MEDIA INFORMATIKA BUDIDARMA, Vol 2, No 4, Oktober 2018 ISSN 2614-5278 (media cetak) ISSN 2548-8368 (media online) Hal 140-149



Penerapan Question Answering System Pada Pembahasan Agama Islam Dengan Pendekatan Metode Pattern Based

Ramadhana Rosyadi, Said Al-Faraby, Adiwijaya

Fakultas Informatika, Universitas Telkom, Bandung, Indonesia Email: ¹ramadrsyd@students.telkomuniversity.ac.id, ²saidalfaraby@telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id

Abstrak

Agama Islam terdapat 25 nabi sebagai pedoman hidup manusia, dokumen yang berisikan informasi mengenai kisah - kisah kehidupan para nabi semasa hidupnya. Penelitian ini bertujuan untuk membangun sistem Tanya jawab yang lebih spesifik dengan menghasilkan jawaban yang relavan tidak dalam bentuk kumpulan dokumen. Question Answering System mampu mengatasi permasalahan dalam sistem Tanya jawab, sistem pencarian informasi dimana jawaban yang dikeluarkan menjadi tepat dengan tanggapan atas permintaan yang disampaikan, tidak dalam bentuk kumpulan dokumen yang mungkin berisi jawaban. Penelitian ini menggunakan metode Pattern Based sebagai mengekstrak potongan kalimat yang menjadi jawaban untuk menemukan jawaban yang sesuai dengan pola yang telah dibuat. Pemilihan dataset menyebabkan sejumlah pertanyaan yang dapat diajukan menjadi terbatas pada informasi yang tersimpan dalam data itu sendiri. Selain itu pertanyaan juga dibatasi berupa kata Tanya yang bersifat Factoid, yaitu Siapa, kapan, dimana, apa dan berapa. Hasil Accuracy yang didapatkan dengan menggunakan metode Pattern Based pada Question Answering System sebesar 39,36%.

Kata kunci: Question Answering System, Pattern Based, Accuracy, 25 Nabi.

Abstract

Islam has 25 prophets as guidelines for human life, documents containing information about the stories of the lives of the prophets during their lifetime. This study aims to build a more specific question and answer system by generating relevant answers not in the form of documents. Question Answering System is able to overcome problems in the Question and answer system, information retrieval systems where the answers issued are correct with responses to requests submitted, not in the form of documents that may contain answers. This study uses the Pattern Based method as extracting sentence pieces which are the answers to find answers that match the patterns that have been made. The selection of datasets causes a number of questions that can be submitted to be limited to information stored in the data itself. Besides that, questions are also limited in the form of Question words that are Factoid, namely Who, when, where, what and how. Accuracy results obtained using the Pattern Based method on Question Answering System are 39.36%.

Keywords: Question Answering System, Pattern Based, Accuracy, And 25 Prophets

1. PENDAHULUAN

Peningkatan pesat dalam penyimpanan pada teknologi informasi dengan menggunakan media web dan sebagainya, yang memungkinkan pada peneliti atau pengguna dapat menyimpan sebuah informasi agar dapat tersedia untuk pengguna umum. Namun, akibat besarnya jumlah data dan informasi yang berdampak dalam pencarian sebuah informasi yang tepat. Karena kesulitan ini banyak pengembang membuat suatu sistem yang dapat menjawab suatu pertanyaan dalam Bahasa alami dan mengembalikan jawaban yang tepat sesuai dengan pertanyaan yang di sampaikan, bukan lagi berupa kumpulan jawaban yang di anggap relevan.

Question Answering System merupakan solusi yang baik dalam menyelesaikan masalah tersebut dengan mengajukan sebuah pertanyaan dan mengembalikan respon jawaban yang relevan dengan mengembalikan dalam bentuk daftar dokumen teks singkat atau frase sebagai jawaban tanpa harus disaring lagi oleh pengguna untuk menentukan dokumen yang mengandung jawaban yang tepat. Question Answering System merupakan pertanyaan dalam bahasa natural, yaitu bahasa yang digunakan dalam percakapan sehari-hari.

Pattern Based Approach merupakan salah satu metode yang dapat di implementasikan pada Question Answering System ini. Metode ini memanfaatkan pola kalimat yang telah ditentukan maupun pola jawaban dalam proses menemukan jawaban yang relevan. Metode ini dipilih karena mampu menghasilkan tingkat Accuracy jawaban yang baik pada Question Answering System berbahasa Indonesia [6]. Maka dari itu dalam tugas akhir ini akan mengimplementasikan Question Answering System ini dengan menggunakan Pattern Based sebagai metode yang digunakan dengan pembahasan seputar tentang agama islam yaitu 25 nabi - nabi dengan menggunakan berbahasa Indonesia.

