

**APLIKASI PENGELOMPOKAN *CONTENT* HALAMAN *WEB*
SECARA OTOMATIS MENGGUNAKAN METODE
*FUZZY C-MEANS CLUSTERING***

SKRIPSI

Oleh:

DANDI PAMARDI

NIM. 09650188



**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2016**

**APLIKASI PENGELOMPOKAN *CONTENT* HALAMAN *WEB* SECARA
OTOMATIS MENGGUNAKAN METODE
*FUZZY C-MEANS CLUSTERING***

SKRIPSI

Diajukan Kepada:

Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh

**DANDI PAMARDI
NIM. 09650188**

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2016**

**APLIKASI PENGELOMPOKAN *CONTENT* HALAMAN *WEB* SECARA
OTOMATIS MENGGUNAKAN METODE
*FUZZY C-MEANS CLUSTERING***

SKRIPSI

Oleh

**DANDI PAMARDI
NIM. 09650188**

Telah Diperiksa dan Disetujui Diuji:

Tanggal: 30 Juni 2016

Dosen Pembimbing I,

Dosen Pembimbing II,

**Zainal Abidin, M.Kom
NIP. 19760613 200501 1 001**

**Fatchurrochman, M. Kom
NIP. 19700731 200501 1 002**

**Mengetahui,
Ketua Jurusan Teknik Informatika**

**Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008
HALAMAN PENGESAHAN**

**APLIKASI PENGELOMPOKAN *CONTENT* HALAMAN *WEB* SECARA
OTOMATIS MENGGUNAKAN METODE
*FUZZY C-MEANS CLUSTERING***

SKRIPSI

Oleh

**DANDI PAMARDI
NIM. 09650188**

Telah Dipertahankan di Depan Dewan Penguji Skripsi dan
Dinyatakan Diterima Sebagai Salah Satu Persyaratan Untuk
Memperoleh Gelar Sarjana Komunikasi (S.Kom)

Tanggal 30 Juni 2016

Susunan Dewan Penguji	Tanda Tangan
1. Penguji Utama :Ririen Kusumawati, M.Kom NIP. 19720309 200501 2 002	()
2. Ketua :Dr. M. Amin Hariyadi, M.T NIP. 19670118 200501 1 001	()
3. Sekretaris :Zainal Abidin, M.Kom NIP. 19760613 200501 1 004	()
4. Anggota :Fatchurrochman, M.Kom NIP. 19700731 200501 1 002	()

Mengesahkan,
Ketua Jurusan Teknik Informatika

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

**SURAT PERNYATAAN
ORISINALITAS PENELITIAN**

Saya yang bertanda tangan di bawah ini:

Nama : Dandi Pamardi

NIM : 09650188

Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika

Judul Penelitian : APLIKASI PENGELOMPOKAN *CONTENT*
HALAMAN

*WEB SECARA OTOMATIS MENGGUNAKAN METODE FUZZY C-MEANS
CLUSTERING*

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut

Malang,

Yang Membuat Pernyataan,

Dandi Pamardi

NIM. 09650188

HALAMAN MOTTO



“Awali setiap keadaan dengan mengucap bismillah dan lanjutkan dengan senyuman”

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan untuk :

- Ibu, bapak, dan kakak yang selalu memberikan doa setiap saat dan menyayangi serta memberikan semangat dalam keadaan apapun
- Keluarga besar dari ibu dan bapak yang selalu mendo'akan dan memberikan nasehat untuk memberikan semangat
- Teman terbaik yang berada dikota malang yang selalu ada dalam kondisi apapun, terutama teman "LC-T34M" yang selalu memberikan masukan-masukan dan juga semangat yang luar biasa



KATA PENGANTAR

Assalaamu 'alaikum Warahmatullaahi Wabaarakaatuh

Syukur alhamdulillah penulis haturkan kepada Allah SWT atas rahmat, taufik serta hidayah-Nya, sehingga penulis mampu menyelesaikan penyusunan skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang teknik informatika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penulis menyadari adanya banyak keterbatasan yang penulis miliki dalam proses penyusunan skripsi ini, sehingga penulis banyak mendapat bimbingan dan arahan dari berbagai pihak. Untuk itu ucapan terima kasih yang sebesar-besarnya dan penghargaan setinggi-tingginya penulis sampaikan terutama kepada :

1. Bapak Zainal Abidin, M.Kom., selaku pembimbing dalam skripsi ini yang telah memberikan bimbingan dan pengarahan dalam proses penyelesaian skripsi ini.
2. Bapak Fatchurrochman, M.Kom., selaku pembimbing dalam skripsi ini yang telah memberikan bimbingan dan pengarahan dalam proses penyelesaian skripsi ini.
3. Ibu, Ayah, Kakakku dan seluruh keluarga besar yang selalu memberikan do'a dan motivasi dalam penyelesaian skripsi ini.
4. Prof. Dr. H. Mudjia Rahardjo, M.Si., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.

5. Dr. Hj. Bayyinatul Muchtaromah., drh., M.Si selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
6. Bapak Dr. Cahyo Crysdiyan, selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
7. Bapak Dr. Suhartono, M.Kom., selaku dosen wali memberikan bimbingan dan pengarahan dalam skripsi ini.
8. Segenap sivitas akademika Jurusan Teknik Informatika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.
9. Seluruh teman-teman Jurusan Teknik Informatika khususnya angkatan 2009.
10. Sahabat-sahabat penulis yang telah memotivasi dan membantu dalam proses penyelesaian skripsi ini.
11. Semua pihak yang tidak mungkin penulis sebutkan satu-persatu, atas segala yang telah diberikan, penulis ucapkan terima kasih yang sebesar-besarnya.

Sebagai penutup, penulis menyadari dalam skripsi ini masih banyak kekurangan dan jauh dari sempurna, untuk itu penulis selalu menerima segala kritik dan saran dari pembaca. Harapan penulis, semoga karya ini bermanfaat bagi kita semua.

Wasslaamu'alaikum Warahmatullahi Wabarakaatuh

Malang,

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
SURAT PERNYATAAN ORISINALITAS PENELITIAN.....	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
ABSTRAK	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	5
1.6 Metode Penelitian	6
1.7 Sistematika Penulisan Skripsi	6
BAB II TINJAUAN PUSTAKA	8
2.1 Information Retrieval.....	8
2.1.1 Definisi <i>Information Retrieval</i>	8
2.1.2 Arsitektur <i>Information Retrieval System</i>	9
2.1.3 Proses Indexing	12
2.2 Model Ruang Vektor (<i>Vector Space Model</i>).....	16
2.2.1 Definisi Model Ruang Vektor	16
2.2.2 Model Ruang Vektor Dalam <i>Information Retrieval</i>	16
2.2.3 Pembobotan Kata (<i>TF-IDF</i>)	18
2.2.4 Fuzzy C-Means	21
2.3 Penelitian Terkait.....	22
BAB III ANALISIS DAN PERANCANGAN SISTEM	23
3.1 Desain Sistem.....	23

3.1.1 Deskripsi Umum Sistem.....	23
3.2 Rancangan <i>Url Getter</i>	25
3.3 Rancangan Ekstraksi.....	26
3.4 Tahap Preprosesing	28
3.5 Pembobotan Frekuensi Kata	30
3.6 PerhitunganFuzzy C-Means.....	31
3.7 Analisa Kelas	34
3.8 Studi Kasus	34
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Implementasi	40
4.1.1 Ruang Lingkup Perangkat Keras	40
4.1.2 Ruang Lingkup Perangkat Lunak	41
4.1.3 Implementasi Desain Antarmuka.....	41
4.1.4 Implementasi Aplikasi.....	43
4.2 Hasil Uji Coba	60
4.2.1 Uji Coba.....	60
4.2.2 Hasil Evaluasi dan Pembahasan	91
4.3 Hubungan Penelitian yang Dilakukan dengan Kaidah Islam.....	91
BAB V PENUTUP	94
DAFTAR PUSTAKA	

DAFTAR TABEL

Tabel 3.1 Koleksi kata.....	30
Tabel 3.2 Koleksi Stop Word Studi Kasus	34
Tabel 3.3 Data Studi Kasus	34
Tabel 3.4 Hasil Perhitungan tf	36
Tabel 3.5 Pembentukan Vektor Index	36
Tabel 3.6 Pembentukan Model Ruang Vektor.....	37
Tabel 3.7 Normalisasi Model Ruang Vektor	37
Tabel 3.8 Pembentukan Vektor Keanggotaan Awal	38
Tabel 3.9 Menghitung Pusat Cluster	38
Tabel 3.10 Perubahan Matriks.....	39
Tabel 3.11 Hasil Proses Klasifikasi	39
Tabel 4.1 Source Code Pengaturan Koneksi Database	44
Tabel 4.2 Source Code Koneksi dan Olah Data Database	44
Tabel 4.3 Source Code Penentuan Tempat Hasil Unduhan.....	47
Tabel 4.4 Source Code Unduh Halaman Indeks Web.....	47
Tabel 4.5 Source Code Cek Kevalidan Url	49
Tabel 4.6 Source Code Pembentukan Koleksi Link	50
Tabel 4.7 Source Code Ekstraksi Judul dan Isi (Content).....	53
Tabel 4.8 Source Code Bentuk Koleksi Stop Word dan Stop Word Removal.....	55
Tabel 4.9 Source Code Simpan Bobot tf ke Database	55
Tabel 4.10 Source Code Pembentukan vector data dan inisialisasi variable awal	57
Tabel 4.11 Source Code Membangkitkan bilangan random	58
Tabel 4.12 Source Code Hitung pusat cluster	58
Tabel 4.13 Source Code Menghitung Fungsi Objektif	59
Tabel 4.14 Source Code Hitung Perubahan Matriks Partisi	59
Tabel 4.15 Source Code Cek Kondisi Berhenti.....	60
Tabel 4.16 Hasil Crawl dan Klasifikasi.....	62
Tabel 4.17 Hasil Klasifikasi Setelah Perubahan Data.....	77

DAFTAR GAMBAR

Gambar 2.1 Arsitektur dasar sistem IR.....	10
Gambar 2.2 Proses dalam sistem IR	11
Gambar 2.3 Contoh lima tahap indexing berbasis content	16
Gambar 2.4 Contoh model ruang vektor dengan dua dokumen.....	18
Gambar 2.5 Contoh matriks term-document	18
Gambar 3.1 Activity Diagram Url Getter	25
Gambar 3.2 Kode sumber halaman web sebelum ekstraksi	26
Gambar 3.3 Kode halaman judul	26
Gambar 3.4 Konten setelah diekstraksi	27
Gambar 3.5 Konten setelah dihilangkan karakter symbol tidak penting.....	28
Gambar 3.6 Konten setelah diubah ke karakter huruf kecil	29
Gambar 3.7 Konten setelah dihilangkan stopwords.....	29
Gambar 3.8 Activity Diagram perhitungan TF.....	31
Gambar 3.10 Flowchart Fuzzy C-means	33
Gambar 4.1 Form Fuzzy C-means Classification	41
Gambar 4.2 Form Crawl.....	42
Gambar 4.3 Form Classification Only.....	43

ABSTRAK

Pamardi, Dandi. 2016. **Aplikasi Pengelompokan *Content* Halaman Web Secara Otomatis Menggunakan Metode *Fuzzy C-Means Clustering***. Skripsi. Jurusan Teknik Informatika Fakultas Sains Dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim, Malang.
Pembimbing : (I) Zainal Abidin, M.Kom (II) Fatchurrochman, M.Kom

Kata Kunci: Pengelompokan, Klasifikasi, *Content Halaman Web*, *Fuzzy C-means*.

Klasifikasi adalah proses pengelompokan objek yang memiliki ciri yang mirip menjadi golongan unit tertentu.

Penelitian ini membahas mengenai pengelompokan halaman *web* berdasarkan kandungan intisari yaitu isi (*content*) menggunakan metode *Fuzzy C-Means*. Tujuan yang ingin didapatkan adalah mengelompokan halaman *web* secara otomatis kedalam group-group dan juga untuk mempermudah proses pencarian informasi yang diinginkan dengan cepat. Aplikasi pengelompokan ini dibangun menggunakan bahasa Java sebagai bahasa pemrogramannya. Untuk uji cobanya dilakukan dengan meng-*crawl* 1 *web* yang sudah ditentukan.

Dari hasil uji coba pada proses uji coba, didapatkan bahwa proses *crawl* menghasilkan 240 dokumen *web* dan dikelompokan secara random untuk keanggotaan awal mejadi 5 group. Kemudian oleh aplikasi dokumen tersebut diklasifikasi atau dikelompokan lanjutan. Beberapa dokumen berubah groupnya. Ada pula yang tidak berubah groupnya. Dokumen yang berubah groupnya sebesar 56% dari jumlah total dokumen dengan beberapa perulangan yang telah dilakukan oleh aplikasi.

ABSTRACT

Pamardi, Dandi. 2016. **Application Grouping Content Web Pages Automatically Using Fuzzy C-Means Clustering Method**. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang,
Supervisor: (I) Zainal Abidin, M.Kom dan (II) Fatchurrochman, M.Kom

Keywords : *Grouping, Classification, Web page content, Fuzzy C-means*

Classification is the process of grouping objects that have similar characteristics into groups of units.

This study discusses the clustering of web pages based on the content of the essence of that content (content) using Fuzzy C-Means. Objectives to be obtained is automatically classify web pages into group-group and also to simplify the process of finding the desired information quickly. This grouping of applications built using the Java language as a programming language. To test try to do the crawl first web that has been determined.

From the test results on the test, it was found that the process of crawl produces 240 web documents and randomly grouped to form the initial membership of five group. Then by the application documents are classified or grouped continued. Some documents changed its group. There also does not change its group. Documents which changed its group amounted to 56% of the total number of documents with multiple loops that have been performed by the application.

المخلص

فامردي، داندي. ٢٠١٦. التجمع التطبيقى محتوى صفحة ويب تلقائيا بطريقة Fuzzy C-Means Clustering. شعبة علوم المعلوم والتكنولوجيا جامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج.
المشرف: (١) زين العابدين الماجستير ، (٢) مفتح الرحمن الماجستير

كلمات البحث: تجمع، تصنيف، صفحة ويب المحتوى، Fuzzy C-Means Clustering
التصنيف هو عملية تجميع الكائنات التي لها خصائص مماثلة إلى مجموعات من الوحدات.

تتناول هذه الدراسة تجميع صفحات الويب على أساس محتوى جوهر هذا المحتوى (المضمون) باستخدام طرق Fuzzy Fuzzy C-Means . الأهداف التي يمكن الحصول عليها وتصنيف صفحات الويب إلى جماعة مجموعة تلقائيا، وكذلك لتبسيط عملية العثور على المعلومات المطلوبة بسرعة. هذه المجموعة من التطبيقات التي تم بناؤها باستخدام لغة Java كما لغة البرمجة. لاختبار تحاول أن تفعل Crawl ١ الويب التي تم تحديدها.

من نتائج الاختبار على الاختبار، تبين أن عملية تنتج ٢٤٠ الوثائق الزحف على شبكة الإنترنت وتجميعها بشكل عشوائي لتشكيل عضوية الأولية من المجموعة الخامسة. ثم من وثائق الطلب يتم تصنيفها أو تجميعها تابع. تغيرت بعض الوثائق مجموعتها. هناك أيضا لا يغير مجموعتها. بلغت الوثائق التي غيرت فريقها إلى ٥٦٪ من إجمالي عدد الوثائق مع الحلقات المتعددة التي تم تنفيذها من قبل التطبيق.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut sejarahnya website adalah sejumlah halaman web yang memiliki topik saling terkait, terkadang disertai pula dengan berkas-berkas gambar, video, atau jenis-jenis berkas lainnya. Sebuah situs web biasanya ditempatkan setidaknya pada sebuah server web yang dapat diakses melalui jaringan seperti internet, ataupun jaringan wilayah lokal (LAN) melalui alamat internet yang dikenali sebagai URL (id.wikipedia.org). Dari awal kemunculan website ada banyak sekali situs web yang ada di internet.

Sekarang ini website lebih dikenal sebagai media penyampai informasi (Komang Wiswakarma 2009). Seperti halnya beberapa situs web yang menyediakan layanan informasi adalah okezone.com, detik.com, kompas.com antaranews.com dan masih banyak lagi yang lainnya. Dalam suatu web tidak hanya menyampaikan satu atau dua informasi tapi ratusan atau bahkan ribuan informasi ditampilkan dalam satu web. Terkadang konsumen kesulitan untuk mencari informasi yang dibutuhkannya karena dengan banyaknya informasi yang disediakan dalam web tersebut. Dan mungkin kita hanya membuang-buang waktu percuma untuk mendapatkan informasi yang mungkin tidak terlalu banyak. Atau bahkan kita hanya mendapatkan informasi yang kurang penting. Seperti ayat-ayat yang ada dalam al-quran sedikit membahas bahwa informasi itu sangat penting dan bermanfaat bagi banyak orang.

وَذَكِّرْ فَإِنَّ الذِّكْرَى تَنْفَعُ الْمُؤْمِنِينَ

Artinya :

“Dan tetaplah peringatan, karena sesungguhnya peringatan itu bermanfaat bagi orang-orang yang beriman”. (QS. Adz-Dzariyaat ayat 55 diambil dari software *Quran in Word*)

Bagi Islam, penguasaan akan suatu informasi dapat mempercepat penyebaran gagasan secara massal serta dapat menjadi sarana yang ampuh menghadapi pengaruh negatif. Sehingga media informasi dan jaringan komunikasi yang ada dapat dijadikan sarana strategis. Dengan adanya ayat tersebut sedikit membuktikan bahwa informasi begitu penting bagi siapapun dan sekarang informasi yang diberikan media begitu banyak sehingga kita harus bias memilih informasi yang sesuai dengan kebutuhan.

Dan untuk mengatasi masalah ini mungkin salah satu solusinya adalah pengelompokan halaman web berdasarkan kategori yang sesuai untuk memudahkan pengklasifikasian informasi setelah dibentuk kategori ini. Serta memudahkan pencarian informasi yang dibutuhkan secara cepat dan tepat. Data untuk pengelompokan ini diambil dari hasil ekstraksi halaman web yang kemudian diolah dengan langkah-langkah tertentu dan kemudian dihasilkanlah kategori-kategori. Aplikasi ini akan dibuat secara otomatis dengan menyesuaikan informasi-informasi yang ada dalam web tersebut untuk dijadikan dalam satu bagian yang mempunyai karakter sama disetiap informasinya.

يُرِيدُ اللَّهُ بِكُمْ الْيُسْرَ وَلَا يُرِيدُ بِكُمْ الْعُسْرَ...

Artinya :

“Allah menghendaki kemudahan bagimu, dan tidak menghendaki kesukaran bagimu”. (QS. Al-Baqarah ayat 185 diambil dari software *Quran in Word*)

Seperti yang telah terlihat pengelompokan informasi berdasarkan kategori ini sangat penting untuk memudahkan pencarian informasi yang dibutuhkan dengan cepat dan tepat. Seperti yang telah dijelaskan sebelumnya bahwa dalam satu website tidak hanya diberikan satu atau dua informasi saja tetapi banyak sekali informasi yang diberikan. Seperti halnya salah satu website seperti antaranews.com, situs ini memberikan begitu banyak informasi dari yang dahulu kala maupun yang terbaru. Untuk mencari informasi yang kita butuhkan pun terkadang masih keliru dengan adanya beberapa informasi yang belum sesuai dengan kategori. terkadang kita menjumpai juga berita olahraga yang ada kaitannya dengan ekonomi, politik ataupun yang lainnya. Maka disinilah dibutuhkannya pembuatan aplikasi pengelompokan ini.

Dalam penelitian ini aplikasi yang akan dibuat adalah aplikasi pengelompokan halaman web secara otomatis menggunakan metode *Fuzzy C-Means*. Metode ini khusus digunakan untuk pengklasteran tapi sebelum melakukan proses pengklasteran mungkin ada proses-proses yang harus dilakukan.

Mungkin salah satunya adalah text processing dimana proses ini dilakukan untuk memfilter isi dari data tersebut agar dapat dilakukan proses pengklasteran dengan menggunakan metode *fuzzy c-means*. Metode *fuzzy c-means* ini ditujukan untuk membantu dalam pembentukan kelompok-kelompok serupa yang lebih kecil dari koleksi content yang lebih besar (Sara Nasser, Rawan Alkhaldi, Gregory Vert. 2006). *Fuzzy c-means* merupakan salah satu metode pengelompokan yang sering digunakan karena menghasilkan kelompok-kelompok yang cukup efektif (Shihab,2000). Dalam menentukan kelompok atau membuat pengelompokan metode ini didasarkan pada derajat keanggotaan. (Bezdek,2005).

1.2 Rumusan Masalah

- a. Apakah algoritma *fuzzy c-means* dapat digunakan untuk pengelompokan *content* halaman *web*?
- b. Seberapa baik algoritma *fuzzy c-means* dalam mengelompokkan *content* halaman *web*?
- c. Bagaimana membangun aplikasi pengelompokkan *content* halaman *web* menggunakan metode *fuzzy c-means clustering*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah membuat aplikasi pengelompokan *content web* secara otomatis menggunakan metode *fuzzy c-means*. Dan juga membuktikan bahwa metode *fuzzy c-means* dapat digunakan untuk mengelompokkan *content* halaman *web* berita.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

- a. *Content* halaman *web* yang diekstrak adalah halaman *web* berita berbahasa Indonesia.
- b. Inputan data awal yang akan diproses adalah alamat *url* dari *web* berita yang akan diunduh.
- c. *Web extraction* yang digunakan sebagai bahan penelitian adalah halaman *web* yang berisi tentang berita bahasa Indonesia.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan akan memberikan manfaat sebagai berikut:

- a. Menyediakan aplikasi yang membantu dalam pengelompokan dokumen berita.
- b. Menyediakan aplikasi yang dapat memudahkan user dalam menemukan berita yang diinginkan secara cepat.
- c. Menyediakan aplikasi yang dapat digunakan dalam pengelompokan *content* halaman *web* secara otomatis.

