International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Fuzzy Object-Oriented Database for Big Data

Bhawana Chaudhary, Santosh Kumar

Abstract*: Due to the huge amount of uncertain data collection, mankind facing a colossal amount and fast data having such a composite configurations. The data can be transact on the web and manually exchanges, interpersonal organizations and through our everyday life exercises. A very helpful result of designing an appropriate Big Data in different zones, for example, medical services, human services, executives and administrators. Managing the Big Data efficiently, new investigation perspective has been secured at this point the viewpoints about huge information (Big Data) requires extra-long stretch innovative interests. The Fuzzy Logic s have been implemented here for Big Data because of their capacities to handle the vagueness and uncertainties in the information. A couple of imaginative approach for Big Data processing is presented previously. To abbreviate the current duties and extant a point of view of further enhancements. We survey the different examinations and concentrated on that there are constrained Fuzzy systems have been adopted in large information (Big Data preparing. We also examine the various advantages of Fuzzy Logic s in large information (Big Data) issues. Therefore in this paper a Fuzzy object-oriented database is designed for Big Data and perform some Fuzzy queries to check the performance of the designed Fuzzy object-oriented database. We focused on the continuously propelled augmentations of Fuzzy sets and their blends in with various contraptions could offer a novel promising planning condition.

Keywords* : Big Data, Fuzzy Logic , Database, Fuzzy Object-Oriented Database (FOOD).

I. INTRODUCTION

In the era of Big Data, a tremendous volume of data is rapidly produced. As the data is gathered from various sources which has various verity and should be expounded in almost real time. There are three major factors that plays an important role in Big Data that volume, velocity and verity. These unique three-V is as yet substantial. Actually Big Data might be inadequately accurate in finding data. However the data that are used to examine may be exploited in several context like health care applications, business applications, IOT based applications etc.

Revised Manuscript Received on April 21, 2020.
* Correspondence Author
Bhawana Chaudhary*, School of Computer Science, Maharishi university of Information Technology, Lucknow, India. Email: rcbhawana844@gmail.com.
Santosh Kumar, School of Computer Science, Maharishi university of Information Technology, Lucknow, India. Email: sant7783@hotmail.com.

Extracting the valuable information from a huge and complexed database is little bit difficult but in context of Big Data is not a tricky task. Due to its various properties such as volume, redundancy, diversity and complexity. The Big Data has the humungous volume of data sets which cannot be possible to store or load into a single machine, that restrict the several algorithms such as data mining, machine learning and sequential algorithms. Due to the huge volume of Big Data that leads the uncertainty also. To keep in mind this, a Fuzzy Logic is implemented on the Big Data to make information more accurate. The Fuzzy Logic is based on the vague of uncertain values which lies between 0 and 1. These values express the fractional truth, whereas the Boolean logic based on either true or false values i.e. 0 (false) or 1 (true). Therefore, the Fuzzy Logic has the ability to manage the information that are ambiguous. These qualities can be handle with Big Data. A number of commitments on Fuzzy Logic for Big Data has presented before by several researchers which has explained in literature review section, many of them focused on design and implementation of Fuzzy Logic on Big Data.

In this work, we focused on the uncertainty and vagueness in result of queries of Big Data. A fuzzy object-oriented database is designed here and implemented some fuzzy queries on Big Datasets to compute the performance of the designed fuzzy object-oriented database for Big Data. Therefore, a fuzzy object-oriented database is modeled to handle that kind of objects and fuzzy data.

II. RELATED WORK

As there is limited research is done on fuzzy object-oriented database; let us first describe the important work done previously related to the fuzzy object-oriented database along with Big Data. H. wang et al [1] have rreviewed the recent studies from two distinct views. The first perspective spotlights on what kinds of fuzzy set strategies have been embraced. It recognizes clear patterns with regards to the use of fuzzy sets in enormous information handling (Big Data processing).Babanli M. B. [2] has applied the fuzzy set hypothesis to information mining from large information (Big Data) on material qualities and proposed fuzzy clustering created If-Then principles as a reason for computer amalgamation of new materials. Soha S. L. [3] has implemented classification techniques using the map reduce structure utilizing fuzzy and certain techniques, likewise to orchestrate an investigation that can thoroughly analyze the results of the recommended frameworks against the strategies evaluated in the archived works. Shweta T.

et al [4] have built up another calculation to deal with the characterization by utilizing fuzzy standards on this present reality informational collection. the proposed calculation caters in taking care of confirmation of understudies to different colleges by ordering them into three groups conceded, dismissed and the individuals who most likely would get the affirmation. Ikhlas A. et al [5] have used a hybrid approach utilizing fuzzy rationale and MapReduce to deliver another form of MapReduce which comprise of four layers. Golnar and Shahriar [6] have utilized information (data) anonymization method for affiliation rule stowing away, while parallelization and adaptability highlights are likewise inserted in the proposed model, so as to accelerate enormous information mining process. Pietro D. et al [7] have given an outline of the

latest conveyed learning calculations for producing fuzzy grouping models for Big Data. Gayatri and Shankar [8] have chipped away at order of understudy information to assess the exhibition of the understudy, Predictive Analytics in Higher Education, Mining Social Media Data for Understanding Students‟ Learning Experiences, assessing and anticipating understudy execution before admission to the school just as assessing the reasonableness of the passage tests. Betty and Ganesh [9] have examined about the audit on information based dynamic that are gotten from these two methods and how proficiently these two procedures gives the dynamic in their own specific manner and the means engaged with these systems for decision making. Rathore and Jayanthi [10] have applied Fuzzy Inference framework to foresee understudy execution which will assist with recognizing execution of the understudies and furthermore gives a chance to improve to performance. Gandikota et al [11] have proposed the structure is accomplishing the information anonymization by utilizing Fuzzy Logic by supporting huge information mining. Osama and Hesham [12] have presented an information digging approach for Big Data dependent on coordinating Fuzzy sets and unpleasant sets speculations. The proposed approach gives a novel granular information digging for Big Data that permit removing valuable information and rules from tremendous information to improve the dynamic procedure. Bhuvaneswari and Deepamalar [13] have concentrated on the effect of Fuzzy based Big Data investigation for the e-administration. Getting in sight to the consequences of the prescient examination can give tremendous advantages to the e-administration. Golnar and Shahria [14] have considered, information anonymization is utilized to fit the proposed model for Big Data mining. Aparna and Shubham [15] have concentrated on finding a superior specialized examination technique which would give quantitative outcomes rather than the paired one. Galindo et al [17] have introduced a theoretical Fuzzy model, purported Fuzzy EER, and a case tool (Fuzzy CASE), to help the database architects to fabricate the applied model for Fuzzy databases which are spoken to in the book Fuzzy Databases: Modeling, Design and Implementation and this book has a genuine impact in the direction of the databases course. Zadeh [16] has presented the hypothesis of Fuzzy sets and Fuzzy Logic , two ideas that established the framework of probability hypothesis in 1977. As indicated by him "the hypothesis of fluffy sets is a stage toward a rapprochement between the exactness of old style arithmetic and the inescapable imprecision of this present reality a rapprochement conceived of the unending human mission for a superior comprehension of mental procedures and insight. ZHANG [18] has managed Fuzzy inquiries with loads dependent on Fuzzy Logic conjunctions and give the outcome sets as indicated by coordinating degree.

III. EXPERIMENTAL STUDY

In this section a sample fuzzy object oriented database has been intended for the health care system i.e. a huge data set for several kinds of diseases. A fuzzy object-oriented database has been designed here for viral disease named COVID-19. An architecture for Big Data is presented in figure 3.1.1 below and Fuzzy Logic or fuzzy set has been implemented to remove the uncertainty or vagueness in the resultant values. Therefore, for implementing the logics a fuzzy object-oriented database has been designed here and performed some fuzzy queries to evaluate the performance of the designed database.

A. Architecture for Big Data Processing

An architecture for Big Data processing is represented in figure 3.1.1. an uncertain data is collected from the users and corrected it by applying some Fuzzy Logic s and then process it to make a decision.

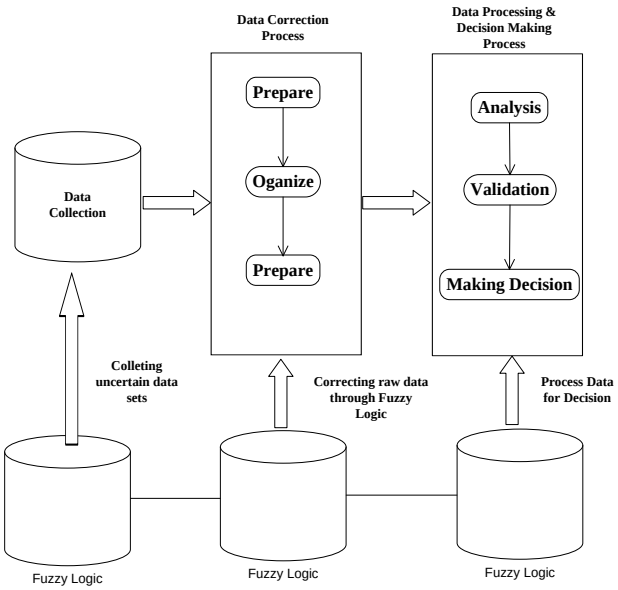


Fig 3.1.1 An Architecture for Big Data Processing

B. Modelling Of Fuzzy Object-Oriented Database for Big Data

In the field of clinical sciences, there are such a lot of dubiousness and vulnerability found in the clinical records. In this way, it is exceptionally hard to locate an exact arrangement of a given issue based on dubious information. The procedure of analysis of any infection by utilizing the Fuzzy Logic has a few phases like dubious estimations of manifestations of sicknesses, pathology test results for example surrendered to run, infections signs and so on.

The fuzzy object-oriented database gives the opportunity to speak to and launch that sort of questionable and exact properties esteems and reinforced the relations. As per the famous scientist **Galindo [17]**,

the fuzzy object-oriented database model is an augmentation of diagram based item model that oversee both certain and ambiguous values/data by using fuzzy set hypothesis and plausibility hypothesis. Therefore, the conceptual schema is characterized as quintuple {C, T, A, P, N}:

Where:

a. C is a limited arrangement of class names (certain and unclear classes). Fuzzy classes accumulate objects, which can have a partial enrolment to the class.
b. T is a limited arrangement of type names (clear and vague sorts). Indistinct sorts of segments demonstrate sets of questionable and free characteristics.
c. A denotes the characteristics names. Attributes are exceptionally basic enough as the zone is their sort and gotten unpredictable as the territory of properties
d. become their class. Also, both single-regarded and multivalued attributes exist.
e. P is the property connection. P relates one class with its property names and both with the area. This space might be another class or type.
f. H is the legacy connection. On the off chance that the legacy connection is fuzzy, at that point one name indicates the degree to which the occurrences of the subclass are additionally occasions of the super class

As per the famous scientist **Zadeh [16],** the Fuzzy Logic is an approximation rather than the exact reasoning. Therefore; a fuzzy set is defined as: let S be a fuzzy set over vast expanse of discourse X, i.e. a limited or unbounded break inside which the fluffy set can take a worth having a lot of sets:

$$S = \{\mu_s(x)/x: x \in X, \mu_s(x) \in [0, 1] \in R\} \quad (1)$$

The membership function or degree of the component x to the Fuzzy set S is represented by μ_s and the universe of discourse X of qualities being considered can be of two sorts:

1. Limited or tactual universe of discourse

$X = \{x_1, x_2, \ldots\ldots, x_n\}$, where set *S* can be represented by:

$$S = \mu_1 / x_1 + \mu_2 / x_2 + \ldots + \mu_n / x_n. \quad (2)$$

2. Limitless universe of discourse, where set An over X can be represented by:

$$S = \int \mu S(x) / x \quad (3)$$

Therefore, a set S that has linguistic variables is defined for its input values such that disease, signs, symptoms and pathology tests results. For input values a membership function μ_s is defined that are applied to determine the truth values. The input values are represented through set $S = \{x_1, x_2, \ldots\ldots, x_n\}$ is { signs, symptoms, pathology tests results}. Therefore, the set is defined as:

$$S = \{x_i, \mu_s(x_i)\} \quad (4)$$

Where i = 1, 2, 3, 4,…….n.

Therefore, a table is designed here that representing linguistic variables and range values for the person who are suffer from COVID-19. In membership function each variable described through the linguistic variables associated with the elements of the fuzzy set S i.e. Normal, Moderate, Critical, Very Critical. The output parameters are in light of the etymological factors and their related fuzzy value range which is given in **Table I.**

Table I. Linguistic Variables along with their Range values

Linguistic Variables		*Range values*
	Dyspnoea respiratory	12-15 per min
Mild	Blood oxygen saturation	95-100%
	PaO_2/FiO_2 ratio	200-300
	Lung infiltrate	14-40%
Moderate	Dyspnoea respiratory	15-30 per min
	Blood oxygen saturation	<=95%
	PaO_2/FiO_2 ratio	100-200
	Lung infiltrate	24-50%
Sever	Dyspnoea respiratory	>=30/minutes
	Blood oxygen saturation	88-92%
	PaO_2/FiO_2 ratio	<100
	Lung infiltrate	>50%
Critical	Dyspnoea respiratory	>30/minutes
	Blood oxygen saturation	<92
	PaO_2/FiO_2 ratio	<100
	Lung infiltrate	>50%

By incorporating these linguistic variables and their range values, the Fuzzy Object-Oriented Database (FOOD) has been designed for the patient who suffers from deadly disease called novel corona virus i.e. COVID-19. There are several major fields like COP_ID, COP_Name, Symptoms, COPPatient_Status, Disease, Linguistic Variable and Fuzzy Value are shown in the database **Table II**.

Table II. Fuzzy object-oriented database for Healthcare System

DESKTOP-V6GDTR7...op_patient_data | DESKTOP-V6GDTR7...op_patient_data*

cop_id	cop_name	cop_age	cop_gender	cop_symptoms	cop_deseases	cop_state	cop_patientstatu
1	harsha	34	F	drycoughmhighf...	COVID-19	andhrapradesh	Recovered
2	Mohammad ashr...	64	M	drycoughmhighf...	COVID-19	andhrapradesh	Recovered
3	vijayanand	55	M	drycoughmhighf...	COVID-19	andhrapradesh	Death
4	nirmal singh	105	M	drycoughmhighf...	COVID-19	bihar	Death
5	keshaw kumar	31	M	drycoughmhighf...	COVID-19	Lucknow	Cured
6	rama singh	19	F	drycoughmhighf...	COVID-19	Lakhimpur	Death
7	sunita	46	F	drycoughmhighf...	COVID-19	Lucknow	Recovered
8	kanika kapoor	41	F	drycoughmhighf...	COVID-19	Lucknow	Recovered
9	ramakant	22	M	drycoughmhighf...	COVID-19	Chandigharh	Death
10	shahil siddaqui	65	M	drycoughmhighf...	COVID-19	Mumbai	Death

After designed the fuzzy database some fuzzy queries have been performed here.

The general structure for a fuzzy query is given as:

SELECT attributes
FROM sourceTableName
WHERE condition 1…….. condition n
WITH α

where:

attributes: attributes are the fields in the database

sourceTableName: it is name of the table from where the records redrived.

Condition: it is to be checked, it may be complexed, composed and simple.

α: it is a threshold which is a connection between different conditions.

WHERE Clause: it can be represented with quantifiers in fuzzy queries. There are two types of quantifiers: **absolute quantifier** (it represent at least two or more fuzzy subset of nonnegative numbers) and **relative quantifier** (it represented as fuzzy subset of at least, most and etc of unit interval.

one of the important part of the fuzzy query is membership function or degree of attributes. Therefore, there are different types of membership function/degree (for example Age, Dyspnoea respiratory, Blood Oxygen Saturation etc). here we take a membership function as:

membership function/degree for Age:

Table III. Membership degree for Age

Linguistic Variable	*Range Value*
Young	<=25
Middle	40-60
Old	>=60

$$\mu_{young}(x) = \begin{cases} 1, & x \le 25 \\ \frac{(25-x)}{10}, & x \ge 20 \end{cases}$$

$$\mu_{middle}(x) = \begin{cases} 1, & x \ge 25 \\ \frac{(60-x)}{10}, & 40 \le x \le 60 \\ \frac{x-40}{10}, & 25 \le x \le 40 \end{cases}$$

$$\mu_{Old}(x) = \begin{cases} 1, & x \ge 60 \\ \frac{(60-x)}{10}, & 60 \le x \le 55 \end{cases}$$

