



- (51) **International Patent Classification:**
G06Q 10/06 (2012.01) G06F 17/20 (2006.01)
- (21) **International Application Number:**
PCT/IB2012/056514
- (22) **International Filing Date:**
16 November 2012 (16.11.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** DATA2TEXT LIMITED [GB/GB]; Meston Building, University of Aberdeen, Aberdeen AB24 3UE (GB).
- (72) **Inventors:** SRIPADA, Gowri Somayajulu; 9 Brodiach Court, Westhill AB32 6QY (GB). BURNETT, Neil; 11 Whiteley Well Place, Inverurie AB51 4FP (GB).
- (74) **Agents:** CARTER, Stephen et al.; Mewburn Ellis LLP, 33 Guter Lane, London EC2V 8AS (GB).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))



WO 2014/076525 A1

(54) **Title:** METHOD AND APPARATUS FOR EXPRESSING TIME IN AN OUTPUT TEXT

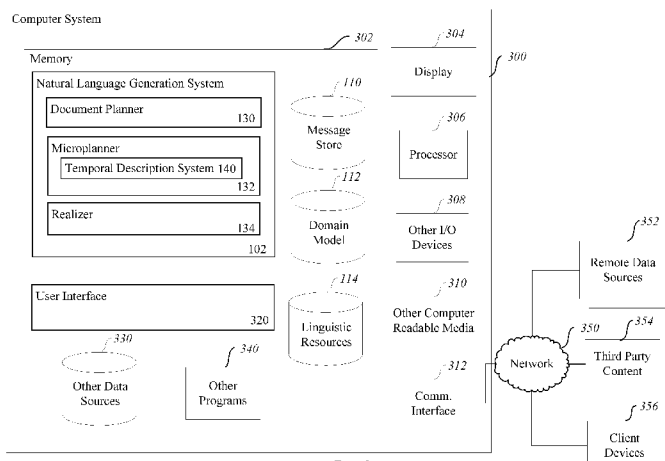


FIG. 3

(57) **Abstract:** Methods, apparatuses, and computer program products are described herein that are configured to express a time in an output text. In some example embodiments, a method is provided that comprises identifying a time period to be described linguistically in an output text. The method of this embodiment may also include identifying a communicative context for the output text. The method of this embodiment may also include determining one or more temporal reference frames that are applicable to the time period and a domain defined by the communicative context. The method of this embodiment may also include generating a phrase specification that linguistically describes the time period based on the descriptor that is defined by a temporal reference frame of the one or more temporal reference frames. In some examples, the descriptor specifies a time window that is inclusive of at least a portion of the time period to be described linguistically.

METHOD AND APPARATUS FOR EXPRESSING TIME IN AN OUTPUT TEXT

TECHNOLOGICAL FIELD

Embodiments of the present invention relate generally to natural language generation technologies and, more particularly, relate to a method, apparatus, and computer program product for expressing time in an output text.

BACKGROUND

In some examples, a natural language generation (NLG) system is configured to transform raw input data that is expressed in a non-linguistic format into a format that can be expressed linguistically, such as through the use of natural language. For example, raw input data may take the form of a value of a stock market index over time and, as such, the raw input data may include data that is suggestive of a time, a duration, a value and/or the like. Therefore, an NLG system may be configured to input the raw input data and output text that linguistically describes the value of the stock market index; for example, "securities markets rose steadily through most of the morning, before sliding downhill late in the day."

Data that is input into a NLG system may be provided in, for example, a recurrent formal structure. The recurrent formal structure may comprise a plurality of individual fields and defined relationships between the plurality of individual fields. For example, the input data may be contained in a spreadsheet or database, presented in a tabulated log message or other defined structure, encoded in a 'knowledge representation' such as the resource description framework (RDF) triples that make up the Semantic Web and/or the like. In some examples, the data may include numerical content, symbolic content or the like. Symbolic content may include, but is not limited to, alphanumeric and other non-numeric character sequences in any character encoding, used to represent arbitrary elements of information. In some examples, the output of the NLG system is text in a natural language (e.g. English, Japanese or Swahili), but may also be in the form of synthesized speech.

BRIEF SUMMARY

Methods, apparatuses, and computer program products are described herein that are configured to linguistically describe a time period detected in a data structure in an output text generated by a natural language generation system. In some example embodiments, a method is provided that comprises identifying the time period to be described linguistically in an output text. The method of this embodiment may also include identifying a communicative context for the output text. The method of this embodiment may also include determining one or more temporal reference frames that are applicable to the time period and are appropriate for the domain defined by the communicative context. The method of this embodiment may also include generating a phrase specification that linguistically describes the time period based on the descriptor that is defined by a temporal reference frame of the one or more temporal reference frames. In some examples, the descriptor specifies a time window that is inclusive of at least a portion of the time period to be described linguistically.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus described embodiments of the invention in general terms, reference will now be made to the accompanying drawings, which are not necessarily drawn to scale, and wherein:

Figure 1 is a schematic representation of natural language generation environment that may benefit from some example embodiments of the present invention;

Figure 2 illustrates an example expression of a time using a temporal description system according to some example embodiments described herein;

Figure 3 illustrates a block diagram of an apparatus that embodies a natural language generation system in accordance with some example embodiments of the present invention; and

Figure 4 illustrates a flowchart that may be performed by a temporal description system in accordance with some example embodiments of the present invention.

DETAILED DESCRIPTION

Example embodiments will now be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all, embodiments are shown. Indeed, the embodiments may take many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Like reference numerals refer to like elements throughout. The terms "data," "content," "information," and similar terms may be used interchangeably, according to some example embodiments, to refer

to data capable of being transmitted, received, operated on, and/or stored. Moreover, the term “exemplary”, as may be used herein, is not provided to convey any qualitative assessment, but instead merely to convey an illustration of an example. Thus, use of any such terms should not be taken to limit the spirit and scope of embodiments of the present invention.

Natural language generation systems may be configured to describe a time of an event, happening or the like in an output text. The time of the event may, in some cases, be described by its numerical time, such as “11:02 am”, but more often, the time of the event may be referred to by a description of that time period, such as “late morning”, “before lunch”, “before dawn”, “early in the semester” or the like. As such, to generate a description of a time, a natural language generation system may need external information or a communicative context (e.g. the domain of the event, the location of the reader of the output text, the time the output text is generated in comparison to the timing of the event and/or the like) of the output text in order to generate the preferred or otherwise appropriate description of the time. For example, a local time of 6:00 am may be before sunrise in some areas of the world whereas in other locations 6:00 am may be after sunrise. In other examples, “early morning” may be 6 am for a soldier, whereas “early morning” may be 9 am for a professional.

As is described herein and according to some example embodiments, a temporal description system is provided that enables the generation of a linguistic description, based on communicative context, in the form of a phrase specification for a time period that can be incorporated in an output text. A time period is a space of seconds, minutes, hours, days, weeks, months or years with an established beginning date and ending date. A time period may also include and/or be used interchangeably with, for example, a time point, a time window, a duration of time, an instance of time and/or the like.

In some examples, a microplanner may receive a document plan tree that contains or otherwise refers to a message that includes reference to a time period in a slot of the message. In order to convert the time period in the message into a phrase specification or a syntactic constituent for use in a phrase specification that may be processed by a microplanner, the microplanner may call or otherwise access a temporal description system, such as the temporal description system described herein. In some example embodiments, the temporal description system may reference or otherwise utilize one or more temporal reference frames that are aligned to a current communicative context and define one or more descriptors (e.g. linguistic words or phrases that describe a time window) to linguistically describe the time period. A temporal reference frame is a means of partitioning a given timeline into a set of time partitions called descriptors that can be refined based on a hierarchy. For example, a temporal reference frame may relate

to seasons and have descriptors called winter, spring, summer and fall. Those descriptors may be refined, such as by using a modifier “early,” “middle” and/or “late.” As such, the temporal description system may select one of the descriptors that at least partially include the time period received to describe the time period, such as “during the spring.” In some examples, a temporal relationship between a time period and a descriptor may be also be linguistically described to provide a more precise time reference, for example “early spring”.

Figure 1 is an example block diagram of example components of an example natural language generation environment 100. In some example embodiments, the natural language generation environment 100 comprises a natural language generation system 102, message store 110, a domain model 112 and/or linguistic resources 114. The natural language generation system 102 may take the form of, for example, a code module, a component, circuitry and/or the like. The components of the natural language generation environment 100 are configured to provide various logic (e.g. code, instructions, functions, routines and/or the like) and/or services related to the natural language generation system, the microplanner and/or a temporal description system.

A message store 110 or knowledge pool is configured to store one or more messages that are accessible by the natural language generation system 102. Messages are language independent data structures that correspond to informational elements in a text and/or collect together underlying data, referred to as slots, arguments or features, which can be presented within a fragment of natural language such as a phrase or sentence. Messages may be represented in various ways; for example, each slot may consist of a named attribute and its corresponding value; these values may recursively consist of sets of named attributes and their values, and each message may belong to one of a set of predefined types. The concepts and relationships that make up messages may be drawn from an ontology (e.g. a domain model 112) that formally represents knowledge about the application scenario. In some examples, the domain model 112 is a representation of information about a particular domain. For example, a domain model may contain an ontology that specifies the kinds of objects, instances, concepts and/or the like that may exist in the domain in concrete or abstract form, properties that may be predicated of the objects, concepts and the like, relationships that may hold between the objects, concepts and the like, a communicative context and representations of any specific knowledge that is required to function in the particular domain.

In some examples, messages are created based on a requirements analysis as to what is to be communicated for a particular scenario (e.g. for a particular domain or genre). A message typically corresponds to a fact about the underlying data (for example, the existence of some observed event) that could be expressed via a simple sentence

(although it may ultimately be realized by some other linguistic means). For example, to linguistically describe a weather event, such as a rain storm, a user may want to know the location of the rain storm, when it will reach the user's location, the last time rain was detected and/or the like. In some cases, the user may not want to know about a weather event, but instead want to be warned in an instance in which the weather presents a danger in a particular area; for example, "high winds predicted this evening."

In some examples, a message is created in an instance in which the raw input data warrants the construction of such a message. For example, a wind message would only be constructed in an instance in which wind data was present in the raw input data. Alternatively or additionally, while messages may correspond directly to observations taken from a raw data input, others, however, may be derived from the observations by means of a process of inference or based on one or more detected events. For example, the presence of rain may be indicative of other conditions, such as the potential for snow at some temperatures.

Messages may be instantiated based on many variations of source data, such as but not limited to time series data, time and space data, data from multiple data channels, an ontology, sentence or phrase extraction from one or more texts, a text, survey responses, structured data, unstructured data and/or the like. For example, in some cases, messages may be generated based on text related to multiple news articles focused on the same or similar news story in order to generate a news story; whereas, in other examples, messages may be built based on survey responses and/or event data.

Messages may be annotated with an indication of their relative importance; this information can be used in subsequent processing steps or by the natural language generation system 102 to make decisions about which information may be conveyed and which information may be suppressed. Alternatively or additionally, messages may include information on relationships between the one or more messages.

