

# 1. Pendahuluan

## 1.1 Latar belakang

Berkembangnya dunia teknologi dan informasi yang semakin pesat menyebabkan kebutuhan mendapatkan informasi juga harus cepat. Informasi merupakan hal yang penting dan sangat dibutuhkan oleh orang banyak. Salah satu bentuk informasi adalah berupa artikel dan berita *online*. Proses untuk mendapat informasi dari sebuah dokumen membutuhkan waktu yang lama karena manusia harus melakukan ekstraksi/abstraksi manual terhadap dokumen. Hal ini mengurangi efektivitas dan efisiensi perolehan informasi di tengah data yang melimpah. Untuk menanggapi pertumbuhan yang pesat dari informasi maka peringkasan teks dapat menolong manusia untuk mengekstrak inti dari informasi tersebut.

Peringkasan teks otomatis (*automatic text summarization* atau *ATS*) adalah pembuatan ringkasan dari sebuah teks secara otomatis dengan memanfaatkan aplikasi yang dijalankan pada komputer. Terdapat dua pendekatan pada peringkasan teks, yaitu ekstraksi dan abstraksi. Pendekatan ekstraksi, pertama kali diperkenalkan oleh Luhn (1958)[7]. Luhn menggunakan teknik statistik sederhana untuk menentukan kalimat yang paling signifikan dalam suatu dokumen. Kalimat ini kemudian diekstrak dari dokumen dan dijadikan ringkasan.

Tugas akhir ini metode peringkasan berbasis graf menggunakan pendekatan ekstraksi. *Graph-based summarization algorithm* atau peringkasan teks berbasis graf merupakan suatu metode peringkasan teks yang *language independent* dan dapat menghasilkan ringkasan ekstraktif. Teks sumber direpresentasikan menjadi sebuah graf sehingga disebut graf tekstual [13]. *Node* pada graf tersebut dapat berupa unit-unit teks seperti kata-kata, kalimat-kalimat, atau paragraf-paragraf dalam teks. *Edge* dalam graf menunjukkan keterhubungan antar *node*. Keterhubungan dapat berupa *similarity* antar kalimat ataupun hubungan leksikal atau gramatikal antar kata/frasa. Konsep *similarity* antar unit teks digunakan dalam proses pembangunan graf tekstual.

Beberapa metode untuk ekstraksi, kalimat yang diekstrak berdasarkan relevansi dengan topik dokumen. Akan tetapi metode-metode tersebut mungkin menghasilkan ringkasan dengan keterbacaan rendah[2]. Ringkasan yang baik adalah setiap kalimat dalam ringkasan terkait dengan topik, mudah dibaca dan koheren (yaitu dapat dilihat sebagai dokumen kecil). Dalam sebuah ringkasan mungkin saja kalimat tersebut sangat tinggi keterhubungannya dengan topik dan keterbacaannya sangat rendah sehingga sulit dibaca. Akan tetapi ada juga ringkasan yang mempunyai keterhubungan dengan topik sangat rendah dan keterbacaannya sangat tinggi. Karena adanya proses *trade offs* antara ketiga faktor tersebut maka pembuatan ringkasan dapat dikatakan sebuah masalah optimasi[2].

Pada kasus peringkasan teks ini merupakan masalah optimasi maka dapat digunakan algoritma optimasi. Salah satu algoritma optimasi adalah algoritma *Harmony Search* (HS). *Harmony Search* adalah algoritma metaheuristik yang berbasis populasi. HS menirukan evolusi yang terjadi pada proses pertunjukan musik, misalnya improvisasi jazz yang berusaha mencari harmoni lebih baik. Dengan analogi tersebut, HS melakukan proses optimasi untuk mendapatkan

keadaan terbaik (optimum global) dengan cara mengevaluasi fungsi objektif. Dengan menggunakan himpunan pola-pola titik nada (*pitches*) yang dikeluarkan oleh alat-alat music, fungsi objektif pada HS dihitung menggunakan himpunan nilai-nilai yang diberikan untuk variable-variabel keputusan (*decision variables*). Jika kualitas suara estetika dapat diperbaiki melalui latihan-latihan demi latihan maka nilai fungsi objektif juga dapat terus ditingkatkan dari iterasi ke iterasi [11]. Algoritma *Harmony Search* ini dapat digunakan untuk menemukan solusi ringkasan yang optimum global.

Dengan menggunakan algoritma HS kalimat-kalimat yang ada pada dokumen akan diekstrak sesuai panjang ringkasan yang diinginkan. Dengan menggunakan himpunan kalimat-kalimat ringkasan yang dikombinasikan, fungsi objektif dihitung untuk menentukan kualitas ringkasan. Kualitas dari ringkasan dapat ditingkatkan dengan mengekstrak kalimat-kalimat ringkasan yang baru sehingga nilai fungsi objektif dapat ditingkatkan dari iterasi ke iterasi.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang diangkat pada tugas akhir ini :

1. Bagaimana Algoritma *Harmony Search* dapat menghasilkan peringkasan dokumen dengan menggunakan pendekatan ekstraksi.
2. Bagaimana performansi karakteristik parameter *Harmony Search* yaitu *Harmony Memory Considering Rate* dalam menemukan solusi ringkasan yang optimal.

## 1.3 Tujuan

Tujuan tugas akhir ini adalah:

1. Mengimplementasikan algoritma *Harmony Search* untuk pembuat peringkasan teks otomatis pada dokumen tunggal berbahasa Indonesia, khususnya artikel berita.
2. Menganalisa performansi karakteristik parameter *Harmony Search* yaitu *Harmony Memory Considering Rate* dalam hal optimasi untuk mendapatkan ringkasan.

## 1.4 Hipotesa

Pada peringkasan teks untuk parameter *Harmony Memory Considering Rate* (HMCR) semakin mendekati 1 hasil *F-Measure* yang diperoleh tidak terlalu bagus.

## 1.5 Batasan Masalah

Terdapat beberapa batasan dalam penelitian tugas akhir ini, antara lain:

1. Teks sumber yang akan diringkas adalah artikel berita berbahasa Indonesia yang diperoleh dari *website* berita.
2. Peringkasan dilakukan secara *offline*.
3. Peringkasan yang dilakukan adalah peringkasan dokumen tunggal.
4. Hasil peringkasan berupa ekstraksi dari teks sumber.

5. Tidak melakukan *stemming* terhadap teks masukan.
6. Tanpa mengeliminasi *stopwords*.
7. Evaluasi dilakukan dengan membandingkan *content overlap* antara peringkasan otomatis dan ringkasan referensi dengan menggunakan ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) *evaluation toolkit* [1].

## 1.6 Metodologi penyelesaian masalah

Metodologi yang dilakukan untuk menyelesaikan permasalahan adalah sebagai berikut:

1. Identifikasi masalah, yakni melakukan identifikasi masalah yang akan di analisa pada tugas akhir ini
2. Studi literatur, tahapan eksplorasi dan studi literatur terhadap *automatic text summarization* dengan algoritma *Harmony Search*. Tahap ini melakukan studi literatur seperti jurnal, website, maupun artikel dan bacaan yang relevan.
3. Analisa dan pengumpulan data yang digunakan sebagai inputan :  
Data yang digunakan sebagai inputan berupa single dokumen teks. Dokumen tersebut akan direpresentasikan sebagai *Directed Acyclic Graph* (DAG). Setiap kalimat akan direpresentasikan sebagai *node* dan *edge* sebagai penghubung.
4. Analisa dan perancangan pembangunan Algoritma *Harmony Search* berdasarkan langkah-langkah berikut :
  - a. Inisialisasi permasalahan dan parameter *Harmony Search* (HS)  
Pada tahap ini dilakukan pendefinisian permasalahan yaitu meminimalisasi nilai  $f(x)$  dan pendefinisian parameter-parameter yang dibutuhkan algoritma *Harmony Search*, parameter-parameter tersebut antara lain : *Harmony Memory Search* (HMS), *Harmony Memory Considering Rate* (HMCR), *Pitch Adjustment Rate* (PAR), dan *Number of Improvisation* (NI)
  - b. Inisialisasi *Harmony Search*.  
Solusi yang dihasilkan *Harmony Search* akan disimpan dalam suatu memori yang disebut dengan *Harmony Memory* (HM).
  - c. Improvisasi *Harmony* Baru.
  - d. Update *Harmony Memori*.
5. Implementasi dan pembangunan sistem.
6. Pengujian dan analisa hasil
  - a. *Testing* sistem, menguji sistem dengan menggunakan data berupa single dokumen teks.
  - b. Analisa hasil, performansi ringkasan ditinjau dengan membandingkan antara hasil peringkasan otomatis dan ringkasan referensi dengan menggunakan ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) *evaluation toolkit*.
  - c. Analisa parameter HS dapat mengoptimasi ringkasan dengan optimal.

7. Pembuatan laporan, melakukan pelaporan hasil pengerjaan tugas akhir berupa analisis sistem yang dibangun beserta dokumentasinya serta kesimpulan akhir.