

A New Entropy Based Link Optimization Algorithm

Shivankur Thapliyal*, Soni Chamoli**

*(Computer Science and Engineering, Uttarakhand Technical University/Doon Institute of Engineering and Technology, and Rishikesh, Uttarakhand

Email: satshiv07@gmail.com)

** (Computer Science and Engineering, Uttarakhand Technical University/Doon Institute of Engineering and Technology, and Rishikesh, Uttarakhand

Email: satshiv07@gmail.com)

Abstract:

In the preferences of search engine [1] scenarios, the increasing and the large availability of datasets create a very huge concern that how we optimize and stored these vast and large amounts of data sets, generally storage and optimization [6] both plays a very bigger role in search engine optimization [7]. But we focused on the optimization techniques of these data sets that at any time 24 * 7 data availability are present here for the user , but we know data is in very vast form such as sky survey cataloguing data, Genomics data, scientific data, we trained data using various fundamental Data mining Strategies but a one major concern the availability of these data for the user at very short instant of time with very fast and enormous speed but how to achieve this, now the realistic answer of this is to a very strong reliable and sophisticated and adequate optimization techniques[2].

Keywords —SEO Link Optimization, Link Optimization Algorithm, Entropy based Link Optimization Algorithm.

I. INTRODUCTION

Search Engine Optimization [1] are a major challenging area in today’s information age context. A very precise and accurate information are to find in the web space are still a very challenging and tedious tasks. There are various links are to be available for any search query, but to find those links that contains a very precise and significant information are still very typical and most time consuming NP-hard problem. To find precise and accurate information are totally dependent on the page rank [4] of any website, a page which belongs to particular website contains a maximum page rank index, then it’s must contains a very

precise and significant information, but In today’s scenarios page rank [2] dependent on some parameters about that page corresponding to that query such as: the total number of links available on that page, and the maximum times that the total number of words used in that page , which belongs to particular query, and total number of hits on that page, viewing time of that page across world-wide. Web mining [3] are responsible to filter any web page corresponding to that query based on some parameters, but here we proposed a new method of filtering any web page, which provides rapid results and filter most accurate web page corresponding to that query.

II. CONCEPTUALLY THEORY OF ENTROPY BASED LINK OPTIMIZATION ALGORITHM

This algorithms worked with the links (inbound/incoming links, outbound/outgoing links).

- ✓ Worked with the pattern of links as inbound links or outbound links.
- ✓ Very efficient for search engines with respect to processing and computation.
- ✓ Using probabilistic techniques and finding the entropy of each page separately.
- ✓ At the end all pages entropy summation and counts the whole web application (websites entropy).
- ✓ And a website which have a highest entropy rank it first.

III. ENTROPY BASED LINK OPTIMIZATION ALGORITHM

1. Start.
2. Take any websites and count its pages , such as N pages contains any websites.
3. Initial Population of Websites n.
4. Take first page of A1 website from n pages
5. 1 page $\leftarrow A1(N)$.
6. Counting the total links which contains any websites such as TOTLINK.
7. Counting the separately total links for any webpage such as TOTLINKPAGE1.
8. Counting 1 page inbound or incoming link such as INLINK.
9. Find the probability such as $P(\text{INLINK}) = \text{INLINK} / \text{TOTLINKPAGE1}$.
10. Now Counting the entropy of 1st page such as , ENTROPY OF INLINK(1st page)= $\text{LOG}_2(P(\text{INLINK}))$.
11. Similarly , Counting 1 page outbound or outgoing link such as OUTLINK.

12. Find the probability such as $P(\text{INLINK}) = \text{OUTLINK} / \text{TOTLINKPAGE1}$.
13. Now Counting the entropy of 1st page such as , ENTROPY OF OUTLINK(1st page)= $\text{LOG}_2(P(\text{OUTLINK}))$.
14. Now combined entropy of 1st page such as : $(\text{ENTROPY OF INLINK} + \text{ENTROPY OF OUTLINK})/\text{ENTROPY OF TOTLINK}$.
15. Now similarly count N pages entropy of website 1st.
16. At last summation such as Total Website Combined Entropy (WCE) = $\text{Entropy}(\text{Page1})+\text{Entropy}(\text{Page2})+\dots\dots\dots +\text{Entropy}(\text{Page N})$.
17. A Website which have highest WCE selected for ranking and placed at top rank.
18. Exit.

IV. ESSENTIAL MEASUREMENTS

Time Complexity : $O(\log n)$.
Space Complexity : $O(\log n)$.

V. A NUMERIC ANALYSIS OF THIS NEW PROPOSED METHOD

Numeric example:
A data set which are given as follows:
For one website:

For one websites

Pages	Incoming links	Outgoing links	Total Links
Page 1 st	12	16	28
Page 2 nd	13	14	27
Page 3 rd	10	17	27
Page 4 th	16	10	26
Total	51	57	108

Now Numerical computation
Page 1st = $P(\text{INLINK}) = 12/28=0.4285$.
Page 1st = $P(\text{OUTLINK}) = 16/28 = 0.5714$.

ENTROPY(INLINK) = LOG2(0.4285) = -1.22246.

ENTROPY(OUTLINK) = LOG2(0.5714) = -0.807222.

ENTROPY(Page 1st) = ENTROPY(INLINK)+ENTROPY(OUTLINK) / ENTROPY(TOTLINK).

ENTROPY(TOTLINK) = LOG2(108) = 6.75488.

ENTROPY(Page 1st) = (-1.22246+(-0.807222))/6.75488

ENTROPY(Page 1st) = -0.30047.

Similarly we do for remaining 3 pages, And the tabular form are as follows :

Table 2nd : Significant data regarding INLINK and OUTLINK

Pages	P (INLINK)	P (OUTLINK)	ENTROY (INLINK)	ENTROPY (OUTLINK)	ENTROPY (PAGES)
Page 1 st	0.4285	0.5714	-1.22246	-0.807222	-0.30047
Page 2 nd	2.33314	1.2345	2.345	1.346	1.22223
Page 3 rd	1.4567	2.3456	1.2345	1.89655	3.4666
Page 4 th	1.2345	2.3456	1.456	1.5678	2.3456

Page 2 , Page 3 and Page 4 results are predicted because calculations are similar to page 1st. Now at the end

Website Combined Entropy (WCE) = -0.30047+1.22223+3.4666+2.3456 = 6.73396.

VI. RESULT ANALYSIS

Query : 11379 : world war 2				
Links	Google Rank	ELOA entropy	ELO A Rank	Description
https://en.wikipedia.org/wiki/World_War_II	1 st	0.3999	5 th	Under Review
https://www.history.com/topics/world-war-ii/world-war-ii-history	2 nd	0.3789	6 th	Under Review
https://www.history.com/topics/world-war-ii	3 rd	0.4207 4	3 rd	Under Review
https://www.britannica.com/event/World-War-II	4 th	0.4279	2 nd	Under Review
https://www.theatlantic.com/photo/2011/10/world-war-ii-after-the-war/100180/	5 th	0.4114 7	4 th	Under Review
https://www.vox.com/2014/11/13/7148855/40-maps-that-explain-world-war-ii	6 th	0.2961 8	7 th	Under Review

https://www.nationalww2museum.org/students-teachers/student-resources/explore-wwii-history	7 th	0.5035 2	1 st	Under Review
---	-----------------	-------------	-----------------	--------------

Query : government space agencies of global world				
Link	Google Rank	ELOA Entropy	ELOA Rank	Description
https://en.wikipedia.org/wiki/List_of_government_space_agencies	1 st	0.3442 7	5 th	Moderate Informative
https://www.unoosa.org/oosa/en/ourwork/space-agencies.html	2 nd	0.3188 5	6 th	Moderate Informative
https://spacenews.com/op-ed-global-government-space-budgets-continues-multiyear-rebound/	3 rd	0.4752 3	2 nd	More Informative
https://www.forbes.com/sites/quora/2017/05/22/which-space-agencies-are-considered-the-best-in-the-world/#174133c15245	4 th	0.3780 6	4 th	More Informative
https://www.statista.com/statistics/947300/leading-space-agencies-by-government-budget-worldwide/ https://www.statista.com/statistics/947300/leading-space-agencies-by-government-budget-worldwide/	5 th	0.4482	3 rd	More Informative
https://www.nasa.gov/mission_pages/station/cooperation/index.html	6 th	0.0000	9 th	Very Less Informative
https://www.aerospace-technology.com/features/featurethe-worlds-largest-space-agencies-4743900/	7 th	0.5497	1 st	More informative
https://www.weforum.org/agenda/2016/01/which-countries-spend-the-most-on-space-exploration/	8 th	0.2919	7 th	Less Informative
https://www.rferl.org/a/space-agencies-and-their-budgets/29766044.html	9 th	0.2753	8 th	Less Informative
https://www.industry.gov.au/sites/default/files/2019-03/global_space_industry_dynamics_research_paper.pdf	10 th	0.0000 0	10 th	Very Less Informative

VII. FUTURISTIC APPLICATIONS

- ✓ Data Assessment
- ✓ Data Searching
- ✓ Data Tracking
- ✓ Data Fetching
- ✓ Data Pattern Recognition and evaluation.

VIII. ACKNOWLEDGEMENTS

We are very much grateful to all respected professors of DIET Rishikesh for their kind help, lasting encouragement, valuable suggestion throughout the entire period of our project work. We are highly indebted to his astute guidance, sincere support and boosting confidence to make this Research successful.

The acknowledgment will be incomplete if we fail to express our obligation and reverence to our family members and friends whose moral support is great factor in doing this research.

IX. FUNDING

This study was not funded by any other profitable or non-profitable organisations.

X. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

XI. CONCLUSION:

This new entropy based link optimization algorithm gives a very satisfactory results and opt those page or to rank those page which contains most precise and significant information corresponding to that query, and it's fetching time also very enormous and fast, because it's time and space complexities are very optimum in nature. This algorithm provides very best and accurate results compare to google page rank algorithm. This algorithm utilizes all

links and predict an appropriate results based on that links. There are various traditional approach based page rank algorithms which uses some traditional or orthodox patters to find out the significance of that page, which is also a most time consuming in nature, but using this link optimization algorithm, at a very small slice of time, we find out a better and appropriate results, which contains most significant information corresponding to that query.

XII. REFERENCES

- [1]. Beel, Jöran and Gipp, Bela and Wilde, Erik (2010). "Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co." . Journal of Scholarly Publishing. pp. 176–190. Retrieved April 18, 2010.
- [2]. Brian Pinkerton. "Finding What People Want: Experiences with the WebCrawler" (PDF). The Second International WWW Conference Chicago, USA, October 17–20, 1994. Retrieved May 7, 2007.
- [3]. Danny Sullivan (June 14, 2004). "Who Invented the Term "Search Engine Optimization"?". Search Engine Watch. Retrieved May 14, 2007. See Google groups thread.
- [4]. "Search King, Inc. v. Google Technology, Inc., CIV-02- 1457-M" (PDF). docstoc. com. May 27, 2003. Retrieved May 23, 2008.
- [5].Stefanie Olsen (May 30, 2003). "Judge dismisses suit against Google". CNET. Retrieved May 10, 2007.
- [6]. "Technology & Marketing Law Blog: KinderStart v. Google Dismissed—With Sanctions Against KinderStart's Counsel". blog.ericgoldman.org. Retrieved June 23, 2008.
- [7]. "Technology & Marketing Law Blog: Google Sued Over Rankings—

KinderStart.com v. Google".
blog.ericgoldman.org. Retrieved June 23,
2008.