

SISTEM TEMU KEMBALI KELAS BUKU UNTUK MENENTUKAN NOMOR KLASIFIKASI BUKU DI PERPUSTAKAAN

Oleh

Putu Tika Parmawati (Perpustakaan)
Kadek Surya Mahedy (Jurusan Manajemen Informatika, FTK)
UNDIKSHA

ABSTRAK

Penelitian ini mengembangkan Sistem Temu Kembali Informasi dalam bidang perpustakaan, yang bertujuan menghasilkan rancangan dan aplikasi Sistem Temu Kembali Informasi Kelas Buku sebagai alternatif untuk membantu analisis subyek buku dalam menentukan nomor klasifikasi buku di perpustakaan.

Sistem Temu Kembali Informasi Kelas Buku dikembangkan menggunakan metode Model Ruang Vektor untuk memodelkan dokumen dan *query*. Tiap dokumen dan *query* akan direpresentasikan dalam bentuk vektor-vektor. Elemen vektor adalah *term-term* yang terdapat pada dokumen dan *query* dengan nilai bobot masing-masing *term*. Perhitungan bobot tiap *term* dilakukan dengan konsep *Term Frequency* dan *Inverse Document Frequency*. Vektor yang terbentuk dari dokumen dan *query*, tingkat kesesuaian ditentukan dengan menggunakan *Cosine Similarity Measure* melalui proses pencocokan.

Penelitian ini menghasilkan sebuah perangkat lunak bernama Sistem Temu Kembali Informasi Kelas Buku. Sistem Temu Kembali Informasi Kelas Buku merupakan perangkat untuk membantu proses menganalisis subyek sebuah buku dalam menentukan nomor klasifikasi buku di perpustakaan. Sistem ini mengelola masukan *query* pengguna yang akan dicocokkan dengan koleksi dokumen yang dimiliki sistem. Koleksi dokumen adalah kumpulan judul dan daftar isi dari beberapa buku disimpan dalam basis data. Hasil pencocokan antara *query* pengguna dan koleksi dokumen menghasilkan kelas buku yang relevan atau sesuai dengan *query* pengguna.

Kata-kata kunci : Perpustakaan, Sistem Temu Kembali Informasi, Model Ruang Vektor.

ABSTRACT

This study developed Information Retrieval System for library, by producing Information Retrieval System design and application as an alternative to analyze the subject of references and to give book classification number in library.

Information Retrieval System was developed through the method of Vector Space Model for modeling document and query. Each document and query was represented in the form of vectors. The elements of vector are terms existing in the document and query with their particular weight. The calculation of the weight of

each term was carried out through Term Frequency and Inverse Document Frequency Concepts. The relevance level of vector formed through document and query was measured through Cosine Similarity Measure.

The software produced in this study can be used to analyze the subject of references and give book classification number in library. This system organizes the query entry of the user that will be matched with document collection available in the system. Document collection is a collection of title and table of content of some books that are saved in the database. The matching result between user query and document collection shows a type of book that is relevant to user query.

Key words: Library, Information Retrieval, Vector Space Model.

1. PENDAHULUAN

Perpustakaan adalah institusi yang bertugas mengelola bahan pustaka, baik berupa buku maupun bukan berupa buku (*non book material*) sehingga dapat digunakan sebagai sumber informasi oleh setiap pemakainya. Bahan pustaka atau bahan informasi lain yang ada di perpustakaan disusun berdasarkan system klasifikasi. Klasifikasi adalah alat untuk mempermudah pencarian buku atau dokumen. Sistem klasifikasi mempunyai sandi/kode yang dapat berupa gambar, huruf, angka, atau yang lain, yang telah disepakati sebagai pengganti sesuatu.

Sistem klasifikasi yang banyak digunakan oleh perpustakaan adalah *Dewey Decimal Classification* (DDC). Pemberian sandi atau kode pada bahan pustaka menggunakan sistem klasifikasi DDC berdasarkan subyek buku. Pengklasifikasian berdasarkan subyek diawali dengan suatu kegiatan yang disebut "analisis subyek". Kegiatan analisis subyek ini merupakan kegiatan yang sangat penting dan memerlukan kemampuan intelektual, karena di sinilah ditentukan pada subyek apa suatu bahan pustaka ditempatkan atau menetapkan isi bahan pustaka. Oleh karena itu, analisis ini harus dikerjakan secara akurat dan konsisten.

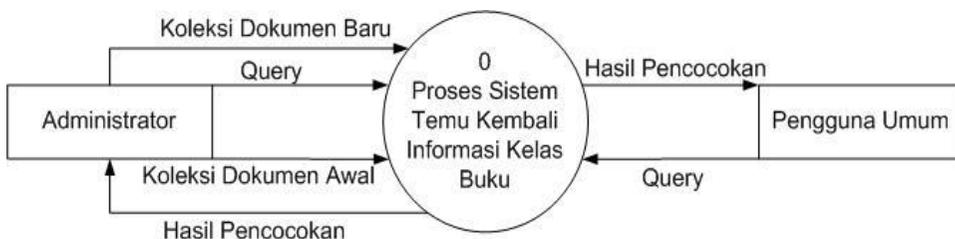
Dalam pelaksanaan pekerjaannya seorang pustakawan masih sering mengalami kesulitan dalam mengklasifikasi bahan pustaka dan memberikan nomor kelas buku, hal ini dikarenakan minimnya pengetahuan dasar mengenai sistem klasifikasi. Terkadang pustakawan lupa bagaimana mekanisme analisis subyek suatu bahan pustaka yang akan diklasifikasi serta kurangnya konsistensi dalam pemberian nomor kelas buku.

Kemajuan teknologi komputer yang pesat sangat membantu kehidupan manusia bahkan di dalam bidang-bidang di luar disiplin ilmu komputer. Sistem temu kembali informasi (*information retrieval system*) adalah suatu proses untuk mengidentifikasi, kemudian memanggil (*retrieve*) suatu dokumen dari suatu simpanan (*file*), sebagai jawaban atas permintaan informasi. Aplikasi Sistem Temu Kembali Informasi dikembangkan dalam berbagai bidang, termasuk dalam bidang perpustakaan. Pengembangan aplikasi sistem temu kembali informasi kelas buku dipilih menjadi alternatif untuk membantu analisis subyek dalam penentuan kelas buku dan nomor klasifikasi buku di perpustakaan. Pemanfaatan sistem temu kembali diharapkan dapat memberikan banyak keuntungan bagi perpustakaan diantaranya: memudahkan dalam menganalisis subyek buku berdasarkan informasi yang tersimpan di basis data, memudahkan proses klasifikasi bahan pustaka di perpustakaan, basis data kelas buku yang disimpan akurat dan mudah dimutakhirkan. Dari paparan di atas maka dipandang perlu untuk mengembangkan sebuah aplikasi Sistem Temu Kembali Informasi Kelas Buku untuk Menentukan Nomor Klasifikasi Buku di Perpustakaan.

2. METODOLOGI

Kebutuhan fungsional perangkat lunak Sistem Temu Kembali Informasi Kelas Buku dijelaskan dengan aliran data yang digambarkan dengan DAD: Diagram Aliran Data aras 0 (DFD : *Data FlowDiagram*).

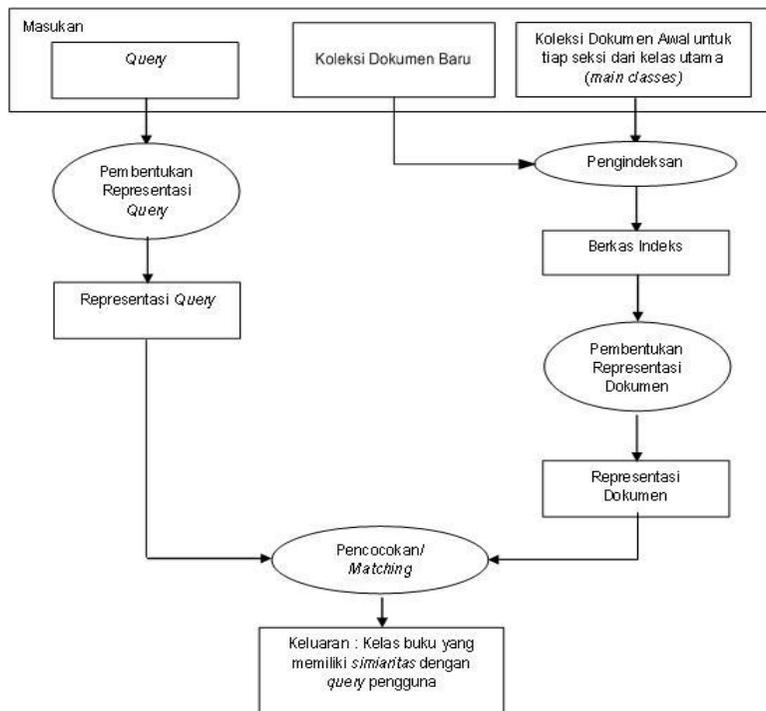
Entitas luar yang dalam Sistem Temu Kembali Informasi Kelas Buku ada dua buah, yaitu kedua karakteristik pengguna : Administrator dan Pengguna Umum sesuai dengan Gambar 1.



Gambar 1 Diagram Konteks Data

Pengguna Administrator akan memasukkan koleksi dokumen awal, koleksi dokumen dan *query* ke sistem maka sistem akan memberikan output hasil pencocokan dokumen dengan *query*. Pengguna umum hanya dapat memasukkan *query* ke sistem dan sistem akan memberikan output hasil pencocokan dokumen dengan *query*

Aktivitas Sistem Temu Kembali Kelas Buku sesuai Diagram Alir Data aras 0 dapat digambarkan dengan arsitektur sistem pada gambar 2.



Gambar 2 Arsitektur Sistem Temu Kembali Informasi Kelas Buku

Berdasarkan gambar arsitektur sistem diatas, tahapan aktivitas dalam Sistem Temu Kembali Informasi Kelas Buku yaitu:

- Koleksi Dokumen

Koleksi dokumen yang diterima oleh sistem merupakan kumpulan banyak dokumen untuk tiap seksi dari kelas utama (*main classes*) buku. Setiap dokumen terdiri dari judul dan daftar isi dari beberapa buku yang masing-masingnya dapat terdiri dari satu atau lebih *term*.

- Pengindeksan

Pengindeksan / *indexing* dilakukan untuk mendapatkan bobot dari setiap *term* dalam dokumen. Penghitungan bobot tersebut dilakukan dengan melakukan perhitungan terhadap *Term Frequency (Tf)* dan *Document Frequency (Df)* dari tiap *term* yang terdapat di koleksi dokumen. Nilai *Df* selanjutnya akan diproses menjadi nilai *Inverse Document Frequency (Idf)* yang akan digunakan dalam perhitungan bobot *term*. Formula perhitungan *Tf* yang akan digunakan pada Penelitian ini adalah formula *Logarithmic Tf*. Formula ini berbentuk:

$$1 + \ln(\text{Term Frequency}) \dots\dots\dots (1)$$

dimana *Term Frequency* adalah jumlah kemunculan sebuah *term* dalam dokumen atau *query*. *Df* suatu *term* adalah banyaknya dokumen dalam koleksi, yang mengandung kemunculan *term* tersebut. *Idf* suatu *term* adalah invers dari *Df*. Formula untuk *Inverse Document Frequency term i* adalah:

$$\text{Idfi} = \ln(N / \text{Dfi}) \dots\dots\dots (2)$$

dimana N = jumlah seluruh dokumen dalam koleksi, dan Dfi = *Document Frequency term i*. Formula untuk bobot *term* dalam dokumen atau *query* adalah:

$$w_{ij} = \text{Tf}_{ij} \cdot \text{Idfi} \dots\dots\dots (3)$$

dimana w_{ij} = bobot *term i* dalam dokumen j , $\text{Idfi} = \text{Idf term } i$, $\text{Tf}_{i,j} = \text{Tf term } i$ dalam dokumen j .

- Representasi *Query*

Representasi *query* dalam sistem ini berupa kumpulan dari beberapa kata kunci maupun pernyataan kebutuhan pengguna umum. *Query* akan mengalami proses pengindeksan sebelum dicocokkan dengan koleksi dokumen. Pada proses pengindeksan akan dihitung nilai *Tf*, *Df*, *Idf* dan bobot pada *term query*.

- Pencocokan/ *Matching*

Proses pencocokan *query* pengguna dengan koleksi dokumen menggunakan konsep *Cosine Similarity Measure*. *Cosine Similarity Measure* memberikan tingkat kemiripan antara vektor dokumen d_i dan *query q* dengan melakukan perhitungan besar kosinus dari sudut yang dibentuk oleh dua vektor. Ukuran kosinus sudut antara kedua vektor dapat dinyatakan sebagai :

$$\text{CosSim}(d_j, q) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2 \cdot \sum_{i=1}^t w_{iq}^2}} \dots\dots\dots (4)$$

3. HASIL DAN PEMBAHASAN

Aplikasi Sistem Temu Kembali Kelas Buku ini dikembangkan untuk membantu pustakawan dalam menganalisis subyek buku dan menentukan kelas buku.

Aplikasi Sistem Temu Kembali Informasi Kelas Buku dapat melakukan pencocokan *query* pengguna dengan koleksi dokumen yang tersimpan dalam sistem. Koleksi dokumen adalah kumpulan judul dan daftar isi buku yang tersimpan di *database* dan *query* pengguna adalah masukan pengguna yang akan dicocokkan dengan koleksi dokumen. Sistem ini juga perlu memenuhi kebutuhan informasi pengguna dengan memberikan hasil pencocokan (*matching*) berupa kelas buku yang relevan atau sesuai *query* pengguna.

Metode yang akan digunakan dalam perepresentasian *query* dan dokumen adalah metode Model Ruang Vektor. Vektor yang terbentuk dari dokumen dan *query*, tingkat kesesuaian atau relevansinya akan ditentukan dengan menggunakan *Cosine Similarity Measure* melalui proses pencocokan (*matching*). Dalam mempersiapkan koleksi dokumen untuk proses perepresentasiannya menjadi sebuah vektor, dilakukan langkah-langkah pemrosesan awal yaitu :

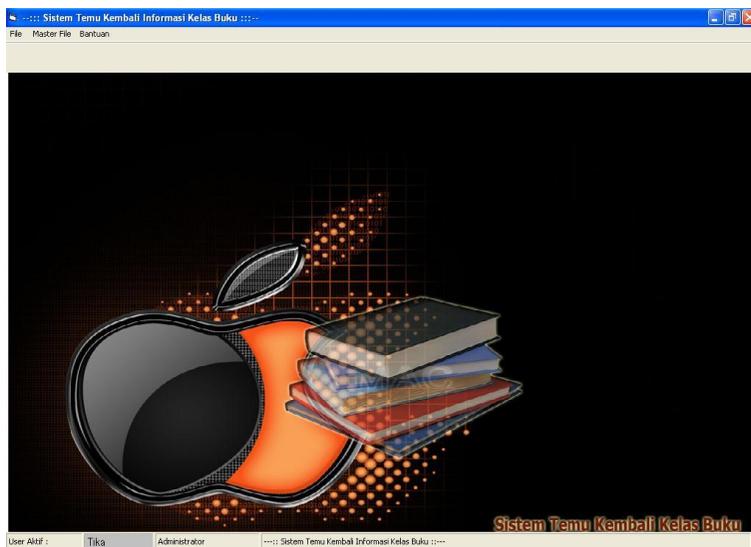
1. Dokumen dipisahkan menjadi *term-term*
2. Penghilangan *stopword* dari dokumen
3. Membangun indeks dokumen (kumpulan *term* dengan nilai *Term Frequency*, *Document Frequency*, *Invers Document Frequency* dan bobot masing-masing *keyterm*).

