

# Anotasi Semantik Untuk Pembangunan Korpus Sistem Tanya Jawab Non-Factoid

*by 08032023 Purwatiningtyas*

---

**Submission date:** 08-Mar-2023 02:39PM (UTC+0700)

**Submission ID:** 2031959693

**File name:** 3.\_Anotasi\_Semantik\_Untuk\_Pembangunan.pdf (668.35K)

**Word count:** 2907

**Character count:** 18162

# Anotasi Semantik Untuk Pembangunan Korpus Sistem Tanya Jawab Non-Factoid

Dyah Anjani Ayuningtyias<sup>1</sup>, Fenty Eka Muzayyana Agusin<sup>2</sup>, Ria Hari Gusmita<sup>3</sup>, Purwatiningsy<sup>4</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Universitas Islam Syarif Hidayatullah Jakarta  
Jl. Ir. H. Juanda No. 95, Ciputat, Tangerang Selatan, Banten

<sup>1</sup>dyahanjania@gmail.com

<sup>2</sup>fenty.eka@uinjkt.ac.id

<sup>3</sup>ria.gusmita@uinjkt.ac.id

<sup>4</sup>diba.ian@yahoo.com

<sup>4</sup>Program Studi Teknik Informatika, Universitas Stikubank Semarang  
Jl. Tri Lomba Juang No. 1, Semarang, Jawa Tengah

**Abstrak**— Pembangunan korpus untuk sistem tanya jawab dengan memberi informasi linguistik memiliki hasil yang lebih baik dari sistem tanya jawab yang tidak diberi informasi linguistik. Informasi linguistik yang digunakan adalah dengan metode anotasi semantik. Anotasi merupakan metodologi untuk menambahkan informasi ke dokumen pada tingkat tertentu sedangkan semantik anotasi adalah proses melampirkan tag metadata atau kelas ontologi ke segmen teks, sebagai alat untuk mengambil suatu pengetahuan. Dokumen yang diambil untuk membangun korpus adalah surah terjemahan Al baqarah yang memiliki 286 ayat, dan yang mendukung hanya 14 ayat karena ada beberapa ayat yang tidak mengandung pola yang sesuai sehingga tidak dapat digunakan. Korpus dipecah menjadi 286 file dalam bentuk .txt. Uji coba sistem dilakukan menggunakan bahasa perl dengan membuat sistem tanya jawab sederhana. Sistem tanya jawab merupakan sistem penelusuran informasi yang mengizinkan para user untuk memasukkan pertanyaan dalam bahasa alami

**Kata kunci**— korpus, anotasi semantik, non-factoid, sistem tanya jawab, expected answer type

## I. PENDAHULUAN

Membangun sebuah korpus dengan jumlah yang besar untuk digunakan dalam analisis linguistik dalam bentuk digital dapat dianggap sebagai hasil dari sebuah rujukan, dikumpulkan dan disusun dalam bentuk tulisan tangan selama puluhan tahun. Istilah korpus saat ini paling banyak digunakan untuk merujuk kepada sekumpulan data linguistik yang dikumpulkan untuk tujuan analitik tertentu, dengan anggapan bahwa hal ini akan disimpan, dikelola, dan dianalisa dalam bentuk digital. Namun demikian, linguistik berbasis korpus secara luas masih dianggap sebagai pusat penelitian dengan banyak aspek dalam sifat dan fungsi bahasa Indonesia, dengan aplikasi di bidang-bidang yang beragam seperti leksikografi, pemrosesan

bahasa alamiah, mesin terjemah, dan pembelajaran bahasa. [1].

*Question Answering System* merupakan sebuah sistem yang langsung mengembalikan jawaban kepada pengguna berdasarkan pertanyaan yang diberikannya. Dalam hal ini pengguna tidak memerlukan waktu dan usaha yang lebih untuk menemukan informasi yang dibutuhkan. *Question Answering System* dapat dikelompokkan berdasarkan pertanyaan yang dihasilkan. Tipe pertanyaan yang dapat diproses pada sebuah sistem *Question Answering System* terbagi atas 5 jenis pertanyaan yaitu *factoid*, *non-factoid*, *yes-no*, *list* dan opini [2].

Pada penelitian [3] persentase pertanyaan yang di uji dengan menggunakan beberapa passages, yang memberikan informasi bahwa sistem menjawab 37 pertanyaan dengan benar dari 60 pertanyaan yang masing-masing 30 pertanyaan dengan masing-masing persentase 61,67% benar, 1,67% tidak tepat, 1,67% tidak mendukung, 13,33% salah, 22% tidak terjawab, pada penelitian yang dilakukan Lestari [4] persentase dari pengembangan sistem yang dilakukan secara tata bahasa diperoleh persentase 32% jawaban benar, 8% jawaban salah, dan 60% jawaban tidak terjawab sedangkan pada penelitian Nadisah [5], persentase hasil dari *factoid* 100% jawaban benar dan *non factoid* 80% jawaban benar, 6,6% jawaban salah, 6,6% jawaban tidak terjawab dan 6,6% jawaban tidak tepat. Berdasarkan hasil analisa terhadap ketiga penelitian tersebut diketahui bahwa ketiganya masih menggunakan korpus yang tidak dilengkapi dengan informasi linguistik, hal ini mempengaruhi kinerja dari komponen perolehan jawaban pada setiap *Question Answer System*.

Salah satu untuk meningkatkan kinerja *Question Answer System* adalah dengan memberikan informasi linguistik terhadap korpusnya dan sudah dilakukan hal tersebut untuk *Factoid Question Answer System* sebagai usaha untuk meningkatkan *Question Answer System* yang sudah

di bangun terlebih dahulu [6]. Namun hal ini belum ada upaya untuk membangun korpus dengan informasi linguistik pada *Non-Factoid Question Answer System*. Pada penelitian ini penulis akan membangun korpus dengan memberikan informasi linguistik untuk digunakan pada *Non-Factoid Question Answer System*.

## II. METODOLOGI

Penelitian ini dilakukan dengan mempelajari penelitian-penelitian yang terkait. Membangun korpus terjemahan surah Al Baqarah untuk kata tanya *Non Factoid* dengan anotasi semantik dilakukan untuk mengetahui kinerja korpus, lebih baik atau lebih buruk dengan korpus yang tidak diberi anotasi semantik. Langkah yang dilakukan untuk membangun korpus ini, yaitu Analisis dokumen terjemah Al Baqarah untuk menemukan pola kalimat dalam surah terjemahan Al-Baqarah.

Analisis dilakukan secara manual dengan tahapan: (1) membaca dan memahami setiap terjemahan ayat pada surat Al Baqarah; (2) menentukan apakah konten ayat tersebut menjelaskan jawaban dari jenis kata tanya non-factoid yaitu apa, mengapa, dan bagaimana hal ini dikarenakan korpus di peruntukkan untuk sistem tanya jawab *non-factoid*; (3) Ayat-ayat yang terpilih dikelompokkan berdasarkan *Expected Answer Type* (EAT) dari kata tanya apa, mengapa, dan bagaimana.

Langkah selanjutnya adalah membuat definisi model anotasi semantik. Model anotasi semantik yaitu proses menentukan simbol dan pemisahan kata yang tepat pada surah Al Baqarah. Simbol yang digunakan pada tahap model anotasi semantik adalah < >, pemilihan simbol tersebut diambil karena untuk memudahkan penulis membedakan tanda-tanda yang terdapat pada surah Al Baqarah contohnya ( ), simbol tersebut juga digunakan saat tahapan model semantik. Masing-masing kata tanya memiliki anotasi yang berbeda-beda. Setelah proses model anotasi dilakukan kemudian di dapat hasil dari model tersebut, maka dibuatlah proses anotasi, dalam hal ini proses anotasi dilakukan secara manual terhadap 286 dokumen menggunakan notepad++, anotasi dilakukan dengan menganalisa terlebih dahulu dokumen yang digunakan kemudian memberi tanda bagian-bagian yang akan di anotasi.

## III. HASIL DAN PEMBAHASAN

Terjemahan Surat Al Baqarah yang 286 ayat itu turun di Madinah yang sebagian besar diturunkan pada tahun Hijriah, kecuali ayat 281 diturunkan di Mina pada *Hajji wadaa'* (haji Nabi Muhammad s.a.w. yang terakhir). Seluruh ayat dari surat Al Baqarah termasuk golongan Madaniyyah, yang merupakan surat terpanjang di antara surat-surat Al Quran. Pada surah Al Baqarah terdapat pula ayat yang terpanjang yang tertera pada ayat 282. Surat ini dinamai Al Baqarah karena di dalamnya disebutkan kisah penyembelihan sapi betina yang diperintahkan Allah kepada Bani Israil (ayat 67 sampai dengan 74), yang

dijelaskan watak orang Yahudi pada umumnya. Dinamai *usthaatul-Quran* (puncak Al Quran) karena memuat beberapa hukum yang tidak disebutkan dalam surat yang lain. Dinamai juga surat *alif-laam-miim* karena surat ini dimulai dengan *Alif-laam-miim*.

Kata tanya *non factoid* merupakan bentuk kata tanya yang memiliki jawaban uraian. pada terjemahan Al Baqarah banyak penjelasan yang bersifat uraian namun tidak sebanyak jawaban yang bersifat factoid atau pernyataan. Kata tanya non factoid terdiri dari apa, mengapa, bagaimana masing-masing kata tanya memiliki beberapa kandidat dan karakter yang berbeda. Berikut ini merupakan contoh kandidat ayat yang memiliki ciri EAT berupa: (1) definisi (*Definition*), (2) metode (*method*), dan (3) alasan (*Reason*).

Dari 286 ayat yang selanjutnya disebut dokumen korpus terjemahan surah al baqarah, diperoleh 31 kandidat ayat yang akan dilakukan anotasi semantik. Kandidat ayat tersebut dirinci sebanyak 16 dokumen yang mengandung ciri definisi (kata tanya "Apa"). Dokumen yang memiliki ciri alasan yaitu yang mengandung kata tanya "Mengapa" terdiri dari 8 ayat. Dokumen yang memiliki ciri metode atau kata tanya "Bagaimana" terdiri dari 7 ayat.

Salah satu contoh terjemah ayat nomor 61 pada surat Al Baqarah:

*dan (ingatlah), ketika kamu berkata: "hai musa, kami tidak bisa sabar (tahan) dengan satu macam makanan saja. sebab itu mohonkanlah untuk kami kepada tuhanmu, agar dia mengeluarkan bagi kami dari apa yang ditumbuhkan di bumi yaitu sayur - mayurnya, ketimunnya, bawang putihnya, kacang adasnya, dan bawang merahya". musa berkata: "maukah kamu mengambil yang rendah sebagai pengganti yang lebih baik? pergilah kamu ke suatu kota, pasti kamu memperoleh apa yang kamu minta". lalu ditimpahkanlah kepada mereka nista dan kehinaan, serta mereka mendapat kemurkaan dari allah. hal itu (terjadi) karena mereka selalu mengingkari ayat-ayat allah dan membunuh para nabi yang memang tidak dibenarkan. demikian itu (terjadi) karena mereka selalu berbuat durhaka dan melampaui batas.*

Dari terjemah ayat 61 tersebut diperoleh analisis sebagai berikut:

*Clue:* Yaitu

*Term:* ditumbuhkan di bumi

Contoh Pertanyaan: Apa yang ditumbuhkan di bumi?

