

## PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini tidak mengandung karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak mengandung karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka

Yogyakarta, 6 September 2018



Selvi Oktaria

## KATA PENGANTAR

Penulis bersyukur dengan segenap hati kepada Allah SWT yang telah memberikan segalanya sehingga penulisan karta Tugas Akhir yang berjudul TEKNIK SCRAPING DAN CRAWLING UNTUK MENGEKSTRAKSI REVIEW HOTEL ONLINE PADA WEBSITE TRAVELOKA (BERBASIS AJAX) telah terselesaikan dengan baik. Pengembangan sistem bertujuan untuk memberikan informasi kepada pemerintah tentang pariwisata untuk menentukan kebijakan-kebijakan terkait pengembangan sektor pariwisata.

Dalam melakukan pengembangan sistem dan penyusunan laporan skripsi penulis telah mendapatkan banyak dukungan dan bantuan dari berbagai pihak. Penulis mengucapkan terima kasih yang tak terhingga kepada:

1. Bapak M. Helmi Zain Nuri, S.T., M.T. selaku pembimbing pertama yang telah membantu dalam proses pengembangan sistem dan proses penulisan laporan akhir dengan penuh kesabaran dan ketulusan.
2. Bapak Chayadi Oktomy Noto S., S.T., M.Eng., ITILF. selaku pembimbing kedua yang telah meluangkan waktu membimbing penulis dengan penuh kesabaran dan ketulusan dalam penulisan skripsi.
3. Bapak Dr. Ir. Dwijoko Purbohadi, M.T. selaku dosen penguji yang telah memberikan masukan berarti, serta dengan semangat dan pengertian dalam ilmu dan pengalaman selama masa perkuliahan penulis.
4. Para dosen dan staf Teknik Informatika yang telah memberikan ilmunya selama penulis belajar di Universitas Muhammadiyah Yogyakarta.
5. Kedua orang tua penulis Supriono dan Siti Kalimah dan saudara kandung penulis Arif Nugroho yang selalu memberikan semangat, nasehat, cinta, perhatian, materi, dan kasih sayang serta doa untuk penulis.
6. Erise Juliana, Deantari Dwikan P.A dan Arief Fadillah Ritonga yang selalu memberikan semangat dan membantu dalam setiap proses pengerjaan tugas akhir ini.
7. Barista-barista Starbucks Galleria Mall yang selalu menyajikan baverage dengan sepenuh hati untuk penulis sebagai teman setia yang mendampingi penulis selama proses pengerjaan tugas akhir.
8. Bapak-bapak penjaga parkir Galleria Mall yang selalu setia menjaga kendaraan penulis dan selalu mengingatkan pulang tepat waktu.
9. Bapak driver Gojek dan Rumah Ori yang selalu sedia membuatkan dan mengantarkan makanan untuk penulis dikala lapar tengah malam.
10. Semua pihak yang tidak dapat penulis sebutkan satu-persatu, yang telah membantu secara langsung maupun tidak langsung.

Yogyakarta, 9 Agustus 2018

Selvi Oktaria

## DAFTAR ISI

HALAMAN PENGESAHAN.....	i
PERNYATAAN.....	ii
KATA PENGANTAR .....	iii
INTISARI.....	iv
<i>ABSTRACT</i> .....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR .....	viii
DAFTAR TABEL.....	ix
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan masalah.....	2
1.3. Batasan masalah .....	3
1.4. Tujuan .....	3
1.5. Manfaat .....	3
BAB II DASAR TEORI .....	5
2.1. Struktur Web .....	5
2.1.1. HTML.....	5
2.1.2. CSS .....	8
2.1.3. JavaScript .....	8
2.2. AJAX .....	9
2.3. Scraping.....	10
2.3.1. Scrapy.....	11
2.3.2. Selenium.....	12
BAB III PENGEMBANGAN SISTEM.....	14
3.1. Metodologi Pengembangan Sistem.....	14
3.2. Analisa Kebutuhan.....	15
3.3. Kebutuhan Fungsional .....	16
3.4. Mendefinisikan Sumber Data.....	16
3.5. Analisa Sumber Data.....	17

3.6.	Analisa Sistem.....	18
3.7.	Rancangan Sistem .....	19
3.8.	Implementasi dan Pengujian .....	21
3.9.	Proses Ekstraksi .....	21
3.10.	Penyimpanan Data Hasil Ekstraksi .....	22
	<b>BAB IV IMPLEMENTASI dan PEMBAHASAN .....</b>	<b>23</b>
4.1.	Model Data.....	23
4.2.	Analisa Web Traveloka.....	24
4.3.	Instalasi Scrapy Framework.....	33
4.4.	Pengembangan scrapy.....	35
4.4.1.	Activity Diagram.....	35
4.4.2.	Konfigurasi Scrapy.....	37
4.4.3.	Spider Url Hotel .....	38
4.4.4.	Spider Review Hotel .....	40
4.4.5.	Konfigurasi Modul Elasticsearch .....	43
4.5.	Pengujian.....	44
4.5.1.	Pengujian Fungsional Scraping dan Crawling .....	44
4.5.2.	Pengujian Proses Scraping dan Crawling.....	49
4.6.	Hasil .....	50
	<b>BAB V KESIMPULAN .....</b>	<b>53</b>
	<b>DAFTAR PUSTAKA .....</b>	<b>53</b>

## DAFTAR GAMBAR

Gambar 2-1. Struktur HTML .....	7
Gambar 2-2. Arsitektur Ajax.....	9
Gambar 2-3. Tahap-tahap scraping.....	10
Gambar 2-4. Arsitektur Scrapy (Scrapy Documentation, 2016).....	11
Gambar 2-5. Arsitektur Selenium WebDriver .....	13
Gambar 3-1. Tahapan pengembangan sistem scraping dsn crawling .....	15
Gambar 3-2. Halaman review Traveloka.....	17
Gambar 3-3. Arsitektur sistem.....	20
Gambar 4-1. Struktur web Traveloka.....	24
Gambar 4-2. Halaman pencarian hotel.....	25
Gambar 4-3. Form pencarian hotel .....	25
Gambar 4-4. Halaman daftar hotel.....	26
Gambar 4-5. Halaman hotel Traveloka.....	26
Gambar 4-6. Review hotel Traveloka .....	27
Gambar 4-7. Elemen url hotel.....	28
Gambar 4-8. Elemen membuka halaman hotel .....	28
Gambar 4-9. Elemen url hotel.....	29
Gambar 4-10. Elemen canonical .....	29
Gambar 4-11. Elemen data hotel.....	30
Gambar 4-12. Elemen nama hotel.....	30
Gambar 4-13 Elemen rating hotel.....	30
Gambar 4-14. Elemen bintang hotel .....	30
Gambar 4-15. Elemen alamat hotel.....	30
Gambar 4-16. Elemen review hotel .....	31
Gambar 4-17. Elemen review .....	32
Gambar 4-18. Elemen button next .....	32
Gambar 4-19. File directory scrapy project .....	34
Gambar 4-20. Activity diagram .....	36
Gambar 4-21. Pengujian ekstraksi list url hotel.....	45
Gambar 4-22. Pengujian ekstraksi data hotel.....	46
Gambar 4-23. Pengujian ekstraksi review hotel .....	47
Gambar 4-24. Pengujian crawling 1.....	47
Gambar 4-25. Pengujian crawling 2.....	48
Gambar 4-26. Pengujian proses concurrent .....	49
Gambar 4-27. Hasil ekstraksi 1 .....	51
Gambar 4-28. Hasil ekstraksi 2.....	51
Gambar 4-29. Hasil ekstraksi (json) .....	52

## DAFTAR TABEL

Tabel 4- 1. Tabel data ekstraksi .....	23
Tabel 4- 2. Tabel XPath 1 .....	32
Tabel 4- 3. Tabel XPath 2 .....	33