Mechanism of Using Natural Language Processing Techniques in Developing Interactive E-Book

آلية استخدام تقنيات معالجة اللغة الطبيعية في تطوير الكتاب الإلكتروني التفاعلي

Prof.Dr. Elsaeed Elsaeed Mohamed Abd El-Razek

Computer Teacher Preparation Department
Faculty of Specific Education
Damietta University

Dr. Safaa Mahmoud Mustafa Al-Atwi

Computer Teacher Preparation Department
Faculty of Specific Education
Damietta University

Tasneem Mostafa Gaweesh

Computer Teacher Preparation Department
Faculty of Specific Education
Damietta University

Mechanism of Using Natural Language Processing Techniques in Developing Interactive E-Book

Elsaeed E. AbdElrazek, Safaa M.ElAtawy, Tasneem M.Gaweesh

Computer Teacher Preparation Department, Faculty of Specific Education, Damietta University

Abstract:

This research describes how to use natural language processing techniques to develop an interactive e-book using Latent Dirichlet Allocation (LDA) technology, which is a widely used method to analyze, understand, and summarize large sets of textual data and discover the underlying topics in them. The purpose of the proposed system is to extract keywords from the paragraphs of the interactive e-book as they represent a summary of educational text data. The proposed system is an educational platform through which the student can view or download the interactive e-book and learn about the extracted keywords. The proposed system was tested on a sample of 40 male and female students from the Computer department at the College of Specific Education, the results showed positive student attitudes towards the proposed system, the importance of using natural language processing techniques in the educational process to enhance student participation interests, and interaction with the content.

Keywords: NLP, Interactive E-Book, LDA**,** Keywords.

1. Introduction:

Universities have recently experienced a multitude of rapid changes in the realm of scientific and technological advancements. These changes have compelled them to adapt their programs, methodologies, and teaching strategies to adhere to quality standards and foster an educational system that aligns with the realities of various domains. Just like other sectors, higher education has embarked on a quest to explore technology, particularly harnessing the potential of artificial intelligence (AI) technology, which has pervaded every facet of our lives [1].

The convenience of e-books has transformed them into a viable substitute for printed or photocopied paper books. Their accessibility and swift distribution have made them immensely popular among users. With just a few clicks, any book can be accessed and circulated rapidly by individuals worldwide, regardless of their proximity to the physical book's location. Moreover, e-books offer the added advantage of being available in multiple languages, allowing readers to translate and read them in their preferred language. This flexibility has further enhanced the appeal and utility of e-books in the digital age [2].

Natural language processing (NLP) is an interdisciplinary approach that merges language, computer science, and artificial intelligence. It is referred to as "natural language" because it involves human-to-smart device communication using everyday language [3]. The applications of NLP include machine translation, question answering, text summarization, sentiment analysis, named entity recognition, and designing/implementing models, systems, and algorithms to solve practical problems in understanding human languages [4].

Keyword formation is of utmost importance as it provides a concise representation of the content within data. Furthermore, it plays a pivotal role in information retrieval systems and search engine optimization. By facilitating more efficient information searches, keywords assist users in locating relevant information quickly. Additionally, keywords aid readers in making informed decisions about whether to read a document or not. Their strategic implementation enhances the accessibility and discoverability of information, ensuring that users can easily find and assess the relevance of documents [5].

Latent Dirichlet Allocation (LDA) is a probabilistic model based on unsupervised learning, which assumes each document in a corpus is a random mixture of latent topics, and each topic has a probability distribution over all words in the vocabulary. LDA is based on the idea that each document contains

several hidden topics, each of which contains a set of words related to the topic [6] [7].

By examining previous studies such as [8] [9], it becomes evident that these studies have demonstrated the effectiveness of e-books in teaching academic subjects, particularly with regard to academic achievement. These studies have emphasized the significance of e-books in enriching the educational process. In a study conducted by [10], the importance of utilizing artificial intelligence techniques in education was highlighted. Such techniques aim to provide students with optimal learning experiences and equip them with scientific skills. Furthermore, a study conducted by [11] highlighted the importance of integrating natural language processing into interactive e-books due to its profound impact and usefulness in the learning process.

In this research, a platform for an interactive e-book was developed, utilizing natural language processing techniques. The goal was to extract keywords from the interactive e-book, aiding students in comprehending the core concepts and content of the e-book they are reading. The application of natural language processing facilitated the identification and extraction of important keywords, enhancing the overall learning experience for students.

2. Objectives of the study:

The primary goals of this study were:

- Reaching the stages of developing an interactive e-book based on natural language processing techniques.
- Providing teachers with an easy-to-use smart software tool.
- Providing an interactive e-book based on natural language processing techniques.

The importance of research also stems from importance of the e-book, as it is a means that can be used to obtain information quickly that enables it to keep pace with technological developments in human cognitive field [12], and to benefit from artificial intelligence techniques, especially NLP, because of its advantages, the most important of which is ability to learn, possibility of collecting and analyzing data, knowing relationship between them, and making decisions based on the process of analyzing information. It is also relatively permanent less expensive and effortless [13].

3. Research Problem:

The sense of the problem stemmed from previous studies, such as the study by [14], emphasizes the effectiveness of natural language processing (NLP), particularly text mining, in extracting valuable information from educational texts. Second, research by [1] [2] demonstrates the positive impact of interactive e-books on student learning, as evidenced by improved cognitive achievement and pre/post-test performance. Finally, an exploratory study conducted by the researcher with 30 fourth-year students during the 2020-2021 academic year revealed that 85% of their electronic courses lacked engaging elements and did not leverage artificial intelligence (AI) techniques. These findings illuminate a significant gap in current e-learning resources, highlighting the need for an AI-powered, NLP-based interactive e-book to enhance the educational process.

The research problem is the need to develop an interactive e-book based on NLP techniques.

This problem can be formulated in the following question:

- What are the stages of developing an interactive e-book based on natural language processing techniques?
- What are students' attitudes toward the proposed system?

4. Theoretical framework:
4.1 Interactive e-book:

E-books, a technologically enhanced form of publishing, are at the forefront of advancing educational systems [15]. They provide easier and higher-quality access to knowledge compared to traditional printed materials [16]. Furthermore, their digital format allows for interactive learning experiences, making them valuable tools for in-class lectures through navigation and interaction with the content [17].

The study conducted by [18] emphasized significance of incorporating interactive elements into e-books. This transformation allows reader to become an active participant in reading process. By introducing interactive features, e-books provide individuals with a tool that enables them to engage in various functions within the e-book, without the need for extensive programming knowledge. This approach enhances the user experience, making e-books more dynamic and engaging for readers.

A study by [19] investigated the use of e-books in learning through a survey. The findings indicated positive student responses and potential for future

development in both e-book use and content. However, the research also acknowledged that some students still prefer printed books.

4.2 Natural language processing:

Natural language processing encompasses a range of techniques used to automatically analyze and represent human languages. However, the ability of machines to truly understand natural language remains a distant goal. While there have been notable examples of NLP applications such as aggregation, online information retrieval, and question-answering, these have primarily relied on algorithmic approaches. Algorithms excel at tasks like text retrieval (IR), spell-checking, tokenization, and word-level analysis. However, they face significant challenges when it comes to analyzing sentences and paragraphs comprehensively. The current capabilities of these algorithms are still limited in terms of interpreting sentences and extracting meaningful information [20].

Study [21] presented a system that utilizes NLP to generate summaries and extract keywords from websites. This approach empowers users to select websites efficiently by providing summaries of both individual links and aggregated collections based on their specific needs. Furthermore, the system's keyword extraction function plays a crucial role in delivering the essence of entire documents or websites within seconds.

A recent study [11] explored the potential of using large language models (LLMs) like GPT-3 and OpenAI Codex to generate code explanations for an interactive e-book on web software development. The system provided access to explanations through buttons, offering students three options: line-by-line explanations, lists of key concepts, and high-level code summaries. User interaction data revealed that students viewed all explanation types. While the majority found the explanations helpful, engagement varied based on factors such as code complexity, explanation type, and code length.

Study [22] proposed an interactive reading platform that leverages Natural Language Processing (NLP) Transformer models, such as T5, DistilBERT, and DistilBART, to enhance the reading experience. Users can download e-books, generate automatic text summaries, and answer system-generated questions. This platform aims to reduce reading time by providing concise summaries and engaging questionnaires, all powered by existing, high-accuracy NLP algorithms.

5. Methodology:

Reviewing literature and previous studies on natural language processing and its use in education and developing interactive e-books. LDA technology was used to develop the interactive e-book by extracting keywords to identify the content of the educational text. In this context, the methodology consists of two main stages: producing the interactive e-book and extracting keywords. Each stage of this methodology is described in more detail in the following section:

5.1 Proposed system:

The educational platform offers access to a variety of learning resources for both students and teachers. Users can log in to the main page using the credentials provided to them.

When a student user selects educational content to learn, an interactive e-book is displayed. The student can download the e-book to their device and utilize various features, such as:

- Adding comments

- Reviewing the educational objectives

- Taking assessments

- Participating in activities

To enhance the learning experience, the content is divided into sentences, and the key words are identified. The platform provides solved examples for reference, as well as opportunities for the student to practice with unsolved examples. Students can also take notes and communicate with their teachers to obtain additional learning resources.

