

OTOMATISASI PEMBANGKITAN PERTANYAAN UNTUK BAHASA INDONESIA (SYSTEMATIC LITERATURE REVIEW)

Mohammad Farid Naufal¹, Selvia Ferdiana Kusuma^{*2}

¹Universitas Surabaya, Surabaya, ²Politeknik Elektronika Negeri Surabaya
Email: ¹faridnaufal@staff.ubaya.ac.id, ^{2*}selvia@pens.ac.id

*Penulis Korespondensi

(Naskah masuk: 02 Agustus 2022, diterima untuk diterbitkan: 28 Februari 2023)

Abstrak

Penelitian tentang otomatisasi pembangkitan pertanyaan terus berkembang. Berbagai metode telah coba diterapkan pada berbagai bahasa. Setiap bahasa memiliki karakteristik yang berbeda beda. Oleh karena itu, metode yang digunakan untuk membangkitkan pertanyaan juga harus disesuaikan dengan bahasa yang digunakan. Otomatisasi pembangkitan pertanyaan untuk bahasa Indonesia juga sudah mulai berkembang sejak 2015. Hasil penelitian-penelitian tersebut perlu dianalisis agar dapat mengetahui kelebihan maupun kekurangan dari setiap metode yang pernah digunakan. Oleh karena itu, jurnal ini membahas tentang Systematic Literature Review (SLR) pembangkitan pertanyaan pada bahasa Indonesia. SLR yang dibangun ini dapat digunakan untuk bahan pertimbangan optimalisasi penelitian tentang pembangkitan pertanyaan menggunakan bahasa Indonesia di kemudian hari. Tahapan yang dilakukan dalam pembentukan SLR adalah perencanaan *literature review*, kemudian melakukan *literature review* dan terakhir adalah pelaporan hasil *literature review*. Pencarian pada google scholar menghasilkan 27 penelitian yang relevan dengan kata kunci. Penerapan kriteria inklusi dan eksklusi menghasilkan 15 penelitian yang relevan. Kemudian proses *backward* dan *forward snowballing* yang dilakukan menghasilkan 2 penelitian tambahan. Total penelitian yang dianalisis berjumlah 17 penelitian. Proses selanjutnya adalah penilaian kualitas penelitian. Hasil penilaian kualitas penelitian menunjukkan bahwa keseluruhan penelitian yang berjumlah 17 penelitian tersebut memiliki kualitas yang baik untuk dianalisis. Hasil analisis yang dilakukan menunjukkan bahwa penelitian awal terkait pembangkitan pertanyaan untuk bahasa Indonesia masih memiliki beberapa celah. Diantaranya terkait dataset yang belum memadai, model pertanyaan yang kurang beragam, belum adanya penanganan/preproses model kalimat *unstructured*, dan belum adanya pembangkitan pertanyaan yang berasal dari gabungan beberapa informasi.

Kata kunci: bahasa Indonesia, otomatisasi pembangkitan pertanyaan, *systematic literature review*.

AUTOMATIC OF QUESTIONS GENERATION IN INDONESIAN (SYSTEMATIC LITERATURE REVIEW)

Abstract

Research on automated question generation is constantly evolving. Various methods have been tried to be applied in various languages. Each language has different characteristics. Therefore, the method used to generate questions should be adapted based on the language. The automation of question generation for Indonesian has also begun to develop since 2015. The result of these studies need to be analyzed to find out the advantages and disadvantages of each method that has been used. Therefore, this journal discusses the Systematic Literature Review (SLR) for generating questions in Indonesian. The SLR that was built can be used for consideration of optimizing research on generating questions using Indonesian in the future. The steps taken in this SLR are planning analysis, then carrying out the analysis and finally reporting the analysis. A search on Google Scholar yielded 27 studies that were relevant to the keyword. The application of inclusion and exclusion criteria resulted in 15 relevant studies. Then the backward and forward snowballing processes carried out resulted in 2 additional studies. Total research analyzed amounted to 17 studies. The next process is the assessment of research quality. The results of the research quality assessment showed that the overall 17 studies had good quality for analysis. The results of the analysis carried out indicate that the initial research related to question generation for Indonesian still has some gaps. For examples about datasets, question models, handling unstructured sentence models, and generating questions from a combination of some information.

Keywords: automated question generation, Indonesian, *systematic literature review*

1. PENDAHULUAN

Pembuatan atau pembangkitan pertanyaan secara manual membutuhkan waktu yang lama. Pembangkitan pertanyaan secara manual juga pekerjaan yang melelahkan bagi pembuat pertanyaan karena harus mempertimbangkan berbagai variasi pertanyaan dari sebuah materi atau informasi. Oleh karena itu *Automatic Question Generation* (AQG) atau otomatisasi pembangkitan pertanyaan mendapat perhatian besar dari para peneliti di bidang linguistik komputasi (Soni, Kumar, & Saha, 2019). Otomatisasi proses pembangkitan pertanyaan memiliki manfaat untuk mempercepat proses pembuatan pertanyaan. Hasil otomatisasi pembangkitan pertanyaan dapat digunakan untuk menunjang berbagai sistem. Sistem *Massive Online Open Courses* (MOOC) menggunakan hasil pembangkitan pertanyaan untuk menunjang proses evaluasi hasil pembelajaran. Selain itu, sistem bantuan otomatis, mesin pencari (*search engine*), sistem *chatbot* untuk interaksi pelanggan juga menggunakan hasil otomatisasi pembangkitan pertanyaan sebagai data latih pembangunan interaksi antara sistem dan pengguna. Otomatisasi proses pembangkitan pertanyaan saat ini semakin berkembang (Amidei, Piwek and Willis, 2018). Setiap bahasa memiliki karakteristik yang berbeda, oleh karena itu jika *basic* pembangkitan pertanyaan adalah *natural language processing*, maka perlu ada penyesuaian metode sesuai dengan bahasa yang digunakan.

Sampai saat ini sudah banyak metode yang digunakan untuk mengembangkan otomatisasi pembangkitan pertanyaan. Tidak semua metode sesuai jika diterapkan pada bahasa yang berbeda. Oleh karena itu perlu ada pembahasan khusus untuk mengetahui sejauh mana penelitian tentang otomatisasi pembangkitan pertanyaan ini sudah berkembang di suatu bahasa. Solusinya adalah pembuatan *Systematic Literature Review* (SLR) agar dapat mengetahui perkembangan otomatisasi pembangkitan pertanyaan untuk Bahasa Indonesia. Tujuan dari pembuatan SLR ini adalah mendapatkan informasi mengenai kelebihan dan kekurangan dari setiap metode yang digunakan pada penelitian sebelumnya. Selain itu, SLR ini juga dapat digunakan untuk mengetahui celah kekurangan dari penelitian sebelumnya, sehingga dapat dilakukan optimalisasi pada penelitian selanjutnya.

2. METODE PENELITIAN

Ada 3 tahapan yang dilakukan pada pembuatan SLR ini. Tahap pertama adalah perencanaan *literature review*, kemudian melakukan *literature review* dan terakhir adalah pelaporan hasil *literature review*. Detail dari setiap proses dijelaskan pada sub bab 3.1 sampai dengan sub bab 3.3. Metode yang digunakan pada pembuatan SLR ini serupa dengan penelitian (Raharjana, Siahaan, & Fatichah, 2021).

2.1. Perencanaan *Literature Review*

Tahapan perencanaan *literature review* berisi tentang penentuan tujuan dan *research question*. Kemudian penentuan strategi pencarian, kriteria inklusi dan eksklusi kemudian terakhir proses *backward and forward snowballing* untuk memastikan bahwa semua penelitian serupa masuk pada list *literature review*. Detail setiap tahapan dijelaskan pada sub bab A sampai dengan D.

A. Penentuan tujuan dan *research question*

Penelitian terkait pembangkitan pertanyaan semakin marak. Setiap bahasa memiliki karakteristik yang berbeda-beda. Oleh karena itu metode yang digunakan juga disesuaikan dengan bahasa yang digunakan untuk membangkitkan pertanyaan. Perkembangan penelitian di bidang pembangkitan pertanyaan khususnya pada bahasa Indonesia perlu diketahui. Oleh karena itu diperlukan SLR yang membahas tentang pembangkitan pertanyaan khusus untuk Bahasa Indonesia. *Research question* yang dibentuk untuk memenuhi tujuan dari SLR ini ada tiga, ditunjukkan pada *Research Question 1* (RQ1) sampai dengan *Research Question 3* (RQ3).

RQ1: Bagaimana perkembangan pembangkitan pertanyaan untuk bahasa Indonesia?

RQ2: Apa saja metode yang sudah digunakan untuk membangkitkan pertanyaan?

RQ3: Apa tantangan otomatisasi pembangkitan pertanyaan pada Bahasa Indonesia?

B. Strategi Pencarian

Pencarian penelitian tentang pembangkitan pertanyaan dilakukan melalui google scholar. Kata kunci yang digunakan adalah “pembangkitan pertanyaan”, “pembangkitan soal” dan “*question generation*”. Ketiga kata kunci tersebut digunakan bergantian untuk memaksimalkan pencarian penelitian terkait pembangkitan pertanyaan.

C. Kriteria inklusi dan eksklusi

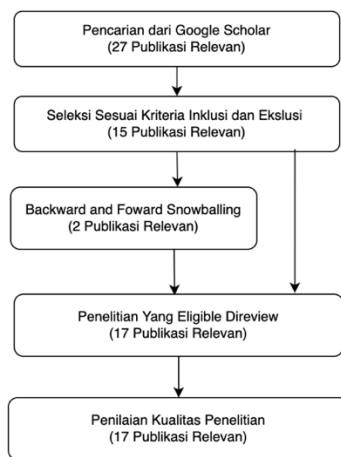
Kriteria inklusi adalah kriteria yang digunakan sebagai syarat penelitian-penelitian yang dianalisis. Kriteria inklusi pada SLR ini adalah merupakan penelitian tentang pembangkitan pertanyaan dan menggunakan bahasa Indonesia. Penelitian diterbitkan pada rentang waktu antara 2015 sampai dengan 2021. Sedangkan kriteria eksklusi yang digunakan adalah penelitian tentang *question answering*, berbahasa selain bahasa Indonesia dan bukan penelitian yang bersifat tertutup / tidak dapat diunduh.

D. *Backward & forward snowballing*

SLR ini juga menggunakan metode *backward and forward snowballing* untuk memaksimalkan pencarian penelitian menggunakan referensi yang pernah digunakan pada penelitian serupa. Cara ini juga pernah digunakan pada jurnal SLR milik (Raharjana, Siahaan, & Fatichah, 2021)

2.2. Proses literature review

Hasil pencarian data menghasilkan sebanyak 27 dokumen yang sesuai dengan kata kunci. Dokumen berasal dari buku laporan, jurnal dan konferensi. Namun ada 10 data berbentuk buku laporan skripsi dan tidak dapat diunduh. Oleh karena itu kesepuluh data tidak ditindaklanjuti. Setelah dilakukan seleksi menggunakan kriteria inklusi dan eksklusi ternyata ada 2 penelitian yang tidak sesuai. Sehingga pada akhirnya ada 15 penelitian yang sesuai dengan kriteria inklusi dan eksklusi. Selanjutnya, dilakukan proses *backward and forward snowball*. Proses *backward and forward snowball* menghasilkan 2 tambahan penelitian. Sehingga penelitian yang dianalisis berjumlah 17. Ada 6 konferensi dan 11 jurnal. Ilustrasi proses pencarian penelitian ditunjukkan pada Gambar 1.



Gambar 1. Ilustrasi Proses Pencarian Jurnal Relevan

Sebelum masuk pada proses *review*, semua penelitian akan dinilai kualitasnya. Hanya penelitian yang memiliki kualitas yang sesuai yang masuk pada proses analisis lanjutan. Metode analisis kualitas penelitian ini di adopsi dari penelitian (Raharjana, Siahaan, & Fatichah, 2021). Kriteria penilaian kualitas ada 5 seperti ditunjukkan pada Tabel 1. Penilaian memiliki 3 rentang nilai yaitu -1 untuk jawaban “tidak”, 0 untuk jawaban “cukup/ sebagian” dan 1 untuk jawaban “ya”.

Tabel 1. Kriteria penialain kualitas penelitian	
No	Kriteria Penilaian
1	Apakah tujuan penelitian jelas
2	Apakah solusi dijelaskan secara detail?
3	Apakah solusi yang ditawarkan divalidasi?
4	Apakah penelitian itu menunjukkan opini penulis (penilaian terbalik)
5	Apakah penelitian tersebut disitasi penelitian lain?

Berdasarkan hasil penilaian diketahui bahwa penilaian untuk kriteria pertama mencapai 100% penelitian telah menunjukan tujuan yang jelas. Penilaian kriteria kedua menunjukan 82% penelitian menjelaskan solusi secara detail. Nilai dari penilaian kriteria ketiga yaitu 88% penelitian telah melakukan

evaluasi dan validasi dari metode yang diusulkan. Kemudian nilai dari kriteria keempat adalah 0% karena tidak ada penelitian dijelaskan secara subjektif, semua penelitian dijelaskan secara objektif dengan pembuktian melalui hasil eksperimen. Hasil penilaian kriteria kelima menunjukkan bahwa 76% penelitian disitasi oleh penelitian selanjutnya. Hasil analisis kualitas penelitian menunjukkan bahwa semua penelitian yang relevan memiliki kualitas yang baik dan layak untuk masuk proses analisis lanjutan.

Komponen yang digunakan untuk menganalisis setiap penelitian ditunjukkan pada Tabel 2. Ada 11 komponen yang digunakan untuk mengekstrak informasi dari setiap penelitian. Setiap informasi sudah disesuaikan dengan RQ yang dibentuk.

Tabel 2. Komponen analisis penelitian

No	Komponen yang dianalisis	RQ yang relevan
1	Judul	Gambaran awal
2	Author	Gambaran awal
3	Tahun	Gambaran awal
4	Tipe artikel	Gambaran awal
5	Tujuan	RQ1
6	Data dan Domain	RQ2
7	Metode	RQ2
8	Validasi	RQ2
9	Tools	RQ2
10	Kekurangan dan tantangan	RQ3
11	Penelitian selanjutnya	RQ1

2.3. Pelaporan hasil literature review

Hasil review dilaporkan dengan menjelaskan ringkasan studi dan menjawab setiap RQ. Deskripsi setiap RQ didasarkan pada hasil ekstraksi data.

3. PENELITIAN TERDAHULU

Penelitian terkait *Literature Review* (LR) untuk pembangkitan pertanyaan ini sudah pernah dilakukan oleh beberapa peneliti terdahulu. Namun belum ada yang membahas spesifik pada pembangkitan pertanyaan untuk bahasa Indonesia.

LR yang dibuat oleh (Soni, Kumar, & Saha, 2019) tidak berbentuk SLR namun melakukan *comparative study* terhadap penelitian yang ada, mulai dari prosedur yang digunakan dan jenis pertanyaan yang dibangkitkan. Ada 14 jurnal yang dibandingkan pada LR ini. Seluruh jurnal yang dibahas pada LR ini menggunakan bahasa Inggris. LR ini membandingkan metode-metode yang digunakan disetiap jurnal.

LR yang dibuat oleh (Gumaste et al., 2019) menjelaskan secara singkat metode-metode yang pernah digunakan dalam proses pembangkitan pertanyaan. Penelitian ini tidak membahas tentang perbandingan metode-metode yang pernah digunakan. Penelitian ini juga tidak secara spesifik menjelaskan bagaimana cara mendapatkan datanya dan melakukan proses studi literaturnya.

LR yang dibuat oleh (CH and Saha, 2020) menjelaskan secara detail cara mendapatkan jurnal, dan data yang digunakan untuk melakukan *review*. Namun fokusnya hanya pembangkitan pertanyaan

untuk pilihan ganda. Jumlah total jurnal yang dianalisis setelah melakukan proses *include* dan *exclude* sejumlah 86 jurnal. SLR ini menjelaskan metode metode yang digunakan di tiap jurnal. SLR ini juga memberikan review atau peluang yang masih bisa diperbaiki dari setiap tahapan. Namun belum ada proses evaluasi terhadap hasil pembuatan SLRnya.

LR yang dibuat oleh (Kurdi et al., 2020) berbentuk SLR yang berfokus untuk pendidikan. Ada 93 jurnal yang direview. Tujuan dari SLR ini adalah memberikan gambaran umum tentang komunitas *Automatic Question Generation* (AQG) dan aktivitasnya, meringkas tren dan kemajuan saat ini di bidang AQG, menyoroti perubahan yang telah dialami area tersebut dalam beberapa tahun terakhir, dan menyarankan area untuk perbaikan dan peluang masa depan untuk AQG khususnya di bidang Pendidikan.

4. HASIL DAN PEMBAHASAN

Informasi tentang perkembangan penelitian terkait pembangkitan pertanyaan menggunakan bahasa Indonesia didapatkan berdasarkan RQ yang telah dibuat. Temuan yang didapatkan dari setiap RQ dijelaskan sebagai berikut.

RQ1: Bagaimana perkembangan pembangkitan pertanyaan untuk Bahasa Indonesia?

Berdasarkan studi literatur yang telah dilakukan diketahui bahwa pembangkitan pertanyaan berbahasa Indonesia sudah banyak dikembangkan dan dipublikasikan dalam bentuk jurnal maupun konferensi internasional. Total ada 11 jurnal dan 6 konferensi internasional. Detailnya ditunjukkan pada Tabel 3.

Tabel 3. Jenis Publikasi Topik Penelitian tentang Pembangkitan Pertanyaan

Jenis Publikasi	Jumlah	Penelitian
Jurnal	11	(Saefurochman, 2017; Suwarningsih, Supriana, & Purwarianti, 2015b; Ginanjar and Purnamasari, 2017; Aminudin et al., 2018; Saripudin and Purnamasari 2017; Basuki and Kusuma, 2018; Kusuma and Alhamri, 2018; Wijanarko et al., 2020; Sumantri et al., 2020; Wijanarko et al., 2021; Flores, Jasa, & Hartati, 2021)
Conference	6	(Suwarningsih, Supriana, & Purwarianti, 2015a; Kusuma et al., 2018, 2019; Kusuma, Siahaan, & Faticah, 2020; Muis and Purwarianti, 2020; Wisnu Prabowo, Budi, & Budi Santoso, 2021)

Berdasarkan Tabel 4 diketahui bahwa otomatisasi pembangkitan pertanyaan ini telah diuji coba untuk berbagai domain. Domain pendidikan paling dominan karena memang proses otomatisasi pembangkitan pertanyaan dapat digunakan untuk menunjang proses pembangkitan soal evaluasi pembelajaran. Materi pendidikan khususnya ilmu

pengetahuan alam atau biologi mendominasi domain yang digunakan untuk pengujian proses pembangkitan pertanyaan. Namun sejauh ini belum ada data yang benar-benar terbuka dan dapat dijadikan acuan uji coba berbagai metode pembangkitan pertanyaan. Setiap penelitian masih menggunakan data yang berbeda-beda. Sehingga metode yang diterapkan belum benar-benar diujicobakan dengan berbagai data.

Tabel 4. Domain pada Pembangkitan Pertanyaan

Domain	Jumlah	Penelitian
Kesehatan	2	(Suwarningsih, Supriana, & Purwarianti, 2015a, 2015b)
Ilmu Pengetahuan Alam	4	(Saefurochman, 2017; Kusuma et al., 2019; Kusuma, Siahaan, & Faticah, 2020; Flores, Jasa, & Hartati, 2021)
Biologi	3	(Aminudin et al., 2018; Basuki and Kusuma, 2018; Wisnu Prabowo, Budi, & Budi Santoso, 2021)
Software engineering	4	(Kusuma and Alhamri, 2018; Kusuma et al., 2018; Wijanarko et al., 2020; 2021)
Cerita Anak	1	(Sumantri et al., 2020)
SQUAD + TyDiQA	1	(Muis and Purwarianti, 2020)
Open Domain	1	(Basuki and Kusuma, 2018)
Tidak dijelaskan	2	(Ginanjar and Purnamasari, 2017; Saripudin and Purnamasari, 2017)

Berdasarkan data pada Tabel 5 diketahui jenis pertanyaan yang pernah dibangkitkan gabungan dari jenis *factoid* dan *non factoid*. Belum ada yang mencoba membangkitkan pertanyaan untuk jenis pilihan ganda. Perlu *effort* yang lebih ketika membangkitkan pertanyaan berbentuk pilihan ganda. Pilihan jawaban juga mempengaruhi tingkat kesulitan soal. Semakin mirip jawaban pada pilihan ganda maka semakin sulit tingkat kesulitan soalnya. Begitu juga sebaliknya.

Tabel 6 menunjukkan jumlah data, hasil pertanyaan dan akurasi yang dihasilkan oleh setiap penelitian. Namun tidak semua penelitian menjelaskan dengan detail berapa data yang digunakan dan berapa jumlah pertanyaan yang dihasilkan. NA pada Tabel 6 artinya adalah *Not Available*. NA digunakan ketika penelitian tersebut tidak menjelaskan berapa hasil pertanyaan yang dihasilkan.

Tabel 5. Jenis Pertanyaan pada Pembangkitan Pertanyaan

Jenis Pertanyaan	Jumlah	Penelitian
<i>Factoid</i>	3	(Sumantri et al., 2020; Muis & Purwarianti, 2020; Wisnu Prabowo, Budi, and Budi Santoso, 2021)
<i>Non Factoid</i>	4	(Suwarningsih, Supriana, & Purwarianti, 2015a, 2015b; Saefurochman, 2017; Aminudin et al., 2018)
Gabungan	10	(Saripudin and Purnamasari, 2017; Ginanjar and Purnamasari, 2017; Basuki & Kusuma 2018; Kusuma and Alhamri, 2018; Wijanarko et al. 2020, 2021; Kusuma et al., 2018, 2019; Kusuma, Siahaan, & Fatichah, 2020; Flores, Jasa, & Hartati, 2021)

Jenis data yang digunakan pada setiap penelitian cukup beragam. Ada yang menggunakan data berupa materi, teks/paragraf dan kalimat. Hal tersebut membuat tidak dapat dilihatnya efektifitas metode yang diusulkan. Suatu metode dikatakan efektif kalau dengan data yang sedikit dapat membangkitkan pertanyaan yang lebih banyak.

Akurasi dapat menggambarkan keberhasilan metode namun tidak menggambarkan keefektifan metode yang diusulkan untuk membangkitkan pertanyaan. Detail dari jumlah data, jumlah pertanyaan yang dihasilkan, dan akurasi yang dihasilkan oleh setiap penelitian di tunjukkan pada Tabel 6.

Semakin lama, data yang digunakan semakin banyak. Tapi rata-rata masih sekitar 1000 kalimat. Ada yang mencapai puluhan ribu kalimat namun itu hasil penerjemahan dataset dari bahasa lain. Kelemahannya adalah hasil kalimat yang digunakan tidak natural. Padahal sebenarnya struktur kalimat yang digunakan untuk teks berbahasa Indonesia terkadang sangat komplek. Akurasi yang dihasilkan mayoritas sudah cukup tinggi, namun sayangnya akurasi hanya dilakukan untuk data yang sama. Proses uji coba tidak dilakukan dengan data-data dari penelitian lain.

RQ2: Apa saja metode yang sudah digunakan untuk membangkitkan pertanyaan?

Tabel 7 menunjukkan metode-metode yang pernah digunakan dalam proses pembangkitan pertanyaan untuk masing-masing penelitian. Metode yang sering digunakan masih berbasis templat. *Named Entity Recognize* (NER) juga dapat digunakan untuk mencari kata kunci pembangkitan pertanyaan. Penelitian (Wisnu Prabowo, Budi and Budi Santoso, 2021) dan (Flores, Jasa and Hartati, 2021) menggunakan NER untuk mengenali adanya *noun*, *adverb*, *person*, dan *place*. Informasi mengandung "*noun*" akan diubah menjadi kalimat tanya yang diawali dengan kata tanya "apa". Informasi mengandung "*adverb*" akan diubah menjadi kalimat

tanya yang diawali dengan kata tanya "kapan". Informasi mengandung "*person*" akan diubah menjadi kalimat tanya yang diawali dengan kata tanya "siapa". Informasi mengandung "*number*" akan diubah menjadi kalimat tanya yang diawali dengan kata tanya "berapa" dan informasi mengandung "*place*" akan diubah menjadi kalimat tanya yang diawali dengan kata tanya "dimana".

Tabel 6. Hasil Akurasi Pembangkitan Soal

Penelitian	Jumlah data	Pertanyaan	Akurasi
(Saefurochman, 2017)	30 materi	343	70%
(Suwarningsih, Supriana, & Purwarianti, 2015b)	100 kalimat	NA	Tidak Ada
(Ginanjar and Purnamasari, 2017)	35 teks	NA	71%
(Aminudin, SN and Ahmad, 2018)	510 kalimat	NA	91%
(Saripudin and Purnamasari, 2017)	50 teks	413	90%
(Basuki and Kusuma, 2018)	978 kalimat	386	88,6%
(Kusuma and Alhamri, 2018)	278 kalimat	654	82%
(Wijanarko et al., 2020)	1.432 kalimat	60.000	95%
(Sumantri et al., 2020)	1 cerita	17	76%
(Wijanarko et al., 2021)	1.347 kalimat	92.608	Tidak ada
(Flores, Jasa & Hartati, 2021)	18 materi	748	89%
(Suwarningsih, Supriana, & Purwarianti, 2015a)	200 kalimat	NA	Tidak ada
(Kusuma et al., 2018)	214 kalimat	563	81%
(Kusuma et al., 2019)	50 kompetensi	412	Tidak ada
(Kusuma, Siahaan, & Fatichah, 2020)	2 materi	1.711	86%
(Muis and Purwarianti, 2020)	536 artikel	161.550	Tidak ada
(Wisnu Prabowo, Budi, & Budi Santoso, 2021)	1.027 kalimat	5.149	

Konsep pembangkitan pertanyaan berdasarkan pola kata juga dilakukan pada penelitian (Wijanarko et al., 2020). Penelitian (Wijanarko et al., 2020) mengenali sebuah objek pertanyaan berdasarkan *noun* pada kalimat tersebut. Proses pengenalan *noun* dilakukan menggunakan metode parse tree berbasis *Context Free Grammar* (CFG). Namun metode ini juga belum dapat membangkitkan pertanyaan dari relasi informasi. Pertanyaan yang dibangkitkan hanya berdasarkan informasi dari satu kalimat. Metode yang terbaru adalah penggunaan deep learning (Muis and Purwarianti, 2020) dan ontologi (Kusuma, Siahaan, & Fatichah 2020). Pembangkitan pertanyaan

menggunakan deep learning masih terkendala dengan dataset. Penggunaan hasil penerjemahan dari bahasa Inggris ke bahasa Indonesia membuat beberapa kalimatnya tidak natural.

Tabel 7. Metode pada Pembangkitan Pertanyaan

No	Metode	Jumlah	Penelitian
1	Template + Rule	12	(Suwarningsih, Supriana, & Purwarianti, 2015a, 2015b; Saefurochman et al, 2017; Saripudin and Purnamasari, 2017; Ginanjar and Purnamasari 2017; Basuki and Kusuma, 2018; Aminudin et al. 2018; Kusuma and Alhamri, 2018; Kusuma et al., 2018; Sumantri et al., 2020; Flores, Jasa, & Hartati, 2021)
2	Pattern Keyphare	1	(Wijanarko et al., 2020)
3	Template + Standard Language Regulation	1	(Wisnu Prabowo, Budi, & Budi Santoso 2021)
4	NER	4	(Suwarningsih, Supriana, & Purwarianti, 2015b; Saripudin and Purnamasari, 2017; Ginanjar and Purnamasari, 2017; Wisnu Prabowo, Budi, & Budi Santoso, 2021)
5	Deteksi majemuk	1	(Saripudin and Purnamasari, 2017)
6	Deep Learning	1	(Muis and Purwarianti, 2020)
7	Mind map + template	1	(Kusuma et al., 2019)
8	Ontology + template	1	(Kusuma, Siahaan, & Fatichah, 2020)

Metode-metode yang ada dapat dikembangkan untuk membuat dataset. Kemudian data tersebut dapat digunakan untuk *training* model pada metode deep learning. Metode baru lain yang diterapkan untuk membentuk pertanyaan untuk bahasa Indonesia adalah ontologi. Konsep pembangkitan pertanyaan menggunakan ontologi ini memiliki kelebihan, yaitu dapat membangkitkan pertanyaan yang secara implisit. Ontologi dapat menghasilkan pertanyaan yang tidak dibentuk dari kalimat-kalimat yang sudah ada. Ontologi memungkinkan ada informasi tambahan yang tidak ada dari struktur teks. Namun pengubahan informasi ke bentuk ontologi sendiri memerlukan usaha yang tidak mudah. Tidak semua

orang dapat melakukan hal tersebut. Oleh karena itu, perlu model ontologi baru yang dapat digunakan untuk menyimpan informasi dengan lebih mudah.

Mayoritas metode bersanding dengan basis templat. Templat yang digunakan haruslah bersifat open domain, sehingga templat tersebut dapat digunakan untuk membangkitkan pertanyaan dari berbagai domain. Jika menggunakan templat, maka templat perlu terus di *upgrade* agar sesuai dengan semua kondisi.

Setiap penelitian harus dievaluasi untuk membuktikan bahwa metode yang diusulkan benar benar dapat menyelesaikan masalah yang diangkat. Pada penelitian tentang pembangkitan pertanyaan ini, mayoritas evaluasinya menggunakan penilaian dari ahli. Namun penilaian ahli sendiri sebenarnya tidak bisa dijadikan standart karena setiap ahli memiliki tingkat pemahaman yang berbeda. Perlu ada tambahan metode penentu kesamaan jawaban dari ahli, misalkan dengan metode Gwet AC1 (Kusuma, Siahaan and Fatichah, 2022) atau Kappa (Wang et al., 2019). Tabel 8 menunjukkan metode evaluasi yang digunakan oleh setiap penelitian.

Tabel 8. Evaluasi pada Pembangkitan Pertanyaan

No	Evaluasi	Jumlah	Penelitian
1	Ahli	12	(Suwarningsih, Supriana, & Purwarianti, 2015a; Saefurochman 2017; Saripudin and Purnamasari, 2017; Ginanjar and Purnamasari, 2017; Aminudin et al., 2018; Basuki and Kusuma, 2018; Kusuma and Alhamri, 2018; Kusuma et al., 2019; Wijanarko et al., 2020; Sumantri et al., 2020; Wijanarko et al., 2021; Flores, Jasa, & Hartati, 2021)
2	Matrix	3	(Muis and Purwarianti, 2020; Wijanarko et al., 2020; 2021)
3	Tanpa Evaluasi	2	(Suwarningsih, Supriana, and Purwarianti, 2015b; Wisnu Prabowo, Budi, & Budi Santoso, 2021)

RQ3: Apa tantangan otomatisasi pembangkitan pertanyaan pada Bahasa Indonesia?

Ada 3 tantangan yang dapat dijadikan fokus pada pembangkitan pertanyaan. Pertama yaitu tidak ada dataset yang memadai. Sehingga tidak ada *ground truth*. Evaluasi dari banyak ahli dengan berbagai latar belakang pendidikan tentunya juga mempengaruhi hasil validasi. Jika ada dataset yang banyak, tentunya metode *deep learning* dapat digunakan. Seharusnya ada penelitian yang mengembangkan dataset ini.

Kedua yaitu struktur data. Data yang digunakan dalam pemrosesan bahasa alamiah adalah rangkaian kalimat. Terdapat berbagai karakter kalimat deklaratif dalam dokumen teks, misalnya: adanya konten co-referensi dan penggunaan kata ganti. Rangkaian kalimat pada bahasa Indonesia cenderung kompleks dan terkadang tidak terstruktur sehingga terkadang metode berbasis aturan gagal mengekstrak komponen kalimat. Oleh karena itu, rangkaian kalimat tersebut perlu dipraproses. Otomatisasi praproses untuk simplifikasi kalimat pada bahasa Indonesia ini masih hanya pemisahan kalimat majemuk. Praproses untuk kalimat kompleks yang lainnya perlu dilakukan otomatisasi. Kalimat pada bahasa Indonesia sendiri juga sering menggunakan pronomina (kata ganti). Komputer akan kesulitan memahami rangkaian kalimat jika menggunakan pronomina. Padahal untuk membangkitkan soal yang bersifat HOTS mestinya informasi tersebut saling berkaitan.

Ketiga yaitu jenis pertanyaan. Pembangkitan pertanyaan dapat dibentuk dari rangkuman paragraf. Namun tentunya metode peringkasan juga harus tepat agar tidak mempengaruhi hasil pembangkitan pertanyaan (Saefurochman, 2017). Sampai saat ini pertanyaan yang dibangkitkan adalah pertanyaan berjenis *factoid* dan *non factoid*. Belum pernah ada penelitian menggunakan bahasa Indonesia yang membuat pertanyaan bertipe pilihan ganda. Selain itu, juga belum pernah ada yang mencoba pembangkitan pertanyaan untuk matematika, berbentuk tabel maupun gambar.

5. KESIMPULAN

Penelitian tentang otomatisasi pembangkitan pertanyaan terus berkembang. Metode yang digunakan juga sangat beragam. Perlu adanya SLR untuk mengetahui perkembangan penelitian-penelitian di bidang pembangkitan pertanyaan. Oleh sebab itu, penelitian ini berfokus pada pembuatan SLR di bidang pembangkitan pertanyaan khususnya pada bahasa Indonesia. Terdapat 17 penelitian yang sesuai dengan kriteria inklusi dan eksklusi serta memiliki kualitas yang baik untuk dianalisis. Hasil analisis yang dilakukan menunjukkan bahwa penelitian awal terkait pembangkitan pertanyaan untuk bahasa Indonesia masih memiliki beberapa celah. Diantaranya terkait dataset yang belum memadai, model pertanyaan yang kurang beragam, belum adanya penanganan/preproses model kalimat *unstructured*, dan belum adanya pembangkitan pertanyaan yang berasal dari gabungan beberapa informasi.

DAFTAR PUSTAKA

AMIDEI, J., PIWEK, P. & WILLIS, A., 2018. Evaluation methodologies in automatic question generation 2013-2018. *INLG 2018 - 11th International Natural Language*

Generation Conference, Proceedings of the Conference, pp.307–317. <https://doi.org/10.18653/v1/w18-6537>.

AMINUDIN, A., SN, A. & AHMAD, B., 2018. Automatic Question Generation (AQG) Dari Dokumen Teks Bahasa Indonesia Berdasarkan Non-Factoid Quesion. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(2), pp.2–8. <https://doi.org/10.25126/jtiik.201852664>.

BASUKI, S. AND KUSUMA, S.F., 2018. Automatic Question Generation For 5w-1h Open Domain Of Indonesian Questions By Using Syntactical Template-Based Features From Academic Textbooks. *Journal of Theoretical and Applied Information Technology*, 96(12), pp.3908–3923.

CH, D.R. & SAHA, S.K., 2020. Automatic Multiple Choice Question Generation From Text : A Survey. In: *14 IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. IEEE*. pp.14–25.

FLORES, V.A., JASA, L. & HARTATI, R.S., 2021. Pembangkit Pertanyaan Otomatis pada Materi Pelajaran Ilmu Pengetahuan Alam Berbahasa Indonesia di Tingkat Sekolah Dasar Berdasarkan Revisi Taksonomi Bloom. *Majalah Ilmiah Teknologi Elektro*, 20(2), p.343. <https://doi.org/10.24843/mite.2021.v20i02.p19>.

GINANJAR, D.D. & PURNAMASARI, K.K., 2017. Pembangkit Pertanyaan Otomatis Untuk Teks Berbahasa Indonesia Yang Mengandung Kalimat Majemuk. *Elib Unikom*, pp.1–7.