Rumusan masalah dalam penelitian ini adalah Bagaimana mengimplementasikan *Question Answering System* dengan topik pembahasan seputar agama islam yaitu 25 nabi - nabi dengan berbahasa Indonesia, dan bagaimana pengaruh yang didapat dengan struktur dan pengembangan pola kalimat terhadap performansi *Question Answering System*.

Terdapat batasan masalah dalam penelitian ini, yaitu pertanyaan telah disediakan dan sistem menanyakan pertanyaan seputar agama islam yaitu 25 nabi - nabi dengan berbahasa Indonesia. Pertanyaan yang dapat diproses hanya pertanyaan yang bersifat *Factoid*, seperti siapa, dimana, apa, berapa, dan kapan. Dilakukan penghilangan *stopwords* pada *preprocessing dataset*, dan menggunakan bantuan *tools* dalam penelitian ini.



Tujuan dari pembuatan *Question Answering System* pada penelitian ini yaitu mampu mengimplementasikan *Question Answering System* dengan pembahasan seputar agama islam yaitu 25 nabi - nabi dengan berbahasa Indonesia, dan melakukan analisis terkait pengaruh yang didapat dengan struktur dan pengembangan pola kalimat terhadap performansi *Question Answering System*.

Pada jurnal ini, terdapat studi terkait yang menjelaskan mengenai metode penelitian sebelumnya yang berkaitan dengan metode penelitian yang diterapkan pada penelitian ini. Pada bagian ketiga, terdapat penjelasan mengenai sistem *Question Answering System* yang dibangun dan di implementasikan pada penelitian ini yaitu pada dokumen 25 Nabi pada Agama Islam dalam bahasa Indonesia dengan melewati proses *preprocessing dataset*, dan penggunaan Metode *Pattern Based*. Pada bagian ke empat, terdapat evaluasi hasil skenario pengujian dan analisis hasil pengujian yang telah didapatkan. Pada bagian terkahir, terdapat kesimpulan dan saran dari penelitian yang telah dilakukan.

2. TEORITIS

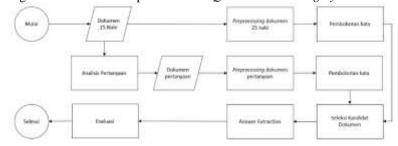
Gunawan dan Gita Lovina (2006) membuat *Question Answering System* dan menerapkannya pada alkitab. Kesimpulan yang berhasil diperoleh dari penilitian ini mengenai QA yaitu QA system merupakan bagian dari *information retrieval*. Sistem ini dapat dipandang sebagai bentuk pengembangan terhadap kemampuan yang dimiliki oleh *search engine* dengan mengembalikan respon jawaban terhadap kueri yang berupa pertanyaan *natural language*. QA system dapat dikembangkan pada domain yang beranekaragam, tergantung dari tujuan pengembangan sistem. Domain yang dimasuk disini erat kaitannya dengan sumber informasi yang akan digunakan untuk menjawab pertanyaan, yang secara langsung akan membatasi jenis pertanyaan sesuai dengan dataset yang dapat dipertanyakan. Teknik yang digunakan dalam mengembangkan sebuah QA system dapat bervariasi mulai dari teknik yang paling sederhana seperti halnya pencocokan pola, sampai dengan teknik – teknik lain yang lebih kompleks, tergantung dari keterbatasan dan kebutuhan sistem. Tahapan proses yang dijalankan oleh sebuah aplikasi QA system tidak akan terlepas dari tiga tahapan proses utama yang menyusun arsitektur umum QA system, yaitu analisis pertanyaan, memilih kandidat dokumen atau segmen dokumen, dan ekstraksi jawaban [1].

Kartina (2010) membuat analisis pertanyaan berBahasa Indonesia pada *Question Answering System*. Hasil dari penelitian yang telah beliau lakukan adalah pembentukan frase pada tahap analisis pertanyaan dapat meningkatkan relevansi dokumen yang di temu kembalikan, dan semakin tepat *Top passage* yang diperoleh, maka semakin tepat jawaban yang dikembalikan oleh sistem QA. Oleh karena itu jika dokumen yang relevan tetapi dalam metode pembobotan *passage* belum tepat, maka berakibat pada jawaban yang diperoleh belum tepat [10].

Hapnes Toba dan Mirna Adriani (2009) membuat *Question Answering System* bahasa Indonesia dengan menggunakan metode *Pattern Based*. Memperoleh hasil dari uji coba dengan menggunakan pendekatan pola pertanyaan yang hasilnya menunjukan bahwa pada QA sistem sangat menjanjikan jika diimplementasi pada pertanyaan seputar dokumen berbahasa Indonesia, akan tetapi ada kelemahan utama pada pendekatan pola pembelajaran yaitu pola pertanyaan yang diperlukan harus dikembangkan secara spesifik dan fase ekstraksi jawaban yang besar [6].

3. ANALISA DAN PEMBAHASAN

Berikut ini merupakan gambaran umum dari proses sistem Question Answering System:



Gambar 1. Diagram Alur Sistem

3.1 Dokumen 25 nabi

Pada penelitian ini, peneliti menggunakan *Dataset* yang berisikan tentang 25 Nabi pada Agama Islam yang menceritakan kisah-kisah nabi. Sumber dataset yang digunakan yaitu dari Wikipedia. Terdapat 25 dokumen dari setiap nabi didalamnya, yang merupakan penjelasan dari kisah teladan nabi, dengan total kalimat sebanyak 34,702. Dataset yang ada akan di proses lebih lanjut untuk kebutuhan penelitian.



Tabel 1. Dokumen NabiDokumenJumlahNabiKalimatNabi Adam2018Nabi Idris509Nabi Nuh1295

Total Kalimat 34702

3.2 Analisi Pertanyaan

Pertanyaan yang dibuat berdasarkan sumber Wikipedia dengan pembahasan 25 nabi dalam ajaran agama Islam dengan Bahasa Indonesia. Pertanyaan menghasilkan target, konteks dan properti. Pola pertanyaan dan pola jawaban yang saling berkaitan. Target jawaban berdasarkan kata tanya seperti, siapa (*Person*), kapan (*Time*), dimana (*Location*), apa (*Object*) dan berapa (*Count*).

Tabel 2. Analisis Pertanyaan dan Dokumen Pertanyaan

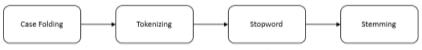
Kata Tanya	Pertanyaan	Jawaban	Pola Pertanyaan	Pola Jawaban dan Target Named Entity Recognition
Siapa	Siapa nabi pertama ?	nabi pertama adalah nabi adam	(Siapa) <c><t></t></c>	(PERSON) <c><t>Adalah<p></p></t></c>
Kapan	kapan Hawa di lahirkan ?	hawa di lahirkan pada tahun 3890 SM	(Kapan) <c> <t>Hawa lahir</t></c>	(TIME) <c><t>Tahun<p></p></t></c>
Dimana	Dimana Hawa di turunkan ?	Hawa diturunkan di Arabia	(Dimana) <c><t>Diturunkan di bumi</t></c>	(LOCATION) <c><t>di<p></p></t></c>
Apa	Apa tujuan Nabi Adam di turunkan ?	tujuan nabi adam di turunkan adalah sebagai khalifah (pemimpin) di muka bumi	(Apa) <c><t>Di Turunkan</t></c>	(OBJECT) <c><t>Adalah<p></p></t></c>
Berapa	Berapa tinggi badan Nabi Adam ?	tinggi badan Nabi Adam Adalah 27,432 meter	(Berapa)Tinggi Badan <c><t></t></c>	(COUNT) <c><t>Adalah<p></p></t></c>
	•••••			

3.3 Dokumen Pertanyaan

Dokumen pertanyaan berisikan kumpulan pertanyaan yang telah di analisis pada tahap sebelumnya, diperoleh 282 pertanyaan bersifat *factoid* yaitu pertanyaan yang mempunyai jawaban yang singkat atau terdiri dari beberapa kata. Penjelasan dalam tabel berada pada lampiran.

3.4 Preprocessing dataset

Proses *preprocessing* yaitu proses pembersihan teks dengan cara membuang kata-kata yang tidak terpakai, dilakukan juga restrukturisasi dengan cara memisahkan tiap kata, menghilangkan imbuhan pada tiap kata yang ada, dan melakukan proses penghilangan kata yang tidak relevan pada dataset dokumen nabi maupun pertanyaan yang akan menghasilkan data yang bersih dan siap diolah dengan tahapan sebagai berikut.



Gambar 2. Preprocessing dataset

- 1. Case Folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil, agar sama untuk setiap kata yang ada. Pada penilitian ini menggunakan bantuan fungsi dari *str.lower()* untuk *Case Folding* pada bahasa pemrograman *Python*.
- 2. Tokenizing adalah proses pemotongan *string* berdasarkan setiap kata yang menyusunnya. pada penelitian ini digunakan bantuan *library* yang ada pada bahasa pemrograman *Python* yaitu *Natural Language Toolkit* (NLTK).
- 3. Stopword adalah tahap menghilangkan kata-kata yang tidak memberikan informasi penting. Pada penilitian ini menggunakan bantuan kamus list Bahasa Indonesia yaitu *Stopword list* dari Tala pada bahasa pemrograman *Pyhton*.
- 4. Stemming adalah proses untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk yang berbeda karena mendapatkan imbuhan yang berbeda. Pada penilitian ini menggunakan bantuan library yang ada pada bahasa pemrograman Python yaitu Sastrawi stemmer untuk Bahasa Indonesia.