The proposed system was applied to a sample of 40 male and female students from the third year of the Computer Department at the Faculty of Specific Education - Damietta.
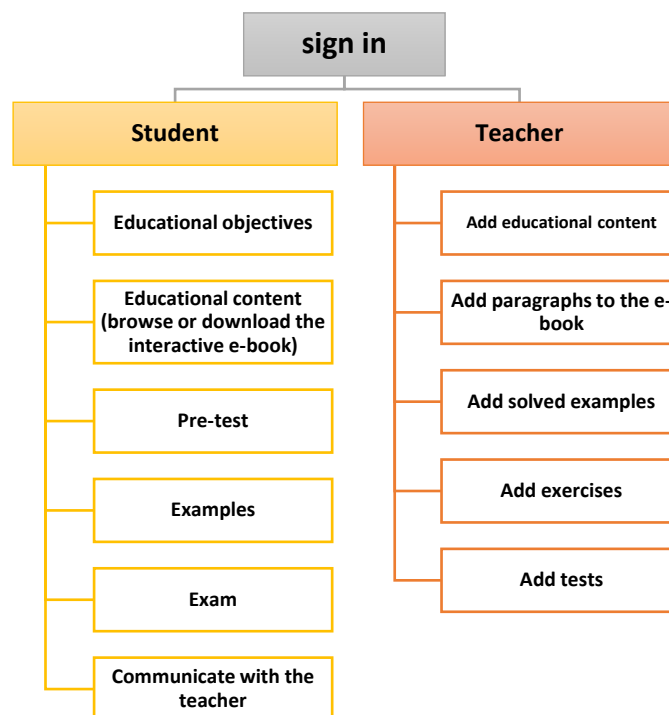
Figure 1 General structure of the proposed system

5.2 Implementation of the system:

Flip PDF Professional program was used to produce the e-book that was uploaded to the educational platform. The platform was built using the Django framework for the backend, with a MySQL database. HTML, CSS, and JavaScript were leveraged for the front-end pages, while Python was employed to extract keywords.

5.2.1 First: Producing the interactive e-book:

The e-book was produced according to the following steps:

1. **Designing interactive e-book pages:** This involved adhering to e-book design standards in terms of simplicity, balance of elements, clarity of pages and links, multimedia efficiency, and the use of various elements.

2. **Identifying the educational resources necessary to produce the interactive e-book:** These included written texts, which were authored using Microsoft Office Word. Additionally, educational videos and images were incorporated to express the content of the e-book.

3. **Producing the interactive e-book in its initial form:** After preparing the content, the e-book was produced in a format that could be played on any computer. The initial version was then presented to a sample of reviewers to approve the content and provide feedback on the electronic form of the book.

5.2.2 Second: Extracting keywords:

The process of extracting keywords is considered one of the important steps in the e-book production. This process involves highlighting important words, phrases, and expressions within the specific content. The extraction of keywords is facilitated through the use of natural language processing techniques [21].

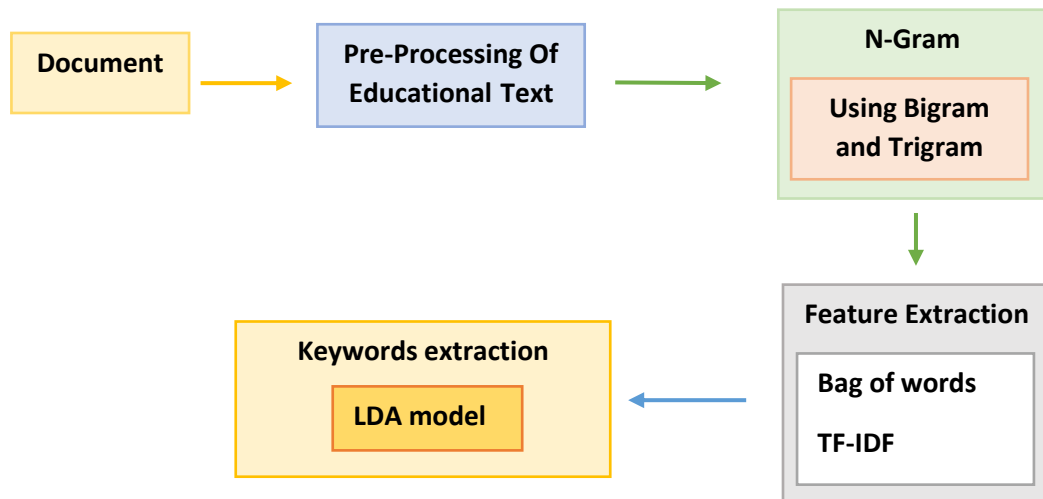The following figure shows the steps for extracting keywords:

Figure 2: Keyword Extraction Phases Using LDA Model

5.2.2.1 Data:

The computer applications book was designed as an interactive e-book and uploaded to the educational platform for use in instruction. The content of the book was stored in a text file format, which enabled the extraction of keywords from the content.

5.2.2.2 Pre-processing of educational text:

The educational text data provided was in its raw form, which can contain noise, errors, and unwanted content that may impact the efficiency of analysis. To obtain better results, it is necessary to preprocess the textual data [23]. This involves several steps to extract accurate hidden information and improve the data for further understanding and analysis. Text preprocessing steps include: Dividing the text into sentences, removing stop words, special characters, numbers, and punctuation marks and reducing words to their root forms.

Properly preprocessing the raw educational text data helps to remove irrelevant information and prepare the content for more accurate keyword extraction and analysis.

These preprocessing tasks can be accomplished by importing the Re and NLTK libraries in Python. The step-by-step approach to this textual data preprocessing is illustrated in the figure.
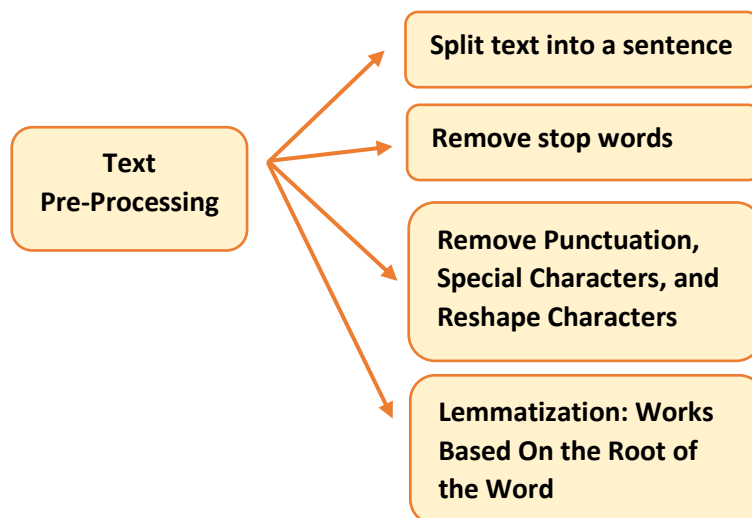
```
                          ┌──────────────────────────────┐
                          │  Split text into a sentence  │
                          └──────────────────────────────┘
                          ┌──────────────────────────────┐
                          │      Remove stop words        │
┌──────────────┐          └──────────────────────────────┘
│    Text      │          ┌──────────────────────────────┐
│ Pre-Processing│──┬──────│  Remove Punctuation,          │
└──────────────┘          │  Special Characters, and      │
                          │  Reshape Characters           │
                          └──────────────────────────────┘
                          ┌──────────────────────────────┐
                          │  Lemmatization: Works         │
                          │  Based On the Root of         │
                          │  the Word                     │
                          └──────────────────────────────┘
```

Figure 3 Pre-processing steps of educational text

5.2.2.3 N-Gram:

N-gram is a set of n elements that represent a sequence, which is used to model text data [24]. N-gram analysis must be considered in the statistical processing of keywords, as they often consist of one or more words within documents. Degree of a unigram is equivalent to the frequency or weight of a single word, as a unigram comprises one word. In contrast, a bigram consists of two words, while a trigram is composed of three words. Gensim library was utilized to map and analyze bigrams and trigrams within the text [25]. Incorporating n-gram analysis provides a more comprehensive understanding of the keyword structure and relationships present in the educational content.

5.2.2.4 Feature extraction:

Feature extraction is where the document is represented as an ordered vector of words. This vector is then converted into numerical features that characterize the text; Features range in complexity from length and word frequency to word order and patterns [26].

- **Bag of Words (BOW):**

BOW model offers a simplified approach to representing educational text. It treats each unique word as an independent entity, focusing on word frequency

rather than order, grammar, or semantic relationships. It is a very easy algorithm that simply counts the number of times each word appears in the text [27].

- **Term Frequency - Inverse Document Frequency (TF-IDF):**

It is a statistical measure to evaluate the relevance of a word to a document in a set of documents. Each word is assigned a weight indicating its importance in the document. It checks how important the word is throughout the group. This weight is calculated by multiplying two factors: Term Frequency (TF) and Inverse Document Frequency (IDF). TF simply counts how often a word appears in the document, while IDF considers how rare the word is across the entire corpus. Words that are frequent within a single document but uncommon overall receive a higher TF-IDF weight, indicating their relevance to that specific document [28]. TFIDF appears in Eq1.

$$w_{a,b} = tf_{a,b} \, xLog \, \frac{N}{df_b}$$

Where TF (a, b) is the number of occurrences of a (word) in b (document), DF (b) is the number of documents containing a, and N is the total number of documents (collection) [29].

