

Sistem Terintegrasi Portal Web Perguruan Tinggi Agama Islam Negeri Menggunakan Metode Grabbing

Sumarsono, Ahmad Syihab Husin

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi
UIN Sunan Kalijaga

e-mail : sumarsonoyk@yahoo.com

Abstract

Website is one of the tools used State Islamic University to facilitate and share information - their current information. However, with so many websites makes the user (users) difficult to obtain such information at a time. Users are required to open the website one by one which takes quite a long time. This study uses web grabbing technique or what is often referred to as web scrapping / web fetching, the technique of taking content from a server or website and display it in the system. In its implementation, the system uses the PHP cURL library which is used to retrieve the content of the website. Integrated web portal system is able to retrieve contents the latest information of website and be able to display it back on the system without need to enter or change the data. This system is very useful for users who need the latest informations.

Keywords: *Web grabbing, web scrapping, web fetching, Integrated System, cURL*

PENDAHULUAN

Indonesia memiliki banyak sekali perguruan tinggi, mulai dari universitas, institut maupun sekolah tinggi islam.. Salah satu perguruan tinggi di Indonesia adalah Perguruan Tinggi Agama Islam Negeri (PTAIN). PTAIN tersebut memiliki website – website mereka sendiri yang ditujukan untuk menyediakan berbagai informasi seputar kegiatan mereka.

Informasi tersebut bisa berupa berita tentang seminar, wisuda, ataupun kegiatan – kegiatan terkini seputar civitas akademik PTAIN tersebut. Namun, dengan banyaknya website – website tersebut membuat *user* (pengguna) kesulitan untuk mendapatkan informasi – informasi tersebut dalam satu waktu. Untuk mendapatkan informasi dari suatu perguruan tinggi pengguna perlu

mengetahui alamat dari website dari perguruan tinggi yang bersangkutan.

Sistem terintegrasi merupakan suatu sistem informasi yang melibatkan berbagai sumber daya yang memungkinkan suatu organisasi atau institusi dapat mengintegrasikan dan mengkoordinasikan proses bisnis yang dimiliki.

Teknik grabbing merupakan suatu cara untuk memujudkan keinginan mengintegrasikan semua informasi yang ada di website. Grabbing merupakan suatu teknik untuk mengambil teks atau secara umum obyek yang ada di suatu situs lain kemudian diletakkan ke situs kita. Grabbing menggunakan command line cURL yang berfungsi mengambil atau mengirim file melalui sebuah *URL syntax*. Semenjak *curl* menggunakan *libcurl*, *curl*

mendukung internet protokol yang didukung oleh *libcurl*

Dengan adanya suatu teknologi yang memiliki kemampuan mengambil informasi dari halaman web secara real time, sementara disisi lain setiap perguruan tinggi memiliki informasi penting untuk diketahui oleh masyarakat umum. Namun masyarakat mungkin memiliki keterbatasan pengetahuan terhadap alamat web dari seluruh perguruan tinggi islam di Indonesia.

Dengan adanya permasalahan tersebut diatas diperlukan adanya suatu sistem informasi yang mampu mengintegrasikan, mewedahi dan menampung informasi terkini dari website – website PTAIN tersebut serta menampilkannya kembali dalam satu alamat web.

METODE PENELITIAN

Pada metode pengembangan sistem, penulis menggunakan metode pengembangan sistem SDLC (*System Development Life Cycle*) yaitu suatu metode yang digunakan untuk pengembangan perangkat lunak. Adapun langkah – langkah yang dilakukan antara lain :

1. Analisis

Mengenali segenap permasalahan yang muncul pada pengguna dengan mendekomposisi dan merealisasikan *use case* diagram lebih lanjut. Mengenali komponen – komponen sistem atau perangkat lunak, objek – objek, hubungan antar objek, dan sebagainya.

2. Perancangan

Tahap perancangan disini merupakan tahap desain merupakan solusi permasalahan dari tahap analisis. Desain disini *meliputi* desain sistem dan desain perangkat lunak. Desain sistem dan perangkat lunak disini menggunakan *tools* UML yakni *Use Case* Diagram, *Class*

Diagram, *Sequence* Diagram, *Activity* Diagram.

