

MAKALAH SISTEM TEMU KEMBALI INFORMASI (INFORMATION RETRIEVAL SYSTEM)

OLEH :

KARINA PURI DAYANTI (15.01.55.0009)

Abstrak

Information retrieval system adalah system yang digunakan dalam menemukan kembali informasi yang pernah tersimpan sebelumnya. Jumlah dokumen dengan jenis file yang beragam dan banyak menjadikan proses dalam pencarian dokumen menjadi sulit dan membutuhkan waktu lama. Maka dilakukan penerapan information retrieval system, yang diharapkan dapat menghasilkan pencariaan dokumen yang relevan dan akurat sesuai kategorinya. Dan dengan menggunakan proses tersebut pencarian sebuah data menjadi lebih cepat sehingga pekerjaan dapat terselesaikan dengan tepat waktu.

Kata Kunci : tokenisasi, stopword removal, stemming

BAB I

PENDADULUAN

**1.1. Latar Belakang**

Sistem Temu Kembali Informasi (*information retrieval system)* digunakan untuk menemukan kata/kalimat/informasi tertentu dari berbagai sumber relevan yang dicari atau dibutuhkan. Selain itu *information retrieval* juga bisa didefinisikan sebagai ilmu pencarian informasi pada dokumen, bisa berupa pencarian untuk dokumen itu sendiri, untuk menjelaskan dokumen, atau mencari di dalam *database* berupa teks, suara, gambar atau data.

Kebutuhan akan suatu informasi dari berbagai jenis file yang bisa jadi berasal dari banyak sumber terkadang memakan waktu yang cukup lama untuk mendapatkan informasi yang relevan dan akurat. Oleh sebab itu, maka perlu adanya penerapan *information retrieval system* yang diharapkan dapat menghasilkan pencarian dokumen yang relevan dan akurat sesuai kategorinya. Sehingga kinerja menjadi lebih cepat dan hasil yang didapat juga sesuai dengan kebutuhan.

**1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang akan dibahas dalam makalah ini adalah:

1. Apa yang dimaksud dengan sistem temu kembali atau *information retrieval system ?*
2. Bagaimana jalannya proses sebuah sistem temu kembali atau *information retrieval system ?*
3. Apa saja komponen sebuah sistem temu kembali atau *information retrieval system ?*
4. Apa peranan dari *information retrieval system ?*

BAB II

LANDASAN TEORI

* 1. **Pengertian Information Retrieval System**

*Information Retrieval*System atau sering disebut “system temu kembali infromasi” adalah ilmu yang mempelajari prosedur-prosedur dan metode-metode untuk menemukan kembali informasi yang tersimpan dari berbagai sumber (*resource*s) yang relevan atau koleksi sumber informasi yang dicari atau dibutuhkan. Dengan tindakan  index (*indexing*), panggilan (*searching*), pemanggilan data kembali (*recalling*).

Dalam pencarian data, beberapa jenis data dapat ditemukan diantaranya texts, table, gambar (*image*), video, audio. Adapun tujuan dari Infromation Retrieval ialah untuk memenuhi informasi pengguna dengan cara meretrieve dokumen yang relevan atau menguragi dokumen pencarian yang tidak relevean.

Proses yang berlangsung dalam i*nformation retrieval system* terdiri dari 2 bagian utama, yaitu *indexing subsystem*, dan *searching subsystem* (*matching system*). Proses *indexing* dilakukan untuk membentuk basis data terhadap koleksi dokumen yang dimasukkan, atau dengan kata lain, *indexing* merupakan proses persiapan yang dilakukan terhadap dokumen sehingga dokumen siap untuk diproses. Proses *indexing* sendiri meliputi 2 proses, yaitu *document indexing* dan *term indexing*. Dari term indexing akan dihasilkan koleksi kata yang akan digunakan untuk meningkatkan performansi pencarian pada tahap selanjutnya.

