



TEKNOLOGI INFORMASI

UNIVERSITAS DARMA PERSADA

BAB 2

LANDASAN TEORI

2.1 Similarity Index

Similarity Index merupakan tingkat kemiripan yang ada pada artikel atau tulisan. Yang berasal dari informasi yang dibandingkan. Untuk dinilai tingkat kemiripannya berdasarkan persentase. Sedangkan plagiasi merupakan tindakan pengambilan karangan, tulisan atau karya orang lain sebagian atau keseluruhan tanpa mencantumkan sumber referensi pemilik aslinya. Kalimat dinyatakan dibajak sebesar <30% untuk tingkat ringan, 30-70% untuk tingkat sedang, dan >70% tingkat berat.

(Agustiawan, 2022).

2.2 Algoritma

Istilah algoritma berasal dari kata “algorism” yang telah ditemukan oleh ahli matematika muslim di abad ke-19 bernama Ibnu Al-Khwarizmi. Konsep pemikirannya dalam memecahkan permasalahan secara logis serta matematis yang dapat dijadikan sebagai metode dalam pemrosesan kerja komputer. Algoritma merupakan langkah-langkah untuk membenahi suatu persoalan secara terstruktur sehingga dapat memberi solusi yang tepat untuk suatu masalah dengan tepat.

Salah satu contoh sederhana penerapan algoritma dalam kehidupan sehari-hari adalah menghubungkan perangkat dengan jaringan wifi yaitu:

1. Pilih jaringan wifi yang tersedia.
2. Masukkan password.
3. Jika tidak sesuai, sistem menampilkan gagal koneksi.

4. Jika sesuai, koneksi internet dapat digunakan.

Adapun beberapa karakteristik dari algoritma ialah:

1. Menerima beberapa inputan.
2. Memproses inputan melalui langkah-langkah yang berurutan.
3. Setiap langkah mampu didefinisikan dengan jelas, sederhana, dan efektif.
4. Urutan langkah tersebut harus terbatas dan berhenti.
5. Menghasilkan keluaran atau output.

Algoritma terbaik akan menghasilkan output yang benar, tepat guna (efektif) dalam waktu yang relatif singkat dan penggunaan penyimpanan yang relatif sedikit. (Indahyanti, Rahmawati, 2020:2).

2.2.1 Winnowing

Winnowing adalah algoritma yang digunakan untuk memproses dokumen fingerprinting dengan langkah yang digunakan untuk mendeteksi kemiripan antar dokumen atau hanya sebagian teks saja. Prinsip kerja dari langkah dokumen fingerprinting ini yaitu menggunakan sebuah fungsi untuk mengubah setiap string menjadi angka. Proses ini bertujuan untuk membantu mengidentifikasi kesamaan, termasuk bagian kecil dari dokumen yang serupa.

Tahapan pada Algoritma Winnowing:

1. Pembuangan karakter yang tidak diperlukan.

Pada tahap ini proses yang dilakukan yaitu penghapusan tanda baca, spasi dan simbol-simbol seperti @, #,\$,*,(,),!,- ,_,”,+,>.

2. Pembentukan rangkaian K-gram

Pada pembentukan rangkaian K-gram dilakukan dengan cara membentuk rangkaian karakter sepanjang K dari hasil pembuangan karakter yang tidak diperlukan dengan ukuran K. Misalnya, $K = 3$.

3. Perhitungan fungsi hash untuk setiap K-gram

Maksudnya yaitu melakukan perhitungan nilai-nilai hash dari setiap gram.

4. Pembentukan window dari nilai hash

Nilai hash yang terbentuk pada tahap sebelumnya dipecah menjadi window berukuran w . Window pertama berisi nilai hash pertama hingga nilai hash ke- w . Window kedua dibentuk dari nilai hash kedua ke nilai hash $w+1$ dan seterusnya hingga seluruh nilai hash menjadi window.

5. Pemilihan fingerprint dari setiap window

Menentukan nilai fingerprintteks ditentukan dengan memilih nilai hash terkecil dari setiap window.

6. Persamaan Jaccard Coeficient

$$\text{Similarity} = \text{intersection} / (\text{union} - \text{intersection}) \times 100\%$$

(Mustaqiem, Winarni, 2021).

Pada penelitian ini metode winnowing akan digunakan untuk mengukur kemiripan pada data pengajuan surat domisili yang sudah diubah kedalam text secara keseluruhan yang akan dibandingkan dengan data yang sudah ada.

