1. PENCARIAN DOKUMEN TEKS ARSIP SURAT DENGAN METODE INDEXING DAN QUERY

Dalam sebuah organisasi, berkas surat yang sedikit biasnya belum menjadi masalah namun setelah data surat yang sudah dibuat bertumpuk - tumpuk bahkan sampai bertahun -tahun dengan jumlah yang banyak menjadi bingung. Permasalah tersebut maka dilakukan dengan proses penemuan kembali dengan indeks agar pada saat-saat tertentu data surat dapat ditemukan dengan baik. Metode indeks prinsipnya mengambil indeks yang ada pada buku. Buku yang tebal mempunyai indeks langsung dapat menemukan di dalam buku tersebut ada kata atau kalimat apa saya dapat ditemukan. Pada pengarsipan surat maka dapat dilakukan indeks yang sudah tersimpan dengan penyimpanan, dan dapat namun terlebih dahulu harus dioleh ketahapan awal pemecahan kata dalam suatu teks kalimat tersebut dilakukan indeks dan sekaligus tersimpan dalam database dengan tabel penyimpanan. Metodologi yang akan digunakan pada penelitian ini antara lain; Studi Literatur dan SDLC (System Development Life Cycle) yang meliputi tahap Analysis, Design, Implementation,Testing dan Maintenance. Secara garis besar sistem yang dibangun untuk mengetahui kata atau kalimat dalam suatu surat yang dalam bentuk teks dapat di dilakukan proses index kata tersebut atau kalimat tersebut terdapat pada paragrap, baris, frekwensi kemunculan dan dalam file apa tersimpan di dalam database.Hasil penyimpanan tabel disimpan secara permanen untuk dilakukan pengolahan dengan query apabila dicari data kata yang telah terindeks sesuai dengan keingingan. Pengembangan query dapat dilakukan pengolahan yang lebih lanjut dapat dilakuka pengelompokkan berdasarkan kata sehingga dapat terrangkum menjadi simple.

1. PENERAPAN FASILITAS TWAIN UNTUK DIGITALISASI KEARSIPAN PADA PT. PLN AREA FLORES BAGIAN BARAT BERBASIS CLIENT-SERVER

PT. PLN Area Flores Barat adalah salah satu Badan Usaha Milik Negara (BUMN) yang tidak terlepas dari kegiatan surat menyurat dan kearsipan lainnya. Permasalahan yang dialami oleh PT. PLN Area Flores Barat adalah proses penyimpanan arsip surat masih bersifat konvensional dan belum menerapkan sistem pengindeksan serta pemanfaatan teknologi komputer. Hal ini menyebabkan pencarian arsip dalam bentuk fisik menjadi sulit dan lambat, selain itu rusaknya kertas arsip surat dikarenakan penumpukan. Penelitian ini bertujuan untuk membuat sebuah aplikasi bantu yang dapat mempermudah proses pengarsipan dokumen berbasis client-server pada PT. PLN area Flores bagian barat secara digital dengan menerapkan fasilitas TWAIN. Metode yang digunakan dalam penelitian ini adalah metode Waterfall dengan Bahasa pemrograman Microsoft Visual Basic. Hasil penelitian ini dapat membantu dan mempermudah proses pengolahan data kearsipan pada PT. PLN Area Flores Bagian Barat sehingga diharapkan dapat meminimalisasi kehilangan dan kerusakan data serta dapat memberikan informasi yang efektif dan efisien dalam pendukung pengambilan keputusan oleh pimpinan. Dari penelitian yang dilakukan dapat disimpukan bahwa aplikasi yang dibuat dapat berjalan dengan baik dan dapat diterima dengan baik. Dengan adanya aplikasi digitalisasi kearsipan pada PT. PLN Area Flores Bagian Barat berbasis client-server dengan menggunakan fasilitas TWAIN maka akan membantu dan mengolah data kearsipan yang lebih efektif dan efisien

1. RANCANG BANGUN SISTEM INFORMASI KEARSIPAN PADA PT. ZENITH DJAJA PALEMBANG

Dalam proses penyimpanan arsip surat pada PT. Zenith Djaja Palembang masih mengalami banyak kendala, seperti dalam sistem penyimpanan arsipnya saat ini tidak menerapkan sistem pengindeksan, sehingga pencarian arsip dalam bentuk fisik menjadi sulit dan lambat, rusaknya kertas arsip surat dikarenakan penumpukan dan masih besarnya tingkat kehilangan arsip surat, serta kesulitan untuk memindahkan arsip surat yang berdasarkan klasifikasi kode arsip karena tidak tertatanya arsip sesuai dengan klasifikasi kode arsip. Untuk membantu mengidentifikasi, menganalisis dan memecahkan masalah-masalah yang dihadapi PT Zenith Djaja, penulis menggunakan kerangka PIECES (Performance, Information, Economic, Control, Efficiency and Service). Permasalahan yang muncul pada PT Zenith Djaja yaitu Performance Tidak ada sistem indeks dalam penyimpanan arsip. Penyimpanan arsip saat ini tidak terurut sesuai sistem pengindeksan arsip sehingga bagian administrasi memerlukan waktu yang cukup lama untuk menemukan arsip. Berdasarkan hasil analisis yang dilakukan oleh penulis pada PT. Zenith Djaja Palembang, maka dapat mengambil kesimpulan yaitu: Aplikasi pengarsipan yang sudah dibuat dapat menerapkan sistem indeks maka pencarian data arsip surat masuk maupun surat keluar menjadi lebih cepat dan mudah.

1. IMPLEMENTASI WEB SERVICE PADA SISTEM PENGINDEKSAN DAN PENCARIAN DOKUMEN TUGAS AKHIR, SKRIPSI, DAN PRAKTIK KERJA LAPANGAN

Implementasi web service pada sistem pengindeksan dan pencarian dokumen merupakan penyempurnaan sistem yang telah dikembangkan sebelumnya. Adapun kekurangan yang dimaksud meliputi data yang digunakan masih disediakan dan dikelola pada sistem itu sendiri yang menyebabkan manajemen data yang kurang efisien. Selain itu dalam proses pengindeksan dokumen, hasil index yang telah dibuat belum dapat diperbaharui apabila terjadi perubahan maupun penambahan dokumen baru. Berdasarkan hal tersebut, dikembangkan sistem pengindeksan dan pencarian dokumen dengan memanfaatkan web service pada sistem yang berbeda sebagai sumber data dan index yang telah dibuat dapat diperbaharui. Selain itu, sistem ini juga mengimplementasikan teknologi web service sebagai penyedia layanan yang dapat dimanfaatkan oleh pengembang maupun sistem yang berbeda. Layanan yang disediakan berupa layanan pencarian dokumen tugas akhir, skripsi dan praktik kerja lapangan. Keluaran dari layanan pencarian dokumen berupa dokumen yang sudah diberikan skor serta diurutkan berdasarkan relevansi dokumen dengan masukan berupa keywords dari pengguna. Dalam pembuatan sistem ini, terdapat dua langkah penting yang dilakukan yaitu analisis dan peracangan sistem dengan menggunakan notasi UML dimana diagram yang terlibat yaitu use case, activity diagram, class diagram dan sequence diagram. Selanjutnya rancangan tersebut diimplementasikan ke dalam bentuk kode sehingga terbentuk sistem sesuai dengan analisis dan rancangan. Hasil atau keluaran dari layanan pencarian dokumen tersebut berupa daftar dokumen yang sudah diberikan skor dan diurutkan berdasarkan relevansinya dengan keyword pencarian pengguna.

1. APLIKASI SISTEM TEMU KEMBALI DOKUMEN DENGAN METODE VECTOR SPACE MODEL

Penelitian ini akan menjelaskan nilai tambah penggunaan Vector Space Model dan tahapan-tahapan pendukungnya seperti Tokenization dan Stemming dalam upaya untuk mendapatkan nilai balik informasi yang paling relevan diantara kumpulan data cetak yang telah di digitalisasi. Pencarian dokumen dilakukan dengan dua cara, yaitu pencarian penuh (tanpa stemming) dan pencarian normal (dengan stemming). Dari percobaan yang dilakukan diketahui bahwa penggunaan stemming mampu mengurangi jumlah term\_tunggal yang harus di indeks. Tetapi penggunaan indeks dengan proses non stemming, memberikan hasil yang sedikit lebih relevan dibanding penggunaan indeks dengan proses stemming. Vector Space Model kurang dapat bekerja baik saat jumlah dokumen yang ter-indeks di basis data masih sedikit. Sehingga perlu indeks data yang besar untuk mendapatkan hasil yang lebih relevan. Beberapa teknik pendekatan yang digunakan untuk proses stemming teks berbahasa Indonesia ada yang dengan menggunakan kamus namun ada juga yang tidak menggunakan kamus. Dalam penelitian ini akan digunakan skema stemming Porter-like Stemmer. Algoritma ini berdasarkan algoritma stemming untuk Bahasa Inggris yang dimodifikasi sesuai kaidah morfologi Bahasa Indonesia. Algoritma ini diajukan oleh Fadillah Z Tala dalam thesisnya yang berjudul A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia - Master of Logic Project, Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands. kesimpulan dan saran yang berkaitan dengan topik dan tema dalam penelitian ini yaitu belum ada algoritme stemming untuk Bahasa Indonesia yang baik. Algoritme yang ada masih dalam tahap penelitian dan ujicoba oleh masing-masing peneliti. Yang kedua ialah penggunaan stemming mampu mengurangi jumlah term\_tunggal yang harus di indeks.

1. PENERAPAN SISTEM TEMU KEMBALI INFORMASI PADA KUMPULAN DOKUMEN SKRIPSI

Sistem temu kembali informasi (information retrieval system) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunanya, dengan menerapkan sistem tersebut permasalahan pencarian informasi dokumen skripsi bisa memberikan hasil yang relevan sesuai kebutuhan pengguna. Terdapat dua proses utama dalam sistem temu kembali informasi yaitu indexing dan retrieval. Proses indexing adalah proses untuk memberikan bobot pada kata dalam dokumen, metode pembobotan pada penelitian ini menggunakan metode pembobotan TF-IDF. Prosesretrieval adalah proses untuk menghitung kemiripan query terhadap dokumen, perhitungan kemiripan menggunakan konsepvector space modeldengan mencari nilai cosine similarity.Tujuan dari penelitian ini adalah untuk mengembangkan dan mengimplementasikan pengindeksan otomatis untuk membangun sistem pencarian dokumen di dalam sebuah system penyimpanan dokumen dengan konsep temu-kembali informasi. Metode TF-IDF (Term Frequncy Inverse Document Frequency) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot. Berdasarkan hasil penelitian yang dilakukan sistem temu kembali informasi berhasil dibangun dan dapat mengimplementasikan pengindeksan otomatis dengan menggunakan metode pembobotan TF-IDF dan dapat menemukan dokumen yang relevan terhadap kata kunci dengan menghitung nilai cosine similarity. Waktu proses indexing sistem berdasarkan pengujian waktu indexing, waktu proses indexing terus meningkat bersamaan dengan banyaknya dokumen yang di indexdengan waktu pengujian indexing pada 100 dokumen skripsi 328.85 detik

1. ESSENCE: A RESOURCE DISCOVERY SYSTEM BASED ON SEMANTIC FILE INDEXING

Discovering different types of file resources (such as documentation, programs, and images) in the vast amount of data contained within network file systems is useful for both users and system administrators. In this paper we discuss the Essence resource discovery system, which exploits file semantics to index both textual and binary files. By exploiting semantics, Essence extracts keywords that summarize a file, and generates a compact yet representative index. Essence understands nested file structures (such as uuencoded, compressed, ‘‘tar’’ files), and recursively unravels such files to generate summaries for them. These features allow Essence to be used in a number of useful settings, such as anonymous FTP archives. We present measurements of our prototype and compare them to related projects, such as the Wide Area Information Servers (WAIS) system and the MIT Semantic File System (SFS). We demonstrate that Essence can index more data types, generate smaller indexes, and in some cases index data faster than these systems. Our prototype generates WAIS-compatible indexes, allowing WAIS users to take advantage of the Essence indexing methods

1. COMPUTER-AIDED DOCUMENT INDEXING SYSTEM

An enormous number of documents is being produced that have to be stored, searched and accessed. Document indexing represents an efficient way to tackle this problem. Contributing to the document indexing process, we developed the Computer-Aided Document Indexing System (CADIS) that applies controlled vocabulary keywords from the EUROVOC thesaurus. The main contribution of this paper is the introduction of the special CADIS internal data structure that copes with the morphological complexity of the Croatian language. CADIS internal data structure ensures efficient statistical analysis of input documents and quick visual feedback generation that helps indexing documents more quickly, accurately and uniformly than by manual indexing. Computer-Aided Document Indexing System CADIS that provides different statistical and visual information, supporting an efficient and uniform indexing process is described in the paper. Some problems tackled during the design and implementation of the CADIS include the morphological complexity of Croatian language, strongly connected with implemented statistical functions of word and lemma count and collocation finding

1. TAGIT: AN INTEGRATED INDEXING AND SEARCH SERVICE FOR FILE SYSTEMS

Data services such as search, discovery, and management in scalable distributed environments have traditionally been decoupled from the underlying file systems, and are often deployed using external databases and indexing services. However, modern data production rates, looming data movement costs, and the lack of metadata, entail revisiting the decoupled file system-data services design philosophy. In this paper, we present TagIt, a scalable data management service framework aimed at scientific datasets, which is tightly integrated into a shared-nothing distributed file system. A key feature of TagIt is a scalable, distributed metadata indexing framework, using which we implement a flexible tagging capability to support data discovery. The tags can also be associated with an active operator, for pre-processing, filtering, or automatic metadata extraction, which we seamlessly offload to file servers in a load-aware fashion. Our evaluation shows that TagIt can expedite data search by up to 10X over the extant decoupled approach

1. RADLEX: A NEW METHOD FOR INDEXING ONLINE EDUCATIONAL MATERIALS

For several decades, the American College of Radiology’s *Index for Radiological Diagnoses* (known as the ACR Index) has served admirably as an indexing system for radiology teaching files. The ACR Index was originally developed to categorize and organize the image-based interesting cases collected by radiologists, often in paper folders on office shelves. As radiology clinical practice and radiology education move online, there is an increasing need for an indexing system that works equally well in the digital world. The Radiological Society of North America’s (RSNA’s) project RadLex is designed to address that need. This editorial discusses the ACR Index and introduces the RadLex lexicon, a new method for indexing online educational materials for radiologists and educators.