

SKRIPSI

**IT STUDENT PORTAL: PEMANFAATAN *WEB SCRAPING*
UNTUK KUSTOMISASI STUDENT PORTAL UNPAR**



HERFAN HERYANDI

NPM: 2012730012

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2015**

UNDERGRADUATE THESIS

**IT STUDENT PORTAL: EXPLOITING WEB SCRAPING FOR
STUDENT PORTAL UNPAR CUSTOMIZATION**



HERFAN HERYANDI

NPM: 2012730012

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2015**

LEMBAR PENGESAHAN

**IT STUDENT PORTAL: PEMANFAATAN *WEB SCRAPING*
UNTUK KUSTOMISASI STUDENT PORTAL UNPAR**

HERFAN HERYANDI

NPM: 2012730012

Bandung, «tanggal» «bulan» 2015

Menyetujui,

Pembimbing Utama

Pembimbing Pendamping

«pembimbing utama/1»

«pembimbing pendamping/2»

Ketua Tim Penguji

Anggota Tim Penguji

«penguji 1»

«penguji 2»

Mengetahui,

Ketua Program Studi

Thomas Anung Basuki, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

IT STUDENT PORTAL: PEMANFAATAN *WEB SCRAPING* UNTUK KUSTOMISASI STUDENT PORTAL UNPAR

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal «tanggal» «bulan» 2015

Meterai

Herfan Heryandi
NPM: 2012730012

ABSTRAK

«Tuliskan abstrak anda di sini, dalam bahasa Indonesia» Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Kata-kata kunci: «Tuliskan di sini kata-kata kunci yang anda gunakan, dalam bahasa Indonesia»

ABSTRACT

«Tuliskan abstrak anda di sini, dalam bahasa Inggris» Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

Keywords: «Tuliskan di sini kata-kata kunci yang anda gunakan, dalam bahasa Inggris»

«kepada siapa anda mempersembahkan skripsi ini...?»

KATA PENGANTAR

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Bandung, «bulan» 2015

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xviii
DAFTAR TABEL	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Metode Penelitian	3
1.6 Sistematika Penulisan	3
2 DASAR TEORI	5
2.1 jsoup	5
2.1.1 Jsoup	5
2.1.2 Connection	5
2.1.3 Response	6
2.1.4 Document	7
2.1.5 Elements	7
2.1.6 Element	7
2.2 Chrome DevTools	8
2.2.1 Elements	9
2.2.2 Network	9
2.3 Play Framework	9
3 ANALISIS	11
3.1 Analisis Fitur-fitur FTIS Student Portal	11
DAFTAR REFERENSI	13

DAFTAR GAMBAR

1.1	Prasyarat Mata Kuliah[1]	2
1.2	Prasyarat Mata Kuliah Student Portal UNPAR[2]	2
2.1	Chrome DevTools	8

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Student Portal UNPAR[2] merupakan sistem informasi berbasis *web* yang digunakan oleh mahasiswa Universitas Katolik Parahyangan. Fitur-fitur yang dimiliki Student Portal UNPAR yaitu rencana studi, jadwal, nilai dan indeks prestasi, dan pembayaran uang kuliah. Namun, fitur-fitur tersebut masih belum cukup untuk mendukung kebutuhan akademik mahasiswa Program Studi Teknik Informatika.

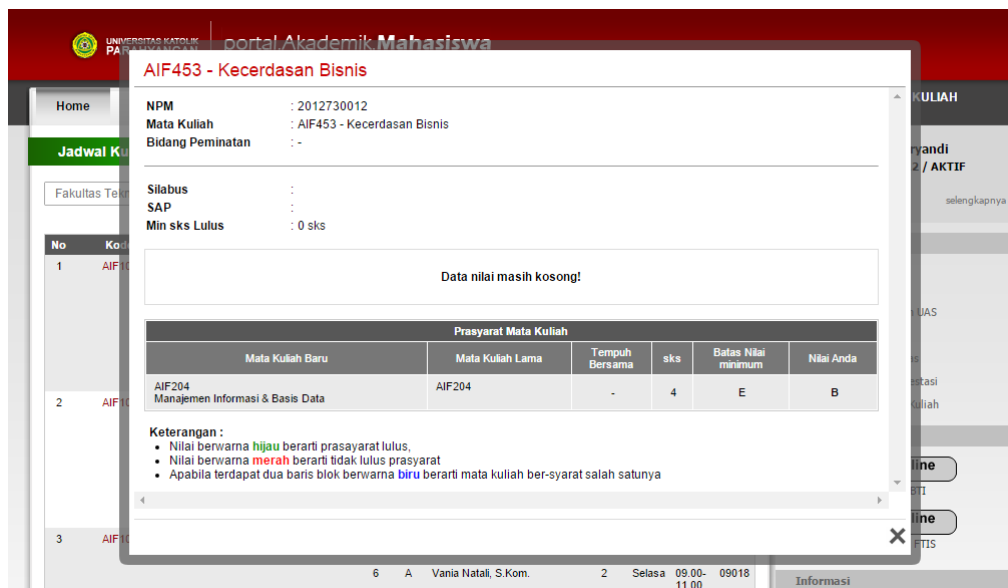
Salah satu fitur yang diperlukan oleh mahasiswa Teknik Informatika UNPAR adalah prasyarat mata kuliah. Dalam Teknik Informatika UNPAR, terdapat beberapa mata kuliah yang membutuhkan prasyarat baik prasyarat tempuh maupun prasyarat lulus. Student Portal UNPAR sudah menyediakan fitur prasyarat mata kuliah namun kurang mendukung karena data yang ditampilkan kurang akurat. Misalnya dalam pengambilan mata kuliah "AIF453 Kecerdasan Bisnis" membutuhkan prasyarat lulus mata kuliah "AIF204 Manajemen Informasi dan Basis Data" atau lulus mata kuliah "AIF102 Algoritma dan Struktur Data" dengan IPK di atas 2.75 (Gambar 1.1). Namun dalam Student Portal UNPAR, prasyarat yang dicantumkan hanya lulus mata kuliah "AIF204 Manajemen Informasi dan Basis Data" (Gambar 1.2). Selain itu, pemeriksaan prasyarat mata kuliah tidak dilakukan secara otomatis sehingga setiap pengambilan mata kuliah tetap dianggap valid meskipun belum memenuhi prasyarat.

jsoup[3] merupakan *library* Java yang digunakan untuk menelusuri suatu situs *web* untuk mendapatkan suatu informasi. Informasi yang didapat berupa HTML yang kemudian diekstrak dan disajikan dalam bentuk *Document Object Model*. Play Framework[4] merupakan sebuah *web framework* berbasis Java dan Scala. Play juga menggunakan *design pattern Model-View-Controller* (MVC) di mana *model* dan *controller* menggunakan bahasa Java sedangkan *view* menggunakan bahasa Scala dan HTML. SIA Models[5] merupakan kelas-kelas dalam bahasa Java yang merepresentasikan Sistem Informasi Akademik UNPAR. Aplikasi akan dibuat dengan menggunakan Play Framework dan jsoup karena aplikasi didukung oleh SIA Models yang tersedia dalam bahasa Java.

Prasyarat Mata Kuliah

4	AIF203	Struktur Diskret	1. T : AIF103 Matematika Diskret
5	AIF205	Arsitektur & Organisasi Komputer	1. T : AIF106 Sistem Digital
6	AIF301	Pengantar Sistem Cerdas	1. T : AIF104 Logika Informatika 2. T : AIF202/292 Desain dan Analisis Algoritma
4	AIF302	Penulisan Ilmiah	1. Lulus minimal 84 SKS
5	AIF303	Pengantar Sistem Informasi	1. T : AIF204/AIF294 Manaj. Informasi & Basis Data
6	AIF305	Jaringan Komputer	1. T : AIF206 Sistem Operasi
	AIF403	Komputer dan Masyarakat	1. Sudah lulus minimal 72 SKS
7	AIF405	Proyek Sistem Informasi 2	1. T: AIF304 Proyek Sistem Informasi 1
8	AIF401	Skripsi 1	1. L : AIF302 Penulisan Ilmiah 2. Lulus minimal 108 SKS
9	AIF402	Skripsi 2	1. L : AIF401 Skripsi 1 ATAU 2. Ambil AIF401 Skripsi 1 & lulus min. 124 SKS
13	APS402	Etika Profesi	1. Sudah lulus minimal 90 SKS
Kuliah Pilihan Wajib			
1	AIF311	Pemrograman Fungsional	1. T : AIF103 Matematika Diskret
2	AIF314	Pemrograman Basisdata	1. T : AIF204/AIF294 Manajemen Informasi & Basis Data
3	AIF315	Pemrograman Berbasis Web	2. AIF204/AIF294 Manaj. Informasi & Basis Data
4	AIF317	Desain Antarmuka Grafis	1. T : AIF210 Interaksi Manusia Komputer
Kuliah Pilihan			
1	AIF438	Penambangan Data	1. T : AIF102/AIF192 Algoritma & Struktur Data 1. T : AIF204/AIF 294 Manaj. Inf. & Basis Data ATAU 2. T : AIF102/AIF 192 Algo. & Struk. Data & IPK >2.75
2	AIF453	Kecerdasan Bisnis	
3	AIF457	Kewirausahaan Berbasis Teknologi	1. Lulus minimal 70 SKS
4	AIF469	Layanan Berbasis Web	1. T : AIF305 Jaringan Komputer 2. T : AIF315 Pemrograman Berbasis Web
5	AIF362	Sistem & Aplikasi Telematika	1. T : AIF305 Jaringan Komputer

Gambar 1.1: Prasyarat Mata Kuliah[1]



Gambar 1.2: Prasyarat Mata Kuliah Student Portal UNPAR[2]

Untuk mendukung kebutuhan akademik mahasiswa Program Studi Teknik Informatika, fitur-fitur yang diperlukan akan dianalisa kemudian diimplementasikan ke dalam program IT Student Portal. Program yang akan dibuat merupakan program berbasis *web* menggunakan Play Framework. Selain itu, data-data yang akan ditampilkan diambil langsung dari Student Portal UNPAR dengan *web scraping* menggunakan *library* jsoup. Untuk melakukan pengambilan data, jsoup harus mengetahui cara kerja dari Student Portal UNPAR. Analisis komunikasi Student Portal UNPAR akan dilakukan dengan menggunakan Chrome DevTools.

1.2 Rumusan Masalah

Rumusan dari masalah yang akan dibahas pada skripsi ini sebagai berikut:

1. Fitur-fitur apa saja yang akan dibuat untuk IT Student Portal?
2. Bagaimana mengimplementasikan *web scraping* menggunakan *library* jsoup?
3. Bagaimana membangun aplikasi IT Student Portal?

1.3 Tujuan

Tujuan-tujuan yang hendak dicapai melalui penulisan skripsi ini sebagai berikut:

1. Mengetahui fitur-fitur yang akan dibuat dalam IT Student Portal.
2. Mengimplementasikan *web scraping* menggunakan *library* jsoup.
3. Membangun aplikasi IT Student Portal.

1.4 Batasan Masalah

Beberapa batasan yang dibuat terkait dengan pengerjaan skripsi ini sebagai berikut:

1. Prasyarat mata kuliah yang tersedia hanya mata kuliah yang didukung SIA Models.
2. Aplikasi akan diuji pada server FTIS sehingga tidak bisa diakses dari luar jaringan FTIS.

1.5 Metode Penelitian

Metode-metode yang dilakukan pada penelitian ini sebagai berikut:

1. Melakukan studi mengenai *library* jsoup, Chrome DevTools, dan Play Framework.
2. Melakukan wawancara.
3. Menganalisis Student Portal UNPAR.
4. Mengimplementasikan *web scraping* menggunakan *library* jsoup.
5. Melakukan eksperimen dan pengujian.

1.6 Sistematika Penulisan

Sistematika penulisan setiap bab pada penelitian ini sebagai berikut:

1. Bab Pendahuluan
Bab 1 berisikan latar belakang, rumusan masalah, tujuan, metode penelitian, dan sistematika penulisan dari penelitian yang dilakukan.

