

Penerapan Graf dalam Algoritma *PageRank*

Mesin Pencari *Google*

Adya Naufal Fikri - 13515130
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13515130@std.stei.itb.ac.id

Abstract—Matematika merupakan salah satu cabang ilmu pengetahuan yang banyak digunakan untuk memodelkan suatu permasalahan. Bukan hanya itu, matematika juga berperan penting untuk menyelesaikan permasalahan tersebut. Salah satunya adalah masalah pencarian informasi di internet. Dengan menggunakan suatu teori matematika, yaitu teori Graf, mesin pencari *Google* dengan algoritma *PageRank*-nya dapat menghasilkan pencarian yang relevan dengan apa yang dimaksud pengguna.

Keywords—*PageRank, Graph Theory, Google, Web.*

I. PENDAHULUAN

Internet merupakan salah satu sarana untuk menunjang pencarian informasi. Dengan banyaknya sumber informasi yang ada di internet, itu memudahkan kita ketika hendak mencari informasi tentang suatu hal. Namun, pada awal mulanya internet ada, para pengguna kebingungan untuk mencari informasi yang mereka butuhkan. Karena mereka tidak tahu harus mulai mencari informasi yang mereka butuhkan dari mana. Karena pada saat itu mesin pencari belum ada, sehingga ketika mereka mencari di internet, maka mereka harus memasukan alamat *Web* dari situs *Web* yang mereka tuju, dan informasi di suatu situs *Web* itu terbatas.

Mulailah dikembangkan yang namanya search engine (mesin pencari). Pada awalnya, search engine dikembangkan di McGill University (Montreal) untuk mencari dan mendapatkan file di sebuah jaringan komputer. Lalu untuk pencarian di internet masih menggunakan *Web* direktori yang diurus secara manual. Dengan perkembangan situs *Web* yang lebih dari 100% saat itu, maka dibutuhkan banyak sekali sumber daya manusia untuk mengurusnya *Web* direktori itu, dan hal itu tidaklah praktis. Dan penggunaan *Web* direktori untuk mencari suatu informasi saat itu, terkadang tidak memberikan informasi yang relevan dengan apa yang dimaksudkan oleh pengguna.

Baru pada tahun 1998, muncullah *Google*, sebuah mesin pencari karya dari Sergey Brin dan Larry Page. *Google* saat itu bahkan hingga saat ini menjadi primadona, karena mengubah cara mencari di internet menjadi lebih

praktis. Dengan menggunakan algoritma *PageRank* yang memanfaatkan teori matematika, yaitu teori Graf, *Google* berhasil mengambil hati para pengguna internet dikarenakan hasil pencariannya yang relevan dengan apa yang dimaksudkan oleh pengguna dan juga hasil pencarian banyak yang menuju ke sumber informasi yang terpercaya.

Dalam makalah ini, penulis akan mencoba menjelaskan bagaimana *Google* menerapkan teori Graf dalam algoritma *PageRank*-nya sehingga mereka bisa mendapatkan hasil pencarian yang relevan dengan apa yang dimaksud oleh pengguna.

II. LANDASAN TEORI

A. Definisi Graf

Graf merupakan suatu struktur yang terdiri dari objek-objek diskrit yang direpresentasikan secara visual dengan titik atau noktah yang disebut simpul, dan hubungan antara objek-objek tersebut yang direpresentasikan dengan garis yang disebut sisi[1]. Graf didefinisikan sebagai pasangan himpunan (V, E) , yang dalam hal ini :

$$V = \text{himpunan tidak-kosong dari simpul-simpul} = \{v_1, v_2, \dots, v_n\}$$

dan

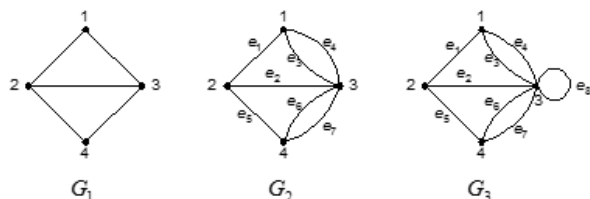
$$E = \text{himpunan sisi yang menghubungkan sepasang simpul} = \{e_1, e_2, \dots, e_n\}$$

atau bisa ditulis singkat notasi $G = (V, E)$. Definisi tersebut menyatakan bahwa V tidak boleh kosong, sedangkan E boleh kosong. Jadi, sebuah graf dimungkinkan tidak mempunyai sisi satu pun, tetapi simpulnya harus ada, minimal satu[2].

Simpul pada graf dapat dinomori dengan huruf, seperti $a, b, c, \dots, v, w, \dots$, dengan bilangan asli $1, 2, 3, \dots$, atau gabungan keduanya. Sedangkan sisi yang menghubungkan simpul v_i , dengan simpul v_j dinyatakan dengan pasangan (v_i, v_j) atau dengan lambang e_1, e_2, \dots . Dengan kata lain, jika e adalah sisi yang menghubungkan sisi v_i dengan simpul v_j , maka e dapat ditulis sebagai

$$e = (v_i, v_j).$$

Secara geometri graf dapat digambarkan sebagai sekumpulan noktah (simpul) di bidang dua dimensi yang dihubungkan dengan sekumpulan garis (sisi).



Gambar 1. Tiga buah graf
(sumber : [2])

Pada G_2 , sisi $e_3 = (1,3)$ dan $e_4 = (1,3)$ dinamakan sisi-ganda karena kedua sisi ini menghubungkan dua buah simpul yang sama, yaitu simpul 1 dan simpul 3, sehingga G_2 disebut graf ganda karena mempunyai sisi ganda. Pada G_3 , sisi $e_3 = (3,3)$ dinamakan gelang atau kalang (*loop*) karena berawal dan berakhir pada simpul yang sama. Graf G_3 disebut graf semu dikarenakan mempunyai sisi gelang[2].

B. Jenis-jenis Graf

Berdasarkan ada tidaknya gelang atau sisi-ganda pada suatu graf, maka secara umum graf dapat digolongkan menjadi dua jenis, graf sederhana (*simple graph*) dan graf tidak-sederhana (*unsimple-graph*). Graf sederhana merupakan graf yang tidak mempunyai sisi-ganda dan sisi gelang, sedangkan graf tidak-sederhana mempunyai sisi gelang atau sisi-ganda atau keduanya.

Sisi pada sebuah graf dapat mempunyai orientasi arah. Berdasarkan orientasi arahnya, graf dibedakan menjadi dua, yaitu graf berarah (*directed graph*) dan graf tak-berarah (*undirected graph*). Graf berarah merupakan graf yang setiap sisinya mempunyai orientasi arah, yang biasa disebut busur. Untuk suatu busur (sisi berarah) yang menghubungkan simpul v_j ke simpul v_k , v_j disebut simpul asal, sedangkan v_k disebut simpul terminal. Pada graf berarah sisi gelang diperbolehkan, sedangkan sisi-ganda tidak diperbolehkan. Sementara itu, graf tak-berarah adalah graf yang sisinya tidak mempunyai orientasi arah.

C. Terminologi Dasar Graf

Di bawah ini merupakan beberapa terminologi yang akan sering digunakan ketika membahas masalah graf :

1. Bertetangga (*Adjacent*)
Dua buah simpul pada graf tak-berarah dikatakan bertetangga, jika keduanya terhubung langsung melalui suatu sisi.
2. Bersisian (*Incident*)
Untuk sembarang sisi $e = (v,u)$, sisi e dikatakan bersisian dengan simpul v dan simpul u .
3. Graf Kosong (*Null Graph*)
Graf yang himpunan sisinya kosong, namun himpunan simpulnya tidak.

4. Derajat (*Degree*)

Pada graf tak-berarah, derajat suatu simpul merupakan banyak sisi yang bersisian dengan simpul tersebut. Sedangkan pada graf berarah, derajatnya terbagi dua, yaitu derajat masuk, dan derajat keluar. Dengan derajat masuk adalah banyak busur yang masuk ke simpul tersebut, sedangkan derajat keluar adalah banyak busur yang keluar dari simpul tersebut.

5. Lintasan (*Path*)

Lintasan yang panjangnya n dari simpul awal v_0 ke simpul tujuan v_n di dalam graf G adalah barisan berselang-seling dari simpul-simpul dan sisi-sisi berbentuk $v_0, e_1, v_1, e_2, \dots, e_n, v_n$ sehingga banyaknya sisi yang terlewati adalah n , dan e_1, e_2, \dots, e_n merupakan sisi-sisi dari graf G .

6. Terhubung (*Connected*)

Graf tak-berarah G disebut graf terhubung jika untuk setiap pasang v_i dan v_j yang merupakan simpul di G , terdapat lintasan yang menghubungkan keduanya. Jika ada yang tidak terhubung maka graf tersebut disebut graf tak-terhubung.

7. Graf Berbobot (*Weighted Graph*)

Graf berbobot adalah graf yang setiap sisinya mempunyai sebuah nilai (bobot).

D. Internet

Internet (kependekan dari *interconnection-networking*) adalah seluruh jaringan komputer yang saling terhubung menggunakan standar protokol *TCP/IP* sebagai protokol pertukaran paket data untuk melayani miliaran pengguna di seluruh dunia. Internet merupakan jaringan yang terdiri dari jaringan pribadi, jaringan publik, jaringan akademik, jaringan bisnis, dan jaringan pemerintah dari lingkup lokal hingga lingkup global.

Internet memiliki berbagai macam sumber informasi dan layanan, seperti dokumen dan aplikasi dalam situs *Web* (aplikasi *Web*), *e-mail*, telepon (*VoIP*) dan jaringan *peer-to-peer* untuk berbagi file. Tetapi layanan internet bisa dikelompokkan menjadi tiga, yaitu *World Wide Web*, komunikasi, dan *data transfer*.

Yang pertama, yaitu *World Wide Web*. Banyak orang menggunakan istilah Internet dan *World Wide Web* (disingkat *Web*), secara bergantian dan mereka menganggapnya sama padahal sebenarnya tidak. *World Wide Web* adalah sebuah aplikasi (layanan) utama yang ada di dalam internet dan setiap hari digunakan miliaran orang. Di dalam *Web* terdapat banyak dokumen, gambar, dan sebagainya yang terkoneksi melalui *hyperlink*. *Hypertext Transfer Protocol (HTTP)* merupakan protokol akses utama ke *Web*. Layanan *Web* juga menggunakan *HTTP* untuk memungkinkan sistem perangkat lunak berkomunikasi untuk berbagi dan bertukar informasi.

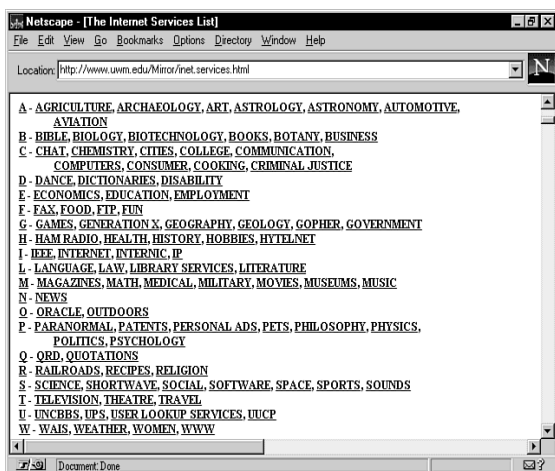
Yang kedua, komunikasi. Internet memungkinkan kita untuk berkomunikasi dengan satu bahkan banyak orang. Cara-cara untuk berkomunikasi tersebut bisa

melalui *VoIP* (telepon internet), *chatting*, *e-mail*, dan lain sebagainya. Yang ketiga, *data transfer*. Berbagi file adalah suatu contoh dari mentransfer data berukuran besar melalui internet. Sebuah file komputer dapat dikirim kepada orang lain melalui *e-mail* sebagai lampiran. Atau bisa juga diunggah ke sebuah situs *Web* atau ke *FTP server* untuk memudahkan orang lain mengunduh file tersebut.

E. Search Engine

Mesin pencari *Web* adalah sebuah sistem perangkat lunak yang dirancang untuk mencari informasi dalam internet (*World Wide Web*). Hasil dari pencarian tersebut umumnya direpresentasikan dalam beberapa baris yang merujuk pada informasi yang relevan dengan kata kunci tersebut. Informasi itu dapat berupa halaman *Web*, gambar, dokumen, dan lain sebagainya. Beberapa mesin pencari juga mengambil data yang tersedia dari *database* atau direktori terbuka. Tidak seperti *Web* direktori yang dijalankan oleh seorang editor, mesin pencari dapat mencari informasi secara langsung dengan menjalankan algoritma pada *web crawler* (pencari informasi di internet).

Pada awalnya mencari sebuah informasi di internet merupakan suatu hal yang sulit bagi pengguna, karena mereka tidak tahu harus mulai mencari informasi dari mana. Hanya ada direktori dari beberapa topik yang diurus oleh *enthusiaist* (ahli), sehingga tidak semua informasi yang ada di internet dapat dicapai dengan mudah waktu itu. Terlebih lagi direktori itu harus diperbaharui secara manual oleh *enthusiaist* tersebut, sedangkan perkembangan internet waktu itu sangat signifikan. Jumlah situs *Web* meningkat lebih dari 100% selama beberapa tahun pertama adanya *World Wide Web*.

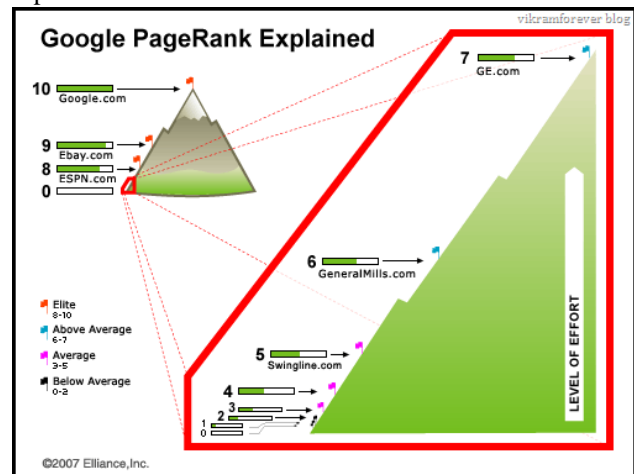


Gambar 2. Salah satu *Web* direktori pada tahun 1995 (<http://www.let.leidenuniv.nl/history/ivh/chap4.htm>)

III. PENERAPAN GRAF DALAM ALGORITMA PAGERANK GOOGLE

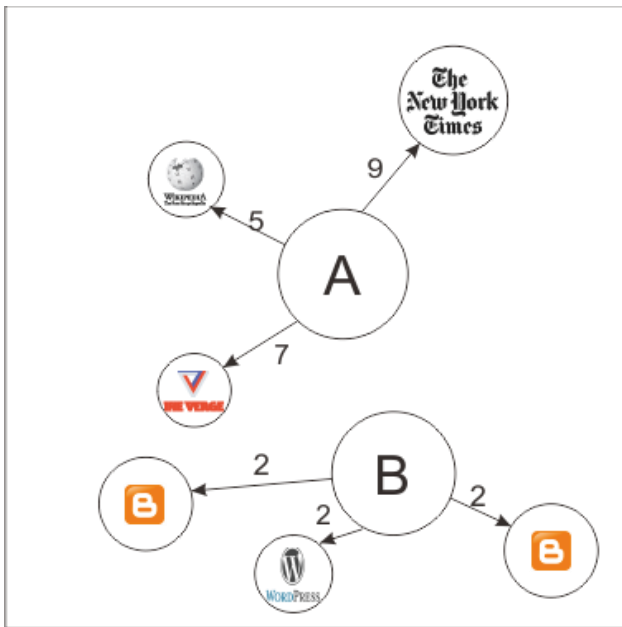
Pada tahun 1998, Sergey Brin dan Larry Page dua orang kandidat Ph.D di Departemen *Computer Science Stanford University*, mengubah bagaimana cara orang untuk berinternet di dunia maya. Mereka membuat suatu sistem perangkat lunak di dalam *Web* (situs *Web*) berupa mesin pencari *Web* yang kelak akan menjadi sebuah alamat *Web* yang wajib dikunjungi ketika hendak mencari informasi. *Google*, begitu karya mereka berdua disebut, merupakan salah satu situs *Web* yang sangat populer dan digunakan di santero dunia. *Google* juga menjadi situs *Web* yang paling banyak dikunjungi di seluruh dunia. Hingga akhirnya mengantarkan *Google* menjadi salah satu perusahaan paling sukses di dunia.

Mesin pencari *Google* didasari oleh sebuah algoritma yang disebut *PageRank*. Algoritma tersebut disusun langsung oleh Larry Page dan Sergey Brin pada tahun 1998. *PageRank* adalah sebuah algoritma untuk mengoptimalisasi hasil pencarian supaya lebih relevan dengan menggunakan graf berarah. Algoritma tersebut secara lebih detail dapat dilihat dalam paper yang dipublikasikan oleh mereka.



Gambar 3. Penjelasan *PageRank* (<http://vikramforever.blogspot.co.id/2011/04/google-pagerank-graph-pagerank.html>)

Graf *PageRank* atau selanjutnya disebut graf *PR*, dihasilkan dengan membuat semua halaman *Web* yang ada di *World Wide Web* sebagai simpul dan setiap link yang ada pada halaman tersebut sebagai sisi yang menghubungkan satu halaman *Web* ke halaman *Web* yang lainnya. Sisi tersebut kemudian dikategorikan menjadi sisi kuat atau lemah dengan memberi bobot untuk setiap sisi. Halaman *Web* yang terhubung dengan banyak sumber yang terpercaya seperti situs *CNN* atau *USA.gov* akan mempunyai bobot sisi yang lebih besar nilainya. Sehingga ketika kita membandingkan dua buah halaman *Web* atau situs *Web* yang memiliki jumlah sisi yang sama, maka *PageRank* akan memberikan situs dengan banyak link ke sumber terpercaya ranking yang lebih atas.



Gambar 4. Contoh graf berbobot
(Sumber Logo : New York Times, The Verge, Wikipedia, Blogspot, Wordpress)

Untuk mengindeks halaman Web (situs Web), mesin pencari Google secara sistematis akan berselancar di internet dimulai dari situs yang diketahui. Mesin pencari Google selain menggunakan algoritma PageRank, juga menggunakan sebuah program yang disebut *web spiders* (atau *crawler* atau *bot*) untuk mengindeks situs Web. *Web spiders* digunakan untuk mengunjungi situs Web dan menganalisa konten dari situs Web tersebut.

Web spiders menggunakan *depth-first searching* dan *breadth-first searching* untuk membuat indeks. Seperti disebut sebelumnya, halaman Web dan hubungan antara mereka dapat dimodelkan oleh graf berarah yang disebut graf Web. Halaman Web direpresentasikan oleh simpul dan link direpresentasikan oleh sisi berarah. Menggunakan *depth-first searching*, halaman Web pertama dipilih, lalu link diikuti ke halaman Web berikutnya (jika ada link tersebut), link pada halaman Web yang kedua diikuti ke halaman Web ketiga, jika ada link baru tersebut, dan seterusnya, sampai halaman tanpa link baru ditemukan. *Backtracking* kemudian digunakan untuk memeriksa link di tingkat sebelumnya untuk mencari link baru, dan seterusnya. (Karena keterbatasan, web spiders memiliki batas kedalaman dalam *depth-first searching* yang mereka lakukan.) Menggunakan *breadth-first searching*, halaman Web awal dipilih dan link pada halaman Web ini diikuti ke halaman Web yang kedua, maka link kedua di halaman Web awal diikuti (jika ada), dan seterusnya, sampai semua link dari halaman Web awal telah diikuti. Kemudian link pada halaman satu tingkat ke bawah diikuti, halaman demi halaman, dan seterusnya.

Akhirnya, ketika seseorang mencari suatu hal dengan memasukan sebuah *query* ke mesin pencari Google, mesin pencari Google akan mengurai masukan dari pengguna lalu akan mengupayakan hasil pencarian yang cocok dengan yang dimaksudkan pengguna. Kemudian hasil pencarian

itu akan disusun berdasarkan ranking yang didapat dari algoritma PageRank dengan situs yang mendapat ranking terbaik muncul paling atas dibandingkan yang lainnya. Itu adalah prosedur sederhana bagaimana mesin pencari Google menjalankan sebuah *query*.

IV. CARA KERJA PAGERANK

PageRank bekerja dengan cara membandingkan kepopuleran suatu situs Web dengan situs Web yang lainnya. Sebuah situs Web akan semakin populer jika semakin banyak situs Web lain yang meletakkan link yang menuju ke situs Web tersebut. Peringkat halaman dihitung dalam skala 1-10. Situs Web yang memiliki peringkat lebih tinggi akan muncul terlebih dahulu dalam hasil pencarian Google.

Banyak cara yang digunakan oleh sebuah *search engine* untuk menentukan kualitas atau ranking dari sebuah halaman Web. Namun, pada PageRank pendekatan yang digunakan adalah sebuah halaman Web dianggap penting jika halaman Web lain memiliki link ke halaman Web tersebut. Sebuah halaman Web juga akan menjadi semakin berkualitas atau populer, jika sebuah halaman Web lain yang memiliki ranking (PageRank) tinggi memiliki link menuju halaman Web tersebut.

Dengan pendekatan ini, maka proses terjadi secara rekursif dimana sebuah ranking akan ditentukan oleh ranking dari halaman Web lain yang rankingnya ditentukan oleh ranking halaman Web lain yang memiliki link ke halaman Web tersebut. Di dunia maya, ada jutaan bahkan milyaran halaman Web. Oleh karena itu sebuah ranking halaman Web ditentukan dari struktur link dari keseluruhan halaman Web yang ada di dunia maya. Sebuah proses yang sangat banyak dan kompleks.

Dari pendekatan yang sudah dijelaskan diatas, Lawrence Page and Sergey Brin membuat algoritma PageRank seperti di bawah:

$$PR(A) = (1-d) + d ((PR(T1) / C(T1)) + \dots + (PR(Tn) / C(Tn)))$$

PR(A) adalah Pagerank halaman Web A

PR(T1) adalah Pagerank halaman Web T1 yang mengacu ke halaman Web A

C(T1) adalah jumlah link keluar (*outbound link*) pada halaman Web T1

d adalah damping factor yang bisa diberi antara 0 dan 1, namun biasa diberi nilai 0,85.

Dari algoritma di atas dapat dilihat bahwa PageRank sebuah halaman Web ditentukan dari PageRank halaman Web yang mengacu kepadanya yang juga menjalani proses penentuan PageRank dengan cara yang sama, jadi proses ini akan berulang sampai ditemukan hasil yang tepat. Akan tetapi PageRank halaman Web T1 tidak langsung diberikan kepada halaman Web yang dituju, akan tetapi sebelumnya dibagi dengan jumlah link yang ada pada halaman Web T1 (*outbound link*), dan PageRank itu akan dibagi rata kepada setiap link yang ada pada halaman Web tersebut.

V. KESIMPULAN

Graf banyak sekali membantu untuk memodelkan persoalan sehingga lebih mudah untuk diselesaikan. Salah satunya adalah persoalan mesin pencarian tersebut. Dengan menggunakan pemodelan graf berbobot pada setiap situs *Web* di *World Wide Web*, kita bisa meranking setiap situs *Web*, sehingga kita bisa mendapatkan hasil yang sesuai dengan apa yang kita harapkan ketika melakukan pencarian di mesin pencari.

REFERENSI

- [1] Rosen, Kenneth H. 2012. *Discrete Mathematics and Its Applications*. McGraw-Hill : New York.
- [2] Munir, Rinaldi. 2006. *Diktat Kuliah IF2120 Matematika Diskrit*. Institut Teknologi Bandung : Bandung.
- [3] B. Sergey and P. Lawrance, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems archive* vol. 30, pp 107-117, April 1998.
- [4] PageRank: The Graph Theory-based Backbone of Google. <https://blogs.cornell.edu/info2040/2011/09/20/pagerank-backbone-of-google>. Diakses pada tanggal 8 Desember 2016 pukul 22.01.
- [5] Di Balik Penelusuran. <https://www.google.com/insidesearch/howsearchworks/> Diakses pada tanggal 9 Desember 2016 pukul 09.28.
- [6] History of The Internet. http://www.let.leidenuniv.nl/history/ivh/frame_theorie.html Diakses pada tanggal 9 Desember 2016 pukul 10.11.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 9 Desember 2016



Adya Naufal Fikri – 13515130