International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-4 Issue-6 December-2012

Novel Approach to Knowledge Extraction Using Conceptual Association

Neha Mittal[1], Kanika Sharma[2] ,V.K.Panchal[3]
[1]GPMCE(I.P.University),[2]GCEW(MD.University),[3]DefenceTerrainResearchLaboratory, Delhi India

Abstract

This paper describes the novel approach to Knowledge Extraction using conceptual association. Web has a huge collection of unstructured information. In this paper we have discovered structured knowledge from unstructured information available on the web browser and examine relationship between different concepts using web. Our concept is OCEAN WATER DESALINATION(OWD).OceanWater Desalination is the process of creating fresh water by removing saline(salt) from large ocean bodies with saline water. We have taken the view of four countries of the world namely INDIA(a country in continent ASIA), AUSTRALIA (a country in continent AUSTRALIA),U.S.A(a country in continent NORTH AMERICA) and AUSTRIA (a country in continent EUROPE) and studied the techniques presently used and proposed for OWD .Based on our study, we have classified techniques into two generations namely first and second generation and then we have attempted to extract the knowledge taking the help of Rough set theory and Fuzzy logic. Using overlap coefficient method to find Co-occurrence number, fuzzy membership function we have defined the concept strength in terms of nodes and arcs in visibility graph. Finally, fuzzy associative matrix depicts how various elements have worked together to yield some useful result.

Keywords

Concept association learning (CAL),Rough set theory, Overlap coefficient method,Co-occurrence matrix, Fuzzy membership function, Fuzzy associative matrix.

1. Introduction

In recent times as the technology is advancing we have come across various conditions where different people have different viewpoints on same concept, that concept can be a process, or a technique. For example:- Consider ship building process in which we collect viewpoints of different countries and we extract the knowledge from the web browser on the basis of which relationships and nodes generated.This paper is organized as follows: In section 2, we summarized the basic concepts concerning Concept association learning(CAL) ,Rough set Theory and fuzzy logic. Section 3 summarizes the problem definition. Section 4 discusses the methodology. Section 5 discusses the proposed algorithm. Section 6 discusses the implementation of methodology in our case study. Section 7 discusses the result. Section 8 discusses the conclusion and future work.

2. Basic Concepts

This section describes the basic concepts regarding the Concept Association learning(CAL) ,Rough set theory and fuzzy logic.

A.*Concept Association learning(CAL)*

Concept association is a process of forming mental connections or bonds between sensations, ideas or memory. Concept association learning is a learning process in which two or more items or concepts become associated with each other, often used in relation to learned stimulus-response relationships. For instance, we associate the concept"dog" with all of the characteristics of dogs (four legs, fur, tail and so on) and are able to generalize the concept about unfamiliar dogs. Eleanor Rosch(1978) have truly said that concepts are learned through examples rather than abstract rules.

B. Rough Set Theory

The first pioneering paper on Rough Set was written by Zdzislaw Pawlak, Polish scientist in 1982. A **rough set,** is a formal approximation of sets which give the *lower* and the *upper* approximation of a crisp set (i.e., conventional set) in terms of a pair of the original set but in other variations, the approximating sets may be fuzzy sets. Rough set philosophy is founded on the assumption that in every object of the universe of discourse some information (data, knowledge) is associated.

C. Fuzzy Logic

Fuzzy Logic was initiated in 1965 by Lotfi A. Zadeh . Basically, Fuzzy Logic (FL) is a multi-valued logic, that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc. As an extension of the case of multi-valued logic, valuations($\mu : V_o \rightarrow W$)of propositional variables (V_o) into a set of membership degrees (W) can be thought of as membership functions mapping predicates into fuzzy sets.

Fuzzy Sets:-Fuzzy Sets are the simple fuzzy variables which have values in the range [0,1]and represent the degree to which a quality is possessed.

3. Problem Definition

In this Section we present you the concept we have considered in this paper. Our concept is **OCEAN WATER DESALINATION(OWD).**Based on our study, we have classified techniques into two generations namely first and second generation and then we have attempted to extract the knowledge taking the help of Rough set theory and Fuzzy logic.The first generation processes are defined as those which are being presently used and Second generation processes are defined as those which are meant to be used in coming future years.

First generation processes are categorized into three categories:

1) Common processes
2) Similar processes.
3) Speciality of a country.

Second generation processes are categorized into three categories:

1) Common processes
2) Similar processes.
3) New Concepts.

Under Common processes, we listed out those processes which are presently being used or proposed to be used in all the countries.Under Similar processes, we listed out those processes which are being used or proposed to be used in two or more countries but not all of them. Both Common and Similar processes are used in first as well as second generation Under Speciality of a country, we listed out those processes which are used in only one country. It is used only in first generation processes. Under New concepts, we have listed the proposed techniques to be used for Second generation.

Table 1 lists description of first generation processes of different countries. Table 2 lists description of second generation processes of different countries. India does not have its own second generation processes, those listed in table 2 are proposed.

Table 1: First Generation Processes

Countries	Common processes	Similar processes	Speciality of a country
India	1).Reverse Osmosis(R.O) 2)Solar Desalination 3)Ion Exchange Method Vapour Compression(V.C) 4)Electrodalysis(E.D) 5)Thermal Desalination (a)Multi-Effect Distillation (b)MultiStage Flash Distillation	1) India-Australia-Austria (a)Clathrate or hydrate Formation 2)India-Australia-U.S.A (a)Forward Osmosis(F.O) (b)Electrodialysis Reversal(EDR) 3)India-U.S.A (a)Nanofiltration Technology	1)Sea water green house technology 2)Geothermal Technology
Australia	1).Reverse Osmosis(R.O) 2)Solar Desalination 3)Ion Exchange Method Vapour Compression(V.C) 4)Electrodalysis(E.D) 5)Thermal Desalination (a)Multi-Effect Distillation (b)MultiStage Flash Distillation	1)Australia-India-Austria (a)Clathrate or hydrate Formation 2) Australia-India-U.S.A (a)Forward Osmosis(F.O) (b)ElectrodialysisReversal(EDR) 3) Australia-U.S.A (a)Freeze Desalination (b)Ultra filtration Technology 4)Australia-Austria (a)Vaccum freeze Desalination	1) Pressure Retarded Osmosis 2) Dew Vaporation 3)Secondary refrigerant freezing(SRF)

U.S.A	1).Reverse Osmosis(R.O) 2)Solar Desalination 3)Ion Exchange Method Vapour Compression(V.C) 4)Electrodalysis(E.D) 5)Thermal Desalination (a)Multi-Effect Distillation (b)MultiStage Flash Distillation	1)U.S.A-India-Australia (a)Forward Osmosis(F.O) (b)ElectrodialysisReversal(EDR) 2)U.S.A-India (a)NanoTechnology 3) Australia-U.S.A (a)Freeze Desalination (b)UltraFiltration Technology	1) Microfiltration Technology
Austria	1).Reverse Osmosis(R.O) 2)Solar Desalination 3)Ion Exchange Method Vapour Compression(V.C) 4)Electrodalysis(E.D) 5)ThermalDesalination (a)Multi-Effect Distillation (b)MultiStage Flash Distillation	1)Austria-India-Australia (a)Clathrate or hydrate formation 2)Austria-Australia (a)Vaccum freeze Desalination	1)Solvent Extraction 2)Pervaporation

Table 2: Second Generation Processes

Countries	Common processes	Similar processes	New concepts
India	1) Membrane Distillation	1)India-Australia a) Reverse Osmosis(R.O)+ Vapour Compression(V.C) b)Electrodialysis(ED)+Reverse Osmosis 2)India-U.S.A (a) Magnetic Technology (b)Ultrasonic+Supersonic Technologies 3)India-Austria (a)MultiEffectDistillation+Vapour Compression(V.C) 4)India-Austria-U.S.A (a) Wind Technology (b)Fossil fuel Technology 5)India-Australia-U.S.A (a) Capacitive Desalination	1)Reverse Osmosis(R.O)+Vapour Compression (V.C) 2)Ultra+Supersonic Technology 3) Fossil fuel Technology 4)Electrodialysis+Reverse Osmosis 5)Magnetic Technology 6)Vapour Compression+Multiple-Effect Distillation 7)Capacitive Desalination 8) Wind (Hydro)Technology
Australia	1) Membrane Distillation	1)Australia-India (a)ReverseOsmosis(R.O)+Vapour Compression(V.C) (b)Electrodialysis(ED)+Reverse Osmosis 2)Australia-India-U.S.A (a) Capacitive Desalination 3) Australia-U.S.A (a) Biomemtric Technology	1)Reverse Osmosis(R.O)+Vapour Compression (V.C) 2)Electrodialysis(ED)+Reverse Osmosis(R.O) 3)Capacitive Desalination 4)Biomemtric Technology 5)Gas hydrate (clathrate) Desalination
U.S.A	1) Membrane Distillation	1)U.S.A-India (a)Magnetic Technology (b)Ultrasonic+Supersonic Technologies 2)U.S.A-India-Australia (a) Capacitive Desalination 3)U.S.A-India-Austria (a) Wind Technology (b)Fossil fuel Technology 4)U.S.A-Australia (a)Biomemtric Technology 5)U.S.A-Austria (a)Solar ponds	1)Magnetic Technology 2)Capacitive Desalination 3)Superconcentrate Technologies 4)Ultrasonic+Supersonic Technologies 5)Solar salt ponds 6)Watershed based salinity management 7)Waterharvesting from air 8)Biomemtric Technology 9)Wind Technology 10) Ion Sorption (a)Zeolite Crystallisation 11Fossil fuel Technology 12)Biomass Technology

			13)Ion 3 Stage exchange:-Desal Process 14)Puraq:-Liquid-Liquid Extraction Process 15)Gravacuotron:-Vaccum Distillation Process
Austria	1) Membrane Distillation	1)Austria-U.S.A-India (a) Wind Technology (b)Fossil fuel Technology 2))Austria-India (a)MultiEffectDistillation+Vapour Compression(V.C) 3) Austria-U.S.A (a)Solar ponds	1)Nuclear Reactor Technology 2)Solar salt ponds 3)Power Towers 4)Paint Focus dishes 5)MultiEffectDistillation+Vapour Compression(V.C) 6)Wind(hydro) Technology 7)Biomass Technology 8)Heat consuming vapour compression(V.C)+ Thermal or Freeze process 9)Multi Stage flash Distillation(MSF)+ Vertical tube Evaporation(VTE)

4. Methodology

In this Section we have presented a brief discussion on the methods which we have used in our implementation.

(A) Overlap Coefficient method to find Co-occurrence

The overlap co-efficient method is a similarity measure that computes the overlap between two sets which is defined as follows:-

$$\text{overlap}(X,Y) = \frac{|X \cap Y|}{\min(|X|,|Y|)}$$

Figure 1

(B) Fuzzy Membership Graph

The **membership function** of a fuzzy set is a generalization of the indicator function in classical sets and were introduced by Zadeh. For any set **X**, a membership function on **X** is any function from **X** to the real unit interval [0,1].The membership function which represents a fuzzy set **~A** is usually denoted by $\mathbf{u_A}$. (*membership degree* of **x** in the fuzzy set **~A)**.

Figure 2:-Membership function of a fuzzy set

(C) Visibility graph

A visibility graph is a graph of intervisible locations, typically for a set of points and obstacles in the Euclidean plane. Each node in the graph represents a point location, and each edge represents a visible connection between them. Visibility graph analysis (VGA) is a method of analyzing the inter-visibility connections within buildings or urban networks and was developed from the architectural theory of space syntax by Turner, and is applied through construction of a visibility graph within the open space of a plan.

(D) *Fuzzy Associative matrix(FAM)*

The concept of fuzzy associative matrix introduced by Kosko expresses fuzzy logic rules in tabular form. These rules usually take two variables as input, mapping cleanly to a two-dimensional matrix, although theoretically a matrix of any number of dimensions is possible. It works in an iterative manner by extracting functional rules from a data set which consisted of input parameters. At each iteration, a sort of "fuzzy contingency table" (the fuzzy associative matrix, FAM) was updated with new information.

5. Algorithm

This section discusses the algorithm to be used in implementation of case study of our concept that is Ocean Water Desalination(OWD).
STEP 1:-Start.
STEP 2:- Categorize the available techniques into first and second generation techniques respectively.
STEP 3:-Try to find out common techniques in both first and second generations.

STEP 4 :- Try to find out similar techniques in both first and second generations.
STEP 5:- Try to find out specialty of a country in first generation.
STEP 6:- Try to find out new concepts for second generation.
STEP7:-Prove existence of similar concept(techniques) using Rough set theory.
STEP 8:-Using Overlap coefficient method,find the co-occurrence number.Algorithm to return Co-occurrence number Let us consider two countries denoted by A and B
O_1:Country("A"); O_2:Country("B");
O_{12}:Overlap function(O_1,O_2); Return (O_{12});
STEP 9:-Define fuzzy membership graph and plot concept strength on it to which class of strength it belongs to and then accordingly we will add edges between the nodes using a colour combination {for very low relation :-red,low relation:-blue,medium relation:-green,high relation:-yellow and very high relation:-black}.
STEP 10:-Draw the visibility graph and state the results in the associative matrix.
STEP 11:-Stop.

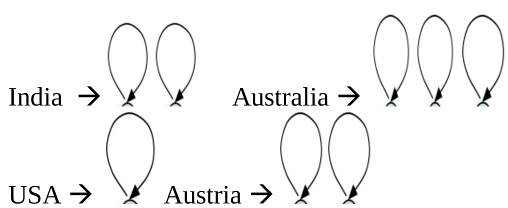
6. Implementing Methodologies in Case Study

This section discusses the implementation of the methodologies in the case study of our concept, that is Ocean Water Desalination (OWD) taken into consideration. To testify the effectiveness of proposed method we have proved the existence of similar concepts using rough set theory and then we have taken the overlap coefficient method to find the co-occurrence number. We have found the concept strength of a Tie(edge) by assuming some nodes (countries) as in input and based on those nodes we will extract knowledge and analyze it.

Table 3: Proving existence of similar concept using Rough set theory

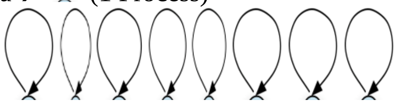
1st GENERATION	1st COMMON PROCESSES	1st NEW CONCEPT SIMILAR	1st SPECIALITY OF A COUNTRY
Object	P_1	P_2	P_3
O_1(INDIA)	6	4	2
O_2(AUSTRALIA)	6	6	3
O_3(USA)	6	5	1
O_4(AUSTRIA)	6	2	2
2nd GENERATION	**2nd COMMON PROCESSES**	**2nd CONCEPT SIMILAR**	**2nd NEW CONCEPT SPECIALITY**
Object	P_4	P_5	P_6
O1(INDIA)	1	7	8
O2(AUSTRALIA)	1	4	5
O3(USA)	1	7	15
O4(AUSTRIA)	1	4	9

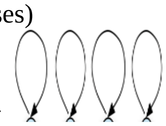
From the above set O_1,O_2,O_3,O_4 are distinguishable objects. In 1st generation $P_1O_1=P_2O_2=P_1O_3=P_1O_4$ are equal they are equivalent structures they represents common processes.$x_p=P=\{P_2\}\rightarrow$ No Equivalent structure indiscremebility $x_p=P=\{P_3\}=\{O_1,O_4\}$ means both P_3O_1 and P_3O_4they have equal number of speciality processes.
$x_{p=}P=\{P_4\}=\{O_1,O_2,O_3,O_4\}=1$ they resemble indistinguishable sets of 2nd generation
$x_p=P=\{P_5\}=1$ means no common processes.
$x_p=P=\{P_6\}=$ not equal number of new concepts
We are considering only 4 nodes O_1, O_2, O_3,O_4 where O_1 represents INDIA,O_2 represents AUSTRALIA.O_3 represents USA,O_4 represents AUSTRIA.
P_1= Common Processes of 1st generation
Overlap $(O_1,O_2)= (O_1\cap O_2)/\min(O_1,O_2)=6/6 = 1$
Rest is all similar for P_1 since O1,O2,O_3,O_4 are similar i.e 6
P_2= Similar Processes of 1st generation
Overlap $(O_1,O_2)= (O_1\cap O_2)/\min(O_1,O_2)= 3/4 =0.75$
Overlap $(O_1,O_3)= (O_1\cap O_3)/\min(O_1,O_3)= 3/4 =0.75$
Overlap $(O_1,O_4)= (O_1\cap O_4)/\min(O_1,O_4)= 1/2 =0.50$
Overlap $(O_2,O_3)= (O_2\cap O_3)/\min(O_2,O_3) = 3/5 =0.60$
Overlap $(O_2,O_4)= (O_2\cap O_4)/\min(O_2,O_4)= 1/2 =0.50$
Overlap $(O_3,O_4)= (O_3\cap O_4)/\min(O_3,O_4) = 0/2 = 0$
P_3= Speciality Processes of 1st generation we will denote special processes with the help of self loop.



P_4= Common Processes of 2nd generation
Overlap $(O_1,O_2)= (O_1\cap O_2)/\min(O_1,O_2)=1/1=1$
Rest is all similar for P_2 since, O1,O2,O_3,O_4 are similar i.e 1
P_5 =New Concept: Similar Processes of 2nd generation.
Overlap $(O_1,O_2)= (O_1\cap O_2)/\min(O_1,O_2) =3/4 = 0.75$
Overlap $(O_1,O_3)= (O_1\cap O_3)/\min(O_1,O_3)=4/7 = 0.57$
Overlap $(O_1,O_4)= (O_1\cap O_4)/\min(O_1,O_4) =3/5 = 0.60$
Overlap $(O_2,O_3)= (O_2\cap O_3)/\min(O_2,O_3)=7/4 = 0.50$
Overlap $(O_3,O_4)= (O_3\cap O_4)/\min(O_3,O_4) =3/5 = 0.60$
Overlap $(O_2,O_4)= (O_2\cap O_4)/\min(O_2,O_4) =0/4 = 0$
P_6= New Concept: Speciality Processes of 2nd generation we will denote special processes

with the help of self-loop.India → Speciality NIL ;Australia→  (1 Process)

USA → 
(8 Processes)

Austria →  (4 Processes)

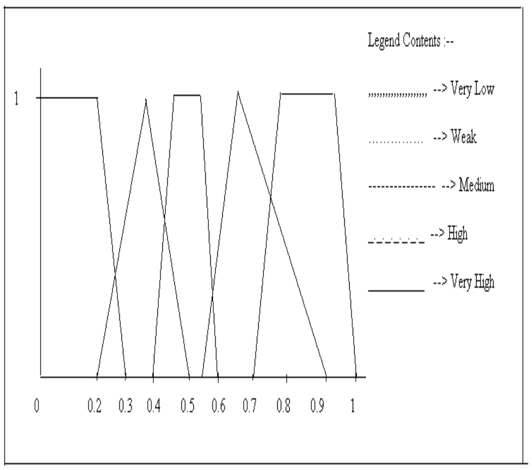


Figure 3: shows fuzzy membership graph which defines strength of a tie(Edge). Five variables are used :-Very Weak, Weak, medium, High, Very High

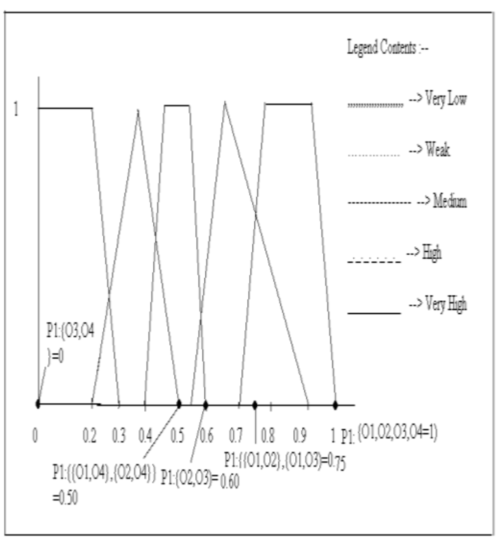


Figure 4:1st generation fuzzy membership graph

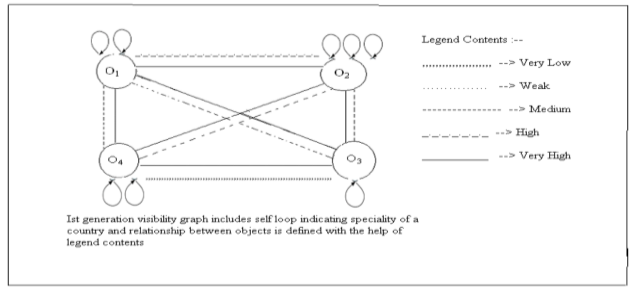


Figure 5 :1st generation visibility graph with help of self loops and legend constraints

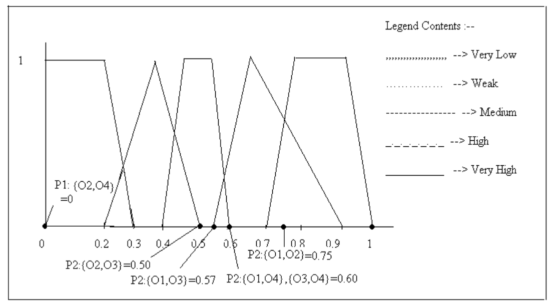


Figure 6: 2nd generation fuzzy membership graph

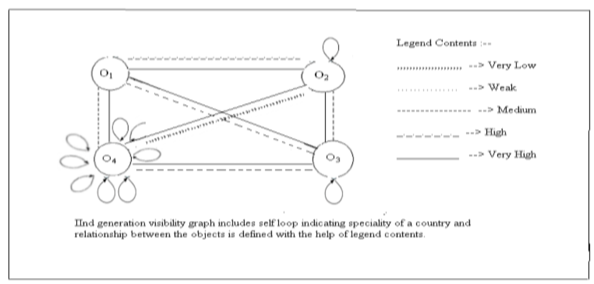


Figure 7: 2nd generation visibility graph with the help of self-loops and legend constraints

7. Result

To obtain the result of the above problem, we make use of associative matrix and finally we would define the five constraints so as to reach the final optimality. To begin with we have defined five constraints as :- effort, progress, safety, efficiency and development. The levels used to markup these constraints are very low, low , medium, high, very high that is according to these 5 levels all the 5 constraints would be classified . The associative matrix that we are using takes account of concept strength of both first generation as well as second generation processes .

Table 4

SECOND	....	FIRST	GENERATION	CONCEPT STRENGTH	....
GENERATION CONCEPT STRENGTH	EFFORT *Very Low*	SAFETY *Low*	EFFICIENCY *Medium*	PROGRESS *High*	DEVELOPMENT *Very High*
Very Low	no effort	no progress	no safety	no efficiency	no development
Low	good effort	good progress	good safety	good efficiency	good development
Medium	very good effort	good progress	very good safety	very good efficiency	very good development
HIgh	excellent effort	very good progress	excellent safety	excellent efficiency	excellent development
Very High	outstanding effort	outstanding progress	outstanding safety	outstanding efficiency	outstanding development

8. Conclusion and future work

This paper describes a novel approach to knowledge extraction using conceptual network analysis. We discussed important issues like extracting unstructured knowledge from structured knowledge and examining relationship between different concepts using web browser. We have proved the existence of similar concept using rough set theory and then defined concept strength of a link using fuzzy approach. In the future we will apply the proposed method to different data (information) available on the internet. We will also focus on other aspects of Artificial Intelligence.

References

[1] http://www-bisc.cs.berkeley.edu/Zadeh-1965.pdf
[2] www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzysets.html
[3] http://www.tandfonline.com/doi/abs/10.1080/03081079008935107
[4] http://www.sciencedirect.com/science/article/pii/S0377221700001673
[5] http://www.questia.com/googleScholar.qst?docId=5000372893
[6] http://www.indaindia.org/intro_desalination.html
[7] geography.about.com/od/waterandice/a/WaterDesalination.html
[8] http://www.scribd.com/doc/34722153/UsingDesalination-Technologies-for-Water-Treatment
[9] www.environment.gov.au/water/.../urban/.../desalination-summary.pdf
[10] www.alazhar.edu.ps/arabic/.../Researches/Deslination%20Thesis.pdf
[11] www.moraassociates.com/.../0712%20Water%20desalination.pdf
[12] http://peakenergy.blogspot.in/2008/05/solving-our- water- problems-desalination.html
[13] desalination.edu.au/wp.../02/NCEDA-ResearchRoadmap- 2011.pdf
[14] See.xidian.edu.cn/faculty/xbgao/FuzzySystem/Introduction/Fuzzy.pdf
[15] www.mrga.netau/documents/MERGA-33-Haiyne Wong.pdf
[16] www.rehab.research.vec.gov/jour/98/35/1/pdf/lafrente.pdf
[17] bcpw.bg.bw.edu.pl/Context/2059/RS-theory.pdf
[18] zsuraj@wsiz.rzeszow.pl
[19] www.cognitiveatlas.org/term/association_learning
[20] http://www2.futurelab.org.uk/projects/why-dont-you/e xplore/word-concept-association
[21] turing.cs.washington.edu/papers/Kcatofpo 5-banko.pdf
[22] deberg,MarkVanKrevald,Marc,Overmars,mark; Schwarzkopf,ofried(2WO),"Chapter15:Visibility Graph ,"Computational Geometry.

The author is a student of B.tech of Guru Premsukh Memorial College of Engineering (GPMCE) in Information Technology. Did my second year training from Sify Corporation,Okhla,Delhi and 3rd year Summer Training from Defence Research and Development organisation(DRDO),Delhi and it was the training at DRDO the research paper was made.

The author is a student of Gurgaaon College of Engineering For Women (GCEW) in Information Technology (IT) . Did my 2nd year training from NIIT institute and 3rd year Summer training from Defense Research and Development Organization(DRDO) and it was the training at DRDO the research paper was made.

Dr. V.K. Panchal is Add. Director at Defence Terrain Research Lab, New Delhi.Associate Member of IEEE (ComputerSociety) and Life Member of IndianSociety of Remote Sensing.He has done Ph.D in Artificial Intelligence and is currently working as Scientist 'G'atDRDO,Delhi.Chaired sessions & delivered invited talks at many national & international conferences.