# ALGORITMA JARO WINKLER UNTUK MENGUKUR SIMILARITY BERITA ONLINE

**SKRIPSI**


disusun oleh

**Teguh Efriyanto**

**18.11.2388**



**PROGRAM SARJANA**
**PROGRAM STUDI INFORMATIKA**
**FAKULTAS ILMU KOMPUTER**
**UNIVERSITAS AMIKOM YOGYAKARTA**
**YOGYAKARTA**


**2022**

# ALGORITMA JARO WINKLER UNTUK MENGUKUR SIMILARITY BERITA ONLINE

**SKRIPSI**

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

**Teguh Efriyanto**

**18.11.2388**

**PROGRAM SARJANA**
**PROGRAM STUDI INFORMATIKA**
**FAKULTAS ILMU KOMPUTER**
**UNIVERSITAS AMIKOM YOGYAKARTA**
**YOGYAKARTA**
**2022**

## PERSETUJUAN

## SKRIPSI

**ALGORITMA JARO WINKLER UNTUK MENGUKUR SIMILARITY BERITA ONLINE**

yang dipersiapkan dan disusun oleh

**Teguh Efriyanto**

**18.11.2388**

telah disetujuai oleh Dosen Pembimbing Skripsi

pada tanggal 08 Oktober 2021

**Dosen Pembimbing,**

**Mardhiya Hayaty, S.T., M.Kom.**

**NIK. 190302108**

## PENGESAHAN

## SKRIPSI

**ALGORITMA JARO WINKLER UNTUK MENGUKUR SIMILARITY BERITA ONLINE**

yang dipersiapkan dan disusun oleh

**Teguh Efriyanto**

**18.11.2388**

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Maret 2022

**Susunan Dewan Penguji**

| Nama Penguji | Tanda Tangan |
|---|---|
| **Heri Sismoro, M.Kom**<br>**NIK. 190302057** | |
| **Yuli Astuti, M.Kom**<br>**NIK. 190302146** | |
| **Mardhiya Hayaty, S.T., M.Kom.**<br>**NIK. 190302108** | |

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal ... Maret 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, M.Kom**
**NIK. 190302096**

# PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 22 Maret 2022

Teguh Eriyanto

NIM. 18.11.2388

# MOTTO

"Tidak ada kesuksesan tanpa kerja keras. Tidak ada keberhasilan tanpa kebersamaan. Tidak ada kemudahan tanpa doa."

( Ridwan Kamil )

"Sukses adalah guru yang buruk. Sukses menggoda orang yang tekun berpikir bahwa mereka tidak bisa gagal."

( Bill Gates )

"Angin tidak berhembus untuk menggoyangkan pepohonan, melainkan menguji kekuatan akarnya."

( Ali bin Abi Thalib )

# PERSEMBAHAN

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan lancar. Selama pengerjaan skripsi ini banyak sekali bantuan, dukungan, dan do'a dari berbagai pihak, sehingga penulis sampaikan rasa terimakasih sedalam-dalamnya kepada :

1. Kedua orang tua dan saudara-saudara yang telah memberikan do'a dan dukungan kepada penulis secara moril maupun materil sehingga skripsi ini dapat selesai dengan baik.
2. Dosen pembimbing, Ibu Mardhiya Hayaty, S.T., M.Kom. yang telah membimbing dalam penyusunan skripsi ini hingga selesai.
3. Calon istri, Rifa Indriyani yang telah senantiasa memberikan motivasi dan membantu dalam segala hal terutama telah bersedia menyediakan alat untuk saya bisa menyelesaikan skripsi ini.
4. Teman-teman Informatika angkatan 2018 khususnya kelas IF-09 yang senantiasa memberi motivasi dan berjuang bersama selama menjadi mahasiswa.
5. Semua pihak yang telah banyak membantu dalam penyusunan skripsi ini yang tidak bisa penulis sebutkan satu-persatu.

# KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT, karena rahmat dan hidayah-Nya sehingga skripsi yang berjudul "Algoritma Jaro Winkler Untuk Mengukur Similarity Berita Online" ini dapat selesai dengan baik sebagai salah syarat untuk dapat menempuh ujian sarjana pada Fakultas Ilmu Komputer (FIK) program studi Informatika di Universitas Amikom Yogyakarta.

Pengerjaan skripsi ini tidak lepas dari bantuan beberapa pihak. Oleh karena itu, penulis ingin menyampaikan rasa hormat dan terimakasih kepada :

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya Dhuhita, M.Kom. selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
3. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang telah membantu dan membimbing penulis dengan saran dan waktunya.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan sehingga saran dan kritik yang membangun sangat penulis harapkan dan penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi.

Yogyakarta,

Teguh Efriyanto

# DAFTAR ISI

# DAFTAR TABEL

# DAFTAR GAMBAR

# INTISARI

Perkembangan zaman di era industri 4.0 telah membuat media berita online menjadi sangat penting di kehidupan sehari-hari masyarakat untuk mencari sumber informasi. Hal ini berdampak pada wartawan media berita online untuk dapat mencari informasi berita yang cepat dan akurat, tidak menutup kemungkinan wartawan yang bekerja secara bersama-sama di lapangan melakukan tindakan plagiarisme ke wartawan lain atau mengambil bahan berita dari situs media berita lain dan menggunakannya untuk dimuat dimedianya tanpa mencantumkan sumbernya. sehingga untuk mengatasi hal tersebut dibutuhkan sebuah algoritma yang dapat mengukur similarity berita online untuk mengetahui tingkat plagiarisme antar berita online.

Salah satu algoritma yang dapat digunakan untuk mengetahui tingkat plagiarisme dengan menghitung nilai similarity adalah algoritma jaro winkler, Pada algoritma jaro winkler semakin tinggi nilai jaro winkler pada dua string, maka semakin mirip string tersebut. Nilai 1 menandakan kesamaan antar string dan nilai 0 menandakan ketidaksamaan antar string. Data yang digunakan adalah data isi berita dari 20 situs media berita online daerah Kalimantan Tengah, yang diperoleh dengan proses scraping yang disorting dengan Google Custome Search Engine dengan memanfaatkan Custome Search JSON API dan menggunakan keyword untuk mendapatkan berita dengan topik yang sama dan dilakukan proses text preprocessing. Dan pada metode perancangan sistem pada penelitian ini menggunakan metode prototype.

Pada penelitian ini menghasilkan nilai rata-rata similarity antar berita online sebesar 74,49% dengan data sebanyak 55 data berita, dimana diperoleh 43 data berita dengan tingkat plagiarisme berat dan 12 data berita dengan tingkat plagiarism sedang dan terdapat kelemahan pada algoritma jaro winkler dalam menghitung nilai similarity pada data yang diperoleh, yang mana terdapat beberapa data yang tidak terdeteksi yang seharunya tingkat plagiarismenya berat namun tidak berat dan sebaliknya.

**Kata Kunci:** Jaro Winkler, Similarity, Plagiarisme, Text Preprocessing, Berita Online

## *ABSTRACT*

*The development of the era in the industrial era 4.0 has made online news media very important in people's daily lives to find sources of information. This has an impact on online news media journalists to be able to find news information quickly and accurately, it is possible for journalists who work together in the field to commit plagiarism to other journalists or take news material from other news media sites and use it to be published in their media. without citing the source. so that to overcome this we need an algorithm that can measure the similarity of online news to determine the level of plagiarism between online news.*

*One of the algorithms that can be used to determine the level of plagiarism by calculating the similarity value is the Jaro Winkler algorithm. In the Jaro Winkler algorithm, the higher the Jaro Winkler value on two strings, the more similar the strings are. The value 1 indicates the similarity between the strings and the value 0 indicates the inequality between the strings. The data used is news content data from 20 online news media sites in the Central Kalimantan area, which is obtained by a scraping process which is sorted by Google Custome Search Engine by utilizing the Custome Search JSON API and using keywords to get news with the same topic and text preprocessing is carried out. . And the system design method in this study uses the prototype method.*

*In this study, the average value of similarity between online news was 74.49% with 55 news data, which obtained 43 news data with severe plagiarism levels and 12 news data with moderate plagiarism levels and there were weaknesses in the Jaro Winkler algorithm in calculating the similarity value in the data obtained, where there are some undetected data which should have a heavy plagiarism level but not heavy and vice versa.*



***Keyword:*** *Jaro Winkler, Similarity, Plagiarism, Text Preprocessing, Online News*