Design and Implementation of SMS Extraction and Analysis System

Yu-Yang WAN, Hui XU* ,and Rong-Rong FANG

College of computer science and technology, Nantong University, 226009,Nantong, China
** Corresponding Author: xu.h@ntu.edu.cn*

Abstract. With more and more applications of smart phone, more and more text messages in mobile phone, and then more and more information possibly contained in text messages. How to dig out useful information form SMS? This paper discusses techniques of the SMS extraction and analysis. Taking the bank SMS as example, key information is extracted to inform, is formatted to story in APP database, and then be analysed and statistic result shown in chart. The APP with the function is run well on Android phone and has Practical value. This technology helps to expand the application of SMS.

1 Introduction

"Intelligent" of smart phone greatly extends the functionality of the phone[1], the phone call is not just a call tool, it has become a good assistant and an indispensable tool in people work, live, entertainment, communication and transaction processing. Meanwhile, SMS(Short Message Service) is no longer a text communication tool. As one of the guarantees real-name system, SMS gradually become a tool of receiving verification information, accessing to the notification, and making transaction confirmation. With the development of fast payment such as online banking, Paypal, WeChat pay, etc., people are more inclined to quickly and easily done through mobile payment, during this process, SMS link in the whole play, timely notification, preserve transaction records and play an important role. So, it is cleared that SMS-based treatment processing, statistical analysis becomes significant and promising.

2 System Design

2.1 Business Process

Data is extracted automatically in the background. When the client first starts, SMS inbox is scanned automatically, key information is extracted from the corresponding text by regular expressions, including transaction time, transaction amount, transaction type, transaction banking, and etc., and automatically recorded in the APP database. When new message arrives, critical information also extracted by regular expressions [2], simplified refining and give a notification. Business process is shown in Figure 1.

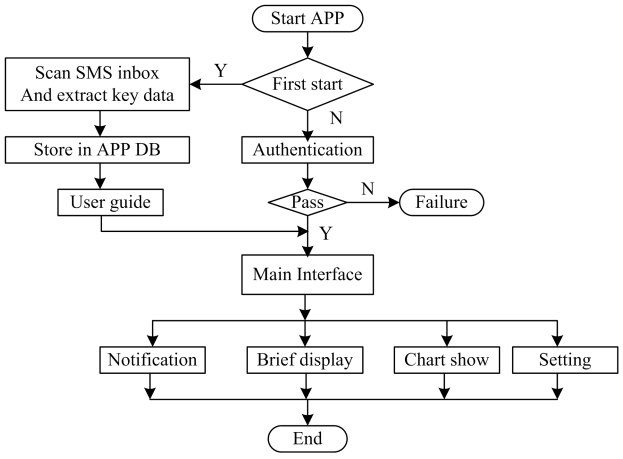

Figure 1. Business Process

2.2 System Function

This system based on Android platform[1], the system function module include: (1) User guide module;(2) Authentication module;(3) Data porcessing module; (4) Regex module; (5) SMS broadcast receiver module; (6) SMS summary module;(7) Chart module;(8) QR code scanning.

2.2.1 User Guide

The main function of this module is to show the main function and characteristics of the APP for the first use. This module is to be beautiful with elegant design and natural smooth transition animation. Leave a great impression to the users, meanwhile hiding the database processing module processing in the background.

2.2.2 Authentication

The main function of this module is to ensure the safety of personal privacy [3]. Because APP involves important information such as bank card, belong to the category of personal financial privacy, therefore, before entering the APP take two authentication mechanisms, including the Pin code and fingerprint recognize authentication.

Pin code validation:4 digit Pin code is used as the authentication information, and be set by the user.

Fingerprint authentication: fingerprintsis used as authentication schemes. a system level verification, access to key information in the TrustZone hardware, security is guaranteed.

2.2.3 Data Processing

Data extraction occurs at two moments: one is the first start of the application; another is the arrival of a new message after the starts. For the first case, the inbox database is scanned. In both case, data are acquired through match by regular expression, then are formatted and stored into APP database for analysis.

Database processing flow of the first case is shown in figure 1.

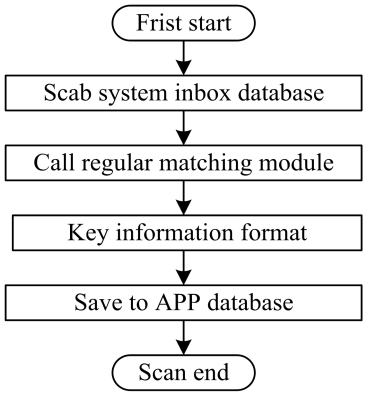

Figure 2. The Flow of Data Processing

2.2.4 Regular Expression

The main function of this module is to get key information using a regular expression. Regular Expression[4] (abbreviated as regex, regexp or RE), is syntax rules described by single string which is used to match a series of sentence. In many text editor, Regular expression are typically used to retrieve, replace the text that fits a pattern.

Different types of message take different regular expressions [5], Part of the regular expression, function and the matching results are shown in table 1. the key message is not the same, use Bank and Express message as examples.

Bank SMS: the bank name, bank card number and transaction amount, transaction time and transaction details information are necessary. Some information needs to format to show in charts.

Express SMS: Courier name, delivery time, delivery take the key information are necessary to ensure information notice.

Table 1. Example of Some Regular Expression

Function	Regex expression	Result
Extract Card	Suffix\w+	Suffix:6666
Extract Bank	(?<=[)[\u4e00-\u9fa5]+(? =])	the bank name
Extract Left Money	Balance\d+.\d+Yuan	Balance1000.00 Yuan

2.2.5 SMS Broadcast Receiver

The main function of this module is to get messages broadcast and analyze the message transmitting and text messages when phone received new messages, if the message is from bank or express , call the regular key information processing module to extract data, and call the information notification module of key information to make a notice, at the same time to write the key information into the database for storage. The business flow chart is shown in figure 3.

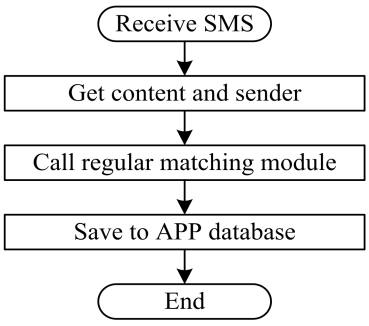

Figure 3. Receiving and Process of SMS

2.2.6 SMS Summary

The main function of this module is to show the information of the database. Using CardView and RecyclerView layout, only show the key information, Such

as bank logo, bank card, transaction type, trade time, trade amount, etc. The key technologies including dynamic hidden and loading step by step.

dynamic hidden:Hide Toolbar when move up to show more data information.

loading step by step:Load 20 items while o the interface is initialized, cooperate with the drop-down refresh and tensile load, more data to be loaded when the action triggers, reduce page rendering time. While adding a refresh and load the animation, in line with the intuitive operation.

2.2.7 Statistic Chart

Charts display visually the user's income and expenditure. For data stored in the APP database give statistics daily expenses and income; illustrate result monthly time periods, illustrating statistical results for each month. Expenditure and revenue respectively below and above the horizontal axis, a head understand. The results of the monthly income and expenditure are given in text.

2.2.8 QR Code Scanning

The main function of this module is to read the Courier number, at the same time for the APP to join more extensions provide reliable interface.

3 System Implementation

3.1 Guide Interface

The interface uses a light background color with a flat Schematic diagram. It introduces APP features, while completing the database scan task in the background. The running effect as show in Figure 4.

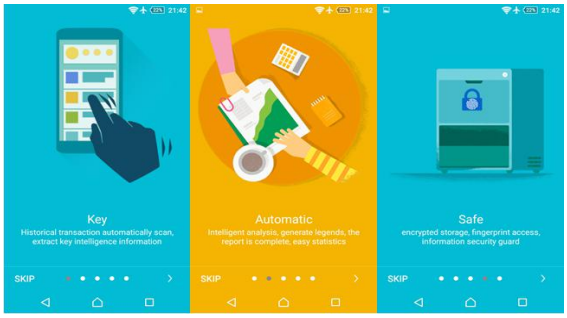

Figure 4 Guide Interface

3.2 Authentication Interface

The interface provides two authentication ways: fingerprint verification and password verification, shown in Figure 5.

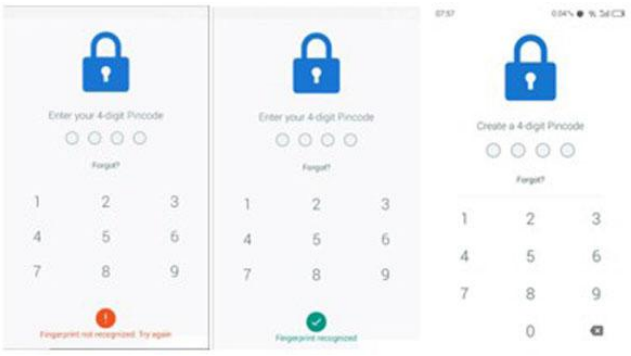

Figure **5.** Authentication Interface

In Fig. 5, the left one for fingerprint verification error, the center one for fingerprint authentication is successful interface, and the right one to enter the Pin code validation interface.

3.3 SMS Summary

Use the card layout, left for conventional interface, center for the drop-down refresh, on the right for tensile load and dynamic hide the ToolBar, shown in Figure 6.

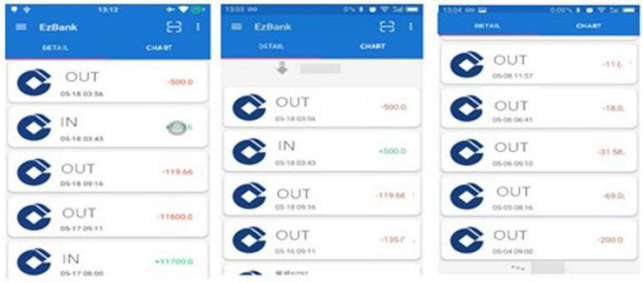

Figure 6. Infromation Display Interface

3.4 Chart Interface

Income and expenditure are labeled in different color and different position. Income for the direction of the green, red spending a negative direction. It supports two-finger zoom, zoom in to view details of the data. Statistics of the month shows below the table including the total revenue and total expenditure of current month with all auto-complete without manual settings. Graphics are made by google charts plugin.

Operation interface as shown in figure 7, support in switch, left with statistical figure of May and April respectively. Spending double refers to scaling, center and right are the same as for April, the width of the histogram.

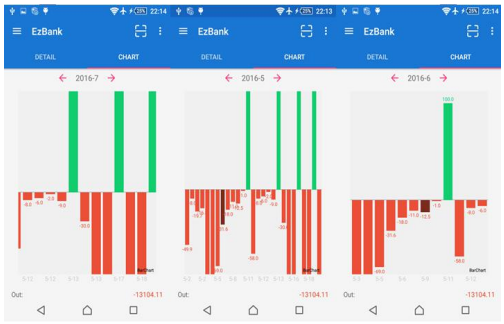

Figure 7. Chart Interface

The key code is as follow:

```
private void setDataFromDB(String sYear, String sMonth) {
    toShowData = DataSupport.where("year = ?", sYear).where("month = ?", sMonth).order("id
desc").find(BankCard.class);
    if (toShowData.size() > 0) {
        ……
        for (int m = 0; m < toShowData.size(); m++) {
            listFmoney = toShowData.get(m).getFmoney();
            if (listFmoney > 0) {
                in_sum = in_sum + listFmoney;
            } else {
                out_sum = out_sum + listFmoney;
            }
            all_sum = all_sum + listFmoney;
            listDate = toShowData.get(m).getMonth() + "-" + toShowData.get(m).getDay();
            dataToShow.add(new ChartData(m, listFmoney, listDate));
             }
        setData(dataToShow);
      ……
```

4 Conclusion

This system developed on the Android[6] platform with Java language and MVC [7] architecture. The system develop a complete text extraction, stored, ,analysis and gracphic display static result. System interface has the stytle of Google official Material Design style, with a good visual effect and interactive experience. Using a regular expression to extract text information, has high flexibility and extensibility.

References

[1] Research and trend analysis reports of 2016 China's smart phone market outlook Situation [EB/OL]. http://wenku.baidu.com/view/5076-968858fafab068dc024b,2016

[2] Q. Y. Qiu. Patented infromattion extration based on requlr expression 2007,18(19): 271-350

[3] S. Biswas, W. Haipeng, J. Rashid. Android Permissions Management at App Installing[J]. International Journal of Security and Its Applications, 2016,**10**(3):223-232

[4] J. E. Fried, S Yu. Mastering regular expressions [M]. Beijing: Electronic industry Press,2007

[5] M. C. Lian. POS endowed Corpus retrieval and that the regular expression [J]. Chinese Language Education, 2009, 2:65-73

[6] Android [EB/OL], https://zh.wikipedia.org/wiki/Android.

[7] MVC framework in Android mode of use, [EB / OL], http://blog.csdn.com/fxds/detail/324632.