Volume 3, No. 4, April 2012
Journal of Global Research in Computer Science
RESEARCH PAPER
Available Online at www.jgrcs.info
ISSN-2229-371X

RESULT DOWNLOAD MANAGER AND ANALYZER SYSTEM

Atul Kumar[*1], Vikas Tyagi[2], Mrinal Trivedi[3], Krishna Kumar Mishra[4] and Mukul Kumar[5]
*[1] Department Of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India
atul0411@yahoo.com[1]
[2, 5] Department Of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India
[2]itengg.vikas@gmail.com, [3]mrinaltrivs27@hotmail.com, [4]krrish.1212@gmail.com, [5]mukul.kvarshney@yahoo.in

Abstract*:* Result Download Manager and Analyzer System (RDMAS) is a real world application utility that aims to download the University results and per-forms various analysis over them. This utility seems to be very useful as it will reduce a lot of manual work of result analysis and its storage .It will provide functionality of downloading the complete result of any number of students by extracting the data from html page of result available on University web-site and storing them into our local database. It will also facilitate to download the results of selected range of roll numbers on a single click of a button. After downloading the results a complete analysis of results can be performed by using the queries provided within the utility. The comparison of results can be done as per the requirements of user.

INTRODUCTION

Downloading the results for few students from the University website in .htm format is not very complicated task and we can easily analyze the results if the number of student is small .But when number of students getting increasing day by day it is going to be a very hectic task to download the results and analyze them manually. So the macro functionality of MS Excel is used by various colleges for purpose of downloading and analyzing them which reduces a lot of manual work. Macros allow the automation of repetitive tasks such as printing, for-matting, configuring, or otherwise manipulating data in Excel [1].

As technologies advancing at very alarming rate due to which threats for data security becomes very common activity. So considering the data security various Universities had implemented some security constraints. This results in improper functioning of macro utility of MS Excel. This drawback of macro function leads to develop such a real world application which should be an automation of macro of MS Excel and can perform all the functionalities in much efficient manner with same level of accuracy.

The Result Download Manager and Analyzer System is a real world application developed to overcome the drawbacks of macro of MS Excel. This utility works according to new security constraints of University for the data security. This aims to download the University results of students on a single click and perform different analysis over it like comparison of internal and external marks of two students, topper of the branch, average number of marks scored in a particular subject and many more. Here results are downloaded from the html page of student result which is available on University website by crawling the html page and copy the content of that page into our local database which is used for result analysis purpose. The phenomenon of crawling is performed by using Standard DOM Parser designed for PHP (Preprocessor Hypertext). This DOM extension of PHP is enabled by default [2].

Standard DOM Parser:

The Document Object Model (DOM) Parser is a cross-platform and language-independent convention for representing and interacting with objects in HTML, XHTML and XML documents. Aspects of the DOM (such as its "Elements") may be addressed and manipulated within the syntax of the programming language in use. The public interface of a DOM is specified in its application programming interface (API) [3].

A html DOM parser written in PHP5+ let us to manipulate html in very easy way. It requires PHP5+, supports invalid html. Finds tags in html page with selectors just like jQuery. It extract contents from html in a single line [4].

In order to analyze a web page for content extraction, the page is passed through an HTML parser that creates a Document Object Mod-el tree. The algorithm begins by starting at the root node of the DOM tree (the <HTML> tag), and proceeds by parsing through its children using a recursive depth first search function [5].

One of the ways is to segment the web pages into *zones* based on its HTML structure. Once these zones are identified, attribute based analysis of the content can be carried out [6].

```
//create DOM from URL or file
$html= file_get_html('http://www.google.com/');
//dump contents (without tags) from html
Echo file_get_html('http://www.google.com/')->plaintext;
```

RELATED WORK DONE

Before implementation of new security constraints by different universities to protect their data from being hacked by some unauthorized access, results are being downloaded by using Macro of MS Excel and stored in MS Excel and then statistical analysis is performed.

For downloading results from University website first we record the Macro for a single roll number. Now we have result of single student in our database. After this we format the downloaded result in a specified manner according to our needs. Now we stop the recording of macro and go to the source code of macro which is programmed in Visual Basic. We insert the range of roll numbers in the visual basic code for which we have to download the results.

Now start the Macro again by copying and paste the URL of the web-page of result from University website to the Macro and complete result get downloaded and stored in our database in specified format.

Now by using the features of MS Excel statistical analysis is performed and results of analysis are displayed in graphical forms.
For example-

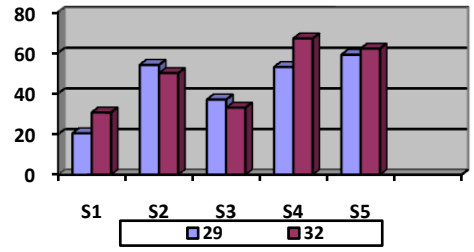

Figure 1 Statistical analysis of marks

PROPOSED WORK

Methodology Used The Result Download Manager and Analyzer System perform its task by first accessing the specific URL for a given student's roll number. On accessing this URL DOM parser goes through the background HTML content of the desired page from which result has to be extracted. PHP's DOM extension reads in the file and creates a walkable object tree in memory. Starting with the document or an element (called nodes in the DOM) of a document we can get or set the children, parents, and text content of each part of the tree [7].

Now, the RDMAS identifies and extracts the required data elements using background scripts from the DOM created earlier. The scripting language used here is PHP (Hypertext Preprocessor). The DOM contains the entire data collection in the form of an array. The extraction logic is entirely dependent on PHP's foreach () loop that helps in easy navigation over Data Object to identify any html tag.
foreach($object_name->find('html_tag _name') as $key => $info){}

After extraction, the data extracted using the Data object is stored in local database. Now, the Data Object is unset for the current roll number and a new object is created for the next roll number.

Now, we are ready with the data to be analyzed. Based on the standard SQL queries various analysis is performed over the data being collected just with a single button click. A number of basic analysis options are provided by the utility like `retrieving the roll number of overall topper`, `retrieving average marks obtained in a specific subject`, `retrieving a comparison of external versus internal marks obtained`, ` obtaining the list of students whose marks are below/above a specified percentage of marks`, and the likes.

Besides, the above mentioned features the RDMAS supports a number of other features some of which are:
Security of administrator account whose password is stored as an MD5 hash.

Security of data by not allowing the user to directly access the original page loaded on server just by accessing the parent/root directory.

Robustness against situations when a given HTML page does not exist or the server is overloaded by providing a timeout deadline.

ALGORITHMIC STEPS

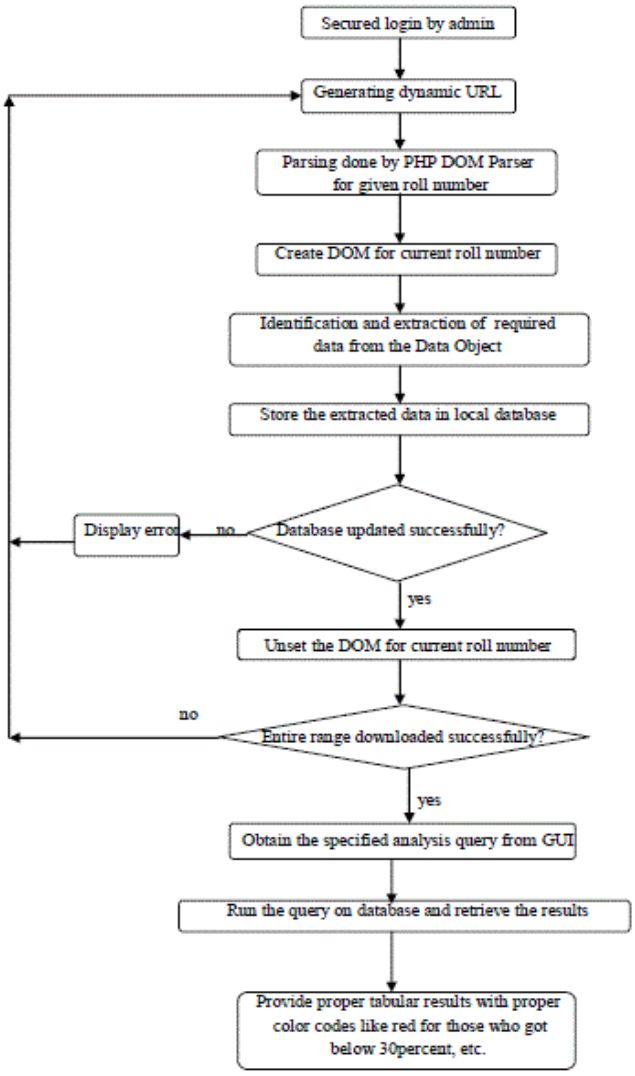

Figure: 2

LIMITATIONS

This utility is specifically designed for the pattern of UPTU only but it can be further extended for other University examination pattern.

This utility cannot tolerate very slow internet connection because of timeout of 30 seconds after which DOM Parser will get expired. But this timeout is implemented because for various security constraints.(For accurate measurement please view the page setup of sample paper)

CONCLUSION

The key motive to publish this paper is to give a glimpse of understanding the "Result Download Manager and Analyzer System" as new concept of downloading and analyzing the University results with enhanced functionalities as compared to other download systems. It gives a new aspect to automate the downloading and result analysis procedures at various college levels. It reduces a lot of manual work and time. Here we can also perform the comparison of results between various colleges, departments, sections. It becomes easy to retrieve the list of subject toppers, branch toppers, external toppers, average marks scored in each subject etc. moreover it provides results of analysis in graphical forms which are very useful.

ACKNOWLEDEGEMENT

We feel it our principal obligation to thank all our colleagues who helped us in identifying the need to work in this area. We are also thankful to all those who appreciated this work, gave proper guidance and provided us a boost to think and to work more effectively in this area of automation.

REFERENCES

[1]. Collins, J. C. "Microsoft Excel Functions, Macros and Data Commands". Atlanta, Georgia: ASA Research.

[2]. manual/en/book.dom.php.(n.d.). Retrieved from http://www.php.net: php.net

[3]. wiki/Document_object_model. (2001, january 15). (Wikimedia Foundation) Retrieved from http://en.wikipedia.org: wikipedia.org.

[4]. http://simplehtmldom.sourceforge.net/

[5]. Suchit Gupta, G. K. (2003). "DOM-based Content Extraction of HTML Documents,". Budapest, Hungary.

[6]. A. F. R. Rahman, H. A. (2001). "Content Extraction from HTML Documents". 1st Int. Workshop on Web Document Analysis (WDA2001), (p. 4).

[7]. Tim Converse, J. P. PHP5 and MySQL Bible.John Wiley and Sons.

Short Biodata of All the Authors

 Atul Kumar Associate Professor, Computer Science and Engineering Department, IMS Engineering College Ghaziabad, India, Mobile- +91931381482

 Vikas Tyagi Assistant Professor, Computer Science and Engineering Department, IMS Engineering College Ghaziabad, India, Mobile- +919313061237

 Mrinal Trivedi B.Tech , Final Year, Computer science and Engineering Department, IMS Engineering College Ghaziabad, India, Mobile- +919451505855

 Krishna Kumar Mishra B.Tech , Final Year, Computer Science and Engineering Department, IMS Engineering College Ghaziabad, India, Mobile- +919457670753, +918791871874

 Mukul Kumar B.Tech , Final Year, Computer Science and Engineering Department, IMS Engineering College Ghaziabad, India Mobile- +919650227750

