SYSTEMATIC REVIEW OF EFFORT ESTIMATION AND COST ESTIMATION

Qamar Alam[1], Preeti Bhatia[2] and Saoud Sarwar[3]

[1]Lecturer, Institute of Management studies, Roorkee
Email: qamaralam83@gmail.com
[2]Lecturer, N.G.F.C.E.T. , Palwal
Email: bhatiapreet@gmail.com
[3]HOD, AFSET, Dhauj, Faridabad
Email: saoud.phd@gmail.com

ABSTRACT

Accurate software effort and cost estimation is critical for developers. The costs of development are primarily the costs of the effort involved, so the effort computation is used in both the cost and the schedule estimate. For most projects, the dominant cost is the effort cost. In this paper authors study and analyze effort and cost estimation using details of various software projects. Authors collected 15-20 software projects details such as Backend, Frontend, No. of front end forms, No. of Pages documented etc. and then predict effort and cost for that software projects. We use various methods for estimation and the standardized COCOMO Intermediate Model.

***Keywords*:** Effort estimation, Cost estimation, methods, model, software project details.

INTRODUCTION

Software estimation is important because it decides the execution boundaries of the project. It also tells whether or not the project is considered feasible in terms of cost and the ability to meet customer's requirement. The purpose of the estimation is that the project manager need to be prepare quotation before completing the design. Estimation includes various estimations such as cost, effort, time, size. Effort estimation is the key of cost and schedule estimation [1]. To estimate the effort the analyses of requirements, design of front end forms, no. of pages documented etc.. tasks earlier performed. Analyses of the project data details must be carried out before using any available estimation method. For any project, it is best to approach the estimation from more than angle using more than one model. If the estimates arrived at from the all models same or almost similar, then we can have a greater confidence in the estimation. In this paper we have taken different projects and estimate their effort using various models. We also used fuzzy logic approach.

Details of various software projects

Authors collected general data and information from various organization about the projects. We have not gone into technical details in deep. The main aim was to utilize the data and to

estimate effort required for the completion of software project .The data given was also used to estimate effort by other available methods for comparison.

Table 1.

Sl. No	Project Title.	Organization
1	Docphin	Smart data Enterprises
2	Oder management System	Smart data Enterprises
3	Booktup	Smart data Enterprises
4	Pahali Job Portal	Alleviatetech, Delhi
5	Teacher Evaluation Software System	USTECH INDIA INC New Delhi
6	Service Connect Product	Binary Info technology Noida
7	Libsoft	Ats infotech Pvt. Ltd. Noida
8	Institute Information System Portal	Gmass Pvt .Ltd , Gaziabad
9	Council of artech	Ats infotech Pvt. Ltd. Noida
10	Superb Star	Ats infotech Pvt. Ltd. Noida
11	Load Balancing	Gurunank Trust
12	Sales Tax Management	Ats infotech Pvt. Ltd. Noida
13	Customer relation ship	Krishvatika , bhiwadi,
14	Erpassid	Single system solution, bhiwadi
15	Erpproprt	Single system solution, bhiwadi
16	Sezjobs.com	Acreaty Management Consultant Pvt. Ltd
17	Portfolio Management System	EXCEL InfoTech Pvt. Ltd
18	Jobshopee.com	Acreaty Management Consultant Pvt. Ltd
19	Gharondaa.com	Acreaty Management Consultant Pvt. Ltd
20	Human Resource Management System	EXCEL InfoTech Pvt. Ltd

Table 2(i) contains the genral details of software projects

Sl. No	Project Title.	No of Persons in Team.	Front end	Database	KLOC
1	Docphin	3	C# .NET	SQL	6
2	Oder management System	2	ASP. NET	SQL	8
3	Booktup	2	C# .NET	SQL	15
4	Pahali Job Portal	5	ASP. NET	SQL	20
5	Teacher Evaluation Software System	4	Java	ORACLE9I	27
6	Service Connect Product	3	Java	ORACLE9I	38
7	Libsoft	4	VB 6.0	SQL	45
8	Institute Information System Portal	5	ASP. NET	SQL	60
9	Council of artech	6	ASP .NET	SQL	80
10	Superb Star	3	ASP. NET	SQL	90
11	Load Balancing	2	Java	ORACLE9I	100
12	Sales Tax Management	4	Java	ORACLE9I	120
13	Customer relation ship	3	PHP	MY SQL	140
14	Erpassid	2	ASP. NET	SQL	155

Table 2(i) contains the general details of software projects(Contd..)

Sl. No	Project Title.	No of Persons in Team.	Front end	Database	KLOC
15	Erpproprt	2	C# .NET	SQL	170
16	Sezjobs.com	2	ASP. NET	SQL	180
17	Portfolio Management System	3	C# .NET	SQL	200
18	Jobshopee.com	2	ASP. NET	SQL	220
19	Gharondaa.com	3	ASP. NET	SQL	225
20	Human Resource Management System	4	C# .NET	SQL	250

Table 2(ii)

Sl. No	Project Title.	Estimated time	Actual time	NFF	NPD	Tables
1	Docphin	6	6	15	20	4
2	Oder management System	5	5	22	15	10
3	Booktup	10	13	18	30	8
4	Pahali Job Portal	16	15	34	45	9
5	Teacher Evaluation Software System	14	16	10	25	12
6	Service Connect Product	20	20	60	80	30
7	Libsoft	30	32	70	95	25
8	Institute Information System Portal	50	50	220	160	40
9	Council of artech	36	38	90	60	36
10	Superb Star	19	21	80	70	20
11	Load Balancing	70	75	45	50	28
12	Sales Tax Management	60	60	50	75	30
13	Customer relation ship	100	110	330	220	70
14	Erpassid	15	18	25	48	12
15	Erpproprt	6	7	20	40	14
16	Sezjobs.com	16	18	35	64	22
17	Portfolio Management System	140	145	660	480	120
18	Jobshopee.com	24	26	30	66	16
19	Gharondaa.com	20	20	40	54	22
20	Human Resource Management System	90	90	110	90	55

Analysis of data

It can be observed that the tables 1, 2(i) and 2(ii) that there are vast variations in data. Kloc varies from 6-250. Team size also varies. The important one that is the estimation time for the completion of the project varies from 6 PM to 140 PM. the no. of front end forms also varies from 10-660. The project 'Portfolio Management System' have large no. of front end forms i.e 660 ant its estimated time is 140 PM. Based on the data it can be observed that actual effort is not proportional to either Kloc or no. of front end forms to be developed. Using Kloc of various projects we evaluate efforts by using methods COCOMO model and showed their effort values in table.

Based on the data it can be observed that actual effort is not proportional to either Kilo lines of code or Number of front end forms developed. It can be easily analyzed that effort values are not only the function of size that is KLOC,but also of many other parameters such as complexity of design, experience and maturity level of developers, environment, team cohesion, etc and most of these parameters are intangibles.[4]

Sl. No	Kloc	Orgainic	Semi	Embeded
1	6	18.8995	30.9089	33.6563
2	8	25.5644	36.9626	47.5328
3	15	49.4638	74.7351	101.0637
4	20	66.9097	103.1468	142.7322
5	27	91.6905	144.3543	204.6079
6	38	131.2699	211.670399	308.3375
7	45	156.7709	255.7999	377.6948
8	60	212.0563	353.04649	533.4178
9	80	286.8381	487.2628	753.3445
10	90	324.5989	555.9735	867.7714
11	100	362.5705	625.6082	984.4659
12	120	439.06902	767.3358	1225.4724
13	140	516.2106	911.9392	1474.4827
14	155	574.4348	1022.0543	1666.0349
15	170	632.9419	1133.4576	1861.3359
16	180	672.0918	1208.3916	1993.4854
17	200	750.0713	1359.7406	2262.1534
18	220	829.729	1512.9197	2536.2571
19	225	849.5405	1551.4826	2605.584
20	250	948.8919	1745.8033	2956.7464

Summary of effort estimation using different methods for the projects

Project→ Method ↓	1	2	3	4	5
Actual Effort	6	5	13	15	16
Cocomo Intermediate(Organic)	18.8995	25.5644	49.4638	66.9097	91.6905
Cocomo Intermediate(Semidetached)	30.9089	36.9626	74.7351	103.1468	144.3543
Cocomo Intermediate(Embedded)	33.6563	47.5328	101.0637	142.7322	204.6079
Bailey-Basil	11.334	13.645	22.388	29.0785	38.896
Barry-Boehm	20.999	28.404	54.959	74.3415	101.878
Doty	34.515	46.647	90.086	121.750	166.697
Waslton Felix	26.553	34.499	61.128	79.421	104.362

Project→ Method ↓	6	7	8	9	10
Actual Effort	20	32	50	38	21
Cocomo Intermediate(Organic)	131.2699	156.7709	212.0563	286.8381	324.5989
Cocomo Intermediate(Semidetached)	211.67039	255.7999	353.04649	487.2628	555.9735
Cocomo Intermediate(Embedded)	308.3375	377.6948	533.4178	753.3445	867.7714
Bailey-Basil	55.144	65.901	89.829	123.235	140.472
Barry-Boehm	145.855	174.189	235.618	318.079	360.665
Doty	238.410	284.580	384.605	519.788	537.401
Waslton Felix	142.431	166.121	215.834	280.424	312.150

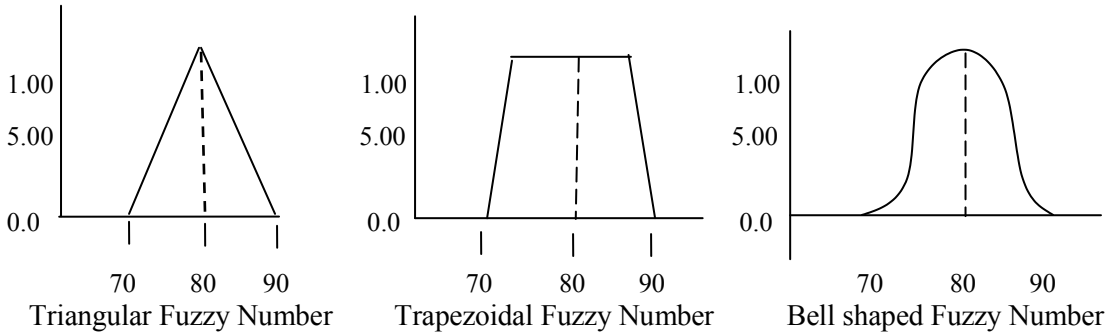
Project→ Method ↓	11	12	13	14	15
Actual Effort	75	60	110	18	7
Cocomo Intermediate(Organic)	362.5705	439.06902	516.2106	574.4348	632.9419
Cocomo Intermediate(Semidetached)	625.6082	767.3358	911.9392	1022.0543	1133.4576
Cocomo Intermediate(Embedded)	984.4659	1225.4724	1474.4827	1666.0349	1861.3359
Bailey-Basil	158.018	193.940	230.836	259.075	287.756
Barry-Boehm	402.856	487.854	573.567	638.260	703.268
Doty	656.585	794.6835	933.872	925.1834	1032.672
Waslton Felix	343.560	405.556	466.637	511.923	556.816

Projects→ Method ↓	11	12	13	14	15
Actual Effort	75	60	110	18	7
Cocomo Intermediate(Organic)	362.5705	439.06902	516.2106	574.4348	632.9419
Cocomo Intermediate(Semidetached)	625.6082	767.3358	911.9392	1022.0543	1133.4576
Cocomo Intermediate(Embedded)	984.4659	1225.4724	1474.4827	1666.0349	1861.3359
Bailey-Basil	158.018	193.940	230.836	259.075	287.756
Barry-Boehm	402.856	487.854	573.567	638.260	703.268
Doty	656.585	794.6835	933.872	925.1834	1032.672
Waslton Felix	343.560	405.556	466.637	511.923	556.816

Project→ Method ↓	16	17	18	19	20
Actual Effort	18	145	26	20	90
Cocomo Intermediate(Organic)	672.0918	750.0713	829.729	849.5405	948.8919
Cocomo Intermediate (Semidetached)	1208.3916	1359.7406	1512.9197	1551.4826	1745.8033
Cocomo Intermediate (Embedded)	1993.4854	2262.1534	2536.2571	2605.584	2956.7464
Bailey-Basili	291.842	346.314	386.156	396.210	447.002
Barry-Boehm	746.768	834.125	921.921	943.933	1054.355
Doty	1214.959	1356.656	1499.0219	1534.710	1713.699
Waslton Felix	586.544	645.565	704.0572	718.603	790.913

Effort estimation model1 using fuzzy technique

A fuzzy number is a quantity whose value is imprecise, rather than exact as in the case of ordinary single valued numbers. Any fuzzy number can be thought of as a function, called membership function, whose domain is specified, usually the set of real numbers, and whose range is the span of positive numbers in the closed interval [0, 1]. Each numerical value of the domain is assigned a specific value and 0 represents the smallest possible value of the membership function, while the largest possible value is 1. The curve in figure 1 is a triangular fuzzy number, the curve in figure 2 is a trapezoidal fuzzy number, and the curve in figure 3 is bell shaped fuzzy number.[3]

