PENGELOMPOKAN ARTIKEL BERBAHASA INDONESIA DENGAN MENGGUNAKAN REDUKSI FITUR INFORMATION GAIN THRESHOLDING DAN K-MEANS

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:
Novia Agusvina
NIM: 145150201111108

PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

PENGELOMPOKAN ARTIKEL BERBAHASA INDONESIA DENGAN MENGGUNAKAN REDUKSI FITUR INFORMATION GAIN THRESHOLDING DAN K-MEANS

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :
Novia Agusvina
NIM: 145150201111108

Skripsi ini telah diuji dan dinyatakan lulus pada
17 Januari 2018
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I	Dosen Pembimbing II
Indriati, S.T, M.Kom	Nurudin Santoso, S.T., M.T
NIP:19831013 201504 2 002	NIP: 197409162000121001

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP:19710518 200312 1 001

IDENTITAS TIM PENGUJI

PENGUJI 1

Nama : Lailil Muflikhah, S.Kom, M.Sc

Kantor : FILKOM UB, Jl. Veteran No 8 (Gedung A, Lt.1, R. A1.5)

Email : Lailil [At] Ub [Dot] Ac [Dot] Id

PENGUJI 2

Nama : Wibisono Sukmo Wardhono, S.T, M.T

Kantor : FILKOM UB, Jl. Veteran No 8 (Gedung C, Lt.1, R. C1.2)

Email : Wibiwardhono [At] Ub [Dot] Ac [Dot] Id

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 17 Januari 2018

Novia Agusvina

NIM: 145150201111108

Novia Agusvina

noviaagusvina@gmail.com
082257251248
Jalan Sukarno Hatta Kartika 1 no 16, Kediri

Informasi Personal

Nama Lengkap	Novia Agusvina
Tempat & Tanggal Lahir	Kediri, 16 November 1996
Skype	noviaagusvina
LINE id	noviagusvina
Jurusan dan Angkatan	Teknik Informatika 2014

Pendidikan

2014-2018 Timur — Universitas Brawijaya — Malang, Jawa

- Mahasiswa di Fakultas Ilmu Komputer, jurusan Teknik Informatika jalur SNMPTN

2012-2014 Timur — SMA Negeri 3 Kediri — Kediri, Jawa

- Jalur akselerasi dengan jurusan IPA

2013-2014 Timur — English First — Kediri, Jawa

- Lulus Level Frontunner dengan Predikat Baik

2009-2012 Timur — SMP Negeri 3 Kediri — Kediri, Jawa

2002-2009 Timur — SD Negeri Banjaran 2 — Kediri, Jawa

Pengalaman Organisasi, Kepanitiaan, dan Bekerja

2014-2016 Timur — Tergabung di AIESEC in Universitas Brawijaya — Malang, Jawa

- Sebagai staff of Talent Management 14/15
- Sebagai Talent Member Cordinatoor Manager of Talent Development 15/16

Feb- Ags 2015 Timur	Tergabung Entrevolution Project AIESEC	Malang, Jawa

- Sebagai OrginizingCommittee Vice President Program and Comunication

2015	Tergabung di National Conference Aiesec	Solo, Jawa Tengah

- Sebagai peserta

2015 Timur	Asisten Praktikum Pemrogaman Lanjut	Malang, Jawa
2015-2016 Timur	Tergabung di “Himpunan Informatika” in Fakultas	Malang, Jawa

- Sebagai staff Pengembangan Sumber Daya Mahasiswa

Februari 2016 Timur	Tergabung di Hometown Project Aiesec Indonesia	Kediri, Jawa

- Sebagai pembicara tentang kesadaran air bersih

2015-2016 Timur	Tergabung di Informatics Radio	Malang, Jawa

- Sebagai anggota Tim Kreatif

2016-2017 Timur	Tergabung di Laboratorium Basis Data FILKOM	Malang, Jawa

- Sebagai Asisten Praktikum
- Sebagai Anggota dari Tim Soal

2016-2017 Timur	Asisten Praktikum Analisis dan Perancangan Sistem	Malang, Jawa
2017-2018	Tergabung di Ruang Belajar Aqil	

- Sebagai anggota Kelompok Riset Sahaja +
- Sebagai Relawan

UCAPAN TERIMAKASIH

Penulis berterimakasih kepada semua pihak yang telah berperan dalam menyusun skripsi ini, diantaranya :

1. Indriati, S.T, M.Kom, selaku dosen pembimbing pertama yang membimbing penulis dalam menyelesaikan penelitian ini.

2. Nurudin Santoso, S.T., M.T, selaku dosen pembimbing kedua yang telah membimbing dalam penulisan skripsi ini.

3. Bapak Wayan Firdaus Mahmudy, S.Si, M.T, Phd. Selaku Dekan Fakultas Ilmu Komputer, Universitas Brawijaya Malang, beserta jajarannya.

4. Tri Astoto Kurniawan, S.T, M.T, Ph.D, selaku Ketua Jurusan Teknik Informatika Universitas Brawijaya.

5. Seluruh dosen jurusan Teknik Informatika yang selama ini memberikan ilmunya kepada penulis

6. Orang tua penulis, Bagus Sunarko dan Nanik Widayani dan kakak

penulis, Susamto Anandani, yang selalu memberikan dukungan terbesar

hingga akhir dalam keadaan apapun.

7. Seluruh sahabat sahabat saya yang telah membantu baik secara dukungan maupun waktu.

8. Teman- teman Ruang Belajar Aqil yang telah meluangkan waktunya untuk berdiskusi.

9. Dan seluruh pihak yang membantu dalam penyusunan penelitian ini.

ABSTRAK

Artikel online merupakan sumber informasi yang banyak tersebar di situs internet. Semakin banyaknya artikel yang tersebar di situs internet, menyulitkan pengguna dalam menemukan artikel yang diinginkan. Salah satu penyedia layanan artikel online adalah Kompas.com. Untuk menghadapi persaingan antar industri media massa, langkah yang dilakukan Kompas.com adalah memberikan fitur yang memudahkan pengguna, seperti fitur rekomendasi artikel terkait. Namun, dalam penerapannya Kompas.com masih kurang maksimal sehingga tetap kalah dengan media massa online lainnya. Pada penelitian ini, peneliti mengimplementasikan metode reduksi fitur Information Gain Thresholding dan K-Means untuk membuat kelompok artikel terkait. Tujuan dari penelitian ini adalah untuk memperbaiki sistem artikel terkait dari Kompas.com.

Data yang digunakan dalam penelitian ini adalah artikel dari Kompas.com dari kategori Lifestyle. Dalam pengimplementasian digunakan bahasa java. Pada tahap awal dilakukan preprocessing untuk mengurangi gangguan dalam data, selanjutnya dilakukan reduksi fitur untuk mengurangi fitur yang digunakan agar proses kebih cepat, kemudian dilakukan pembobotan sebagai dasar untuk menghitung jarak antar dokumen, setelah menemukan nilai jarak awal atau centroid , pengelompokan dapat dilakukan.

Hasil menunjukan bahwa pengelompokan artikel dengan metode Information Gain Thresholding dan K-Means mampu menghasilkan kelompok dokumen yang baik dengan nilai *silhouette coeffiecient* sebesar 0.9595 dan *purity measure* sebesar 0.75 dengan penggunaan 3 cluster dan batas ambang untuk reduksi fitur terbaik adalah 0.04 dengan waktu eksekusi lebih cepat sebanyak 103 menit dibandingkan tanpa reduksi fitur.

Kata kunci: artikel online, reduksi fitur, information gain thresholding, pengelompokan, K-Mean

ABSTRACT

Online articles are a source of information that is widely spread on the internet site. The increasing number of articles on the website makes an activity to find the desired article difficult for the user. A simple example of this online article service provider is Kompas.com. Since competition among mass media industry is getting more difficult, Kompas.com needs to find a way to keep up, one of the ways is by giving a "recommended articles related" feature. However, in its application Kompas.com still less than the maximum so it remains inferior to other online mass media. In this study, researcher implemented two methods called "Information Gain Threshold" and "K-Means" to create a group of related articles and to improve the searching activity in Kompas.com.

The data used in this research is an article taken from Kompas.com in "Lifestyle" category. In the implementation, the programming language used is Java and as for the early stages of preprocessing, to reduce the disturbance in the data, feature reduction is used to give faster processing of the data. Then, it is weighted to do the basic calculation of the distance between documents. Afterward, making the data into group can be done.

The results show that the clustering of articles using Information Gain Threshold and K-Means is good enough, has criteria of silhouette coefficient of 0.9595 *and a purity measure of 0.75 with 3 clusters and 0.04 threshold limit, this conclude that it gives faster execution time in 103 minutes of time compared to without feature reduction.*

Keywords : online articles, feature reduction, information gain thresholding, clustering, K-Means

KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan karuniaNya sehingga skripsi dengan judul "Pengelompokan Artikel Berbahasa Indonesia dengan Menggunakan Reduksi Fitur Information Gain Thresholding dan K-Means" dapat diselesaikan sebagai sebagian persyaratan memperoleh gelar Sarjana Komputer. Penulis juga berterimakasih kepada semua pihak yang telah berperan dalam menyusun proposal ini, diantaranya :

1. Indriati, S.T, M.Kom, selaku dosen pembimbing pertama yang membimbing penulis dalam menyelesaikan penelitian ini.
2. Nurudin Santoso, S.T., M.T, selaku dosen pembimbing kedua yang telah membimbing dalam penulisan skripsi ini.
3. Bapak Wayan Firdaus Mahmudy, S.Si, M.T, Phd. Selaku Dekan Fakultas Ilmu Komputer, Universitas Brawijaya Malang, beserta jajarannya.
4. Tri Astoto Kurniawan, S.T, M.T, Ph.D, selaku Ketua Jurusan Teknik Informatika Universitas Brawijaya.
5. Seluruh dosen jurusan Teknik Informatika yang selama ini memberikan ilmunya kepada penulis
6. Orang tua penulis, Bagus Sunarko dan Nanik Widayani dan kakak penulis, Susamto Anandani, yang selalu memberikan dukungan terbesar hingga akhir dalam keadaan apapun.
7. Seluruh sahabat sahabat saya yang telah membantu baik secara dukungan maupun waktu.
8. Teman- teman Ruang Belajar Aqil yang telah meluangkan waktunya untuk berdiskusi.
9. Dan seluruh pihak yang membantu dalam penyusunan penelitian ini.

Penulis hanya bisa berdoa semoga Allah SWT membalas kebaikan yang semua pihak telah berikan kepada penyusun. Penyusun menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna. Akhir kata kami mengucapkan terimakasih. Kami juga menerima kritik dan saran sebagai perbaikan dimasa-masa yang akan datang.

Malang, 17 Januari 2018

Penulis

noviaagusvina@student.ub.ac.id

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR