JURNAL

IMPLEMENTASI *HIERARCHICHAL CLUSTERING* PADA PENGELOMPOKAN KARYA TULIS ILMIAH (STUDI KASUS : PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS NUSANTARA PGRI KEDIRI)

HIERARCHICHAL CLUSTERING IMPLEMENTATION IN THE CLASSIFICATION OF SCIENTIFIC PAPERS (CASE STUDY: INFORMATICS ENGINEERING STUDY PROGRAM OF UNIVERSITY NUSANTARA PGRI KEDIRI)

Oleh:

ALDILA ANGGUN SASMITA

12.1.03.02.0208

Dibimbing oleh :

1. **Resty Wulanningrum, M.Kom**
2. **Daniel Swanjaya, M.Kom**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2017

Artikel Skripsi
Universitas Nusantara PGRI Kediri

SURAT PERNYATAAN

ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Aldila Anggun Sasmita
NPM	: 12.1.03.02.0208
Telepun/HP	: 08563699982
Alamat Surel (Email)	: aldilasasmita08@gmail.com
Judul Artikel	: Implementasi Hierarchichal Clustering Pada Pengelompokan Karya Tulis Ilmiah (Studi Kasus : Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri)
Fakultas – Program Studi	: Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jl. K.H. Ahmad Dahlan No. 76, Mojoroto, Kediri, Jawa Timur 64112

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 30 Januari 2017
Pembimbing I	Pembimbing II	Penulis,
Resty Wulanningrum, M.Kom NIDN 0719068702	Daniel Swanjaya, M.Kom NIDN 0723098303	Aldila Anggun Sasmita NPM 12.1.03.02.0208

Aldila Anggun Sasmita | 12.1.03.02.0208 simki.unpkediri.ac.id
Fakultas Teknik – Teknik Informatika || 1||

IMPLEMENTASI HIERARCHICHAL CLUSTERING PADA PENGELOMPOKAN KARYA TULIS ILMIAH (STUDI KASUS : PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS NUSANTARA PGRI KEDIRI)

Aldila Anggun Sasmita
12.1.03.02.0208
Fakultas Teknik – Teknik Informatika
Email : aldilasasmita08@gmail.com
Resty Wulanningrum, M.Kom dan Daniel Swanjaya, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Aldila Anggun Samita: Implementasi Hierarchichal Clustering pada Pengelompokan Karya Tulis Ilmiah (Studi Kasus: Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri), Skripsi, TI, FT UNP Kediri, 2017.

Penelitian ini membahas tentang *text mining* berbasis pengelompokan skripsi menggunakan metode *Hierarchichal Clustering.* Metode ini digunakan untuk mengelompokan skripsi pada Prodi Teknik Informatika Universitas Nusantara PGRI Kediri. Pada Prodi Teknik Informatika Universitas Nusantara PGRI Kediri data skripsi mahasiswa masih belum tertata dengan baik. Sehingga sering terjadi skripsi yang memiliki kesamaan, kesulitan dalam pencarian dan kesulitan dalam pengecekan tema skripsi. Saat ini untuk mencari dokumen yang memiliki kemiripan harus dilakukan dengan cara manual. Ketika dosen melakukan pengecekan tentang tema skripsi juga dilakukan secara manual. Pencarian maupun pengecekan yang dilakukan akan memakan waktu yang panjang dan sudah pasti itu sangat merugikan.

Permasalahan penelitian ini adalah bagaimana mengimplementasikan metode *Hierarchichal Clustering* pada pengelompokan karya tulis ilmiah ?

Pada sistem ini menggunakan pendahuluan/latar belakang sebagai *input*-an. Selanjutnya setiap kata yang ada dalam pendahuluan/latar belakang akan dipecah dengan menggunakan *Tokenizing,* tujuannya agar mudah diproses untuk langkah selanjutnya. Kata-kata yang telah dipecah akan di*stemming*kan dan dikembalikan ke bentuk kata dasar. Proses selanjutnya yaitu pembobotan menggunakan metode TF-IDF, bobot ini yang digunakan untuk menghitung jarak antar dokumen. Selanjutnya data akan diproses untuk mencari keberadaan kata dalam sebuah dokumen, hasilnya dibentuk dalam sebuah tabel vektor fitur. Dari tabel tersebut lanjut ke proses perhitungan kemiripan menggunakan algoritma *cosine similarity*. Ketika semua proses tersebut selesai, maka selanjutnya hasil dari proses *cosine similarity* akan diolah menggunakan *Hierarchichal Clustering*.

Kesimpulan dari penelitian ini adalah metode *Hierarchichal Clustering* bisa digunakan untuk mengelompokkan dokumen skripsi dan hasil yang diperoleh dapat langsung dilihat.

KATA KUNCI : *Cosine Similarity, Hierarchichal Clustering.*

I. LATAR BELAKANG

Penelitian merupakan suatu investigasi yang terorganisasi, yang dilakukan untuk menyajikan suatu informasi dan memecahkan masalah. Lebih lanjut, Kerlinger (Kerlinger 1986), menyatakan bahwa " Penelitian ilmiah merupakan

investigasi sistematik, terkendali, bersifat empiric, serta kritis mengenai fenomena alami (*natural phenomena*) yang dibimbing oleh teori dan hipotesis mengenai "hubungan-hubungan" yang diduga sebelumnya mengenai fenomena tersebut.

Karya tulis ilmiah berupa paparan tulisan atau hasil penelitian mahasiswa yang menempuh pendidikan jenjang Strata 1 yang membahas suatu permasalahan atau fenomena dalam bidang ilmu tertentu dengan menggunakan kaidah – kaidah yang berlaku. Dalam karya tulis ilmiah tentunya terdapat sebuah latar belakang yang menceritakan hal-hal yang melatarbelakangi peneliti dalam pemilihan judul penelitian.

Dengan banyaknya karya tulis ilmiah yang telah selesai dibuat, maka dibutuhkan pengelompokan karya tulis ilmiah mahasiswa sesuai dengan bidang konsentrasi masing – masing sebagai bahan penelitian dosen. Pada penelitian sebelumnya yang telah dilakukan oleh Kestrilia Rega .P dan Hendra Wijaya pada tahun 2014 dengan judul "Aplikasi *Text Mining* untuk Automasi Penentuan Tren Topik Skripsi dengan Metode *K-Means Clustering*". Pada penelitian ini metode *K-Means Clustering* digunakan untuk penentuan tren topik skripsi. Setelah diuji coba diperoleh rata-rata nilai *purity* sekitar 76% dokumen berhasil dikelompokkan.

Sampai saat ini Prodi Teknik Informatika, Universitas Nusantara PGRI Kediri belum memiliki sistem pengelompokan tersebut. Oleh karena itu penulis mengajukan Penelitian untuk mengelompokan skripsi dengan menggunakan metode *Hierarchical Clustering* sehingga dari hasil pengelompokan tersebut dapat diketahui tingkat kesamaan antar skripsi dan mencegah plagiasi di masa yang akan datang. Dari uraian di atas penulis mengajukan judul Penelitian "**Implementasi *Hierarchichal Clustering* pada Pengelompokan Karya Tulis Ilmiah (Studi Kasus : Prodi Teknik Informatika Universitas Nusantara PGRI Kediri)** ".

II. METODE

Hierarchichal Clustering merupakan salah satu metode dalam data mining yang digunakan untuk mengelompokan data. Edy

Irwansyah & Muhammad Faisal menjelaskan bahwa "Pada *hierarchichal clustering* data dikelompokkan melalui suatu bagan yang berupa hirarki, dimana terdapat penggabungan dua grup yang terdekat disetiap iterasinya ataupun pembagian dari seluruh set data kedalam *cluster* " (Irwansyah,2015).

