August 2021

REVISION AND QUALITY ANALYSIS OF LIBRARY AND INFORMATION SCIENCE CONCEPTS IN WIKIPEDIA

Chaitali Ghosh
Jadavpur University, chaitali.ghosh89@gmail.com

Prof. Udayan Bhattacharya
Jadavpur University, ubhattacharyya@libsc.jdvu.ac.in

REVISION AND QUALITY ANALYSIS OF LIBRARY AND INFORMATION SCIENCE CONCEPTS IN WIKIPEDIA

Chaitali Ghosh
Jadavpur University, Department of Library and Information Science, Main Campus, 188, Raja S.C Mallick Rd, Kolkata, West Bengal-700032, India,
chaitali.ghosh89@gmail.com

Prof. Udayan Bhattacharya
Professor, *Jadavpur University*, *Department of Library and Information Science, Main Campus, 188, Raja S.C Mallick Rd, Kolkata, West Bengal-700032, India,*
ubhattacharyya@libsc.jdvu.ac.in

ABSTRACT

Wikipedia is public encyclopedia contains articles on different subjects and different topics. The objective of the study is to evaluate the standard and status of revision of the library related articles. History page of Wikipedia article contains statistics of page from where the status and standard of the page has been derived. It has been concluded that Wikipedia covers 67 concepts of DDC 23rd edition arrived in 020 class (Library and Information Science Class). The content of Library and Information Science (LIS) related articles in Wikipedia was not enough informative as per Wikiproject. The concepts related to information science i.e. World Wide Web has been edited more times and it has been edited by more editors.

Keywords: Library and Information Science, Wikipedia, Revision of Wikipedia

INTRODUCTION

Wikipedia is an online public encyclopedia and it is multilingual, open-content. It is written collaboratively by people with varying degrees of expertise. Anyone can access, edit Wikipedia's content and may revise any of its content. It contains a large number of articles on Library & Information Science related topics, but the quality of this information, frequency of revision are unknown. This study has been done to know the changing of wikipedia's article till now and the standard of Wikipedia's article on library and information science by taking the viewpoint of Wikiproject members.

The universe of knowledge is changing its direction and dimension gradually. This continuous change of the structure of the subject has led to change the structure of the article in Wikipedia. Collections, Organization, dissemination are the main three pillar of library concept. But the dimension of the library concept is changing with the development of technology. Wikipedia helps changing their article on library and information science day by day with author

collaboration. Wikipedia is ideal for such activities because it helps people to track information on different subject.

SCOPE AND COVERGE OF RESEARCH

The concepts have been taken from the class of 020 of Dewey Decimal Classification (DDC) Scheme. All concepts of DDC 020 class numbers are not written in Wikipedia but 67 concepts are enlisted in Wikipedia which has been taken for the study. The standard of articles has been determined and frequency of revision has been determined by the study.

LITERATURE REVIEW

Stvilia worked on the organization of IQ assurance work in a large-scale, open, collaborative encyclopedia—Wikipedia. As per Stvilia, Wikipedia as a resource where the quality discussions and processes are strongly connected to the data itself and are accessible to the general public. This openness makes it particularly easy for researchers to study a particular kind of collaborative work that is highly distributed and that has a particularly substantial focus, not just on error detection but also on error correction. (Stvilia et al., 2008)

The online encyclopedia, Wikipedia provides multitude of network structures such as the citation network of its pages or the co-authorship network of users. It can be observed that the edit volume of Wikipedia pages varies strongly over time. The study established normalization method on wikipedia's article. (Brandes et al., 2007)

Wilkinson et al. studied the editorial activity of high quality Wikipedia articles and concluded that Wikipedia was being edited by people frequently and the edition a stochastic process, where the number of new edits to a given article in a given period of time is a randomly varying percentage of the total number of previous edits. (Wilkinson et al., 2007)

Wikipedia is the "free online encyclopedia that anyone can edit", having a huge impact on gathering information about the world by people. So, epistemologists and information scientists should certainly be trying to identify changes to Wikipedia that will bring about even better epistemic consequences. He suggested that to improve Wikipedia, for that it is to clarify what epistemic values are and to better understand why Wikipedia works as well as it does. (Fallis, 2008)

STATEMENT OF THE PROBLEM

Revision of Wikipedia is the stage of writing process or editing process. It contains addition and removal of content. It is ongoing process for professional writers to go through many drafts and

revisions before successfully creating an essay. Article in Wikipedia also has been revised by many stages. The articles on library and information science have been selected for our study. The development of the content of library science is being developed day by day with the change of technology. The previous concepts are going to obsolete. The article in Wikipedia is changing accordingly. The study has been made to know the following problems

- How the revision of Wikipedia has been taken place?
- How many editors are working on revision of article?
- How many times the article is edited and average edit per editor?
- Which article is currently updated and mostly revised?
- How many times the pages are viewed?
- How many link to the page and link from the page attached with the article?
- What is the status of articles of Library and Information Science in Wikipedia in qualitative concerned?

OBJECTIVE OF RESEARCH

Objective of this study is to develop quantitative measurement to determine the quality of articles in Wikipedia. The revision, addition and completeness of the information on Wikipedia are done by experts and non-experts can contribute to the Wikipedia's content. Thus, the intention of my research is to systematically evaluate the quality and completeness, revision of a sample of library-related articles on Wikipedia. The main objectives has been as follows-

- To know the process of revision on wikipedia
- To determine the frequency of edit of the Library and Information Science (LIS) article and to identify the most revised article on the concepts of library and information science
- To determine the average time between edit, average edit per editor, date of last revision and first revision, number of page views, link to the page and link from the page of Wikipedia article which can represent the utility of page
- To determine the quality of LIS article as per member of Wikiproject

METHODOLOGY

The Wikipedia has been searched for articles on library and information science. The articles have been taken for determining quality of content, links, readability, frequency of updating for my study. Articles have been collected on the concepts of library and information science mentioned in DDC 020 class number. All concepts on DDC are not written in Wikipedia till now. 67 written concepts have been taken for evaluation study of quality of articles and utility of article of Library and Information Science in Wikipedia.

Aims of WikiProject Systems is to improve Wikipedia's coverage of theory and practice. Members of WikiProjects also assist in performing quality assessments of articles pertaining to their specific topic area. Quality grading is as follows-

- Featured articles(FA) ⭐: professional and outstanding writing from reliable source
- A-class articles(A) 🅐 : very useful for readers
- Good articles(GA) ⊕: useful for almost all readers.
- B-class articles(B) 🅑 Contents may not complete enough to satisfy serious students and researchers
- C-class articles(C) ©: useful to a casual readers
- Start-class articles (SC)⬤: provides some meaningful content but writing is very short, readers need more
- Stub Class articles: provides very little meaningful content
- Featured List: writing having professional standard, having set format
- List : there is no set format for writing

From these labeling done by Wikipedia itself has been easily identified quality of article by categorization

DATA COLLECTION AND INTERPRETATION

Table 1 : Showing Total edits, Editors, Average time between edit, Average edit per user, First edit, Latest edit, Page Views within 60days, Link to the page ad Link from the page of Articles in Wikipedia

Serial No	Concepts	Total Edits [A]	Editors [B]	Average time b/w edit [C]	Average Edit per User [D]	First Edit [E]	Latest Edit [F]	Page views (60days) [G]	Link to the page [H]	Link from the page [I]

1	Library technical services	52	33	102.3	1.6	3.9.2006	29.3.2021	1459	108	121
2	Information retrieval	1104	637	6.4	1.7	22.11.2001	29.3.2021	30209	1015	166
3	World wide web	8514	4442	0.8	1.9	1.11.2001	19.04.2021	258601	6094	739
4	Web sites	6852	3543	1	1.9	1.11.2001	17.3.2001	223802	10376	299
5	Searching	611	327	1.9	1.9	24.0`1.2018	30.3.2021	67603	84	68
6	Retrieval	36	30	147.6	1.2	21.04.2006	5.11.2020	481	28	14
7	Search Engine	4292	2120	1.3	2	13.02.2006	19.04.2021	118169	1655	316
8	Semantic Web	2330	1336	3	1.7	7.12.2001	16.04.2021	7184	1264	392
9	Reprography	98	76	67.9	1.3	13.12.2002	1.03.2021	41245	56	24
10	Collection development	133	73	43.3	1.8	18.05.2005	27.02.2021	4818	116	113
11	Censorship	5676	2929	1.3	1.9	5.04.2001	19.04.2021	63053	6002	864
12	Weeding	483	275	13.4	1.8	15.06.2003	18.3.2021	14161	397	356
13	Bibliography	2182	1228	3	1.8	2.03.2003	1.04.2021	41259	3697	146
14	Cataloguing in publication	43	39	136.7	1.1	19.01.2005	24.02.2021	997	95	29
15	ISBN	2040	1130	3.5	1.8	18.01.2001	14.04.2021	78184	10407	587
16	Catalogue	132	107	44.8	1.2	19.01.2005	5.04.2021	2205	77	32
17	Online catalogue	29	22	174.5	1.3	25.06.2006	3.05.2020	282	13	4
18	Descriptive catalogue	62	28	77.1	2.2	14.08.2007	17.09..2020	219	199	226
19	MARC	352	230	17.1	1.5	15.06.2004	9.12.2020	9066	226	115
20	Authority control	445	283	13.3	1.6	18.01.2005	14.04.2021	6234	506	211
21	Uniform Title	42	30	117.5	1.4	17.09.2007	21.03.2021	382	23	35
22	Cataloguing	278	126	2.2	0.1	17.11.2005	20.02.2021	9871	233	193
23	Classification	413	271	15.8	1.5	23.05.2003	1.04.2021	15190	239	40
24	Indexing	948	616	6.1	1.5	4.03.2001	5.06.2021	35741	358	108
25	Manuscript	798	481	9.1	1.7	20.05.2001	14.04.2021	26191	4328	298

26	Pamphlets	370	273	17.6	1.4	10.07.2003	22.04.2021	14057	978	183
27	Serial	178	118	32.3	1.5	17.07.2005	21.04.2021	2814	133	35
28	Report	778	504	7	1.5	10.04.2006	13.02.2021	38079	488	73
29	Maps	3984	2158	1.8	1.8	7.11.2001	20.04.2021	75184	4014	384
30	Atlases	1012	638	7	1.6	21.11.2001	23.04.2021	26794	761	281
31	Globe	1546	990	4.2	1.6	11.08.2003	27.04.2021	28545	1132	131
32	Graphical projection	537	295	12	1.8	7.09.2003	20.04.2021	29014	217	216
33	Film	3535	1840	2	1.9	26.09.2001	15.04.2021	180486	18487	516
34	Slide	281	182	22.8	1.5	25.09.2003	8.04.2021	1777	53	58
35	Video	2158	1298	3.3	1.7	26.09.2001	24.04.2021	114094	3086	774
36	Sound recording	1319	670	4.4	2	9.08.2005	27.04.2021	34923	6517	478
37	Realia	48	29	75.4	1.7	7.01.2010	5.12.2019	469	23	40
38	Braille	2433	1380	2.9	1.8	4.11.2001	20.11.2021	52591	2325	948
39	Microform	511	333	11	1.5	5.10.2005	7.02.2021	13357	373	160
40	Game	3784	1971	1.9	1.9	1.11.2001	2.05.2021	149821	2184	349
41	Toy	3685	1956	1.8	1.9	22.11.2002	5.05.2021	39324	5554	280
42	Abstract	207	151	29.7	1.4	27.03.2004	26.01.2021	2139	150	14
43	Natural language processing	1441	831	5	1.7	22.09.2001	26.04.2021	110358	1659	252
44	Dewey Decimal Classification	2001	1181	3.6	1.7	25.10.2001	10.04.2021	99970	7190	143
45	Universal Decimal Classification	319	206	22	1.5	4.02.2002	22.04.2021	11270	377	166
46	Library of Congress Classification	661	331	11.7	1.8	25.10.2001	6.05.2021	18694	727	442
47	Bliss Bibliographic Classification	117	65	59.9	1.8	19.12.2001	24.02.2021	1269	57	73
48	Colon Classification	179	129	39.4	1.4	25.10.2001	19.02.2021	8529	127	83
49	Tag	752	500	8.3	1.5	14.01.2001	18.03.2021	6196	153	82

50	SDI (Selective Dissemination of Information)	47	35	109.6	1.3	30.12.2006	7.02.2021	1570	19	12
51	Library Circulation	59	44	88.9	1.3	23.09.2006	30.01.2021	3598	150	171
52	Interlibrary Loan	303	173	21.6	1.8	18.01.2003	27.12.2020	3924	387	48
53	Preservation	116	75	45.4	1.5	5.11.2004	13.04.2019	1062	68	25
54	Library	3985	2189	1.8	1.8	9.11.2001	3.05.2021	66632	5785	390
55	Archives	860	503	7.5	1.7	12.09.2003	15.04.2021	30603	1475	342
56	Information Center	258	135	23.5	1.9	16.07.2004	3.03.2021	3749	538	266
57	Private Library	217	117	26	1.9	27.08.2005	28.01.2021	3056	142	264
58	Public Library	1105	553	6	2	11.02.2003	12.04.2021	24949	2670	396
59	National Library	293	203	21.2	1.4	1.05.2004	29.4.2021	9328	468	225
60	Special Library	228	112	21.7	2	25.08.2007	23.03.2021	8036	179	159
61	Prison Libraries	123	64	27.8	1.9	28.11.2011	7.04.2021	911	216	282
62	Academic library	255	153	21	1.7	9.09.2006	7.05.2021	10089	258	147
63	School Library	307	175	17.7	1.8	13.04.2006	1.03.2021	13570	195	138
64	Library Stack	64	25	39.6	2.6	7.05.2014	14.04.2021	1783	38	96
65	Lighting in Libraries	21	15	205.7	1.4	11.02.2007	10.12.2018	320	7	6
66	Library management	169	107	41.7	1.6	9.11.2001	19.02.2021	4348	139	131
67	Human resource management	2254	1450	2.7	1.8	27.09.2004	2.05.2021	95372	1036	321

Table 1 shows the total edit of the page, editors, average time between edit, average edit per users, first edit, latest edit, page views within 60 days, link to page and link from the page. Article on World Wide Web has been edited more and more number of editors is working to edit the contents. World Wide Web and websites are viewed more and number of links to the page having article on website and ISBN is high and number of links from the page contains article on censorship is high. Most of the articles are edited currently in 2021.
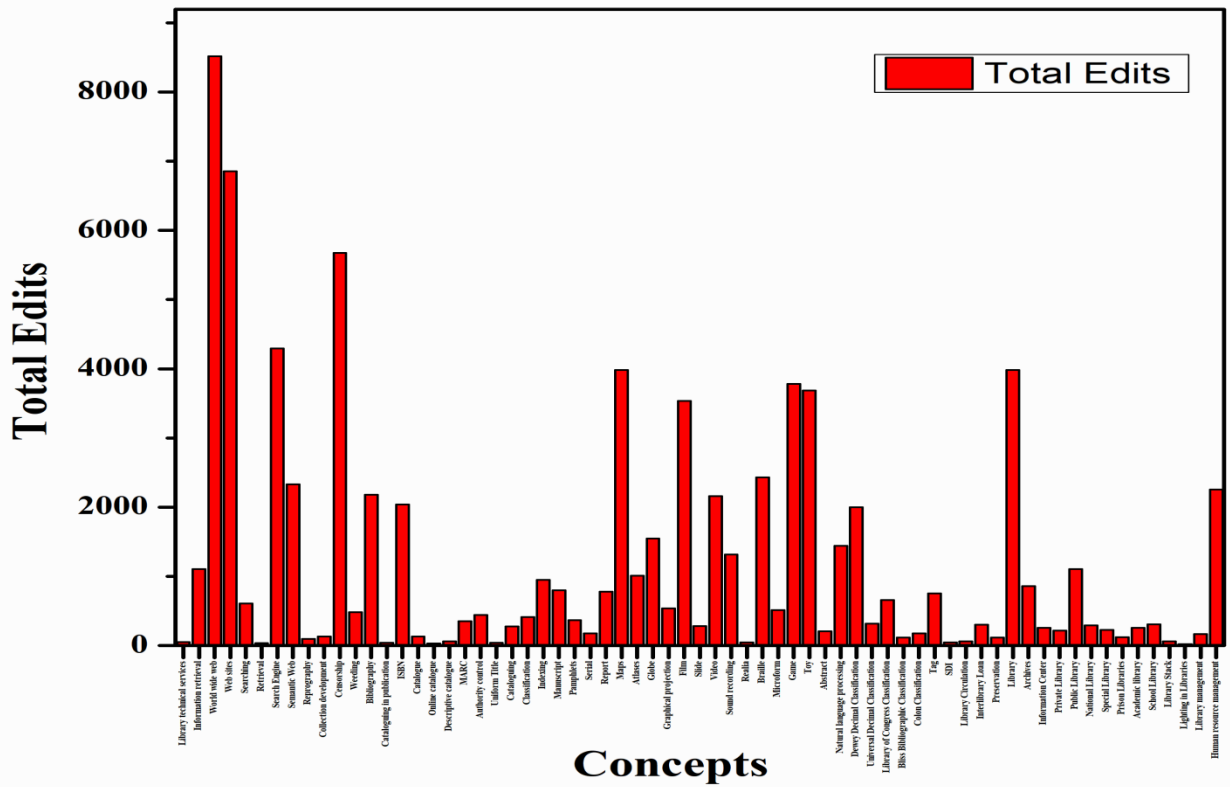
FIGURE 1 : Showing total edits of article on library and information science related concepts

FIGURE 2: Showing editors working on revision of article on library and information science concepts

FIGURE 3: Showing Page views by readers within 60 days

FIGURE 4 : Showing link to the page and link from the page of article

Table2: Showing quality of article

SL	Concepts	Class
1	Library technical services	Stub Class
2	Information retrieval	Start class
3	World wide web	Featured article
4	Web sites	C-Class

5	Searching	Not mentioned
6	Retrieval	Not mentioned
7	Search Engine	C-Class
8	Semantic Web	Start class
9	Reprography	Not mentioned
10	Collection development	Stub-Class
11	Censorship	C-Class
12	Weeding	Start class
13	Bibliography	Start class
14	Cataloguing in publication	Stub Class
15	ISBN	C-Class
16	Catalogue	Not mentioned
17	Online catalogue	Not mentioned
18	Descriptive catalogue	Start class
19	MARC	Not mentioned
20	Authority control	B-Class
21	Uniform Title	Stub Class
22	Cataloguing	Start class
23	Classification	Stub Class
24	Indexing	Not mentioned
25	Manuscript	C-Class
26	Pamphlets	Not mentioned
27	Serial	Not mentioned
28	Report	Start class
29	Maps	Start class
30	Atlases	Start class
31	Globe	Start class
32	Graphical projection	Start class
33	Film	C-Class

34	Slide	Not mentioned
35	Video	C-Class
36	Sound recording	C-Class
37	Realia	Start class
38	Braille	B-Class
39	Microform	Start class
40	Game	C-Class
41	Toy	C-Class
42	Abstract	Not mentioned
43	Natural language processing	C-Class
44	Dewey Decimal Classification	GA Class
45	Universal Decimal Classification	Start class
46	Library of Congress Classification	List Class
47	Bliss Bibliographic Classification	Start class
48	Colon Classification	Start class
49	Tag	Not mentioned
50	SDI (Selective Dissemination of Information)	Not mentioned
51	Library Circulation	Start class
52	Interlibrary Loan	Start class
53	Preservation	Start class
54	Library	C-Class
55	Archives	Start class
56	Information Center	Not mentioned
57	Private Library	Stub class
58	Public Library	C-Class
59	National Library	Start class
60	Special Library	Stub Class

61	Prison Libraries	Start class
62	Academic library	C-Class
63	School Library	Start class
64	Library Stack	Start class
65	Lighting in Libraries	Start class
66	Library management	Start class
67	Human resource management	C-Class

Table2 presents the standard of articles mentioned by Wikipedia itself. It represents only 1 article in under featured class, 1 in GA class, 1 in List class, 2 in B class, 15 in C class, 26 in start class, and 7 in stub class. Rest of the articles' classes is not mentioned. It has been graphically presented by Figure 5.

TABLE 2.1 : Number of articles belonging in different class

Number of article under FA	1
Number of article under class B	2
Number of article under class C	15
No of article under class GA	1
No of article under List Class	1
No of article class not mentioned	14
No of article under start class	26
No of article under stub class	7
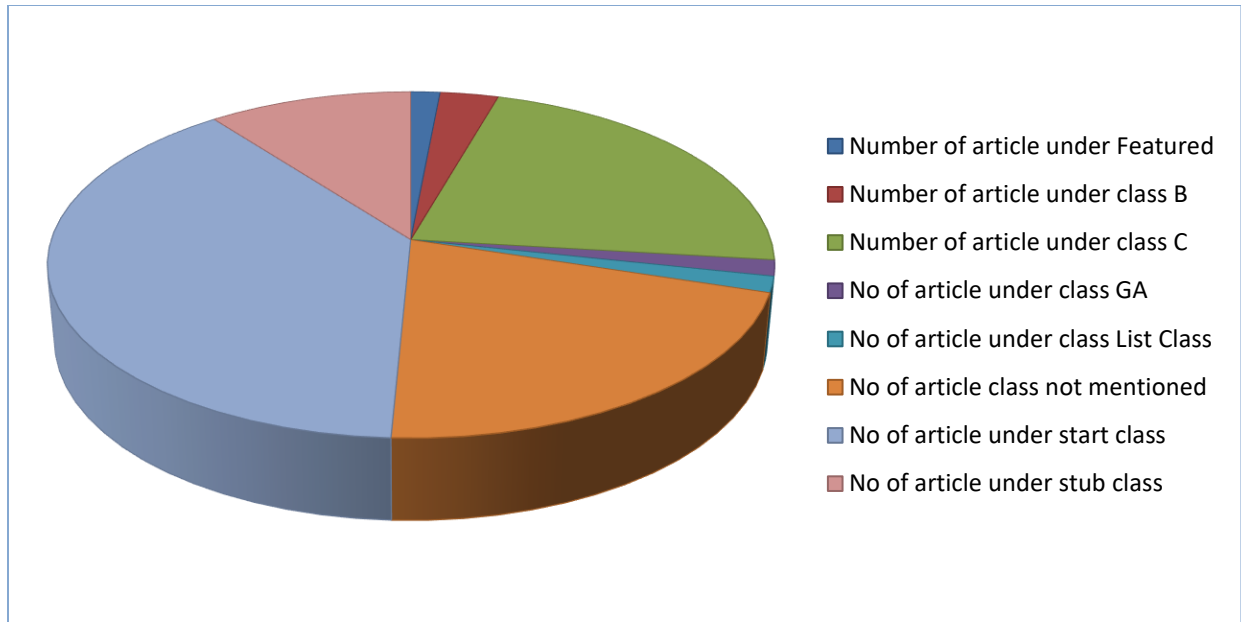
FIGURE 5: Showing distribution of articles according class

FINDINGS

Revision of Wikipedia is taken place by editor. General people can edit the Wikipedia article by giving proper reference. Column B in Table no 1 and Figure 2 identifies the number of editors working on revision of Wikipedia article about library and information science concepts.Column A in Table no 1 and Figure 1 identifies total edits and column D identifies average edit per users. From the data, it is seen that world wide web is mostly revised article among all selected articles. "Tag", the term firstly appeared as wikipedia's article and academic library are edited currently Column G in Table no 1 shows the page views within 60 days and it is represented by Figure 3. The page view indicates the utility of the page. The article on World Wide Web is viewed more times. Column H and I in Table no 1 shows the number of links to the page and from the page and that is represented graphically by Figure-4. The standard of the content of articles are not informative. Only one article out of 67 belongs to featured class. Most of the articles i.e. 26 articles are listed under start class that means contents are meaningful but readers want more.

CONCLUSION

Wikipedia contains articles on different topics or concepts. 67 Library and Information science related concepts among all concepts in DDC 020 class are also available in Wikipedia and those are also revised frequently and so many editors are working on revision purpose. The qualities of content of articles are not informative but only one article i.e. World Wide Web is listed under Featured class. Most of the articles belong to start class. Readers are not fully satisfied, they need

more while reading the library and information science articles. Moreover, it can be said that articles on information science related are more informative and frequency of revision is high.

REFERENCES

Brandes, U.& Lerner, J.(2007). Revision and Co-revision in Wikipedia: Detecting Clusters of Interest.Proceedings of International Workshop Bridging the Gap Between Semantic Web and Web 2.0, 4th European Semantic Web Conference.

Brandes, U., Kenis, P., Lerner, J., & Van Raaij, D. (2009, April). Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th international conference on World wide web* (pp. 731-740).

Brandes, U., & Lerner, J. (2007). Revision and co-revision in Wikipedia: Detecting clusters of interest. In *ESWC*.

Cabonero,D.A., &Dolendo,R.B.(2013). Cataloging and Classification Skills of Library and Information Science Practitioners in their Workplaces: A Case Analysis.*Library Philosophy and Practice* [e-journal]

Fallis, D.(2008).Toward an Epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*. 59(10),1662–1674.

Ferschke, O., Zesch, T., & Gurevych, I. (2011, June). Wikipedia revision toolkit: Efficiently accessing wikipedia's edit history. In *Proceedings of the ACL-HLT 2011 System Demonstrations* (pp. 97-102).

Jurgens, D., & Lu, T. C. (2012, May). Temporal motifs reveal the dynamics of editor interactions in Wikipedia. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 6, No. 1).

McDonough Dolmaya, J. (2015). Revision history: Translation trends in Wikipedia. *Translation Studies*, *8*(1), 16-34.

Mittal, R.(2011). Library and Information Science Research trends in India.*Annals of Library and Information Studies*. 58, 319-325

Nunes, S., Ribeiro, C., & David, G. (2008, September). WikiChanges: exposing Wikipedia revision activity. In *Proceedings of the 4th International Symposium on Wikis* (pp. 1-4).

Purdy, J. P. (2009). When the tenets of composition go public: A study of writing in Wikipedia. *College Composition and Communication*, *61*(2), W351.

Stvilia, B., Twidale, M. B., Smith, L. C., &Gasser, L.(2016).Information quality work organization in Wikipedia.*Journal of the American Society for Information Science and Technology*.59(6).983–1001

West, A. G., Kannan, S., & Lee, I. (2010, April). Detecting wikipedia vandalism via spatio-temporal analysis of revision metadata?. In *Proceedings of the Third European Workshop on System Security* (pp. 22-28).

Wilkinson, D. M., & Huberman, B. A. (2007, October). Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis* (pp. 157-164).

Wilkinson, D.M.,&Huberman, B.A.(2007) Cooperation and quality in wikipedia. In WikiSym '07: Proceedings of the 2007 international symposium on Wikis, 157 Conference proceeding of ACM,Newyork

Yamangil, E., & Nelken, R. (2008, June). Mining Wikipedia revision histories for improving sentence compression. In *Proceedings of ACL-08: HLT, Short Papers* (pp. 137-140).