ISSN 2548-8368 (media online)

Hal 140-149



3.5 Pembobotan Kata

Setelah didapatkan data bersih melalui proses *Preprocessing*. Tahap pertama yang dilakukan adalah melakukan pembobotan kata (*Term Weighting*) yaitu jumlah kemunculan kata dalam suatu dokumen. Selanjutnya melakukan perhitungan *Inverse Document Frequency* (IDF) yaitu bagaimana *term* didistribusikan secara luas pada koleksi dokumen, semakin sedikit jumlah dokumen yang mengandung *term*, maka nilai IDF semakin besar. Berikut merupakan persamaan untuk menghitung bobot *term*:

$$TFIDF(d,t) = TF(d,t) \cdot IDF(t) = TF(d,t) \cdot log(\frac{D}{df_i})$$
 (1)

Keterangan:

TFIDF(d,t) : bobot kata(t) terhadap dokumen(d).

TF(d,t) : jumlah kemunculan kata(t) dalam dokumen(d).
D : jumlah semua dokumen dalam koleksi.
Df_i : jumlah dokumen yang mengandung kata.

3.6 Seleksi Kandidat Dokumen

Tahap selanjutnya adalah proses seleksi kandidat dokumen untuk setiap pertanyaan atau *Query* dengan menggunakan metode *Vektor Space Model. Vektor Space Model* merupakan suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu *Query. Query* dan dokumen dianggap sebagai *Vektor-Vektor* pada n-dimensi, dimana t adalah jumlah dari seluruh term (kata) yang ada dalam leksikon. Menghitung kesamaan antara dokumen dan *Query* di ukur berdasarkan sudut *cosinus* dari *Vektor* dokumen dan *Vektor Query* (*Cosine Measure*) dari nilai pembobotan yang telah diperoleh, berikut ini merupakan persamaan dari *Cosine Similarity*:

$$Cosine(d_j, q) = \frac{(d_j \cdot q)}{||d_j|| \times ||q||}$$
(2)

Keterangan:

q: jumlah *term* keyword d: jumlah *term* dokumen

|q| : panjang vektor dari keyword

|d| : panjang vektor dari dokumen

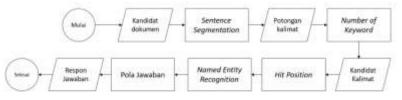
Maka diperoleh hasil perhitungan *Cosine Similarity* dan dilakukan perankingan berdasarkan nilai *Cosine Similarity* tertinggi lalu di ambil 3 nilai tertinggi untuk setiap pertanyaan.

Tabe	<u> 13.</u>	Cosine	Simil	arity

Pertanyaan	Rank	Dokumen	Cosine
	1	10	1
1	2	1	0,894427191
	3	4	0,894427191
	1	25	0,962576835
2	2	1	0,872425573
	3	3	0,872425573
••••		•••	•••

3.7 Answer Extraction

Proses selanjutnya yaitu *Answer Extraction*, proses menemukan kandidat kalimat yang mengandung jawaban. Berikut ini adalah tahap-tahapan dalam proses *Answer Extraction*:



Gambar 3. Answer Extraction

3.7.1 Sentence Segmentation

Kandidat jawaban yang masih berupa dokumen dipisahkan menjadi dalam bentuk kumpulan kalimat. Pada proses pemisahan kalimat ini digunakan penanda sebagai batas antara satu kalimat dengan kalimat lainnya, yaitu menggunakan titik.



3.7.2 Number of Keyword

Proses selanjutnya yaitu menghitung jumlah *Number of Keyword* yang ada pada setiap kalimat-kalimat yang telah di pecah dari proses sebelumnya. *Keyword* merupakan kata kunci atau pertanyaan yang digunakan untuk pencarian dokumen pada proses sebelumnya, contoh pada pertanyaan 1 dengan menghilangkan kata tanya, "siapa nabi pertama?" menjadi "nabi pertama" yang mempunyai jumlah 2 kata pada *keyword*, kemudian jumlah *keyword* yang ditemukan didalam setiap kalimat dimasukkan ke dalam persamaan sebagai berikut:

$$M \ge \left[\sqrt{K} - 1\right] + 1\tag{3}$$

Keterangan:

K : Jumlah Keyword pada pertanyaan.

M : Jumlah Keyword yang terdapat pada kalimat kandidat jawaban.

Kalimat-kalimat kandidat jawaban yang terpenuhi persamaan tersebut, akan menjadi kalimat jawaban dengan skor berdasarkan nilai M tersebut. sebaliknya, jika tidak terpenuhi maka akan di eliminasi kalimat yang tidak memenuhi persamaan tersebut. Agar skor tidak terlampaui jauh maka perlu di normalisasikan terlebih dahulu pada skor tersebut dengan perhitungan persamaan sebagai berikut:

$$Skor = \frac{M}{K} \times 10 \tag{4}$$

3.7.3 Hit Position

Pada proses *Number of Keyword* didapatkan skor pada masing – masing kandidat kalimat. Pada proses *Hit Position* kandidat kalimat jawaban tersebut diurutkan berdasarkan skor tertinggi yang telah di peroleh dari perhitungan diatas, diambil 3 nilai tertinggi untuk setiap pertanyaan. Untuk penjelasan dalam bentuk tabel berada pada lampiran.

Tabel 4. Hit Position

Pertanyaan ke	Kandidat kalimat	Skor
Siapa nabi pertama	Unta Nabi Shaleh dibunuh	15
pertuna	nabi pertama adalah nabi adam Nabi Shaleh memberi waktu	15 15
Siapa anak nabi	 Dari Ibnu Abi Hatim Abu	16,56854
adam	Perbedaan Nabi dan rasulanak nabi adam adalah Qabil	16,56854 12,42641

3.7.4 Named Entity Recognition

Proses selanjutnya Named Entity Recognition menggunakan bantuan tools anago[11] yaitu melakukan pengenalan sejumlah entity dari setiap kandidat kalimat seperti Person, Location, Time, Object dan Count berdasarkan kata tanya yang ada seperti siapa mencari entity Person dimana mencari entity Location, kapan mencari entity Time, apa mencari entity Object dan berapa mencari entity Count. Setelah proses Named Entity Recognition maka didapatkan hasil kandidat kalimat baru, jika terdapat entity yang tidak memenuhi syarat berdasarkan kata tanya, maka tidak akan masuk ke proses selanjutnya.

Tabel 5 Named Entity Recognition

Tabel 3 Ivamed Entity Recognition				
Pertanyaan Ke		Kandidat kalimat	Named Entity Recognition	
Siapa	nabi	Unta Nabi Shaleh dibunuh	[Unta = Object], [Nabi Shaleh =	
pertama			Person]	
		nabi pertama adalah nabi	[Nabi adam = Person] [Nabi	
		adam	pertama = Person	
		Nabi Shaleh memberi	[Nabi shaleh = $Person$]	
		waktu		



3.7.5 Pola Jawaban dan Respon Jawaban

Selanjutnya proses mencocokan pola jawaban berdasarkan kata yang mewakili target dan konteks pada kandidat kalimat. Contoh pertanyaan 1, dengan merubah "nabi" menjadi tag target<T> dan "pertama" menjadi *context*<C>. Sesuai dengan pola jawaban dari pertanyaan yang telah di tentukan :

Contoh implementasi pada pertanyaan 1 :

- 1. Kandidat Kalimat 1 = Unta <T> Shaleh dibunuh
- 2. Kandidat Kalimat $2 = \langle T \rangle \langle C \rangle$ adalah nabi adam
- 3. Kandidat Kalimat $3 = \langle T \rangle$ Shaleh memberi waktu

Setiap kandidat kalimat jawaban dibandingkan dengan pola jawaban yang memiliki tipe properti yang sesuai dengan hasil interpretasi pertanyaan. Dari semua pola jawaban yang didapatkan, kemudian di cari pola yang paling mirip dengan kalimat kandidat jawaban. Untuk dapat menentukan kemiripan dilakukan perbandingan dengan posisi <T>, dan posisi <C>, serta kalimat yang menyertai *tag* tersebut. Berikut pola jawaban yang mewakili tipe properti PERSON:

Pada kandidat kalimat 1, 2, dan 3, hanya kalimat 2 yang mewakili target dan konteks dari pertanyaan. Maka kalimat kandidat 2 akan dijadikan kandidat jawaban untuk diproses ditahap selanjutnya. Sehingga di dapatkan kandidat kalimat baru yaitu:

Kandidat Kalimat $2 = \langle T \rangle \langle C \rangle$ adalah nabi adam

Dari pencocokan pola jawaban dengan kandidat kalimat didapat hanya kandiat kalimat 2 yang ditemukan pola jawaban yang cocok dengan properti PERSON yaitu :

Pada pola pertanyaan tersebut *tag* <P> mewakili potongan kalimat yang akan di ekstrak dan akan menghasilkan respon jawaban dari pertanyaan. Sehingga dari kandidat kalimat 2 untuk pertanyaan pertama di dapatkan jawaban dari hasil ekstrak <P> dengan kata setelah "adalah" yaitu "nabi adam", maka respon jawaban yang diberikan sebagai berikut:

Respon Jawaban: nabi pertama adalah nabi adam.

