

Sistem Pembangkit Pertanyaan Otomatis Dengan Metode *Template-Based*

M. Fachrurrozi
Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya
Indralaya
mfachrz@unsri.ac.id

Novi Yusliani
Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya
Indralaya
novi_yusliani@unsri.ac.id

Abstrak—Fokus dari penelitian ini adalah untuk membangun sebuah sistem pembangkit pertanyaan Bahasa Indonesia otomatis. Terdapat tiga proses utama yang terjadi di dalam *question generation system*. Pertama adalah mengekstrak kalimat di dalam dokumen. Setelah itu, kalimat yang telah diekstrak akan diklasifikasi berdasarkan kata-kata yang terkandung di dalam kalimat tersebut. Pengklasifikasian kalimat bertujuan untuk menentukan kategori pertanyaan yang bisa dibangkitkan berdasarkan kalimat tersebut. Kategori pertanyaan yang dibangkitkan adalah pertanyaan ‘*non-factoid*’. Proses terakhir adalah membangkitkan pertanyaan berdasarkan kalimat tersebut dengan metode *template-based*. Metode *template-based* adalah salah satu metode yang dapat digunakan untuk membangkitkan pertanyaan. Metode ini menggunakan *template-template* yang telah didefinisikan sebelumnya untuk membangkitkan pertanyaan. Dengan menggunakan 30 dokumen sumber, terdapat 1871 pertanyaan ‘*non-factoid*’ yang berhasil dibangkitkan oleh sistem.

Keywords—*component; question generation, pertanyaan non-factoid, template based*

I. PENDAHULUAN

Pada peristiwa belajar dan pembelajaran, hasil atau prestasi yang diperoleh oleh siswa berbeda-beda, walaupun siswa dibimbing oleh guru dengan bahan pelajaran, waktu, tempat, dan metode yang sama. Guru pada dasarnya bertanggung jawab atas keseluruhan proses pembelajaran di kelas. Oleh karena itu, bantuan mengatasi kesulitan belajar dan memperkaya hasil belajar diperlukan untuk membantu proses belajar di luar sekolah. Salah satu bantuan untuk mengatasi masalah ini yaitu dengan memberikan latihan soal kepada siswa. Hal ini diharapkan dapat mempermudah siswa dalam memahami materi yang diberikan. Akan tetapi, jika pertanyaan yang dibuat banyak dengan mempertimbangkan waktu yang efisien maka tidaklah mudah.

Dengan adanya tuntutan untuk memenuhi kebutuhan membuat pertanyaan secara cepat dan tepat, maka dikembangkanlah sistem pembangkit pertanyaan otomatis. Penelitian dalam bidang ini berkaitan dengan bagaimana sistem dapat membangkitkan pertanyaan berdasarkan informasi atau teks yang ada. Selain itu, sistem ini diharapkan bekerja seperti halnya manusia yang dapat membuat pertanyaan ketika diberikan sebuah teks. Manusia dapat membuat pertanyaan dikarenakan manusia dapat memahami teks yang diberikan dan berdasarkan pengetahuan-pengetahuan yang dimiliki.

Penelitian di bidang sistem pembangkit pertanyaan otomatis sudah pernah dilakukan [7] [8]. Kontribusi yang diberikan dalam penelitian tersebut adalah membangkitkan pertanyaan secara otomatis dari sebuah teks. Pembangkitan pertanyaan dilakukan dengan melihat kalimat yang telah diekstraksi dari teks yang ada. Kategori pertanyaan yang dibangkitkan didasarkan pada hasil klasifikasi kalimat tersebut. Teks yang digunakan oleh mereka adalah teks berbahasa Inggris.

Beberapa bahasa seperti bahasa Inggris, bahasa-bahasa di benua Asia bagian Timur, dan bahasa-bahasa di benua Eropa telah tersedia banyak sumber yang dapat digunakan untuk membantu penelitian sistem pembangkit pertanyaan otomatis. Sedangkan bahasa-bahasa di benua Asia selain bahasa Jepang, masih menyediakan sedikit sumber yang dapat digunakan untuk membantu penelitian sistem pembangkit pertanyaan otomatis termasuk salah satunya bahasa Indonesia. Karena itu, penelitian yang dilakukan yaitu membangun sebuah sistem yang dapat membangkitkan pertanyaan otomatis dalam teks berbahasa Indonesia dengan menggunakan metode *template-based*.

Secara umum, pertanyaan dapat diklasifikasikan menjadi lima kategori, yaitu pertanyaan yang bersifat ‘*factoid*’ (*factoid question*), pertanyaan yang bersifat ‘*non-factoid*’ (*non-factoid question*), *list question*, *yes/no question*, dan *opinion question*. *Question generation system* (sistem pembangkit pertanyaan) bermanfaat dalam pembuatan soal-soal latihan secara otomatis yang dapat membantu siswa atau mahasiswa (pembelajar) belajar. Sistem ini dapat membangkitkan pertanyaan secara otomatis dengan memanfaatkan sekumpulan dokumen yang ada.

II. METODOLOGI

Question Generation berfungsi untuk membangkitkan pertanyaan dari dokumen yang diberikan kepada sistem. Proses yang dilakukan pada proses ini adalah membangkitkan pertanyaan dengan menggunakan kata tanya yang sesuai dengan kategori kalimat. Adapun proses yang harus dilakukan untuk pembangkitan pertanyaan adalah:

1. Mengetahui sintaksis yang membentuk pertanyaan. Sintaksis untuk ‘*non-factoid question*’ pada umumnya berbeda dengan sintaksis yang menyusun ‘*factoid-question*’. Kata tanya dalam bahasa Indonesia

diantaranya apa, mengapa, siapa, bagaimana, kapan, dimana, dan sebagainya. Pada penelitian ini, pertanyaan yang dapat dibangkitkan merupakan pertanyaan ‘*non-factoid question*’ dengan kata tanya apa, mengapa, dan bagaimana.

2. Mengetahui kata khusus dari kalimat yang mencirikan kategori pertanyaan yang akan dibangkitkan

Kata khusus untuk setiap kategori kalimat berbeda. Kata khusus yang diperlukan dapat dilihat pada tabel 1. Kata khusus dapat berfungsi sebagai kata penghubung atau konjungsi pada sebuah kalimat. Kalimat alasan merupakan kalimat yang mengandung sebab-akibat terhadap sesuatu. Kata penghubung yang biasa digunakan dalam kalimat ini adalah ‘karena’. Kata tanya yang membutuhkan jawaban berupa alasan adalah ‘mengapa’. Sehingga pertanyaan yang dapat dibangkitkan dari kalimat yang menyatakan sebab-akibat dengan kata khusus ‘karena’ adalah pertanyaan alasan, yaitu dengan kata tanya ‘mengapa’. Contoh pertanyaan untuk tiap kategori dapat dilihat pada tabel 2.

TABEL I
DAFTAR KATA KHUSUS SETIAP KATEGORI

Kategori	Kata khusus Sebelum target kata untuk pertanyaan	Kata khusus Setelah target kata untuk pertanyaan
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

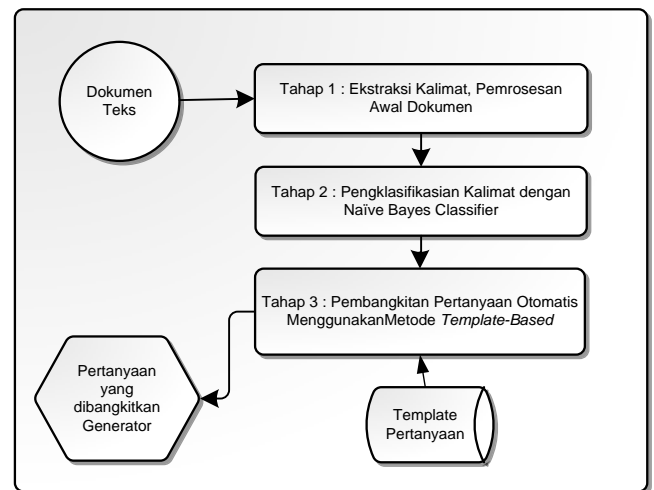
TABEL II
CONTOH PERTANYAAN SETIAP KATEGORI

Kategori	Contoh Pertanyaan
definisi	Apa yang dimaksud dengan pencernaan kimiawi ? Apa yang dimaksud dengan fotosintesis ?
alasan	Mengapa fotosintesis dapat terjadi pada siang maupun malam hari ? Mengapa astronot dapat melayang-layang di bulan ?
metode	Bagaimana cara mengubah protein menjadi asam amino ? Bagaimana cara dna membentuk rna ?

Gambar 1 merupakan arsitektur dari sistem pembangkit pertanyaan otomatis. Masukan sistem ini berupa dokumen teks yang telah disediakan dalam format teks (.txt) yang kemudian diekstrak ke dalam kumpulan kalimat. Keluaran dari sistem ini berupa sekumpulan pertanyaan yang dibangkitkan berdasarkan kalimat yang telah diklasifikasikan.

Tahap pertama yaitu ekstraksi kalimat dan *preprocessing* awal dokumen. Pada tahap ini, seluruh kalimat yang ada pada dokumen teks diekstrak, kemudian diuraikan dengan pemberian label klasifikasi pada setiap kata yang disebut dengan *POS Tagger*. Tahap kedua yaitu melakukan klasifikasi kalimat dengan *Naive Bayes Classifier*. Terdapat tiga kategori kalimat yang digunakan dalam penelitian ini yaitu definisi,

alasan, dan metode. Kalimat definisi merupakan kalimat yang menjelaskan suatu istilah. Kalimat alasan merupakan kalimat yang mengandung hubungan sebab-akibat. Kalimat metode merupakan kalimat yang menjelaskan bagaimana cara sesuatu dilakukan atau bekerja. Pengklasifikasian kalimat pada tahap ini menggunakan kata khusus yang mewakili masing-masing kategori kalimat.



Gambar 1 Arsitektur Sistem

Jika kalimat termasuk ke dalam salah satu kategori yang telah ditentukan, maka kalimat tersebut merupakan salah satu kandidat sumber pertanyaan yang akan dibangkitkan. Tahap terakhir yaitu pembangkitan pertanyaan otomatis menggunakan metode *template-based*. Pada tahap ini, pertanyaan akan dibangkitkan dari kalimat yang telah diklasifikasikan pada tahap sebelumnya. Kalimat tersebut kemudian dimasukkan ke dalam *template-template* pertanyaan yang telah disediakan sesuai kategorinya, sehingga proses pembangkitan pertanyaan dapat dilakukan. *Template* pertanyaan yang digunakan pada penelitian ini dapat dilihat pada gambar 2.

- | |
|---|
| <ol style="list-style-type: none"> 1. Apa yang dimaksud dengan 2. Mengapa 3. Bagaimana cara |
|---|

Gambar 2 Template Pertanyaan

III. HASIL DAN PEMBAHASAN

Pertanyaan yang dibangkitkan oleh sistem sebanyak 1871. Pertanyaan ini bersumber dari 30 dokumen teks yang masing-masing digunakan sebagai masukan. Contoh pertanyaan yang dibangkitkan oleh sistem dapat dilihat pada gambar 3.

Pengujian dilakukan terhadap pertanyaan yang telah dibangkitkan oleh sistem melalui kuisioner yang diberikan kepada 30 orang mahasiswa. Pertanyaan yang dinyatakan dapat diterima harus memenuhi salah satu ataupun sebagian dari kategori yang telah ditentukan. Kategori yang dapat

menyatakan sebuah pertanyaan dapat diterima atau tidak, jika pertanyaan tersebut memenuhi yaitu:

1. Pertanyaan yang dibangkitkan sistem dituliskan dengan benar.
2. Maksud dari pertanyaan yang dibangkitkan sistem dapat dimengerti.
3. Pertanyaan yang dibangkitkan sistem sesuai dengan konteks.
4. Jika pembaca membuat pertanyaan dari teks yang diberikan, pertanyaan yang pembaca buat sama seperti pertanyaan yang telah dibangkitkan sistem.
5. Pertanyaan yang dibangkitkan sistem merupakan pertanyaan yang bermanfaat.

No	Nama Artikel	Pt	Pd	Pr
16	Alat Pernapasan	39	26	66.66
17	Susunan Bumi	40	27	67.5
18	Gaya dan Gerak	40	29	72.5
19	Gaya	41	29	70.73
20	Pembentukan Tanah	36	25	69.44
21	Pelestarian Makhluk Hidup	30	22	73.33
22	Dampak Peristiwa Alam	18	9	50
23	Sifat Bahan	31	18	58.06
24	Energi Listrik	47	34	72.34
25	Pengantar Listrik	20	12	60
26	Perubahan pada Benda	60	45	75
27	Struktur dan Fungsi Jaringan pada Hewan	81	39	48.14
28	Ekosistem (2)	150	105	70
29	Alat Peredaran Darah	65	35	53.84
30	Organisasi Kehidupan	52	26	50

Tabel 3 menunjukkan jumlah pertanyaan yang berhasil dibangkitkan oleh sistem berdasarkan dokumen yang digunakan sebagai masukan. Pt merupakan jumlah pertanyaan yang dibangkitkan oleh sistem. Pd merupakan jumlah pertanyaan yang diterima berdasarkan kategori yang telah ditentukan. Pr merupakan persentase yang dihasilkan antara jumlah pertanyaan yang dapat diterima dari seluruh pertanyaan yang dibangkitkan sistem. Nilai Pr didapatkan dari $Pd/Pt \times 100$. Dari 1871 pertanyaan yang dibangkitkan oleh sistem, 1047 diantaranya dapat diterima oleh *tester*, sehingga persentasenya adalah 55.95%.

IV. KESIMPULAN

Sistem pembangkit pertanyaan otomatis dalam bahasa Indonesia yang dibangun menggunakan *template-based method*. Pembangkitan pertanyaan dari sekumpulan dokumen teks yang disediakan dapat dilakukan dengan mengekstrak kalimat dari dokumen kemudian melakukan pengklasifikasi kalimat berdasarkan kata khusus yang terkandung dalam teks tersebut. Dengan adanya penelitian ini, maka dihasilkan sebuah sistem yang dapat membangkitkan pertanyaan '*non-factoid*' dalam jumlah banyak dengan mempertimbangkan waktu yang efisien.

Sistem pembangkit pertanyaan otomatis dalam bahasa Indonesia ini masih terdapat beberapa kekurangan, sehingga apabila dikembangkan maka sebaiknya menambahkan istilah biologi pada kamus kata dasar agar istilah tersebut dapat diberi kategori, menambahkan *template* pertanyaan untuk jenis pertanyaan yang lain, serta dapat menghasilkan pertanyaan yang tidak berulang dengan menerapkan penyaringan semantik.

DAFTAR PUSTAKA

- [1] Adriani, M., Jelita, A., Bobby, N., Tahaghoghi, S., & F. W., H. (2006). Stemming Indonesian: A Confix-Stripping Approach. *ACM Transactions on Asian Language Information Processing Vol. 6, No. 4.*

<p><u>Kalimat yang diekstrak:</u> Flagela adalah struktur tambahan pada tubuh berupa bulu cambuk yang berfungsi sebagai alat gerak.</p> <p><u>Kata khusus:</u> adalah</p> <p><u>Kategori:</u> definisi</p> <p><u>Template:</u> Apa yang dimaksud dengan</p> <p><u>Pertanyaan yang dibangkitkan:</u> <i>Apa yang dimaksud dengan flagela?</i></p>
<p><u>Kalimat yang diekstrak:</u> Bumi dapat tetap berada pada orbitnya karena gaya gravitasi yang bekerja antara bumi dan matahari.</p> <p><u>Kata khusus:</u> karena</p> <p><u>Kategori:</u> alasan</p> <p><u>Template:</u> Mengapa</p> <p><u>Pertanyaan yang dibangkitkan:</u> <i>Mengapa bumi dapat tetap berada pada orbitnya?</i></p>
<p><u>Kalimat yang diekstrak:</u> Organisme pengurai memperoleh makanan dengan cara merombak sisa produk organisme dan organisme yang mati dengan enzim pencernaan yang dimilikinya.</p> <p><u>Kata khusus:</u> dengan cara</p> <p><u>Kategori:</u> metode</p> <p><u>Template:</u> Bagaimana cara</p> <p><u>Pertanyaan yang dibangkitkan:</u> <i>Bagaimana cara organisme pengurai memperoleh makanan?</i></p>

Gambar 3 Contoh Pertanyaan yang Dibangkitkan Sistem

TABEL III
EVALUASI PERTANYAAN YANG DIBANGKITKAN SISTEM

No	Nama Artikel	Pt	Pd	Pr
1	Tata Surya	124	57	45.96
2	Evolusi	99	34	34.34
3	Metabolisme	181	76	41.98
4	Ekosistem	101	18	17.82
5	Ciri-ciri Khusus Makhluk Hidup	35	18	51.42
6	Cahaya	28	20	71.42
7	Klasifikasi Makhluk Hidup	61	41	67.21
8	Bioteknologi	77	42	54.54
9	Perubahan Sifat Benda	21	10	47.61
10	Virus	91	41	45.05
11	Peristiwa Alam	18	11	61.11
12	Kerja Ilmiah	51	22	43.13
13	Alat Pencernaan	69	45	65.21
14	Keanekaragaman Makhluk Hidup	109	95	87.15
15	Sistem Tubuh	56	36	64.28

- [2] Agusta, L. (2009). Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief dan Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika*. Bali.
- [3] Ali, H., Chali, Y., & Hasan, S. A. (2010). Automatic Question Generation from Sentences. *TALN 2010, Montréal*.
- [4] Ayache, C., Grau, B., & Vilnat, A. (2006). EQueR : the French Evaluation campaign of Question Answering system EQueR/EVALDA. *Proceedings of the 5th international Conference on Language Resources and Evaluation*, 1157-1160.
- [5] Barakbah, A. R. (2010). Natural Processing Language.
- [6] Iftene, A., Diana, T., Maria, H., & Mihai, A. (2010). *Question Answering on Romanian, English, and French Languages*. Padua Italia: CLEF 2010 LABs and Workshops, Notebook Papers.
- [7] Liu, M., & Calvo, R. (2012). G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support. *Dialogue and Discourse, School of Electrical and Information Engineering, University of Sydney, Sydney NSW 2006, Australia*, 101–124.
- [8] Manning, C. D., Prabhakar R., & Hinrich S. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- [9] Mori, T. (2007). A Monolithic Approach and a Type-by-Type Approach for Non-Factoid Question-Answering. *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*.
- [10] Niu, Y. (2007). *Analysis of Semantic Classes: Toward Non-Factoid Question Answering*. Toronto: Department of Computer Science, University of Toronto.
- [11] Saxena, A. K. (2007). *IITD-IBMIRL System for Question Answering using Pattern Matching, Semantic Type, and Semantic Category Recognition*. India: IBM India Research Lab.
- [12] Sukanto, R. A. (2009). *Penguraian Bahasa Indonesia Dengan Menggunakan Pengurai Collins*. Bandung: Institut Teknologi Bandung.
- [13] Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Universiteit van Amsterdam.
- [14] Wang, R. C. (2008). Automatic Set Expansion for List Question Answering. *Proceeding EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- [15] Purwarianti, A., & Yusliani, N. (2012). SISTEM QUESTION ANSWERING BAHASA INDONESIA UNTUK PERTANYAAN NON-FACTOID. *Jurnal Ilmu Komputer dan Informasi*, 4(1), 10.