International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2015): 6.391

A New Approach on Restructuring and Reranking of User Search Results by using Feedback Sessions

Priti Alok Mishra[1], Bidisha Roy[2]

[1]Student, M.E. (Computer Engineering), St. Francis Institute of Technology, Mumbai

[2]Associate Professor, St. Francis Institute of Technology, Mumbai

Abstract: ***To enhance the significance of web search engine, it is very essential to meet the user's requirement of results. These user search results can be generated with the help of feedback sessions. The framework in this approach firstly creates Feedback session out of user click-through logs. The click-through data will later be preprocessed for precise information retrieval. The original search results will then be restructured and they would be re ranked according to user's domain of interest. Lastly, to evaluate the performance of search results, Average Precision will be used. The main advantage of this framework is that it will provide the most optimum search result in short period of time. Also it will reduce those noisy URLs that degrade the performance of web search engines on large scale.***

Keywords: Click-through logs; Feedback Session; Re ranked search results; Average Precision

1. Introduction

The Current era is called "Internet Era" because of extensive use of Internet in day to day life by millions of users. Different users make use of search engine to get different required information from internet. In spite of that, the users are not satisfied with results at the end. Web search engines cover a broad topic with multiple users browsing over Internet. To enhance the significance of web search engine, it is very essential to meet the user's requirement of results. So a framework is proposed and implemented below which will not only help retrieve the most optimum search result in shorter period of time but also reduce the noisy URLs that degrade the performance of web search results.

A lot of research has been done in the past in the field of data mining applications. Many research scholars had proposed innovative approaches to get the desired search results. Before proceeding towards the details of proposed framework, let us have a look at some of the loopholes that were identified in user search results inference [1]:

a) Some URLs might get skipped by users as they appeared similar to previous ones. Also the unclicked URLs could wrongly reduce the weight of some terms in pseudo-documents.
b) There were many noisy URLs in the search results that were seldom clicked by users. When search results were clustered with noisy ones its performance degraded greatly.
c) In situation where the number of clustered URLs that were clicked by user appeared small, to segment them precisely was a difficult task.

2. Literature Framework

In [2], Kota Vamsee Krishna et.al proposed a desktop-based application which will re-rank and display search results, for end users, using user goals. Considering login/logout of a single session, user logs will be captured, feedback sessions and pseudo-documents will be generated to re-rank and optimize search results. Based on the query related matching information which is currently available in the database, search engine will display the results and user will have to scan through them to find out the website of his interest.

In [3], a framework was proposed that captured user's preferences to provide query suggestions. Web search personalization used collaborative filtering that adapted a generic search engine for the needs of a community of users. The main objective of work was to improve the retrieval of information by expanding the user query and to rank the result list based on the users domain of interest.

K.N. Vimal Shankar et.al [4] proposed a framework in which user profile was created for each individual and it got authenticated by client. The user interest was registered in the database and according to the change in interest of user the database got updated. Feedback sessions were generated and they were clustered dependent on the keywords. The original search results were then restructured and they were based on user search goals.

In [5], user search logs were generated from feedback session and accordingly a framework was proposed. This resulted in Pseudo-document being created which represented users search interest. An innovative approach known as Pattern Matrix was used. It consisted of documents and patterns were calculated as inputs. And later the documents were clustered. In this technique a new Clustering method called as semantic clustering was used to cluster documents. And by using TF*IDF matrix user search goal text were analyzed. Though TF*IDF is a good weighting algorithm, it clustered documents that had similar keywords so it identified near identical documents but it failed miserably for all those topics that shared one keyword even if they shared the same topic. Also for effective semantic clustering, the clusters should be correctly specified by the programmer because in this approach the clusters were created on demand and were not specified by the user.

In [6], Zheng Lu et.al proposed two methods for inferring user search goals. The first method deals with clustering of search results. In this method, a program is created that submits the queries to search engine. First 100 results are crawled up including the title and snippets. Then each search result is mapped to feature vector. The search results are clustered by means of K-means clustering and the most optimal K is selected by means of Classified Average Precision Criterion. But in this technique, search results are clustered. However, this technique resulted in noisy URLs which were randomly clicked by users also getting selected, thus degrading the performance.