Properti	Pola pertanyaan	Pola Jawaban
Person	(Siapa) <t><c></c></t>	(PERSON) <t><c>Adalah<p></p></c></t>
	(Siapa)anak <t><c></c></t>	
	(Siapa)makhluk	
	sebelum <t><c></c></t>	
T:	(Variation) /To (C) di labinhan	(TIME) AT A COMMING AD
Time	(Kapan) <t><c>di lahirkan</c></t>	(TIME) <t><c>tahun<p></p></c></t>
	(Kapan) <t><c>menjadi nabi</c></t>	
	(Kapan) <t><c>menjadi rasul</c></t>	
Location	(Dimana) <t><c></c></t>	(LOCATION) <t><c>di<p></p></c></t>
	(Dimana) <t><c>di turunkan</c></t>	(=======,)
	(Dimana) <t><c>bertemu</c></t>	
Object	(Apa) <t><c></c></t>	(OIECT) <t> <c> Adoloh <d></d></c></t>
Object		(OJECT) <t><c>Adalah<p></p></c></t>
	(Apa) julukan <t><c></c></t>	
	(Apa)mukjizat <t><c></c></t>	
Count	(Berapa)usia <t><c></c></t>	(COUNT) <t><c>Adalah<p></p></c></t>
	(Berapa)jumlah <t><c></c></t>	
	(Berapa)anak <t><c></c></t>	

3.8 Metode Evaluasi

Tahap terakhir yaitu mengevaluasi hasil *Question Answering* System yang telah di bangun pada penelitian ini dengan perhitungan *Accuracy*. *Accuracy* dipilih sebagai perhitungan evaluasi untuk mengukur performasi dari *Question Answering System* dengan pertanyaan sebanyak 282, semakin tinggi *Accuracy*, maka performansi dalam penggunaan metode *Pattern Based* semakin baik dalam kasus *Question Answering System*. Berikut merupakan perhitungan *Accuracy*:

$$Accuracy = \frac{jumlah \ pertanyaan \ yang \ benar}{jumlah \ pertanyaan} \times 100 \%$$
 (6)



4. IMPLEMENTASI

Performansi dari sebuah *Question Answering System* dengan menggunakan metode evaluasi dengan menggunakan perhitungan *Accuracy* dari keberhasilan sistem yang mampu mengembalikan jawaban yang sesuai dengan jawaban yang telah disediakan sebelumnya. Pada penelitian ini telah disediakan pertanyaan berikut jawaban, pertanyaan terdiri dari kata tanya siapa berjumlah 84, kapan berjumlah 59, dimana berjumlah 51, apa berjumlah 50, dan berapa berjumlah 38 pertanyaan, dengan total 282 pertanyaan. Hasil pengujian penelitian ini berupa evaluasi dari jawaban yang sesuai, dengan menggunakan metode *Accuracy* sebagai evaluasi dari keberhasilan sistem mengembalikan jawaban yang benar.

Berikut ini ada beberapa skenario yang dibuat dalam penelitian ini :

penelitian dari skenario ke 2 :

1. Skenario pengujian pengaruh penggunaan proses Stemming, Number of Keyword dan Named Entity Recognition pada Question Answering System.

Pada skenario pengujian pengaruh dengan penggunaan proses Stemming, Named Entity Recognition dan Number of Keyword pada Question Answering System. Stemming merupakan proses mengubah kata berimbuhan menjadi kata dasar. Named Entity Recognition merupakan komponen utama dari information extraction yang bertujuan untuk mendeteksi dan mengklasifikasikan Named Entity pada suatu teks[11], dengan mendeteksi PERSON, LOCATION, TIME, OBJECT, dan COUNT dari kata pada setiap kalimat. Number of Keyword untuk menghitung jumlah kata yang sama dari pertanyaan terhadap dokumen jawaban. Pada penelitian ini bertujuan untuk mengetahui peran penting dalam penggunaan Stemming, Named Entity Recognition dan Number of Keyword pada Question Answering System. Diperoleh hasil dari skenario pengujian ini dengan nilai Accuracy sebesar 17,73 %.

Tabel 7. skenario 1

Izoto tonzo	jawaban		
kata tanya	Benar	Salah	
Siapa	7	77	
Kapan	12	47	
Dimana	7	44	
Apa	10	40	
Berapa	14	24	
Total	50	232	
total			
pertanyaan	aan 282		
Accuracy	17,73049645		

Berdasarkan tabel 7 hasil penelitian dengan menggunakan *Stemming* berpengaruh pada hasil *Accuracy* dengan menghilangkan kata imbuhan menjadi kata dasar. Pengaruh Penggunaan *Number of keyword* pada penelitian ini menghitung jumlah kata yang sama dari pertanyaan terhadap dokumen jawaban. Pengaruh penggunaan *Named Entity Recognition* dapat mendeteksi lebih baik pada pola pertanyaan dengan membandingkan kata tanya terhadap target pola jawaban.

 Skenario pengujian pengaruh penggunaan proses Number of Keyword dan Named Entity Recognition pada Question Answering System dengan menghilangkan proses Stemming.
 Pada skenario pengujian pengaruh penggunaan proses Number of Keyword dan Named Entity Recognition pada Question Answering System dengan menghilangkan proses Stemming. Berikut ini merupakan hasil dari

Tabel 8. skenario 2

Iroto tonvo	Jawaban		
kata tanya	Benar	Salah	
siapa	4	80	
kapan	7	52	
dimana	7	44	
Apa	11	39	
berapa	12	26	
total	41	241	
total			
pertanyaan	282		
Accuracy 14,53		00709	



Berdasarkan penggujian ini diperoleh hasil *Accuracy* dengan menghilangkan proses *Stemming* sebesar 14,53%. Dapat disimpulkan, bahwa dengan menghilangkan proses *Stemming* dapat mengurangi hasil *Accuracy*, dikarenakan pada proses *Stemming* berfungsi untuk mengurangi dimensi kata dengan menghilangkan kata imbuhan menjadi kata yang bermakna sama atau kata dasar.

3. Skenario pengujian pengaruh penggunaan proses *Stemming* dan *Number of Keyword* pada *Question Answering System* dengan menghilangkan proses *Named Entity Recognition*.

Pada skenario pengujian pengaruh penggunaan proses *Stemming* dan *Number of Keyword* pada *Question Answering System* dengan menghilangkan proses *Named Entity Recognition*. Berikut ini merupakan hasil dari penelitian dari skenario ke 3:

Tabel 8. skenario 3			
kata tanya	Jawaban		
kata tanya	Benar	Salah	
siapa	7	77	
kapan	12	47	
dimana	6	45	
apa	8	42	
berapa	8	30	
total	41	241	
total			
pertanyaan	28	32	
Accuracy	14,53900709		

Berdasarkan pengujian ini diperoleh hasil dengan menghilangkan proses *Named Entity Recognition* sebesar 14,53%. Dapat disimpulkan, bahwa dengan menghilangkan proses *Named Entity Recognition* dapat mengurangi hasil *Accuracy*, dikarenakan saat proses pencocokan berdasarkan kata tanya dengan target jawaban yang mengakibatkan terjadi kesalahan dalam mengambil jawaban.

4. Skenario pengujian pengaruh penggunaan proses *Stemming* dan *Named Entity Recognition* pada *Question Answering System* dengan menghilangkan proses *Number of Keyword*.

Tabel 9. skenario 4

Pada skenario pengujian dengan pengaruh penggunaan proses *Stemming* dan *Named Entity Recognition* pada *Question Answering System* dengan menghilangkan proses *Number of Keyword*. Berikut ini merupakan hasil dari penelitian dari skenario ke 4:

 kata tanya
 jawaban

 siapa
 30
 54

 kapan
 22
 37

 dimana
 20
 31

apa 23 27
berapa 16 22
total 111 171
total
pertanyaan 282

Accuracy

Berdasarkan pengujian ini diperoleh hasil dengan menghilangkan proses *Number of Keyword* sebesar 39,36%. Dapat disimpulkan, bahwa dengan menghilangkan proses *Number of Keyword* dapat meningkatkan hasil *Accuracy*, dikarenakan saat proses *Number of Keyword* kata nabi lebih dominan pada setiap kalimatnya yang berakibat terjadi kesalahan dalam mengambil pada kandidat kalimat jawaban.

39,36170213

4.1 Analisis Hasil Pengujian

Dari hasil analisis pengujian yang telah dilakukan dengan 282 pertanyaan dengan berbagai kata tanya, didapat jawaban benar sebanyak 111 dan jawaban yang tidak sesuai sebanyak 171 pertanyaan. Diperoleh hasil *Accuracy* terbaik dengan nilai 39,36% dengan menghilangkan proses *Number of Keyword*. Berikut ini yang menyebabkan tidak ditemukannya dokumen relevan atau kalimat yang relevan dengan pertanyaan sehingga pertanyaan dianggap tidak berhubungan dengan dokumen dan sistem juga tidak menghasilkan jawaban karena beberapa faktor yang menyebabkan tidak ada dokumen yang dapat diproses untuk mendapatkan jawaban.

Sistem mengeluarkan jawaban yang salah dikarena kan dalam proses *Number of Keyword* yaitu menentukan kandidat kalimat yang sesuai dengan pertanyaan dari setiap dokumen yang telah menjadi kandidat

MEDIA INFORMATIKA BUDIDARMA, Vol 2, No 4, Oktober 2018 ISSN 2614-5278 (media cetak) ISSN 2548-8368 (media online) Hal 140-149



dokumen yang didapatkan, dikarenakan *Query* yang dikeluarkan pada saat pencarian kandidat kalimat lebih banyak sehingga kalimat yang diambil meskipun kalimat tersebut tidak mengandung jawaban, seperti pada kata nabi yang menjadi konteks, selalu ada di setiap dokumen dan berkali kali muncul, sedangkan targetnya seharusnya tertutupi oleh kata nabi, yang dapat menimbulkan salah dalam mengambil kandidat kalimat, hal ini bisa di atasi dengan menghilangkan proses *Number of Keyword* dengan menghitung jumlah kata yang sama hanya terhitung satu.

Pada saat proses *Named Entity Recognition* ada beberapa kata yang salah dalam *tag* kata yang tidak sesuai, seperti kata *Object* menjadi *Person*, hal itu dapat mempengaruhi dalam proses pola ekstraksi jawaban. Hal ini bisa diatasi jika dikembangkan lagi dalam proses *Learning* dalam model *Named Entity Recognition*. Permasalahan dengan pola jawaban, karena tidak ditemukannya pola jawaban yang mirip dengan susunan kalimat kandidat jawaban, untuk mengurangi permasalahan ini dapat dilakukan dengan pegembangan dalam menemukan pola jawaban yang baik.

5. KESIMPULAN

Setelah dilakukan analisis *Question Answering System* dengan metode *Pattern Based* pada dokumen 25 Nabi dalam Agama Islam maka dapat disimpulkan bahwa peran pola jawaban sangat penting dalam ekstraksi jawaban. Sistem akan lebih mudah mengenali pola jika pola lebih spesifik dan lebih unik. Peran memberi *tag* pada kata mempengaruhi dalam menemukan target jawaban yang sesuai dengan kalimat tanya. Sistem akan lebih mudah jika *tag* yang di berikan lebih akurat. Proses *Number of Keyword* terdapat masalah jika *Query* atau kata selalu muncul dalam semua dokumen, yang dapat mengakibatkan kesalahan dalam melakukan pengambilan kandidat kalimat yang relevan.

Saran untuk penelitian selanjutnya, dalam pencarian dokumen yang membahas nabi lebih diperluas. Untuk di kembangkan lagi dalam proses analisa tentang *Natural Language Processing* (NLP) terkait *Named Entity Recognition* dan memperbanyak dalam pembuatan pola jawaban yang lebih spesifik dan lebih unik.

REFERENCES

- [1] Gunawan dan Gita Lovina "QUESTION ANSWERING SYSTEM DAN PENERAPANNYA PADA ALKITAB".
- [2] Hirschman, L, dan Gaizauskas, R, *Natural language question asnwring: the view from here*. 2001. Cambrige University Press, United Kingdom.
- [3] Bintang Kristina Situmorang "Analisis *Question Answering System* pada dokumen Bahasa Indonesia Menggunakan *Pattern Based* Approach".
- [4] H. T. a. M. Adriani, "Pattern Based Approach in Indonesian Question Answering System" 2010.
- [5] Abdelghani BOUZIANE, Djelloul BOUCHIHA, Noureddine DOUMI and Mimoun MALKI "Question Answering Systems: Survey and Trends"
- [6] Hapnes Toba and Mirna Adriani "Pattern Based Indonesian Question Answering System".
- [7] Junichi Fukumoto, Noriaki Aburai, dan Ryosuke Yamanishi "Interactive Document Expansion for Answer Extraction of Question Answering System".
- [8] Amit Mishra, dan Sanjay Kumar Jain "A survey on Question Answering Systems with classification".
- [9] Sanjay K Dwivedi, dan Vaishali Singh "Research and reviews in Question Answering System".
- [10] Kartina "Analisis Pertanyaan Berbahasa Indonesia Pada Question Answering System".
- [11] Wibisono, Yudi. NER (Named Entity Recognition) dengan anaGo (Python Keras) https://yudiwbs.wordpress.com/2018/03/29/ner-named-entity-recognition-dengan-anago-python-keras/. Diakses pada 16 juli 2018