

# Role of SPARQL in Leveraging Sematic Web Technology

Poornima G. Naik, Kavita S. Oza



Abstract: Semantic web is not just a matter of translation from HTML to RDF/OWL languages. It is a matter of understanding the content of the web through knowledge graphs. Entities need to be related with relationships. This content is composed of resources (web pages) that contain, for example, text, images and audio. Thus, there is the need of extracting entities from these resources. Currently, most of the web content is in HTML5 format which is a W3C recommendation which enables describing the structure marginally with the help of annotations. The main challenge here is to transform unstructured data from plain HTML files to structured data (e.g RDF or OWL). The current work provides the first hand information for dealing with unstructured heterogeneous data residing on web using Twinkle, a Java tool for SPARQL query execution on FOAF (Friend Of A Friend) document.

Keywords : Filter, FOAF , Twinkle, RDF, Projection, Ontology, SPARQL.

#### I. INTRODUCTION

#### 1.1 Current State of Web

The current state of the web is highly unstructured and consists of vast repository of interconnected documents which are presented to end users as a collection of huge inter-linked documents. Extracting a structure from such highly unstructured web poses a big challenge to a researcher. Further, since the content is available for public access, quality of the content posted on WWW cannot be validated and guaranteed to be reliable. Also, the persistence of documents cannot be uniformly guaranteed. HTML's simplicity comes at a cost of interoperability which implies HTML documents are human readable but extensive ground work is desirable to make them machine readable and inter-operable by different software's. This is how XML emerged adding structuredness to unstructured HTML data in the form of DTD and Schema. The current state of the web is mature enough owing to the new technologies such as XML, Ontology, SPARQL etc. to name a few which strive to ingest some sort of structuredness and semantics to the otherwise unstructured and heterogeneous web.

Revised Manuscript Received on February 15, 2020. \* Correspondence Author

Poornima G. Naik, Professor in the Department of Computer Studies, CSIBER, Kolhapur.

Kavita S. Oza, assistant professor at Department of Computer Science, Shivaji University, Kolhapur

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

#### **1.2 Introduction to SPARQL**

SPARQL plays a key role in executing queries against heterogeneous data sources employing its native RDF format or which is transformed into RDF format by some middleware application. SPARQL operates on RDF graphs and mainly employs the logical operations conjunction and disjunction for unleashing the unknown relationships between the data and generates the results which can be result sets or themselves be RDF graphs.

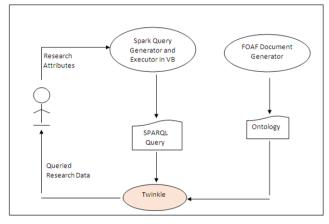
SPARQL is the query language of the Semantic web which enables

- Pulling the data from both structured as well as semi-structured data
- Data exploration by querying unknown hidden relationships between data
- Performing complex joins of heterogeneous databases employing a simple query
- Transforming RDF data from one vocabulary to another

The following section describes Twinkle, a Java-based tool for the execution of SPARQL queries.

#### 1.3 Working of Spark query generator and Executor Model.

Model is developed for execution of Spark query which accepts FOAF document, generates Spark query on fly and employs Twinkle, a Java based tool for execution of Spark queries. Figure 1 depicts working of SPARQ Query Generator and Executor model.



#### Figure 1 Working of SPARQ Query Generator and **Executor module**

 $\checkmark$  Depending on the subject area selected by an end user HTML document generator initially searches the local file system for the required FOAF file.



Retrieval Number: C5161029320/2020©BEIESP DOI: 10.35940/ijeat.C5161.029320 Journal Website: www.ijeat.org

Published By:

& Sciences Publication

The naming convention adopted for FOAF document is <subject name>FOAF.xml which is in XML file format.

- ✓ If the file exists then it is used otherwise FOAF document is dynamically generated by invoking HTML Page Extractor and Multi-threaded Web Crawler module.
- ✓ The foaf document created in the above step is input to spark query generator tool which generates dynamically a spark query.

The spark query is executed by employing a java based tool twinkle which generates the required output.

# II. RESULTS AND DISCUSSIONS

#### 2.1 Generating a Spark query based on end user input.

A GUI has been designed for the purpose in VB6 for accepting the research journal specific attributes from an end user. Currently, the user can query the research data

pertaining to the following attributes:Journal type which can be either national or international

- Information about the journal corresponding to one or more of the following fields:
- Title
- Volume
- Issue
- Charges
- e-ISSN
- p-ISSN
- UGC Recommended Journals
- Maximum impact factor in the available journals
- Impact factor in the given range.
- Minimum processing charges
- Processing charges in the given range.

In each case the user can view the generated Spark query.

# 2.2 Executing the generated Spark query using Twinkle

TWINKLE is the most popular tool used to execute SPARQL Query. To use the tool it requires jdk1.5 or higher to be installed on the system. Figure 2 shows command-line for the execution of Twinkle tool.



#### Figure 2 Execution of Twinkle Tool Through Command-Line

Most forms of SPARQL query contain a set of triple patterns called a basic graph pattern. Triple patterns are like RDF triples except that each of the subject, predicate and object may be a variable. A basic graph pattern matches a subgraph of the RDF data when RDF terms from that subgraph may be

substituted for the variables and the result is RDF graph equivalent to the subgraph. The SPARQL query along with the FOAF document generated is input to a java tool, Twinkle which fires SPARQL query on FOAF document to generate a desired output to an end user. Figure 3 depicts the "Spark Query" main menu structure.

3 Ipml	- 0 ×
Decide RDF Document View Spark Overy View Report Demo	
Generate Machine Readable Ontology (POH)	
View Machine Leadable Continger (CAR)	
Garantes Speek Carry Proceeding of the State	
terrestant Corry Ling matching Amain data SEMANTIC WEB	
(Barda D reserved and a second data a	
metadata semantice web	
ontaking mapping XM	
And	
Select Subject Area - Add New Area	
Processing	

### Figure 3. Structure of 'Spark Query' Menu of Semantic Web Application

Figure 4 depicts the GUI for SPARQL Query Generator.

Form3		- 0
bit Direct Proof George Other Sectors 10 Direct	Generation of Spark Query	
Ueloode UR	<ul> <li>○ National</li> <li>◎ International National</li> <li>○ Both</li> </ul>	C All
Impact Processing C	Factor : > C Maximum	Select Fields F (Trile)
Generate Spa	rk Query View Generated Spark Query	□ e-ISSN □ p-ISSN

# Figure 4. SPARQL Query Generator

The following section highlights execution of few sample SPRQL queries employing projection, selection and rewriting rule (employing Filter variable). The generated SPARQL queries are further executed by Twinkle, java tool for executing SPARQL queries.Figure 5(a)-5(b) depict the execution of SPARQL in Twinkle.

text table







🖹 Save 下 Run 🔞 Ca	un nel	514
harmonie	incer	- 0 ¢
ase URI		
ata URL	file:/F:/semantic_project/Managementfoaf.rdf	File
	://www.w3.org/1999/02/22-rdf-ayntax-ns≢> p//xmlns.com/foaf/0.1/>SELECT ?title tle .	
International Journal of International Research International Journal of	Management Studies (JMS) Business and Management Studies (JBMS) ournal of Management Sociology and Humanities Business, Management and Allied Sciences (JBMAS)	
	Management Practice(IJMP)	
	nal of Economics and Management Studies ( SSRG - IJEMS Journal of Management Sciences(IRJMS)	
	Management Excellence (IDME),	
International Journal of		
MERC Globals Internatio	n of Engineering and Management Education (IAEME)	

#### Figure 5(a)-5(b) Execution of SPARQL for Projection in Twinkle in Table Format for Research Journals in Computer and Management.

