Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 8 No. 1, March, 2017 - www.cisdijournal.net

# Automated Career Guidance Expert System Using Case-Based Reasoning Technique

**Ezenkwu, C.P**
Department of Electrical/Electronic and Computer Engineering
University of Uyo
Uyo, Akwa Ibom State, Nigeria
Email: chineduezenkwu@uniuyo.edu.ng
Mobile: +2347034471962

**Johnson, E.H.**
Department of Electrical and Electronic Engineering
Akwa-Ibom State University, Nigeria
Email: enyenihijohnson@aksu.edu.ng
Mobile: +2348127725225

**Jerome, O.B.**
Department of Electrical/Electronic & Computer Engineering
University of Uyo, Uyo, Akwa Ibom State, Nigeria
Email: jeromeblackq@gmail.com
Mobile: +2348138204517

## ABSTRACT

This paper is aimed at developing an Automated Career Guidance Expert System (AC-GES) using case-based reasoning (CBR) technique. Notably, a majority of tertiary education aspirants are faced with the challenge of selecting courses that will best match their academic capabilities. This has posed problems such as poor students' performances in their chosen areas of study, the inability of some students to complete their studies within the stipulated time and so on. In view of this, the need for an expert's advice that will guide these tertiary education aspirants in choosing suitable courses of study becomes pertinent. AC-GES is to assist high school students in choosing career paths that best suit their abilities based on their previous performances in some selected subjects, using Nigerian students as a case study. The case-base consists of 800 cases while two hundred (200) known cases were used for testing. The similarity metric and the neighbourhood parameter k used in the CBR algorithm were determined experimentally. With Euclidean distance, as the similarity metric, and 10 as the k-parameter, the CBR algorithm achieved a 0% classification error on the test-case. The system was finally prototyped as a MATLAB GUI application.

**Keywords***: Career; Guidance; Counseling; Expert System; Case Base Reasoning; similarity metric*

**CISDI Journal Reference Format**
Ezenkwu, C.P., Johnson, E.H & Jerome, O.B. (2017): Automated Career Guidance Expert System Using Case-Based Reasoning Technique Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 8 No 1. Pp 81-88.
Available online at www.cisdijournal.net

## 1. INTRODUCTION

According to Ivan Prelovský (as cited by Arumugam & Nagalingam, 2015 ), "career guidance is recognized as training and motivating individuals and students to plan their education, training and work; to take responsibility for their own educational and occupational opportunities, providing help to students for further education, flexible professional development, career management, as well as professional promotion". Furthermore, it plays a key role in helping labor markets, work and education systems meet their goals. The concept of career guidance has been widely used in the past and in several varieties. According to Herr and Cramer (1996), "It is largely a verbal process which involves the counselor and counselee(s) developing a dynamic and collaborative relationship, focused on identifying and acting on the counselee's goals, in which the counselor employs a repertoire of diverse techniques and processes, to help bring about self-understanding, understanding of behavioral options available, and informed decision making in the counselee, who has the responsibility for his or her own actions". Remarkably, guidance and counseling is an integral practice in schools and colleges, the intent of which is to provide pieces of advice to students in order to enable them improve in their academic performances, choose a befitting career and maximise their individual potentials.

However, the ubiquity of the Information and Communication Technology (ICT) has changed the way this activity is carried out in most developed countries. Motivated by the aforesaid, the aim of this paper is to automate the process of career guidance using an Artificial Intelligence approach, specifically known as Case-based reasoning (CBR). CBR is a general artificial intelligence paradigm for reasoning from experience. It is a natural way of solving problems that are similar to previously solved problems. This technique is employed by lawyers when they refer to previous case files in order to handle a current case which is similar to it. Moreover, other professionals such as medical doctors, engineers, etc. often use CBR in their daily activities. The idea behind this technique is to have a case-base of all cases experienced and their corresponding solutions such that if a new case is encountered, a similarity metric is used to compare its similarities with the cases in the case-base and the solution to the most related cases is applied to the new case. Some of the similarity metrics often used include; Euclidean Norm, Manhattan distance, Hamming Distance and so on. CBR is also known as Instance-based Reasoning or K-Nearest Neighbor algorithm (KNN).

