



Visualisasi Kumpulan Berita Dalam Bentuk Peta Digital Dengan Metode Term Frequency- Inverse Document Frequency dan Gazetteer

Wasis Saefudin, Agus Komarudin, Ridwan Ilyas*

Fakultas Sains dan Informatika, Program Studi Informatika, Universitas Jenderal Achmad Yani
Jl. Terusan Jenderal Sudirman 148 Cimahi, Jawa Barat, Indonesia
Email: ¹itsmewasis@gmail.com, ²agus.komarudin@lecture.unjani.ac.id, ³ilyas@lecture.unjani.ac.id

Abstrak

Seiring berkembangnya teknologi internet tidak hanya digunakan sebagai media hiburan saja tetapi juga dapat digunakan sebagai sumber informasi, sumber informasi yang dapat diperoleh di internet salah satunya adalah melalui berita. Di era saat ini membaca berita sudah menjadi kebutuhan informasi setiap orang dengan kemudahan membaca berita disetiap portal berita. Namun dengan banyaknya berita yang terbit setiap harinya membuat sebagian berita tidak terbaca. visualisasi berita dapat dilakukan untuk memetakan banyaknya berita berdasarkan lokasi berita. Untuk memecahkan masalah tersebut digunakan wordcloud untuk meyajikan sekumpulan berita agar didapat berita treding pada waktu dan lokasi berita. Penyajiannya menampilkan kata penting pada berita dengan menunjukkan frekuensi kemunculannya. Kata penting yang muncul dibedakan dengan ukuran hurufnya. kemunculan berita divisualkan menggunakan peta yang dipadukan dengan gazetteer untuk menentukan kordinat lokasi. Hasil yang diperoleh melalui 20 naracoba, perangkat lunak ini menghasilkan ketertarikan secara visual kepada pembaca sebanyak 68%, kesesuaian data yang ditampilkan wordcloud 65,33% dan 52,66% membantu dalam memudahkan menyimpulkan berita dalam jumlah banyak dan melihat tren berita. Sehingga perangkat lunak ini cukup baik dan masih membutuhkan beberapa penambahan dalam aplikasinya.

Kata Kunci: berita, portal berita, wordcloud, visualisasi, tf-id, gazeteer

1. PENDAHULUAN

Seiring berkembangnya teknologi berdampak pada penggunaan internet yang menjadikan internet sebagai kebutuhan pokok setiap manusia. *Internet* tidak hanya digunakan sebagai media hiburan saja tetapi juga dapat digunakan sebagai sumber informasi, Sumber tersebut dapat diperoleh di internet salah satunya adalah melalui berita. Sumber berita dapat berupa portal berita yang menyediakan kumpulan berita yang beraneka ragam setiap harinya, berita berkembang dengan pesat mengakibatkan semakin banyaknya jumlah berita dengan jumlah yang besar.

Berita yang tersebar di berbagai lokasi dan dimunculkan pada setiap portal berita berdampak pada pembaca yang sulit untuk mengetahui seluruh berita yang terjadi dan yang paling sering terjadi di lokasi tertentu. Oleh karena itu banyaknya berita merupakan masalah penting dan cukup rumit yang dapat diselesaikan dengan cara penyampaian yang baik.

Salah satu cara untuk menyajikan data dengan baik adalah menggunakan visualisasi informasi yang dapat mengasimilasikan berbagai data dalam satu waktu dalam bentuk gambar, peta, diagram ataupun animasi dalam satu tempat. Tidak semua bentuk visualisasi cocok untuk menunjukkan lokasi pemetaan, visualisasi peta adalah salah satu cara untuk menyajikan data pemetaan lokasi agar lebih interaktif seperti pada fasilitas penunjang wisata religi menggunakan map [1], pemanfaatan peta untuk pemetaan ruang ujian [2]. visualisasi pejualan dan lokasi penjualan pada sentral telepon otomatis [3]. Pemetaan pertambangan batu bara [4], pemetaan sinyal base transceiver station [5]. Pemetaan peta digital untuk kawasan bencana alam Kabupaten Polewali Mandar [6]. Dapat digunakan untuk menggambarkan pemetaan tempat terjadinya tindak kejahatan multilevel [7]. Namun pengambilan gambaran mengenai garis besar berita untuk mengambil keputusan belum dapat diperoleh, salah satu cara adalah menggunakan wordcloud melalui kolerasi antara masing-masing kata yang dapat terlihat.

Wordcloud adalah visualisasi teks dengan memunculkan susunan kata sebagai citra visual terkait frekuensi kemunculan kata dalam sebuah dokumen. Sehingga dapat menjadi alat bantu dalam melakukan analisis terhadap sebuah tulisan [8] [9]. analisis sentimen pelanggan *wifiid* pada *twitter* yang divisualisasikan menggunakan *wordcloud* dengan pembobotan menggunakan TF IDF [10].

TF-IDF dapat diimplementasikan untuk peringkasan multi dokumen berita yang pembobotannya berdasarkan *trending issue* [11]. TF-IDF dapat gunakan untuk pencarian karya ilmiah yang diukur dengan kemiripan dengan *keyword* yang didapat dari ekstraksi teks dokumen [12]. TF-IDF dapat digunakan untuk peringkasan *tweet* berdasarkan *trending topic twitter* [13].

2. METODOLOGI PENELITIAN

Metode penelitian terdiri dari lima tahap yaitu, pengambilan dan pengumpulan data, ekstraksi dokumen dan perhitungan bobot, perancangan perangkat lunak dan pembuatan perangkat lunak, pengujian dan evaluasi, pelaporan dan publikasi ilmiah.

Tahap pertama yaitu pengambilan data dilakukan dengan mengumpulkan data dari portal berita yang dikenal dengan web scraping kemudian dipilah menjadi beberapa komponen. Tahap kedua ekstraksi dokumen dan perhitungan bobot dilakukan dengan menghilangkan kata tidak penting pada isi berita, kemudian perhitungan bobot pada setiap dokumen berita dengan menggunakan metode *TF-IDF*, Tahap ketiga visualisasi wordcloud dan pemetaan lokasi, Tahap keempat perancangan perangkat lunak dan pembuatan sistem, Tahap kelima pengujian dan evaluasi.

2.1 Pengumpulan Data

Tahap pertama pengambilan data pada visualisasi berita menjelaskan proses data dikumpulkan dari portal berita. Berita diambil menggunakan teknik *webscraping*, berita yang diambil berupa satu halaman dan masih memiliki *noise*.

2.2 Ekstraksi Dokumen dan Perhitungan Bobot

Pada tahap ekstraksi dokumen data yang sudah diambil dilakukan *preprocessing* yang merupakan tahap dimana data yang akan digunakan dibersihkan dari *noise* atau dari hal yang tidak mempunyai pengaruh. *Preprocessing* yang dilakukan dalam tahap ini antara lain:

- Casefolding* merubah semua huruf yang terdapat pada dokumen menjadi huruf kecil atau huruf besar.
- Tokenizing* memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata.
- Filtering* menghilangkan kata tidak penting menggunakan algoritma *stoplist/stopword*

Tahap selanjutnya adalah Pembobotan kata dilakukan menggunakan TF-IDF untuk menentukan bobot dari tiap kata didalam teks berita, terdapat beberapa tahap untuk mendapatkan nilai bobot, yaitu:

- Menghitung term frekuensi pada dokumen, diasumsikan term memiliki kepentingan yang sebanding pada setiap dokumen.

$$TF(d,t) = f(d, t) \quad (2.1)$$

Dimana $f(d, t)$ adalah frekuensi kemunculan term t pada dokumen d .

- Menghitung inverse document frequency pada dokumen, Semakin sedikit jumlah dokumen yang mengandung term yang dimaksud, maka nilai IDF semakin besar.

$$IDF_t = LOG \left(\frac{D}{df_t} \right) \quad (2.2)$$

Dimana $df(t)$ jumlah dokumen yang mengandung term t dan D adalah jumlah dokumen yang mengandung kata kunci

- Menghitung nilai bobot (w) dari suatu dokumen yang diambil dari nilai TF dan nilai IDF

$$W(d,t) = TF(d,t) \times DF(t) \quad (2.3)$$

Dimana w adalah bobot dari setiap dokumen yang mengandung kata kunci

2.3 Tahap visualisasi Wordcloud dan Pemetaan Lokasi

Pada tahap ini term yang sudah memiliki bobot ditampilkan dalam bentuk wordcloud kemudian ditampilkan berdasarkan bobotnya jika bobot besar maka term yang muncul akan semakin besar.

Pada tahap selanjutnya setelah visualisasi wordcloud adalah pemetaan lokasi berita yang dilakukan dengan menampilkan polygon lokasi berita pada peta digital.

3. ANALISA DAN PEMBAHASAN

Analisis data masukan pada visualisasi berita merupakan data berita dari satu halaman portal berita yang masih memiliki *noise*, kemudian dipilah menjadi beberapa bagian saja seperti judul, penulis, penerbit, isi. Contoh data berita yg sudah dipilah seperti pada Table 1.

Tabel 1. Konten berita

Judul berita	Jateng, Provinsi Pertama yang Terapkan Penggunaan Obligasi Daerah
Penerbit	Kompas
Penulis	Kontributor Semarang, Nazar Nurdin
Lokasi	Semarang
Isi	Pemerintah Provinsi Jawa Tengah Jateng serius menjadikan obligasi daerah sebagai salah satu sumber pembiayaan untuk proyek-proyek infrastrukturnya. Saat ini, Jawa Tengah telah resmi dijadikan pilot proyek pertama daerah yang akan menerapkan obligasi daerah. Pemerintah pusat sudah menetapkan Jateng sebagai pilot project obligasi daerah. Ini yang pertama kali di Indonesia, setelah dulu Jawa Barat dan DKI Jakarta hampir berhasil," Yogyakarta Bambang Kiswono saat Ngopi bareng Media di Semarang, Senin Ia mengatakan, keinginan mencari pembiayaan alternatif itu Jateng telah membentuk tim percepatan yang ditarget selesai di akhir tahun 2018. Pada 2019, Jateng ditarget sudah mendapat pembiayaan dari pasar modal "2019 diharap nanti bisa jual. Tim percepatan itu dari OJK, Kemendagri, Kemenkeu dan Pemprov Jateng itu sendiri. Sudah ada perusahaan penilai yang melakukan penilaian," ujarnya. "Kami harap Januari sudah dapat dipasarkan," pungkasnya.

Analisis *Preprocessing* merupakan langkah awal dilakukan pemrosesan teks guna membuat teks menjadi lebih terstruktur, Pertama dilakukan casefolding, Tokenizing dan filtering.

Tabel 2. Casefolding

Artikel Berita
pemerintah provinsi jawa tengah jateng serius menjadikan obligasi daerah sebagai salah satu sumber pembiayaan untuk proyek-proyek infrastrukturnyasaat ini, jawa tengah telah resmi dijadikan pilot proyek pertama daerah yang akan menerapkan obligasi daerah pemerintah pusat sudah menetapkan jateng sebagai pilot project obligasi daerah. ini yang pertama kali di indonesia, setelah dulu jawa barat dan dki jakarta hampir berhasil," ujar kepala ojk regional 3 jawa tengah dan daerah istimewa yogyakarta bambang kiswono saat ngopi bareng media di semarang. senin ia mengatakan, pemprov jateng telah melangkah cukup jauh untuk merealisasikan keinginan mencari pembiayaan alternatif itu.jateng telah membentuk tim percepatan yang ditarget selesai di akhir tahun 2018. pada 2019, jateng ditarget sudah mendapat pembiayaan dari pasar modal"2019 diharap nanti bisa jual. tim percepatan itu dari ojk, kemendagri, kemenkeu dan pemprov jateng itu sendiri. sudah ada perusahaan penilai yang melakukan penilaian," ujarnya. "kami harap januari sudah dapat dipasarkan," pungkasnya.

Proses *tokenizing* kata yaitu menghilangkan karakter pemisah (*delimiter*) yang menyusunnya berupa karakter spasi. Berikut hasil dari *tokenizing* terdapat pada tabel 3.

Table 3. Tokenizing

infrastrukturnya	daerah	yogyakarta	membentuk	percepatan
saat	ini	bambang	tim	itu
ini	yang	kiswono	percepatan	dari
jawa	pertama	saat	yang	ojk
tengah	kali	ngopi	ditarget	kemendagri
telah	di	bareng	selesai	kemenkeu
resmi	indonesia	media	di	dan
dijadikan	setelah	di	akhir	pemprov
pilot	dulu	semarang	tahun	jateng
proyek	jawa	senin	2018	itu
pertama	barat	ia	pada	sendiri
daerah	dan	mengatakan	2019	sudah
yang	dki	pemprov	jateng	ada
akan	jakarta	jateng	ditarget	perusahaan
menerapkan	hampir	telah	sudah	penilai
obligasi	berhasil	melangkah	mendapat	yang
daerah	ujar	cukup	pendanaan	melakukan
pemerintah	kepala	jauh	dari	penilaian
pusat	ojk	untuk	pasar	ujarnya
sudah	regional	merealisasikan	modal	kami
menetapkan	3	keinginan	2019	harap
jateng	jawa	mencari	diharap	januari
sebagai	tengah	pendanaan	nanti	sudah
pilot	dan	alternatif	bisa	dapat
project	daerah	itu.jateng	jual	dipasarkan
obligasi	istimewa	telah	tim	pungkasnya

Filtering menggunakan algoritma *Stopword Removal* merupakan penghilangan kata dalam sebuah dokumen, seperti kata "dari," "adalah," "atau," "sebuah", dan lain-lain dalam dokumen bahasa Indonesia. Hasil dari *tokenizing* kemudian dilakukan proses *stop word removal* ini. Berikut ini adalah hasil dari proses *stop word removal* pada tabel 4.

Tabel 4. Filtering

infrastruktur	perusahaan	alternatif	jateng
jawa	berhasil	jateng	perusahaan
tengah	ujar	membentuk	penilai
resmi	kepala	tim	penilaian
dijadikan	ojk	percepatan	januari
pilot	regional	dipasarkan	pungkasnya
proyek	jawa	ditarget	jateng
daerah	tengah	selesai	penilaian
menerapkan	daerah	akhir	dipasarkan



obligasi	istimewa	tahun
daerah	yogyakarta	2018
pemerintah	bambang	2019
pusat	kiswono	jateng
menetapkan	ngopi	ditarget
jateng	bareng	pembiayaan
januari	media	pasar
pilot	semarang	modal
project	senin	2019
obligasi	pungkasnya	diharap
daerah	pemprov	jual
pertama	jateng	tim
indonesia	melangkah	percepatan
jawa	merealisasikan	ojk
barat	keinginan	kemendagri
dki	mencari	kemenkeu
jakarta	pembiayaan	pemprov

Data yang sudah melewati tahap preprosesing kemudian diberi bobot dengan menggunakan TF-IDF, langkah pertama mencari *term frequency* dan didapat hasil seperti tabel 5.

Tabel 5. mencari TF

Term	Kemunculan term	
	Berita 1	Berita 2
infrastruktur	1	0
perusahaan	2	0
jawa	3	0
jateng	6	0
berhasil	1	0
tengah	2	0
ujar	1	0
membentuk	1	0
penilai	1	0
resmi	1	0
kepala	1	0
tim	2	0
nilaian	1	0
dijadikan	1	0
ojk	2	0
percepatan	2	0
januari	2	0
pilot	2	0
regional	1	0
dipasarkan	2	0
pungkasnya	2	0
proyek	1	0
ditarget	2	0
daerah	4	0
selesai	1	0
penilaian	1	0
menerapkan	1	0
obligasi	2	0
istimewa	1	0
tahun	1	0
yogyakarta	1	0
menteri	0	2
kelautan	0	2
perikanan	0	2
susi	0	4
pudjiastuti	0	1



kekhawatiran	0	1
bocornya	0	1
garam	0	22
impor	0	8
pasaran	0	1
menteri	0	2
kelautan	0	2
perikanan	0	2
susi	0	4

Langkah pembobotan selanjutnya setelah mencari TF adalah mencari DF dan didapat hasil seperti pada tabel 6

Tabel 6 Mencari DF

Term	Kemunculan Term	Term	Kemunculan Term
infrastruktur	1	paparnya	1
perusahaan	1	rapat	1
alternatif	1	kerja	1
jateng	1	kementerian	1
jawa	1	komisi	1
berhasil	1	dikutip	1
tengah	1	kontan	1
ujar	1	melanjutkan	1
membentuk	1	kondisi	1
penilai	1	pelik	1
resmi	1	terbitnya	1
kepala	1	peraturan	1
tim	1	nomor	1
nilaian	1	melepas	1
dijadikan	1	wewenang	1
ojk	1	kkp	1
percepatan	1	mengendalikan	1
januari	1	industri	1
pilot	1	padahal	1
regional	1	sesuai	1
dipasarkan	1	uu/2016	1
pungkasnya	1	mengatur	1
proyek	1	rekomendasi	1
ditarget	1	ikan	1
daerah	1	berdasar	1
selesai	1	kali	1
penilaian	1	menikmati	1
menerapkan	1	dalam	1
akhir	1	pertemuan	1
obligasi	1		

Langkah selanjutnya menghitung nilai bobot dari TF dan IDF hasilnya seperti pada tabel 7

Tabel 7 Menghitung Bobot

Term	Bobot
infrastruktur	0.69314718055995
perusahaan	1.3862943611199
alternatif	0.69314718055995
jateng	4.1588830833597
jawa	2.0794415416798
berhasil	0.69314718055995
tengah	1.3862943611199
ujar	0.69314718055995
membentuk	0.69314718055995
penilai	0.69314718055995
resmi	0.69314718055995

kepala	0.69314718055995
tim	1.3862943611199
nilaian	0.69314718055995
dijadikan	0.69314718055995
ojk	1.3862943611199
percepatan	1.3862943611199
januari	1.3862943611199
pilot	1.3862943611199
regional	0.69314718055995
dipasarkan	1.3862943611199
pungkasnya	1.3862943611199
proyek	0.69314718055995
ditarget	1.3862943611199
daerah	2.7725887222398
selesai	0.69314718055995
penilaian	0.69314718055995
menerapkan	0.69314718055995
akhir	0.69314718055995
obligasi	1.3862943611199
istimewa	0.69314718055995
tahun	0.69314718055995
yogyakarta	0.69314718055995
2018	0.69314718055995
pemerintah	1.3862943611199
bambang	0.69314718055995
2019	0.69314718055995
pusat	0.69314718055995
kiswono	0.69314718055995
menetapkan	01.3862943611199
ngopi	00.69314718055995

4. KESIMPULAN

Penelitian ini telah menghasilkan sebuah perangkat lunak visualisasi kumpulan berita menggunakan TF-IDF sebagai pembobotan dan *gazetteer* sebagai acuan lokasi berita untuk memetakan lokasi berita. Hasil yang diperoleh melalui 20 naracoba, perangkat lunak ini menghasilkan ketertarikan secara visual kepada pembaca sebanyak 68%, kesesuaian data yang ditampilkan wordcloud 65,33% dan 52,66% membantu dalam memudahkan menyimpulkan berita dalam jumlah banyak dan melihat tren berita

REFERENCES

- [1] M. F. Ilham, A. L. Nugraha dan H. S. Firdaus, "Visualisasi Peta Fasilitas Penunjang Wisata Religi Kabupaten Demak Menggunakan Aplikasi Carrymap," *Jurnal Geodesi Undip*, vol. Volume 6 Nomor 4, 2017.
- [2] M. b. Wardana dan M. Jazman, "Rancang Bangun Sistem Informasi Geografis Pemetaan Ruang Ujian Menggunakan Bootstrap dan Leaflet.js," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri*, 2017.
- [3] R. A. Hendrawan, C. Aldino dan . E. Suryani, "Visualisasi Penjualan dan Lokasi Penjualan," Vol. %1 dari %2Volume 11, Nomor 1, 2013.
- [4] k. m. wibowo, i. kanedi dan j. jumadi, "Sistem Informasi Geografis (Sig) Menentukan Lokasi Pertambangan Batu Bara Di Provinsi Bengkulu Berbasis Website," *Jurnal Media Infotama*, vol. vol. 11 no. 1, 2015.
- [5] S. Rani, "Pemanfaatan Google Maps Api untuk Visualisasi Data Base Transceiver Station," *seminar nasional universitas islam indonesia*, Vol. %1 dari %2Vol 22, No 2, 2016.
- [6] H. A. Qashlim dan S. , "Sistem Informasi Geografis Kawasan Bencana Alam," *Jurnal Ilmiah Ilmu Komputer*, pp. Vol. 3, No. 2, 2017.
- [7] S. Nurdianti, B. Barus dan . D. Prasetyo, "Pengembangan Sistem Informasi Geografis Tindak Kejahatan," *Jurnal Ilmu Komputer*, Vol. %1 dari %2Vol 6, No 1, 2008.
- [8] S. Lohmann, J. Ziegler dan L. Tetzlaff, "Comparison of Tag Cloud Layouts: Task-related," *International Conference on Human-Computer Interaction*, pp. 392-404, 2009.
- [9] I. Adiyana dan F. Hakim, "Implementasi Text Mining pada Mesin Pencarian Twitter untuk Menganalisis Topik Topik terkait "KPK dan Jokowi,"" *Seminar Nasional universitas muhammadiyah*, 2015.
- [10] D. Apriandi, M. . I. Afandi dan E. D. Wahyuni, "Analisis Sentimen Pelanggan Wifi.Id Pada Twitter Dengan Support Vector Machine," *Jurnal Sistem Informasi Dan Bisnis Cerdas*, Vol. %1 dari %2Vol. 9, No. 1, 2016.
- [11] N. Hayatin, C. Fatichah dan D. Purwitasari, "Pembobotan kalimat Berdasarkan Fitur Berita dan Trending Issue untuk Peringkasan Multi Dokumen Berita," *Jurnal Ilmiah Teknologi Informasi*, Vol. %1 dari %2Vol 13, No 01, 2015.
- [12] A. A. Maarif, "Penerapan Algoritma TF-IDF untuk Pencarian Karya Ilmiah," *Journal of Electrical and Computer Engineering*, 2015.
- [13] A. Y. Munarko dan Y. Azhar, "Peringkasan Tweet Berdasarkan Trending Topic Twitter dengan Pembobotan TF-IDF dan Single Linkage Angglomerative Hierarchical Clustering," *KINETIK*, Vol. %1 dari %2Vol.1, No.1, 2016.