

KARYA ILMIAH

**FTIS 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**

DAFTAR ISI

DAFTAR ISI	iii
DAFTAR GAMBAR	iv
DAFTAR TABEL	v
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	1
1.4 Batasan Masalah	2
1.5 Metode Penelitian	2
1.6 Sistematika Penulisan	2
2 DASAR TEORI	3
2.1 <i>Web Scraping</i>	3
2.2 <i>CSS Selector</i>	3
2.3 <i>Document Object Model</i>	6
2.4 Chrome DevTools	6
2.5 jsoup	10
3 ANALISIS	11
3.1 Analisis Fitur-fitur FTIS Student Portal	11
3.2 Analisis <i>Web Scraping</i>	12
3.3 Analisis Aplikasi yang Akan Dibuat	12
DAFTAR REFERENSI	13

DAFTAR GAMBAR

2.1	Elements Panel [1]	6
2.2	Console Panel [1]	7
2.3	Sources Panel [1]	7
2.4	Network Panel [1]	8
2.5	Audits Panel [1]	8
2.6	Timeline Panel [1]	9
2.7	Profiles Panel [1]	9
2.8	Resources Panel [1]	10

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Student Portal UNPAR merupakan sistem informasi akademik (SIA) berbasis *web* yang digunakan oleh 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 sistem informasi akademik Fakultas Teknologi Informasi dan Sains (FTIS). FTIS membutuhkan fitur-fitur tambahan untuk mendukung kebutuhan sistem informasi akademik.

Fitur-fitur yang diperlukan FTIS akan ditambahkan dengan melakukan kustomisasi Student Portal UNPAR menjadi FTIS Student Portal. Kustomisasi ini dilakukan karena kebutuhan sistem informasi akademik setiap fakultas berbeda-beda. Jika seluruh kustomisasi fakultas diimplementasikan pada Student Portal UNPAR, sistem informasi akademik yang disediakan akan menjadi rumit sehingga pengguna akan mengalami kesulitan dalam menggunakannya. Metode yang akan digunakan dalam kustomisasi Student Portal UNPAR adalah *web scraping*. Dengan menggunakan *web scraping*, fitur-fitur baru dapat ditambahkan tanpa merubah fitur-fitur yang sudah dimiliki Student Portal UNPAR.

1.2 Rumusan Masalah

Rumusan dari masalah yang akan anda bahas pada skripsi ini sebagai berikut.

- Fitur-fitur apa saja yang perlu dikustomisasi FTIS Student Portal?
- Bagaimana cara kerja *Web Scraping*?
- Bagaimana melakukan kustomisasi Student Portal UNPAR dengan *Web Scraping*?

1.3 Tujuan

Tujuan-tujuan yang hendak dicapai melalui penulisan karya ilmiah ini sebagai berikut.

- Mengetahui fitur-fitur yang diperlukan dalam FTIS Student Portal.
- Mengetahui informasi seputar *Web Scraping*.
- Mengetahui upaya kustomisasi Student Portal UNPAR dengan *Web Scraping*.

- Mengetahui analisis perancangan aplikasi FTIS Student Portal.

1.4 Batasan Masalah

Berikut ini adalah beberapa batasan yang dibuat terkait dengan pengerjaan penelitian ini.

- Aplikasi yang dibuat merupakan aplikasi berbasis *web*.
- Penelitian dilakukan dalam Sistem Operasi Windows.
- *Developer tools* yang digunakan adalah *developer tools* Google Chrome yaitu Chrome DevTools.

1.5 Metode Penelitian

Berikut ini adalah metode-metode yang dilakukan pada penelitian ini:

1. Melakukan studi pustaka mengenai *Web Scraping*.
2. Melakukan studi pustaka mengenai *CSS Selector*.
3. Melakukan studi pustaka mengenai *Document Object Model*.
4. Melakukan studi mengenai cara kerja Chrome DevTools.
5. Melakukan studi mengenai jsoup.
6. Mengamati kinerja jaringan Student Portal UNPAR.
7. Melakukan eksperimen *Web Scraping*.

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 *Web Scraping*, *CSS Selector*, *Document Object Model*, Chrome DevTools, jsoup.
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 *Web Scraping*

Web Scraping [2] adalah teknik yang digunakan untuk mendapatkan informasi dari sebuah *website* secara otomatis bukan menyalin secara manual. Tujuan dari *Web Scraping* adalah untuk mencari jenis informasi tertentu, kemudian mengekstraknya untuk menjadi *website* baru. *Web Scraping* berfokus pada transformasi data tidak terstruktur dalam *website*, biasanya format HTML. Kemudian data tersebut diubah menjadi data terstruktur yang dapat disimpan dan dianalisis. Saat ini *Web Scraping* banyak digunakan untuk perbandingan harga, pemantauan cuaca, deteksi perubahan *website*, dan penelitian *website*.

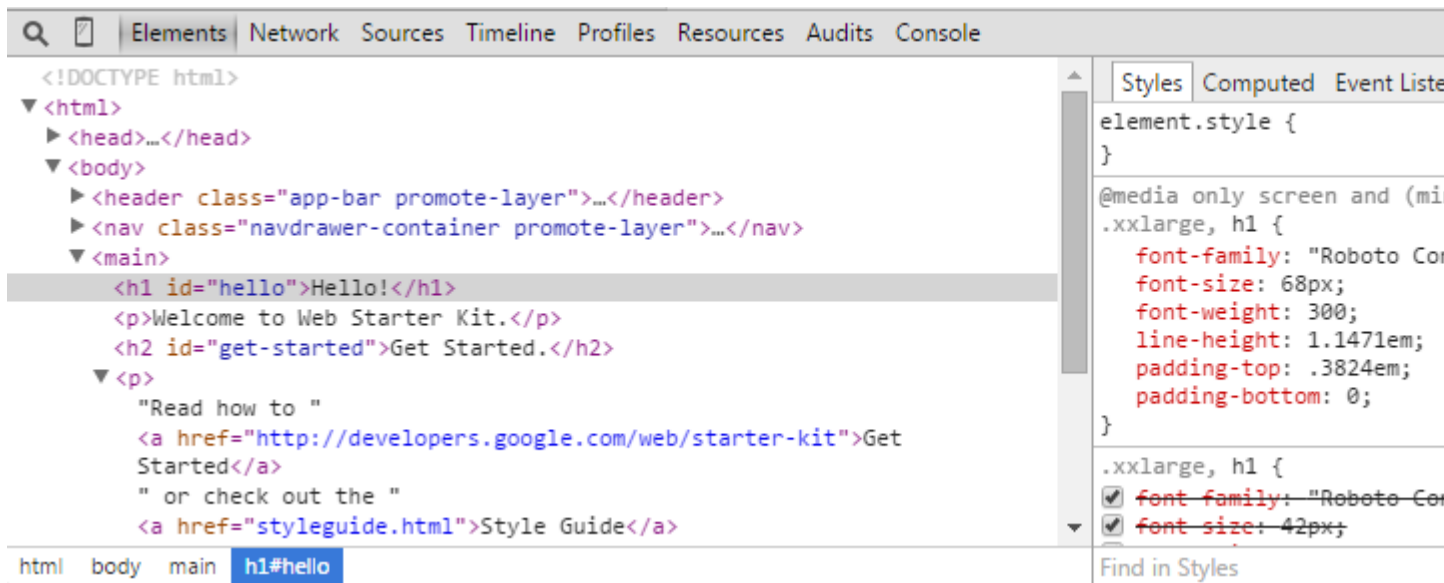
2.2 *CSS Selector*

Cascading Style Sheets (CSS) [3] adalah suatu mekanisme untuk mendefinisikan bagaimana suatu elemen HTML ditampilkan. Dalam CSS, terdapat *selector* yaitu pola yang digunakan untuk memilih elemen yang akan diubah tampilannya. Pada dasarnya, *CSS Selector* memberikan *style* kepada elemen HTML yang dipilihnya.

Jenis-jenis **CSS Selector** [4] dapat dilihat pada tabel berikut.

SELECTOR	DESKRIPSI
.class	Memilih elemen dengan kelas tertentu
#id	Memilih elemen dengan id tertentu
*	Memilih semua elemen
element	Memilih elemen dengan tag html tertentu
element1,element2	Memilih beberapa elemen sekaligus
element1Âœelement2	Memilih element2 di dalam suatu element1
element1>element2	Memilih elemen di mana element1 adalah parent element2
element1+element2	Memilih elemen2 di mana element1 berada setelah element1
element1~element2	Memilih elemen2 yang didahului elemen1
[attribute]	Memilih semua elemen yang memiliki atribut tertentu
[attribute=value]	Memilih semua elemen dengan atribut yang memiliki nilai tertentu
[attribute=̂value]	Memilih semua elemen dengan atribut yang mengandung nilai tertentu
[attribute =value]	Memilih semua elemen dengan atribut yang dimuali dengan nilai tertentu
[attribute^=value]	Memilih elemen tertentu dengan atribut yang diawali dengan nilai tertentu
[attribute\$=value]	Memilih elemen tertentu dengan atribut yang diakhiri dengan nilai tertentu

SELECTOR	DESKRIPSI
[attribute*=value]	Memilih elemen tertentu dengan atribut yang mengandung nilai tertentu
:active	Memilih active element
::after	Menyisipkan konten setelah elemen tertentu
::before	Menyisipkan konten sebelum elemen tertentu
:checked	Memilih checked element
:disabled	Memilih disabled element
:empty	Memilih elemen yang tidak memiliki child
:enabled	Memilih enabled element
:first-child	Memilih elemen yang merupakan child pertama
::first-letter	Memilih huruf pertama elemen dengan tertentu
::first-line	Memilih baris pertama elemen tertentu
:first-of-type	Memilih elemen yang merupakan type pertama
:focus	Memilih focus element
:hover	Memilih hover element
:in-range	Memilih elemen dengan nilai yang berada pada range tertentu
:invalid	Memilih elemen dengan nilai yang tidak valid
:lang(language)	Memilih elemen dengan nilai tertentu pada atribut lang
:last-child	Memilih elemen yang merupakan child terakhir
:last-of-type	Memilih elemen yang merupakan type terakhir
:link	Memilih link element
:not(selector)	Memilih elemen selain elemen tertentu
:nth-child(n)	Memilih elemen yang merupakan child ke-n
:nth-last-child(n)	Memilih elemen yang merupakan child ke-n, dihitung dari child terakhir
:nth-last-of-type(n)	Memilih elemen yang merupakan type ke-n, dihitung dari child terakhir
:nth-of-type(n)	Memilih elemen yang merupakan type ke-n
:only-of-type	Memilih elemen yang merupakan child dengan type tertentu
:only-child	Memilih elemen yang merupakan child
:optional	Memilih elemen yang tidak memiliki atribut "required"
:out-of-range	Memilih elemen dengan nilai yang berada di luar range tertentu
:read-only	Memilih elemen yang memiliki atribut "readonly"
:read-write	Memilih elemen yang tidak memiliki atribut "readonly"
:required	Memilih elemen yang memiliki atribut "required"
:root	Memilih root elemen
::selection	Memilih elemen dalam selection pengguna
:target	Memilih elemen yang sedang aktif
:valid	Memilih elemen dengan nilai yang valid
:visited	Memilih visited element

Gambar 2.1: *Elements Panel* [1]

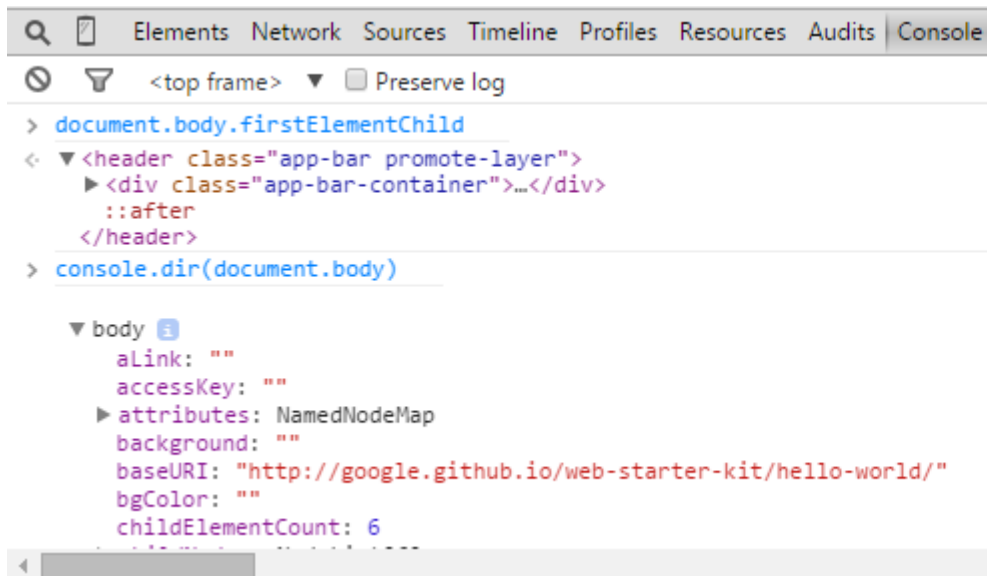
2.3 Document Object Model

Document Object Model (DOM) [5] adalah model yang merepresentasikan dokumen HTML dalam bentuk pohon. Setiap objek pada HTML merupakan elemen yang memiliki hubungan *parent-child* dengan elemen lainnya. Oleh karena itu, DOM direpresentasikan dalam bentuk pohon yang disebut dengan *DOM Tree*.

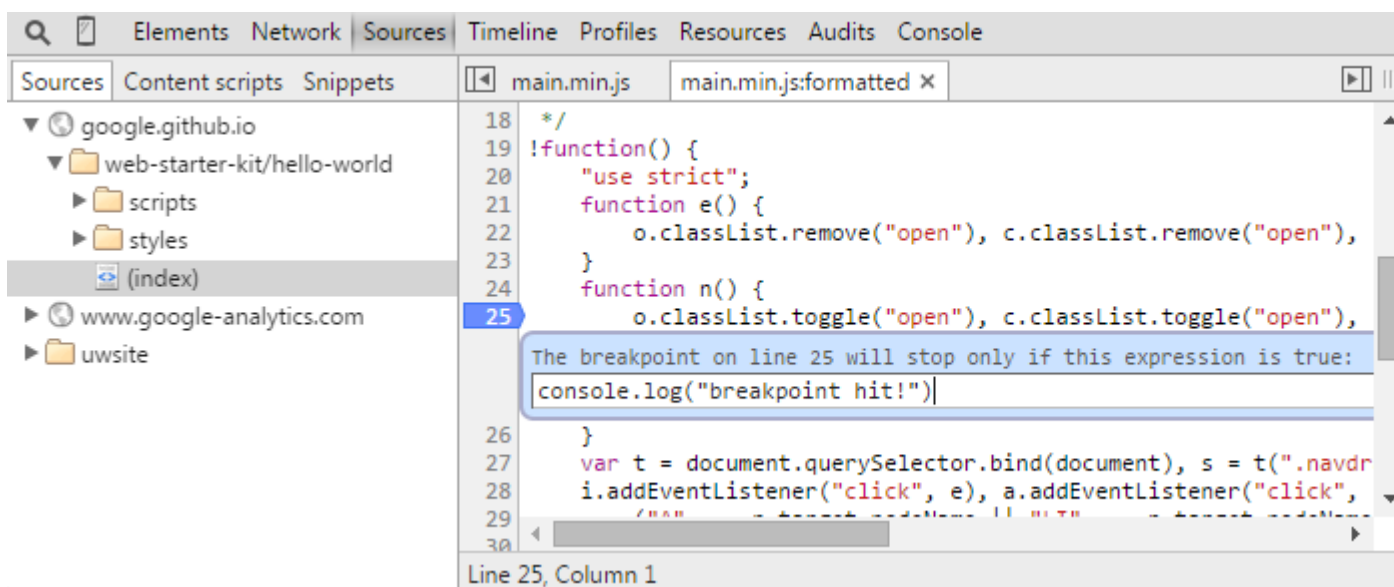
2.4 Chrome DevTools

Chrome Developer Tools (DevTools) [1] adalah seperangkat alat *web authoring* dan *debugging* yang dimiliki Google Chrome. *Developer tools* sendiri berfungsi untuk melakukan *debugging* terhadap *website* buatan sendiri. DevTools dapat melakukan *debugging* terhadap semua *website* yang dikunjungi. Fitur-fitur yang dimiliki DevTools antara lain:

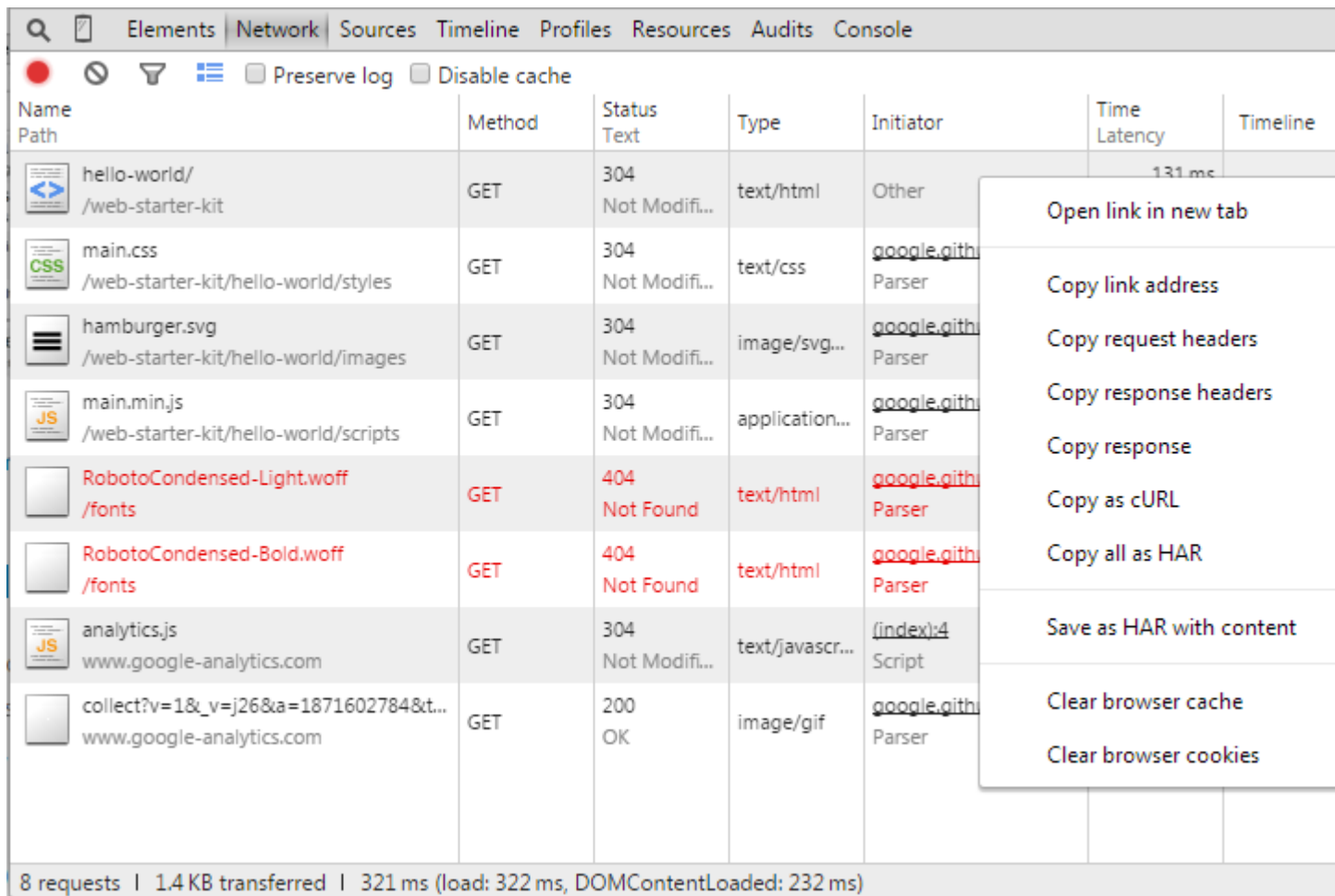
1. Elements, memeriksa dan mengubah DOM dan *style* dari suatu *website*.
2. Console, mendapatkan informasi pengembangan dan berinteraksi dengan dokumen.
3. Sources, melakukan *debugging* pada JavaScript dengan menentukan *breakpoint*.
4. Network, memantau kinerja jaringan pada *website* secara *real-time*.
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*.



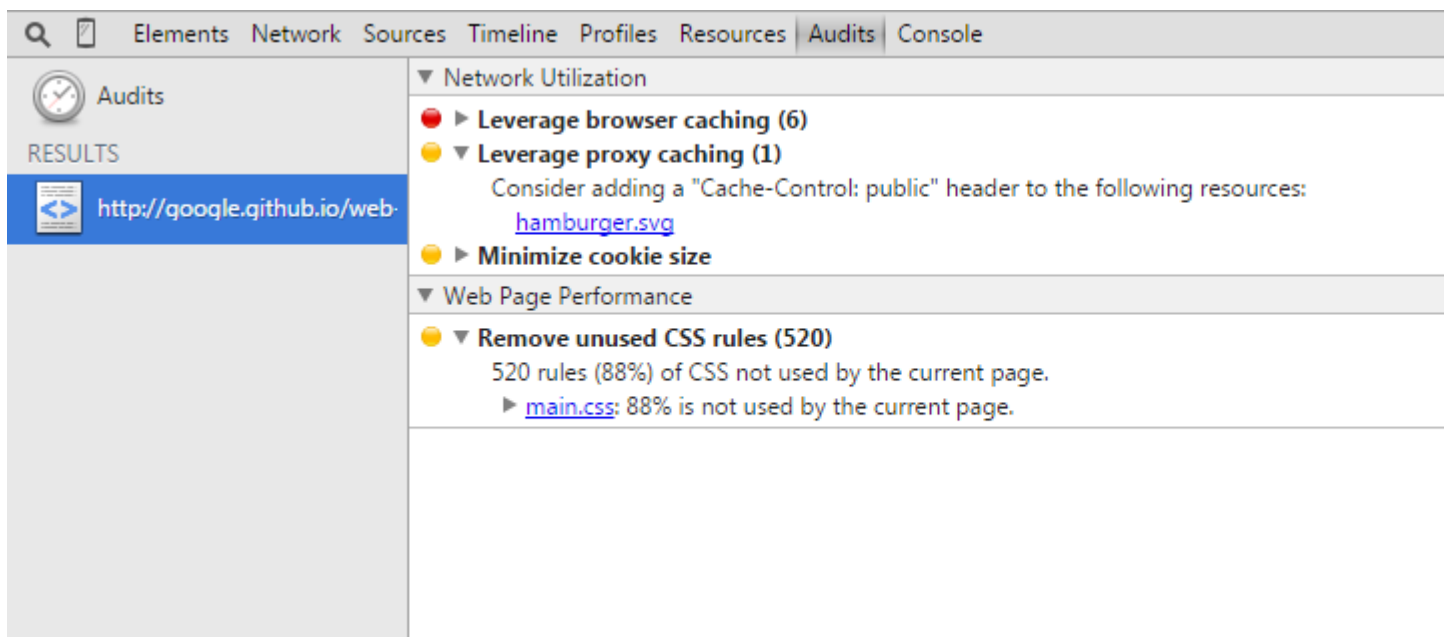
Gambar 2.2: Console Panel [1]



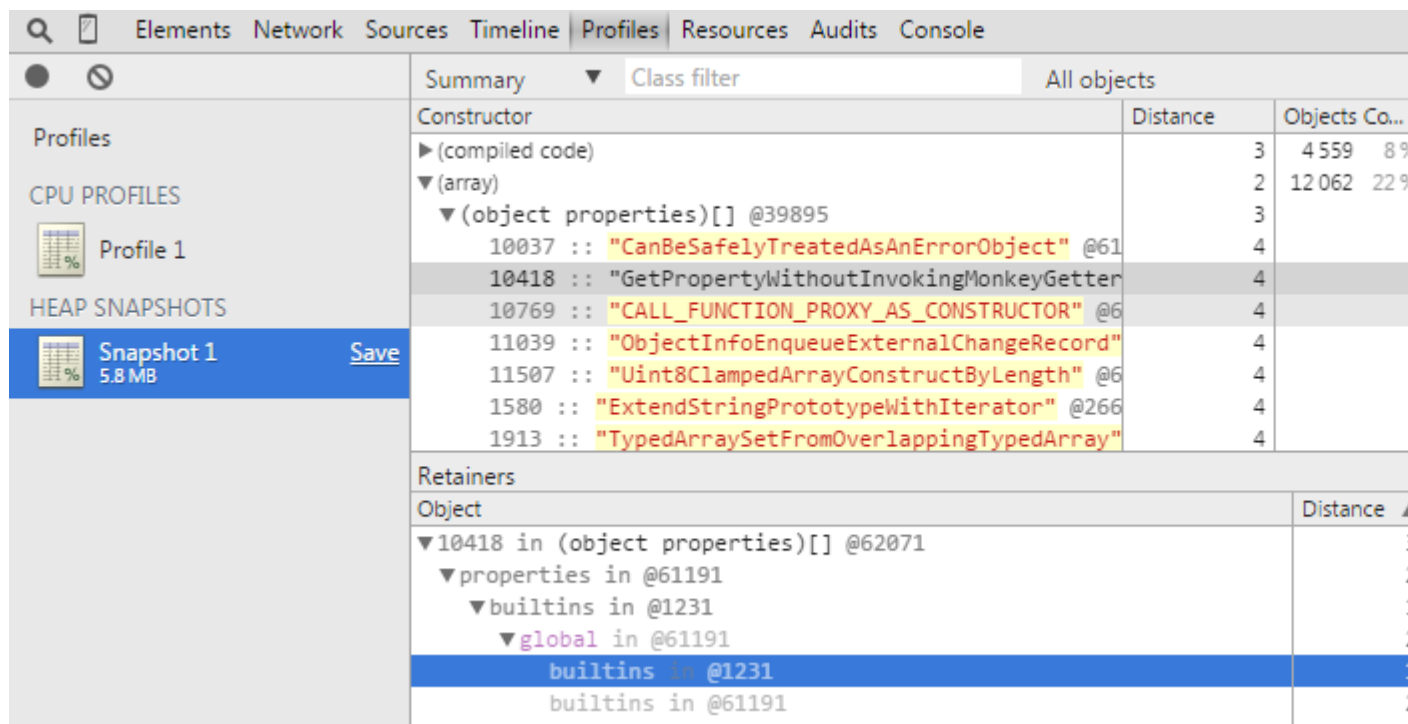
Gambar 2.3: Sources Panel [1]

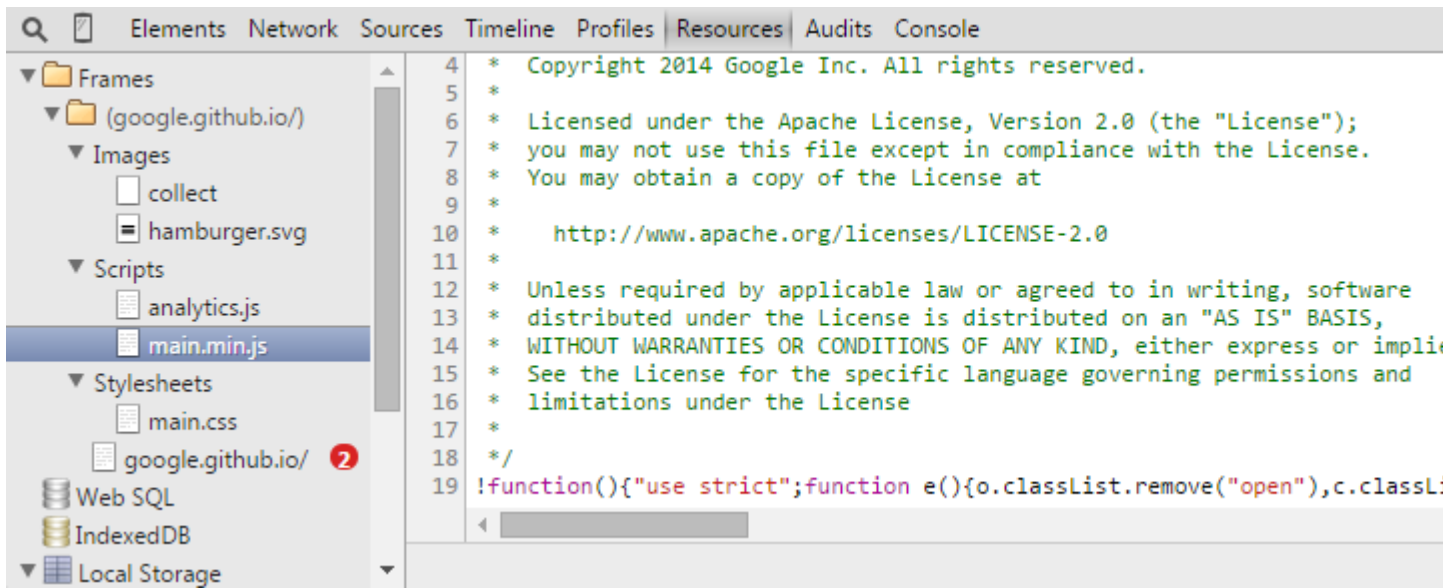


Gambar 2.4: Network Panel [1]



Gambar 2.5: Audits Panel [1]

Gambar 2.6: *Timeline Panel* [1]Gambar 2.7: *Profiles Panel* [1]

Gambar 2.8: *Resources Panel* [1]

2.5 jsoup

jsoup [6] adalah *library* Java yang bekerja dengan *real-world* HTML. jsoup juga menyediakan API untuk mengekstrak dan memanipulasi data menggunakan DOM, CSS, dan method seperti jquery. Selain itu, jsoup mengimplementasikan spesifikasi HTML5 dari *Web Hypertext Application Technology Working Group* (WHATWG) yaitu suatu organisasi yang mengembangkan HTML. jsoup merupakan sebuah proyek *open source* yang didistribusikan di bawah lisensi Massachusetts Institute of Technology (MIT).

jsoup melakukan *parsing* HTML menjadi DOM dengan teknik yang sama dengan teknik *browser* masa kini yaitu dengan [6] :

- *Scraping* dan *parsing* HTML dari sebuah URL, *file*, atau *string*.
- Menemukan dan mengekstrak data menggunakan penelusuran DOM atau CSS *Selector*
- Memanipulasi elemen, atribut, dan teks HTML

BAB 3

ANALISIS

3.1 Analisis Fitur-fitur FTIS Student Portal

Berdasarkan penelitian dan pengalaman penulis, fitur-fitur yang diperlukan oleh FTIS untuk kustomisasi Student Portal UNPAR adalah sebagai berikut:

1. Deskripsi Mata Kuliah

Dalam pengambilan mata kuliah, mahasiswa sering kali mengambil mata kuliah tanpa pertimbangan. Penyebabnya adalah kurangnya penjelasan mengenai mata kuliah tersebut. Akibatnya, terjadi ketidakcocokan mahasiswa dengan mata kuliah tersebut. Meskipun mahasiswa dapat melakukan Perubahan Rencana Studi (PRS), tidak terjamin mahasiswa dapat mengambil kuliah lain karena ada kemungkinan kelas mata kuliah lain yang akan diambil sudah penuh.

Sekretaris jurusan sebenarnya sudah mengumumkan deskripsi mata kuliah, namun deskripsi tersebut kurang lengkap. Tidak semua mata kuliah memiliki deskripsi karena tidak semua dosen membuat deskripsi tersebut. Pengumuman tersebut juga diumumkan melalui Facebook yang memungkinkan mahasiswa tidak membaca pengumuman tersebut. Seharusnya mahasiswa siaga dengan pengumuman sekretaris jurusan dalam minggu-minggu menjelang FRS, namun tetap saja terjadi kelalaian yang mengakibatkan mahasiswa tidak membaca pengumuman.

Untuk mengatasi masalah tersebut, penulis memutuskan untuk menambahkan fitur deskripsi mata kuliah. Fitur ini memberikan informasi tentang mata kuliah apa saja yang akan dibuka semester yang akan datang beserta penjelasannya. Dengan demikian, mahasiswa dapat mempertimbangkan mata kuliah apa saja yang akan diambil.

2. Prasyarat Mata Kuliah

Masalah lain yang sering dialami FTIS adalah kesalahan dalam mengisi Formulir Rencana Studi (FRS). Meskipun sudah diberi daftar mata kuliah beserta prasyaratnya, 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 ditambahkan dalam FTIS 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. Dengan adanya fitur prasyarat mata kuliah ini, beban dosen wali dan sekretaris jurusan akan berkurang.

3.2 Analisis *Web Scraping*

Sebelum melakukan implementasi *Web Scraping*, pengamatan kinerja jaringan perlu dilakukan. Untuk mengamati kinerja jaringan, kita membutuhkan fitur dari DevTools yaitu Network. Melalui *network panel*, kita bisa melihat *traffic* jaringan dari Student Portal UNPAR. Kita harus mengamati perpindahan halaman tersebut. Dalam berpindah halaman, terdapat pengiriman data seperti *method*, *form* dan *cookies*.

Setelah melakukan pengamatan kinerja jaringan, kita dapat mengimplementasikan *Web Scraping* menggunakan jsoup. Teknik *Web Scraping* yang digunakan oleh jsoup adalah HTML *parser* dan DOM *parsing*. Dengan jsoup kita bisa melakukan koneksi terhadap halaman-halaman yang dilewati. Setelah melakukan koneksi, data-data yang diperlukan oleh halaman selanjutnya akan dimasukkan sebagai parameter. Data yang berasal dari *form* dapat diperoleh dengan melakukan sedikit *Web Scraping* yaitu dengan mendapatkan elemen *form* tersebut menggunakan CSS *Selector* kemudian elemen tersebut direpresentasikan ke dalam DOM dengan tipe data *Document*. Melalui *Document* tersebut, kita bisa mendapatkan nilai dari *form* tersebut. Setelah memperoleh data tersebut, kita akan melakukan koneksi ke halaman selanjutnya dengan *method* POST atau GET sesuai dengan yang dilakukan Student Portal UNPAR.

Jika sudah terkoneksi dengan halaman yang diinginkan, kita dapat memperoleh elemen-elemen yang dibutuhkan dalam kustomisasi Student Portal UNPAR. Selain itu, kita juga bisa menambahkan fitur-fitur yang diperlukan oleh FTIS. Setelah melakukan *Web Scraping*, hasil kustomisasi akan ditampilkan ke dalam *website* baru yaitu FTIS Student Portal.

3.3 Analisis Aplikasi yang Akan Dibuat

Aplikasi yang akan dibuat merupakan aplikasi berbasis *web* menggunakan Play Framework yaitu *framework* dengan basis MVC (*Model View Controller*) dengan bahasa Java dan Scala. Namun bahasa yang akan digunakan adalah Java. Aplikasi ini juga akan menggunakan *library* jsoup. Dalam aplikasi ini, jsoup merupakan komponen utama untuk melakukan *Web Scraping*. Hasil dari *Web Scraping* akan dibuat sebagai HTML dari FTIS Student Portal. Kustomisasi fitur-fitur FTIS akan diterapkan melalui penambahan HTML pada hasil *Web Scraping* dan pemrograman bahasa Java untuk mengimplementasikan fungsi dari fitur-fitur kustomisasi. Hasil *Web Scraping* yang sudah dikustomisasi akan dijadikan sebagai *view*. *Model* yang akan digunakan adalah mahasiswa dan mata kuliah yang akan direpresentasikan dalam kelas Mahasiswa dan kelas MataKuliah. *Controller* berfungsi untuk menghubungkan *view* dan *model* sekaligus mengatur pemanggilan *method* yang mengimplementasikan fitur-fitur yang dimiliki FTIS Student Portal.

DAFTAR REFERENSI

- [1] Google, “Chrome devtools.”
- [2] 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.
- [3] “Css.”
- [4] “Css selector.”
- [5] “Dom.”
- [6] “jsoup.”