FORM 2

THE PATENTS ACT, 1970

(39 of 1970)

COMPLETE SPECIFICATION

(See section 10; rule 13)

1. TITLE OF THE INVENTION: **RETRIEVAL-AUGMENTED GENERATION SYSTEM FOR VOCABULARY ACQUISITION THROUGH ADAPTIVE LEARNING AND METHOD THEREOF**

2. APPLICANT (S):

(a) NAME : PRAVEEN KUMAR ANWLA

(b) NATIONALITY : INDIAN

(c) ADDRESS : KRISHNA BHAWAN, SABZI MANDI, ALANPUR, SAWAI MADHOPUR, RAJASTHAN 322021, INDIA

THE FOLLOWING SPECIFICATION PARTICULARLY DESCRIBES THE INVENTION AND THE MANNER IN WHICH IT IS TO BE PERFORMED.

FIELD OF INVENTION

[0001]Embodiments of the present invention relate to educational technology, and more particularly relate to a retrieval-augmented generation (RAG) system for vocabulary acquisition through adaptive learning, with a primary goal of enhancing vocabulary learning experiences through information retrieval and generation.

BACKGROUND

[0002]In the realm of educational technology, the significance of addressing challenges encountered by students and individuals, especially those from vernacular backgrounds, cannot be overstated. Learning unfamiliar words in a brief period, particularly for the individuals from vernacular backgrounds, poses a considerable challenge. This challenge is compounded by the fact that people may forget words quickly if they are not regularly utilized in daily communication. Conventional methods of vocabulary acquisition, utilizing tools like flashcards, books, and Portable Document Format (PDF) guides, have proven insufficient in fostering long-term retention and practical application of newly acquired English words. For instance, students preparing for GRE, GMAT, CAT, etc., often grapple with the need for effective vocabulary retention given the fast-paced nature of exam preparation.

[0003]One of the predominant issues with traditional learning approaches is the tendency of individuals to forget newly learned words, particularly when not regularly incorporated into daily written or verbal communication. This challenge is exacerbated for non-native English speakers facing the additional pressure of quickly assimilating a substantial number of words within a limited timeframe for competitive exams.

[0004]Students resorting to rote memorisation through flashcards, books, or PDF guides encounter a common obstacle – the transience of their newly acquired vocabulary. The need for effective and enduring vocabulary learning tools becomes increasingly evident in scenarios where individuals' primary language is not English, and time constraints intensify the pressure for rapid and efficient learning.

[0005]Moreover, the psychological barrier hindering individuals from incorporating newly acquired words into their daily communication poses a significant impediment. The fear of incorrectly using words in verbal or written exchanges further compounds the challenges faced by learners. This apprehension is particularly pronounced among those with limited exposure to the English language, impeding not only their learning journey but also their personal and professional growth.

[0006]In the existing technology, a vocabulary learning support system and method are disclosed. The method for enhancing the vocabulary of a learner involves accessing a personalised vocabulary profile containing both unstable terms that the learner is in the process of learning and stable terms that the learner has already mastered. The method further includes identifying a term within an electronic document displayed on a device as an unstable term if it corresponds to one of the terms in the learner's unstable vocabulary profile. Subsequently, the method involves capturing input from the learner to ascertain their understanding of the marked term. Based on the received input, the system determines whether the learner comprehends the marked unstable term and updates the personalised vocabulary profile accordingly. However, the vocabulary learning support system relies on a static approach to vocabulary learning, where terms are categorised as either stable or unstable. The vocabulary learning support system does not consider the dynamic nature of a learner's understanding of words over time.

[0007]There are various technical problems with the educational technology in the prior art. In the existing technology, learners often encounter limitations that hinder the effectiveness of vocabulary learning. One notable drawback lies in the static nature of vocabulary profiling. The existing technology relies on a fixed categorisation of terms as either stable or unstable, overlooking the dynamic evolution of the understanding of words over time. Furthermore, the prior art tends to overlook the psychological aspects of language acquisition, particularly the apprehension learners feel in incorporating newly acquired words into their daily communication. This hesitation is a significant barrier, impacting the practical application of learned vocabulary. Additionally, the conventional methods often lack the integration of advanced natural language processing techniques, limiting their ability to provide a responsive and personalised learning environment.

[0008]Therefore, there is a need for a system to address the aforementioned issues by providing a dynamic retrieval-augmented generation (RAG) during a learning stage.

SUMMARY

[0009]This summary is provided to introduce a selection of concepts, in a simple manner, which is further described in the detailed description of the disclosure. This summary is neither intended to identify key or essential inventive concepts of the subject matter nor to determine the scope of the disclosure.

[0010]In order to overcome the above deficiencies of the prior art, the present disclosure is to solve the technical problem by providing a retrieval-augmented generation (RAG) system for vocabulary acquisition through adaptive learning, with a primary goal of enhancing vocabulary learning experiences through information retrieval and generation.

[0011]In accordance with an embodiment of the present disclosure, the retrieval-augmented generation (RAG) system for vocabulary acquisition through adaptive learning is disclosed. The RAG system comprises a server, a data repository unit, one or more hardware processors, and a plurality of subsystems. The plurality of subsystems comprises a resource data-obtaining subsystem, a data processing subsystem, a query receiving subsystem, a query encoding subsystem, and a response generation subsystem.

[0012]In an embodiment, the server is configured with the one or more hardware processors. The data repository unit is operatively coupled to the one or more hardware processors. The data repository unit comprises the plurality of subsystems in form of programmable instructions executable by the one or more hardware processors.

[0013]In an embodiment, the resource data-obtaining subsystem is configured to obtain learning resource data for storing in a database. The learning resource data comprises at least one of: flashcards, books, Portable Document Format (PDF) documents with primary and secondary meanings and examples, and documents with a list of vocabulary words.

[0014]In an embodiment, the data processing subsystem is configured to extract vocabulary information from the obtained learning resource data for organising the extracted vocabulary information into a first set of dense vectors to store in a vector database.

[0015]In an embodiment, the query receiving subsystem is configured to receive one or more search queries from a user associated with a user profile for which one or more responses are required to be generated. The query receiving subsystem is configured with a retrieval-augmented generation (RAG) module to retrieve the associated first set of dense vectors from the vector database information based on a dense vector similarity search method. The retrieval-augmented generation (RAG) module is configured with one or more libraries comprises at least one of a: Lama-Index, LangChain, and Haystack.

[0016]In an embodiment, the query encoding subsystem is configured to provide a second set of dense vectors by converting the received one or more search queries into fixed-dimensional vectors for retrieving the associated first set of dense vectors from the vector database. The query encoding subsystem is configured with transformer-based encoder blocks configured to convert the received one or more search queries into the second set of dense vectors for retrieval of the first set of dense vectors.

[0017]In an embodiment, the response generation subsystem is configured to generate the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through the adaptive learning. The response generation subsystem is configured with Large Language Models (LLMs) to generate the one or more responses. The Large Language Models (LLMs) comprise one of a: Mistral 7B, LlaMA 2, BLOOM, Falcon, Generative Pre-trained Transformer (GPT) models, Bidirectional Encoder Representations from Transformers (BERT), eXtreme Learning with Transformers (XLNet), Text-To-Text Transfer Transformer (T5), and Robustly optimised BERT approach (RoBERTa).

[0018]The response generation subsystem is configured to adapt to user preferences and learning history dynamically, providing personalised one or more responses associated with the one or more search queries based on the user profile. The response generation subsystem is configured to prioritise generated the one or more responses based on a user feedback collected through a user interface subsystem. The user interface subsystem is configured to present a visual representation of a word cloud displaying words used by the user as the one or more responses and enabling users to access the history of usage for words.

[0019]In an embodiment, the RAG system comprises a reinforcement learning subsystem. The reinforcement learning subsystem is configured to continuously train and refine the response generation subsystem based on the user feedback received through the user interface subsystem. The RAG system comprises a lexical relations graph generation subsystem. The lexical relations graph generation subsystem is configured to analyse the one or more search queries to retrieve the first set of dense vectors with compiled information. The compiled information comprises at least one of: synonyms, antonyms, words in analogous contexts, and instances of usage in discrete contexts.

[0020]In accordance with an embodiment of the present disclosure, a retrieval-augmented generation (RAG) method for vocabulary acquisition through adaptive learning is disclosed. In a first step, the RAG method includes, obtaining, by the resource data-obtaining subsystem, the learning resource data to store in the database. In the next step, the RAG method includes extracting, by the data processing subsystem, vocabulary information from the obtained learning resource data to organise the extracted vocabulary information into the first set of dense vectors to store in the vector database.

[0021]In the next step, the RAG method includes receiving, by the query receiving subsystem, the one or more search queries from the user associated with the user profile for which the one or more responses are required to be generated. In the next step, the RAG method includes providing, by the query encoding subsystem, the second set of dense vectors by converting the received one or more search queries into the fixed-dimensional vectors to retrieve the associated first set of dense vectors from the vector database. In the next step, the RAG method includes generating, by the response generation subsystem, the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through adaptive learning.

[0022]To further clarify the advantages and features of the present invention, a more particular description of the invention will follow by reference to specific embodiments thereof, which are illustrated in the appended figures. It is to be appreciated that these figures depict only typical embodiments of the invention and are therefore not to be considered limiting in scope. The invention will be described and explained with additional specificity and detail with the appended figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023]The disclosure will be described and explained with additional specificity and detail with the accompanying figures in which:

[0024]FIG. 1 illustrates an exemplary block diagram representation of a network architecture of a retrieval-augmented generation system for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure;

[0025]FIGs. 2A and 2B illustrate an exemplary block diagram representation of the retrieval-augmented generation system as shown in **FIG. 1** for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure;

[0026]FIG. 3 illustrates an exemplary flow chart depicting a retrieval-augmented generation method for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure;

[0027]FIG. 4 illustrates an exemplary user interface depicting the retrieval-augmented generation system for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure; and

[0028]FIG. 5 illustrates an exemplary snapshot of a third-party application configured with the retrieval-augmented generation system for vocabulary acquisition, in accordance with an embodiment of the present disclosure.

[0029]Further, those skilled in the art will appreciate that elements in the figures are illustrated for simplicity and may not have necessarily been drawn to scale. Furthermore, in terms of the method steps, and parameters used herein may have been represented in the figures by conventional symbols, and the figures may show only those specific details that are pertinent to understanding the embodiments of the present disclosure so as not to obscure the figures with details that will be readily apparent to those skilled in the art having the benefit of the description herein.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0030]For the purpose of promoting an understanding of the principles of the disclosure, reference will now be made to the embodiment illustrated in the figures and specific language will be used to describe them. It will nevertheless be understood that no limitation of the scope of the disclosure is thereby intended. Such alterations and further modifications in the illustrated system, and such further applications of the principles of the disclosure as would normally occur to those skilled in the art are to be construed as being within the scope of the present disclosure.

[0031]The terms "comprises", "comprising", or any other variations thereof, are intended to cover a non-exclusive inclusion, such that a process or method that comprises a list of steps does not include only those steps but may include other steps not expressly listed or inherent to such a process or method. Similarly, one or more components, compounds, and ingredients preceded by "comprises... a" does not, without more constraints, preclude the existence of other components or compounds or ingredients or additional components. Appearances of the phrase "in an embodiment", "in another embodiment" and similar language throughout this specification may, but not necessarily do, all refer to the same embodiment.

[0032]Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by those skilled in the art to which this disclosure belongs. The system, methods, and examples provided herein are only illustrative and not intended to be limiting.

[0033]In the following specification and the claims, reference will be made to a number of terms, which shall be defined to have the following meanings. The singular forms “a”, “an”, and “the” include plural references unless the context clearly dictates otherwise.

[0034]Embodiments of the present disclosure relate to a retrieval-augmented generation system for vocabulary acquisition through adaptive learning, with a primary goal of enhancing vocabulary learning experiences through information retrieval and generation.

[0035]FIG. 1 refers to an exemplary block diagram representation of a network architecture 100 of the retrieval-augmented generation system 102 for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure.

[0036]According to an exemplary embodiment of the present disclosure, **FIG. 1** depicts the network architecture 100 may include the retrieval-augmented generation system 102 (hereinafter referred to as the system 102), a database 104, and one or more communication devices 106. The system 102 may be communicatively coupled to the database 104 via a communication network 108. The communication network 108 may be a wired communication network and/or a wireless communication network. The database 104 may include, but not limited to, storing, and managing data related to learning resource data. The database 104 may be any kind of database such as, but not limited to, relational databases, non-relational databases, document databases, dedicated databases, dynamic databases, monetised databases, scalable databases, cloud databases, distributed databases, any other databases, and a combination thereof. The database 104 is configured to support the functionality of the system 102 and enables efficient data retrieval and storage for various aspects associated with an effective learning environment. The one or more communication devices 106 may be digital devices, computing devices and/or networks. The one or more communication devices 106 may include, but not limited to, a mobile device, a smartphone, a Personal Digital Assistant (PDA), a tablet computer, a phablet computer, a wearable computing device, a laptop, a desktop, and the like.

[0037]This integrated network architecture 100 facilitates seamless communication and data exchange, enabling the system 102 to operate cohesively for the dynamic retrieval-augmented generation during the learning stage. The system 102 capability to provide a personalised learning experience to each user is underpinned by the effective collaboration among the system 102, the database 104, and the one or more communication devices 106 within the communication network 108.

[0038]Further, the system 102 may be implemented by way of a single device or a combination of multiple devices that may be operatively connected or networked together. The system 102 may be implemented in hardware or a suitable combination of hardware and software. The system 102 includes a server 116 and a data repository unit 112. The server 116 may include one or more hardware processors 110. The data repository unit 112 may include a plurality of subsystems 114. The system 102 may be a hardware device including the one or more hardware processors 110 executing machine-readable program instructions for the dynamic retrieval-augmented generation during the learning stage. Execution of the machine-readable program instructions by the one or more hardware processors 110 may enable the system 102 to dynamically recommend course of action sequences for the dynamic retrieval-augmented generation during the learning stage. The course of action sequences may involve various steps or decisions taken for data-obtaining, data processing, query receiving, query encoding, and response generating. The “hardware” may comprise a combination of discrete components, an integrated circuit, an application-specific integrated circuit, a field-programmable gate array, a digital signal processor, or other suitable hardware. The “software” may comprise one or more objects, agents, threads, lines of code, subroutines, separate software applications, two or more lines of code, or other suitable software structures operating in one or more software applications or on one or more processors.

[0039]The server 116 may be an optimal-performance computing unit configured with the one or more hardware processors 110 (e.g., CPUs or GPUs) designed to handle complex computations and efficiently process large datasets. The server 116 is configured to orchestrate various subsystems, execute algorithms, and manage the overall functionality of the system 102 for vocabulary learning. The server 116 is configured as a computational backbone, processing user queries, retrieving information, and generating contextually relevant responses to enhance the vocabulary learning experience.

[0040]The one or more hardware processors 110 may include, for example, microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuits, and/or any devices that manipulate data or signals based on operational instructions. Among other capabilities, the one or more hardware processors 110 may fetch and execute computer-readable instructions in the data repository unit 112 operationally coupled with the system 102 for performing tasks such as data processing, input/output processing, and/or any other functions. Any reference to a task in the present disclosure may refer to an operation being or that may be performed on data.

[0041]Though few components and subsystems are disclosed in **FIG. 1**, there may be additional components and subsystems which is not shown, such as, but not limited to, ports, routers, repeaters, firewall devices, network devices, databases, network attached storage devices, servers, assets, machinery, instruments, facility equipment, emergency management devices, image capturing devices, any other devices, and combination thereof. A person skilled in the art should not be limiting the components/subsystems shown in **FIG. 1**.

[0042]Those of ordinary skilled in the art will appreciate that the hardware depicted in **FIG. 1** may vary for particular implementations. For example, other peripheral devices such as an optical disk drive and the like, local area network (LAN), wide area network (WAN), wireless (e.g., wireless-fidelity (Wi-Fi)) adapter, graphics adapter, disk controller, input/output (I/O) adapter also may be used in addition or place of the hardware depicted. The depicted example is provided for explanation only and is not meant to imply architectural limitations concerning the present disclosure.

[0043]Those skilled in the art will recognise that, for simplicity and clarity, the full structure and operation of all data processing systems suitable for use with the present disclosure are not being depicted or described herein. Instead, only so much of the system 102 as is unique to the present disclosure or necessary for an understanding of the present disclosure is depicted and described. The remainder of the construction and operation of the system 102 may conform to any of the various current implementations and practices that were known in the art.

[0044]FIGs. 2A and 2B refer to an exemplary block diagram representation 200 of the retrieval-augmented generation system 102 as shown in **FIG. 1** for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure.

[0045]In an exemplary embodiment, the system 102 comprises the one or more hardware processors 110, the data repository unit 112, and the database 104. The one or more hardware processors 110, the data repository unit 112, and the database 104 are communicatively coupled through a system bus 202 or any similar mechanism. The data repository unit 112 is operatively coupled to the one or more hardware processors 110. The data repository unit 112 comprises the plurality of subsystems 114 in the form of programmable instructions executable by the one or more hardware processors 110.

[0046]In an exemplary embodiment, the plurality of subsystems 114 comprises a resource data-obtaining subsystem 204, a data processing subsystem 206, a query receiving subsystem 208, a query encoding subsystem 210, and a response generation subsystem 212. The plurality of subsystems 114 dynamically retrieves and augments data during the learning stage, thereby ultimately generating the one or more responses based on user queries. The division of tasks among the plurality of subsystems 114 assists in achieving a modular and efficient approach to the overall functionality of the system 102.

[0047]The one or more hardware processors 110, as used herein, means any type of computational circuit, such as, but not limited to, a microprocessor unit, microcontroller, complex instruction set computing microprocessor unit, reduced instruction set computing microprocessor unit, very long instruction word microprocessor unit, explicitly parallel instruction computing microprocessor unit, graphics processing unit, digital signal processing unit, or any other type of processing circuit. The one or more hardware processors 110 may also include embedded controllers, such as generic or programmable logic devices or arrays, application-specific integrated circuits, single-chip computers, and the like.

