Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluation of Critical Factors Affecting The Efficiency of Searching Techniques

Olabiyisi S.O*
C.S.E. Dept, LAUTECH, Nig.
Nigeria

Aladesae T.S. , Oyeyinka F.I. , Oladipo Oluwasegun
C.I.T.M., YabaTech, Nigeria
Nigeria

Abstract— Searching for data is one of the fundamental fields of computing. Often, the difference between a fast and slow one is the use of good algorithm for the data set. The efficiency of searching techniques used in a task will determine how fast such task can be completed. In this work, the efficiency of linear and binary searching techniques in terms of running time, memory usage and the numbers of comparison was evaluated. The search time, memory used and numbers of comparison were used as decision variables to evaluate their efficiencies. Experimental results for the decision variables were generated from a software tool written using Java programming language in which the arrays of number searched were varied for the four different searching techniques of linear and binary. The results were subjected to Factor analysis to test the level at which each factor affects the searching techniques. Eigenvalues were used to indicate how well each of the extracted factors fits the data from the experimental results. The result showed that numbers of comparison contributed; 84.459% and 72.876% for linear and binary searching techniques respectively. The time taken came second, contributing 15.538% and 27.041% while memory used was the least of all contributing 0.003% and 0 .082% for linear and binary search respectively. It can be concluded that numbers of comparison is the most critical factor affecting the searching techniques. Binary search is the most efficient of all the searching techniques considered.

Keywords— Factor Analysis, Searching techniques, Decision Variables, Eigenvalue, Principal Components, Communality, Correlation

1. Introduction

Searching for data is one of the fundamental operations in computing. Often, the difference between a fast program and a slow one is the use of a good search algorithm for the dataset. Searching is a process of discovering something by examining, somebody for concealed items or computer (files, disk, database, or network) for information [12] while an algorithm is a tool for solving a well-specified computational problem [19]. Search is said to be successful if item does appear in data and unsuccessful otherwise. Google claims it searches greater than 3 billion web pages (Punjab university library). We have a huge number of search algorithms available for us today, many programmers try to select the best out of the available alternatives by making a guess of the input data; all the search algorithms work well on a set of data. Although may have a set of data for which the performance is bad. The concept of efficiency (or complexity) is important when comparing algorithms for operations, like searching, that are repeated frequently. Efficiency of searching techniques used in a task will determine how fast such task can be completed**.** This research work determines the efficiency of searching techniques in terms of these three factors time taken to search data, memory usage and number of numbers of comparison using statistical analysis. Statistical analysis has been a great tool in research, and which cut across all field of study. The most widely used of the statistical tool is Statistical Package for the social Science (SPSS) which have different statistical tool like ANOVA, paired t-test, chi square test, exploratory factor analysis (EFA), survivals, forecasting, missing value analysis, multiple imputation, complex samples, quality control and so on which can be applied on the dissertation. It helps researcher to make the research work more scientific and reliable.

In this work, factor analysis which is a branch of dimension reduction method is used. It seeks to discover if the observed variables can be interpreted in a more compressed form with few numbers of variables called factors. The numbers of factors used depend on the researcher and the expected results. It should be noted that in statistical packages, the volume of data used for factor analysis must be a minimum of fifty (50). The use of lesser volume of data might likely not give a good result.

2. Literature Review

The searching algorithms of Linear and binary got from various sources like textbooks, internet, articles and journals were implemented using Java programming language and the data searched were randomly generated matriculation numbers. The types of search algorithm: **Linear Search Algorithm:** Linear Search is simply searching through a line of objects in order to find a particular item. Assume an array called list, Check the first item, then the second item, and so on until you find the target item or reach the end of the list, that's **linear** (or **sequential**) search. **Binary Search Algorithm:** Is an algorithm for locating the position of an element in a sorted list. It inspects the middle element of the sorted list: if equal to the sought value, then the position has been found; otherwise, the upper half or lower half is chosen for further

searching. Statistical Package for Social Sciences (Factor Analysis) by Principal Component Analysis was used to check the efficiency of the search algorithm by checking the numbers of comparison, Time and memory utilized by the search algorithms.

3. Related Works

Reference [12] presented a paper on an Exploratory study of factors affecting the efficiency of searching Techniques using Linear, Binary, Fibonacci and Interpolative in other to study their efficiency in terms of the running time and memory usage. The experimental results for the variables were generated from an algorithm implemented in JAVA in which the amounts of number sorted were varied for the different searching techniques. Factor analysis by principal components of the obtained experimental data was carried out using Statistical package for Social Scientists (SPSS) for the purpose of estimating the contribution of each factor to the success of the searching algorithms and one factor was extracted. Further statistical analysis was carried out to generate eigenvalue of the extracted factor. The eigenvalue forms the basis for estimating the contribution of the extracted factor. A system of linear equation was used to estimate the assessment of each assessor of the sorting techniques is proposed. Reference [14] presented a paper on the overview of Factor Analysis. The paper stated the different between Exploratory Factor Analysis (EFA), Principal Component Analysis (PCA) and Confirmatory factor analysis (CFA). Reference [8] in their paper, investigated sorting problems and their solutions. The paper explained the most popular algorithms that are useful for sorting lists. They are: Bubble sort, Selection sort, Insertion sort, Shell sort, Merge sort, Heapsort, Quicksort and Bucket sort. Algorithms were represented with perfect descriptions and also tried to indicate their computational complexity in the worst, middle and best cases. At the end, implementation code was placed. Reference [4] in their paper,"Factor Analysis of the Performance Indices and Communications Technology Projects in Public Sector of the Nigerian Economy". They proposed fifty three performance indices of ICT projects and a tool for evaluating them. The ICT projects in the public sector of Nigerian economy were surveyed and completed questionnaires were received from these entire sector. The data collected were subjected to factor analysis by principal components using statistical package for social science (SPSS). Reference [9] "Development and Factor Analysis of Questionnaire to measure Patient Satisfaction with Injected and Inhaled Insulin for Type 1 Diabetes". Used exploratory factor analysis to evaluate the data collated from the questionnaire. Reference [5], Factor Analysis of the Effects of Academic Staff Profile on the Investment Portfolio of a University". The research being reported in this paper take a study of the three factors teaching, research and community development as a major investment foci of a university loaded a number of academic profile related decision variables. The result obtained from the factor analysis provides a basis for deriving a system of equations desirable for evaluating the contributions of each academic staff to teaching, research and community development. Reference [3] "An empirical Model for Information Retrieval System Evaluation: The User's perspective". This paper took an in depth study to one of the major issue hindering easy access to information retrieval in other to come up a better and efficient paradigm with much ease using a mathematical model and factor analytic method. Reference [10] "Factor Analysis of Post-Implementation Review of Student Information Systems in Nigeria". In their paper, research was carried out in four Universities in other to evaluate the critical factors that may enhance the success and failures encountered after implementation of the student information systems. An on-line questionnaire was designed using Macromedia Dreamweaver, PHP and Mysql to gather data from the users of the system. Analysis of the data gathered was carried out using SPSS 10 and Factor analysis was used for the assessment of system's attributes identified.

The aim of this research work is to discover critically factors affecting the efficiency of Searching Techniques among are searching time, memory usage and the numbers of comparison.

The objectives of this research include:

i. To carry out an exploratory study of critical factors affecting the efficiency of Binary and Linear algorithms.
ii. To conduct experiments in order to determine the efficiency of these searching techniques mentioned in (i) above in term of execution time, memory used and the number of comparison
iii. To subject the result obtained in (ii) above to factor analysis by Statistical Package for the Social Sciences(SPSS)

4. MATERIALS AND METHODS

Factor analysis by principal component was adopted in the evaluation of the efficiencies of these searching techniques. The experimental result was analysed using SPSS (Statistical package for social sciences). To generate eigenvalue of the extracted factor and further statistical analysis were carried out. The eigenvalue form the basis for finding the contribution of the extracted factors. The following statistical tools were used for the purpose of achieving the above objectives:

a. Descriptive statistics
b. Correlation matrix
c. Barlets's test and Kaiser-Mayer Olkin(KMO).
d. Communality
e. Initial factor loadings.
f. Eigenvalue.

4.1 This chapter presents numerical analysis and interpretation of data obtained from the Factor Analysis by Principal Component and discussion on how the three factors considered have effect on the search Techniques. The factors are as follows:

1. Numbers of Comparison

2. Time Taken (nano seconds)
3. Memory Used (Bits)

The above three factors were being considered in the four search algorithms (Linear and Binary search) in other to know the most critical out of them all. Once SPSS is activated the user is presented with the SPSS Screen. A new Data-file is opened automatically. This file contains two Windows, a Data View and a Variable View.

Variable View: The variable view contains a grid of rows and columns. In this Window the rows represent the variables in the analysis and the columns are used to define the characteristics of each variable of the factors considered to affect search algorithm

Data View: This Data View consists of a grid of Columns and Rows, similar to a spreadsheet. The Columns represent Variables and the Rows represent Cases. The numerical data from the factors affecting the search techniques are typed in the window

4.2 Result : The tables below give the summary and analysis of the data got from factor analysis by principal component on each factor affecting the efficiency of the search algorithm

Descriptive Statistics

Table: 3.1a Linear search

	Mean	Std. Deviation	Analysis N	Missing N
Numbers of Comparison	1941.00	3.742	6	0
Time Taken	2503598.67	4378650.844	6	0
Memory Used	1104674.67	322284.063	6	0

Table: 3.1b Binary search

	Mean	Std. Deviation	Analysis N	Missing N
Numbers of Comparison	14.83	3.061	6	0
Time Taken	480088.00	338356.600	6	0
Memory Used	1098693.33	328465.539	6	0

The descriptive statistic present s the mean and standard deviation of the raw score of each performance indices given by each assessor; data collected exhibits the mean and standard deviation of the rating of the impact of numbers of comparison, time taken and memory used on the efficiency of the searching techniques by the experimental resulted generated. The mean and standard deviation of the rating on numbers of comparison, Time taken and memory used by linear search Techniques (Table 4.1a) are 1941.17 and 3.742, 2503598.67 and 4378650.844, respectively. For Binary(Table 4.1b), the mean and standard deviation of the rating on numbers of comparison, Time taken and memory used are 14.83 and 3.061, 480088.00 and 338356.600, 1098693.33 and 328465.539 respectively.

Kaiser-Meyer Olkin (KMO) and Bartlett's Test: The KMO measures the sampling adequacy which should be greater than 0.5 for a satisfactory factor analysis to proceed. Looking at the table 4.2a below, the KMO measure is 0.522. From the same table, it can be seen that the Bartlett's test of sphericity is significant. That is, its associated probability is less than 0.05. In fact, it is actually 0.000. This means that the correlation matrix is not an identity matrix and highly corrected to provide a reasonable basis for factor analysis. The same is applicable to table 4.2b

Table: 3.2a Linear search

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.522
Bartlett's Test of Sphericity	Approx. Chi-Square	28.876
	Df	3
	Sig.	.000

Table: 3.2b Binary search

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.545
Bartlett's Test of Sphericity	Approx. Chi-Square	17.193
	Df	3
	Sig.	.001

Communality: The tables below reports communalities of the performance indices generated for searching Techniques with PCA as the extraction method which indicate the amount of variance in each variable that is accounted for.

	Initial	Extraction
Numbers of Comparison	1.000	1.000
Time Taken	1.000	1.000
Memory Used	1.000	1.000

Component Correlation Matrix: The correlation matrix computes coefficient of the columns of a matrix. Row i and column j of the correlation matrix is the correlation between column i and column j of the original matrix. The diagonal elements of the correlation matrix will be 1 since they are the correlation of a column with itself. The correlation matrix is also symmetric since the correlation of column i with column j is the same as the correlation of column j with column i. it can be achieved by formulating a linear equation of the form:

$$C_{i,j} - \sum_{k=1}^{3} b_{k,j} s_{1,k} \quad i - 1,2,..n_j \quad \text{j-1} \qquad \text{eqn (4.3)}$$

Where C_{ij} represents the contribution of i^{th} assessor to j^{th} factor; $b_{k,j}$ represents the component score coefficient of k_{th} decision variable for j^{th} factor; $S_{i,k}$ represents score of i^{th} assessor for k_{th} decision variable and n represents the number of sampled assessors. $S_{i,k}$ is estimated by

$$S_{i,k} = A + \frac{(x_1 + y_i)}{d_1} \qquad \text{eqn (4.4)}$$

Where A represents the allowable minimum raw score for decision variable; in this instance, it is 1; x_i represents the raw score for decision variable; y_i represents the mean of the raw scores of i^{th} decision variable. For each sampled Assesor, the system of linear equations for the single extracted factor can be represented as follows;

$$b_{1,1}S_{i,1} + b_{2,1}S_{i,2} + b_{4.1}S_{i,4} = C_{i,1} \qquad \text{eqn (4.5)}$$

Scree Plot: It Plots all the eigenvalues in their decreasing order. The plot looks like the side of a mountain, and "scree" refers to the debris fallen from a mountain and lying at its base. So the scree test proposes to stop analysis at the point the mountain ends and the debris (error) begins

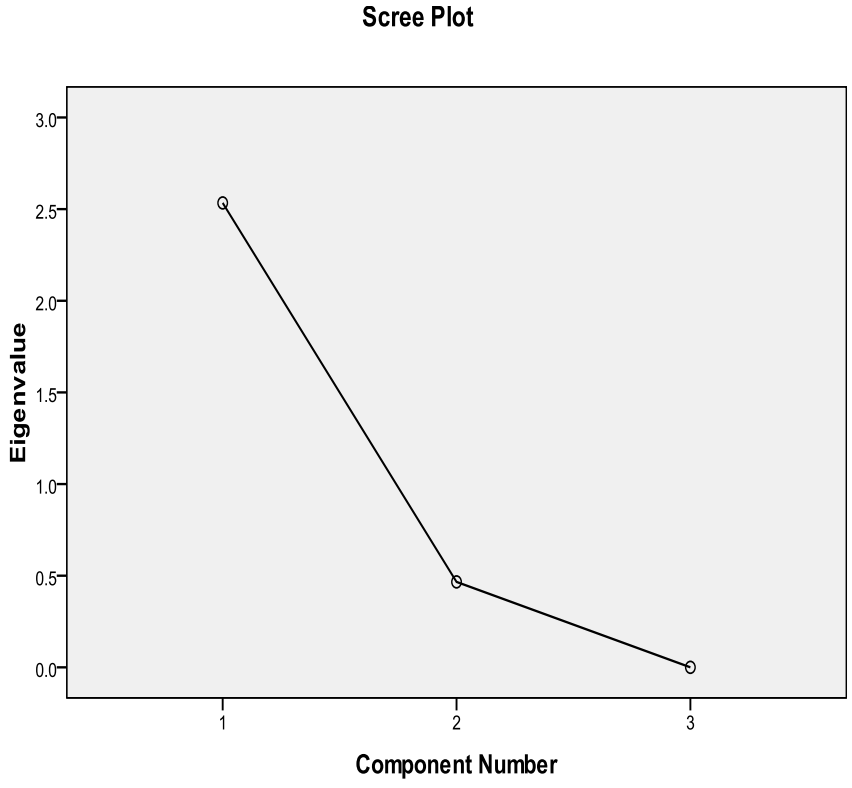

Fig: 3.3a Linear Search

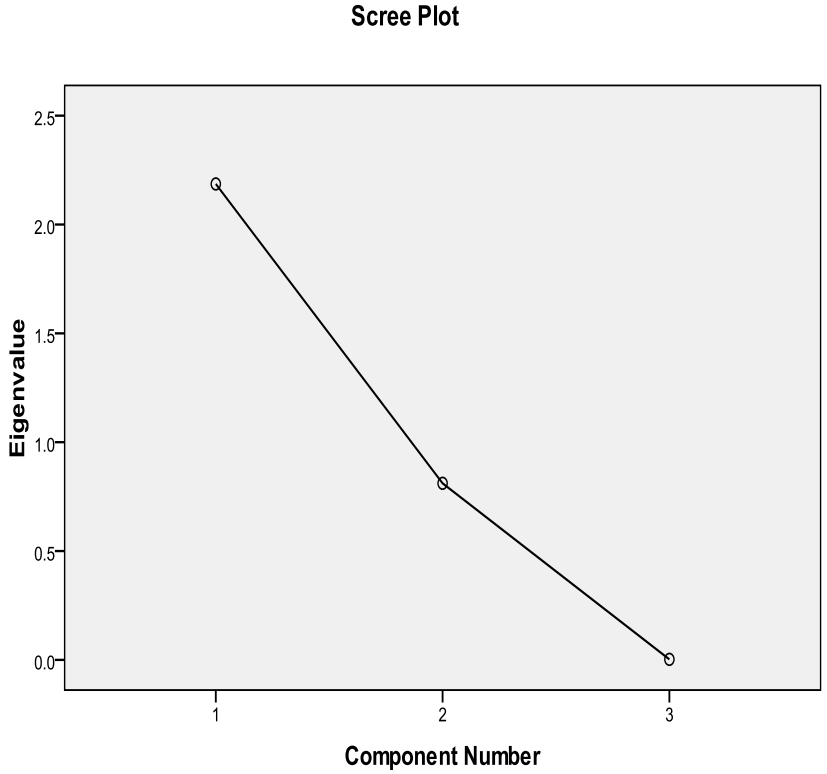

Fig:3.3b Binary Search

	1	2	3
Linear Search	2.534	0.466	9.28E-05
Binary Search	2.186	0.811	0.002

Fig: 4.11 Comparison of Searching Techniques

Factor - The initial number of factors is the same as the number of variables used in the factor analysis. However, not all 3 factors have effect on the efficiency of the search Techniques.

Initial Eigenvalues - Eigenvalues are the variances of the factors. Because we conducted our factor analysis on the correlation matrix, the variables are standardized so the total variance is equal to the number of variables used in the analysis. The eigenvalue of each factor is generated by the formula below where j^{th} factor is denoted by 'E_j' and calculated by:

$$E_j = \sum_{k=1}^{3} X^2{}_{i,j} \qquad i = 1,2,3; \quad j = 1 \qquad \text{eqn (4.6)}$$

$$P = 100\left(\frac{E_j}{n}\right) \qquad \text{eqn (4.7)}$$

Total - This column contains the eigenvalues. The first factor accounts for the most variance (and hence have the highest eigenvalue), and the next factor account for as much of the left over variance as it can, and so on. Hence, each successive factor will account for less and less variance.

% of Variance - This column contains the percent of total variance accounted for by each factor.

Cumulative % - This column contains the cumulative percentage of variance accounted for by the current and all preceding factors. In table 4.6a the first row shows a value of 84.459%, second row 15.538% and third .003%. This means that the first three factors together account for 100% of the total variance.

Extraction Sums of Squared Loadings - The number of rows in this panel of the table correspond to the number of factors retained. In this research, three factors are used, so there are three rows, one for each retained factor. The values in this panel of the table are calculated in the same way as the values in the left panel, except that here the values are based on the common variance.

