

Meta-Analysis of Anita Desai's Novels: A Comparative Study of Digital Humanities Tools for Topic Modelling

Thamil Selvi V.¹, Vaishnavi B.²

Department of English, Dhanalakshmi Srinivasan University, Samayapuram, Tiruchirappalli, India.

E-mail: ¹thamilselviv.set@dsuniversity.ac.in, ²vaishnavib.phd2023@dsuniversity.ac.in

Abstract

Digital Humanities, being an integration of computing and humanities, has transformed literary text analysis through the power of making textual analysis fast and accurate. The option of meta-analysis of an author's writings within a given time period has become possible. Researchers are now able to view the big picture and establish new relationships between a given novel and among all the novels of a specific writer, something that was not even feasible earlier. By using proper tools when analyzing data on the internet, researchers, particularly those from the department of humanities, are able to perform quality research without any fear of time deficiency or any fabricated data gathered. Topic modeling is also among the digital humanities tools that search over massive data to identify and compare underlying themes, stylistic and linguistic patterns, and narrative forms using computational approaches such as Latent Semantic Analysis (LSA). This paper is a comparison between MALLET (MACHINE Learning for Language Toolkit) and Orange, the two topic modeling packages. The two tools are compared based on ease of use and feature support in text mining for theme evaluation in the selected novels of Anita Desai, an Indian feminist writer. The input to be mined is the selected novels of Anita Desai. The paper examines how digital humanities tools like text mining, sentiment analysis, and word frequency analysis can be utilized by researchers to simplify the analysis process and make it more accurate. The objective of the research is to analyze Anita Desai's novels to understand the predominant themes in her works. Subsequently, the findings of the paper suggest that MALLET software is more efficient than the Orange Data Mining Tool. By employing experimental methods in humanities research,

this investigation presents an avenue for a unifying research partnership between computer scientists and humanities scholars.

Keywords: Digital Humanities, Anita Desai, Textual Analysis, Topic Modelling, Text Mining.

1. Introduction

Digital Humanities is transforming the work of classical humanities research by integrating computational means into traditional disciplines. When coupled with literature and digitalization, scholars can compute large volumes of information, graphically represent elaborate patterns, and form links that would be inaccessible through close reading otherwise. The digitalization of books allows us to utilize computational tools like topic modeling to investigate themes, authorship, generic transformation, and trends in the humanities. Digitalization forces us to reconsider the creation, classification, and distribution of knowledge through digital conduits. Interdisciplinary methods based on instruments provide quantitative analysis and qualitative insight. Computer humanities have simplified work to benefit society by automating and conserving time in addressing responsibilities. Literature is humanity's repository of information. Literature is a collection of human feelings and experiences, dispersed across various timeframes. Speed and accuracy are what computer humanities introduce to literature. Literature and computational studies are twin fields that require significant bridging. While the two disciplines would otherwise rely on one another, the intersection of sciences and humanities can be contentious. The novelty of this paper arises from a thematic analysis of 11 novels using topic modeling techniques under a methodological approach. Statistics and Latent Dirichlet Allocation (LDA) are employed in the essay as the primary research methods, rather than the traditional close reading method. Topic modeling is a technique primarily used in sentiment analysis and trending topic extraction for social media tracking. Although topic modeling has been utilized in the humanities as a new wave of research, it is still in its early stages. Topic modeling is a human-made model of natural language processing that has matured into a sophisticated method for discovering shared themes in a document corpus. As in the paper "A Comprehensive Overview of Topic Modelling: Techniques, Applications, and Challenges," topic models are developed with the promise of predicting a set of documents' latent topics under two light assumptions: (i) documents in a collection are represented as bags of many topics, and (ii) a topic is represented as a word distribution. That is, a document is a distribution over topics of a given size, and the

topics are a distribution over words of probabilities. [1] The topic modelling method hence seems to be a very resilient approach in literary work analysis, where thematic research is the cornerstone of scholarly research. Unscrambled from close reading, topic modeling supports a fresh kind of distant reading that examines hundreds, even thousands, of texts in the aggregate to uncover thematic dynamics, conceptual bundles, and even stylistic changes over the *longue durée*. It does not attempt to replace the close reading of an individual text but to supplement it with a bird's-eye view of a literary corpus that can be tested on finer points with the conventional method once identified. MALLET (MACHINE Learning for Language Toolkit) and the Orange data mining tool are applied to topic modelling in this instance. MALLET is an open-source, free language classification, clustering, and topic modeling toolbox. Andrew McCallum started developing the project with assistance from some professors and University of Massachusetts Amherst graduate students. MALLET is maintained by David Mimno, who also assisted with the topic modeling package. MALLET employs the Latent Dirichlet Allocation (LDA) technique to identify thematic hidden structures in a document corpus. MALLET is quite fast and delivers highly coherent topic distributions. MALLET is written in Java and Command Prompt syntax. The reliability and effectiveness of MALLET have ensured that it is the go-to tool for digital humanities researchers, particularly those interested in decomposing thematic trends from literary or historical corpora. Orange is a highly used data mining tool. It is GUI-based and allows users to define workflows in a drag-and-link format by drawing and linking widgets representing different analysis operations. Orange Software has also been utilized to educate clustering, classification, and text mining for technically non-expert students in order to close the gap between computation and interpretation. Orange enables easy text analysis, data trend visualization, and interactive literary analysis pipelines. Literary analysis using Digital Humanities has added potential for cultural background and text interpretation. It is not a fashion shift or theory turn but a philosophical transformation in humanities knowledge production and communication. It is a transformation constantly remapping the terms and limits of literary scholarship in the age of the digital network. Scholars can capitalize on literature as not just discontinuous works of art but nodes in vast linguistic, historical, and social webs. With topic modeling methods applied in literary study, the writer can emphasize more authorial voice, intertextuality, author or genre invention, and thematic analysis. It provides new ways of investigating canonicity, authorship, and the representation of a book or author. Orange and MALLET are some of the software through which researchers can conduct data-driven examinations on a brushstroke level without compromising the depth

of literary scholarship. This paper aims to compare the efficiency of MALLET and the Orange Data Mining Tool in the thematic analysis of Anita Desai's novels using topic modeling.

Problem Statement

The acceleration of cyber information and the increased availability of means of computation have sped up changes in the world of the humanities. However, despite digital techniques becoming ubiquitous across humanities research, there is still a colossal gap between knowing and applying these computational tools. Not every tool can be utilized effectively by researchers from outside the Computer Science field, as most of them require expertise in programming languages such as Java and Python. This paper attempts to bridge the gap through the analysis of two computer programs, MALLET and Orange Data Mining Software. It examines their functionality and usability in the thematic analysis of the novels of Anita Desai.

2. Literature Review

Digital Humanities then emerged as the flagship interdisciplinary program at the confluence of humanistic inquiry and computational models. Digital Humanities, Helmi speaks about in his "Digital Humanities: A Paradigm for the 21st Century," "appeals to the visual. New research abilities can thereby be acquired and applied, new literacies are needed, and new modes of representing/visualizing/narrating the findings are called for." DH is the future, and the humanities allow us to understand our world, but DH has a crucial role to play in our understanding of our living in and contribution to the digital world. [2] Fangli Su and Shihao Li reflect on the humanities' contribution to Digital Humanities and include, "Humanities is the center of digital humanities." Without the humanities, studies on digital humanities will be baseless. Therefore, training and practice in digital humanities should provide problem-solving strategies for humanities researchers conducting research in the digital age.

