UNIVERSITAS INDONESIA

IMPLEMENTASI SISTEM INFORMASI SITASI JURNAL ELEKTRONIK INDONESIA DARI BANYAK SUMBER *E-JOURNAL*

SKRIPSI

SUBHAN MUBARAK
0706199956

FAKULTAS TEKNIK UNIVERSITAS INDONESIA
PROGRAM STUDI TEKNIK ELEKTRO
DEPOK
JUNI 2010

UNIVERSITAS INDONESIA

IMPLEMENTASI SISTEM INFORMASI SITASI JURNAL ELEKTRONIK INDONESIA DARI BANYAK SUMBER *E-JOURNAL*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

SUBHAN MUBARAK
0706199956

FAKULTAS TEKNIK UNIVERSITAS INDONESIA
PROGRAM STUDI TEKNIK ELEKTRO
DEPOK
JUNI 2010

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	**: Subhan Mubarak**
NPM	**: 0706199956**
Tanda Tangan	**: .....**
Tanggal	**: 30 Juni 2010**

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :
Nama : Subhan Mubarak
NPM : 0706199956
Program Studi : Teknik Elektro
Judul Skripsi : **IMPLEMENTASI SISTEM INFORMASI SITASI JURNAL ELEKTRONIK INDONESIA DARI BANYAK SUMBER E-JOURNAL**

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Teknik pada Program Studi Elektro Fakultas Teknik, Universitas Indonesia.

DEWAN PENGUJI

Pembimbing : **Prof. Dr. Ir. Riri Fitri Sari M.Sc., MM.** (.....)
Penguji : **Prof. Dr. Ir. Bagio Budiardjo M.Sc** (.....)
Penguji : **Prof. Dr. Ir. Kalamullah Ramli M.Eng** (.....)

Ditetapkan di : Depok
Tanggal : 30 Juni 2010

UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Allah SWT, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik pada Fakultas Teknik Universitas Indonesia. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan seminar ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

(1) Prof. Dr. Ir. Riri Fitri Sari,MSc.,MM., selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini;
(2) orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral; dan
(3) sahabat yang telah banyak membantu saya dalam menyelesaikan skripsi ini.

Akhir kata, saya berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Depok, 30 Juni 2010

Subhan Mubarak

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Subhan Mubarak
NPM : : 0706199956
Program Studi : Teknik Elektro
Departemen : Elektro
Fakultas : Teknik
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

"IMPLEMENTASI SISTEM INFORMASI SITASI JURNAL ELEKTRONIK INDONESIA DARI BANYAK SUMBER *E-JOURNAL*"

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok
Pada tanggal : 30 Juni 2010
Yang menyatakan

(Subhan Mubarak)

ABSTRAK

Nama : Subhan Mubarak
Program Studi : Teknik Elektro
Judul : Implementasi Sistem Informasi Sitasi Jurnal Elektronik Indonesia dari Banyak Sumber E-Journal

Seiring dengan berkembangnya teknologi informasi dan komunikasi, maka publikasi atau penyebaran informasi yang semula melalui kertas (cetak/*hard-copy*) kini mulai berganti menjadi bentuk elektronik (*soft-copy*). Perubahan teknologi ini membuka peluang besar bagi penyebaran informasi ilmiah tersebut, terutama jika dapat diakses secara *online*. Fokus tugas akhir ini adalah perancangan dan implementasi dari aplikasi sistem informasi sitasi jurnal elektronik dari banyak sumber e-journal. Sistem ini bertujuan untuk memudahkan seseorang dalam mengelola pengumpulan serta pencarian artikel jurnal ilmiah. Sumber e-journal yang dipakai pada sistem ini dari 12 universitas yang ada di Indonesia.

Aplikasi sistem informasi sitasi ini bekerja dengan menggunakan beberapa metode web. Metode web ekstraksi untuk mengutip data atau informasi dari suatu website, metode mashup untuk menggabungkan data dari dua atau lebih sumber data yang telah dikutip, dan metode indeks sitasi untuk menunjukan keterkaitan asal-usul atau sumber suatu kutipan yang tercantum didalam suatu karya tulis. Menggunakan program PHP untuk bahasa pemograman yang dipakai dan MySQL sebagai database.

Aplikasi sistem informasi sitasi ini telah di uji coba oleh pemakai. Hal-hal yang menjadi topik pengujian adalah pengujian pencarian (*searching*) kata kunci pada halaman utama, lama waktu proses ekstraksi memerlukan waktu tercepat 38 detik dan waktu terlama 6899 detik. Untuk memori database memerlukan kapasitas 10 Mb. Pada analisa pressicion di dapatkan data bahwa ketepatan suatu pencarian tidak dapat mencapai 100%, semakin banyak data jurnal yang di miliki dalam database, semakin tidak dapat mencapai ketepatan 100% (persen).

Kata kunci : E-journal, Ekstraksi Web, Jurnal, *Mashup,* Indek sitasi, *Pressicion*.

ABSTRACT

Name : Subhan Mubarak
Study Program : Electrical Engineering
Title : Implementation of Indonesia Electronic Journal Citation Information Systems using Multiple E-Journal Sources

Along with the development of information and communication technology, the publication or dissemination of information, initially through the paper (printed/hard-copy) is now turning into an electronic form (soft-copy). Technological change opens great opportunities for the dissemination of scientific information, especially if it can be accessed online. The focus of this thesis is to design and implement application of systems journal citation information from many sources of electronic e-journal. This system aims to facilitate a person to manage the collection and the search for scientific journal articles. E-journal sources used in this system of 12 universities in Indonesia.

This citation information system works by using several methods on the web. Web extraction method for citing data or information from a website, mashup methods for combining data from two or more data sources that have been cited, and citation index method to show the linkage origin or source of a quote contained in a paper. Using PHP program that is used for programming language and MySQL as database.

The application of citation information system has been tested. The topic of testing is the search keywords on the main page. The fastest query extraction process takes 38 seconds and the longest takes 6899 seconds. The memory capacity required for the database is 10 Mbytes. The pressicion analysis in getting the data show that the accuracy of the search. The more data in journals database, they become less able to achieve 100% accuracy.

Keywords : E-journal, Web Extraction, Journal, Mashup, Citation index, Pressicion.

DAFTAR ISI

Halaman

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Jurnal sebagai salah satu hasil pengetahuan yang terwujud dan terangkum dalam tulisan-tulisan ilmiah memungkinkan siapa saja yang mempunyai karya tulis dapat memasukkan karyanya. Jurnal yang kita kenal biasanya berupa buletin atau majalah ilmiah yang diterbitkan oleh institusi tertentu. Namun terdapat kelemahan dalam jurnal konvesional tersebut yaitu, terbatasnya karya ilmiah yang akan dimuat sehingga membuat karya ilmiah yang diterima harus diseleksi terlebih dahulu dan terbatasnya pendanaan dalam penerbitan. Ini membuat jurnal konvensional tidak dapat terbit secara berkala dalam waktu yang singkat sehingga jelas membatasi tersampaikannya karya ilmiah.

Untuk peningkatan investasi penelitian serta pengembangan ilmu dan teknologi maka diperlukan pengelolaan pengetahuan yang tepat. Pengelolaan pengetahuan (*knowledge management)* adalah upaya bagaimana manusia dapat mengumpulkan aset pengetahuan *(knowledge asset)* dan kemudian menggunakannya untuk mendapatkan keunggulan kompetitif. Dengan demikian maka teknologi informasi dan komunikasi sangat berperan besar dalam membuat masyarakat menjadi pintar. Upaya pemgembangan ilmu dilakukan melalui pertukaran pengetahuan dengan mudah dan cepat yang pada gilirannya akan membuat pengetahuan terus berkembang.

Dengan pengelolaan ilmu yang tepat di dunia pendidikan, maka akan meningkatkan kualitas Sumber Daya Manusia. Penulisan karya-karya ilmiah, dan hasil penelitian tentunya membutuhkan suatu wadah publikasi yang dapat mengakomodasi secara cepat, merata dan mudah diperbarui. Hal ini dimaksudkan agar informasi ilmiah yang terkandung di dalamnya dapat tersampaikan ke masyarakat luas sebagai upaya pemberdayaan dan peningkatan mutu SDM khususnya dalam dunia pendidikan.

Seiring dengan berkembangnya teknologi informasi dan komunikasi, maka publikasi atau penyebaran informasi yang semula melalui kertas (cetak/*hard-copy*) kini mulai berganti menjadi bentuk elektronik (*soft-copy*). Perubahan

teknologi ini membuka peluang besar bagi penyebaran informasi ilmiah tersebut, terutama jika dapat diakses secara *online* dan dibangun pusat informasinya. Kemampuan dan kemudahan teknologi tersebut memberi peluang yang sangat luas bagi terbangunnya pusat publikasi karya ilmiah berbasis TIK atau akan kita sebut sebagai *e-journal*. Dan demikian, akan memberi peluang bagi peningkatan mutu sumber daya manusia dengan memberi kemudahan dalam memperoleh ilmu pengetahuan tersebut.

Langkah awalnya yaitu kita membuat portal web sendiri yang di dalamnya adalah kumpulan-kumpulan dari portal jurnal yang ada di universitas di Indonesia, dengan cara mengekstraksi portal-portal jurnal yang ada, lalu menyimpan hasil dari ekstraksi tersebut dalam database tersendiri, yang nantinya hasil dari ekstraksi ini sebagai kata kunci dalam pencarian keterkaitan jurnal.

Dalam metode ekstraksi ini merupakan cara yang sangat rapi dalam mengumpulkan jurnal-jurnal, yang sebelumnya tidak teratur, dan memudahkan dalam pencarian.

1.2 Perumusan Masalah

Untuk memudahkan pengolahan data, database dibuat dalam satu dokumen yang sama. Sistem ekstraksi ini di jalankan secara otomatis, hasil ekstraksi yang diambil sebagai keterkaitan jurnal adalah referensi atau daftar pustaka dari jurnal-jurnal yang telah diekstrak, Sedangkan untuk memudahkan pencarian menggunakan penulis atas judul dari jurnal sebagai kata kunci.

1.3 Tujuan

Tujuan yang ingin dicapai dari hasil pembuatan dan penulisan skripsi ini adalah membuat suatu sistem yang dapat membantu dalam proses pencarian keterkaitan kepustakaan (referensi) antara satu dokumen hasil karya penelitian dengan dokumen lainnya yang didapat dari internet.

1.4 Batasan Masalah

Pembatasan masalah pada penyusunan skripsi ini adalah mencari dokumen yang berbentuk artikel jurnal atau karya ilmiah. Dokumen tersebut dalam format pdf, jurnal-jurnal tersebut tidak termasuk jurnal-jurnal yang tidak bisa diakses bebas, dan jurnal yang digunakan adalah adalah jurnal-jurnal yang dikeluarkan beberapa institusi pendidikan di dalam negeri yaitu Indonesia.

1.5 Metodelogi

Metode penulisan yang digunakan adalah sebagai berikut :

1. Studi Pustaka

Yaitu dengan mencari serta membaca dan mempelajari literatur yang berhubungan tentang masalah diatas, yang dapat membantu dalam penyusunan skripsi.

2. Perancangan Sistem

Yaitu melakukan proses penggunaan berbagai teknik dan prinsip yang didapat dari studi pustaka untuk tujuan mendefinisikan proses atau sistem secara detail. Dimana perancangan ini berfokus pada pada alur kerja dari sistem yang dibuat dan pembagian fungsi dari hal-hal (*tools*) yang telah dipelajari.

Perancangan sistem yang dibuat dibagi menjadi dua bagian yaitu sistem ekstraksi yang digunakan untuk mendapatkan informasi yang diperlukan dari situs penyedia jurnal, dan sistem tampilan yang berfungsi untuk menampilkan keterkaitan sitasi antara artikel jurnal yang telah didapat informasinya. Dalam hal ini ditentukan juga jumlah situs dan alamat situs penyedia jurnal yang akan dijadikan sebagai sumber data artikel jurnal.

1.6 Sistematika Pembahasan

- BAB 1 PENDAHULUAN

  Membahas latar belakang masalah, perumusan masalah, tujuan, batasan masalah, metodologi dan sistematika pembahasan.

- BAB 2 TEORI PENUNJANG

  Membahas teori-teori penunjang yang akan digunakan pada proses perancangan.

- BAB 3 PERANCANGAN

  Membahas tahapan-tahapan perancangan yang dilakukan dan proses pengerjaan / implementasi sistem yang dibuat.

- BAB 4 PENGUJIAN DAN ANALISA

  Membahas pengujian serta analisa perhitungan dari proses berjalannya sistem yang dibuat.

- BAB 5 PENUTUP

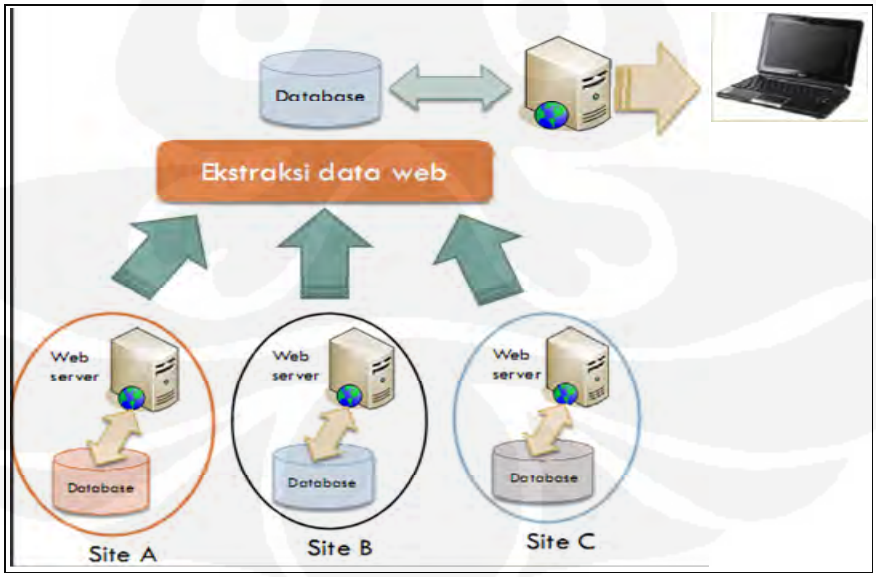
  Berisi mengenai kesimpulan dari semua tahapan yang dilalui dalam proses perancangan sistem

BAB 2
TEORI PENUNJANG

2.1. Teknik Ekstraksi Data Web

Screen scraping adalah suatu teknik pengambilan isi dari web untuk diolah tanpa fasilitas sindikasi dimana suatu program dalam komputer yang memungkinkan data dari monitor komputer dapat dibaca dan digunakan untuk diimpor ke yang lain, program yang mendasarinya. Dan program yang melakukannya disebut *screen scrapper* [1]. Yang membedakannya dengan *parsing* biasa adalah dimana untuk *screen scraping* ini datanya lebih diperuntukkan untuk ditampilkan ke pengguna akhir daripada untuk inputan program lain. *Screen scraping* sering mengabaikan data biner (biasanya foto atau data multimedia) dan format elemennya, sehingga cenderung pada data penting berupa teks [3].

Awalnya *screen scraping* digunakan untuk membaca data teks dari tampilan layar komputer. Hal ini dilakukan dengan membaca terminal memori dan dengan menggunakan *port* tambahan. Alternatif lainnya menjadikan *output port* dari suatu komputer menjadi *input* bagi *port* komputer lainnya.



Gambar.2.1. Alur kerja ekstraksi jurnal [7]

Umumnya transfer data antara program dilakukan dengan struktur data yang cocok untuk diproses secara otomatis dengan komputer, seperti pada pertukaran format dan protokol yang berstruktur kaku, di dokumentasikan dengan baik, dan minimum ambigu. Seringnya transmisi ini tidak dibaca manusia sama sekali. Tetapi untuk *output* yang berkebalikan dengan hal di atas seperti label yang berlebih atau komentar yang berlebih atau informasi lainnya yang tidak dapat dilakukan dengan proses otomasi. Akan tetapi, meskipun *output* yang tersedia adalah sebuah tampilan untuk manusia, *screen scraping* menjadi suatu cara untuk mengerjakan transfer data tersebut.

Web Scrapping atau *Web harvesting* atau ekstraksi data web adalah suatu teknik untuk mengutip data atau informasi dari suatu *website* menggunakan *software* dengan program tertentu. Biasanya program dalam *software* tersebut mensimulasikan eksplorasi manusia terhadap suatu web dengan menggunakan *low-level* HTTP atau menggunakan *full-fledged web* tertentu seperti internet *explorer* dan *Mozilla* [1].

Gambar.2.2. Tag data yang akan di ekstraksi [7]

Web scraping berhubungan dengan pengindeks-an web yang merupakan suatu teknik universal yang dipakai hampir semua *search engine*. Perbedaannya *web scraping* lebih berfokus pada tranformasi dari suatu isi web yang tidak terstruktur, umumnya dalam format HTML menjadi suatu format data terstruktur yang dapat disimpan dan dianalisa pada *database* atau lembar kerja.

Web scraping juga terkait dengan otomasi web, yang mensimulasikan aktivitas *web browsing* dari manusia menggunakan perangkat lunak komputer.

Contoh penggunaannya antara lain :

- Perbandingan harga online / katalog produk
- monitoring data cuaca
- deteksi perubahan website
- penelitian web
- integrasi data web
- *web content mashup*
- Mengumpulkan informasi seputar perumahan (nama, lokasi, harga, kontak, dan lain-lain)
- Mengkliping artikel (judul, isi, kata kunci,sumber referensi, dan lain-lain)
- Otomatisasi situs lelang
- Mengekstraksi undian judi, dan lainnya.

