***Interpretable* STS Terjemahan Ayat Alquran dengan Menggunakan *Chunking*  otomatis**

**Interpretable STS Koran verse translation Using Automated *Chunking***

**Tugas Akhir**

**Maulisye Audina Ulfa**

**1103134354**

**Diajukan untuk memenuhi sebagian dari syarat untuk memperoleh gelar sarjana Teknik Informatika Fakultas Informatika**

**Universitas Telkom**



**Program Studi Sarjana Teknik Informatika**

**Fakultas Informatika**

**Universitas Telkom**

**Bandung**

**2017**

# Lembar Pernyataan

Saya yang bertanda tangan dibawah ini :

Nama : Maulisye Audina Ulfa

NIM : 1103134354

Email : maulisyeaudinaulfa@gmail.com

HP :081221952007

Menyatakan bahwa Tugas Akhir dengan judul " Interpretable STS Terje-mahan Ayat Alquran Menggunakan *Chunking* otomatis " merupakan karya orisinal saya sendiri. Saya tidak melakukan penjiplakan kecuali melalui pengutipan sesuai dengan etika keilmuan yang berlaku. Saya bersedia menanggung sanksi yang dijatuhkan kepada saya apabila ditemukan pelanggaran terhadap etika keilmuan dalam karya ini, atau ditemukan bukti yang menunjukan ketidakaslian karya ini.

Bandung, 12 Januari 2017

Yang membuat pernyataan,

Maulisye Audina Ulfa

NIP: 1103134354

# Lembar Persetujuan

***Interpretable* STS Terjemahan Ayat Alquran Menggunakan *Chunking*  otomatis**

***Interpretable STS* Koran *verse translation Using Automated Chunking***

**Maulisye Audina Ulfa**

**NIM : 1103134354**

Tugas Akhir ini diterima dan disahkan untuk memenuhi sebagian dari syarat untuk memperoleh gelar pada Program Studi Sarjana Teknik Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 12 Januari 2017

Menyetujui,

|  |  |
| --- | --- |
| Pembimbing 1  Moch. Arif Bijaksana, Ph.D NIP. 03650312-4 | Pembimbing 2  Said Al Faraby, S.T., M.Sc. NIP. 15891889-1 |

|  |
| --- |
| Mengesahkan,  Kepala Program Studi Teknik Informatika  Moch. Arif Bijaksana, Ph.D NIP. 03650312-4 |

# Abstrak

Penelitian dibidang Linguistik memiliki daya tarik yang cukup besar saat ini, salah satunya adalah penelitian *interpretable* STS (*Semantic Textual Simillarity*) yang merupakan penelitian untuk mengukur tingkat kesamaan makna antara dua buat teks dengan cara pemberian label keterhubungan dan skor kesamaan antar teks berdasarkan proses *alignment*nya. Tujuan Tugas Akhir ini yaitu membuat program untuk menghitung tingkat kesamaan makna antara pasangan ayat Alquran dalam terjemahan bahasa Inggris versi *saheeh international*  dengan proses *Chunking* yang otomatis*.* Untuk melakukan penelitian ini, dibuat sistem berupa *unsupervised system*. Garis besar tahapan yangdilakukan dalam penelitian ini yaitu mengumpulkan data pasangan terjemahan ayat Alquran, selanjutnya datamelalui proses *Chunking* secara otomatis oleh sistem lalu masuk pada proses *alignment.* Selanjutnyadiberikan label keterhubungan antar potongan ayat Alquran, dan tahap terakhir yangdihitung skor kesamaan antara potongan ayat Alquran. Adanya penelitian ini diharapkan dapat mengetahui secara optimal mengenai kesamaan makna pasangan potongan terjemahan Alquran melalui nilai kesamaan dan label keterhubungan yang dimiliki. Penelitian ini memiliki nilai akurasi untuk sistem *chunking* sebesar === dan

**Kata Kunci :** *Semantic Textual Similarity****,*** *Interpretable STS,**unsupervised system, Chunking, alignment, Alquran*

# Abstrack

Research in the field of Linguistics has an appeal that is large enough at the moment, one of which is the research interpretable STS (Semantic Textual Simillarity) which is a study to measure the level of similarity of meaning between the two create a text by labeling of connectedness and similarity score between the text based alignment process. In this final project will create a program to calculate the degree of similarity of meaning between a couple verses of the Koran in English translation Saheeh international version with automated Chunking. To conduct this research, the system will be made in the form of unsupervised system. Outlines steps to be taken in this study is to collect data translation couple verses of the Koran, then the data will go through the process Chunking system automatically and entered in the alignment process. Next will be given label connectivity between pieces of verses of the Koran, and the last stage will be counted scores of similarity between pieces of verses of the Koran. The existence of this study are expected to know about the similarity of meaning optimally couple pieces of translations of the Koran through similarity value and a label owned. This study refers to *SemEval2016*  or Semantic Evaluation 2016 is one of the world competition in the field of Semantic Similarity Textual (STS)[[1]](#footnote-1).

**Keywords:** Semantic Textual Similarity,Interpretable STS, unsupervised system, Chunking, alignment

# Lembar Persembahan

Proses pengerjaan Tugas Akhir tidak mencapai tujuannya jika tidak ada Sang Pencipta dan orang-orang hebat dibelakangnya. Untuk berbagai pihak yang telah memberikan dukungan sehingga Tugas Akhir dapat diselesaikan tepat waktu. Penulis ingin mengucapkan terima kasih kepada:

1. Allah Subhanahu wa Ta’ala yang sepanjang waktu senantiasa memberikan rahmat kasih sayang, ridho-Nya, dan menjadi naungan bagi penulis selama proses pengerjaan Tugas Akhir . Serta terima kasih karena selalu memberikan nikmat iman dan islam kepada penulis.
2. Baginda Nabi Besar Muhammad Sallallahu’alaihi wa Salam yang telah memberikan pembelajaran terbaik dan menuntun umat dari zaman kebodohan menuju zaman yang penuh cahaya. Terima kasih karena engkau telah menjadi media bagaimana sebuah peradaban islam tumbuh berkembang.
3. Alm ibu tercinta yang telah mengajarkan makna sesungguhnya kehidupan ini, terima kasih karena engkau tak pernah meninggalkan sedetikpun anakmu ini. "Bolehlah tubuh berpisah, ketika kenangan senantiasa bersemayam dilubuk hati"
4. Bapak tercinta yang tak pernah padam semangatnya untuk membimbing anak-anaknya menuju kebaikan hidup. Terima kasih bapak, karena bapak adalah manusia yang tak pernah mengeluh dalam hidupnya.
5. mas Maulana saiful imron yang selalu mengajarkan arti kerja keras yang sesungguhnya, mas yang selalu memotivasi untuk menjadi manusia yang lebih baik dari waktu ke waktu.
6. dek Avida Salasyatin A.N yang merupakan adekku satu-satunya, terima kasih karena kamu adalah orang yang selalu mengatakan "BISA" ketika sangat kecil kemungkinan harapan bisa terjadi.
7. Keluargaku tercinta : keluarga Besar Bani Basri , semoga kita semua selalu dimudahkan jalannya baik di dunia maupun diakhirat.
8. Bapak Moch. Arif Bijaksana selaku pembimbing I dan serta Bapak Said Al Faraby pembimbing II penulis yang selalu memberikan ilmu, arahan dan motivasi kepada penulis dalam menyelesaikan Tugas Akhir. Terima kasih telah meluangkan waktu, tenaga dan pikiran kepada penulis. Semoga ilmu yang telah diberikan serta hasil penelitian Tugas Akhir ini menjadi manfaat bagi umat dan menjadi Amal jariyah. Aamiin
9. Bapak Anditya arifianto selaku dosen wali IF-37-04 yang selalu memberikan arahan terbaik bagi anak-anak walinya.
10. Laboratorium BASIS DATA yang telah menjadi bagian hidup saya selama tiga tahun ini. Mulai dari Study grup, Riset grup sampai menjadi aslab, semua kenangan indah yangterpatri dalam hati. Terima kasih untuk teman-teman lab saya tercinta yang selalu menjadi media meluapkan emosi muehehe, selalu menjadi media pembelajaran profesionalitas yang baik. Terima kasih Habib fikri, Matias, Helmi, Saepul, Fanny, Diska, Gina yang selalu mengingatkan saya untuk mengerjakan Tugas Akhir ini. Terima kasih untuk Tiara anissa dionti, Eriza, Daud sandy, Dwitika diah yang dari awal masuk lab basdat sampai akhir tetap setia dengan lab tercinta ini.
11. RSPQ ( Rumah Singgah Pencinta Alquran ) yang selalu menjadi rumah untuk hati-hati yang luar biasa, berkumpul bersama, belajar bersama, bermain bersama, melewati suka duka mendidik pelurus bangsa yang hebat. Memori-memori ini yang sangat berharga dan takpernah terlupakan. Terima kasih untuk kalian pengajar hebat yangsela-lu hebat Fitriana Istiqomah, Tau k Abdurahman, Aulia, Lina, Rahma, Audina, dan masih banyak lagi. Terima kasih juga kepada pengajar terdahulu yang memberikan kesempatan kepada kami untuk memegang tongkat estafet kepimpinan RSPQ Triadi N, Wahyu M, Rizal, Hassan, Achmad, dan Dwijayanti selaku wanita yang luar biasa tangguhnya di RSPQ.
12. AL-FATH adalah lembaga dakwah kampus yang sangat istimewa, selama tiga tahun penulis besar dilingkungan yang aman, lingkungan yang se-nantiasa melindungi dan menjaga amalan yaumiah, terima kasih Al-fath. " Setiap masa ada penerusnya dan setiap penerus ada masanya"
13. Ignatius Nicolas, Ika Monika, Febriana Sawitri, Bening Suryani yang selalu ada disaat penulis mengalami masa-masa sulit ketika mengerjakan Tugas Akhir ini, “ tanpa kalian apa jadinya aku “.
14. BFF Maulina Istiana Astari , Fi nella Rahma, Dyah Ayu Cintya Dewi, Dwitika Diah, Maya Budi Rahayu terima kasih karena telah memberikan pertemanan yang menyenangkan
15. Geng Finance al-fath 2013 yaitu kak Desty, Hening, Umi Nurjannah yang telah memberikan warna baru dalam kehidupan selama di kampus ini. Apapun yang terjadi sekali saudara tetaplah saudara.
16. IF3704 merupakan kelas paling gila yang pernah ada, kelas yang paling susah buat kompak, kelas yang punya hastag paling aneh "KM-LEMAH", kelas yang paling menjadi kenangan, kelas yang terdiri dari orang-orang hebat. terima kasih kepada semua teman-teman IF3704 yaitu Bunga dwipa, Atikha novesy, Desri apriliza, Indah mekar sari, Maya budi, Dwitika diah , Fifinella, Cincin, Rayi Widhayanti, Sarah andianti, Dinda, Maulina Istiana, Archie atrie, Moch Arif , Aria Octa, Fajar su-noto, Rahman, Andi, Rizky, William, Ananda Bayu, Faruq, Galih cita, Wahyu kurniawan, Hafidz ali, Al , Oktarifan Agung, M Agung, Ghass-an Amanullah, Ikhsan Ade, Ifendo, El rizky, Gerry noor, Fachri sinaga, Wiraka, Pranesha Wahyu Sinatria.
17. Geng Jawa ( Aini Rahmawati, Agustyna Dyah Sri Kamulyan, Aria Octa Cahyadi) matur nuwun kalian udah membuat aku jadi gila seperti kalian, terima kasih karna kalian selalu membuatku tersenyum.
18. seluruh saudaraku muslim didunia ini yang membuat saya membuka mata bahwa dunia tidak selebar daun kelor, tidak seindah surga, dan membuat saya patang menyerah untuk selalu mencari Ridho-Nya, selalu tegakkan keadilan didunia ini.

# Kata Pengantar

Bismillahirrahmanirrahim

Assalamu’alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah, puji dan syukur senantiasa penulis panjatkan kepada Allah SWT, serta shalawat dan salam penulis sampaikan kepada Nabi Muhammad SAW. karena dengan karunia serta nikmat Islam yang dibawakan Rasulullah SAW atas izin-Nya, penulis dapat menyelesaikan Tugas Akhir yang berjudul “Interpretable STS Terjemahan Ayat Alquran menggunakan *Chunking* otomatis " sebagai syarat kelulusan program studi Sarjana Teknik Informatika Universitas Telkom. Terima kasih kepada seluruh pihak yang telah membantu penulis selama proses pengerjaan Tugas Akhir. Begitu banyak kejadian yang penulis alami selama proses pengerjaan Tugas Akhir, dan kejadian itu mem-berikan hikmah yang luar biasa bagi penulis untuk selalu bersyukur dan selalu berusaha mendapatkan hasil yang terbaik walaupun hasilnyadiserahkan kepada Allah. karena penulis menyadari bahwa kesusahan yang dihadapi bisa teratasi apabila mau mengendalikan cara pandang hidup, berpikiran positif dan memaksimalkan kesungguhan dan ketekunan.

"Karena itu, ingatlah kamu kepada-Ku niscaya Akuingat kepadamu dan bersyukurlah kepada-Ku, dan janganlah kamu mengingkari-Ku."

(Q.S.Al-Baqarah:152)

Penulis menyadari bahwa pada penulisan Tugas Akhir ini tidak luput dari kesalahan. Oleh karena itu, penulis mengharapkan adanya masukan berupa kritik dan saran yang membangun bagi penulis agar mendapat perbaikan yang dapat menjadikan Tugas Akhir ini lebih baik. Penulis berharap Tugas Akhir ini dapat dikembangkan lebih lanjut dan pengembangan penelitian tentang teks Islami terutama teks Alquran dapat terus meningkat supaya dapat membawa manfaat untuk umat, agama, bangsa dan negara.

Wassalamu’alaikum Warahmatullahi Wabarakatuh.

# Daftar Isi

[Lembar Pernyataan ii](#_Toc471851144)

[Lembar Persetujuan iii](#_Toc471851145)

[Abstrak iv](#_Toc471851146)

[Abstrack v](#_Toc471851147)

[Lembar Persembahan vi](#_Toc471851148)

[Kata Pengantar ix](#_Toc471851149)

[Daftar Isi x](#_Toc471851150)

[Daftar Gambar xiii](#_Toc471851151)

[Daftar Tabel xiv](#_Toc471851152)

[Daftar Istilah xv](#_Toc471851153)

[Bab 1 Pendahuluan 1](#_Toc471851154)

[1.1. Latar Belakang 1](#_Toc471851155)

[1.2. Perumusan masalah 2](#_Toc471851156)

[1.3. Tujuan 2](#_Toc471851157)

[1.4. Batasan Masalah 2](#_Toc471851158)

[1.5. Metodologi Penelitian 3](#_Toc471851159)

[1.6. Sistematika Penulisan 5](#_Toc471851160)

[2. Kajian Pustaka 6](#_Toc471851161)

[2.1 *Interpretable Semantic Textual Similarity* (STS) 6](#_Toc471851162)

[2.1.1 *Chunking* 6](#_Toc471851163)

[2.1.2 Skor Kesamaan 6](#_Toc471851164)

[2.1.3 Label Keterhubungan 7](#_Toc471851165)

[2.2 Alquran 10](#_Toc471851166)

[2.3 Qursim 12](#_Toc471851167)

[2.4 *Text Preprocessing* 13](#_Toc471851168)

[2.4.1 *Tokenization* 13](#_Toc471851169)

[2.4.2  *Lemmatization* 14](#_Toc471851170)

[*2.5* *IXA pipes* 14](#_Toc471851171)

[2.6 Stanford Core NLP 15](#_Toc471851172)

[2.7 Algoritma Hungarian 16](#_Toc471851173)

[2.8 Perhitungan skor kesamaan 17](#_Toc471851174)

[2.9 Evaluasi 18](#_Toc471851175)

[2.9.1.Evaluasi *Chunk* 18](#_Toc471851176)

[2.9.2. Matriks Evaluasi 19](#_Toc471851177)

[Bab III Metodologi dan Desain Sistem 21](#_Toc471851178)

[3.1 Gambaran Umum Sistem 21](#_Toc471851179)

[3.2 Pengumpulan Data 24](#_Toc471851180)

[3.2.1. Data Pasangan Ayat 24](#_Toc471851181)

[3.2.2. Anotasi Data 25](#_Toc471851182)

[3.2.3. Gold Standard 25](#_Toc471851183)

[3.3 Pembangunan chunking otomatis 25](#_Toc471851184)

[3.4 Analisis Kebutuhan Fungsionalitas Sistem 26](#_Toc471851185)

[3.4.1 Kebutuhan Fungsional 26](#_Toc471851186)

[3.4.2 Spesifikasi Perangkat Keras 26](#_Toc471851187)

[3.4.3 Spesifikasi Sistem 26](#_Toc471851188)

[Bab IV Pengujian dan Analisis 27](#_Toc471851189)

[4.1 Set Data 27](#_Toc471851190)

[4.1.1 Data Pasangan Ayat 27](#_Toc471851191)

[4.1.2 Data Gold Standard 30](#_Toc471851192)

[4.2 Pengujian Sistem 31](#_Toc471851193)

[4.2.1 Tujuan Pengujian 31](#_Toc471851194)

[4.2.2. Skenario Pengujian 31](#_Toc471851195)

[4.3 Analisis 31](#_Toc471851196)

[4.3.1 Analisis ketepatan chunk 31](#_Toc471851197)

[4.3.2 Analisis perbandingan F1 32](#_Toc471851198)

[Bab V Kesimpulan 35](#_Toc471851199)

[5.1 Kesimpulan 35](#_Toc471851200)

[5.2 Saran 35](#_Toc471851201)

[Daftar Pustaka 36](#_Toc471851202)

[Lampiran 22](#_Toc471851203)

# Daftar Gambar

[Gambar 2 - 1 Q.S. Al-Baqarah:185 10](#_Toc472225690)

[Gambar 2 - 2 Q.S. AL-A'raf :52-53 11](#_Toc472225691)

[Gambar 4 Sistem *Chunking* 21](#_Toc472225692)

[Gambar 5 sistem interpretable STS 22](#_Toc472225693)

[Gambar 6 sistem keseluruhan 23](#_Toc472225694)

[Gambar 7 Grafik Akurasi chunk level 31](#_Toc472225695)

[Gambar 8 Grafik F1 Ali 31](#_Toc472225696)

[Gambar 9 Grafik F1 Type 32](#_Toc472225697)

[Gambar 10 Grafik F1 Score 32](#_Toc472225698)

[Gambar 11 Grafik F1 Type+Score 33](#_Toc472225699)

[Gambar 12 Grafik F1 secara keseluruhan 33](#_Toc472225700)

# Daftar Tabel

[Tabel 1 Matriks Evaluasi 20](#_Toc471188671)

[Tabel 2 Data Pasangan Ayat 24](#_Toc471188672)

[Tabel 3 Anotasi Data 25](#_Toc471188673)

[Tabel 4 -1 :Data pasangan chunk ayat 27](#_Toc471188674)

# Daftar Istilah

|  |  |  |
| --- | --- | --- |
| Alquran | : | Kitab suci umat Islam yang diturunkan oleh Tuhan.. karakteristik Alquran yaitu memiliki struktur yang terdiri dari ayat yang panjang dan ayat yang pendek dan biasanya ayat yang cenderung berulang di berbagai surat yang berbeda . Alquran mempunyai ayat-ayat yang memiliki kesamaan makna. |
| Ayat Alquran | : | Kata dalam bahasa Arab untuk tanda atau keajaiban yang tertulis dalam Alquran. |
| Gold Standard | : | Metode yang digunakan untuk menjadi pembanding hasil dari sistem. |
| Semantik | : | Ilmu yang berkaitan dengan makna kata dan kalimat. |
| Semantic Relatedness | : | Keterkaitan semantik antar teks. |
| Semantic Similarity | : | Kesamaan semantik antar teks. |
| Semantic Label | : | Label keterhubungan antar teks |
| Alignment | : | Sebuah metode penyelarasan kata/frasa yang memiliki keselarasan makna atau keterkaitan tertentu. |
| Feature Alignment | : | Karakteristik atau informasi yang relevan untuk memecahkan tugas komputasi yang terkait dengan pengaplikasian alignment. |
| Interpretable STS | : | Metode untuk mengetahui tingkat kesamaan makna antar teks dengan melihat semantic labels dan semantic similarity berdasarkan alignment. |
| Semantic textual simillarity | : | Metode yang digunakan untuk mengukur kesamaan teks. |
| Interpretable | : | Istilah kata yang berartikan penafsiran |
| Chunking | : | Istilah kata yang berartikan pemotongan kata. Chunk merupakan potongan dari satu kata atau lebih atau konstituen yang berarti gabungan dari satu kata atau lebih. |

# Pendahuluan

## Latar Belakang

Umat Islam memiliki kitab kepercayaan untuk panduan hidup mereka, kitab tersebut bernama Alquran. Alquran sendiri terdiri dari teks yang biasa disebut dengan ayat. Kitab Alquran terdiri dari ayat-ayat yang tersebar dalam 114 surat [1]. Alquran memiliki banyak ayat yang berulang dalam berbagai surat yang masih memiliki keterhubungan atau topik yang saling berkaitan. Hal ini menyebabkan kesulitan bagi orang awam untuk mencari ayat yang memiliki informasi yang saling berkait dalam Alquran. Oleh karena itu,dilakukan implementasi konsep *Interpretable* STS pada terjemahaan ayat Alquran baik itu ayat-ayat yang memiliki kesamaan baik dari segi tulisan maupun dari segi topik yang dibahas. Hal ini bertujuan untuk melakukan pengukuran kesamaan makna dengan cara menentukan label keterhubungan dan skor kesamaan berdasarkan proses *alignment*nya.

Untuk pengukuran kesamaan antar ayat yang dilakukan pada penelitian ini berpanduan pada *SemEval2016 Task2*, hal ini dikarenakan *SemEval2016* *Task2* merupakan salah satu sarana penelitian mengenai pengukuran kesamaan semantik yang telah menjadi acuan para peneliti dibidang linguistik. *SemEval2016 Task2* memiliki kelebihan daripada *SemEval2015* yaitu untuk pengukuran kesamaan pada *SemEval2015* hanya dibatasi 1:1, artinya setiap potongan hanya dapat disejajarkan dengan satu potongan, jika ada dua pilihan untuk menyelaraskan, potongan yang terkuat yang terpilih untuk disejajarkan[5]. Sedangkan untuk pengukuran pada *SemEval2016 Task2* setiap potongan ayat bisa disejajarkan lebih dari satu potongan. Oleh karena itu, pengukuran dalam penelitian ini berpedoman terhadap *SemEval2016 Task2*, untuk proses pengukuran kesamaan makna yang berdasarkan pada *chunk* dan *alignment*. Sehingga dalam penelitian tugas akhir ini menggunakan proses *Chunking* otomatis. Hal ini dilakukan karena metode *Chunking* otomatis ini adalah metode yang sesuai dengan kebutuhan penelitian ini dan pada *SemEval2016*  metode ini memiliki tingkat performansi yang cukup tinggi dan mendapatkan peringkat pertama untuk kategori *Interpretable* STS [[2]](#footnote-2). Pembuatan Tugas Akhir ini diharapkan dapat membantu dalam mengetahui tingkat kesamaan makna antar potongan ayat Alquran dengan melihat nilai kesamaannya serta juga label keterhubungannya.

## Perumusan masalah

Berdasarkan latar belakang masalah tersebut, maka permasalahan yang dapat dirumuskan sebagai berikut :

1. Bagaimana tahapan proses yang dilakukan untuk menentukan kesamaan makna antar potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis?
2. Bagaimana hasil analisis terhadap pengukuran kesamaan makna terhadap potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis?
3. Bagaimana evaluasi dari pengukuran kesamaan makna antar kedua potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis?

## Tujuan

Berdasarkan rumusan masalah tersebut, maka tujuan yang dicapai adalah sebagai berikut :

1. Mengetahui tahapan secara lengkap proses untuk menentukan kesamaan makna antar potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis .
2. Memahami bagaimana hasil analisis terhadap pengukuran kesamaan makna terhadap potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis.
3. Menganalisis evaluasi pengukuran kesamaan makna antar kedua potongan terjemahan ayat Alquran dalam bahasa Inggris dengan metode *Chunking* otomatis.

## Batasan Masalah

Berdasarkan tujuan yang ingin dicapai, maka batasan permasalahan dalam penellitian ini adalah sebagai berikut

1. Pengukuran kesamaan makna hanya digunakan untuk teks yang relatif pendek, hal ini berdasarkan hasil penelitian yang ada di *SemEval2016*  untuk melakukan perhitungan nilai kesamaan baru bisa digunakan untuk teks yang relatif pendek .
2. Penelitian ini hanya untuk mengetahui tingkat kesamaan makna dengan menentukan jenis label dan skor kesamaan dari setiap *alignment* potongan ayat Alquran terjemahan bahasa Inggris.
3. Terjemahan ayat Alquran yang digunakan yaitu terjemahan dari Saheeh International terjemahan bahasa inggris yang diambil dari quranindex.info
4. Sistem yang dibangun dibagi menjadi dua sistem yaitu sistem untuk melakukan proses pemotongan (*chunking)* secara otomatis dan sistem untuk mengukur tingkat kesamaan makna.

## Metodologi Penelitian

Metodologi yang digunakan dalam melakukan penelitian ini adalah sebagai berikut:

1.Identikasi Permasalahan

Mengidentifikasi suatu permasalahan merupakan salah satu metode terpenting yang harus ada dalam suatu penelitian. Hal ini dikarenakan untuk membuat suatu penelitian, haruslah ada masalah yang ingin di teliti. Dalam penelitian ini, metode yang pertama kali di lakukan adalah mengidentifikasi permalasahan, pengidentifikasikan masalah yang diangkat dari salah satu task yang ada di *SemEval2016* mengenai *Interpretable STS (Semantic Textual Similarity)*.

2. Identifikasi Metode Penyelesaian Masalah

Tahapan selanjutnya yaitu mengidentifikasi metode yang digunakan dalam menyelesaikan permasalahan dalam penelitian yang dilakukan ini. Dalam pemilihan metode haruslah sesuai dengan permasalahan yang dimiliki. Oleh karena itu, dalam pemilihan metode penyelesaian permasalahan harus membanding-bandingan terlebih dahulu antara metode penyelesaian yang ada. Setelah melakukan analisis didapatkan metode yang tepat untuk penelitian ini yaitu menggunakan *Chunking* otomatis. Metode ini diterapkan karena, pada *SemEval2016 Task 2* metode *Chunking* otomatis mendapatkan peringkat tertinggi.

3. Pengumpulan Data

Tahapan ini digunakan untuk mengumpulkan data pasangan ayat Alquran untuk penelitian ini. Ada tiga tahapan yang dilakukan untuk pembangunan data yaitu pengumpulan pasangan ayat Alquran dan pembuatan gold standard untuk *chunk* pasangan ayat Alquran dan untuk data yang sudah diberi alignment, label keterhubungan beserta skor kesamaan.

* 1. Pengumpulan pasangan ayat : Dilakukan pengumpulan data pasangan ayat Alquran secara manual
  2. Pembuatan gold standard

Pembuatan gold standard untuk *chunk* pasangan ayat Alquran dilakukan secara manual yaitu memotong ayat secara manual. untuk pembuatan gold standard untuk sistem *Interpretable* STS juga dilakukan secara manual dengan melakukan proses *alignment,* menentukan label keterhubungan dan mendapatkan nilai kesamaan antardua potongan ayat yang digunakan dalam mengevaluasi sistem yang dibangun. Kesamaan antara pasangan ayat diberikan nilai dari skala 0 hingga 5 oleh anotator.

4. Pemodelan Sistem

Tahapan yang digunakan untuk memodelkan sistem yang dibangun, sehingga dapat dengan mudah merepresentasikan cara kerja sistem yang dibuat kedepannya.

5. Implemetasi

Tahapan yang digunakan untuk mengimplementasi pengoperasian suatu model sistem yang telah dibuat, secara sederhana tahapan ini berarti *coding* atau membuat aplikasi sistem yang telah dirancang.

6. Pengujian dan analisis

Tahapan yang digunakan untuk menguji aplikasi sistem yang telah dibuat apakah sudah bisa digunakan atau belum. Dalam tahapan ini dapat menganalisis apakah aplikasi sistem masih mengandung *bug* atau tidak.

7. Pembuatan Laporan

Tahapan ini merupakan yang terakhir, tujuan tahapan ini yaitu digunakan untuk mendokumentasikan hasil dari penelitian yang dikerjakan dalam bentuk laporan.

## Sistematika Penulisan

Sistematika penulisan dari Tugas Akhir ***“ Interpretable STS*  terjemahan ayat Alquran "** adalah sebagai berikut.

* Pendahuluan

Bab 1 menjelaskan tentang gambaran umum dari penelitian ini seperti latar belakang masalah yang menjadi dasar penelitian, identifikasi terhadap rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian serta sistematika penulisan pada penelitian yang dilakukan.

* Kajian Pustaka

Bab 2 memaparkan mengenai dasar teori pendukung dan penelitian-penelitian terkait yang telah dilakukan sebelumnya yangdigunakan sebagai landasan dalam pengembangan sistem.

* Perancangan Sistem

Bab 3 memaparkan mengenai perancangan sistem yangdibangun dan alur proses kerja dari sistem seperti proses pembangunan data, pemotongan data secara manual, pembuatan gold standard, proses chun-king otomatis deskripsi proses evaluasi yangdipakai pada sistem. Perancangan sistem disusun berdasarkan dasar teori yang telah diuraik-an pada Bab 2. Perancangan sistem mende nisikan setiap proses dari penelitian ini.

* Pengujian dan Analisis

Bab 4 menjelaskan tujuan dari pengujian sistem, skenaario pengujian sistem dan hasil dari pengujian sistem yang dilakukan.

* Kesimpulan dan Saran

Bab 5 menjelaskan mengenai kesimpulan dari hasil analisis penelitan yang dilakukan, saran yang terkait untuk pengembangan penelitian ke-depannya.

# Kajian Pustaka

Pada bab ini, akan dipaparkan mengenai

## 2.1 *Interpretable Semantic Textual Similarity* (STS)

*Interpretable Semantic Textual Similarity (ISTS)* yaitu salah satu *task* yang terdapat di dalam *SemEval2015* dan *SemEval 2016*. *Interpretable STS* merupakan salah satu metode STS untuk proses pengukuran kesamaan makna antara kalimat dengan menggunakan potongan-potongan (*Chunk)[[3]](#footnote-3)* dan pemberian label keterhubungan serta skor kesamaan berdasarkan *alignment*. Contoh penerapan ISTS pada Alquran sebagai berikut.

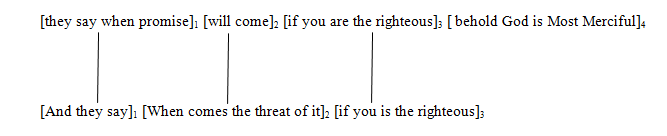


Figure ‑ contoh implementasi ists

### *Semantic Textual Similarity*

*Semantic Textual Similarity* adalah metode yang digunakan untuk proses pergukuran kesamaan semantik dengan cara membandingkan antara dua buah teks yang bisa berupa kata, kalimat pendek maupun dokumen. Semantik adalah suatu ilmu bahasa yang mempelajari mengenai interpretasi kata, makna yang dapat merepresentasikan pemahaman pembaca dalam menafsirkan suatu kalimat, karena setiap pembaca dapat menafsirkan suatu kalimat secara berbeda-beda. Hasil dari proses pengukuran kesamaan semantik merupakan salah satu faktor yang digunakan dalam penelitian *natural language processing*. Penelitian ini adalah penelitian dasar yang dikembangkan untuk proses penelitian yang lain seperti *Text Summarization, Question Answering, Short Answering Scoring, word sense disambiguation, information retrival*.

### *Chunking*

*Chunk* yang dalam bahasa Indonesia berarti potongan. *Chunk*  dapat diartikan sebagai bagian dari konstituen atau gabungan dari berbagai kata. Karena *chunk* terdiri dari satu atau lebih dari satu kata. Untuk menentukan *chunk* pada semEval 2016 dapat mengikuti panduan pada CONLL 2000[[4]](#footnote-4). Berikut cara menentukan potongan berdasarkan panduan CO-NLL 2000.

1. Membagi klausa dan bawahan klausa utama dalam menjadi potongan yang lebih kecil seperti NP, verb chain,PP, adverbs, dan expressions.
2. Mengambil PP’s secara keseluruhan pasangan

Contoh *chunking*  seperti berikut.

