

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kebutuhan manusia terhadap teknologi memacu cepatnya perkembangan teknologi itu sendiri. Kemajuan teknologi yang semakin meningkat didukung dengan sarana dan prasarana yang memadai, ini membuktikan bahwa dewasa ini teknologi menjadi kebutuhan pokok bagi kehidupan manusia.

Dinamika perkembangan Ilmu Pengetahuan dan Teknologi saat ini sangat cepat salah satunya adalah teknologi informasi dibidang *search engine*. Pembuatan jurnal merupakan salah satu syarat untuk kelulusan bagi mahasiswa mahasiswi Universitas maupun sekolah tinggi, sehingga membutuhkan pengelolaan yang lebih baik, meski saat ini sudah banyak penyedia portal jurnal.

Sudah banyak peneliti melakukan riset tentang *search engine*, memakai dan mengembangkan metode-metode *search* dengan melihat pentingnya fungsi dari *search* itu sendiri. Banyak metode yang bisa dipakai dan dikembangkan untuk mengoptimalkan penelitian tentang *search* dan sistem rekomendasi, seperti yang sudah dilakukan oleh S. Palgunadi[1] yang menerapkan metode n-gram dan implementasi algoritma *tanimoto cosine*.

Di lingkungan Sekolah Tinggi Elektronika dan Komputer (STEKOM), pengelolaan tugas akhir dan skripsi belum dikelola dengan baik, dimana dalam kasus tertentu, terjadi banyak *copy paste* terhadap penulisan karya ilmiah yang tanpa mencantumkan referensi yang jelas. Untuk itu maka perlu adanya sebuah sistem yang dapat mengelola tugas akhir dan skripsi agar lebih baik.

Penelitian ini membangun sistem untuk menghitung nilai kemiripan antara satu artikel atau jurnal dengan kata kunci yang lain dengan menggunakan metode *n-gram* dan *tanimoto coefficient*. Mengacu pada penelitian sebelumnya yang dilakukan oleh D. Purwitasari[2] dan penelitian yang dilakukan oleh S. Palgunadi[1] yang menerapkan n-gram serta menggabungkan algoritma *weighted tree similarity* dan *tanimoto cosine*, menghasilkan akurasi hingga 100%. Berawal dari itu, penelitian ini menerapkan n-gram yang merupakan salah satu cara untuk memecah kalimat atau kata menjadi n-karakter termasuk tanda baca dan spasi, kemudian pencocokan menggunakan algoritma *tanimoto coefficient* atau bisa juga disebut *Jaccard coefficient*. *Tanimoto cosine* merupakan algoritma gabungan dari *tanimoto similarity* dengan *cosine similarity*, dengan harapan mampu menghasilkan perhitungan

presisi, akurasi serta recall yang lebih baik jika dibandingkan dengan penggunaan satu algoritma. Pemilihan algoritma *tanimoto coefficient* mengacu pada penelitian sebelumnya yang ditulis oleh S. Palgunadi[1], penelitian tersebut mengacu pada hasil dari penelitian oleh Sharma(2011), dalam tulisan tersebut dikatakan bahwa penggunaan *tanimoto cosine* terbukti paling akurat[3].

## **1.2 Rumusan Masalah**

Berdasarkan uraian yang dikemukakan diatas, dapat dirumuskan permasalahan yaitu sebagai berikut :

1. Bagaimana menghitung kemiripan antara satu artikel atau jurnal dengan artikel atau jurnal yang lain.
2. Bagaimana membangun serta mengembangkan sistem rekomendasi pencarian artikel berbahasa indonesia.

## **1.3 Tujuan Penelitian**

Adapun yang menjadi tujuan dari penelitian adalah :

1. Menghitung kemiripan antara satu artikel dengan kata kunci yang lain dengan menggunakan n-gram dan *tanimoto coefficient* diharapkan dapat menghasilkan akurasi, presisi dan recall yang lebih baik.
2. Mengembangkan *system* rekomendasi pencarian artikel berbahasa Indonesia.

## **1.4 Batasan Masalah**

Adapun yang menjadi ruang lingkup permasalahan adalah sebagai berikut:

1. Sistem rekomendasi khusus untuk internal STEKOM cabang Kendal-Weleri yaitu segenap Civitas Akademik.
2. Analisa pengujian sistem menggunakan *confusion matrix*

## 1.5 Keaslian Penelitian

Mengacu pada penelusuran terhadap penelitian tentang sistem rekomendasi pencarian artikel berbahasa Indonesia menggunakan *n*-gram dan *tanimoto coefficient*, belum ditemukan judul penelitian yang sama, beberapa penelitian sebelumnya yang sudah pernah dilakukan sebelumnya yang berkaitan dengan perancangan sistem rekomendasi dan penerapan algoritma yang dipakai dalam penelitian ini adalah :

1. Dalam penelitian D. Purwitasari (2011) menerapkan *n*-gram dan algoritma winnowing untuk mencari dokumen *fingerprinting* dengan mengubah rangkaian *N*-gram dari teks menjadi sekumpulan nilai-nilai *hash*.
2. Penelitian berikutnya menggabungkan antara algoritma *weighted tree similarity* dengan *cosine similarity* dan efektif diimplementasikan untuk pencarian semantic, penelitian tersebut dilakukan oleh Riyanarto sarno (2008).
3. Penelitian lain yang berkaitan dengan penelitian ini ditulis oleh Wayan Firdaus Mahmudy (2007), penelitian tersebut menggunakan algoritma gabungan yaitu *decision tree* dan algoritma *k-means clustering*. Pengukuran tingkat akurasi dilakukan dengan membandingkan data asli dengan hasil prediksi yang didapatkan.
4. Adi dan Palgunadi (2014) menerapkan Kombinasi Algoritma *Weighted Tree Similarity* dengan *Tanimoto Cosine* untuk Pencarian Semantik yang menyimpulkan jika hasil penelitian yang telah dilakukan dan dipaparkan bahwa kombinasi *weighted tree similarity* dengan *Tanimoto Cosine* mencapai 100% dan recall 84,44%.

Banyaknya penelitian yang berkaitan dengan penelitian ini maka penulis perlu melakukan penekanan dan pembeda dari penelitian terkait sistem rekomendasi pencarian artikel berbahasa Indonesia menggunakan *n*-gram dan *tanimoto coefficient*.

## **1.6 Sistematika Penulisan**

Tesis ini terdiri dari 5 (lima) bab yang disusun sedemikian rupa dengan materi pembahasan yang saling berhubungan dengan sistematika sebagai berikut:

### **BAB I : PENDAHULUAN**

Merupakan latar belakang, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian S sistematika penulisan.

### **BAB II : KAJIAN PUSTAKA**

Berisi tentang kajian teori yang digunakan untuk mendukung penyusunan thesis, antara lain : teori secara umum, teori secara khusus, dan teori pendukung.

### **BAB III : METODOLOGI PENELITIAN.**

Berisi uraian tentang data dan metode yang digunakan untuk merancang system rekomendasi pencarian artikel berbahasa Indonesia menggunakan n-gram dan tanimoto cosine.

### **BAB IV : HASIL PENELITIAN DAN PEMBAHASAN**

Berisi uraian tentang hasil analisis dan hasil penelitian serta pengujian system rekomendasi pencarian artikel berbahasa indonesia.

### **BAB V : KESIMPULAN DAN SARAN**

Berisi uraian singkat tentang kesimpulan dari pembahasan penelitian dan saran untuk pengembangan penelitian berikutnya.

## **DAFTAR PUSTAKA**

## **LAMPIRAN**