

Aplikasi Pencarian Produk Jual Mobile Devices dari Berbagai Situs E-commerce

Robby Christianto Wijaya¹, Justinus Andjarwirawan², Henry Novianus Palit³

Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031) – 2983455, Fax. (031) - 8417658

E-mail: robbchrizzz13@gmail.com¹, justin@petra.ac.id², hnpalit@petra.ac.id³

ABSTRAK

Situs-situs *e-commerce* di Indonesia menawarkan banyak produk dengan harga yang berbeda-beda. Konsumen akan kesulitan untuk melakukan perbandingan produk dengan membuka setiap *web* dari *e-commerce* yang ada. Oleh karena itu, dibuat sebuah *website* portal untuk konsumen agar dapat mencari dan melakukan perbandingan produk yang dicari dari setiap *e-commerce* yang ada.

Pengerjaan skripsi ini berfokus pada fitur untuk perbandingan produk antar *e-commerce*. Fitur perbandingan produk memerlukan adanya *server* yang dapat mengambil data produk dari setiap situs *e-commerce* dan *Gsmarena*. Kemudian proses pengambilan dilanjutkan dengan *parsing* dan pengolahan pada setiap data produk *e-commerce* yang sudah didapatkan untuk disesuaikan dengan produk *Gsmarena*.

Berdasarkan pengujian yang sudah dilakukan, waktu pengambilan data produk *e-commerce* dan *Gsmarena* bergantung pada *bandwidth* yang dimiliki *server* dan *response time* yang diberikan oleh masing-masing situs. Produk pada setiap *e-commerce* memiliki nama produk yang berbeda dengan *Gsmarena*, sehingga diperlukan perhitungan untuk menyesuaikan nama produk antara *e-commerce* dan *Gsmarena*.

Kata Kunci: *Web Scraping, DOM Parser, Regex, CURL, Aplikasi Perbandingan*

ABSTRACT

E-commerce sites in Indonesia offers many products with different prices. Consumers will find it hard to do a comparison of products by opening many existing e-commerce web. Therefore, a website portal is created for consumers to find and do a comparison of product from any existing e-commerce.

The research focused on the feature for comparison between e-commerce products. Product comparison feature requires the server can retrieve product data from any e-commerce site and GSMarena. Then, process followed by parsing and processing on any e-commerce product data which has been established to be adapted to the product GSMarena.

Based on the research, the time to get data e-commerce products and GSMarena products relies on the server's bandwidth and response time provided by each site. Products on each e-commerce have different product name from GSMarena, so the calculations is needed for the suitability of the product name between e-commerce and GSMarena.

Keywords: *Web Scraping, DOM Parser, Regex, CURL, Application Comparison*

1. PENDAHULUAN

Perkembangan teknologi informasi dari hari ke hari semakin cepat. Beberapa media yang dapat digunakan untuk mendapatkan informasi adalah koran, radio, televisi, internet, dan lain-lain. Dari berbagai media yang ada, internet memiliki kelebihan untuk menyediakan informasi yang cepat. Konsumen dapat mudah mengoperasikan dan menggunakan internet sepanjang waktu dengan *smartphone, tablet, notebook*, dan komputer yang dimiliki. Kelebihan itu yang menjadi faktor pendorong berkembangnya internet menjadi gaya hidup di seluruh penduduk dunia termasuk Indonesia.

Kemudahan yang ditawarkan layanan internet memberikan dampak positif kepada para pebisnis untuk menawarkan atau menjual produk atau jasa mereka melalui internet. Salah satu manfaat yang diperoleh adalah dapat mengetahui atau mencari berbagai macam produk atau jasa yang ditawarkan dan dibutuhkan. Hal ini yang mendorong banyaknya *e-commerce* bermunculan di Indonesia atau yang biasa disebut *online shop*.

Situs-situs *e-commerce* di Indonesia menawarkan setiap produk dengan harga yang berbeda-beda. Dimana konsumen dapat memilih dan melakukan pengecekan produk yang dicari dengan melihat setiap *e-commerce* yang ada. Konsumen akan kesulitan untuk melakukan perbandingan produk dengan membuka setiap *web* dari *e-commerce* yang ada.

Sebagai solusi untuk mengatasi permasalahan di atas maka dibuat sebuah *website portal* untuk konsumen agar dapat mencari dan melakukan perbandingan produk yang dicari dari setiap *e-commerce* yang ada. Oleh karena itu, skripsi ini mengusulkan aplikasi pencarian yang dapat memberikan informasi produk yang dibutuhkan konsumen dari berbagai *e-commerce*.