2.2.2 Jaccard Similarity

Pada tahapan ini untuk mencari nilai kemiripan pada sebuah hash atau string menggunakan rumus Jaccard Similarity dengan melihat irisan dan gabungan dari fingerprint antar dua dokumen. Persamaan merupakan persamaan Jaccard

Similarity dari dokumen teks X dan Y. Irisan dari dokumen X dan Y dibagi dengan gabungan dari dokumen X dan Y kemudian dikali 100% maka menghasilkan persentase kemiripan dari dokumen X dan Y.

$$\text{Similarity (X,Y)} = \frac{|X \cap Y|}{|X \cup Y|} \times 100\%$$

(Desena, Solichin, 2021).

2.2.3 Image Processing

Pengolahan citra digital (image processing) merupakan proses pengolahan piksel pada citra digital untuk tujuan tertentu. Suatu gambar dapat dikatakan sebagai citra digital jika gambar yang dihasilkan berasal dari tangkapan sebuah kamera, scanner atau proses perangkat elektronik lainnya. Pengolahan citra digital diproses oleh komputer menggunakan algoritma tertentu.

(Putri, 2020).

Pada penelitian ini teknik image processing yang digunakan adalah Optical Character Recognition (OCR). Untuk mengekstrak informasi yang ada di dalam gambar menjadi sebuah string.

2.2.4 Optical Character Recognition (OCR)

Pengenalan karakter optik (Optical Character Recognition) adalah proses menerjemahkan dalam bentuk teks yang dapat diedit dan disimpan. Teknologi ini memungkinkan menyimpan materi-materi cetakan dalam bentuk format teks yang lebih menghemat ruang pada hard disk.

(Humonggio, Abdullah, 2019).

OCR adalah metode pengenalan karakter tulisan mesin atau mengekstrak teks ke dalam format digital yang dapat disesuaikan. Metode OCR melibatkan tiga proses utama pre-processing, pengenalan (recognition) dan post-procesing.

1. Pre-processing: tahapan ini berfungsi untuk meningkatkan akurasi dalam pengenalan karakter yang meliputi binerisasi, penyusutan derau, normalisasi rasio dan skala, serta pemotongan citra.
2. Character recognition: terdapat empat jenis algoritma OCR yang utama, yaitu Template Matching, Pendekatan Statistik, Analisa Struktural, dan Jaringan Syaraf Tiruan.
3. Post-processing: tahapan pengolahan data hasil pengenalan karakter untuk diproses lebih lanjut.

(Ibnutama, Suryanata, 2020).

2.2.5 Tesseract OCR

Tesseract adalah mesin OCR open source yang dikembangkan oleh Hewlett-Packard antara tahun 1984 dan 1994.

(Mursari & Wibowo, 2021).

Tahapan yang digunakan untuk memenuhi kebutuhan sistem adalah sebagai berikut:

1. Atur skala gambar

Mengatur ukuran skala gambar minimal 300 dpi (standar tesseract).

2. Binarisasi

Tahap ini bekerja untuk mempertajam kualitas gambar dengan mengubah gambar menjadi hitam putih.

3. Noise Removal

Pada tahap ini merupakan variasi acak dari kecerahan atau warna pada gambar, yang dapat membuat teks pada gambar sulit untuk dibaca. Tesseract tidak dapat menghilangkan jenis noise tertentu dalam tahap binarisasi, yang dapat mengurangi akurasi.

4. Dilation & Erosion

Pada tahap ini pelebaran dan pengikisan tepi karakter dengan latar belakang yang sama untuk membesar (Dilation) atau menyusut (Erosion). Erosion dapat digunakan untuk mengecilkan karakter kembali ke struktur mesin normalnya.

(Official documentation: tesseract-ocr.github.io, 2022).

Pada penelitian ini peneliti menggunakan library *thiagoalessio/tesseract_ocr* yang digunakan untuk *image processing*.

2.3 Arsitektur Program

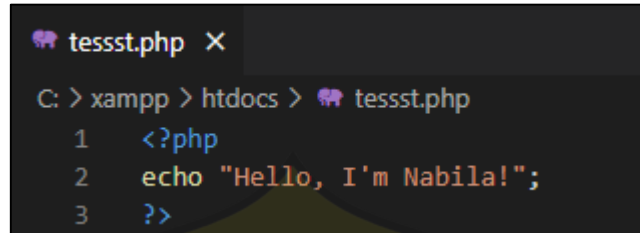
Berikut merupakan landasan teori aritektur program yang akan dibangun.

2.3.1 PHP

PHP (*Hypertext Preprocessing*) merupakan bahasa script pemrograman yang dapat ditanamkan atau disisipkan kedalam HTML. Php banyak dipakai pemrograman web dinamis yang dapat digunakan untuk membangun Content

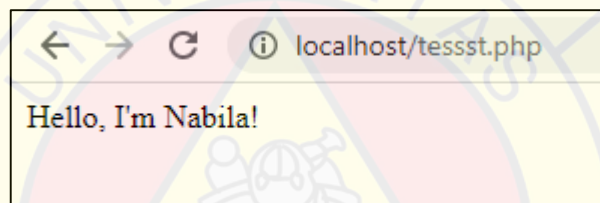
Management System (CMS). Salah satu contoh aplikasi yang dibangun menggunakan bahasa php adalah phpmysql untuk mengelola database mysql.

(Haqi, Satria, 2019:9).



```
tesst.php X
C: > xampp > htdocs > tesst.php
1  <?php
2  echo "Hello, I'm Nabila!";
3  ?>
```

Gambar 2.1 Contoh Code Php Sederhana.



Gambar 2.2 Hasil Output Code Php Pada Browser.

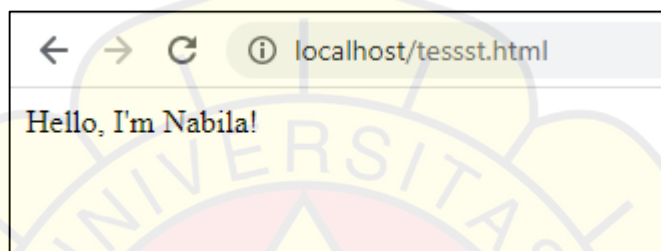
2.3.2 HTML

HTML adalah singkatan dari *Hyper Text Markup Language* yang merupakan bahasa pemrograman dasar dalam membuat website sebagai pondasi atau struktur dari web yang menjadi bahasa pemrogramannya yaitu PHP dan Javascript. HTML merupakan bahasa markup yang berjalan pada sisi client (*client side*) yang di tampilkan pada browser.

(Adi, AP. 2019:10).


```
tesst.html X
C: > xampp > htdocs > tesst.html > ...
1 <html>
2 <body>
3 Hello, I'm Nabila!
4 </body>
5 </html>
```

Gambar 2.3 Contoh Code HTML Sederhana.



Gambar 2.4 Output Code HTML Pada Browser.

2.3.3 Cascading Style Sheet (CSS)

CSS atau singkatan dari *Cascading Style Sheet* adalah suatu aturan untuk mengatur tampilan dari website sehingga tampilan dalam web lebih terstruktur yang dapat dikatakan sebagai konfigurasi tampilan dari suatu tag pada website. CSS dapat merubah text, warna, background dan posisi dari suatu tag.

(Marlina, Masnur, 2021).

2.3.4 Bootstrap

Bootstrap merupakan salah satu framework CSS yang dirancang untuk mendesain tampilan web agar terlihat menarik dan menyediakan class hingga komponen yang sudah siap dipakai atau dapat disebut seperti template. Dengan

adanya bootstrap tampilan website dapat lebih responsive ke semua device pengguna desktop maupun smartphone.

(Saputra, Sulistyanto, 2021).

2.3.5 Javascript

Javascript adalah bahasa pemrograman *client side* yang dinamis, mendekati bahasa manusia atau bisa dikatakan bahasa yang mudah dipahami oleh manusia. Tag *script* merupakan kode Javascript yang disisipkan dalam halaman web.

(Supardi, Ir. Yuniar, 2020:1).

2.3.6 JQuery

JQuery merupakan pustaka atau library JavaScript yang dapat berjalan pada HTML, PHP, dan JSP. JQuery memudahkan developer web dengan meringkas sebuah kode program Javascript dalam pembuatan website.

(Adi, AP. 2019:3).

2.3.7 MySQL

MySQL merupakan server untuk manajemen atau mengolah database dengan pemrograman perintah khusus (query SQL) yang bersifat multiuser dan bahasa yang digunakan yaitu Structured Query Language (SQL).

(Enterprise, Jubilee 2018:2).

2.3.8 PhpMyadmin

PhpMyadmin adalah bagian untuk mengelola database mysql yang ada pada server *local* 127.0.0.1 atau localhost dalam komputer.

(Haqi, Satria, 2019:10).

2.3.9 Composer

Composer merupakan aplikasi package manager untuk bahasa pemrograman PHP yang format standar dependency untuk mengelola *library* yang diperlukan. *Composer* dikembangkan oleh Nils Aderman dan Jordi Boggiano dan rilis pertama pada tahun 1 Maret 2012. *Repository* utama *composer* bernama *packagist*. (*Composer publisher*).

Pada penelitian ini *composer* digunakan untuk pemanggilan *library* yang dibuat secara terpisah dengan program aplikasi. Dengan tujuan membuat rapih dalam management program.

2.3.10 Imagick

Imagick adalah pengembangan php asli untuk membuat dan memodifikasi gambar menggunakan API ImageMagick.

