Army Institute of Technology Publications (3.3.2)

Summary Sheet 2015-16

Total Number of papers						TOTAL
SCI/SCIE	ESCI	SCOPUS	WOS	UGC	OTHERS	
02	01	02			21	26



International Journal of Advanced Research in Computer and Communication Engineering Vol. 4. Issue 11. November 2015

Ant Colony Optimization Algorithm for Composition of Web Service using Mobile agents based Semantic, WSDL and QOS analysis

Mr. Sunil R Dhore¹, Prof. Dr M U Kharat²

Member IEEE, Research Scholar Sant Gadge Baba Amaravati University, Amaravati, M.S. India¹ MET's Institute of Engineering, Bhujbal Knowledge City, Nashik, M.S. India²

Abstract: Building a composite web service as per the user's complex need requires using the multiple online web services available. Selecting the best web service for composition will become combinatorial problem leads to be the NP complete problem. To solve this problem we are proposing a framework for composite web service using ant colony optimization algorithm. For choosing the proper path selection in ACO, we are negotiating with individual web service to select it for the participation during composite web service. For negotiation we are using intelligent agents for analyzing web services for based on semantic, WSDL and QOS description. As per the profile of the users the negotiation agents are built for building composite web services.

Keywords: Mobile agent, web service composition, QOS, SOA, ACO, WSDL.

1. INTRODUCTION

Aim of building composite web service is to satisfy the customer"s complex need. To fulfill the customer"s complex need, single webservice will not be sufficient. So single composite web service may require multiple numbers of abstract web services. And for the single basic requirement multiple numbers of abstract web services are available. So it is challenging to select the best abstract web service out of these.

Also building a composite web service is not the straightforward .It is a combination of different abstract web services together. And these abstract web services are combining together in sequence, parallel or in combination. So to build such composite web service we have to construct directed acyclic graph (DAG) to combine different abstract web services in sequence/parallel/combination.

To construct a DAG, we are proposing Ant colony optimization algorithm. And to select the proper abstract web service for participation in composite web service and to move forward during DAG creation using ACO, we are building agents to analyze abstract web service based on its features like WSDL Contents, Semantics and QOS parameters.

2. WEB SERVICE COMPOSITION

Sometime need of the end user will not be satisfy by the single abstract web service, so we need to combine multiple number of abstract web services together. The process of building such composite web service is called as web service composition (Fig, 01).

While building such composite web services, the combination of various abstract web services together are in sequentially one after the other or they may be combining in parallel or in combination of both.

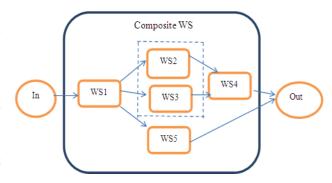


Fig.01: Composition of Web Service

3. SOFTWARE AGENTS [06]

A **Software Agent** is a computer program that acts for a user or other program in a relationship of agency. Software agents possess the following properties:

- **1. Locality Affiliation:** Agents are locality affiliated. Mobile agents are moved from one place to another during run time and task execution.
- **2. Role, Service Capacity:** Describes the kind of result an agent can produce It represents its functionality for task execution. It is sub divided into action type and task type.
- **3. Communication Behavior :** Each agent belonging to an instance of an agent system can communicate within its name space according to its behavior at any moment. The behavior of an agent determines whether it carries outtasks delegated to it in co-operation with other agents or whether it is capable of doing this on its own.
- **4. Negotiation Ability :** Negotiation ability describes the properties of an agent to execute a task collaborating with other agents and to negotiate this co-operation.
- **5. Delegation Ability:** Taking into consideration authority, agents can place and take on tasks. Delegation means that

ISSN (Online): 2277-5420 www.IJCSN.org

Impact Factor: 0.417

Analyzing Quality of Service Parameters of Abstract Web Services Using Software Agents for Building Composite Web Services

¹ Sunil R Dhore, ² Dr. M U Kharat

¹ Research Scholar SantGadge Baba Amaravati University, Amaravati, M.S. India

²MET's Institute of Engineering, Bhujbal Knowledge City, Nashik, M.S. India

Abstract - Composite web service is a combination of multiple abstract web services. To build a composite web service we have to combine multiple abstract web services together may be in sequence or in parallel. On the fly to select the best web service from the multiple abstract web services available for the same purpose is the challenging job and leads to the NP complete problem. So we are proposing an analyzing model of the abstract web services using software agents. Due to seamless integration of web services and mobile agents, agent based framework is a natural choice. We are proposing the analysis part of web services and users with the help of software agents. And decision taken by this analyzing agent will be used to select the best abstract web service which will be the part of composite web service. Analyzing agents use the QOS parameter and Characteristic of web services for selecting best web service.

Keywords - Mobile Agent, Web Service Composition, QOS, SOAP, WSDL, Restful Web Services, Abstract Web Service.

1. Introduction

Aim of building composite web service is to satisfy the customer's complex need. To fulfill the customer's complex need, singleabstract webservice will not be sufficient. So single composite web service may require multiple numbers of abstract web services. And for the single basic requirement multiple number of abstract web services are available. So it is challenging to select the best abstract web service out of these. Also building a composite web service is not the straightforward. It is a combination of different abstract web services together. And these abstract web services are combining together in sequence, parallel or in combination. So to build such composite web service we have to construct directed acyclic graph (DAG) to combine different abstract web services sequence/parallel/ combination.

In directed acyclic graph, each node is related to individual abstract web service which is the part of composite web service. To select individual abstract web service for constructing composite web service on the fly is difficult, because for each individual subtask of composite web service numbers of similar abstract web services are available. So we are taking the advantage of seamless integration of web services and software agents. We are proposing various mobile agents which will analyze individual abstract web services. To analyze abstract web services, we are using the unique features of web services based on its description, type and various quality of service possess by individual abstract web service. These intelligent agents are also using to priorities an users need of quality of service. As per the users expectation coordinating agents select the proper individual abstract web service to take part in composing a composite web service.

2. Related Work

Authors of Ref [01] concentrates on combining optimization and ranking based on non-functional QoS parameters to evaluate its quality. Ref[02] However use innovative idea for selecting best web service using the reliability issue for designing QoS-aware optimal selection scheme for web services with a trusted environment. We can also compare the web services based on the different types of web services, explain in Ref[05]. Author of Ref[06] presents a comparative performance evaluation of two Web service implementations: one is based on SOAP and the other on Representational State Transfer (REST). We utilized response time and throughput metrics to compare the performance of these Web services. They found that, on average, REST has better performance

e- ISSN: 2348-4470 p- ISSN: 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 4, April -2016

Learning Non-linear Dynamical Systems From Raw Images

Amit Chahar¹, Anil Kumar²,Rohit Rajput³, Venkata Ramesh L⁴, Prof. Sunil Dhore⁵

1-5Computer Engineering, Army Institute of Technology

Abstract —We introduce a method for model learning and control of non-linear dynamical systems from raw pixel images. It consists of a deep generative model, belonging to the family of variational autoencoders, that learns to generate image trajectories from a latent space in which the dynamics is constrained to be locally linear. Our model is derived directly from an optimal control formulation in latent space, supports long-term prediction of image sequences and exhibits strong performance on a variety of complex control problems. For capturing the information of non-linear object's behavior, we need to use high-dimensional data. Processing the high-dimensional data is expensive and not feasible. So, in this model, first Auto-encoder is used for dimensionality reduction, and after prediction method (transition mapping) is used, and the imagereconstructed. We demonstrate that our model enables learning good predictive models of dynamical systems from pixel information only.

Keywords-machine learning, autoencoder, neural networks, latent space, non-linear systems, prediction, dynamical systems.

I. INTRODUCTION

Dynamical systems are mathematical objects used to model physical phenomena whose state (or instantaneous description) changes over time. These models are used in financial and economic forecasting, environmental modeling, medical diagnosis, industrial equipment diagnosis, and a host of other applications. If we have two short movies of billiards balls rolling around on a table without friction, we could not tell which was recorded first. Hence this system is stationary. On the other hand, if there is friction, then we are in the non-stationary situation, because the balls will slow down as time progresses, and their speed gives us a way of deducing when the observation was made.

A key challenge is system identification, i.e. finding a mathematical model of the dynamical system based on the information provided by measurements from the underlying system. In the context of state-space models this includes finding two functional relationships between (a) the states at different time steps (prediction/transition model) and (b) states and corresponding measurements (observation/ measurement model)[1].

Control of non-linear dynamical systems with continuous state and action spaces is one of the key problems in robotics and, in a broader context, in reinforcement learning for autonomous agents. A prominent class of algorithms that aim to solve this problem are model-based locally optimal (stochastic) control algorithms. When combined with recedinghorizon control, and machine learning methods for learning approximate system models, such algorithms are powerful tools for solving complicated control problems [3, 4, 5]; however, they either rely on a known system model or require the design of relatively low-dimensional state representations. For real autonomous agents to succeed, we ultimately needalgorithms that are capable of controlling complex dynamical systems from raw sensory input (e.g. images) only. In this paper we tackle this difficult problem.

II. NON-LINEAR DYNAMICAL SYSTEMS

A dynamical system will be defined to be a system in which the present state (the values of all of the variables and all of their derivatives) is somehow dependent on previous states of the system. A deterministic system will be taken to be a system in which the present state is entirely dependent on previous states of the system. A linear system is a system in which all of the dependence of the current state on previous states can be expressed in terms of a linear combination. A linear stochastic system is a system in which all of the dependence of the current state on previous states can be expressed in terms of a linear combination and the residual unpredictable portions can be expressed as additive, independent, identically distributed, random variables[2].

A nonlinear system is a system in which the dependence of the current state onprevious states cannot be expressed entirely as a linear combination; even if some of the dependence can be captured in a linear combination of the previous states, something extra is required to capture all of the dependence.

Simulations of nonlinear dynamical systems have shown that nonlinear time series can be entirely deterministic, that is generated without any random component, and yet exhibit behavior which appears to have an error variance when analyzed by linear statistical methods. This work will present a variety of techniques for the analysis of nonlinear time series which have the potential to be modeled as signal portions of time series that are often discarded as noise. Learning non-linear dynamical models from veryhigh-dimensional sensor data is even more challenging. First, finding (non-linear) functional relationships in very high dimensions is hard (un-identifiability, local optimal, over-fitting, etc.); second, the amount ofdata required to find a good function approximation is enormous. Fortunately, high-dimensional data often



International Journal of Advance Engineering and Research
Development

e-ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 2, February -2016

Augmented Reality in Computer Education using Android

Deep Narain Sodhi¹, Anubhav Singh², Vineet Chaurasiya³, Juhi Choubey⁴, Prof. N.K.Bansode⁵

^{1,2,3,4,5}Army Institute Of Technology

Abstract — Augmented Reality is a concept wherein one's perception of environment is improved by superimposing digital information like graphical, textual, or audio content, as well as objects onto a display screen. It is in between realand virtual reality, wherein virtual objects are seen in real world. The application will be mobile based.

The idea is to allow the user to view the virtual object in education in the real world using a marker based AR system. We propose that the system should be used in computer education by augmented animations of various parts of computer. It is not expensive as the user does not actually need to open parts of a computer, but instead just needs to view the animation.

The Android Application will scan the marker and the animation will be seen on the display screen of the mobile device along with the real environment.

Keywords-virtual reality, real, marker, augmented, android

I. INTRODUCTION

Augmented Reality known as Mixed Reality combines virtual and real scenes to achieve that virtual ones are belong to the real world. Because of this integration it is being used in various applications like medical, education, and entertainment.

There are 2 ways in which this can be implemented:

Marker Based:

There are different types of AR markers which can be detected by a camera and used with software as the location for virtual objects placed in a scene. They are image descriptors or black and white images (features + key points)..Simple augmented reality markers can consist basic shapes made up of black squares on a white background. Different Markers can be created using simple images that are still read properly by a camera, and these codes can even take the form of tattoos.

The simplest types of AR markers are black and white images that consist of two dimensional (2D) barcodes.

Marker less:

A marker less AR application recognizes images which were not provided to the application beforehand. This is more difficult to implement because the recognition algorithm running in the AR application should identify patterns, colors or some other "features" that may exist in camera frames

II. MARKER DETECTION, DESIGN AND REGONITION METHOD

Markers are square and have a black thick border and black graphics within its white internal region. The advantage of using black and white color is to separate the marker from the background easily. In terms of projective geometry, the square markers in real world could not be a square after projecting it onto an image plane, in other words, the graphics which are internal in the markers often display in distortion. So it is necessary to unwrap these markers when we recognize them

The procedure of unwrapping these image is shown in the figure

After detecting the grabbed frame the calculation of the marker unwrapping can be described as follows: (xc, yc), i = 1,2,3,4 as the four corners of a marker which are acquired. These positions in the real world of the four corners are given by (xm, ym), i = 1,2,3,4. Homography matrix H is calculated as shown in figure:



International Journal of Advance Engineering and Research Development

e-ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 4, April -2016

Analysis of Cloud Computing Utilization from user's perspective using intelligent tool

M B Lonare, Mayur Indalkar, Rajnish Kumar, Shubham Verma, Vivek Mahto

Computer Department, Army Institute of Technology

ABSTRACT- Cloud computing is a way of computing which includes sharing computing resources rather than having local servers or some personal devices to handle applications or store data. It is a prime need of today's IT world and in order to increase benefits, organizations are rapidly shifting to cloud computing. In this form of computing resources such as servers, storage and applications are made available to organizations through Internet. The vendor complete the request of the user through a package, the user procure the resources in the beginning and they don't know how much they are using ,so they end up paying for the whole package. In this paper we have made a novel attempt to propose a technique to calculate the real time utilization of cloud resources. Further, in this paper we have proposed an algorithm/tool for continuous updating of cloud resource's use and thus increasing the user's profit.

Index Terms- Cloud Computing, Cloud Resources, Cloud Resources allocation, Cloud Resources analysis

I. INTRODUCTION

Cloud computing is the most prevailing technology that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Users of cloud computing request for different computing resources and the cloud service providers pick up the required resources from cloud resource pool and provide these as per the requirement of individual user. In order to accomplish as many requests as possible, it is very much necessary to utilize cloud resources effectively and wisely. A calculated use of cloud resources helps the organization or an individual user to save their money by using resources only they required and also enhances the performances.

Cloud computing [1] is one of the prime need of today's IT world. Organizations are shifting to the cloud rapidly in order to increase their overall benefits. Cloud computing has Virtualization as its backbone. Cloud resources are provided on demand using the Internet and an on-running migration of resources is done by the cloud service provider.

So it is very important to monitor the cloud resource allocation and utilization. The very basic motivation behind this is a particular user procure certain resources from any vendor as per their use [2], but at the vendor end only that much resources are deployed which the user is actually using, so it is a notable loss at users end and user is unaware of it.

The major problem with the traditional monitoring system is that, they provide only domain specific information. For example, a networking tool monitors only the network packets. Therefore, these performance tools report resource utilization statistics and good view of individual components. But they fail to give a good view of entire cloud infrastructure.

The end users look for intelligent reports to actually point to what, where and when action is needed. So, it is essential to have a monitoring system that correlates all the components and give a consolidated report about the resource availability and consumption of resources in a cloud infrastructure so that, customers believe that they pay only for what they are getting.

This brings us to the problem how to measure the actual resource utilization as per user prospective, so as to provide user with an insight of their actual requirement which would in turn not only minimize their cost overhead and maximize their profit margins but also save them from probable exploits [3].

In this paper we have proposed algorithm to continuously monitor the utilization of cloud resources as per utilization and also provided with a solution, a tool, which would monitor the actual use of resource and provide user with useful insights.

II. RELATED WORK

This section discusses prior work related to our research.

i) Cloudyn

Cloudyn is able to complete various aspects of the proper cloud computing utilization. It optimizes the cloud and ensure that the cloud deployment is optimally priced and utilized, it has also helped in cutting the costs and eliminating the shocking monthly bills. Gaining insight and visualizing the usage and performance trends all in one place is the new feature of the Cloudyn and it also helps in comparing clouds and research the deployment's performance.



International Journal of Advance Engineering and Research Development

ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 2, February -2016

Docker Management Using Libvirt API

Amit Chahar¹, Ajay Shekhawat², Ankit Mishra³, Komal Kumari⁴, Prof. P.R. Sonawane⁵

^{1,2,3,4,5}Department of Computer Engineering, Army Institute of Technology, Pune, India

Abstract—Docker automates the deployment of applications inside software containers by providing an additional layer of abstraction. Docker implements a high-level API to provide lightweight containers that run processes in isolation. It uses resource isolation features of the Linux kernel such as cgroups and kernel namespaces to allow independent containers to run within a single Linux instance, avoiding the overhead of starting and maintaining virtual machines.

Libvirt is used by the most clouds to give user access to the cloud and has bindings in other languages also like java, python, ruby. It support LXC (linux containers) but does not support docker containers. Initially docker used LXC as driver but because of it being managed by open source community, docker could not rely on it. So, docker developed its own driver libcontainer in go language.

Docker is todays emerging technology but clouds can support docker only by using their own apis specially designed for docker since libvirt does not support libcontainer. This increases the complexity of the cloud. Proposed solution is to implement docker api in c and integrate it with the libvirt api. Thus, clouds will have to give access to only libvirt without using any special api for docker. On the other hand, the docker interface will be generic for all clouds, thus user does not have to face the difficulty while migrating from one cloud to another.

Keywords—libvirt; linux containers; docker containers; libcontainer; cloud computing

I. INTRODUCTION

Most commercial cloud computing systems, both services and cloud operating system software products use hypervisors. Enterprise VMware installations, which can rightly be called early private clouds, use the ESXi Hypervisor. Some public clouds (Terremark, Savvis, and Bluelock, for example) use ESXi as well. Both Rackspace and Amazon Web Services (AWS) use the XEN Hypervisor, which gained tremendous popularity because of its early open source inclusion with Linux. Because Linux has now shifted to support KVM, another open source alternative, KVM has found its way into more recently constructed clouds (such as ATT, HP, Comcast, and Orange). KVM is also a favorite hypervisor of the OpenStack project and is used in most OpenStack distributions (such as RedHat, Cloudscaling, Piston, and Nebula). Microsoft uses its Hyper-V hypervisor underneath both Microsoft Azure and Microsoft Private Cloud.

However, not all well-known public clouds use hypervisors. For example, Google, IBM/Softlayer, and Joyent are all examples of extremely successful public cloud platforms using containers, not VMs. Docker is an open-source project that automates the deployment of applications inside software containers, by providing an additional layer of abstraction and automation of operating-systemlevel virtualization on Linux. Docker uses resource isolation features of the Linux kernel such as cgroups and kernel namespaces to allow independent containers to run within a single Linux instance, avoiding the overhead of starting and maintaining virtual machines. Docker implements a high-level API to provide lightweight containers that run processes in isolation. Building on top of facilities provided by the Linux kernel (primarily cgroups and namespaces), a Docker container, unlike a virtual machine, does not require or include a separate operating system. Instead, it relies on the kernels functionality and uses resource isolation (CPU, memory, block I/O, network, etc.) and separate namespaces to isolate the applications view of the operating system. Docker accesses the Linux kernels virtualization features either directly using the libcontainer library, which is available since Docker 0.9, or indirectly via libvirt, LXC (Linux Containers) or systemd-nspawn.

The libvirt project develops a virtualization abstraction layer, which is able to manage a set of virtual machines across different hypervisors. The goals of libvirt are to provide a library that offers all necessary operations for hypervisor management without implementing functionalities, which are tailored to a specific virtualization solutions and which might not be of general interest. Additionally, the long-term stability of the libvirt API helps these management solutions to be isolated from changes of hypervisor APIs.

Most of the clouds use libvirt api for hypervisor management as libvirt gives a common interface to the user for all hypervisors. Since docker has stopped supporting LXC drivers because LXC drivers are maintained by open source community and therefore docker developers had to made frequent changes to the implementation. Libvirt only supports LXC drivers, therefore it is difficult to use docker on clouds. Also some clouds have separate apis for docker management. Therefore having separate apis of docker and libvirt are inconvient to both the cloud service provider and the user.

Accessing and Modifying Sqlite Remotely for Catering Multi-Client Access

Sharayu Lokhande Department of Computer Engineering Army Institute of Technology, Pune Vaishali Ganganwar Department of Computer Engineering Army Institute of Technology, Pune

Abstract: SQLite is a lightweight database management system and a Stable serverless database with almost zero difficulty in installations. SQLite does not support client server facility due to the write lock issue. For expedite multi-client access to the central database, multiple instances of the database on the central system can be created and later integrating these instances to give the resultant product. Accessing these instances remotely would be a solution to the write lock issue. As a result of creating multiple instances of the database on the same system, there might be a heavy traffic which could lead to reduce performance. To handle this cloud computing concept of High Availability which refers to a system or component that is continuously operational for a desirably long length of time.

Keywords— Remotely Access, SQLite, High Availability

1. Introduction

A lightweight database system is a high- performance,

application-specific Database Management system. It differs from a general- purpose (heavyweight) [1] DBMS in that it omits one or more features and specializes in the implementation of its features to maximize performance. Although heavyweight monolithic and extensible DBMS might be able to emulate LWDB capabilities, they cannot match LWDB performance.

SQLite is a software library that implements a SQL engine. It has been used with great success as on-disk file format: allows the developer to handle data in a simple way, but also have the use of database features (such as undo, redo, etc.). In embedded deviceenvironment, in which there is low-concurrency and there are little ormedium size datasets, SQLite is the right choice. If we want to save the data in a common place, i.e., Remote Server until now there is no easy mechanism to implement this.

The need for storing information in remote server exists to have centralized access to data by the users. The idea of storing information in remote server is implemented using Web Services (plugin) which can save the data in the Remote database like SQL Server and retrieve as and when required. When a project is developed, a group of developers/testers are involved. They will needconcurrent information for development which can be done using a centralized database. For example feedback is collected from different customers for a product and it is more feasible to store it ina centralized repository that can be used by the entire for improvements and further development. So we require a remote access to SQLite [2]

to be used by all of them. The relevant changes need to be reflected and others discarded. SQLite has write lock issues which have to resolve by creation of different instances of the database. Testers can access and debug the problems directly and provide the information without having to install the entire system or database files.

1.1 High availability

Virtualization, a technique to run several operating systems simultaneously on one physical server, has become a core concept in modern data centers, mainly driven by benefit of application isolation, resource sharing, fault tolerance, portability and cost efficiency. A special middleware, hypervisor, abstracts from physical hardware resources and provides so called virtual machines actinglike real computers with their own (virtual) hardware resources. High availability system [3] design approach and associated service implementation that ensures a prearranged level of operational performance will be met during a contractual measurement period. Enabling high availability we can detect any point of failure to propagate reliable crossover, if needed. High availability is a characteristic of a system. The definition of availability is Ao = up time / total time. If (total time - down time) is substituted for up time then you have Ao = (total time - down time) / total time. Determining tolerable down time is practical. From that, the required availability may be easily calculated. Here a small network has made with a master, slave (replica of master) backing up data, controller and a user virtual machine. Controller will be constantly checking the master for downtime and doing crossover to slave in case tolerable down time is exceeded. For this purpose we will use open source tools like heartbeat, pacemaker and DRBD.

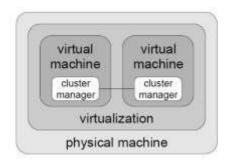


Fig. 1 Virtualization

www.ijsea.com 350



Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

User Analytics on Twitter Stream Data

Sagar Rane, Manik Hendre, Sharayu Lokhande

.....

Assistant Professor, Dept. of CSE, Army Institute of Technology, Savitribai Phule Pune University, Pune, India.

Manuscript Info

Manuscript History:

Received: 14 December 2015 Final Accepted: 19 January 2016 Published Online: February 2016

......

Key words:

Hadoop, Big Data, Map Reduce, Twitter, HDFS, Tweets, Sentimental Analysis, Flume.

*Corresponding Author

Sagar Rane.

Abstract

Twitter, is one of the largest social media site that receives tweets in millions of data in every day in range of Petabytes per year. Big Data is a pool of information that is outsized and difficult to progression by data processing applications, Hadoop is a disseminated archetype used to handle the huge quantity of documents. It grasps the vast quantity of documents and carry out the procedures like documents analysis, outcome analysis, and records analytics. It is highly scalable computing platform. Productive E-commerce sites, Facebook, Twitter one of the largest social media site receives comments, tweets or customer reviews in millions every day in the range of terabyte or petabytes per day. Ideas and opinions of people are influenced by the opinions of other people. Lot of research is going on analysis of reviews given by people. We can collect the data from the social media site by using BIGDATA eco-system using online streaming tool Flume. This huge amount of raw data can be used for industrial or business. This Analytics paper provides a way of analyzing of big data such as Twitter data using Apache Hadoop which will process and analyze the tweets on a Hadoop clusters. In this paper, we are going to talk how effectively sentiment analysis is done on the data which is collected from the Twitter using Flume. Twitter is an online web application which contains huge amount of data that can be a structured, semi-structured and un-structured data. Twitter is also difficult due to language that is used for comments. So here we are taking sentiment analysis, for this we are using Hive and its queries to give the sentiment data based up on the groups that we have defined in the HQL (Hive Query Language). Here we have categorized this sentiment analysis into 3 groups like comments that are having positive, moderate and negative comments.

Copy Right, IJAR, 2016,. All rights reserved.

Introduction:-

Big data is recycled ubiquitously at the present in disseminated archetype on web. BIG data is the group of collections of massive volume of data. So Big data came into picture in the real time business analysis of processing data. Some well-known internet companies like Google, Amazon, LinkedIn, Yahoo! etc. have generated a huge amount of structured and unstructured data every day. This exponential growth of data leads to some challenges like processing of large data sets, extraction of useful information from online generated data sets etc. Twitter.com is a popular microblogging website. Each tweets is 140 characters in length. Tweets are frequently used to express a twitter's emotion on particular subject. The upcoming of operational societal mass media and communication machineries has activated a quick rise in the stream of user produced content of several forms. Persons are precise their responses, desires and preferences through societal mass media via means of word-based piece of short nature relatively scripting extensive writing. We appeal groups complete through this old-fashioned way as G-friends, which stands for topographical geo location based groups. So Big data came into picture in the real time business analysis of processing data. Some well-known internet companies like Google, Amazon, LinkedIn, Yahoo! etc. have generated a huge amount of structured and unstructured data every day. This exponential growth of data leads to



International Journal of Advance Engineering and Research Development

ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 4, April -2016

User Interface to evaluate Fuzzy Matching Results

Chirag Nagpal¹, Neha Malik², Ajit Singh³, Shushma Shirke⁴, Shinsmon Varghese⁵

¹⁻⁵Dept. Of Computer Engineering, Army Institute of Technology

Abstract —Entity Resolution is a challenging problem, even more relevant in todays' time of big data. In this paper, we present a User Interface framework in order to perform and evaluate results of the Entity Resolution process. The user interface provides an intuitive experience to an end user to experiment with various parameters, and machine learning algorithms to perform fuzzy clustering based Entity Resolution.

Keywords-component; formatting; style; styling; insert (key words) (minimum 5 keyword require) [10pt, Times new roman, Italic, line spacing 1.0]

I. INTRODUCTION

Entity Resolution is a challenging problem, Entity Resolution is a challenging Computer Science problem. In today's world of Big Data, and the Semantic Web, entity resolution has become an even more important problem to solve. Today, the internet provides extensive amount of data. This data, might be in the form of structured formats like SQL Tables, or maybe completely unstructured with multiple hierarchical structures.

This unstructured data, is a modern feature of today's world of Big Data, and any computational task requiring the exploitation of this data requires it to be arranged, systematically, such that standard algorithms can be easily deployable across a wide range of data sources. The unstructured and noisy nature of this data, also necessitates newer more robust approaches to solve the Entity Resolution problem.

Traditional ER approaches employ extremely naive comparisons between individual features in order to solve the ER problem. Thus a traditional ER approach, might involve the comparison of a combination of certain specific features from two individual databases. In case of a match, these records are grouped together, or in other words resolved into a single entity.

A common extension to such an approach is the 'Merge-and Purge' strategy in which records are compared, pairwise, and if found to be matched, they are merged into a single record. Such a strategy is computationally more tractable as compared to naïve pairwise comparison. [1]

While such an approach might work well, for well structured datasets, however, determining such crisp rules, to resolve unstructured datasets is a challenging task. This challenge stems from the fact that it is extremely difficult to determine a single rule or combination of rules, that can effectively predict a match with perfect confidence, and such a rule can obviously not be determined manually, owning to the possibility of large dimensionality of the data. On the other hand, the dataset under consideration might be incomplete, that is, might be missing in certain features, and hence a specific rule or set of rules may not be applicable. This is especially true in case of data crawled from the internet, like for example microblogs.

A case in example could be microblogs. Let us consider an Entity Resolution problem involving the resolution of all tweets posted at from a particular location. While in some cases, microblogs may contain specific features like the geo-location of posting, in other cases it might contain this particular feature. Thus a simple rule involving the resolution of all tweets posted at a particular area may not be adequate. Here, the match function would require a more sophisticated approach of looking at other features, like if the post is from the same user and the time difference between the various posts is minimal, we can predict that the posting location is the same.

Such complex relations between various features can not be modelled easily with simple feature subsets and require more complicated approaches. A reasonable approach here would be to leverage machine learning approaches in order to learn such complicated features to predict a match. Incase we tried to model the problem above using a standard supervised learning model like decision tree, it would have been able to learn the complicated relations between various individual features in the features, thus able to come up with a representation for the problem. In the most obvious case a decision tree for this task would have looked at the location information, incase not present, it would then have tried to look for if the user is the same and then, the posting difference. In case this corresponded to the previously learnt values for these features, the algorithm would have been able to compute if there is a match.

In such a specialized, specific application, we require to model this problem in an unsupervised or semi supervised setting, in order to carry out the ER process, without the need of labelled training data. [2] In this paper, we present a framework to carry out the semi supervised pipeline to perform entity resolution. Our semi supervised pipeline depends



International Journal of Advance Engineering and Research **Development**

e-ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 3, March -2016

Network Application Testing Platform using Openstack and OpenDaylight

Ritesh Singh Parihar¹, Sarwan Kumar Rai², Janardan Singh Parihar³, Yogita T. Hambir⁴

1,2,3,4Department of Computer Engineering, Army Institute of Technology

Abstract —The networking environment today is increasingly complex. The convergence of data center, business expansion and new service deployments need to combine several different technologies - including physical and virtual devices - into a single network infrastructure. These networks must handle massive amounts of traffic with critical applications on both issues to the point and multicast network strategies. You need high-capacity network devices to route and switch huge amounts of traffic. Network equipment manufacturers are positive that their equipment can not only handle the needs of a complex network infrastructure, but also easily expand - often at massive scales. suppliers and network service companies must be able to highlight their infrastructure to see if they can handle the traffic volumes and provide the best quality service.

With our project, we will make an assessment and gather the requirements to provide our developers, multi-node OpenStack test environments on demand where they can, in a whiteboard like fashion, configure different network functions, edit them in all different types of distributions. The main objectives are the ease and speed of the test lab configuration, safety, repeatability and cost.

Our benchmarks conduct functional and performance testing and probing the weaknesses and vulnerabilities. As such, we have to push the changes to a variety of environments with multiple physical and virtual components.

Keywords- Opendaylight; Openstack; RESTful API; Virtual Machine; Devstack

I. INTRODUCTION

A centralized platform for testing and analyzing the performance of network application is going to developed. The platform will provide the administrator with features to test:

- 1. Multiple OS support
- 2. Direction and rate of traffic flow
- 3.Resources required and usage
- 4.Security
- 5.Cost
- 6. Integration for already present and future technologies

It is done by using Opendaylight Restful and neutron APIs which is the local API support provided for Opendaylight and Openstack Integration.

II. RELATED WORK

OpenDaylight is an open platform for network programmability to enable Software-Defined Networking (SDN) deployments on heterogeneousmodern multi-vendor networks. OpenDaylight provides a platform abstraction model oriented service that allows users to easily write applications that work across a variety of hardware and South terminal protocols.

OpenStack is a platform for free and open-source software cloud computing, mostly deployed as an infrastructure -as-aservice (IaaS). The software platform consists of interrelated elements that control pools of processing equipment, storage, and network resources across a data center. Users manage either through a dashboard based on the Web, through command line tools, or through a RESTful API.

Z COS

e-ISSN (O): 2348-4470 p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 5, May - 2016

IMAGE COMPRESSION FOR EFFICIENT TRANSMISSION AND STORAGE OVER NETWORK

Diwakar kumar singh¹, Nitin Yadav², Jitender Singh³, Satender Shakya⁴, Asst. Prof. Vaishali Ganganwar

5

- ¹Department of Computer Engineering, Army Institute of Technology, Pune
- ²Department of Computer Engineering, Army Institute of Technology, Pune
- ³Department of Computer Engineering, Army Institute of Technology, Pune
- ⁴Department of Computer Engineering, Army Institute of Technology, Pune
- ⁵Department of Computer Engineering, Army Institute of Technology, Pune

ABSTRACT: In recent' years the satellite communication has been growing for the development of rapid and efficient techniques for the storage and transmission of satellite images. The science of reducing the number of bits required to represent the image is called image compression. In the Existing approaches to reduce the size of the image is being transmitted and it is reconstructed in the receiving side so that the compressed image is being decompressed to obtain the original image by using the wavelet to improve the compression ratio as well as visual quality which is implemented by the algorithm which are a combination of spatial and Coding algorithm. In the Proposed System to compress the Original image by using the two techniques which called as Huffman and Run Length Coding Compression is defined to compress the Patches of an images Based on the PSNR and MSE value is to find the compression ratio. In this approaches Huffman compression is implemented to compress the still image. The 2D discrete Cosine transform (DCT) has been applied and the detail matrices from the information matrix of the image have been estimated the image as a Patches Conversion. The reconstructed image is synthesized using the estimated detail matrices and information matrix provided. The Patches of images is to be compressed by using these techniques. The quality of the compressed images has been evaluated using some factors like Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE).

1. Introduction:

Data compression is defined as the process of encoding data using a representation that reduces the overall size of data. This reduction is possible when the original dataset contains some type of redundancy. Digital image compression is a field that studies methods for reducing the total number of bits required to represent an image. A commonly image contain redundant information i.e. because of neighboring pixels which are correlated and Contain redundant information. The main objective of image compression [1][9] is redundancy and irrelevancy reduction. It needs to represent an image by removing redundancies as much as possible, while keeping the resolution and visual quality of compressed image as close to the original image. Decompression is the inverse processes of compression i.e. get back the original image from compressed image.



International Journal of Advance Engineering and Research Development

e-ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 5, May -2016

Web Framework for Data Deduplication at Scale

Chirag Nagpal¹, Neha Malik², Ajit Singh³, Shushma Shirke⁴, Shinsmon Varghese⁵

1,2,3,4,5 Dept. Of Computer Engineering, Army Institute of Technology,

Abstract —In todays time of multiple heterogenous sources of data, data dedupliction is a difficult challenege. We present a Web Based framework that utilizes the popular MVC paradigm to provide the end user with a functionality allowing Data Deduplication. In the backend the framework allows the use of distributed computing using a Map Reduce strategy to scaleu p our de deuplication schme.

Keywords-component; formatting; style; styling; insert (key words) (minimum 5 keyword require) [10pt, Times new roman, Italic, line spacing 1.0]

I. INTRODUCTION

We live in the day and age of big data with multiple heterogeneous sources of data from the Internet of Things to the Semantic Web. With multiple such data records, there is inherent possibility of data redundancy in these strict schema driven databases. Thus a need is felt to utilize recent trends in machine learning techniques in order to resolve such inherent redundancies arising from various sources including human error.

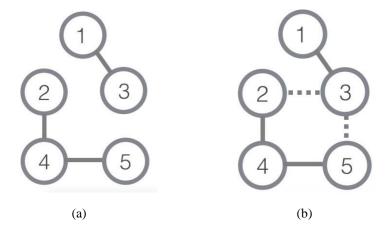


Fig. 1 (a) The graph nodes represent the entities before Data Deduplication. (b) Represents the records after Deduplication the dashed edges represents the records that are found to be similar using the deduplication framework.

In this paper we present a web based framework, along with a Map Reduce backend in order to perform this large scale fuzzy matching based Entity Resolution. The framework allows users without any prior domain knowledge of machine learning or distributed computing to perform data deduplication on these datasets. The web framework is flexible and allows decoupling of the underlying Map Reduce backend with a Model-View-Controller based frontend.

II. MVC FRAMEWORK

The MVC[1] is an extremely popular design pattern encountered in software development. MVC frameworks are the preferred for various large scale web platforms. MVC frameworks allows the decoupling of components into the application into three basic parts – The Model, The View and the Controller.

2.1. Model

It refers to the data of the application. The model comprises of the rules and logic that govern how data is stored and manipulated. Popular MVC frameworks support various different Model paradigms like RDBMS, where data is stored in strict schema driven relational databases, such a strategy is employed by SQL or Structured Query Language. Newer MVC frameworks also support the use of paradigms like NoSQL databases, which are extremely flexible in terms of the schema of the data to be stored.



International Journal of Advance Engineering and Research Development

e-ISSN (O): 2348-4470

p-ISSN (P): 2348-6406

Volume 3, Issue 3, March -2016

Hybrid App Suite for Medical Diagnosis using Naive Bayes' Classifier

Praveen Hore, Assistant Professor, Army Institute of Technology, Pune Anuj Tomar, Ashutosh Thakur, Nikhil Judge, Vishal Yadav Department of Computer Engineering, Army Institute of Technology, Pune

Abstract—Medical and Healthcare industry comprises of sec- tor of industries that focus on patient well- being and medical research. These industries make up about 10% of the GDP of any developing nation. In India, the worth of medical industry is about \$78 billion and it is expected to rise to \$200 billion by 2020 [15]. IT industry has touched almost every domain of industries, medical industry remains to be exploited. IT industry will help in automation, better record management, and research in the healthcare industry. With the traditional methods, chances of medical negligence and misdiagnosis are prominent. We can use information technology to facilitate medical practitioners by providing them an interface to make calculations on-the-go, retrieve medical records and provide diagnosis by eliminating guesswork. By making calculations automated, degree of error is drastically reduced, which increases precision. Chances of medical record loss can be handled by maintaining a central server. Also, confidentiality of medical records can be achieved. In the current scenario, applications have to be developed for every platform separately. This consumes both time and money. Our solution proposes a hybrid application that will run on every platform. The application development is faster, simpler, and more rapid and the application is easier to maintain. You can change platforms anytime you need, Cordova lets you build your application for more than one platform just by one adding line of code.

Keywords—Hybrid application, Prediction, Cross-platform porta-bility

I. INTRODUCTION

Medical practitioners require a lot of complex calculations in which the amount of precision expected is very high. Such calculations might involve medical inputs such as sugar level, haemoglobin, etc. The first part of our app suite facilitates that, providing on-the-go calculations to medical practitioners. This can help them to make medical decisions. The second part is related to medical diagnosis using Naive Bayes'classifier. Prediction can be done using this machine learning technique.

Both these parts come under the umbrella of a hybrid app suite which facilitates the use of the app on multiple platforms without the need of writing large pieces of extra code.

A. Motivation

The future for medical apps is bright. There will coexist both native and web based apps. App stores will continue to thrive and the market for consumer directed health apps will thrive. However, apps directed at physicians will become increasingly better connected to patients clinical records and will more likely be distributed by hospitals and other provider institutions.

According to industry surveys, by 2018, more than 1.7 billion smartphone and tablet users will have downloaded at least one mobile medical app (MMA) [16]. This level of acceptance and adoption of medical apps means new trends in innovation, and low cost medical services. These apps often control human physiology and work on sensitive health data, thus it is necessary to have evidences of their trustworthiness before actual marketing.

B. Literature Survey

Over the last fifteen years, IEC 62304 has become the benchmark standard for the development of medical device software, whether standalone software or otherwise, in both the EU and the US [17]. Leading industry innovation in software technologies has led key industry leaders and government regulators to recognise the emergence of numerous standalone medical software products that operates as medical devices. This has been reflected in regulatory changes. Example: European medical devices directive MDD/93/42, updated in 2007. In Europe, a guidance document has been published on this subject. [18] The quality management system requirements for manufacturing a software medical

International Journal of Computing Science and Communication Vol. 1 Issue 1 2016 www.ijcsc.co.in

Effective Searching Mechanism with Feedback Session

Sagar Rane

Department of Computer Engineering,

Army Institute of Technology, Pune, MH, India

ABSTRACT:

These days Internet is broadly utilized by clients to fulfil different data needs. In any case, questionable inquiry/theme submitted to web crawler doesn't fulfil client data needs, on the grounds that distinctive clients may have diverse data needs on differing endless supply of same inquiry/point to web index. So finding diverse client seek objectives gets to be confused analysing client look objective is fundamental to give best result to which the client searches for in the web. Input sessions have been grouped to take in a few clients investigate destinations fora question. Number one, we propose a system to taking care of the every current issue viably. Number two, we propose a novel way to deal with tackling the current issues and create pseudo reports great approach to speak to that records. Last one, we propose another variable "Characterized Average Precision (CAP)" to taking care of the current issues and this strategy predominantly utilization of the execution is extremely viable that is the reason this technique utilizing as a part of this anticipate.

KEYWORDS - User search goals, implicit feedback sessions, pseudo-documents, restructuring search results, k-means clustering, Keyword search.

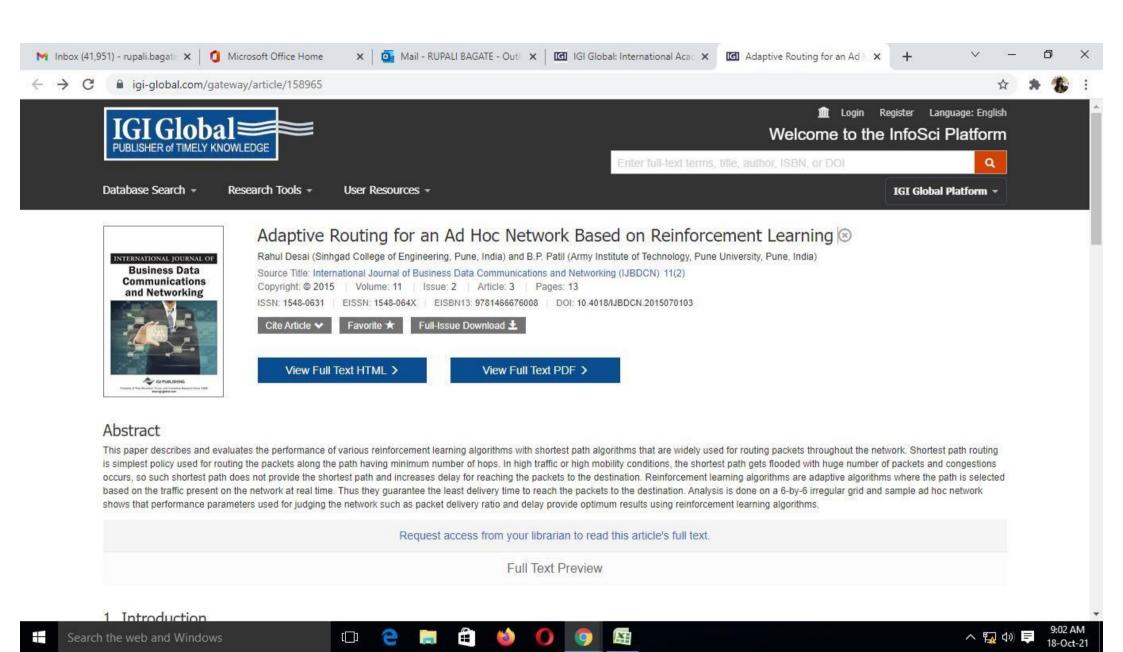
1.INTRODUCTION

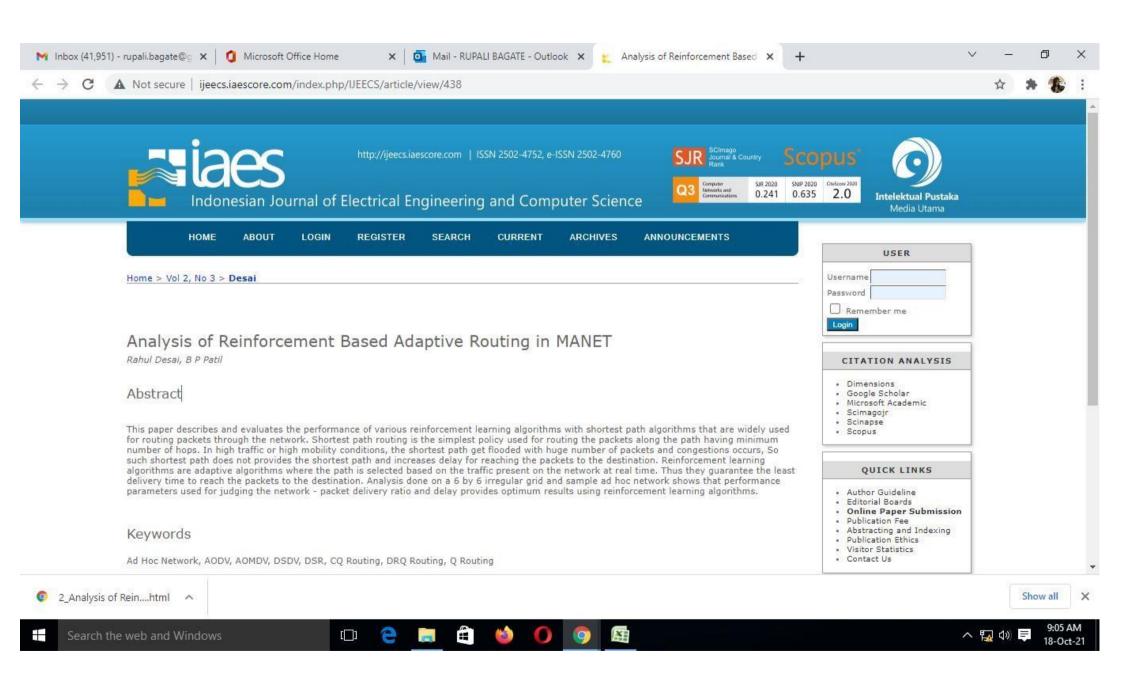
Today web has turned into the biggest wellspring of recovering the data. Be that as it may, because of the meagre condition of the web, it has been hard to locate the applicable data. For instance, when client presents an inquiry "java" to web search tool, a few clients are intrigued to know data

about programming dialect and a few clients need to know data about island of Indonesia. In this way it is important to catch the distinctive quest objectives for apropos data recovery. In this paper, we present the criticism session developed from the client navigate logs, then guide the input sessions

www.ijcsc.co.in Page 16

Page 14

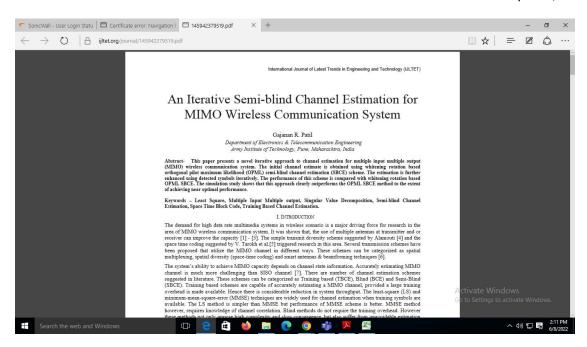




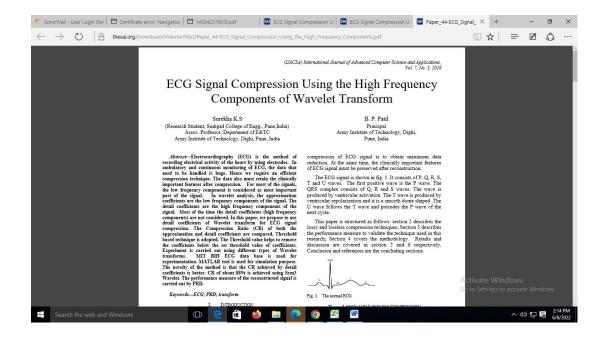
3.3.2 - List of UGC Care Journals

I. AY 2015-16

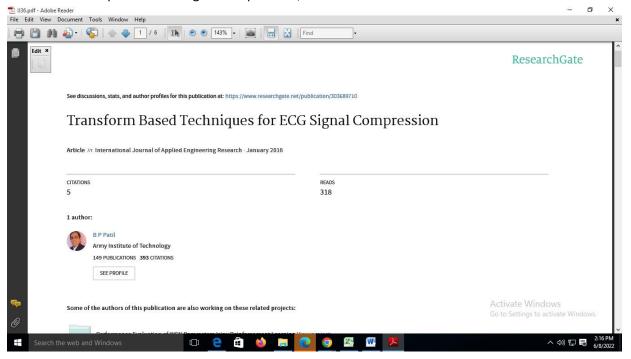
1. An Iterative Semi-blind Channel Estimation for MIMO Wireless communication System, Dr G R Patil



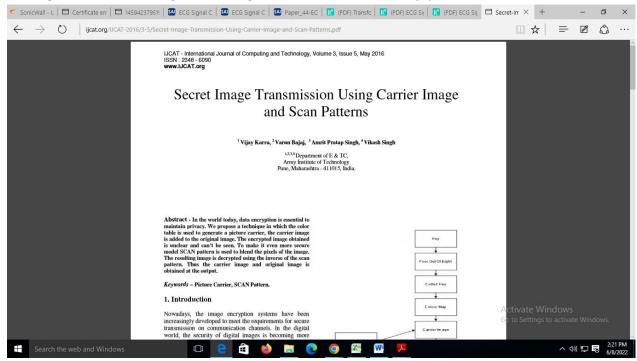
ECG Signal Compression Using the High Frequency Components of Wavelet Transform, Dr Surekha K S

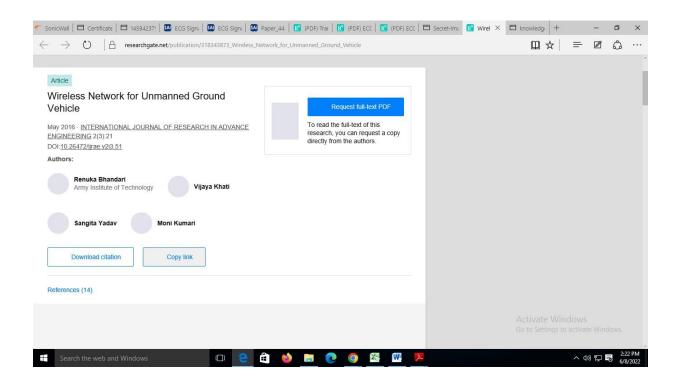


Transform Based Techniques for ECG Signal compression, Dr Surekha K S

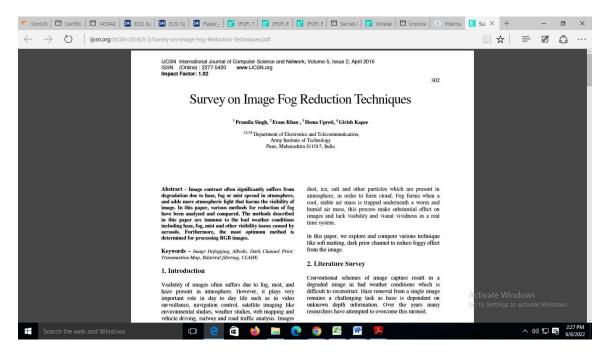


Secret Image Transmission using Carrier Image & Scan Patterns, Mr. K. Vijaykumar





Survey on Image Fog Reduction Techniques, Mr Girish Kapase



Technical Survey of Topics and Trends - Internet of Things, Mr Girish Kapse



