

MAXIMUM MARGINAL RELEVANCE UNTUK PERINGKASAN TEKS OTOMATIS SINOPSIS BUKU BERBAHASA INDONESIA

Aida Indriani¹⁾

¹⁾ Teknik Informatika STMIK PPKIA Tarakanita Rahmawati Tarakan
Jl Yos Sudarso 8, Tarakan 77111
Email : m31d1t4_03@yahoo.com¹⁾

Abstrak

Peringkasan teks otomatis merupakan salah satu bidang Information Retrieval System yang saat ini banyak digunakan untuk mendapatkan sebuah ringkasan (summary) terhadap suatu dokumen. Proses belajar mengajar tidak lepas dari beberapa sumber yang biasanya digunakan sebagai bahan referensi, salah satunya adalah buku. Untuk mencari buku yang sesuai dengan pilihan (query) dapat dilakukan dengan membaca sinopsis dari buku tersebut. Banyaknya kata-kata atau kalimat yang terdapat pada sinopsis buku terkadang tidak mengandung query yang diinginkan. Manfaat dari peringkasan teks dapat digunakan untuk kelancaran proses belajar mengajar, yaitu dengan melakukan peringkasan teks otomatis terhadap sinopsis buku secara cepat dan tepat.

Dalam proses peringkasan teks otomatis sinopsis buku dilakukan dengan beberapa tahapan dari pre-processing yaitu: pemecahan kalimat, tokenizing, penghapusan stop-words dan stemming. Langkah selanjutnya yaitu melakukan perhitungan pembobotan $tf\ idf$, query relevance dan bobot similarity kalimat. Hasil akhir ringkasan diperoleh dengan menggunakan metode maximum marginal relevance.

Untuk uji coba digunakan 20 data uji yang diambil dari koleksi buku perpustakaan STMIK PPKIA Tarakanita Rahmawati yang berbahasa Indonesia. Hasil pengujian dibandingkan dengan ringkasan manual yang menghasilkan rata-rata recall 73,6% dan precision 62,4%.

Kata kunci: Peringkasan, Pre-processing, $Tf\ Idf$, BM25, Cosine Distance Measure, Maximum Marginal Relevance

1. Pendahuluan

1.1 Latar Belakang

Dunia Pendidikan tidak lepas dari proses belajar mengajar. Banyak fasilitas yang mendukung dalam proses belajar mengajar, salah satunya adalah buku. Buku merupakan salah satu sumber yang dapat digunakan sebagai referensi dalam proses belajar mengajar. Untuk mencari buku yang sesuai dengan pilihan, tidak cukup hanya dengan melihat judul dari buku tersebut, dikarenakan judul buku terkadang tidak mengandung isi dari buku tersebut. Bagian dari buku

yang menjelaskan secara singkat mengenai isi buku adalah sinopsis buku yang biasanya terletak dibagian belakang buku.

Sinopsis adalah ikhtisar karangan yang biasanya diterbitkan bersama-sama dengan karangan asli yang menjadi dasar sinopsis itu (Kamus Besar Bahasa Indonesia, 2013) [1]. Sinopsis buku biasanya terdiri atas beberapa paragraph dan beberapa baris kalimat yang letaknya dibagian belakang buku. Untuk mencari buku yang sesuai dengan pilihan, biasanya dengan memasukkan suatu kata kunci (query) agar menemukan buku yang sesuai (relevan) dengan query. Tetapi tidak menutup kemungkinan baris-baris kalimat yang terdapat pada sinopsis buku tidak mengandung query yang dimasukkan. Agar menemukan baris-baris kalimat pada sinopsis buku yang sesuai dengan query, maka dibutuhkan suatu ringkasan.

Ringkasan digunakan untuk memperoleh inti atau isi dari sinopsis buku yang relevan dengan query. Menurut Olivia (2009:30) menyatakan bahwa "ringkasan adalah suatu cara yang efektif untuk menyajikan suatu karangan yang panjang dalam bentuk yang singkat" (Fuji Mulia, 2013) [2]. Ringkasan adalah mengambil isi yang paling penting dari sumber informasi yang kemudian menyajikannya kembali dalam bentuk yang lebih ringkas bagi penggunaanya (Mustaqhfi, Abidin dan Kusumawati, 2011) [3].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan pada poin 1.1, maka bagaimana caranya melakukan peringkasan yang akan menghasilkan suatu ringkasan (summary) terhadap sinopsis buku. Dengan menggunakan metode pembobotan BM25, cosine distance measure dan maximum marginal relevance dapat mengatasi solusi dalam peringkasan sinopsis buku yang digunakan sebagai referensi dalam proses belajar mengajar dengan cepat dan tepat.

1.3 Tujuan

Peringkasan sinopsis buku berbahasa Indonesia ini bertujuan untuk membantu user (Mahasiswa atau Siswa) yang membutuhkan buku yang dapat digunakan sebagai referensi dalam proses belajar mengajar secara cepat dan tepat serta relevan dengan query yang dimasukkan oleh user.

1.4 Tinjauan Pustaka

1.4.1 Peringkasan Teks

Peringkasan teks (*text summarization*) merupakan suatu proses penyulingan sebagian besar informasi penting dari sumber (beberapa sumber) untuk menghasilkan suatu ringkasan bagi pemakai atau pekerjaan tertentu (Mani, House, Klein dan dkk, 1999) [4]. Ketika proses peringkasan teks dilakukan oleh komputer secara otomatis, maka disebut sebagai *Automatic Text Summarization* (peringkasan teks otomatis – PTO). Peringkasan teks otomatis menghasilkan 2 (dua) keluaran yaitu Ekstraksi (*extract*) dan Abstraksi (*abstract*). Adapun tipe peringkasan teks otomatis terbagi menjadi 2 (dua) tipe yaitu Ringkasan yang umum (*generic summary*) dan Ringkasan berpusat pada pemakai (*query-driven*) yang merupakan peringkasan bersandar pada spesifikasi kebutuhan informasi pemakai, seperti topik atau *query*. (Susanto, 2013) [5].

Penelitian peringkasan teks otomatis dipelopori oleh Luhn sejak tahun 1958. Teknik-teknik yang digunakan dalam ringkasan antara lain (1) teknik pendekatan statistika, (2) teknik pendekatan dengan *natural language analysis* yang salah satunya yaitu dengan menggunakan metode *maximum marginal relevance* (Mustaqhfi, Abidin dan Kusumawati, 2011) [3].

1.4.2 Pre-processing

Information Retrieval System atau temu kembali informasi merupakan proses pencarian tidak terstruktur untuk memenuhi kebutuhan informasi dari sekumpulan koleksi yang besar (Manning, 2009) [6]. Secara garis besar ada dua pekerjaan yang ditangani oleh sistem IR, yaitu melakukan *pre-processing* terhadap dokumen dan kemudian menerapkan metode tertentu untuk menghitung kedekatan (relevansi atau *similarity*) antara dokumen dengan *query*.

Tugas pokok pada tahapan *pre-processing* di dalam IR adalah membangun *index* dari koleksi dokumen. *Index* adalah himpunan *term* yang menunjukkan isi atau topik yang dikandung oleh dokumen. Ekstraksi *term* biasanya melibatkan tiga operasi utama antara lain (1) Pemisahan rangkaian *term* (*tokenization*). *Tokenization* adalah tugas memisahkan deretan kata di dalam kalimat, paragraph atau halaman menjadi *token* atau potongan kata tunggal atau *termed word*. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua token ke bentuk huruf kecil (*lower case*), (2) Penghapusan *stop-words*. *Stop word* didefinisikan sebagai *term* yang tidak berhubungan (*irrelevant*) dengan subyek utama dari database meskipun kata tersebut sering kali hadir di dalam dokumen. Contoh *stop words* adalah “ada”, “adalah”, “adanya”, “adapun” “agak”, dll, (3) *Stemming*. Kata-kata yang muncul di dalam dokumen sering mempunyai banyak varian morfologik. Karena itu, setiap kata yang bukan *stop-words* direduksi ke *stemmed word* (*term*) yang cocok yaitu kata tersebut distem untuk

mendapatkan bentuk akarnya dengan menghilangkan awalan atau akhiran (Ilyas, 2010) [7].

