

# Analysis on Backlinks and PageRank of Automotive Company Websites Using Crawler Tool

Dina Agusten, ST., MMSI  
Department of Informatics Engineering  
Gunadarma University  
Depok, Indonesia

Winarti, SKom., MMSI  
Department of Information System  
Gunadarma University  
Depok, Indonesia

**Abstract** —The marketing strategy for a company is very important for gaining higher sale over time. Many companies utilized the Internet to develop marketing strategies, amongst which was a website that represent the company image. To attract as many visitors as possible, the website required optimization of its backlink/s and pagerank. How good the backlinks and how high the pagerank can be tested using a webcrawler. By using this tool, the company can improve its marketing strategy worldwide. This study gives result a backlink and pagerank information based on experiments done on the tool crawler. So that the company knows its position in web search site. Company makes better marketing strategies in the use of internet based on this study.

**Keywords-** component; tool crawler; analysis URL address; backlink; pagerank.

## I. INTRODUCTION

Growing automotive companies worldwide that each company to launch new innovations to keep their track on competition. Marketing department is important factor to ensure consumers getting to know the company's products. One of the company's marketing strategy to consumers worldwide is the utilization of Internet and company website.

Considering the high intensity on competition among automotive companies, each of the company tried to reach the market faster by utilizing website. The website would represent company image and introduce their products to potential buyers at global scale. In turn, this opened a very good opportunity to develop a marketing strategy related to company website. This analysis was intended to carry out a test of crawler tool as a mean for developing better website marketing strategy:

1. Provide information on pagerank and backlinks of the companies' websites, this will give the data about the position of the web on search engine.
2. Creating a marketing strategy by improving the position of the website on search engine.

## II. STATE OF THE ART

“An important part of a search engine is a crawler. Web crawler is a program that collects information to be placed on the database”[1]. “A web crawler in charge of searching the web and collect documents data in it. Then, the web crawler will sort the documents and build a list of index” [2]. “General-purpose web crawlers collect and process the entire

contents of the Web in a centralized location, so that it can be indexed in advance to be able to respond to many user queries [3]”. “Web crawler is a computer program that perform a search on the world wide web and store information that is available to a storage in an order” [4]. Web crawler architecture shown in figure 1.

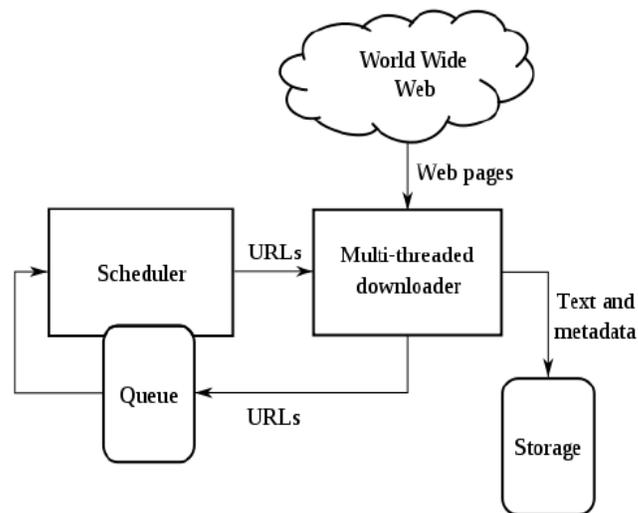


Figure 1. Architecture Crawler [3]

After the data is stored on the storage, indexing search engine will do the collecting, classifying, and storing data more regularly to facilitate the users' searching process on search engines more quickly and effectively. “Links are crawled in two ways: in a depth-first or breadth-first” [5].

“Depth-First crawling, employs a narrow, but deep, way of traversing the hypertext structure. This is in contrast to the wide and shallow traversal in Breadth-First approach. Starting from the seed page, the robot picks the first link on the page and follows it, then the first link on the second page, and so on until it cannot go deeper, returning recursively” [6][7].

“A Breadth-First crawler is the simplest strategy for crawling. It uses the frontier as a FIFO queue, crawling links in the order in which they are encountered. Note that when the frontier is full, the crawler can add only one link from a crawled page”[8].

**Metrics / Crawler Software Measurement**

“Not all web pages receive the same attention from a crawler. For example, if the crawler used to build a database that is specific to a particular topic, then the page that refers to a more important topic, and it should be visited as early as possible. If given a web page p, defined the importance of a page I (p)” [2][9].

According to [9] a crawler has important metrics consisting of; similarity query, forward link count, count backlinks, pagerank and location. Here is an explanation of each metric.

In conducting information search in the search engines, there are 2 types of queries similarity [10][11], namely:

- Document-query similarity is important for identifying similarity of user query to documents.
- Document-document similarity is important for finding similar documents in the document pool of a search engine. Some search engines represent this feature.

A (query) Q control crawling process, and I (p) Important page is defined as textual similarity between p (page) and Q (query). To calculate the similarity, we can see the web pages of each document as a vector of dimension n ( $w_1, w_2, \dots, w_n$ ). Term  $w_i$  is a vector that represents word in the literature. If  $w_i$  does not appear in the document, then  $w_i$  is zero. If it appears,  $w_i$  is set to represent the significance of the word. One common way to calculate the significance of  $w_i$  is by multiplying the number of the word in the document with the inverse document frequency (idf) of the word. Idf factor is one divided by the number of words in the overall appearance of the collection [12].

Metrics IF(p) or important forward link(page) counts the number of links coming from p (page). Using this metric, a page with many outgoing links are very valuable, because it is a Web directory. This metric can be calculated directly from p. In these measurements, a document judged by the number of links in the document. Forward links can be traced to the concept of breadth-first [13][2].

Inbound links or backlinks originated from other sites directing or leading to our site. Backlinks are very important because it can be a measure of the popularity of a website on the internet.

Every page has some number of forward links (outedges) and backlinks (inedges) (see Figure 2). We can never know

whether we have found all the backlinks of a particular page but if we have downloaded it, we know all of its forward links at that time [14].

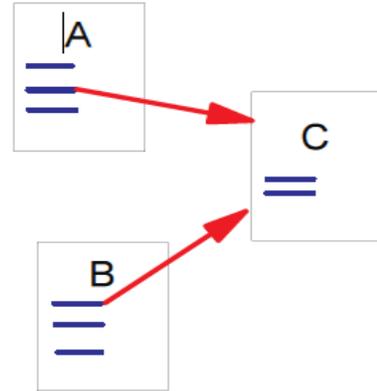


Figure 2. A and B are Backlink of C [14]

Web pages vary greatly in terms of the number of backlinks they have. For example, the Netscape home page has 62,804 backlinks in our current database compared to most pages which have just a few backlinks. Generally, highly linked pages are more “important” than pages with few links. Simple citation counting has been used to speculate on the future winners of the Nobel Prize [15]. PageRank provides a more sophisticated method for doing citation counting.

The reason that PageRank is interesting is that there are many cases where simple citation counting does not correspond to our common sense notion of importance. For example, if a web page has a link o. the Yahoo home page, it may be just one link but it is a very important one. This page should be ranked higher than many pages with more links but from obscure places. PageRank is an attempt to see how good an approximation to “importance” can be obtained just from the link structure [14].

Pagerank has the same basic concept of link popularity, but it does not only consider the “amount” of inbound and outbound links. The approach used is a page would be necessary if other pages have a link to that page. A page will also become increasingly important if other pages have a rank (pagerank) refers to the page height. According to [16], early algorithms are

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

And another published algorithm

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

Where:

- PR (A) is the PageRank page A
- PR (T1) is the PageRank of page T1 that refer to pages A
- C (T1) is number of outbound links on page T1
- d is a damping factor which can be between 0 and 1.

- N is the total number of web pages.

From the above algorithm can be seen that PageRank is determined for each searched page, not the whole website. PAGERANK of a page is determined by PageRank of the page(s) that refer to it which is also undergoing a process of determining the pagerank in the same way, so this process will be repeated until you find the right results.

However, pagerank page A is not given directly to the intended page, but formerly divided by the number of links on the page T1 (outbound links), and that pagerank will be divided equally to every link on the page. Likewise, any other pages "Tn" which refers to page "A".

After all pagerank obtained from other pages that refer to page "A" are sum, the value is then multiplied by the damping factor value between 0 and 1. This is to make sure not overall PageRank values of page T distributed to page A [16][9].

IL (p) where IL is Important Location and p is the page, a search result or crawling by looking at the important pages that match the criteria in the desired URL address corresponding to the presence server, instead of the content. For example, a URL that ends with ".com" is considered more useful than URLs with other endings or characters that contain "home" may be more of a priority than other URL [9].

### III. METHODOLOGY

For sampling data that will be tested, we take 10 site addresses of automotive companies from Japan, based on the fact that those companies have worldwide recognition, including in Indonesia. They are Toyota, Honda, Nissan Mazda, Suzuki, Hino, Daihatsu, Mitsubishi, Yamaha, Kawasaki. Other reason is because Japan is considered more innovative than other countries in the automotive industry. Speaking of the quality, Japan made on is not inferior to other countries, proven by the growing number of consumers who purchase these products at a relatively affordable price when compared to other countries. Do it and make sure the website address 10 Japanese automotive company is active. Overview of the research steps can be seen in figure 3.

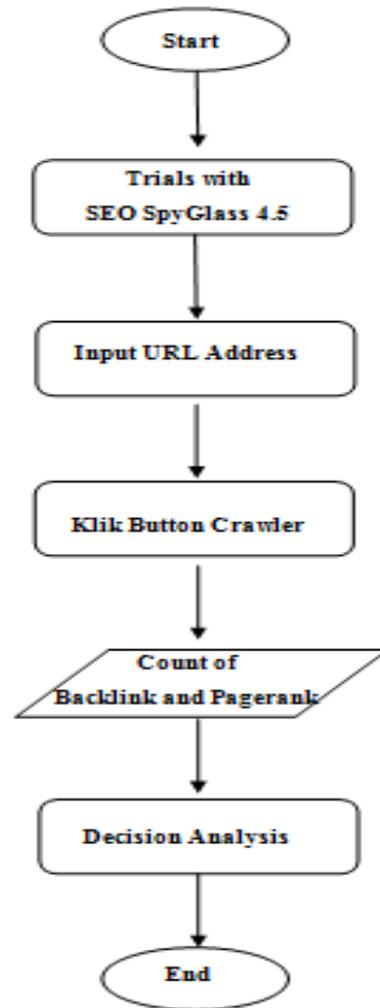


Figure 3. Research Steps

First of all we need is to ensure the web addresses of 10 Japanese automotive companies, and they are active pages. After that, the first step to enter the url address for the first sampling data. Click the start / crawl / next to the crawling / search. Crawling results will automatically provide information on the number of backlinks and pagerank. From these results we can can be analyzeand decide what strategy should be taken to enhance the company's position in order to get the first position.

### IV. RESULT AND DISCUSS

We use backlink and pagerank as crawler metrics. Backlink considered very important as web measurement because it comes from other sites that direct or to our site. Backlinks are very important because it can measure the popularity of a site on the internet. PageRank can be used to predict which URL's are qualified to be visited first. PageRank has search characteristics of depth first search and breadth-first search [1]. By using the two measurements, a maximum performance of a company website can be decided. PageRank will be taken based on google search google, since Google is the most popular search engine used by consumers.

There are many crawler tools available, from full proprietary licensed to the open source ones. Some of them are XenoCrawler, Win Web Crawler, Spideye, SEO SpyGlass, Screaming Frog Spider Tool, etc. In this study, as a test tool for the 10 web addresses of the Japanese automotive companies website, we use SEO SpyGlass, part of SEO Powersuite.

SEO SpyGlass helps the users to see both pagerank and backlinks of competitors' sites in a matter of minutes. Simply by entering the URL address, then of many important metrics in the crawler will appear and the user can also compare it with the company's own website or any other competitor. Thus, it provides helpful data for the company to set a good strategy in achieving its targets and goals. SEO SpyGlass can also help sort out the keywords that have been included so that users can know how many times the keywords will appear. SEO SpyGlass has search capabilities of 132 countries, where each country has at least one search engine. SEO SpyGlass is open source software or non proprietary. SEO SpyGlass developers also offer paid service if you want to optimize the performance of the crawler.

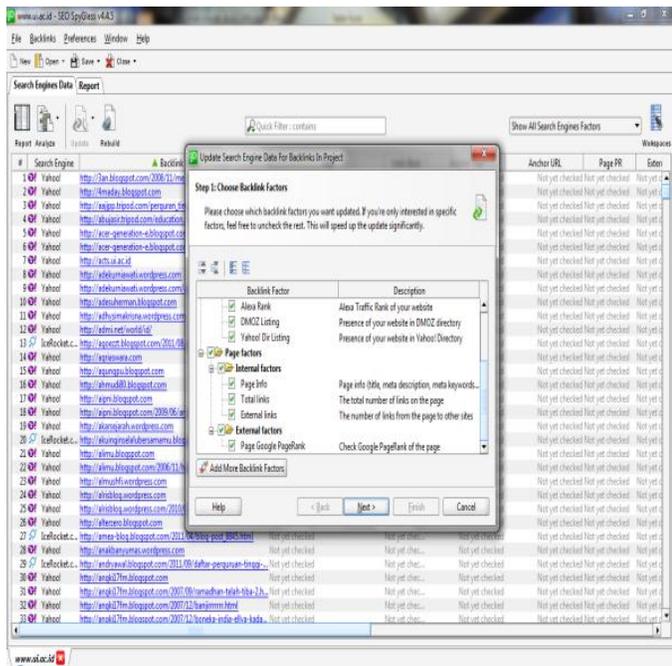


Figure 4. Software SEO SpyGlass [17]

Before the test, preparation should be done to avoid failure. This study tested 10 addresses of Japanese automotive company websites with SEO SpyGlass tool crawler by analyzing the results, i.e backlinks and pagerank. SEO SpyGlass is used based on the results of earlier studies for thesis writing [18]. The study used 10 automotive company website address specified. Those 10 addresses in table 1 selected on the basis of popularity among the masses. The obtained results would be analyzed to give a decision on the marketing strategy for the automotive company that owns the website.

Testing scenario is as follows:

1. Ensuring crawler tool is connected to the Internet.

2. Checking the URL address before the test.
3. Conducting the test using SEO SpyGlass, input web address
4. Determining the criteria that will be generated in the trial.
5. Analyzing the results after all of the 10 URL addresses being tested.

The test results of 10 automotive company website address will be included in a table according to the table 1. These results will provide clear information for automotive companies to improve and develop their respective website.

The test results on the 10 site addresses a Japanese automotive company that has been done by using SEO SpyGlass is presented in Table 2. At trial results pagerank 2 sources can be generated from the search engine Google and Alexa, the results can be seen in table 1.

TABLE I. PAGERANK COMPARISON DATA

NO	Software	URL Address	Google Pagerank	Alexa Rank
1	SEO SpyGlass	<a href="http://www.toyota.com/">http://www.toyota.com/</a>	7	4947
2		<a href="http://www.honda.com/">http://www.honda.com/</a>	6	4915
3		<a href="http://www.nissan.co.jp/EN/">http://www.nissan.co.jp/EN/</a>	6	8131
4		<a href="http://www.hino.com/">http://www.hino.com/</a>	5	1324131
5		<a href="http://www.daihatsu.com/">http://www.daihatsu.com/</a>	5	247927
6		<a href="http://www.mitsubishicars.com/">http://www.mitsubishicars.com/</a>	6	54966
7		<a href="http://www.mazda.com/">http://www.mazda.com/</a>	6	91131
8		<a href="http://www.suzuki.com/">http://www.suzuki.com/</a>	6	189217
9		<a href="http://www.yamaha-motor.com/">http://www.yamaha-motor.com/</a>	6	43381
10		<a href="http://www.kawasaki.com/">http://www.kawasaki.com/</a>	6	62081

Looking at the compared pagerank results in table 1, it is clear that the higher rank on Alexa is equivalent with Pagerank on Google. Backlink and Pagerank is part of the five key metrics of a crawler software. From both these metrics companies can obtain information of the number of web sites linking to it as well as its position on search engine. By using two metrics a company also can compare its position with a rival company.

The results of this study in addition to taking the test results from 18 search engines backlinks were found in 132 countries also take the results by searching google pagerank. The decision took the google pagerank of results because google is a popular search engine today. Based on table 2 it can be seen that the results obtained at the most backlinks by toyota company, but the resulting PageRank is at the lowest position when compared to 9 other automotive companies.

From the test results it is clear that the number of backlinks is not a guarantee that the web gets higher pagerank. Based on backlink and pagerank, the test results placed [www.daihatsu.com](http://www.daihatsu.com) at first rank although the address doesn't have as many backlink as 9 other sites tested. The reason could occur because some pages were no longer active, the page server down while testing occurs or keywords of the page few recognizable as information retrieval.

Based on research conducted by [19], a keyword is very important to show the position of pagerank. Decisions to enhance company website in order to improve marketing strategy can be drawn using our test result. The company can multifold the active backlinks as well as put as many useful keywords as possible in the website. This will make search engine users reach the website easier because its better position on search engine result. In turn this will increase company image, and consecutively, its sales and income.

TABLE II. TEST RESULT DATA

NO	Software	URL Address	Backlink	Page Rank	Rank
1	SEO SpyGlass	<a href="http://www.toyota.com/">http://www.toyota.com/</a>	16989	7	10
2		<a href="http://www.honda.com/">http://www.honda.com/</a>	16245	6	3
3		<a href="http://www.nissan.co.jp/EN/">http://www.nissan.co.jp/EN/</a>	4948	6	8
4		<a href="http://www.hino.com/">http://www.hino.com/</a>	704	5	2
5		<a href="http://www.daihatsu.com/">http://www.daihatsu.com/</a>	2166	5	1
6		<a href="http://www.mitsubishicars.com/">http://www.mitsubishicars.com/</a>	5357	6	7
7		<a href="http://www.mazda.com/">http://www.mazda.com/</a>	7795	6	5
8		<a href="http://www.suzuki.com/">http://www.suzuki.com/</a>	4641	6	9
9		<a href="http://www.yamaha-motor.com/">http://www.yamaha-motor.com/</a>	7981	6	4
10		<a href="http://www.kawasaki.com/">http://www.kawasaki.com/</a>	7761	6	6

Based on research conducted [1], the results of the test to backlinks and PageRank is still done manually by using a specific formula and it is unknown how many searches carried out by the search engines. In the test, automotive companies site address has been able to be tested with accurate results and clear information. From the results of this study, we can also see clearly that both metrics (backlinks and PageRank) determine final results. This is in line with information on figure 5.

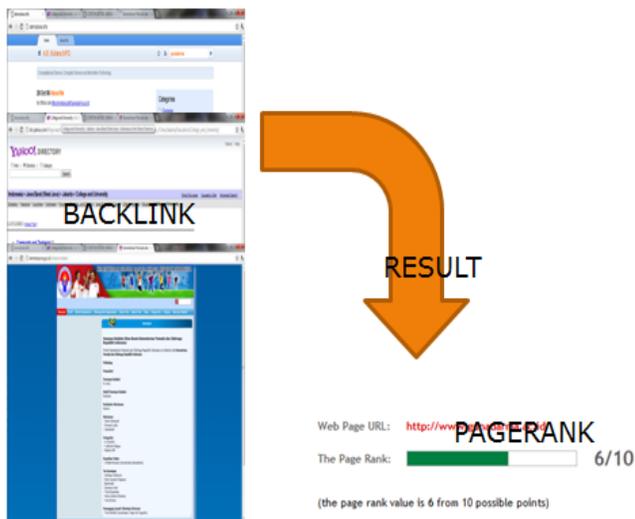


Figure 5. The influence of the Pagerank Backlink

According to [20] OICA correspondents survey report on world ranking of manufacturers that produced the largest production by toyota. Toyota also has a majority stake in Daihatsu and Hino thus making the company has have this clean up billions of yen. Based on the establishment of the company, Daihatsu first established that first recognized by the world market. By [21] that the distribution is expressed quite successful when considering the benefits and provide opportunities Micromarketing and control over the marketing relationship. Based on figure 5 which shows the number of backlinks so influential pagerank positions used by automotive companies to have a web of data sampling, so that if the company has a web highest pagerank or first position will have a great chance to be visited first by the consumer or information seekers.

V. CONCLUSION

The results of tests on 10 addresses of Japanese automotive company websites using SEO SpyGlass is also influenced by the factors of time and an internet connection is not stable and the growth of new web, so the number of backlinks to change the results in seconds. Based on the results of testing and analysis, the conclusions are:

1. The first rank is occupied by an automotive company website with address www.daihatsu.com with page rank position is 5 and has 2166 backlinks
2. Information comes from this test will become a useful input for the companies to develop a marketing strategy so they can improve the websites. The strategy will employ an increase of backlinks as well keywords in order to obtain top positions of pagerank

REFERENCES

- [1] Budianto, A. Z. Arifin, S. Lili. "Designing and Making Software Web crawler (Web Crawler) Using PageRank Algorithm", Proceedings of the National Seminar on Information Technology and Application, Tenth of November Institute Technology, Surabaya, 2003.
- [2] J. Cho, H. Garcia, M. L. Page. "Efficient Crawling Through URL Ordering", Computer Networks and ISDN Systems 30.1: 161-172. 1998.
- [3] S.S. Dhenakaran and K. T. Sambanthan, "Web Crawler-An Overview", International Journal of Computer Science and Communication, vol. 2, no.1, pp.265-267, January-June 2011.
- [4] G. Desrianti. "Accuracy Searching in Search Engines", Papers System and Information Technology Studies Program, School of Electrical Engineering and Informatics, Bandung Institute of Technology, 2010.
- [5] Sulastri and E. Zuliarso. "Crawler Web Applications Based on Breadth First Search and Back-Links", Faculty of Information Technology, University of Stikubank Semarang, 2010.
- [6] Dr. H. H. Saleh & Dr I. T. Ali, "Effective Web Page Crawler", Eng. & Tech. Journal, Vol.29, No.3, 2011.
- [7] Hafri, Y. Gjeraba, C., "Dominos: A New Web Crawler's Design", Ecole Polytechnique de Nates, September 16, 2004. Available at: <http://iww.europarchive.org/04/Harfi.pdf>.
- [8] R. Preziosi, "An Information Guided Spidering: A Domain Specific Case Study (Thesis)", Universita Degli Studi Di Salerno, 2007-2008.
- [9] J. Cho. "Crawling the Web: Discovery and Maintenance of a Large-Scale Web Data", Ph.D. Dissertation, Department of Computer Science, Stanford University, November 2001.
- [10] A.A Barfouroush, H. R. M. Nezhad, M.L. Anderson, D. Perlis. "Information Retrieval on the World Wide Web and Active Logic: A Survey and Problem Definition", University of Maryland, 2002.

- [11] M. Kobayashi and K. Takeda, Information Retrieval on the web: Selected Topics, IBM Research, Tokyo Research Laboratory, IBM, Japan, 1999.
- [12] A. Arasu, J. Cho, H. Garcia-molina, A. Paepcke, S. Raghavan "Searching the Web" ACM Transactions On Internet Technology - TOIT, 2001. Available at <http://dbpubs.stanford.edu/pub/2000-37>.
- [13] E. Zuliarso, K. Mustofa. "Crawling Web Based On Ontology", Journal of Information Technology DYNAMIC, vol. XIV, no.2, July 2009.
- [14] L. Page, S. Brin, R. Motwani, T. Winograd. "The PageRank Citation Ranking: Bringing Order To The Web". Technical Report Stanford Digital Library Technologies Project SIDL-WP-1999-0120, 1998.
- [15] Neeraja, Sankaran. Speculation In The Biomedical Community Abounds Over Likely Candidates For Nobel. The Scientist, 9.19., Oct 1995. [http://www.the-scientist.library.upenn.edu/yr1995/oct/nobel\\_951002.html%](http://www.the-scientist.library.upenn.edu/yr1995/oct/nobel_951002.html%).
- [16] S. Brin and L. Page. "The anatomy of a large-scale hypertextual Web search engine". Computer Networks and ISDN Systems 30: 107-117. doi:10.1016/S0169-7552(98)00110-X. ISSN 0169-7552, 1998.
- [17] <http://www.seopowersuite.com/download.html>, visited on 10 March 2012.
- [18] D. Agusten. "Crawler Software Performance Analysis Case Study: Universities Website (MMSI. Thesis)", Gunadarma University, March 2012.
- [19] E. S. O. Purba, Ismail, T. N. Adi. "Analysis of Implementation of SEO (Search Engine Optimization) Website PT. Polypet Karyapersada", Program Information Management Studies, Polytechnic Telkom Bandung, 2011.
- [20] "World Motor Vehicle Production: World Ranking of Manufacturers Year 2010" (PDF). OICA. accessed on February 16, 2013.
- [21] K. Peattie, , "The Marketing Mix in the Third Age of Computing", "Marketing Intelligence and Planning, 15(3):142-150, 1997.

#### AUTHORS PROFILE

**Dina Agusten, ST., MMSI**, obtaining an engineering degree (ST) Program of Study Informatics Engineering Gunadarma University graduated in 2005. In 2012 obtaining Master of Information Systems Management (MMSI) of Information Systems Software Program Gunadarma University. Today is as permanent staff secretariat information systems majors and as lecturer at Gunadarma University.

**Winarti, SKom., MMSI**, obtaining a computer science degree (S.Kom) Program of Study Information Systems Gunadarma University graduated in 2005. In 2012 obtaining Master of Information Systems Management (MMSI) from Business Information Systems Program Gunadarma University. Today is as permanent staff PSMA Online and as lecturer at Gunadarma University.