Library Philosophy and Practice (e-journal)   Libraries at University of Nebraska-Lincoln

DDC-wise subject mapping of NDLI indexed multileveled, multilingual resources: A special reference with Indian Languages

Debasis Majhi
Indraprastha Institute of Information Technology, Delhi, debasismajhidlis@gmail.com

Sibsankar Jana
University of Kalyani, sibs_jana@yahoo.com

Amit Nath
University of Kalyani, India, amitnathdlis@gmail.com

Soumik Kerketta
University of Kalyani, India, soumik.kerketta.bdy@gmail.com

DDC-wise subject mapping of NDLI indexed multileveled, multilingual resources: A special reference with Indian Languages

Debasis Majhi

Library and information Assistant (IIIT-Delhi)

debasismajhidlis@gmail.com

Sibsankar Jana (Mentor)

Department of Library and Information Science, University of Kalyani, Kalyani, West Bengal, India. E-mail: sibs_jana@yahoo.com

Amit Nath

Research Scholar, Dept. of Library and Information Science, University of Kalyani.
Email- amitnathdlis@gmail.com

Soumik Kerketta

Dept. of Library and Information Science, University of Kalyani
Email- soumik.kerketta.bdy@gmail.com

Abstract: The study helped with the subject mapping of the resources on 22 languages included in Indian constitution that were indexed in the National Digital Library of India from 2018 and covering the different fields of arts and humanities, social sciences, science, engineering, technology, and history. The subject mapping is based on the Dewey Decimal Classification (DDC) 22nd edition. The relevant/required data was collected from the official website of the National Digital Library of India (https://ndl.iitkgp.ac.in/) on 4th April 2021, Sunday by browsing subject-wise.
(https://ndl.iitkgp.ac.in/result?q={"t":"subject","b":{"browse":"subject","filters":[]}}) and filtering one by one of 22 officially recognized Indian languages in the Indian constitution. A total number of 398793 documents on different subjects indexed in NDLI till 4th April 2021, The maximum number of documents indexed in Computer science, Information & general works (000) subject [88.98%] and the larger number of documents indexed in the Bengali language [40.71%].

Keywords: NDLI, Subject mapping, Indian languages, virtual learning platform, IIT Kharagpur.

Introduction

The National Digital Library of India (NDLI) is a project under the Ministry of Human Resource Development (HMRD) of the Government of India. The objective of the study is to collect and collate metadata and provide a full-text index from several national and international digital libraries, as well as other relevant sources.["National Digital Library of India, 2021"]. NDLI has indexed the maximum Indian language document in different subjects. More than 19,500 languages or dialects are spoken in India as mother tongues, according to the latest analysis of a census 2011, and there are 121 languages which are spoken by 10,000 or more people in India, which has a population of 121 crores ["Languages of India, 2021"]. The Eighth Schedule to the Constitution of India lists the 22 official languages of the Republic of India in census 2011. These 22 recognized languages are Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kashmiri, Kannada, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santhali, Sindhi, Tamil, Telugu, and Urdu. But Hindi is the most widely spoken language in India (43.63%), and the second is the Bengali language (8.03%).

Review of literature

There have been a few studies conducted on individual journals. Some of the relevant studies in the aforesaid directions are worthy of review. Das et al. (2016) analyzed in a study that how the NDLI will serve as a pan-India virtual teaching-learning-evaluation-knowledge discovery and innovation platform which will help personalized, self-paced, new-age multimedia education at all levels, act as a platform that can bring a fundamental shift in the paradigm of education and research and thus become a true national asset. Later Mangurkar & Chaudhari (2018) have analyzed different languages covered in NDLI, educational degrees covered, and different subjects available on the National Digital Library of India. They analyzed a total of 15,302,524 documents by contents, a total of 5,760,00 9 documents by educational level, a total of 2,592,258 documents language-wise, and a total of 16,730,585 by subject distribution based on data available from 22nd February to 15th March 2018.

Bashir, Nasreen & Loan (2019) have analyzed in the study an overview of the National Digital Library of India (NDL India) to understand its advantages, features, and collection in the global

digital space and explored the service architecture, use case, content types and file formats of the National Digital library of India and enumerated its salient features and issues. In another study, Nasreen, N. et al. (2019) found that ten broad subjects covered under the collection of the WDL are Computer Science, Information and general works, Philosophy and Psychology, Religion, Social Sciences, Language, Natural science and Mathematics, Technology (Applied Sciences), The Arts; fine and decorative arts, Literature and Rhetoric, History and Geography. The maximum number of resources belongs to History and Geography (16135) followed by Computer Science Information and general works (3560), Social Sciences (3163), The Arts; fine and decorative arts (3124) and the minimum number is in Philosophy and Psychology (140).

Moid, Raza & Jahan (2020) examined the study of the grey literature (GL) in the National Digital Library of India (NDLI) with particular reference to the synopsis, thesis, report, and patent. As a result, the majority of GL resources were extracted, and the maximum number of resources is available in the English language. Mishra (2019) has explored the types of content available at the NDLI and their pattern of the growth of learning resources. Also, she has covered the educational patterns of learning resources with language coverage. From the literature review, it is found that there is no specific study of subject mapping on the documents of Indian languages indexed in the National Digital Library of India (NDLI). Therefore, this is the research gap we have identified from this literature review.

Objectives of the study

The main objective of this study is to find out the subject-wise distributions of Indian language documents that were indexed in NDLI. More specific objectives are as follows:

- ❖ To examine the growth of Indian language documents on different subjects
- ❖ To identify the subject-wise distribution of Indian language documents
- ❖ To identify the access restriction of Indian language documents
- ❖ To identify the distribution of educational level of Indian language documents
- ❖ To identify the content-wise distribution of Indian language documents

Methodology

The present study is based on subject mapping of those documents that were published in the Indian language and indexed in the National Digital Library of India (NDLI). The data was collected by subject-wise browsing from the official website of the National Digital Library of India (https://ndl.iitkgp.ac.in/) which is shown in Figure-1. NDLI has classified all types of indexed documents according to the DDC classification scheme. In the first level of subject NDLI has used the first level of summary of DDC classification scheme and this method helps to subject-wise required documents collection (https://ndl.iitkgp.ac.in/result?q={"t":"subject","b":{"browse":"subject","filters":[]}}).
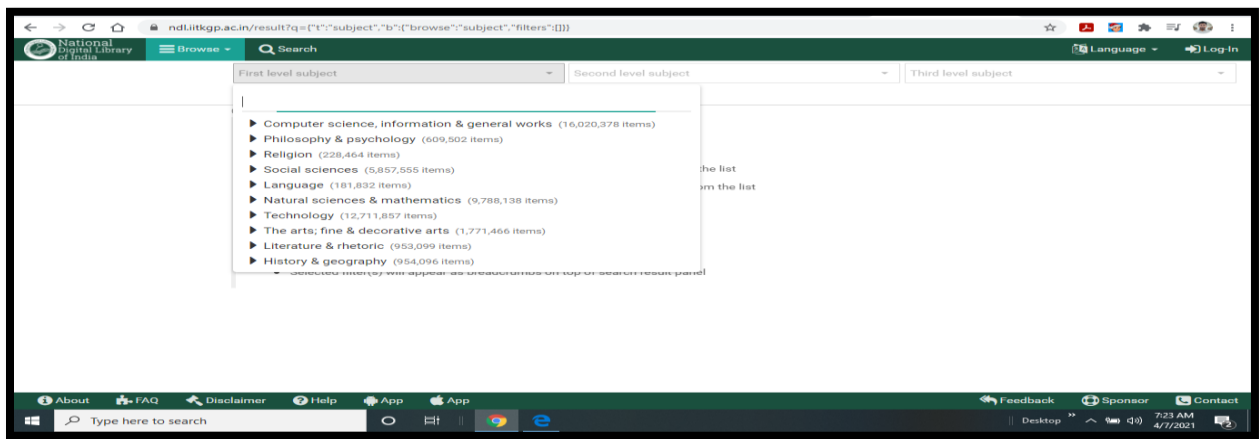

Figure-1: Subject classification of NDLI in the first level of subject

Analysis of Data

We analyzed collected data based on different criteria viz subject-wise distribution of Indian language on NDLI, subject-wise distribution of Indian language documents by Access restriction, subject-wise distribution of Indian language documents by Educational level, and subject-wise distribution of Indian language books by Content Types.

A total of 398793 numbers of Indian 22 recognized language documents are indexed in the National Digital Library of India from 2018 of ten subjects.

Table-1 : Subject-wise distribution of Indian language on NDLI

	Computer science (000)	Philosophy & Psychology (100)	Religion (200)	Social sciences (300)	Language (400)	Natural science (500)	Technology (600)	Arts (700)	Literature (800)	History & Geography (900)	Total
											Subject-wise distribution of Indian language on NDLI
Assamese	841	-	1	262	645	903	1	-	109	97	2859 (0.72%)
Bengali	133082	212	2355	958	556	2268	2187	285	18689	1774	162366 (40.71%)
Bodo	39	-	-	8	62	63	-	-	21	9	202 (0.05%)
Dogri	-	-	-	-	34	-	-	-	13	-	47 (0.01%)
Gujarati	4075	83	23	1650	1455	2791	971	44	2479	101	13672 (3.43%)
Hindi	6368	1115	272	16156	19271	19754	5371	3109	21577	5879	98872 (24.79%)
Kannada	1789	10	3	1146	2426	2137	1571	30	1618	158	10888 (2.73%)
Kashmiri	100	-	-	1	45	-	-	-	39	1	186 (0.47%)
Konkani	10	-	-	42	303	67	-	-	260	-	682 (0.17%)
Maithili	2	-	-	-	69	-	-	-	-	-	71 (0.02%)

Malayalam	807	1	4	534	872	921	553	1	487	78	4258 (1.07%)
Manipuri	361	-	-	7	104	11	-	4	19	7	513 (0.13%)
Marathi	1740	24	10	3559	4656	4886	614	264	3025	1081	19859 (4.98%)
Nepali	314	-	-	1	63	37	11	-	21	-	447 (0.11%)
Oriya	398	-	-	344	1048	1717	-	2	1593	945	6047 (1.52%)
Punjabi	-	-	-	-	-	-	-	-	-	-	-
Sanskrit	10649	688	135	80	4076	63	14	4	3780	19	19508 (4.89%)
Santali	1	-	-	-	16	-	-	-	16	-	33 (0.01%)
Sindhi	6	-	18	42	48	54	-	-	41	41	250 (0.06%)
Tamil	3677	59	28	2634	2122	5705	6980	102	2812	358	24477 (6.14%)
Telugu	5274	19	53	1476	1071	2403	1527	40	1261	142	13266 (3.32%)
Urdu	4907	137	479	1391	3580	3376	243	78	5299	800	20290 (5.09%)
Total	174440 (43.74%)	2348 (0.09%)	3381 (0.85%)	30291 (7.60%)	42522 (10.66%)	47156 (11.82%)	20043 (5.02%)	3963 (0.99%)	63159 (15.84%)	11490 (2.88%)	398793 (100%)

Table-1 depicts the subject-wise distribution of the documents on Indian 22 constitutional recognized language that were indexed NDLI. It shows that a maximum number of documents are indexed on Computer science subjects (000) 174440 documents, 43.74%, and the second number is Literature subject (800) 63159 documents, 15.84% and followed by Natural science and Mathematics subject (500) 47156 documents, 11.82%, Language (400) 42522 documents 10.66%. Similarly, a maximum number of documents are indexed in the National Digital Library of India (NDLI) of Bengali language (40.71%) approx162366 documents and second most is the Hindi language (24.79%) approx98872 which is clearly shown in Figure- in Figure-2. The maximum languages covered most of the subject and sequentially distributed their resources, but there is no data in the Punjabi language.
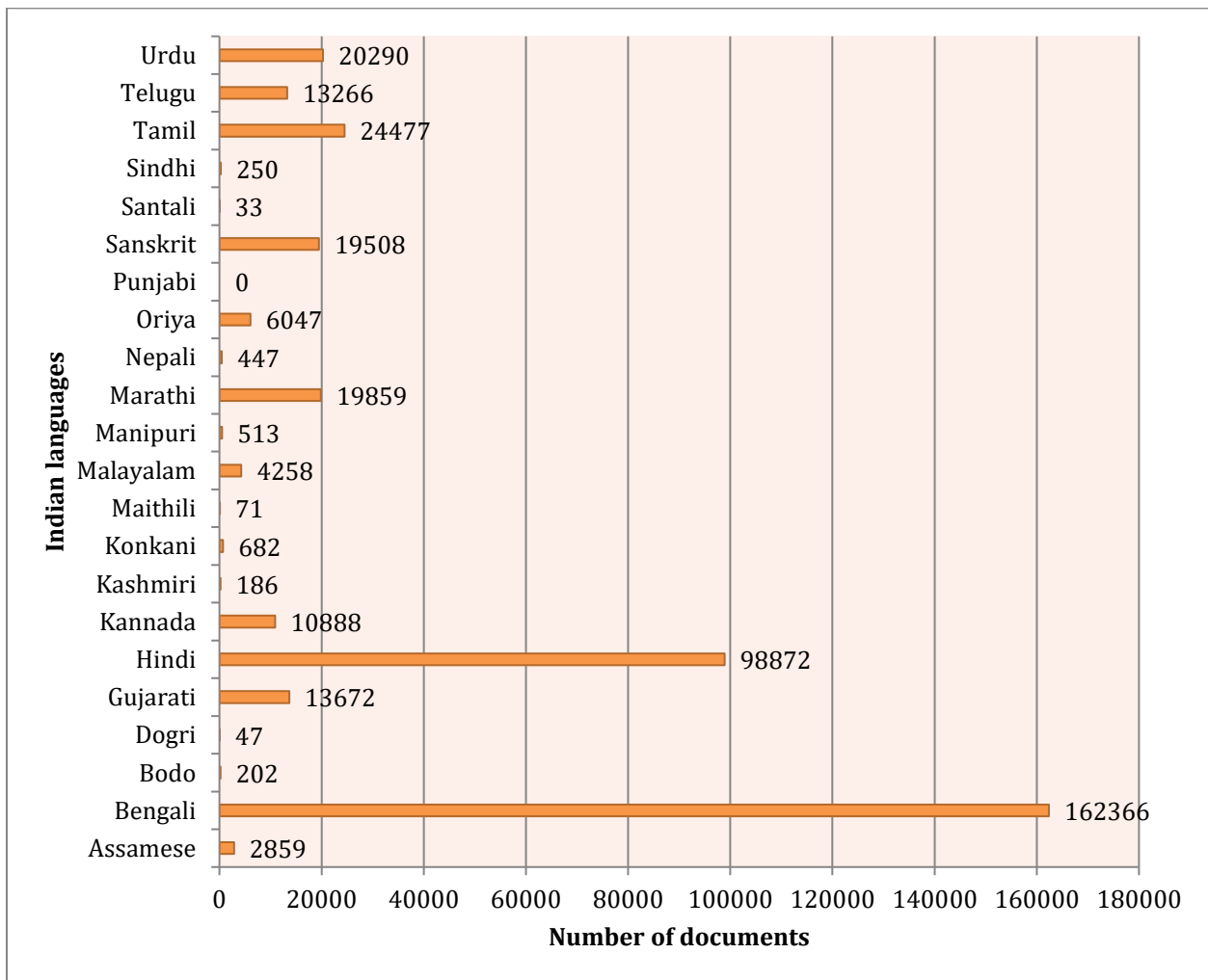
Figure-2: Number of documents in different Indian languages

Table 2: Subject-wise distribution of Indian language documents by Access restriction :

Subject-wise distribution of Indian language documents by Access restriction						
	Open	NDLI	Subscribed	Limited	Authorized	Total
Computer science, information & general works (000)	142461 (81.67%)	17336 (9.94%)	14643 (8.39%)	-	-	174440 (100%)
Philosophy & Psychology (100)	2233 (95.10%)	53 (2.26%)	-	-	62 (2.64%)	2348 (100%)
Religion (200)	1775 (52.50%)	1598 (47.26%)	4 (0.12%)	2 (0.06%)	2 (0.06%)	3381 (100%)
Social sciences (300)	29576 (97.64%)	438 (1.44%)	252 (0.83%)	-	25 (0.09%)	30291 (100%)
Language (400)	42242 (99.34%)	99 (0.23%)	1	-	180 (0.43%)	42522 (100%)
Natural science and Mathematics (500)	46828 (99.30%)	224 (0.47%)	3 (0.01)	-	101 (0.22%)	47156(100%)
Technology (600)	19364 (96.61%)	240 (1.20%)	3 (0.01%)	-	436 (2.18%)	20043 (100%)
Arts (700)	3794 (95.73%)	169 (4.27%)	-	-	-	3963 (100%)
Literature, rhetoric & criticism (800)	56047 (88.74%)	6534 (10.34%)	241 (0.38%)	3 (0.01%)	334 (0.53%)	63159 (100%)
History & Geography (900)	10542 (91.75%)	861 (7.49%)	2 (0.02%)	-	85 (0.74%)	11490 (100%)

	354862 (88.98%)	27552 (6.91)	15149 (3.80%)	5 (0.0%)	1225 (0.31%)	398793 (100%)
Total						

Table-2 indicates the subject-wise distribution of access restriction or availability of NDLI indexed Indian language documents. The above table clearly shows the maximum number of documents that are openly available for access, there is no barrier to read or download. Users can easily offline access in his/her system by downloading. However, NDLI has categories of its resources in the following access method.
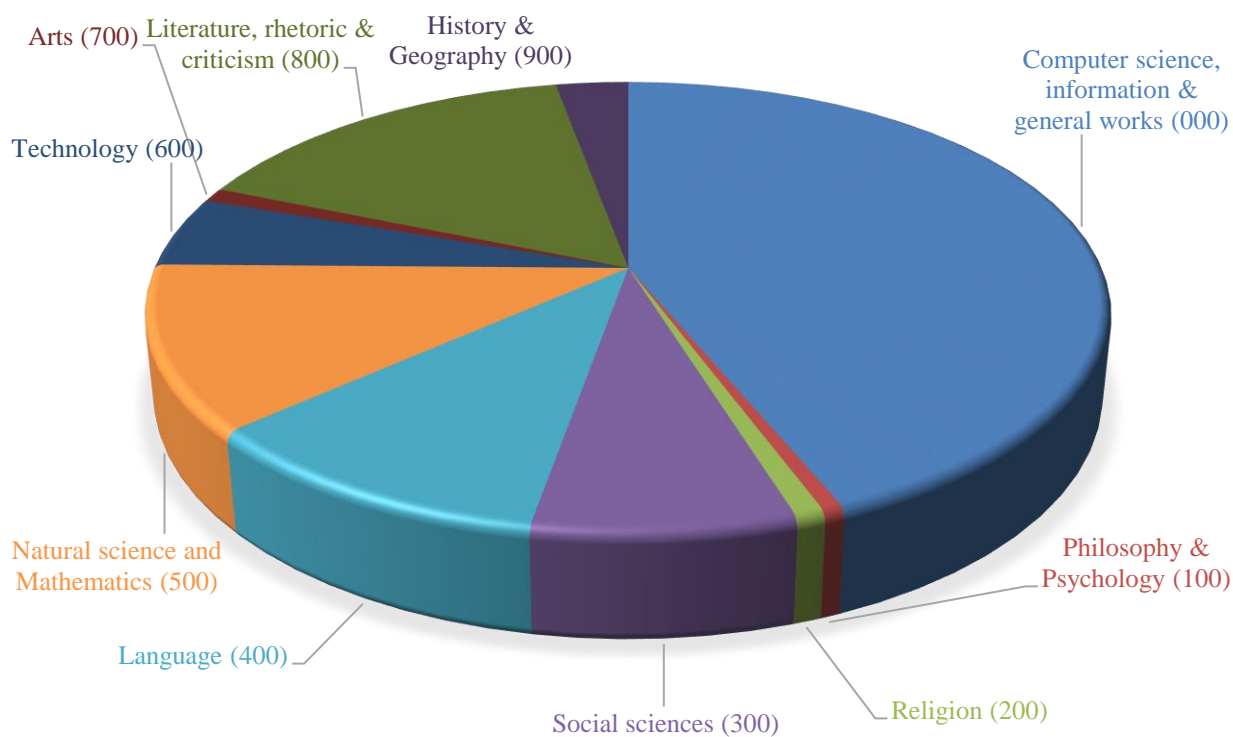
Figure-3: Number of documents in different subjects categories

- 🔓Open: full-text available to all (Example: NCERT)
- 🛡Limited: part of the text is available but full-text requires authorization by Source Authority (Example: IISER, Bhopal)
- 🔓Subscribed: full-text available for institutions that have subscribed to the Source (Example: Springer)

- 🔒 Authorized: full-text access requires authorization by Source authority and separate login to the Source (Example: IIT Jodhpur)
- 🔒 NDLI: This signifies that the document can be accessed in full-text mode by logging into NDL
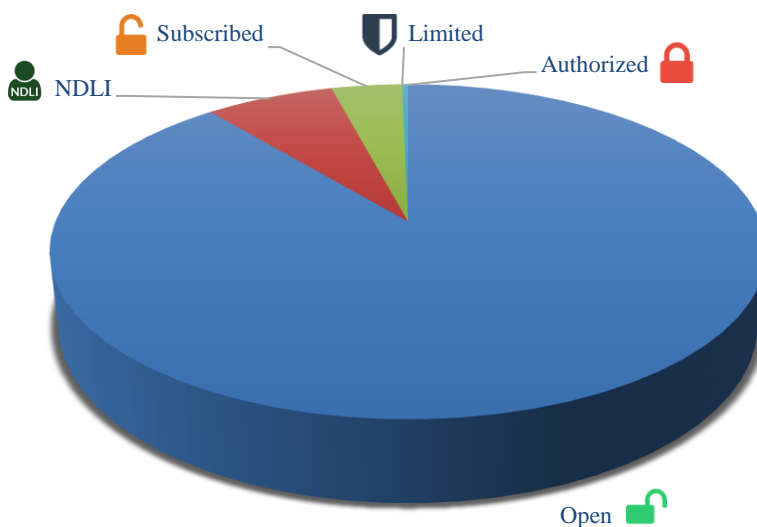

Figure-4: Number of documents based on different access restrictions

If we see the subject-wise restriction, it is clearly displayed in Figure-4 that more than eighty-five percentages (85%) of documents are no restriction for use to open for all, and 6.91 percent of documents are no restriction for NDLI users. Approximately, ninety-five percent (95%) of documents are accessible free of cost. Only five percent (5%) of documents are under limited and subscribed access.Table-3: Subject-wise distribution of Indian language documents by educational level.

Subject-wise distribution of Indian language documents by Educational level						
	Primary	**Secondary**	**H. Secondary**	**UG/PG**	**Others**	**Total**
Computer science,	56 (0.03%)	379 (0.22%)	734 (0.42%)	6260 (3.59%)	167011 (95.74%)	174440 (100%)

information & general works (000)						
Philosophy & Psychology (100)	17 (0.76%)	33 (1.40)	266 (11.32%)	804 (34.24%)	1228 (52.28%)	2348 (100%)
Religion (200)	-	-	-	92 (2.72%)	3289 (97.28%)	3381 (100%)
Social sciences (300)	2715 (8.96%)	8909 (29.41%)	2594 (8.56%)	2965 (9.79%)	13108 (43.28%)	30291 (100%)
Language (400)	7326 (17.23%)	18984 (44.64%)	6754 (15.88%)	1798 (4.24%)	7660 (18.02%)	42522 (100%)
Natural science and Mathematics (500)	6582 (13.96%)	27271 (57.83%)	7119 (15.10%)	1259 (2.66%)	4925 (10.45%)	47156 (100%)
Technology (600)	52 (0.26%)	729 (3.63%)	302 (1.50%)	5918 (29.53%)	13042 (65.08%)	20043 (100%)
Arts (700)	229 (5.78%)	1747 (44.08%)	1026 (25.89%)	155 (3.91%)	806 (20.34%)	3963 (100%)
Literature, rhetoric & criticism (800)	4545 (7.20%)	18853 (29.85%)	3207 (5.07%)	5692 (9.01%)	30862 (48.87%)	63159 (100%)
History & Geography (900)	130 (1.13%)	4697 (40.88%)	1322 (11.50%)	254 (2.21%)	5087 (44.28%)	11490 (100%)
Total	21652 (5.43%)	81602 (20.46%)	23324 (5.85%)	25197 (6.32%)	247018 (61.94%)	398793 (100%)

The distribution of subject mapping according to educational level is given in table-3. NDLI has classified all resources in their content-wise, i.e., up to class I, I to IV, V to VII, IX to X, XI to XII, UG and PG, Adult education, Career/Technical Standard, etc. For this analysis, educational levels have been classified into five groups. Up to class I and class I to IV consider as the primary standard, class V to VIII and class IX to X considered as a secondary standard, class XI to XII considered as a higher secondary standard, UG and PG considered as UG/PG education level standard and adult education and career/ technical standard considered as other educational level standards. In this subject mapping, the majority shows in others' education level (61%). The second highest in secondary education level (21%). The Higher Secondary and UG/PG both have 6% documents respectively and at the last, the primary education level encompasses 5% documents (Figure-5).
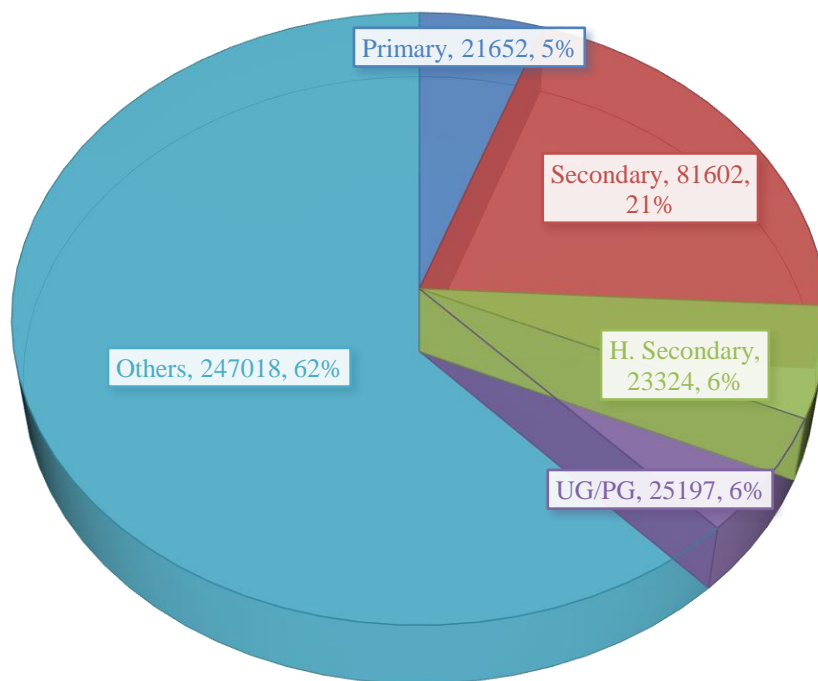

Figure-5: Number of documents in education levels

The maximum NDLI resources are under the other (adult and career/technical) educational level except for language, science, and arts subjects. More than 50% of resources of every subject are career or technical-related. In subject language (400), Science (500), and Arts (700), a large

number of resources are under the secondary level (Class V to VIII and IX to X) and higher
secondary level (Class XI to XII) given in Table-4.

Table-4: Subject-wise distribution of Indian language books by Content Types

Subject-wise distribution of Indian language books by Content Types						
	Text	Video	Audio	Image	Others	Total
Computer science, information & general works (000)	168323 (96.49%)	6064 (3.48%)	37 (0.02%)	16 (0.01%)	-	174440
Philosophy & Psychology (100)	2168 (92.33%)	170 (7.25%)	8 (0.34%)	1 (0.04%)	1 (0.04%)	2348
Religion (200)	3357 (99.29%)	24 (0.71%)	-	-	-	3381
Social sciences (300)	24914 (82.25%)	5045 (16.65%)	324 (1.07)	8 (0.03%)	-	30291
Language (400)	31786 (74.75%)	8572 (20.16%)	2049 (4.82%)	45 (0.11%)	70 (0.16%)	42522
Natural science and Mathematics (500)	24501 (51.95%)	22513 (47.75%)	117 (0.25)	20 (0.04)	5 (0.01)	47156
Technology (600)	18619 (92.90%)	1201(5.996%)	217 (1.08%)	5 (0.02%)	1 (0.004%)	20043
Arts (700)	2526 (63.73%)	1238 (31.24%)	140 (3.53%)	38 (0.95%)	21 (0.55%)	3963
Literature, rhetoric & criticism (800)	52818 (83.63%)	9835 (15.57%)	497 (0.79%)	9 (0.01%)	-	63159

History & Geography (900)	8756 (76.20%)	2546 (22.16%)	188 (1.64%)	-	-	11490
Total	337768 (84.70%)	57208 (14.34%)	3577 (0.90%)	142 (0.04%)	98 (0.02%)	398793(100%)

Table. 4 shows the content types of NDLI indexed documents. Not only textual contents are available, but you can also get materials in non-textual forms like audio, video, image, animation, simulation and presentation besides textual materials. Select the 'By Type' option on 'Browse' or the material type from the right-hand bar on 'Search' to get materials in your desired form [9]. For analysis purposes, we categories text, audio, video, image and others (animation, simulation and presentation) content types.
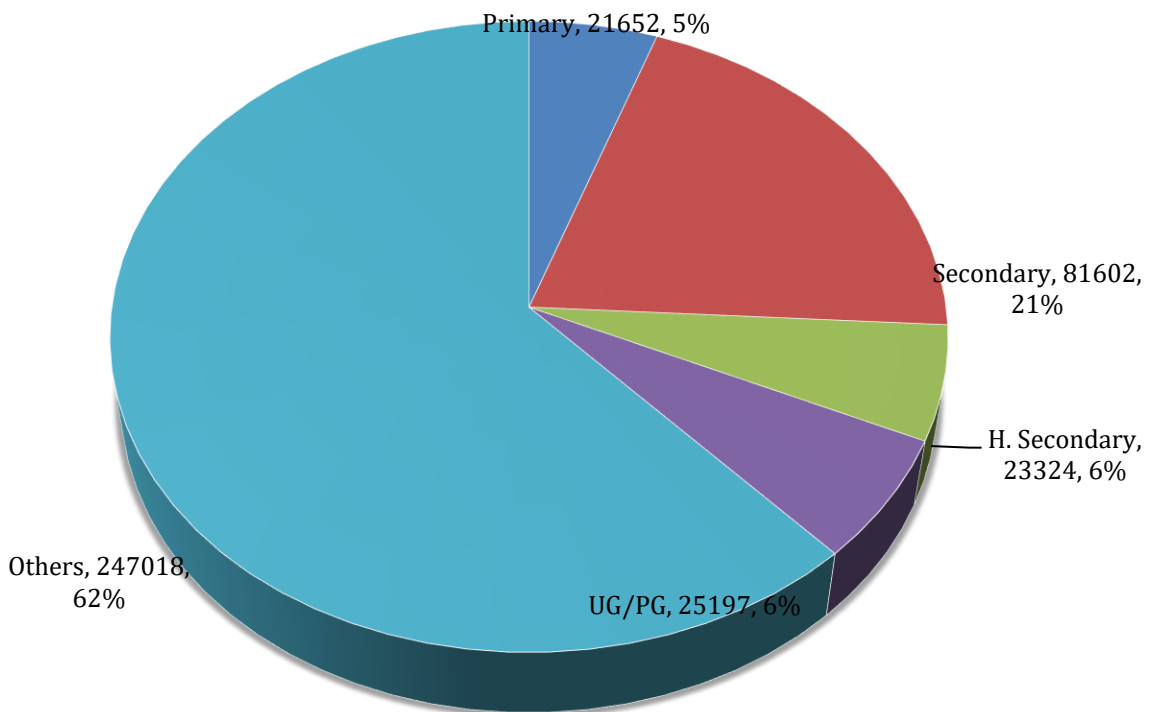

Figure-6: Number of documents in different formats

It is observed that 337768 out of the 398793 documents have a textual format, more than eighty-four percent (84.70%) followed by documents on video format 57208 documents (14.34%),

audio format 3577 (0.90%). The rest of the documents are images and other formats that are very few numbers.

If we see according to subject-wise distribution, more than eighty to ninety percent of documents are available on textual formats of each subject except Science (500) arts (700) subjects and ten to twenty percent of documents are available on video formats. The rest of one to two percent are video, image, animation, and presentation formats.

Conclusions

The NDLI project is all about knowledge dissemination, and there is no fee for accessing the digital library. It can be accessed by an individual or an institute member equally well. However, access to full-text content in many cases may require an institutional subscription or may have to be subscribed by the member personally. NDLI is a very good initiative to overcome the various barriers like communication, region, language, and money which is the main obstacle in our education systems and provides 24x7 services.

More or less Indian 22 constitutional recognized language documents are indexed in the NDLI except for the Punjabi language. So, it is recommended that more efforts should be made for initiatives for other languages and try to index the maximum numbers of all regional documents too. Because the National Digital Library (NDLI) is one kind of digital archival platform.

References

Bashir, B., Nasreen, N., & Loan, F. (2019). National Digital Library of India: An Overview. *Library Philosophy and Practice.* 2019. https://digitalcommons.unl.edu/libphilprac/2601/.

Chander, H., & Singh, K. P. (2021). Subject mapping of Punjabi language books: a bibliometric study. *Annals of Library and Information Studies* (ALIS), 67(4), 251-256. http://op.niscair.res.in/index.php/ALIS/article/view/34755.

Languages of India (2021, April 04) In *Wikipedia.* https://en.wikipedia.org/wiki/Languages_of_India

Mangurkar, M. R., & Chaudhari, Y. (2018). Content Analysis of National Digital *Library.*
Library Philosophy and Practice. 2018. https://digitalcommons.unl.edu/libphilprac/1883/.

Mehler, A., & Waltinger, U. (2009). Enhancing document modeling by means of open topic
models: Crossing the frontier of classification schemes in digital libraries by example of the
DDC. *Library Hi Tech*, 27(4), 520-539.

Moid, A., Raza, M. & Jahan, K. (2020). *A study of grey literature in National Digital Library
of India: analyses and trends. Collection and Curation*. ahead-of-print. 10.1108/CC-12-
2019-0041.

Nasreen, N., Bashir, B., & Loan, F. (2019). World Digital Library: An Analysis of
Collection.   *Library   Philosophy   and   Practice*.   2019.
https://digitalcommons.unl.edu/libphilprac/2606/.

National   Digital   Library   of   India   (2021,   April   04)   In   *Wikipedia*.
https://en.wikipedia.org/wiki/National_Digital_Library_of_India

National Digital Library of India. *FAQ (Website)*. https://ndl.iitkgp.ac.in/faq.

O'Regan, P., & O'Donnell, D. (2000). Mapping intellectual resources: insights from critical
modernism. *Journal of European Industrial Training*, 24(2/3/4), 118-127.

Singh, R., & Kumar, S. (2019). Subject Mapping of Information Literacy Competency
Levels of Social Science Researchers in National Capital Region, India. *Library Herald*,
57(2), 163-182.