The dataset used in this paper is an excel file in .xls format containing one thousand (1000) different cases with ten (10) features and one output. The features are selected important subjects offered in senior secondary schools. The subjects include mathematics, English language, physics, chemistry, biology, geography, Literature, accounting, Christian Religious Knowledge (CRK), and Government. For each case, the output consists of a human counselor's career recommendation based on the students performances in the ten (10) subjects. Using eight hundred (800) cases from the dataset for the case-base, a GUI- based MATLAB application is developed to recommend careers to new user leveraging the human counselor's experience captured in the case-base. The other two hundred (200) cases were used for testing.

## 2. LITERATURE REVIEW

### 2.1 Expert Systems for Career Guidance

Advice helps students to interpret information and apply it to their own situation; guidance is the in-depth support provided by qualified practitioners to help students explore their options and make informed choices that are best for them (CDI, 2014). In other words, counselling is a transformative process of helping people to learn all that are to be learnt both in and outside the school (Egbo 2013). According to Odu (2004), "the main aim of guidance and counselling is to assist the student to develop physically, mentally, emotionally, morally and educationally to cope with the learning situations within and outside the school environment". Akinade(2012) opined that there is limited number of trained counsellors in Nigerian schools and the ones already trained choose to go into non-school settings.

This makes the need for computer-based counsellors in schools extremely imperative. The role of ICT in guidance can be seen in three ways: as a tool, as an alternative, or as an agent of change (Oye *et al*, 2012). Ojenge and Lawrence(2008) recommended the use of expert systems for career guidance. According to Satvika *et al*(2010), "It is concluded that while expert systems in education have great potential, they remain un-established as a useful technology due to lack of research and documentation. Expert system for career selection can be developed using Fuzzy logic, neural network( Waghmode & Jamsandekar, 2015), decision tree and other Artificial Intelligence techniques for guiding students to select proper career stream.

Waghmode & Jamsandekar (2015) presented some expert systems used in educational sector for career guidance, few of which have been reviewed in this paper

**iAdvice:** This expert system uses features such as reasoning ability, providing explanations, alternative solutions, uncertainty and probability measures, questioning ability and also forward chaining, backward chaining and rule based inference in designing expert system. Past examination performance, student preferences and skills, industry alignment with subjects, are the main factors considered by a human expert in providing career guidance. The system was designed by Hendahewa *et al* (2006) .

**PAS**: Post graduate Advisor Expert System (PAS) was proposed and developed by Al-ghamdi (2012) for advising post graduate students of computer science major in King Abdulaziz University (KAU) to select suitable course during their PG programme.

**CAM** (**Career Advisor Model)** - is a model proposed to provide right advice to students. This model is based on machine learning technique and rule based decision support capability. It was presented by Mundra(2014)

**CMS (Career Master System):** Balogun *et al*( 2009) present development of career master DSS for counsellors to assist students in selecting the right discipline for secondary school leaving students of Nigeria who have problems with their choice of careers as they intend to study at tertiary institutions of their choice. This career master system was implemented using Visual Basic. This system is designed for desktop and counsellors, and system recommendation were based on parameters like ability, skills, Intelligent Quotient, interest, parents and friends influence, preferences, parent occupation and hobbies, past academic performance.

**CACGS**: Computer Assisted Career Guidance System (CACGS) is a career system which helps clients from secondary to post-secondary students of Malaysia. It is a interventions and various print and media-based support resources that are used within organizations to assist individuals in making career decisions. It was presented by (Adawiyah & Ismail ,2006)

**2.2 Case-Based Reasoning**

According to Bergamann *et al*(2009), "The core of every case-based problem solver is the case-base, which is a collection of previously made and stored experience items, called cases". A case-based problem solver solves new problems primarily by reuse of solutions from the cases in the case base. CBR is also referred to as k-nearest neighbor (KNN) algorithm, instance-base algorithm or memory-base learning. The idea of CBR is presented in the simple CBR life cycle shown in figure 1. The CBR life cycle comprises four steps – Retrieve, Reuse, Revise and Retain. The Retrieve step involves the selection of one or more cases from the case-base that have similarity with the case under consideration. To achieve this, a similarity metric is required to carry out this comparison. In the reuse step, the solutions to the retrieved similar cases are considered for the new case. In the reused case, the proposed solution or solutions can be revised or adapted to be well-suitable to the new case. Finally, the new case is made part of the case-base in the retain step.