In some example embodiments, a natural language generation system, such as natural language generation system 102, is configured to generate words, phrases, sentences, text or the like which may take the form of a natural language text. The natural language generation system 102 comprises a document planner 130, a microplanner 132 and/or a realizer 134. The natural language generation system 102 may also be in data communication with the message store 110, the domain model 112 and/or the linguistic resources 114. In some examples, the linguistic resources 114 include, but are not limited to, text schemas, communicative context, aggregation rules, reference rules, lexicalization rules and/or grammar rules that may be used by one or more of the document planner 130, the microplanner 132 and/or the realizer 134. Other natural language generation systems may be used in some example embodiments, such as a

natural language generation system as described in *Building Natural Language Generation Systems* by Ehud Reiter and Robert Dale, Cambridge University Press (2000), which is incorporated by reference in its entirety herein.

The document planner 130 is configured to input the one or more messages from the message store 110 and to determine how to arrange those messages in order to describe one or more patterns in the one or more data channels derived from the raw input data. The document planner 130 may also comprise a content determination process that is configured to select the messages, such as the messages that contain a representation of the data that is to be output via a natural language text.

The document planner 130 may also comprise a structuring process that determines the order of messages to be included in a text. In some example embodiments, the document planner 130 may access one or more text schemas for the purposes of content determination and document structuring. A text schema is a rule set that defines the order in which a number of messages are to be presented in a document. For example, a rain message may be described prior to a temperature message. In other examples, a wind message may be described after, but in a specific relation to, the rain message.

The output of the document planner 130 may be a tree-structured object or other data structure that is referred to as a document plan. In an instance in which a tree-structured object is chosen for the document plan, the leaf nodes of the tree may contain the messages, and the intermediate nodes of the tree structure object may be configured to indicate how the subordinate nodes are related (e.g. elaboration, consequence, contrast, sequence and/or the like) to each other. An example document plan is shown with respect to document plan 202 of Figure 2. An example message is also shown in Figure 2 as message 206.

The microplanner 132 is configured to construct a text specification based on the document plan from the document planner 130, such that the document plan may be expressed in natural language. In some example embodiments, the microplanner 132 may convert the one or more messages in the document plan into one or more phrase specifications in a text specification. In some example embodiments, the microplanner 132 may perform aggregation, lexicalization and referring expression generation. In some examples, aggregation includes, but is not limited to, determining whether two or more messages can be combined together linguistically to produce a more complex phrase specification. For example, one or more messages may be aggregated so that both of the messages can be described by a single sentence. In some examples, lexicalization includes, but is not limited to, choosing particular words for the expression of concepts and relations. In some examples, referring expression generation includes, but is not

limited to, choosing how to refer to an entity so that it can be unambiguously identified by the reader.

In some example embodiments, the microplanner 132 may embody or otherwise may be in data communication with a temporal description system 140. The microplanner 132 may interact with the temporal description system 140 in an instance in which the microplanner detects a time period (e.g. a time point, a time window, a duration or the like) in a slot of a message in the document plan tree received or otherwise accessed via the document planner 130. As such, the temporal description system 140 is configured to determine or otherwise identify the communicative context of an output text as is provided via the domain model 112 and/or the linguistic resources 114. A communicative context is a factor or combinations of factors of the environment in which the events to be described are occurring and which have an influence on the output text. In some example embodiments, the factor or combination of factors may include a domain for which the text is to be generated (e.g. medical, weather, academic, sports and/or the like), a location of a reader of the output text or a location described by the output text (e.g. Scotland may have a later sunrise when compared to Italy), the current time that the output text is being generated (e.g. "6 am" may be an appropriate descriptor for event tomorrow, but "in the morning next month" may be a more appropriate descriptor to identify an event in the future), the time of the event (e.g. in order to set the tense of a verb), user or reader preferences (e.g. "early morning," "6am" or "0600"), language preferences (e.g. descriptors chosen based on regional dialects), and/or the like.

In some example embodiments, the temporal description system 140 may be configured to output a phrase specification that describes the input time period. The temporal description system 140 may linguistically describe the time period in the phrase specification using generic time descriptors, such as a day name, a date, an hour or the like. However, in other example embodiments, the temporal description system 140 may linguistically describe the time period using a descriptor that is defined by a temporal reference frame. A temporal reference frame is a set of time partitionings that are used to describe time in a particular domain (e.g. trimesters in pregnancy, semesters in a university and/or the like). A descriptor is the linguistically describable name of the various partitionings within a temporal reference frame (e.g. first, second and third trimesters of a pregnancy, fall semester and spring semester in a university and/or the like). The various partitionings may be further refined by using modifiers, may be refined based on a hierarchy and/or the like. In some examples, the temporal reference frames are configured to be aligned to or otherwise instantiated based on the communicative context (e.g. first trimester aligned to the conception date of a pregnancy, fall semester aligned to the fall start date, morning tied to a sunrise time and/or the like) by the

temporal description system 140. Descriptors may also be aligned within the temporal reference frames in some example embodiments.

The temporal description system 140, in some example embodiments, may then select a temporal reference frame and a descriptor from the available temporal reference frames. In some examples, the temporal reference frame and a descriptor may be chosen based on whether a particular descriptor describes the entire time period. In other examples, a domain model or the linguistic resources may provide a preferred ordering of temporal reference frames to be used in a particular output text. In other examples, a temporal reference frame and descriptor may be chosen based on previously chosen temporal reference frames or previously generated phrase specifications. By way of example, a previously referred-to temporal reference frame may be solar movement (e.g. before sunrise) and, as such, the temporal description system 140 may subsequently use a day period (e.g. morning or afternoon) temporal reference frame to add to readability and/or variety in the output text.

In some examples, the temporal reference frame and descriptor may be chosen based on a scoring system. The scoring system may be based on the ability of a descriptor to describe the entire time period, the detection of false positives (e.g. describing a time period that does not include the event) or false negatives (e.g. failing to describe a time period that contains the event) based on the descriptor and/or the like. In some examples, the temporal reference frame may be selected by the temporal description system 140 randomly. Alternatively or additionally, multiple descriptors within a reference frame may be used to describe a time period.

Once a temporal reference frame and one or more descriptors have been identified, the temporal description system 140 is configured to determine a temporal relationship between the time period and the one or more descriptors. For example, if a single descriptor, such as day period, is used, a temporal relationship may be defined as "early," "late" or "mid" if the time period relates to the descriptor in such a way (e.g. 8 am is "early in the day"). In an instance in which two or more descriptors are used, the temporal relationship between the time period and the descriptors may be defined by describing the time period as "between" or "overlapping" (e.g. "between lunch and dinner" or "shift 1 overlaps shift 2").

Using the temporal reference frame, the one or more descriptors and the determined relationship, the temporal description system 140 is configured to generate a phrase specification subject to one or more constraints. Constraints may include, but are not limited to, constraints imposed by a domain model, user preferences, language constraints, output text length constraints, readability and variety constraints, previously referred to temporal reference frames, previous phrase specifications and/or the like. The

temporal description system may then output or otherwise provide the phrase specification to the microplanner 132 to be used in a text specification as its own phrase specification or more likely incorporated inside another phrase specification. The output of the microplanner 132, in some example embodiments, is a tree-structured realization specification whose leaf-nodes are phrase specifications, and whose internal nodes express rhetorical relations between the leaf nodes.

A realizer 134 is configured to traverse a text specification output by the microplanner 132 to express the text specification in natural language. The realization process that is applied to each phrase specification and further makes use of a grammar (e.g. the grammar of the linguistic resources 114), which specifies the valid syntactic structures in the language and further provides a way of mapping from phrase specifications into the corresponding natural language sentences. The output of the process is, in some example embodiments, a natural language text. In some examples, the natural language text may include embedded mark-up.

Figure 2 illustrates an example expression of a time period using a temporal description system according to some example embodiments described herein. In some examples, and by using the systems and methods described with respect to Figure 1, a time period described in a message may be linguistically described via a temporal description system. One or more non-limiting examples of various operations of the temporal description system will now be described with respect to Figure 2.

In one example, the timing of weather events may make use of domain specific temporal reference frames, such solar movement (e.g. before or after sunrise), day period (e.g. morning, afternoon, evening), mealtime (e.g. around lunch), time (e.g. before 5pm) and/or the like. As such, and according to some example embodiments, a microplanner may detect a message, such as message 206 in document plan 202 that contains a time period such as a start time (e.g. StartTime = 0600 11 October 2012) and an end time (e.g. EndTime = 1500 11 October 2012). The microplanner, when converting the message to a phrase specification, is configured to call or otherwise access the temporal description system 140 to convert the time period defined by the StartTime and the EndTime into a phrase specification, such as phrase specification 208 that is part of phrase specification 210 in the text specification 204 (e.g. "from early morning into the afternoon"). Using the message 206 and the phrase specification 210, a text may be generated (e.g. via the realizer) such as: "rain is expected from early morning into the afternoon."

In further examples, a document plan may have multiple messages having references to one or more time periods. As such, multiple temporal reference frames may be used to describe the multiple time periods. For example, the following text uses four

temporal reference frames: “showers are likely to develop before sunrise before clearing into the morning. There should be some sunny spells around lunchtime lasting until around 5pm.” The four example temporal reference frames selected by the temporal description system, in this example, are: solar movement (“before sunrise”), day period (“into the morning”), mealtime (“around lunchtime”) and time (“around 5pm”). In other examples, the temporal description system 140 may also select alternate temporal reference frames. Alternatively or additionally, the temporal reference frames may be used in a different order, such as is shown in the following text: “Showers are likely to develop during the early morning before clearing later. There should be some sunny spells from midday lasting into the evening.”

By way of another example in a medical domain, a time period used to describe a key event during a pregnancy could be described as occurring during the following non-exhaustive list of temporal reference frames: month number (e.g. “in the 6th month”), trimester (e.g. “during the 2nd trimester”), month name (e.g. “during July”), week number (e.g. “in week 30”), day number (e.g. “around day 210”) as well as simply giving a date or time (e.g. “Friday”, “July 5”, “5am” or the like). By way of further example, the temporal reference frame TRIMESTER may include example descriptors “first trimester”, “second trimester” or “third trimester”. As such, using the domain-specific temporal reference frames, an example output text may include three descriptors that belong to three different temporal reference frames: “During the 2nd trimester, the baby developed as expected. There were concerns during April that the baby was starting to turn but these concerns have subsided due to the baby turning back to the normal position on day 90 of this pregnancy.” In this example, the baby’s development has been detected as fine for a 3-month-long period. Without the introduction of the TRIMESTER temporal reference frame, another descriptor may have been used, such as “May to July,” but this may, in some examples, result in fairly repetitive and limited text. By enabling the temporal description system 140 to describe the same time span using multiple descriptors, advantageously, for example, a greater variation in descriptors may be achieved in the resultant output text. Alternatively or additionally, temporal reference frames may be reused in some example embodiments.

In some examples, the temporal description system 140 is configured to use descriptors such as sunrise and sunset times for a specific location along with the time of day and span of an input time period to generate an output text to linguistically describe a time period. By making use of sunrise and sunset times, it is also possible to take into account the time of year when creating the linguistic description of the input time period. For example, during the summer, the night time is shorter than in the middle of winter. Therefore the time period that can be linguistically described as “morning” covers a longer

time period if the input time period is during the summer than if the input time period was during the winter. Further, during the winter, the sun may only be up in the morning for a few hours in some locations, therefore the temporal description system 140 may not need to describe a relationship between the time period and the descriptor and, as such, may drop "early" from "early morning."

Figure 3 is an example block diagram of an example computing device for practicing embodiments of an example temporal description system. In particular, Figure 3 shows a computing system 300 that may be utilized to implement a natural language generation environment 100 having a natural language generation system 102 including, in some examples, a document planner 130, a microplanner 132 and/or a realizer 134 and/or a user interface 320. One or more general purpose or special purpose computing systems/devices may be used to implement the natural language generation system 102 and/or the user interface 320. In addition, the computing system 300 may comprise one or more distinct computing systems/devices and may span distributed locations. In some example embodiments, the natural language generation system 102 may be configured to operate remotely via the network 350. In other example embodiments, a pre-processing module or other module that requires heavy computational load may be configured to perform that computational load and thus may be on a remote device or server. For example, the realizer 134 or the temporal description system may be accessed remotely. In other example embodiments, a user device may be configured to operate or otherwise access the natural language generation system 102. Furthermore, each block shown may represent one or more such blocks as appropriate to a specific example embodiment. In some cases one or more of the blocks may be combined with other blocks. Also, the natural language generation system 102 may be implemented in software, hardware, firmware, or in some combination to achieve the capabilities described herein.

In the example embodiment shown, computing system 300 comprises a computer memory ("memory") 302, a display 304, one or more processors 306, input/output devices 308 (e.g., keyboard, mouse, CRT or LCD display, touch screen, gesture sensing device and/or the like), other computer-readable media 310, and communications interface 312. The processor 306 may, for example, be embodied as various means including one or more microprocessors with accompanying digital signal processor(s), one or more processor(s) without an accompanying digital signal processor, one or more coprocessors, one or more multi-core processors, one or more controllers, processing circuitry, one or more computers, various other processing elements including integrated circuits such as, for example, an application-specific integrated circuit (ASIC) or field-programmable gate array (FPGA), or some combination thereof. Accordingly, although

illustrated in Figure 3 as a single processor, in some embodiments the processor 306 comprises a plurality of processors. The plurality of processors may be in operative communication with each other and may be collectively configured to perform one or more functionalities of the temporal description system as described herein.

The natural language generation system 102 is shown residing in memory 302. The memory 302 may comprise, for example, transitory and/or non-transitory memory, such as volatile memory, non-volatile memory, or some combination thereof. Although illustrated in Figure 3 as a single memory, the memory 302 may comprise a plurality of memories. The plurality of memories may be embodied on a single computing device or may be distributed across a plurality of computing devices collectively configured to function as the natural language system, the microplanner and/or the temporal description system. In various example embodiments, the memory 302 may comprise, for example, a hard disk, random access memory, cache memory, flash memory, a compact disc read only memory (CD-ROM), digital versatile disc read only memory (DVD-ROM), an optical disc, circuitry configured to store information, or some combination thereof.

In other embodiments, some portion of the contents, some or all of the components of the natural language generation system 102 may be stored on and/or transmitted over the other computer-readable media 310. The components of the natural language generation system 102 preferably execute on one or more processors 306 and are configured to enable operation of a temporal description system, as described herein.

Alternatively or additionally, other code or programs 340 (e.g., an administrative interface, one or more application programming interface, a Web server, and the like) and potentially other data repositories, such as other data sources 330, also reside in the memory 302, and preferably execute on one or more processors 306. Of note, one or more of the components in Figure 3 may not be present in any specific implementation. For example, some embodiments may not provide other computer readable media 310 or a display 304.

The natural language generation system 102 is further configured to provide functions such as those described with reference to Figure 1. The natural language generation system 102 may interact with the network 350, via the communications interface 312, with remote data sources 352 (e.g. remote reference data, remote performance data, remote aggregation data, remote knowledge pools and/or the like), third-party content providers 354 and/or client devices 356. The network 350 may be any combination of media (e.g., twisted pair, coaxial, fiber optic, radio frequency), hardware (e.g., routers, switches, repeaters, transceivers), and protocols (e.g., TCP/IP, UDP, Ethernet, Wi-Fi, WiMAX, Bluetooth) that facilitate communication between remotely situated humans and/or devices. In some instances, the network 350 may take the form

of the internet or may be embodied by a cellular network such as an LTE based network. In this regard, the communications interface 312 may be capable of operating with one or more air interface standards, communication protocols, modulation types, access types, and/or the like. The client devices 356 include desktop computing systems, notebook computers, mobile phones, smart phones, personal digital assistants, tablets and/or the like. In some example embodiments, a client device may embody some or all of computing system 300.

In an example embodiment, components/modules of the natural language generation system 102 are implemented using standard programming techniques. For example, the natural language generation system 102 may be implemented as a "native" executable running on the processor 306, along with one or more static or dynamic libraries. In other embodiments, the natural language generation system 102 may be implemented as instructions processed by a virtual machine that executes as one of the other programs 340. In general, a range of programming languages known in the art may be employed for implementing such example embodiments, including representative implementations of various programming language paradigms, including but not limited to, object-oriented (e.g., Java, C++, C#, Visual Basic.NET, Smalltalk, and the like), functional (e.g., ML, Lisp, Scheme, and the like), procedural (e.g., C, Pascal, Ada, Modula, and the like), scripting (e.g., Perl, Ruby, Python, JavaScript, VBScript, and the like), and declarative (e.g., SQL, Prolog, and the like).

