paper content.pdf

Unsupervised Topic Modeling with LDA for Textbook Content Comprehension, a qualitative survey

Abstract

In this research LDA driven exploratory analysis is shown to depict various keywords related to subtle topic in context. It identifies latent topics within textbook lessons, uncovers coherent themes from textual data, aims to improve the curriculum provided English textbook content synthesis and acquisition skill of learners. It is anticipated extracted topics enable readers to comprehend the curriculum material more efficiently. Extensive analysis is conducted to visualize, high impact keyword, co-occurrence patterns and correlation of extracted topics. A prototype mobile app is developed which incorporates topic modeling extracted keywords. Furthermore, qualitative research survey is undertaken to evaluate its effectiveness on end-users, course instructors of Bangladesh's higher secondary school. The challenges, future potential of LDA extracted content integrated mobile app into the learning process is explored. After collecting feedback, word clouds were used to analyze the participants' recommended terms, and the LIWC approach is used to estimate overall sentiment. LIWC score showed positive sentiment and survey process enticed the participants, demonstrates learners eager to use NLP technology driven topic modeling approach in teaching and learning, and there are tremendous opportunities.

Introduction

In Bangladesh there is lacking in effective acquisition, synthesis skill of English language from National Curriculum and Textbook Board (NCTB) curriculum provided textbook [1]–[4]. Specially in the rural area in National Board examinations like SSC, HSC most of the student get poor marks in English subject. It is anticipated students have lacking in understanding context. Topic modeling can play a significant role for context understanding for curriculum provided English textbook. Topic modeling is a subfield of natural language processing and machine learning, offers a promising unsupervised approach to identify latent topics within provided documents. It can help identify the main themes, concepts, and topics within the textbook's content, enabling instructors to tailor the learning experience to individual students. LDA is one the prominent algorithm which can be used for topic modeling. It provides coherent topics, dominant keywords, latent combination of features that characterizes similarities between topics. In this research LDA extracted keywords are rearranged and incorporated into a mobile app to observe user experience. Some renowned mobile apps are Duolingo, Busuu, Babel, Voxy etc [25]. Across all English learning applications via digital media, 55% have activities for vocabulary learning and other exercises are about 41% [18], [19] includes quizzes, exercises, and game for enhancing learners' comprehension and self-checks [17]. One caveat is of these apps are these are not based on Curriculum Board provided Textbook for learning English hence, could not able to attract a large number of pupils in Bangladesh who are mostly depended on NCBI textbook.

To grasp the English language knowledge from curriculum provided textbook a novel approach LDA based unsupervised Topic modeling using textbook corpus is adopted and exploratory analysis is demonstrated in this study. Our anticipation is through this way student can able to interpret meaningful information facilitates students to understand the correlated topics and important keywords related to that topics leads to understand the subtle meaning of the textbook context.

Research Overview

LDA is a widely used probabilistic topic modeling technique which can automatically identify underlying themes or topics within a corpus of given text. It leverages learning experience, improve interpretation and knowledge acquisition. The synthesis of existing research sheds light on the potential of topic modeling to create adaptive and tailored learning experiences, ultimately improving student engagement, comprehension, and knowledge retention. LDA doesn't directly account for student engagement in interactive activities such as learning tasks in mobile apps. Hence a prototype app is developed and qualitative survey is conducted to observe the sentiment impact.

Qualitative survey research is undertaken to evaluate the effectiveness of unsupervised topic modeling LDA Bangladesh's National Curriculum Textbook Board (NCTB) provided English Textbook for Higher secondary school education. This study seeks to ascertain if students can learn English better if a mobile app is introduced which includes NLP's LDA driven topic modeling applied extracted keywords and analysis. A prototype mobile app is developed to incorporate the topic modeling extracted keywords into the app. This article presents the key findings and insights from the survey, shedding light on the prospective of learners especially instructors. In the survey questions, it was indicated whether the students, teachers/instructors, and government organizations would find it acceptable and appreciated if textbook information were made available through a mobile app and presented in interactive format. To demonstrate the mobile app idea during the interrogation survey session a prototype is also prepared. Participants were asked for suggestions on how to make the app better and about any shortcomings. After collecting feedback, word clouds were used to analyze the frequency of the participants' recommended terms, and the LIWC approach was used to estimate overall sentiment. The survey's findings show that teachers are eager to use NLP provided extracted keywords technology in teaching and learning, and there are tremendous opportunities.

Topic Modeling

Different techniques have been developed to perform topic modeling in unsupervised topic modelling in NLP, having their own strengths and limitations. Apart from LDA, Mallet LDA, STM (Structural Topic Model), and HDP (Hierarchical Dirichlet Process) etc are also prevailing. Algorithms like Non-Negative Matrix Factorization (NMF) or Latent Semantic Analysis (LSA) can also be considered.

Topic models comparative analysis

While some variations of LDA, like Mallet LDA, focus on scalability, LDA in general can still be efficiently applied to moderately sized corpora. If large corpus needs to analyze, Mallet LDA might be more suitable. Analyzing topics within the context of metadata, STM could be a better fit. Hierarchical Dirichlet Process (HDP) can be useful when we cannot guess the number of topics in advance. However, as a baseline model LDA is often considered one of the most prominent choices. In this study Textbook corpus is divided into lessons which is a mixture of topics and using LDA expecting to determine which word in the lesson belong to Lesson's topics.

Latent Dirichlet Allocation (LDA):

LDA is a probabilistic model that assumes documents are mixtures of topics, and each topic is a distribution over words. The goal is to infer the hidden topic assignments and the topic-word distributions that best explain the observed documents. The joint distribution of LDA model can be expressed as

$$p(\theta_d, z, w \vee \alpha, \beta) = p(\theta_d \vee \alpha) \prod_{n=1}^{N} p(z_{d,n} \vee \theta_d)\, p(w_{d,n} \vee z_{d,n}, \beta)$$

Where $w_{d,n}$ The n_{th}word in document d, $z_{d,n}$ the topic assigned to the n_{th} word in document d, α, β are the Dirichlet LDA model parameters. controls per-document topic distribution, and per topic word distribution. θ_d represent the topic distribution. $p(\theta_d|\alpha)$ Dirichlet distribution representing the document-topic distribution, $p(z_{d,n}|\theta_d)$ is the word topic assignment for the n_{th} word in document d, $p(w_{d,n} \vee z_{d,n}, \beta)$ is the distribution representing the observed word given a topic.

Baseline Model:

LDA produces interpretative results for exploratory topic analysis. The identified topics are represented as distributions over words, making it easy to assign meaningful labels to topics. Provided by most of the libraries and tools, making it easy to implement and can be integrated into existing workflows. Hence, LDA serves as a solid baseline for topic modeling tasks. We have chosen LDA for baseline statistical topic modeling tool. However, how many topics are ideal it is needed to determine and also topic modeling quality needs to measure.

Determining Optimal Topics with Coherence

Coherence score measure how coherent or interpret the words in that topic and estimates number of topic clusters. Coherence score assess the quality of the topics produced by LDA and ensures that the topics generated are statistically significant. Coherence C_{topic} can be expressed as follows

$$C_{topic}=\sum_{i=1}^{N}\frac{1}{N(N-1)}\sum_{j=1}^{i}PMI(w_i,w_j)$$

Where, $PMI(w_i,w_j)$ represent pointwise mutual information statistical association between two words occurring together. PMI score indicates that the two words are more closely related within a topic. $PMI(w_i,w_j)$ can expressed as

$$PMI(w_i,w_j)=\log\frac{P(w_i,w_j)}{P(w_i)P(w_j)}$$

where $P(w_i,w_j)$ is joint probability of occurrence of words w_i and w_j.

To calculate the coherence score genism library provides range of options such as $u_{mass}, c_v, c_{uci}, c_{npmi}$. u_{mass} and c_vThese two methods are most popular. For given topic with words $\{w_1,w_2,w_3,\ldots\ldots,w_n\}$ a fixed context window size is provided (default size 10 words) then coherence score is calculated using an equation $\sum_{j=1}^{i}PMI(w_i,w_j)$ which provides negative coherence score. c_v can be expressed as

$$c_v=\frac{1}{N(N-1)}\sum_{j=1}^{i}similarity(w_i,w_j)$$

in which $similarity(w_i,w_j)$ represent the pairwise similarity between terms based on $PMI(w_i,w_j)$ scores. c_v provides a positive coherence score.

Higher coherence values indicate that the topics are more coherent and representative of meaningful themes within the text data. Coherence score 0.5 are fairly good [43].

Literature Review

It is a probabilistic model that assumes each document is a mixture of a small number of topics and that each word's presence in a document is attributable to one of the document's topics. The goal of LDA is to uncover these latent topics from a collection of documents without needing any prior labeling or categorization of the content. LDA based topic modeling has been used for semantic search, ontology exploration, classification, dominant keywords searching in many research studies.

LDA for Textbook Content

For curriculum based textbook study it could be an option is not revealed from rigorous search in online repositories. In 2021 Krishna Raj [41] explained human brain has a tendency to overlook a number of minor details about the events in the book. The model can scan massive amounts of text in the book and display interesting ideas derived from the given book. LDA Machine learning and lexical approaches can be used to analyze literary works. LDA model can aid in the easy and accurate adaptation of a book, making the learning process much simpler and precise. Research by Rani In 2020 [42] topic modeling text summarization approach for Hindi novels and stories was proposed. The proposed model is implemented by infusing linguistic features into LDA based topic modeling to discover a set of topic-words from the provided document. Educational content-based topic modeling for an Intelligent system to develop a tutoring system is proposed in [34] by researcher Stefan Slater in 2017. It is proposed a personalized learning system using correlated topic modeling, a natural language processing approach, to analyze the linguistic content of mathematics problems. For solving mathematics problems, a range of potentially meaningful and useful topics within the context is explored. They showed that correlated topic modeling is an effective approach for automatically labeling for personalized learning system. For key terms detection within articles LDA based topic modeling has been used in many research studies in which dominant keywords reveals future research trends or most prominent topics. Investigation of Julio Guerra in 2013 showed how LDA model can be used for textbook content linking. It can be further applied to facilitate content modeling and context understanding for collections of reference books of same subject [33]. The dataset was collections of textbooks of two domains Elementary Algebra and Information Retrieval. They inferred LDA topic modeling recommendation and navigation support for e-educational systems is promising.

LDA for Dominant Keywords Determination

In 2020 Kazi Masudul Alam demonstrated that LDA-based topic modeling can observe the trend of Bangla news [35]. Their experiments prove that proper corpus and labeled LDA is a good model for news topic modeling and articles become more human-readable if LDA tagged Labels are assigned. Research trends related to COVID-19 and sports' were analyzed using topic modeling in 2022 by J. Lee [36]. The LDA topic modeling technique revealed the latent knowledge dimension and structures in 'Sports-COVID-19' articles. In 2022 Rahul Gupta [37] involves the analysis of research trends in 3269 articles published under “Applied Intelligence” from 1991 to 2021, using the application of LDA. Topic modeling using BoW as well as using TF-IDF score was performed. Their analysis showed that BoW outperforms the TF-IDF. In this research BoW is also used for LDA model. In 2019 Wafa Shafqat [38] proposed an architecture model for better understanding of crowdfunding comments posted by the investors to understand their motive to classify whether comments are scam or legitimate comments. Deep neural network Language modeling either LSTM or RNN encoded vectors are fed into a LDA based topic modeling model to understand the context of discussion trends. Afterwards compared the results with simple Neural Networks (NNs) and non LDA based approach which shows their model can play a substantial role in a better understanding of crowdfunding comments context understanding.

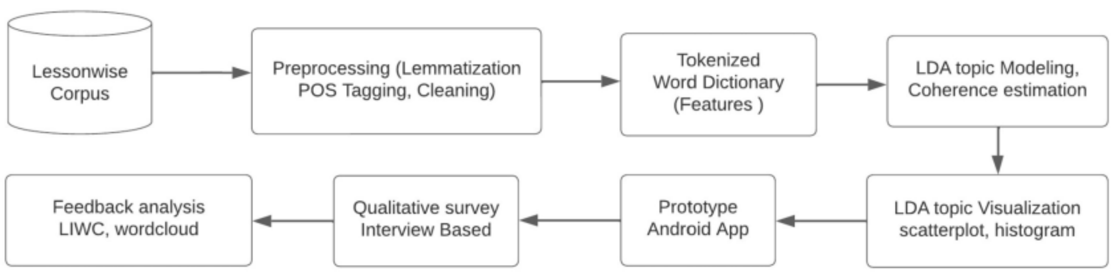
LDA based Sentiment Analysis

Sentiment analysis has been a key research area of NLP based research domain, where LDA has been applied to determine significant features and those features contributes to segregate sentiments and provide recommendations. LDA based topic modeling has been used in sentiment analysis task. In 2021 Y. Cho published research study of LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model is proposed in this article [39]. This article proposed TDS novel approach that combines LDA-based topic modeling for sentiment analysis within documents. The authors used LDA-based topic modeling and sentiment analysis to explore the Chinese public's perception of Omicron variants on social media Sina Weibo 121,632 pieces of omicron data [40]. From topic analysis they realized omicron's impact, infection situation, pandemic prevention and control geographically. Hence, it is actually revealed LDA based topic modeling can be used for understanding subtle topics and exploring various facts.

Hence, from the above literature review we can infer LDA could help analyze the content of the textbook and identify the main topics covered. This information could then be used to enhance the learning experience.

Methodology:

The methodology involves preprocessing textual data, training the LDA model on the preprocessed text, and subsequently interpreting and visualizing the generated topics. Data is collected from NCBI's English Textbook for class 9 of higher secondary school. Data mining approach is applied to segregate into subsequent lessons. Then textual data is preprocessed to remove noise, text standardizing, followed by the application of LDA to identify underlying topics, then coherence measurements are applied. Research overview is depicted as follows



Workflow of research

LDA algorithm automatically discovers latent topics within the documents based on word co-occurrences. Each topic will be represented by a set of words. Interpret these words to understand the main concepts associated with each topic. Extensive exploratory analysis is conducted to visualize the topic modeling outputs. Analyze the topics generated by the model require manual review and adjustment to ensure the topics make sense.

Data Processing and Feature Extraction

a) First NLP's data processing or data mining techniques are applied for meaningful token or feature extraction. Text is converted to Lowercased and Normalized to ensure consistent pre-processing.

i) Data cleaning: unwanted characters, punctuation and special character removed and stop words (such as "and," "the," "is," etc) are removed. Spacy library's English word model and NLTK's stopwords list are used together. Also, words less than two characters are removed such as: I, Hi, Oh etc. Hence, Noise is removed and irrelevant characters, symbols, or data artifacts that have been introduced during data collection or scraping from pdf file to text file generation are separated. Hence, we found a cleaned corpus.
ii) Lemmatization: Root words are collected words to their dictionary form (lemma) is extracted using NLTK's WordNetLemmatizer package. Stemming Reduce words to their base or root form is not used since sometimes it changes the actual words.
iii) Part-of-Speech Tagging: Spacy's English model 'en_core_web_sm' is used to extract interested words (such as noun, verb, adjective) and excluded (CCONJ, AUX, DET, INTJ, PART etc which are Coordinating Conjunction, Auxiliary, Determinator, Interjection, Particle etc) thereby token is collected for only which are not punctuation, conjunction, symbol etc.

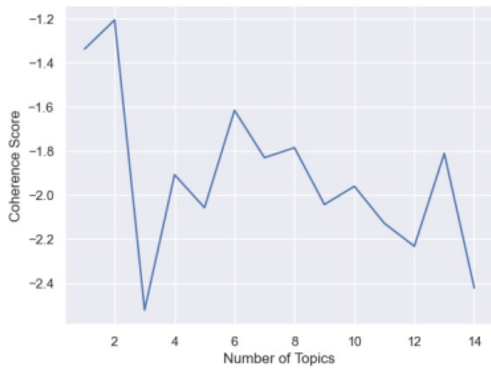
Exploratory analysis of Textbook content

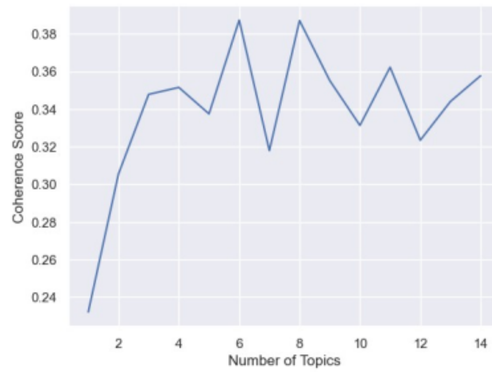
Here whole book is segregated into Lessons and we wanted to explore the important topics within the content. Similar topic words remain together. Therefore, assumptions are, it helps students to understands the words, sentences and context of the book. Ideal number of topics are determined using coherence score.

Coherence for LDA model

To measure coherence in the context of LDA, following steps are followed:

1. Cleaned document samples are prepared using python's NLP data mining techniques explain in detailed in data reprocessing section. Prepared $T(d_i)$ set of tokens in Documents d_i for i^{th} document samples in corpus D.
2. Doc to BOW corpus dictionary is prepared with Doc2Bow vector. This vector x_d can be represented as where $n(w_i, d)$ denotes the count of words w_i for the document d.
3. Trained LDA Model: During the training phase gensim's MulticoreLDA model with four CPU worker thread is set. Doc2Bow dictionary is applied along with 20 iterations is invoked. The rest of the parameters for LDA model training was default parameter settings of gensim library.
4. Calculate Coherence: To Calculate the coherence score for each LDA model for n number of topics step 3 is iterated for $n=15$ times.
5. Iteration result coherence score for n number of topics are saved in a list and plotted using seaborn.





From the chart we can see that six topics are dominant in our provided corpus. The chart shown at the left shows the coherence score for u_{mass} and the right chart represents the score for c_v for multiple iterations. Using 6 topics we can see the output of corresponding topic and top 10 words in a topic.

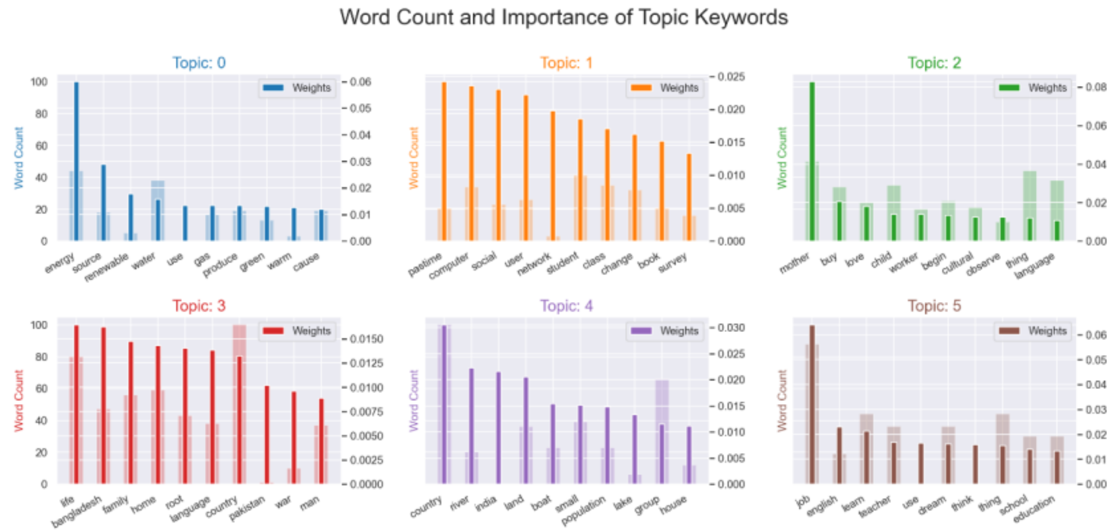
Topic: 01 ["energy" 0.060, "source" 0.029, "renewable" 0.018, "water" 0.016, "use" 0.013, "gas" 0.013, "produce" 0.013, "green" 0.013, "warm" 0.013, "cause" 0.012,]
Topic: 2 ["pastime" 0.024, "computer" 0.024, "social" 0.023, "user" 0.022, "network" 0.020, "student" 0.019, "class" 0.017, "change" 0.016, "book" 0.015, "survey" 0.013,]
Topic: 3 ["mother" 0.083, "buy" 0.021, "love" 0.018, "child" 0.014, "worker" 0.014, "begin" 0.013, "cultural" 0.012, "observe" 0.012, "thing" 0.012, "language" 0.011,]
Topic: 4 ["life" 0.016, "Bangladesh" 0.016, "family" 0.015, "home" 0.014, "root" 0.014, "language" 0.014, "country" 0.013, "Pakistan" 0.010, "war" 0.010, "man" 0.009,]
Topic: 5 ["country" 0.031, "river" 0.022, "India" 0.022, "land" 0.021, "boat" 0.015, "small" 0.015, "population" 0.015, "lake" 0.013, "group" 0.012, "house" 0.011,]
Topic: 6 ["job" 0.064, "English" 0.023, "learn" 0.021, "teacher" 0.017, "use" 0.016, "dream" 0.016, "think" 0.016, "thing" 0.015, "school" 0.014, "education" 0.013]

Word count vs Relative Importance measurement

Word frequency $n(w_j, d_i)$ in each document D is measured as below which identifies the most frequent words within each document and across the entire corpus.

$$D = \sum_{d_i \in D} \begin{cases} 1, n(w_i, d_i) > 0 \\ 0, n(w_i, d_i) = 0 \end{cases}$$

we can visualize relative importance of any keywords in terms of frequency and plotted inclined with LDA provided weights.



Dominant topic and contribution

In LDA models, each document is composed of multiple topics. But typically, some specific topics are dominant. The following experiment extracts this dominant topic for each sentence and shows the relative weight of the topic and the keywords. It estimated which document belongs predominantly to which topic. How frequently the words have appeared in the documents and the weights of each keyword in the same chart, words that occur in multiple topics and the ones whose relative frequency is more than the weight.

Topic-Term Matrix Visualization and Inter-Topic Distance Map

Visualizing the topics and their relationships in a topic model Python library PyLDAvis is used provides an interactive web-based interface to explore and analyze the LDA results of topic modeling. PyLDAvis itself abstracts away much of the underlying mathematical complexity and provides a user-friendly way to generate visualizations and interactively explore topics and their relationships. Key components distance among topics and salient terms are explained below:

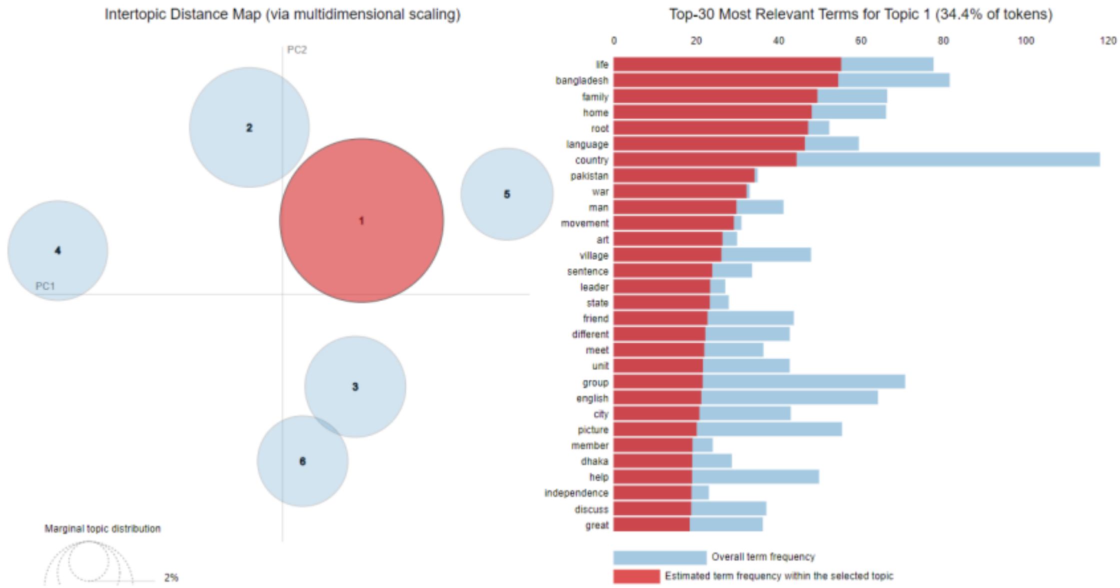
Inter-Topic Distance Map

Distance among topics refers to the measurement of similarity between topics in a high dimensional space matrix provided by the LDA model. PyLDAvis library is used to conserve dimensionality reduction using PCA and for calculating distance between topics metric like Euclidean distance or Cosine Similarity ¿. Topic-topic distribution matrix $Q(t_1,t_2)$ for topic t_1 and topic t_2, distance Dbetween t_1,t_2 can be represented as

$$D(t_1,t_2)=¿(Q[t_1,:],Q[t_2,:])$$

Salient Terms or dominant keywords

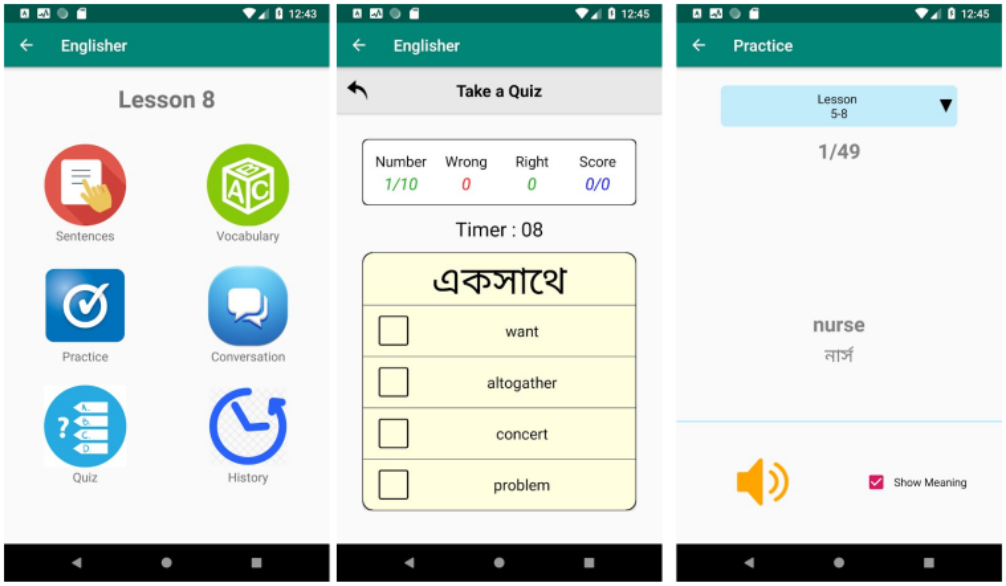
Salient Terms in a topic are words W that are most strongly associated with specific topic. The mathematical expression for finding salient terms w for a topic t involves, extraction of top n words that poses the highest probability scores for topic t in the topic-term matrix $P[t,w]$.

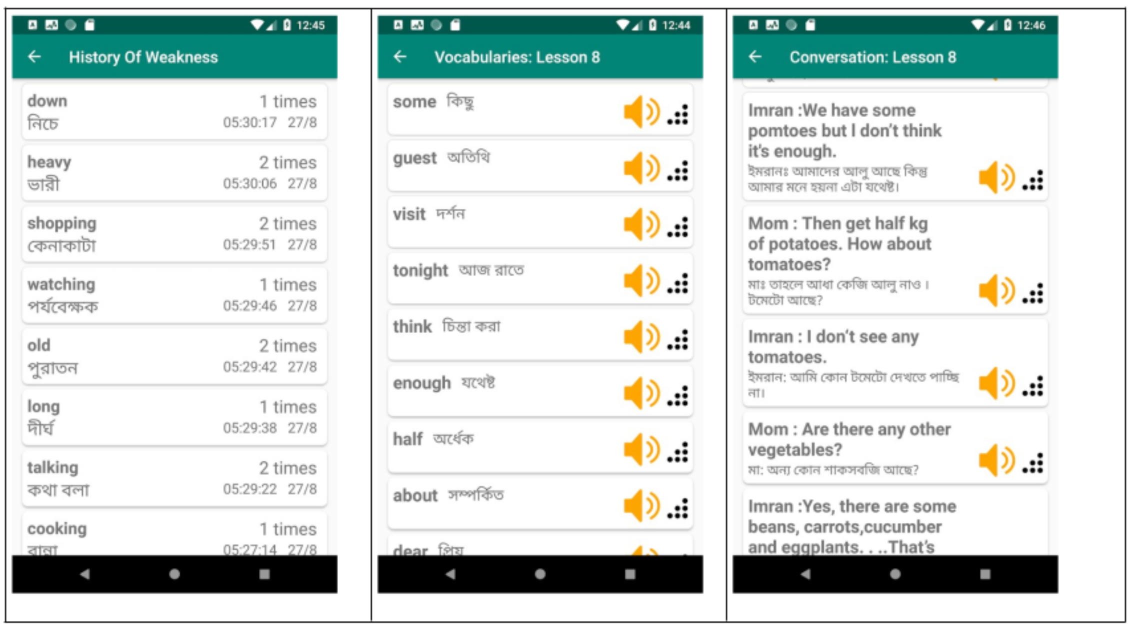


Top 30 most salient terms are showed at right in the bar chart histogram and left figure shows inter topic distance, their size etc. PCA dimensionality reduction technique is applied here to embed the LDA result into a 2D plain scale. Projected the data in lower-dimensional subspace by computing eigenvalues reduced the circle overlapping. Topics that are closer together in the map are more similar in terms of the distribution of words.

Englisher Mobile App:

A mobile application (Englisher) is being developed with content from the NCTB's English Textbook for class 9. The extracted keywords are organized into lessons and furthermore quiz is introduced as an exercise. Each sentence's and word's Bengali meaning is provided in accordance with the lesson. Students can take quizzes, and their results are recorded in the history so that history can be reviewed and performance can be improved by more practice in the future.





Qualitative survey

In the survey questions, it was indicated whether the students, teachers/instructors, and government organizations would find it acceptable and appreciated if textbook information were made available through a mobile app and presented in interactive format. To demonstrate the mobile app idea during the interrogation survey session prototype app Englisher is prepared.

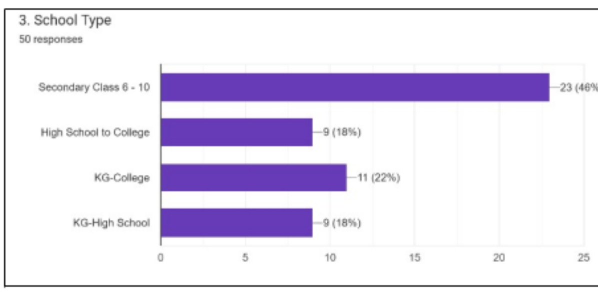
Participants were asked for suggestions on how to make the app better and specify shortcomings. Presumably It provides an insight of teacher's emotion about inclusion of mobile technology in higher secondary English education system.

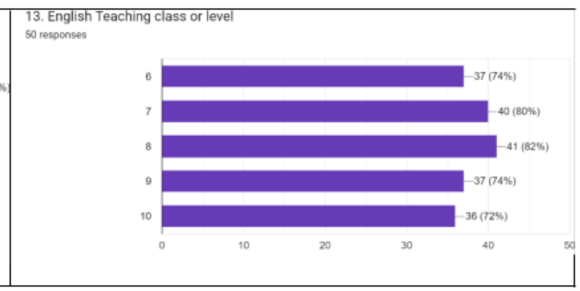
Survey Planning

The survey was conducted over a period of four weeks, with 50 High schools in Dhaka and Bogura district of Bangladesh. It encompasses only English subject areas Teachers who teaches in high schools from class six to class Ten and teaches regularly in the school. A questionnaire was distributed to teachers allowing us to gather questionnaire answer.

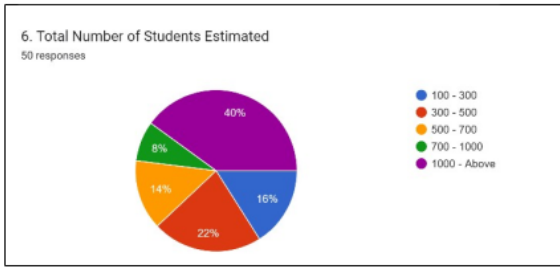
Survey Results:

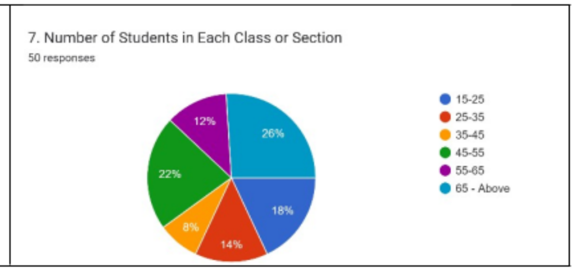
We have done extensive analysis with the survey data collected. In our data collection highest priority is given for the secondary class student teachers who teach between 6-10th class about 46%. High school, KG college and KG High school. Details about the statistics are depicted in the following figure. Adjacent chart explains the percentage of teachers who teach in which class. Hence, from these two figures we can get a vivid image of collected dataset resources about the participating teachers.





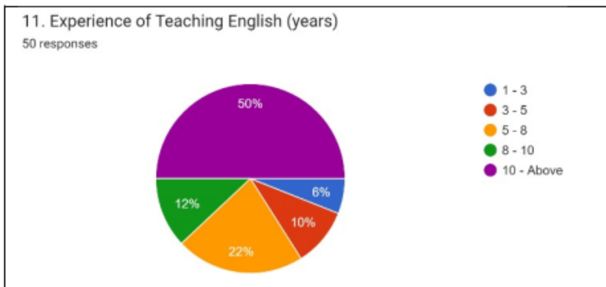
The infrastructure's overall quality and condition, which is generally above average (76% good and 26% considered as average), are shown in the following statistics. The majority of school owners are privately held 45%, yet there are some of variable quality. 32% of which are MPO institutes—non-government educational institutions that receive funding from the government nonetheless—and 22% of which are government institutes. There are quite a few students overall. Over 1000 students attend almost 40% of the institutions. A sizable number of pupils are present in each section and class. A significant percentage of classes—38%—have a size greater than 50. So, we can presume that the participating teachers have quite a bit of experience teaching a lot of children.

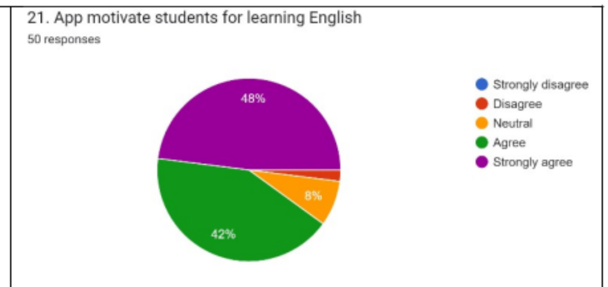


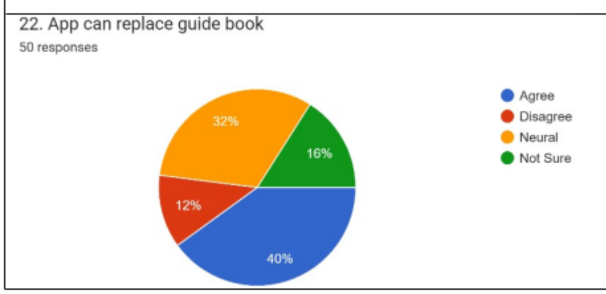


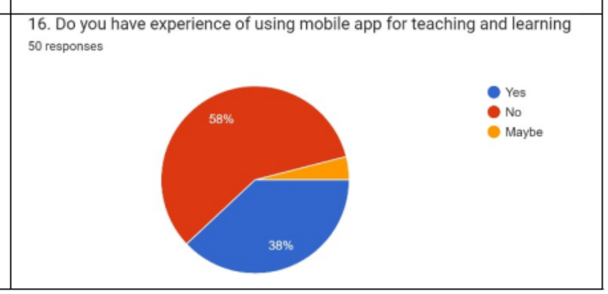
For the data privacy and security issues Teachers were reluctant to provide their social website address to the surveyor. Among 50 participants only 12 has, that means 24% attendees provided their social sites address to use them for research purpose.

The following graphs give an overview of the English teaching experiences of the teachers as well as the general consensus regarding the use of digital content and mobile apps in everyday teaching and learning. Almost 62% of teachers have been teaching for more than 8 to 10 years, and some of them have been teaching for decades in higher secondary education. 32% of teachers have three to eight years of experience, while just 6% are fresh to the profession. Around 83.7% of teachers the language of instruction during their graduation was English, and their major was also English. Very few teachers 13.3% graduation major is something other than English yet teaching English in secondary schools probably have sufficient English language proficiency.









Analysis Facts:

More than half of teachers, or 58%, have no prior experience utilizing mobile apps or technology for teaching, but 90% of them agree, and more than 45% strongly agree, that it encourages pupils to engage actively in their learning. However, they (almost 60%) also hold the opinion that a notebook cannot be completely replaced, despite the fact that mobile apps may solve many problems and provide technological support for teaching and learning. Promisingly optimistic approximately 40%, although thinking that the notebook-based content memorizing learning method can be replaced, feel that mobile app-based learning can replace it permanently.

Questions	***yes***	***No***
Do you use digital content for teaching or digital medium for teaching and learning	84%	16%
Have you ever used Internet or Mobile app to teach students or asked students to find learning materials from internet or Mobile App	76%	24%
Education during graduation was English and English was used for learning	83.70%	16.30%
Customized mobile app for Learning and Teaching English		
Do you think teacher will use the topic model mobile app for teaching	92%	8%
Do you think students will use the topic model app for learning?	80%	20%
Do you think topic model based mobile app-based learning can improve English proficiency of students	86%	14%
Do you think Govt should promote these types of innovation for education sector	98%	2%

This study proposes the Englisher mobile app and presents it to the participating teachers to gather their insightful feedback. 92% of teachers reported that they would use this type of mobile app for teaching if it were made available after using the trial version of the offered customized Englisher app. Teachers anticipate that 80% of students will utilize this app during class. 86% of respondents believed it may help students' English proficiency, and 98% agreed that the government should support this kind of innovation in the education sector.

Linguistic Inquiry and Word Count (LIWC) for Qualitative sentiment

In this research LIWC is used to ascertain the general sentiment of the responses given by the survey participants. Linguistic Inquiry and Word Count (LIWC) [31], [32] is a text analysis tool to measure psychological or emotional characteristics. It aims to quantify sentiment by examining the frequencies of different linguistic terms within given text based on predefined dictionary of words associated with various categories. Let's assume text as a sequence of words: $\{w_1, w_2, \ldots, w_n\}$ and M different linguistic categories: $\{C_1, C_2, \ldots, C_m\}$. Proportion of words in each category $P[i] = \frac{w[i]}{T}$ where T is the total number of Text. Now, a matrix $X[i, j]$ can be formed, where w_i represents the frequency of the word in the linguistic category C_j. LIWC vector containing the proportions of words in each linguistic category can be expressed as

$$P_{total}[j] = \frac{\sum_{i=1}^{T} X[i,j]}{T}.$$

Traditional LIWC Dimension	Answer Text	Standard Commercial Language	Answer Text	Standard for Formal Language	Answer Text	Standard for story language
Positive Tone	2.54	3.96	3.91	2.33	3.22	2.18
Negative Tone	0	1.1	0	1.38	0	1.75
Social Words	2.54	6.87	5.65	6.54	4.08	10.5
Cognitive Processes	13.56	9.35	18.26	7.95	15.88	8.7
Allure	2.54	7.79	3.04	3.58	2.79	5.48
Moralization	0	0.2	0	0.3	0	0.21

From this LIWC table higher proportion of words related to positive emotions indicate a positive emotional tone in the text in the answer for the questions related to "How this app can be improved" and "How English learning can be improved using Mobile App". LIWC is applied for three different categories "commercial writing", "Formal language", and "story language" and in all the categories answer text showed highly positive sentiment from the survey user. Though respondents had a mix of optimism and skepticism regarding the use of mobile apps in teaching and learning. During the interrogation session, their tone was positive and enticed participants.

Visualize participants response with Word cloud:

LIWC involves linguistic analysis using mathematical expression but using word cloud survey answers can be visualized vividly in interpretable interactive format. Word cloud consider a set of words $\{w_1, w_2, \ldots, w_n\}$ extracted from Text document and associated frequencies $\{f_1, f_2, \ldots, f_n\}$, s_i represent the proportional size of the word in the cloud can be expressed as

$s_i = \frac{f_i}{\sum_{j=1}^{n} f_j}$ where normalized frequency f_i.

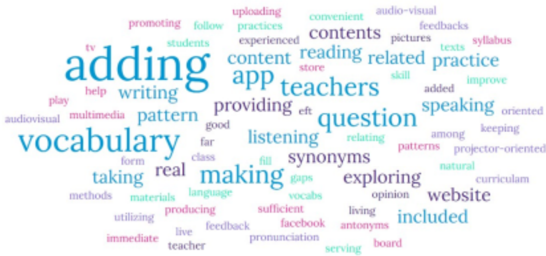


Fig: word cloud for the question regarding English learning app improvement

Fig: word cloud for the question regarding app improvement

The word cloud is generated from the answers provided question “how the app can be improved” and followed by more generalized question “how English learning can be improved using an App”. The participants narrated varieties of viewpoints for the questions. From the word cloud it is inferred adding graphical content would enhance the apps' usefulness and make them more visually appealing to users. More practice resources for listening and exercise would be helpful. Another suggestion is to include synonym, and syllabus-related instances as well as audiovisual engagement with the app. Add additional vocabulary and involve more experience teachers who have greater experience in digital learning and teaching.

Results Discussion and outcomes:

The results demonstrate that LDA-based topic modeling significantly enhances content comprehension by providing concise summaries of the learning material. Readers can grasp the main ideas and connections between topics, aiding in retention and knowledge acquisition. From the qualitative survey it is revealed LDA based topic modeling approach based extracted keywords within mobile apps seem effective as it provides more contextual meaning to the learners. Most of the participating teachers are enthusiastic about topic modeling based contexual resources learning related to technology incorporating into pedagogy. Participants appreciated the cutting edge NLP’s learning resources available through mobile devices. Teachers admitted that available digital resources facilitated a deeper understanding of topics and catered to different learning styles, nurturing more engaging learning environment. This will positively impacted student motivation and overall engagement and hence boost overall learning. Some crucial suggestions were improving the graphics of the app so that it becomes interactive and guardian involvement can be introduced. Based on the survey results, it is revealed the potential for digital mobile-based learning in school is immense. Government should take initiatives to incorporate it into course curriculum syllabus and could impose ordinance to adopt mobile app based learning teaching in the school.

Apps need to be improved by including collaborative form of learning. Additionally, the interactive interface receives positive feedback for its user-friendly design and utility in assisting readers' navigation through the textbook. Recommendation is to make it specific for NCTB Books only for particular class. This approach is also our goal considering NCTB Books. Including interactive e-books, dictionary, educational apps, and multimedia content.

Limitations

LDA could play a role in understanding the topics covered in an English textbook and potentially aiding in content customization and topic relevance for personalized learning. In a personalized learning context, the goal is to tailor the educational experience to the individual needs and preferences of each learner. This involves understanding the learner's strengths, weaknesses, interests, and learning style. While LDA could be useful in some aspects of this process, it might not directly address all the requirements of personalized learning for an English textbook. In this research study we showed that LDA based topic modeling could be a solution to enhance the context understanding of the learners. However from the survey it is revealed that app was not sufficient. Learners oriented topic-document distribution to identify which topics are most relevant to a specific student can be provided. App can provide additional explanations, examples, or resources to cater to their individual learning style. Assessments and exercises focused on the topics that need reinforcement for each student. Monitor their progress and adjust the learning path accordingly. Analyze students' performance, engagement, and feedback to refine the topic modeling process and its integration into the learning environment. More sophisticated approaches, such as adaptive learning systems and AI-based tutoring, might be needed to truly personalize the learning experience in a comprehensive manner.

Special Remarks:

For the data privacy and security issues many Teachers were reluctant to provide their social website address to the surveyor. Among all the participants only 24% attendees provided their social sites address to use them publicly for research purposes.

Conclusion:

By employing topic modeling in a personalized learning context, educators can create more engaging and effective learning experience. This approach allows enhancement understanding and retention of the Textbook context. The school survey with prototype app reaffirmed its potential in learning experiences. It was revealed teachers/instructors would find it acceptable and appreciated if textbook information were presented using cutting edge NLP technology driven algorithms like LDA topic modeling in mobile app. The study concludes apps seem effective as they provide a personal and learner-centered learning opportunity ubiquitously. Reveal to user as a complementary essential material to learn English Textbook quickly and effectively. However, apps need to be improved by including collaborative form of learning.

References

[1] S. Report, "Schools of 0," *The Daily Star*, May 09, 2018. https://www.thedailystar.net/frontpage/schools-0-1573576 (accessed Jul. 28, 2023).

[2] W. B. Habib and T. S. Adhikary, "English, maths drag results down again," *The Daily Star*, May 07, 2018. https://www.thedailystar.net/frontpage/ssc-examination-result-2018-bangladesh-english-maths-drag-results-down-again-1572613 (accessed Jul. 28, 2023).

[3] "Bangladesh Education Statistics 2021." http://banbeis.portal.gov.bd/sites/default/files/files/banbeis.portal.gov.bd/npfblock/Bangladesh%20Education%20Statistics%202021_compressed-1-235.pdf (accessed Jul. 28, 2022).

[4] B. B. of E. I. and Statistics, *Bangladesh Educational Statistics 2016*, First Edition. Bangladesh Bureau of Educational Information and Statistics, 2017.

[5] J. Sandberg, M. Maris, and K. de Geus, "Mobile English learning: An evidence-based study with fifth graders," *Comput. Educ.*, vol. 57, no. 1, pp. 1334–1347, Aug. 2011, doi: 10.1016/j.compedu.2011.01.015.

[6] S. Hu, K. Laxman, and K. Lee, "Exploring factors affecting academics' adoption of emerging mobile technologies-an extended UTAUT perspective," *Educ. Inf. Technol.*, vol. 25, no. 5, pp. 4615–4635, Sep. 2020, doi: 10.1007/s10639-020-10171-x.

[7] R. Shadiev, T. Liu, and W.-Y. Hwang, "Review of research on mobile-assisted language learning in familiar, authentic environments," *Br. J. Educ. Technol.*, vol. 51, no. 3, pp. 709–720, 2020, doi: 10.1111/bjet.12839.

[8] S. F. Isamiddinovna, "Mobile Applications As A Modern Means Of Learning English," in *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, Nov. 2019, pp. 1–5. doi: 10.1109/ICISCT47635.2019.9011897.

[9] M. M. Elaish, L. Shuib, N. A. Ghani, and E. Yadegaridehkordi, "Mobile English Language Learning (MELL): a literature review," *Educ. Rev.*, vol. 71, no. 2, pp. 257–276, Mar. 2019, doi: 10.1080/00131911.2017.1382445.

[10] B. Klimova, "Impact of Mobile Learning on Students' Achievement Results," *Educ. Sci.*, vol. 9, no. 2, Art. no. 2, Jun. 2019, doi: 10.3390/educsci9020090.

[11] M. L. Bernacki, J. A. Greene, and H. Crompton, "Mobile technology, learning, and achievement: Advances in understanding and measuring the role of mobile technology in education," *Contemp. Educ. Psychol.*, vol. 60, p. 101827, Jan. 2020, doi: 10.1016/j.cedpsych.2019.101827.

[12] S. Criollo-C, A. Guerrero-Arias, Á. Jaramillo-Alcázar, and S. Luján-Mora, "Mobile Learning Technologies for Education: Benefits and Pending Issues," *Appl. Sci.*, vol. 11, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/app11094111.

[13] X. Chen, "Evaluating Language-learning Mobile Apps for Second-language Learners," *J. Educ. Technol. Dev. Exch.*, vol. 9, no. 2, Dec. 2016, doi: 10.18785/jetde.0902.03.

[14] V. N. Hoi, "Understanding higher education learners' acceptance and use of mobile devices for language learning: A Rasch-based path modeling approach," *Comput. Educ.*, vol. 146, p. 103761, Mar. 2020, doi: 10.1016/j.compedu.2019.103761.

[15] K. R. M. Rafiq, H. Hashim, and M. M. Yunus, "Sustaining Education with Mobile Learning for English for Specific Purposes (ESP): A Systematic Review (2012–2021)," *Sustainability*, vol. 13, no. 17, Art. no. 17, Jan. 2021, doi: 10.3390/su13179768.

[16] Z. Yu, W. Xu, and P. Sukjairungwattana, "Motivation, Learning Strategies, and Outcomes in Mobile English Language Learning," *Asia-Pac. Educ. Res.*, vol. 32, no. 4, pp. 545–560, Aug. 2023, doi: 10.1007/s40299-022-00675-0.

[17] Z. Xu, Z. Chen, L. Eutsler, Z. Geng, and A. Kogut, "A scoping review of digital game-based technology on English language learning," *Educ. Technol. Res. Dev.*, vol. 68, no. 3, pp. 877–904, Jun. 2020, doi: 10.1007/s11423-019-09702-2.

[18] Y. Hao, K. S. Lee, S.-T. Chen, and S. C. Sim, "An evaluative study of a mobile application for middle school students struggling with English vocabulary learning," *Comput. Hum. Behav.*, vol. 95, pp. 208–216, Jun. 2019, doi: 10.1016/j.chb.2018.10.013.

[19] B. Klímová and A. Berger, "Evaluation of the Use of Mobile Application in Learning English Vocabulary and Phrases – A Case Study," in *Emerging Technologies for Education*, T. Hao, W. Chen, H. Xie, W. Nadee, and R. Lau, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 3–11. doi: 10.1007/978-3-030-03580-8_1.

[20] C.-H. Chen and C.-C. Tsai, "In-service teachers' conceptions of mobile technology-integrated instruction: Tendency towards student-centered learning," *Comput. Educ.*, vol. 170, p. 104224, Sep. 2021, doi: 10.1016/j.compedu.2021.104224.

[21] I. García-Martínez, J. M. Fernández-Batanero, D. Cobos Sanchiz, and A. Luque de la Rosa, "Using Mobile Devices for Improving Learning Outcomes and Teachers' Professionalization," *Sustainability*, vol. 11, no. 24, Art. no. 24, Jan. 2019, doi: 10.3390/su11246917.

[22] H. Oz, "An Investigation of Preservice English Teachers' Perceptions of Mobile Assisted Language Learning," *Engl. Lang. Teach.*, vol. 8, no. 2, pp. 22–34, 2015.

[23] J. Kacetl and B. Klímová, "Use of Smartphone Applications in English Language Learning—A Challenge for Foreign Language Education," *Educ. Sci.*, vol. 9, no. 3, Art. no. 3, Sep. 2019, doi: 10.3390/educsci9030179.

[24] Z. Jie and Y. Sunze, "Investigating pedagogical challenges of mobile technology to English teaching," *Interact. Learn. Environ.*, vol. 31, no. 5, pp. 2767–2779, Jul. 2023, doi: 10.1080/10494820.2021.1903933.

[25] R. Metruk, "The Use of Smartphone English Language Learning Apps in the Process of Learning English: Slovak EFL Students' Perspectives," *Sustainability*, vol. 13, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/su13158205.

[26] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020," *Comput. Assist. Lang. Learn.*, vol. 36, no. 3, pp. 517–554, Mar. 2023, doi: 10.1080/09588221.2021.1933540.

[27] P. Poláková and B. Klímová, "Mobile Technology and Generation Z in the English Language Classroom—A Preliminary Study," *Educ. Sci.*, vol. 9, no. 3, Art. no. 3, Sep. 2019, doi: 10.3390/educsci9030203.

[28] R. Kaliisa, E. Palmer, and J. Miller, "Mobile learning in higher education: A comparative analysis of developed and developing country contexts," *Br. J. Educ. Technol.*, vol. 50, no. 2, pp. 546–561, 2019, doi: 10.1111/bjet.12583.

[29] A. Kao and S. R. Poteet, *Natural Language Processing and Text Mining*. Springer Science & Business Media, 2007.

[30] P. M. McCarthy and C. Boonthum-Denecke, Eds., *Applied Natural Language Processing: Identification, Investigation and Resolution*. IGI Global, 2012. doi: 10.4018/978-1-60960-741-8.

[31] "Welcome to LIWC-22." https://www.liwc.app/ (accessed May 06, 2023).

[32] "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods - Yla R. Tausczik, James W. Pennebaker, 2010." https://journals.sagepub.com/doi/abs/10.1177/0261927x09351676 (accessed Jul. 28, 2023).

[33] J. Guerra, S. Sosnovsky, and P. Brusilovsky, "When One Textbook Is Not Enough: Linking Multiple Textbooks Using Probabilistic Topic Models," in *Scaling up Learning for Sustained Impact*, D. Hernández-Leo, T. Ley, R. Klamma, and A. Harrer, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 125–138. doi: 10.1007/978-3-642-40814-4_11.

[34] S. Slater, R. Baker, Ma. V. Almeda, A. Bowers, and N. Heffernan, "Using correlational topic modeling for automated topic identification in intelligent tutoring systems," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver British Columbia Canada: ACM, Mar. 2017, pp. 393–397. doi: 10.1145/3027385.3027438.

[35] K. M. Alam, Md. T. H. Hemel, S. M. Muhaiminul Islam, and A. Akther, "Bangla News Trend Observation using LDA Based Topic Modeling," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Dec. 2020, pp. 1–6. doi: 10.1109/ICCIT51783.2020.9392719.

[36] J. W. Lee, Y. Kim, and D. H. Han, "LDA-based topic modeling for COVID-19-related sports research trends," *Frontiers in Psychology*, vol. 13, 2022, Accessed: Aug. 26, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1033872

[37] R. K. Gupta, R. Agarwalla, B. H. Naik, J. R. Evuri, A. Thapa, and T. D. Singh, "Prediction of research trends using LDA based topic modeling," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 298–304, Jun. 2022, doi: 10.1016/j.gltp.2022.03.015.

[38] "Topic Predictions and Optimized Recommendation Mechanism Based on Integrated Topic Modeling and Deep Neural Networks in Crowdfunding Platforms." https://www.mdpi.com/2076-3417/9/24/5496 (accessed Aug. 29, 2023).

[39] A. Farkhod, A. Abdusalomov, F. Makhmudov, and Y. I. Cho, "LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model," *Applied Sciences*, vol. 11, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/app112311091.

[40] H. Wang, K. Sun, and Y. Wang, "Exploring the Chinese Public's Perception of Omicron Variants on Social Media: LDA-Based Topic Modeling and Sentiment Analysis," *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, Art. no. 14, Jan. 2022, doi: 10.3390/ijerph19148377.

[41] K. Raj P M and J. Sai D, "Sentiment analysis, opinion mining and topic modelling of epics and novels using machine learning techniques," *Materials Today: Proceedings*, vol. 51, pp. 576–584, Jan. 2022, doi: 10.1016/j.matpr.2021.06.001.

[42] R. Rani and D. K. Lobiyal, "An extractive text summarization approach using tagged-LDA based topic modeling," *Multimed Tools Appl*, vol. 80, no. 3, pp. 3275–3305, Jan. 2021, doi: 10.1007/s11042-020-09549-3.

[43] David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing Semantic Coherence in Topic Models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 262–272, Edinburgh, Scotland, UK.. Association for Computational Linguistics.

paper content.pdf

ORIGINALITY REPORT

27%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.researchgate.net Internet	204 words — 3%
2	www.mdpi.com Internet	108 words — 1%
3	download.atlantis-press.com Internet	75 words — 1%
4	www.machinelearningplus.com Internet	72 words — 1%
5	link.springer.com Internet	62 words — 1%
6	downloads.hindawi.com Internet	53 words — 1%
7	www.jatit.org Internet	46 words — 1%
8	beei.org Internet	42 words — 1%
9	dspace.bracu.ac.bd Internet	42 words — 1%
10	aclanthology.org Internet	

41 words — 1%

11 journals.riverpublishers.com
Internet
40 words — 1%

12 www.frontiersin.org
Internet
40 words — 1%

13 e-journal.uajy.ac.id
Internet
38 words — 1%

14 Divakar Yadav, Rishabh Katna, Arun Kumar Yadav, Jorge Morato. "Feature Based Automatic Text Summarization Methods: A Comprehensive State-of-the-Art Survey", IEEE Access, 2022
Crossref
36 words — < 1%

15 Rahul Kumar Gupta, Ritu Agarwalla, Bukya Hemanth Naik, Joythish Reddy Evuri, Apil Thapa, Thoudam Doren Singh. "Prediction of research trends using LDA based topic modeling", Global Transitions Proceedings, 2022
Crossref
35 words — < 1%

16 diglib.eg.org
Internet
35 words — < 1%

17 Bandi Rupendra Reddy, Darukumalli Sai Tharun Reddy, M C Sandeep Preetham, A H N Rajasekhar, R Subramani. "Comparative Study Analysis on News Articles Categorization using LSA and NMF Approaches", 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022
Crossref
34 words — < 1%

18	Catherine Akoth Ongoro, Yong-Yi Fangjiang. "Digital Game-Based Technology for English Language Learning in Preschools and Primary Schools: A Systematic Analysis", IEEE Transactions on Learning Technologies, 2023 Crossref	34 words — < 1%
19	www.ijair.id Internet	34 words — < 1%
20	journals.iium.edu.my Internet	31 words — < 1%
21	vbs.rs Internet	29 words — < 1%
22	www.hindawi.com Internet	28 words — < 1%
23	www.politesi.polimi.it Internet	28 words — < 1%
24	www.atlantis-press.com Internet	27 words — < 1%
25	www.slideshare.net Internet	27 words — < 1%
26	"Scaling up Learning for Sustained Impact", Springer Nature, 2013 Crossref	26 words — < 1%
27	ccs.in Internet	25 words — < 1%
28	eudl.eu Internet	24 words — < 1%

29	msrit-bucket.s3.us-west-2.amazonaws.com Internet	23 words — < 1%
30	www.ijraset.com Internet	23 words — < 1%
31	Lydia Maketo, Tomayess Issa, Theodora Issa, S. Zaung Nau. "M-Learning adoption in higher education towards SDG4", Future Generation Computer Systems, 2023 Crossref	22 words — < 1%
32	hal.inria.fr Internet	22 words — < 1%
33	insights2techinfo.com Internet	22 words — < 1%
34	ouci.dntb.gov.ua Internet	22 words — < 1%
35	socj.telkomuniversity.ac.id Internet	22 words — < 1%
36	www.ijert.org Internet	21 words — < 1%
37	Shadiev Rustam, Liu Jiawen, Cheng Pei-Yu. "The Impact of Mobile-Assisted Social Language Learning Activities on Speaking Skills and Self-Efficacy Development", IEEE Transactions on Learning Technologies, 2023 Crossref	20 words — < 1%
38	avesis.hacettepe.edu.tr Internet	20 words — < 1%

39	repository.uin-malang.ac.id Internet	20 words — < 1%
40	Alexandru Garaba, Nummra Aslam, Francesco Ponzio, Abdullah Al Awadhi et al. "Radiomics for differentiation of gliomas from primary central nervous system lymphomas: a systematic review and meta-analysis", Research Square Platform LLC, 2023 Crossref Posted Content	18 words — < 1%
41	Oyshi, Uttamasha Anjally. "Topic Modeling and Prediction of Aid Data in Development Studies Using LDA and BERT", University of Arkansas at Little Rock, 2023 ProQuest	18 words — < 1%
42	www.saf21.eu Internet	18 words — < 1%
43	Jia-Cing Liang, Gwo-Jen Hwang. "A robot-based digital storytelling approach to enhancing EFL learners' multimodal storytelling ability and narrative engagement", Computers & Education, 2023 Crossref	17 words — < 1%
44	Lan Thi Nguyen, Issara Kanjug, Grichawat Lowatcharin, Theeradej Manakul et al. "How teachers manage their classroom in the digital learning environment – experiences from the University Smart Learning Project", Heliyon, 2022 Crossref	16 words — < 1%
45	csap.snu.ac.kr Internet	16 words — < 1%
46	export.arxiv.org Internet	16 words — < 1%

47 Harada, Tomohiko, Kazuhiko Tsuda, Nobuo Suzuki, and Yoshikatsu Fujita. "The estimate method of the omission of Japanese inquiry texts using an LDA algorithm", International Journal of Computer Applications in Technology, 2015.
Crossref
15 words — < 1%

48 Kazi Masudul Alam, Md. Tanvir Hossain Hemel, S.M. Muhaiminul Islam, Aysha Akther. "Bangla News Trend Observation using LDA Based Topic Modeling", 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020
Crossref
15 words — < 1%

49 Mohamed Trabelsi, Zhiyu Chen, Brian D. Davison, Jeff Heflin. "Relational Graph Embeddings for Table Retrieval", 2020 IEEE International Conference on Big Data (Big Data), 2020
Crossref
15 words — < 1%

50 Spitters, Martijn, Stefan Verbruggen, and Mark van Staalduinen. "Towards a Comprehensive Insight into the Thematic Organization of the Tor Hidden Services", 2014 IEEE Joint Intelligence and Security Informatics Conference, 2014.
Crossref
15 words — < 1%

51 scholar.afit.edu
Internet
14 words — < 1%

52 Samson Ebenezar Uthirapathy, Domnic Sandanam. "Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model.", Procedia Computer Science, 2023
Crossref
13 words — < 1%

53	"Open Learning and Teaching in Educational Communities", Springer Science and Business Media LLC, 2014 Crossref	12 words — < 1%
54	Sarah Hussain, Anubha Mahender Singh, Priyakrushna Mohanty, Mahender Reddy Gavinolla. "Next generation employability andcareer sustainability inthehospitality industry 5.0", Worldwide Hospitality and Tourism Themes, 2023 Crossref	12 words — < 1%
55	ikee.lib.auth.gr Internet	12 words — < 1%
56	aisel.aisnet.org Internet	11 words — < 1%
57	Jeya Amantha Kumar, Sharifah Osman, Mageswaran Sanmugam, Rasammal Rasappan. "Mobile Learning Acceptance Post Pandemic: A Behavioural Shift among Engineering Undergraduates", Sustainability, 2022 Crossref	10 words — < 1%
58	archiv.ub.uni-heidelberg.de Internet	10 words — < 1%
59	repository.bsi.ac.id Internet	10 words — < 1%
60	Clark, Ryan D.. "Digital Scriptures, Perceived Spiritual and Learning Experiences: A Qualitative Descriptive Study", Grand Canyon University, 2023 ProQuest	9 words — < 1%
61	patents.justia.com Internet	9 words — < 1%

62 Fumiya Okubo, Tetsuya Shiino, Tsubasa Minematsu, Yuta Taniguchi, Atsushi Shimada. "Adaptive Learning Support System Based on Automatic Recommendation of Personalized Review Materials", IEEE Transactions on Learning Technologies, 2022 — 8 words — < 1%

Crossref

---

63 G. Manning Richardson, Janet Bowers, A. John Woodill, Joseph R. Barr, Jean Mark Gawron, Richard A. Levine. "Topic Models: A Tutorial with R", International Journal of Semantic Computing, 2014 — 8 words — < 1%

Crossref

---

64 Liu, Shixia, Michelle X. Zhou, Shimei Pan, Yangqiu Song, Weihong Qian, Weijia Cai, and Xiaoxiao Lian. "TIARA : Interactive, Topic-Based Visual Text Summarization and Analysis", ACM Transactions on Intelligent Systems and Technology, 2012. — 8 words — < 1%

Crossref

---

65 M. Ileana Ruiz-Cantisani, Jorge Membrillo-Hernández, Patricia Caratozzolo, Vianney Lara-Prieto. "Towards the gender equity in Education: Strategies to develop equity leadership in higher education institution", 2023 IEEE Global Engineering Education Conference (EDUCON), 2023 — 8 words — < 1%

Crossref

---

66 Manju Bhardwaj, Priya Mishra, Shikha Badhani, Sunil K. Muttoo. "Sentiment analysis and topic modeling of COVID-19 tweets of India", International Journal of System Assurance Engineering and Management, 2023 — 8 words — < 1%

Crossref

---

67 Xiaosheng Liu, Jia Zeng, Xi Yang, Jianfeng Yan, Qiang Yang. "Scalable Parallel EM Algorithms for Latent Dirichlet Allocation in Multi-Core Systems", Proceedings — 8 words — < 1%

of the 24th International Conference on World Wide Web - WWW '15, 2015
Crossref

68	ajet.org.au Internet	8 words — < 1%
69	documentserver.uhasselt.be Internet	8 words — < 1%
70	icrera.org Internet	8 words — < 1%
71	osuva.uwasa.fi Internet	8 words — < 1%
72	uknowledge.uky.edu Internet	8 words — < 1%
73	www.aclweb.org Internet	8 words — < 1%
74	www.adamhousepress.com.au Internet	8 words — < 1%
75	www.escholar.manchester.ac.uk Internet	8 words — < 1%
76	"Technology in Education. Innovative Solutions and Practices", Springer Science and Business Media LLC, 2018 Crossref	7 words — < 1%
77	Krishna Raj P M, Jagadeesh Sai D. "Sentiment analysis, opinion mining and topic modelling of epics and novels using machine learning techniques", Materials Today: Proceedings, 2021 Crossref	7 words — < 1%

78	Lise Claiborne, Vishalache Balakrishnan. "Moving towards Inclusive Education", Brill, 2020 Crossref	7 words — < 1%
79	Zhixia Yang, Haiyan Men, Richard Ingham. "Media Representation of China in COVID-19 Reports: Text Mining of the Language of New York Times", 2022 European Conference on Natural Language Processing and Information Retrieval (ECNLPIR), 2022 Crossref	7 words — < 1%
80	"Implicit aspect based sentiment analysis for restaurant review using LDA topic modeling and ensemble approach", International Journal of Advanced Technology and Engineering Exploration, 2023 Crossref	6 words — < 1%
81	Alexandre Rekeraho, Daniel Tudor Cotfas, Petru Adrian Cotfas, Titus Constantin Bălan, Emmanuel Tuyishime, Rebecca Acheampong. "Cybersecurity challenges in IoT-based smart renewable energy", Research Square Platform LLC, 2023 Crossref Posted Content	6 words — < 1%
82	Govind Sharma. "Mining Sentiments from Songs Using Latent Dirichlet Allocation", Lecture Notes in Computer Science, 2011 Crossref	6 words — < 1%
83	Priyanka Thakral, Praveen Ranjan Srivastava, Sanket Sunand Dash, Sajjad M. Jasimuddin, Zuopeng (Justin) Zhang. "Trends in the thematic landscape of HR analytics research: astructural topic modeling approach", Management Decision, 2023 Crossref	6 words — < 1%