Rotation Sums of Squared Loadings - The values in this panel of the tables below represent the distribution of the variance after the Promax rotation. Promax rotation tries to maximize the variance of each of the factors, so the total amount of variance accounted for is redistributed over the three extracted factors

Table: 4.6a Linear search

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.534	84.459	84.459	2.534	84.459	84.459	2.396
2	.466	15.538	99.997	.466	15.538	99.997	1.793
3	9.280E-5	.003	100.000	9.280E-5	.003	100.000	1.277

Table: 4.6b Binary Search

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.186	72.876	72.876	2.186	72.876	72.876	2.109
2	.811	27.041	99.918	.811	27.041	99.918	1.223
3	.002	.082	100.000	.002	.082	100.000	.028

5. Conclusion:

The efficiency with which search techniques is carried out has a significant impact on the overall efficiency of a program. The efficiency of linear and binary techniques in terms of numbers of comparison, time taken and memory used were used as decision variable to evaluate their efficiencies. Experimental results of the decision variables were generated from software tool in which data searched were varied for four different search techniques. The results were subjected to factor analysis using Statistical Packages for social scientist (SPSS) to test the level at which each of the factors affects the search techniques. The Kaiser-Meyer-Olkin and Bartlett test of sphericity indicate that the data used are adequate and can proceed to factor analysis. The KMO of the data used is more than 0.50 and the significance value is less than 0.05. The communality shows that reliable amount of variance was extracted and accounted for by returning value 1. Also, component correlation matrix indicates high level of correlation because it gives a diagonal matrix not identity matrix. Eigenvalue were used to indicate how well each of the extracted factors fit data from the experimental result. From the analysis results, the main factor affecting the search techniques was numbers of comparison. It contributed 84.459%, and 72.876% for Linear and Binary search techniques respectively. The Time taken came second contributing 15.538 and 27.041% for Linear and Binary respectively. Total memory consumed was the least of all contributing factors and has negligible percentages for the three searching techniques.

In summary, 'numbers of comparison' to search techniques is the main factor affecting the efficiency of searching techniques. It was also observed that in numbers of comparison binary search comes first, and linear search second respectively. While in Time taken, binary operation time is shorter followed linear search and finally, linear came first in memory usage followed by binary. From the above stated analysis and numbers of comparison is to be considered most critical, binary works better.

Contribution to Knowledge:

The research has significantly contributed to the body of knowledge in the following ways:

i. Using factor analysis by principal component, the contribution of each of the factors time taken, memory used, and numbers of comparison, to the efficiency of searching techniques is analyzed and established, hence the most critical factor in achieving efficiency of searching algorithm is determined.

ii. This thesis will assist researchers, developer or code writers to know which of the factors time, memory and numbers of comparison is critical to the searching techniques they use.

References

[1] Addison-Wesley, Searching an Ordered Table, ISBN 0-201-89685-0. Section 6.2.1:, pp.409–426. (1997).

[2] Aho A., Hopcroft J., and Ullman J.,(1974)- *The Design and Analysis of Computer Algorithms*, Addison Wesley.

[3] Akhigbe, Afolabi B., and Adagunodo E.(2011) - An Empirical Model for Information Retireval System Evaluation: The User's perspective. Computer Engineering and Intelligent System Vol 2 No. 4.
[4] Akinyokun O., Angaye O., and Ubaru O. (2009) -Factor Analysis of the Performance Indices and Communications Technology Projects in Public Sector of the Nigerian Economy. Journal of Technology Research, Vol. 1.
[5] Akinyokun O., Uzoka F.(2007) -Factor Analysis of the Effects of Academic Staff Profile on the Investment Portfolio of a University. International Journal of The Computer, the Internet and Management Vol. 15 1.
[6] Anupam Shukla and Rahul Kala 2009 International Journal of Computer Science and Network Security, VOL.8 No.6, June 2008 344
[7] Bruegge, B., Dutoit, A.H. (2004): Object-Oriented Software Engineering – Using UML, Patterns, and Java, 2nd International Edition, Prentice Hall
[8] Canaan C., Garai M,. and Daya M., (2011) - Popular Sorting Algorithms, Information institute Chireddzi, Zimbabwe World Applied Programming, ISSN: 2222-2510 Vol (1), No (1), pp 42-50
[9] Cappelleri C., Gerber A., Kourides A., and Gelfand A., (2000) - Development and Factor Analysis of Questionnaire to measure Patient Satisfaction with Injected and Inhaled Insulin for Type 1 Diabetes. Pfizer Inc. (J.C.C., R.A. Ger., R.A. Gel), Global Research And Development, Groton, Connecticut; and Pfizer Inc. (I.A.K.), New York.
[10] Chiemeke C., and Osazuwa W.,(2008) - Factor Analysis of Post-Implementation Review Of Student Information Systems in Nigeria. Research Journal for Applied Science Vol. 3 No 5.
[11] Donald E. (1998). - The Art of Computer Programming, Volume 3, Sorting and Searching. Addison-Wesley, Reading, Massachusetts.
[12] Encyclopedia." Microsoft Encarta 2009 [DVD]. Redmond, WA: Microsoft Corporation.
[13] Folorunsho O., Vicent O., Salako O.,(2010) - An Exploratory Study of Critical Factors Affecting the Efficiency of Sorting Techniques (Shell, Heap and Treap). Anale. SeriaInformatica. Vol. VIII fasc. 1.
[14] Jamie DeCoster (2011) Overview of Factor Analysis, Department of Psychology University of Alabama 348 Gordon Palmer Hall
[15] Johnson Richard A., Wichern, Dean W.(J992): Applied Multivariate Statistical Analysis. Jersey. Prentice Hall.
[16] Kruse, Robert L.(1999): "Data Structures and Program Design in C++", Prentice-Hall, ISBN 0 13-768995-0, page 280.
[17] Misra, S., Akman, I.(2008): A Complexity Metric based on Cognitive Informatics, Lecture Notes in Computer Science, Vol. 5009, pp.620-627.
[18] Roger S. P. (2005): Software Engineering – A practitioner's approach, 6th Edition. McGraw-Hill pp 34-37
[19] Thomas H., Charles E., Ronald L., and Clifford S., (2001), Introduction to Algorithms *Second Edition*, The MIT Press Cambridge, Massachusetts London, England McGraw-Hill.
[20] Vandana., Singh A., Monika., & Kaur S., (2011) – "Assortment of Different Sorting Algorithms". Asian Journal of Computer science and Technology, Swami Vivekananad Engineering College, Punjab, India
[21] Weisstein, Eric W. "Eigenvalue, (2011)" From Math World--A Wolfram Web Resource. http://mathworld.wolfram.com/Eigenvalue.html This page was last modified on 24 December 2011 at 10:32

Websites Accessed

[22] http://cryptodox.com/Binary_search_algorithm#Equal_elements This page was last modified on 24 December 2011 at 10:32am
[23] https://www.math.duke.edu//education/webfeatsII/Lite_Applets/Eigenvalue/start.html This page was last modified on 15 Jan 2011 at 10:32
[24] http://www.algolist.net/Algorithms/Binary_search Algorithms and Data Structures with implementations in Java and C++ This page was last modified on 24 December 2011 at 10:32am
[25] http://www.fon.hum.uva.nl/praat/manual/Principal_component_analysis.html © djmw, This page was last modified on November 10, 2010 at 09:09
[26] http://users.ics.tkk.fi/jhollmen/dippa/node30.html Jaakko Hollmen Fri Mar 8 13:44:32 EET 1996
[27] http://css.dzone.com/articles/algorithm-week-interpolation (Marcus and Minc 1988, p. 144) This page was last modified on 24 Jan. 2010 at 10:32
[28] www.wikipedia/interpolation..... This page was last modified on 20 November 2011 at 16:47

