

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Social Media Analytics* saat ini sedang menjadi trend di dunia, banyak yang menggunakan untuk internal perusahaan sendiri, atau dijadikan *Software as a Service (SAAS)*. Tujuan menggunakan *Social Media Analytics* sendiri, yaitu untuk mengetahui, sejauh mana *Brand* suatu perusahaan itu berefek di masyarakat, apakah baik atau buruk, berapa orang yang suka atau tidak, dan siapakah *Influencer* pada suatu *Brand* tersebut.

PT. Mediawave Analytic adalah perusahaan IT & digital marketing yang didirikan pada tahun 2009, yang berfokus pada *Social Media Analytics*. Mereka memiliki mesin sendiri untuk melakukan *Social Media Analytics*, sumber data utama untuk *Analytic* didapat dari *twitter, facebook, blog, news*, sebagian memanfaatkan *API (Application Programming Interface)* yang *Open* salah satunya dari *Google* untuk mendapatkan data dari *blog, news, images* dan *video*. PT. Mediawave Analytic sendiri merupakan perusahaan pertama di Indonesia yang memulai *Social Media Analytics Service*. Di luar Indonesia, beberapa perusahaan yang memiliki *Social Media Analytics Service* yaitu *Radian6*, dan *Uber VU*.

Untuk mendapatkan data berupa *News, Blogs, Images*, dan *Video*, PT Mediawave Analytic memanfaatkan *API Searching* milik *Google*. Dan saat ini, *API* mereka sudah dibatasi, bahkan tidak bisa diakses. Maka sumber data berupa *News*, tidak bisa didapat dengan mudah. Maka cara untuk mendapatkannya yaitu

mengambil (*crawl*) dari *Website* penyedia berita. *Web crawler* adalah program yang mengeksplorasi struktur grafik *Web* untuk berpindah dari halaman ke halaman. Dalam tahap awal pengembangan aplikasi, bisa disebut juga dengan *Spiders, wanderers, fish*, atau *worms*.

Pembuatan aplikasi *News Crawler* adalah langkah awal untuk mendapatkan data berupa *News* yang sangat kompleks. Dengan begitu aplikasi yang dibangun dapat menjadi sumber data yang bisa dengan mudah diakses dan dapat dipergunakan dalam *Analytic Tool* milik PT Mediawave Analytic, sehingga dapat menjawab permasalahan yang ada. Dari analisa di atas, maka skripsi ini mengambil judul : “***News Crawler dengan Metode XPATH***”.

## 1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang, maka masalah yang akan dibahas yaitu :

1. Bagaimana membuat mesin pencari berita (*News*) dengan memanfaatkan metode *Web crawling* sehingga didapatkan data mentah (*raw data*) yang dapat diolah kembali.
2. Bagaimana membuat proses pencarian berita (*News*) dan *input* data ke dalam *database* bisa sangat cepat.

## 1.3 Tujuan Penelitian

Sesuai dengan rumusan masalah, maka tujuan penelitian adalah:

1. Membuat mesin pencari berita.

2. Membuat *database* yang besar tentang berita (*news*) sendiri dan dapat dipergunakan untuk *analytic* bagi PT Mediawave Analytic maupun *Searching* data bagi *user public*.

#### 1.4 Batasan Penelitian

Dalam penulisan skripsi ini, diberikan batasan penelitian sebagai berikut :

1. Aplikasi ini hanya mendapatkan data saja (menyimpan dalam *database*). Aplikasi ini berbasis *Python*, menggunakan *Web Frame work django*, *database My SQL*, dan akan berjalan otomatis selama 24 jam/ hari, 7 hari/ minggu di *server VPS* yang berbasis *linux* milik PT Mediawave Analytic.
2. Data yang dicari hanya data berupa *content* berita (*news*), dengan memanfaatkan *RSS* dari penyedia berita sebagai data awalnya.
3. Sumber berita didapat dari sumber yang sudah ditentukan (*detik news*, *kompas*, dan *viva news*).
4. Aplikasi dapat digunakan untuk melakukan pencarian berita dan melihat *statistic* dengan output berupa *chart* (grafik).

#### 1.5 Metodologi Penelitian

Setiap mengadakan penelitian dan analisa maka tidak terlepas dari berbagai masalah yang perlu dihadapi dan harus dipecahkan. Agar lebih praktis digunakan metode ilmiah sehingga dapat diperoleh jalan keluar yang baik, efektif serta mudah.

### 1.5.1 Metodologi Pengumpulan Data

Metode pengumpulan data yang dipergunakan adalah sebagai berikut.

a. Wawancara langsung

Dalam penelitian ini digunakan teknik wawancara, yaitu dengan cara melakukan tanya jawab secara lisan dan langsung dengan analis yang ada di PT Mediawave Analytic sehubungan dengan masalah yang diteliti.

b. Observasi

Teknik observasi yaitu melakukan pengamatan langsung ke lokasi penelitian, yaitu di PT Mediawave Analytic.

c. Dokumentasi

Data lainnya diperoleh dari dokumen-dokumen perusahaan yang berhubungan dengan masalah yang diteliti.

d. Studi Pustaka/ Literatur

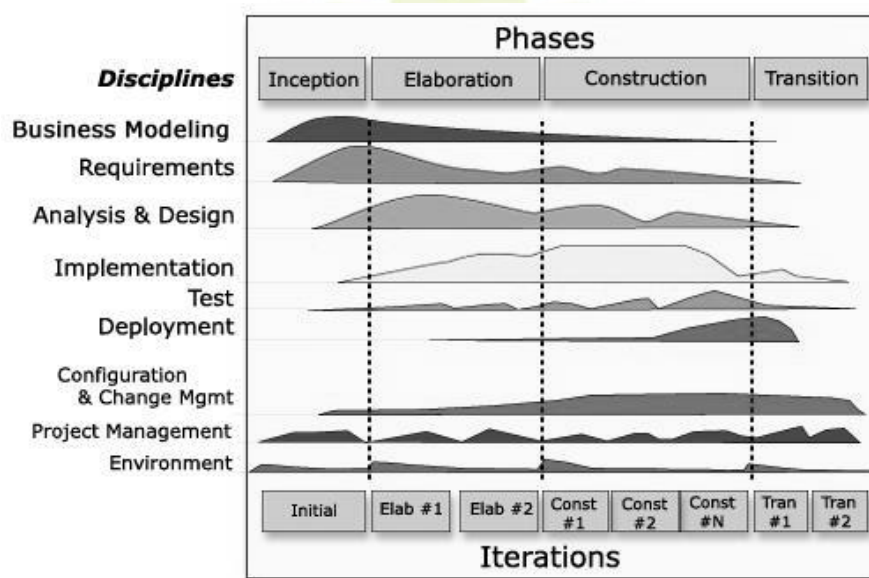
Penelitian dilakukan melalui buku-buku literatur atau jurnal yang berhubungan dengan masalah yang diteliti sebagai bahan acuan dalam menyusun skripsi ini.

### 1.5.2 Metode Pengembangan Perangkat Lunak

*Rational Unified Process* (RUP) adalah salah satu proses Rekayasa Perangkat Lunak. RUP ini menyediakan pendekatan disiplin untuk menugaskan tugas dan tanggung jawab dalam organisasi pembangunan. Tujuannya adalah untuk memastikan perangkat lunak yang berkualitas tinggi dan memenuhi kebutuhan (*end-user*) pengguna akhir, yang sesuai dengan jadwal dan anggaran yang diprediksi. [1]

RUP memiliki 2 dimensi [1], yaitu:

1. **Dimensi pertama** digambarkan secara horizontal. Dimensi ini mewakili aspek-aspek dinamis dari pengembangan perangkat lunak. Aspek ini dijabarkan dalam tahapan pengembangan atau fase.
2. **Dimensi kedua** digambarkan secara vertikal. Dimensi ini mewakili aspek-aspek statis dari proses pengembangan perangkat lunak yang dikelompokkan ke dalam beberapa disiplin.



SUNAN GUNUNG DJATI  
 Tabel 1.1. Arsitektur *Rational Unified Process*

Fase dalam RUP [1], antara lain :

- a) *Inception*/ Insepsi : Menentukan Ruang lingkup proyek; Membuat '*Business Case*'; Menjawab pertanyaan "apakah yang dikerjakan dapat menciptakan '*good business sense*' sehingga proyek dapat dilanjutkan
- b) *Elaboration*/ Elaborasi : Menganalisa berbagai persyaratan dan resiko; Menetapkan '*base line*'; Merencanakan fase berikutnya yaitu *construction*

- c) *Construction/ Kontruksi*: Melakukan sederetan iterasi; Pada setiap iterasi akan melibatkan proses berikut: analisa desain, implementasi dan testing
- d) *Transistion/ Transisi* : Membuat apa yang sudah dimodelkan menjadi suatu produk jadi; Dalam fase ini dilakukan : beta dan performance testing, membuat dokumentasi tambahan seperti; training, *userguides* dan sales kit, membuat rencana peluncuran produk ke komunitas pengguna.

## 1.6 Tempat Penelitian

Penelitian dilakukan di PT Mediawave Analytic, Jl. Reog No. 4D, Bandung dilakukan pada periode Desember – Februari 2012.

## 1.7 Sistematika Penulisan

Secara garis besar gambaran mengenai hal-hal yang dibahas dalam skripsi ini terdiri dari enam bab yaitu:

### Bab I Pendahuluan

Dalam bab ini menguraikan tentang beberapa sub bahasan yaitu Latar Belakang, Identifikasi Masalah, Batasan Penelitian, Tujuan dan Manfaat Penelitian, Metodologi Penelitian, Waktu dan Tempat Penelitian dan Sistematika Penulisan.

### Bab II Landasan Teori

Menjelaskan tentang konsep dasar dan teori-teori yang digunakan untuk menjelaskan dan menyelesaikan permasalahan yang akan dikaji.

### Bab III Tinjauan Umum Perusahaan

Menjelaskan tentang PT Mediawave Analytic, apa yang dilakukan, apa

yang diteliti, dan pengembangan sistem *Analytic* nya.

#### Bab IV Analisis dan Perancangan

Menganalisis masalah dari penelitian yang sudah dilakukan untuk memperlihatkan keterkaitan antar variabel yang diteliti serta model analisisnya. Dan merancang sistem yang akan di implementasikan pada tahap berikutnya.

#### Bab V Implementasi dan Pengujian

Implementasi sistem merupakan penerapan dari analisis dan desain sistem yang telah dilakukan menjadi suatu aplikasi. Serta melakukan tahap pengujian setelah implementasi selesai.

#### Bab VI Kesimpulan Dan Saran

Merupakan bagian akhir penyusunan skripsi yang berisi kesimpulan dan saran.





uin

UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG