© 2017 IJSRST | Volume 3 | Issue 7 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Android App with PaaS Solution for Web Mining of E-Mail Addresses and its Triggering Mail

Neha Gupta[1], Ankur Varshney[2]
nehaaagupta1991@gmail.com[1], ankur.varshan@gmail.com
ITM, Aligarh, Karsua Uttar Pradesh, India

ABSTRACT

Web is a collection of inter-related files on one or more web servers and the dimension of World Wide Web (The Internet) is in billions in terms of web pages and increasing rapidly. The web data includes web pages, web links, objects on the web and web logs. Web mining is one of the data mining domains where data mining techniques are used for extracting information from the web servers. Web mining is used to understand the customer behavior, evaluate a particular website based on the information which is stored in web log files. Web mining is evaluated by using data mining techniques, namely classification, clustering, and association rules. It has some beneficial areas or applications such as Electronic commerce, E-learning, E-government, E-policies, E-democracy, Electronic business, security, crime investigation and digital library. Data mining techniques and applications are very much needed in the cloud computing paradigm. The implementation of data mining techniques through Web mining under the cloud computing framework will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage. We can say, cloud mining is one of the most recent concept that is likely to grab attention in the field of advance technology. Since, due to the diversity of web pages available on the web, the high degree relevant information retrieval becomes a major issue. Retrieving the required web page from the web efficiently and effectively becomes a challenging task because web is made up of unstructured data, which delivers the large amount of information and increase the complexity of dealing information from different web service providers. Extracting the relevant or the required set of data from such a large number of scattered information available on the web makes the computation complex and time consuming and furthermore makes it prone to risk of retrieval failure of the required dataset. The present work proposes to develop an application that could help users not only mine data conveniently but also store it effectively in a data bank for a specific purpose or future usage and also to make use of mobile cloud computing, the latest in the field.
Keywords: World Wide Web; Web Mining; Cloud Computing; Data Mining

I. INTRODUCTION

1.1 CLOUD COMPUTING

The term **"**cloud**"**, as used in this perspective, appears to have its origins in network diagrams that represented the internet, or various parts of it, as schematic clouds. **"**Cloud computing**"** was coined for what happens when applications and services are moved into the internet **"**cloud.**"**

Cloud computing is not something that suddenly appeared overnight; in some form it may trace back to a time when computer systems remotely time-shared computing resources and applications. More currently though, cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications.

Many companies are delivering services from the cloud. Some notable examples as of 2010 include the following:

• **Google** — Has a private cloud that it uses for delivering many different services to its users, including email access, document applications, text translations, maps, web analytics, and much more.

• **Microsoft** — Has Microsoft® Sharepoint® online service that allows for content and business intelligence tools to be moved into the cloud, and Microsoft currently makes its office applications available in a cloud.

• **Salesforce.com** — Runs its application set for its customers in a cloud, and its Force.com and Vmforce.com products provide developers with platforms to build customized cloud services.

1.1.1 Characteristics

Cloud computing has a variety of characteristics, with the main ones being:

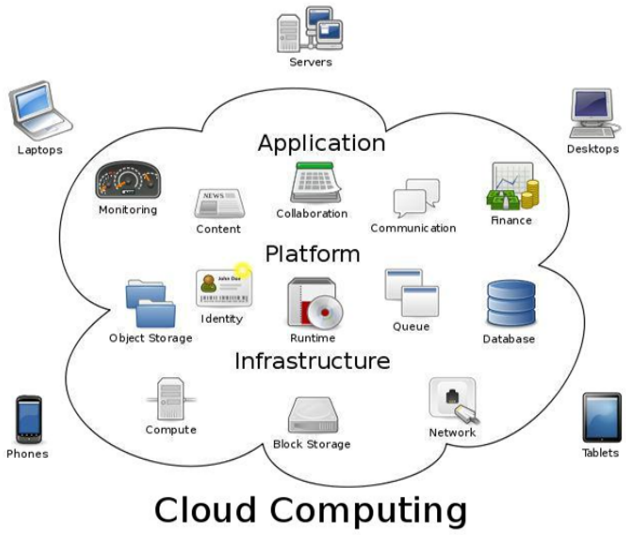
a) Shared Infrastructure — Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.

b) Dynamic Provisioning — Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.

c) Network Access — Needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smart phones.

d) Managed Metering — Uses metering for managing and optimizing the service and to provide reporting and billing information. In this way, consumers are billed for services according to how much they have actually used during the billing period.

In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.



1.1.2 Advantages

The following are some of the possible benefits for those who offer cloud computing-based services and applications:

a) Cost Savings — Companies can reduce their capital expenditures and use operational expenditures for increasing their computing capabilities. This is a lower barrier to entry and also requires fewer in-house IT resources to provide system support.

b) Scalability/Flexibility — Companies can start with a small deployment and grow to a large deployment fairly rapidly, and then scale back if necessary. Also, the flexibility of cloud computing allows companies to use extra resources at peak times, enabling them to satisfy consumer demands.

c) Reliability — Services using multiple redundant sites can support business continuity and disaster recovery.

d) Maintenance — Cloud service providers do the system maintenance, and access is through APIs that do not require application installations onto PCs, thus further reducing maintenance requirements.

e) Mobile Accessible — Mobile workers have increased productivity due to systems accessible in an infrastructure available from anywhere.

1.2 Web Mining

Internet has become an integral part of our lives. Today, we use web in virtually all the spheres of our daily routine activities like Searching, Communicating, Playing, Calculating etc. The biggest advantage of internet is easy & swift access to the information in the very moment that we desire. Easy, Accurate and On-demand information within split second is one of the powerful discoveries ever made by mankind which empowers everyone of us to remain connected with everyone and every event, anytime.

Searching of updated information requires an automatic procedure that intelligently works robotically to fetch, collect and update latest information of the events happening around the world. This automatic procedure is called crawling or mining. A **crawler** is a program which owes the task of dipping into the large database of information and collecting the updated information about the events and tasks that happen every second, globally. The World Wide Web or WWW is a global, large repository of information like text, e-mail addresses, documents, images, multimedia, movies, high definition data and much other information, referred to as information resources. A large amount of new information is posted on the WWW every second, every day.

Crawler is an essential part of search engines or more specifically, searching activities which is used by all of us to find any information on the web. These programs persistently work in the back-end in order to collect updated data from World Wide Web and store them in the large repositories or database which is used by search engines like Google, Bing, MSN, etc. in order to fulfill the search queries of global users.

We know that web is a mesh of interconnected or correlated information. In other words, every web page is usually connected to some other web page which facilitates the user to search not only for the information that he/she requires but also, more related or other interesting information.

Nowadays, corporate and government exploit the capabilities of internet through collecting specific or general, latest or archived information on the web related to preferences, activities or other personal / official data of anybody or anything. Such vital magnitude of data could be aimed towards performing a background check of a person or investigation – oriented or it could even be related to marketing and publicity of a brand or a product. We propose to develop an application that could help users not only mine data conveniently but also store it effectively in a data bank for a specific purpose or future usage. For ex: If a company has to run a bulk e-mail marketing campaign, it would prefer to possess a data bank of latest e-mail addresses which are posted or updated on the internet in order to target those chunk of consumers who are the latest rather than mailing dummy or inactive mail addresses from an archived data bank.

As WWW has plethora of information available, in order to fulfill our objective, we wish to focus ourselves on mining some specific set of data, like, e-mail addresses which are of worth to generally every corporate or government establishment for the purpose of their own marketing or social awareness. At the same time, we also propose to store all the crawled or mined e-mail addresses in order to build a web databank that helps the users afterwards.

1.3 Android Platform

Android is a software stack for mobile devices that includes an operating system, middleware and key applications. The Android SDK provides the tools and APIs necessary to begin developing applications on the Android platform using the Java programming language.

1.3.1 Features

1. **Application framework** enabling reuse and replacement of components
2. **Dalvik virtual machine** optimized for mobile devices
3. **Integrated browser** based on the open source WebKit engine
4. **Optimized graphics** powered by a custom 2D graphics library; 3D graphics based on the OpenGL ES 1.0 specification (hardware acceleration optional)
5. **SQLite** for structured data storage

- **Media support** for common audio, video, and still image formats (MPEG4, H.264, MP3, AAC, AMR, JPG, PNG, GIF)
- **GSM Telephony** (hardware dependent)

- **Bluetooth, EDGE, 3G, and WiFi** (hardware dependent)
- **Camera, GPS, compass, and accelerometer** (hardware dependent)
- **Rich development environment** including a device emulator, tools for debugging, memory and performance profiling, and a plugin for the Eclipse IDE.

II. Literature Survey

With the increasing popularity of Cloud computing, researchers studied the performance of Clouds for different types of applications such as scientific computing, e-commerce and web applications.

Web mining has three classifications namely, web content mining, web structure mining and web usage mining. Each classification is having its own algorithms and tools. Web content mining is nothing but the discovery of valuable information from web documents and these web documents may contain text, image, hyperlinks, metadata and structured records. It is used to look at the information by search engine or web spiders i.e. Google, Yahoo. It is the process of retrieving the useful information from the web content or web documents. Web structure mining is also process of discovering structured information from the websites. The structure of a graph consists of web pages and hyperlinks where the web pages are considered as nodes and the hyperlinks are edges and these are connecting between related pages. Web usage mining is also called as web log mining.[19] It reflects the user's behavior which can catch the meaningful patterns from one or more web localities

Web usage mining is also called as web log mining which is used to analyze the behavior of online users [12]. It fed into two types of tracking; one is general access tracking and another one is customize usage tracking [9]. The general access tracking is used to predict the customer behavior on the web and it identifies the user while the user interacts with the web. It can store the data automatically when the web server log and application log [15]. The web log is located in three different locations they are web server log, web proxy server and client browser and it contains only plain text file (.txt). The large amounts of irrelevant data are available in the web log file because it contains noisy data, large amount of incomplete, eroded and unnecessary information[17].Web server log files are used to identify the errors and failed requests were given by the web master and the system administrator. Web usage mining is to extract the data which are stored in server access logs, referrer logs, agent logs and error logs. Web usage mining generally uses basic data mining algorithms such as association rule mining, sequential rule mining, clustering, and classification. It has several tools to analyze the behavior of the user.

2.2 Background preliminaries

According to a research paper on data mining in cloud computing, [1] Data mining technologies provided through Cloud computing is an absolutely necessary characteristic for today's businesses to make proactive, knowledge driven decisions, as it helps them have future trends and behaviors' predicted. This paper provides an overview of the necessity and utility of data mining in cloud computing. As the need for data mining tools is growing every day, the ability of integrating them in cloud computing becomes more and more stringent.

Whereas another describes a cloud-based infrastructure designed for data mining large distributed data sets over clusters connected with high performance wide area networks. [2]Sector/Sphere is open source and available through Source Forge. It is used as a basis for several distributed data mining applications.

The infrastructure consists of the Sector storage cloud and the Sphere compute cloud. We have described the design of Sector and Sphere and showed through experimental studies that Sector/Sphere can process large datasets that are distributed across the continental U.S. with a performance penalty of approximately 80% compared to the time required if all the data were located on a single rack. [2] Sector/Sphere utilizes a specialized networking layer to achieve this performance.

It also describes a Sector/Sphere application to detect emergent behavior in network track and showed that for this application Sector/Sphere can compute clusters on over 300,000 distributed.[2] Finally, they performed experimental studies on a wide area test bed and demonstrated that Sector/Sphere is approximately 2.4 {2.6 times faster than Hadoop using the Tera sort

benchmark supplied with Hadoop. Using a benchmark we developed call Tera split that computes a single split in a classification and regression tree, we found that Sector/Sphere was about 1.6 -1.9 times faster than Hadoop.

Cloud computing infrastructures can be effectively used to run data intensive applications. [3]To help users in this task, simple but high-level environments must be provided. This paper presented the Data Mining Cloud App framework that has been designed to support the efficient execution of parameter sweeping data mining applications in a Cloud. The framework has been implemented using the Windows Azure platform and evaluated through a set of parameter sweeping clustering and classification applications [3]. The user interface is very simple and hides the complexity of the Cloud infrastructure used to run applications. The experimental results discussed in the paper demonstrates the effectiveness of the proposed framework, as well as the scalability that can be achieved through the parallel execution of parameter sweeping applications on a pool of virtual servers.

Other than supporting users in designing and running parameter sweeping data mining applications we intend to exploit Cloud computing platforms for running service oriented knowledge discovery processes designed as a combination of several data analysis steps to be run in parallel on Cloud computing elements.[7] To achieve this goal, we are currently extending the framework for supporting also workflow-based KDD applications, in which complex data analysis applications are specified as graphs that link together data sources, data mining algorithms, and visualization tools.
[4]We provide a survey about the research in the area of Web mining‘s today structure and tomorrow view. It points towards some confusion between data mining and web mining. Web data is growing at a significant rate. Web Mining is fertile area of research. Many Successful applications exist.

It also suggests the subtask of web mining & future of web mining. Now we also work for the process mining and try to combine usage mining with structure mining. We also go for the mining from cloud. Whenever we work on mining over cloud computing that time we hesitate for the cost but that come very less by cloud mining. So, we can say that cloud mining can be seen as future of web mining.

Paper suggests that Cloud computing is an architecture which is known for its powerful capability of computation and storage and resource sharing.[5] These features make cloud computing favorable to data mining service in network environment. It discusses association rule mining in cloud environment and various parallel and distributed mining algorithms.

The implementation of association rule in the distributed systems can be efficiently done on Hadoop. Further, the data transfer among the nodes and the situations like node failure etc are taken care of by Hadoop. This adds robustness and scalability to the system. [5]

Various parameters affect the performance of algorithms such as time taken to generate frequent item sets, inter site communication cost, number of scans through the database etc. [3]The integration of association rule mining and cloud computing in this paper is at the initial stage of our research on data mining service in cloud environment and requires further improvement.

A distributed architecture was proposed to eliminate this threat. But overheads were still prevalent in the system. Hence cache memory concept was implemented in our system by generating frequent item sets using any data mining tool.

Android is an operating system (OS) developed by the Open Handset Alliance (OHA). The Alliance is a coalition of more than 50 mobile technology companies ranging from handset manufactures and service providers to semiconductor manufacturers and software developers, including Acer, ARM, Google, eBay, HTC, Intel, LG Electronics, Qualcomm, Sprint and T-Mobile. The stated goal of the OHA is to "accelerate innovation in mobile and offer consumers a richer, less expensive and better mobile experience". The java platform and the SDK tools were available in October 2008. There is single mobile phone that runs the Android OS which was G1 from T Mobile. According to the Android website the platform is based into the four core features.

2.3 Summary of literature review

After a thorough study of all the related paperwork, several bottlenecks and unresolved issues are found under the related topic. There are several issues regarding web-mining and bigger problem arises when we talk about combining two different aspects of web, cloud computing and web mining. To yield better scope in future, we need to resolve the issues related to these two separately first.

In the future instead of having a separate cache for every provider we can have a single cache for all the providers which will store the frequently accessed client data therefore enhancing the efficiency of the current system. Therefore, after considering all the above research papers, we concluded that an application which can effectively mine resourceful data from the web that could be further utilized for the benefit of users is required. The application should not only be efficient in terms of its execution and turnaround times but also be less burdensome on the essential operational metrics and environment.

Now is an exciting time for mobile developers. Android also offers an equal alternative. Android is an open source architecture that includes the operating system, middleware and its key applications along with a set of API libraries for writing mobile applications that can shape the look, feel, and function of mobile handsets. Mobile developers can now expand into the Android platform to enhance existing products. Without any artificial barriers, Android developers write applications that take full advantage of increasingly powerful mobile hardware. Mobile applications are a rapidly growing segment of the global mobile market. In this paper, we discuss on Android mobile platform for the mobile application development, layered approach for android.

Google released Android which is an open source mobile phone operating system which is Linux based. Android becomes the most widely used OS on mobile phones. Android is mobile operating systems designed for increasingly powerful mobile hardware. Windows Mobile and Apple's iPhone provide simplified development environment for mobile applications. Android is built on proprietary operating systems that often prioritize applications those are created by third parties and restrict communication among applications and native phone data .Android offers possibilities for mobile applications by offering an open development environment built on an open source Linux kernel. Hardware access is available through a series of API libraries, and application interaction.

Android Mobile Application Development is based on Java language codes. It allows developers to write codes in the Java language. These codes can control mobile devices via Google- enabled Java libraries. It provides the platform to develop mobile applications using the software stack provided in the Google Android SDK. Android mobile OS provides a flexible environment for Android Mobile Application Development as the developers can not only make use of Android Java Libraries but it is also possible to use Java IDEs. The software developer in Mobile Development has expertise in developing applications based on Android Java Libraries and other important tools. Android Mobile Application Development can be used to create innovative applications. Mobile

Development has worked extensively on projects gaming software, organizers, media players, picture editors devices and more.

Nowadays, both hardware and software of mobile devices get greater improvements than before, some smart-phones such as iPhone, Android serials, window mobile phones and blackberry, are no longer just traditional mobile phones with conversation, SMS, Email and website browser, but are daily necessities to user. However at any given cost and level of technology, considerations such as weight, size, battery life, ergonomics and heat dissipation exact a severe penalty in computational resources such as processor speed, memory size, and disk capacity. Therefore three approaches have been proposed for mobile cloud applications:

1) Extending the access to cloud services to mobile devices. In this approach users use mobile devices often through web browsers, to access software/applications as services offered by cloud. The mobile cloud is most often viewed as a Software-as-a-service (SaaS) cloud and all the computation and data handling are usually performed in the cloud.

2) Enabling mobile devices to work collaboratively as cloud resource providers. This approach makes use of the resource at individual mobile devices to provide a virtual mobile cloud, which is useful in an ad hoc networking environment without the use of internet cloud.

3) Augmenting the execution of mobile applications on portable devices using cloud resources.

This approach uses the cloud storage and processing for applications running on mobile devices. The mobile cloud is considered as an Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS) cloud. In this partial offloading of computation and data storage is done to cloud from the mobile devices.

As mobile devices have become our primary data processing devices nowadays, mobile cloud computing has emerged as a great extension to cloud computing field. There were certain issues regarding the execution and security of database. There have been paper published before analyzing the in-depth survey of research work done in mobile cloud computing. Open issues have also been covered, with some primary issues being discussed along with the research done around them.

Section III [11] contains the detailed survey around the key categories of mobile cloud computing which points out at some of the approaches focusing on collaborative working of mobile devices, migrating the execution from mobile devices to resource rich platforms and partitioning of applications for offloading them to the cloud.

Thus it concludes that for a number of applications local resources are not sufficient to execute on mobile device and also that the available local resources of a group of devices residing in the same areacan be used to form a virtual cloud to overcome the resource constraints of our mobile devices. This way, need of internet availability can also be suppressed. In computation intensive applications, sometimes the local resources cannot provide enough support to deliver the required quality of service. Such applications can be migrated to be executed on cloud.

Application partitioning approaches can also be used to augment the execution of certain mobile applications on cloud resources. While some of the discussed approaches might seem quite complicated, these fields offer some promising scope of research for future.

III. Proposed Work

In the proposed work, the idea is to launch an application on an Android device where a user will input any URL, (For example: http://www.yahoo.com or http://www.gmail.com, or http://www.keepvid.com, etc.). Then our application will visit that URL address via HTTP (3G mobile internet) and crawl (collect) all the e-mail addresses hosted on it. The URL which had been inputted may contain images, graphics, other text, etc. but this application shall collect e-mail addresses only.

In this work, we have proposed to implement the technical logic of performing crawling though parsing the index HTML page of the inputted parent URL using regular expressions. The crawling process makes use of the third type of web mining known as web usage mining. All the collected e-mail IDs shall be henceforth stored on a web database in real-time for which I had propose to make use of a free cloud service provider. Thereafter, the application will enable the user to trigger an automated e-mail in one-go to collectively all the addresses stored on the web/cloud database from any specific sender e-mail address.

The purpose of triggering this e-mail message to mass recipients could be related to corporate etc who can utilize this for marketing of their company, brand or product promotion or spreading a social awareness message and this well defines the scope of this work.

One can consider this as a malware, but it's actually upon the user to use any software application in a constructive way or destructive way. Thus, although it is more of a malware but it can also be used in business purposes to broadcast messages with just one click on a simple android oriented smart phone.

3.1 Proposed System Design Approach

3.1.1 Architectural Design

Under architectural design, after defining the whole system into a set of objectives & further subdividing them into functions, we defined the basic dependency & communication between them. This means that all the prime functions, their required inputs, expected output/behavior interdependency between other functions were clearly defined. The corresponding interfaces for the user for each function were designed to ensure user-friendliness. We actually addressed the system-level problems here and made a conscious effort to build a robust design which can result in an effective communication within itself and with the system in terms of raw data or processed information. All the primary database design for data storage was also done in this phase.

3.1.2 Detailed Design

In this phase, we further subdivided every function into a set of modules & defined required inputs & expected behavior for each of them. All the minute correlations, interdependencies, communication between the modules were clearly defined. The source, usage & processing of data for every module was carefully done. The database design was also normalized at this stage to ensure that the data is efficiently stored & retrieved.

3.2 Design and Implementation

Cloud computing is broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world. In this section, we compare and contrast our work with previous research work for evaluating and comparing the performance of different Cloud services.

With the increasing popularity of Cloud computing, many researchers studied the performance of Clouds for different types of applications such as scientific computing, e-commerce and web applications. For instance, some have analyzed the performance of many-task applications on Clouds. Similarly, many performance monitoring and analysis tools are also proposed in the literature. Our work complements these previous works by utilizing these tools and data to rank and measure the Quality of System of various Cloud services according to users' applications. Other works such have proposed frameworks to compare the performance of different Cloud services such as Amazon EC2, Windows Azure, Rackspace, etc. These works again focused on comparing the low level performance of Cloud services such as CPU and network throughput.

- There would be 2 tiers of the project namely Android & Cloud tier.
- Both the tiers would be connected wirelessly using HTTP, independent of their geographical location. The application's database is stored on cloud.
- Using the android application on the phone user can enter the URL of desired website and can crawl the e-mail ids present on that webpage.
- The crawled email addresses are stored on the users' account on cloud.
- An email from the users' id will be send to the crawled email addresses, or in other words, a message via mail shall be broadcast to those email ids.
- This whole process is called triggering of mail in real-time cloud computing system and application is known as 'spambot'.

IV. RESULT AND ANALYSIS

4.1 Application - Spambot

A bot is an abbreviated form for robot. By Bots, or Internet robots one can conclude that they are software applications build to perform a repetitive task. They are also known as spiders, crawlers, and web bots. While they may be utilized to perform repetitive jobs, such as indexing a search engine, they often come in the form of malware which are used to gain total control over a computer. Now the question comes if it's a malware how can it be used in a constructive way. Therefore, it's necessary to find the advantages of a bot.

One of the typical "good" bot uses is to gather information. Bots in such guises are called web crawlers. Another "good" use is automatic interaction with instant messaging, instant relay chat, or assorted other web interfaces. Dynamic interaction with websites is yet another way bots are used for positive purposes.

So after talking about a general idea of bot, it becomes easy to define a spambot. A spambot is a computer application designed to send spam emails automatically in large quantities. It automatically collects email addresses from various sources on the Internet. Using the large number of email addresses collected, a spambot creates mailing lists and sends junk mail, also known as spam. Spambots may be used by hackers, also known as spammers, to carry out certain attacks on a website or servers. Spambots create fake accounts and send unsolicited messages through them for purposes which could include advertising, hacking or even fraudulent businesses. Many websites and hosts try to protect their websites from spam by using anti-spam programs.

Spambots are automated computer applications that are capable of sending huge numbers of emails to a large mailing list. Spammers use these spambots for various purposes. A spam mail may contain a virus attachment and spread malware. It may also be used to post demeaning content onto somebody's inbox or as a method of advertisement.

Spambots can be classified broadly into the following types:

Email spambots — These are the most common spambots. They are easy to code and are able to identify email addresses based on their format. These spambots are web crawlers that can collect email addresses from various sources such as newsgroups, forums, websites, chat rooms and Web posts. There are many techniques used to counter the actions of these spambots. One of the most popular techniques involves altering the email addresses in a way spambots are not able to recognize, such as "me [at] mydomain [dot] com."

Forum spambots — these spambots look into guest books, wikis, blogs, forums and other types of Web forms to submit fake details. The CAPTCHA technique used in many websites is helpful in thwarting these spambots.

Spam messages sent by these spambots may also include hyperlinks posted in websites to boost a site's search engine ranking.

Twitter spambots — Also known as twitter bots, these bots are used to make automated posts on the social media site Twitter. It can be used to re-tweet a tweet repeatedly and post replies to posts.

Spambot Application Overview:

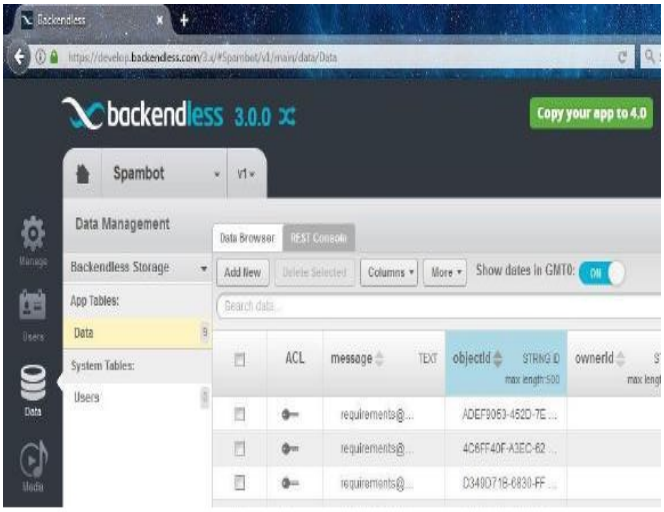
The resulted application is the final output of this project. It is an android application which is supported by latest android version (6.2 and above) and it not only involves the android environment to run, it also contains a cloud tier-backendless is used for this purpose. The account is thus created on cloud and all the data is stored on this account.

To download this app, the APK file of the app is copied from the PC to the smart-phone used.

This page appears as the application initiates. The cursor is placed in the white box and the user needs to enter URL of the website he or she wish to search the email addresses on. It should be taken into account that the URLs should be of such web pages which have the maximum possibility of having lot of email-ids. Here is one significant thing to note that the crawler shall search only the email addresses written in text form and not in the pictures or jpeg format.

This is because the search takes place using web-mining as a concept and when w search the internet using the concept of web-mining, we can only input a certain pattern and it shall search the similar information in accordance to the pattern input and match it before searching. Hence, it won't be able to fetch addresses if they are given in any other format. As we enter the correct URL and press submit, loading starts and fetching of email-ids take place. Sometimes when there is no email id on the 'searched' page, then 'no email addresses available is displayed'. And when the URL entered have set of email-ids present than it displays the list of email addresses crawled. Now if the user clicks on the button, a test mail or whatever mail one wish to send shall be sent to all of the email addresses fetched in just one click. The body of the mail is customizable. But for this project the body of the mail contains "Test Mail" as the message. The following shows the sent mail from the email id of the user. Out of all the time the application have been used and tested, a few times the email-id on which test mails have been sent, have replied back too. Thus, this accounts for the success of the application. If you don't want to send mails and just wish to store these email addresses for future use you can skip clicking the button and email address would be stored in the users' cloud account simultaneously. The cloud account on 'backendless' platform is shown below. The application record is maintained on this account and it keeps getting updated whenever the user uses the application on his/her android smart-phone. The backendless account can also be opened on a mobile phone and as well on the PC.

A major role in this app is of cloud computing technology. All the data of the application is being stored on a cloud account. To get a cloud account, after lot of search one which provided free cloud usage for a limited period of time was a provider calld as "backendless". Backendless is a universal server-side used by mobil and desktop developers to build applications. It provides an API for developing and enhancing the applications.



One can easily integrate user registration and login into their web application or mobile application using backendless. It provides user registration and authentication APIs in all SDKs (Android, iOS, plain JS, Angular, Typescript, .NET and REST). Registered users can be managed using an intuitive interface which supports:

query-based search;
changing user password;
disabling user accounts;
modifying any of the user properties or creating new properties;
Setting user's permission.

V. CONCLUSION

Cloud computing represents both the software and the hardware delivered as services over the Internet. Cloud Computing is a new concept that defines the use of computing as a utility, that has recently attracted significant attention. Data mining is defined as a Type of database analysis that attempts to discover useful patterns or relationships in a group of data. Web mining is extended version of data mining. However, Web mining does not only mean applying data mining techniques to the data stored in the Web. The algorithms have to be modified such that they better suit the demands of the Web. Web data is growing at a significant rate. Web Mining is fertile area of research and many Successful applications exist. Our proposed work makes use of techniques like crawling, web usage mining, etc to extract out the email addresses of the users hosted on a particular URL using Parsing on

mobile cloud computing framework. It's a successful attempt to make the work cost efficient by using cloud storage and also to make it user friendly as it's in the form of a simple android application.

VI. REFERENCES

[1]. Ruxandra-Ştefania PETRE, "Data mining in Cloud Computing' Bucharest Academy of Economic Studies, Database Systems Journal vol. III, no. 3/2012

[2]. Yunhong Gu, Robert Grossman, "Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere' University of Illinois at Chicago and Open Data Group

[3]. Fabrizio Marozzo, "A Cloud Framework for Parameter Sweeping Data Mining Applications' DEIS, University of Calabria, Rende (CS), Italy and Domenico Talia, ICAR-CNR, DEIS, University of Calabria Rende (CS), Italy and Paolo Trunfio, DEIS, University of Calabria, Rende (CS), Italy

[4]. Kaikala Anjani Sravanthi and Yalamarthi Madhavi Lata, "Web Mining Using Cloud Computing' International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 3, Issue 4, April 2013

[5]. Zeba Qureshi, Jaya Bansal and Sanjay Bansal, "A Survey on Association Rule Mining in Cloud Computing' International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 3, Issue 4, April 2013

[6]. Srishti Sharma, Harshita Mehta, "Improving Cloud Security Using Data Mining' IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727, Volume 16, Issue 1, Ver. II (Jan. 2014)

[7]. Dr.S. Vijiyarani and Ms. E. Suganya 'RESEARCH ISSUES IN WEB MINING' International Journal of Computer-Aided Technologies (IJCAx) Vol.2, No.3, July 2015

[8]. C Mayank 'Implementing Web Mining into Cloud Computing' International Journal of Scientific & Engineering Research, Volume 5, Issue 3, March-2014 ISSN 2229-5518

[9]. Saurabh Kumar Garg, Steve Versteeg, Rajkumar Buyya'A framework for ranking of cloud computing services' Future Generation Computer Systems 29 (2013) 1012–1023

[10]. HaiLong Li ZhenQi Wang, "Research of massive Web log data mining based on cloud computing' 2013 International Conference on Computational and Information Sciences

[11]. Kaikala Anjani Sravanthi, Yalamarthi Madhavi Lata, "Web Mining Using Cloud Computing' International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013)

[12]. Michael Jennings, "What are the major comparisons or differences between Web mining and data mining?' Information Management Online, June 25, 2002.

[13]. Chen, M. S, Han, J. and Yu, P. S. "Data Mining: An overview from a database perspective', IEEE transaction on knowledge and data engineering, Vol. 08, No. 6, pp: 866-883, 1996.

[14]. Karan Bhalla & Deepak Prasad,' Data Preparation and Pattern Discovery For Web Usage Mining'

[15]. Amit Pratap Singh1, Dr. R. C. Jain 2,' A Survey on Different Phases of Web Usage Mining for Anomaly User Behavior Investigation' International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)Volume 3, Issue 3, May – June 2014

[16]. M. G. da Gomes Jr. and Z.Gong, "Web Structure Mining: An Introduction', Proceedings of the IEEE International Conference on Information Acquisition, 2005.

[17]. Dalibor Fiala, "Web Mining and Its Applications to Researchers Support', Technical Report No. DCSE/TR-2005-06, April 2005

[18]. Wang jicheng, Huang Yuan,Wu Gangshan, Zhang Fuyan, "Web mining: Knowledge discovery on the Web Systems", Man and Cybernetics 1999 IEEE SMC 99 conference Proceedings. 1999 IEEE International conference

[19]. C.Gomathi, M. Moorthi,' Web Access Pattern Algorithms in Education Domain' Computer and information science journal vol. 1, No.4, November 2008

[20]. Android Wireless Application Development, S. Conder and L. Darcey, Addison-Wesley (2010).