

---

# Enhancing Knowledge and Bounce Rate in SERPs Using Micro-Data

MUHAMMAD IKRAMULLAH LALI\*, KINZA MEHR\*, WAQAR ASLAM\*\*, RAZA UL MUSTAFA\*\*\*,  
AND UMAIR ALI\*

RECEIVED ON 28.11.2016 ACCEPTED ON 21.02.2017

## ABSTRACT

Internet has revolutionized the human life. SEs (Search Engines) are one of the major tools being used for finding information over the Internet. SEs enlist the information into links as per relevance to the searched query. A searcher usually visits the top web links retrieved on SERPs (Search Engine Results Pages) in response to a search query. With the evolving nature of Internet and the increasing number of competitors; it is hard to maintain high ranking in SERPs even for professional correspondents. However, correspondents can apply the techniques of web micro-data to achieve high CTR (Click through Rate) in SERPs. Ranking in major SEs is still a critical factor, although in certain cases such as movies, books, recipes rich snippets proved profitable for webmasters. This study aims to address the gap in micro-data moving from top category such as *Animals* to their limited scope. Animals with information such as name, price, category will have high CTR and hence more user satisfaction for specified result will lead to high ranking in SERPs.

**Key Words:** Search Engine Optimization, Visibility Enhancement, Use of Meta-Data, Semantic Searching, Search Engine Results Pages.

## 1. INTRODUCTION

Internet searching has now become a routine activity of many people around the globe. Searches are executed due to customary interest, originated through a live event, news, social report or new scientific discovery [1,2]. To extract related data set from the Internet, SE queries are applied. Searched query results have revealed that SE is a primary element in extracting the true information. For example people search for facts concerned to health on SE for numerous reasons. Queries linked to recognition of infectious diseases are requested through SE that

provides valuable health-related information, might also yield useful health statistics [3].

Lali et. al. [3] show that epidemics trend can be detected with information available on Web. They used SE queries and twitter tweets to detect influenza epidemics in a particular geographical area. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza. Eysenbach [4] aimed to determine the prevalence of

---

Corresponding Author (E-Mail: drikramullah@uos.edu.pk)

\* Department of Computer Science & Information Technology, University of Sargodha.

\*\* Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur.

\*\*\* Department of Information Science, COMSATS Institute of Information Technology, Sahiwal.

health-related searches on the web by analyzing search terms entered by people into popular SEs. They also made some preliminary attempts in qualitatively describing and classifying these searches. Occasional difficulties in determining what constitutes a health-related search led us to propose and validate a simple method to automatically classify a search string as health-related.

Information can be extracted from the different SEs in many ways, either by typing a topic name, phrase, keyword or an appropriate question. Most popular SEs are Google, Yahoo, Bing and Yandex which fulfill the searching requirements [5]. Since the last few decades, SE developers are working to provide with more accurate, reliable and relevant information. Keeping this in mind, web masters and bloggers should acquire the skills to boost the designing and learn marketing tactics to attract the visitors to their website. Most Internet searching is performed after office timing or majorly on weekends [6]. According to iprospect more than 60% users chosen to hit the top results appeared on the initial page of SERPs. Only 5% searchers go beyond 2<sup>nd</sup> page and just on 2% visited beyond 3<sup>rd</sup> Page [7].

Hence, it is important to maintain top position in SERPs for continuous flow of websites visitors to run a business through it. In modern era, where business grooming techniques like online trade, marketing and advertisement have a great impact to increase the revenue from the business [8]. It has been recognized as a better way to generate revenue by advertising via websites. Many professionals are working to develop number of websites for advertising businesses. Each website is developed in a particular language and style influenced by its developer/user's geographical location and civilization. However, it develops a sense of competition among online businesses that attracts the most visitors to the website.

SE optimization is the technique of optimizing a website to gain a distinguished ranking in SERPS. Initially its idea

was proposed by Sullivan and is progressing since then. SEM (Search Engine Marketing) provides techniques to increase the ranking of a website [9].

Therefore, different SEs use various searching algorithms. Every SE has adopted its own ranking criteria to present the specific result higher in the SERPs. With the passage of time, the IT industry has become used to these techniques. Meanwhile, the developers have acquired the knowledge to improve the algorithms. Hence, these algorithms improved the effectiveness of the results regarding the relevant websites ranking.

However, the continuous improvements of these algorithms were not properly communicated to the industry for verifying their practical implementation. Therefore, one can not rely on these but can focus on their basic context, available material and the way it contains information to address his search queries.

SEs intend to give more importance to those pages where relevant information is present. For this, SEs such as Google, Yahoo, Microsoft, Pinterest, Yandex has introduced structured data on the Internet [10,11].

- (1) Articles
- (2) Book Reviews
- (3) Events
- (4) Local Business
- (5) Movies
- (6) Products
- (7) Restaurants
- (8) Software Applications
- (9) TV Episodes
- (10) TV Episodes with Ratings

This article focused to find out the gap in the existing micro-data, to give more organic results to users. These rich snippets

are not only used for ranking signals, but they will be used in future for ranking in SERPs [12]. Section 2 presents the related work and important approaches and techniques. Section 3 discusses the methodology and section 4 concludes the article.

## 2. RELATED WORK

Most people use SEs for finding the relevant material and information via Internet services. The information technology has become the basic need for survival in the modern age. As according to comSource, it's over than 12 billion searches are done every month from January 2009, almost 400 billion searches are accomplished each day [13]. Another research by Google Search Statistics, approximately 4 billion searches were performed every day [14]. SEs work systematically, which crawls information from various sources on Internet. Each time user request for information by simply typing the query, SE responds on it by providing the relevant information. While entering a search query mostly SEs help the users with the topic prediction to assist him [15].

SEs are mainly comprised of four components; crawler, indexer, database and GUI (Graphical User Interface). An exceptional SE should have four features; quick, correct, easy and strong. SE optimization is a broader term which covers a huge area on the Web. When a user inputs a query in the form of keywords on the SE, the results listed are organic in nature. Excluding those results having the adverts sponsored links (pay-per-click ads) [16].

Sergey Brine and Larry Page designed a link analysis algorithm known as "PageRank", which is used by SEs. The PageRank values are pre-calculated and stored for all pages acknowledged to IR system. Hence, every web page has its own rank free from the query terms. The page rank score shows the importance of different web pages having the same query terms [17,18]. The maximum value which is given

by the SE e.g. Google toolbar is 10. The method to calculate the page rank is shown in Equation (1) [19].

$$PR(B)=(1-d)+d(PR(S1)/C(S1))+\dots+PR(Sn)/C(Sn) \quad (1)$$

A website could only be visible and optimized when it is ranked accordingly and positioned among the first page of the search results. SEO (Search Engine Optimization) is a method and set of concepts, tools and procedures applied to rank websites and place them at first page in the organic search results. The various SEO techniques that can be used, are discussed in the following sections along with their effects.

### 2.1 On-Site Optimization

This is also referred as on-page optimization, performed directly on the page with the code of that website [20]. It is further divided into a number of techniques such as code, content and link/directory structure.

### 2.2 Page Title

The most important and favorable factor for the SE results is the relevancy of title of web pages with the title tags. The position of the title tag in the code section is at the <head> tag section of your web coding document and that is a key way of utilizing meta information for increasing relevancy and rank of the website [21]. The topic and other major keywords are also shown with the help of this tag. That is the result of combining together the string and the words, including in the title of HTML (Hyper Text Markup Language) document.

### 2.3 Description Tag

This tag contains HTML attributes that provide a useful explanation of the contents in a web page. The SE results in the (SERP) contain this information right after the title of the web page. In some cases, without opening the entire website, it can be a great way of providing information about the contents of the website for increasing its traffic and marketing [22].

## 2.4 Meta Keywords Tag

The words which are the most relevant and directly related to the content of website title are included in the Keywords Tag along with the body contents. Embedding the most important and relevant keywords first can positively improve the SERP [23].

## 2.5 Using Sitemaps

All the files and URL's used in a website are saved in a list with the help of sitemaps like ISTIMAGE.JPEG, FILE.PDF. The technique of how SE prioritizes the web pages and how they crawl the website is accomplished with the help of sitemaps. It also offers robots to efficiently crawl the website [24].

## 2.6 Directory/URL Structure

This structure is simple, logically linked with the contents of the website and also easily understandable by the user for efficiency purposes. For example a URL like <http://www.animalshleterinpakistan.com/freeservices>; helps you to decide but a URL like <http://www.yhissite.com/index.php> etc. leads towards confusion. Most importantly use punctuation in URLs [25].

All the relative techniques which have been discussed for the SEO focused and enlightened the ways of getting the web page higher position in the SE results. However, they did not emphasis on the approach; how the related content which best matches the query has to be shown in the proper position. SEs retain limited knowledge about the web page and its contents but there may be other inherent meanings inside every page which user understands when they read it.

Fig. 1 shows the results of a searched query phrase “Jaguar” in the SE Google. The search results, in response to the keyword “Jagur” were animal’s pictures and its relevant information. In actual this was a “Jaguar USA Car company”, so there were inherent meanings inside the website that could not be understood by the SE.

Therefore, it is important to address this problem by enhancing the web visibility using its internal structure and information stored in it. This article is focused on how microdata can solve this problem and motivated to find out the gap in the existing microdata to optimize the performance. By identifying the gap, we have participated in the international community as a standard platform. The key objective is to attain the knowledge to get the maximum visibility in SERPs with the help of internal structure and information already present inside of the

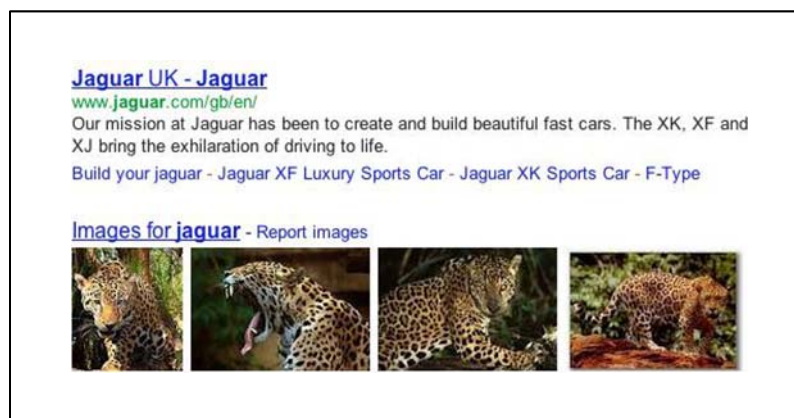


FIG. 1. QUERY SEARCHED IN GOOGLE FOR INFORMATION RELATED TO JAGUAR

webpage. Maximum CTR will boost ranking. Both CTR and Ranking are codependent variables. There is obviously a relationship between these two. If you beat average CTR for a given position by 20%, you're likely to appear in position 1.

### 3. METHODOLOGY

We use analytic approach for our study by using data along with the information that is already available to perform a thorough quantitative analysis. This aims at finding an immediate problem faced by the society or industrial/business organization. We have proposed some new micro-data which does not exist in the schema.org dictionary [26].

The collaborative effort of the three widely used SEs, Google, Bing and Yahoo has created online data repository. The repository has all the major entities and their lot of subcategories. Whenever the web developers and other online business developers need to develop their websites can directly tag all the data using this dictionary. This online combined dictionary is available in the form of Schema.org. More than 10 million websites use this schema.org for marking up their web pages [26].

The vocabularies are maintained and managed by open community groups under the platform of schema.org. The different communities discuss this emerging technique of enhancing the visibility of web pages. Discussions lead towards the addition of new vocabularies in the already published list. These discussions occur by two methods.

- (a) Using the mailing list public-schema@w3.org
- (b) Through github

This online vocabulary provides a shared common platform through which the various web developers and entrepreneurs can markup their web pages. These markups will lead to a better understanding of the web contents inside the web page for the major SEs during their search for a particular query or phrase [26].

The information in the web page is used to create the Rich Snippets by major SEs which produce information in the search results for the user. For example *lasagna recipe* has 4.8 rating, and 9,542 reviews. Which is clearly indicator that the information present on this page is liked by users as shown in Fig. 2.

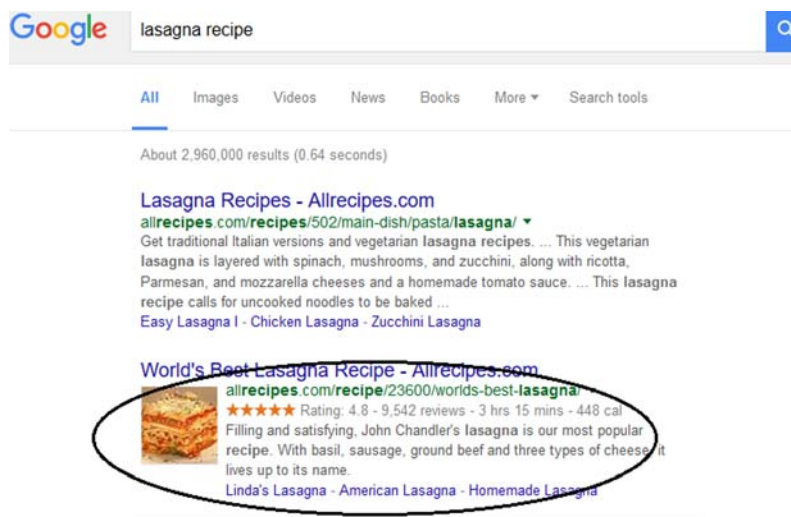


FIG. 2. SEARCH RESULTS OF LASAGNA RECIPE

These search results help the user to get the most relevant results so the CTR will be higher. Higher CTR is directly proportional to the revenue and user interests on the page [27]. There are three techniques for implementing this technique for getting the underlying meanings.

- (i) Microdata
- (ii) RDFa (Resource Description Framework in Attributes)
- (iii) JSON-LD (JavaScript Object Notation for Linked Data)

The syntax for using the major tags and their distribution within their subtypes is listed in Fig. 3. Starting from the tags having information about the item is enclosed inside Fig. 3. By implementing the “itemtype” for a recipe, it is linked up with the schema.org vocabulary. Furthermore,

in every inside tag of “itemprop” the property of the item type is included. This results in directly linking that specified item type with its specific category shown in Fig. 3.

There are lot of specific fields and the global scope of the different tags. They could be used by directly joining their major scope and then moving down towards the limited scope, hence enclosing it within their markups [22]. Starting from the most common type and moving towards the local business, the most limited scope of the types [26]. For example, the first search result listed in Fig. 4 contains both a star rating and a publication date. Both of these can be added using schema micro-data which is coded in Fig. 5.

Adding Schema markup to HTML tags, improves the way your page displays in SERPs by enhancing the rich snippets which are displayed beneath the page title.

```
<div itemscope itemtype="http://schema.org/Recipe">
  <span itemprop="name"> Bread</span>
</div>
Another example
<div itemscope itemtype="http://schema.org/Movie">
  <span itemprop="name">Actor</span>
</div>
```

FIG. 3. EXAMPLE OF MICRO-DATA FOR RECIPE AND MOVIE



FIG. 4. STAR RATING AND REVIEWS GIVEN BY USERS

```
<div itemprop="aggregateRating" itemscope
itemtype="http://schema.org/AggregateRating">
  <span itemprop="ratingValue"> [Aggregate rating given] </span> stars
  <span itemprop="reviewCount"> [Number of reviews] </span> reviews
</div>
```

FIG. 5. CODE FOR THE PLACEMENT OF RATINGS AND REVIEWS IN HTML

The schema.org is an online vocabulary available for marking up the tags but it is still under developed. The vocabulary of schema.org is enhanced by the contributions from the open community who are discussing the new entities that have not been included yet. We have contributed in it, by signing up as a member at the date of 21<sup>st</sup> May, 2015 and then performed a detail study on all the available vocabularies to find the gaps. We discovered that there are many parts still missing in available online dictionary for the webmasters. It restrain the benefits, due the maximum unavailability of the item, their types and their properties. We observed that there was a major category missing relating the “Animal terminology” despite of the fact that there are thousands of search queries against this phrase and keywords appear daily. Some of queries searched in year 2015 are listed in Table 1. We have collected this information using Google Adwords Keyword Planner [28].

There should be a better way to handle this “Animal terminology” through micro-data. The number of searches on the web master tool and the frequency of animal related queries have increased, therefore, it is

important to address this with the help of micro-data format especially recommended by Google.

Proposal for the category of animals recommended on the date 24<sup>th</sup> August, 2015 on GitHub shown in Fig. 6, which is an open community discussion group for the acceptance and releasing of new vocabularies in existing vocabularies.

Our proposal was considered initially for discussion by the different active members. These members include from different interest groups and fields of sub emerging types of web engineering throughout the out world listed in Fig. 7.

Further detailed about this open-ended community could be acquired with the live web address at: <https://github.com/schemaorg/schemaorg/issues/746>. Complete information and opinions received by different people is available at this link.

The detailed discussion with the different members of the Git Hub community on the proposed topic of “Animal Terminology” also guided us to cover identified gap. The resulting acceptance of the proposal is shown in Fig. 8.

TABLE 1. ANIMALS RELATED QUERIES SEARCHED IN GOOGLE FROM JANUARY 2015 TO DECEMBER 2015

Search Query	Avg. Monthly Searches (Exact Match Only)	Search Query	Avg. Monthly Searches (Exact Match Only)
Animals	1-10M	information of animals	100-1K
wild animals	100-1M	list of wild animals	1-10K
wild animals information	1-10K	information on animals	100-1K
cruelty to animals	1-10K	a to z animals	1-10K
rare animals	10-100K	wild animals name	1-10K
wild animals list	1-10K	nature animals	1-10K

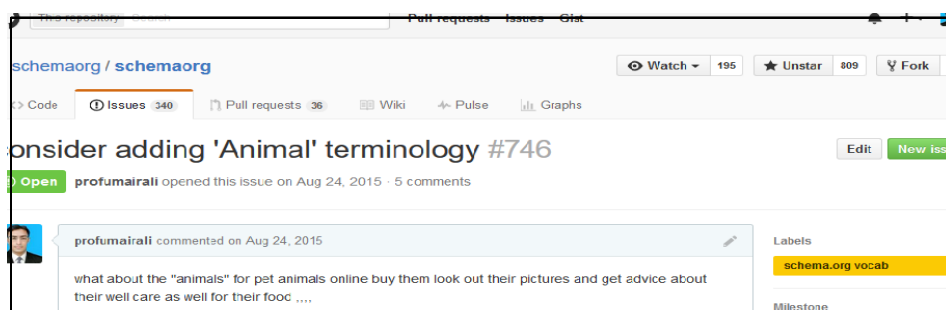


FIG. 6. ISSUE# 746 FOR CONSIDERING ANIMAL TERMINOLOGY

Our identified “Animal Terminology” proposal has been accepted. Proposed micro-data tags hierarchy is shown in Fig. 9.

Animals have been classified in to their respective categories. List of supposed micro-data for animals is listed in Table 2.

TABLE 2. SUPPOSED MICRODATA PROPERTIES FOR ANIMALS

```
<div itemscope itemtype="http://schema.org/Animals">
  <span itemprop="name">Panda</span>
  <span itemprop="price">$19.95</span>
  <span itemprop="type">Vertebrates / Invertebrates</span>
  <span itemprop="Category">Fist / Birds / Worms etc</span>
  <span itemprop="description">One of the unique animal on
    earth</span>
</div>
```



FIG. 7. PARTICIPATION OF OTHER CONTRIBUTORS

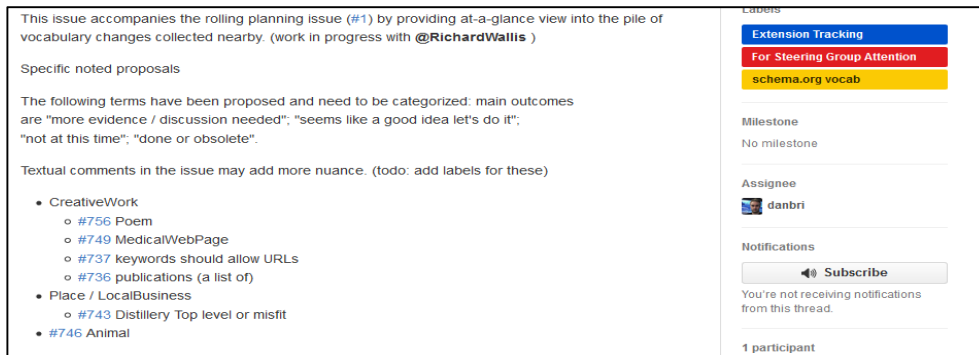


FIG. 8. ACCEPTANCE OF PROPOSAL #746

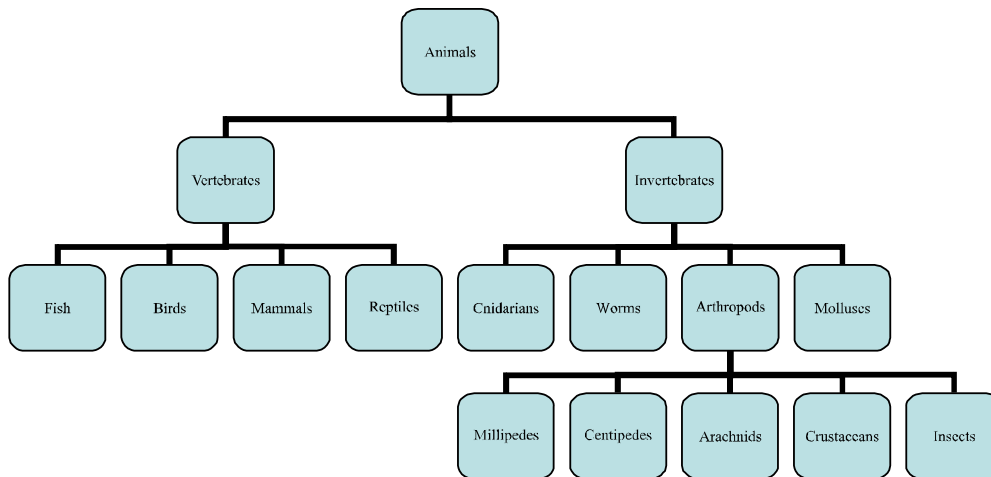


FIG. 9. CLASSIFICATIONS OF ANIMALS AND PROPOSED MICRO-DATA ACCORDING TO THEIR CLASS



## 4. CONCLUSION

In recent years, the SERP's have been augmented with new markup elements that introduce seamlessly additional semantic information. Nowadays, good content is king and also leads to more likes and shares, tweets and as a result returns visitors to your website. Therefore, you have to do some extra things to maximize the SEO friendliness of your post, but most important is that user likes your content. The user must get what he is looking for. Ranking is still an important factor, but introducing micro-data from top to down regarding all necessary information a casual user is looking, will enhance CTR and definitely ranking. In the future we are interested to show some region oriented techniques to rank local in business in local SERPs.

## ACKNOWLEDGEMENT

The authors are very grateful to the University of Sargodha, The Islamia University of Bahawalpur, and COMSATS Institute of Information Technology, Sahiwal, Pakistan, for providing their resources to carry out this research.

## REFERENCES

- [1] Signorini, A., Segre, A.M. and Polgreen, P.M., "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the US during the Influenza A H1N1 Pandemic", PLoS One, Volume 6, No. 5, pp. 1-10, May, 2011.
- [2] Lali, M.I., Mustafa, R.U., Saleem, K., Nawaz, M.S., Zia, T., and Shahzad, B., "Finding Healthcare Issues with Search Engine Queries and Social Network Data", International Journal on Semantic Web and Information Systems, Volume 13, No. 1, pp. 48-62, January, 2017.
- [3] Mustafa, R.U., Nawaz, M.S., and Lali, M.I., "Search Engine Optimization Techniques to Get High Score in SERP's Using Recommended Guidelines", Science International (Lahore), Volume 27, No. 6, pp. 5079-5086, December, 2015.
- [4] Eysenbach, G., and Kohler, C., "What is the Prevalence of Health-Related Searches on the World Wide Web? Qualitative and Quantitative Analysis of Search Engine Queries on the Internet", Proceedings of AMIA Annual Symposium, pp. 225-229, 2003.
- [5] Zhang, S., and Cabage, N., "Does SEO Matter? Increasing Classroom Blog Visibility through Search Engine Optimization", Proceedings of IEEE 46<sup>th</sup> Hawaii International Conference on System Sciences, pp. 1610-1619, January, 2013.
- [6] Shahzad, B., and Alwagait, E., "Does a Change in Weekend Days Have an Impact on Social Networking Activity?", Journal of Universal Computer Science, Volume 20, No. 15, pp. 2068-2079, 2014.
- [7] Dinesh, C., Rewatkar, P., and Sweta, K., "Search Engine Optimization", Asian Journal of Computing Updates and Trends, Volume 2, No. 1, pp. 38-47, 2011.
- [8] Silverman, G., "Secrets of Word-of-Mouth Marketing: How to Trigger Exponential Sales through Runaway Word of Mouth", AMACOM Div American Management Association, 2011.
- [9] Olbrich, R., and Schultz, C., "Search Engine Marketing and Click Fraud", Department of Business Administration and Economics, Fern University in Hagen, 2008.
- [10] Thelwall, M. "Commercial Web Site Links", Internet Research, Volume 11, No. 2, pp. 114-124, 2001.
- [11] Wilkinson, D., and Thelwall, M., "Search Markets and Search Results: The Case of Bing", Library & Information Science Research, Volume 35, No. 4, pp. 318-325, 2013.
- [12] English Google Webmaster Central office-hours hangout [Online] (Last Accessed: 4<sup>th</sup> January, 2017). <https://www.youtube.com/watch?v=QWL864V1W71>
- [13] Balnaves, M., and Caputi, P., "Introduction to Quantitative Research Methods: An Investigative Approach", Sage, 2001.
- [14] Ochoa, E.D., "Analysis of the Application of Selected Search Engine Optimization (SEO) Techniques and their Effectiveness on Google's Search Ranking Algorithm", Doctoral Dissertation, California State University, Northridge, 2012.

- [15] Daud, A., Shaikh, A.M., and Muhammad, F., "Research Community Mining via Generalized Topic Modeling", *Mehran University Research Journal of Engineering & Technology*, Volume 31, No. 4, pp. 599-612, Jamshoro, Pakistan, October, 2012.
- [16] Gleick, J., "Chaos: Making a New Science (Enhanced Edition)" Open Road Media, 2011.
- [16] Beaudoin, C.E., and Thorson, E., "Social Capital in Rural and Urban Communities: Testing Differences in Media Effects and Models", *Journalism & Mass Communication Quarterly*, Volume 81 No. 2, pp. 378-399, 2004.
- [18] Gifford, M.L. "A Global Review of Cookstove Programs", Master's Thesis, UC-Berkeley, 2010.
- [19] Balay, S., Buschelman, K., Gropp, W.D., Kaushik, D., Knepley, M.G., McInnes, L.C., and Zhang, H., "PETSc web Page 2001", 2004.
- [20] Fleischner, M.H., "SEO Made Simple: Strategies for Dominating the World's Largest Search Engine", *SEO Made Simple*, 2009.
- [21] Abazajian, K.N., Adelman-McCarthy, J.K., Agüeros, M.A., Allam, S.S., Prieto, C.A., and Bailer-Jones, C.A., "The Seventh Data Release of the Sloan Digital Sky Survey", *The Astrophysical Journal Supplement Series*, Volume 182, No. 2, pp. 543. 2009.
- [22] Silva, N., and Aguiar, A., "Web Site Optimization for Search Engines: An Empirical Study", 9th Iberian Conference on Information Systems and Technologies, pp. 1-7, June, 2014.
- [23] Tate, A., "Using goal Structure to Direct Search in a Problem Solver", 1976.
- [24] Blankson, C., Kalafatis, S.P., Cheng, J.M.S., and Hadjicharalambous, C., "Impact of Positioning Strategies on Corporate Performance", *Journal of Advertising Research*, Volume 48 No. 1, pp. 106-122, 2008.
- [25] Smoot, M.E., Ono, K., Ruschinski, J., Wang, P.L., and Ideker, T., "Cytoscape 2.8: New Features for Data Integration and Network Visualization", *Journal of Bioinformatics*, Volume 27, No. 3, pp. 431-432, 2011.
- [26] Web Metadata Schema.org website [Online] (Lasted Accessed: 10<sup>th</sup> April, 2015). Available <https://schema.org/docs/full.html>
- [27] Bidgoli, H., "The Handbook of Technology Management", *Supply Chain Management, Marketing and Advertising, and Global Management*, Volume 2, 2010.
- [28] Mari-Carmen, M., Gavin, F., and Arapakis, I., "Effect of Snippets on User Experience in Web Search", *Proceedings of 16<sup>th</sup> ACM International Conference on Human Computer Interaction*, 2015.