[3] Deny Yadava's "The Role of Artificial Intelligence in Literary Analysis: A Computational Approach to Understand Literary Styles" introduces computational literary analysis as an application and is optimistic: "AI can effectively deal with large corpora, identify repeated patterns, and offer a quantitative basis for analysis." The quantitative dimension adds to traditional literary scholarship and broadens the scope of study and choices for the scholar. The objectivity and computability of AI also eliminate human biases. [4] Dr. Preethi states in "Quantitative Analysis of Literary Texts: Computational Approaches in Digital Humanities

Research" that "Not only do the new computational approaches promise to reach deeper into subtleties of text information, but they also reveal patterns, trends, as well as emotional subtleties that hitherto were not available through the so-called conventional qualitative approaches alone." [5]

Among the Digital Humanities philosophy problems is how to handle and define uncertainty. Panagiotidou et al. (n.d.) provide the concept and presentation of uncertainty in Digital Humanities research. For them, "In digital humanities research, ambiguity is acceptable because it is something to hold onto. Taxonomies of ambiguity visualization are ignored or recognized as uncertainty markers not actively managed [14, 65]." Interpretivist paradigms socialize humanists; however, these will include partial, indeterminate accounts of information and knowledge [38]. The flexibility secret, they promise to their process [6]. In the preservation and reuse of data, Asselborn et al. (n.d.) discover more and more sought-after flexible, lightweight information systems in terms of facilitating humanities research data long-term reuse. They also state, "The successful testing and application of this method in the EDAK project proves its ability to rationalize the development process and generate the necessary scalability in the humanities and beyond." These effective EIS solutions are adopted by organizations or businesses so that they are able to function efficiently and bring efficiency to their research or business operations [7]. In a similar manner, Hawkins recommends, "Linked Data offers a low-barrier route to making digitised and born-digital archives more useful, to creating integrated, enriched, and interoperable large-scale archival datasets for reuse across multiple modes. Publication of archival metadata as Linked Data has been the most prevalent area of Archival Linked Data activity to date. But it is the content of digitised and born-digital documents that holds most use for researchers, and hence it is more important than ever that archive services datafy their collections" [8].

Web page and social media topic modeling is also plagued by issues. Caitlin Doogan, Poet Laureate, in the paper "A systematic review of the use of topic models for short text social media analysis," finds model construction to be unsuitable for use in the real world for short texts like social media. Their own critical examination of 189 studies indicates a need for better guidelines and improved communication between users in the field and computational designers to bring about productive results [9]. Shang, Ma, and Moulaison-Sandy (n.d.) are trying to find out this question through a study on AI representation in Digital Humanities scholarship. In their bibliometric study, "Algorithmic processes have changed the face of DH

research, with researchers seriously considering the advancements in AI technology. That is a very bold statement when it comes to AI adoption research in DH" [10].

Recent research investigates the broader intersection of generative AI and humanities scholarship. Priyanka Gupta et al. in "Generative AI: A Systematic Review Using Topic Modelling Techniques" performed a systematic review using BERTopic to examine over one thousand GAI articles and found top themes such as content generation, privacy, and the educational use of GPT. Their work needs further innovation, particularly in explainability, data ethics, and multi-modal AI systems.[11] Eze and Shamir (n.d.) also critically analyze the abusive use of generative AI, focusing on phishing emails generated by AI. Their paper includes a hand-curated dataset release on such texts and provides benchmark detection using machine learning, finding that emails generated by AI have weak linguistic signals. This paper places the highest emphasis on using proper detection frameworks to defend against cyber attacks involving AI.[12]

"The 'Mallet vs GenSim: Topic Modelling Evaluation Report'" by Islam and Arango compares the performance of the two most widely used topic modeling tools against the 20 Newsgroups dataset. MALLET is commended for its ease of installation and use but criticized for the absence of a parameter adjustment feature. Meanwhile, the open-source Python library GenSim is ostensibly more complex but requires superior programming skills. Their comparison demonstrates how data pre-processing and parameter choice can affect topic model output. This study is useful in steering research emphasis among researchers toward technical proficiency and need-based tool selection. [13] Similar to the research on the Orange data mining application software by Zdena Dobešová, this study is also education-oriented, under the guise of geoinformatics. The GUI and support provided by Orange for GIS programming using software like ArcGIS Pro create an effective learning environment for users with lower-level programming skills. While less interactive than text-conditioned software like GenSim, Orange is pedagogically solid and more suited to user interface needs. The research justifies the use of visual data mining software in enabling machine learning across disciplines, particularly in hybrid discipline spaces like digital humanities crossovers.[14]

Outside the literary humanities, Matthias R. Fengler and Tri Minh Phan apply the new paradigm of topic modeling to earnings releases i.e., the MD&A section of 10-K filings. Using word-embedding techniques in a bid to find sentimentally based topics and thereafter proceeding to add onto them temporality and sentiment, the study, done in the guise of a

business model, validates the higher expressiveness of topic modeling and how literary critics would like to analyze trends of sentiment, shifts of discourse, or even genre evolution in corpora.[15]

Ning Wang outlines the epistemological transformation towards digital humanities in literary studies. Emphasized is the far reading, and Wang brings in computationally informed methods as a supplement, not a replacement, for close reading. The essay seeks the confluence of humanist investigation and data science, and also challenges academics to make use of electronic media as a means to widen their analytical vision.[16] One scholar who continues this argument is Gunther Martens, in an essay on how literature, or, for that matter, encyclopedic forms react and respond to technological advancements. His treatment of Robert Musil's *The Man Without Qualities* returns the digital humanities to the tradition of traditional knowledge patterns, i.e., encyclopedic and intertextual relationships. Martens' study argues that literacy theory must be broadened to encompass electronic authorship and connectivity on a continuum, rather than viewing it as a break in literary history.[17] The scholarly monograph *Vita Ædwardi Regis* by Dockray-Miller, Drout, et al. is the best example of the recuperation of lost information on matters of authorship and patronage via the digital. Enabled by Lexomics and other computer programs for text analysis, the article exemplifies the multi-authorship of medieval text production and Queen Edith's patronage. The study attests to Digital Humanities practice exposing power structures and structural gendered agency in medieval textual culture, ascribing a single meaning of authorship and queering the medieval canon to include non-dominant texts.[18] Finally, Josh Brown's article places on center stage the less contentious issue of linguistic diversity in Digital Humanities. The multilingual, hybrid structure of previous languages is charted in a taxonomy by Brown in the case study *Dictionnaire de la langue franque*. He avoids the Anglocentrism that characterizes most Digital Humanities efforts and provides examples for the employment of hybrid linguistic material. It is a contribution to broader debates regarding methodological pluralism and participation in digital humanities.[19]

MALLET Workflow

MALLET is a JavaScript topic modeling tool and therefore needs to be installed before actually downloading the software. The MALLET program is an open-source application and can be freely downloaded from the website. Latent Dirichlet Allocation (LDA) was used to generate 10 topics for topic modeling. The data files should be in .txt format so that MALLET

can read them. Once the PDF documents were converted to .txt format, they were saved on the C drive in the mydocs directory of MALLET. They were then run through the following input code for topic modeling:

```
cd C:\mallet
```

```
bin\mallet import-dir --input mydocs --output topic-input.mallet --keep-sequence --remove-stopwords
```

```
bin\mallet train-topics --input topic-input.mallet --num-topics 10 --output-state topic-state.gz --output-topic-keys topic-keys.txt --output-doc-topics doc-topics.txt
```

The output generated was saved using the following Command

```
--output-topic-keys topic_keys.txt
```

```
--output-doc-topics doc_topics.txt
```

Figure 1 shows the output generated by the MALLET Software

```

0      0.5  deven nanda raka nur kaul murad Ilah House poetry das chapter siddiqui poet college ram lal don't tape urdu door
1      0.5  nirode gautama don't amla ing jit dharma sonny david monisha love it's calcutta air long i'm rain aunt house garden
2      0.5  eyes long white hand left dark stood hands life great mother hair till made told good looked black open trees
3      0.5  face night turned small voice began felt man young heard father knew work give light asked world rose find evening
4      0.5  bim tara raja aunt baba don't bakul veranda mira hyder girls bed it's door garden bim's ali tara's can't family
5      0.5  matteo sophie laila krishna mother india ashram dance don't hands love woman door light indian air master small isabel water
6      0.5  eric family asked men company woman town find brought art light man village mexico appeared open stones table people told
7      0.5  back head made room time make day house home feet found put children water child thought hear making red past
8      0.5  uma sarah adit dev mama mrs arun indian papa india patton tea aruna die door miramasi emma tlie don't england
9      0.5  baumgartner hugo boy lotte camp war farrokh flat german cats table door men don't knew baumgartner's found felt chimanlal
india

```

Figure 1. Showing MALLET Key Topics Output

The --output-topic-keys command topic_keys.txt output 10 topics, and one of them is a collection of words which occur together or are similar to one another in all 11 topics. All of the words in a topic describe a specific theme. The topic label of one of them is as stated below:

With this data, the themes of Anita Desai's novels can be identified. Also, based on the percentage distribution of these themes in a novel, we can compare the dominant themes of each novel. The above output provides the percentage distribution of each theme in a novel.

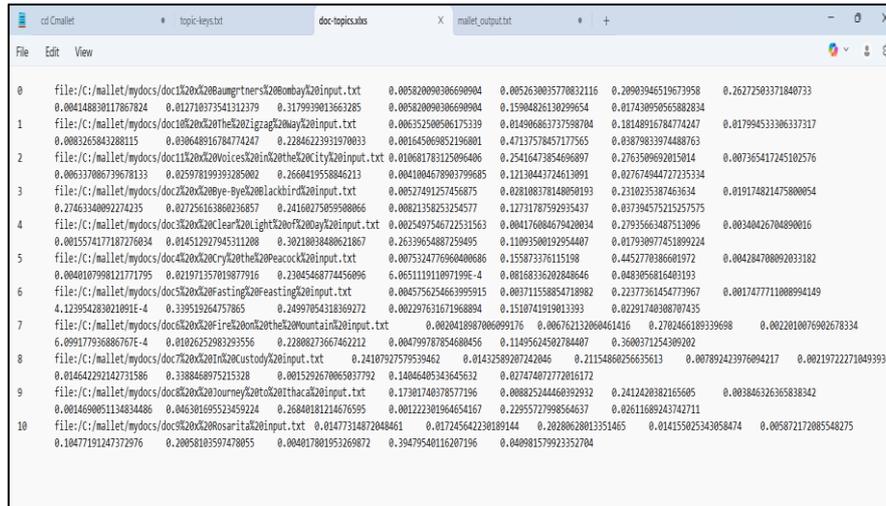


Figure 2. Showing MALLET Topic Distribution Output

For a better understanding of the output data, an Excel worksheet was used to decode and transcribe the data. The table below shows the distribution of each topic in the novel along with the percentage of usage. Using this data, recurring themes in the corpus of Anita Desai’s novel can be analysed.

Table 1. Showing Topic Distribution by MALLET Software

Topic Distribution by MALLET Software										
	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
Baumgartner's Bombay	1%	1%	21%	26%	0%	1%	32%	1%	16%	2%
The Zigzag Way	1%	1%	18%	2%	1%	3%	23%	0%	47%	4%
Voices in the City	1%	25%	28%	1%	1%	3%	27%	0%	12%	3%
Bye-Bye Blackbird	1%	3%	23%	2%	27%	3%	24%	1%	13%	4%
Clear Light of Day	0%	0%	28%	0%	0%	1%	30%	26%	11%	2%
Cry, the Peacock	1%	16%	45%	0%	0%	2%	23%	0%	8%	5%
Fasting, Feasting	0%	0%	22%	0%	0%	34%	25%	0%	15%	2%
Fire on the Mountain	0%	1%	27%	0%	0%	1%	23%	0%	11%	36%
In Custody	24%	1%	21%	1%	0%	1%	34%	0%	14%	3%
Journey to Ithaca	17%	1%	24%	0%	0%	5%	27%	0%	23%	3%
Rosarita	1%	2%	20%	1%	1%	10%	20%	0%	39%	4%

Orange Workflow

Similar data can be generated using Orange Software. The Orange Data Mining tool [21] requires the Python programming language to function within the system. After installing Python, the Orange tool was downloaded from the site. The software is executable, and there is no need for any coding language. Researchers can utilize it with a simple drag-and-drop method. After installing Orange Software, the Text Mining add-on was downloaded. There are

various widgets in the software. With the help of the connecting widgets, different data mining operations are possible. Since Orange is capable of reading PDFs, data pre-processing to transform the file type was not performed. The text tool was utilized for topic modeling and the following widgets were selected and configured to receive output.

Import Documents → Process text → Topic Modelling → Data Table → Save Data.

The following image shows the workflow created in the Orange Software

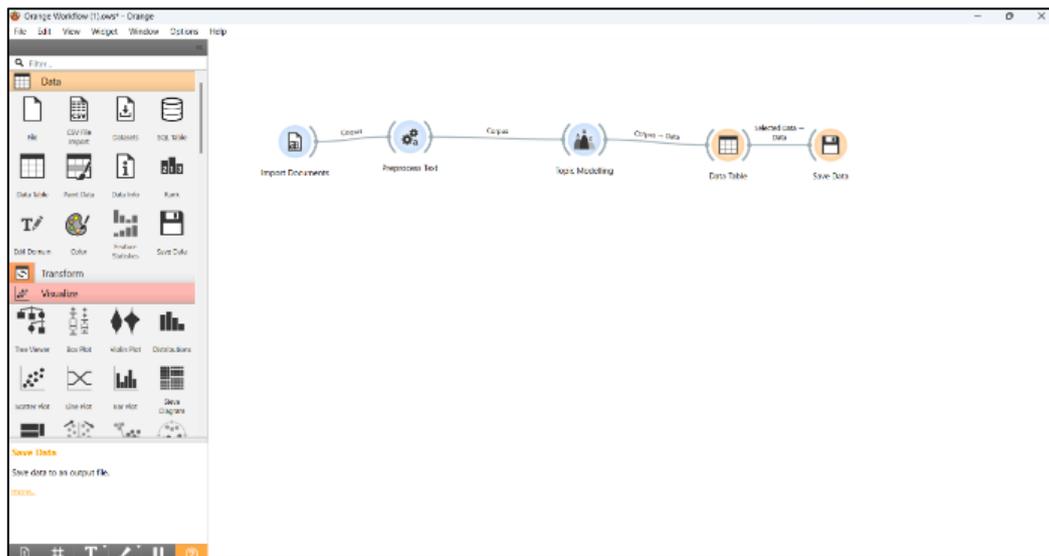


Figure 3. Showing Orange Workflow -1

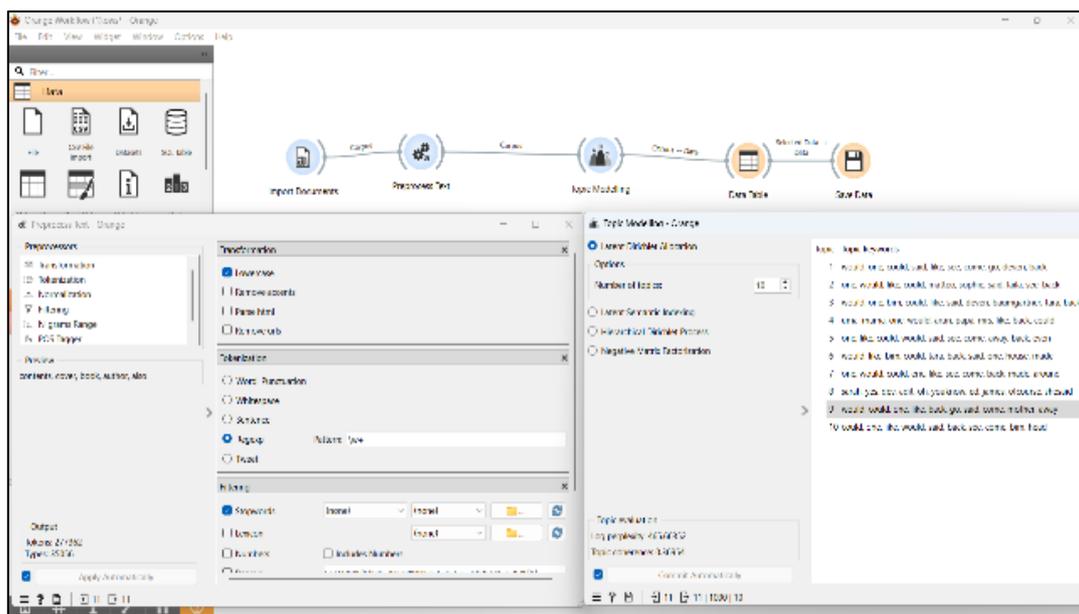


Figure 4. Showing Orange Workflow -2

Orange output can be exported directly to an Excel sheet. The table below shows the output offered by Orange Software. Data can be displayed both in the orange palette and in an Excel workbook. The books "Cry the Peacock" and "Voices in the City" were scanned books and, therefore, read as '0' because it was not possible for Orange Software to read from them.

Table 2. Showing Topic Distribution by Orange Data Mining Software

Name	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
Baumgartners Bombay	3817.97	-201.092	25.5778	-155.371	-43.058	-69.1801	-7.16632	23.727	1.13992
Bye-Bye Blackbird	37.8032	5.07991	-5.78464	-9.6536	-1.6786	1.5675	1.81649	1.2466	-0.7205
Clear Light of Day	4602.57	246.519	-262.991	-7.39947	92.4187	-23.6051	-5.29066	-0.48177	1.55599
Cry the Peacock	0	0	0	0	0	0	0	0	0
Fasting, Feasting	2909.79	-77.297	-44.1332	-17.8556	-55.2197	119.083	88.0048	4.49682	1.66599
Fire on the Mountain	2270.04	-59.9058	-0.78163	-67.6837	81.0823	-89.4314	-74.3578	-55.359	-3.79817
In Custody	3459.84	-263.852	-85.1811	133.812	-102.636	-243.936	-321.575	-25.1885	-0.98245
Journey to Ithaca	5898.75	233.828	232.737	-37.3356	-51.3868	-5.94338	-2.69202	-1.75128	-0.32987
Rosarita	1053.91	-90.2557	40.2234	42.4486	46.3597	54.8032	45.3623	85.1124	4.48062
The Zigzag Way	2627.25	-144.107	42.6379	34.7939	145.297	-43.2463	66.7409	-10.874	-1.97478
Voices-In the City	0	0	0	0	0	0	0	0	0

3. Methodology

This research uses the comparative method to establish the performance, usability, and results of two topic modeling programs: MALLET (MACHINE Learning for LANGUAGE Toolkit)

and Orange Data Mining Software. The two packages were installed on the system to test them. The installation process and functions of MALLET and Orange are easy for beginners compared to other software that involves intricate coding. The purchase of a programming language poses a problem for Digital Humanities researchers; MALLET and Orange are beginner-friendly software for Digital Humanities. Ease of installation is a significant factor in software choice. Beyond this, the comparison is organized based on preprocessing in the software, usability, output quality, and interpretability. The main aim of the study is to gain an understanding of the operation of topic modeling tools in the literary analysis of novels. The novels of Anita Desai were compiled, processed, and uploaded as corpus data.

Compilation: A list of 11 novels was compiled. The books comprised Baumgartner's Bombay,[22] The Zigzag Way,[23] Voices in the City,[24] Bye-Bye Blackbird,[25] Clear Light of Day,[26] Cry, the Peacock,[27] Fasting, Feasting,[28] Fire on the Mountain,[29] In Custody,[30] Journey to Ithaca,[31] and Rosarita.[32]

Processing: The data were processed according to the software package specifications. To prepare the data to be analyzed using MALLET, the data were converted into plain text files. In the case of the Orange Data Mining Tool, it was converted into a PDF file. The correctness of the results generated using the software relies on the efficient removal of stopwords. Stopwords are high-frequency words that continuously repeat themselves, like "a," "the," "and," "an," etc., and can destabilize the output of the data. When words tend to occur together with high frequencies, the machine learning algorithm of the topic modeling tool perceives them to be meaningful and groups them under a single topic. If stop words aren't eliminated, they become interference or semantic noise.

Analysis: Both software outputs were decoded, transcribed, and visualized with the help of an Excel workbook. Raw data presented by the software were translated into percentages, and generated keywords were analyzed in order to determine the dominating theme based on the keywords. The data were graphically represented in MS Excel in the form of a stacked column graph and a line graph.

4. Findings and Discussion

The research compared the performance of MALLET and Orange software in supporting the thematic analysis of the novels by Anita Desai. Upon rigorous analysis of the

output from these software programs, it was determined that MALLET is relatively easier to use and provides authentic data. The primary reason behind this is that MALLET contains an inbuilt coding that automatically eliminates stopwords before presenting the output. In Orange, a text file containing stopwords needs to be added to the preprocessing text widget. Even after uploading the file, the output provided is not trustworthy. This can be seen by looking at the topics provided by both software programs. The following table includes keywords and topic names provided by both MALLET and Orange.

Table 3. Showing Top Keywords by MALLET Software

Top Keywords by MALLET		
Topics	Key Words	Topic Labels
0	devan, mateo, sophie, laila, mur, jashima, musad, ashram, don't, poet, dance, siddiqui, india, poetry, face, made, college, small, loud, tape	Literary Characters & Poetic India
1	rirode, dorit, gautama, amita, rje, made, rit, pharm, sorry, david, monisha, city, youth, cigarette, rain, love, its, aunt, i'm, calcutta	Modern Youth & Urban Philosophy
2	long, face, wife, eyes, life, voice, mother, night, thought, still, great, small, felt, father, air, room, head, sat, world, and	Personal Emotions & Inner Life
3	baumgartner, hugo, made, boy, lotte, camp, table, man, business, farrokh, cats, german, war, police, baumgartner's, men, chinmamal, street, don't, fat	War Memory & Diaspora Identity
4	sarah, adit, dev, indian, india, die, mrs, lea, emma, london, life, english, don't, england, roscommon, james, british, dear, man, country, green	Colonial Cross-Culture & Identity
5	uma, mama, mrs, papa, asra, don't, patton, aruna, family, water, stands, eat, miramasi, made, room, cook, school, tars, dinner, parents	Family Roles & Domestic Life
6	back, time, head, hands, house, left, looked, stood, dark, began, door, hand, knew, make, heard, away, people, beneath, good, day	Reflections on Time & Daily Life
7	bim, tara, raja, aunt, baba, don't, veranda, babu, made, mira, hyder, birds, room, garden, it's, bim's, all, tara's, school, had	Sisters & Childhood Memories
8	made, brought, tale, family, found, maisie, young, find, woman, children, got, told, left, city, fear, arise, cried, child, brine, male	Family Structure & Nurturing
9	randa, raja, laxmi, house, asia, child, chapter, das, made, rani, wind, lal, trees, garden, hear, sky, ate, summer, veranda, children	Women & Nature in Domestic Spaces

Table 4. Showing Top Keywords by Orange Software

Top Keywords by Orange		
Topic	Keywords	Suggested Topic Label
1	would, one, could, said, like, see, come, go, deven, back	General Actions and Dialogue
2	one, would, like, could, matteo, sophie, said, laila, see, back	Character Interaction (Matteo & Sophie)
3	would, one, bim, could, like, said, deven, baumgartner, tara, back	Character Interaction (Tara & Deven)
4	uma, mama, one, would, arun, papa, mrs, like, back, even	Family and Relationships
5	one, like, could, would, said, see, come, away, back, even	General Movement and Conversation
6	would, like, bim, could, tara, back, said, one, house, made	Home and Personal Spaces
7	one, would, could, eric, like, see, come, back, made, around	Narrative and Surroundings
8	sarah, yes, dev, adit, oh, youknow, ed, james, ofcourse, she said	Casual Talk and Named Characters
9	would, could, one, like, back, go, said, come, mother, away	Separation or Departure
10	could, one, like, would, said, back, see, come, bim, head	Thoughts and Perceptions

The keywords provided by MALLET are more precise and extensive, encompassing the entire corpus of texts. By sorting the themes based on the percentage distribution of the novel (as shown in Tables 5 & 6), we can narrow it down to the top three dominant themes of the novel. By doing a manual cross-verification or by reading the summaries of all the novels, it was observed that the output given by MALLET is in agreement with that provided by manual inspection. However, the output from Orange Software is not distinct, and the themes are not based on a single topic. There are some words that appear in all 10 topics, which makes the data questionable.

Table 5. Showing Topic Ranking in MALLET

Ranking	Topic Ranking by MALLET									
	1	2	3	4	5	6	7	8	9	10
Baumgartner's Bombay	Topic 6	Topic 3	Topic 2	Topic 8	Topic 9	Topic 5	Topic 0	Topic 7	Topic 1	Topic 4
	32%	26%	21%	16%	2%	1%	1%	1%	1%	0%
The Zigzag Way	Topic 8	Topic 6	Topic 2	Topic 9	Topic 5	Topic 3	Topic 1	Topic 4	Topic 0	Topic 7
	47%	23%	18%	4%	3%	2%	1%	1%	1%	0%
Voices in the City	Topic 2	Topic 6	Topic 1	Topic 8	Topic 9	Topic 5	Topic 0	Topic 3	Topic 4	Topic 7
	28%	27%	25%	12%	3%	3%	1%	1%	1%	0%
Bye-Bye Blackbird	Topic 4	Topic 6	Topic 2	Topic 8	Topic 9	Topic 1	Topic 5	Topic 3	Topic 7	Topic 0
	27%	24%	23%	13%	4%	3%	3%	2%	1%	1%
Clear Light of Day	Topic 6	Topic 2	Topic 7	Topic 8	Topic 9	Topic 5	Topic 1	Topic 3	Topic 0	Topic 4
	30%	28%	26%	11%	2%	1%	0%	0%	0%	0%
Cry, the Peacock	Topic 2	Topic 6	Topic 1	Topic 8	Topic 9	Topic 5	Topic 0	Topic 3	Topic 4	Topic 7
	45%	23%	16%	8%	5%	2%	1%	0%	0%	0%
Fasting, Feasting	Topic 5	Topic 6	Topic 2	Topic 8	Topic 9	Topic 0	Topic 1	Topic 7	Topic 3	Topic 4
	34%	25%	22%	15%	2%	0%	0%	0%	0%	0%
Fire on the Mountain	Topic 9	Topic 2	Topic 6	Topic 8	Topic 5	Topic 1	Topic 7	Topic 3	Topic 0	Topic 4
	36%	27%	23%	11%	1%	1%	0%	0%	0%	0%
In Custody	Topic 6	Topic 0	Topic 2	Topic 8	Topic 9	Topic 5	Topic 1	Topic 3	Topic 4	Topic 7
	34%	24%	21%	14%	3%	1%	1%	1%	0%	0%
Journey to Ithaca	Topic 6	Topic 2	Topic 8	Topic 0	Topic 5	Topic 9	Topic 1	Topic 3	Topic 4	Topic 7
	27%	24%	23%	17%	5%	3%	1%	0%	0%	0%
Rosarita	Topic 8	Topic 2	Topic 6	Topic 5	Topic 9	Topic 1	Topic 0	Topic 3	Topic 4	Topic 7
	39%	20%	20%	10%	4%	2%	1%	1%	1%	0%

Table 6. Showing Topic Ranking in Orange Software

Topic Ranking by Orange									
Ranking	1	2	3	4	5	6	7	8	9
Baumgartners Bombay	Topic 1	Topic 3	Topic 8	Topic 9	Topic 7	Topic 5	Topic 6	Topic 4	Topic 2
	3817.97	25.5778	23.727	1.13992	-7.16532	-43.0358	-69.18	-155.37	-201.09
Bye-Bye Blackbird	Topic 1	Topic 8	Topic 2	Topic 7	Topic 6	Topic 4	Topic 3	Topic 5	Topic 9
	37.8032	12.4686	5.07991	1.81649	1.5675	0.39509	-5.7846	-6.1876	-60.721
Clear Light of Day	Topic 1	Topic 2	Topic 5	Topic 8	Topic 9	Topic 7	Topic 4	Topic 6	Topic 3
	4602.57	246.519	29.914	16.4722	1.55579	-5.29936	-7.3995	-23.601	-262.99
Cry the Peacock	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
	0	0	0	0	0	0	0	0	0
Fasting, Feasting	Topic 1	Topic 6	Topic 7	Topic 8	Topic 9	Topic 4	Topic 3	Topic 5	Topic 2
	2909.79	119.093	88.0048	4.49682	1.66599	-17.8556	-44.133	-55.22	-77.297
Fire on the Mountain	Topic 1	Topic 6	Topic 5	Topic 3	Topic 9	Topic 8	Topic 2	Topic 4	Topic 7
	2270.04	89.431	71.0852	-0.78123	-3.79817	-55.395	-59.965	-67.684	-74.376
In Custody	Topic 1	Topic 4	Topic 9	Topic 6	Topic 8	Topic 7	Topic 5	Topic 3	Topic 2
	3454.89	133.812	-0.92884	-24.6314	-25.1885	-35.2597	-62.764	-85.118	-263.85
Journey to Ithaca	Topic 1	Topic 2	Topic 3	Topic 4	Topic 9	Topic 6	Topic 7	Topic 8	Topic 5
	5898.75	233.828	232.737	39.7353	-0.29972	-5.94338	-6.926	-7.1572	-31.806
Rosarita	Topic 1	Topic 8	Topic 6	Topic 5	Topic 4	Topic 3	Topic 9	Topic 7	Topic 2
	1053.91	98.1241	54.0032	46.3976	42.4486	40.5071	4.48062	-45.835	-90.256
The Zigzag Way	Topic 1	Topic 5	Topic 7	Topic 3	Topic 4	Topic 9	Topic 8	Topic 6	Topic 2
	2627.25	145.297	66.769	46.3279	34.7939	-1.97478	-10.803	-43.243	-144.11
Voices in the City	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
	0	0	0	0	0	0	0	0	0

5. Topic Modelling

The topic modeling graph comparing the distribution of topics in the novels helps us understand the most prevailing topics. The most prevailing topics according to MALLET's output are Topics 2, 6, and 8 (as shown in figure 5). Topic 2 deals with Personal Emotions & Inner Life, Topic 6 deals with Reflections on Time & Daily Life, and Topic 8 deals with Family Structure & Nurturing. By studying the novels, it is found that Anita Desai's works revolve around these themes. The finding in the research paper "Major Themes and Techniques in the Novels of Anita Desai" is validated by this research, and the writer states, "Her sensitive portrayal of the inner feelings of her female characters is excellent." The majority of Anita Desai's novels engage in a discussion about the conflicts among family members and the isolation of middle-class women.[33] Dr. Sadhana Agrawal, in "A Study of Thematic Vision in Anita Desai's Novels," notes that alienation, existential loneliness, displacement, disharmony in marriage, cultural uprooting, and the psychological suffering of women in a patriarchal, urbanized, indifferent society are prevalent in the novels.[34] Existential loneliness and alienation fall under the category of Personal Emotions & Inner Life; cultural displacement and uprooting fall under Reflections on Time & Daily Life; and marital discord and the psychological suffering of women in a patriarchal society fall under the category of Family

Structure & Nurturing. Therefore, the information provided by the MALLET software is verifiable as well as authentic.

However, as figure 6 shows, Topic 1, General Actions and Dialogue, prevails across all the novels. General Actions and Dialogue such as "would," "one," "could," "said," "like," "see," "come," "go," "even," and "back" cannot determine the theme or tone of the novel. The data is therefore rendered invalid, and the result cannot be considered valid.

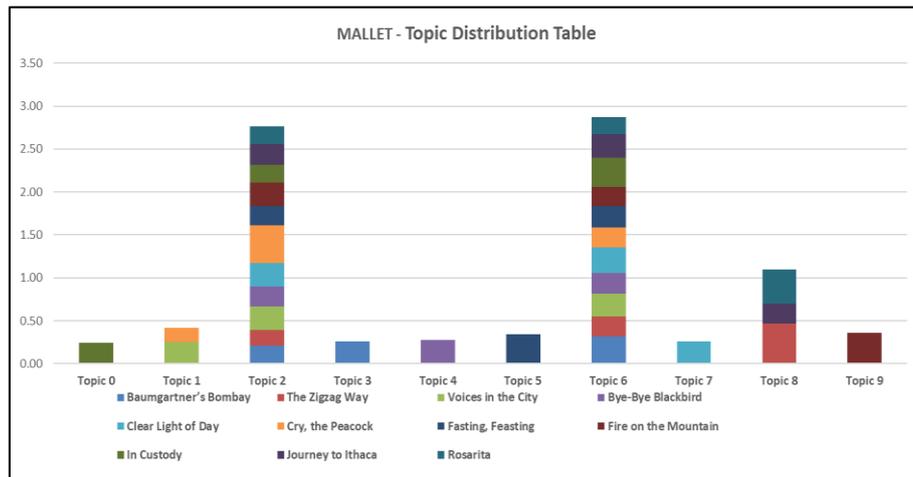


Figure 5. Showing Topic Distribution Across the Novels Generated by MALLET

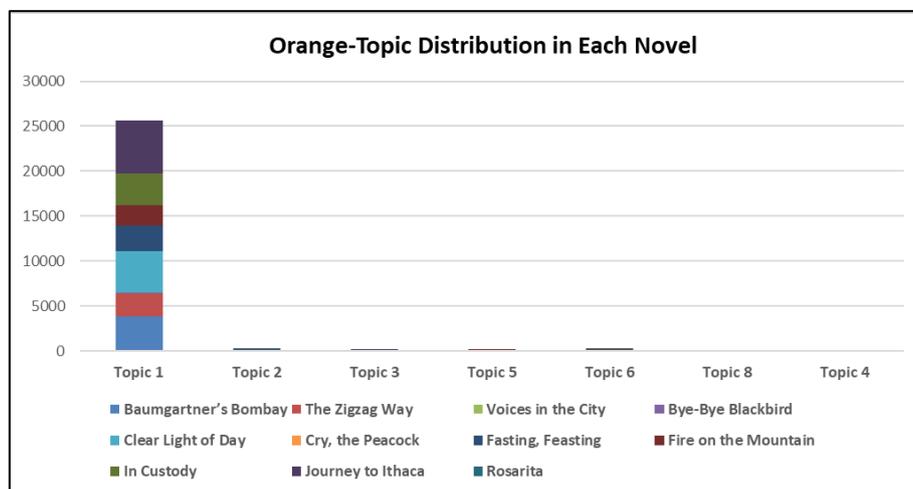


Figure 6. Showing the Topic Distribution Across the Novels Generated by Orange

Comparing Figures 7 and 8 further solidifies the finding that MALLET is a better topic modelling software compared to Orange. In Graph 3, there is an even distribution of topics, indicating the richness of Anita Desai's writing. Figures 9 and 10 help in visualising the thematic link between all the selected novels.

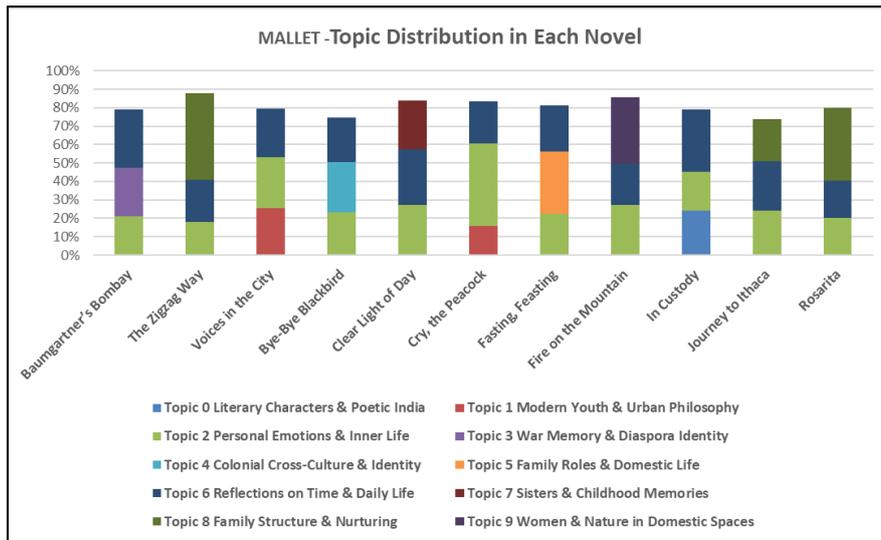


Figure 7. Showing Distribution of the Top 3 Topics in Each Novel (MALLET)

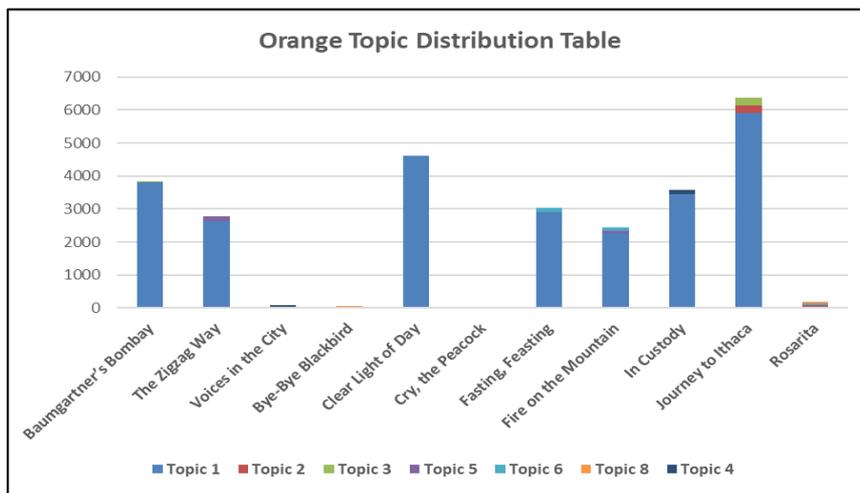


Figure 8. Showing the Distribution of the Top 3 Topics in Each Novel (Orange)

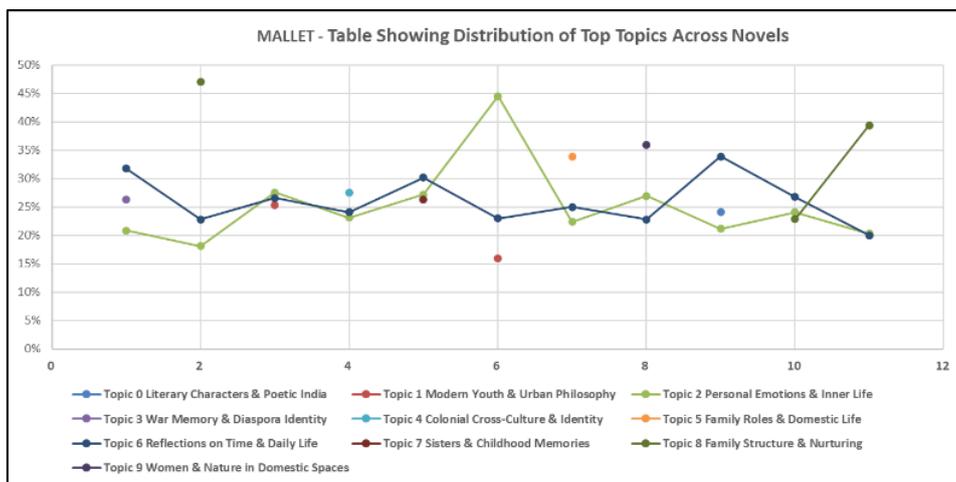


Figure 9. Showing the Underlying Themes in the Novels (MALLET)

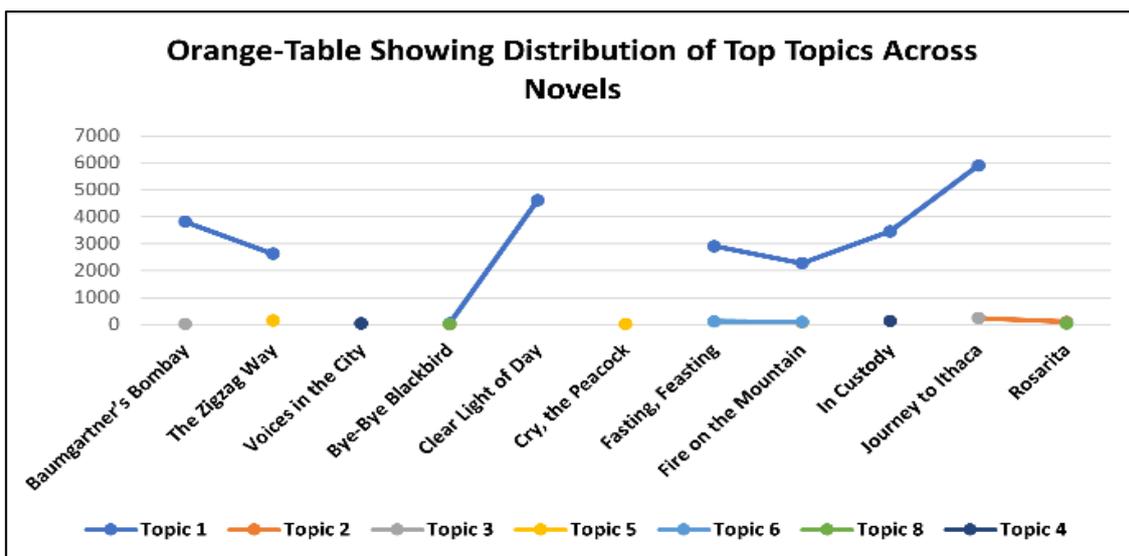


Figure 10. Showing the Underlying Themes in the Novels (Orange)

6. Conclusion

Topic modeling software can indeed be an incredibly helpful tool if applied to the thematic analysis of literary works. However, one must remember that topic modeling is no substitute for traditional analysis. It is merely a tool that facilitates the process by providing an interesting experience when working with a vast corpus of text. One of the largest drawbacks of using digital humanities tools is that coding remains an issue for students of the humanities. Even given the availability of multiple tutorials online, the ability to understand and implement the code with precision requires skill and patience. Based on the research, it is questionable whether the MALLET program is much improved compared to the Orange program. Following the process of installation and the basics of programming, the desired result is relatively simple to achieve. Orange software makes extensive use of manual data preprocessing. Stopword removal is one of its features, which is particularly difficult in Orange software. MALLET assists in removing stopwords, making the processed data more reliable and precise. For a beginner trying to pull data, MALLET software is a great choice. While the coding is done in the command prompt and data is limited, it can be easily plotted using a basic insert chart option in an Excel document. Digital humanities is not something that will save a beginner's time; learning the software is time-consuming and burdensome. Even with knowledge of general concepts and working methodologies, using tools like Orange makes the task all the more difficult. MALLET is a great tool for topic modeling, but it is inflexible. There are no provisions for sentiment analysis or other text analyses in it. Topic modeling is only one step

toward computational literary analysis. Scholars must download certain packages to carry out critical study, which requires computing machines with sufficient processing capacity. The application of digital humanities in literary work analysis deepens the analysis by extracting sub-textual themes and correlations from the data corpus. Computer literacy is no longer the preserve of engineers or computer science students in this era of explosive AI development. In every new development, if we don't support the change, the change will dominate us and plunge us into oblivion. Literary students can do nothing better than welcome this new digital age and AI with open arms and work on themselves if they step forward to learn the finer aspects of digital humanities. Digital humanities neither clip nor cut the wings of literary research but give unimaginable power to the research. It is then the researcher's duty to channel that power and knowledge to optimize the scope of the research work output.

References

- [1] Hankar, Mustapha, Mohammed Kasri, and Abderrahim Beni-Hssane. "A comprehensive overview of topic modeling: Techniques, applications and challenges." *Neurocomputing* (2025): 129638.
- [2] Helmi, Laila CA. "Digital humanities: A paradigm for the 21st century." *BAU Journal-Society, Culture and Human Behavior* 2, no. 2 (2021): 5.
- [3] Su, Fangli, Shihao Li, Qinghua Liu, and Yin Zhang. "A bibliometric study of digital humanities research in China from 2012 to 2021." *Information Research an international electronic journal* 28, no. 4 (2023): 62-82.
- [4] Yadav, Deny. "The role of Artificial Intelligence in literary analysis: a computational approach to understand literary styles." *International Journal of Emerging Knowledge Studies* 3, no. 9 (2024): 558-565.
- [5] Preeti, Dr. "Quantitative Analysis of Literary Texts: Computational Approaches in Digital Humanities Research." *Educational Administration: Theory and Practice* 30, no. 5 (2024): 5234-5240.
- [6] Panagiotidou, Georgia, Houda Lamqaddam, Jeroen Poblome, Koenraad Brosens, Katrien Verbert, and Andrew Vande Moere. "Communicating uncertainty in digital humanities

- visualization research." *IEEE Transactions on Visualization and Computer Graphics* 29, no. 1 (2022): 635-645.
- [7] Asselborn, Thomas, Ralf Möller, and Sylvia Melzer. "Implementation of information systems for the long-term reuse of data in humanities research." *Procedia Computer Science* 256 (2025): 86-92.
- [8] Hawkins, Ashleigh. "Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web." *Archival Science* 22, no. 3 (2022): 319-344.
- [9] Laureate, Caitlin Doogan Poet, Wray Buntine, and Henry Linger. "A systematic review of the use of topic models for short text social media analysis." *Artificial Intelligence Review* 56, no. 12 (2023): 14223-14255.
- [10] Shang, Wenyi, Rongqian Ma, and Heather Moulaison-Sandy. "How does digital humanities research talk about AI? A bibliometric analysis." *Information Research an international electronic journal* 30, no. iConf (2025): 635-645.
- [11] Gupta, Priyanka, Bosheng Ding, Chong Guan, and Ding Ding. "Generative AI: A systematic review using topic modelling techniques." *Data and Information Management* 8, no. 2 (2024): 100066.
- [12] Eze, Chibuiké Samuel, and Lior Shamir. "Analysis and prevention of AI-based phishing email attacks." *Electronics* 13, no. 10 (2024): 1839.
- [13] Akef, Islam, Juan S. Munoz Arango, and Xiaowei Xu. "Mallet vs GenSim: Topic modeling for 20 news groups report." *Univ. Ark. Little Rock Law J* 2, no. 39205 (2016): 10-13140.
- [14] Z. Dobesova, "Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics," *Computer Applications in Engineering Education*, vol. 32, no. 4, Jul. 2024, doi: 10.1002/cae.22735.
- [15] M. R. Fengler and T. M. Phan, "Unveiling themes in 10-K disclosures: A new topic modeling perspective," *International Review of Financial Analysis*, vol. 103, Jul. 2025, doi: 10.1016/j.irfa.2025.104121.

- [16] N. Wang, “The rise of a new paradigm of literary studies: The challenge of digital humanities,” *New Techno Humanities*, vol. 2, no. 1, pp. 28–33, Jun. 2022, doi: 10.1016/j.techum.2022.11.001.
- [17] G. Martens, “Literature, digital humanities, and the age of the encyclopedia,” *CLCWeb - Comparative Literature and Culture*, vol. 15, no. 3, Sep. 2013, doi: 10.7771/1481-4374.2241.
- [18] M. Dockray-Miller, M. D. C. Drout, S. Kinkade, and J. Valerio, “The Author and the Authors of the ‘Vita Ædwardi Regis:’ Women’s Literary Culture and Digital Humanities,” *Interfaces: A Journal of Medieval European Literatures*, no. 8, pp. 160–213, Dec. 2021, doi: 10.54103/interfaces-08-09.
- [19] J. Brown, “Whose Language? Whose DH? Towards a taxonomy of definitional elusiveness in the digital humanities,” *Digital Scholarship in the Humanities*, vol. 38, no. 2, pp. 501–514, Jun. 2023, doi: 10.1093/lc/fqac072.
- [20] Andrew McCallum and Andrew Kachites., “Mallet: MACHine Learning for Language Toolkit,” 2002.
- [21] [21] University of Ljubljana and Jožef Stefan Institute, “Orange: Data Mining Toolbox ,” 3.38.1.
- [22] A. Desai, *Baumgartner’s Bombay*. Penguin Random House UK, 1988.
- [23] A. Desai, *The Zigzag Way*. Penguin Random House UK, 2004.
- [24] A. Desai, *Voices in the City*. Helmrich & Blair Ltd\, 1965.
- [25] A. Desai, *Bye-Bye, Blackbird*. HIND POCKET BOO’KS PVT. LTD, 1971.
- [26] A. Desai, *Clear Light of Day*. First Mariner Books Edition, 2000.
- [27] A. Desai, *Cry the Peacock*. St Stephen’s (Bristol) Press Ltd, 1991.
- [28] A. Desai, *Fasting, Feasting*. Chatto & Windus, 1999.
- [29] A. Desai, *Fire on the Mountain*. UK: Vintage Random House, 1999.
- [30] A. Desai, *In Custody*. Heinemann, 1984.

- [31] A. Desai, *Journey to Ithaca*. Vintage, 2001.
- [32] A. Desai, *Rosarita*. Picador, 2024.
- [33] D. Gurjar, "Major Themes and Techniques in the Novels of Anita Desai," *International Conference on Socio-Cultural Dimensions of English Studies*, vol. 8, no. 4, Feb. 2023, [Online]. Available: www.vidhyayanaejournal.org
- [34] S. Agrawal, "A Study of Thematic Vision in Anita Desai's Novels," *An International Refereed e-Journal of Literary*, vol. 3, 2015, [Online]. Available: www.researchscholar.co.in