Analisis dan Implementasi *Similarity* dengan *Word Alignment* pada Nabi Abraham dalam Alkitab dan Al-quran

Grace Duma Tambunan[1], Moch. Arif Bijaksana[2], Said AL Faraby[3]
[123] Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung
Jalan Telekomunikasi 1, Dayeuh Kolot, Bandung 40257
[1] **gracedt@student.telkomuniversity.ac.id,** [2]**arifbijaksana@telkomuniversity.ac.id,**
[3]**said.al.faraby@gmail.com**

Abstrak

Agama Kristen dan Islam merupakan agama yang sah di Indonesia. Pada kitab suci kedua agama terdapat kisah nabi Abraham dan memiliki kesamaan. Hal tersebut telah menjadi topik yang banyak dibahas. Oleh karena itu, perlu pengetahuan untuk melihat sisi kesamaan dan kemiripan kisah nabi tersebut. Dengan menerapkan konsep penambangan teks dan metode *word alignment* akan dilakukan perbandingan kisah antara kedua kitab. Metode ini dipilih karena telah banyak digunakan dalam penelitian pemrosesan teks, dan juga telah menjadi topik yang banyak digunakan dalam kompetisi SemEval dan menjadi metode yang terbaik pada seri SemEval 2014 dan 2015 yaitu *Sultan Aligner*. Dalam penelitian ini, digunakan Kisah Nabi Abraham sebagai dataset. Data input kisah Nabi Abraham diambil dari buku Kisah para Nabi dan Rasul karangan Ibnu Katsir yang merujuk pada ayat-ayat Al-Quran dan dari kisah Nabi Abraham pada Alkitab. Kisah yang didapatkan dari buku karangan Ibnu Katsir ini telah dibandingkan dengan Al-Quran secara manual sehingga dapat dipastikan keakuratan datanya. Hasil dari penelitian berupa nilai korelasi dengan cara membandingkan hasil kesamaan semantik yang dihasilkan sistem dengan menggunakan *word alignment* dan *gold standard* yang dibangun secara manual dengan intuisi manusia. Dari penelitian tugas akhir diperoleh nilai korelasi sebesar 0,8025 pada *single proportion* dan 0,7991 pada *separate proportion*.

Kata Kunci: Nabi Abraham, kesamaan, kemiripan, *word alignment*, korelasi, Alkitab, Al-quran, *gold standard*

Abstract

Christianity and Islam are legitimate religions in Indonesia. In the scriptures both religions have the story of the prophet Abraham and have similarities. It has become a widely discussed topic. Therefore, it is necessary to see the similarity and resemblance of the Prophet's story. By applying the concept of text mining and word alignment method will be done comparison of the story between the two books. This method was chosen because it has been widely used in textual research, and has also become a widely used topic in the SemEval competition and became the best method in SemEval 2014 and 2015 series namely Sultan Aligner. In this study, the story of Prophet Ibrahim is used as a dataset. The input data of the story of Prophet Ibrahim is taken from the book of Acts of Prophets and Apostles by Ibn Kathir who is alive in the verses of the Qur'an and from the story of the Prophet Abraham in the Bible. The story produced from the book by Ibn Kathir has been compared with the Al-Quran manually so it can be ascertained the accuracy of the data. The results of the study by comparing the results of semantic similarities produced by using the word alignment and the gold standard built manually with human intuition. From the results of the study. With an average proportion and 0.7991 in a separate proportion.

Keywords: Prophet Ibrahim, similarity, likeness, word alignment, Bible, Al-quran, gold standard

1. Pendahuluan

Terdapat beberapa agama dan kepercayaan yang telah diakui oleh pemerintah Indonesia dan secara sosial atau pengakuan dalam masyarakat sebagai bagian dari adat dan budaya. Setiap agama memiliki ajaran-ajaran yang menjadi pedoman dan pegangan bagi penganutnya yang tertulis dalam masing-masing kitabnya. Sebagian besar ajaran tersebut mencakup pengalaman hidup beberapa nabi. Beberapa kisah nabi dalam kitab suatu agama, juga tertulis dalam kitab agama yang lain. Sebagai contoh ialah Nabi Abraham dalam Agama Kristen yang tertulis dalam Alkitab dan Nabi Ibrahim dalam Agama Islam yang tertulis dalam Al-quran. Kisah Nabi Abraham tersebut memiliki kemiripan dan kesamaan.