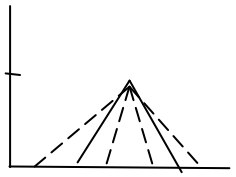


Triangular Fuzzy Number Trapezoidal Fuzzy Number Bell shaped Fuzzy Number

Fuzzy Effrot Estiation

We take the input as size S, output as effort E, then a triangular fuzzy number, K(S), is defined as follows:

$$K(S)=\begin{cases} 0, & S \le 0 \\ \dfrac{S-\alpha}{m-\alpha}, & \alpha \le S \le m \\ \dfrac{\beta - S}{\beta - m}, & m \le S \le \beta \\ 0, & S \ge \beta \end{cases}$$



Representation of k(s)

K(S) = TFN (α, m, β), F = 0.1

K(S) = TFN (α, m, β), F= 0.2

K(S) = TFN (α, m, β), F = 0.3

Where α, m and β are the parameters of the membership function, K(S),m is the model value, α and β are the right and left boundary respectively.

Let (m, 0) divides, internally, the base of the triangle in ratio k : 1, where k is real positive number. [3]

So that $m = \frac{\alpha + k\beta}{k+1}$

As by definition of fuzziness

$F = \frac{\beta - \alpha}{2m}$

So, $\alpha = 1\text{-}2kF/\ k+1\ *m$

$\beta = (1 + 2k\ /F+1)* m$

Table 1, gives values of α and β for F=0.1, 0.2 and 0.3 for various values of k using equations (4) and (5), where m, size estimate in KDLOC.

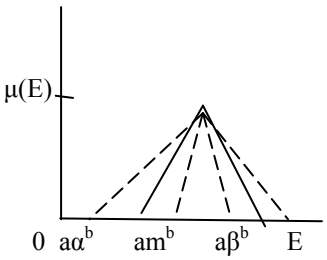
Similarly, the TFN μ(E) is defined as,

$0, E \le a\alpha^b$

$\frac{(E/a)^{1/b} - \alpha}{m-a}, \quad a\alpha^b \le E \le a m^b$

$\frac{\beta - (E/a)^{1/b}}{\beta - m}, \quad a m^b \le E \le a\beta^b$

$0, E \ge a\beta^b$



μ (E) = TFN (a α ^{b}, a m ^{b}, β^b), F=0.1

μ (E) = TFN (a α ^{b}, a m ^{b}, β^b), F=0.2

μ (E) = TFN (a α ^{b}, a m ^{b}, β^b), F=0.3

Table1. Values of α and β for different values of F and k

F	k=1	k=2	k=3
0.1	α=0.9m β=1.1m	α=(2.6/3)*m β=(3.2/3)*m	α=.85m β=1.05m
0.2	α=0.8m β=1.2m	α=(2.2/3)*m β=(3.4/3)*m	α=.7m β=1.1m
0.3	α=0.7m β=1.3m	α=0.6m β=1.2m	α=0.55m β=1.15m

Defuzzification:

The single output, fuzzy estimate of E,can be computed as a weighted average of the optimistic($a\alpha^b$), most likely (am^b) and pessimistic estimate($a\beta^b$)

$$\text{Effort } E = \frac{w_1(a\alpha^b) + w_2(am^b) + w_3(a\beta^b)}{w_1+w_2+w_3}$$

Where w1, w2 and w3 are weights of the optimistic, most likely and pessimistic estimate respectively. Maximum weight should be given to the most expected estimate.

Proposed Model:

The single output, fuzzy estimate of E, can be computed as a weighted average of the optimistic ($a\alpha^b$), most likely (am^b) and pessimistic estimate($a\beta^b$)

$$\text{Effort } E = \frac{w_1(a\alpha^b) + w_2(am^b) + w_3(a\beta^b)}{w_1+w_2+w_3}$$

Where w1, w2 and w3 are weights of the optimistic, most likely and pessimistic estimate respectively. Maximum weight should be given to the most expected estimate.