In the second method different clicked URLs are clustered to infer user search goals. User click-through logs are created by combining different single sessions. These URLs, which were very few, were selected along with their titles and snippets. A feature vector is assigned to every clicked URL. Finally, results are clustered. But in this method, instead of feedback sessions the clicked URLs are clustered. So the number of URLs clicked differently is reduced. Thus, precise segmentation becomes difficult. Based on the gaps identified in the existing systems, a new method is proposed for restructuring and re ranking user search results with the help of feedback sessions. The Keywords predefined in the database will help perform categorization and also re-rank the search results.

3. Proposed System

The previous systems used for optimization of search results mostly made use of Pseudo-documents for enhancing user search goals. Also for evaluation of search results a metric "Average Precision" was used. Average Precision evaluated implicit search results only according to user feedback. It was not effective for restructured search results [6].

In our proposed system, user-click logs are generated and user search goals are inferred by clustering feedback sessions. The search results will then be restructured. Specific set of Keywords (Computer Systems, Automative, Animals) are predefined in the database. With the help of those keywords, categorization based re-ranking would be performed. A metric "Classified Average Precision" is used for evaluation of search results. The proposed new criterion evaluates search results by calculating the following:

a) Average Precision
b) Voted Average Precision (VAP)
c) Risk
d) Classified Average Precision (CAP)

The proposed system represented its block diagram as shown below in Fig.1:

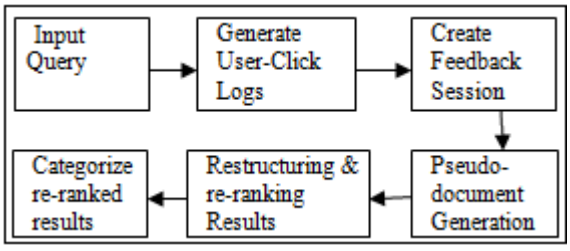

Figure 1: Block Diagram of Proposed System

3.1 Input Query

First, an application is made where the user can login and search for the term. The user profile is created for each individual. The user's need for information gets registered in the database as soon as user searches for term.

3.2 Generate User-Click Logs

The information is retrieved when user enters a particular term. Latter the links that are generated are stored in database. The Fig. 2 below depicts the terms entered by the user along with their visited links:

User_id	Query	VisitedLink	Date
8	jaguar	https://www.jaguarusa.com/index.html	2017-10-21 12:33:03.173
8	jaguar	https://en.wikipedia.org/wiki/Jaguar_Cars	2017-10-21 12:33:13.787
8	BMW	https://en.wikipedia.org/wiki/BMW	2017-10-21 16:36:05.880
8	BMW	https://www.bmw.com/en/home.html	2017-10-21 16:39:55.960
8	BMW	http://cpo.bmwusa.com/	2017-10-21 16:40:03.660
8	BMW	https://twitter.com/bmw?lang=en	2017-10-21 16:40:29.057
8	dog	https://en.wikipedia.org/wiki/Dog	2017-10-18 14:37:19.190
8	dog	http://www.animalplanet.com/pets/dogs/	2017-10-18 14:37:35.890
8	mahindra	https://mahindrausa.com/	2017-10-18 15:23:27.133
8	mahindra	http://www.mahindra.com/	2017-10-18 15:23:57.017
8	mahindra	http://www.mahindratuv300.com/	2017-10-18 15:24:14.693

Figure 2: User Search Goals

3.3 Create Feedback Session

In this module, a feedback session is created which consists of both the clicked and un-clicked URLs. Feedback session usually depicts what the user requires and what user does not require. The Fig. 3a & 3b below gives a diagrammatic representation of feedback session respectively wherein "0" indicates unclicked URLs and "1" indicates the clicked ones.

Query	Title	Link	Content	Sequence
jaguar	Jaguar USA: Jag...	https://www.jagu...	The official home...	0
jaguar	Jaguar - Wikipedia	https://en.wikipe...	The jaguar (Pant...	0
jaguar	Build Yours | Jag...	https://www.jagu...	Build your Jaguar...	0
jaguar	Jaguar Cars - Wi...	https://en.wikipe...	Jaguar is the luxu...	0
jaguar	2018 Jaguar F-T...	https://www.jagu...	Experience the la...	1
jaguar	Jaguar | FUTUR...	https://www.yout...	Sep 7, 2017 ... In...	0
jaguar	Jaguar | Basic Fa...	http://www.defe...	The jaguar is the ...	0
jaguar	Jaguar | E-type Z...	https://www.yout...	Sep 7, 2017 ... E-...	0
jaguar	Jaguars vs. Texans	http://www.jagu...	Links to Action N...	2
jaguar	Certified Pre-Own...	http://www.jagu...	Learn about the ...	0

Figure 3a: Feedback Session

Query	Title	Link	Content	Sequence
dog	Dog - Wikipedia	https://en.wikipe...	The domestic do...	0
dog	Complete Guide t...	http://www.vetst...	Your dog's online...	1
dog	Dog Protecting B...	https://www.yout...	Nov 14, 2016	0
dog	Dog Health Cent...	http://pets.webm...	WebMD veterina...	4
dog	Jealous Dog Wa...	https://www.yout...	Jan 20, 2017	0
dog	Dog Supplies | D...	https://www.dog...	Dog.com is your ...	3
dog	Dog	http://mashable....	Student asks to b...	0
dog	#dogs hashtag o...	https://twitter.co...	@dogs . We lov...	0
dog	Dog: Dog Breeds...	https://www.petfi...	Everything you n...	2
dog	Down Dog: Great...	https://play.goog...	The Down Dog A...	0

Figure 3a: Feedback Session

3.4 Generate Pseudo-Documents

For each distinct query, in each feedback session the following pre-processing steps are carried out-

a) Stop words Removal: Removes frequently occurring keywords like a, an, the, on, over etc. from each pseudo-document.
b) Non-words Removal: Removes non-words i.e. symbols like @, #, %, etc. from each pseudo-document.

The Fig.4 below gives a diagrammatic representation of Pseudo Document for Query "Dog "(Fig. 3b)

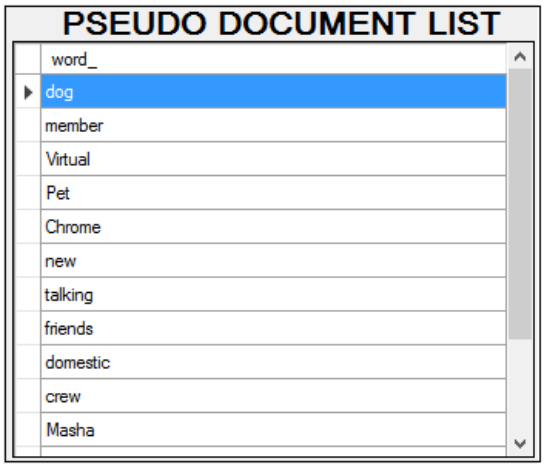

Figure 4: Pseudo Document

3.5 Restructuring Web Search Results

After the data is preprocessed, restructured web search results would be generated. Based on the user-click pattern the results will be stored in database. The search results will be recorded. And the re-ranking will also be performed based on users past history. The Fig.5 below shows the restructured results for Query" Mahindra":

Figure 5: Restructured Web Search Result

3.6 Categorize the Re-Ranked Results

For categorization, certain categories will be pre-defined in database. Depending upon the Keywords generated the re-ranked search results will be categorized. The Fig.6 below shows the final categorized re-ranked results:

Title	Query	link	Contlink	UserId	Category
Jaguar | Definition of Jaguar by Merriam-Webster	jaguar ...	https://www.merriam-webster.com/diction...	0	8	Animals
Jaguar Conquers Caiman In Epic Battle At Brazilian Ri...	jaguar ...	http://www.huffingtonpost.com/entry/jag...	0	8	Computer Systems
Drivers & Downloads | Dell US	dell	http://www.dell.com/drivers	2	8	Computer Systems
Mahindra Racing - Official Website of Mahindra Raci...	mahindra	http://www.mahindraracing.com/	0	8	Computer Systems
Jaguar Cars | 2017 Jaguar Models and Prices | Car an...	jaguar	https://www.caranddriver.com/jaguar	1	8	Computer Systems
View All Honda Car Models & Types | Honda	honda	https://automobiles.honda.com/vehicles	1	8	Automotive
Jaguar: Home	jaguar	http://www.jaguarspot.jaguarhg.com/	0	8	Animals
All Models	BMW	https://www.bmw.com/en/all-models.html	0	8	Computer Systems
Bayerische Motoren Werke AG: ETR:BMW quotes & ...	BMW	http://www.google.com/finance?q=ETR:...	0	8	Automotive
BMW Motorrad USA	BMW	http://www.bmwmotorcycles.com/us/en/...	0	8	Automotive
BMW Motorrad USA	BMW	http://www.bmwmotorcycles.com/promoti...	0	8	Automotive

Figure 6: Categorized Results

4. Performance Analysis

Evaluation of user search results on a large scale becomes a very tedious task because of single sessions in user click-through logs. In our paper we will evaluate the search results using CAP (Classified Average Precision) which gives the most optimized re-ranked search results. The search results are evaluated by calculating the following:

4.1 Average Precision (AP)

Average Precision is a method that computes user search

results based on users feedback at each point of retrieved document. The formulae for calculation "AP" is as follows:

$$AP = \frac{1}{N^+} \sum_{r=1}^{N} rel\ (r) \frac{R_r}{r}$$

- N^+ is the number of relevant (or clicked) documents in the retrieved ones,
- r is the rank
- N is the total number of retrieved documents
- rel () is a binary function on the relevance of a given rank
- R_r is the number of relevant retrieved documents of rank r or less.

For example in Feedback Session for Query "Dog" (Fig. 3b) the Average Precision is calculated as follows:
A.P. = 1/4 [0 + 1/2 + 0 + 4/4 + 3/6 + 0 + 2/9] = 0.55

But Average Precision (A.P.) fails for calculating restructured search results. To further optimize search results Voted Average Precision is calculated.

4.2 Voted Average Precision (VAP)

URLs in a single feedback session are re-structured into two classes where the un-boldfaced ones are clustered into class 1 and boldfaced ones are clustered into class 2. Voted Average Precision (VAP) is the average precision of the class including more clicks namely votes.

For example, the VAP of the restructured search results is the AP of Query "Dog" (Fig. 3b), calculated by:

VAP = 1/4 [1 + 1 + 1 + 1] = 1

A good restructured result will always have a higher VAP (close to1).

But VAP is not a good criterion. If each URL in the click session is categorized into one class, VAP will always be the highest value namely 1 no matter whether users have so many search goals or not. To avoid classifying search results into too many classes by error, Risk is proposed.

4.3 Risk

$$\text{Risk} = \frac{1 + 1 + 1}{C^2_m}$$

Where m = Number of clicks

The Risk calculated for Query "Dog" (Fig. 3b) is

$$\text{Risk} = \frac{1 + 1 + 1}{C^2_4} \qquad (m = 4)$$

$C^2_4 = 4(4 - 1) / 2 = 6$

Risk = 0.5

4.4 Classified Average Precision

CAP selects the Average Precision of the class that user is interested in (i.e., with the most clicks/votes. The influence of Risk on CAP is adjusted. The proposed CAP depends on both of Risk and VAP. The CAP for Query "Dog" (Fig 3b):
$CAP = VAP \times (1 - Risk)^{\gamma}$ ($\gamma = 0.7$)

$CAP = 0.55 * (1 - 0.5)^{0.7}$
$= 0.192$

5. Experimental Results

To implement the desktop based application, Google API was downloaded and added to the reference library so that application could fetch URLs from Google.

We compare our method with the existing method for 15 random queries. The Existing System [5] evaluated the search results by extracting Feedback sessions and mapping them to Pseudo-documents. (Feature representation of Pseudo-documents- computing similarity using cosine score).

Further the documents were clustered using K-means clustering; the performance of restructured search results were evaluated using the criterion "CAP".

The user search results can be evaluated by either resampling feedback sessions and clustering Pseudo-documents or Generating Feedback session and restructuring search results by processing Feedback session and then reranking search results.

Our paper generates Feedback Session, processes the pseudo-document (Stop-word removal and Non-word Removal Technique) restructures results and then evaluates search results by using Metric "Classified Average Precision". The past work analysis on user search goals has generated very few restructured results using Re-ranking and Categorization. The Fig.7 below gives the Graphical representation of existing method with our method:

