

**IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN
COSINE SIMILARITY UNTUK SISTEM REKOMENDASI
BUKU BERBASIS KONTEN**

Skripsi



oleh
YOHANES RUDY HARYANTO
71110018

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

**IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN
COSINE SIMILARITY UNTUK SISTEM REKOMENDASI
BUKU BERBASIS KONTEN**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

YOHANES RUDY HARYANTO
71110018

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN COSINE SIMILARITY UNTUK SISTEM REKOMENDASI BUKU BERBASIS KONTEN

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi ke sarjana di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar ke sarjana saya.

Yogyakarta, 18 Juni 2015



YOHANES RUDY HARYANTO
71110018

HALAMAN PERSETUJUAN

Judul Skripsi : IMPLEMENTASI ALGORITMA FUZZY C-MEANS
DAN COSINE SIMILARITY UNTUK SISTEM
REKOMENDASI BUKU BERBASIS KONTEN

Nama Mahasiswa : YOHANES RUDY HARYANTO

N I M : 71110018

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Genap

Tahun Akademik : 2014/2015


Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 18 Juni 2015

Dosen Pembimbing I



Antonius Rachmat C., S.Kom.,M.Cs.

Dosen Pembimbing II



Junius Karel, M.T.

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN COSINE SIMILARITY UNTUK SISTEM REKOMENDASI BUKU BERBASIS KONTEN

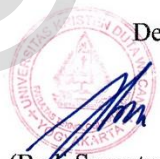
Oleh: YOHANES RUDY HARYANTO / 71110018

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 8 Juni 2015

Yogyakarta, 18 Juni 2015
Mengesahkan,

Dewan Penguji:

1. Antonius Rachmat C., S.Kom., M.Cs.
2. Junius Karel, M.T.
3. Ignatia Dhian E K R, S.Kom, M.Eng
4. Yuan Lukito, S.Kom., M.Cs.



Dekan


(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi


(Gloria Virginia, Ph.D.)

UCAPAN TERIMAKASIH

Puji syukur penulis panjatkan ke hadirat Tuhan Yesus Kristus yang telah melimpahkan anugerah dan rahmatnya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Algoritma Fuzzy C-Means dan Cosine Similarity untuk Sistem Rekomendasi Buku Berbasis Konten” dengan tepat waktu.

Dalam menyelesaikan penelitian ini, penulis menyadari banyak menerima masukan dan saran dari berbagai pihak, baik secara langsung maupun secara tidak langsung. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terimakasih kepada:

1. Antonius Rachmat C, S.Kom., M.Cs selaku dosen pembimbing serta koordinator tugas akhir dan Junius Karel, M.T. selaku dosen pembimbing yang telah membimbing sekaligus memberikan banyak nasehat kepada penulis dalam menyelesaikan penelitian ini.
2. Keluarga terkasih yang selalu mendoakan, mendukung, dan memberikan motivasi kepada penulis selama ini.
3. Teman-teman yang selalu mendukung penulis selama mengadakan penelitian ini.
4. Semua pihak yang telah membantu penulis dalam penyelesaian penelitian ini yang tidak dapat penulis sebutkan satu per satu.

Akhir kata penulis ingin meminta maaf apabila ada kesalahan dalam penyusunan laporan ini. Terimakasih.

Yogyakarta, 20 Mei 2015

Penulis

INTISARI

Mempelajari sesuatu hanya dari sebuah buku jangkit pinjam dari perpustakaan terkadang tidak cukup. Pembahasan yang belum terlalu mendalam ataupun konten yang tidak terlalu lengkap menjadi alasan untuk meminjam buku lebih dari satu. Namun mengetahui buku lain yang terkait dengan buku yang sedang kita baca juga merupakan hal yang tidak mudah. Sebuah sistem rekomendasi buku diperlukan untuk mengatasi permasalahan ini. Sistem rekomendasi yang umum dijumpai pada umumnya mengakses catatan peminjaman buku dari peminjam. Namun, pada kasus-kasus tertentu catatan peminjaman buku dari peminjam belum tentu dapat diakses ataupun tidak ada catatan sama sekali.

Pada penelitian ini akan dibahas mengenai sistem rekomendasi yang merekomendasikan buku berdasarkan konten dari buku tersebut sehingga masalah dari catatan peminjaman buku dapat diatasi. Sistem ini menggunakan algoritma *Fuzzy C-Means* untuk mengelompokkan buku-buku ke dalam *cluster-cluster* terlebih dahulu sebelum menghitung kedekatan antar buku menggunakan *cosine similarity*. Pengelompokan buku dimaksudkan agar penghitungan kedekatan dari suatu buku tidak dilakukan ke semua kumpulan buku, melainkan hanya pada kelompok-kelompok yang sudah dibentuk saja.

Hasil rekomendasi buku berbasis konten yang menggunakan algoritma *Fuzzy C-Means* dan *Cosine Similarity* berdasarkan survei terhadap para responden untuk 14 sampel buku menunjukkan tingkat kesesuaian hasil rekomendasi yang cukup baik, yaitu sebesar 89,394%.

Kata kunci: Cluster, Fuzzy C-Means, Rekomendasi, Cosine Similarity

DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMAKASIH.....	vi
INTISARI.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Metodologi Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Tinjauan Pustaka.....	6
2.2 Landasan Teori.....	7
2.2.1 Text Mining.....	7
2.2.2 Preprocessing.....	7
2.2.3 Pembobotan TF-IDF.....	8
2.2.4 Clustering.....	9
2.2.5 Fuzzy C-Means.....	9
2.2.6 Sistem Rekomendasi.....	12
2.2.7 Vector Space Model.....	12
BAB 3 ANALISIS DAN PERANCANGAN SISTEM.....	13
3.1 Spesifikasi Kebutuhan Perangkat Keras dan Perangkat Lunak.....	13

3.1.1 Spesifikasi Kebutuhan Perangkat Keras.....	13
3.1.2 Spesifikasi Kebutuhan Perangkat Lunak.....	13
3.2 Arsitektur Sistem.....	14
3.3 Use Case Diagram.....	15
3.4 Flowchart.....	16
3.5 Skema Basis Data.....	20
3.6 Kamus Data.....	21
3.6.1 Tabel Books.....	21
3.6.2 Tabel Preprocessings.....	21
3.6.3 Tabel Recommendations.....	22
3.6.4 Tabel Stopwords.....	22
3.6.5 Tabel Token_Titles.....	23
3.6.6 Tabel Token_Prefaces.....	23
3.6.7 Tabel Token_Contents.....	24
3.6.8 Tabel Token_All.....	25
3.7 Perancangan Antarmuka.....	26
3.8 Perancangan Pengujian Sistem.....	30
3.8.1 Profil Buku.....	30
3.8.2 Pengujian Hasil Rekomendasi.....	32
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....	36
4.1 Implementasi Sistem.....	36
4.1.1 Tahapan Memasukkan Buku Baru.....	36
4.1.2 Tahapan Preprocessing, Tokenisasi, dan Pembobotan TF-IDF...	37
4.1.2.1 Tahapan Preprocessing.....	37
4.1.2.2 Tahapan Tokenisasi.....	38
4.1.2.3 Tahapan Pembobotan TF-IDF.....	38
4.1.3 Tahapan Normalisasi Bobot Token.....	39
4.1.4 Tahapan Clustering.....	41
4.1.5 Tahapan Penghitungan Cosine Similarity.....	46
4.2 Analisis Sistem.....	47
4.2.1 Analisis Hasil Rekomendasi.....	47
BAB 5 KESIMPULAN DAN SARAN.....	49

5.1 Kesimpulan.....	49
5.2 Saran.....	49
DAFTAR PUSTAKA.....	50
LAMPIRAN	

©UKDWN

DAFTAR TABEL

Tabel 3.1	Struktur Tabel Books.....	21
Tabel 3.2	Struktur Tabel Preprocessings.....	21
Tabel 3.3	Struktur Tabel Recommendations.....	22
Tabel 3.4	Struktur Tabel Stopwords.....	22
Tabel 3.5	Struktur Tabel Token_Titles.....	23
Tabel 3.6	Struktur Tabel Token_Prefaces.....	23
Tabel 3.7	Struktur Tabel Token_Contents.....	24
Tabel 3.8	Struktur Tabel Token_All.....	25
Tabel 3.9	Profil Buku.....	31
Tabel 4.1	Tabel Hasil Evaluasi Rekomendasi.....	48

DAFTAR GAMBAR

Gambar 2.1	Tahapan dalam Text Mining.....	7
Gambar 3.1	Arsitektur Sistem Rekomendasi Buku Berbasis Konten.....	14
Gambar 3.2	Use Case Diagram Sistem.....	15
Gambar 3.3	Flowchart Sistem.....	16
Gambar 3.4	Sub-Flowchart Preprocessing.....	17
Gambar 3.5	Sub-Flowchart Tokenisasi.....	18
Gambar 3.6	Sub-Flowchart Pembobotan TF-IDF dan Normalisasi.....	19
Gambar 3.7	Skema Diagram Basis Data.....	20
Gambar 3.8	Perancangan Halaman Penginputan Konten Buku.....	26
Gambar 3.9	Perancangan Halaman Preprocessing.....	26
Gambar 3.10	Perancangan Halaman Hasil Pembobotan.....	27
Gambar 3.11	Perancangan Halaman Clustering.....	28
Gambar 3.12	Perancangan Halaman Hasil Clustering.....	28
Gambar 3.13	Perancangan Halaman Hasil Pencarian.....	29
Gambar 3.14	Perancangan Halaman Menampilkan Hasil Rekomendasi.....	30
Gambar 4.1	Halaman Memasukkan Buku Baru.....	36
Gambar 4.2	Halaman Preprocessing.....	37
Gambar 4.3	Halaman Normalisasi Bobot.....	39
Gambar 4.4	Halaman Clustering.....	41
Gambar 4.5	Halaman Hasil Clustering.....	45
Gambar 4.6	Halaman Set Rekomendasi.....	46

DAFTAR LAMPIRAN

Listing Code <i>process_insertnewbook.php</i>	A-1
Listing Code <i>lowercase.php</i>	A-1
Listing Code <i>title_tokenization.php</i>	A-3
Listing Code <i>title_tfidf.php</i>	A-4
Listing Code <i>preface_tokenization.php</i>	A-5
Listing Code <i>preface_tfidf.php</i>	A-7
Listing Code <i>content_tokenization.php</i>	A-8
Listing Code <i>content_tfidf.php</i>	A-9
Listing Code <i>normalization.php</i>	A-11
Listing Code <i>fcm.php</i>	A-14
Listing Code <i>similarity.php</i>	A-19

©UKUDN

INTISARI

Mempelajari sesuatu hanya dari sebuah buku jangkit pinjam dari perpustakaan terkadang tidak cukup. Pembahasan yang belum terlalu mendalam ataupun konten yang tidak terlalu lengkap menjadi alasan untuk meminjam buku lebih dari satu. Namun mengetahui buku lain yang terkait dengan buku yang sedang kita baca juga merupakan hal yang tidak mudah. Sebuah sistem rekomendasi buku diperlukan untuk mengatasi permasalahan ini. Sistem rekomendasi yang umum dijumpai pada umumnya mengakses catatan peminjaman buku dari peminjam. Namun, pada kasus-kasus tertentu catatan peminjaman buku dari peminjam belum tentu dapat diakses ataupun tidak ada catatan sama sekali.

Pada penelitian ini akan dibahas mengenai sistem rekomendasi yang merekomendasikan buku berdasarkan konten dari buku tersebut sehingga masalah dari catatan peminjaman buku dapat diatasi. Sistem ini menggunakan algoritma *Fuzzy C-Means* untuk mengelompokkan buku-buku ke dalam *cluster-cluster* terlebih dahulu sebelum menghitung kedekatan antar buku menggunakan *cosine similarity*. Pengelompokan buku dimaksudkan agar penghitungan kedekatan dari suatu buku tidak dilakukan ke semua kumpulan buku, melainkan hanya pada kelompok-kelompok yang sudah dibentuk saja.

Hasil rekomendasi buku berbasis konten yang menggunakan algoritma *Fuzzy C-Means* dan *Cosine Similarity* berdasarkan survei terhadap para responden untuk 14 sampel buku menunjukkan tingkat kesesuaian hasil rekomendasi yang cukup baik, yaitu sebesar 89,394%.

Kata kunci: Cluster, Fuzzy C-Means, Rekomendasi, Cosine Similarity

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perpustakaan merupakan salah satu tempat yang untuk mendapatkan informasi pada zaman sekarang karena di dalamnya terdapat banyak sekali buku yang bermanfaat. Masing-masing buku tersebut tentu memiliki topik bahasannya masing-masing. Namun, ternyata ada keterkaitan antara topik bahasan buku yang satu dengan yang lain (Subagyo, 2009). Keterkaitan antar buku ini bermanfaat untuk saling melengkapi informasi yang dibutuhkan oleh peminjam buku, karena terkadang referensi dari satu buku saja belum cukup memenuhi kepuasan peminjam buku yang mencari informasi lebih lanjut.

Keterkaitan topik bahasan antar buku sebenarnya bisa disimpulkan oleh peminjam buku yang sudah benar-benar paham mengenai topik bahasan tersebut. Namun, tidak menutup kemungkinan ada kasus di mana calon peminjam buku adalah orang yang benar-benar awam tentang topik bahasan dari buku yang akan dipinjamnya, sehingga orang tersebut tidak mengetahui buku mana saja yang kira-kira masih berkaitan dengan topik bahasan tersebut (Han & Karypis, 2005). Oleh karena itu perlu adanya sistem untuk merekomendasi buku agar peminjam buku mengetahui buku-buku mana saja yang terkait dengan buku yang akan dicari. Dengan adanya rekomendasi dari buku-buku yang berkaitan dengan buku utama yang dicari, maka peminjam buku diharapkan dapat mendapatkan pengetahuan yang lebih mendalam lagi dari referensi buku lain yang tepat.

Metode rekomendasi buku yang akan dibahas dalam penelitian ini memanfaatkan metode *clustering*, yang nantinya sistem akan menampilkan beberapa judul buku yang paling berkaitan dalam satu kelompok / *cluster* yang telah terbentuk sebelumnya. Dengan memanfaatkan *cluster-cluster* yang sudah terbentuk, diharapkan hasil rekomendasi akan lebih relevan dan cepat. Proses

perekomendasi yang dibahas di penelitian ini dilakukan berdasarkan pendekatan berbasis konten buku, yang meliputi judul buku, daftar isi, abstraksi buku, dan nama penerbit. Perekomendasi berdasarkan konten dipilih oleh penulis untuk mengatasi kelemahan perekomendasi berbasis *collaborative filtering* yang pendekatannya menggunakan catatan aktivitas pengguna sistem yang untuk beberapa kasus tertentu belum tentu dapat diizinkan pengaksesannya.

Pada penelitian ini, penulis mengembangkan sistem rekomendasi buku berbasis konten dengan menggunakan algoritma *clustering Fuzzy C-Means* serta metode *Vector Space Model*. Pemanfaatan algoritma *clustering* ini diharapkan selain dapat menguji algoritma itu sendiri, juga dapat mempercepat proses rekomendasi berbasis konten yang pengukurannya menggunakan prinsip *Vector Space Model*, yaitu *Cosine Similarity*.

1.2 Perumusan Masalah

Rumusan masalah yang dapat dibuat dalam tugas akhir ini adalah

- a) Bagaimana mengimplementasikan *Fuzzy C-Means* dan *Cosine Similarity* untuk menghasilkan rekomendasi buku berbasis konten ?
- b) Seberapa tepatkah hasil rekomendasi sistem dalam menampilkan buku-buku yang terkait ?

1.3 Batasan Masalah

Parameter-parameter pembatas dalam penelitian ini adalah sebagai berikut:

- a) Buku-buku yang akan digunakan untuk penelitian adalah 60 buku berbahasa Inggris terkait dengan bidang teknologi informasi yang ada di perpustakaan Universitas Kristen Duta Wacana.
- b) Konten-konten yang diproses saat *clustering* adalah judul, daftar isi, dan kata pengantar buku.

- c) Proses *stemming*¹ pada tahap *preprocessing* tidak dilakukan karena sampai saat ini belum ada metode *stemming* yang mampu menemukan kata dasar dari suatu kata secara sempurna. Begitu juga dengan *Wordnet*, karena terdapat istilah-istilah di bidang teknologi informasi yang bisa jadi punya arti lain dalam kata sehari-hari, misalkan: *bash*, *shell*, dan lain-lain.
- d) Bobot pangkat (*m*) dalam algoritma *Fuzzy C-Means* yang dipakai adalah 2.
- e) Sistem rekomendasi secara *default* hanya akan merekomendasikan 3 buku yang paling berkaitan dengan buku acuan.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menguji keefektifan algoritma Fuzzy C-Means dalam melakukan proses *clustering*, serta Vector Space Model pada sistem rekomendasi buku berbasis konten berbahasa Inggris dengan studi kasus di perpustakaan Universitas Kristen Duta Wacana.

1.5 Metodologi Penelitian

- a) Studi Pustaka
Mempelajari teori dan implementasi melalui buku, artikel, jurnal, dan referensi lain yang berhubungan dengan *clustering* dokumen dan algoritma *Fuzzy C-Means*.
- b) Pengumpulan *Sample Data*
Proses pengumpulan data berupa fitur dari 60 buku teks berbahasa Inggris terkait dengan bidang teknologi informasi di perpustakaan Universitas Kristen Duta Wacana.

¹ Proses mendapatkan kata dasar dari suatu kata dengan memotong imbuhan kata tersebut yang dapat terdiri dari awalan dan akhiran.

- c) Implementasi Sistem
Melakukan *preprocessing*, *clustering*, dan rekomendasi menggunakan *Vector Space Model* terhadap fitur-fitur buku yang sudah didapatkan dari pengumpulan *sample data*.
- d) Evaluasi Rekomendasi
Evaluasi sistem rekomendasi dilakukan dengan memberikan kuesioner kepada 11 sukarelawan. Kuesioner berisi 14 buku acuan yang masing-masing buku acuan tersebut memiliki 3 rekomendasi buku. Sukarelawan dipersilahkan untuk menandai masing-masing rekomendasi buku tersebut apakah sudah sesuai dengan buku acuan atau tidak. Penulis lalu menjumlah persentase kesesuaian hasil rekomendasi untuk masing-masing responden untuk kemudian dirata-rata untuk menyimpulkan keakuratan rekomendasi sistem yang dibangun.
- e) Penyusunan Laporan
Membuat laporan penelitian dari hasil kinerja algoritma terhadap studi kasus.

1.6 Sistematika Penulisan

Penulis membagi laporan tugas akhir ini menjadi lima bab. Bab pertama adalah bab Pendahuluan yang berisi tentang penjelasan umum mengenai penelitian yang akan diteliti oleh penulis. Bab pendahuluan ini terdiri dari enam sub bab, antara lain : Latar Belakang Masalah, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Metode Penelitian, dan Sistematika Penulisan.

