

**QUESTION ANSWERING SYSTEM MENGGUNAKAN
SEMANTIC WEB DAN ALGORITMA PORTER
SEBAGAI SISTEM MERKATA**

SKRIPSI

Oleh :

DZAKIYATUR ROSYIDAH

NIM . 10650077



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK
IBRAHIM MALANG**

2014

HALAMAN PENGANTAR

**QUESTION ANSWERING SYSTEM MENGGUNAKAN
SEMANTIC WEB DAN ALGORITMA PORTER
SEBAGAI SISTEM MERKATA**

SKRIPSI

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**oleh :
DZAKIYATUR ROSYIDAH
NIM . 10650077 / S-1**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2014**

HALAMAN PERSETUJUAN

QUESTION ANSWERING SYSTEM MENGGUNAKAN SEMANTIC WEB DAN ALGORITMA PORTER SEBAGAI STEMMER KATA

SKRIPSI

Oleh :

Nama : Dzakiyatur Rosyidah
NIM : 10650077
Jurusan : Teknik Informatika
Fakultas : Sains Dan Teknologi

Telah Diperiksa dan Disetujui:
Tanggal: 26 November 2014

Dosen Pembimbing I

Dosen Pembimbing II



A'la Syauqi, M.Kom

Fatchurrochman, M.Kom

NIP. 19771201 200801 1 007

NIP. 19700731 200501 1 002

Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. Cahyo Crys dian

NIP. 19740424 200901 1 008

HALAMAN PENGESAHAN

**QUESTION ANSWERING SYSTEM MENGGUNAKAN SEMANTIC WEB
DAN ALGORITMA PORTER SEBAGAI STEMMER KATA**

SKRIPSI

Oleh :

**Dzakiyatur Rosyidah
NIM. 10650077 / S-1**

Telah Dipertahankan Di Depan Dewan Penguji Skripsi
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)

Tanggal 26 November 2014

Susunan Dewan Penguji:

- | | | Tanda Tangan |
|--------------------|--|---|
| 1. Penguji Utama | : <u>Roro Inda Melani, M.Sc</u>
NIP. 19780925 200501 2 008 | () |
| 2. Ketua Penguji | : <u>Dr. M. Amin Hariyadi, M.T</u>
NIP. 19670118 200501 1 001 | () |
| 3. Sekretaris | : <u>A'la Syaqui, M.Kom</u>
NIP. 19771201 200801 1 007 | () |
| 4. Anggota Penguji | : <u>Fatchurrochman, M.Kom</u>
NIP. 19700731 200501 1 002 | () |

Mengetahui,
Ketua Jurusan Teknik Informatika



**Dr. Cahyo Crvsdian
NIP. 19740424 200901 1 008**

**HALAMAN PERNYATAAN
ORISINALITAS PENELITIAN**

Saya yang bertandatangan di bawah ini:

Nama : Dzakiyatur Rosyidah
NIM : 10650077
Fakultas/Jurusan : Sains Dan Teknologi / Teknik Informatika
Judul Penelitian : Question Answering System Menggunakan
Semantic Web dan Algoritma Porter sebagai
Stemmer Kata

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut

Malang, 11 November 2014

Yang Membuat Pernyataan,



DzakiyaturRosyidah

10650077

MOTTO

“Do your best at any moment
that you have, cause every
effort has it rewards”

“Tidak ada usaha yang sia-sia”

PERSEMBAHAN

Alhamdullillah, segala puji syukur atas rahmat, hidayah dan inayah Allah SWT, karya ini ku persembahkan untuk orang-orang istimewa dalam hidupku

| Ayahanda dan Ibunda Tercinta |

Harun Nur Rasyid dan Zulia ti

Terima kasih atas kasih sayang, semangat, dan keikhlasan hati telah mendidik dan membimbingku

| Adik-adikku Tersayang |

Ahmad Mubarakul Wajid

Faiz Nur Ihsan Arif

Fauzan Rosyidi

M. Rizqy Rosyid

Terima kasih atas semangat dan doanya

| Para Dosen TI UIN Maliki Malang |

Terima kasih telah membimbing dan mendidik saya semoga ilmu yang saya terima barokah dunia akhirat

| INFINITY (TI 2010) |

Kita lebih dari teman seperjuangan menuntut ilmu,

kita adalah keluarga

KATA PENGANTAR



Assalamua'alaikum W r. W b.

Segala puji bagi Allah SW T yang telah melimpahkan rahmat, hidayah dan karunia-Nya kepada penulis. Shalawat serta salam juga tetap tucurahkan kepada Nabi kita Muhammad SAW yang telah membimbing umatnya dari kegelapan menuju terangnya cahaya Islam. Sehingga penulis dapat menyelesaikan skripsi dengan judul "Question Answering System Menggunakan Semantic Web dan Algoritma Porter sebagai Stemmer Kata" dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. A'la Syauqi, M. Kom, selaku dosen pembimbing skripsi I, yang telah meluangkan waktu untuk membimbing, memotivasi, serta memberikan pengarahan, masukan dan pengalaman berharga.
2. Fatchurrohman, M.Kom, selaku dosen pembimbing skripsi II, yang telah memberikan saran dan masukan dalam penyusunan skripsi.
3. Dr. Cahyo Crysdiyan, selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, yang mendukung dan mengarahkan dalam pengerjaan skripsi ini.

4. Segenap sivitas akademika Jurusan Teknik Informatika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingan yang telah diberikan.

5. Semua pihak yang tidak mungkin penulis sebutkan satu-persatu, atas segala yang telah diberikan kepada penulis dan dapat menjadi pelajaran.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap semoga skripsi ini dapat memberikan manfaat kepada para pembaca. Amin Ya Rabbal Alamin.

W assalam u'alaikum W r. W b.

Malang, 9 November 2014

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
ABSTRAK	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Batasan Masalah	8
1.4 Tujuan Penulisan	9
1.5 Manfaat Penulisan	9
1.6 Sistem Penulisan	9
1.7 Metode Penelitian	10
BAB II TINJAUAN PUSTAKA	
2.1 Penelitian Terkait	11
2.2 <i>Semantic Web</i>	14
2.3 Komponen-komponen <i>Semantic Web</i>	15
2.4 <i>Ontology</i>	18
2.5 Temu Kembali Informasi (<i>Information Retrieval</i>)	19
2.6 <i>Question Answering</i>	20
2.7 DBpedia	21
2.7.1 <i>Linked Data</i>	21
2.7.2 <i>SPARQL Endpoint</i>	21
2.7.3 <i>RDF Dumps</i>	22
2.8 <i>Parsing</i>	22
2.9 <i>Stemming</i>	22
2.10 <i>SPARQL</i>	23
BAB III ANALISIS DAN PERANCANGAN SISTEM	
3.1 Desain Sistem	25
3.1.1 <i>Text Processing</i>	26
3.1.1.1 Kalimat Tanya	28
3.1.1.2 <i>Casefolding</i>	28
3.1.1.3 Menghapus Tanda Tanya	28
3.1.1.4 <i>Parsing</i>	28

3.1.1.5 Menghapus Stop words	29
3.1.1.6 Porter Stemming	29
3.1.2 Data Mining	34
3.1.2.1 Cari Resource di Wikipedia	35
3.1.2.2 SPARQL Query	35
3.1.3 Menampilkan Jawaban	36
3.2 Desain Interface	37
3.3 Kebutuhan Sistem	38
BAB IV HASIL DAN PEMBAHASAN	
4.1 Implementasi Sistem	40
4.1.1 Implementasi Porter Stemming	42
4.1.2 Implementasi Proses Pencarian Resource	45
4.1.3 Implementasi SPARQL pada DBpedia Indonesia	46
4.2 Implementasi Tampilan	47
4.3 Uji Coba	50
4.4 Analisa Hasil	51
4.5 Integrasi Question Answering System dengan Islam	68
BAB V PENUTUP	
5.1 Kesimpulan	71
5.2 Saran	71
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 1.1 Hasil Pencarian dari Yahoo <i>Search Engine</i>	3
Gambar 1.2 Hasil Pencarian dari Wolfram Alpha <i>Question Answering System</i>	5
Gambar 1.3 Hasil Perankingan Wikipedia di Alexa	7
Gambar 1.4 Demografi Pengguna di Wikipedia	7
Gambar 2.1 Layer <i>Semantic Web</i>	16
Gambar 2.2 <i>Meaning Triangle</i>	18
Gambar 3.1 Blok Diagram <i>Question Answering System</i>	26
Gambar 3.2 Blok Diagram <i>Text Processing</i>	27
Gambar 3.3 Blok Diagram <i>Porter Stemming</i>	30
Gambar 3.4 Blok Diagram <i>Data Mining</i>	35
Gambar 3.5 Struktur Kueri SPARQL	36
Gambar 3.6 Contoh Kueri SPARQL	36
Gambar 3.7 Hasil dari Kueri SPARQL	36
Gambar 3.8 Desain <i>Interface Question Answering System</i>	37
Gambar 4.1 Implementasi <i>User Interface</i>	48
Gambar 4.2 Sistem Menampilkan Jawaban kepada Pengguna	49

DAFTAR TABEL

Tabel 3.1 Aturan untuk <i>Inflection Particle</i>	31
Tabel 3.2 Aturan untuk <i>Inflection Possesive Pronoun</i>	31
Tabel 3.3 Aturan untuk <i>First Order Derivational Prefix</i>	31
Tabel 3.4 Aturan untuk <i>Second Order Derrivational Prefix</i>	32
Tabel 3.5 Aturan untuk <i>Derrivational Suffix</i>	32
Tabel 4.1 Rekapitulasi Uji Coba	51



ABSTRAK

Rosyidah, Dzakiyatur. 2014. **Question Answering System Menggunakan Semantic Web dan Algoritma Porter sebagai Stemmer Kata**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) A'la Syauqi, M. Kom (II) Fatchurrochman, M. Kom

Kata Kunci: *Natural Language Processing, Question Answering, Wikipedia, Dbpedia, SPARQL*

Question Answering System (QAS) merupakan bagian dari penelitian mengenai *Natural Language Processing (NLP)*, dimana sistem ini mengizinkan pengguna menyatakan kebutuhan informasinya dalam bentuk *natural language question* (pertanyaan dalam bahasa alami), dan mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban. Saat ini, kebutuhan untuk mendapatkan informasi secara cepat dan akurat semakin meningkat diiringi pula dengan peningkatan data yang ada pada web. Untuk ini pada penelitian ini, peneliti membangun sebuah *Question Answering System* menggunakan algoritma Porter sebagai stemmer kata. Algoritma Porter digunakan untuk stemming pada kalimat tanya yang diinputkan oleh pengguna. Sumber data yang akan digunakan adalah *DBpedia Indonesia*, dimana *DBpedia* merupakan basis data dari *Wikipedia* yang menampung jutaan informasi. Peneliti juga menggunakan *SPARQL* untuk melakukan kueri pada *DBpedia Indonesia* untuk memperoleh jawabannya. Dari uji coba yang telah dilakukan, *QAS* yang dibuat memberikan nilai akurasi sebesar 71,50%.

ABSTRACT

Rosyidah, Dzakiyatur. 2014. **Question Answering System Using Semantic Web and Porter Algorithm for Word Stemmer**. Thesis. Informatics Department of Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University, Malang.

Adviser: (I) A'la Syauqi, M.Kom, M.Kom (II) Fatchurrochman, M.Kom

Keywords: *Natural Language Processing, Question Answering, Wikipedia, Dbpedia, SPARQL*

Question Answering System (QAS) is part of the research on the Natural Language Processing (NLP), where the system is allow users to express their information needs in the form of natural language question (questions in natural language), and returns the text or a phrase as the answer. Currently, the need to get information quickly and accurately increasingly accompanied by an increase in the existing data on the web. For this study, researchers are building a Question Answering System using the Porter algorithm for stemming. Porter stemming algorithm is used to stem sentence entered by the user. Source of data to be used is the DBpedia Indonesia, where DBpedia is a database that holds millions of Wikipedia information. Researchers also use SPARQL to query on DBpedia Indonesia to get the answer. From the experiments that have been carried out, QAS give an accuracy 71,50 % .

الملخص

الرشيدة، ذكية. 2014. النظام السؤال والجواب عن طريق الدلالي الويب والخوارزميات بورتر منالة تغيير الكلمة (steemer) أطروحة. قسم المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: أعلى شوقي الماجستير وفتح الرحمن الماجستير

الكلمات البحث: معالجة اللغات الطبيعية، السؤال الجواب، ويكيبيديا، SPARQL ، Dbpedia

النظام السؤال والجواب (QAS) هو جزء من البحث حول معالجة اللغات الطبيعية (NLP)، حيث النظام هو السماح للمستخدمين بالتعبير عن احتياجاتهم من المعلومات في شكل معالجة اللغة الطبيعية (الأسئلة في اللغة الطبيعية)، وإرجاع مقترحات نصية قصيرة أو حتى عبارة ما الجواب. في الوقت الحاضر، والحاجة إلى الحصول على المعلومات بسرعة وبدقة زاد تيرافق زيادة في البيانات الموجودة على شبكة الإنترنت. لهذه الدراسة، وبني الباحث النظام السؤال والجواب باستخدام خوارزمية بورتر من الة تغيير الكلمة. الخوارزمية بورتر النابعة المستخدمة للجملة لاستفهام المدخلة بالمستخدم. مصدر البيانات لاستخدامها هو Dbpedia اندونيسيا، حيث Dbpedia هي قاعدة بيانات من ويكيبيديا التي تحمل ملايين من المعلومات. الباحث أيضا استخدام SPARQL الاستعلام على Dbpedia اندونيسيا للحصول على الجواب. من التجارب التي تم القيام بها، الأمر الذي جعل QAS تعطي دقة 71.50%.

الملخص

الرشيدة، ذكية. 2014. النظام السؤال والجواب عن طريق الدلالي الويب والخوارزميات بوتر من الة تغير الكلمة (stemmer) أطروحة. قسم المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: أعلى شوقي الماجستير و فتح الرحمن الماجستير

الكلمات البحث : معالجة اللغات الطبيعية، السؤال الجواب، ويكيبيديا، Dbpedia، SPARQL، النظام السؤال والجواب (QAS) هو جزء من البحث حول معالجة اللغات الطبيعية (NLP) ، حيث النظام هو السماح للمستخدمين التعبير عن احتياجاتهم من المعلومات في شكل معالجة اللغة الطبيعية (الأسئلة في اللغة الطبيعية)، وإرجاع مقتطفات نصية قصيرة أو حتى عبارة ما الجواب. في الوقت الحاضر، والحاجة إلى الحصول على المعلومات بسرعة وبدقة زادت يرافقه زيادة في البيانات الموجودة على شبكة الإنترنت. لهذه الدراسة، وبنى الباحث النظام السؤال والجواب باستخدام خوارزمية بوتر من الة تغير الكلمة. الخوارزمية بوتر النابعة المستخدمة للجملة الاستفهام المدخلة بالمستخدم. مصدر البيانات لاستخدامها هو DBpedia اندونيسيا، حيث DBpedia هي قاعدة بيانات من ويكيبيديا التي تحمل الملايين من المعلومات. الباحث أيضا استخدام SPARQL الاستعلام على DBpedia اندونيسيا للحصول على الجواب. من التجارب التي تم القيام بها، الأمر الذي جعل QAS تعطي دقة 71.50٪.

ABSTRACT

Rosyidah, Dzakiyatur. 2014. **Question Answering System Using Semantic Web and Porter Algorithm for Word Stemmer**. Thesis. Informatics Department of Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University, Malang.

Adviser: (I) A'la Syauqi, M.Kom, M.Kom (II) Fatchurrochman, M.Kom

Keywords: *Natural Language Processing, Question Answering, Wikipedia, Dbpedia, SPARQL*

Question Answering System (QAS) is part of the research on the Natural Language Processing (NLP), where the system is allow users to express their information needs in the form of natural language question (questions in natural language), and returns the text or a phrase as the answer. Currently, the need to get information quickly and accurately increasingly accompanied by an increase in the existing data on the web. For this study, researchers are building a Question Answering System using the Porter algorithm for stemming. Porter stemming algorithm is used to stem sentence entered by the user. Source of data to be used is the DBPedia Indonesia, where DBPedia is a database that holds millions of Wikipedia information. Researchers also use SPARQL to query on DBPedia Indonesia to get the answer. From the experiments that have been carried out, QAS give an accuracy 71,50 %.

ABSTRAK

Rosyidah, Dzakiyatur. 2014. **Question Answering System Menggunakan Semantic Web dan Algoritma Porter sebagai Stemmer Kata**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) A'la Syauqi, M. Kom (II) Fatchurrochman, M.Kom

Kata Kunci: *Natural Language Processing, Question Answering, Wikipedia, Dbpedia, SPARQL*

Question Answering System (QAS) merupakan bagian dari penelitian mengenai *Natural Language Processing (NLP)*, dimana sistem ini mengizinkan pengguna menyatakan kebutuhan informasinya dalam bentuk natural language question (pertanyaan dalam bahasa alami), dan mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban. Saat ini, kebutuhan untuk mendapatkan informasi secara cepat dan akurat semakin meningkat diiringi pula dengan peningkatan data yang ada pada web. Untuk ini pada penelitian ini, peneliti membangun sebuah *Question Answering System* menggunakan algoritma Porter sebagai stemmer kata. Algoritma Porter digunakan untuk stemming pada kalimat tanya yang diinputkan oleh pengguna. Sumber data yang akan digunakan adalah DBPedia Indonesia, dimana DBPedia merupakan basis data dari Wikipedia yang menampung jutaan informasi. Peneliti juga menggunakan SPARQL untuk melakukan kueri pada DBPedia Indonesia untuk memperoleh jawabannya. Dari uji coba yang telah dilakukan, QAS yang dibuat memberikan nilai akurasi sebesar 71,50%.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dewasa ini internet telah tumbuh menjadi media elektronik yang menyediakan berbagai kebutuhan informasi untuk segala bidang keahlian. Sebagian besar masyarakat di dunia memanfaatkan internet sebagai sarana menuntut ilmu. Agama Islam juga mengajarkan umatnya untuk senantiasa menuntut ilmu, bahkan jika menginginkan kebahagiaan dunia akhirat maka wajib menuntut ilmu, seperti hadits yang diriwayatkan oleh Turmudzi berikut ini.

مَنْ أَرَادَ الدُّنْيَا فَعَلَيْهِ بِالْعِلْمِ, وَمَنْ أَرَادَ الْآخِرَةَ فَعَلَيْهِ بِالْعِلْمِ, وَمَنْ أَرَادَهُمَا فَعَلَيْهِ بِالْعِلْمِ

Artinya: “Barang siapa yang menghendaki kehidupan dunia maka wajib baginya memiliki ilmu, dan barang siapa yang menghendaki kehidupan akhirat, maka wajib baginya memiliki ilmu, dan barang siapa menghendaki keduanya maka wajib baginya memiliki ilmu”. (HR. Turmudzi)

Dari hadits tersebut jelas bahwa sebagai umat muslim wajib untuk menuntut ilmu. Ada berbagai cara untuk memperkaya ilmu pengetahuan dan memperoleh informasi, antara lain dengan membaca, berguru kepada seorang ahli, ataupun secara otodidak melalui media elektronik, seperti internet, televisi, radio, dan sebagainya. Semua hal tersebut dapat dilakukan sebagai media pembelajaran untuk memperkaya ilmu pengetahuan dan informasi. Di dalam Al Qur'an surat Al 'Alaq ayat 3 – 5 juga termaktub ayat mengenai media pembelajaran sebagai berikut.

أَقْرَأُ وَرَبُّكَ الْأَكْرَمُ ﴿١﴾ الَّذِي عَلَّمَ بِالْقَلَمِ ﴿٢﴾ عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ ﴿٣﴾

Artinya:

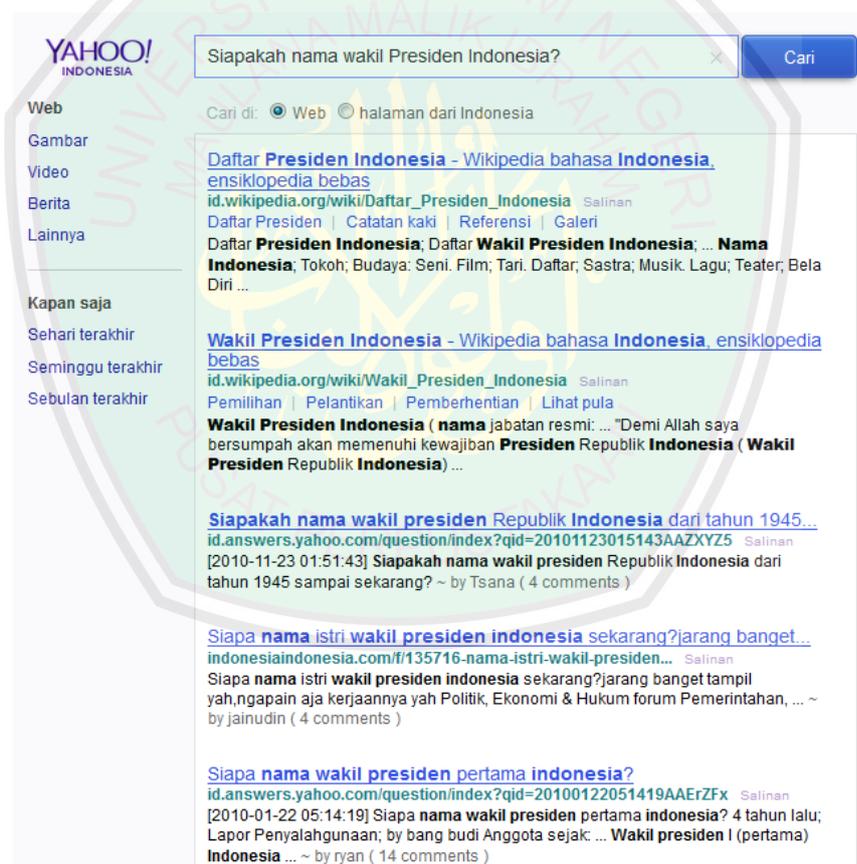
3. Bacalah, dan Tuhanmulah yang Maha pemurah,
4. Yang mengajar (manusia) dengan perantaran kalam,
5. Dia mengajar kepada manusia apa yang tidak diketahuinya.

Menurut terjemah tafsir jalalain, makna bacalah pada lafal ayat ini ialah mengukuhkan makna lafal pertama yang sama (dan Rabbmulah Yang Paling Pemurah) artinya tiada seorang pun yang dapat menandingi kemurahan-Nya. Lafal ayat ini sebagai Haal dan Dhamir yang terkandung di dalam lafal Iqra'. Selanjutnya, terjemah dari ayat keempat adalah yang mengajar (manusia) dengan perantaraan kalam, manusia pertama yang menulis dengan memakai kalam atau pena ialah Nabi Idris a.s. Pada ayat kelima, (Dia mengajarkan kepada manusia) atau jenis manusia (apa yang tidak diketahuinya) yaitu sebelum Dia mengajarkan kepadanya hidayah, menulis dan berkreasi serta hal-hal lainnya (As Suyuthi, 2008).

Dari ayat tersebut, dapat diilhami bahwa proses pembelajaran atau proses pentransferan pengetahuan kepada manusia dari yang semula tidak tahu menjadi tahu menggunakan perantara berupa pena. Pena yang dimaksud disini adalah baca tulis. Sehingga ayat ini mengandung makna bahwa penyampaian ilmu pengetahuan kepada manusia tidak secara langsung begitu saja. Akan tetapi Allah memberikan pengetahuan melalui perantara. Sehingga Allah sudah

mengisyaratkan melalui surat Al ‘Alaq ini bahwa penggunaan media pembelajaran untuk pentransferan ilmu pengetahuan memang penting.

Mesin pencari (*search engine*) merupakan salah satu fasilitas yang sering digunakan untuk mencari informasi, contohnya: Google, Yahoo, Altavista, Bing dan lain sebagainya. Cara kerja *search engine* adalah sistem akan menampilkan daftar dokumen yang berkaitan dengan *keyword* yang telah diinput oleh pengguna (Gunawan, 2006). Gambar 1.1 menunjukkan daftar dokumen dari hasil pencarian menggunakan *search engine*.



Gambar 1.1 Hasil Pencarian dari Yahoo Search Engine
(sumber: Yahoo, 10 Maret 2014)

Search engine tidak dapat melakukan analisis linguistik terhadap teks kueri secara mendalam. *Search engine* akan menampilkan daftar dokumen atau website. Untuk mendapatkan informasi yang benar-benar dibutuhkan, pengguna harus memeriksa dan membaca setiap dokumen di daftar tersebut. Tidak jarang sebagai pengguna terkadang mengalami kesulitan dalam menemukan informasi yang tepat sesuai dengan *keyword* dicari. Hal ini dipengaruhi oleh tingginya peningkatan data pada web, sehingga semakin banyak informasi dan pengetahuan atau data yang ada pada internet, maka *search engine* akan mengembalikan daftar dokumen yang banyak pula. Hal yang demikian kurang efisien dari segi waktu dan akurasi informasi yang dibutuhkan pengguna.

Dengan demikian, maka perlu dibangun sebuah *Question Answering System* (QAS). Ada beberapa faktor yang mendorong dikembangkannya *Question Answering System* (QAS), antara lain: tingginya peningkatan data web, kemajuan informasi dan teknologi, dan banyaknya permintaan dari pengguna untuk mendapatkan informasi dengan cepat dan akurat (Tahri, 2013, hal. 23). Gambar 1.2 menunjukkan hasil pencarian menggunakan *Question Answering System* yaitu Wolfram Alpha. Wolfram Alpha merupakan salah satu contoh *Question Answering System* menggunakan bahasa Inggris.

Enter what you want to calculate or know about:

who is agnes monica? ☆

Examples Random

Assuming "agnes monica" is a person | Use "agnes" as a given name instead

Input interpretation:
Agnes Monica (singer/songwriter, etc.)

Basic information:
 date of birth Tuesday, July 1, 1986 (age: 27 years)

Image:

 No results available.
 Try your query again.

Gambar 1.2 Hasil Pencarian dari Wolfram Alpha *Question Answering System* (sumber: Wolfram Alpha, 10 Maret 2014)

Question Answering System adalah sistem yang memungkinkan pengguna menyatakan kebutuhannya dalam bentuk *natural language question* (pertanyaan dalam bahasa alami), dan mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban (Gunawan, 2006). *Question Answering System* (QAS) atau sistem tanya jawab berupaya mencari jawaban spesifik terhadap pertanyaan yang diajukan dalam bahasa natural. Dibandingkan dengan mesin pencari (*search engine*) yang memberikan himpunan dokumen hasil penelusuran, sebuah QAS harus dapat menghadirkan jawaban akhir kepada pengguna. Secara nilai kecerdasan, QAS harus memiliki intelegensi yang lebih dibandingkan dengan mesin pencari biasa (Toba, 2010).

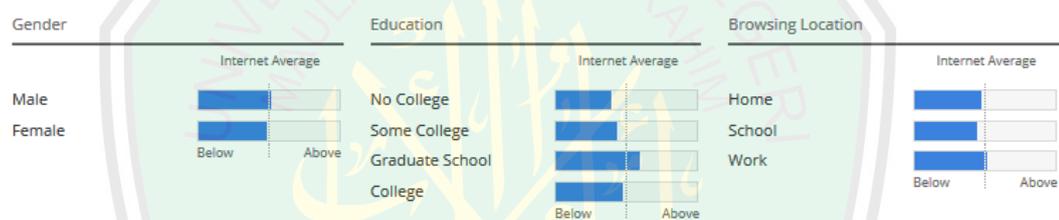
Ada beberapa penelitian yang mengembangkan QAS dengan menggunakan tema atau topik tertentu, misalnya: QAS seputar NBA (*National Basketball Association*). QAS ini dibangun oleh Calvin Irwan, Dr. Eng. Ayu Purwarianti, S.T, M. T. QAS ini mempunyai pengetahuan (*knowledge*) seputar liga bola basket pria di Amerika Serikat yang paling bergengsi di dunia. QAS lain dikembangkan oleh Gunawan dan Gita Lovina. QAS ini menggunakan Alkitab sebagai pengetahuan (*knowledge*). Peneliti ingin mengembangkan sebuah QAS berbahasa Indonesia yang dapat menjawab pertanyaan-pertanyaan umum sehingga QAS yang akan dibangun tidak terpaku pada topik atau tema tertentu seperti pada penelitian sebelumnya.

Dengan adanya data yang besar pada web dikombinasikan dengan teknologi *Information Retrieval* (IR) dan *Natural Language Processing* (NLP) dapat dikembangkan sebuah *Question Answering System* menggunakan Wikipedia sebagai basis pengetahuan atau *knowledge*-nya. Wikipedia digunakan sebagai basis pengetahuan karena merupakan *free* ensiklopedia terbesar yang menampung berjuta-juta informasi yang dikelola oleh berjuta-juta kontributor dari seluruh dunia. Berdasarkan hasil perankingan di Alexa.com, Wikipedia berada di peringkat keenam secara global dan di United States (US). Wikipedia masuk ke dalam *top 10 sites*. Hasil perankingan tersebut dapat dilihat pada Gambar 1.3.



Gambar 1.3 Hasil Perankingan Wikipedia di Alexa
(sumber: Alexa, 10 Maret 2014)

Jumlah pengguna yang mengakses Wikipedia juga beragam, datanya dapat dilihat demografinya berdasarkan jenis kelamin, tingkat pendidikan dan lokasi pencarian (*browsing*) pada Gambar 1.4.



Gambar 1.4 Demografi Pengguna di Wikipedia
(sumber: Alexa, 10 Maret 2014)

Berdasarkan data tersebut, maka Wikipedia dipilih sebagai *knowledge* dari *Question Answering System* (QAS) yang akan dibuat. *Resource* yang akan digunakan yaitu DBPedia. DBPedia adalah sebuah komunitas yang bergerak untuk mengekstrak informasi terstruktur dari Wikipedia dan menyediakan informasi tersebut dalam sebuah web. DBPedia yang akan digunakan pada penelitian ini adalah DBPedia Indonesia. DBPedia Indonesia merupakan web yang menyediakan data hasil ekstraksi dari Wikipedia Indonesia. Sehingga dapat dilakukan pencarian yang kompleks terhadap data di Wikipedia. Ada 3 cara yang dapat dilakukan untuk mengakses *knowledge* dari DBPedia Indonesia, yaitu

Linked Data, *SPARQL Endpoint* dan *RDF Dumps*. *Linked Data* adalah mengakses sebuah entitas dengan menggunakan URI *resource* dari entitas tersebut, contoh URI *resource* yaitu <http://id.dbpedia.org/resource/Indonesia>. *SPARQL Endpoint* adalah mengakses DBPedia dengan cara melakukan *query* pada *knowledge base*. Sedangkan *RDF Dumps* merupakan hasil dari ekstraksi Wikipedia Indonesia, dimana ini terbagi menjadi 26 berkas dalam format *N-Triple*. Pada pembuatan QAS ini akan dilakukan pengaksesan DBPedia Indonesia dengan menggunakan *SPARQL Endpoint*. Dengan adanya QAS ini diharapkan dapat membantu pengguna untuk menemukan informasi dengan cepat dan akurat.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini sebagai berikut:

- a. Apakah *Question Answering System* yang dibuat dapat memberikan jawaban yang akurat?
- b. Berapa tingkat akurasi jawaban yang dihasilkan dari *Question Answering System* yang dibuat?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut:

- a. Pengetahuan atau *knowledge* yang digunakan untuk sumber jawaban adalah DBPedia
- b. Pertanyaan yang diajukan oleh pengguna adalah sesuatu mengenai Indonesia

1.4 Tujuan Penelitian

Tujuan pada penelitian ini sebagai berikut:

- a. Untuk membangun *Question Answering System* menggunakan algoritma Porter sebagai *stemmer* kata.
- b. Untuk mengukur tingkat akurasi jawaban yang dihasilkan pada *Question Answering System*.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat membantu pengguna untuk mendapatkan informasi secara tepat dan akurat melalui *Question Answering System* yang telah dibuat.

1.6 Sistematika Penulisan

Penulisan skripsi ini tersusun dalam lima bab dengan sistematika penulisan sebagai berikut :

BAB I Pendahuluan

Pendahuluan, membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penyusunan tugas akhir, metodologi, dan sistematika penyusunan tugas akhir.

BAB II Landasan Teori

Landasan teori berisikan beberapa teori yang mendasari dalam penyusunan tugas akhir ini. Adapun yang dibahas dalam bab ini adalah dasar teori yang berkaitan dengan pembahasan tentang penelitian terkait, *semantic web*,

komponen-komponen *semantic web*, *ontology*, temu kembali informasi (*Information Retrieval*), *Question Answering*, *DBPedia*, *parsing*, *stemming*, dan *SPARQL*.

BAB III Analisa dan Perancangan

Menganalisa kebutuhan sistem untuk membangun *Question Answering System* dan langkah-langkah pembuatannya.

BAB IV Implementasi dan Pembahasan

Menjelaskan tentang implementasi sistem berdasarkan rancangan yang telah disusun sebelumnya serta pengujian *Question Answering System* yang telah dibuat beserta pembahasannya.

BAB V Penutup

Berisi kesimpulan dan saran.

1.7 Metode Penelitian

Berdasarkan jenis dan analisis data, penulisan skripsi ini menggunakan metode penelitian kuantitatif untuk mengukur akurasi jawaban dari QAS yang akan dibuat.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Sebelumnya pernah dilakukan penelitian mengenai QAS dengan menggunakan metode *Rule-Based* pada terjemahan Al Qur'an surat Al Baqarah oleh Meynar pada tahun 2007. Proses awal pada QAS adalah dengan mengambil dokumen-dokumen (ayat-ayat) berekstensi (*.txt) yang terdapat pada satu direktori dan menyimpan nama masing-masing dokumen dalam tabel penyimpanan dokumen. Kemudian dokumen-dokumen tersebut dipecah menjadi kalimat-kalimat yang dimasukkan ke dalam tabel penyimpanan kalimat. Selanjutnya dilakukan *parsing* dengan menggunakan titik (.) sebagai pemisah (*separator*), penghilangan *stopwords*, dan *stemming* yang akan menghasilkan token-token kalimat. Token-token tersebut disimpan dalam tabel penyimpanan. Pengguna memasukkan kueri berupa kalimat pertanyaan, selanjutnya dilakukan *parsing*, penghilangan *stopwords*, dan *stemming* yang akan menghasilkan token-token kueri. Kemudian dilakukan proses *WordMatch* yaitu membandingkan token-token pada setiap kalimat dokumen dan kalimat kueri. Setiap token yang sama dari kalimat kueri akan memberikan nilai *clue* (+3) pada kalimat dokumen tersebut. Selanjutnya masuk ke dalam proses *rule* sesuai dengan tipe kueri yang diberikan. Dari proses tersebut, masing-masing kalimat akan memperoleh nilai berdasarkan nilai *WordMatch* dan *rule*. Kalimat yang dikembalikan sebagai jawaban adalah kalimat yang memiliki nilai paling tinggi. Akurasi rata-rata *rule*

terhadap kueri oleh peneliti adalah 85,69 % dan akurasi rata-rata *rule* terhadap kueri yang diberikan pengguna umum adalah 53,14 %. Kelebihan dari penelitian ini antara lain: sistem ini berbasis web, kuerinya menggunakan bahasa alami, dan *indexing* dokumen hanya dilakukan sekali ketika di awal pembangunan sistem. Adapun kekurangan dari penelitian ini antara lain: jika terdapat penambahan dokumen, maka harus dilakukan proses pengindeksan ulang, tidak dilakukan kajian terhadap hubungan makna semantik dalam dokumen, sehingga kandungan terjemahan yang bersifat prosedural belum dapat ditemukembalikan, dan penelitian ini belum menggunakan *thesaurus* dan hanya menggunakan *simple matching* (Dwi, 2007).

Penelitian seputar QAS lainnya yaitu penelitian mengenai aplikasi *Question Answering* (QA) tentang *National Basketball Association* (NBA). QA ini merupakan aplikasi yang menjawab pertanyaan seputar NBA. Aplikasi ini dibangun dengan 4 modul, yaitu antarmuka, *Question Analyzer*, *Query Processor*, dan *Database Generator*. Teknik *Rule Based* diterapkan pada *Database Generator* untuk pembuatan aturan-aturan yang memanfaatkan teknik pencocokan string dilanjutkan dengan pemotongan string untuk pengambilan data basis data. Pada modul *Question Analyzer*, digunakan pencocokan string untuk mendeteksi *stopwords*, *Expected Answer Type* (EAT), dan *keywords*. Selanjutnya pada modul *Query Processor* digunakan query basis data biasa. *Query* menggunakan EAT sebagai kolom elemen tabel dan *keywords* untuk dicocokkan pada elemen tabel untuk pengembalian jawaban. Hasil pengujian dari aplikasi ini menunjukkan bahwa akurasi kebenaran aplikasi berdasarkan skenario uji mencapai 93,18%,

sementara hasil pengujian responden menunjukkan akurasi 91, 67% (Irwan, 2012).

Penelitian lain seputar *Question Answering* yaitu *Question Answering System* pada Alkitab menggunakan metode *Rule Based*. Arsitektur aplikasinya menggunakan arsitektur umum dari QAS yang dibuat oleh Monz, yaitu *Question Analysis*, *Document Collection Preprocessing*, *Candidate Document Selection*, *Candidate Document Analysis*, *Answer Extraction*, dan *Response Generation* (Monz, 2003). Pada tahap *Question Analysis* menghasilkan dua buah representasi pertanyaan sebagai output. Representasi pertama berupa sebuah *query* yang akan diteruskan pada tahap *Candidate Document Selection* dan representasi kedua adalah representasi semantik yang mengarah pada jenis informasi atau tipe jawaban yang diinginkan oleh pertanyaan yang dijadikan input. Tahap *Candidate Document Selection* akan mengidentifikasi dokumen-dokumen yang mengandung jawaban menggunakan metode *SurfaceText Patterns*. Lalu tahap *Candidate Document Analysis* dilakukan untuk memperkecil ukuran dokumen. Kemudian tahap *Answer Extraction* bertugas mencocokkan kandidat dokumen dengan representasi semantik dari pertanyaan sehingga menghasilkan daftar jawaban. Selanjutnya menyajikan jawaban, contoh penyajiannya bisa berupa daftar dokumen, daftar paragraf, daftar kalimat, atau daftar frase jawaban sebagai respon. Berdasarkan hasil pengujian melalui kuisisioner, prosentase rata-rata evaluasi program adalah 77.2% (Handojo, 2012).

Penelitian terkait lainnya yaitu sebuah arsitektur baru untuk mengembangkan *factoid Question Answering System* menggunakan DBPedia

ontologi dan DBPedia *extraction framework*. Arsitektur dari *Question Answering System* ini terdiri dari 3 proses utama, yaitu *Question Classification and Decision Model Generation*, *Question Processing* dan *Query Formulation and Execution*. *Question Classification* adalah tahap awal untuk mengklasifikasikan pertanyaan. Ada 7 kelas utama untuk klasifikasi pertanyaan, yaitu *human*, *location*, *entity*, *description*, *abbreviation*, *number*, dan *boolean*. Karena ada beberapa kelas untuk mengklasifikasikan pertanyaan, maka peneliti menggunakan algoritma *Support Vector Machine* (SVM) untuk *multi-class SVM* yaitu *Directed Acrylic Graph*. Tahap selanjutnya yaitu memproses pertanyaan, mengekstraksi *resource*, mengekstraksi *keywords*, formulasi *query* menggunakan SPARQL dan eksekusi. Kernel SVM yang digunakan pada penelitian ini adalah linear karena kernel ini mampu bekerja dengan cepat dalam menangani data yang besar. Penggunaan kernel ini memiliki tingkat akurasi 87% untuk pengklasifikasian pertanyaan. Dari penelitian yang dilakukan, masih ada beberapa kekurangan, yaitu kurangnya informasi yang tersedia pada DBPedia ontologi dan terkadang terjadi kesalahan dalam pengklasifikasian pertanyaan (Tahri, 2013).

2.2 Semantic Web

Semantik di dalam bahasa Indonesia berasal dari bahasa Inggris *semantics*, dari bahasa Yunani *sema* (nomina) ‘tanda’: atau dari verba *samaino* ‘menandai’, ‘berarti’. Istilah tersebut digunakan para pakar bahasa untuk menyebut bagian Ilmu bahasa yang mempelajari tentang makna (Djajasudarman, 1993).

Sedangkan web sendiri yaitu bagian visual dari internet. Web didasarkan pada prinsip-prinsip *Web page* multimedia yang bentuknya seperti halaman-

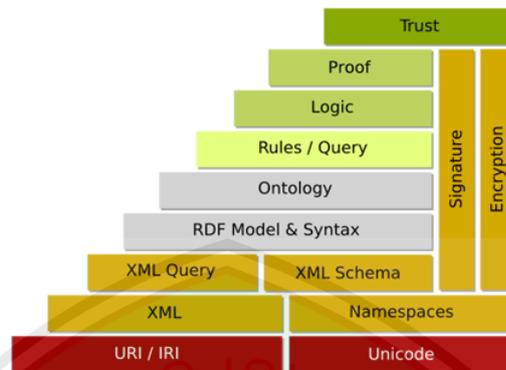
halaman majalah yang dilengkapi dengan gambar, teks, dan foto. Pembuat web, *European Particle Physics Laboratory* (CERN) mengembangkan bahasa komputer yang disebut dengan *Hypertext Markup Language* atau HTML, yang mana menjadi dasar dari web (Ause, 1995).

Semantic Web merupakan perluasan dari *World Wide Web* yang memungkinkan orang untuk berbagi konten melampaui batas-batas suatu aplikasi dan website. *Semantic web* merupakan web data yang memungkinkan mesin untuk memahami suatu informasi pada *World Wide Web* sesuai maknanya atau secara struktural. Ini dapat memperluas jaringan *hyperlink* suatu halaman web dengan memasukkan mesin pembaca metadata mengenai halaman yang terdapat dalam web tersebut dan bagaimana mereka saling berhubungan satu sama lain. Hal tersebut memungkinkan suatu agen untuk otomatis mengakses web secara lebih cerdas dan melakukan tugas-tugas atas nama pengguna (Berners, 2001). *Semantic web* sedikit berbeda apabila digambarkan seperti web pada umumnya. Penggunaan *semantic web* merupakan suatu pergeseran paradigma saat ini dalam penggunaan web sehari-hari. Maka dari itu, *semantic web* telah mengilhami banyak orang untuk menciptakan aplikasi dan teknologi semantik yang inovatif.

2.3 Komponen-komponen *Semantic Web*

Terdapat beberapa standar yang memungkinkan dalam pembuatan *Semantic Web* yang telah dikoordinasi oleh *World Wide Web Consortium* (W3C). Beberapa standar pokok dalam pembangunan suatu *Semantic Web* adalah *Extensible Markup Language* (XML), XML Schema, *Resource Description Framework* (RDF), RDF Schema, *Web Ontology Language* (OWL), dan SPARQL. Layer dari

Semantic Web sebagaimana direkomendasikan oleh W3C dapat dilihat pada Gambar 2.1 (Koivunen, 2002).



Gambar 2.1 Layer *Semantic Web*

a. *Unicode* dan URI

Unicode adalah standar representasi karakter komputer. Dalam hal ini, *Semantic Web* seharusnya dapat merepresentasikan dokumen dalam bahasa manusia yang berbeda dengan pemahaman mesin. URI (*Uniform Resource Identifier*) merupakan standar untuk lokasi dan identitas suatu *resource* (misalnya *web page*) dimana URI sebagai suatu penamaan unik untuk identifikasi *semantic web*.

b. XML dan *Namespaces*

Extensible Markup Language (XML) dan *Namespaces* serta *Schema*, merupakan sintaks yang berfungsi untuk menyajikan struktur data pada web. XML sendiri merupakan *Markup Language* yang memungkinkan penciptaan dokumen-dokumen yang tersusun dari struktur data dan *Namespaces* yang menyediakan caranya untuk me-*markup* dari banyak *sources*. *Semantic web*

merupakan sistem yang menghubungkan banyak data sehingga memerlukan *source* dalam pengolahan dokumen-dokumen tersebut.

c. RDF dan RDF *Schema*

Resource Description Framework (RDF) adalah sebuah *framework* yang dibuat oleh W3C untuk merepresentasikan informasi dengan menggunakan sekumpulan format sintaks. Ide dasar dari RDF adalah bagaimana dapat membuat pernyataan mengenai sebuah *resource* Web dalam bentuk ekspresi tertentu.

“Subjet-Predikat-Objek”. Dalam *terminology* RDF, SPO ini seringkali disebut dengan istilah *N-triple*. Subjek mengacu pada *resource* yang ingin dideksripsikan. Predikat menggambarkan kelakuan atau karakteristik dari *resource* tersebut dan mengekspresikan hubungan antara subjek dan objek.

RDF Schema merupakan sebuah kamus data yang menyediakan dasar-dasar *vocabulary* untuk mendeskripsikan *properties* dan *classes* dari *resources* RDF. Hal ini dapat digunakan dalam pembuatan hirarki *properties* dan *classes*-nya.

d. *Query*

Sebuah *query* digunakan untuk mengakses data pada *repository* yang tersedia.

e. *Ontology Vocabulary*

Bahasa ontology yang direkomendasikan oleh W3C pada 10 Februari 2004 adalah OWL *Web Ontology Language*, merupakan bahasa yang lebih kaya dan kompleks untuk mendeskripsikan *resource*.

f. *Logic dan Proof*

Layer ini berupa *rule* dan sistem untuk melakukan *reasoning* pada ontology sehingga dapat disimpulkan apakah suatu *resource* memenuhi syarat tertentu.

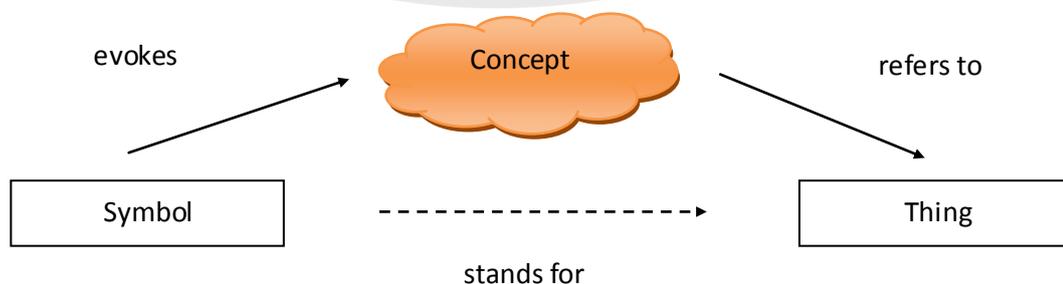
g. *Trust*

Layer dari *semantic web* yang memungkinkan pengguna *web* untuk mempercayai suatu informasi pada *web* (Pollock, 2009).

2.4 Ontology

Definisi ontologi dalam ilmu komputer yang sering dirujuk berasal dari Tom Gruber (2007) yang menyatakan “*An ontology is an explicit and formal specification of a conceptualization of domain of interest*”.

Dalam memahami definisi tersebut digunakan istilah *semantic*, *logic*, *controlled vocabulary*, *taxonomy* dan *thesauri* sebagai sudut pandang untuk memulai pembahasan tentang pengertian ontologi. Untuk memberikan pengertian *semantic*, dirujuk permasalahan dalam konteks komunikasi. Dalam konteks komunikasi dengan ataupun tanpa *ontology* dikenal suatu penggambaran yang disebut *meaning triangle* yang dapat dilihat pada Gambar 2.2.



Gambar 2.2 *Meaning Triangle*
(sumber: Sarno, 2012)

Ada tiga komponen pada *meaning triangle* yang berupa *symbol*, *concept* dan *thing*. *Symbol* merupakan kata atau istilah yang digunakan dalam bahasa (*syntax*).

Thing (referent) adalah sesuatu yang diwakilkan oleh *symbol*, namun tidak terdapat relasi langsung dari *symbol* menuju *thing*. Untuk memahami *thing* yang dimaksud, diperlukan *concept* yang memberikan *meaning (semantic)* sehingga terdapat *reference* menuju *thing* tersebut.

Suatu kosakata terkontrol (*controlled vocabulary*) adalah suatu daftar istilah (*term*) atau konsep yang dinyatakan secara eksplisit. Semua istilah dalam kosakata terkontrol memiliki definisi yang jelas dan tidak redundansi.

Taxonomy adalah koleksi dari kosakata terkontrol yang diorganisasikan secara hierarki berdasarkan hubungan generalisasi. *Thesauri* adalah jaringan koleksi dari kata atau frase dengan satu set relasi *linguistic*. *Thesauri* juga menggunakan relasi asosiatif selain menggunakan relasi induk-anak sehingga *thesauri* lebih kuat dalam memberikan *semantic* bila dibandingkan dengan *taxonomy*. Relasi asosiatif dalam *thesauri* misalnya “*related-to*”, “*broader*”, “*narrower*” dan sebagainya.

Tujuan dibuatnya *ontology semantic* adalah untuk meningkatkan otomatisasi pemrosesan teks dengan menyediakan representasi konsep yang ada di dunia secara *language independent* dan *meaning-based* (Sarno, 2012).

2.5 Temu Kembali Informasi (*Information Retrieval*)

Temu kembali informasi berkaitan dengan representasi, penyimpanan, pengorganisasian, dan pengaksesan informasi. Sistem temu kembali informasi

menyediakan kemudahan akses informasi bagi pengguna. Pengguna harus menerjemahkan kebutuhan informasinya dalam bentuk kueri. Dengan adanya kueri yang diberikan pengguna, tujuan utama dari sistem temu kembali informasi adalah mengembalikan informasi yang relevan dengan kueri dan informasi yang tidak relevan sesedikit mungkin.

2.6 Question Answering

Question Answering merupakan aplikasi nyata dari teknologi *Natural Language Processing* (NLP). Tujuan utama dari *Question Answering* (QA) adalah menampilkan jawaban atas pertanyaan yang diberikan pengguna. Ide utama QA adalah (Dolarosa, 2008):

- Menentukan tipe semantik jawaban yang diharapkan
- Menentukan tipe dokumen-dokumen yang memiliki *keywords* seperti pada pertanyaan
- Mencari entitas dengan tipe yang sesuai dengan pertanyaan, yang dekat dengan *keywords*

2.7 DBPedia

DBPedia Indonesia merupakan sebuah komunitas yang bergerak untuk mengekstrak informasi terstruktur dari Wikipedia dan menyediakan informasi tersebut dalam sebuah web. DBPedia Indonesia merupakan sebuah web yang menyediakan data hasil ekstraksi dari Wikipedia Indonesia. Dengan DBPedia dapat dilakukan pencarian yang kompleks terhadap data dari Wikipedia. Adapun

cara pengaksesan *knowledge* DBPedia Indonesia ada 3 cara, yaitu *Linked Data*, *SPARQL Endpoint* dan *RDF Dumps*.

2.7.1 *Linked Data*

Cara pengaksesan entitas dengan menggunakan URI *resource* dari entitas tersebut. Setiap artikel Wikipedia dengan URL <http://id.wikipedia.org/wiki/JUDULARTIKEL> akan memiliki URI *resource* <http://id.dbpedia.org/resource/JUDULARTIKEL>. Di bawah ini contoh dari beberapa URI *resource*, antara lain:

- http://id.dbpedia.org/resource/Universitas_Indonesia
- <http://id.dbpedia.org/resource/Indonesia>
- <http://id.dbpedia.org/resource/Soekarno>
- http://id.dbpedia.org/resource/Jawa_Tengah
- http://id.dbpedia.org/resource/The_Beatles

2.7.2 *SPARQL Endpoint*

Ini merupakan cara pengaksesan *knowledge base* DBPedia Indonesia dengan menggunakan *query*. Dalam hal ini, untuk melakukan *query* sebaiknya mempelajari *ontology* dan *property-property* yang digunakan di DBPedia. Berikut contoh struktur *query* SPARQL.

```
select ?gubernur ?wakilGubernur
  where {
    dbpedia-id:Daerah_Khusus_Ibukota_Jakarta
    dbpedia-owl:leaderName ?gubernur.
    dbpedia-id:Daerah_Khusus_Ibukota_Jakarta
    dbpedia-owl:viceLeader ?wakilGubernur.
  }
```

Hasil dari *query* tersebut akan menampilkan nama dari gubernur dan wakil gubernur DKI Jakarta.

2.7.3 RDF Dumps

DBpedia Indonesia juga menyediakan RDF *dumps* yang merupakan hasil ekstraksi dari Wikipedia Indonesia. RDF *dumps* terbagi menjadi 26 berkas dalam format *N-Triple* yang dikompresi dengan gzip.

2.8 Parsing

Parsing adalah proses untuk menguraikan kalimat menjadi per kata. Atau dengan kata lain *parsing* adalah proses *parser* atau pemisahan dari elemen-elemen pembentuk dokumen ke dalam satuan terkecil, yaitu kata (Zainal, 2009). Pemisah antar kata yang umum digunakan dalam sebuah kalimat adalah spasi. Sehingga pada penelitian ini akan digunakan spasi sebagai pemisah untuk menguraikan kalimat tanya yang diinputkan oleh pengguna menjadi per kata.

2.9 Stemming

Stemming merupakan suatu proses yang terdapat dalam sistem *Information Retrieval* (IR) yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Sebagai contoh, kata bersama, kebersamaam, menyamai, akan distem ke root wordnya yaitu “sama”. Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris (Keke, 2012). Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi:

Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1

Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan (Agusta, 2009).

2.10 SPARQL

SPARQL *Protocol and RDF Query Language* (SPARQL) adalah sebuah protocol dan bahasa *query* untuk *Semantic Web's resources*. Sebuah *query* yang menggunakan SPARQL dapat terdiri atas *triple patterns*, konjungsi (*or*) dan disjungsi (*and*). Berikut ini adalah contoh *query* yang menghasilkan semua ibu kota di Indonesia:

```
PREFIX abc: <http://mynamespace.com/exampleOntologie#>
SELECT ?capital ?province
WHERE {
  ?x abc:cityname ?capital.
  ?y abc:provincename ?province.
  ?x abc:isCapitalOf ?y.
  ?y abc:isInCountry abc:indonesia.
}
```

Untuk menjalankan SPARQL dapat menggunakan beberapa *tools* dan APIs seperti: *ARQ*, *Rasqal*, *RDF::Query*, *twingql*, *Pellet*, dan *KAON2*. *Tools* tersebut memiliki API yang memungkinkan pemrogram untuk memanipulasi hasil *query* dengan berbagai aplikasi yang ada. Namun, sebagai standar dapat digunakan SPARQL *Query Results XML Fomat* yang direkomendasikan oleh W3C (Beckett, 2013). Hasil dari *query* di atas adalah (Ibrahim, 2007):

```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
<head>
<variable name="capital"/>
<variable name="province"/>
```

```
</head>
<results>
<result>
<binding name="capital">
<literal
datatype="http://www.w3.org/2001/XMLSchema#string">
Bandung
</literal>
</binding>
<binding name="province">
<literal
datatype="http://www.w3.org/2001/XMLSchema#string">
Jawa Barat
</literal>
</binding>
</result>
<!-- more results -->
</results>
</sparql>
```



BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Desain Sistem

Question Answering System (QAS) merupakan sistem yang memungkinkan pengguna menyatakan kebutuhan informasinya dalam bentuk *natural language question* (pertanyaan dalam bahasa alami), dan mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban.

Sistem yang akan dibuat menggunakan DBPedia Indonesia sebagai sumber pengetahuan. Ada 3 proses utama di dalamnya, yaitu:

1. *Text Processing*

Ini merupakan tahap awal untuk mengolah kalimat tanya yang diinput oleh pengguna. Kalimat tersebut diubah menjadi huruf kecil semua (*lower*), dihilangkan tanda bacanya, dihapus stopwords-nya, kemudian dicari bentuk kata dasarnya menggunakan algoritma *Porter stemming*.

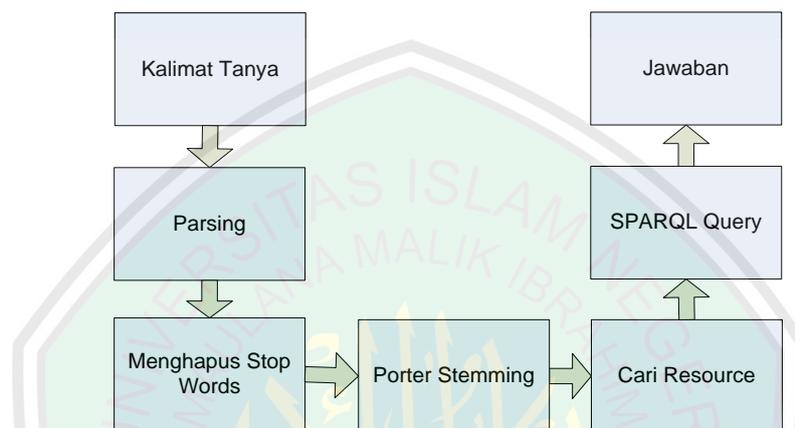
2. *Data Mining*

Data mining merupakan proses untuk menggali kata-kata yang akan digunakan dalam kueri SPARQL. Kandidat kata-kata yang digunakan adalah kata yang telah dilakukan *text processing* pada tahap pertama.

Data mining digunakan untuk mencari *resource* di DBPedia Indonesia yang sesuai dengan pertanyaan yang diinputkan oleh pengguna.

3. Menampilkan jawaban

Setelah ditemukan *resource* di DBPedia yang sesuai dengan pertanyaan dari pengguna. Maka selanjutnya adalah menampilkan hasil jawaban kepada pengguna. Blok diagram dari *Question Answering System* yang akan dibuat dapat dilihat pada Gambar 3.1.



Gambar 3.1 Blok Diagram *Question Answering System*

Gambar 3.1 merupakan blok diagram dari *Question Answering System* yang akan dibangun. Penjelasan dari masing-masing proses pada blok diagram di atas dijelaskan pada subbab berikutnya.

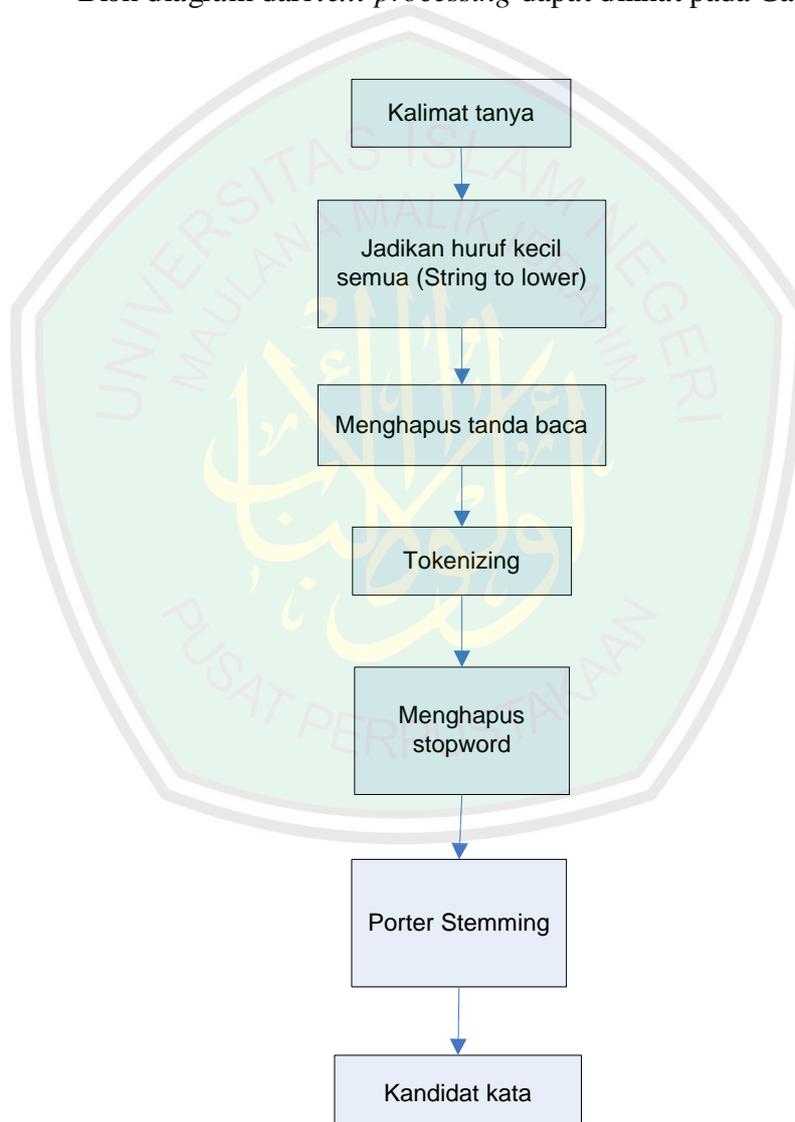
3.1.1 *Text Processing*

Pada *Question Answering System* (QAS), yang dijadikan inputan adalah string pertanyaan yang diinputkan oleh pengguna. String ini diolah terlebih dahulu untuk mempermudah proses pencarian kandidat kata yang dapat digunakan sebagai *resource*. Pengolahan string meliputi:

- Mengubah string menjadi huruf kecil semua (*lower*)
- Menghilangkan tanda tanya (?)

- Memecah kalimat tersebut menjadi per kata (*tokenizing*)
- Melakukan penghapusan *stopwords*. *Stopwords* adalah kata yang tidak memiliki makna dan sebaiknya dihilangkan saja.
- Mencari kata dasar dari masing-masing kata dengan menggunakan algoritma *Porter Stemming*.

Blok diagram dari *text processing* dapat dilihat pada Gambar 3.2.



Gambar 3.2 Blok Diagram *Text Processing*

3.1.1.1 Kalimat tanya

Ini merupakan tahap dimana pengguna menginputkan kalimat tanya ke dalam sistem. Kalimat tanya yang diinputkan dapat menggunakan beberapa macam kata tanya, antara lain: apa atau apakah, siapa atau siapakah, dimana atau dimanakah, kapan atau kapankah, mengapa, bagaimana dan berapa.

Misalnya: Siapakah Presiden Indonesia?

3.1.1.2 Casefolding

String pertanyaan dari pengguna diubah menjadi huruf kecil semua. Proses ini juga dapat dinamakan sebagai *case folding* (membuat semua huruf pada teks menjadi huruf kecil).

3.1.1.3 Menghapus Tanda Tanya

Setelah diubah menjadi huruf kecil semua, tanda tanya (?) yang ada pada pertanyaan akan dihapus.

3.1.1.4 Parsing

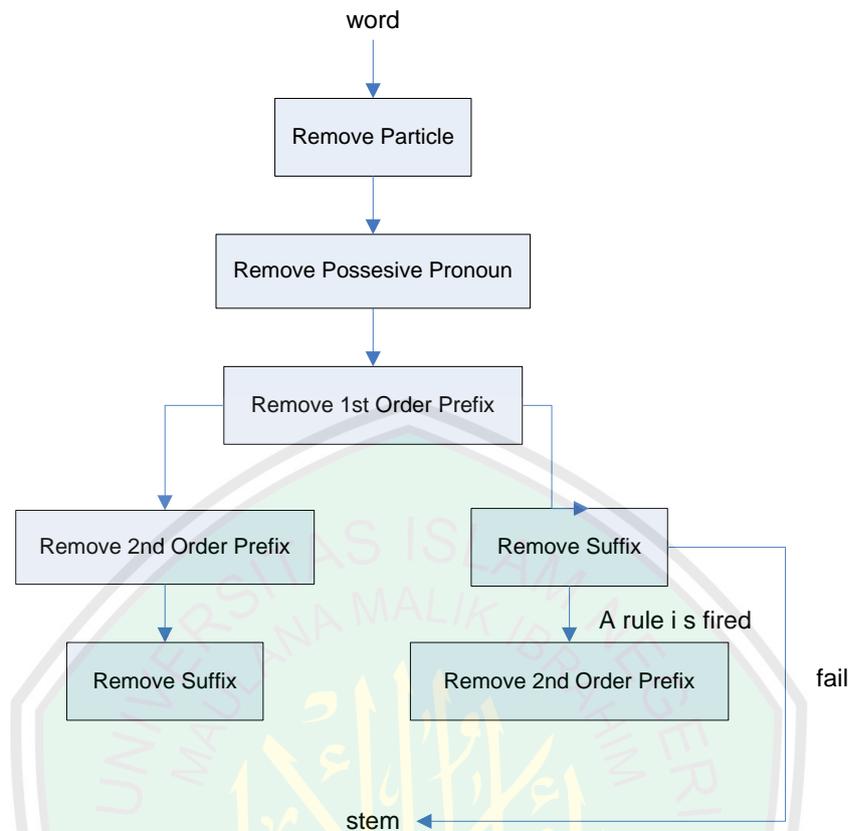
Proses yang ada dalam *parsing* adalah *tokenizing*, yaitu memecah kalimat yang telah diinput oleh pengguna menjadi per kata. Proses *parsing* menggunakan spasi sebagai separator. Sebelum dilakukan *parsing*, dilakukan penghapusan tanda tanya (?) untuk memudahkan proses selanjutnya jika pengguna memasukkan tanda tanya ketika menginputkan pertanyaan. Proses ini akan menghasilkan token-token untuk kemudian dilanjutkan dengan proses penghapusan *stopwords*.

3.1.1.5 Menghapus *Stopwords*

Proses penghapusan *stopwords* dilakukan untuk menghilangkan kata-kata yang tidak mempunyai makna, misalnya: di, yang, pada, ke, dan lain sebagainya. Kata-kata yang termasuk ke dalam *stopwords* tersebut disimpan ke dalam suatu tabel pada database dan dicocokkan satu persatu dengan kata yang diinputkan oleh pengguna. Daftar *stopwords* dapat dilihat pada lampiran. Apabila terdapat kata-kata yang mengandung *stopwords* dari kalimat yang diinput oleh pengguna maka kata tersebut akan dihapus.

3.1.1.6 *Porter Stemming*

Stemming adalah proses untuk mentransformasikan kata-kata yang telah dipecah sebelumnya menjadi *root word* (kata dasarnya). Pada penelitian ini, *stemming* menggunakan algoritma *Porter*. Algoritma ini ditemukan oleh Martin Porter pada tahun 1980. Algoritma tersebut digunakan untuk *stemming* kata berbahasa Inggris, kemudian karena proses *stemming* bahasa Inggris berbeda dengan bahasa Indonesia, maka dikembangkan algoritma *Porter* khusus untuk bahasa Indonesia oleh W.B. Frakes pada tahun 1992. *Stemming* kata dilakukan pada kata yang diindikasikan sebagai *keyword*. Blok diagram dari algoritma *Porter* dapat dilihat pada Gambar 3.3 (Agusta, 2009):



Gambar 3.3 Blok diagram *Porter Stemming*
(sumber: Agusta, 2009)

Adapun langkah- langkah algoritma ini adalah sebagai berikut:

1. Hapus *Particle*
2. Hapus *Possesive Pronoun*
3. Hapus awalan pertama. Jika tidak ada lanjutkan ke langkah 4a, jika ada cari maka lanjutkan ke langkah 4b
4. (a) Hapus awalan kedua, lanjutkan ke langkah 5a
(b) Hapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai *root word*. Jika ditemukan maka lanjutkan ke langkah 5b

5. (a) Hapus akhiran. Kemudian kata akhir diasumsikan sebagai *root word*

(b) Hapus awalan kedua. Kemudian kata akhir diasumsikan sebagai *root word*

Terdapat 5 kelompok aturan pada algoritma *Porter* untuk bahasa Indonesia ini. Aturan tersebut dapat dilihat pada Tabel 3.1, Tabel 3.2, Tabel 3.3, Tabel 3.4 dan Tabel 3.5.

Tabel 3.1 Aturan untuk *Inflection Particle*

Akhiran	Replacement	Measure Condition	Additional Condition	Contoh
-kah	NULL	2	NULL	Bukukah
-lah	NULL	2	NULL	Pergilah
-pun	NULL	2	NULL	Bukupun

Tabel 3.2 Aturan untuk *Inflection Possesive Pronoun*

Akhiran	Replacement	Measure Condition	Additional Condition	Contoh
-ku	NULL	2	NULL	Bukuku
-mu	NULL	2	NULL	Bukumu
-nya	NULL	2	NULL	Bukunya

Tabel 3.3 Aturan untuk *First Order Derivational Prefix*

Awalan	Replacement	Measure Condition	Additional Condition	Contoh
Meng-	NULL	2	NULL	Mengukur → ukur
Meny-	S	2	V...*	Menyapu → sapu
Men-	NULL	2	NULL	Menduga → duga
Mem-	P	2	V...	Memaksa → paksa
Mem-	NULL	2	NULL	Membaca → baca
Me-	NULL	2	NULL	Merusak → rusak
Peng-	NULL	2	NULL	Pengukur → ukur
Peny-	S	2	V...	Penyapu → sapu
Pen-	NULL	2	NULL	Penduga → duga
Pem-	P	2	V...	Pemaksa → paksa
Pem-	NULL	2	NULL	Pembaca → baca
Di-	NULL	2	NULL	Diukur → ukur
Ter-	NULL	2	NULL	Tersapu → sapu
Ke-	NULL	2	NULL	Kekasih → kasih

Tabel 3.4 Aturan untuk *Second Order Derivational Prefix*

Awalan	Replacement	Measure Condition	Additional Condition	Contoh
Ber-	NULL	2	NULL	Berlari → lari
Bel-	NULL	2	Ajar	Belajar → ajar
Be-	NULL	2	K*er	Bekerja → kerja
Per-	NULL	2	NULL	Perjelas → jelas
Pel-	NULL	2	Ajar	Pelajar → ajar
Pe-	NULL	2	NULL	Pekerja → kerja

Tabel 3.5 Aturan untuk *Derrivational Suffix*

Akhiran	Replacement	Measure Condition	Additional Condition	Contoh
-kan	NULL	2	Prefix bukan anggota {ke, peng}	Tarikkan → tarik Mengambilkan → ambil
-an	NULL	2	Prefix bukan anggota {di, meng, ter}	Makanan → makan Perjanjian → janji
-i	NULL	2	Prefix bukan anggota {ber, ke, peng}	Tandai → tanda Mendapati → dapat

Stemming adalah proses menghilangkan imbuhan dari suatu kata pada suatu dokumen. Permasalahan utama dalam *stemming* adalah bagaimana mendapatkan kata dasar yang benar dari hasil pemenggalan imbuhan dari suatu kata jadian. Efektifitas algoritma *stemming* dipengaruhi oleh beberapa faktor (Mandala, 2004):

- a. Kesalahan dalam proses pemenggalan imbuhan dari kata dasarnya. Kesalahan ini dapat berupa:
 - *Overstemming*: yaitu pemenggalan imbuhan yang melebihi dari yang seharusnya. Contoh: kata masalah menjadi masa. Kesalahan ini dapat timbul karena bentuk kata dasar yang menyerupai imbuhan.

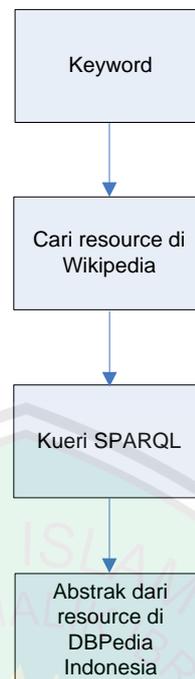
- *Understemming*: yaitu pemenggalan imbuhan yang terlalu sedikit dari yang seharusnya. Contoh: kata belajar menjadi lajar. Kesalahan ini dapat timbul karena kekurangan pada aturan pola imbuhan yang didefinisikan.
 - *Unchange*: yaitu kasus khusus dari *understemming*, dimana tidak terjadi pemenggalan imbuhan sama sekali. Contoh: kata telapak, setelah pemenggalan kata dasar yang didapat tetap telapak. Kesalahan ini dapat ditimbulkan karena kekurangan pada aturan pola imbuhan yang didefinisikan.
 - *Spelling exception*: yaitu huruf pertama kata dasar yang didapat tidak benar yang diakibatkan dari pemenggalan awalan. Contoh: kata memukul menjadi ukul. Kesalahan ini dapat timbul karena ada beberapa imbuhan yang berubah bentuk ketika ditempelkan pada suatu kata dasar. Misalnya awalan berR-, meN-, teR-, peR-, akan bergantung pada huruf pertama kata dasar dimana imbuhan tersebut ditempelkan. Contoh: ber- + ajar = belajar, pen- + lihat = penglihatan, pen- + sakit = penyakit. Atau sebaliknya ada imbuhan yang mengakibatkan huruf pertama kata dasar yang ditempelinya menjadi luluh. Misalnya meng- / peng- meluluhkan huruf 'k'. Contoh: mengarang menjadi meng- dan karang atau men- / pen- meluluhkan huruf 'p'. Contoh: menuai dari men- dan tuai.
- b. Kekurangan dalam perumusan aturan penambahan imbuhan pada kata dasar. Hal ini dapat terjadi karena morfologi bahasa Indonesia yang kompleks, sehingga sangat sulit atau bahkan tidak mungkin untuk merumuskan aturan yang sempurna.

- c. Jumlah total aturan imbuhan yang didapat berhubungan dengan efektifitas proses temu kembali. Dimana semakin banyak pola penambahan imbuhan yang dapat dirumuskan, maka proses temu kembali akan semakin efektif.

Kata-kata yang di-*stemming* adalah kata-kata yang bukan termasuk ke dalam *stopwords*. Setelah dilakukan *stemming*, maka akan diperoleh kata dasar atau akar kata yang akan diproses ke tahap selanjutnya. Dari beberapa langkah pada *text processing* ini selanjutnya diperoleh kandidat kata. Kandidat kata ini selanjutnya digunakan sebagai *keywords* untuk melakukan proses selanjutnya yaitu *data mining*. *Data mining* ini dilakukan untuk mencari *resource* di Wikipedia.

3.1.2 *Data Mining*

Data mining adalah proses penggalian data, *data mining* yang dimaksud pada QAS ini adalah untuk menemukan kata yang dapat digunakan sebagai *resource* dari pertanyaan yang diinputkan oleh pengguna. Data yang akan di-*mining* adalah data pada DBPedia Indonesia. Pencarian *resource* pada DBPedia Indonesia menggunakan kandidat kata dari *text processing* pada proses sebelumnya. Sehingga pada tahap ini dibagi lagi menjadi 2 proses di dalamnya, yaitu: mencari *link resource* terlebih dahulu di Wikipedia Indonesia dan kueri SPARQL dengan *endpoint* DBPedia Indonesia. Blok diagram dari *data mining* dapat dilihat pada Gambar 3.4:



Gambar 3.4 Blok diagram *Data Mining*

3.1.2.1 Cari *Resource* di Wikipedia

Pencarian *resource* dilakukan dengan menggunakan Google API yaitu CSE (*Custom Search Engine*). *Endpoint* pencariannya adalah <http://id.wikipedia.org>. Hasil dari pencarian *resource* ini selanjutnya digunakan untuk *parsing* kueri SPARQL. Hasil dari proses pencarian ini adalah sebuah link untuk URI *resource*, misalnya: <http://id.wikipedia.org/wiki/Indonesia>. Yang dibutuhkan untuk proses selanjutnya adalah kata Indonesia nya

3.1.2.2 SPARQL Query

Struktur dari kueri SPARQL yang akan digunakan dapat dilihat pada Gambar 3.5.

```
select distinct ?Concept where {[] a ?Concept}
```

Gambar 3.5 Struktur Kueri SPARQL

Contoh kueri SPARQL dengan menggunakan *endpoint* DBpedia Indonesia dapat dilihat pada Gambar 3.6.

```
select ?jawaban
  where {
    dbpedia-id:Ekosistem dbpedia-owl:abstract
    ?jawaban.
  }
```

Gambar 3.6 Contoh Kueri SPARQL

3.1.3 Menampilkan Jawaban

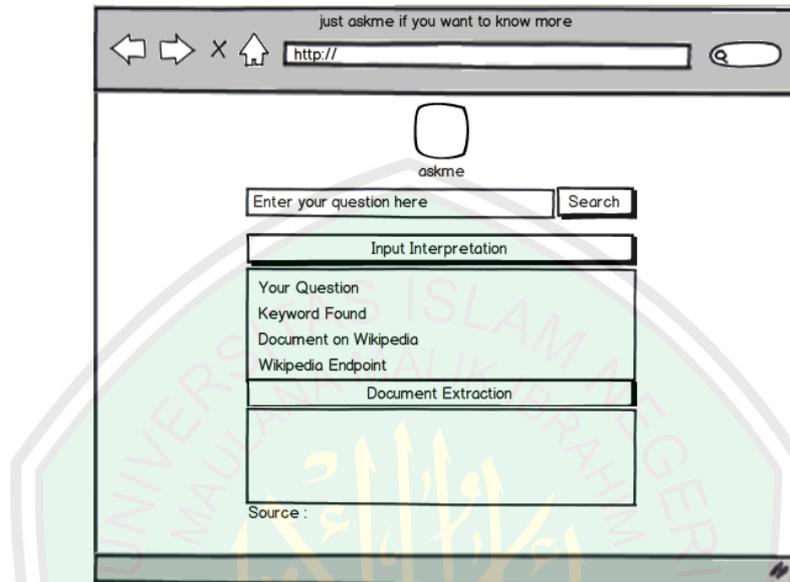
Setelah dilakukan kueri menggunakan SPARQL, maka akan diperoleh jawaban yang akan ditampilkan ke pengguna. Jawaban yang ditampilkan kepada pengguna adalah *abstract* Wikipedia dari *resource* yang telah ditemukan. Hasil dari kueri pada Gambar 3.6 dapat dilihat pada Gambar 3.7.

Ekosistem adalah suatu sistem ekologi yang terbentuk oleh hubungan timbal balik tak terpisahkan antara makhluk hidup dengan lingkungannya. Ekosistem bisa dikatakan juga suatu tatanan kesatuan secara utuh dan menyeluruh antara segenap unsur lingkungan hidup yang saling memengaruhi. Ekosistem merupakan penggabungan dari setiap unit biosistem yang melibatkan interaksi timbal balik antara organisme dan lingkungan fisik sehingga aliran energi menuju kepada suatu struktur biotik tertentu dan terjadi suatu siklus materi antara organisme dan anorganisme. Matahari sebagai sumber dari semua energi yang ada. Dalam ekosistem, organisme dalam komunitas berkembang bersama-sama dengan lingkungan fisik sebagai suatu sistem. Organisme akan beradaptasi dengan lingkungan fisik, sebaliknya organisme juga memengaruhi lingkungan fisik untuk keperluan hidup. Pengertian ini didasarkan pada Hipotesis Gaia, yaitu: "organisme, khususnya mikroorganisme, bersama-sama dengan lingkungan fisik menghasilkan suatu sistem kontrol yang menjaga keadaan di bumi cocok untuk kehidupan". Hal ini mengarah pada kenyataan bahwa kandungan kimia atmosfer dan bumi sangat terkendali dan sangat berbeda dengan planet lain dalam tata surya. Kehadiran, kelimpahan dan penyebaran suatu spesies dalam ekosistem ditentukan oleh tingkat ketersediaan sumber daya serta kondisi faktor kimiawi dan fisis yang harus berada dalam kisaran yang dapat ditoleransi oleh spesies tersebut, inilah yang disebut dengan hukum toleransi. Misalnya: Panda memiliki toleransi yang luas terhadap suhu, namun memiliki toleransi yang sempit terhadap makanannya, yaitu bambu. Dengan demikian, panda dapat hidup di ekosistem dengan kondisi apapun asalkan dalam ekosistem tersebut terdapat bambu sebagai sumber makanannya. Berbeda dengan makhluk hidup yang lain, manusia dapat memperlebar kisaran toleransinya karena kemampuannya untuk berpikir, mengembangkan teknologi dan memanipulasi alam.

Gambar 3.7 Hasil dari Kueri SPARQL

3.2 Desain Interface

Desain *interface* untuk tampilan awal dari sistem dapat dilihat pada Gambar 3.8 ini:



Gambar 3.8 Desain Interface Question Answering System

Pada bagian pertama, terdapat *input field* untuk menginputkan pertanyaan ke dalam sistem. Berikut penjelasan untuk kolom *Your Question*, *Keyword Found*, *Document on Wikipedia*, *Wikipedia Endpoint* dan *Document Extraction*:

- *Your Question* : menampilkan pertanyaan yang diinputkan oleh pengguna
- *Keyword Found* : menampilkan kandidat kata yang digunakan sebagai *keyword* untuk pencarian *resource* di Wikipedia Indonesia. Kata yang ditampilkan diperoleh dari proses *text processing* yaitu: *case folding*, *parsing*, menghapus *stopwords* dan Porter

stemming.

- *Document on Wikipedia* : menampilkan *link resource* URI dari Wikipedia
- *Wikipedia Endpoint* : menampilkan kata yang akan diproses untuk SPARQL
- *Document Extraction* : menampilkan *abstract* DBPedia Indonesia dari kata pada Wikipedia *endpoint*.

3.3 Kebutuhan Sistem

Berikut ini beberapa perangkat keras maupun lunak yang dibutuhkan untuk mendukung pembuatan dan uji coba *Question Answering System (QAS)*.

- a) Perangkat Keras (Hardware)
 - Peneliti menggunakan PC / Laptop dengan spesifikasi processor Intel(R) Core(TM) i3-2328M CPU @2.20GHz 2.20 GHz dan RAM 2,00 GB
- b) Perangkat Lunak (Software)
 - XAMPP versi 1.8.3
 XAMPP digunakan sebagai web server yang berdiri sendiri (localhost), yang terdiri atas program Apache HTTP Server, MySQL database dan penerjemah bahasa yang ditulis dengan bahasa pemrograman PHP dan Perl
 - Notepad++
 Notepad++ digunakan sebagai *text editor*.
 - Mozilla Firefox

Mozilla Firefox merupakan browser yang digunakan untuk mengeksekusi aplikasi

- Adobe Photoshop

Photoshop digunakan untuk membuat desain logo dan tampilan web.



BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang implementasi dari perancangan yang telah dibuat sebelumnya.

4.1 Implementasi Sistem

Sistem yang dibuat terdiri dari 3 proses utama, yaitu *text processing*, *data mining* dan menampilkan jawaban. *Text processing* terdiri dari beberapa langkah, yaitu:

- Mengubah string menjadi huruf kecil semua (*lower*) atau *case folding*
- Menghilangkan tanda tanya (?)
- Memecah kalimat tersebut menjadi per kata (*tokenizing*)
- Melakukan penghapusan *stopwords*. *Stopwords* adalah kata yang tidak memiliki makna dan sebaiknya dihilangkan saja.
- Mencari kata dasar dari masing-masing kata yang telah di-*parsed* dengan menggunakan algoritma *Porter stemming*.

Berikut ini kode program dari *text processing*:

```
$soal = $_POST['soal'];

//jadikan huruf kecil semua
$lower = strtolower($soal);

//menghilangkan tanda tanya
$ubah = str_replace("?", " ", $lower);

//memecah kalimat menjadi per kata (tokenizing)
$pecah = explode(" ", $ubah);

$i = 0;
while($split[$i] != null) {
```

```

$dicari=
hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(hap
uspp(hapuspartikel($pecah[0])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[1])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[2])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[3])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[4])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[5])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[6])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[7])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[8])))).'
'.hapusStopword(hapusakhiran(hapusawalan2(hapusawalan1(ha
puspp(hapuspartikel($pecah[9]))));

$i++;
}

```

Berikut kode program untuk menghapus stopwords:

```

function hapusStopword ($kata) {
// cari di database
$sql = mysql_query("SELECT * from tb_stoplist where
stoplist ='$kata'");

if(mysql_num_rows($sql) == 1) {
$stopword = $kata;
$katabaru = str_replace($kata," ", $kata);
return $katabaru;
}

else {
$katabaru = $kata;
return $katabaru;
}}

```

4.1.1 Implementasi *Porter Stemming*

Stemming pada QAS ini digunakan untuk mentransformasikan kata-kata yang telah dipecah sebelumnya menjadi *root word* (kata dasarnya). Blok diagram dari *Porter stemming* dapat dilihat pada Gambar 3.3. Langkah-langkah algoritma ini adalah sebagai berikut:

1. Hapus *Particle*
2. Hapus *Possesive Pronoun*
3. Hapus awalan pertama. Jika tidak ada lanjutkan ke langkah 4a, jika ada cari maka lanjutkan ke langkah 4b
4. (a) Hapus awalan kedua, lanjutkan ke langkah 5a
(b) Hapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai *root word*. Jika ditemukan maka lanjutkan ke langkah 5b
5. (a) Hapus akhiran. Kemudian kata akhir diasumsikan sebagai *root word*
(b) Hapus awalan kedua. Kemudian kata akhir diasumsikan sebagai *root word*

Berikut kode program dari *Porter stemming*:

```
<?php
function cari($kata){
    include "koneksi.php";
    $hasil = mysql_num_rows(mysql_query("SELECT *
FROM tb_katadasar WHERE katadasar='$kata'"));
    return $hasil;
}

//langkah 1 - hapus partikel
function hapuspartikel($kata){
if(cari($kata)!=1){
```

```

if((substr($kata, -3) == 'kah' )||( substr($kata, -
3) == 'lah' )||( substr($kata, -3) == 'pun' )){
    $kata = substr($kata, 0, -3);

    }
}
return $kata;
}

//langkah 2 - hapus possessive pronoun
function hapuspp($kata){
if(cari($kata)!=1){
    if(strlen($kata) > 4){
        if((substr($kata, -2)== 'ku')||(substr($kata,
-2)== 'mu')){
            $kata = substr($kata, 0, -2);
        }else if((substr($kata, -3)== 'nya')){
            $kata = substr($kata,0, -3);
        }
    }
}
return $kata;
}

//langkah 3 hapus first order prefiks (awalan
pertama)
function hapusawalan1($kata){
if(cari($kata)!=1){

    if(substr($kata,0,4)=="meng"){

        if(substr($kata,4,1)=="e"||substr($kata,4,1)=
=="u"){
            $kata = "k".substr($kata,4);
        }else{
            $kata = substr($kata,4);
        }
    }else if(substr($kata,0,4)=="meny"){
        $kata = "s".substr($kata,4);
    }else if(substr($kata,0,3)=="men"){
        $kata = substr($kata,3);
    }else if(substr($kata,0,3)=="mem"){
if(substr($kata,3,1)=="a" || substr($kata,3,1)=="i"
|| substr($kata,3,1)=="e" || substr($kata,3,1)=="u"
|| substr($kata,3,1)=="o"){
            $kata = "p".substr($kata,3);
        }else{
            $kata = substr($kata,3);
        }
    }
}
}

```

```

}else if(substr($kata,0,2)=="me"){
    $kata = substr($kata,2);
}else if(substr($kata,0,4)=="peng"){
    if(substr($kata,4,1)=="e" ||
substr($kata,4,1)=="a"){
        $kata = "k".substr($kata,4);
    }else{
        $kata = substr($kata,4);
    }
}else if(substr($kata,0,4)=="peny"){
    $kata = "s".substr($kata,4);
}else if(substr($kata,0,3)=="pen"){
    if(substr($kata,3,1)=="a" ||
substr($kata,3,1)=="i" || substr($kata,3,1)=="e" ||
substr($kata,3,1)=="u" || substr($kata,3,1)=="o"){
        $kata = "t".substr($kata,3);
    }else{
        $kata = substr($kata,3);
    }
}else if(substr($kata,0,3)=="pem"){
    if(substr($kata,3,1)=="a" ||
substr($kata,3,1)=="i" || substr($kata,3,1)=="e" ||
substr($kata,3,1)=="u" || substr($kata,3,1)=="o"){
        $kata = "p".substr($kata,3);
    }else{
        $kata = substr($kata,3);
    }
}else if(substr($kata,0,2)=="di"){
    $kata = substr($kata,2);
}else if(substr($kata,0,3)=="ter"){
    $kata = substr($kata,3);
}else if(substr($kata,0,2)=="ke"){
    $kata = substr($kata,2);
}
}
return $kata;
}

```

//langkah 4 hapus second order prefiks (awalan kedua)

```

function hapusawalan2($kata){
if(cari($kata)!=1){

    if(substr($kata,0,3)=="ber"){
        $kata = substr($kata,3);
    }else if(substr($kata,0,3)=="bel"){
        $kata = substr($kata,3);
    }else if(substr($kata,0,2)=="be"){
        $kata = substr($kata,2);
    }
}
}

```

```

}else if(substr($kata,0,3)=="per" && strlen($kata)
> 5){
    $kata = substr($kata,3);
}else if(substr($kata,0,2)=="pe" &&
strlen($kata) > 5){
    $kata = substr($kata,2);
}else if(substr($kata,0,3)=="pel" &&
strlen($kata) > 5){
    $kata = substr($kata,3);
}else if(substr($kata,0,2)=="se" &&
strlen($kata) > 5){
    $kata = substr($kata,2);
}
}
return $kata;
}
////langkah 5 hapus suffiks
function hapusakhiran($kata){
if(cari($kata)!=1){
    if (substr($kata, -3)== "kan" ){
        $kata = substr($kata, 0, -3);
    }
    else if(substr($kata, -1)== "i" ){
        $kata = substr($kata, 0, -1);
    }else if(substr($kata, -2)== "an"){
        $kata = substr($kata, 0, -2);
    }
}
return $kata;
}
?>

```

4.1.2 Implementasi Proses Pencarian *Resource*

Untuk proses pencarian resource, peneliti memanfaatkan Google API yaitu *Google Custom Search Engine* (GCSE). Berikut kode program untuk proses pencarian *resource*:

```

$kata = $dicari;
$carilink = str_replace(" ", "+", $kata);
$cariUrl =
file_get_contents("https://www.googleapis.com/custo
msearch/v1?q=.'$carilink' .&cx=016089238520837785579
:mf4ije6qmcm&key=AIzaSyDrOMqyZkmuoKk-40xuVhWa-
qHeHaH_tS8");

```

```

$jas = json_decode($cariUrl, true);
$linkwiki = $jas["items"][0]
["link"];
$linkwiki2 = $jas["items"][1]
["link"];
$linkwiki3 = $jas["items"][2]
["link"];
$pecah = explode("/", $linkwiki);
$pecah2 = explode("/", $linkwiki2);
$pecah3 = explode("/", $linkwiki3);
$resource = $pecah[4];
$resource2 = $pecah2[4];
$resource3 = $pecah3[4];

```

4.1.3 Implementasi SPARQL pada DBpedia Indonesia

Setelah dilakukan pencarian *link resource*, maka didapatkan kata yang akan dimasukkan ke dalam kueri SPARQL dengan *endpoint* DBpedia

Indonesia. Berikut kode programnya:

```

<?php
function cariUrlDBpediaAbstract($resource)
{
    $format = 'json';
    $query =
    "select ?jawaban
        where {
            dbpedia-id:". $resource ."          dbpedia-
owl:abstract ?jawaban
        }";
    $searchUrl = 'http://id.dbpedia.org/sparql?'
    .'query='.urlencode($query)
    .'&format='.$format;
    return $searchUrl;
}
function request($url){
    // is curl installed?
    if (!function_exists('curl_init')){
        die('CURL is not installed!');
    }
    // get curl handle
    $ch= curl_init();
    // set request url
    curl_setopt($ch,
    CURLOPT_URL,
    $url);
    // return response, don't print/echo

```




ask me

Enter your question here Search

Input Interpretation

Your Question	
Keyword Found	
Document on Wikipedia	
Wikipedia Endpoint	

Document Extraction

Source:

Lama Proses Pencarian Jawaban : 2.040116071701 detik

Gambar 4.1 Implementasi *User Interface*

Hasil dari pencarian QAS akan menampilkan 3 kandidat jawaban teratas karena jika menggunakan salah satu saja, jawaban yang ditampilkan terkadang tidak cocok dengan pertanyaan yang dimasukkan oleh pengguna. Hal ini dilakukan karena jawaban yang ditampilkan kepada pengguna hanyalah abstrak dari DBPedia Indonesia sehingga peneliti menggunakan 3 kandidat jawaban teratas.

Pada Gambar 4.2 ditampilkan implementasi ketika pengguna memasukkan pertanyaan dan aplikasi menampilkan jawaban. Pada kolom *your question* menampilkan pertanyaan yang dimasukkan oleh pengguna, selanjutnya pada kolom *Keyword Found* menampilkan kata kunci untuk proses pencarian *link resource* di Wikipedia Indonesia. Kata kunci diperoleh dari proses *casefolding*, *parsing*, menghapus *stopwords* dan *Porter stemming*. Pada kolom *Document on Wikipedia* adalah hasil pencarian *link* menggunakan *Google Custom Search Engine* (GCSE). Kata yang ditampilkan pada *Wikipedia endpoint* digunakan pada kueri SPARQL untuk proses pencarian jawaban di DBPedia Indonesia.



Search

Input Interpretation

Your Question	Apa yang dimaksud dengan sholat jamak?
Keyword Found	m aksud sholat jamak
Document on Wikipedia	http://id.wikipedia.org/wiki/Salat_Jamak http://id.wikipedia.org/wiki/Salat_Qashar http://id.wikipedia.org/wiki/Majelis_Tafsir_AI_Quran
Wikipedia Endpoint	Salat_Jamak Salat_Qashar Majelis_Tafsir_AI_Quran

Document Extraction

Salat Jamak yaitu salat yg dilaksanakan dengan mengumpulkan dua salat wajib dalam satu waktu, seperti salat Zuhur dengan Asar dan salat Magrib dengan salat Isya (khusus dalam perjalanan). Adapun pasangan salat yang bisa dijamak adalah salat Dzuhur dengan Ashar atau salat Maghrib dengan Isya. Salat jamak dibedakan menjadi dua tipe yakni: Jama' Taqdim penggabungan pelaksanaan dua salat dalam satu waktu dengan cara memajukan salat yang belum masuk waktu ke dalam salat yang telah masuk waktunya (seperti penggabungan pelaksanaan salat Asar dengan salat Zuhur pada waktu salat Zuhur atau pelaksanaan salat Isya dengan salat Magrib pada waktu salat Magrib). Jama' Ta'khir penggabungan pelaksanaan dua salat dalam satu waktu dengan cara mengundurkan salat yang sudah masuk waktu ke dalam waktu salat yang berikutnya (seperti penggabungan pelaksanaan salat Zuhur dengan salat Asar pada waktu salat Asar, atau pelaksanaan salat Magrib dengan salat Isya pada waktu salat Isya)

Salat Qashar adalah melakukan salat dengan meringkas/mengurangi jumlah raka'at salat yang bersangkutan. Salat Qashar merupakan keringanan yang diberikan kepada mereka yang sedang melakukan perjalanan. Adapun salat yang dapat diqashar adalah salat dzuhur, ashar dan isya, dimana raka'at yang aslinya berjumlah 4 dikurangi/diringkas menjadi 2 raka'at saja.

Majlis Tafsir Al-Quran atau disingkat MTA adalah lembaga dakwah dalam bentuk yayasan yang didirikan oleh Al-Ustadz Abdullah Thufail Saputra pada tanggal 19 September 1972. Pendirian Yayasan MTA ini selanjutnya dikukuhkan dengan akte notaris R. Soegondo Notodisoerjo, nomor 23, tanggal 23 Januari 1974 di Surakarta.

