International Journal of Engineering Research in Computer Science and Engineering Engineering (IJERCSE) Vol 2, Issue 3, March 2015

Increasing the trustworthy of user by Multiple's Method

[1]Abhaypratap M. Singh,[2]Aniket Raut,[3]Akshay Tule,[4]Mithil Wasnik
[1][2][3]Student, Computer Technology, Rajiv Gandhi college of Engineering and Research, Nagpur, Maharashtra, INDIA
[4]Lecturer, Computer Technology ,Rajiv Gandhi college of Engineering and Research, Nagpur, Maharashtra, INDIA

Abstract: **Accessing the data from various website are day to day life work for the various purposes. When we access data from various websites there have no guaranty that provided data is correct. So t he user are always in confusion that provided data on website are correct or not some of the famous website also even provided sometime incorrect data. So to provide the true data to user with multiple checking of same data to different website concept are introducing in this paper. With help of this method not fully we satisfied the condition but fulfilled at the some instinct level.**

I. INTRODUCTION

Web crawling are our daily routine work for developing of technology into various field, searching of data for projects, data to know about current information, latest technology and various different purpose. In current condition half of the people are used website to find out various kind of information about anything they want to find out. To find out the different kind of information from different website we need trustworthy of website to take data for some important work. Suppose we use data for the research work, or to find out the data for the data about any new cell phone introducing into the market, compare the various price value of the same thing on different website. We don't get the correct data from the different website on which we trust. Because different websites provides different varieties of data on same things, which we are searching into the web crawler.

To increase the trustworthy of user on website data, we introduce some method to resolve this problem. With the help of multiple comparisons of same data on different website increased the trust of user on website data. The same data values of the different website are collected into the file and after the comparison of those data values we provide the trusted website on the basis of different data value comparison provided by different websites.

Nowadays, people use websites for online shopping of books, cell phones, cloths, plot, etc. When they compare the same item on different website they get the different specification of same things. So how they will trust on the different specification on various websites which will provided the true information. So to overcome this problem we introduce the mechanism which we called as Conflict

Resolve by Multiple Facts. With help of multiple facts we provide some level of trust on the website data which we used for different purpose. So by comparison of multiple facts we provide the correct website which is providing the true data even if the most popular website gives the wrong information. We cannot justify that which website is providing correct data and which one is incorrect data provider.

I. LITERATURE SURVEY

In the various papers we have studied that the web crawling is important aspect of nowadays life. So to reduce the conflicts on various data values they provide various mechanisms. In one paper mention that various facts on different website lead to the veracity problem, so to reduce the veracity they define the some restriction towards the crawling.

When the data are crawling on websites, on that time various facts get generated so they have to manage into the proper format of same or different kind of data values of every websites data. On the time of comparing various facts we have to properly manage the result of that data, on that time the time factor leads to be the problem may caused. When deals with some of the uncommon facts than it should be managed into the extra data values and the plus facts of that websites. But on the time to comparing we do not have to neglect

those facts, but it may be happen that some of the websites give the same data values.

So with the help of this survey we introduce the facts comparing between the times and give the best results to the end user. So on the time of the crawling we providing the number of facts are in certain limit which is depend on the accuracy required by the user on the time of crawling. Depend on the user requirement we providing them accuracy of the data which he/she searched so the proposed method providing below.

II. Web Crawling Definition And Analysis

A. Web Crawling:

Science and technology both required the web crawler to get the data in easier way. A web crawler is the interface between the user and the world data. It provided the easiest way to find out the data from the Internet. To find out the data from the various website is nothing but a crawling with the help of the web (internet) is known as "web crawling".

Web crawling is the part of our way to find the data from the one part of globe to the other part. Web provided the various types of data and every type of data available but there have no guaranty that provided data are correct we cannot easily justified. But with the help of our algorithm it is easily justified that which data is correct or incorrect. It is the techniques which provide the mechanism on the basis of multiple facts we justified the correct or incorrect data from web provider.

B. Analysis:

Web is the easiest way to find out the data, so due to that everyone try to use the web to find out the various types of data in single click. So the user must needed the correct data from the websites but it is not possible because every website have their own way to collect the data so the chances of data variation are always be there. But user must need the correct information from the website so that purpose it needed some kind of application who provided the right to take correct information from the website.

Our algorithm is fulfilled the requirement of this problem. Our mechanism provided proper way to justify the correct data from the website. By compared multiple facts on single object, from various different website. It reduced the chances to give the wrong output. We are working on the basis of polling mechanism condition.

III. ALGORITHM AND METHOD

In this algorithm data values of the object are compared with the data value of same object on different websites.

A. Algorithm:

W – websites, F- facts, O –Output

Step1: Collect the data values of on object from website w1.

Gather into the file and mark from them the f1, f2, f3, ..fn.

Step2: Use the f1 and compare with the website w2,w3,w4,..wn.

Step3: Take the o1 for that comparison.

Step4: Increase the website Wj, output Oj Go to step 3.

Step5: Increase the fact Fi. Go to step 2.

Step6: Perform the operation until the websites not end.

After the comparison of the data values from the various website it will result some output which give the good website that gave the correct information for the crawled object. Thus the comparison should be stopped otherwise it take to much time to give the result, so we have to bound the crawling into specific bounding as per the requirement.

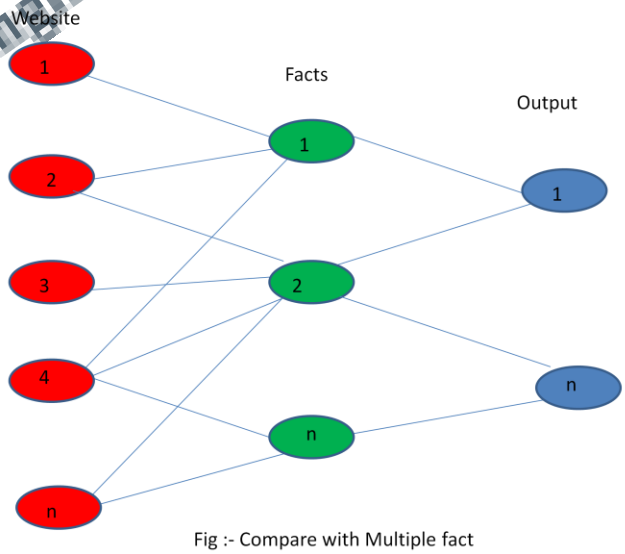

Fig :- Compare with Multiple fact

It is just an overview of how we comparing the data from the different websites. With the help of fig u can easily understand the mechanism it shows that there have the number of websites, number of facts, number

of output website are generated. On the output it depend on the user which website he would like to access for the correct data. There have an N no of websites which compare the data of same facts as well as on the basis of different fact provided by the various websites. On the analysis of this process output would be generated.

When we compare the Wi data values to the Wi+1 data values on the different facts F1,F2,… Fn. The output present gives the some Boolean values which we stored into the file and at the end of this process. We get the some value for every website on that basis we decided that which website is more preferable to provide the correct data to the user. As per the requirement the values would be changed not always there have single website which provide always correct, as we early say that there have most visited websites also sometimes provided the incorrect data.
With the help of this method we also checked out the most popular websites.

Let take an example, suppose we have to find out the author of book from the web, than when we got the result we mostly preferred the first page link provided by the search engine, we are not sure that data provided on the first link are correct so to reduce the most visitation of pages. It saves the time and effort of the person who want to search data with the help of websites. Data of the various website are comparing with the other website data than produce the result. For doing this thing we have to some important constraint on our mind i.e. time complexity of the method because user needed quick result so that is why we bound the crawling into certain limit to produce quick answer.

B. Method:
We are using the fact of various website mechanisms to generate the solution for this problem. When we collecting the data from the websites than those data values are stored in a file and some new fact also store in the file. On the basis of similar facts and un-similar facts of the various website, we preferred that website whose produce the most correct fact as well as some new fact to produce the result. Due to this the result are quite good and the most popular site incorrect data are prevent from the user. The lack of knowledge person who doesn't understand more about the website they also got the correct data from the website.

From of the various websites W1,W2,W3,…,Wn with the help of various facts F1,F2,F3,…Fn, on the time of comparison it perform like

W1 ----- F1,F2,F3,…,Fn
W2 ----- F1,F2,F3,…,Fn,Fn+1
W3 ----- F1,F2,F3,…,Fn-2
W4 ----- F1,F2,F3,…,Fn-1
. .
. .
. .
Wn-1----- F1,F2,F3,…,Fn,Fn+1,Fn+2,Fn+3
Wn ----- F1,F2,F3,…,Fn

Due to this kind of comparisons data produce on the results are good and useful for user for accessing the correct data from the websites. The resulted data produce the link for to user from which he need to use for access the correct data. In the comparison who have looked that on the website W2 in this the fact comparison are more than the other website, because we early mention that the common and uncommon fact are comparing on the different website. With the help of uncommon data we generate the website which having the more data than other website to give more priority than the other website when using the data by the user. But it doesn't mean that website which having not the correct data for common but have more data than other website it not have the higher priority to gave the information.

We have some kind of conflict occur on the time of preferring the website more are as follow as.

a) When the website have correct data as well as more data than other websites.
b) When the website having correct data but doesn't have the more data than other website.
c) When the websites having correct data but more data are not the correct one.
d) When the websites having the incorrect data but have more data than other website.
e) When the websites having no correct data as well as not having more data.

So with the help of various conflict occur on the time of priority of website we prefers the following sequence which we mention above for the priority of the website for output. Due to this mechanism user got correct data as well as more data from the websites. But the resultant value of the every website should be different and according to the link of most preferable with correct data provider link have highest priority to show the resultant data value to the user.

Data which is shown by user should be depend on the user requirement as well as the data present on the various website. Suppose in the thousand of website have the same data but the data are incorrect and one website have the correct data than the priority of the other website are more than that single website. So this kind of condition not satisfied by our mechanism, but guaranty that we provided user correct data on the basis of highest polling mechanism. Due to the use of this mechanism the other low priority website also got the chance to display the data to user for their use.

So with the help of above mechanism we take experiment on some data to analysis the working of the method which is successful or not for the user.

IV. EXPERIMENT

Suppose we have to find the author of the book there some different type of result occur on different websites are

a) Author name Stan lee, cost 500, published by S.K.
b) Author name S. lee, cost 510, published by Swiden.K.
c) Author name Stan, cost 500, published by Swiden.K.
d) Author name Stan lee, cost 510, published by S.K.
e) Author name Stan lee, cost 510, published by S.Knight, etc detail also mentioned

So this type of result occur for the single data search now from the experiment we clearly justify that the Author name Stan lee, cost 510, published by S.K. with the help of most used data by the different website. So on the basis of our result we chose one website who fully satisfy our condition and also provide the extra data to the user and have the first priority to user for using the data for their different kind of work.

Now let us take another example, suppose we have to find out the specification of the mobile so the result occur are as follow as

a) Company name Nokia, price 11,109, Camera 5mp, Operating system windows, ram 2 gb, processor 2.2ghz, sound up to 25db.
b) Company name Nokia, price 11,119, Camera 5mp, Operating system windows, ram 1.96 gb, processor 2.216ghz, sound up to 24db.
c) Company name Nokia Pvt. Lmt., price 11,109 , Camera 5mp, Operating system windows, ram 2gb, processor 2.216ghz, sound up to 25db, screen gorilla glass, touch sensitive.
d) Company name Nokia, price 11,109, Operating system windows, ram 2gb, processor 2.2ghz, touch sensitive.
e) Company name Nokia, price 11,009, Camera 5mp, Operating system windows, ram 2gb, processor 2.2ghz,
f) Company name Nokia, price 11,229, Camera 5mp, Operating system windows, ram 2gb, processor 2.2ghz, sound up to 25db,touch sensitive, cache memory 1 mb, One year warranty, free services for one year, 2600 mA battery, with exciting price.
g) Company name Nokia Pvt. Lmt, price 11,109, Operating system windows, ram 2gb, processor 2.2ghz, cache memory 1 mb, 2600 mA battery.

In the above experiment it is clearly mentioned that the data of various website should be vary according to the condition. So the most used website having lots of data and have chance to provide more correct data to the user.

With the help of this mechanism the data are correctly provided to the user and avoid various conflicts occur which we have mentioned above get removed. But our some mentioned conflicts will not be removed up to that time but it will be removed in future. With the help of this method we solve the problem of data trustworthy of user from various websites. Due to this mechanism data nature should be improved, and the usage data should be trustable for user and no one claim on that available data.

CONCLUSION

With the above experiment we conclude that we get the result in the required time and according to the requirement of the accuracy we need time to maintain the accuracy of data to the user. So the behalf of accuracy and requirement we provide data to user properly on to the requirement basis of user.

FUTURE RESEARCH AND DEVELOPMENT

In this paper we already mention that on the basis of various website polling we decide the trust of data but we had not currently able to solve the problem if a single website have only correct data and other have not that on that condition how we justify that the correct is which one. It is not easy to describe that who said correct but with help of authentication certification provided by the website than we can say that data are correct. It is not easy to identify this kind of strategy to

find out the correct data but after the development of this type of method or mechanism we are proving the data are fully correct and it will lead toward the different way of correctness of data for the user and also for the owner.

With this problem solved it give the great relief to the user who have not fully trusted on the websites, this lead to fully satisfy the condition of person requirement. This research will support toward the trustworthy of user to the various website and also lead to the e-learning of student are more. It leads to be development of science and technology into the great pace and also the development of society.

REFERENCES

[1] Xiaoxin Yin, Jiawei Han (2008) : Truth Discovery with Multiple Conflicting Information Providers on the Web, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 6, JUNE 2008

[2] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized Trust Management," Proc. IEEE Symp. Security and Privacy (ISSP '96), May 1996.

[3] A. Borodin, G.O. Roberts, J.S. Rosenthal, and P. Tsaparas, "LinkAnalysis Ranking: Algorithms, Theory, and Experiments," ACM Trans. Internet Technology, vol. 5, no. 1, pp. 231-297, 2005.

[4] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," technical report, Microsoft Research, 1998.

[5] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of Trust and Distrust," Proc. 13th Int'l Conf. World Wide Web(WWW), 2004.

[6] B. Amento, L.G. Terveen, and W.C. Hill, "Does 'Authority' Mean Quality? Predicting Expert Quality Ratings of Web Documents," Proc. ACM SIGIR '00, July 2000.

[7] Anonymous, Net users distrust corporate privacy policiesstudy, available at http://www.newsbytes.com/cgibin/udt/im..ble?client.id newsbytesandstory.id – 174596

[8] Ba, S., Establishing online trust through a community responsibility system. Decision Support Systems, Vol. 31, pp. 323–336, Feb 2002.

[9] Fang Liu, Clement Yu, Weiyi Meng, Personalized Web Search for Improving Retrieval Effectiveness, IEEE transactions on knowledge and data engineering, Vol. 16 (1), January 2004.

[10]Christian Bizera, Richard Cyganiakb, Quality-driven information filtering using the WIQA policy framework,

[11] R. Cooley, B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the World Wide Web. In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997.

[12] Sigmoid Function from Wolfram MathWorld, http://mathworld. wolfram.com/SigmoidFunction.html, 2008.

[13] R.Y. Wang and D.M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," J. Management Information Systems, vol. 12, no. 4, pp. 5-34, 1997.

[14] http://www.sciencedaily.com/articles/w/web_crawler.htm

[15] https://code.google.com/p/crawler-commons/

[16]https://code.google.com/p/ldspider/