

工學碩士 學位 論文

**A Study on the Restructure of Multi-Platform Databases for
MIRAS**

指導教授 辛 昌 勳

2002年 8月

韓國海洋大學校 大學院

物流 工學科

柳 秉 武

本 論 文 柳秉武 工學碩士 學位 論文 認 准

委員長 申 宰 榮 印

委 員 金 載 熏 印

委 員 辛 昌 勳 印

2002年 8月

韓國海洋大學校 大學院

物流 工學科

柳 秉 武

.....	1
1.	1
2.	1
3.	2
.....	3
1.	3
1.1	3
1.2 -	3
2.	4
2.1	4
2.2	4
2.3	5
2.4	6
3. MIRAS	7
MIRAS	10
1. MIRAS	10
2. MIRAS 가.....	11
3. (事前)	12
4.	14
.....	16
1.	16
2.	17
3.	20
MIRAS 가.....	22
1. MIRAS	22
2. MIRAS	24
3. MIRAS	28
.....	32

1 -

..... 6

1 -	4
2 -	8
3 - MIRAS	9
4 - MIRAS	10
5 - MIRAS	10
6 -	12
7 -	13
8 -	14
9 -	14
10 -	15
11 - MIRAS	16
12 -	17
13 -	17
14 -	18
15 -	18
16 -	PROC VARCLUS	19
17 -	17	20
18 -	22
19 -	23
20 -	23
21 - MIRAS	24
22 -	25
23 -	25
24 -	26
25 -	27
26 -	28
27 -	29
28 -	30
29 -	30

A Study on the Restructure of Multi-Platform Databases for MIRAS

Pyungmu Ryu

**Department of Logistics Engineering
Graduate School of Korea Maritime University**

Abstract

It is important to retrieve information that a user requires on the web. The web is an open system. The amount of information is increasing rapidly. While each of information was compiled into the database piece at a single platform in the past, it is now compiled into complicated structure at a multi-platform. Restructuring the multi-platform database is needed to efficiently retrieve information. MIRAS (Meta Information Retrieval Agent System) has a multi-platform database on the web. This study applies the classification of the existing sites' categories to restructure the database systematically. The empirical analysis shows that the suggested method is effective for information retrieval and multi-platform database restructuring. This study helps users to save on time-cost of searching information.

Keyword:

Meta Information Retrieval Agent System; Single Platform Database; Multi-Platform Database; Restructure; Category

1.

(Web) 가 , 가 .
 (Index) (Single Platform) .
 (Multi-Platform)
 가 .

2.

(MIRAS)
 가 .
 , MIRAS .
 .
 .
 MIRAS .
 가 .
 .
 51 (Category keyword) .
 51 가 가 가 .

1 51 60
 6 2 20
 (Site URL), (Title), (Web Directory),
 (Directory URL), (Engine name), (Description)
 51 (Degree of
 Relation)
 (Hierarchical Cluster Method)
 (Proportion Explained)
 가
 , MIRAS (Recall ability), (Overlap), (Time-cost)
 MIRAS (Cache System)
 가

3.

가
 가
 , MIRAS
 가 (事前)
 , MIRAS 가

1
 2 , , , , ,

1.

1.1

(Agent)

(Tool)

가 90

1.2 -

HTTP(Hyper Text Transfer Protocol) RPC

(Remote Procedure Call)

[, , 1997].

가

가

가

가

(Information Retrieval Agent System)

(Recommendation Agent)[Gerald, Valerie,

1999],

(New-contents Agent),

(Search Agent),

(Customized Agent),

(Personal-status Agent)

[lai, Yang, 1998].

2.

2.1

[Salton, McGill, 1983]. $W_{ij} = \frac{TF_{ij} \times IDF_i}{DF_i}$ (1)

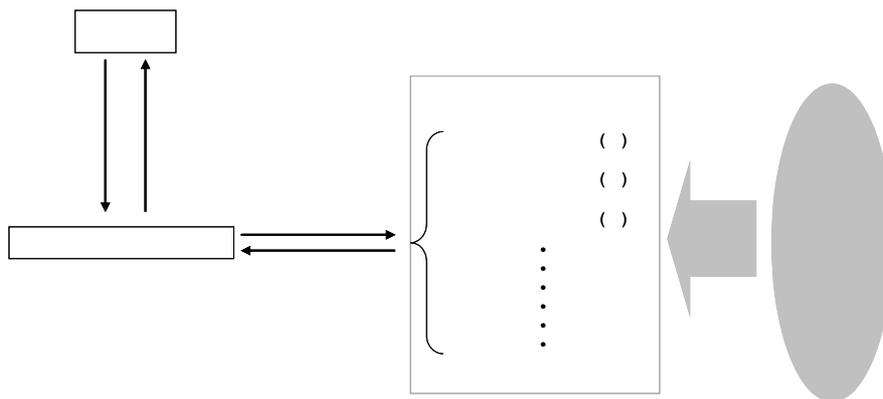
· TF_{ij} T_i 가 D_j , IDF_i 가 DF_i

$$W_{ij} \equiv TF_{ij} \times IDF_i, \quad IDF_i \equiv \log(N/DF_i), \quad N \equiv \text{ } \quad (1)$$

, IDF_i , T_i 가 DF_i 가 [Mauldin, Leavitt, 1994].

2.2

가



I -

³ <http://www.lycos.com>

1

4

2.3

가 ,

가

가

Bicchieri(1998)

(Database Selection Problem)

, 가

가

Prospero File System[Neuman, 1992], Gopher[Michael , 1992]

가

가 가

가

가

, WAIS[Kahle, Medlar, 1991]

(Directory Of Service)

, ALIWEB⁵

가

가

1

⁴ Meta-Retrieval Service

	System		
	SMART[a]		(Centroid Term-Vector)
			가
	Gloss[b]		가
			가
	Savvy[c]		
			가
	HOMIRS[d]		3
			가

: [a] Salton, 1971, [b] Gravano, Garcia-Molina, 1995, [c] Howe, Dreilinger, 1997, [d] , 2000

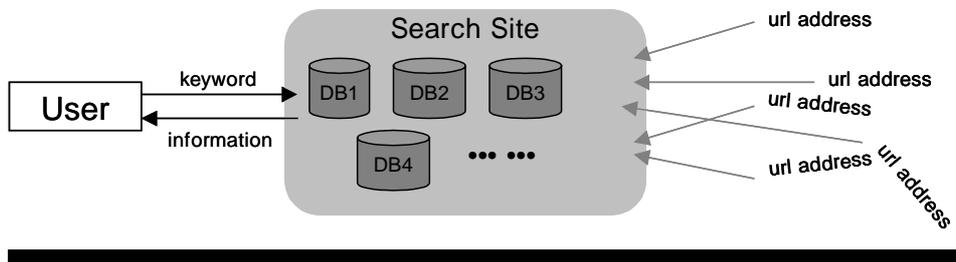
2.4

⁵ <http://aliweb.emnet.co.uk/searchform.html>

¹⁷가

가 가 URL

2



2 -

가 가

가

MIRAS

3 MIRAS

MIRAS

가

(Profile),

(Access record)

(事前)

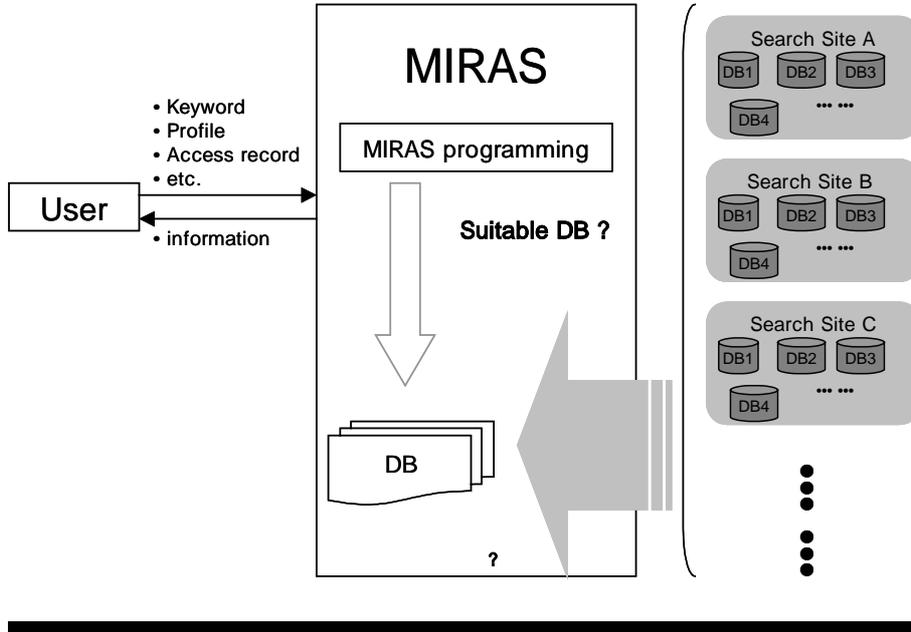
, MIRAS

(Programming) 가

가

¹⁷ <http://www.wakano.co.kr/>

MIRAS



MIRAS

1. MIRAS

MIRAS (事前) 4 MIRAS가



4 - MIRAS

5

	Category A	Category B	Category C	Category D	Category E
Keyword A	32	18	32	2	2	-
Keyword B	28	10	25	3	1	-
Keyword C	10	32	5	18	17	-
....	-	-	-	-	-	-

5 - MIRAS

MIRAS가

가 ,

, MIRAS

(3)

$$\equiv \text{Intersept} + \beta * \text{DBi} + \gamma * \text{Ri} + \varepsilon$$

β :

DBi :

γ :

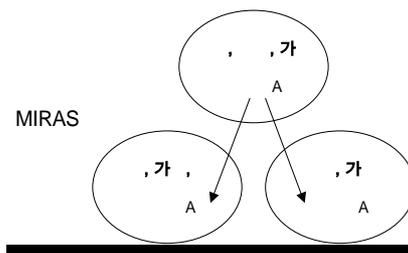
Ri :

ε :

(3)

6

(overlap)



6 -

가

(4)

(overlap)

\equiv

(a)

-

(4)

3. (事前)

가

(b1), (b9), (b16), (b24), (b31), (b38), (b45),	(b2), (b17), (b25), (b32), (b39), (b46),	(b3), (b10), (b18), (b26), (b33), (b40), (b47),	(b4), (b11), 가 (b19), (b26), (b34), (b41),	(b5), (b12), (b20), (b27), (b35), (b42), (b48),	(b6), (b13), (b21), (b28), (b36), (b43), (b49),	(b7), (b14), 가 (b22), (b29), (b37), (b44), (b50),	(b8), (b15), (b23), (b30), (b37), (b44), (b51).
--	---	---	---	---	---	---	---

8 -

4.

MIRAS

9

50

30

1							
2							
3							
4	2						
5							
6							
7							
.....							
48							
49							
50							

2001. 10

9 -

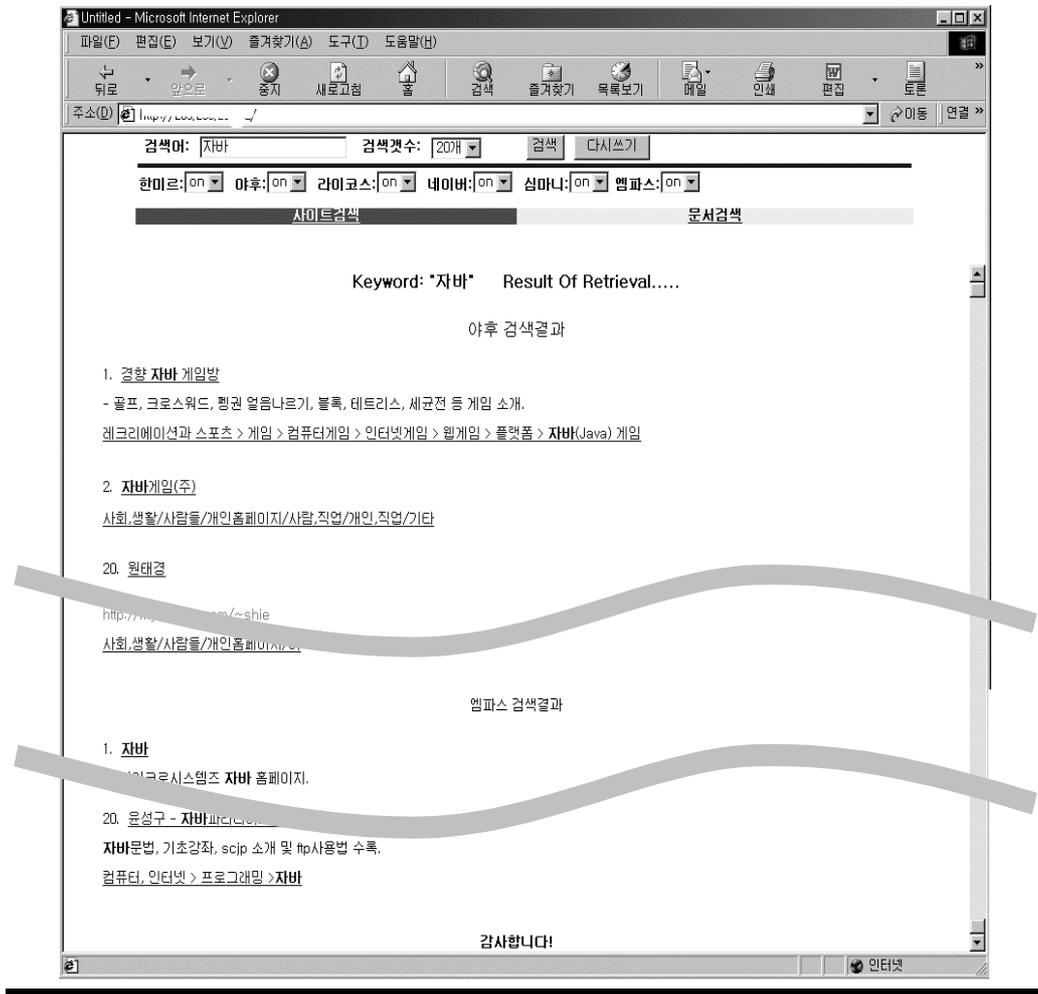
“ (3)” “ 2(4)”, “ (5)”

1.

(Java Web Server 2.0)
(Microsoft Access)

(Java Servlet)

11 MIRAS



11 - MIRAS

(Java Web Server 2.0)

(Java Servlet)

(Microsoft Access)

keyw	url	title	dirname	dirad	Engin	descrip	now	b1	b2	b3	b4	b5
업기	http://www.r	업기 패	엔터테인먼트, 예술	>유	http://dir.naver	네이버	 - 업기	Thu Dec 13 13:2	0	0	0	0
개인방	http://www.s	셀프티비	지역정보	>서울특별시	http://dir.naver	네이버	 - 개인	Thu Dec 13 21:5	0	0	0	0
성인	http://www.1	119 성인	소풍, 생활, 가족	> 인터넷	http://dir.empa	엠펙스	성인	Mon Dec 17 16:1	0	0	0	0
업기	http://??	wo.ti <b style="back <td>엔터테인먼트, 예술</td> <td>유대, 지</td> <td>http://dir.hanm</td> <td>한미르</td> <td><b style="back <td>Thu Dec 13 13:2</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td> </td>	엔터테인먼트, 예술	유대, 지	http://dir.hanm	한미르	<b style="back <td>Thu Dec 13 13:2</td> <td>0</td> <td>0</td> <td>0</td> <td>0</td>	Thu Dec 13 13:2	0	0	0	0
과외	http://007stud	007study 과	비즈니스, 경제/취업/분	업	http://dir.simm	삼미리	<font class="f	Mon Dec 17 14:1	0	0	1	1
공포	http://04yo.cc	공포 요	레크리에이션 & 스포츠	&	http://search	라이코스	공포	Tue Dec 18 00:1	0	0	0	0
시저	http://100.ami	엠펙스 백과(하)	(하) 시저(하)	카기카리	http://dir.ama	엠펙스	(h)	Thu Dec 27 03:0	0	0	0	0

12 -

12

2.

