PENDETEKSIAN KALIMAT NON-DEFINISI PADA PEMBANGKIT PERTANYAAN OTOMATIS UNTUK TEKS BERBAHASA INDONESIA

Wahyu Hidayat[1], Ken Kinanti Purnamasari[2]

[1,2] Teknik Informatika, Universitas Komputer Indonesia
Jalan Dipatiukur No. 112-116, Coblong, Lebakgede, Bandung, Kota Bandung, Jawa Barat 40132, Indonesia
E-mail : wahjoe16@gmail.com[1], kinantiken@gmail.com[2]

ABSTRAK

Penelitian ini merupakan pengembangan penelitian sebelumnya tentang membangun pembangkit pertanyaan otomatis untuk teks berbahasa Indonesia yang mengandung kalimat majemuk. Penelitian pembangkit pertanyaan pada kalimat majemuk dilakukan untuk menghasilkan pertanyaan *non-factoid* yaitu pertanyaan mengenai definisi, alasan, dan cara serta untuk menghasilkan pertanyaan *factoid* yaitu pertanyaan mengenai apa, siapa, dimana/kemana, berapa, dan kapan. Akan tetapi pada penelitian tersebut masih ditemukan pertanyaan yang tidak sesuai dengan konteks kalimat yang dijadikan masukan. Hal ini terjadi pada kalimat dengan kata kunci *non-factoid* dengan kategori definisi seperti kata ‘adalah’ dan ‘ialah’ selalu terdeteksi sebagai kandidat pertanyaan *non-factoid* dengan kategori definisi padahal tidak semua kalimat yang mengandung kata ‘adalah’ dan ‘ialah’ merupakan kalimat definisi.

Proses keseluruhan yang akan dilakukan oleh sistem terdiri dari dua tahap yaitu *preprocessing* yang diadopsi dari penelitian sebelumnya dan proses utama. *Preprocessing* adalah tahap awal sistem untuk mendapatkan jenis kata, entitas bernama, jenis frasa, dan fungsi gramatikal untuk setiap kata pada kalimat, sedangkan proses utama adalah proses deteksi kalimat non-definisi yang menggunakan metode *rule-based* dan proses pembangkit pertanyaan dengan metode *template*.

Dengan menggunakan 48 teks masukan yang diambil dari buku mata pelajaran bahasa Indonesia sekolah elektronik Kemendikbud RI, diperoleh nilai akurasi pertanyaan sebesar 68,84% dengan total pertanyaan yang berhasil dibangkitkan sebanyak 398 pertanyaan.

Kata kunci : pembangkitan pertanyaan, kalimat definisi, deteksi kalimat non-definisi, kalimat majemuk, *rule-based*.

1. PENDAHULUAN

Penelitian tentang sistem pembangkit pertanyaan telah dilakukan sebelumnya dengan metode sintaksis dan metode statistik[1][2][3]. Untuk bahasa Indonesia penelitian tentang pembangkit pertanyaan dilakukan dengan metode *template*[4][5] dan metode semantik[6]. Adapun pembangkit pertanyaan dalam bahasa Indonesia dilakukan untuk menghasilkan pertanyaan *non-factoid* yaitu pertanyaan mengenai definisi, alasan, dan cara[4][5] serta untuk menghasilkan pertanyaan *factoid* yaitu pertanyaan mengenai apa, siapa, dimana/kemana, berapa, dan kapan[7]. Pembangkit pertanyaan berguna untuk membantu proses evaluasi dalam kegiatan belajar-mengajar baik untuk peserta didik maupun pengajar. Pembangkit pertanyaan juga dapat digunakan untuk membangun *help system* seperti *Frequently Ask And Question* (FAQ)[1][5].