2. LANDASAN TEORI

2.1 Web Scraping

Web Scraping [9] adalah adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain. Aplikasi web scraping berfokus pada cara memperoleh data melalui pengambilan dan ekstraksi data dengan ukuran data yang bervariasi. Web scraping memiliki sejumlah langkah meliputi :

- (1) *Create Scraping Template:* Pembuat program mempelajari dokumen HTML dari website yang akan diambil informasinya untuk tag HTML yang mengapit informasi yang akan diambil.
- (2) *Explore Site Navigation:* Pembuat program mempelajari teknik navigasi pada website yang akan diambil

informasinya untuk ditirukan pada aplikasi web scraper yang akan dibuat.

- (3) *Automate Navigation and Extraction*: Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi web scraper dibuat untuk mengotomatisasi pengambilan informasi dari website yang ditentukan.
- (4) *Extracted Data and Package History*: Informasi yang didapat dari langkah 3 disimpan dalam tabel database. Cara kerja web scraping dapat dilihat pada Gambar 1. [2]



Gambar 1. Cara Kerja Web Scraping

2.2 CURL

CURL [5] merupakan sebuah *library* dan baris perintah untuk melakukan transfer data ke dan dari *server*. CURL memungkinkan untuk saling berkomunikasi antar *server* dengan protokol yang berbeda. CURL menyediakan dua macam fungsi yaitu *libcurl* dan *curl*. *Libcurl* merupakan sebuah *library* handal dan *portable* yang memudahkan untuk mengakses ke berbagai protokol yang berbeda, sedangkan *curl* merupakan baris perintah untuk mendapatkan dan mengirimkan data dengan penggunaan sintaks URL.

CURL digunakan untuk metode pengambilan data dari situs *e-commerce*. Instalasi ekstensi *libcurl* diperlukan untuk dapat menggunakan CURL pada pemrograman PHP. Tetapi pada sebagian besar pada instalasi WAMP (Windows Apache, MySQL, PHP) sudah mendukung CURL.

Client dapat memberikan *request* yang dibuat menggunakan HTTP untuk mendapatkan URL yang berisi data dan file. Terdapat beberapa fungsi perintah yang digunakan untuk *client* memberikan *request* seperti yang terlihat pada Gambar 2.

cURL Functions	
• <code>curl_close</code>	— Close a cURL session
• <code>curl_copy_handle</code>	— Copy a cURL handle along with all of its preferences
• <code>curl_errno</code>	— Return the last error number
• <code>curl_error</code>	— Return a string containing the last error for the current session
• <code>curl_escape</code>	— URL encodes the given string
• <code>curl_exec</code>	— Perform a cURL session
• <code>curl_file_create</code>	— Create a cURLFile object
• <code>curl_getinfo</code>	— Get information regarding a specific transfer
• <code>curl_init</code>	— Initialize a cURL session
• <code>curl_multi_add_handle</code>	— Add a normal cURL handle to a cURL multi handle
• <code>curl_multi_close</code>	— Close a set of cURL handles
• <code>curl_multi_exec</code>	— Run the sub-connections of the current cURL handle
• <code>curl_multi_getcontent</code>	— Return the content of a cURL handle if CURLOPT_RETURNTRANSFER is set
• <code>curl_multi_info_read</code>	— Get information about the current transfers

Gambar 2, Beberapa Fungsi CURL

2.3 DOM Parser (Document Object Model Parser)

DOM (*Document Object Model*) [6]-[7] merupakan interface yang menggambarkan XML atau HTML dokumen. DOM didefinisikan sebagai struktur logis dari dokumen dan cara dokumen agar dapat diakses dan dimanipulasi. XML digunakan sebagai cara untuk

mewakili berbagai jenis informasi yang dapat disimpan dalam sistem yang beragam dan menyediakan data sebagai dokumen. DOM dapat digunakan untuk mengelola data XML.

DOM Parser [1] dapat dituliskan dalam bahasa PHP5+ dan bertujuan untuk membantu memanipulasi elemen HTML. Kelebihan DOM Parser adalah dapat mengambil konten HTML dalam satu baris, dapat mencari tag pada HTML dengan selector seperti jQuery, dan memanipulasi elemen HTML yang invalid. Contoh beberapa fungsi DOM Parser yang dapat dilihat pada Gambar 3.

Helper functions	
Name	Description
object <code>str_get_html (string \$content)</code>	Creates a DOM object from a string.
object <code>file_get_html (string \$filename)</code>	Creates a DOM object from a file or a URL.

