

ANSWER FINDING PADA INDONESIA QUESTION  
ANSWERING SYSTEM MENGGUNAKAN  
REGULAR EXPRESSION DAN COSINE SIMILARITY

*Diajukan Untuk Mengajukan Skripsi  
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI*



Oleh :

Nabilah Isyraq Alfadhilah  
NIM : 09021381823145

**Jurusan Teknik Informatika**  
**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**  
**2024**

LEMBAR PENGESAHAN SKRIPSI

ANSWER FINDING PADA INDONESIA QUESTION  
ANSWERING SYSTEM MENGGUNAKAN  
REGULAR EXPRESSION DAN COSINE SIMILARITY

Oleh :

Nabilah Isyraq Alfadhilah

090213818123145

Palembang, 10 Juli 2024

Pembimbing I

  
Novi Yusliani, M.T.

NIP. 198211082012122001

Pembimbing II,

  
Desty Rodiah, M.T.

NIP. 198912212020122011

Mengetahui,

Ketua Jurusan Teknik Informatika

  
  
Dr. M. Fachrurrozi, S.Si., M.T.  
NIP. 198005222008121002

## TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Rabu tanggal 10 Juli 2024 telah dilaksanakan ujian sidang skripsi oleh jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nabilah Isyraq Alfadhilah

NIM : 09021381823145

Judul : Answer Finding pada Indonesia Question Answering System  
Menggunakan Regular Expression dan Cosine Similarity

### 1. Pembimbing I

Novi Yusliani, M.T.  
NIP. 198211082012122001

### 2. Pembimbing II

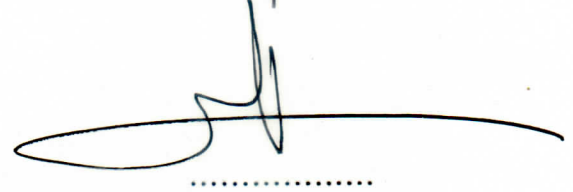
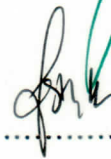
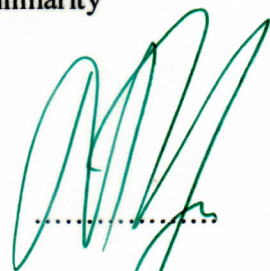
Desty Rodiah, M.T.  
NIP. 198912212020122011

### 3. Ketua Penguji

Yunita, M.Cs  
NIP. 198306062015042002

### 4. Penguji

Dr. Abdiansah, M.Cs  
NIP. 198410012009121005



Mengetahui,  
Ketua Jurusan Teknik Informatika

Dr. M. Fachrurrozi, S.Si., M.T  
NIP. 198005222008121002

## HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Nabilah Isyraq Alfadhilah  
NIM : 09021381823145  
Program Studi : Teknik Informatika  
Judul Skripsi : Answer Finding pada Indonesia Question Answering System Menggunakan Regular Expression dan Cosine Similarity

Hasil Pengecekan Software *iThenticate/Turnitin* : 7 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan /plagiat. Apabila ditemukan unsur penjiplakan/plagiat dala projek ini, maka saya siap menerima sanksi akademik ari Universitas Sriwijaya sesuai dengan ketentuan berlaku.

Demikian pernyataan ini saya buat dengan sebenar-benarnya dan tidak ada paksaan dari pihak manapun.



Palembang, 10 Juli 2024



Nabilah Isyraq Alfadhilah

09021381823145

Motto :

Allah tidak membebani seseorang itu melainkan sesuai dengan kesanggupannya

(Q.S. Al-Baqarah : 286)

Penelitian ini dipersembahkan kepada :

Orang tua dan teman-teman TIBILC 2018

## ABSTRACT

A Question Answering System (QAS) is an automated system designed to provide accurate answers to questions posed using natural language. This research aims to determine the ability of the Indonesian QAS built to perform answer finding or predict the answers submitted to the system using regular expression and cosine similarity methods. The steps taken include filtering the system's documents that have the same type of interrogative word as the question submitted to the system using regular expression, after which the similarity value is calculated using cosine similarity. The system documents that have the highest similarity value to the question asked will be taken as the result of the system's answer finding. The system documents used consist of a collection of questions and answers divided into three types of interrogative words: 'what', 'who', and 'where', all themed around Indonesian history. The number of system documents is 150 pairs of questions and answers, consisting of 50 pairs for the 'what' type, 50 pairs for the 'who' type, and 50 pairs for the 'where' type. The system documents were obtained from websites and then documented in .sql file format. The system testing was conducted based on three scenarios, differentiated by the three types of interrogative words, in predicting the answers to 30 test questions for each scenario. The system resulted in an average precision(0.7), recall(0.7), and F-1 score(0.7) in predicting correct answers. In predicting answers, the system still predicted some incorrect answers, which is due to the number of words in the system documents being greater than the number of words in the questions asked.

Keywords : *Answer Finding, Cosine Similarity, Indonesia Question Answering System, Regular Expression*

## ABSTRAK

*Question Answering System (QAS)* adalah sistem otomatis yang dirancang untuk memberikan jawaban yang tepat terhadap pertanyaan yang diajukan menggunakan bahasa alami. Penelitian ini bertujuan untuk mengetahui kemampuan sistem Indonesia QAS yang dibangun dalam melakukan *answer finding* atau memprediksi jawaban yang diajukan ke sistem menggunakan metode *regular expression* dan *cosine similarity*. Tahapan yang dilakukan adalah melakukan pemfilteran dokumen sistem yang memiliki jenis kata tanya yang sama dengan pertanyaan yang diajukan ke sistem menggunakan *regular expression* yang setelah itu akan dihitung nilai *similarity*-nya menggunakan *cosine similarity*. Dokumen sistem yang memiliki nilai *similarity* yang paling tinggi dengan pertanyaan yang diajukan akan diambil sebagai hasil *answer finding* sistem. Dokumen sistem yang digunakan berupa kumpulan pertanyaan dan jawaban yang terbagi menjadi 3 jenis kata tanya ‘apa’, ‘siapa’, dan ‘dimana’ dan bertema sejarah Indonesia. Jumlah dokumen sistem adalah 150 pasangan pertanyaan dan jawaban yang terdiri dari 50 pasang untuk jenis kata tanya ‘apa’, 50 pasang untuk jenis kata tanya ‘siapa’, dan , 50 pasang untuk jenis kata tanya ‘dimana’. Dokumen sistem diperoleh dari website yang kemudian didokumentasikan dalam bentuk file .sql. Pengujian sistem dilakukan berdasarkan 3 skenario yang dibedakan berdasarkan 3 jenis kata tanya dalam memprediksi jawaban pertanyaan yang diajukan sebanyak 30 pertanyaan uji pada masing-masing skenario. Sistem menghasilkan nilai rata-rata *precision(0,7)*, *recall(0,7)*, dan *F-1 score(0,7)* dalam memprediksi jawaban benar. Dalam memprediksi jawaban sistem masih memprediksi beberapa jawaban yang salah, ini disebabkan karena jumlah kata pada dokumen sistem lebih banyak daripada jumlah kata yang terdapat pada pertanyaan yang diajukan.

Kata Kunci : *Answer Finding, Cosine Similarity, Indonesia Question Answering System, Regular Expression*

## KATA PENGANTAR

Alhamdulillahirabbiláalamiin. Segala puji dan syukur penyusun haturkan kepada Allah SWT atas rahmat dan karunia-Nya penyusunan skripsi yang berjudul **“ANSWER FINDING PADA INDONESIA QUESTION ANSWERING SYSTEM MENGGUNAKAN REGULAR EXPRESSION DAN COSINE SIMILARITY”** dapat diselesaikan dengan baik. Skripsi ini dimaksudkan untuk memperoleh gelar sarjana strata satu (S1) Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu:

1. Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak M. Facurrozi, M.T.
2. Sekretaris Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya Bapak Osvari Arsalan, M.T. yang telah mengarahkan dan memotivasi penyusun untuk menyelesaikan penelitian ini.
3. Kedua orang tua, tercinta yang selalu memberikan dukungan, motivasi, dan semangat kepada penyusun.
4. Dosen Pembimbing 1, Ibu Novi Yusliani, S.Kom., M.T. yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memotivasi penyusun dalam menyelesaikan penelitian ini.



5. Dosen Pembimbing 2, Ibu Desty Rodiah, M.T. yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing, mengarahkan, serta memotivasi penyusun dalam menyelesaikan penelitian ini.
6. Seluruh dosen Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu pengetahuan kepada penyusun.
7. Teman-teman mahasiswa kelas TIBIL C Jurusan Teknik Informatika 2018 Fakultas Ilmu Komputer Universitas Sriwijaya yang selalu memberikan dukungan dan semangat kepada penyusun.

Penyusun sadar penelitian ini jauh dari kata sempurna dan masih terdapat kekurangan. Untuk itu penyusun memohon maaf serta menerima kritik dan saran yang dapat membantu membangun penelitian ini agar dapat dikembangkan menjadi lebih baik lagi.

Palembang, Juni 2024

Penyusun

## DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN.....	ii
TANDA LULUS SIDANG SKRIPSI.....	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT .....	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT .....	vi
ABSTRAK .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN.....	xv

### BAB I PENDAHULUAN

1.1	Pendahuluan .....	I-1
1.2	Latar Belakang Masalah .....	I-1
1.3	Rumusan Masalah.....	I-3
1.4	Tujuan Penelitian.....	I-3
1.5	Manfaat Penelitian.....	I-4
1.6	Batasan Masalah.....	I-4
1.7	Sistematika Penulisan .....	I-4
1.8	Kesimpulan .....	I-6

### BAB II KAJIAN LITERATUR

2.1	Pendahuluan .....	II-1
2.2	Landasan Teori.....	II-1
2.2.1	Question Answering System.....	II-1
2.2.2	Pre-Processing.....	II-2
2.2.3	Regular Expression .....	II-6

2.2.4	Cosine Similarity.....	II-6
2.2.5	Confusion Matrix .....	II-8
2.2.6	Metode Rational Unified Process.....	II-9
2.3	Penelitian Lain yang Relevan.....	II-10
2.4	Kesimpulan .....	II-11

### BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Question Answering System.....	III-1
3.2.2	Pre-Processing.....	III-1
3.3	Tahapan Penelitian .....	III-1
3.3.1	Kerangka Kerja Penelitian.....	III-1
3.3.2	Kriteria Pengujian .....	III-3
3.3.3	Format Data Pengujian .....	III-3
3.3.4	Alat Bantu Pelaksanaan Penelitian.....	III-4
3.3.5	Pengujian Penelitian .....	III-4
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan .....	III-5
3.4	Metode Pengembangan Perangkat Lunak.....	III-5
3.4.1	Inception .....	III-5
3.4.2	Elaboration.....	III-6
3.4.3	Construction.....	III-6
3.4.4	Trasition.....	III-6
3.5	Manajemen Proyek Perangkat Lunak .....	III-6
3.6	Kesimpulan .....	III-9

### BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan .....	IV-1
4.2	Metode Pengembangan Perangkat Lunak.....	IV-1
4.2.1	Tahap <i>Inception</i> .....	IV-1
4.2.2	Tahap <i>Elaboration</i> .....	IV-2

4.2.3 Tahap <i>Construction</i> .....	IV-8
4.2.4 Tahap <i>Trasition</i> .....	IV-11
4.3 Kesimpulan .....	IV-12

## BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan .....	V-1
5.2 Data Hasil Penelitian .....	V-1
5.2.1 Konfigurasi Percobaan .....	V-1
5.2.2 Data Hasil Konfigurasi I.....	V-1
5.2.3 Data Hasil Konfigurasi II.....	V-6
5.2.4 Data Hasil Konfigurasi III .....	V-10
5.3 Analisis Hasil Penelitian .....	V-15
5.4 Kesimpulan .....	V-19

## BAB VI KESIMPULAN DAN SARAN

6.1 Kesimpulan .....	VI-1
6.2 Saran .....	VI-1

DAFTAR PUSTAKA .....	xvi
----------------------	-----

## LAMPIRAN

## DAFTAR TABEL

	Halaman
II-1	Contoh Tahap <i>Case Folding</i> ..... II-3
II-2	Contoh Tahap <i>Tokenization</i> dan <i>Punctuation</i> ..... II-3
II-3	Contoh Tahap <i>Stemming</i> ..... II-4
II-4	Tabel <i>Confusion Matrix</i> ..... II-7
III-1	Format Data Pengujian ..... III-4
III-2	Alat Bantu Penelitian ..... III-4
III-3	Tabel Data Skenario Pengujian ..... III-5
III-4	Tabel Hasil Analisis Pengujian ..... III-5
III-5	Rencana Pengembangan Perangkat Lunak ..... III-7
IV-1	Kebutuhan Fungsional Sistem ..... IV-1
IV-2	Kebutuhan Non-Fungsional Sistem ..... IV-1
IV-3	Alat Bantu Penelitian ..... IV-2
IV-4	Skenario <i>Use Case</i> Memberikan Jawaban pada Pertanyaan Uji ..... IV-3
IV-5	Implementasi Kelas pada Kode Program ..... IV-8
IV-6	Alat Bantu Pengujian Sistem ..... IV-11
IV-7	Skenario Rencana Pengujian <i>Use Case</i> Memberikan Jawaban pada Pertanyaan Uji ..... IV-12
IV-8	Hasil Pengujian <i>Use Case</i> Memberikan Jawaban pada Pertanyaan Uji ..... IV-12
V-1	Hasil Pengujian Skenario I ..... V-2
V-2	Hasil Pengujian Skenario II ..... V-6
V-3	Hasil Pengujian Skenario III ..... V-11
V-4	Tabel Hasil Analisis Skenario I ..... V-15
V-5	Tabel Hasil Analisis Skenario II ..... V-16
V-6	Tabel Hasil Analisis Skenario III ..... V-17
V-7	Tabel Hasil Analisis Pengujian ..... V-18

## DAFTAR GAMBAR

	Halaman
II-1 Alur Pengembangan IQAS .....	II-2
III-1 Diagram Alur Proses Umum Sistem .....	III-2
IV-1 <i>Use Case Diagram</i> Perangkat Lunak .....	IV-2
IV-2 <i>Activity Diagram</i> Memberikan Jawaban pada Pertanyaan Uji .....	IV-4
IV-3 <i>Sequence Diagram</i> Memberikan Jawaban pada Pertanyaan Uji .....	IV-6
IV-4 <i>Class Diagram</i> Sistem.....	IV-7
IV-5 Rancangan Antarmuka Sistem.....	IV-8
IV-6 Implementasi Antarmuka Sistem.....	IV-11
V-1 Grafik Rentang Nilai <i>Similarity</i> Jawaban Prediksi Sistem per Skenario .....	V-18

## **DAFTAR LAMPIRAN**

- Lampiran 1. Kumpulan Data Dokumen Sistem dan Pertanyaan Uji
- Lampiran 2. Kode Program Kelas Main
- Lampiran 3. Kode Program Kelas Database
- Lampiran 4. Kode Program Kelas QASystem
- Lampiran 5. Kode Program Kelas Preprocessing
- Lampiran 6. Kode Program Kelas RegexFilter
- Lampiran 7. Kode Program Kelas CosineSimilarity
- Lampiran 8. Hasil Uji Kemiripan Dokumen
- Lampiran 9. Form Perbaikan Skripsi

# BAB I

## PENDAHULUAN

### 1.1 Pendahuluan

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah dan sistematika penulisan.

### 1.2 Latar Belakang Masalah

*Question Answering System (QAS)* adalah pendekatan otomatis untuk mendapatkan jawaban yang tepat untuk merespon pertanyaan yang diajukan oleh manusia dalam *natural language* atau bahasa alami. Dalam melakukan *answer finding* atau pencarian jawaban pada QAS metode statistikal paling sering digunakan untuk menemukan jawaban yang memiliki tingkat kemiripan yang tinggi. *Cosine similarity* sebagai salah satu metode statistikal memiliki kelebihan untuk menganalisa pertanyaan berdasarkan fitur *similarity* dalam menentukan kedekatan kandidat jawaban terhadap pertanyaan tersebut (Berger et al., 2000). Untuk menunjang fitur *answer finding* agar lebih maksimal, diperlukan metode untuk mengklasifikasikan pertanyaan seperti metode *pattern-matching*. *Regular expression* merupakan salah satu metode *pattern-matching* memiliki kelebihan untuk mengklasifikasikan pertanyaan berdasarkan kata tanya sehingga pengklasifikasian jawaban menjadi lebih mudah (Dwivedi & Singh, 2013).

Sejalan dengan semakin berkembangnya penelitian mengenai QAS semakin banyak pula metode yang yang dikembangkan maupun yang diterapkan.



Salah satunya, penelitian oleh Jovita et al. (2015) yang melakukan penelitian QAS menggunakan *Vector Space Model* (VSM) untuk mempresentasikan *knowledge* menjadi vektor, lalu menggunakan *cosine similarity* untuk membandingkan vektor dokumen dengan vektor *query*. Pada penelitian ini dijelaskan bahwa VSM dengan *cosine similarity* berhasil mengungguli penelitian sebelumnya yang menggunakan *rule-based*, N-gram, *template-driven response*, dan *reversible transformation* dengan *score* rata-rata *precision* (0,547), *recall* (0,662), dan *f-measure* (0,580).

Demikian juga penelitian yang dilakukan oleh Purbasari et al. (2018) yaitu penelitian mengenai QAS menggunakan VSM dengan *cosine similarity* yang memiliki dataset 100 pasangan pertanyaan dan jawaban. Metode VSM pada penelitian ini menggunakan TF-IDF sebagai metode untuk menghitung bobot masing-masing dokumen yang memiliki *term* yang sesuai dengan topik *domain*, kemudian *cosine similarity* akan menghitung jarak antara *query* dan dokumen. Hasil dari penelitian ini mempunyai *score* rata-rata *precision* (0,413), *recall* (0,893), dan *f-measure* (0,563). Pada penelitian ini juga menyebutkan bahwa jumlah dataset pertanyaan dan jawaban memengaruhi keandalan QAS.

Dalam penelitian lain yang ditulis oleh Kowsher et al. (2019) yaitu penelitian mengenai pengembangan *Bangla Intelligence Question Answering System* (BIQAS) dengan 3 prosedur matematika dan statistika yaitu *cosine similarity*, *Jaccard similarity*, dan algoritma Naive Bayes. Pada penelitian tersebut *cosine similarity* memperoleh presentasi akurasi jawaban tertinggi yaitu 93,22% dibandingkan *Jaccard similarity* (84,64%) dan algoritma Naive Bayes (91,31%).

Penelitian lainnya yang dilakukan oleh Cooper & Rüger (2001) yaitu

penelitian mengenai QAS menggunakan *regular expression* untuk mengklasifikasikan pertanyaan (*when, where, why, describe, define, who, whom, what, which, how*), setelah pertanyaan berhasil diklasifikasi maka tipe jawaban untuk pertanyaan tersebut baru bisa ditentukan. Selanjutnya jawaban-jawaban yang memenuhi tipe tersebut akan dicari bobotnya dan diranking menggunakan *heuristic score*.

Berdasarkan latar belakang tersebut, maka penelitian dilakukan untuk mengembangkan Indonesia *Question Answering System* menggunakan *regular expression* dan *cosine similarity* yang diharapkan dapat memberikan jawaban cepat, tepat, serta tanpa perantara.

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang di atas, adapun rumusan masalah yang akan diselesaikan dalam penelitian ini adalah sebagai berikut :

1. Bagaimana membangun IQAS untuk *answer finding* menggunakan *regular expression* dan *cosine similarity*?
2. Bagaimana kinerja metode *regular expression* dan *cosine similarity* untuk *answer finding* pada IQAS?

### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah diatas tujuan dilakukannya penelitian ini adalah sebagai berikut :

1. Menghasilkan IQAS untuk *answer finding* menggunakan *regular*

*expression* dan *cosine similarity*.

2. Mengetahui kinerja metode *regular expression* dan *cosine similarity* untuk *answer finding* pada IQAS.

### **1.5 Manfaat Penelitian**

Manfaat yang diberikan penelitian ini adalah :

1. Diharapkan bisa menjadi rujukan untuk penelitian beikutnya.
2. Memudahkan untuk menemukan jawaban mengenai Sejarah Indonesia.

### **1.6 Batasan Masalah**

Batasan yang diterapkan dalam melakukan penelitian ini adalah :

1. Sistem Indonesia QAS yang dibuat berupa *closed domain* untuk Sejarah Indonesia.
2. Pertanyaan yang diterima sistem harus mengandung kata tanya *apa*, *siapa*, dan *dimana*.

### **1.7 Sistematika Penulisan**

Sistematika penulisan skripsi ini adalah sebagai berikut :

#### **BAB.I PENDAHULUAN**

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah dan sistematika penulisan.

**BAB.II KAJIAN LITERATUR**

Pada bab kajian literatur ini akan membahas tentang dasar teori dalam menyelesaikan masalah – masalah pada penelitian ini. Kajian literatur bertujuan sebagai landasan dalam pemecahan masalah pada penelitian.

**BAB.III METODOLOGI PENELITIAN**

Pada bab metodologi penelitian ini akan membahas tahapan – tahapan apa saja yang akan dilakukan pada penelitian ini. Rencana tahapan penelitian akan dijelaskan secara rinci dengan mengacu pada kerangka kerja dan perancangan proyek perangkat lunak.

**BAB.IV PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini akan dibahas mengenai perancangan dan implementasi perangkat lunak pada penelitian.

**BAB.V HASIL DAN ANALISI PENELITIAN**

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

**BAB.VI KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan penelitian ini.

## 1.8 Kesimpulan

Pada bab ini disimpulkan bahwa penelitian ini dilakukan untuk membangun sebuah perangkat lunak berupa Indonesia *Question Answering System* menggunakan *regular expression* dan *cosine similarity*. Dengan dilakukannya penelitian ini, diharapkan perangkat lunak yang dibuat dapat memberikan informasi yang cepat, tepat serta tanpa perantara.

## DAFTAR PUSTAKA

- Berger, R. Caruana, D. Cohn, D. Freitag, and V. Mittal. 2000. Bridging the lexical chasm: statistical approaches to answer-finding. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval:192–199.
- Cooper, R.J., and Ruger, S.M. 2000. A Simple Question Answering System. Proceedings of the 9th Text REtrieval Conference.
- Dwivedi, Sanjay K., and Singh, Vaishali. 2013. Research and Reviews in Question Answering System. Procedia Technology(10):417-424.
- Jovita, Linda, A. Hartawan, and D. Suhartono. 2015. Using Vector Space Model in Question Answering System. Procedia Computer Science(59):305–311.
- Jurafsky, D., and Martin. 2009. Speech and Language Processing. Prentice Hall, Englewood Cliffs : New Jersey.
- Kohavi, Ron and Provost, Foster. 1998. “Glossary of Terms.” Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, Machine Learning(30):271–274, Kluwer Academic Publishes.
- Kowsher, et. al. 2019. Bangla Intelligence Question Answering System Based on Mathematics and Statistics.
- Kroll, P., et. al. 2003. The Rational Unified Process Made Easy : a practitioner’s guide to the RUP. Pearson Education : Canada
- Lende, Sweta P., and Raghuwanshi, M. M. 2016. Closed Domain Question Answering System Using NLP Techniques.

- Nazief, B., & Adriani, M. 2007. Stemming Indonesian: A Confix-Stripping Approach. *ACM Transactions on Asian Language Information Processing*.
- Purbasari, Intan Yuniar, et. al. 2018. Replies Identification in Question Answering System using Vector Space Model. *Proceedings of the International Conference on Science and Technology*.
- Shitu, Tanzim Tamanna, et. al. 2020. Domain Specific Factoid Question Answering by Regular Expression Generation.
- Soricut, Radu and Brill, E. 2006. Automatic question answering using the web: Beyond the Factoid.
- Yu, Hong, et. al. 2007. Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. *Journal of Biomedical Informatics*(40):236-251.