

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan Teknologi khususnya internet sangat berperan dalam kehidupan sehari-hari. Dengan adanya internet, informasi dapat dengan mudah disebarluaskan dan diakses oleh banyak orang. Banyaknya informasi yang beredar tentu membuat kebutuhan akan informasi yang relevan semakin meningkat. Salah satu cara yang bisa digunakan untuk mendapat informasi yang relevan adalah dengan informasi yang relevan adalah dengan menggunakan sistem temu kembali informasi.

Judul-judul Tugas Akhir Teknik Elektro, selama ini belum terkoordinasi dengan baik, dimana pada calon mahasiswa tugas akhir (Semester 8) biasanya mempunyai kendala dalam hal membuat judul Tugas Akhir, terkadang judul tersebut diajukan di meja hijau, ternyata judul tersebut sudah ada, hal ini yang mendasari penulis untuk mengangkat permasalahan yang ada selama ini

Pembuatan search engine tugas akhir ini akan mencari data tugas akhir berdasarkan judul yang akan di upload dan di simpan ke dalam database yang kemudian akan dilakukan pembobotan dengan menggunakan Metode TF/IDF sehingga data yang di dapatkan dapat di tampilkan berdasarkan bobot dari hasil perhitungan TF/IDF.

Berdasarkan tujuan permasalahan diatas maka saya mengangkat judul Aplikasi Mesin Pencari Tugas Akhir Politeknik Negeri Manado Jurusan Elektro menggunakan Metode TF/IDF

1.2. Rumusan Masalah

Perumusan masalah dalam pembuatan Tugas Akhir pada ini adalah sebagai berikut:

1. Bagaimana membuat sebuah aplikasi yang dapat digunakan untuk menyimpan data tugas akhir.
2. Bagaimana membuat aplikasi yang dapat melakukan pencarian data tugas akhir berdasarkan data judul yang di simpan.
3. Bagaimana melakukan proses Indexing data judul yang telah di simpan.
4. Bagaimana mengimplementasikan metode TF/IDF pada mesin pencari tugas akhir.

1.3. Batasan Masalah

Agar pembahasan tentang aplikasi *Search Engine* Tugas Akhir ini tidak terlalu meluas, namun hasil yang di harapkan dapat tercapai secara optimal, maka ruang lingkup penelitian dibatasi pada proses :

1. Membahas tentang proses menyimpan dokumen Tugas Akhir pada aplikasi.
2. Membahas tentang proses menyimpan data judul Tugas Akhir kedalam database.
3. Membahas tentang proses *Indexing* yang hanya meliputi *Stemming, Filtering, Tokenizing* data judul tugas akhir yang telah disimpan ke dalam database.
4. Membahas tentang proses pembobotan pencarian dengan menggunakan metode TF/IDF.
5. Aplikasi tidak dapat melakukan pembacaan file PDF secara langsung.

1.4. Manfaat Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah :

1. Mempermudah dalam pencarian informasi Tugas Akhir karena dapat mempresentasikan hasil informasi secara terurut berdasarkan kemiripan antara query dengan informasi yang ada pada dokumen Tugas Akhir.

2. Bagaimana tidak terjadi tumpang tindih judul Tugas Akhir Mahasiswa, seperti kemiripan, fungsi dan metode.

1.5. Tujuan Penelitian

Tujuan yang Akan dicapai Dalam Penelitian ini adalah:

1. Aplikasi ini akan menjadi suatu patokan ketika setiap mahasiswa kesulitan atau bimbang dengan judul yang ada pada mahasiswa tingkat akhir, sudah pernah di angkat di tahun-tahun sebelumnya.
2. Menerapkan algoritma TF-IDF yang dapat digunakan untuk mencari Tugas Akhir sebagai pengukur tingkat similaritas antara dokumen dengan *keyword* yang didapat dari ekstraksi teks pada dokumen sehingga mendapatkan data yang terurut dari kemiripannya (tingkat similaritas) paling tinggi sehingga pencarian tugas akhir menjadi lebih efisien sebagai informasi yang relevan.

1.6. Metodologi Penelitian

Metodologi yang digunakan dalam pembuatan aplikasi ini adalah :

1. Studi Literatur
Berupa studi kepustakaan dan kajian dari buku - buku teks pendukung.
2. Analisa dan Perancangan Sistem
Perancangan aplikasi merubah sistem manual yang sudah lama dan mengubah menjadi sistem yang baru dengan lebih terstruktur dan mudah di dapat.

1.7. Sistematika Penulisan

Untuk memberikan gambaran yang lebih jelas dan sistematis, skripsi ini dibagi menjadi lima bab dan tiap bab memiliki beberapa sub bab dengan urutan sebagai berikut:

- BAB I Pendahuluan

Bab ini menguraikan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian dan sistematika penulisan.

- BAB II Landasan Teori

Pada bab ini akan menjelaskan tentang teori-teori dengan pembuatan aplikasi ini.

- BAB III Metodologi Penelitian

Pada bab ini akan menjelaskan waktu dan tempat penelitian, dan menjelaskan tahap-tahap analisa dan perancangan Aplikasi Search Engine Tugas Akhir Politeknik Negeri Manado Jurusan Elektro dengan menggunakan Metode TF/IDF.

- BAB VI Pembahasan

Pada bab ini akan diuraikan hasil pengujian dari perancangan aplikasi dari setiap implementasi serta menyajikan informasi pengujian sistem.

- BAB V Kesimpulan dan saran

Bab ini berisi kesimpulan dari hasil yang didapat dan juga saran untuk pengembangan bagi peneliti lain tentang pengembangan dan penyempurnaan aplikasi

BAB II

TINJAUAN PUSTAKA

2.1. Data

Data merupakan fakta atau bagian dari fakta yang mengandung arti, yang dihubungkan dengan kenyataan, gambar-gambar, kata-kata, angka-angka, huruf atau simbol-simbol menyatakan suatu ide objek kondisi atau situasi dan lain-lain.

Data adalah kenyataan yang menggambarkan suatu kejadian-kejadian dan kesatuan nyata. Kejadian adalah sesuatu yang terjadi pada saat tertentu, contohnya: transaksi. Kesatuan nyata adalah berupa suatu objek nyata seperti tempat, benda dan orang yang betul-betul ada dan terjadi.

Data adalah bentuk yang masih mentah yang belum dapat bercerita banyak, sehingga perlu diolah lebih lanjut. Data diolah melalui suatu model untuk dihasilkan informasi.

Jenis-jenis data dapat dibagi berdasarkan sifatnya, sumbernya, cara memperolehnya, dan waktu pengumpulannya. Menurut sifatnya, jenis-jenis data yaitu:

- Data Kualitatif: data kualitatif adalah data yang tidak berbentuk angka, misalnya: Kuesioner Pertanyaan tentang suasana kerja, kualitas pelayanan sebuah rumah sakit atau gaya kepemimpinan, dan lain-lain.
- Data Kuantitatif: data kuantitatif adalah data yang berbentuk angka, misalnya: harga saham, besarnya pendapatan, dan lain-lain.

Jenis-jenis data menurut sumbernya, antara lain:

- Data Internal: data internal adalah data dari dalam suatu organisasi yang menggambarkan keadaan organisasi tersebut. Contohnya: suatu perusahaan, jumlah karyawannya, jumlah modalnya, atau jumlah produksinya, dan lain-lain.
- Data Eksternal: data eksternal adalah data dari luar suatu organisasi yang dapat menggambarkan faktor-faktor yang mungkin mempengaruhi hasil

kerja suatu organisasi. Misalnya: daya beli masyarakat mempengaruhi hasil penjualan suatu perusahaan.

Jenis-jenis data menurut cara memperolehnya, antara lain:

- Data Primer (primary data): data primer adalah data yang dikumpulkan sendiri oleh perorangan/suatu organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan yang dapat berupa interview, observasi.
- Data Sekunder (secondary data): data sekunder adalah data yang diperoleh/dikumpulkan dan disatukan oleh studi-studi sebelumnya atau yang diterbitkan oleh berbagai instansi lain. Biasanya sumber tidak langsung berupa data dokumentasi dan arsip-arsip resmi.

2.2. Informasi

Informasi adalah rangkaian data yang mempunyai sifat sementara, tergantung dengan waktu, mampu memberi kejutan atau *surprise* pada yang menerimanya. Informasi dapat juga dikatakan sebagai data yang telah diproses, yang mempunyai nilai tentang tindakan atau keputusan.

2.3. Basis Data

Basis data adalah sekumpulan data tersebar yang berhubungan secara logis, dan penjelasan dari data ini dirancang untuk memenuhi kebutuhan informasi dari suatu organisasi.

Basis data adalah sekumpulan data yang saling berhubungan yang disimpan (biasanya dengan redundansi yang terkontrol dan terbatas) berdasarkan skema. Sistem Informasi

Alasan diperlukan Basis data :

- Salah satu komponen penting dalam sistem informasi, karena merupakan dasar dalam menyediakan informasi

- Menentukan kualitas informasi : akurat, tepat pada waktunya dan relevan. Informasi dapat dikatakan bernilai bila manfaatnya lebih efektif dibandingkan dengan biaya mendapatkannya.
- Mengurangi duplikasi data (*data redundancy*).
- Hubungan data dapat ditingkatkan (*data reliability*).
- Mengurangi pemborosan tempat simpanan luar.

2.4. Perancangan Sistem

Perancangan sistem berarti menyusun suatu sistem yang digunakan untuk maksud tertentu dalam mencapai tujuan, biasanya sistem tersebut mempunyai kelebihan-kelebihan diantaranya lebih efisien, tepat waktu dan relevan. Adapun alat-alat bantu dalam perancangan sistem yaitu

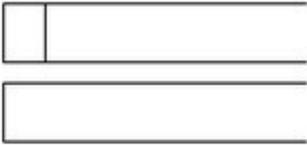
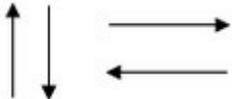
2.4.1. Diagram Alir Data (DAD atau DFD)

Diagram alir data digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru akan dikembangkan secara logika tanpa mempertimbangkan lingkungan fisik dimana data mengalir dan data disimpan. Diagram alir data merupakan alat yang digunakan pada metodologi pengembangan sistem yang terstruktur

Menurut Raymond Me Leod, Jr. Diagram Alir Data merupakan alat pembuatan model yang memungkinkan profesional sistem untuk menggambarkan sistem sebagai suatu jaringan proses fungsional yang dibutuhkan satu sama lain dengan alur data baik secara manual atau terkomputerisasi. Data Flow Diagram atau Diagram Arus Data adalah suatu gambaran grafis dan suatu sistem yang menggunakan sejumlah bentuk-bentuk symbol untuk menggambarkan bagaimana data mengalir melalui suatu proses yang berkaitan. Diagram alir data digunakan dalam diagram context dan diagram level n

Simbol atau lambang yang digunakan dalam membuat diagram alir data yang lazim digunakan, terdiri dari empat buah symbol yaitu :

Tabel 2.1 Simbol-simbol dalam DAD

Simbol	Keterangan
	<i>External Entity</i> , merupakan kesatuan di lingkungan luar sistem yang bisa berupa orang, organisasi atau sistem lain.
	<i>Process</i> , merupakan proses seperti perhitungan aritmatik penulisan suatu formula atau pembuatan laporan
	<i>Data Store (Simpan Data)</i> , dapat berupa suatu file atau database pada sistem komputer atau catatan manual
	<i>Data Flow (arus data)</i> , arus data ini mengalir diantara proses, simpan data dan kesatuan luar

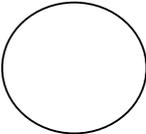
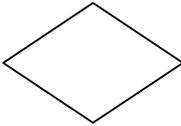
2.4.2. ERD

Entity Relationship Diagram adalah suatu model konseptual yang mendeskripsikan hubungan antara penyimpanan data (file data). ERD digunakan untuk memodelkan struktur data dan hubungan antar data serta untuk menguji model dengan menyampaikan proses yang harus dilakukan. Adapun Pendapat menurut para ahli :

- **Pengertian ERD Menurut Nugroho (2002)** Adalah diagram yang memperlihatkan entitas-entitas yang terlibat dalam suatu sistem serta hubungan-hubungan (relasi) antar entitas tersebut
- **Pengertian ERD Menurut Jogiyanto (2001 : 700)** Adalah suatu komponen-komponen himpunan entitas dan himpunan relasi yang masing-masing dilengkapi dengan atribut yang mempresentasikan seluruh fakta dari dunia nyata yang ditinjau. Entity Relation Diagram menggambarkan data dan

hubungan antar data secara global dengan menggunakan Entity Relation Diagram.