5.2.2.5 Extract keywords:

Latent Dirichlet Allocation (LDA)

LDA is a probabilistic and generative model and is one of the modeling algorithms commonly used in text mining. The term "latent" in LDA is related to the discovery of the semantic content of documents through the analysis of latent semantic structures within documents. The generative approach in LDA is defined as the process of mapping words in a document to random variables and grouping them semantically using a recursive probabilistic process based on a Dirichlet distribution [30]. The LDA model is a model for unsupervised machine learning. For LdaModel module of Genism library in Python, the topic number K, the initial value of the topic distribution a, and the prior value of the distribution of topic words n involved in topic modeling must be determined in advance.
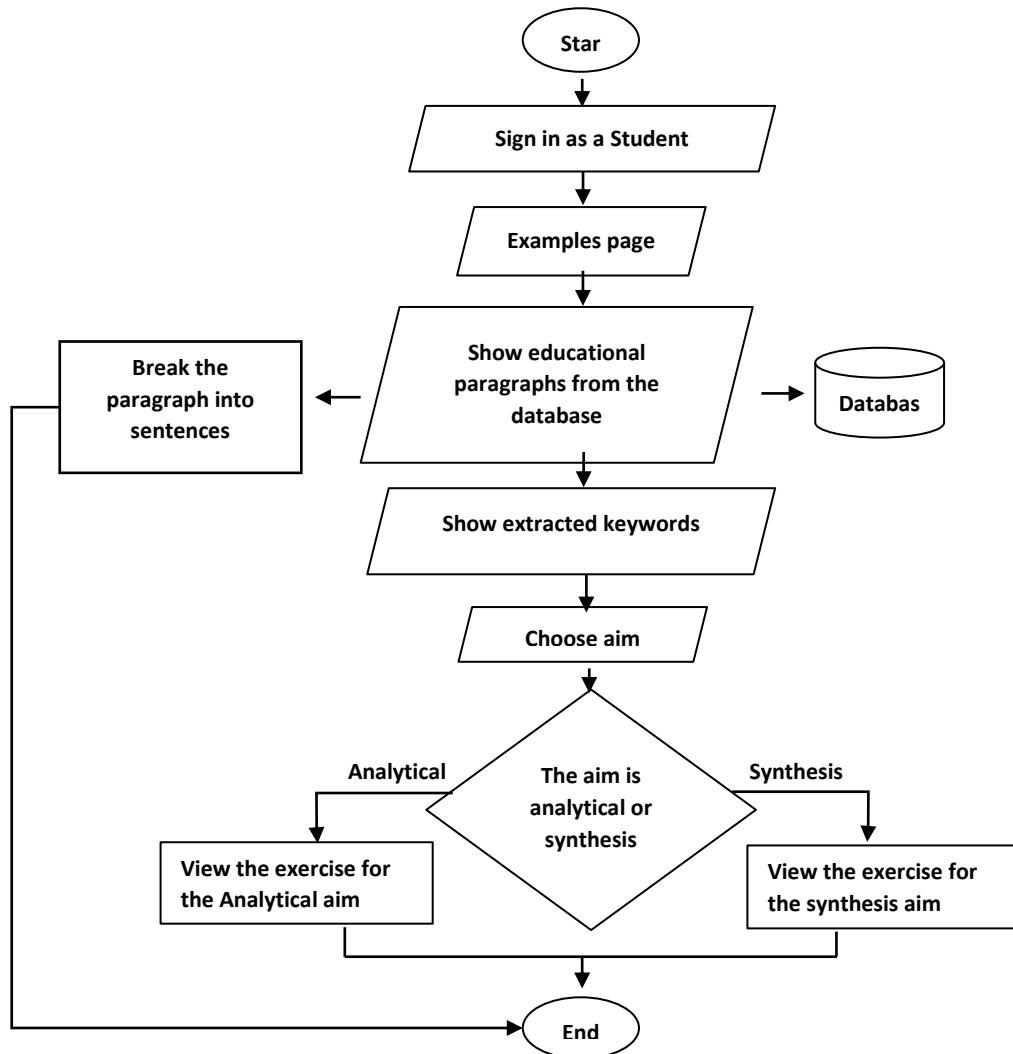
Figure 4 Flowchart for extracting keywords from interactive e-book paragraphs

6. Results and Discussion:

- To answer first question which states: What are the stages of developing an interactive e-book based on natural language processing techniques?

This question was answered within the proposed system for the study, and the researcher reached the stages of developing an E-book based on NLP techniques, which are as follows:

1- **Producing Interactive E-Book**: where the educational content was determined, the multimedia such as videos and images were determined, the e-book pages were designed, and then the e-book was initially produced.

2- **Extracting Keywords:** Four stages were reached to extract keywords from the educational text, which are as follows:

- **Pre-Processing Of Educational Text:** At this stage, we need to clean the data to make it free of noise and to improve performance and obtain the best results. Pre-processing of the text includes dividing the text into sentences and removing paused words, punctuation marks, and special letters and numbers.
- **N-Gram**: At this stage, we perform statistical analysis of the educational text, by dividing the sentence into parts of words, which are also used to represent the educational text.
- **Feature Extraction:** As it is one of the important stages of converting educational text data into quantitative vectors, a different set of features was used, namely (BOW), and (TF-IDF).
- **Keyword Extraction**: We used LDA model, which is an unsupervised probabilistic model that extracts latent topics from text documents.

- **To answer the second question: What are students' attitudes toward the proposed system?**

   After reviewing previous studies, the researcher constructed a questionnaire for students' attitudes toward the proposed system that included (10) items. The measurement was corrected according to the three-way Likert system, where three alternatives were specified for each item in the questionnaire, ranging from approval to disapproval. A weight was placed for each alternative. Of the alternatives, they were recorded as follows:

Alternatives	Agree	Neutral	Disagree
Correction degree	٣	٢	١

The answers were classified into three levels of equal range through the following equation:

Calculating range = largest value - lowest value

Calculating category length = dividing the range by the number of classes

Therefore, length of category is (0.66), which is added to the smallest value in measurement, which is (1), to obtain the first category, so the first category was from (1 to 1.66), then adding length of the first category to second category to obtain it, as in the table the next :

Average value categories	Attitude
1 to 1.66	**Disagree**
1.67 to 2.32	**Neutral**
2.33 to 2.89	**Agree**

- The validity of questionnaire was verified by presenting it to a group of arbitrators specialized in field of computer science and use of computers in education. The arbitrators made important observations, and necessary

amendments were made. The paragraphs on which arbitrators agreed were valid so that number of questionnaire elements became (10) elements.

- The reliability of attitudes questionnaire was calculated according to Cronbach's alpha equation using SPSS program, where the value of reliability coefficient was (0.74), and this value is considered acceptable in terms of degree of confidence in the scale.
- To judge the attitude, the researcher dealt with categories (disagree, neutral) as indicating a negative attitude if mean fell within them, and with category (Agree) as indicating a positive attitude if mean fell within them. To verify research question, arithmetic means were calculated. The standard deviations of responses of students in research sample to the attitudes questionnaire are as shown below:

Table 1 : Mean and standard deviation of the attitude measure of the proposed system.

No.	statements Attitude	Agree		Neutral		Disagree		mean	standard deviation	Rank	General Attitude
		F	%	F	%	F	%				
1	Were you able to use the proposed system easily؟	38	95.0	2	5.0	0	0	2.95	0.220	1.5	**Agree**
2	The proposed system helps to understand and comprehend the content	37	92.5	2	5.0	1	2.5	2.90	0.378	6.5	**Agree**
3	The proposed system contributed to understanding the information	38	95.0	2	5.0	0	0	2.95	٠.220	1.5	**Agree**
4	The proposed system helped link learning to reality	28	70.0	11	27.5	1	2.5	2.76	0.525	10	**Agree**
5	The system allows you to review lessons quickly and easily	37	92.5	3	7.5	0	0	2.92	٠.266	4	**Agree**
6	The proposed system helps to identify and color keywords in educational paragraphs	37	92.5	3	7.5	0	0	2.92	٠.266	4	**Agree**
7	The proposed system provides feedback	33	82.5	7	17.5	0	0	2.82	0.381	8	**Agree**
8	The proposed system gains the skill of dividing content into its simplest form	37	92.5	3	7.5	0	0	2.92	.266	4	**Agree**
9	The proposed system helps organize information	36	90.0	4	10.0	0	0	2.90	٠.303	6.5	**Agree**
10	Do you recommend using the proposed system in all courses؟	33	82.5	6	15.0	1	2.5	2.80	٠.464	9	**Agree**
General Mean Dimension									2.87		**Agree**
Standard Deviation									0.195		

It is noted from Table (1) overall that the mean attitudes of the students in the research sample towards the proposed system, as indicated by the total score of the questionnaire, were high. The mean of their attitudes was 2.88, which falls within approval category. This indicates a positive inclination among the students in the research sample towards the proposed system. This can be attributed to the following reasons:

- Students are convinced of the ease of using the proposed system as it helps improve learning and enables remote communication anytime and anywhere.
- The proposed system provides an educational environment that captures learners' attention.
- The proposed system helps simplify information.

These findings reflect the positive reception of the proposed system among the students and their favorable attitudes towards its implementation.

7. Conclusion:

In this study, we propose a method for developing an interactive e-book using NLP techniques. Our approach involves utilizing LDA to uncover hidden semantic structures within a specific text corpus and extracting keywords from educational texts. These keywords play a crucial role in aiding students in comprehending the essence and substance of interactive e-book, as they provide emphasis on important terms within educational texts. The process of keyword extraction consists of four stages. The initial stage focuses on preprocessing data to ensure the cleanliness and removal of any noise present in the text documents. The second stage utilizes N-Gram analysis for statistical evaluation. The third stage involves feature extraction to convert textual data into numerical values. Lastly, in the fourth stage, we extract keywords from educational content. Additionally, results of our study demonstrate a positive reception from students towards proposed system.

Future work could concentrate on incorporating more interactive features into e-book. This could involve generating text summaries and questions based on educational text, collecting and analyzing student interaction data, and utilizing NLP for providing customized content. By exploring these areas, we can enhance interactivity of e-book and provide a more tailored learning experience for students.

6. References:

[١] رياض زروقي ، أميرة فالتة . (٢٠٢٠) . دور الذكاء الاصطناعي في تحسين جودة التعليم العالي . **المجلة العربية للتربية النوعية** ـ المجلد الرابع – العدد ١٢ – أبريل ٢٠٢٠.

[٢] أمينة حسنى . (٢٠٢٠) . الكتاب الإلكترونى ودوره في التحصيل العلمى والمعرفى عند طلبة الجامعات . **مجلة العربية** ، المجلد ٧ ، الجزائر.

[٣] ريهام سامى . (٢٠٢١) . اتجاهات مستخدمى مواقع التواصل الاجتماعى نحو جائحة كورونا (كوفيد – ١٩) تحليل من المستوى الثانى لدراسات مدخل معالجة اللغة الطبيعية . **مجلة البحوث الإعلامية** – جامعة الأزهر – كلية الإعلام – العدد السادس والخمسون.

[4] Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: Models, techniques, and tools. Neurocomputing, 470, 443-456.

[5] Firoozeh, N., Nazarenko, A., Alizon, F., & Daille, B. (2020). Keyword extraction: Issues and methods. Natural Language Engineering, 26(3), 259-291.

[6] Huaijin, P., Jing, W., & Qiwei, S. (2019, January). Improving text models with latent feature vector representations. In 2019 IEEE **13th International Conference on Semantic Computing** (ICSC) (pp. 154-157). IEEE.

[7] Liang, Q., Wu, P., & Huang, C. (2019, June). An efficient method for text classification tasks. In Proceedings of the 2019 **International Conference on Big Data Engineering** (pp. 92-97).

[8] Zhang, X., Tlili, A., Shubeck, K., Hu, X., Huang, R., & Zhu, L. (2021). Teachers' adoption of an open and interactive e-book for teaching K-12 students Artificial Intelligence: a mixed methods inquiry. Smart Learning Environments, 8, 1-20.

[9] Septikasari, A. N., Maison, M., & Nazarudin, N. (2021). Interactive e-book for physics learning: analysis of students' characters and conceptual understanding. Indonesian Journal of Science and Mathematics Education, 4(1), 25–36. https://doi.org/10.24042/ ijsme.v4i1.7664

[10] Mu, Ping (2019). Research on Artificial Intelligence Education and Its Value Orientation. **1st International Education Technology and Research Conference** (IETRC 2019). UK: francis Academic Press. 771-775.

[11] MacNeil, S., Tran, A., Hellas, A., Kim, J., Sarsa, S., Denny, P., ... & Leinonen, J. (2023, March). Experiences from using code explanations

generated by large language models in a web software development e-book. In Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (pp. 931-937).

[١٢] محمد مالك مبارك . (٢٠١٩) . فاعلية كتاب إلكترونى تفاعلى في اكتساب طلبة كلية التربية تطبيقات google التعليمية.

[١٣] سيد نوح سيد عبد الجواد . (٢٠١٩) . أثر نمط التغذية الراجعة المقدمة من خلال برنامج قائم على الذكاء الاصطناعى في تنمية مهارات البرمجة لدى طلاب الصف الثالث من الحلقة الثانية من التعليم الأساسى .**مجلة جامعة الفيوم للعلوم التربوية والنفسية** . العدد ١٢- الجزء الرابع – ٢٠١٩.

[14] Zhao, K., Shi, N., Sa, Z., Wang, H. X., Lu, C. H., & Xu, X. Y. (2020). Text mining and analysis of treatise on febrile diseases based on natural language processing. **World journal of traditional Chinese medicine**, 6(1), 67-73.

[15] Al-Qatawneh, S., Alsalhi, N., Al Rawashdeh, A., Ismail, T., & Aljarrah, K. (2019). To E-textbook or not to E-textbook? A quantitative analysis of the extent of the use of E-textbooks at Ajman University from students' perspectives. Education and Information Technologies, 24, 2997-3019.

[16] Tang, K. Y. (2021). Paradigm shifts in e-book-supported learning: Evidence from the Web of Science using a co-citation network analysis with an education focus (2010–2019). Computers & Education, 175, 104323.

[17] Liu, Y., Chou, P. L., & Lee, B. O. (2020). Effect of an interactive e-book on nursing students' electrocardiogram-related learning achievement: A quasi-experimental design. Nurse Education Today, 90, 104427.

[18] Sarasa-Cabezuelo, A. (2020). A model for creating interactive eBooks for eLearning. Future Internet, 12(12), 223.

[19] Tuah, T., Herman, N. D., & Maknun, J. (2019, February). E-books in teaching and learning process. In 5th UPI International Conference on Technical and Vocational Education and Training (ICTVET 2018) (pp. 281-287). Atlantis Press.

[20] Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.

[21] Payak, A., Rai, S., Shrivastava, K., & Gulwani, R. (2020). Automatic text summarization and keyword extraction using natural language processing. Paper presented at the 2020 **International Conference on Electronics and Sustainable Communication Systems** (ICESC).

[22] Patil, S., Chavan, L., Mukane, J., Vora, D., & Chitre, V. (2022). State-of-the-art approach to e-learning with cutting-edge NLP transformers: Implementing text summarization, question and distractor generation, and question answering. **International Journal of Advanced Computer Science and Applications**, 13(1).

[23] Anandarajan, M., Hill, C., & Nolan, T. (2019). Practical text analytics. Maximizing the Value of Text Data.(Advances in Analytics and Data Science. Vol. 2.) Springer, 45-59.

[24] Šaloun, P., Andršič, D., Cigánková, B., & Anagnostopoulos, I. (2020, October). Crowd Sourcing as an Improvement of N-Grams Text Document Classification Algorithm. In 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA (pp. 1-6). IEEE.

[25] Suleiman, D., Awajan, A. A., & Al Etaiwi, W. (2019, October). Arabic text keywords extraction using word2vec. In 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS) (pp. 1-7). IEEE.

[26] Anglin, K.(2019).Gather-narrow-extract: A framework for studying local policy variation using web-scraping and natural language processing. J Res Educ Eff 2019; 12 (4): 685–706.

[27] Winder, E. (2023). Natural Language Processing Algorithms.

[28] Armouty, B., & Tedmori, S. (2019, April). Automated keyword extraction using a support vector machine from Arabic news documents. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 342-346). IEEE.

[29] Hoblos, J. (2020). Experimenting with latent semantic analysis and latent Dirichlet allocation on automated essay grading. Paper presented at the 2020 **Seventh International Conference on Social Networks Analysis, Management and Security** (SNAMS).

[30] Gurcan, F., & Cagiltay, N. E. (2019). Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling. **IEEE Access**, 7, 82541-82552.

آلية استخدام تقنيات معالجة اللغة الطبيعية في تطوير الكتاب الإلكتروني التفاعلي

ا.دالسعيد السعيد محمد عبد الرازق ، د/ صفاء محمود العطوي ، تسنيم مصطفى جاويش

قسم إعداد معلم الحاسب الآلي – كلية التربية النوعية – جامعة دمياط

ملخص البحث باللغة العربي

يصف هذا البحث كيفية استخدام تقنيات معالجة اللغة الطبيعية لتطوير كتاب إلكتروني تفاعلي باستخدام تقنية (Latent Dirichlet Allocation (LDA، وهي طريقة مستخدمة على نطاق واسع لتحليل وفهم وتلخيص مجموعات كبيرة من البيانات النصية واكتشاف الموضوعات الأساسية فيها. الغرض من النظام المقترح هو استخراج الكلمات الرئيسية من فقرات الكتاب الإلكتروني التفاعلي كونها تمثل ملخصاً لبيانات النص التعليمي. النظام المقترح عبارة عن منصة تعليمية يستطيع الطالب من خلالها عرض أو تحميل الكتاب الإلكتروني التفاعلي والتعرف على الكلمات الرئيسية المستخرجة ، تم اختبار النظام المقترح على عينة مكونة من ٤٠ طالباً وطالبة من قسم الحاسب الألى بكلية التربية النوعية، وأظهرت النتائج اتجاهات الطلبة الإيجابية نحو النظام المقترح، وأهمية استخدام تقنيات معالجة اللغة الطبيعية في العملية التعليمية لتعزيز اهتمامات مشاركة الطلاب، والتفاعل مع المحتوى.

كلمات البحث : – معالجة اللغة الطبيعية – الكتاب الإلكتروني التفاعلي – الكلمات الرئيسية – LDA