

2001

2002

論文 碩士學位 論文 提出

2002 年 1月

梨花女子大學校 科學技術大學院

學科 成 炫 姪

# 碩 士 學 位 論 文 認 准

指導教授

審查委員

_____	_____
_____	_____
_____	_____

梨 花 女 子 大 學 校 科 學 技 術 大 學 院

	-----	vi
I .	-----	1
1.1	-----	1
1.2	-----	2
II .	-----	4
2.1	-----	4
2.2	-----	5
2.3	-----	6
2.4	-----	7
III .	-----	9
3.1	-----	10
3.2	-----	11
3.3	-----	17
3.4	-----	21
3.4.1	-----	21
3.4.2	-----	21
3.4.3	-----	22
3.5     가	-----	25

IV .	-----	27
4.1	-----	27
4.2	-----	28
4.3	-----	30
4.4	-----	32
4.4.1	-----	32
4.4.2	-----	34
4.4.3	-----	34
4.4.4	-----	35
V .	-----	37
	-----	39

[ 2.1]	webTrends	-----	5
[ 2.2]	webCANVAS	-----	6
[ 3.1]		-----	9
[ 3.2]		-----	10
[ 3.3]		-----	12
[ 3.4]		-----	16
[ 3.5]		-----	16
[ 3.6]		-----	16
[ 3.7]		-----	18
[ 3.8]		-----	19
[ 3.9]		-----	20
[ 3.10]		-----	20
[ 3.11]	가	2 -----	26
[ 4.1]		-----	29
[ 4.2]	Import log	-----	31
[ 4.3]		-----	32
[ 4.4]		-----	33
[ 4.5]		-----	34
[ 4.6]		-----	35
[ 4.7]		-----	36

[ 3.1] common log format	-----	13
[ 3.2]	-----	14
[ 3.3]	-----	15
[ 3.4]	-----	15
[ 3.5]	-----	17
[ 3.6]	-----	23
[ 3.7]	-----	24
[ 4.1]	-----	27
[ 4.2]	-----	28





# I.

## 1.1

가 ,  
 (Web content mining),  
 (Web structure mining),  
 ( ) (Web usage  
 mining)

가 가 가  
 가

1.2



가

[1]

가

가

2

3

4가

4

가

5

## II.

가

(2.1, 2.2, 2.3)

(2.4)

### 2.1

가

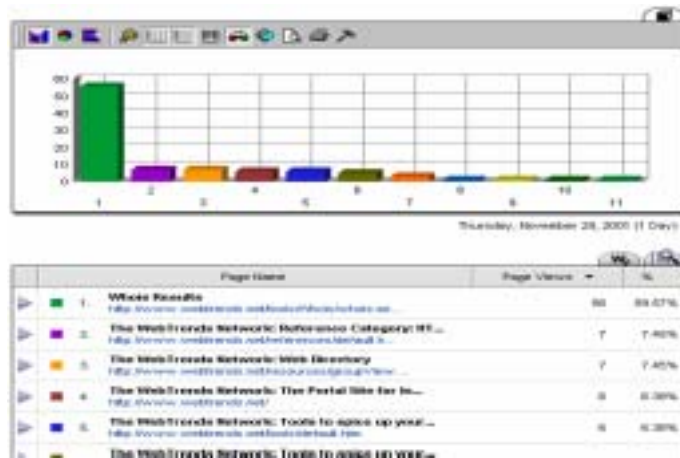
가

가

2.1 WebTrends

X Y (hit)

가



### 2. 1 WebTrends

### 2.2 (Association Rule)

(association rule)

가

가

Apriori

[2]

suffix 가

## 2.3 (Clustering)

(clustering)

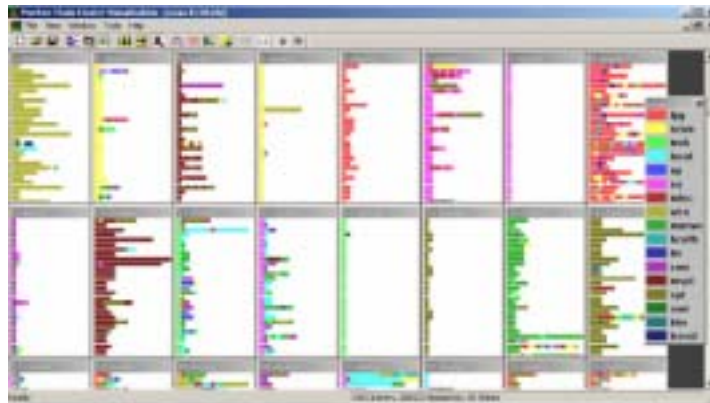
가

2.2

webCANVAS(web clustering analysis and visualization of sequence)

가

[3].



2. 2 webCANVAS

## 2.4

[4]

가

1 가 가

가

(iterative deepening search)

WUM(Web Utilization Miner)

가

가 MINT

[5].

2.4

ID

, ( )

가

1.2

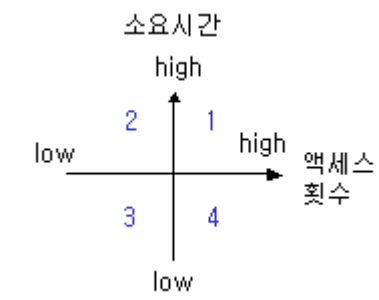
3



### III.

3.1

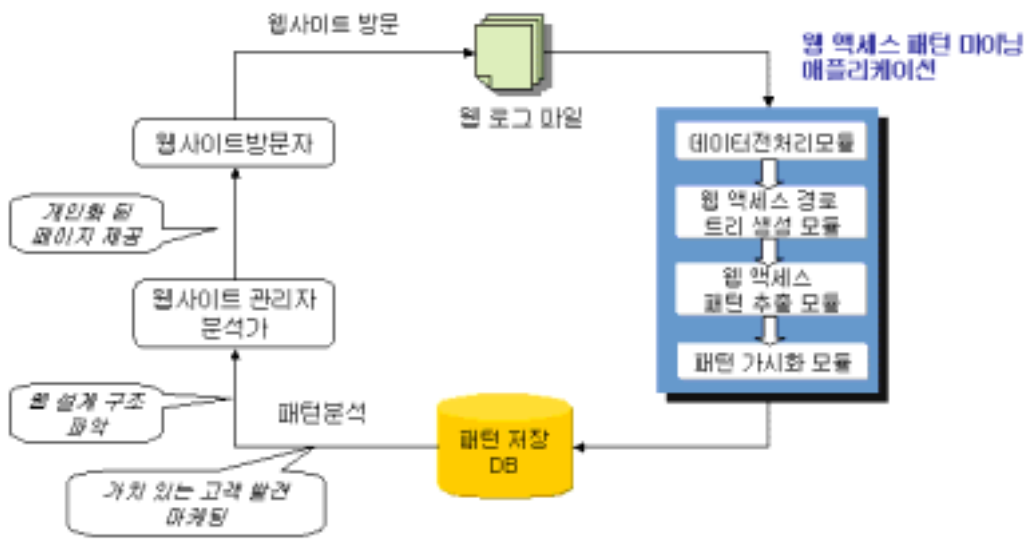
X Y 가 ,  
 4 1 . 1  
 가 . 3  
 , 가  
 가 , 2  
 가  
 가  
 가  
 가 2



3. 1

### 3.1

( 3.2) 가 ,  
 IP( ) 가 가  
 가 .



3. 2

□

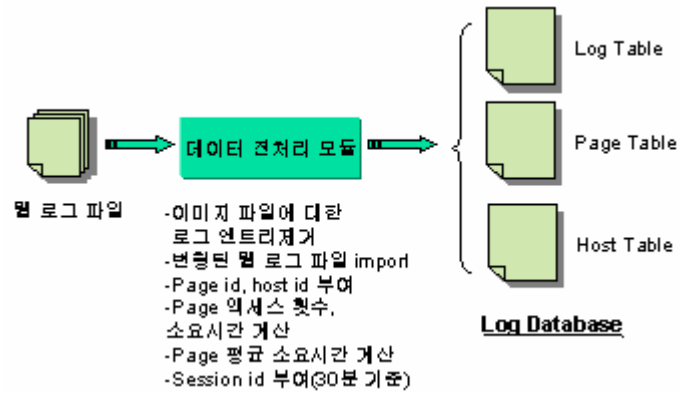
:

□ : .  
 가 .  
 □ : 가  
 □ 가 : 가 가  
 , 가 가  
 가 가  
 , 가  
 ,

### 3.2

가 .

3.3



3. 3

가 common log

가

```

in24.inetnebr.com - - [01/Aug/1995:00:00:01 -0400] "GET /shuttle/missions/sts-68/news/sts-68-
mcc-05.txt HTTP/1.0" 200 1839

uplherc.upl.com - - [01/Aug/1995:00:00:07 -0400] "GET / HTTP/1.0" 304 0

uplherc.upl.com - - [01/Aug/1995:00:00:08 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0"
304 0

```

1

*remotehost rfc931 authuser [date] "request" status bytes "referrer" "user\_agent"*

3.1

### 3. 1 common log format

Remotehost	DNS ( IP )
Rfc931	
Authuser	
Date	
Request	( : )
Status	HTTP status
Bytes	
Referrer	URL
User_agent	

가 가 ,  
 Remotehost, Date, Request 가 . ,  
 가  
 ID . 가 가  
 ID .  
 가 가  
 가 .

가

30

가

ID

가

3.2,

3.3, 3.4

(support)

0

1

가

common log

MS IIS log

### 3.2

ID		
Host		
Timestamp	/	
Page_name		
Page_id		ID
Duration		

Session_id		ID
------------	--	----

### 3.3

Page_id		ID
Page_name		
Avg_duration		
Access		
Support		/

### 3.4

Host_ID		ID
Host_name		

3.4

host timestamp .

ID	host	timestamp	page_name	page_id	duration	session_id
25446	128,110,29,180	00-11-02 오전 8:37:20	/sportsScience/sportsscience_Index.asp	536	8	1
25449	128,110,29,180	00-11-02 오전 8:37:28	/toem/javascript/login.js	187	11	1
25452	128,110,29,180	00-11-02 오전 8:37:39	/SportsScience/psychology/psychology_Inc	528	11	1
25453	128,110,29,180	00-11-02 오전 8:37:50	/SportsScience/psychology/Psychology_4.c	527	6	1
25455	128,110,29,180	00-11-02 오전 8:37:56	/SportsScience/psychology/Psychology_2.c	525	3	1
25457	128,110,29,180	00-11-02 오전 8:37:59	/SportsScience/psychology/Psychology_3.c	526	-1	1
17732	128,134,121,184	00-11-01 오전 7:20:20	/SportsCenter/diet/diet.asp	301	27	2
17739	128,134,121,184	00-11-01 오전 7:20:47	/SportsCenter/diet/0015.asp	285	16	2
17742	128,134,121,184	00-11-01 오전 7:21:03	/SportsCenter/diet/0024.asp	294	32	2
17743	128,134,121,184	00-11-01 오전 7:21:35	/SportsCenter/diet/0001.asp	271	-1	2
3166	128,134,130,11	00-10-28 오후 4:07:53	/Index.asp	199	-1	3
8328	128,134,130,11	00-10-29 오후 4:07:09	/Index.asp	199	-1	4
20406	128,134,130,11	00-11-01 오후 4:36:05	/Index.asp	199	-1	5
28311	128,134,130,11	00-11-02 오후 4:20:30	/Index.asp	199	-1	6
35291	128,134,130,11	00-11-03 오후 4:27:25	/Index.asp	199	-1	7
41190	128,134,130,11	00-11-04 오후 4:20:28	/Index.asp	199	-1	8
26457	128,134,220,199	00-11-02 오전 11:43:20	/SportsScience/Marketing/SportsMarketing_	450	-1	9
14772	128,134,221,130	00-10-30 오후 2:37:20	/toem/board_event.asp	183	14	10
14778	128,134,221,130	00-10-30 오후 2:37:34	/toem/javascript/login.js	187	39	10
14781	128,134,221,130	00-10-30 오후 2:38:12	/SportsCenter/...	921	20	10

3. 4

3.5, 3.6

page_id	page_name	avg_duration	access	support
1	-	12	1	2,388117E-05
2	/_vti_bin/shtml,	0,375	8	0,0001910493
3	/_vti_inf.html	1,25	8	0,0001910493
4	/admin/admin_	24,66667	18	0,000429861
5	/admin/Admin_	5,846154	13	0,0003104552
6	/admin/content	1,833333	13	0,0003104552
7	/admin/GameD	65,63636	11	0,0002626928
8	/admin/GameD	5	1	2,388117E-05
9	/admin/GameD	4,857143	7	0,0001671682
10	/admin/GameD	13	7	0,0001671682
11	/admin/GameD	9,4	5	0,0001194058
12	/admin/Login_E	4,5	2	4,776234E-05
13	/admin/Login_ç	3,466667	15	0,0003582175
14	/admin/logoff,a	2	1	2,388117E-05
15	/Admin/Membe	386	1	2,388117E-05

3. 5

ID	host
1	128,110,29,180
2	128,134,121,184
3	128,134,130,11
4	128,134,220,199
5	128,134,221,130
6	128,134,253,133
7	128,134,253,136
8	128,134,253,44
9	128,134,33,90
10	128,134,35,186
11	128,134,37,190
12	128,134,69,150
13	128,134,98,61
14	128,134,98,62
15	128,173,34,190

3. 6



### 3.3

3.5

ID

가                    가                    가

ID

### 3.5

ID	
1	1 2 3
2	1 3 1 2
3	1 2 4 5
4	2 3 1
5	2 4
6	1 3 1
7	1 3 5
8	2 4 3 5

,    4                    2                    ,                    3    1



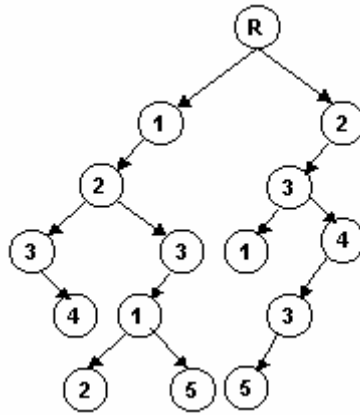
가 , 가  
 , 가 가  
 1 가 가 ( )가  
 . 가 2  
 . ( )

3.7

3.8

가

가



3. 8

가

C++

( 3.9)

( 3.10)

```

typedef struct node * treePointer;
typedef struct node {
    int page_id;           // ID
    long page_access;     //
    long page_duration;  //
    treePointer left_child; // 가
    treePointer right_child; // 가
}

```

3. 9

```

Tree_generation() {
    for(;;) {
        가
        1
        가
        가 NULL break
    }
}

```

3. 10

### 3.4

3.3

가

가

가 가

가 . ,

가

가

#### 3.4.1

가

가

가

#### 3.4.2

가 .

가 ,

가 ,

가 ,

가 .

가 .

가 .

### 3.4.3

ID

(string)

3.6

ID parent\_page\_id

ID child\_page\_id

pattern\_id

, 3.3

가 가 ,

. ,

가

3.6

가 .

**3. 6**

Pattern_id		ID
Parent_page_id		ID
Child_page_id		ID

3.5

3.7 ,

3.7 .

**3.7**

Pattern_id	Parent_page_id	Child_page_id
1	1	2
1	2	3
2	1	2
2	2	4
2	4	5
3	1	3
3	3	1
3	1	2
4	1	3
4	3	5
5	2	3
5	3	1
6	2	4
6	4	3
6	3	5



.	5	2 4	8	2 4 3 5
,	6	1 3 1	2	1 3 1 2

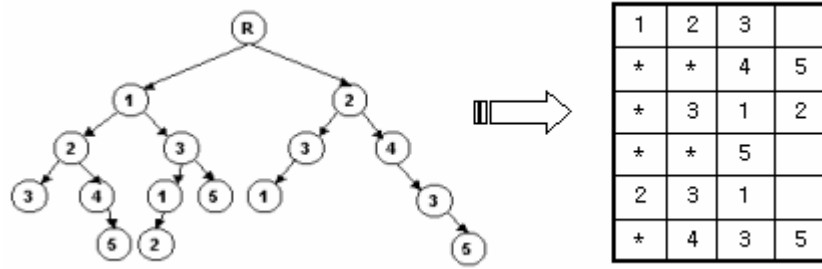
가  
50% 가 ,

### 3.5 가

가 가 가 . 3.7  
2 ,  
가 가 . 3.11 가 2  
가  
2 . 1  
ID , \* ID  
\*  
ID

, ID

5



3. 11 가

2

## IV.

### 4.1

1 MS IIS

1 ex[ , , ]

4.1

#### 4.1

	(MB)
Ex001028	15.1
Ex001029	20.1
Ex001030	21.8
Ex001101	22.6
Ex001102	29
Ex001103	23.9
Ex001104	22.6

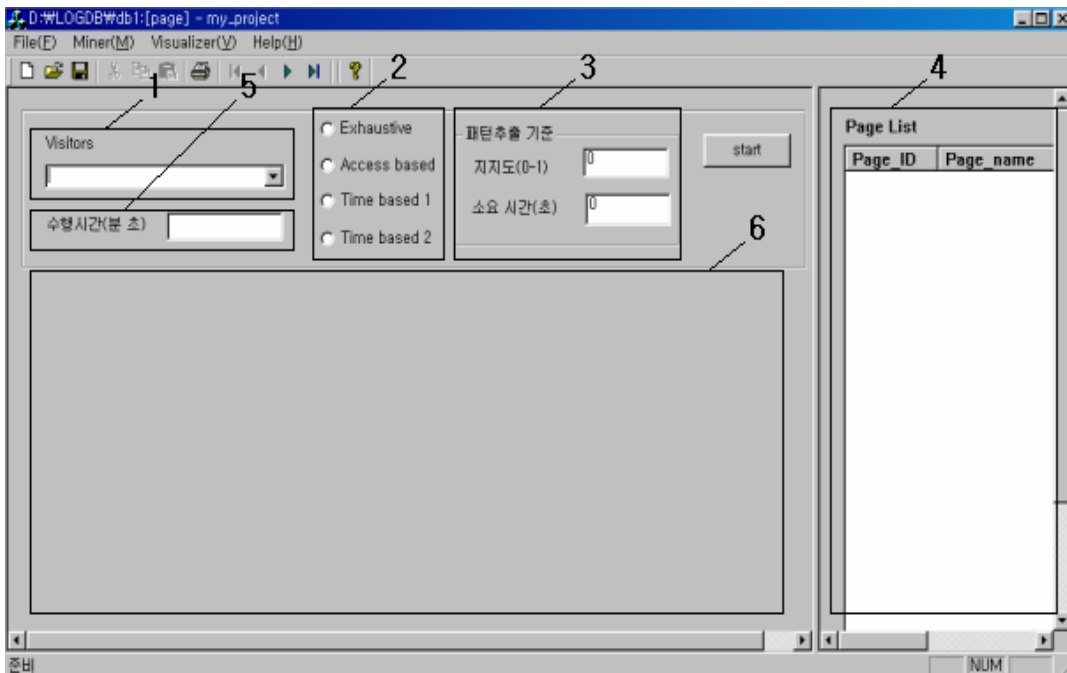
41874 ,  
10% .  
646 , 4744 가

**4. 2**

	Windows 2000 server
	MS SQL 2000
	Visual C++ 6.0

**4.2**

4.1



4.1

1 : ip

2 : 4가

- Exhaustive :
- Access based :
- Time based 1 :
- Time based 2 :

3 : 가 (0 - 1 )

4 : ID

5 :

6 : page\_id (page\_access,  
page\_duration)

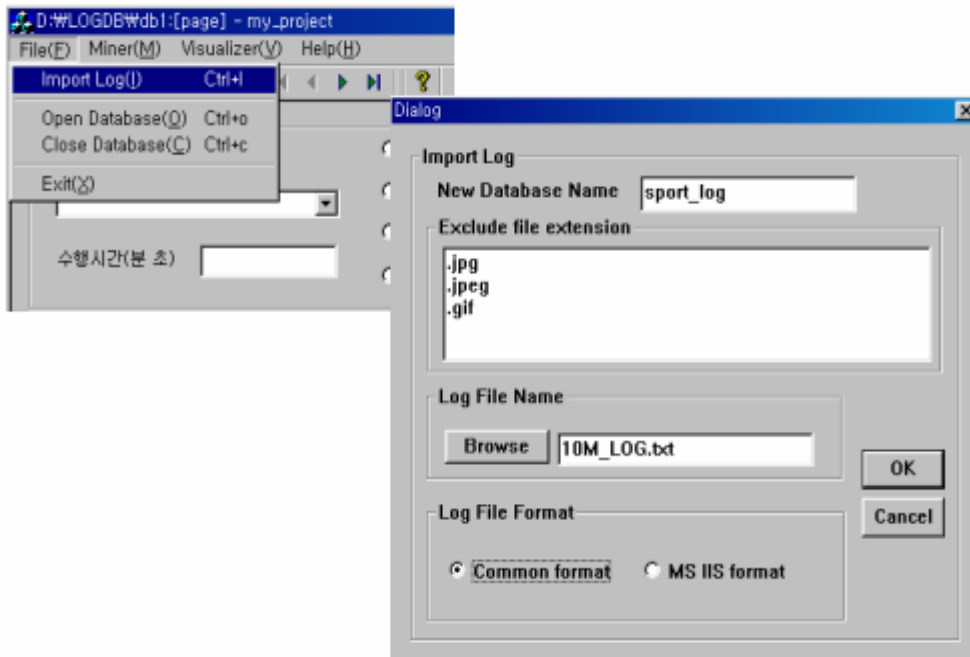
5

1 가

### 4.3

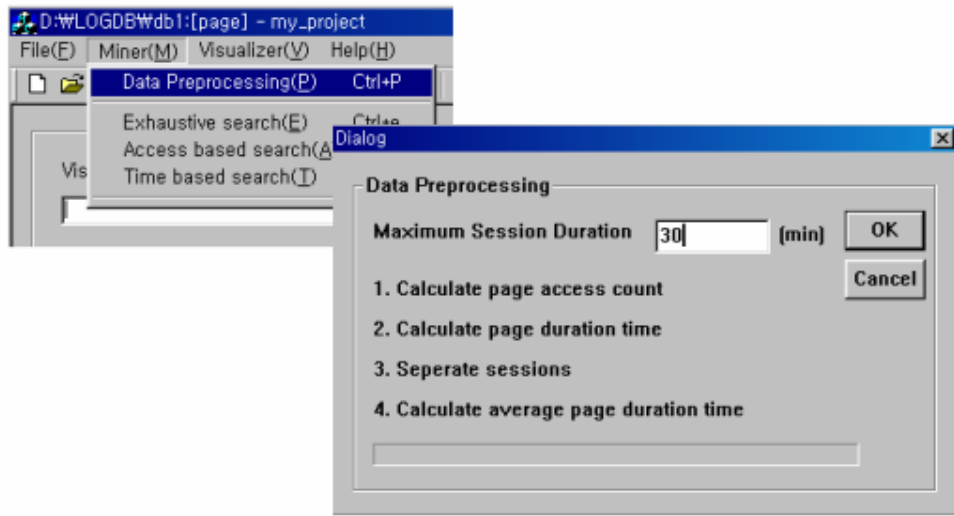
가 File

Import log 4.2 가



#### 4. 2 Import log

OK 가 가  
가 가 .  
File open database close database  
Miner Data Preprocessing 4.3 가  
가 .



### 4.3

가 Success 가 Miner  
가  
가 .

### 4.4

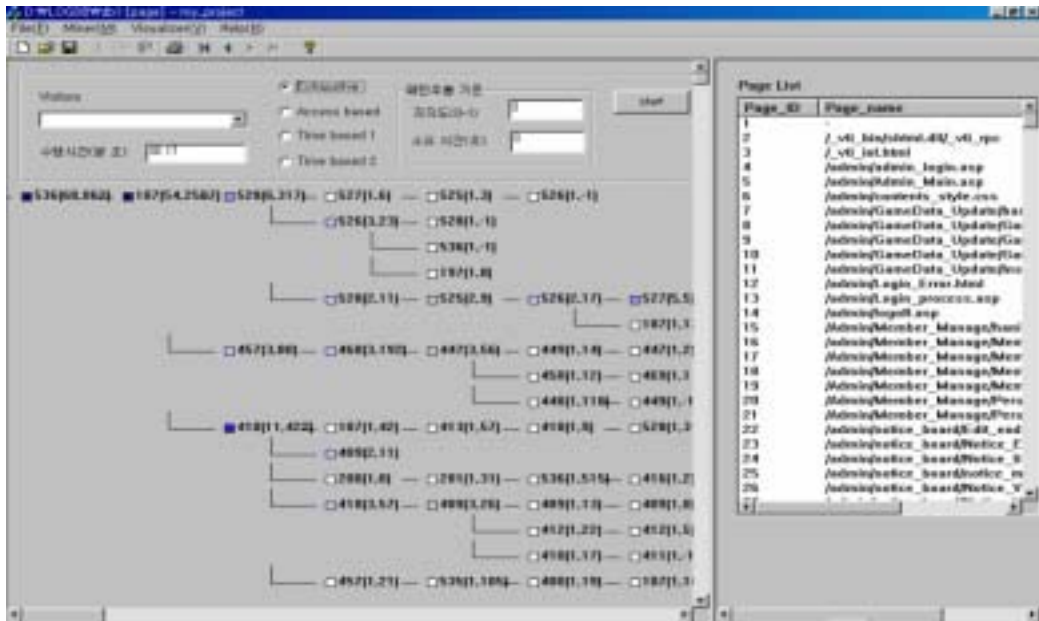
#### 4.4.1

4.4 가  
. Visitors  
. exhaustive start



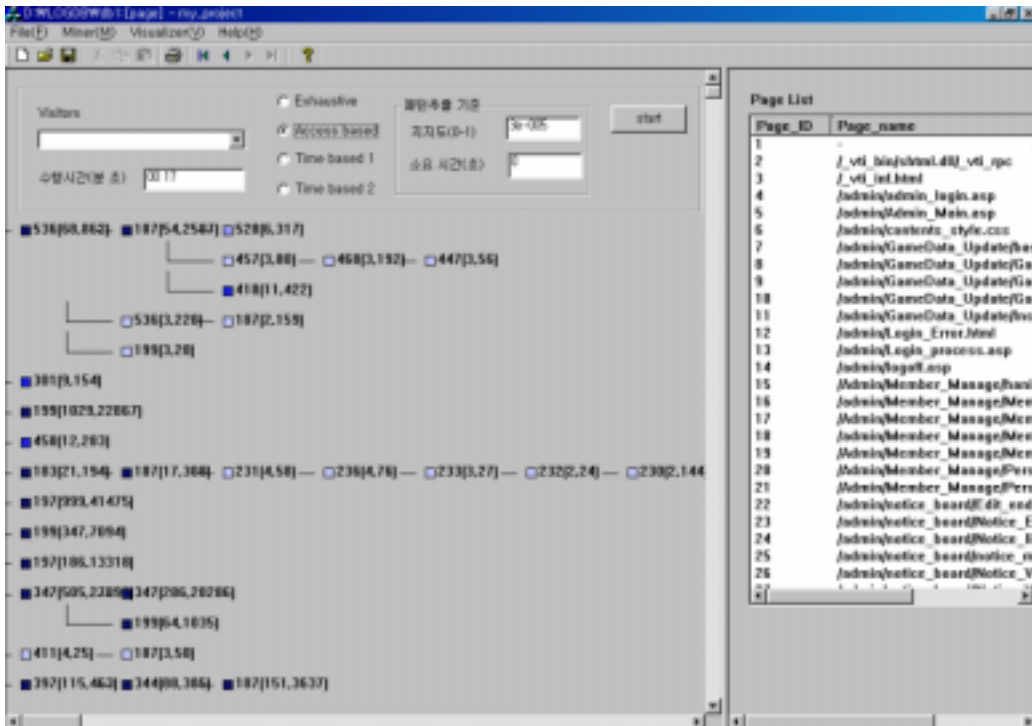
4.3

4.4 가



4.4





#### 4.6

#### 4.4.4

가

Time based 1

4.6

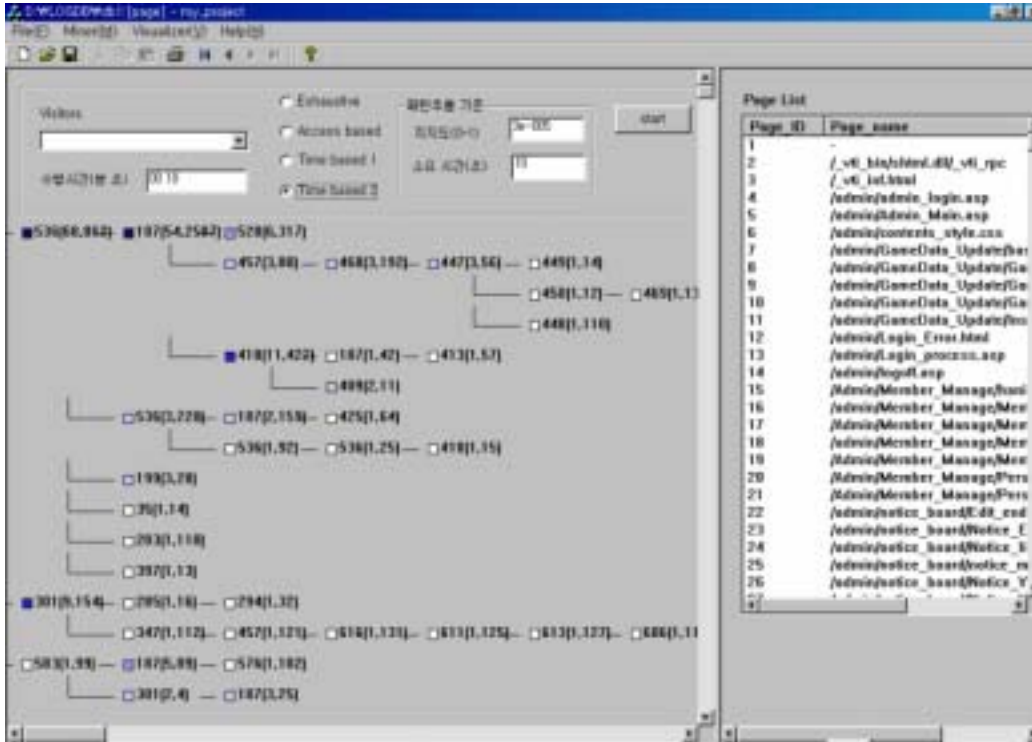
가

4.7 Time based 2

가

가 0.00003

10



4.7

4.6

10

ID 201,536

가

4.4.3

가

10

# V.

.  
 ,  
 가 가 가  
 . 가  
 가 ,  
 가  
 .  
 4가 .  
 ,  
 , 가 가  
 .  
 .  
 .

. , 가  
 가 가  
 가  
 . , 가  
 . 가 가  
 , .  
 가  
 , .

- [1] Mark Claypool, Phong Le, Makoto Waked and David Brown, "Implicit Interest Indicators", *IUI'01*, January
  
- [2] Jian Pei, Jiawei Han, Behzad Mortazavi-asl, and Hua Zhu, "Mining access patterns efficiently from web log", *2000 Pacific-Asia Conf. On Knowledge Discovery and A Data Mining*
  
- [3] Igor Cadez, David Heckerman, "Visualization of navigation patterns on a web site using model-based clustering.", *ACM SIGKDD 2000 Conf*
  
- [4] J.Borges and M.Levine, "A fine grained heuristic to capture web navigation pattern", *ACM SIGKDD 2000 Conf*, Volume 2, Issue 1, pp. 40-50
  
- [5] Myra Spiliopoulou, Lukas Faulstich and Karsten Winkler, "A data miner analyzing the navigational behaviour of web users", *In Proc. Of the workshop on Machin Learning in User Modelling of the ACAI'99 Int. Conf., Creta, Greece, July 1999*
  
- [6] Rodney Fuller, Johannes J. de Graaff, "Measuring users motivation from server log files", *Microsoft Whitepaper 1996*