[0048]The data repository unit 112 may be a non-transitory volatile memory and a non-volatile memory. The data repository unit 112 may be coupled to communicate with the one or more hardware processors 110, such as being a computer-readable storage medium. The one or more hardware processors 110 may execute machine-readable instructions and/or source code stored in the data repository unit 112. A variety of machine-readable instructions may be stored in and accessed from the data repository unit 112. The data repository unit 112 may include any suitable elements for storing data and machine-readable instructions, such as read-only memory, random access memory, erasable programmable read-only memory, electrically erasable programmable read-only memory, a hard drive, a removable media drive for handling compact disks, digital video disks, diskettes, magnetic tape cartridges, memory cards, and the like. In the present embodiment, the data repository unit 112 includes the plurality of subsystems 114 stored in the form of machine-readable instructions on any of the above-mentioned storage media and may be in communication with and executed by the one or more hardware processors 110.

[0049]In an exemplary embodiment, the resource data-obtaining subsystem 204 is configured to acquire the learning resource data, which is subsequently stored in the database 104. The learning resource data may include, but not limited to at least one of: flashcards, books, Portable Document Format (PDF) documents with primary and secondary meanings and examples, documents with a list of vocabulary words, and the like. By efficiently gathering and organising the multifaceted learning resource data, the system 102 enhances its capacity to provide comprehensive and contextually relevant one or more responses during the learning stage.

[0050]In an exemplary embodiment, the data processing subsystem 206 is configured to extract vocabulary information from the acquired learning resource data. The data processing subsystem 206 subsequently organises the extracted vocabulary information into a first set of dense vectors to store in a vector database. By systematically parsing and categorising the extracted vocabulary information, the data processing subsystem 206 contributes to the establishment of the first set of dense vectors, thereby ensuring efficient storage and retrieval of the first set of dense vectors in the vector database. The vector database enhances the capability of the system 102 to respond contextually and comprehensively during the learning stage, thereby providing users with a valuable resource for language acquisition. The vector databases such as, but not limited to, chroma, Qdrant, Pinecone, Weaviate, and the like.

[0051]In an exemplary embodiment, the query receiving subsystem 208 is configured to receive one or more search queries from a user associated with a user profile for which the one or more responses are required to be generated. The query receiving subsystem 208 is configured with a retrieval-augmented generation (RAG) module to retrieve the associated first set of dense vectors from the vector database based on a dense vector similarity search method. The retrieval-augmented generation (RAG) module is configured with one or more libraries that comprises, but not limited to, at least one of a: Lama-Index, LangChain, Haystack and the like. The RAG module enables the system 102 to conduct sophisticated searches, ensuring accuracy and relevance in retrieving the vocabulary information.

[0052]The dense vector similarity search method is a technique used in information retrieval and natural language processing, particularly when dealing with large datasets of fixed-dimensional vectors. Words and terms are represented as the dense vectors, where each dimension of the dense vectors corresponds to a specific feature and aspect of the word. The dense vector representation captures the semantic meaning and relationships between words in a continuous vector space. The dense vector similarity search method is configured to determine the similarity between dense vectors to be measured using mathematical metrics. These mathematical metrics quantify how close or similar two dense vectors are in the high-dimensional space. The system 102 employs the dense vector similarity search method to retrieve relevant vocabulary information. For example, when the user inputs the one or more search queries, the system 102 converts the query into a second set of dense vectors and searches the vector database for words within the first set of dense vectors.

[0053]In an exemplary embodiment, the query encoding subsystem 210 is configured to provide the second set of dense vectors by converting received the one or more search queries of the users into the fixed-dimensional vectors, thereby facilitating the efficient retrieval of the associated first set of dense vectors from the vector database. This encoding process involves the use of transformer-based encoder blocks. The transformer-based encoder blocks are configured to convert the received one or more search queries into the fixed-dimensional vectors. The transformer-based encoder blocks are known for their effectiveness in capturing complex linguistic patterns and relationships within the one or more search queries.

[0054]By employing the transformer-based encoder blocks, the query encoding subsystem 210 enhances the ability of the system 102 to understand and process the semantics of the one or more search queries effectively. The resulting fixed-dimensional vectors serve as a structured and informative representation of the one or more search queries, thereby enabling seamless matching and retrieval of the associated first set of dense vectors from the vector database. This approach ensures that the system 102 may retrieve the relevant vocabulary information with precision, contributing to the overall accuracy and responsiveness of the system 102.

[0055]In an exemplary embodiment, the response generation subsystem 212 plays a critical role in generating the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through the adaptive learning. The response generation subsystem 212 is adept at leveraging the second set of dense vectors, which represents the one or more search queries converted into the fixed-dimensional vectors, for the dynamic retrieval-augmented generation during the learning stage. Moreover, the response generation subsystem 212 is adaptive and dynamic, capable of adjusting to user preferences and learning history. This dynamic adaptation enhances the user experience by delivering the one or more responses that align with the user's one or more search queries and learning trajectory, thereby fostering a more effective and personalised language learning environment.