Banyak perbincangan dan pembahasan mengenai kisah Nabi Abraham dan kisah nabi lain dalam Alkitab dan Al-quran yang memiliki kesamaan dengan berbagai sumber. Untuk mengetahui kesamaan kedua kisah nabi tersebut, terlebih dahulu harus membaca dan memahami ayat-ayat yang terdapat dalam Alkitab maupun Al-quran. Selain itu, perlu dilakukan pengumpulan ayat-ayat yang menggambarkan kisah Nabi Abraham karena ayat pada Al-quran yang berhubungan dengan kisah Nabi Abraham belum dikelompokkan dan tidak terurut sedangkan pada Alkitab telah terkelompokkan dan terurut. Permasalahan lainnya ialah antara ayat dalam Alkitab dan Al-quran atau sebaliknya beberapa memiliki keterhubungan umum ke khusus. Artinya, dalam ayat Alkitab suatu kisah dijabarkan secara umum sementara dalam ayat Al-quran dijabarkan secara rinci dan detail atau sebaliknya. Salah satu contoh, dalam Alkitab terdapat silsilah Nabi Abraham dan dijabarkan dalam beberapa ayat, sedangkan dalam Al-quran menjelaskan ayah Nabi Abraham.

Hal di atas merupakan bukti bahwa perlunya pengetahuan untuk melihat dari sisi kesamaan kedua kitab. Dengan menerapkan konsep Penambangan Teks dan Pemrosesan Bahasa Natural, dapat dibandingkan tingkat kemiripan isi kedua kisah Nabi Abraham dari Agama Islam dan Kristen.

Salah satu konsep dalam pengukuran kesamaan pasangan teks adalah *Semantics Textual Similarity* (STS). Pada penelitian tugas akhir ini, konsep STS dapat dilakukan dengan pendekatan *word alignment*. Pendekatan

word alignment dipilih karena merupakan salah satu metode yang cukup banyak digunakan pada *unsupervised system*. Pada salah satu kompetisi di bidang STS yaitu SemEval, *word alignment* merupakan metode yang sederhana dan menunjukkan performansi yang paling baik dalam dalam salah satu running di task STS pada SemEval 2014 dan SemEval 2015 yaitu *Sultan Aligner*[1]. *Sultan Aligner* mempertimbangkan kata identik dan sinonim yang teridentifikasi pada pasangan teks yang dikelompokkan pada *word similarity*. Selain itu *Sultan Aligner* juga mempertimbangkan makna kontekstual pasangan ayat dengan mengidentifikasi entitas nama (nama orang, tempat, perusahaan, dll) dan sekuen kata (minimal 2 kata secara terurut) yang dikelompokkan pada *contextual similarity*. Kedua kelompok ini akan digabungkan dengan perhitungan *single proportion* (tanpa memperhitungkan proporsi kata pasangan teks) dan *separate proportion* (memperhitungkan proporsi kata pasangan teks) sebagai hasil nilai kesamaan semantik [2][3]. Sehingga dengan menggunakan *Sultan Aligner* dapat mengidentifikasi kesamaan berdasarkan kata dan juga secara kontekstual.

2. Dasar Teori dan Perancangan Sistem

2.1. *Word Alignment*

Word Alignment adalah metode yang digunakan untuk menemukan kesamaan dalam pasangan teks. *Alignment* mampu menghasilkan informasi penting mengenai seberapa besar dua kalimat saling berkaitan. Pada penelitian ini, metode yang digunakan ialah *Sultan Aligner* [3]. Terdapat dua pendekatan yaitu *word similarity* dan *contextual similarity*.

