DESIDOC Journal of Library & Information Technology, Vol. 43, No. 4, July 2023, pp. 266-275, DOI : 10.14429/djlit.43.4.19237

Book Purchase Model for Academic Libraries Using Digital Humanities Tool

Aysha Zakhiya A.*, Aparna K.S. and Syamili C.

Department of Library and Information Science, University of Calicut, Kerala-673 001, India
**E- mail: ayshazakhiyaa@gmail.com*

ABSTRACT

Most libraries don't buy books based on a set of rules that reflect the research done at their affiliated institutions. A structured method for evaluating evolving research areas is a must have process in libraries, especially academic libraries. So, in this study, with the help of the digital humanities tool, the authors developed a book purchasing algorithm for an academic library that functions in accordance with the core research subjects and evolving research areas of that institution. To conduct this study, the researchers searched the SCOPUS database for articles published by the "University of Calicut" as of February 13, 2023, yielding 3305 publications. Voyant tools have been used to analyse the collected data. Significant changes in the area of research at the University of Calicut are observed from its beginning to its recent history (1967–2023). The analysis revealed a discrepancy between the primary subject areas of research and the book collection at the Calicut University library. The findings of this paper will aid librarians in their acquisition processes, and the institution will be able to support the research community. This study proposes a novel "Book Purchase Model for Academic Libraries" using a digital humanities tool that can be used in professional practice.

Keywords: Digital humanities; Voyant tools; University of Calicut; SCOPUS; Research trends; Collection development; Digital humanities tools

1. INTRODUCTION

Digital Humanities (DH) is a vast research and multidisciplinary area were technology crossover humanities disciplines like literature, history, statistics, art, media studies etc. It is a two way approach that-(i) a systematic use of digital resources in humanities discipline and (ii) humanistic analysis of application of digital resources.[1] A proper definition of digital humanities given by Oxford dictionary is 'an academic field concerned with the application of computational tools and methods to traditional humanities disciplines such as literature, history, and philosophy.[2] Discussion of digital humanities and its connection to libraries has grown rapidly in the past several years. Sula[3] in their study defined that, Stephen Ramsay has linked digital humanities to one of the oldest functions of the library, namely knowledge organisation.

Libraries can use digital humanities tools (Voyant tool, Omeka, Tableau, Word press, Demo tool) to digitise their archives and make them accessible online. These tools can also be used to analyse large sets of text, such as books or academic articles. This can help researchers identify patterns and themes, and can also be used to create visualisations that help to communicate complex information.

In this study, the researchers applied digital humanities tools in University Libraries of University of Calicut, for identifying research areas and to assist librarians to purchase books in a more effective way. University of Calicut is the largest university in Kerala which was established in 1968. CHMK Library, the central library of University of Calicut has a collection of about 95000 books and it subscribed to more than 125 journals and ten newspapers. The library has the facility of Online Public Access Catalogue (OPAC), which was used by the researchers of this study for collecting data.

A structured method for evaluating research areas is a must have process in libraries, especially academic libraries. Digital humanities can assist libraries for this. It helps in analysing the subjects using various digital humanities tools and scientifically finding out the most required subject fields. So, in this study, with the help of digital humanities tool, the authors developed a novel book purchasing model for academic libraries that functions in accordance with the core research subjects and evolving research area of that institution.

Received : 25 February 2023, Revised : 01 April 2023
Accepted : 30 June 2023, Online published : 04 September 2023

2. REVIEW OF RELATED LITERATURE

Even though digital humanities is an evolving topic in research, the available studies show that it can be used to analyse and interpret various incidents that happens in our daily life. Jueman Zhang & others[4] analysed the role of twitter in COVID-19 and vaccination. The results show that the public shows much interest in getting information about the pandemic and vaccination. Idler and others[5] conducted a similar study about COVID-19 and its relation with religion. The results showed that there are both narratives and counter narratives on this. Lu and others[6] did their study collecting data from SinaWeibo, a Chinese social media platform. Analysis shows that there are both positive and negative comments, with most negative comments being anger, sadness and anxiety.

In 2016, Grandjean[7] conducted a study that analysed digital humanities through twitter. The author did an effort to define the field of digital humanities, and it was discovered that linguistic groupings are important criteria to explain clustering inside a network with comparable characteristics.

Elo[8] in 2022 attempted to understand unstructured textual documents like speeches. For this the author conducted a text analysis and visualisation of New Year speeches held by politicians from Germany, Switzerland and Austria from 2000 to 2021. The results revealed that the use of concepts related to crisis and insecurity in their speeches has increased dramatically since 2008. Similarly, in 2022, Mlambo and others[9] used the digital humanities tools such as Voyant tool and multilingual concordancer (ParaConc) to outline the approaches taken for translating the South African Constitution from English into Xitsonga, Siswati, and isiNdebele when those target languages lacked terminology equivalents.

Asarta & Mendez-Carbajo[10] in 2021 examined the abstract of articles of an economic education journal using digital humanities tools. The results show there are changes in research overtime. A study by Anson and Moskovitz[11] in 2021 showed the word patterns in the research output of applied science faculties. Bernau[12], through his study in 2018, provided keyword trends and authorship to flagship sociology journals using data from JSTOR. The research conducted by Fallucchi and others[13] in 2022 shows how digital humanities tools help users to identify structural and linguistics patterns and research.

Wang & others[14] conducted a study in 2021 on Qing Dynasty HetuDangse and used text analysis and visualisation to conduct data mining analysis of historical document data. The study by Ullah and others[15] focused on visualising 'Never Let Me Go' using Voyant tool, and generated word cloud for text analysis, comparison, prediction, keyword extraction etc. But the major drawback of the study was hidden meanings and messages cannot be understood. Basak and Roy[16] conducted a bibliometric study using Scopus data for mapping the literature on digital humanities. The authors observed the considerable increase in the research publications on digital humanities. The main contributing countries in research are Germany, USA and UK. Sula[3], in their paper, analysed the current areas of digital humanities work, presented a cultural informatics model of libraries and the digital humanities.

3. RESEARCH QUESTIONS

The following are the major research questions

1. To evaluate the existing book purchasing system in academic libraries.
2. To identify the core research area and research trends of University of Calicut using digital humanities tools.
3. To visualise research productivity of University of Calicut using digital humanities tool.
4. To identify the missing gap between the core research areas of University of Calicut and existing book collection in University libraries with the help of digital humanities tools.
5. To explore how digital humanities tools can be used to evaluate book collection and frame book purchase model for university libraries.
6. To propose 'Book Purchase Decision Making Model for Academic Libraries' using digital humanities tools.

4. ANALYSIS AND INTERPRETATION

4.1 Evaluation of Book Purchasing Principles Followed in Academic Libraries

There are many book purchasing principles, among which, Drury's principles, Dewey's principle, McColvin's principle and Ranganathan's laws[17] are most popular. Drury's principle states "Right book to right reader at the right time". It shows the importance of making the information reach the user when they need it the most. Right time for a subject in this context means, when research is high in that subject. Dewey's principle says "Best reading for largest number at least cost". Digital humanities can be applied to identify the research trends of the concerned institution and help libraries realise the most needed books and invest money in them. McColvin's Principle is "Books in themselves are nothing". So the best way to make the books useful is to purchase the books that are actually in demand. In his five laws of Library Science, Ranganathan states "Books are for use" as first law. That is unused books stay irrelevant. In an academic library, it's important for the librarians to understand the research field to make their library collection more useful to them. Digital humanities tools can be used for this purpose as it easily shows the areas that are most involved in research.

Central Library of University of Calicut usually ask Department librarians about the needed books that must be included in the central library. Department librarians, from their experience and personal knowledge recommend some books that are important in respective fields. But, no scientific examination of research trends were done or

research areas are evaluated scientifically. So a structured method for evaluating research areas is a must have process in libraries, especially academic libraries. Hence digital humanities can be a solution for this. It helps in analysing the subjects using various digital humanities tools (Voyant tools, etc.) and thus scientifically finding out the most required subject fields.

4.2 Core Research Area and the Research Trends of University of Calicut

This Research question is analysed in two phases. First phase where the core research area of University of Calicut is identified, and in next phase research trends of University of Calicut are examined.

4.2.1 Core Research Area of University of Calicut Methodology

The data used for the study is research publications affiliated to University of Calicut extracted from SCOPUS database, by focusing on Title, Abstract, Keywords and Year of publication of the articles. A total of 3305 results were generated on 13/02/2023. The data is categorised based on the decades the articles are published, as shown in Table 1

Table 1. Research output of University of Calicut

Classification	Period	No. of documents
2020s	2020 - 13/02/2023	854
2010s	2010-2019	1551
2000s	2000-2009	483
1990s	1990-1999	205
1980s	1980-1989	151
Less than 1980s	1967-1979	61

While analysing the number of research outputs each year, a massive growth in number of research articles can be seen. 2020s only had 3 years and 1.5 month so far, but within short amount of time more than 800 articles are published, which is impressive. Since the data is extracted from SCOPUS database, it will be of high standard and quality. The manual verification of number of articles published in each year resulted in arriving at the conclusion that the research output in University of Calicut is progressing positively and in very hopeful manner.

This extracted data is now uploaded into Voyant tool. A result page with 5 sections was generated.

Section i: **Cirrus**: word cloud of most repeated terms in corpus

Section ii: **Reader**: the documents uploaded in plain text

Section iii: **Trends**: graphical representation of most repeated words

Section iv: **Summary**: details on terms density, length, count etc.

Section v: **Contexts**: left and right occurring sentences of all frequently used terms.

From 'Terms' option in Section i, most repeating 50 terms are noted.

Next step is to find stopwords and exclude them from the list of most repeated terms. To identify stopwords, the repeated terms must be examined manually to understand the subject in which those terms appear the most. Since covering all the terms manually is impossible, a sample is taken for analysing.

Sample Selection:

5 % of all selected terms (50) are determined in each decade with respect to their total occurrences.

For example:

In a similar way, the sample size for remaining 49 terms are found. After long calculations, the total sample size is fixed as 2419. The sample units are randomly selected and analysed from 'Contexts' option in section v. The left and right hand side of selected sample units are manually checked and the subjects were determined. If the subject area is same for more than 80 % of sample for each term, then the term is accepted, else added to the list of stopwords. The list of accepted terms and stop words are given in Table 3

Result:

It is clear from the list that the core research area of University of Calicut is mainly Biology, Chemicals, Natural resources, Technology, Astrophysics etc. which belong to Science field. From top 50 repeated terms, other than common terms, numerals (2022, ii) are also considered as stop words. Even though India and Kerala are general terms, they're mostly used in scientific researches, hence included in the list of Accepted terms.

4.2.2 Recent Research Trends of University of Calicut

Methodology:

The research output affiliated to University of Calicut during 2010s and 2020s are considered for identifying research trends. This data is uploaded into Voyant tool. Stop words are excluded and the obtained results are analysed.

From 'Terms' option from Section i, the most frequently used terms in recent research can be listed.

Result:

From Table 4, it can be inferred that the recent trends in research affiliated to University of Calicut is mainly in science area. When we compare the result with the previous result of overall core subject area, the findings are same. That means, the research areas are mostly dominated by recent research.

For example, the term "species" repeated 2015 times throughout six decades, and in last 13 years (2010 – 13/02/2023) is 1525 times. This indicates, for the first four decades from the sample, the term "species" have been used only 490 times. Hence, even though the term species was used from the very beginning, it was studied in-depth in recent decades.

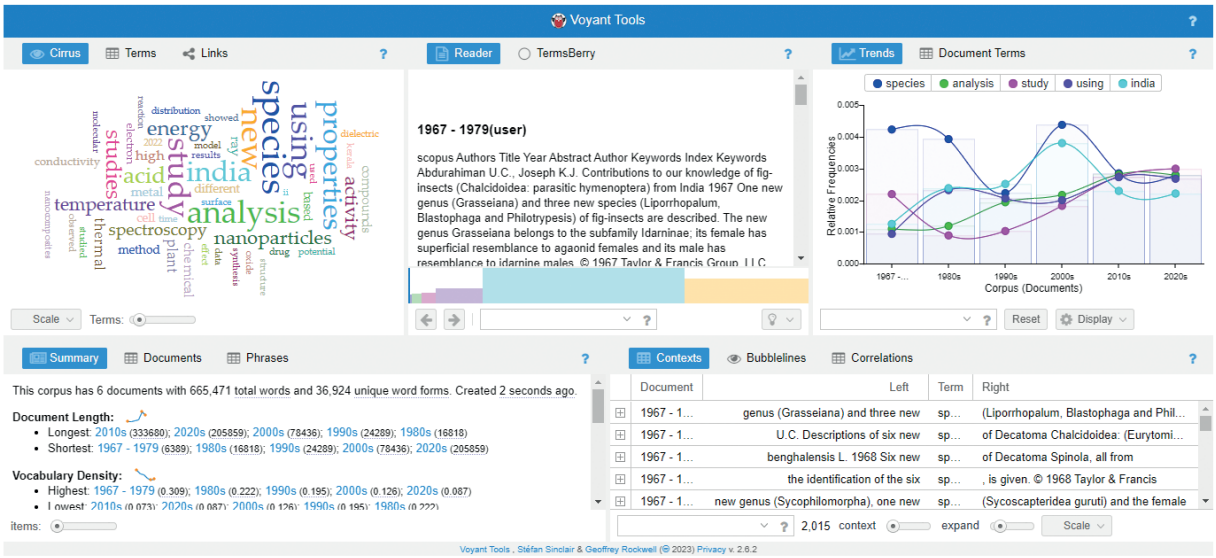

Figure 1. Result page of Voyant tool.

Figure 2. Word cloud representation of most repeating 50 terms.

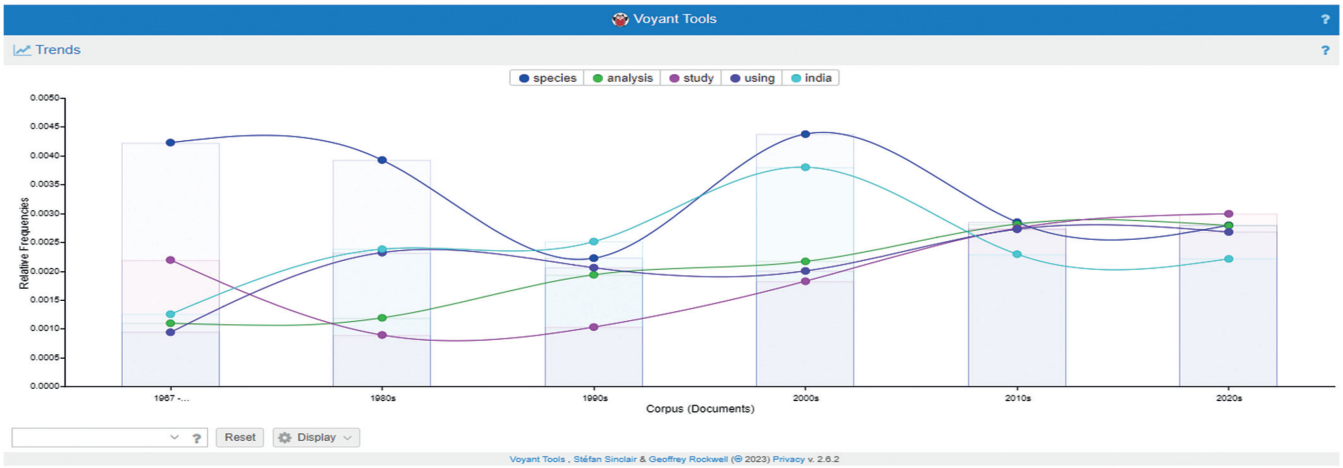

Figure 3. Graphical representation of most repeating 50 terms.

Table 2. Sample size selection method

Species	Step	Sample size
<1980s (27)	$\frac{5}{100}$ X 27	1
1980S (66)	$\frac{5}{100}$ X 66	3
1990s (54)	$\frac{5}{100}$ X 54	3
2000s (343)	$\frac{5}{100}$ X 343	17
2010s (950)	$\frac{5}{100}$ X 950	48
2020s (575)	$\frac{5}{100}$ X 575	29
Total		**101**

Table 3. Accepted words and stopwords

Accepted words	Stopwords
Species	Analysis
India	Study
Acid	Using
Energy	Properties
Nanoparticles	New
Temperature	Studies
Spectroscopy	Activities
Thermal	Ray
Plant	High
Chemicals	Different
Cell	Method
Metal	Based
Conductivity	Results
Compounds	Ii
Electron	Effect
Dielectric	Used
Surface	Showed
Drug	Data
Synthesis	Time
Molecular	Reaction
Model	Structure
Potential	2022
Oxide	Observed
Nanocomposites	Distribution
Kerala	Studied

Table 4. Most used terms in recent research

Terms (2010^{+})	Frequency
Species	1525
India	1220
Nanoparticles	1128
Energy	1015
Acid	985
Temperature	970
Spectroscopy	939
Conductivity	821
Cell	818
Dielectric	774
Electron	773
Chemical	750
Surface	740
Thermal	723
Compound	699
Nanocomposites	691
Plant	671
Molecular	669
Metal	665
Oxide	652
Potential	644
Model	592
Structure	584
Corrosion	582
Drug	576

4.3 Visualisation of Research Productivity of University of Calicut

Methodology:

The total count of research articles published in each year from beginning (1967) to recent (13/02/2023) were noted. The results are plotted in a line graph with Year on X-axis and Number of publications in each year on Y-axis.

The title, abstract and keywords of all research articles published in SCOPUS in each year is saved in separate excel files (year wise), and those file are uploaded to Voyant tools to visualise the results.

Results:

The Trends graph of the uploaded data is shown in Figure 6

This graph shows the research progress in University of Calicut from SCOPUS database. There were not many trending terms during early years, so the term 'species' is prominent during that period. But now, researches are happening in different areas unlike

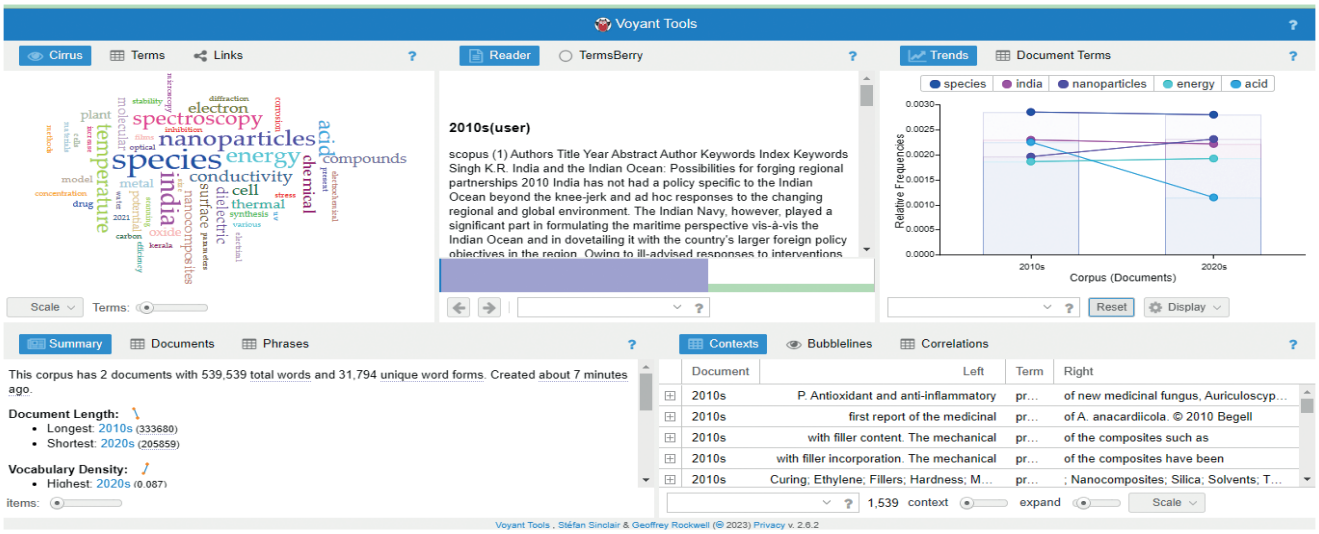

Figure 4. Voyant tool results of research outputs of University of Calicut during 2010s and 2020s.

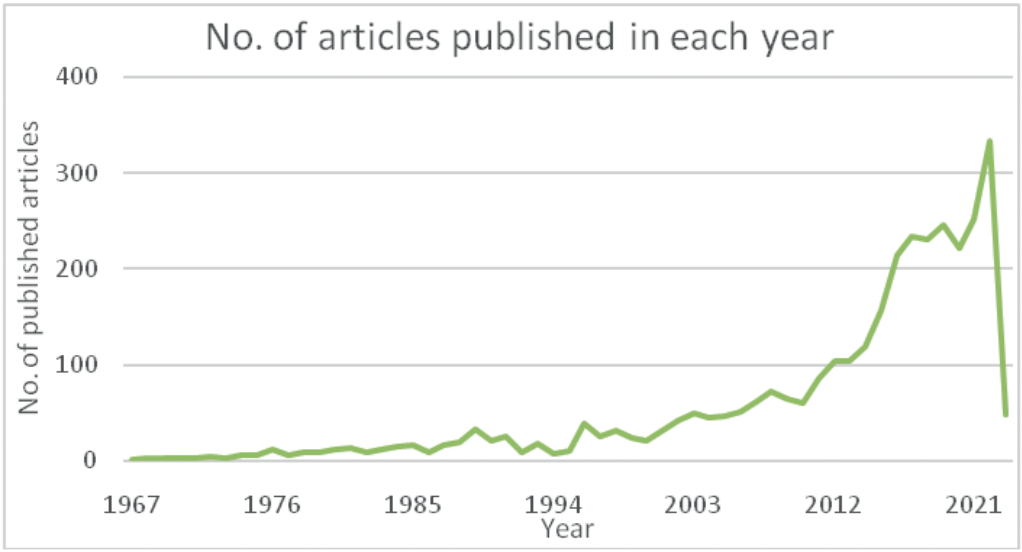

Figure 5. Research articles published in each year under UOC.

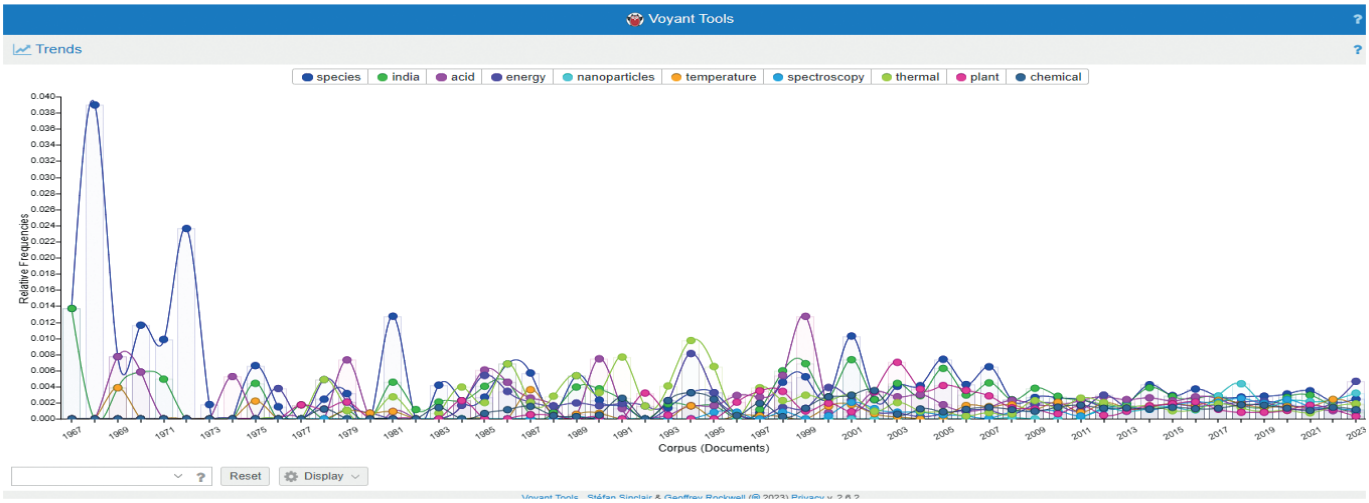

Figure 6. Graph showing trends of research.

before, so there are various terms that are important along with 'species'.

4.4 Identification of the Missing Gap Between the Core Research Areas of University of Calicut and Existing Book Collection in University Libraries

Methodology:

OPAC system of CHMK Library was used to count book collection. Accepted 25 terms are used as keyterm. Number of books appearing for each term is noted. To compare the missing gap, percentages of both Voyant tool and OPAC results are calculated.

- For OPAC, take the percentage of each entry with respect to total number of books appeared collectively for all keyterms.
- For Voyant tool, take the percentage of number of times each keyterm repeated in results retrieved from SCOPUS database, with respect to total number of times all selected 25 terms are repeating

Compare the two percentages to get the missing gap between the core research area and existing book collection in CHMK library

Draw a graph by plotting accepted terms on X-axis, Voyant tool and OPAC percentage results on Y-axis and compare the results.

Results:

After comparing the results from both Voyant tool and CHMK central library book collection, it can be observed that there exists a gap in certain areas of research.

4.5 To Frame a Book Purchase Rule for CHMK Library

Results:

The research publications of University of Calicut from 1967 to 13/02/2023 is 3305. When we interpret these researches using Digital Humanities tool (here Voyant tool), the recent trends of research progress and

Table 5. Books available at CHMK library on core areas of research

Terms	Available no. of books	Frequency
Species	104	2015
India	26167	1627
Acid	218	1274
Energy	479	1178
Nanoparticles	42	1130
Temperature	57	1075
Spectroscopy	352	1006
Thermal	89	965
Plant	2372	957
Chemical	63	902
Cell	1341	893
Metal	296	882
Conductivity	5	858
Compounds	447	855
Electron	1423	845
Dielectric	8	803
Surface	204	796
Drug	304	762
Synthesis	316	755
Molecular	1748	727
Model	19	716
Potential	169	709
Oxide	14	695
Nanocomposites	21	691
Kerala	3466	641

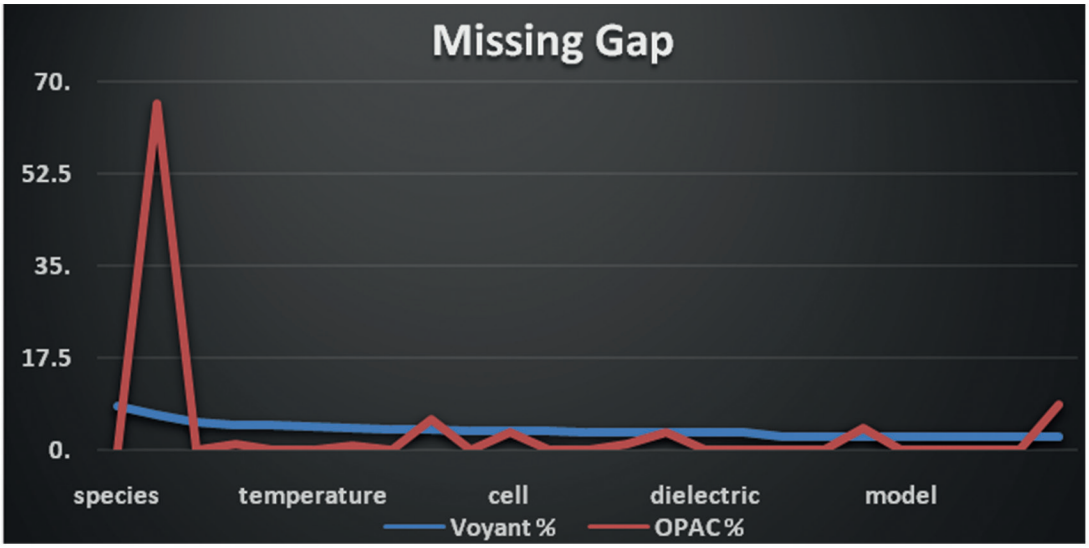

Figure 7. Comparison between core research area and available books in library (in %).

the missing gap in CHMK library can be found. Since CHMK library is the major source of information for researchers under University of Calicut, it is important to provide the users with all required needs of information. Since, the most researches are happening on science field, it's important for the library to update their books and resources in science section. Also, there is a chance that the lack or less researches in humanities fields is due to the unavailability of necessary books from library. So the library must initiate to bring latest books on humanities subjects to encourage the users to do more publications.

4.6 Book Purchase Decision Making Model for Academic Libraries

The following are the major steps followed in a book purchase modal for university of Calicut.

The research outputs are collected using different database (eg; SCOPUS, Web of Science) by using necessary key terms. Printed collection of library can be retrieved by searching in OPAC using criteria and key terms or by manually (for libraries with no OPAC facility). These data are uploaded in any Digital Humanities tools (here Voyant tool) for analysis. Results are collected

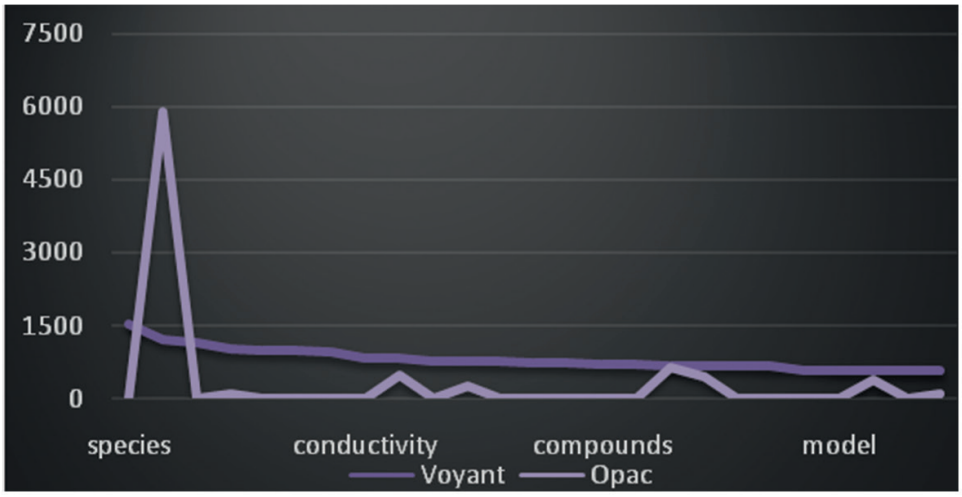

Figure 8. Comparison graph of most used research terms and available books (in counts).

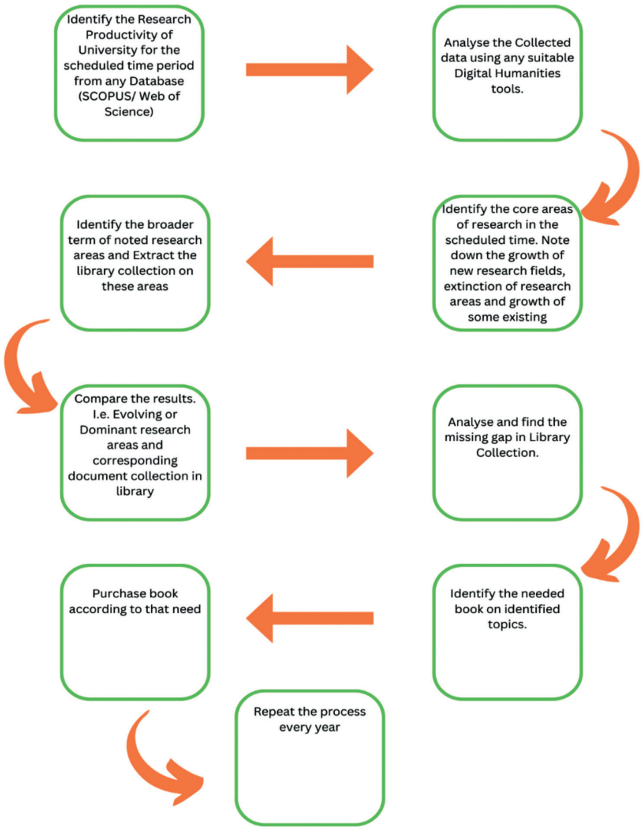

Figure 9. Steps involved in deducing areas to which book to purchase.

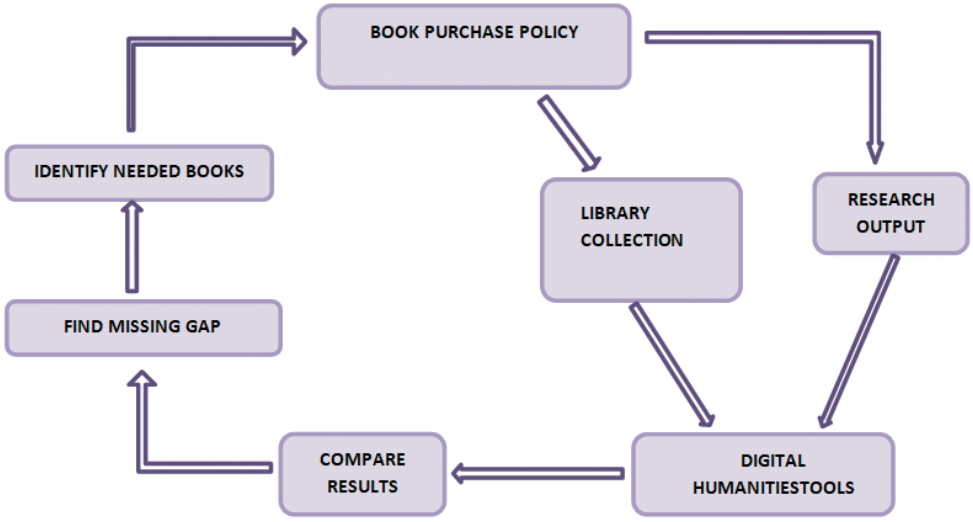

Figure 10. Book purchase decision making model for academic libraries using DH tools.

separately and compared. This is helpful in identifying the difference in collection of a library and research output of corresponding institution. Which means it will help to understand if there exists a stable proportion between the research output and book collection. If not, the library should focus on improving its quality by purchasing books and other resources for improving the proportion between library collection and research publications. This process should repeat once in a year for keeping the library up-to-date.

4.7 Advantages and Limitations of the Book Purchasing Model

The model is a novel idea that reflect in book purchase, is cost and time effective. It is easy-to-understand and further developments are possible

But the model only for academic libraries that focus on research productivity and not for public libraries. The model is based on SCOPUS database only, means the subjects that are not covered by SCOPUS are excluded.

5. FINDINGS AND DISCUSSION

This study was conducted with an aim to propose a book purchasing model for libraries. It is high time the traditional book purchasing methods of libraries are replaced by new methods that consider research outputs of the institution. Collections of an academic library must reflect its parent organisation's research publications. The authors, through this study discuss how to use Digital Humanities tool for this purpose.

The doctoral dissertation by Tignor[18], observed that even though in many libraries librarians won't participate in any projects that are officially conducted as a part of digital humanities, but they do get involved in projects that comes under digital humanities. Zhang and others[19] conducted a study that used digital humanities for knowledge creation. Two open-access products were created by the authors: a bilingual dictionary and an online crowdsourcing translation tool. Giri and others[20], used an observational methodology to determine the appropriate number of copies to purchase books in library collection.

6. CONCLUSION

Digital humanities is an emerging subject. Library and information science and humanities shows equal interest to Digital Humanities (DH). This study proposed a model for purchasing books by taking research output of the institution into account. In order to carry out this study, data of articles published under University of Calicut from SCOPUS was collected. The collected data is analysed using Voyant tool. The analysed data is visually represented using Voyant tool and Excel. From result we got that research fields of University had changed a lot. This paper suggest a model for purchasing books for library. This makes book purchasing a methodological process. This model can be used to increase effectiveness in book purchase. By doing this it will enrich library collection by that there may occur potential changes in research output and enhance research. Digital humanities as a subject is evolving continuously, and by considering the importance it have in many diverse subjects, it have the potential to rise to higher level in research fields, and more applications of the subjects are yet to come.

REFERENCES

1. What are the Digital Humanities. Retrieved February 17, 2023, from https://github.com/binipringle/intro-to-dh/readings/1-definition/.
2. Egerton, F. *Oxford LibGuides: Digital Humanities: Home*. Retrieved February 17, 2023, from https://libguides.bodleian.ox.ac.uk/digital-humanities/home.
3. Sula, C.A. Digital humanities and libraries: A conceptual model. *Journal of Library Administration*, 2013, **53**(1), 10–26. Doi: 10.1080/01930826.2013.756680.
4. Zhang, J.; Wang, Y.; Shi, M. & Wang, X. Factors driving the popularity and virality of Covid-19 vaccine discourse on Twitter: Text mining and data visualisation

study. *JMIR Public Health and Surveillance*, 2021, **7**(12).
Doi: 10.2196/32814.
5. Idler, E.; Bernau, J.A. & Zaras, D. Narratives and counter-narratives in religious responses to COVID-19: A computational text analysis. *PLoS ONE*, 2022, **17**(2 February).
Doi: 10.1371/journal.pone.0262905.
6. Lu, S.; Zhao, L.; Lai, L.; Shi, C. & Jiang, W. How do Chinese people view cyberbullying? a text analysis based on social media. *International Journal of Environmental Research and Public Health*, 2022, **19**(3).
Doi: 10.3390/ijerph19031822.
7. Grandjean, M. A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts and Humanities*, 2016, **3**(1).
Doi: 10.1080/23311983.2016.1171458.
8. Elo, K. A text network analysis of discursive changes in German, Austrian and Swiss New Year's speeches 2000-2021. *Digital Humanities Quarterly*, 2022, **16**(1).
9. Mlambo, R.; Matfunjwa, M. & Skosana, N. Contrastive analysis of word-formation strategies in the translated South African Constitution. *Literator- Journal of Literary Criticism, Comparative Linguistics and Literary Studies*, 2022, **43**(1).
Doi.org/10.4102/lit.v43i1.1861.
10. Asarta, C.J. & Méndez-Carbajo, D. Bringing down the walls in business and management education: A textual analysis of economic education articles. *J. Education for Business*, 2021, **96**(1), 10–21.
Doi: 10.1080/08832323.2020.1726268.
11. Anson, I.G. & Moskovitz, C. Text recycling in STEM: A text-analytic study of recently published research articles. *Accountability in Research*, 2021, **28**(6), 349–371.
Doi: 10.1080/08989621.2020.1850284.
12. Bernau, J.A. Text Analysis with JSTOR Archives. *In* Socius- sociological research for a Dynamic World, 4, SAGE Publications Inc., 2018.
Doi: 10.1177/2378023118809264.
13. Fallucchi, F.; Ghattas, B.; Spielhaus, R. & de Luca, E.W. Digital qualitative and quantitative analysis of Arabic textbooks. *Future Internet*, 2022, **14**(8).
Doi: 10.3390/fi14080237.
14. Wang, Z.; Wu, J.; Yu, G. & Song, Z. Text analysis and visualisation research on the hetudangse during the qing dynasty of china. *In* Information Technology and Libraries 2021, **40**(3). American Library Association.
Doi: 10.6017/ital.v40i3.13279.
15. Ullah, Z.; Uzair, M. & Mahmood, A. Extraction of key motifs as a preview from 2017 nobel prize winning novel, "Never Let Me Go": An interactive word cloud study. *Lecturer in English*, 2019, **7**(2).
16. Basak, M. & Roy, S.B. Mapping the literature on digital humanities: A bibliometric study using Scopus data. *DESIDOC J. Libr. Inf. Technol.*, 2022, **42**(6), 354–363.
Doi: 10.14429/djlit.42.6.18271.
17. Sakkaravarthi, K. (2013). *Book Selection Principles—Librarians Portal*. Book Selection Principles. https://newonlinelibrary.blogspot.com/2013/10/book-selection-principles.html?m=1.
18. Tignor, M. *Digital humanities & community college libraries: Opportunities and challenges*, 2022. https://stars.library.ucf.edu/etd2020/1300.
19. Zhang, Y.; Xue, S. & Xue, Z. From collection curation to knowledge creation: Exploring new roles of academic librarians in digital humanities research. *J. Academic Librarianship*, 2021, **47**(2).
Doi: 10.1016/j.acalib.2021.10232.
20. Giri, R.; Sen, B.K. & Mahesh, G. Collection development in Indian academic libraries: An empirical approach to determine the number of copies for acquisition. *DESIDOC J. Library Inf. Technol.*, 2015, **35**(3), 184–192.
Doi: 10.14429/djlit.35.3.7806.

CONTRIBUTORS

Ms Aysha Zakhiya A. is a PhD student at Department of Library and Information Science, University of Calicut, Kerala. Her areas of interest include: Human information behaviour, Digital humanities and Social media
Her contribution for the present work are: Developed the idea of research, Installation of the software and contributed to the collaborative writing and drafting of the manuscript.

Ms Aparna K.S. is a postgraduate student at Department of Library and Information Science, University of Calicut, Kerala. She is interested in social media metrics and sentiment analysis. Her contribution for the present work are: Collaborative collection and analysis of data, valuation, and collaborative writing and drafting of the manuscript and addressing the revisions.

Dr Syamili C. is currently an Assistant Professor at Dept. of Library & Information Science, Calicut University. She holds a Master's degree from Calicut University and a Doctoral degree from Pondicherry Central University. She has published several papers in leading Journals and Conferences of International Repute. Much of her research focuses on semantic web, ontology, indigenous knowledge preservation and altmetrics. She for the present work, conceived the idea of developing the proposal and designing the methodology, constructive feedback, reviewing and editing.