BAB I
PENDAHULUAN

I.1 Latar Belakang Masalah

Web Content Mining (WCM) merupakan salah satu bagian dari *Web Mining* yang digunakan untuk menemukan informasi yang berguna dari sekian banyak data yang terdapat di internet seperti gambar, suara, video, dan teks. WCM sendiri memiliki beberapa teknik atau metode yang bisa digunakan dalam memperoleh informasi-informasi tersebut. WCM biasanya menggunakan teknik *text mining* karena sebagian besar data yang berada di *World Wide Web* (WWW) berupa teks. Berdasarkan hasil penelitian dari Insyafie Yuliansyah [1], menyebutkan bahwa WCM dapat diterapkan dalam menemukan pola kecendrungan berita dengan tingkat akurasi sebesar 90% yang dihasilkan pada uji coba ke-5 dalam 10 kali percobaan yang diuji secara fungsional keseluruhan. Penelitian tersebut menggunakan metode *text preprocessing* dengan *regex* untuk pengolahan datanya. Maka dari penelitian tersebut, metode WCM ini akan digunakan kembali ke sebuah sistem untuk mengumpulkan dan mencari informasi berita berdasarkan inputan kata kunci dari pengguna atau *Information Retrieval* (IR).

Agregator berita (*news aggregator*) merupakan salah satu jenis sistem agregator (pengumpul) yang mengumpulkan berita dari berbagai sumber, kemudian menyajikannya kembali kepada pengguna dalam satu kesatuan. Sehingga, pengguna tidak perlu lagi menjelajah ke berbagai situs berita untuk sekedar mencari informasi. Sistem *news aggregator* ini dapat mengurangi waktu dan upaya yang dibutuhkan untuk mengecek situs berita dalam mencari sebuah informasi. Menurut penelitian dari Rubi Henjaya [2] dan Arie Karhendana [3] bahwa sistem *news aggregator* “polos” atau yang hanya mengumpulkan berita saja memiliki kelemahan, yaitu saat menggabungkan semua berita hasil perolehannya dalam satu kumpulan besar dapat mengakibatkan efek *information*

overload bagi pengguna karena belum tentu pengguna tertarik dengan semua berita yang terdapat pada *news aggregator*.

Berdasarkan masalah tersebut, dapat disimpulkan bahwa pengguna hanya mencari berita tertentu dari beberapa situs berita yang ada. Maka dari itu, diperlukannya metode WCM dalam pengolahan sistem *news aggregator* ini agar dapat menampilkan informasi berita sesuai dengan yang pengguna cari dari beberapa situs berita yang ada. Pada tugas akhir ini akan memilih pendekatan *partitional clustering* untuk pengelompokkan berita-berita dengan topik yang sama, yaitu dengan algoritma K-Means karena berdasarkan penelitian dari Deddy W. Suliantoro, Irya Wisnubhadra dan Ernawati [4] algoritma ini cocok digunakan untuk *clustering* dokumen teks dengan tingkat *precision* dan *recall* yang tinggi serta waktu kerja yang relatif singkat oleh sistem.

Berdasarkan penjelasan diatas, diharapkan dapat selesainya tugas akhir ini yang akan dikembangkan ke dalam bentuk skripsi berjudul "**Penerapan *Web Content Mining* Menggunakan *Partitional Clustering* K-Means Pada *News Aggregator***".

I.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan sebuah masalah yaitu bagaimana menerapkan *Web Content Mining* dengan menggunakan *Partitional Clustering* K-Means pada *News Aggregator*.

I.3 Maksud dan Tujuan

Maksud dari tugas akhir ini adalah untuk menerapkan *Web Content Mining* dengan menggunakan *Partitional Clustering* K-Means pada sistem *News Aggregator.*

Sedangkan tujuannya untuk mengetahui apakah sistem *news aggregator* akan bekerja dengan baik dalam menampilkan informasi berita yang telah dikelompokkan dari beberapa situs berita berdasarkan inputan kata kunci dari pengguna.

I.4 Batasan Masalah

Batasan Masalah dimaksudkan untuk membatasi ruang lingkup pekerjaan yang dilaksanakan dengan tujuan untuk memperkecil masalah yang ada. Adapun batasan masalah dalam penelitian tugas akhir ini adalah sebagai berikut:

1. Jenis data berupa teks.
2. Topik berita yang digunakan dalam penelitian ini adalah berita seputar pendidikan.
3. Berita yang digunakan adalah berita *online* bahasa indonesia yang diambil dari 3 situs berita, yaitu Tribun, Republika, dan Kompas.
4. Fungsional pada sistem *News Aggregator* ini menggunakan pendekatan agen atau *Information Retrieval* dimana pengguna akan memasukkan *query* atau kata kunci berita yang akan dicari dari kumpulan berita.
5. Teknik yang digunakan adalah *Text Mining* karena sumber data yang digunakan adalah teks.
6. Algoritma untuk *stemming* adalah algoritma Nazief dan Andriani.
7. Informasi berita yang diambil berupa deskripsi berita atau *headlines*.
8. Untuk pengambilan informasi berita dari beberapa situs media *online* menggunakan teknik *Web Crawler* dengan *max depth* sampai level 1.

I.5 Metodologi Penelitian

Dalam tugas akhir ini digunakan metode penelitian deskriptif yang menggambarkan fakta-fakta dan informasi secara sistematis, faktual, dan akurat. Penelitian ini memiliki dua metode penelitian, yaitu metode pengumpulan data dan metode pembangunan perangkat lunak.

I.5.1 Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini adalah dengan melakukan studi literatur yang di mana melakukan pengumpulan literatur, jurnal, *paper* dan bacaan-bacaan yang ada kaitannya dengan judul penelitian.

I.5.2 Metode Pembangunan Perangkat Lunak

Untuk membangun sistem *news aggregator* yang akan dikembangkan ini menggunakan pendekatan *prototype*, yang meliputi beberapa proses didalamnya :

1. Pengumpulan Kebutuhan

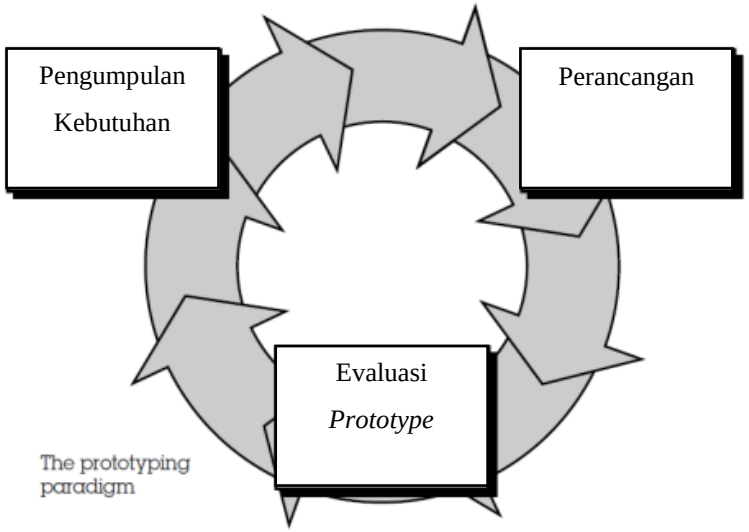
   Pengumpulan data dengan cara mengumpulkan literatur, jurnal, *paper* dan bacaan-bacaan yang ada kaitannya dengan judul penelitian.

2. Perancangan

   Perancangan dilakukan cepat dan rancangan mewakili semua aspek *software* yang diketahui. Rancangan ini akan menjadi dasar pembuatan *protoype*.

3. Evaluasi *Prototype*

   Pengujian terhadap *software* yang dibuat dan digunakan untuk memperjelas kebutuhan *software*.



Gambar I.1 Model *Protoype* [5]

I.6 Sistematika Penulisan

Sistematika penulisan penelitian tugas akhir ini disusun untuk memberikan gambaran umum tentang penelitian yang dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini menguraikan secara umum latar belakang masalah, rumusan masalah, maksud dan tujuan penelitian, batasan masalah, metodologi penelitian serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas landasan teori yang digunakan dalam pembuatan tugas akhir ini, yaitu teori mengenai *Web Mining*, *Text Mining*, *Clustering*, *K-Means*, *Web Crawler*, PHP, *Unified Modelling Language* (UML), serta teori-teori pengujian yang akan dilakukan terhadap sistem yang dibangun.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan analisis terhadap seluruh spesifikasi sistem yang mencakup analisis masalah, analisis data, analisis metode, analisis kebutuhan non fungsional dan fungsional, serta analisis basis data. Selain analisis sistem, bab ini terdapat juga perancangan antarmuka untuk sistem yang akan dibangun sesuai dengan hasil analisis yang telah dibuat.

BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini menjelaskan tentang implementasi dan pengujian dari sistem yang dibangun berdasarkan analisis dan perancangan sistem yang telah dilakukan. Hasil dari implementasi kemudian dilakukan pengujian fungsional dari sistem tersebut menggunakan metode *blackbox* yang terdiri dari pengujian alpha dan beta sehingga sistem yang dibangun sesuai dengan analisis dan perancangan yang telah dilakukan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran yang sudah diperoleh dari hasil penulisan tugas akhir yang perlu diperhatikan bagi pengembangan perangkat lunak di masa yang akan datang.