

***Sistem Temu Kembali untuk Pasal-Pasal KUHP pada Hukum  
Indonesia***

***Dokumen Komponen dan Arsitektur Sistem Temu Kembali  
Informasi***

Ditulis untuk Memenuhi Tugas Mata Kuliah Temu Kembali



Informasi

**Anggota Kelompok**

**Ferdiansyah Dwi Nurcahyo**

**(20/459265/PA/19926)**

**Muny Safitri**

**(22/506376/NUGM/01039)**

**Raden Mas Garda**

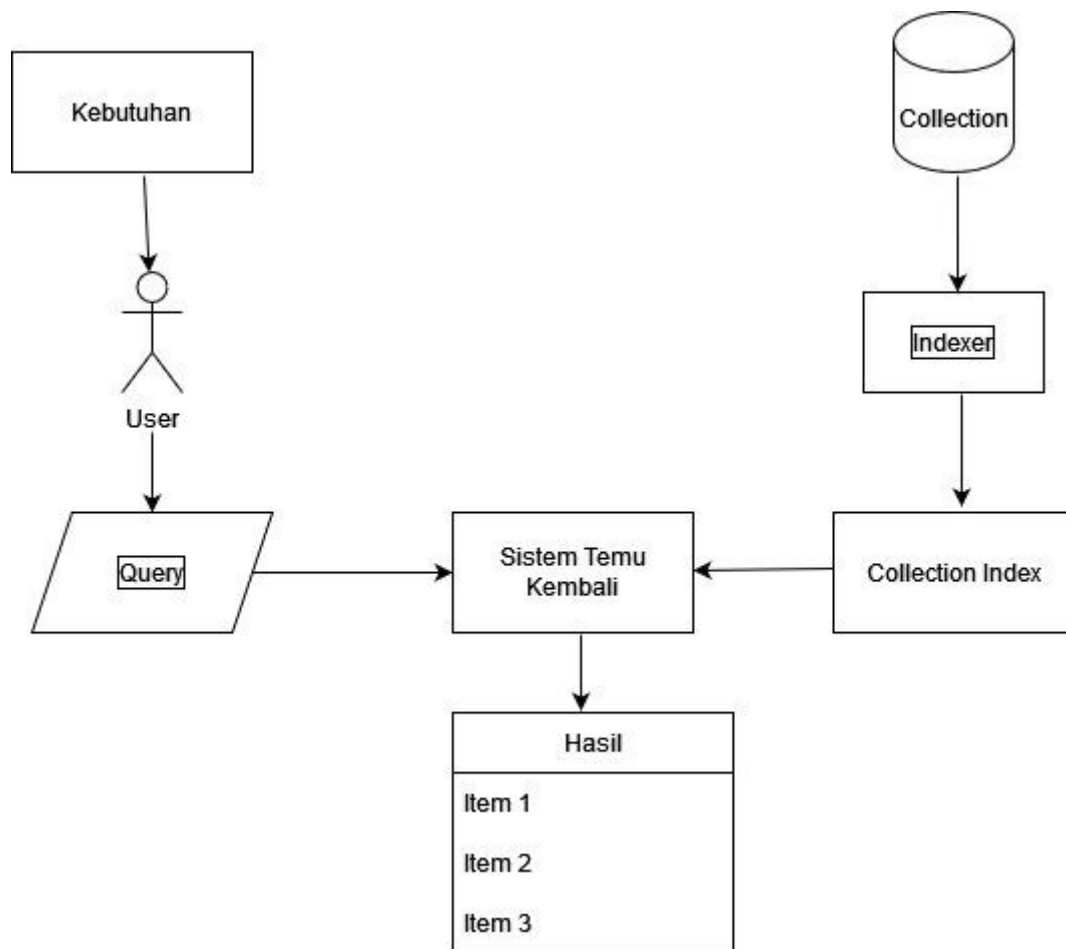
**(20/459275/PA/19996)**

**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS GADJAH MADA**

## DAFTAR ISI

1.	Arsitektur.....	1
2.	Database.....	2
3.	Pembobotan Kata dan Vector Space Model.....	3
4.	Query: Praproses dan <i>Constraint</i> .....	5
5.	Komponen Pendukung.....	6

## 1. Arsitektur



Arsitektur pada sistem Temu Kembali Informasi (TKI) berawal dari kebutuhan pengguna untuk mengetahui pasal-pasal KUHP apa yang terkait dengan permasalahannya. kebutuhan user tersebut kemudian diterjemahkan menjadi query yang akan dikirim ke antarmuka sistem TKI. Setelah menerima query, sistem TKI segera men-*generate* query untuk mengambil semua data dari database. Data dari database tersebut kemudian dilakukan indexing dengan dilakukan text preprocessing, pembuatan struktur data berkaitan (ex: inverted index dan postings list), dan pembuatan ranking oleh indexer. Hasilnya adalah sekumpulan index dan posting yang berada pada suatu struktur data tertentu dan terurut yang disebut collection index. Collection index ini kemudian dikirimkan ke antarmuka sistem TKI berupa list hasil.

## 2. Database



Database yang kami buat menggunakan MySQL. Dari database ini, kami membuat tabel Bab yang berisi `id_bab`, `judul_bab`, dan `update`. Kemudian tabel Pasal yang berisi `id_pasal`, `isi_pasal`, dan `update`. Serta tabel Ayat yang berisi `id_ayat`, `isi_ayat` dan juga `update`. Primary key yang digunakan adalah `id_bab`, `id_pasal`, dan `id_ayat`. Relation yang terjadi antara masing masing tabel adalah one to many pada Bab ke Pasal, one to many pada pasal ke ayat, dan many to one pada ayat ke bab.

### 3. Pembobotan Kata dan Vector Space Model

#### 3.1 Text Processing

Merupakan tahap persiapan yang dilakukan untuk menyiapkan dokumen sebelum diolah. penggunaan *text preprocessing* dilakukan karena dokumen teks tidak dapat diproses langsung oleh algoritma pencarian, sehingga diperlukan proses untuk menghasilkan data numerik yang akan digunakan dalam perhitungan. Tahapan Text Preprocessing pada meliputi:

- Penghapusan format dan markup dalam dokumen
- Tokenizing
- Filtering
- Stemming

#### 3.2 Cosine Similarity

Dokumen dan query dipresentasikan dalam bentuk vektor

$$Q = [q_1, q_2, \dots, q_n]$$

Kesamaan antara dokumen dan query diukur berdasarkan sudut cosinus dari vektor dokumen dan vektor query (*cosine measure*)

$$\begin{aligned} \cos(Q, D) &= \frac{(Q \cdot D)}{\|Q\| \|D\|} \\ &= \frac{\sum_{i=1}^n q_i d_i}{\sqrt{\sum_{i=1}^n q_i^2} \sqrt{\sum_{i=1}^n d_i^2}} \end{aligned}$$

keterangan

$q_i$  = bobot query

$|Q|$  = panjang query

$d_i$  = bobot dokumen

$|D|$  = panjang dokumen

dimana  $d_{ij}$  adalah bobot term dalam dokumen dengan rumus

$$d_{ij} = \frac{t_{ij}}{\sum_{j=1}^n t_{ij}}$$

dan  $q_{ij}$  adalah bobot query dengan rumus

$$q_{ij} = \frac{h_{ij}}{\sum_{i=1}^n h_{ij}} \cdot \frac{1}{\sum_{j=1}^n q_{ij}}$$

### **3.3 Vector Space Model (VSM)**

Vector Space Model (VSM) merupakan model yang digunakan untuk mengukur kemiripan antara dokumen dan query yang mewakili setiap dokumen dalam sebuah koleksi sebagai sebuah titik dalam ruang . Dalam metode Vector Space Model dihitung weighted dari setiap term yang terdapat dalam semua dokumen dan query dari user. Term adalah kata atau kumpulan kata yang merupakan ekspresi verbal dari suatu pengertian, kumpulan kata disini adalah kata dari pasal-pasal KUHP di Indonesia.

Pada implementasi metode VSM akan membandingkan data query yang sudah melalui proses preprocessing dengan membandingkan data dokumen pasal yang ada di database. Pada perhitungan VSM digunakan pembobotan TF-IDF dan perhitungan similarity dengan menggunakan Cosine Similarity. Metode TF-IDF adalah cara untuk memberikan bobot hubungan suatu term terhadap dokumen. Metode ini menggabungkan dua konsep perhitungan bobot yaitu frekuensi kemunculan kata dalam suatu dokumen dan inverse dari frekuensi yang mengandung kata tersebut .



#### 4. Query: Praproses dan *Constraint*

Proses preprocessing query input bertujuan untuk menyiapkan query input yang akan dibandingkan dengan dokumen yang ada di database. Secara umum, Query dibentuk menjadi vektor token yang berfungsi untuk mencari token yang berhubungan dengan melihat kesamaan vektor dokumen query berdasarkan jarak vektor.

Untuk menekankan user-centered experience, perlu dilakukan tindakan untuk meningkatkan *robustness* sistem TKI dan mencegah *error by input*. *Error by input* bisa disebabkan oleh *user* yang menuliskan kata kunci query secara tidak sempurna (salah eja atau kata kunci dituliskan tidak sampai selesai). Pada pengamatan terhadap sistem TKI yang sudah ada, kemudahan yang diberikan oleh sistem TKI yang sudah terbangun dengan baik (*well-established*) membiasakan user untuk tidak merasa perlu menuliskan kata kunci secara sempurna.

Untuk memproses query yang berdasarkan pada kata kunci tidak sempurna ini, dapat dilakukan dengan mengganti query yang berasal dari user dengan query yang berasal dari mesin (*machine-generated query*) yang dapat berupa *regular expression* (regex) atau bisa juga dilakukan proses pencocokan kata kunci dengan term pada term dictionary yang memiliki permutasi huruf paling mendekati. Proses ini mungkin membutuhkan pembelajaran mesin untuk secara optimal melihat kebiasaan user spesifik secara khusus dan populasi user pada sistem TKI secara umum. Penghapusan markup format untuk menghilangkan beberapa format HTML yang tidak digunakan untuk perhitungan juga diperlukan..



## **5. Komponen Pendukung**

### **5.1 Website**

Website adalah sebuah sistem layanan informasi dengan menggunakan konsep hyperlink yang dapat memudahkan pengguna untuk memperoleh informasi dengan cukup mendapatkan suatu link berupa teks ataupun gambar, informasi tersebut berupa teks ataupun gambar yang ditampilkan secara lebih rinci dihalaman browser serta dapat diakses dari berbagai perangkat gadget seperti komputer pc/laptop ataupun smartphone

### **5.2 Framework Laravel**

Laravel adalah sebuah framework PHP yang dirilis dibawah lisensi MIT, dibangun dengan konsep MVC (model view controller). Laravel adalah pengembangan website berbasis MVP yang ditulis dalam PHP yang dirancang untuk meningkatkan kualitas perangkat lunak. Laravel menyediakan sintaks yang ekspresif dan elegan, serta memudahkan pengembang dalam mengembangkan sebuah proyek-proyek web seperti authentication, routing, sessions, dan caching. Framework ini bertujuan untuk memberikan kenyamanan bagi pengembang tanpa mengorbankan fungsionalitas aplikasi . MVC adalah sebuah 6 pendekatan perangkat lunak yang memisahkan aplikasi logika dari presentasi. MVC memisahkan aplikasi berdasarkan komponen- komponen aplikasi, seperti : manipulasi data, controller, dan user interface.

1. Model, mewakili struktur data. Biasanya model berisi fungsi-fungsi yang membantu seseorang dalam pengelolaan basis data seperti memasukkan data ke basis data, pembaruan data dan lain-lain.
2. View, adalah bagian yang mengatur tampilan ke pengguna. Bisa dikatakan berupa halaman web.
3. Controller, merupakan bagian yang menjembatani model dan view

### 5.3 MySQL

MySQL sudah tidak asing lagi sebagai database server yang digunakan dalam pemrograman PHP MYSQL, Berfungsi sebagai penyimpanan data dalam database dan memanipulasinya. Manipulasi data tersebut seperti menambah, mengatur, dan menghapus data yang berdata dalam database. Database sering disebut juga sebagai sekumpulan data, secara teknis yang berada dalam sebuah database adalah kumpulan tabel. Tujuan utamanya untuk memudahkan dalam mengakses data secara efektif cepat.

### 5.4 Pengembangan

Dokumen yang telah tersimpan di Database (Pasal KUHP), Satu Bab Pasal diwakili dengan satu dokumen. Kemudian, akan dilakukan pembuatan indeks sesuai jumlah *term* (kata) pada pasal. Lalu akan dilakukan pembobotan(*weight*) dan akan dilakukan perhitungan vektor, algoritma yang digunakan adalah TF IDF. Metode TF-IDF adalah cara untuk memberikan bobot hubungan suatu term terhadap dokumen. Metode ini menggabungkan dua konsep perhitungan bobot yaitu frekuensi kemunculan kata dalam suatu dokumen dan invers dari frekuensi yang mengandung kata tersebut . setelah itu, dilakukan perhitungan vektor atau panjang vektor setiap dokumen. Pada laman TKI (*Retrieval Information*) ditampilkan dokumen yang berhubungan dengan key yang diinput oleh *user* berdasarkan angka kemiripan yang tertinggi dengan kata kunci yang dimasukkan.