

BAB II

TINJAUAN PUSTAKA

Pada bab ini akan membahas mengenai uraian singkat hasil-hasil penelitian atau analisis terdahulu yang ada hubungannya dengan permasalahan yang akan ditinjau dalam Tugas Akhir.

Contoh *text mining* yang telah digunakan yaitu Implementasi *text mining* untuk mendeteksi kemiripan dokumen (Evelin, 2006). Dalam dunia akademis maupun non akademis, terdapat banyak dokumen dalam bentuk teks. Dari sekian banyak dokumen teks yang ada, banyak diantara dokumen-dokumen tersebut yang memiliki isi yang mirip. Pembaca dapat beberapa kali membaca dokumen yang kurang lebih memiliki isi sama hanya karena judul yang berbeda. Kemiripan antar dokumen dapat dilakukan dengan menganalisis dokumen-dokumen tersebut secara harafiah, yaitu setiap dokumen dibaca, lalu dikelompokkan. Pencarian kemiripan dokumen secara manual, pasti membutuhkan waktu dan tenaga yang sangat banyak, apalagi dokumen yang dianalisis cukup banyak dan tebal. Untuk mengatasi masalah tersebut, dibutuhkan komputerisasi untuk menganalisis kemiripan dokumen teks. Metode yang dapat digunakan untuk mencari kemiripan dokumen adalah *text mining*. *Text mining* merupakan suatu teknologi untuk menemukan suatu pengetahuan yang berguna dalam suatu koleksi dokumen teks sehingga diperoleh tren, pola, atau

kemiripan teks bahasa alamiah. Salah satu metode text mining adalah *Frequent Term-Based Text Clustering*. Metode ini mengelompokkan dokumen teks berdasarkan kemiripan antar dokumen. Dokumen teks dikatakan mirip jika pasangan dokumen tersebut memiliki banyak kata yang sama. Proses yang dilakukan adalah menghilangkan stopword, lalu cari keterikatan kata-kata dalam koleksi dokumen dan memasukkan dokumen kedalam kelompok-kelompok awal. Pada pengelompokan awal ini, satu dokumen dapat masuk ke dalam beberapa kelompok. Oleh sebab itu perlu adanya perhitungan bobot suatu dokumen terhadap kelompok awal sebelum dokumen dimasukkan ke dalam kelompok akhir. Dokumen teks yang berada di dalam satu kelompok akhir memiliki kemiripan yang lebih tinggi dibandingkan dengan dokumen teks yang berada di kelompok lain. Dengan mengimplementasikan metode *Frequent Term-Based Text Clustering*, diharapkan pembaca tidak perlu lagi membaca dokumen teks yang hampir sama berulang-ulang.

Contoh *text mining* berikutnya yaitu Sistem klasifikasi dan pencarian dokumen teks dengan algoritma naive bayes classifier dan vector space model (Entot, 2013). Kebutuhan pengklasifikasian dan pencarian informasi secara otomatis semakin diperlukan untuk menghemat waktu dan sumber daya manusia dalam penggolongan dokumen teks serta mempermudah pencarian informasi mengenai suatu kejadian atau bidang tertentu. Penelitian ini bertujuan untuk membangun sistem klasifikasi dan pencarian dokumen teks secara otomatis. Metode yang digunakan untuk membangun sistem tersebut

adalah menggabungkan dua algoritma *Na ve Bayes Classifier* dan *Vector Space Model*. Sistem ini dapat melakukan klasifikasi dokumen teks dan pencarian dokumen teks berbahasa indonesia secara lengkap dan tepat sesuai dengan kata kunci.

Contoh aplikasi *text mining* berikutnya yaitu Mesin pencari dokumen dengan pengklasteran secara otomatis (Etin, 2010). *Web mining* untuk pencariann berdasarkan kata kunci dengan pengklasteran secara otomatis adalah suatu metode pencarian dokumen dengan cara mengelompokkan atau mengklaster dokumen dari dokumen-dokumen berdasarkan kata kuncinya. Selanjutnya dilakukan pengklasteran dengan metode *centroid linkage hierarchical method (CLHM)* terhadap jumlah kata kunci yang diperoleh dari masing-masing dokumen. Dalam pengklasteran, umumnya harus dilakukan inisialisasi jumlah klaster yang ingin dibentuk terlebih dahulu, padahal pada beberapa kasus pengklasteran, user bahkan tidak tahu berapa banyak klaster yang bisa dibangun. Untuk itu, diaplikasikan metode *Valley Tracing* sebagai constraint yang akan melakukan identifikasi terhadap pergerakan varian dari tiap tahap pembentukan klaster dan menganalisa polanya untuk membentuk suatu klaster secara otomatis (*automatic clustering*).

Dari penelitian yang sudah dilakukan sebelumnya, maka penulis mencoba membangun Aplikasi pencarian dokumen dengan menggunakan *text mining* untuk membantu pencarian dokumen teks. Dalam penelitian ini dibangun menggunakan SQL Server 2005 dan Microsoft Visual Studio 2008.

Tabel 2.1. Tabel perbandingan sistem yang telah ada dengan yang akan dibangun.

Indikator	Evelin (2006)	Entot (2013)	Etin (2010)	Penulis
Judul	Implementasi text mining untuk mendeteksi kemiripan dokumen	Sistem klasifikasi dan pencarian dokumen teks dengan algoritma <i>naive bayes classifier</i> dan <i>vector space model</i>	Mesin pencari dokumen dengan pengklasteran secara otomatis	Pembangunan aplikasi pencarian dokumen dengan menggunakan <i>text mining</i>
Web Base	-	-	√	√
Text Mining	√	√	√	√
TF-IDF	-	-	-	√
VSM	-	√	-	√
<i>Cosine Similarity</i>	-	-	-	√