Web Information Retrieval For Gold and Silver Prices: A Design of A Wrapper Program

Norliza Katuk¹, and Afiq Aisamuddin Mohamad Sharif²

Universiti Utara Malaysia (UUM) Malaysia, ¹k.norliza@uum.edu.my; ²maxis5049876@yahoo.com

ABSTRACT

This paper presents a research on information retrieval for gold prices that intends to help users to get the information easily from the Internet. In gold trading, users usually visit more than one website to get information and compare the prices. Comparing prices from the individual websites is a time-consuming job because each website presents a lot of information about the respective companies as well as their products. Hence, users have to spend a lot of time eliminating unnecessary information. In information retrieval, users usually perform tasks like searching, filtering and organizing information in the Internet. The tasks can be automated by a program called wrapper. This research proposes a wrapper model that has been demonstrated in a mobile application called Gold-Trader Application (GTA). The application provides information about gold and silver prices to sellers and buyers. It helps the users to compare the prices from multiple traders more effectively. A usability evaluation on GTA was carried out and presented in this paper.

Keywords: web information retrieval, gold prices, wrapper programs, information gathering process

I INTRODUCTION

Gold and silver are classified as valuable commodities which many people invested into nowadays. Search engines over the Internet have become the medium to get the latest market price of these commodities. However, users are always overloaded with other information when search engines are used. This caused difficulty for them to identify relevant and accurate information that fits their needs (Bergamaschi, Guerra, & Leiba, 2010).

This paper presents a research on designing a tool known as a wrapper program that intends to help users to obtain information about gold and silver prices easier than the use of search engines. Section II of this paper gives a brief overview to wrapper programs and their roles in web information retrieval. This section also describes a wrapper model that is demonstrated in a system called Gold Trader Application (GTA). Section III explains the evaluation of GTA while the results are presented in Section IV. Section V concludes the paper.

II WRAPPER PROGRAMS FOR WEB INFORMATION RETRIEVAL

A. Wrapper Programs

Wrappers produce structured data that are ready for post-processing during information extraction process (Baumgartner, Henze, & Herzog, 2005; Chang, Kayed, Girgis, & Shaalan, 2006). A wrapper provides a query interface to access information from multiple web sources usually through pattern matching rules. It must be able to recognize the required data among many other non-required pieces of text (Gregg & Walczak, 2006). A wrapper can be developed manually using any programming languages or through the use of a wrapper induction system; a tool that automates the generation of wrappers. In this paper, we specifically focus our discussion on manually-generated wrappers. In addition, the paper also focuses on wrappers for site-specific rather than the generic type of it. Some examples of wrapper techniques were discussed by Simon and Lausen (Simon & Lausen, 2005). Other similar works can be found in (Badr, Merialdo, & Crescenzi, 2013; Younsi, Quafafou, Ouzegane, & Tari, 2013).

B. A Wrapper Model for Gold-Trader Application (GTA)

Gold Trader Application (GTA) is a web-based application that runs on Android devices. It was developed to demonstrate the proposed wrapper model. The main function of GTA is to allow users to view the latest gold and silver prices that were automatically gathered from different websites.

The application extracts and utilizes information from five gold and silver trader’s websites. Four of them are based in Malaysia, and the other one is an international company. The four websites that based in Malaysia were employed and the prices of their gold and silver were extracted. The international company was used for extracting the gold and silver spot price. All of the websites are as below:

i. World Islamic Mint
   (http://www.worldislamicmint.my)
   (http://gsr2u.com.my/sistem)
Presented into a structured table. The users can view the information and get the update every time they load the system.

All of these components are combined in a wrapper model. The wrapper runs on a web server to collect the information. The users can access the information via smartphones, computer tablets, and desktop that connected to the Internet. GTA application is accessible through http://gold-trader.skyrocket.my/. Figure 1 shows the wrapper model from the view of the overall architecture of GTA.

The overall process of the wrapper model is illustrated in Figure 2. The process consists of eight iterated activities. First, the wrapper identifies the trader’s Uniform Resource Locators (URLs) where gold and silver prices will be extracted from. Then, it extracts the web content of the specified URLs and stores the web content into a temporary file and converts them into strings. After that, the wrapper filters the strings for matching strings or keywords such as price, selling and buying and selects them. Then, it stores the matching selected string in an array. As some of the websites provide price in USD, the wrapper converts it to MYR for consistency. It is also important to note that the international gold spot price is in ounce (oz.) which is a different metric unit used in Malaysia. In Malaysia gold and silver are traded in grams (g). Hence, the wrapper performs conversion of the two metric units to get the local price. Then, the wrapper stores the local prices in variables and finally displays them in a table. The wrapper is set to allow GTA to provide up-to-date information; hence, the activities will be iterated every hour in automated mode.

GTA focuses on five important functions. First, users can view the live updates of gold and silver prices in graph as shown in Figure 3. The graph is set on a main page, so, users can see the daily trend of the price going up or down, just by looking at the graph. The application also allows the users to view the information of the gold and silver prices from a single page. Users can see the selling and buying prices for gold and silver, then, compare them side by side. Users are no longer required to open several websites to compare the prices as shown in Figure 4. The third function is to display the dealers’ contact information all around Malaysia. Another function is to link group of gold and silver marketplace in Facebook. The last function

v. Kitco (http://www.kitco.com/market/)

Besides, Bank Negara website was used as a reference to convert currencies from U.S. Dollar (USD) to Malaysian Ringgit (MYR).
provides a service to calculate zakat for gold and silver investment. Users must provide the amount of their current gold or silver savings, and then the application will calculate the total zakat automatically.

Figure 3. The GTA Mainpage.

Figure 4. The Interface for Gold Prices.

III EVALUATION

As explained in the previous section, GTA was developed to demonstrate the proposed wrapper model. The effectiveness and usability of the wrapper model are evaluated through an experimental study on real users (i.e., gold and silver traders). The following sub-sections describe the study in detail.

A. Objective and Method

The objective of the study was to investigate the usability of GTA. An online experimental study was conducted in November 2013 to meet the objective.

B. Participants

Invitation emails were sent to users of private groups in a Facebook, VIP PelaburanSilver.com, selected public gold dealers, college students and other participants that were interested in gold or silver. Hundred emails were sent out to invite participants to try out GTA and answer short questionnaire. However, only 23 participants responded to the invitation. All of the data were used in the following analysis. 17 (74%) male and 6 (26%) female participated in this study. More than half of the respondents were aged between 21 to 30 years, (i.e., 14 participants).

C. Materials

An online web-based questionnaire and GTA prototype were the materials used to facilitate the data collection of this usability study. The URL for online survey is http://gold-trader.skyrocket.my/questionnaire/.

The questionnaire has three sections and in between the sections, the participants experienced working with GTA prototype on a virtual smartphone (i.e., android emulator). The three sections consist of demographic information, respondent behaviours towards gold and silver, and usability.

On the demographic information section (i.e., Section A), the participants were asked about their age, gender, occupation, household income and their saving every month. For section B, which is about respondents’ behaviours towards gold and silver, the participants were asked about their interest and their enlightenment about gold and silver. How much they save, where they buy, how they get the latest price, and further about their interest in gold and silver. The last section (i.e., Section C) is about the usability of GTA. This section measured the usability of the application in three dimensions. Those three dimensions are: (i) Usefulness, (ii) Ease of Use and (iii) Satisfaction. These three dimensions were measured after all participants completed Section A, Section B, and browse GTA. The questionnaire was in a 5-point Likert scale, one represented ‘strongly disagree’ and five represented ‘strongly agree’. The questions were adapted from System Usability Scale (SUS)(Brooke, 1996; Sauro, 2011) which has been
validated in many other studies (Bangor, Kortum, & Miller, 2008).

Figure 5 shows the GTA application that the participants will see on the virtual smartphones. The virtual smartphone emulates the real smartphone and the participants will feel like they are using the application through their smartphone.

D. Procedure

The invitations of the survey were sent out to selected Facebook groups and to several participants that interested in buying gold and silver. They were supplied with the URL of the experimental study. The participants need to start the survey via web browser.

Before the participants agreed to start the survey, they were presented with an information sheet page. The page contained all of the information about the researchers, what the research is about, what the participant should do and how the data have been collected will be used in the analysis. After that, the participants were directed to the consent form page. Then, the participants were redirected to instruction page. The participants were briefed about the steps to complete the study.

After completing all of the introduction part, the participants were directed to Section A. In Section A, the participants were asked about their demographic information. Then, after completing Section A, the participants were redirected to section B on respondents’ behaviours towards gold and silver.

When both sections were completed, participants were redirected to the GTA prototype. The GTA prototype was displayed on a virtual smartphone. The reasons GTA prototype was presented on virtual smartphone is to let the participants feel the GTA application as a mobile application and to reduce the time for downloading and installing the application in their mobile devices.

After the interaction with GTA prototype, the participants were directed to section C that involved in evaluating the usability of the GTA prototype. At the end of the survey, the participants were redirected to thank you page.

IV RESULTS AND DISCUSSIONS

This section describes the results of the usability evaluation. The results are presented according to the components of the usability questionnaire.

A. Demographic Information

From the 23 participants, 19 (83%) of them were full-time employees and 4 (17%) were students. 52% of working participants were from private sector and the rest were government staff. 12 of them had monthly income between RM2000 to RM4999 and 3 of the participants had their monthly income more than RM10000. 17 respondents saved more than 25% of their income. Figure 6 presents the monthly savings graph of the participants.

B. Respondents’ Behaviors towards Gold and Silver

21 participants were interested in buying gold and silver; many of the participants buy gold jewellery more than other gold products. From the analysis, it shows that a total of 12 (52%) participants buy gold jewellery, 6 (26%) participants buy dinar and 3 (13%) participants never buy any kind of gold. In addition, silver jewellery, dirham and silver bar share the same total participants, which are 3 participants in each type. The other 14 (61%) participants never buy any kind of silver. The analysis also shows that the participants were more interested on saving or buying gold rather than silver. 20 participants possessed gold and only 10 participants possessed silver. The reason why the participants were not interested in silver is because they do not know the potential of the silver in a long term.
Gold has also been selling online. However, only 6 (26%) of the participants bought online, and the rest were buying through agents or stores. The famous product that participants buy online is gold jewellery (13%); followed by dinar (9%), gold bar (4%) and gold investment account (4%). Trader’s store (44%) is the top choice of place where the participants buy the precious metal; followed by online trading (30%) and then through agents (26%). Although many of the participants purchase gold and silver through the trader's store, most of them used the Internet to check the prices.

C. The Usability of GTA

17 (74%) of the participants never have a problem to get the information from the application. It means that the purpose of the application was successful. Only 6 (26%) of the participants experienced problems to get the intended information. The reason of this problem was due to difficulty to load the application. Hence they would have had a trouble to load the prices information. This usually happens if the participants use a slow Internet connection.

The participants’ responses on the questionnaire were recorded in a spreadsheet and analyzed using a statistical package; IBM SPSS Version 19. We first conducted a reliability test on the data. The test suggested that the Cronbach’s alpha coefficient for the thirteen items is 0.860 which indicates a high level of internal consistency of scales for the usability analysis.

We calculated the means and standard deviations for the usability items. Table 1 indicates the means and standard deviation for all of the Usability scale. The mean for usefulness, ease of use and satisfaction is 3.57, 3.61 and 3.76 respectively. Therefore, the total mean for the overall usability is 3.66 suggesting that GTA has an acceptable level of usability.

<table>
<thead>
<tr>
<th>Usability Scale</th>
<th>Number of Item</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>4</td>
<td>3.57</td>
<td>0.742</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>3</td>
<td>3.61</td>
<td>0.763</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>6</td>
<td>3.76</td>
<td>0.709</td>
</tr>
<tr>
<td>Overall</td>
<td>13</td>
<td>3.66</td>
<td>0.573</td>
</tr>
</tbody>
</table>

D. Discussions

The preceding sections described the wrapper model that has been demonstrated in GTA, and the usability evaluation of it. The results of the usability evaluation suggested that GTA has an acceptable level of usability with users’ satisfaction obtained the highest rank.

As mentioned earlier, the aim of the wrapper model was to simplify the process of searching, filtering, organizing and presenting gold and silver prices from the Internet. It also aimed to reduce the time that users spent on performing the information retrieval process. These two main objectives were partially achieved through the users’ evaluation.

The study suggested that the application is useful and easy to use, while giving users a substantial level of satisfaction. The obvious advantage of the system is that, it provides a single interface that combines the price information from different traders. This definitely eliminates many of the existing steps in the common information retrieval process. It also supplies users with the prices for further decision making process. The users can compare the price of different traders side by side without the need to open several websites. This helps the users in their decision making process more effectively because the application already eliminates the time they spent on searching, filtering and organize the information from the web.

The major limitation of the wrapper model is that it will unable to load price information if the structure of trader’s website is changed. This is because, the wrapper need to identify a pattern that have been redefined and rewritten if the website structure is changed. Despite that, it only affects the corresponding trader’s data, while the others are working well.

V CONCLUSION AND FUTURE WORKS

Site-specific wrapper programs are important for web information retrieval because it can simplify the processes that need to be done repeatedly to achieve the perfect and accurate information. The best wrapper program is the one that is customized for specific domains. By configuring and performing string pattern matching, it will help the users to get the right information easily.

The wrapper model that we proposed in this paper has undergone usability evaluation by gold and silver traders. For future research, we plan to study in depth on the string matching rules and evaluated the performance of the rules empirically. We are also interested on utilizing the wrapper model in other products and domain of study such as airline tickets, accommodation, groceries and cinema schedule. This could be an interesting research and beneficial to many people and users of the particular domains.
REFERENCES