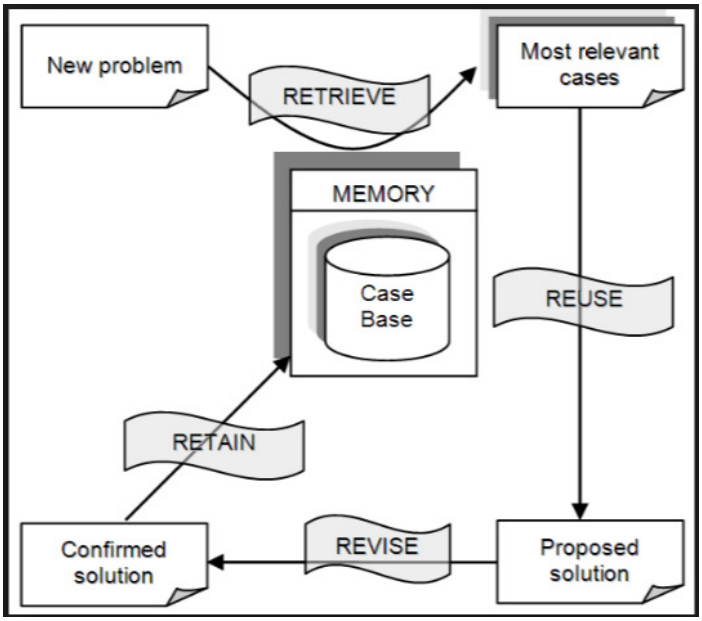


**Figure 1. CBR life cycle (Hugo *et al*, 2011)**

In recent time, CBR has been applied in solving devise problems in engineering, sciences, biology, sociology and so on.

Adebayo , Adekoya & Ekwonna (2014) proposed a Temperament and Mood Detection system which employs Case-Based Reasoning technique to help psychologists solve part of the problems in their complex domain.

Maha *et al*(2014) compared the performances of Decision Tree J48 and CBR in designing an automated counseling system for students in Pakistan. Out of the two algorithms tested, CBR gave the higher accuracy than the Decision Tree J-48.

**2.3 Similarity Measure**

Distance or similarity measures are essential to solving many pattern recognition problems such as classification, clustering, and retrieval problems (Sung-Hyuk, 2004). They are real-valued functions that quantify how similar two objects are.

A well-known similarity measure is the Minkowoski distance metric.

Let $i = < x_{i1}, x_{i2}, \ldots\ldots, x_{ip} >$ and $j = < x_{j1}, x_{j2}, \ldots\ldots, x_{jp} >$ be two p-dimensional data objects. Suppose that the dissimilarity of the two data objects is measured upon all the dimensions. The Minkowski distance between objects *i* and *j*.is defined as

$$d(i,j) = \sqrt[q]{|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \cdots + |x_{ip} - x_{jp}|^q} \quad (1)$$

Where q is a positive number. Three versions of the metric are in practical use according to the value of q. When q is set to 1, the version is known as the Manhattan or City-block distance. When q is set to 2, the version is known as Euclidean distance. When q is set to ∞, the version is known as Chebyshev or Supremum distance.

Manhattan distance: $d_1(i,j) = \sum_{t=1}^{p} |x_{it} - x_{jt}|$ (2)

Euclidean distance: $d_2(i,j) = \sqrt{\sum_{t=1}^{p} |x_{it} - x_{jt}|^2}$ (3)

Chebyshev distance: $d_\infty = \max_t(|x_{it} - x_{jt}|)$ (4)

Others include, Hamming distance, cosine distance and so on.

## 3. METHODOLOGY

### 3.1 Data Collection

An excel dataset in .xls format containing one thousand (1000) different cases with ten(10) features representing the performance of SS 1 students in a cross section of important selected subjects offered in senior secondary schools in Nigeria was collected from a secondary school. For each case the output contains a human counselor's career recommendation based on the students' performances in the ten subjects.

### 3.2 Data pre-processing

The data set was normalized using the min-max normalization technique represented by the formula;

$x_{normalised} = \frac{x - \min(X)}{\max(X) - \min(X)}$, where $x \in X$ (5)

This normalization produces values between 0 and 1. Normalization is a scaling or mapping technique used at data pre-processing stage, where a new range is found from an existing one. When it is necessary to maintain the large variation of prediction and forecasting, a normalization technique is required to make them closer. Some of the widely used normalisation techniques include Min-Max, Z-score & Decimal scaling. Min-Max Normalization is used here because the technique provides linear transformation on original range of data and also keeps the relationship among original data.

### 3.3 Data Splitting

The dataset is split in such a way that 80% of the dataset was used to create the case-base while 20% was used as the test-case for evaluating the retrieval ability of the system. The splitting of the dataset was done in random fashion to ensure that each subset contains similar distribution of the cases.

### 3.4 Implementation of the CBR algorithm

#### 3.4.1 CBR algorithm

The CBR algorithm is as follows:

```
algorithm CBR (Cb: Case base; k: integer; var n: New case)
begin
        for each case c in Cb do
                calculate distance(c, n) upon descriptive attributes
        end for;
        select into D k nearest neighbours of n according to their distances to n;
        n.class := majority class in D
end;
```

The proximity between the new case and each case in the case base are measured. The nearest k neighbours of the new case are selected. The recommendation to the new case is determined by the majority recommendation of the k nearest neighbours.

### 3.4.2 Comparison of different similarity metric at different k-parameters

The major issues in CBR algorithm are how to choose the optimum value of the neighborhood parameter k and the suitable similarity metric for the dataset. These two are crucial because the classification performance of the CBR algorithm depends on the neighbourhood parameter k and the similarity metric used. The error percentages on the 200 test cases for different similarity metrics at various values of k ranging from 10 to 100 were obtained as shown in Table 1.

The error is computed following the formula;

$$\text{Error (\%)} = \frac{\text{number of misclassification}}{\text{total number of Classification}} * 100 \quad (6)$$

**Table 1. Comparison of the performances different similarity metrics at different k-parameters**

| | | Values of k | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Similarity Metric** | **10** | **20** | **30** | **40** | **50** | **60** | **70** | **80** | **90** | **100** |
| **Error (%)** | Euclidean Distance | 0 | 1 | 3 | 2 | 1 | 1.5 | 2 | 1 | 1.5 | 0 |
| | Hamming Distance | 39 | 47 | 45 | 49 | 48.5 | 56 | 56.5 | 62.5 | 61.5 | 62.5 |
| | Cosine Distance | 23 | 19 | 22 | 21 | 26.5 | 24 | 23.5 | 26 | 27 | 29 |
| | Correlation Distance | 18.5 | 14 | 19.5 | 20 | 22 | 26.5 | 26 | 28.5 | 29.5 | 28.5 |
| | Chebyshev Distance | 11 | 10.5 | 11 | 12 | 12 | 12 | 11.5 | 11 | 11 | 11 |

The Table shows the comparison of these similarity metrics at various values of k.

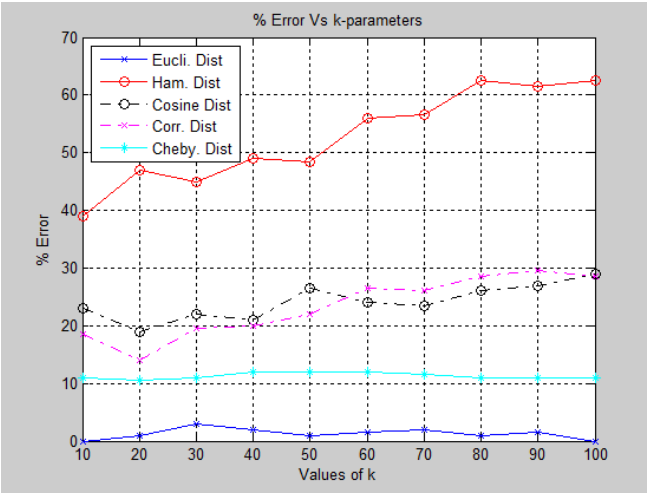


**Figure 2. Graph of % error against values of k for different similarity metrics**

From the graph, the Euclidean distance showed the least percentage error at all values of k. Moreover, % errors at k = 10 and k = 100 are 0%, implying that the CBR algorithm with Euclidean distance as similarity metric and a neighbourhood k parameters of 10 or 100 gives 100% accuracy in classifying new cases correctly. Informed by the outcome of this experiment, the Euclidean distance and k = 10 were selected for the system implementation.

.

### 3.5 Implementation of AC-GES

The system was prototyped as a MATLAB GUI application. Figure 4 presents AC-GES.

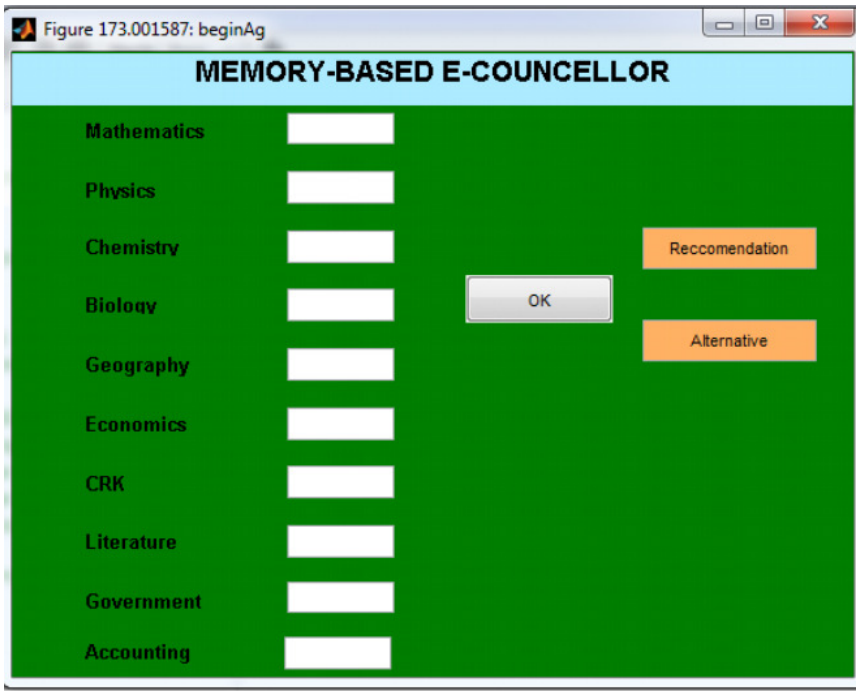


**Figure 4. AC-GES GUI**

A student's grades in the above subjects are keyed in. When the Ok button is clicked, the system recommends two career paths to the students, in the order of their preferences.

Figure 5 shows recommendations made to a student based on the grades s/he provided in the subjects. It provides first and second choice recommendation. The first choice recommendation is the majority recommendation of the k-nearest neighbours and the alternative recommendation is the second majority recommendation for the neighbours.

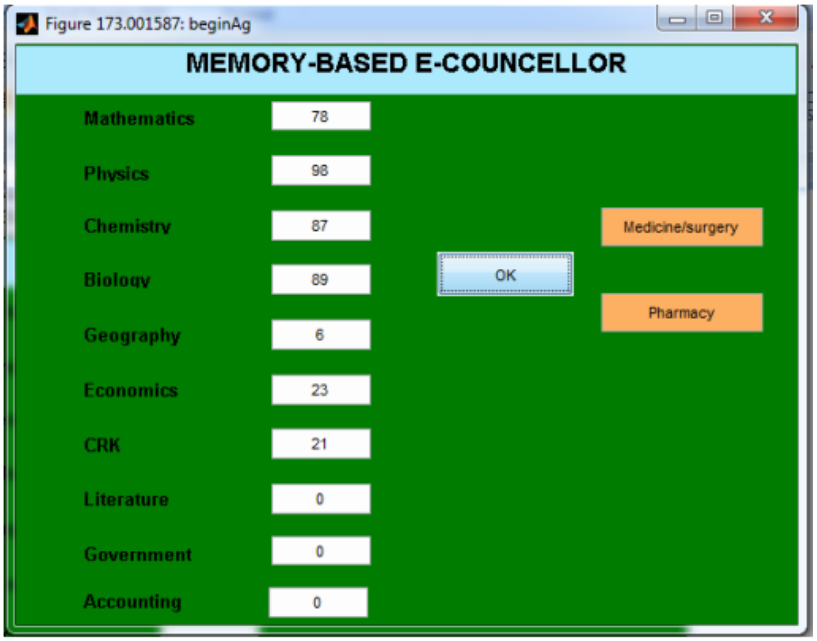


**Figure 5. AC-GES recommending career paths to a student based grades in subjects**

## 4. CONCLUSIONS

In this paper, an Automated Career Guidance Expert System (AC-GES) has been developed using case-based reasoning (CBR) technique**.** AC-GES is to assist high school students in choosing career paths that best suit their abilities based on their previous performances in some selected subjects, using Nigerian students as a case study. The case-base used in this paper consists of eight hundred (800) cases while two hundred (200) known cases were reserved for testing. Each of the cases, both in the case-base and the test-case, consists of 10 attributes, which represent a student's performances in ten (10) selected subjects and one(1) output, which is the human counselor's recommendation to the student. The similarity metric and the neighbourhood parameter k used in the CBR algorithm were determined experimentally. With Euclidean distance, as the similarity metric, and 10 as the k-parameter, the CBR algorithm achieved a 0% classification error on the test-case.

**REFERENCES**


1. Arumugam B, Nagalingam S. Career Guidance in Medical Education: Is It Mandatory? Int J Sci Stud 2015;3(7):1-2.
2. Herr, E. L.; Cramer, S. H. 1996. Career Guidance and Counseling through the Life Span. New York, Longman.
3. Career Development Institute(CDI) .2014. Careers Guidance in Schools and Colleges: A Guide to Best Practice. Retrieved from http://www.thecdi.net/write/final-best_practice_guide-.pdf on 04 Dec. 2016
4. Egbo A.C. 2013. The Role of Guidance and Counselling in Effectve Teaching and Learning in Schools: The Nigerian Perspective. The European Conference on Education. Retrieved from http://iafor.org/archives/offprints/ece2013-offprints/ECE2013_0392.pdf on 04 Dec. 2016
5. Odu, B.K. (2004) counselling intervention and crisis management as medium of improving psychological well being of the students journal of educational research and evaluation 4 (i) 140-148.
6. Akinade, E.A (2012), Modern Behaviour modification, principles and practices. Ibadan Bright way publishers
7. Oye N. D., Obi M. C., Mohd T. N., and Bernice, A.2012. Guidance and Counseling in Nigerian Secondary Schools: The Role of ICT. I.J. Modern Education and Computer Science, 8, 26-33
8. Satvika K., Akhil K. and Manoj B. 2010. Expert Systems Advances in Education. National Conference on Computational Instrumentation CSIO Chandigarh, India.
9. Borchert M, "Career Choice Factors of High School Students," pp. 1–82, 2002
10. Waghmode M. L and Jamsandekar P.P .2015. A Study of Expert System for Career Selection: Literature Review . International Journal of Advanced Research in Computer Science and Software Engineering. ISSN: 2277 128X. Available online at: www.ijarcsse.com
11. Hendahewa C, Dissanayake M, Samaraweera S, Ruwanpathirana A, and Karunananda A.S. 2006. Artificial Intelligence Approach to Effective Career Guidance, no. September, pp. 32–42, 2006.
12. A. Al-ghamdi et. 2012. An Expert System for Advising Postgraduate Students," vol. 3, no. 3, pp. 4529–4532.
13. A. Mundra, A. Soni, S. K. Sharma, P. Kumar, and D. S. Chauhan . 2014. Decision Support System for Determining : Right Education Career Choice, Elsevier, vol. 4, no. Table 1, pp. 8–17.
14. V. F. Balogun, A. F. Thompson, and O. State. 2009. Career Master : A Decision Support System ( DSS ) for Guidance and Counseling. vol. 10, no. 2, pp. 337–354.
15. Adawiyah A and Ismail B. 2006. Career Guidance System for Secondary School Students. Universiti Teknologi MARA.
16. Ralph Bergmann, Klaus-Dieter Althoff, Mirjam Minor, Meike Reichle, and Kerstin Bach .2009. Case-Based Reasoning - Introduction and Recent Developments. *In: Künstliche Intelligenz, 23(1),* pp. 5-11.
17. Hugo López-Fernández, Florentino Fdez-Riverola, Miguel Reboiro-Jato, Daniel Glez-Peña and José R. Méndez .2011. Using CBR as Design Methodology for Developing Adaptable Decision Support Systems, Efficient Decision Support Systems - Practice and Challenges From Current to Future, Prof. Chiang Jao (Ed.), ISBN: 978-953-307-326-2, InTech, Available from: http://www.intechopen.com/books/efficient-decisionsupport-systems-practice-and-challenges-from-current-to-future/using-cbr-as-design-methodology-fordeveloping-adaptable-decision-support-systems
18. Adebayo K. J, Adekoya A. M. and Ekwonna C. 2014. Temperament and Mood Detection Using Case-Based Reasoning and the work. I.J. Intelligent Systems and Applications, 03, 50-61. DOI: 10.5815/ijisa.2014.03.05
19. Akhil jabbar M.,Deekshatulu B.L and Chandra P. 2013. Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm. International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA)
20. Maha N, Anum A, Unsa T, Jannat F.S, Rabia A, and Maria T . 2014 . Automated Career Counseling System for Students using CBR and J48. J. Appl. Environ. Biol. Sci., 4(7S)113-120, 2014. ISSN: 2090-4274
21. Sung-Hyuk C . 2004. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. International Journal of Mathematical Models and Methods in Applied Sciences. Issue 4, Volume 1.
22. Hongo D .2010. Data Mining Techniques and Applications: An Introduction. Cengage Learning EMEA.