

BAB II TINJAUAN PUSTAKA

Kebutuhan manusia akan informasi membuat banyaknya penyedia informasi seperti situs berita *online*. Terlalu banyaknya situs berita *online* membuat pengguna kesulitan dalam menyerap informasi yang berbobot. Terkadang situs berita yang satu berbeda dengan yang lainnya tentang apa yang diberitakan. Berbedanya informasi berita di setiap situs membuat pengguna membuka banyak situs berita. Kebiasaan pengguna membuka lebih dari satu situs berita untuk mendapatkan informasinya dapat memakan banyak waktu. Pada penelitian ini, akan dibangun sebuah aplikasi yang digunakan untuk membaca berita dari berbagai sumber. Teknologi *web scraping* digunakan dalam penelitian ini untuk mengambil data dari berbagai situs berita. Selain *web scraping*, aplikasi yang dibangun ini juga bisa melakukan peringkasan berita dengan menggunakan metode TF-IDF.

Bab tinjauan pustaka ini akan menjabarkan beberapa penelitian yang terlebih dahulu dilakukan sebelum pembangunan aplikasi *web scraper* situs berita dan peringkasan berita. Berikut adalah beberapa penelitian lain yang menjadi acuan penulis untuk membangun aplikasi ini.

Penelitian yang dilakukan oleh (Riyadi, 2013) dalam tugas akhirnya yang berjudul "Rancang Bangun *REST Web Service* untuk Perbandingan Harga Pengiriman dengan Metode *Web Scrapping* dan Pemanfaatan API". Dalam penelitian ini dibangun aplikasi untuk membandingkan

harga pengiriman dari TIKI, JNE dan POS Indonesia. Pengguna tidak perlu membuka situs web penyedia layanan pengiriman tersebut satu persatu dan tidak perlu menginputkan data berulang-ulang. Pembangunan aplikasi pada penelitian ini menggunakan teknologi REST (*Representational State Transfer*) *web service*, *web scraping* dan pemanfaatan API (*Application Programming Interface*) ongkir.info. Pemanfaatan API ongkir.info pada penelitian ini digunakan untuk mendapatkan data harga pengiriman barang dan layanan kurir JNE. Teknik *web scraping* pada penelitian ini digunakan untuk mendapatkan harga pengiriman barang pada situs TIKI dan POS Indonesia. Setelah data harga pengiriman barang didapatkan kemudian dimasukkan ke dalam *database*. REST *web service* digunakan untuk mengakses data harga pengiriman barang pada *database*.

(Utomo, 2012) melakukan penelitian dengan judul "Implementasi PHP sebagai Penghasil Konten Otomatis pada Halaman Situs". Dalam penelitian ini dibangun sebuah aplikasi berbasis *web* yang digunakan untuk menghasilkan konten secara otomatis. Sistem secara otomatis menghasilkan konten dan ditampilkan pada halaman *website* berdasarkan pada kata kunci. Sistem juga akan memasukkan hasil yang berupa judul, isi, rangkuman dan kata kunci. Dengan memanfaatkan *Web scraping* aplikasi ini dibuat untuk mengambil konten dari wikipedia, yahoo, dan bing berdasarkan *keyword* yang dimasukkan oleh pengguna aplikasi. Jika artikel ditemukan maka akan tampil pada halaman *web* dan jika tidak ditemukan maka akan menampilkan pesan *not found*. Metode pengambilan konten yang digunakan adalah

fetching dan *cURL*. Menurut penulis agar mendapatkan rangking yang bagus oleh mesin pencari dibutuhkan sebuah konten yang unik, sedangkan dengan teknik pencari konten hasilnya tidak unik.

(Rifangi, 2012) melakukan penelitian dalam tugas akhirnya dengan judul "Rancang Bangun Aplikasi *Dekstop Client* Transkrip Nilai pada *Domain* "krs.amikom.ac.id" dengan Teknik *Web Scraping*". Pada penelitian ini dibangun sebuah aplikasi untuk simulasi transkrip nilai IPK berbasis *dekstop*. Penggunaan *web scraping* pada penelitian ini adalah untuk mengambil data yang ada pada krs.amikom.ac.id. Dengan data tersebut pengguna bisa melakukan akses nilai maupun melakukan simulasi IPK menggunakan aplikasi *dekstop*.

(Josi, et al., 2014) melakukan penelitian dengan judul "Penerapan Teknik *Web Scraping* pada Mesin Pencari Artikel Ilmiah". Penelitian ini dilakukan pada awal Desember 2013 sampai akhir Januari 2014 dengan observasi pada portal jurnal gratis. Penelitian, perancangan, dan pengimplementasian tentang aplikasi pencarian artikel ilmiah dengan bahasa pemrograman PHP dan MySQL sebagai *databasenya*. Proses pencarian artikel dengan menggunakan *web scraping*. *User* memasukkan *keyword* yang akan dicari, jika ada maka akan ditampilkan pada halaman *web* dan jika tidak maka akan muncul pesan *not found*. Teknik *web scraping* adalah teknik yang bermanfaat untuk mendapatkan data dari halaman-halaman *web* secara cepat.

Berdasarkan beberapa penelitian yang sudah dilakukan sebelumnya, penulis dapat menyimpulkan bahwa faktor menggunakan *web scraping* adalah banyaknya situs

yang harus dibuka untuk mendapatkan informasi. Keuntungan menggunakan *web scraping* adalah hemat waktu. Salah satu keuntungan memanfaatkan *web scraping* adalah data atau konten dari beberapa sumber atau *website* dapat dimasukkan dalam satu aplikasi saja.

Penulis mencoba membangun *web scraper* untuk situs berita dengan maksud memudahkan dan menghemat waktu pembaca berita dalam mengakses berita. Aplikasi yang akan dibangun dibagi menjadi dua yaitu aplikasi untuk *administrator* dan aplikasi untuk *user* atau pengguna. Aplikasi *admin* berbasis *website* berfungsi untuk pengaturan *scraping* konten berita berupa URL dan *rules*. Aplikasi untuk *user* berbasis *mobile* berfungsi untuk membaca berita. Selain berfungsi untuk menampilkan berita, aplikasi *mobile* yang dibuat juga bisa meringkas beberapa berita sesuai kata kunci yang dimasukan pengguna. Penulis juga membuat *web service* yang berfungsi untuk mengambil data berita pada *database* yang nantinya ditampilkan pada aplikasi *mobile*. *Database* yang digunakan adalah MySQL.

Tabel 2.1 Tabel Perbandingan Sistem

No	Pembanding	Riyadi (2013)	Utomo (2012)	Rifangi (2012)	Josi, Abdillah, & Suryayusra (2014)	*Penulis (2015)
1	Judul	Rancang Bangun REST <i>Web Service</i> untuk Perbandingan Harga Pengiriman dengan Metode <i>Web Scrapping</i> dan Pemanfaatan API	Implementasi PHP sebagai Penghasil Konten Otomatis pada Halaman Situs	Rancang Bangun Aplikasi Dekstop <i>Client</i> Transkrip Nilai pada Domain "krs.amikom.ac.id"	Penerapan Teknik <i>Web scraping</i> pada Mesin Pencari Artikel Ilmiah	Pembangunan <i>Web Scraper</i> Situs Berita dan Peringkat Berita
2	Tujuan	Membangun REST <i>Web Service</i> untuk perbandingan harga pengiriman barang dengan <i>Web scraping</i> dan Pemanfaatan API	Membangun aplikasi penghasil konten yang dapat membangun konten pada <i>web</i> yang ber CMS <i>wordpress</i> .	Membangun aplikasi dekstop <i>Client</i> transkrip nilai pada <i>domain</i> "krs.amikom.ac.id"	Membangun aplikasi <i>Web scraping</i> pada mesin pencari artikel ilmiah	Membangun aplikasi <i>web scraper</i> situs berita untuk mengambil data berita dan peringkat berita

3	Basis Data	Ya	Ya	Tidak	Ya	Ya
4	Platform	Web	Web	Dekstop	Web	Web dan Moblie
5.	Bahasa Pemrograman	PHP	PHP	Actionscript	PHP	PHP