

PEMERINGKATAN HALAMAN WEB PADA MESIN PENCARI INTERNET MENGGUNAKAN RANTAI MARKOV

DANNY IRWAN, MAHDHIVAN SYAFWAN, MONIKA RIANTI HELMI

*Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : dannyirwan7@gmail.com*

Diterima 22 Juni 2019 Direvisi 6 Juli 2019 Dipublikasikan 4 Agustus 2019

Abstrak. Pemerinkatan halaman web didasarkan pada banyaknya jumlah kunjungan ke halaman web tersebut. Namun cara ini tidak efektif dilakukan untuk *webgraph* dengan skala besar. Untuk mengatasi permasalahan tersebut, digunakan rantai Markov. Masalah baru muncul ketika suatu *webgraph* memuat beberapa halaman web yang tidak mempunyai tautan luar ke halaman lainnya, sehingga hasil pemerinkatannya menjadi tidak realistis. Oleh karena itu, teknik dasar yang digunakan mesin pencari internet dalam pemerinkatan halaman web yaitu dengan metode rantai Markov menggunakan faktor redaman. Pada makalah ini dibahas langkah-langkah dalam pemerinkatan halaman web. Sebagai contoh, metode rantai Markov dengan faktor redaman diterapkan pada *webgraph* yang diperoleh dari *EECS Instructional and Electronic Support University of California Berkeley*. Dari hasil pemerinkatan diperoleh halaman dengan situs *encyclopedia.com* memiliki peringkat tertinggi karena berisi rangkuman informasi dari semua cabang ilmu pengetahuan.

Kata Kunci: Rantai Markov, Webgraph, Faktor Redaman

1. Pendahuluan

Pada era globalisasi ini, manusia dapat mengakses informasi dengan mudah melalui internet. Internet telah menjadi kebutuhan penting untuk mendapatkan informasi dengan cepat. Cara untuk mencari informasi dengan mudah di internet adalah dengan menggunakan *search engine* atau mesin pencari. Dengan mesin pencari ini, seseorang dapat menemukan informasi yang dibutuhkan dengan mengetik *keywords* atau kata kunci yang relevan. Dalam hal ini mesin pencari melakukan penyaringan jutaan halaman di WWW yang cocok dengan kata kunci. Peringkat dari setiap halaman web yang cocok dengan kata kunci didasarkan pada banyaknya jumlah kunjungan ke setiap halaman web tersebut.

Perpindahan setiap halaman web yang terjadi karena melakukan proses pencarian pada mesin pencari mengakibatkan perbedaan jumlah kunjungan ke setiap halaman web. Hal tersebut membuat setiap halaman web memiliki peringkatnya masing-masing. Jika jumlah halaman web masih relatif sedikit, maka pe-

meringkatan halaman web dapat dilakukan berdasarkan jumlah kunjungan pada halaman-halaman web yang ada. Namun, masalah muncul saat mesin pencari ingin menentukan kecocokan pencarian dan membuat pemeringkatan dengan jumlah halaman web yang sangat banyak. Untuk mengatasi permasalahan tersebut, digunakan rantai Markov.

2. Landasan Teori

2.1. Rantai Markov

Rantai Markov adalah proses kejadian dimana peluang bersyarat kejadian yang akan datang tergantung pada kejadian sekarang. Analisis Markov merupakan suatu bentuk khusus dari model probabilitistik yang lebih umum dikenal dengan proses stokastik.

Definisi 2.1. [3] *Proses stokastik $X_n(t)$ merupakan serangkaian peubah acak yang berubah terhadap waktu pengamatan $t \in T$.*

Definisi 2.2. [3] *Suatu proses stokastik $X_n(t)$ dikatakan memiliki sifat Markov jika $Pr(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = Pr(X_{n+1} = j \mid X_n = i)$ untuk waktu $n = 0, 1, \dots$ dan untuk setiap $j, i, i_{n-1}, \dots, i_1, i_0$.*

Definisi 2.3. [1] *Jika sebuah **rantai Markov** mempunyai k kemungkinan keadaan, maka peluang bahwa sistem berada dalam keadaan i pada sebarang pengamatan setelah sistem tersebut berada dalam keadaan j pada pengamatan sebelumnya dilambangkan dengan p_{ij} dan disebut peluang transisi dari keadaan j ke keadaan i . Matriks $P = [p_{ij}]$ disebut matriks transisi peluang (disingkat matriks transisi), dengan $p_{ij} \geq 0$ dan $\sum_{i=1}^k p_{ij} = 1$ untuk setiap $j = 1, 2, \dots, k$.*

Definisi 2.4. [1] *Vektor keadaan (state vector) untuk sebuah pengamatan pada suatu rantai Markov yang mempunyai k keadaan adalah sebuah vektor kolom \mathbf{x} dimana komponen ke- i , yakni x_i , menyatakan peluang bahwa sistem berada dalam keadaan ke- i .*

3. Pembahasan

Diasumsikan bahwa seorang pengunjung tidak harus mengikuti tautan di halaman yang sedang dikunjungi, tetapi juga dapat memilih setiap halaman dalam *web-graph* dengan suatu peluang tertentu. Misalkan δ menyatakan peluang seseorang mengunjungi halaman web berikutnya dengan memilih suatu tautan yang ada pada halaman yang sedang dikunjungi. Dengan demikian peluang seseorang memilih halaman berikutnya secara acak (tanpa mengikuti tautan) adalah $1 - \delta$. Nilai δ disebut **faktor redaman (damping factor)**. Jika terdapat n halaman di dalam suatu *webgraph*, maka peluang seseorang memilih halaman berikutnya secara acak yaitu $\frac{1 - \delta}{n}$. Vektor keadaan yang dihasilkan dari sejumlah klik berturut-turut akan membentuk barisan

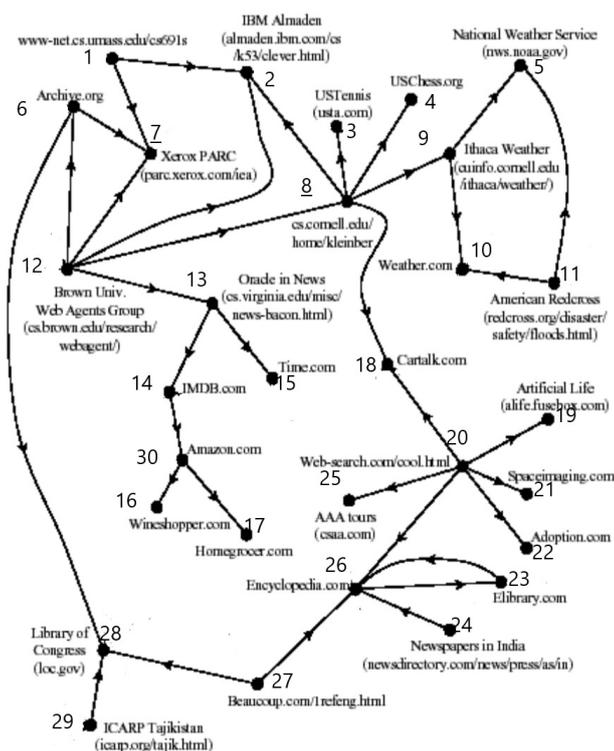
$$\mathbf{x}^{(k)} = B\mathbf{x}^{(k-1)}, k = 1, 2, 3, \dots \quad (3.1)$$

Entri-entri dari \mathbf{x} tersebut dapat diinterpretasikan sebagai proporsi dari jumlah kunjungan ke setiap halaman web ketika banyaknya klik menuju ke takhingga.

Karena $\mathbf{x}^{(k-1)}$ adalah vektor keadaan dimana entri-entrinya bernilai taknegatif dan berjumlah 1, maka persamaan terakhir menjadi

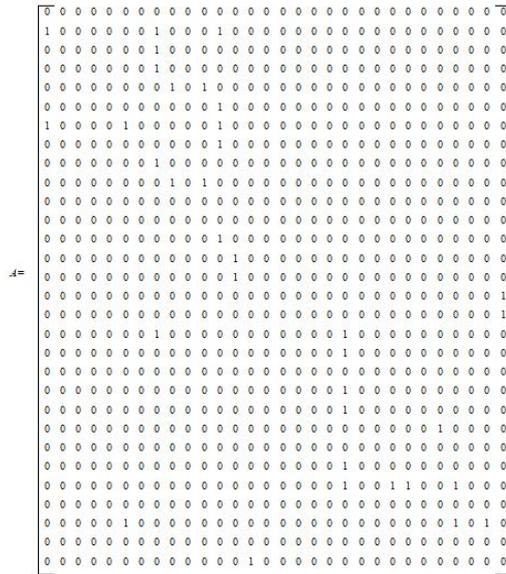
$$\mathbf{x}^{(k)} = \delta B\mathbf{x}^{(k-1)} + \frac{1-\delta}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.2}$$

Diberikan sebuah *webgraph* pada Gambar 1 yang diperoleh dari [2] yang menunjukkan lintasan berarah dari halaman-halaman web yang memuat topik tentang pendidikan yang dibuat oleh *EECS Instructional and Electronic Support University of California Berkeley*.

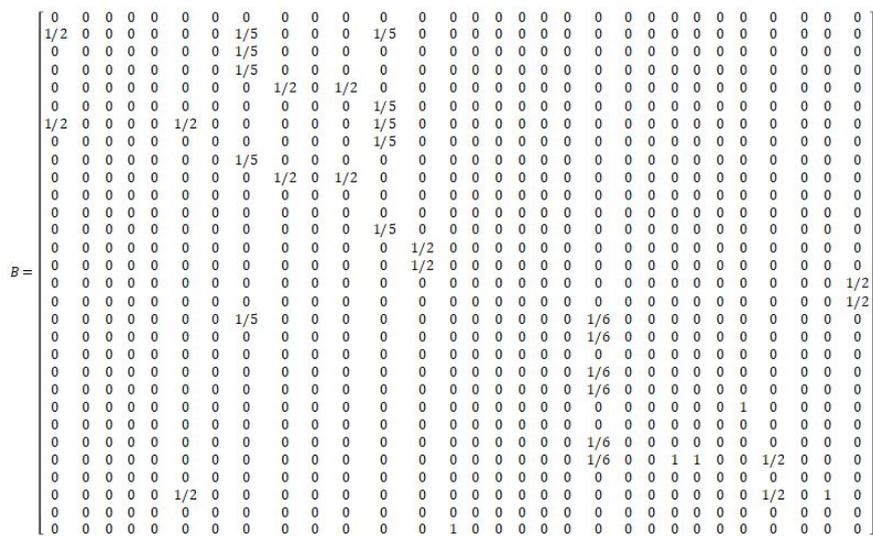


Gambar 1. [2] Webgraph

Berdasarkan *webgraph* tersebut dapat dibuat matriks keterhubungan dan matriks transisi peluang berturut-turut seperti pada Gambar 2 dan Gambar 3.

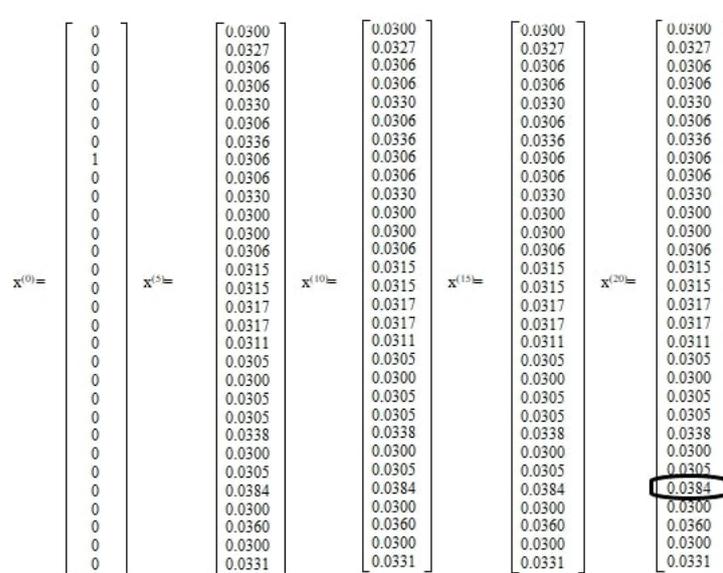


Gambar 2. Matriks Keterhubungan



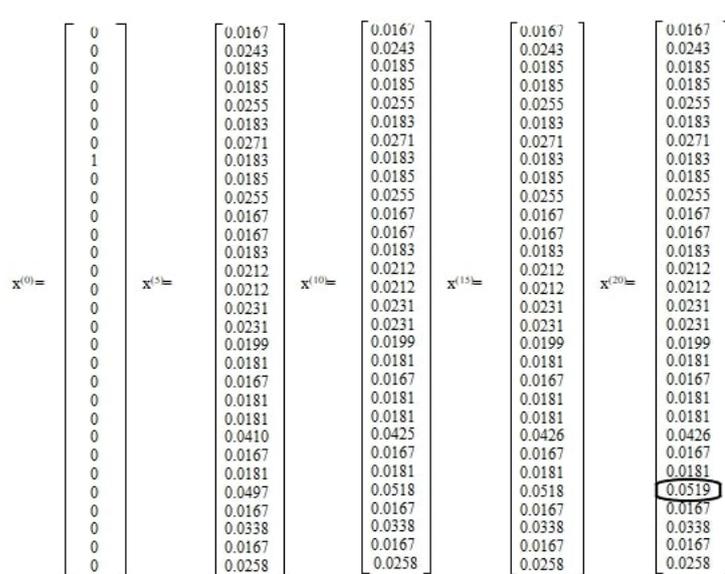
Gambar 3. Matriks Transisi

Misalkan seseorang memulai dari halaman 8. Vektor keadaan awal $\mathbf{x}^{(0)}$ dan perhitungan untuk vektor keadaan $\mathbf{x}^{(k)}$ menggunakan Persamaan (3.2) dengan memilih $\delta = 0.1$ untuk $k = 5, 10, 15$ dan 20 memberikan hasil pada Gambar 4.



Gambar 4. $\delta = 0.1$, $k = 5, 10, 15$ dan 20

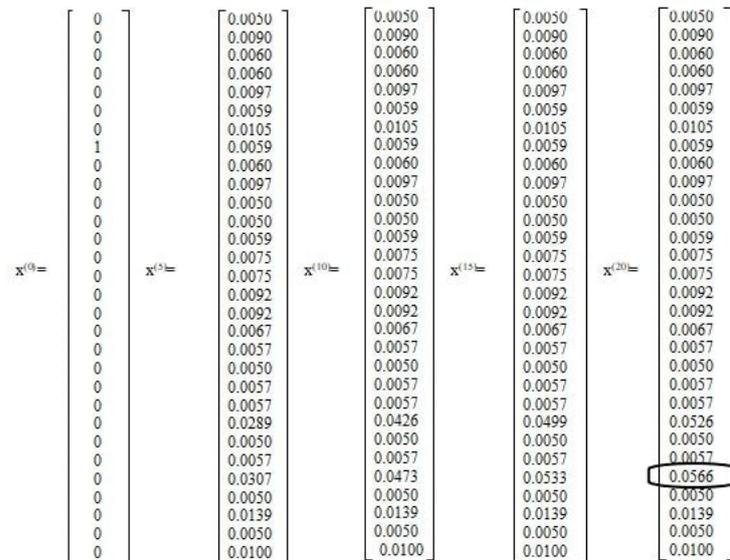
Dengan memilih $\delta = 0.5$ untuk $k = 5, 10, 15$ dan 20 , diperoleh hasil pada Gambar 5.



Gambar 5. $\delta = 0.5$, $k = 5, 10, 15$ dan 20

Selanjutnya dengan memilih $\delta = 0.85$ untuk $k = 5, 10, 15$ dan 20 , diperoleh

hasil pada Gambar 6.



Gambar 6. $\delta = 0.85$, $k = 5, 10, 15$ dan 20

Dari ketiga pilihan faktor redaman, terlihat bahwa halaman 26, yaitu situs *encyclopedia.com*, memiliki peringkat pertama dalam proses pencarian. Hal ini dinilai wajar karena situs tersebut berisi tentang rangkuman informasi dari semua cabang ilmu pengetahuan yang tentu saja memiliki peluang yang lebih besar dikunjungi oleh sivitas akademika di University of California Berkeley dibandingkan dengan halaman-halaman web yang lain. Jika nilai faktor redaman yang dipilih semakin besar, maka akan diperoleh nilai proporsi setiap halaman juga semakin besar. Peringkat dari halaman web yang dikunjungi disajikan pada Tabel 1.

Tabel 1. Tabel Peringkat Halaman

Peringkat	Halaman Web yang dikunjungi	Peringkat	Halaman Web yang dikunjungi
1	26	16	20
2	7	17	19
3	28	18	14
4	6	19	3
5	30	20	2
6	11	21	25
7	5	22	17
8	29	23	13
9	10	24	27
10	9	25	12
11	4	26	24
12	8	27	16
13	18	28	15
14	23	29	21
15	22	30	1

4. Kesimpulan

Teknik dasar yang digunakan mesin pencari internet dalam pemeringkatan halaman web yaitu dengan metode rantai Markov menggunakan faktor redaman. Adapun langkah-langkahnya sebagai berikut :

- (i) Konstruksi matriks keterhubungan dan matriks transisi peluang dari suatu *webgraph* yang diberikan.
- (ii) Tetapkan sebuah vektor keadaan awal dan pilih nilai faktor redaman, kemudian gunakan Persamaan (3.2) untuk mencari vektor keadaan di setiap banyak klik.
- (iii) Entri dari vektor keadaan yang mempunyai nilai tertinggi menunjukkan halaman yang memiliki peringkat tertinggi.

5. Ucapan Terima kasih

Ucapan terima kasih disampaikan kepada Dr. Ferra Yanuar, Dr. Ahmad Iqbal Baqi, Dr. Susila Bahri yang telah memberi masukan dan saran sehingga penelitian ini dapat terselesaikan.

Daftar Pustaka

- [1] Anton, H. 2013. *Elementary Linear Algebra*. 11th Edition. Willey Global Education, New Jersey.
- [2] EECS. 2014. *Example World Wide Web*. [http://www-inst.eecs.berkeley.edu/cs61bl/r/cur/graphs/world-wide-web.](http://www-inst.eecs.berkeley.edu/cs61bl/r/cur/graphs/world-wide-web/), diakses pada 28 Februari 2019.
- [3] Ross, S M. 1982. *Stochastic Process*. John Willey and Sons Inc, Berkeley.