

## IMPLEMENTASI *WEBCRAWLER* PADA SOCIAL MEDIA MONITORING

Dewi Rosmala<sup>[1]</sup>, Rizqi Rivani Syafei<sup>[2]</sup>

Jurusan Teknik Informatika  
Institut Teknologi Nasional Bandung

### ABSTRAK

*Social media monitoring adalah proses menghimpun, memahami, dan merespon opini tentang merek, produk, reputasi, atau opini user di social media. Hal tersebut dilakukan untuk menjaga brand image dari produk itu sendiri. Brand image dibangun guna menciptakan loyalitas konsumen terhadap suatu produk karena dengan membangun brand image berarti membangun dan menjaga keuntungan (laba investasi) serta kelangsungan hidup perusahaan. Caranya, dengan menjelajahi web dengan memanfaatkan web crawler untuk mencari aktivitas dan pembicaraan yang sedang terjadi dan menentukan bagaimana cara yang tepat untuk mempengaruhi dan membentuk opini di social media. Web crawler merupakan sebuah perangkat lunak yang digunakan untuk menjelajahi halaman di internet dan akan mengambil informasi yang tersedia. Atas dasar tersebut, penulis akan membangun aplikasi social media monitoring dengan memanfaatkan web crawler. Aplikasi ini memungkinkan para pelaku bisnis untuk memantau opini masyarakat terhadap suatu produk. Dengan adanya aplikasi ini diharapkan mampu membantu pengguna dalam menjaga brand image. Selain itu diharapkan dengan memanfaatkan aplikasi ini tidak terbatas pada memantau saja, melainkan social media monitoring pun bisa menjadi sumber key learning bagaimana menciptakan strategi promosi yang sukses.*

**Kata Kunci :** *Social Media Monitoring, Brand Image, Web Crawler*

### ABSTRACT

*Social media monitoring is a process to collect, understand, and respond to opinions about brands, products, reputation, or opinion on social media. This is done to maintain the brand image of the product itself. Brand image was built in order to create consumer loyalty to a product, because by establishing a brand image means built and keep the profit (return on investment) as well as the company's survival. By surfing the web using a web crawler to find activities and conversations that are happening and determine what the proper way to influence and form an opinion in social media. Web crawler is a software which is used to explore the pages on the internet and will take the an available information. Based on that case, was built a social media monitoring application by take an advantage of web crawler. This application allow business people to monitor public opinions towards a product. With this application should be able to assist business people in maintaining a brand image. In addition, are expected by take an advantage of this application is not limited to monitoring only, but social media monitoring can be a source of key learning how to create a successful campaign strategy.*

**Keyword :** *Social Media Monitoring, Brand Image, Web Crawler*

## **PENDAHULUAN**

### **Latar Belakang Masalah**

*Social Media* kini sudah menjadi bagian kehidupan sehari-hari bagi pengguna internet di Indonesia. Dengan 43,6 juta jumlah pengguna *Facebook* dan 19,5 juta jumlah pengguna *Twitter*, Indonesia merupakan negara dengan tingkat penggunaan *social media* yang besar di dunia pada tahun 2011 [Dwiki Setiyawan, 2012]. Dari data dan fakta yang ada dapat disimpulkan kesempatan yang dapat diambil dari fenomena tersebut bahwa pangsa pasar di Indonesia sangat menggiurkan. Dengan jumlah pengguna *social media* di Indonesia yang diperkirakan akan mencapai angka 100 juta orang pengguna pada tahun 2014 berbagai produsen berlomba-lomba menggarap potensi ekonomi sebesar ini [Telkomsolution, 2011]. Tidak sedikit pemasar yang memanfaatkan *social media* untuk mengencarkan kampanye produk berupa barang atau jasa, atau brand mereka. [Rahadian P. Paramita, 2012].

Pemanfaatan tersebut dilakukan karena *social media* dapat membuat masyarakat berbicara apa saja tentang suatu *brand* dengan mudah. Dengan penyebaran yang sangat cepat, pengguna-pengguna *social media* dapat dengan mudah menerima opini ataupun ekspresi terhadap suatu produk. Sehingga akan terjadi satu kesamaan pemikiran antar pengguna dan terjadi sebuah gerakan sosial atau *social movement* yang dapat mempengaruhi sebuah pemikiran secara luas dan tidak terkendali. Jika tidak diawasi, maka opini tersebut dapat dengan mudah menghancurkan suatu *brand image* dalam sekejap. Contohnya seperti banyak keluhan konsumen yang di sampaikan melalui *social media*, jika tidak ditanggapi secara tepat maka akan menimbulkan kerugian bagi *brand image* itu sendiri. Maka diperlukan data dan informasi mengenai

kondisi *brand image* dari sebuah produk yang sedang terjadi.

Dalam menjaga *brand image* untuk kepentingan bisnis, para pelaku pasar memanfaatkan data dan informasi yang tersebar di internet dengan menggunakan *search engine*. Beberapa *search engine* seperti Google, MSN, Bing, Yahoo Search memanfaatkan *web crawler* sebagai *engine* untuk memperoleh informasi. *Web crawler* merupakan sebuah perangkat lunak yang digunakan untuk menjelajahi halaman internet dan akan mengambil informasi yang tersedia. Informasi tersebut berupa halaman-halaman *web* yang diambil secara otomatis lalu isi setiap halaman tersebut diuraikan berdasarkan data yang dibutuhkan misalnya mengambil data teks dan gambar. Data dari halaman *web* tersebut disimpan kedalam sebuah database untuk digunakan selanjutnya.

Pada tugas akhir ini penulis akan membangun sebuah aplikasi *Social Media Monitoring*, aplikasi ini merupakan *tools* untuk menghimpun opini tentang merek, produk, reputasi, atau opini masyarakat di *social media*. Caranya, dengan menjelajahi *web* dengan memanfaatkan *web crawler* untuk mencari aktivitas dan pembicaraan yang sedang terjadi dan menentukan bagaimana cara yang tepat untuk mempengaruhi dan membentuk opini di *social media*.

### **Rumusan Masalah**

Berdasarkan penjelasan latar belakang diatas, maka dapat di rumuskan masalah yang timbul yaitu, bagaimana cara memantau *brand image* dari sebuah produk melalui *social media* dengan menggunakan teknologi *webcrawler* ?

### **Tujuan**

Tujuan yang ingin dicapai pada Tugas Akhir ini adalah membangun *social media monitoring* dengan memanfaatkan *web*

*crawler* sebagai *engine* untuk menghimpun data dan informasi pada *social media* lalu direpresentasikan kedalam sebuah *web*.

## Batasan Masalah

Dari penelitian yang dilakukan, penulis membatasi ruang lingkup yang dibahas pada Tugas Akhir ini, yaitu sebagai berikut :

1. Aplikasi ini tidak membahas mengenai pengolahan kalimat semantik atau *Natural Language Processing* sebagai *Artificial Intelligence*.
2. Media sosial yang menjadi bahan penelitian adalah *Twitter*.
3. Data yang dipergunakan dalam implementasi aplikasi ini terbatas pada data akun *Twitter* yang ada. Tidak termasuk usia, gender, dan lokasi.
4. *Output* yang dihasilkan adalah tabel hasil *crawling* berdasarkan *keyword* yang telah ditentukan sebelumnya.

## Tinjauan Pustaka

Pawestri Dwi Utami, Royyana Muslim I, dan Henning T.C (2009) dalam penelitiannya yang berjudul “**Perancangan Dan Pembuatan Web Crawler Aplikasi Panduan Pembelian Spesifikasi Komputer Memanfaatkan Google Gears**” menyimpulkan hasil dari penelitiannya bahwa, penggunaan *crawling* dalam aplikasi *web crawler* ini merupakan solusi yang efisien dalam panduan pencarian spesifikasi komputer rakitan online. Hal tersebut dikarenakan proses *insert* serta *update* barang tidak dilakukan secara manual, namun dilakukan otomatis oleh *crawling*. Berjalannya proses *crawling* di *background process* tidak mengganggu kinerja user ketika mengakses aplikasi *web*. Konsep yang diambil dari penelitian ini adalah konsep *background process* dimana penelitian yang akan diimplementasikan nantinya akan menggunakan *daemon thread* dalam bahasa Java. *Daemon thread* merupakan *thread* yang tidak akan selesai

(*shutdown*) bila semua proses pada JVM telah selesai.

Ardy Purnama P, Pennyka Trifikta R dan Rizky (2011) dalam penelitiannya yang berjudul “**Sistem Monitoring Tugas Akhir**” mengemukakan bahwa, aplikasi akandiimplementasikan berbasis *web* sehingga mahasiswa mampu mengajukan judul dan perkembangan tugas akhirnya melalui system ini tanpa bertatap muka dengan dosen pembimbing. Dilain pihak, secara otomatis dokumentasi dari aktivitas user akan tersimpan dalam database *admin*. Hal yang diambil dari penelitian ini adalah konsep pada dokumentasi *online* sehingga tidak diperlukan lagi tatap muka dalam pencarian sebuah informasi yang sifatnya penting. Konsep tersebut diimplementasikan pada pencarian sebuah data berbentuk opini tanpa disadari oleh responden pada *social media* *Twitter*.

Fitra Razab B dan Dewi Rosmala (2011) dalam penelitiannya yang berjudul “**Implementasi Gravitation Based Model Dalam Sistem Temu Balik Informasi (STBI) Search Engine**” mengemukakan bahwa pemodelan sistem temu balik informasi terbagi kedalam dua bagian utama proses, yaitu *indexing* dan *searching*. *Indexing* adalah proses yang merepresentasikan koleksi data dalam bentuk index sehingga memudahkan dan mempercepat proses pencarian data. Sedangkan *searching (matching)* merupakan proses menemukan kembali data yang relevan dengan query. Konsep yang diambil dari penelitian ini adalah proses *searching* dan *indexing* dimana pada implementasi penelitian ini *query* akan terbentuk dari metoda *searching* serta *indexing* untuk memudahkan pencarian kata.

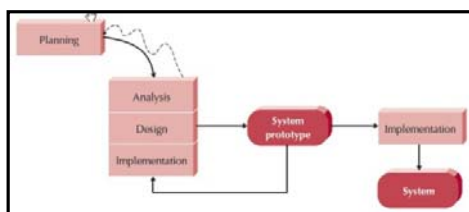
## Metodologi Penelitian

Metode-metode yang digunakan dalam tugas akhir ini adalah sebagai berikut:

1. Studi pustaka, yaitudilakukan dengan cara membaca berbagai referensi buku,

literatur, diktat, dan catatan perkuliahan maupun catatan lain yang berhubungan dengan permasalahan yang sedang diteliti, serta mencari data yang berhubungan di internet. Penelitian ini dimaksudkan untuk memperoleh data, teori dan konsep sehingga penulis dapat memperoleh informasi yang lebih luas yang berhubungan dengan pokok permasalahan.

2. Metodologi pengembangan sistem yang digunakan adalah *prototype*.



Gambar 1 Life Cycle Prototype Model

Dari Gambar 1 terlihat fase-fase yang terdapat pada kerangka kerja *prototype* dari setiap fase akan terhubung satu sama lainnya dan tidak terpisah. Pada subbab selanjutnya akan dijelaskan secara spesifik apa saja yang di lakukan.

## LANDASAN TEORI

### Web Crawler<sup>[1]</sup>

*Web crawler* adalah sebuah perangkat lunak yang digunakan untuk menjelajah serta mengumpulkan halaman-halaman *web* yang selanjutnya diindeks oleh mesin pencari (Gatjal E, 2005). Sering juga disebut dengan *web spider* atau *web robot*. Ide dasarnya sederhana dan mirip dengan saat menjelajahi halaman *website* dengan menggunakan *browser*. Bermula pada point awal berupa sebuah link alamat *website* dan dibuka pada *browser*, lalu *browser* melakukan *request* dan men-download data dari *webserver* melalui protokol HTTP. Setiap *hyperlink* yang ditemui pada konten yang tampil akan dibuka lagi pada *windows/tab browser* yang baru, demikian proses terus berulang. Sebuah *web crawler* mengotomatiskan pekerjaan itu.

