

Sistem *Question Answering* untuk Bahasa Bali menggunakan Metode *Rule-Based* dan *String Similarity*

Question Answering System for Balinese using Rule-Based and String Similarity Method

Made Agus Putra Subali¹, Puritan Wijaya²

^{1,2}Fakultas Informatika dan Komputer, Program Studi Sistem Informasi,
Institut Teknologi dan Bisnis STIKOM Bali, Denpasar, Indonesia
E-mail: ¹madeagusputrasubali@gmail.com, ²puri@stikom-bali.ac.id

Abstrak

Sistem *question answering* merupakan kemampuan sistem untuk memberikan jawaban atas kalimat tanya yang diberikan oleh *user*. Sampai saat ini penelitian tentang sistem *question answering* pada bahasa Bali belum pernah dilakukan. Pada penelitian ini kalimat tanya yang digunakan adalah kalimat tanya biasa, sebagai contoh kalimat tanya “*akuda memene ngubuh siap?*” Dalam bahasa Indonesia “berapa ibumu memelihara ayam?” Data yang digunakan dalam penelitian ini merupakan lima puluh dokumen berbahasa Bali. Sedangkan pada tahap pengujian dilakukan dengan menggunakan dua puluh kalimat tanya. Adapun metode yang diusulkan dalam penelitian ini dimulai dari memberi *input* pertanyaan, mencari dokumen paling relevan berdasarkan pertanyaan yang diberikan, dan memperoleh jawaban berdasarkan *rules* untuk setiap pertanyaan. Berdasarkan pengujian pada kedua puluh pertanyaan yang diberikan metode yang diusulkan memperoleh akurasi sebesar 40% terkait kebenaran respons atau jawaban yang diberikan.

Kata kunci: Sistem *Question Answering*, *Rule-Based*, *Cosine Similarity*, *N-Gram Similarity*

Abstract

The question answering system is the system's ability to answer the question given by the user. Currently, research on the question answering system in Balinese has never been conducting. In this study, the sentence used is the usual interrogative sentence in Balinese, for example, "akuda memene ngubuh siap?" In Indonesian "berapa ibumu memelihara ayam?" The data have obtained from 50 documents in the Balinese. The testing phase has carried out using 20 interrogative sentences. The proposed method in this study begins with providing the input questions, searching for the most relevant documents, and getting the answers. Based on the testing result's obtained an accuracy of 40% of the 20 questions given.

Keywords: *Question Answering System*, *Rule-Based*, *Cosine Similarity*, *N-Gram Similarity*

1. PENDAHULUAN

Sistem *question answering* merupakan kemampuan sistem untuk memberikan jawaban atas kalimat tanya yang diberikan oleh *user* [1]. Kalimat tanya dalam bahasa Bali dapat dibedakan menurut jawaban yang diinginkan dan menurut cakupan isi atau penekanannya [2]. Metode pada sistem *question answering* memiliki tiga komponen utama yaitu *question analysis*, *passage retrieval*, dan *answer extraction* [3]. Penelitian tentang sistem *question answering* pernah dilakukan oleh Yusliani [4] menggunakan tipe pertanyaan *non-factoid* (definisi, alasan, dan metode), dimana pada tahap pengambil dokumen dilakukan proses pencarian lima belas dokumen paling relevan menggunakan persamaan *cosinus* tertinggi. Untuk memperoleh jawaban terbaik dilakukan dengan mencocokkan dengan *rules* kalimat yang telah ditetapkan, hasil dari penelitian tersebut memperoleh rerata akurasi 76% untuk pertanyaan definisi, 59%

untuk pertanyaan alasan, dan 57% untuk pertanyaan metode. Sedangkan, penelitian yang dilakukan oleh Purwarianti [5] menggunakan tipe pertanyaan *factoid* [6], [7] dengan metode atau kerangka sistem *question answering* yang sama dengan penelitian [4] hasil dari penelitian tersebut memperoleh rerata akurasi 51%. Baik penelitian [4], [5], [7] bahasa yang digunakan adalah bahasa Indonesia sehingga *rules* kalimat untuk memperoleh jawaban terbaik tidak cocok dengan bahasa Bali, hasil dari penelitian sistem *question answering* terdahulu memberikan akurasi yang cukup baik terutama pada pertanyaan definisi. Penelitian lain yang serupa dengan topik sistem *question answering* yaitu *chat bot* pernah dilakukan oleh [8] dengan menggunakan *line bot application* dalam pembuatan kamus penerjemah dari bahasa Indonesia ke bahasa Bali, namun pada penelitian tersebut metode *chat bot* yang diterapkan sangat sederhana, *input* hanya menggunakan kata tunggal sekaligus sebagai kata kunci, kemudian respons yang diberikan adalah kata terjemahan yang paling sesuai dengan kata kunci, apabila kata kunci tidak ditemukan maka tidak ada hasil yang diberikan. Sampai saat ini penelitian tentang sistem *question answering* pada bahasa Bali belum pernah dilakukan.

Pada sistem *question answering* selain menerima kata kunci *input*, dilakukan tahap pengambilan dokumen yang paling relevan berdasarkan kata kunci, serta tahapan pencarian jawaban diperoleh berdasarkan *rules* di setiap kata tanya pada kata kunci. Pembentukan *rules* dibentuk berdasarkan kata tanya yang digunakan. Kalimat tanya dalam Bahasa Bali, dibedakan menjadi dua bentuk yaitu kalimat tanya biasa dan kalimat tanya retorik. Kalimat tanya biasa adalah kalimat tanya yang meminta jawaban sesuai dengan pertanyaan, sedangkan kalimat tanya retorik adalah kalimat tanya yang tidak memerlukan jawaban karena penanya sudah mengetahui jawabannya [2]. Pada penelitian ini kalimat tanya yang akan digunakan adalah kalimat tanya biasa, seperti *akuda, apa, dija, kenken, kuda, dan nyen*, sebagai contoh kalimat tanya “*akuda memene ngubuh siap?*” memiliki arti “berapa ibumu memelihara ayam?”

Data yang digunakan dalam penelitian ini merupakan lima puluh dokumen berbahasa Bali dari buku pelajaran sekolah dasar, sedangkan pada tahap pengujian dilakukan dengan menggunakan dua puluh kalimat tanya dengan variasi tipe pertanyaan definisi dalam bahasa Bali. Adapun metode yang diusulkan dalam penelitian ini dimulai dari (a) memberi *input* pertanyaan, (b) mencari sepuluh dokumen paling relevan berdasarkan pertanyaan yang diberikan dengan memperhatikan nilai kesamaan *cosinus* tertinggi menggunakan metode *cosine similarity*, serta (c) memperoleh jawaban berdasarkan *rules* untuk setiap pertanyaan, *rules* untuk setiap pertanyaan ditentukan berdasarkan keputusan *expert*. Hasil yang diperoleh dalam penelitian ini adalah sistem *question answering* yang diusulkan mampu merespons jawaban dengan akurasi kebenaran sebesar 40%.

2. METODE PENELITIAN

Pada Gambar 1 merupakan metode usulan yang terdiri dari beberapa tahapan, antara lain: *input* pertanyaan, *data preprocessing*, analisis pertanyaan, ambil dokumen, cari jawaban, dan *output* jawaban.



Gambar 1 Metode Usulan

2.1 Input Pertanyaan

Pertanyaan yang dimasukkan adalah kalimat tanya bertipe definisi dalam bahasa Bali, seperti kata: *akuda, apa, dija, kenken, kuda, dan nyen* [2].

2.2 Data Preprocessing

Adapun tahapan data *pre-processing* yang dilakukan adalah *tokenize*, hapus *stop word* dengan menghilangkan kata seperti *anggan, sane, ring, miwah, puniki, olih*, [9] serta tambahan

kata *di* dan *i*, kemudian *stemming* bahasa Bali menggunakan [10] dari *input* pertanyaan.

2.3 Analisis Pertanyaan

Analisis pertanyaan digunakan untuk memperoleh kata kunci dari tahap data *pre-processing*, kata kunci yang dimaksud adalah kata benda, kata kerja, dan kata tanya [4], [5].

2.4 Ambil Dokumen

Tahap ambil dokumen digunakan untuk mengambil sepuluh dokumen paling relevan dari lima puluh dokumen, dengan cara mengukur tingkat kemiripan setiap dokumen dengan kata kunci yang diperoleh menggunakan metode *cosine similarity* [11], seperti pada persamaan (1).

$$similarity(d_j, q) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_{i=1}^n (W_{i,j} \cdot W_{i,q})}{\sqrt{\sum_{i=1}^n W_{i,j}^2 \cdot \sum_{i=1}^n W_{i,q}^2}} \quad (1)$$

Keterangan:

$W_{i,j}$ merupakan bobot kata *i* pada dokumen *j*.

$W_{i,q}$ merupakan bobot kata *i* pada pertanyaan *q*.

Untuk memperoleh bobot kata dihitung menggunakan *tf · idf* kata tersebut.

2.5 Cari Jawaban

Tahap cari jawaban digunakan untuk memperoleh jawaban terbaik berdasarkan dokumen yang diperoleh pada tahap ambil dokumen. Dokumen paling relevan dari lima puluh dokumen akan diukur tingkat kemiripannya dengan *rules* atau aturan pencari jawaban dengan menggunakan kombinasi persamaan (2) dan persamaan (3). Penggunaan persamaan (3) digunakan untuk mengukur tingkat kemiripan struktur kalimat pada dokumen dengan *rules* pencari jawaban [12]. Adapun beberapa *rules* yang diusulkan terlihat pada Tabel 1.

$$IF \ p \ THEN \ q \quad (2)$$

Keterangan:

IF menerangkan suatu kondisi *p*.

THEN menerangkan aksi *q*.

$$dc = (2 \cdot c) \div (a + b) \quad (3)$$

Keterangan:

dc merupakan hasil perhitungan tingkat kemiripan.

a merupakan bentuk *n*-gram pada kata pertama.

b merupakan bentuk *n*-gram pada kata kedua.

c merupakan bentuk *n*-gram yang sama antara *a* dan *b*.

Tabel 1 *Rules* Pencari Jawaban Bahasa Bali

| Tipe Pertanyaan | Rule Pencari Jawaban |
|--------------------|---|
| <i>Akuda</i> | Kata Kunci + Sufiks <i>-Ne</i> + Kata Khusus (<i>Ada</i>) + Jawaban. |
| <i>Apa</i> | Kata Khusus (<i>Sawireh</i>) + Kata Kunci + Jawaban. Kata Kunci + Kata Khusus (<i>Saja Mirib</i>) + Jawaban. |
| <i>Dija</i> | Kata Khusus (<i>Awake</i>) + Kata Kunci + Jawaban. Jawaban + Kata Khusus (<i>Ada</i>) + Kata Kunci. |
| <i>Kenken</i> | Kata Kunci + Jawaban. Kata Kunci + Kata Kunci + Prefiks <i>-Ma</i> + Jawaban. Kata Khusus (<i>Makejang</i>) + Kata Kunci + Jawaban. |
| <i>Kuda/Kudang</i> | Kata Khusus (<i>Ada</i>) + Jawaban + Kata Kunci. Kata Khusus (<i>Ada</i>) + Jawaban. |
| <i>Nyen</i> | Jawaban + Kata Kunci. Kata Kunci + Jawaban + Kata Kunci. |

2.6 Output Jawaban

Satu jawaban terbaik yang diperoleh dari dokumen paling relevan dari lima puluh dokumen yang digunakan akan ditampilkan.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Data penelitian yang digunakan dalam penelitian ini adalah 50 dokumen berbahasa Bali. Pada setiap dokumen berisi informasi yang berkorelasi dengan setiap topik dokumen. Seluruh dokumen diperoleh dari buku mata pelajaran sekolah dasar. Sedangkan dalam tahap pengujian digunakan 20 pertanyaan beserta jawabannya sebagai acuan dalam menentukan hasil penelitian. Pada Tabel 2 merupakan kedua puluh pertanyaan.

Tabel 2 Kedua Puluh Pertanyaan

| No. | Pertanyaan |
|-----|---|
| 1 | <i>dadong dauh ngelah siap mabulu mawarna kenken?</i> |
| 2 | <i>kuda siap putihe mataluh?</i> |
| 3 | <i>nyen sane ngelah siap putih?</i> |
| 4 | <i>apa krana i kekua berang aking?</i> |
| 5 | <i>dija i lutung nepukin tongos melah?</i> |
| 6 | <i>i kancil dadi juru kemit apa?</i> |
| 7 | <i>dugas i malu nyen sane matanduk?</i> |
| 8 | <i>kenken warnan bok i kaki?</i> |
| 9 | <i>akuda made santi ngelah kreta?</i> |
| 10 | <i>kenkenang patutne plialiane suud anggo?</i> |
| 11 | <i>dija i wayan susila malajah ngigel?</i> |
| 12 | <i>anak sane kenken madan ubuh?</i> |
| 13 | <i>jam kuda murid kelas duane masuk?</i> |
| 14 | <i>kudang wuku muride maan libur?</i> |
| 15 | <i>kali kenken i siap selem makeber?</i> |
| 16 | <i>nyen dadi ketua kelas di kelas dua?</i> |
| 17 | <i>nyen adan guru kelas duane?</i> |
| 18 | <i>makire kija putri meseh penganggo?</i> |
| 19 | <i>nyen ane teka tur nyaplok batu?</i> |
| 20 | <i>dija i doglagan medem?</i> |

3.2 Analisis Pertanyaan

Berdasarkan kedua puluh *input* pertanyaan yang diberikan tahap analisis pertanyaan mampu membedakan kata kunci berupa kata tanya dan kata kunci lain. Pada Tabel 3 merupakan detail kata kunci yang diperoleh dari kedua puluh dokumen. Pada kolom *pertanyaan* merupakan nomor pertanyaan sesuai pada Tabel 2, kolom *kata tanya* merupakan kata tanya pada setiap *input* pertanyaan, kolom *kata kunci lain* merupakan kata kunci yang bukan merupakan kata tanya dan bukan kata pada daftar *stop word*.

Tabel 3 Detail Kata Kunci Kedua Puluh Pertanyaan

| No. | Pertanyaan | Kata Tanya | Kata Kunci Lain | Stop Word |
|-----|------------|------------------|--|----------------|
| 1 | (1) | <i>kenken</i> | <i>dadong, dauh, ngelah, siap, mabulu, mawarna</i> | - |
| 2 | (2) | <i>kuda</i> | <i>siap, putihe, mataluh</i> | - |
| 3 | (3) | <i>nyen</i> | <i>ngelah, siap, putih</i> | <i>sane</i> |
| 4 | (4) | <i>apa</i> | <i>krana, kekua, berang, aking</i> | <i>i</i> |
| 5 | (5) | <i>dija</i> | <i>lutung, nepukin, tongos, melah</i> | <i>i</i> |
| 6 | (6) | <i>apa</i> | <i>kancil, dadi, juru, kemit</i> | <i>i</i> |
| 7 | (7) | <i>nyen</i> | <i>dugas, malu, matanduk</i> | <i>sane, i</i> |
| 8 | (8) | <i>kenken</i> | <i>warnan, bok, kaki</i> | <i>i</i> |
| 9 | (9) | <i>akuda</i> | <i>made, santi, ngelah, kreta</i> | - |
| 10 | (10) | <i>kenkenang</i> | <i>patutne, plialiane, suud, anggo</i> | - |
| 11 | (11) | <i>dija</i> | <i>wayan, susila, malajah, ngigel</i> | <i>i</i> |
| 12 | (12) | <i>kenken</i> | <i>anak, madan, ubuh</i> | <i>sane</i> |

| | | | | |
|----|------|---------------|--|-----------|
| 13 | (13) | <i>kuda</i> | <i>jam, murid, kelas, duane, masuk</i> | - |
| 14 | (14) | <i>kudang</i> | <i>wuku, muride, maan, libur</i> | - |
| 15 | (15) | <i>kenken</i> | <i>kali, siap, selem, makeber</i> | <i>i</i> |
| 16 | (16) | <i>nyen</i> | <i>dadi, ketua, kelas, kelas, dua</i> | <i>di</i> |
| 17 | (17) | <i>nyen</i> | <i>adan, guru, kelas, duane</i> | - |
| 18 | (18) | <i>kija</i> | <i>makire, putri, meseh, penganggo</i> | - |
| 19 | (19) | <i>nyen</i> | <i>ane, teka, tur, nyaplok, batu</i> | - |
| 20 | (20) | <i>dija</i> | <i>dija, doglagan, medem</i> | <i>i</i> |

3.3 Ambil Dokumen

Pada Tabel 4 merupakan dokumen paling relevan dari kelima puluh dokumen untuk setiap *input* pertanyaan. Pada kolom *pertanyaan* merupakan nomor pertanyaan dari *input* pertanyaan sesuai pada Tabel 2, kolom *dokumen (skor)* merupakan nomor dokumen sesuai dengan *dataset*, sedangkan skor atau tingkat kemiripan dari dokumen tersebut dengan kata kunci dari *input* pertanyaan menggunakan persamaan (1).

Tabel 4 Dokumen Paling Relevan pada setiap *Input* Pertanyaan

| No. | Pertanyaan | Dokumen (Skor) |
|-----|------------|--|
| 1 | (1) | 0 (0.39); 1 (0.04); 10 (0.07); 19 (0.13); 23 (0.06); |
| 2 | (2) | 0 (0.31); |
| 3 | (3) | 0 (0.26); 4 (0.10); 10 (0.07); 13 (0.12); 19 (0.10); 22 (0.07); 23 (0.06); |
| 4 | (4) | 1 (0.23); 2 (0.04); 9 (0.03); |
| 5 | (5) | 0 (0.10); 1 (0.33); 2 (0.13); 10 (0.10); 18 (0.10); 25 (0.10); |
| 6 | (6) | 2 (0.39); 14 (0.03); 24 (0.03); |
| 7 | (7) | 3 (0.20); 13 (0.11); 14 (0.04); |
| 8 | (8) | 4 (0.31); |
| 9 | (9) | 0 (0.12); 10 (0.12); 12 (0.05); 14 (0.11); 17 (0.06); 19 (0.17); 23 (0.31); |
| 10 | (10) | 1 (0.03); 11 (0.04); 14 (0.09); 15 (0.06); 18 (0.10); 20 (0.08); |
| 11 | (11) | 8 (0.11); 11 (0.08); 14 (0.06); 25 (0.13); |
| 12 | (12) | 0 (0.16); 9 (0.11); |

| | | |
|----|------|--|
| | | 10 (0.17); 16 (0.18); |
| 13 | (13) | 12 (0.24); 14 (0.04); 16 (0.07); 20 (0.11); 23 (0.14); 25 (0.04); |
| 14 | (14) | 9 (0.06); 11 (0.26); 12 (0.09); |
| 15 | (15) | 0 (0.22); 29 (0.27); |
| 16 | (16) | 2 (0.02); 12 (0.25); 13 (0.09); 14 (0.10); 23 (0.19); 24 (0.02); |
| 17 | (17) | 11 (0.05); 12 (0.15); 13 (0.13); 14 (0.05); 23 (0.08); 24 (0.05); |
| 18 | (18) | 6 (0.17); 14 (0.17); |
| 19 | (19) | 2 (0.03); 8 (0.07); 9 (0.04); 13 (0.12); 16 (0.15); 19 (0.11); 21 (0.05); 23 (0.06); 24 (0.09); 28 (0.02); |
| 20 | (20) | 29 (0.28); |

3.4 Cari Jawaban

Pada Tabel 5 merupakan hasil dari proses pencarian jawaban pada setiap *input* pertanyaan di setiap dokumen paling relevan. Pada kolom *pertanyaan* merupakan nomor pertanyaan dari *input* pertanyaan sesuai pada Tabel 2, kolom *jawaban* merupakan jawaban yang diperoleh dari metode yang diusulkan. Terdapat sembilan pertanyaan yang direspons sedangkan sebelas pertanyaan yang tidak dapat direspons oleh metode yang diusulkan.

Sebagai contoh berikut merupakan tahapan proses dalam mencari jawaban pada baris nomor dua dengan jawaban “*ada limolas taluh*” pada Tabel 5:

1. Pada tahap analisis pertanyaan diperoleh kata kunci “*siap, putihe, mataluh*” dan kata tanya “*kuda*” dari *input* pertanyaan “*kuda siap putihe mataluh?*”.
2. Berdasarkan *rules* pencari jawaban pada tipe pertanyaan “*kuda/kudang*” pada Tabel 1 diperoleh *rules* sebagai berikut: *ada {jawaban} <kata kunci>*. Berdasarkan *rules* tersebut diperoleh struktur kandidat jawaban sesuai dengan kata kuncinya sebagai berikut:
 - a. *ada {jawaban} siap.*
 - b. *ada {jawaban} putih.*
 - c. *ada {jawaban} taluh.*
3. Struktur kandidat jawaban yang diperoleh kemudian diukur tingkat kemiripannya dengan struktur kalimat pada dokumen yang paling relevan. Untuk memperoleh struktur kalimat yang akan diukur dilakukan dengan mengambil tiga kata dengan struktur: *ada {jawaban} <kata kunci>* dari dokumen, sebagai berikut: “*dadong dauh ngelah siap putih suba mataluh reko minab ada limolas taluhne nanging lacur ada nak nepukin anak cerik-cerik kaliwat rusit ipun*”, terdapat dua kalimat yang mengandung kata khusus “*ada*”

yang telah ditandai dengan garis bawah. Pada proses untuk mengukur tingkat kemiripannya dilakukan menggunakan persamaan (3). Berikut adalah hasil dari proses perhitungan:

- a. *ada limolas taluhne* : *ada limolas siap* = 0.67
 - b. *ada limolas taluhne* : *ada limolas putih* = 0.65
 - c. *ada limolas taluhne* : *ada limolas taluh* = 0.94
 - d. *ada nak nepukin* : *ada nak siap* = 0.56
 - e. *ada nak nepukin* : *ada nak putih* = 0.62
 - f. *ada nak nepukin* : *ada nak taluh* = 0.54
4. Berdasarkan hasil perhitungan tingkat kemiripan dengan skor tertinggi diperoleh kandidat jawaban “*ada limolas taluh*” dengan skor 0.94.

Tabel 5 Hasil Pencarian Jawaban pada setiap *Input* Pertanyaan

| No. | Pertanyaan | Jawaban |
|-----|------------|------------------------------|
| 1 | (1) | - |
| 2 | (2) | <i>ada limolas taluh</i> |
| 3 | (3) | - |
| 4 | (4) | <i>aking saja mirib tuna</i> |
| 5 | (5) | <i>tukad ada tongos</i> |
| 6 | (6) | <i>sawireh dadi juru</i> |
| 7 | (7) | - |
| 8 | (8) | <i>kaki mabok putih</i> |
| 9 | (9) | <i>kretane ada dadua</i> |
| 10 | (10) | - |
| 11 | (11) | - |
| 12 | (12) | - |
| 13 | (13) | - |
| 14 | (14) | <i>ada duang wuku</i> |
| 15 | (15) | <i>selem makeber petenge</i> |
| 16 | (16) | - |
| 17 | (17) | - |
| 18 | (18) | - |
| 19 | (19) | - |
| 20 | (20) | <i>awake medem duur</i> |

3.5 Evaluasi Hasil

Berdasarkan kesembilan pertanyaan yang telah diberikan respons, kemudian dilakukan evaluasi apakah respons atau jawaban yang diberikan sudah sesuai atau belum. Evaluasi dilakukan oleh *expert* dalam bidang ilmu Bahasa Bali. Pada pertanyaan nomor enam respons yang diberikan dianggap salah, sehingga hasil akurasi kesesuaian respons yang diberikan sebesar 40% diperoleh dari persamaan (4) untuk mengukur akurasi [13].

Rendahnya akurasi yang diperoleh disebabkan adanya beberapa pertanyaan yang tidak dapat memberikan respons, hal ini dikarenakan *rules* pada pencari jawaban tidak menemukan kecocokan dengan struktur kata pada kalimat atau isi dokumen, seperti sulitnya menemukan jawaban yang berupa frasa, tidak adanya kata khusus pada *rules* pencari jawaban, dan tidak adanya *rules* untuk menangani struktur kalimat pasif.

Sebagai contoh pada pertanyaan “*dadong dauh ngelah siap mabulu mawarna kenken?*” ketiga *rules* pencari jawaban pada tipe pertanyaan *kenken* tidak ada yang sesuai dengan struktur kata pada isi dokumen “*dadong dauh ngelah siap putih suba mataluh reko minab ada limolas taluhne nanging lacur ada nak nepukin anak cerik-cerik kaliwat rusit ipun...*”. Pada pertanyaan dan isi dokumen tersebut terdapat kata kunci yang sama yaitu *dadong*, *dauh*, *ngelah*, *siap* namun pada isi dokumen tidak ditemukan prefiks *ma* dimana jawaban terdapat setelah kata dengan prefiks *ma* yaitu *putih*. Apabila susunan isi dokumen seperti “*...ngelah siap mabulu putih...*” atau “*...ngelah siap mawarna putih*” pertanyaan akan direspons dengan benar.

$$s = \frac{t}{n} \times 100 \quad (4)$$

Keterangan:

s merupakan hasil perhitungan akurasi.

t merupakan total respons yang sesuai dari keseluruhan pertanyaan.

n merupakan total pertanyaan.

4. KESIMPULAN DAN SARAN

Sistem *question answering* yang diusulkan telah mampu merespons pertanyaan yang diberikan. Berdasarkan kedua puluh pertanyaan, metode yang diusulkan mampu memperoleh akurasi sebesar 40% atau metode yang diusulkan dapat merespons sembilan pertanyaan dan delapan diantaranya memberikan respons yang sesuai dengan penilaian *expert*. Pada *rules* pencari jawaban yang diusulkan belum dapat mencari atau menangani struktur kata yang lebih kompleks seperti jawaban yang berupa frasa ataupun jawaban yang berada pada struktur kalimat pasif. Pada penelitian selanjutnya hasil akurasi akan ditingkatkan dengan membentuk *rules* yang dapat menangani struktur bahasa yang lebih kompleks serta memperhatikan tanda baca pada isi dokumen untuk mencari jawaban yang berupa frasa.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Institut Teknologi dan Bisnis STIKOM Bali yang telah memberi dukungan finansial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] S. Nalawade, S. Kumar, and D. Tiwari, "Question Answering System," *International Journal of Science and Research (IJSR)*, vol. 3, no. 5, pp. 439–444, 2014.
- [2] I. W. O. Granoka *et al.*, *Tata Bahasa Baku Bahasa Bali*. Denpasar: Balai Penelitian Bahasa Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan, 1996.
- [3] J. Prager, J. Chu-Carroll, E. W. Brown, and K. Czuba, "Question Answering By Predictive Annotation," in *Advances in Open Domain Question Answering*, T. Strzalkowski and S. M. Harabagiu, Eds. Dordrecht: Springer, 2008, pp. 307–347.
- [4] N. Yusliani and A. Purwarianti, "Sistem Question Answering Bahasa Indonesia untuk Pertanyaan Non - Factoid," *Jurnal Ilmu Komputer dan Informasi*, vol. 4, no. 1, pp. 10–14, 2011.
- [5] A. Purwarianti, "Developing Cross Language Systems for Language Pair with Limited Resource: Indonesian - Japanese CLIR and CLQA," Toyohashi University of Technology, Japan, 2007.
- [6] N. P. Er and I. Cicekli, "A Factoid Question Answering System Using Answer Pattern Matching," in *International Joint Conference on Natural Language Processing (IJCNLP)*, 2013, pp. 854–858.
- [7] R. Rosyadi, S. Al-Faraby, and Adiwijaya, "Penerapan Question Answering System Pada Pembahasan Agama Islam Dengan Pendekatan Metode Pattern Based," *Jurnal Media Informatika Budidarma*, vol. 2, no. 4, pp. 140–149, 2018.
- [8] I. P. G. A. Sudiarmika, I. M. A. W. Putra, K. H. S. Dewi, and I. K. B. Aryawan, "Line Bot Implementation for Automation Balinese Language Dictionary," in *International Conference on Cybernetics and Intelligent System (ICORIS)*, 2019, pp. 227–232.
- [9] I. B. G. W. Putra, M. Sudarma, and I. N. S. Kumara, "Klasifikasi Teks Bahasa Bali dengan Metode Supervised Learning Naive Bayes Classifier," *Teknologi Elektro*, vol. 15, no. 2, pp. 81–86, 2016.
- [10] M. A. P. Subali and C. Fatichah, "Kombinasi Metode Rule-Based dan N-Gram

- Stemming untuk Mengenali Stemmer Bahasa Bali,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 2, 2019.
- [11] M. A. Fauzi, A. Z. Arifin, and A. Yuniarti, “Term Weighting Berbasis Indeks Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab,” *Lontar Komputer*, vol. 5, no. 2, pp. 435–442, 2014.
- [12] T. M. Sembok and Z. A. Bakar, “Effectiveness of Stemming and N-Grams String Similarity Matching on Malay Documents,” *International Journal of Applied Mathematics and Informatics*, vol. 5, no. 3, pp. 208–215, 2011.
- [13] M. S. HUSAIN, “An Unsupervised Approach to Develop Stemmer,” *International Journal on Natural Language Computing (IJNLC)*, vol. 1, no. 2, pp. 15–23, 2012.