Project Completion Report

SCIENTOMETRIC AND ONTOLOGICAL MAPPING OF SOLID WASTE MANAGEMENT

Implemented by

PI

DR. SHWETMALA KASHYAP

RAMAIAH PUBLIC POLICY CENTER

Co-PI

DR HN CHANAKYA, CST, INDIAN INSTITUTE OF SCIENCE

DR ARKALGUD RAMAPRASAD, RAMAIAH PUBLIC POLICY CENTER

DR CHETAN SINGAI, RAMAIAH PUBLIC POLICY CENTER

DST PROJECT NO. DST/NSTMIS/05/06/2019-20

Study Sponsored by

National Science and Technology Management Information System (NSTMIS) Department of Science & Technology Government of India

(February 2022)

NSTMIS Division
Department of Science &Technology
Ministry of Science &Technology
Technology Bhawan, New Mehrauli Road, NewDelhi-110016
Phone:91-011-26567373
Website: www.nstmis-dst.org/

About NSTMIS:
The National Science and Technology Management Information System (NSTMIS), a division of the Department of Science and Technology (DST) has been entrusted with the task of building the information based on a continuous basis on resources devoted to scientific and technological activities for policy planning in the country.

Citation:
The report may be cited as DST (2022): **Report on Scientometric and Ontological Mapping of Solid Waste Management**; **Dr. Shwetmala Kashyap**, **Ramaiah Public Policy Center, Bengaluru**, **Karnataka.**
New Delhi, Govt. of India

Disclaimer:
Every care has been taken to provide authenticated information. However, the onus of the authenticity of data rests with the PI of the project.

RPPC 2022
Scientometric and Ontological Mapping of Solid Waste Management
Sanctioned During 2019-20 to 2021-22 Period
Bengaluru: Ramaiah Public Policy Center.
[Project Report No. DST/NSTMIS/05/06/2019-20]

For more information
Ramaiah Public Policy Center
Ramaiah Medical College Building,
Upper basement floor,
M S Ramaiah Nagar,
MSRIT Post, Bengaluru, Karnataka 560054
India.
Phone: +91 80 2360 1742

Contents

ACKNOWLEDGMENTS

The project team is grateful to the National Science and Technology Management Information System (NSTMIS) Division of the Department of Science and Technology (DST), Government of India for providing an opportunity to undertake this study on 'Scientometric and Ontological mapping of Solid Waste Management'.

We are thankful to Dr. Thant Syn, Florida Gulf University, USA for helping us to use the Convolutional Neural Networks (CNN), an Artificial Intelligence tool for analysis of the corpus of research on solid waste management.

That RPPC was able to successfully complete this project with whole-hearted support and cooperation extended by the project investigator, heads of the institution, and project coordinators. The project team duly acknowledges their support.

Finally, a word of thanks also goes to the project team members Mr. Gowrish R and Ms. PT Lekshmi for their effort and hard work and to other RPPC administrative professionals who contributed towards the timely completion of the project.

Shwetmala Kashyap
(Project Investigator)

Bengaluru
February 2022

i. EXECUTIVE SUMMARY

The study has analyzed all the peer-reviewed literature on Solid Waste Management (SWM) during the period 1960 to 2020. This report includes the process of data collection, curation, and analysis of the data. Recommendations and suggestions that emerged from the study are given at the end of the report. Simultaneously, a full-text analysis of 1000 articles was carried out with the text mining software called QDA Miner lite.

All the downloaded peer-reviewed literature indexed in Scopus, Web of Science, and PubMed were subjected to data collation and curation. The Zotero corpus was updated with all the available information and has been manually cross-checked for errors and corrected. This step is important to make sure all the information regarding each article is available and valid. Further, these articles were scrutinized for cleaning and integrating the corpus. The articles in a non-English language, with no abstract, with no author name, and duplicates were identified and removed manually to clean the data sets. The detailed method used for cleaning the data sets is explained in the methodology chapter. After completing the task of data curation, 12519 articles were retained for further data analysis.

A sample of 1000 articles was used for comparing the result of full paper analysis vs title, abstract, and keyword analysis using QDA miner lite software. Then using Convolutional Neural Networks (CNN) method with an accuracy of 84% was used for all the 12519 articles text data analysis.

Findings at a glance

Full paper analysis and comparison

- Full paper analysis of 1000 articles using QDA miner lite software shows that among all the dimensions outcome, type, and source of solid waste are most emphasized. Among the outcomes, the focus has been on environmental, followed by heath, and social outcomes. Among the types of solid waste, the focus has been on non-biodegradable waste, followed by biodegradable waste. Among the source of solid waste generation, the focus has been on non-residential industrial sources, non-residential municipal services, and non-residential commercial sources.
- Comparing the 1000 papers descriptive metadata coding done using CNN and QDA miner lite software show similar results. Among all the elements results are more emphasized on the environmental outcome, processing functional element, non-biodegradable solid waste, legislative policy instrument, and municipal/local government stakeholder. The frequency values were different for CNN and QDA miner lite software results, but the pattern of emphasis is approximately the same.

CNN analysis of the total literature

- There is a significant increase in the number of publications for the period of 1961-2020.
- The CNN analysis result shows that among all the dimensions, function, outcome, and solid waste categories are emphasized most, whereas among all the elements the environmental outcome is emphasized most.

Study at a Glance	
Period of Study	January 2020-2022
Funding agency	NSTMIS, DST, Government of India
Data Collection	
No of Databases	3
Number of publications downloaded	19864
Scopus	14389
Web of Science	4268
PubMed	1207
Time period	1960-2020
Data Curation	
Data Management Software	Zotero
Scopus Articles retained	11102
Web of Science articles retained	1167
PubMed articles retained	250
No of articles for analysis	12519
Data Analysis and Visualization	
Software used for metadata analysis (binary coding)	CNN
Software used for full-text coding of 1000 articles	QDA lite miner
Visualization software	MS Excel and VOS Viewer

1.0 INTRODUCTION

Solid Waste Management (SWM) is a critical issue globally, and nationally, in India. The SWM system is complex as it entails many pathways to manage the waste. It involves many elements and sub-elements as shown in Figure 1.

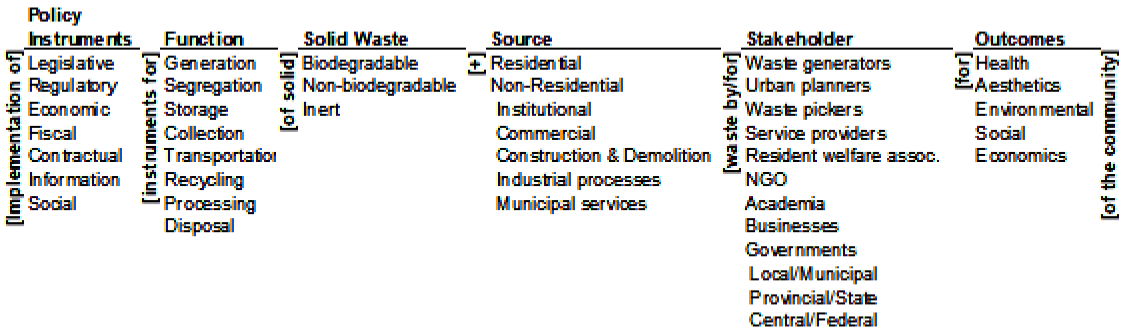

Figure 1: Ontological framework of urban solid waste management (Kashyap et al., 2019)

A typical SWM system encompasses the following functional elements: generation, segregation, storage, collection, transportation, recycling, processing, and disposal of waste. The success of SWM will depend not only on these functional elements, but also on the waste management policies, waste composition, selection of technology, and involvement of stakeholders.

In the past few decades, the capabilities of SWM technologies have evolved and their complexity has increased. Their effective development and implementation in the future will be critical for the sustainability of the villages, cities, and the country. There is a wide variation in the knowledge of, decisions about, implementation, and assessment of SWM globally. Knowing and understanding the past trajectory of SWM practices globally will help to direct their future in India.

This study assesses the practice and performance of SWM globally throughout 1960-2020 using scientometric and ontological analysis. The scientometric mapping highlights the nucleation of key SWM technologies and their evolution globally using the citation structure of research articles on the subject. The ontological mapping highlights the most frequently, infrequently, and unused technological pathways for SWM, and the evolution of the pathways longitudinally. Both results are presented graphically and analytically. Such a systematic analysis of the SWM research literature will help to discover innovative and inventive solutions for the future.

1.1 Objective

1. Describe the performance and practice of SWM and the role of research and academia in waste management
    - (a) Construction and validation of the ontology of SWM
    - (b) Mapping of the performance of SWM
2. Explain the reasons for not achieving complete SWM
    - (a) Analysis of gaps between practices and performance of SWM
    - (b) Finding the antecedents and consequences of the gaps between aspirations and performance of the institutions

3. Develop a roadmap for the changes required in research, policy, and practice
    (a) Recommending pathways for bridging the gaps to achieve complete SWM

2.0 REVIEW OF LITERATURE

Solid waste is a heterogeneous mixture of solid materials which does not have any further use to society (UNFCCC, 2012). It includes all the solid wastes generated from residential and non-residential sources located in a municipal or notified area in an urban locality (MSW Rule, 2000; SWM Rule, 2016).

With rapid urbanization, the quantity of solid waste is also increasing, necessitating systematic and systemic solid waste management. Indian cities' average annual per capita increase in waste generation rate is 1-1.33% (Shekdar, 1999; Pappu et al., 2007; Sharholy et al., 2008). It requires a detailed study of past and present practices and performances of SWM.

Scientometrics is a quantitative approach to study the history of science. It uses bibliometric measurements for evaluating a scientific domain (in terms of progress, development, and social relevance) and for finding the impact of the applications of scientific knowledge. Bibliometrics quantitatively analyses the bibliographic material to provide a general overview (Broadus, 1987).

The ontological analysis is used extensively in computer science, medicine, and philosophy. It is used to develop domain concepts and categories and the relationships between them. An ontology is an explicit specification of a conceptualization (Gruber, 1995). It provides a holistic vision of the journal impact domain and is used to systematically map, analyze, and synthesiz e a given corpus of knowledge. Both scientometric and ontological approaches are used to understand the complete system based on peer-reviewed literature and such approaches are gaining attention. However, deep scientometrics and ontological study are not yet to be conducted for SWM.

3.0 METHODOLOGY

This session explains systematic strategies undertaken during the study. The target population includes all the SWM publications which are globally available over the past six decades. The data collection method is focused on peer-reviewed literature on SWM. It includes articles, reviews, and conference proceedings published from 1960 to 2020 but excludes earlier ones. The study searches for the bibliographic materials through Scopus, Web of Science, and PubMed. These databases curate and index peer-reviewed scientific literature with the aim of archiving all the leading scientific material that is published globally. The data were analyzed using CNN and visualized and interpreted using VOS viewer maps and MS excel graphs.

For ontological analysis, the title, keywords, and abstract are used for coding. The articles were mapped to the ontological framework using binary coding of '1' for present and '0' for absent. The ontological map will show the frequency of occurrence of each ontological element in the corpus of the research paper.

Scientometric analysis was based on a wide range of indicators including the number of documents published, the number of citations, the citations per paper, and the citation thresholds.

A total of 19864 articles were downloaded from all three databases. These articles have been further filtered based on certain criteria to arrive at 12519 articles for further analysis. The PRISMA flow diagram in section 3.9 shows the filtration of data through different phases. Simultaneously, a full-text analysis of 1000 articles were carried out with a text mining software called QDA Miner lite.

3.1 Data Search Strategies

Multiple trials with various relevant search terms were carried out in Scopus to arrive at the most appropriate search term (Table 1). The maximum document results were found by search term TITLE-ABS-KEY "Solid Waste Management". The search results showed a 10% overlapping of Urban Solid Waste Management and Municipal Solid Waste Management with Solid Waste Management. Urban Solid Waste Management and Municipal Solid Waste Management are sub-categories of Solid Waste Management. Thus, an all-inclusive search term "Solid Waste Management" covers results for both Urban Solid Waste Management and Municipal Solid Waste Management. The same search term was employed for Web of Science and PubMed to search for documents.

Date	**Time**	**Search**	**Year**	**Document Results**
22nd June 2020	10.30AM	(TITLE-ABS-KEY ("Urban Solid Waste Management"))	1960-2020	123
22nd June 2020	10.35AM	(TITLE-ABS-KEY ("Municipal Solid Waste Management"))	1960-2020	1479
22nd June 2020	10.40AM	(TITLE-ABS-KEY ("Solid Waste Management"))	1960-2020	14161
22nd June 2020	11.40AM	(TITLE-ABS-KEY ("Solid Waste Management"))	1989-2019	9786
29th June 2020		(TITLE-ABS-KEY ("Solid Waste Management")) AND NOT (TITLE-ABS-KEY ("Urban Solid Waste Management")) AND NOT (TITLE-ABS-KEY ("Municipal Solid Waste Management"))	1960-2020	12579
29th June 2020		(TITLE-ABS-KEY ("Solid Waste Management")) AND NOT (TITLE-ABS-KEY ("Urban Solid Waste Management"))	1960-2020	14054

29th June 2020		(TITLE-ABS-KEY ("Solid Waste Management")) AND NOT (TITLE-ABS-KEY ("Municipal Solid Waste Management"))		12692
30th June 2020		(TITLE-ABS-KEY ("Municipal Solid Waste Management")) AND NOT (TITLE-ABS-KEY ("Urban Solid Waste Management"))		1475
30th June 2020		(TITLE-ABS-KEY ("Urban Solid Waste Management")) AND (TITLE-ABS-KEY ("Municipal Solid Waste Management"))		10

Table 1: Scopus search results

Details about the selected search

There is a significant dip in the number of documents around the years 1989 and 2000 and after that an increase in the succeeding years (Figure 2).

Year 1966 to 2020

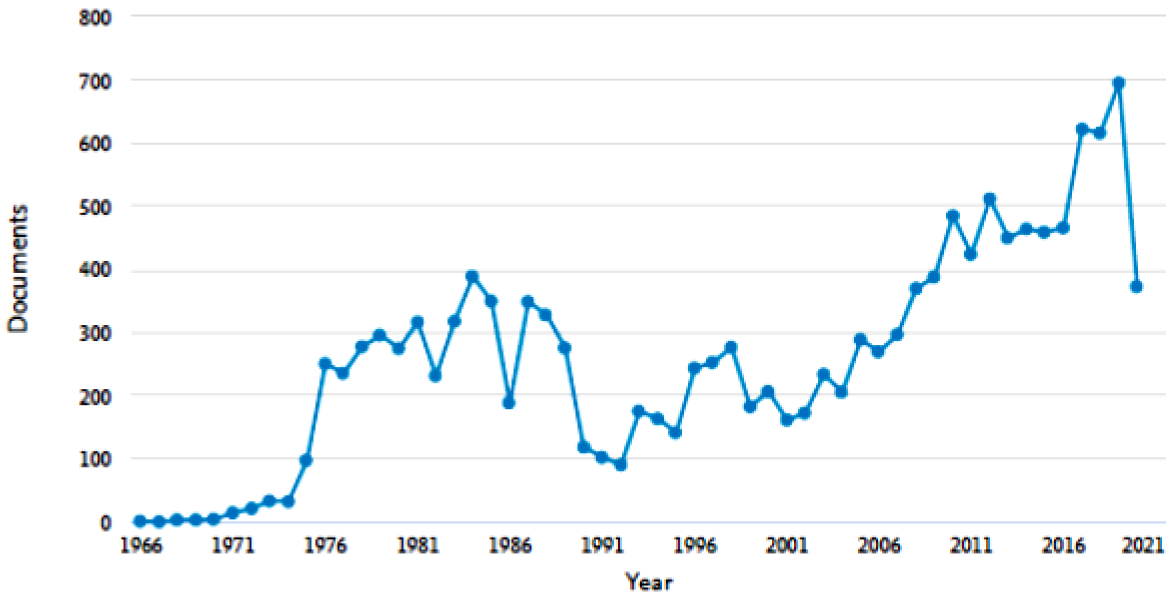

Figure 2: Year wise documents

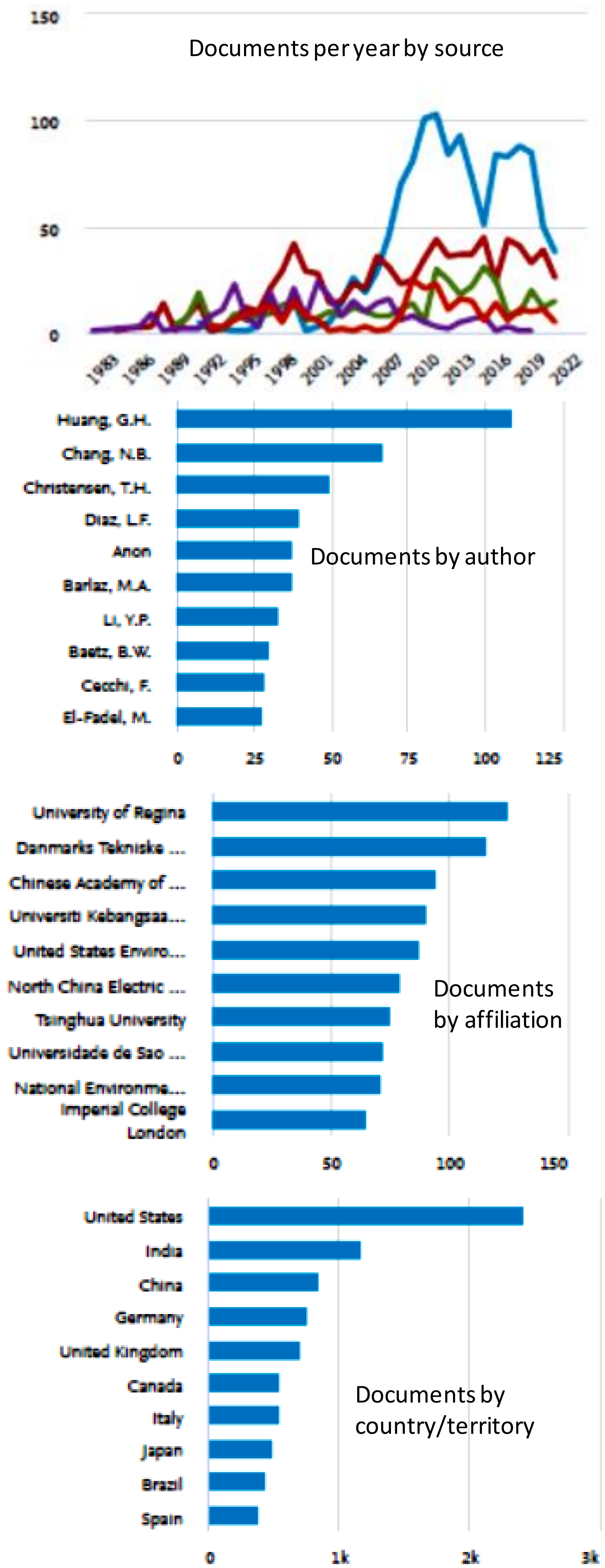

Figure 3: Documents by source, author, affiliation, and country

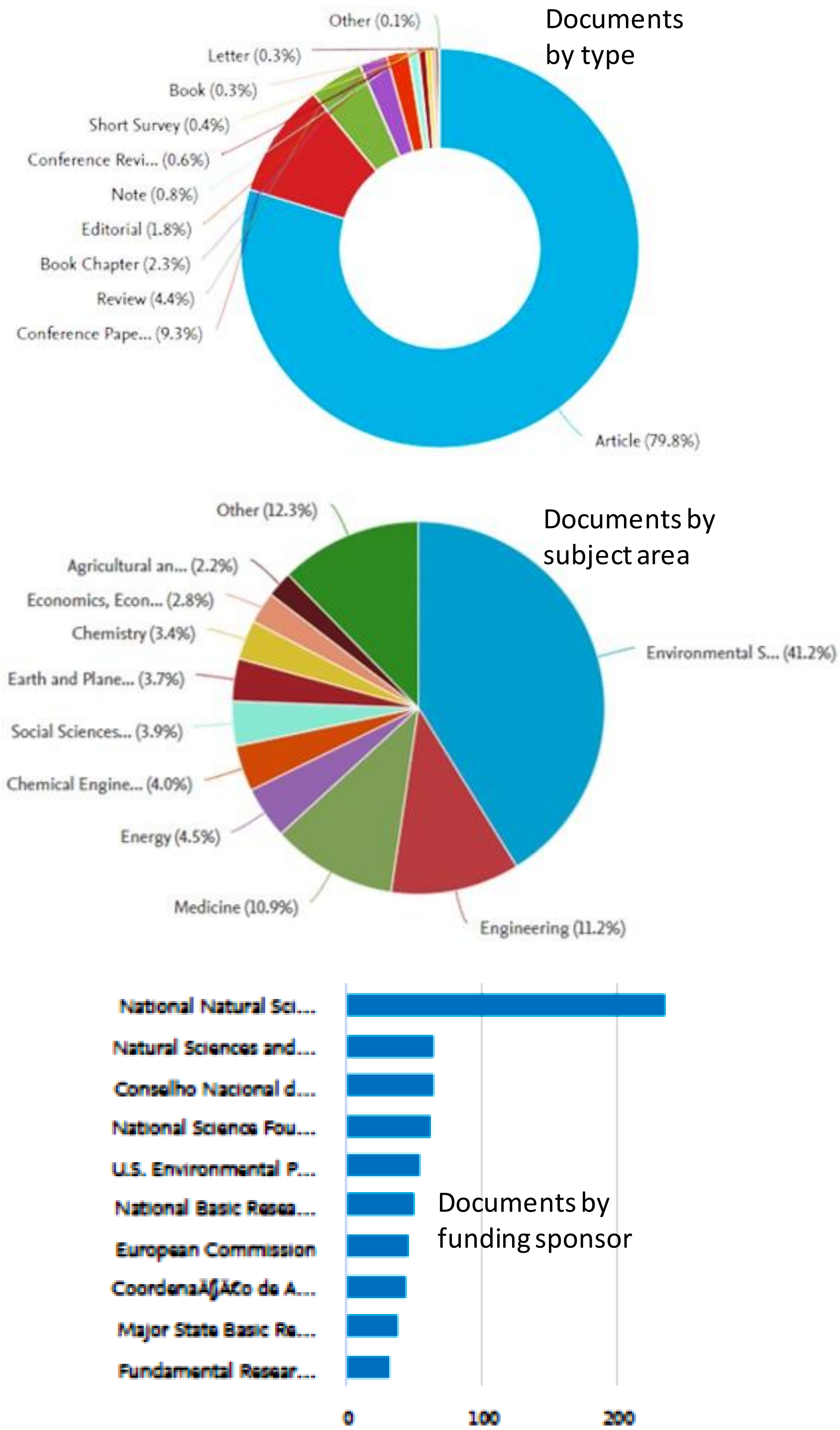

Figure 4: Documents by type, subject area, and funding sponsor

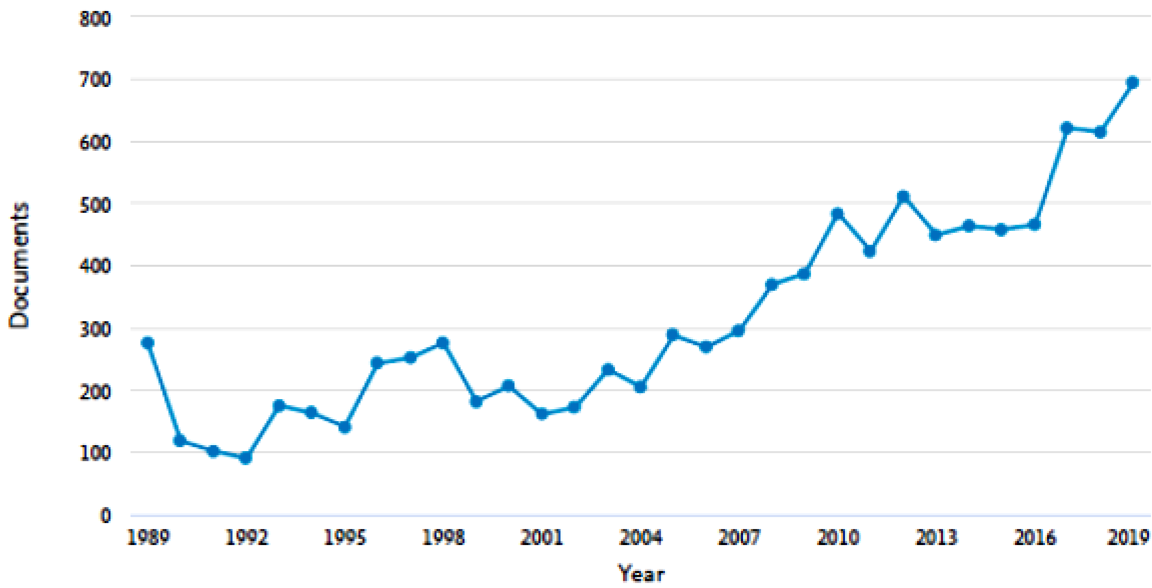

Figure 5: Documents by year

Year 1989 to 2019

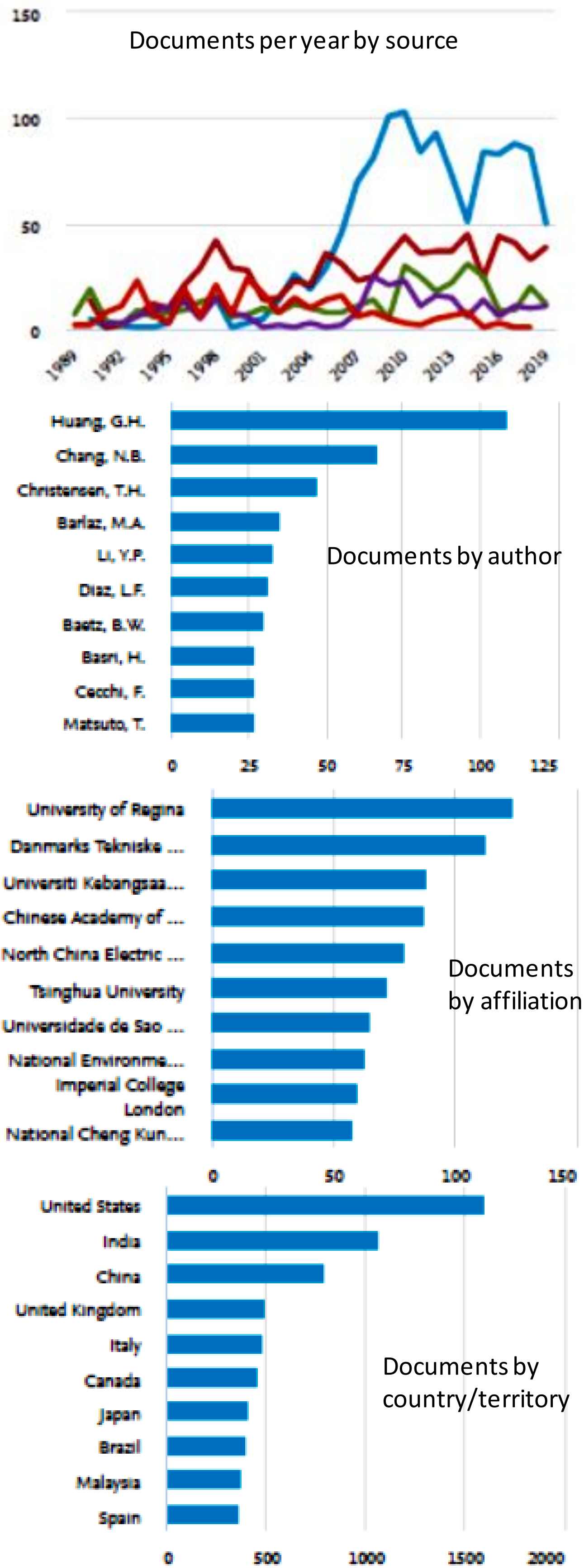

Figure 6: Documents by source, author, affiliation, and country

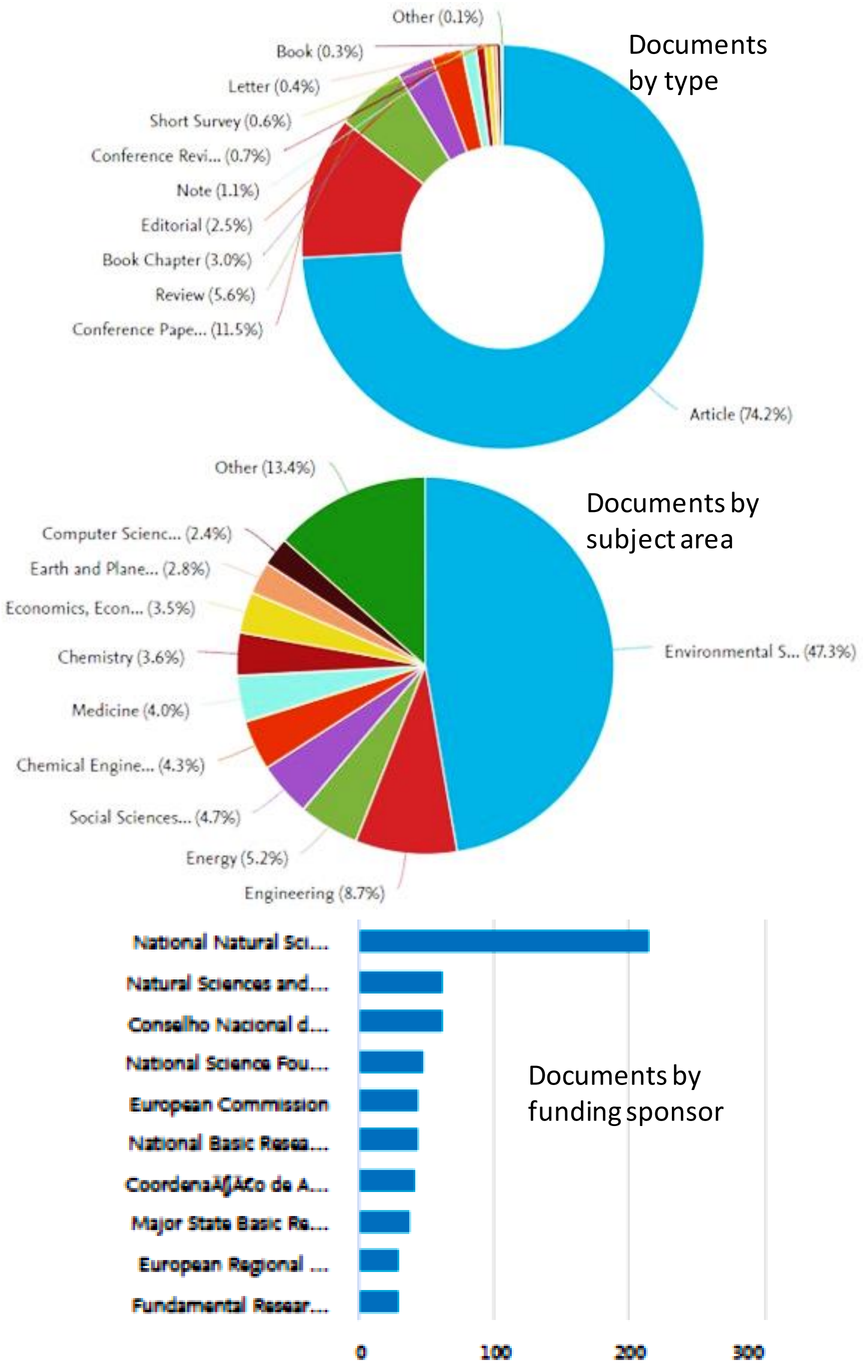

Figure 7: Documents by type, subject area, and funding sponsor

3.2 Download Strategy

The TITLE-ABS-KEY "Solid Waste Management" search term was used for downloading the documents from Web of Science, PubMed, and Scopus. Various trials were employed to figure out the optimal method for downloading documents from each database. The documents were downloaded using a plugin to the data management software called Zotero. Following are the details about the download from all three databases:

Databases	Documents Downloaded	Date of Download
Web of Science	4268	28th Sept 2020
PubMed	1207	21st Sept 2020
Scopus	14389	21st and 22nd Sept 2020
Total	19864	

Table 2: Total literature downloaded from databases

3.2.1 Web of Science: Total number of documents downloaded from Web of Science was 4268.

Year	Documents
1988	1
1989	1
1990	12
1991	2
1992	8
1993	15
1994	17
1995	15
1996	44
1997	46
1998	50
1999	38
2000	67
2001	42
2002	45
2003	58
2004	46
2005	68
2006	72
2007	94
2008	122
2009	170
2010	181
2011	183
2012	297
2013	184
2014	195
2015	303
2016	313
2017	423
2018	402
2019	416
2020	288
No year	50
Total	**4268**

Table 3: Year-wise number of downloaded documents from Web of Science

3.2.2 PubMed: Total number of downloaded documents from PubMed was 1207. Following are the details about the documents.

Year	Documents
1966	1
1968	1
1971	3
1972	1
1973	1
1974	1
1977	1
1978	2
1981	1
1987	1
1989	1
1991	7
1992	2
1993	2
1994	4
1995	1
1996	1
1997	2
1998	5
1999	3
2000	2
2001	9
2002	10
2003	18
2004	8
2005	28
2006	35
2007	40
2008	38
2009	66
2010	49
2011	42
2012	53
2013	53
2014	75
2015	70
2016	79
2017	112
2018	131
2019	110
2020	135
2021	3
Total	**1207**

Table 4: Year-wise number of downloaded documents from PubMed

3.2.3 Scopus: Total number of downloaded documents from Scopus was 14389. Following are the details about the documents:

Year	Documents
1966-1974	111
1975	97
1976	250
1977	235
1978	277
1979	295
1980	274
1981	316
1982	231
1983	317
1984	388
1985	349
1986	188
1987	348
1988	327
1989	275
1990	118
1991-1992	192
1993	175
1994	163
1995	141
1996	243
1997	252
1998	276
1999	182
2000	206
2001	161
2002	172
2003	234
2004	205
2005	287
2006	268
2007	295
2008	369
2009	388
2010	484
2011	423
2012	511
2013	449
2014	465
2015	459
2016	466
2017	623
2018	624
2019	694
2020	586
Total	**14389**

Table 5: Year-wise number of downloaded documents from Scopus

3.3 Data Curation

This section explains the systematic data collation methods followed to convert downloaded raw data into well-organized data. All the peer-reviewed literature (articles, conference proceedings, reviews, and book chapters) indexed in Scopus, Web of Science, and PubMed were downloaded. After downloading the corpus of the articles in Zotero software, it was checked for the available information of title, author, abstract, publication, keywords, date, and other relevant information. The appropriate corrections and modifications were made.

These articles were further scrutinized for cleaning and integrating the corpus. The articles with non-English language, no abstract, no author name, and duplicates were identified and removed manually. The data collation of Web of Science, PubMed, and Scopus are explained below:

3.3.1 Web of Science: Total 4268 articles were downloaded from Web of Science (WOS). After excluding the 8 articles which are book reviews, 254 articles without abstract, 9 articles without author name, and 6 duplicates, only 3991 articles were eligible from the Web of Science database for further analysis (Figure 8).

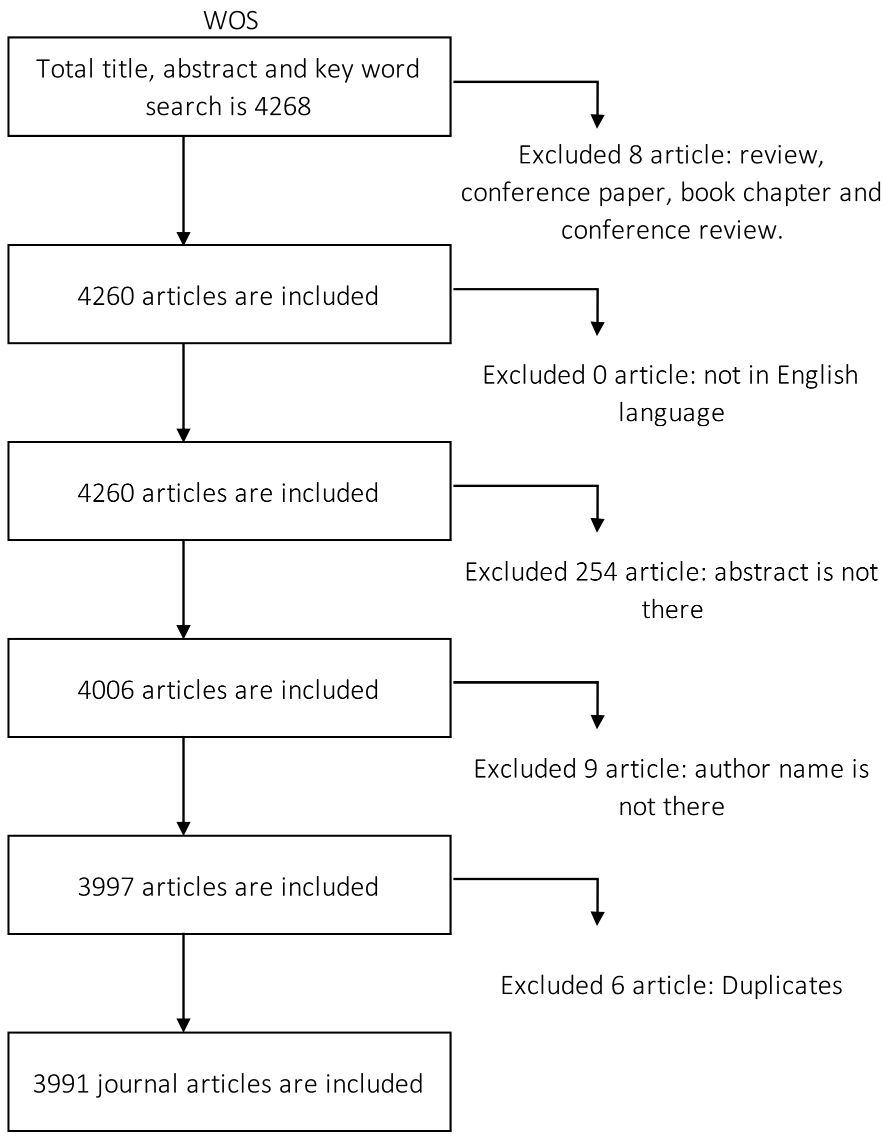

Figure 8: PRISMA diagram of data curation process of Web of Science literature

3.3.2 PubMed: Total 1207 articles were downloaded from PubMed. After excluding the 49 articles without abstract, and 1 article without author name, only 1157 articles were eligible from the PubMed database for further analysis (Figure 9).

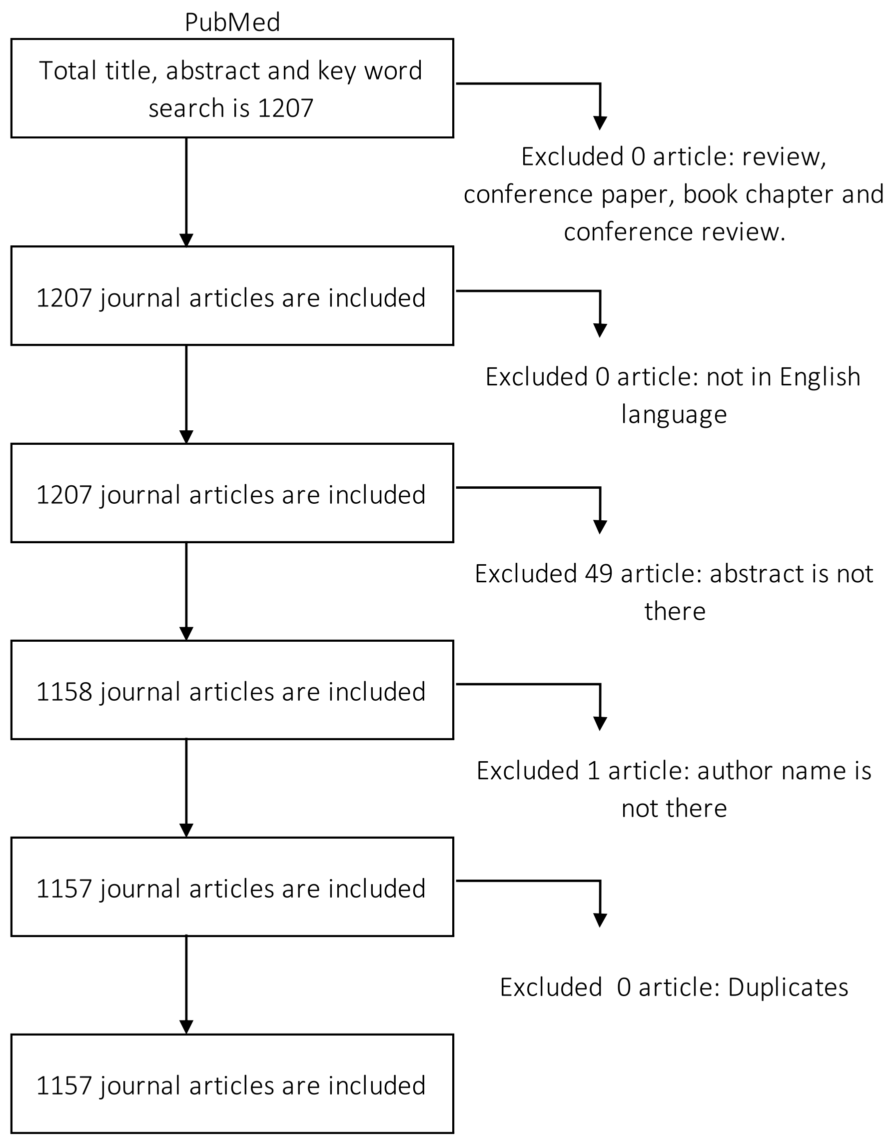

Figure 9: PRISMA diagram of data curation process of PubMed literature

3.3.3 Scopus: Total 14389 articles were downloaded from Scopus. After excluding the 8 non-English articles, 2919 articles without abstract, 248 articles without author name, and 9 duplicates, only 11205 articles were eligible from the Scopus database for further analysis.

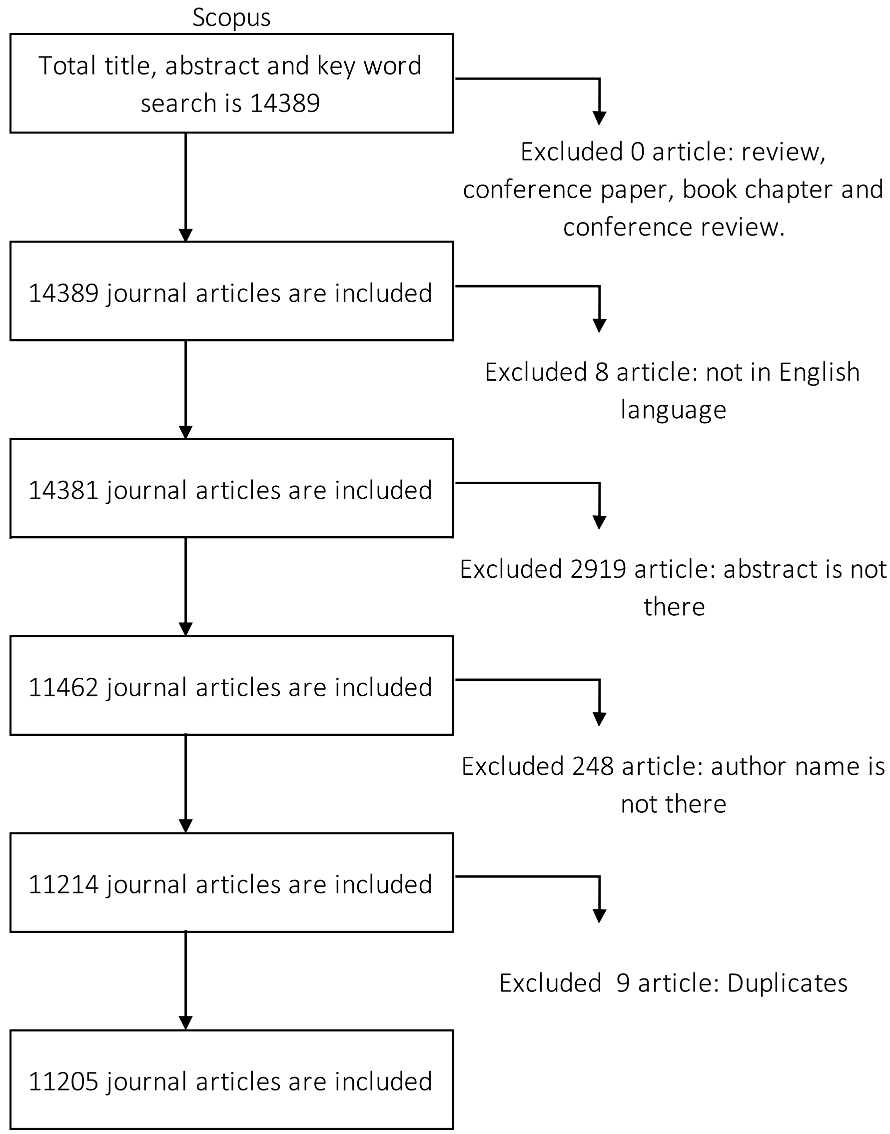

Figure 10: PRISMA diagram of data curation process of Scopus literature

3.4 Data Integration

Total 16353 articles were integrated from Web of Science (3991), PubMed (1157), and Scopus (11205). Duplicates were checked across these three databases based on title, author's name, and abstract. Scopus articles were retained, and Web of Science and PubMed were excluded if there is any overlap.

A total of 3834 articles were identified as duplicates across the databases and 12519 articles were retained for further analysis. For the papers published in both conference and the journal, the journal article was retained. The papers with the same title, abstract, and authors published in both conference and as a book chapter, the Scopus document was preferred to retain, if the duplicate is in Web of Science and PubMed the book chapter was retained. The final list of documents had 12519 literature from Scopus, Web of Science, and PubMed.

Databases	Documents Used for Analysis
Scopus	11102
Web of Science	1167
PubMed	250
Total	12519

Table 6: Total number of literature from Scopus, Web of Science, and PubMed

The downloaded literature was integrated, cleaned, and duplicates were removed as shown in the PRISMA diagram (Figure 11).

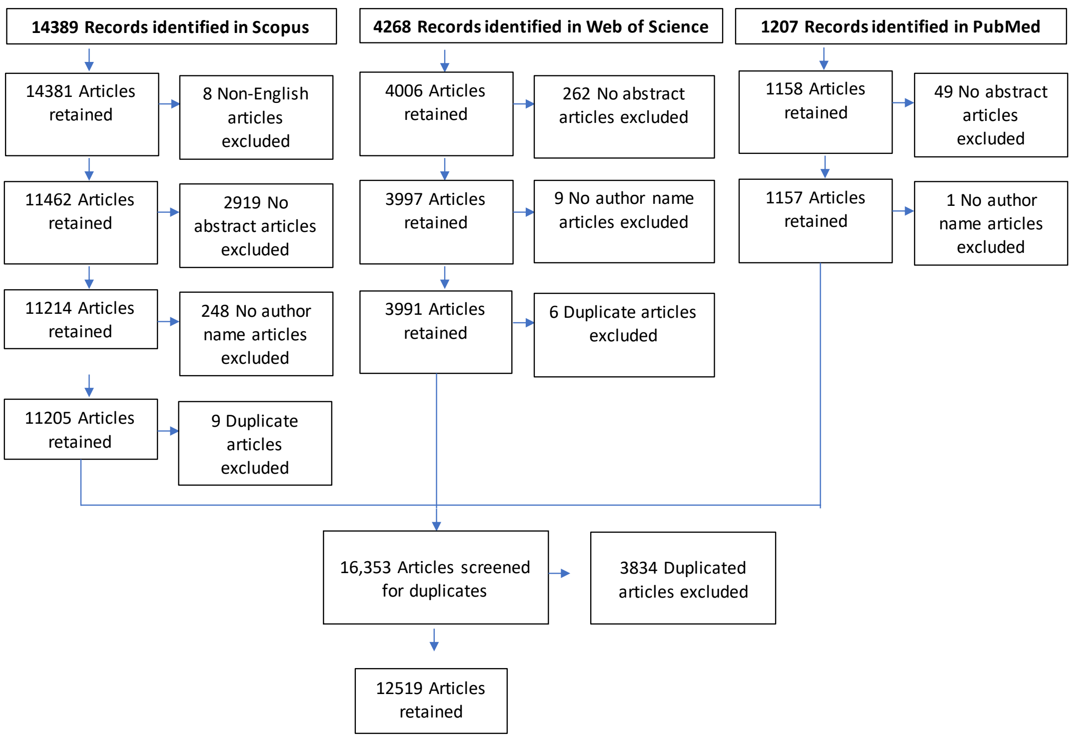

Figure 11: Final PRISMA diagram after integrating the databases

3.5 Full Article analysis.

3.5.1 The selection of articles with full text

The articles were looked at in the corpus of 12519 articles for the availability of full text in PDF format. Total 1000 articles were downloaded and attached to the Zotero files, and the same files were kept in the local computer to use for further analysis.

3.5.2 Analysis of full-text of the articles

These 1000 articles were downloaded in PDF format and coded on the SWM ontological elements using QDA miner coding software. This was a text coding software where it detects the text and synonyms used in the PDF file based on the list of ontological dimensions, elements, and their synonyms. The results of the coding are represented using the monads map.

3.6 Data Trial Sets for CNN

This section explains the methodology followed to prepare the data sets required to train the CNN software. A sample set of manually coded articles was required to train the CNN so that it can code remaining all the articles.

3.6.1 Creating the training set

The first training set of 100 articles was prepared for CNN analysis. Again 4 sets of 50 articles were prepared to increase the accuracy of CNN analysis. These samples were selected randomly from 12519 articles. The comma-separated values (CSV) file of the selected samples was converted and saved as an Excel sheet workbook. The set of 100 documents and another 4 sets of 50 documents were manually coded on the ontological elements.

3.6.2 Manual coding of training sets for CNN

The training sets saved in the Excel sheet workbook were coded manually to increase the accuracy of CNN coding. The first trial set (P1) was containing 100 documents and consecutive sets (P2, P3, P4, and P5) were containing 50 each. The coding sheet contains the title, abstract, and keywords of all the documents against the elements of Ontology. Binary coding was carried out separately by 2 individuals and with the help of a consensus coding sheet, the differences were discussed and finalized. When the set 100 of documents (P1) was used in CNN the accuracy rate was low (around 81%).

To increase the accuracy rate, a new set of 50 more documents (P2) were selected randomly from the total literature after removing 100 documents (P1) and manually coded. This new set P2 was coded by CNN based on our P1 set that we feed to CNN. The coding results of the P2 set came out with an accuracy rate of 82%. To further increase the accuracy of coding we carried out the reconcile coding of P2 set CNN coding with manual coding of P2. The manual (M) and CNN coding are compared for learning purposes and generated coding sheet based on reconciled coding as shown below:

Manual	CNN	Final	Description
1	1	1	Agree with M and CNN
1	1	0	Disagree with M and CNN
1	0	1	Agree with M disagree with CNN
1	0	0	Agree with CNN disagree with M
0	1	1	Agree with CNN disagree with M
0	1	0	Agree with M disagree with CNN
0	0	1	Disagree with M and CNN
0	0	0	Agree with M and CNN

Table 7: Description of CNN and manual coding (M)

Again 50 documents of P1 were coded manually and were fed to the CNN P1 (100) + P2 (50) + P3 (50). The same procedure was repeated as the accuracy of the P3 trial set came up to 86.5%. Additional 50 documents were coded manually, and all trial sets were fed to the CNN P1 (100) + P2 (50) + P3 (50) +P4 (50). The coding results of the P4 trial set came out with an accuracy rate of 85.3%. To increase the accuracy, we carried out the reconciliation coding of P4 trial set CNN coding with manual coding of P4. A final set of random coding sheets of 50 documents i.e., P5 trial set were manually coded, and all trial sets were fed to the CNN P1 (100) + P2 (50) + P3 (50) +P4 (50) +P5 (50).

3.7 Data Analysis Using CNN

3.7.1 Coding and analysis of first set (T1).

The first set of 100 articles was mapped onto the ontology by two individuals through consensus coding. The coding was binary indicating the presence (1) or absence (0) of an element in the title, abstract, and keywords of the articles, the coders used an indigenously developed Excel tool. The two coders first coded independently on dimensions and elements of the ontology, then reconciled the differences through discussion to arrive at a consensus and to assure the reliability and validity of the coding. This coding sheet was fed to CNN software to train the software, after training the same 100 articles were coded by the CNN software and this was called a pilot run 1 (P1). The comparison sheet between manual coding and CNN coding was generated to calculate the accuracy results. The results were calculated based on the total number of coding, for 100 articles coded for 40 ontological elements, and a total of 4000 codes. In the first set of coding the CNN accuracy was 88.55%. The details about this are given below:

Coding	Count	%	CNN Statistics	
M	370	57.99%	Accuracy	88.55%
C	88	13.79%	False Omission Rate	9.91%
MC	180	28.21%	False Discovery Rate	32.84%
			False Positive Rate/Fall-Out	2.55%
			False Negative Rate/Miss Rate	67.27%
			Negative Predictive Value	90.09%
			Positive Preditive Value/Precision	67.16%
			True Negative Rate/Specificity	97.45%
			True Positive Rate/Sensitivity, Recall, Hit Rate	32.73%

Table 8: Accuracy results for the first set of coding and CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

3.7.2 Coding and analysis of second set (T2)

Coding and analysis as explained in section 3.7.1 were repeated with a new set of 50 articles (P2) to check the accuracy level of CNN analysis. A new 50 articles were randomly selected, coded, reconciled, and used for CNN analysis. The comparison sheet was generated between CNN coding and manual coding for this second set. In the second set of coding the CNN accuracy was 81.20%. The details about this are given below:

Coding	Count	%		CNN Statistics	
M	86	14.96%		Accuracy	81.20%
C	290	50.43%		False Omission Rate	5.69%
MC	199	34.61%		False Discovery Rate	59.30%
				False Positive Rate/Fall-Out	16.91%
				False Negative Rate/Miss Rate	30.18%
				Negative Predictive Value	94.31%
				Positive Preditive Value/Precision	40.70%
				True Negative Rate/Specificity	83.09%
				True Positive Rate/Sensitivity, Recall, Hit Rate	69.82%
				Index (Average of ACC, TNR, TPR)	78.04%

Table 9: Accuracy results for the second set of coding and CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

To increase the accuracy level of coding reconciliation of CNN coding and manual coding were taken and the final coding sheet is used further as a T2 to train the CNN. The changes in the reconciliation work of the set were mentioned below.

Manual	CNN	Final	Description	Total
1	1	1	Agree with M & CNN	197
1	1	0	Disagree with M & CNN	2
1	0	1	Agree with M disagree with CNN	61
1	0	0	Agree with CNN disagree with M	25
0	1	1	Agree with CNN disagree with M	12
0	1	0	Agree with M disagree with CNN	277
0	0	1	Disagree with M & CNN	1
0	0	0	Agree with M & CNN	1425
				2000

Table 10: Reconciliation work of trial run (T2) and pilot run 2 (P2)

3.7.3 Coding and analysis of third set (T3).

Coding and analysis as explained in section 3.7.1 were repeated with a new third set of 50 articles. A new 50 articles were randomly selected, coded, and reconciled. CNN trained with T1+T2, Pilot run 3 is carried on T3. The comparison sheet was generated for the CNN coding of T3 and manual coding of T3 to calculate the accuracy levels. In the third set coding, the CNN accuracy was 86.50%. The details about this are given below:

Coding	Count	%		CNN Statistics	
M	47	10.11%		Accuracy	86.50%
C	223	47.96%		False Omission Rate	2.97%
MC	195	41.94%		False Discovery Rate	53.35%
				False Positive Rate/Fall-Out	12.68%
				False Negative Rate/Miss Rate	19.42%
				Negative Predictive Value	97.03%
				Positive Preditive Value/Precision	46.65%
				True Negative Rate/Specificity	87.32%
			True Positive Rate/Sensitivity, Recall, Hit Rate		80.58%
				Index (Average of ACC, TNR, TPR)	84.80%

Table 11: Accuracy results for third set of coding and CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

To increase the accuracy level of coding reconciliation of CNN coding and manual coding were taken and the final coding sheet was used further as a T3 to train the CNN. The changes in the reconciliation work of the set were mentioned below.

Manual	CNN	Final	Description	Total
1	1	1	Agree with M & CNN	193
1	1	0	Disagree with M & CNN	2
1	0	1	Agree with M disagree with CNN	38
1	0	0	Agree with CNN disagree with M	8
0	1	1	Agree with CNN disagree with M	15
0	1	0	Agree with M disagree with CNN	209
0	0	1	Disagree with M & CNN	3
0	0	0	Agree with M & CNN	1532
				2000

Table 12: Reconciliation work of trial run (T3) and pilot run 3 (P3)

3.7.4 Coding and analysis of fourth set (T4)

Coding and analysis as explained in section 3.7.1 were repeated with a new fourth set of 50 articles. A new 50 articles were randomly selected, coded, and reconciled. CNN trained with T1+T2+T3, Pilot run 4 was carried on T4. The comparison sheet was generated for the CNN coding of T3 and manual coding of T4 to calculate the accuracy levels. In the fourth set of coding the CNN accuracy was 85.30%. The details about this are given below:

Coding	Count	%		CNN Statistics	
M	40	8.53%		Accuracy	85.30%
C	254	54.16%		False Omission Rate	2.55%
MC	175	37.31%		False Discovery Rate	59.21%
				False Positive Rate/Fall-Out	14.23%
				False Negative Rate/Miss Rate	18.60%
				Negative Predictive Value	97.45%
				Positive Preditive Value/Precision	40.79%
				True Negative Rate/Specificity	85.77%
				True Positive Rate/Sensitivity, Recall, Hit Rate	81.40%
				Index (Average of ACC, TNR, TPR)	84.16%

Table 13: Accuracy results for the fourth set of coding and CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

To increase the accuracy level of coding reconciliation of CNN coding and manual coding were taken and the final coding sheet was used further as a T4 to train the CNN. The changes in the reconciliation work of the set were mentioned below.

Manual	CNN	Final	Description	Total
1	1	1	Agree with M & CNN	175
1	1	0	Disagree with M & CNN	0
1	0	1	Agree with M disagree with CNN	38
1	0	0	Agree with CNN disagree with M	3
0	1	1	Agree with CNN disagree with M	16
0	1	0	Agree with M disagree with CNN	238
0	0	1	Disagree with M & CNN	0
0	0	0	Agree with M & CNN	1530
				2000

Table 14: Reconciliation work of trial run (T4) and pilot run 4 (P4)

3.7.5 Coding and analysis of fifth set (T5)

Coding and analysis as explained in section 3.7.1 were repeated with a new fifth set of 50 articles. A new 50 articles were randomly selected, coded, and reconciled. CNN trained with T1+T2+T3+T4, Pilot run 5 is carried on T5. The comparison sheet was generated for the CNN coding of T5 and manual coding of T5 to calculate the accuracy levels. In the fifth set of coding the CNN accuracy was 84.00%. The details about this are given below:

Coding	Count	%		CNN Statistics	
M	53	10.23%		Accuracy	84.00%
C	267	51.54%		False Omission Rate	3.45%
MC	198	38.22%		False Discovery Rate	57.42%
				False Positive Rate/Fall-Out	15.27%
				False Negative Rate/Miss Rate	21.12%
				Negative Predictive Value	96.55%
				Positive Preditive Value/Precision	42.58%
				True Negative Rate/Specificity	84.73%
			True Positive Rate/Sensitivity, Recall, Hit Rate		78.88%
				Index (Average of ACC, TNR, TPR)	82.54%

Table 15: Accuracy results for the fifth set of coding and CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

To increase the accuracy level of coding reconciliation of CNN coding and manual coding were taken and the final coding sheet was used further as a T5 to train the CNN. The changes in the reconciliation work of the set were mentioned below.

Manual	CNN	Final	Description	Total
1	1	1	Agree with M & CNN	198
1	1	0	Disagree with M & CNN	0
1	0	1	Agree with M disagree with CNN	53
1	0	0	Agree with CNN disagree with M	0
0	1	1	Agree with CNN disagree with M	17
0	1	0	Agree with M disagree with CNN	250
0	0	1	Disagree with M & CNN	2
0	0	0	Agree with M & CNN	1480
				2000

Table 16: Reconciliation work of trial run (T5) and pilot run 5 (P5)

Coding and analysis as explained in section 3.7.1 were repeated with a new fifth set of 50 articles. A new 50 articles were randomly selected, coded, and reconciled. CNN trained with T1+T2+T3+T4, Pilot run 5 is carried on T5. The comparison sheet was generated for the CNN coding of T5 and manual coding of T5 to calculate the accuracy levels. In the fifth set of coding the CNN accuracy was 84.00%. The details about this are given below:

By comparing the accuracy levels of five continuous runs for CNN analysis, the CNN software has reached the saturation level. Now the trained CNN software was used to code all the 12519 articles. The comparison results of all pilot runs were given in the table below.

	Measure	P2	P3	P4	P5
Comparison	Count M	86	47	40	53
	Count C	290	223	254	267
	Count MC	199	195	175	198
	% M	14.96%	10.11%	8.53%	10.23%
	% C	50.43%	47.96%	54.16%	51.54%
	% MC	34.61%	41.94%	37.31%	38.22%
CNN	True Positive	199	195	175	198
	True Negative	1,425	1,535	1,531	1,482
	False Positive	290	223	254	267
	False Negative	86	47	40	53
	Accuracy	81.20%	86.50%	85.30%	84.00%
	False Omission Rate	5.69%	2.97%	2.55%	3.45%
	False Discovery Rate	59.30%	53.35%	59.21%	57.42%
	False Positive Rate/Fall-Out	16.91%	12.68%	14.23%	15.27%
	False Negative Rate/Miss Rate	30.18%	19.42%	18.60%	21.12%
	Negative Predictive Value	94.31%	97.03%	97.45%	96.55%
	Positive Preditive Value/Precision	40.70%	46.65%	40.79%	42.58%
	True Negative Rate/Specificity	83.09%	87.32%	85.77%	84.73%
	True Positive Rate/Sensitivity, Recall, Hit Ra	69.82%	80.58%	81.40%	78.88%
	Index (Average of ACC, TNR, TPR)	78.04%	84.80%	84.16%	82.54%

Table 17: Comparison of accuracy results for all the pilot runs of CNN analysis
Note: M= Manual coding of the presence of the element, C= CNN coding of the presence of the element, MC= Manual and CNN coding of the presence of the element

3.8 Bibliometric Analysis

Bibliometric analysis was conducted using the VOS viewer. It allows reading the data from RIS and CSV format files. The term map analysis and co-citation analysis were conducted for the literature.

For term map analysis, the data from all three databases were transferred directly from Zotero (12519) in RIS format to VOS viewer to generate a term map.

For co-citation analysis, the Scopus data from 1960-2020 (14674 publications on 27-12-2021) was used as among all the databases Scopus is the largest database. It is converted to CSV file and used in the VOS viewer to generate the co-citation map based on Source, Reference, and author.

4.0 RESULTS

4.1 Total Literature Monad Map

The CNN coded results are represented using a monad map. The monad map numerically and visually summarizes the frequency of occurrence of each ontological dimension and element in all the literature available from the year 1960 to 2020 (Figure 12). The number adjacent to the dimension and the element is the frequency of occurrence of the same across the title, abstract, and keywords of all the literature. The bar below of each element is proportional to the frequency relative to the maximum frequency among all the elements. Since each item can be coded to multiple elements of a dimension, the sum of the frequency of occurrence of elements may exceed the frequency of occurrence of the dimension to which the elements belong.

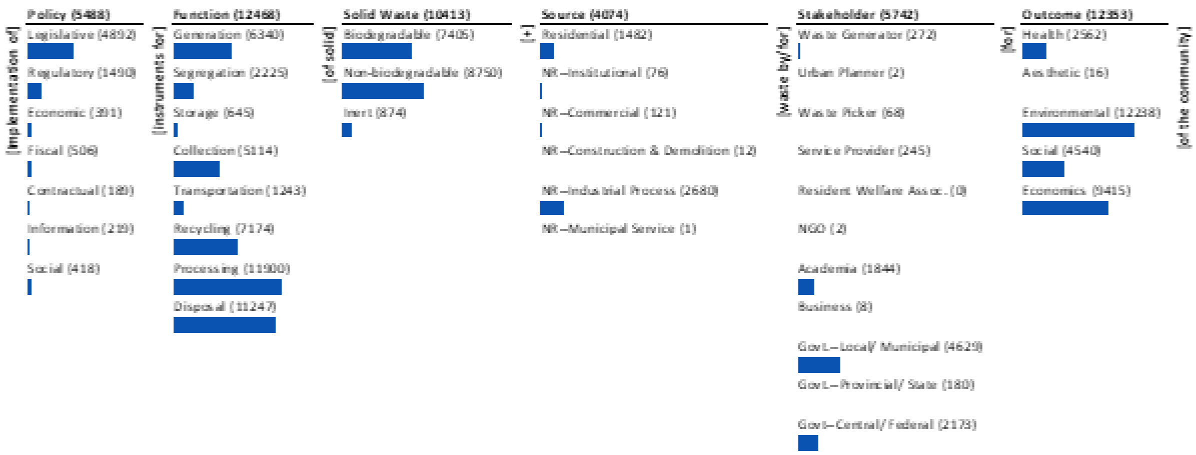

Figure 12: Monad map of CNN coded analysis

The predominant focus of the literature is on the dimensions of Function (12468), Outcome (12353), and Solid Waste (10413) of waste management. There is considerable focus on the Stakeholders (5742) involved and the implementation of Policy Instruments (5488). The least focus is on the Source of waste (4074).

Among all the elements, the dominant focus is on the Environmental (12238) element. There is substantial emphasis on Economics (9415) and some emphasis on Social (4540) and Health (2562). The least emphasis is on Aesthetics (16). The SWM functions range from generation to disposal. The literature prioritizes Processing (11900), Disposal (11247), Recycling (7174), Generation (6340), and Collection (5114) stages of SWM. There is a medium focus on the Segregation (2225) process. The least-focused function is Storage (645). A substantial amount of literature considers the types of solid waste that are generated. The highest focus is on the Non-Biodegradable (8750) wastes, followed by Biodegradable (7405) wastes. The least focus is on Inert (874) wastes.

Among the stakeholders, the major focus is on the role of Govt.--local/ municipal (4629) and Govt--central/ federal (2173) in implementing the policy instruments. There is a medium focus on the role of Academia (1844). The less focused are the roles of the Waste Generator (272), Service Provider (245), Govt.--provincial/ state (180), and the Waste Picker (68). A little or no focus on the roles of Business (8), Urban Planner (2), and NGO (2). There is zero focus on the element Resident Welfare Association. Amongst the elements of policy instruments, a high priority is on the implementation of Legislative (4892) and Regulatory (1490) policy instruments. Elements like Fiscal (506), Social (418), and Economics (391) received moderate attention.

Contractual (189) policy instruments are the least focused element. Compared to all the other dimensions, the source of waste received very less attention. Among the sources, Non-Residential Industrial Process (2680) and Residential (1482) elements have high priority. Non-residential sources of waste like NR--Commercial (121) and NR--Institutional (76) have been moderately focused. However, NR--Construction & Demolition (12) and NR--Municipal Service (1) have little or no attention.

4.1.1 Decade-wise Bar graphs of literature from 1961-70 to 2011- 2020.

The results are broken down into decades and visualized using a monad map and bar graphs for detailed analysis. Decade-wise analysis using bar graphs helps in understanding the changes in the literature over years. Figure 1 summarizes the frequency of occurrence of each ontological dimension and element in the literature available in the decade 1961 to 1970.

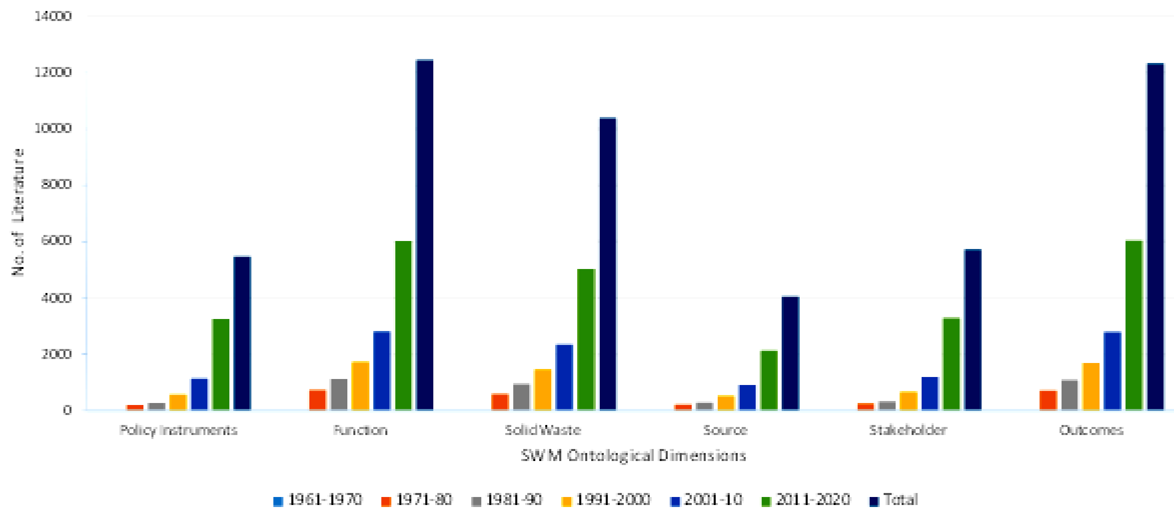

Figure 13: Decade wise comparison of SWM ontological dimensions

The above graph shows the consistent increase in the focus for all the dimensions, namely, Policy Instruments, Function, Solid Waste, Source, Stakeholder, and Outcomes as decades progress from 1961-1970 to 2011-2020. However, from 1971-1980 to 2011-2020, the priority of the state of research in SWM is inclined towards Function related papers followed by outcomes and then Solid Waste. Out of the total of 12519 papers from all the decades, only 4063 talks about the source of solid waste, making it the least focused topic of research within SWM. On the other hand, 12444 papers talk about Function related aspects.

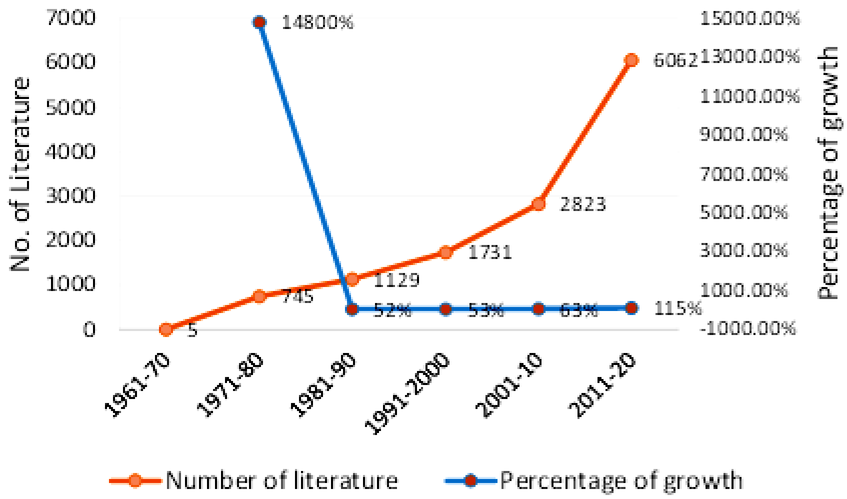

Figure 14: Growth of Literature over decades (1961-70 to 2011-20)

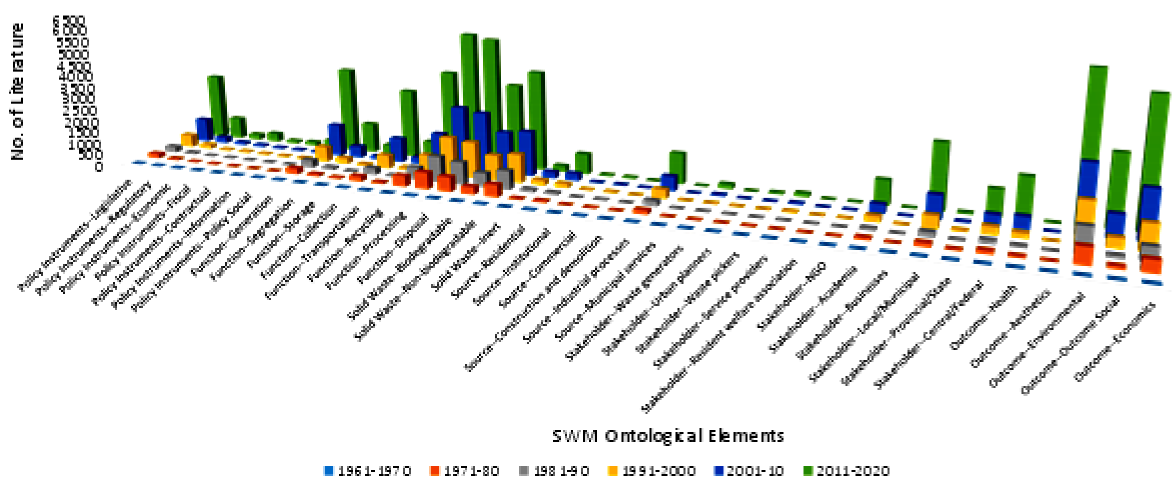

Figure 15: Decade wise trends comparison of SWM ontological Elements

Figure 15 provides insights into the decade-wise trends of ontological elements. However, Figure 14 depicts the absolute increase in the number of publications and the percentage increase over each decade. Hence it can help in comparing the occurrence of each ontological element. First, the total number of literature published in each decade varies drastically. In 1961-70, the number of publications was 5 but by the decade 1971-80, it reached 745, which is a whopping 14800% progress. The absolute number of literature progressed till the decade 2011-20 but the percentage growth decreased to 52% in 1981-90. The percentage growth increased gradually over decades and reached 115% in 2011-20.

Though function dimension is the priority of the state of research, the outcome element, environmental (12214), is the highest prioritized topic of research when it comes to the total number of publications (12495) till 2020. The occurrence of environmental outcomes in each decade's (1961-70 to 2011-20) publications progressed from 80% to 99%. The economic outcome has been fairly focused since 9393 publications talk about the same out of the total literature. The occurrence of economic outcomes in each decade's (1981-90 to 2011-20) publication progressed from 52% to 86%. The social outcome has below average representation of 4525 publications overall. The occurrence of social outcomes in each decade's (1961-70 to 2011-20) publication slightly progressed from 40% to 46% with a dip of 10% in the decade 1981-90. Health outcomes have a very small number of publications (2555). The occurrence of health outcomes in each decade's (1961-70 to 2011-20) publication progressed from 0% to 29%. The least focused element is aesthetics, with only 15 publications overall. Hence the percentage increase is negligible.

Among the functions, processing is highly prioritized with 11877 publications. The occurrence of topics regarding processing in each decade's (1961-70 to 2011-20) publications progressed from 80% to 94% with a peak of 97% in the decade 1981-90. Disposal is also near-equal to 11224 publications. However, the occurrence of disposal element in each decade's (1961-70 to 2011-20) publications diminished from 100% to 92%. Recycling has gotten off to a good start in the early decades and has 7154 publications discussing the same. The occurrence of topics of recycling in each decade's publications diminished from 100% to 65% with a steep drop of 45% in the decade of 1991-2000. Generation and collection also have similar trends as recycling. The occurrence of topics of generation and collection in each decade's publications diminished from 80% to 61% and 80% to 49%, respectively. There is a huge drop in focus when it comes to segregation across all the decades compared to earlier mentioned function elements. However, the occurrence of topics of segregation in each decade's publications increased from 0% to 21%. The occurrence of topics of transportation and storage in each decade's publications diminished from 40% to 12% and 40% to 6%.

Among the solid waste elements, non-biodegradable wastes have high priority with 8733 publications. The occurrence of non-biodegradable wastes in each decade's (1961-70 to 2011-20) publications increased from 40% to 70% with a stagnant growth over the last 4 decades. The occurrence of biodegradable wastes in each decade's publications increased from 20% to 60%. There are only a few publications (873) on inert waste. The occurrence of inert wastes in each decade's publications slightly improved from 0% to 4%.

Among the policy instruments, the legislative holds higher priority with 4877 publications. However, the occurrence of legislative policy in each decade's (1961-70 to 2001-2010) publications increased from 40% to 37% and slight progress to 47% in 2011-20. The occurrence of regulatory policy in each decade's publications decreased from 20% to 15%. The case with fiscal and social policy is that it shows slight improvement over decades. The occurrence of fiscal and social policy in each decade's publications slightly improved from 0% to 6% and 0% to 5%. Unlike other elements which saw a rise in their occurrence in literature over decades, economic policy reduced from 40% to 4%. Both information and contractual policy instruments received modest attention with just 217 and 188 publications overall, respectively.

Among the stakeholders, local/municipal is prioritized with 4614 publications. The occurrence of local/municipal in each decade (1961-70 to 2001-2010) publications marginally increased from 40% to 46%. Central/Federal and academia have had a similar pattern over decades. For both, the occurrence in each decade's publications decreased from 40% to 18% and 20% to 17%. Waste generators and service providers have gained attention only in recent decades. For both, the occurrence in each decade's publications moderately increased from 0% to 3% and 0% to 2%. Provincial/ State saw a decline from 4% to 1% over 1971-80 to 2011-20. The share of publications discussing waste pickers is nearly zero in almost all the decades except for the latest decade with just 1%. The total number of publications for business, urban planners, and NGOs are 8, 2, and 2, respectively. The role of Resident Welfare Organizations is not discussed over the decades.

Among the Sources, the industrial process has a moderate focus on publications. The occurrence of industrial processes in each decade's (1961-70 to 2001-2010) publications has sharply declined from 60% to 22%. The occurrence of residential in each decade's publications has improved from 0% to 14%. Commercial and institutional have only 1% representation in the literature over the decades. Construction and demolition and Municipal services have overall 12 and 1 publications, respectively.

4.2 Full-Text Analysis- Comparison between QDA miner and CNN analysis.

This section explains the comparison results of QDA miner and CNN analysis using 1000 publications randomly selected from the total literature. The full paper text mining was conducted using QDA miner lite software. Simultaneously, for the same set of 1000 publications, a text mining analysis of title, abstract, and keywords is conducted using CNN software. The results from both analyses are visualized using monad maps and a bar graph which shows the comparison between QDA miner and CNN analysis.

4.2.1 Text mining of full paper using QDA miner lite software

QDA miner is a qualitative text data analysis software, and it is used for text mining of full papers. The PDFs of 1000 publications were fed into QDA miner software. The software converts each PDF file into a text document. Ontological elements were coded into the software to analyze the documents. The accuracy of analysis was increased using synonyms of a few elements. These synonyms were manually entered into the software. The list synonyms used for elements are given below:

Generation	Produce
Segregation	Sorting
Transportation	Transfer
Processing	Composting, Biogas, Biomethanation, Anaerobic Degradation, Aerobic Degradation
Disposal	Landfill
Biodegradable	Organic Fraction of Municipal Solid Waste, Wet Waste, Organic waste, Vegetable and fruit peels, Tea leaves, Cooked food, Garden Waste
Non-biodegradable	Dry Waste , Platic, Paper, Cardboard, Metal, Glass, Rubber, Cloth, Leather, Wood
Inert	Dust
Residential	Houses
Waste pickers	Rag pickers
Service providers	Contractors
Aesthetics	Smell
Economics	Cost

Table 18: Synonyms used for QDA miner analysis

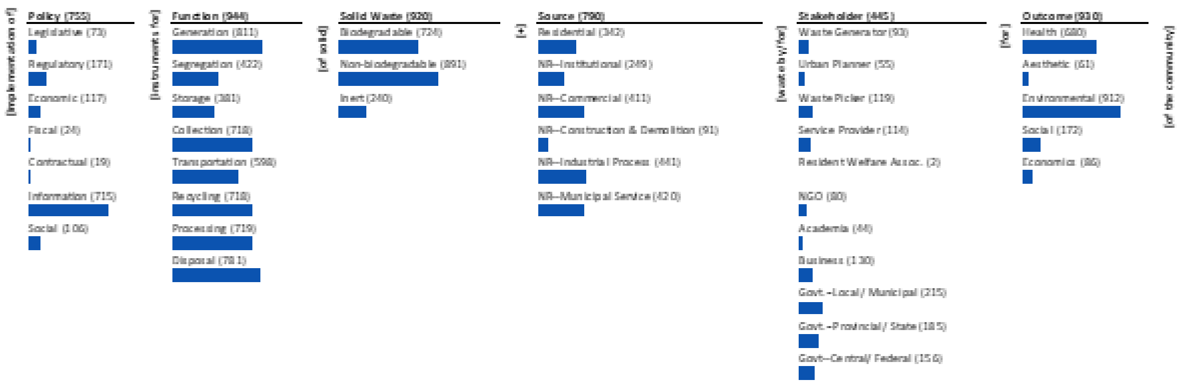

Figure 16: Monad map of QDA miner analysis

Figure 16 shows the results from QDA miner analysis which is overlayed on the SWM ontology. This monad map shows the frequency of occurrence of the elements. Among the dimensions, Function (944) dimension received more emphasis followed by Outcomes (930), Solid Waste (920), and Source (790). Among the elements, Environmental (921) outcome received major emphasis in the study, followed by Health (680) outcome and Social (172) outcome. Economics (86) and Aesthetic (61) outcomes received less emphasis. Non-Biodegradable (891) and Biodegradable (724) waste received the highest attention among the Solid Wastes dimension and Inert (240) with less emphasis.

Among the Function (944), Generation (811), Disposal (781), Processing (719), Collection (718), and Recycling (718) received maximum attention with average attention on Transportation (598), Segregation (422), and Storage (381). Out of all the Policy Instruments (755), Information (715) received maximum emphasis. However, the emphasis on all the other elements is below average. In Source (790) dimension, NR-Industrial process (441), NR-Municipal services (420), and NR-Commercial (411) received more emphasis than other elements. From the Stakeholder (445) dimension, Municipal/ Local Government (215), Provincial/ State Government (185), and Central/ Federal Government (156) received more emphasis compared to other elements.

4.2.2 Text mining of descriptive metadata using CNN.

The titles, abstracts, and keywords from 1000 publications were used for text mining analysis in CNN. The results are represented in the monad map as given below:

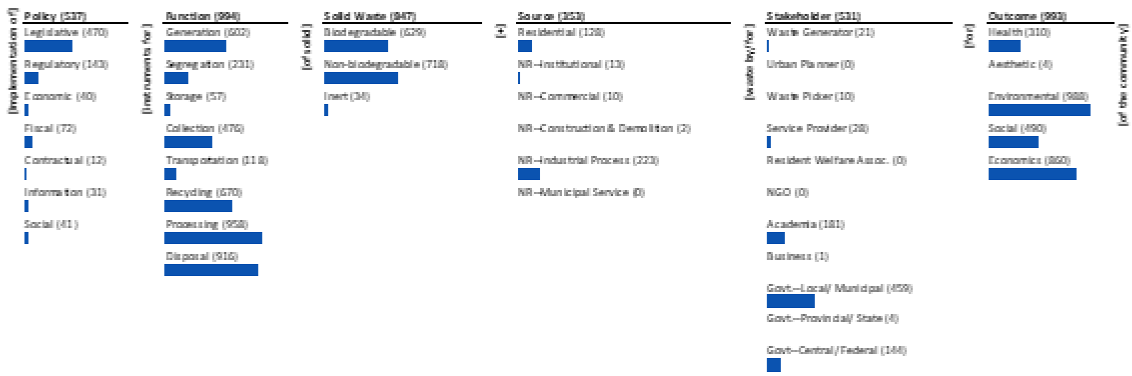

Figure 17: Monad map of CNN analysis

Figure 17 shows that the Outcome (903) elements like Environmental (988), and Economics (860) elements received major emphasis, followed by Social outcome (490), and Health outcome (310). Aesthetic outcome (4) has received very less emphasis. Among the Function (994) dimensions, Processing (958), Disposal (916), Recycling (617), Generation (602), and Collection (476) received maximum attention. Whereas Segregation (213), Transportation (118), and Storage (57) received less emphasis.

Among solid waste (847) dimensions Non-Biodegradable (718) and Biodegradable (629) waste received the highest attention. Whereas Inert (240) received less emphasis.

Among Policy Instruments (537), Legislative (470), and Regulatory (143) received maximum emphasis. However, all others received very less emphasis. From the Stakeholders (531) dimension, Municipal/ Local government (459), Academia (181), and Central/ Federal government (144) received maximum emphasis than other elements. In Source (353) dimension, NR-Industrial process (223), and Residential (128) received higher emphasis. Whereas all others with very less attention.

4.2.3 Comparative visualization and interpretation of both QDA miner and CNN analysis.

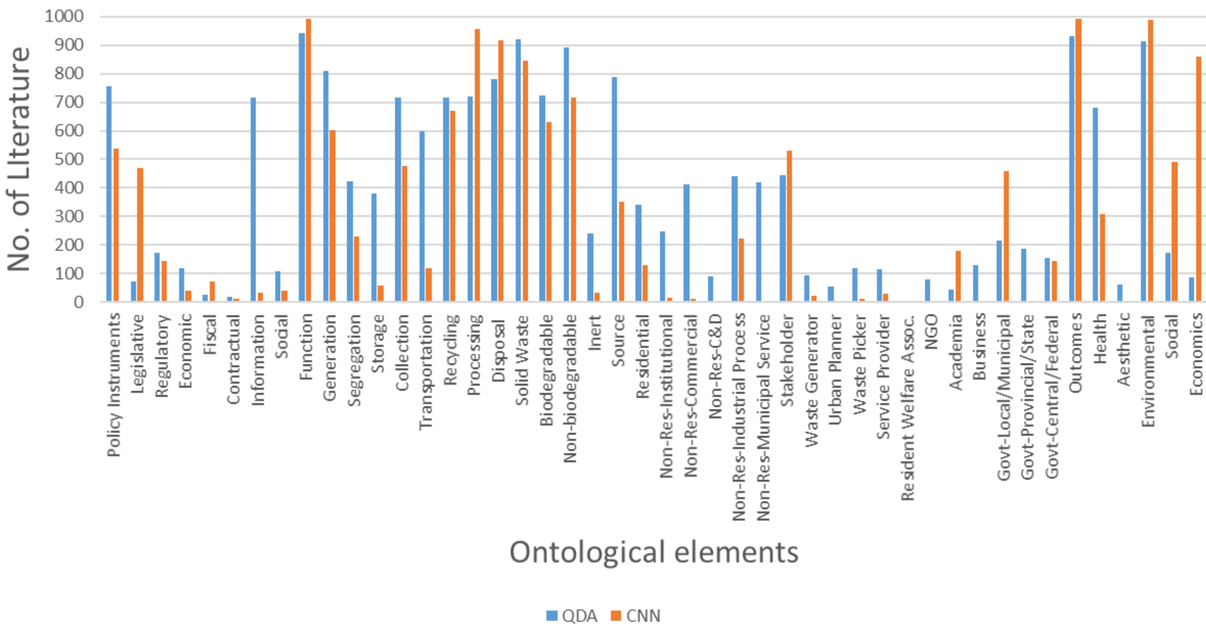

Figure 18: Comparison of QDA miner and CNN analyses.

Figure 18 shows the comparison of results for both QDA miner and CNN analysis of 1000 publications. Both analyses revealed that the environmental outcome element received maximum emphasis. However, the CNN analysis showed that the maximum occurrence of environmental outcome is 988 and 912 from the QDA miner analysis. Economics (860), and Social (490) outcome received higher emphasis after environmental outcome in the CNN analysis. whereas in QDA miner analysis Health (680) outcome received higher emphasis.

Among the functional dimension, Processing (958), Disposal (916), Recycling (670), and Generation (602) carry dominance in CNN analysis. On the other hand, Generation (811), Disposal (781), Processing (719), Recycling (718), and Collection (718) when it comes to QDA miner analysis. All the elements in solid waste dimension followed the same pattern of emphasis in both CNN and QDA miner analysis. Though the number of occurrences of elements differs between both analyses, results show a similar pattern when it comes to the emphasis of elements across publications.

The QDA miner results show that the elements of source dimension have high occurrence frequencies compared to CNN analysis. In QDA miner results NR--Industrial Process (441), and NR--Municipal Service (420) both have received maximum attention whereas in CNN analysis results NR--Industrial Process (223) has got higher emphasis. However, other elements from the same dimension received less attention in both QDA miner and CNN analysis. The stakeholder dimension has shown an almost similar pattern in both analyses with higher emphasis on local/municipal government. The QDA miner results showed higher emphasis for Information (715) element under policy instruments dimension whereas CNN results showed a higher emphasis on Legislative (470) element.

4.3 Term network map analysis and Co-citation analysis using VOS Viewer

4.3.1 Term network map analysis

Figure 19 is created by analyzing titles and abstracts of 19519 Publications on SWM from the year 1960-2020 using natural language processing techniques. A total of 170502 relevant terms were identified from titles and abstracts of all the publications using binary counting. Further to increase the relevancy of the terms, a minimum number of occurrences of a term is set at 20, or in other words, the terms which are found in at least 20 publications are selected. Around 2710 publications met the threshold. The software uses its relevancy score calculations, the default is to select 60 percent of the most relevant terms which is 1626 terms. The list of terms was further scrutinized manually for shortlisting the terms to 1000 that are relevant to the SWM ontology.

These terms are shown in the term map visualization provided in Figure 19. The following map has 1000 terms mapped into various clusters. Each cluster has a specific color (Cluster 1- red, cluster 2- green, Cluster 3- blue, cluster 4- yellow, and cluster 5- black). This map is used to split the SWM literature into various subfields. The red cluster or cluster 1 on the left side of the map broadly covers functional elements and outcomes of the SWM. The green cluster or cluster 2 on the right side of the map broadly covers SWM policies and their implementation by various stakeholders. The blue, yellow, and black clusters in the middle area cover the environmental outcomes related to SWM. The red and green clusters are the major clusters with a maximum link strength and terms. The terms like concentration (Red cluster), and waste management (Green cluster) have the highest number of occurrences and links.

The circles and labels denote each term on the map. The size of the circle for each term denotes the publications in which the term was found, and the distance between the two terms offers an approximate indication of the relatedness of the terms. Co-occurrences determines the relatedness of terms. For instance, the terms in the green cluster, 'Waste Management' and 'City' have a link strength of 923 which means both terms co-occur 923 times or in 923 publications, hence a strong relationship between the terms. In the map, both clusters and curved lines indicate the strongest relations between the terms. An interactive version of the term map visualization is presented in Figure 19.

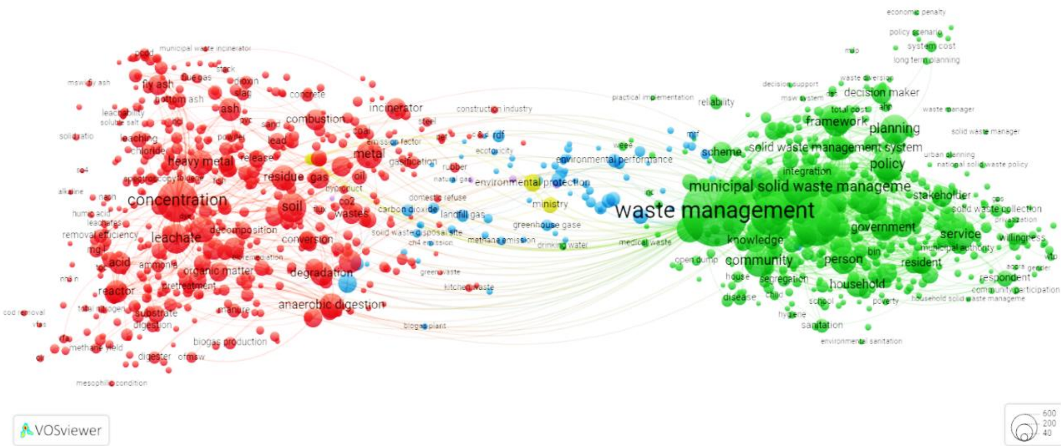

Figure 19. VOS viewer term map visualization.
Source: Visualization in VOS viewer. The minimum number of occurrences: 20, Binary counting, 1000 items, 5 clusters, and 135186 links.

4.3.2 Co-Citation analysis based on authors

Figure 20 is created by analyzing citations of authors from the 14674 publications. A total of 232437 authors were identified. The minimum number of citations of an author is kept at 20 to make the results concise. Hence, only 4934 authors met the threshold. For each of the 4934 authors, the total strength of the co-citation links with other authors is calculated. Hence by default, 1000 authors with the greatest total link strength are selected.

Author co-citation analysis measures the number of times a particular group of authors was cited together within the collection. In this visualization, the clusters show the pivotal authors whose works were cited in various publications.

Authors in the five clusters (red, blue, green, yellow, and purple) appear as the main scholars, contributing to the literature of SWM for the past few decades. Author Huang, G.H. in the purple cluster has the most co-authorship link of 902 and citation of around 2314. Followed by works from authors such as Christensen, T.H., from the yellow cluster, Wilson, D.C., from the green cluster, Zhang, Y., from the red cluster, and Barlaz, M.A., from the blue cluster are the major contributors to the field of SWM.

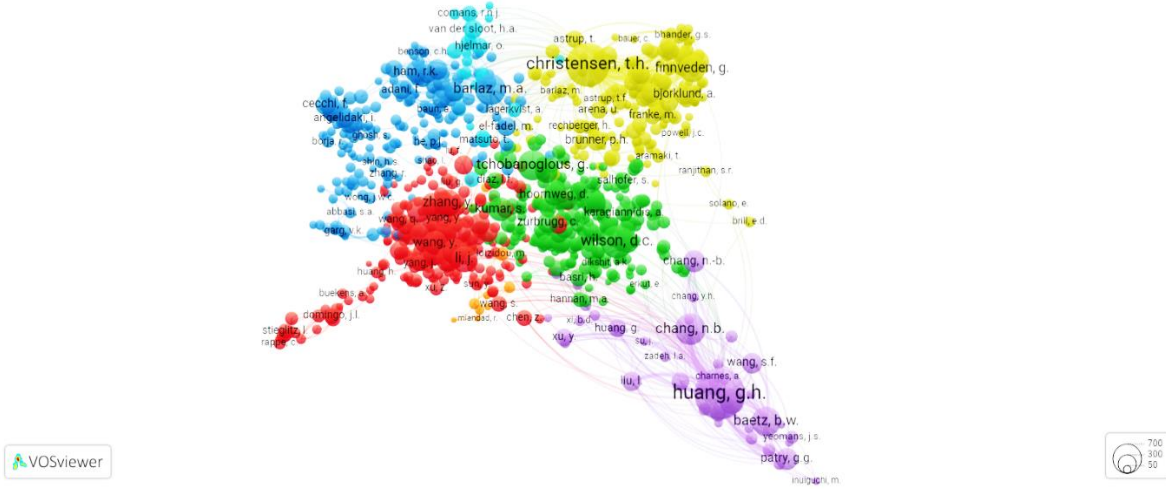

Figure. 20 Author Co-citation network
Source: Visualization in VOS viewer from Scopus. Minimum number of citations: 20, 1000 items, 7 clusters, and 316617 links.

4.3.3 Co- Citation analysis based on cited sources

Figure 21 represents the co-citation analysis based on the unit of cited source. The minimum number of citations of a source is kept at 20. Out of the 104948 sources only 1120 sources met the threshold. Out of 1120 sources some of the items have no connection hence only 998 sources with connections are plotted in Figure 21. Software plotted the data into seven clusters with total link strength of 1592375..

Source co-citation analysis measures the number of times a group of sources or journals cited together in the collection. Seven clusters were identified (red, green, blue, algal green, and purple). Source waste management from cluster two (Green colour) showed the maximum link strength of 129319 with the co-citations of about 7421. Followed by waste management from cluster 4 (Algal green colour), chemosphere from cluster one (Red colour), bioresourse technology from cluster 3 (Blue colour), and waste management from cluster five (Purple colour).

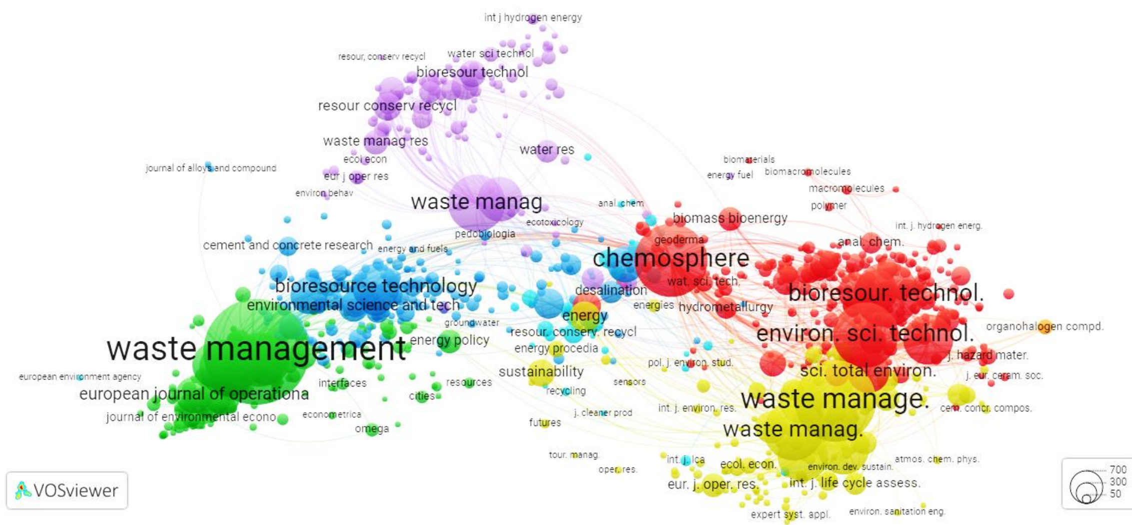

Figure 21. Cited source co-citation network.
Source: Visualization in VOS viewer. Minimum number of occurrences: 20, Full counting, 1120 items, 7 clusters, and 93493 links

4.3.4 Co-Citation analysis based on cited references

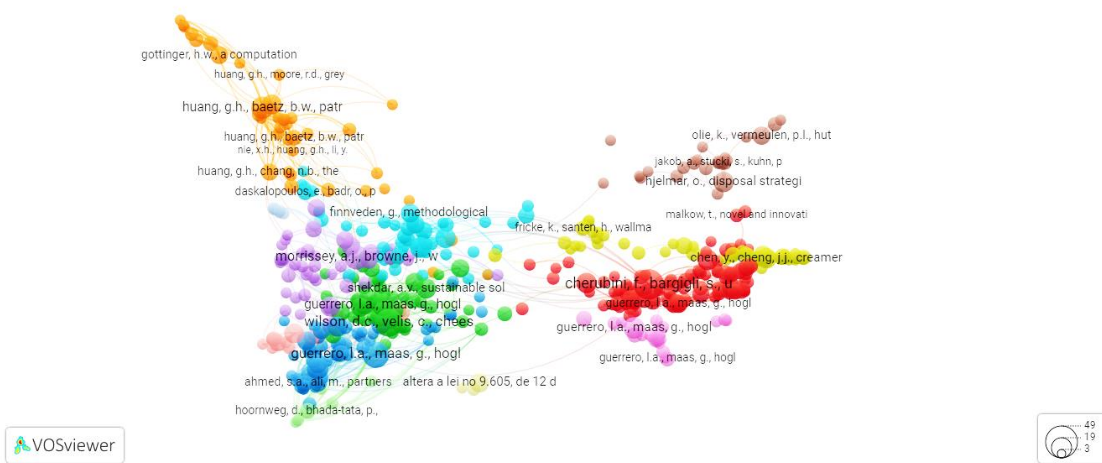

Figure 22. Cited reference co-citation network.
Source: Visualization in VOS viewer. The minimum number of occurrences: 6, Full counting, 541 items, 14 clusters, and 3805 links.

Figure 22 represents the co-citation analysis based on the unit of a cited reference. The minimum number of citations are kept at 6 since a higher threshold would eliminate a good amount of elements. Out of the 294920 cited references, 541 met the threshold. Out of 541 items some items have no connection hence only 510 references with connections were plotted in this map. Software mapped them in 14 clusters with total link strength of 5637.

This visualization shows the dominant co-cited references within the collection, the reference from cluster 11 (lite green color) Hoornweg, D., Bhada-tata, P., (2012) showed maximum link strength of 252 with 12 citations followed by Cherubini, F., Bargigli, S., Ulgiati, S., life cyc. (Red color), Huang, G.H., Baetz, B.W., Patry (orange), Guerrero, L.A., Maas, G., Hogland, W., solid waste (Blue colour), and Banar, M., Cokaygil, Z., Ozkan, a., life cycle ass. (Sky blue Colour).

5.0 FINDINGS/SUMMARY AND RECOMMENDATIONS

The study revealed that's there in fact a wide variation in the knowledge of, decisions about, implementation, and assessment of SWM globally. The report gives a comprehensive analytical understanding of the past trajectory of SWM practices globally and which in turn will help to direct its future in India.

1. Overall, the predominant focus of the literature is on the functional elements and environmental outcomes of a typical SWM system.
2. The data shows a whopping absolute and relative increase in the number of publications over each decade.
3. Across all the decades, there is inconsistent focus on the policy instruments, source of waste, and stakeholders of the SWM system.
4. The comparative analysis of full paper coding and descriptive metadata coding study revealed that the result from both methods followed the same pattern. But the descriptive metadata coding gives more options like manual coding, data availability, and data management for quality analysis. It is recommended to follow descriptive metadata coding rather than full paper analysis.
5. The ontology and the analysis of the present SWM corpus can help comprehend the pathways that are emphasized, not emphasized, and absent during the past few decades.
6. The study put forward an agenda for research to fill the gaps in the state-of-the-research, -policies, and –practices in the field of SWM.
7. The term network map using VOS viewer discloses the prominence of terms like 'Concentration', 'Waste management', 'City', etc. indicates the bias in the SWM literature towards certain fractions of the topic.
8. The co-citation analysis revealed the literature used for analysis is qualitatively and quantitively significant. The cited author Huang, GH with 2314 citations, cited source waste management with 7421 citations, and cited reference Hoornweg, D, Bhada-tata, P (2012) with 12 citations, we can notice a significant citation number.

6.0 FUTURE PLAN

The current study has focused on the research gaps in the understanding of the SWM system globally. It helps to put forward an agenda for research in the state-of-the-research in the field of SWM. It also analyses the significant contribution of the research corpus. Further work may involve:

1. Country-wise understanding of the research contribution in the area of SWM. We can also find India's contribution compared to the global scenario.
2. Analysis of policy to find the research gaps to achieve an efficient SWM system. The concept and perception of waste are different from one country to another country.

Policy analysis can help to put forward our agenda of an efficient waste management system.

3. We plan to analyze the projects, and funds invested in the SWM system. As the research funding should go towards not only stating the problems but also to find the solution.
4. The study of research, policy, and funding will help to understand the gaps in the SWM system. And develop the road map for solid waste management systematically for the next few decades.

7.0 EXPECTED PUBLICATIONS

1. Techniques for big text data mining technique to analyze SWM research globally, Tentative outlet: Waste Management Journal
2. A comparative analysis of SWM using text mining of descriptive meta data and full paper, Tentative outlet: Journal of Material Cycling and Waste Management
3. Ontological visualization and interpretation of big data in the field of Solid Waste Management, Tentative outlet: Waste Management Journal
4. A systematic analysis of Solid Waste Management using CNN analysis, Tentative outlet: Environmental Sustainability

8.0 REFERENCES

Broadus, R.N. (1987). Toward a definition of "Bibliometrics". Scientometrics, 12, 373–379.

Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. Int. J. Human-Computer Stud, 43, 907-928.

Kashyap, S., Ramaprasad, A., Singai, B. C. (2019). An ontological analysis of challenges involved in Urban Solid Waste Management (Abstract). 2nd International Conference on Environmental Geotechnology, Recycled Waste Materials and Sustainable Engineering (EGRWSE-2019), UIC, Chicago).

MSW (Management and Handling) Rules. (2000). Ministry of Environment. Government of India. New Delhi.

Pappu, A., Saxena, M., Asokar, S.R. (2007). Solid waste generation in India and their recycling potential in building materials. Journal of building and environment, 42(6), 2311-2324.

Sharholy, M., Ahmed, K., Mahmood, G., Trivedi, R.C. (2008). Municipal solid waste management in India cities – A review. Waste management, 28, 459-467.

Shekdar, A.V. (1999). Municipal solid waste management – the Indian perspective. Journal of Indian Association for Environmental Management, 26 (2), 100-108.

Solid Waste Management Rules. (2016). Ministry of Environment. Government of India. New Delhi.

UNFCCC, The United Nations Framework Convention on Climate Change. (2012). Methodological tool "emissions from solid waste disposal sites". EB 66 Report Annex 46, version 06.0.1.