[0056]Furthermore, the response generation subsystem 212 is tailored to prioritise the generation of the one or more responses based on a user feedback collected through a user interface subsystem 214. This strategic integration allows the system 102 to adapt and refine the one or more responses according to the user preferences and effectiveness. The user interface subsystem 214 complements this process by providing a visual representation of a word cloud, showcasing words actively employed by the user. This visual aid not only aids users track their language usage patterns but also provides access to the history of usage words, thereby contributing to a more personalised and informed language learning experience. The integration of the user feedback ensures that the response generation subsystem 212 delivers the one or more responses that align with the learning needs and preferences of the user, fostering an interactive and effective learning environment.

[0057]The response generation subsystem 212 is configured with Large Language Models (LLMs) to generate the one or more responses. The Large Language Models (LLMs) comprise one of a: Mistral 7B, LlaMA 2, BLOOM, Falcon, Generative Pre-trained Transformer (GPT) models, Bidirectional Encoder Representations from Transformers (BERT), eXtreme Learning with Transformers (XLNet), Text-To-Text Transfer Transformer (T5), and Robustly optimised BERT approach (RoBERTa). The GPT models are configured to enhance the quality and contextuality of the generated one or more responses. Particularly, a GPT-4 language model generates substantial improvements in natural language understanding and generation capabilities. Leveraging the advancements of the GPT-4, the system 102 achieves more nuanced and contextually relevant responses, aligning with the user's evolving learning needs. Furthermore, the response generation subsystem 212 is tailored to prioritize the generation of a top three responses associated with the one or more responses based on user feedback collected through a user interface subsystem 214.

[0058]In an exemplary embodiment, the system 102 further incorporates a reinforcement learning subsystem 216. The reinforcement learning subsystem 216 is designed to continually train and refine the response generation subsystem 212 by leveraging the user feedback collected through the user interface subsystem 214. This reinforcement learning mechanism allows the system 102 to learn from user interactions, adapting and optimising the one or more response generation strategies over time.

[0059]In an exemplary embodiment, the coordinated operation of the query receiving subsystem 208, the query encoding subsystem 210, and the response generation subsystem 212 collectively construct a Retrieval-Augmented Generation (RAG) model. The orchestrated RAG model is responsible for retrieving relevant words, their meanings, example sentences, and the like, based on the one or more search queries of the users by employing the dense vector similarity search method.

[0060]In an exemplary embodiment, the system 102 is configured with a lexical relations graph generation subsystem 218. The lexical relations graph generation subsystem 218 is configured to systematically analyse the one or more search queries to retrieve the first set of dense vectors with compiled information. For every word in the first set of dense vectors, the lexical relations graph generation subsystem 218 is programmed to retrieve and compile the vocabulary information such as, but not limited to, synonyms, antonyms, words in similar contexts of the second set of dense vectors, and instances of usage in different contexts. This vocabulary information is then organized into a comprehensive knowledge graph. In this comprehensive knowledge graph, each word serves as a node, and the various linguistic relationships, such as synonyms, antonyms, and words in similar contexts, are represented as edges. For instance, the "synonym" relationship edge connects the synonym nodes to the corresponding word nodes, establishing a transparent and interconnected representation of the linguistic relationships associated with each word in the vocabulary. This lexical relations graph generation subsystem 218 enhances the depth of understanding and connectivity within the linguistic knowledge database, facilitating a more nuanced exploration of the language and its semantic intricacies. Additionally, the knowledge graphs are used to fine-tune the LLMs in the RAG module. This integration enhances the capability of the system 102 to generate more contextually appropriate content.

[0061]FIG. 3 refers to an exemplary flow chart depicting a retrieval-augmented generation method 300 for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure.

[0062]According to an exemplary embodiment of the present disclosure, the method 300 (hereinafter referred to as the method 300) for vocabulary acquisition through adaptive learning is disclosed. At step 302, the method 300 includes the resource data-obtaining subsystem, which acquires the learning resource data to store in the database. This initial step establishes a foundation for effective learning resource management. Subsequently, at step 304, the method 300 includes the data processing subsystem to retrieve the vocabulary information from the obtained learning resource data. The data processing subsystem organises the extracted vocabulary information into the first set of dense vectors to store in the vector database.

[0063]At step 306, the method 300 includes the query receiving subsystem, which receives the one or more search queries from a user. The one or more search queries are intricately linked to the user profile to which the one or more responses are required to be generated. Advancing further, at step 308, the method 300 includes the query encoding subsystem, which obtains the second set of dense vectors by converting the received one or more search queries into the fixed-dimensional vectors, thereby retrieving the associated first set of dense vectors from the vector database.

[0064]Concluding at step 310, the method 300 includes the response generation subsystem, generating the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through adaptive learning. The generation of the one or more responses is based on the user feedback collected through the user interface subsystem.

[0065]FIG. 4 refers to an exemplary user interface 400 depicting the retrieval-augmented generation system for vocabulary acquisition through adaptive learning, in accordance with an embodiment of the present disclosure.

[0066]The user interface 400 of the system 102 serves as a comprehensive platform for user interaction. The users initiate the process by uploading the learning resource data through the "Choose File" section, thereby providing to the system 102 with the learning resources data. Additionally, the users are configured to input the one or more search queries in “Input Text” based on their specific requirements. Upon clicking on the "Generate Response", the system 102 employs the response generation subsystem 212 to dynamically generate the one or more responses tailored to the one or more search queries. The generated one or more responses may employ the words from the list of words associated with the learning resource data. Furthermore, the system 102 presents a feature that generates the word cloud encountered in the last week. Upon clicking the words in the word cloud, the users may view the history of usage of the words, thereby providing the users with a convenient overview of recently encountered vocabulary.

[0067]FIG. 5 refers to an exemplary snapshot 500 of a third-party application depicting configured with the retrieval-augmented generation system 102 for vocabulary acquisition, in accordance with an embodiment of the present disclosure.

[0068]The dynamic retrieval-augmented generation functionality of the system 102 is suitable for integration as a plugin within third-party applications. This seamless integration enables the incorporation of advanced vocabulary learning capabilities into diverse software environments, thereby enhancing the linguistic aspects of the third-party applications. The adaptability of the system 102 allows for a flexible and efficient augmentation of vocabulary learning features, contributing to an enriched user experience within the diverse software environments. The third-party applications such as Facebook Messenger®, WhatsApp®, and any other communication channel thereof. This transformative approach aligns with the concept of enhancing linguistic interactions not only through a dedicated website but also by providing real-time response suggestions within popular communication applications. In an alternative exemplary embodiment, the dynamic retrieval-augmented generation functionality of the system 102 extends to offer real-time suggestions while users are typing within any digital platform. This digital platform includes, but is not limited to, search engines, word-processing applications, and similar environments. The system intelligently analyses the context of the ongoing user input and suggests one or more responses that align with the extracted vocabulary from the learning resource data. By seamlessly integrating into various digital platforms, users receive instant language enhancement suggestions as they compose text, fostering an efficient and continuous learning experience.

[0069]Numerous advantages of the present disclosure may be apparent from the discussion above. In accordance with the present disclosure, the system for providing dynamic retrieval-augmented generation of vocabulary is provided. The system provides a personalised experience to enhance the learning experience of the users. The system accelerates the learning of the users who are preparing for competitive exams including GRE, GMAT, CAT, and the like. The users are configured to employ the plugins to respond to the one or more search queries by employing the generated one or more responses on the spot within the third-party applications. The system assists the users in learning the set of words consciously and effortlessly. The system is able to leverage pre-trained large language models (LLMs), such as GPT-4, to incorporate out-of-vocabulary words in exceptional scenarios where the RAG pipeline encounters difficulty in finding suitable terms within the learning resource data. This capability allows the system to efficiently generate responses using words not present in the documents fed to the RAG pipeline, drawing on the extensive knowledge acquired during pre-training on vast corpora. The potential of LLMs, including GPT-4, Llama2, Falcon and the like, to seamlessly handle out-of-vocabulary scenarios enhances the adaptability and effectiveness of the system in diverse language contexts. The system also displays the words employed in the last week along with the usage history of the words to enhance the learning experience of the users.

[0070]While specific language has been used to describe the invention, any limitations arising on account of the same are not intended. As would be apparent to a person skilled in the art, various working modifications may be made to the method in order to implement the inventive concept as taught herein.

[0071]The figures and the foregoing description give examples of embodiments. Those skilled in the art will appreciate that one or more of the described elements may well be combined into a single functional element. Alternatively, certain elements may be split into multiple functional elements. Elements from one embodiment may be added to another embodiment. For example, order of processes described herein may be changed and are not limited to the manner described herein. Moreover, the actions of any flow diagram need not be implemented in the order shown; nor do all of the acts need to be necessarily performed. Also, those acts that are not dependent on other acts may be performed in parallel with the other acts. The scope of embodiments is by no means limited by these specific examples.

I/We claim:

1. A retrieval-augmented generation (RAG) system (102) for vocabulary acquisition through adaptive learning, comprising:

   a server (116) configured with one or more hardware processors (110),

   a data repository unit (112) operatively coupled to the one or more hardware processors (110), wherein the data repository unit (112) comprises a plurality of subsystems (114) in form of programmable instructions executable by the one or more hardware processors (110), wherein the plurality of subsystems (114) comprises:

      a resource data-obtaining subsystem (204) configured to obtain learning resource data for storing in a database (104);

      a data processing subsystem (206) configured to extract vocabulary information from the obtained learning resource data for organising the extracted vocabulary information into a first set of dense vectors to store in a vector database;

      a query receiving subsystem (208) configured to receive one or more search queries from a user associated with a user profile for which one or more responses are required to be generated;

      a query encoding subsystem (210) configured to provide a second set of dense vectors by converting the received one or more search queries into fixed-dimensional vectors for retrieving the associated first set of dense vectors from the vector database; and

      a response generation subsystem (212) configured to generate the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through the adaptive learning.

2. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the learning resource data comprises at least one of: flashcards, books, Portable Document Format (PDF) documents with primary and secondary meanings and examples, and documents with a list of vocabulary words.

3. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the query receiving subsystem (208) is configured with a retrieval-augmented generation (RAG) module to retrieve the associated first set of dense vectors from the vector database information based on a dense vector similarity search method,

the retrieval-augmented generation (RAG) module is configured with one or more libraries comprises at least one of a: Lama-Index, LangChain, and Haystack.

4. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the query encoding subsystem (210) is configured with transformer-based encoder blocks configured to convert the received one or more search queries into the second set of dense vectors for retrieval of the first set of dense vectors.

5. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the response generation subsystem (212) is configured with Large Language Models (LLMs) to generate the one or more responses,

the Large Language Models (LLMs) comprises one of a: Mistral 7B, LlaMA 2, BLOOM, Falcon, Generative Pre-trained Transformer (GPT) models, Bidirectional Encoder Representations from Transformers (BERT), eXtreme Learning with Transformers (XLNet), Text-To-Text Transfer Transformer (T5), and Robustly optimised BERT approach (RoBERTa).

6. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the response generation subsystem (212) is configured to adapt to user preferences and learning history dynamically, providing personalised one or

more responses associated with the one or more search queries based on the user profile.

7. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the response generation subsystem (212) is configured to prioritise the generated one or more responses based on a user feedback collected through a user interface subsystem (214).

8. The retrieval-augmented generation (RAG) system (102) as claimed in claim 7, wherein the user interface subsystem (214) is configured to present a visual representation of a word cloud displaying words used by the user as the one or more responses and enabling users to access the history of usage for words.

9. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the system (102) comprises a reinforcement learning subsystem (216),

   the reinforcement learning subsystem (216) configured to continuously train and refine the response generation subsystem (212) based on the user feedback received through the user interface subsystem (214).

10. The retrieval-augmented generation (RAG) system (102) as claimed in claim 1, wherein the retrieval-augmented generation (RAG) system (102) comprises a lexical relations graph generation subsystem (218),

   the lexical relations graph generation subsystem (218) is configured to analyse the one or more search queries to retrieve the first set of dense vectors with compiled information,

   the compiled information comprises at least one of: synonyms, antonyms, words in analogous contexts, and instances of usage in discrete contexts.

11. A retrieval-augmented generation (RAG) method (300) for vocabulary acquisition through adaptive learning, comprising:

   obtaining, by a resource data-obtaining subsystem (204), learning resource data to store in a database (104);

extracting, by a data processing subsystem (206), vocabulary information from the obtained learning resource data to organise the extracted vocabulary information into a first set of dense vectors to store in a vector database;

receiving, by a query receiving subsystem (208), one or more search queries from a user associated with a user profile for which one or more responses are required to be generated;

providing, by a query encoding subsystem (210), a second set of dense vectors by converting the received one or more search queries into fixed-dimensional vectors to retrieve the associated first set of dense vectors from the vector database; and

generating, by a response generation subsystem (212), the one or more responses by comparing the second set of dense vectors with the first set of dense vectors for vocabulary acquisition through the adaptive learning.

Dated this 25th day of January, 2024

VIDYA BHASKAR SINGH NANDIYAL
Patent Agent No. 2912
IPExcel Private Limited
AGENT FOR APPLICANTS

ABSTRACT

RETRIEVAL-AUGMENTED GENERATION SYSTEM FOR VOCABULARY ACQUISITION THROUGH ADAPTIVE LEARNING AND METHOD THEREOF

The present invention discloses a retrieval-augmented generation system for vocabulary acquisition through adaptive learning and method thereof. The system (102) comprises a resource data-obtaining subsystem (204), a data processing subsystem (206), a query receiving subsystem (208), a query encoding subsystem (210), and a response generation subsystem (212). The coordinated operation of the query receiving subsystem (208), the query encoding subsystem (210), and the response generation subsystem (212) collectively construct a Retrieval-Augmented Generation (RAG) module. The orchestrated RAG module is responsible for retrieving relevant words, their meanings, and example sentences based on one or more search queries by employing a dense vector similarity search method. The system (102) provides personalised responses to enhance the learning experience of users. Furthermore, the system (102) is tailored to prioritize the generation of one or more responses based on a user feedback collected through a user interface subsystem (214).

FIG. 2