

## IMPLEMENTASI ALGORITMA SENTENCE SIMILARITY TERHADAP CHATBOT SEPUTAR AMIKOM

<sup>1</sup>Dicky Andhika Rizaldi, <sup>2</sup>Galih Adhi Kuncoro Rosyad, <sup>3</sup>Anggit Dwi Hartanto

<sup>1,2,3</sup>Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

Email: <sup>1</sup>[dicky.1428@students.amikom.ac.id](mailto:dicky.1428@students.amikom.ac.id), <sup>2</sup>[galih.1997@students.amikom.ac.id](mailto:galih.1997@students.amikom.ac.id), <sup>3</sup>[anggit@amikom.ac.id](mailto:anggit@amikom.ac.id)

DOI: <https://doi.org/10.46880/jmika.Vol4No1.pp10-14>

### ABSTRAK

Sebuah informasi merupakan suatu hal yang sangat penting terhadap manusia dalam mengetahui masalah ataupun situasi yang ada. Kecepatan dalam mendapatkan sebuah informasi memudahkan seseorang untuk menyelesaikan maupun mengantisipasi dalam kondisi yang ada. Sebuah informasi yang akurat sangat dibutuhkan dalam kehidupan manusia, dengan adanya perkembangan teknologi yang semakin berkembang penyampaian informasi dapat disebarkan dengan berbagai macam cara. Salah satunya adalah dengan *Chatbot* yang menggunakan algoritma *Sentence Similarity*, dimana dengan algoritma ini dapat mengumpulkan informasi berdasarkan pengelompokan kata maupun kalimat secara akurat, karena algoritma ini memiliki tingkat keakuratan yang sangat tinggi. Selain itu banyaknya pola kata/kalimat yang terkumpul dan telah disimpan membuat chatbot memiliki keakuratan yang lebih baik.

**Kata Kunci:** *Chatbot, Sentence Similarity, Informasi*

### PENDAHULUAN

Dalam kehidupan Informasi digunakan dalam mengetahui masalah atau situasi yang ada. Kecepatan dalam mendapatkan informasi memudahkan seseorang untuk menyelesaikan/ mengantisipasi sebuah kondisi yang ada. Informasi akurat sangat dibutuhkan dalam kehidupan. Di era ini peran teknologi dengan informasi sangatlah erat. Berbagai media dalam penyampaian teknologi semakin berkembang dari media cetak hingga sosial media. Tidak hanya dengan media tersebut informasi disebarluaskan penerapan lain terhadap perkembangan teknologi di bidang informasi salah satunya dengan Chatbot. Chatbot itu sendiri merupakan aplikasi chat dengan menambahkan Kecerdasan Buatan (*Artificial Intelligence*). Dimana penambahan Artificial Intelligence mampu membuat orang yang mengirim chat ke aplikasi itu dapat menerima jawaban dari aplikasi tanpa bantuan manusia.

Namun dalam penerapannya penggunaan chatbot mengalami banyak polemik yang ada seperti tidak akurasi terhadap informasi yang ada. Selain itu dari penggunaannya banyak diantara kita yang terbiasa untuk menanyakan sebuah hal yang kita tidak ketahui ke orang lain/orang yang memahami masalah tertentu. Disinilah masalah terjadi yakni terjadinya transmisi dari pola yang manual (sistem taya jawab) berubah menjadi chat.

Penggunaan *Artificial Intelligence* dalam chatbot tergolong NLP (*Natural Language*

*Processing*) Sedangkan penggunaannya dapat dilakukan dengan berbagai metode/algoritma yang ada salah satunya dengan menggunakan Algoritma *Sentence Similarity*. Algoritma *Sentence Similarity* merupakan algoritma yang bekerja dengan mencocokkan sebuah kata/kalimat terhadap sebuah data yang telah ada. Penggunaan algoritma *Sentence Similarity* memiliki keakuratan yang sangat tinggi (Deddy Winarsono, 2009). Penggunaan algoritma ini nantinya akan menanggulangi masalah terhadap keakuratan sebuah informasi. Nantinya jika diterapkan chatbot dalam sebuah lembaga/tempat akan memudahkan dalam mengatasi minimnya informasi terhadap lembaga/tempat tersebut.

