

IMPLEMENTASI WEB CRAWLING UNTUK MENGUMPULKAN INFORMASI WISATA KULINER DI BANDAR LAMPUNG

Raidah Hanifah¹, Isye Susana Nurhasanah²

¹ Program Studi Teknik Informatika, Institut Teknologi Sumatera

² Program Studi Perencanaan Wilayah dan Kota, Institut Teknologi Sumatera
Email: ¹raidah.hanifah@if.itera.ac.id, ²isye.susana@pwk.itera.ac.id

(Naskah masuk: 6 Juni 2018, diterima untuk diterbitkan: 23 Oktober 2018)

Abstrak

Bandar Lampung merupakan ibu kota Provinsi Lampung yang terletak di ujung pulau Sumatera. Berbeda dengan beberapa Kabupaten di Provinsi Lampung yang memiliki banyak potensi wisata alam, Bandar Lampung sebagai pusat kota tidak banyak memiliki potensi wisata alam. Namun, kota Bandar Lampung masih memiliki potensi untuk mengembangkan wisata kuliner. Untuk menangkap peluang ekonomi dari wisata kuliner, salah satu caranya adalah dengan membuat pemetaan lokasi-lokasi wisata kuliner strategis yang telah dan berpotensi untuk berkembang, serta membuat perencanaan dan program pengembangan wisata kuliner beserta dengan promosinya yang efektif dan optimal. Sebelum hal tersebut dilakukan, tentunya dibutuhkan informasi terkait kondisi wisata kuliner yang ada saat ini, seperti berapa jumlah wisata kuliner dan letak persebaran lokasinya. Sayangnya, informasi tersebut belum diakomodasi oleh pemerintah kota Bandar Lampung. Informasi terkait kuliner di Bandar Lampung justru tersedia di berbagai *website* dan media sosial lainnya, mengingat saat ini orang senang berbagi informasi melalui internet. Oleh karena itu, penelitian ini bertujuan untuk mengumpulkan informasi kuliner di Bandar Lampung yang tersebar di *web*. Untuk mencapai tujuan tersebut, dilakukan penggalian *web* (*web mining*) untuk mendapatkan informasi terkait wisata kuliner di Bandar Lampung dengan metode *web crawling*. Hasil *web crawling* dari *website* TripAdvisor, didapatkan berbagai macam informasi seperti nama restoran, alamat, nomor telepon, *rating* serta jumlah orang yang *me-review* kuliner tersebut. Informasi yang telah dikumpulkan tersebut dapat dianalisis lebih lanjut untuk berbagai kebutuhan, seperti analisis spasial persebaran objek wisata kuliner, analisis rasio perkembangan jumlah wisata kuliner terhadap jumlah wisatawan yang datang, perencanaan pengembangan wisata kuliner di suatu wilayah secara spesifik maupun komprehensif, serta lokasi dan jenis kuliner yang paling diminati.

Kata kunci: *web mining, web crawling, wisata kuliner, bandar lampung, tripadvisor.*

WEB CRAWLING IMPLEMENTATION FOR COLLECTING CULINARY TOURISM INFORMATION IN BANDAR LAMPUNG

Abstract

Bandar Lampung is the capital city of Lampung Province which is located on the edge of the Sumatera island. Unlike some regencies in Lampung Province which have a lot of potential natural tourism, Bandar Lampung as a capital city does not have much. However, Bandar Lampung city still has the potential to develop its culinary tourism. To capture the economic opportunities of culinary tourism, one of the way is to map strategic culinary tourism locations that have developed and potentially develop, and create a comprehensive planning and programs for culinary tourism development along with effective and optimal promotion. But before that, information regarding the current conditions of culinary tourism is needed; for instance, how many and where the location of the culinary tourism are in Bandar Lampung area. Unfortunately, this information has not accomodate yet by the municipility government. Culinary information in Bandar Lampung is actually available on various websites and other social media, considering that currently people like to share information via the internet. So, this study aim is to collect culinary tourism information in Bandar Lampung that sprawl in web. In accomplishing the goal, this research conducts web mining to get information related to culinary tourism in Bandar Lampung by web crawling method. The results of web crawling from the TripAdvisor website such as restaurant name, address, telephone number, rating and the number of people who reviewed the culinary. The result of this research can be use for further analysis with difference approach and goals, such as spatial analysis of culinary tourism sprawl, analysis the impact of culinary tourism development to the number of tourists coming, culinary tourism development and planning in certain region, as well as the type and location of favorable culinary in Bandar Lampung.

Keywords: *web mining, web crawling, culinary tourism, bandar lampung, tripadvisor.*

1. PENDAHULUAN

Provinsi Lampung dapat dikatakan sebagai gerbang Pulau Sumatera. Secara geografis provinsi Lampung terletak di paling selatan pulau Sumatera, sehingga tidak mengherankan jika provinsi Lampung menjadi penghubung sekaligus pintu masuk ke pulau Sumatera dari pulau Jawa. Saat ini Provinsi Lampung sudah dikenal sebagai alternatif tempat untuk berwisata. Beberapa tempat wisata terkenal di Provinsi Lampung, yaitu Pulau Pahawang, Teluk Kilauan, Taman Nasional Way Kambas, Taman Nasional Bukit Barisan Selatan, dan Gunung Krakatau. Jumlah kunjungan wisatawan, baik domestik maupun asing sudah mencapai 7,5 juta di tahun 2016. Angka tersebut naik 2 juta dari yang sebelumnya di tahun 2015 sebanyak 5,5 juta wisatawan. (Traveling.bisnis.com, 2017)

Ibu kota Provinsi Lampung adalah kota Bandar Lampung. Berbeda dengan beberapa Kabupaten di Provinsi Lampung yang memiliki banyak potensi wisata alam, Bandar Lampung sebagai pusat kota tidak banyak memiliki potensi wisata alam. Namun, kota Bandar Lampung masih memiliki peluang untuk mengembangkan wisata kuliner. Sebagaimana yang disampaikan Dinas Pariwisata kota Bandar Lampung bahwa dalam rangka mendongkrak angka kunjungan wisatawan, maka akan dimaksimalkan potensi wisata kota Bandar Lampung dengan fokus kepada promosi wisata kuliner dan budaya (Lampungpro.com, 2017).

Saat ini, wisata kuliner sudah menjadi gaya hidup masyarakat perkotaan (Pandean & Hansun, 2018). Orang senang untuk makan di luar rumah sambil menikmati suasana yang berbeda. Wisata kuliner juga menjadi hiburan tersendiri bagi para wisatawan. Ditambah lagi, banyak tayangan di televisi yang membahas tentang kuliner dari berbagai daerah di Indonesia, baik itu terkait rasa, tempat, suasana hingga harga. Hal tersebut membuat wisata kuliner semakin populer dan mendorong masyarakat untuk lebih mengenal masakan khas daerah di Indonesia.

Untuk menangkap peluang ekonomi dari wisata kuliner, salah satu caranya adalah membuat pemetaan lokasi-lokasi wisata kuliner strategis yang telah dan berpotensi untuk berkembang, serta membuat perencanaan dan program pengembangan wisata kuliner beserta dengan promosinya yang efektif dan optimal. Sebelum hal tersebut dilakukan, tentunya dibutuhkan informasi terkait kondisi wisata kuliner yang ada saat ini, seperti berapa jumlah wisata kuliner dan letak persebaran lokasinya. Sayangnya, informasi tersebut belum diakomodasi oleh pemerintah kota Bandar Lampung. Oleh karena itu, penelitian ini bertujuan untuk mengumpulkan informasi kuliner di Bandar Lampung dengan menggunakan *web* sebagai sumber data.

Saat ini, *World Wide Web* (WWW) atau yang biasa disebut dengan *web*, telah menjadi sumber informasi yang sangat besar dan sangat dikenal orang. Ketika seseorang ingin mencari suatu informasi, yang

pertama kali diakses adalah *web*. *Web* terdiri dari miliaran dokumen yang saling berhubungan dan dikelola oleh jutaan orang. *Web* mudah untuk diakses, sehingga pencarian informasi melalui *web* dapat dilakukan dimana saja dan kapan saja. (Liu, 2011)

Penggalian informasi dari *web* atau *web mining* bisa menjadi salah satu solusi untuk mengumpulkan informasi wisata kuliner yang memanfaatkan *web* sebagai sumber data. Penggunaan cara tradisional, seperti survei, wawancara ataupun kuisioner, seringkali terkendala masalah dana dan geografis (Peng, Li, Wang, & Chen, 2018). Di era digital saat ini, dimana banyak informasi yang dibagikan orang melalui internet, *web mining* diharapkan dapat menjadi solusi yang efektif untuk mencari informasi ataupun mengumpulkan informasi.