#### **Applying Projection**

**Query 1:** The following query will find the subject, predicate (properties) and object of the research journals.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> SELECT \* WHERE {?s ?p ?o}

Figure 6. depicts the execution of SPARQL in Twinkle.

	er "test2.rdf		0.9
	Save Run 🙆 Cancel		
📢 General 🛞	Base URI		
Write Simple Query			
C white surple Query	Data URL	fle:,D:/research1.rdf	Fic
	PREFIX rdf: <http: td="" www<=""><td>w.w3.org/1999/02/22-rdf-syntax-ns#&gt;</td><td></td></http:>	w.w3.org/1999/02/22-rdf-syntax-ns#>	
In Hemory	PREFIX foaf: <http: td="" xmlns<=""><td>.com/foaf/0.1/&gt;</td><td></td></http:>	.com/foaf/0.1/>	
Periodic Table	SELECT . WHERE (7s 7p 7o)		
Planet RDF Feed & Blogrol			
A Interencing			
Planet Feed (RDFS)			
Sector Se			
Persistent Stores 🛞			
		p	
Remote Services (2)	Sec///Dc/research1.edf#	http://www.com/foaf/0.1/bissn	e-555N : 0976-5697
B. House Friday C.	fie://D:kesearch1.rdf#	http://www.com/foaf/0.1/charges	2000
Offordia.org	fie:///D:/research1.rdf#	http://mins.com/foef/0.1/mpectfactor	Cosmos Impact Factor - 4.490
Gev/Track.up	Be://D:/research1.rdfm Be://D:/research1.rdfm	http://www.com/toat/u_lympactractor http://www.com/foat/0_lybean	p-199N : 2319-3778
ReynLon	fie:///D:/research1.rdf#	http://www.com/roat/u.igeon	p-100W 2 2319-3778
	fie:///D:/research1.rdf #	http://www.com/toat/u.t/losue http://www.com/foat/0.t/lide	2 International Journal of Advanced Research in Computer Scien
	file:///D:/research1.rdf#i	http://www.com/foaf/0.1/url	https://www.ijtet.org/
	file:///Dt/research1.rdf#i	http://www.com/foef/0.1/url	http://jarcs.info/
	file:///D:,kesearch1.rdf#i	http://www.com/foaf/0.1/eison	e-555N : 0976-5697
	file:///D:,/research1.rdf#i	http://www.com/foaf/0.1/bite	International Journal of Latest Trends in Engineering and Techn
	file:///Dt/research1.rdf#i	http://hmins.com/foaf/0.1/eissn	e-155N : 2278-621X
	file:///D:/research1.rdf#i	http://bmins.com/foaf/0.1/charges	2500
	file:///D:/research1.rdf#i	http://bmins.com/foaf/0.1/issue	9
	file:///D:/research1.rdf#i	http://wins.com/feef/0.1/impactfactor	Scientific Journal Impact Factor - 7.234
	file:///D:/research1.rdf#i	http://bmins.com/feaf/0.1/vol	9
	file:///D:/research1.rdf#i	http://www.com/fisaf/0.1/ugcapproved	yes
	file:///D:/research1.rdf#i	http://www.com/foaf/0.1/vol	a
	file:///Dt/Hesearch1.rdf#i	http://www.w3.org/1999/02/22+df-syntax-ns#type	http://wins.com/foaf/0.1/Journal
	fle:///D:/research1.rdf	http://www.com/wot/0.1/assurance	file:///D:/webwho.xrdf.asc
			file:///Dukesearch1.rdf#i
	file:///D:/research1.rdf	http://xmlns.com/foaf/0.1/primaryTopic	
	file:///D:/research1.rdf file:///D:/research1.rdf	http://www.com/toat/o.1/primary1opic http://www.com/foat/o.1/maker	fle:///D:/research1.rdf#i

#### Figure 6. Execution of SPARQL in Twinkle for Generating RDF Triplet

**Query 2:**The following query will find the title, volume, issue number, charges, UGC approved status, impact factor, eissn, pissn and URL of the research papers in various journals.

PREFIX rdf:

<http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?title ?vol ?issue ?charges ?ugcapproved ?impactfactor ?eissn ?pissn ?url WHERE { ?x foaf:title ?title . ?x foaf:vol ?vol . ?x foaf:issue ?issue. ?x foaf:charges ?charges. ?x foaf:ugcapproved ?ugcapproved.

- ?x foaf:impactfactor ?impactfactor.
- 2x foaf eisen 20isen

?x foaf:eissn ?eissn.

Retrieval Number: C5161029320/2020©BEIESP DOI: 10.35940/ijeat.C5161.029320 Journal Website: <u>www.ijeat.org</u> ?x foaf:pissn ?pissn. ?x foaf:url ?url.

**Query 3:**The following query will find the pattern matching using regular expressions for filtering research papers in international journals.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> SELECT ?title WHERE { ?x foaf:title ?title . FILTER regex(?title, "^Int").

# Figure 7. depicts the execution of SPARQL in Twinkle for pattern matching.

🖹 Save ▷ Run	🔯 Cano	el							- 7
Base URI									
Data URL	1	ile:/F:/sem	antic_project/	Computerfoaf.r	df			File	
PREFIX rdf:	<http: <="" td=""><td>/www.w3.0</td><td>org/1999/0</td><td>2/22-rdf-sy</td><td>ntax-ns#&gt;</td><td></td><td></td><td></td><td></td></http:>	/www.w3.0	org/1999/0	2/22-rdf-sy	ntax-ns#>				
PREFIX foaf:	<http:< td=""><td>//xmlns.</td><td>com/foaf/0</td><td>.1/&gt;</td><td></td><td></td><td></td><td></td><td></td></http:<>	//xmlns.	com/foaf/0	.1/>					
SELECT ?titl	e								
WHERE {									
?x foaf:titl	a 25151	•							
FILTER regex									
FILIER regex	(stitle	, ""Int")	-						
}									
title									
	rnal of Sc	ientific Rese	arch in Comp	uter Science. E	ngineering an	d Informati	on Tech		
International Jou						d Informati	on Tech		
International Jou International Jou	rnal of Lat	est Trends	in Engineering	and Technolo		d Informati	on Tech		
International Jou International Jou International Res	rnal of Lat æarch Jou	est Trends rnal of Com	in Engineering puter Science	and Technolo (IRJCS)	9Ÿ		on Tech		
International Jou International Jou International Res International Res	rnal of Lai æarch Jou æarch Jou	est Trends rnal of Com rnal of Com	in Engineering puter Science puter Science	and Technolo (IRJCS) and Information	gy on Systems (IF		on Tech		
International Jou International Jou International Res International Res International Jou	rnal of Lat æarch Jou æarch Jou rnal of Co	est Trends rnal of Com rnal of Com mputer Trei	in Engineering puter Science puter Science nds and Tech	and Technolo (IRJCS) and Informatio tology ( IJCTT	gy on Systems (IF		on Tech		
International Jou International Jou International Res International Res International Jou International Res	rnal of Lat earch Jou earch Jou rnal of Co earch Jou	est Trends rnal of Com rnal of Com mputer Trei rnal of Com	in Engineering puter Science puter Science nds and Tech puter Science	and Technolo (IRJCS) and Informatio hology ( IJCTT (IRJCS)	gy on Systems (IF )		on Tech		
International Jou International Jou International Res International Res International Jou International Jou	rnal of Lai earch Jou earch Jou rnal of Co earch Jou rnal of Co	est Trends rnal of Com rnal of Com mputer Tren rnal of Com mputer Scie	in Engineering puter Science puter Science nds and Tech puter Science nce and Infor	and Technolo (IRJCS) and Informatio ology ( IJCTT (IRJCS) mation Security	9y on Systems (IF ) / (IJCSIS)		on Tech		
International Jou International Jou International Res International Res International Jou International Jou International Jou	rnal of Lat earch Jou earch Jou rnal of Co earch Jou rnal of Co rnal of Co	est Trends rnal of Com rnal of Com mputer Trei rnal of Com mputer Scie mputer Scie	in Engineering puter Science puter Science nds and Tech puter Science ince and Infor	and Technolo (IRJCS) and Information tology ( IJCTT (IRJCS) mation Security mation Techno	y on Systems (IF ) ( (DCSIS) logy (DCSIT)		on Tech		
International Jou International Jou International Res International Res International Jou International Jou International Jou International Jou	rnal of Lat search Jou search Jou rnal of Co search Jou rnal of Co rnal of Co rnal of Co	est Trends rnal of Com mputer Trei rnal of Com mputer Scie mputer Scie mputer Scie	in Engineering puter Science puter Science nds and Techn puter Science ence and Infor ence and Infor	and Technolo (IRJCS) and Informatio hology (IJCTT (IRJCS) mation Security mation Techno vare Engineerin	y on Systems (IF ) ( (DCSIS) logy (DCSIT)		on Tech		
International Jou International Jou International Res International Res International Jou International Jou International Jou	rnal of Lat search Jou search Jou rnal of Co search Jou rnal of Co rnal of Co rnal of Co	est Trends rnal of Com mputer Trei rnal of Com mputer Scie mputer Scie mputer Scie	in Engineering puter Science puter Science nds and Techn puter Science ence and Infor ence and Infor	and Technolo (IRJCS) and Informatio hology (IJCTT (IRJCS) mation Security mation Techno vare Engineerin	y on Systems (IF ) ( (DCSIS) logy (DCSIT)		on Tech		
International Jou International Jou International Res International Res International Jou International Jou International Jou International Jou	rnal of Lat search Jou search Jou rnal of Co search Jou rnal of Co rnal of Co rnal of Co	est Trends rnal of Com mputer Trei rnal of Com mputer Scie mputer Scie mputer Scie	in Engineering puter Science puter Science nds and Techn puter Science ence and Infor ence and Infor	and Technolo (IRJCS) and Informatio hology (IJCTT (IRJCS) mation Security mation Techno vare Engineerin	y on Systems (IF ) ( (DCSIS) logy (DCSIT)		on Tech		
International Jou International Jou International Res International Res International Jou International Jou International Jou International Jou	rnal of Lat search Jou search Jou rnal of Co search Jou rnal of Co rnal of Co rnal of Co	est Trends rnal of Com mputer Trei rnal of Com mputer Scie mputer Scie mputer Scie	in Engineering puter Science puter Science nds and Techn puter Science ence and Infor ence and Infor	and Technolo (IRJCS) and Informatio hology (IJCTT (IRJCS) mation Security mation Techno vare Engineerin	y on Systems (IF ) ( (DCSIS) logy (DCSIT)		on Tech		

