

# Research on Olympics-oriented Mobile Game News Ordering System

Yonggui Yang, Lei Li

CISTR, Beijing University of Posts and Telecommunications, 100876

E-mail: [ygyang410@sohu.com](mailto:ygyang410@sohu.com) [lilei@nlu.caai.cn](mailto:lilei@nlu.caai.cn)

**Abstract.** The Olympics-oriented mobile game news ordering system introduced in the paper involved the technology of natural language processing, machine learning and information extraction. We collected text information from the major Internet media websites about the table-tennis and badminton games of the 2004 Athens Olympics; then we transformed these texts information into structured data through information extraction. Based on this source data-base, we can provide game news to the end users in the form that they can read them on their mobile phones easily.

**Keywords:** machine learning, information extraction, Short Message Service (SMS)

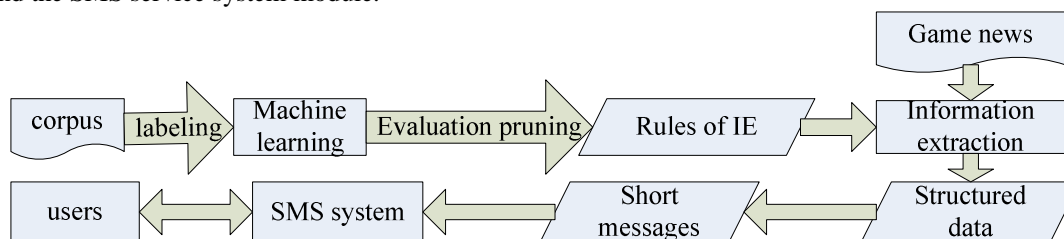
## 1. Introduction

2008 Beijing Olympics raise the aim of being a “digital Olympics”. In 2008, a great number of athletes, spectators and tourists all over the world will come to China, and that providing them with convenient mobile order service as well as query service can embody the feature of 2008 Beijing Olympics. Based on it, we try to develop this Olympics-oriented mobile game news ordering system. At present, the research on this system mainly aims at the items of table-tennis and badminton. Our final goal is to establish a quick and convenient game news information service system.

The paper is structured as follows. In the next section the overall frame of the system is described. In section 3 we introduce machine learning and information extraction functions. Section 4 describes the designing and realization of the SMS service system. Finally, we will say a word about the prospects and promotion points of our game news ordering system.

## 2. The overall Frame of the system

The system is mainly made up by two modules, the machine learning and information extraction module and the SMS service system module.



**Fig. 1.** The overall frame of the system

The task of the first module is to process the free Chinese game news texts to produce structured data as follows: With the machine learning we obtain the initial rules, after being evaluated and pruned, the rules will be put into the knowledge base of information extraction, and then the texts can be changed

into structured data through information extraction. The main task of the second module is to interact with the users and to receive and send SMS. It is made up by the server side and client side.

### **3. Machine learning and information extraction**

#### **3.1 Machine learning**

##### **3.1.1 Introduction of the program**

To generate rules, we used the machine learning method of from special to common and from bottom to top. Under our supervision, machine trained with the labeled corpus, then the initial rules are produced. More generalized rules are obtained through generalizing some items in the initial rules, and finally the exact and recapitulative rules are generated.

During the first processing, the program distilled the labeled sentence from text. Then we used the technology of shallow parsing to delete unimportant words. Then we specially designed a learning module which learned from the trained corpus and generated initial rules. Finally, we used Laplacian operator to evaluate the initial rules and to make sure whether they were the eligible rule.

##### **3.1.2 Corpus labeling and training**

Corpus labeling is the foundation of machine learning. We worked out a pattern for the labeling by which each label process is carried out step by step. The pattern consists of target, range, level and form to be labeled and the standard of the corpus. We collected 48 Chinese table tennis and badminton game news about 2004 Athens Olympics from the Internet media websites and all the game news was saved in the form of HTML.

At first the corpus was segmented automatically. For better effect in segmentation, we added some concrete names of sports meeting, names of items, nationalities and so on to the knowledge base. Then we labeled the items like sports meeting, contest, player, nationality, rival, rivals nationality, result, etc.

In order to train the corpus, we numbered the frequency of each part of speech in every labeled sentence, and then we fixed on the PRI (priority level) of the part of speech and exposed the result in a PRI table. This table was very important for it was directly related to the effect of shallow parsing, and it affected the quality of the rules. Then the PRI table was added to the program of machine learning. The program learned and trained with the labeled corpus and the initial rules were generated.

In order to heighten the veracity of information extraction, the initial rules still need some improvement before being added into the rule base. We used Laplacian operator to evaluate the initial rules and fixed on whether the rules are eligible based on the value of evaluation. Sometimes, we should do some manual pruning, especially for double matches. With the machine learning we obtained 20 eligible rules of information extraction, and 9 of these were manually pruned.

#### **3.2 Information extraction**

The rules which have been evaluated and pruned could be added to the knowledge base of information extraction. The information extraction module could use these rules and transform the game news into structured data. Then the results of IE were organized and a source data-base was built. We can provide SMS service to the end users in the form that they can read them on their mobile phones easily.

In Fig.3, there was a very refined quick message. If we would send it straightly away to a user, we would need two short messages; while with the information extraction (as in Fig.3 demonstrated), we need only one short message. The SMS the user received was as follows: “雅典奥运会乒乓球男子单打铜牌争夺战中国王励勤战胜瑞典瓦尔德内尔”.

The SMS contained the essential information of the original quick message, but the text is shortened drastically and is convenient for sending with mobile phone. If the game news would be longer – as it generally is the case – the effect of the information extraction would be a lot more evident.

[央视国际首页](#) > [新闻频道](#) > [体育新闻](#) > [正文](#)  
 王励勤获得乒乓球男子单打铜牌  
 央视国际 (2004年08月23日 18:36)  
 CCTV.com消息：在刚刚结束的雅典奥运会乒乓球男子单打铜牌争夺战中，中国选手王励勤4-1战胜瑞典选手瓦尔德内尔，五局比分分别是10-12，11-3，11-8，11-7，11-9。  
 责编：宋璐

```

抽取结果1:
sports-meeting: 雅典奥运会
item: 乒乓球男子单打
level: 铜牌争夺战
nationality: 中国
name: 王励勤
result: 战胜
opp-nationality: 瑞典
opponent: 瓦尔德内尔
  
```

Fig. 2. A corpus and the result of IE

## 4. SMS service system

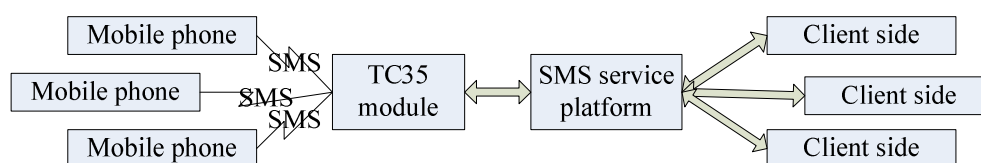


Fig. 3. SMS service system

### 4.1 Server side

The main task of the server side is to send and receive SMS between computers connected to our LAN and mobile phones of the wireless communication network, so it has to control the wireless module, to bargain on the communication interface with the client side, to process the SMS in order to conform to the transmission format. In addition, server side should save the system parameter existing or new and establish a log file. Server side was divided into five modules: serial interface processing module, SMS processing module, client side interface module, log processing module and parameter setup module.

Serial interface processing module is responsible for interactions between the server and the wireless module. SMS processing module receives SMS coming from the serial processing module and turns them into UTF16-code, then delivers them to the client side interface module and log processing module. SMS processing module also sends SMS coming from the client side module to the serial interface processing module. Client side interface module delivers the SMS coming from client side to the SMS processing module. In addition, it receives the SMS coming from SMS processing module and sends them packed according to the stated format to client side. Log processing module should log all the operations. The administrator can administered the system more effectively according to the log file. The task of parameter module is to set parameters of the system and SMS devices.

### 4.2 Client side

The main frame of client side is: it receives information packages from the server side, and then unpacks them, queries the texts and sends information, etc. Client side was divided into the SMS receiving module, the SMS sending module and the packing and unpacking module.

SMS receiving module could receive SMS from the client side. It receives the data packages sent by server and delivers the SMS coming from server to the SMS sending module. SMS sending module could process the data packages from the SMS receiving module and to transmit data packages to server. The data packages will be delivered to the packing and unpacking module and be unpacked, then there is an operation of querying and the results have been packed will be transmitted to server. Packing and unpacking module could unpack the data packages from the SMS sending module and to distill the contents of SMS. Later on it has to pack the processed contents of SMS and to deliver them to the SMS sending module.

### 4.3 An example

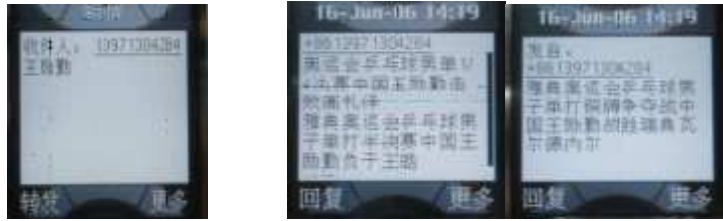


Fig. 4. Intercepted figure of mobile phone

A user was interested in the match situation about Wang Liqin, he or she just needed send the keywords “Wang Liqin” to the server number. Then the system will automatically query all the information about Wang Liqin in the source data-base generated by IE and send the result in form of SMS to the user.

### 5. Test

With a close test about the 48 texts we collected, we found that the information extraction result of 22 texts about table-tennis and 20 texts about badminton were right, while the result of other 6 texts about badminton were wrong. We achieved an information extraction accuracy of 87.5%.

Through an analysis, we found that the 6 texts with incorrect information extraction results were the game news on comments, and did not represent the results of matches. So we could see that the quality of corpus was important to the information extraction results.

### 6. Conclusions and expectation

In this paper, we introduced the overall frame and the realization of our developing Olympics-oriented mobile game news ordering system.

As the system is still in the early stage of its development, there is a great need for improvement and promotion: (1) The model of information extraction lacks high efficiency. As it is designed based on the corpus about table-tennis and badminton, it does not fit for items like swimming, tracking and field, etc. In the future, we will use the theory of comprehensive information, based on the characteristic of game news, to promote the efficiency and accuracy of the model. Then we will also think about providing integrative service for the users. (2) The present ordering function of the system is still very simple. The user can only send the name of an athlete. In the future, we will promote the ordering function, including increasing the service of customization, etc. (3) The present system is static. All the corpus are pre-collected and can't be real-time updated. During our next research steps we will use the technology of search engine to monitor the major Internet media websites to real-time update our data-base.

### References

1. Li Lei, Zhou Yanquan: Research on Comprehensive Information Based Chinese Information Extraction System. IEEE NLP-KE 2005, Wuhan, Oct. 2005.
2. Li Lei, Zhou Yanquan: Research on A New Intelligent Mobile Information Service. CAAI-11, Oct. 2005.
3. Li Lei, Zhou Yanquan, Wang Jinghua: Center for Intelligence Science and Technology Reaearch of BUPT, “Chinese Information Extraction System and Application Based on Comprehensive Information”, Transaction of Beijing University of Posts and telecommunications, Volume 28, Issue 6, Dec.2005, Page 48-51.
4. Zhang Xiurong, Beijing University of Posts and telecommunications: “Analysis, Design for Multilingual Mobile Terminal Prototype System—Design, Implementation for SMS Service System”, 2004