2.1.1. *Word similarity*

Word similarity merupakan untuk mengidentifikasi kata atau frasa yang memiliki nilai kesamaan tanpa mempertimbangkan konteks kata. Berdasarkan penelitian Sultan[2] perhitungan *alignment* berdasarkan semantik kata penyusun kalimat terdapat beberapa aturan, yakni:

1. Jika terdapat 2 kata atau frasa yang identik (secara struktural kata) maka akan diberikan skor 1.
2. Jika terdapat 2 kata atau frasa yang memiliki kesamaan makna berdasarkan PPDB maka akan diberikan skor 0.9^2.
3. Jika kata atau frasa tidak memiliki kesamaan akan diberi skor 0.

Pada *word similarity* terdapat dua buah fitur *alignment* yang digunakan untuk penelitian ini, yaitu *identical word* dan *alignment* Sinonim.

2.1.2. *Contextual similarity*

Contextual Similarity merupakan *alignment* untuk mengidentifikasi kemiripan dan kesamaan antara dua kalimat berdasarkan konteks kata-kata yang menyusun kalimat tersebut. Pada *contextual similarity* terdapat dua buah fitur *alignment* yang digunakan untuk penelitian ini, yaitu *name entity* dan *word sequence*.

2.2. Perhitungan *Similarity*

Pada penelitian Sultan terdapat dua perhitungan nilai kesamaan semantik yaitu, *single proportion* dan *separate proportion* dan kedua perhitungan tersebut juga digunakan dalam penelitian ini. Berikut persamaan perhitungannya.

2.2.1. *Single Proportion*

$$sts(S^{(1)}, S^{(2)}) = \frac{n_c^a(S^{(1)}) + n_c^a(S^{(2)})}{n_c(S^{(1)}) + n_c(S^{(2)})} \qquad (2.1)$$

dimana $n_c^a(S^{(i)})$ merupakan jumlah token yang pada kalimat *i* yang ter-*align* atau tersejajarkan dengan token pada kalimat pasangannya, dan $n_c(S^{(i)})$ merupakan jumlah kata yang terdapat dalam kalimat *i*.

2.2.2. *Separate Proportion*

$$prop_{Al}^{(1)} = \frac{|\{i : [\exists j : (i,j) \in Al] \; and \; w_i^{(1)} \in C\}|}{|\{i : w_i^{(1)} \in C\}|} \qquad (2.2)$$

dimana *C* adalah himpunan semua kata-kata konten dalam bahasa Indonesia dan *Al* adalah kata yang teridentifikasi align pada ayat kedua terhadap ayat pertama. Sedangkan proporsi pada ayat kedua juga dapat dihitung dengan cara yang sama [5]. Kemudian perhitungan nilai kesamaan semantik pada metode perhitungan diperoleh dengan cara sebagai berikut :

$$sts(S^{(1)}, S^{(2)}) = \frac{2 \; x \; prop_{Al}^{(1)} x \; prop_{Al}^{(2)})}{prop_{Al}^{(1)} + prop_{Al}^{(2)}} \qquad (2.3)$$

3.

Adapun Perhitungan nilai kesamaan semantik dari hasil pendekatan *word similarity* dan *contextual similarity* dilakukan dengan cara sebagai berikut :

$$f(simW, simC) = 0.9 \times simW + 0.1 \times simC \qquad (2.4)$$

dimana *simW* merupakan nilai kesamaan semantik menggunakan pendekatan *word similarity* dan *simC* merupakan nilai kesamaan semantik menggunakan pendekatan *contextual similarity.*

2.3. *Gold Standard*

Gold Standard merupakan penilaian kesamaan pasangan teks yang didasarkan pada intuisi manusia. Penilaian ini akan menjadi pembanding terhadap nilai kesamaan yang dihasilkan oleh sistem yang akan dibangun. *Gold Standard* dibangun dengan cara mengumpulkan pendapat beberapa 10 responden dan nilai akan merepresentasikan tingkat kemiripan pasangan teks tersebut. Skor penilaian dalam range [0-5], dimana 0 menyatakan tidak terdapat kemiripan, 1 menyatakan skor paling rendah kemiripan hingga 5 menyatakan pasangan ayat 100% mirip atau dapat dikatakan sama.

2.4. Evaluasi

Evaluasi sistem dilakukan dengan menggunakan korelasi Pearson. Korelasi digunakan untuk melihat keterhubungan *gold standard* dengan hasil kesamaan sistem yaitu *single proportion* dan *separate proportion.* Nilai korelasi yang tinggi (mendekati 100%) menyatakan sistem yang dibangun semakin baik.

2.5. Dataset

Dataset yang digunakan pada penelitian adalah kisah Nabi Abraham yang terdapat pada Alkitab dan kisah Nabi Ibraham yang terdapat pada Alquran terjemahan bahasa Indonesia. Kisah Nabi Abraham dari Al-quran didapatkan dari buku Kisah Nabi dan Rasul Karangan Ibnu Katsir yang bersumber dari ayat-ayat Al-quran dan telah dibandingkan langsung dengan ayat-ayat Al-quran secara manual oleh penulis. Sedangkan kisah Nabi Abraham dari Alkitab, didapatkan dari kisah yang termuat dalam kitab digital agama Kristen telah terverifikasi oleh Lembaga Alkitab Indonesia (LAI).

Dataset yang sudah terkumpul kemudian dipasangkan secara manual oleh penulis. Dataset sebelum dipasangkan terdiri dari 181 ayat Al-quran dan 375 ayat Alkitab (kitab Kejadian). Kemudian setelah dipasangkan, terdapat 18 pasang ayat, dan 356 ayat Alkitab yang tidak berpasangan dan 163 ayat Al-quran yang tidak berpasangan. Data yang telah dihasilkan kemudian disimpan dalam file dengan format .txt dengan tombol *tab* sebagai tanda pemisah ayat dan *enter* sebagai pemisah pasangan ayat.

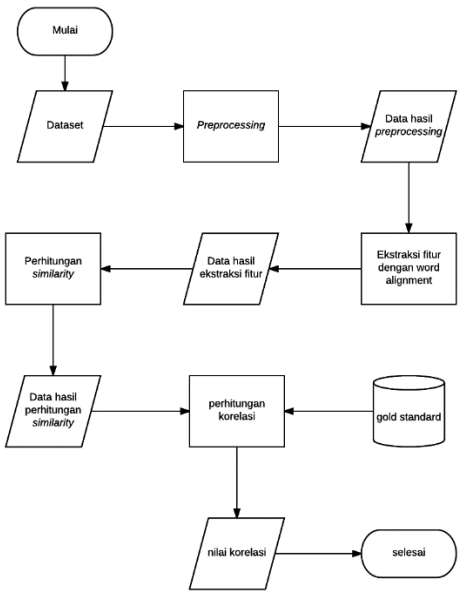
2.5.1. Al-quran

Al-quran merupakan kitab suci umat Islam yang berisi firman Allah yang diturunkan kepada Nabi Muhammad SAW dengan perantaraan malaikat Jibril untuk dibaca, dipahami, dan diamalkan sebagai petunjuk atau pedoman hidup bagi umat manusia[4]. Alquran terdiri dari 6236 ayat, 114 surat, 30 Juz. Alquran merupakan pedoman utama umat Islam dimana jumlah kaum muslimin merupakan urutan kedua paling banyak di dunia yaitu sekitar 1.6 milyar jiwa. Alquran mencatat Nabi Ibrahim yang disebut sebanyak 62 kali dalam 24 surat. Nabi Ibrahim merupakan salah satu Nabi besar dan menjadi salah satu dari 5 Nabi Ulul Azmi yaitu Nabi yang memiliki ketabahan yang luar biasa dalam menyebarkan ajaran Allah.

2.5.2. Alkitab

Alkitab merupakan kitab suci umat Kristen yang terdiri dari Perjanjian Lama(PL) dan Pernjanjian Baru(PB)[4]. Berdasarkan statistik salah satu penyedia Alkitab Daring, Alkitab mempunyai 66 kitab (PL: 39, PB: 27), 1.189 pasal (PL: 929 / PB: 260) dan 31.102 ayat (PL: 23.145 / PB: 7.957). Alkitab telah menjadi pedoman dan dasar kehidupan umat Kristen. Yesus Kristus merupakan tokoh utama dalam Alkitab karena seluruh kitab pada dasarnya adalah mengenai Dia. Nama Nabi Abraham mempunyai Arti Bapa bangsa-bangsa. Alkitab mencatat beberapa profil Abraham, Abraham merupakan keturunan Terah, berpasangan dengan Sarai, yang memperanakkan Isak dan Ismael, lahir di Urkasdim dan meninggal di Hebron dan dikubur di Gua Makhpela bersama istrinya. Nabi Abraham merupakan orang yang taat dan setia pada Allah, sehingga Allah menjanjikannya akan mempunyai keturunan sebanyak bintang di langit.

2.6. Perancangan Sistem



Gambar 1 Gambaran umum sistem

Berikut penjelasan tahapan dan proses utama sistem pada Gambar 1:

1. Sistem akan menerima inputan dataset berupa pasangan ayat Alkitab dan Alquran terjemahan bahasa Indonesia.
2. Sistem melakukan *preprocessing* pada data inputan, yaitu berupa tokenisasi untuk memotong setiap string, *stopword removal* untuk menghapus kata yang tidak mempengaru *alignment* seperti kata hubung, penghapusan tanda baca, *stemming* untuk menormalisasi kata yang mengandung imbuhan dan termasuk penghapusan spasi. Kemudian dihasilkan data bersih hasil *preprocessing*.
3. Sistem melakukan ekstraksi fitur dengan menggunakan fitur *alignment*, yaitu *identical word*, sinonim, *name entity*, *word sequence*. Ekstraksi dilakukan dengan mensejajarkan setiap pasangan ayat, kemudian melakukan pengecekan setiap kata pada setiap ayat yang mengandung kata identik, sinonim, entitas nama dan sekuen kata.
4. Sistem melakukan perhitungan nilai kesamaan semantik hasil *alignment* yang telah dilakukan masing-masing persamaan *word similarity dan contextual similarity* kemudian melakukan perhitungan kesaman semantik dengan *single proportion* dan *separate proportion*
5. Sistem akan melakukan evaluasi terhadap hasil perhitungan kesamaan semantik dengan perhitungan nilai korelasi Pearson. Korelasi Pearson akan mengevaluasi *gold standard* dengan hasil STS yang dihitung oleh sistem sebelumnya yaitu *single proportion dan separate proportion*. Pada tahapan ini sistem akan menghasilkan output berupa nilai korelasi dalam bentuk persentasi korelasi antara kedua data tersebut.

3. Pembahasan

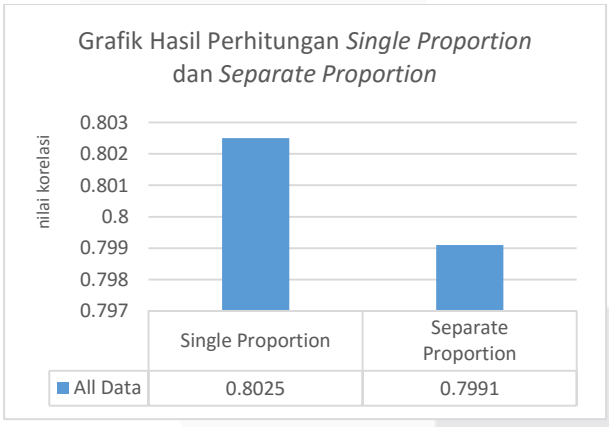
Sistem dibangun sesuai perancangan sistem. Sistem diimplementasikan untuk mengukur kesamaan semantik pada dataset kisah Nabi Abraham dalam Alkitab dan Al-quran.

3.1. Skenario Pengujian

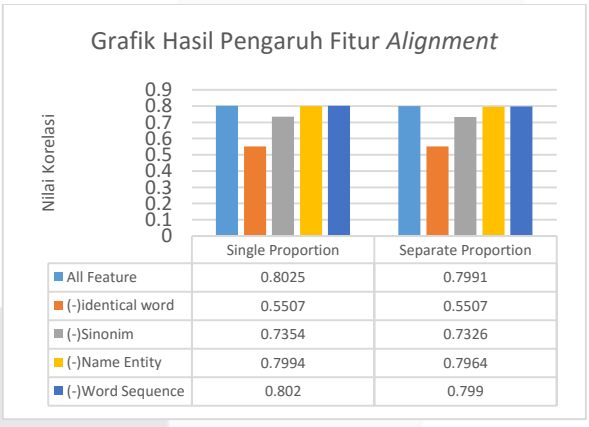
1. Membandingkan dan menganalisis hasil korelasi sistem dengan hasil korelasi penelitian lain. Penelitian lain yang dimaksudkan adalah penelitian yang menggunakan metode *word alignment* dan dengan penggunaan dan pengembangan fitur-fiturnya dan menggunakan dataset yang berbeda.
2. Menganalisis pengaruh perhitungan *alignment word similarity* dan *contextual similarity* terhadap nilai korelasi sistem.
3. Menganalisis pengaruh metode perhitungan *single proportion* dan *separate proportion* terhadap nilai korelasi sistem.
4. Menganalisis pengaruh setiap fitur *alignment* yang diterapkan terhadap nilai korelasi. Pada pengujian ini, fitur *alignment* akan diuji satu per satu untuk mengetahui pengaruh nya terhadap nilai korelasi.

3.2. Hasil Pengujian

Berikut merupakan hasil pengujian yang telah dilakukan.



Gambar 2 Hasil Pengaruh fitur Alignment



Gambar 3 Perhitungan single proportion dan separate proportion

3.3. Analisis

3.3.1. Analisis Perbandingan hasil penelitian dengan Baseline

Hasil perhitungan penelitian ini akan dibandingkan dengan hasil perhitungan penelitian lain yang mempunyai topik yang sama yaitu, *Semantic Textual Similarity* (STS). Berikut beberapa penelitian yang dijadikan sebagai pembanding dan terurut dari yang terkecil hingga yang terbesar.

Tabel 1 Korelasi Penelitian SemEval

Korelasi	Judul Penelitian SemEval
70,0	Yao et al.(2013)
73,4	*Back to basic for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidance*
77,7	Madnani et al.(2013)

77,8	NUIG-UNLP at SemEval Task 1 : *Soft Alignment and Deep Learning for Semantic Textual Similarity* SemEval 2016
79,42	ExB Themis: *Extensive Feature Extraction from Word Alignment for Semantic Textual Similarity* Semeval 2015
80,15	DLS@CU: *Sentence Similarity from Word Alignment and Semantic Vector Composition SemEval 2015*
89,8	BIT NeRoSim: *A System for Measuring and Interpreting Semantic Textual Similarity* SemEval 2015

Hasil perhitungan penelitian tugas akhir ini mendapatkan nilai korelasi *single proportion* sebesar 0,8025 dan *separate proportion* sebesar 0,7991. Berdasarkan data pada Tabel 1 dapat diketahui bahwa hasil penelitian yang telah dilakukan dengan menerapkan *word alignment* pada dataset kisah Nabi Abraham pada Alkitab dan Al-quran tidak jauh berbeda dengan hasil penelitian SemEval. Apabila diurutkan pada hasil penelitian di atas maka hasil penelitian tugas akhir ini berapa pada peringkat kedua untuk nilai *single proportion*. Dengan demikian dapat dilihat bahwa sistem yang dibangun cukup baik untuk mengukur tingkat kesamaan teks karena dengan menggunakan konsep dan metode yang sama yaitu *word alignment* dengan dataset yang berbeda tetap menghasilkan korelasi yang cukup tinggi.

3.3.2. Analisis Pengaruh Metode Perhitungan *Single Proportion* dan *Separate Proportion*

Berdasarkan hasil pengujian sistem yang terdapat pada Gambar 2 diketahui bahwa hasil metode perhitungan *single proportion* lebih tinggi dibandingkan metode perhitungan *separate proportion*. Hal ini dikarenakan hasil perhitungan *similarity* dengan *single proportion* lebih mendekati skor *gold standard*. Berdasarkan persaman perhitungan single proportion, dapat dilihat bahwa hasil perhitungan cenderung konstan, karena hanya membutuhkan jumlah kata yang ter-*align* kemudian dibagi dengan total string pada pasangan ayat. Berbeda dengan perhitungan *separate proportion* yang memperhatikan proporsi kata yang ter-align terhadap jumlah kata pada pasangan ayatnya. Jumlah kata yang ter-align dan jumlah kata pada setiap ayat sangat mempengaruhi hasil perhitungan.