3. Implementasi Sistem

Dalam *tahap* ini, merupakan tahap implementasi sistem dimana penulis mengimplementasikan perancangan sistem ke situasi dunia nyata.

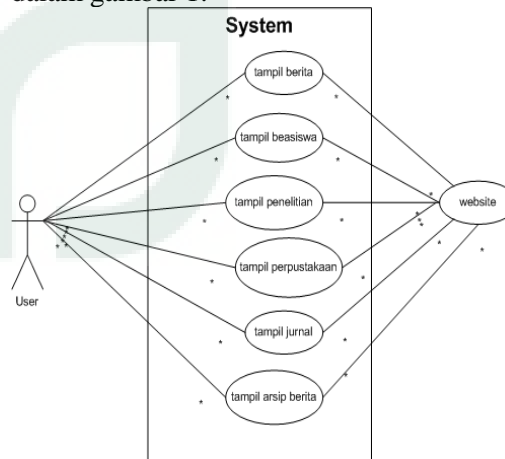
4. Pengujian Sistem

Pengujian sistem dilakukan dengan menggunakan metode *black box*, yaitu melihat hasil pengeluaran apakah sudah berjalan sebagaimana yang diharapkan. Tahap ini bertujuan untuk mengoptimalkan kinerja dari perangkat lunak itu sendiri.

PERANCANGAN

Use Case Diagram

Desain menggunakan use case diagram akan memudahkan dalam menggambarkan atau mengilustrasikan interaksi antara pengguna (*actor*) dan sistem (*use case*). Desain use case diagram ini menjelaskan tentang diagram dari sistem terintegrasi portal web ditampilkan dalam gambar 1.

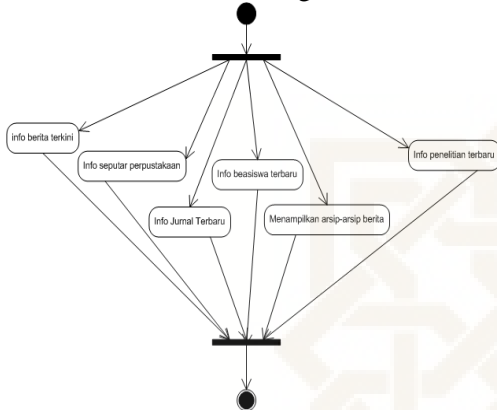


Gambar 1 Use Case Sistem

Activity Diagram

Setelah didapatkan *use case* diagram, langkah selanjutnya adalah membuat *activity* diagram. Perbedaan

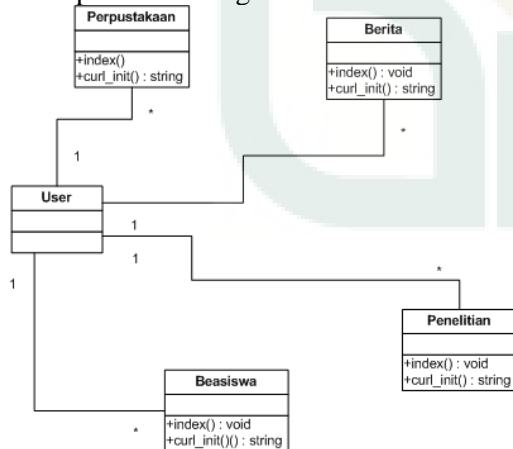
mendasar antar *use case* diagram dengan *activity* diagram adalah *activity* diagram menggambarkan proses yang berjalan, sementara *use case* menggambarkan bagaimana aktor menggunakan sistem untuk melakukan aktifitas. Gambar 2 berikut menjelaskan aktifitas yang dapat diambil dari *use case* diagram:



Gambar 2 Activity Diagram

Class Diagram

Class diagram menggambarkan class – class yang digunakan dalam penelitian ini. Desain class diagram ditampilkan dalam gambar 3.



Gambar 3 Class Diagram

HASIL DAN PEMBAHASAN

Untuk mewujudkan konsep pengembangan sistem secara teknis

dibutuhkan beberapa langkah konfigurasi yaitu konfigurasi cURL di PHP.

Konfigurasi cURL di PHP. Supaya apache web server dapat mendukung cURL, maka diharuskan mengaktifkan cURL pada PHP terlebih dahulu. Pengaktifan cURL ini dilakukan dengan cara menghilangkan tanda *semicolon* (;) *Libcurl* PHP merupakan *library cURL* di PHP. *Library* ini digunakan untuk melakukan grabbing dari suatu website dan mengambil konten – konten dari website tersebut.

Sumber Data dan Informasi Website

Data yang digunakan adalah menggunakan informasi yang ada dalam perguruan tinggi islam di Indonesia dengan 2 kategori yaitu Universitas Islam Negeri (UIN) dan Institut Agama Islam Negeri (IAIN) yaitu :

- UIN Syarif Hidayatullah Jakarta
- UIN Sunan Kalijaga Yogyakarta
- UIN Maulana Malik Ibrahim Malang
- UIN Sultan Syarif Kasim Riau
- UIN Bandung
- UIN Alaudin Makasar
- IAIN Antasari Banjarmasin
- IAIN Ar-Raniry Banda Aceh
- IAIN Raden Fatah Palembang
- IAIN Raden Intan Lampung
- IAIN Jambi
- IAIN Sunan Ampel Surabaya
- IAIN Syekh Nurjati Cirebon
- IAIN Walisongo Semarang

Implementasi Halaman Berita

Halaman berita merupakan halaman yang menampilkan informasi terkini seputar berita – berita dari kampus PTAIN yang diambil dari website PTAIN. Halaman berita menampilkan berita – berita terkini dari berbagai PTAIN yang diambil dari website resmi PTAIN.

Halaman berita dapat diakses dengan mengetikkan *url* = *“localhost/client/”*. Berdasarkan informasi yang ada dapat ditemukan bahwa seluruh obyek penelitian yaitu website perguruan tinggi memiliki informasi tentang “berita”. Adapun implementasi halaman berita ditampilkan pada gambar 4.



Gambar 4 Halaman Berita

Implementasi Halaman Beasiswa

Halaman beasiswa merupakan halaman dari sistem terintegrasi portal web yang menampilkan informasi – informasi terkini seputar beasiswa. Informasi beasiswa ini diambil dari website – website PTAIN yang menyediakan informasi beasiswa di website resminya. Untuk implementasinya, halaman beasiswa sama dengan halaman berita. Yakni, mengambil *url-url web server* dari database dan melakukan *grabbing* dengan fungsi *grab(\$url)*, lalu setelah itu hasil *grabbing* disimpan dalam database dan menampilkannya kembali pada sistem. Halaman beasiswa dapat dilihat dengan mengakses *“localhost/client/index.php/beasiswa”*.

Implementasi Halaman Penelitian

Halaman penelitian merupakan halaman dari sistem yang menampilkan informasi – informasi terkini seputar penelitian yang di-*publish* di website – website resmi PTAIN. Halaman ini dapat

diakses dengan mengetikkan *url localhost/client/index.php/penelitian*.

Halaman penelitian menampilkan informasi penelitian dari 3 PTAIN, yakni dari UIN Syarif Hidayatullah Jakarta, IAIN Raden Fatah Palembang dan IAIN Sunan Ampel Surabaya.



Gambar 5 Halaman Penelitian

Implementasi Halaman Perpustakaan

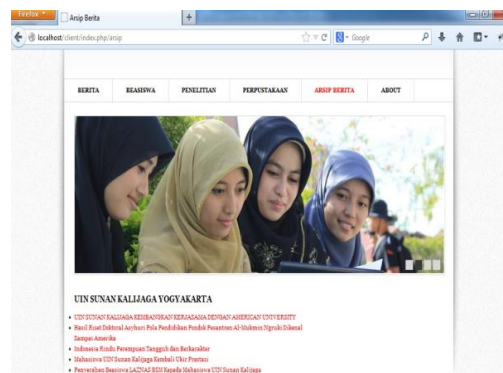
Halaman perpustakaan merupakan halaman yang menampilkan informasi – informasi terkini seputar berita perpustakaan PTAIN. Berita perpustakaan ini diambil dari website – website perpustakaan PTAIN. Halaman perpustakaan ini dapat diakses dengan menuliskan *url “localhost/client/index.php/perpus”* seperti pada gambar 6 dan gambar 7