GUMASTE, P.S., JOSHI, S.S., KHADPEKAR, S.A. & MALI, S.R., 2019. Automated Question Generator System: a Review. *International Journal of Engineering Applied Sciences and Technology*, 04(08), pp.171–176. <https://doi.org/10.33564/ijeast.2019.v04i08.027>.

KURDI, G., LEO, J., PARSIA, B., SATTLER, U. & AL-EMARI, S., 2020. A Systematic Review of Automatic Question Generation for Educational Purposes. *International Journal of Artificial Intelligence in Education*, 30(1), pp.121–204. <https://doi.org/10.1007/s40593-019-00186-y>.

KUSUMA, S.F. & ALHAMRI, R.Z., 2018. Generating Indonesian Question Automatically Based on Bloom ' s Taxonomy Using Template Based Method. *KINETIK: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 3(2), pp.145–152.

- KUSUMA, S.F., ALHAMRI, R.Z., SIAHAAN, D.O., FATICHAH, C. & NAUFAL, M.F., 2018. Indonesian Question Generation Based on Bloom ' s Taxonomy Using Text Analysis. *2018 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, (August), pp.269–274. <https://doi.org/10.1109/ISITIA.2018.8711015>.
- KUSUMA, S.F., SIAHAAN, D.O. & FATICHAH, C., 2020. Automatic Question Generation in Education Domain Based on Ontology. In: *CENIM 2020 - Proceeding: International Conference on Computer Engineering, Network, and Intelligent Multimedia 2020*. pp.251–256. <https://doi.org/10.1109/CENIM51130.2020.9297991>.
- KUSUMA, S.F., SIAHAAN, D.O. & FATICHAH, C., 2022. Automatic question generation with various difficulty levels based on knowledge ontology using a query template. *Knowledge-Based Systems*, 249, p.108906. <https://doi.org/10.1016/j.knosys.2022.108906>.
- KUSUMA, S.F., SIAHAAN, D.O., FATICHAH, C. & NAUFAL, M.F., 2019. Automatic Question Generation with Classification Based on Mind Map. *JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence*, pp.336–341. <https://doi.org/10.1109/JCSSE.2019.8864208>.
- MUIS, F.J. & PURWARIANTI, A., 2020. Sequence-to-Sequence Learning for Indonesian Automatic Question Generator. *2020 7th International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2020*. <https://doi.org/10.1109/ICAICTA49861.2020.9429032>.
- RAHARJANA, I.K., SIAHAAN, D. & FATICHAH, C., 2021. User Stories and Natural Language Processing: A Systematic Literature Review. *IEEE Access*, 9, pp.53811–53826. <https://doi.org/10.1109/ACCESS.2021.3070606>.
- SAEFUROCHMAN, S., 2017. Pembangkit Pertanyaan Berdasarkan Template Dari Dokumen Ringkasan Menggunakan Metode Lexical Chain Dan Korelasi Antar Kalimat.
- SARIPUDIN, P. & PURNAMASARI, K.K., 2017. Pendeteksian Keterkaitan Antar Kalimat Dengan Metode Template - Based Dalam Pembangkit Pertanyaan Otomatis. pp.1–7.
- SONI, S., KUMAR, P. & SAHA, A., 2019. Automatic Question Generation: A Systematic Review. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3403926>.
- SUMANTRI, WAHAB, S.A., SETYANINGSIH, E.R. & SUTIKSNO, H., 2020. Question Generator Menggunakan Pendekatan Rule Based Heuristik Pada Cerita Anak-Anak. *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, 5(1), pp.1–6.
- SUWARNINGSIH, W., SUPRIANA, I. & PURWARIANTI, A., 2015a. Indonesian medical sentence transformation for question generation. In: *IEEE Asia Pacific Conference on Wireless and Mobile*. pp.68–71. <https://doi.org/10.1109/APWiMob.2015.7374929>.
- SUWARNINGSIH, W., SUPRIANA, I. & PURWARIANTI, A., 2015b. Towards a framework for an indonesian medical question generator. *Telkomnika (Telecommunication Computing Electronics and Control)*, 13(1), pp.357–363. <https://doi.org/10.12928/TELKOMNIKA.v13i1.648>.
- WANG, W., FENG, S., WANG, D. & ZHANG, Y., 2019. Answer-guided and semantic coherent question generation in open-domain conversation. In: *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*. pp.5066–5076. <https://doi.org/10.18653/v1/d19-1511>.
- WIJANARKO, B.D., HERYADI, Y., TOBA, H. & BUDIHARTO, W., 2020. Automated question generating method based on derived keyphrase structures from bloom's taxonomy. *ICIC Express Letters*, 14(11), pp.1059–1067. <https://doi.org/10.24507/icicel.14.11.1059>.
- WIJANARKO, B.D., HERYADI, Y., TOBA, H. & BUDIHARTO, W., 2021. Question generation model based on key-phrase, context-free grammar, and Bloom's taxonomy. *Education and Information Technologies*, 26(2), pp.2207–2223. <https://doi.org/10.1007/s10639-020-10356-4>.
- WISNU PRABOWO, M., BUDI, I. & BUDI SANTOSO, H., 2021. Developing Question Generation System for Bahasa Indonesia Using Indonesian Standard Language Regulation. *ACM International Conference Proceeding Series*, pp.258–261. <https://doi.org/10.1145/3457784.3457824>.