

# Sistem Temu Kembali Informasi Keadaan Lalu Lintas Persimpangan di Daerah Istimewa Yogyakarta

Muhammad Fahmi Zulfikar

Fakultas Matematika dan Ilmu Pengetahuan  
Alam

Universitas Gadjah Mada, Yogyakarta  
muhammadfahmizulfikar@mail.ugm.ac.id

Ryan Maulana

Fakultas Matematika dan Ilmu Pengetahuan  
Alam

Universitas Gadjah Mada, Yogyakarta  
ryanmaulana2001@mail.ugm.ac.id

## ABSTRACT

Pemanfaatan sistem *information retrieval* telah banyak digunakan pada berbagai bidang yang berkaitan dengan pencarian dan akuisisi data. Pada kesempatan kali ini kami mengimplementasikannya dalam bidang prasarana lalu lintas. Dimana pada prosesnya, kami mengumpulkan seluruh informasi mengenai kondisi lalu lintas pada jalan-jalan dan persimpangan di Daerah Istimewa Yogyakarta. Dengan memanfaatkan laporan hasil pemantauan ATCS DIY yang terdapat di twitter, kami mengumpulkan perkembangan kondisi lalu lintas di Daerah Istimewa Yogyakarta. Kemudian informasi tersebut diolah sehingga dapat menghasilkan sebuah data mengenai jalan manakah yang harus dilalui untuk menghindari kemacetan serta menampilkan jam-jam tertentu yang sering terjadi kemacetan.

## Author Keywords

Lalu lintas, Persimpangan, Machine Learning, Information Retrieval, Sistem Temu Kembali Informasi

## 1. PENDAHULUAN

Sistem temu balik informasi (*information retrieval system*) digunakan untuk menemukan kembali informasi-informasi yang relevan terhadap kebutuhan pengguna dari satu himpunan informasi secara otomatis. Salah satu aplikasi umum dari sistem temu balik informasi adalah *search-engine* atau mesin pencarian yang terdapat pada jaringan internet. Menggunakan mesin pencarian ini pengguna dapat mencari halaman Web yang dibutuhkan melalui mesin tersebut secara cepat. Dalam *Information Retrieval*, untuk mendapatkan dokumen yang relevan tidaklah cukup. Tujuan yang harus dipenuhi adalah bagaimana mendapatkan dokumen yang relevan dan tidak mendapatkan dokumen yang tidak relevan<sup>1</sup>.

<sup>1</sup>[https://id.wikipedia.org/wiki/Sistem\\_temu\\_balik\\_informasi](https://id.wikipedia.org/wiki/Sistem_temu_balik_informasi)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHI '20, April 25–30, 2020, Honolulu, HI, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-6708-0/20/04...\$15.00

DOI: <https://doi.org/10.1145/3313831.XXXXXX>

## 2. LATAR BELAKANG

Di Daerah Istimewa Yogyakarta setidaknya terdapat 34 titik rawan macet dan 4 diantaranya rawan terjadi kecelakaan. Laporan dari ATCS menunjukkan bahwa titik kemacetan terparah terjadi pada jalan-jalan utama. Kemacetan terparah biasanya terjadi hanya pada saat jam-jam sibuk atau saat jam pulang kantor, namun kondisi jalan yang kurang memadai dapat memperburuk keadaan. Pada saat musim penghujan terdapat beberapa jalan yang tergenang oleh air, hal tersebut dapat mengurangi kenyamanan dalam berkendara.

### 2.1 Peningkatan Volume Kendaraan

Saat ini semakin banyak masyarakat yang memiliki kendaraan pribadi dan kurangnya kesadaran masyarakat akan pentingnya moda transportasi umum. Hal ini mengakibatkan bertambahnya volume kendaraan di jalan raya. Akibatnya terjadi penghambatan di beberapa titik ruas jalan raya ditambah dengan kondisi lalu lintas dan kondisi jalan yang tidak menentu di Daerah Istimewa Yogyakarta

## 3. CRAWLING

Sebelum kita lanjut ke pembahasan mengenai crawling, pertama kita harus menetapkan darimana informasi akan diambil, bagaimana data akan dikumpulkan, dan bagaimana caranya data ditampilkan sesuai dengan permintaan pengguna?

### 3.1 Platform Media Sosial

Pada proyek Sistem Temu Kembali Informasi ini kami berfokus pada pengambilan informasi dari media sosial. Media sosial adalah Media sosial atau yang dikenal juga dengan jejaring sosial merupakan bagian dari media baru. Jelas karena bahwa muatan interaktif dalam media baru sangatlah tinggi. Media sosial, dikutip dari Wikipedia, didefinisikan sebagai sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia.<sup>2</sup>

B.K. Lewis dalam karyanya yang berjudul *Social Media and Strategic Communication Attitudes and Perceptions among College Students* yang terbit pada tahun 2010 menyatakan, bahwa media sosial merupakan suatu label yang merujuk pada

<sup>2</sup>[http://id.wikipedia.org/wiki/Media\\_sosial](http://id.wikipedia.org/wiki/Media_sosial)

teknologi digital yang berpotensi membuat semua orang untuk saling terhubung dan melakukan interaksi, produksi dan berbagi pesan. [6]. Selanjutnya, pada tahun 2010, Chris Brogan dalam bukunya yang berjudul *Social Media 101: Tactics and Tips to Develop Your Business*, menyebutkan bahwa media sosial adalah suatu perangkat alat komunikasi yang memuat berbagai kemungkinan untuk terciptanya bentuk interaksi gaya baru. [1]

### 3.2 Media Sosial Twitter

Twitter adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan (tweet). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan "pesan singkat dari Internet."<sup>3</sup>

### 3.3 Metode Eksplorasi

Sosial media adalah website berisi informasi yang informasinya bertambah setiap harinya dengan jumlah yang tidak sedikit. Pada proyek ini pengambilan informasi akan dilakukan dengan cara focused crawling.

3.3.1 Focused Crawling adalah program/perangkat lunak atau skrip terprogram, yang secara otomatis mengumpulkan halaman web yang relevan dengan topik tertentu. Biasanya berisi empat modul utama, yaitu *Seed Uniform Resource Locators* (URL), topik yang diberikan, perhitungan relevansi dan penetapan prioritas URL. [5]

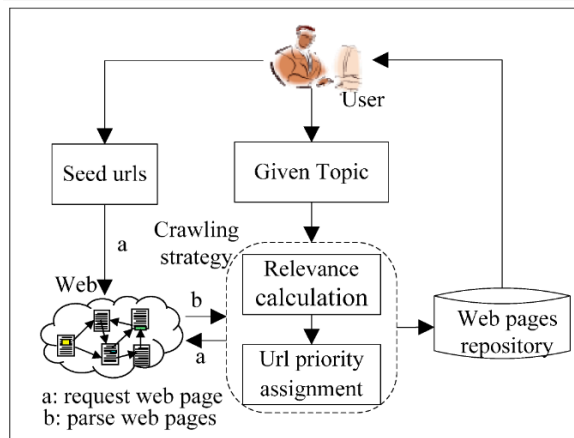


Figure 1. Proses Focused Crawler pada Web Crawling

3.3.2. Cara kerja focused crawler terbagi menjadi empat bagian. Pertama pengguna memasukkan topik dan target pada crawler, pada proyek ini topik yang dicari adalah kondisi lalu lintas dengan target berasal dari akun twitter ATCS Dinas Perhubungan DIY, kedua dilakukan penguraian isi dari tweet tersebut dan dihitung nilai relevansi tweet dengan

topik yang diinput. Ketiga mengulang operasi di atas dan mengumpulkannya menjadi satu dokumen. Terakhir, untuk meningkatkan efisiensi seluruh proses perayapan, beberapa algoritme peringkat diterapkan untuk menetapkan prioritas sub-tautan, yang dapat mengontrol daftar permintaan. Konsep pada Figure 1 adalah focused crawler pada Web Crawling, yang mana tidak jauh berbeda cara kerjanya dengan focused crawler media sosial.

3.3.3. Pada proyek ini digunakan Twint sebagai crawler untuk mengumpulkan tweet dari akun ATCS Dinas Perhubungan DIY. Cara Kerja Twint adalah dengan melakukan pencarian topik pada search bar Twitter serta menggunakan fitur operator pencarian Twitter untuk membantu keakuratan pencarian. Setelah twint menemukan tweet yang relevan dan sesuai dengan topik, Twint akan mengumpulkan Tweet kedalam sebuah dokumen excel atau csv.

## 4. PENGAMBILAN INFORMASI

Headings of subsections should be in Helvetica or Arial 9-point bold with initial letters capitalized. For sub-sections and sub-subsections, a word like *the* or *of* is not capitalized unless it is the first word of the heading.

### 4.1 Pembersihan Data

Proses selanjutnya adalah proses pembersihan data menggunakan Python dengan cara menyaring isi dokumen untuk hanya menampilkan data dengan kata-kata yang berhubungan dengan kondisi lalu lintas.

```
key_value = ["jalan",
             "lalin",
             "lalu lintas",
             "lancar",
             "macet",
             "padat",
             "ramai",
             "tersendat"]
```

Figure 2. Kata Kunci yang Berhubungan Dengan Keadaan Lalu Lintas

Selanjutnya melakukan penyaringan lebih lanjut dengan memberikan variabel pendukung yang berisi kondisi, peristiwa, dan cuaca.

Kondisi	Peristiwa	Cuaca
lancar	kecelakaan	berawan
ramai lancar	tabrakan	hujan
padat	banjir	gerimis
padat merayap	demonstrasi	cerah
tersendat	mogok	
macet	genangan	
	pohon tumbang	

Figure 3. Data Non Jalan (Pendukung Keywords)

<sup>3</sup><http://id.wikipedia.org/wiki/Twitter>

Selain dilakukan pembersihan pada tweet, sekarang dilakukan pembersihan zona waktu (timezone) dengan melakukan konversi dari UTC +00:00 ke GMT +07:00.

Setelah melakukan konversi waktu maka lanjut ke proses terakhir yaitu proses splitting. Proses splitting dilakukan untuk mendapatkan inti dari informasi secara cepat. Ketika seseorang melakukan sebuah pencarian maka hasil langsung menunjukkan informasi secara detail tanpa kata-kata penghubung atau kata-kata yang tidak terlalu penting.

	Date	Year	Month	Day	Waktu	Dari
0	2022-Oct-03	2022	October	Monday	08:10:58	MM UGM
1	2022-Oct-03	2022	October	Monday	08:10:35	Demangan
2	2022-Oct-03	2022	October	Monday	08:10:11	UIN
3	2022-Oct-03	2022	October	Monday	08:05:19	Kentungan
4	2022-Oct-03	2022	October	Monday	08:04:45	Condongcatur
..	...	...	...	...	...	...
77	2022-Sep-24	2022	September	Saturday	17:16:13	Galeria
78	2022-Sep-24	2022	September	Saturday	15:15:34	Bandara
79	2022-Sep-24	2022	September	Saturday	15:14:57	Gramedia
80	2022-Sep-24	2022	September	Saturday	07:26:14	Gramedia
81	2022-Sep-24	2022	September	Saturday	07:25:52	MM UGM

Figure 4. Hasil Splitting

## 5. INDEXING

Indexing adalah daftar terurut dari artikel yang dikutip, masing-masing disertai dengan daftar artikel yang mengutip. Artikel yang mengutip diidentifikasi sebagai sumber dan artikel yang dikutip sebagai referensi. Dengan indexing isi jurnal dapat dicari dengan menggunakan judul subjek (kata kunci, nama penulis, judul, abstrak, dll.) di database yang tersedia. [2]

### 5.1 Algoritma Indexing

Terdiri beberapa algoritma untuk membuat index, namun single pass in memory indexing lebih efisien karena Bagian dari algoritme yang mem-parsing dokumen dan mengubahnya menjadi aliran pasangan term-docID, yang biasa disebut token telah dihilangkan. SPIMI-INVERT dipanggil berulang kali pada aliran token hingga seluruh koleksi diproses.

Alih-alih mengumpulkan semua pasangan termID-docID terlebih dahulu dan kemudian menyortirnya (seperti yang kami lakukan di BSBI), setiap daftar posting bersifat dinamis (yaitu, ukurannya disesuaikan saat bertambah) dan segera tersedia untuk mengumpulkan posting<sup>4</sup>.

### 5.2 Inverted Index

Inverted index adalah indeks basis data yang menyimpan pemetaan dari konten, seperti kata atau angka ke lokasinya dalam tabel atau dokumen atau kumpulan dokumen. Tujuan dari inverted index adalah untuk memungkinkan pencarian teks lengkap yang cepat, dengan biaya pemrosesan yang meningkat ketika dokumen ditambahkan ke database. Pengindeksan ini adalah metode pengstruktur data paling populer yang digunakan dalam sistem pengambilan dokumen (information retrieval) [8].

Indexing dilakukan dengan pertama meletakkan semua file dokumen hasil crawling ke dalam satu folder dan menyatukan

<sup>4</sup><https://nlp.stanford.edu/IR-book/html/htmledition/single-pass-in-memory-indexing-1.html>

dokumen tersebut menjadi satu dokumen besar. Karena pencarian utamanya adalah nama persimpangan, maka cukup dilakukan pembuatan index dengan kata-kata nama persimpangan dan arahnya.

Simpang	Arah
Tempel	Timur
Kidsfun	Barat
Giwangan	Selatan
Gamping	Utara
MM UGM	
Demangan	

Figure 5. Nama-nama persimpangan dan arahnya

Dari sini akan menghasilkan posisi-posisi index dari nama-nama persimpangan tersebut. Untuk memudahkan pencarian oleh manusia, maka dibentuk juga indexing dengan menampilkan pada dokumen apa jalan tersebut muncul.

Pada indexing ini tidak dilakukan kompresi index karena kamus index disini tidak berasal dari token kata per kata dari tweet dalam dokumen besar yang diindex, tetapi menggunakan kamus dari luar yang isinya hanya berfokus pada nama persimpangan [4, 7].

```
{
  'Maguwoharjo': [3, 18, 29, 31, 40, 67],
  'Bandara': [4, 19, 25, 30, 32, 41, 47, 57, 68],
  'Gramedia': [5, 8, 10, 13, 36, 43, 58, 59, 63, 69],
  'Galeria': [7, 56],
  'Pogung': [9, 54, 64],
  'Seturan': [11, 42, 49, 62, 70],
  'Jati Kencana': [15, 33, 35],
  'MM UGM': [16, 20, 38, 51, 55, 60, 65],
  'Condongcatur': [17, 24, 37, 61],
  'Kentungan': [23],
  'Tugu': [26, 45, 52, 66],
  'Kidsfun': [34]}
{'Seturan': ['2022-Sep-30', '2022-Sep-28', '2022-Sep-29', '2022-Sep-26', '2022-Oct-01']}
```

Figure 6. Hasil indexing

## 6. PENCARIAN INFORMASI

Dengan menggabungkan metode pencarian, pembersihan, dan pengindeksan, maka didapat ribuan data tweet. Meskipun sudah diurutkan berdasarkan waktu tentu masih sangat sulit untuk mencari suatu data spesifik. Lalu bagaimana cara mencari data yang spesifik tersebut?

Pertama dibuat dulu sebuah testing pencarian agar dapat dipastikan bahwa data sudah bisa dicari. Untuk testing ini masih dilakukan dengan mencari kata yang masih case sensitive.

Coba masukkan nama jalan

Coba masukkan nama jalan

Setelah sudah bisa dilakukan sebuah pencarian, maka dilanjutkan dengan proses pembuatan antarmuka (interface) agar memudahkan pencarian secara visual.

```

Yay! Saya menemukan beberapa Tweet!

Date: 2022-Oct-01
Year: 2022
Month: October
Day: Saturday
Time: 07:31:25
Tweet: 07:31 wib lalin sp seturan arah timur dan barat ramai lancar arah selatan terpantau lancar
From: Seturan
To: Timur, Barat
Condition: ramai lancar
Cause: -

Date: 2022-Sep-30
Year: 2022
Month: September
Day: Friday
Time: 09:08:25
Tweet: 09:07 wib lalin sp seturan arah timur dan barat ramai lancar arah selatan terpantau lancar
From: Seturan
To: Timur, Barat
Condition: ramai lancar
Cause: -

```

Figure 7. Pencarian Dengan Terminal

## 6.1 Antarmuka

Menurut Wilbert O. Galitz, user interface atau antarmuka adalah bagian dari komputer dan perangkat lunak yang dapat dilihat, didengar, disentuh, diajak bicara, dan yang dapat dimengerti secara langsung oleh manusia [3].

Desain antarmuka pada proyek ini adalah berbasis web dengan bahasa pemrograman html dan php. Antarmuka disini cukup sederhana yaitu hanya terdiri dari judul, kotak pencarian, dan tabel hasil pencarian. Pengguna hanya cukup memasukkan nama persimpangan pada kotak pencarian dan hasilnya akan ditunjukkan di dalam tabel dibawahnya. Informasi yang ditunjukkan yaitu nomor, tanggal, waktu, dari mana asalnya, ke arah mana, kondisi jalan, hambatan jika terjadi kemacetan yang disebabkan oleh kecelakaan, demonstrasi, dan sebagainya, serta link tweet yang bisa diakses dengan mengeklik tweet dalam tabel.

Informasi Kondisi Lalu Lintas Jalan D.I. Yogyakarta via tweet ATCS D.I.Y							
<p>Hi! Saya akan mengambil informasi lalu lintas dari tweet @ATCS_DISHUB_DIY. Coba tulis nama jalannya disini ya!</p> <p>Masukkan Nama Persimpangan <input type="text"/></p> <p>Informasi jalan akan muncul dalam tabel di bawah ini</p>							
No	Tanggal	Waktu	Dari	Arah	Kondisi	Hambatan	Tweet (Klik untuk melihat Tweet)
<p>Informasi Kondisi Lalu Lintas Jalan D.I. Yogyakarta via tweet ATCS D.I.Y</p> <p>Hi! Saya akan mengambil informasi lalu lintas dari tweet @ATCS_DISHUB_DIY. Coba tulis nama jalannya disini ya!</p> <p>Masukkan Nama Persimpangan <input type="text"/></p> <p>Yay! Saya menemukan beberapa tweet!</p>							
1	2022-Sep-15	12:55:46	Maguwoharjo	Barat	lancar	-	12.56 wib kondisi lalin sp maguwo arah barat terpantau lancar
2	2022-Sep-13	13:27:23	Maguwoharjo	Barat	ramai lancar	-	13.27 wib kondisi lalin sp maguwo terpantau ramai lancar dari arah barat
3	2022-Sep-12	20:17:08	Maguwoharjo	Barat	lancar	-	20.17 wib kondisi lalin sp maguwo arah barat terpantau lancar
4	2022-Sep-12	16:06:59	Maguwoharjo	Barat	ramai lancar	-	16.08 wib lalin sp maguwo arah barat terpantau ramai lancar
5	2022-Sep-12	14:52:56	Maguwoharjo	Barat	lancar	-	14.52 wib kondisi lalin sp maguwo arah barat terpantau lancar
6	2022-Sep-10	11:06:02	Maguwoharjo	Barat	ramai lancar	-	12.26 wib kondisi lalin sp maguwo arah barat terpantau ramai lancar
7	2022-Sep-07	12:35:02	Maguwoharjo	Barat	lancar	-	12.35 wib lalin sp maguwo arah barat terpantau lancar
8	2022-Sep-07	10:02:21	Maguwoharjo	Barat	lancar	-	10.02 wib lalin sp maguwo arah barat terpantau lancar
9	2022-Oct-03	08:02:54	Maguwoharjo		padet	-	08.02 WIB Arus lalu lintas simpang Maguwo padat dari semua arah
10	2022-Oct-01	08:22:27	Maguwoharjo	Barat	lancar	-	08.22 wib kondisi lalin sp maguwo arah barat terpantau lancar
11	2022-Sep-30	07:09:35	Maguwoharjo	Barat	lancar	-	07.09 wib lalin sp maguwo arah barat terpantau lancar
12	2022-Sep-28	08:33:20	Maguwoharjo	Barat	lancar	-	08.33 wib kondisi lalin sp maguwo arah barat terpantau lancar
13	2022-Sep-27	18:44:07	Maguwoharjo	Barat	ramai lancar	-	18.44 wib lalin sp maguwo arah barat terpantau ramai lancar

Figure 8. Antarmuka (interface) mesin pencarian

## 7. KESIMPULAN

Pada paper ini telah dijabarkan mengenai Sistem Temu Kembali Informasi yang informasinya berasal dari media sosial dengan cara crawling, indexing, dan pengambilan informasi. Kami menyatukan metode-metode tersebut menjadi sebuah program mesin pencarian berbasis tulisan dengan fungsi mencari data situasi persimpangan jalan raya di Daerah Istimewa Yogyakarta.

## REFERENCES

- [1] C. Brogan. 2010. *Social Media 101: Tactics and Tips to Develop Your Business Online*. Wiley.  
<https://books.google.co.id/books?id=in7BSJlU7z4C>
- [2] Haq R. U. Dhammi, I. K. 2016. What is indexing. (2016). <https://doi.org/10.4103/0019-5413.177579>
- [3] W.O. Galitz. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. Wiley.  
[https://books.google.co.id/books?id=Q3Xp\\_Awu49sC](https://books.google.co.id/books?id=Q3Xp_Awu49sC)
- [4] Raidah Hanifah, Suhono Supangkat, and Ayu Purwarianti. 2014. Twitter Information Extraction for Smart City. 295–299. DOI:  
<http://dx.doi.org/10.1109/ICTSS.2014.7013190>
- [5] Dongyang Hou, Hao Wu, Jun Chen, and Ran Li. 2014. A Focused Crawler for Borderlands Situation Information with Geographical Properties of Place Names. *Sustainability* 6, 10 (2014), 6529–6552. DOI:  
<http://dx.doi.org/10.3390/su6106529>
- [6] Bobbi Kay Lewis. 2010. Social Media and Strategic Communication: Attitudes and Perceptions Among College Students. *Public Relations Journal* 4 (2010).
- [7] Jaka Eka Sembodo, Erwin Setiawan, and Abdurahman Baizal. 2016. Data Crawling Otomatis pada Twitter. 11–16. DOI:  
<http://dx.doi.org/10.21108/INDOSC.2016.111>
- [8] Justin Zobel, Alistair Moffat, and Kotagiri Ramamohanarao. 1998. Inverted files versus signature files for text indexing. *ACM Trans. Database Syst.* 23 (1998), 453–490.