Adapun tahap-tahap yang terjadi pada proses *indexing*, yaitu (Harjanto, 2012):

1. *Tokenisasi* dokumen, yaitu   proses mengubah dokumen menjadi kumpulan  *term*  dengan  cara  menghapus  semua  karakter  tanda  baca yang terdapat pada *token*. Hingga pada akhirnya yang diperoleh hanya kumpulan kata-kata dari suatu teks/dokumen.
2. *Stopword removal* dokumen, yaitu kata-kata yang sering muncul dalam dokumen namun artinya tidak deskriptif dan tidak memiliki keterkaitan dengan tema tertentu. Pada Bahasa Indonesia, *stopword* disebut juga sebagai  kata  yang  tidak  penting,  misalnya  “di”,  “oleh”,  “pada”, “sebuah”, “karena” dan lain sebagainya.
3. *Stemming* dokumen,   yaitu   tahap penghilangan   imbuhan   sehingga didapatkan kata dasar dari *term-term* dokumen inputan.
4. *Term Weighting*, yaitu proses pembobotan pada setiap *term* (kata) yang ada didalam dokumen

**2.2 Tujuan Sistem Temu Kembali Informasi**

Information Retrieval System memiliki kegunaan yang banyak untuk user. Kita bisa melihat fungsinya di mesin pencari untuk mencari informasi, atau di perpustakaan, di apotik dan lain sebagainya. Itu semua adalah karena jasa IR. Information Retrieval mempunyai peran untuk:

1. Menganalisis isi sumber informasi dan pertanyaan pengguna.

2. Mempertemukan pertanyaan pengguna dengan sumber informasi untuk mendapatkan dokumen yang relevan.

**2.3 Deskripsi masalah**

Pada deskripsi masalah penulis akan menjabarkan 2 tahap yang terjadi pada proses indexing yaitu :

Tokenisasi adalah proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu. Sebagai contoh, tokenisasi dari kalimat “Aku baru saja makan bakso pedas” menghasilkan enam token, yakni: “Aku”, “baru”, “saja”, “makan”, “bakso”, “pedas”. Biasanya, yang menjadi acuan pemisah antar token adalah spasi dan tanda baca. Tokenisasi seringkali dipakai dalam ilmu linguistik dan hasil tokenisasi berguna untuk analisis teks lebih lanjut

Istilah penting   :

* + Tokenisasi : proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu.
  + Stopword : kata-kata yang disaring sebelum atau sesudah pengolahan data bahasa alami (teks).
  + Stopword removal : Tahapan untuk menghilangkan stopword.

Perintah program yang baru :

1. move\_uploaded\_file($lokasi\_file,”$folder”) : …..
2. mysqli\_connect(“localhost”,”root”,””,”upload”) :……..
3. include(‘class.pdf2text.php’) : ……………………..
4. strtolower(trim($teks)):
5. explode(” “, $teks):
6. array\_diff($myArray, $astoplist) :

Alur program :  
Mulai -> Mengunggah file pdf -> Simpan info ke tabel upload -> Baca file pdf -> Tokenisasi -> Simpan

Tujuan pengujian :

1. Mengunggah file pdf
2. Melakukan tokenisasi
3. Melakukan stopword removal

Input :

1. file pdf Undang-undang Republik Indonesia Nomor 35 Tahun 2009 tentang Narkotika
2. file pdf Undang-undang St. No. 419 tgl. 22 Desember 1949 tentang Obat Keras

Luaran diharapkan  : berhasilnya file dengan format pdf dan hasil tokenisasinya masuk ke dalam tabel yang berada di database yang sudah dibuat dan tidak gagal saat upload.  
Luaran aktual :

1. Terkadang file tidak dapat diupload sehingga file pdf harus diganti dengan yang lain
2. Masih terdapat kesalahan pada script sehingga menyebabkan eror.
3. kesalahan pada penulisan scrip sehingga menyebabkan eror  
   Status :
   * 1. *Gagal upload*
     2. *Sukses upload*
     3. *Tokenisasi*

Analisa:

1. *Gagal upload*

Keterangan : eror terjadi pada file php hasil\_upload karena ada beberapa script yang belum di nonaktifkan

1. *Sukses upload*

Keterangan : Script program yang tidak dibutuhkan  tersebut telah dinonaktifkan sehingga file sukses di upload ke database

1. *Tokenisasi*

Tokenisasi berhasil, tidak terjadi error pada saat tokenisasi dan file tokenisasi masuk di tabel dalam database sehingga pada proses tokenisasi berhasil.

Stemmer merupakan aplikasi pemotongan imbuhan pada kata berimbuhan (awalan, akhiran, sisipan, kombinasi) yang dijalankan dengan algoritma tertentu.

Algoritme yang pertama dikembangkan adalah oleh Martin Porter, yang mengolah pemotongan imbuhan dalam bahasa inggris. Pemotongan imbuhan lebih banyak digunakan pada aplikasi penelusuran data seperti pada mesin pencari (misalnya google, vivisimo, dan lainnya).

Pengolahan queri yang mengandung imbuhan pada mesin pencari di atas (dalam bahasa Inggris), akan memberikan hasil pencarian halaman-halaman yang memuat kata dasarnya atau kombinasi imbuhan lainnya.  
Istilah penting   :

* + Stemming : Stemming merupakan suatu proses untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan semua imbuhan (affixes) baik yang terdiri dari awalan (prefixes), sisipan (infixes),akhiran (suffixes) dan confixes (kombinasi dari awalan dan akhiran) pada kata turunan.
  + Stoplist : Stop list merupakan fitur yang dimiliki sistem information retrieval (search engine), dimana term/kata-kata yang dianggap umum akan diabaikan dalam proses indexing (contoh : I, and, or, is, of, was).
  + Stopword removal : Tahapan untuk menghilangkan stopword.

Perintah program yang baru :

1. move\_uploaded\_file($lokasi\_file,”$folder”) : …..
2. mysqli\_connect(“localhost”,”root”,””,”upload”) :……..
3. include(‘class.pdf2text.php’) : ……………………..
4. strtolower(trim($teks)):
5. explode(” “, $teks):
6. array\_diff($myArray, $astoplist) :

Alur program :  
Mulai -> Mengunggah file pdf -> Simpan info ke tabel upload -> Baca file pdf -> Tokenisasi -> Stemming -> Simpan

Tujuan pengujian :

1. Mengunggah file pdf
2. Melakukan tokenisasi
3. Melakukan stopword removal
4. Melakukan stemming

Input :

1. file pdf Undang-undang Perlindungan Anak
2. file pdf Undang-undang Informasi dan Transaksi Elektronik
3. file pdf Undan-undang Ketenagakerjaan

Luaran diharapkan  : berhasilnya file dengan format pdf dan hasil tokenisasinya masuk ke dalam tabel yang berada di database yang sudah dibuat dan tidak gagal saat upload.  
Luaran aktual :

1. Terkadang file tidak dapat diupload sehingga file pdf harus diganti dengan yang lain
2. Masih terdapat kesalahan pada script sehingga menyebabkan eror.
3. Kesalahan pada penulisan script sehingga menyebabkan eror.

Status :

* + 1. *Gagal upload*
    2. *Sukses upload*
    3. *Tokenisasi*

Analisa:

1. *Gagal upload*

Keterangan : eror terjadi pada file php hasil\_upload karena ada beberapa script yang belum di nonaktifkan

1. *Sukses upload*

Keterangan : Script program yang tidak dibutuhkan  tersebut telah dinonaktifkan sehingga file sukses di upload ke database

1. *Tokenisasi*
2. *Stemming*

Proses Stemming berhasil dan tidak terjadi error pada saat upload file pdf dan proses penyimpanan di database berhasil.

**2.4 Koding Program**

Koding dari proses indexing diatas dapat diakses pada link github penulis yaitu <https://github.com/karinapurid/information-retrieval>

BAB III

* 1. KESIMPULAN

Dapat disimpulkan bahwa Information Retrieval System ini berguna sebagai alat pencarian informasi dengan waktu singkat dengan data yang didapat merupakan data yang relevan dan akurat. Information Retrieval System ini merujuk pada keseluruhan kegiatan yang meliputi pembuatan wakil informasi (representation), penyimpanan (storage), pengaturan (organization) sampai kepada pengambilan (access). Jadi tujuan akhir dari system tersebut adalah memberikan kemudahan dan kepuasan bagi user atau pengguna system.

3.2 DAFTAR PUSTAKA

http://karinapurid.blogspot.com

<https://ligiaprapta17.wordpress.com/2015/03/03/pengertian-information-retrieval-ir-peranan-ir-dan-contoh-contoh-ir/>

<https://radenkanjengmazzbroo.blogspot.com/2017/09/makalah-information-retrieval.html>