Contoh metode *hierarchical clustering: single linkage, complete linkage, average linkage* dan *average group linkage*. Dalam kasus ini menggunakan *complete linkage*. Berikut ini contoh perhitungan dari *complete linkage* :

Terdapat matriks jarak antara lima buah obyek yaitu :

Tabel 2.1 : Tabel Matriks Jarak

	A	B	C	D	E
A	0.0	1.0	5.0	6.0	8.0
B	1.0	0.0	3.0	8.0	7.0
C	5.0	3.0	0.0	4.0	6.0
D	6.0	8.0	4.0	0.0	2.0
E	8.0	7.0	6.0	2.0	0.0

Langkah penyelesaian:

1. Mencari obyek dengan jarak minimum A dan B mempunyai jarak terdekat, yaitu 1.0 maka obyek A dan B bergabung menjadi satu *cluster*.
2. Menghitung jarak antara *cluster* AB dengan obyek lainnya.

$D_{(AB)C}$ = max {d_{AC} , d_{BC}}= d_{AC}= 5.0

$D_{(AB)D}$ = max {d_{AD} , d_{BD}}= d_{BD}= 8.0

$D_{(AB)E}$ = max {d_{AE} , d_{BE}}= d_{AE}= 8.0

Dengan demikian terbentuk matriks jarak yang baru

Tabel 2.2 : Tabel Matriks Jarak Baru

	AB	C	D	E
AB	0.0	5.0	8.0	8.0
C	5.0	0.0	4.0	6.0
D	8.0	4.0	0.0	2.0
E	8.0	6.0	2.0	0.0

3. Mencari obyek dengan jarak terdekat. D dan E mempunyai jarak yang terdekat yaitu 2.0 maka obyek D dan E bergabung menjadi satu *cluster*.
4. Menghitung jarak antar cluster dengan obyek lainnya.

$D_{(AB)C}$ = max {d_{AC} , d_{BC}}= d_{AC}= 5.0

$D_{(AB)(DE)}$ = max {d_{AD} , d_{AE}, d_{BD}, d_{BE}}= d_{AE}=d_{BD}= 8.0

$D_{(DE)C}$ = max {d_{CD} , d_{CE}}= d_{CE}= 6.0

Maka terbentuklah matriks jarak yang baru, yaitu :

Tabel 2.3 : Tabel Matriks Jarak Terbaru

	AB	C	DE
AB	0.0	5.0	8.0
C	5.0	0.0	6.0
DE	8.0	6.0	0.0

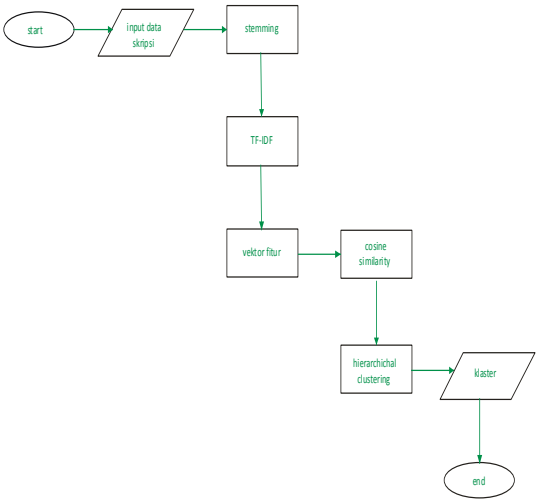
5. Mencari jarak terdekat antara cluster dengan obyek dan diperoleh obyek C bergabung dengan cluster AB.

6. Pada langkah yang terakhir cluster ABC bergabung dengan DE sehingga terbentuk cluster tunggal.

III. HASIL DAN KESIMPULAN

A. Hasil

Dari penjelasan sebelumnya, berikut ini merupakan gambaran dari alur sistem yang telah dibuat :



Gambar 3.1. *Flowchart* Sistem

Dalam *flowchart* Gambar 3.1 menunjukkan bahwa alur dimulai dari menginputkan data Skripsi. Yang dimaksud data skripsi yaitu: judul skripsi, NPM penulis, nama penulis dan latar belakang. Data tersebut masuk ke dalam proses *stemming*, dalam hal ini menggunakan *stemmer* TALA. Setelah proses *stemming* selesai, hasil dari proses tersebut selanjutnya dihitung bobotnya menggunakan algoritma TF-IDF. Dari proses ini dapat pula kita ketahui *threshold* atau batasannya. Selanjutnya hasil bobot tersebut dijadikan patokan untuk membuat sebuah tabel vektor fitur. Dari tabel tersebut lanjut ke proses penghitungan kemiripan, menggunakan algoritma *cosine similarity*. Pada bagian ini dapat pula diketahui seberapa mirip antar dokumen, dengan kata lain dapat mengetahui plagiasi. Ketika semua proses tersebut selesai, maka hasilnya akan diolah menggunakan *hierarchichal clustering* untuk dijadikan klaster dari dokumen Skripsi.

Misalkan terdapat beberapa sampel data yang berbentuk tabel *Feature Vektor* dari dokumen Skripsi sebagai berikut :

Tabel 2.1 : Tabel *Feature Vektor*

Term	Dokumen				
	d1	d2	d3	d4	d5
t1	1	0	1	1	1
t2	0	1	0	0	0
t3	1	0	1	0	1
t4	0	1	1	1	0
t5	1	0	0	0	0
t6	1	1	1	0	0
t7	0	0	1	0	0
t8	1	1	1	1	1
t9	0	1	1	0	0
t10	1	0	0	1	0

Dari tabel tersebut kemudian akan di hitung menggunakan algoritma *cosine similarity*. Berikut tabel yang menunjukkan penghitungan dengan rumus dari algoritma *cosine similarity* :

Tabel 2.2 : Tabel algoritma *cosine similarity*

A	B	A.B	A^2	B^2
1	1	1	1	1
0	0	0	0	0
0	1	0	0	1
1	0	0	1	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
1	1	1	1	1
0	0	0	0	0
1	0	0	1	0
1	0	0	1	0
Jumlah		2	5	3
penarikan akar			2,236068	1,732051
hasil menghitung dengan rumus *cosine similarity*				0,516398

Dari hasil menghitung dengan algoritma cosine similarity kemudian dibentuk tabel hierarki sebagai berikut :

Tabel 2.3 : Tabel Hirarki

	d1	d2	d3	d4	d5
d1	0	0,34	0,57	0,57	0,66
d2		0	0,62	0,41	0,24
d3			0	0,53	0,61
d4				0	0,52
d5					0

Dari tabel tersebut dipilih angka yang terbesar, didapat 0,66 yang berada pada d5. Maka d5 di klaster dengan d1. Lalu bandingkan jarak lain yang tidak termasuk klaster tersebut.

$d_{(15)2}$ = max (d_{12},d_{52}) = max (0,34 , 0) = 0,34

$d_{(15)3}$ = max (d_{13},d_{53}) = max (0,57 , 0) = 0,57

$d_{(15)4}$ = max (d_{14},d_{54}) = max (0,57 , 0) = 0,57

Lalu hasil perbandingan tersebut dimuat dalam tabel seperti dibawah ini :

Tabel 2.4 : Tahap Pertama

	d_1d_5	d_2	d_3	d_4
d_1d_5	0	0,34	0,57	0,57
d_2	0,34		0,62	0,41
d_3	0,57			0,53
d_4	0,57			

Pilih lagi angka yang terbesar, didapat 0,62 yang berada pada d_3. Maka d_3 di klaster dengan d_2. Lalu ulangi tahap membandingkan jarak.