3.3.3. Analisis Perhitungan *Pendekatan Word Similarity* dan *Contextual Similarity*

Pengujian ini dilakukan untuk mengetahui pengaruh perhitungan *word similarity* dan *contextual similarity* pada hasil korelasi. Proses perhitungan *alignment word similarity* terdiri dari *align identical word* dan *align* sinonim. Sedangkan perhitungan *contextual similarity* terdiri dari yaitu *align name entity* dan *align word sequence*. Kedua bagian *alignment* ini diterapkan secara bergantian pada sistem untuk mengetahui pengaruh dari masing-masing bagian alignment terhadap nilai korelasi.

Tabel 2 Hasil Pengujian *Word Similarity* dan *Contextual Similarity*

Feature *Alignment*	*Single Proportion*	*Separate Proportion*
Full Feature	0,8025	0,7991
Word Similarity	0,8011	0,7972
Contextual Similarity	0,2755	0,2755

Berdasarkan Tabel 2 dan Gambar 3, dapat dilihat bahwa penggunaan komponen *alignment* dalam *word similarity* lebih tinggi dari *contextual similarity* baik pada *single proportion* maupun *separate proportion*. Hal ini dapat terjadi karena kata yang teridentifikasi align lebih banyak ditemukan pada *word similarity*. Sedangkan kata yang ter-*align* pada *contextual similarity* hanya sedikit. Hal ini dipengaruhi oleh karakteristik pasangat ayat pada dataset yang lebih banyak mengandung kata identik, dan sinonim dibandingkan dengan kata yang mengandung entitas nama maupun sekuen kata atau *word sequence*. Sehingga hal tersebut sangat mempengaruhi nilai korelasi yang dihasilkan sistem.

3.3.4. Analisis Pengaruh Fitur *Alignment*

Pada Gambar 3 dapat dilihat bahwa setiap fitur mempengaruhi nilai korelasi. Menghilangkan atau tidak menyertakan salah satu fitur membuat nilai korelasi menurun. Berikut daftar selisih nilai korelasi antara penggunaan semua fitur atau menghilangkan salah satu fitur.

Tabel 3 Selisih nilai Korelasi saat setiap fitur tidak diikutsertakan

Feature *Alignment*	*Single Proportion*	*Separate Proportion*
Full Feature	0,8025	0,7991
(-) Align Identical Word	0,5507	0,5507
(-) Align Synonim	0,7354	0,7326
(-) Align Name Entity	0,7994	0,7964
(-) Align Word Sequence	0,802	0,799

Berdasarkan Gambar 3 dan Tabel 3 dapat dilihat bahwa semua fitur mempengaruhi nilai korelasi, namun tidak semua fitur yang mempunyai pengaruh besar. Dari tabel 3 dapat disimpulkan bahwa fitur yang paling mempengaruhi nilai korelasi yaitu fitur yang memiliki paling banyak penurunan nilai korelasi adalah

identical word. Tanpa melibatkan fitur ini, nilai korelasi yang dihasilnya jauh lebih rendah dari nilai korelasi dengan melibatkan seluruh fitur alignment. Nilai korelasi yang dihasilkan tanpa melibatkan fitur *identical word* mengalami penurunan sebesar 0,2518 untuk perhitungan dengan *single proportion* dan 0,2484 untuk perhitungan dengan *separate proportion*. Penurunan nilai korelasi yang cukup besar ini karena fitur *identical word* ini melakukan pengecekan terhadap kata yang identik. Pada dataset yang digunakan dalam penelitian ini pasangan ayat mengandung lebih banyak kata yang identik daripada kata sinonim, entitas nama dan word sequence. Sehingga banyak kata yang ter-*align* dengan menggunakan fitur *identical word* dan mempengaruhi nilai korelasi. Kemudian pada tabel 3 dapat dilihat jika fitur sinonim tidak digunakan maka terjadi penurunan nilai korelasi sebesar 0,0671 pada *single proportion* dan sebesar 0,0665 pada *separate proportion*. Penurunan nilai korelasi karena *align* sinonim melakukan pengecekan setiap kata dalam kalimat dengan mempertimbangkan konteks kalimat tersebut. Sehingga *align* sinonim ini juga cukup berpengaruh pada perhitungan nilai korelasi meskipun kata yang teridentifikasi *align* tidak sebanyak kata yang teridentifikasi *identical word*. Sedangkan untuk fitur-fitur yang lain yaitu *align name entity* dan *align word sequence*, tidak terlalu berpengaruh terhadap nilai korelasi. Berdasarkan data dalam 3, nilai korelasi yang dihasilkan tanpa melibatkan salah satu fitur tersebut mengalami penurunan yang tidak jauh dengan nilai korelasi dengan menerapkan seluruh fitur *alignment*. Hal ini dapat terjadi karena, jumlah kata align yang terdapat dalam dataset tidak cukup banyak. Pada *align name entity* pasangan ayat dataset yang mengandung kesamaan entitas hanya sedikit. Sehingga kata-kata yang teridentifikasi *align* juga hanya sedikit, yang mengakibatkan tidak berpengaruh besar terhadap nilai korelasi. Dalam dataset, hanya terdapat lima pasang kata yang teridentifikasi *align*. Begitu juga dengan *word sequence* yang paling sedikit mempengaruhi nilai korelasi. Hal ini dapat terjadi karena tidak ditemukan pasangan sekuen kata pada pasangan ayat.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan dalam penelitian ini dapat disimpulkan:

1. Implementasi *similarity* dengan *word alignment* pada Kisah Nabi Abraham dalam Alkitab dan Al-quran mempunyai nilai korelasi tertinggi 0,8025 dengan menggunakan perhitungan kesamaan *single proportion*. Metode *word alignment* pada penelitian ini bekerja dengan cukup baik dapat dilihat dari sistem yang mampu mengidentifikasi kata pada setiap fitur dan dapat digunakan dengan jenis dataset yang berbeda.
2. Dengan menggunakan perhitungan *single proportion* menghasilkan nilai korelasi yang lebih tinggi dibandingkan dengan perhitungan *separate proportion*. Hal ini karena pengaruh proporsi kata terhadap pasangan ayat.
3. Fitur *alignment* yang memiliki pengaruh tertinggi pada nilai korelasi adalah *alignment* dengan *identical word* yang terdapat pada kelompok word similarity. Hal ini karena sistem lebih banyak mengidentifikasi kata yang identik dibandingkan dengan kata secara kontekstual.

4.2. Saran

Saran penulis yang diperlukan untuk pengembangan Tugas Akhir ini adalah:

1. Melakukan penelitian STS berbahasa Indonesia dengan metode yang lain untuk membandingkan hasilnya dan dapat menemukan metode yang paling optimal.
2. Menambahkan daftar kata sinonim yang dibangun, sehingga semakin banyak dataset yang dapat menggunakan sistem tanpa menurunkan performansi sistem.
3. Mempertimbangkan fitur-fitur baru word alignment seperti Hiponim, Hipernim dll untuk dataset berbahasa Indonesia.
4. Memperbanyak dataset dari kisah-kisah Nabi lainnya pada Alkitab dan Alquran yang memiliki kemiripan.

5. Daftar Pustaka

[1]	Md Arafat Sultan, Steven Bethard, and Tamara Sumner, "Sentence Similarity from Word Alignment and Semantic Vector Composition," *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 148-153, June 2015.
[2]	Md Arafat Sultan, Steven Bethard, and Tamara Sumner, "Sentence Similarity from Word Alignment," *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 241-246, August 2014.
[3]	Md Arafat Sultan, Steven Bethard, and Tamara Sumner, "Back to Basics for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidence," in *Transactions of the Association for Computational Linguistics*, 2014, pp. 219–230.

[4]	Tafsir Ibnu Katsir. Tafsir Ibnu Katsir. Pustaka Al-Kautsar, 2011.
[5]	Hak Cipta Badan Pengembangan dan Pembinaan Bahasa. Kamus Besar Bahasa Indonesia.