Membership Function/degree of Dyspnoea respiratory

Table IV. Membership degree for Dyspnoea respiratory

Linguistic Variable	*Range Value*
Mild	12-14 per minute
Moderate	15-30 per minute
Sever	>=30 per minute

$$\mu_{mild}(x) = \begin{cases} 1, & x < 14 \\ \frac{(14-x)}{2}, & x > 12 \\ \frac{x-12}{2}, & 14 > x \end{cases}$$

$$\mu_{moderate}(x) = \begin{cases} 1, & x < 30 \\ \frac{(30-x)}{2}, & x > 15 \\ \frac{x-15}{2}, & 30 > x \end{cases}$$

$$\mu_{sever}(x) = \begin{cases} 1, & x < 14 \\ \frac{(14-x)}{2}, & x > 12 \\ \frac{x-12}{2}, & 14 > x \end{cases}$$

Hence, a fuzzy query is constructed here for retrieving the record of person who have covid-19 symptoms.

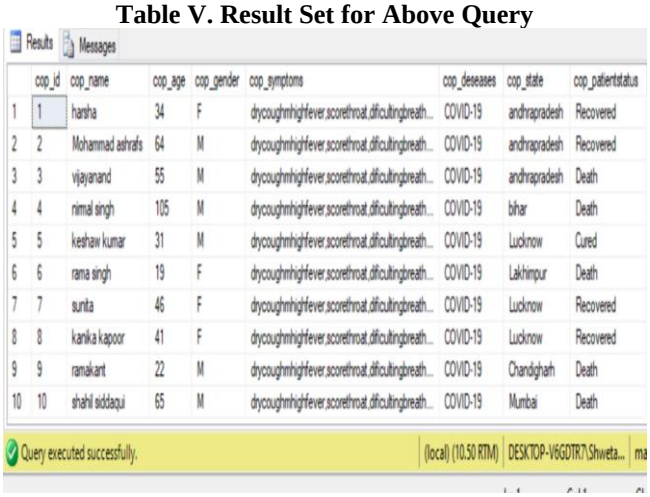
SELECT * FROM Patient WHERE((Age=0.5 AND Age>=0.7) AND Dysp_Resp>=0.3 AND Dysp_Resp<=0.1))

The above query can be expressed as:

SELECT * FROM Patient WHERE(Age="Old" AND Dysp_Resp = "High")

From the above query the result set is as bellow:

Table V. Result Set for Above Query


Results Messages

	cop_id	cop_name	cop_age	cop_gender	cop_symptoms	cop_deseases	cop_state	cop_patientstatus
1	1	harsha	34	F	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	andhrapradesh	Recovered
2	2	Mohammad ashrafs	64	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	andhrapradesh	Recovered
3	3	vijayanand	55	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	andhrapradesh	Death
4	4	nimal singh	105	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	bihar	Death
5	5	keshaw kumar	31	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Lucknow	Cured
6	6	rana singh	19	F	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Lakhimpur	Death
7	7	sunita	46	F	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Lucknow	Recovered
8	8	kanika kapoor	41	F	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Lucknow	Recovered
9	9	ramakant	22	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Chandigharh	Death
10	10	shahil siddaqui	65	M	drycoughmhighfever,sorethroat,difficultingbreath...	COVID-19	Mumbai	Death

Query executed successfully. (local) (10.50 RTM) DESKTOP-V6GDTR7\Shweta...

IV. CONCLUSION

From the above work, it is measured that the Fuzzy Logic is can be implemented easily on the Big Data. The present work is an attempt to represent that the fuzziness in Big Data can be eliminated by applying the fuzzy techniques to get a crisp result which provide help to make a valid decision in the fields where the decision making is quite difficult. A Fuzzy Database is designed here for healthcare system. This work can be further extended for designing the 3D data cube that anyone can extract the information easy and faster manner about the patient details and other related details.

ACKNOWLEDGMENT

Authors are grateful to Vice-Chancellor, Maharishi University of Information Technology Lucknow for giving the incredible facilities in the processing lab of Maharishi college of Information Technology, Lucknow, India. Much appreciated are additionally because of University Grant Commission, India for help to the University.

REFERENCES

1. Hai w a n g , Zeshui Xu and witold pedryz, "An overview on the roles of fuzzy set techniques in Big Data processing: Trends, challenges and opportunities ", Knowledge-Based Systems, Vol. 118, 15 February 2017, pp. 15-30.
2. Babanli M. B, "Synthesis of new materials by using fuzzy and Big Data concepts", 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August 2017, Budapest, Hungary, Procedia Computer Science, Vol. 120, pp.104–111.
3. Soha S. L., "A comparative study to classify Big Data using fuzzy techniques", 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 6-8 Dec. 2016, UAE.

4. Shweta T., Bhawna S., Himanshu N., Anchit J., Akshay K. and Sachin G., "A new approach for data classification using Fuzzy Logic ", 6th International Conference - Cloud System and Big Data Engineering (Confluence), 14-15 Jan. 2016, Noida, India.
5. Ikhlas A., Wael A. and Mohamad A. , "Hybrid Approach Using Fuzzy Logic and MapReduce to Achieve Meaningful Used Big Data", International Journal of Engineering & Technology, Vol. 7, Issue 4, pp. 6997-700, 2018.
6. Golnar A. A. and Shahriar M., "Privacy Preserving Big Data Mining: Association Rule Hiding", Journal of Information Systems and Telecommunication, Vol. 4, No. 2, pp. 70-77, April-June 2016.
7. Pietro D., Michela F. and Francesco M., "An overview of recent distributed algorithms for learning fuzzy models in Big Data classification", Journal of Big Data, Vol. 7, Issue 19, pp. 1-29, 2020.
8. Gayatri N. and Shankar M. P., "Classification System to Predict Universities for Students using Fuzzy Logic ", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 3, pp. 424-427, March 2017.
9. J. Betty J. and E. N. Ganesh, "A Review On Big Data With Machine Learning And Fuzzy Logic For Better Decision Making", International Journal Of Scientific & Technology Research Volume 8, Issue 10, October 2019.
10. R. K. Rathore And J. Jayanthi, "Student Prediction System For Placement Training Using Fuzzy Inference System", Ictact Journal On Soft Computing, Vol. 07, Issue 03, pp. 1443-1446, April 2017.
11. Gandikota R., Soumya M., Appawala J., Somasekar J. and K. K. Baseer, "Protecting Big Data mining association rules using fuzzy system", TELKOMNIKA Telecommunication, Computing, Electronics and Control Vol.17, No.6, pp.3057-3065, December 2019.
12. Osama S. A. and Hesham A. H., "A Novel Data Mining Approach for Big Data Based on Rough Sets and Fuzzy Logic ", International Journal of Intelligent Engineering and Systems, Vol.12, No.6, pp. 133-146, 2019.
13. S. Bhuvaneswari and M. Deepamalar, "Enhanced and Secured Big Data Analytics using Fuzzy Set for Effective E-Governance", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 6, pp. 10582-10591, June 2017.
14. Golnar A. A. and Shahriar M., "Privacy preserving Big Data mining: association rule hiding using Fuzzy Logic approach", The Institution of Engineering and Technology, Volume 12, Issue 1, January 2018, p. 15 – 24.
15. Aparna S. and Shubham C., "Big Data Analysis Using Fuzzy Logic and K-mean Clustering Algorithm", National Conference on Emerging Technologies for Sustainable Engineering & Management (NCETSEM'18)-2018, pp. 1-5.
16. Zadeh L.A., 1965. Fuzzy Sets, Information and control, Vol. 8, pp. 338-353.
17. Galindo J., Urrutia A. and Piattini M., "Fuzzy Databases: Modeling, Design and Implementation", Idea Group Publishing.
18. ZHANG P. , "A Study on Database Fuzzy Query Method in SQL", International Conference on Advances in Engineering, procedia Engineering, 24, 2011, pp. 340-344.

AUTHORS PROFILE

Ms. Bhawana Choudhary, PhD Scholar from Maharishi University of Information Technology , Lucknow had completed her MCA from Pt. Ravishankar Shukla University, Raipur and more than 10 years in IT services in different MNC in India.

Dr. Santosh Kumar, Associate Professor & Dy. Dean of School of Computer Science, in Maharishi University of Information Technology, Lucknow, INDIA. He has awarded Doctorate in Computer Science from Babasaheb Bhimrao Ambedkar University (A Central University) Lucknow, INDIA, in the area of Software Engineering along with data mining. He has 8 years of Teaching and 5 years of research experiences in the field of software engineering, Data Mining, Fuzzy object-oriented database and cloud computing. He has guided more than 4 Ph.D. Scholars and published more than 44 International/National research papers in various refereed journals and attended more than 10 International/National conferences.