Selanjutnya diberikan contoh analisis pada ayat 68:

*Musa menjawab: "sesungguhnya Allah berfirman bahwa sapi betina itu adalah sapi betina yang tidak tua dan tidak muda.*

Dari terjemah ayat 68 diperoleh analisis sebagai berikut:

Clue: Adalah

Term: Sapi Betina

Contoh Pertanyaan: Apa yang dimaksud dengan sapi betina?

Surah Al Baqarah ayat 61 dan 68 memiliki *clue word* "yaitu" dan "adalah" yang merupakan acuan dari kalimat yang mengandung definisi (kata tanya "Apa"). *Term*, digunakan sebagai acuan pada setiap kata tanya yang diberikan user. Dengan cara yang sama dapat diperoleh *clue word* pada kandidat ayat yang terangkum pada tabel berikut.

TABEL I  
EAT KATA TANYA NON-FACTOID

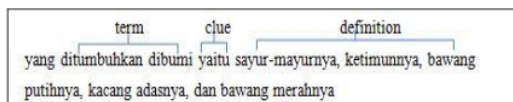
EAT	Pertanyaan	Clue word
Definition	Apa, Apakah	Adalah, ialah, yaitu, merupakan, artinya, berarti, arti, maksud, dikenal, makna
Reason	Mengapa	Karena, oleh karena itu, penyebab, sebab, akibat, oleh sebab itu
Method	Bagaimana	Dengan cara, sebagaimana, dengan, proses, prosedur

Dalam proses pengelompokkan tidak semua terjemahan Al Baqarah memiliki kata tanya yang tepat atau jawaban yang sesuai, contohnya pada terjemahan ayat 3 :

(yaitu) mereka yang beriman kepada yang ghaib, yang mendirikan shalat, dan menafkahkan sebahagian rezeki yang Kami anugerahkan kepada mereka.

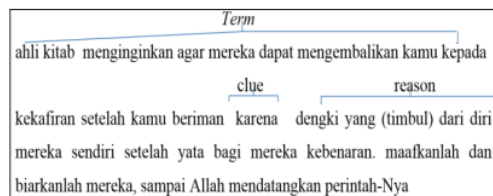
Pada terjemahan tersebut menjelaskan bahwa ayat tersebut memiliki *clue word* yang sesuai dengan kata tanya apa (yaitu), namun kata ganti "mereka" menunjukkan jawaban untuk kata tanya siapa atau kata tanya *factoid*.

Proses selanjutnya adalah anotasi atau memberikan tanda pada kata-kata unik yang akan di jadikan sebagai: <clue>, <term>, <definition/method/reason>.



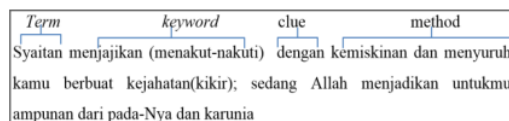
Gambar 1. Contoh Anotasi Kategori Definition

Gambar 1 menunjukkan contoh proses anotasi pada dokumen terjemah Al Baqarah untuk kategori *definition* (kata tanya "Apa").



Gambar 2. Contoh Anotasi Kategori Reason

Gambar 2 menunjukkan *clue* dengan kata "karena" yang merupakan ciri kata tanya yang mengharap jawaban berupa alasan atau kategori pertanyaan *reason* (kata tanya "Mengapa"). Sedangkan kategori pertanyaan dengan jawaban berupa metode dapat dilihat pada gambar dibawah ini.



Gambar 3. Contoh Anotasi Kategori Method

Dari ketiga model anotasi semantik yang mengandung definisi, metode dan sebab akibat, dapat disimpulkan bahwa ketiganya memiliki pola anotasi semantik yang berbeda. Pola anotasi semantik tersebut dapat dilihat pada tabel dibawah ini.

TABEL III  
POLA MODEL ANOTASI SEMANTIK

APA	MENGAPA	BAGAIMANA
Term+clue+definition. Keyword+clue+term+definition.. Term+keyword+clue+definition.. Keyword+term+clue+definition..	Term+keyword+clue+reason.. Keyword+reason+clue+term+keyword.. Term+clue+reason.. Keyword+term+clue+reason..	Term+keyword+method(clue).. Term+keyword+clue+method.. Term+keyword+clue+method+term..

