The Case Study for Price Extracting of Mobile Phone Sell Online

Quanyin Zhu, Yunyang Yan, Jin Ding and Jin Qian
Faculty of Computer Engineering
Huaiyin Institute of Technology
Huaian, Jiangsu Province, China
hyitzqy@126.com

Abstract—The commodities price is a very important data for shopkeepers of shop online. By using the Web mining can get more and more data in everywhere such as e-supermarkets and e-commerce. Many theories and algorithms are reported for Web mining. This paper shows a case study for the price extracting of mobile phone of shop online. The MVC design model is opted to build the application system. The opened developing tools: PHP language and MySQL database are used to code the system program and construct the system database respectively. The system functions are described in detailed. The functions coded by PHP are designed for implantation the price extracting of mobile phone, comparative analyzing, and system management. Experiment demonstrates its performance and proves price data is meaningful and useful for the shopkeepers selling online.

Keywords- price extracting; mobile phone; shop online; case study

I. INTRODUCTION

Since the early 1970s, Decision Support System (DSS) concept was put forward, and developed rapidly. Many researchers focus on DSS research, and it has gradually by management science, economics, applied mathematics, engineering technology, information science and other important areas [1]. DSS architecture includes three parts: the database, the model base and user interface [2].

Data Mining, also known as data extracting, is extract hidden in which people do not know beforehand but is potentially useful information and knowledge of the process from a lot of incomplete, noisy, fuzzy, random data. Internet data mining is a personal, business, website extract available information resources from the Website, according to their different purposes and characteristics, extract relevant data and implicit in the rules and knowledge then handed further access the process of the data [3]. Data warehouse is subject-oriented, integrated, non-volatile and the time-variant for the management and decision-making data sets [4]. It is different from traditional databases is that data warehouse is a database service in the high-level decision-making, which not only collect, organize and store a large number of dispersed geographically with the constructed data from different sources of information, but also through the history of these data processing and transformation, capture a series of data for decision analysis, and these data can provide users with better decision support [5].

As we know, the Web mining can be used to discover the news [6], analyze email communications [7], and require skill sets for computing Jobs [8] and even to latent topics from web sites of terrorists or extremists [9] and so on. When shopkeepers want to know more information about the customer and recommend the new products to them using e-mails, but the most important data is the commodities price. So, our group designed a Web price extracting system of mobile phone for the shopkeeper’s online shop.

II. SYSTEM DESIGN

A. System Infrastructure

The application system includes three basic entries, which are Web server, Database Server, and Client. Figure 1 shows the system infrastructure.

Figure 1  System infrastructure

B. Database

In the application system, there are six data tables: db_user is used to store the users’ information; db_website data table is used to store the Web sites’ information; db_brand data table is used to store each mobile phone’s brand information; db_breakpoint data table is used to store each mobile phone’s brand information; db_bigclass data table is used to store the mobile phone belong which classes’ information; and db_product data table is used to store each mobile phone’s information. The relationship of some of them show as in the Figure 2.

C. Flow Chart

After user login in the system and select the data mining model. If the user input the key word “cell phone”, then should
to select the brand of cell phone which you want to extract it, you can select one brand or more than one brand. In the process of extract price, system will detect the network environment automatic at any time. After system access a webpage, system will record the URL into the database. If network interrupt at this time, the user need to refresh the page, and the application system can find the last URL before the network interrupt to extract information continue. Figure 3 show as the flow chart for data mining process.

### D. Extracting Based on Participle

There are two methods for data extract of Web mining in a general way, one is regular expression and another is participle. The method of participle is used the principle of participle to extract data in the Website.

Let n represents the biggest matching character length; Length represents the character string length for extracted; w represents the character length for marched; y is the character string residual length.

First judge the length of string bigger than n or not, if smaller than n continue to participle, if not divide it again. Second confirm first n-1 world is existed in database, if true find the information, if not confirm first n-2 world is existed in database, if true find the information, and unconfirmed middle n-2 world is existed in database, it continuous so on until the string residual length equal zero. Figure 4 shows flow chart for data extract by participle.

### E. DFD

The extracted price data and the analyzed data can be taken as a statistics. The application system can be extracted price for any mobile phone depend on the shopkeeper requirement. The Data Flow Diagram (DFD) of data mining shows as Figure 5.

### F. Class Design

There are two methods to extract price data. So there are two kinds of class design. Figure 6 shows as the class of
III. MAIN FUNCTIONS DESIGN

The MVC architecture is used to develop the price extracting system of mobile phone. The main functions are more than forty; some of them are introduced as follow:

1) get_brand()
   Belong to: JingDong, TaoBao
   Function: Extract all the brand of cell phone in the website and insert into database.
   Achieve method: Determine URL of a website, for example http://search.360buy.com/Search?key=手机号码, invoke file_get_contents() to get the source code, invoke eregi() to find the string between string “brand” and “price”, then insert the information which match the regular express into the database.
   Return: Null.

2) get_nextpage($url)
   Belong to: JingDong, TaoBao
   Function: Get the URL of the next webpage of current webpage.
   Achieve method: invoke file_get_contents() to get the source code of current webpage, invoke eregi() to find the string between string “var_product_addTime” and “<!--filter end-- “, then find the hyperlink that there is a “nextpage” behind it.
   Return: URL of nextpage of current webpage.

3) get_product_url($url)
   Belong to: JingDong, TaoBao
   Function: Get the URL of all the product of current webpage.
   Achieve method: invoke file_get_contents() to get the source code of current webpage, invoke eregi() to find the string between string “<ul class="list-h clearfix”” and “<div class="in clearfix””, then find the hyperlink that there is a “nextpage” behind it.
   Return: URL of nextpage of current webpage.

4) get_info ($str)
   Belong to: JingDong, TaoBao
   Function: Get detail information of a product.
   Achieve method: invoke file_get_contents() to get the source code of current webpage, invoke eregi() to find the string between string “brand” and “type” find the brand information, invoke change_format() modify the information. Then find the information of type, net standard, shape, market time and so on.
   Return: An array include all the information of product.

5) checkproduct($type)
   Belong to: JingDong, TaoBao
   Function: Check the product extracted is existed in the database or not.
   Achieve method: select all the records that brand is same as $brand and type is same as $type, then calculate the count of record, if record >0 means that product is existed in the database, if not means the product is not existed in the database.
   Return: $rows>0 return false, $rows<0 return true.

6) search_url($url)
   Belong to: get_url
   Function: Find the URL of webpage that show the all product or contain the “nextpage”.
   Achieve method: invoke eregi() to find the webpage that contain the key word “nextpage” if find means current webpage is the basement, it can be the first webpage to start extract.
   Return: An array include all the information of product.

7) find_brand($str,$brand_array,$MaxLength)
   Belong to: get_url
   Function: Get all the brands in the current webpage.
   Achieve method: invoke geturldata() and convert_array() to find all the words in the current webpage, then invoke lessthan() and morethan() to find all the brands store in a array.
   Return: array of brands in current webpage.

8) find_price($beg,$fcontent)
   Belong to: get_url
   Function: Get the price of product.
   Achieve method: invoke file_get_contents() to get the source code of current webpage, invoke eregi() to find the string between string “<ul class="list-h clearflix"” and “<div class="in clearfix””, then find the hyperlink that there is a “nextpage” behind it.
   Return: URL of nextpage of current webpage.
9) **find_info ($data,$array,$para)**
   Belong to: newsina,newsohu
   Function: find the corresponding detail parameter of product check currently.
   Achieve method: find all the words from the array $data and can be found in the $array.
   Return: array contain all the detail parameter of product.

10) **goto_mobile_index ($str)**
    Belong to: newsina,newsohu
    Function: find level2 webpage from the $str.
    Achieve method: invoke preg_match() to find the hyperlink of key word “cell phone” from the $str and invoke geturldata() to get the source code of that webpage.
    Return: source code of the hyperlink of key word “cell phone”.

11) **find_money($i,$data)**
    Belong to: newsina,newsohu
    Function: find price of current cell phone
    Achieve method: find the string from that array $data that index is $i, then find the string that behind that string, if string after that string is null, go on to find, if not return string.
    Return: price of current cell phone.

IV. IMPLEMENTATION

We select the http://search.360buy.com, http://www.taobao.com and www.dangdang.com as the test Websites. The products name of mobile phone can be extracted and as the products database, and then the shopkeepers can choose their interesting to extract the price. The products database shows as Figure 8, and the price extracted of mobile phone selected shows as Figure 9.

![Figure 8 The product choosing](image)

![Figure 9 Price extracted of mobile phone](image)

V. CONCLUSION AND FUTURE WORK

Our team developed an application system using the case above. The PHP language is opted to code the system program, the two methods are all studied to compare the accuracy rate. The regular expression can be achieved 83% and it depend on the network status, but the participle can be reached more than 99% accuracy.

The market of mobile phone change very fast. So the price extracting is pay more attention by the shopkeepers. Our future interesting work is on the decision support system for the mobile phone sale online.

ACKNOWLEDGMENT

This work is supported by the Industry-University Collaboration Project of Huaiian City, China (HAC201002); the fund of Huaiian Industry Science and Technology, China(HAG2010064, HAG08009); and Qing Lan Project of Jiangsu, China.

REFERENCES