2-1-2020

Evaluating the Effectiveness of Keyword, Phrase, Boolean Operator [AND] in Bibliographic Data Access and Retrieval for the Subject Chemistry

saravanan T
Annamalai University, sarlib@live.in

Evaluating the Effectiveness of Keyword, Phrase, Boolean Operator [AND] in Bibliographic Data Access and Retrieval for the Subject Chemistry

Dr. Saravanan, T.

Associate Professor, Department of Library and Information Science, Annamalai University, Tamilnadu, India. E-mail*: sarlib@live.in / tsarlib@gmail.com*
ORCID ID: https://orcid.org/0000-0003-0123-6210

Abstract

The present study focuses on information search and retrieval of the bibliographical data related to the key terms that fall in the subject chemistry. The search techniques include simple keywords, phrase terms and Boolean operator "AND". Further, study encompasses the Lancaster's 2 x 2 Table matrix to trace the recall (R) and precision (P) ratio for the obtained records that are retrieved from the bibliographic database using the same search techniques. To fulfil the aims of the present study, a bibliographic database which contains the doctoral theses belonging to various subjects submitted to Annamalai University has been vouched. The database consists of 1641 records and was built using Unesco's Java CDS/ISIS (Jean-Claude Dauphin, 2016). The queries used in this study are acid, correlation, kinetics, oxidation, kinetics of oxidation, (acid AND chemistry), (kinetics AND oxidation), and (kinetics AND oxidation AND correlation).

Keywords: Evaluation, Search, Retrieval, chemistry, keyword, phrase, Boolean, AND, unesco, *Java* CDS/ISIS, *J-ISIS*, Jean-Claude Dauphin, Database, Lancaster, Recall, Precision

Introduction:

Evaluation can be conducted in two different ways: namely managerial based and users based. According to Cleverdon (1966), precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of the total amount of relevant instances that were actually retrieved. Vickery (1970) states that recall is the proportion of such references that are retrieved in a search while precision is the ability of the system to screen our irrelevant references. Lancaster (1979) proposed five evaluation criteria where recall and precision were defined as ability of the system to retrieve wanted items (i.e. recall) and ability of the system to avoid retrieval of unwanted items (i.e. precision). Information scientists usually conduct the users' centred evaluation. Users' needs can be fulfilled by way of providing a good information retrieval system. Saravanan,T and others (2010b) conducted a study on search techniques awareness of the PG students and revealed that 102 (13.26%) respondents were aware of the AND operator which received first position with the Mean score 16.83. However, satisfying all the users' demands won't be possible as users' requirements always differ from person to person. Technological advancements make the

users' data access and retrieval easy and compact. Ability of the modern information retrieval systems brings the broader terms into narrow form. Search queries may be refined as per the needs of the information seekers. Information retrieval systems offer the features to its users in the form of simple search and advanced level search. Information seekers may choose the search options based on their access and retrieval skills.

***Java* CDS/ISIS:**

Java CDS/ISIS was initiated by UNESCO and further developed by Jean-claude Dauphin. According to Jean-Claude Dauphin (2016), J-ISIS is not an Integrated Library System (ILS) as ABCD, it's a non relational (No SQL) database management system that uses the ISIS concepts and that is particularly well suited for the storage and retrieval of bibliographic information. A Bibliographic database with the records count 1641 has been built using J-ISIS and used to test the effectiveness of key words, phrase and Boolean operator "AND" along with the phrase search. The Bibliographic database was structured according to the standards of CCF (Common Communication Format). The database encompasses 17 fields, subfields and repeatable field.

Webopedia gives the definitions for keyword, phrase and Boolean search as shown below.

Keyword Search: A type of search that looks for matching documents that contain one or more words specified by the user.

For example, a keyword search could be ***kinetics***. This would limit the search results to only those documents containing the two keywords.

Phrase Search: A type of search that allows users to search for documents containing an exact sentence or phrase, rather than single keywords.

For example, a phrase search could be ***kinetics of oxidation***. This would limit the search results to only those documents containing the phrase.

Boolean Search: Boolean search is a type of search allowing users to combine keywords with operators (or modifiers) such as AND, NOT and OR to further produce more relevant results.

For example, a Boolean search could be ***Chemistry* AND *India***. This would limit the search results to only those documents containing the two keywords.

Objectives:

1. To search and retrieve the bibliographic records related to the key terms, phrase, and Boolean Operator [AND] used combined terms belonging to the subject Chemistry.

2. To analyze the Recall and Precision values for the search results rendered by the key terms, phrase, and Boolean Operator [AND].

Research Design:

A bibliographic theses database was built by Java CDS/ISIS for the present study. Database encompasses 1641 doctoral bibliographic records related to various subjects. A few queries were applied by the select users from *Chemistry*, and the obtained results were analyzed using Lancaster Table Matrix (Recall & Precision). Simple phrase search and Boolean search technique were used in this study. The Boolean search included 'AND' operator only and no other operator is used. Other operators and records table for the key terms *acid, and chemistry* are not considered here in order to keep the length of this paper within the limitation. For illustration purpose a few tables have been generated for the key terms *kinetics, oxidation,* and *correlation.* Throughout the study wherever the term 'AND' (Alphabet Capital) exists that refers to the Boolean operator.

Lancaster Table Matrix:

The recall (R) and precision (P) values can be calculated using the given Lancasters' 2x2 matrix as shown in the Table-1.

Table-1: Lancaster Table Matrix (2 x 2):

Records	Relevant	Non-Relevant	Total
Retrieved	a	b	(a+b)
Not Retrieved	c	d	(c+d)
Total	(a+c)	(b+d)	(a+b+c+d)

The small alphabets are represented as shown below.

a = hits, b = noise, c = missed and d = rejected

The formula for Recall and Precision is given below.

Recall (R) = a/(a+c) x 100 ; Precision (P) = a/(a+b) x 100

So, in this case, precision is "how useful the search results are", and recall is "how complete the results are".

A few Users from the discipline of Chemistry were asked to apply a few search queries using the key terms, phrase and Boolean operator "AND" to measure the recall and precision values for the search results. The select search queries are as shown here.

Queries applied:

Acid, correlation, kinetics, oxidation, kinetics of oxidation, (acid AND chemistry), (kinetics AND oxidation), and (kinetics AND oxidation AND correlation).

Query-1: *acid*

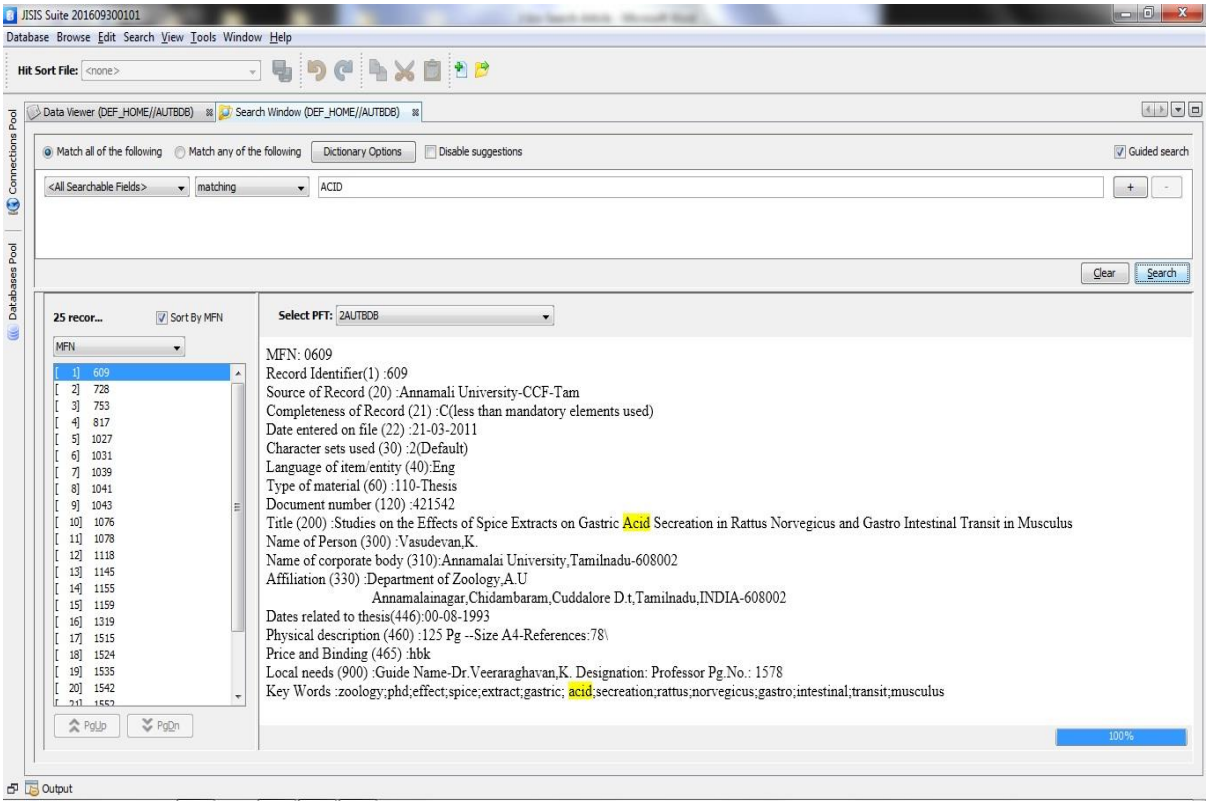

Figure-1: Query- *acid*

The term *"acid"* is applied by the user from the subject *Chemistry*, as a search query that retrieves 25 records from the database (Figrue-1) where 11 records are relevant to the users' demand while rest of the 14 records are found as non-relevant which fall in different subjects namely *Bio-chemistry (8 records), physics (3 records), agri-chemistry (1 record), zoology (1 record) and marine biology (1 record).*

$R = [a/(a+c)] \times 100$; $R = [11/(11+0) \times 100] = 100\%$; Recall= 100% ,

$P = [a/(a+b)] \times 100$; $P = [11/(11+14) \times 100] = 78.57\%$; Precision= 78.57%

For this query the observed results indicate that there is 100% completeness of output, and relevance of output has secured 78.57%.

Query-2: *acid* AND *chemistry*

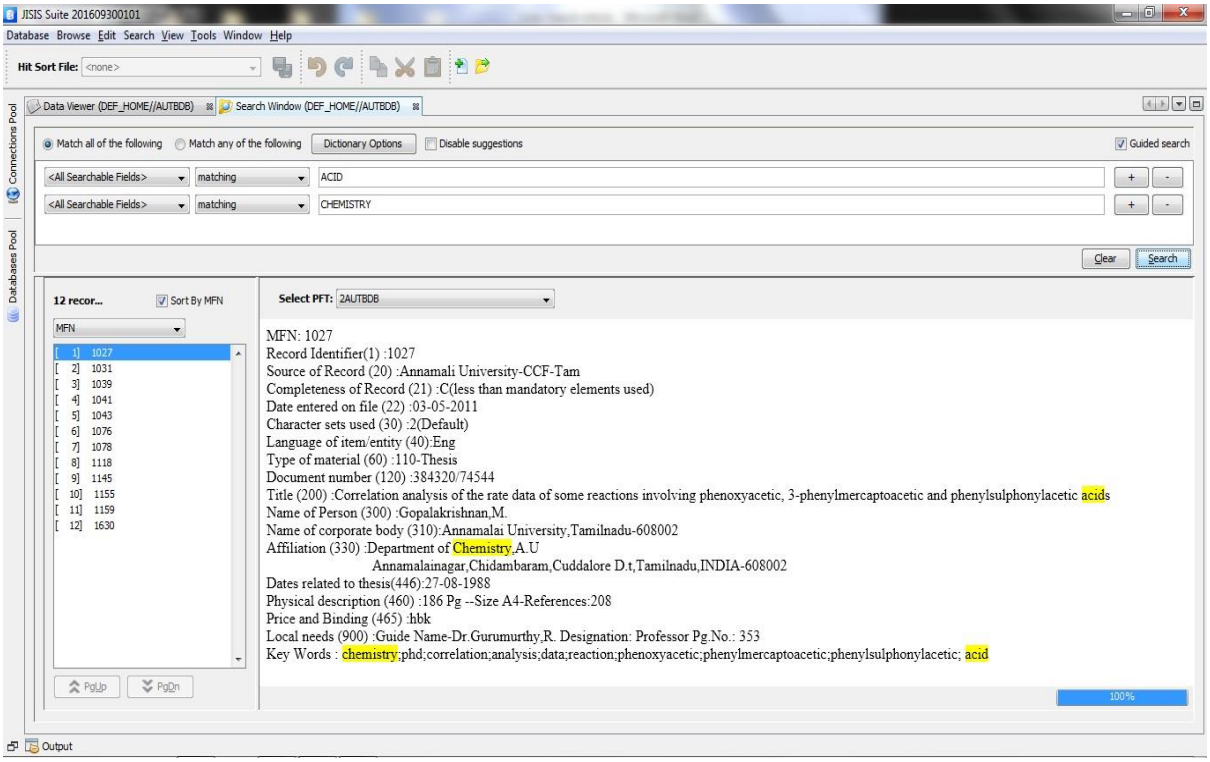

Figure-2: Query- *acid* AND *chemistry*

The previous query is refined by adding one more key term *"chemistry"* with the term *"acid"* (Figure-3). This query retrieved 12 records where 11 records are relevant while 1 record falls under non-relevant. The term *"acid"* is found in 25 records (Figure-1), while the term *"chemistry"* is traced in 199 records. After filtering the 11 records which are related to the subject chemistry, the total number of records is 188 (199-11). For the term *"acid"* the remaining records are 14 (25-11). Hence, the total number of records in the system is 202 (188+14):

$$R = [a/(a+c)] \times 100 ; \quad R=[11/(11+191) \times 100] = 5.45\% ; \quad \text{Recall} = 5.45\% ,$$
$$P = [a/(a+b] \times 100 ; \quad P=[11/(11+1) \times 100] = 91.66\%; \quad \text{Precision} = 91.66\%$$

The combined terms have retrieved the items where the ratio for the completeness of output is traced as 5.45%, and the relevance of output is 91.66%.

Queries-3: *correlation, chemistry* [subject: chemistry]

The users from chemistry subject applied the term *"correlation"* that retrieved 25 records (Figure-3), where 22 records are relevant to the users' demand while rest of the 3 records are

found as non-relevant [MFN: 793, 817, and 845]. The term *"chemistry"* is traced in 199 records. After filtering the 22 records which are related to the subject chemistry, the total number of records is 177 (199-22). However, there are still 3 more records related to the term *"correlation"* that falls under some other subject. Hence, the total number of records in the system is 180 (177+3)

R= [*a*/ (*a*+*c*)] x 100 ; R=[22/(22+158) x 100] = 13.92% ; Recall= 13.92% ,
P= [*a*/ (*a*+*b*] x 100 ; P=[22/(22+3) x 100] = 88% ; Precision= 88%

It is observed that the completeness of output is 13.92%, whereas the relevance of output depicts 88%.

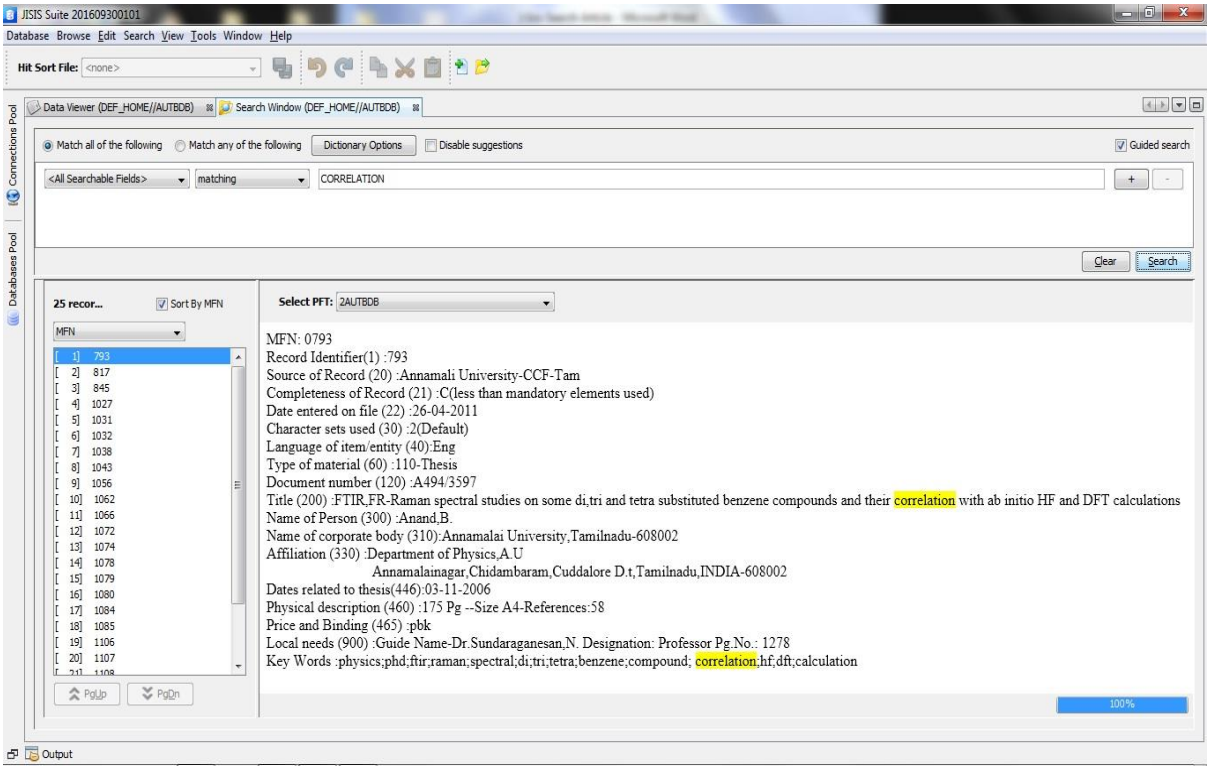

Figure-3: Query- *correlation*

Query-4: *correlation* AND ***chemistry***

The query *"correlation* AND *chemistry"* (Figure-4) is applied to search the required items. This query retrieves 22 records, where 15 are identified as relevant to the users' demand while rest of the 7 records are found as non-relevant. It is observed from the previous search results that the term *"correlation"* exists in 25 records, where 22 records are relevant to the subject chemistry. Also, it is clear from the previous results that the term *"chemistry"* is available in 199 records. After filtering the 22 records which are related to the subject chemistry, the total number of records is 177 (199-22). However, there are still 3 more

records related to the term *"correlation"* that falls under some other subject. Hence, the total number of records in the system is 180 (177+3):

R= [*a*/ (*a*+*c*)] x 100 ; R=[15/(15+165) x 100] = 8.33% ; Recall= 8.33% ,
P= [*a*/ (*a*+*b*] x 100 ; P=[15/(15+7) x 100] = 88% ; Precision=68.18%

It is observed that the completeness of output is 8.33%, whereas the relevance of output depicts 68.18%.

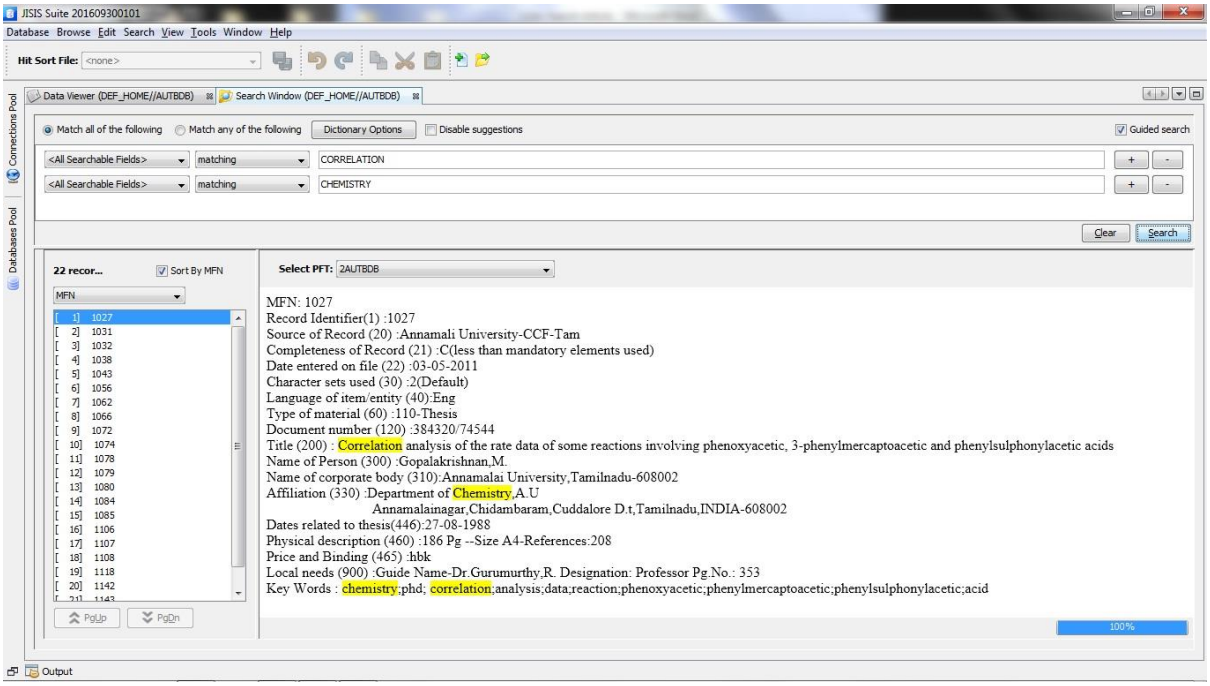

Figure-4: Query- *correlation* AND *chemistry*

Query-5: *kinetics, oxidation, kinetics* AND *oxidation* (excludes *"correlation"*)

A search query using the term "*kinetics*" is used to retrieve the relevant records from the database. In return, 17 relevant hits are retrieved (Figure-5). One more Query "*oxidation*" is applied to retrieve the records which are related to the key term. 41records are matched and retrieved (Figure-6). However, those terms do not satisfy the users' demands. Hence, combinations of both the terms "*kinetics* AND *oxidation*" are used to retrieve the relevant items (Figure-7). This query retrieves 12 records, where 6 records are identified as relevant to the users' demand and rest of the 6 records fall out of the users' demands. The details associated with the records counts are given in Table-2.

Search query: *Kinetics* [Matched with 17 terms available in 17 records]

Search query: *Oxidation* [Matched with 41terms available in 41 records]

Of the 17 records, 12 records have the term *"oxidation"*. Hence, a simple calculation may help to trace the number of terms and number of records (17-12=5 records [MFNs: 674, 999, 1106, 1107, and 1154] & 41-12= 29 records). See the Table-2 for a better understanding of

single as well as combination of the two different search terms and their matched items. The matched records are indicated by *bold* and *.

29 records with the same MFN (Master File Number) for both the terms *"oxidation"* and *"kinetics"* are traced. In addition to that 5 more records with the MFNs 674, 999, 1106, 1107, and 1154 have been traced where the term *"kinetics"* exists and gives rise to a sum of 34 records *(*29+5=34). i.e. both 58 terms (41+17) are found in 34 records.

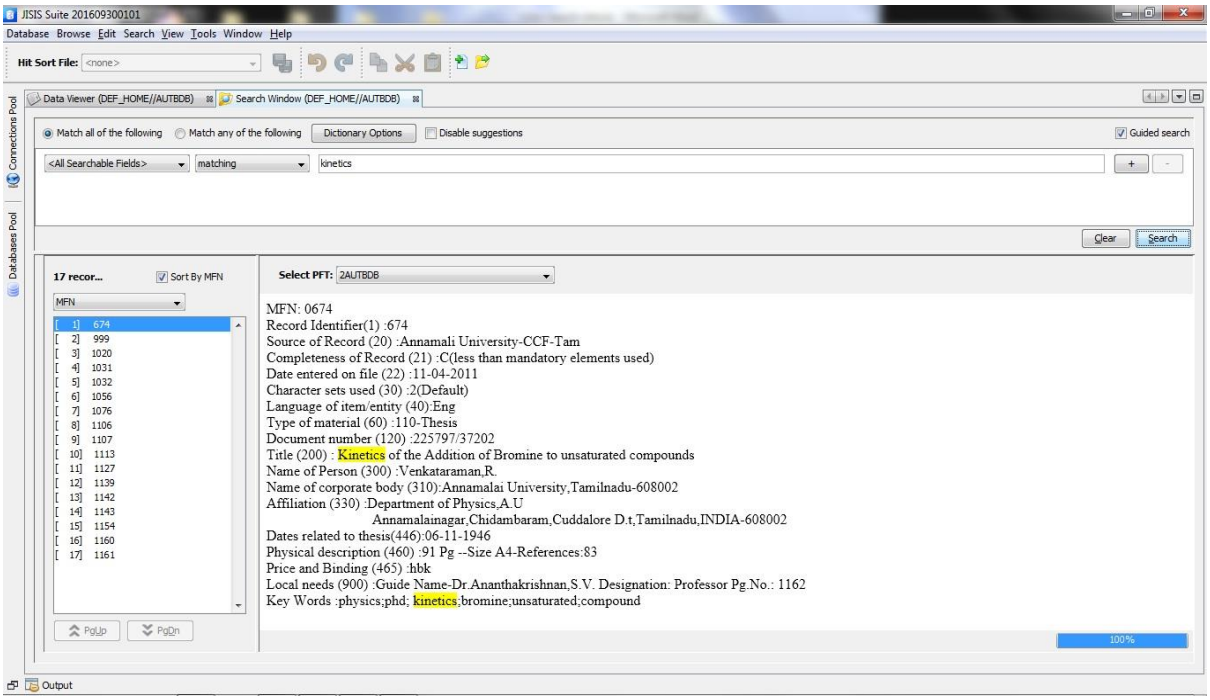

Figure-5: Query- *kinetics*

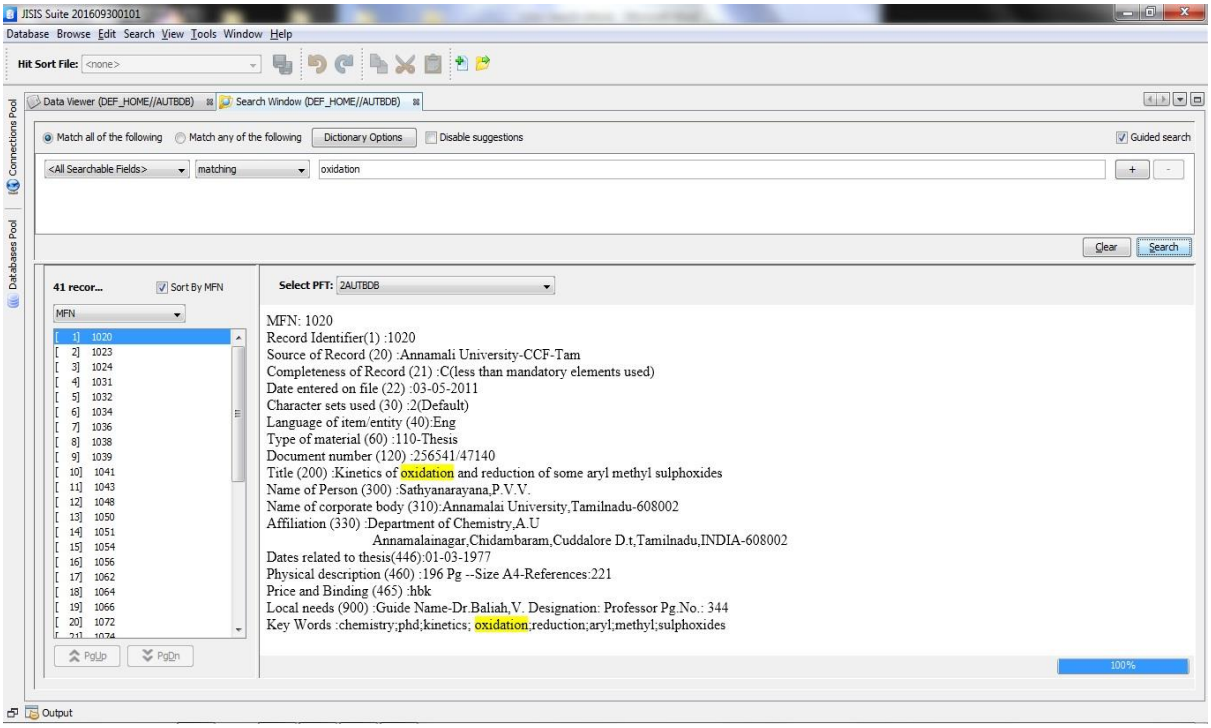

Figure-6: Query- *oxidation*

Table-2: MFNs traced for the Queries [*kinetics*, *oxidation*, *kinetics* AND *oxidation*]

S.No.	kinetics	oxidation	kinetics+oxidation
	MFN	MFN	MFN
1	674	**1020***	1020
2	999	1023	1031
3	**1020***	1024	1032
4	**1031***	**1031***	1056
5	**1032***	**1032***	1076
6	**1056***	1034	1113
7	**1076***	1036	1127
8	1106	1038	1139
9	1107	1039	1142
10	**1113***	1041	1143
11	**1127***	1043	1160
12	**1139***	1048	1161
13	**1142***	1050	
14	**1143***	1051	
15	1154	1054	
16	**1160***	**1056***	
17	**1161***	1062	
18		1064	
19		1066	
20		1072	
21		1074	
22		1075	
23		**1076***	
24		1078	
25		1079	
26		1080	
27		1085	
28		1110	
29		**1113***	
30		**1127***	
31		**1139***	
32		**1142***	
33		**1143***	
34		1145	
35		1149	
36		1157	
37		**1160***	
38		**1161***	
39		1164	
40		1166	
41		1559	

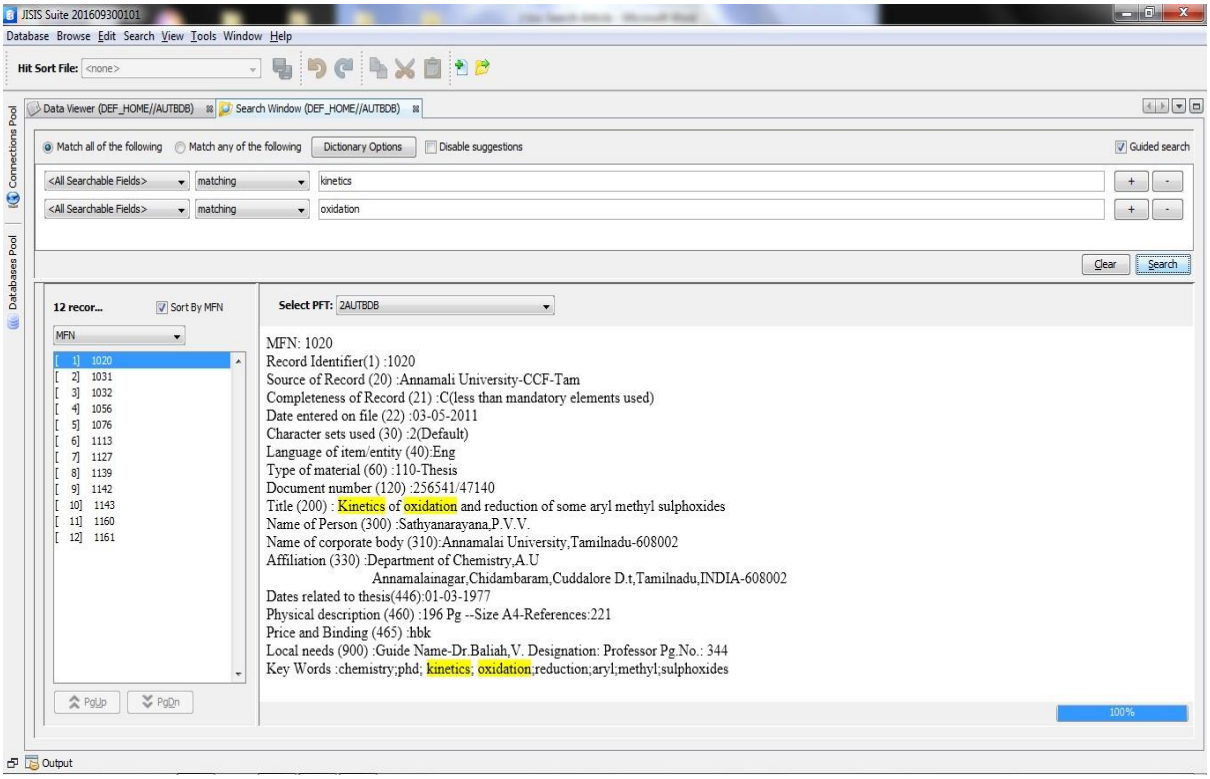

Figure-7: Query- *kinetics* AND *oxidation*

Of the retrieved 12 records, 6 records are traced as relevant, while the remaining 6 records are not relevant to the users' demands. This irrelevance is caused by the term *"correlation"* which is not supposed to be retrieved. The query is further refined in the next attempt where one more key term *"correlation"* is further added to confirm the non-relevant retrievals (i.e. 6 records with the additional term "*correlation"*).

R= [*a*/ (*a*+*c*)] x 100 ; R=[6/(6+28) x 100] = 17.64% ; Recall= 17.64% ,
P= [*a*/ (*a*+*b*] x 100 ; P=[6/(6+6) x 100] = 50% ; Precision= 50%

For the said three key terms, the observed ratio for the completeness of output is 17.64%, whereas the relevance of output is 50%.

Query-6: *kinetics* AND ***oxidation*** AND ***correlation***

The records that are relevant to the each key term are discussed here.

First key word : *Kinetics* [Matched with 17 terms available in 17 records *(See Figure-5)*]

Second key word: *Oxidation* [Matched with 41 terms available in 41 records *(See Figure-6)*]

Third key word: *Correlation* [Matched with 25 terms available in 25 records *(See Figure-3)*]

Of the 17 records, 12 records are found with the term *"oxidation"*. (The previous query explains the way of calculation)

Calculation:

Kinetics=17 records; Oxidation =41 Records; *Correlation*=25 records; *kinetics* and *oxidation* = 12 records (So, 41-12= 29 & 17-12=5); *combination of 3 key terms*=6 records (5 relevant records and 1 non-relevant).

All the 6 records' MFNs encompass the terms "*kinetics"* AND "*oxidation"*.

The 3 key words are found in 83 (41+17+25) records.

The term *"oxidation"* is available in 41 records and the term *"kinetics"* is found in 12 records that fall in the MFNs where the term *"oxidation"* exists. So, 12 records related to *kinetics* need to be ignored in this context and the remaining 5 records may be added to 41, giving rise to 46 records. Let us see the term *"correlation*" which was actually found in 25 records. However, 6 records are found in the MFNs where the terms *"oxidation"* and *"kinetics"* exist. Hence, those 6 records need to be ignored in this context and the remaining 19 records may be added to 46 records and give rise to 65 (i.e. 83-18=65 records) records that exist in the system.

The combinations of all 3 key terms are used to test the retrieval of relevant items. The details associated with the records counts are given in the Table-3.

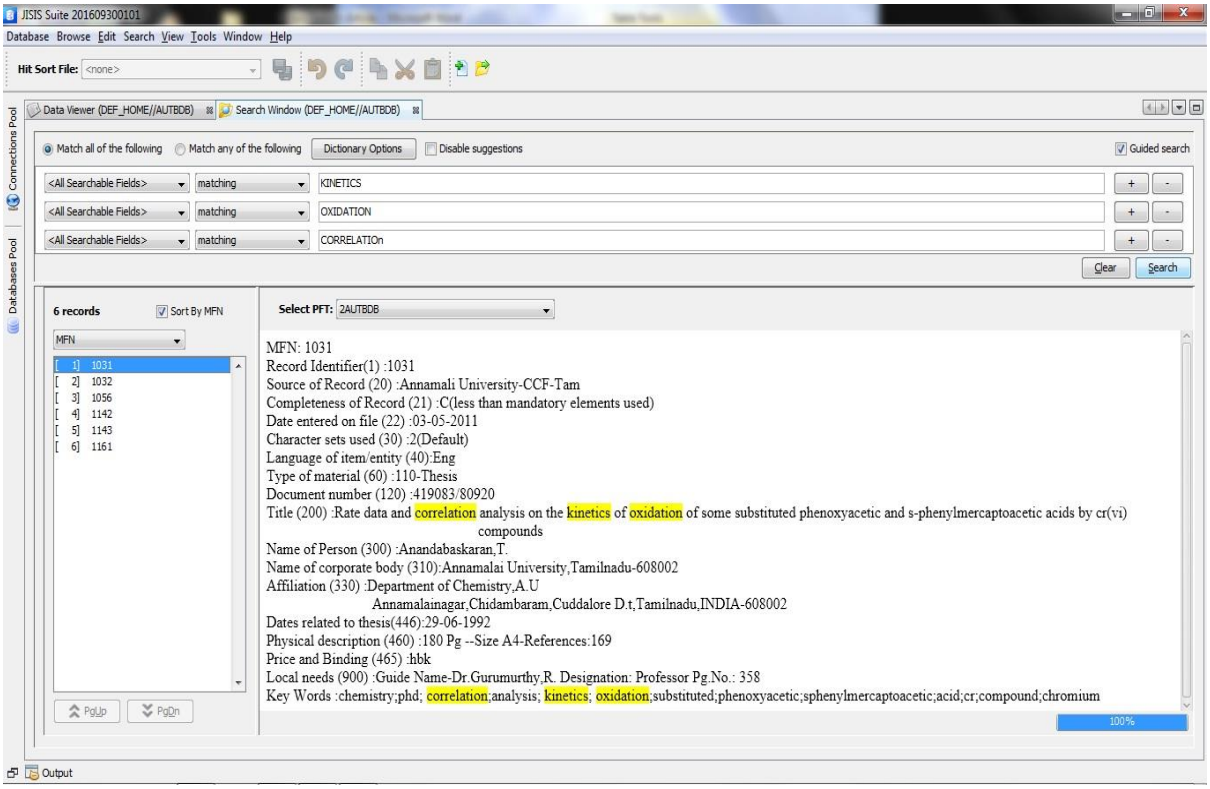

Figure-8: Query- *kinetics* AND *oxidation* AND *correlation*

Table-3: MFNs traced for the Queries [*kinetics* AND *oxidation* AND *correlation*]

S.No.	kinetics	oxidation	correlation	kinetics+oxidation+correlation
	MFN	**MFN**	**MFN**	**MFN**
1	674	1020	793	1031
2	999	1023	817	1032
3	1020	1024	845	1056
4	**1031***	**1031***	1027	1142
5	**1032***	**1032***	**1031***	1143
6	**1056***	1034	**1032***	1161
7	1076	1036	1038	
8	1106	1038	1043	
9	1107	1039	**1056***	
10	1113	1041	1062	
11	1127	1043	1066	
12	1139	1048	1072	
13	**1142***	1050	1074	
14	**1143***	1051	1078	
15	1154	1054	1079	
16	1160	**1056***	1080	
17	**1161***	1062	1084	
18		1064	1085	
19		1066	1106	
20		1072	1107	
21		1074	1108	
22		1075	1118	
23		1076	**1142***	
24		1078	**1143***	
25		1079	**1161***	
26		1080		
27		1085		
28		1110		
29		1113		
30		1127		
31		1139		
32		**1142***		
33		**1143***		
34		1145		
35		1149		
36		1157		
37		1160		
38		**1161***		
39		1164		
40		1166		
41		1559		

Although the key term *"correlation"* is found in 25 records, the users' search matched with 6 records (Figure-8) where 5 records are traced as relevant. Table-3 depicts the results for the 3 key terms where the matched records are highlighted by bold & *.

R= [*a*/ (*a*+*c*)] x 100 ; R=[5/(5+60) x 100] = 8.33% ; Recall= 8.33% ,
P= [*a*/ (*a*+*b*] x 100 ; P=[5/(5+1) x 100] = 83.33% ; Precision= 83.33%

It is observed that the completeness of output shows 8.33%, whereas the relevance of output indicates 83.33%.

Query-7: *Kinetics of oxidation [Phrase]*

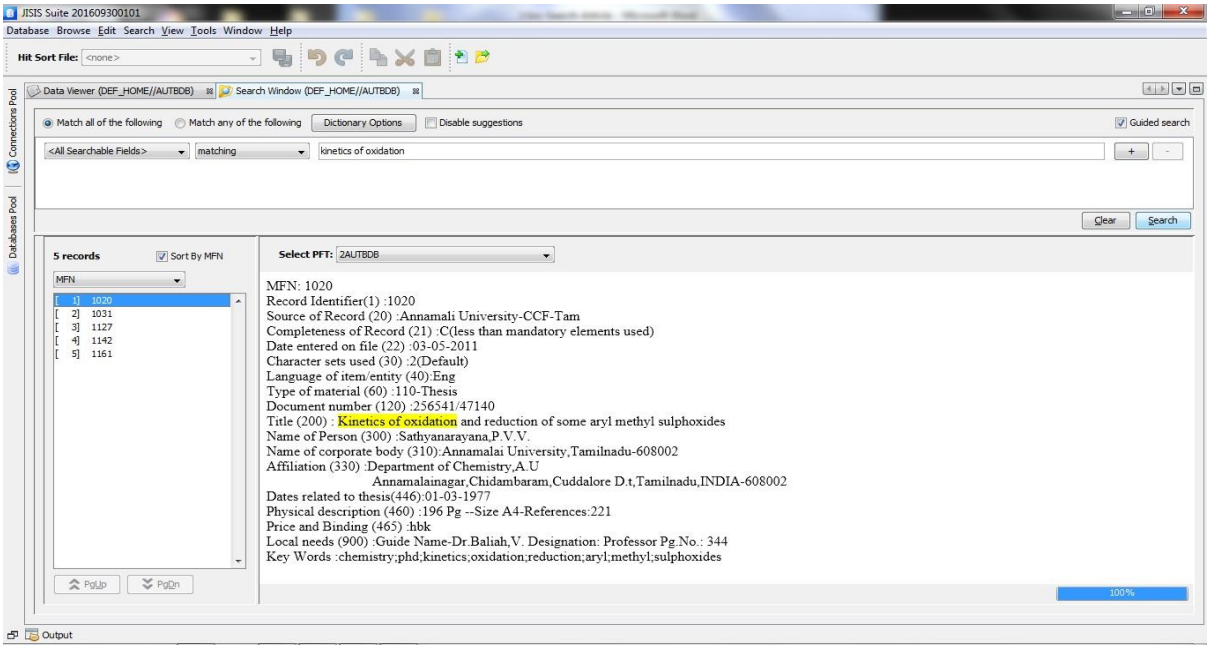

Figure-9: Query- *kinetics of oxidation (phrase search)*

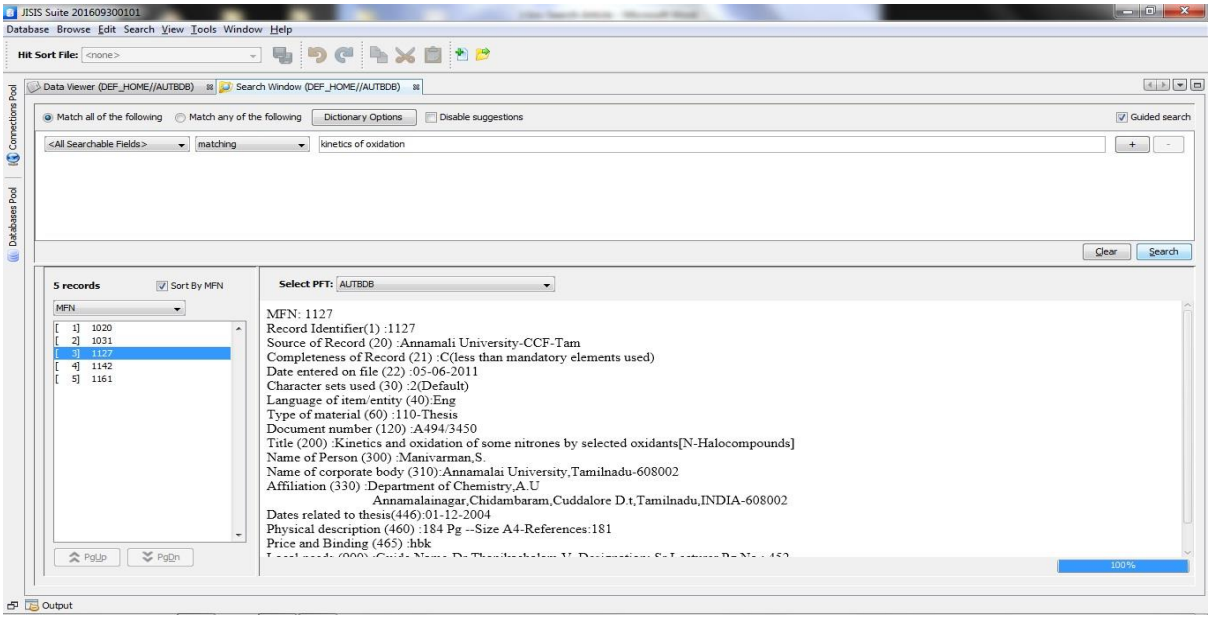

Figure-10: Query- *kinetics and oxidation (phrase search)*

Table-4: MFNs traced for the Query [*kinetics of oxidation*]

S.No.	kinetics	oxidation	Kinetics of oxidation
	MFN	**MFN**	**MFN**
1	674	**1020***	1020
2	999	1023	1031
3	**1020***	1024	*1127*
4	**1031***	**1031***	1142
5	1032	1032	1161
6	1056	1034	
7	1076	1036	
8	1106	1038	
9	1107	1039	
10	1113	1041	
11	**1127***	1043	
12	1139	1048	
13	**1142***	1050	
14	1143	1051	
15	1154	1054	
16	1160	1056	
17	**1161***	1062	
18		1064	
19		1066	
20		1072	
21		1074	
22		1075	
23		1076	
24		1078	
25		1079	
26		1080	
27		1085	
28		1110	
29		1113	
30		**1127***	
31		1139	
32		**1142***	
33		1143	
34		1145	
35		1149	
36		1157	
37		1160	
38		**1161***	
39		1164	
40		1166	
41		1559	

A search query is applied to retrieve the records that are relevant to the *term "kinetics of oxidation"*. Table-4 depicts the results of the key term where the matched records are highlighted by bold & *.

First key word : *Kinetics* [Matched with 17 terms available in 17 records]

Second key word: *Oxidation* [Matched with 41 terms available in 41 records]

Total records with both the terms = 48 **[**excluding the records repetitions)

Records retrieved for the combination of the key terms using a single query.

i.e. *"kinetics of oxidation"* is applied to retrieve the relevant records (Figure-9). Total records retrieved are 5 where relevant items are 4 and the rest of the single item is traced as not relevant. [MFN:1127 (Figure-10). This record includes the terms "*kinetics and oxidation" [Note: Here, the word "and" is not an operator]* and not *"kinetics of oxidation"*. The 4th column of the Table-4 shows the (MFN: 1127) non-relevant record which is indicated by *Italic* format]

R= [*a/ (a+c*)] x 100 ; R=[4/(4+44) x 100] = 8.33% ; Recall= 8.33% ,
P= [*a/ (a+b*] x 100 ; P=[4/(4+1) x 100] = 80% ; Precision= 80%

It is observed that the completeness of output shows 11.76% whereas the relevance of output indicates 80%.

Table-5: Summary

Order	Query	Recall	Precision	Terms
1.	*acid*	100%	78.57%	Single
2.	*acid* AND *chemistry*	5.45%	91.66%	Combined with Boolean
3.	*correlation, chemistry*	13.92%	88%	Single
4.	*correlation* AND *chemistry*	8.33%	68.18%	Combined with Boolean
5.	*kinetics* AND *oxidation*	17.64%	50%	Combined with Boolean
6.	*kinetics* AND *oxidation* AND *correlation*	8.33%	83.33%	Combined with Boolean
7.	*kinetics* of *oxidation*	8.33%	80%	phrase

Table-5 explores the queries, recall value, precision value, and various search terms which have been used in the study.

Determinations:

According to Lancaster, it is always possible to get 100 per cent recall if we retrieve enough of the total collection; if we retrieve the entire collection, we certainly achieve 100 per cent recall. Unfortunately, however, precision would be extremely low in this situation because, for any typical request, the great majority of the items in the collection are not relevant. This is confirmed with the results obtained towards the order number 1 for the recall ratio. The precision ratio is high for the order numbers 2, 3, 6 and 7. The success rate of data retrieval always depends on the search features of the retrieval system as well as the familiarity of users who access the data using the relevant search terms. Java CDS/ISIS adopts Berkeley DB to store all the records of a database and Lucene for creating Inverted files. Advanced indexing techniques are covered by Java CDS/ISIS for data processing and retrieval. It would be difficult for the novice users while retrieving the required data using these search techniques. They need to learn the techniques associated with the search and browse so that they can retrieve the required relevant items. However, the present study doesn't focus on the barriers of users while retrieving the data using the search key words, phrase and Boolean operator "AND". The effectiveness of the search techniques is good, which can be observed from the summary table generated for the recall and precision.

References:

Andrew, Buxton and Alan Hopkinson.(September 2001).*The CDS/ISIS for Windows Handbook.* Paris: UNESCO.

De Smet, E. (2009). *The abc of ABCD: the Reference Manual Version 1.0,* Accessed January 17, 2018. http://bvsmodelo.bvsalud.org/download/abcd/manuais/ ABCofABCD _atua lizado.pdf.

Jean-claude Dauphin. *Java CDS/ISIS.* (2015). Retrieved March 17, 2019 from https://github.com/J-ISIS/jisis

Jean-claude Dauphin. *J-ISIS 10 March 2016 Release.* (2016). Retrieved March 17, 2019 from https://github.com/J-ISIS/jisis

Lancaster, F W. (1979). *Information retrieval systems: characteristics, testing, and evaluation*. 2nd Ed. New York: John Wiley

Saravanan, T. (2007). Harnessing Database for Indian Journal of Information, Library & Society (IJILIS) using WINISIS: An Attempt. *IJILIS 20*(3-4),196-203.

Saravanan, T. (2008). A Visualization Interface Supporting Bibliographic Retrieval in Higher Educational Sector. *DLIBCOM 3* (1), 3-4.

Saravanan, T. (2008). A LIS Dissertation Abstracts Database: An Experience with WINISIS. *Information Studies, 14*(3),181-199.

Saravanan,T. (2009).Emerging Self learning Graphical Interface in LIS Discipline: A Glanc*e. Library Progress 29*(1),21-32.

Saravanan, *T.* (2010a) .Use of WinISIS for Compiling Abstract of dissertations submitted by all faculties of Annamalai University Tamilnadu: An Attem*pt. SRELS 47*(1), 71-74.

Saravanan, T et al. (2010b).Google Use and Users: A Survey. *Information Studies, 16*(1),49-64.

Saravanan, T. (2011). WinISIS: The information storage and retrieval application for faculty member's research performances. *Humanities 47*(20), 393-400.

Saravanan, T. (2012). Designing an Academic Database Using WinISIS. *Humanities XLVIII*(21),45-48.

Saravanan, T. (2012). *Information Architecture: A Multidisciplinary Text Book.* New Delhi: EssEss.

Saravanan T. (2013). *Library Automation (Automation Series)*. CDM: RTSV

Saravanan T. (2015). *Library Automation (Automation Series)*. New Delhi: APH.

Saravanan, T. (2018a). UNESCO's WinISIS: A Light behind the darkness. *IJNGLT 4*(1), 1-18.

Saravanan, T. (2018b, March). *Java CDS/ISIS*: Are the Librarians Ready To Rejuvenate the Libraries?. Paper presented at the meeting of *International Conference on Rejuvenating Libraries for Information Access in the Digital Era*, Annamalai University, Department of Library and Information Science: Annamalai University, Available at https://github.com/J-ISIS/J-ISIS/blob/master/J-ISIS%20Article%20 Saravanan % 2CT..pdf

Saravanan, T. (2018c). *Java CDS/ISIS*: One more precious stone in the crown of CDS/ISIS family *IJNGLT*, 4(3),1-24.

Saravanan, T. & Gana, G.S.,(2019). Bibliographic Database Shifting: An experience gained from Java CDS/ISIS. In *International Conference on Enhancement of Technology and Innovations in contemporary libraries* (pp. 493–499). Karaikudi, TN: Alagappa University.

Saravanan, T. (2019a). *Java CDS/ISIS*: A Red Carpet for Public Libraries, *IJILIS,* 32(1-2), 86-99.

Saravanan, T. (2019b). *Java CDS/ISIS*: *Java CDS/ISIS*: A Rigorous Tool for Effective and Efficient Data Sorting, *Library Philosophy and Practice (e-journal)*. https://digitalcommons.unl.edu/libphilprac/2647/.

Vickery, B.C. (1970). *Techniques of information retrieval*. London: Butterworth

https://en.wikipedia.org/wiki/Cyril_Cleverdon

https://www.webopedia.com/TERM/K/keyword_search.html

https://www.webopedia.com/TERM/P/phrase_search.html

https://www.webopedia.com/TERM/B/Boolean_search.html

***