REAL TIME PRICE COMPARING SYSTEM FOR ROOM RESERVATION: A WEB CONTENT MINING APPROACH

M.J.Ahamed Sabani[1] S.L.Roshan Rukshana[2]

[1]Department of ICT, South Eastern University of Sri Lanka, Sri Lanka
[2]Faculty of Arts and Culture, South Eastern University of Sri Lanka, Sri Lanka

[1]mjasabani@seu.ac.lk, [2]roshanslr93@gmail.com

ABSTRACT: *Online reservation systems provide facility to find rooms in hotels via internet. But all users don't get full satisfaction, because the availability and rate of rooms have possibility of variation between hotels and between the third party Reservation System and the hotels' own Reservation System sites. The main reasons would be the third party's additional charge or the databases are not in up-to-date. In other hand hotel owners have headache of giving privileges to third party Reservation System to access their database for advertise and make reservation. As the solutions this prototypical system has the capability of comparing room rates of different hotel Reservation System web sites without accessing their databases. It fetches information about available rooms in different hotels directly from their own Reservation System sites on real-time. Web content mining approach with PHP (cURL library), JavaScript and other web development techniques were used and will give comparable output to the customers*.

Keywords: web content mining, online reservation system, PHP-cURL, JavaScript

1. INTRODUCTION

Nowadays there are several systems which supplying online reservations to hotels, airlines, apartments, inbound tour operators, travel portal websites and others selling accommodation over the Internet.

According to (Richard Bemile, Akwasi Achampong, et al, 2014) A hotel reservation system, commonly known as a central reservation system (CRS) is a computerized system that stores and distributes information of a hotel, resort or other lodging facilities.

If a customer needs to find desirable hotel without using any reservation agencies, he/she has to go through all the own web sites of firm to compare and find the best matches. It is very difficult to compare prices using different views and will take more time. As well as finding a suitable hotel using a travel agencies or hotel booking agencies, almost all of them are not much efficient. Prices in reservation system might vary from the real prices on their own reservation website. And some reservation agencies shows limited number of available rooms categories.

At the same time hotel owners facing a problem that they need to give privileges to access their database to reservation agencies. Or else when reservation agencies keep a backup of database, it may be vary from real time information. At most demanded solution for these problems, this study supposed to develop a prototypical web based system having the capability of comparing prices of different hotel web sites on real-time using the technique of web content mining without accessing/ using any database.

Web mining is a multi-disciplinary effort that draws techniques from fields like information retrieval, statistics, machine learning, natural language processing, and others. Web content mining is a sub-area of web mining which is application of Data Mining techniques to unstructured or semi-structured text, typically HTML-documents (Claudia Elena Dinucă and Dumitru Ciobanu, 2012).

A hotel's Reservation System website provides data to access by all over the world. For this prototypical system we have included some selected hotels in Sri Lanka which have their own Reservation System. By using this proposed system users can full fill their expectation from single site on a short time period. At the same time hotel owners do not need to give database privileges to the system.

2. OBJECTIVES

The main objective of this system is to compare and show room prices of different hotels without accessing their own databases directly. Through this main objective this system will achieve the following goals
- Save users time and ignore unwanted risk by comparing prices of different hotel Reservation System web sites using a single web site.
- Reduce the hotel owner's risk about their database privileges.

3. LITERATURE REVIEW

In the arena of global competition, organizations in all over the world are competing through the use of the most comprehensive and advanced technological feature. Various industries use the information technologies and the advancements of software and internet to maintain and monitor their business transactions. In the application of the informative systems, reservation systems take a leading position. The airline and hotel industries are the most common users of such system. The purpose of the application of Reservation System is to easily manage and organize all the reservations and bookings of the clients and gain the competitive advantage.

3.1 Online Reservation System

The word "reservation" means that the written record or promises of an arrangement by which accommodations are secured in advance (WordWeb ver.8, 2016). Online system is any electronic interactive system that delivers information to users via telephone lines to personal computers or via cables to terminals. Such a service provides information, usually in text form, about news, education, business, entertainment, shopping, and more. Some also provide message services and graphic

and audio information (Encyclopedia Britannica, 2009). Online hotel reservations are a popular method for booking hotel rooms. Travelers can book rooms on a computer by using online security to protect their privacy and financial information and by using several online travel agents to compare prices and facilities at different hotels (retrieved from Wikipedia, 2017). The room rate has a significantly negative effect on the average number of online bookings, and that hotels in larger cities tend to receive more online bookings (Qiang, Y. et al, 2010).

3.2 Available approaches

There are several systems which supplying an online Reservation System to hotels, airlines, apartments, inbound tour operators, travel portal websites and others selling accommodation over the internet. But almost all of them are not much efficient. Users may have to go through all the hotel web sites to find out the actual prices and best price matches.

Expedia.com is a travel website owned by Expedia Inc. which can be used to book airline tickets, hotel reservations, car rentals, cruises, vacation packages and various attractions and services via the internet or telephone agents. The site uses multiple global distribution systems like Amadeus or the Sabre reservation systems for flights and for hotels, Worldspan and Pegasus, along with its own hotel reservation system for contracted, bulk-rate reservations (retrieved from Wikipedia, 2017). When user request, many Reservation System will load all details, but not only the required information. It makes complication to take decisions. And some Reservation System do not compare prices while showing all prices on the web pages. Also there might be some differences between the prices on hotel's own Reservation System sites and the prices shown on these travel agents' site. It might be the cause of service charge or the deviation of database records. Some travel agent site shows less information such limited number of rooms while showing more rooms on real site.

Following figure1 and figure 2 clearly illustrate the price different between one of the popular travel agent's site and the hotel's own site for same requirement. Also travel agent site shown only eight types of rooms, while hotel Reservation System site showing fourteen types of rooms.

Figure 1: Rooms price list of a hotel in a popular Reservation System site

Figure 2: Rooms price list of hotel's own Reservation System site

The data mining is defined as the process of discovering useful patterns or knowledge from data repositories such as in the form of databases, texts, images, the Web, etc.(Malarvizhi R., Saraswathi. K.,2013). The data mining, especially web content mining technics will be applied to find and filter necessary information such room category and prices from the fetched web pages without accessing their databases. Web structure mining mines the structures like HTML or XML tags and gets information from the actual organization of the page.(Faustina Johnson, Santosh Kumar Gupta, 2012) Using the technics in PHP, HTML, XML and JSON, the information can be identified and fetched

4. METHODOLOGY

Water fall model was used to develop the system since the requirement is well defined at the beginning of this study. The waterfall model is a popular model of the systems development life cycle model for software engineering. This is a web based client-server side system and it runs with online internet connection. This system was developed using server side

scripting language PHP for server side development and JavaScript and CSS for web design.

4.1 Technology adopted

Server-side programming languages are scripts that are executed on the server, and are then translated into HyperText Markup Language (HTML) which can be viewed by all web browsers. The two most popular server-side scripting languages are PHP: Hypertext Processor and Active Server Pages (ASP). Additionally, there are numerous other languages like AJAX and Coldfusion (retrieved from Cyberindian.com, 2017). PHP can run on both UNIX and Windows servers, which makes it more accessible than its Windows counterpart, Active Server Pages (ASP). PHP- Hypertext Preprocessor is a widely used, server side scripting language was chosen to develop this system.

PHP has a very powerful library libcurl that specifically designed to fetch data safely from remote web sites. Libcurl is a library created by Daniel Stenberg that allows to connect and communicate too many different types of servers with many different types of protocols. cURL is a command line tool for getting or sending files using URL syntax. It supports the http, https, ftp, gopher, telnet, dict, file, and ldap protocols. libcurl also supports HTTPS certificates, HTTP POST, HTTP PUT, FTP uploading (this can also be done with PHP's ftp extension), HTTP form based upload, proxies, cookies, and user-password authentication. This library was used to fetch needed information from hotel's own Reservation System web site as web content mining approach.

It is important to have the right tools to filter needed information from the whole page that has been retrieved. PHP Simple HTML DOM Parser was used which a library is written in PHP that gives dictator like control over HTML and let manipulate HTML in a very easy way. It finds tags on an HTML page with selectors just like JQuery and extract contents from HTML in a single line as need. In this system this library was included to filter out necessary information.

Firebug is a most popular and powerful web development tool which can be integrated with Firefox to put a wealth of web development tools at your fingertips while you browse. You can edit, debug, and monitor CSS, HTML, and JavaScript live in any web page (retrieved from Firebug, 2017). Required data from retrieved web site were analyzed and filtered out using PHP simple DOM parser with the aid of firebug tool which installed on Firefox browser.

HttpFox is an add-on for Mozilla based browsers that monitors and analyzes all incoming and outgoing HTTP traffic between the browser and

the web servers (retrieved from httpfox, 2017). This add-on shows following information of each request:

- Request and response headers
- Sent and received cookies
- Query string parameters
- POST parameters
- Response body

Every website sends request as query string parameters which gets manipulated in between and most times end up appending values which are really not required. This certainly adds to the GET or POST overhead. HttpFox helps to know different page requests and what all pages get loaded and parameters upended to it. Hotel page submission method, parameters that passed with the request, redirections and etc. were observed using this tool to send request from the system to retrieve the required data from the hotel's Reservation System web sites according to user requirement.

JavaScript is a prototype-based scripting language that is dynamic, weakly typed and has first-class functions. It is a multi-paradigm language, supporting object-oriented, imperative, and functional programming styles. jQuery is a cross-browser JavaScript library designed to simplify the client-side scripting of HTML. It simplifies HTML document traversing, event handling, animating, and Ajax interactions for rapid web development. jQuery is designed to change the way that you write JavaScript.

JSON decoding was used when filtering out some web pages that contain required data in JSON format. JSON or JavaScript Object Notation is a lightweight text-based open standard designed for human-readable data interchange. It is derived from the JavaScript scripting language for representing simple data structures and associative arrays, called objects.

4.2 Approaches

This is a web based system which is comparing prices of available rooms from retrieved web pages from hotel's own Reservation System websites by passing parameters given by the user to multiple Reservation System websites. As a result it gives compared prices in a user friendly manner. As a web content mining approach, PHP-cURL method was used to retrieve the required information from web pages. Because it is a powerful library for work with third parties website using queries. Normally web pages contain lots of information that not needed for analyzing. PHP simple DOM parser library and JSON decoding were used to filter required information from retrieved web pages as needed. Look and feel

of user interface was developed using HTML, JavaScript and CSS to attract the user and facilitate their search. JavaScript was also used to validate the forms and fields in this system. Moreover Firebug and Httpfox tools were used to observe the request and parameter passing and POST / GET methods request to create cURL request. Also these tools were used in the filtering phase to analyze the source code of retrieved web pages.

5. ANALYZE AND DESIGN

Requirement analysis is the process of determining user expectations for a new or modified product. In software development, requirements are often called functional specifications which has to be fulfilled to get a complete solution. Considering the stated problems, there are three main functional requirements should be satisfied by the implementation.

1. Get user requirement through a web interface.
2. Mine the required web content from different relevant Reservation System web sites.
3. Analyze and show rates of available rooms.

Figure 3 shows the top level diagram of the system. It describes the path for passing requests and getting responses.
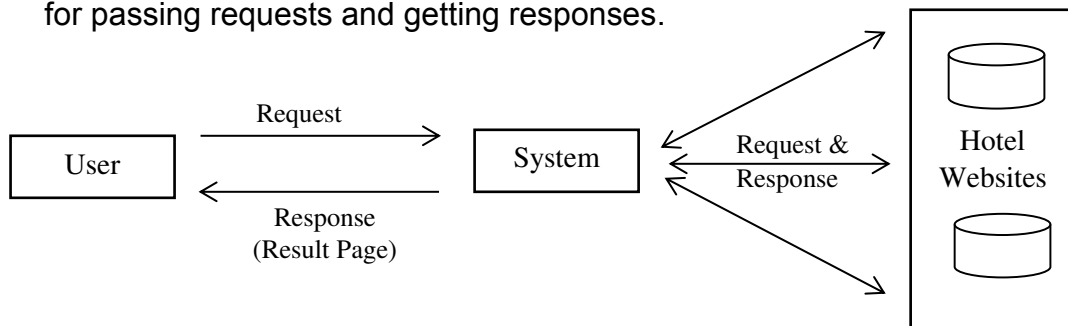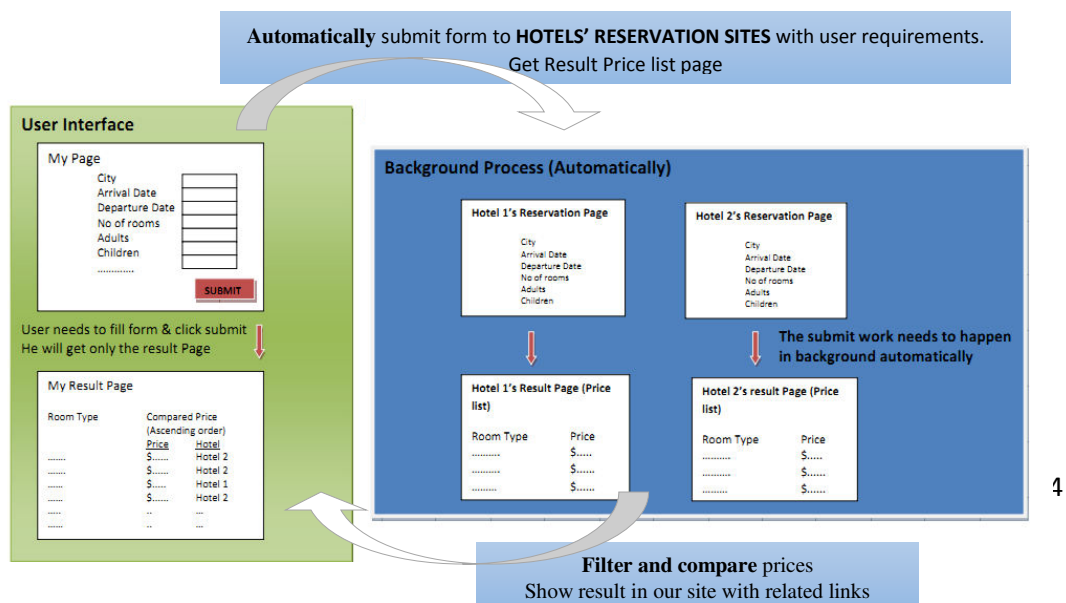

Figure 3: The top level diagram of the system

The whole processes can be divided into two areas such user interface and background process. In user interface users need to provide their requirement then the analyzed and compared results will be shown in same interface according to their requirement. Background processes are

4

the major part of this system. The system needs to pass requirement parameters to each Reservation System of hotels included in the system. Using the parameters system should produce prices of available rooms with some customization options to the users. Figure 4 illustrates the full model of the system in detail.

Figure 4: Detailed Diagram of the system

6. IMPLEMENTATION

The system implementation consists of following areas.

- Preparing environment for the development.
- Web user interface development.
- Development of main Reservation System price comparing system.
- System validation.

PHP server side scripting language was used to implement this system. Apache server was selected as the platform to run PHP scripts locally. The WAMP server which has PHP and Apache server was used to implement and test the system. Even WAMP server has already configured and ready to run PHP scripts, some additional configuration such php_curl and php_openssl extensions were needed to enable. Firebug and httpfox are add-ons that was integrated with firefox to observe the request and parameter passing and cURL requests made by web sites.

The web user interface was developed using Html, JavaScript and CSS. The main client side part of this system interface is the form to get user requirement. The form to get required information such city, check in and check out dates, number of rooms, number of children and adults in each room and other sorting options were designed using some form elements such date picker, drop-down list and radio buttons to avoid incorrect input values.

The main web content mining functional phase was implemented using PHP and JavaScript. The user preferences for search were needed to pass to all included Reservation System of hotels to get information about available rooms and its rates. And these input requirements were submitted in POST form submission method.

Before start to create cURL request function, the request passing method, parameters and the response for that request of each included hotel Reservation System sites were observed using httpfox tool to create cURL query. These parameters were varied between Reservation System web sites. The following things were identified when observing request and responses;
- Result page may receive data in GET method or POST method or using both methods
- Many request were done in a sequence.
- Sometimes many number of redirections were happened

cURL function and query string were created carefully by considering above mentioned factors for each hotel Reservation System site.

The function get_url_contents($url) was created to retrieve whole contents of result page using cURL query request by passing URL string. The URL string was created according to the observation made using httpfox tool. This function was used only for the hotel Reservation System site which use GET method only.

```php
function get_url_contents($url, $opts = array()) {
    $urls = (array) $url;
    $crl = curl_init();
    $timeout = 5;
    $useragent = $_SERVER['HTTP_USER_AGENT'];
    curl_setopt($crl, CURLOPT_USERAGENT, $useragent);
    curl_setopt($crl, CURLOPT_FOLLOWLOCATION, TRUE);
    curl_setopt($crl, CURLOPT_RETURNTRANSFER, 1);
    curl_setopt($crl, CURLOPT_CONNECTTIMEOUT, $timeout);
```

This function was modified using curl_setopt() function with different parameters to create different cURL request for different situations mentioned above.

For the sites use POST method or both GET and POST methods, this function

```php
if($method == 'post'){
    curl_setopt($crl, CURLOPT_POST, true);
    curl_setopt($crl, CURLOPT_POSTFIELDS, $data);
```

was changed to include parameter for method and relevant contents.

When more than one request needed, cURL was executed many times

and for redirection the maximum redirection time was set to curl_setopt()

```php
$nexturl= curl_getinfo($crl,CURLINFO_EFFECTIVE_URL);

curl_setopt ($crl, CURLOPT_MAXREDIRS, 10);
```

function.

curl_getinfo() function was used to get information such received URL, whether redirection is occurred or not, uploaded size, downloaded size, etc. for each request. The URL string was created with users' parameter according to the observed pattern using httpfox.

A URL string used to create cURL request with GET method as follows,

```php
14  $html = get_url_contents(
    "http://www.ramada.com/Ramada/control/Booking/check_avail?areaCode=&brandCode=RA,
    nation={$vals['city']}&state=NA&country=LK&checkInDate={$vals['checkInRamada']}&n
    &checkOutDate={$vals['checkOutRamada']}&numberChildren=0&numberBigChildren={$vals
```

And it was different for POST method,

The retrieved hotel RS's result pages using cURL request were filtered to get only required information. The PHP library calls PHP simple DOM parser was

```php
$tajParams =
"rate_text=-1&timeZoneFullName=Asia/Colombo&ccpSelectCountry=LK-XX&ccpSelectCity=CMB001&ccpSel
ectHotel=TJ-HLTCMBTS&calArrivalDateField={$vals['arDate']}&calDepartureDateField=
{$vals['dpDate']}&arrivalDateField={$checkInTaj[0]}&departureDateField={$checkOutTaj[0]}
&prsRoomCount=2&occupancy={$occupancy}&prsNumChildren={$vals['$childT'][0]}&prsNumChildren=
{$vals['$childT'][1]}&{$prsRoomAge}
accountId=&rateAccessCode=&_flowExecutionKey=_c7330FF8F-AD10-20C4-98E9-79A118D4FA04_kA855A1B4-
29CC-2D16-CFA9-8AC7F1412727&_eventId=submit";

$html = get_url_contents_taj("https://www.yourreservation.net/ibe/02_taj/main.htm", 'post',
$tajParams);
```

used to filter data. Firebug was used to discover and inspect the tags and elements in retrieved html page. Xpath of an element could be observed using this tool. Different methods were used in different situation to filter data. In some situation descending selectors was used to filter data in the html based result page. In other some cases regular expressions and JSON decode were also used. This type of filtering was used when the required data available in JavaScript and in JSON format of retrieved web pages. Normally JavaScript does not have tags and elements like html.

The filtered data were stored in an array to analyze and compare. The functions compare_price_ASC(), compare_price_DES() and compare_hotel() were used to compare and sort results to display. The retrieved results were kept in cache using SESSION to sort and modify the result further by users. The time out was set to retrieve web pages again without use stored data in cache. Because the data that retrieved from the hotel websites may be differed with earlier one. When a user do search in two different tabs of same browser simultaneously with different requirement, the cache will keep last retrieved data. To ignore this problem, session with id was used to confirm the correct result. If cache has result for different request, then system will retrieve data again.

In addition to reservation check page, system has hotel page to provide more information about hotels and other related sources. The contact page contains Google map of the office location and email feature to send comments and feedbacks to the system administration. Reservation checking form and the feedback form were validated using JavaScript.

7. RESULTS AND DISCUSSION

Testing is an integral and important part of the system development process. This part of the process ensures that defects are recognized as early as possible. The system was tested online to check and evaluate each functional requirement. The form submission was tested to make sure the correct parameter pass using httpfox.

Finally the whole system was tested by comparing the result with the real hotels' Reservation System web site. The results were compared and tested by giving same user requirement in both developed system and related hotel Reservation System sites. The obtained results from developed system and relevant hotel Reservation System sites were matched without any conflicts. The following figure 5 and figure 6 are show the results of developed system and from one of the hotel Reservation System site.

Hotel	Room Type	Price(per Room/per Day)	Total Price
Hotel Ramada	2 Double Beds Room	153.06	612.24
Hotel Ramada	2 Double Beds Room	153.06	612.24
Hotel Ramada	1 King Bed Room	159.34	637.34
Hotel Ramada	1 King Bed Room	159.34	637.34
Hilton Colombo	King Room-best Available Rate	228.50	914.00
Hilton Colombo	Twin Room-best Available Rate	228.50	914.00
Hotel Taji	Superior Room Twin Bed City View	230.00	920.00

Figure 5: Result page of the system

Figure 6: Result page of real hotel RS

Other hotel's results were also compared with this system. From the comparison and observation, the conclusion was made that the system's main functionality has been working correctly. The sorting criteria and limit options were verified. And comment and email part were also tested by sending email to a Gmail address.

8. CONCLUSION

This system was developed with user friendly interfaces to access easily. It gives solution to both customers and hotel owners. User can save their valuable time and avoid taking unwanted risk to check and compare available rooms to reserve. Also they can find the availability with the correct price in real-time. In other hand, from the hotel owners' side of view it gives solution for their headache of giving database privileges to third party RS. This system does not access any database directly. Using web content mining approach it accesses only the hotel websites that are available to access all over the world via the internet.

For each hotel PHP cURL function needs to be used in different ways. Therefor when including every hotel in to this system, it needs to be analyzed the request and create cURL requests carefully in certain way. However as a prototypical system, it has achieved the functionalities and objectives for included hotels in the system.

9. REFERENCES

Malarvizhi R., Saraswathi. K.,International, Web Content Mining Techniques Tools & Algorithms – A Comprehensive Study, Journal of Computer Trends and Technology (IJCTT), vol. 4 , no. 8, August 2013

Claudia Elena Dinucă, Dumitru Ciobanu, Web Content Mining, Annals of The University of Petroşani, Economics, Vol. Xii - Part I, 2012

WordWeb (Version 8), [Computer Software]. (2016).

Online System in Encyclopedia Britannica 2009 Student and Home Edition, (Version 2009.00.00.000000000) [Computer Software]. (2009).

Online Hotel Reservations in Wikipedia, Retrieved March 19, 2017, from https://en.wikipedia.org/wiki/Online_hotel_reservations

Qiang, Y., Rob, L., Bin, G.& Wei, C. (2010). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Computers in Human Behavior

Expedia (Website) in Wikipedia, Retrieved April 06, 2017, from https://en.wikipedia.org/wiki/Expedia_(website)

Faustina Johnson, Santosh Kumar Gupta,Web Content Mining Techniques: A Survey, International Journal of Computer Applications, vol 47, no.11, June 2012

Richard Bemile, Akwasi Achampong and Emmanuel Danquah, Online Hotel Reservation System, International Journal of Innovative Science, Engineering & Technology, vol. 1, no. 9, November 2014

Server side programming languages, Retrieved January 11, 2017, from http://www.cyberindian.com/web-designing/web-development/server-side-web-programming-languages.php

What is firebug?, Retrieved January 15, 2017, from http://getfirebug.com/whatisfirebug

httpfox, Retrieved January 15, 2017, from https://code.google.com/archive/p/httpfox/