Nantinya penggunaan chatbot dengan algoritma *Sentence Similarity* memanfaatkan tingkat akuratnya dari algoritma tersebut serta dalam *Sentence Similarity* terdapat 2 pencocokan kata sebagai berikut:

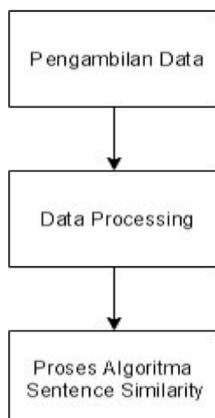
1. Kemiripan Sintatik antar kata adalah pencocokan berdasarkan pengambilan sebuah kata, cara kerja algoritma disini adalah dengan menghitung tingkat ketepatan berdasarkan total transformasi antar kata, yaitu penambahan, penggantian, dan penghapusan karakter.
2. Kemiripan Semantik antar kalimat adalah pencocokan berdasarkan pengambilan sebuah kalimat dengan cara menghitung ketepatan dari arti semua kata yang ada pada kalimat lain yang

kemudian akan diambil kalimat yang identik sama secara semantik.

Berdasarkan acuan pada jurnal (Said Al-Fakhri, 2019) terdapat metode penelitian serta hasil pembahasan yang sangat kompleks, serta hasil dari chatbot nantinya akan diimplementasikan pada sebuah aplikasi Line Messenger dimana pengaplikasiannya menggunakan API Line. Pada penelitian (Manalu, 2016) dengan menggunakan algoritma genetika menggunakan Posi Formulation diperoleh hasil kemiripan dokumen sebesar 0.24 sampai 0.11. Sedangkan pada jurnal yang peneliti buat pada bagian hasil dan pembahasan kurang kompleks karena hanya menerangkan dan menjelaskan alur ketika chatbot tersebut bekerja, tidak adanya hasil akhir serta gambar menjadi salah satu kekurangan dari jurnal yang peneliti lakukan. Peneliti (Zhao Jingling, 2014) penggunaan Algoritma *Sentence Similarity* menggunakan pencocokan sebuah kata terdekat. Dimana setiap kata yang diinputkan akan mencari kata yang memiliki kedekatan. Data yang diperoleh dari (Deddy Winarsono, 2009) menunjukkan penggunaan algoritma *Sentence Similarity* memiliki deviasi dengan rentang tertingginya 0.19 hingga terendahnya 2.61 dengan rata-rata deviasi 1.33 data tersebut menunjukkan tingkat keakuratan dari algoritma tersebut. Namun dari semua keakuratan algoritma ini memiliki kekurangan seperti halnya yang dipaparkan dalam (Pantulkar Sravanthi, 2017)

**METODE PENELITIAN**

Metode yang digunakan peneliti untuk menyelesaikan masalah terhadap pembuatan chatbot ditunjukkan pada Gambar 1 berikut:



**Gambar 1.** Metode Pembuatan Chatbot

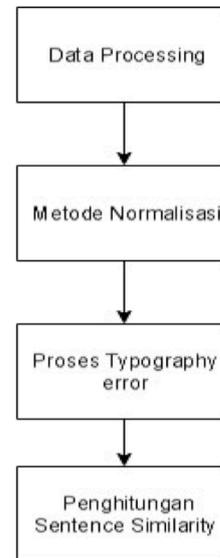
**Pengambilan Data**

Dalam tahapan ini dilakukan pengumpulan data menggunakan metode Observasi. Dimana peneliti terjun langsung dalam pengambilan sebuah data yang

nantinya akan dijadikan sebuah objek. Objek-objek tersebut akan dibagi menjadi beberapa bagian. Bagian-bagian objek tersebut antara lain : objek klu(objek kunci), objek induk

**Data Processing**

Pemrosesan data pada penelitian ini ditunjukkan pada Gambar 2 berikut:



**Gambar 2.** Data Processing

Proses Data diperoleh dari metode yang digunakan, dalam pembuatan chatbot ini menggunakan proses normalisasi dan pencocokan pola.

**Metode Normalisasi**

Proses normalisasi berfungsi sebagai penyesuaian sebuah kalimat sebelum dilakukan pencocokan pola, dalam proses ini terdapat dua sub proses yaitu:

- 1) Typography error, membenarkan setiap ejaan kata maupun kalimat yang salah dalam penulisan, sehingga menggantikannya dengan kata atau kalimat ejaan yang baik dan benar.

No	Input	Hasil
1.	Dmn Ruang Pengjran	Dimana Ruang Pengajaran
2.	Ada berpa jurusan di Amikom	Ada berapa jurusan di Amikom
3.	Dimana Amikom Brada	Dimana Amikom Berada

- 2) Penghitungan Sentence Similarity

Ditahap ini kalimat/kata yang diinputkan akan mengalami tahapan penghitungan kemiripan dengan data yang ada, proses penghitungan melibatkan string (dalam hal ini data yang dicari) dan Bigram. Inti dari Bigram adalah dua rangkaian yang saling berurutan baik itu kalimat maupun huruf, hal demikian juga telah dijelaskan (wikipedia).

Berikut ini merupakan rumus penghitungan kalimat dengan rumus (Said Al-Fakhri, 2019)

$$S = \frac{\text{count}(b1 \in b2) + \text{count}(b2 \in b1)}{\text{count } b1 + \text{count } b2}$$

Penjelasan Variabel yang ada di rumus sebagai berikut:

b1 merupakan bigram yang dihasilkan dari string 1  
 .....(1)

b2 merupakan bigram yang dihasilkan dari string 2  
 .....(2)

(b1 ∈ b2) merupakan operasi yang menandakan anggota b1 juga anggota b2

(b2 ∈ b1) merupakan operasi yang menandakan anggota b2 juga anggota b1

Berikut ini adalah penerapan rumus dengan menggunakan 2 buah kata “Informatika” dan “Informatika”

- String 1 : “Informatika”
- String 2 : “Informatika”
- Bigram 1 : “In, nf, fo, or, rm, ma, at, ti, ik, ka”(10)
- Bigram 2 : “In, nf, fo, or,rr, rm, ma, at, ti, ik, ka”(11)
- (b1 ∈ b2) : “In, nf, fo, or, rm, ma, at, ti, ik, ka”(10)
- (b2 ∈ b1) : “In, nf, fo, or, rm, ma, at, ti, ik, ka”(10)

$$S = (10+10)/(10+11) = 0.95238095238$$

Maka nilai similarity string1 dan string2 adalah 0.95238095238

### 3) Proses Sentence Similarity

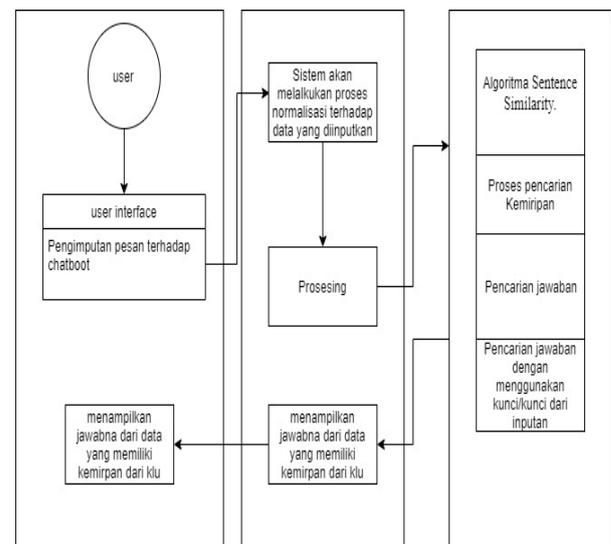
Ditahapan ini data yang diinputkan akan dicocokkan dengan data yang telah disimpan, data yang di inputkan berupa kalimat akan diibaratkan menjadi tag. Tag inilah yang menjadi kunci, sedangkan klu khusus akan menjadi pintu untuk menuju jawaban dari tag.



Gambar 3. Ilustrasi Proses Sentence Similarity

## HASIL DAN PEMBAHASAN

Proses dari penelitian ini ditampilkan pada Gambar 4 berikut:



Gambar 4. User Interface

### Penginputan

Dalam tahap ini terjadi penginputan pesan yang berisi pertanyaan kedalam aplikasi chatbot seputar Amikom. Pesan yang diinputkan tersebut seperti “Kpan tanggl pendaftan maba”

### Normalisasi

Kemudian setelah terjadi penginputan kalimat maka akan terjadi normalisasi berupa membenarkan kata ejaan serta menghilangkan tanda baca (?) atau typography error dari kata yang telah diinputkan akan berubah menjadi “Kapan tanggal pendaftaran maba”. Maka diperoleh sebuah kata tanya yakni “Kapan” dan “Pendaftaran Amikom” sebagai klu

### Algoritma Sentence Similarity

Saat inputan memasuki algoritma akan mengalami beberapa proses sebagai berikut :

**Pencocokan Data**

Inputan akan mengalami pencocokan data, data yang ada tersusun sebagai berikut:

**Tabel 1.** Pencocokan Data

File	Tipe data	Keterangan
Idklu	INT	ID dari klu
Klu	VARCHAR(100)	Klu untuk menjadi penghubung antara inputan dan jawaban
jwb	VARCHAR(100)	Hasil dari klu

Dari pasangan tabel tersebut maka sampel data akan menjadi seperti berikut:

**Tabel 2.** Pasangan Sampel Data

Idklu	Klu	Jwb
01	Letak Amikom	Ringrod utara
02	Rektor Amikom	Prof. M. Suyanto
03	Letak Kanti	Sebelah Gedung 3
04	Letak Ingkubator Amikom	Di Gedung 5
05	Selamat Pagi	Pagi
06	Pendaftaran Amikom	29 oktober

Kemudian inputan tersebut akan mengalami proses pencocokan data dengan mencari klu yang terdapat dalam inputan dan diperoleh pedaftran amikom. Maka selanjutnya klu akan menjadi penghubung dari sebuah jawaban.

**Pengujian Algoritma Sentence Similarity**

Setelah klu diperoleh maka telah ditemukan penghubung untuk jawaban namun untuk memastikan klu tersebut sebagai jawaban yang benar maka akan dilakukan pengujian terhadap klu dengan menggunakan penghubung dalam hal ini kata tanya dan klu

	Subyek	kalimat	Nilai Kemiripan
No	Kata penghubung + Kata Tanya	Kapan Pendaftaran Amikom	-
1	Inputan	Kapan pedaftraan Amikom	1
2		Pendaftan Amikom	0.85
3		Amikom Didaftar	0

Dengan menggunakan rumus (Said Al-Fakhri, 2019) yang telah di terapkan ke sebuah aplikasi chatbot maka nilai kemiripan paling akurat terdapat pada no 1.

**KESIMPULAN**

Algoritma Sentence Similarity dapat digunakan dalam sebuah aplikasi chatbot dengan memanfaatkan kemiripan sebuah objek. Dalam penghitungan Sentence Similarity sebuah kalimat tanda baca akan mempengaruhi hasil dari penghitungannya.

**DAFTAR PUSTAKA**

Manalu, D. R. (2016). Pengujian tingkat kemiripan skripsi mahasiswa dengan algoritma genetika menggunakan posi formulation. *Jurnal METHODIKA*, 2 (2), 175-182.

Mas`udia, P. E., Atmadja, M. D., & Mustafa, L. D. (2017). Information Retrieval Tugas Akhir dan Perhitungan Kemiripan Dokumen mengacu pada Abstrak Menggunakan Vector Space Model. *Jurnal SIMETRIS*, 8 (1), 355-362

Pantulkar, S. D. B. (2017). Semantic similarity between sentences. *International Research Journal of Engineering and Technology (IRJET)*, 04, 156-161.

Pratama, B.P. dan Pamungkas, S.A. (2016). Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks. *Jurnal LOG!K@*, 4 (2), 131-143.

Said Al-Fakhri, H. U. (2019). Aplikasi Chatbot Informasi Kampus Polban Menggunakan Aplikasi LINE Messenger. *IRWNS*. Bandung.

Saputra, J.S., Fachrurrozi, M., & Yunita. (2017). Peringkasan Teks Berita Berbahasa Indonesia Menggunakan Metode Latent Semantic Analysis (LSA) dan Teknik Steinberger & Jezek. *ARS*. Sriwijaya.

Sugiyanto, Surarso, B., & Sugiharto, A. (2014). Analisa Performa Metode Cosine dan Jacard pada Pengujian Kesamaan Dokumen. *Jurnal Masyarakat Informatika*, 5 (10), 1-8.

Wikipedia. (n.d.). Retrieved 11 14, 2019, from <https://en.m.wikipedia.org/wiki/Bigram>

- Winarsono, D., Siahaan, D., & Yuhana, U. L. (2009).  
Sistem penilaian otomatis kemiripan kalimat  
menggunakan syntactic-semantic similarity pada  
sistem e-learning. *KURSOR*. Surabaya.
- Zhao Jingling, Z. H. (2014). Sentence Similarity  
Based on Semantic Vector Model . *Ninth  
International Conference on P2P* (pp. 499-503).  
Beijing, Cina: CPS.