🛓 spark_query5		
🖹 Save ▷ Run 🔇 Can	cel	
Base URI		
Data URL	file:/F:/semantic_project/Managementfoaf.rdf	File
title		
	urnal of Management Sciences(IRJMS)	
International Research jo	urnal of Management Sociology and Humanities usiness, Management and Allied Sciences (DBMAS)	
	lanagement Studies (DMS)	
	lanagement Excellence (IJME),	
	usiness and Management Studies (IJBMS)	
	of Engineering and Management Education (IAEME)	
	lanagement Practice(IJMP)	
text table		

Figure 7. Execution of SPARQL in Twinkle for Pattern Matching



28

Query 4: The following query shows the method for converting string to integer or for employing an expression in a numeric for finding research papers in various journals with charges less than 3000.

	PREFIX foaf: <http: th="" xr<=""></http:>
CONVERTING STRING TO INTEGER	PREFIX xsd: <http: td="" ww<=""></http:>
PREFIX rdf:	SELECT ?impactfactor
<http: 02="" 1999="" 22-rdf-syntax-ns#="" www.w3.org=""></http:>	WHERE { ?x foaf:impa
PREFIX foaf: <http: 0.1="" foaf="" xmlns.com=""></http:>	ORDER BY DESC(xsd
PREFIX xsd: <http: 2001="" www.w3.org="" xmlschema#=""></http:>	Similarly, for finding re
SELECT ?charges	select the first item aft
WHERE {	order.
?x foaf:charges ?charges .	
FILTER(xsd:integer(?charges) < 2500).	FINDING MINIMUM CHA
}	

The execution of the query in Twinkle is shown in Figure 8.

