

BAB 2

TINJAUAN PUSTAKA

Perkembangan informasi teks digital telah tumbuh sangat cepat. Saat ini diperkirakan 80% teks digital dalam bentuk tidak terstruktur. Tingginya volume dokumen teks ini dipicu oleh aktivitas dari berbagai sumber berita dan aktivitas akademis dari kegiatan riset, konferensi dan pertemuan ilmiah yang makin meningkat (Hamzah, 2012). Perkembangan teknologi internet berdampak bertambahnya jumlah situs berita berbahasa Indonesia dan menciptakan ledakan informasi. Hal tersebut menuntut semua informasi bisa diakses dengan cepat dan tidak harus membutuhkan banyak waktu dalam sebuah *headline* berita (Mustaqhfiri, et al., 2011).

Berita-berita diberbagai situs terkadang memiliki inti yang sama tetapi narasi yang berbeda, maka dari itu untuk mendapatkan informasi pengguna menghabiskan banyak waktu untuk membaca beberapa berita dari situs yang berbeda. Pada penelitian ini, akan dibangun sebuah aplikasi yang digunakan untuk meringkas berita dari berbagai sumber. Metode *Maximum Marginal Relevance* digunakan untuk melakukan peringkasan berita.

Bab tinjauan pustaka ini akan menjabarkan beberapa penelitian yang terlebih dahulu dilakukan sebelum pembangunan aplikasi web peringkasan berita. Berikut adalah beberapa penelitian lain yang menjadi acuan penulis untuk membangun aplikasi ini.

Evan membuat aplikasi web untuk memenuhi kebutuhan untuk meringkas dokumen karena dirasa sangat penting saat ini sudah memasuki era dimana banyak sekali terdapat informasi yang disajikan. Evan menyajikan aplikasi peringkas dokumen otomatis dari banyak sumber berdasarkan pada identifikasi dan ekstraksi kalimat penting dari dokumen yang menjadi sumber informasi menggunakan *sentence scoring* dengan metode TF-IDF. Aplikasi ini Evan bangun berbasis website agar mempermudah pengguna dalam mengakses aplikasi dimanapun dan kapanpun (Evan, 2014).

Yulita membangun aplikasi ini dengan metode *Maximum Marginal Relevance* untuk peringkasannya dan pembobotan katanya menggunakan TF-IDF-DF untuk melihat keefektifan suatu kata. Aplikasi ini dibangun berbasis web dan dengan menggunakan bahasa pemrograman PHP. Inputan teks berupa artikel berita sebagai *single document* yang merupakan bahan mentah untuk menghasilkan ringkasan. Inputan terlebih dahulu menjalankan *text preprocessing*. Setelah itu dilakukan penghitungan pembobotan kata dengan menggunakan TF-IDF, penghitungan MMR, perankingan kalimat, dan penyusunan kalimat (Yulita, 2015).

Afifah membuat penelitian untuk membuat peringkasan dokumen otomatis menggunakan metode *Maximum Marginal Relevance* (MMR) dan fitur kata untuk dokumen skripsi. Metode ini menggabungkan relevansi antara kalimat dengan query dan kalimat dengan kalimat yang telah terpilih sebagai ringkasan (Afifah, 2015).

Muztahid membuat penelitian untuk pembuatan peringkasan dokumen otomatis menggunakan fitur kata dan metode KMeans. Ringkasan dokumen otomatis dapat digunakan untuk mendapatkan ringkasan teks dengan cepat sehingga memudahkan pengguna untuk mendapatkan informasi utama dari sebuah dokumen. Penelitian ini melakukan peringkasan dokumen dengan menggunakan fitur kata dan metode k-means (Muztahid, 2015).

Luthfiarta membangun penelitian dengan model integrasi peringkasan dokumen otomatis menggunakan algoritma LSA (*Latent Semantic Analysis*) untuk proses *clustering* dokumen. Algoritma ini akan dibandingkan dengan algoritma *Feature Based* yang sudah ada dalam penelitian sebelumnya untuk meningkatkan akurasi hasil *clustering* dokumen. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi dari *clustering* dokumen dengan mengintegrasikan peringkasan dokumen otomatis (Luthfiarta, et al., 2013).

Ringkasan berita diartikan sebagai teks yang dihasilkan dari satu atau lebih kalimat yang menyampaikan informasi penting dari berita. Salah satu fase penting dalam peringkasan adalah pembobotan kalimat (*sentence scoring*). Hayatin membuat penelitian yang bertujuan untuk meringkas multi berita menggunakan metode pembobotan berdasarkan Trending Issue dengan tetap mempertimbangkan fitur penting berita, yaitu *word frequency*, TF-IDF, posisi kalimat, dan kemiripan kalimat terhadap judul (Hayatin, et al., 2015).

Tabel 2.1 Tabel Perbandingan

Penulis	Metode	Algoritma
(Evan, 2014)	Sentence Scoring	Input dokumen → Pemecahan dokumen menjadi kalimat → <i>text preprocessing</i> → pembobotan kata → penghitungan TF IDF → pembobotan kalimat → menyusun kalimat dari <i>sentence scoring</i> → hasil ringkasan
(Yulita, 2015)	<i>Maximum Marginal Relevance</i>	Memasukkan teks dokumen dan <i>query</i> → <i>Text preprocessing</i> → pembobotan TF-IDF-DF → Hitung nilai <i>similarity</i> → Hitung bobot MMR → ringkasan.
(Afifah, 2015)	<i>Maximum Marginal Relevance</i>	Pengumpulan dokumen → Pengindeksan → Proses peringkasan pembobotan TF-ISF → Penghitungan <i>cosine similarity</i> → seleksi menggunakan MMR → hasil ringkasan sistem
(Muztahid, 2015)	K-Means	Menentukan banyaknya <i>cluster</i> → menentukan centroid → menghitung jarak centroid → mengelompokan dari jarak terdekat.
(Luthfiarta, et al., 2013)	<i>Latent Semantic Analysis</i>	Metode peringkas LSA menggunakan SVD, dimana SVD digunakan untuk mengurangi matrik term dokumen ke dalam matrik yang mempunyai dimensi lebih rendah.
(Penulis, 2016)	<i>Maximum Marginal Relevance</i>	Menentukan <i>cluster</i> → melakukan peringkasan tiap <i>cluster</i> , peringkasan tunggal dengan <i>sentence scoring</i> → peringkasan multidokumen dengan MMR