1.6 Metode Penelitian

Untuk mencapai tujuan yang telah ditentukan sebelumnya, maka metodologi pengumpulan data yang dilakukan dalam penulisan skripsi ini adalah *library research* yaitu suatu cara penelitian dan pengumpulan data teoritis dari buku-buku, artikel, jurnal dan berbagai literatur yang mendukung penyusunan skripsi. Pengumpulan data juga dilakukan dengan melakukan uji coba aplikasi.

1.7 Sistematika Penulisan Skripsi

Sistematika dalam penulisan skripsi ini akan dibagi menjadi beberapa bab sebagai berikut:

BAB I Pendahuluan

Pada bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan skripsi ini.

BAB II Tinjauan Pustaka

Landasan teori menjelaskan beberapa teori yang menjadi dasar penyusunan skripsi ini. Teori-teori yang dibahas dalam bab ini adalah teori yang berkaitan dengan pembahasan tentang *Information Retrieval System*, *TF IDF*, *Fuzzy C-Menas*.

BAB III Analisis, dan Perancangan Sistem

Bab ini menjelaskan mengenai analisa pengklasifikasian data yang sudah ada sebelumnya dan perancangan pengembangan aplikasi pengklasifikasian ini seperti apa nantinya.

BAB IV Hasil Dan Pembahasan

Bab ini membahas tentang implementasi dari algoritma *fuzzy c-means* pada pengelompokan data atau *content web*.

BAB V Penutup

Bab ini berisi kesimpulan dari keseluruhan dari laporan skripsi dan saran yang diharapkan dapat bermanfaat untuk pengembangan pembuatan sistem selanjutnya.

Daftar Pustaka

Seluruh materi referensi yang ada dalam penulisan skripsi ini, akan dicantumkan dalam bab ini.

Lampiran

Data pendukung untuk melengkapi uraian yang telah disajikan dalam bagian utama di tempatkan di bagian ini.

BAB II

TINJAUAN PUSTAKA

2.1 Information Retrieval

2.1.1 Definisi *Information Retrieval*

Information Retrieval System atau Sistem Temu Balik Informasi merupakan seni dan ilmu untuk mencari sebuah informasi pada dokumen. Dapat juga dikatakan mencari dokumen sendiri, mencari sebuah metadata dengan gambaran berbentuk dokumen, atau mencari dalam data base, apakah itu hubungan data base yang berdiri sendiri atau *hiperteks* jaringan database seperti internet atau intranet, untuk teks, suara, gambar atau data. Sistem temu balik informasi merupakan sistem yang mampu melakukan penyimpanan, pencarian, dan pemeliharaan informasi. Informasinya merupakan sebuah teks (termasuk data numerik dan tanggal), gambar, audio, video, dan objek multimedia lainnya (Kowalski, 1997).

Dalam beberapa dekade terakhir Sistem Temu Balik Informasi berperan besar dalam mengatur ledakan informasi dalam literatur ilmiah. Tidak sedikit juga universitas dan perpustakaan umum menggunakan IR sistem untuk menyediakan akses ke buku, jurnal, dan dokumen lainnya. IR sistem seringkali dihubungkan kepada objek dan *query*. Tujuan utama dari *Information Retrieval System* (IRS) adalah menemukan kembali informasi yang mungkin berguna atau relevan dengan informasi yang dicari oleh pengguna dalam sekumpulan dokumen baik itu relevan ataupun tidak diwaktu bersamaan (Murad, 2007). Dalam penerapannya sistem IR yang baik adalah system yang dapat menangani kebutuhan informasi pengguna secara cepat dan akurat. Model *Information*

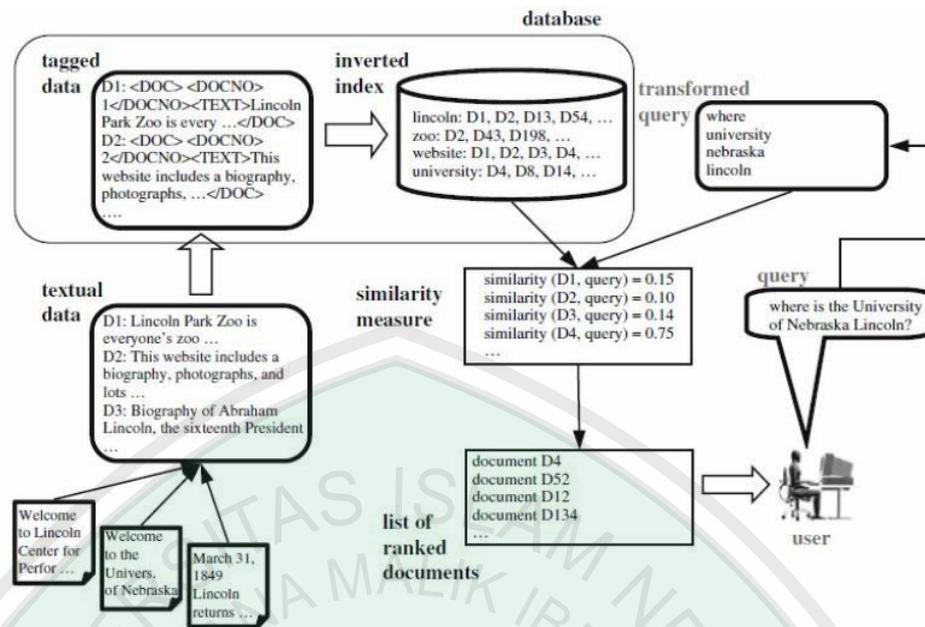
Retrieval adalah model yang dilakukan untuk melakukan pencocokan antara informasi yang dicari oleh pengguna dengan kumpulan dokumen yang ada (Kowalski, 1997).

Model yang terdapat dalam Information retrieval terbagi dalam 3 model besar, yaitu:

- a. *Set-theoretic models*, yaitu model yang merepresentasikan dokumen sebagai himpunan kata atau frase. Contoh model ini ialah *standard Boolean model* dan *extended Boolean model*.
- b. *Algebraic model*, yaitu model yang merepresentasikan dokumen dan query sebagai vektor atau matriks *similarity* antara vektor dokumen dan vektor *query* yang direpresentasikan sebagai sebuah nilai skalar. Contoh model ini ialah *vektor space model* (model ruang vektor) dan *latent semantic indexing* (LSI).
- c. *Probabilistic model*, yaitu model melakukan proses pengambilan dokumen sebagai sebuah *probabilistic inference*. Contoh model ini ialah penerapan teorema bayes dalam model probabilistik.

2.1.2 Arsitektur Information Retrieval System

Pekerjaan utama yang ditangani oleh system ini adalah melakukan pre-processing dan penerapan metode sesuai **Gambar 2.1** (Cios,2007). Pre-processing diaplikasikan ke database yang telah tersedia sebagai bahan untuk penerapan metode. Selanjutnya adalah penerapan metode tertentu untuk menghitung kedekatan (relevansi atau similarity) dokumen pada database yang telah tersedia pada tahap sebelumnya. Query pengguna adalah bagian pokok pada sistem ini.

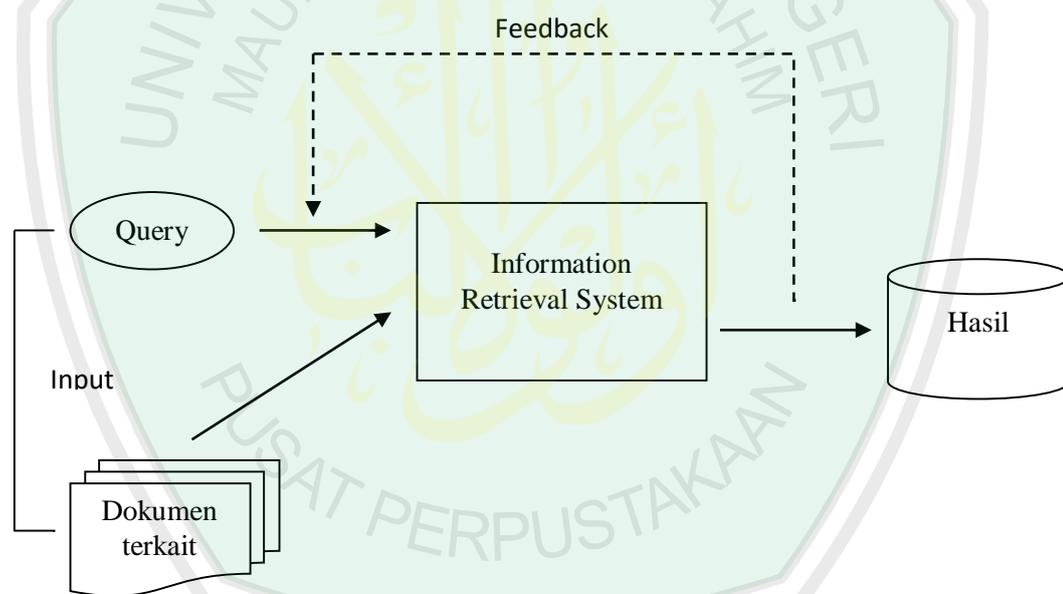


Gambar 2.1 Arsitektur dasar sistem IR (cios,2007)

Pada tahapan *pre-processing*, sistem yang berurusan dengan dokumen *semi-structured* biasanya memberikan tag tertentu pada *term - term* atau bagian dari dokumen, sedangkan pada dokumen tidak terstruktur proses ini dilewati dan membiarkan term tanpa imbuhan tag. *Query* yang dimasukkan pengguna dikonversi sesuai aturan tertentu untuk mengekstrak term - term penting yang konsisten dengan term - term yang sebelumnya telah diekstrak dari dokumen dan menghitung relevansi antara *query* dan dokumen berdasarkan pada term - term tersebut. Hasilnya, sistem mengembalikan suatu daftar dokumen terurut *descending (ranking)* sesuai nilai kemiripannya dengan *query* pengguna (Cios, 2007).

Setiap dokumen (termasuk *query*) direpresentasikan menggunakan model *bag-of-words* yang mengabaikan urutan dari kata-kata di dalam dokumen, struktur sintaktis dari dokumen dan kalimat. Dokumen ditransformasi ke dalam suatu "task" berisi kata-kata *independen*. *Term* disimpan dalam suatu

database pencarian khusus yang ditata sebagai sebuah *invertedindex*. Index ini merupakan konversi dari dokumen asli yang mengandung sekumpulan kata ke dalam daftar kata yang berasosiasi dengan dokumen terkait dimana kata-kata tersebut muncul. Secara singkat proses dalam *Information Retrieval* dapat digambarkan dengan proses mendapatkan dokumen terkait dari kumpulan dokumen yang ada melalui pencarian *query* yang di-*input*-kan oleh *user* (pengguna) seperti terlihat pada **Gambar 2.2**. Proses yang terjadi dalam sistem *InformationRetrieval* sendiri terdiri dari 2, yaitu *Indexing subsystem* dan *Searching subsystem (matching system)*.



Gambar 2.2 Proses dalam sistem IR

2.1.3 Proses Indexing

Suatu bahasa indeks adalah bahasa yang digunakan untuk menguraikan dokumen dan permintaan. Unsur-unsur dari bahasa indeks adalah terminologi indeks, yang mungkin diperoleh dari teks dokumen untuk diuraikan, atau mungkin dengan bebas. Bahasa indeks dapat diuraikan menjadi pre-coordinate atau post-coordinate, yang pertama menunjukkan bahwa terminologi dikoordinir ketika mengindeks dan ketika dalam pencarian. Secara lebih rinci, dalam indeks pre-coordinate suatu kombinasi logis tentang segala terminologi indeks mungkin digunakan sebagai suatu label untuk mengidentifikasi suatu kelas dokumen, sedangkan di dalam indeks post-coordinate kelas yang sama akan dikenali pada waktu pencarian dengan mengombinasikan kelas dokumen berlabel dengan terminologi indeks individu.

Pembangunan index dari koleksi dokumen merupakan tugas pokok pada tahapan *pre-processing* di dalam IR. Kualitas index mempengaruhi efektifitas dan efisiensi sistem IR (Chu, 2002). *Index* dokumen adalah himpunan *term* yang menunjukkan isi atau topik yang dikandung oleh dokumen. *Index* akan membedakan suatu dokumen dari dokumen lain yang berada di dalam koleksi. Ukuran index yang kecil dapat mengakibatkan hasil buruk dan mungkin dapat kehilangan beberapa item yang relevan. Index yang besar memungkinkan retrieval banyak dokumen bermanfaat sekaligus dapat menaikkan jumlah dokumen yang tidak relevan dan juga dapat menurunkan kecepatan pencarian (*searching*), (Hyusein, 2003).

Pembuatan *inverted index* harus melibatkan konsep *linguistic processing* yang bertujuan mengekstrak *term - term* penting dari dokumen yang direpresentasikan sebagai *bag-of-words*. Ekstraksi *term* biasanya melibatkan dua operasi utama berikut (Cios, 2007).

- a. Penghapusan ***stopwords***. Stopword merupakan kata-kata yang sangat sering muncul dalam dokumen. Stopword ini tidak bisa dikatakan sebagai kata-kata yang tidak berpengaruh terhadap proses kategorisasi. Misalnya adalah kata penghubung seperti ‘dan’, ‘atau’, ‘kemudian’, dan seterusnya. Selain itu, kata depan juga merupakan stopwords yang tidak memiliki arti penting bagi dokumen itu sendiri. Oleh karena itu, stopwords harus dihilangkan. Penghilangan stopwords ini dilakukan berdasarkan kamus kata tertentu yang disebut dengan database stopwords.
- b. ***Stemming***. Stemming dilakukan untuk mengubah kata berimbuhan menjadi kata dasarnya. Misalnya ‘berlari’ menjadi ‘lari’, ‘pemrosesan’ menjadi ‘proses’ dan seterusnya. Dengan demikian, kualitas informasi meningkat. Kualitas informasi yang dimaksud adalah hubungan antar kata itu sendiri, misalnya ‘memberi’, ‘diberi’, ‘memberikan’, yang semula adalah kata yang berbeda, dengan adanya stemming, kata tersebut menjadi ‘beri’, sehingga ada hubungan antara satu sama lain. Selain itu, space yang digunakan untuk penyimpanan juga menjadi lebih kecil. Terdapat 5 langkah pembangunan *inverted index*, yaitu (Manning, 2009):

- **Penghapusan format dan markup dari dalam dokumen**

Jika dokumen yang digunakan bukan berupa teks murni maka tahap ini dilakukan. Karena dokumen teks yang biasanya kita lihat berupa format non teks seperti html, pdf atau dalam bentuk word. Format-format ini mengharuskan sebuah teks dilengkapi unsur-unsur tambahan untuk dapat menghasilkan tampilan yang friendly dimata kita. Informasi-informasi itu dihilangkan karena dianggap tidak perlu dan tidak mencerminkan isi sebuah dokumen teks.

- **Pemisahan rangkaian kata (*tokenization*)**

Proses tokenizing adalah proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini memisahkan setiap kata yang menyusun dokumen. Pada umumnya setiap kata teridentifikasi dengan kata yang lain oleh karakter spasi, sehingga proses tokenizing mengandalkan karakter spasi pada dokumen untuk pemisahan kata. Pada proses tokenizing biasanya juga ditambahkan informasi jumlah kemunculan kata pada kalimat tersebut.

- **Penyaringan (*filtration*)**

Proses Filtering adalah proses pengambilan kata-kata yang dianggap penting atau mempunyai makna saja. Pada proses ini kata-kata yang dianggap tidak mempunyai makna seperti kata sambung akan dihilangkan. Pada proses ini biasanya digunakan daftar stopword yang tersimpan dalam

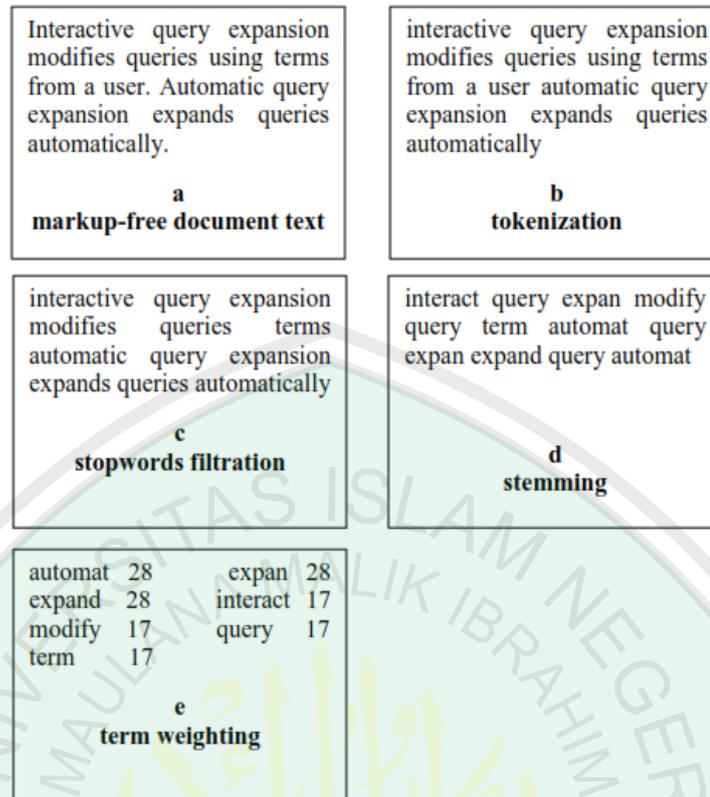
suatu tabel basis data, yang nantinya digunakan sebagai acuan penghilangan kata. Daftar stopword berbeda untuk setiap bahasanya.

○ **Konversi *term* ke bentuk dasar (*stemming*)**

Stemming adalah proses konversi *term* ke bentuk umumnya, sebagaimana dijelaskan sebelumnya. Dokumen dapat pula diekspansi dengan mencarikan sinonim bagi term-term tertentu di dalamnya. Sinonim adalah kata-kata yang mempunyai pengertian serupa tetapi berbeda dari sudut pandang morfologis. Seperti *stemming*, operasi ini bertujuan menemukan suatu kelompok kata terkait. Akan tetapi sinonim bekerja berdasarkan pada *thesaurus*, tidak berbagi-pakai term stem. Jika pengguna memasukkan query “heart disease” maka query diekspansi untuk mengakomodasi semua sinonim dari *disease* seperti *ailment*, *complication*, *condition*, *disorder*, *fever*, *ill*, *illness*, *infirmity*, *malady*, *sickness*, dan lain-lain (Cios, 2007).

○ **Pemberian bobot terhadap term (*weighting*)**

Dimulai dengan perhitungan jumlah kata dalam setiap dokumen, yang kemudian akan dihitung menggunakan skema pembobotan yang dikehendaki. Setiap *term* diberikan bobot sesuai skema pembobotan yang dipilih, apakah pembobotan lokal, global atau kombinasi keduanya. Banyak aplikasi menerapkan pembobotan dengan perkalian bobot lokal *term frequency* dan *global inverse document frequency*, ditulis *tf .idf*.



Gambar 2.3 Contoh lima tahap indexing berbasis content (Ruthven, 2003)

2.2 Model Ruang Vektor (*Vector Space Model*)

2.2.1 Definisi Model Ruang Vektor

Dalam sistem temu balik informasi Vektor Space Model mengibaratkan masing – masing query dan dokumen sebagai sebuah vector n-dimensi. Satu term mewakili tiap dimensi pada vector. *Term* yang digunakan biasanya berpatokan kepada *term* yang ada pada *query* atau *keyword*, sehingga *term* yang ada pada dokumen tetapi tidak ada pada *query* biasanya diabaikan.

2.2.2 Model Ruang Vektor Dalam *Information Retrieval*

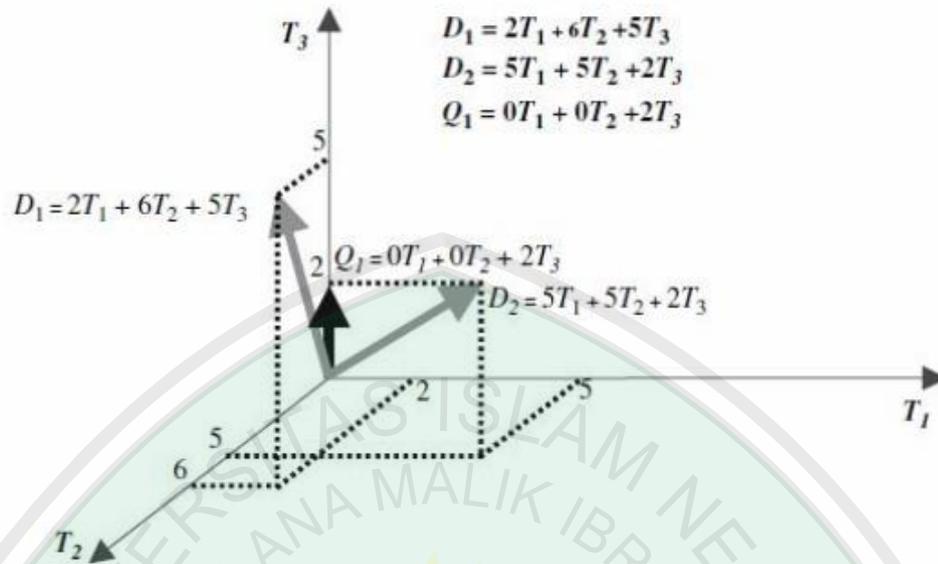
Model Ruang Vektor adalah salah satu dari beberapa metode searching yang diterapkan pada Information Retrieval System. Model ruang vektor dibuat berdasarkan pemikiran bahwa isi dari dokumen ditentukan oleh kata-kata yang digunakan dalam dokumen tersebut. Model ini menentukan kemiripan

(*similarity*) antara dokumen dengan query dengan cara merepresentasikan dokumen dan *query* masing-masing ke dalam bentuk vektor. Elemen vector diperoleh dari tiap kata yang ditemukan pada dokumen dan query yang melewati proses pembobotan kata.

