Preprints.org

Article

A Hybrid Text Summarization Technique of Student Open-Ended Responses to Online Educational Surveys

Nikos Karousos , George Vorvilas [*] , Despoina Pantazi , Vassilios S. Verykios [*]

Article

A Hybrid Text Summarization Technique of Student Open-Ended Responses to Online Educational Surveys

Nikos Karousos, George Vorvilas, Despoina Pantazi and Vassilios S. Verykios *

Hellenic Open University; karousos@eap.gr (N.K.); vorvilas@eap.gr (G.V.); pandadesp@eap.gr (D.P.)

* Correspondence: verykios@eap.gr

Abstract: This study investigates the effectiveness of a text summarization method applied to open-ended student evaluations at the Hellenic Open University, aiming to improve the analysis of qualitative feedback from online educational surveys. To address the challenges of processing large volumes of student feedback, an automated summarization technique utilizing advancements in Natural Language Processing (NLP) and text mining is proposed. Using the TextRank and the Walktrap algorithms, student comments on positive aspects, study challenges, and improvement suggestions were analyzed. The analysis revealed that while some students expressed satisfaction with tutor-student interactions and the organization of educational material, there were also negative comments about outdated content in some modules and scheduling issues. The findings highlight the importance of qualitative feedback for education quality, providing actionable insights for improving curricula and teaching effectiveness. This research responds to the growing need for effective qualitative data analysis in higher education and contributes to ongoing discussions about student satisfaction in distance learning environments. By effectively summarizing open-ended responses, university staff can better understand student experiences and make informed decisions to enhance the educational process.

Keywords: Online surveys; student opinions; text summarization; TextRank; Walktrap

1. Introduction

The landscape of academic education has evolved rapidly in recent years. During the COVID-19 pandemic, many higher education institutions (HEIs) worldwide were forced to switch to technologies to support the educational process remotely, which greatly boosted the spread of academic distance education and e-learning [1]. Under these circumstances, enhancing the quality of academic education services provided was a priority for many HEIs. To achieve this, student performance was evaluated through various methods, focusing on students' grades and their interaction with learning management systems, their engagement with online resources, the frequency and quality of their participation in virtual discussions, and their ability to complete assignments and assessments within the platform [2–4]. Among from the above indicators, for assessing the quality of academic education, are the evaluation by students of the effectiveness of teaching, the quality of student attendance, the content and design of curricula, etc. In this context, the systematic collection and analysis of student feedback along these dimensions have emerged as key practices in the operational and strategic decision-making processes of HEIs. Therefore, the emphasis on students' evaluation of their studies do not only contribute to curriculum improvement and faculty performance evaluation but is also an important indicator for assessing the overall educational experience of students.

Recent developments have made online surveys a widespread method for collecting educational evaluation data. This method provides many significant advantages over traditional pen-and-paper surveys, including faster data collection, increased accessibility via mobile devices, and the ability to explore a broader student sample. In addition, online survey platforms provide researchers with the

flexibility to use a variety of question formats, enhancing the richness and depth of the data collected. However, despite these advantages, implementing online surveys often presents issues such as low response rates, potential biases in sample representation, and the complexity of analyzing qualitative responses to large data sets [5]. These challenges underscore the need for continued examination of methodologies and practices for collecting and interpreting educational evaluation data to improve the educational process of HEIs.

The prevalence of closed-ended questions in survey designs has lent a degree of objectivity and convenience to quantitative analyses. However, this reliance can bring limitations. Quantitative data often carry on predetermined research patterns, potentially limiting the exploration of emerging themes and insights that could enrich the understanding of students' experiences [6]. In contrast, open-ended questions offer an effective way to more systematically elicit students' views and experiences in their own words. However, the inherent challenges of processing and analyzing qualitative data can pose difficulties, especially in large-scale studies. Processing qualitative data requires rigorous coding and thematic analysis, which can be particularly challenging and time-consuming with large samples [7].

In recent years, significant developments in text mining and Natural Language Processing (NLP) have greatly enhanced the structured analysis of data collected from large text corpora. A promising method that can contribute to the systematic feedback of HEIs with qualitative data for education evaluation is automatic text summarization [8]. The present study aims to contribute to the broader debate on enhancing the quality of education by applying the above technique in the context of the Hellenic Open University (HOU). The HOU is the only HEI in Greece that exclusively offers distance learning. Students' evaluations at HOU investigate key dimensions such as tutor effectiveness, the quality of educational content, and the quality of the design of academic modules. Findings showed that students were satisfied with tutor-student interactions and the organization of educational material, but also there were negative comments about outdated content in some modules, and scheduling issues. Regarding the proposed text summarization technique, it seems to be less time consuming than human summarization, but requires thorough data preparation, anonymization, and normalization. In conclusion, the findings elicited with this technique can provide valuable insights not only for HOU but also for HEIs in general, regarding the experiences of attendance and the factors that positively or negatively affect it.

2. Analyzing University Students' Opinions from On-Line Educational Surveys: Benefits, Challenges, and Methods

Students' opinions in academic education is an important factor in investigating the quality of services provided by HEIs. Students' opinions are mainly concerned with evaluations of faculty effectiveness and programme quality [9–11], as well as student satisfaction [12]. These evaluations are routinely carried out using either internationally recognized and valid questionnaires or questionnaires constructed and adapted to the requirements of the institution concerned. HEIs attach particular importance to the student opinions they collect because it helps them to make decisions regarding the recruitment or promotion of academic staff, changes in course content, and improvements in the administrative and technical services provided [13].

In recent years, conducting evaluations in HEIs to collect feedback from students' opinions has been greatly enhanced by the use of online surveys. Several online tools are commercially available, but many HEIs have also developed infrastructures that support their in-house online surveys. The major advantages of online surveys are the speed of data collection, the flexibility of location and time of completion, the ability to participate in the survey through mobile devices, the ability to reach a large sample of participants, etc. In addition, they provide researchers with ease of creating and structuring questionnaires through a range of question types and functionalities for data visualization and monitoring during the collection of the responses [5,14,15]. Their drawbacks include, often low response rates, research design arbitrariness regarding sample representativeness, unclear questions or instructions, collection of unwanted responses (e.g. insulting comments), and privacy assurance issues [5,16,17].

In higher education, online evaluation questionnaires with closed-ended questions are preferred by far. These questions are indicators for measuring factors such as student-tutor interactions, programme materials, course organization, and instructional design. Online questionnaires are effective because they allow for the rapid collection and analysis of data from almost the entire population and comparing results over time [18]. In addition, including rating items allows for easier interpretation of findings regarding tutor effectiveness and other dimensions of the educational process, and subsequent decision-making [19].

However, it has been pointed out that quantitative methods are characterized by their objectively closed nature, meaning that if a topic/dimension is not anticipated in the survey design, it will not appear in the results. More generally, the tendency of researchers to rely on previous quantitative studies and to confirm existing research hypotheses, often leads to the replication of a closed variable structure that is systematically used in different contexts. This approach can simply recycle existing knowledge without any innovation, thereby limiting the possibility of generating new knowledge [6]. Although the use of open-ended questions can provide valuable qualitative data on student feedback, they are often considered time-consuming and demanding to process. As a result, they may be unsuitable for many researchers conducting large-scale surveys [11].

Including open-ended responses in online surveys seems to have many additional benefits compared to closed-ended responses. Open-ended questions provide valuable information and complement the quantitative findings in the survey. They are useful when exploring new topics or concept dimensions as they may reflect real-life experiences [20]. Through open-ended responses, respondents are not limited by a predetermined set of closed options, they can express their thoughts in their own words, often producing rich information [21]. In addition, respondents can create themes and sub-themes themselves, which may contradict or extend the theoretical assumptions of the researchers on the issue under investigation [6,22]. However, research has shown that completion rates for open-ended responses to electronic questionnaires are often lower than for closed-ended responses [15]. Issues such as the design of the questionnaire, the available space for typing, the instructions for completion [23,24], the interest of the respondent and the cognitive load required to answer [7] can affect the length and quality of open-ended responses [21].

In Large-scale online surveys, open-ended responses posit significant challenges to researchers. Processing qualitative data requires their coding into categories, through the identification of recurring patterns and ideas [25]. The categories of analysis are shaped by the relevant literature review and the data itself, ensuring logical consistency. This process is time-consuming, requiring time to review the literature, preliminary testing of the coding scheme, and training raters in its use. Raters' training is essential to ensure agreement between them in coding the data and thus for the reliability of the survey [26]. In small samples, coding can be done manually, with two or more people coding independently to assess inter-rater reliability. In large-scale qualitative surveys, coding often requires the ad-hoc creation of broad categories, leaving a wealth of data unused [27]. Furthermore, this type of survey involves the division of labor by creating multiple teams with specific responsibilities, which may facilitate survey management but often create problems of communication and translation of information when they do not communicate effectively with each other [27,28]. In addition, processing and coding qualitative data from online surveys becomes particularly challenging when responses are unclear [29].

The use of computer-assisted qualitative data analysis software (CAQDAS), such as Atlas.ti, NVivo or MAXQDA, can help reduce the issues associated with analyzing qualitative data, but requires careful oversight to ensure the accuracy of the classification. Although these tools help with text tagging, the coding process remains time-consuming and demanding for large samples. In addition, these tools cannot be fully trained to automatically identify the appropriate pieces of text for coding. Therefore, individual decisions by researchers become crucial in this case, and agreement among raters on an acceptable coding scheme remains a common strategy to minimize subjective errors and bias [28].

Researchers have tried to automate the above process using various text mining methods through machine learning. Literature distinguishes between supervised and unsupervised machine

learning methods [30]. In supervised methods, initially, human raters encode a subset of training data, which in turn is used to train an algorithm for predicting uncategorized text responses. Although these methods require modeling and prediction expertise, they offer low cost, and fast execution [7]. Unsupervised methods extract data from text mainly at an exploratory analysis level to search for common underlying themes and do not require manual effort. These themes lead to a better understanding of the conceptual structure of the corpus, cleaning it of unwanted information and coding it to highlight key concepts in the analysis. For a brief presentation of these approaches see also [31].

In addition to the above methods, a popular method in recent years is automated text summarization. This is the process of condensing a large amount of text data into a shorter, more concise form, allowing for easier apprehension of the main points without reading the entire text. It is a text mining technique that aims to summarize the content of large texts using appropriate algorithms. This technique saves time when large volumes of text bodies must be managed. The procedure of automated text summarization usually includes three stages: pre-processing, processing, and post-processing. In pre-processing, words, sentences, and structural elements of the text are identified as input units. In processing, the input text is converted into a summary using a summary method. Finally, in post-processing any problems in the summary created are corrected [32].

There are different types of automatic text summaries depending on the criteria applied to produce them [8,33]. We distinguish between single and multi-document summaries depending on whether the summary results from one or more documents. We also speak of language-specific summary if all documents and their summaries are written in the same language (e.g., Greek), or multilingual, if for example the documents are written in Greek and English but the summary must be in Greek, or cross-lingual, if all documents are written in Greek and the summary is in English. Summaries can also be flat (a single summary is produced with no intermediate summaries) or hierarchical (multiple levels of summaries are made, e.g., abstract and extended summary, allowing for zooming in and out of the text), general purpose or query-based (highlighting the parts of the text related to a query). Finally, summaries can be extractive or abstractive. In extractive summary, the most important sentences from the original text are selected and prioritized into a shorter summary; in abstractive summary, the most important text information is reworded to create the final summary [34]. Hybrid summarization is a combination of extractive and abstractive techniques.

Although abstractive summarization often performs better than extractive [35], it is usually difficult to implement. The emergence of transformer-based technology has dramatically advanced the field of abstractive summarization. However, the high hardware requirements and specifications, as well as the need for large amounts of training data, can act as a deterrent. This is because the required data may not be readily available in various domains, or the volume of data may not need to be particularly large. As [32] states, due to the complexity required to produce an abstract summary, recent research in the field has focused on extractive techniques.

3. Context and Research Questions of the Present Study

The research was conducted at the Hellenic Open University (HOU). HOU is the only HEI in Greece that offers exclusively distance education at undergraduate and postgraduate levels. With approximately 40,000 active tuition-paying students and an average age of 35 years, HOU is the largest institution of its kind in the country. Academic studies at HOU provide autonomous and self-directed learning which is facilitated by tutor-student, student-student, and student-content interactions, within a technologically supported environment [36,37]. HOU offers distance academic education in various subjects, with study programmes structured in annual or semester modules. For each module, students have access to specific educational content (educational materials such as books and other digital educational resources, as well as educational activities such as written assignments, projects, and laboratory exercises). The learning path is defined by study guides and time schedules that accompany the educational content. It is the learning material that plays the primary instructional role in the student's learning process and not the faculty members [38]. Also,

activities supported by adequate feedback play a vital role in helping students to learn [39]. Faculty members, in this distance learning context, assume the role of tutors, who do not lecture but advise and guide students in their studies [40]. More specifically, the role of tutors is to provide clear explanations to student queries, assist students with the comprehension of educational material, guide them through educational activities, offer constructive written feedback on their work, and maintain effective communication with their students. Additionally, tutors should encourage and motivate students, while effectively utilizing the educational platforms to enhance the learning experience. Attendance at HOU is accompanied by online tutor-student meetings during which tutors are able to guide students through the activities, resolve their questions, and encourage them to continue their studies [41]. In addition to these online meetings, interaction with tutors and educational content is covered asynchronously through the use of educational platforms which are customized versions of the Moodle LMS.

In the aforementioned context, students are invited to anonymously evaluate their experience of the module or modules they attended to through an online questionnaire, which includes four evaluation axes: a) the role of the tutor, b) the educational content (educational material and activities), c) the design of the module and d) the administrative and technical services. These axes include a total of 20 closed-ended questions. In addition, the questionnaire includes three open-ended questions that encourage students to articulate the most important positive and negative aspects of the module, as well as propose suggestions for improvements. These questions are included in order to gather additional information about students' learning experiences, which the closed-ended structure of the questionnaire may do not anticipate.

The large volume of qualitative data resulting from student feedback, as well as its analysis, remains a challenge for HOU. Various text mining techniques have previously been applied to either detect faculty member and student satisfaction with remote online examinations [42,43] or student satisfaction with master's thesis supervision [44]. However, a comprehensive approach in terms of methodology and scale regarding the processing and analysis of qualitative data collected from online surveys is lacking. The findings resulting from such an approach will provide valuable feedback to the academic community and HOU administration. This feedback can assist in making decisions regarding collaboration with tutors, updates of educational materials and activities, redesign of study programmes and modules, and improvement of administrative and technical services.

More specifically, the beneficiaries of text summarization results will be HOU's Study Programme Directors (S.P.Ds.). An S.P.D. is a faculty member whose main goal is to ensure objective assessment of student progress, promote scientific research, and develop technology and methodology in lifelong distance learning. The S.P.D. assigns Thesis projects, suggests other faculty members for the role of module coordinators and coordinator assistants, collaborates with administrative services, and contributes to the improvement of educational material. Additionally, the S.P.D. is responsible for regular communication with module coordinators and tutors They collect reports to inform the University Administration about the progress of the study program, highlighting issues and proposing improvements. The S.P.D. also coordinates the study programme certification processes according to H.O.U. standards, ensuring the implementation of guidelines from the Quality Assurance Unit (Q.A.U.) and the Internal Evaluation Teams (I.E.T.).

Considering the context of HOU's evaluation of the distance education process and the challenges posed by open-ended questions, the aim of this study is to trace the satisfaction of students of an annual study programme from the distance academic education provided by HOU by analyzing student feedback from open-ended responses to an online evaluation questionnaire through the application of hybrid text summarization. More specifically, the research questions posed are as follows:

1) What are the students' positive experiences of taking their modules?
2) What aspects of module attendance do students find challenging or unsatisfactory?
3) How can the modules be improved?

4. Research Design and Participants

The present research was based on a cross-sectional design [45]. More specifically, students were requested to evaluate the dimensions of the educational process through an online questionnaire shortly before completing their studies in a study programme. Students are informed that evaluation is not compulsory, but are encouraged through frequent reminders to participate, pointing out that they can contribute with their participation to the improvement of the course they are already attending. Students have the choice before the final submission of the questionnaire to consent or not to the further use of their personal data for surveys conducted by the Internal Evaluation Unit (IEU). They are also informed that the results of the evaluation are anonymously communicated to the tutors after the completion of the re-examinations. The administration of the evaluation questionnaire and the subsequent collection and processing of the data were carried out by IEU of HOU. The target population comprised students enrolled in the annual study programmes provided in the academic year 2023-2024. The sample for this study consisted of students who attended the twenty modules of a graduate study programme, in that year. In total, 1.197 out of 2.618 student evaluations were anonymously submitted (45.72%).

5. Method

Figure 1 shows the proposed model for the analysis of the responses to the open-ended questions. The model describes the workflow that is followed, and which consists of 5 distinct steps. In each step, one or more processes are executed, the results of which feed the next steps. The first 3 steps are concerning pre-processing while the next 2 are about the mechanisms for extractive and abstractive summarization. The first group of steps can be performed iteratively until its result is at a satisfactory level of quality.
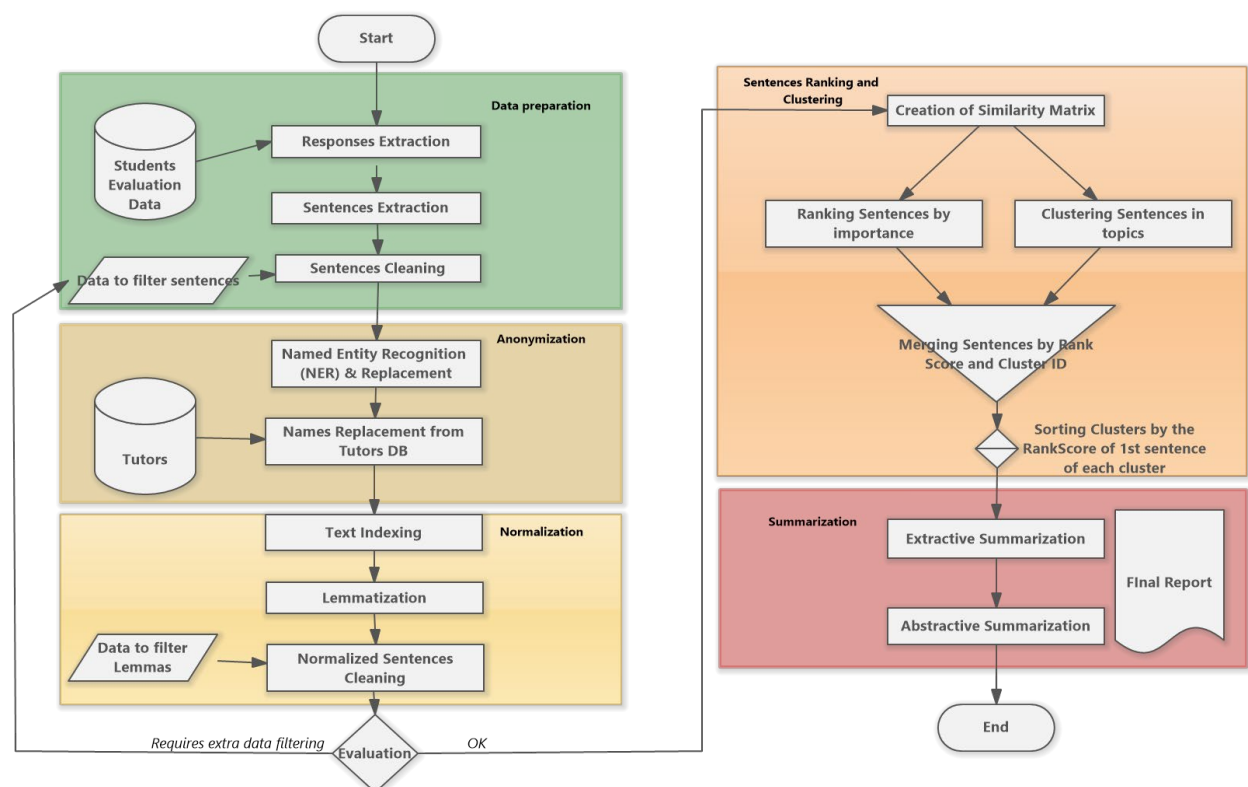
Figure 1. The proposed automated text-summarization model.

5.1. Data Preparation

This process aims to collect, separate and clean the student responses to produce an initial dataset, suitable for the anonymization as well as the lemmatization processes that follow.

5.1.1. Responses Extraction

A relational database maintained by the Institution is used to store participants' responses for every completed questionnaire. When data preparation begins, all responses to a particular question are collected in a dataset. Each response may include multiple sentences. Next, for each response a unique identifier is assigned. Metadata regarding school, study programme, and module are attributed to each response. Usually, each response is categorized as positive, negative, or suggesting improvements based on the open-ended questions in the evaluation questionnaire. However other categories of questions may also be supported depending on the application field.

5.1.2. Sentence Extraction

A sentence in each student response is considered as the coding unit of analysis, and the full student response as the context unit [46]. Therefore, the dataset from stage 5.1.1 is shaped into a dataset in which each row contains a single sentence (with a unique ID), along with the relevant metadata, and the student's full responses, about where exactly the sentence is located.

5.1.3. Sentence Cleaning

The sentences are reviewed to remove those that are irrelevant to the open-ended questions of the questionnaire or do not provide meaningful insights. For example, in response to the question, "Please mention the most significant positive experiences from your studies in the module," answers like "everything is good," "everything is perfect," or "I am completely satisfied" are excluded. Similarly, for the question, "Please mention the most significant negative experiences from your studies in the module," responses such as "I found nothing negative", "no negative aspects, everything is good!" are also excluded. Likewise, for the question, "Please mention any improvements you would like to see in the module," responses like "no improvements needed" or "I have no suggestions" are excluded too. Moreover, extra filters concerning generic rules for sentence removal may be applied here. For example, rules that exclude one-word sentences, or sentences that have less than a predefined number of required characters may also be applied. Finally, it is possible to exclude sentences that are written in other languages than Greek. The entire set of rules and filters is fed by external datasets of specific stop-words and other similar parameters that are generated once by administrators and domain experts and can be updated later.

5.2. Anonymization

The purpose of this process is to detect and replace the tutors' names that may be mentioned in the sentences. This process does not guarantee absolute success and in case of a relevant requirement, human intervention is possible. It is worth noting that both 5.1 and this procedure are preferred to be carried out within the Institution because they may involve the processing of personal data.

5.2.1. Named Entity Recognition and Replacement

This process takes as input the prepared dataset and the intended result of its execution is the recognition of all words that are likely to be the name of a tutor. For this purpose, Named Entity Recognition (NER) is used. NER provides tools to recognize mentions of rigid designators from text belonging to predefined semantic types such as person, location, organization [47]. NER methods can be dictionary-based, rule-based, machine learning-based, deep learning-based, or hybrid combinations of them [48]. After the names are identified, these names are replaced by the word "tutor". This creates a new dataset in which a significant percentage of names have been identified

and replaced. Due to the requirement for optimal name replacement the entire process may additionally include one or more NER models.

5.2.2. Custom Name Replacement from Tutor DB

In this scenario the entire set of tutors that are responsible for the students' classes are already known by the Institute. Hence, a custom-made procedure for detecting and replacement of already known tutors' names is being executed. This procedure considers on the one hand the students' classes and their tutors and on the other hand the class in which each student response was found. Thus, it performs a very specific search-and-replace technique by searching in each sentence for a very limited number of names. Initial tests showed that completing both steps of the name matching had over 90% efficiency.

5.3. Normalization

Text normalization refers to the standardization of the elements (tokens) of a document to achieve greater homogenization aiming at a robust content analysis. To this end, two approaches are mentioned in literature: stemming and lemmatization [47]. In stemming, the endings of words are removed, and their roots are kept in order to achieve homogenization of the information (e.g., detailed turns into detail). In lemmatization, words are replaced by their base form, i.e., the form in which they appear in a dictionary entry (e.g., students turns into student). In the steps presented below, lemmatization has been adopted as a normalization method. The evaluation of this stage is very important since it can lead to the detection of inaccuracies and unnecessary lemmas which should be removed immediately or even redirected to the initial filtering stage and to repeat the process.

5.3.1. Text Indexing

Text indexing concerns the process of converting a text into a list of words and includes the following steps [33], [49]: a) Tokenization: Each sentence (V_i) are broken down into individual tokens, presenting the sentence as a vector of words, $V_i = \{W_1, W_2, \dots, W_n\}$; where there are N total sentences with a total number of n_i words for each sentence, $i \in \mathbb{Z} \cap [1, N]$. As a result, each sentence is represented as a collection of word tokens; b) Stop-words' removal: Common word tokens with high frequency in the text, such as articles, pronouns, and conjunctions, are removed as they do not carry substantial information. All stop-words that must be removed can be given as external datasets filled by domain expert users; c) Part-of-Speech (POS) Tagging: Word tokens are tagged with the grammatical category they belong to (e.g., noun, adjective, verb).

5.3.2. Lemmatization

Tokens are converted to their dictionary form to achieve better language normalization. In turn, users may decide which lemmas should be kept for text analysis, based on the POS they belong to. Such decisions are guided by the research goals in question. POS descriptive statistics is also available in this step to get better insight.

5.3.3. Normalized Sentence Cleaning

Normalized methods are not always completely accurate, so often is required from reviewers to verify the normalization process and make manual corrections to ensure correct lemma/stem formulation, rectify misclassified lemmas/stems, and ultimately exclude lemmas/stems with no contribution to the subsequent analysis. These correction data are recorded in specific datasets and may then update the initial datasets into the cleaning mechanism. The process is repeated until the desired outcome is reached.

5.4. Sentence Ranking and Clustering

The aim of this procedure is to recognize both the representativeness and the cluster to which each sentence may belong. These two factors will play a critical role to the final step in which summarization will take place, since they will guide the summarization models in including the most representative and significant trends coming from the set of responses. Combining clustering with the sentence ranking procedure has a limitation in addressing the disadvantages of the algorithm used to yield significant sentences. In particular, this approach lacks understanding of the meaning and is influenced by the length of the sentences [50]. The distribution of the sentences into groups mitigates the bias that may arise in the measurement of their rank importance. Primarily, the extracted trends will offer a comprehensive and broad variety of topics on which students focus their responses.

5.4.1. Creation of a Similarity Matrix

The similarity matrix is formed via the Jaccard similarity measure, which assesses their overlap in terms of unique lemmas among sentences. Each sentence becomes a row and column in the matrix, with values indicating their similarity. Other similarity algorithms may also be tried for performance reasons when dealing with a large number of sentences. Generally, the Jaccard measure between two different sentences (vectors of words, Sec. 5.3.1) V_i and V_j is shown by:

$$J(V_i, V_j) = \frac{|V_i \cap V_j|}{|V_i \cup V_j|} \tag{1}$$

for $i, j \in \mathbb{Z} \cap [1, N]$ where $|V_i \cap V_j|$ is the number of common words between V_i and V_j and $|V_i \cup V_j|$ is the number of total unique words in both sentences.

5.4.2. Ranking Sentences by Importance

The TextRank algorithm [50] is used to compute the relative importance scores for each sentence and then rank them. It is an implementation of the PageRank algorithm [51] for sentences. The PageRank algorithm was introduced by [48] as a way to measure the probability of going to other pages, given that a user is on one web page (for brevity is referred as page). PageRank is a graph-based ranking model and considers a graph G(V,E); where V are the vertices-nodes (pages) and E are the edges (links of pages). The PageRank (PR) for a random vertex of the graph, page A, is shown by:

$$PR(A; t + 1) = \frac{(1 - d)}{N} + d * \left(\frac{PR(T_1; t)}{L(T_1)} + \cdots + \frac{PR(T_k; t)}{L(T_k)} \right) \tag{2}$$

Here, assuming a web with N total pages and a random node $k \in \mathbb{Z} \cap [0, N]$, $T_{1\ldots} T_k$ are the pages that are linked to page A, namely the incoming pages to A (for instance references to A) with the approximate number of outgoing links from these pages to others, $L(T_1),_{\ldots}, L(T_k)$, d is the damping factor (residual probability) to take into account that the user may not continue to click on a page that links to A (usually set to 0.85) and t is the time. It is obvious that PageRank score needs an initial PR score and is typically set to 1/N and that the algorithm focuses on incoming links to each page. The algorithm stops in a time length, t, where the convergence is achieved, i.e., for page A when $|PR(A; t + 1) - PR(A; t)| < e$, assuming a convergence threshold, e.

TextRank is also a graph-based ranking model and is an extended version of PageRank by considering the sentences as nodes and the edges as weighted connections between pairs of nodes (similarity measure). For that purpose, the similarity matrix [5.4.1], needs to be transformed into a column-stochastic one (each column totals one) to rely on incoming links. The TextRank (TR) for a random vertex of the graph, sentence, V_i, is given by:

$$TR(V_i; t + 1) = \frac{(1 - d)}{N} + d * \left(\frac{w_{1i} TR(V_1; t)}{W_1} + \cdots + \frac{w_{ki} TR(V_k; t)}{W_k} \right) \tag{3}$$

Here, $V_{1\ldots} V_k$ are the sentences that are linked to the sentence V_i (k total incoming links to V_i, for $k \in \mathbb{Z} \cap [0, N]$), w_{ji} is the weight of the edge from the node V_j to the node V_i and W_j is the sum of the weights of all outgoing edges from the node V_j; $W_j = \sum_{l=0}^{m} w_{jl}$ (m total outgoing links from V_j, $l, m \in \mathbb{Z} \cap [0, N]$).

The calculation of the TextRank is closely linked to the calculation of the similarity index [4.1] used, in a proportional way. Indeed, the degree of change in a similarity measure (weight) affects the numerator of TextRank more directly than the denominator. In a parallel perspective, long sentences tend to have more common words with other sentences. Therefore, a problem that may arise is the bias of the TextRank in favor of long sentences, as they are more likely to have a high similarity with the other sentences, although it may be good when considering smoothness [52]. This is differentiated by the similarity measure used in each case and can be significantly limited, i.e., by normalizing the similarity index by the length of each sentence [50]. Moreover, using a clustering method reduces the potential of promoting long sentences by spreading sentences across communities and extracting a selected number of significant sentences from each group rather than the whole set.

5.4.3. Clustering Sentences in Topics

As TextRank can lead to distributions that may contain sentences with the same meaning in consecutive positions in the ranking, it is preferable to cluster the sentences so that eventually sentences from different clusters appear in the appropriate order, giving a better reflection of the overall trend. The similarity matrix (5.4.1) is transformed to zero diagonals to avoid self-loops in nodes, and a graph, G(V,E), is constructed and inputted on the Walktrap algorithm. Walktrap algorithm uses the Random Walk (RW) procedure's properties (Markovian Property and Stochasticity). A random walk in a graph describes a repetitive path of random steps in space, commencing from a given vertex and progressing to a randomly selected neighboring vertex until a termination condition is satisfied.

Walktrap is used [53] to simulate transition probabilities as random walks (5.4.3) on an undirected graph G(V,E) where the algorithm starts at each sentence and moves to a neighboring sentence based on the unweighted or weighted edges in the graph (probabilities of transitioning from the starting sentences to others or remaining at the initial sentence). The basic idea of the Walktrap is based on the fact that random walks have the tendency to get "trapped" into densely connected parts of the graphs (community detection). Namely, the algorithm identifies regions of the graph where the random walks tend to stay within certain clusters of sentences for longer periods of time. These regions are interpreted as communities or clusters of closely related sentences. Given a weighted graph, the Walktrap algorithm begins by computing at each time step, t, the transition probability (p_{ij}) from vertex i to vertex j as:

$$p_{ij} = \frac{s_{ij}}{d(i)} \tag{4}$$

where s_{ij} is the i-row, j-column element of the zero-diagonal similarity matrix and $d(i)$ is the degree of vertex i given by $d(i) = \sum_j s_{ij}$, namely the sum of i neighbors' weights. To initialize the algorithm (t = 0) based on the starting node i, the initial transition probabilities for N total nodes are set as a vector of length N, $P_i^{(0)} = [p_{i1}^{(0)}, \ p_{i2}^{(0)}, ..., p_{ii}^{(0)}, ..., p_{iN}^{(0)}]$ where, $p_{ii}^{(0)} = 1$ at node-position i and $p_{ij}^{(0)} = 0$ at all other positions for j ≠ i. Assuming the transition matrix P of the random walk process, the follow applies for t-steps (length):

$$P_i^{(t)} = P_i^{(t-1)} \times P^{(t)} \tag{5}$$

where $P_i^{(t)} = P_i^t$. It repeats this process several times (the length of the algorithm is defined and equals 4 in this study). Indeed, in Walktrap, the random walks are short in time (best empirical results for length 3 ≤ t ≤ 8, with a specific good compromised length t=4 or t=5) due to the fact that when time, t, converges to infinity, each node's transition probability equals an equilibrium probability [53]. This equilibrium probability of being on a vertex j depends only on the degree of vertex j and not on the degree of starting vertex i. Walktrap relates to spectral approach on a graph as it can be proved that the equilibrium transition probability vector corresponds to the eigenvector of the largest eigenvalue of matrix P (the transition matrix of the random walk process) and results in only one community (very small differences in transition probabilities). Subsequently, the Walktrap algorithm calculates the distances between sentences using the probability distribution in a metric based on Euclidean distance normalized with respect to the degree of node k (to reduce the influence of the total weights of neighbors), where k is any random node from 1 to N:

$$d_{ij}^{(t)} = \sqrt{\sum_{k=1}^{n} \frac{(p_{ik}^{(t)} - p_{jk}^{(t)})^2}{d(k)}} \qquad (6)$$

Using these probabilistic distances, Walktrap reduces the problem of finding communities through the technique of agglomerative clustering and the Ward method (nodes with the smallest distance are first merged into one community). The process stops when the maximum value of modularity, a common metric for the quality of a division of a network into communities [54], is achieved.

Graph clustering, which is part of graph theory, is an interesting and challenging research problem that has received much attention in recent years. Clustering on a large graph aims to partition the graph into several densely connected groups, called communities (community detection). A major difference between graph clustering and traditional relational data clustering is that graph clustering measures vertex closeness based on connectivity (e.g., the number of possible paths between two vertices) and structural similarity (e.g., the number of common neighbors of two vertices), while relational data clustering measures distance mainly based on attribute similarity [55]. Although many of the techniques involved in graph clustering are closely related to the task of finding clusters within a given graph, Walktrap has a lot of advantages compared to them.

Spectral clustering is an important technique in the domain of community detection as it considers the overall structure of the graph. The problem that spectral analysis faces, however, is the time needed to compute the eigenvalues, especially for a large-scale graph. Walktrap uses random walks to find communities quickly by making use of transition probabilities, avoiding the explicit computation of eigenvectors [53].

The hierarchical clustering method has some main drawbacks that make it inappropriate for large real-word networks [56]: that it cannot clearly determine how many communities the graph will be divided into, despite the fact that it does not need to set the a-priori number (as in k-means), and that it fails to identify communities with different scales (the resolution problem). In particular, agglomerative clustering fails to find the correct communities when the structure of communities on the graph is known and tends to identify only the core of a community and omit the periphery [54]. The Walktrap offers a good solution as it avoids the problems that exist by unilaterally using hierarchical or spectral clustering and achieves the identification of communities through random walks and transition probabilities.

5.4.4. Updating Ordered Sentences with Cluster Information

Ranked sentences in descending order [4.2] according to their TextRank scores are inserted in a data frame along with their corresponding cluster ID. Essentially, the dataset brings together information about the cluster each sentence belongs to and how important each sentence is within its cluster. This unified dataset enables a comprehensive understanding of the relationship between sentences within clusters and their relative importance based on TextRank scores. Consequently, this approach enhances the ability to derive meaningful insights from the data, facilitating more informed decision-making, according to next step.

5.5. Summarization

This final step of summarization is performed in order to produce a well-structured summary of participants' responses. This summary will be in terms of extractive and abstractive one. Considering extractive summarization, a list of the most important representative sentences that participants made is required to be presented. On the other hand, abstractive summarization will produce a synopsis of the trends that are being recognized by the extractive analysis.

5.5.1. Extractive Summarization

Assuming clusters C_i, sentences V_{ij}; the sentence V_j in C_i cluster, $i \in \mathbb{Z} \cap [1, c]$,

$j \in \mathbb{Z} \cap [1, N]$ and TextRank score for each sentence $TR(V_{ij})$, the process involves descending the clusters by their top-ranked [5.4.4] sentence, namely the sentence with the highest TextRank score within each cluster. In each ordered cluster C_i, z top-ranked sentences are selected using:

$$V_{ij} = \{v_{ij} \in C_i | TR(v_{ij}) \geq TR_{median}\} \tag{7}$$

Here, TR_{median} is the median TextRank score of all sentences across all clusters and $|V_{ij}| \leq z$. The choice of median is made to ensure that the most representative and unbiased trend from each community is obtained. Subsequently, x clusters that satisfy the condition of including sentences with a TextRank score value as defined are included in the resulted list with z sentences, where $x \leq c$ for a given c number of clusters resulted from Walktrap (5.4.3).

5.5.2. Abstractive Summarization

This step involves generating an abstractive summary by concatenating all selected cluster sentences (5.5.1) and passing them for elaboration to the OpenAI API in order to enhance readability and meaning for the end user. The final summary is generated in a hierarchical manner by first generating the summaries of each cluster and then generating a summary of all cluster summaries.

Ater the final summary generation, this process produces a final summary report that consists of three parts. The first part presents quantitative data concerning the number of open-ended responses per question type, the number of sentences extracted, and the number of pre-processed sentences used for summarization. The second part presents the abstractive summary, and the third part presents a list of 5-10 representative comments, according to (5.4.4).

6. Implementation

The implementation of the above flowchart took place into the environment of the HOU university trying to bring most of the models inside the university's area and limit the dependencies of 3rd party models in the cloud due to the restrictions of anonymization and the institute data policy. The future goal is to build a system that could be used for the analysis of all open-ended questions of all evaluations that are taking place in the university regardless of the study programme. However, a pilot study was focused on only one postgraduate study programme in the university.

Two were the core modules of the system: The *Responses Preparation Module* and the *Survey Summarization Module.* They were implemented using Python as RestFull services, utilizing the Flask python library [57]. A MySQL database management system was used as storage of all the required data. The administration and the production of summaries were provided through a web-based application built on PHP programming language.

The first module moderates the data preparation steps and it takes as input raw data of the students' responses and returns a well-formed dataset that contains the initial response, the sentences of each response, and the lemmas of each sentence. The dataset is filtered and cleaned appropriately. This module includes sentence preparation, anonymization and lemmatization internal components. The Survey Summarization Module is responsible for the production of a summary of the responses to open-ended questions. It takes as input the dataset that contains clean lemmatized sentences and then computes the similarity matrix between the sentences to produce a graph. Next this module invokes the Sentence Ranking together with the Sentence Clustering internal components. These two components will result in the ranking of all the sentences and in their clustering as well. In this way it will be able to highlight the most important sentences and to prepare a dataset of clustered sentences for an abstractive summarization. Finally, the module will produce the final report with the summary of the responses. More specifically, it presents an instructive quantitative paragraph in relation to the number of answers as well as the number of sentences that arose from them and the number of clean sentences that were finally used in the summarization process. Next an abstractive summarization follows that briefly describes the trends found from the analysis of the responses and finally a list of indicative and representative sentences selected from the students' responses is presented. When summarizing must be abstractive, then the service will invoke an external AI summarization service.

Concerning the required input data for the whole process, they consist of the following datasets:

- Students' responses of an open-ended question
- Replacement rules in the form of pairs find:replace
- Words to be excluded from NER analysis
- List of study programmes, course modules and tutors of each module (as an input for the custom Name-Replacement algorithm)
- Lemmas to be excluded from lemmatization
- Lemmas to be corrected before lemmatization
- Some POS custom definitions in particular lemmas

Only two datasets of the above are being replaced each time a new question is analyzed (students' responses, and sometimes the list of study programmes, modules and tutors) while other datasets are domain specific and are updated rarely.

Some critical libraries and external services were also integrated in the system. First concerning data cleaning the langdetect 1.0.9 [58] was used to detect responses in languages other than modern Greek that system is using. This library is a direct port of Google's language-detection library from Java to Python. Concerning NER, an already pre-trained model in Greek Language named Spacy [36,37] was used and configured to recognize only persons' names. The implementation of lemmatization in Greek Language was done by the spacy-udpipe library [61]. Both the TextRank algorithm and the Jaccard similarity computation were implemented in python, however the latter used a CountVectorizer algorithm of the scikit-learn library [62] to optimize matrix production. In the same way the WalkTrap algorithm was implemented inside the module using the igraph library [63], a network analysis package. Regarding abstractive summarization that is performed in case of the production of first the clusters' summaries and next the final summary of all the clusters, the GPT4o mini model [64] was used via the OpenAI API.

It's worth mentioning that the development approach supports future integration of new models in case they can enhance the current system functionality.

7. Results

Regarding quantitative analysis, for the study programme of the present research, a total of 1197 evaluations were conducted out of a total of 2618 potential evaluations (participation rate 45.72 %). The means per evaluation dimension are presented in Figure 2, where the maximum value is 5 and the minimum is 1. As can be seen, students tend to be satisfied with their study in this curriculum with means ranging above the middle of the five-point scale. However, they seem less satisfied with the educational content (materials and activities).
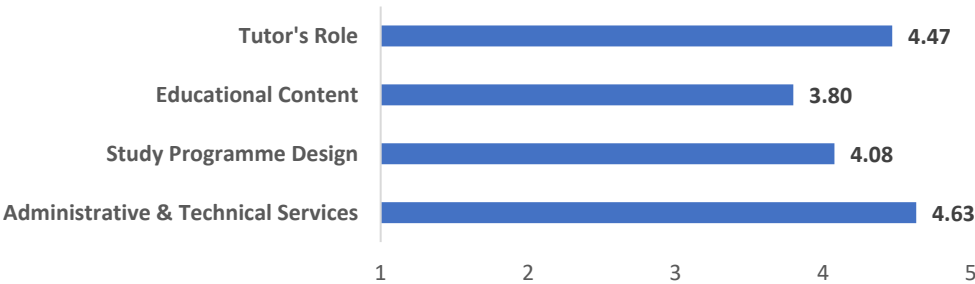
Figure 2. Results of student satisfaction per dimension of the study programme.

886 responses were given to the three open-ended questions of the research tool. A response is the filling of free text in a submission form. Students could optionally provide one to three responses - one for each comment type (positive comments, negative comments, and suggestions for improvement). Each response included 1-n sentences. From these sentences, a subset was selected after pre-processing to form the units of analysis for the automatic summarization. In Table 1 we can see by comment type, the distribution of the responses, and the sentences used for the automated summarization. We observe an equal distribution of responses per comment type. At the level of sentence distribution by comment type, we observe that students tend to write more sentences for

improvements and fewer for negatives. The same is true for the final sentences used for automated summarization.

Table 1. Distribution of student comments.

