

Ranking Algorithm

Viraj Anchan¹, Sarang Deshpande², Deep Doshi³, Akshat Kedia⁴

Student, Computer Engineering Department, MPSTME, Mumbai, India^{1,2,3,4}

Abstract: In this paper, different ranking methods have been compared along with the applications using them. We have also proposed our own ranking algorithm that can be used in any Android application to rank the different advertisements or posts, posted within the application. We have also considered further possibilities that could be used within our algorithm.

Keywords: Ranking, Android, User Rating, Sorting.

I. INTRODUCTION

This research paper serves two purposes: (1) Comparison of various ranking algorithms currently used by websites like Quora and Stack Overflow. (2) Proposing our own ranking algorithm. Various websites use different ways to rank the posts, the users or any other content on their websites. Our ranking algorithm is ideal for user posts-based applications. Our ranking algorithm will not only help the users make good choices within the application, but will also improve the search results and the way they will be shown. The different parameters required for the algorithm were analysed first.

The criterion for ranking was decided based on which parameter has to be given more weight age and which has to be given less weight age. Accordingly, the formula was designed such that it could be changed according to the need of the application

II. OVERVIEW OF EXISTING ALGORITHM

A. Ranking Algorithm – Quora

Quora is a website which allows user to register and then post questions on the website. Different users can post answers to it and they are ranked from time to time. According to Quora, the company uses certain parameters to rank the answers posted by the users for the questions on the website. The parameters are listed as:

- More the number of upvotes, higher the rank.
- If number of downvotes is higher, it is ranked low.
- Votes of users writing good answers are considered of more value than the others.
- Answers written by users who have written good answers in the past are ranked higher.
- Users spamming or bullying the system, their vote will not be considered.
- Ranks are not affected if the user is an administrator or a reviewer.

Quora mentions that, to get a good rank, the users should focus on writing good answers.

The algorithm is similar to that of Page Rank, which is used by Google. The algorithm helps to influence the users to add appropriate content to the site. [1][2]

B. Ranking Algorithm – Stack Overflow

Stack Overflow is a website where users post questions related to any programming language and different users can post answers to the question. The answers are then ranked using an algorithm. Stack Overflow website has an algorithm to rank the answers posted for any programming related question on the website. It mainly uses the two parameters, which are, upvotes and downvotes. It subtracts the downvotes from the upvotes to calculate the rank. But, in addition to it, there are certain other factors that play a role in determining the rank of the user on Stack Overflow:

- Score of the user posting the question
- Views for the question
- Answers for the question
- Scores for the answers posted
- Date and Time of posting the question
- Date and Time of last viewing the question
- Reputation of the user asking the question
- Reputation of the users answering the question

Post consideration of these requirements, the website uses a formula to calculate the final rank of the question and accordingly the rank of the user. Based on this, the question is displayed on the website.

We have also observed that posting redundant questions also incur a penalty in the user's score and frequent posting of such questions can also impose a ban on the user from using the Stack Overflow system. [3]

III. RANKING ALGORITHM

As mentioned, we have developed this algorithm specific to Android applications that display posts or advertisements and may have a need to display them in certain order. Two formulae have been devised to be used for ranking. One of the formulae contains two parameters, while the other one takes into consideration three parameters.

This algorithm has been written and tested using the Java Programming language. This algorithm is very concise, calculates the rank in minimal number of steps, thus saving lot of memory during execution of the algorithm.

A. Identifying The Constraints

We figured out the primary constraints that are required to calculate the rank of the posts or advertisements. The most important parameters are the date of posting the post or advertisement and the number of views for it. The algorithm calculates the rank every time any user views the advertisement or, as, the number of days since posting increase.

The second formula considers another parameter (along with the other two used in the first formula), which is, favorites. This means, whenever, any user adds the advertisement as favorite, the favorite count for that advertisement increases. This makes the algorithm run again and calculate the rank again. [4]

B. Steps to Formulate the Ranking Algorithm

- In the Java Program written to implement the algorithm, the inputs are taken from the user of the various parameters used.
- First, the date of posting the advertisement is taken as input. (It will be parsed from within the application when the application will run on a device).
- Second, the number of views is taken as input (Similar to the date, it will also be parsed when the application runs on a device).
- These parameters are stored in an array. These values are passed to a function to calculate the rank.
- Since the date of posting is in a proper date format, the difference in that date and current date cannot be determined. Hence, using a function, both the dates are converted to milliseconds. These are number of milliseconds elapsed since January 1, 1970.
- Once the number of milliseconds for both the dates has been calculated, the difference of one from the other will produce the result of difference between the two dates.
- Since this difference is also in milliseconds, it has to be converted to number of days using standard conversion formulae.
- After obtaining the difference in the current date and date of posting, it can be used in the formula.
- When the rank of the advertisement is calculated based on these two parameters, the rank is stored in an array.
- The above steps are repeated for all the advertisements posted by users and are stored in the same array.
- When all ranks have been calculated, the array is sorted using selection sort and the array is returned to the main function. In the application, the array will be returned to the client and the advertisement list will appear sorted to the client.
- If the second formula is used, a third input, favorite count, has to be taken and the rank will be calculated after that.

C. Algorithm

If the difference in the number of days and current date is very less and the number of views is very high, the rank will be high. Conversely, if the difference is high and the

number of views is moderately high, the rank will not be as high. Similarly, if the difference is high and the number of views is low, the rank will be very low.

The equation along with the algorithm to calculate the rank is given below:

```
if(days>0)
{
    user_rank[j]=(float)((((10.0/days)*10)*(view/30.0));
}
else
{
    user_rank[j]=(float)(10*(view/5.0));
}
return user_rank[j];
```

The 'days' variable depicts the difference in the date of posting and the current date. The 'view' variable is used to store the number of views for an advertisement. The 'user_rank' variable is an array that stores rank of individual advertisements. In the above piece of code, if the date of posting is the current date, then only the number of views will be considered and the rank will be calculated. If there is a difference in the date of posting and the current date, the difference will be taken into consideration. Various pieces of the equation are explained below:

10/days: This is performed because the difference is inversely proportional to the rank. It is multiplied by a factor of 10 to make the rank a bigger number or else it will show a rank in a very low range of numbers.

View/30: This is performed considering that the average number of views for advertisements will be 30. If the difference in days is zero, the average number of views assumed is 5. These values can be changed based on assumption of number of views expected.

The algorithm for the second formula, which includes the favorite count is similar to the above algorithm and is as mentioned below:

```
if(days>0)
{
    user_rank[j]=(float)((((10.0/days)*10)*(view/30.0))*(favorite/10.0);
}
else
{
    user_rank[j]=(float)(10*(view/5.0)*(favorite/5.0));
}
return user_rank[j];
```

In the above algorithm,

Favorite/10.0: This depicts that the favorite count is expected to be 10, on an average (if there is difference in days). Else, it is expected to be 5.

After the rank is calculated, it is returned to the main function and the algorithm is once again implemented for another advertisement. Once all the ranks are calculated,

the whole array is sent as parameters to a sorting function. This function uses selection sort to sort the ranks. Finally, the sorted array is returned. [3]

D. Implementation

We implemented the algorithm using NetBeans IDE and the result obtained is as follows:

```
Enter number of views and date in format dd-MM-yyyy
50
20-03-2015
35
25-03-2015
28
25-03-2015
40
21-03-2015
32
30-03-2015
43
18-03-2015
65
11-03-2015
55
14-03-2015
20
30-03-2015
36
08-03-2015
```

Fig. 1. Input to the function

```
Rank: 64.0
Views: 32
Date: Mon Mar 30 00:00:00 IST 2015
Rank: 40.0
Views: 20
Date: Mon Mar 30 00:00:00 IST 2015
Rank: 23.333334
Views: 35
Date: Wed Mar 25 00:00:00 IST 2015
Rank: 18.666666
Views: 28
Date: Wed Mar 25 00:00:00 IST 2015
Rank: 16.666666
Views: 50
Date: Fri Mar 20 00:00:00 IST 2015
Rank: 14.814815
Views: 40
Date: Sat Mar 21 00:00:00 IST 2015
Rank: 11.944445
Views: 43
Date: Wed Mar 18 00:00:00 IST 2015
Rank: 11.458333
Views: 55
Date: Sat Mar 14 00:00:00 IST 2015
Rank: 11.403509
Views: 65
Date: Wed Mar 11 00:00:00 IST 2015
Rank: 5.454545
Views: 36
Date: Sun Mar 08 00:00:00 IST 2015
```

Fig. 2. Output

The date on which the program was implemented was 30th March 2015. This is taken as the current date by the algorithm for calculating the ranks.

As it can be seen in the input, the date is given as 'DD-MM-YYYY' format. The output shows the rank of the advertisement along with its number of views and the date

of posting the advertisement. The date is still stored in the same variable and can be displayed as it is. Once the user enters the date, it is stored in a string variable and is converted to be stored in the Date type variable. This is the reason for the date being displayed in such a format as it can be seen in the output. It also assumes that the time is 00:00 and the time zone is set to IST.

The rank calculated is stored as a float variable and is hence displayed with decimal places. This is done in order to sort the advertisements more accurately. If the rank is not taken as a float value, then the ranks will be rounded off to integers. Because of that, many advertisements may have same ranks and would be difficult to sort.

- We have only shown the implementation of the algorithm using the first formula. The implementation of the algorithm using second formula, considering the favorite count is similar to this implementation.

IV. COMPARISON

To compare our ranking algorithm with the current ranking algorithms implemented by various websites like Quora and Stack Overflow, we consider the parameters and the user base of the different systems.

- Parameters considered by the other systems include the user's current rating, the ratings based on the answers or upvotes/downvotes. In our application, we have included the favorite count, which is similar to the upvotes/downvotes.
- As mentioned in the earlier section, Stack Overflow subtracts the downvotes from the upvotes so as to calculate the rank. Applications for which this algorithm is targeted do not have any downvotes. There is no negative voting towards an advertisement. Instead, the favorite count is taken directly as a parameter and implemented in the formula.
- Another huge comparison that can be made to these systems is the user base. The user base of such applications is very huge. There are approximately millions of users who are registered with these applications. Our algorithm is designed for applications with a small number of user base. If the user base of these applications increases, then we can consider changing the weight given to various parameters in the algorithm or consider adding new parameters.
- Also, because of a small user base, the server that is chosen is also not capable of handling large number of requests at the same time.

V. FUTURE WORK

We have developed a ranking algorithm that is sufficient for the application that we will be using it for. On the other hand, there could be certain modifications to the algorithm that could help to calculate the rank in a better way.

There are two elements that could be integrated to the current algorithm.

First, there could be tagging included. This can be implemented if the user who is posting it tags the advertisements with certain words. The tags can then be used to rank the advertisement. The tags can be checked against a list of current trending tags. If the tags are present in the list, the rank will be higher as compared to, if the tags are not present in the current list of trending tags.

Secondly, clustering of words can be done. Certain words can be combined into one cluster. Each cluster can be assigned a certain rank or probability. If the words in the advertisement match any of the words in the cluster, the rank will be higher. Conversely, if the words do not match the ones in the cluster, the advertisement will be ranked lower.

VI. CONCLUSION

In this research paper, we have identified various ranking algorithms that have helped us to devise our own ranking algorithm. We identified the constraints and then designed our algorithm in two ways, one having two parameters and the other having three parameters. We have also compared our algorithm to the others currently being implemented by companies running web applications.

ACKNOWLEDGMENT

We would like to express our special thanks to our mentor, Prof. Krishna Palod for her support and guidance. We take this opportunity to express gratitude to all of the Department faculty members for their help and support. We would also like to extend our sincere thanks to the University and individuals who have helped us.

REFERENCES

- [1]. Audrey Watters. (2011, Feb 6). Quora's New Algorithm for Ranking Answers.[Online]. Available: http://readwrite.com/2011/02/06/quoras_new_algorithm_for_ranking_answers
- [2]. Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H. & Zhao, B. Y. (2013), Wisdom in the social crowd: an analysis of quora., in Daniel Schwabe; Virgilio A. F. Almeida; Hartmut Glaser; Ricardo A. Baeza-Yates & Sue B. Moon, ed., "WWW", International World Wide Web Conferences Steering Committee / ACM, , pp. 1341-1352 .
- [3]. Kartik Bajaj, Karthik Pattabiraman, Ali Mesbah, "Mining questions asked by web developers", in MSR 2014 Proceedings of the 11th Working Conference on Mining Software Repositories, Association for Computing Machinery, New York, NY, USA, 2014, pp. 112-121
- [4]. Batra, N. ; Kumar, A. ; Singh, D. ; Rajotia, R.N., "Content based hidden web ranking algorithm(CHWRA)", in Advance Computing Conference (IACC), 2014 IEEE International, Gurgaon, 2014, pp. 586-589.