$d_{(23)15}$ = max (d_{21}, d_{25}, d_{31}, d_{35}) = max (0, 0,24 , 0 , 0,61) = 0,61

$d_{(23)4}$ = max (d_{24},d_{34}) = max (0,41 , 0,53) = 0,53

Lalu hasil perbandingan tersebut dimuat kembali dalam tabel seperti berikut :

Tabel 2.5 : Tahap Kedua

	d_2d_3	d_1d_5	d_4
d_2d_3	0	0,61	0,53
d_1d_5	0,61	0	0,57
d4	0,53	0,57	0

Dari tabel tersebut dapat dipilih lagi angka terbesar yang selanjutnya akan di klaster, yaitu 0,61 yang terdapat pada d_1d_5. Lakukan klaster dengan d_4 dan bandingkan jaraknya.

$d_{(154)23}$ =max (d_{12},d_{13},d_{52},d_{53},d_{42},d_{43}) = max (0,34 , 0,57 , 0,0,0,0)= 0,57

Tabel 2.6 : Hasil Akhir

	$d_1d_5d_4$	d_2d_3
$d_1d_5d_4$	0	0,57
d_2d_3	0,57	0

Berdasarkan tabel tersebut serta proses klaster yang telah dijalani dapat disimpulkan bahwa klaster yang terbuat yaitu $d_1d_5d_4$ dan terjadi ketika jarak 0,57.

B. Kesimpulan

Melalui pembahasan yang telah diuraikan pada bab sebelumnya, maka dapat ditarik kesimpulan yaitu:

Penelitian ini telah mampu mengelompokkan skripsi yang ada, dimana terbentuk 5 kelompok, masing-masing beranggotakan rata-rata 8 sampai 9 dokumen.

IV. DAFTAR PUSTAKA

Ambarwati & Winarko, Edi. 2013. *Pengelompokan Berita Indonesia Berdasarkan Histogram Kata Menggunakan Self-Organizing Map*, 8 (1). (Online), tersedia: http://genome.tugraz.at/MedicalInformatics2/SOM.pdf, diunduh 4 Oktober 2016

Anhar, ST. 2010. *Panduan Menguasai PHP & MySQL Secara Otodidak*. Jakarta: Mediakita.

Budhi, Gregorius S., Rahardjo, Arlinah I. & Taufik, Hendrawan. 2008. *Hierarchichal Clustering untuk Aplikasi Automated Text Integration.* (Online), tersedia: http://fportfolio.petra.ac.id/user_files/02-030/Hierarchical%20Clustering%20untuk%20Integrasi%20Dokumen%20(Greg-Petra).doc, diunduh 13 januari 2016.

Cooper, Donald R.;& Schindler, P.S. 2006. *Business Research Methods* (8th ed). Boston: McGraw-Hill.

Irwansyah, Edy & Faisal, Muhammad. 2015. *Advanced Clustering Teori dan Aplikasi.* Yogyakarta: Deepublish.

Irwansyah, Edy & Faisal, Muhammad. 2015. *Advanced Clustering*. Yogyakarta: Deepublish.

Kerlinger, Fred N. 1986. *Foundation of Behavioral Research.* Forth Worth: Holt Rinehart and winston.

Prilianti, Kestrilia Rega & Wijaya, Hendra. 2014. *Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering*, 2 (1). (Online), tersedia: http://www.cybermatika.stei.itb.ac.id, diunduh 11 Januari 2016.

Supriyatno. 2010. *Pemrograman Database Menggunakan Java dan MySQL untuk Pemula.* Jakarta: Mediakita.

Tala F. Z.2004. A Study of Stemming Effects on Information Retrieval in

Bahasa Indonesia. Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands.

Tan, P. e*t al*. 2006. *Introduction to Data Mining*. Boston: Pearson Education.

Utomo, Mardi Siswo. 2013. *Implementasi Stemmer Tala pada Aplikasi Berbasis Web*, 18 (1). (Online), tersedia: http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1673/607, diunduh 25 Januari 2016.

Wijaya, Rahmadi. 2007. *Pengertian Penelitian Rekayasa*. Jakarta : Mediakita.