Perbandingan Peformansi Terhadap Algoritma *Breadth First Search (BFS) & Depth First Search (DFS)* Pada Web Crawler

Aditya Eka Wibowo[1], Dr.Kemas Muslim Lhaksmana, S.T., M.ISD [2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]aditekaw@students.telkomuniversity.ac.id, [2]kemasmuslim@telkomuniversity.ac.id

Abstrak

Seiring berkembang pesatnya dunia internet dan kebebasan dari sesorang untuk membuat suatu halaman web maka mengakibatkan halaman web bekembang jumlahnya dengan sangat pesat dan hal tersebut menjadi suatu permasalahan untuk seseorang melakukan pencarian data yang memang dibutuhkan dari suatu halaman web. Banyak pengguna yang mencari suatu berita akan tetapi masih meragukan akan suatu isi dari informasi tersebut. Maka diperlukan suatu scan atau "crawl" ke semua halaman-halaman Internet untuk membuat index dari data yang dicarinya. Dalam penelitian ini mengkaji hasil pengujian menunjukkan bahwa BFS dan DFS berhasil *crawling* URL dengan baik serta menganalisa *backlinks* terhadap SEO yang berbeda.

Kata kunci : ***Web Crawler, Breadth First Search, Depth First Search, backlink***

Abstract

Along with the rapid development of the internet and the freedom of someone to create a web page, the number of developing web pages is very rapid and it becomes a problem for someone to search for data that is really needed from a web page. Many users are looking for a news but still doubt about the contents of that information. Then a scan or "crawl" is needed on all Internet pages to create an index of the data that it is looking for. In this study, reviewing the results of the test shows that BFS and DFS successfully crawling URLs well and analyzing backlinks to different SEOs.

Keywords: ***Web Crawler, Breadth First Search, Depth First Search, backlink***

1. Pendahuluan

Latar Belakang

Pertumbuhan *World Wide Web* yang eksplosif membuat sukar menemukan informasi yang sesuai dengan keinginan pemakai. Terlalu banyak server dan halaman yang harus dilihat dan dilakukan secara on line tetap merupakan tugas yang mengkonsumsi waktu. Hal inilah yang disebut masalah penemuan sumberdaya internet (internet resource discovery problem). Seiring dengan meningkatnya pertumbuhan jumlah halaman web, tentunya berkembang pula kita dalam mengumpulkan dan mengolah suatu data. Untuk datanya bisa berbagai bidang baik dalam bidang pendidikan, bisnis, olahraga, maupun sosial [2].

Dengan semakin banyaknya kebutuhan informasi dan publikasi terhadap suatu data, sehingga dengan mudah mengakses data yang beredar di internet kemudia dijadikan bahan acuan atau referensi oleh pihak pihak yang berkaitan di dunianya. Dengan berdasarkan data yang tersebar, maka untuk mengetahui isi data tersebut, maka dibutuhkan cara untuk dapat membaca isi dari suatu data yang berada pada halaman web. Cara tersebut umunya disebut *Web crawler*. *Web crawler*, juga sering dikenal sebagai Web Spider atau Web Robot adalah salah satu komponen penting dalam sebuah mesin pencari modern. Fungsi utama *Web crawle*r adalah untuk melakukan penjelajahan dan pengambilan halaman-halaman Web yang ada di Internet. Hasil pengumpulan situs Web selanjutnya akan diindeks oleh mesin pencari sehingga mempermudah pencarian informasi di Internet [2].

Pada tugas akhir ini, akan membandingkan 2 buah algoritma, yaitu BFS, DFS, dan Backlink untuk di uji peformansinya. Untuk BFS, *Web crawler* akan menelusuri dokumen-dokumen atau file global terlebih dahulu lalu menelusuri dokumen-dokumen atau file lokal. Berbeda dengan DFS, dimana *Web crawler* menelusuri dokumen-dokumen atau file lokal dahulu. Kemudian menelusuri dokumen-dokumen atau file pada web lain. Salah satu metode penelusuran seperti ini adalah berdasarkan banyaknya jumlah *backlink* [8].

Perumusan Masalah

Berikut rumusan masalah dalam pengerjaan tugas akhir:

1. Bagaimana implementasi *Web Crawler* berbasis *Breadth First Search, Depth First Search & backlink* ?
2. Bagaimana perbandingan peformansi dari kedua algoritma tersebut berdasarkan parameter waktu yang dibutuhkan untuk melakukan *crawling* dan analisa *backlinks* terhadap SEO yang berbeda ?

Berikut hal-hal yang dibatasi pada tugas akhir ini:

1. Tugas akhir ini hanya terbatas pada pengujian peformansi BFS,dan DFS dengan tergantung pada banyaknya *Branching Factor/data* dan juga menganalisa *backlink* terhadap dua S.E.O yang berbeda.
2. Data yang diambil untuk Tugas Akhir ini dari ‘http://corpus.quran.com /’.
3. Branching factor yang akan di uji sebanyak 250, 500, 750.
4. Dua SEO untuk analisa *backlink* yaitu Moz dan Wayback.

Tujuan

Berikut tujuan yang ingin dicapai dalam pengerjaan tugas akhir:

1. Mengetahui hasil peformansi dari setiap kedua algoritma tersebut.
2. Menganalisis performansi *backlinks* terhadap S.E.O yang berbeda.

Organisasi Tulisan

Jurnal tugas akhir ini disusun sebagai berikut, pada Bagian dua, akan dijelaskan lebih lanjut tentnag *Web Crawler*, *Breadth First Search*, dan *Depth First Search*, dan analisa *backlink* pada dua SEO yang berbeda. Bagian tiga membahas tentang rancangan sistem dan skenario percobaan yang akan diujikan. Hasil dan analisis dibahas pada bagian empat. Terakhir, kesimpulan ditulis pada bagian lima.

2. Studi Terkait

Web Crawler

Sebuah web crawler (atau dikenal juga dengan nama lain web spider atau web robot atau web scutter) adalah sebuah program atau skrip otomatis yang menjelajah ke area world wide web untuk membuat copy dari sebuah halaman web yang sudah dikunjungi untuk selanjutnya oleh search engine akan dilakukan indexing agar proses pencarian menjadi cepat. Crawler dapat juga digunakan untuk melakukan maintenance otomatis seperti memeriksa link-link atau validasi kode html [3]. Walaupun banyak aplikasi untuk Web crawler, pada intinya semuanya secara fundamental sama.

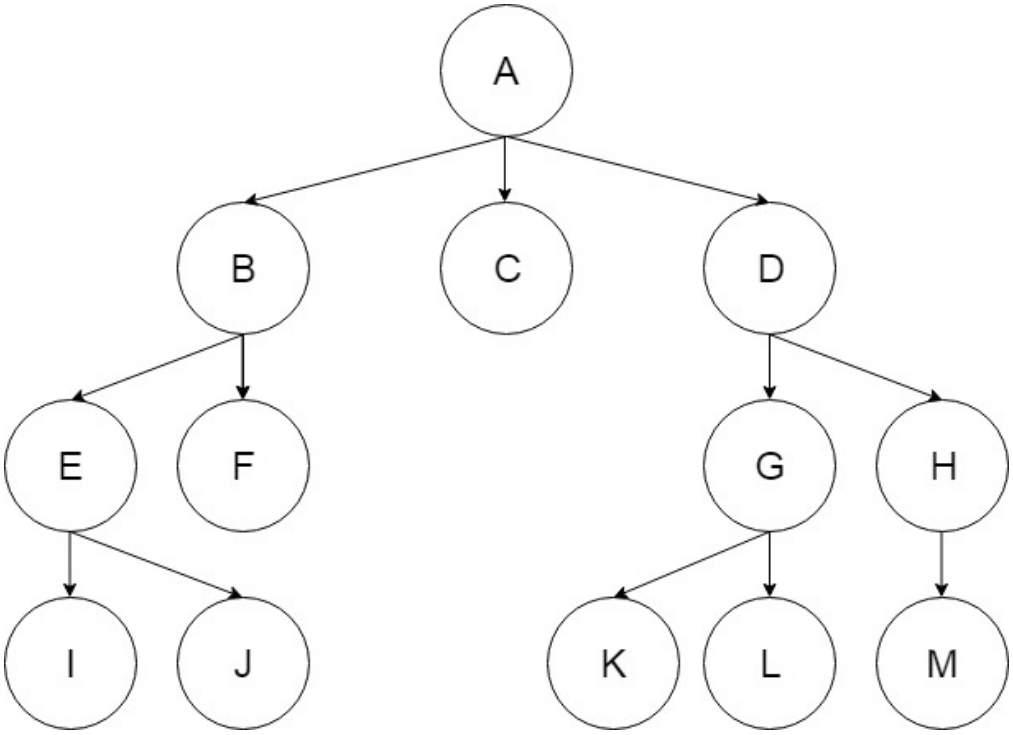
Dalam langkah pertama, sebuah web crawler mengambil URL dan mengunduh halaman dari Internet berdasarkan URL yang diberikan. Seringkali halaman yang diunduh disimpan ke sebuah file atau ditempatkan di basisdata [6]. Dengan menyimpan halaman web, maka crawler atau program yang lain dapat memanipulasi halaman itu untuk diindeks (dalam kasus mesin pencari) atau untuk pengarsipan untuk digunakan oleh pengarsip otomatis. Tahap kedua, Web crawler memparsing keseluruhan halaman yang diunduh dan mengambil link-link ke halaman lain. Tiap link dalam halaman didefinisikan dengan sebuah penanda HTML yang serupa dengan yang ditunjukkan disini [1]:

```
<A
HREF="http://www.host.com/directory/file.html
">Link</A>
```

Setelah crawler mengambil link dari halaman, tiap link ditambahkan ke sebuah daftar untuk dicrawler. Langkah ketiga dari Web crawling adalah mengulangi proses.Semua crawler bekerja dengan rekursif atau bentuk perulangan, tetapi ada dua cara berbeda untuk menanganinya. Link dapat dicrawl dalam cara Depth-first atau Breadth-first [1].

Breadth First Search (BFS)

Algoritma Breadth-First Search (BFS) atau dikenal juga dengan nama algoritma pencarian melebar adalah algoritma yang melakukan pencarian secara melebar yang mengunjungi simpul secara *preorder* yaitu mengunjungi suatu simpul kemudian mengunjungi semua simpul yang bertetangga dengan simpul tersebut terlebih dahulu. Selanjutnya, simpul yang belum dikunjungi dan bertetangga dengan simpul-simpul yang tadi dikunjungi, demikian seterusnya. [10].

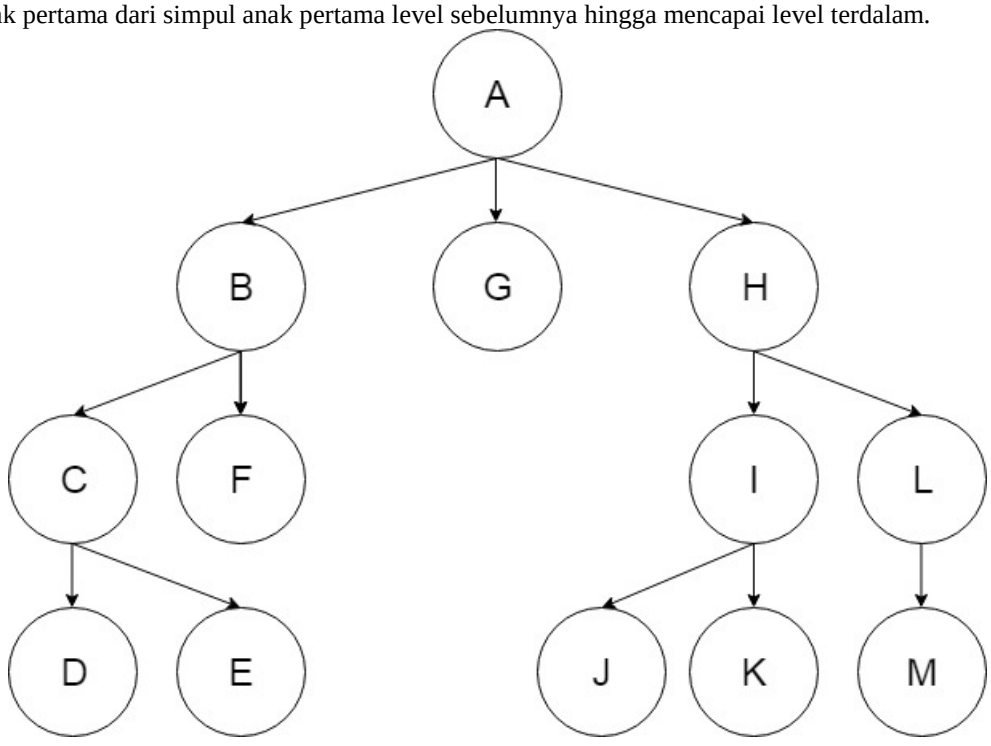


Gambar 2.1 Urutan langkah dari BFS.

Dengan menggunakan contoh pada **Gambar 2.1**, penelusuran dimulai dari **A** dan telusuri simpul miliknya; **B**,**C**, dan **D**. Setelah melakukan penelusuran terhadap simpul **A,B,C,** dan **D**, Lalu lanjutkan pencarian dari anak simpul **B** (**E** dan **F**), lalu anak simpul dari **D** (**G** dan **H**), dan seterusnya.

Depth First Search (DFS)

DFS (Depth-First-Search) adalah salah satu algoritma penelusuran struktur graf / pohon berdasarkan kedalaman. Simpul ditelusuri dari *root* kemudian ke salah satu simpul anaknya (misalnya prioritas penelusuran berdasarkan anak pertama [simpul sebelah kiri]), maka penelusuran dilakukan terus melalui simpul anak pertama dari simpul anak pertama level sebelumnya hingga mencapai level terdalam.



Gambar 2.2 Urutan langkah dari DFS.

Setelah sampai di level terdalam, penelusuran akan kembali ke 1 level sebelumnya untuk menelusuri simpul anak kedua pada pohon biner [simpul sebelah kanan] lalu kembali ke langkah sebelumnya dengan menelusuri simpul anak pertama lagi sampai level terdalam dan seterusnya [12].

Dengan menggunakan contoh pada **Gambar 2.2**, penelusuran dimulai dari **A** dan telusuri simpul miliknya; **B**,**C**, dan **D**. Setelah melakukan penelusuran terhadap simpul **A,B,C,** dan **D**, Lalu lanjutkan pencarian dari simpul disebelah kanan **E**, **F, G** dan seterusnya.

Backlink

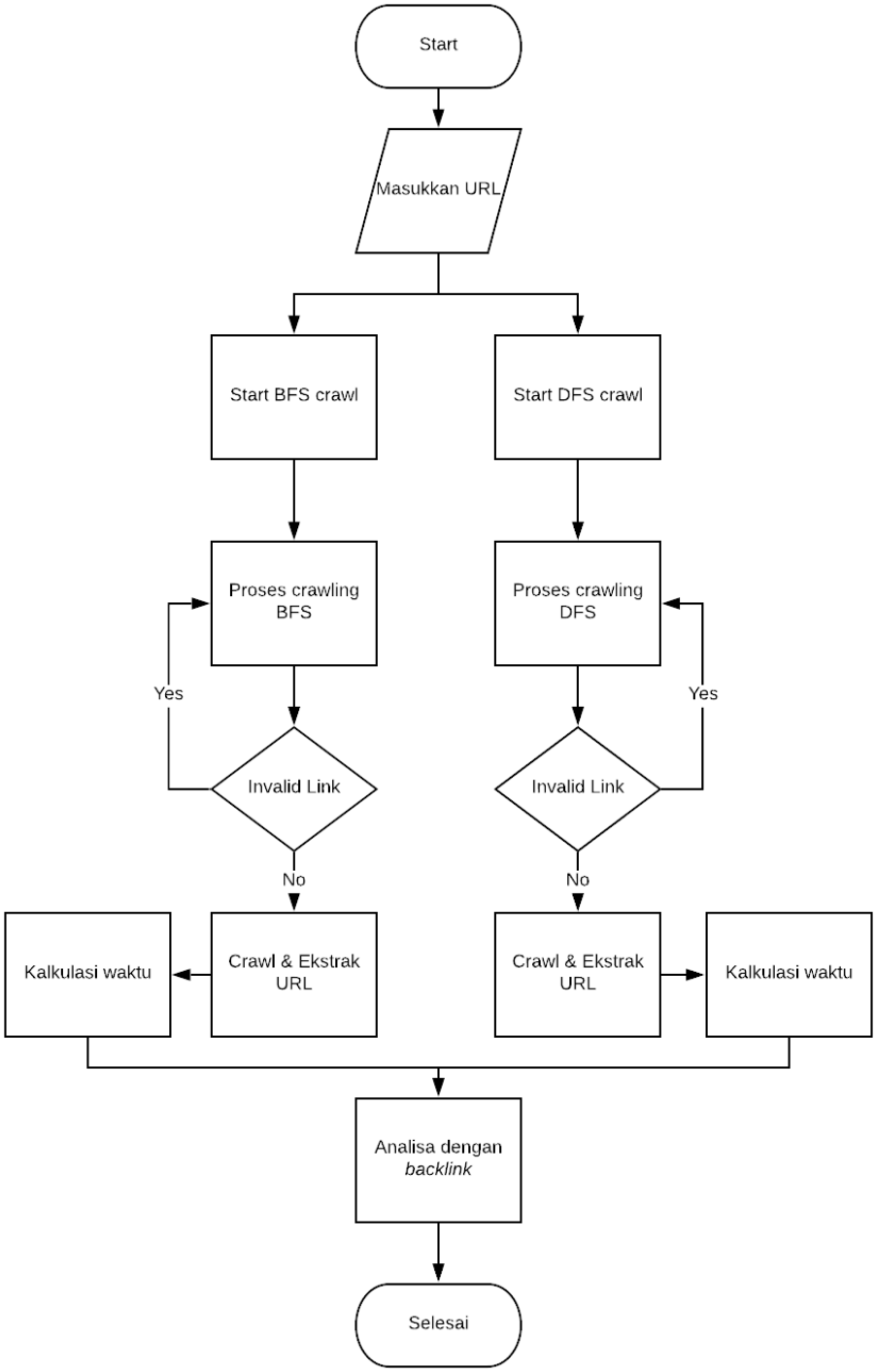
Backlink adalah hyperlink yang menghubungkan dari halaman Web, kembali ke halaman Web website itu sendiri. **Backlink** juga disebut sebagai *Inbound Link* (IBL). Link ini penting dalam menentukan popularitas (atau kepentingan) dari suatu website [13].

Backlink juga membantu anda dalam meningkatkan Google Page Rank dan Alexa Ranking [14]. Beberapa mesin pencari, termasuk Google akan mempertimbangkan website dengan backlink yang relevan dalam halaman hasil pencarian. Semakin banyak jumlah backlink yang mengarah ke website yang kita cari maka semakin tinggi kesempatan untuk menempati halaman satu atau peringkat satu di Google Search [13].

3. Desain Sistem

Deskripsi Sistem

Alur sistem data pengerjaan tugas akhir secara umum terdiri dari *input* URL, crawling dengan menggunakan *Breadth First Search* dan *Depth First Search*, kemudian ekstrak URL dan kalkulasi waktu, dan Analisa hasil dengan *backlink*. Alur rancangan sistem secara umum dapat dilihat pada gambar 3.1, pemaparan lebih lanjut dilanjutkan pada sub-bagian selanjutnya.



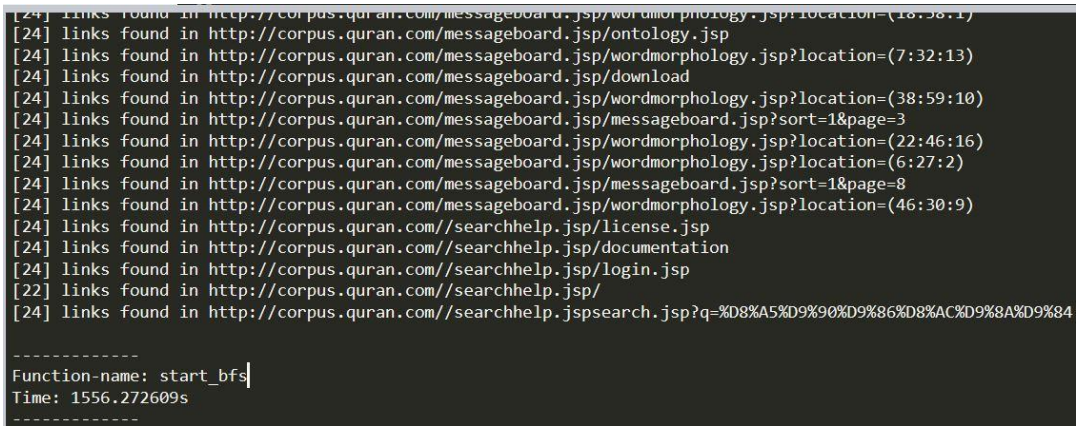
Gambar 3.1 *Flowchart* Deskripsi Sistem

Waktu proses

Waktu proses *crawling* adalah lamanya proses web crawler dalam melakukan *crawling* terhadap sekumpulan URL maupun dokumen web. Waktu proses yang diperlukan dalam melakukan proses *crawling* merupakan faktor sangat penting dalam pemilihan algoritma yang tepat untuk menangani data web yang semakin banyak.

Crawl* dan Ekstrak *URL

Sesuai gambar 3.1, Proses crawling BFS dan DFS dilakukan secara bersamaan. Adapun data yang akan diuji coba pada tugas akhir ini adalah jumlah *branching factors*. Jumlah data yang diuji coba adalah 250, 500, 750.

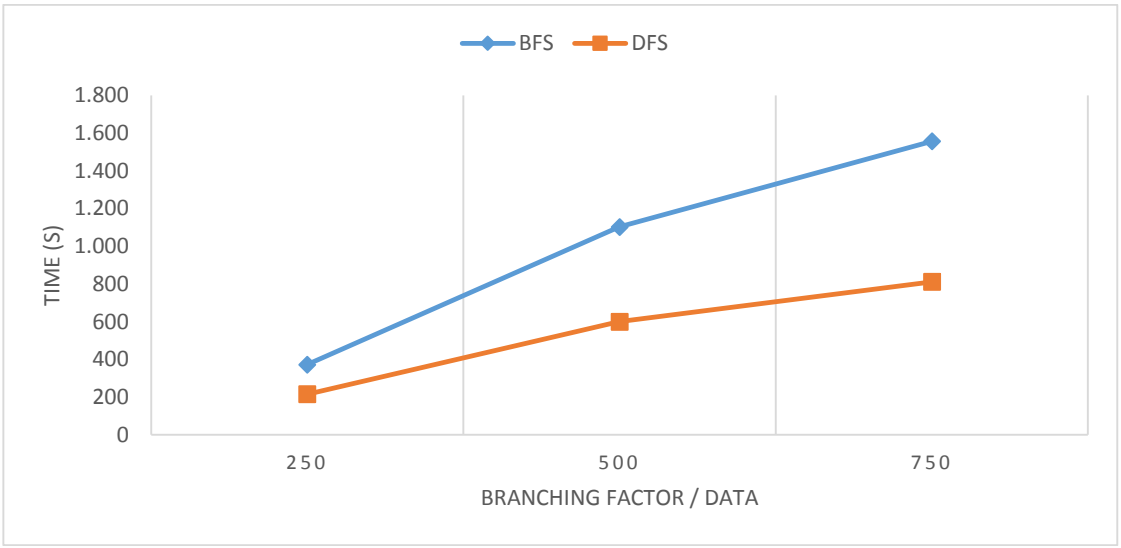


Gambar 3.1 Proses crawling pada BFS.

Setelah *crawling* selesai, hasil kalkulasi terhadap dua algoritma tersebut di analisa untuk diujikan pada data *backlink*. Hasil analisa data tersebut kemudian dievaluasi dengan SEO **Moz** dan **Wayback**.

4. Hasil Pengujian dan Analisis

Pertama-tama, dilakukan *input* URL. Setelah itu dilakukan dengan *crawling* dengan 2 algoritma tersebut. Berdasarkan tampilan grafik pada Gambar 4.1, pada *Branching Factor* 250 terlihat BFS memakan waktu sebanyak 373.300427s (6 menit 21 detik) dan DFS sebanyak 215.957440s (3 menit 6 detik). Lalu pada *Branching Factor* 500 BFS memakan waktu sebanyak 1102.421112s (18 menit 36 detik) dan DFS sebanyak 600.107345s (10 meni). Dan yang terakhir pada *Branching Factor* 750 BFS memakan waktu sebanyak 1556.272609s (26 menit 33 detik) dan DFS sebanyak 810.849228s (13 menit 51 detik). Dari hasil perbandingan tersebut diketahui bahwa algoritma BFS memakan waktu yang lama. Hasil kedua ditampilkan pada tabel 4.2. Analisa pada *backlink* pada tabel 4.3 dan tabel 4.4, menunjukan bahwa analisa *backlink* yang telah diuji dari SEO **Moz** dan **Wayback.**



Gambar 4.1 Grafik pengujian crawling pada BFS dan DFS.

Jumlah	Metode		Konversi dari detik ke menit	
	BFS	DFS	BFS	DFS
250	373.300427s	215.957440s	6min 21s	3min 6s
500	1102.421112s	600.107345s	18min 36s	10min
750	1556.272609s	810.849228s	26min 33s	13min 51s

Tabel 4.2 Perbandingan waktu antara BFS dan DFS.

Position	URL	Title	Descriptio	Page Authority	Total Links to Page	Total Linking Root Domains to Page	Domain Authority
34	http://cor	The Quran	Welcome	52	49928		70
41	http://cor	The Quran	The prope	48	42823		70
0	http://cor	The Quran	The Quran	55	29470		70
16	http://cor	English Tra	Welcome	48	862		70
14	http://cor	Syntactic T	Welcome	45	235		70
69	http://cor	Quranic Gr	This sectio	44	151		70
46	http://cor	The Quran	The list of	42	91		70
15	http://cor	Ontology c	The Quran	45	77		70
56	http://cor	The Quran	The list of	41	45		70
3	http://cor	Java API -	JQuranTre	42	42		70
65	http://cor	Java API -	Buckwalte	43	39		70
33	http://cor	The Quran	Where is t	43	37		70
30	http://cor	The Quran	We are cu	41	30		70
48	http://cor	The Quran	The topic i	42	28		70
79	http://cor	Part-of-sp	Traditiona	40	28		70
23	http://cor	The Quran	Yousuf M.	41	23		70
2	http://cor	Quranic Ar	To downlo	36	14		70
58	http://cor	Quranic Ar	1st May, 2	39	13		70
61	http://cor	Concept -	Concept is	38	12		70
60	http://cor	The Quran	You can us	40	10		70
1	http://cor	Document	Welcome	39	9		70
59	http://cor	The Quran	The Quran	39	9		70
63	http://cor	The Quran	You can sig	38	8		70
72	http://cor	Morpholog	Arabic has	37	8		70
57	http://cor	The Quran	You can se	38	7		70
77	http://cor	Dependen	The syntax	39	6		70
42	http://cor	Bibliograp	A Referenc	38	5		70
19	http://cor	The Quran	Version 3,	36	2		70
75	http://cor	Java API -	JQuranTre	36	2		70

Tabel 4.3 Analisa *backlinks* pada **Moz.**

BACKLINK REPORT

SOURCE URL	TARGET URL	CITATION FLOW	DATE	ANCHOR TEXT	NF	IMG
http://www.waqt.org/	http://quran.com/	62	Mar 20, 2019	qur'an	--	--
http://www.waqt.org/	http://corpus.quran.com/wordbyword.jsp	62	Mar 20, 2019	word by word	--	--
http://www.waqt.org/	http://alpha.quran.com/	62	Mar 20, 2019	new : alpha.quran.com	--	--
http://www.waqt.org/	http://alpha.quran.com/contributions	62	Mar 20, 2019	contribute	--	--

Tabel 4.4 Analisa *backlinks* pada **Wayback.**

5. Kesimpulan

Seluruh hasil pengujian menunjukkan bahwa BFS dan DFS berhasil *crawling* URL dengan baik. Adapun saran yang dapat diberikan pada tugas akhir ini, sebaiknya sistem diujikan kembali pada data URL yang berbeda dan dibandingkan dengan algoritma *searching* PageRank dan tools SEO yang sudah ada. Dan untuk kedepan, hasil analisa untuk *backlink* bisa di uji peformansi nya dengan algoritma *searching* lainnya.

Daftar Pustaka

[1] Suhartanto, M. (2017). pembuatan website sekolah menengah pertama negeri 3 delanggu dengan menggunakan php dan mysql. *Speed-Sentra Penelitian Engineering dan Edukasi*, *4*(1)..

[2] Sulatri, S., & Zuliarso, E. (2010). Aplikasi Web crawler Berdasarkan Breadth First Search dan Back-Link. *Dinamik-Jurnal Teknologi Informasi*, *15*(1).

[3] Juliasar, N., Juliasar, N., Sitompul, J. C., & Sitompul, J. C. (2013). Aplikasi Search Engine dengan Metode Depth First Search (DFS). *JURNAL MAHASISWA TI SI*.

[4] Henzinger, Monika R., "Hyperlink Analysis for The Web. California: IEEE Internet Computing, 2000.

[5] Henzinger, Monika R., "Link Analysis in Web Information Retrieval", California, 2001.

[6] Suchomel, V., & Pomikálek, J. (2012). Efficient web crawling for large text corpora. In *Proceedings of the seventh Web as Corpus Workshop (WAC7)* (pp. 39-43).

[7] Arifin, A. Z., Budianto, Lili, S., "Perancangan dan PembuatanPerangkat Lunak Penelusur Web (Web Crawler) Menggunakan Algoritma Pagerank", Teknik Informatika, Institut Teknologi Sepuluh November, Surabaya, 2003.

[8] Cho, Junghoo., Gracia-Molina, Hector., Page, Lawrence., "Efficient Crawling Through URL Ordering", Computer Networks and ISDN Systems, vol. 30 (1-7), pp. 161-172, 1998.

[9] Kustanto, Cynthia, Mutia S, Ratna, Viqarunnisa, Pocut, "Penerapan Algoritma Breadth-first Search dan Depth-first Search Pada FTP Search Engine for ITB Network", Teknik Informatika, Institut Teknologi Bandung, Bandung.

[10] Munir, Rinaldi. Strategi Algoritmik Diktat Kuliah IF2251. Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika, institut Teknologi Bandung, bandung. 2006.

[11] Google. Depth First Search. Available at: https://saungkode.wordpress.com/2014/04/16/penelusuran-pohon-biner-berdasarkan-kedalaman-dengan-algoritma-dfs-stack-dan-secara-melebar-level-order-dengan-algoritma-bfs-queue-dan-implementasinya-dalam-bahasa-c/. Diakses tanggal 23 April 2018.

[12] Google. Algoritma BFS. Available at: https://5onbuble.wordpress.com/2011/05/26/6/. Diakses tanggal 23 April 2018.

[13] Google. Backlink. Available at: http://www.boc.web.id/apa-itu-pengertian-backlinks/. Diakses tanggal 23 April 2018.

[14] Google. Backlink. Available at: https://www.infosaku.com/2013/02/cara-membuat-backlink.html. Diakses tanggal 23 April 2018.

[15] Wikipedia. Web crawler. Available at: https://en.wikipedia.org/wiki/File:WebCrawlerArchitecture.svg. Diakses tanggal 23 April 2018.