Bab kedua adalah bab Tinjauan Pustaka. Bab ini terdiri dari dua sub bab, yaitu: Tinjauan Pustaka dan Landasan Teori. Pada sub bab Tinjauan Pustaka membahas tentang berbagai teori atau penelitian dari sumber-sumber pustaka yang digunakan untuk penyusunan laporan tugas akhir, seperti penjelasan tentang *clustering* data dan *Fuzzy C-Means*. Pada sub-bab Landasan Teori membahas tentang penjabaran konsep-konsep yang digunakan dalam penelitian.

Bab ketiga adalah bab Analisis dan Perancangan Sistem yang mencakup perancangan sistem yang akan dibuat, yakni mengenai kebutuhan *hardware* dan *software*, spesifikasi sistem, diagram sistem, arsitektur sistem, skema basis data, dan perancangan antar muka.

Bab keempat adalah Implementasi dan Analisis Sistem yang memuat hasil implementasi dan pembahasan mengenai pengujian sistem yang akan dibuat oleh penulis. Hasil dari sistem yang sudah dibangun dan analisis dari sistem yang telah dibuat juga dimuat dalam bab ini.

Bab kelima adalah bab Kesimpulan dan Saran memuat tentang kesimpulan dari hasil penelitian yang dilakukan oleh penulis dan saran untuk memberikan analisis dan gambaran tentang pengembangan sistem yang lebih baik lagi pada masa mendatang tentang topik yang serupa.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil implementasi dan analisis sistem yang dilakukan oleh penulis, maka didapat kesimpulan-kesimpulan sebagai berikut:

- a) Metode *Fuzzy C-Means* dan *Cosine Similarity* dapat diimplementasikan untuk menghasilkan rekomendasi buku berbasis konten dengan menerapkan *threshold* derajat keanggotaan pada saat *clustering* serta menghitung *cosine similarity* buku acuan terhadap buku-buku lain yang ada di *cluster-cluster* yang sama.
- b) Dalam analisis evaluasi hasil rekomendasi terhadap 11 responden dan 14 sampel buku, didapati bahwa persentase rata-rata kesesuaian hasil rekomendasi dengan buku acuan adalah sebesar 89,394%.

5.2 Saran

Demi pengembangan dan perbaikan sistem ke depan, penulis memberikan beberapa saran sebagai berikut:

- a) Dapat mencoba pereliasian daftar token ke *database Wordnet* tetapi dengan menghilangkan kata-kata yang merupakan istilah-istilah di bidang teknologi informasi untuk menghindari kata-kata yang ambigu dengan kata sehari-hari, misalkan: *bash*, *shell*, dll.
- b) Dapat mencoba proses *lemmatization* pada saat *preprocessing*.
- c) Untuk bagian kata pengantar buku, dapat dilakukan proses peringkasan teks terlebih dahulu agar data yang didapat merupakan kalimat-kalimat yang efektif.

DAFTAR PUSTAKA

- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The Fuzzy C-Means Clustering Algorithm. *Computers & Geosciences*, 10, 191-203.
- Deepa, M., & Revathy, P. (2012). Validation of Document Clustering based on Purity and Entropy Measures. *International Journal of Advanced Research in Computer and Communication Engineering*, 147-152.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Book: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.
- Grover, N. (2014). A Study of Various Fuzzy Clustering Algorithms. *International Journal of Engineering Research*, 3(3), 177-181.
- Han, E. H., & Karypis, G. (2005). Feature-Based Recommendation System. *Conference on Information and Knowledge Management*, 446-452.
- Konchady, M. (2006). *Text Mining Application Programming*. Boston: Charles River Media.
- Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2010). *Recommender Systems Handbook*. New York: Springer.
- Srinivasa, K. G., Venugopal, K. R., & Patnaik, L. M. (2006). Feature Extraction using Fuzzy C-Means Clustering for Data Mining System. *International Journal of Computer Science and Network Security*, 230-236.
- Suganya, R., & Shanti, R. (2012). Fuzzy C-Means Algorithm - A Review. *International Journal of Scientific and Research Publications*, 2(11), 1-3.