Annals of Library and Information Studies
Vol. 70, September 2023, pp. 132-145
DOI: 10.56042/alis.v70i3.2492

Retrieval effectiveness of Scopus and Web of Science based on three types of subject metadata: a comparative study in Economics

Kalyan Sundar Samanta[a] and Durga Sankar Rath[b]

[a]Librarian, Prabhu Jagatbandhu College, Howrah – 711302, West Bengal, India

[b]Professor, Department of Library & Information Science, Vidyasagar University, Midnapore -721102, West Bengal, India

Email: dsrath@mail.vidyasagar.ac.in, kalyansundarsamanta@gmail.com

Received: 8 June 2023; accepted: 17 July 2023

The study tried to measure the retrieval effectiveness of Scopus and Web of Science databases and, simultaneously, to measure the resource description capabilities of social tags in comparison with standard subject metadata like LCSH descriptors and SLSH terms. The study results show that both databases produce more or less the same type of Average Mean Precision (AMP), but comparatively, WOS (AMP=2.44) has a higher AMP than Scopus (AMP=2.30). In the case of Average Mean Relative Recall (AMRR), the study indicates that Scopus (AMRR=0.73) has a higher AMRR than WOS (AMRR=0.44). The study also reveals that social tags (MP=2.24) have lower Mean Precision (MP) than other metadata in Scopus, but social tags (2.46) have higher MP in the WOS database. In the case of Mean Relative Recall (MRR), the study found that social tags (MRR=0.62 in Scopus & MR=0.38 in WOS) have the same MRR as LCSH descriptors in Scopus and WOS databases, where SLSH terms have higher MRR than other subject metadata. Moreover, the study concludes that social tags can describe and retrieve Economics documents like standard metadata (LCSH descriptors and SLSH terms).

Keywords: Retrieval evaluation of Scopus, Retrieval evaluation of Web of Science, Recall and Precision, information retrieval

1. Introduction

Scientific publications have made a dramatic change in our society. As a result, most critical decisions and policies of a state or a nation are based on research and scientific output. Earlier, retrieving those publications was a challenge among researchers. Subject metadata has remained researchers' prime choice, leading them to access those publications easily. Even to augment subject access, the library community has been applying adequate subject metadata over the years[1-8]. With the change of time and technology over years, the information needs of users has been changed. Earlier, library users used library card catalogues for book searches by preferring subject entries and accessing print-based collections like print journals, thesauri, glossaries, dictionaries, and encyclopedias to meet their information needs. The development of the internet and communication technologies has radically changed information gathering, storing, organizing, accessing and retrieving. It transformed those print-based sources into online information sources like transforming the card catalogue into a web-based Online Public Access Catalogue (Web OPAC) and transforming other sources into online information sources like online encyclopedias, online thesaurus, subject directories, subject gateways, web search engines and online databases to access the subject-based information. In that context, creating online databases like Scopus and Web of Science is a significant achievement for providing many users with subject-based reliable and scholarly information[2, 9].

Online databases are sources of authentic and updated scholarly information in various fields and disciplines[10]. It is mainly a collection of electronic information sources from multiple areas by the publishers[11]. The information is organized in such a way that it can be accessed, managed and updated over the network. Online databases provide a simple and advanced search interface with few search syntaxes to retrieve exact information among millions of data. The primary purpose of those online databases is to share scientific ideas, study results, subject-based discussions and information among the academic community globally[2-3, 12]. The success of those online databases depends on the use of adequate subject metadata during document descriptions. The use of proper metadata can lead to the effective retrieval of information. Subject metadata is generated through the procedure of subject cataloguing. In traditional libraries, librarians use subject headings

like Library of Congress Subject Headings (LCSH) and Sears List of Subject Headings (SLSH) etc., for document description. In contrast to that, a new concept of subject cataloguing emerged as 'social, which allows users to assign subject metadata for web resources[1].

With that context, it is essential to measure how online databases retrieve information based on controlled (LCSHs and SLSHs) and uncontrolled metadata (social tags) and simultaneously to measure the quality of that uncontrolled metadata for document description in Economics.

2. Review of Related Literature

Many researchers conducted studies to measure the retrieval effectiveness of online databases. Ritchie, Banyas, and Sevin conducted a similar kind of study where they compared the retrieval performance of eight literature databases like AGRICOLA, AGRIS, BIOSIS, CAB Direct, FSTA, Google Scholar, Scopus, and Web of Science, based on three topics from different agricultural disciplines. They preferred three criteria, Precision, Recall and Uniqueness, for the evaluation. The study found that AGRICOLA has the highest Precision and it produced more relevant results than other databases, whereas CAB & Web of Science has the highest Recall. Also, it was found that four databases, Google Scholar, AGRICOLA, BIOSIS and CAB, similarly produced the most unique and relevant content. However, in the case of Google Scholar, the relevancy of half of the content was judged[13].

Stokes et al. also conducted a similar kind of analysis, where they compared four bibliographic databases like BNI, CINAHL, EMBASE and MEDLINE of health sciences. In total, nine nursing students participated in the evaluation process, where they were asked to search topics on those databases and to judge the relevancy of retrieved results. The study used three criteria for the evaluation, Precision and Recall for effectiveness and Uniqueness for database efficiency. The study found that the four databases are different from each other based on the Precision, originality and availability of retrieved resources. Besides, odds ratio tests also reveal that BNI is the most effective and accessible database, whereas CINAHL is the most effective and efficient database[14].

Walters also compared Google Scholar with eleven other databases to measure the retrieval performance for retrieving literature on later-life migration. The study preferred simple keyword searches, and two popular criteria, Recall and Precision, were used for the evaluation. The entire retrieval performance was evaluated in comparison to a set of 155 relevant articles that were identified before searching. The study reveals that Google Scholar performs far better than other databases in producing literature on the topic[15].

Sewell conducted a study on CAB Abstract, the premier database for veterinary medical literature. The study compared the retrieval performance of four vendor databases, CAB Direct, EBSCO host, OvidSP and ISI, used to access CAB Abstract. The study used recall and precision criteria to measure retrieval performance. The study found no statistical differences in Recall and Precision in the search performance among four databases[16].

3. Objectives of the Study

The prime focus of the study is to evaluate the retrieval effectiveness of Scopus and Web of Science databases based on three different sets of subject metadata as search queries in the field of Economics. To fulfil the main purpose, the present study wants to identify (a) which database has the best retrieval performance and (b) which set of subject metadata has the best quality for document description as well as document retrieval.

4. Methodology

The study intends to evaluate the retrieval performance of Scopus and WOS databases based on three sets of subject metadata in Economics. We have taken two broad categories of subject metadata in the study, (i) controlled metadata like LCSH descriptors and SLSH terms and (ii) uncontrolled metadata like social tags. For collecting social tags, we preferred popular social cataloguing tools like LibraryThing (https://www.librarything.com), and for collecting LCSH descriptors, we chose the Library of Congress Online Catalogue (https://catalog.loc.gov/). To carry out the evaluation, initially, we have selected a thousand books on Economics that have been catalogued by both the LibraryThing and the Library of Congress database. Based on those thousand books, we collected social tags assigned by users and LCSH descriptors assigned by the experts. Further, as there is no ready-made database for collecting SLSH terms, we have generated SLSH terms using the Sears List of

Subject Headings (19th ed.). Though we collected a good number of unique social tags (2983), unique LCSH descriptors (804) and unique SLSH terms (246) for the initial evaluation, we have selected only the top thirty (30) subject-based unique tags and terms (mentioned in Table 1) as queries from each set for conducting the retrieval evaluation process.

Further, we have selected two worldwide popular online databases like Scopus (https://www.scopus.com) and Web of Science (https://clarivate.com/webofsciencegroup/solutions/web-of-science), where the entire search process was conducted.

We followed a specific search string for Scopus *{TITLE("public policy") AND (LIMIT-TO (LANGUAGE "English")}* and for WOS database *{(TI=("public policy")) AND LANGUAGE: (English)}* (mentioned in Fig. 1 & 2). We evaluated the retrieved results for measuring the effectiveness of online databases. The effectiveness depends on how many resources have been indexed by any query under a given online database. We have preferred the top fifty (50) retrieved results for evaluation in our study[17].

Besides, we have categorized the retrieved results based on their relevance to the queries. Relevance indicates how much a search query is related to the search result[18]. For determining the relevance, we have preferred the human relevance judgement process[19-20]. Human relevance judgement is a process where selected individuals judge the relevance of the retrieved results based on a query. We have preferred 'Four Point Scale' to measure the relevance of retrieved results effectively[21]. In Four Point Scale, the relevance is categorized into four scales, (a) *Highly Relevant (HR)* – document discusses the themes of the topic exhaustively (level 3), (b) *Relevant (R)* – the document contains more information than the topic description, but the presentation is not exhaustive (level 2), (c) *Marginally Relevant (MR)* – document only points to the topic. It does not contain more or

Table 1 — Top thirty social tags, LCSH descriptors & SLSH terms used as queries for evaluation

Query	Social tags	LCSH descriptors	SLSH terms
Q1	economics	economics	economics
Q2	business	economc conditions	economic conditions
Q3	finance	capitalism	capitalism
Q4	capitalism	economic policy	finance
Q5	ecomomy	economic history	economic policy
Q6	economic history	financial crises	economic development
Q7	political economy	economic aspects	investments
Q8	globalization	economic development	socialism
Q9	money	investments	financial crises
Q10	marxism	globalizations	marxism
Q11	socialism	finance	free enerprise
Q12	labor	international economic relations	economic forecasting
Q13	investment	money	money
Q14	poverty	free enterprise	international economic relations
Q15	economic theory	labor unions	macroeconomics
Q16	markets	monetary policy	labor economics
Q17	economic policy	international finance	labor unions
Q18	financial	socialism	economic history
Q19	economic development	economic forecasting	monetary policy
Q20	international relations	income distribution	globalization
Q21	public policy	poverty	employment
Q22	wealth	working class	labor
Q23	business & economics	wealth	wealth
Q24	financial crisis	macroeconomics	personal finance
Q25	business history	sustainable development	microeconomics
Q26	labor history	business cycles	sustainable development
Q27	history of economics	microeconomics	economic growth
Q28	macroeconomics	government policy	business cycles
Q29	austrian economics	capital	business enterprises
Q30	stock market	employment	industrial revolution

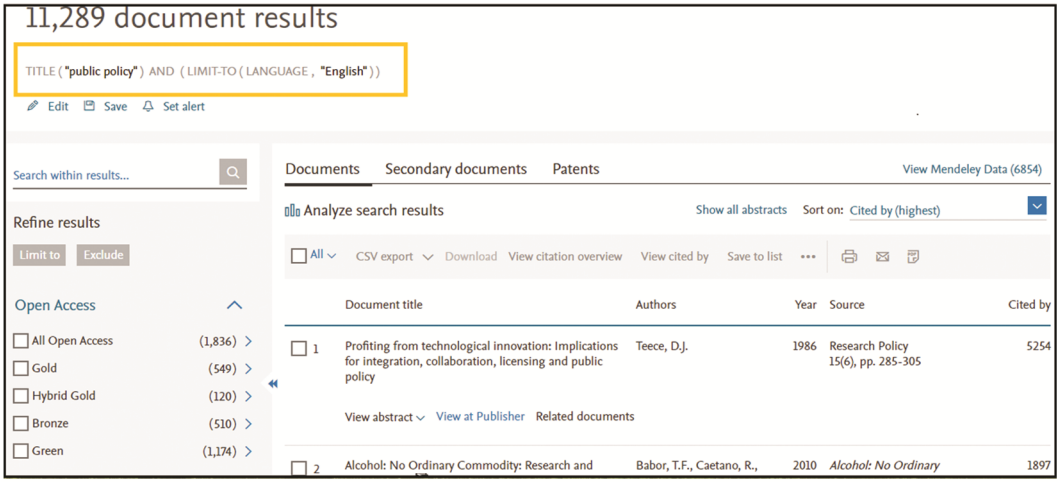

Fig. 1 — Sample Scopus search (TI=("public policy")) *AND* LANGUAGE: (English)

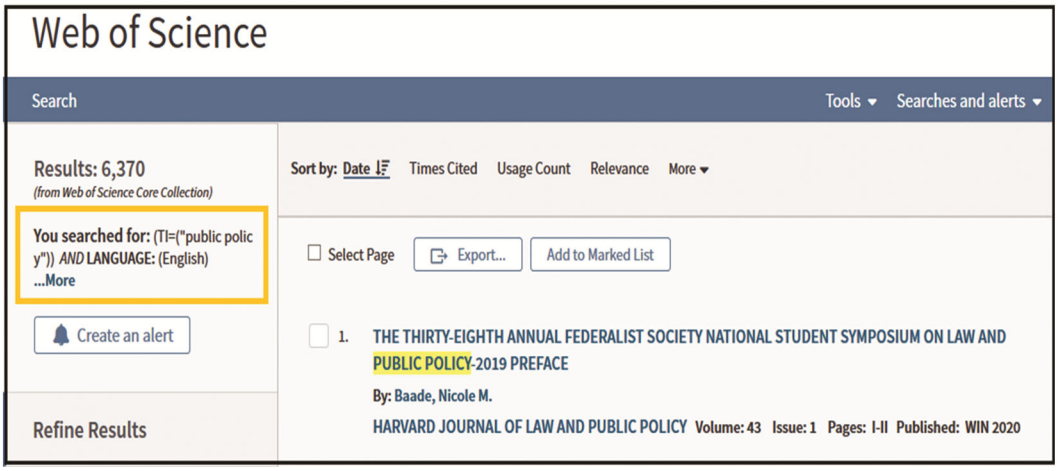

Fig. 2 — Sample Web of Science search (TI=("public policy")) *AND* LANGUAGE: (English)

other information than the topic description (level 1), (d) *Irrelevant (I)* – the document contains no information about the topic (level 0).

After categorizing the retrieved results, we have preferred two popular information retrieval criteria, Precision and Recall, for measuring the retrieval effectiveness of online databases. Precision is the fraction of a search output relevant to a particular query, and its calculation hence requires knowledge only of the relevant and non-relevant in the retrieved set of documents.

$$Precision = \frac{\text{Sum of the scores of search results by an online database}}{\text{Total number of search results evaluated}}$$

Further, Recall is the faction of the relevant documents for a particular query that is retrieved. But it is difficult to calculate the exact Recall for online databases[22], because it is required to know the relevant items in the collection to calculate the Recall. Relevant items can be measured when the collection size is small, but it is difficult when the collection size is large, like for online databases. To avoid this problem, Information Scientists prefer Relative Recall to compare the relative performance of online databases[23].

$$Relative\ Recal = \frac{\text{Total number of search results retrieved by an online database}}{\text{Sum of sites retrieved by the two online databases}}$$

Based on the retrieval parameters, we evaluated the entire retrieval performance of databases. We also prepared the Mean Precision (MP) and the Mean Relative Recall (MRR) for both databases separately to identify the effective metadata and database[24-25]. Even, we calculated the Pearson Correlation Coefficient (r) to measure the consistency of the Precision under both databases[26-27].

5. Data Analysis

5.1 Precision Analysis

5.11 Precision Analysis of Scopus and WOS Database

Precision is an essential tool to evaluate the effectiveness of any information retrieval system. High Precision indicates that the system can effectively retrieve relevant documents to the query. In contrast, low Precision suggests the system is less effective and produces irrelevant documents to the query[24, 28-30]. The present study measures the Precision of Scopus and WOS databases under three different types of metadata like social tags, LCSH descriptors and SLSH terms used for document description in Economics. Table 2 indicates the Precision of the Scopus and WOS databases based on social tags as search queries. Table 2 reveals that the search query 'economic policy' has the highest Precision (2.92); in contrast, the search query 'labor history' has the lowest Precision (0.58) and overall, Total Precision (TP = 69.66) and Mean Precision (MP = 2.32) respectively in Scopus database. Another way, Table 2 also reveals that the search query 'stock market' has the highest Precision (2.9) and the search query 'economics' has the lowest Precision (1.6). Overall, it reveals that TP = 73.94 and MP = 2.46, respectively, out of the top thirty search queries in the WOS database.

[NRR – No. of Resources Retrieved, NRE – No. of Resources Evaluated, HR – Highly Relevant, R – Relevant, MR – Marginally Relevant, I – Irrelevant, P – Precision, TP – Total Precision and MP – Mean Precision]

Table 2 — Precision Analysis of Scopus & WOS database based on Social tags

Query	Scopus					P@50	WOS					P@50
	NRR	HR	R	MR	I		NRR	HR	R	MR	I	
Q1	42481	29	13	0	8	2.26	26756	22	6	2	20	1.6
Q2	119846	37	8	0	5	2.54	48268	34	6	0	10	2.28
Q3	16850	37	10	0	3	2.62	8352	33	5	0	12	2.18
Q4	8060	43	5	0	2	2.78	8819	41	3	0	6	2.58
Q5	79075	30	14	0	6	2.36	41818	27	12	0	11	2.1
Q6	1245	45	3	0	2	2.82	1536	42	6	0	2	2.76
Q7	10005	39	7	0	4	2.62	9280	41	8	0	1	2.78
Q8	19446	35	9	0	6	2.46	10753	36	7	0	7	2.44
Q9	23027	36	11	0	3	2.6	16067	35	10	0	5	2.5
Q10	1970	44	6	0	0	2.88	1623	40	8	0	2	2.72
Q11	3720	40	8	0	2	2.72	4147	36	8	0	6	2.48
Q12	78789	30	10	0	10	2.2	37093	27	6	0	17	1.86
Q13	54995	35	5	0	10	2.3	23801	36	8	0	6	2.48
Q14	23520	35	13	0	2	2.62	16983	37	6	0	7	2.46
Q15	1478	31	12	0	7	2.34	850	33	10	0	7	2.38
Q16	183377	25	11	0	14	1.94	29980	22	8	0	20	1.64
Q17	2709	46	4	0	0	2.92	1657	44	3	0	3	2.76
Q18	73582	29	14	0	7	2.3	37110	35	7	0	8	2.38
Q19	11420	37	11	0	2	2.66	5888	39	6	0	5	2.58
Q20	3575	45	5	0	0	2.9	3219	44	5	0	1	2.84
Q21	10549	0	21	11	18	1.06	6272	30	8	5	7	2.22
Q22	8918	7	36	4	3	1.94	7141	32	6	5	7	2.26
Q23	129	26	12	5	7	2.14	50	30	16	3	1	2.5
Q24	7567	12	32	6	0	2.12	3281	44	5	1	0	2.86
Q25	402	6	35	8	1	1.92	336	39	9	0	2	2.7
Q26	473	0	7	15	28	0.58	351	37	8	5	0	2.64
Q27	85	4	38	8	0	1.92	94	37	12	1	0	2.72
Q28	1229	37	8	0	5	2.54	986	37	9	3	1	2.64
Q29	156	10	34	6	0	2.08	63	36	13	1	0	2.7
Q30	11091	2	39	9	0	1.86	3959	45	5	0	0	2.9
	TP				69						73.94	
	MP (where n= 30)				2.3						2.46	

[NRR – No. of Resources Retrieved, NRE – No. of Resources Evaluated, HR – Highly Relevant, R – Relevant, MR – Marginally Relevant, I – Irrelevant, P – Precision, TP – Total Precision and MP – Mean Precision]

Table 3 indicates the Precision of Scopus and WOS databases based on LCSH descriptors as search queries. Table 3 reveals that the search query 'international finance' has the highest Precision (2.96), and 'working class' & 'government policy' both have the lowest Precision (0.98). Overall, the study reveals that TP = 69.66 and MP = 2.32, out of the top thirty search queries in the Scopus database. Again, Table 3 also reveals that the search query 'international finance' has the highest Precision (2.92) and 'working class' has the lowest Precision (1.34). Overall, it reveals that TP = 72.86 and MP = 2.43, respectively, out of the top thirty search queries in the WOS database.

[NRR – No. of Resources Retrieved, NRE – No. of Resources Evaluated, HR – Highly Relevant, R – Relevant, MR – Marginally Relevant, I – Irrelevant, P – Precision, TP – Total Precision and MP – Mean Precision]

Table 4 indicates the Precision of Scopus and WOS databases based on SLSH terms as search queries. Table 4 highlights that the search query 'economic policy', 'financial crises', 'international economic relations' and 'monetary policy' have the highest Precision (2.92), whereas 'sustainable development' has the lowest Precision. Overall, the study reveals that TP = 70.22 and MP = 2.34 out of the top thirty search queries in the Scopus database. Again, Table 4 reveals that the search query 'monetary policy' and 'financial crises' have the highest Precision (2.88) and 'economics' has the lowest Precision (1.6). Further, the study reveals that TP = 72.7 and MP = 2.42,

Table 3 — Precision Analysis of Scopus & WOS database based on LCSH descriptors

Query	Scopus						WOS					
	NRR	HR	R	MR	I	P@50	NRR	HR	R	MR	I	P@50
Q1	42481	29	13	0	8	2.26	26756	22	6	2	20	1.6
Q2	956	36	5	0	9	2.36	477	34	6	0	10	2.28
Q3	8060	43	5	0	2	2.78	8819	41	3	0	6	2.58
Q4	2709	46	4	0	0	2.92	1657	44	3	0	3	2.76
Q5	1245	45	3	0	2	2.82	1536	42	6	0	2	2.76
Q6	7437	46	4	0	0	2.92	604	44	6	0	0	2.88
Q7	2297	16	16	0	18	1.6	772	22	18	0	10	2.04
Q8	11420	37	11	0	2	2.66	5888	39	6	0	5	2.58
Q9	54995	35	5	0	10	2.3	23801	36	8	0	6	2.48
Q10	19446	35	9	0	6	2.46	10753	36	7	0	7	2.44
Q11	16850	37	10	0	3	2.62	8352	33	5	0	12	2.18
Q12	55	48	1	0	1	2.92	51	37	6	0	7	2.46
Q13	23027	36	11	0	3	2.6	16067	35	10	0	5	2.5
Q14	184	39	9	0	2	2.7	154	42	3	0	5	2.64
Q15	499	38	10	0	2	2.68	250	41	6	0	3	2.7
Q16	6612	46	4	0	0	2.92	4147	44	6	0	0	2.88
Q17	236	48	2	0	0	2.96	182	46	4	0	0	2.92
Q18	3720	40	8	0	2	2.72	4147	36	8	0	6	2.48
Q19	149	35	10	0	5	2.5	77	38	6	0	6	2.52
Q20	2037	40	9	0	1	2.76	926	41	8	0	1	2.78
Q21	23520	35	13	0	2	2.62	16983	37	6	0	7	2.46
Q22	2961	0	17	15	18	0.98	3199	8	16	11	15	1.34
Q23	8918	7	36	4	3	1.94	7141	32	6	5	7	2.26
Q24	1229	37	8	0	5	2.54	986	37	9	3	1	2.64
Q25	19288	0	22	11	17	1.1	8600	28	10	7	5	2.22
Q26	3875	3	30	16	1	1.7	1093	45	3	0	2	2.82
Q27	355	4	37	8	1	1.88	266	32	12	3	3	2.46
Q28	1935	0	21	7	22	0.98	586	24	10	10	6	2.04
Q29	58747	1	34	14	1	1.7	32860	30	9	5	6	2.26
Q30	36458	3	34	11	2	1.76	21407	24	10	3	13	1.9
		TP				69.66						72.86
		MP (where n= 30)				2.32						2.43

[NRR – No. of Resources Retrieved, NRE – No. of Resources Evaluated, HR – Highly Relevant, R – Relevant, MR – Marginally Relevant, I – Irrelevant, P – Precision, TP – Total Precision and MP – Mean Precision]

Table 4 — Precision Analysis of Scopus & WOS database based on SLSH terms

Query	Scopus						WOS					
	NRR	HR	R	MR	I	P@50	NRR	HR	R	MR	I	P@50
Q1	42481	29	13	0	8	2.26	26756	22	6	2	20	1.6
Q2	956	36	5	0	9	2.36	477	34	6	0	10	2.28
Q3	8060	43	5	0	2	2.78	8819	41	3	0	6	2.58
Q4	16850	37	10	0	3	2.62	8352	33	5	0	12	2.18
Q5	2709	46	4	0	0	2.92	1657	44	3	0	3	2.76
Q6	11420	37	11	0	2	2.66	5888	39	6	0	5	2.58
Q7	54995	35	5	0	10	2.3	23801	36	8	0	6	2.48
Q8	3720	40	8	0	2	2.72	4147	36	8	0	6	2.48
Q9	7437	46	4	0	0	2.92	604	44	6	0	0	2.88
Q10	1970	44	6	0	0	2.88	1623	40	8	0	2	2.72
Q11	184	39	9	0	2	2.7	154	42	3	0	5	2.64
Q12	149	35	10	0	5	2.5	77	38	6	0	6	2.52
Q13	23027	36	11	0	3	2.6	16067	35	10	0	5	2.5
Q14	55	48	1	0	1	2.92	51	37	6	0	7	2.46
Q15	1229	37	8	0	5	2.54	979	34	8	0	8	2.36
Q16	140	39	9	0	2	2.7	107	37	7	0	6	2.5
Q17	499	38	10	0	2	2.68	250	41	6	0	3	2.7
Q18	1245	45	3	0	2	2.82	1536	42	6	0	2	2.76
Q19	6612	46	4	0	0	2.92	4147	44	6	0	0	2.88
Q20	19446	35	9	0	6	2.46	10753	36	7	0	7	2.44
Q21	36458	3	34	11	2	1.76	21407	24	10	3	13	1.9
Q22	78789	30	10	0	10	2.2	37093	27	6	0	17	1.86
Q23	8918	7	36	4	3	1.94	7141	32	6	5	7	2.26
Q24	186	2	36	8	4	1.72	89	30	11	6	3	2.36
Q25	355	4	37	8	1	1.88	266	32	12	3	3	2.46
Q26	19288	0	22	11	17	1.1	8600	28	10	7	5	2.22
Q27	15402	0	37	5	8	1.58	7792	32	4	4	10	2.16
Q28	3875	3	30	16	1	1.7	1093	45	3	0	2	2.82
Q29	381	0	32	7	11	1.42	115	31	12	5	2	2.44
Q30	1536	4	33	5	8	1.66	1180	27	5	5	13	1.92
TP						70.22						72.7
MP (where n= 30)						2.34						2.42

[NRR – No. of Resources Retrieved, NRE – No. of Resources Evaluated, HR – Highly Relevant, R – Relevant, MR – Marginally Relevant, I – Irrelevant, P – Precision, TP – Total Precision and MP – Mean Precision]

respectively, out of the top thirty search queries in the WOS database.

5.12 Correlation Coefficient Analysis based on the Precision of Scopus and WOS database

To measure the consistency among the Precision of Scopus and WOS databases, we used Pearson Correlation Coefficient. Correlation measures the relationship between two variables. It closely measures how two variables are related to each other. Even we also wanted to identify whether there is a positive or negative correlation between them. A positive correlation indicates they are closely related, and negative means they are different.

Pearson Correlation Coefficient (r) =

$$r = \frac{\sum(\mathrm{x}-\bar{\mathrm{x}})\,(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2.\ 2(y-\bar{y})^2]}} \quad \ldots (1)$$

Here, we determine the Precision of social tags as X, the Precision of LCSH descriptors as Y and the Precision of SLSH terms as Z in the study[26-27].

Table 5 compares the correlation among the Precision of the top thirty frequent social tags (X), LCSH descriptors (Y) and SLSH terms (Z) under the Scopus database. Table 5 also reveals that the correlation between X & Y = 0.36; between Y & Z = 0.66, and between X & Z = 0.72. That means the study indicates a positive and strong correlation among the Precision of three kinds of metadata.

Table 6 compares the correlation among the Precision of the top thirty frequent social tags (X),

Table 5 — Correlation of Scopus search

Query	X	Y	Z	Correlation Coefficient		
				X & Y	Y & Z	X & Z
Q1	2.26	2.26	2.26	0.36	0.66	0.72
Q2	2.54	2.36	2.36			
Q3	2.62	2.78	2.78			
Q4	2.78	2.92	2.62			
Q5	2.36	2.82	2.92			
Q6	2.82	2.92	2.66			
Q7	2.62	1.6	2.3			
Q8	2.46	2.66	2.72			
Q9	2.6	2.3	2.92			
Q10	2.88	2.46	2.88			
Q11	2.72	2.62	2.7			
Q12	2.2	2.92	2.5			
Q13	2.3	2.6	2.6			
Q14	2.62	2.7	2.92			
Q15	2.34	2.68	2.54			
Q16	1.94	2.92	2.7			
Q17	2.92	2.96	2.68			
Q18	2.3	2.72	2.82			
Q19	2.66	2.5	2.92			
Q20	2.9	2.76	2.46			
Q21	1.06	2.62	1.76			
Q22	1.94	0.98	2.2			
Q23	2.14	1.94	1.94			
Q24	2.12	2.54	1.72			
Q25	1.92	1.1	1.88			
Q26	0.58	1.7	1.1			
Q27	1.92	1.88	1.58			
Q28	2.54	0.98	1.7			
Q29	2.08	1.7	1.42			
Q30	1.86	1.76	1.66			

Table 6 — Correlation of WOS search

Query	X	Y	Z	Correlation Coefficient		
				X & Y	Y & Z	X & Z
Q1	1.6	1.6	1.6	0.13	0.49	0.22
Q2	2.28	2.28	2.28			
Q3	2.18	2.58	2.58			
Q4	2.58	2.76	2.18			
Q5	2.1	2.76	2.76			
Q6	2.76	2.88	2.58			
Q7	2.78	2.04	2.48			
Q8	2.44	2.58	2.48			
Q9	2.5	2.48	2.88			
Q10	2.72	2.44	2.72			
Q11	2.48	2.18	2.64			
Q12	1.86	2.46	2.52			
Q13	2.48	2.5	2.5			
Q14	2.46	2.64	2.46			
Q15	2.38	2.7	2.36			
Q16	1.64	2.88	2.5			

(contd.)

Table 6 — Correlation of WOS search (*contd.*)

Query	X	Y	Z	Correlation Coefficient		
				X & Y	Y & Z	X & Z
Q17	2.76	2.92	2.7			
Q18	2.38	2.48	2.76			
Q19	2.58	2.52	2.88			
Q20	2.84	2.78	2.44			
Q21	2.22	2.46	1.9			
Q22	2.26	1.34	1.86			
Q23	2.5	2.26	2.26			
Q24	2.86	2.64	2.36			
Q25	2.7	2.22	2.46			
Q26	2.64	2.82	2.22			
Q27	2.72	2.46	2.16			
Q28	2.64	2.04	2.82			
Q29	2.7	2.26	2.44			
Q30	2.9	1.9	1.92			

LCSH descriptors (Y) and SLSH terms (Z) under the WOS database. Table 6 reveals the correlation between X & Y = 0.13; between Y & Z = 0.49, and between X & Z = 0.22. That means the study indicates a positive correlation among all but found a strong correlation only between Y & Z.

5.2 Relative Recall (RR) Analysis

Recall denotes how well a system performs in retrieving the relevant information based on the search query. Relative Recall (RR) indicates the fraction of relevant documents based on the total number of relevant documents in the collection[24, 28-30]. Table 7 shows the Relative Recall (RR) based on the top thirty queries under Scopus and WOS database, where Total Relative Recall (TRR) and Mean Relative Recall (MRR) are higher in Scopus (TRR = 18.47 & MRR = 0.62) than WOS database (TRR = 11.53 & MRR = 0.38).

[NSR – No. of sites retrieved, RR – Relative Recall, TRR – Total Relative Recall and MRR – Mean Relative Recall]

Table 8 indicates the Relative Recall (RR) based on the top thirty queries (LCSH descriptors) under Scopus and WOS databases. Table 8 reveals that both databases have good Relative Recall (RR), but comparatively, Scopus (TRR = 18.74 & MRR = 0.62) has a higher Relative Recall than the WOS database (TRR = 11.26 & MRR = 0.38).

Table 9 indicates the Relative Recall (RR) based on the top thirty queries (SLSH terms) under Scopus (TRR = 18.74 & MRR = 0.94) and WOS database, where it is found that comparatively, Scopus has higher Relative Recall than the WOS database (TRR = 11.26 & MRR = 0.56).

5.3 Mean Precision (MP) and Mean Relative Recall (MRR) Analysis

We also compared the Mean Precision (mentioned in Table 10 & Fig. 3) and Mean Relative Recall (mentioned in Table 10 & Fig. 4) of both databases based on three sets of metadata in Economics. It is found that MP is almost the same for three types of metadata but comparatively higher in SLSH terms (MP = 2.34) than others under the Scopus database. Again, MP is more or less the same for three types of metadata, but comparatively, social tags have higher MP (MP = 2.46) than others under the WOS database. Again, in the case of Mean Relative Recall, the study reveals that social tags (MRR = 0.62 in both databases) and LCSH descriptors (MR = 0.62 in both databases) have similar MRR, but comparatively, SLSH terms (MR = 0.94 in Scopus and 0.56 in WOS) have higher MRR in both databases.

[MP – Mean Precision, MRR – Mean Relative Recall, AMP – Average Mean Precision & AMRR – Average Mean Relative Recall]

6. Results and Discussion

Overall, the study reveals that the Mean Precision of social tags, LCSH descriptors and SLSH terms are more or less the same under Scopus and WOS databases. Social tags have comparatively lower Mean Precision in the Scopus search and higher Mean Precision in the WOS search. Again, the study found a positive correlation among three metadata sets in both databases. Even the study found a strong positive

Table 7 — Relative Recall of top thirty Social tags based on Scopus and WOS Database

		Scopus		WOS	
Query No.	Query	NSR	RR	NSR	RR
1	economics	42481	0.61	26756	0.39
2	business	119846	0.71	48268	0.29
3	finance	16850	0.67	8352	0.33
4	capitalism	8060	0.48	8819	0.52
5	ecomomy	79075	0.65	41818	0.35
6	economic history	1245	0.45	1536	0.55
7	political economy	10005	0.52	9280	0.48
8	globalization	19446	0.64	10753	0.36
9	money	23027	0.59	16067	0.41
10	marxism	1970	0.55	1623	0.45
11	socialism	3720	0.47	4147	0.53
12	labor	78789	0.68	37093	0.32
13	investment	54995	0.70	23801	0.30
14	poverty	23520	0.58	16983	0.42
15	economic theory	1478	0.63	850	0.37
16	markets	183377	0.86	29980	0.14
17	economic policy	2709	0.62	1657	0.38
18	financial	73582	0.66	37110	0.34
19	economic development	11420	0.66	5888	0.34
20	international relations	3575	0.53	3219	0.47
21	public policy	10549	0.63	6272	0.37
22	wealth	8918	0.56	7141	0.44
23	business & economics	129	0.72	50	0.28
24	financial crisis	7567	0.70	3281	0.30
25	business history	402	0.54	336	0.46
26	labor history	473	0.57	351	0.43
27	history of economics	85	0.47	94	0.53
28	macroeconomics	1229	0.55	986	0.45
29	austrian economics	156	0.71	63	0.29
30	stock market	11091	0.74	3959	0.26
	TRR		18.47		11.53
	MRR		0.62		0.38

[NSR – No. of sites retrieved, RR – Relative Recall, TRR – Total Relative Recall and MRR – Mean Relative Recall]

Table 8 — Relative Recall of top thirty LCSH descriptors based on Scopus and WOS Database

		Scopus		WOS	
Query No.	Query	NSR	RR	NSR	RR
1	economics	42481	0.61	26756	0.39
2	economc conditions	956	0.67	477	0.33
3	capitalism	8060	0.48	8819	0.52
4	economic policy	2709	0.62	1657	0.38
5	economic history	1245	0.45	1536	0.55
6	financial crises	7437	0.92	604	0.08
7	economic aspects	2297	0.75	772	0.25
8	economic development	11420	0.66	5888	0.34
9	investments	54995	0.70	23801	0.30
10	globalizations	19446	0.64	10753	0.36
11	finance	16850	0.67	8352	0.33
12	international economic relations	55	0.52	51	0.48
13	money	23027	0.59	16067	0.41
14	free enterprise	184	0.54	154	0.46

(*contd.*)

Table 8 — Relative Recall of top thirty LCSH descriptors based on Scopus and WOS Database (*contd.*)

		Scopus		WOS	
15	labor unions	499	0.67	250	0.33
16	monetary policy	6612	0.61	4147	0.39
17	international finance	236	0.56	182	0.44
18	socialism	3720	0.47	4147	0.53
19	economic forecasting	149	0.66	77	0.34
20	income distribution	2037	0.69	926	0.31
21	poverty	23520	0.58	16983	0.42
22	working class	2961	0.48	3199	0.52
23	wealth	8918	0.56	7141	0.44
24	macroeconomics	1229	0.55	986	0.45
25	sustainable development	19288	0.69	8600	0.31
26	business cycles	3875	0.78	1093	0.22
27	microeconomics	355	0.57	266	0.43
28	government policy	1935	0.77	586	0.23
29	capital	58747	0.64	32860	0.36
30	employment	36458	0.63	21407	0.37
	TRR		18.74		11.26
	MRR		0.62		0.38

[NSR – No. of sites retrieved, RR – Relative Recall, TRR – Total Relative Recall and MRR – Mean Relative Recall]

Table 9 — Relative Recall of top thirty SLSH terms based on Scopus and WOS Database

		Scopus		WOS	
Query No.	Query	NSR	RR	NSR	RR
1	economics	42481	0.61	26756	0.39
2	economc conditions	956	0.67	477	0.33
3	capitalism	8060	0.48	8819	0.52
4	economic policy	2709	0.62	1657	0.38
5	economic history	1245	0.45	1536	0.55
6	financial crises	7437	0.92	604	0.08
7	economic aspects	2297	0.75	772	0.25
8	economic development	11420	0.66	5888	0.34
9	investments	54995	0.70	23801	0.30
10	globalizations	19446	0.64	10753	0.36
11	finance	16850	0.67	8352	0.33
12	international economic relations	55	0.52	51	0.48
13	money	23027	0.59	16067	0.41
14	free enterprise	184	0.54	154	0.46
15	labor unions	499	0.67	250	0.33
16	monetary policy	6612	0.61	4147	0.39
17	international finance	236	0.56	182	0.44
18	socialism	3720	0.47	4147	0.53
19	economic forecasting	149	0.66	77	0.34
20	income distribution	2037	0.69	926	0.31
21	poverty	23520	0.58	16983	0.42
22	working class	2961	0.48	3199	0.52
23	wealth	8918	0.56	7141	0.44
24	macroeconomics	1229	0.55	986	0.45
25	sustainable development	19288	0.69	8600	0.31
26	business cycles	3875	0.78	1093	0.22
27	microeconomics	355	0.57	266	0.43
28	government policy	1935	0.77	586	0.23
29	capital	58747	0.64	32860	0.36

(*contd.*)

Table 9 — Relative Recall of top thirty SLSH terms based on Scopus and WOS Database (*contd.*)

			Scopus		WOS	
30	employment		36458	0.63	21407	0.37
		TRR		18.74		11.26
		MRR		0.94		0.56

[NSR – No. of sites retrieved, RR – Relative Recall, TRR – Total Relative Recall and MRR – Mean Relative Recall]

Table 10 — Mean Precision (MP) and Mean Relative Recall (MRR)

Database	Social tags		LCSH descriptors		SLSH terms			
	MP	MRR	MP	MRR	MP	MRR	AMP	AMRR
Scopus	2.24	0.62	2.32	0.62	2.34	0.94	2.3	0.73
WOS	2.46	0.38	2.43	0.38	2.42	0.56	2.44	0.44

[MP – Mean Precision, MRR – Mean Relative Recall, AMP – Average Mean Precision & AMRR – Average Mean Relative Recall]

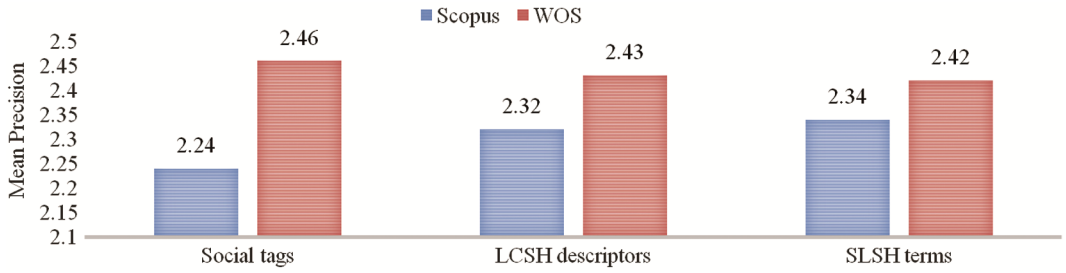

Fig. 3 — Mean Precision

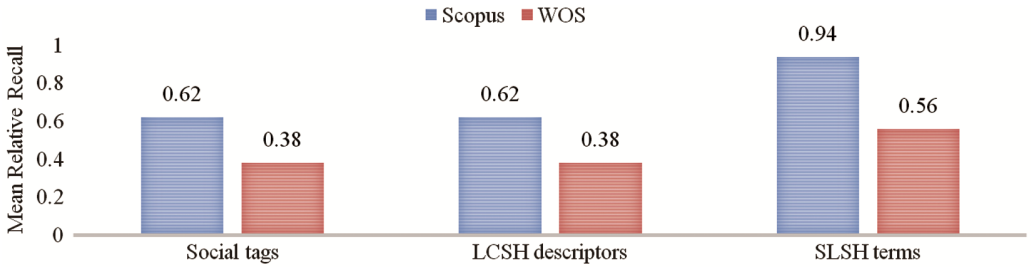

Fig. 4 — Mean Relative Recall

correlation between the Precision of LCSH descriptors and SLSH terms and between the Precision of social tags and SLSH terms in the Scopus database, whereas between the Precision of LCSH descriptors and SLSH terms in the WOS database. Further, in the case of Relative Recall, the study found more or less the same Mean Relative Recall (MRR) for social tags and LCSH descriptors but comparatively higher for SLSH terms in both databases. So, based on that fact, it is identified that both databases are more or less equally effective, but WOS is more effective because of higher Precision, and Scopus is more effective for producing Recall under Economics. Further, it is also identified that social tags are similar and effective to standard subject metadata for generating the same level of Precision and Relative Recall under both databases.

7. Conclusion

Finding the correct information from the vast amount of information is a big challenge for all of us nowadays. Sometimes, it's very difficult to identify which information is accurate, reliable and trustworthy. Online databases like Scopus and WOS are the ones that provide us with reliable scholarly information subject or topic-wise. As web-based resources augment, it is more challenging for information professionals to describe and retrieve those metadata with suitable metadata. With that context, the present study tried to evaluate the best online database that produces the best Precision and relative Recall and the effectiveness of social tags as metadata compared with standard metadata like LCSH descriptors and SLSH terms for document retrieval.

Moreover, the study found that both databases are equally effective, but Scopus has higher Recall and Web of Science has higher Precision comparatively. Further, social tags perform more or less the same standard metadata as LCSH descriptors and SLSH terms in Economics. Before carrying out the study, it ascertains that LCSH descriptors and SLSH terms would be more effective than social tags in document retrieval because of their controlled structure and nature. In contrast, social tags come from an uncontrolled environment and have a lot of terminological issues in document descriptions. But still, the present study reveals that social tags are equally effective metadata as other standard metadata in document retrieval. In that context, the present study recommends that social tags can be used in libraries along with LCSH descriptors and SLSH terms for document description.

References

1 Samanta K S and Rath D S, Measuring the applicability of user-generated social tags along with expert-generated LCSH descriptors in Sociology: a heuristic study, *Annals of Library and Information Studies*, 68(March) (2021) 28-38.

2 Abramo G, Revisiting the scientometric conceptualization of impact and its measurement, *Journal of Informetrics,* 12(3) (2018) 590-597. https://doi.org/10.1016/j.joi.2018.05.001

3 Agarwal A et al., Bibliometrics: tracking research impact by selecting the appropriate metrics, *Asian Journal of Andrology*, 18(2) (2016) 296-309. https://doi: 10.4103/1008-682X.171582

4 Aksnes D W, Langfeldt L and Wouters P, Citations, citation indicators, and research quality: An overview of basic concepts and theories, *Sage Open*, 9(1) (2019)

5 https://doi.org/10.1177/2158244019829575

6 Moed H F and Halevi G, Multidimensional assessment of scholarly research impact, *Journal of the Association for Information Science and Technology*, 66(10) (2015) 1988-2002.

7 Bianco M, Gras N and Sutz J, Academic evaluation: Universal instrument? Tool for development? *Minerva*, 54 (2016) 399-421. https://doi.org/10.1007/s11024-016-9306-9

8 De Rijcke S et al., Evaluation practices and effects of indicator use—a literature review, *Research evaluation*, 25(2) (2016) 161-169. https://doi.org/10.1093/reseval/rvv038

9 Kun Á, Publish and who should perish: you or science? *Publications*, 6(2) (2018) 18. https://doi.org/10.3390/publications6020018

10 Dukić D, Online databases as research support and the role of librarians in their promotion: The case of Croatia, *Library Collections, Acquisitions, and Technical Services*, 37(1-2) (2013) 56-65. https://doi.org/10.1016/j.lcats.2013.09.005

11 Morrison P J, Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web, *Information Processing & Management*, 44(4) (2008) 1562-1579.

12 https://doi.org/10.1016/j.ipm.2007.12.010

13 Ani O E and Ahiauzu B, Towards effective development of electronic information resources in Nigerian university libraries, *Library Management*, 29(6-7) (2008) 504-514.

14 https://doi.org/10.1108/01435120810894527

15 Pranckute R, Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World, *Publications*, 9 (2021) 12.

16 https://doi.org/10.3390/publications9010012

17 Ritchie S, Banyas K and Sevin C, A Comparison of Selected Bibliographic Database Search Retrieval for Agricultural Information, *Issues in Science and Technology Librarianship*, (2019) 93. https://doi.org/10.29173/istl48

18 Stokes P, Foster A and Urquhart C, Beyond relevance and recall: testing new user-centred measures of database performance, *Health Information & Libraries Journal*, 26(3) (2009) 220-231. https://doi.org/10.1111/j.1471-1842.2008.00822.x

19 Walters W H, Google Scholar search performance: Comparative Recall and Precision, *Portal: Libraries and the Academy*, 9(1) (2009) 5-24. https://doi.org/10.1353/pla.0.0034

20 Sewell R R, Comparing four CAB abstracts platforms from a veterinary medicine perspective, *Journal of electronic resources in medical libraries*, 8(2) (2011) 134-149. https://doi.org/10.1080/15424065.2011.576608

21 CheshmehSohrabi M and Sadati E A, Performance evaluation of web search engines in image retrieval: An experimental study, *Information Development*, 38(4) (2022) 522-534.

22 https://doi.org/10.1177/02666669211010211

23 Greisdorf H, Relevance thresholds: a multi-stage predictive model of how users evaluate information, *Information Processing & Management*, 39(3) (2003) 403-423.

24 https://doi.org/10.1016/S0306-4573(02)00032-8

25 Razikin K, Goh D H, Chua A Y K and Lee C S, Social Tags for Resource Discovery: A Comparison between Machine Learning and User-Centric Approaches, *Journal of Information Science*, 37(4) (2011) 391-404.

26 Harter S P, Variations in Relevance Assessments and the Measurement of Retrieval Effectiveness, *Journal of the American Society for Information Science*, 47(1) (1996) 37-49.

27 https://doi.org/10.1002/(SICI)1097-4571(199601)47:1<37::AID-ASI4>3.0.CO;2-3

28 Sormunen E A, Method for Measuring Wide Range Performance of Boolean Queries in Full Text Databases, PhD thesis, University of Tampere. 2000.

29 https://trepo.tuni.fi/bitstream/handle/10024/67002/951-44-4732-8.pdf?sequence=1 (Accessed on 10 December 2022)

30 Chu H and Rosenthal M, Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology, *Proceedings of the Annual Meeting-American Society for Information Science*, 33 (1996) 127-35.

31 Saracevic T et al., Letters to the Editor, *Information Processing & Management*, 39(1) (2003) 153-56.

32 Gul S, Ali S and Hussain A, Retrieval performance of Google, Yahoo and Bing for navigational queries in the field of life science and biomedicine, *Data Technologies and Applications*, 54(2) (2020) 133-150. https://doi.org/10.1108/DTA-05-2019-0083

33 Su L T, Value of search results as a whole as the best single measure of information retrieval performance, *Information Processing & Management*, 34(5) (1998) 557-579.

34 https://doi.org/10.1016/S0306-4573(98)00023-5
35 Usmani T A, Pant D and Bhatt A K, A comparative study of Google and Bing search engines in context of Precision and relative recall parameter, *International Journal on Computer Science and Engineering*, 4(1) (2012) 21.
36 Su L T, The relevance of Recall and Precision in user evaluation, *Journal of the American Society for Information Science*, 45(3) (1994) 207-217.
37 https://doi.org/10.1002/(SICI)1097-4571(199404)45:3<207::AID-ASI10>3.0.CO;2-1
38 Kumar B S and Prakash J N, Precision and relative recall of search engines: A comparative study of Google and Yahoo, *Singapore Journal of Library & Information Management*, 38(1) (2009) 124-137.
39 Shafi S M and Ali S, Retrieval performance of select search engines in the field of physical sciences, *Annals of Library and Information Studies*, 66 (September) (2019) 117-122.
40 CheshmehSohrabi M and Sadati E A, Performance evaluation of web search engines in image retrieval: An experimental study, *Information Development*, 38(4) (2022) 522-534.
41 https://doi.org/10.1177/02666669211010211