



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Awal mula fungsi komputer berkaitan dengan pengolahan informasi, di mana pengolahan informasi hampir eksklusif berhubungan dengan masalah aritmatika, tetapi komputer sekarang telah dipakai untuk banyak tugas yang tidak berhubungan dengan masalah aritmatika saja (Setyaji, 2009). Komputer dewasa ini telah menjadi perangkat pengolah informasi yang berkembang sangat pesat baik dari sisi perangkat keras / *hardware* maupun perangkat lunak / *software*.

Seiring berkembangnya teknologi dalam bidang komputer, terdapat berbagai macam *file* teks dengan format maupun ekstensi yang berbeda. Semakin banyaknya dokumen teks digital, diperlukan suatu metode *Information Retrieval* untuk menyaring data dari suatu dokumen yang diperlukan. *Information Retrieval* (pencarian informasi) merupakan suatu metode untuk menemukan kembali suatu data yang tersimpan pada sekumpulan dokumen, kemudian menyediakan informasi mengenai subyek yang dibutuhkan. Tujuan dari sistem *IR* ini adalah memenuhi kebutuhan informasi pengguna dengan dan pada waktu yang sama untuk mendapatkan dokumen yang dibutuhkan dengan waktu yang sesingkat mungkin (Pardede, 2013).

Metode *Cosine Similarity* dapat digunakan untuk mengukur kemiripan antara dua objek berdasarkan sudut yang dibentuk dari dua object pada *vector* (Zhang, 2007). Aturan *Cosine Similarity* juga merupakan bagian dari basis *IR* yang

dapat digunakan untuk menghitung relevansi dua buah objek yang dibobotkan dengan perhitungan dalam bentuk *vektor*.

Stemming merupakan suatu proses yang terdapat dalam sistem *IR / Information Retrieval* yang mentransformasikan kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya atau kata dasarnya (*root word*) dengan menggunakan aturan – aturan tertentu (Agusta, 2009).

Dalam pembuatan penelitian ini tentunya tidak lepas dari referensi jurnal-jurnal pendukung dari mahasiswa lain yang mendukung pembuatan penelitian ini. Penelitian-penelitian sejenis yang berhubungan dengan penelitian ini yaitu oleh Jansen.(2012) yang berjudul “Implementasi Algoritma Porter Pada Aplikasi Pencarian Kata”, Ledy Agusta.(2009) yang berjudul “Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia”.

Penelitian dari Jansen tersebut yang mendorong penulis untuk melakukan penelitian ini. Dari penelitian Jansen tersebut, peneliti ingin mengembangkan aplikasi tersebut dengan membuat tidak hanya aplikasi pencarian kata saja, namun dapat dikembangkan menjadi aplikasi relevansi / kesamaan kata dalam dokumen digital berbahasa Indonesia dengan metode *stemming* Nazief Adriani karena sesuai dengan hasil penelitian Ledy Agusta, algoritma ini memiliki presisi *stemming* lebih baik untuk Bahasa Indonesia, serta penggabungan dengan aturan *Cosine Similarity* untuk menilai suatu relevansi isi dokumen tersebut berdasarkan kata dasar yang membangun dokumen tersebut, tipe dokumen digital yang dapat dibaca terdapat dokumen bertipe *docx*, *rtf* dan *txt*.

1.2 Rumusan Masalah

Dengan latar belakang tersebut maka peneliti mendapatkan masalah-masalah seperti berikut ini

1. Bagaimana cara menilai relevansi / kesamaan isi antar dokumen digital satu dengan lainnya ?
2. Bagaimanakah penerapan metode Nazief – Adriani untuk proses *stemming* dalam Bahasa Indonesia ?
3. Bagaimana cara menilai ketepatan suatu algoritma *stemming* ?

1.3 Batasan Masalah

Dalam pembuatan aplikasi nanti terdapat batasan-batasan yang peneliti tetapkan sebagai berikut:

1. Teknik *Stemming* dilakukan untuk dokumen berbahasa Indonesia.
2. Dokumen yang dapat dihitung relevansi isinya merupakan bentuk dokumen digital berformat *.txt*, *.docx*, dan *.rtf*.
3. Konten dokumen digital yang akan diproses berupa teks.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. Menilai persentasi relevansi isi antar dua dokumen digital.
2. Menerapkan algoritma nazief-adriani pada relevansi kata berbahasa Indonesia dalam suatu dokumen digital.
3. Menerapkan aturan Paice dalam menentukan ketepatan index nilai suatu algoritma *stemming*.

1.5 Manfaat Penelitian

Manfaat dalam penelitian ini antara lain supaya pengguna dapat mengetahui relevansi antar dua dokumen digital berbahasa Indonesia dengan berbagai format, sehingga tingkat persentasi kemiripan isi antar dokumen digital tersebut dapat diketahui.

Bagi peneliti sendiri penelitian ini berguna dapat untuk mempelajari sebuah materi baru, yakni algoritma Nazief dan Adriani, aturan *Cosine Similarity*, serta evaluasi Paice yang baru di ketahui oleh penulis.

Bagi pembaca, penelitian ini dapat memberikan pengetahuan dan wawasan mengenai salah satu penerapan teknik *stemming* Nazief-Adriani, aturan *Cosine Similarity* serta evaluasi *Paice* dalam menilai suatu ketepatan algoritma *Stemming*. Selain itu dapat memberikan wawasan mengenai perancangan aplikasi dengan menggunakan bahasa pemrograman C#.

UMMN