



WHAT DO USERS WRITE IN ONE FILE CABINET DATABASES: AN ANALYSIS BASED ON TEXT MINING

Natalis Ransi^{*1}, La Surimi², Jumadil Nangi³, Adha Mashur Sajiah⁴, Arman⁵, Edi Cahyono⁶

^{1,2}Department of Computer Science, Universitas Halu Oleo

^{3,4}Department of Informatic Engineering, Universitas Halu Oleo

^{5,6}Department of Mathematic, Universitas Halu Oleo

e-mail: ^{*1}natalis.ransi@uho.ac.id, ²lasurimi@uho.ac.id, ³jumadilnangi87@gmail.com,

⁴adha.m.sajiah@gmail.com, ⁵arman.mtmk@uho.ac.id, ⁶edi.cahyono@uho.ac.id

Abstrak

One file cabinet adalah sistem terintegrasi yang salah satu fungsinya untuk merekam catatan kerja (*logbook*) dosen dan tenaga kependidikan di Universitas Halu Oleo. Dalam *logbook*, setiap hari kerja dosen dan tenaga kependidikan menuliskan deskripsi kerja yang telah mereka lakukan dalam sistem *database*. Artikel ini membahas analisis kumpulan kata-kata yang dituliskan oleh dosen dan tenaga kependidikan pada *logbook* mereka. Metode yang digunakan adalah penambangan teks kata-kata (*text mining*). Hasil analisis diperlukan oleh pimpinan (*top managers*) Universitas Halu Oleo dalam memperoleh gambaran global kinerja dosen dan tenaga kependidikan. Informasi ini penting untuk menentukan kebijakan dalam pengembangan sumber daya manusia.

Kata kunci; *Database, Logbook, One File Cabinet, Sumber Daya Manusia, Text Mining*

Abstract

One file cabinet is an integrated system that one of its functions is to record work notes (logbook) lecturers and educational personnel at Halu Oleo University. In the logbook, every weekday lecturers and education workers write a description of the work they have done in the database system. This article discusses the analysis of a collection of words written by lecturers and educational personnel in their logbooks. The method used is text mining and the results of the analysis are required by the leaders (top managers) of Halu Oleo University in obtaining a global picture of the performance of lecturers and educational personnel. This information is important for determining policies in human resource development.

Keywords; *Database, Logbook, One File Cabinet, Human Resources, Text Mining*

1. PENDAHULUAN

Transformasi data menjadi informasi pada dokumen teks, misalnya membuat ringkasan masih merupakan kegiatan yang terus mendapat perhatian bagi praktisi di teknologi informasi dan komunikasi. Penerapannya pada aplikasi perangkat lunak komputer yang mengumpulkan document teks secara masal, misalnya *social media* masih

terus dilakukan [1], [2], [3], [4], [5], [6]. Demikian juga pada aplikasi-aplikasi yang dikembangkan secara mandiri oleh beberapa organisasi [7], [8]. Hal tersebut dimulai dengan merekayasa desain tempat penyimpanan data sampai dengan mengaplikasikan metode transformasi data menjadi informasi menggunakan algoritma tertentu.

Salah satu tahapan dalam melakukan transformasi data menjadi informasi atau



pengetahuan adalah metode *data mining*. *Data mining* yang lebih khusus menangani data berupa teks disebut *text mining*. *Text mining* sudah diterapkan pada beberapa bidang ilmu yang ada saat ini [9], [10].

Selain membuat ringkasan, tugas *text mining* yang lain adalah membantu peneliti untuk memahami objek yang sedang diteliti [11], [12]. Pengaplikasian lebih spesifik tentang penerapan *text mining* pada bidang kesehatan [13], biomedis [14], marketing [4], pendidikan [15], energi [16], transportasi [17], [5], juga telah dilaporkan. Dokumen teks yang dimiliki dimanfaatkan untuk memperoleh pengetahuan baru. Pengetahuan baru itu digunakan sebagai data dukung dalam memutuskan tindakan yang diambil ataupun dalam merencanakan sebuah strategi yang akan diterapkan.

Text mining juga dilakukan untuk membantu proses evaluasi laporan, misalnya pada pengembangan metode dalam pemeriksaan konsistensi laporan sebuah organisasi, ekstraksi pola dari sekumpulan dokumen teks telah membantu peneliti di bidang ekonomi untuk mengetahui kata, topik serta pengelompokannya dalam mewakili ukuran inflasi dan beberapa sektor seperti jasa atau hotel dan akomodasi serta jenis barang yang tahan lama atau tidak. Dokumen teks yang digunakan adalah laporan *The Monthly Price Development Reports (MPDR) of the Central Bank of Republic of Turkey (CBRT)* antara tahun 2006-2018. Hasil analisis menunjukkan bahwa MPDR konsisten dengan angka *consumer price index (CPI)* tahunan [18].

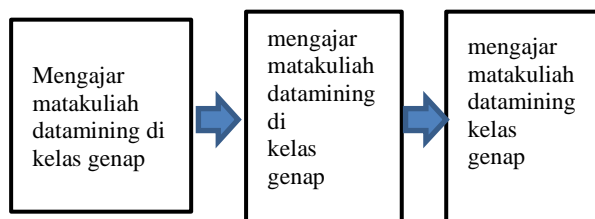
Kepuasan publik dalam hal layanan yang diberikan oleh pemerintah sebuah negara atau kota telah juga memanfaatkan *text mining*. Seperti halnya yang dilakukan oleh pemerintahan di Beijing. Mereka menggunakan *text mining* untuk memahami petisi warga sebagai dasar untuk meningkatkan tata kelola pemerintahan. *Text mining* membantu meringkas dan mengeksplorasi kekhawatiran umum yang disuarakan oleh warga. Hal ini dilaporkan oleh peneliti dari *Beijing Academy of Social Sciences* dan *Beijing Official Portal Management Center*, melaporkan bahwa pendaftaran rumah tangga (30,3%), konstruksi ilegal (22,6%), pendidikan (13,4%), perusakan (11,8%), manajemen kota (9,8%), perumahan (7,7%) dan lalu lintas (4,4%) merupakan petisi

warga yang dirangkum dari sebuah aplikasi yang disebut *Petition Network Platform* [19].

Beberapa penerapan yang telah dijelaskan menggunakan teknologi basis data untuk melakukan pengelolaan datanya. *One File Cabinet (OFC)* merupakan basis data teintegrasi yang dikembangkan oleh Universitas Halu Oleo (UHO) sebagai media pengelolaan data. Saat ini OFC menjadi tempat untuk memanajemen data sumber daya manusia, daftar hadir *online*, *logbook* mahasiswa, tenaga pendidik dan tenaga kependidikan. Pada makalah ini kami melaporkan hasil penerapan metode *text mining* untuk memperoleh kata apa saja yang sering digunakan oleh pengguna OFC.

2. METODE PENELITIAN

Adapun langkah atau alur penelitian yang dilakukan adalah studi pustaka, pengambilan data, *text mining*, penarikan kesimpulan. Khusus untuk *text mining* langkah penelitian yang dilakukan secara umum metode mengadopsi tahapan pada *Knowledge Discovery in Database (KDD)* yaitu *data selection, preprocessing, transformation, text mining, visualization*.



Gambar 1 Metode penelitian

2.1 Text Mining

Text mining merupakan variasi dari data mining yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. *Text mining* didefinisikan sebagai yang berupa teks yang biasanya sumber data didapatkan dari dokumen, dengan tujuan adalah mencari kata-kata yang dapat mewakili isi dari dokumen tersebut yang nantinya dapat dilakukan analisis hubungan antar dokumen.

Tahapan *text mining* secara umum dibagi menjadi beberapa tahapan umum [22].

a. Text Preprocessing

Text Preprocessing merupakan tahapan awal dari *text mining* yang bertujuan mempersiapkan teks menjadi data yang akan mengalami pengolahan pada tahap selanjutnya. Pada *text mining*, data mentah

yang berisi informasi memiliki struktur yang sembarang, sehingga diperlukan proses pengubahan bentuk menjadi data yang terstruktur sesuai kebutuhan, yaitu biasanya akan menjadi nilai-nilai numerik. Proses ini disebut *Text Preprocessing*. Pada tahap ini, tindakan yang dilakukan adalah *toLowerCase*, dengan mengubah semua karakter huruf menjadi huruf kecil, dan *tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat mejadi kata-kata kemudian menghilangkan delimiter-delimiter seperti tanda koma (,), tanda titik (.), spasi, dan karakter angka yang terdapat pada kata tersebut.

b. Seleksi Fitur (*Feature Selection*)

Pada tahap ini akan dilakukan seleksi dengan mengurangi jumlah kata-kata yang dianggap tidak penting dalam dokumen tersebut untuk menghasilkan proses pengklasifikasian yang lebih efektif dan akurat. Tahapan ini adalah dengan melakukan penghilangan *stopword* dan juga mengubah kata-kata kedalam bentuk dasar terhadap kata yang berimbuhan. *Stopword* merupakan kosakata yang bukan merupakan ciri atau kata unik dari suatu dokumen seperti kata sambung. Yang termasuk *stopword* yaitu “ di”, “pada”, ”sebuah”, ”karena”, ”oleh” dan sebagainya. Sebelum memasuki tahapan penghilang *stopword*, daftar *stopword* harus dibuat terlebih dahulu. Jika kata-kata yang termasuk *stopword* masuk dalam *stoplist*, maka kata tersebut akan dihapus dari deskripsi sehingga sisanya dianggap sebagai kata-kata yang mencirikan isi dokumen atau *keywords*.

c. *Stemming*

Stemming adalah proses pemetaan dan penguraian berbagai bentuk dari suatu kata menjadi kata dasarnya. Tujuan dilakukannya proses *stemming* adalah menghilangkan imbuhan-imbuhan berupa *prefix*, *suffix*, maupun konfiks yang terdapat pada setiap kata. Apabila imbuhan tadi tidak dihilangkan maka setiap kata akan disimpan di dalam *database*, sehingga nantinya akan menjadi beban di dalam *database*.

2.2 Orange dan *word cloud*

Orange adalah aplikasi perangkat lunak *machine learning* dan *data mining* untuk analisis data dengan menggunakan bahasa pemrograman Python dan pemrograman visual. Aplikasi ini didesain sedemikian rupa sehingga memiliki komponen berupa *toolbox/widget* dengan tingkat fleksibilitas yang tinggi [20]. Salah satu *widget* yang digunakan pada penelitian ini adalah *data management and preprocessing*. Fitur ini ditujukan untuk *data input and output*, *data filtering and sampling*, *feature manipulation* (*discretization*, *continuization*, *normalization*, *scaling* and *scoring*), dan *feature selection*.

2.3 *Word Cloud*

Word Cloud merupakan salah satu *widget* yang sangat powerful untuk keperluan visualiasi hasil preprocessing data. *Widget* ini menampilkan *corpus* yang terseleksi dari parameter yang diberikan kepada *document text*. *Widget* ini menerima inputan berupa *corpus* dari himpunan dokumen dan menghasilkan *output* berupa *corpus* yang memenuhi parameter yang diberikan dilengkapi dengan nilai frekuensi kemunculan setiap *corpus*. Selanjutnya *word cloud* menampilkan token (*corpus* terseleksi) dengan ukuran dan warna yang terasosiasi dengan frekuensi token tersebut [21].

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Penelitian ini menggunakan basis data OFC khususnya *logbook* mahasiswa, tenaga kependidikan dan tenaga pendidik. Periode data yang digunakan adalah *logbook* Bulan Juni-Oktober 2020.

Tabel 1 Data Penelitian

<i>Logbook</i>	Jumlah Dok.	<i>Segment</i> (karakter)	<i>Corpus</i> (kata)	<i>Tokens</i>
Tenaga Pendidik	77.849	5.661.218	776.489	241.025
Tenaga Kependidikan	36.806	2.307.853	306.798	85.866
Mahasiswa	40.498	7.931.699	1.124.052	221.253



- Marketing: A text mining and topic modeling based literature analysis,” *Eur. Res. Manag. Bus. Econ.*, vol. 24, no. 1, pp. 1–7, 2018.
- [5] E. Sugawara and H. Nikaido, “Properties of AdeABC and AdeIJK efflux systems of *Acinetobacter baumannii* compared with those of the AcrAB-TolC system of *Escherichia coli*,” *Antimicrob. Agents Chemother.*, vol. 58, no. 12, pp. 7250–7257, 2014.
- [6] A. Imam and H. Fajriab, “Implementasi Text Mining pada Mesin Pencarian Twitter untuk Menganalisis Topik - Topik Terkait ‘KPK dan Jokowi,’” *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS 2015*, pp. 570–581, 2015.
- [7] E. Nilawati, “Kecenderungan Isi Sistem Informasi Desa Kabupaten Gunungkidul (Village Information System Content Trends in Gunungkidul Regency),” *J. IPTEKKOM J. Ilmu Pengetah. Teknol. Inf.*, vol. 21, no. 2, p. 169, 2019.
- [8] J. N. Apriliana, Natalis Ransi, “Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naïve Bayes Classifier,” *Semant. Vol.3, No.2, Jul-Des 2017*, vol. 3, no. 2, pp. 187–194, 2017.
- [9] K. He *et al.*, “Understanding the patient perspective of epilepsy treatment through text mining of online patient support groups,” *Epilepsy Behav.*, vol. 94, pp. 65–71, 2019.
- [10] S. Chen, X. Guo, T. Wu, and X. Ju, “Exploring the Online Doctor-Patient Interaction on Patient Satisfaction Based on Text Mining and Empirical Analysis,” *Inf. Process. Manag.*, vol. 57, no. 5, p. 102253, 2020.
- [11] R. Raja, A. Coelho, S. Hemaiswarya, P. Kumar, I. S. Carvalho, and A. Alagarsamy, “Applications of microalgal paste and powder as food and feed: An update using text mining tool,” *Beni-Suef Univ. J. Basic Appl. Sci.*, vol. 7, no. 4, pp. 740–747, 2018.
- [12] S. Tachibana and K. Tsuda, “A study of a method to understand the intention of taste expressions through text mining,” in *Procedia Computer Science*, 2020.
- [13] B. K. Olorisade, P. Brereton, and P. Andras, “Reproducibility of studies on text mining for citation screening in systematic reviews: Evaluation and checklist,” *J. Biomed. Inform.*, vol. 73, pp. 1–13, 2017.
- [14] F. Zhu *et al.*, “Biomedical text mining and its applications in cancer research,” *Journal of Biomedical Informatics*, vol. 46, no. 2, pp. 200–211, 2013.
- [15] N. Süzen, A. N. Gorban, J. Levesley, and E. M. Mirkes, “Automatic short answer grading and feedback using text mining methods,” in *Procedia Computer Science*, 2020, vol. 169, 2019.
- [16] L. T. Zhao, S. Q. Guo, and Y. Wang, “Oil market risk factor identification based on text mining technology,” in *Energy Procedia*, 2019.
- [17] A. Serna and S. Gasparovic, “Transport Analysis Approach Based on Big Data and Text Mining Analysis from Social Media,” *Transp. Res. Procedia*, vol. 33, pp. 291–298, 2018.
- [18] H. B. Eskici and N. A. Koçak, “A text mining application on monthly price developments reports,” *Cent. Bank Rev.*, vol. 18, no. 2, pp. 51–60, 2018.
- [19] Z. Wang and Y. Zhong, “What were residents’ petitions in Beijing- based on text mining,” *J. Urban Manag.*, vol. 9, no. 2, pp. 228–237, 2020.
- [20] J. Demšar *et al.*, “Orange: Data mining toolbox in python,” *J. Mach. Learn. Res.*, vol. 14, pp. 2349–2353, 2013.
- [21] Biolab, “Orange3 Text Mining Documentation,” 2020. <https://readthedocs.org/projects/orange3-text/downloads/pdf/latest/>.