Classification of search engine optimization techniques: A data mining approach

Nitin Duklan¹, Diwakar², Dr. Himanshu Bahuguna³

¹Student, J B Institute of Technology, Dehradun

²Assistant Professor, Department of Computer Science and Engineering, J B Institute of Technology, Dehradun

³Associate Professor Department of Computer Science and Application, Uttaranchal University, Dehradun (UK)

¹nitinduklan@gmail.com ²sdkmourya01@gmail.com

³himanshu1081@gmail.com

Abstract— Data mining is the application of sophisticated analysis to large amounts of data in order to discover new knowledge in the form of patterns, trends, and associations. For finding suitable information Search engines play an important role in retrieval of information on the web as they had become the entry point for accessing the web. Search engines analyse various aspects of the web page like its content and other attributes and display them accordingly .So it is important for the webmasters to develop and create those webpage's that fulfill the requirements of search engine but there is no such standard for ranking the webpage as it vary from one search engine to another. So webmasters uses different search engine optimization techniques like Keyword selection, Directory submission, Social bookmarking, Target market, Content, Keyword density in contents etc. to promote their webpage's.

This paper focuses in analysing different search engine optimization techniques and finding those techniques that makes maximum impact in the ranking of the web page for this purpose researcher had used k-means cluster analysis for clustering various SEOT.

Keywords— Search engine optimization, Data mining, K-means cluster analysis, weka

I. INTRODUCTION

The growth of the Internet, its usage and dependency, leads towards various challenges. The Internet has opened up vast possibilities by opening the doors to Data control and access. It allows users to share their information and data through social networking site or simply by creating some WebPages using different languages and technology.

Search engines have a unique policy for indexing information in an efficient manner, and it is essential to optimize web-pages in a specific way to enhance their search ranking. Search engine optimization is about modifying the webpage accordingly to different parts of the website. These

small modifications when viewed individually might make exponential improvements.

Search engine are used to fetch the information through the World Wide Web and thus many challenges arises in the working of search engine. Search engine are used to fetch the list of websites with respect to the keywords query inputted by the user .Webmasters Optimizes their websites so as to get best position in the SERP. They use various optimization techniques to improve their rank in Search engine. Thus a need was found to analyse various search engine optimization techniques.

Various researchers has worked on the subject and founded various Search engine optimization techniques to improve the rank of a website in search engine result page .In the past decade SEOT has drawn attention of webmasters since the use of search engine is preferred by the users.

So there is a need of proper classification of Search Engine Optimization Techniques (SEOT) that will help to improve the knowledge base and further leads to increase the efficiency of search engine.

The purpose of study is to give guidelines and a better understanding to the persons engaged in the field of Search engine optimization and web developers that how to implement search engine optimization practices so as to improve their website ranking.

Search engine optimization practices have grown increasingly important and this study suggest

College of Computing Sciences & Information Technology (CCSIT), Teerthanker Mahaveer University, Moradabad, India

some techniques that have to be focused while promoting a website or a webpage.

Data mining - also known as knowledge-discovery in databases (KDD) is process of extracting potentially useful information from raw data. A software engine can scan large amounts of data and automatically report interesting patterns without requiring human intervention. Other

Author details must not show any professional title (e.g. knowledge discovery technologies are Statistical Analysis, OLAP, Data Visualization, and Ad hoc queries.

Unlike these technologies, data mining do not require a human to ask specific questions. Partitioning a set of objects into homogeneous clusters is a fundamental operation in data mining. The operation is needed in a number of data mining tasks, such as unsupervised classification and data summation, as well as segmentation of large heterogeneous data sets into smaller homogeneous subsets that can be easily managed, separately modelled and analysed. Clustering is a popular approach used to implement this operation.

Clustering methods partition a set of objects into clusters such that objects in the same cluster are more similar to each other than objects in different clusters according to some defined criteria.

II. OVERVIEW OF SEARCH ENGINE Optimization Techniques

Search Engine Optimization (SEO) is the process of improving visibility of a website in search engine's results page (SERP) in response to a keyword query. Search Engine Optimization technique is basically used to improve organic search results rather paid results. This paper focuses on optimizing the ranking outcome of the organic result in search engines through the use of important optimization techniques.

Search engine optimization can be broadly categorized into two different categories:

- 1. On-page optimization
- 2. Off-Page optimization

Some popularly used Search engine optimization

techniques

- Link Exchanges
- Meta tags
- Directory Submission
- Website tracking
- W3c standards for creating website
- Xml files
- Social bookmarking submission

III. AIMS AND OBJECTIVES

The Main objective of the paper is to classify the search engine optimization techniques.

Further sub objectives include:

• Perform a literature review to get an overview of what has been done in the area of SEO.

• To analyse various Search engine optimization techniques practiced by the software professionals.

• To study the classification techniques used in SEOT for improving Page Rank.

• To classify the SEOT from users perspectives applying Data mining approach

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

IV. LITERATURE REVIEW

As popularity of WWW increases incrementally, millions of people use various search engines to discover information for the various web servers. But majority of users are interested only in few top listed result pages. Here comes the role of Search Engine optimization and hence promoting a website in search engine result page is a major task in website development and maintenance. Website ranking in search result strongly depends on how Search engine optimization (SEO) is implemented [1].

Search engine becomes an integral part of everyone's life to search information. The users rely on search engines to provide us right information at right time. To satisfy users need search engine must find and filter most relevant information matching a user query and display that information to the user [2].

Search Engine Optimization (SEO) is a process/activity that relates with of optimizing websites/web-pages to achieve higher raking in the SERP. Major search engines rank individual web-pages or websites based on certain factors. The focused methodology used in Search engine optimization is to update both content and associated coding of the website/webpage to improve its visibility and ranking in organic searches made by the search engines[3].

Search engine optimization increases the ranking of search results in the Internet marketing and they elaborated in their research that the rank of the motel sites and its bandwidth increased for Internet marketing after the implementation SEO techniques. The authors used several techniques of SEO to increase the bandwidth and ranking of search results including text title, label text, picture note, and HTML modification [4].

Search Engines are inefficient if they are not able to find relevant information, to make search engines efficient SEO practitioners should use optimization of web pages as well as algorithm and search engine. Optimization is the process of improving the performance with the focus on time and accuracy. There are many techniques that are been proposed for optimizing the search engines. In this paper, researcher considered the best techniques existed with certain limitations and now overcoming them with suitable solutions using parallel processing and k -means clustering [5]. In particular, they presented the importance of parallel processing the vector space algorithm for document indexing and incremental updating of means in kmeans clustering algorithm in the real world scenario.

Cluster analysis is an important data mining technique used to find data segmentation and pattern information. By clustering the data, people can obtain the data distribution, observe the character of each cluster, and make further study on particular clusters. In addition, cluster analysis usually acts as the pre-processing of other data mining operations. Therefore, cluster analysis has become a very active research topic in data mining [6].

V. PROPOSED TECHNIQUES FOR CLUSTERING

The proposed approach is divided into two sections in section one the data collection is done by different software professionals and then the sub processes like data cleaning is done. This process involves removing of irrelevant data which can be termed as noise, thereafter missing and this irrelevant data is formatted according to the required format and a database or data warehouse is evolved. Second section generates the clusters according to the choices of the attributes. In this paper we had used k-means cluster analysis for creating the clusters [7].



Figure 1: Displaying Clustering Process

VI. K-MEANS CLUSTER ANALYSIS

The initial step of k-means clustering is quite simple. It is based on minimum sum of squares to assign observations in a cluster .In the starting, we determine number of cluster which is represented as K and we assume the centroid of these clusters.

College of Computing Sciences & Information Technology (CCSIT), Teerthanker Mahaveer University, Moradabad, India

We can take any random objects as the initial centroids or the first K objects can also serve as the initial centroids.

Then the K means algorithm will do the three steps below until convergence. Iterate until stable or no object move group:

- 1. Determine the centroid coordinate
- 2. Determine the distance of each object to the Centroids.

$$d_{ik} = \sqrt{\sum_{j=1}^{p} (x_{ij} - \bar{x}_{kj})^2}$$

3. Group the object based on minimum distance (find the closest centroid)



Figure 15: Showing clustering technique

I. ANALYSIS OF STUDY

During the study different search optimization techniques were studied and analysed, further a popular data mining technique (Cluster Analysis) is used to analyse different search engine optimization techniques so as to identify some crucial SEOT that will be helpful for the webmaster

In this paper k-means cluster analysis is used to create the clusters using a well known tool Weka 3.6. Weka provides us facilities so as to implement different data mining algorithms in very easy way.

Cluster analysis a techniques that is used to classify cases into small individual groups that are homogeneous within themselves and heterogeneous among each other, on the basis of their similar attributes and this classification is known as clusters [8].

The clusters were created using partitioning or non-hierarchical clustering method i.e. K-means cluster analysis. In this method we are free to make number of cluster to be formed in the sample but as stated earlier in the research that by simply testing different parameter settings depending on the research data set and expected use of the results in our research we had created 6 clusters using weka 3.6 that are shown in the **Table No. 1**.All the missing values are adjusted by the algorithm automatically so as to get the expected results.

Following are the obtained from the study:

Number of Cases in each Cluster						
	0	15%				
Cluster	1	19%				
	2	6%				
	3	7%				
	4	38%				
	5	15%				

Table 1: Showing the Clusters and Number of case in each cluster

ne al les	1 24 5- Snpel Vers								- 5
nen Gent, Arm sonnen isenanteren feinte									
T Contraction of the second	- mini alt embration at test sp	ii Jii.							
Simplifying the Common Department of the of 1911 Str.	2018 - 1000 D 6 (0.1917 - 1017								
	bb:::								
114 Enedikin									
	further of therations: S	future of constants S							
Schemenses of schemers freedowing of 7 of these room for international strength of 111 of 11	allefart, stratter ser af apparet erner	: L1.1886	SETURE.						
Alation wet	discing values globally replaced at	e merende							
WE ATACH: 105									
Artilates 2	Cluster certraides								
Jarget dasles			Crent						
Desit see statut	ACARDONE .	5:11 Eth	(100	2	4	4	5	
Begenet generations		1.570	15	135	(10)		(23)	(24)	
United									
Sussare_I_vésile	Terra libeara	2.5:55	1.8%	2.0404	2.5	3.77%	1, 1617	1.0.7	
Nu stankuls jurguestin, velatie	Entrie the stiniter	24.5	2.7		13	3.4575	1.18	2.15	
ne Beseris in attent	Daniel miret at	1.118	1,1675		21	2,8190	1.7.0	1.093	
n n n n n n n n n n n n n n n n n n n	20070	1.4.15	2.35	1 1917	-	1 85.0	1.30	1,5212	
itter vedale ripph	Dimension of scholars	1.00		1 1324		3 2424	2.11	1 (50)	
Note cape	The eventues day sectors warran	1.10		1.7.7		1.2.15	1 1017	1 01 7	
Predict Data	Transfer to sender to sender	1.025	3.5.95	· · · ·				1 (15)	
mele ti des euro i la la	a set of the set of th		1.16					1.000	
Free Gl	1. 34	2.2.4	1.17	100			S INST	27.36	
24 arm	CLASS IN	1. 1.	1.1.8	A200		1.13		2.1985	
En millo	40.000	2,4(2)	1.12	2.281	1	1.2.2	5,001	2.1925	
Tort Endporter	and de	2. 24	1.65		14	1.111	1.14	2.7.85	
Pressure Powentier	and title and alt tage	37.37	1.1.5	2.440	1	7.4.4	1000		
Part of Bertlerier	OF THET	3.5 94	2.2.5	2.340	1	3.60	. 4.2		
Service Minuted and State	trait: thir	2,621	1.1.5	2.440	2.1	3	1.10	2.017	
Servi bedreiter nier min	£		2.82	2.500	1.5	7.44	. 8.5	2.0412	
Wr. Ste. Indext	Hot means	2,623	1.6.1	1.450	-	1.2.5	12	2.(%)	
(acCronni	provinciality	2.040	7.01	1.000	1	1.7.7	. 411	3.1 0 1	
Text modercplict 600 truin, semaining text	Directory Schrowich	2.074	2.62	1. H	ы	1.855	.12	2.31	
	Aitri+ "hrein	2,5,01	1.9.2	1. 80	:	1.41	. 12	2,407	
Booel and evaluation on tradinity set	Stored lettertioning Street	2.004	1.92	1.13 5	1	7.115		2.11	
	increal promenting sources or	3.0.25	7.5	1.39	1.1	2.77	. 40	2.697	
	es fate Treorag		1.61	5. 14	1.4	1.77	1.1.1	2.381	
	100110-010	175	r	1	115	p	***	112	
re 2: Showing the run information on different									





Figure 3: Showing different clusters on test data





Figure 5: Visualization of final clust

From the obtained results it is possible to find some SEOT for improving rank of the webpage in SERP .Further it can be stated that best search engine optimization techniques falls in cluster 5 as these techniques may improve the rank of the webpage. While cluster 0,1 and 3 can be classified as Negative value .

During the study following results are obtained:

- 1.) Cluster 5 which represents about 15% focuses on xml files and image tags as an important aspect also they give importance to keyword selection, structure of website ,inter webpage linking, meta tags, and of the opinion that by giving importance to these techniques and other similar techniques to improve the ranking of the webpage.
- 2.) Cluster 4 which also represents about 38 % of population give importance to xml files, keyword selection, website tracking, blog creation, Meta tags, directory submission, target market as most relevant and can help improve rank
- 3.) Cluster 1 which represent about 19 % gives least importance to link exchange, directory submission, article submission and are not sure about the techniques used to optimize a webpage.

In this study researchers also find certain limitations of SEOT like use of these techniques is not able to assume the exact Rank also if these techniques are used in excess there in a possibility that site would be Penalized .It was also found that ranking of website changes with the use of different search engines.

VII. SUGGESTIONS

It was found during the study that some areas of Search engine optimization techniques could be analysed further. Some of the related topics are:

- 1.) Different Data mining algorithm for clustering
- 2.) New and trending Search engine optimization techniques
- 3.) Strategies for Search engine optimization. CONCLUSION

Data mining is implemented for finding some useful facts and patterns from different data sources. Use of data mining technique can help to understand and analyse data and information in proper way so that is will be helpful in different sectors. Search engine optimization is also a sector where there is a requirement of identifying some selective search engine optimization techniques from various SEOT so as to achieve a better rank in search engine result page.

This study uses k-means clustering technique for clustering. The final results demonstrate that the

proposed approach revealed those techniques that can help webmasters to achieve better rank in search engine result page.

The purpose of study is to give guidelines and a better understanding to the persons engaged in the field of Search engine optimization and web developers that how to implement search engine optimization practices so as to improve their website ranking.

I. REFERENCES

- Patil Swati, Pawar B.V., Patil Ajay S, "Search Engine Optimization: A Study", Research Journal of Computer and Information Technology Sciences Vol. 1(1), 10-13, February (2013)
- [2] Cho J. And Roy S., "Impact of search engines on page popularity", Proc. 13th International conference on World Wide Web, 20-29 (2004)
- [3] Rehman Khalil ur and Khan Ahmed Naeem Muhammad, "The Foremost Guidelines for Achieving Higher Ranking in Search Results through Search Engine Optimization", International Journal of Advanced Science and Technology Vol. 52, March, 2013
- [4] H. L. Hsing, H. J. Chen, L. Me. Huang and H. H. Yi, "The application of search engine optimization for internet marketing: An example of the motel web sites", ICCAE Chung HuaUniversity Taiwan, vol. 1, (2010), pp. 380-383.
- [5] Krishna Kiran Kattamuri and RupaChiramdasu,"Search engine with parallel processing and incremental k-means for fast search and retrieval", International Journal of Advances in Engineering & Technology (Jan 2013),ISSN: 2231-1963
- [6] Sharma Depika and Mehla Sandeep," A Study of Data Mining Clustering Techniques", International Journal of Advanced Research inComputer Science and Software Engineering) Volume 4, Issue 3, March 2014, ISSN: 2277 128X
- [7] http://www.originlab.com/doc/Origin-Help/KCA-Algorithm
- [8] https://en.wikipedia.org/wiki/Cluster_analysis