DOM methods & properties	
Name	Description
void <code>__construct ([string \$filename])</code>	Constructor, set the filename parameter will automatically load the contents, either text or file/url.
string <code>plaintext</code>	Returns the contents extracted from HTML.
void <code>clear ()</code>	Clean up memory.
void <code>load (string \$content)</code>	Load contents from a string.
string <code>save ([string \$filename])</code>	Dumps the internal DOM tree back into a string. If the \$filename is set, result string will save to file.
void <code>load_file (string \$filename)</code>	Load contents from a file or a URL.
void <code>set_callback (string \$function_name)</code>	Set a callback function.
mixed <code>find (string \$selector [, int \$index])</code>	Find elements by the CSS selector. Returns the nth element object if \$index is set, otherwise return an array of object.

Gambar 3. Beberapa Fungsi DOM Parser

2.4 Regex (Regular Expression)

Regular expression [3] merupakan sebuah pola yang menggambarkan dan menyamakan sejumlah teks. Penggunaan Regex dapat digunakan untuk menemukan, mengganti, dan membagi teks sesuai dengan pola tertentu. Selain itu, dapat digunakan untuk memverifikasi *input* data dan sejumlah teks yang terlihat seperti *email*, nomor telepon, tanggal, kode negara, dan lainnya. Regex terdiri dari *string* yang merupakan kombinasi antara karakter normal dan metakarakter khusus. Dalam implementasi dari kombinasi ini, karakter normal mewakili karakter itu sendiri. Sedangkan metakarakter khusus adalah karakter yang merepresentasikan maksud tertentu.

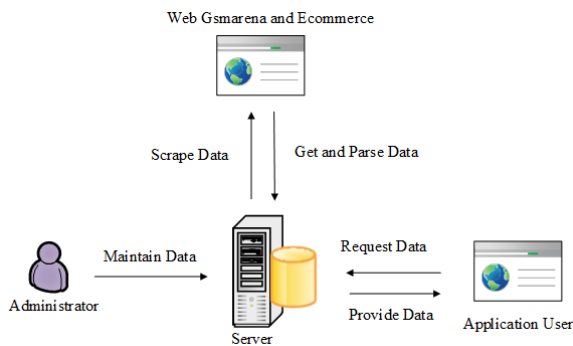
Regex merupakan bagian dari bahasa pemrograman dan *library* yang terus berkembang. Bahasa pemrograman yang dapat menggunakan regex yaitu .NET, Java, JavaScript, Perl, PHP, Python, Java, Javascript, Ruby, dan lainnya. Selain itu dapat digunakan pada Unix melalui *utility ed, sed, vi editors, grep, AWK*, dan lainnya.

Contoh pola menggambarkan alamat *email* yaitu `"\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,4}\b"`. Penggunaan pola tersebut merupakan pola yang kompleks. Pola menggambarkan kumpulan yang terdiri dari huruf, angka, titik, garis bawah, simbol persen, dan tanda hubung. Kemudian diikuti oleh sebuah titik tunggal di antara 2-4 huruf. [4]

3. DESAIN SISTEM DAN ANALISA

3.1 Grafik Hirarki Aplikasi

Pada Gambar 4 menjelaskan desain sistem secara keseluruhan aplikasi PriceMPT. *Server* melakukan *scraping* pada situs Gsmarena dan *e-commerce*. *Server* akan memilah data yang didapatkan. *Admin* melakukan pengelolaan data pada *database* berupa penambahan data, *pengeditan* data, dan penghapusan data. Aplikasi *user* dapat melakukan *request data* pada *database server* untuk mendapatkan data.



Gambar 4. Desain Sistem Secara Keseluruhan

Aplikasi server dibagi menjadi 2 bagian, yaitu proses *data scraping product Gsmarena* dan proses *data scraping product e-commerce*. Dua proses memiliki fungsi yang sama untuk memasukkan produk baru dan melakukan perbaruan data produk. Pada setiap situs Gsmarena dan *e-commerce* memiliki struktur HTML yang berbeda-beda. Oleh karena itu, diperlukan membaca dan mengerti struktur HTML setiap situs untuk memilih data mana yang akan diambil.

Proses *data scraping product Gsmarena* digunakan untuk mengambil info secara umum dan spesifikasi produk pada *link Gsmarena*. Proses *data scraping product e-commerce* digunakan untuk mengambil info secara umum dan detail produk *mobile devices* pada setiap situs *e-commerce* yaitu Lazada, Dinomarket, Blibli, Matahari Mall, dan Erafone. Produk *e-commerce* yang masuk akan dihitung perhitungan relevansi kemiripan dengan produk Gsmarena yang disebut dengan perhitungan *mapping product*.

Data yang telah diambil akan disimpan pada *database* dan akan dilakukan *update* secara berkala yang telah ditentukan oleh *admin*. Proses *update* dijalankan secara berkala dengan menggunakan *scheduler cronab* pada *server*.

3.2 Mapping Product

Pada proses *insert product and price* terdapat proses *mapping product* untuk menghitung kesamaan antara *title* pada setiap *e-commerce* dan Gsmarena pada awal proses *insert product and price*. *Admin* telah menentukan jumlah persentase relevan pada awal proses *insert product and price*. Proses ini akan melakukan pengecekan pada keseluruhan data produk. Umumnya penulisan *title mobile devices* pada setiap *e-commerce* lebih panjang daripada *title mobile devices* pada Gsmarena. Contoh penulisan *title mobile devices* pada *e-commerce* yaitu “Samsung Galaxy J5 - SM-J500 - 8GB - Putih” dan penulisan *title mobile devices* pada Gsmarena yaitu “Samsung Galaxy J5”. Oleh karena itu, dilakukan perbandingan dengan mencocokkan setiap kata pada *title e-commerce* apakah sama dengan setiap kata pada *title Gsmarena*. Terdapat dua perhitungan persentase meliputi:

- **Title Produk Terkandung**

Pada produk *title Gsmarena* yang secara keseluruhan terkandung dalam *title e-commerce* memiliki nilai yang lebih dan relevan yang lebih tinggi daripada *title Gsmarena* yang secara keseluruhan tidak terkandung. Perhitungan persentase dilakukan dengan membagi *counter* jumlah kata yang sama dengan jumlah kata yang dimiliki *title* produk Gsmarena dan ditambahkan dengan nilai bobot (yang bergantung pada panjang karakter *title Gsmarena*). Penambahan bobot ini mengakibatkan jumlah persentase yang melebihi dari seratus persen. Rumus perhitungan dapat dilihat sebagai berikut:

$$\text{similarPercent} = \frac{\text{swF}}{\text{gwF}} \times \left(1 + \frac{\text{gcF}}{\text{ecF}}\right) \times 100 \quad (1)$$

Keterangan :

swF = jumlah kata sama

gwF = jumlah kata pada *title Gsmarena*

gcF = jumlah karakter *title Gsmarena*

ecF = jumlah karakter *title e-commerce*

- **Title Produk Tidak Terkandung**

Perhitungan persentase dilakukan dengan membagi *counter* jumlah kata yang sama dengan jumlah kata yang dimiliki *title* produk Gsmarena. Rumus perhitungan dapat dilihat sebagai berikut:

$$\text{similarPercent} = \frac{\text{swF}}{\text{gwF}} \times 100 \quad (2)$$

Keterangan :

swF = jumlah kata sama

gwF = jumlah kata pada *title Gsmarena*

4. IMPLEMENTASI SISTEM

Aplikasi dibuat dengan menggunakan bahasa HTML5, PHP dan Javascript. *Database server* menggunakan MySQL.[8]

4.1 Proses Data Scraping Product Gsmarena

Proses ini berguna untuk mengambil data produk yang disediakan oleh situs-situs setiap *brand Gsmarena*. *Grabber* akan mengambil *link* yang sudah tersimpan pada tabel *brand* di *database* seperti yang terlihat pada Gambar 5.

id	nameBrand	href
1	Samsung	http://www.gsmarena.com/samsung-phones-9.php
2	Apple	http://www.gsmarena.com/apple-phones-48.php
5	LG	http://www.gsmarena.com/lg-phones-20.php
6	Xiaomi	http://www.gsmarena.com/xiaomi-phones-80.php
7	Sony	http://www.gsmarena.com/sony-phones-7.php
8	Asus	http://www.gsmarena.com/asus-phones-46.php

Gambar 5. Brand Link yang sudah Disimpan dalam Database

Scraping akan mengunduh data umum produk yang terdapat pada halaman utama dan *pagination* berikutnya. Setelah berhasil memasukkan data umum produk, maka proses selanjutnya adalah mengunduh data spesifikasi produk yang terdapat pada setiap *link* produk Gsmarena. Data spesifikasi produk dilakukan pemecahan untuk memudahkan dalam fitur *filter* pencarian.

Scraping akan berjalan sesuai dengan *scheduler* yang sudah diatur pada *server*. Produk yang sudah dilakukan *scraping* akan disimpan pada tabel *product* dan spesifikasi produk disimpan pada tabel *specification* di *database*. Dapat dilihat pada Gambar 6 spesifikasi produk

id	title	label	value
1	Network	Technology	GSM/HSPA/LTE
2	Network	2G bands	GSM 850/900/1800/1900
3	Network	3G bands	HSDPA 850/900/1900/2100
4	Network	4G bands	LTE
5	Network	Speed	HSPA 42.2/5.76 Mbps
6	Network	Speed	LTE Cat6 300/50 Mbps
7	Network	GPRS	Yes
8	Network	EDGE	Yes
9	Launch	Announced	2015
10	Launch	Announced	October
11	Launch	Status	Coming soon
12	Launch	Status	Exp
13	Launch	Status	release 2015, November 27
14	Body	Dimensions	451.8 x 275.8 x 11.9 mm (17.79 x 10.86 x 0.47 in)
15	Body	Weight	2.65 kg (5.84 lb)
16	Body	SIM	Nano-SIM

Gambar 6. Data Spesifikasi Produk

4.2 Proses Data Scraping Product Gsmarena

Proses ini berguna untuk mengambil data produk yang disediakan oleh situs-situs setiap *e-commerce*. *Scraping* mengambil *link* yang sudah tersimpan pada tabel *ecommerce* di *database* seperti yang terlihat pada Gambar 7.

id	nameEcommerce	href
1	Lazada	http://www.lazada.co.id/beli-handphone/
2	Blibli	https://www.blibli.com/handphone/54593
3	Erafone	http://erafone.com/beli-handphone
5	MatahariMall	https://www.mataharimall.com/p-2/handphone
6	Dinomarket	http://www.dinomarket.com/modules/vbrand/vbrand010...
8	Erafone Tablet	https://erafone.com/beli-tablet
9	Dinomarket Tablet	http://www.dinomarket.com/modules/vbrand/vbrand010...
10	Lazada Tablet	http://www.lazada.co.id/beli-tablet-2/
11	MatahariMall Tablet	https://www.mataharimall.com/p-5/tablet
12	Blibli Tablet	https://www.blibli.com/tablet/54593

Gambar 7. *E-commerce Link* yang sudah Disimpan dalam *Database*

Scraping akan mengunduh data umum produk *e-commerce* yang terdapat pada halaman utama dan *pagination* berikutnya yang tersedia. Setelah berhasil memasukkan data umum produk, maka proses selanjutnya adalah mengunduh data detail produk yang terdapat pada *link* setiap produk *e-commerce*. Data detail produk yang diambil yaitu harga, nilai *rating*, nama toko penjual, dan ketersediaan produk

Scraping akan berjalan sesuai dengan *scheduler* yang sudah diatur pada *server* untuk memasukkan produk *e-commerce* baru dalam setiap *e-commerce*. Produk *e-commerce* dan detail produk yang sudah dilakukan *scraping* akan disimpan pada tabel *product_ecommerce* di *database*. Dapat dilihat pada Gambar 8 data produk *e-commerce*.

nameProduct	href	src	availability	lmt_productecommerce	finalPrice
Oppo Neo 5 R831L - 4GB - Hitam	http://www.lazada.co.id/oppo-neo-5-r831l-4gb-hitam...	http://id-live-03.slatiic.net/p/oppo-neo-5-r831l-4g...	0	2015-12-02 15:16:31	0
OPPO Neo 5 4 GB - Putih	http://www.lazada.co.id/oppo-neo-5-4-gb-putih-1816...	http://id-live-02.slatiic.net/p/oppo-neo-5-4-gb-put...	1	2015-12-02 15:16:31	1925000
Oppo Joy 3 A11w - 4GB - Abu-abu	http://www.lazada.co.id/oppo-joy-3-a11w-4gb-abu-ab...	http://id-live-01.slatiic.net/p/oppo-joy-3-a11w-4gb...	1	2015-12-06 16:52:51	1499000
Oppo Joy 3 A11w - 4GB - Putih	http://www.lazada.co.id/oppo-joy-3-a11w-4gb-putih-...	http://id-live-02.slatiic.net/p/oppo-joy-3-a11w-4gb...	1	2015-12-06 16:52:51	1499000
Oppo Neo 5 R1201 - 4GB - Hitam	http://www.lazada.co.id/oppo-neo-5-r1201-4gb-hitam...	http://id-live-01.slatiic.net/p/oppo-neo-5-r1201-4g...	1	2015-12-02 15:16:31	1999000

Gambar 8. Data Produk *E-commerce*

4.3 Aplikasi Website

User dapat melakukan pencarian produk dari kotak pencarian yang disediakan. Aplikasi PriceMPT menyediakan fitur *autocomplete* untuk memudahkan *user* dalam memilih produk seperti yang terlihat pada Gambar 9. hasil pencarian produk terdapat harga terendah dan harga tertinggi seperti pada Gambar 10.



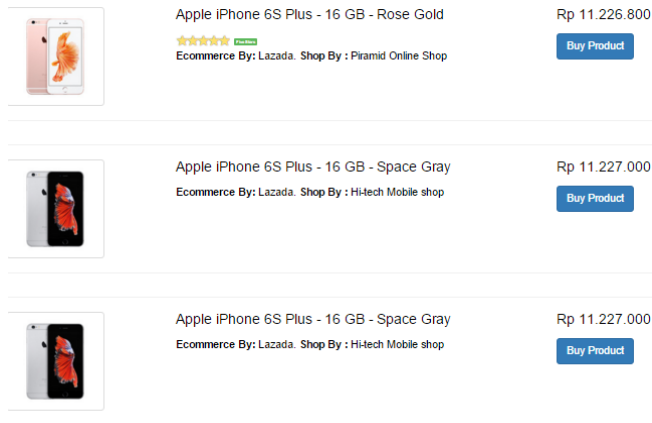
Gambar 9. *Autocomplete Search*



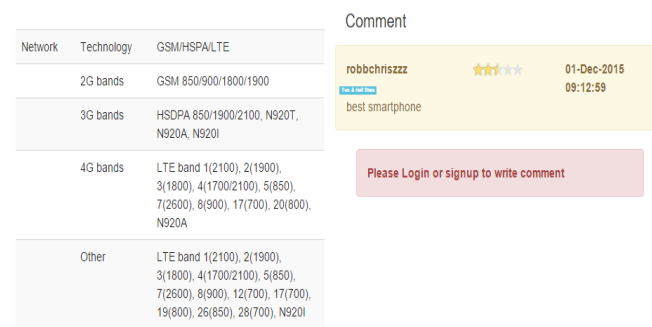
Gambar 10. Hasil Pencarian Produk

Tampilan untuk setiap produk menunjukkan nama produk, gambar produk, kisaran harga, *e-commerce* yang menjual, spesifikasi produk, komentar, dan *rating member*. *User* dapat melihat daftar *e-commerce* yang menjual produk dan dapat melakukan *filter* berdasarkan nama *e-commerce* yang menjual dan melakukan urutan berdasarkan harga termurah atau termahal seperti yang terdapat pada **Error! Reference source not found.** *User* juga

dapat melihat spesifikasi yang dimiliki oleh produk dan komentar yang diberikan oleh *member* seperti pada Gambar 12.



Gambar 11. Daftar E-commerce yang Menjual Produk



Gambar 12. Spesifikasi Produk dan Komentar Member

User dapat melakukan filter pencarian berdasarkan spesifikasi produk dan *brand* dari produk. Spesifikasi meliputi ukuran layar, berat produk, kamera utama, memori internal, memori RAM, ukuran baterai, dan sistem operasi. User dapat mencari produk sesuai dengan keinginan spesifikasi yang diinginkan seperti pada Gambar 14.

User dapat melakukan perbandingan spesifikasi produk dengan produk yang lain. Terdapat 3 kotak pencarian produk untuk user melakukan pencarian produk yang diinginkan seperti yang terlihat pada Gambar 15.

4.4 PENGUJIAN WAKTU

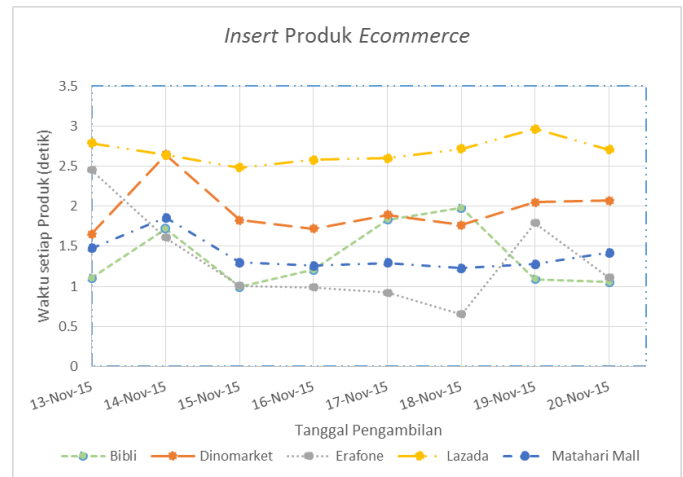
Bagian ini menunjukkan lama waktu yang dilakukan untuk proses pengambilan data produk *e-commerce*. Pengujian dilakukan secara *scheduler* dengan menggunakan *crontab* pada *server* Universitas Kristen Petra. *Server* Universitas Kristen Petra adalah *opensource.petra.ac.id* dengan kecepatan internet yang dimiliki yaitu *bandwidth* 1.1 Gbps (600 Mbps dan 500 Mbps dari *provider* yang berbeda). Pengujian dilakukan dengan jumlah waktu yang berbeda-beda untuk mengetahui waktu pengambilan data.

Lama waktu pengambilan dalam pengambilan produk bergantung pada *bandwidth* yang dimiliki oleh *server* dan waktu respon dari setiap *e-commerce* dalam memberikan dataproduk. *Response time* yang diberikan dari setiap *e-commerce* berbeda-beda. Gambar 13 merupakan grafik yang menggambarkan tanggal melakukan pengujian dan waktu pengambilan yang didapatkan untuk memproses 1 produk. *Response time* pada *e-commerce* Lazada lebih lama dibandingkan *e-commerce* yang lain, karena dapat dipengaruhi oleh beberapa faktor seperti jumlah pengguna yang

melakukan akses secara bersamaan, jumlah produk yang lebih banyak, *bandwidth* yang dibatasi oleh *e-commerce*, konfigurasi *web server*, dan *web hosting* yang digunakan.

Tabel 1. Lama Pengambilan Waktu setiap Produk E-commerce

Tanggal	Bibli (detik)	Dinomarket (detik)	Erafone (detik)	Lazada (detik)	Matahari Mall (detik)
13-Nov-15	1.107	1.656	2.452	2.792	1.478
14-Nov-15	1.715	2.646	1.61	2.644	1.86
15-Nov-15	0.991	1.827	1.009	2.482	1.298
16-Nov-15	1.208	1.719	0.987	2.581	1.26
17-Nov-15	1.833	1.895	0.919	2.6	1.293
18-Nov-15	1.981	1.767	0.653	2.717	1.227
19-Nov-15	1.091	2.053	1.794	2.966	1.279
20-Nov-15	1.056	2.072	1.108	2.707	1.422

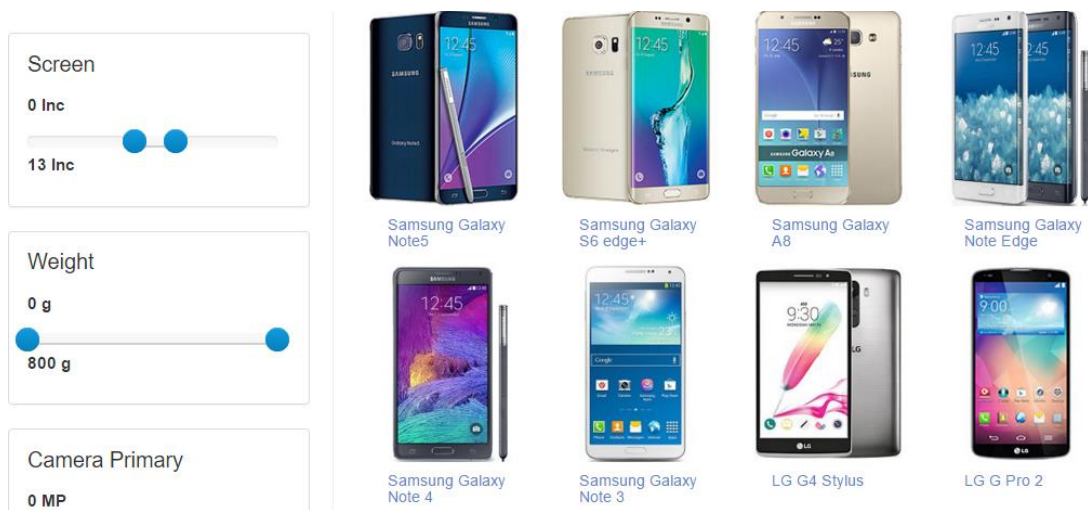


Gambar 13. Grafik Pengujian Waktu E-commerce

5. KESIMPULAN DAN SARAN

Dari proses perancangan, pembuatan, dan hasil pengujian aplikasi pencarian produk jual *mobile devices* dari berbagai situs *e-commerce*, dapat diambil kesimpulan sebagai berikut:

- Penggunaan metode *scraping* untuk mengambil data produk dari situs *Gsmarena* dan *e-commerce*.
- Hasil waktu pengambilan setiap produk *e-commerce* bergantung pada *bandwidth* yang dimiliki oleh *server* dan *response time* yang diberikan oleh masing-masing *e-commerce*.
- Situs *e-commerce* memiliki format struktur HTML yang berbeda-beda. Hal ini menyebabkan adanya program *Parser* yang dibuat secara spesifik untuk masing-masing *e-commerce* dalam pengambilan produk.
- Situs *e-commerce* yang mengalami *maintenance* mengakibatkan data produk tidak dapat diambil secara keseluruhan, tetapi untuk pengambilan data produk di hari esok dapat dilakukan kembali dan terdapat produk baru.
- Situs *e-commerce* memiliki produk baru untuk setiap harinya, sehingga terdapat produk baru yang masuk.



Gambar 14. Filter Pencarian

Samsung Galaxy Note5



Samsung Galaxy Note 4



Apple iPhone 6s



Network	Technology	GSM/HSPA/LTE
2G bands	GSM	850/900/1800/1900
3G bands	HSDPA	850/1900/2100, N920T, N920A, N920I
4G bands	LTE band	1(2100), 2(1900), 3(1800), 4(1700/2100), 5(850), 7(2600), 8(900), 17(700), 20(800), N920A
Other	LTE band	1(2100), 2(1900), 3(1800), 4(1700/2100)

Network	Technology	GSM/HSPA/LTE
2G bands	GSM	850/900/1800/1900
3G bands	HSDPA	850/900/1900/2100
Other	HSDPA	850/1700/1900/2100, N910W8
4G bands	LTE band	1(2100), 2(1900), 3(1800), 4(1700/2100), 5(850), 7(2600), 8(900), 17(700), 20(800), N920A

Network	Technology	GSM/CDMA/HSPA/EVDO/LTE
2G bands	GSM	850/900/1800/1900
Other	CDMA	800/1700/1900/2100
3G bands	HSDPA	850/900/1700/1900/2100
Other	CDMA2000	1xEV-DO
4G bands	LTE band	1(2100), 2(1900), 3(1800), 4(1700/2100), 5(850), 7(2600), 8(900)

Gambar 15. Perbandingan Spesifikasi Produk

- Hasil dari perhitungan *mapping product* antara produk Gsmarena dan produk *e-commerce* secara keseluruhan terdapat beberapa produk yang tidak sesuai dengan produk yang dimaksud. Penyebabnya adalah penulisan title dari setiap *e-commerce* dengan title pada Gsmarena berbeda-beda. Terdapat juga penulisan title produk yang salah dan salah masuk kategori produk dari situs *e-commerce*. Hal ini mengakibatkan diperlukan keputusan dari admin untuk memutuskan relevansi dari beberapa produk dan bantuan dari *member* untuk memberikan report produk apabila tidak sesuai.

6. DAFTAR PUSTAKA

- [1] Chen, S.C. 2015. *PHP Simple HTML DOM Parser*. URI=<http://simplehtmldom.sourceforge.net/>.
- [2] Computer Advisor. 2015. *Web Site Scraper The Most Effective Tool for Web Data Extraction*. URI=<http://www.thecomputeradvisor.net/web-site-scraper-the-most-effective-tool-for-web-data-extraction/>
- [3] Fitzgerald, M. 2012. *Introducing Regular Expressions*. Sebastopol: O'Reilly Media, Inc.
- [4] Goyvaerts, J. 2015. *Regular Expressions Tutorial Learn How to Use and Get The Most Out of Regular Expressions*. URI=<http://www.regular-expressions.info/tutorial.html>.
- [5] Khalid, M. I. 2006. *PHP/CURL Book with Examples Version 1.8*. Pakistan: Imran & LAMP Technologies.
- [6] Lindley, C. 2013. *DOM Enlightenment*. Sebastopol: O'Reilly Media, Inc.
- [7] Marini, J. 2002. *The Document Object Model: Processing Structured Documents*. New York: The McGraw-Hill Companies, Inc.
- [8] Nixon, R. 2014. *Learning PHP, MySQL, Javascript, CSS&HTML5* (3rd ed.). Sebastopol: O'Reilly Media, Inc.
- [9] Turland, M. 2010. *PHP/Architect's Guide to Web Scraping with PHP*. Canada: Marco Tabini & Associates, Inc.