# Google Newsm- Java Rest Service Based Webcrawling

[1]R.Senthil Kumar,[2]Dr.S.Saravanakumar,[3]Prabhu J,[4]Delin David D,[5]Shahul Hameed M
[1]Associate Professor,[2]Professor,[3][4][5]Final Year IT Student
Department of Information Technology,Panimalar Institute of Technology
[1]dr.senthil2002@gmail.com,
[3]prabhubalaji94@gmail.com,[4]delin2david@gmail.com,[5]hameedshahul171@gmail.com          .

*Abstract* — **The heightened demand in the usage of android mobiles has paid a way for the evolution of multifold android applications that will be in use in our routine activities. In this android era, the application that will be used in our daily life is the Google map application. Google map is widely used in several fields such as object positioning, traffic monitoring and automated vehicle transportation. Large groups of people are not spending their time in reading day to day news, in order to make them as a social participant, we introduce application that will plot news on the Google map and provide a descriptive module for each news with respect to their place of occurrences. To empower college students and share holders, the proposed android application will crawl information from the websites regarding symposium, conferences, job interviews and share market status equip them with vital information and mark on the Google map, which when elected provide the compiled information as a dialog prompt flourished with requisite information.**

*Keywords:* **Java, Android, Spring Tool Suite, Maven Dependencies**

## I. INTRODUCTION

The habitude of android applications and its impact on society has been tremendously increasing. To envision the proposed services, a Web Service named Java Rest Service will be designed according to each module prescription (i.e.) whether it is for College events (or) Job interviews (or) Daily News (or) Stock market info. Since the JRS that is devised has the ability to run on different platform i.e. OS independent this web service will be conformable to all types of accessing through desktop, tablet or Mobile phones (for all OS).In consideration of operation on Daily news it will not be available in a predefined format so a peculiar process named Natural Language Processing (NLP) Algorithm will be used to operate on the news site and discriminate the news with respect to their place of occurrences and compile details related to characters involved in the incidents.

*Server side:*

The Java Rest Web Service will be designed using IDE named Spring Tool suite, which is indispensably built to foothold the development of spring projects. The Spring Tool Suite has a unique feature named

Maven dependencies, which include several jar files URL that will be downloaded from the proposed website when the project gets instantiated. During pioneer execution it will be delayed due to downloading, as on later stages it will be executed instantly. A Framework named spring will be used to flourish our designed modules. The spring is known for its Aspect Oriented Programming concept. Another Framework named Spring Boot is been used to make the execution faster due to its unique capability of having an inbuilt Tomcat Server. The Java Rest Service is a logic that can be executed on any type of platform namely Android (or) IOS (or) Windows...Etc.

*Client side:*

The end user to access this web service must have a suitable communication medium. The proposed project uses Android language to foster the clientside environment. The user interface will have four modules namely Daily news, College event, Job Interviews and share market information. All progress of the client side request will be processedusing Asynchronous task to perform the

functioning in rear.

## I. Project Capsule:

**1.** Existing System

**2.** Proposed System

**3.** Conclusion

**4.** References

## II. Existing System:

The Web application that is developed to determine no of garden in and around United States. At Pioneer stage, information about garden is collected through an online registration Process from people belonging to US and it is maintained in a secure server. When the user request the corresponding webpage to provide details about the gardens, two options will be displayed whether for whole nation (or) for a particular State. If it is for the entire Nation, only the information about no of gardens in each state will be displayed and for State complete description about the garden will be provided including Garden area, location, variety of fruits and vegetables planted.
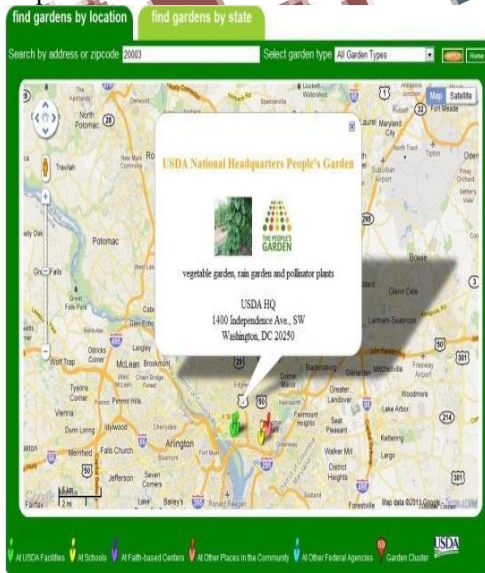


## III. Proposed System:

The contemplated project can be realized into an artifact usable to the society by developing the vital modules at every stage of communication. The architecture diagram for the proposed plan is laid out below with the control flow for the
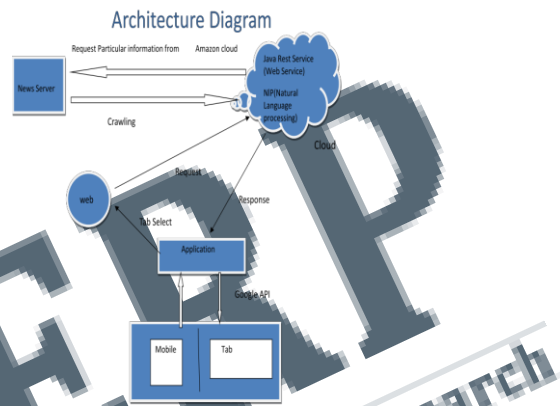


a.   Request initiation b. Server Processing
c.   Map coordinates to Mobile d. Feature Enhancement

**a.   Request initiation:**
The usage of our product gets instantiated by the client request from the Android mobile/tablet. The Product consists of several tabs such as Daily News, College Events, Job Interviews and stock market info. With respect to the corresponding tab selection, a request will be generated to the server.

**b.   Server Processing:**
The Java Rest Service (JRS) residing in ourserver framework wangle the request fromclient and execute corresponding logic to obtain data source from the prescribed URL. After extortion of data source, it will be hoarded into Document object. Based on the structure of data hoarded correspondinglogic will be drawn. For example in grabbing college events information        the        selected        URL        is http://campus.knowafest.com/fest-  departments/IT/  , which provides details about the college symposiums, conferences that are scheduled to take place in INDIA . The webpage is provided with the table name, table

rows and columns. By referring the table class name each and every rows and columns of the table is selected and the corresponding object info is stored in the List with data type College info, which is a bean class with necessary fields such as date, event Name, college Name, dept Name, place, details URL. The JRS will create a object for every events constituted in the page with the associative bean class information. The output of JRS service will be provided in the form of JSON data,

which is advancement over traditional XML. The usage of JSON is elevated because it provides an elegant way of generating documents, more verbose than XML and not follows conventional opening and closing tags. The generated JSON data can also be visualized using Online JSON visualizer, which generates a tree format for every object in the link promoting for easier

   access.

***Server JRS code for College Events***:

```
package com.coll.event; import
java.io.IOException; import
java.text.ParseException;
import java.text.SimpleDateFormat;
import java.util.ArrayList;
import java.util.Date;
import java.util.List;
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element; import
org.jsoup.select.Elements; import
org.springframework.web.bind.annotation.Requ
estMapping;
import
org.springframework.web.bind.annotation.RestC
ontroller;

@RestController
public class FetchCollegeInfoController
  {
  // http://localhost:8080/geteventinfo
     @RequestMapping(value="/geteventinfo")
      public List<CollegeInfo> getCollegeInfo()
    {

List<CollegeInfo> collegeList = new
ArrayList<CollegeInfo>();
// College info is a Bean class with a necessary getter
and setter methods
Document doc = null;
```

```
SimpleDateFormat format = new
SimpleDateFormat("ddMMMyyyy");
 // crawler

 try {
//crawl info from the prescribed site doc
= Jsoup.connect(
"http://campus.knowafest.com/fest-
departments/IT/").get(); Elements table =
doc.getElementsByAttributeValue("class",
"dataTableClass");
Element tableElement = table.get(0);
Elements trElements =
tableElement.getElementsByTag("tr");
for (Element tr : trElements)
{
Elements tdElements =
tr.getElementsByTag("td");
int count = 0;
CollegeInfo collegeInfo = new CollegeInfo();
for (Element td : tdElements)
{
if(count == 0)
{
String dateStr
ing = td.text();
 String splittedDate[] = dateString.split("-");
// split the date according to '-' String dat =
splittedDate[0].substring(0,splittedDate[0].lengt h()-
2);
//in the first string for e.g. "28th "eliminate the
last two characters to get the integer "28"
Date date = null;
try
{ System.out.println(dat+splittedDate[1]+splitted
Date[2]);
date =
format.parse(dat+splittedDate[1]+splittedDate[2
]);
}

catch (ParseException e)
 {
        e.printStackTrace();
}

collegeInfo.setDate(date);
}
else if(count == 1)
{
Element element = td.select("a").first();
String url = element.attr("href");
```

```
String value = element.text();
collegeInfo.setDetailsUrl(url);
collegeInfo.setEventName(value);
}
else if(count == 2)
{
collegeInfo.setType(td.text());
}
else if(count == 3)
{
collegeInfo.setCollegeName(td.text());

}
else if(count == 4)
{
collegeInfo.setPlace(td.text());
} collegeList.add(collegeInfo);
count++;
}

}}

catch (IOException e)

 {
 e.printStackTrace();
 }

     return collegeList;
     }
```

**c. Map coordinates to mobile:** From the JSON data generated for every entity in the webpage, a unique identifier will be used to determine the latitude and longitude of the grabbed news. Before accessing Google map over Android mobile for evaluating the geographic location several permissions must be procured for accessing the Google map, Internet Connectivity….etc. The Service provider must have got temporary (or) Permanent API key from the Google map application.
In case of College events [College name will be declared as an identifier], Job Interviews [Interview Location] and Daily News [Incident Location]

**d. Feature Enhancement:**
In order to increase the accessing speed of information, Server Caching approach will be worned in our design, which increases the performance of the system. For realizing the concept of Server Caching, a special file called Property File will be maintained

in our Server for stockpiling of the current news information. During pioneer access, a delay will be exerted by our service due to server determining the Geographical Location for each and every entity through single requisition. During second entry into our service, at first Property File would be accessed to determine whether there is any previously stored data, In case of absenteeism in the Property File, only for that entity a request will be generated to figure out the location and for the rest data from the Property file will be used to pinpoint on the Google Map.

**Output for College events:**
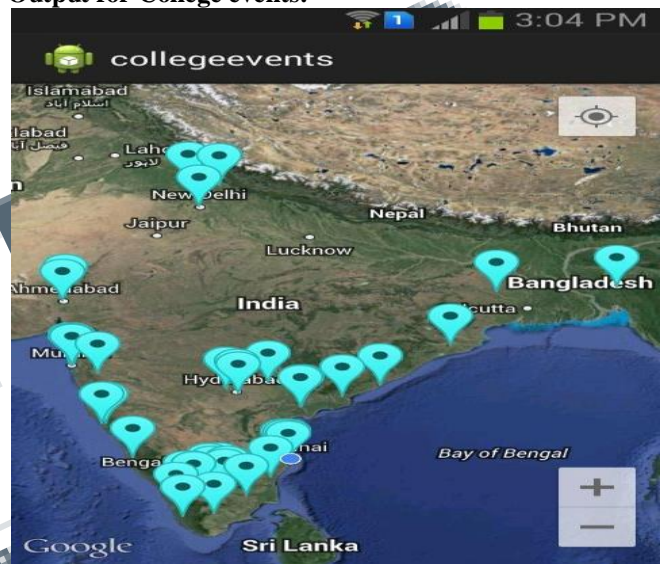


**CONCLUSION:**
This paper has demonstrated the creation of new service that is involved in the crawling of information from the webpage and processing it using Java Rest Service residing on the Server. From the Processed information, a prescribed unique identifier will be used to determine the Geographic Location of the incident and then plot it against the Google Map API in our Android Mobile.

**REFERENCE:**
 1. Online Map Application Development Using Google Maps API, SQL Database and ASP.NET Shunfu Hu , Ting Dai
2. Plotting Surveying Data on Google Earth

D M STILLMAN

3.   Eclipse, 2009. About the Eclipse Foundation.<http://www.eclipse.org/org/>

4.   Google Earth, 2009. Google Earth. http://earth.google.com/tour.html

5.   Peterson, M. P. (2008). International Perspectives on Maps and the Internet: An Introduction, In M. P. Peterson (Ed.), International Perspectives on Maps and the Internet (pp. 3-10), Springer.

6.   www.developersbook.com/spring/spr ing-tutorials/spring-tutorials.php

7.   http://www.journaldev.com/2433/spring -mvc-tutorial-for-beginners-with-spring- tool-suite

8.   Peterson, M. P. (2012). Online Mapping with APIs, Online Maps with APIs and Mapservices (M. P. Peterson, ed.), Springer, pp. 3-12.

● ● ●