Kemiripan antar dokumen didefinisikan berdasarkan representasi *bag-of-words* dan dikonversi ke suatu model ruang vektor (*vector space model*, VSM). Model ini diperkenalkan oleh Salton dan telah digunakan secara luas. Pada VSM, setiap dokumen di dalam database dan *query* pengguna direpresentasikan oleh suatu vektor multi-dimensi (Cios, 2007). Dimensi sesuai dengan jumlah term dalam dokumen yang terlibat Pada model ini:

- a. *Vocabulary* merupakan kumpulan semua *term* berbeda yang tersisa dari dokumen setelah *preprocessing* dan mengandung *termindex*. *Term-term* ini membentuk suatu ruang vektor.
- b. Setiap *term i* di dalam dokumen atau *query j*, diberikan suatu bobot (*weight*) bernilai *real* W_{ij} .
- c. Dokumen dan *query* diekspresikan sebagai vektor t dimensi $d_j = (W_1, W_2, \dots, W_t)$ dan terdapat n dokumen di dalam koleksi, yaitu $j = 1, 2, \dots, n$.

Contoh dari model ruang vektor tiga dimensi untuk dua dokumen D1 dan D2, satu query pengguna Q1, dan tiga term T1, T2 dan T3 diperlihatkan pada **Gambar 2.4**. Dalam model ruang vektor, koleksi dokumen direpresentasikan oleh matriks *term-document* (atau matriks *term-frequency*). Setiap sel dalam matriks bersesuaian dengan bobot yang diberikan dari suatu term dalam dokumen yang ditentukan. Nilai nol berarti bahwa *term* tersebut tidak hadir di dalam dokumen (Cios, 2007).



Gambar 2.4 Contoh model ruang vektor dengan dua dokumen(Polettini,2004)

$$\begin{pmatrix}
 & T_1 & T_2 & \dots & T_t \\
 D_1 & W_{11} & W_{21} & \dots & W_{t1} \\
 D_2 & W_{12} & W_{22} & \dots & W_{t2} \\
 \dots & \dots & \dots & \dots & \dots \\
 D_n & W_{1n} & W_{2n} & \dots & W_{tn}
 \end{pmatrix}$$

Gambar 2.5 Contoh matriks *term-document*(Cios,2007)

2.2.3 Pembobotan Kata (*TF-IDF*)

Dalam menentukan kemiripan antara dokumen dengan query, Pembobotan kata sangatlah berpengaruh. Hasil perhitungan teks akan menghasilkan kemiripan yang baik jika bobot tiap kata dapat ditentukan dengan tepat. Keberhasilan dari model VSM ini ditentukan oleh skema pembobotan terhadap suatu *term* baik untuk cakupan lokal maupun global, dan faktor normalisasi (Polettini, 2004). Pembobotan lokal hanya berpedoman pada

frekuensi munculnya term dalam suatu dokumen dan tidak melihat kemunculan term tersebut di dalam dokumen lainnya. Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah cara pemberian bobot hubungan suatu kata (term) terhadap dokumen. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu Term frequency (TF) merupakan frekuensi kemunculan kata (t) pada kalimat (d). Document frequency (DF) adalah banyaknya kalimat dimana suatu kata (t) muncul. TF juga disebut Local Term Weight, didefinisikan sebagai frekuensi kemunculan TF tersebut dalam sebuah dokumen. Algoritma TF sendiri bermacam-macam seiring perkembangan zaman. Intinya algoritma TF hanya berkorelasi dengan frekuensinya sendiri. IDF (Inverse Document Frequency) adalah pengurangan dominasi term yang sering muncul di berbagai dokumen. Hal ini diperlukan karena term yang banyak muncul dapat dianggap sebagai term umum sehingga tidak penting nilainya. Sebaliknya factor kejarangmunculan kata harus diperhatikan dan dipandang sebagai kata yang lebih penting. IDF juga disebut Global Term Weight, didasarkan pada penghitungan jumlah dokumen dalam suatu kumpulan yang dicari yang diindeks oleh suatu istilah (Tian dan Tong, 2010).

Pada dasarnya TF-IDF bekerja dengan menentukan frekuensi relatif suatu kata dalam dokumen tertentu dan dibandingkan dengan proporsi kebalikan dari kata itu pada seluruh dokumen. Secara intuitif, perhitungan ini menentukan seberapa relevan kata yang disisipka dalam suatu dokumen tertentu. Kata-kata yang umum dalam satu atau sekelompok kecil dokumen cenderung memiliki TF-IDF tinggi (Ramos, 2008). Frekuensi kemunculan kata

di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen. Pada Metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai TF dan IDF. Adapaun rumus perhitungan TF-IDF yaitu (Christopher , 2008) :

$$\begin{aligned} Idf(t,d) &= \log \frac{N}{\{d \in N : t \in d\}} \\ &= \log \frac{N}{df} \end{aligned} \quad (1)$$

$$\begin{aligned} W(t,d) &= tf(t,d) \times idf(t,d) \\ &= tf(t,d) \times \left(\log \frac{N}{df} \right) \end{aligned} \quad (2)$$

Dimana :

t : Kata

d : Kalimat

W(t,d) : Bobot kata t pada kalimat d.

Tf(t,d) : Jumlah kemunculan kata t dalam kalimat d.

Idf(t,d) : Pengurangan dominasi kata t yang sering muncul di kalimat d

N : Jumlah semua kalimat yang ada

df : Jumlah kalimat yang mengandung kata t

Bisa disimpulkan bahwa metode ini adalah penjumlahan bobot dari kata kunci yang terkandung. Kemudian, penjumlahan dari bobot dibuat untuk menentukan kemungkinan setiap kata yang diambil kata yang dicari (Sierra, 2008).

2.2.4 Fuzzy C-Means

Model ruang vektor dan pembobotan tf-idf digunakan untuk merepresentasikan nilai numerik dokumen sehingga kemudian dapat dihitung kedekatan antar dokumen. *Fuzzy c-means* merupakan deretan pusat cluster dan beberapa derajat keanggotaan. Berikut adalah langkah penghitungan algoritma fuzzy c-means :

- a) *Input* data yang akan dicluster X , berupa matriks berukuran $n \times m$ (n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, \dots, m$).
- b) Tentukan :
 - Jumlah *cluster* = c ;
 - Pangkat = w ;
 - Maksimum iterasi = MaxIter ;
 - Error terkecil yang diharapkan = ξ ;
 - Fungsi objektif awal = $P_0 = 0$;
 - Iterasi awal = $t = 1$;
- c) Bangkitkan bilangan random μ_{ik} , $i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$; sebagai elemen-elemen matriks partisi awal U . Hitung jumlah setiap kolom :

$$Q_i = \sum_{k=1}^c \mu_{ik}$$

- d) Hitung pusat *cluster* ke- k : V_{kj} , dengan $k = 1, 2, \dots, c$; dan $j = 1, 2, \dots, m$.

$$V_{kj} = \frac{\sum_{i=1}^n ((u_{ik})^w * X_{ij})}{\sum_{i=1}^n (u_{ik})^w}$$

e) Hitung fungsi objektif pada iterasi ke-t, P_t

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (u_{ik})^w \right)$$

f) Hitung perubahan matriks partisi

$$u_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}$$

g) Cek kondisi berhenti :

- Jika : $(|P_t - P_{t-1}| < \xi)$ atau $(t > \text{MaxIter})$ maka berhenti;
- Jika tidak : $t = t + 1$, ulangi langkah ke-4

2.3 Penelitian Terkait

Penelitian yang mempunyai hubungan dengan penelitian ini adalah penelitian oleh Mahasiswa Universitas Dian Nuswantoro (UDINUS) Semarang yaitu Bahar. Penelitiannya adalah Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means. Selain itu ada penelitian yang juga berkaitan dengan penelitian ini yaitu penelitian yang dilakukan oleh dua Mahasiswa Universitas Diponegoro Semarang yaitu Budi Setiyono dan R. Rizal Isnanto. Penelitiannya adalah Pembagian Kelas Kuliah Mahasiswa Menggunakan Algoritma Pengklasteran Fuzzy C-Means. Dalam penelitian ini penulis akan menggunakan metode fuzzy c-means untuk mengelompokkan halaman web dalam bentuk kategori, karena banyak penelitian pengklasteran menggunakan metode ini. Dibandingkan metode lain metode inilah yang dirasa paling efektif dalam pengelompokan.

BAB III

ANALISIS DAN PERANCANGAN SISTEM

Pada bab desain dan implementasi ini akan mengulas tentang tahap yang digunakan dalam penelitian pembuatan pengelompokan halaman web secara otomatis dengan menggunakan metode *fuzzy c-means clustering*.

3.1 Desain Sistem

Pada subbab ini pembahasan akan dititik beratkan pada unsur-unsur yang berkaitan dengan penelitian pengelompokan halaman web secara otomatis seperti downloader, ekstraksi dan membuat pengelompokan yang sesuai dengan isi berita.

3.1.1 Deskripsi Umum Sistem

Aplikasi pengelompokan halaman web yang akan dibuat merupakan sistem yang membuat pengelompokan berdasarkan *content* halaman *web* dan akan melakukan pengelompokan ke dalam kategori berdasarkan ciri-ciri dari *content* berita dari hasil ekstraksi. Karena data yang akan dikelompokkan merupakan dokumen berita online, maka untuk mengunduh halaman situs bisa menggunakan *web crawler* sederhana, dan untuk pengelompokan *content* halaman situs maka menggunakan metode *fuzzy c-means clustering* dan dihitung berdasarkan nilai pembobotan frekuensi kemunculan kata. Alur berjalannya proses pengunduhan dan ekstraksi halaman situs adalah sebagai berikut :

- a. Sistem mengatur antrian *link* yang ada dalam situs yang sudah ditentukan untuk untuk proses pengunduhan
- b. Selanjutnya sistem menyimpan data hasil unduhan dalam *database*

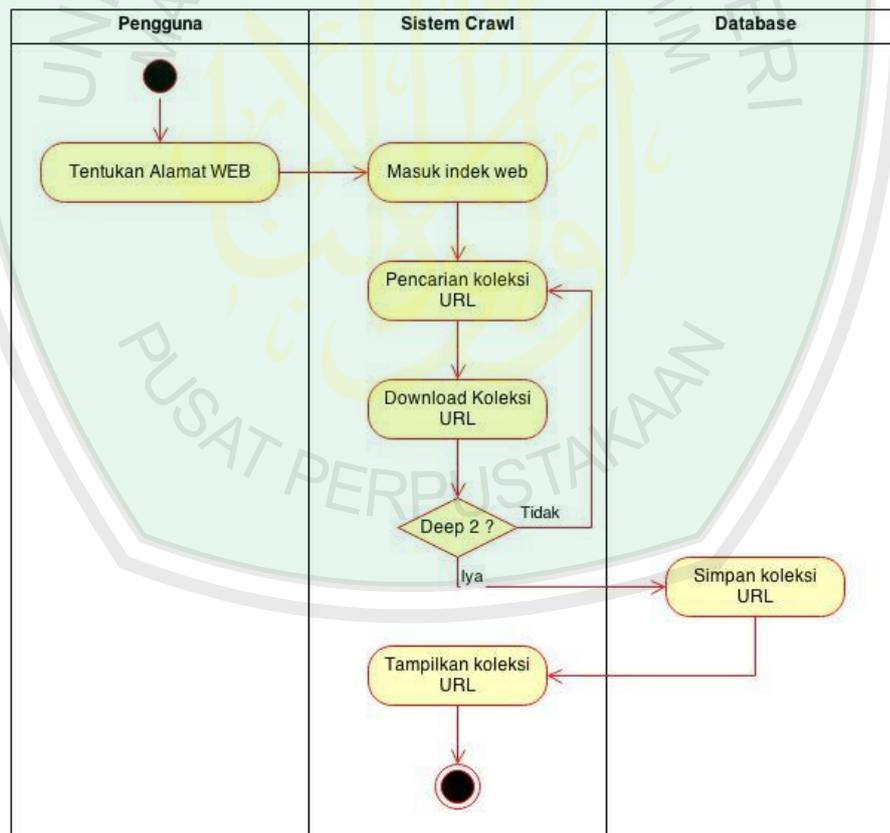
c. Setelah selesai terunduh semua data dari *link* situs tersebut dan disimpan di database, sistem langsung mengekstraksi data untuk diambil kontennya

Proses pengelompokan dibagi menjadi 2 proses, yaitu proses untuk mendapatkan dan mengekstrak halaman web berita dan proses pengelompokan halaman berdasarkan konten hasil ekstraksi halaman berita. Suatu dokumen yang akan diklasifikasi akan melalui beberapa proses berikut :

- a. Pengguna memasukkan *url* situs yang diinginkan untuk diambil datanya dengan *crawler*.
- b. Setelah memasukkan *url yang diinginkan* user kemudian menentukan kedalaman indeks yang akan diunduh.
- c. Kemudian sistem akan memulai proses unduhan dan membuat antrian link data yang ada disitus.
- d. Setelah data berhasil diunduh, sistem mulai proses ekstraksi dokumen halaman situs untuk mendapatkan judul dan konten dan disimpan dalam *database*.
- e. Dari hasil ekstraksi kemudian sistem akan memulai pada proses pembobotan frekuensi kata pada data yang disebut dengan metode tf-idf.
- f. Dan kemudian proses akhir adalah klasifikasi menggunakan metode fuzzy c-means, dengan langkah awal membuat *vector* dari tf-idf yang sudah terbentuk.

3.2 Rancangan *Url Getter*

Tahap ini membahas pengunduhan halaman web secara keseluruhan. Untuk tahapnya bisa dilihat pada **Gambar 3.2**. Proses pertama sistem akan melakukan pengunduhan halaman index yang diinputkan oleh user dan dijadikan dokumen web. Sistem akan menyimpan nama dan URL asli dan dimasukkan ke dalam folder link index. Nama, url index dan folder penyimpanan dimasukkan ke database lalu kemudian link indek akan dianalisa dan akan diambil sublink sampai pada kedalaman sub kedua agar link yang terambil tidak terlalu banyak. Kemudian link tersebut dimasukkan kedalam koleksi link system. Koleksi link system inilah yang akan dicrawl oleh system.



Gambar 3.1 Activity Diagram Url Getter

Pontianak (ANTARA News) - Dirjen Bimas Buddha Dasikin menegaskan bahwa tempat ibadah, khususnya Tempat Ibadah Tri Dharma (TITD) atau yang biasa disebut Klenteng, harus ditata dengan tertib administrasi. Salah satu indikatornya adalah berbadan hukum yayasan keagamaan Buddha, serta terdaftar di Kementerian Agama, dan Kemenhukham. Hal ini disampaikan Dasikin saat membuka Sosialisasi Regulasi Tempat Ibadah Tri Dharma (TITD) Wilayah Kalimantan di Pontianak, Selasa (19/4) malam. Hadir dalam kesempatan ini, Direktur Urusan dan Pendidikan Agama Buddha Paniran, Kakanwil Kemenag Provinsi Kalimantan Barat Syahril Yadi, serta Pembimas Buddha Kalimantan Barat Saryono. Selain tertib administrasi, lanjut Dasikin sebagaimana dikutip dari laman Kemenag, Rabu, rumah ibadah Tri Dharma juga harus dapat menjalankan empat fungsi pokoknya. TITD harus menjadi pusat keagamaan dan meditasi bagi umat Buddha. Lebih dari itu, TITD juga harus bisa menjadi pusat pendidikan dan pengembangan kebudayaan, serta sebagai tempat sosial kemasyarakatan yang nyaman dan kondusif. Kegiatan Sosialisasi Regulasi TITD ini diikuti oleh 50 peserta dari perwakilan pengurus Rumah Ibadah TITD wilayah Kalimantan. Selaku tuan rumah, Kepala Kanwil Kemenag Prov. Kalimantan Barat Syahril Yadi menyampaikan bahwa Singkawang terkenal dengan Kota 1.000 Klenteng. Agar lebih tertib administratif, Syahril Yadi berharap ke depan perlu ditertibkan kembali papan nama atau prasasti Klenteng di Kalbar, tentunya di bawah binaan Ditjen Bimas Buddha. Karenanya, Syahril Yadi memandang kegiatan sosialisasi ini penting untuk peningkatan pemahaman regulasi. Senada dengan Syahril Yadi, Ketua Panitia kegiatan, Parwadi dalam laporannya menyampaikan bahwa sosialisasi regulasi ini bertujuan meningkatkan pemahaman pengurus lembaga/organisasi keagamaan Buddha tentang Peraturan Pemerintah terkait Rumah Ibadah Tri Dharma, pengelolaan dan perlindungan terhadap aset aset, serta penataan administrasi manajemen dan pelayanan pada umat Buddha.

Gambar 3.4 Konten setelah diekstraksi

Untuk memperoleh isi dokumen (konten) system akan menghilangkan seluruh tag HTML dan mengambil baris terpanjang (Simple and fast algorithm) dari kode sumber yaitu **Gambar 3.5**. Tahap ini akan terus dilakukan sampai semua link yang tersimpan dalam system terekstrak semua.

3.4 Tahap Preprocessing

Bagian ini membahas tentang preprocessing yaitu tahap dimana dokumen web akan dihilangkan karakter simbolnya, merubah karakter huruf dijadikan huruf kecil semua, menghilangkan stopwords, memecah dokumen menjadi koleksi kata. Hasil dari preprocessing nantinya akan dirposis ke tahap selanjutnya. Contoh langkah preprocessing adalah sebagai berikut :

- a. Dokumen yang diproses adalah konten pada **Gambar 3.4**.
- b. Isi konten akan dibebaskan dari simbol yang tidak penting seperti terlihat pada **Gambar 3.5**.

Pontianak (ANTARA News) - Dirjen Bimas Buddha Dasikin menegaskan bahwa tempat ibadah, khususnya Tempat Ibadah Tri Dharma (TITD) atau yang biasa disebut Klenteng, harus ditata dengan tertib administrasi. Salah satu indikatornya adalah berbadan hukum yayasan keagamaan Buddha, serta terdaftar di Kementerian Agama, dan Kemenhukham. Hal in disampaikan Dasikin saat membuka Sosialisasi Regulasi Tempat Ibadah Tri Dharma (TITD) Wilayah Kalimantan di Pontianak, Selasa (19/4) malam. Hadir dalam kesempatan ini, Direktur Urusan dan Pendidikan Agama Buddha Paniran, Kakanwil Kemenag Provinsi Kalimantan Barat Syahril Yadi, serta Pembimas Buddha Kalimantan Barat Saryono. Selain tertib administrasi, lanjut Dasikin sebagaimana dikutip dari laman Kemenag, Rabu, rumah ibadah Tri Dharma juga harus dapat menjalankan empat fungsi pokoknya. TITD harus menjadi pusat keagamaan dan meditasi bagi umat Buddha. Lebih dari itu, TTID juga harus bisa menjadi pusat pendidikan dan pengembangan kebudayaan, serta sebagai tempat sosial kemasyarakatan yang nyaman dan kondusif. Kegiatan Sosialisasi Regulasi TTID ini diikuti oleh 50 peserta dari perwakilan pengurus Rumah Ibadah TTID wilayah Kalimantan. Selaku tuan rumah, Kepala Kanwil Kemenag Prov. Kalimantan Barat Syahril Yadi menyampaikan bahwa Singkawang terkenal dengan Kota 1.000 Klenteng. Agar lebih tertib administratif, Syahril Yadi berharap ke depan perlu ditertibkan kembali papan nama atau prasasti Klenteng di Kalbar, tentunya di bawah binaan Ditjen Bimas Buddha. Karenanya, Syahril Yadi memandang kegiatan sosialisasi ini penting untuk peningkatan pemahaman regulasi. Senada dengan Syahril Yadi, Ketua Panitia kegiatan, Parwadi dalam laporannya menyampaikan bahwa sosialisasi regulasi ini bertujuan meningkatkan pemahaman pengurus lembaga/organisasi keagamaan Buddha tentang Peraturan Pemerintah terkait Rumah Ibadah Tri Dharma, pengelolaan dan perlindungan terhadap aset aset, serta penataan administrasi manajemen dan pelayanan pada umat Buddha.

Gambar 3.5 Konten setelah dihilangkan karakter symbol tidak penting

- c. Setelah menghilangkan simbol yang kurang penting kemudian konten dirubah seluruhnya kedalam huruf kecil seperti terlihat pada **Gambar 3.6**.

pontianak (antara news) - dirjen bimas buddha dasikin menegaskan bahwa tempat ibadah, khususnya tempat ibadah tri dharma (titd) atau yang biasa disebut klenteng, harus ditata dengan tertib administrasi. salah satu indikatornya adalah berbadan hukum yayasan keagamaan buddha, serta terdaftar di kementerian agama, dan kemenhukham. hal in disampaikan dasikin saat membuka sosialisasi regulasi tempat ibadah tri dharma (titd) wilayah kalimantan di pontianak, selasa (19/4) malam. hadir dalam kesempatan ini, direktur urusan dan pendidikan agama buddha paniran, kakanwil kemenag provinsi kalimantan barat syahril yadi, serta pembimas buddha kalimantan barat saryono. selain tertib administrasi, lanjut dasikin sebagaimana dikutip dari laman kemenag, rabu, rumah ibadah tri dharma juga harus dapat menjalankan empat fungsi pokoknya. titdharus menjadi pusat keagamaan dan meditasi bagi umat buddha. lebih dari itu, ttid juga harus bisa menjadi pusat pendidikan dan pengembangan kebudayaan, serta sebagai tempat sosial kemasyarakatan yang nyaman dan kondusif. kegiatan sosialisasi regulasi ttid ini diikuti oleh 50 peserta dari perwakilan pengurus rumah ibadah ttid wilayah kalimantan. selaku tuan rumah, kepala kanwil kemenag prov. kalimantan barat syahril yadi menyampaikan bahwa singkawang terkenal dengan kota 1.000 klenteng. agar lebih tertib administratif, syahril yadi berharap ke depan perlu ditertibkan kembali papan nama atau prasasti klenteng di kalbar, tentunya di bawah binaan ditjen bimas buddha. karenanya, syahril yadi memandang kegiatan sosialisasi ini penting untuk peningkatan pemahaman regulasi. senada dengan syahril yadi, ketua panitia kegiatan, parwadi dalam laporannya menyampaikan bahwa sosialisasi regulasi ini bertujuan meningkatkan pemahaman pengurus lembaga/organisasi keagamaan buddha tentang peraturan pemerintah terkait rumah ibadah tri dharma, pengelolaan dan perlindungan terhadap aset aset, serta penataan administrasi manajemen dan pelayanan pada umat buddha.

Gambar 3.6 Konten setelah diubah ke karakter huruf kecil

- d. Konten yang telah diubah menjadi huruf kecil kemudian konten akan dihilangkan *stopwordnya* seperti terlihat pada **Gambar 3.7**.

pontianak news dirjen bimas buddha dasikin ibadah, ibadah tri dharma titd klenteng, ditata tertib administrasi. salah indikatornya berbadan hukum yayasan keagamaan buddha, terdaftar kementerian agama, kemenhukham. in dasikin nbsp membuka sosialisasi regulasi ibadah tri dharma titd wilayah kalimantan pontianak, selasa 19 4 malam. hadir kesempatan ini, nbsp direktur urusan pendidikan agama buddha paniran, kakanwil kemenag provinsi kalimantan barat syahril yadi, pembimas buddha kalimantan barat saryono. tertib administrasi, dasikin dikutip laman kemenag, nbsp rabu, nbsp rumah ibadah tri dharma menjalankan fungsi pokoknya. titd nbsp nbsp pusat keagamaan meditasi umat buddha. itu, ttid nbsp pusat nbsp pendidikan pengembangan nbsp kebudayaan, nbsp sosial kemasyarakatan nyaman kondusif. kegiatan sosialisasi regulasi ttid diikuti 50 peserta perwakilan pengurus rumah ibadah ttid wilayah kalimantan. tuan rumah, nbsp kepala kanwil kemenag prov. kalimantan barat nbsp syahril yadi nbsp singkawang terkenal kota 1.000 klenteng. tertib administratif, syahril yadi berharap nbsp ditertibkan papan nama prasasti klenteng kalbar, nbsp binaan ditjen bimas buddha. karenanya, syahril yadi memandang kegiatan sosialisasi peningkatan pemahaman regulasi. senada syahril yadi, ketua panitia kegiatan, parwadi laporannya sosialisasi regulasi bertujuan nbsp meningkatkan pemahaman pengurus lembaga organisasi keagamaan buddha peraturan pemerintah terkait rumah ibadah tri dharma, pengelolaan perlindungan aset aset, penataan administrasi manajemen pelayanan umat buddha.

Gambar 3.7 1Konten setelah dihilangkan *stopwordnya*

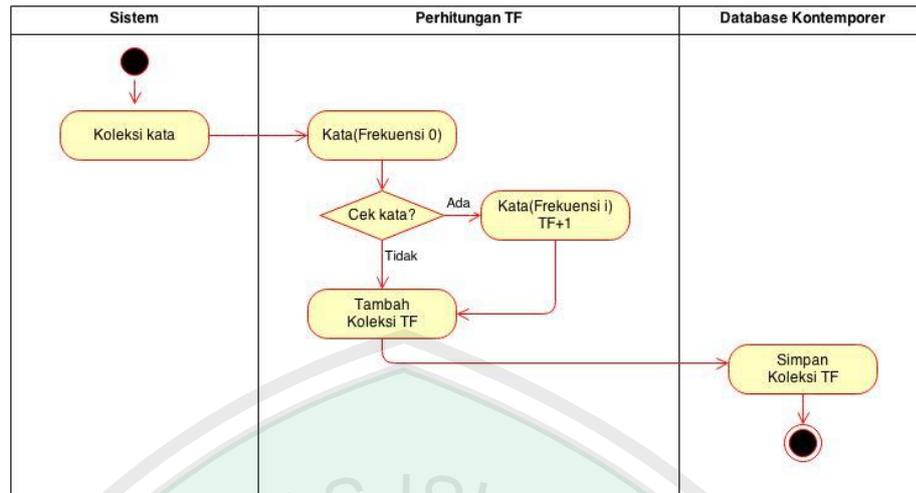
- e. Setelah konten berhasil dihilangkan *stopwordnya* maka hasilnya adalah koleksi kata seperti terlihat pada **tabel 3.1**.

Tabel 3.1 Koleksi kata

Pontianak	Indikatornya	Selasa	Syahril
News	Berbadan	19	Yadi
Dirjen	Hukum	4	Pembimas
Bimas	Yayasan	Malam	Saryono
Buddha	Keagamaan	Hadir	Administrasi
Dasikin	Buddha	Kesempatan	Dikutip
Ibadah	Terdaftar	Urusan	Laman
Tri	Kementerian	Direktur	Rabu
Dharma	Agama	Pendidikan	Rumah
Titd	Membuka	Agama	Menjalankan
Klenteng	Sosialisasi	Paniran	Fungsi
Ditata	Regulasi	Kakanwil	Pokoknya
Tertib	Wilayah	Kemenag	Pusat
Administrasi	Kalimantan	Provinsi	Meditasi
Salah	Pontianak	Barat	Umat
Pengembangan	Rumah	Kebudayaan	Kepala
Social	Kanwil	Kemasyarakatan	Prof
Nyaman	Yadi	Kondusif	Singkawang
Kegiatan	Terkenal	Sosialisasi	Kota
Diikuti	1000	50	Klenteng
Peserta	Administratif	Perwakilan	Berharap
Perwakilan	Berharap	Pengurus	Diterbitkan
Kalimantan	Papan	Tuan	Nama

3.5 Pembobotan Frekuensi Kata

Pada bagian ini dilakukan proses pembobotan pada konten yang telah melalui tahap preprocessing seperti pada **Gambar 3.9**.



Gambar 3.8 2Activity Diagram perhitungan TF

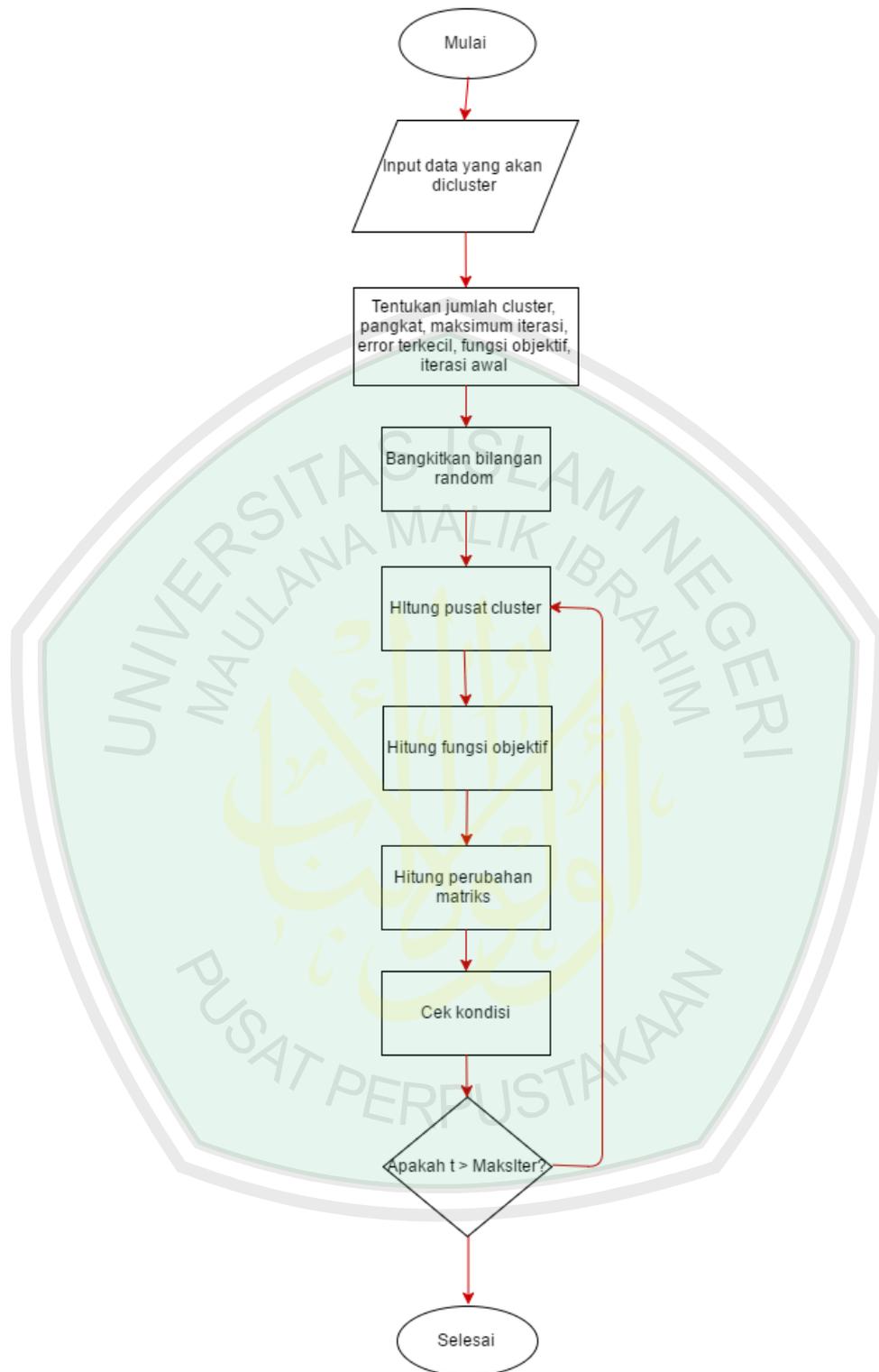
Pada tahap ini system akan melakukan proses pembobotan yang hasilnya akan dimasukkan ke dalam database. Proses yang terjadi adalah system tidak akan mengambil kata yang sudah ada dua kali agar tidak terdapat duplikasi kata. Kata acuan adalah kata yang terdapat pada paragraph pertama. Setelah koleksi kata terbentuk maka koleksi kata akan dihitung terhadap kata yang ada pada paragraph pertama. Setelah selesai maka koleksi kata beserta frekuensinya akan disimpan ke database.

3.6 Perhitungan Fuzzy C-Means

Algoritma *fuzzy c-means* memiliki beberapa komponen yang dibutuhkan, yaitu pusat kelompok, derajat keanggotaan, bobot, objek dan fungsi objek. Proses perhitungan *fuzzy c-means* adalah sebagai berikut :

- 1) Proses pertama adalah *input* data yang akan di *cluster*, data berupa matriks berukuran $n \times m$. Dimana n adalah jumlah sampel data, m adalah atribut setiap data.

- 2) Kemudian dilakukan inisialisasi dengan menentukan jumlah *cluster*, pangkat, maksimum iterasi, *error* terkecil yang diharapkan, fungsi objektif awal, serta iterasi awal.
- 3) Selanjutnya dilakukan proses pembangkitan bilangan random, sebagai elemen-elemen matriks partisi awal.
- 4) Setelah selesai pembangkitan bilangan random maka dilanjutkan dengan menghitung pusat *cluster*.
- 5) Dari hasil perhitungan pusat cluster kemudian menghitung fungsi objektif pada iterasi.
- 6) Setelah ditemukan fungsi objektifnya lalu dilanjutkan dengan menghitung perubahan matriks partisi.
- 7) Langkah terakhir adalah mengecek kondisi berhenti.



Gambar 3.10 Flowchart Fuzzy C-means

3.7 Analisa Kelas

Kelas yang diperlukan dalam pembuatan aplikasi peringkasan ini adalah sebagai berikut:

a. Kelas Proses

Kelas ini memuat antarmuka proses crawling web kemudian memuat proses penentuan link berita terkait dari website portal berita dan memuat proses pembacaan berita hingga proses perhitungan cosine untuk peringkasan tiap berita dari tiap link terkait yang terbaca oleh proses crawler.

b. Kelas Perhitungan

Kelas ini memuat antarmuka proses crawling isi berita dari web kemudian memuat proses pembacaan berita hingga proses perhitungan cosine untuk peringkasan tiap berita dari web.

3.8 Studi Kasus

Terdapat 3 *data* dengan masing-masing memiliki kategori yang unik. Dalam studi kasus ini terdapat koleksi *stop word* sebagai berikut:

Tabel 3.2 Koleksi Stop Word Studi Kasus

A	F	K
B	G	L
C	H	M
D	I	N
E	J	O

Tabel 3.3 Data Studi Kasus

	Kategori	Data Training
D1	X	dAta > A a !mine J Course@ k L
D2	Y	Study# F Text H &mine D. text E mine subField g Data mine^ G o M
D3	Z	Mine& A text interest? n interest

Adapaun langkah pengklasifikasian dokumen baru tersebut sebagai berikut :

1) Menghilangkan simbol yang tidak penting.

```
Untuk Data :
- D1 : dAta A a mine J Course k L
- D2 : Study F Text H mine D text E mine subField g Data
mine G o M
- D3 : Mine A text interest n interest
```

2) Setelah dihilangkan symbol yang tidak penting, selanjutnya mengubah seluruh isi karakter menjadi huruf kecil

```
Untuk Data :
- D1 : data a a mine j course k l
- D2 : study f text h mine d text e mine subfield g data
mine g o m
- D3 : mine a text interest n interest
```

3) Selesai proses mengubah karakter kemudian proses menghilangkan *stop word*

```
Untuk Data :
- D1 : data mine course
- D2 : study text mine text mine subfield data mine
- D3 : mine text interest interest
```

4) Selanjutnya dilakukan proses *tokenizing* (merubah dokumen menjadi koleksi kata)

```
Dokumen dipisah menurut white space (spasi)
Untuk Data :
- D1 : data,mine ,course
- D2 : study,text,mine,text,mine,subfield,data,mine
- D3 : mine,text,interest,interest
```

5) Kemudian menuju proses perhitungan pembobotan *tf* pada data, hasilnya pada **tabel 3.4**.

Tabel 3.4 Hasil Perhitungan tf

D1		D2		D3		Q	
Kata	TF	Kata	TF	Kata	TF	Kata	TF
Data	1	Study	1	Mine	1	data	2
Mine	1	Text	2	Text	1	mine	1
Course	1	Mine	3	Interest	2	course	1
		Subfield	1			from	1
		Data	1			interest	1

6) Setelah proses perhitungan tf kemudian lanjut ke proses pembuatan vektor

index terhadap *data*, hasilnya dapat dilihat pada **tabel 3.5**.

Total Data (n) = 3

$$idf_i = \log\left(\frac{n}{df_i}\right)$$

Tabel 3.5 Pembentukan Vektor Index

Kata	DF	IDF
Course	1	0.477
Data	2	0.176
Interest	1	0.477
Mine	3	0
Study	1	0.477
Subfield	1	0.477
Text	2	0.176

7) Kemudian proses pembentukan model ruang vektor terhadap *data*,

hasilnya ada pada **tabel 3.6**.

$$w_{ij} = tf_{ij} \times idf_i$$

Tabel 3.6 Pembentukan Model Ruang Vektor

D1				D2				D3			
Kata	T F	IDF	W=TF * IDF	Kata	T F	IDF	W=TF * IDF	Kata	T F	IDF	W=TF * IDF
Course	1	0.47 7	0.47 7	Course	0	0.47 7	0	course	0	0.47 7	0
Data	1	0.17 6	0.17 6	Data	1	0.17 6	0.17 6	data	0	0.17 6	0
Intere st	0	0.47 7	0	Intere st	0	0.47 7	0	intere st	2	0.47 7	0.95 4
Mine	1	0	0	Mine	3	0	0	mine	1	0	0
Study	0	0.47 7	0	Study	1	0.47 7	0.47 7	study	0	0.47 7	0
Subfie ld	0	0.47 7	0	Subfie ld	1	0.47 7	0.47 7	subfie ld	0	0.47 7	0
Text	0	0.17 6	0	Text	2	0.17 6	0.35 2	text	1	0.17 6	0.17 6

- 8) Selanjutnya proses normalisasi terhadap model ruang vektor *data*, yang hasilnya ada pada **tabel 3.7**.

$$X_{ij} = \frac{w_{ij}}{\sqrt{\sum_{j=1} w_{ij}^2}}$$

Tabel 3.7 Normalisasi Model Ruang Vektor

D1			D2			D3		
Kata	W=TF * IDF	X	Kata	W=TF * IDF	X	Kata	W=TF * IDF	X
Course	0.477		Course	0		course	0	
Data	0.176		Data	0.176		Data	0	
Interest	0		Interest	0		interest	0.954	
Mine	0		Mine	0		Mine	0	
Study	0		Study	0.477		Study	0	
subfield	0		Subfield	0.477		Subfield	0	
Text	0		Text	0.352		Text	0.176	

9) Lalu akan diteruskan pada proses inialisasi *variable* awal

- Jumlah *cluster* = $c = 2$;
- Pangkat = $w = 2$;
- Maksimum iterasi = $\text{MaxIter} = 50$;
- Error terkecil yang diharapkan = $\xi = 0.05$;
- Fungsi objektif awal = $P_0 = 0$;
- Iterasi awal = $t = 1$;

10) Setelah inialisasi selesai kemudian dilakukan proses pembentukan partisi / vektor keanggotaan awal, dapat dilihat pada **tabel 3.8**.

$$Q_i = \sum_{k=1}^c \mu_{ik}$$

Hitung :

$$u_{ik} = \frac{u_{ik}}{Q_i}$$

Tabel 3.8 Pembentukan Vektor Keanggotaan Awal

U_{ik}	k_1	k_2
i_1	0.535	0.278
i_2	0.342	0.942
i_3	0.499	0.067

11) Setelah itu menuju proses menghitung pusat cluster, hasilnya ada pada **tabel 3.9**.

$$V_{kj} = \frac{\sum_{i=1}^n ((u_{ik})^w * X_{ij})}{\sum_{i=1}^n (u_{ik})^w}$$

Tabel 3.9 Menghitung Pusat Cluster

V_{kj}	j_1	j_2	j_3	j_4	j_5	j_6	j_7
k_1	0.412	0.192	0.375	0	0.109	0.109	0.150
k_2	0.785	0.234	0.004	0	0.559	0.559	0.414

12) Setelah pusat cluster terbentuk maka lanjut ke proses perhitungan fungsi obyektif

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (u_{ik})^w \right)$$

$$P_{\text{next}}: 0.4536886410149075$$

13) Dari proses fungsi objektif kemudian data dibuat untuk menghitung perubahan matriks, hasilnya ada pada **tabel 3.10**.

$$u_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}$$

Tabel 3.10 Perubahan Matriks

U_{ik}	k_1	k_2
i_1	0.761	0.239
i_2	0.014	0.986
i_3	0.738	0.261

14) Cek kondisi berhenti

- Jika : $(|P_0 - P_{\text{next}}| < \xi)$ atau $(t > \text{MaxIter})$ maka berhenti;

$$(|0 - 0.4536886410149075|) = 0.4536886410149075$$

- Jika tidak : $t = t + 1$, ulangi langkah ke-11.

15) Keterangan

Dikasu ini iterasi berulang 3 kali dan menghasilkan pengelompokan *data* dengan selisih nilai objektif sebesar 0.005729942378, hasil ada pada **tabel**

3.11.

Tabel 3.11 Hasil Proses Klasifikasi

Data	Klasifikasi
D1	1
D2	2
D3	1

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas tentang implementasi dari rancangan yang telah dibuat pada bab sebelumnya. Selain itu juga akan dijelaskan analisa dari uji coba aplikasi yang dibuat sehingga akan diketahui apakah aplikasi telah berjalan sesuai dengan yang diharapkan. Baik dari implementasi algoritma maupun kegunaan bagi penggunaanya.

4.1 Implementasi

Implementasi merupakan proses transformasi representasi rancangan ke dalam bahasa pemrograman yang dapat dimengerti oleh komputer. Implementasi sistem juga merupakan sebuah proses pembuatan dan penerapan sistem secara utuh baik dari sisi perangkat keras maupun perangkat lunaknya. Pada subbab implementasi ini akan dibahas hal-hal yang berkaitan dengan implemenatsi sistem pengklasifikasian halaman *web* berdasarkan *content*, yaitu meliputi lingkungan perangkat keras, lingkungan perangkat lunak, implementasi desain antar muka, serta implementasi aplikasi.