The embodiments described above may also use synchronous or asynchronous client-server computing techniques. Also, the various components may be implemented using more monolithic programming techniques, for example, as an executable running on a single processor computer system, or alternatively decomposed using a variety of structuring techniques, including but not limited to, multiprogramming, multithreading, client-server, or peer-to-peer, running on one or more computer systems each having one or more processors. Some embodiments may execute concurrently and asynchronously, and communicate using message passing techniques. Equivalent synchronous embodiments are also supported. Also, other functions could be implemented and/or performed by each component/module, and in different orders, and by different components/modules, yet still achieve the described functions.

In addition, programming interfaces to the data stored as part of the natural language generation system 102, such as by using one or more application programming interfaces can be made available by mechanisms such as through application programming interfaces (API) (e.g. C, C++, C#, and Java); libraries for accessing files, databases, or other data repositories; through scripting languages such as XML; or through Web servers, FTP servers, or other types of servers providing access to stored

data. The message store 110, the domain model 112 and/or the linguistic resources 114 may be implemented as one or more database systems, file systems, or any other technique for storing such information, or any combination of the above, including implementations using distributed computing techniques. Alternatively or additionally, the message store 110, the domain model 112 and/or the linguistic resources 114 may be local data stores but may also be configured to access data from the remote data sources 352.

Different configurations and locations of programs and data are contemplated for use with techniques described herein. A variety of distributed computing techniques are appropriate for implementing the components of the illustrated embodiments in a distributed manner including but not limited to TCP/IP sockets, RPC, RMI, HTTP, Web Services (XML-RPC, JAX-RPC, SOAP, and the like). Other variations are possible. Also, other functionality could be provided by each component/module, or existing functionality could be distributed amongst the components/modules in different ways, yet still achieve the functions described herein.

Furthermore, in some embodiments, some or all of the components of the natural language generation system 102 may be implemented or provided in other manners, such as at least partially in firmware and/or hardware, including, but not limited to one or more ASICs, standard integrated circuits, controllers executing appropriate instructions, and including microcontrollers and/or embedded controllers, FPGAs, complex programmable logic devices ("CPLDs"), and the like. Some or all of the system components and/or data structures may also be stored as contents (e.g., as executable or other machine-readable software instructions or structured data) on a computer-readable medium so as to enable or configure the computer-readable medium and/or one or more associated computing systems or devices to execute or otherwise use or provide the contents to perform at least some of the described techniques. Some or all of the system components and data structures may also be stored as data signals (e.g., by being encoded as part of a carrier wave or included as part of an analog or digital propagated signal) on a variety of computer-readable transmission mediums, which are then transmitted, including across wireless-based and wired/cable-based mediums, and may take a variety of forms (e.g., as part of a single or multiplexed analog signal, or as multiple discrete digital packets or frames). Such computer program products may also take other forms in other embodiments. Accordingly, embodiments of this disclosure may be practiced with other computer system configurations.

Figure 4 is a flowchart illustrating an example method performed by a temporal description system in accordance with some example embodiments described herein. As is shown in operation 402, an apparatus may include means, such as the microplanner

132, the temporal description system 140, the processor 306, or the like, for receiving a time period to be described linguistically in an output text. In some example embodiments, the time period may define a time point or a series of time points. As is shown in operation 404, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for identifying a communicative context for the output text.

As is shown in operation 406, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for determining one or more temporal reference frames that are applicable to the time period and are appropriate for the domain defined by the communicative context. For example, the temporal reference frame that is partitioned into trimesters that define a pregnancy term would likely only be appropriate for the medical domain and not for the weather domain. As is shown in operation 408, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for instantiating or otherwise aligning the one or more temporal reference frames to the communicative context. For example, if the temporal reference frame is partitioned into trimesters that define a pregnancy term, then an example first trimester would be instantiated with the conception date and a date three months later.

As is shown in operation 410, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for selecting a descriptor within a temporal reference frame of the one or more temporal reference frames based on the time period and one or more previously generated phrase specifications. As is described herein, a temporal reference frame may be chosen at random, may be selected by a user, may be dictated by a domain model and/or the like. In some examples, a temporal reference frame may be different than the previously referred-to temporal reference frame. In some example embodiments, each temporal reference frame of the one or more temporal reference frames are partitioned, such that a descriptor is configured to linguistically describe each partition (e.g. each partition or descriptor has a linguistically-expressible name).

As is shown in operation 412, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for determining a temporal relationship between the time period and the descriptor. For example the time period may be at the beginning of a descriptor. As such, the relationship may, for example, be "early" in the descriptor "March." As is shown in operation 414, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for generating a phrase specification based on the descriptor and the temporal relationship between the time period and the descriptor.

As is shown in operation 416, an apparatus may include means, such as the microplanner 132, the temporal description system 140, the processor 306, or the like, for verifying the phrase specification using one or more constraints. A constraint may include, but is not limited to, verifying that the same descriptor is not used consecutively in the same sentence, that a descriptor is not used incorrectly within a communicative context (e.g. 6 am is not described as "morning" when the sun has not yet risen), context based constraints, constraints due to the text length and/or the like.

Figure 4 illustrates an example flowchart of the operations performed by an apparatus, such as computing system 300 of Figure 3, in accordance with example embodiments of the present invention. It will be understood that each block of the flowchart, and combinations of blocks in the flowchart, may be implemented by various means, such as hardware, firmware, one or more processors, circuitry and/or other devices associated with execution of software including one or more computer program instructions. For example, one or more of the procedures described above may be embodied by computer program instructions. In this regard, the computer program instructions which embody the procedures described above may be stored by a memory 302 of an apparatus employing an embodiment of the present invention and executed by a processor 306 in the apparatus. As will be appreciated, any such computer program instructions may be loaded onto a computer or other programmable apparatus (e.g., hardware) to produce a machine, such that the resulting computer or other programmable apparatus provides for implementation of the functions specified in the flowchart's block(s). These computer program instructions may also be stored in a non-transitory computer-readable storage memory that may direct a computer or other programmable apparatus to function in a particular manner, such that the instructions stored in the computer-readable storage memory produce an article of manufacture, the execution of which implements the function specified in the flowchart's block(s). The computer program instructions may also be loaded onto a computer or other programmable apparatus to cause a series of operations to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions which execute on the computer or other programmable apparatus provide operations for implementing the functions specified in the flowchart's block(s). As such, the operations of Figure 4, when executed, convert a computer or processing circuitry into a particular machine configured to perform an example embodiment of the present invention. Accordingly, the operations of Figure 4 define an algorithm for configuring a computer or processor, to perform an example embodiment. In some cases, a general purpose computer may be provided with an instance of the processor which performs the

algorithm of Figure 4 to transform the general purpose computer into a particular machine configured to perform an example embodiment.

Accordingly, blocks of the flowchart support combinations of means for performing the specified functions and combinations of operations for performing the specified functions. It will also be understood that one or more blocks of the flowchart, and combinations of blocks in the flowchart, can be implemented by special purpose hardware-based computer systems which perform the specified functions, or combinations of special purpose hardware and computer instructions.

In some example embodiments, certain ones of the operations herein may be modified or further amplified as described below. Moreover, in some embodiments additional optional operations may also be included. It should be appreciated that each of the modifications, optional additions or amplifications described herein may be included with the operations herein either alone or in combination with any others among the features described herein.

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Moreover, although the foregoing descriptions and the associated drawings describe example embodiments in the context of certain example combinations of elements and/or functions, it should be appreciated that different combinations of elements and/or functions may be provided by alternative embodiments without departing from the scope of the appended claims. In this regard, for example, different combinations of elements and/or functions than those explicitly described above are also contemplated as may be set forth in some of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

THAT WHICH IS CLAIMED:

1. A method comprising:
 - identifying a time period to be described linguistically in an output text;
 - identifying a communicative context for the output text;
 - determining one or more temporal reference frames that are applicable to the time period and a domain defined by the communicative context; and
 - generating a phrase specification that linguistically describes the time period based on the descriptor that is defined by a temporal reference frame of the one or more temporal reference frames, wherein the descriptor specifies a time window that is inclusive of at least a portion of the time period to be described linguistically.
2. A method according to Claim 1, further comprising:
 - instantiating or aligning the one or more temporal reference frames to the communicative context; and
 - selecting the descriptor defined by the temporal reference frame of the one or more temporal reference frames based on the time period and one or more previously generated phrase specifications.
3. A method according to Claim 1, further comprising:
 - determining a temporal relationship between the time period and the descriptor, wherein the phrase specification is further based on the temporal relationship between the time period and the descriptor.
4. A method according to Claim 3, further comprising:
 - verifying the phrase specification using one or more constraints.
5. A method according to Claim 1, further comprising:
 - identifying another time period to be described linguistically; and

determining a descriptor that is defined by another temporal reference frame of the one or more temporal reference frames.

6. A method according to Claim 5, wherein the descriptor for the another time period is different from the descriptor used for the time period.

7. A method according to Claim 1, wherein each temporal reference frame of the one or more temporal reference frames are partitioned, such that a descriptor is configured to linguistically describe each partition.

8. A method according to Claim 7, wherein the communicative context is defined by at least one of a domain model, a current time or a current location.

9. A method according to Claim 1, wherein the time period comprises at least one of a time point or a series of time points.

10. A method of Claim 1, further comprising:

determining that a single descriptor does not specify a time window that is inclusive of the time period to be described linguistically; and
determining two or more descriptors that are defined by a temporal reference frame of the one or more temporal reference frames, wherein the two or more descriptors specify a time window that is inclusive of the time period to be described linguistically.

11. An apparatus comprising:

at least one processor; and

at least one memory including computer program code, at least one memory and the computer program code configured to, with the at least one processor, cause the apparatus to: at least:

identify a time period to be described linguistically in an output text;

identify a communicative context for the output text;

determine one or more temporal reference frames that are applicable to the time period and a domain defined by the communicative context; and

generate a phrase specification that linguistically describes the time period based on the descriptor that is defined by a temporal reference frame of the one or more temporal reference frames, wherein the descriptor specifies a time window that is inclusive of at least a portion of the time period to be described linguistically.

12. An apparatus according to Claim 11, wherein the at least one memory including the computer program code is further configured to, with the at least one processor, cause the apparatus to:

align the one or more temporal reference frames to the communicative context; and

select the descriptor defined by the temporal reference frame of the one or more temporal reference frames based on the time period and one or more previously generated phrase specifications.

13. An apparatus according to Claim 11, wherein the at least one memory including the computer program code is further configured to, with the at least one processor, cause the apparatus to:

determine a temporal relationship between the time period and the descriptor, wherein the phrase specification is further based on the temporal relationship between the time period and the descriptor.

14. An apparatus according to Claim 13, wherein the at least one memory including the computer program code is further configured to, with the at least one processor, cause the apparatus to:

verify the phrase specification using one or more constraints.

15. An apparatus according to Claim 11, wherein the at least one memory including the computer program code is further configured to, with the at least one processor, cause the apparatus to:

identify another time period to be described linguistically; and
determine a descriptor that is defined by another temporal reference frame of the one or more temporal reference frames.

16. An apparatus according to Claim 15, wherein the descriptor for the another time period is different from the descriptor used for the time period.

17. An apparatus according to Claim 11, wherein each temporal reference frame of the one or more temporal reference frames are partitioned, such that a descriptor is configured to linguistically describe each partition.

18. An apparatus according to Claim 17, wherein the communicative context is defined by at least one of a domain model, a current time or a current location.