Where a = 3.41, b = 0.795 and m represents the size in KLOC

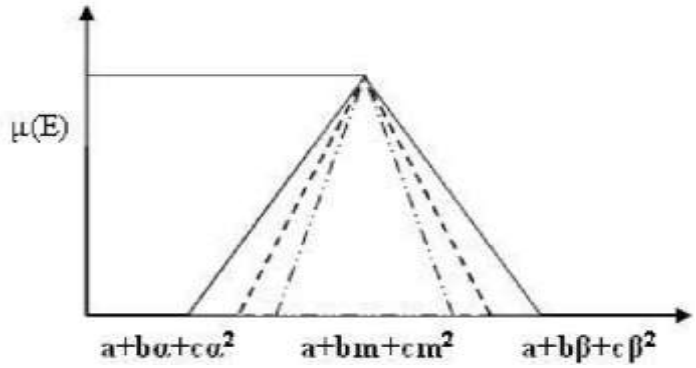
$$\alpha = \left(1 - \frac{2KF}{K+1}\right)m$$

$$\beta = \left(1 + \frac{2F}{K+1}\right)m$$

Here K=1, F=0.2, w1=1, w2=10 and w3=10, α=0.8m and β=1.2m are arbitrary constants. The effort is estimated in man months (MM).

Polynomial regression model effort estimation using fuzzy

$$\mu(E) = \begin{cases} 0 & if\ E \le a + b\alpha + c\alpha^2 \\ \dfrac{\left(-b+\frac{\sqrt{b^2+4cE}}{2c}-\alpha\right)}{m-\alpha} & if\ a + b\alpha + c\alpha^2 \le E \le a + bm + cm^2 \\ \dfrac{\left(\beta-\frac{-b+\sqrt{b^2+4cE}}{2c}\right)}{\beta-m} & if\ a + bm + cm^2 \le E \le a + b\beta + c\beta^2 \\ 0 & if\ E \ge a + b\beta + c\beta^2 \end{cases}$$



μ (E) = TFN($a+b\alpha+c\alpha^2$, $a+bm+cm^2$, $a+b\beta+c\beta^2$) F = 0.1

μ (E) = TFN($a+b\alpha+c\alpha^2$, $a+bm+cm^2$, $a+b\beta+c\beta^2$) F = 0.2

μ (E) = TFN($a+b\alpha+c\alpha^2$, $a+bm+cm^2$, $a+b\beta+c\beta^2$) F = 0.3 [2]

Defuzzification:

The output, fuzzy estimate of E, can be computed as a weighted average of the optimistic ($a+b\alpha+c\alpha^2$), most likely ($a+bm+cm^2$),and pessimistic estimate ($a+b\beta+c\beta^2$) Fuzzy effort estimate (E) is given as

$$E = \frac{w1(a + b\alpha + c\alpha^2) + w2(a + bm + cm^2) + w3(a + b\beta + c\beta^2)}{w1 + w2 + w3}$$

Where w1, w2 and w3 are weights of the optimistic, most likely and pessimistic estimate respectively. Maximum weight should be given to the most expected estimate.

Where a = 135.918, b = -0.769 and c = 0.004 and m represents the size in KLOC

$$\alpha = \left(1 - \frac{2KF}{K+1}\right)m$$

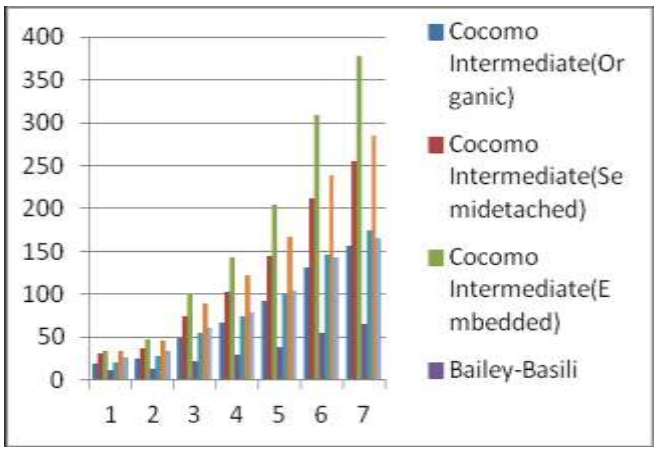
$$\beta = \left(1 + \frac{2F}{K+1}\right)m$$

Here K=1, F=0.3, w1=50, w2=2 and w3=1 are arbitrary constants. The effort is estimated in man months (MM).

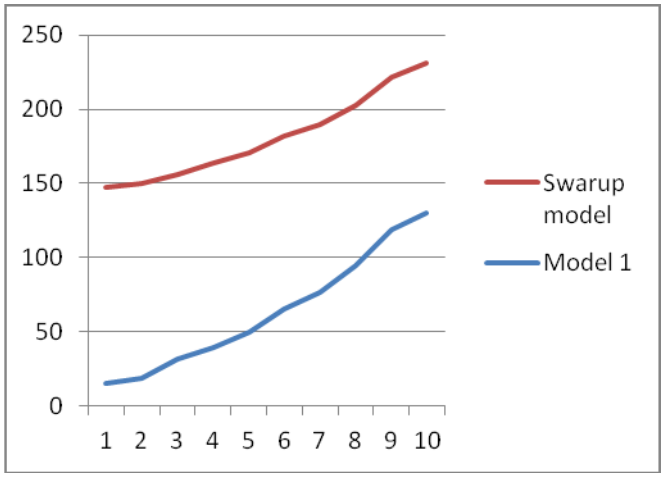
Effort Estimation using Fuzzy Technique

Project↓ Method →	Model 1	Swarup Model
1	15.1133	132.2534
2	18.9770	131.0747
3	31.31277	124.2875
4	39.3593	124.4383
5	49.96469	120.9440
6	65.5627	116.3389
7	76.1070	113.1279
8	94.26691	107.9335
9	118.4910	103.1237
10	103.1237	101.1705
11	141.4910	99.9844
12	153.7067	99.2170
13	184.4884	97.5083
14	213.1065	103.5889
15	215.7426	102.918
16	225.7223	109.4543
17	245.4979	113.2659
18	264.8226	127.8221
19	269.5964	136.7674
20	293.1509	145.9783

Comparision between different effort model model



Comparision between model 1 and Swarup model



CONCLUSION

In this paper we estimated the effort using various model such as Cocomo Intermediate (Organic, Semidetached, Embedded),Bailey-Basili, Barry-Boehm, Doty, Waslton Felix. Authors taken various project details and then estimated the effort by using various models. Authors also used the fuzzy technique for the estimation. As we known choice of estimation method and level of accuracy depends on many factors and the information associated with projects. It is absolutely necessary to use more than one method of estimation and analyze the values for decision making. Authors uses the fuzzy model and one of the proposed model. This paper also contains the two comparison graphs, first graph shows the comparison between fuzzy model1 and Swarup model, another graph shows the comparison among various effort estimation models.

REFERENCES

1. Swapna Kishore, rajesh Naik, Software Requirements and Estimation, Tata McGraw-Hill
2. Fuzzy logic for Software Effort Estimation Using Polynomial Regression as Firing Interval , Aravind Mandala et al, Int. J. Comp. Tech. Appl., Vol 2 (6),1843-1847
3. Optimization Criteria for Effort Estimation using Fuzzy Technique, CLEI ELECTRONIC JOURNAL, VOLUME 10, NUMBER 1, PAPER 2, JUNE 2007
4. AM Nageswara Yogi, Mala V Patil, 'Software Effort estimation models and performance analysis with case studies', International Journal of Computer Applications in Engineering, Technology and Sciences(IJ-CA-ETS), Vol. 1 no 2, , 2008.
5. Barry Boeham, Chris Abts, Sunitha Chulani 'Software Development Cost Estimation Approaches -A survey', Annalsof software Engineering,Vol.10, no.1-4 ,pp 177-205,2000
6. Alaa F. Sheta, 2006, Estimation of the COCOMO Model Parameters Using Genetic Algorithms for NASA Software Projects. Journal of Computer Science 2(2):118-123.

7. Bailey, J.W. and Basili, 1981. A Meta model for software development resource expenditure. Proc. Intl. Conf. Software Engineering, pp: 107-115
8. Boehm, B.,1981. Software Engineering Economics, Englewood Cliffs, NJ. Prentice-Hall.
9. L.A. Zadeh, 2002, From Computing with numbers to computing with words-from manipulation of Measurements to manipulation of perceptions, Int. J. Appl. Math.Comut.Sci.,2002, Vol.12,No.3,307-324.
10. Jose Galindo".Handbook of Research in Fuzzy Information Processing In Databases", Information science Reference, 2008..
11. B.Boehm, Software Engineering Economics Englewood Cliffs, NJ, Prentice Hall, 1981.
12. B. Boehm., Cost Models for Future Life Cycle Process: COCOMO2. Annals of Software Engineering. 1995
13. Pankaj jalote, ―An Integrated Approach for Software Engineering.‖ , Third Edition. ISBN: 978-81-7319-702-4.