PEMBANGKIT PERTANYAAN BERDASARKAN *TEMPLATE* DARI DOKUMEN RINGKASAN MENGGUNAKAN METODE *LEXICAL CHAIN* DAN KORELASI ANTAR KALIMAT

Sani Saefurochman

Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : Saefurochman.sani@gmail.com

ABSTRAK

Perkembangan teknologi yang semakin pesat berpengaruh terhadap semua bidang termasuk bidang pendidikan. Pembuatan pertanyaan dari sebuah teks untuk mengevaluasi pemahaman pembaca mengenai apa yang disampaikan didalam teks merupakan hal yang sudah tidak asing lagi bagi terutama bagi seorang pengajar untuk melakukan evaluasi terhadap siswanya. Akan tetapi manusia memiliki keterbatasan dalam membuat pertanyaan dari teks yang berisi kalimat yang banyak. Hal ini lah yang membuat para peneliti melakukan penelitian tentang bagaimana pemanfaatan mesin dalam pelaksanaan *intelligent tutoring*.

Question Generation (QG) merupakan proses yang dilakukan untuk membangkitkan pertanyaan oleh mesin. Salah satunya adalah dengan menggunakan metode *template based*. Akan tetapi metode ini dinilai memiliki keterbatasan terutama dalam kesesuaian pertanyaan yang dihasilkan dengan konteks materi. Oleh Karena itu dilakukan penelitian dengan menambahkan peringkasan otomatis materi sehingga dapat membatasi pertanyaan agar sesuai dengan konteks. Metode peringkasan otomatis yang digunakan adalah metode *lexical chain* dan korelasi antar kalimat.

Dari hasil penelitian, didapat akurasi pertanyaan yang dihasilkan yaitu sebesar 70,26% dengan akurasi dari peringkasan sebesar 83%.

Kata kunci : *Question Generation, intelligent tutoring, template based, lexical chain*.

1. PENDAHULUAN

Intelligent Tutoring merupakan salah satu bahasan dibidang kecerdasan buatan yang dikaji oleh para ilmuwan dibidang computer science dengan harapan bahwa manusia bisa belajar dengan lebih cepat mendapatkan pemahaman menggunakan bantuan mesin [1]. Pembangkit pertanyaan otomatis termasuk kedalam studi yang dibahas dalam intelligent tutoring dengan tujuan untuk mengevaluasi apakah pengguna mampu memahami isi dari sebuah teks yang telah dipelajari.

Question Generation (QG) merupakan proses yang dilakukan untuk membangkitkan pertanyaan yang dapat diterima oleh manusia dari sebuah input dengan bentuk terstruktur (*database*) ataupun tidak terstruktur (*text*) [2]. *Question Generation* merupakan spesialisasi dari *Natural Language Processing* (Pemrosesan Bahasa Alami) yang didalamnya mempelajari bagaimana mesin bisa mengerti dan berinteraksi dengan menggunakan bahasa manusia.

Berdasarkan hasil penelitian sebelumnya dari M. Fachrurrozy dan Novi Yusliani, studi mengenai pembuat pertanyaan otomatis sudah lama menjadi topik dalam Pemrosesan Bahasa Alami. Dari pertanyaan yang dibangkitkan secara otomatis oleh mesin Question Generator yang dibuat, 55,95% dari pertanyaan tersebut dapat diterima [3]. Pertanyaan yang dihasilkan oleh QG memuat pertanyaan yang *nonfactoid* dengan kendala bahwa kesesuaian dengan konteks materi yang dimasukkan sebagai input masih memiliki kekurangan sehingga menurunkan akurasi pertanyaan yang benar. Untuk itu *dibangun Question generator* yang mampu membuat ringkasan dari materi agar sesuai dengan konteks kemudian dibuat pertanyaan dari hasil peringkasannya.

Dalam penelitian yang akan dilakukan untuk menambah presentase pertanyaan yang dapat diterima agar tidak keluar dari konteks yang dibahas dari input dengan membuat ringkasan dari dokumen input terlebih dahulu. Sehingga diharapkan pertanyaan yang dihasilkan oleh *Semantic Question Generator* (SQG) lebih akurat dalam menghasilkan Pertanyaan yang berhubungan denga inti dari dokumen inut. Dengan tujuan mengevaluasi apakah pengguna memahami inti dari dokumen yang dipelajari dan dibuat sebagai input didalam penggunaan SQG.

Automatic Text Summarization adalah sebuah proses pembuatan ringkasan dari sebuah dokumen secara otomatis dengan mengambil inti dari sebuah dokmen. Untuk mendapatkan ringkasan yang baik dari sebuah text, digunakan metode *lexical chain* dan

correlation of sentences yang dinilai mampu membuat ringkasan lebih baik dengan tingkat keakuratan (precision) yang tinggi [4].

Berdasarkan uraian latar belakang yang telah dijelaskan sebelumnya, maka rumusan masalah yang ditemukan adalah sebagai berikut :

1. Bagaimana membuat pertanyaan yang dapat diterima dalam Bahasa Indonesia secara otomatis dari sebuah dokumen text?
2. Bagaimana *Automatic Text Summarization* dapat diimplementasikan dalam pembangkitan pertanyaan secara otomatis?

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Dapat membangkitkan pertanyaan yang dapat diterima dari sebuah teks berbahasa Indonesia.
2. Mengetahui tingkat akurasi pertanyaan dengan mengimplementasikan peringkasan teks otomatis terhadap pembangkit pertanyaan otomatis.

2. ISI PENELITIAN

2.1. *Question Generator*

Question Generation (QG) merupakan proses yang dilakukan untuk membangkitkan pertanyaan yang dapat diterima oleh manusia dari sebuah input dengan bentuk terstruktur (*database*) ataupun tidak terstruktur (*text*). Tugas Pembangkit pertanyaan berisi beberapa subarea [2]. Biasanya, pendekatan yang diambil untuk QG tergantung pada tujuan aplikasi QG.

Pembangkitan pertanyaan dapat dilakukan dengan memanfaatkan sekumpulan dokumen yang ada di internet (open domain) atau sekumpulan dokumen yang ada di dalam sistem itu sendiri (closed domain). Secara umum, pertanyaan dapat diklasifikasikan menjadi lima kategori, yaitu pertanyaan yang bersifat factoid (factoid question), pertanyaan yang bersifat nonfactoid (non-factoid question), list question, yes/no question, dan opinion question. Hal ini menandakan bahwa pertanyaan yang dibangkitkan oleh sistem pembangkit pertanyaan otomatis dapat berupa salah satu, beberapa, atau semua kategori pertanyaan [5]. Pertanyaan yang bersifat non factoid adalah pertanyaan dengan jawaban fakta yang cukup panjang yang biasanya terdiri dari satu klausa atau lebih.

Dalam pembuatan pertanyaan berdasarkan template yang telah ditentukan sebelumnya, terdapat tiga klasifikasi kata khusus yang nantinya akan mengacu pada template mana yang akan dipakai untuk membangkitkan pertanyaan. Berikut ini adalah kata khusus yang digunakan untuk membangkitkan pertanyaan.

2.2. *Automatic Text Summarization*

Automatic Text Summarization adalah proses meringkas dokumen teks menggunakan bantuan program komputer dengan tetap mempertahankan inti dokumen asli [6]. Pada umumnya, *automatic summarization* memiliki tiga teknik peringkasan, yaitu *extraction-based summarization*, *abstraction-based summarization*, dan *maximum entropy-based summarization*. Teknik yang sering digunakan dalam kasus meringkas teks ini yaitu *extraction-based summarization*. Teknik ini memiliki fungsi untuk memberikan hasil ringkasan yang lebih pendek dari dokumen aslinya. Teknik selanjutnya adalah teknik abstraksi yang merupakan cara mengambil intisari dari teks sumber kemudian membuat ringkasan dengan menciptakan kalimat-kalimat baru yang merepresentasikan intisari teks sumber dalam bentuk berbeda.

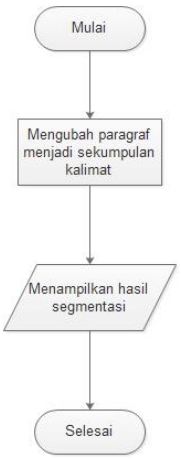
2.2.1. *Preprocessing*

Text mining dapat didefinisikan sebagai keilmuan di mana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu menggunakan *tools* analisis. Dengan menganalogikan kedalam data mining, teks mining digunakan untuk mengekstrak informasi yang bermanfaat dari sumber data melalui identifikasi dan mengeksplorasi pola yang menarik. Dalam kasus teks mining, sumber data adalah koleksi dokumen, dan pola yang menarik yang ditemukan tidak di antara catatan database formal tetapi dalam Data tekstual terstruktur di dokumen dalam koleksi ini.

a. Segmentasi

Tahap segmentasi kalimat yaitu proses memisahkan string teks dokumen menjadi kumpulan kalimat - kalimat [7]. Proses pemisahan dokumen menjadi kalimat-kalimat ini menggunakan fungsi split(), dengan tanda titik ".", "?","!" sebagai delimiter untuk memotong string dokumen.

Berikut ini adalah flowchart dari tahap segmentasi kalimat :



Gambar 1. Flow chart segmentasi kalimat

b. Tokenisasi

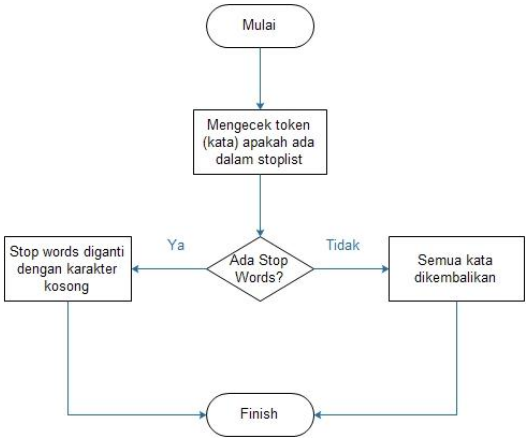
Tahap *tokenization* merupakan tahap pemotongan string input menjadi kumpulan kata yang menyusunnya [7].



Gambar 2. Proses tokenisasi

c. Penghapusan *Stop Words*

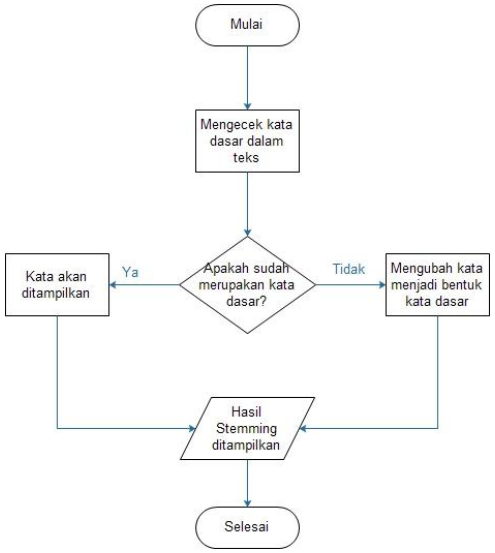
Tahap removing stop words merupakan tahap penghilangan kata-kata yang tidak terlalu penting [7]. Setiap kata hasil dari tokenization dilakukan pengecekan dengan stop list yang berisi kata-kata yang tidak terlalu penting. Jika ditemukan kata pada hasil tokenization sama dengan kata yang ada di dalam stop list, maka kata tersebut akan hilang.



Gambar 3. Proses penghapusan *stop words*

d. *Stemming*

Tahap *stemming* merupakan tahap mengubah kata dari hasil *removing stop words* menjadi kumpulan kata dasar [7]. Tahap ini tidak jauh berbeda dengan hasil *stop words* karena jika hasil kata dasar sudah ada dan tidak ada perubahan maka hasilnya akan sama. Tahapan ini akan menentukan kata dasar dari setiap kata yang terdapat didalam kalimat sehingga pada saat proses perhitungan nantinya yang akan dihitung adalah bobot bobot dari kata yang sama. Tahap stemming memiliki beberapa macam dan yang digunakan disini adalah proses stemming dari penelitian Nazief-Adriani. Berikut ini adalah proses dari stemming dapat dilihat pada gambar 4.



Gambar 4. Proses stemming

2.2.2. Fitur Ekstraksi Kalimat

Extraction feature merupakan proses pengolahan kalimat menggunakan beberapa fitur sehingga menghasilkan nilai-nilai yang nantinya akan digunakan untuk meringkas dokumen.adapun fitur yang digunakan antara lain :

a. Panjang Kalimat

Panjang kalimat merupakan proses untuk mengetahui rasio dalam jumlah kata yang terjadi dalam kalimat atas jumlah kata yang terjadi dalam kalimat dokumen terpanjang. Rumus panjang kalimat dapat dilihat sebagai berikut:

$$S_F1 = \frac{Jumlah\ Kata\ dalam\ Kalimat}{Kalimat\ Terpanjang} \quad (1)$$

b. Bobot Kalimat

Bobot kata merupakan proses menghitung frekuensi kejadian di dalam dokumen yang telah sering digunakan untuk menghitung kata yang terdapat di dalam kalimat. Bobot yang dilakukan adalah dengan menggunakan perhitungan tf-idf (*term frequency, inverse document frequency*). Bobot kalimat dapat dihitung dengan rumus sebagai berikut:

$$S_F2 = \mathbf{wi} = \mathbf{TF} \times (\log \frac{\mathbf{N}}{ni}) \quad (2)$$

c. Kemiripan Kalimat

Kemiripan antar kalimat merupakan kata yang muncul di dalam sebuah kalimat sama dengan kata yang muncul di dalam kalimat lain.untuk menghitung kemiripan, dibutuhkan hasil perhitnugan jarak kalimat dengan rumus sebagai berikut :

$$Ki, Kj = \frac{\Sigma Wit \times Wij}{\sqrt{\sum Wit^2} \times \sqrt{\sum Wij^2}} \quad (3)$$

$$S_F4(K)=\frac{\Sigma Sim(Ki,Kj)}{Max(\Sigma Sim(Ki,Kj))} \quad (4)$$

d. *Lexical Chain*

Lexical chains adalah sebuah teknik penting dalam pengolahan bahasa alami. Lexical chains berkaitan dengan kalimat yang berelasi dengan kata benda, kata kerja, kata sifat [6]. Jika kalimat tidak berkaitan satu sama lain di dalam teks, maka relasi antar leksikal berelasi satu sama lain. Hal ini merupakan konsep yang penting. Skor dari sebuah rantai leksikal secara langsung berkaitan dengan panjang rantai yang hubungannya dengan kata benda. Peringkas menggunakan *lexical chains* untuk memilih kalimat yang paling representatif.

$$CStrength=Clength\times\{\sum\frac{\Sigma((Tf-Isfwi\times Swi))+(Tf-Isfwj\times Swj)}{\text{distance between wi and wj}}\}+TotalNumberOfPairs$$

Gambar 5. Rumus *Lexical Chain*

2.2.3. Logika *Fuzzy*

Logika fuzzy pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dasar logika fuzzy adalah teori himpunan fuzzy. Pada teori himpunan fuzzy, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Logika fuzzy dapat dianggap sebagai kotak hitam yang menghubungkan antara ruang input menuju ruang output [8].

Ada beberapa hal yang perlu diketahui dalam sistem fuzzy, diantaranya:

a. Variabel *Fuzzy*

Variabel Fuzzy adalah variabel yang hendak dibahas dalam suatu sistem variabel *fuzzy*. Contoh : umur, temperatur, permintaan, dsb.

b. Himpunan *fuzzy*

Himpunan Variabel *fuzzy* adalah suatu grup yang mewakili suatu kondisi atau keadaan tertentu dalam suatu variabel *fuzzy*. Himpunan *fuzzy* memiliki dua atribut, yaitu linguistic dan numeris.

c. Semesta Pembicaraan

Semesta pembicaraan adalah himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai semesta pembicaraan dapat berupa bilangan positif atau negatif.

d. Domain

Domain adalah himpunan bilangan real yang senantiasa naik (bertambah) secara monoton dari kiri ke kanan. Nilai domain dapat berupa bilangan positif atau negatif.

Logika fuzzy memiliki fungsi keanggotaan. Fungsi keanggotaan adalah suatu kurva yang menunjukkan pemetaan titik-titik input data ke dalam nilai keanggotaannya (derajat keanggotaan) yang memiliki interval antara 0 sampai 1 [8]. Salah satu cara yang dapat digunakan untuk mendapatkan nilai keanggotaan adalah dengan melalui pendekatan fungsi.

Beberapa fungsi yang dapat digunakan adalah pemetaan input ke derajat keanggotaannya digambarkan sebagai suatu garis lurus. Bentuk ini paling sederhana dan menjadi pilihan yang baik untuk mendekati suatu konsep yang kurang jelas. Ada 2 keadaan himpunan fuzzy yang linear. Pertama, representasi linear naik merupakan kenaikan himpunan dimulai pada nilai domain yang memiliki derajat keanggotaan nol (0) bergerak ke kanan menuju ke nilai domain yang memiliki derajat keanggotaan lebih tinggi. Berikut ini merupakan fungsi keanggotaannya:

$$\mu(x)=\begin{cases}0; & x\le a\\ \frac{(x-a)}{(b-a)}; & a\le x\le b\\ 1; & x\ge b\end{cases} \quad (5)$$

Kedua, representasi linear turun merupakan kebalikan yang pertama. Garis lurus dimulai dari nilai domain dengan derajat keanggotaan tertinggi pada sisi kiri, kemudian bergerak menurun ke nilai domain yang memiliki derajat keanggotaan lebih rendah. Berikut ini merupakan fungsi keanggotaannya:

$$\mu(x)=\begin{cases}\frac{(b-x)}{(b-a)}; & a\le x\le b\\ 0; & x\ge b\end{cases} \quad (6)$$

2.2.4. Korelasi Antar Kalimat

Korelasi antar kalimat merupakan hubungan antara satu kalimat dengan yang lainnya dimana kalimat kalimat saling berkaitan. Apabila sebuah kalimat memiliki keterkaitan dengan kalimat yang lainnya maka kalimat tersebut bisa menjadi induk klaimat ataupun anak kalimat bagi kalimat sebelum dan sesudahnya. Anak kalimat berguna untuk menjelaskan kalimat induk sehingga mempersempit arti dari kalimat induk tersebut. Keterkaitan antar kalimat ditandai dengan kata penghubung sebagai berikut;

Agaknya, ... Akan tetapi, ... Akhinya, ... Akiatnya, Artiya, ... Biarpun begitu, ... Biarpun demikian, ... Berkaitan dengan itu, ... Dalam hal ini, ... Dalam hubungan ini, ... Dalam konteks ini, ... Dengan demikian, ... Dengan kata lain, ... Disamping itu, ... Disatu pihak, ... Dipihak lain, ... Jadi,... Jika demikian, ... Kalau begitu, ... Kalau tidak salah, ... Kecuali itu, ... Lagi pula, ... Meskipun begitu, ... Mekipun demikian, ... Namun, ... Oleh karena itu, ... Oleh sebab itu, ... Pada dasarnya, ... Pada hakikatnya, ... Pada prinsipnya, ... Sebagai kesimpulan, ...

Sebaiknya, ... Sebaliknya, ... Sebelumnya, ... Sebenarnya, ... Sebetulnya, ...
Contoh peringkasan materi input adalah sebagai berikut.

> Teratai adalah tumbuhan yang hidup di air atau sebagai tanaman air. Oleh Karena itu Apabila kamu perhatikan, bunga teratai nampak indah mengapung di atas permukaan air. Meskipun bunganya mengapung, sebenarnya akar bunga teratai berada di dasar air. Jadi bunga teratai tidak semata mengambang di atas air, melainkan tumbuh dari dasar air. Tangkai daun cenderung tumbuh menjalar, sehingga helai daun selalu terlihat mengapung. Teratai memiliki daun berbentuk bundar dan berukuran sangat lebar.

Gambar 6. Contoh teks masukan

Dari teks masukan yang terdapat pada gambar 2, hasil peringkasan berdasarkan metode yang digunakan yang didapat adalah :

Tabel 1. Hasil peringkasan yang didapat

Ki	Kalimat
K01	Teratai adalah tumbuhan yang hidup di air atau sebagai tanaman air
K02	Oleh karena itu apabila kamu perhatikan, bunga teratai nampak indah mengapung di atas permukaan air
K05	Tangkai daun cenderung tumbuh menjalar, sehingga helai daun selalu terlihat mengapung

Dari tabel 1 dapat dilihat bahwa hasil ringkasan dari masukan yang diproses terdapat 3 kalimat yakni kalimat 1, kalimat 2, dan kalimat 5. Peringkasan dilakukan dengan menggunakan aturan ringkasan 30%. Dimana dari 6 kalimat diambil 2 kalimat inti, kemudian setelah itu diambil anak kalimat atau kalimat induk dari dua kalimat pokok tersebut berdasarkan metode korelasi antar kalimat. Sehingga kalimat 2 menjadi kalimat inti Karena merupakan anak kalimat dari kalimat pertama. Setelah dilakukan pengujian, dari 30 materi yang diujikan terhadap 10 orang guru dihasilkan data yang dapat dilihat pada tabel 2.

Tabel 2. Hasil pengujian ringkasan

Materi ke -	Judul Materi	Jumlah kalimat terpilih	Jumlah kalimat benar
1	Bagian Hewan yang Sering Digunakan	12	9
2	Perkembangbiakan Vegetatif Buatan	7	6
3	Manfaat Sumber Daya Alam bagi Manusia	10	8
4	Bunga Pada Tumbuhan	7	6
5	Penebangan Hutan Secara Liar	10	10
6	Hewan yang Terancam Punah di Indonesia	16	13
7	Kemampuan Bertahan Hidup Bunglon	7	7
8	Teratai Sebagai Tanaman Air	8	7
9	Kemampuan Ekolokasi Kelelawar	7	7
10	Kerusakan Lingkungan Hidup	7	6
11	Perkembanganbiakan Vegetatif Alami	13	11
12	Pelapukan pada Batuan	9	9
13	Perubahan Energi Listrik Menjadi Energi Lain	9	9
14	Gaya dan Gerak	7	6
15	Ekosistem	9	7
16	Ciri Khusus Tanaman Kaktus	9	9
17	Pemanfaatan Konduktor dan Isolator	10	7
18	Pelestarian Lingkungan Hidup	9	7
19	Gaya pada Ketapel	6	6
20	Planet	6	6
21	Pengembangbiakan Hewan dan Tumbuhan yang Hampir Punah	8	6
22	Energi Listrik	5	3
23	Penghematan Energi	9	7
24	Sistem Tata Surya	9	7
25	Ciri Khusus Ikan Lele	8	6
26	Jenis Tumbuhan Dikotil	7	5
27	Bagian Tumbuhan yang sering dimanfaatkan	11	8
28	Pembuangan Limbah	6	5
29	Planet Jupiter	7	5
30	Planet Bumi	7	5

Dari total kalimat yang diujikan, didapat 255 kalimat yang dihasilkan oleh sistem dan 213 kalimat diantaranya dinyatakan benar merupakan kalimat inti dari materi yang diujikan. Sehingga dapat dikatakan peringkasan yang dilakukan memiliki tingkat akurasi sebesar 83%.

2.3. Kata khusus

Kata khusus menggambarkan kalimat jawaban untuk suatu kategori pertanyaan. Kalimat yang mengandung kata "karena" dapat digunakan untuk membangkitkan pertanyaan kategori alasan. Kalimat yang mengandung kata "adalah" dapat digunakan untuk membangkitkan pertanyaan kategori definisi. Sedangkan kalimat yang mengandung kata "dengan cara" dapat digunakan untuk membangkitkan pertanyaan kategori metoda. Sistem akan melakukan klasifikasi terlebih dahulu untuk mencari kata kata khusus yang terdapat didalam input untuk menentukan *template* mana yang akan dipakai nantinya. Tabel 3 menunjukkan kata khusus setiap kategori.

Tabel 3. Kategori kata khusus

Kategori	**Kata Khusus Sebelum Target Kata**	**Kata Khusus Setelah Target Kata**
Definisi	disebut, dikenal, dinamakan, mendefinisikan	adalah, yaitu, ialah, merupakan, diartikan
Alasan	oleh sebab itu, jadi, memungkinkan adanya, , dengan demikian, maka, dikatakan, penyebab terjadinya, sehingga, mengapa, walau demikian, namun demikian	sebab, karena, bertujuan
Metode	dengan cara	berfungsi untuk, berguna untuk

2.4. Template Pertanyaan

Template pola kalimat pertanyaan diperoleh melalui proses analisis terhadap sekumpulan pertanyaan. Pada penelitian ini, pertanyaan-pertanyaan yang dibangkitkan adalah pertanyaan-pertanyaan yang termasuk ke dalam kategori pertanyaan *non-factoid* Bahasa Indonesia, yaitu definisi, alasan, dan metode. Karena itu, analisis dilakukan pada sekumpulan pertanyaan *non-factoid* dalam Bahasa Indonesia. Tabel 2 menunjukkan template pola kalimat pertanyaan *non-factoid* dalam Bahasa Indonesia.

Tabel 4. Tabel Template Pertanyaan

Kategori	**Template Pertanyaan**
Definisi	Apakah yang dimaksud dengan + target kata Apa yang anda ketahui mengenai + target kata
Alasan	Mengapa + target kata
Metode	Bagaimana cara + target kata

Kategori pertanyaan yang dibangkitkan tergantung pada hasil pengklasifikasian kalimat. Informasi-informasi yang diperlukan untuk pembangkitan pertanyaan adalah:

1. Ringkasan dari masukkan yang berupa kalimat kalimat penting yang akan dibuat pertanyaan.
2. Klasifikasi setiap kata dalam kalimat yang mengandung kata khusus. Hal ini dilakukan untuk mengetahui posisi kata khusus dan kata target pertanyaan dalam kalimat.
3. Mengetahui klasifikasi kalimat yang mengandung kata khusus.
4. Mengetahui template pola kalimat pertanyaan untuk setiap kategori.

Keempat informasi tersebut sangat penting untuk proses pembangkitkan pertanyaan. Apabila ketiga informasi tersebut diperoleh proses pembangkitkan kalimat dapat dilakukan. Proses pembangkitan kalimat dilakukan dengan cara :

1. Menentukan template pola kalimat pertanyaan yang digunakan didapat dari hasil penentuan kata khusus dan kategori yang telah ditentukan pada kalimat terpilih.

2. Mengambil kata yang berada di sebelum atau setelah kata khusus untuk menjadi target kata pertanyaan. Proses ini dapat dilakukan Karena adanya informasi klasifikasi setiap kata dalam kalimat, posisi kata khusus, dan kata khusus yang digunakan.

3. Menempatkan target kata ke dalam template pola kalimat pertanyaan. Berikut ini adalah contoh pertanyaan yang telah dibangkitkan :

Tabel 5. Hasil pertanyaan yang dibangkitkan

Pi	Pertanyaan
P1	Apakah yang dimaksud dengan teratai ?
P2	Apa yang anda ketahui mengenai teratai ?
P3	Mengapa oleh ?
P4	Mengapa helai daun selalu terlihat mengapung ?

2.5. Pengujian Pertanyaan yang Dihasilkan

Pengujian pertanyaan yang dihasilkan oleh sistem akan dilakukan dengan cara memberikan kuisioner terhadap 10 orang guru SD yang akan memeriksa beberapa parameter , diantaranya :

1. Pertanyaan yang dibangkitkan sistem dituliskan dengan benar.
2. Pertanyaan yang dibangkitkan memiliki jawaban
3. Pertanyaan yang dibangkitkan sistem sesuai dengan konteks.

Ketiga parameter tersebut harus terpenuhi semuanya untuk dapat menentukan apakah pertanyaan yang dihasilkan oleh sistem dinyatakan benar. Sama seperti sebelumnya, terdapat 30 materi yang diujikan dapat dilihat pada tabel 6.

Tabel 6. Hasil pengujian pertanyaan yang dihasilkan

Materi Ke	**Jumlah Pertanyaan**	**Jumlah Pertanyaan Benar**
1	12	9
2	9	7
3	9	8
4	10	10
5	9	10
6	10	9
7	15	10
8	12	5
9	10	10
10	10	7
11	12	10
12	7	9
13	11	7
14	16	9
15	11	9
16	15	4
17	15	12
18	9	9
19	11	5
20	14	4
21	10	8
22	13	6
23	5	8
24	10	10
25	15	7
26	11	5
27	12	7
28	17	9
29	14	9
30	9	9
∑	**343**	**241**

Dari 30 Materi yang diujikan, sistem menghasilkan 343 pertanyaan dan 241 diantaranya dinyatakan benar. Sehingga presentase akurasi pertanyaan yang dapat diterima adalah Jumlah pertanyaan benar dibagi Jumlah pertanyaan yang dibangkitkan dikalikan 100 persen. (241/343)*100% = 70,26%. Jadi dari tabel 4.44 dapat disimpulkan bahwa sistem pembangkit pertanyaan yang dibangun menghasilkan 343 pertanyaan dan 241 pertanyaan benar dengan presentase 70,26%. Pembangkit pertanyaan otomatis ini masih bisa memiliki tingkat akurasi yang tinggi dengan menyeleksi kalimat relevan yang lebih akurat serta dengan memperkaya kata khusus yang digunakan dan menambah template pertanyaan yang digunakan.

3. PENUTUP

Kesimpulan yang didapatkan dari penelitian tugas akhir tentang pembangkit pertanyaan otomatis berdasarkan template dari dokumen ringkasan menggunakan metode *lexical chain* dan Korelasi antar kalimat

1. Tingkat akurasi pemilihan kalimat ringkasan dari materi dengan menggunakan metode *lexical chain* dan korelasi antar kalimat yang akan dibuat pertanyaan adalah 83% yang diperoleh dari pilihan para ahli dalam menentukan kesesuaiannya.
2. *Automatic Text Summarization* atau peringkasan otomatis yang diterapkan terhadap proses pembangkitan pertanyaan menghasilkan akurasi sebesar 70,36%.

Adapun saran yang dapat diberikan untuk meningkatkan tingkat akurasi pertanyaan dan peringkasan yang dilakukan dalam penelitian ini adalah :

1. Memperkaya kamus kata hubung untuk peringkasan, kata khusus untuk menentukan target kata, dan template pertanyaan.
2. Memperkaya kamus kata dasar dan *stop list* untuk mendapatkan ringkasan yang lebih baik serta menghubungkan *chain* kalimat tidak hanya dengan kata yang sama tetapi kata yang memiliki arti yang sama.
3. Menggunakan teknik *stemming* yang lebih baik agar peringkasan memiliki tingkat akurasi yang lebih tinggi.

UCAPAN TERIMA KASIH

Ucapan terimakasih yang sebesar – besarnya kepada Allah Subhanahuwata'ala yang telah memberikan rahmat dan karunianya kepada penulis sehingga dapat menyelesaikan penelitian ini. Terimakasih juga saya ucapkan kepada Bapak Galih Hermawan, S.Kom., M.T. selaku pembimbing yang telah memberikan arahan nya kepada penulis sehingga penulis mampu menyelesaikan penelitian ini. Kepada semua pihak yang telah membantu penelitian ini yang tidak dapat penulis sebutkan satu persatu , penulis mengucapkan terimakasih yang sebesar besarnya.

DAFTAR PUSTAKA

[1] A. C. Graesser, "Intelligent Tutoring System with COnversational Dialogue," *AAAI,* vol. 22, p. 14, 2001.

[2] G. B. Y. Z. Xuchen Yao, "Semantic-Based Question Generation And Implementation," p. 32, 2012.

[3] M. Fachrurrozy, Novi Yusliani, "SISTEM PEMBANGKIT PERTANYAAN OTOMATIS," *Journal of Research in Computer Science and Applications,* vol. Vol.2, no. No.1, p. 6, 2013.

[4] Kulkarni A.R, S.S.Apte, "An Automatic Text Summarization using Lexical Cohesion and Correlation of Sentences," *International Journal of Research in Engineering and Technology,* vol. 03, p. 8, 2014.

[5] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval, England: ACM Press, 1999.

[6] M. L. Murphy and M. Ltd, Lexical Meaning, New York: Cambridge, 2010.

[7] F. Damerau, Handbook of Natural Language Processing Second Edition, New Jersey: Taylor and Francis Group, 2010.

[8] H. S. Kusumadewi, Aplikasi Logika Fuzzy Untuk Pendukung Keputusan, Yogyakarta, Yogyakarta: Graha Ilmu, 2010.

[9] S. Upasana Pandey, Chakraverty, B. Juneja, A. Arora and P. Jain, "Semantic Document Classification using Lexical Chaining & Fuzzy Approach," *International Journal of Soft Computing and Engineering (IJSCE),* vol. 01, p. 5, 2011.