Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 50 (2015) 433 – 439

Procedia Computer Science

2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Experiencing company's popularity and finding correlation between companies in various countries using Facebook's insight data

Harish Hothi[a], Dr. Saleena B[b], Prakash B [c]*

[a]School of Computing Sciences and Engineering, VIT University Chennai,Tamilnadu 600127, India
[b]School of Computing Sciences and Engineering, VIT University Chennai,Tamilnadu 600127, India
[c]School of Computing Sciences and Engineering, VIT University Chennai,Tamilnadu 600127, India

Abstract

The aim of this research was to analyse and experience the various electronics company profiles in various countries using giant social media, Facebook. This analysis was performed with the Insight data of Facebook's page which provide 4 different count values named day, day_28, week and lifetime respectively. To analyze the company's performance in various countries, aggregation was performed to find total users in a country those are engaged with different Facebook pages. All these four counts were used to compare various companies popularity using various measures like Total Country in which people knows about Company, Top-K Country and Least-K country, Count Comparison, Country wise Standard Deviation, Correlation between two companies in a country. Analysis results proved that Samsung was more popular in most of the country compared to all other companies. These findings will definitely help the companies in improving their popularity in social media, which intern will improve their business.

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

Keywords: Facebook; Correlation; Standard deviation; Page insight data, NoSQL

* [a] Corresponding author. Tel.: 919374794079.
[a]*E-mail address:* hothi.harish2013@vit.ac.in
[b] *E-mail address:* saleena.b@vit.ac.in
[c] *E-mail address:* prakash.bala@vit.ac.in

Peer-review under responsibility of scientific committee of 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)
doi:10.1016/j.procs.2015.04.010

1. Introduction

Grouping of individuals, organizations based on common interest like learning, research, business and communicating (socializing) between them through online medium led to the formation of Social networks. In 20th Century's digital world rising of social media gave new ways and opportunities to the people to organize themselves. Recent peak growth of social media gives new opportunity for growing business. Nowadays popular social networks are Facebook, twitter and Google plus. According to US Census Bureau, the statistics of people engaged with social media as on January 2014 were 1,856,680,860. Looking towards this high number which is around 26% of the whole world's population; companies consider social media as a new track for marketing their products. In 2011 76% companies planned to build good popularity in Facebook.

Facebook provides user profile to represent a user. It also provides facility of Facebook page to represent public figures, brands, organization and business. To get updates from company, Facebook users can engage with company's Facebook page by liking the company's page. This represents a bonding between the user and company. Users those are engaged with page can give their feedback to page's update by posting comments or giving likes to a post. This interaction between the page and user helps the company to know its popularity in Facebook and give a chance to improve and grow their business. Facebook is open to all, so companies can easily find their competitors through Facebook.

Most of the companies have their business across many countries. So it is very crucial to find company's popularity in various countries to experience where they are lagging and where they have high crowd in social media. The motivation for this work comes from this truth.

For each page Facebook provides Insight data, provided that page has at least 30 users associated with it. Insight data provide valuable and rich information on the performance of the page in various countries. Insight data has 4 different counts. **a) Day count** represents the number of unique users talking about the Page daily by user country. **b) Days_28 count** represents the number of unique users talking about the Page in past 27 days by user country. **c) Week count** represents the number of unique users talking about the page in a week by user country. **d) Lifetime count** represents Lifetime: Aggregated Facebook location data, sorted by country, about the Unique Users who like a Page. By getting these counts for different country we can compare different company's popularity in various countries. By finding all pages those are related to a company and getting all pages insight data then performing aggregation based on the country will give total users in a country those are engaged with facebook.

2. Related Work

Martin Grančay et al. [1] present content analysis of 250 official Facebook pages for 250 different airlines. This paper also uses only 'like' count. Author provides how various airlines use Facebook to handle their business. Author describes how airline companies operate their Facebook page by providing reply to customer query on Facebook, how frequently airline update their flights update.

Finding the correlation between various Facebook communities using page's like count was discussed by Nikos Salamanos et al. [2]. BFS algorithm was used to collect the data from various Facebook pages. This paper uses only like count to get the correlation between Communities vs. Population and between communities, top-k page. The problem with like count is, it does not give actual information since some users do not regularly use Facebook.

Sean Munson et al. [3] discussed the study of two facebook page named "3 good things" and "goal post". Author study user behaviour those are associated with page by analysing user's post.

Shohei Ohsawa et al. [4] provides like prediction model for facebook page using DBphidia and facebook like count for page.

Unlike other previous papers our study includes all 4 count provided by Facebook to find country wise correlation between two companies, top-k and least-k countries for a given company, standard deviation for a given company and country, graphical count comparison between companies make more robust and good informatics analysis.

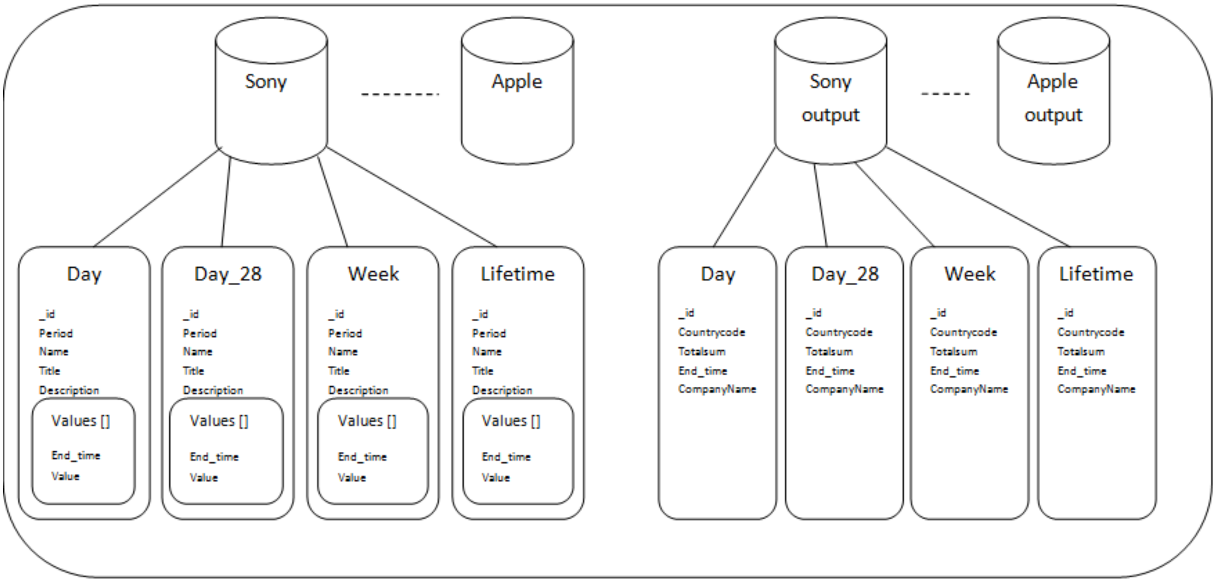

Fig. 1. Storage Model

3. Data Gathering, Storing and analyzing

Our study involved seven giant electronics companies named Apple, Acer, Dell, HP, Lenovo, Sony and Samsung. A java based application was designed for data collection, storing and analyzing. Using the core Java API, restfb library is used to fetch data from Facebook, JFree chart API to visualize data through charts. MongoDB as database to store collected data into the database.

3.1. Page Searching

For data collection from Facebook, when application gets a request to find pages for ex. page named Samsung, application logged in to Facebook from our created Facebook application and search given keyword in Facebook through restfb library which internally uses Facebook graph API. As a result of searching, a set of relevant and irrelevant pages were discovered for our study. For example, when a search for Apple is done, the search results in some irrelevant pages which are of apple fruit and not Apple Company, here filtering have to be done on acquired search result.

3.2. Page Filtering

All acquired pages are filtered by identifying their category defined by the Facebook. To get a page category, the page's basic information is fetched using restfb library and compared with its category to check whether a page is relevant to our study or not. To filter pages by their category we include following Facebook category in our consideration "Business Services, Professional Services, Community Organization, Company, Computer Technology, Consulting Business Services, Internet Software, CameraPhoto, Computers, Electronics, PhoneTablet, ProductService,Software,Website, Shoppingretail,Local business, Computerstechnology".

3.3. Collecting and Preprocessing Insight Data

After fetching a set of relevant pages, the insight data for each page was extracted using restfb library and respective page-ids. Then each count was separated from Insight data and converted into BasicDBObject which is directly inserted into MongoDB's database's respective collection as a document. A MongoDB collection for each count was created in the database. MongoDB stores these documents in a binary representation called BSON (Binary JSON).

3.4. Storage Model

Primary requirement of proposed system is to retrieve relevant facebook page insight data and store it in database then perform aggregation or map-reduce jobs and other analysis. The structured approach of RDBMS database like SQL slows down performance as data volume or size gets bigger and it is also not scalable to meet the needs of analyzing large volume of social data. So uses of MongoDB NoSql design helps to overcome manipulation of large data and achieve the primary requirement of proposed system. A MongoDB storage model for application includes seven databases for each company and each database has four collections named as day, day_28, week and lifetime which corresponds to counts in Insight data. Figure 1, shows storage model for application.

4. Analysis and Results

This section describes the various analysis performed on the gathered data. Results were implemented using Java programming language. Various analysis on gathered data are performed using the aggregation operation. Aggregations are done on each company and its collection. Aggregation operation on 'day' collection of Samsung database gives the total number of users in each country, who operates on any Facebook page of Samsung company on a particular date. The term operation here means liking any post or liking any page or give comments on any post. Similarly aggregation on ’day_28’ and 'week' collection gives total number active users in past 28 days and the total number of active users in the last week for each country. Aggregation on 'lifetime' collection gives the total number of likes in each country. Output of aggregation operation is stored in respective output database. The aggregation output of Samsung’s lifetime collection stored in Samsung output databases lifetime collection.

4.1. Top-K countries and Least-K countries

All the documents from each collection are sorted according to their ‘TotalSum’ field, the output of aggregation operation. After that the first K countries are selected which give Top-K country for that company. Top-k operation on ‘day’ collection gives Top-K countries having highest active user for giving date. Similarly Top-k operation on ’day_28’ and ’week’ collection gives Top-k countries having highest active users in last 28 days and in last 7 days . Table 1 shows Top-5 countries with their day count for each company. Opposite of Top-K countries, Least-K countries gives least K countries for a given company. Least-K countries show where company is less popular , so that they can promote their products to grow their business in those countries.

Table 1. Top 5 Countries for all companies based on their day count

Top-5 country	Apple	Acer	Dell	HP	Lenovo	Sony	Samsung
1	IN-6556	TH-602	IT-93317	IN-4413	MY-7741	ID-46130	IN-73080
2	US-3145	MX-455	US-2120	US-3818	PH-2407	IN-22441	PH-40944
3	PK-1848	PE-425	IR-2098	MX-2910	VN-1892	BR-10140	ID-35217
4	BR-1529	IN-405	BR-1985	CO-2805	IN-1732	KR-9474	TH-34764
5	MX-1421	CA-389	IN-1928	BR-2649	ID-886	TW-8942	BR-23087

4.2. Total Country in which People Know About Company

This result analysis shows the total countries where people know about a particular Company. Seven companies as listed in Table 2 was for our research purpose and the analysis shows that Samsung has highest count of 220 that means in 220 different countries people know about Samsung. At same time Dell has least count. We can also interpret this result as; Samsung has business in different 220 countries since people know about Samsung in those countries.

Table 2. Distinct countries covered by company

Company Name	Total Distinct Country
Apple	210
Acer	196
Dell	194
HP	201
Lenovo	181
Sony	211
Samsung	220

4.3. Count Comparison

This analysis gives bar chart representation of each count i.e. day, day_28, week and lifetime count. Bar charts are generated in two ways. **a) Bar chart for a particular country**, where X axis represents the active users of a particular country in different days and group of companies for which we want to visualize and compare the various Facebook page count and Y axis as respective count values. The chart below depicts how many people were associated with different companies from a particular country. It was identified that we have 220 different countries using Samsung products so there will be 220 different plots. Figure 2, 3 and 4 shows the number of daily active users of various companies in Canada, Italy and India respectively. **b) Bar chart for a particular date and given company** can also be generated. This type of bar chart represents how many people from various countries are associated with a company. Here the X axis represents different countries and Y- axis represents the different counts for a given day. Since our study involves seven companies so we have 7 different chart each having one company.

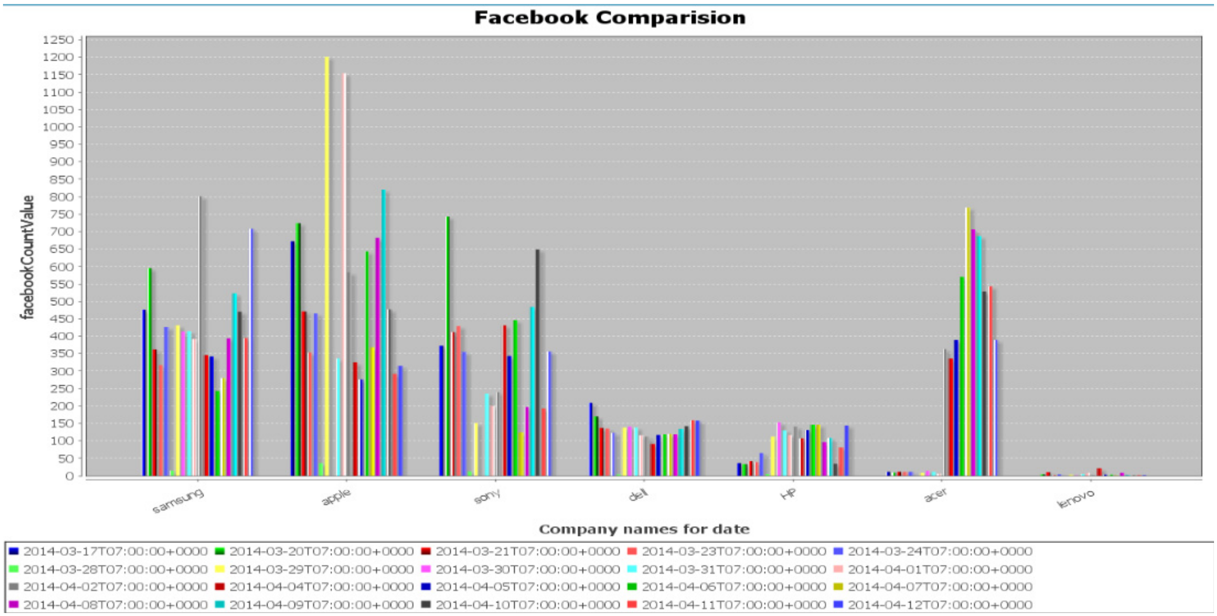

Fig. 2. Daily active users of Canada

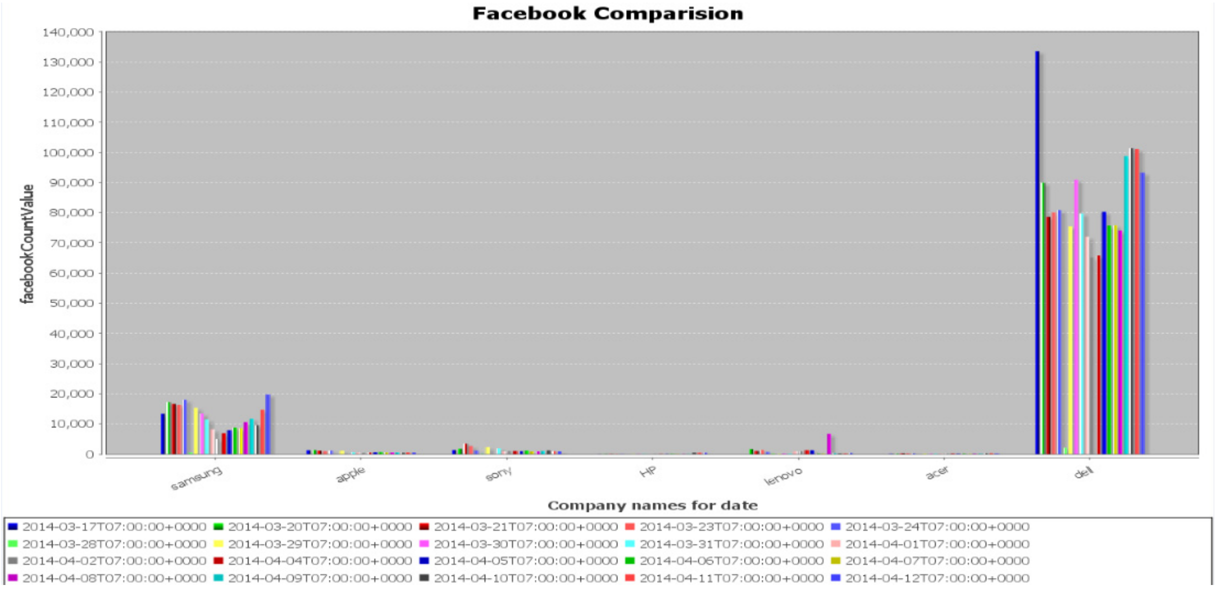

Fig.3. Daily active users of Italy

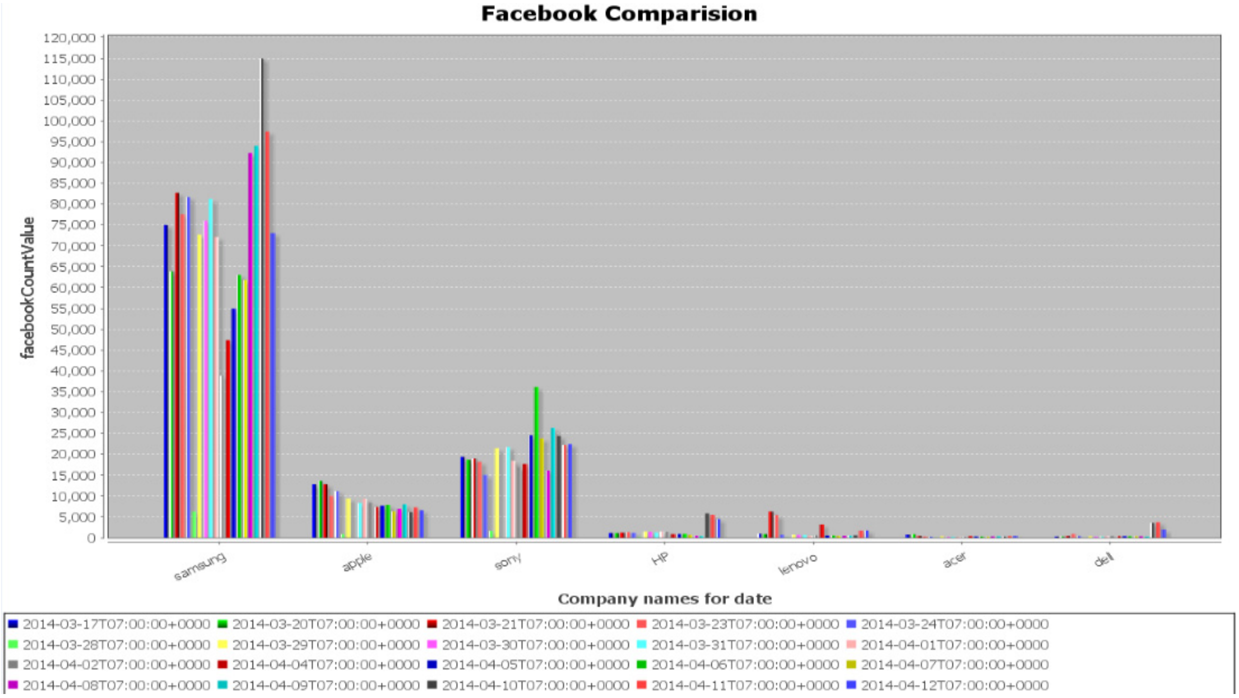

Fig.4. Daily active users of India

4.4. Standard Deviation

To measure how the numbers are spread the standard deviation measure is used. In our study, the standard deviation for each company is measured with respect to their Facebook count. The standard deviation on ’day’ count for a given country and company shows how people’s interactions with different Facebook pages are deviated in social media. Similarly, Standard deviation on ’day_28’ and ’week’ count for a company shows during the past 28 days and during the past week how people's interaction is deviated in a given country. The result of the Standard deviation on 'day_28' collection is more important compared to the standard deviation on 'day' collection since there are some users that do not engage with company's page daily. 'Lifetime' collection shows how many total people are associated with a Company in Facebook for a given country, here it does not matter whether people are active or not. The standard deviation on 'lifetime' collection shows the deviation of newly engaged people of a country.

4.5. Correlation

Correlation analysis is useful to check how two variables are related. Popularity in Facebook can be defined as how many people are associated with any Facebook pages of a company. All 4 counts directly reflect the popularity. Correlation analysis is included in our study to check how two company's popularity growth in Facebook is related to each other's. Using Pearson correlation, country wise correlation analysis was done, for a country 'i' and two company 'k' and 'j' we finds 4 correlation results using each count. Correlation using day count gives a measure of daily correlation between company 'k' and 'j'. Similarly using day_28 and week count give a measure of past 28 days correlation and past week's correlation. Correlation analysis on lifetime count represents exciting results since lifetime count shows how many users are engaged with your page. So country wise correlation using lifetime count cover all user for a country 'i' and then give correlation between company 'j' and 'k'. Results show high correlation between two companies 'j' and 'k' in some countries and vice versa.

5. Conclusion

This paper investigates the company's popularity in Facebook as well as the correlation between two companies. Results shows that Samsung is more popular in almost every country compared to other companies. Our study experienced that there is a huge crowd of active Facebook users who are daily engaged with Samsung in some countries like India, Philippines. The experimental results on correlation analysis show high correlation between companies in some countries and low correlation in some countries. Interacting with the users by posting relevant updates in their pages will help the companies to upgrade their reputation in social media and improve their business. Our Future work aims to build country wise 'like' prediction model for a company which helps company to predict its popularity in various countries.

Acknowledgements

I would like to extend my special thanks to my guide for providing excellent guidance throughout my work. Also we would like to thank all the reviewers for providing constructive comments which helped us in refining the paper to this extent.

References

1. Martin Granay, Airline Facebook pages-a content analysis, European Transport Research Review, 213-223, (2013).
2. Nikos Salamanos, Elli Voudigari, Theodore Papageorgiou, Michalis Vazirgiannis, Discov- ering Correlation between Communities and Likes in Facebook, Green Computing and Communications (GreenCom) 2012 IEEE International Conference , 368-371 (2012).
3. Sean Munson, Beyond the share button: making social network sites work for health and wellness, Potentials-IEEE, 30 issue 5, 42-47 (2011).
4. Shohei Ohsawa, Yutaka Matsuo, Like Prediction: Modeling like Counts by Bridging Facebook Pages with Linked Data, International World Wide Web Conference Committee (IW3C2), 541-548 (2013).
5. https://developers.facebook.com/docs/graph-api
6. https://developers.facebook.com/docs/insights
7. http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient