JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

NAMA	**:**	**Idham Mardi Putra**
NRP	**:**	**5110100024**
DOSEN WALI	**:**	**Dr. Eng. Nanik Suciati, S.Kom., M.Kom.**
DOSEN PEMBIMBING	**:**	**1. Dr. Royyana Muslim Ijtihadi, S.Kom, M.Kom**
		2. Baskoro Adi Pratomo, S.Kom, M.Kom

2. JUDUL TUGAS AKHIR

"Rancang Bangun Aplikasi Pendeteksi URL Berbahaya Pada *Stream* Twitter"

3. LATAR BELAKANG

Twitter merupakan salah satu media sosial dan layanan berbagi informasi yang memberikan fasilitas bagi pengguna untuk saling bertukar pesan kurang dari 140-karakter dengan memanfaatkan jaringan internet, yang biasa disebut dengan *tweets*. Pengguna dapat mendistribusikan *tweet*-nya otomatis ke semua pengguna lain yang telah terdaftar sebagai teman dari pengguna tersebut, yang biasa disebut dengan *followers*. Selain mendistribusikan *tweet*-nya ke semua *followers*, pengguna tersebut juga dapat mengirim sebuah *tweet* ke salah satu pengguna dengan menyebutkan nama pengguna lain yang didahului tanda "@" pada depan nama pengguna tersebut seperti @bob. Pada Twitter, hal ini biasa disebut dengan *mention*. Pengguna melakukan *mention* tidak harus kepada teman yang terdaftar sebagai *followers*, namun juga dapat melakukannya kepada pengguna lain yang bukan terdaftar sebagai *followers*.

Ketika para pengguna twitter ingin berbagi sebuah URL yang diselipkan pada *tweet*-nya, mereka sering kali memanfaatkan layanan untuk menyusutkan URL dengan mengurangi panjang dari karakter sebuah URL, seperti *bit.ly* dan *tinyurl.com*. hal ini dikarenakan keterbatasan dari jumlah karakter pada tweet tersebut. Sedangkan Twitter sendiri juga mempunyai layanan untuk penyusutan URL secara otomatis. Artinya

terdapat suatu rantai URL yang saling mengalihkan hingga ke suatu domain situs tertentu, dalam hal ini URL tersebut bisa dikategorikan URL yang dicurigai berbahaya.

Dengan memanfaatkan twitter sebagai salah satu media sosial yang populer di seluruh penjuru dunia, para penyerang dunia maya (*attacker*) mengambil peluang dengan mendistribusikan URL berbahaya tersebut kepada semua pengguna twitter. Perlu diketahui bahwa, *spam* yang merupakan salah satu bentuk paling umum dari serangan web, menjadi bermunculan di twitter. URL yang telah disisipkan oleh *attacker* pada *tweet* merupakan suatu rantai dari URL-URL lain yang mengarah pada halaman tertentu.

Sistem pendeteksi URL Berbahaya pada media sosial seperti twitter ini sudah pernah diusulkan dengan menggunakan *social honeypot* yang memanfaatkan suatu agen pada *honeyclient* yang biasa disebut dengan *crawler* untuk menangani kasus semacam itu [2]. Namun para *attacker* tidak tinggal diam, mereka menggunakan beberapa teknik untuk menghindari para *crawler* dan meneruskan URL tersebut ke URL yang umum, seperti google.com. Tidak hanya itu, seorang *attacker* dapat dengan mudah memalsukan suatu akun pada twitter.

Oleh karena itu, pada tugas akhir ini, dirancang dan diimplementasikan suatu aplikasi sistem pendeteksi URL berbahaya pada *stream* Twitter yang diambil dari Twitter *stream* API dengan mengamati rantai URL dari suatu *tweet*. Kemudian setelah data dikumpulkan, dilakukan tahapan ekstraksi fitur dengan cara mengelompokan identitas domain dari dua URL atau lebih. Selanjutnya pengekstrakan didasari oleh dua fitur, yakni fitur pertama yang berasal dari korelasi antar rantai URL mulai dari panjang rantai URL, posisi dari *entry point* URL, banyaknya jumlah inisial URL yang berbeda, dan banyaknya jumlah URL *destination* yang berbeda, dan fitur yang kedua berdasarkan konteks informasi dari *tweet* itu sendiri seperti, jumlah *sources* URL yang berbeda, jumlah akun twitter yang berbeda, standar deviasi dari kapan dibuatnya akun pengirim *tweet*, dan kesamaan isi teks pada *tweet*.

4. RUMUSAN MASALAH

Rumusan masalah yang akan diangkat dalam Tugas Akhir ini dapat dipaparkan sebagai berikut:

1. Bagaimana proses pengambilan *tweets* yang muncul dengan URL secara terus menerus yang bersifat *spam*?
2. Bagaimana cara menyimpan data yang telah diambil untuk dianalisa URLnya?
3. Apakah URL satu dengan yang lain yang memiliki kesamaan *IP Address* antar domain?
4. Berapa jumlah frekuensi dari rantai URL yang didapat dari suatu *tweet?*
5. Apakah URL yang telah dianalisa mempunyai pengalihan kondisi dari URL yang bercabang?
6. Bagaimana cara mencari letak *entry point* pada tengah rantai URL?

7. Bagaimana membedakan akun pengguna twitter yang berstatus *suspended* dan yang aktif?
8. Bagaimana mencari kesamaan teks pada *tweet?*
9. Bagaimana mengambil kesimpulan akhir untuk menentukan apakah URL yang telah didistribusikan melalui *tweet* berbahaya atau tidak?
10. Bagaimana cara mengirimkan suatu notifikasi apabila ditemukan suatu URL yang berbahaya?

5. BATASAN MASALAH

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki beberapa batasan, diantaranya sebagai berikut:

1. Aplikasi ini menggunakan *crawler* yang statis, sehingga hanya mampu menangani HTPP *redirection*. Hal ini menjadikan sistem ini tidak efektif dalam penanganan halaman yang mengandung *embedded dynamic redirection* seperti *JavaScript* atau *Flash redirection*.
2. Untuk melakukan pengambilan *tweet* yang mengandung URL, aplikasi ini membutuhkan *history* dari *tweet* sebelumnya sehingga aplikasi ini tidak bisa berjalan secara *real time*, tetapi harus mengambil dan mengumpulkan data tweet yang mengandung URL
3. Hanya 12 fitur yang digunakan sebagai *feature vector*.

6. TUJUAN PEMBUATAN TUGAS AKHIR

Tugas Akhir ini memiliki tujuan yang rinciannya dapat dituliskan sebagai berikut:

1. Membuat layanan untuk mendeteksi adanya URL berbahaya pada *tweet*.
2. Sebagai pengembangan dari usul yang sebelumnya yang masih terbilang lemah dalam pendeteksian URL berbahaya pada sosial media dengan menggunakan *honeypot client*.
3. Mencegah terjadinya pendistribusian *tweet* yang berkepanjangan yang digunakan oleh akun twitter yang berstatus *suspended*.

7. MANFAAT TUGAS AKHIR

Dengan dibangunnya aplikasi sistem pendeteksi ini, diharapkan mampu mengurangi kejahatan pada dunia maya yang menyerang pada sosial media melalui *tweet* yang mengandung unsur URL yang berrantai yang dicurigai sebagai URL yang berbahaya.

8. TINJAUAN PUSTAKA

8.1 Analisis *URL Redirect Chains*

URL *Redirect Chains* merupakan suatu URL yang terlihat bukanlah URL yang sebenarnya ingin dituju. URL tersebut melakukan banyak pengalihan halaman yang telah ditentukan oleh seorang *attackers*. Biasanya mereka memanfaatkan layanan untuk menyusutkan URL dengan mengurangi panjang dari karakter sebuah URL, seperti *bit.ly* dan *tinyurl.com*. selain itu pada twitter itu sendiri juga mempunyai layanan memperkecil ukuran karakter dari suatu URL seperti t.co. dan mereka mempunyai halaman yang dapat mengalihkan lebih dari satu tujuan agar membedakan pengunjung yang menggunakan *browser* yang normal atau *crawler*.

8.2 Twitter

Twitter adalah layanan jejaring sosial yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (*tweet*). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di internet. Di Twitter, pengguna tak terdaftar hanya bisa membaca *tweet*, sedangkan pengguna terdaftar bisa mengirim tweet melalui antarmuka situs *web*, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler. Melalui twitter seorang pengguna / pemilik akun dapat melakukan update status atau berinteraksi dengan sesama pemilik akun. Jika membutuhkan privasi dalam berinteraksi dengan sesama, pengguna dapat mengirimkan pesan pribadi ke pengguna lain yang dituju. Di dalam twitter, terdapat istilah *follow* dan *followers* yang di mana follow berarti kita dapat mengikuti kebiasaan pengguna melalui *tweet – tweet* yang dikirimkan pengguna tersebut. Secara otomatis, *tweet* dari penggguna yang kita *follow* akan muncul di *timeline* kita. Sedangkan *followers* merupakan pengguna lain yang mem-*follow* akun kita. Twitter mengalami pertumbuhan yang pesat dan dengan cepat meraih popularitas di seluruh dunia. Hingga bulan Januari 2013, terdapat lebih dari 500 juta pengguna terdaftar di Twitter, 200 juta di antaranya adalah pengguna aktif [3].

Pada awal 2013, pengguna Twitter mengirimkan lebih dari 340 juta *tweet* per hari, dan Twitter menangani lebih dari 1.6 miliar permintaan pencarian setiap hari. Hal ini menyebabkan posisi Twitter naik ke peringkat kedua sebagai situs jejaring sosial yang paling sering dikunjungi di dunia, dari yang sebelumnya menempati peringkat dua puluh dua Seiring dengan bertambah naiknya jumlah pengguna twitter, angka kejahatan di dunia maya yang memanfaatkan twitter sebagai media ikut bertambah tinggi.

8.3 Twitter Steaming APIs

Application Programming Interface atau yang biasa disebut dengan API merupkan sekumpulan perintah, fungsi, dan protokol yang dapat digunakan oleh programmer saat membangun perangkat lunak untuk sistem operasi tertentu. API

memungkinkan programmer untuk menggunakan fungsi standar untuk berinteraksi dengan sistem operasi.

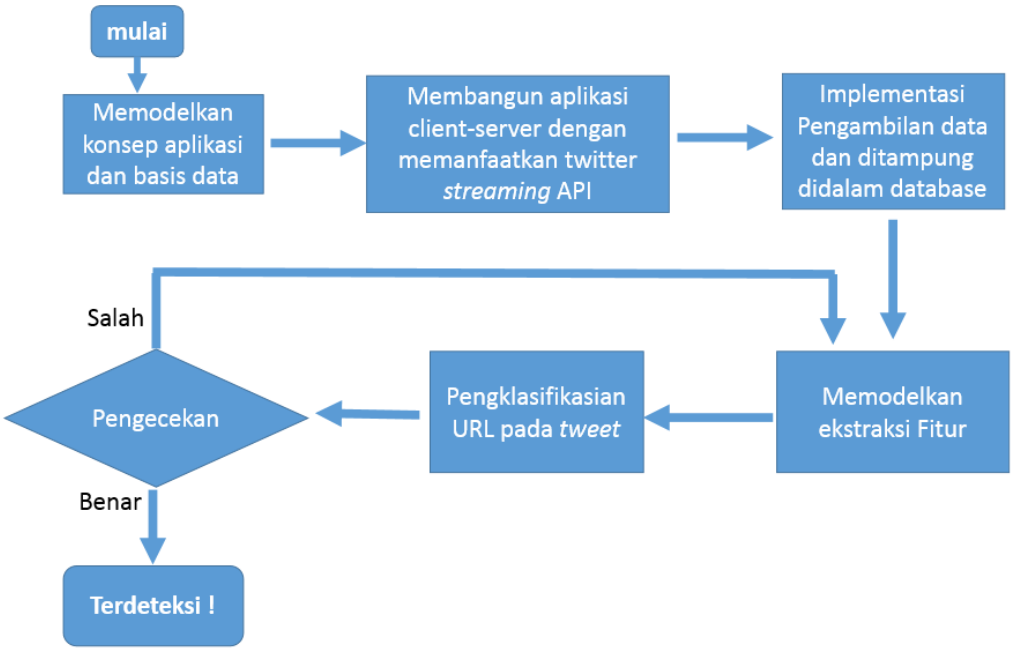
Twitter *Streaming* API merupakan sekumpulan perintah dan fungsi yang digunakan untuk membangun perangkat lunak yang berhubungan atau berinteraksi dengan fungsi yang terdapat pada aliran Twitter [5].

9. RINGKASAN ISI TUGAS AKHIR

Twitter dapat memberikan fasilitas bagi pengguna untuk menyebarkan *tweet* berbahaya dengan konten berupa URL dan didistribusikan ke seluruh penjuru dunia dengan cara *spam*. Sementara itu sudah diusulkan skema untuk pendeteksian fitur yang digunakan khusus untuk akun pada Twitter seperti rasio dari tweet yang mengandung URL dan tanggal pembuatan akun, yang dapat memberikan status *suspended* atau masih aktif. Dengan menggunakan *Honeypot Client* seperti *Capture-HPC* [4], *HoneyMonkey*, pendeteksian URL yang menyebabkan terjadinya *spam* tersebut masih terbilang lemah. Oleh karena itu diperlukan cara lain yang mampu mendeteksi secara akurat apakah *tweet* yang mengandung URL berbahaya tersebut,

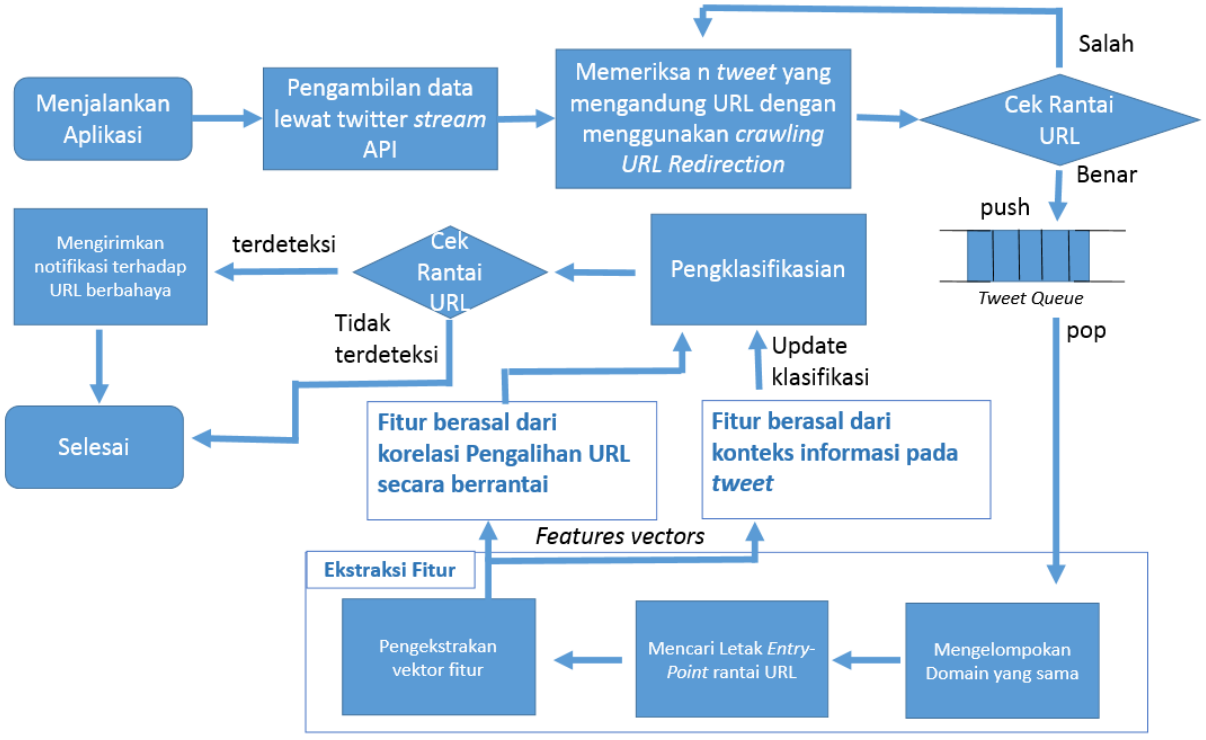
Tugas akhir ini mencoba memberikan satu solusi baru untuk menentukan apakah *tweet* yang mengandung URL berbahaya dapat dideksi oleh aplikasi ini. Apabila terdeteksi adanya *tweet* yang mengandung unsur URL yang berbahaya maka akan dikirimkan suatu bentuk notifikasi dan meminimalisir *stream* twitter agar mencegah terjadinya *spam* yang dapat membuat server pada twitter terancam dan membuat pengguna lain menjadi nyaman.

Langkah-langkah pengerjaan tugas akhir ini akan dijelaskan pada Gambar 1.



Gambar 1. Alur pengerjaan proposal tugas akhir

Berikut alur kerja dari aplikasi:



Gambar 2. Alur kerja aplikasi

Berdasarkan gambar 2. Alur kerja aplikasi ini adalah sebagai berikut:

1. Aplikasi dijalankan dan melakukan pengambilan data *tweet* yang mengandung URL lewat twitter *stream* API
2. *Tweets* dipush kedalam penampung *tweet Queue* dalam basis data dan memulai penecekan dengan mengelompokan domain yang menggunakan *IP-Address* yang sama.
3. Mencari letak *entry point* dengan melihat frekuensi kemunculan URL pada rantai URL satu dengan yang lainnya.
4. Pengekstrakan *feature vector* yang berdasarkan 2 faktor, yakni berdasarkan rantai URL, dan konteks informasi pada *tweet*.
5. Proses pengklasifikasian, bila terdapat informasi pada pencocokan status akun, maka klasifikasi akan diperbaharui
6. Cek Rantai URL tersebut, ada tidaknya deteksi akan dikirimkan suatu notifikasi.

Dalam proses pembentukan *feature vector*, berikut adalah fitur yang akan digunakan berdasarkan 2 faktor tersebut, yaitu:

A. Fitur berdasarkan korelasi URL yang berantai
   1. Panjang dari suatu rantai URL

2. Frekuensi dari *entry point* URL
3. Letak dari *entry point* URL
4. Jumlah titik pada alamat situs
5. Jumlah titik pada semua tautan

B. Fitur berdasarkan konteks informasi pada *tweet*

6. Jumlah pada *source tweet* yang berbeda
7. Jumlah pada akun pengirim yang berbeda
8. Standar deviasi pada tanggal pembuatan akun pengirim
9. Standar deviasi pada *followers* akun pengirim
10. Standar deviasi pada *following* akun pengirim
11. Standar deviasi pada rasio dari *followers-following* akun pengirim
12. Jumlah kemiripan isi dari *tweet* tersebut.

10.METODOLOGI

a. Penyusunan proposal tugas akhir

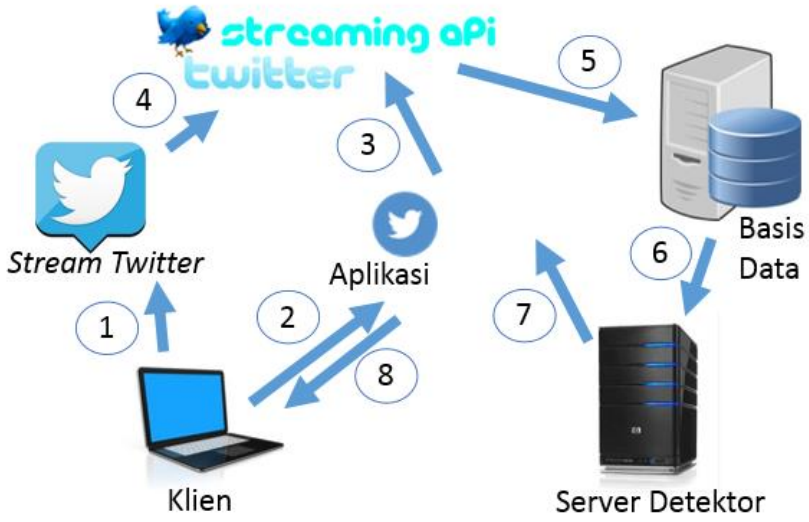
Tahap awal untuk memulai pengerjaan tugas akhir ini adalah penyusunan proposal tugas akhir. Pada proposal ini penulis mengajukan gagasan dan rancang sistem pendeteksi suatu *tweet* yang dikategorikan mengandung unsur URL yang berbahaya dengan menggunakan pencocokan *entry point* antar rantai URL. *Entry point* didapat dari freksuensi kesamaan URL pada penemuan di tengah rantai URL tersebut dengan penemuan di tengah rantai URL yang lain. Apabila terdeteksi adanya *tweet* yang mengandung unsur URL yang berbahaya maka akan dikirimkan suatu bentuk notifikasi dan meminimalisir *stream* twitter.

b. Studi literatur

Tugas akhir ini menggunakan literatur paper beserta artikel dari internet. Paper yang digunakan adalah "*WARNING BIRD: Detecting Suspicious URLs in Twitter Stream"* [1]. Paper tersebut menjadi acuan utama dan dasar dalam pengerjaan tugas akhir ini.

c. Analisis dan desain perangkat lunak

Dalam aplikasi ini digunakan arsitektur *client - server*. Sehingga dibutuhkan server dalam konteks ini server meliputi: Server basis data sebagai pengolahan data, Server detektor sebagai sistem pendeteksi URL tersebut, web server sebagai memberikan layanan untuk mengakses situs URL tersebut.



Gambar 3. Arsitektur Jaringan

Berdasarkan gambar 3. Arsitektur jaringan alur kerja pada aplikasi ini adalah sebagai berikut:

1. Klien login sebagai akun twitter dan melihat isi dari *stream* Twitter.
2. Kemudian aplikasi dijalankan oleh klien.
3. Aplikasi mengambil data dari *stream* twitter menggunakan Twitter *Streaming* APIs
4. Kemudian layanan Twitter *Streaming* APIs tersebut mengambil dan mengoleksi sejumlah *n tweet*.
5. Dari data yang dikumpulkan tersebut, dimasukan kedalam basis data.
6. Data tersebut di-*push* kedalam server detektor dan menghapus semua data saat itu dari basis data.
7. Setelah dianalisa oleh server detektor, data tersebut disaring dan di-*push* ke dalam aplikasi tersebut.
8. Aplikasi mengirimkan notifikasi sekumpulan *tweet* yang mengandung URL yang berbahaya.

d. Implementasi perangkat lunak

Dalam pembuatan aplikasi, digunakan beberapa teknologi untuk dapat mengaplikasikan rancangan yang sudah ada, diantaranya:

a. Twitter *Stream* API
   Twitter *Stream* API digunakan sebagai library dan fungsi tambahan yang memungkinkan program yang dibuat bekerja [5].
b. Bahasa Pemrograman Aplikasi

Aplikasi ini dibangun dengan menggunakan bahasa pemrograman Java. Penggunaan bahasa pemrograman diharapkan dapat membantu menangani kebutuhan aplikasi terutama kemudahan untuk konektivitas dengan basis data dan kebutuhan lainnya.

c. Basis Data
   Basis data pada server digunakan untuk menampun seluruh data yang dibutuhkan dari *tweet* yang mengandung unsur URL. Dalam sistem ini akan digunakan basis data MySQL
d. IDE
   Pengembangan aplikasi ini menggunakan Netbeans 7.2 sebagai IDE
e. Modeling Tools
   Beberapa modeling tools yang digunakan untuk mengembangkan aplikasi ini Power Designer 15.00, StarUML, Microsoft Visio 2013

e. Pengujian dan evaluasi

Pada tahap ini dilakukan uji coba terhadap aplikasi yang telah dibuat.Tujuan uji coba perangkat lunak adalah untuk menemukan kesalahan-kesalahan (*bug*) sedini mungkin sehingga dapat diperbaiki sesegera mungkin.

f. Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
   a. Latar Belakang
   b. Rumusan Masalah
   c. Batasan Tugas Akhir
   d. Tujuan
   e. Metodologi
   f. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

11. JADWAL KEGIATAN

Tahapan	Tahun 2013																			
	Oktober				Nopember				Desember				Januari					Februari		
Penyusunan Proposal																				
Studi Literatur																				
Perancangan sistem																				
Implementasi																				
Pengujian dan evaluasi																				
Penyusunan buku																				

12. DAFTAR PUSTAKA

[1] S. Lee and J. Kim, "WarningBird: Detecting Suspicious URLs in Twitter Stream" in symposium on Netwrok and Distributed System Security (NDSS), 2012

[2] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: Social honeypots + machine learning

[3] "Twitter." Wikipedia, [Online]. Availabe: http://id.wikipedia.org/wiki/Twitter. [Accessed 4 October 2013]

[4] Capute-HPC. https://projects.honeynet.org/capture-hpc

[5] "Twitter Streaming APIs" Available: https://dev.twitter.com/docs/streaming-apis [Accessed 9 October 2013]