

# 1. Pendahuluan

## 1.1 Latar Belakang

Kata kunci memiliki peranan penting dalam memahami isi teks secara keseluruhan. Kata kunci akan membantu pembaca menemukan inti/topik dari suatu teks dengan cepat tanpa harus membaca keseluruhan isi teks. Karena kata kunci adalah unit terkecil yang mengekspresikan makna keseluruhan dokumen [19]. Tujuan dari mengekstrak kata kunci adalah menemukan kumpulan kata-kata yang merepresentasikan makna dari sebuah dokumen. Sehingga dengan mengetahui kata kuncinya saja, diharapkan pembaca dapat mengetahui arah alur topik yang sedang dibahas dalam dokumen tersebut. Banyak aplikasi mengambil keuntungan ini seperti pengindeksan secara otomatis, peringkasan teks, *Information Retrieval*, klasifikasi, *clustering*, *filtering*, katalogisasi, deteksi topik, dan lain-lain. [19]

Dalam melakukan ekstraksi kata kunci, akan ditemukan kemungkinan-kemungkinan bahwa kata kunci tidak selalu dalam bentuk satu kata, melainkan bisa juga dalam 2 atau 3 kata, yang sering disebut sebagai frase. Pada Tugas Akhir ini, dilakukan penanganan identifikasi frase. Frase yang digunakan hanyalah frase nomina saja. Frase Nomina (FN) adalah frase *endosentris* berinduk satu yang induknya nomina dengan kata lain frase nomina adalah frase yang intinya kata benda. Contoh : Pohon cemara, pidato presiden, bahasa Indonesia, sekolah dasar, sepeda gunung, baju kebaya, tas sekolah, dunia maya, demam berdarah, algoritma genetika, dan lain-lain.

Ada banyak metode yang dapat dilakukan dalam mengekstrak kata kunci diantaranya yaitu *Naive Bayes*, *Text Rank*, *Support Vector Machine (SVM)*, *Lexical Chain*, *Hidden Markov Model (HMM)*, dan lain-lain. Pada Tugas Akhir ini, digunakan Algoritma *Hidden Markov Model (HMM)* untuk menentukan kata kunci (baik dalam bentuk kata, maupun frase). *Hidden Markov Model (HMM)* merupakan sebuah model dari sebuah sistem yang diasumsikan sebuah proses *Markov* dengan parameter yang tak diketahui, dan tantangannya adalah menentukan parameter-parameter tersembunyi (*hidden*) dari parameter-parameter yang dapat diamati (*observer*). Berdasarkan *state-state* yang dapat diamati, dapat diketahui *state-state* yang tersembunyi. Pada Tugas Akhir ini, *state* yang menjadi *observed state* adalah tiap kata pada abstrak *paper* hasil *preprocessing*, sedangkan *hidden state* adalah kata kunci dari abstrak *paper* tersebut.

## 1.2 Perumusan masalah

Dari latar belakang di atas, permasalahan yang akan diteliti yaitu:

1. Bagaimana mengidentifikasi bahwa suatu kata merupakan kata kunci atau bukan.
2. Bagaimana keakuratan kata kunci yang dihasilkan menggunakan metode *Hidden Markov Model (HMM)*.

Adapun batasan masalah dari Tugas Akhir ini adalah sebagai berikut :

1. Pencarian kata kunci hanya dilakukan pada teks berbahasa Indonesia yang menggunakan bahasa baku.
2. Teks yang digunakan adalah abstrak *paper* pada [digilib.ittelkom.ac.id](http://digilib.ittelkom.ac.id) yang dilakukan secara manual.
3. Menggunakan proses *pos tagging* yang sudah ada.
4. Sistem yang dibangun tidak terhubung ke internet.
5. Data input untuk proses *preprocessing* adalah file.txt sedangkan untuk proses *training* dan proses *testing* juga berbentuk file.txt
6. Pencarian *dataset* frase dilakukan secara manual berdasarkan data abstrak yang ada.
7. Frase yang digunakan hanyalah frase nomina (frase kata benda).
8. Proses pembagian data *training* dan data *testing* dilakukan secara manual dan random.

### 1.3 Tujuan

Tujuan yang ingin dicapai pada Tugas Akhir ini adalah sebagai berikut :

- a. Men-*generate* kata kunci baik berbentuk kata maupun frase dari teks berbahasa Indonesia menggunakan *Hidden Markov Model*.
- b. Mengevaluasi hasil akurasi ekstraksi kata kunci yang dihasilkan metode *Hidden Markov Model*.

### 1.4 Metodologi penyelesaian masalah

Metodologi yang akan digunakan dalam menyelesaikan Tugas Akhir ini adalah :

1. Studi Literatur  
Mempelajari landasan teori dan literature dari jurnal, *paper*, artikel, dan buku yang berkaitan dengan *keyword extraction*, *keyphrase extraction*, *information extraction*, *text mining*, *Hidden Markov Model (HMM)*, *summarization*, *Viterbi Algorithm* dan hal lain yang berkaitan dengan Tugas Akhir ini.
2. Pengumpulan Data  
Data yang dikumpulkan berupa abstrak *paper* pada [digilib.ittelkom.ac.id](http://digilib.ittelkom.ac.id) yang dilakukan secara manual.
3. Perancangan Sistem dan Implementasi  
Pada tahap ini dilakukan perancangan sistem menggunakan metode terstruktur dalam pembangunan sistem.
4. Pengujian dan Analisis Hasil  
Pada tahap ini dilakukan pengujian terhadap data yang telah terkumpul dan menganalisa *Hidden Markov Model (HMM)* dalam ekstraksi kata kunci.
5. Pengambilan Kesimpulan dan Pembuatan Laporan  
Membuat kesimpulan dari hasil yang telah dilakukan serta membuat laporan sesuai dari langkah-langkah pertama sampai akhir dalam melakukan penelitian Tugas Akhir ini.