IJCSIS Vol. 13 No. 4, April 2015
ISSN 1947-5500

International Journal of Computer Science & Information Security

Pennsylvania, USA

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2015 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site
Notification: see web site
Revision: see web site
Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge
Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site https://sites.google.com/site/ijcsis/

For more information, please visit the journal website (https://sites.google.com/site/ijcsis/)

Editorial
Message from Managing Editor

Over the past several decades, we have witnessed significant research and innovation in several domains including network security, cloud computing and virtualization. The purpose of this edition is to gather novel experimental and theoretical evidence from both industry and academia in the broad areas of Computer Science, ICT & Security and further bring together people who work in the relevant areas. The ***International Journal of Computer Science and Information Security*** *(IJCSIS) promotes research publications which offer significant contribution to the computer science knowledge and which are of high interest to a wide academic/research/practitioner audience. Coverage extends to several main-stream and state of the art branches of computer science and security. As a scholarly open access peer-reviewed journal, IJCSIS mission is to provide an outlet for quality research & academic publications. It aims to promote universal access for international scientific community to scientific knowledge; and the creation and dissemination of scientific and technical information.*

IJCSIS archives all publications in major academic/scientific databases. Indexed by the following International agencies and institutions: Google Scholar, CiteSeerX, Cornell's University Library EI, Scopus, DBLP, DOAJ, ProQuest and EBSCO. Moreover, Google Scholar reported increased in number cited papers published in IJCSIS ***(No. of Cited Papers: 555, No. of Citations: 1305)****. Abstracting/indexing/reviewing process, editorial board and other important information are available online on homepage. We help researchers to succeed by providing high visibility, prestige and efficient publication process.*

IJCSIS editorial board, consisting of international experts, guarantees a rigorous peer-reviewing process. We look forward to your collaboration. For further questions please do not hesitate to contact us at ***ijcsiseditor@gmail.com****.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 13, No. 4, April 2015 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Bibliographic Information
ISSN: 1947-5500
Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:
IJCSIS Managing Editor
(ijcsiseditor@gmail.com)
Pennsylvania, USA
Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

Fuzzy Logic is described in detail. Finally, an analysis of the system's performance as applied to a synthetic static scenario is presented. The simulation results are acceptable and fine and show that this model is reliability.

Keywords-component; Threat Assessment; Fuzzy Knowledge Based System; Decision Support System; Weapons Assignment; Threat Evaluation Fuzzy Model

4. Paper 31031518: A New Data Fusion Instrument for Threat Evaluation Using of Fuzzy Sets Theory (pp. 19-32)

Ehsan Azimirad, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran
Javad Haddadnia, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran

Abstract — This paper represents an intelligent description of the threat evaluation process in 3 level of JDL model using of fuzzy sets theory. The degree of threat is evaluated for all possible targets to improve the situational awareness and decision making in command and control system and is calculated to precise weapon assignment. Taking into consideration the amount of uncertainty involved in the process of threat evaluation for dynamic targets, the fuzzy set theory turns out to be a good candidate to model this problem. In this approach, based on a fuzzy logic, is making it possible to handle imperfect observations. The structure of the fuzzy expert system based on fuzzy number approach is described in detail. Finally, an analysis of the system's performance as applied to multiple dynamic scenarios is presented. The simulation results show the correctness, accuracy, reliability and minimum errors in the system is designed.

Keywords-component; Fuzzy Number, JDL Model, Decision Support System, Dynamic Air Targets, Multi Sensor Data Fusion; Weapons Assignment

5. Paper 31031526: Optimizing TCP Vegas for Optical Networks: a Fuzzy Logic Approach (pp. 33-45)

Reza Poorzare, Department of Computer Science Young Researchers Club, Ardabil Branch, Islamic Azad University, Ardabil, Iran
Shahram Jamali, Department of Computer Engineering University of Mohaghegh Ardabili, Ardebil, Iran

Abstract — Performance of TCP is reduced over buffer-less optical burst switched (OBS) networks by misunderstanding of the congestion status in the network. In other words, when a burst drop occurs in the network and we cannot distinguish congestion and burst contention in the network, TCP wrongly decreases the congestion window size (cwnd) and causes significant reduction of the network performance. This paper employs the fuzzy logic to solve this problem. By using the fuzzy logic we provide a framework to distinguish whether the burst drop is due to the congestion or is due to the burst contention. The full approach, for detecting state of network, relies on Round-Trip-Time (RTT) measurement only. So, this is an end-to-end scheme which only end nodes are needed to cooperate. Extensive simulative studies show that the proposed algorithm outperforms other TCP flavors such as TCP Vegas, TCP Sack and TCP Reno, in terms of throughput, packet delivery count and fairness.

Keywords — Fuzzy Logic, Optical Burst Switching, TCP Vegas, Transport Control Protocol (TCP).

6. Paper 31031532: Improved Algorithm for fusion of Satellite Images Using Combined DWT-FDCT Transforms (pp. 46-50)

Manjushree B S, CSE, DBIT/VTU, Bangalore, India
Shruthi G, CSE, DBIT/VTU, Bangalore, India

Abstract - Image fusion based on the Fourier and wavelet transform methods retain rich multispectral details but less spatial details from source images. Wavelets perform well only at linear features but not at non linear discontinuities because they do not use the geometric properties of structures. Curvelet transforms overcome such difficulties in feature representation. A novel fusion rule via high pass modulation using Local Magnitude Ratio (LMR) in Fast Discrete Curvelet Transforms (FDCT) domain and Discrete wavelet transforms (DWT) is defined. For experimental

study of this method Indian Remote Sensing (IRS) Geo satellite images are used for Pan and MS images. This fusion rule generates HR multispectral image at high spatial resolution. This method is quantitatively compared with Wavelet, Principal component analysis (PCA) fusion methods. Proposed method spatially outperforms the other methods and retains rich multispectral details.

Keywords: Image Fusion, Fast Discrete Curvelet Transforms, Discrete wavelet transforms, Local Magnitude Ratio (LMR)

7. Paper 31031534: Biometric Student Record Management System (pp. 51-62)

Onuiri Ernest E., Oludele Awodele, Oshilagun Ibukun, Yadi Chukwuemeka and Etuk Otobong
Department of Computer Science, Babcock University, Ilishan-Remo, Ogun State; Nigeria

Abstract – Information is an important part of any system. In the academic world, information is especially very important and essential. Students have to register for courses, take attendance, quizzes, and exams and as well as check their scores. Years after graduating from the school, students come back asking for transcript. It is therefore very important to handle students' records in a way that is accessible, maintainable and secure. The manual method of cumulating and storing student record is often prone to various degrees of human error and is also unsecured making it exposed to unauthorized personnel. This paper presents the design and development of a biometric student record management that provides an interface between student and the institution to enable prompt checking of grades, as well as track their progress and efficiently record each student's attendance for every lecture attended through the use of a biometric device. The methodology used in developing this system is the waterfall methodology and this was used because it is a one dimensional model, meaning it is very easy to implement and also the documentation is done at the beginning of the software development. During the course of this research, it was realized that developing a biometric student record management system was a herculean task. This system was given to random students to use and 90% of them loved the interactive nature of the system. A projection of record growth in relation to student population and system requirement was carried out in the study.

Keywords – Fingerprint, Biometrics, Biometric Student Record Management System (BSRMS), Student Information Management System (SIMS), Grade Point Average (GPA), Students

8. Paper 31031538: Classification Framework Based on C4.5 Algorithm for Medicinal Data (pp. 63-67)

Karthik Ganesan, Department of Computer Science and Engineering, College of Engineering, Anna University, Chennai, India

Abstract - This study proposes a framework with preprocessing techniques namely Missing value replacement, Discretization, Principal Component Analysis (PCA) to extract the key features and then applying c4.5 classifier algorithm to enhance the classification of medicinal data. The input data gets subjected to missing data imputation through any one of the standard methods like mean, mode, constant and manual input. The dataset is then subjected to Discretization to formalize a reasonable set of discrete bins. PCA is then applied on the dataset to identify the principal components of the dataset, which attribute to the mean data inference. C4.5 algorithm has been used to construct a decision tree based on the information gain of the training set. This work used Cleveland heart disease dataset, obtained from UCI machine learning repository. The dataset is composed of details of about 303 patients and helps to predict presence or absence of cardio vascular disorder based on 75 attributes. The proposed framework was applied on this dataset and exhibited an accuracy of about 77.73%.

Keywords — PCA, Discretization, C4.5, classification

9. Paper 31031539: Energy Efficiency of IEEE 802.15.6 MAC Access Modes for Remote Patient Monitoring Applications (pp. 68-77)

Anas Bouayad, Nour El Houda Chaoui, Mohammed El Ghazi, Molhime El Bekkali

1234 TTI Laboratory, USMBA, FEZ, MOROCCO

Abstract - The progress that has been made over the last decade in the medical field was focused on integrating communication and information technology especially Wireless Body Area Networks (WBANs) in healthcare systems for remote patient monitoring (RPM) applications. WBANs have shown great potential in improving healthcare quality, allowing continuous patient to be remotely monitored and diagnosed by doctors. WBAN operates in close vicinity to, on, or inside a human body and supports a variety of medical applications. Energy consumption is a key WBANs since energy-constrained sensors monitor the vital signs of human beings in healthcare applications. In this work, we are interested in evaluating access methods and access mechanism used in MAC layer of the IEEE 802.15.6 standard and the proposition of suitable access methods and parameters should be used to decrease the energy consumption. Performance evaluation will be based on the simulation of a short range wireless Body Area Network based solution implementing the IEEE 802.15.6. Simulation will be performed on OMNet++ with the Castalia simulator.

Keywords: RPM, Wireless Body Area Networks, IEEE 802.15.6, (MAC) protocols, access methods, polling, CSMA/CA, Energy consumption.

10. Paper 31031540: Designing a jitter buffer for QoS improvement in VoIP networks (pp. 78-83)

Negar Chehregosha, Dept. of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
Mohammad Ali Pourmina, Dept. of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract -- Today main challenge in IP networks engineering is simultaneous support of different applications such as sending voice, video and data, with appropriate quality of service. The generated traffic by IP telephone, voice and video conference and on line applications, are real time and time sensitive. Jitter is an usual problem in quality of service of VoIP network. The purpose of this paper is to reduce jitter to improve quality of service. Achieve Real time voice quality is required jitter smoothing in receiver that usually is done by jitter buffer mechanism. Here we introduce an algorithm to design jitter buffer. We simulate one VoIP network by OPNeT simulator and Matlab software is used to implement the algorithm; then we compare simulation results before and after applying the algorithm and the effects of changes in buffer size on delay and jitter are checked. Output voice quality will be measured based on PESQ, according to ITU-T P.862 recommendation. The results show packet buffering reduces packets delay and makes values of them become closer together.

Keywords: VoIP, Jitter, Jitter buffer, Delay, Quality of Service

11. Paper 31031541: Trend Analysis in Academic Journals in Computer Science Using Text Mining (pp. 84-88)

Adebola K. Ojo and Adesesan B. Adeyemo
Department of Computer Science, University of Ibadan, Ibadan, Nigeria

Abstract — Text mining is the process of discovering new, hidden information from texts- structured, semi-structured and unstructured. There are so many benefits, valuable insights, discoveries and useful information that can be derived from unstructured or semi- unstructured data. In this study, text mining techniques were used to identify trends of different topics that exist in the text and how they change over time. Keywords were crawled from the abstracts in Journal of Computer Science and Technology (JCST), one of the ISI indexed journals in the field of Computer Science from 1993 to 2013. Results of our analysis clearly showed a varying trend in the representation of various subfields in a Computer Science journal from decade to decade. It was discovered that the research direction was changing from pure mathematical foundations, Theory of Computation to Applied Computing, Artificial Intelligence in form of Robotics and Embedded Systems.

Keywords-component; Computer Science, Text Mining, mathematical foundations, applied computing, Robotics, Embedded Systems

12. Paper 31031543: d-HMAC — An Improved HMAC Algorithm (pp. 89-96)

Mohannad Najjar, University of Tabuk, Tabuk, Saudi Arabia

Abstract—The keyed-hash message authentication code (HMAC) algorithm is a security tool primarily used to ensure authentication and data integrity in information systems and computer networks. HMAC is a very simple algorithm, and relies on hash functions that use a secret key. HMAC's cryptographic strength is based on the use of effective cryptographic characteristics such as balancing and the avalanche effect. In this study, we develop a new algorithm, entitled dynamic HMAC (d-HMAC), to improve and enhance the cryptographic characteristics of HMAC. The improved algorithm provides stronger resistance against birthday attacks and brute force attacks. To achieve this objective, HMAC constant values *ipad* and *opad* are dynamically calculated in d-HMAC. Values for *ipad* and *opad* will be obtained from the HMAC input message, the public key of the receiver, and a substitution-box (*S*-box) table with enhanced security characteristics specifically created for this purpose. We demonstrate that the improved d-HMAC algorithm is more resistant to known cryptographic attacks, and prove that it exhibits similar or better cryptographic characteristics than HMAC.

Keywords-cryptography; data integrity; authentication; MAC; HMAC; hash functions; SHA-256

13. Paper 28021519: Hadoop Architecture and its Functionality (pp. 97-103)

Dr. B V Ramana Murthy,Jyothishmathi College of Engg and Technology, Shamirpet, India
Mr. V Padmakar, Guru Nanak Institutions Technical Campus, Hyderabad
Mr. M Abhishek Reddy, Jyothishmathi College of Engg and Technology, Shamirpet, India
Abstract - Hadoop is nothing but a "framework of tools" and it is a java based programming framework (In simple terms it is not software). The main target of hadoop is to process the large data sets into smaller distributed computing. It is part of the Apache project sponsored by the Apache Software Foundation. As we observe in database management system, all the data are stored in organized form by following the rules like normalization , generalizations etc., and hadoop do not bother about the DBMS features as it stores large amount of data in servers. We are studying about Hadoop architecture and how big data is stored in servers by using this tools and the functionalities of Map Reduce and HDFS (Hadoop File System).

Keywords: Big Data, HDFS, Map Reduce Task Tracker, Job Tracker, Data Node, and Name Node.

14. Paper 31031509: Data mining methodologies to study student's academic performance using the C4.5 algorithm (pp. 104-113)

Hythem Hashim (1) , Ahmed A. Talab (2) , Ali Satty (3) , and Samani A. Talab (1)
(1) Faculty of Computer Science and Technology, Alneelain University, Khartoum, Sudan.
(2) White Nile College for Science and Technology, White Nile state ,Kosti.
(3) School of Statistics and Actuarial Sciences, Alneelain University, Khartoum, Sudan.

Abstract - The study placed a particular emphasis on the so called data mining algorithms, but focuses the bulk of attention on the C4.5 algorithm. Each educational institution, in general, aims to present a high quality of education. This depends upon predicting the unmotivated students before they entering in to final examination. Data mining techniques give many tasks that could be used to investigate the students performance. The main objective of this paper is to built a classication model that can be used to improve the students academic records in Faculty of Mathematical Science and Statistics. This model has been done using the C4.5 algorithm as it is a well-known, commonly used data mining technique. The importance of this study is that predicting student performance is useful in many different settings. Data from the previous students academic records in the faculty have been used to illustrate the considered algorithm in order to build our classification model.

Keywords: Data mining, The C4.5 algorithm, Prediction, Classification algorithms.

15. Paper 31031533: Resource Modeling for the Development of a Decision-making System - Applied HCEFLCD of Morocco (pp. 114-118)

K. Oubedda, M. Khalfaoui, A. Ettahir
Systems Analysis Laboratory of Information Processing and Integrated Management (LASTIMI) School of Sale Technology, University Mohammed V Agdal

Abstract - The aim of this work and to develop a decision support system for the operation of a model including the main stakeholders of the High Commissioner for Water, Forests and Desertification Control (maker / managers, administrative, customers). This system is based on the relationship between the actors and their activities and their needs vary by contribution in time. It aims to make available to managers a set of dashboards that can improve the quality of provided services. We begin by modeling the actors up and clean process for studying both their organizations and their activities and needs. The first applications of this work has focused on data for the Directorate of Planning, Information System and Cooperation, and the Directorate of Forest Estate, Legal Affairs and Litigation. The results are encouraging.

Keywords: Information systems, decision support systems, dashboards, databases, modeling.

Trend Analysis in Academic Journals in Computer Science Using Text Mining

* Adebola K. Ojo and Adesesan B. Adeyemo
[1,2]Department of Computer Science
University of Ibadan
Ibadan, Nigeria
*Corresponding author

***Abstract*— Text mining is the process of discovering new, hidden information from texts- structured, semi-structured and unstructured. There are so many benefits, valuable insights, discoveries and useful information that can be derived from unstructured or semi- unstructured data. In this study, text mining techniques were used to identify trends of different topics that exist in the text and how they change over time. Keywords were crawled from the abstracts in Journal of Computer Science and Technology (JCST), one of the ISI indexed journals in the field of Computer Science from 1993 to 2013. Results of our analysis clearly showed a varying trend in the representation of various subfields in a Computer Science journal from decade to decade. It was discovered that the research direction was changing from pure mathematical foundations, Theory of Computation to Applied Computing, Artificial Intelligence in form of Robotics and Embedded Systems.**

Keywords-component; Computer Science, Text Mining, mathematical foundations, applied computing, Robotics, Embedded Systems

I. INTRODUCTION

Text mining is the discovery by computer of new, previously unknown information, by automatically extracting information from a usually large amount of different unstructured textual resources [1]. *Previously unknown* implies discovering genuinely new information. *Unstructured* means free naturally occurring texts as opposed to HyperText Markup Language (HTML), eXtensible Markup Language (XML), and other scripting languages. In text mining, the goal is to discover unknown information, something that no one yet knows and so could not have yet written down [2]. [3] referred to it as a collection of methods used to find patterns and create intelligence from unstructured data.

Text mining techniques are used to draw out the occurrences and instances of key terms in large blocks of text, such as articles, web pages, complaint forums, or Internet chat rooms and identify relationships among the attributes [4]. Often used as a preparatory step for data mining, text mining often translates unstructured text into a useable database-like format suitable for data mining for further and deeper analysis [5]. [3] also described text mining as an emerging technology that can be used `to augment existing data in corporate databases by making unstructured text data available for analysis.

Generally research publications have been on the increase globally. As a result, different areas are being covered such as science, engineering, agriculture, medicine and education. However, it is a time consuming task to determine manually the areas being focused on by authors who are publishing in these journals. In this study, a framework for discovering the research trends in Computer Science in the last three decades was developed using text mining techniques.

This work used an ISI indexed journal called Journal of Computer Science and Technology (JCST). The trending of topics published in papers in JCST across two decades was explored.

There are three types of textual data for text mining. These include Title of the paper, Abstract of the paper and complete body of the paper [6]. The data used in this study were the abstracts with the keywords. Analysing the abstract of a paper is appropriate since it contains the detailed objective of a paper and did not contain extraneous items such as tables and images [6]. Three data sets were created with the number of observations (that is, paper abstracts) as shown in Figure 1.

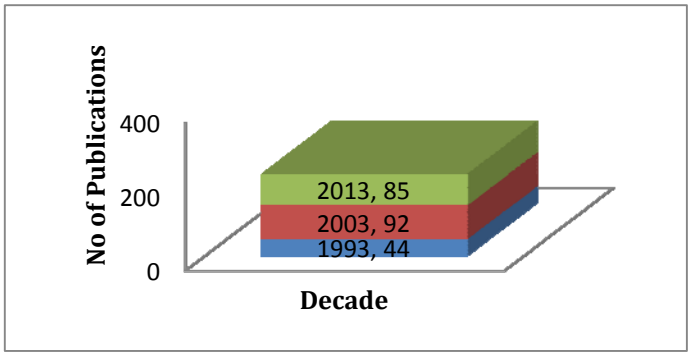

Figure 1: Number of Paper Abstracts in each Period

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. Materials and Method

The framework for Research Trend discovery is presented in Figure 2.

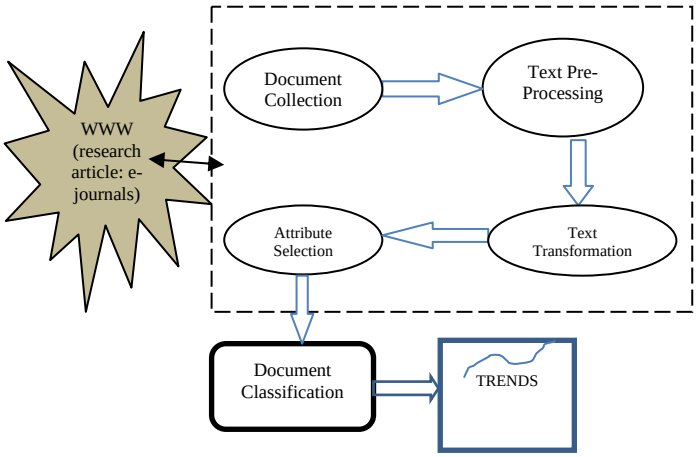

Figure 2: Framework for Academic Journal Articles (Trend Discovery)

Document Collection: The first phase in Figure 2 is the document collection. In this phase, the abstracts of the academic journal (JSCT) were *extracted* as documents (doc, pdf and html) crawled from the internet. These text documents were stored in different formats (pdf, doc, txt, html and xls). This depends on the nature and type of the text data. These contained the list of the abstracts of the volumes published in the years 2013, 2003 and 1993.

Text Pre-processing: This is also known as tokenization or text normalization. It involves the process of text clean-up (advertisements from the web pages are removed as well as the tables, figure and formulas) and tokenization (splitting up a string of characters into a set of tokens). During term extraction, character text was first parsed into words. This process also stripped away words that conveyed no meaning. Adjectives, adverbs, nouns and multi-word were extracted from the document. Noisy data, such as, tags, punctuation marks, white spaces, special characters and digits were extracted as well. Also, certain words occurred very frequently in text data. Examples included "the" and "a". These words were removed from the term collection because they had no meaningful content.

Text Transformation/Attribute Selection: By creating a list of stop words and eliminating them, the number of indicator variables created was reduced. After removal of stop words, stemming was performed. Word frequency and inverse document frequency were two parameters used in filtering terms. Low term frequency (TF) and document frequency (DF) terms were removed from the indexing of those documents. In "Bags of words" representation each word is represented as a separate variable having numeric weight.

III. Results and Discussion

The abstracts were extracted from http://link.springer.com/journal/volumesAndIssues/11390. One well known subject classification system for Computer Science is the ACM Computing Classification System devised by the Association for Computing Machinery [7] [8]. Computer Science was divided into ten (10) subfields. The subfields included Algorithm and Data Structures, Artificial Intelligence, Communication and Security, Computer Architecture, Computer Graphics, Databases, Programming Languages and Compilers, Scientific Computing, Software Engineering, and Theory of Computation. Table 1 presents the article classifications for 1993, 2003 and 2013 while Table 2 shows percentage distribution of article classifications.

In the Text Extraction Process, all the abstracts were parsed into independent words. This process also stripped away words that conveyed no meaning. Adjectives, adverbs, nouns and multi-word are extracted from the document. Noisy data, such as, tags, punctuation marks, white spaces, special characters and digits were extracted as well. Table 1 shows the clusters generated with the term frequencies and weights. Our suggested cluster labels were based on the descriptive terms and corresponding fields.

TABLE 1: ARTICLE CLASSIFICATIONS FROM 1993 TO 2013

No	Descriptive Terms	1993	2003	2013
1	Algorithms, Data Structures	33	27	24
2	Artificial Intelligence, Automated Reasoning, Computer Vision, Natural Language Processing, Machine Learning, Robotics	13	37	112
3	Networking, Computer Security, Cryptography, Concurrent, Parallel & Distributed Systems	51	128	221
4	Computer Architecture, Operating Systems	17	70	157
5	Computer Graphics, Image Processing	19	49	67
6	Relational Databases, Data Mining	12	41	41
7	Compiler Theory, Programming Language Pragmatics, Programming Language Theory, Formal Semantics	49	6	20
8	Computational Science, Numerical Analysis, Symbolic Computation, Computational Chemistry, Bioinformatics & Computational Biology, Computational Neuroscience	1	1	10
9	Software Engineering, Formal Methods, Algorithm Design, Computer Programming, Human-Computer Interaction, Reverse Engineering	0	5	34
10	Theory of Computation, Automata Theory, Computability Theory	33	50	91

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE 2: PERCENTAGE DISTRIBUTION OF ARTICLE CLASSIFICATIONS FROM 1993 – 2013

No	Descriptive Terms	1993	2003	2013
1	Algorithm and Data Structures	14.5	6.5	3.1
2	Artificial Intelligence	5.7	8.9	14.4
3	Communication and Security	22.4	30.9	28.4
4	Computer Architecture	7.5	16.9	20.2
5	Computer Graphics	8.3	11.8	8.6
6	Databases	5.3	9.9	5.3
7	Programming Languages and Compilers	21.5	1.4	2.6
8	Scientific Computing	0.4	0.2	1.3
9	Software Engineering	0.0	1.2	4.4
10	Theory of Computation	14.5	12.1	11.7
	TOTAL (%)	**100**	**100**	**100**

A. Trends

Trend analysis is used for identifying trends in documents collected over a period of time [2]. Identification of meaningful patterns and trends and the extraction of potential knowledge in large volumes of text data is an important task in various fields [9][10]. The appearances of specific terms across the two decades are used to understand the trends and research patterns of sub fields in Computer Science. A frequency value of 'n' for a term means that particular term was mentioned in the abstracts of 'n' distinct journals. Figure 3 shows the percentage of papers contributed for the ten disciplines across the period.

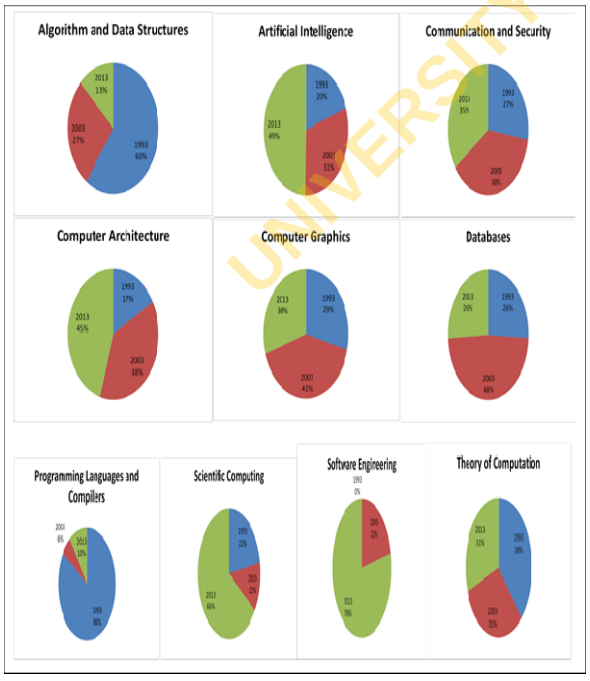

Figure 3: Percentage of Papers Contributed For Ten Computer Science Disciplines From 1993 To 2013

In Figures 6, most of the papers in 1993s (22.4%) were presented on Communications and Security followed by Programming Languages and Compilers (21.5%), Theory of Computation and Algorithms and Data Structures (14.5%), Computer Graphics (8.3%), Computer Architecture (7.5%), Artificial Intelligence (5.7%), Databases (5.3%), Scientific Computing (0.4) and Software Engineering (0.0%). The same trend was not observed in the following years. Percentage of papers published in Artificial Intelligence, Communication and Security, Computer Architecture, Computer Graphics, Databases and Software Engineering gradually increased from 1993 through 2003 (as shown in Figure 5) while publications in Algorithm and Data Structures, Programming Languages and Compilers, Scientific Computing and Theory of Computation reduced drastically (as shown in Figure 6). However In 2013, there was a great increase in papers published in Artificial Intelligence, Computer Architecture and Software Engineering while in there was relatively fewer numbers of papers published in the other disciplines. This shows that the research direction is changing from pure mathematical foundations, Theory of Computation to applied computing, Artificial Intelligence in form of Robotics and embedded systems.

Figure 4 shows the trend plot of the sub fields in Computer Science across the two decades.

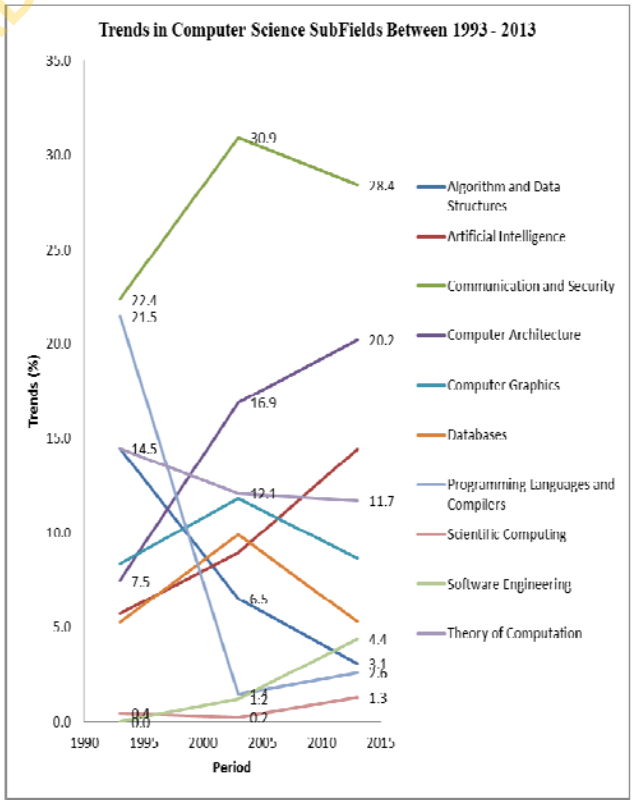

Figure 4: Trend Plot of the Sub Fields in Computer Science across the two decades

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

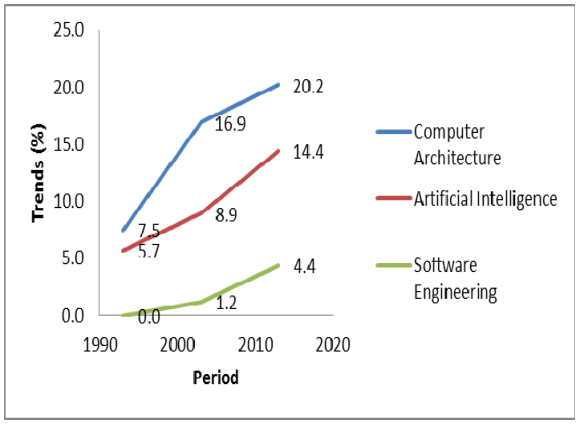

Figure 5: Gradual increase of some sub fields across the two decades

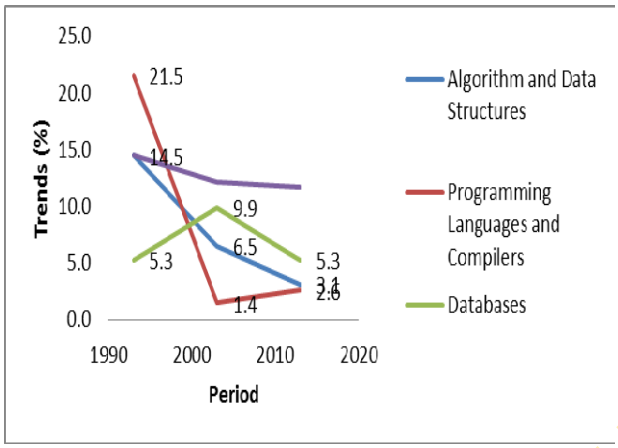

Figure 6: Gradual decrease of some sub fields across the two decades

Figure 6 shows the trend plot of various sub fields in Computer Science with much less representation in the papers compared to the large scale representation of the discipline as shown in Figure 5.

TABLE 3: PERCENTAGE INCREASE OVER THE TWO DECADES

Descriptive Terms	% Increase over Decades
Algorithm and Data Structures	21.34
Artificial Intelligence	252.81
Communication and Security	127.16
Computer Architecture	271.00
Computer Graphics	103.47
Databases	100.26
Programming Languages and Compilers	11.98
Scientific Computing	293.44
Software Engineering	>300.44
Theory of Computation	80.92

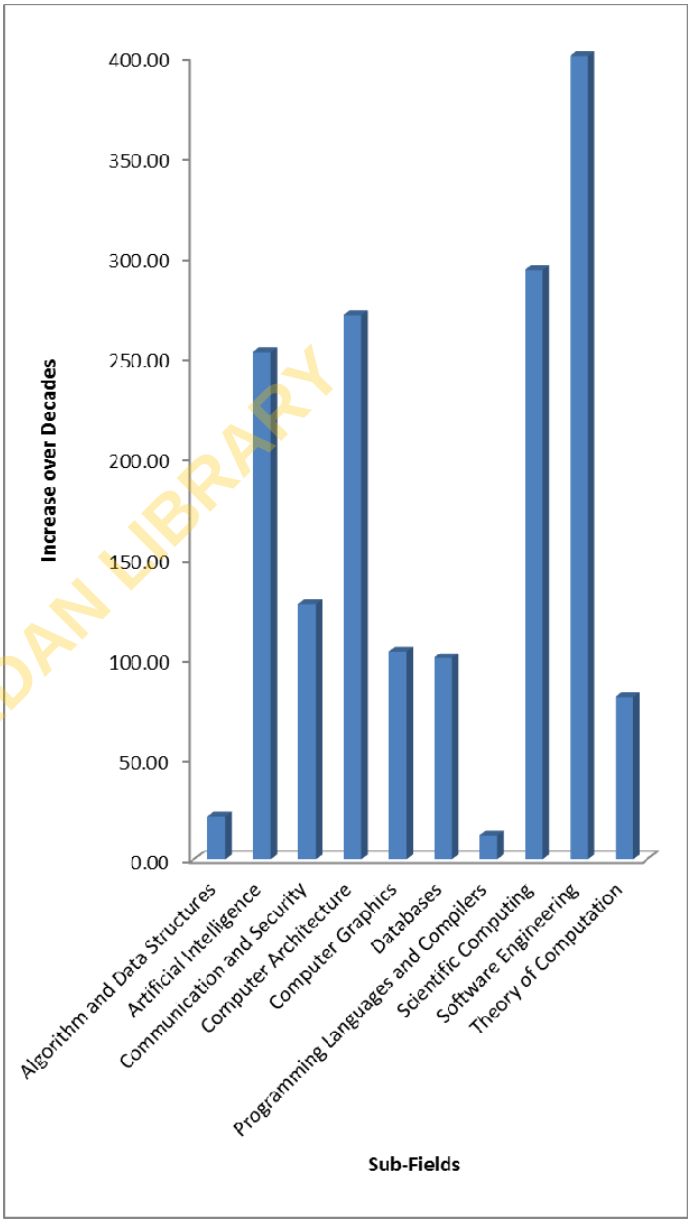

Figure 7: Percentage Increase over the Two Decades

Table 3 and Figure 7 showed percentage increase over the two decades of various disciplines in JCST. Computer architecture, artificial intelligence, scientific computing, software engineering and communication and securities were the disciplines where highest percentage increase was recorded. The least were databases and computer graphics. It is widely known that the growth of computer hardware: processors, embedded systems (such as, mobile devices) and controllers occurred in 2000s and hence we can expect more papers published in Computer Architecture and Artificial Intelligence

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

during the decade of 2003. It is gratifying to observe that trend in the plot (Figure 7).

IV. CONCLUSION

In this work, text mining is applied to figure out trends in research topics related to various subfields in Computer Science academic journal articles within the period of two decades. This analysis can also be extended to find trends in research topics related to other disciplines in the academic journal articles. A similar approach can also be used to analyse many academic electronic journal articles (corpus) in other fields. Text mining has tremendous potential in identifying trending topics during a period of time.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of [1] M. Hearst. *Untangling Text Data Mining*, in the Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, 1999.

[2] V. Gupta and G. S. Lehal, *A Survey of Text Mining Techniques and Applications*, Journal of Emerging Technologies in Web Intelligence, Vol. 1, No.1. August 2009.

[3] F. Louise, and M. Flynn, *Text Mining Handbook*, Casualty Actuarial Society E-Forum, 2010.

[4] D. Robb , *Taming Text*, Retrieved from http://vnweb.hwwilsonweb.com/hww/jumpstart.jhtml?recid=0bc05f7a67b1790e8bd354a88a41ad89a928d23360302a4959035699f17e2ba8a63e2dd032c73f8a7fmt=H, 2005

[5] P. Cerrito, *Inside Text Mining*, Retrieved from http://wilsontxt.hwwilson.com/pdffull/06619/275n6/gs9.pdf, 2005.

[6] Z. Shaik, S. Garia, and G. Chakraborty, *SAS® Since 1976: An Application of Text Mining to Reveal Trends*, Proceedings of the SAS Global Forum 2012 Conference, SAS Institute Inc., Cary. support.**sas**.com/resources/papers/proceedings12/135-2012.pdf, 2012

[7] B. Mirkin, S. Nascimento, L. M. Pereira, *Representing a Computer Science Research Organization on the ACM Computing Classification System*, in Eklund, Peter; Haemmerlé, Ollivier, Supplementary Proceedings of the 16th International Conference on Conceptual Structures (ICCS-2008), CEUR Workshop Proceedings**354**, RWTH Aachen University, pp. 57–65, 2008.

[8] Wikipedia: http://en.wikipedia.org/wiki/Outline_of_computer_science

[9] A. Kao and S. R. Poteet, (Eds), *Natural Language Processing and Text Mining*, Springer, London, UK, 2007.

[10] S. G. Cho, and S. B. Kim, *Identification of Research Patterns and trends Through Text Mining*, International Journal of Information and Educational Technology, 2(3). June 2012.

AUTHORS PROFILE

Adebola K. OJO is a PhD student in the Department of Computer Science, University of Ibadan, Nigeria. She is a registered member of the Computer Professional of Nigeria (CPN) and Nigeria Computer Society (NCS). She had her BSc in Computer Engineering from Obafemi Awolowo University, Nigeria. She also obtained her Masters of Science Degree in Computer Science from University of Ibadan, Nigeria. Her research interests are in Digital Computer Networks, Data Mining, Text Mining and Computer Simulation. She is also into data warehouse architecture, design and data quality via data mining approach.

Dr. Adesesan Barnabas Adeyemo is a Senior Lecturer at the Computer Science Department of the University of Ibadan, Nigeria. He obtained his PhD, M.Tech., and PGD Computer Science degrees at the Federal University of Technology, Akure. His research activities are in Data Mining, Data Warehousing & Computer Networking. He is a member of the Nigerian Computer Society and the Computer Professionals Registration Council of Nigeria. Dr. Adeyemo is a Computer Systems and Network Administration specialist with expertise in Data Analysis and Data Management.