4.1.1 Ruang Lingkup Perangkat Keras

Perangkat keras yang digunakan dalam pengembangan aplikasi pengelompokan halaman *web* berdasarkan *content* ini adalah menggunakan piranti sebagai berikut:

- 1) Prosesor Genuine Intel(R) Core(TM) i5-2410M CPU @ 2.30GHz
2.30GHz
- 2) RAM 4 GB
- 3) Hardisk dengan kapasitas 500GB

- 4) Monitor 14"
- 5) Keyboard
- 6) TouchPad/Mouse PS2

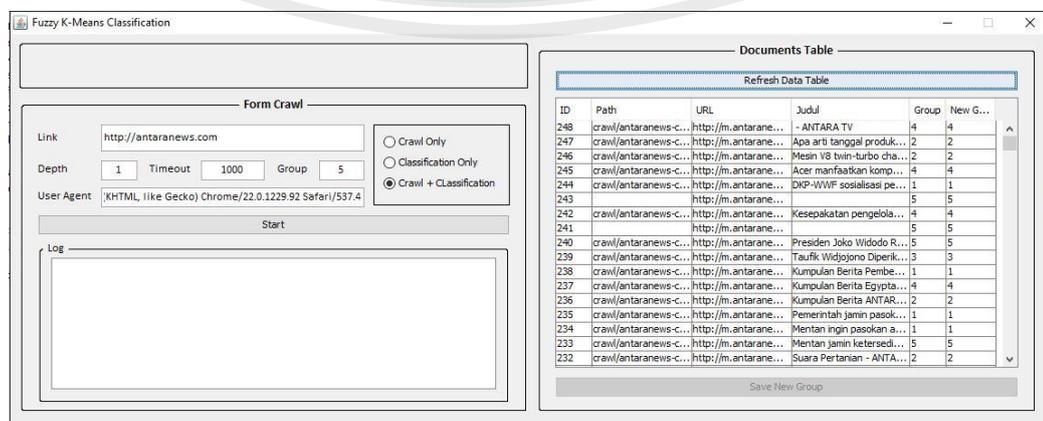
4.1.2 Ruang Lingkup Perangkat Lunak

Perangkat lunak yang digunakan dalam pengembangan aplikasi pengklasifikasian halaman *web* berdasarkan *content* ini adalah sebagai berikut:

- 1) Sistem Operasi Microsoft Windows 10 64-bit
- 2) JDK 1.7.0_51
- 3) JRE 1.7.0_51
- 4) IDE NetBeans 7.0.1
- 5) XAMPP 3.2.2
- 6) Apache 2.4.18
- 7) MySQL 5.0.12
- 8) PHP 7.0.4
- 9) PhpMyAdmin 4.5.1

4.1.3 Implementasi Desain Antarmuka

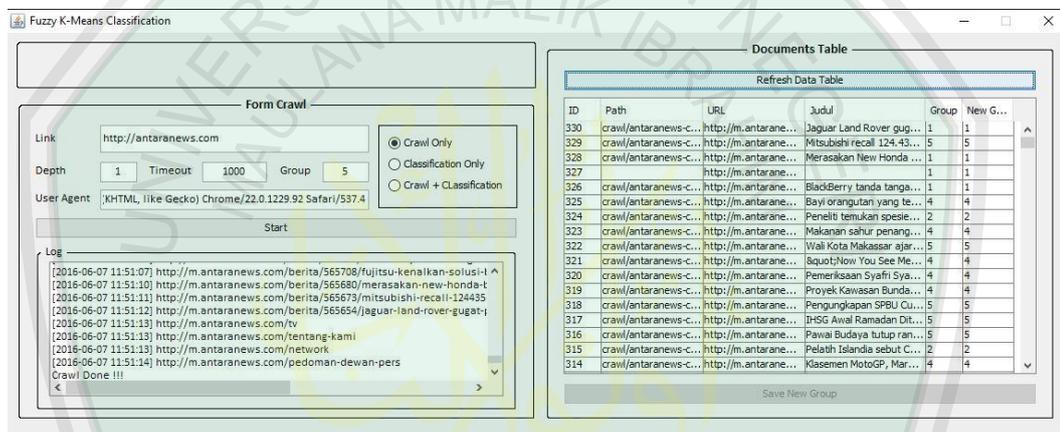
Berikut implementasi desain antarmuka aplikasi pengelompokan halaman *web* berdasarkan *content* :



Gambar 4.1 3Form Fuzzy C-means Classification

Gambar 4.1 merupakan form utama setelah aplikasi dijalankan. Pada form utama terdapat beberapa tombol dan juga radio button. Pertama adalah *radiobutton,button* ini berfungsi untuk menentukan pilihan yang diinginkan sebelum memulai menjalankan aplikasi. Kemudian tombol “Start”, tombol ini berfungsi untuk menjalankan aplikasi tersebut setelah memilih pilihan pada radio *button*.

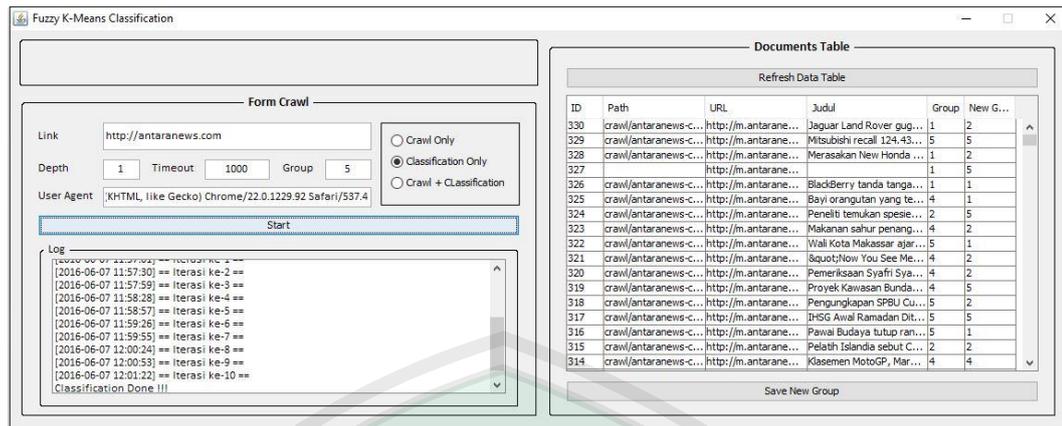
Untuk pilihan *Crawl Only*, maka aplikasi hanya akan mengunduh data yang ada pada *link* yang telah di inputkan, prosesnya dapat dilihat pada **gambar 4.2**.



Gambar 4.2 Form *Crawl*

Disebelah kiri bawah terdapat text area “log” yang berfungsi untuk mengetahui proses secara detail tahap-tahap berjalannya aplikasi. Dan disebelah kanan merupakan table “document” yang berfungsi untuk menampilkan data yang sudah dicrawl oleh aplikasi. Diatas table juga terdapat tombol “refresh” yang berfungsi untuk merefresh jika ada data baru.

Terdapat juga radio button “*Classification Only*” yang berfungsi untuk mengklasifikasikan data yang sudah terdownload dan disimpan dalam database, proses *classification* dapat dilihat pada **gambar 4.3**.



Gambar 4.3 Form *Classification Only*

Disebelah kiri bawah merupakan text area “log” yang berfungsi untuk menampilkan proses berjalannya pengklasifikasian data. Setelah selesai maka data pada table “document” akan berubah berdasarkan proses klasifikasi yang sudah berlangsung, dalam *field* “new group” akan tampak perbedaan dari data sebelum diklasifikasi.

4.1.4 Implementasi Aplikasi

Pada tahap implementasi merujuk dalam bab 3, dimana proses pengklasifikasian dilakukan dengan beberapa tahapan yaitu bermula dengan memasukkan alamat web, koneksi dengan database, cek alamat IP website sesuai apa tidak, mengunduh seluruh *link* yang ada pada halaman *index* (awal) alamat IP website sampai kedalaman yang ditentukan, kemudian ekstraksi seluruh isi (*content*) *file HTML* yang telah diunduh dan disimpan dalam computer yang kemudian disimpan kedalam database, *preprocessing* yang mana akan dilakukan pemecahan teks, lalu perhitungan frekuensi kata dan disimpan ke database, pembentukan vektor, pembentukan ruang vektor sesuai dokumen data training. Perhitungan *fuzzy c-means*, *input* data yang akan *dicluster*, berupa matriks

data. Untuk lebih jelasnya akan dibahas pada bahasan selanjutnya mengenai implementasi dari langkah-langkah tersebut.

4.1.4.1 Implementasi Koneksi Database

Langkah pertama yang dilakukan dalam tahap ini adalah sistem akan berkoneksi dengan database. Seluruh aktivitas yang berhubungan dengan database akan dihubungkan dengan kelas `dbQuery.java`. Konfigurasi database dimasukkan dalam *file* berformat `txt`. Adapun isi dari *file* tersebut sebagai berikut :

Tabel 4.1 Source Code Pengaturan Koneksi Database

```
com.mysql.jdbc.Driver
db_kmeans
root

jdbc:mysql://localhost/
c:/xampp/mysql/bin/mysqldump
```

Baris pertama menunjukkan jenis *driver* program *java* untuk berkoneksi dengan database dengan database *mySQL*. Baris kedua menunjukkan nama database. Baris ketiga menunjukkan *username* database. Baris keempat menunjukkan *password* database. Baris terakhir menunjukkan lokasi database.

Tabel 4.2 Source Code Koneksi dan Olah Data Database

```
public class dbQuery {
    Connection conn;
    Statement st;
    ResultSet rs;
    public String driver,db,username,pass,url;
    String namaFile="koneksi.txt";
    String [] dbArr= new String[5];
    private boolean DatabaseKonek=false;
    public dbQuery(){
        try{
            UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());
        }
        catch(Exception e){}
        baca();
    }
    private void baca() {
        int i=0;
        BufferedReader baca;
        try {
            baca = new BufferedReader(new FileReader(namaFile));
            while(i<dbArr.length){
                String tampung;
                tampung = baca.readLine();
                dbArr[i]=tampung;
                i++;
            }
        }
    }
}
```

```

    }
    if(dbArr[0]==null || dbArr[1]==null || dbArr[2]==null ||
    dbArr[4]==null){
    int cek = Integer.parseInt("fas");
    }
    else{
    baca.close();
    setVarDB(dbArr);
    }
    }
    catch (FileNotFoundException ex) {
    JOptionPane.showMessageDialog(null, "Pengaturan Software Salah !!!" +
    "\nFile '"+namaFile+"' tidak dapat ditemukan. \n\n" +
    "Pesan Kesalahan : \n\n" "+ex+" \n", "Pesan Kesalahan Pengaturan", 0);
    System.exit(0);
    }
    catch (IOException ex) {
    JOptionPane.showMessageDialog(null, "Pengaturan Software Salah !!!" +
    "\nTidak dapat membaca file '"+namaFile+"'. \n\n" +
    "Pesan Kesalahan : \n\n" "+ex+" \n", "Pesan Kesalahan Pengaturan", 0);
    System.exit(0);
    }
    catch (Exception ex) {
    JOptionPane.showMessageDialog(null, "Pengaturan Software Salah !!!" +
    "\nIsi format file '"+namaFile+"' salah. \n\n" +
    "Petunjuk isi format file '"+namaFile+"' : " +
    "\nBaris ke-1 : gunakan value ini jika menggunakan database mySQL => \n
    com.mysql.jdbc.Driver \n (tanpa petik)" +
    "\nBaris ke-2 : nama database Anda." +
    "\nBaris ke-3 : pengguna(username) dari database Anda." +
    "\nBaris ke-4 : kata sandi(password) dari database Anda." +
    "\nBaris ke-5 : default software ini untuk database mySQL => \n
    jdbc:mysql://localhost/ \n (tanpa petik)" +
    "\n" +
    "\n*) Catatan : pastikan Anda HANYA menulis 5 baris diatas."
    , "Pesan Kesalahan Pengaturan", 0);
    System.exit(0);
    }
    }
    private void setVarDB(String [] arr){
    driver=arr[0];
    db=arr[1];
    username=arr[2];
    pass=arr[3];
    url=arr[4]+db+"?user="+username+"&password="+pass;
    }
    public Connection connect(){
    try{
    Class.forName(driver);
    conn=DriverManager.getConnection(url);
    st = conn.createStatement();
    DatabaseKonek=true;
    }
    catch(Exception e){
    DatabaseKonek=false;
    }
    return conn;
    }
    public boolean getKonekStatusDB(){return DatabaseKonek;}
    private String [][] duplicArr(String [][] old , String [] dataBaru){
    String [][] arr = new String[old.length+1][dataBaru.length];
    for(int i=0;i<arr.length;i++){
    for(int j=0;j<arr[0].length;j++){
    if(i==arr.length-1) arr[i][j]=dataBaru[j];
    else arr[i][j]=old[i][j];
    }
    }
    }

```

```

    }
    }
    return arr;
    }
    public String [][] select(String query) {
        String [][] data = null;
        try{
            rs= st.executeQuery(query);
            ResultSetMetaData rsmd = rs.getMetaData();
            int kolom=rsmd.getColumnCount();
            data =new String[0][kolom];
            String [] arr = new String[kolom];
            while(rs.next()){
                for(int i=0;i<arr.length;i++) arr[i] = rs.getString(i+1);
                data = duplicArr(data, arr);
            }
        }
        catch(Exception e){
            e.printStackTrace();
        }
        return data;
    }
    public void cetak(String [][] array){
        for(int i=0;i<array.length;i++){
            for(int j=0;j<array[0].length;j++){
                System.out.print(array[i][j]+"\t");
            }
            System.out.println();
        }
    }
    public boolean iud(String query){
        boolean hasil=true;
        try{
            st.executeUpdate(query);
            hasil=true;
        }
        catch(Exception e){
            hasil=false;
            System.out.println(e);
        }
        return hasil;
    }
    public static void main(String args[]){
        dbQuery d = new dbQuery();
        Connection cd = d.connect();
        String [][] tes=d.select("SELECT * FROM tb_stopword");
        d.cetak(tes);
    }
}

```

Setiap kelas yang ingin berkoneksi dengan database akan memanggil kelas tersebut untuk dijadikan sebuah objek. *Method baca()* merupakan *method* untuk membaca isi dari file “koneksi.txt” yang kemudian isinya ditampung sementara dalam variable *arrayarr[]* yang diproses lewat *method setVarDB()*. Kemudian untuk mengkoneksikan program dengan database menggunakan

method connect() yang dipanggil tiap kelas. Jika ingin mengambil data dari dalam database (*query select*) cukup dengan memanggil *method select()*.

4.1.4.2 Implementasi *Downloader* dan Ekstraksi

Pada tahap ini akan dilakukan proses pengunduhan dari halaman *index website* serta halaman-halaman yang berada satu *link* dengan halaman *index* sesuai dengan kedalaman yang telah ditentukan.

Tabel 4.3 Source Code Penentuan Tempat Hasil Unduhan

```

...
private String siteName(String site){
    String remove = site.toString().replaceAll("http://|https://",
    "");
    String filename = remove.replaceAll("[^A-Za-z0-9]", "-");
    return filename;
}
...
private void CreateFolder(String folder){
    File theDir = new File(folder);
    if (!theDir.exists()) {
        boolean result = theDir.mkdirs();
    }
}
...

```

Langkah awal dalam tahap ini adalah membuat tempat penyimpanan *file* hasil unduhan kedalam *folder*. Nama *folder* sesuai dengan alamat *url* yang menjadi *input* proses menggunakan *method siteName()* dan **CreateFolder()**. Selanjutnya sistem melakukan *download* (unduh) terhadap *file index website* menggunakan *method downloader_page()*.

Tabel 4.4 Source Code Unduh Halaman Indeks Web

```

...
private String downloader_page(String folder, String site, String
customName){
    String letakFile="";
    try{
        TrustManager[] trustAllCerts = new TrustManager[]{
            new X509TrustManager() {
                public java.security.cert.X509Certificate[]
getAcceptedIssuers() {
                    return null;
                }
                public void checkClientTrusted(
                    java.security.cert.X509Certificate[]
certs, String authType) {
                }
            }
        };
    }
}

```

```

        public void checkServerTrusted(
            java.security.cert.X509Certificate[]
certs, String authType) {
        }
    };
    SSLContext sc = SSLContext.getInstance("SSL");
    sc.init(null, trustAllCerts, new
java.security.SecureRandom());

    HttpURLConnection.setDefaultSSLSocketFactory(sc.getSocketFactory(
));
    //////////////////////////////////////
    //////////////////////////////////////
    URL url = new URL(site);
    URLConnection uc = url.openConnection();

    uc.setRequestProperty("User-Agent", hUserAgent);
    uc.addRequestProperty("Accept-Language", "en-US,en;q=0.8");
    uc.addRequestProperty("Referer", "google.com");
    uc.setReadTimeout(TimeOut);

    ReadableByteChannel rbc =
Channels.newChannel(uc.getInputStream());
    String remove =
url.toString().replaceAll("http://|https://", "");
    String filename = remove.replaceAll("[^A-Za-z0-9]", "-");
    if(filename.length() > 100){
        filename.substring(0, 100);
    }
    String to_file=filename.trim();
    if(!customName.equals(""))to_file=customName;

    letakFile=folder+"/"+to_file+".html";
    FileOutputStream fos = new FileOutputStream(letakFile);
    fos.getChannel().transferFrom(rbc, 0, 1 << 24);
}
catch(Exception e){
}
return letakFile;
}
...

```

Alamat akan dicek menggunakan metode *response header* terlebih dahulu kevalidannya. Apakah alamat tersebut diteruskan/diubah (*redirect*) atau tidak menggunakan *method* **checkRealAddress()**. Kemudian aturan penamaan *file* hasil unduhan adalah menghilangkan kata “http://” dan “https://”, menghilangkan seluruh karakter symbol dan yang tersisa hanya karakter *alphabet* dan *numeric* menggunakan *method* **replaceAll()**. Kemudian ditambahkan pula ekstensi yaitu “.html” untuk memberikan jenis bahwa *file* tersebut adalah *file* hasil unduhan *website*.

Tabel 4.5 Source Code Cek Kevalidan Url

```

.....
private String checkRealAddress(String site){
String realAddress="";
try{
    URL url = new URL(site);
    URLConnection uc = url.openConnection();

    InputStream is = uc.getInputStream();
    realAddress = uc.getURL().toString();
    is.close();

    HttpURLConnection http =
(HttpURLConnection)url.openConnection();
    http.setRequestProperty("Accept-Language", "en-US,en;q=0.8");
    http.setRequestProperty("User-Agent", hUserAgent);
    http.setRequestProperty("Referer", "google.com");
    int statusCode = http.getResponseCode();
    if(statusCode==301 || statusCode==302 || statusCode==303){
        realAddress = http.getHeaderField("Location");
        String cookies = http.getHeaderField("Set-Cookie");
        url = new URL(realAddress);

        http = (HttpURLConnection)url.openConnection();
        http.setConnectTimeout(TimeOut);
        http.setRequestProperty("Cookie", cookies);
        http.setRequestProperty("Accept-Language", "en-US,en;q=0.8");
        http.setRequestProperty("User-Agent",hUserAgent);
        http.setRequestProperty("Referer", "google.com");
        System.out.println(realAddress);
    }
}
catch(Exception e){}

return realAddress;
}
.....

```

Setelah *file index* selesai diunduh maka file akan dibaca untuk melihat *link* apa saja yang ada didalam *file index* tadi menggunakan *method crawl()*. Pembacaan berdasarkan kalimat yang memiliki *tag HTML* “a” menggunakan *method ambil_A()*. Kemudian *link* ditentukan dengan cara mengambil kata setelah *tag HTML* “href” menggunakan *method ambil_href()*. Kemudian terdapat pula *method pembenahanURL()* yang fungsinya untuk koreksi terhadap *link website* yang tidak sempurna.

Tabel 4.6 Source Code Pembentukan Koleksi Link

```

...
private List<String> crawl(String file , String link_root){
    List<String> hasil = new ArrayList<String>();
    String fileName = file;
    if(!fileName.equals("")){
        String text="";
        try {
            RandomAccessFile aFile = new
RandomAccessFile(fileName,"r");
            FileChannel inChannel = aFile.getChannel();
            long fileSize = inChannel.size();
            byte[] buffer = new byte[(int) fileSize];

            FileInputStream inputStream = new
FileInputStream(fileName);

            int total = 0;
            int nRead = 0;
            while((nRead = inputStream.read(buffer)) != -1) {
                text+=new String(buffer);
                total += nRead;
            }
            inputStream.close();
        }
        catch(FileNotFoundException ex) {
            System.out.println("Unable to open file '" +
fileName + "'");
        }
        catch(IOException ex) {
            System.out.println("Error reading file '" +
fileName + "'");
        }
        hasil = ambil_A(text,link_root);
    }
    return hasil;
}
...
private List<String> ambil_A(String text, String link_root){
    List<String> hasil = new ArrayList<String>();
    String tag="a";
    for(int i=0 ;i<text.length();i++){
        if (String.valueOf(text.charAt(i)).equals("<") &&
i!=text.length()-1){
            int a=i;
            int b=0;
            boolean iya=false;
            for(int j=a;j<text.length();j++){

                if (String.valueOf(text.charAt(j)).equals(">")){
                    b=j;
                    iya=true;
                    break;
                }
            }
            if(iya){
                String [] cek = kl.splitTanda(" ",
kl.ambilText(a+1, b-1, text));
                if(cek[0].toLowerCase().equals(tag)){
                    i=b+1;
                    String abc =
ambil_href(kl.ambilText(a, b, text),link_root);
                    if(!kl.SelectFromList(hasil, abc) &&
!kl.SelectFromList(list_link, abc)) hasil.add(abc);

```

```

    }
    }
    }
    return hasil;
}
...
private String ambil_href(String text, String link_root){
    String abc="";
    String [] arr = kl.splitTanda(" ", text);
    for(int i=1;i<arr.length;i++){
        if(arr[i].length()>=4 &&
String.valueOf(arr[i].charAt(0)).toLowerCase().equals("h") &&
String.valueOf(arr[i].charAt(1)).toLowerCase().equals("r") &&
String.valueOf(arr[i].charAt(2)).toLowerCase().equals("e") &&
String.valueOf(arr[i].charAt(3)).toLowerCase().equals("f")
){
    if(arr[i].length()==4) abc=arr[i+2];
    else if(arr[i].length()==4) abc=arr[i+1];
    else{
        String [] temp = kl.splitTanda("=", arr[i]);
        abc=temp[1];
    }
    if(!abc.equals("")){
        String [] temp;

        if(String.valueOf(abc.charAt(0)).equals("\\")){
            abc=kl.strHapusKarakterAwal("\\"", "");
            temp = kl.splitTanda("\"", abc);
        }

        if(String.valueOf(abc.charAt(0)).equals("\'"))
            temp = kl.splitTanda("\'", abc);
        else
        if(String.valueOf(abc.charAt(0)).equals("\""))
            temp = kl.splitTanda("\"", abc);
        else{
            temp = new String[1];
            temp[0]=abc;
        }
        abc=pembenahanURL(temp[0], link_root);
    }
    break;
}
}
return abc;
}
...
private String pembenahanURL(String text , String link_root){
    String [] cek = kl.splitTanda(":", text);
    if (String.valueOf(cek[0].charAt(0)).equals("/")){
        text=link_root+text.substring(1);
    }
    else if (cek[0].equals("http") || cek[0].equals("https")){
        if(!String.valueOf(cek[1].charAt(cek[1].length()-
1)).equals("/")) text+="/";
    }
    else{
        text=link_root+text;
    }
}

```

```

        return text;
    }
    ...
private void a_list_all_depth(/*List<String> link_a , */int depth,
String folderSimpan, String real_ip){
    int [] tanda = new int[depth];
    int dp_now=0;
    for(int i=0;i<list_link.size();i++){
        if(!move) break;
        String file=downloader_page(folderSimpan,
list_link.get(i), "");
        if(move) {
            if(!kl.SelectFromList(list_link_db ,
list_link.get(i))){
                new extract(list_link.get(i), file, group);
                refresh_table();
            }
            toLog(list_link.get(i));
        }
        if(dp_now+1<depth || i==0){
            List<String> link_a=crawl(file , list_link.get(i));
            list_link.addAll(link_a);
            list_link=filter_list_link(list_link , real_ip);
        }
        if(i==0){ tanda[dp_now]=list_link.size(); }
        if(i==tanda[dp_now] ){
            dp_now++;
            if(dp_now==depth) break;
            else tanda[dp_now]=list_link.size();
        }
    }
}
    ...

