

# 1. Pendahuluan

## 1.1 Latar Belakang

Peringkasan terhadap sebuah teks merupakan salah satu cara untuk mendapatkan informasi, dengan berkembangnya teknologi peringkasan teks dapat dilakukan secara otomatis. Peringkasan teks dilakukan untuk mendapatkan teks yang lebih pendek dari teks aslinya agar tercipta sebuah ringkasan yang mempertahankan poin yang paling penting dari dokumen asli sehingga pembaca dengan mudah menangkap inti dari teks dan tetap berkaitan dalam kategori teks tersebut.

Majalah, artikel, atau teks tentang olahraga terkadang memiliki sedikit kalimat yang memang berkaitan dalam bidang olahraga, sehingga pembaca perlu membaca keseluruhan isi teks untuk mendapatkan inti dari teks tersebut. Dengan melakukan peringkasan maka berita atau teks tentang olahraga tersebut dapat diperoleh dengan mudah. Kebanyakan orang meringkas hanya melihat kata yang sering muncul tanpa memperhatikan keterkaitan dari kata tersebut, namun ada juga yang melakukan peringkasan yang hanya melihat keterkaitan kata tersebut terhadap kategori tertentu.

Peringkasan yang dilakukan dengan metode TF (*Term Frequency*) melakukan peringkasan berdasarkan nilai kemunculan dari sebuah kata saja, namun hasil dari ringkasan kemungkinan tidak sesuai dengan kategori, sedangkan metode yang lain seperti IDF (*Inverse Document Frequency*) memperhatikan kemunculan kata pada kumpulan dokumen. Acuan penelitian ini berdasarkan penelitian *Term Frequency* oleh Hans Peter Luhn pada tahun 1957 dan penelitian *Inverse Document Frequency* oleh Karen Sparck Jones pada tahun 1972. Dalam penelitian ini digunakan metode *Extractive Summarization* yaitu dengan cara mengambil kalimat asli dari teks yang diringkas. Pada *Text Summarization* terdapat dua pendekatan dalam menangani hal ini yaitu, NLP (*Natural Language*

*Processing*) dan IR (*Information Retrieval*). Sebagian besar pendekatan yang dilakukan adalah IR karena komputasi yang lebih cepat dari NLP sehingga penulis menggunakan pendekatan IR. Untuk mendapatkan hasil ringkasan yang memiliki kemunculan kata yang banyak dan berkaitan dengan kategori tertentu dapat berdasarkan metode TF dan IDF. Berdasarkan kedua metode tersebut sangat sulit menentukannya karena kedua hal tersebut sangat penting dalam melakukan peringkasan. Maka dari itu, *Fuzzy System* diterapkan di dalam peringkasan teks ini.

Terdapat beberapa pendekatan yang dilakukan dalam *Text Summarization* misalnya *Graph-based* yang merepresentasikan teks sebagai *graph*. Selain itu ada metode *Lexical-Chain* yang membuat ikatan antar kata untuk merepresentasikan teks. *Fuzzy System* adalah metode penalaran yang menyerupai metode penalaran manusia yang dapat menangani hal yang tidak pasti. *Fuzzy System* pertama kali diperkenalkan oleh Lotfi Zadeh pada tahun 1965. Di tahun 2015, ada jurnal *Text Summarization Using Fuzzy Relational Clustering Algorithm* yang membahas mengenai penggunaan *cluster* pada kata yang berkaitan, dilakukan dengan menggabungkan kata yang memiliki keterkaitan di dalam *cluster* yang sama. Pendekatan *Fuzzy System* mengimitasi metode penalaran manusia yang melibatkan nilai kemungkinan diantara “ ya ” dan “ tidak “. Pendekatan ini umumnya diterapkan pada masalah yang mengandung unsur ketidakpastian atau ketidaktepatan. Kemampuan *Fuzzy System* untuk memilih *term* yang dianggap penting sebagai hasil ringkasan dapat diterapkan pada *Text Summarization*. Hal ini dikarenakan *Fuzzy System* mengimitasi metode penalaran manusia, misalnya dalam *range possibilities* yang meliputi *Certainly Yes*, *Possibly Yes*, *Cannot Say*, *Possibly No*, dan *Certainly No*. Kapabilitas dari *Fuzzy System* ini dimanfaatkan pada *Text Summarization* untuk menghitung dan menentukan hasil akhir yang berupa ringkasan yang dikutip dari majalah, artikel, atau teks tentang olahraga tersebut.

Dalam tugas akhir ini, penulis mencoba untuk menganalisis kualitas ringkasan teks dengan menggunakan metode *Fuzzy System* berdasarkan *Term Frequency* dan *Inverse Document Frequency*.

## 1.2 Perumusan Masalah

Dari penjelasan di atas maka masalah yang dicari penyelesaiannya dalam tugas akhir ini adalah :

1. Bagaimana cara mendapatkan informasi pada sebuah teks yang panjang dengan mudah?
2. Bagaimana cara meringkas isi sebuah bacaan?

Dalam tugas akhir ini yang dibahas adalah *TextSummarization* menggunakan *FuzzySystem* berdasarkan *TermFrequency* dan *InverseDocumentFrequency* dengan batasan masalahnya sebagai berikut :

1. Aplikasi ini dikembangkan dengan Java.
2. Peringkasan teks yang dilakukan adalah artikel atau teks bahasa Indonesia yang berkaitan dengan bidang olahraga dengan kategori sepak bola dan basket.
3. Peringkasan yang dilakukan bersifat *ExtractiveSummarization*.
4. Pengujian yang dilakukan menggunakan skenario dengan *user judgement*.

## 1.3 Tujuan

Berdasarkan masalah yang telah diidentifikasi di atas, maka tujuan yang ingin dicapai dalam tugas akhir ini adalah :

1. Mendapatkan suatu teks yang lebih ringkas dari teks yang asli.
2. Hasil ringkasan tetap berkaitan dengan kategori dalam bidang olahraga.

## 1.4 Metodologi penyelesaian masalah

Pada aplikasi ini telah dilakukan beberapa langkah seperti :

1. Persiapan data

Data yang dipersiapkan adalah kata-kata yang berkaitan dalam bidang olahraga yang berisikan nilai keterkaitannya dan beberapa referensi yang menjadi acuan aplikasi ini.

2. Pengembangan aplikasi serta pembangunan model

Aplikasi ini dikembangkan dengan batasan yang telah ditentukan dan melalui tahapan seperti, *preprocessing*, *processing* dan proses *output*.

3. Implementasi

Implementasi aplikasi ini disimulasikan dengan Java.

4. Pengujian

Pengujian dilakukan dengan melakukan peringkasan teks atau artikel yang berkaitan dengan bidang olahraga sehingga mendapatkan ringkasan yang lebih pendek dan menggunakan skenario *userjudgement*.

5. Analisa hasil

Hasil dari ringkasan tersebut telah dianalisa dengan memperhatikan panjang teks yang asli dengan *output* yang dihasilkan serta menjaga inti dari kalimat agar lebih spesifik dan jelas.

6. Pembuatan laporan

Setelah melakukan semua tahapan yang telah ditentukan tahapan berikutnya adalah pembuatan laporan mengenai *Text Summarization* berdasarkan implementasi, pengujian, analisis serta teori-teori yang terkait dalam pengembangan aplikasi ini.