🛓 spark_query6			- D ×		
🖹 Save ▷ Run 😫 Cancel					
Base URI	JRI				
Data URL	file:/F:/semantic_project/Computerfoaf.rdf File				
<pre>PREFIX foaf: <http PREFIX xsd: <http: SELECT ?title ?cha WHERE { ?x foaf:title ?tit ?x foaf:charges ?c</http: </http </pre>	le .				
title		charges			
International Journal of C	omputer Science and Informati				
	cientific Research in Computer				
Research Journal of Com	puter and Information Technol	1550			
International Journal of C	omputer Trends and Technolo	2200			
International Journal of Computer Science and Software 90					
International Research Journal of Computer Science and 0					
Journal of Computer Science and Information Technology 0					
	International Journal of Advanced Research in Computer 2000				
International Journal of C	omputer Science and Informati	0			
text table					

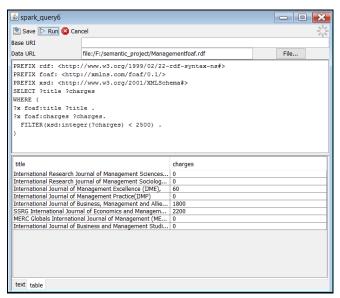


Figure 8. Execution of Filter SPARQL in Twinkle for Filtering

The logic for finding the research journals with highest impact factor is selecting the first item after sorting the impact factors in descending order.

FINDING MAXIMUM CHARGE PREFIX rdf: <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a> mlns.com/foaf/0.1/> /ww.w3.org/2001/XMLSchema#> actfactor ?impactfactor } d:integer(?impactfactor)) LIMIT 1 research journals with lowest charges, fter sorting the charges in ascending

ARGE

PREFIX rdf:

<http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> SELECT ?charges WHERE { ?x foaf:charges ?charges } ORDER BY ASC(xsd:integer(?charges)) LIMIT 1

#### **III. CONCLUSION AND SCOPE FOR FUTURE** WORK

The current research provides the first hand information for dealing with unstructured heterogeneous data residing on web with an emphasis to Twinkle, a Java tool for SPARQL query execution on FOAF document. The research can be extended further to retrieve the text from the images employing OCR tools. Also, image scraper or web scraper can be adopted for extracting large amounts of information from the website which involves downloading several web pages or the entire website which may include text from pages or HTML or both HTML and images. Some of the best web scraper tools are import.io, webhose.io, scrapehub, parsehub, visualscraper, spinn3r etc.

# REFERENCES

- Matthias Palmér "Learning Applications based on Semantic Web 1. Technologies." Doctoral thesis, Stockholm, Sweden 2012. Available at: https://www.diva-portal.org/smash/get/diva2:564709/FULLTEXT01.p
- Marcelo Arenas, Jorge Pérez "Querying Semantic Web Data with 2 SPARQL." Department of Computer Science, Universidad de Chile, Available at:
- 3 http://www.csd.uoc.gr/~hy561/papers/formalization/Querying%20Sem antic%20Web%20Data%20with%20SPARQL.pdf
- 4 Prasad Kulkarni "Distributed SPARQL engine using query MapReduce." University of Edinburgh , 2010. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.186.2063&re p=rep1&type=pdf
- Mulugeta Mammo "Distributed SPARQL over Big RDF Data, A 5.
- 6. Comparative Analysis Using Presto and MapReduce." Arizona State
- 7 University, December 2014, Available at:
- https://repository.asu.edu/attachments/143388/content/Mammo\_asu\_00 8. 10N 14524.pdf



Published By:

& Sciences Publication



#### **AUTHORS PROFILE**



Dr. Poornima G. Naik, received her M.Sc. degree in Physics and Mathematics and Ph.D. degree in physics from Karnataka University, Dharwad. She received MCA degree from IGNOU with first class Distinction. Currently, she is working as Professor in the Department of Computer Studies, CSIBER,

Kolhapur. Her areas of interest are network security, soft computing and cloud computing. She has participated in several national and international conferences, authored 20+ books on various cutting edge technologies in IT and has published more than 70 papers in International and national journals of repute. She is a prolific technical writer with excellent communication, analytical and technical skills. She is a recipient of prestigious Dr. APJ Abdul Kalam Life Time Achievement National Award for remarkable achievements in the field of Teaching, Research & Publications awarded by International Institute for Social and Economic Reforms, Bangalore.



Dr. Kavita S. Oza has received her PhD from Shivaji University in Computer Science. Currently she is working as assistant professor at Department of Computer Science, Shivaji University, Kolhapur. She is life member of CSI. Her research interest is machine learning, algorithms and Text mining. She has more 30 research publications to her credit. Two students have been awarded PhD under her guidance and four

are pursuing the same.