```

Setelah mendapatkan seluruh *link* yang terdapat didalam *file index*, berikutnya sistem akan mulai mengunjungi satu per satu daftar *link* dilanjutkan dengan mengunduh *file* menggunakan *method crawl()*, serta mengecek kembali apakah halaman tersebut memiliki *link* juga. Kemudian sistem membaca content dengan menggunakan *method Ambil_A()* dan *method ambil_href()*. Serta sistem akan langsung melakukan pembenahan *link* dengan menggunakan *method pembenahanURL()*. Proses ini berulang sesuai dengan kedalaman yang ditentukan oleh pengguna. Aturankedalaman dimasukkan menggunakan *method all_list_all_depth()*.

Selanjutnya pada tahap ini akan dilakukan proses ekstraksi isi (*content*) dari *file HTML*. Metode ekstraksi yang peneliti pakai adalah mengambil teks pada baris tertentu yang memiliki jumlah karakter terpanjang.

Tabel 4.7 Source Code Ekstraksi Judul dan Isi (Content)

```

...
private String bacaFile(String files){
    String text="";
    try {
        RandomAccessFile aFile = new RandomAccessFile(files,"r");
        FileChannel inChannel = aFile.getChannel();
        long fileSize = inChannel.size();
        byte[] buffer = new byte[(int) fileSize];
        FileInputStream inputStream = new FileInputStream(files);
        int total = 0;
        int nRead = 0;
        while((nRead = inputStream.read(buffer)) != -1) {
            text+=new String(buffer);
        }
        inputStream.close();
    }
    catch(Exception ex) {
        ...
    }
    return text;
}
...
private String title_web(String web){
    String a="";
    Pattern gTitle = Pattern.compile("<title>(.*?)</title>");
    Matcher mjudul = gTitle.matcher(web);
    if(mjudul.find()){
        a = mjudul.group(1);
    }
    a = a.replaceAll("'", "&#039;");
    a = a.replaceAll("\\t+", "");
    a = a.replaceAll(" +", " ");
    return a;
}
...
private String simpleNfast(String web){
    web = web.replaceAll("\\<.*?>", " ");
    web = web.replaceAll("\\t+", "");
    web = web.replaceAll(" +", " ");
    String [] arr = web.split("\n");
    String content="";
    if(arr.length!=0){
        ArrayList<String> ListContent = new ArrayList<String>();
        int maxChar=0 , idChar=0;
        for(int q=0;q<arr.length;q++){
            String text= kl.strHapusKarakterAwal(" ", "", arr[q]);
            if(arr.length!=0){
                ListContent.add(text);
                if(text.length()>=maxChar) {
                    maxChar=text.length();
                    idChar=q;
                }
            }
        }
        content=ListContent.get(idChar);
    }
    return content;
}

```

Langkah pertama adalah *file HTML* dibaca menggunakan *method bacaFile()*. Hasil dari pembacaan *file* disebut dengan dokumen oleh penulis. Kemudian proses

berikutnya mencari *tag HTML* “title” untuk menentukan judul dari dokumen menggunakan *method title_web()*. Didalam *method title_web()* dokumen akan dibagi menjadi tiga bagian yaitu bagian sebelum kata “<title>”, bagian sesudah kata “</title>”, dan bagian diantara kata “<title>” dan “</title>” menggunakan *method compile()*. Judul dari dokumen terletak pada bagian ketiga yaitu bagian diantara kata tadi. Setelah itu judul akan dikoreksi dengan membuang karakter \t dan karakter “ ” (spasi ganda) .

Langkah kedua dokumen dicek ulang menggunakan *method simpleNfast()*. Didalamnya dokumen akan dihilangkan *tag HTML*, membuang karakter \t, dan karakter “ ” (spasi ganda). Setelah itu dokumen akan dipisah berdasarkan karakter \n (ganti baris) menggunakan *method split()* menjadi kumpulan baris dan disimpan dalam koleksi baris. Kemudian setiap baris akan dicek panjang karakternya. Disini akan diambil baris yang memiliki karakter terpanjang untuk menentukan isi dari dokumen tersebut.

4.1.4.3 Implementasi Text Preprocessing

Pada tahap ini akan dilakukan pemecahan dokumen menjadi kata. Serta menghitung frekuensi kata pada masing-masing dokumen. Sebelum melakukan pemecahan dokumen menjadi kumpulan kata maka dokumen harus melewati langkah *preprocessing*. Diantaranya merubah seluruh karakter menjadi *lowercase* (huruf kecil), penghilangan simbol yang dianggap tidak penting, menghilangkan *stopword*.

Tabel 4.8 Source Code Bentuk Koleksi Stop Word dan Stop Word Removal

```

...
private String [][] StopWordList;
...
private void setStopWordList(){
    StopWordList=dq.select("SELECT sw_list FROM tb_stopword ORDER BY
        id_sw ASC", 1);
}
...
String bersih = text.replaceAll("[^A-Za-z0-9.,]", " ");
bersih = bersih.replaceAll(" +", " ");
...
private String preprocessing(String text , String [][] StopWord){
    String a=text.replace("\t", "");
    a = a.toLowerCase();
    for(int i=0;i<StopWord.length;i++){
        a = a.replace(" "+StopWord[i][0]+" ", " ");
    }
    return a;
}
...

```

Langkah pertama adalah mengambil data *stopword* dari dalam database ke dalam koleksi *stopword* menggunakan *method* **setStopWordList()**. Sebelum masuk *method* **preprocessing()**, isi dokumen dihilangkan terlebih dahulu symbol yang tidak penting. Kemudian didalam *method* **preprocessing()**, seluruh isi dokumen diganti menjadi huruf kecil menggunakan *method* **toLowerCase()**. Setelah seluruh dokumen diganti ke huruf kecil kemudian seluruh *stopword* yang sesuai dengan koleksi *stopword* yang ada didalam dokumen dihapus sesuai menggunakan *method* **replace()**. Kemudian seluruh data mulai dari judul, isi dokumen, hasil *preprocessing*, di simpan kedalam database.

Langkah berikutnya adalah memotong dokumen menjadi kumpulan kata dengan menggunakan *method* **split()**. Pemisahan kata berdasarkan karakter “ ” (spasi) dan disimpan ke dalam koleksi kata.

Tabel 4.9 Source Code Simpan Bobot tf ke Database

```

...
private void saveTF_Dokumen(String id_Dokumen){
    String [][] dokumen = dq.select("SELECT ringkasan_dok FROM tb_dokumen
        WHERE id_dok='"+id_Dokumen+"'", 1);
    String [] doc = dokumen[0][0].split(" ");
    ...
    String [][] index = new String[0][0];
    for(int i=0;i<doc.length;i++){
        if(index.length==0){
            String [] DataBaru = {doc[i],"1"};
            index = kl.duplicArrDobel(index, DataBaru);
        }
        else{
            int temp=index.length;
            boolean isis=true;
            int ada=0;
            for(int j=0;j<temp;j++){
                if(index[j][0].equals(doc[i])){
                    isis=false;
                    ada=j;
                    break;
                }
            }
            else isis=true;
        }
        if(isis){
            String [] DataBaru = {doc[i],"1"};
            index = kl.duplicArrDobel(index, DataBaru);
        }
        else
            index[ada][1]=String.valueOf
                (Integer.parseInt(index[ada][1])+1);
        }
    }
    String [][] cek = dq.select("SELECT id_tf FROM tb_tf_dok WHERE
        id_dok='"+id_Dokumen+"'", 1);
    if(cek.length==0){
        for(int i=0;i<index.length;i++){
            dq.iud("INSERT INTO tb_tf_dok VALUES
                (null, '"+id_Dokumen+"', '"+index[i][0]+'', '"+index[i][1]+'')");
        }
    }
}
...

```

Setelah dokumen diubah menjadi koleksi kata maka langkah berikutnya adalah menghitung frekuensi kata kemudian disimpan dalam database menggunakan *method saveTF_Dokumen()*.

4.1.4.4 Implementasi Klasifikasi Menggunakan Algoritma *Fuzzy C-means*

Pada tahap ini akan dilakukan proses pengklasifikasian dokumen yang telah disimpan dalam database. Adapun langkah- langkah pengklasifikasian meliputi input data yang akan *dicluster*, tentukan (jumlah *cluster*, pangkat, maksimum iterasi, error terkecil, fungsi objektif, iterasi awal), bangkitkan

bilangan random, hitung pusat *cluster*, hitung fungsi objektif, hitung perubahan matriks, cek kondisi. Langkah pertama menentukan inisialisasi kebutuhan *variable* awal. Setelah itu vector data dibentuk melalui proses tf-idf dengan menggunakan *method* **CreateTableVektorData()**.

Tabel 4.10 Source Code Pembentukan vector data dan inisialisasi variable awal

```

...
private String use_tb="train_";
private String [][] dokumen_data;
private String [] idx_dok;
private double [][] data;
private double [][] keanggotaan;
private int jumlah_cluster;
private int bobot=2;
private int max_loop=50;
private double stopped_mse=0.05;
private double p_awal=0;
private double [][] pusat_cluster;
...
private void main_proses(int group){
    this.jumlah_cluster=group;
    dokumen_data = dq.select("SELECT * FROM "+use_tb+"dokumen WHERE
status='1' ORDER BY id_dok");
    idx_dok=new String[dokumen_data.length];
    for(int i=0;i<dokumen_data.length;i++)
idx_dok[i]=dokumen_data[i][0];
listWord = dq.select("SELECT DISTINCT(b.word) FROM "+use_tb+"dokumen a ,
"+use_tb+"tf_dok b WHERE a.id_dok=b.fid_dok AND a.status='1' ORDER BY
b.word");
    word_index=word_index();
    keanggotaan=new double[dokumen_data.length][jumlah_cluster];

    data=CreateTableVektorData();
    ...
}
...
private double [][] CreateTableVektorData(){
    double [][] tabel = new double[0][0];
    double [] blank_word=new double[listWord.length];
    for(int i=0;i<listWord.length;i++) {
        blank_word[i]=0;
    }

    for(int i=0;i<dokumen_data.length;i++){
        double [] arr_tf=new double[blank_word.length];
        System.arraycopy( blank_word, 0, arr_tf, 0, blank_word.length
);

        String [][] frek_word = dq.select("SELECT * FROM
"+use_tb+"tf_dok WHERE fid_dok='"+dokumen_data[i][0]+''' ORDER BY word");
        for(int j=0;j<frek_word.length;j++){
            int idx=word_index.get(frek_word[j][2]);
            arr_tf[idx]=Double.parseDouble(frek_word[j][3]);
        }
        tabel = kl.duplicArrDobel(tabel, arr_tf);
    }
    return tabel;
}
...

```

Langkah kedua membangkitkan bilangan random sebagai elemen-elemen matriks partisi menggunakan *method* **CreateKeanggotaanAwal()**.

Tabel 4.11 Source Code Membangkitkan bilangan random

```

...
private double [][] CreateKeanggotaanAwal(double [][] source){
    for(int i=0;i<dokumen_data.length;i++){
        int idx_kategori=-1;
        int random=0;
        for(int j=0;j<jumlah_cluster;j++){

            if(j+1==Double.parseDouble(dokumen_data[i][6]))
idx_kategori=j;
                int bangkit_rand=(int) (Math.random()*999)+1;
                source[i][j]=Double.parseDouble(String.valueOf(bangkit_ran
d))/1000;
                if(random<bangkit_rand) random=bangkit_rand;
            }
            if(idx_kategori!=-1)
source[i][idx_kategori]=Double.parseDouble(String.valueOf(((int)
(Math.random()*(999-random))+1+random))/1000;
            idx_kategori=-1;
            random=0;
        }
        return source;
    }
...

```

Langkah ketiga hitung pusat cluster menggunakan *method* **Hitung_variable_cluster()**.

Tabel 4.12 Source Code Hitung pusat cluster

```

...
private void Hitung_variable_cluster(){
    for(int a=0;a<pusat_cluster.length;a++){
        for(int b=0;b<pusat_cluster[0].length;b++){
            double atas=0;
            double bawah=0;
            for(int c=0;c<data.length;c++){ //iterasi sesuai
jumlah data
                atas+=Math.pow(keanggotaan[c][a],
2)*data[c][b];
            }
            for(int d=0;d<data.length;d++){ //iterasi sesuai
jumlah data
                bawah+=Math.pow(keanggotaan[d][a], 2);
            }
            pusat_cluster[a][b]=atas/bawah;
        }
    }
}
...

```

Langkah keempat adalah menghitung fungsi objektif pada iterasi dengan menggunakan *method* **Hitung_fungsi_objektif()**.

Tabel 4.13 Source Code Menghitung Fungsi Objektif

```

...
private double Hitung_fungsi_obyektif(){
double p=0;
for(int a=0;a<data.length;a++){
double L=0;
for(int b=0;b<pusat_cluster.length;b++){
double jarak=0;
for(int c=0;c<data[0].length;c++){
jarak+=Math.pow((data[a][c]-
pusat_cluster[b][c]), 2);
}
L+=Math.pow(jarak*keanggotaan[a][b],2);
}
p+=L;
}
return p;
}
...

```

Langkah kelima adalah menghitung perubahan matriks partisi dengan menggunakan *method* **next_keanggotaan()**.

Tabel 4.14 Source Code Hitung Perubahan Matriks Partisi

```

...
private void next_keanggotaan(){
for(int a=0;a<data.length;a++){
for(int b=0;b<pusat_cluster.length;b++){
double var_atas=0;
double var_bawah=0;
for(int c=0;c<data[0].length;c++){
var_atas+=Math.pow((data[a][c]-pusat_cluster[b][c]), 2);
}
var_atas=Math.pow(var_atas, (-1/(bobot-1)));

for(int d=0;d<pusat_cluster.length;d++){
double jarak=0;
for(int c=0;c<data[0].length;c++){
jarak+=Math.pow((data[a][c]-
pusat_cluster[d][c]), 2);
}
var_bawah+=Math.pow(jarak, (-1/(bobot-1)));
}
keanggotaan[a][b]=var_atas/var_bawah;
}
}
}
...

```

Langkah terakhir adalah mengecek kondisi berhenti system dengan menggunakan *method* **proses()**.

Tabel 4.15 Source Code Cek Kondisi Berhenti

```

...
private void proses(){
for(int i=1;i<=max_loop;i++){
...

for(int a=0;a<keanggotaan.length;a++){
    double large=0;
    int idx=0;
    for(int b=0;b<keanggotaan[0].length;b++){
        if(large<keanggotaan[a][b]){
            large=keanggotaan[a][b];
            idx=b;
        }
    }
    System.out.println("D"+(a+1)+" : "+(idx+1));
}
double obyektif=Math.abs(p_awal-next_p);
System.out.println("|Pbefore - Pnext | = | "+p_awal+" -
"+next_p+" | = "+obyektif);
p_awal=next_p;
if(obyektif<=stopped_mse) break;
}
}
...

```

4.2 Hasil Uji Coba

Pada subbab ini akan disajikan hasil uji coba yang dilakukan terhadap sistem dan pembahasan mengenai hasil evaluasi dari klasifikasi yang dihasilkan sistem.

4.2.1 Uji Coba

Pada subbab ini pengujian dilakukan terhadap aplikasi pengelompokan halaman *web* berdasarkan *content* dengan penerapan algoritma *Fuzzy C-means*.

Sistematis pengujian sebagai berikut:

- 1) Menginputkan alamat situs yang ingin dicrawl, disini menggunakan contoh www.antaraneews.com, diakses pada tanggal 8 juni 2016 pada pukul 02.00.
- 2) Menentukan kedalaman *link* = 1
- 3) Memilih beberapa pilihan radiobutton *crawl* dan *classification*

- 4) Selanjutnya mengklik tombol *start* untuk memulai proses aplikasi
- 5) Proses berjalan sampai tidak ada *link* lagi pada halaman *index* dengan kedalaman 1
- 6) Hasil dari proses *crawl* dan *classification* ditampilkan dalam bentuk table pada sisi kanan aplikasi

Setelah melalui proses yang mungkin cukup lama maka akan dihasilkan data seperti yang ditampilkan dalam **tabel 4.16** berikut ini. Hasil *crawl* dan *classification* aplikasi disimpan dalam bentuk data tabel yang dapat dilihat pada **tabel 4.16**.



Tabel 4.16 Hasil Crawl dan Klasifikasi

No	Daftar halaman	Klasifikasi
1	ANTARA News : Portal Berita Indonesia	4
2	Top News - ANTARA News	4
3	Berita Terkini - ANTARA News	4
4	Rilis Pers - ANTARA News	2
5	RSS ANTARA News	1
6	Ketentuan Penggunaan - Antara News	1
7	Tentang Kami - ANTARA News	4
8	Nasional - ANTARA News	2
9	Nasional Umum - ANTARA News	4
10	Nasional Polkam - ANTARA News	2
11	Nasional Hukum - ANTARA News	2
12	Nasional Pendidikan - ANTARA News	2
13	Nasional Kesehatan - ANTARA News	2
14	Nasional Naker - ANTARA News	1
15	Nasional Unik - ANTARA News	1
16	Internasional - ANTARA News	2
17	Internasional Asiaoseania - ANTARA News	2
18	Internasional Amerika - ANTARA News	1
19	Internasional Eropa - ANTARA News	1
20	Internasional Timur Tengah - ANTARA News	4
21	Internasional Afrika - ANTARA News	1
22	Ekonomi - ANTARA News	2

23	Ekonomi Makro - ANTARA News	2
24	Ekonomi Moneter - ANTARA News	2
25	Ekonomi Bisnis - ANTARA News	2
26	Ekonomi Bursa - ANTARA News	2
27	Olahraga - ANTARA News	2
28	Olahraga Sepakbola - ANTARA News	2
29	Olahraga Tenis - ANTARA News	1
30	Olahraga Bulutangkis - ANTARA News	2
31	Olahraga Basket - ANTARA News	1
32	Olahraga Tinju - ANTARA News	1
33	Olahraga Balap - ANTARA News	2
34	Olahraga Umum - ANTARA News	2
35	Hiburan - ANTARA News	2
36	Hiburan Sinema - ANTARA News	2
37	Hiburan Musik - ANTARA News	2
38	Hiburan Selebriti - ANTARA News	2
39	Hiburan Seni Budaya - ANTARA News	2
40	Hiburan Gaya Hidup - ANTARA News	2
41	Teknologi - ANTARA News	2
42	Teknologi Sains - ANTARA News	2
43	Teknologi Komputer - ANTARA News	2
44	Teknologi Gadget - ANTARA News	2
45	Teknologi Internet - ANTARA News	1

46	Warta Bumi - ANTARA News	1
47	Warta Bumi Pemanasan Global - ANTARA News	1
48	Warta Bumi Flora Fauna - ANTARA News	2
49	Artikel - ANTARA News	1
50	Artikel Komentar - ANTARA News	2
51	Artikel Opini - ANTARA News	1
52	Artikel Buku - ANTARA News	1
53	Artikel Jeda - ANTARA News	1
54	Artikel Visi - ANTARA News	1
55	Galeri Foto ANTARA News	1
56	ANTARA TV - ANTARA News	1
57	ANTARA News : Portal Berita Indonesia	4
58	Polda Metro ringkus pemutilasi wanita hamil - ANTARA News	2
59	Harga minyak dunia melonjak dipicu penurunan produksi AS - ANTARA News	2
60	Madrid bekuk Villarreal tiga gol tanpa balas - ANTARA News	3
61	MU atasi Palace 2-0, dekati zona Liga Champions - ANTARA News	3
62	Wapres Buka Innacraft 2016 - Foto ANTARA News	1
63	Suarez caturgol dan tiga assist, Barcelona libas Deportivo 8-0 - ANTARA News	3
65	Mari momok Darurat Narkoba itu kita jadikan momentum bersih- bersih - ANTARA News	2
66	Presiden Jokowi tiba di Brussels - ANTARA News	2

67	Info Keagamaan - ANTARA News	1
68	Kemenag upayakan kuota haji ditambah 10 ribu - ANTARA News	4
69	Tempat ibadah Tri Dharma Buddha harus tertib administrasi - ANTARA News	1
70	Kumpulan Fokus Berita - ANTARA News	4
71	Fokus Reklamasi Teluk Jakarta - ANTARA News	4
72	Fokus Resuffle Kabinet - ANTARA News	4
73	Industry Update - ANTARA News	1
74	Indonesia genjot industri tekstil pada perundingan kelima D8 - ANTARA News	2
75	Eksistensi industri kerajinan jaga regenerasi perajin - ANTARA News	2
76	Empat kebijakan Kemenperin dongkrak industri MRO - ANTARA News	2
77	Info Perdagangan - Ekonomi - Antara News - ANTARA News	1
78	INACRAF 2016 momentum peningkatan ekspor kerajinan - ANTARA News	2
79	ITPC Budapest promosikan potensi Indonesia - ANTARA News	2
80	ITPC : Pameran Lagos buka peluang transaksi potensial - ANTARA News	2
81	Pasar saham Australia dibuka lebih tinggi - ANTARA News	2
82	Bursa saham Tiongkok dibuka lebih rendah - ANTARA News	2
83	19 mahasiswa Indonesia dievakuasi ke Fukuoka - ANTARA News	4
84	Semalaman diguyur hujan, genangan landa sejumlah titik Jakarta -	1