14.3 Pembobotan TF IDF

Pembobotan dapat diperoleh berdasarkan jumlah kemunculan suatu *term* dalam sebuah dokumen *term frequency* (tf) dan jumlah kemunculan *term* dalam koleksi dokumen *inverse document frequency* (idf). Nilai idf sebuah *term* (kata) dapat dihitung menggunakan persamaan (1) sebagai berikut:

$$IDF = \text{Log} \left(\frac{D}{d_{ft}} \right) \dots\dots\dots (1)$$

D adalah jumlah dokumen yang berisi *term* (t) dan dft adalah jumlah kemunculan (frekuensi) *term* terhadap D. Adapun algoritma yang digunakan untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci (*query*) dapat dilihat pada persamaan (2), yaitu:

$$W_{d,t} = t f_{d,t} \times IDF_t \dots\dots\dots (2)$$

Keterangan:

- d = dokumen ke-d
- t = term ke-t dari kata kunci
- tf = term frekuensi/frekuensi kata
- W = bobot dokumen ke-d terhadap term ke-t (Mustaqhfi, Abidin dan Kusumawati, 2011) [3].

1.4.4 Pembobotan BM25

Pembobotan BM25 atau Okapi adalah pembobotan yang mengurutkan set dokumen berdasarkan *term* kueri yang muncul pada setiap dokumen koleksi. Hubungan antara *term* kueri dengan dokumen dipengaruhi oleh parameter k1 (parameter untuk kalibrasi skala frekuensi term) dan parameter b (parameter untuk kalibrasi skala panjang dokumen). Nilai parameter yang optimal untuk pembobotan BM25 adalah k1=1.2 dan b=0.75. Penghitungan bobot suatu dokumen berdasarkan term t dinyatakan dalam persamaan (3) sebagai berikut:

$$\text{score}(Q, D_i) = \sum_{t \in q} \log \left[\frac{N}{d_{ft}} \right] \cdot \frac{(k_1 + 1) t f_{ti}}{K + t f_{ti}} \dots\dots\dots (3)$$

dengan $K = k_1 ((1 - b) + (b \times (L_{di} / L_{ave})))$;

Keterangan:

- $\log \left[\frac{N}{d_{ft}} \right]$ adalah inverse document frequency
- tf_{ti} adalah frekuensi term t pada dokumen i
- L_{di} dan L_{ave} adalah panjang dokumen Di dan rata-rata panjang dokumen dalam koleksi
- K1 dan b adalah parameter penskalaan terhadap tf dan panjang dokumen (Paskianti, 2011) [8].

1.4.5 Cosine Distance Measure

Cosine distance measure digunakan untuk menghitung pendekatan relevansi antara kalimat yang satu terhadap kalimat lainnya. Penghitungan cosine distance measure dinyatakan dalam persamaan (4) sebagai berikut:

$$\text{Cos}(W_1, W_2) = \frac{\sum_{i=1}^n p_i q_i}{\sqrt{\sum_{i=1}^n p_i^2 \sum_{i=1}^n q_i^2}} \dots\dots\dots (4)$$

(Sun, 2007) [9].

1.4.6 Maximum Marginal Relevance

Algoritma maximum marginal relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (*extractive summary*) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen. MMR meringkas dokumen dengan menghitung kesamaan (*similarity*) antara bagian teks.

Pada peringkasan dokumen dengan metode MMR dilakukan proses segmentasi dokumen menjadi kalimat dan dilakukan pengelompokkan sesuai dengan gender kalimat tersebut. MMR digunakan dengan mengkombinasikan matriks cosine similarity untuk merangking kalimat-kalimat sebagai tanggapan pada *query* yang diberikan oleh *user*. Penghitungan MMR dinyatakan dalam persamaan (5) sebagai berikut:

$$\text{MMR} = \text{argmax} [\lambda * \text{Sim}_1(S_i, Q) - (1 - \lambda) * \text{max Sim}_2(S_i, S')] \dots\dots\dots (5)$$

S_i adalah kalimat di dokumen, sedangkan S' adalah kalimat yang telah dipilih atau diekstrak. Koefisien digunakan untuk mengatur kombinasi nilai untuk memberi penekanan bahwa kalimat tersebut dan untuk mengurangi redudansi. Nilai parameter adalah mulai dari 0 sampai dengan 1 (range [0,1]). Untuk peringkasan *small dokumen*, seperti pada berita atau sinopsis, menggunakan nilai parameter = 0.7 atau = 0.8, karena akan menghasilkan ringkasan yang baik. Sim_1 adalah matrik similarity kalimat S_i terhadap *query* yang diberikan oleh *user* sedangkan Sim_2 adalah matrik similarity kalimat S_i terhadap kalimat yang telah diekstrak sebelumnya (Mustaqfiri, Abidin dan Kusumawati, 2011) [3].

1.4.7 Recall dan Precision

Efektifitas kinerja dari suatu IR system dilihat dari seberapa banyak IR system tersebut menghasilkan dokumen yang relevan dengan query. Ukuran efektifitas kinerja yang banyak digunakan di bidang IR yaitu *Precision* dan *Recall*. *Precision* adalah perbandingan antara jumlah dokumen relevan yang ditampilkan terhadap jumlah seluruh dokumen yang ditampilkan. *Recall* adalah perbandingan antara jumlah dokumen relevan yang ditampilkan terhadap jumlah seluruh dokumen yang relevan. Penotasian *precision* dan *recall* dalam bentuk *contingency table* ditunjukkan pada Tabel 1.

Tabel 1. *Contingency Table Precision dan Recall*

	Relevant	Nonrelevant
Retrieved	TP	FP
Not Retrieved	FN	TN

Dari Tabel 1 diperoleh sebuah persamaan (6) untuk menghitung *precision* dan persamaan (7) untuk menghitung *recall*.

$$\text{Precision} = \frac{tp}{(tp+fp)} \dots\dots\dots (6)$$

$$\text{Recall} = \frac{tp}{(tp+fn)} \dots\dots\dots (7)$$

Keterangan:

True Positives (tp) = Jumlah dokumen yang ditampilkan oleh sistem dan yang relevan dengan query.
 False Positives (fp) = Jumlah dokumen yang ditampilkan oleh sistem tetapi tidak relevan dengan query.
 False Negatives (fn) = Jumlah dokumen yang tidak ditampilkan oleh sistem tetapi relevan dengan query.
 True Negatives (tn) = Jumlah dokumen yang tidak ditampilkan oleh sistem dan yang tidak relevan dengan query. (Manning, 2009) [6].

2. Pembahasan

Peringkasan teks otomatis sinopsis buku pada penelitian ini dengan menggunakan inputan single dokumen dan menghasilkan sebuah ringkasan (summary) teks sinopsis yang *relevan* dengan inputan *query*. Ada beberapa tahapan yang harus dilakukan dalam proses peringkasan teks otomatis sinopsis buku, antara lain: pre-processing, pembobotan tf idf, pembobotan relevance query, pembobotan similarity kalimat, pembobotan MMR dan ekstraksi ringkasan.

Gambar 1. *Sinopsis Buku dan Query*

Buku Kreasi Desain Produk, Distro dan fashion 3D dibuat berdasarkan perkembangan industri-industri di indonesia saat ini, di mana dalam buku ini diajarkan desain-desain produk sederhana, produk yang banyak di produksi di industri skala kecil dan menengah. Materi-materi yang diajarkan merupakan bidang keahlian AutoCAD dan 3DS Max yang sangat dibutuhkan oleh industri-industri di tanah air saat ini. Banyak lowongan kerja terbuka bagi anda yang ahli di bidang desain produk. Pembahasan diberikan secara lengkap, mulai dari 2D, 3D, hingga operasi rendering. Bonus di dalam CD terdapat file-file pendukung dan file latihan. Bagi anda yang sudah cukup ahli, anda dapat langsung belajar mengembangkan desain yang telah ada di dalam CD untuk dibuat menjadi lebih atraktif.

Query: desain distro 3d

Gambar 1 merupakan sinopsis buku yang berjudul “Kreasi Desain Produk Distro dan Fashion 3D” dengan inputan query berupa “desain distro 3d”. Setelah memasukkan teks sinopsis buku dan query, langkah selanjutnya adalah sebagai berikut:

2.1 Pemecahan Kalimat

Pemecahan kalimat digunakan untuk membagi teks sinopsis buku menjadi beberapa bagian kalimat dengan menghilangkan tanda akhir kalimat (delimiter). Tanda baca akhir kalimat seperti tanda titik “.”, tanda Tanya “?”, dan tanda seru “!” (Mustaqfiri, Abidin dan Kusumawati, 2011) [3]. Kalimat-kalimat yang diperoleh setelah dilakukan pemecahan kalimat dapat dilihat pada Tabel 2.

Tabel 2. Kumpulan Pemecahan Kalimat

No.	Kalimat
1	Buku Kreasi Desain Produk, Distro dan fashion 3D dibuat berdasarkan perkembangan industri-industri di Indonesia saat ini, di mana dalam buku ini diajarkan desain-desain produk sederhana, produk yang banyak di produksi di industri skala kecil dan menengah
2	Materi-materi yang diajarkan merupakan bidang keahlian AutoCAD dan 3DS Max yang sangat dibutuhkan oleh industri-industri di tanah air saat ini
3	Banyak lowongan kerja terbuka bagi anda yang ahli di bidang desain produk
4	Pembahasan diberikan secara lengkap, mulai dari 2D, 3D, hingga operasi rendering
5	Bonus di dalam CD terdapat file-file pendukung dan file latihan
6	Bagi anda yang sudah cukup ahli, anda dapat langsung belajar mengembangkan desain yang telah ada di dalam CD untuk dibuat menjadi lebih atraktif

Pemecahan kalimat yang dilakukan pada sinopsis buku yang terdapat pada Gambar 1 menghasilkan 6 (enam) kalimat.

2.2 Pre-processing

Proses pre-processing seperti yang telah dijelaskan pada poin 1.5.2 terdapat 3 tahapan utama yaitu:

2.2.1 Tokenizing (tokenisasi)

Tahap ini diawali dengan melakukan pemisahan kata per kata pada setiap kalimat disertai dengan menghilangkan karakter tertentu seperti tanda baca “~@#%&* _+=\|[]{}<>.!?-'"/+();0” dan diakhiri dengan mengubah semua token ke bentuk huruf kecil (lower case). Bentuk hasil dari tokenisasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Tokenisasi Kalimat ke-1

No.	Kalimat
1	buku kreasi desain produk distro dan fashion 3d dibuat berdasarkan perkembangan industri industri di Indonesia saat ini di mana dalam buku ini diajarkan desain desain produk sederhana produk yang banyak di produksi di industri skala kecil dan menengah

2.2.2 Penghapusan stop-words

Tahap ini digunakan untuk membuang kata-kata yang tidak mempunyai hubungan dengan sinopsis. Contoh *stop words* adalah ada, adalah, adanya, adapun, agak, dia, saya, bagaimana, dan lainnya. Bentuk hasil penghapusan stop-words ditunjukkan pada Tabel 4.

Tabel 4. Hasil Penghapusan Stop-Words Kalimat ke-1

No.	Kalimat
1	kreasi desain distro fashion 3d perkembangan diajarkan desain desain sederhana produksi skala menengah

2.2.3 Stemming

Pada tahap ini dilakukan pembuangan awalan atau akhiran dari suatu kata sehingga mendapatkan bentuk kata dasarnya. Sebagai contoh, kata terdaftar, pendaftaran berbagi-pakai term stem umum daftar. Bentuk hasil stemming ditunjukkan pada Tabel 5.

Tabel 5. Hasil Stemming Kalimat ke-1

No.	Kalimat
1	kreasi desain distro fashion 3d kembang ajar desain desain sederhana produksi skala menengah

2.3 Pembobotan TF IDF

Setelah dilakukan tahapan pre-processing, langkah selanjutnya adalah melakukan perhitungan pembobotan tf idf untuk setiap term (kata) dan query. Sebelum melakukan perhitungan tf idf, tahapan yang harus dilakukan adalah membuat sebuah matriks antara kalimat dan setiap term. Matriks antara kalimat dan setiap term untuk beberapa kalimat dan term dapat ditunjukkan pada Tabel 6.

Tabel 6. Matriks Term dan Kalimat

No.	Term	KK	K1	K2	K3	K4
1	Kreasi		1	0	0	0
2	Desain	1	3	0	1	0
3	Distro	1	1	0	0	0
4	Fashion		1	0	0	0
5	3d	1	1	0	0	1
6	Kembang		1	0	0	0

Setelah didapat sebuah matriks antara term dan kalimat, kemudian dilakukan perhitungan tf idf dengan menggunakan persamaan (1) dan (2). Hasil tf idf untuk setiap kalimat terhadap masing-masing query ditunjukkan pada Tabel 7.

Tabel 7. Nilai TF IDF

K1	K2	K3	K4	K5	K6
2,16	0	0,30	0,48	0	0,30

Pada Tabel 7 terdapat nilai yang sama yaitu nilai tf idf untuk kalimat 3 dan kalimat 6. Jika ada nilai tf idf yang sama, maka belum bias dilakukan pengurutan (sorting)

kalimat. Pengurutan kalimat dapat dilakukan dengan menggunakan pembobotan BM25.

2.4 Pembobotan Query Relevance

Pada tahap ini dilakukan perhitungan bobot query relevance yang digunakan untuk proses perhitungan MMR (maximum marginal relevance). Pembobotan query relevance dilakukan dengan cara membandingkan kemiripan (similarity) antara query dengan setiap kalimat. Perhitungan pembobotan query relevance dapat dilakukan dengan menggunakan persamaan (3). Hasil dari perhitungan pembobotan query relevance ditunjukkan pada Tabel 8.

Tabel 8. Bobot Query Relevance (QR)

K1	K2	K3	K4	K5	K6
31,31	0	2,90	5,57	0	3,71

Dari perhitungan pembobotan query relevance diperoleh urutan kalimat dari nilai yang terbesar sampai nilai yang terkecil yaitu : Kalimat 1, Kalimat 4, Kalimat 6, Kalimat 3, Kalimat 2 dan Kalimat 5.

2.5 Pembobotan Similarity Kalimat

Langkah berikutnya adalah melakukan perhitungan pembobotan similarity kalimat yang akan digunakan untuk proses perhitungan MMR (maximum marginal relevance). Pembobotan similarity kalimat dilakukan dengan cara membandingkan kemiripan (similarity) antara kalimat yang satu dengan kalimat lainnya sehingga membentuk sebuah matriks bobot similarity kalimat. Perhitungan pembobotan similarity kalimat dapat dilakukan dengan menggunakan persamaan (4). Hasil dari perhitungan pembobotan similarity kalimat ditunjukkan pada Tabel 9.

Tabel 9 Matriks Bobot Similarity Kalimat

No.	K1	K2	K3	K4	K5	K6
1	1	0,06	0,11	0,08	0	0,16
2	0,06	1	0,14	0	0	0,09
3	0,11	0,14	1	0	0	0,35
4	0,08	0	0	1	0	0
5	0	0	0	0	1	0,1
6	0,16	0,09	0,35	0	0,1	1

2.6 Pembobotan MMR

Setelah mendapatkan hasil bobot query relevance dan matriks bobot similarity kalimat, langkah berikutnya adalah melakukan perhitungan pembobotan MMR dengan menggunakan persamaan (5). Perhitungan MMR dilakukan dengan perhitungan iterasi antara bobot query relevance dan bobot similarity kalimat. Pada penelitian ini nilai $\alpha = 0.8$. Kalimat yang memiliki nilai MMR paling tinggi dalam setiap iterasi akan terpilih sebagai ringkasan yang pertama dan seterusnya. Iterasi akan berhenti apabila perhitungan MMR menghasilkan nilai 0 (nol) atau minus (-). Hasil MMR untuk setiap iterasi ditunjukkan pada Tabel 10.

Tabel 10 Hasil Iterasi MMR

ke	K1	K2	K3	K4	K5	K6
1	25,02	-0,03	2,25	4,44	-0,20	2,90
2	-	-0,05	1,73	3,54	-0,04	2,25
3	-	-0,07	1,31	-	-0,05	1,73
4	-	-0,08	0,98	-	-0,06	-
5	-	-0,09	-	-	-0,07	-

2.7 Ekstraksi Ringkasan

Tahap terakhir adalah ekstraksi ringkasan dari hasil pembobotan MMR kalimat. Ekstraksi ringkasan diperoleh dengan mengambil kalimat yang memiliki nilai MMR paling tinggi untuk setiap iterasi. Hasil ekstraksi kalimat yang dihasilkan mulai dari nilai MMR tertinggi sampai terendah ditunjukkan pada Tabel 11.

Tabel 11. Hasil Ekstraksi

No.	Kalimat
K1	Buku Kreasi Desain Produk, Distro dan fashion 3D dibuat berdasarkan perkembangan industri-industri di indonesia saat ini, di mana dalam buku ini diajarkan desain-desain produk sederhana, produk yang banyak di produksi di industri skala kecil dan menengah
K4	Pembahasan diberikan secara lengkap, mulai dari 2D, 3D, hingga operasi rendering
K6	Bagi anda yang sudah cukup ahli, anda dapat langsung belajar mengembangkan desain yang telah ada di dalam CD untuk dibuat menjadi lebih atraktif
K3	Banyak lowongan kerja terbuka bagi anda yang ahli di bidang desain produk

2.8 Hasil Uji Coba

Data sinopsis buku yang digunakan sebagai uji coba sejumlah 20 sinopsis buku. Data uji coba diperoleh dari Perpustakaan STMIK PPKIA Tarakanita Rahmawati. Uji coba dilakukan dengan memasukkan teks yang terdapat pada sinopsis buku dan query yang dimasukkan oleh pengguna (user). Hasil ringkasan kalimat untuk beberapa data uji coba ditunjukkan pada Tabel 12.

Tabel 12. Hasil Ekstraksi dari Data Uji Coba

Nomor Buku	Jumlah Seluruh Kalimat	Jumlah Ringkasan Kalimat	Ringkasan (No. Kalimat)
1	5	4	1,3,5,4
2	7	3	5,6,4
3	9	4	6,8,5,4
4	8	5	4,7,8,3,5
5	8	4	5,3,8,4
6	8	6	8,4,7,1,2,5
7	6	5	4,5,1,2,3
8	4	3	1,3,4
9	7	2	1,3
10	4	2	1,3

2.9 Pengukuran Kualitas Ringkasan

Kualitas ringkasan teks otomatis sinopsis buku diukur dengan membandingkan terhadap ringkasan manual. Perhitungan kualitas ringkasan diperoleh dari nilai precision dan recall dengan menggunakan persamaan (6) dan persamaan (7). Dari ringkasan yang dilakukan secara manual diperoleh sebanyak 2 (dua) kalimat yaitu kalimat 1 dan 4. Dari hasil ringkasan ekstraksi dengan MMR dan manual dibuatkan sebuah contingency table seperti yang ditunjukkan pada Tabel 13.

Tabel 13. Contingency Table Precision dan Recall antara Ringkasan Manual dan MMR

	Relevant	Nonrelevant
Retrieved	2	2
Not Retrieved	0	2

Dari Contingency Table pada Tabel 14 dilakukan perhitungan nilai Precision dan Recall sebagai berikut:

Diketahui : tp (True Positives) = 2
 fp (False Positives) = 2
 fn (False Negatives) = 0

Precision = $tp / (tp + fp) = 2 / (2 + 2) = 2/4 = 0,50 = 50\%$

Recall = $tp / (tp + fn) = 2 / (2 + 0) = 2/2 = 1 = 100\%$

Nilai yang diperoleh dari hasil perhitungan Precision dan Recall adalah 50% untuk nilai Precision dan 100% untuk nilai Recall. Hasil dari pengukuran kualitas ringkasan untuk beberapa data uji coba ditunjukkan pada Tabel 14.

Tabel 14. Hasil Pengukuran Kualitas Ringkasan

Nomor Buku	Recall	Precision
1	100%	50%
2	67%	67%
3	100%	75%
4	100%	40%
5	80%	80%
6	80%	67%
7	100%	40%
8	100%	33%
9	50%	100%
10	100%	100%

Dari nilai precision dan recall pada Tabel 14 untuk 20 data uji coba diperoleh rata-rata pengukuran kualitas ringkasan untuk nilai precision yaitu 62,4% dan nilai recall 73,6%.

3. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian yaitu penggunaan metode MMR dapat digunakan untuk meringkas teks otomatis terhadap sinopsis buku dengan menggunakan beberapa metode pendukung yaitu pembobotan BM25 untuk menghasilkan bobot query relevance dan cosine distance measure untuk

menghasilkan bobot similarity antar kalimat, hasil dari 20 data uji coba diperoleh nilai pengukuran kualitas ringkasan dengan nilai precision sebesar 62,4% dan nilai recall sebesar 73,6%.

Hasil ringkasan yang diperoleh pada penelitian ini dengan menggunakan beberapa metode yaitu metode pembobotan BM25, cosine distance measure dan MMR. Pengembangan lebih lanjut dapat menggunakan metode-lainnya agar menghasilkan nilai precision dan recall yang lebih besar. Untuk pengukuran kualitas ringkasan, dapat juga dengan menggunakan perhitungan *f-measure*.

Daftar Pustaka

- [1] "Kamus Besar Bahasa Indonesia", <http://kbbi.web.id/sinopsis>, (diakses tanggal 11 Nopember 2013).
- [2] Fuji Mulia, <http://www.trigonalworld.com/2013/05/pengertian-dan-tujuan-ringkasan.html>, (diakses tanggal 11 Nopember 2013).
- [3] M. Mustaqfiri, Z. Abidin, R. Kusumawati, "Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance", in *n nitro PDF professional*, pp. 134-147.
- [4] I. Mani, D. House, G. Klein dkk, "The TIPSTER SUMMAC Text Summarization Evaluation", in *Proc. Of EACL 1999*, pp.77, 1999.
- [5] Budi Susanto, "Text Summarization", in *Sebuah Pengantar TI UKDW*, Februari 4, 2013.
- [6] C.D. Manning, P. Raghavan dan H. Schütze, "An Introduction to Information Retrieval", in *Cambridge University Press*, pp.26, April 1, 2009.
- [7] Husni Ilyas, "Unified Messaging System Information Retrieval & Klasifikasi Teks", in *Komputasi / Suatu Permulaan Data Mining & IR*, pp.6-8, Januari 22, 2010.
- [8] Kristina Paskianti, "Klasifikasi Dokumen Tumbuhan Obat menggunakan Algoritma KNN Fuzzy", in *Thesis Fakultas Matematika dan Ilmu Pengetahuan Alam IPB Bogor*, pp.4, 2011.
- [9] Yixing Sun, "Using the Organizational and Narrative Thread Structures in an e-Book to Support Comprehension", in *A Dissertation submitted to The Robert Gordon University*, pp.67, Agustus, 2007.

Biodata Penulis

Nama Lengkap Penulis Pertama, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknologi Informasi Sekolah Tinggi Teknik Surabaya, lulus tahun 2013. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati Tarakan.