BAB I
Pendahuluan

1.1. Latar Belakang

Sitasi adalah daftar pustaka dari sejumlah dokumen yang dirujuk atau yang dikutip oleh sebuah dokumen dan setiap daftar pustaka dokumen tersebut dimuat dalam bibliografi dokumen yang mengutip, yang secara khusus mengkaji pengarang dan karya-karya lain. Bisa juga di definisikan untuk menunjukkan asal-usul atau sumber suatu kutipan, mengutip pernyataan atau menyalin/mengulang pernyataan seseorang dan mencantumkannya di dalam suatu karya tulis yang dibuat, namun tetap mengindikasikan bahwa kutipan tersebut itu adalah pernyataan orang lain [1].

Kalimat sitasi sering digunakan pada artikel atau paper ilmiah dengan maksud mereferensikan pekerjaan ilmiah seseorang yang telah dipublikasikan seperti, jurnal, buku, paper, dan makalah. Salah satu contoh sitasi adlah jika sebuah kalimat diikuti dengan kurung siku maka kalimat tersebut adalah sebuah kalimat sitasi [2]. Kalimat sitasi juga mempunyai peran penting dalam perkembangan ilmu pengetahuan. Ketika sebuah referensi muncul di sebuah artikel ilmiah, biasanya diikuti dengan teks yang berisi tentang peranan penting paper yang telah dikerjakan [3].

Pengidentifikasian kalimat sitasi berguna untuk membantu para akademisi untuk mengecek kembali jurnal ilmiah yang ditulis. Kecerobohan akademisi untuk menggunakan sitasi dapat menjadi fatal karena dapat dianggap melakukan plagiarisme. Selain itu pengidentifikasian sitasi juga berguna bagi ilmuwan untuk melihat seberapa jauh sebuah makalah ilmiah telah berkembang.

Jurnal ilmiah menjadi aset suatu institusi atau lembaga pendidikan sebagai wadah untuk mempublikasikan hasil penelitian dari dosen, mahasiswa, dan praktisi. Jika penelitian pengidentifikasian sitasi didalami lebih lanjut, ilmuwan dapat mengetahui apakah makalah ilmiah yang dia tulis sering direferensikan. Semakin sering sebuah penelitian direferensikan oleh penelitian lain maka penelitian tersebut berguna dalam pengembangan topik yang sedang diteliti.

Supervised learning adalah pembelajaran yang menggunakan dua jenis data yaitu *training dataset* dan *testing dataset*. *Supervised learning* mempelajari data *training* sehingga mendapatkan hasil yang paling baik di *data testing* [4]. Pada penelitian ini supervised learning yang digunakan adalah *naive bayes* (NB) dan *support vector machine* (SVM)

1.2. Perumusan Masalah

Perumusan masalah yang diangkat dalam penelitian tugas akhir ini adalah:

A. Bagaimana performa dari penggunaan fitur dalam identifikasi kalimat sitasi?
B. Apa saja fitur yang baik untuk mengidentifikasi kalimat sitasi?

1.3. Batasan Masalah

Batasan – batasan permasalahan yang digunakan dalam penelitian ini antara lain:

A. Data yang digunakan adalah ACL *Anthology Reference Corpus* berjumlah 10 *paper*.

B. *Classifier* yang digunakan adalah yang *classifier* yang telah ada di aplikasi weka.

1. *Support Vector Machine.*(SVM)
2. *Naive Bayes.*(NB)

C. Pelabelan dibantu menggunakan *Regular Expression*.

1.4. Tujuan Penelitian

Untuk menjawab beberapa permasalahan yang ada, maka tujuan dari penelitian ini antara lain:

A. Mengetahui dan menganalisis performansi fitur yang dari pelatihan kalimat sitasi.

B. Mengetahui fitur dan *classifier* yang tepat untuk mengidentifikasi kalimat sitasi.

1.5. Hipotesis

Hipotesis dalam penulisan tugas akhir ini adalah penambahan fitur dapat memperbaik performansi dalam pengindentifikasian kalimat sitasi

1.6. Metodologi Penyelesaian Masalah

Untuk menyelesaikan tugas akhir ini digunakan metodologi:

A. Studi Literatur

Mencari dan menemukan referensi yang berhubungan dengan pengidentifikasian kalimat sitasi dan fitur-fitur yang dapat mengekstraksi kalimat pada suatu file yang telah ada. Referensi yang digunakan didapat dari jurnal, buku, dan sumber yang didapat di intenet.

B. Pengumpulan Data

Pencarian data dari kumpulan paper ACL-ARC yang didapat dari: http://web.eecs.umich.edu/~lahiri/acl_arc.html. Data tersebut akan dibagi menjadi dua bagian, kalimat yang mempunyai sitasi sebagai *positive instance* dan yang tidak memiliki sitasi sebagai *negative instance*.

C. Analisis dan Perancangan Sistem

Perancangan sistem identifikasi sitasi ini akan menggunakan *Regular Expression* untuk membantu melabeli kalimat sitasi dan akan dilakukan pengamatan secara manual agar data training yang dipunya menjadi lebih baik. Analisis yang dilakukan adalah analisis ekstraksi fitur, dan analisis performansi fitur dalam pengklasifikasian kalimat sitasi.

D. Implementasi Sistem

Pembangunan sistem dilakukan sesuai dengan hasil dari perancangan sistem yang didapatkan dengan hasil analisis sehingga dapat memberikan fitur dengan akurasi yang paling tinggi untuk memberikan fitur yang terbaik untuk mengidentifikasi kalimat sitasi.

E. Analisis Hasil Implementasi

Menganalisis fitur yang paling tepat untuk mengidentifikasi kalimat sitasi dan menganalisis tingkat akurasi yang didapatkan dari pengklasifikasian yang telah dilakukan.

F. Pembuatan Laporan

Dari analisis dan kesimpulan yang telah didapatkan, akan dibuat sebuah laporan tugas akhir.

1.7. Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi penjelasan tentang latar belakang dilakukannya penilitian, perumusan masalah batasan masalah, tujuan penelitian, metodologi penyelesaian maalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi dasar-dasar teori yang menjadi dasar dasar dari penilitian, yaitu tentang analisis penggunaan fitur pada identifikasi kalimat sitasi pada paper ilmiah.

BAB III PERANCANGAN SISTEM

Bab ini memaparkan tahapan yang dilakukan untuk membangun sistem.

BAB IV ANALISIS HASIL PENGUJIAN

Bab ini menjelaskan tentang proses pengujian dan analisis yang didapat dari sistem.

BAB V KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan dan saran untuk pengembangan selanjutnya.

1.8. Jadwal Kegiatan

No	Kegiatan	Bulan 1				Bulan 2				Bulan 3				Bulan 4				Bulan 5				Bulan 6			
1	Studi Literatur																								
2	Pengumpulan Data																								
3	Perancangan Sistem																								
4	Implementasi Sistem																								
5	Analisis Hasil Implementasi Sistem																								
6	Penyusunan Laporan																								