पेटेंट कार्यालय
शासकीय जर्नल

OFFICIAL JOURNAL
OF
THE PATENT OFFICE

निर्गमन सं. 29/2022	शुक्रवार	दिनांक: 22/07/2022
ISSUE NO. 29/2022	**FRIDAY**	**DATE: 22/07/2022**

पेटेंट कार्यालय का एक प्रकाशन
PUBLICATION OF THE PATENT OFFICE

(12) PATENT APPLICATION PUBLICATION
(19) INDIA
(22) Date of filing of Application :01/07/2022

(21) Application No.202241038011 A
(43) Publication Date : 22/07/2022

(54) Title of the invention : METHOD FOR UTILIZING MACHINE LEARNING (ML) MODELS FOR AUTOMATICALLY CORRECTING ERRORS IN GLOBAL HISTORICAL DATA

(51) International classification	:G06N0020000000, G16H0010600000, H04L0012240000, G06K0009620000, G06Q0030020000
(86) International Application No Filing Date	:PCT// :01/01/1900
(87) International Publication No	: NA
(61) Patent of Addition to Application Number Filing Date	:NA :NA
(62) Divisional to Application Number Filing Date	:NA :NA

(71)**Name of Applicant :**
1)DR. KOLIPAKA SRINIVAS (Head and BoS Chairperson)
Address of Applicant :DEPARTMENT OF HISTORY GOVERNMENT DEGREE COLLEGE(A), SIDDIPET, SIDDIPET DISTRICT, TELANGANA STATE, INDIA-502 103 e-mail: drsrinu75@gmail.com Phone+91- 9866611947 Siddipet ----------- -----------
2)KOLIPAKA AKSHAYA SREEJA (Student)
3)D. PAVAN SRIVASTA (Student)
4)DR. D SURESH BABU (Head & BoS Chairperson)
Name of Applicant : NA
Address of Applicant : NA
(72)**Name of Inventor :**
1)DR. KOLIPAKA SRINIVAS (Head and BoS Chairperson)
Address of Applicant :DEPARTMENT OF HISTORY GOVERNMENT DEGREE COLLEGE(A), SIDDIPET, SIDDIPET DISTRICT, TELANGANA STATE, INDIA-502 103 e-mail: drsrinu75@gmail.com Phone+91- 9866611947 Siddipet ----------- -----------
2)KOLIPAKA AKSHAYA SREEJA (Student)
Address of Applicant :B. TECH III YEAR MECHANICAL ENGINEERING NATIONAL INSTITUTE OF TECHNOLOGY (NIT) WARANGAL, TELANGANA STATE, INDIA-506004 e-mail: Phone: +916301748945 Siddipet ----------- -----------
3)D. PAVAN SRIVASTA (Student)
Address of Applicant :B. TECH (EEE) III YEAR VASAVI COLLEGE OF ENGINEERING(AUTONOMOUS) HYDERABAD, TELANGANA STATE, INDIA-507001. e-mail: Phone+918979108983 Hyderabad ----------- -----------
4)DR. D SURESH BABU (Head & BoS Chairperson)
Address of Applicant :DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001. e-mail: Phone+919849664111 Khammam ----------- -----------

(57) Abstract :
METHOD FOR UTILIZING MACHINE LEARNING (ML) MODELS FOR AUTOMATICALLY CORRECTING ERRORS IN GLOBAL HISTORICAL DATA ABSTRACT The present invention provides an approach for automatically correcting errors in global historical data. The invention utilizes one or more Machine Learning (ML) models that are trained using global history data, for identifying absurd, errors, and mistakes in the data related o history of plurality of kingdoms and countries, and automatically attempts to rectify them by utilizing the trained data. The ML models are trained using multiple history databases that are available online via various digital libraries and online historical sources. Upon identifying the errors in the existing global history data in a database, the ML model provides an alert to the administrator of the database regarding the error and automatically rectifies the error with correct datasets or information. Upon correcting the information, a second alert or notification is sent to the administrator of the database regarding the corrected details for confirmation.

No. of Pages : 22 No. of Claims : 5

FORM 2

THE PATENT ACT, 1970

(39 OF 1970)

&

THE PATENT RULES, 2003

COMPLETE SPECIFICATION

[SEE SECTION 10 AND RULE 13]

TITLE: METHOD FOR UTILIZING MACHINE LEARNING (ML) MODELS FOR AUTOMATICALLY CORRECTING ERRORS IN GLOBAL HISTORICAL DATA	
APPLICANTS:	
DR. KOLIPAKA SRINIVAS	DEPARTMENT OF HISTORY GOVERNMENT DEGREE COLLEGE(A), SIDDIPET, SIDDIPET DISTRICT, TELANGANA STATE, INDIA-502 103
KOLIPAKA AKSHAYA SREEJA	MECHANICAL ENGINEERING NATIONAL INSTITUTE OF TECHNOLOGY (NIT) WARANGAL, TELANGANA STATE, INDIA-506004
D. PAVAN SRIVASTA	VASAVI COLLEGE OF ENGINEERING(AUTONOMOUS) HYDERABAD, TELANGANA STATE, INDIA-507001.
DR. D SURESH BABU	DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS SR & BGNR GOVT. ARTS & SCIENCE COLLEGE(A), KHAMMAM, KHAMMAM DISTRICT, TELANGANA STATE, INDIA-507001.

The following specification particularly describes the invention and the manner in which it is to be performed

METHOD FOR UTILIZING MACHINE LEARNING (ML) MODELS FOR AUTOMATICALLY CORRECTING ERRORS IN GLOBAL HISTORICAL DATA

FIELD OF THE INVENTION

The present invention generally relates to historical data. More particularly, the present invention relates to utilizing machine learning (ML) models for automatically correcting errors in global historical data.

BACKGROUND OF THE INVENTION

Linguistics is the scientific study of language. One aspect of linguistics is the application of computer science to human natural languages, such as english. Computer applications for linguistics are increasing due to the tremendous increase in processor speed and memory capacity. For example, computer-enabled analysis of speech utterances facilitates many applications, such as automated agents, that can answer questions from users. It is becoming increasingly popular to use "chat robots" (chatbots) and agents (agents) to answer questions, facilitate discussions, manage conversations, and provide social promotions (social promotions). To meet this need, a wide range of techniques have been developed that include synthetic semantics. Such techniques may support automated brokering in the case of simple, short queries and responses.

However, because existing solutions fail to match answers to questions due to insufficient lexical analysis (rhelogical analysis), such solutions fail to answer questions, perform dialog management, provide suggestions, or implement a "chat bot" system with rich speech-related information. More specifically, statistical-based solutions fail to separate the task of determining topics from sentences from resolving the lexical agreement (rhetorial agreement) between sentences and answers. Statistical-based solutions either do not consider the lexicographic structure of the problem and response at all, or attempt to resolve both topic and lexicographic consistency without properly resolving the lexicographic consistency. Without sufficient lexicographic analysis, a question that may have any lexicographic structure cannot be matched with an appropriate answer that may also have any lexicographic structure.

BRIEF DESCRIPTION OF FIGURES

FIG. 1 illustrate an exemplary classification environment, in accordance with an embodiment of the invention.

FIG. 2 illustrates an example of a tree representation model in accordance with an embodiment of the invention.

FIG. 3 illustrates an example of a discourse tree of the global historical data in accordance with an exemplary embodiment of the invention.

FIG. 4 illustrates schemas, in accordance with an embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

In general, the system, apparatus and method of the present invention are related to computing a lexical interrelationship(s) between one or more sentences of global historical data. In an example, a computer-implemented method accesses a sentence that includes a segment for identifying absurd, errors, and mistakes in the data related o history of plurality of kingdoms and countries, and automatically attempts to rectify them by utilizing the trained data.

At least one segment includes a verb and a word. Each word can serve as a role within multiple words within a segment. Each segment is a basic speech unit. The method generates a speech tree that represents the lexical interrelationships between sentence fragments. The utterance tree includes nodes including non-terminating nodes and terminating nodes, each non-terminating node representing a thesaurus interrelationship between two sentence fragments, and each terminating node in the nodes of the utterance tree is associated with one of the sentence fragments. The method matches each segment with a verb signature, thereby creating a communication utterance tree.

Matching includes accessing a plurality of verb signatures. Each verb signature includes a sequence of verbs and a subject role (textual role) of the segment. The method determines, for each verb signature in the verb signatures, a number of topic roles in the corresponding signature that match the roles of the words in the segment. The method selects the particular verb signature from the verb signatures based on the particular verb signature including the highest number of matches. The method associates a particular verb signature with the fragment.

Aspects disclosed herein provide technological improvements to the field of computer-implemented linguistics. More specifically, aspects described herein represent a thesaurus interrelationship between one or more sentences in a communication utterance tree.

The "communication speech tree" or "CDT" includes a speech tree supplemented with communication actions. Communicative actions are collaborative actions taken by individuals based on mutual negotiations and demonstrations.

Other aspects disclosed herein enable an improved automated agent or chat robot that can answer questions received from a user by using a communication utterance tree. By using a communication-utterance tree, aspects overcome the limitations of previous systems, which generally fail to separate the task of determining topics from sentences from resolving thesaurus concordance between sentences and answers.

In an example, a thesaurus classification application executing on a computing device receives a question from a user. The thesaurus classification application generates a communication utterance tree for the question. The communication speech tree is a kind of speech tree, or a speech tree containing communication actions. The thesaurus classification application accesses a database of potential answers to the question. Using the predictive model, the thesaurus correspondence application determines a level of complementarity between the question and each potential answer. In response to determining that the level of complementarity is above the threshold, the thesaurus consistency classifier provides an answer to the user, e.g., via a display device.

Technical advantages of some aspects include improved autonomous agents (such as chat robots) and improved search engine performance over traditional statistical-based approaches. Conventional statistical keyword-based methods either (i) fail to solve the subject of the problem, or (ii) fail to solve the thesaurus consistency between the problem and the answer. Thus, existing autonomous agent solutions are only capable of scripting or limited response to user problems. Such a solution cannot determine whether the answer is fully responsive to the question.

For example, aspects described herein use a communication utterance tree. The communication speech tree combines the retrieval information with the communication action. By integrating tags that recognize communication actions, learning of the communication utterance tree can occur over a richer feature set rather than just the grammar and the lexical relationships (relationships) of the basic utterance unit (EDU). With this feature set, additional techniques such as classification can be used to determine a level of thesaurus correspondence between questions and answers or request-response pairs, thereby enabling an improved automated agent. By doing so, the computing system implements autonomous agents that can intelligently answer questions and other messages.

FIG. 1 illustrates an exemplary lexicographic classification environment in accordance with an aspect. FIG. 1 depicts a thesaurus classification computing device 101, an input question 130, an output question 150, a data network 104, a server 160, and a mobile device 170. The thesaurus classification computing device 101 includes one or more of a thesaurus classification application 102, an answer database 105, a thesaurus correspondence classifier 120, and training data 125. The thesaurus classification application 102 includes one or more of a question exchange utterance tree 110, an answer exchange utterance tree 110.

The mobile device 170 may be any mobile device, such as a mobile phone, a smartphone, a tablet, a laptop, a smart watch, and so forth. The mobile device 170 communicates with the server 160 or the pruning classification computing device 101 via the data network 104. In this manner, the mobile device 170 may provide the question 171, for example, from the user to the server 160 or the thesaurus computing device 101. In an example, the thesaurus classification computing device 101 determines a suitable answer 172 and provides the answer 172 to the mobile device 170 over the data network 104.

The data network 104 may be any public or private network, wired or wireless network, wide area network, local area network, or the internet.

In an example, the thesaurus classification application 102 answers questions received via chat. More specifically, the thesaurus classification application 102 receives an input question 130, which may be a single question or a stream of questions such as chat. The thesaurus classification application 102 creates a question exchange utterance tree 110 and selects one or more candidate answers. The answers may be obtained from an existing database, such as answer database 105, or from a server 160 communicating over data network 104. The server 160 may be a public or private internet server, such as a public database of user questions and answers.

The thesaurus classification application 102 determines the most appropriate answer from the candidate answers. As further explained herein, different approaches may be used. In an aspect, the thesaurus classification application 102 may create a candidate answer communication utterance tree for each candidate answer and compare the question communication utterance tree 110 to each candidate utterance tree. The thesaurus classification application 102 identifies the best match between the question alternating utterance tree and the candidate answer alternating utterance tree. The thesaurus classification application 102 then accesses or queries a database for text from the best communicating utterance tree. The thesaurus classification application 102 then sends the text associated with the second communication utterance tree to the mobile device.

In another aspect, the thesaurus classification application 102 creates an answer communication utterance tree 111 for each candidate answer. The thesaurus classification application 102 then creates a question-answer pair comprising the question 130 and the candidate answer for each candidate answer.

The thesaurus classification application 102 provides the question-answer pairs to a predictive model, such as a thesaurus consistency classifier 120. Using the trained thesaurus consistency classifier 120, the thesaurus classification application 102 determines whether the question-answer pair is above a threshold level of matching, e.g., indicating whether the answer solves the question. If the answer does not solve the question, the thesaurus classification application 102 continues to analyze other pairs including the question and different answers until a suitable answer is found. By using the communication utterance tree, the correspondence of the revisions and the communication actions between the question and the answer can be accurately modeled.

The thesaurus classification application 102 provides the answer as an output answer 150. For example, as shown in FIG. 1, an agent implemented by the thesaurus classification application 102 provides the text "here is song personal list of songs" in response to a chat history involving two users discussing a singing of a song.

Structural theory of revising and lexical tree

Linguistics is the scientific study of language. For example, linguistics may include the structure (syntax) of a sentence, e.g., subject-verb-object; meaning of sentences (semantics), e.g., dog versus dog; and what the speaker does in the conversation, i.e., an analysis of utterances or a linguistic analysis outside of the sentence.

The theoretical basis of utterances, the modified structure Theory (RST), can be attributed to Mann, William and Thompson, Sandra, "Rheological structure Theory: A Theory of Text organization", Text-Interdisciplinery Journal for the student of science, 8(3): 243-. RST helps to enable the analysis of utterances, similar to how the syntax and semantics of programming language theory help to enable modern software compilers. More specifically, RST places structural blocks on at least two levels, such as a first level of nuclearity and lexicography, and a second level of structure or schema. The utterance parser or other computer software can parse the text into an utterance tree.

The thesaurus structure theory models the logical organization of text, which is the structure adopted by the author and depends on the relationships between parts of the text. RST simulates text coherence by forming a hierarchical connected structure of text through a speech tree. The pruning relationships are divided into coordination (coordination) and subordinate (subordinate) categories; these relationships are maintained across two or more text sections, thus achieving consistency. These text sections are called basic speech units (EDUs). Clauses in the sentence and sentences in the text are logically connected by the author. The meaning of a given sentence is related to the meaning of preceding and succeeding sentences. This logical relationship between clauses is called a coherent structure of text. RST is one of the most popular utterance theories, and is based on a tree-like utterance structure, the utterance tree (DT). The leaves of DT correspond to EDUs, i.e. consecutive atomic text segments. Adjacent EDUs are connected by a coherent relationship (e.g., cause, order) to form a higher-level speech unit. These units are then also constrained by this relationship linkage. EDUs linked by relationships are then distinguished based on their relative importance: the core is the core part of the relationship and the satellites are the peripheral parts. As discussed, to determine accurate request-response pairs, both topic and thesaurus consistency are analyzed. When a speaker answers a question, such as a phrase or sentence, the speaker's answer should be directed to the subject of the question. In the case of a question posed implicitly via the seed text of a message, it is desirable to not only maintain the topic but also an appropriate answer that matches the generalized cognitive state of the seed.

Relations between revisions and expressions

As discussed, aspects described herein use a communication utterance tree. The pruning relations may be described in different ways. For example, Mann and Thompson describe 23 possible relationships. Mann, William and Thompson, Sandra (1987) ("Mann and Thompson"). A Theory of textile Theory. Other numbers of relationships are also possible.

Fig. 2 depicts an example of a utterance tree according to an aspect. Fig. 2 includes a speech tree 200. The utterance tree includes a text section 201, a text section 202, a text section 203, a relationship 210, and a relationship 228. The numbers in fig. 2 correspond to three text sections. Fig. 3 corresponds to the following example text with three text sections numbered 1, 2, 3:

honolulu, Hawaii will be site of the 2017Conference on Hawaiian History (Hawaii fire Nuoluu will become the place of the 2017 Hawaii History)

It is expected that 200historians from the U.S. and Asia will participate in the study (200 historians from the United states and Asia are expected to be present)

The conference with The second connected with The method of The connected to Hawaii (The conference will focus on how The people in Borisia navigate to Hawaii)

For example, relationship 210 (or illustration) describes the interrelationship between text section 201 and text section 202. Relationship 228 depicts the interrelationship (elaboration) between text sections 203 and 204. As depicted, text sections 202 and 203 further illustrate text section 201. In the example above, text section 1 is the core given the goal of informing readers of the meeting. Text sections 2 and 3 provide more detailed information about the conference. In fig. 2, horizontal numbers (e.g., 1-3, 1, 2, 3) cover sections of text (possibly made up of further sections); the vertical lines indicate one or more cores; and the curves represent the lexicographic relations (exposition) and the direction of the arrow points from the satellite to the core. If the text segment is used only as a satellite and not as a core, deleting the satellite will still leave coherent text. If the core is removed from FIG. 2, then text sections 2 and 3 will be difficult to understand.

Fig. 3 depicts another example of a utterance tree in accordance with an aspect. FIG. 3 includes components 301 and 302, text segment 305 and 307, relationship 310 and relationship 328. Relationship 310 depicts the interrelationship 310, enable, between components 306 and 305 and 307 and 305. Fig. 3 relates to the following text sections:

the new Tech Report extract now in The journal area of The library near The reduced dictionary

Please sign your name by means of means that is said you would of the same name as in search (Please sign your name in a way you are interested in seeing.)

Last day for sign-ups is 31May (the Last day signed is 31 months.)

As can be seen, relationship 328 depicts the correlation between entities 307 and 306 (the correlation is enabled). Fig. 3 illustrates that although the cores may be nested, there is only one most core text section.

Constructing a speech tree

The utterance tree can be generated using different methods. A simple example of a method of constructing DTs from the bottom up is:

(1) dividing the utterance text into units by:

(a) the cell size may be different depending on the target of the analysis

(b) Typically, the units are clauses

(2) Check every cell and its neighbors? for relationships between them

(3) If so, the relationship is marked.

(4) If not, the cell may be located at the boundary of a higher level relationship. Look at the relationships maintained between larger units (sections).

(5) And continues until all elements in the text are considered.

Mann and Thompson also describe a second level of building block structure, called schema application. In RST, the pruning relation is not directly mapped to the text; they are assembled onto structures called schema applications, and these structures are in turn assembled into text. Schema applications are derived from a simpler structure called a schema (as shown in fig. 4). Each mode indicates how a particular unit of text is broken down into other smaller units of text. The prune structure tree or DT is a hierarchical system of schema applications. The schema application links multiple consecutive text sections and creates a complex text section that can in turn be linked by a higher level schema application. RST assertion, the structure of each consecutive utterance can be described by a single prune structure tree whose top mode creates a segment that covers the entire utterance.

FIG. 4 depicts an illustrative schema in accordance with an aspect. FIG. 4 shows that the federation mode is a list of items consisting of cores without satellites. FIG. 4 depicts mode 401-406. Schema 401 depicts the environmental relationship between text sections 410 and 428. Schema 402 depicts the sequence relationship between text sections 420 and 421 and the sequence relationship between text sections 421 and 422. Schema 403 depicts the contrasting relationship between text sections 430 and 431. Schema 404 depicts the joint interrelationship between text sections 440 and 441. The schema 405 depicts the motivational interrelationship between 450 and 451, and the enablement interrelationship between 452 and 451. Schema 406 depicts the joint interrelationship between text sections 460 and 462. An example of a federation mode for the following three text sections is shown in FIG. 4:

skies will be partial sunny in the New York metropolitan area today (the sky in New York City district will be partially sunny.)

It with more than humid, with temperature in the middle of 80's (the weather is more humid)

Tonight with be most mortar close, with the low temperature between 65 and 70. (most cloudy today, low temperatures between 65 and 70.)

Although fig. 2-4 depict some graphical representations of the utterance tree, other representations are possible.

Automatic utterance segmentation can be performed in different ways. For example, given a sentence, the segmentation model identifies the boundaries of the composite base utterance unit by predicting whether boundaries should be inserted before each particular symbol (token) in the sentence. For example, one framework considers each symbol in a sentence sequentially and independently. In this framework, the segmentation model scans the sentence symbol-by-symbol and uses a binary classifier (such as a support vector machine or logistic regression) to predict whether it is appropriate to insert a boundary before the symbol being examined. In another example, the task is a sequential tagging puzzle. Once the text is segmented into basic speech units, sentence-level speech parsing can be performed to construct a speech tree. Machine learning techniques may be used.

Further, the above two utterance parsers, i.e., corenlpp processor and fastnlp pprocessor, perform syntax parsing using Natural Language Processing (NLP). For example, Stanford Core NLP gives the basic form of words, parts of their speech, whether they are the names of companies, people, etc.; standardizing date, time and numerical quantity; tagging the structure of the sentence according to the phrase and syntactic dependency; indicating which noun phrases refer to the same entity. In fact, RST is still the theory that may work in many cases of utterances but may not work in some cases.

The thesaurus classification application 102 may determine whether a given answer or response, such as an answer obtained from the answer database 105 or a public database, is responsive to a given question or request. More specifically, the thesaurus classification application 102 analyzes whether a request and a response pair are correct or incorrect by determining one or both of (i) a correlation or (ii) a thesaurus correspondence between the request and the response. The thesaurus consistency can be analyzed without considering the correlation, which can be processed orthogonally.

The thesaurus classification application 102 may use different methods to determine the similarity between question-answer pairs. For example, the thesaurus classification application 102 may determine a level of similarity between individual questions and individual answers. Alternatively, the thesaurus classification application 102 may determine a measure of similarity between a first pair comprising the question and the answer and a second pair comprising the question and the answer.

Computer-readable storage media 3022 containing the code or portions of code may also include any suitable media known or used in the art, including storage media and communication media such as, but not limited to, volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage and/or transmission of information. This may include tangible, non-transitory, computer-readable storage media, such as RAM, ROM, electrically erasable, programmable ROM (eeprom), flash memory or other memory technology, CD-ROM, Digital Versatile Disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or other tangible, computer-readable media. When designated, this may also include non-tangible, transitory computer-readable media, such as data signals, data transmissions, or any other medium that may be used to transmit the desired information and that may be accessed by the computing system 3000.