	ANTARA News	
85	Saham Tokyo dibuka melonjak didukung kenaikan minyak - ANTARA News	2
86	Pasar rakyat kuliner semarakkan HUT Depok - ANTARA News	4
87	5 kiat menjaga ban tetap awet - ANTARA News Otomotif	1
89	Ruas jalan Beteleme-Kolonodale Morowali Utara putus - ANTARA News	4
90	Obama bertemu Raja Saudi sebelum pertemuan puncak AS-GCC - ANTARA News	4
91	30 cedera dalam ledakan kompleks petrokimia di Meksiko - ANTARA News	1
92	Polda Metro ringkus pemutilasi wanita hamil - ANTARA News	2
93	Pasar saham Spanyol ditutup naik 1,96 persen - ANTARA News	2
94	Indeks FTSE 100 Inggris berakhir naik 0,08 persen - ANTARA News	2
95	Samsung rilis varian Pink Gold Galaxy S7/ S7 edge - ANTARA News	2
96	WNI korban gempa Kumamoto tiba di Tokyo - ANTARA News	4
97	Kenaikan harga minyak angkat saham-saham AS - ANTARA News	2
98	Kartini masa kini dimata Krisdayanti - ANTARA News	2
99	Skyland Petroleum pasok minyak dan gas ke Tiongkok - ANTARA News	2
100	Pekan ke-34 La Liga Spanyol, persaingan juara kian ketat - ANTARA News	3

101	Dolar AS menguat didukung data ekonomi positif - ANTARA News	2
102	Emas relatif stabil di tengah penguatan dolar AS - ANTARA News	2
103	Kalah 1-2 dari Udinese, Fiorentina gagal masuk empat besar - ANTARA News	3
104	Wall Street reli di tengah laporan laba dan kenaikan minyak - ANTARA News	2
105	Harga minyak dunia melonjak dipicu penurunan produksi AS - ANTARA News	2
106	Atasi Heerenveen 2-0, Ajax kembali puncak Eredivisie - ANTARA News	3
107	Totti bawa Roma berbalik kalahkan Torino 3-2 - ANTARA News	3
108	Inter tersandung di kandang Genoa, kalah 0-1 - ANTARA News	3
109	Ringkasan hasil pekan ke-34 Liga Italia Seri A - ANTARA News	3
110	Madrid bekuk Villarreal tiga gol tanpa balas - ANTARA News	3
111	Atasi Sochaux 1-0, Marseille ke final Piala Prancis - ANTARA News	3
112	Valencia kuliti Eibar 4-0, trigol Alcacer sempurna - ANTARA News	3
113	West Ham menang meyakinkan 3-1 lawan Watford - ANTARA News	3
114	MU atasi Palace 2-0, dekati zona Liga Champions - ANTARA News	3
115	Liverpool cukur 10 pemain Everton 4-0 - ANTARA News	3

116	Hantam Lazio 3-0, Juventus kembali perlebar jarak di puncak klasemen - ANTARA News	3
117	Torres antar Atletico atasi Bilbao 1-0 - ANTARA News	3
118	Libas Berlin 3-0, Dortmund hadapi Muenchen di final Piala Jerman - ANTARA News	3
119	Liverpool sementara unggul 2-0 atas Everton - ANTARA News	3
120	Feyenoord imbang 2-2 kontra Heracles - ANTARA News	3
121	Busana daerah dan adat perkawinan Indonesia pukau wanita Roma - ANTARA News	4
122	Pemkot Bekasi alokasikan Rp120 miliar bangun ruang kelas senilai Rp300 juta/kelas - ANTARA News	2
123	Origi pemimpin Liverpool dalam Derby Merseyside - ANTARA News	3
124	Goodyear TripleMax sasar pengguna mobil SUV - ANTARA News Otomotif	1
125	Menjajal Goodyear TripleMax ban jagoan di jalan basah - ANTARA News Otomotif	1
126	Alasan Goodyear tidak produksi ban motor - ANTARA News Otomotif	1
127	Ahok santaianggapi syarat baru KPU dukungan harus bermaterai - ANTARA News	1
128	Ahok bantah adiknya notaris - ANTARA News	4
129	Manchester United terus buru Cristiano Ronaldo - ANTARA News	1
130	Jenazah kepala sekolah ditemukan mengapung di perairan Teluk	2

	Balikpapan - ANTARA News	
131	Amerika Serikat akan tambah tentara ke Irak perangi ISIS - ANTARA News	4
132	Kumpulan Berita Skandal Panama Papers - ANTARA News	1
133	Kumpulan Berita Perburuan Teroris Santoso - ANTARA News	1
134	Kumpulan Berita Rio Haryanto F1 - ANTARA News	1
135	Kumpulan Berita Pembongkaran Kalijodo - ANTARA News	1
136	Kumpulan Berita Liga Champions 2016 - ANTARA News	1
137	Kumpulan Berita Pencabutan Sanksi Iran - ANTARA News	1
138	Kumpulan Berita Revisi UU KPK - ANTARA News	1
139	Kumpulan Berita Kereta Cepat Jakarta-Bandung - ANTARA News	1
140	Kumpulan Berita Misteri Kematian Mirna - ANTARA News	1
141	Kumpulan Berita Pimpinan Baru KPK - ANTARA News	1
142	Seminar Nasional Kompolnas - Foto ANTARA News	1
143	Sidang Pimpinan DPRD Muba - Foto ANTARA News	1
144	Pemeriksaan M.Sanusi - Foto ANTARA News	1
145	Operasi Narkoba Bali - Foto ANTARA News	1
146	27 pemuda ASEAN ikuti program pertukaran pemuda di Korea - ANTARA News	2
147	Melly Goeslaw tidak paksa anak bermusik - ANTARA News	2
148	Film "Rudy Habibie" ditargetkan raup 10 juta penonton - ANTARA News	2
149	Indonesia dapat dana hibah Rp11 miliar untuk pelestarian dugong	2

	- ANTARA News	
150	Ini komentar Walhi soal reklamasi Teluk Jakarta - ANTARA News	1
151	Separuh Karang Penghalang Besar mati atau sekarat - ANTARA News	1
152	Soal IoT, Panasonic tunggu kesiapan infrastruktur Indonesia - ANTARA News	1
153	Panasonic: 4K TV masih jadi tren 2016 - ANTARA News	2
155	Tentang Kami - ANTARA News	4
156	Jaringan Antaranews.com - ANTARA News	4
157	Pedoman Dewan Pers - ANTARA News	4
158	Walkot Bekasi sambut baik perluasan trayek Transjakarta - ANTARA News	5
159	Kapolri berharap "May Day" diisi bakti sosial - ANTARA News	5
160	Riyadh pecat Menteri Pengairan karena buruk dalam penerapan kebijakan - ANTARA News	2
161	Kebakaran Kali baru - Foto ANTARA News	1
162	Rossi juarai balap MotoGP Spanyol - ANTARA News	1
163	Pengamat: Singapura dicurigai jegal RUU Tax Amnesty - ANTARA News	2
164	Empat semifinalis Liga Champions menangi laga liga - ANTARA News	3
165	Ini curhat Lorenzo soal gagal raih pole MotoGP Jerez - ANTARA News	1

166	Menag sambut ajakan Saudi perangi terorisme - ANTARA News	1
167	Wakil Ketua MPR: dai perkokoh empat pilar - ANTARA News	1
168	Kemandirian galangan kapal nasional dipercepat - ANTARA News	2
169	Lewati Watford 2-1, Palace hadapi MU di final Piala FA - ANTARA News	3
170	Kemenperin dukung industri kertas berharga nasional rajai pasar - ANTARA News	2
171	Pembangunan pabrik stainless steel di Morowali capai 80 persen - ANTARA News	2
172	Mendag: negosiasi IEU-CEPA segera dimulai - ANTARA News	2
173	Paket Ekonomi tarik kerja sama perdagangan-investasi - ANTARA News	2
174	Mendag harap Indonesia masuk 40 besar kemudahan berbisnis - ANTARA News	2
175	Kalahkan Nishikore, Rafael Nadal juarai Barcelona Terbuka - ANTARA News	1
176	Ringkasan laga dan klasemen Liga Inggris, Leicester di ambang juara - ANTARA News	3
177	Alcacer selamatkan Valencia dari kekalahan - ANTARA News	3
178	Feyenoord juarai Piala Belanda - ANTARA News	3
179	Bekuk Lorient 2-0, Saint-Etienne jaga peluang ke Liga Champions - ANTARA News	3
180	Bantai Swansea 4-0, Leicester makin dekat dengan gelar juara - ANTARA News	3

181	Lewati Watford 2-1, Palace hadapi MU di final Piala FA - ANTARA News	3
182	Bentrok antarkelompok masyarakat terjadi di Karawang - ANTARA News	4
183	Kerber akhiri kisah manis Siegemund di Stuttgart - ANTARA News	1
184	Sevilla atasi Betis 2-0 - ANTARA News	3
185	Sampdoria kalahkan Lazio 2-1, jauhi zona degradasi - ANTARA News	3
186	Walkot Bekasi sambut baik perluasan trayek Transjakarta - ANTARA News	5
187	Atasi Hoffenheim 3-1, Gladbach bayangi zona Liga Champions - ANTARA News	3
188	Tahanimbang Arsenal 0-0, Sunderland tinggalkan zona degradasi - ANTARA News	3
189	Menkumham jelaskan kronologi kasus kematian Undang Kosim - ANTARA News	1
190	Kipchoge pertahankan gelar juara dunia lari marathon - ANTARA News	2
191	Undian putaran final IGC ditetapkan - ANTARA News	4
192	Polisi amankan 14 ton bawang merah ilegal - ANTARA News	1
193	Kapolri berharap "May Day" diisi bakti sosial - ANTARA News	5
194	Ditahanimbang Nantes 1-1, Marseille Lewati 11 laga beruntun tanpa kemenangan - ANTARA News	3

195	Indonesia kalahkan Singapura 79-61 di kejuaraan SEABA U-18 - ANTARA News	2
196	Filipina menang mudah atas Laos 133-52 - ANTARA News	2
197	TNI AL gagalkan penyelundupan 60 ton bawang merah di Aceh - ANTARA News	5
198	Protes "Up All Night" berubah rusuh di Paris - ANTARA News	1
199	Spektronic-XI ITS juarai "Chem-E-Car" Indonesia 2016 - ANTARA News	2
200	Riyadh pecat Menteri Pengairan karena buruk dalam penerapan kebijakan - ANTARA News	2
201	Pasukan Afghanistan tewaskan lima gerilyawan di provinsi utara - ANTARA News	2
202	Klasemen MotoGP, Rossi persempit jarak dengan Lorenzo - ANTARA News	1
203	Yasonna akui ada pemaksaan pemeriksaan Undang Kosim - ANTARA News	4
204	Warga korban banjir Karawang mulai bersihkan rumah - ANTARA News	5
205	Sriwijaya FC uji coba lawan Gresik 2-2 - ANTARA News	4
206	BPBD Karawang: banjir mulai surut - ANTARA News	5
207	Brad Binder menangi balap Moto3 Jerez - ANTARA News	1
208	Yasonna: kerugian kebakaran Lapas Banceuy Rp6 miliar - ANTARA News	1
209	Hasil GP Moto2 Spanyol, Lowes tercepat - ANTARA News	1

210	Bulog berjanji perbaiki kualitas raskin - ANTARA News	5
211	Dubes RI diundang dukung langsung Daud Yordan Juni mendatang - ANTARA News	1
212	Warga Karimun keluhkan jalan protokol tergenang air - ANTARA News	5
213	BMKG imbau warga Babel waspadaai hujan lebat disertai petir Senin - ANTARA News	5
214	300 pemain muda ikuti Football Clinic Day - ANTARA News	1
215	Toyota akan meluncurkan mobil plug-in hybrid di China tahun 2018 - ANTARA News Otomotif	1
216	Trail Blazers hidupkan persaingan dengan rebut game ketiga - ANTARA News	2
217	Dayung pastikan dua tiket ke Olimpiade Brasil - ANTARA News	1
218	Munas Pertina diwarnai walk out, kepesertaan Bali dan Sulut dipertanyakan - ANTARA News	1
219	Arema lawan Metro FC dalam uji coba terakhir - ANTARA News	4
220	IHSG naik 1,89 persen dalam 18-22 April - ANTARA News	1
221	Empat parpol siapkan koalisi hadapi Pilkada Jatim - ANTARA News	2
222	Ketua MPR sebut TMII perkuat identitas bangsa - ANTARA News	1
223	Hasil kualifikasi MotoGP Spanyol, Rossi terdepan - ANTARA News	1
224	Pakar: reklamasi tambah disparitas pembangunan antarpulau - ANTARA News	2

225	Hasil kajian reklamasi akan dipublikasikan dua bulan mendatang - ANTARA News	5
226	Waka KPAI: Agus pelaku mutilasi layak dihukum mati - ANTARA News	4
227	Singapura klarifikasi soal perjanjian ekstradisi dengan RI - ANTARA News	2
228	Lasykar Wanita Anti Narkoba - Foto ANTARA News	1
229	Rekonstruksi Pembunuhan Mutilasi Batal di Gelar - Foto ANTARA News	1
230	Indonesia Chem-E-Car Competition - Foto ANTARA News	1
231	Indonesia pameran kuliner di Nanyang Culture Festival 2016 - ANTARA News	2
232	Jenazah Prince dikremasi - ANTARA News	1
233	Pameran "Reborn" di Borobudur diikuti 24 perupa - ANTARA News	2
234	Sineas ditantang kembalikan kejayaan Makassar - ANTARA News	2
235	Jimbaran Festival 2016 serangkaian peringatan Hari Bumi - ANTARA News	1
236	Kebakaran lahan di Bengkalis belum bisa dipadamkan - ANTARA News	2
237	Kebun Raya LIPI tanam 300 pohon langka - ANTARA News	1
238	Ini strategi perempuan atasi kekerasan seksual online - ANTARA News	2
239	Anggota dewan jajak kapal riset LIPI - ANTARA News	1

240	Orangtua alasan utama perempuan berkarir di iptek - ANTARA News	1
241	Komunitas scooter Kutu Bandung rayakan ulang tahun - ANTARA News Otomotif	1
242	Pajero dan Delica Ramaikan Gala Premiere Film AADC - ANTARA News Otomotif	1
243	- ANTARA TV	1

Data tersebut adalah hasil dari proses crawl dan klasifikasi. Untuk mengetahui akurasi dari aplikasi maka dilakukan perubahan data pada field klasifikasi. Kemudian dilakukan klasifikasi ulang pada data yang telah diubah. Hasilnya dapat dilihat pada **tabel 4.17**.

Tabel 4.17 Hasil Klasifikasi Setelah Perubahan Data

No	Daftar halaman	Klasifikasi
1	ANTARA News : Portal Berita Indonesia	5
2	Top News - ANTARA News	1
3	Berita Terkini - ANTARA News	1
4	Rilis Pers - ANTARA News	2
5	RSS ANTARA News	2
6	Ketentuan Penggunaan - Antara News	1
7	Tentang Kami - ANTARA News	5
8	Nasional - ANTARA News	3
9	Nasional Umum - ANTARA News	3
10	Nasional Polkam - ANTARA News	1
11	Nasional Hukum - ANTARA News	4
12	Nasional Pendidikan - ANTARA News	4
13	Nasional Kesehatan - ANTARA News	4
14	Nasional Naker - ANTARA News	1
15	Nasional Unik - ANTARA News	1
16	Internasional - ANTARA News	2
17	Internasional Asiaoseania - ANTARA News	2
18	Internasional Amerika - ANTARA News	1
19	Internasional Eropa - ANTARA News	1
20	Internasional Timur Tengah - ANTARA News	4
21	Internasional Afrika - ANTARA News	1
22	Ekonomi - ANTARA News	2

23	Ekonomi Makro - ANTARA News	2
24	Ekonomi Moneter - ANTARA News	2
25	Ekonomi Bisnis - ANTARA News	2
26	Ekonomi Bursa - ANTARA News	2
27	Olahraga - ANTARA News	2
28	Olahraga Sepakbola - ANTARA News	2
29	Olahraga Tennis - ANTARA News	1
30	Olahraga Bulutangkis - ANTARA News	2
31	Olahraga Basket - ANTARA News	3
32	Olahraga Tinju - ANTARA News	3
33	Olahraga Balap - ANTARA News	3
34	Olahraga Umum - ANTARA News	3
35	Hiburan - ANTARA News	2
36	Hiburan Sinema - ANTARA News	2
37	Hiburan Musik - ANTARA News	2
38	Hiburan Selebriti - ANTARA News	2
39	Hiburan Seni Budaya - ANTARA News	2
40	Hiburan Gaya Hidup - ANTARA News	2
41	Teknologi - ANTARA News	4
42	Teknologi Sains - ANTARA News	4
43	Teknologi Komputer - ANTARA News	2
44	Teknologi Gadget - ANTARA News	2
45	Teknologi Internet - ANTARA News	1

46	Warta Bumi - ANTARA News	1
47	Warta Bumi Pemanasan Global - ANTARA News	1
48	Warta Bumi Flora Fauna - ANTARA News	2
49	Artikel - ANTARA News	1
50	Artikel Komentar - ANTARA News	2
51	Artikel Opini - ANTARA News	1
52	Artikel Buku - ANTARA News	1
53	Artikel Jeda - ANTARA News	1
54	Artikel Visi - ANTARA News	1
55	Galeri Foto ANTARA News	1
56	ANTARA TV - ANTARA News	2
57	ANTARA News : Portal Berita Indonesia	5
58	Polda Metro ringkus pemutilasi wanita hamil - ANTARA News	1
59	Harga minyak dunia melonjak dipicu penurunan produksi AS - ANTARA News	1
60	Madrid bekuk Villarreal tiga gol tanpa balas - ANTARA News	3
61	MU atasi Palace 2-0, dekati zona Liga Champions - ANTARA News	3
62	Wapres Buka Innacraft 2016 - Foto ANTARA News	1
63	Suarez caturgol dan tiga assist, Barcelona libas Deportivo 8-0 - ANTARA News	3
65	Mari momok Darurat Narkoba itu kita jadikan momentum bersih- bersih - ANTARA News	2
66	Presiden Jokowi tiba di Brussels - ANTARA News	2

67	Info Keagamaan - ANTARA News	2
68	Kemenag upayakan kuota haji ditambah 10 ribu - ANTARA News	5
69	Tempat ibadah Tri Dharma Buddha harus tertib administrasi - ANTARA News	2
70	Kumpulan Fokus Berita - ANTARA News	2
71	Fokus Reklamasi Teluk Jakarta - ANTARA News	5
72	Fokus Resuffle Kabinet - ANTARA News	1
73	Industry Update - ANTARA News	2
74	Indonesia genjot industri tekstil pada perundingan kelima D8 - ANTARA News	2
75	Eksistensi industri kerajinan jaga regenerasi perajin - ANTARA News	2
76	Empat kebijakan Kemenperin dongkrak industri MRO - ANTARA News	2
77	Info Perdagangan - Ekonomi - Antara News - ANTARA News	1
78	INACRAF 2016 momentum peningkatan ekspor kerajinan - ANTARA News	2
79	IIPC Budapest promosikan potensi Indonesia - ANTARA News	2
80	IIPC : Pameran Lagos buka peluang transaksi potensial - ANTARA News	2
81	Pasar saham Australia dibuka lebih tinggi - ANTARA News	2
82	Bursa saham Tiongkok dibuka lebih rendah - ANTARA News	2
83	19 mahasiswa Indonesia dievakuasi ke Fukuoka - ANTARA News	1
84	Semalaman diguyur hujan, genangan landa sejumlah titik Jakarta -	1

	ANTARA News	
85	Saham Tokyo dibuka melonjak didukung kenaikan minyak - ANTARA News	2
86	Pasar rakyat kuliner semarakkan HUT Depok - ANTARA News	4
87	5 kiat menjaga ban tetap awet - ANTARA News Otomotif	1
89	Ruas jalan Beteleme-Kolonodale Morowali Utara putus - ANTARA News	4
90	Obama bertemu Raja Saudi sebelum pertemuan puncak AS-GCC - ANTARA News	4
91	30 cedera dalam ledakan kompleks petrokimia di Meksiko - ANTARA News	1
92	Polda Metro ringkus pemutilasi wanita hamil - ANTARA News	5
93	Pasar saham Spanyol ditutup naik 1,96 persen - ANTARA News	5
94	Indeks FTSE 100 Inggris berakhir naik 0,08 persen - ANTARA News	2
95	Samsung rilis varian Pink Gold Galaxy S7/ S7 edge - ANTARA News	2
96	WNI korban gempa Kumamoto tiba di Tokyo - ANTARA News	4
97	Kenaikan harga minyak angkat saham-saham AS - ANTARA News	2
98	Kartini masa kini dimata Krisdayanti - ANTARA News	2
99	Skyland Petroleum pasok minyak dan gas ke Tiongkok - ANTARA News	2
100	Pekan ke-34 La Liga Spanyol, persaingan juara kian ketat - ANTARA News	3

101	Dolar AS menguat didukung data ekonomi positif - ANTARA News	2
102	Emas relatif stabil di tengah penguatan dolar AS - ANTARA News	2
103	Kalah 1-2 dari Udinese, Fiorentina gagal masuk empat besar - ANTARA News	3
104	Wall Street reli di tengah laporan laba dan kenaikan minyak - ANTARA News	2
105	Harga minyak dunia melonjak dipicu penurunan produksi AS - ANTARA News	2
106	Atasi Heerenveen 2-0, Ajax kembali puncak Eredivisie - ANTARA News	3
107	Totti bawa Roma berbalik kalahkan Torino 3-2 - ANTARA News	3
108	Inter tersandung di kandang Genoa, kalah 0-1 - ANTARA News	3
109	Ringkasan hasil pekan ke-34 Liga Italia Seri A - ANTARA News	3
110	Madrid bekuk Villarreal tiga gol tanpa balas - ANTARA News	3
111	Atasi Sochaux 1-0, Marseille ke final Piala Prancis - ANTARA News	3
112	Valencia kuliti Eibar 4-0, trigol Alcacer sempurna - ANTARA News	3
113	West Ham menang meyakinkan 3-1 lawan Watford - ANTARA News	3
114	MU atasi Palace 2-0, dekati zona Liga Champions - ANTARA News	3
115	Liverpool cukur 10 pemain Everton 4-0 - ANTARA News	3

116	Hantam Lazio 3-0, Juventus kembali perlebar jarak di puncak klasemen - ANTARA News	3
117	Torres antar Atletico atasi Bilbao 1-0 - ANTARA News	3
118	Libas Berlin 3-0, Dortmund hadapi Muenchen di final Piala Jerman - ANTARA News	3
119	Liverpool sementara unggul 2-0 atas Everton - ANTARA News	3
120	Feyenoordimbang 2-2 kontra Heracles - ANTARA News	3
121	Busana daerah dan adat perkawinan Indonesia pukai wanita Roma - ANTARA News	4
122	Pemkot Bekasi alokasikan Rp120 miliar bangun ruang kelas senilai Rp300 juta/kelas - ANTARA News	2
123	Origi pimpin Liverpool dalam Derby Merseyside - ANTARA News	3
124	Goodyear TripleMax sasar pengguna mobil SUV - ANTARA News Otomotif	1
125	Menjajal Goodyear TripleMax ban jagoan di jalan basah - ANTARA News Otomotif	1
126	Alasan Goodyear tidak produksi ban motor - ANTARA News Otomotif	1
127	Ahok santai tanggap syarat baru KPU dukungan harus bermaterai - ANTARA News	1
128	Ahok bantah adiknya notaris - ANTARA News	4
129	Manchester United terus buru Cristiano Ronaldo - ANTARA News	2
130	Jenazah kepala sekolah ditemukan mengapung di perairan Teluk	2

	Balikpapan - ANTARA News	
131	Amerika Serikat akan tambah tentara ke Irak perangi ISIS - ANTARA News	4
132	Kumpulan Berita Skandal Panama Papers - ANTARA News	5
133	Kumpulan Berita Perburuan Teroris Santoso - ANTARA News	2
134	Kumpulan Berita Rio Haryanto F1 - ANTARA News	1
135	Kumpulan Berita Pembongkaran Kalijodo - ANTARA News	5
136	Kumpulan Berita Liga Champions 2016 - ANTARA News	2
137	Kumpulan Berita Pencabutan Sanksi Iran - ANTARA News	5
138	Kumpulan Berita Revisi UU KPK - ANTARA News	1
139	Kumpulan Berita Kereta Cepat Jakarta-Bandung - ANTARA News	1
140	Kumpulan Berita Misteri Kematian Mirna - ANTARA News	1
141	Kumpulan Berita Pimpinan Baru KPK - ANTARA News	1
142	Seminar Nasional Kompolnas - Foto ANTARA News	1
143	Sidang Pimpinan DPRD Muba - Foto ANTARA News	1
144	Pemeriksaan M.Sanusi - Foto ANTARA News	1
145	Operasi Narkoba Bali - Foto ANTARA News	1
146	27 pemuda ASEAN ikuti program pertukaran pemuda di Korea - ANTARA News	2
147	Melly Goeslaw tidak paksa anak bermusik - ANTARA News	2
148	Film "Rudy Habibie" ditargetkan raup 10 juta penonton - ANTARA News	2
149	Indonesia dapat dana hibah Rp11 miliar untuk pelestarian dugong	2

	- ANTARA News	
150	Ini komentar Walhi soal reklamasi Teluk Jakarta - ANTARA News	1
151	Sepuluh Karang Penghalang Besar mati atau sekarat - ANTARA News	1
152	Soal IoT, Panasonic tunggu kesiapan infrastruktur Indonesia - ANTARA News	1
153	Panasonic: 4K TV masih jadi tren 2016 - ANTARA News	2
155	Tentang Kami - ANTARA News	4
156	Jaringan Antaranews.com - ANTARA News	4
157	Pedoman Dewan Pers - ANTARA News	4
158	Walkot Bekasi sambut baik perluasan trayek Transjakarta - ANTARA News	5
159	Kapolri berharap "May Day" diisi bakti sosial - ANTARA News	5
160	Riyadh pecat Menteri Pengairan karena buruk dalam penerapan kebijakan - ANTARA News	2
161	Kebakaran Kali baru - Foto ANTARA News	1
162	Rossi juarai balap MotoGP Spanyol - ANTARA News	1
163	Pengamat: Singapura dicurigai jejal RUU Tax Amnesty - ANTARA News	2
164	Empat semifinalis Liga Champions menangi laga liga - ANTARA News	3
165	Ini curhat Lorenzo soal gagal raih pole MotoGP Jerez - ANTARA News	1

166	Menag sambut ajakan Saudi perangi terorisme - ANTARA News	1
167	Wakil Ketua MPR: dai perkokoh empat pilar - ANTARA News	1
168	Kemandirian galangan kapal nasional dipercepat - ANTARA News	2
169	Lewati Watford 2-1, Palace hadapi MU di final Piala FA - ANTARA News	3
170	Kemenperin dukung industri kertas berharga nasional rajai pasar - ANTARA News	2
171	Pembangunan pabrik stainless steel di Morowali capai 80 persen - ANTARA News	2
172	Mendag: negosiasi IEU-CEPA segera dimulai - ANTARA News	2
173	Paket Ekonomi tarik kerja sama perdagangan-investasi - ANTARA News	2
174	Mendag harap Indonesia masuk 40 besar kemudahan berbisnis - ANTARA News	2
175	Kalahkan Nishikore, Rafael Nadal juarai Barcelona Terbuka - ANTARA News	1
176	Ringkasan laga dan klasemen Liga Inggris, Leicester di ambang juara - ANTARA News	3
177	Alcacer selamatkan Valencia dari kekalahan - ANTARA News	3
178	Feyenoord juarai Piala Belanda - ANTARA News	3
179	Bekuk Lorient 2-0, Saint-Etienne jaga peluang ke Liga Champions - ANTARA News	3
180	Bantai Swansea 4-0, Leicester makin dekat dengan gelar juara - ANTARA News	3

181	Lewati Watford 2-1, Palace hadapi MU di final Piala FA - ANTARA News	3
182	Bentrok antarkelompok masyarakat terjadi di Karawang - ANTARA News	4
183	Kerber akhiri kisah manis Siegemund di Stuttgart - ANTARA News	1
184	Sevilla atasi Betis 2-0 - ANTARA News	3
185	Sampdoria kalahkan Lazio 2-1, jauhi zona degradasi - ANTARA News	3
186	Walkot Bekasi sambut baik perluasan trayek Transjakarta - ANTARA News	5
187	Atasi Hoffenheim 3-1, Gladbach bayangi zona Liga Champions - ANTARA News	3
188	Tahanimbang Arsenal 0-0, Sunderland tinggalkan zona degradasi - ANTARA News	3
189	Menkumham jelaskan kronologi kasus kematian Undang Kosim - ANTARA News	1
190	Kipchoge pertahankan gelar juara dunia lari marathon - ANTARA News	2
191	Undian putaran final IGC ditetapkan - ANTARA News	4
192	Polisi amankan 14 ton bawang merah ilegal - ANTARA News	1
193	Kapolri berharap "May Day" diisi bakti sosial - ANTARA News	5
194	Ditahanimbang Nantes 1-1, Marseille lewati 11 laga beruntun tanpa kemenangan - ANTARA News	3

195	Indonesia kalahkan Singapura 79-61 di kejuaraan SEABA U-18 - ANTARA News	2
196	Filipina menang mudah atas Laos 133-52 - ANTARA News	2
197	TNI AL gagalkan penyelundupan 60 ton bawang merah di Aceh - ANTARA News	5
198	Protes "Up All Night" berubah rusuh di Paris - ANTARA News	1
199	Spektronic-XI ITS juarai "Chem-E-Car" Indonesia 2016 - ANTARA News	2
200	Riyadh pecat Menteri Pengairan karena buruk dalam penerapan kebijakan - ANTARA News	2
201	Pasukan Afghanistan tewaskan lima gerilyawan di provinsi utara - ANTARA News	2
202	Klasemen MotoGP, Rossi persempit jarak dengan Lorenzo - ANTARA News	1
203	Yasonna akui ada pemaksaan pemeriksaan Undang Kosim - ANTARA News	4
204	Warga korban banjir Karawang mulai bersihkan rumah - ANTARA News	5
205	Sriwijaya FC uji coba lawan Gresik 2-2 - ANTARA News	4
206	BPBD Karawang: banjir mulai surut - ANTARA News	5
207	Brad Binder menangi balap Moto3 Jerez - ANTARA News	1
208	Yasonna: kerugian kebakaran Lapas Banceuy Rp6 miliar - ANTARA News	4
209	Hasil GP Moto2 Spanyol, Lowes tercepat - ANTARA News	1

210	Bulog berjanji perbaiki kualitas raskin - ANTARA News	3
211	Dubes RI diundang dukung langsung Daud Yordan Juni mendatang - ANTARA News	5
212	Warga Karimun keluhkan jalan protokol tergenang air - ANTARA News	5
213	BMKGimbau warga Babel waspadai hujan lebat disertai petir Senin - ANTARA News	2
214	300 pemain muda ikuti Football Clinic Day - ANTARA News	4
215	Toyota akan meluncurkan mobil plug-in hybrid di China tahun 2018 - ANTARA News Otomotif	1
216	Trail Blazers hidupkan persaingan dengan rebut game ketiga - ANTARA News	2
217	Dayung pastikan dua tiket ke Olimpiade Brasil - ANTARA News	1
218	Munas Pertina diwarnai walk out, kepesertaan Bali dan Sulut dipertanyakan - ANTARA News	1
219	Arema lawan Metro FC dalam uji coba terakhir - ANTARA News	4
220	IHSG naik 1,89 persen dalam 18-22 April - ANTARA News	5
221	Empat parpol siapkan koalisi hadapi Pilkada Jatim - ANTARA News	2
222	Ketua MPR sebut TMII perkuat identitas bangsa - ANTARA News	1
223	Hasil kualifikasi MotoGP Spanyol, Rossi terdepan - ANTARA News	1
224	Pakar: reklamasi tambah disparitas pembangunan antarpulau - ANTARA News	2

225	Hasil kajian reklamasi akan dipublikasikan dua bulan mendatang - ANTARA News	5
226	Waka KPAI: Agus pelaku mutilasi layak dihukum mati - ANTARA News	4
227	Singapura klarifikasi soal perjanjian ekstradisi dengan RI - ANTARA News	2
228	Lasykar Wanita Anti Narkoba - Foto ANTARA News	1
229	Rekonstruksi Pembunuhan Mutilasi Batal di Gelar - Foto ANTARA News	1
230	Indonesia Chem-E-Car Competition - Foto ANTARA News	1
231	Indonesia pameran kuliner di Nanyang Culture Festival 2016 - ANTARA News	2
232	Jenazah Prince dikremasi - ANTARA News	1
233	Pameran "Reborn" di Borobudur diikuti 24 perupa - ANTARA News	2
234	Sineas ditantang kembalikan kejayaan Makassar - ANTARA News	2
235	Jimbaran Festival 2016 serangkaian peringatan Hari Bumi - ANTARA News	1
236	Kebakaran lahan di Bengkalis belum bisa dipadamkan - ANTARA News	2
237	Kebun Raya LIPI tanam 300 pohon langka - ANTARA News	1
238	Ini strategi perempuan atasi kekerasan seksual online - ANTARA News	2
239	Anggota dewan jajak kapal riset LIPI - ANTARA News	1

240	Orangtua alasan utama perempuan berkarir di iptek - ANTARA News	1
241	Komunitas scooter Kutu Bandung rayakan ulang tahun - ANTARA News Otomotif	1
242	Pajero dan Delica Ramaikan Gala Premiere Film AADC - ANTARA News Otomotif	1
243	- ANTARA TV	1

Dari hasil klasifikasi ulang tersebut, dapat dilihat berapa tingkat keakurasiannya.

Tingkat keakurasiannya dihitung dengan membagi jumlah data yang berubah dengan total data, kemudian dikali dengan 100%. Maka hasil dari keakurasiannya aplikasi dapat diketahui.

4.2.2 Hasil Evaluasi dan Pembahasan

Dari hasil pengklasifikasi pada **Tabel 4.1** didapatkan bahwa beberapa dokumen yang terunduh berjumlah 243 dokumen. Pada table hasil *crawl* dan *classification* dapat terlihat ada 5 group yang terbentuk berdasarkan penentuan di awal saat akan melakukan prosesnya. Dari hasil uji coba tampak bahwa algoritma *fuzzy c-means* sangat cocok digunakan untuk pengklasifikasi sekumpulan dokumen. Hal ini dikarenakan adanya pembobotan yang kemudian dilanjutkan dengan derajat keanggotaan pada algoritma ini.

4.3 Hubungan Penelitian yang Dilakukan dengan Kaidah Islam

Allah SWT dalam Al-Quran berfirman bahwa umat islam dianjurkan untuk mencari ilmu. Sebagaimana yang tertulis dalam firman Allah SWT berikut :

﴿ خَيْرٌ تَعْمَلُونَ بِمَا وَاللَّهُ دَرَجَاتٍ الْعِلْمَ أُتُوا وَالَّذِينَ مِنْكُمْ ءَامَنُوا الَّذِينَ اللَّهُ

“Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu pengetahuan beberapa derajat. Dan Allah Maha Mengetahui apa yang kamu kerjakan.” (QS. Al-Mujadlah: 11).

Dari ayat tersebut dijelaskan bahwa tidak hanya orang yang beriman yang ditinggikan derajatnya namun Allah SWT akan meninggikan derajat orang-orang yang memiliki ilmu dan senantiasa dalam kehidupannya mencari ilmu pengetahuan. Mencari ilmu di dunia baik ilmu yang berhubungan dengan dunia maupun ilmu yang berhubungan dengan akhirat merupakan suatu ibadah yang diwajibkan oleh Allah SWT. Dan oleh Allah SWT dicatat sebagai amal ibadah pahala untuk kelak diakhirat. Betapa meruginya orang yang tidak mencari ilmu di dunia.

Dalam persepektif islam, ilmu merupakan hal yang menjadikan manusia berbeda dan lebih unggul atau lebih tinggi derajatnya dibanding makhluk lain sebagaimana yang dijelaskan dalam ayat berikut:

ن كُنْتُمْ إِنْ هَتُّوْا بِأَسْمَاءِ أَنْبِيَائِي فَقَالَ الْمَلَكُ عَلَى عَرَضِهِمْ ثُمَّ كُلَّهَا أَلْأَسْمَاءِ آدَمَ وَعَلَّمَ

الْحَكِيمُ الْعَلِيمُ أَنْتَ إِنْ كُنْتَ عَلِمْتَ مَا إِلَّا لَنَا عِلْمٌ لَا سُبْحَانَكَ قَالُوا ۖ صَدَقَ

“Dan Dia mengajarkan kepada Adam nama-nama (benda-benda) seluruhnya, kemudian mengemukakannya kepada para Malaikat lalu berfirman: "Sebutkanlah kepada-Ku nama benda-benda itu jika kamu mamang benar orang-orang yang benar!". Mereka menjawab: "Maha Suci Engkau, tidak ada yang kami ketahui selain dari apa yang telah Engkau ajarkan kepada kami; sesungguhnya Engkaulah Yang Maha Mengetahui lagi Maha Bijaksana.” (Q.S Al-Baqarah: 31-32)

Ayat tersebut menjelaskan bahwa ilmu memberikan keunggulan dan derajat yang lebih diantara makhluk Allah SWT yang lainnya, maka dalam Kamus Besar Bahasa Indonesia yang juga dibahas bahwa definisi ilmu dapat diartikan sebagai

pengetahuan tentang suatu bidang yang disusun secara bersistem menurut metode tertentu, yang dapat digunakan untuk menerangkan gejala tertentu di bidang (pengetahuan) itu; atau pengetahuan atau kepandaian tentang soal duniawi, akhirat, lahir, batin, dsb. Untuk mendapatkan ilmu kita sebagai manusia bisa memanfaatkan informasi yang beredar luas di dunia ini. Segala sesuatu yang menjadi fakta dalam dunia ini bisa dijadikan suatu informasi yang bisa dijadikan sumber ilmu pengetahuan.

Sehubungan dengan dalam rangka mencari ilmu yang merupakan perintah ibadah dari Allah SWT melalui informasi, penulis berinisiatif membuat suatu aplikasi pengklasifikasian halaman *web* dengan keluaran aplikasi berupa pokok isi (*content*) dari halaman *web* di tambah dengan pengklasifikasiannya. Dengan aplikasi ini mempermudah kita sebagai umat manusia untuk mendapatkan suatu informasi yang merupakan sumber untuk mendapatkan ilmu pengetahuan di dunia. Karena dalam aplikasi ini akan diberikan suatu klasifikasi dari suatu informasi yang didapat dari halaman *web* sehingga dapat diketahui kandungan jenis dari informasi atau cabang keilmuan apakah yang ada didalam informasi tersebut.

BAB V

PENUTUP

5.1 Kesimpulan

Dari hasil implementasi dan uji coba yang dilakukan dapat disimpulkan bahwa metode *Fuzzy C-means Clustering* merupakan algoritma yang mampu dijadikan sebagai pengelompokan *content* halaman *web*. Hal ini terbukti dengan seluruh *content* halaman *web* yang sudah diunduh menjadi dokumen *web* berhasil dikelompokkan dengan tingkat akurasi klasifikasi yang dilakukan oleh aplikasi sebesar 56%.

5.2 Saran

Berdasarkan kesimpulan di atas, masih banyak kekurangan dalam penelitian aplikasi pengklasifikasian ini. Oleh karena itu penulis menyarankan beberapa hal untuk bahan pengembangan selanjutnya, diantaranya:

- 1) Menambahkan data training untuk proses pengelompokan untuk mempengaruhi hasil akurasi pengelompokan
- 2) Mengembangkan algoritma khusus untuk melakukan ekstraksi dari dokumen *web* ke isi *web*.
- 3) Menambahkan koleksi *stop word* untuk memfilter isi dari dokumen *web*.
- 4) Menambah proses *stemming* untuk menambah tingkat akurasi pengklasifikasian.

DAFTAR PUSTAKA

- A. K. Jain and R. C. Dubes, *Algorithm for clustering Data*. Englewood, NJ; Prentice-Hall, 1998.
- Bahar. 2011. Penentuan Jurusan Sekolah Menengah Atas dengan Algoritma Fuzzy C-Means. Teknik Informatika Universitas Dian Nuswantoro. Semarang.
- Beze, Husmul. 2008. Karakterisasi Pelanggan PLN Menggunakan Algoritma Fuzzy C-Means. Pascasarjana Institut Pertanian Bogor, Bogor
- Bezdek J. C., (1981) *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum Press, New York.
- Cios, Krzysztof J. Etc. 2007. *Data Mining A Knowledge Discovery Approach*. Springer
- J. C. Bezdek, R. J. Hathaway, M. J. Sabin, and W. T. Tucker, (1987) *Convergence theory for fuzzy-c-means: Counterexamples and repair*, IEEE.
- Kusrini, Emha Taufiq Luthfi. 2009. Algoritma Fuzzy C-Means. Jurusan Sistem
- Hyusein, Byurhan, Patel, Ahmad. 2003. *Web Document Indexing and Retrieval*. LNCS 2588 pp. 573-579, Springer Verlag Berlin
- Cios, Krzysztof J. Etc. 2007. *Data Mining A Knowledge Discovery Approach*. Springer
- Kowalski, Gerald. 1997. *Information Storage and Retrieval Systems Theory and Implementation*. United Kingdom.
- Manning, Christopher D, Ragnavan Prabhakar, Schutze, Hinrich. 2009. *Introduction to Information Retrieval*. Cambridge University Press
- Polettini, Nicola (2004) *The Vector Space Model in Information Retrieval – Term Weighting Problem*
- Ruthven I., Lalmas M. (2003) *A survey on the use of relevance feedback for information access systems*, *Knowledge Engineering Review*

Ramos, Juan. 2008. *Using TF-IDF to Determine World Relevance in Document Queries*. Department of Computer Science, New Jersey
Salton, Gerrad, Christopher Buckley. 1988. *Information Processing and Management* . Department of Computer Science, Cornell University. USA

Saputra, Iwan Pahendra Anto. 2011. Penggunaan Algoritma TFIDF Dalam Proses Hierarchical Template Maching. *School of Electric Engineering and Informatics-ITB*. Bandung

Setiyono, Budi dan R. Rizal Isnanto. 2008. Pembagian Kelas Kuliah Mahasiswa Menggunakan Algoritma Pengklasteran Fuzzy C-Means. Teknik Elektro Universitas Diponegoro. Semarang.

Shihab, A. (2000). *Fuzzy Clustering Algorithm and Their Application to Medical Image Analysis*. Dissertation, University of London, London.

W Zhao, R Chellappa, A Rosenfeld, P J Phillips. 2000. *Face Recognition : A Literature Survey*. National Institute of Standards and CVL Technical Report, University of Marylan, Columbia.

Wiswakarma, Komang. 2009. Membuat Katalog Online dengan PHP dan CSS