Pola model anotasi ini dilakukan pada awal analisa menggunakan model semantik, pola model anotasi memiliki perubahan dikarenakan jawaban yang akan diberikan sistem kepada *user* tidak mengandung jawaban yang mengandung unsur definisi, sebab dan metode. Oleh karena itu ditemukan kembali model anotasi semantik yang baru.

TABEL IIIII  
MODIFIKASI MODEL ANOTASI SEMANTIK

APA	MENGAPA	BAGAIMAN A
<term> + <clue> + <definition> <keyword> + <clue> + <term> + <definition> <keyword> + <term> + <clue> + <definition>	<term> + <keyword> + <clue> + <reason>	<term> + <keyword> + <clue> + <method>

Modifikasi model anotasi semantik ini karena banyak jawaban yang tidak mengandung penjelasan dari kata tanya yang diberikan *user*, namun hanya terdapat *clue word* yang merupakan kelompok dari kata tanya *non factoid*.

Selanjutnya, dokumen terjemah Al Baqarah diubah menjadi 286 file dengan bahasa pemrograman perl. Sedangkan untuk proses memberikan *tagging* dilakukan secara manual. Berikut contoh proses anotasi yang dilakukan:

dan (ingatlah), ketika kamu berkata: "hai musa, kami tidak bisa sabar (tahan) dengan satu macam makanan saja. sebab itu mohonkanlah untuk kami kepada tuhanmu, agar dia mengeluarkan bagi kami dari apa yang <term>ditumbuhkan di bumi</term><clue>yaitu </clue><definition>sayur - mayurnya, ketimunnnya, bawang putihnya, kacang adasnya, dan bawang merahnya</definition>". musa berkata: "maukah kamu mengambil yang rendah sebagai pengganti yang lebih baik? pergilah kamu ke suatu kota, pasti kamu memperoleh apa yang kamu minta". lalu ditimpahkanlah kepada mereka nista dan kehinaan, serta mereka mendapat kemurkaan dari allah. hal itu (terjadi) karena mereka selalu mengingkari ayat-ayat allah dan membunuh para nabi yang memang tidak dibenarkan. demikian itu (terjadi) karena mereka selalu berbuat durhaka dan melampaui batas.

Gambar 4. Anotasi Semantik Pada Dokumen

Cara kerja sistem tanya jawab adalah sebagai berikut: (1) *user* memasukkan pertanyaan, kemudian pertanyaan tersebut dianalisa untuk mendapatkan tipe jawaban; (2) Tipe jawaban diperoleh dengan memeriksa kata tanya yang digunakan *user* dengan cara melakukan pencocokan pada kalimat pertanyaan. Jika ditemukan kata tanya yang sesuai maka sistem akan menyimpan tipe jawaban yang sesuai dengan kalimat pertanyaan tersebut. Contoh : " Apa definisi bulan ramadhan". Dari kalimat pertanyaan tersebut sistem akan mendapatkan kata tanya yang digunakan adalah " Apa" sehingga dapat diketahui tipe jawaban dari kata tanya ini adalah "definisi"; (3) dari kalimat pertanyaan tersebut sistem menghilangkan kata tanya dan kata penghubung dari kalimat pertanyaan tersebut. Pada contoh tersebut dihasilkan: "definisi bulan

ramadhan". Dari kalimat tersebut untuk menghasilkan *keyword* dengan menghilangkan kata penghubung, jika kalimat tersebut memiliki kata penghubung maka kalimat pertanyaan hanya tersisa memiliki *keyword* saja yaitu: "ramadhan". (4) *Keyword* yang telah didapat akan disimpan oleh sistem untuk di cocokkan ke dalam korpus. Sistem akan mencari dokumen yang sesuai dengan *keyword* dari kalimat pertanyaan tersebut.

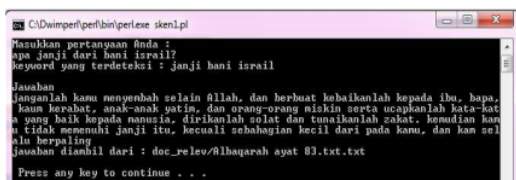
Dokumen tersebut akan di masukkan dalam satu folder kemudian pada tahapan perolehan jawaban. Pengambilan jawaban dilakukan dengan ketentuan (1) temukan bagian *term* di setiap dokumen; (2) bandingkan *keyword* dari pertanyaan dengan bagian *term*; (3) jika *keyword* dan *term* sama, maka kembalikan bagian definisi/metode/alasan sebagai jawaban. Dari hasil perolehan jawaban tersebut dilakukan pengujian yaitu dengan membangun sistem tanya jawab dan menguji sistem tersebut dengan memanfaatkan korpus yang sudah dianotasi.

Pengujian sistem tanya jawab sederhana menggunakan bahasa perl. Sistem meminta *user* mengajukan pertanyaan, sistem akan mengambil kata kunci dari pertanyaan tersebut. *Keyword* yang cocok dengan yang ada dalam korpus akan dipilih oleh sistem untuk diberikan jawaban yang sesuai. Pengujian skenario 1 yaitu menguji sistem tanya jawab menggunakan data *Benchmark*. Data *Benchmark* adalah daftar pertanyaan yang jawabannya sudah pasti ada pada dokumen.

No	Daftar pertanyaan
1	Apakah ditumbuhkan di bumi?
2	Apakah dimaksud sapi betina?
3	Apakah janji dari bani israil?
4	Apakah perumpamaan orang yang kafir?
5	Apakah dimaksud kebajikan?
6	Apakah pengganti jika tidak berpuasa?
7	Apakah bulan ramadhan itu?
8	Apakah dimaksud bulan sabit?
9	Apakah dimaksud berfidyah?
10	Apakah dimaksud musim haji?
11	Apakah dimaksud haidh?
12	Mengapa ahli kitab menginginkan agar dapat mengembalikan kekafiran setelah beriman?
13	Bagaimana Musa memohon air untuk kaumnya?
14	Bagaimana syaitan menjanjikan serta menak-nakuti?
15	Apakah tanda mukjizat itu?
16	Apakah semua makanan halal atau haram bagi israil?
17	Apakah perumpamaan harta yang dinafkahkan di dunia?
18	Apakah itu al quran?
19	Mengapa orang-orang kafir diberi rasa takut?
20	Bagaimana siksa untuk orang-orang yang kafir?

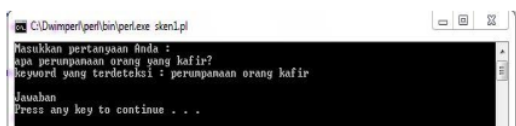
Gambar 5. Contoh Data Benchmark

Gambar 5 merupakan contoh data *benchmark*, yaitu daftar pertanyaan yang sudah disediakan jawabannya di korpus. Pada skenario 1 sistem tanya jawab akan memilah jawaban menjadi 2 kategori, yaitu Benar dan Tidak Terjawab. Sebagai contoh, *user* menuliskan pertanyaan "apa perumpamaan orang kafir". Sistem akan mencari jawaban yang sesuai dengan pertanyaan pada korpus, jika sistem tidak mengembalikan jawaban, maka pertanyaan tersebut termasuk kategori tidak terjawab. Berikut ini tampilan pertanyaan yang menghasilkan jawaban Benar.



Gambar 6. Skenario 1 Jawaban Benar

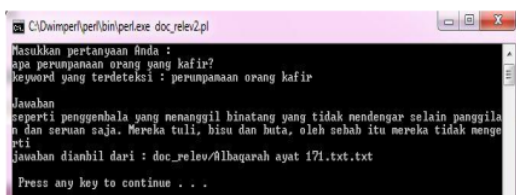
Gambar 6 menunjukkan sistem dapat mengembalikan jawaban dengan benar karena *keyword* yang terdeteksi cocok dengan *term* yang ada pada dokumen korpus. Sedangkan pada gambar dibawah ini adalah contoh sistem yang tidak dapat mengembalikan jawaban dari pertanyaan *user*. Sedangkan pada gambar 7 dibawah ini menunjukkan pertanyaan yang Tidak Terjawab.



Gambar 7. Skenario 1 Pertanyaan yang Tidak terjawab

Pengujian skenario 1 sistem tanya jawab memberikan hasil jawaban Benar sebanyak 55%, dan Tidak Terjawab 45%. Hal ini dapat disimpulkan bahwa sistem belum mampu menjawab semua pertanyaan dengan benar. Pertanyaan yang tidak terjawab disebabkan *keyword* pada pertanyaan tidak sama dengan *term* pada data *benchmark* sehingga sistem tidak mampu mengambil jawaban dari pertanyaan tersebut.

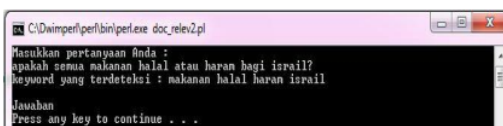
Selanjutnya dilakukan perbaikan pada sistem tanya jawab, dan melakukan pengujian dengan skenario kedua. Skenario 2 dilakukan dengan menerapkan perbandingan kemiripan antara *keyword* dan *term*. Perbandingan berdasarkan kemiripan adalah *keyword* pada pertanyaan dan *term* di dokumen korpus dinilai mirip ketika semua kata pada *string keyword* muncul pada *string term*. Contoh pengujian skenario 2 adalah memperbaiki hasil pertanyaan pada gambar 7 diatas.



Gambar 8. Skenario 2 Menghasilkan Jawaban Benar

Pada gambar 8 diatas merupakan hasil perbaikan sistem dari skenario 1, gambar diatas adalah perbaikan dari gambar 7 yaitu *keyword* tidak terjawab karena adanya kata alias pada *term*. Setelah dilakukan perbaikan skenario pengujian, diperoleh hasil jawaban yang Benar sebanyak

90%, dan yang tidak terjawab 10%. Salah satu contoh pertanyaan yang tidak terjawab adalah sebagai berikut.



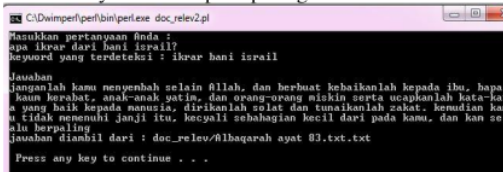
Gambar 9. Skenario 2 yang Tidak Terjawab

Untuk memperbaiki sistem tanya jawab ini, dibuatlah skenario 3 dengan pengembangan kata yang mengandung sinonim. Dibawah ini adalah contoh pertanyaan sinonim.

No	Daftar pertanyaan sinonim
1	Apa yang ditumbuhkan di alam?
2	Apa yang dimaksud lembu betina?
3	Apa ikrar dari bani israil?
4	Ibarat apakah orang yang kafir?
5	Mengapa ahli kitab mengharapkan agar dapat mengembalikan kekafiran setelah beragama?
6	Bagaimana Musa meminta air untuk kaumnya ?
7	Apa saja tanda kebenaran itu ?
8	Apakah ungkapan untuk harta yang dinafkahkan di dunia?
9	Mengapa orang-orang kafir diberi rasa cemas?
10	Bagaimana azab untuk orang-orang yang kafir?

Gambar 10. Contoh Pertanyaan Sinonim

Gambar 10 adalah daftar pertanyaan sinonim dari daftar pertanyaan dalam data *benchmark*. Sebagai contoh kata bumi diberi sinonim alam, kata janji diberi sinonim ikrar. Dengan daftar ini dilakukan pengujian sistem kembali, dan hasilnya adalah seperti pada gambar dibawah ini.



Gambar 11. Skenario 3 Menghasilkan Jawaban Benar

Pada gambar diatas sinonim yang diberikan adalah "ikrar" yang merupakan kata sinonim dari "janji". Jawaban yang diberikan oleh sistem adalah benar dan sesuai dengan skenario 3. Dari ketiga hasil skenario di atas penulis dapat mengambil kesimpulan bahwa korpus beranotasi semantik dapat menjawab dengan benar dan dapat digunakan dengan kata yang mengandung sinonim.

#### IV. KESIMPULAN

Korpus dengan anotasi semantik dilakukan dengan beberapa tahap yaitu menganalisa dokumen, menentukan kandidat korpus, pengelompokan dokumen berdasarkan kategori definisi, metode dan alasan. Setelah membangun korpus dilanjutkan membuat model anotasi semantik dan melakukan proses anotasi. Korpus dibangun terdiri 286 ayat yang dikelompokkan dalam 18 kategori definisi, 6 kategori metode dan 7 kategori alasan.

Berdasarkan hasil analisa pengujian terhadap korpus beranotasi semantik, korpus dengan informasi linguistik memiliki kinerja lebih baik dibanding korpus yang tidak memiliki informasi linguistik. 3. Hasil dari uji coba pada sistem tanya jawab yang di bangun berdasarkan data korpusnya yang sama-sama dari dokumen terjemahan, diberikan 20 pertanyaan yang mengandung kata tanya apa, mengapa dan bagaimana memiliki presentase 90% jawaban terjawab dengan benar dan 10% sistem tidak menjawab hal ini membuktikan bahwa korpus yang diberi informasi linguistik mampu memberikan hasil yang baik.

#### REFERENSI

- [1] Jati, Sasongko. "Aplikasi untuk Membangun Corpus dari Data Hasil Crawling dengan Berbagai Format Data Secara Otomatis". *J. Dinamik*, Vol 15 No. 1, 2010.
- [2] A. Purwarianti and N. Yuslimi, "Sistem Question Answering Bahasa Indonesia untuk Pertanyaan Non-Factoid," *J. Ilmu Komput. dan Inf.*, vol. 4, no. 1, pp. 10–14, 2012
- [3] Z. Naf'an and R. H. Gusmita, "Development of an Indonesian Question Answering System about Khulafaur Rasyidin's History," in *Proceedings of The 1st International Conference on Cyber & IT Service Management in Conjunction with the ITIL v. 3 Workshop, Training and Certification.*, 2012.
- [4] Lestari, T.H., "Pengembangan Sistem Tanya Jawab Tentang Khulafaur Rasyidin", UIN Syarif Hidayatullah Jakarta (Unublish), 2014.
- [5] Nadisah H.A., 2015. "Pengembangan Sistem Tanya Jawab Untuk Pertanyaan Factoid dan Non-Factoid (Kasus: Sejarah Khulafaur Rasyidin)", UIN Syarif Hidayatullah Jakarta, (Unpublish), 2015.
- [6] Sukmana, Husni Teja, Ria Hari Gusminta, Yusuf Durachman, and Asep Fajar Firmansyah. "Semantically annotated corpus model of Indonesian Translation of Quran: An effort in increasing question answering system performance." In *2016 4th International Conference on Cyber and IT Service Management*, pp. 1-5. IEEE, 2016.

# Anotasi Semantik Untuk Pembangunan Korpus Sistem Tanya Jawab Non-Factoid

---

## ORIGINALITY REPORT

---

9%

SIMILARITY INDEX

8%

INTERNET SOURCES

2%

PUBLICATIONS

4%

STUDENT PAPERS

---

## MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

3%

★ [www.hidayatullah.or.id](http://www.hidayatullah.or.id)

Internet Source

---

Exclude quotes  On

Exclude matches  < 1%

Exclude bibliography  On