* NP [The boy] / [Maulisye Audina and Muhammad]
* verb chain [is chatting] / [does not like]
* PP [at a time] / [with the computer] / [the house] [of that man] adverbs [of course]
* expressions [once upon a time] / [by the way][[5]](#footnote-5)

### Skor Kesamaan

Untuk menentukan skor kesamaan dan keterkaitan pada penelitian ini harus dilakukan sebelum memberikan label pada pasangan ayat. Nilai kesamaan diberikan untuk setiap *alignment*. Menurut sumber dari *detail task description* *SemEval2016*, untuk menentukan nilai kesamaaan, yang dilakukan terlebih dahulu adalah mengidentifikasi setiap potongan ayat. Kemudian tentukan parameter untuk perhitungan keselarasaan. Kategori nilai skor kesamaan dan keterkaitan antar potongan yaitu sebagai berikut.

Skor kesamaan memiliki *range* dari 5 sampai 0, dimana angka lima merupakan skor maksimal sedangkan nol merupakan skor minimal.

* [5] jika makna dari kedua potongan ayat itu sama atau selaras

Contohnya : [unbelievers] ⬄[an unbelievers]

* [4,3] jika makna dari kedua potongan ayat itu sangat mirip atau terikat

Contohnya : [will doom ] ⬄[ will torment]

* [2,1] jika makna dari kedua potongan ayat itu sedikit mirip atau sedikit terkait

Contohnya : [on the house] ⬄[ on horse]

* 0 yang direpresentasikan NIL makna dari kedua potongan ayat itu sama sekali tidak mirip atau sama sekali tidak berhubungan.

### Label Keterhubungan

Label keterhubungan (*alignment type*) bertujuan untuk menentukan tipe dari *alignment*  potongan kalimat, apakah potongan tersebut mengandung tipe yang seperti apa. Berikut label yang digunakan untuk kesamaan antar potongan.

* EQUI : kedua potongan memiliki arti yang sama, dan memiliki nilai semantik yang setara. Contohnya : Hospital dengan At Hospital.
* OPPO : arti kedua potongan saling bertentangan satu sama lain. Contohnya : in southern indonesia dengan in northern indonesia.
* SPE1 : antar potongan memiliki arti yang sama, tetapi potongan di kalimat 1 lebih spesifik daripada potongan di kalimat lain. Contohnya : by a shop dengan outside a bookstore.
* SPE2 : seperti SPE1, tetapi yang berbeda adalah potongan dalam kalimat 2 yang lebih spesifik. Contohnya holy book dengan Koran.
* SIMI : antar potongan hanya memiliki arti yang sama, namun tidak EQUI, OPPO, hubungan SPE1 ataupun SPE2. Contohnya Palestina dengan Indonesia.
* REL : antar potongan tidak dianggap sama , namun saling terkait erat dengan beberapa hubungan ( tidak ada hubungan EQUI, OPPO, SPE1, SPE2, atau SIMI).
* NOALI : potongan ini tidak memiliki potongan yang sesuai dalam kalimat lainnya [3].

Selain menggunakan tujuh label tersebut, dapat juga menggunakan dua label berikut.

* FAKTA: faktualitas dipotongan (yaitu apakah potongan merupakan suatu pernyataan, fakta atau hanya spekulasi).
* POL: polaritas dalam potongan (yaitu pendapat yang dikemukakan, pendapat bisa berupa pendapat positif,negatif, ataupun netral).

Contoh ilustrasi yang bersumber dari *SemEval2016*, Interpretable STS *Annotation Guidelines* yaitu sebagai berikut.

[12]1 [killed ]2  [in bus accident ]3 [in Pakistan]4

[10]1 [killed ]2  [in road accident ]3 [in NW Pakistan]4

Pemberian urutan label dilakukan secara manual oleh manusia.

2 ⇔ 2 (EQUI **5**),

3 ⇔ 3 (SPE1 **4**),

4 ⇔ 4 (SPE2 **4**),

1 ⇔ 1 (SIMI **4**)

Dari pemberian urutan label, maka penyelarasaan potongan menjadi seperti berikut ini.

1 ⇔ 1 (SIMI **4**), 2 ⇔ 2 (EQUI **5**), 3 ⇔ 3 (SPE1 **4**), 4 ⇔ 4 (SPE2 **4**)

Keterangan : untuk angka berwarna *bold* menjelaskan skor kesamaan yang di miliki oleh potongan. Adapun beberapa kondisi dalam menentukan label dalam potongan ayat adalah sebagai berikut.

1. Kesejajaran antara kedua potongan ayat memiliki arti yang sama dan saling terkait dalam satu konteks pembahasan

[2 car bombs]1 [kill ]2  [8]3 [in southern iraq]4

[car bombing]1 [kills ]2  [14]3 [in northern iraq]4

Sehingga , penyelarasan dalam potongan menjadi seperti berikut.

1 ⇔ 1 (SPE1 **4**), 2 ⇔ 2 (EQUI **5**), 3 ⇔ 3 (SIMI **3**), 4 ⇔ 4 (OPPO **4**)

1. Dibeberapa kondisi dalam sebuah kasus, perlu memahami peristiwa yang ada untuk mendeskripsikan kalimat dan mengetahui peranan yang dimainkan oleh potongan yang disejajarkan. Biasanya potongan yang selarasmemainkan peran yang sama

[Gunmen]1 [abduct ]2  [seven foreign workers]3 [in Nigeria]4

[Seven foreign workers ]1 [kidnapped ]2  [in Nigeria]3

Sehingga , penyelarasan dalam potongan menjadi seperti berikut ini.

1 ⇔ 1 (SPE1 **4**), 2 ⇔ 2 (EQUI **5**), 3 ⇔ 3 (SIMI **3**), 4 ⇔ 4 (OPPO **4**)

* Kondisi ketika potongan memainkan peran yang berbeda , namun saling terkait.

[Hundreds]1 [of Bangladesh clothes factory workers ]2  [ill]3

[Hundreds]1 [fall ]2  [sick]3 [in Bangladesh factory]4

penyelarasan dalam potongan menjadi seperti berikut.

1 ⇔ 1 (EQUI **5**), 2 ⇔ 4 (SPE1 **3**), 3 ⇔ 2,3 (EQUI **5**)

* Kondisi ketika kalimat merujuk pada peristiwa yang berbeda, maka potongan dapat di sejajarkan bahkan jika memiliki peran yang berbeda.

[Saudis]1 [to permit ]2  [women]3 [to complete]4 [in Olympics]5

[women]1 [are confronting ]2  [a glass ceiling]3

penyelarasan dalam potongan menjadi seperti berikut.

1 ⇔ ϴ (NOALI), 2 ⇔ϴ (NOALI), 3 ⇔ 1 (SPE1 **4**), 4 ⇔ ϴ (NOALI), 5⇔ ϴ (NOALI), ϴ ⇔ 3 (NOALI)

Untuk penelitian ini hanya digunakan tujuh label keterhubungan yaitu SIMI, EQUI, OPPO, SPE1, SPE2, NOALI, REL

## Alquran

Alquran merupakan kitab suci umat islam yang di bawa oleh seorang nabi yaitu nabi Muhammad SAW. Kitab suci Alquran dipercaya oleh umat muslim sebagai kabar gembira yang diberikan oleh Allah, dan umat muslim percaya apabila umat islam mengerjakan kebaikan maka umat islam mendapatkan pahala yang diberikan oleh Allah. Seperti dalam kutipan surat Al-Isra’ ayat 9 yang berbunyi : “Sungguh Alquran ini memberi petunjuk ke jalan yang paling lurus dan memberi kabar gembira kepada orang mukmin yang mengerjakan kebajikan, bahwa mereka mendapatkan pahala yang besar”. Dan ketika umat islam membaca Alquran mereka percaya, merekaterhindar dari godaan syetan, seperti yan sudah dijelaskan di Q.S.An-Nahl:98 yan berbunyi :” Dan jika kamumembaca Alquran, maka mintalah perlindungan kepada Allah dari (godaan-godaan) syetan yang terkutuk.”

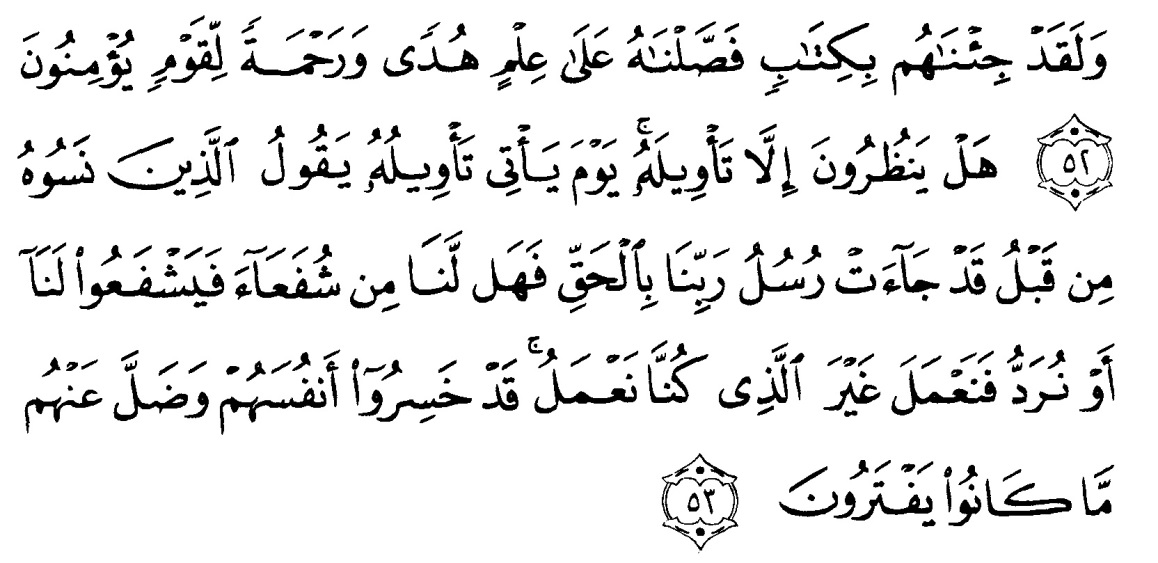


Figure ‑ Q.S. AL-A'raf :52-53

Artinya : “ Dan sesungguhnya Kami telah mendatangkan sebuah Kitab (al-Koran) kepada mereka, yang Kami telah menjelaskannya atas dasar penge-tahuan Kami; menjadi petunjuk dan rahmat bagi orang orang yang beriman. (QS. 7:52) Tiadalah mereka menunggu-nunggu, kecuali (terlaksananya kebe-naran) Alquran itu. Pada hari datangnya kebenaran pemberitaan al-Koran itu, berkatalah orang-orang yang melupakannya sebelum itu: `Sesungguhnya telah datang Rasul-Rasul Rabb kami membawa yang haq, maka adakah bagi kami pemberi syafa'at bagi kami, atau dapatkah kami dikembalikan (ke du-nia) sehingga kami dapat beramal yang lain dari yang pernah kami amal-kan?' Sesungguhnya mereka telah merugikan diri sendiri dan telah lenyaplah dari mereka apa yang mereka ada-adakan[1](QS. 7:53).

Dari berbagai penjelasan diatas dapat diketahui bahwa Alquran adalah ki-tab suci umat muslim yang digunakan untuk pedoman hidup, kitab ini di bawa oleh nabi Muhammad SAW yang diturunkan oleh malaikat jibril, dan mereka percaya yang mempelajarinya mendapatkan pahala dan syafa'at pada hari akhir. Proses penurunan Alquran berlangsung selama 22 tahun 2 bulan 22 hari, selama proses penurunan Alquran dibagi menjadi 2 periode yaitu periode penurunan Alquran yang berlangsung di kota Mekkah sehingga surat yang diturunkan disebut dengan surat Makkiyah dan periode penurunan Alquran di Madinah sehingga suratnya disebut dengan surat Madaniyah.

Pada masa nabi Muhammad SAW, Alquran tidak terdokumentasikan dalam bentuk teks melainkan hanya dalam bentuk lisan. Namun, setelah zaman khalifah Utsman bin Affan mulailah terjadi proses pembukuan Alquran. Alquran diturunkan menggunakan bahasa Arab namun seiring perkembangan zaman Alquran ditransformasi kedalam berbagai bahasa terjemahan supaya memudahkan umat manusia untuk memahami isi kandungan Alquran. Dalam penelitian Tugas Akhir ini, penyusun menggunakan Alquran sebagai bahan penelitian dikarenakan, karakteristik yang di miliki oleh Alquran yaitu.

1. Pada dasarnya Alquran terbentuk dari ayat-ayat, dimana ayat-ayat ini terdiri dari ayat-ayat pendek dan ayat-ayat sebagai kalimat pendek.
2. Penjelasaan topik yang ada dalam Alquran biasanya tidak dibahas dalam lebih dari satu ayat saja , namun dibahas di berbagai ayat yang tersebar dalam surat yang berbeda.
3. Alquran cenderung memiliki ayat-ayat yang sama karena pada Alquran ayat cenderung berulang di berbagai surat yang berbeda .
4. Alquran mempunyai ayat-ayat yang memiliki kesamaan makna.

## Qursim[13]

Salah satu penelitian dalam bidang quran mining adalah Qursim. Qursim yaitu corpus untuk melakukan evaluasi kesamaan dan keterkaitan teks pendek pada Alquran. Adapun data yang digunakan pada penelitian ini adalah data ayat Alquran hasil eksplorasi dari tafsir Ibn Katsir. Tafsir Ibn Katsir merupakan salah satu buku tafsir yang sangat popular dan banyak dijadikan referensi tafsir Alquran. Untuk mendapatkan data pasangan ayat dilakukan dengan cara melakukan ekstraksi pada ayat secara otomatis. Selanjutnya dilakukan pengecekan secara manual untuk menghilangkan sebagian data yang tidak konsisten sehingga proses ini menghasilkan 7.679 pasangan ayat Alqur-an. Kemudian dilakukan pengecekan kembali karena dalam tafsir Ibn Katsir pembahasan suatu ayat mengarahkan pada pembahasan yang sangat luas. Setelah dilakukan pengecekan kembali secara manual dihasilkan 6.796 pasangan ayat yang terbagi dengan tiga level berbeda yaitu level 0, level 1, level2 [13].

1. level 0

Pasangan ayat memiliki keterhubungan yang sangat rendah dan sulit un-tuk dipahami,sehingga untuk memahami keterhubungannya diharuskan melihat dari konteks dalam buku tafsir tersebut. Berikut contoh data level 0 seperti berikut.

**Ayat 1** : They request from you a [legal] ruling. Say, "Allah gives you a ruling concerning one having neither descendants nor ascendants [as heirs]." If a man dies, leaving no child but [only] a sister, she will have half of what he left. And he inherits from her if she [dies and] has no child. But if there are two sisters [or more], they will have two-thirds of what he left. If there are both brothers and sisters, the male will have the share of two females. Allah makes clear to you [His law], lest you go astray. And Allah is Knowing of all things (Q.S.An-Nisa :176)

**Ayat 2** : And do not invoke with Allah another deity. There is no deity except Him. Everything will be destroyed except His Face. His is the judgement, and to Him you will be returned (Q.S.Al-Qasas :88).

1. Level 1

Pasangan ayat ini memiliki keterkaitan rendah namun dapat di-pahami oleh manusia, tapi masih mungkin sulit untuk digunakan sebagai data training pada algoritma pembelajaran mesin. Berikut contoh data level 1.

**Ayat 1** : In the name of Allah, the Entirely Merciful, the Especially Merciful.(Q.S.Al-Baqarah :1)

**Ayat 2** : So exalt the name of your Lord, the Most Great (Q.S.Al-Waqiah:74).

1. Level 2

Pasangan ayat ini memiliki keterkaitan yang tinggi dan sangat mudah dipahami oleh pembaca. Data ini cocok untuk digunakan sebagai data training pada algoritma pembelajaran mesin. Berikut contoh data level 2.

**Ayat 1** : This is the Book about which there is no doubt, a guidance for those conscious of Allah (Q.S.Al-Baqarah :7)

**Ayat 2** :O mankind, there has to come to you instruction from your Lord and healing for what is in the chest and guidance and mercy for the believers (Q.S.Yunus:57).

## *Text Preprocessing*

Preprocessing merupakan tahapan awalan untuk memproses suatu teks, supaya teks dapat lebih cocok atau lebih siap untuk digunakan dalam pemrosesan tahap lanjut. Dengan adanya preprocessing dapat meningkatkan analisis terhadap waktu dan kualitas suatu teks untuk diolah. Tahap preprocessing dilakukan karena data teks yang diolah biasanya masih banyak yang belum sesuai dengan parameter untuk proses analisis. Untuk melakukan proses preprocessing untuk data dilakukan dengan beberapa tahapan seperti *Tokenization* dan  *Lemmatization.*

### *Tokenization*

Tokenisasi merupakan teknik untuk memecah beberapa karakter dalam suatu teks ke dalam suatu kata. Untuk melakukan pemecahan dapat menggunakan karakter pemisah kata seperti semikolon (;), titik(.) ataupun spasi ( ). Contoh dari tokenisasi yaitu sebagai berikut.

Input : aku anak Indonesia cinta kebudayaan Indonesia

Output tokenisasi, terdapat enam token yaitu [aku] [anak] [Indonesia] [cinta] [kebudayaan][Indonesia]

### 2.4.2 *Lemmatization*

*Lemmatization* merupakan suatu metode yang digunakan mendapatkan bentuk dasar dari transformasi suatu kata. Proses *lemmatization* hampir sama dengan *stemming,* namun untuk *stemming* hanya mengubah kata ke bentuk suffix katanya. Adapun contoh dari suatu *lemmatization* yaitu sebagai berikut.

Toy, toys, toy’s => toy

Was, were => be

## *IXA pipes*

*IXA pipes* saat ini menyediakan anotasi linguistik sebagai berikut : *Sentence segmentation, tokenization, Part of Speech (POS) tagging, lemmatization, Named Entity Recognition and Classification (NERC), constituent parsing and coreference resolution*. Setiap modul yang dimiliki bekerja dalam bahasa Inggris dan Spanyol. *IXA pipes* adalah satu set modul dalam tools NLP Natural Language Processing yang menyediakan akses yang mudah dalam teknologi NLP untuk beberapa bahasa. *IXA pipes* dibuat dengan tujuan untuk mengefisiensikan dan menguatkan penjelasan mengenai linguistik untuk peneliti dan ahli yang tidak menguasai NLP. Hal ini dimaksudkan untuk mempermudah peneliti dapat memahami mengenai teknologi NLP. Sehingga hambatan dalam penggunaan teknologi NLP menjadi berkurang. *IXA pipes* dikembangkan oleh grup IXA NLP dari University of the Basque Country. untuk lebih mengerti lebih spesik mengenai IXA pipes, berikut kelebihan yang dimiliki oleh ixa pipes[[6]](#footnote-6).

1. Simple and ready to use: sangat sederhana dan mudah untuk digunakan
2. Portable: sangat portable sehingga tidak diperlukan untuk melakukan sistem kon gurasi yang rumit. modul ini dapat dijalankan pada platform apapun selama JVM 1.7+ , Python 2.7 tersedia.
3. Modular:tidak seperti toolkit NLP lainnya, yang seringkali dibangun dalam arsitektur monolitik, *IXA pipes* dibangun dalam arsitektur data centric sehingga modul dapat diubah. Modul berperilaku seperti Unix pipes, tools mengambil input standard, melakukan beberapa proses an-notation, dan menghasilkan keluaran yangmenjadi inputan untuk modul selanjutnya. Arsitektur data-centric dari *IXA pipes* membuat seti-ap modul independen dan karenanya dapat dikolaborasi penggunaannya dengan toolkit lainnya.
4. Efisien: Proses yang digunakan memiliki tingkat kecepatan yang tinggi
5. Multilingual: NLP annotation untuk saat ini sudah dapat digunakan untuk bahasa Inggris dan Spanyol.
6. Accurate: tingkat akurasi untuk ixa pipes tinggi

## Stanford Core NLP

Stanford core NLP toolkit adalah tools yang digunakan untuk membantu pemprosesan dalam Natural Language Processing. Stanford CoreNLP menye-diakan satu set alat analisis bahasa alami. Hal ini dapat memberikan bentuk dasar kata-kata, part of speech, apakah mereka adalah nama-nama perusahaan (Named Entity Recognition), menormalkan tanggal, waktu, dan jumlah numerik, mark up struktur kalimat dalam hal frase dan kata dependensi, meng-indikasikan yang nomina frasa merujuk pada entitas yang sama, menunjukkan sentimen, ekstrak hubungan tertentu atau membuka hubungan kelas antara entitas[[7]](#footnote-7) . Kelebihan dari tools ini adalah sebagai berikut :

* Sebuah toolkit yang terintegrasi dengan rentang yang baik dari alat analisis gramatikal
* Cepat dalam melakukan analisis teks
* Secara keseluruhan analisis teks yang dilakukan memiliki kualitas tinggi Tersedia dalam berbagai bahasa manusia yang berbeda-beda
* Interface tersedia untuk sebagian besar bahasa pemrograman utama yang modern
* Memiliki kemampuan untuk menjalankan sebagai layanan web sederhana

Tujuan Stanford CoreNLP adalah untuk membuat alat analisis linguistik sangat mudah untuk digunakan dalam sepotong teks. CoreNLP dirancang untuk menjadi sangat eksibel dan extensible. Stanford CoreNLP mengintegrasikan banyak alat NLP seperti *part-of-speech* (POS) *tagger, the named entity recognizer (NER), parser, the coreference resolution system, sentiment analysis, bootstrapped pattern learning, and the open information extraction tools*. Analisis CoreNLP ini menyediakan fondasi untuk tingkat yang lebih tinggi dan pemahaman teks aplikasi domain lebih spesik.

Pada penelitian ini yangdigunakan yaitu stanford parser dan stanford NER. Stanford parser bekerja diluar struktur gramatikal kalimat, stanford parser ini merupakan implementasi dari  *probabilistic natural languge parsers* baik PCFG parser lexicalized maupun lexicalized dependency parsers. Inputan untuk stanford parser yaitu berupa teks sedangkan keluarannya berupa pohon struktur kalimat[[8]](#footnote-8). Berikut contoh inputan dan outputan dari stanford parser .

Masukan : Hello my name is lee and now i reading quran

Keluaran :

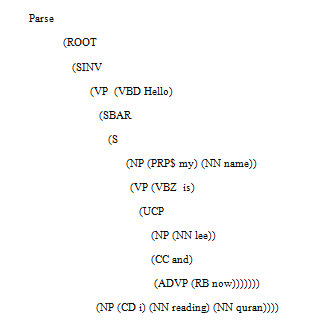


Figure ‑ contoh proses stanford parser

## Algoritma Hungarian

Algoritma Hungarian ini digunakan untuk menyelesaikan permasalahan yang menyangkut assignment. Metode Hungarian adalah algoritma optimasi kombinatorial yang memecahkan masalah. Algoritma ini dikembangkan dan diterbitkan pada tahun 1955 oleh Harold Kuhn, yang memberi nama "metode Hungarian", pemberian nama ini merujuk dari karya-karya sebelumnya dari dua ahli matematika Hungarian yaitu Denes Konig dan Jeno Egerváry. Oleh karena itu dinamakan metode Hungarian. Selanjutnya pada 1957 ada peneliti yang bernama James Munkres yang memodifikasi algoritma hungarian menjadi lebih baik. Sehingga nama algoritma hungarian sekarang ini lebih dikenal dengan nama algoritma Kuhn-Munkres [[9]](#footnote-9). Prosedur penggunaan metode Kuhn-Munkres dalam matriks yaitu sebagai berikut :

1. Untuk setiap baris dalam matriks, cari elemen yang memiliki nilai paling minimum. Lakukan proses pengurangan untuk semua elemen pada baris tersebut, pengurangan dilakukan dengan nilai yang minimum.
2. Untuk setiap kolom dalam matriks, cari elemen yang memiliki nilai paling minimum. Lakukan proses pengurangan untuk semua elemen pada baris tersebut, pengurangan dilakukan dengan nilai yang minimum.
3. Buat garis yang mencakup setiap elemen yang bernilai nol.

## Perhitungan skor kesamaan

Untuk proses perhitungan skor kesamaan pada penelitian ini menggunakan *rules* dari penelitian nerosim[12]. Berikut *rules* yang digunakan.

1. EQUI *Rules*

* Jika kedua chunks memiliki token yang samadiberi skor kesamaan yaitu 5, misalnya to compete dengan To Compete
* Jika kedua chunks memiliki kontents kata yang samadiberi skor kesamaan yaitu 5, misalnya *in Olympics* dengan *At Olympics*
* Jika kedua chunks memiliki konten yang sama berdasarkan sinonimnya diberi skor kesamaan yaitu 5, misalnya *to permit* dengan *Allowed*.

1. OPPO *Rules*

* Jika kedua chunks memiliki konten berupa antonim atau lawan kata,diberi skor kesamaan yaitu 4, misalnya *in southern palestina* dengan *in northern palestina.*

1. SPE *Rules*

* Untuk SPE1 yaitu jika chunk pertama merupakan hiponim dari chunk kedua. Hiponim adalah kata-kata khusus yang biasa di representasikan dengan kata yang lain. Untuk kondisi ini diberi skor kesamaan yaitu 4, misalnya *by a shop* dengan *outside a bookstore*.
* Untuk SPE2 yaitu jika chunk pertama merupakan hipernim dari chunk kedua. Hipernim adalah kata-kata yang mewakili kata yang lain. Untuk kondisi ini diberi skor kesamaan yaitu 4, misalnya *holy book* dengan *Koran*.

1. SIMI Rules

* Jika kedua chunks memiliki token dengan *Named Entity Recognition*  yang sama seperti *location type, date typr, time type* yang sama. Untuk kondisi inidiberi skor kesamaan yaitu 3, misalnya *Palestina* dengan *Indonesia*.

1. REL Rules

* Jika kedua chunks memiliki keterhubungan satu sama lain, untuk mengetahuinya menggunakan *Mikolov*, namun pada penelitian ini metode *Mikolov* diganti dengan *path similarity* pada *Wordnet*. Jika *path*  pendek [0.5 sampai 1.0] maka skornya bernilai 4, *path* [0.4 sampai 0.5] maka skor bernilai 3.

1. NOALI Rules

* Jika salah chunks tidak memiliki pasangan yang selaras ( *not aligned).* Untuk kondisi ini diberi skor kesamaan yaitu 0 tapi dalam penelitian ini nilai 0 direpresentasikan dengan NIL, misalnya *holy book* dengan  *-* maka skor bernilai NIL

## Evaluasi

Evaluasi adalah metode yang digunakan untuk mengetahui seberapa baik sistem yang telah dibuat. Pada penelitian ini untuk evaluasi dibedakan menjadi dua yaitu evaluasi *chunk* dan matriks evaluasi. Evaluasi *chunk* digunakan untuk membandingkan *chunk* oleh sistem dengan *chunk*  manual, sehingga pada evaluasi *chunk* hanya dihitung nilai akurasi sistem. Sedangkan matriks evaluasi digunakan untuk membandingkan skor kesamaan dan label keterhubungan berdasarkan *alignment* oleh sistem dengan *gold standard*. Pada matriks evaluasi yang dihitung adalah nilai F1.

### Evaluasi *Chunk*

Pada sistem *chunk*ing otomatis yang telah dibangun, cara mengukur ketepatan proses *chunk*ing dengan menghitung akurasi sistem. Perhitungan akurasi yang dilakukan berdasarkan dari akurasi *chunk level.* Berikut merupakan rumus untuk menghitung akurasi *chunk level*.

(2-1)

Keterangan :

X = Jumlah chunk sistem yang sesuai dengan gold standard

N= Jumlah keseluruhan chunk

### 2.9.2. Matriks Evaluasi

Matriks evaluasi adalah metode evaluasi yang di terapkan dalam *SemEval2016 Task 2*. Untuk matriks evaluasi dilakukan perhitungan F1 atau F *measure* yang digunakan untuk mengukur performansi dengan cara membandingkan hasil dari sistem dengan manual yang dibuat dari ahli. Untuk perhitungan F1 terlebih dahulu dilakukan perhitungan presisi dan recall. Untuk penelitian ini digunakan matriks evaluasi yang sesuai dengan *SemEval2016 Task 2.* Pengukuran F1 dalam penelitian ini di bedakan menjadi dua yaitu pengukuran F1 untuk sistem dan pengukuran F1 untuk *gold Chunk*  (dataset yang sudah di potong). Adapun persamaan yang digunakan untuk perhitungan presisi, recall dan juga F1 yaitu sebagai berikut.

(2-2)

(2-3)

(2-4)

Keterangan untuk persamaan tersebut yaitu sebagai berikut.

* *True Positive* (TP) : Nilai yang akan didapat ketika hasil prediksi *true*(bernilai 1) dan nilai sebenarnya *true*(bernilai 1).
* *False Positive* (FP) : Nilai yang akan didapat ketika hasil prediksi *true*(bernilai 1) dan nilai sebenarnya *false*(bernilai 0).
* *False Negative* (FN) : Nilai yang akan didapat ketika hasil prediksi *false*(bernilai 0) dan nilai sebenarnya *true*(bernilai 1).

Ada empat pengukuran yang dilakukan dalam evaluasi yaitu sebagai berikut.

* F1 Ali: dimana perhitungan tidak melibatkan label keterhubungan dan skor, dengan kata lain label keterhubungan dan skor diabaikan.
* F1 Type: dimana perhitungan hnaya melibatkan label keterhubungan dan skor diabaikan.
* F1 Score : dimana perhitungan hanya melibatkan skor dan label keterhubungan diabaikan.
* F1 Type+Score : di mana pengukuran dilakukan berdasarkan kecocokan antara label keterhubungan dan juga skor.

Adapun matriks evaluasi yang digunakan untuk penelitian ini yaitu dapat dilihat pada tabel sebagai berikut .

Tabel 2‑1 Matriks Evaluasi

| System | | | | |
| --- | --- | --- | --- | --- |
| Dataset | F1 Ali | F1 Type | F1 Score | F1 Type+ Score |
| Data pasangan ayat berdasarkan indeks temantik |  |  |  |  |
| Data pasangan ayat dari Qursim |  |  |  |  |
| Gabungan data indeks temantik dan Qursim |  |  |  |  |

# Perancangan Sistem

Pada bab ini dipaparkan mengenai perancangan sistem, skema pengujian yangdilakukan, deskripsi perangkat lunak yang dibangun dan analisis kebutuhan.

## 3.1 Gambaran Umum Sistem

Pada dasarnya sistem *Interpretable* STS yang dibangun membutuhkan data inputan berupa pasangan ayat yang dibandingkan dan dicari nilai kesamaan dan label keterhubungan berdasarkan *alignment* antara kedua potong ayat terjemahan tersebut. Selanjutnya sistem yang dibangun mengeluarkan outputan berupa nilai kesamaan dan jenis label yang dimiliki oleh kedua potongan ayat terjemahan Alquran dalam terjemahan bahasa Inggris berdasarkan *alignment*nya. Kemudian sistem menampilkan jenis label yang ada berdasarkan aturan label yang telah ditentukan dalam penelitian ini. Setelah itu, sistem akan mengevaluasi perhitungan nilai kesamaan yang dikelola oleh sistem, dan di korelasikan dengan nilai gold standard yang merupakan penilaian secara manual dari perhitungan manusia yang termasuk dalam penelitian ini yaitu penulis,beserta orang yang ahli dalam bidang Alquran. Sistem yang dibangun dibagi menjadi dua yaitu sistem untuk *chunking* dan sistem untuk *Interpretable* STS. Berikut alur proses sistem secara umum.



Figure ‑ Alur proses sistem secara umum

Penjelasan mengenai tahapan dan proses pada Gambar 3-1 adalah sebagai berikut.

1. Sistem membaca masukan yang berupa data pasangan ayat Alquran terjemahan yang terdiri dari data Qursim dan data indeks tematik.
2. Sistem melakukan *preprocessing*  data seperti melakukan menghilangkan simbol dan tanda baca yang tidak perlu terhadap data input pasangan ayat Alquran terjemahan bahasa Inggris.
3. Sistem melakukan pemotongan atau *chunking* terhadap data input pasangan ayat Alquran terjemahan bahasa Inggris.
4. Hasil dari pemotongan dilanjutkan pada proses *alignment*
5. Sistem memberikan label keterhubungan berdasarkan data yang terhubung (*alignment)*
6. Sistem menghitung skor kesamaan berdasarkan data yang terhubung (*alignment)*
7. Sistem melakukan evaluasi menggunakan matriks evaluasi yang dibandingkan dengan *gold standard.* Pada tahap ini sistem menghasilkan keluaran berupa nilai F1, F1 score, F1 Type, F1 Type+Score.

Adapun gambaran untuk sistem *chunking*  yaitu sebagai berikut.



Figure ‑ Alur proses chunking

Penjelasan mengenai tahapan dan proses pada Gambar 3-2 adalah sebagai berikut.

1. Sistem membaca masukan yang berupa data pasangan ayat Alquran terjemahan yang terdiri dari data Qursim dan data indeks tematik.
2. Sistem melakukan *preprocessing*  data seperti melakukan menghilangkan simbol dan tanda baca yang tidak diperlukan oleh data input pasangan ayat Alquran terjemahan bahasa Inggris.
3. Sistem melakukan pemotongan atau *chunking* terhadap data input pasangan ayat Alquran terjemahan bahasa Inggris.
4. Hasil dari proses *chunking*  akan disimpan ke dalam database data *chunking* pasangan ayat.



Figure ‑ Alur proses untuk Interpretable STS

Adapun penjelasan untuk Gambar 3-3 yaitu sebagai berikut.

1. Sistem membaca inputan data yang berupa data *chunking* pasangan ayat Alquran
2. Data tersebut akan dilakukan proses *alignment*.
3. Sistem memberikan label keterhubungan
4. Sistem memberikan skor kesamaan
5. Sistem akan melakukan evaluasi dengan membandingkan hasil dari sistem dengan *gold standard*

## 3.2 Pengumpulan Data

Pada tahapan ini dilakukan proses pengumpulan dan pembuatan data pasangan ayat Alquran, data *chunking* pasangan ayat secara manual, dan data *gold standard.*

### 3.2.1. Data Pasangan Ayat

Data yang digunakan pada penelitian ini yaitu 400 pasang ayat terjemahan Alquran, dimana 200 pasang ayat didapatkan dari data *Qursim* hasil penelitian keterkaitan *corpus* Alquran berdasarkan tafsir Ibn katsir oleh Abdul-Baquee M. Sharaf dan Eric S [13]. Atwell yang dimuat dalam situs web *textminingthequran.com* dan 200 pasang ayat lainnya diambil dari data indeks tematik pada Alquran terjemahan Kementrian Agama RI dan situs web Pusat Kajian Hadist Al-Mughni *alquranalhadi.com.*

Tabel ‑ : Contoh Data Pasangan Ayat

|  |  |  |
| --- | --- | --- |
| Data Indeks Temantik | | |
| Posisi Ayat | Ayat 1 | Ayat 2 |
| 7 :155 – 20:65 | the magicians said,” O Moses How dare who will cast the first or we the sorcerers who will throw”. | after they gathered they the sorcerers said,” O Moses choose whether ye that cast first or speakest the sorcerers who first threw”. |
| 2 : 34 – 7: 11 | And remember when We God said to the angels,” Adore ye angel to Adam and they bowed down angel except Satan he Satan reluctant and haughtiness and is it Satan including the the unbelievers”. | Behold We God has created you Adam and We God form of body then We God say to the angels,” Adore ye angel to Adam they angel prostrate except Satan He Satan not including them who prostrate”. |
| Data Qursim | | |
| Posisi Ayat | Ayat 1 | Ayat 2 |
| 18 :28—6: 52 | And keep Muhammad patient [by being] with those who call upon believer Lord in the morning and the evening, seeking Allah countenance. | And do not send away those who call upon who fear Lord morning and afternoon, seeking Allah countenance. |
| 34 : 1—28 : 70 | [All] praise is [due] to Allah, to whom belongs whatever is in the heavens and whatever is in the earth, and to Allah belongs [all] praise in the Hereafter. | To Allah is [due all] praise in the first [life] and the Hereafter. |

### 3.2.2. Anotasi Data

Untuk 400 data pasangan ayat yang telah terkumpul terlebih dahulu dilakukan pemotongan ayat dan anotasi manual dengan menandai beberapa unit semantik yang terdapat pada data pasangan ayat Alquran. Adapun anotasi manual yang dilakukan yaitu menambahkan kata ganti atau *pronoun* pada suatu objek data terjemahan ayat Alquran tersebut. Sedangan *chunking* atau pemotongan dilakukan berdasarkan hubungan kata. Adapun aturan anotasi yang diterapkan dalam data pasangan ayat Alquran adalah seperti pada tabel

Tabel ‑ Anotasi Data

|  |  |  |
| --- | --- | --- |
| Unit Semantik | Simbol | Keterangan |
| *Chunk* | [ ] | Pemotongan kata yang dilakukan berdasarkan konstituen atau kumpulan kata yang sesuai aturan. |
| Kata ganti | < > | Kata ganti dari subjek atau objek yang terdapat dalam potongan ayat, seperti kata they yang dimaksud apakah God, human atau objek lain. Contoh : they < unbelievers > |

### 3.2.3. Gold Standard

Data *gold standard*  pada penelitian ini dibuat berdasarkan intuisi manusia sebagai seorang yang ahli. Data *gold standard*  dibuat secara manual oleh penulis, pembuatan Data *gold standard*  yaitu dengan cara memberikan nilai kesamaan semantik dan label keterhubungan antar *alignment*  pada masing-masing pasangan ayat terjemahan Alquran. Sebelum melakukan proses pembuatan *gold standard* ada beberapa kriteria yang harus dipenuhi oleh penulis sebagai pembuat *gold standard.* Berikut kriteria yang harus dipenuhi oleh penulis.

1. Berusia diatas 18 tahun
2. Beragama islam
3. Berakal sehat
4. Sudah pernah khattam Alquran
5. Interaksi dengan Alquran relatif sering ( minimal satu juz per hari).

## Pembangunan *chunking* otomatis

Untuk proses pembangunan proses *chunking* otomatis pada 400 pasangan data terjemahan Alquran

## 3.4 Analisis Kebutuhan Fungsionalitas Sistem

### 3.4.1 Kebutuhan Fungsional

Beberapa fungsionalitas dari sistem yang harus dipenuhi dalam penelitian tugas akhir ini adalah.

1. Sistemmenampilkan GUI yang dapat dijalankan oleh user dengan mudah.
2. Sistem dapat membaca inputan dataset dari user yang berupa dataset pasangan potongan terjemahan ayat
3. Sistem dapat melakukan perhitungan kesamaan makna secara semantik
4. Sistem dapat melakukan evaluasi performansi dari sistem dengan *gold standard.*.

### 3.4.2 Spesifikasi Perangkat Keras

Spesifikasi perangkat keras yang digunakan dalam pembangunan sistem yaitu.

1. Processor Intel ***R*** Core***TM*** i5-3210M
2. 2.5GHz RAM 4.00 GB
3. HDD 500 G

### 3.4.3 Spesifikasi Sistem

Spesifikasi sistem yang digunakan dalam pembangunan sistem yaitu.

1. Sistem Operasi : Linux 32 bit.
2. Bahasa Pemograman : python
3. IDE : JetBrains PyCharm
4. Library dan Corpus : Stanford ner, Stanford CoreNLP,WordNet 2.1

# Bab IV Pengujian dan Analisis

Pada bagian inidipaparkan mengenai pengujian apa saja yang dilakukan, skenario pengujian, hasil pengujian dan juga analisis terhadap hasil pengujian.

## 4.1 Set Data

### 4.1.1 Data Pasangan Ayat

Data yang digunakan sebanyak 400 pasang ayat terjemahan Alquran, dimana 200 pasang ayat didapatkan dari data *Qursim* [13] hasil penelitian keterkaitan *corpus* Alquran berdasarkan tafsir Ibn katsir oleh Abdul-Baquee M. Sharaf dan Eric S. Atwell yang dimuat dalam situs web *textminingthequran.com* dan 200 pasang ayat lainnya diambil dari data indeks tematik pada Alquran terjemahan Kementrian Agama RI dan situs web Pusat Kajian Hadist Al-Mughni alquranalhadi.com. Selanjutnya data potong-potong (*chunk*) sesuai dengan intuisi manusia, karena tahapan ini proses pemotongan untuk data dilakukan secara manual. Dari hasil data yang sudah di potongdilakukan perubahan kata ganti sesuai dengan subjek yang berperan.

Tabel 0‑1 -1 : Contoh Data pasangan ayat

|  |  |  |  |
| --- | --- | --- | --- |
| Letak Ayat 1 | Letak Ayat 2 | Ayat1 | Ayat2 |
| 23 : 13 | 77 : 21 | Then We <God> placed him water mani stored in place that firmly | then We <God> put him in place that firmly |
| 22: 5 | 23 : 14 | o human if ye in doubt about resurrection from the grave behold! verily We <God> has made you from land then from a drop of sperm then from congealed blood then from a lump of meat that perfect occurrence and not perfect so We <God> explain to you and we <God> specify in uterus what We <God> travel until time is specified then We <God> remove you as baby then with gradually ye came to maturity and in between you no take and therewith in between you that elongated age until mild forget fulness so that he does not know more nothing that once he knew. | Then the sperm We <God> made congealed blood and congealed blood We <God> made a lump of meat and a lump of meat was We <God> made bones and bones We <God> sachet with flesh Then We <God> made him living that form Then Most that makes God Creator Best |
| 30 : 21 | 16 : 72 | and in between His signs God is He <God> create you wives from type own so that you tend to and feel quiet him and made him <God> between compassion and compassion Behold on that really are signs for the who think | God make for you wives from type of you own and make you from wives you children and grandchildren and your sustenance from good. Then why they believe to confound and reject the favors of Allah |
| 2 : 34 | 7 :11 | And remember when We <God> said to the angels Adore ye angel to Adam and they bowed down angel except Satan he Satan reluctant and haughtiness and is it Satan including the the unbelievers | Behold We <God> has created you Adam and We God form of body then We God say to the angels Adore ye angel to Adam they angel prostrate except Satan He Satan not including them who prostrate |

### 4.1.2 Data Gold Standard

Data gold standard adalah data yang dipergunakan untuk menjadi data pembanding untuk hasil skor kesamaan semantik yang dihasilkan oleh sistem. Pembuatan *gold standard*  disesuaikan dengan cara pembuatan *gold standard*  pada semEval 2016 task 2. Berikut merupakan salah satu *gold standard*  untuk satu pasangan ayat.

|  |  |  |  |
| --- | --- | --- | --- |
| Letak Ayat 1 | Letak Ayat 2 | Chunk Ayat1 | Chunk Ayat2 |
| 23 : 13 | 77 : 21 | [Then We <God>] [placed him] [water mani] [stored] [in] [place] [that firmly] | [then We <God> put him] [in] [place] [that firmly] |
| 22: 5 | 23 : 14 | [o human] [if] [ye] [in] [doubt] [about] [resurrection from the grave] [behold verily We <God> has made you] [from] [land] [then] [from] [a drop of sperm] [then] [from] [congealed blood] [then] [from] [a lump of meat] [that perfect occurrence] [and not perfect] [so We <God>] [explain] [to you] [and we <God>] [specify] [in] [uterus] [what] [We <God>] [travel] [until] [time] [is specified] [then] [We <God>] [remove you] [as baby] [then with gradually] [ye] [came to maturity] [and in between you] [no] [take] [and therewith in between you] [that elongated] [age until mild forgetfulness] [so that he does not] [know more] [nothing] [that once he knew] | [Then] [the sperm] [We <God> made] [congealed blood] [and] [congealed blood] [We <God> made] [a lump of meat] [and] [a lump of meat was] [We <God> made] [bones] [and] [bones] [We <God> sachet] [with flesh] [Then] [We <God> made him] [living] [that] [form] [Then Most that makes] [God] [Creator] [Best] |
| 30 : 21 | 16 : 72 | [and in between] [His signs <God>] [is He <God> create] [you] [wives] [from] [type own] [so that you tend to and feel quiet] [him] [and made him <God>] [between] [compassion] [and compassion] [Behold on that really are signs] [for the] [who think] | [God] [make] [for you] [wives] [from] [type of you own] [and make] [you] [from] [wives you] [children] [and grandchildren] [and your sustenance] [from] [good] [Then why they believe to confound and reject] [the favors of Allah] |

<sentence id="2" status="">

Ayat 1 : say muhammad i afraid will doom a great day if i disobeyed my lord

Ayat 2: say muhammad i afraid will torment a great day the last day if i disobeyed my lord

<source>

1 <==> 1 // EQUI // 5 // say <==> say

2 3 4 <==> 2 3 4 // EQUI // 5 // muhammad i afraid <==> muhammad i afraid

5 7 8 9<==> 5 7 8 9 // EQUI // 5 // will // a great day <==> will a great day

10 11 12 13 14 <==> 12 13 14 15 16 // EQUI // 5 // if i disobeyed my lord <==> if i disobeyed my lord

0 <==> 10 11 // NOALI // NIL // -not aligned- <==> the last day

6 <==> 6 // SIMI // 3 // doom <==> torment

6</alignment>

## 4.2 Pengujian Sistem

Skenario pengujian sistem yang dilakukan pada penelitian ini dilakukan sesuai dengan tujuan dari penelitian ini yaitu untuk menganalisis efektitas *chunk* yang telah dibuat oleh sistem dan menganalisis faktor apa saja yang sangat mempengaruhi skor kesamaan pada alignment.

### 4.2.1 Tujuan Pengujian

Adapun tujuan pengujian yang ingin dicapai dalam pengujian sistem ada-lah sebagai berikut :

1. Menganalisis efektifitas sistem pemotongan (*chunking* ) yang dibuat berdasarkan set data yang diuji.
2. Menganalisis pengaruh faktor apa yang berpengaruh terhadap pemberian label keterhubungan
3. Menganalisis faktor apa saja yang sangat mempengaruhi F1.

### 4.2.2. Skenario Pengujian

Proses pengujian dilakukan dengan menggunakan set data pasangan ayat yang telah dibangun yaitu set data pasangan ayat Alquran berdasarkan indeks tematik, *Qursim,* dan gabungan antara indeks tematik, *Qursim*. Skenario pengujian sistem yang dilakukan yaitu sebagai berikut.

1. Perhitungan ketepatan antara *chunk*  manual dengan *chunk* oleh sistem dengan melihat nilai akurasi pada *chunk level* adalah cara untuk mengetahui tingkat efektifitas sistem *chunking* yang telah dibuat oleh penulis.
2. Membandingkan nilai F1 berdasarkan penggunaan feature yang digunakan.

## Eksperimen

### Analisis

* **Analisis ketepatan chunk**

Pengujian dilakukan dengan cara membandingkan nilai akurasi *chunk level*  pada setiap data uji yang terdiri dari data tematik, data Qursim dan data gabungan antara data tematik dan data Qursim.

Gambar 0‑1 Grafik Akurasi chunk level

Perbandingan nilai akurasi dapat dilihat dari Gambar 7, dapat diketahui bahwa nilai akurasi antara ketiga data. Data yang memiliki akurasi chunk level tertinggi adalah data gabungan dari data Qursim dan data indeks tematik.

* **Analisis perbandingan F1**

Perhitungan F1 dibedakan menjadi 4 yaitu sebagai berikut .

1. F1 Ali

Gambar 0‑2 Grafik F1 Ali

1. F1 Type

Gambar 0‑3 Grafik F1 Type

1. F1 Score

Gambar 0‑4 Grafik F1 Score

1. F1 Type + Score

Gambar 0‑5 Grafik F1 Type+Score

Gambar 0‑6 Grafik F1 secara keseluruhan

Dari Grafik F1 dapat diketahui bahwa nilai F1 paling tinggi adalah F1 Ali. F1 Ali yaitu perhitungan F1 yang hanya memperhatikan kecocokan alignment antar ayat.

# Bab V Kesimpulan

## 5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan pada Tugas Akhir ini, dapat disimpulkan bahwa :

1. Untuk menentukan kesamaan makna antar potongan terjemahan ayat Alquran ada tahapan proses yang dilakukan yaitu
2. Untuk perbandingan Akurasi untuk proses chunking berdasarkan chunk level dapat diketahui bahwa akurasi paling tinggi adalah nilai akurasi untuk data gabungan antara data Qursim dan data indeks tematik.
3. Untuk perbandingan F1 antara F1 ali, F1 Type, F1 Score, F1 Type +Score dapat diketahui bahwa F1 ali memiliki nilai paling tinggi diantara F1 lainnya pada semua data uji yang dicoba.

## 5.2 Saran

Saran yang diperlukan untuk pengembangan pada Tugas Akhir ini adalah sebagai berikut :

1. Untuk penelitian interpretable STS terjemahan ayat Alquran sebaiknya menggunakan metode yang lain dan dataset yang dipergunakan sebaiknya lebih diperbanyak lagi.
2. Untuk pembuatan proses chunking otomatis harus menggunakan metode lain.

# ­­­­­­Daftar Pustaka

1. Muhammad Sya'ban Al-Ha zh Abdul Aziz Ahmad, M.Si. Robbani Al-Quran per Kata, Tajwid Warna.
2. Abney, S. (1991). Parsing by *Chunk* s. In: Robert Berwick, Steven Abney and Carol Tenny (eds.), Principle-Based Parsing. Kluwer Academic Publishers, Dordrecht.
3. Banjade Rajendra, dkk (2015). A System for Measuring and Interpreting Semantic Textual Similarity .In Proceedings of the 9th International Workshop on Semantic Evaluation.
4. Eneko Agirre, Mona Diab, Daniel Cer, and Aitor Gonzalez-Agirre. 2012. Semeval-2012 task 6: A pilot on semantic textual similarity. In Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, pages 385–393. Association for Computational Linguistics.
5. Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, andWeiwei Guo. 2013. sem 2013 shared task:Semantic textual similarity, includin g a pilot on typedsimilarity.In In\* SEM 2013: The Second Joint Conferenceon Lexical and Computational Semantics. Association for Computational Linguistics. Citeseer.
6. Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre,Weiwei Guo,Iigo Lopez-Gazpio, Montse Maritxalar, Rada Mihalcea,German Rigau, Larraitz Uria, and Janyce Wiebe.2015. SemEval-2015 Task 2: Semantic Textual Similarity,English, Spanish and Pilot on Interpretability. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, June. Association for Computational Linguistics.
7. Claudia Leacock and Martin Chodorow. 1998. Combininglocal context and wordnet similarity for word sense identification. WordNet: An electronic lexical database, 49(2):265–283.
8. Fellbaum, Christiane (2005). WordNet and wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier, 665-670
9. Christiane Fellbaum. 1998. WordNet. Wiley Online Library.
10. Dan Melamed. (1998) Manual annotation of translational equivalence: The blinker project. Technical Report 98-07, Institute for Research in Cognitive Science, Philadelphia
11. Banjade, Rajendra, Nabin Maharjan, Nobal B Niraula, and Vasile Rus. 2016. “DTSim at SemEval-2016 Task 2 : Interpreting Similarity of Texts Based on Automated *Chunk*ing , *Chunk* Alignment and Semantic Relation Prediction,” 809–13.
12. Banjade, Rajendra, Nobal B Niraula, Nabin Maharjan, Vasile Rus, Dan Stefanescu, Mihai Lintean, and Dipesh Gautam. 2015. “NeRoSim: A System for Measuring and Interpreting Semantic Textual Similarity.” *SemEval2015*, no. SemEval: 164–71.
13. Sharaf, M., and Eric S. Atwell. 2012. “Qursim: A Corpus for Evaluation of Relatedness in Short Texts.” *8th International Conference on Language Resources and Evaluation (LREC’12)*, no. January 2012: 2295–2302. doi:10.13140/2.1.4007.0088.
14. Susanto, Alvin. 2008. “Penggunaan Algoritma Hungarian Dalam Menyelesaikan Persoalan Matriks Berbobot,” no. 10.

# Lampiran

1. **Data Pasangan Ayat Alquran dari Indeks Temantik dengan proses pemotongan secara manual**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **NO** | **Posisi Ayat** | | | | **Pasangan Ayat** | | ***Chunk* Pasangan Ayat** | |
| **Ayat 1** | | **Ayat 2** | | **Ayat 1** | **Ayat 2** | **Ayat 1** | **Ayat 2** |
| 1 | 2 | 162 | 3 | 88 | They unbelievers eternal life in the curse of will not lightened punishment from them unbelievers and not again they unbelievers reprieved | they unbelievers eternal life in it the not lightened punishment from them unbelievers and not again they unbelievers reprieved | [They <unbelievers> eternal life] [in the curse of] [will not] [lightened] [punishment] [from them <unbelievers>] [and not again] [they <unbelievers>] [reprieved] | [they <unbelievers>] [eternal life] [in it] [not] [lightened] [punishment] [from them <unbelievers>] [and not again] [they <unbelievers>] [reprieved] |
| 2 | 6 | 15 | 39 | 13 | Say Muhammad I afraid will doom a great day if I disobeyed My Lord | Say Muhammad I afraid will torment a great daythe last day if I disobeyed My Lord | [Say <Muhammad>] [I] [afraid] [will doom] [a great day] [if] [I disobeyed] [My Lord] | [Say] [Muhammad] [I] [afraid] [will torment] [a great day] [the last day] [if] [I disobeyed] [My Lord] |

1. http://at.qrci.org/semeval2016/task2 [↑](#footnote-ref-1)
2. http://alt.qcri.org/semeval2016/task2/index.php?id=results [↑](#footnote-ref-2)
3. http://alt.qcri.org/semeval2016 [↑](#footnote-ref-3)
4. http://www.clips.ua.ac.be/conll2000/chunking/ [↑](#footnote-ref-4)
5. http://alt.qcri.org/semeval2016/task2/annotationguidelinesinterpretablests2016v2.2.pdf [↑](#footnote-ref-5)
6. http://ixa2.si.ehu.es/ixa-pipes/ [↑](#footnote-ref-6)
7. http://stanfordnlp.github.io/CoreNLP/ [↑](#footnote-ref-7)
8. http://nlp.stanford.edu/software/lex-parser.shtml [↑](#footnote-ref-8)
9. http://en.wikipedia.org/wiki/Hungarian\_algorithm [↑](#footnote-ref-9)