## Social Media<sup>[2]</sup>

*Social Media* adalah media online yang mendukung interaksi sosial. *Social media* menggunakan teknologi berbasis *web* yang mengubah komunikasi menjadi dialog interaktif. Andreas Kaplan dan Michael Haenlein mendefinisikan media sosial sebagai "sebuah kelompok aplikasi berbasis internet yang membangun di atas dasar ideologi dan teknologi *Web 2.0*, dan yang memungkinkan penciptaan dan pertukaran *user-generated content*".

## Monitoring<sup>[3]</sup>

Monitoring (bahasa Indonesia: pemantauan) adalah pemantauan yang dapat dijelaskan sebagai kesadaran (*awareness*) tentang apa yang ingin diketahui, pemantauan berkadar tingkat tinggi dilakukan agar dapat membuat pengukuran melalui waktu yang menunjukkan pergerakan ke arah tujuan atau menjauh dari itu.

## Twitter<sup>[4]</sup>

Twitter adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh *Twitter Inc.*, yang menawarkan jejaring sosial berupa mikroblog sehingga memungkinkan penggunaanya untuk mengirim dan membaca pesan yang disebut kicauan (*tweets*). *Tweets* adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna.

## ANALISIS DAN PERANCANGAN

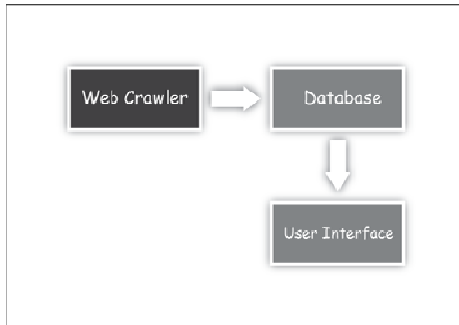
### Identifikasi Sistem

Identifikasi sistem yang dilakukan adalah bagaimana mengimplementasikan sebuah *web crawler* yang dapat dimanfaatkan untuk menghimpun opini masyarakat pada media sosial khususnya *Twitter*. Dengan memanfaatkan *Twitter Search API* sebagai input data yang diolah lebih lanjut dengan menggunakan *parser*, sehingga data dapat disimpan ke dalam database dan direpresentasikan sebagai output akhir yang dapat ditampilkan ke dalam aplikasi *web* sebagai informasi.



**Analisa Kebutuhan Sistem**

Berdasarkan hasil analisa terhadap identifikasi sistem, maka disusunlah sebuah analisa kebutuhan dari sistem tersebut. Berikut adalah hasil analisa kebutuhan sistem yang akan dibangun yang digambarkan pada Gambar 2.



Gambar 2 Diagram Blok Aplikasi

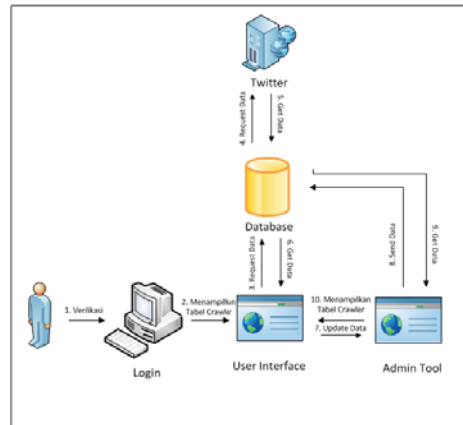
Berikut dijelaskan mengenai diagram blok aplikasi yang terbagi kedalam tiga bagian :

- Membangun *web crawler* sebagai engine dari sistem *social media monitoring*.
- Merancang database sebagai tempat penyimpanan hasil data *crawler*.
- Merancang *user interface* sebagai media untuk menampilkan informasi hasil dari aktifitas *crawler*.

**Analisa Perancangan Sistem**

Tahapan selanjutnya adalah membuat analisa perancangan sistem menggunakan metode *UML*, secara berturut-turut dinyatakan dalam bentuk *workflow*, *flowchart*, *class diagram* dan *sequence diagram*.

**Perancangan Workflow**

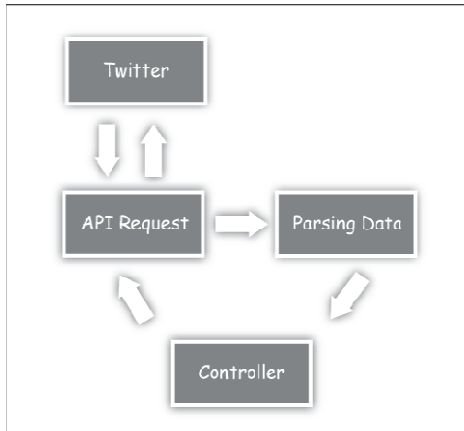


Gambar 3 Workflow Sistem

*Workflow* dari *social media monitoring* ini berdasarkan asumsi awal dari *user* telah ditentukan sebelumnya, sehingga otoritas pada fungsi *login* berlaku untuk pengguna yang terdaftar dalam database. Berdasarkan Gambar 3, pengguna yang sudah diverifikasi otoritasnya menuju halaman utama *web crawler* yaitu tabel informasi *crawler*. *Crawler* akan terus melakukan *update data* dengan cara melakukan *request API* Twitter secara berkala. Setelah *request* tersebut dipenuhi, *crawler* mengambil data tersebut lalu mengolah data menjadi informasi yang bisa ditampilkan pada *user interface*. Jika pengguna tersebut adalah *admin* maka dapat mengolah informasi *crawler* dengan mengakses halaman khusus *admin*.

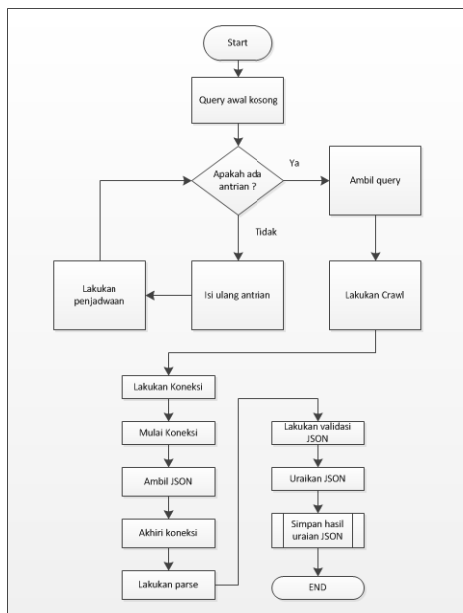
**Perancangan Web Crawler**

Pada perancangan *web crawler*, hal pertama yang dilakukan adalah mendefinisikan proses yang di butuhkan sehingga *crawler* tersebut dapat berfungsi sesuai dengan permasalahan dan pemanfaatan yang ada. Secara garis besar gambaran perancangan *web crawler* sebagai *engine* dari sistem dapat dilihat pada arsitektur aplikasi pada Gambar 4.



Gambar 4 Arsitektur Aplikasi

Proses *crawler* yang akan diimplementasikan terbagi ke dalam beberapa proses yaitu, proses *API request*, proses *parsing data*, dan proses *controlling*. Berikut adalah *flowchart* dari aplikasi *social media monitoring* yang ditunjukkan pada Gambar 5



Gambar 5 Flowchart Crawler

Berikut adalah penjelasan *flowchart* dari Gambar 5.

- Input* awal adalah *query* API yang telah tersimpan dalam database.
- Proses yang berlangsung adalah proses *queue*, proses *scheduling*, proses *API request*, proses *parsing data*.

c. Output yang dihasilkan adalah data pada database yang dapat direpresentasikan sebagai informasi pada *user interface*

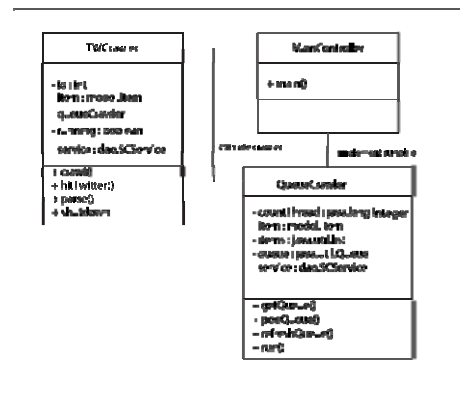
**Perancangan Class Diagram**

Berdasarkan hasil analisis arsitektur aplikasi pada Gambar 4, kelas inti yang akan diimplementasikan dalam *crawler* terbagi dalam tiga kelas. Pada Tabel 1 menunjukkan tanggung jawab untuk masing-masing kelas.

Tabel 1 Tabel Analisis Kelas

Nama Kelas	Nama File	Keterangan
TWCrawler	TWCrawler.class	Kelas yang menyediakan engine inti dari aplikasi ini.
MainController	MainController.class	Kelas kontrol yang mengendalikan sistem.
QueueCrawler	QueueCrawler.class	Kelas kontrol yang mengendalikan sistem Queue

Gambar 6 memperlihatkan diagram kelas perancangan sistem *crawler* berbentuk struktur kelas lengkap beserta atributnya.

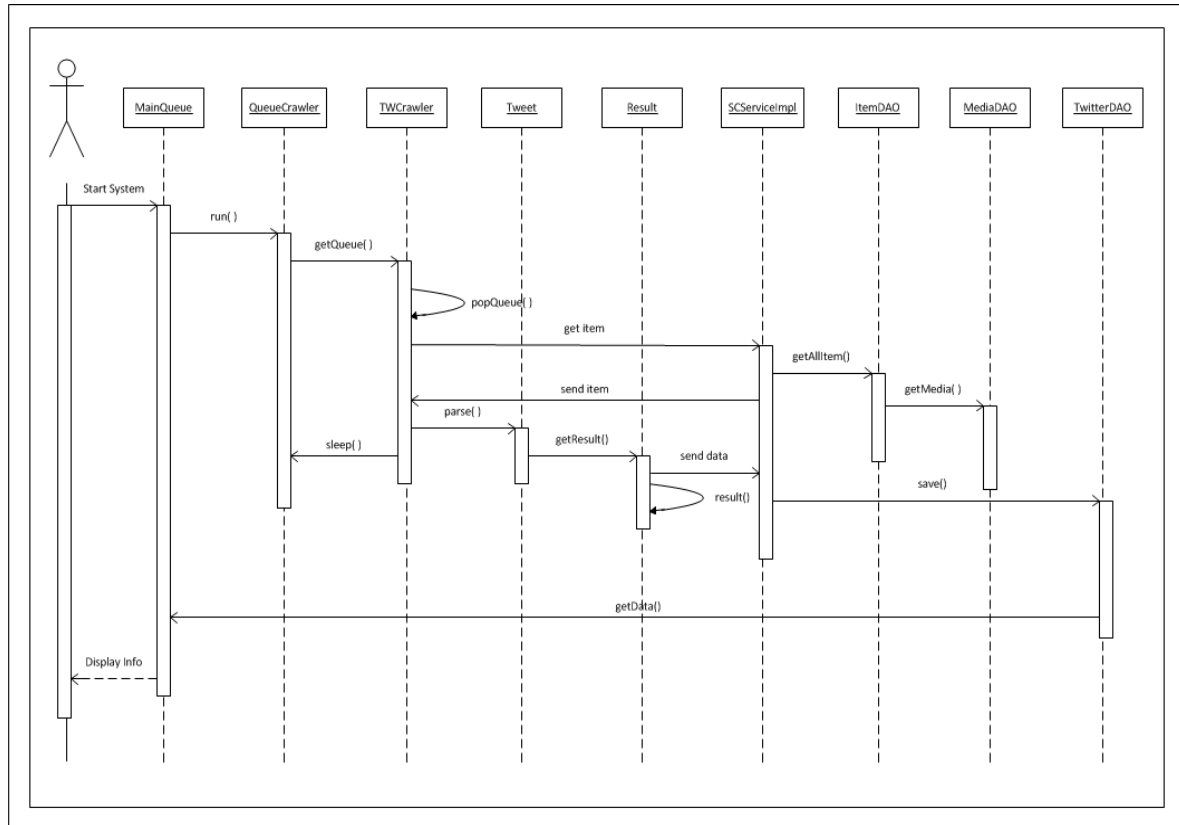


Gambar 6 Diagram Class Crawler

## Perancangan *Sequence Diagram*

Berdasarkan perancangan kelas yang dijabarkan pada Gambar 6, maka *sequence diagram* akan menggambarkan interaksi antar objek di dalam dan di sekitar

sistem, digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respons dari sebuah proses untuk menghasilkan *output* tertentu. Berikut adalah *sequence diagram crawler social media monitoring*.



Gambar 7 *Sequence Diagram Crawler*

## Perancangan Database

Perancangan database yang dimaksud adalah bagaimana data yang telah dimiliki sebagai penunjang pembangunan sistem dapat tersimpan dengan baik di dalam sebuah database. Perancangan database meliputi *Entity Relationship Diagram* (ERD) dan *Table Relationship Diagram* (TRD).

## Perancangan User Interface

Pada pembangunan *user interface* dari *prototype* sistem, *user interface* di rancang dan ditampilkan dalam bentuk *interface* sebuah *website*, karena aplikasi

yang diterapkan pada aplikasi ini berbasis *web*.

## **IMPLEMENTASI DAN PENGUJIAN**

### Implementasi Sistem

Aplikasi *Social Media Monitoring*, memerlukan dukungan perangkat lunak lain dalam implementasinya seperti, sistem operasi Windows, bahasa pemrograman Java, *tools* manajemen database HeidiSQL 7.0.0.4051, Netbeans IDE 7.1.2 sebagai *tools* lingkungan pengembangan Java serta Mozilla Firefox 9.0.1 sebagai *browser* aplikasi.

## Teknik Pengujian

Teknik pengujian yang digunakan adalah teknik pengujian *blackbox testing*. *Blackbox testing* atau disebut juga *behavioral testing*, memiliki fokus pada kebutuhan fungsional dari perangkat lunak. *Blackbox testing* memungkinkan pemrogram untuk

memperoleh sekumpulan kondisi masukan (*input*) yang akan secara penuh menjalankan semua kebutuhan fungsional untuk sebuah program. Dalam hal ini diambil satu butir uji, yaitu menguji fungsi *crawler* yang dibuktikan dalam tabel database seperti yang ditunjukkan pada Tabel 2.

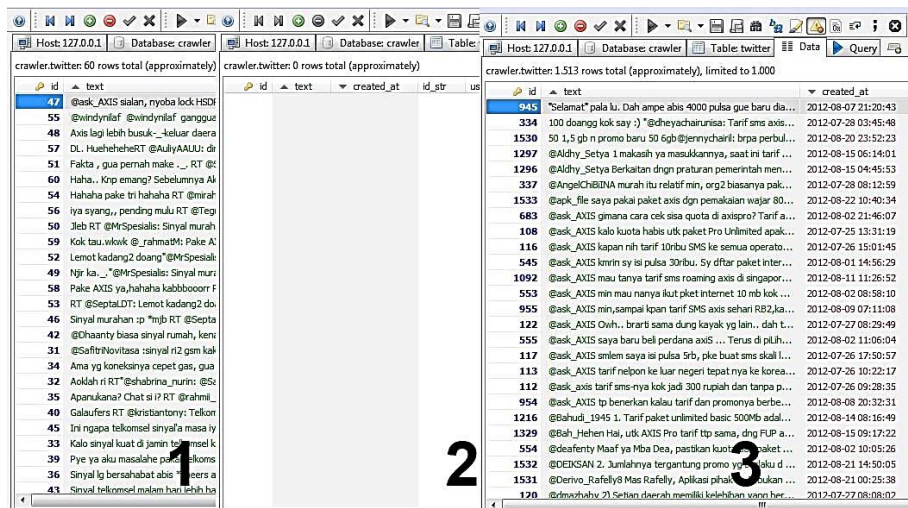
Tabel 2

<b>Identifikasi</b>	<b>CR-01</b>		
Nama Butiran Uji	Uji Implementasi <i>Web Crawler</i>		
Tujuan	Untuk menguji fungsi <i>crawler</i> pada aplikasi terhadap data yang telah berhasil dihimpun.		
Aktor	<i>Admin</i>		
Kondisi Awal	Koneksi <i>crawler</i> dalam keadaan non-aktif dan data dari tabel twitter sudah terisi data <i>crawler</i> .		
Pengujian			
<b>Skenario Uji</b>			
a. Menghapus data dari tabel twitter untuk membuktikan bahwa sistem telah melakukan <i>input</i> data ke dalam database. b. Lalu mengaktifkan koneksi <i>crawler</i> .			
<b>Kriteria Evaluasi Hasil</b>			
a. <i>Crawler</i> dapat melakukan update data terbaru. b. Database dapat menyimpan data hasil <i>crawler</i> .			
Kondisi Akhir	Data dari table twitter terisi data <i>crawler</i>		
<b>Kasus dan Hasil Uji</b>			
Masukan	Harapan	Pengamatan	Kesimpulan
Data tabel twitter	Data pada tabel twitter terhapus dan terisi kembali oleh data <i>crawler</i> terbaru	Koneksi <i>crawler</i> diaktifkan dan data pada tabel twitter terisi oleh data baru.	[x] Terima [ ] Tolak
Hasil yang didapat	Data <i>crawler</i> yang terhimpun pada table twitter sebanyak 1415 data yang selanjutnya dapat ditampilkan pada <i>user interface</i> .		



Hasil pengujian fungsi *crawler* yang dibuktikan dalam tabel database oleh

*admin* direpresentasikan pada Gambar 8.



Gambar 8 Diagram Class Crawler

Hasil pengujian memperlihatkan bahwa aplikasi yang dibangun bekerja dengan baik sesuai dengan spesifikasi kebutuhan, analisis, dan perancangan. Hasil pengujian ini tidak ditemukan kesalahan jika prosedur pelaksanaan operasi dijalankan dengan benar.

**PENUTUP**

**Kesimpulan**

Berdasarkan evaluasi hasil implementasi dan pengujian yang telah dijabarkan pada Tabel 2, tujuan dari pelaksanaan tugas akhir yakni implementasi *web crawler* pada aplikasi *social media monitoring* telah berhasil dilakukan berdasarkan analisa dan perancangan yang telah dilakukan. Fitur-fitur aplikasi yang ada telah bisa digunakan sehingga memudahkan *user* memantau *issue-*

*issue* yang sedang terjadi terhadap sebuah brand image dari sebuah produk

**DAFTAR PUSTAKA**

1. Ensiklopedia bebas, Wikipedia, "Web Crawler", 2012 dalam situs en.wikipedia.org/wiki/Web\_crawler
2. Ensiklopedia bebas, Wikipedia, "Social Media", 2012 dalam situs en.wikipedia.org/wiki/Social\_media
3. Ensiklopedia bebas, Wikipedia, "Monitoring", 2012 dalam situs id.wikipedia.org/wiki/Monitoring
4. Ensiklopedia bebas, Wikipedia, "Twitter", 2012 dalam situs id.wikipedia.org/wiki/Twitter
5. Khanna, Rajiv A., dan Kasliwal, 2007. Sourabh, "Designing a Web Crawler.