By way of example, the computer-readable storage media 3022 may include a hard disk drive that reads from or writes to non-removable, nonvolatile magnetic media, a magnetic disk drive that reads from or writes to a removable, nonvolatile magnetic disk, and removable, nonvolatile optical disks such as CD ROMs, DVDs, and the like Disk or other optical media) to which it is read or written. The computer-readable storage medium 3022 may include, but is not limited to, drives, flash memory cards, Universal Serial Bus (USB) flash drives, Secure Digital (SD) cards, DVD disks, digital audio bands, and the like. The computer-readable storage medium 3022 may also include non-volatile memory based Solid State Drives (SSDs) (such as flash memory based SSDs, enterprise flash drives, solid state ROMs, etc.), volatile memory based SSDs (such as solid state RAM, dynamic RAM, static RAM), DRAM based SSDs, magneto resistive RAM (MRAM) SSDs, and hybrid SSDs that use a combination of DRAM and flash memory based SSDs. The disk drives and their associated computer-readable media may provide non-volatile storage of computer readable instructions, data structures, program modules and other data for computer system 3000.

Communication subsystem 3024 provides an interface to other computer systems and networks. The communication subsystem 3024 serves as an interface for receiving data from other systems and transmitting data from the computer system 3000 to other systems. For example, the communication subsystem 3024 may enable the computer system 3000 to connect to one or more devices via the internet. In some aspects, the communication subsystem 3024 may include Radio Frequency (RF) transceiver components (e.g., using cellular telephone technology, advanced data network technologies such as 3G, 4G, or EDGE (enhanced data rates for global evolution), WiFi (IEEE 802.28 family of standards), or other mobile communication technologies, or any combination thereof), Global Positioning System (GPS) receiver components, and/or other components for accessing wireless voice and/or data networks. In some aspects, the communication subsystem 3024 may provide a wired network connection (e.g., ethernet) in addition to, or in place of, a wireless interface.

In some aspects, the communication subsystem 3024 may also receive input communications in the form of structured and/or unstructured data feeds 3026, event streams 3028, event updates 3030, and the like, on behalf of one or more users who may use the computer system 3000.

As an example, the communication subsystem 3024 may be configured to receive unstructured data feeds 3026 from users of social media networks and/or other communication services in real-time, such as Feeding, Updates, web feeds such as Rich Site Summary (RSS) feeds, and/or real-time updates from one or more third-party information sources.

Further, the communication subsystem 3024 may also be configured to receive data in the form of a continuous data stream, which may include an event stream 3028 and/or event updates 3030 that may be continuous or unbounded in nature with real-time events that do not terminate explicitly. Examples of applications that generate continuous data may include, for example, sensor data applications, financial tickers, network performance measurement tools (e.g., network monitoring and traffic management applications), clickstream analysis tools, automotive traffic monitoring, and so forth.

The communication subsystem 3024 may also be configured to output structured and/or unstructured data feeds 3026, event streams 3028, event updates 3030, and the like to one or more databases, which may be in communication with one or more streaming data source computers coupled to the computer system 3000.

The computer system 3000 may be one of various types, including a hand-portable device (e.g., a cellular phone, Computing tablet, PDA), wearable device (e.g., Google) Head mounted display), a PC, a workstation, a mainframe, a kiosk, a server rack, or any other data processing system.

Due to the ever-changing nature of computers and networks, the description of computer system 3000 depicted in the drawings is intended only as a specific example. Many other configurations are possible with more or fewer components than the system depicted in the figures. For example, custom hardware may also be used and/or particular elements may be implemented in hardware, firmware, software (including applets), or combinations thereof. In addition, connections to other computing devices, such as network input/output devices, may also be employed. Based on the disclosure and teachings provided herein, one of ordinary skill in the art will recognize other ways and/or methods to implement the various aspects.

In the foregoing specification, aspects of the present invention have been described with reference to specific aspects thereof, but those skilled in the art will recognize that the present invention is not limited thereto. Various features and aspects of the above-described invention may be used alone or in combination. Further, aspects may be used in any number of environments and applications beyond those described herein without departing from the broader spirit and scope of the specification. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

I/WE CLAIM:

1. A method for utilizing machine learning (ML) models for automatically correcting errors in global historical data, the method comprising:

    accessing, by one or more processors, global historical data;

    pre-processing, by the one or more processors, the accessed global historical data to scrutinize unwanted data and/or error prone data;

    sending, by the one or more processors, the pre-processed data to one or more data filters to scan the data for identifying correctness of the historical data;

    mapping, by the one or more processors, links to details of various heirs present in digital libraries and thesaurus;

    implementing, one or more ML models to intelligently predict any errors that are existing between the links that are mapped in the aforementioned step;

    sending the identified errors, absurd, and mistakes in the data related o history of plurality of kingdoms and countries to text processing models;

    automatically rectifying errors in global historical data using text processing models.

2. The method for utilizing machine learning (ML) models for automatically correcting errors in global historical data, as claimed in claim 1, wherein a role comprises an entity type or a subject.

3. The method for utilizing machine learning (ML) models for automatically correcting errors in global historical data, as claimed in claim 1, wherein a verb signature includes an ordered list of thematic roles, matching a role of a word in the elementary discourse unit to a respective thematic role.

4. The method for utilizing machine learning (ML) models for automatically correcting errors in global historical data, as claimed in claim 1, wherein the global historical data includes countries, continents, kingdoms, monarchies, etc.

5. The method for utilizing machine learning (ML) models for automatically correcting errors in global historical data, as claimed in claim 1, wherein the global historical data includes an adverb, or a noun.

METHOD FOR UTILIZING MACHINE LEARNING (ML) MODELS FOR AUTOMATICALLY CORRECTING ERRORS IN GLOBAL HISTORICAL DATA

ABSTRACT

The present invention provides an approach for automatically correcting errors in global historical data. The invention utilizes one or more Machine Learning (ML) models that are trained using global history data, for identifying absurd, errors, and mistakes in the data related o history of plurality of kingdoms and countries, and automatically attempts to rectify them by utilizing the trained data. The ML models are trained using multiple history databases that are available online via various digital libraries and online historical sources. Upon identifying the errors in the existing global history data in a database, the ML model provides an alert to the administrator of the database regarding the error and automatically rectifies the error with correct datasets or information. Upon correcting the information, a second alert or notification is sent to the administrator of the database regarding the corrected details for confirmation.

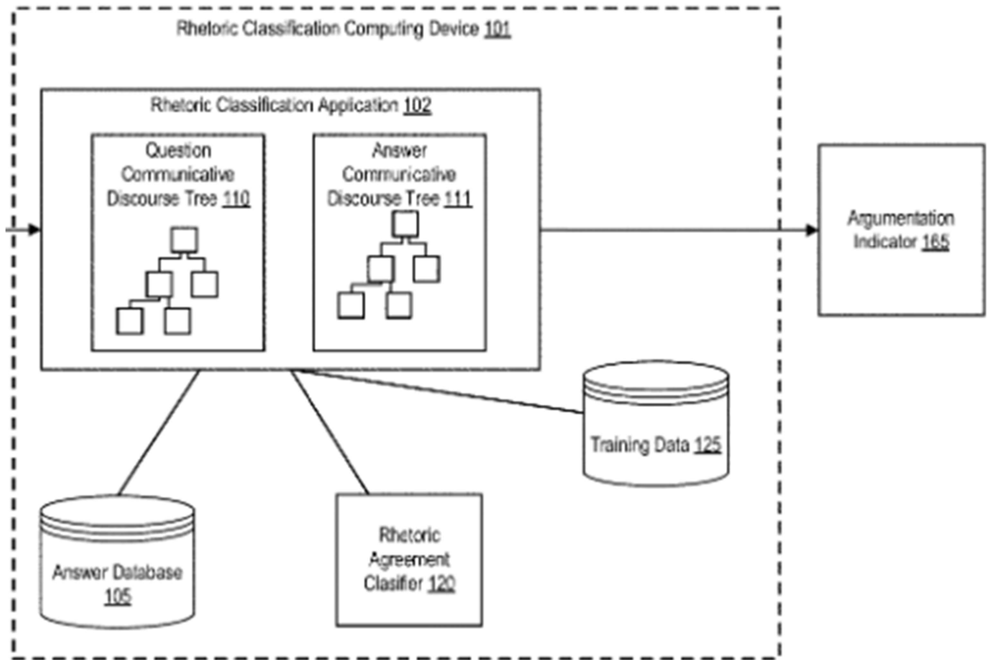
Rhetoric Classification Computing Device 101
Rhetoric Classification Application 102
Question Communicative Discourse Tree 110
Answer Communicative Discourse Tree 111
Argumentation Indicator 165
Training Data 125
Answer Database 105
Rhetoric Agreement Clasifier 120

FIG. 1

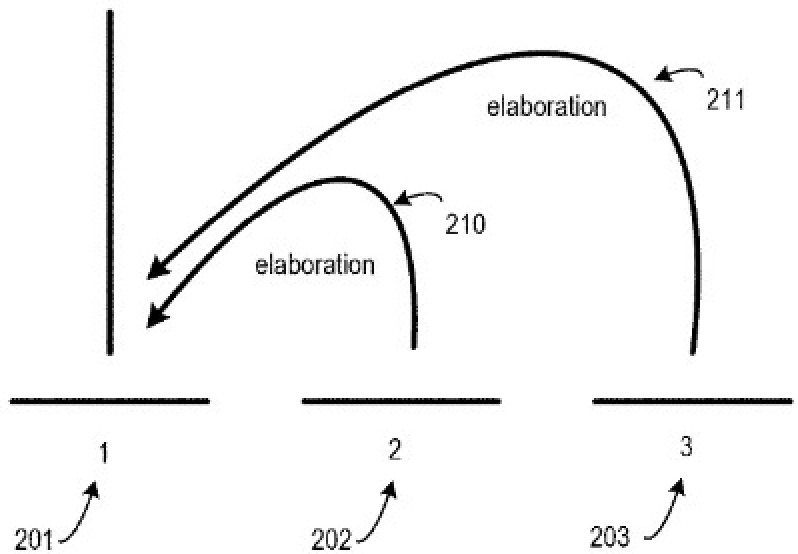

FIG. 2

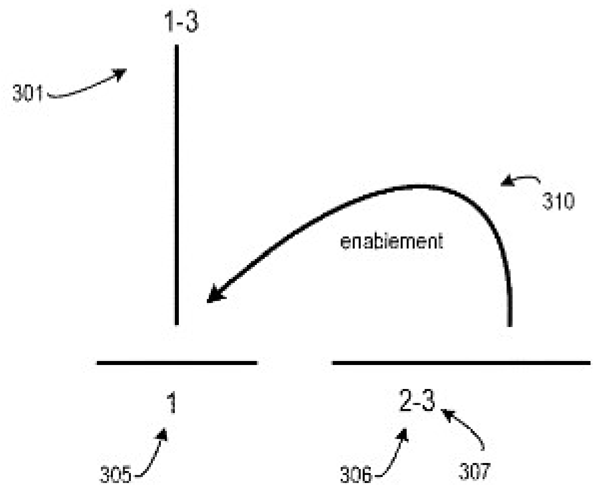

FIG. 3

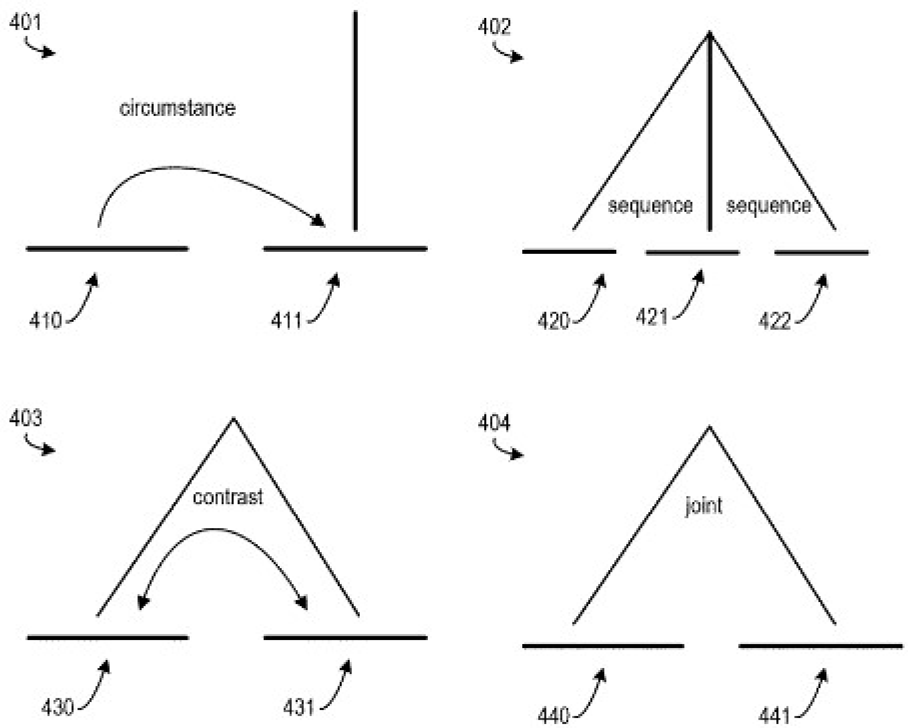
401
circumstance
410
411
402
sequence
sequence
420
421
422
403
contrast
430
431
404
joint
440
441

FIG. 4