Web *scraping* merupakan area yang cukup banyak dikembangkan. Proses otomasi pengumpulan data atau informasi web merupakan tujuan bersama dari semantik web. Pengolahan teks, pengertian semantik, kecerdasan buatan dan interaksi manusia dengan komputer, adalah hal yang banyak diperhatikan.

Web scraping menjadi semacam solusi praktis yang berdasarkan teknologi yang ada meskipun beberapa solusi masih khusus. Karena itu ada beberapa level dari otomasi yang tersedia pada *web scraping* antara lain :

- *Text grepping and regular expression matching*: Sebuah pendekatan sederhana namun canggih untuk mengambil informasi dari halaman Web dapat berdasarkan *unix grep* perintah dan kalimat biasa cocok dengan menggunakan *perl* atau *python* bahasa pemrograman.
- *Human copy-and-paste*: Sering terjadi bahwa teknologi Web *Scraping* tidak bisa menggantikan manusia dari pemeriksaan manual dan meng-*copy-paste*, Kadang-kadang hal ini dapat menjadi satu-satunya solusi yang ada ketika situs web secara eksplisit terdapat hambatan untuk mencegah mesin otomatisasi.
- *HTTP programming*: statis dan dinamis halaman web dapat diambil dengan permintaan HTTP ke *server web* yang jauh (*remote*) menggunakan pemrograman *socket*.
- *DOM parsing*: dengan menambahkan suatu *full-fledged Web browser*, seperti Internet *Explorer* atau *Mozilla*, program dapat mengambil isi dinamis yang dihasilkan dari skrip sisi klien.
- *HTML parsers*: beberapa bahasa *query* data semi berstruktur, seperti XML *query language* (XQL) dan *hyper-text query language* (HTQL), dapat digunakan untuk mem-*parsing* halaman HTML dan untuk mengambil konten dan mentransformasi Web.
- *Web-scraping software*: ada banyak perangkat lunak *web scraping* software yang dapat digunakan untuk solusi web *scraping*. Perangkat lunak tersebut mungkin menyediakan antarmuka untuk merekam web sehingga menghilangkan kebutuhan untuk secara manual menulis kode untuk *web scrapping*, atau beberapa *skrip* dari fungsi yang dapat digunakan untuk mengekstrak dan mentransformasi isi web, dan antarmuka basis data yang dapat menyimpan data yang diambil ke database lokal.

Untuk mengekstraksi data dari suatu *website* perlu dilakukan beberapa hal seperti:

- menemukan halaman HTML sasaran dalam sebuah situs dengan mengikuti *hyperlinks*.
- ekstraksi potongan-potongan data yang relevan dari halamannya,
- penyaringan dan pemprosesan data.

Dalam mendapatkan data yang relevan dari halaman suatu web dapat dilakukan dengan *scanning* bagian-bagian efektif pada sebuah dokumen seperti kolom kutipan, pengarang, judul buku, tanggal dan lain sebagainya (ini dapat disesuaikan dengan tujuan dan kebutuhan penggunaan teknik ekstraksi web yang dibangun). Ada kalanya informasi-informasi mengenai konten suatu dokumen tidak dapat begitu saja diperoleh. Hal ini bisa disebabkan misalnya informasi penting disajikan dalam dokumen non HTML, seperti yang umumnya digunakan saat ini yaitu PDF dan *Flash clips*.

Bervariasinya dokumen yang ada sehingga dimungkinkan ada dokumen yang tidak mengikuti format yang dapat secara otomatis diketahui maksudnya oleh sistem ekstraksi web. Contohnya dalam penulisan kolom referensi, nama pembuat dokumen, judul, dan tanggal pembuatan yang ditulis berbeda tidak sesuai template standar yang diketahui sistem. Dengan adanya perubahan format atau karena format yang digunakan tidak standar, maka teknik ekstraksi web ini harus dapat mengatasi hal tersebut dengan menyediakan kemungkinan-kemungkinan perubahan format yang dapat terjadi.

2.2. Mashup

Mashup merupakan istilah dalam pengembangan web yang mendefinisikan sebuah aplikasi web yang menggabungkan data dari dua atau lebih sumber data. Atau alat bantu pada aplikasi web untuk mendapatkan data dari banyak sumber, Wikipedia menyebut *mashup* sebagai sebuah aplikasi web hibrida. Web mashup merupakan aplikasi web yang mengkomposisi ulang data dari berbagai macam sumber dan menampilkannya kembali dalam bentuk yang berbeda [8].

Mashup dibangun dengan menggabungkan data yang sudah ada di internet baik itu Open API, RSS maupun layanan informasi lainnya. Dengan menggunakan teknologi mashup, data bisa didapatkan dengan mudah melalui penyedia layanan di interenet tanpa harus menyediakan sendiri. Penggunaan data kembali ini menghemat waktu dan biaya yang dikeluarkan ketika data yang dibutuhkan aplikasi harus disediakan sendiri. Salah satu contoh adalah ketika sebuah aplikasi membutuhkan data spasial atau peta. Menggunakan teknologi *mashup*, data spasial dapat didapatkan dengan mudah dan gratis dari *Google Map* atau Microsoft *Virtual Earth*. Dan hal ini jelas lebih efisien daripada menyediakan data spasial sendiri.

Mashup menggunakan sumber data yang terkoneksi langsung dengan internet sehingga informasi berkembang sesuai dengan perkembangan sumber data tersebut di internet. Misalkan sebuah sistem informasi geografis perkebunan me-*mashup* data dari *Google Map*. Awalnya peta perkebunan di daerah jawa timur tidak lengkap karena *Google Map* belum menyediakan. Ketika *Google Map* menyediakan peta jawa timur yang lebih lengkap, sistem informasi geografis itu pun secara otomatis akan ter-*update* data spasialnya.

Sebagian besar penyedia layanan internet seperti Google, Yahoo, Flickr, Ebay dan Amazon sudah menyediakan *Open API* untuk mengakses konten dan data yang mereka miliki sehingga semakin mendorong perkembangan *mashup*. *Open Api* menggunakan protokol *http* yang tidak terikat sistem operasi maupun bahasa pemrograman seperti *REST* dan hasil pengembaliannya berupa xml atau json yang fleksibel. Selain itu sebagian besar *OpenAPI* tersedia secara gratis dan terdokumentasi dengan baik.

Teknologi *mashup* menggunakan data langsung dari internet oleh karena itu data yang di *mashup* tidak perlu lagi disimpan kedalam database. *Storage* database yang seharusnya di bebankan kepada aplikasi dialihkan kepada penyedia layanan data sehingga aplikasi menjadi lebih ringan.

Contoh *Mashup* yang ada adalah penggunaan data kartografis dari *Google Maps* untuk menambah informasi lokasi ke data *real-estate*, dengan membuat layanan web yang baru dan berbeda dari yang sudah ada. Konten *Mashup* biasanya diambil dari pihak ketiga melalui *Interface publik* atau *Application*

Programming Interface (API) atau *Web Services.* Selain itu metode mendapatkan konten dari Mashup meliputi web feeds (RSS atau Atom) dan *screen scraping*. Banyak orang bereksperimen dengan *Mashup* menggunakan Amazon, eBay, Flickr, Google, Microsoft, Yahoo, atau YouTube API, yang menuju kepada terbentuknya editor Mashup. Tabel 2.1 berikut ini merupakan beberapa perbedaan antara *Mashup* dengan *portal*.

Tabel 2.1. Perbedaan *Portal* dengan *Mashup* [8]

Parameter	Portal	Mashup
Klasifikasi	Teknologi lama.	Teknologi web terbaru (Web 2.0)
Metode aggregasi	Membagi konten web menjadi segmen lalu mengaggregasinya melalui web server.	Konten dilayani menggunakan *Open API*.
Dependensi konten	Konten berupa fragmen yang bercampur dengan data tampilan web (HTML, WML).	Konten yang didistribusikan berupa data murni berbentuk XML atau JSON.
Dependensi Konten	Dilakukan pada sisi server.	Dapat dilakukan pada sisi client.
Style aggregasi	Konten diaggregasi dengan terkotak kotak.	Konten bisa dibentuk sesuai dengan keinginan pengguna.
Model *event*	Menggunakan spesifikasi *portlet*.	Operasi CRUD menggunakan REST.
Standar yang digunakan	JSR 168, JSR 286, dan WSRP.	Menggunakan standar XML dan *web service*.

Ada beberapa tipe dari Mashup seperti *consumer Mashup*, *data Mashup* dan juga *business mashup,* dan yang paling umum digunakan adalah *consumer Mashup* yang dipergunakan dan membantu untuk kalangan umum, misalnya pada *google map*. Mashup yang mengkombinasikan antara beberapa informasi atau

media yang sama dari berbagai sumber dan mengintegrasikannya dalam satu sistem disebut juga sebagai data mashup. Selain itu, *business mashup* lebih berfokus pada suatu sistem untuk kerjasama suatu perusahaan.

Secara arsitektur, *Mashup* ini terbagi menjadi 2 yaitu *Web-based* dan *Server-based*. *Mashup* berbasis web umumnya berupa *web browser* pengguna yang melakukan penggabungan dan memformat ulang data. *Mashup* berbasis server akan melakukan analisa dan format ulang data yang ada dilakukan pada server dan mengirimkannya ke *browser* pengguna.

Kekurangan

- **Aplikasi bergantung kepada sumber data *mashup***

Data pada aplikasi yang menggunakan mashup terdapat pada penyedia layanan internet. Sehingga terjadi ketergantungan antara aplikasi dengan penyedia data. Ketik penyedia data mengalami masalah, maka aplikasi juga tidak akan dapat berjalan. Masalah ini dapat di minimasi dengan cara memilih penyedia layanan data yang terpercaya sehingga kemunginan ketidak tersediaan data dapat berkurang.

- **Aplikasi bergantung kepada kualitas koneksi internet**

Teknologi *mashup* membutuhkan koneksi internet yang baik karena data tersimpan pada penyedia jasa diluar aplikasi. Setiap kali aplikasi melakukan *load* data maka koneksi internet dilakukan. Hal ini menjadi masalah terutama di negara yang koneksi internetnya relatif mahal dan *bandwith* terbatas seperti Indonesia dan negara berkembang lainnya.

2.3. Sitasi dan Indeks Sitasi

Sitasi menunjukan asal-usul atau sumber suatu kutipan, mengutip pernyataan, atau menyalin/mengulang pernyataan seseorang dan mencantumkannya di dalam suatu karya tulis yang dibuat, namun tetap mengindikasikan bahwa kutipan tersebut itu adalah pernyataan orang lain atau merupakan suatu rujukan terhadap suatu buku, artikel, jurnal, halaman web ataupun bentuk-bentuk publikasi lainnya.

Isi sitasi biasanya terdiri dari nama penulis, judul buku, atau artikel, penerbit, tahun publikasi, dan URL juga tanggal karya tersebut diakses. Ada beberapa metode penulisan sitasi yang dibuat dan diterbitkan oleh berbagai asosiasi atau individu yang digunakan oleh penulis, yang berbentuk pustaka konvensional, misal Kate L.Turabian dan Robert A.Day, misalnya Chicago style dan Turabian style yang digunakan untuk semua bidang. *Modern association Language* (MLA) yang digunakan untuk seni, kesusastraan, dan humaniora, Serta Pustaka elektronik, *American Psycological Association* (APA) yang digunakan untuk psikologi, pendidikan, dan ilmu sosial lainnya, *American Medical Association* (AMA) untuk bidang kedokteran, kesehatan, dan biologi. Standar penulisan lainnya seperti *National Library Of Medicine* (NLM), *American Chemical Society* (ACS), *American Polotical Science Association* (APSA), *Councils Of Biology Editors* (CBE), IEEE style, *American Sociological Association* (ASA), columbia style, dan *Modern Humanities Research Association* (MHRA).

Sitasi atau kutipan terhadap suatu karya ilmiah ataupun dokumen dilakukan karena dokumen yang dikutip tersebut menyediakan informasi yang relevan terhadap penelitian atau tulisan yang dikerjakan oleh penulis. Dengan demikian dapat diketahui bahwa makin sering sebuah dokumen dikutip, maka semakin besarlah dokumen tersebut memberikan kontribusi informasi, dan semakin besarlah pula pengaruhnya pada penelitian atau penulisan yang sedang dilaporkan di dalam dokumen pengutip. Ukuran dari pengaruh atau dampak (*impact*) dari sebuah dokumen memberikan suatu informasi tergantung pada jumlah pengutipan terhadap dokumen tersebut.

Dengan mengetahui berapa kali sebuah dokumen dikutip dalam satu rentang waktu tertentu menunjukkan berapa banyak informasi di dalam dokumen tersebut berguna untuk sebuah penelitian atau penulisan. Dimana apabila frekuensinya menurun, maka dokumen tersebut semakin tidak relevan, sampai akhirnya menjadi usang alias *obsolete.* Apabila terdapat dua dokumen bersama-sama dikutip oleh suatu dokumen, maka kedua dokumen tersebut bersama-sama memberi sumbangan informasi yang saling terkait. Sehingga semakin sering dua

dokumen dikutip bersama (*co-cited*), maka semakin dekatlah hubungan kedua dokumen tersebut.

Indeks sitasi merupakan suatu indeks dari sitasi-sitasi antara berbagai penerbitan, yang memungkinkan pengguna dengan mudah mendapatkan dokumen lebih baru mana yang mensitasi dokumen lebih lama yang mana. indeks sitasi yang pertama adalah untuk *legal citation*, seperti Shepard’s sitasi yang dibuat pada tahun 1873. Kemudian tahun 1960, Eugene Garfield dari *Institute For Scientific Information* (ISI) memulai indeks sitasi untuk *paper* yang diterbitkan di jurnal akademik, diawali dengan *Science Citation Index* (SCI) lalu kemudian *Social Sciences Citation Index* (SSCI) dan *Art And Humanities Index* (AHCI).

Sekarang ini telah banyak yang menyediakan layanan indeks sitasi baik yang berbayar ataupun dapat digunakan secara gratis, bebarapa diantaranya antara lain :

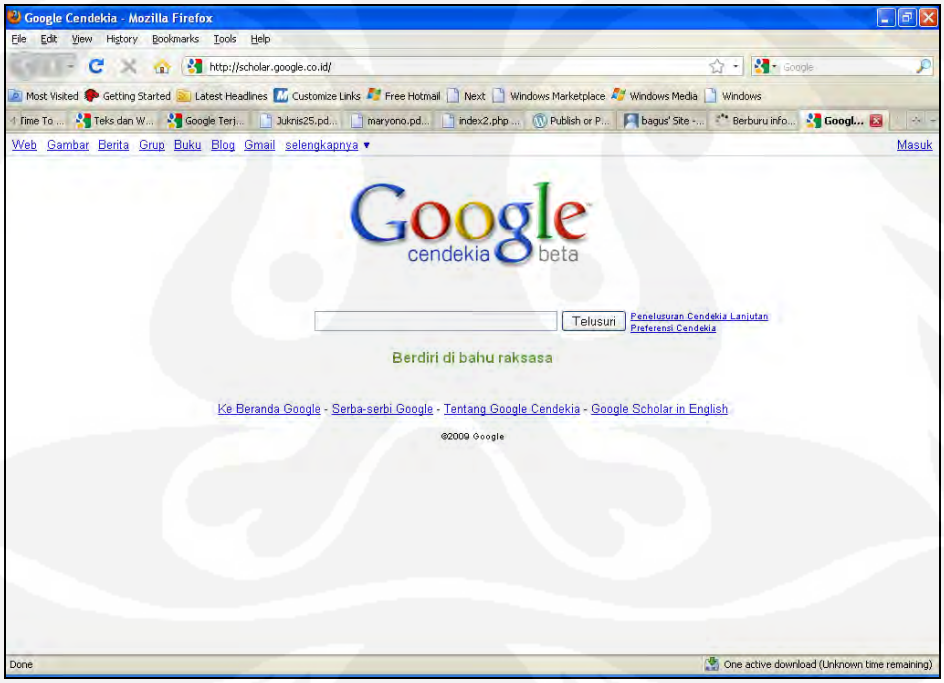
- **ISI** merupakan layanan berbayar yang merupakan bagian dari *Thomson Scientific* adalah penyelia jasa sitasi yang paling utama. Indeks sitasi ISI oleh Thomson Scientific sampai sekarang masih disediakan dalam format cetakan dan CD ROM. Sekarang ISI bisa diakses lewat web dengan nama *Web of Science*. *Web of science* ini merupakan sebuah layanan akademik *online* yang dapat diakses dengan menggunakan ISI *Web of Knowledge* (WoK), dimana WoK menyediakan akses ke *database* dan sumber lainnya. Salah satunya adalah *Web Of Science* yang menyediakan indeks sitasi seperti *Science Citation Index* (SCI), *Social Sciences Citation Index* (SSCI), *Arts and Humanities Citation Index* (A&HCI), *Index Chemicus, Current Chemical Reactions, Conference Proceedings Citation Index* untuk Science dan *Social Science and Humanities*. Semuanya tersedia melalui jasa *database Web of Knowledge* dari ISI. *Database* ini memungkinkan peneliti mengidentifikasi artikel yang paling sering disitasi, dan siapa yang telah menyitirnya. ISI juga menerbitkan tahunan ***Journal Citation Report***, yang mendaftar peringkat *impact factor* untuk tiap jurnal yang diindeksnya[12].

- ***CiteseerX*** merupakan penyedia indeks sitasi yang dapat digunakan gratis untuk *scientific* dan akademik *paper* yaitu bidang komputer dan ilmu informatika. Dasar dari *citeseerX* ini adalah *citeseer* yang kemudian dibuat dengan infrastruktur *open source* baru yaitu *Seersuite* dan algoritma yang baru pada penerapannya. Sehingga bertujuan melanjutkan *citeseer* untuk mengambil dokumen *scientific* dan akademik untuk dijadikan indeks sitasi yang kemudian dapat digunakan untuk merangking dokumen dengan menggunakan *impact of citation*. Berikut merupakan tampilan halaman web dari *citeseerX* :

Gambar 2.3. Tampilan Halaman Web Layanan *CiteSeerX*

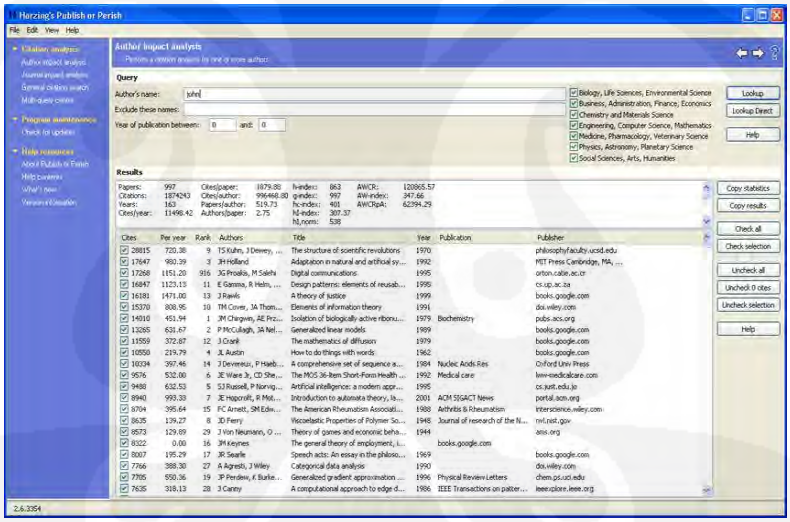
- ***Google Scholar*** merupakan salah satu penyedia indeks sitasi yang dapat digunakan gratis. *Google Scholar* ini dirilis versi beta-nya pada bulan November 2004. Fungsi dari *google scholar* ini mirip seperti Scopus, citeseerX, *Web of Science*. Apabila *Thomson Scientific* dan Scopus membuat laporan *Citation Indexes* berdasarkan data primer (dari *database* mereka),

maka Google Scholar memanfaatkan artikel-artikel yang tersedia bebas di internet (umumnya dari artikel serupa yang disimpan dalam website pribadi penulis ataupun *repository* universitasnya) ataupun dari literatur abu-abu seperti buku, *proceeding*, monograf, *website* penulis, dan lain sebagainya. Walau demikian ketepatan perhitungan Google Scholar cukup tinggi, terlebih lagi untuk artikel-artikel yang terbit setelah tahun 2004. Sekarang ini *Google Scolar* menyediakan hitungan sitasi (*citation count*) yang dapat diakses gratis melalui Internet sehingga semua orang kini dapat menyiapkan laporan *citation count*, *citation index*, ataupun *impact factor* tanpa harus berlangganan ke jasa-jasa komersial seperti *Thomson Scientific* atau *Elsevier*. Dengan demikian kemungkinan produk berbayar yang ada sebelumnya bisa saja tidak terpakai lagi dimasa depan. Berikut tampilan halaman web dari *Google Scholar*:



Gambar 2.4. Tampilan Halaman Web Layanan *Google Scholar*

- ***Publish or Perish*** adalah sebuah perangkat lunak gratisan untuk menghitung analisa sitasi. Pada perangkat lunak ini, dasar perhitungannya diambil dari *data citation* di *google scholar*. Keuntungannya, dengan *google scholar*, kita dapat memperoleh *data citation* yang lebih lengkap mencakup kutipan dari buku, *book chapter*, jurnal-jurnal yang tidak termasuk dalam master list ISI web of science dll. Maksudnya dokumen yang masuk di google scholar. Indeks sitasi ini dapat digunakan untuk merangking karya ilmiah-karya ilmiah yang ada, baik dari tema ataupun penulisnya. Selain itu *Publish or Perish* dapat dijadikan juga acuan bagi penulis atau peneliti untuk mengetahui informasi terkini mengenai karya ilmiah yang telah dihasilkannya, terutama mengenai berapa jumlah peneliti atau acuan dalam artikel atau hasil karya ilmiah lainnya yang telah mengacu pada hasil karyanya[12].



Gambar 2.5. Tampilan Halaman *Publish or Perish*

Perangkat lunak ini dibuat oleh Harzing.com, yaitu *website* professor Anne-Wil Harzing. Professor Harzing bekerja di departemen International Management at the University of Melbourne, Australia. Biasanya kesimpulan yang bisa didapat dari adanya indeks sitasi ini yaitu *impact factor* (IF) yang merupakan ukuran dari sitasi (*citation*) terhadap jurnal-jurnal ilmu pengetahuan alam (*science*) dan ilmu pengetahuan social (*social*

science) dan sering kali digunakan sebagai ukuran terhadap pentingnya suatu jurnal dalam bidangnya. *Impact Factor* diciptakan oleh Eugene Garfield dari *Institute of Scientific Information* (ISI), kini bagian dari Thomson Scientific) pada tahun 1960 dengan menghitung indeks sitasi (*citation index*) dari jurnal-jurnal yang diindeks oleh Thomson ISI dan dilaporkan setiap tahun dalam *Journal Citation Report* (JCR).

Setalah mengunduh filenya dan kemudian menginstalnya, saya mencoba melakukan test dengan sejumlah nama peneliti CIFOR. Test ini sekedar ingin mengetahui bagaimana cara menggunakan perangkat lunak ini. Sebagai contohnya, saya menggunakan nama David Kaimowitz, kemudian memberi tanda centang pada semua subjek. Hasilnya H-index David adalah 14. So, kalo David mau jadi Nobel winner, dia mesti mengejar di atas angka 40an, atau setidaknya kalau mau jadi *full professor* segera sabet angka h-index.

2.4. Web Semantik

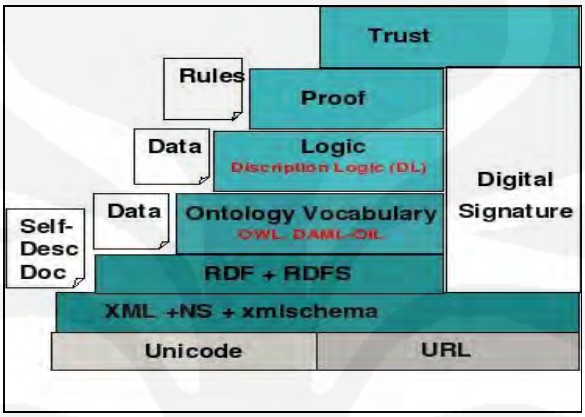
Web semantik adalah suatu data web yang dapat diproses secara langsung atau tidak langsung oleh mesin, yang merupakan pengembangan *world wide web* yang memungkinkan web untuk memahami dan dapat memenuhi permintaan dari manusia dan mesin untuk menggunakan isi web. Langkah pertama adalah meletakan data pada web dalam suatu bentuk sehingga mesin dapat memahami, atau mengubah menjadi format tertentu [9].

Data dapat terdiri dari beberapa domain dan digolongkan dalam taksonomi hierarkis. Penggolongan dapat digunakan untuk penemuan data, hubungan yang sederhana antara kategori dalam tsksonomi dapat digunakan untuk menghubungkan dan dengan demikian mengkombinasi data. Sehingga, data kini cukup cerdas untuk dengan mudah ditemukan.

Web semantik itu sendiri diperkenalkan oleh Tim Berners-Lee, penemu World Wide Web. Sekarang, prinsip web semantik disebut-sebut akan muncul pada Web 3.0, generasi ketiga dari World Wide Web. Bahkan Web 3.0 itu sendiri sering disamakan dengan web semantik. Tujuan dari web semantik adalah web

menjadi media yang universal untuk data, informasi dan saling bertukar pengetahuan [9].

Web semantik terdiri dari standar dan *tools* diantaranya XML, XMLS (XML Schema), RDF, *Resources Description Network Schema* (RDFS) dan OWL yang diatur dalam Stak Semantik web stack seperti pada Gambar 2.5 [4]:

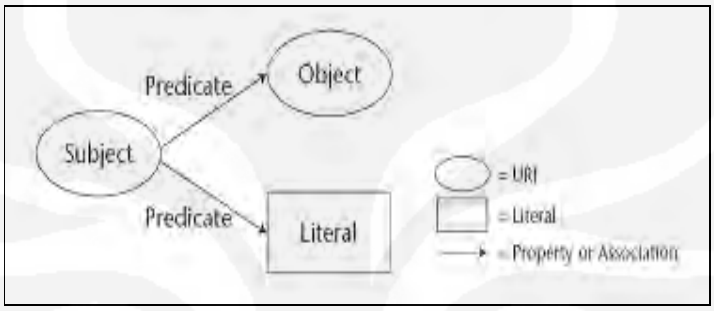


Gambar 2.6. Piramid Web Semantik

Masing-masing standar dan tools yang ada tersebut mempunyai fungsi dan keterkaitan berikut :

- XML adalah lapisan pondasi yang syntactic dari sematik web. Kebutuhan teknologi web yang lain (seperti RDF) menjadi lapisan paling atas dari XML menjamin suatu tingkatan dasar dari *interoperabilitas* [9].
- XML *Schema* adalah sebuah definisi dari XML untuk memberikan aturan untuk sebuah dokumen XML, teknologi XML dibangun atas *Unicode characters* dan *uniform Resource Identifier*. *Unicode characters* mengijinkan XML untuk menggunakan karakter internasional. Uri digunakan sebagai *unik identifiers* untuk konsep di sematic web tersebut [9].
- RDF adalah *layer* untuk merepresentasikan sematik dari isi halaman tersebut, yang merupakan dasar untuk pemprosesan metadata, dimana metadata dalam web dapat di kode kan, dipertukaranan dan dipergunakan. Model RDF adalah suatu *triple* yang dinamakan statement; satu sumber daya (subject) yang dihubungkan ke sumber daya yang lain atau satu literal (object) melalui satu *arc* dari sumber daya ke tiga, predikat [9]. Dimana bisa disebut juga sebagai:

- *Resources*, adalah bagian dari sumber informasi, dalam era Internet di representasikan dalam alamat web atau URL, ini disebut subyek atau obyek [9].
- *Property*, adalah sebuah karakteristik dari atribut atau relasi untuk menjelaskan sumber, ini disebut juga predikat [9].



Gambar 2.7. RDF *triple* [9]

- RDF *Schema* adalah sebuah lapisan diatas RDF, dan merupakan sebuah *set standard* sederhana dari sumber RDF yang memungkinkan untuk membuat *vokabulari* RDF sendiri. Model dari RDFS memiliki kemiripan dengan yang digunakan oleh *objectoriented*, yaitu dengan memiliki *class, relation, property dan instance*. *Class* adalah kumpulan dari obyek yang memiliki kesamaan karakter. *Relation* adalah sifat hubungan antar kelas. *Property* adalah karakter dari sebuah kelas. *Instances* adalah sebuah obyek yang merupakan anggota sebuah kelas [5].
- OWL menambahkan lebih banyak daftar kosa kata untuk menggambarkan properti-properti dan kelas-kelas, seperti hubungan antar kelas, keutamaan, persamaan, cara penulisan properti, karakteristik properti dan jumlah kelas.
- SPARQL adalah protokol dan bahasa untuk web semantik data

2.5. Tools Ekstraksi Data Web

Pada saat ini kita dengan dapat menemukan banyak tools yang digunakan untuk proses ekstraksi data web. Tools yang tersedia dapat membantu kita untuk lebih memudahkan dalam memproses sebuah ekstraksi data web, karena tools yang ada saat ini tidak banyak menggunakan skrip code, akan tetapi lebih banyak

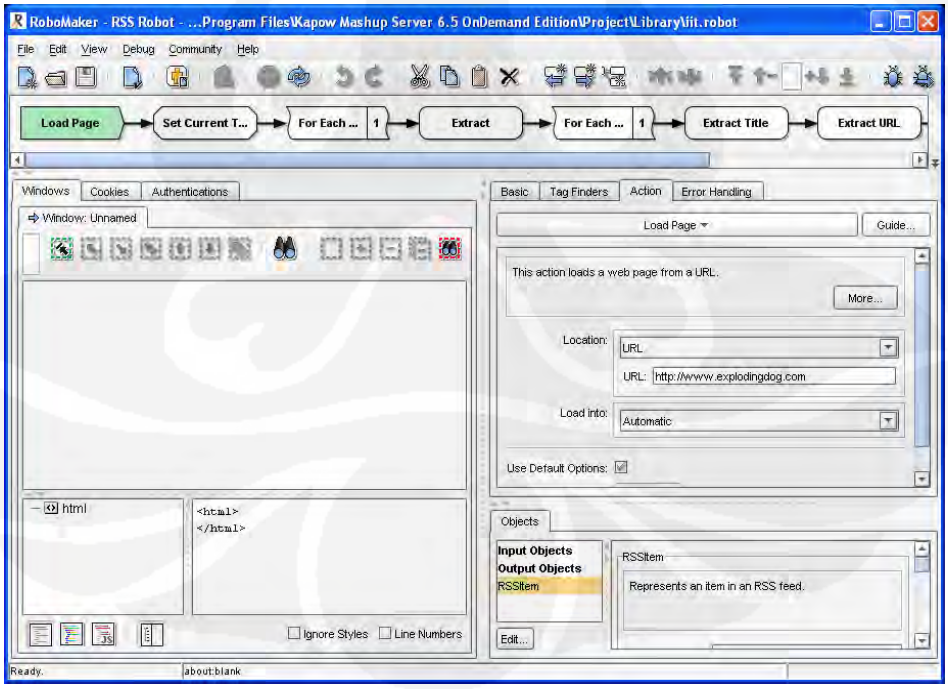
menggunakan visual atau *tag frame* seperti *drag and drop*. *Tools* yang dikenal saat ini ada yang gratisan (*open source*) ada juga yang berbayar, tergantung dari mana keluarannya. Beberapa *tools* tersebut antara lain:

2.5.1. Kapow Mashup Server 6.5 Robomaker

Kapow Mashup Server 6.5 Robomaker yang suatu *open service platform* dari openkapow, dimana pengguna dapat menggunakan suatu program tersendiri dan menjalankannya dari www.openkapow.com atau diinstalkan pada computer masing-masing dengan gratis.

Konsep penting dari Kapow Mashup Server 6.5 Robomaker adalah robot, dimana robot ini adalah suatu program yang didesain untuk menjalankan suatu tugas tertentu, biasanya terkait *website*. Pada dasarnya robot dapat diprogram untuk secara otomatis semua yang dapat dilakukan pada sebuah browser.

Robomaker adalah lingkungan untuk pemrograman robot dalam sebuah bahasa pemrograman dengan kegunaan khusus dengan sintak dan semantik tersendiri. Lingkungan pemrograman dalam robomaker dapat terlihat seperti pada Gambar 2.8 berikut :



Gambar 2.8. Tampilan Utama Jendela Robomaker

Robot view terletak di bawah *icon toolbar* dari robomaker main window. *Robot view* ini akan menampilkan tahapan (*step*) dan koneksi dari tahapan tersebut sehingga membentuk sebuah robot. Tahapan yang ada pada *robot view* mempunyai beberapa elemen seperti nama, list tag, dan aksi atau akivitas yang dikerjakan tahapan tersebut. Rangkaian tahapan dapat terlihat seperti berikut:

Gambar 2.9. Contoh Rangkaian Tahapan Robomaker

State *view* terleak dibawah robot *view* di sebelah kiri Jendela Utama Robomaker main window. Pada *state view* dapat terlihat keadaan robot aktual, atau tampilan dari yang sedang dilakukan robot, misalnya membuka salah satu halaman web. *State view* ini terdiri dari beberapa bagian diantaranya *tag path view, browser view, tree view* dan *source view* seperti Gambar 2.8 berikut :

Gambar 2.10. Tampilan Jendela Utama *Page View* Robomaker

Step view terletak di sebelah kanan *state view*. *Step view* ini memperlihatkan konfigurasi dari tahapan aktual. Pada *step view* ini juga dapat dilakukan konfigurasi untuk tahapan yang akan dibuat pada *Robot view*.

Objek *view* terletak di bawah *step view*. Objek *view* ini memperlihatkan objek dari keadaan tahapan aktual baik input ataupun output dari robot.

2.6. PHP

PHP merupakan singkatan dari *PHP Hypertext preprocessor*, merupakan suatu bahasa berbentuk skrip yang ditempatkan dan diproses di server yang hasilnya akan dikirimkan ke klien untuk ditampilkan di browser. PHP dirancang untuk membentuk suatu aplikasi web yang dinamis, yaitu membentuk suatu tampilan berdasar kebutuhan terkini/aktual, prinsipnya PHP sama dengan skrip lainnya seperti ASP, Cold fusion, ataupun Perl [2].

PHP terbentuk saat Rasmus Lerdof pada tahun 1994 membuat sejumlah script menggunakan perl yang dapat digunakan untuk mengamati siapa saja yang telah melihat riwayat hidupnya, kemudian skrip ini dikemas menjadi *tool* yang disebut "*Personal Homepage*". Paket inilah yang menjadi cikal bakal PHP, tahun 1995 rasmus menciptakan PHP/FI versi 2, pada versi inilah pemrograman dapat menempelkan kode terstruktur pada tag *Hypertext Markup Language* (HTML). Kode PHP ini juga dapat berkomunikasi dengan database dan melakukan perhitungan sambil jalan.

PHP dapat berfungsi pada server-server yang berbasis LINUX, UNIX, Windows, dan macintosh, awalnya dirancang untuk diintegrasikan dengan web server apache. Namun kemudian php juga dapat bekerja dengan *web server* seperti *Personal Web Server* (PWS), *Internet Information Server* (IIS), dan Xitami.

Script PHP berkedudukan sebagai *tag* dalam bahasa HTML, dimana HTML ini adalah salah satu bahasa standar dalam membuat web yang disimpan dengan ekstensi .htm atau .html. Berikut contoh kode php dalam script HTML, dimana kode PHP ini diawali dengan *tag* *<?php* dan diakhiri dengan *?>* [2] :

```
<HTML>
<HEAD>
<TITLE>Latihan Pertama</TITLE>
```

```
</HEAD>
<BODY>
<?php
  printf("Tgl. Sekarang: %s ", Date("d F Y"));
?>
</BODY>
</HTML>
```

Salah satu kelebihan PHP adalah mampu berkomunikasi dengan database sehingga dapat menampilkan data secara dinamis yang diambil dari database. Pada saat ini PHP dapat berkomunikasi dengan berbagai database seperti DBM, Oracle, Oracle, FilePro, PostgreSQL, Informix, Sybase, Ingres, Microsoft Acces, MSQL dan MySQL.

2.6.1. Elemen dasar PHP

Sama seperti pemrograman lainnya PHP dalam penggunaanya menggunakan beberapa elemen dasar yang digunakan dalam skripnya, antara lain:

- Karakter

  Karakter banyak dijumpai dalam tiap bahasa pemrograman, yang dapat berupa huruf, angka, spasi, tanda kontrol ataupun simbol.

- *Identifier*

  Identifier atau pengenal merupakan sesuatu yang digunakan oleh bahasa pemrograman untuk memberi nama fungsi, variabel atau kelas.

- Tipe data

  Tipe data dalam PHP ada 4 macam yaitu *integer, double, string*, dan *boolean.*

- Konstanta

  Merupakan sesuatu yang digunakan untuk menyatakan nilai yang tetap dalam program.

- Variabel

  Variabel berfungsi untuk menyimpan suatu nilai dimana nilai didalamnya dapat diubah sewaktu-waktu sesuai kebutuhan. Variabel dalam PHP selalu diawali dengan tanda $ diikuti dengan nama variabel tersebut.

- Operator

  Operator adalah simbol dalam program yang digunakan untuk melakukan suatu operasi [2].

2.7. SQL (*Structured Query Language*)

SQL adalah kepanjangan dari *Structured Query Language* merupakan bahasa yang banyak digunakan dalam berbagai produk *database*. SQL ini adalah sebuah bahasa yang dipergunakan untuk mengakses data dalam basis data relasional. Saat ini hampir semua server basis data yang ada mendukung bahasa ini untuk melakukan manajemen datanya[13].

Fungsi paling dasar dari SQL adalah menampilkan data dari *database*. Data tersebut dapat difilter dan dimanipulasi sesuai kebutuhan aplikasi. Secara umum perinta-perintah pada SQL terdiri dari dua kelompok yaitu :

- *Data Definition Language* (DDL)
  digunakan untuk mendefinisikan, mengubah, serta menghapus basis data dan objek-objek yang diperlukan dalam basis data, misalnya tabel, view, user, dan sebagainya.
- *Data Manipulation Language* (DML)
  digunakan untuk memanipulasi data yang ada dalam suatu tabel.

Selain mengambil data dari *database*, DML dapat juga melakukan berbagai perhitungan terhadap data tersebut, seperti penjumlahan, perkalian, pembagian dan pengurangan. Sekumpulan fungsi yang agregat yang merupakan gabungan dari beberapa fungsi matematis di atas juga dapat dilakukan DML.

2.7.1. MySQL

MySQL merupakan salah satu *database* server yang sangat terkenal, ini disebabkan MySQL menggunakan SQL sebagai bahasa dasar untuk mengakses databasenya. Selain itu MySQL bersifat *open source*, kecuali untuk jenis *enterprise* yang bersifat komersial. MySQL termasuk jenis *relational database management system* (RDBMS), sehingga istilah seperti tabel, baris dan kolom digunakan dalam MySQL. Pada MySQL sebuah *database* dimungkinkan mengandung satu atau sejumlah tabel [2].

Sebagai *database server*, MySQL dapat dikatakan lebih unggul dibandingkan database *server* lainnya dalam *query* data. Hal ini terbukti untuk *query* yang dilakukan oleh *single user*, kecepatan *query* MySQL bisa sepuluh kali

lebih cepat dari *PostgreSQL* dan lima kali lebih cepat dibandingkan Interbase.

Selain itu MySQL juga memiliki beberapa keistimewaan, antara lain dalam hal *Portability, Open Source, Multiuser, Performance tuning, Column types, Command dan functions, Security, Scalability dan limits, Connectivity, Localisation, Interface, Clients dan tools,* dan *Struktur tabel.*

2.7.2. Tipe data pada MySQL

Pada MySQL mendukung untuk beberapa tipe data yang digunakan seperti untuk tipe data bilangan [2]:

- `TINYINT, SMALLINT, INT, INTEGER, BIGINT, FLOAT, DOUBLE`, dan lainnya

Untuk tipe data tanggal dan jam antara lain :

- `YEAR, DATETIME, DATE, TIMESTAMP, TIME.`

Untuk tipe data karakter antara lain :

- `Char (M), VARCHAR (M), TINYBLOB, TINYTEXT`, dan lainnya.

2.7.3. Fungsi-Fungsi MySQL

Sejumlah fungsi yang berawalan mysql_ digunakan untuk mengakses database server MySQL, berikut beberapa fungsi yang ada pada MySQL [2]:

- mysql_connect (*host, nama_pemakai, password*)
  digunakan untuk membuat hubungan ke database server MySQL yang terdapat pada suatu hostmysql_close (*pengenal_hubungan*)
  digunakan untuk menutup hubungan ke database MySQL.
- mysql_db (*database, pengenal_hubungan*)
  digunakan untuk memilih database.
- dan fungsi-fungsi lainnya yang tersedia.

2.8. Regular Expression

Regular Expression (regex) adalah suatua aturan tertentu yang berguna untuk menentukan valid tidaknya suatu subjek. Mesin regex terdiri dari dua jenis *Text-directed engine* dan *Regex direct engine* atau ada juga yang mengatakan DFA (*Deterministic Finite Automaton*) dan NFA (*Nondeterministic Finite*

Automaton) *engine*. Namun jenis mesin engine yang lebih banyak di minati adalah *regex-directed engine*, disamping itu *feature* nya lebih hebat dari *text-directed engine*.

Regular expression atau yang sering disebut sebagai *Regex* adalah sebuah formula untuk pencarian pola suatu kalimat/*string.* Sering kali orang beranggapan bahwa *regex* susah dan membingungkan. Namun sebenarnya *regex* sangatlah membantu dalam menemukan pola-pola kalimat. Sehingga percobaan terhadap semua kemungkinan pola kalimat tidak perlu dilakukan [10].

Regular expression umumnya digunakan oleh banyak pengolah kata/*text editor* dan peralatan lainnya untuk mencari dan memanipulasi kalimat dengan berdasarkan kepada suatu pola tertentu. Banyak bahasa pemrograman yang mendukung *regular expression* seperti misalnya PHP, perl, VB dan Tcl. Sebuah alasan yang sangat bagus untuk menggunakan *regex* adalah karena *regex* sangatlah *powerfull*. Pada level rendah *regex* dapat mencari sebuah penggalan kata. Pada level tinggi *regex* mampu melakukan kontrol terhadap data. Baik mencari, menghapus dan merubah. Mari kita pikirkan bagaimana cara untuk mencari sebuah file di hard disk. Seringkali digunakan karakter "?" dan "*". Penggunaan karakter "?" mengandung arti bahwa sedang dicari sebuah sebuah file yang mengandung sebuah karakter tertentu dan karakter "*" mengandung arti sedang dicari nol atau lebih karakter [10]. Sebagai contoh : pencarian dengan menggunakan pola "*file?.dat***"** akan menghasilkan beberapa contoh sebagai berikut :

- file1.dat
- file2.dat
- file3.dat
- file.dat
- fileN.dat

Tabel 2.2. Pola umum pada Regex [10]

No.	Penjelasan
[ ]	ekspresi kurung. cocok dengan satu karakter yang berada dalam kurung, misal pattern "a[bcd]i" cocok dengan string "abi", "aci", dan "adi". penggunaan range huruf dalam kurung diperbolehkan, misal : pattern "[a-z]" cocok dengan salah satu karakter diantara string "a" sampai "z". pattern [0-9] cocok dengan salah satu angka. jika ingin mencari karakter "-" juga, karakter tersebut harus diletakkan di depan atau di belakang kelompok, misal: "[abc]".
[^]	cocok dengan sebuah karakter yang tidak ada dalam kurung, berlawanan dengan yang diatas. misal: pattern "[^abc]" cocok dengan satu karakter apa saja kecuali "a", "b", "c".
?	cocok dengan nol atau satu karakter sebelumnya. misal: pattern "died?" cocok dengan string "die" dan "died".
+	cocok dengan satu atau lebih karakter sebelumnya. misal: "yu+k" cocok dengan "yuk", "yuuk", "yuuuk", dan seterusnya.
*	cocok dengan nol atau lebih karakter sebelumnya. misal: pattern "hu*p" cocok dengan string "hp", "hup", "huup" dan seterusnya.
{ x }	cocok dengan karakter sebelumnya sejumlah x karakter. misal: pattern "[09]{3}" cocok dengan bilangan berapa saja yang berukuran 3 digit.
{x, y}	cocok dengan karakter sebelumnya sejumlah x hingga y karakter. misal: pattern "[a-z]{3,5}" cocok dengan semua susunan huruf kecil yang terdiri dari 3 sampai 5 huruf
!	jika diletakkan di depan pattern, maka berarti "bukan". misal pattern "!a.u" cocok dengan string apa saja kecuali string "alu", "anu", "abu", "asu", "aiu", dan seterusnya
^	jika diletakkan di depan pattern, akan cocok dengan awal sebuah string.
$	jika diletakkan di belakang pattern, akan cocok dengan akhir sebuah string
( )	gruping. digunakan untuk mengelompokkan karakter-karakter menjadi single unit. string yang cocok dalan pattern yang berada dalam tanda kurung dapat digunakan pada operasi berikutnya. semacam variable.
\	escape character. mengembalikan fungsi metacharacter menjadi karakter biasa. pada beberapa system dapat berarti sebaliknya, yaitu metacharacter menggunakan escape character didepannya

2.9. Portable Document Format

Portable Document Format (PDF) adalah suatu format file yang digunakan untuk pertukaran dokumen digital, yang dibuat pada tahun 1993 oleh *Adobe system*. PDF tidak memiliki *encoding* yang ditujukan untuk *hardware* atau software tertentu sehingga bisa dibuka diberbagai macam *platform*. PDF ini digunakan untuk merepresentasikan dokumen dua dimensi pada aplikasi perangkat lunak, *hardware* ataupun sistem operasi yang independen, dimana file PDF ini terdiri dari kumpulan deskripsi untuk *layout* dua dimensi seperti text, jenis huruf, citra, dan vektor grafik dua dimensi [11].

Awalnya penggunaan PDF tidak terlalu banyak karena perangkat lunak untuk membuat dan membacanya tidak tersedia secara gratis dan tidak mendukung untuk hiperlink ekstrenal. Ukurannya yang besar juga menjadikan PDF ini kurang populer. Pada saat itu juga PDF harus bersaing dalam tingkat penggunaannya dengan format lain seperti Envoy, *Common Ground Digital Paper*, dan *PostScript* (.ps). *PostScript* adalah format yang juga diciptakan oleh Adobe dan sebagian fungsinya diimplementasikan pada PDF. Laju peningkatan penggunaan dokumen PDF meningkat dengan pesat setelah Adobe mulai mendistribusikan perangkat lunak *Acrobat Reader* secara gratis dan membebaskan pembuatan aplikasi pembuat maupun pembaca dokumen PDF tanpa perlu membayar royalti kepada *Adobe System* selaku pemegang hak paten PDF. Beberapa versi PDF sejak pertamakali dibuata antara lain [11] :

- Tahun 1993 – PDF 1.0 atau Acrobat 1.0.
- Tahun 1994 – PDF 1.1 atau Acrobat 2.0 dengan fitur *Passwords, device-independent color, threads and links.*
- Tahun 1996 – PDF 1.2 atau Acrobat 3.0 dengan fitur *Interactive page elements, mouse events, multimedia types, Unicode, advanced color features and image proxying.*
- Tahun 1999 – PDF 1.3 atau Acrobat 4.0 dengan fitur *Digital signatures; ICC and DeviceN color spaces; JavaScript actions.*
- Tahun 2001 – PDF 1.4 atau Acrobat 5.0 dengan fitur JBIG2; *transparency; OCR text layer.*

- Tahun 2003 – PDF 1.5 atau Acrobat 6.0 dengan fitur JPEG2000; *linked multimedia.*
- Tahun 2005 – PDF 1.6 atau Acrobat 7.0 dengan fitur *Embedded multimedia; XML forms; AES encryption.*
- Tahun 2006 – PDF 1.7 atau Acrobat 8.0.
- Tahun 2008 – PDF 1.7, Adobe Extension Level 3 / Acrobat 9.0.

Pada dasarnya PDF mengkombinasikan tiga teknologi yaitu :

- *Sub-set* dari pemrograman deskripsi halaman *PostScript* untuk menghasilkan tampilan dan grafik.
- Sistem penempatan/pemindahan huruf untuk mengijinkan perpindahan huruf di dalam dokumen.
- Sistem penyimpanan terstruktur untuk menempatkan dan mengkompresi elemen-elemen dokumen ke dalam satu berkas.

BAB 3
PERANCANGAN

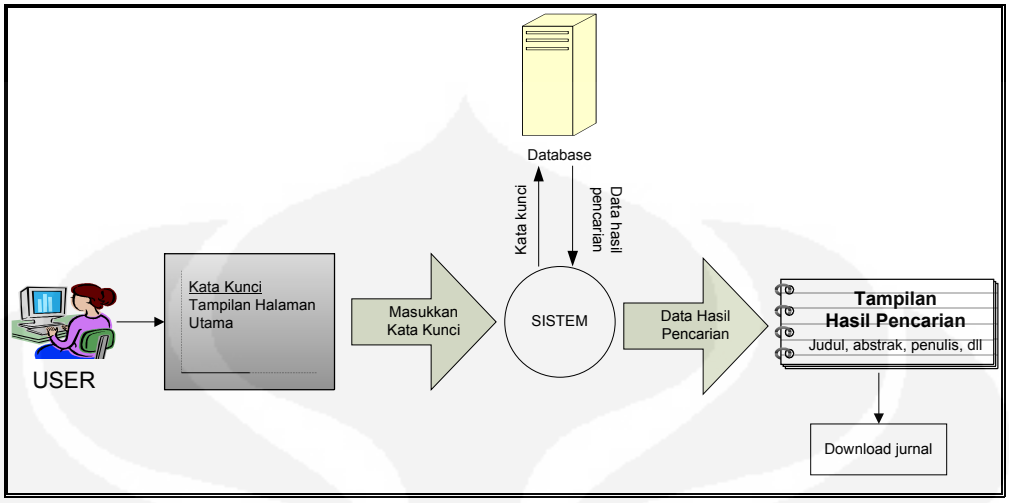
3.1. Prinsip Dasar Aplilkasi

Sistem yang dirancang merupakan sistem berbasis web yang berfungsi untuk mencari dan menampilkan indeks sitasi dari jurnal-jurnal yang bersumber dari universitas yang ada di Indonesia. Jurnal yang diproses oleh sistem adalah jurnal yang dapat diakses bebas dan dalam format file pdf. Sistem menampilkan indeks sitasi dengan menggunakan suatu antarmuka pemakai yang berbentuk halaman web yang dapat diakses bebas memalui internet.

Sistem ini hampir sama dengan mesin pencari atau *search engine* yang sudah kita kenal, seperti *Pdf search engine* atau pun *google search engine*. Pengguna harus memasukan suatu kata kunci berdasarkan judul ataupun penulis pada kolom yang telah di sediakan, dengan terlebih dahulu mengklik *radio button* pada judul ataupun penulis. Setelah perintah pencarian dieksekusi oleh pengguna, maka sistem akan mencari data sesuai kata kunci tersebut pada database lalu menampilkannya kepada pengguna sebagai hasil pencarian.

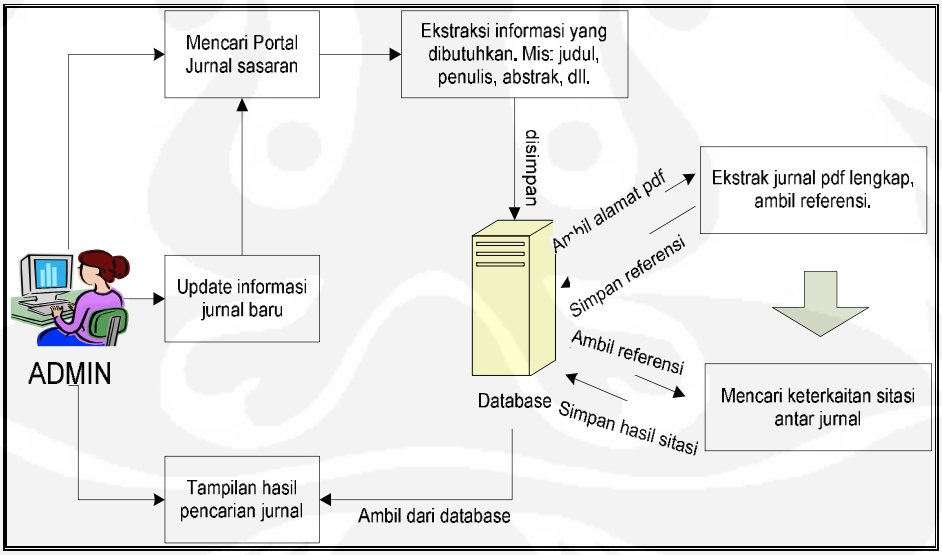
Untuk database sendiri, memakai database lokal yang dibuat menggunakan MySQL, dengan membuat kolom sesuai dengan yang diinginkan, seperti kolom : judul, penulis, alamat, abstrak, dan yang lainnya. Database ini dapat di lakukan *update* atau pembaruan data jurnal yang ada, dengan melihat perubahan isi jurnal yang ada pada portal web yang sudah pernah di jadikan database ataupun adanya penambahan data jurnal dari portal jurnal yang belum dimasukkan menjadi database pada sistem ini.

Untuk menggambarkan urutan kegiatan proses pada sistem ini, maka di buatkan diagram alur kegiatan, dari mulai pengguna memasukan kata kunci sampai pengguna mendapatkan file artikel jurnal ilmiah yang di inginkan, serta urutan proses pembuatan *database*, dari mengekstrak halaman web yang dituju sampai memanggil data pada *database* sesuai pada kata kunci tersebut.



Gambar 3.1. Gambar skema mencari artikel jurnal

Pada Gambar 3.2 adalah gambar diagram alur pembuatan database yang di dapat dari portal web universitas yang kita inginkan untuk dijadikan sumber informasi dari artiket jurnal, yang mempunyai akses bebas bagi publik.



Gambar 3.2. Gambar skema *Mashup*

Tahapan selanjutnya adalah perancangan halaman web yang akan digunakan sebagai antar muka bagi pengguna, halaman web yang dibuat diharapkan memenuhi kriteria berikut :

- Tampilan/*layout* menarik
- Fungsi-fungsi yang ada mudah dipahami dan digunakan sehingga memperkecil tingkat kesalahan pemakaian web oleh pengguna.
- Maksud dari isi yang ditampilkan jelas dan mudah dimengerti juga sesuai dengan tujuan awal pembuatan sistem.

Halaman web yang akan dibuat ini akan terdiri dari tiga bagian utama dengan fungsi dan spesifikasi isi sebagai berikut :

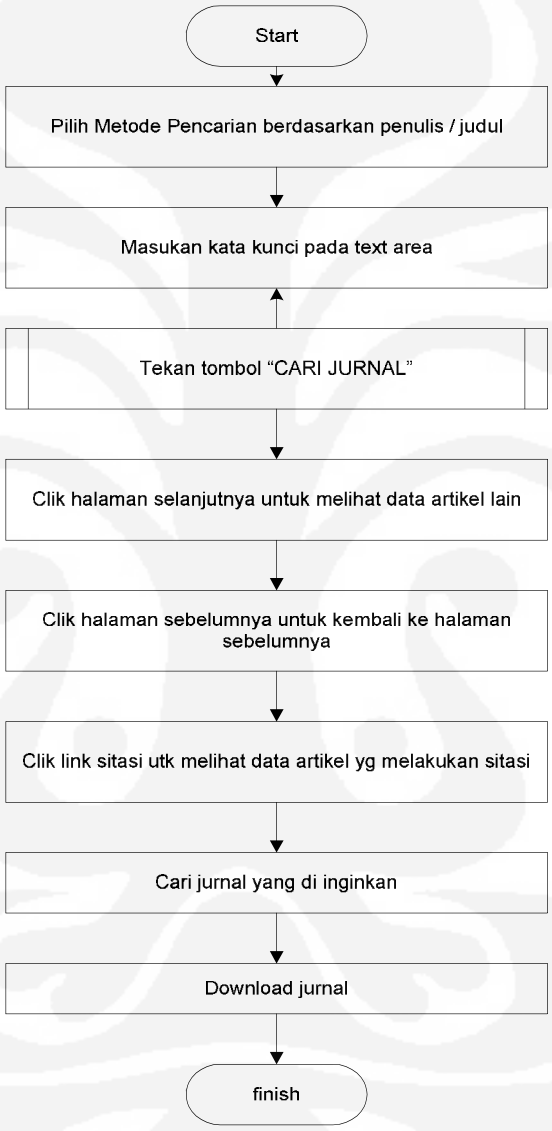
- Kepala halaman

  Kepala halaman ini terdiri dari :
  - Judul / nama web.
  - Kotak teks untuk mengisikan kata kunci pencarian jurnal.
  - Tombol pilihan untuk pemilihan tipe kata kunci (judul atau penulis).
  - Tombol untuk memulai pencarian.
  - *Link* untuk melihat halaman berikutnya.
  - *Link* untuk melihat halaman sebelumnya.
  - Teks untuk menampilkan hasil jumlah pencarian.
  - Teks untuk menampilkan urutan jumlah hasil pencarian yang ditampilkan.
- Badan Halaman

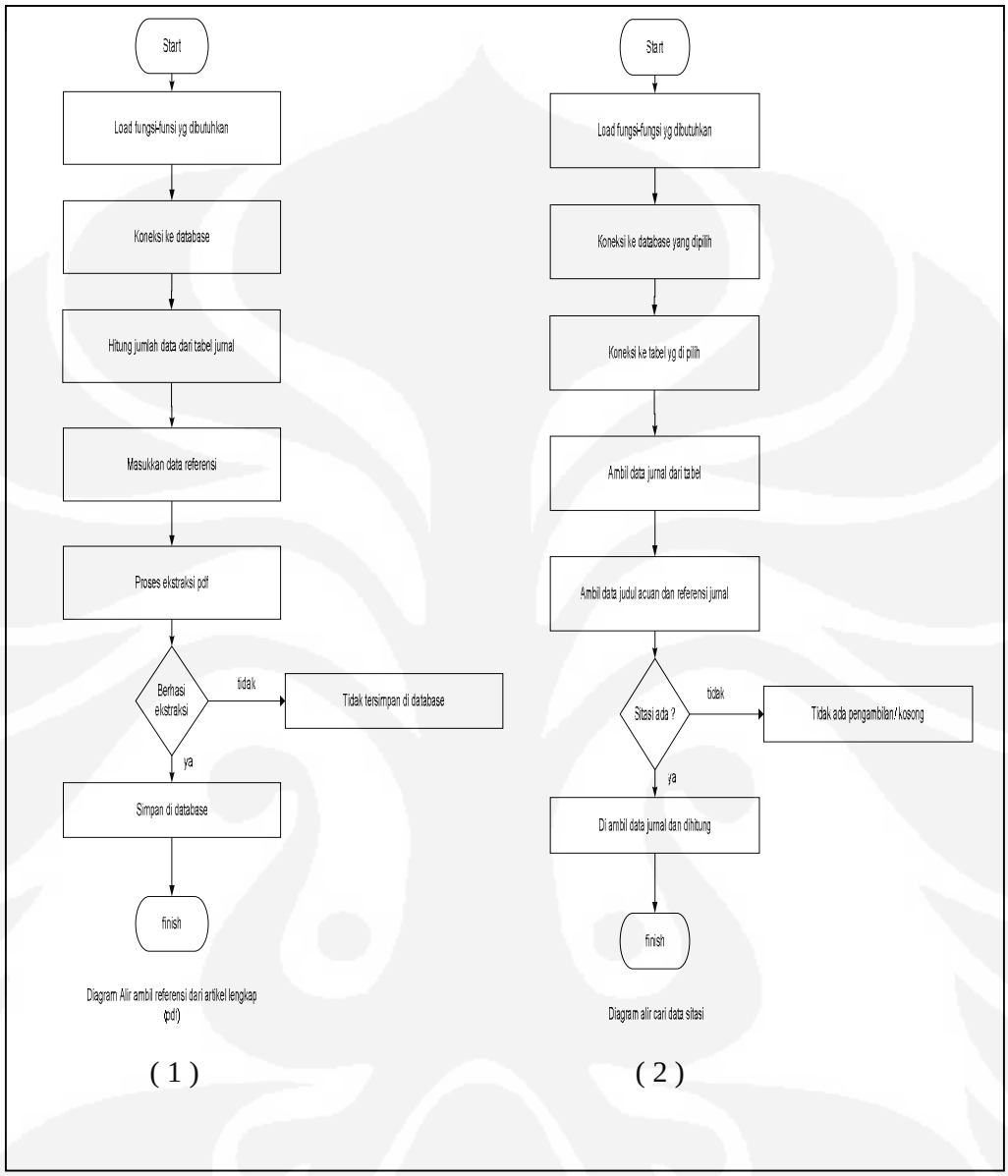
  Badan halaman ini berisi data-data hasil pencarian yang terdiri dari :
  - Judul artikel jurnal (beserta *link* ke halaman sebenarnya).
  - Penulis.
  - Penerbit.
  - Institusi penulis.
  - Potongan abstraksi.
  - Jumlah sitasi terhadap judul jurnal diatas.
  - Jumlah data jurnal yang ditampilkan per halaman adalah 10 data artikel.
- Kaki halaman

  Berupa grafik atau tulisan pelengkap identitas web.

Setelah mengetahui gambar skema dari pencarian artikel jurnal dan skema mashup untuk membuat *database* lokal secara garis besar, maka untuk menggambarkan alur kerja dari fungsi-fungsi yang akan dibuat pada halaman web tersebut digunakan diagram alir seperti terlihat pada Gambar 3.3 berikut :



Gambar 3.3 Alir kerja pencarian halaman utama



Gambar 3.4. Diagram alir main program halaman web

Gambar

Diagram alir untuk ambil data referensi dari artikel lengkap (1), Diagram alir untuk cari data sitasi (2).

Setelah semua perancangan yang terkait dengan cara kerja sistem secara keseluruhan didefinisikan, baik itu mengenai interaksi antara pengguna dengan sistem, urutan komunikasi antar komponen yang ada pada sistem, variabel dan data yang digunakan pada sistem dan juga alur atau langkah kerja dari sistem (sesuai dengan fungsi masing-masing komponen yang ada pada sistem), maka proses implementasi siap dikerjakan. Halaman web yang akan digunakan sebagai antarmuka dengan pengguna beserta masing-masing fungsinya. Selanjutnya pembuatan sistem dengan melakukan *coding* untuk fungsi-fungsi yang terdapat pada masing-masing komponen sehingga keseluruhan sistem dapat terealisasi seluruhnya, dan dapat bekerja dan dipergunakan sesuai dengan yang diharapkan dan tujuan awal pembuatan sistem. Untuk pembuatannya digunakan bahasa PHP dan MySQL sebagai *database*.

BAB 4
PENGUJIAN DAN ANALISA

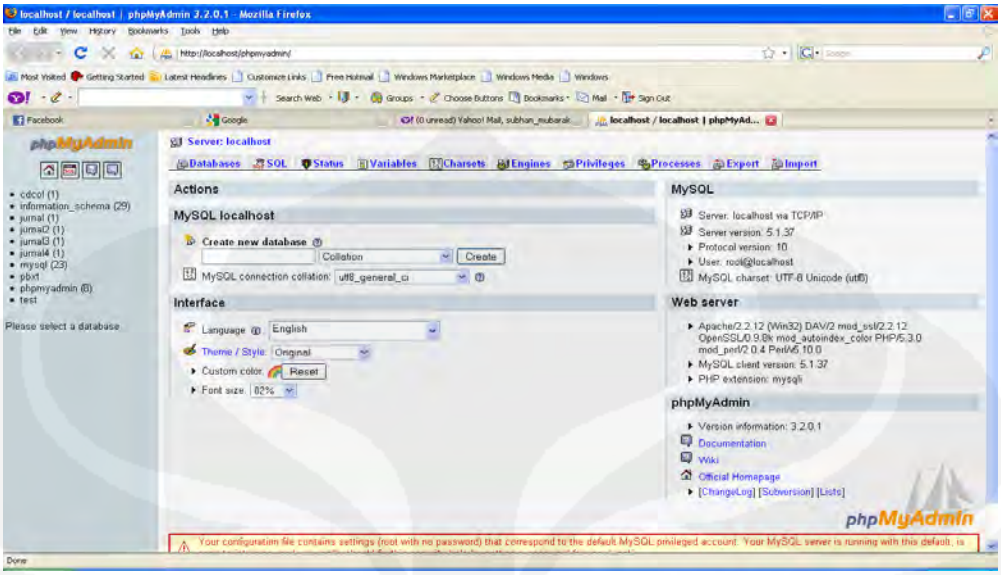
4.1. UJI COBA APLIKASI

Bab 4 ini akan membahas implementasi dari aplikasi ekstraksi web untuk keperluan indeks sitasi. Sistem bekerja dengan mengekstraksi halaman-halaman web penyedia jurnal dengan situs-situs berikut :

- http://repository.gunadarma.ac.id (Universitas Gunadarma)
- http://journal.unair.ac.id (Universitas Airlangga)
- http://journal.uii.ac.id (Universitas Islam Indonesia)
- http://ejurnal.gunadarma.ac.id (Universitas Gunadarma)
- http://journal.ui.ac.id (Universitas Indonesia)
- http://digilib.unsri.ac.id (Universitas Sriwijaya)
- http://ejournal.unud.ac.id (Universitas Udayana)
- http://puslit2.petra.ac.id (Universitas Kristen Petra)
- http://i-lib.ugm.ac.id (Universitas Gajah Mada)
- http://eprints.undip.ac.id (Universitas Diponogoro)
- http://proceeding.itb.ac.id (Institut Teknologi Bandung)
- http://jurnal.ump.ac.id (Universitas Muhammadiyah Purwokerto)

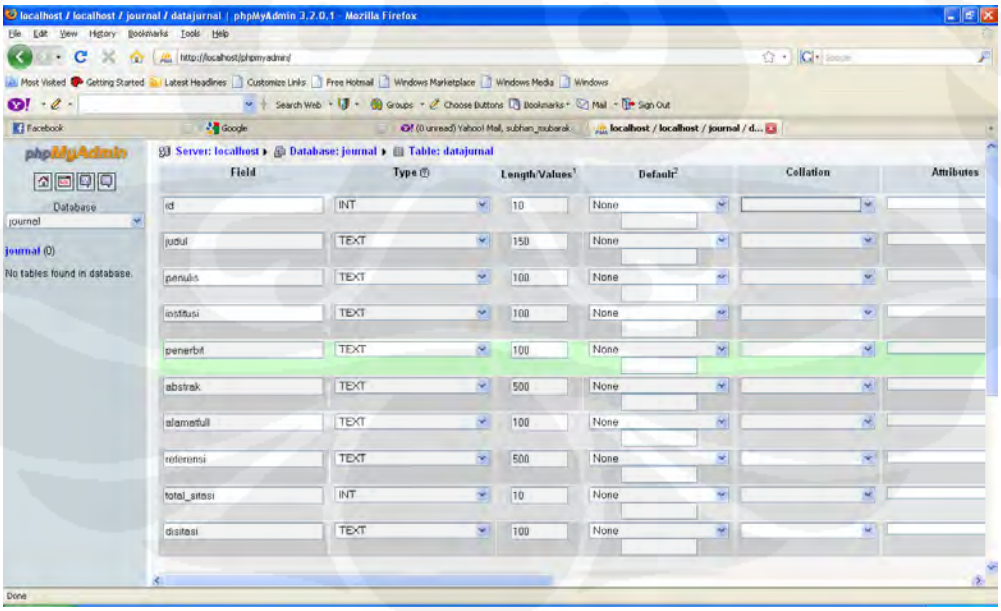
4.2. LANGKAH – LANGKAH PENGUJIAN

Langkah Pertama yang harus dilakukan dalam melakukan pengujian ini adalah dengan membuat atau mempersiapkan kolom *database* pada *database* MySQL. Pembuatan *database* sangat penting sebagai sarana tempat menyimpan data-data yang telah kita ekstrak. Pembuatan *database* ini dapat dibuat dengan cara manual maupun dengan cara otomatis dengan membuat bahasa *query* pada *notepad*, lalu diproses pada MySQL. Sebelumnya kita harus mengaktifkan dahulu *localhost* atau *phpmyadmin*. Berikut adalah cara membuat kolom *database* secara manual :



Gambar 4.1 Tampilan pembuatan tabel *database* di MySQL phpmyadmin

Setelah MySQL *localhost* diaktifkan, kita harus membuat nama *database*, nama tabel dan jumlah kolom yang diperlukan, setelah kolom terbentuk lalu masukkan judul atau kepala kolom sesuai yang di inginkan, dan *setting* tipe data dari masing-masing kolom dari tabel yang ingin dibuat.



Gambar 4.2. Pengaturan tipe kolom tabel

Pada kolom “Id” di *setting* sebagai *Primary Key* dan *Auto Increment*, sebagai pusat atau induk untuk pemanggilan dan penghitungan database. Jika ingin cara cepat membuat kolom dapat dengan meng-*import query* yang sebelumnya di buat pada *notepad*, dengan syarat mengetahui format penulisan perintah *query*, sebagai contoh pada Gambar 4.3. berikut ini :

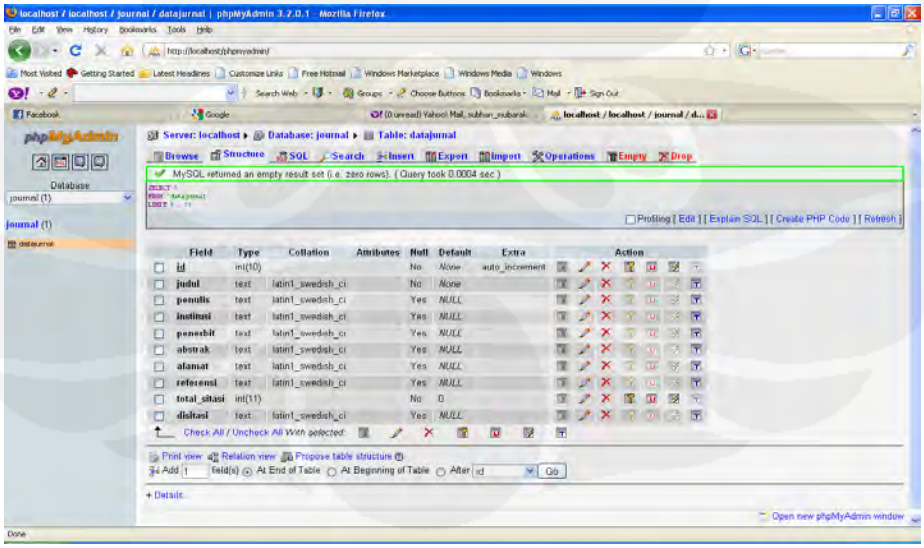
```
-- phpMyAdmin SQL Dump
-- version 3.2.0.1
-- http://www.phpmyadmin.net
--
-- Host: 127.0.0.1:3306
-- Generation Time: Mar 31, 2010 at 04:38 PM
-- Server version: 5.1.36
-- PHP Version: 5.3.0

SET SQL_MODE="NO_AUTO_VALUE_ON_ZERO";

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8 */;
--
-- Database: `jurnal4`
--
-- --------------------------------------------------------
--
-- Table structure for table `datajurnal`
--
CREATE TABLE IF NOT EXISTS `datajurnal` ( `id` int(10) NOT NULL AUTO_INCREMENT,  `judul` text NOT NULL,
`penulis` text,  `institusi` text,  `penerbit` text,  `abstrak` text,  `alamat` text,  `referensi` text,  `total_sitasi` int(11) NOT NULL
DEFAULT '0',  `disitasi` text,  PRIMARY KEY (`id`),  UNIQUE KEY `id` (`id`)) ENGINE=MyISAM  DEFAULT
```

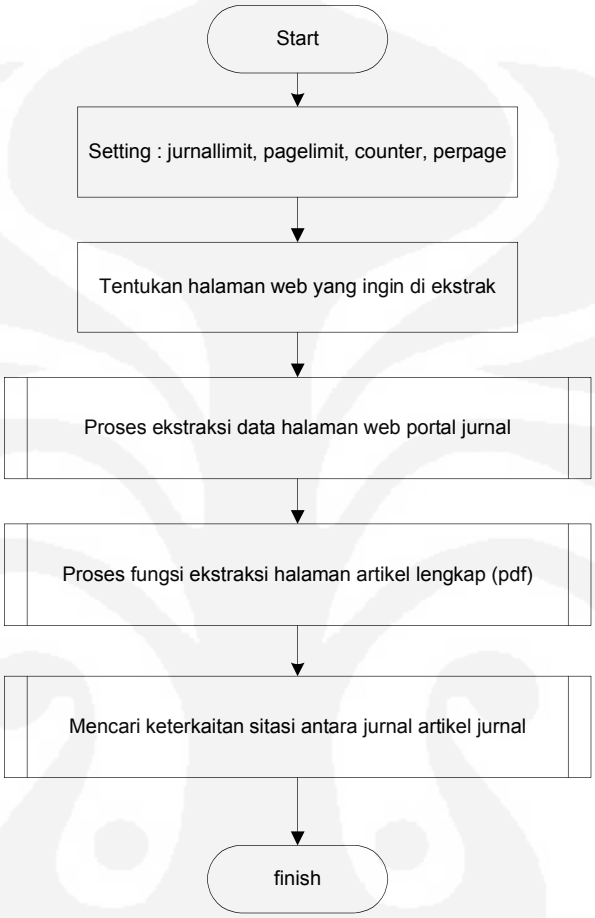
Gambar 4.3 Contoh skrip perintah *query*

Setelah Penulisan *query* sudah dianggap benar, selanjutnya *import* data di proses. Jika benar maka akan ada pernyataan bahwa *import query* berhasil dan akan menghasilkan kolom yang diinginkan, seperti pada Gambar 4.4 :



Gambar 4.4. Tampilan Hasil Pembuatan Judul Kolom Database

Langkah berikutnya adalah proses ekstraksi *web journal*. Pada proses ini bagian yang akan di ekstrak akan menempati kolom yang telah ditentukan. Gambar 4.5 berikut adalah gambar diagram alur dari proses ekstraksi web.

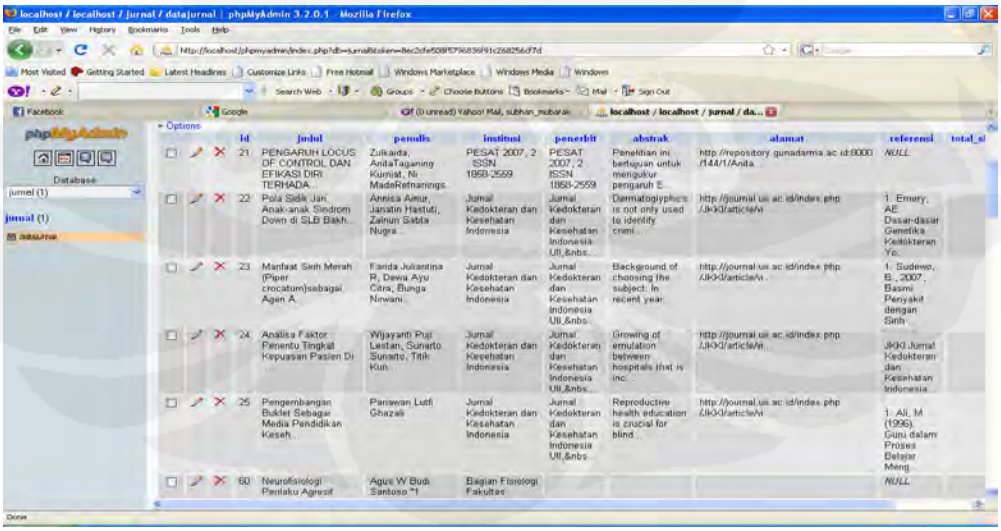


Gambar 4.5 . Diagram Alir Fungsi Ekstraksi Halaman Web

Eksekusi aplikasi ekstraksi halaman web sesuai diagram alir Gambar 4.5, akan menghasilkan informasi atau data yang berhasil diekstraksi dari situs penyedia jurnal, dan juga informasi tambahan terkait eksekusi aplikasi program tersebut. Kemudian akan ditampilkan pada halaman *browser*, seperti terlihat pada Gambar 4.6.

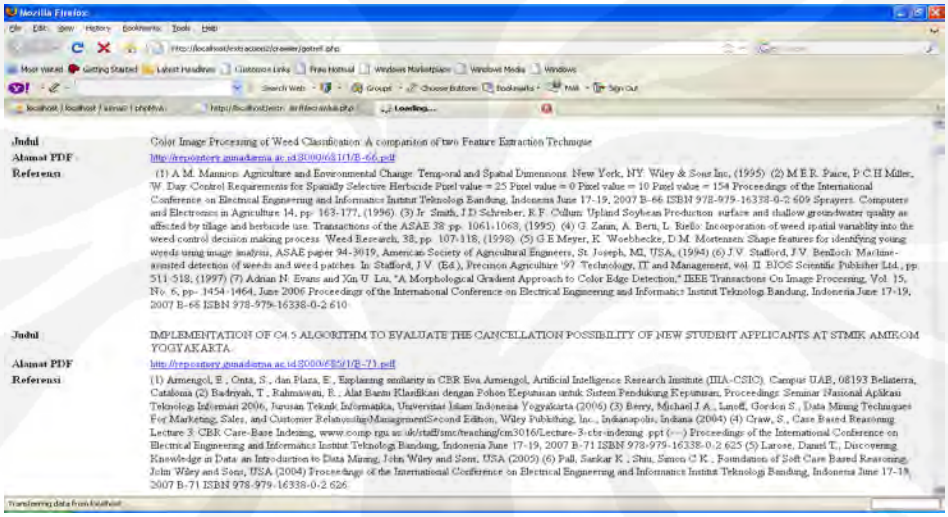
Gambar 4.6 . Hasil Tampilan Aplikasi Program Ekstraksi Halaman Web

Hasil tampilan proses ekstraksi diatas akan di masukkan pada tabel database yang telah dibuat sebelumnya secara otomatis. Jumlah data artikel jurnal baru yang berhasil diekstraksi dari proses eksekusi aplikasi ekstraksi halaman web adalah sebanyak 16631 dengan rincian pada Gambar Tabel 4.7 berikut :



Gambar 4.7. Kolom database hasil ekstraksi

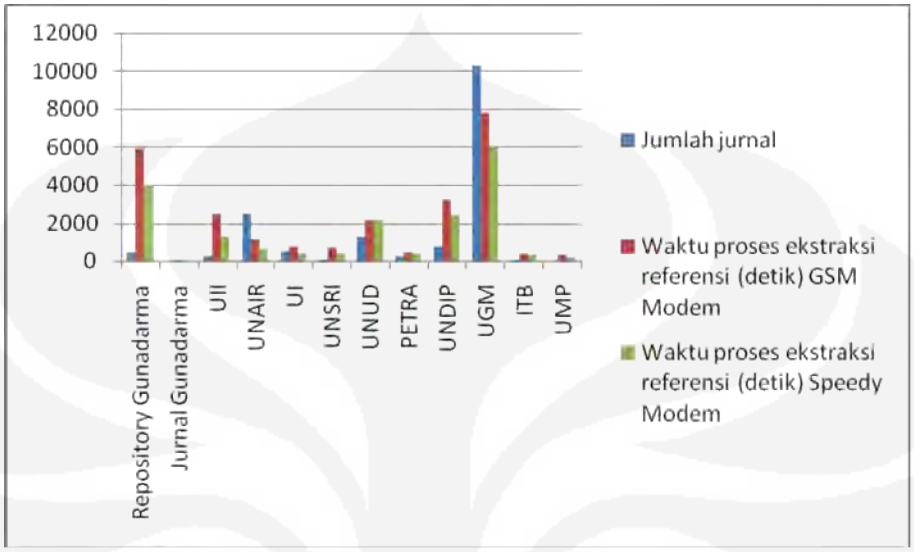
Setelah melakukan ekstraksi pertama, yaitu ekstraksi judul, penulis, institusi, penerbit, abstrak dan alamat jurnal. Maka selanjutnya adalah mengekstraksi referensi, proses ekstraksi referensi ini dilakukan dengan proses yang berbeda dari proses ekstraksi pertama, karena pada sistem ini menggunakan software pdftotext untuk mengkoversi bentuk pdf menjadi text. Hasil konversi tersebut di letakkan ke dalam database. Hasil dari ekstraksi referensi di gunakan untuk melakukan proses berikutnya, yaitu proses sitasi. Berikut Gambar hasil proses ekstraksi.



Gambar 4.8. Hasil proses ekstraksi referensi

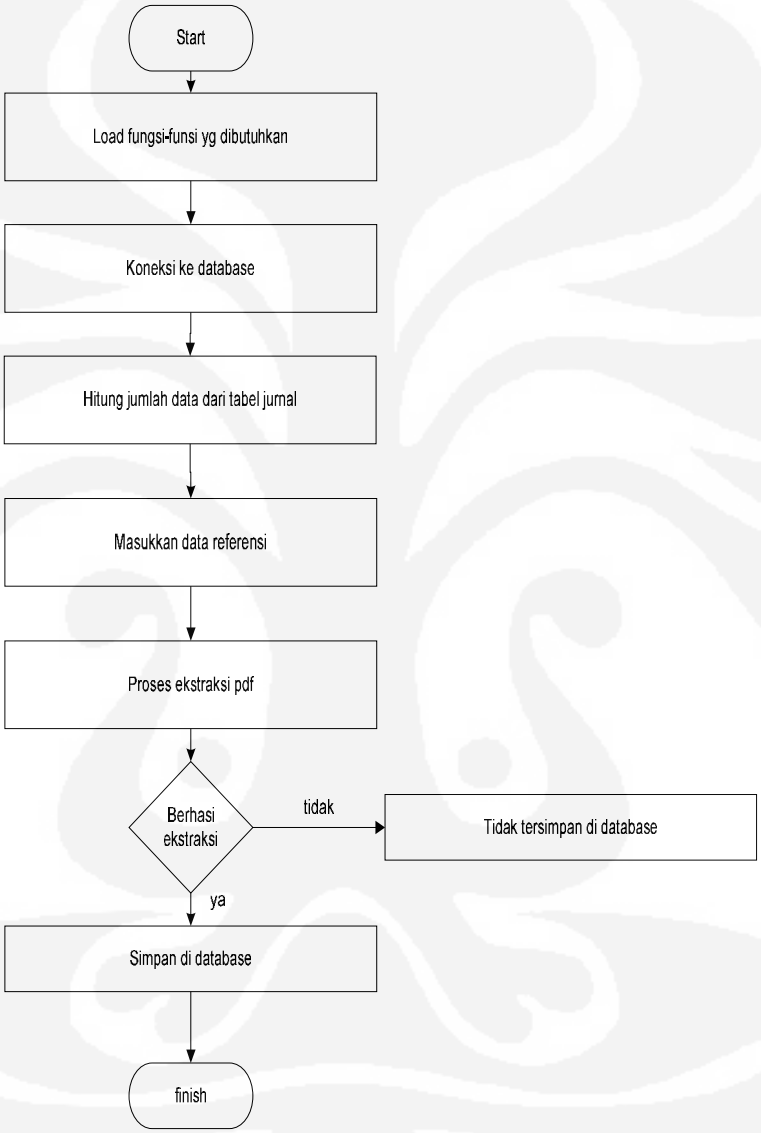
Tabel 4.1.Hasil Pengujian proses ekstraksi keseluruhan referensi jurnal.

Parameter	Jumlah jurnal	Waktu proses ekstraksi referensi (detik)	
		GSM Modem	Speedy Modem
Repository Gunadarma	440	5899	3949
Jurnal Gunadarma	22	48	27
UII	301	2483	1302
UNAIR	2474	1117	632
UI	524	732	402
UNSRI	97	697	386
UNUD	1287	2120	2210
PETRA	264	434	396
UNDIP	763	3220	2437
UGM	10307	7820	5989
ITB	106	367	375
UMP	46	332	232



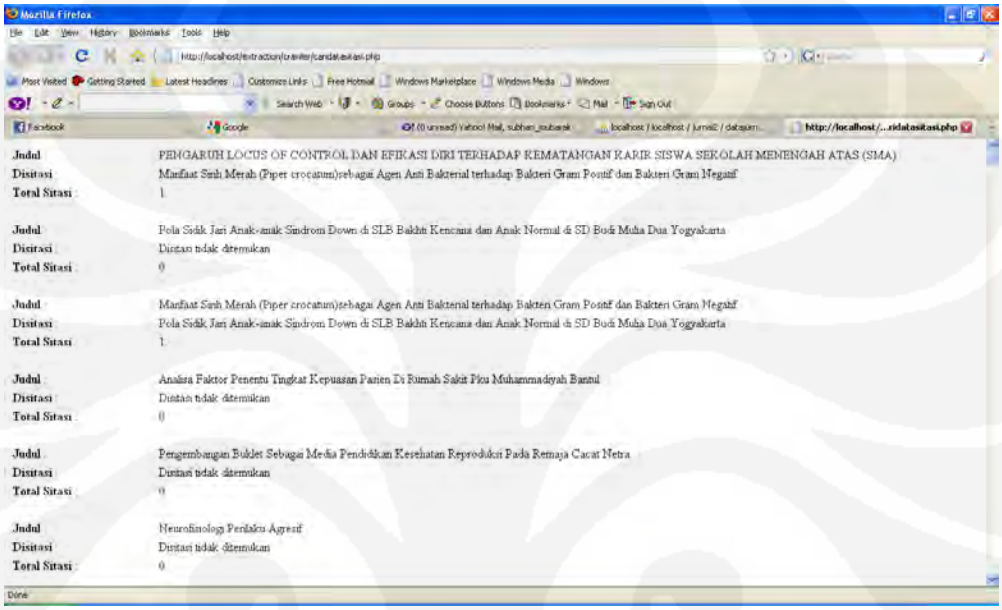
Gambar 4.9. Grafik hasil pengujian proses ekstraksi keseluruhan referensi jurnal

Hasil dari pengujian ekstraksi referensi, dalam hal ini lamanya waktu yang di peroleh dalam proses ini dipengaruhi oleh, banyaknya referensi yang di ekstrak, ditemukan atau tidaknya, alamat pdf tersebut, serta banyaknya jumlah jurnal yang di ekstrak. Gambar 4.10 menjelaskan diagram alir proses ekstraksi referensi.



Gambar 4.10. Diagram alir proses ekstraksi referensi

Proses berikutnya adalah proses sitasi, yaitu proses mencari keterkaitan dari referensi atau daftar pustaka yang di pakai oleh dua jurnal atau lebih yang berbeda judul tetapi mempunyai kemiripan pembahasan. Hasil dari proses sitasi seperti gambar berikut :



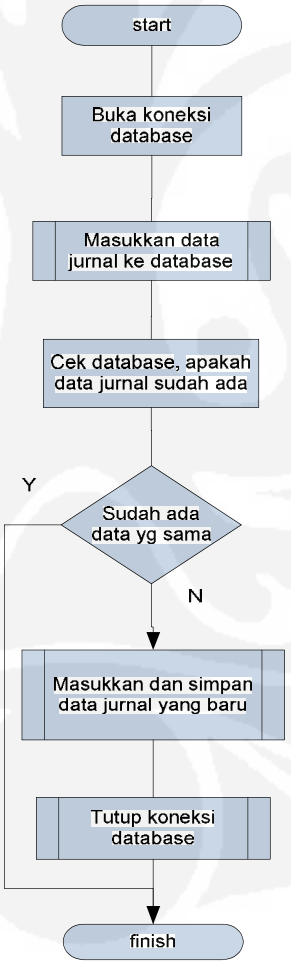
Gambar 4.11. Tampilan hasil proses sitasi

Tabel.4.2. Sumber Dan Jumlah Artikel Jurnal Yang Berhasil Diekstraksi

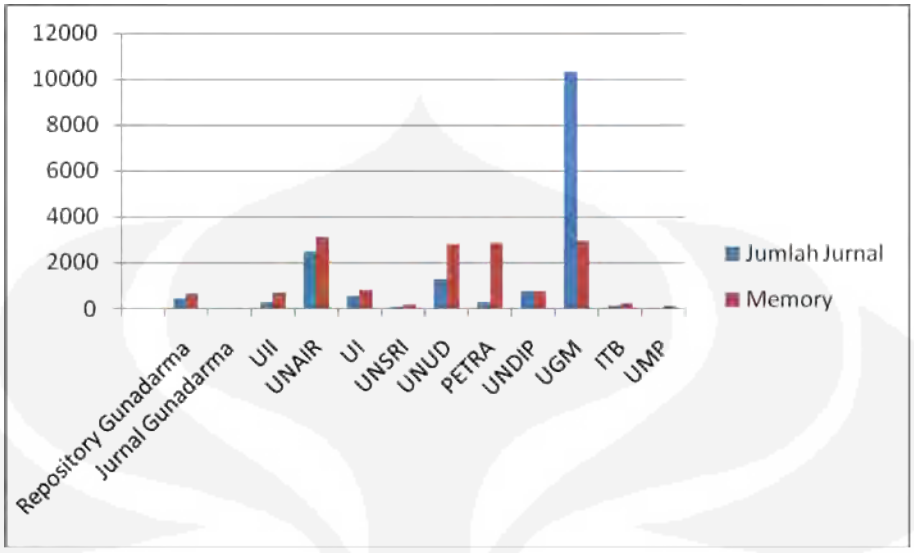
Sumber	***jumlah artikel hasil ekstraksi***
Universitas Gunadarma repository	440
Universitas Airlangga	2474
Universitas Islam Indonesia	301
Universitas Gunadarma	22
Universitas Indonesia	524
Universitas Sriwijaya	69
Universitas Udayana	1287
Universitas Kristen Petra	484
Universitas Diponogoro	763
Universitas Gajah Mada	10307
Institut Teknologi Bandung	106
Universitas Muhamammadiah Purwokerto	46

Tabel 4.3. Waktu Eksekusi untuk Aplikasi Fungsi Ekstraksi Halaman Web

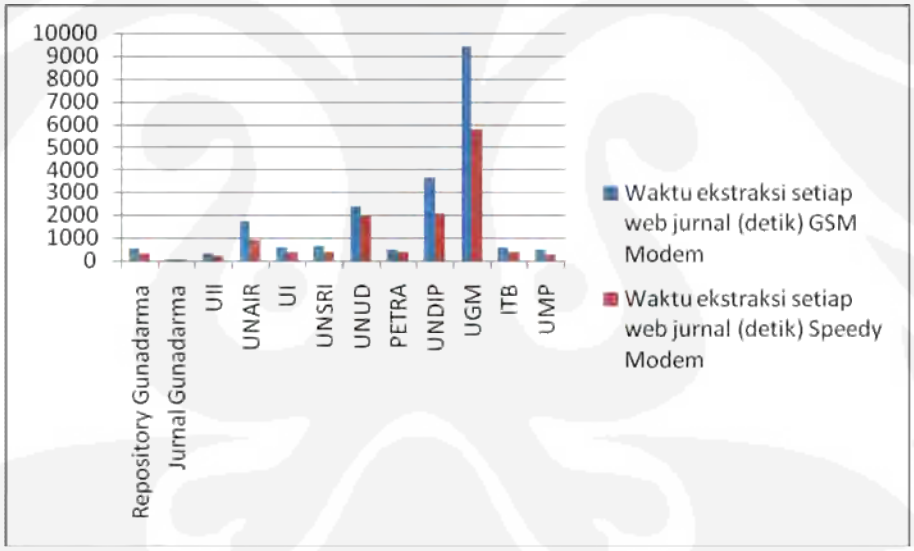
Parameter	Jumlah Jurnal	Memory	Waktu ekstraksi setiap web jurnal (detik)	
			GSM Modem	Speedy Modem
Repository Gunadarma	440	636,9	507,6	267
Jurnal Gunadarma	22	27,8	52,79	28,2
UII	301	701	312	162
UNAIR	2474	3135,3	1734	897
UI	524	822,8	569	332
UNSRI	97	164,2	645	345
UNUD	1287	2787,1	2400	1914
PETRA	264	2865,3	484	322
UNDIP	763	754,2	3660	2067
UGM	10307	2975	9420	5746
ITB	106	252,5	569	332
UMP	46	124	452	216



Gambar 4.12. Diagram alir fungsi input ke *Database*



Gambar 4.13. Grafik perbandingan jumlah jurnal dengan memory yang terpakai



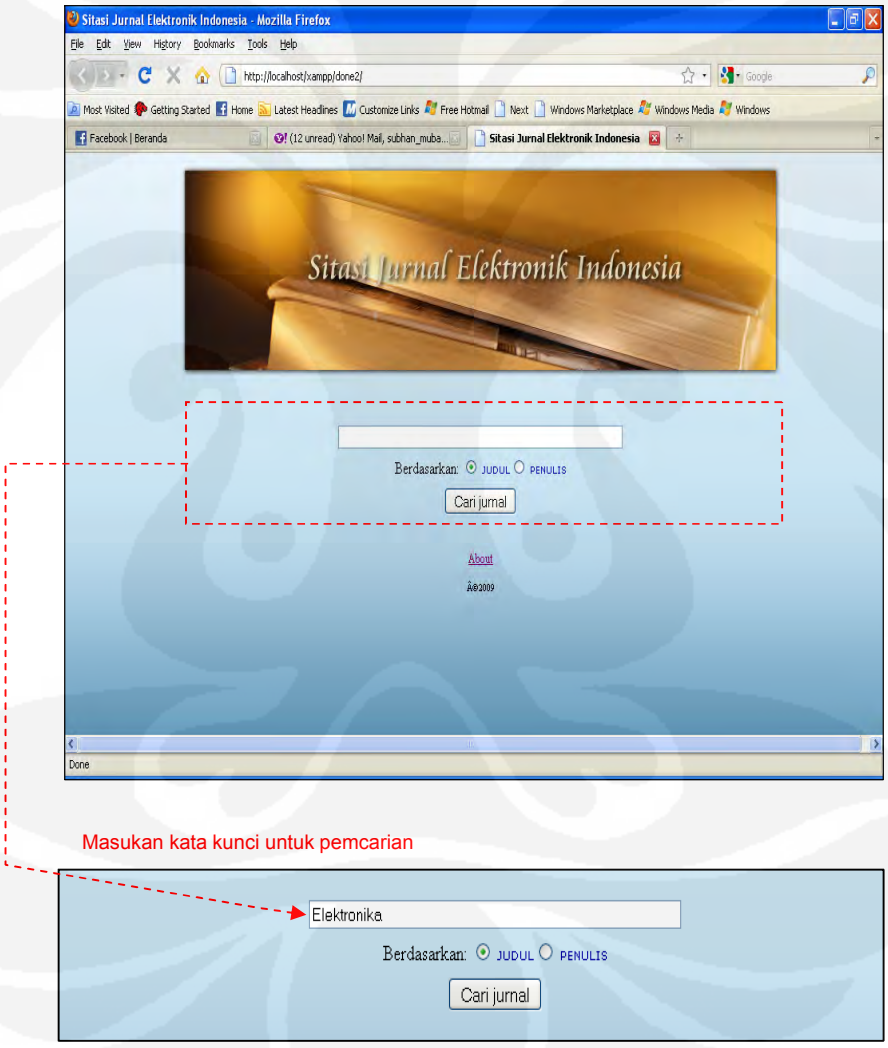
Gambar 4.14. Grafik fungsi ekstraksi halaman web dengan Speedy Modem

Gambar grafik diatas adalah hasil dari perbandingan jumlah jurnal yang didapat dengan dengan memori yang di terpakai, serta dibandingkan dengan waktu ekstraksi dari modem yang berbeda.

4.3. Pengujian Tampilan Searching.

Pengujian Tampilan pencarian atau sistem halaman utama dilakukan juga dengan cara menjalankan aplikasi program. Aplikasi program yang dijalankan meliputi semua fungsi yang terdapat pada halaman antarmuka, baik halaman antamuka pembuka ataupun halaman antarmuka utama.

Hasil tampilan halaman antarmuka pengguna, yaitu halaman pembuka dapat dilihat pada Gambar 4.15 di bawah ini.



Gambar 4.15. Hasil tampilan halaman antarmuka pembuka

Tampilan halaman antarmuka utama yang digunakan untuk menampilkan hasil pencarian data artikel jurnal, melakukan proses pencarian selanjutnya, dan melihat data artikel jurnal yang melakukan sitasi terhadap artikel jurnal lainnya, dapat dilihat pada Gambar 4.16 berikut ini:

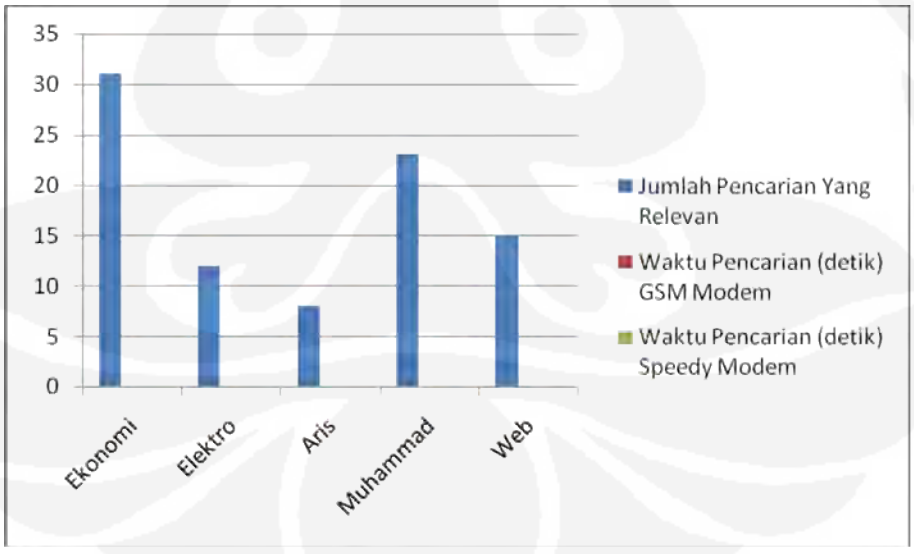
Gambar 4.16. Tampilan hasil pencarian halaman utama

Pengujian yang dilakukan meliputi beberapa fungsi berikut:

- Pengujian fungsi pencarian dengan kata kunci judul.
- Pengujian fungsi pencarian dengan kata kunci penulis.
- Pengujian fungsi melihat halaman selanjutnya.
- Pengujian fungsi melihat halaman sebelumnya.
- Pengujian fungsi melihat data artikel yang melakukan sitasi.
- Pengecekan hasil tampilan yang ada pada halaman antarmuka dan *link* untuk melihat artikel lengkap dari data artikel yang ditampilkan.

Tabel 4.4. Pengujian pencarian kata kunci

Kata Kunci	Jumlah Pencarian Yang Relevan	Waktu Pencarian (detik)	
		GSM Modem	Speedy Modem
Ekonomi	31	0,091	0,091
Elektro	12	0,081	0,081
Aris	8	0,034	0,034
Muhammad	23	0,038	0,038
Web	15	0,039	0,039

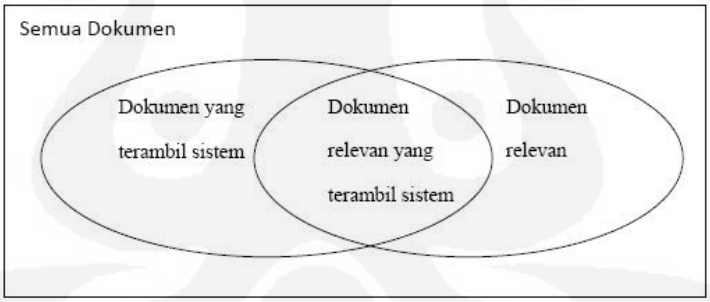


Gambar 4.17. Grafik pengujian kata kunci

Dari data table diatas dapat kita lihat, pada umumnya lama waktu yang di dapat dalam pencarian kata kunci untuk ditampilan pada tampilan utama, tergantung dari jumlah jurnal yang relevan yang ditemukan, semakin banyak jumlah kata kunci yang relevan ditemukan, semakin lama waktu yang di butuhkan. Tetapi pada kasus lain ada juga yang dapat menyebabkan lamanya waktu pencarian, yaitu banyaknya abstrak yang ditampilkan dan penempatan urutan *database* dari jurnal tersebut.

4.4. Analisa Data

Pada pengujian ini menggunakan sistem *search subsystem* (*matching*), yang merupakan proses menemukan kembali informasi (dokumen) yang relevan terhadap *query* yang diberikan. Tidak semua dokumen yang diambil (*retrieved*) oleh sistem merupakan dokumen yang sesuai dengan keinginan user (*relevant*). Gambar 4.18 dibawah ini menunjukkan hubungan antara dokumen relevan, dokumen yang terambil oleh sistem, dan dokumen relevan yang terambil oleh system [6].



Gambar 4.18. Relevansi dokumen

Pengukuran ini dikenal dengan perhitungan *recall*, *precision*, dan *Interpolate Average Precision* (IAP) [6].

$$\textbf{\textit{Precision}} = \frac{\textit{Jumlah dokumen yg relevan dan terambil}}{\textit{Jumlah seluruh dokumen terambil}} \times 100\%$$

$$\textbf{\textit{Recall}} = \frac{\textit{Jumlah dokumen yg relevan dan terambil}}{\textit{Jumlah Dokumen yg relevan}} \times 100\%$$

Precision (ketepatan) ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen yang terambil oleh sistem baik relevan maupun tidak relevan. *Recall* (kelengkapan) ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen relevan yang ada dalam koleksi dokumen (terambil ataupun tak terambil oleh sistem) [6].

Nilai performansi dari aplikasi *Information Retrieval* (IR) menunjukkan keberhasilan dari suatu *Information Retrieval System* (IRS) dalam mengembalikan informasi yang dibutuhkan oleh *user*. Untuk mengukur performansi dari IRS, digunakan koleksi uji. Koleksi uji terdiri dari tiga bagian, yaitu koleksi dokumen, *query*, dan *relevance judgement*. Koleksi dokumen adalah kumpulan dokumen yang dijadikan bahan pencarian oleh sistem. *Relevance judgement* adalah daftar dokumen-dokumen yang relevan dengan semua *query* yang telah disediakan.

Yang mana *Information Retrieval* merupakan bagian dari ilmu komputer yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Berdasarkan referensi dijelaskan bahwa *Information Retrieval* merupakan pencarian informasi berupa dokumen yang didasarkan pada suatu *query* (*input*an *user*) yang diharapkan dapat memenuhi keinginan *user* dari kumpulan dokumen yang ada. Definisi *query* dalam *Information Retrieval* adalah sebuah formula yang digunakan untuk mencari informasi yang dibutuhkan oleh *user*, dalam bentuk yang paling sederhana, sebuah *query* merupakan suatu *keywords* (kata kunci). Dokumen yang mengandung *keywords* merupakan dokumen yang dicari dalam IRS [6].

Tabel 4.5. Perhitungan *Pressicion* dan *Recall*

No.	Total Jurnal	Kata Kunci					Pressicion	Recall
		Pencarian 1	Juml terambil	Pencarian 2	Juml relevan	Relevan &terambil		
1.	2787	Elektro	12	Elektroforesis	4	4	0,33%	100%
2.		Ekonomi	31	Ekonomi mikro	4	4	0,12%	100%
3.		Web	15	Website	3	0	0,2%	0%
4.		Muhammad	23	Muhammad Ridwan	2	2	0,008%	100%
5.		Aris	8	Aris Junaidi	1	1	0,125	100%

4.5 Keterbatasan Sistem

Dari hasil *output* yang dihasilkan sistem sudah dapat mengerjakan kekurangan dari sistem terdahulu, yang diketahui sistem terdahulu belum dapat mengekstraksi semua tipe file pdf (yaitu pdf diatas tipe 1.4). Tetapi sistem yang baru ini, masih belum dapat mengekstrak portal jurnal yang memerlukan *password* untuk login dahulu (tidak terbuka). Sumber situs penyedia jurnal elektronik yang diekstrak terbatas sebanyak 12 situs.

4.6 Pengembangan

Dari hasil pengujian sistem, diperlukan pengembangan lanjutan dari sistem yang telah ada. Pengembangan dilakukan dengan mengurangi keterbatasan yang ada pada sistem ini, seperti dapat mengekstrak web jurnal yang di *password*, dapat mengekstrak atau pun men-*download* jurnal yang mempunyai format tulisan tidak hanya format PDF saja, seperti *word, wordpad,* dll. Dan menambahkan lebih banyak lagi web jurnal yang diekstrak.

Agar tercapai sistem sitasi jurnal elektronik Indonesia yang lengkap, diperlukan pengembangan dalam hal ekstraksi untuk semua tipe file pdf, diperlukan penambahan sumber-sumber institusi penerbit jurnal, dan aplikasi ekstraksi halaman web dengan cara yang berbeda tetapi dapat mengikuti perubahan pada halaman web penyedia jurnal. Agar dapat mengekstraksi portal jurnal yang di login, walaupun izin terlebih dahulu.

BAB 5
KESIMPULAN

Setelah dilakukan perancangan, implementasi, uji coba dan analisa aplikasi Sistem Informasi Sitasi Jurnal Elektronik Indonesia Dari Banyak Sumber *E-Journal* yang dilakukan dengan mengambil informasi dari portal penyedia jurnal, dan membuat *mashup* gabungan informasi, dapat diambil beberapa kesimpulan :

- Sistem informasi sitasi jurnal elektronik ini memudahkan seseorang dalam mengelola pengumpulan serta pencarian artikel jurnal ilmiah.
- Program ini akan berproses jika database, dalam hal ini MySQL di aktifkan terlebih dahulu.
- Dalam pengujian dan analisa program mempunyai kesimpulan :
  1. Untuk pengujian ekstraksi judul, penulis, penerbit, abstrak dan alamat jurnal di pengaruhi oleh banyaknya jurnal yang di ekstraksi dan *traffic* jaringan internet.
  2. Pada pengujian ekstraksi referensi dipengaruhi oleh banyaknya referensi yang di ekstrak, ditemukan atau tidaknya alamat pdf tersebut, serta banyaknya jumlah jurnal yang di ekstrak.
  3. Pengujian Sitasi dipengaruhi oleh berhasil atau tidaknya proses ekstraksi referensi ke database.
  4. pengujian pencarian (*searching*) kata kunci pada halaman utama, lama waktu proses ekstraksi memerlukan waktu tercepat 38 detik dan waktu terlama 6899. Untuk memori database memerlukan kapasitas 10 Mb. Pada analisa *pressicion* di dapatkan data bahwa ketepatan suatu pencarian tidak dapat mencapai 100%, semakin banyak data jurnal yang di miliki dalam database, semakin tidak dapat mencapai ketepatan 100%. Sedangkan dalam kelengkapan (*Recall*) di mungkinkan sampai kelengkapan 100 %.
- Secara umum sistem dapat bekerja dengan baik, dan bisa menampilkan artikel jurnal dengan mengeksekusi kata kunci.
- Sistem ini belum dapat mengekstrak web jurnal yang dilogin. Mengurutkan hasil sitasi yang didapat, dari yang terbanyak.

DAFTAR REFERENSI

[1] *Screen scraping*.http://en.wikipedia.org/wiki/Screen_scraping, diakses terakhir 3 Maret 2009.

[2] Kadir, Abdul. Dasar Pemrograman Web PHP. Maret, 2008.

[3] Widyaseno, Zulfikar, FX Ferdinand, Ruki Harwahyu dan Reza hadi S. Penggunaan Teknik Ekstraksi Web dalam Pengolahan referensi kepustakaan. Departermen Teknik Elektro Universitas Indonesia

[4] Wulandari, Lily, I wayan s wicaksana. S*emantic web.*

[5] Palmer, sean B. *the semantic web: an introduction.* http://infomesh.net/2001/swintro/, diakses terakhir 9 Februari 2009.

[6]Mustaqim, Taufik.*Information Retrieval.* http://www.ittelkom.ac.id/library/index.php, diakses terakhir Januari 2010

[7] Kurniawan, Aan. Riset Dalam bidang Komputer. Juni 2009.

[8] Mashup*(web_application_hybrid*). http://en.wikipedia.org/wiki/Mashup, di akses terakhir Maret 2010

[9] Daconta, c Michael., Leo J Obsrt. Web Sematic.

[10] Muliantara, Agus. Penerapan *Regular expression* dalam melindungi alamat email dari spam robot pada *content wordpress*.

[11] *Portable Document Format*. http://en.wikipedia.org/wiki/PDF, diakses terakhir 01 Juni 2009

[12] *Publish or Perish.* http://harzing.com/resource.htm#/pop.htm, diakses terakhir 3 Januari 2010

[13] *SQL*. http://id.wikipedia.org/wiki/SQL, diakses terakhir 13 Maret 2010

[14] Kurniawan, Agung. Implementasi sistem sitasi jurnal elektronik Indonesia berbasis teknik ekstraksi web. Juni 2009

Lampiran 2. Skrip PHP Aplikasi Program Ekstraksi.

```
<?php
/*
	Script ini digunakan untuk mengekstrak data-data jurnal yang terdapat pada website http://repository.gunadarma.ac.id:8000/view/subjects/
	Tahapan program:
	1. Load fungsi-fungsi yang dibutuhkan
	2. Download halaman utama (http://repository.gunadarma.ac.id:8000/view/subjects/)
	3. Ekstrak semua alamat seri jurnal yang ditemukan pada halaman utama dengan menggunakan Perl Regex
	4. Proses tiap alamat seri yang ditemukan
		a. Download alamat seri
		b. Ekstrak semua alamat jurnal yang ditemukan pada halaman seri
		c. Proses tiap alamat jurnal yang ditemukan dengan menggunakan fungsi saveJurnalRepGunadarma
*/

set_time_limit( 16000 );
$startTime = microtime(true);
//============================== ( 1 ) ======================================
include_once( "../allfunction.php" );
include_once( "../setting.php" );

?>
<head>
<style type="text/css">
	#error { color: red }
	#clear { color: blue }
	#warn { color: green }
</style>
</head>
<body>
<?php
//=============================================================================================
//======================================== ( 2 ) ========================================
$website = 'http://repository.gunadarma.ac.id:8000/view/subjects/';

$isi = file_get_contents( $website );

if ( !$isi )
    die( '<p id="error">Website ' . $website . ' tidak dapat dibuka</p>' );

//=============================================================================================
//================================== ( 3 ) ============================================
preg_match( "/(?<=ep_view_menu\">).+(?=<\/div>\s+<map)/s", $isi, $isiTengah ); //ekstrak kontent bagian tengah halaman

$seriPattern = "/(?<=href=\")[^<>]+(?=\")/";                    // pola regex alamat seri
```

```
preg_match_all( $seriPattern, $isiTengah[ 0 ], $alamatSeriArray, PREG_SET_ORDER );          // ambil alamat untuk
tiap seri menggunakan regex yang disimpan sebagai array pada var $alamatSeriArray

if ( !$alamatSeriArray )
        die( '<br /><p id="error">Alamat seri tidak ditemukan dari ' . $website . '</p>' );
//=============================================================================
//============================== ( 4 ) ================================================
foreach( $alamatSeriArray as $alamatSeri ) {

        $alamatSeri[ 0 ] = $website . $alamatSeri[ 0 ];
        echo( '<br /> <h1 style="color: blue;">seri : ' . $alamatSeri[ 0 ] . '</h1>' );

        //========================================== ( 4a ) ================================
        $isiSeri = file_get_contents( $alamatSeri[ 0 ] );
        //=============================================================================

        if ( !$isiSeri ) {
                echo( '<br /><p id="error">Website ' . $alamatSeri[ 0 ] . ' tidak dapat dibuka</p>');
        }
        else {

                preg_match( "/<\/div>\s+<p>.+<\/div>\s+<map/s", $isiSeri, $isiSeriTengah );          //ekstrak
kontent bagian tengah halaman

                if ( $isiSeriTengah ) {
        //============================================== ( 4b ) ======================
                        $jurnalPattern = "/(?<=href=\")[^<>]+(?=\">)/";
                        preg_match_all(     $jurnalPattern,     $isiSeriTengah[     0     ],     $alamatJurnalArray,
PREG_SET_ORDER );
        //=============================================================================

                        if ( $alamatJurnalArray ) {

        //========================================== ( 4c ) ================================
                        foreach ( $alamatJurnalArray as $alamatJurnal ) {
                        echo '<br /><h2>Alamat Jurnal : ' . $alamatJurnal[ 0 ] . '</h2>';
                        saveJurnalRepGunadarma( $alamatJurnal[ 0 ] );          //Ambil dan simpan data jurnal

        // Bagian untuk membatasi jurnal yang akan diextract
        // ======================================================
                                        if ( $jCnt == $jurnalLimit ) {
                                                printf( "<br />%d Jurnal(s) have been extracted in
<b>%.3f</b> seconds.", $jCnt, microtime( true ) - $startTime );
                                                die( '<br /><p id="error">Jurnal limit reached</p>' );
                                        }
        // ======================================================
                                }
        //======================================================
                        }
```

```
                                else {
                                        echo '<p id="error">Alamat Jurnal tidak ditemukan dari ' . $alamatSeri[ 0 ] .
'</p>';
                                }
                        }
                        else {
                                echo '<p id="error">Isi tidak ditemukan dari ' . $alamatSeri[ 0 ] . '</p>';
                        }
                }
}
//=============================================================================================
/*      Fungsi untuk mengextract data jurnal (judul, penulis, institusi, abstrak, alamat pdf) kemudian menyimpannya
pada database */

function saveJurnalRepGunadarma( $alamatJurnal ) {

        global $jCnt;
        global $jurnalLimit;
        global $website;

        $isiJurnal = file_get_contents( $alamatJurnal );

        if ( !$isiJurnal ) {
                echo '<p id="error">Website ' . $alamatJurnal . ' tidak dapat dibuka</p>';
        }
        else {

                preg_match( "/center\">.+<\/div>\s+<map/s", $isiJurnal, $content );                     //ekstrak
kontent utama halaman jurnal

                $info = '<table>';

                if ( !$content ) {
                        $info .= exError( 'Content', $alamatJurnal );
                }
                else {
                        $isiJurnal = $content[ 0 ];

        //=============================================================================
                        preg_match( "/(?<=pagetitle\">).+(?=<\/h1>)/", $isiJurnal, $judul );            //ekstrak
judul
                        if ( !$judul ) {
                                $info .= exError( 'Judul', $alamatJurnal );
                                $judul[ 0 ] = '';
                        }
                        else {
                                $info .= infoRow( 'Judul', $judul[ 0 ] );
                        }
```

```
		//=========================================================================
		//=========================================================================
				preg_match_all(	"/(?<=name\">)[^<>-]+(?=<\/span>)/",	$isiJurnal,	$penulisArray,
PREG_SET_ORDER );		//ekstrak penulis
				if ( !$penulisArray ) {
					$info .= exError( 'Penulis', $alamatJurnal );
					$penulis[ 0 ] = '';
				}
				else {
					$penulis[ 0 ] = '';

					foreach ( $penulisArray as $penulisArr ) {
						$penulis[ 0 ] = $penulis[ 0 ] . $penulisArr[ 0 ];
					}
					$info .= infoRow( 'Penulis', $penulis[ 0 ] );

				}

		//=========================================================================
		//=========================================================================
				preg_match( "/(?<=<\/em>)[^<>]+(?=<\/p>)/", $isiJurnal, $institusi );
			//ekstrak institusi
				if ( !$institusi ) {
					$info .= exError( 'Institusi', $alamatJurnal );
					$info .= exError( 'Penerbit', $alamatJurnal );
					$institusi[ 0 ] = '';
				}
				else {
					$info .= infoRow( 'Institusi', $institusi[ 0 ] );
					$info .= infoRow( 'Penerbit', $institusi[ 0 ] );
				}

		//=========================================================================
		//=========================================================================
				// preg_match( "/(?<=<h2>)[^<>]+(?=<\/h2>)/i", $isiJurnal, $penerbit );		//ekstrak
penerbit
				// if ( !$penerbit ) {
					// $info .= exError( 'Penerbit', $alamatJurnal );
					// $penerbit[ 0 ] = '';
				// }
				// else {
					// $penerbit[ 0 ] = preg_replace( '/,/', ' UII,', $penerbit[ 0 ], 1 );
					// $info .= infoRow( 'Penerbit', $penerbit[ 0 ] );
				// }

		//=========================================================================
		//=========================================================================
```

```
                                preg_match( "/Abstract.+<\/p><\/div>/s", $isiJurnal, $abstrakContent );          //ekstrak
kontent abstrak
                                if ( !$abstrakContent ) {
                                        $info .= exError( 'Abstrak content', $alamatJurnal );
                                        $abstrak[ 0 ] = '';
                                }
                                else {

                                        $abstrakContent[ 0 ] = preg_replace ( "/&#xD;/", " ", $abstrakContent[ 0 ] );
                                        preg_match( "/(?<=\">).+(?=<\/p>)/s", $abstrakContent[ 0 ], $abstrak );
                        //ekstrak abstrak dari kontent abstrak
                                        if ( $abstrak ) {
                                                $info .= infoRow( 'Abstrak', $abstrak[ 0 ] );
                                        }
                                        else {
                                                $info .= exError( 'Abstrak', $alamatJurnal );
                                                $abstrak[ 0 ] = '';
                                        }
                                }
                //=============================================================================
                //=============================================================================
                                preg_match( "/(?<=href=\")[^<>]+pdf(?=\")/", $isiJurnal, $alamatFull );
                //ekstrak alamat pdf jurnal
                                if ( !$alamatFull ) {
                                        $info .= exError( 'Alamat PDF', $alamatJurnal );
                                        $alamatFull[ 0 ] = '';
                                }
                                else {
                                        //$alamatFull[ 0 ] = $website . $alamatFull[ 0 ];
                                        $info .= infoRow( 'Alamat PDF', $alamatFull[ 0 ] );
                                        //$referensi = getReference( $alamatFull[ 0 ] );
                                }
                //=============================================================================
                        }
                }

                $info .= '</table>';
                echo convertLink( $info );

                insertGeneralData( $judul[ 0 ], $penulis[ 0 ], $institusi[ 0 ], $institusi[ 0 ], $abstrak[ 0 ], $alamatFull[ 0 ] );
//simpan di database

                $jCnt++;

}
?>
```

```
<?php
/**
 *  Getref2.php
 *  File untuk memproses referensi yang dimiliki oleh data.
 */

error_reporting(E_ALL);
set_time_limit(0); // unlimited, hahahah
$startTime = microtime(true);

//ini_set('memory_limit', '800M');

include_once( "../allfunction.php" );
include_once( "../setting.php" );
include_once( "../simple_html_dom.php");

/// sql related
$conn = connectDb();
$sql = "SELECT id,judul,alamat FROM `datajurnal` where alamat like :url";
$sth = $conn->prepare($sql);

/// petra
$petra_url = 'http://puslit2.petra.ac.id/ejournal%';
$sth->bindParam(':url',$petra_url,PDO::PARAM_STR);
$sth->execute();
petra_ref($sth);

/// ugm
$ugm_url = 'http://i-lib.ugm.ac.id/jurnal/download.php?dataId=%';
$sth->bindParam(':url',$ugm_url,PDO::PARAM_STR);
$sth->execute();
ugm_ref($sth);

$sth = null;

/// other
$other_sql = "select * from datajurnal where alamat not like 'http://i-lib.ugm.ac.id/jurnal/download.php?dataId=%' and
alamat not like 'http://puslit2.petra.ac.id/ejournal%'";
$other_sth = $conn->prepare($other_sql);
$other_sth->execute();
general_ref($other_sth);

printf( "<br />%d Jurnal(s) have been extracted in <b>%.3f</b> seconds.", $jCnt, microtime( true ) - $startTime );

function general_ref($sth) {
   while ($row = $sth->fetch(PDO::FETCH_ASSOC)) {
      if(strlen($row['alamat']) < 10) {
         continue;
      }
```

```
        $info = '<table>';
        $info .= infoRow( 'Judul', $row['judul'] );
        $info .= infoRow( 'Alamat PDF', $row['alamat'] );

        $referensi = getReference2( $row['alamat'] );
        $info .= $referensi[ 'status' ];

        //jika referensi berhasil diekstrak
        if ( isset( $referensi[ 0 ] ) ) {
            $update_sql = "UPDATE datajurnal SET referensi =:text where id=:id";
            /// @var $update_sth PDOStatement
            $update_sth = $conn->prepare($update_sql);
            $update_sth->bindParam(':id',   $row['id'],    PDO::PARAM_INT);
            $update_sth->bindParam(':text', $referensi[0], PDO::PARAM_STR);
            $update_sth->execute();
            if($update_sth->rowCount()==0) {
                $info = '<tr><td colspan="2" id="error">Data tidak dapat diupdate</td></tr>';
            }
        }

        $info .= '</table>';
        echo $info;
    }
}

function ugm_ref($sth){
    while ($row = $sth->fetch(PDO::FETCH_ASSOC)) {
        $info = '<table>';
        $info .= infoRow( 'Judul', $row['judul'] );
        $info .= infoRow( 'Alamat PDF', $row['alamat'] );

        $alamat = str_replace(
            'http://i-lib.ugm.ac.id/jurnal/download.php?dataId=',
            'http://i-lib.ugm.ac.id/jurnal/detail.php?dataId=',
            $row['alamat']
            );

        $isi = htmlget($alamat);
        if(!$isi){
            echo "tidak dapat membuka $alamat";
            continue;
        }

        $dom = str_get_html($isi);

        $e = $dom->find('strong',1);
        if( !is_null($e) && $e->plaintext == 'Referensi'){
            $text = "";
            $e = $e->parent()->parent();
```

```
            while($e = $e->next_sibling()){
               $text .= $e->plaintext;
            }
            $text = str_replace(' ', '', $text );

            $info .= infoRow("Referensi", $text);
            $info .= saveRef($row['id'], $text);
         }

         $info .= '</table>';
         echo $info;
      }
   }
   function petra_ref($sth) {
      while ($row = $sth->fetch(PDO::FETCH_ASSOC)) {
         if(strlen($row['alamat']) < 10) {
            continue;
         }
         $info = '<table>';
         $info .= infoRow( 'Judul', $row['judul'] );
         $info .= infoRow( 'Alamat PDF', $row['alamat'] );

         $referensi = getReference2( str_replace('shop', 'viewFile', $row['alamat'] ));
         $info .= $referensi[ 'status' ];

         //jika referensi berhasil diekstrak
         if ( isset( $referensi[ 0 ] ) ) {
            saveRef($row['id'],$referensi[0]);
         }

         $info .= '</table>';
         echo convertLink( $info );
      }
   }
   function saveRef($id,$text){
      global $conn;
      $update_sql = "UPDATE datajurnal SET referensi =:text where id=:id";
      /* @var $update_sth PDOStatement */
      $update_sth = $conn->prepare($update_sql);
      $update_sth->bindParam(':id',   $id,   PDO::PARAM_INT);
      $update_sth->bindParam(':text', $text, PDO::PARAM_STR);
      $update_sth->execute();

      if($update_sth->rowCount()==0) {
         return '<tr><td colspan="2" id="error">Data tidak dapat diupdate</td></tr>';
      }
      return '';
   }
   ?>
```