Gambar 6 Halaman Perpustakaan



Gambar 7 Halaman Perpustakaan



Gambar 7 Halaman Arsip Berita

Implementasi Halaman Arsip Berita

Halaman arsip berita merupakan halaman yang menampilkan arsip – arsip berita dari website – website resmi PTAIN yang berupa *link – link* informasi berita terdahulu. Dalam implementasinya, halamann arsip berita juga sama dengan halaman lainnya. Yaitu, mengambil *url – url web server* dari database lalu melakukan proses *grabbing* dan menyimpannya dalam database. Setelah itu, sistem akan menampilkannya kembali. Implementasi halaman berita ditampilkan pada gambar 7.

Implementasi Halaman Jurnal

Halaman jurnal diimplementasikan dengan mengambil data – data informasi seputar jurnal PTAIN yang terdapat dalam website PTAIN. Navigasi jurnal ditampilkan dalam *sidebar* pada halaman berita, beasiswa, penelitian, arsip berita dan perpustakaan yang nantinya akan di-*redirect* ke halaman jurnal. Namun, tidak semua PTAIN menyediakan seputar jurnalnya di website mereka. Untuk mengakses halaman jurnal bisa menuliskan '*localhost/client/index.php/jurnal*' pada *toolbar web browser* kita.

Pengujian Sistem

Pengujian pada sistem terintegrasi portal web bertujuan untuk menunjukkan bahwa sistem terintegrasi ini telah bekerja sesuai dengan spesifikasi dan kebutuhan yang akan dicapai.

Pengujian sistem terintegrasi ini dilakukan dengan menggunakan *system test*, yaitu menguji kemampuan keseluruhan yang dilakukan aplikasi dengan melihat integrasi dari semua paket. Pengujian *system test* ini menggunakan metode *black box*, yaitu melihat hasil pengeluaran apakah telah berjalan sebagaimana yang diharapkan.

KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, maka didapatkan beberapa kesimpulan sebagai berikut ;

1. Sistem terintegrasi portal web Perguruan Tinggi Agama Islam Negeri berhasil dirancang dan dibangun dengan menggunakan bahasa pemrograman PHP dengan menggunakan PHP *framework* CodeIgniter dan library *cURL PHP*;
2. Sistem terintegrasi portal web Perguruan Tinggi Agama Islam Negeri berhasil mengambil konten – konten dari website resmi Perguruan Tinggi Agama Islam Negeri menggunakan teknik *grabbing*.

3. Seorang user mampu melihat seluruh informasi yang terkait dengan berita, perpustakaan dan beasiswa dari perguruan tinggi melalui sebuah website terintegrasi ini.

Dalam implementasinya sistem belum memiliki *exception handling* yang baik dalam menangani *error*. Masih ada beberapa konten yang ditampilkan di sistem berupa *error*. Hal ini disebabkan website perguruan tinggi yang menjadi acuan data dan informasi mengalami *error* atau tidak online.

