

# Recommender System for Multiple Databases Based on Web Log Mining



**Pusat Pengajian  
Pengkomputeran**

School of Computing

---

Universiti Utara Malaysia

Nurul Farhana Ismail  
Wan Hussain Wan Ishak

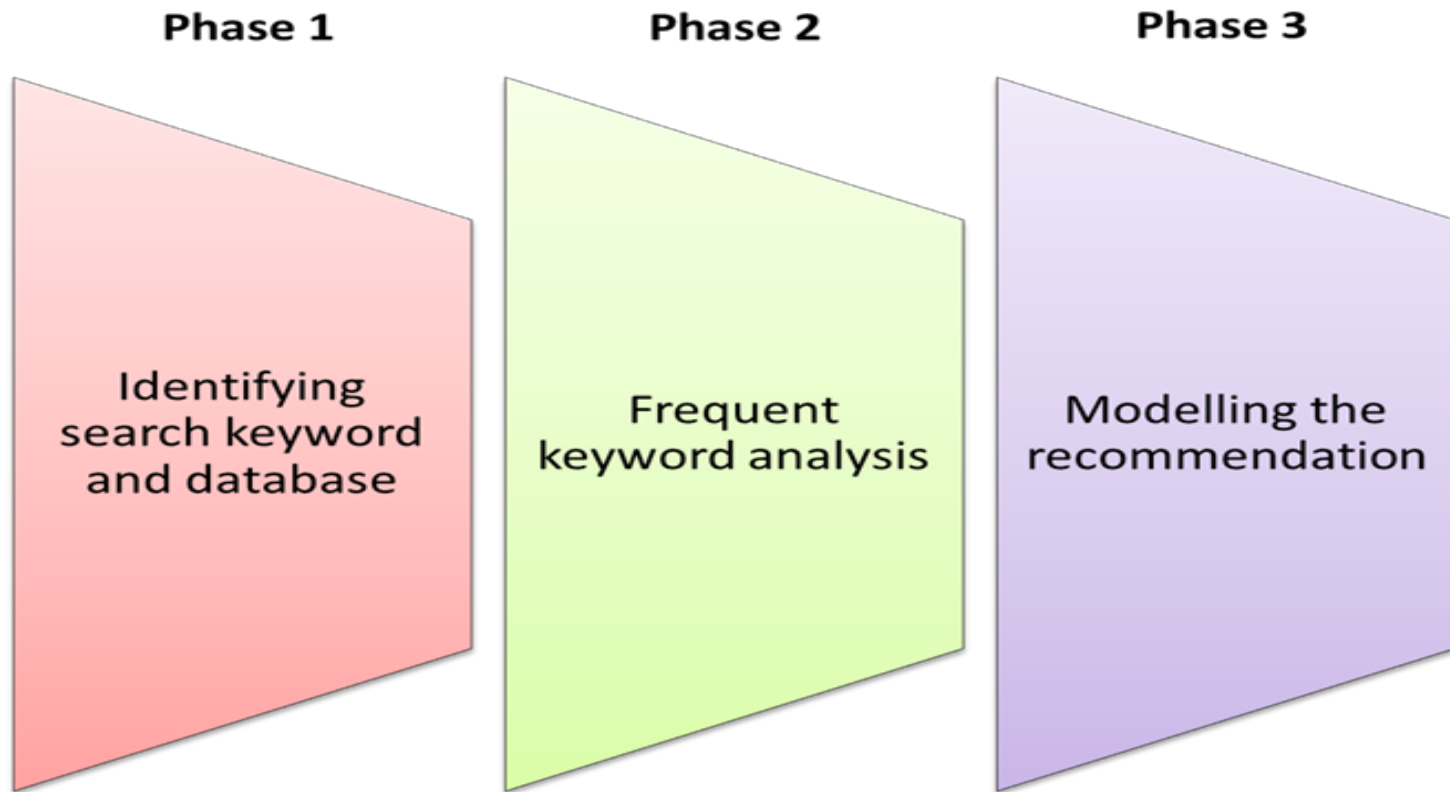
# Introduction

- Search log is an important source of information
- contains information related to the users' activities and behaviour while using the search system
- can be used to study the users' searching pattern.

# Introduction

- Information on the previous successful searchers activities can be used to assist other searchers to find what they want
- Recommender system - recommending items that are related to the user's interest
- Personalized content

# Methodology



**Phases of study**

# Methodology

- Online database (e-Resources) from Perpustakaan Sultanah Bahiyah UUM
- Data: Jan – Jun 2012 (6 months)
- Data preprocessing and query/keyword identification.

# Methodology

```
ESERV-20120100.txt - Notepad
File Edit Format View Help
10.63.253.9      58jB2AbToj3Dpwg      -      [01/Jan/2012:00:00:15 +0800]      "GET
http://syndic8.scopus.com:80/getMessage?registrationId=GAJGHDRGHCJOOBNI IAKHHAKOGANGGCNMMSLUVGJPMS HTTP/1.1"      200
27853
10.63.253.9      Acph9VLms57xDm6      801209145308      [01/Jan/2012:00:00:23 +0800]      "POST
http://web.ebscohost.com:80/ehost/resultsadvanced?sid=16b6a981-8449-43c5-a9b7-8abffb2cfece
%40sessionmgr115&vid=6&hid=106&bquery=(ethnic+AND
+competence)&bdata=JmRiPWEzaCZKyj1idGgmZGI9bmxlymsmZGI9cGRoJmRiPXBzewgmZGI9YndoJmRiPTI4aCZ0eXBlPTEmc2l0ZT1laG9zdC1saXZlJnN
jb3BlPXPndGU%3d HTTP/1.1"      302      593
10.63.253.9      -      -      [01/Jan/2012:00:00:23 +0800]      "HEAD http://eserv.uum.edu.my:80/ HTTP/1.1"      302      0
10.63.253.9      -      -      [01/Jan/2012:00:00:24 +0800]      "HEAD http://eserv.uum.edu.my:80/login HTTP/1.1"
200      6147
10.63.253.9      Acph9VLms57xDm6      801209145308      [01/Jan/2012:00:00:26 +0800]      "GET
http://web.ebscohost.com:80/ehost/pdfviewer/pdfviewer?vid=7&hid=106&sid=16b6a981-8449-43c5-a9b7-8abffb2cfece
%40sessionmgr115 HTTP/1.1"      200      31346
10.19.74.76      -      -      [01/Jan/2012:00:00:31 +0800]      "GET http://eserv.uum.edu.my:80/login?
url=http://www.emeraldinsight.com/journals.htm?issn=1741-038X&volume=19&issue=4&articleid=1723279&show=pdf HTTP/1.1"
200      6147
10.63.253.9      Acph9VLms57xDm6      801209145308      [01/Jan/2012:00:00:31 +0800]      "GET
http://content.ebscohost.com:80/ContentServer.asp?
T=P&P=AN&K=36532003&S=R&D=a3h&EbscoContent=dGJyMmHX8kSeqa84wtvhOLCmr0mep69Srqa4TbSxwXs&ContentCustomer=dGJyMPPq34rfset55%
2B55feb18YwA HTTP/1.1"      302      778
10.63.253.9      Acph9VLms57xDm6      801209145308      [01/Jan/2012:00:00:58 +0800]      "GET
http://content.ebscohost.com:80/pdf9/pdf/2009/ESF/01Mar09/36532003.pdf?
T=P&P=AN&K=36532003&S=R&D=a3h&EbscoContent=dGJyMmHX8kSeqa84wtvhOLCmr0mep69Srqa4TbSxwXs&ContentCustomer=dGJyMPPq34rfset55%
2B55feb18YwA HTTP/1.1"      200      323054
10.2.4.189      -      -      [01/Jan/2012:00:01:18 +0800]      "GET http://eserv.uum.edu.my:80/ HTTP/1.1"      302      0
10.63.253.9      yQnbh43lFm9ANy      A01073984      [01/Jan/2012:00:01:29 +0800]      "GET
http://www.jstor.org:80/abs/addons?uri=http%3A%2F%2Fwww.jstor.org.eserv.uum.edu.my%2Fstable
%2Fi314133&groups=anonymous&scripts=plantlinks&_ =1325347351898 HTTP/1.1"      200      6586
10.63.253.9      yQnbh43lFm9ANy      A01073984      [01/Jan/2012:00:01:30 +0800]      "GET http://www.jstor.org:80/rx?
```

Raw Data

# Methodology

GET http://www.emeraldinsight.com:80/search.htm?st1=sukuk+structure&ct=all&ec=1&bf=1 HTTP/1.1
GET http://proquest.umi.com:80/pqdweb?RQT=302&COPT=REJTPUcyODcrM2lxMCZJTIQ9MCZWRVI9Mg==&clientId=28929&cfc=1 H
GET http://www.emeraldinsight.com:80/journals.htm?issn=1753-8394&volume=2&issue=4&articleid=1826941&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com:80/journals.htm?issn=1759-0817&volume=1&issue=1&articleid=1858421&show=pdf HTTP/1.1
GET http://proquest.umi.com:80/pqdweb?SQ=continuous+education+and+retention&DBId=G647&date=ALL&onDate=&beforeDate=&af
GET http://www.emeraldinsight.com:80/search.htm?st1=sukuk+structure&ec=1&bf=1&ct=jnl HTTP/1.1
GET http://proquest.umi.com:80/pqdweb?SQ=continuous+education+factors&DBId=G647&date=ALL&onDate=&beforeDate=&afterDate
GET http://eserv.uum.edu.my:80/connect?session=s5kK6uj4JmxDFOGY&url=menu HTTP/1.1
GET http://eserv.uum.edu.my:80/menu HTTP/1.1
GET http://www.emeraldinsight.com:80/journals.htm?issn=0828-8666&volume=26&issue=1&articleid=1846113&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com:80/search.htm?st1=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=2 HTTP/1.1
GET http://eserv.uum.edu.my:80/public/rightmenu.htm HTTP/1.1
GET http://www.emeraldinsight.com:80/search.htm?st1=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=2 HTTP/1.1
GET http://www.emeraldinsight.com:80/journals.htm?issn=1753-8394&volume=2&issue=2&articleid=1795358&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com:80/journals.htm?issn=1759-0817&volume=1&issue=1&articleid=1858422&show=pdf HTTP/1.1
GET http://www.emeraldinsight.com:80/search.htm?st1=sukuk+structure&ec=1&bf=1&ct=jnl&nolog=167626&page=3 HTTP/1.1
GET http://www.emeraldinsight.com:80/search.htm?st1=attitude&ec=1&bf=1&ct=jnl&nolog=814577&page=2 HTTP/1.1
GET http://www.emeraldinsight.com:80/search.htm?st1=changing+attitude&ct=all&ec=1&bf=1 HTTP/1.1
GET http://eserv.uum.edu.my:80/connect?session=sjZEm3XIn7jzCwY3&url=menu HTTP/1.1

Keyword Detection

# Methodology

- The recommender system model is design and develop to illustrate the functionality of the actual recommender system.

Recommender System for E-Resources

Enter a keyword:

Most user find articles related to "management" from : Serial Solution

Alternatively, user may find related articles from the following database:

1. ieee
2. ScienceDirect
3. Emerald
4. OnlineLibrary
5. JSTOR
6. SCOPUS
7. Eric
8. Ebscohost
9. JournalCambridge
10. XREF



# Findings

NO	DATABASE	NUMBER OF KEYWORDS	FREQUENT KEYWORDS	SUM OF KEYWORDS
1	ebscohost	1552	217	1835
2	Emerald	912	74	1011
3	Eric	462	341	1081
4	IEEE	<b>3418</b>	<b>808</b>	<b>4862</b>
5	JournalCambridge	1080	67	1148
6	JSTOR	1321	150	1503
7	OnlineLibrary	956	80	1044
8	ScienceDirect	1561	255	1906
9	SCOPUS	1226	124	1379
10	SerialSolution	2408	363	3043
11	XREF	307	18	334
	TOTAL	6143	2497	<b>19146</b>

Summary of the keywords

# Findings

NO.	DATABASE	KEYWORD FREQUENCIES									
		1	2	3	4	5	6	7	8	9	>10
1	ebscohost	1335	168	40	6	1	1	0	0	1	0
2	Emerald	838	54	16	3	1	0	0	0	0	0
3	Eric	121	121	194	7	8	10	0	1	0	0
4	IEEE	2610	516	151	60	39	18	9	3	3	9
5	JournalCambridge	1013	66	1	0	0	0	0	0	0	0
6	JSTOR	1171	130	11	6	3	0	0	0	0	0
7	OnlineLibrary	876	75	3	1	1	0	0	0	0	0
8	ScienceDirect	1306	196	41	11	3	2	2	0	0	0
9	SCOPUS	1102	107	8	6	3	0	0	0	0	0
10	SerialSolution	2045	240	59	36	9	9	3	1	1	5
11	XREF	289	12	3	3	0	0	0	0	0	0
	TOTAL	12706	1685	527	139	68	40	14	5	5	14

Keywords Frequencies

# Findings

NO.	KEYWORD	NUMBER OF SOURCE DATABASES	KEYWORD FREQUENCY				
			TOTAL	MIN	MAX	AVERAGE	MEDIAN
1	management	11	68	1	20	6.18	4
2	organization	9	44	2	12	4.89	5
3	network	10	39	1	15	3.90	2
4	technology	10	37	1	10	3.70	2
5	performance	11	36	1	7	3.27	3
6	entrepreneur	11	35	1	10	3.18	2
7	communication	10	34	1	14	3.40	2
8	organizational	10	34	1	6	3.40	3
9	information	10	32	1	9	3.20	2.5
10	system	11	32	1	11	2.91	2

10 Highest Frequent Keywords

# Conclusion

- The most popular database is IEEE database
- searchers are actively using an online database to search and access information especially academic articles
- searchers are not just focusing on one database