Source:
http://id.wikipedia.org/wiki/Salat_Jamak
http://id.wikipedia.org/wiki/Salat_Qashar
http://id.wikipedia.org/wiki/Majelis_Tafsir_AI_Quran
 Lama Proses Pencarian Jawaban : 4.3243958950043 detik

Gambar 4.2 Sistem Menampilkan Jawaban kepada Pengguna

Implementasi sistem pada Gambar 3.4 menggunakan contoh pertanyaan ‘Apa yang dimaksud dengan sholat jamak?’. Setelah dilakukan proses *casefolding*, *parsing*, menghapus stopwords dan *Porter stemming*, *keyword* yang diperoleh adalah maksud sholat jamak. Selanjutnya *keyword* ini digunakan untuk

mencari *link* Wikipedia Indonesia dengan menggunakan *Google Custom Search Engine* (GCSE). Tiga link teratas yang diperoleh yaitu:

http://id.wikipedia.org/wiki/Salat_Jamak

http://id.wikipedia.org/wiki/Salat_Qashar

http://id.wikipedia.org/wiki/Majelis_Tafsir_Al_Quran

Setelah itu dilakukan kueri SPARQL dan jawaban pada tampilan kepada pengguna ditampilkan pada *Document Extraction*. Paragraf pertama menampilkan abstrak dari *link* Wikipedia yang pertama yaitu mengenai salat jamak, paragraf kedua menampilkan abstrak dari *link* Wikipedia yang kedua yaitu salat qasar, dan paragraf ketiga menampilkan abstrak dari *link* Wikipedia yang ketiga yaitu majelis tafsir Al Qur'an.

4.3 Uji Coba

Pengujian aplikasi ini menggunakan sejumlah beberapa pertanyaan dan dicek ketepatan jawabannya. Ada 200 pertanyaan yang diujikan dengan beberapa kata tanya yang berbeda, yaitu: apa atau apakah, siapa atau siapakah, dimana atau dimanakah, berapa atau berapakah, kapan atau kapankah, mengapa dan bagaimana. Hasil uji coba dicantumkan pada lampiran.

Jawaban dari hasil uji coba terbagi menjadi 2 yaitu, jawaban tepat dan jawaban tidak tepat. Jawaban tepat adalah kolom untuk jawaban yang dijawab dengan tepat oleh sistem, sedangkan kolom jawaban tidak tepat adalah untuk jawaban yang salah atau sistem tidak mengembalikan jawaban sama sekali kepada pengguna.

Berdasarkan hasil uji coba terhadap 200 pertanyaan di atas, jumlah pertanyaan yang dijawab dengan tepat adalah 143 pertanyaan, sedangkan jawaban yang tidak tepat sebanyak 57 pertanyaan. Rekapitulasi dari uji coba dapat dilihat pada Tabel 4.1.

Tabel 4.1 Rekapitulasi Uji Coba

Jumlah Pertanyaan	Jawaban Tepat		Jawaban Tidak Tepat	
	Jumlah	Prosentase	Jumlah	Prosentase
200	143	71,50 %	57	28,5 %

4.4 Analisa Hasil

Dari hasil uji coba terhadap 200 pertanyaan, nilai akurasi jawaban yang didapat dari Question Answering System yang dibuat sebesar 71,50 %.

Berikut hasil analisa dari uji coba terhadap 200 pertanyaan:

- a. Kesalahan pada *stemming* yaitu *spelling exception*. *Spelling exception* adalah huruf pertama kata dasar yang didapat tidak benar ini diakibatkan dari pemenggalan awalan. Kesalahan ini terjadi pada beberapa pertanyaan, antara lain:
 - Pertanyaan no 3 (Apa pengertian dari *hardware*?). Hasil dari *stemming* dan penghapusan *stopwords* adalah kerti *hardware*. Kata dasar dari pengertian seharusnya adalah ngerti bukan kerti. Pengertian merupakan kata yang mempunyai awalan pen- dan akhiran -an.
 - Pertanyaan no 11 (Apa nama planet yang memiliki cincin?). Hasil dari *stemming* dan penghapusan *stopwords* adalah planet pilik cincin. Kata dasar dari memiliki seharusnya adalah milik. Memiliki merupakan kata yang mempunyai awalan me- dan akhiran -i.

- Pertanyaan no 49 (Kapan Nagita Slavina memulai karir?). Hasil dari *stemming* dan penghapusan *stopwords* adalah nagita slavina pulai karir. Kata dasar dari memulai seharusnya adalah mulai. Memulai merupakan kata yang mempunyai awalan me- dan akhiran -i.
- Pertanyaan no 82 (Bagaimana cara mengukur presisi?). Hasil dari *stemming* dan penghapusan *stopwords* adalah kukur presisi. Kata dasar dari mengukur adalah ukur. Mengukur mempunyai awalan me-. Kesalahan sama juga terjadi pada pertanyaan no 84.
- Pertanyaan no 156 (Berapa hari ayam mengerami telurnya?). *Keyword* yang ditemukan yaitu berapa ayam keram telur. Kata mengerami mengalami *spelling exception*, seharusnya kata dasar dari mengerami adalah eram. Kata mengerami mempunyai awalan meng- dan akhiran -i. Wikipedia *endpoint* yang ditemukan tidak berhubungan sama sekali dengan pertanyaan.
Anafilaksis, Vitamin, Sindrom_prahaid
- b. Kesalahan pada *stemming* berupa *overstemming*. *Overstemming* adalah pemenggalan imbuhan yang melebihi dari yang seharusnya, seperti yang terjadi pada beberapa pertanyaan, antara lain:
 - Pertanyaan no 43 (Siapa penemu listrik?). Hasil dari *stemming*-nya adalah te listrik. Kata dasar dari penemu seharusnya adalah temu. Penemu merupakan kata yang mempunyai awalan pe-. *Keyword* yang ditemukan adalah te listrik sehingga Wikipedia *endpoint* yang ditemukan tidak sesuai dengan pertanyaan yaitu:
Pretty_Man, Tarif_dasar_listrik, Sistem_listrik_pesawat_terbang

- Pertanyaan no 52 (Kapan perayaan natal berlangsung?). Hasil dari *stemming* dan penghapusan *stopwords* adalah aya natal. Kata dasar dari perayaan seharusnya adalah raya. Perayaan merupakan kata yang memiliki awalan pe- dan akhiran -an.
- Pertanyaan no 53 (Kapan masa jabatan Jokowi sebagai Gubernur DKI Jakarta?). Hasil dari *stemming* dan penghapusan *stopword* adalah jabat jokow bagai dk jakarta. Kata Jokowi dan DKI seharusnya tetap.
- Pertanyaan no 57 (Kapan PKI dibubarkan?). *Keyword* yang ditemukan pk bubar. Kata PKI seharusnya tetap, tidak perlu di-*stemming*.
- Pertanyaan no 58 (Kapan TNI dibentuk?). *Keyword* yang ditemukan tn bentuk. Kata TNI seharusnya tetap, tidak perlu di *stemming*.
- Pertanyaan no 79 (Bagaimana cara menanam jagung?). *Keyword* yang ditemukan anam jagung. Kata dasar dari menanam seharusnya adalah tanam. Menanam adalah kata yang memiliki awalan me-. Kesalahan yang sama juga terjadi pada pertanyaan no 89.
- Pertanyaan no 122 (Siapa yang pertama kali menemukan benua Amerika?). *Keyword* yang ditemukan mengalami *overstem* yaitu emu benua amerika. Kata dasar dari menemukan seharusnya adalah temu. Wikipedia *endpoint*-nya juga tidak berhubungan dengan pertanyaan, yaitu:

Emu, Harimau_Tasmania, Australia
- Pertanyaan no 141 (Dimana Jokowi dilahirkan?). *Keyword* yang ditemukan adalah jokow lahir. Kata Jokowi seharusnya tidak perlu dilakukan

stemming. Wikipedia *endpoint* yang ditemukan hanya 1 dan tidak berhubungan dengan pertanyaan yaitu Gamawan_Fauzi.

- Pertanyaan no 149 (Berapa luas Pulau Kalimantan?). *Keyword* yang ditemukan yaitu berapa luas pulau kalimant. Kata kalimantan seharusnya tidak perlu di-*stemming*. Tiga *Wikipedia endpoint* yang diperoleh yaitu:

Kalimantan, Kalimantan_(pulau), Daftar _provinsi_di_Indonesia

Ketiga abstrak dari *Wikipedia endpoint* tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 163 (Kapan hari Pendidikan?). *Keyword* yang ditemukan adalah didi dan kata ini mengalami *overstem*. Kata dasar dari pendidikan adalah didik. Kata pendidikan memiliki awalan pe- dan akhiran -an. *Wikipedia endpoint* yang ditemukan tidak berhubungan sama sekali dengan pertanyaan, yaitu:

Didi_Mahardika, Didi_Riyadi, Didi,_Dusun_Timur, Barito_Timur

- Pertanyaan no 177 (Mengapa Al Qur'an diturunkan secara berangsur-angsur?). *Keyword* yang ditemukan yaitu al qur' turun angsur-angsur. Kata Al Qur'an seharusnya tidak perlu di-*stemming*. *Wikipedia endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Al-Qur'an, Surah_Al-Furqan, Pemeluk_Islam_pertama

- c. Informasi yang kurang lengkap pada abstrak DBPedia dan *Wikipedia endpoint* yang tidak sesuai, seperti yang terjadi pada beberapa pertanyaan, antara lain:

- Pertanyaan no 8 (Apakah judul lagu kebangsaan Indonesia?). *Keyword* yang didapatkan sudah benar yaitu judul lagu bangsa Indonesia, sedangkan 3 Wikipedia *endpoint* teratas yang didapatkan yaitu:
Daftar_lagu_nasional_Indonesia, Satu_Nusa_Satu_Bangsa,
Lagu_Bangsa_Johor
Dari ketiga *endpoint* tersebut tidak ada jawaban yang sesuai dengan pertanyaan.
- Pertanyaan no 77 (Bagaimana cara membuat tempe?). Tiga Wikipedia *endpoint* yang ditemukan yaitu:
Tempe, Tempe_(disambiguasi), Blok_Tempe
Dari ketiga *endpoint* tersebut, tidak ada jawaban yang sesuai.
- Pertanyaan no 81 (Bagaimana cara merawat kucing anggora?). *Keyword* yang ditemukan rawat kucing anggora dan Wikipedia *endpoint* yang ditemukan tidak memiliki abstrak yang sesuai dengan jawaban yaitu:
Anggora_Turki, Kucing, Anak_kucing
- Pertanyaan no 104 (Berapa warna yang ada pada pelangi?). *Keyword* yang ditemukan berapa warna pelangi dan Wikipedia *endpoint* yang ditemukan tidak memiliki abstrak yang sesuai dengan pertanyaan yaitu:
Pelangi, Ular-air_pelangi, Spektrum
- Pertanyaan no 118 (Apa saja dampak dari bom nuklir?). *Keyword* yang didapatkan adalah dampak bom nuklir. Wikipedia *endpoint* yang ditemukan antara lain:

Daya_ledak_senjata_nuklir,

Serangan_bom_atom_di_Hiroshima_dan_Nagasaki, Senjata_nuklir

Abstrak yang ditampilkan dari ketiga *endpoint* tersebut masih kurang sesuai menjawab pertanyaan.

- Pertanyaan no 127 (Siapa pendiri organisasi NU?). *Keyword* yang ditemukan sesuai yaitu organisasi nu. Wikipedia *endpoint* nya juga sesuai, yaitu:

Nahdlatul_'Ulama, Muslimat_Nahdlatul_Ulama, Gerakan_Pemuda_Ansor.

Akan tetapi dari ketiganya, abstrak yang ditampilkan tidak dapat menjawab pertanyaan.

- Pertanyaan no 128 (Siapa pencipta lagu Rasa Sayange?). *Keyword* yang ditemukan adalah cipta lagu rasa sayange. Tiga Wikipedia *endpoint* yang didapat yaitu:

Kopi_darat, Putri_Ayu_Silaen, Daftar_lagu_anak_di_Indonesia

Dari ketiga Wikipedia *endpoint* tersebut, abstrak yang disajikan belum menjawab pertanyaan.

- Pertanyaan no 130 (Siapa istri Hanung Bramantyo?). *Keyword* yang ditemukan adalah istri hanung bramantyo. Wikipedia *endpoint* yang didapatkan yaitu:

Hanung_Bramantyo, %3F_(film), Cinta_Tapi_Beda

Akan tetapi abstrak yang ditampilkan dari ketiga Wikipedia *endpoint* tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 139 (Dimana pertama kali diadakan SEA GAMES?). *Keyword* yang ditemukan adalah sea games. Kata tersebut mengalami *understemming*. Kata dasar dari diadakan seharusnya adalah ada. Wikipedia *endpoint* yang ditemukan yaitu:

Pesta_Olahraga_Asia_Tenggara, Pesta_Olahraga_Asia_Tenggara_2013,

Pesta_Olahraga_Asia_Tenggara_2011

Abstrak yang ditampilkan dari ketiga Wikipedia *endpoint* tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 152 (Berapa luas negara Indonesia?). *Keyword* yang ditemukan yaitu berapa luas negara Indonesia. Ketiga Wikipedia *endpoint* yang diperoleh yaitu:

Daftar_negara_menurut_luas_wilayah, Indonesia,

Daerah_Khusus_Ibukota_Jakarta

Abstrak dari ketiga Wikipedia *endpoint* tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 155 (Berapa jumlah album Arie Lasso?). *Keyword* yang ditemukan adalah berapa jumlah album arie lasso. Kata-katanya tetap tidak mengalami perubahan. Wikipedia *endpoint* yang ditemukan juga sesuai yaitu:

Arie_Lasso, Keseimbangan_(album_Arie_Lasso),

Selalu_Ada_(album_Arie_Lasso)

Akan tetapi abstrak yang ditampilkan belum bisa menjawab pertanyaan.

- Pertanyaan no 158 (Berapa suhu di matahari?). *Keyword* yang ditemukan yaitu berapa suhu matahari. Wikipedia *endpoint* yang ditemukan sesuai, antara lain:

Matahari, Neptunus, Pembicaraan:Matahari

Akan tetapi abstrak yang ditampilkan belum bisa menjawab pertanyaan.

- Pertanyaan no 162 (Kapan hari kelahiran kota Malang?). *Keyword* yang ditemukan yaitu lahir kota malang. Wikipedia *endpoint* yang ditemukan sudah sesuai namun abstraknya belum bisa menjawab pertanyaan.

Kota_Malang,

Daftar_kota_di_Indonesia_menurut_hari_jadi, Muhammad_Anton

- Pertanyaan no 164 (Kapan Rumah Sakit Syaiful Anwar dibangun?). *Keyword* yang ditemukan yaitu rumah sakit syaiful anwar bangun. Wikipedia *endpoint* yang ditemukan antara lain:

Rumah_Sakit, Masjid_Istiqlal, Daftar_rumah_sakit_di_Indonesia

Abstrak dari ketiga Wikipedia *endpoint* tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 166 (Kapan diadakan Konferensi Asia Afrika). *Keyword* yang ditemukan yaitu adakan konferensi asia afrika. Kata adakan mengalami *understemming*. Kata dasar dari diadakan adalah ada. Kata diadakan mempunyai awalan di- dan akhiran -kan. Wikipedia *endpoint* yang ditemukan sudah sesuai dengan pertanyaan. Namun data yang tersaji pada abstraknya belum bisa menjawab pertanyaan.

Konferensi_Asia% E2%60%93 Afrika,

Konferensi_Tingkat_Tinggi_Asia% E2%80%93 Afrika_2005, Gerakan_Non-Blok

- Pertanyaan no 174 (Bagaimana gejala penyakit demam berdarah?). *Keyword* yang ditemukan yaitu gejala sakit demam berdarah. Wikipedia *endpoint* yang ditemukan sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Demam_berdarah_dengue, Penyakit_virus_Ebola, Chikungunya

- Pertanyaan no 175 (Bagaimana mencegah osteoporosis?). *Keyword* yang ditemukan yaitu cegah osteoporosis. Wikipedia *endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Osteoporosis, Susu_kedelai, Semanggi

- Pertanyaan no 181 (Mengapa bisa terjadi obesitas?). *Keyword* yang ditemukan yaitu obesitas. Wikipedia *endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun DBPedia Indonesia tidak memiliki datanya, sehingga tidak ada jawaban yang ditampilkan kepada pengguna.

Kegemukan, Obesitas_abdominal, Berkas:Obesitas.jpg

- Pertanyaan no 183 (Mengapa boraks berbahaya jika digunakan di makanan?). *Keyword* yang ditemukan yaitu boraks bahaya guna makan. Wikipedia *endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Bleng, Asam_borat, Tembaga(II)_sulfat

- Pertanyaan no 187 (Apa yang dimaksud dengan shalat?). *Keyword* yang ditemukan yaitu maksud shalat. Wikipedia *endpoint* yang ditampilkan masih memiliki kaitan dengan pertanyaan namun belum bisa menjawab pertanyaan.

Salat_Istikharah, Salat_Rawatib, Salat_Jamak

- Pertanyaan no 191 (Mengapa umat Muslim yang mampu wajib membayar zakat di bulan Ramadhan?). *Keyword* yang ditemukan yaitu umat muslim wajib bayar zakat bulan. Wikipedia *endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Zakat_Fitrah, Zakat, Ramadhan

- Pertanyaan no 193 (Apa saja macam-macam zakat?). *Keyword* yang ditemukan yaitu macam-macam zakat. Wikipedia *endpoint* yang ditemukan ada yang sesuai dengan pertanyaan namun abstrak yang disajikan belum bisa menjawab pertanyaan.

Zakat, Zakat_Mal, Nisab

- Pertanyaan no 199 (Bagaimana cara melakukan tayammum?). *Keyword* yang ditemukan yaitu laku tayammum. Wikipedia *endpoint* yang ditemukan belum bisa menjawab pertanyaan.

Wudu, Pertempuran_Badar, Talak

- d. Wikipedia *endpoint* yang didapatkan tidak sesuai dengan pertanyaan, seperti yang terjadi pada beberapa pertanyaan, antara lain:

- Pertanyaan no 38 (Siapa pencetus teori gravitasi?). *Keyword* yang didapatkan sudah sesuai yaitu cetus teori gravitasi. Tetapi Wikipedia *endpoint* yang ditemukan tidak sesuai yaitu:

Keberatan_terhadap_evolution, Galaksi_katai, Sagittarius_A*

Dengan demikian, sistem tidak menyajikan jawaban yang sesuai terhadap pengguna.

- Pertanyaan no 46 (Kapan Indonesia dinyatakan merdeka). *Keyword* yang didapatkan yaitu indonesia nyata merdeka. Tiga Wikipedia *endpoint* teratas yaitu:

Merdeka_17805, Merah_Putih_(film),

Daftar_negara_dengan_pengakuan_terbatas

Dari ketiga *endpoint* tersebut tidak ada jawaban yang sesuai.

- Pertanyaan no 52 (Kapan perayaan natal berlangsung?). *Keyword* yang ditemukan aya natal. Tiga Wikipedia *endpoint* yang didapatkan yaitu:

Gals!, Buku_Harian_Nayla, Putri_Ayu_Silaen

Dari ketiga *endpoint* tersebut tidak ada jawaban yang sesuai.

- Pertanyaan no 53 (Kapan masa jabatan Jokowi sebagai Gubernur DKI Jakarta?). *Keyword* yang ditemukan mengalami *overstem* yaitu jabat jokowi bagi guberbur dk jakarta. Tiga Wikipedia *endpoint* yang didapatkan tidak sesuai dengan pertanyaan yaitu:

Johannes_Suryo_Prabowo, Universitas_Syiah_Kuala,

Usulan_ibu_kota_Indonesia_masa_depan

- Pertanyaan no 54 (Kapan hari Kebangkitan Nasional diperingati?). *Keyword* yang ditemukan yaitu bangkit nasional ingat. Tiga Wikipedia *endpoint* yang ditemukan tidak sesuai yaitu:
Parkour, Bahasa_Cirebon, Proyek_Manhattan
- Pertanyaan no 57 (Kapan PKI dibubarkan?). *Keyword* yang ditemukan pk bubar, sehingga 3 Wikipedia *endpoint* yang didapatkan tidak sesuai yaitu:
Partai_Keadilan_Sejahtera, Park_Jung_Min, Adam_Air
- Pertanyaan no 58 (Kapan TNI dibentuk?). *Keyword* yang ditemukan tn bentuk, sehingga 3 Wikipedia *endpoint* yang didapatkan tidak sesuai yaitu:
Jurnal_umum, Taman_Nasional_Gunung_Ciremai, Menara_Eiffel
- Pertanyaan no 59 (Kapan terjadinya perjanjian hudaibiah?). *Keyword* yang didapatkan sudah benar yaitu janji hudaibiah, sedangkan 3 Wikipedia *endpoint* teratas yang didapatkan yaitu:
Ba'at_'Aqabah_Kedua, Ar_Rahman, Hassan_al-Hudaybi
Ketiga *endpoint* tersebut tidak ada yang sesuai dengan pertanyaan.
- Pertanyaan no 61 (Mengapa Jepang menyerah terhadap Belanda tahun 1945?). *Keyword* yang ditemukan yaitu jepang serah hadap belanda 1945, sedangkan Wikipedia *endpoint* yang ditemukan tidak sesuai yaitu
Sumatera_Barat_pada_masa_pendudukan_Jepang.
- Pertanyaan no 64 (Mengapa Jogja dijuluki sebagai daerah istimewa?). *Keyword* yang ditemukan jogja juluk bagai daerah istimewa, sedangkan 3 Wikipedia *endpoint* yang ditemukan tidak sesuai dengan pertanyaan yaitu:
Diponegoro, Mangkunegara_I, Purwokerto

- Pertanyaan no 79 (Bagaimana cara menanam jagung?). *Keyword* yang ditemukan anam jagung, sedangkan 3 Wikipedia *endpoint* yang ditemukan tidak ada yang sesuai dengan pertanyaan yaitu:
Kinali,_Pasaman_Barat,Sudungdewo,_Kertek,_Wonosobo,Kabupaten_Hulu
_Sungai_Tengah
- Pertanyaan no 80 (Bagaimana cara menginstal ulang Windows?). *Keyword* yang ditemukan sesuai yaitu instal ulang windows, sedangkan 3 Wikipedia *endpoint* yang ditemukan tidak sesuai dan tidak ada jawabannya yaitu:
Windows_95, Windows_Registry, Microsoft_Exchange_Server
- Pertanyaan no 84 (Bagaimana cara mengukur akurasi?). *Keyword* yang ditemukan kukur akurasi dan Wikipedia *endpoint* yang ditemukan tidak berhubungan sama sekali dengan pertanyaan yaitu Hidrosefalus. Dalam kasus ini, GCSE hanya menemukan 1 link Wikipedia.
- Pertanyaan no 89 (Bagaimana cara menanam padi?). *Keyword* yang ditemukan yaitu anam padi. Wikipedia *endpoint* yang didapatkan tidak sesuai dengan pertanyaan yaitu:
Kinali,_Pasaman_Barat, Kabupaten_Hulu_Sungai_Tengah,
Tokoh_dari_Jombang
- Pertanyaan no 115 (Apakah sungai yang terpanjang di dunia?). *Keyword* yang ditemukan sungai dunia dan Wikipedia *endpoint* tidak ada yang sesuai yaitu:
Dunia_bawah, Sungai_Amazon, Sungai_Reprua

- Pertanyaan no 116 (Apa alat pernapasan dari serangga?). *Keyword* yang ditemukan adalah alat napas serangga. Wikipedia *endpoint* dan abstrak yang ditemukan tidak sesuai dengan pertanyaan yaitu:
Respirometer, Kumbang, dan Hewan
- Pertanyaan no 121 (Siapa yang disebut sebagai *three musketeer*?). *Keyword* yang didapatkan adalah sebut bagai *three musketeer*. GCSE hanya menemukan satu Wikipedia *endpoint* yaitu Perancis dan tidak berhubungan dengan pertanyaan.
- Pertanyaan no 137 (Dimana letak Kepulauan Seribu?). *Keyword* yang ditemukan adalah letak pulau. Wikipedia *endpoint* yang didapat yaitu:
Kepulauan_Bangka_Belitung, Pulau_Bangka, Pulau_Tidore.
Ketiga Wikipedia *endpoint* tersebut tidak berhubungan dengan pertanyaan.
- Pertanyaan no 145 (Berapa jumlah kadar air dalam tubuh manusia?). *Keyword* yang ditemukan adalah berapa jumlah kadar air tubuh manusia.
Tiga Wikipedia *endpoint* yang ditemukan yaitu:
Homeostasis, Habitat_bakteri, Adaptasi
Abstrak yang ditampilkan dari ketiga Wikipedia *endpoint* tersebut tidak dapat menjawab pertanyaan.
- Pertanyaan no 156 (Berapa hari ayam mengerami telurnya?). *Keyword* yang ditemukan yaitu berapa ayam keram telur. Wikipedia *endpoint* yang ditemukan tidak berhubungan sama sekali dengan pertanyaan.
Anafilaksis, Vitamin, Sindrom_prahaid

- Pertanyaan no 182 (Mengapa bumi makin panas?). *Keyword* yang ditemukan yaitu bumi makin panas. Wikipedia *endpoint* yang ditemukan tidak berhubungan dengan pertanyaan.

Bumi_Makin_Panas, Suzanna, Efek_Tyndall

- Pertanyaan no 190 (Kapan kita dapat melaksanakan haji?). *Keyword* yang ditemukan yaitu laksana haji. Wikipedia *endpoint* yang didapat tidak berhubungan sama sekali dengan pertanyaan.

Laksana,_Pakuhaji,_Tangerang, Pakuhaji,_Tangerang,

Raden_Aria_Wira_Tanu_II

- e. *Google Custom Search Engine* (GCSE) tidak dapat menemukan Wikipedia *endpoint*-nya, sehingga prosesnya terhenti pada menemukan *keyword* saja seperti yang terjadi pada:

- Pertanyaan no 74 (Mengapa edelweiss disebut bunga abadi?). *Keyword* yang ditemukan edelweiss sebut bunga abadi.
- Pertanyaan no 82 (Bagaimana cara mengukur presisi?). *Keyword* yang ditemukan kukur presisi.
- Pertanyaan no 83 (Bagaimana aturan permainan gobak sodor?). *Keyword* yang ditemukan atur main gobak sodor.

- f. Kesalahan pada stemming berupa *understemming*. *Understemming* yaitu pemenggalan imbuhan yang terlalu sedikit dari yang seharusnya. Contoh: kata belajar menjadi lajar. Kesalahan ini dapat timbul karena kekurangan pada aturan pola imbuhan yang didefinisikan. Kesalahan seperti ini dijumpai pada beberapa pertanyaan yaitu:

- Pertanyaan no 139 (Dimana pertama kali diadakan SEA GAMES?). *Keyword* yang ditemukan adalah sea games. Kata adakan mengalami *understemming*. Kata dasar dari diadakan seharusnya adalah ada. Wikipedia *endpoint* yang ditemukan yaitu:

Pesta_Olahraga_Asia_Tenggara, Pesta_Olahraga_Asia_Tenggara_2013,

Pesta_Olahraga_Asia_Tenggara_2011

Abstrak yang ditampilkan dari ketiga Wikipedia endpoint tersebut belum bisa menjawab pertanyaan.

- Pertanyaan no 166 (Kapan diadakan Konferensi Asia Afrika). *Keyword* yang ditemukan yaitu adalah konferensi asia afrika. Kata adakan mengalami *understemming*. Kata dasar dari diadakan adalah ada. Kata diadakan mempunyai awalan di- dan akhiran –kan. Wikipedia *endpoint* yang ditemukan sudah sesuai dengan pertanyaan. Namun data yang tersaji pada abstraknya belum bisa menjawab pertanyaan.

Konferensi_Asia% E2% 60% 93 Afrika,

Konferensi_Tingkat_Tinggi_Asia% E2% 80% 93 Afrika_2005, Gerakan_Non-Blok

Dari analisa beberapa pertanyaan tersebut, algoritma *Porter stemming* memiliki 3 kekurangan yaitu:

(1) *Spelling Exception*

Huruf pertama kata dasar yang didapat tidak benar yang diakibatkan dari pemenggalan awalan. Contoh: kata memukul menjadi ukul. Kesalahan ini dapat timbul karena ada beberapa imbuhan yang berubah bentuk ketika

ditempelkan pada suatu kata dasar. Pada uji coba yang telah dilakukan, kasus *spelling exception* dijumpai pada kata dengan pola sebagai berikut:

- ✓ Awalan pe- dan akhiran -an seperti pada kata pengertian mengalami *spelling exception* menjadi kerti.
- ✓ Awalan me- dan akhiran -i seperti: kata memiliki mengalami *spelling exception* menjadi pilik, kata memulai mengalami *spelling exception* menjadi pulai, kata mengerami mengalami *spelling exception* menjadi keram.

(2) *Overstemming*

Pemenggalan imbuhan yang melebihi dari yang seharusnya. Contoh: kata masalah menjadi masa. Kesalahan ini dapat timbul karena bentuk kata dasar yang menyerupai imbuhan. Pada uji coba yang telah dilakukan kasus *overstemming* dijumpai pada kata dengan pola berikut:

- ✓ Awalan pe-, seperti kata penemu mengalami *overstemming* menjadi te.
- ✓ Awalan pe- dan akhiran -an, seperti: kata perayaan mengalami *overstemming* menjadi aya, dan kata pendidikan menjadi didi.
- ✓ Awalan me-, seperti kata menanam mengalami *overstemming* menjadi anam.

(3) *Understemming*

Pemenggalan imbuhan yang terlalu sedikit dari yang seharusnya. Contoh: kata belajar menjadi lajar. Kesalahan ini dapat timbul karena kekurangan pada aturan pola imbuhan yang didefinisikan. Pada uji coba yang telah

dilakukan, kasus *understemming* dijumpai pada kata dengan pola sebagai berikut:

- ✓ Awalan pe- dan akhiran -an, seperti pada kata diadakan mengalami *understemming* menjadi adakan.

4.5 Integrasi *Question Answering System* dengan Islam

Belajar atau menuntut ilmu merupakan kewajiban bagi seluruh umat muslim. Seperti apa yang telah dikatakan pada hadits yang diriwayatkan oleh Turmudzi berikut ini.

مَنْ أَرَادَ الدُّنْيَا فَعَلَيْهِ بِالْعِلْمِ, وَمَنْ أَرَادَ الْآخِرَةَ فَعَلَيْهِ بِالْعِلْمِ, وَمَنْ أَرَا دَهُمَا فَعَلَيْهِ بِالْعِلْمِ

Artinya:

Barang siapa yang menghendaki kehidupan dunia maka wajib baginya memiliki ilmu, dan barang siapa yang menghendaki kehidupan akhirat, maka wajib baginya memiliki ilmu, dan barang siapa menghendaki keduanya maka wajib baginya memiliki ilmu. (HR. Turmudzi)

Hadits tersebut secara jelas juga menerangkan bahwa jika kita sebagai umat muslim menginginkan kebahagiaan hidup didunia maka kita harus menuntut ilmu. Begitu pula jika kita menginginkan kebahagiaan hidup di akhirat, maka kita juga wajib menuntut ilmu. Ada berbagai cara menuntut ilmu, antara lain: berguru secara langsung kepada ahlinya atau belajar melalui media pembelajaran, seperti: internet, televisi, radio, dan lain-lainnya. Di dalam Al Qur'an surat Al A'laq ayat 3-5 termaktub ayat mengenai media pembelajaran, yaitu:

يَعَلِّمَ مَا لَا نَسْنَعُ بِالْقَلَمِ عَلَّمَ الَّذِي الْأَكْرَمُ وَرَبُّكَ أَقْرَأُ

Artinya:

3. Bacalah, dan Tuhanmulah yang Maha pemurah,
4. Yang mengajar (manusia) dengan perantaran kalam,
5. Dia mengajar kepada manusia apa yang tidak diketahuinya.

Berdasarkan terjemah ayat di atas, dapat diilhami bahwa proses pembelajaran atau pentransferan ilmu kepada manusia dari yang semula tidak tahu menjadi tahu menggunakan perantara berupa pena. Menurut terjemah tafsir jalalain, makna bacalah pada lafal ayat ini ialah mengukuhkan makna lafal pertama yang sama (dan Rabbmulah Yang Paling Pemurah) artinya tiada seorang pun yang dapat menandingi kemurahan-Nya. Lafal ayat ini sebagai Haal dan Dhamir yang terkandung di dalam lafal Iqra'. Selanjutnya, terjemah dari ayat keempat adalah yang mengajar (manusia) dengan perantaraan kalam, manusia pertama yang menulis dengan memakai kalam atau pena ialah Nabi Idris a.s. Pada ayat kelima, (Dia mengajarkan kepada manusia) atau jenis manusia (apa yang tidak diketahuinya) yaitu sebelum Dia mengajarkan kepadanya hidayah, menulis dan berkreasi serta hal-hal lainnya (As Suyuthi, 2008).

Berdasar pada ayat tersebut, maka peneliti mengembangkan Question Answering System dengan tujuan sebagai media pembelajaran atau sarana menuntut ilmu. Sehingga, aplikasi ini diharapkan dapat dijadikan media pendamping untuk belajar atau memperoleh informasimelalui buku (media cetak lainnya) maupun dari seorang guru (ahlinya).

Question Answering System yang telah dibuat oleh peneliti dapat digunakan menjadi salah satu media pembelajaran atau sarana untuk memperoleh ilmu tentang Islam. Berdasarkan hasil uji coba, QAS yang dibuat dapat menjawab beberapa pertanyaan seputar agama Islam seperti: kapan sholat Idul Adha berlangsung?, kapan terjadinya perang badar?, apa yang dimaksud dengan haji?, apa yang dimaksud dengan shalat?, apa yang dimaksud dengan zakat?, apa yang dimaksud dengan puasa?, kapan waktu yang diharamkan untuk berpuasa?, dimana letak masjid nabawi?, apa saja rukun islam itu?, apa saja rukun iman itu?, kapan waktu shalat dhuha?, apa yang dimaksud dengan shalat istikharah?, dan apa yang dimaksud dengan shalat jamak?





BAB V

PENUTUP

5.1 Kesimpulan

Dari proses perancangan, implementasi hingga uji coba yang telah peneliti lakukan dapat disimpulkan bahwa *Question Answering System* yang telah dibuat menggunakan algoritma *Porter* sebagai *stemmer* kata memberikan nilai akurasi sebesar 71,50 % dari 200 pertanyaan yang telah diujikan. Algoritma *Porter* sebagai *stemmer* kata mempunyai andil dalam menentukan *keyword* atau kandidat kata untuk proses pencarian jawaban. Jika *keyword* yang diperoleh dari proses *stemming* tidak mengalami kesalahan, maka dapat diperoleh jawaban yang sesuai dengan pertanyaan.

5.2 Saran

Dari beberapa kelemahan yang telah dipaparkan pada bab sebelumnya, sistem ini masih memiliki beberapa kekurangan. Untuk penelitian ke depan, perlu dilakukan perbaikan pada pengolahan katanya (*text processing*). Salah satunya adalah perbaikan pada algoritma *stemming*-nya. Untuk meningkatkan akurasi jawaban juga perlu ditambahkan pengklasifikasian pada pertanyaannya sehingga semua *property* pada DBPedia dapat digunakan. Dengan adanya penambahan klasifikasi diharapkan dapat menyajikan jawaban akhir kepada pengguna dengan lebih tepat dan akurat.

BAB V

PENUTUP

5.1 Kesimpulan

Dari proses perancangan, implementasi hingga uji coba yang telah peneliti lakukan dapat disimpulkan bahwa *Question Answering System* yang telah dibuat menggunakan algoritma *Porter* sebagai *stemmer* kata memberikan nilai akurasi sebesar 71,50 % dari 200 pertanyaan yang telah diujikan. Algoritma *Porter* sebagai *stemmer* kata mempunyai andil dalam menentukan *keyword* atau kandidat kata untuk proses pencarian jawaban. Jika *keyword* yang diperoleh dari proses *stemming* tidak mengalami kesalahan, maka dapat diperoleh jawaban yang sesuai dengan pertanyaan.

5.2 Saran

Dari beberapa kelemahan yang telah dipaparkan pada bab sebelumnya, sistem ini masih memiliki beberapa kekurangan. Untuk penelitian ke depan, perlu dilakukan perbaikan pada pengolahan katanya (*text processing*). Salah satunya adalah perbaikan pada algoritma *stemming*-nya. Untuk meningkatkan akurasi jawaban juga perlu ditambahkan pengklasifikasian pada pertanyaannya sehingga semua *property* pada DBPedia dapat digunakan. Dengan adanya penambahan klasifikasi diharapkan dapat menyajikan jawaban akhir kepada pengguna dengan lebih tepat dan akurat.

DAFTAR PUSTAKA

- Agusta, L. 2009. Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem dan Informatika, November 14*. Bali: Universitas Kristen Satya Wacana.
- As Suyuthi, J., & Muhammad Ibnu, J. 2008. *Tafsir Jalalain*. Tasikmalaya: Pustaka Al-Hidayah.
- Ause, W. 1995. *Instant HTML Web Pages*. Chicago: Ziff-Davis Press.
- Beckett, D., & Broekstra, J. 2013. *SPARQL Query Results XML Format*. Dipetik Maret 5, 2014, dari <http://www.w3.org/TR/rdf-sparql-XMLres/>
- Berners-Lee, T., Hendler, J., & Lassila, O. 2001. *The Semantic Web*. American Scientific.
- Djasudarman, & Fatimah. 1993. *Semantik 1: Pengantar ke Arah Ilmu Makna*. Bandung: PT. REFIKA.
- Dolarosa, R. 2008. *Implementasi Question Answering System dengan Metode Rule-Based Pada Banyak Dokumen Berbahasa Indonesia*. Bogor : Institut Pertanian Bogor (IPB).
- Gunawan, & Lovina, G. 2006. Question Answering System dan Penerapannya Pada Alkitab. *Jurnal Informatika Vol. 7 No. 1 Mei* .
- Handojo, A., Lie, E., & Intan, R. 2012. Aplikasi Question Answering System dengan Metode Rule-Based Question Answering System pada Alkitab. *Jurnal Informatika Vol. 11 No. 1 Mei* .
- Ibrahim, N. 2007. Pengembangan Aplikasi Semantic Web Untuk Membangun Web yang Lebih Cerdas. *Jurnal Informatika, Vol. 3 No. 1 Juni* , 27-39.
- Irwan, C., & Ayu Purwarianti, E. 2012. Studi dan Implementasi Question Answering System Bahasa Indonesia pada Perangkat Bergerak Domain National Basketball Association. *Jurnal Sarjana Institut Teknologi Bandung bidang Teknik Elektro dan Informatika Vol. 1, No. 3 Oktober* .
- Keke, D., Chikita, R., & Prayogo, A. D. 2012. Sistem Temu Balik Informasi "Algoritma Nazief dan Adriani". *Jurusan Ilmu Komputer, Fakultas MIPA, Universitas Gadjah Mada* .
- Koivunen, M. R., & Miller, E. 2002. W3C Semantic Web Activity. *Semantic Web Kick-Off in Finland (Vision, Technologies, Research, and Applications)* (hal. 27-43). Helsinki: HIIT Publications.
- Monz, C. 2003. From Document Retrieval to Question Answering. *ILLC Dissertation Series* .
- Pollock, J. T. 2009. *Semantic Web For Dummies*. Indianapolis: Wiley Publishing.

- Sarno, R., Anistyasari, Y., & Fitri, R. 2012. *Semantic Search (Pencarian Berdasarkan Konten)*. Yogyakarta: ANDI.
- Tahri, A., & Tibermacine, O. 2013. DBPedia Based Factoid Question Answering System. *International Journal of Web & Semantic Technology (IJWesT) Vol. 4, No. 3 July* .
- Toba, H. 2010. *Analisis Semantik dengan Representasi "First Order Logic" dalam Sistem Tanya Jawab*. Jakarta: Technical Report Natural Language Processing, Program Doktor Ilmu Komputer Fakultas Ilmu Komputer.
- Zainal, R. 2009. Abstraksi Dokumen Menggunakan Analisis Kausalitas Knowledge Graph.



LAMPIRAN

1. Daftar Stopword

yang, mampu, tentang, di, setelah, semua, hampir, juga, am, antara, dan, ada, seperti, jadi, karena, sudah, tetapi, oleh, bisa, tidak, sayang, melakukannya, lakukan, memang, baik, lain, pernah, setiap, untuk, dari, mendapatkan, punya, telah, memiliki, dia, miliknya, bagaimana, bagaimanapun, i, jika, ke, dalam, itu, sama, paling, biarkan, mungkin, aku, sebagian, besar, harus, saya, off, sering, pada, hanya, atau, kita, sendiri, agak, kata, begitu, beberapa, mereka, kemudian, sana, ini, tis, sungguh, keinginan, adalah, kami, apa, kapan, mana, sementara, siapa, mengapa, akan, dengan, belum, anda, sebuah, atas, menurut, sesuai, seberang, sebenarnya, sekali, lagi, terhadap, memungkinkan, sendirian, bersama, meskipun, selalu, apapun, siapapun, anyways, saja, selain, muncul, menghargai, tepat, sekitar, samping, bertanya, terkait, tersedia, jauh, sangat, menjadi, sebab, sebelum, sebelumnya, belakang, percaya, bawah, terbaik, lebih, luar, keduanya, singkat, namun, ayo, c, datang, cant, penyebab, tertentu, tentu, perubahan, jelas, co, com, akibatnya, mempertimbangkan, berisi, mengandung, saat, pasti, dijelaskan, berbeda, melakukan, dilakukan, selama, edu, misalnya, delapan, tempat, cukup, seluruhnya, terutama, et, dll, bahkan, orang, semuanya, mana-mana, mantan, kecuali, sedikit, kelima, pertama, lima, diikuti, berikut, sebagainya, empat, lanjut, mendapat, diberikan, memberi, pergi, salam, terjadi, halo, membantu, maka, sini, inilah, selanjutnya, disini, sesudah, dirinya, hi, nya, mudah-mudahan, howbeit, yaitu, diabaikan, langsung, lantaran, inc, menunjukkan, ditunjukkan, batin, sejauh, sebaliknya, bukan, terus, tahu, diketahui, terakhir, akhir-akhir, kurang, supaya, mari, suka, kemungkinan, kecil, lihat, mencari, tampak, ltd, banyak, berarti, apalagi, diriku, nama, nd, dekat, perlu, butuh, kebutuhan, baru, berikutnya, sembilan, satu, biasanya, novel, sekarang, oh, ok, oke, tua, keluar, secara, keseluruhan, khususnya, per, ditempatkan, silahkan, plus, menyediakan, que, qv, rd, kembali, benar-benar, mengenai, tanpa, memandang, menganggap, relatif, masing-masing, benar, melihat, berkata, mengatakan, kedua, tampaknya, dilihat, masuk, akal, dikirim, serius, tujuh, seharusnya, boleh, sejak, enam, seseorang, entah, sesuatu, kadang, kadang-kadang, suatu, segera, maaf, ditentukan, tentukan, menentukan, masih, sub, sup, t, menerima, mengambil, katakan, cenderung, th, terima, kasih, thanx, thats, situ, demikian, dalamnya, theres, d, berpikir, ketiga, teliti, seksama, tiga, melalui, seluruh, sehingga, bersama-sama, arah, mencoba, dua, kali, un, sayangnya, sampai, kepada, penggunaan, digunakan, berguna, menggunakan, nilai, berbagai, vs, ingin,

cara, selamat, whereafter, sedangkan, dimana, pun, apakah, utuh, mau, bertanya-tanya, ya, kau, milikmu, dirimu, nol, memperhatikan, a, aduh, agar, akhir, antaranya, apa-apa, awal, ayolah, b, bagian, bahwa, bapak, barang, tadi, baru-baru, barusan, bekas, beliau, berakhir, berbalik, bergerak, besok, biasa, bu, buat, bulat, co, con, contoh, dahulu, lain-lain, depan, detail, detik, manapun, disisi, tengah-tengah, abaikan, diambil, dibuat, didapat, dimanakah, dinyatakan, diperlukan, ditindaklanjuti, puluh, belas, e, ex, g, giliran h, hadir, hadirat, hadirin, hai, haiyaa, hal, hamba, hari, hello, heran, ibu, inc, iya, j, jangan, jatuh, jenis, juta, k, kanan, karenanya, kemana, kemari, kebanyakan, kelompok, kemanakah, kemarin, kenapa, kepentingan, keterangan, ketika, kira, kira-kira, kiranya, kosong, kursus, l, la, pula, lainnya, laki, laki-laki, lalu, lanjutkan, lawan, layak, tinggi, lewat, lusa, m, malah, malahan, masa, masalah, masing, melanjutkan, membuat, membuka, meminta, mempunyai, menarik, menerus, mengisi, menjaga, menyebabkan, menyuruh, merasa, mestinya, miliar, milikkita, mr, mrs, mulai, n, nampak, nomer, nomor, non, nyonya, o, on, p, paduka, pak, panjang, penting, penuh, penyajian, perihal, persis, point, pria, r, rata, rata-rata, ratus, rendah, ribu, runcing, rupiah, s, saudara, saudari, sebagai, sebagaimana, sebelas, segala, sejumlah, sekelompok, selama-lamanya, semoga, seorang, sepanjang, sepenuhnya, sepuluh, sesungguhnya, setengah, setidaknya, siapakah, sisi, nanti, sulung, sungguh-sungguh, tahun, tak, satupun, seorangpun, tambah, tanya, tapi, tebal, terbaru, terbelakang, terbesar, terbuka, terkecil, terlalu, terlihat, ternyata, terpanjang, tertarik, tertinggi, terus-menerus, dapat, tipis, titik, triliun, tuan, u, umum, umumnya, v, w, www, x, y, oleh, meski, sebanyak, cuma, percuma, para, ia, rp, bagi, bahwa, tersebut, merupakan,

2. Tabel Uji Coba

No	Pertanyaan	Jawaban Tepat	Jawaban Tidak Tepat
1	Apa yang dimaksud dengan smartphone?	✓	
2	Apa warna bendera negara Jerman?	✓	
3	Apakah pengertian dari <i>hardware</i> ?		✓
4	Apa nama maskot kota Surabaya?	✓	
5	Apa arti semboyan Bhineka Tunggal Ika?	✓	
6	Apa nama tarian khas Bali?	✓	

7	Apa makanan khas Yogyakarta?	✓	
8	Apakah judul lagu kebangsaan Indonesia?		✓
9	Apa teori yang ditemukan oleh Albert Einstein?	✓	
10	Apa nama planet yang paling dekat dengan matahari?	✓	
11	Apa nama planet yang memiliki cincin?		✓
12	Apa nama kota yang disebut kota hujan?	✓	
13	Apa makanan khas Lamongan?	✓	
14	Apa nama gunung di kota Batu?	✓	
15	Apa nama jembatan di kota Palembang?	✓	
16	Dimana letak sungai Musi?	✓	
17	Dimana letak tugu Monas?	✓	
18	Dimana letak museum Radya Pustaka?	✓	
19	Dimana letak makam Bung Karno?	✓	
20	Dimanakah letak Jembatan Ampera?	✓	
21	Dimanakah Pangeran Diponegoro dimakamkan?	✓	
22	Dimana letak kerajaan Majapahit?	✓	
23	Dimana letak Gunung Bromo?	✓	
24	Dimana tugu pahlawan berada?	✓	
25	Dimana tempat ibadah umat muslim?	✓	
26	Dimana letak jembatan suramadu?	✓	
27	Dimana letak pulau Raja Ampat?	✓	
28	Dimana lokasi Institut Teknologi Telkom?	✓	
29	Dimana lokasi Taman Safari Indonesia 2?	✓	
30	Dimana kota asal bahasa Osing?	✓	
31	Siapa nama gubernur Jawa Timur?	✓	

32	Siapa nama istri Bung Karno?	✓	
33	Siapa pembuat sosial media Facebook?	✓	
34	Siapakah nama walikota Surabaya?	✓	
35	Siapakah CEO Apple Inc?	✓	
36	Siapakah nama kiper Arema?	✓	
37	Siapa pencipta lagu Indonesia Raya?	✓	
38	Siapa pencetus teori gravitasi?		✓
39	Siapa pencetus teori relativitas?	✓	
40	Siapakah CEO Microsoft saat ini?	✓	
41	Siapakah proklamator kemerdekaan Indonesia?	✓	
42	Siapa vokalis Ungu band?	✓	
43	Siapa penemu listrik?		✓
44	Siapa nama istri SBY?	✓	
45	Siapa nama gubernur Bank Indonesia?	✓	
46	Kapan Indonesia dinyatakan merdeka?		✓
47	Kapan Gus Dur mulai menjabat sebagai presiden?	✓	
48	Kapan Dude Harlino lahir?	✓	
49	Kapan Nagita Slavina memulai karir?		✓
50	Kapan tanggal lahir Bunga Citra Lestari?	✓	
51	Kapan sholat Idul Adha berlangsung?	✓	
52	Kapan perayaan natal berlangsung?		✓
53	Kapan masa jabatan Jokowi sebagai Gubernur DKI Jakarta?		✓
54	Kapan Hari Kebangkitan Nasional diperingati?		✓
55	Kapan Apollo 11 tiba di Bulan?	✓	
56	Kapan Soeharto berhenti menjadi presiden?	✓	

57	Kapan PKI dibubarkan?		✓
58	Kapan TNI dibentuk?		✓
59	Kapan terjadinya perjanjian hudaibiah?		✓
60	Kapan terjadinya perang badar?	✓	
61	Mengapa Jepang menyerah terhadap Belanda tahun 1945?		✓
62	Mengapa terjadi perang sampit?	✓	
63	Mengapa Idul Adha disebut juga hari raya haji?	✓	
64	Mengapa Jogja dijuluki sebagai daerah istimewa?		✓
65	Mengapa terjadi gerhana bulan?	✓	
66	Mengapa terjadi gerakan Aceh merdeka?	✓	
67	Mengapa Bogor disebut kota hujan?	✓	
68	Mengapa Bali disebut pulau Dewata?	✓	
69	Mengapa Pontianak disebut kota Khatulistiwa?	✓	
70	Mengapa terjadi krisis moneter tahun 1998?	✓	
71	Mengapa Korea berpisah menjadi 2 wilayah?	✓	
72	Mengapa Malin Kundang dikutuk menjadi batu?	✓	
73	Mengapa Surabaya disebut kota Pahlawan?	✓	
74	Mengapa Edelweiss disebut bunga abadi?		✓
75	Mengapa Timor Leste berpisah dari Indonesia?	✓	
76	Bagaimana proses metamorfosis berlangsung?	✓	
77	Bagaimana proses membuat tempe?		✓
78	Bagaimanakah proses terbentuknya pelangi?	✓	

79	Bagaimana cara menanam jagung?		✓
80	Bagaimana cara menginstal ulang Windows?		✓
81	Bagaimanakah cara merawat Kucing Anggora?		✓
82	Bagaimana cara mengukur presisi?		✓
83	Bagaimana aturan permainan gobak sodor?		✓
84	Bagaimana cara mengukur akurasi?		✓
85	Bagaimana aturan permainan catur?	✓	
86	Bagaimana proses terjadinya hujan?	✓	
87	Bagaimana proses terjadinya fotosintesis?	✓	
88	Bagaimana cara mencangkok tumbuhan?	✓	
89	Bagaimana cara menanam padi?		✓
90	Bagaimana proses terjadinya gerhana matahari?	✓	
91	Berapa panjang jalan tol Surabaya-Gempol?	✓	
92	Berapa tinggi tugu Monas?	✓	
93	Berapa banyak pulau di Indonesia?	✓	
94	Berapa luas Kebun Raya Purwodadi?	✓	
95	Berapa banyak suku yang ada di Indonesia?	✓	
96	Berapa tinggi gunung Bromo?	✓	
97	Berapa luas wilayah Taman Nasional Bromo Tengger Semeru?	✓	
98	Berapa diameter bunga Raflesia?	✓	
99	Berapa jarak Bumi dengan Matahari?	✓	
100	Berapa lama proses rotasi Bumi?	✓	
101	Berapa jarak dari surabaya ke jakarta?	✓	
102	Berapa personel JKT48?	✓	
103	Berapa tekanan darah normal manusia?	✓	

104	Berapa warna yang ada pada pelangi?		✓
105	Berapa jumlah episode Running Man?	✓	
106	Apa yang dimaksud dengan ekosistem?	✓	
107	Apa nama Jakarta tempo dulu?	✓	
108	Apa yang dimaksud dengan web semantik?	✓	
109	Apa provinsi di Indonesia yang mempunyai keistimewaan?	✓	
110	Apa makanan khas kota Palembang?	✓	
111	Apa fungsi dari bank?	✓	
112	Apa fungsi dari ginjal?	✓	
113	Apa saja contoh simbiosis mutualisme?	✓	
114	Apa yang dimaksud dengan down syndrome?	✓	
115	Apa sungai terpanjang di dunia?		✓
116	Apa alat pernapasan dari serangga?		✓
117	Apa yang dimaksud dengan teleskop	✓	
118	Apa saja dampak dari bom nuklir?		✓
119	Apa fungsi dari Posyandu?	✓	
120	Siapa pendiri organisasi Budi Utomo?	✓	
121	Siapa yang disebut sebagai three musketeer?		✓
122	Siapa yang pertama kali menemukan benua Amerika?		✓
123	Siapa saja yang disebut sebagai khulafaurrasyidin?	✓	
124	Siapa yang mengetik naskah teks Proklamasi?	✓	
125	Siapa Ibnu Sina?	✓	
126	Siapa penulis buku Habis Gelap Terbitlah Terang?	✓	
127	Siapa pendiri organisasi NU?		✓

128	Siapa pencipta lagu Rasa Sayange?		✓
129	Siapa Anne Avantie?	✓	
130	Siapa istri Hanung Bramantyo?		✓
131	Dimana Rasulullah dilahirkan?	✓	
132	Dimana terjadi Perang Salib?	✓	
133	Dimana letak Candi Borobudur?	✓	
134	Dimana letak Candi Prambanan?	✓	
135	Dimana letak Laut Merah?	✓	
136	Dimana suku Badui tinggal?	✓	
137	Dimana letak Kepulauan Seribu?		✓
138	Dimana terjadi Perang Puputan?	✓	
139	Dimana pertama kali diadakan SEA GAMES?		✓
140	Dimana letak Danau Singkarak?	✓	
141	Dimana Jokowi dilahirkan?		✓
142	Dimana letak sungai Batanghari?	✓	
143	Dimana letak Gunung Raung?	✓	
144	Dimana letak Patung Pancoran?	✓	
145	Berapa jumlah kadar air dalam tubuh manusia?		✓
146	Berapa macam bahasa yang ada di Indonesia?	✓	
147	Berapa tinggi gunung Kilimanjaro?	✓	
148	Berapa jumlah provinsi di Indonesia?	✓	
149	Berapa luas pulau Kalimantan?		✓
150	Berapa jumlah kabupaten di Indonesia?	✓	
151	Berapa tinggi gunung Salak?	✓	
152	Berapa luas negara Indonesia?		✓
153	Berapa jumlah korban Tsunami di Aceh	✓	

	tahun 2004?		
154	Berapa jumlah korban bom Hiroshima?	✓	
155	Berapa jumlah album Ari Lasso?		✓
156	Berapa hari ayam mengerami telurnya?		✓
157	Berapa lama rotasi planet Mars?	✓	
158	Berapa suhu di Matahari?		✓
159	Kapan Rasulullah wafat?	✓	
160	Kapan terjadi peristiwa Rengasdengklok?	✓	
161	Kapan band Dewa dibentuk?	✓	
162	Kapan hari kelahiran kota Malang?		✓
163	Kapan hari Pendidikan?		✓
164	Kapan Rumah Sakit Syaiful Anwar dibangun?		✓
165	Kapan Masjid Istiqlal dibangun?	✓	
166	Kapan diadakan Konferensi Asia Afrika?		✓
167	Kapan Indonesia menjadi tuan rumah SEA GAMES?	✓	
168	Bagaimana terbentuknya stalaktit?	✓	
169	Bagaimana terjadinya petir?	✓	
170	Bagaimana proses terjadinya hujan?	✓	
171	Bagaimana siklus menstruasi yang normal?	✓	
172	Bagaimana sistem pemerintahan di Indonesia?		✓
173	Bagaimana bisa terbentuk batuan sedimen?	✓	
174	Bagaimana gejala penyakit demam berdarah?		✓
175	Bagaimana mencegah osteoporosis?		✓
176	Bagaimana suara bisa terdengar di telinga kita?		
177	Mengapa Al Qur'an diturunkan secara		✓

	berangsur-angsur?		
178	Mengapa bisa terjadi abrasi?	✓	
179	Mengapa bisa terjadi hujan?	✓	
180	Mengapa bisa terjadi inflasi?	✓	
181	Mengapa bisa terjadi obesitas?		✓
182	Mengapa bumi semakin panas?		✓
183	Mengapa boraks berbahaya jika digunakan di makanan?		✓
184	Mengapa air laut asin?	✓	
185	Mengapa bisa terjadi radiasi?	✓	
186	Apa yang dimaksud dengan haji?	✓	
187	Apa yang dimaksud dengan shalat?		✓
188	Apa yang dimaksud dengan zakat?	✓	
189	Apa yang dimaksud dengan puasa?	✓	
190	Kapan kita dapat melaksanakan haji?		✓
191	Mengapa umat Muslim yang mampu wajib membayar zakat di bulan Ramadhan?		✓
192	Kapan waktu yang diharamkan untuk puasa?	✓	
193	Apa saja macam-macam zakat?		✓
194	Dimana letak masjid Nabawi?	✓	
195	Apa saja rukun islam itu?	✓	
196	Apa saja rukun iman itu?	✓	
197	Kapan waktu shalat dhuhah?	✓	
198	Apa yang dimaksud dengan shalat istikharah?	✓	
199	Bagaimana cara melakukan tayamum?		✓
200	Apa yang dimaksud dengan shalat jamak?	✓	