Istilah *web mining* pertama kali digunakan oleh Oren Etzioni. *Web mining* didefinisikan oleh Oren Etzioni sebagai penerapan teknik *data mining* ke data *web* untuk mengekstraksi informasi yang relevan dari sumber daya yang tersedia di *web*, sumber daya dapat berupa dokumen atau *web service*. (Sellamy, Fakhri, Boulaknadel, Moumen, & Hafed, 2018). Proses *web mining* mirip dengan proses *data mining*, yang membedakan biasanya dalam proses pengumpulan datanya. Dalam proses *data mining*, data sudah dikumpulkan dalam *database* atau *data warehouse*. Sedangkan dalam *web mining*, pengumpulan data dari *web* menjadi tugas penting. Untuk mengunduh sejumlah halaman *web* secara otomatis, maka perlu dilakukan *crawling*. *Crawling* adalah proses menjelajahi *web* dan mengunduh halaman *web* secara otomatis untuk mengumpulkan informasi. Program yang khusus bertugas melakukan *crawling* disebut *Crawler*. (Liu, 2011)

Penelitian ini akan mengimplementasikan metode *web crawling* untuk mengumpulkan informasi wisata kuliner di Bandar Lampung. Sumber data yang digunakan adalah *website* TripAdvisor. TripAdvisor dipilih karena *website* tersebut merupakan situs wisata terbesar di dunia (Comscore.com, 2018). TripAdvisor memiliki lebih dari 661 juta ulasan dan opini tentang 7,7 juta akomodasi, maskapai penerbangan, objek wisata, dan restoran. Situs TripAdvisor tersedia untuk 49 negara, dan menaungi komunitas wisata terbesar di dunia, dengan rata-rata 456 juta pengunjung setiap bulannya (Tripadvisor.com, 2017).

2. METODE PENELITIAN

Di era big data saat ini, jumlah data dan jenis informasi yang ada di *web* semakin meningkat. Dimana *web* telah menjadi sumber data, orang membutuhkan suatu cara untuk membantu mengumpulkan data dari *web*. *Web crawling* akan mencari halaman *web* melalui URL (*Uniform Resource Locator*), dan mengembalikan data yang bersangkutan ke pengguna secara langsung. Pengguna tidak perlu mengakses informasi dengan menelusuri halaman *web* satu per satu sehingga menghemat waktu dan tenaga serta meningkatkan

akurasi pengumpulan data (Peng et al., 2018). Data yang sudah dikumpul kemudian dapat diolah dan dianalisis lebih lanjut sesuai dengan kebutuhan pengguna.

Banyak penelitian terkait *web mining* yang melakukan *web crawling* untuk mengumpulkan data. Data yang di-*crawling* pun beragam, mulai dari film (Liu, Li, & Wu, 2018), artikel berita (Hu, Wang, Wu, & Acquisition, 2018), tweet (Kim & Ha, 2016), lowongan pekerjaan (Peng et al., 2018), hingga informasi terkait minyak (Ren, 2018). Sumber objek yang dijadikan sumber informasi dapat terdiri dari beberapa *website* (Hu et al., 2018; Ren, 2018), ataupun fokus pada satu *website* seperti mencari informasi terkait film dari douban.com (Liu et al., 2018), serta dapat juga berasal dari media sosial seperti Twitter (Kim & Ha, 2016) dan Weibo (Xia et al., 2015).

Penelitian ini akan melakukan *web crawling* untuk mengumpulkan informasi terkait lokasi atau alamat wisata kuliner di Bandar Lampung dari *website* TripAdvisor. Tidak banyak *website* yang khusus berisi informasi wisata kuliner di Bandar Lampung, namun ada beberapa yang berhasil ditemukan yaitu tripadvisor.com, jelajahlampung.com, dan jajanlampung.com. Akan tetapi, pada penelitian ini, TripAdvisor dipilih karena informasi yang ada didalamnya sudah dalam bentuk semistruktur sehingga lebih mudah untuk dilakukan proses ekstraksi informasi. Kebanyakan penelitian melakukan *web crawling* menggunakan Scrapy dari Python (Hu et al., 2018; Liu et al., 2018; Peng et al., 2018; Ren, 2018), pada penelitian ini *web crawling* dilakukan dengan menggunakan bahasa R.

Langkah-langkah yang dilakukan dalam melakukan *web crawling* dapat dilihat pada Gambar 1.

1. Initialize Seed URL and Fetch The Page

Langkah pertama yaitu menginisialisasi *Seed URL*. *Seed URL* adalah URL awal yang sudah ditentukan oleh pengguna untuk melakukan *crawling*. *Seed URL* dalam penelitian ini adalah *webpage* utama TripAdvisor yang berisi daftar nama restoran yang ada di Bandar Lampung. *Seed URL* tersebut diakses halaman webnya.

2. Link Extraction

Selanjutnya, proses ekstraksi URL yang ada di halaman tersebut sebelumnya. URL yang belum dikunjungi disebut sebagai *Frontier*. Ekstraksi dilakukan dengan menggunakan *Regular Expression*. *Regular Expression* akan mengidentifikasi pola URL dalam data teks. Selain URL, pada tahapan ini diekstraksi juga nama restoran, yang dilakukan dengan cara mem-*parsing* berdasarkan tag HTML "*title*".

3. URL Conversion

URL hasil ekstraksi bentuknya masih belum baku (*relative URL*), sehingga tidak bisa langsung disimpan sebagai *Frontier*. *Relative URL* tersebut

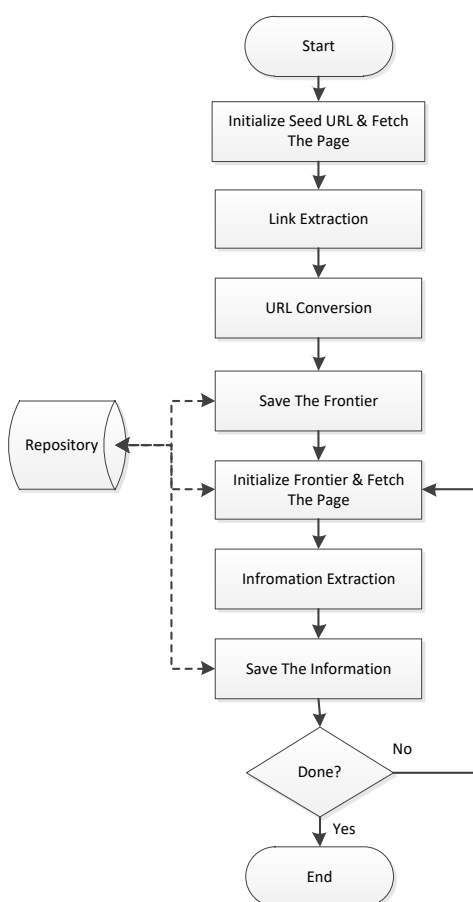
harus diubah terlebih dahulu menjadi bentuk yang baku (*absolute URL*). Konversi dilakukan dengan cara menambahkan URL utama dari *website* TripAdvisor ke depan *relative URL*.

Contoh *relative URL* ;

Restaurant_Review-g297722-d7309448-Reviews-El_s_Coffee_House-Bandar_Lampung_Lampung_Sumatra.html

Setelah Dikonversi menjadi *absolute URL* :

https://www.tripadvisor.co.id/Restaurant_Review-g297722-d7309448-Reviews-El_s_Coffee_House-Bandar_Lampung_Lampung_Sumatra.html



Gambar 1. Metode Web Crawling

4. Save The Frontier

Absolute URL kemudian disimpan ke dalam *repository* untuk digunakan pada tahap selanjutnya.

5. Initialize Frontier and Fetch The page

URL dari *Frontier* kemudian diambil kembali dari *repository* untuk diakses halaman webnya. *Frontier* pada penelitian ini berisi detail informasi terkait sebuah restoran.

6. Information Extraction

Langkah selanjutnya adalah mengekstraksi informasi dengan cara mem-*parsing* berdasarkan struktur halaman HTML, yang dilihat dari *tag* HTML. Pada penelitian ini beberapa informasi yang diekstraksi yaitu alamat, kota, kode pos, *website*, nomor telepon, *rating*, dan jumlah *reviewer*. Tabel 1 menunjukkan jenis informasi yang diekstraksi beserta *tag*-nya di dalam HTML.

Tabel 1. Jenis Informasi dan *Tag* HTMLnya.

No	Informasi	Tag HTML
1	Alamat	<i>street-address</i>
2	Kota	<i>locality</i>
3	Kode Pos	<i>locality</i>
4	<i>Website</i>	<i>website</i>
5	No. Telepon	<i>phone</i>
6	<i>Rating</i>	<i>rating</i>
7	Jumlah <i>Reviewer</i>	<i>rating</i>

Beberapa informasi ada yang tersimpan dalam HTML yang sama, seperti informasi kota dan kode pos yang tersimpan dalam tag "*locality*", serta *rating* dan jumlah *reviewer* yang tersimpan dalam tag "*rating*". Oleh karena itu, perlu dilakukan proses ekstraksi selanjutnya. Ekstraksi dilakukan dengan menggunakan *Regular Expression*.

7. Save The Information

Hasil informasi yang sudah didapatkan kemudian disimpan di dalam *repository*.

3. HASIL DAN PEMBAHASAN

Hasil *web crawling* dari *website* TripAdvisor berhasil mendapatkan 129 informasi terkait restoran atau wisata kuliner yang ada di Bandar Lampung. Namun, ada beberapa rincian informasi yang tidak lengkap, seperti tidak adanya informasi nomor telepon ataupun nama *website*. Tabel 2 menunjukkan persentase kelengkapan jumlah informasi yang didapatkan. Dapat dilihat juga bahwa ada informasi yang tidak ada sama sekali yaitu nama *website*. Hal tersebut disebabkan karena informasi tidak tersedia pada halaman *website* TripAdvisor.

Rata-rata informasi yang lengkap dan berhasil didapatkan adalah sebesar 71.1%. Beberapa hasil informasi yang diperoleh dari *website* TripAdvisor dapat dilihat pada Tabel 3. Hasil informasi yang telah

dikumpulkan tersebut dapat dianalisis lebih lanjut untuk berbagai kebutuhan seperti analisis spasial persebaran objek wisata kuliner, analisis rasio perkembangan jumlah wisata kuliner terhadap jumlah wisatawan yang datang, perencanaan pengembangan wisata kuliner di suatu wilayah secara spesifik maupun komprehensif, serta lokasi dan jenis kuliner yang paling diminati. Analisis-*analisis* tersebut dapat menjadi acuan untuk menentukan kebijakan pengembangan kota yang tepat, dalam rangka meningkatkan ekonomi daerah.

Tabel 2. Hasil Ekstraksi Informasi

No	Jenis Informasi	Jumlah Informasi yang Tersedia	Jumlah Informasi yang Tidak Tersedia	Persentase Kelengkapan Informasi
1	Alamat	126	3	97,67%
2	Kota	126	3	97,67%
3	Kode Pos	74	52	57,36%
4	<i>Website</i>	96	30	74,42%
5	No. Telepon	0	129	0%
6	<i>Rating</i>	110	19	85,27%
7	Jumlah <i>Reviewer</i>	110	19	85,27%
Rata-rata informasi yang tersedia dan berhasil didapatkan				71,1%

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa *web crawling* dapat digunakan sebagai metode untuk mengumpulkan informasi dari *web*. Sumber yang dijadikan obyek *web crawling* dapat berasal dari *website* yang berkaitan dengan bidang tertentu seperti pariwisata, contohnya yaitu *website* TripAdvisor. Hasil dari *web crawling* dari *website* TripAdvisor menunjukkan bahwa ada beberapa rincian informasi yang tidak lengkap, dimana rata-rata informasi yang lengkap dan berhasil didapatkan adalah sebesar 71.1%. Oleh karena itu, untuk pengembangan penelitian selanjutnya, dapat dipertimbangkan untuk menambah sumber *website* yang dijadikan obyek penelitian serta pemanfaatan media sosial seperti Twitter atau Instagram sebagai sumber informasi. Banyaknya obyek yang dijadikan sumber informasi tentunya akan semakin memperkaya informasi terkait wisata kuliner yang diperoleh.