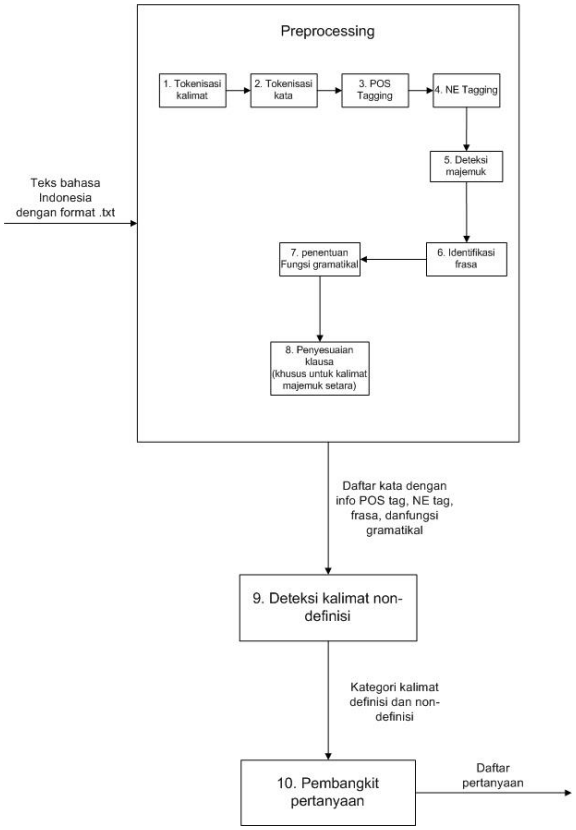
Dalam penelitian sebelumnya sudah dilakukan penelitian pembangkit pertanyaan untuk kalimat sederhana[7][8] dan kalimat majemuk[8]. Hasil pengujian dari penelitian yang tersebut sudah menghasilkan akurasi sampai dengan 64,52%. Pengujian tersebut memakai data masukan sebanyak 35 teks masukan yang mana menghasilkan 327 pertanyaan. Hasil dari 327 pertanyaan tersebut 211 pertanyaan diantaranya menjadi pertanyaan yang akurat dan 116 pertanyaan yang tidak akurat. Diantara pertanyaan-pertanyaan yang tidak akurat pada penelitian tersebut adalah masih ditemukan pertanyaan yang tidak sesuai dengan konteks pada kalimat yang dijadikan masukan sistem. Hal ini terjadi pada kalimat yang mengandung kata kunci pertanyaan *non-factoid* dengan kategori definisi seperti ‘adalah’ dan ’ialah’ selalu terdeteksi sebagai kandidat pertanyaan *non-factoid* dengan kategori definisi, padahal tidak semua kalimat yang mengandung kata kunci pertanyaan *non-factoid* dengan kategori definisi merupakan kalimat definisi.

Untuk mengatasi permasalahan tidak bisa membedakan kalimat defnisi dan non-definisi diperlukan proses deteksi kalimat non-definisi berdasarkan makna yang terkandung dari klausa sebelum dan setelah kata kunci pertanyaan *non-factoid* dengan kategori definisi. Proses deteksi

kalimat non-definisi dilakukan dengan cara pencarian kata yang mengandung makna pelaku (orang), lokasi, kuantitas, dan alat sebelum dan setelah kata kunci pertanyaan *non-factoid* dengan kategori definisi.

2. ISI PENELITIAN

Untuk mengatasi permasalahan pembangkit pertanyaan yang tidak tepat dikarenakan pertanyaan yang seharusnya merupakan pertanyaan *factoid* menjadi pertanyaan *non-factoid*, pada penelitian ini akan dilakukan proses deteksi kalimat non-definisi setelah proses penentuan fungsi gramatikal dan sebelum proses pembangkit pertanyaan. Proses deteksi kalimat non-definisi dilakukan dengan cara melakukan pencarian dari kata pertama hingga kata terakhir kata yang membentuk makna pelaku, kuantitas, dan lokasi sebelum dan setelah kata 'adalah' dan 'ialah'.



Gambar 1. Blok Diagram Alur Sistem

Pada sistem yang akan dibangun ini, *user* mengunggah teks berisi kalimat atau kumpulan kalimat baku dalam bahasa Indonesia, kemudian sistem ini akan mengolah teks masukan yang memiliki alur sebagai berikut; 1) sistem melakukan proses tokenisasi kalimat, 2) sistem melakukan proses tokenisasi kata, 3) sistem melakukan pendeteksian jenis kata (*POS Tagging*), 4) sistem melakukan proses pengenalan entitas bernama (NER), 5) sistem melakukan proses pendeteksian kalimat majemuk berdasarkan keberadaan kata penghubung yang telah didefinisikan, 6) sistem melakukan proses identifikasi frasa berdasarkan pola frasa yang telah didefinisikan, 7) sistem melakukan proses penentuan fungsi gramatikal untuk menentukan subjek, predikat, objek, dan keterangan berdasarkan aturan pengisi fungsi gramatikal, 8) sistem melakukan proses penyesuaian klausa khusus untuk kalimat majemuk setara, 9) sistem melakukan deteksi kalimat non-definisi, 10) sistem melakukan proses pembangkit pertanyaan.

2.1 *Preprocessing*

Preprocessing merupakan proses awal yang dilakukan oleh sistem setelah mendapatkan *file* teks dengan format .txt. Pada tahap ini teks masukan akan dilakukan tokenisasi kalimat, tokenisasi kata, pelabelan jenis kata (POS *tagging*), pengenalan entitas bernama (NER), pendeteksian kalimat majemuk, identifikasi frasa, penentuan fungsi gramatikal, dan penyesuaian klausa untuk kalimat majemuk setara. Proses ini dilakukan pada penelitian sebelumnya[8].

2.1.1 Tokenisasi Kalimat

Teks yang diunggah ke dalam sistem berupa paragraf yang diambil dari buku mata pelajaran bahasa Indonesia. Paragraf tersebut tentunya terdiri dari kumpulan kalimat sehingga diperlukan tokenisasi kalimat untuk memisahkan antar kalimat. Hal ini diperlukan karena pertanyaan yang akan dibangkitkan sistem secara per kalimat. Tokenisasi kalimat dilakukan dengan cara mengenali tanda baca titik (.), tanda seru (!), atau tanda tanya (?).

2.1.2 Tokenisasi Kata

Setelah kalimat dipisahkan dalam proses tokenisasi kalimat, proses selanjutnya adalah tokenisasi kata yang dilakukan untuk memisahkan kata-kata pada setiap kalimat menjadi satuan kata. Tokenisasi kata adalah pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-*scan* kalimat dan memisahkan kata demi kata dengan spasi sebagai *delimiter.*

2.1.3 POS *Tagging*

Proses selanjutnya yang akan dilakukan sistem adalah pendeteksian jenis kata (POS *tagging). Part of Speech Tagging (POS-Tagging)* adalah suatu proses yang memberikan label kelas kata secara otomatis pada suatu kata dalam kalimat[9]. Hasil dari POS ini sangat berpengaruh terhadap keluaran dari proses *Parsing*[10]. POS *tagger* yang dipakai pada penelitian ini adalah POS *tagger* dengan metode HMM[11].

2.1.4 NE *Tagging*

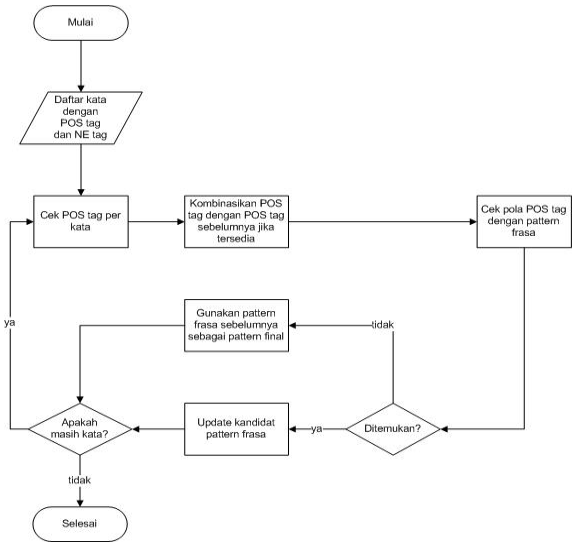
Proses selanjutnya yang akan dilakukan sistem adalah pengenalan entitas bernama (NER). *Named Entity Recognition (NER)* adalah sebuah proses yang digunakan untuk mengenali dan mengklasifikasi istilah dan frasa sebagai entitas nama dari sebuah teks bebas[12]. NER adalah salah satu komponen utama dari *information extration* yang bertujuan untuk mendeteksi dan mengklasifikasikan *named-entity* pada suatu teks. NER umumnya digunakan untuk mendeteksi nama orang, nama tempat dan organisasi dari sebuah dokumen, tetapi dapat juga diperluas untuk identifikasi gen, protein dan lainnya sesuai kebutuhan. Penelitian ini menggunakan aplikasi Indonesia NER, dan menggunakan metode HMM[13][14].

2.1.5 Deteksi Kalimat Majemuk

Proses selanjutnya yang akan dilakukan sistem adalah pendeteksian kalimat majemuk terhadap semua kalimat yang berada pada teks masukan. Kalimat yang diproses akan dikelompokan menjadi tiga, yaitu kalimat majemuk setara, kalimat majemuk bertingkat dan kalimat lainnya. Proses pengelompokan kalimat mejemuk dilakukan berdasarkan pada identifikasi keberadaan kata penghubung di dalam kalimat.

2.1.6 Identifikasi Frasa

Proses selanjutnya yang akan dilakukan sistem adalah identifikasi frasa. Proses identifikasi frasa mengacu pada pola frasa yang telah dijelaskan dan dilakukan pada penelitian yang sebelumnya[8]. Identifikasi frasa pada sebuah klausa dilakukan dengan cara mencocokan kombinasi jenis kata di dalam klausa dengan pola frasa yang telah didefinisikan. Pendeteksian pola frasa dilakukan berurutan dari kata pertama sampai kata terakhir di dalam klausa.



Gambar 2. Flowchart Identifikasi Frasa

Contoh hasil identifikasi frasa untuk klausa "*Ciri utama dari sistem tanam paksa adalah mewajibkan rakyat di Jawa untuk membayar pajak dalam bentuk barang dengan hasil-hasil pertanian yang mereka tanam"*

Hasil Akhir:

1. **Ciri utama (FN)**
2. **dari sistem tanam (FIN)**
3. **paksa adalah (FVB)**
4. mewajibkan (VB)
5. rakyat (NN)
6. **di jawa (FIN)**
7. **untuk membayar (FIN)**
8. pajak (NN)
9. **dalam bentuk barang (FIN)**
10. **dengan hasil-hasil pertanian (FIN)**
11. yang (SC)
12. mereka (PR)
13. tanam (VB)

2.1.7 Penentuan Fungsi Gramatikal

Proses selanjutnya yang akan dilakukan sistem adalah proses penentuan fungsi gramatikal. Proses penentuan fungsi gramatikal dilakukan berdasarkan aturan tata bahasa pembentukan pola kalimat yang telah dilakukan pada penelitian sebelumnya[8]. Penentuan fungsi gramatikal khusus pada kalimat majemuk bertingkat hanya dilakukan pada klausa induk saja sedangkan pada kalimat majemuk setara, penentuan fungsi gramatikal akan dilakukan pada kedua klausa yang telah dipisahkan pada tahap proses deteksi majemuk. Proses penentuan fungsi gramatikal hampir sama dengan proses identifikasi frasa, yaitu dilakukan pengecekan setiap kata maupun frasa yang terdapat pada klausa dari awal sampai akhir secara berurutan.

Tabel 1. Aturan Penentuan Fungsi Gramatikal

No	Aturan
1	Jika kata pertama pada klausa adalah konjungsi, jadikan kata tersebut sebagai klausa *keterangan* sampai tanda baca koma (,).
2	Jika belum ada subjek, dan jenis kata atau jenis frasa sekarang termasuk jenis pembentuk subjek, jadikan kata atau frasa tersebut sebagai subjek.
3	Jika kata atau frasa sebelumnya adalah subjek, dan jenis kata atau jenis frasa sekarang termasuk jenis pembentuk predikat, jadikan kata atau frasa tersebut sebagai predikat.
4	Jika kata atau frasa sebelumnya adalah predikat berjenis verba transitif (VBT) dan jenis kata atau jenis frasa sekarang termasuk jenis pembentuk objek, jadikan kata atau frasa tersebut sebagai objek.

No	Aturan
5	Jika kata atau frasa sebelumnya adalah predikat berjenis verba intransitif (VBI), jadikan kata atau frasa sekarang sebagai pelengkap.
6	Jika fungsi subjek, predikat, objek sudah ada. Dan Jika kata atau frasa sekarang adalah konjungsi, jadikan kata atau frasa tersebut sebagai klausa keterangan sampai akhir klausa.
7	Jika fungsi subjek, predikat, objek sudah ada. Dan Jika kata atau frasa sekarang mengandung kata depan, jadikan kata atau frasa tersebut sebagai keterangan sampai akhir klausa.
8	Jika fungsi subjek, predikat, objek sudah ada. Dan Jika kata atau frasa sekarang **TIDAK** mengandung kata depan, jadikan kata atau frasa tersebut sebagai pelengkap sampai akhir klausa.

Tabel 2 berikut ini adalah contoh penentuan gramatikal pada kasus kalimat *Sistem tanam paksa adalah kebijakan Gubernur Jendral Van den Bosh yang mewajibkan para petani Jawa untuk menanam tanaman-tanaman yang dapat diekspor ke pasaran dunia*.

Tabel 2. Contoh Penetuan Fungsi Gramatikal

No	Proses
1	Frasa / jenis kata pertama *Sistem* merupakan jenis kata NN. Kondisi ini memenuhi aturan ke-2. Maka jenis kata *Sistem* ditentukan sebagai subjek (S)
2	Frasa / jenis kata berikutnya *tanam paksa* merupakan FVB dengan jenis verba transitif (VBT). Kondisi ini memenuhi aturan ke-3. Maka frasa *tanam paksa* ditentukan sebagai predikat (P)
3	Kata atau frasa berikutnya *adalah kebijakan Gubernur Jendral Van den Bosh yang mewajibkan para petani Jawa untuk menanam tanaman-tanaman yang dapat diekspor ke pasaran dunia* merupakan klausa keterangan. Kondisi ini memenuhi aturan ke-6. Maka klausa tersebut ditentukan sebagai keterangan (K)

2.1.8 Penyesuaian Klausa

Proses selanjutnya yang akan dilakukan sistem adalah proses penyesuaian klausa. Proses penyesuaian klausa hanya dilakukan untuk kalimat majemuk setara. Fungsinya adalah untuk melakukan penyesuaian klausa kedua dengan klausa pertama. Kalimat majemuk setara dapat dipecah menjadi klausa-klausa bebas. Kemudian, klausa-klausa bebas ini akan diproses pada modul pembangkit pertanyaan satu persatu secara bergantian sehingga memungkinkan dihasilkannya pertanyaan maupun jawaban yang ambigu.

Tabel 3. Aturan Penyesuaian Klausa

No	Aturan
1	Jika subjek klausa kedua adalah kata ganti (PR), maka kata ganti tersebut mengacu pada subjek di klausa pertama.
2	Jika klausa kedua tidak memiliki subjek, maka subjek klausa kedua sama dengan subjek klausa pertama.
3	Jika objek pada klausa kedua adalah kata milik (NNG), maka objek tersebut sama dengan objek pada klausa pertama.

2.2 Deteksi Kalimat Non-Definisi

Proses deteksi kalimat non-definisi dilakukan dengan cara pencarian kata yang bisa menentukan bahwa kalimat tersebut merupakan kalimat non-definisi. Kata tersebut bisa berupa kata tunggal ataupun penggabungan dua buah kata atau lebih yang membentuk makna tertentu. Makna tersebut bisa membentuk makna tentang lokasi, waktu, pelaku, dll.

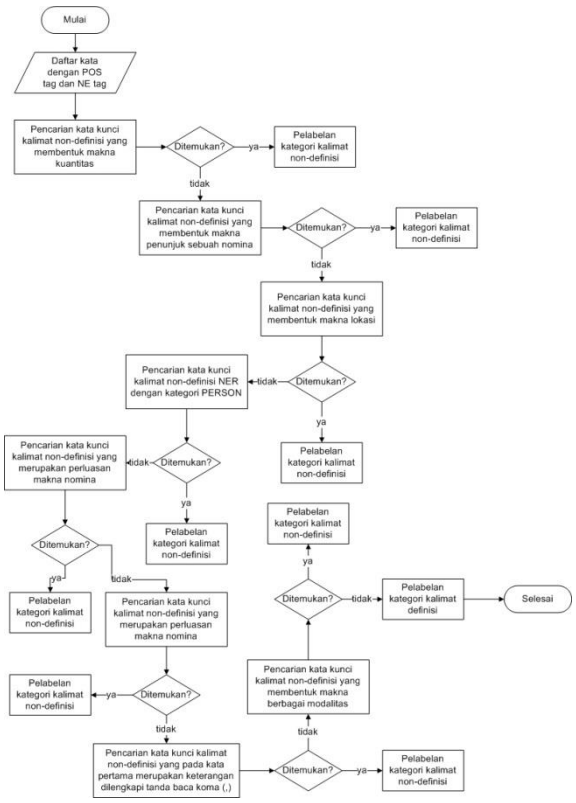
Proses pencarian kata dilakukan secara berurutan dari kata pertama sampai kata terakhir. Penentuan kata yang merupakan penentu kalimat non-definisi bisa dari kata-kata sebelum kata “adalah” ataupun dari kata-kata setelah kata “adalah”. Aturan deteksi kalimat non-definisi secara lengkap dapat dilihat pada tabel 4.

Tabel 4. Aturan Deteksi Kalimat Non-Definisi

No	Aturan
1	Jika terdapat penggabungan kata antara nomina dengan pendamping nomina-nya (beberapa, semua, seluruh, sejumlah, banyak) sebelum kata ‘adalah’, ‘ialah’ dan membentuk makna kuantitas atau jumlah, maka kalimat tersebut merupakan kalimat non-definisi.
2	Jika terdapat penggabungan kata antara nomina dengan pendamping nomina-nya (ini, itu) sebelum kata ‘adalah’, ‘ialah’ dan membentuk makna penunjuk nomina tersebut, maka kalimat tersebut merupakan kalimat non-definisi.
3	Jika terdapat penggabungan kata antara nomina dengan pendamping nomina-nya (di, pada, dari, ke, kepada) sebelum kata ‘adalah’, ‘ialah’ dan membentuk makna lokasi, maka kalimat tersebut merupakan kalimat non-definisi.
4	Jika terdapat NER dengan kategori PERSON sebelum kata ‘adalah’, ‘ialah’ maka kalimat tersebut merupakan kalimat non-definisi.
5	Jika terdapat penggabungan kata antara nomina dengan pendamping nomina-nya (yang) sebelum kata ‘adalah’, ‘ialah’ dan merupakan perluasan makna nomina tersebut, maka kalimat tersebut merupakan

	kalimat non-definisi.
6	Jika pada kata pertama merupakan keterangan yang dilengkapi dengan tanda baca koma (,) dan sebelum kata penghubung ‘adalah’, ‘ialah’, maka kalimat tersebut merupakan kalimat non-definisi.
7	Jika terdapat penggabungan kata antara verba dengan pendamping verba-nya (belum, sedang, akan, boleh, dapat, harus, wajib, mesti, jangan) sebelum kata ‘adalah’, ‘ialah’ dan membentuk makna berbagai modalitas, maka kalimat tersebut merupakan kalimat non-definisi.

Berdasarkan aturan deteksi kalimat non-definisi pada tabel 4, gambar 3 berikut akan menjelaskan alur dari aturan deteksi kalimat non-definisi.



Gambar 3. Flowchart Deteksi Kalimat Non-Definisi

Berikut adalah hasil dari deteksi kalimat non-definisi.

Tabel 5. Contoh Proses Deteksi Kalimat Non-Definisi

No	Proses
1	*<u>Sistem tanam paksa</u>* ***adalah*** *kebijakan Gubernur Jendral Van den Bosh yang mewajibkan para petani Jawa untuk menanam tanaman-tanaman yang dapat diekspor ke pasaran dunia.* Kata-kata sebelum kata “adalah” yaitu *Sistem, tanam,* dan *paksa.* Kondisi ini tidak ada yang memenuhi aturan pada tabel 7, maka kalimat ini merupakan kalimat definisi.
2	*<u>Ciri utama dari sistem tanam paksa</u>* ***adalah*** *mewajibkan rakyat di Jawa untuk membayar pajak dalam bentuk barang dengan hasil-hasil pertanian yang mereka tanam.* Kata-kata sebelum kata “adalah” yaitu *Ciri, utama, dari, sistem, tanam,* dan *paksa*. Kondisi ini memenuhi aturan ke 1 pada tabel 7 yaitu terdapat kata *dari* yang merupakan pendamping nomina yang mendampingi kata *sistem.* Penggabungan dua kata tersebut akan membentuk makna lokasi. Maka kalimat ini merupakan kalimat non-definisi.

2.3 Pembangkit Pertanyaan

Setelah tahapan *preprocessing* dan deteksi kalimat non-definisi selesai dilakukan, maka kalimat yang menjadi masukan sistem akan memiliki berbagai macam informasi penting yang akan digunakan dalam proses pembangkitan pertanyaan. Pertanyaan yang dibangkitkan adalah pertanyaan *factoid* dan *non-factoid.*

Pembangkit pertanyaan dibantu dengan *template* berdasarkan keberadaan kata kunci dan fungsi gramatikal yang didapatkan dari hasil analisis sintaksis. Pembangkitan pertanyaan *non-factoid* akan dilakukan dengan cara mengidentifikasi kata kunci yang terdapat pada kalimat masukan. Setelah kata kunci ditemukan, sistem akan memilih template pertanyaan yang sesuai dengan kategori kata kunci lalu membangkitkan pertanyaan dan jawaban menggunakan template yang telah dipilih tersebut.

2.3.1 Deteksi *Non-factoid*

Proses pelabelan pertanyaan *non-factoid* dilakukan dengan cara mendeteksi keberadaan kata kunci khusus *non-factoid*. Tabel 9 berikut adalah daftar kata kunci untuk pertanyaan *non-factoid*[15].

Tabel 6. Kata Kunci Non-factoid

Kategori non-factoid	Kata Kunci	
	Sebelum target	Sesudah target
Definisi	disebut, dikenal,	adalah, yaitu,

Kategori non-factoid	Kata Kunci	
	Sebelum target	Sesudah target
	dinamakan, mendefinisikan	ialah, merupakan, diartikan
Alasan	oleh sebab itu, jadi, memungkinkan adanya, dengan demikian, maka, dikatakan, penyebab terjadinya, sehingga, walau demikian, namun demikian	sebab, karena, bertujuan
Metode	berfungsi untuk, berguna untuk	dengan cara

2.3.2 Pembangkit pertanyaan *Non-factoid*

Klausa atau kalimat yang terdeteksi sebagai *non-factoid* melalui proses deteksi *non-factoid* akan membuat sistem membangkitkan pertanyaan menggunakan template *non-factoid.* Template *non-factoid* dapat dilihat pada Tabel 10 di bawah ini.

Tabel 7. Template Non-Factoid

Jenis	Template
Definisi	Apa yang dimaksud...?
Alasan	Mengapa ...?
Cara	Bagaimana cara. ..?

2.3.3 Pembangkit pertanyaan *factoid*

Untuk pembangkitan pertanyaan *factoid* akan dilakukan menggunakan bantuan *template* dan analisis sintaksis yang menghasilkan fungsi gramatikal.

Tabel 8. Template Factoid

Target	Template
Subjek	Siapa yang ...? Apa yang ....? Berapa ...?
Objek	Siapa yang ...? Apa yang ....? Berapa ...?
Keterangan	Di mana ...? Ke mana ...? Kapan ...?

3. PENUTUP

Pengembangan pembangkit pertanyaan otomatis untuk teks berbahasa Indonesia berhasil membangkitkan pertanyaan *factoid* dan pertanyaan *non-factoid*. Pengembangan sistem ini dapat menghasilkan pertanyaan dengan akurasi sebesar 68,84%. Dari hasil tersebut dapat disimpulkan bahwa proses deteksi kalimat non-definisi dapat mengatasi kesalahan yang disebabkan karena proses pembangkit pertayaan. Akan tetapi sistem ini masih memiliki tingkat kesalahan yang cukup tinggi sehingga mengurangi akurasi pertanyaan yang dihasilkan sistem.