2. Bab Dasar Teori

Bab 2 berisikan teori-teori yang menunjang penelitian yang dilakukan. Teori yang digunakan dalam penelitian ini, antara lain *Library jsoup*, *Chrome DevTools*, dan *Play Framework*.

3. Bab Analisis

Bab 3 berisikan analisis yang dilakukan pada penelitian ini. Analisis yang dilakukan meliputi: Analisis Fitur-fitur FTIS Student Portal, Analisis *Web Scraping*, dan Analisis dari Aplikasi yang Akan Dibuat.

BAB 2

DASAR TEORI

2.1 jsoup

Web scraping[6] adalah teknik mendapatkan informasi dari sebuah situs *web* secara otomatis. Dalam Java, *web scraping* dapat diimplementasikan menggunakan *library* jsoup[3]. API yang disediakan oleh jsoup dapat digunakan untuk mengekstrak dan memanipulasi data HTML.

Subbab-subbab berikut menjelaskan kelas-kelas dari jsoup yang terkait dengan pengerjaan skripsi ini.

2.1.1 Jsoup

Kelas ini merupakan inti untuk mengakses fungsi jsoup. Salah satu *method* yang dimiliki kelas ini adalah sebagai berikut:

- **public static Connection connect(String url)**
Befungsi untuk membuat koneksi baru dengan suatu situs *web*.

Parameter:

- **url** URL situs *web* dengan protokol HTTP atau HTTPS.

Kembalian: koneksi dengan situs *web*.

2.1.2 Connection

Kelas ini merupakan interface yang menyediakan pengambilan data dari situs *web*. Beberapa *method* yang dimiliki kelas ini adalah sebagai berikut:

- **Connection cookies(Map<String,String> cookies)**
Befungsi untuk menambahkan *cookie*.

Parameter:

- **cookies map** dari *cookie*.

Kembalian: koneksi dengan situs *web*.

- **Connection data(String key, String value)**
Befungsi untuk menambahkan parameter data.

Parameter:

- **key** kunci data.
- **value** nilai data.

Kembalian: koneksi dengan situs *web*.

- **Connection method(Connection.Method method)**

Berfungsi untuk mengatur metode pengiriman.

Parameter:

- **method** metode pengiriman HTTP *request*.

Kembalian: koneksi dengan situs *web*.

- **Connection timeout(int millis)**

Berfungsi untuk mengatur batas waktu *request*. Batas waktu nol akan dianggap sebagai batas waktu yang tak terhingga.

Parameter:

- **millis** banyaknya milisekon sebelum batas waktu.

Kembalian: koneksi dengan situs *web*.

- **Connection validateTLSCertificates(boolean value)**

Berfungsi untuk mengatur pemeriksaan sertifikat TLS untuk HTTPS *request*. Nilai "true" untuk memeriksa dan nilai "false" untuk tidak memeriksa.

Parameter:

- **value** status pemeriksaan sertifikat TLS.

Kembalian: koneksi dengan situs *web*.

- **Connection.Response execute()**

Berfungsi untuk mengirim *request*.

Kembalian: objek Response.

2.1.3 Response

Kelas ini merepresentasikan HTTP *response*. Response merupakan turunan dari kelas Base yang menangani response dan *request*. Beberapa *method* yang dimiliki kelas ini adalah sebagai berikut:

- **Map<String,String> cookies()**

Method ini diturunkan dari kelas Base, berfungsi untuk mendapatkan seluruh *cookies*.

Kembalian: seluruh *cookies*.

- **Document parse()**

Berfungsi untuk *parsing response body* menjadi dokumen.

Kembalian: koneksi dengan situs *web*.

- **String body()**

Berfungsi untuk mendapatkan *response body* berupa *string*.

Kembalian: *response body* dalam bentuk *string*.

2.1.4 Document

Kelas ini merepresentasikan dokumen HTML. Salah satu *method* yang dimiliki kelas ini adalah sebagai berikut:

- **public Elements select(String cssQuery)**

Method ini diturunkan dari kelas Element, berfungsi untuk menemukan elemen HTML yang sesuai dengan kueri CSS.

Parameter:

- **cssQuery** kueri CSS.

Kembalian: elemen-elemen HTML yang sesuai dengan kueri CSS.

2.1.5 Elements

Kelas ini merepresentasikan kumpulan elemen HTML. Beberapa *method* yang dimiliki kelas ini adalah sebagai berikut:

- **public Elements select(String query)**

Method Berfungsi untuk menemukan elemen-elemen yang sesuai dalam *list* elemen.

Parameter:

- **query** kueri *Selector*.

Kembalian: elemen-elemen yang sudah diseleksi sesuai kueri.

- **public String val()**

Method Berfungsi untuk mendapatkan nilai dari elemen pertama.

Kembalian: nilai elemen.

- **public String text()**

Method Berfungsi untuk mendapatkan kombinasi teks dari seluruh elemen yang sesuai.

Kembalian: seluruh teks dalam *string*.

2.1.6 Element

Kelas ini merepresentasikan sebuah elemen HTML yang berisikan *tag*, atribut, dan *child*. Beberapa *method* yang dimiliki kelas ini adalah sebagai berikut:

- **public Element child(int index)**

Method Berfungsi untuk mendapatkan *child element* berdasarkan nomor indeks.

Parameter:

- **index** nomor index.

Kembalian: *child element*.

- **public Element child()**

Method Berfungsi untuk mendapatkan seluruh *child element*.

Kembalian: seluruh *child element*.

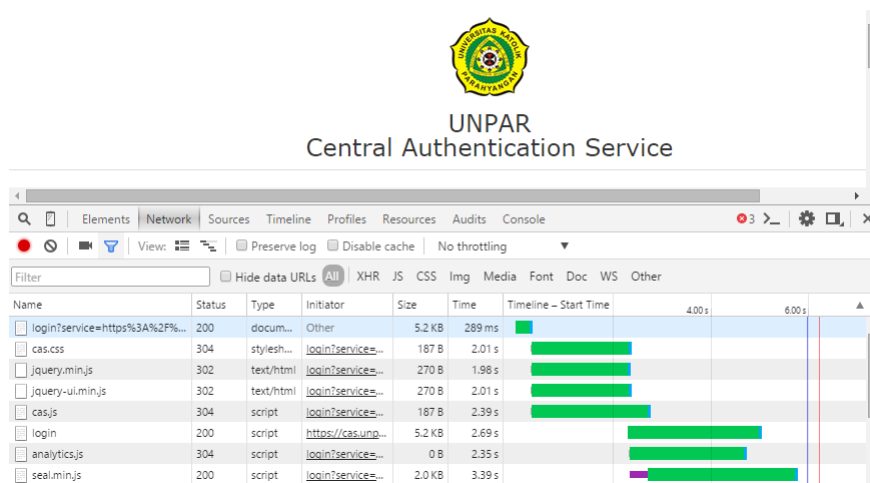
- **public String className()**
Method Berfungsi untuk mendapatkan nama kelas elemen.
Kembalian: nama kelas elemen.
- **public String text()**
Method Berfungsi untuk mendapatkan teks dari elemen.
Kembalian: teks dalam *string*.

2.2 Chrome DevTools

Chrome Developer Tools (DevTools) [7] adalah perangkat *debugging* yang dimiliki Google Chrome. Saat mengunjungi suatu halaman web, pengguna DevTools dapat melakukan *debugging* pada halaman tersebut. DevTools dapat diakses dengan menekan "Ctrl+Shift+I" saat sedang membuka suatu halaman web.

Fitur-fitur yang dimiliki DevTools antara lain:

1. **Elements**, memeriksa dan mengubah elemen HTML dan *style* dari suatu *website*.
2. **Console**, mendapatkan informasi pengembangan dan berinteraksi dengan dokumen.
3. **Sources**, melakukan *debugging* pada JavaScript dengan menentukan *breakpoint*.
4. **Network**, memantau kinerja jaringan pada *website* secara *real-time*.
5. **Audits**, menganalisa halaman yang dimuat.
6. **Timeline**, menampilkan alur waktu saat memuat halaman.
7. **Profiles**, menggambarkan waktu eksekusi dan penggunaan memori saat memuat halaman.
8. **Resources**, memeriksa sumber daya halaman yang dapat berupa basis data, *cookies*, dan *cache*.



Gambar 2.1: Chrome DevTools

Pada subbab-subbab berikut akan dijelaskan mengenai dua panel dari DevTools.

2.2.1 Elements

Panel Elements memungkinkan untuk memperlihatkan informasi yang terstruktur tentang halaman yang sedang dibuka. HTML akan ditampilkan dalam bentuk pohon *Document Object Model* (DOM). Tampilan pohon DOM memperlihatkan struktur DOM dari halaman yang sedang dibuka. Pohon DOM adalah pohon dari node-node yang mewakili setiap elemen HTML seperti `<body>` dan `<p>`. Tampilan struktur pohon DOM menampilkan tag elemen HTML bukan jenis node DOM: misalnya, `<p>` bukan `HTMLParagraphElement`.

2.2.2 Network

2.3 Play Framework

Play Framework [4] merupakan sebuah web framework berbasis Java dan Scala. Play juga menggunakan *design pattern* Model-View-Controller (MVC) di mana *model* dan *controller* menggunakan Java sedangkan *view* menggunakan Scala dan HTML.

BAB 3

ANALISIS

3.1 Analisis Fitur-fitur FTIS Student Portal

Berdasarkan penelitian, fitur-fitur yang diperlukan IT Student Portal adalah sebagai berikut:

1. **Prasayarat Mata Kuliah**

Masalah lain yang sering dialami FTIS adalah kesalahan dalam mengisi Formulir Rencana Studi(FRS). Meskipun sudah diberi daftar mata kuliah beserta prsyaratnya, mahasiswa sering kali lalai dalam memeriksa prasyarat tersebut. Selain itu, tidak semua dosen wali tahu prasyarat mata kuliah yang akan diambil sehingga dosen wali menyetujui mata kuliah yang masih belum boleh diambil. Akibatnya, banyak terjadi kesalahan dalam pengambilan mata kuliah. Kesalahan tersebut akan diperiksa oleh sekretaris jurusan. Artinya, sekretaris jurusan harus memeriksa kartu rencana studi setiap mahasiswa kemudian memeriksa prasyarat dari setiap mata kuliah yang diambil.

Oleh karena itu, fitur prasyarat mata kuliah akan dibuat dalam IT Student Portal. Prasyarat mata kuliah dapat berupa mata kuliah dan sks lulus. Jika pemeriksaan prasyarat mata kuliah dapat dilakukan oleh sistem, maka kesalahan pengambilan mata kuliah semakin berkurang. Selain itu, dosen wali dan sekretaris jurusan tidak perlu melakukan pemeriksaan lebih lanjut.

DAFTAR REFERENSI

- [1] Program Studi Teknik Informatika, “PRASAYARAT MATA KULIAH SEMESTER GANJIL 2015/2016.” <https://tinyurl.com/lionov>, Juli 2015. [Online; diakses 1-Juli-2015].
- [2] Biro Teknologi Informasi UNPAR, “Student Portal UNPAR.” <https://studentportal.unpar.ac.id>, 2012. [Online; diakses 1-Juli-2015].
- [3] J. Hedley, “jsoup: Java HTML Parser.” <http://jsoup.org>, 2009-2015. [Online; diakses 1-Juli-2015].
- [4] N. Leroux and S. de Kaper, *Play for Java*. Manning Publications Co., 2014.
- [5] P. Alfadian, “SIA Models.” <https://github.com/pascalalfadian/SIAModels>, 2015. [Online; diakses 1-Juli-2015].
- [6] E. Vargiu and M. Urru, “Exploiting web scraping in a collaborative filtering-based approach to web advertising,” in *Artificial Intelligence Research*, (Barcelona,Spain), pp. 44–50, 2013.
- [7] Google, “Chrome DevTools.” <https://developer.chrome.com/devtools>, 2013. [Online; diakses 1-Juli-2015].