PENGUKURAN TINGKAT KEMIRIPAN DOKUMEN MENGGUNAKAN ALGORITMA JARO-WINKLER DAN *ENHANCED CONFIX STRIPPING STEMMER*

SKRIPSI

Diajukan untuk memenuhi sebagai persyaratan mendapatkan gelar Strata Satu
Program Studi Informatika

Disusun Oleh:
Anthony Juan Christian
M0512006

PROGRAM STUDI INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA
2017

SKRIPSI

PENGUKURAN TINGKAT KEMIRIPAN DOKUMEN MENGGUNAKAN ALGORITMA JARO-WINKLER DAN *ENHANCED CONFIX STRIPPING STEMMER*

Disusun oleh :
ANTHONY JUAN CHRISTIAN
NIM. M0512006

Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji pada tanggal : 04 Januari 2017

Pembimbing I	Pembimbing II
Dr. Wiranto, M.Kom.,M.Cs.	Abdul Aziz, S.Kom., M.Cs.
NIP. 19661230 199302 1 001	NIP. 19810413 200501 1 001

SKRIPSI

PENGUKURAN TINGKAT KEMIRIPAN DOKUMEN MENGGUNAKAN ALGORITMA JARO-WINKLER DAN *ENHANCED CONFIX STRIPPING STEMMER*

Disusun oleh :
ANTHONY JUAN CHRISTIAN
NIM. M0512006

Skripsi ini telah dipertahankan di hadapan dewan penguji
pada tanggal : 04 Januari 2017

Susunan Dewan Penguji

1. Dr. Wiranto, M.Kom.,M.Cs. ( )
NIP. 19661230 199302 1 001
2. Abdul Aziz, S.Kom., M.Cs. ( )
NIP. 19810413 200501 1 001
3. Ristu Saptono, S.Si.,M.T. ( )
NIP. 19790210 200212 1 001
4. Winarno, S.Si, M.Eng. ( )
NIP. 19820520 200604 1 001

Disahkan Oleh
Kepala Program Studi Informatika

Drs. Bambang Harjito, M.App.Sc., PhD
NIP. 19621130 199103 1 002

MOTTO

“Tugas akhir yang baik adalah yang cepat selesai dan bermanfaat bagi orang lain”

Skripsi ini dipersembahkan untuk :

Universitas Sebelas Maret , yang telah menjadi wadah saya untuk berkembang menjadi seseorang yang akan memberi dampak positif kepada dunia.

KATA PENGANTAR

Segala puji syukur bagi Tuhan Yesus Kristus yang telah melimpahkan kasih, berkat serta hikmat-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul "Pengukuran Tingkat Kemiripan Dokumen Menggunakan Algoritma Jaro-Winkler dan Enhanced Confix Stripping Stemmer".

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu proses pengerjaan Tugas Akhir ini sehingga dapat berwujud sebagaimana yang diharapkan, yaitu kepada :

1. Bapak Walfred Pohan ,Ibu Lince Gultom, dan Keluarga yang tak henti-hentinya memberikan dorongan motivasi bagi penulis
2. Bapak Wiranto, M.Kom.,M.Cs.dan Bapak Abdul Aziz, S.Kom., M.Cs. selaku dosen pembimbing I dan pembimbing II atas ilmu yang diberikan, bimbingan, kebaikan serta kesabaran kepada penulis selama pelaksanaan Tugas Akhir.
3. Enya Blanco dan Keluarga yang tak bosan-bosannya menemani keseharian penulis
4. Keluarga Besar Informatika UNS 2012, Keluarga Besar Kost Griya Khansa (Gori ,Erick, Billy, Abib, Naufal, Jody, Irfan, Hasan, Uzlif, Edo, Dito, Abel, Bahir, Bagus, Bagas, Arinto, Rizal, Dhidit) dan Keluarga Lingsir Wengi (Alfi,Komeng,Rhesa,Yonathan,Adi) yang senantiasa mengisi hari-hari penulis di masa perkuliahan dan membantu penulis dalam pelaksanaan Tugas Akhir.
5. Keluarga Besar Awjackass (Dwiki, Syarif, Ogi, Bobby, Ulil, Hazmi), serta rekan-rekan semasa sekolah menengah yang masih mengingat saya yang tak bisa saya sebutkan satu-satu, yang mau memberikan support jauh-jauh dari Jakarta.

Surakarta, 04 Januari 2017

Penulis

PENGUKURAN TINGKAT KEMIRIPAN DOKUMEN MENGGUNAKAN ALGORITMA JARO-WINKLER DAN *ENHANCED CONFIX STRIPPING STEMMER*

ANTHONY JUAN CHRISTIAN

Program Studi Informatika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Sebelas Maret

ABSTRAK

Plagiarisme merupakan sebuah tindakan mengambil atau menjiplak karya seseorang yang bisa berupa dokumen atau hal lainnya dan menyatakan sebagai milik sendiri, tanpa menyantumkan sumber pengambilan informasi yang bersangkutan. Berdasarkan dari kejadian tersebut, dibutuhkan sebuah sistem yang berfungsi untuk melakukan pendeteksian plagiarisme dokumen teks dengan cara mengukur kecocokan antara dokumen melalui metode similaritas antar string pada dokumen. Algoritma Jaro-Winkler merupakan salah satu algoritma pengukuran similaritas berbasis string yang memperhatikan struktur dari pada kata yang akan di bandingkan, sehingga sesuai bila diterapkan pada dokumen. Data yang akan dipakai dalam pengukuran diambil melalui diglib.uns.ac.id sejumlah 35 data yang berupa dokumen abstrak skripsi S1 Informatika UNS. Pada tahap *preprocessing* digunakan teknik stemming dengan algoritma *Enhanced Confix Stripping (ECS) Stemmer* dengan tujuan meningkatkan keakuratan pencocokan string. Dari lima pengujian yang dilakukan, diketahui nilai similaritas dokumen yang paling memberikan perbedaan signifikan terhadap dua metode yang diuji terletak di pengujian kelima yaitu pemotongan dokumen uji sebanyak 70%, dengan rata-rata persentase yang dihasilkan sebesar 31,27% (dengan ECS) dan 28,64% (tanpa ECS).

Kata kunci: *Enhanced Confix Stripping (ECS) Stemmer*, Jaro-winkler, Plagiarime, Similaritas, *String-based*

DOCUMENT SIMILARITY MEASURE USING JARO-WINKLER ALGORITHM AND ENHANCED CONFIX STRIPPING STEMMER

ANTHONY JUAN CHRISTIAN

Study Program Informatics, Faculty of Mathematics and Natural Sciences,

Sebelas Maret University

ABSTRACT

Plagiarism is an act of taking or plagiarizing the work of others and recognize it as his own handwork, without giving any information about the actual source. Therefore, a system is needed to perform detection of text plagiarism by measuring the similarity between documents using similarity method. Jaro-Winkler Algorithm is one of the string-based similarity method, which focused on structure of the words between two compared strings. In this research , Jaro-winkler algorithm will be applied to measure the similarity between data trial and 35 abstract documents from informatic major. Jaro-Winkler Algorithm will be combined with Enhanced Confix Stripping (ECS) Stemmer Algorithm in the preprocessing stage, to improve the accuracy. based on the five tests which have been performed, the most significant result from these two approaches is on the fifth test which document has been cut by 70%, with average percentage 31,27% (with ECS) dan 28,64% (without ECS).

Keywords*: Enhanced Confix Stripping (ECS) Stemmer*, Jaro-winkler, Plagiarism, similarity, String-based.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN