

AUTOMATIC QUESTION GENERATION (AQG) DARI DOKUMEN TEKS BAHASA INDONESIA BERDASARKAN NON-FACTOID QUESTION

Aminudin¹, Azhari SN², Baaras Ahmad³

^{1,3}Jurusan Teknik Informatika, Universitas Muhammadiyah Malang, Malang, Indonesia

²Departemen Ilmu Komputer dan Elektronika, FMIPA, UGM, Yogyakarta, Indonesia

Email: ¹aminudin2008@umm.ac.id, ²arism.softcmp@gmail.com, ³baaras.ahmad@gmail.com

(Naskah masuk: 26 Februari 2018, diterima untuk diterbitkan: 18 Mei 2018)

Abstrak

Automatic Question Generation (AQG) adalah sistem yang dapat membangkitkan pertanyaan secara otomatis dari teks atau dokumen dengan menggunakan metode atau pola-pola tertentu. Diharapkan sistem AQG yang dikembangkan bekerja seperti halnya manusia membuat pertanyaan setelah diberikan suatu teks. Manusia dapat membuat pertanyaan, dikarenakan manusia dapat memahami teks yang diberikan dan berdasarkan pengetahuan-pengetahuan yang dimilikinya. Untuk mengembangkan sistem AQG penelitian ini, dilakukan kombinasi beberapa metode diantaranya algoritme *Naive Bayes Classifier* untuk mengklasifikasikan kalimat ke dalam jenis kalimat *non-factoid*. Chunking labelling untuk memberikan label pada masing-masing kalimat dari hasil klasifikasi dan pendekatan template untuk mencocokkan hasil kalimat dengan template pertanyaan yang dibuat. Hasil pertanyaan yang dihasilkan oleh sistem akan diukur berdasarkan paramater yang telah ditentukan yang didasarkan atas pengukuran recall, precision dan F-Measure. Dengan adanya sistem AQG ini diharapkan dapat membantu guru mata pelajaran Biologi untuk membuat pertanyaan secara otomatis dan efektif serta efisien.

Kata kunci: *Automatic Question Generation (AQG), Naive Bayes Classifier, Chunking Labelling*

AUTOMATIC QUESTION GENERATION FROM INDONESIAN TEXT DOCUMENT BASED ON NON-FACTOID QUESTION

Abstract

Automatic Question Generation (AQG) is a system can generate question with automatically from text or document by using methods or certain patterns. Expected system AQG developed works like it does humans make create a question after being given a text. Humans can create a question, because humans can understand the given text and based on knowledge assets. To develop the system of AQG in this research, will do a combination of several methods including *Naïve Bayes Clasifier* algorithm to classify the sentence into a kind of non-sentence factoid. Chunking labelling to provide labels on each sentence and template approach to match the right results sentences with question templates created. The results of the question that are generated by the system will be measured based on predetermined parameters required that is based on the measuring precision, recall and F-Measure. With the existence of the AQG system is expected to help teachers of Biology subjects to make the question automatically, effectively and efficiently.

Keywords: *Automatic Question Generation, Naive Bayes Classifier, Chunking Labelling*

1. PENDAHULUAN

Penyediaan soal dalam Proses Belajar Mengajar (PBM) menjadi isu penting karena menyusun soal sebagai bahan evaluasi memerlukan sumber daya baik referensi maupun waktu yang cukup banyak. Pada peristiwa belajar dan mengajar hasil atau prestasi yang diperoleh oleh siswa berbeda-beda walaupun siswa dibimbing oleh guru dengan bahan pelajaran, waktu, tempat, dan metode yang sama. Oleh karena itu, bantuan mengatasi kesulitan belajar dan memperkaya hasil belajar diperlukan untuk membantu proses belajar di luar sekolah. Salah satu bantuan untuk mengatasi masalah ini yaitu dengan

memberikan latihan soal kepada siswa. Akan tetapi, jika pertanyaan yang dibuat banyak dengan mempertimbangkan waktu maka hal tersebut tidaklah mudah, dan salah satu cara dibutuhkannya sistem *Automatic Question Generation (AQG)* yang merupakan salah satu bidang NLP.

Natural Language Processing (NLP) merupakan salah satu area yang paling populer dari kecerdasan buatan dan memiliki berbagai bidang aplikasi seperti *text summarization, machine translation, question answering, Automatic Question Generation (AQG)* (Bednarik dan Kovacs, 2012). *Automatic Question Generation (AQG)* adalah sistem yang dapat membangkitkan pertanyaan berdasarkan

informasi atau teks dokumen dengan menggunakan metode atau algoritme dan pola-pola tertentu. Selain itu, sistem AQG tersebut diharapkan bekerja seperti halnya manusia yang dapat membuat pertanyaan ketika diberikan sebuah teks. Manusia dapat membuat pertanyaan dikarenakan manusia dapat memahami teks yang diberikan dan berdasarkan pengetahuan-pengetahuan yang dimilikinya. AQG merupakan salah satu area pada keilmuan NLP yang terdapat beberapa keilmuan yang serupa misalnya saja *text summarization*, *question answering* dan *intelligent tutoring system*.

Menurut Fattoh dan Haggag (2015) sistem AQG merupakan salah satu komponen penting dalam teknologi pembelajaran seperti *Intelligent Tutoring System* (ITS). Salah satu penerapan sistem AQG dapat digunakan pada salah satu module ITS ialah pada tahap *domain knowledge*, di dalam tahap tersebut sistem AQG dapat digunakan sebagai latihan soal yang dibuat berdasarkan konsep pengetahuan. Pengembangan sistem AQG telah menjadi salah satu isu penelitian yang penting dikarenakan membutuhkan wawasan dari berbagai disiplin ilmu termasuk, *Artificial Intelligence* (AI), NLP, *Machine Learning* (ML), dll.

Salah satu masalah yang paling sulit di dalam mengembangkan sistem AQG adalah menemukan kecocokan antara pertanyaan yang dihasilkan dengan teks atau dokumen yang dimasukkan ke dalam sistem, maka dari itu banyak penelitian yang menggunakan algoritme dan pendekatan baru yang secara aktif menggunakan kalimat dalam dokumen sebagai sumber pertanyaan (Suwarningsih et al, 2014). Salah satu pendekatan yang dilakukan peneliti terdahulu adalah menggunakan pengolahan bahasa alami dengan menggabungkan teknik dan algoritme tertentu yang secara otomatis menghasilkan pertanyaan dari sebuah dokumen teks yang dimasukkan ke dalam sistem (Ali et al, 2010).

Garis besar konseptual sistem AQG menurut Suwarningsih et al. (2014) dibagi menjadi tiga kategori yaitu pendekatan sintaksis, pendekatan semantik dan pendekatan template. Seperti penelitian yang dilakukan oleh Kalady et al (2010); Ali et al (2010) mengusung metode berbasis sintaksis untuk mengembangkan sistem AQG dan penelitian yang dilakukan oleh Mannem et al (2010); Lindberg et al (2013) mengusung metode berbasis semantik untuk mengembangkan sistem AQG. Teknik yang digunakan di dalam kedua pendekatan tersebut adalah NER (*Named Entity Recognition*), parser sintaktik dan SRL (*Semantic Role Labelling*).

Kalady et al (2010) mengembangkan sistem AQG dengan menggunakan pendekatan sintaksis, di dalam penelitian tersebut mengambil sebuah sample kalimat dengan jenis kalimat *factoid question* kemudian dari kalimat tersebut di uraikan menggunakan parser sintaksis dengan memanfaatkan tools *charniak parser*. Hal ini dilakukan untuk menyederhanakan kalimat kompleks, mendeteksi

frase kunci dan menerapkan aturan transformasi sintaksis untuk menghasilkan pertanyaan yang akurat. Sementara itu penelitian yang dilakukan oleh Ali et al (2010) mengembangkan sistem AQG dengan menggunakan pendekatan sintaksis. Penelitian ini menggunakan proses klasifikasi kalimat dengan memanfaatkan NER (*Named Entity Recognition*) untuk mendeteksi jenis kalimat ke dalam lima jenis yaitu manusia, barang, lokasi, waktu dan perhitungan kemudian ketika jenis kalimat tersebut sudah terdeteksi maka akan dilakukan proses pembangkitan kalimat dengan cara mengganti variabel dari jenis kalimat tersebut. Hasil pertanyaan yang di dapatkan dari jenis penelitian ini adalah jenis kalimat *factoid question*.

Untuk mengidentifikasi pendekatan sintaksis tidak selalu cukup. Terkadang pertanyaan sederhana yang menggunakan *what* (apa) dan *who* (siapa) tidak jelas ketika pertanyaan diambil dari konteksnya. Maka diperlukan pendekatan yang lain untuk mengidentifikasi hasil pertanyaan sesuai dengan konteksnya salah satunya pendekatan berbasis semantik. Seperti yang dilakukan oleh Mannem et al (2010) dengan menggabungkan pendekatan semantik dan transformasi sintaksis di dalam membuat sistem AQG. Transformasi sintaksis digunakan untuk mengidentifikasi target potensial pada semua predikat di dalam paragraph dan di masukan ke dalam argument yang telah di tentukan ke dalam konteks SRL dengan memberikan sebuah *argument* pada masing-masing paragraph yang mengandung predikat. Setelah diidentifikasi kalimat kompleks pada kata kerja tertentu maka langkah yang terakhir adalah perankingan di mana perankingan ini memastikan bahwa pertanyaan-pertanyaan yang dihasilkan dari kalimat utama mendapatkan peringkat yang lebih tinggi daripada yang dari anak kalimat.

Berbeda yang dilakukan oleh Mannem et al (2010) di dalam mengembangkan AQG dengan cara mendeteksi kalimat dengan menggunakan SRL, Fattoh (2014) menggunakan metode SRL yang digabungkan dengan NER untuk menghasilkan pertanyaan di dalam sistem AQG. Sistem tersebut disediakan untuk memilih kalimat informatif dan kata kunci untuk ditanyakan berdasarkan *label semantik* dan bernama entitas yang ada dalam kalimat, *distractors* dipilih berdasarkan ukuran kesamaan antara kalimat di dalam *dataset*. Sementara itu pengembangan sistem AQG dengan menggunakan pendekatan berbasis ranking dilakukan oleh Heilman dan Smith (2009) dalam penelitiannya untuk menghasilkan satu himpunan peringkat pertanyaan berdasarkan fakta tentang teks yang diberasal dari artikel. Adapun hasil dari penelitian yang dilakukan oleh penelitian tersebut adalah berupa kerangka kerja AQG.

Menurut Suwarningsih et.al (2014) untuk membuat sistem AQG dengan menggunakan pendekatan sintaksis dan semantik dirasa cukup sulit dikarenakan harus dilakukan dengan menata ulang

permukaan kalimat tetapi kelemahan pada pengembangan sistem AQG berbasis sintaksis dan semantik dapat diperbaiki dengan menggunakan pendekatan template. AQG menggunakan template menawarkan kemampuan untuk mengajukan pertanyaan dengan kata-kata yang tepat dari sumber teks. Template yang dimaksud adalah teks yang telah ditetapkan dengan variabel diganti dengan isi dari teks sumber.

Sistem AQG yang dikembangkan di dalam paper ini menggunakan pendekatan sistem AQG berbasis *template* dan dikombinasikan dengan chunking labelling dan proses klasifikasi menggunakan metode *Naive Bayes Classifier* (NBC) (Ting dan Tsang, 2011) untuk mengklasifikasikan kalimat berdasarkan label kelas *non-factoid* yang terdiri atas kalimat definisi, alasan dan metode. Penelitian ini membangun sebuah sistem AQG yang dapat mengubah teks deskripsi menjadi beberapa pertanyaan yang dikhususkan untuk mata pelajaran Biologi. Hasil dari penelitian ini adalah mengubah dokumen sumber menjadi pertanyaan secara otomatis dengan menggunakan beberapa parameter untuk mencocokkan antara dokumen sumber dengan hasil pertanyaan.

Untuk mengembangkan sistem AQG di dalam paper ini akan diatur sistematika penulisan hal ini digunakan untuk mengukur keberhasilan AQG yang telah dikembangkan bagian (1) pendahuluan yang mengulas permasalahan ketika mengembangkan sistem AQG, kepentingan penelitian AQG, dan solusi yang ditawarkan di dalam penelitian ini, bagian (2) alur pengerjaan atau rancangan sistem AQG dari awal sampai akhir (3) menunjukkan hasil experiment dan beberapa evaluasi dari pengujian sistem, dan bagian (4) kesimpulan yang didapatkan dari hasil pengujian.

2. METODE PENELITIAN

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah kumpulan buku Biologi jenjang SMP dalam bentuk softcopy yang secara bebas di unduh di web <http://www.bse.kemdikbud.go.id>. Buku tersebut dipilih khusus untuk buku MK Biologi dikarenakan data buku tersebut masih menjadi satu dengan IPA terpadu.

2.2 Rancangan Text Preprocessing

Text processing adalah proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan sistem yang akan dibuat. Preprocessing yang dilakukan pada system AQG ini adalah sebagai berikut :

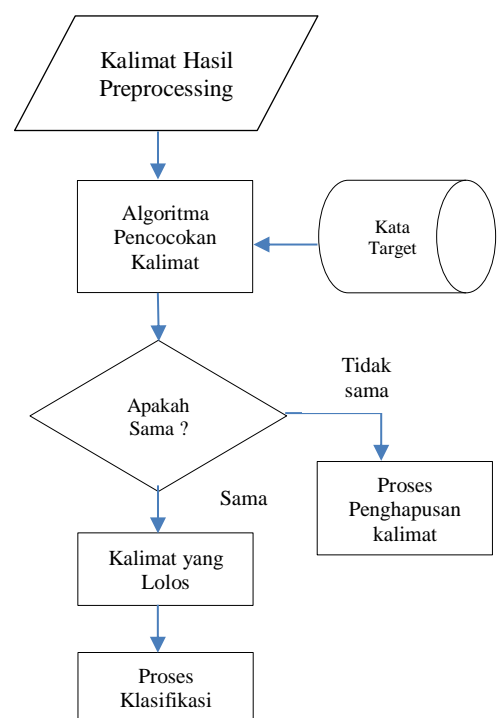
- 1) **Ekstraksi Dokumen** : Proses ini digunakan untuk mengeluarkan isi teks di dalam dokumen untuk direpresentasikan ke dalam input program. Sistem AQG yang akan diinputkan ke dalam sistem adalah sebuah dokumen maka harus dilakukan perubahan ekstraksi dokumen menjadi teks.

- 2) **Case Folding** : Proses ini dilakukan untuk menyamakan case atau teks dalam sebuah dokumen. Oleh karena itu, peran case folding dibutuhkan untuk mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk yang standard.

- 3) **Tokenizing** : Proses ini merupakan pemotongan sebuah dokumen menjadi bagian-bagian tertentu. Tokenizing berfungsi untuk memisah kalimat berdasarkan delimiter yang telah ditentukan.

2.3 Rancangan Eliminasi Kalimat

Proses eliminasi kalimat dilakukan untuk menghapus kalimat yang tidak sesuai dengan kata target yang telah ditentukan, di mana kata target yang dimaksud merupakan kata target yang sudah disesuaikan dengan kata kerja dengan jenis *non-factoid*. Proses eliminasi kalimat pada penelitian ini menggunakan algoritma *string matching* dengan jenis *exact matching*. Cara kerja *exact matching* yaitu memproses semua kemunculan *query* yang selanjutnya disebut *pattern* ke dalam *string* yang lebih panjang (teks). *Pattern* dalam hal ini kata target yang di tanam di dalam tabel di database dan panjang (teks) adalah teks deskripsi yang nantinya sudah di pisahkan berdasarkan kalimat. *Pattern* dan panjang teks akan disamakan, apabila sama dengan kata target maka dapat dikatakan kalimat tersebut adalah mengandung kata target dan jika tidak sama maka kalimat tersebut dihapus. Gambar 1 menunjukkan alur diagram proses eliminasi kalimat.

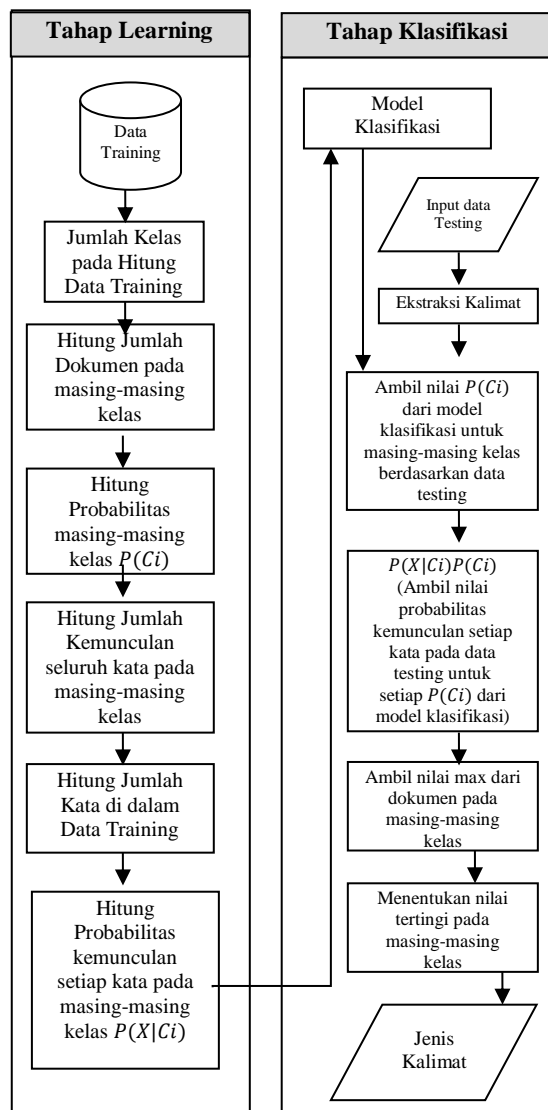


Gambar 1. Alur Proses Eliminasi Kalimat

2.4 Rancangan Naive Bayes Classifier (NBC)

Metode NBC menempuh dua tahap dalam proses klasifikasi kalimat yaitu tahap *learning* dan tahap

klasifikasi. Pada tahap *learning* dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary* yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin merepresentasikan dokumen, pada tahap *learning* terdapat dokumen *training* yang menjadi acuan untuk proses testing pada tahap klasifikasi. Gambar 2 ditunjukkan rancangan proses klasifikasi menggunakan NBC.



Gambar 2. Diagram Proses Klasifikasi dengan Naive Bayes Classifier

2.5 Rancangan Pembangkitan Pertanyaan

Hasil kalimat yang sudah diklasifikasikan menggunakan NBC, langkah selanjutnya yaitu kalimat tersebut akan dibangkitkan pertanyaan dengan cara mendeteksi kata target dan memberikan label kata atau frase pada kalimat tersebut. Kata target yang dimaksud di sini adalah sebuah kata penghubung atau konjungsi sebuah kalimat yang telah ditentukan. Pemberian label kata atau frase pada kalimat akan dicocokkan dengan template pertanyaan berdasarkan kelas yang telah ditentukan.

Tabel 1. Rancangan Pembangkitan Pertanyaan

| No. | Jenis Kalimat | Kata Target | Template Pertanyaan | Keterangan |
|-----|---------------|------------------------------|-----------------------------|--|
| 1. | Definisi | dikenal, dinamakan, disebut | Apa yang dimaksud + VP/NP ? | Pembangkitan pertanyaan setelah kata target |
| 2. | Definisi | adalah, diartikan, merupakan | Apa yang dimaksud + VP/NP ? | Pembangkitan pertanyaan sebelum kata target |
| 3. | Alasan | karena | Mengapa + a? | Pembangkitan pertanyaan sebelum kata target (a karena b) |
| 4. | Alasan | Oleh karena itu | Mengapa + b ? | Pembangkitan pertanyaan sesudah kata target (a karena b) |
| 5. | Metode | Dengan cara | Bagaimana cara + a? | Pembangkitan pertanyaan sebelum kata target (a dengan cara b) |

Tabel 1 menunjukkan bahwa sebelum kalimat di bangkitkan pertanyaan, ada beberapa proses yang menyertainya diantaranya adalah pemberian label pada masing-masing frase atau kata dari kalimat yang sudah diklasifikasikan menggunakan algoritme NBC. Apabila kalimat sudah diklasifikasikan ke dalam kalimat definisi, dan di dalam kalimat definisi tersebut mengandung kata target disebut, dinamakan, dan dikenal maka target kata untuk pertanyaan berada setelah kalimat penghubung. Apabila di dalam kalimat definisi mengandung kata penghubung adalah, merupakan, dan diartikan maka target kata untuk pertanyaan berada sebelum kalimat penghubung tersebut. Proses pembangkitan pertanyaan menggunakan label NP sehingga pembangkitan pertanyaan apa yang dimaksud dengan + NP.

Apabila hasil klasifikasi termasuk jenis kalimat alasan dan di dalam kalimat tersebut mengandung kata karena dan oleh karena itu maka ada aturan tertentu untuk proses pembangkitan pertanyaan. Untuk membuat pertanyaan dengan jenis alasan sebelum kata target di inialisasi sebagai a dan sesudah kata target di inialisasi dengan b. Pada penelitian ini untuk proses pembangkitan pertanyaan pada kalimat alasan dilakukan dengan kata target karena menggunakan rumus a karena b sehingga proses pembangkitan pertanyaan kenapa+a, kemudian jika kalimat alasan dengan kata target oleh karena itu proses pembangkitan pertanyaan menggunakan rumus kenapa+b (Effenberger, 2015). Kalimat metode dalam penelitian ini dideteksi menggunakan kata target dengan cara, sama seperti

kalimat proses pembentukan pertanyaan pada kalimat alasan diatas, pada kalimat metode proses pembentukan pertanyaan menggunakan rumus a dengan cara b maka aturan pembangkitan pertanyaan bagaimana cara + a?. Contoh pembangkitan pertanyaan ditunjukkan pada Tabel 2 sebagai berikut.

Tabel 2. Rancangan Pembangkitan Pertanyaan

| Data Testing | Deteksi Kata Target | Hasil Pertanyaan |
|---|---------------------|---|
| c₁ : Bernapas/VP <u>adalah</u> memasukkan oksigen dari udara yang akan digunakan untuk mengoksidasi makanan serta mengeluarkan sisa hasil oksidasi. | adalah | Apa yang dimaksud dengan bernapas? |
| c₂ : Menghirup udara pernapasan melalui rongga hidung tindakan yang tepat [a] <u>karena</u> di dalam rongga hidung terdapat banyak rambut yang berfungsi sebagai penyaring udara [b]. | karena | Mengapa menghirup udara pernapasan melalui rongga hidung tindakan yang tepat? |
| c₃ : Pernapasan pada manusia berlangsung [a] <u>dengan cara</u> mengubah tekanan udara di dalam paru-paru. | Dengan cara | Bagaimana cara pernapasan pada manusia berlangsung ? |

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Klasifikasi

Untuk menghasilkan pertanyaan yang sesuai dengan kriteria sangat ditentukan dengan hasil klasifikasi. Maka, sebelum dilakukan pengujian pembangkitan pertanyaan, dilakukan terlebih dahulu pengujian klasifikasi dengan data 510 kalimat yang diambil dari buku Biologi jenjang SMP.

Tabel 3. Rancangan Pembangkitan Pertanyaan

| k | Akurasi (%) | Precision (%) | Recall (%) | F-Measures (%) | Error rate (%) |
|-----------------|-------------|---------------|------------|----------------|----------------|
| k ₁ | 96.67 | 96.97 | 96.97 | 96.65 | 3.33 |
| k ₂ | 93.33 | 93.33 | 94.44 | 93.26 | 6.66 |
| k ₃ | 100 | 100 | 100 | 100 | 0 |
| k ₄ | 100 | 100 | 100 | 100 | 0 |
| k ₅ | 93.33 | 93.33 | 94.44 | 93.26 | 6.66 |
| k ₆ | 80 | 80 | 84.39 | 80.52 | 20 |
| k ₇ | 83.33 | 83.33 | 88.89 | 83.74 | 16.67 |
| k ₈ | 93.33 | 93.33 | 93.64 | 93.32 | 6.66 |
| k ₉ | 86.66 | 89.74 | 88.88 | 87.87 | 13.33 |
| k ₁₀ | 83.33 | 83.33 | 88.88 | 82.22 | 16.66 |
| k ₁₁ | 90 | 90 | 91.41 | 89.5 | 10 |
| k ₁₂ | 93.33 | 93.33 | 94.44 | 93.26 | 6.66 |
| k ₁₃ | 100 | 100 | 100 | 100 | 0 |
| k ₁₄ | 93.33 | 93.33 | 94.44 | 93.46 | 6.66 |
| k ₁₅ | 93.33 | 93.93 | 93.33 | 93.12 | 6.66 |
| k ₁₆ | 93.33 | 93.94 | 93.33 | 93.12 | 6.66 |
| k ₁₇ | 90 | 90 | 91.41 | 89.5 | 10 |
| Avg | 91.95 | 92.22 | 93.46 | 91.92 | 8.03 |
| StdDev | 5.90 | 5.80 | 4.31 | 5.86 | 5.89 |

Tabel 3 menunjukkan data hasil pengujian klasifikasi menggunakan *k-fold cross validation* dengan membagi k sebanyak 17. Ada 3 (tiga) k yang akurasi mencapai 100% hal ini dikarenakan adanya penyamarataan antara data testing dan data

training pada masing-masing kelas sehingga hasil akurasi mencapai 100%.

3.2 Pengujian Pembangkitan Pertanyaan

Pengujian pembangkitan pertanyaan ini dilakukan untuk mengukur tingkat ketepatan hasil pertanyaan yang dihasilkan oleh sistem AQG dengan dokumen yang dimasukan kedalam sistem dari buku teks Biologi dengan kriteria yang telah di tentukan. Kriteria dalam pengujian ini adalah sebagai berikut :

- Pertanyaan yang di bangkitkan tepat maksudnya antara pertanyaan yang dihasilkan dengan teks yang diinputkan sesuai.
- Maksud dari pertanyaan yang dibangkitkan sistem dapat dimengerti.
- Pertanyaan yang dibangkitkan sistem sesuai dengan konteks.
- Pertanyaan dihasilkan mempunyai jawaban.

Skenario pengujian ini yaitu membagi dokumen buku Biologi menjadi beberapa bab pada masing-masing kelas untuk dijadikan sebagai data testing kemudian masing-masing bab itu diinputkan ke dalam sistem, hasil pertanyaan dari sistem akan di evaluasi berdasarkan dari kriteria yang telah di tentukan. Hasil dari pertanyaan akan dilakukan pengukuran menggunakan *recall*, *precision* dan *F-measure* (Ali et al, 2010).

$$Recall = Q_a \cap Q_c \quad (1)$$

$$Precision = Q_a \cap Q_b \quad (2)$$

$$F - Measure = 2 (recall * precision) / (recall + precision) \quad (3)$$

Beberapa simbol yang digunakan dalam pengujian pembangkitan pertanyaan sebagai berikut :

- Simbol A adalah jumlah kalimat yang berhasil di ekstraksi.
- B adalah jumlah kalimat yang teridentifikasi benar.
- C adalah jumlah teridentifikasi tidak benar.
- Q_a adalah jumlah pertanyaan yang benar.
- Q_b adalah jumlah pertanyaan yang dihasilkan oleh sistem baik benar ataupun salah.
- Q_c adalah jumlah pertanyaan yang dibuat oleh orang.
- D adalah *recall* maksudnya selisih pertanyaan yang dihasilkan oleh sistem dengan pertanyaan yang dibuat oleh orang.
- E adalah *precision* maksudnya presisi jumlah pertanyaan dari sistem yang benar dengan jumlah pertanyaan yang dihasilkan oleh sistem baik benar ataupun salah.
- F adalah *F-measure* maksudnya mengukur keberhasilan recall dan precision.

Tabel 4. Pengujian Hasil Pertanyaan Kelas VII

| Bab Ke - | Kalimat | | | Pertanyaan | | | Pengukuran (%) | | |
|-----------|---------|----|---|----------------|----------------|----------------|----------------|-----|----|
| | A | B | C | Q _a | Q _b | Q _c | D | E | F |
| I | 9 | 6 | 3 | 6 | 6 | 9 | 66 | 100 | 80 |
| II | 44 | 40 | 4 | 27 | 37 | 35 | 82 | 57 | 67 |
| III | 22 | 22 | 0 | 19 | 23 | 20 | 95 | 70 | 81 |
| IV | 36 | 31 | 5 | 20 | 28 | 27 | 80 | 55 | 65 |
| IV | 48 | 41 | 7 | 27 | 40 | 29 | 95 | 50 | 66 |
| Rata-Rata | | | | | | | 83 | 66 | 72 |

Dari Tabel 4. Menunjukkan data buku Biologi kelas VII SMP bahwa perbandingan antara data yang berhasil diekstraksi oleh sistem dengan hasil pertanyaan yang benar menurut parameter yang telah ditentukan, hasil pertanyaan yang dihasilkan oleh sistem dengan benar masih tinggi. Sehingga dapat dikatakan sistem AQG yang dibangun masih memenuhi standard untuk pengubahan dokumen menjadi pertanyaan. Hal tersebut dapat juga dilihat pengukuran recall dan F-Measure juga masih terbilang bagus. Kemudian data hasil pengujian pertanyaan untuk kelas VIII ditunjukkan pada Tabel 5 dibawah ini.

Tabel 5. Pengujian Hasil Pertanyaan Kelas VIII

| Bab Ke- | Kalimat | | | Pertanyaan | | | Pengukuran (%) | | |
|-----------|---------|----|----|----------------|----------------|----------------|----------------|----|----|
| | A | B | C | Q _a | Q _b | Q _c | D | E | F |
| I | 44 | 34 | 10 | 27 | 34 | 31 | 87 | 65 | 75 |
| II | 49 | 45 | 4 | 38 | 46 | 44 | 86 | 70 | 77 |
| III | 31 | 27 | 4 | 18 | 28 | 22 | 81 | 47 | 60 |
| IV | 19 | 19 | 0 | 19 | 21 | 19 | 100 | 82 | 90 |
| V | 46 | 44 | 2 | 37 | 45 | 37 | 100 | 70 | 82 |
| VI | 47 | 44 | 3 | 33 | 44 | 35 | 94 | 60 | 73 |
| Rata-rata | | | | | | | 91 | 66 | 76 |

Tabel 5. menunjukan hasil pengukuran dari pembangkitan pertanyaan buku teks Biologi SMP untuk kelas VIII misalnya 91% *recall*, 66% *precision* dan 76% *nilai F-measure*. Angka angka tersebut berarti *recall*-nya dikatakan sangat bagus dikarenakan selisih antara pertanyaan yang dihasilkan oleh sistem dengan yang dibuat oleh manusia sangat sedikit sehingga *recall*-nya sangat tinggi, *precision* sedikit kurang baik karena kemungkinan ada kalimat yang tidak beraturan dan terjadi kesalahan pada proses klasifikasi, untuk nilai *F-measure*-nya juga cukup bagus

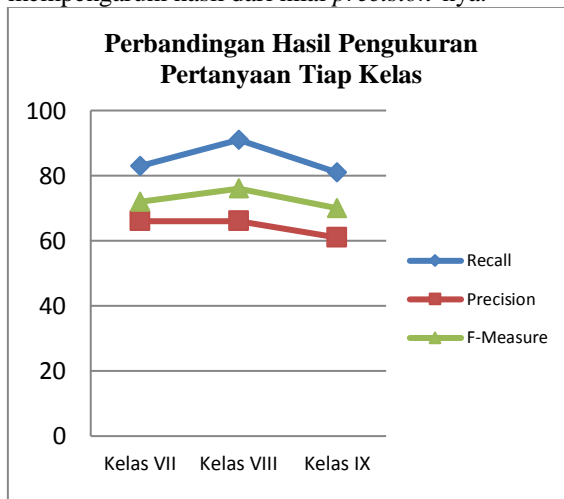
Tabel 6. Pengujian Hasil Pertanyaan Kelas IX

| Bab ke- | Kalimat | | | Pertanyaan | | | Pengukuran (%) | | |
|-----------|---------|----|----|----------------|----------------|----------------|----------------|-----|----|
| | A | B | C | Q _a | Q _b | Q _c | D | E | F |
| I | 21 | 15 | 6 | 9 | 19 | 15 | 60 | 31 | 40 |
| II | 34 | 28 | 6 | 30 | 30 | 31 | 96 | 100 | 98 |
| III | 41 | 33 | 8 | 31 | 36 | 30 | 100 | 75 | 86 |
| IV | 34 | 19 | 15 | 17 | 26 | 16 | 100 | 48 | 65 |
| V | 21 | 20 | 1 | 13 | 20 | 14 | 92 | 48 | 63 |
| VI | 26 | 16 | 10 | 13 | 17 | 16 | 81 | 61 | 70 |
| Rata-rata | | | | | | | 88 | 60 | 70 |

Berdasarkan Tabel 6. menunjukan hasil pengukuran dari pembangkitan pertanyaan buku teks Biologi SMP untuk kelas IX misalnya 88% *recall*, 60% *precision* dan 70% *nilai F-measure*. Angka angka tersebut berarti *recall*-nya dikatakan sangat bagus dikarenakan selisih antara pertanyaan yang dihasilkan oleh sistem dengan yang dibuat oleh manusia sangat sedikit, *precision* sedikit kurang baik karena kemungkinan ada kalimat yang tidak beraturan dan terjadi kesalahan pada proses klasifikasi serta ada beberapa kalimat yang salah ketika proses labelling, untuk nilai *F-measure*-nya juga cukup bagus.

Berdasarkan dari perbandingan pengujian pembangkitan pertanyaan yang ditunjukkan pada Gambar 3. didapatkan bahwa nilai nilai *recall* memiliki nilai tertinggi dikarenakan selisih antara

pertanyaan yang dihasilkan oleh sistem dengan kriteria pertanyaan yang benar yang dibuat oleh user selisihnya tidak terlalu besar sehingga nilai *recall*-nya tinggi. Sebaliknya nilai *precision*-nya rendah dikarenakan selisih antara pertanyaan yang dihasilkan oleh sistem baik itu pertanyaan benar atau salah dengan pertanyaan yang sesuai dengan kriteria selisihnya cukup banyak sehingga sangat mempengaruhi hasil dari nilai *precision*-nya.



Gambar 3. Hasil Perbandingan Pengukuran Pembangkitan Pertanyaan

3.3 Analisa dan Penjelasan Hasil Pertanyaan yang dihasilkan oleh sistem

Ada beberapa jenis pertanyaan yang dihasilkan oleh sistem, tetapi jenis pertanyaan tersebut setelah ditela'ah lebih dalam pertanyaan yang dihasilkan oleh sistem tidak ada di dalam kriteria yang telah ditentukan diatas (point 4.2), adapun jenis pertanyaan tersebut apabila pertanyaan tersebut antara lain :

- 1) Pertanyaan yang mengandung kata ini, tersebut. Contoh "batu tersebut merupakan senyawa kalsium dan penumpukan asam syaraf" dari pertanyaan tersebut sistem akan menghasilkan pertanyaan "Apa yang dimaksud dengan batu tersebut". Sistem menghasilkan pertanyaan tersebut dikarenakan keberagaman teks yang ada dan ketika ditela'ah kembali ternyata pertanyaan tersebut tidak menunjukan salah satu kriteria yang telah disediakan yaitu pertanyaan tersebut tidak mengandung sebuah jawaban.
- 2) Pertanyaan yang dihasilkan kosong. Contoh "Apa yang dimaksud dengan ?" Sistem menghasilkan pertanyaan kosong seperti itu dikarenakan kesalahan pada proses klasifikasi. Pada proses klasifikasi apabila terjadi kesalahan maka kata target dan template yang disediakan tidak ditemukan sehingga pertanyaan yang dihasilkan kosong. Kalau hal ini terjadi maka pertanyaan kosong tersebut akan masuk ke dalam proses pemfilteran dan menjadi pertanyaan dengan jenis pertanyaan salah

- 3) Pertanyaan di dalam kalimat tersebut ada kata yang berulang. Contoh "Apa yang dimaksud dengan kulit kulit ?". Biasanya kalimat itu muncul pada permulaan bab atau sub bab baru dan kemudian setelah bab itu langsung menerangkan pengertian atau definisi dari bab atau sub bab tersebut proses klasifikasi.
4. Pertanyaan ketika ada kalimat yang di dalamnya terdapat kata disebabkan, timbul atau ditimbulkan maka kata tersebut diganti dengan terjadi. Contoh kalimat sumber "Lordosis bisa disebabkan karena perut yang terlalu besar" setelah dilakukan beberapa pemrosesan maka sistem AQG dan menganut beberapa aturan pembentukan pertanyaan maka sistem AQG menampilkan "Mengapa lordosis bisa disebabkan ?" karena contoh tersebut tidak mempunyai makna dan pertanyaan tersebut terkadang tidak memiliki jawaban maka kata 'disebabkan' diganti dengan 'terjadi', jadi pertanyaan berdasarkan dari sumber teks tersebut menjadi "Mengapa lordosis bisa terjadi?".

4. KESIMPULAN

Hasil pengujian dari penelitian ini menunjukan bahwa algoritma Naive Bayes Classifier dan pendekatan template serta chunking labelling yang telah digunakan di dalam membangun sistem AQG dapat berhasil mengubah secara otomatis teks deskripsi menjadi pertanyaan dengan jenis *non-factoid question* pada buku Biologi jenjang SMP. di mana masing-masing kelas mempunyai hasil pertanyaan yang bervariasi berdasarkan dari pengukuran precision, recall dan f-measures.

5. Daftar Pustaka

- AJMERA, J 2014, *Automatic Generation of Question Answer Pairs From Noisy Case Logs*, *IEEE 30th International Conference*.
- Ali, CHALI dan HASAN., 2010, Automatic Question Generation from Sentences, pp. 19–23, *Proc. TALN 2010*.
- AL-YAHYA, M., 2011., A Question Generation Engine for Educational Assesment Based on Domain Ontologies, *11th IEEE International Conference on Advanced Learning Technologies*, pp. 393-395.
- AMINUDIN dan ALWI., 2018, *Analisa Multithreading pada Sistem Rekomendasi Menggunakan Metode Collaborative Filtering dengan Apache Mahout*, *Techno*, vol. 17. No. 1, pp 1-11.
- BEDNARIK. L & KOVACS. L 2012, *Automated EA-type Question Generation from Annotated Texts*, *7th IEEE International Symposium on Applied (computational Intelligence and Informatics)*, pp. 191-195.
- DAS, R, ELIKKOTIL, A & KALADY, S 2010, *Natural Language Question Generation Using Syntax and Keywords*, *Proceedings of the Third Workshop on Question Generation*, Pittsburgh
- EFFENBERGER, T., 2015, "Automatic Question Generation and Adaptive Practice,".
- FATTOH, A. E. ABOUTABL dan HAGGAG., 2015, Semantic Question Generation Using Artificial Immunity, vol. 7, no. 1, pp. 1–8, *Int. J. Mod. Educ. Comput. Sci*.
- FACHRURROZI. M., 2013, *Sistem Pembangkit Pertanyaan Otomatis Dengan Metode Template-Based*.
- GARG, S & GOYAL, V 2010, *System for Generating Questions Automatically From Given Punjabi Text*, *Punjabi University*, p. 324.
- HAN, J dan KAMBER, M 2006, *Data Mining: Concepts and Techniques*, 2nd edn, Morgan Kauffman Publisher, San Francisco.
- HEILMAN, M 2011, *Automatic Factual Question Generation from Text*, *Carnegie Mellon University*, Pittsburgh.
- HEILMAN, M & SMITH, AN 2009, *Question Generation via Overgenerating Transformations and Ranking*, *Carnegie Mellon University*, Pittsburgh.
- KALADY. S, ELIKKOTIL. A dan DAS. R., 2010, *Natural Language Question Generation Using Syntax and Keywords*, p. 91, *Tenth Int. Conf. Intell. Tutoring Syst.*,
- P. MANNEM. P, PRASAD. P dan JHOSI. A., 2010 *Question Generation from Paragraphs at UPenn : QGSTEC System Description*, pp. 84–91 *Proc. QG Third Work. Quest. Gener. Pittsburg, PA.*,
- WIJOYO. S, HERLAMBANG. A, ROZI. F, ISANTA. S., 2017, *Optimasi Suffix Tree Clustering dengan Wordnet dan Named Entity Recognition untuk Pengelompokan Documen*, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 4, No. 4, pp, 263-267
- SUWARNINGSIH.W, SUPRIANA, A. dan PURWARIANTI., 2014, *Tantangan dan Peluang pada Question Generation*, vol. 6, no. 2, pp. 719–730, *J. Sist. Inf*.
- TING, S.L. dan TSANG A. H. C., 2011 "Is Naïve bayes a good classifier for document classification?", vol. 5, no. 3, pp. 37–46 *Int. J. Softw. Eng. its Appl*.
- TAN.P-N, STENIBACH, M dan KUMAR V., 2006, *Introduction to Data Mining*, Addison-Wesley Companion Book Site , United States.

Halaman ini sengaja dikosongkan