Saran untuk penelitian berikutnya untuk meningkatkan akurasi sistem adalah Pertama, diperlukan penelitian lebih mendalam untuk proses pengenalan entitas bernama (NER), karena jika terdapat kesalahan pengenalan entitas bernama maka akan berimbas pada proses deteksi kalimat non-definisi karena tidak terdeteksinya NER dengan kategori PERSON, LOCATION, dll. Kesalahan pengenalan entitas bernama tersebut sering terjadi pada pelabelan NER dengan kategori PERSON dan non-PERSON sehingga menyebabkan kesalahan pembangkitan pertanyaan yang seharusnya menghasilkan pertanyaan ‘siapa’ menjadi ‘apa’ atau sebaliknya.

Kedua, untuk mengatasi pasangan pertanyan-jawaban tidak logis dan kesalahan konteks, pada penelitian selanjutnya bisa ditambahkan proses analisis semantik antar kalimat, sehingga jika terdapat makna yang berkaitan antara satu kalimat dengan kalimat yang lainnya sistem dapat membangkitkan pertanyaan-jawaban yang lebih akurat. Dan jika dilakukan penambahan *template* pertanyaan pada proses pembangkitan pertanyaan, analisis semantik dapat digunakan untuk mencegah kemunculan pertanyaan-pertanyaan yang memiliki konteks yang sama dengan pertanyaan yang lainnya.

DAFTAR PUSTAKA

[1] H. D. A. D. Ali, “Thesis: Automatic question generation : a syntactical approach to the sentence-to-question generation case,” 2012.

[2] M. Heilman, “Automatic Factual Question Generation from Text,” *Dissertation*, no. June, p. 203, 2011.

[3] K. Beulen and H. Ney, “Automatic question generation for decision tree based state tying,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2, no. November, pp. 805–808, 1998.

[4] M. Fachrurrozi, “Sistem Pembangkit Pertanyaan Otomatis Dengan Metode Template-Based,” 2013.

[5] W. Suwarningsih, I. Supriana, and A. Purwarianti, “Indonesian Medical Sentence Transformation for Question Generation,” pp. 68–71, 2015.

[6] F. Ferdian, “Implementation of Semantic Analyzer in Indonesian Text-Understanding Evaluation System,” 2012.

[7] M. R. Iqbal, “Pembangkit Pertanyaan Otomatis Untuk Teks Berbahasa Indonesia

Berdasarkan Template Sintaksis," UNIKOM, 2016.

[8] D. D. Ginanjar *et al.*, "Pembangkit Pertanyaan Otomatis Untuk Teks Berbahasa Indonesia Yang Mengandung Kalimat Majemuk."

[9] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition," *Speech Lang. Process. An Introd. to Nat. Lang. Process. Comput. Linguist. Speech Recognit.*, vol. 21, pp. 0–934, 2009.

[10] R. A. SUKAMTO, "Penguraian Bahasa Indonesia Dengan Rosa Ariani Sukamto Nim :," vol. 23507024, 2009.

[11] A. F. Wicaksono and A. Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia," *4th Int. MALINDO (Malay Indones. Lang. Work.*, no. October 2016, 2010.

[12] I. Budi, S. Bressan, G. Wahyudi, Z. A. Hasibuan, and B. A. A. Nazief, "Named Entity Recognition for the Indonesian language: Combining contextual, morphological and part-of-speech features into a knowledge engineering approach," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3735 LNAI, pp. 57–69, 2005.

[13] M. Fachri, "Named Entity Recognition For Indonesian Text Using Hidden Markov Model," Universitas Gadjah Mada, 2014.

[14] Y. Syaifudin, "Quotations Identification From Indonesian Online News Using Rule-Based Method," Universitas Gadjah Mada, 2016.

[15] a a Zulen and a Purwarianti, "Study and implementation of monolingual approach on indonesian question answering for factoid and non-factoid question," *PACLIC 25 - Proc. 25th Pacific Asia Conf. Lang. Inf. Comput.*, pp. 622–631, 2011.