungsi dari metadata di antaranya :

1. Mengidentifikasi Sumber

Metadata dapat mengidentifikasi sumber yang mempunyai karakteristik unik dan bisa membedakan antara satu sumber dengan sumber yang lain sesuai dengan kriteria yang relevan.

2. Sebagai alat/tool

seperti melakukan monitoring kemajuan pelaksanaan pekerjaan pembangunan data spasial, mendokumentasikan data data yang ada (selesai dikerjakan), menginformasikan data data yang dimiliki untuk dapat dimanfaatkan oleh pihak lain dan melakukan estimasi rencana kerja pengumpulan data di kemudian hari

3. Tracking Penggunaan Sumber

metadata dapat melacak penggunaan sumber yang menghubungkan antara sumber yang satu ke sumber yang lain. Informasi pelacakan ini dikumpulkan secara otomatis oleh sistem dan mengaitkan sumber berdasarkan pelacakan perilaku pengguna sumber.

4. Mengambil Sumber

Metadata dapat membantu user dalam menemukan lokasi sumber informasi. Setelah menemukan sumber yang relevan, user diizinkan untuk mengambil sumber itu, asalkan memenuhi kriteria tertentu. Misalnya, user mencari koleksi fotografi. Setelah foto ditemukan, user bisa mengambil foto tersebut. beberapa tingkatan metadata yang sering digunakan, yaitu :

1. *Exploitation* metadata adalah metadata yang memuat informasi akses data, transfer data, load data, menginterpretasikan data dan penggunaan data untuk suatu aplikasi.
2. *Discovery* metadata adalah informasi minimum yang diberikan untuk menjelaskan isi dari sumber data. Jenis metadata ini tentu saja tidak memenuhi kategori metadata yang bisa diaplikasikan pada tingkat internasional.
3. *Exploration* metadata adalah informasi yang lebih detil yang diberikan dalam

menjelaskan isi dari sumber data. Jenis metadata ini diharapkan dapat membantu pengguna data untuk keperluan analisis

3.6 Kelemahan Metadata

Metadata juga sebagai informasi digital tidak lepas dari kekurangan , hal ini dikarenakan metadata tersebut juga disusun berdasarkan algoritma tertentu. Sehingga Kemungkinan metadata tersebut masih saja bisa dilakukan, untuk mengantisipasi hal tersebut biasanya dilakukan enkripsi didalam metadata tersebut agar dapat teridentifikasi apabila telah dilakukannya sebuah perubahan data.

Ada beberapa indikasi yang dapat dijadikan analisa didalam membantu sebuah file apakah telah melewati proses modifikasi yaitu menganalisa berdasarkan parameter berikut :

1. *LastModifiedTime* : adalah informasi metadata menunjukkan kapan item/file terakhir diubah dalam proses metadata atau juga file dibuat.
2. *lastAccessTime* : adalah kapan terakhir sebuah file dibuka atau dalam proses pengaksesan.
3. *creationTime* : adalah kapan terakhir dalam proses modifikasi (seperti kapan dilakukannya pemindahan data “*Copy File*” dan perubahan file tersebut.

Berdasarkan paramater tersebut untuk meminimalisir perubahan data , didalam aplikasi metadata forensik diberikan logika untuk membaca apakah tanggal *LastModifiedTime* setelah *creationTime*. Berikut hasil menggunakan metadata ekstraktor

1. hasil metadata ketika perubahan menggunakan aplikasi online <https://jimply.com/>



Gambar 3. 1 Hasil Pembacaan Metadata

2. hasil metadata ekstraktor ketika terjadi perubahan menggunakan aplikasi Photoshop



Gambar 3. 2 Hasil Pembacaan Metadata

3.7 Konsep Analissi Metadata Forensik

3.7.1 File

File merupakan sebuah kumpulan berbagai informasi yang saling berhubungan dan juga tersimpan di dalam ruang penyimpanan sekunder seperti hardisk, ssd maupun yang lainnya. File memiliki ekstensi informasi besar file seperti tanggal dan jam penyimpanan, ciri file, nama file, dan atribut tersebut. Ada beberapa jenis file diataranya :

1. File System

merupakan struktur logika yang digunakan untuk mengendalikan akses terhadap data yang ada pada disk, ekstensi dalam file sistem adalah sys, com, bak, bat, tmp, dan exe.

2. File Dokumen

Sebuah dokumen yang mengandung informasi tertentu & dapat dibuka dengan program tertentu, ekstensi pada file dokumen adalah doc, odt, xls, ods, pdf, ppt, txt.

3. File Video

Merupakan berkas komputer yang digunakan untuk menyimpan kumpulan berkas digital seperti video, audio, metadata, informasi, ekstensi pada file video adalah mpg, wmy, mp4, 3gp, avi, flv, kv

4. File Gambar

Merupakan standar atau standardisasi yang digunakan untuk mengatur dan menyimpan gambar dalam format digital, ekstensi file gambar adalah jpg, png, gif, tif, dan lainnya.

5. File Suara

Suatu bentuk jadi yang digunakan dalam menyimpan berkas audio pada sistem komputer, ekstensi file suara adalah wav, mp3, midi, dan rm.

Menurut (Homaidi, Informasi and Informasi, 2017) File merupakan koleksi data atau informasi terkait yang memiliki nama dan tersimpan dalam secondary storage. Biasanya sebuah file menggambarkan data atau program.

Jenis-jenis File

1. Text file, merupakan kumpulan dari karakter-karakter yang diolah sedemikian rupa diatur menjadi barisan dan bahkan mungkin halaman.

2. Source file, merupakan urutan dari berbagai subroutine dan fungsi yang kemudian diatur sebagai deklarasi-deklarasi yang diikuti oleh pernyataan-pernyataan yang dapat dieksekusi.

3. Object file, merupakan urutan dari byte-byte yang diatur menjadi sebuah blok-blok sehingga bisa dipahami oleh penghubung system.

4. Executable file, adalah kumpulan dari kode-kode yang dapat dibawa ke memori dan dijalankan oleh loader.

Sedangkan menurut (Priambodo, 2018) file merupakan kumpulan dari item data yang di atur di dalam suatu record, dimana item-item data tersebut dimanipulasi untuk proses tertentu.

3.7.2 Atribut File

Atribut file adalah suatu kondisi spesifik pada sebuah file atau folder, dimana pada sistem oprasi Windows memperlakukan secara khusus pada file atau folder tersebut. Atribut file digunakan oleh sistem operasi untuk memisahkan tipe-tipe file.

Beberapa atribut file :

1. Tipe File : Menentukan Tipe File
2. Jumlah link : Jumlah link untuk file tersebut.
3. Pemilik (Owner) : Menentukan siapa pemilik file tersebut.
4. Group : Menentukan group yang memiliki file tersebut.
5. Jumlah Karakter : Menentukan ukuran file dalam byte.
6. Waktu Pembuatan : Menentukan kapan file terakhir dimodifikasi.
7. Nama File : Menentukan nama file yang dimaksud.

Pada sistem operasi Windows ada beberapa atribut diantaranya :

1. Archive

File dengan atribut archive akan ditampilkan seperti file dengan atribut lain. Atribut ini berguna sebagai penanda bahwa file ini merupakan back up copy dari file yang asli.

2. Hidden

File dengan atribut Hidden akah disembunyikan atau tidak ditampakkan secara langsung pada windows explorer.

3. ReadOnly

File dengan atribut Read Only hanya dapat dibaca namun tidak dapat dirubah isi filenya sebelum kita menghilangkan atribut Read Only pada file tersebut.

4. System

File dengan atribut system akan disembunyikan dengan prioritas yang lebih diutamakan daripada file dengan atribut hidden.

3.8 Regresi Linier

Menurut (Armstrong, 2012) Analisis regresi merupakan suatu proses statistik untuk mengestimasi hubungan antara variabel-variabel, yakni berupa teknik-teknik memodelkan dan melakukan analisis beberapa variabel atas dasar bentuk hubungan antara satu variabel tak bebas dan satu atau lebih variabel bebas (prediktor).

Dengan analisis regresi peramalan atau perkiraan nilai variabel terikat pada nilai variabel bebas lebih akurat. Selain itu analisis ini untuk mengetahui arah hubungan antara variabel dependen apakah positif dan negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai dari variabel independen mengalami kenaikan atau penurunan variabel independen.

Dengan menggunakan metode ini dalam memprediksi perubahan yang terjadi pada metadata didalam semua file dapat diketahui apakah perubahan tersebut dapat mempengaruhi penambahan ukuran didalam sebuah file yang diukur menggunakan ukuran MegaByte, sehingga diharapkan dapat membantu didalam proses perubahan metadata.

Secara Umum, Perhitungan regresi bisa digambarkan dengan formula:

$$Y = \beta_0 + \beta_1 X, \beta_0 = \bar{y} - \beta_1 \bar{x}$$

Dapat dijelaskan bahwa Y adalah variabel terikat (dependent variable), X adalah variabel bebas (independent variable), serta β_0 dan β_1 adalah koefisien regresi.

$$\beta_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{(\sum_{i=1}^n y_i)(\sum_{i=1}^n x_i)}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

Dapat dijelaskan pada rumus dimana :

- Y_i merupakan nilai dari variabel dependent ke-i
- β_0 dan β_1 merupakan parameter model
- X_i adalah nilai variabel bebas X ke-i
- n adalah banyaknya data (sampel)

3.9 Confusion Matrix

Confusion matrix merupakan salah satu cara dalam melakukan visualisasi terhadap hasil pembelajaran sistem, visualisasi yang ditampilkan memuat dua kategori atau lebih (Rahman, Wiranto and Doewes, 2017). *Confusion matrix* digunakan pada penelitian ini dengan tujuan agar dapat mengukur kinerja suatu metode klasifikasi.

Tabel Table 3.2 *Confusion Matrix*

		Kelas Sebenarnya	
		1	2
Kelas Prediksi	1	True Positive	False Negative
	2	False Positive	True Negative

Keterangan:

- True Positive (TP)* : merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.
- True Negative (TN)* : merupakan jumlah data dengan kelas negative yang diklasifikasikan negati
- False Positive (FP)* : merupakan jumlah data dengan kelas positif yang diklasifikasikan negatif.
- False Negative (FN)* : merupakan jumlah data dengan kelas negatif yang diklasifikasikan positif.

Perhitungan Akurasi yang dilakukan oleh confusion matrix berdasarkan Tabel di atas dapat menggunakan persamaan sebagai berikut: