

1. Pendahuluan

1.1. Latar Belakang

Berita *online* saat ini merupakan sesuatu yang sangat umum di kalangan masyarakat Indonesia bahkan dunia. Hal ini disebabkan karena berita *online* merupakan sarana penyebaran informasi yang mudah dan cepat serta fleksibel, dapat diakses kapan saja, dan dimana saja [1].

Dalam perkembangannya berita *online* yang telah tersimpan pada suatu penyimpanan data mencapai puluhan hingga ratusan miliar berita. Dari miliaran data berita tersebut diperlukan suatu permodelan, agar memudahkan proses pencarian, manipulasi atau pengolahan data tersebut. Salah satu model yang cocok untuk data berita tersebut adalah model *graph*, karena dengan menggunakan model *graph* akan lebih mudah dimengerti oleh pikiran manusia, selain itu model *graph* juga sangat cocok digunakan untuk data yang berskala besar dan juga yang memiliki sifat semi terstruktur [2] [3] seperti data berita *online*.

Pengelompokan berita berdasarkan kemiripan atau keterkaitan isi beritanya adalah salah satu kebutuhan dalam berita *online*, sehingga akan memudahkan pembaca berita *online* tersebut ketika ingin mencari berita yang terkait dengan berita yang sedang dibacanya. Untuk itu diperlukan sebuah metode yang disebut *clustering*, pada dasarnya *clustering* sendiri bertujuan untuk membuat kelompok-kelompok berdasarkan kemiripan elemen di dalamnya [4] [5]. Salah satu teknik *clustering* yang ada adalah *graph clustering*. *Graph clustering* adalah proses pengelompokan *node* dalam sebuah *graph*, sehingga *node* tersebut berada dalam satu *cluster*/kelompok dengan *node* lainnya dengan sifat yang sama [6].

Algoritma *graph clustering* sangatlah banyak, diantaranya MST *Clustering*, *Chameleon*, *Makarov Clustering*, dan *Star Clustering* [7]. Namun, dalam tugas akhir ini penulis memilih menggunakan algoritma *Chinese Whisper* sebagai dasar untuk membentuk *cluster* berita sesuai dengan tingkat keterkaitan isi berita. Hal ini dikarenakan algoritma *Chinese Whisper* mampu membentuk *cluster* dari data *graph* yang besar dengan waktu yang relatif cepat [8], sehingga sangat cocok digunakan untuk kasus *clustering* berita *online*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, dirumuskan beberapa pertanyaan penelitian, antara lain :

1. Bagaimana memodelkan data berita *online* ke dalam model *graph* ?
2. Apakah algoritma *Chinese Whispers* dapat diterapkan untuk pengelompokan berita *online* berdasarkan keterkaitan isi berita ?
3. Apakah hasil *clustering* menggunakan algoritma *Chinese Whispers* mempunyai nilai *intra-cluster* yang lebih besar daripada nilai *inter-cluster*-nya ?
4. Bagaimana akurasi algoritma *Chinese Whispers* dalam pengelompokan berita berdasarkan keterkaitan isi berita ?

1.3. Batasan Masalah

Batasan masalah pada pengerjaan tugas akhir ini antara lain :

1. Data pengujian yang digunakan untuk menguji sistem ini adalah data berita *online* dengan menggunakan bahasa Indonesia.
2. Metode pembobotan keterkaitan berita dengan menggunakan metode *Cosine Similarity*, karena *cosine similarity* adalah metode pengukuran kemiripan dokumen yang paling umum dan paling sering digunakan dalam *text mining* [9].
3. Jenis *graph* yang digunakan adalah *graph* berbobot dan tidak berarah.

1.4. Tujuan

Berdasarkan rumusan masalah diatas, didapatkan beberapa tujuan dari pengerjaan tugas akhir ini, antara lain :

1. Memodelkan file berita online kedalam model *graph*.
2. Menerapkan algoritma *Chinese Whispers* pada pengelompokan berita berdasarkan keterkaitan isi berita.
3. Mengetahui apakah hasil *clustering* dari algoritma *Chinese Whispers* sudah mempunyai nilai *intra-cluster* yang lebih tinggi daripada nilai *inter-cluster*.
4. Mengetahui tingkat akurasi algoritma *Chinese Whispers* dalam pengelompokan berita berdasarkan keterkaitan isi berita.

1.5. Metodologi Penyelesaian Masalah

Adapun tahap-tahap metodologi penyelesaian masalah, yaitu:

1. Studi Literatur
Pada tahap ini dilakukan pencarian beberapa literatur mengenai *Graph Database*, *Graph Clustering*, dan algoritma *Chinese Whispers*.
2. Pengumpulan Data
Pada tahap ini akan dilakukan pengumpulan data dengan *crawling* berita berita online dari beberapa website penyedia berita seperti kompas, okezone, dan lainnya.
3. *Preprocessing*
Pada tahap ini akan dilakukan proses *text mining* guna mencari bobot keterkaitan antar berita yang akan digunakan sebagai acuan dalam pembentukan data berita kedalam model *graph*.
4. Implementasi Metode
Model *graph* yang telah dibuat pada tahap *preprocessing*, selanjutnya akan dilakukan *clustering* sesuai dengan metode/algoritma *Chinese Whispers* sehingga node berita yang tadinya tidak ter-*cluster* menjadi ter-*cluster* sesuai dengan keterkaitan isi beritanya.

5. Pengujian Sistem
Setelah pengimplementasian metode, tahap selanjutnya adalah pengujian sistem. Pada tahap pengujian sistem akan memasukkan data sampel dari populasi data yang ada sebagai bahan untuk tahap analisis dan kesimpulan.
6. Analisis dan Kesimpulan
Setelah melakukan pengujian sistem dengan data sampel, selanjutnya adalah melakukan analisis terhadap hasil pengujian serta penarikan kesimpulan dari hasil analisis tersebut.
7. Pembuatan Laporan
Pada tahap ini dilakukan penyusunan laporan Penelitian yang mencakup hasil analisis yang telah dilakukan. Adapun hasil akhir dari tahap ini adalah Buku Laporan Tugas Akhir.