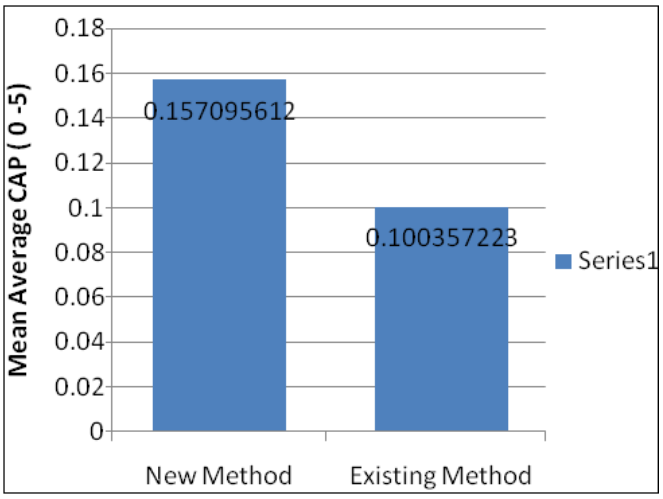

Figure 7: Our Method Vs Existing Method

The above Figure calculates the Classified Average Precision for 15 random Queries based on keywords (Computer Systems, Automotive, Animals) in database.

The X-Axis depicts the comparison between new method and existing method. The Y-Axis evaluates a metric "CAP" the wherein the limits are defined as (0 - 5).

A good restructured result will always have a CAP on a higher side. In the above Fig.6 our method has higher CAP.

In our research, we shortlisted three main categories namely (Computer Systems, Automotive, Animals). The Keywords were used for above three categories and graphs were designed accordingly. The Average Precision (AP) calculated for both the methods is same. The Figures below shows the graphical comparisons for all the keywords used for VAP , Risk and CAP respectively:

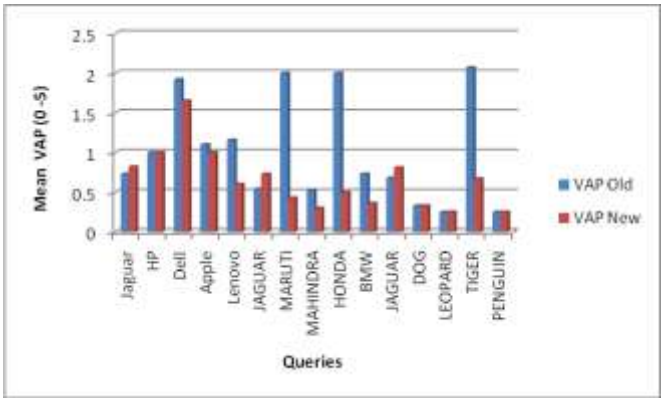

Figure 8a: Mean VAP

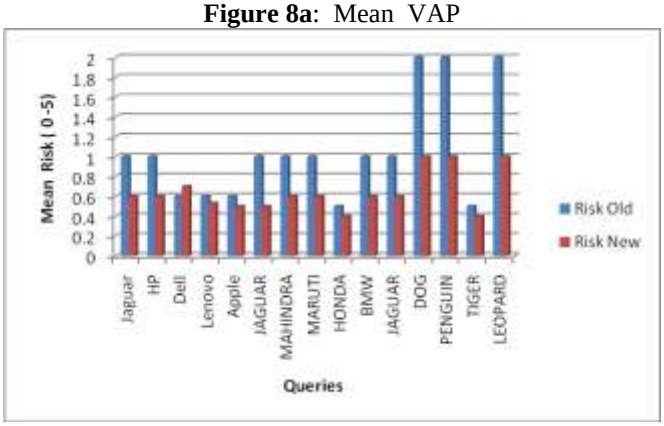

Figure 8b: Mean Risk

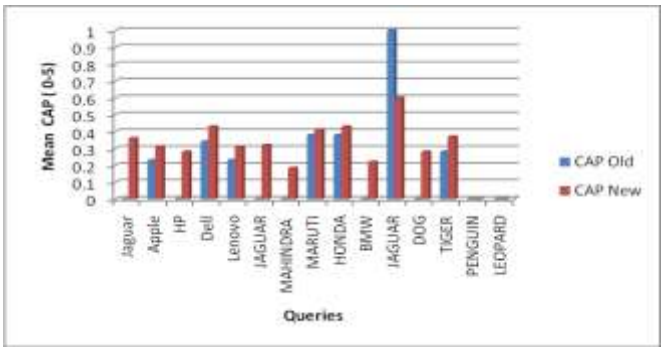

Figure 8c: Mean CAP

Although the VAP and Risk were not good for our system but CAP is on higher side.

6. Conclusion

In reality, our approach can discover user search goals for some popular queries offline at first. Then, when users submit one of the queries, the search engine can return the results that are categorized into different groups according to user search goals online. Thus, users can find what they want conveniently. Also clustering the search results removes the noisy results that degrade the performance of user search goals and helps generate relevant results in shorter period of time. In future the search results could be tested for its accuracy by performing other categorization techniques.

References

[1] Priti Alok Mishra, Bidisha Roy ,"A Novel Technique for Inferring User Search using Feedback Sessions" IEEE International Conference on Intelligent Computing and Control Systems (ICCS), pp.1-5, June 2017.

[2] Kota Vamsee Krishna , Smita Deshmukh,"Search Results Re-ranking using User Goals", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181(www.ijert.org) Vol. 4 Issue 02, pp.732-735, February-2015.

[3] Ms. D. Indumathi ,Dr. A.Chitra, "A Collaborative Search with Query Expansion and Result Re-ranking", IEEE World Congress on Information and Communication Technologies, pp.986-989, December 2011.

[4] S .Preetha , K.N. Vimal Shankar, "Personalized Search Engines On Mining User Preferences Using Click through Data", IEEE International Conference on Information Communication & Embedded systems (ICICES), pp. 01-06 , February 2014.

[5] Shine Sonali Bhaskar, Prof. Bharat Tides, "A New Approach and Compressive Survey on Restructuring User Search Results by using Feedback Session", IEEE International Conference on Computing Communication Control and Automation (ICCUBEA), pp.479-484, February 2015.

[6] Zheng Lu , Hongyuan Zha , Xiaokang Yang , Weiyao Lin ,Zhaohui Zheng, "A New Algorithm for Inferring User Search Goals with Feedback Sessions", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Vol. 25, No. 3, pp.502-512 ,March 2013.

[7] Jiafeng Guo, Lin Ding, Gang Zhang, Yue Liu, Xueqi Cheng, "PSM: A new Re-ranking algorithm for Named-page", *In Proc.* of the Fifteenth Text Retrieval Conference, TREC 2006, Gaithersburg, Maryland, pp.1-5 November 14-17, 2006.

[8] Dasari Amarendra, Kaveti Kiran Kumar, "Inferring User Search Goals with Feedback Sessions using K-means Clustering Algorithm", International Journal of Computer Engineering In Research Trends (IJCERT) Volume 2, Issue 11, pp. 780-784, November 2015.

[9] Sukanya S. Gawade , Gyankamal J.Chhajed, "Using Feedback Sessions for Inferring User Search Goals", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),Vol. 3, Issue 6 , pp.7015-7019, June 2016.

[10] Febna V, Anish Abraham, " User Search Goals Evaluation with Feedback Sessions ", International Conference on Emerging Trends in Engineering , Science and Technology (ICETEST), Procedia Technology (24), pp.1256-1262, Available online at www.sciencedirect.com.

[11] Hemlata Gaikwad, Manisha Darak," Implementation of Intellectual Browser for User Search Goal Discovery and Reordering System", International Journal of

Engineering Science and Computing (IJESC), Volume 6, pp.6535-6538, June 2016.

[12] Harshada P. Bhambure, Mandar Mokashi,"Click Based Inferring of User Search Goals usingPseudo Document", IEEE International Conference on Computer, Communication and Control (IC4),pp.01-07, 10-12 September, 2015.

Author Profile

Priti Mishra : Has received a degree of Bachelor In Computer Engineering (B.E.) from Mumbai University. Currently Pursuing (M.E.) Masters in Computer Engineering from SFIT, Mumbai University. Areas of Interest are IOT, Multimedia , Database, Semantic Web Technology. Has already published 4 papers in IEEE in respective areas of Interest.

Mrs. Bidisha Roy is an Associate Professor in Department of Computer Engineering at St. Francis Institute of Technology. She has received her M.E in Computer Engineering from University of Mumbai. She has more than 15 years of experience in teaching Undergraduate Computer Engineering more than 5 years of experience in teaching Post Graduate Computer Engineering. Her areas of interest include Algorithms and Complexity, Advanced Internet Technologies and Semantic Web, Data Mining, Software Engineering and Computational Intelligence. She is currently pursuing PhD. in Computer Engineering from University of Mumbai.