Tabel 2.2 Simbol-simbol ERD

Nama Simbol	Simbol
<p>ENTITAS</p> <p>Digunakan untuk menggambarkan obyek yang dapat diidentifikasi dalam lingkungan pemakai.</p>	
<p>ATIBUT</p> <p>Digunakan untuk menggambarkan elemen-elemen dari suatu entitas, yang menggambarkan karakter entitas.</p>	
<p>HUBUNGAN</p> <p>Entitas dapat berhubungan satu sama lain. Hubungan ini disebut relasi.</p>	
<p>GARIS</p> <p>Digunakan untuk menghubungkan entitas dan entitas dengan atribut.</p>	

Hubungan Relasi Dalam ERD

- **One To One** : Hubungan antara *file* satu dengan *file* yang kedua adalah satu berbanding satu
- **One To Many** : Hubungan antara *file* yang satu dengan *file* yang kedua adalah satu berbanding banyak atau banyak berbanding satu.

- **Many To Many** : Hubungan antara *file* yang satu antara *file* yang kedua adalah banyak berbanding banyak.

2.4.3. Flowchart

Flowchart adalah penggambaran secara grafik dari langkah-langkah dan urutan prosedur dari suatu program. Flowchart menolong analis dan programmer untuk memecahkan masalah kedalam segmen-segmen yang lebih kecil dan menolong dalam menganalisis alternatif-alternatif lain dalam pengoperasian. Flowchart biasanya mempermudah penyelesaian suatu masalah khususnya masalah yang perlu dipelajari dan dievaluasi lebih lanjut

- **Jenis-jenis Flowchart**

Flowchart terbagi atas lima jenis, yaitu :

1. Flowchart Sistem (*Sistem Flowchart*)
2. Flowchart Paperwork / Flowchart Dokumen (*Document Flowchart*)
3. Flowchart Skematik (*Schematic Flowchart*)
4. Flowchart Program (*Program Flowchart*)
5. Flowchart Proses (*Process Flowchart*)

- **Flowchart Sistem**

Flowchart Sistem merupakan bagan yang menunjukkan alur kerja atau apa yang sedang dikerjakan di dalam sistem secara keseluruhan dan menjelaskan urutan dari prosedur-prosedur yang ada di dalam sistem. Dengan kata lain, flowchart ini merupakan deskripsi secara grafik dari urutan prosedur-prosedur yang terkombinasi yang membentuk suatu sistem.

- **Flowchart Program**

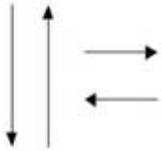
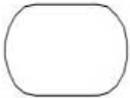
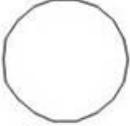
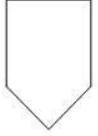
Flowchart Program dihasilkan dari Flowchart Sistem. Flowchart program merupakan keterangan yang lebih rinci tentang bagaimana setiap langkah program atau prosedur sesungguhnya dilaksanakan. Flowchart ini menunjukkan setiap langkah program atau prosedur dalam urutan yang

tepat saat terjadi. Programmer menggunakan flowchart program untuk menggambarkan urutan instruksi dari program komputer. Analisis Sistem menggunakan flowchart program untuk menggambarkan urutan tugas-tugas pekerjaan dalam suatu prosedur atau operasi.

1. Simbol-Simbol Flowchart

Simbol-simbol flowchart yang biasanya dipakai adalah simbol-simbol flowchart standar yang dikeluarkan oleh ANSI dan ISO.

Tabel 2.3 Simbol-simbol Flowchart

Simbol	Arti
	<p>Flow Direction symbol</p> <p>Yaitu simbol yang digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga connecting line.</p>
	<p>Terminator Symbol</p> <p>Yaitu simbol untuk permulaan (start) atau akhir (stop) dari suatu kegiatan</p>
	<p>Connector Symbol</p> <p>Yaitu simbol untuk keluar – masuk atau penyambungan proses dalam lembar / halaman yang sama.</p>
	<p>Connector Symbol</p> <p>Yaitu simbol untuk keluar – masuk atau penyambungan proses pada lembar / halaman yang berbeda</p>

	<p>Processing Symbol</p> <p>Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer</p>
	<p>Simbol Manual Operation</p> <p>Simbol yang menunjukkan pengolahan yang tidak dilakukan oleh komputer</p>
	<p>Simbol Decision</p> <p>Simbol pemilihan proses berdasarkan kondisi yang ada.</p>
	<p>Simbol Input-Output</p> <p>Simbol yang menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya</p>
	<p>Simbol Manual Input</p> <p>Simbol untuk pemasukan data secara manual on-line keyboard</p>
	<p>Simbol Preparation</p> <p>Simbol untuk mempersiapkan penyimpanan yang akan digunakan sebagai tempat pengolahan di dalam storage</p>
	<p>Simbol Predefine Proses</p> <p>Simbol untuk pelaksanaan suatu bagian (sub-program)/prosedure</p>

	<p>Simbol Display</p> <p>Simbol yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya</p>
	<p>Simbol disk and On-line Storage</p> <p>Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk</p>

2.5. Hypertext Markup Language (HTML)

HTML (Hypertext Markup Language) adalah sebuah protokol yang digunakan untuk membuat format suatu dokumen web yang mampu dibaca dalam browser dari berbagai platform komputer. Sifat bahasa HTML ini adalah client script, dimana dokumen tersebut dapat dibuka dalam komputer stand alone yang tidak membutuhkan server untuk dapat menampilkannya di dalam browser.

2.6. Hypertext Preprocessor (PHP)

PHP merupakan kependekan dari kata Hypertext Preprocessor. PHP tergolong sebagai perangkat lunak open source yang di atur dalam aturan general purpose licences (GPL). PHP di khususkan untuk pengembangan web dinamis.

Pada umumnya pembuatan web dinamis berhubungan erat dengan *database* sebagai sumber data yang akan ditampilkan. PHP tergolong juga sebagai bahasa pemrograman yang berbasis server (server side scripting). Ini berarti bahwa semua *script* PHP diletakkan di server dan diterjemahkan oleh web server terlebih dahulu, kemudian hasil terjemahan itu dikirim ke browser client. Secara teknologi, bahasa pemrograman PHP memiliki kesamaan dengan bahasa ASP (Active server page), Cold Fusion, JSP (Java Server Page), atau Perl.

2.7. MySQL

MySQL adalah sebuah perangkat lunak sistem manajemen basis data SQL (*database management sistem*) atau DBMS yang *multithread*, *multi-user*, dengan sekitar 6 juta instalasi di seluruh dunia. MySQL AB membuat MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi *GNU General Public License* (GPL), tetapi mereka juga menjual dibawah lisensi komersial untuk kasus-kasus dimana penggunaannya tidak cocok dengan penggunaan GPL.

MySQL sebenarnya merupakan turunan salah satu konsep utama dalam database sejak lama, yaitu SQL (*Structured Query Language*). SQL adalah sebuah konsep pengoperasian database, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis.

MySQL juga memiliki beberapa kelebihan, antara lain :

1. Portability

MySQL dapat berjalan stabil pada berbagai sistem operasi seperti windows, Linux, FreeBSD, Solaris dan lain-lain.

2. Open Source

MySQL didistribusikan secara open source (gratis), dibawah lisensi GPL sehingga dapat digunakan cuma-cuma.

3. Multi User

MySQL dapat digunakan oleh beberapa user dalam waktu yang bersamaan tanpa mengalami masalah atau konflik.

4. Performance Tuning

MySQL memiliki kecepatan yang menakjubkan dalam menangani query sederhana, dengan kata lain dapat memproses lebih banyak SQL per satuan waktu.

5. Coloumn Types

MySQL memiliki tipe kolom yang sangat kompleks, seperti integer, double, char, text, datedan lain-lain.

6. Command and Function

MySQL memiliki operator dan fungsi secara penuh yang mendukung perintah select dan where dalam query.

7. Security

MySQL memiliki beberapa lapisan sekuritas seperti level subnetmask, nama host, dan izin akses user dengan sistem perizinan yang mendetail serta password terenkripsi.

8. Scability and Limits

MySQL mampu menangani database dalam skala besar, dengan jumlah records lebih dari 50 juta dan 60 ribu tabel serta 5 milyar baris. Selain itu batas indeks yang dapat ditampung mencapai 32 indeks pada tiap tabelnya.

9. Connectivity

MySQL dapat melakukan koneksi dengan clients menggunakan protokol TCP/IP, Unix socket (UNIX) atau Named Pipes (NT).

10. Localisation

MySQL dapat mendeteksi pesan kesalahan pada client dengan menggunakan lebih dari dua puluh bahasa. Meskipun demikian, bahasa Indonesia belum termasuk didalamnya.

11. Interface

MySQL memiliki interface (antar muka) terhadap berbagai aplikasi dan bahasa pemrograman dengan menggunakan fungsi API (Application Programming Interface).

12. Clients and Tools

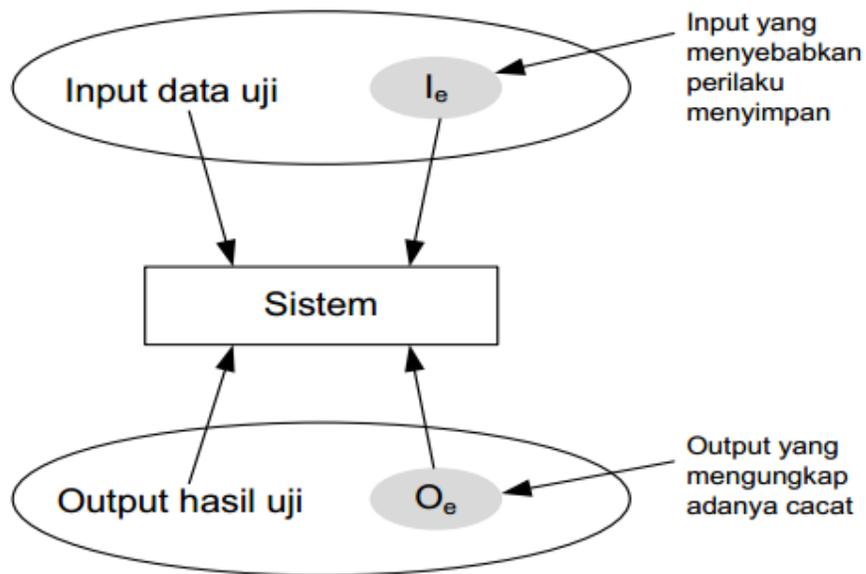
MySQL dilengkapi dengan berbagai tools yang dapat digunakan untuk administrasi database dan pada setiap tool yang ada disertakan petunjuk online.

13. Struktur Tabel

MySQL memiliki struktur tabel yang lebih fleksibel dalam menangani ALTER TABLE, dibandingkan database lainnya semacam PostgreSQL ataupun Oracle.

2.8. Black Box Testing (pengujian kotak hitam)

Pengujian fungsional atau pengujian kotak hitam (black-box testing) merupakan pendekatan pengujian yang ujinya diturunkan dari spesifikasi program atau komponen. Sistem merupakan „kotak hitam“ yang perilakunya hanya dapat ditentukan dengan mempelajari input dan output yang berkaitan. Nama lain untuk cara ini adalah pengujian fungsional karena pengujian hanya berkepentingan dengan fungsionalitas dan buka implementasi perangkat lunak.



Gambar 2.1 Peraga Pengujian Kotak Hitam

Sumber : Buku Software Engineering (Rekayasa Perangkat Lunak)/Edisi 6/Jilid 2, 2003.

2.9. Notepad ++

Notepad++ adalah salah satu text editor yang paling banyak di gunakan untuk menuliskan sebuah program. selain karena ringan dan mudah digunakan , notepad++ juga gratis.

Notepad++ dibuat dengan bahasa C ++ dan menggunakan Win32 API juga STL yang menjamin kecepatan eksekusi lebih tinggi dan ukuran program yang lebih kecil. sehingga dapat mempermudah kita dalam menulis dan menjalankan program yang sedang kita bangun.

Berikut beberapa fitur yang ada di Notepad++ :

- Sintaks highlighted dan Folding
- PCRE (Perl Compatible Regular Expression) Cari / Ganti
- GUI sepenuhnya disesuaikan: minimalis, tab dengan tombol tutup, tab multi-line, tab vertikal dan daftar dokumen vertikal
- Map Documents
- Auto-completion
- Multi-Document
- Multi-View
- WYSIWYG (what you see it what you get)
- Zoom in dan zoom out
- mendukung berbagai bahasa pemrograman
- Bookmark
- Macro recording dan playback

2.10. XAMPP

XAMPP adalah sebuah software web server apache yang didalamnya sudah tersedia database server MySQL dan dapat mendukung pemrograman PHP. XAMPP merupakan software yang mudah digunakan, gratis dan mendukung instalasi di Linux dan Windows.

Keuntungan lainnya adalah cuma menginstal satu kali sudah tersedia Apache Web Server, MySQL Database Server, PHP Support (PHP 4 dan PHP 5) dan beberapa module lainnya.