ImageMagick adalah rangkaian perangkat lunak untuk membuat, mengedit, dan menyusun gambar bitmap yang dapat membaca, mengonversi, dan menulis gambar dalam berbagai format (lebih dari 100) termasuk DPX, EXR, GIF, JPEG, JPEG-2000, PDF, PhotoCD, PNG, Postscript, SVG, dan TIFF.

ImageMagick Studio LLC adalah organisasi non-profit yang didedikasikan untuk membuat solusi pencitraan perangkat lunak tersedia secara gratis.

(Official documentation: php.net/manual/en/intro.imagick.php).

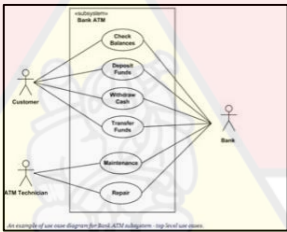
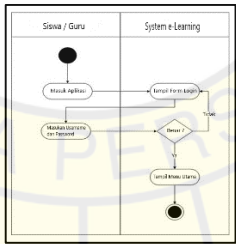
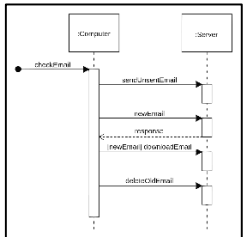
2.4 UML Diagram

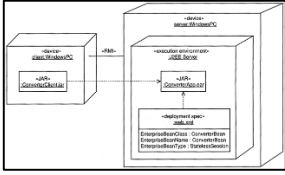
UML (*Unified Modeling Language*) merupakan sekelompok grafik rencana yang membantu dalam mendesain sistem perangkat lunak berorientasi objek (OO).

(Fowler, Martin, 2005:1).

Pada penulisan ini diagram yang digunakan hanya use case diagram, sequence diagram dan activity diagram.

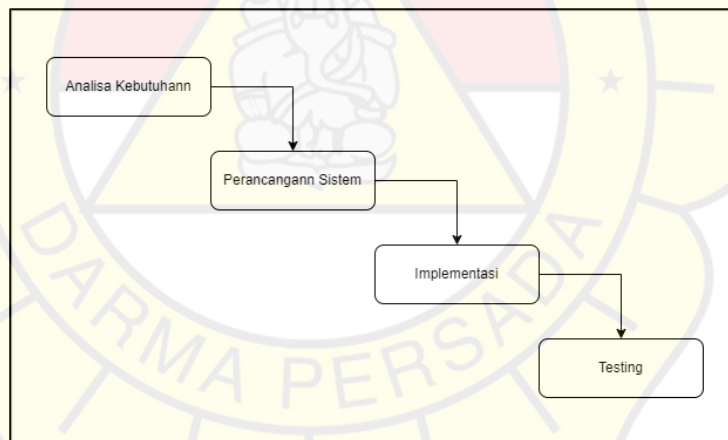
Table 2.1 UML Diagram

Nama	Gambar	fungsi
Use Case diagram		Untuk mendeskripsikan sebuah interaksi satu atau lebih actor dengan sistem informasi yang buat.
Activity Diagram		Untuk menggambarkan workflow aktivitas dari sebuah sistem.workflow aktivitas dari sebuah sistem
Sequence Diagram		Untuk menggambarkan workflow aktivitas dari sebuah sistem.

Deployment diagram		Untuk menggambarkan arsitektur sistem.
--------------------	---	--

2.5 Waterfall

SLDC (Systems Development Life Cycle) merupakan salah satu metode untuk mendevlope sistem aplikasi. Model Waterfall ialah model pertama yang umum digunakan oleh project-project pemerintahan dan perusahaan besar. Model ini juga memfokuskan pentingnya dokumentasi sehingga model ini cocok untuk proyek yang mengutamakan kualitas.



Gambar 2.5 Tahapan Diagram Waterfall.

Metode Waterfall memiliki beberapa tahapan sebagai berikut

1. Analisa Kebutuhan

Pada tahap ini kendala dan permintaan user dikumpulkan untuk dapat dilanjutkan ke tahap berikutnya.

2. Perancangan Sistem

Tahap ini merupakan penerapan terhadap analisa kebutuhan yang di rancang menggunakan software dan hardware.

3. Implementasi

Penerapan dan pelaksanaan pada tahap ini yaitu gabungan dari sistem yang sudah dibangun pada tahap-tahap sebelumnya, diterapkan dalam bentuk implementasi menjadi program satu kesatuan.

4. Testing

Tahap pegujian program untuk melihat apakah sistem siap untuk memenuhi kebutuhan yang di inginkan user.

(Usnaini, Yasin, Sianipar, 2021).