DAFTAR PUSTAKA

- [1] Alfian, G, 2009, 'Rancang Bangun Sistem Rekomendasi Buku Menggunakan Collaborative Filtering', S1, Teknik Informatika - UIN Sunan Kalijaga, Yogyakarta.
- [2] Anhar, S, 2010, *Panduan Menguasai PHP & MySQL Secara Otodidak*, 1st edn, Mediakita, Jakarta Selatan.
- [3] Budhi, Setiabudi & Raharjo, 2006, 'Aplikasi Web Grabber untuk mengambil Halaman Web Sesuai dengan Keyword yang Diinputkan', *Jurnal Informatika*, vol 7 / No. 1, no. Web Grabbing.
- [4] Dhenakaran, SS & Sambanthan, KT, 2011, 'Web Crawler - An Overview', *International Journal of Computer Science and Communication*, Vol. 2, no. 1, pp. 265 - 267.
- [5] Fahrurrozi, I, 2010, 'Sistem Rekomendasi pada Toko Buku Online Menggunakan Item Collaborative Filtering Berbasis Web dan WAP', S1, Teknik Informatika - UIN Sunan Kalijaga, Yogyakarta.
- [6] Fata, HA, 2009, *Rekayasa Sistem Pengenalan Wajah Membangun Sistem Presensi Karyawan Menggunakan Microsoft Visual Basic 6.0 dan Microsoft Access*, 1st edn, C.V Andi Offset, Yogyakarta.
- [7] Khurana, D & Kumar, S, 2012, 'An Improved Approach for Caption Based Image Web Crawler', *International Journal of Computer Science & Management Studies*, Volume 12, no. 02, pp. 2231-5268.
- [8] Mulyanto, A 2009, *Sistem Informasi : Konsep dan Aplikasi*, Pustaka Pelajar, Yogyakarta.
- [9] Marimin, Tanjung, H & Prabowo, H, 2006, *Sistem Informasi Manajemen Sumber Daya Manusia*, Grasindo, Jakarta.
- [10] Mustakini, JH 2005, *Sistem Teknologi Informasi : Pendekatan Terintegrasi*, 2nd edn, Andi Publisher, Yogyakarta.
- [11] Missier, P, Merialdo, P, Crescenzi, V & Mecca, G, 2004, 'An Automatic Data Grabber for Large Web Sites', *30th International Conference on Very Large Data Bases (VLDB)*, Morgan Kaufmann, Toronto, Canada.
- [12] Nugroho, A, 2009, *Rekayasa Perangkat Lunak Menggunakan UML dan Java*, 1st edn, Andi Offset, Yogyakarta.
- [13] Pani, SK, Mohapatra, D & Ratha, BK, 2010, 'Integration of Web Mining and Web Crawler: Relevance and State of Art', *International Journal on Computer Science and Engineering*, Vol. 02, no. 03, pp. 772-776.
- [14] Pant, G, Srinivasan, P & Menczer, F, 2004, 'Crawling the Web', in *Web Dynamics*.
- [15] Perdananugraha, GM, 2010, 'Rancang Bangun dan Implementasi Sistem Informasi Terintegrasi (Integrated Information System) e-Government Berbasis Open source', *Jurnal Teknologi Indonesia*, vol 33, NO. 1, no. Integrated information system, e-government, open source, joomla, internet, pp. 45 - 59.
- [16] Pratama, ANW, 2010, *CodeIgniter: Cara Mudah Membangun Aplikasi PHP*, 1st edn, Mediakita, Jakarta Selatan.
- [17] PHP, 2013. *PHP: explode - Manual*. [Online] Available at <http://php.net/manual/en/function.explode.php> [Accessed 18 May 2013].
- [18] Retyandias, A 2010, 'Sistem Informasi Terintegrasi pada PT. Graha Sarana Gresik', S1 Skripsi, Teknik Informatika - Fakultas Teknologi Industri UPN Veteran Jawa Timur, Jawa Timur.
- [19] Stenberg, D., 2011. *FAQ -- Frequently Asked Questions : What is curl*. [Online] Available at: http://curl.haxx.se/docs/faq.html#What_is_cURL [Accessed 1 May 2013].
- [20] Stenberg, D., 2011. *FAQ -- Frequently Asked Questions : What is libcurl*. [Online] Available at: http://curl.haxx.se/docs/faq.html#What_is_libcurl [Accessed 1 May 2013].
- [21] Schrenk, M, 2012, *Webbots, Spiders and Screen Scrappers A Guide To Developing Internet Agents with PHP / CURL*, 2nd edn, No Starch Press, San Fransisco.
- [22] Simarmata, J, 2010, *Rekayasa Web*, 1st edn, Andi Offset, Yogyakarta.
- [23] Soulemane, M, Rafiuzzaman, M & Mahmud, H, 2012, 'Crawling the Hidden Web: An Approach to Dynamic Web Indexing', *International Journal of Computer Applications (0975 - 8887)*, Vol. 5, no. 01.
- [24] Yuhefizar, SK, Mooduto, IH & Rahmat Hidayat, S, 2006, *Cara Mudah Membangun Website Interaktif Menggunakan Content Management System Joomla*, Revisi edn, PT Elex Media Komputindo, Jakarta.
- [25] Yakub 2012, *Pengantar Sistem Informasi*, Graha Ilmu, Yogyakarta.
- [26] www.php.net, 2013. *PHP: cURL- Manual*. [Online] Available at: <http://php.net/manual/en/book.curl.php> [Accessed 14 May 2013].