Title: FOSTERING OMAN TOURISM USING BIG DATA ANALYTICS BY MINING WEBLOG AND SOCIAL NETWORKING DATA
Proposal ID: Proposal Id not specified
Type of project application: Faculty Mentored Undergraduate Research Award Program (FURAP)

Sector Name
Information Technology and Communication

Team Leader
AL BUSAIDI, SAFA (s26s1533@nct.edu.om)

Faculty Mentor
Gupta, Dr Sanjai (sanjay.gupta@nct.edu.om)

Other Team Members
AL-BUSAIDI, HURIA (s26s16422@nct.edu.om)

AL HINAAI, HANAA (s24s1517@nct.edu.om)

Project Description:
In this research project we will analyze the big volume of social networking data and weblog data using Apache Spark and Apache Hadoop ecosystem. With the social networking data we will do the opinion mining related to Oman tourism, and with weblog data we will try to get different weblog data related to Oman tourism from different related websites. Using these big volumes of data we will do the different analysis which could be Diagnostic, Prescriptive, and Predictive .These different analysis will be used to get the useful information to improve the tourism in Oman. In this r research project priority will be given to practically generated output along with justification and assumption. This project will help greatly in the economy of the country because tourism is an important source in the rise of the economy in the country so it will examine and analyze all the suggestions and criticisms related to tourism in the Sultanate, which was provided by citizens and non-citizens in the Sultanate to determine later solutions appropriate and possible To develop tourism in Oman and attract a large number of citizens and non-citizens to tourism in Oman.
Many of the citizens currently go for tourism outside the Sultanate, despite the presence of many tourist places in the Sultanate so we decided to do a comprehensive study of the reasons that invite citizens to travel outside the country.
In this research project we will generate a report analysis about Oman tourism using Social networking data and weblog data from different Oman tourism websites. Actually we will get the people opinion from the Tweets and feedback given to different websites of Oman Tourism that will explain the people opinion in real time about tourism. The reason we chose social networking data because the social media is gaining popularity for the resident reviews. The Analysis would be certainly reflecting the betterment of the Oman tourism. And show the rating of how much people

said Oman tourism is good and how much of them said bad or not good.
After finishing this research we can provide research report to the concern person/authority which will be useful for them in improving the tourism sector in Oman. Now in our project we will take the people/tourists opinion about all Oman tourism region wise, monument wise, specific place wise etc. to get more accurate result and then we will analyze these result for the improvement to make it more attractive for tourism in Oman.
In this proposed research we will use technologies: Apche Hadoop ecosystem, Apache spark, Rapid Miner and Flume to perform big data analytics by following CRISP methodology.

Project Significance:

- Social network Analysis(SNA) : Is mapping and measuring of relationship and flows between people,groups,organizations,computers or other information/knowledge processing entities.
- Explore big data from social media & Tourism websites in Oman (Weblog data) : Social media indeed has become a main communication network in the daily life of people around the world ,so that social media is genetating insurmountable data . This big data are just waiting to be explored. In fact, social media now embodies the leading and biggest source of consumer data.
- Search problems of the tourisms sites that visitors visit : We can get problems points and try to solve it and make survey that will be useful for tourism sector.
- Social media analytics and velocity Explanation: Data in social media is streaming at exceptional speed that must be dealt witin a well_timed manner .It would be interesting to explore this feature in social media analytics as this is one of the great challenges for many organizations.
- Social media analytics and variety: Data in social media come in all types of formats structured numeric data in traditional database, information generated from line _of_business applications, unstructured text documents ,email,video,audio,stock ticker data and financial transactions all comes in different types of formats.

Methodology

Proposad research project will use **Cross-Industry Standard Process Methodology** along with the following Algorithms**:**

- **Build Twitter application and Weblog application utilizing the Twitter Application Management and Weblog data.**
- **Take the important keys from Twitter application and acquire ,**Consumer key,Consumer Secret , Access Token and Access Token Secret.
- **Get the information utilizing Flume using Oman tourism as Key words. Obtain the** twitter4j-core-4.0.2.jar, twitter4j-stream-4.0.2.jar and twitter4j-media-support-4.0.2.jar and replace into the flume dedicated folder of jar files.
- **Build external table for stacking the information got from twitter as key and esteem utilizing Hive.**
- **Split every tweet into words**
- **Make lexicon content record that contain a large number of positive and negative words with their rating.**
- **Mapping the words split from tweets with words in the lexicon to check what number of positive and negative words.**
- **Tally the level of positive and negative opnion**
- **With aditional KPI's**
- **Visualize the outcome utilizing suitable application.**

Time Line

1. Literature review:

Gathering of essential and optional information that will comprehend the enormous information investigation utilize case and advancements.

1 month

2 Design the architechtural diagram :

In light of the choosen huge information innovation Hadoop, hive and flume and task prerequisite outline the archtechtural chart.

1 month

3 Design detailed process diagram:

In view of the proposed compositional chart deliver the itemized procedure graph.

1 month

4 Data collection:

Gather the information utilizing Twitter Application by suitable keywords in windows condition.

1 month

5 Environment setup:

Hadoop ecosystem system required a devoted setup on unix/linux condition. Need to introduce virtual machine and Hadoop biological community including: Hive , Pig and Flume.

1 month

5 Execute the model :

In light of the system, structural and nitty gritty outline execute the model for the investigation.

1 month

6 Coclusion:

In view of the investigation result, the proposal will be given about the need of changing the activity framework in oman.

15 days

7 Visualization :

In light of the discoveries visualize the outcome utilizing suitable tool.

15 days

8 Documententation :

The finished work will be documentated in the given format.

1 month

Bibliography

1. J. Ashok Kumar1, S. Abirami, 2015,"An Experimental Study Of Feature Extraction Techniques In Opinion Mining", International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol. 4, No. 1.
2. Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori, 2013,"Retrieving Product Features and Opinions from Customer Reviews", Intelligent Systems, IEEE ,Volume:28 , Issue: 3
3. Su Su Htay and Khin Thidar Lynn, 2013,"Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", Hindawi Publishing Corporation, The Scientific World Journal, Volume 2013, Article ID 394758

4. Richa Sharma,Shweta Nigam and Rekha Jain, 2014,"Mining Of Product Reviews At Aspect Level", International Journal in Foundations of Computer Science & Technology , Vol. 4, No. 3.
5. Dim En Nyaung, Thin Lai Lai Thein, 2014,"Feature Based Summarizing From Customer Reviews", International Journal Of Scientific Engineering and Technology Research, Vol. 03, Issue. 46.
6. Madhavi Kulkarni,Mayuri Lingayat, 2015,"Effective Product Ranking Method based on Opinion Mining", International Journal of Computer Applications (0975 – 8887),Volume 120 – No. 18.
7. Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, 2013,"Opinion Mining And Analysis: A Survey", International Journal on Natural Language Computing (IJNLC), Vol. 2, No. 3.
8. Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, 2015, "Amazon Review Classification and Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 6.
9. Asmita Dhokrat, Sunil Khillare, C. Namrata Mahender, 2015,"Review on Techniques and Tools used for Opinion Mining", International Journal of Computer Applications Technology and Research, Volume 4– Issue 6, 419 – 424.
10. Farhan Hassan Khan, Saba Bashir and Usman Qamar, 2013,"TOM: Twitter opinion mining framework using hybrid classification scheme", Elsevier.
11. Jayashri Khairnar and Mayura Kinikar , 2013,"Latent Semantic Analysis Method used for mobile rating and review summarization", International Journal of Computer Science and Telecommunications, Volume 4– Issue 6.
12. Bing Liu, 2012, Sentiment analysis and opinion mining, Morgan & Claypool Publishers
13. J. Ashok Kumar, S. Abirami, S. Murugappan, 2014, "Performance Analysis of the Recent Role of OMSA Approaches in Online Social networks", CS & IT-CSCP.