Metode Model Ruang Vektor dipilih karena merupakan metode yang memodelkan ruang informasi dan ruang kebutuhan secara cukup sederhana. Ukuran

relevansi *Cosine Similarity Measure* dipilih karena sudah memperhitungkan penghilangan pengaruh panjang dokumen pada proses pencocokan.

Aplikasi ini memungkinkan pengguna untuk menambahkan koleksi dokumen baru. Koleksi dokumen baru akan diindeks kembali untuk menentukan kata kunci (*key term*) untuk pencocokan dengan *query* pengguna.

Aplikasi Sistem Temu Kembali Informasi Kelas Buku terdiri dari beberapa *form*, salah satu *form* pada sistem ini adalah *form* utama. *Form* utama memiliki fasilitas menu yang memungkinkan pengguna untuk menentukan fasilitas yang tersedia pada sistem. Gambar 3 dibawah ini adalah tampilan *form* utama dari aplikasi Sistem Temu Kembali Informasi Kelas Buku



Gambar 3 *Form* Utama

4. PENUTUP

Sistem Temu Kembali Informasi Kelas Buku merupakan perangkat lunak untuk membantu proses menganalisis subyek sebuah buku dalam menentukan nomor klasifikasi buku di perpustakaan. Sistem ini mengelola masukan yang menjadi pernyataan kebutuhan informasi pengguna (*query*) yang akan dicocokkan dengan koleksi dokumen yang dimiliki sistem. Koleksi dokumen adalah deskripsi beberapa buku (judul dan daftar isi) disimpan dalam basis data. Sistem akan menampilkan hasil pencocokan (*matching*) antara *query* pengguna dan koleksi dokumen berupa kelas buku yang relevan atau sesuai dengan *query* pengguna.

Sistem Temu Kembali Informasi Kelas Buku dapat dikembangkan menggunakan metode *Information Retrieval* lain, misalnya dengan Model Probabilistik, pendekatan *Neural Network*, dan lain-lain. Dalam Sistem Temu Kembali Informasi Kelas Buku hanya dikembangkan fasilitas penghilangan *stopword*. Fasilitas *stemming* dalam sistem ini tidak dikembangkan karena perlu diperhatikan kata (*term*) dalam sistem klasifikasi buku tidak selalu dalam bentuk dasar.

Pengguna Sistem Temu Kembali Informasi Kelas Buku harus memperhatikan masukan koleksi dokumen agar kata kunci (*key term*) hasil pengindeksan sistem untuk tiap kelas buku bersifat unik.

DAFTAR PUSTAKA

- Bafadal, Ibrahim. 2005. "Pengelolaan Perpustakaan Sekolah". Jakarta: Bumi Aksara.
- Kent, A. 1971. "Information Analysis and Retrieval, 3 rd Edition". New York: Becker and Heys.
- Lancaster, F.W. 1979. "Information Retrieval Systems : Characteristics, Testing, and Evaluation, 2 nd Edition". New York: John Wiley.
- Miswan. 2003. "Klasifikasi dan Katalogisasi : sebuah pengantar". http://www.researchkesos.com/download/jurnal_vol_14.pdf. (Diakses Tanggal 1 Oktober 2009).
- Rila Mandala. 2006. "Evaluasi Kinerja Sistem Penyaringan Informasi Model Ruang Vektor" . <http://journal.uii.ac.id/index.php/Snati/article/viewFile/1547/1323> . (Diakses Tanggal 15 Oktober 2009).
- Salton, G. 1983. "Introduction to Modern Information Retrieval". New York: McGraw-Hill Book Company.
- Soetaminah. 1992. "Perpustakaan Kepustakawanan dan Perpustakaan". Yogyakarta: Kanisius.
- Supono. 2007. "Melvil Dewey, Si Bapak Perpustakaan". <http://supono.wordpress.com/2007/10/17/mevil-dewey-si-bapak-perpustakaan>. (Diakses Tanggal 1 Oktober 2009).