

Design and Implementation of Customer Personalized System Using Web Log and Purchase Database

Jae-hoon Lee* · Hyun-sook Chung** · Sung-joo Lee**

* Dept. of Computer Science, Graduate School, Chosun Univ.

** Dept. of Computer Engineering, Chosun Univ.

Abstract

In this paper, we propose a customer personalized system that presents the web pages to users which are customized to their individuality. It analyzes the action of users who visit the shopping mall, and preferentially supplies the necessary information to them. When they actually buy some items, it forecasts the user's access pattern to web site and their following purchasable items and improves their web page on the bases of their individuality. It reasons the relation among the web documents and among the items by using the log data of web server and the purchase information of DB. For reasoning, it employs Apriori algorithm, which is a method that searches the association rule.

It reasons the web pages by considering the user's access pattern and time by using the web log and reasons the user's purchase pattern by using the purchase information of DB. On the basis of the relation among them, it appends the related web pages to link of user's web pages and displays the inferred goods on user's web pages.

Key Words : Personalized system, Web log, Reasoning

1. Introduction

It needs information about user's preference and access pattern in order to make a marketing strategy and to supply the user-oriented information in Internet. We can supply the dynamic web pages or the link information, which are customized to user's individuality, by using this information.

There have been many researches about this technology. This technology is based on user's convenience and usefulness, and is coming into the spotlight[3].

In this paper, we propose and implement a customer personalized system. It reasons the relation among the web documents and among the items by using the log data of web server and the purchase information of DB. For reasoning, It employs Apriori algorithm, which is a method that searches the association rule.

It reasons the user's access pattern and time by using the web log and reasons the user's purchase pattern by using the purchase information of DB.

On the basis of the relation among them, it appends the related web pages to link of user's web pages and displays the inferred goods on user's web pages.

2. Personalization

2.1 Background of personalization

Many sites make users stay at or revisit their site by using the various marketing methods. One of them is to make users

choose service directly according to their preference degree. It is called personalization that sites supply visitors with information they would prefer[1, 7].

2.2 Advantage of personalization

Definition of personalization is wide concept including Customization, Individualization, and Group Characterization [7].

The companies found methods that users could use more easily huge web site. A method is personalization. Personalization improves familiarity between users and web sites and provides a better service.

Personalization uses the customer information and the various technologies in order to track the electronic commerce interactions between a business and each individual customer. Using information either previously obtained or provided in real-time about the customer, the exchange between the parties is altered to fit customer's needs as well as business's needs based on the available customer information.

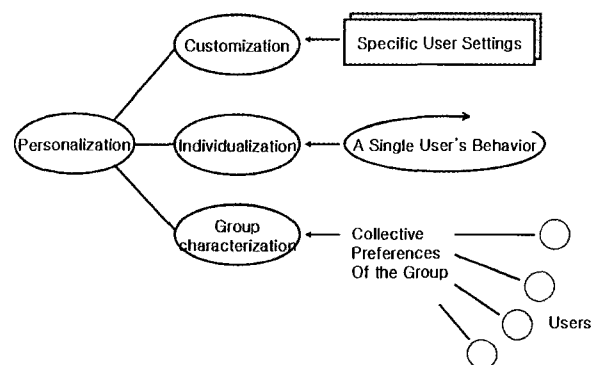


Fig. 1. Presentation form and contents of personalization

Fig. 1. show presentation form and concepts of personalization.

3. Web Log Mining & Apriori Algorithm

3.1 Web Log Mining

In web, basic information about user's activity is automatically collected by web server. Web server stores and manages them in form of 'Web Log'. Web Log data includes a visit information to web page that can recognize user who connect to web server and has characteristics that are different from general statistic data[5].

Basically, users sequentially visit to web page according to time, and log data is used in order to find user's visit pattern to web page mainly. Fig. 2. shows format of Log data.

```

Common Log Format
165.194.11.110 - - [11/Jul/2000:21:06:32 +0900] "GET /image/animimage/head1.jpg HTTP/1.1" 200 7592
165.194.11.110 - - [11/Jul/2000:21:06:32 +0900] "GET /defaultms.asp HTTP/1.1" 200 19514
165.194.11.110 - - [11/Jul/2000:21:06:32 +0900] "GET /image/animimage/statinfo.jpg HTTP/1.1" 200 7376

Extended Log Format
01:30:23 210.117.171.66 GET /jshny/sampling/ 302
01:30:23 210.117.171.66 GET /jshny/sampling/Default.htm 200
01:31:35 210.182.144.220 GET /Emp/Guide/042.html 200
01:31:44 210.182.144.220 GET /Emp/images/midback.gif 200
    
```

Fig. 2. Format of Log data

Web Mining analyzes such data as log file from web site and establishes strategy. It means adding other data to web log file and analyzing them. Here, other data are customer data, account data, electronic commerce data etc.[5, 6].

Usually, Web Mining is fallen into web structure mining, web contents mining and web usage mining according to web data that become target[7, 8].

3.1.1 Web Structure Mining

The purpose of Web structure mining is to get the structural summary information of web site and web page. Structural information of web site is graph structure by hyperlink between web pages.

3.1.2 Web Contents Mining

Web contents mining is technique to draw the meaningful contents from pages which web site is composed of. It is technology that automatically find useful information from wide web data(text, picture, sound etc.) in on-line.

3.1.3 Web Usage mining

Web usage mining is to analyze web-user's use pattern. By using this information, we can understand web-user's action more than connection statistics. By analyzing user's log file, we can perform various analysis such as periodic analysis, user analysis and page analysis as well as user number to visit web site simply.

Web log mining can divided into the preprocessing process, the pattern discovery process and the pattern analysis process. The pattern analysis process analyzes the created rule and pattern but sometimes can be excepted in last analysis step. Fig. 3. shows web log mining's order.

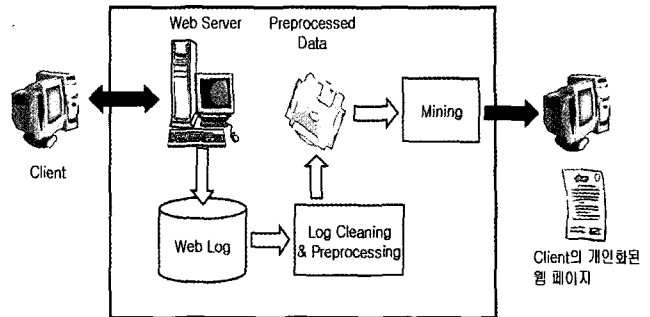


Fig. 3. Web log mining's order

Fig. 4. shows the step of preprocessing of log data.

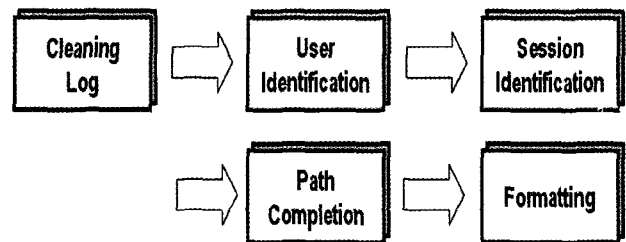


Fig. 4. Step of preprocessing of log data

3.2 Association Rule Inquiry and Apriori Algorithm

Association rule inquiry is to find type of the event that happens frequently in unit transaction in high-capacity database[2].

Apriori algorithm is a representative algorithm of association rule inquiry. Association rule reasoning process that uses this algorithm is consisted of next time two steps.

- Step 1
Search a set of frequently occurred items which have a transaction support more than the predetermined smallest support.
- Step 2
To create association rule from DB, use a set of frequently occurred items.

Fig. 5. shows the search process of set of frequently occurred items in Apriori algorithm, and algorithm is same with Table 1.

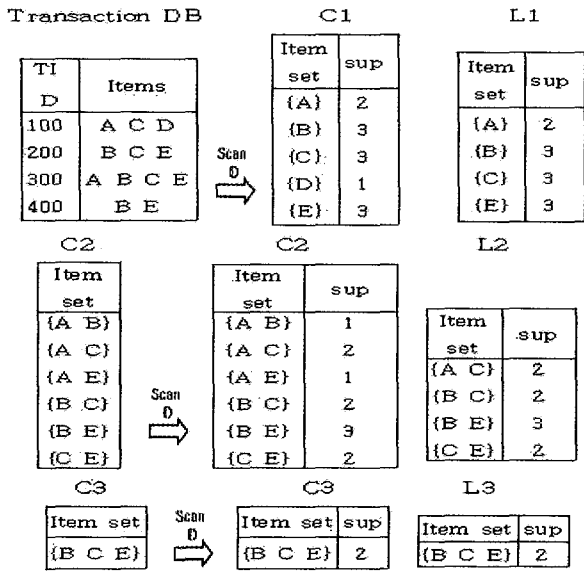


Fig. 5. Creation of candidate item set and frequent occurrence item set

Table 1. Apriori Algorithm

```

L1 = {Large 1-itemsets};
for (k=2; Lk-1 ≠ 0; k++) do begin
  Ck = apriori-gen(Lk-1); //New candidates
  forall transactions t ∈ D do begin
    C1 = subset(Ck, t);
    //Candidates contained in t
    forall candidates c ∈ C1 do
      c.count++;
    end
    Lk = {c ∈ Ck | c.count ≥ minsup}
  end
  Answer = UkLk

```

4. Personalization System

4.1 Abstract of System

In this paper, we present a method to use association rule, which is one of data mining techniques, by using user's purchase information in log data and database of web server basically in order to offer the personalized contents. Therefore, this system is consisted of two parts. In one part, reasoning is achieved with customer's body information and purchase information. Other part grasps customer's running path of web document through log data and use it in reasoning.

In reasoning, Apriori algorithm that inquires the association rule was applied. In reasoning part to use purchase data, algorithm makes ITS(Item Set) by using various kinds of items that user buys, and creates items that user has an ample possibility to purchases on the basis of ITS. While, in reasoning part to use log data, algorithm makes DS(Document Set), and create Documents that user has an ample possibility to approach

on the basis of DS.

Specially, in S-Apriori algorithm that reasons Web documents that user has an ample possibility to approach on the basis of DS, pre-treatment process must find out document running path (P) by user and need to find out MFR (Maximal Forward Reference).

In this paper, we consider time that user stays at web page as well as finding web page that user frequently approaches as existing web

The proposed system finds and recommends patterns having information that user wants by considering time and giving weight to it.

4.2 Composition of system

Fig.6 shows Basic structure of personalization system. Fig. 7 shows Reasoning engine that is the most important part in personalization system. Reasoning engine is fallen into part that uses log file and that uses the purchase recording of database.

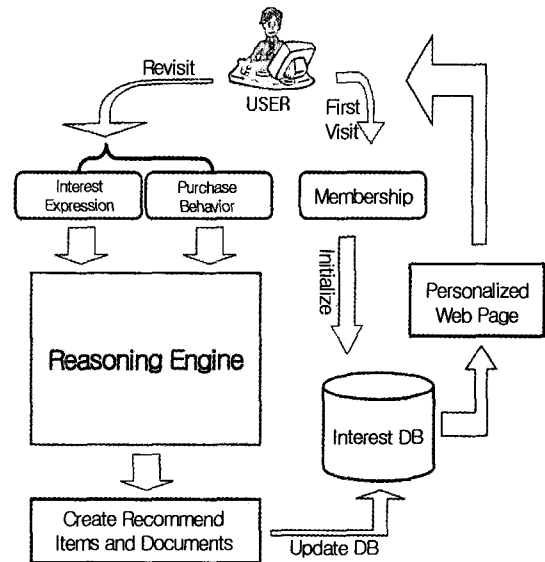


Fig. 6. Flow Chart of Personalization System

Part that uses log file of web server collects web logs and preprocesses them for web log analysis. The preprocessing part finds out running route by user and finds out MFR at the same time because it must run Apriori algorithm. It recommends Documents using Apriori algorithm and the pre-processed data.

Part that uses the purchase recording of database so composes customer Database and purchase Database that connection deduction is possible. It reasons the recommended Items by using the general Apriori algorithm.

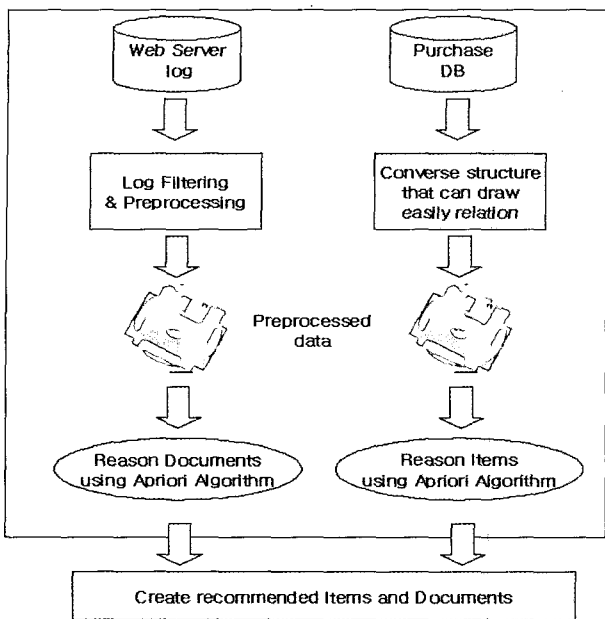


Fig. 7. Reasoning Engine

4.2.1 Reasoning Web Page using log file of Web Server

Reasoning using log files of Web Server collects and filters Web logs from Web server like the general method of web log analysis. It must preprocess next two in order to apply S-Apriori algorithm.

(a) User document running path(P)

The moving path using Fig. 8. is as following.

$$P = \{A, B, H, B, I, B, A, C, J, M, J, C, K, A, D, L\}$$

(b) MFR(Maximal Forward Reference)

It find out web documents that user passes through in case of proceeding except the case of backing, For example, MFR is as follow in Fig. 8.

$$MFR = \{ABH\}, \{ABI\}, \{ACJM\}, \{ACK\}, \{ADL\}$$

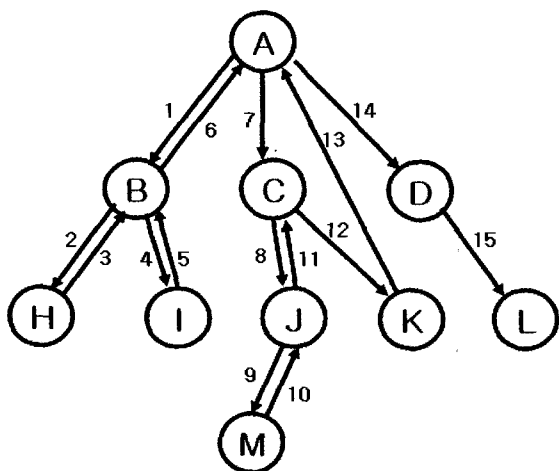


Fig. 8. Running path of web document

If two preprocessing is finished, this system reasons by applying S-Apriori algorithm to the preprocessed data. It adds time-weight to the general S-Apriori algorithm. Adding time-weight increase the confidence of reasoning.

Table 2. Web log data

User ID	(Web Page, Unit Time)
100	(H, 1), (A, 1) (C, 2)
200	(H, 1), (A, 1), (H, 0), (B, 2), (H, 1), (A, 1), (D, 2)
300	(H, 0), (A, 2), (D, 1), (E, 2), (B, 0)
400	(H, 1), (B, 1), (H, 0), (E, 1), (A, 1), (D, 3)
500	(H, 0), (B, 1), (E, 1), (H, 0), (A, 1), (D, 3)

Table 2. is Web log data of some user who visits Web site. First data in parenthesis is Web page that he visits. Second data is a unit time. (A, 2) means that he visits Web page A during 2 unit time. (H, 0) means that he visits Web page H during less time than minimum unit time and that he pass through latter rather than visits Web site.

If pages having the visit time less than 30 seconds is ignored in Table 3. in order to apply S-Apriori algorithm considering time-weight, it become Table 4.

Table 3. Time Weight per Unit

Unit	Time	Weight
0	0 ~ 30 sec.	0
1	30sec ~ 5min	0.2
2	5min ~ 30min	0.3
3	more than 30min	0.5

Table 4. Fine Web log data

User ID	(Web Page, Unit Time)
100	(H, 1), (A, 1) (C, 2)
200	(H, 1), (A, 1), (B, 2), (H, 1), (A, 1), (D, 2)
300	(A, 2), (D, 1), (E, 2)
400	(H, 1), (B, 1), (E, 1), (A, 1), (D, 3)
500	(B, 1), (E, 1), (A, 1), (D, 3)

If it appoints the approval rating as 0.4 (minsup = 0.4) and reasons FOIS(Frequent Occurrence Item Sets) by applying S-Apriori algorithm, result is as following.

(c) 1-FOIS

$$\{H\} = (0.2+0.2+0.2+0.2) \times (3/5) = 0.48$$

$$\{A\} = (0.2+0.2+0.2+0.3+0.2+0.2) \times (5/5) = 1.1$$

$$\{B\} = (0.3+0.2+0.2) \times (3/5) = 0.42$$

$$\{D\} = (0.3+0.2+0.5+0.5) \times (4/5) = 1.2$$

$$\{C\} = (0.3) \times (1/5) = 0.06$$

$$\{E\} = (0.3+0.2+0.2) \times (3/5) = 0.42$$

$$1\text{-FOIS} = \{H, A, B, D, E\} \text{ (minsup}=0.4)$$

(d) 2-FOIS

$$\begin{aligned} \{H, A\} &= (0.2+0.2+0.2+0.2+0.2+0.2) \times (3/5) = 0.72 \\ \{H, B\} &= (0.2+0.3+0.2+0.2) \times (2/5) = 0.36 \\ \{H, D\} &= (0.2+0.3+0.2+0.5) \times (2/5) = 0.48 \\ \{H, E\} &= (0.2+0.2) \times (1/5) = 0.08 \\ \{A, D\} &= (0.2+0.3+0.3+0.2+0.2+0.5+0.2+0.5) \times (4/5) = 1.92 \\ \{A, B\} &= (0.2+0.3) \times (1/5) = 0.1 \\ \{B, A\} &= (0.3+0.2+0.2+0.2+0.2+0.2) \times (3/5) = 0.78 \\ \{A, E\} &= (0.3+0.3) \times (1/5) = 0.12 \\ \{E, A\} &= (0.2+0.3) \times (1/5) = 0.1 \\ \{B, D\} &= (0.3+0.3+0.2+0.5+0.2+0.5) \times (3/5) = 1.2 \\ \{B, E\} &= (0.2+0.2+0.2+0.2) \times (2/5) = 0.32 \\ \{E, D\} &= (0.2+0.5+0.2+0.5) \times (2/5) = 0.56 \end{aligned}$$

$$2\text{-FOIS} = \{H, A\}, \{H, D\}, \{A, D\}, \{B, A\}, \{B, D\}, \{E, D\} \\ (\text{minsup} = 0.4)$$

(e) 3-FOIS

$$\begin{aligned} \{H, A, D\} &= (0.2+0.2+0.3+0.2+0.2+0.5) \times (2/5) = 0.64 \\ \{B, A, D\} &= (0.3+0.2+0.3+0.2+0.2+0.5+0.2+0.2+0.5) \times (3/5) \\ &= 1.56 \end{aligned}$$

$$3\text{-FOIS} = \{H, A, D\}, \{B, A, D\}$$

Because more FOIS than upper cases is not created, {H, A, D}, {B, A, D} item updates DB interest of system and recommends next time web-site to user.

Table 5. compares general S-Apriori algorithm with S-Apriori Algorithm considering time-weight.

Table 5. Compare S-Apriori and Time Weighted S-Apriori

IS (Item Set)	S-Apriori (minsup = 0.4)	S-Apriori considering Time Weight (minsup = 0.4)
1-IS	{H}(0.6), {A}(1), {B}(0.6), {D}(0.8), {E}(0.6)	{H}(0.48), {A}(1.1), {B}(0.42), {D}(1.2), {E}(0.42)
2-IS	{H, A}(0.6) {H, B}(0.4) {H, D}(0.4) {A, D}(0.8) {B, A}(0.6) {E, A}(0.4) {B, D}(0.6) {B, E}(0.4) {E, D}(0.4)	{H, A}(0.72) {H, D}(0.48) {A, D}(1.92) {B, A}(0.78) {B, D}(1.2) {E, D}(0.56)
3-IS	{H, A, D}(0.4) {H, B, D}(0.4) {B, A, D}(0.6) {B, E, A}(0.4) {B, E, D}(0.4) {E, A, D}(0.4)	{H, A, D}(0.64) {B, A, D}(1.56)
4-IS	{B, E, A, D}(0.4)	-

4.2.2 Reasoning Items by using Purchase DB

Reasoning Item by using purchase DB is not more complicated than reasoning web pages by using web log. It collects data from the customer data and the purchase data, and makes

Apriori algorithm use them by preprocessing. It recommends product by general Apriori algorithm and supplies the related information for customer.

Algorithm used for reasoning item is same as one for reasoning Web page by using Web log. However, it does not use time-weight, and the item set is not web page(document set) but product set.

5. Conclusion

By enlargement of business in Internet, the relation with customer become more and more important. For more smooth relation with customer, business constructs not Web-site presented one-sidedly to customer but being able to communicate with customer.

In this paper, we proposed Personalization system for this need. The proposed system presented one method of personalization to construct web-site according to individual's favorite.

This system analyzed user's pattern by using web log data and purchase DB, and reasoned goods that user buys actually and web pages that user visits frequently. It becomes more believable personalization system by this reasoning.

We did not consider new subscribers because information about them was not enough. In cases of new subscribers which system does not accumulate information about, it can use method to use user's professional file on subscribing membership or to use the example based reasoning. The example based reasoning recommends web pages to user by reasoning what products customers who have similar career or interest have bought. However, we presented basic page.

We increased confidence of personalization by using web log information and user's purchase information at the same time and by using Apriori algorithm and the changed Apriori algorithm as the reasoning method. Apriori algorithm is one method of connection rule inquiry.

In part that uses log data of web server, the proposed system analyzed user's access pattern to web page by considering time that user stays at web page as well as web pages that user frequently approaches, and could recommend the high believable web pages to user. In part that recommends goods by using the purchase DB, this system recommended goods by applying Apriori algorithm that uses in traditional connection rule inquiry.

References

- [1] Srikant, Yinghui Yang, "Mining Web Logs to Improve Website Organization", 2001.
- [2] Hyuncheol Kang, Byoung Cheol Jung, "A Study of Web Usage Mining for eCRM", The Korean Communications in Statistics, Vol.8, No.3, pp. 831-840, 2001.
- [3] Jong-Su Park, Yeong-Kung Yu, "Inquiry and Application of Association Rule", Dept. Computer Science, Sungsin Univ.

- [4] Suck-Ki Kim, Jung-Yong Ahn, "A Study on Analysis Method of Web Log data".
 - [5] Suck-Ki Kim, Jung-Yong Ahn, "Applicate Information Using Web Log Data Analysis".
 - [6] Jin-Sung Kim, "Membership Functions and AHP-Based Negotiation Support in Electronic Commerce", Journal of Fuzzy Logic and Intelligent Systems Vol.12, No.4, pp.347-352, 1225-1127, 2002.
 - [7] Sug-Ki Kim, Hyun-Jung Koh, "Customer Support System of Web Based intelligent", Processing of KFIS Fall Conference, 2003.
 - [8] <http://www.webpro.co.kr>
 - [9] Harris Kravatz, "Designing Web Personalization Features".
 - [10] C-H. Lee, Y-H. Kim, P-K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique", Expert System with Applications, pp.131-137, 2001.
-



Sung-Joo Lee

1970 : Dept. of Physics Sciences,
Hannam University(B.S)
1992 : Dept. of Computer sciences,
Kwangwoon University(M.S)
1998 : Dept. of Computer sciences,
Catholic Univ. of Daegu(Ph.D)
1988~1990 : Chief, Computer Center,
Chosun University

1995~1997 : Dean, Information Science, Chosun University
1981~now : Professor in the Dept. of Computer Engineering
Chosun University

Research Interests : Software engineering, Programming Language, Object-oriented software, Rough set



Jae-Hoon Lee

2000 : Dept. of Computer Sciences,
Chosun University(B.S)
2002 : Dept. of Computer sciences,
Chosun University(M.S)
2002~ : Dept. of Computer Science,
Chosun University Doctoral
Student

Research Interests : Software engineering, Programming Language, Object-oriented software, Genetic Algorithm, eCRM



Hyun-Sook Chung

1993 : Dept. of Physics, Catholic
University of Daegu(B.S.)
1995 : Dept. of Computer Science,
Catholic Univ. of Daegu(M.S.)
2003 : Dept. of Computer Science,
Yonsei University(Ph.D.)

2006~ : Dept. of Computer Engineering,
Assistant Professor at Chosun University

Research Interests : character modeling, motion graphics,
e-Learning, Semantic Web

Phone : 062-230-6447
Fax : 017-288-4758
E-mail : hsch@chosun.ac.kr