13

가

SAS Ver.8.0²²

keyword	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13	b14
	32	32	8	17	17	0	0	0	0	1	1	0	0	0
	32	32	9	5	5	0	0	0	0	1	1	0	0	0
	23	23	0	7	17	10	10	0	0	0	0	0	0	0
	0	0	0	30	32	2	2	1	0	8	7	0	0	0

13 -

14

SAS PROC VARCLUS²³

²² Statistical Analysis System

²³

	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
b1	-	1.00	0.88	-0.13	-0.12	-0.05	-0.05	-0.08	-0.07	-0.13
b2	-	-	0.88	-0.13	-0.12	-0.05	-0.05	-0.08	-0.07	-0.13
b3	-	-	-	-0.06	-0.08	-0.13	-0.13	-0.07	-0.06	-0.11
b4	-	-	-	-	0.98	0.69	0.69	-0.11	-0.11	-0.29
b5	-	-	-	-	-	0.81	0.81	-0.11	-0.13	-0.29
b6	-	-	-	-	-	-	1.00	-0.10	-0.14	-0.23
b7	-	-	-	-	-	-	-	-0.10	-0.14	-0.23
b8	-	-	-	-	-	-	-	-	0.97	0.01
b9	-	-	-	-	-	-	-	-	-	0.01

⋮

14 -

15

13

가

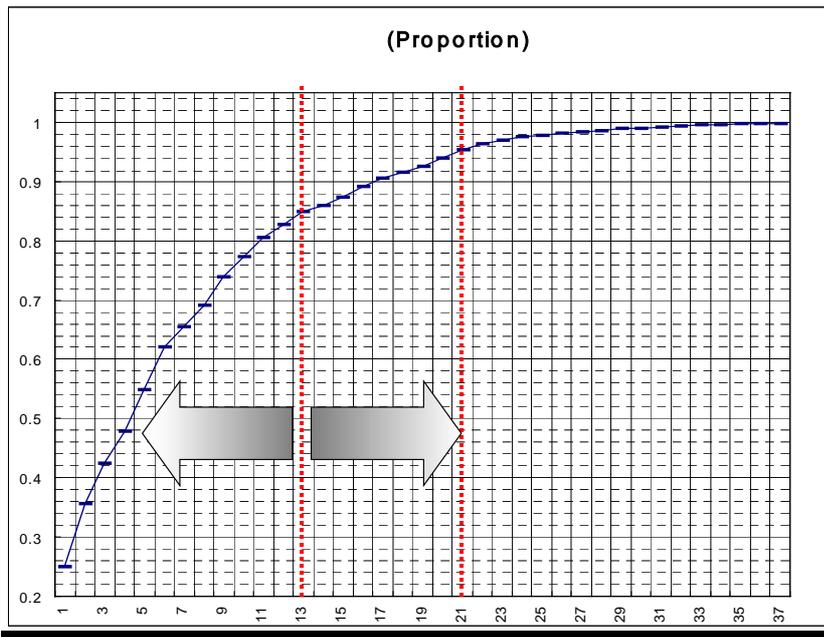
21

95%

가

21

MIRAS



15 -

16

가 가

18

20 . SAS

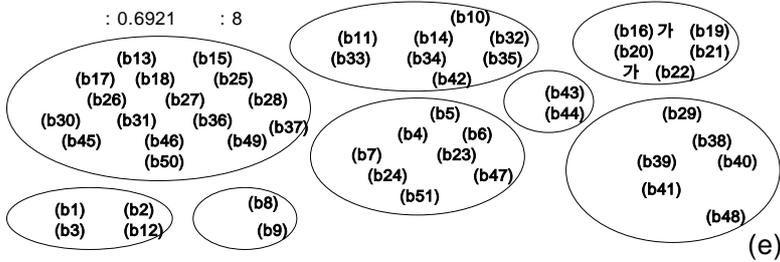
(Varclus)

```

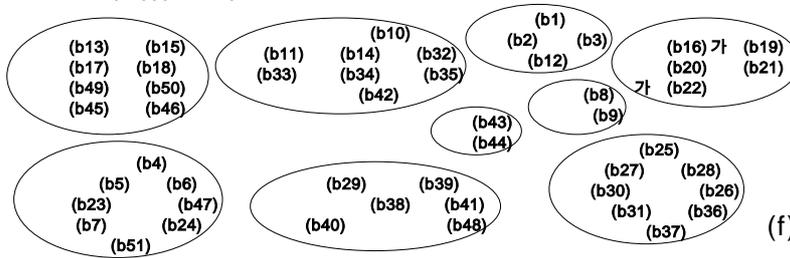
(a) : 0.4238 : 3
      :
      :
      :
(d) : 0.5480 : 7

```

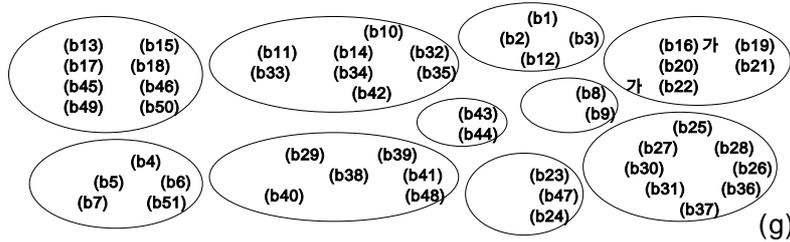
: 0.6921 : 8



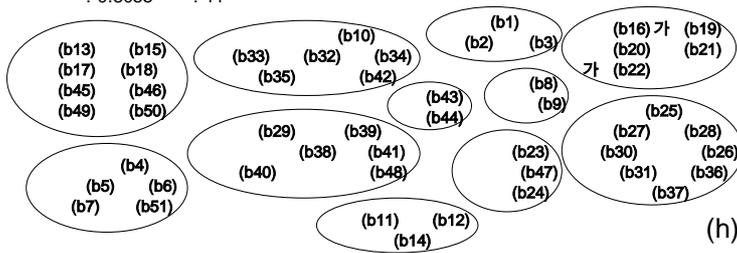
: 0.7390 : 9



: 0.7750 : 10



: 0.8055 : 11



```

(i) : 0.8281 : 12
(j) : 0.8599 : 14
(k) : 0.9060 : 17
(l) : 0.9524 : 21

```

51 (a) 0.4238
 가 3 , (l)
 0.9524 가 21 . MIRAS

가 ,
 가 가 가
 , 가 가 .

3.

. MIRAS

$$(a) \quad (A) \quad \equiv \frac{(A) \quad (a)}{(a)} \quad (5)$$

17 (5)

17

keyword								
1	0.01	0.01	0.01	0.04	0.09	0.05	0.07	0.07
2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.01	0.01	0.00	0.10	0.00	0.00	0.00	0.00
6	0.26	0.11	0.30	0.61	0.34	0.82	0.74	0.74
7	0.62	0.57	0.41	0.00	0.00	0.00	0.00	0.00
8	0.05	0.00	0.01	0.00	0.01	0.00	0.01	0.01
9	0.01	0.01	0.00	0.05	0.00	0.00	0.00	0.04
10	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.00
11	0.01	0.03	0.02	0.02	0.00	0.02	0.02	0.00
12	0.01	0.09	0.05	0.07	0.09	0.02	0.04	0.04
13	0.00	0.11	0.02	0.00	0.16	0.01	0.00	0.00
14	0.01	0.00	0.00	0.02	0.05	0.00	0.01	0.01
15	0.00	0.01	0.00	0.00	0.17	0.00	0.00	0.00
16	0.02	0.04	0.19	0.07	0.04	0.04	0.04	0.04
17	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00

... ..

. , MIRAS , 가
 . (6)
 . , MIRAS ,
 .
 (a) MIRAS $\equiv \Sigma$ (a)
 Q MIRAS가 (a) $\equiv \Sigma$ (a) (6)

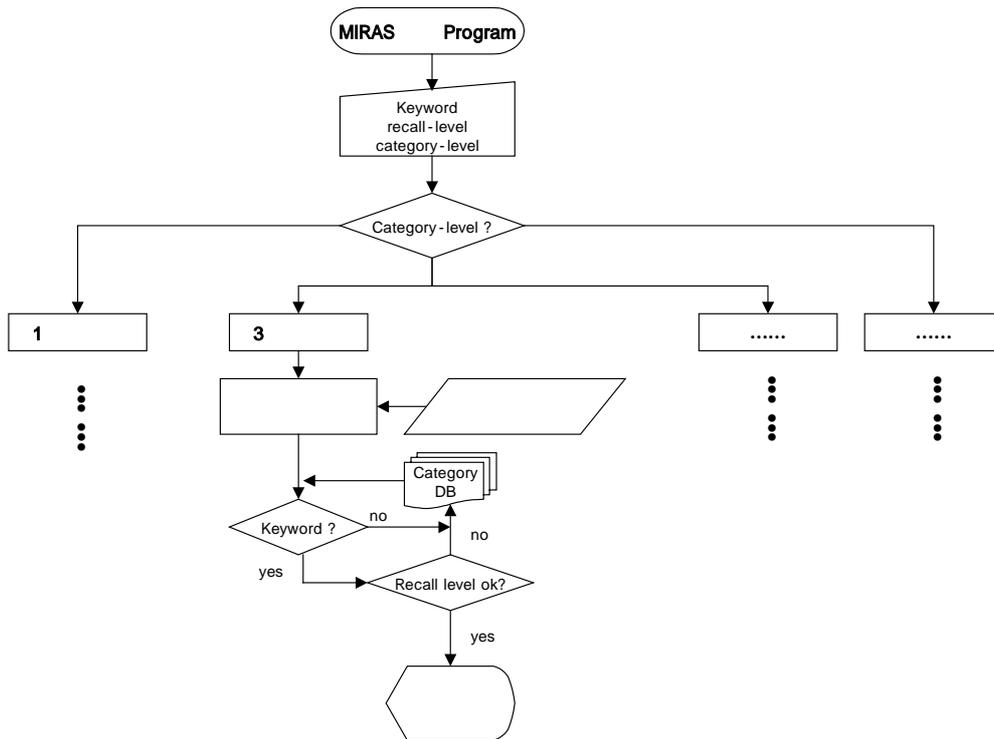
MIRAS

가

1. MIRAS

MIRAS

18



18 -

가 MIRAS 가

가

CPU: Pentium -450

: 256M

2.0

19

ms²⁴

```

Long start = System.currentTimeMillis();

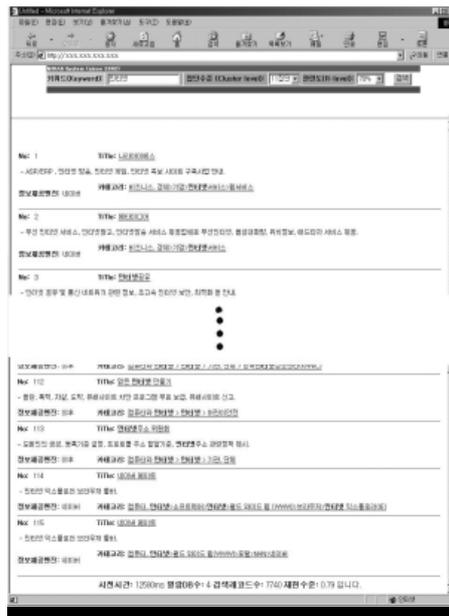
//
... ..
... ..

Long time = System.currentTimeMillis() - start;

```

19 -

20 MIRAS



20 -

가

²⁴ Millisecond(0.001)

2. MIRAS

21

(a-1), (a-2), (a-3), (a-4), (a-5), (a-6) 100%, 90%, 80%, 70%, 60%, 50%

	1	3	5	6	9	11	14	17	21
1	8400	16210	19000	17580	19380	20630	18890	18560	20540
2	8620	16150	18450	17410	18950	19990	18780	18240	19500
3	8400	16090	18620	17520	19110	20380	19390	19000	20980
4	8510	16100	18290	17420	18560	20540	18570	19390	19610
5	8510	15820	18340	17190	18620	20110	19170	18350	19440
	8488	16074	18540	17424	18924	20330	18960	18708	20014
	1	1	1	1	0.99	0.99	0.99	0.99	0.99
DB	1	3	5	5	7	8	9	8	9
	12061	13148	13001	12273	10947	10672	10036	9714	10598
	125	132	131	128	126	126	123	124	125

(a-1) 100%

⋮

(a-2) 90%

(a-3) 70%

(a-4) 60%

⋮

	1	3	5	6	9	11	14	17	21
1	-	5270	2800	2800	2360	2140	2310	2300	2190
2	-	4950	2900	2800	2530	2250	2310	2200	2140
3	-	5220	2800	2810	2250	2250	2580	2190	2140
4	-	5110	2860	2910	2250	2200	2150	2200	2200
5	-	5170	2750	2800	2310	2200	2200	2200	2200
	-	5144	2822	2824	2340	2208	2310	2218	2174
	1	0.66	0.63	0.63	0.62	0.62	0.62	0.62	0.62
DB	1	1	1	1	1	1	2	1	1
	12061	3297	938	938	493	410	410	410	410
	125	86	82	82	78	78	78	78	78

(a-5) 50%

: ms

21 - MIRAS

5

26

27

'DB'

(a)

25
26
27

. MIRAS

가 가 가 .

21 MIRAS

22 .

no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	8488	16074	18540	17424	18924	20330	18960	18708	20014	10496	9798	9424	8972	8116	9196	7438	7594	7132	6122
DB	1	3	5	5	7	8	9	8	9	2	3	3	3	4	3	2	2	2	2
	12061	13148	13001	12273	10947	10672	10036	9714	10598	8730	6639	3559	6114	4824	4824	4221	5747	5302	4012
	1	3	5	6	9	11	14	17	21	3	5	6	9	11	14	21	5	9	11
	125	132	131	128	126	126	123	124	125	122	121	121	117	116	116	116	117	113	112

no.	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
	6130	5954	12063	12211	20154	21254	19926	19018	20882	14092	15524	15864	14048	13350	12326	11910	11952	11150	10764
DB	3	2	4	5	7	8	9	8	10	3	4	5	5	6	5	5	4	4	5
	4012	3901	12063	12211	12401	12089	11453	10264	9661	11171	11399	9896	8269	7726	6748	6291	7649	6442	5902
	14	21	5	6	9	11	14	17	21	5	6	9	11	14	17	21	9	11	14
	112	112	146	146	146	145	145	144	144	145	145	144	139	138	137	132	126	121	121

no.	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
	10446	10260	9246	10620	10476	9646	9008	8906	8780	6392	6018	6296	5842	4986	5004
DB	4	4	2	3	3	3	4	3	3	1	1	2	2	3	2
	5902	5791	6362	6590	7632	6425	5885	5885	5774	4418	4098	3140	2823	2823	2823
	17	21	5	6	9	11	14	17	21	3	5	6	9	14	17
	121	121	101	101	125	120	120	120	120	83	82	99	81	81	81

22 -

23 .

, N = 53

H0: Rho=0

Prob > |r|

		DB			
	1.0000				
DB	0.8883 <.0001	1.0000			
	0.8396 <.0001	0.5842 <.0001	1.0000		
	0.0759 0.5892	0.4066 0.0025	- 0.2647 0.0555	1.0000	
	0.7224 <.0001	0.6041 <.0001	0.7673 <.0001	0.0555 0.6933	1.0000

23 -

, MIRAS

Model (1)

(Intercept)

$$= \text{Intercept} + \beta * DBi + \gamma * Ri + \varepsilon$$

Model(1) :

Intercept : , β :
 DBi : , γ :
 Ri : , ε :

22

Model(1)

24

t

$p < 0.0001$

가

28

0.9452

94.52%

Source	DF	Squares	Square	F Value	Pr>F
Model	2	1192764790	596382395	430.92	<.0001
Error	50	69198651	1383973		
Corrected Total	52	1261963441			

Root MSE : 1176.42383 R-Square : 0.9452
 Dependent Mean : 11929 Adj R-Sq : 0.9430
 Coeff Var : 9.86171

Variable	DF	Estimate	Error	T Value	Pr> t
Intercept	1	733.93645	432.72887	1.70	0.0961
DBi	1	1283.95663	86.73432	14.80	<.0001
Ri	1	0.75991	0.06371	11.93	<.0001

24 -

²⁸ R^2

Model(2) , Model(2)

$$= 733.93645 + 1283.95663 * DBi + 0.75991 * Ri + \epsilon$$

DBi :

Model(2):

Ri :

ϵ :



(7)

1689. 6167

$$Ri = /$$

Ri : 1

(7)



1690

25

Model(2)

90%

Keyword	3	DB			5	DB		
		2	8730	9935.86		3	6639	9630.8
		3	13148	14577.10		3	9845	12067.1
		3	13148	14577.10		4	12109	15071.5
		3	13148	14577.10		3	11171	13074.7
		3	13148	14577.10		3	11171	13074.7
		2	9851	10787.72		2	8907	10070.3
		2	9851	10787.72		3	9799	12032.1
		3	13148	14577.10		3	9845	12067.1
		3	13148	14577.10		3	11171	13074.7

.....

⋮

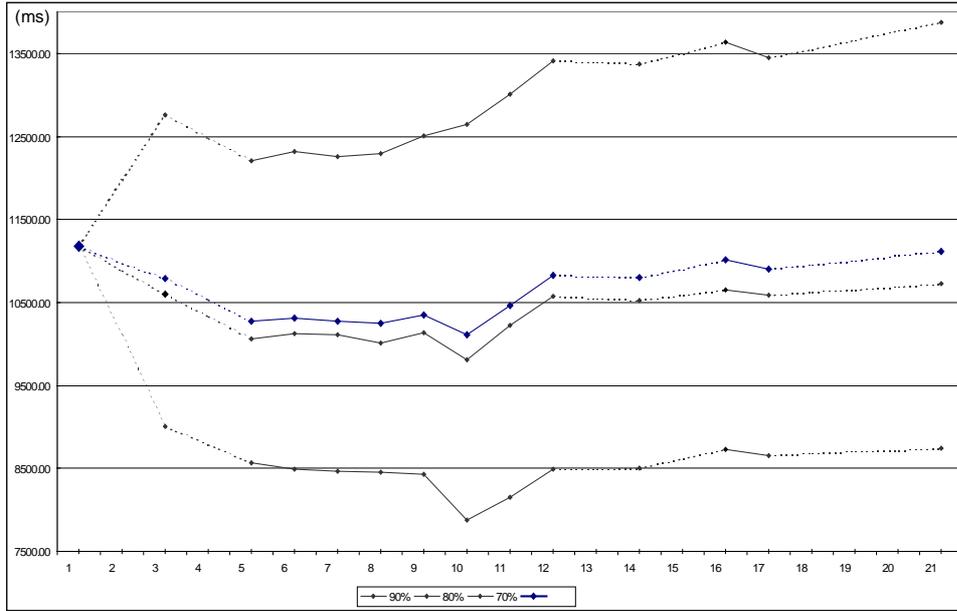


25 -

가 가 가

80% 70%

가 가 가 .



	1	3	5	6	7	8	9	10	11	12	14	16	17	21
100%	11176.33	13744.24	16312.15	17596.11	18880.07	20164.02	21447.98	22731.94	24015.89	25299.85	27867.76	30435.68	31719.63	36855.46
90%	11176.33	12751.88	12202.31	12316.79	12259.01	12290.06	12500.97	12640.70	13014.03	13414.99	13369.30	13638.96	13445.01	13877.13
80%	11176.33	10601.99	10057.81	10121.03	10110.50	10010.01	10129.11	9813.98	10217.06	10569.46	10528.25	10650.60	10587.21	10719.88
70%	11176.33	9009.47	8571.10	8496.36	8464.82	8453.90	8427.04	7882.49	8151.40	8487.70	8504.45	8733.11	8659.21	8737.72
	11176.33	10787.78	10277.07	10308.06	10278.11	10251.32	10352.37	10112.39	10460.83	10824.05	10800.67	11007.56	10897.14	11111.58

27 -

가 , 11 가 10

가 17 가 ,

17 10 .

28 .

12052 가 3 ,

13148 1096 가 가 .

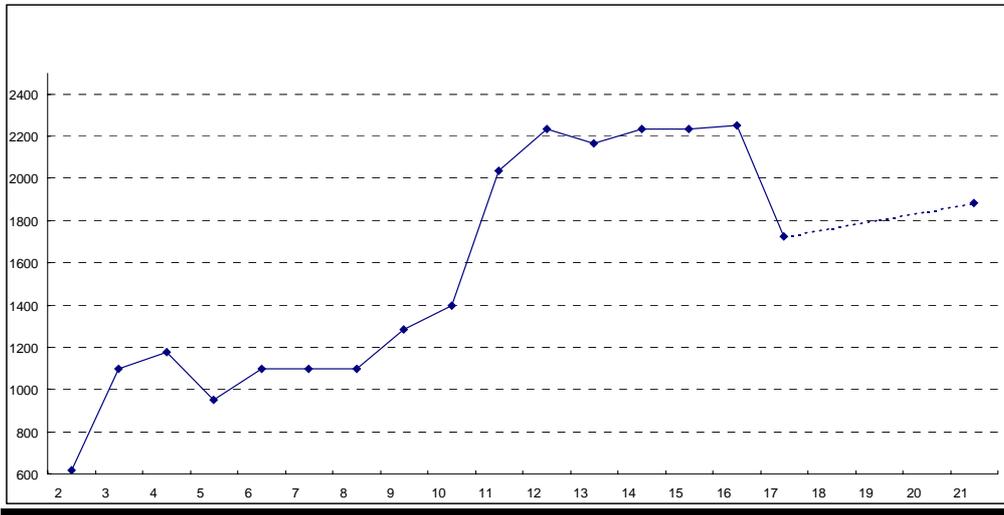
가 .

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	21	
A	12052	4703	4418	4418	4098	3450	3450	3450	1693	1693	1693	1693	1693	1693	1693	1485	1485	1485	
B		7968	5433	2441	2264	2264	2264	2247	2247	2247	1827	1827	1827	1827	1827	1827	846	257	
C			3297	5433	4809	4809	4809	4809	4809	3602	1144	3602	3602	1144	3602	3602	229	229	
D				938	938	938	493	493	493	493	3602	410	410	3602	410	410	225	225	
E					892	812	812	812	812	812	410	812	812	410	812	812	1373	1373	
F						876	876	876	876	876	812	876	796	812	796	796	3602	89	
G							445	445	1947	1947	1315	445	445	1219	445	445	410	500	
H								17	445	445	876	17	17	96	17	17	812	3491	
I									17	17	1947	1487	1487	796	1487	1487	1219	410	
J										1315	445	1315	1219	1487	1219	1219	96	613	
K											17	1144	1144	656	1144	1144	796	320	
L												656	656	445	601	601	1487	1219	
M													113	17	80	80	601	96	
N														80	96	96	445	148	
O															55	55	55	796	
P																	225	17	1487
Q																		80	601
R																			445
S																			55
T																			17
U																			80
	12052	12671	13148	13230	13001	13149	13149	13149	13339	13447	14088	14284	14221	14284	14284	14301	13778	13936	
	0	619	1096	1178	949	1097	1097	1097	1287	1395	2036	2232	2169	2232	2232	2249	1726	1884	

> : = (a) -

28 -

29



29 -

10

11

MIRAS

MIRAS

MIRAS

가

가

12

가

11

가

, 17

10

11

, 17

, MIRAS

가

MIRAS

80%

70% , 10

가

가 가

90%

가 90%

, MIRAS

,
 .
 가 ,
 ,
 (Dynamic System)
 , MIRAS
 .
 , MIRAS (Quality) 가 .
 , 가 가
 . , MIRAS 가
 . MIRAS

References

- 「 」
 , “ ”, 15 3 , 1997.
 , “ ”,
 , 1-144, 2000.
- 「 」
- Blosseville M. and Hebrail G. and Monteil M. and Penot N., “*Automatic Document Classification: Natural Language Processing, Statistical Analysis, and Expert System Techniques Used Together*” SIGIR’92, 51-58, 1992.
- Cristina Bicchieri and Martha E. Pollack, and Carlo Rovelli, “*The Potential for Cooperation among Web Agent*”, Int. J. Human-Computer Studies 48, 9-29, 1998.
- Gerald Häubl and Valerie Trifts, “*Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids*”, Marketing Science 2000 INFORMS Vol. 19, No. 1, 4-21, 2000.
- Gravano, L. and Garcia-Molina H., “*Generalizing GLOSS to Vector-Space Databases and Broker Hierarchies*”, In Proc. of the 21th Int. Conf. on Very Large Data Bases (VLDB), 78-89, 1995.
- Howe, A. and Dreilinger, D., “*Savvy Search A Metasearch Engine That Learns Which Search Engines to Query*”, AI Magazine 18(2), pp. 19 – 25, 1997.
- Hsiangchu lai and Tzyy-ching Yang, “*A System Architecture for Intelligent-Guided Browsing on the Web*”, HICSS (4), 423-432, 1998.
- Kahle B. and Medlar A., “*An Information System for Corporate Users: Wide Area Information Servers*”, The Interoperability Report 5, 11 (Nov.), 2-9, 1991.
- Mauldin M. and Leavitt J., “*Web Agent Related Research at the Center for Machine Translation*”, in Proceedings of ACM SIGNIDR-94, Aug, 1994.
- Michael F. Schwarz, Alan Emtage, Brewster Kahle and Neuman B. C., “*A Comparison of INTERNET Resource Discovery Approaches*”, Computer Systems, 5(4) 461-493, 1992.

Neuman B. C., "*The Prospero File System: A Global File System Based on the Virtual System.*"
Computing Systems, 5(4), 407-432. 1992.

Salton G., "*The SMART Retrieval System—Experiments in Automatic Document Processing*",
Prentice-Hall, Inc., Englewood Cliffs NJ, Chapters 14-17, 1971.

Salton G. and McGill M. J., "*Introduction to Modern Information Retrieval*", MacGraw-Hill
Computer Science Series, New York: McGraw-Hill, 1983.

論文

感謝

，萬端改諭

助言

激勵

가

苦樂

가