

Optimasi Ekstraksi Kata Kunci untuk *Automatic Question Generation* pada Dokumen Ajar Elektronik Menggunakan Kombinasi *TF-IDF* dan *TextRank*

Jonathan Wira Wijaya^{1*}, Eddy Muntina Dharma¹, Anak Agung Gede Adi Mega Putra¹

¹ Program Studi Teknik Informatika, Primakara University, Indonesia

* Corresponding author: jtim.sekawan@gmail.com

Sitasi: J. W. Wijaya, E. M. Dharma, and A. A. G. A. M. Putra, "Optimasi Ekstraksi Kata Kunci untuk *Automatic Question Generation* pada Dokumen Ajar Elektronik Menggunakan Kombinasi *TF-IDF* dan *TextRank*," *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 4, pp. 1044–1055, 2026.

Diterima : 22-08-2026
Direvisi : 26-09-2026
Disetujui : 10-10-2026
Diterbitkan : 19-10-2026

Copyright: © 2026 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-Share Alike 4.0 International License.

Abstrak. Keberadaan instrumen evaluasi yang praktis merupakan pilar penting dalam ekosistem pembelajaran digital. Teknologi *Automatic Question Generation* (AQG) dirancang untuk mereduksi beban kognitif pendidik, namun akurasi ekstraksi kata kunci yang rendah sering kali mendegradasi kualitas soal yang dihasilkan. Pendekatan frekuensi tunggal rentan menyerap derau, sementara algoritma graf mengabaikan kelangkaan kata. Studi ini merumuskan teknik optimasi hibrida dengan memadukan skor *Term Frequency-Inverse Document Frequency* (TF-IDF) dan sentralitas *TextRank* guna menyeleksi kata kunci pada literatur *Internet of Things* (IoT). Pemrosesan dijalankan menggunakan kerangka *unsupervised learning* pada level *unigram*, dengan evaluasi diukur berdasarkan acuan pakar (*Ground Truth*) dan kelayakan tata bahasa soal rumpang. Kalkulasi komputasional pada batas 30 kata teratas membuktikan bahwa metode hibrida meraih F1-Score 40,68%, setara dengan TF-IDF murni dan melampaui *TextRank* (37,29%). Pada tahap validasi pedagogis, model usulan ini mendemonstrasikan keunggulan absolut dengan tingkat kelayakan 86,67% (26 kuis valid), menundukkan TF-IDF yang tertahan di angka 83,33%. Intervensi topologi graf sukses menangkal intrusi kata tak bermakna dan memelihara keutuhan sintaksis, menjadikan formulasi hibrida ini sangat direkomendasikan sebagai motor penggerak bank soal otomatis.

Kata kunci: *Automatic Question Generation*; E-Learning; Ekstraksi Kata Kunci; *TextRank*; TF-IDF

Abstract. The existence of practical evaluation instruments is a vital pillar in the digital learning ecosystem. Automatic Question Generation (AQG) technology is designed to reduce educators' cognitive load, but low keyword extraction accuracy often degrades the quality of generated questions. Single-frequency approaches are susceptible to noise absorption, while graph algorithms ignore word scarcity. This study formulates a hybrid optimization technique by integrating Term Frequency-Inverse Document Frequency (TF-IDF) scores and TextRank centrality to select keywords in Internet of Things (IoT) literature. The processing is executed using an unsupervised learning framework at the unigram level, evaluated against an expert Ground Truth and the grammatical feasibility of cloze tests. Computational calculations at the top 30 word limit proved that the hybrid method achieved an F1-Score of 40.68%, equivalent to pure TF-IDF and surpassing TextRank (37.29%). In the pedagogical validation phase, the proposed model demonstrated absolute superiority with a feasibility rate of 86.67% (26 valid quizzes), subjugating TF-IDF which stalled at 83.33%. The intervention of graph topology successfully deterred the intrusion of meaningless words and maintained syntactic integrity, making this hybrid formulation highly recommended as the driving engine for future question banks.

Keywords: Automatic Question Generation; E-Learning; Keyword Ekstraktion; TextRank; TF-IDF

1. Pendahuluan

Transformasi ekosistem pendidikan menuju era digital menuntut ketersediaan *platform E-Learning* yang interaktif dan komprehensif. Kendati *platform* ini memberikan kebebasan akses yang optimal, tahapan penyediaan bahan evaluasi belajar sering kali menjadi hambatan administratif. Aktivitas perakitan instrumen ujian secara manual menyerap alokasi waktu dan fokus yang sangat masif dari para pengajar [1]. Rutinitas administratif yang repetitif ini menjadi kendala krusial karena menyita energi produktif pendidik, membatasi porsi interaksi pedagogis langsung, serta memperlambat siklus evaluasi materi ajar. Menyikapi defisiensi tersebut, rekayasa bahasa alami atau *Natural Language Processing* (NLP) melalui skema *Automatic Question Generation* (AQG) hadir sebagai terobosan untuk mentransformasi teks materi menjadi butir kuis secara komputasional.

Berbagai tinjauan literatur (*State of the Art*) mengonfirmasi bahwa luaran sistem AQG masih memiliki celah kualitas yang menuntut perbaikan. Pendekatan yang menggunakan arsitektur *Deep Learning* kompleks seperti *Bidirectional Encoder Representations from Transformers* (BERT) atau T5 memang mampu mendongkrak akurasi hingga di atas 80% [1–3]. Namun, implementasi model bahasa berskala besar ini menuntut spesifikasi infrastruktur komputasi (seperti *Graphics Processing Unit* / GPU) yang terlampaui berat, menjadikannya kurang realistis untuk diintegrasikan pada ekosistem *E-Learning* standar institusi. Di sisi lain, metode teringan dan paling lazim diadopsi untuk menyeleksi kata target adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). Sayangnya, algoritma konvensional ini memiliki keterbatasan fundamental berupa “kebutaan” terhadap kohesi semantik antar kata dalam sebuah kalimat [4]. Sebaliknya, teknik *TextRank* beroperasi dengan menakar urgensi istilah melalui jejaring konektivitas graf paragraf [5–7], namun abai terhadap probabilitas keunikan kata secara global di seluruh dokumen. Sebagai alternatif, model representasi kata berbasis jaringan saraf seperti *Word2Vec* memang sanggup memetakan kedekatan semantik antar kata secara presisi. Akan tetapi, implementasi *Word2Vec* menuntut proses pelatihan awal (*pre-training*) pada korpus teks yang masif dan memori komputasi yang lebih besar, sehingga menjadikannya kurang praktis dan tidak efisien jika diterapkan pada ekstraksi dokumen ajar tunggal berskala kecil [8].

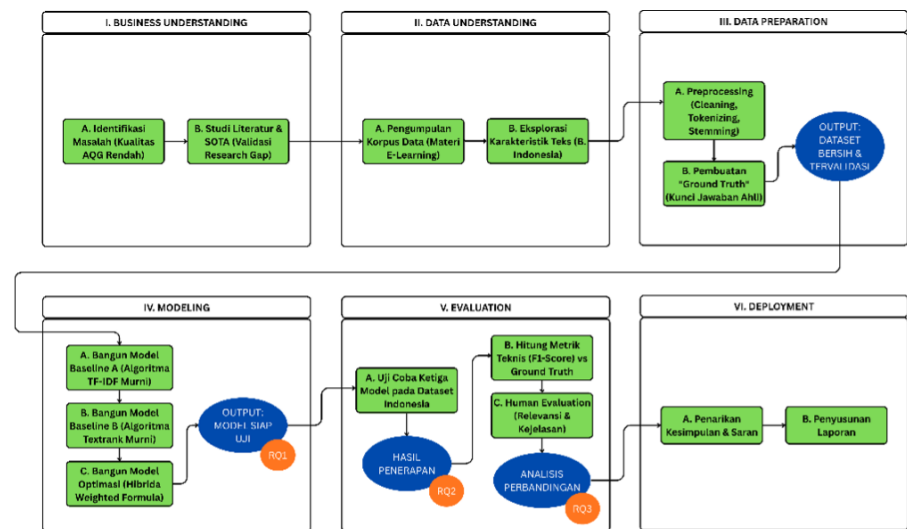
Dalam ekosistem pendidikan digital modern, pemanfaatan *Internet of Things* (IoT) pada *platform E-Learning* berbasis web telah membuka dimensi baru dalam pembelajaran interaktif dan penyampaian materi [9–13]. Integrasi sistem kecerdasan buatan (AI) di dalamnya terus dieksplorasi secara intensif guna memberikan wawasan prediktif dan manajemen sistem yang jauh lebih cerdas [14–18]. Sejalan dengan itu, tantangan krusial dalam sistem *E-Learning* adalah tingginya volume teks, sehingga menuntut adanya otomatisasi evaluasi dokumen. Berbagai pendekatan *Deep Learning* dan pemrosesan bahasa alami (NLP) pun diimplementasikan untuk meningkatkan akurasi klasifikasi teks serta penarikan fitur linguistik [8, 19–22].

Guna menjembatani kesenjangan tersebut, beberapa studi terkini mulai mengeksplorasi penggabungan (fusi) algoritma penambangan teks [23, 24]. Namun, keterbatasan utama dari penelitian-penelitian terdahulu adalah mayoritas model hibrida yang dikembangkan hanya dievaluasi untuk tugas peringkasan dokumen (*text summarization*) atau beroperasi pada domain artikel berita umum. Hingga saat ini, belum ada studi yang merumuskan dan menguji fusi TF-IDF dan *TextRank* secara spesifik untuk memecahkan kendala akurasi ekstraksi kata kunci pada domain *Automatic Question Generation* (AQG). Berangkat dari celah penelitian (*research gap*) tersebut, penelitian ini merancang sebuah pemodelan hibrida yang dioptimasi khusus untuk skenario AQG, menyilangkan bobot frekuensi TF-IDF dengan nilai sentralitas *TextRank*. Pemodelan silang ini dirancang untuk memunculkan ekuilibrium, sehingga peranti lunak tidak sekadar memungut kata yang masif diulang, tetapi juga kata yang secara struktural menjadi pilar penyangga gagasan utama materi. Uji coba dijalankan pada modul pembelajaran

elektronik berbahasa Indonesia di bidang *Internet of Things* (IoT), guna memverifikasi apakah taktik hibrida ini sanggup mematahkan batasan metode konvensional dalam memasok kata kunci untuk pembuatan soal rumpang (*cloze test*). Maka secara spesifik, tujuan penelitian ini adalah mengevaluasi tingkat akurasi ekstraksi kata kunci dari model hibrida usulan dibandingkan dengan algoritma dasarnya. Adapun kontribusi utama dari studi ini adalah menghadirkan arsitektur peranti lunak pembangkit soal otomatis yang efisien untuk diintegrasikan secara langsung pada ekosistem *platform E-Learning*, guna memangkas redundansi tugas administratif pengajar secara praktis.

2. Bahan dan Metode

Rancangan metodologis dalam studi ini dikembangkan dengan mengadopsi kerangka kerja standar industri *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang dieksekusi murni melalui skema pembelajaran tanpa pengawasan (*unsupervised learning*). Sistematika pelaksanaan eksperimen diuraikan secara komprehensif pada [Gambar 1](#) berikut:



Gambar 1. Alur Penelitian

2.1. Business Understanding

Fase inisiasi ini bertumpu pada perumusan objektif riset yang diderivasi dari kendala empiris. Langkah pertama (A) mencakup identifikasi persoalan, yaitu menyoroti degradasi relevansi dan validitas linguistik pada instrumen soal yang dilahirkan oleh sistem AQG saat ini akibat inakurasi seleksi kata kunci. Langkah kedua (B) melibatkan peninjauan literatur dan analisis *State of the Art* (SOTA) guna memvalidasi kesenjangan riset (*research gap*), yang mengonfirmasi esensi penerapan strategi optimasi hibrida.

2.2. Data Understanding

Tahapan ini dikonsentrasikan pada pengumpulan dan pengenalan karakteristik objek studi. Aktivitas awal (A) adalah akuisisi korpus data, di mana peneliti menghimpun dokumen materi instruksional *E-Learning* sebagai bahan baku utama komputasi [25]. Selanjutnya, aktivitas (B) mengeksplorasi anatomi linguistik teks berbahasa Indonesia untuk memastikan kompatibilitas struktur sintaksisnya dengan algoritma yang akan dieksekusi.

2.3. Data Preparation

Fase ini didedikasikan untuk mengonversi teks mentah menjadi kumpulan data yang siap diolah mesin. Langkah (A) merupakan prapemrosesan bertingkat yang mensterilisasi data melalui eliminasi karakter bising (*cleaning*), segmentasi

struktur kalimat (*tokenizing*), serta pemotongan kelas imbuhan (*stemming*) menggunakan pustaka Sastrawi. Sebagai representasi tahapan ini, sebuah kalimat mentah "Sistem IoT mengumpulkan data dari sensor-sensor!" akan dibersihkan dari tanda baca dan diseragamkan menjadi huruf kecil pada tahap *cleaning* ("sistem iot mengumpulkan data dari sensor sensor"). Selanjutnya, kalimat tersebut dipisah menjadi sekumpulan kata mandiri pada tahap *tokenizing* ("sistem", "iot", "mengumpulkan", "data", "dari", "sensor", "sensor"]), dan diakhiri dengan tahap *stemming* untuk mereduksi kata berimbuhan kembali ke akar katanya ("sistem", "iot", "kumpul", "data", "dari", "sensor", "sensor"]). Langkah (B) adalah konstruksi kunci jawaban pakar secara manual. Luaran (*Output*) final dari fase ini adalah dihasilkannya Dataset Bersih dan Tervalidasi.

2.4. Modeling

Tahap ini merepresentasikan jantung eksperimentasi teknis melalui pembangunan tiga arsitektur komputasi. Peneliti membangun model komparatif pertama (A) yang bertumpu murni pada *baseline* statistik TF-IDF. Pendekatan ini dipilih secara spesifik karena efisiensinya dalam memproses dokumen tunggal secara instan tanpa menuntut prapelatihan (*pre-training*) pada korpus masif seperti halnya *Word2Vec*. Eksperimen kemudian disusul oleh model komparatif kedua (B) yang menggunakan *baseline* graf *TextRank*. Sebagai pamungkas, dibangun model optimasi (C) yang meleburkan kedua algoritma tersebut melalui formulasi pembobotan (Hibrida *Weighted Formula*). Luaran dari fase ini adalah Model Siap Uji, yang secara simultan menuntaskan Rumusan Masalah 1 (RQ1) terkait rekayasa metode ekstraksi.

Dalam arsitektur ekstraksi ini, kandidat kata kunci dibatasi secara ketat pada tataran kata tunggal (*unigram*). Filtrasi kata buangan (*noise removal*) tidak bersandar pada identifikasi kelas kata (*Part-of-Speech Tagging*), melainkan dieliminasi secara leksikal menggunakan kamus stopwords kustom yang dirancang agar selaras dengan gaya penulisan literatur *E-Book*. Akurasi komputasional (F1-Score) diukur menggunakan kriteria *Stemmed Match*, di mana prediksi mesin dinilai valid (cocok) terhadap acuan pakar apabila akar katanya identik—sebagai contoh, variasi morfologi seperti 'pemrosesan' dan 'proses' dianggap ekuivalen.

2.4.1. TF-IDF

Pada skenario pertama, penentuan skor bobot TF-IDF didasarkan pada operasi perkalian antara frekuensi internal kata dan tingkat kelangkaan dokumen [1]. Secara matematis, kalkulasi untuk menghitung bobot akhir TF-IDF didasarkan pada Persamaan (1) sebagai berikut:

$$IDF(t, D) = \log \left(\frac{\text{Jumlah total dokumen } |D|}{\text{Jumlah dokumen yang mengandung kata } t} \right) \quad (1)$$

2.4.2. TextRank

Pada skenario kedua, algoritma *TextRank* yang berakar dari mekanisme *PageRank* mengevaluasi urgensi kata melalui analisis konektivitas graf [5, 6]. Secara rekursif, perhitungan bobot simpul kata (V_i) diuraikan pada Persamaan (2) sebagai berikut:

$$S(V_i) = (1 - d) + d \times \sum_{j \in \text{in}(V_i)} \frac{1}{|\text{out}(V_j)|} S(V_j) \quad (2)$$

2.4.3. Hibrida

Untuk skenario ketiga (model usulan), formulasi hibrida mengintegrasikan bobot statistik TF-IDF dan sentralitas graf *TextRank*. Mengingat kedua metode dasar memiliki skala hitung mentah yang berbeda, skor masing-masing disetarakan terlebih dahulu ke dalam rentang $[0, 1]$ menggunakan standardisasi data

(*Min-Max Normalization*) sesuai dengan [Persamaan \(3\)](#) sebagai berikut:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (3)$$

di mana X merupakan skor mentah kata, sedangkan $X_{\min}$ dan $X_{\max}$ berturut-turut merepresentasikan nilai minimum dan maksimum dari himpunan skor kata pada dokumen.

Setelah melalui normalisasi, formulasi hibrida direkayasa melalui penetapan rasio bobot yang diputuskan secara empiris melalui serangkaian uji komparatif (studi ablati). Dalam proses penalaan (*tuning*) tersebut, peneliti menguji beberapa variasi kombinasi rasio secara iteratif (mulai dari 50:50, 70:30, hingga 30:70). Berdasarkan hasil iterasi, kombinasi bobot 40% (0.4) untuk TF-IDF dan 60% (0.6) untuk *TextRank* secara konsisten menghasilkan skor metrik evaluasi yang paling optimal. Skor dari kedua metode dasar disetarakan terlebih dahulu melalui standarisasi data (*Min-Max Normalization*), sebelum diintegrasikan sesuai dengan rumusan pada [Persamaan \(4\)](#) sebagai berikut:

$$\text{Skor Hibrida} = (0.4 \times \text{Skor TF-IDF}) + (0.6 \times \text{Skor Text Rank}) \quad (4)$$

Porsi bobot yang sedikit lebih condong pada *TextRank* (60%) secara sengaja dirancang untuk meredam kemunculan istilah pasaran (derau) berfrekuensi tinggi yang sering luput dari penyaringan TF-IDF murni, tanpa merusak kestabilan matematis yang ada.

Kalimat referensi yang memuat kosakata terpilih kemudian didistribusikan ke dalam mesin pembuat soal berbasis ekstraktif. Algoritma melakukan substitusi (penyamaran) terhadap kata target dengan spasi rumpang. Sistem mengimplementasikan memori *cache* pelacakan kalimat untuk menjamin sebuah teks hanya diekstrak satu kali guna mencegah redundansi kebocoran jawaban (*answer leakage*). Mesin kemudian menyeleksi tiga distraktor (pengecoh) secara acak dari matriks kata kunci yang tersisa untuk melengkapi opsi ganda penyusun instrumen evaluasi.

2.5. Evaluation

Kinerja model diukur secara kuantitatif dan kualitatif guna memvalidasi hipotesis. Langkah (A) menguji ketiga model tersebut secara komparatif pada korpus *E-Learning*, menghasilkan *Output* berupa Hasil Penerapan yang merespons Rumusan Masalah 2 (RQ2). Selanjutnya, kalkulasi metrik teknis (B) dieksekusi dengan membandingkan nilai F1-Score luaran mesin terhadap *Ground Truth*. Pengujian komputasional ini mengadopsi matriks evaluasi berbasis *True Positives* (TP) untuk tebakan yang selaras, *False Positives* (FP) untuk tebakan meleset berupa derau, dan *False Negatives* (FN) untuk kunci pakar yang terlewat. Formulasi perhitungan nilai Precision ditentukan melalui [Persamaan \(5\)](#), nilai Recall pada [Persamaan \(6\)](#), serta nilai pengujian harmonik F1-Score dirumuskan pada [Persamaan \(7\)](#) sebagai berikut:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Langkah penutup (C) adalah *Human Evaluation*, di mana pakar meninjau tingkat relevansi materi dan kejelasan gramatikal kuis. Kolaborasi evaluasi ini membuahkan *Output* Analisis Perbandingan untuk menjawab Rumusan Masalah 3 (RQ3) mengenai signifikansi peningkatan kualitas soal.

2.6. Deployment

Fase purnajual ini merangkum keseluruhan siklus, diawali dengan (A) penarikan konklusi komprehensif dan penyusunan rekomendasi pengembangan, yang kemudian didokumentasikan ke dalam (B) penyusunan laporan akademis yang terstruktur.

3. Hasil

Bagian ini memaparkan catatan komprehensif dari kinerja peranti lunak, baik pada tataran akurasi komputasional ekstraksi kata kunci maupun kualitas final dari instrumen evaluasi yang digenerasi.

3.1. Evaluasi Kinerja Komputasional Ekstraksi

Fase pengujian awal dipusatkan pada kesanggupan algoritma komputasi dalam mencocokkan identifikasinya dengan terminologi esensial yang telah dianotasi oleh ahli materi (*Validator*). Mengingat arsitektur mesin ekstraksi yang dikembangkan beroperasi secara spesifik pada tataran kata tunggal (*unigram*), maka daftar istilah yang dipilih oleh pakar perlu melalui tahap penyesuaian dimensi. Apabila ahli materi menetapkan istilah majemuk yang terdiri dari dua kata atau lebih (sebagai contoh: frasa "*Machine Learning*" atau "*Smart Cities*"), maka frasa tersebut dipisahkan dan disegmentasi menjadi unit kata mandiri ("*Machine*", "*Learning*", "*Smart*", "*Cities*"). Atau jika terdapat dua kata istilah yang sama, maka salah satu kata istilah akan dihapus untuk menghindari terjadinya duplikasi. Ataupun jika terdapat kata istilah yang memiliki makna penting dalam bahasa Inggris dan disetujui oleh ahli materi untuk menambahkan makna tersebut dalam bahasa Inggris dan Indonesia, maka kata istilah tersebut akan dimasukkan dalam dua versi bahasa, yaitu bahasa Indonesia dan Inggris.

Melalui proses segmentasi (*tokenizing*) terhadap seluruh hasil anotasi tersebut, himpunan kata acuan pakar dikonversi secara ekuivalen menjadi 29 terminologi *unigram* yang ditanamkan ke dalam kode program sebagai standar kebenaran mutlak (*Ground Truth*). Cuplikan representasi data acuan dari pakar (sebelum diproses menjadi *unigram*) diilustrasikan pada Tabel 1.

Tabel 1. Cuplikan Standar Kebenaran (*Ground Truth*) Anotasi Pakar

No	Konteks Kalimat	Istilah Penting	Hasil (<i>Ground Truth</i>)
1	" <i>Internet of Things</i> (IoT) adalah sebuah konsep yang memungkinkan perangkat fisik yang terhubung ke Internet untuk mengumpulkan, berbagi, dan memproses data..."	<i>Internet of Things</i> (IoT)	"iot"
2	"...kini memasuki era industri 4.0... yang mengintegrasikan teknologi IoT dengan otomatisasi, analitik, dan kecerdasan buatan (AI) untuk menciptakan pabrik pintar."	industri 4.0	"industri"
3	"...masyarakat beralih ke era society 5.0, yaitu masyarakat cerdas yang akrab dengan internet, big data, dan kecerdasan buatan."	society 5.0	"society"
4	"Sensor mengumpulkan data dari lingkungan fisik, seperti suhu, kelembapan, dan cahaya, dan mengubahnya menjadi sinyal..."	Sensor	"sensor"
5	"Aktuator melakukan tindakan fisik berdasarkan perintah yang diterima dari sistem IoT, seperti menyalakan lampu atau menggerakkan motor."	Aktuator	"aktuator"
6	"Perangkat <i>edge</i> , seperti mikrokontroler dan mikroprosesor, memproses data dari sensor sebelum mengirimkannya ke Cloud..."	Perangkat <i>edge</i>	"perangkat", "edge"

No	Konteks Kalimat	Istilah Penting	Hasil (<i>Ground Truth</i>)
7	"...seperti mikrokontroler dan mikroprosesor, memproses data dari sensor sebelum mengirimkannya ke Cloud..."	mikrokontroler dan mikroprosesor	"mikrokontroler", "mikroprosesor"
8	"IoT mentransmisikan data menggunakan berbagai protokol komunikasi seperti MQTT, CoAP, dan HTTP..."	protokol komunikasi	"protokol", "komunikasi"
9	" <i>Gateway</i> IoT berfungsi sebagai penghubung antara perangkat IoT dan jaringan yang lebih besar..."	<i>Gateway</i> IoT	" <i>gateway</i> "
10	"Cloud menyediakan sumber daya komputasi yang luas untuk penyimpanan dan analisis data skala besar."	Cloud	"cloud"
11	"...dengan teknologi big data dan <i>machine learning</i> digunakan untuk mengenali pola dan tren guna membantu pengambilan keputusan."	big data	"big", "data"
12	"...mengintegrasikan teknologi IoT dengan otomatisasi, analitik, dan kecerdasan buatan (AI)..."	kecerdasan buatan (AI)	"kecerdasan", "buatan", " <i>machine</i> ", " <i>learning</i> ", "ai"
13	"...penggunaan teknologi RFID (<i>Radio Frequency Identification</i>) untuk mengoptimalkan manajemen rantai pasokan perusahaan."	RFID (<i>Radio Frequency Identification</i>)	"rfid"
14	"Jaringan komunikasi memungkinkan perangkat-perangkat ini untuk terhubung dan saling berinteraksi. Teknologi seperti Wi-Fi, Bluetooth, Zigbee, LoRaWAN, dan 5G..."	Jaringan komunikasi	"jaringan"
15	"Teknik enkripsi, autentikasi perangkat, serta manajemen identitas dan akses sangat penting untuk melindungi integritas data..."	enkripsi	"enkripsi", "autentikasi"
16	"IoT memungkinkan otomatisasi dan kontrol yang lebih efisien dalam berbagai aplikasi seperti layanan kesehatan, pertanian, transportasi..."	otomatisasi	"otomatisasi"
17	"Kemampuan IoT untuk mengumpulkan, menganalisis, dan berbagi data secara <i>real-time</i> membawa perubahan signifikan..."	<i>real-time</i>	" <i>real-time</i> "
18	"IoT berperan penting dalam pengembangan <i>smart cities</i> , yang mengintegrasikan teknologi untuk meningkatkan manajemen energi, mengurangi polusi..."	<i>smart cities</i>	" <i>smart</i> ", " <i>cities</i> "
19	"Aplikasi analitik memainkan peran penting dalam mengolah data yang dikumpulkan oleh perangkat IoT."	analitik	"analitik"
Total		29 kata kunci	

Sebagai tahapan awal sebelum mengevaluasi akurasi menggunakan F1-Score, peranti lunak mengeksekusi komputasi pembobotan (*scoring*) terhadap seluruh himpunan kata yang lolos tahap prapemrosesan. Skor dari masing-masing algoritma dasar terlebih dahulu disetarakan menggunakan *Min-Max Normalization*, lalu diintegrasikan melalui persamaan hibrida usulan guna menentukan peringkat akhir kata kunci.

Cuplikan 10 besar (Top-10) kata kunci dengan perolehan skor normalisasi tertinggi divisualisasikan pada [Tabel 2](#).

Tabel 2. Representasi Skor Normalisasi TF-IDF, *TextRank*, dan Hibrida (Top-10)

Kata Kunci	TF-IDF	TextRank	Hibrida
IoT	1	1	1
Data	0,9809	0,8234	0,9021
Perangkat	0,7278	0,7747	0,7513
Aman	0,4468	0,3535	0,4002
Sensor	0,4498	0,3070	0,3784
Teknologi	0,4022	0,3072	0,3547
Komunikasi	0,4204	0,2647	0,3426
Sistem	0,3249	0,3166	0,3207
Komputasi	0,3820	0,2310	0,3065
Platform	0,3772	0,2335	0,3054

Berdasarkan Tabel 2, terlihat mekanisme algoritma usulan dalam menciptakan ekuilibrium (keseimbangan) penilaian. Kata kunci spesifik seperti "Iot" dan "Data" secara konsisten mendominasi probabilitas frekuensi (TF-IDF) sekaligus memiliki kepadatan konektivitas graf (*TextRank*). Skor Hibrida merangkul kedua variabel sentral ini menjadi satu nilai desimal, di mana peringkat teratas dari daftar skor komputasional inilah yang selanjutnya dikirim ke mesin perakit soal untuk diuji kelayakannya terhadap *Ground Truth* pakar.

Maka kemudian komparasi pencapaian angka akurasi ekstraksi antara metode dasar dan arsitektur gabungan pada batas 30 frasa teratas (Top-30) yang divisualisasikan pada Tabel 2, kemudian dilanjutkan dengan pengukuran yang didasarkan pada parameter *True Positives* (TP) atau tebakan selaras, *False Positives* (FP) atau tebakan meleset berupa derau, dan *False Negatives* (FN) atau kunci pakar yang terlewat yang dapat dilihat pada Tabel 3.

Tabel 3. Rekapitulasi Kinerja Ekstraksi Kata Kunci (Top-30)

Model Ekstraksi	TP	FP	FN	Precision	Recall	F1-Score
<i>TextRank</i> Murni	11	19	18	36,67%	37,93%	37,29%
TF-IDF Murni	12	18	17	40,00%	41,38%	40,68%
Hibrida (Usulan)	12	18	17	40,00%	41,38%	40,68%

Merujuk pada parameter di atas, algoritma persilangan berhasil mendeteksi 12 istilah valid, dengan 18 kesalahan prediksi dan 17 kegagalan deteksi. Dari kacamata numerik, sistem hibrida mencatatkan persentase presisi 40,00% dan kelengkapan penemuan (*recall*) 41,38%, yang berakumulasi pada nilai F1-Score sebesar 40,68%. Data komparatif ini menegaskan bahwa strategi hibrida sukses melampaui kemampuan pendekatan graf (*TextRank* murni) yang tertahan di skor 37,29%. Secara perhitungan matematis, perolehan metrik model usulan ini bernilai identik dengan jalur TF-IDF murni (40,68%).

3.2. Validasi Mutu Pedagogis Instrumen Soal

Mengingat catatan akurasi komputasional menunjukkan hasil seri antara model frekuensi (TF-IDF) murni dan gabungan, pembuktian keandalan substantif dari kedua algoritma harus diadili melalui kualitas luaran puncaknya. Evaluasi ini dieksekusi secara manual (penilaian manusia) oleh kepakaran materi, berpautan pada indikator keutuhan konteks, kelogisan tata bahasa, ketiadaan rujukan ganda (anafora), serta kepastian satu-satunya jawaban benar pada opsi ganda. Rekapitulasi penolakan dan penerimaan kuis yang dievaluasi oleh pakar disajikan secara rinci pada Tabel 4.

Tabel 4. Komparasi Kelayakan Instrumen Soal Rumpang (*Cloze Test*)

Kategori Evaluasi	Model Hibrida	TF-IDF Murni	<i>TextRank</i> Murni
Total Dievaluasi	30 Butir	30 Butir	30 Butir
Relevan (Valid)	26 Butir	25 Butir	22 Butir
Tidak Relevan (Ditolak)	4 Butir	5 Butir	8 Butir
Akurasi Kelayakan	86,67%	83,33%	73,33%

Fase verifikasi pedagogis pada Tabel 4 secara konklusif mengesahkan superioritas fungsional dari arsitektur hibrida. Pendekatan ini memperlihatkan ketahanan operasional yang jauh lebih kokoh dengan mencetak 26 pertanyaan sah (tingkat akseptasi 86,67%), mengalahkan produksi sistem TF-IDF murni yang hanya mampu memvalidasi 25 butir (83,33%) dan juga *TextRank* murni yang memvalidasi hanya 22 butir (73,33%).

Untuk memberikan gambaran empiris yang lebih komprehensif mengenai kapabilitas sistem dalam mensintesis evaluasi secara mandiri, representasi dari butir kalimat soal rumpang hasil pembangkitan model hibrida yang telah dinyatakan valid oleh pakar disajikan secara rinci pada Tabel 5.

Tabel 5. Representasi Kalimat Soal Valid Hasil Pembangkitan Sistem Hibrida

No	Konstruksi Pertanyaan (Opsis Tersamar)	Opsis Pilihan Ganda	Kunci Jawaban
1	Pengantar <i>Internet of Things</i> (.....) adalah konsep revolusioner yang menghubungkan berbagai perangkat fisik ke koneksi internet, memungkinkan untuk mengumpulkan, mengolah, berbagi, dan mampu menganalisis data secara <i>real-time</i> (Puspita, 2024)	A. Perangkat B. Sistem C. Data D. Iot	D. Iot
2	Keunggulan utama IoT terletak pada kemampuannya untuk menyediakan yang kaya dan mendalam, serta dapat dianalisis untuk menghasilkan wawasan yang berharga	A. Data B. Sistem C. Aman D. Iot	A. Data
3	Melalui sensor, aktuator, dan jaringan komunikasi maka ini dapat berinteraksi, berkomunikasi, dan berkolaborasi untuk menciptakan ekosistem yang cerdas dan responsif	A. Perangkat B. Jaring C. Sistem D. Aplikasi	A. Perangkat
4	Teknologi keamanan utama, meliputi: enkripsi yang mengamankan data yang ditransmisikan antara perangkat, autentikasi yang memastikan hanya pengguna dan perangkat yang sah dapat mengakses jaringan, <i>firewall</i> dan IDS/IPS yang mampu melindungi jaringan dari serangan siber, dan manajemen sertifikat digital untuk komunikasi yang	A. Aman B. Platform C. Aktuator D. Teknologi	A. Aman
5	 dan Aktuator Dalam konteks <i>Internet of Things</i> (IoT), ... adalah perangkat yang mendeteksi dan mengukur perubahan fisik atau lingkungan, kemudian mengubahnya menjadi sinyal yang dapat dianalisis oleh sistem komputer (Javaid et al.)	A. Sensor B. Aktuator C. Komputasi D. Aman	A. Sensor

4. Pembahasan

Penelusuran lebih mendalam terhadap alasan penolakan pada sisa instrumen kuis berhasil menguak adanya limitasi inheren dalam memroses morfologi bahasa Indonesia secara komputasional. Dari empat soal hibrida yang dianulir oleh evaluator, tiga diantaranya murni diakibatkan oleh fenomena pemotongan radikal (*over-stemming*) yang diproses oleh *library* NLP Sastrawi.

Sebagai contoh empiris dari cacat leksikal ini, algoritma Hibrida dan TF-IDF sama-sama mengalami galat ketika mesin stemmer mereduksi kata benda "Lingkungan" secara ekstrem menjadi "Lingkung", "Jaringan" menjadi "Jaring", serta "Analisis" menjadi "Analis". Hal ini memicu disrupsi fatal pada integritas kalimat rumpang dan tata bahasa opsi ganda, salah satunya memproduksi soal cacat: "Sensor menjadi komponen kunci yang memungkinkan IoT mencapai suatu tujuan untuk menciptakan ----- yang lebih cerdas, responsif, dan terintegrasi." dengan kunci jawaban dipaksa menjadi "Lingkung" alih-alih "Lingkungan". Sebagai contoh bisa dilihat pada Tabel 6.

Tabel 6. Representasi Kasus Kegagalan Ekstraksi (Soal Ditolak)

No	Konstruksi Pertanyaan (Opsi Tersamar)	Opsi Pilihan Ganda	Kunci Jawaban (Cacat kata)
1	Wi-Fi adalah teknologi yang sering digunakan untuk menghubungkan perangkat dalam area lokal (LAN), dengan kecepatan tinggi dan jangkauan yang cukup luas (Priantama, 2017)	A. Iot B. Aman C. Cloud D. Jaring	D. Jaring (<i>Over-stemming</i> pada kata “Jaringan”)
2	Sensor menjadi komponen kunci yang memungkinkan IoT mencapai suatu tujuan untuk menciptakan yang lebih cerdas, responsif, dan terintegrasi	A. Cloud B. Data C. Analisis D. Lingkung	D. Lingkung (<i>Over-stemming</i> pada kata “Lingkungan”)

Akan tetapi, divergensi kualitas yang mencolok antara Hibrida dan TF-IDF terlihat jelas pada tipologi penolakan lainnya. Di luar infeksi leksikal *stemming*, algoritma TF-IDF murni mengidap kelemahan “kebutaan konteks” atau kegagalan membaca kepadatan relasi antar kata. Keterbatasan ini terkonfirmasi secara empiris saat TF-IDF gagal memfilter kata-kata yang kehilangan keutuhan konteks kalimatnya, sehingga menghasilkan soal ambigu yang digugurkan oleh pakar, seperti pada ekstraksi target kata “Internet” maupun “Daya”. Cacat bawaan TF-IDF ini sejalan dengan temuan riset terdahulu yang mengonfirmasi bahwa pendekatan berbasis frekuensi tunggal rentan mengekstrak istilah yang tidak memiliki tautan kohesi semantik di dalam teks [4]. Sebaliknya, fusi perhitungan *TextRank* dalam rumusan hibrida secara meyakinkan sanggup memitigasi kelemahan tersebut. Pemetaan topologi graf bekerja secara presisi sebagai tameng saringan; memastikan bahwa kosakata target mutlak diserap dari kalimat berstruktur naratif utuh (memiliki subjek dan predikat yang kohesif), sekaligus membendung penarikan kata yang miskin konteks spesifik IoT. Superioritas arsitektur graf dalam memelihara relasi tekstual ini turut memperkuat kesimpulan studi-studi sebelumnya yang membuktikan bahwa pemodelan hibrida mampu menekan tingkat ekstraksi derau (*noise*) secara signifikan [11, 24]. Berkat mekanisme proteksi ganda (frekuensi dan semantik) tersebut, instrumen ujian yang dicetak oleh sistem hibrida terwujud dengan konfigurasi tata bahasa yang jauh lebih logis, stabil, dan relevan untuk didayagunakan dalam siklus pendidikan digital. Temuan empiris ini turut mengonfirmasi literatur terdahulu sekaligus mengindikasikan bahwa integrasi antara bobot statistik dan topologi graf dapat menjadi solusi yang andal guna menambal kelemahan algoritma tunggal pada proses ekstraksi kata kunci dokumen ajar.

5. Kesimpulan

Penelitian ini berhasil mengimplementasikan model hibrida TF-IDF (40%) dan *TextRank* (60%) untuk mengoptimalkan ekstraksi kata kunci pada sistem *Automatic Question Generation* (AQG). Secara komputasional (batas Top-30 *unigram*), model hibrida mencatatkan F1-Score sebesar 40,68%, setara dengan performa TF-IDF murni dan mengungguli *TextRank* murni (37,29%). Walaupun secara metrik statistik ekuivalen dengan TF-IDF, pengujian pedagogis melalui validasi pakar materi (*human evaluation*) membuktikan bahwa model hibrida lebih unggul dengan persentase kelayakan soal rumpang (*cloze test*) mencapai 86,67% (26 butir valid) dibandingkan TF-IDF murni (83,33%). Keunggulan ini dicapai karena fitur *TextRank* secara efektif mampu menyaring kata yang miskin konteks dan menjaga keutuhan sintaksis kalimat pertanyaan. Meskipun demikian, sistem ini masih memiliki keterbatasan berupa tingginya ketergantungan pada modul prapemrosesan bahasa, di mana penggunaan pustaka Sastrawi memicu galat pemotongan berlebihan (*over-stemming*) pada kata esensial (seperti “Lingkungan” menjadi “Lingkung”) yang mengganggu tata bahasa luaran soal, serta ruang lingkup pengujian yang masih terbatas pada materi satu bab dokumen ajar. Oleh karena itu, penelitian selanjutnya disarankan untuk mengintegrasikan metode *stemming* yang lebih adaptif, memperluas skala pengujian pada dokumen multibab, serta mengeksplorasi ekstraksi pada level frasa (n-gram) guna mening-

katkan kedalaman dan variasi bank soal otomatis yang dihasilkan.

Pernyataan Penggunaan *Artificial Intelligence*. Penulis menggunakan Gemini Pro AI untuk membantu mencari jurnal referensi dan merapikan data-data penelitian serta memperbaiki tata bahasa yang kurang rapi. Seluruh keluaran yang dihasilkan oleh AI telah diperiksa dan diverifikasi oleh penulis. Penulis bertanggung jawab sepenuhnya atas keakuratan, orisinalitas, dan integritas isi artikel.

Pernyataan Kontribusi Penulis. JWW berkontribusi dalam konseptualisasi, metodologi, pengumpulan data, dan analisis penelitian. EMD berkontribusi dalam pengembangan dan implementasi sistem serta interpretasi hasil penelitian. AAGAMP berkontribusi dalam penyusunan, peninjauan, dan penyuntingan naskah. Seluruh penulis telah membaca, meninjau, dan menyetujui versi akhir artikel serta bertanggung jawab atas seluruh aspek penelitian dan publikasi.

Pernyataan Konflik Kepentingan. Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian, kepengarangan, dan publikasi artikel ini.

Pernyataan Pendanaan. Penelitian ini tidak menerima pendanaan khusus dari lembaga pemerintah, komersial, maupun organisasi nirlaba.

Ucapan Terima Kasih. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Primakara University atas segala fasilitas riset, dukungan infrastruktur, serta bimbingan akademis yang telah diberikan sepanjang pelaksanaan penelitian ini.

Daftar Pustaka

- [1] F. Maheen, M. Asif, H. Ahmad, S. Ahmad, F. Alturise, O. Asiry, and Y. Y. Ghadi, "Automatic computer science domain multiple-choice questions generation based on informative sentences," *PeerJ Comput. Sci.*, vol. 8, p. e1010, 2022. <https://doi.org/10.7717/PEERJ-CS.1010>.
- [2] E. Yehia, N. Hassan, and S. AbdelGaber, "Towards more effective automatic question generation: A hybrid approach for extracting informative sentences," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 9, 2025. <https://doi.org/10.14569/IJACSA.2025.0160920>.
- [3] I. D. Osibajo, O. E. Olorunshola, F. Adamu-Fika, and T. J. Samson, "Hybrid zero-shot nlp pipeline for text summarization and question generation," *Int. J. Res. Innov. Appl. Sci.*, vol. 10, no. 7, pp. 342–354, 2025. <https://doi.org/10.51584/IJRIAS.2025.100700030>.
- [4] Z. Xu and J. Zhang, "Extracting keywords from texts based on word frequency and association features," *Procedia Comput. Sci.*, vol. 187, pp. 77–82, 2021. <https://doi.org/10.1016/j.procs.2021.04.035>.
- [5] M. T. A. Anwar, S. H. Wijoyo, and W. H. N. Putra, "Implementasi metode textrank dan named entity recognition untuk ekstraksi kata kunci pada media online berita," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 5, no. 1, pp. 34–41, 2024. <https://doi.org/10.25126/justsi.v5i1.401>.
- [6] S. Halmahera, "Peringkasan teks artikel berita menggunakan algoritma textrank dengan widf dan lsa," 2025. <http://etheses.uin-malang.ac.id/id/eprint/74957>.
- [7] F. Husniah, "Peringkasan teks otomatis artikel berbahasa indonesia menggunakan algoritma textrank," *Peringkasan Teks Otomatis Artikel Berbahasa Indonesia Menggunakan Algoritma Textrank*, vol. 6, no. 1, 2022. <https://repository.uin-suska.ac.id/64930/1/Repository%20Faizah%20Fix.pdf>.
- [8] Z. H. Amur, Y. K. Hooi, G. M. Soomro, H. Bhanbhro, S. Karyem, and N. Sohu, "Unlocking the potential of keyword extraction: The need for access to high-quality datasets," *Appl. Sci.*, vol. 13, no. 12, p. 7228, 2023. <https://doi.org/10.3390/app13127228>.
- [9] D. G. P. A. Putra, E. M. Dharma, and K. T. Werthi, "Pembangunan e-learning seni budaya bali berbasis web," *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 187, 2022. <https://doi.org/10.37600/tekinkom.v5i2.531>.

- [10] J. Lee and I. Jung, "Instructional changes instigated by university faculty during the covid-19 pandemic: the effect of individual, course and institutional factors," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, p. 52, 2021. <https://doi.org/10.1186/s41239-021-00286-7>.
- [11] L. Zhang, W. Wang, J. Ma, and Y. Wen, "Iwf-textrank keyword extraction algorithm modelling," *Appl. Sci.*, vol. 14, no. 22, p. 10657, 2024. <https://doi.org/10.3390/app142210657>.
- [12] R. A. Rasyid and D. H. U. Ningsih, "Penerapan algoritma tf-idf dan cosine similarity untuk query pencarian pada dataset destinasi wisata," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 170–178, 2024. <https://doi.org/10.35870/jtik.v8i1.1416>.
- [13] D. A. H. Elneel, H. Kahtan, A. S. Fakharudin, M. Abdulhak, A. S. Al-Ahmad, and Y. I. Alzoubi, "Investigating internet of things impact on e-learning system: An overview," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, 2023. <https://doi.org/10.14569/IJACSA.2023.0140572>.
- [14] E. M. Dharma, H. Prabowo, A. Trisetyarso, and T. Wiguna, "The use of artificial intelligence to predict depression through thermal imaging," p. 040002, 2023. <https://doi.org/10.1063/5.0163192>.
- [15] E. Fuadillah, L. S. Riza, and R. Megasari, "Comprehensive review: Transforming self-education through automatic question generation technology," *Master J. Futur. Educ.*, vol. 2, no. 1, pp. 27–42, 2025. <https://doi.org/10.63461/cadikajournal.v2i1.296>.
- [16] D. T. Aga, R. Chintanippu, R. A. Mowri, and M. Siddula, "Exploring secure and private data aggregation techniques for the internet of things: a comprehensive review," *Discov. Internet Things*, vol. 4, no. 1, p. 28, 2024. <https://doi.org/10.1007/s43926-024-00064-7>.
- [17] T. Alam, "Blockchain and big data-based access control for communication among iot devices in smart cities," *Wirel. Pers. Commun.*, vol. 132, no. 1, pp. 433–456, 2023. <https://doi.org/10.1007/s11277-023-10617-8>.
- [18] R. Setiawan *et al.*, "Iot based virtual e-learning system for sustainable development of smart cities," *J. Grid Comput.*, vol. 20, no. 3, p. 24, 2022. <https://doi.org/10.1007/s10723-022-09616-z>.
- [19] K. W. Trisna, J. Huang, H. Liang, and E. M. Dharma, "Fusion text representations to enhance contextual meaning in sentiment classification," *Appl. Sci.*, vol. 14, no. 22, p. 10420, 2024. <https://doi.org/10.3390/app142210420>.
- [20] S. Alhojely and J. Kalita, "Recent progress on text summarization," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1503–1509, IEEE, 2020. <https://doi.org/10.1109/CSCI51800.2020.00278>.
- [21] S. Zhang, Z. Wang, J. Qi, J. Liu, and Z. Ying, "Accurate assessment via process data," *Psychometrika*, vol. 88, no. 1, pp. 76–97, 2023. <https://doi.org/10.1007/s11336-022-09880-8>.
- [22] M. Song, P. Xu, Y. Feng, H. Liu, and L. Jing, "Mitigating over-generation for unsupervised keyphrase extraction with heterogeneous centrality detection," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, (Stroudsburg, PA, USA), pp. 16349–16359, Association for Computational Linguistics, 2023. <https://doi.org/10.18653/v1/2023.emnlp-main.1017>.
- [23] M. F. Fakhrezi, M. A. Bijaksana, and A. F. Huda, "Implementation of automatic text summarization with textrank method in the development of al-qur'an vocabulary encyclopedia," *Procedia Comput. Sci.*, vol. 179, pp. 391–398, 2021. <https://doi.org/10.1016/j.procs.2021.01.021>.
- [24] J. J. Sihombing, A. Arnita, S. I. A. Idrus, and D. Y. Niska, "Implementation of text summarization on indonesian scientific articles using textrank algorithm with tf-idf web-based," *J. Soft Comput. Explor.*, vol. 5, no. 3, pp. 310–319, 2024. <https://doi.org/10.52465/josce.v5i3.475>.
- [25] R. A. Ma'rifat, I. Suraharta, and I. M. I. Jaya, *Internet of Things*, vol. 2. Bali: CV Rey Media Grafika, 2024.