Response type	Comments' distribution per response type	Sentence distribution per response type	Sentences kept for summarization
Positives	610	852	835
Negatives	612	831	758
Improvements	610	933	889
Total	1832	2616	2482

Regarding the students' positive experiences from their study programme, a portion of them mentioned that they were satisfied with the student-tutor interactions. tutors were willing to communicate and collaborate with students and efficiently solved their questions. In the online tutor-student meetings, a positive atmosphere of cooperation and communication was cultivated among students, tutors, and their fellow students. Students felt that they gained new and useful knowledge from the module they attended. Also, the educational material was considered well organized, and the educational activities contributed to its understanding. Finally, the support of their studies by the possibility of attending tutorial online meetings was also positively valued by the students.

Regarding the negative comments of the students, a portion of them referred to issues related to the educational material, such as the large size of the educational material which does not give the time for sufficient understanding, and the outdated contents in some cases. Negative comments were also made by a proportion of students about tutors, focusing on the reluctance of some tutors to work with their students, the lack of adequate student guidance and educational activities' feedback as well as the insufficient question resolution and the lack of examples for better understanding. Finally, students referred to time scheduling issues such as inflexible deadlines for submission of educational activities, the reluctance to grant extensions, the timing of mid-term examinations in some modules, etc.

Regarding the suggestions for improvement, the students stressed that updated educational material for some modules is needed, more understandable, and enriched with digital learning resources (e.g. videos and PowerPoint files). Regarding online tutor-student meetings, several students suggested increasing their number (and consequently the total hours of implementation) so that they will be numerous but of short duration each, to avoid tediousness and to enhance the interaction of the participants in them. Regarding the mid-term examinations, students suggest they should be more and implemented on available dates to allow time for better preparation and performance. Finally, students ask for better cooperation with tutors, through the presentation of practical examples of theory and their guidance and support in the educational activities.

Concerning the proposed model, it is noticeable that, through the finding of thematic communities, Walktrap clearly defined the underlying structure and categorization of related topics within the comments, and provided comprehensive general feedback on their concerns, while using Textrank, the most relatively important comments within each topic were extracted. With this combined approach, a complete comprehension of the student feedback is ensured, exposing both the high-level trends and the subtle insights required for well-informed decision-making and focused programme improvements. Results were also examined by the researchers regarding their plausibility, by a careful reading of the communities' contents and their representative ranked comments. After abstractive summarization, summaries were also carefully red and slightly modified by the researchers to be comprehensible.

8. Discussion

This study aimed to explore students' perceptions in an annual study programme at HOU regarding their distance learning experience by analyzing their open-ended responses in an online questionnaire. The quantitative findings showed that students are in general satisfied with the tutor's role, the study programme design, the administrative - technical services, and to a lesser degree with the educational content. Nevertheless, a richer picture is given with respect to their open-ended

responses. Students had a positive impression of the role of tutors which contributed to their satisfaction with the modules they attended. The findings are in agreement with the literature where it is confirmed that student satisfaction depends to a large extent on the quality of the interaction with the tutor [41,65,66]. Moreover, the desired competencies that tutors should have, include communicating frequently with students [[41], solving their questions and guiding and encouraging them in their studies [67,68], and finally providing them with appropriate feedback on written assignments [69]. Some students expressed their satisfaction with the educational material of the modules. As confirmed by the literature, the quality of student-content interaction (e.g., reading learning articles and textbooks, writing assignments and projects, interacting with multimedia resources) is an important factor that influences the learning experience and the overall satisfaction of students with their studies [70-71].

For a portion of students, the main negative experience they had concerned the poor organization of the educational material of some modules and its outdated information, as well as their problematic interaction with some tutors. In distance education, students spend most of their time reading and writing, systematically studying the educational content to understand the concepts of the subject matter and to further develop their cognitive skills [72]. The educational material therefore should promote students' reflection, discussion, and exchange of information with the aim of collaborative knowledge, searching creatively for solutions, and working out hypotheses to explain and solve various problems, that is, elements that promote critical thinking [73]. The poor quality and structure of the material and the planned activities can significantly limit the development of students' critical thinking skills [74]. Regarding the problematic aspects of tutor-student interaction that were mentioned, literature shows that when tutor-student interaction is effectively practiced it can significantly enhance the understanding of course materials, and encourage community experiences through discussions, group work, and collaborative projects [75]. This aligns with Moore's concept of minimizing transactional distance via engagement and communication [76]. In distance and online educational environments, students value timely feedback, connection, and support from tutors, elements that boost motivation and engagement [77-78]. Feedback on assignments plays a crucial role in developing students' academic skills, motivating students, offering encouragement, and pinpointing areas for improvement and may impact their performance. However, feedback can be problematic due to misconceptions, lack of clarity, and dissatisfaction among students. Issues such as unclear comments affecting self-confidence, delayed delivery hindering progress, and difficulty understanding academic language have been reported as barriers to effective feedback [69]. Finally, students mentioned issues of bad design for some modules (e.g., timetables, deadlines) which had a negative impact on them. The quality of a study progamme's design and structure might drastically affect student satisfaction. The structure of a study programme, encompassing its objectives, teaching and learning methods, and assessment approaches. Key factors influencing the design and structure of a study programme include content presentation, learning paths, personalized guidance, student motivation, etc. Study programmes with rigid structures tend to limit opportunities for communication and interaction in general, whereas more flexible study programmes can reduce transactional distance, fostering better communication and engagement [66].

The approach of automatic text summarization is distinguished based on several methods, the most important of which are what they produce (extractive, abstractive, or hybrid) and whether they need training on the sample data (supervised or unsupervised). The Textrank algorithm combined with Walktrap yields an improved result on the unsupervised extractive summarization, as mentioned. In future work, it is proposed not only to further automate the data preparation and noise removal processes but also to use advanced methods with natural language processing models to improve both extractive and abstractive summaries.

It is suggested that future research delve deeper into the creation of unique natural language processing models that are especially suited to the linguistic subtleties and contextual information present in language student feedback. This could help in resolving the text preprocessing errors discovered by using generic tools such as UDPipe. Exploring the integration of sentiment analysis to

assess the emotional tone of student comments could also be beneficial to gain a more nuanced understanding of the experiences and satisfaction levels of the students. Both UDPipe's Greek lemmatization and comment filtering for detecting the noise in data require manual correction, which is indeed time-consuming. A more automatic technique must be done to speed the process and simultaneously keeping all the important data without overfitting with comments that do not contribute to the analysis. Another interesting prospect is the processing of English comments in parallel and, even better, in a simultaneous analysis with Greek ones.

Advanced Natural Language Processing (NLP) techniques like word embeddings and transformers can be very useful in addressing such issues [79]. Word embeddings will be used in extractive summarization. With this method, words and their context are represented as vectors in a continuous vector space so that words with similar meanings have vectors that are close to each other in the embedding space. This kind of mathematical representation can improve similarity performance and it can be combined with TextRank [72] to improve performance in finding the overall trend, and with Walktrap for more efficient separation of communities. The embeddings will be created through the training process on a large dataset, in this case, the education domain.

Transformers, combined with word embeddings, could be used for abstractive text summarization, producing summaries that resemble human-generated ones without using exact sentences from the original documents. The transformer's architecture, utilizing the attention mechanism, captures long-range dependencies and enables every word in the sentence to pay attention to every other word, regardless of how far apart they are [75]. This surpasses older models like RNNs and LSTMs. Additionally, transformers maintain word order through positional encoding, preserving the original sentence's meaning.

References

[1]   A. R. Masalimova, M. A. Khvatova, L. S. Chikileva, E. P. Zvyagintseva, V. V. Stepanova, and M. V. Melnik, Distance Learning in Higher Education During Covid-19, *Front. Educ.*, vol. 7, Mar. 2022, doi: 10.3389/feduc.2022.822958.

[2]   D. Karapiperis, K. Tzafilkou, R. Tsoni, G. Feretzakis, and V. S. Verykios, A Probabilistic Approach to Modeling Students' Interactions in a Learning Management System for Facilitating Distance Learning, *Information*, vol. 14, no. 8, p. 440, Aug. 2023, doi: 10.3390/info14080440.

[3]   E. Paxinou, E. Manousou, V. S. Verykios, and D. Kalles, Centrality Metrics from Students' Discussion Fora at Distance Education, in *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Volos, Greece: IEEE, Jul. 2023, pp. 1–6. doi: 10.1109/IISA59645.2023.10345914.

[4]   E. Paxinou, G. Feretzakis, R. Tsoni, D. Karapiperis, D. Kalles, and V. S. Verykios, Tracing Student Activity Patterns in E-Learning Environments: Insights into Academic Performance, *Future Internet*, vol. 16, no. 6, p. 190, May 2024, doi: 10.3390/fi16060190.

[5]   J. R. Evans and A. Mathur, The value of online surveys: a look back and a look ahead, *Internet Research*, vol. 28, no. 4, pp. 854–887, Jan. 2018, doi: 10.1108/IntR-03-2018-0089.

[6]   G. Merkys and D. Bubeliene, Quantification of Textual Responses To Open-Ended Questions In Big Data, in *Modeling and Simulation of Social-Behavioral Phenomena in Creative Societies*, N. Agarwal, G. B. Kleiner, and L. Sakalauskas, Eds., Cham: Springer Nature Switzerland, 2023, pp. 191–200. doi: 10.1007/978-3-031-33728-4_14.

[7]   A.-C. Haensch, B. Weiß, P. Steins, P. Chyrva, and K. Bitz, The semi-automatic classification of an open-ended question on panel survey motivation and its application in attrition analysis, *Front. Big Data*, vol. 5, Aug. 2022, doi: 10.3389/fdata.2022.880554.

[8]   A. Fiori, *Trends and Applications of Text Summarization Techniques*. IGI Global, 2019.

[9] H. W. Marsh, SEEQ: A Reliable, Valid, and Useful Instrument for Collecting Students' Evaluations of University Teaching, *British Journal of Educational Psychology*, vol. 52, pp. 77–95, Feb. 1982.

[10] P. Ramsden, Learning to Teach in Higher Education, Routledge & CRC Press. Accessed: Jun. 28, 2024. [Online]. Available: https://www.routledge.com/Learning-to-Teach-in-Higher-Education/Ramsden-Ramsden/p/book/9780415303453

[11] J. T. E. Richardson, Instruments for obtaining student feedback: a review of the literature, *Assessment & Evaluation in Higher Education*, vol. 30, no. 4, pp. 387–415, Aug. 2005, doi: 10.1080/02602930500099193.

[12] F. Martin and D. U. Bolliger, Developing an online learner satisfaction framework in higher education through a systematic review of research, *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 50, Sep. 2022, doi: 10.1186/s41239-022-00355-5.

[13] S. L. Baglione and Z. Smith, Grade inflation: undergraduate students' perspective, *Quality Assurance in Education*, vol. 30, no. 2, pp. 251–267, Jan. 2022, doi: 10.1108/QAE-08-2021-0134.

[14] P. Spooren, B. Brockx, and D. Mortelmans, On the Validity of Student Evaluation of Teaching: The State of the Art, *Review of Educational Research*, vol. 83, no. 4, pp. 598–642, 2013.

[15] C. Zucco, C. Paglia, S. Graziano, S. Bella, and M. Cannataro, Sentiment Analysis and Text Mining of Questionnaires to Support Telemonitoring Programs, *Information*, vol. 11, no. 12, Art. no. 12, Dec. 2020, doi: 10.3390/info11120550.

[16] M.-J. Wu, K. Zhao, and F. Fils-Aime, Response rates of online surveys in published research: A meta-analysis, *Computers in Human Behavior Reports*, vol. 7, p. 100206, Aug. 2022, doi: 10.1016/j.chbr.2022.100206.

[17] D. Wallace, G. Cesar, and E. C. Hedberg, The Effect of Survey Mode on Socially Undesirable Responses to Open-ended Questions: A Mixed Methods Approach, *Field Methods*, vol. 30, no. 2, pp. 105–123, May 2018, doi: 10.1177/1525822X18766284.

[18] D. P. Stergiou and D. Airey, Using the Course Experience Questionnaire for evaluating undergraduate tourism management courses in Greece, *Journal of Hospitality, Leisure, Sport & Tourism Education*, vol. 11, no. 1, pp. 41–49, Apr. 2012, doi: 10.1016/j.jhlste.2012.02.002.

[19] H. W. Marsh, Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness, in *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, R. P. Perry and J. C. Smart, Eds., Dordrecht: Springer Netherlands, 2007, pp. 319–383. doi: 10.1007/1-4020-5742-3_9.

[20] G. Chung, M. Rodriguez, P. Lanier, and D. Gibbs, Text-Mining Open-Ended Survey Responses Using Structural Topic Modeling: A Practical Demonstration to Understand Parents' Coping Methods During the COVID-19 Pandemic in Singapore, *Journal of Technology in Human Services*, vol. 40, no. 4, pp. 296–318, Oct. 2022, doi: 10.1080/15228835.2022.2036301.

[21] V. Baburajan, J. D. A. E Silva, and F. C. Pereira, Open-Ended Versus Closed-Ended Responses: A Comparison Study Using Topic Modeling and Factor Analysis, *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 4, pp. 2123–2132, Apr. 2021, doi: 10.1109/TITS.2020.3040904.

[22] C. E. Neuert, K. Meitinger, and D. Behr, Open-ended versus Closed Probes: Assessing Different Formats of Web Probing, *Sociological Methods & Research*, vol. 52, no. 4, pp. 1981–2015, Nov. 2023, doi: 10.1177/00491241211031271.

[23] N. Maloshonok and E. Terentev, The impact of visual design and response formats on data quality in a web survey of MOOC students, *Computers in Human Behavior*, vol. 62, pp. 506–515, Sep. 2016, doi: 10.1016/j.chb.2016.04.025.

[24] A. Hofelich Mohr, A. Sell, and T. Lindsay, Thinking Inside the Box: Data from an Online Alternative Uses Task with Visual Manipulation of the Survey Response Box, Sep. 2016, Accessed: Jul. 15, 2024. [Online]. Available: https://hdl.handle.net/11299/172116

[25] J. Saldana, *The Coding Manual for Qualitative Researchers*. SAGE, 2021.

[26] K. A. Neuendorf, *The Content Analysis Guidebook*. SAGE Publications, Inc, 2017. doi: 10.4135/9781071802878.

[27] G. Hunt, M. Moloney, and A. Fazio, Embarking on large-scale qualitative research: reaping the benefits of mixed methods in studying youth, clubs and drugs, *Nordisk Alkohol Nark*, vol. 28, no. 5–6, pp. 433–452, Dec. 2011, doi: 10.2478/v10199-011-0040-1.

[28] M. S. González Canché, Machine driven classification of open-ended responses (MDCOR): An analytic framework and no-code, free software application to classify longitudinal and cross-sectional text responses in survey and social media research, *Expert systems with applications*, vol. 215, pp. 119265-, 2023, doi: 10.1016/j.eswa.2022.119265.

[29] R. Popping, Analyzing Open-ended Questions by Means of Text Analysis Procedures, *BMS: Bulletin of Sociological Methodology / Bulletin de Méthodologie Sociologique*, no. 128, pp. 23–39, 2015.

[30] C. Zhai and S. Massung, *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Morgan & Claypool, 2016.

[31] M. Allahyari *et al.*, A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques, Jul. 28, 2017, *arXiv*: arXiv:1707.02919. doi: 10.48550/arXiv.1707.02919.

[32] S. Alhojely and J. Kalita, Recent Progress on Text Summarization, in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2020, pp. 1503–1509. doi: 10.1109/CSCI51800.2020.00278.

[33] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*, 1st ed. 2019 edition. New York, NY: Springer, 2018.

[34] D. Saikumar and P. Subathra, Two-Level Text Summarization Using Topic Modeling, in *Intelligent System Design*, S. C. Satapathy, V. Bhateja, B. Janakiramaiah, and Y.-W. Chen, Eds., Singapore: Springer, 2021, pp. 153–167. doi: 10.1007/978-981-15-5400-1_16.

[35] N. Giarelis, C. Mastrokostas, and N. Karacapilidis, Abstractive vs. Extractive Summarization: An Experimental Review, *Applied Sciences*, vol. 13, no. 13, Art. no. 13, Jan. 2023, doi: 10.3390/app13137620.

[36] M. G. Moore, Editorial: Three types of interaction, *American Journal of Distance Education*, vol. 3, no. 2, pp. 1–7, Jan. 1989, doi: 10.1080/08923648909526659.

[37] E. Katsarou and P. Chatzipanagiotou, A Critical Review of Selected Literature on Learner-centered Interactions in Online Learning, *Electronic Journal of e-Learning*, vol. 19, no. 5, Art. no. 5, Nov. 2021, doi: 10.34190/ejel.19.5.2469.

[38] D. Rowntree, *Exploring Open and Distance Learning*. Kogan Page, 1992.

[39] D. Rowntree, *Preparing Materials for Open, Distance & Flexible Learning: An Action Guide for Teachers and Trainers*, 1st edition. London: Kogan Page, 1993.

[40] A. Lionarakis, S. Papadimitriou, and V. Ioakimidou, The Hellenic Open University: Innovations and Challenges in Greek Higher Education, *The Hellenic Open University: Innovations and Challenges in Greek Higher Education*, no. 1, pp. 6–22, Sep. 2019, doi: https://doi.org/10.12681/jode.20211.

[41] P. Anastasiades and C. Iliadou, Communication between Tutors--Students in DL: A Case Study of the Hellenic Open University, *European Journal of Open, Distance and E-Learning*, 2010, doi: 10/Panagiotis_Chrysoula.pdf.

[42] G. Vorvilas, A. Liapis, A. Korovesis, D. Aggelopoulou, N. Karousos, and E. Efstathopoulos, CONDUCTING REMOTE ELECTRONIC EXAMINATIONS IN DISTANCE HIGHER EDUCATION: STUDENTS' PERCEPTIONS, *TOJDE*, vol. 24, no. 2, Art. no. 2, Apr. 2023, doi: 10.17718/tojde.971889.

[43] G. Vorvilas, A. Liapis, N. Karousos, L. Theodorakopoulos, E. Lagiou, and A. Kameas, Faculty members' Perceptions of Remote Electronic Examinations in Distance Academic Education, in *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Jul. 2023, pp. 1–6. doi: 10.1109/IISA59645.2023.10345883.

[44] G. Vorvilas, A. Liapis, A. Korovesis, V. Athanasakopoulou, N. Karousos, and A. Kameas, *Evaluating the quality of master's thesis supervision in academic distance education: Hellenic Open University students' perceptions*. 2022. doi: 10.5281/zenodo.7330856.

[45] A. Bryman, E. Bell, J. Reck, and J. Fields, *Social Research Methods*, 1st edition. New York, NY: Oxford University Press, 2021.

[46] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. SAGE, 2004.

[47] A. Kumar and A. Paul, *Mastering Text Mining with R: Extract and recognize your text data*. Birmingham Mumbai: Packt Publishing, 2016.

[48] Warto *et al.*, Systematic Literature Review on Named Entity Recognition: Approach, Method, and Application, *Statistics, Optimization & Information Computing*, vol. 12, no. 4, Art. no. 4, Feb. 2024, doi: 10.19139/soic-2310-5070-1631.

[49] M. Anandarajan, C. Hill, and T. Nolan, *Practical Text Analytics: Maximizing the Value of Text Data*. Springer, 2018.

[50] R. Mihalcea and P. Tarau, TextRank: Bringing Order into Text, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, D. Lin and D. Wu, Eds., Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 404–411. Accessed: Apr. 02, 2024. [Online]. Available: https://aclanthology.org/W04-3252

[51] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, presented at the The Web Conference, Nov. 1999. Accessed: Apr. 19, 2024. [Online]. Available: https://www.semanticscholar.org/paper/The-PageRank-Citation-Ranking-%3A-Bringing-Order-to-Page-Brin/eb82d3035849cd23578096462ba419b53198a556

[52] K. S. Thakkar, R. V. Dharaskar, and M. B. Chandak, Graph-Based Algorithms for Text Summarization, in *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, Nov. 2010, pp. 516–519. doi: 10.1109/ICETET.2010.104.

[53] P. Pons and M. Latapy, Journal of Graph Algorithms and Applications Computing Communities in Large Networks Using Random Walks, Accessed: Jun. 06, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Journal-of-Graph-Algorithms-and-Applications-in-Pons-Latapy/51e4f920c54cc8794f0fe68f0bfe1d6e122c19ff

[54] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E*, vol. 69, no. 2, p. 026113, Feb. 2004, doi: 10.1103/PhysRevE.69.026113.

[55] Y. Zhou, H. Cheng, and J. X. Yu, Graph clustering based on structural/attribute similarities, *Proc. VLDB Endow.*, vol. 2, no. 1, pp. 718–729, Aug. 2009, doi: 10.14778/1687627.1687709.

[56] M. E. J. Newman, Detecting community structure in networks, *Eur. Phys. J. B*, vol. 38, no. 2, pp. 321–330, Mar. 2004, doi: 10.1140/epjb/e2004-00124-y.

[57] Welcome to Flask — Flask Documentation (3.0.x). Accessed: Jul. 22, 2024. [Online]. Available: https://flask.palletsprojects.com/en/3.0.x/

[58] M. M. Danilak, *langdetect: Language detection library ported from Google's language-detection.* Python. Accessed: Jul. 22, 2024. [OS Independent]. Available: https://github.com/Mimino666/langdetect

[59] spaCy · Industrial-strength Natural Language Processing in Python. Accessed: Jun. 10, 2024. [Online]. Available: https://spacy.io/

[60] EntityRecognizer · spaCy API Documentation, EntityRecognizer. Accessed: Jun. 10, 2024. [Online]. Available: https://spacy.io/api/entityrecognizer

[61] TakeLab, *spacy-udpipe: Use fast UDPipe models directly in spaCy*. Python. Accessed: Jun. 10, 2024. [OS Independent]. Available: https://github.com/TakeLab/spacy-udpipe

[62] scikit-learn: machine learning in Python — scikit-learn 1.5.1 documentation. Accessed: Jul. 22, 2024. [Online]. Available: https://scikit-learn.org/stable/

[63] python-igraph stable — igraph stable documentation. Accessed: Jul. 22, 2024. [Online]. Available: https://python.igraph.org/en/stable/

[64] GPT-4o mini: advancing cost-efficient intelligence. Accessed: Jul. 22, 2024. [Online]. Available: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

[65] E. Anagnostopoulou, I. Mavroidis, Y. Giossos, and M. Koutsouba, Student Satisfaction in The Context of A Postgraduate Programme of The Hellenic Open University, *Turkish Online Journal of Distance Education*, vol. 16, no. 2, Art. no. 2, Apr. 2015, doi: 10.17718/tojde.52944.

[66] V. Gavrilis, I. Mavroidis, and Y. Giossos, Transactional Distance and Student Satisfaction in a Postgraduate Distance Learning Program, *Turkish Online Journal of Distance Education*, vol. 21, no. 3, pp. 48–62, Jul. 2020.

[67] N. de Metz and A. Bezuidenhout, An importance–competence analysis of the roles and competencies of e-tutors at an open distance learning institution, *Australasian Journal of Educational Technology*, vol. 34, no. 5, Art. no. 5, Nov. 2018, doi: 10.14742/ajet.3364.

[68] S. Li, J. Zhang, C. Yu, and L. Chen, Rethinking Distance Tutoring in e-Learning Environments: A Study of the Priority of Roles and Competencies of Open University Tutors in China, *irrodl*, vol. 18, no. 2, pp. 189–212, 2017, doi: 10.19173/irrodl.v18i2.2752.

[69] G. Kreonidou and V. Kazamia, Assignment feedback in distance education: How do students perceive it?", *Research Papers in Language Teaching and Learning*, vol. 10, no. 1, pp. 134–153, 2019.

[70] Y.-C. Kuo, A. E. Walker, K. E. E. Schroder, and B. R. Belland, Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses, *The Internet and Higher Education*, vol. 20, pp. 35–50, Jan. 2014, doi: 10.1016/j.iheduc.2013.10.001.

[71] P. Kumar, C. Saxena, and H. Baber, Learner-content interaction in e-learning- the moderating role of perceived harm of COVID-19 in assessing the satisfaction of learners, *Smart Learning Environments*, vol. 8, no. 1, p. 5, Apr. 2021, doi: 10.1186/s40561-021-00149-8.

[72] Y.-C. Kuo, Interaction, Internet Self-Efficacy, and Self-Regulated Learning as Predictors of Student Satisfaction in Distance Education Courses, *All Graduate Theses and Dissertations, Spring 1920 to Summer 2023*, May 2010, doi: https://doi.org/10.26076/157a-7cf6.

[73] M. Mena, New pedagogical approaches to improve production of materials in distance education, International Journal of E-Learning & Distance Education / Revue internationale du e-learning et la formation à distance. Accessed: Jul. 22, 2024. [Online]. Available: https://www.ijede.ca/index.php/jde/article/view/510

[74] V. Giannouli and G. Vorvilas, Barriers in fostering critical thinking in higher distance education: faculty members' perceptions, *Mediterranean Journal of Education*, vol. 3, no. 1, Art. no. 1, Jan. 2023, doi: 10.26220/mje.4144.

[75] B. Delaney and K. Betts, Addressing Transactional Distance Through Teaching Presence Strategies in Online Journalism and Mass Communication Courses, *Journalism & Mass Communication Educator*, vol. 77, no. 1, pp. 5–23, Mar. 2022, doi: 10.1177/10776958211001214.

[76] M. Moore, Keegan, D., ed. "Theoretical Principles of Distance Education, Routledge, 1997, pp. 22–38.

[77] D. Keržič *et al.*, Academic student satisfaction and perceived performance in the e-learning environment during the COVID-19 pandemic: Evidence across ten countries, *PLoS One*, vol. 16, no. 10, p. e0258807, 2021, doi: 10.1371/journal.pone.0258807.

[78] S. B. Eom and N. Ashill, The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Update, *Decision Sciences Journal of Innovative Education*, vol. 14, no. 2, pp. 185–215, 2016, doi: 10.1111/dsji.12097.

[79] D. S. Asudani, N. K. Nagwani, and P. Singh, Impact of word embedding models on text analytics in deep learning environment: a review, *Artif Intell Rev*, pp. 1–81, Feb. 2023, doi: 10.1007/s10462-023-10419-1.

21

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.