Tabel 3. Contoh Informasi Hasil *Web Crawling* dari *Website* TripAdvisor

No	Nama Kuliner	Alamat	Kode Pos	No Telepon	<i>Website</i>	<i>Rating</i>	Jumlah <i>Reviewer</i>
1	Bakso Sonhaji Sony	Jln. Wolter Monginsidi no. 42A	35119	085764139899	+ Tambah situs web	4,0	187
2	Pempek 123	Jl. Jend. Sudirman No. 15	NA	0721255791	+ Tambah situs web	4,0	173
3	Jumbo Kakap	Jln. Ikan Kakap No. 45	35223	0721476159	+ Tambah situs web	4,0	102

No	Nama Kuliner	Alamat	Kode Pos	No Telepon	Website	Rating	Jumlah Reviewer
4	Taman Santap Rumah Kayu	Jl. Arief Rahman Hakim No. 45, Way Halim	35141	21700666	Situs Web	4,0	95
5	Begadang Resto Convention Hall	Jl. Diponegoro no. 1	NA	0721474059	Situs Web	4,0	116

DAFTAR PUSTAKA

- COMSCORE.COM., 2018. Top 50 Multi-Platform Properties (Desktop and Mobile). [online] Tersedia di: <<https://www.comscore.com/Insights/Rankings>> [Diakses 15 April 2018]
- HU, C., WANG, Y., WU, L., & ACQUISITION, A. D., 2018. Analysis of Hot News Based on Big Data, *17th International Conference on Computer and Information Science (ICIS)*, pp. 678–681.
- KIM, S., & HA, Y., 2016. Automated Discovery of Small Business Domain Knowledge Using Web Crawling and Data Mining, *Big Data and Smart Computing (BigComp)*, pp. 481–484.
- LAMPUNGPRO.COM., 2017. Promosi Wisata Bandar Lampung Fokus pada Kuliner dan Budaya. [online] Tersedia di : <<https://lampungpro.com/post/2833/promosi-wisata-bandar-lampung-fokus-pada-kuliner-dan-budaya>> [Diakses 27 Agustus 2018]
- LIU, B., 2011. Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data. 2nd ed. Chicago : Springer.
- LIU, X., LI, C., & WU, D., 2018. Film and TV Actors Recommendation Based on SALSALSA Algorithm. *17th International Conference on Computer and Information Science (ICIS)*, pp. 372–376.
- PANDEAN, S. S., & HANSUN, S., 2018. Aplikasi Web untuk Rekomendasi Restoran Menggunakan Weighted Product. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, pp. 87-95.
- PENG, D., LI, T., WANG, Y., & CHEN, C. L. P., 2018. Research on Information Collection Method of Shipping Job Hunting Based on Web Crawler. *Eighth International Conference on Information Science and Technology (ICIST)*, pp. 57–62.
- REN, Y., 2018. A Framework of Petroleum Information Retrieval System Based On Web Scraping With Python. *15th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6.
- SELLAMY, K., FAKHRI, Y., BOULAKNADEL, S., MOUMEN, A., & HAFED, K., 2018. Web Mining Techniques and Applications : Literature Review and a Proposal Approach to Improve Performance of Employment for Young Graduate in Morocco. *Intelligent System and Computer Vision (ISCV)*, pp. 1-5.
- TRAVELLING.BISNIS.COM., 2017. Lampung Berbenah Siapkan Objek Wisata Unggulan. [online] Tersedia di <<http://traveling.bisnis.com/read/20170319/224/638262/lampung-berbenah-siapkan-objek-wisata-unggulan>>
- TRIPADVISOR.COM., 2017. Tentang TripAdvisor. [online] Tersedia di: <<https://tripadvisor.mediaroom.com/id-about-us>> [Diakses 21 Agustus 2018]
- XIA, J., WAN, W., LIU, R., CHEN, G., & FENG, Q., 2015. Distributed Web Crawling: A Framework for Crawling of Micro-Blog Data. *International Conference on Smart and Sustainable City and Big Data (ICSSC)*, pp. 62-68.

Halaman ini sengaja dikosongkan