19. An apparatus according to Claim 11, wherein the time period comprises at least one of a time point or a series of time points.

20. An apparatus of Claim 11, wherein the at least one memory including the computer program code is further configured to, with the at least one processor, cause the apparatus to:

determine that a single descriptor does not specify a time window that is inclusive of the time period to be described linguistically; and
determine two or more descriptors that are defined by a temporal reference frame of the one or more temporal reference frames, wherein the two or more descriptors specify a time window that is inclusive of the time period to be described linguistically.

21. A computer program product comprising:

at least one computer readable non-transitory memory medium having program code instructions stored thereon, the program code instructions which when executed by an apparatus cause the apparatus at least to:

identify a time period to be described linguistically in an output text;

identify a communicative context for the output text;

determine one or more temporal reference frames that are applicable to the time period and a domain defined by the communicative context; and

generate a phrase specification that linguistically describes the time period based on a descriptor that is defined by a temporal reference frame of the one or more temporal reference frames, wherein the descriptor specifies a time window that is inclusive of at least a portion of the time period to be described linguistically.

22. A computer program product according to Claim 21, further comprising program code instructions configured to:

align the one or more temporal reference frames to the communicative context; and

select the descriptor defined by the temporal reference frame of the one or more temporal reference frames based on the time period and one or more previously generated phrase specifications.

23. A computer program product according to Claim 21, further comprising program code instructions configured to:

determine a temporal relationship between the time period and the descriptor, wherein the phrase specification is further based on the temporal relationship between the time period and the descriptor.

24. A computer program product according to Claim 23, further comprising program code instructions configured to:

verify the phrase specification using one or more constraints.

25. A computer program product according to Claim 21, further comprising program code instructions configured to:

identify another time period to be described linguistically; and

determine a descriptor that is defined by another temporal reference frame of the one or more temporal reference frames.

26. A computer program product according to Claim 25 wherein the descriptor for the another time period is different from the descriptor used for the time period.

27. A computer program product according to Claim 21, wherein each temporal reference frame of the one or more temporal reference frames are partitioned, such that a descriptor is configured to linguistically describe each partition.

28. A computer program product according to Claim 27, wherein the communicative context is defined by at least one of a domain model, a current time or a current location.

29. A computer program product according to Claim 21, wherein the time period comprises at least one of a time point or a series of time points.

30. A computer program product of Claim 21, further comprising program code instructions configured to:

determine that a single descriptor does not specify a time window that is inclusive of the time period to be described linguistically; and

determine two or more descriptors that are defined by a temporal reference frame of the one or more temporal reference frames, wherein the two or more descriptors specify a time window that is inclusive of the time period to be described linguistically.

100 ↘

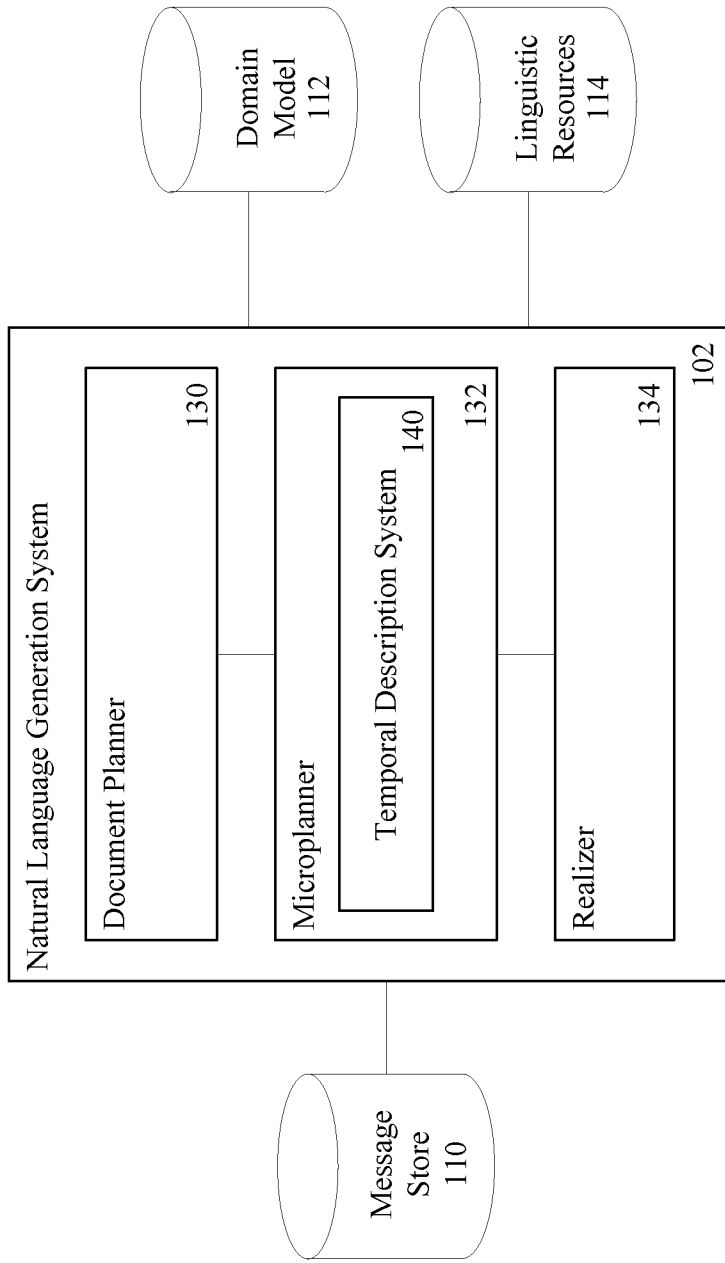


FIG. 1

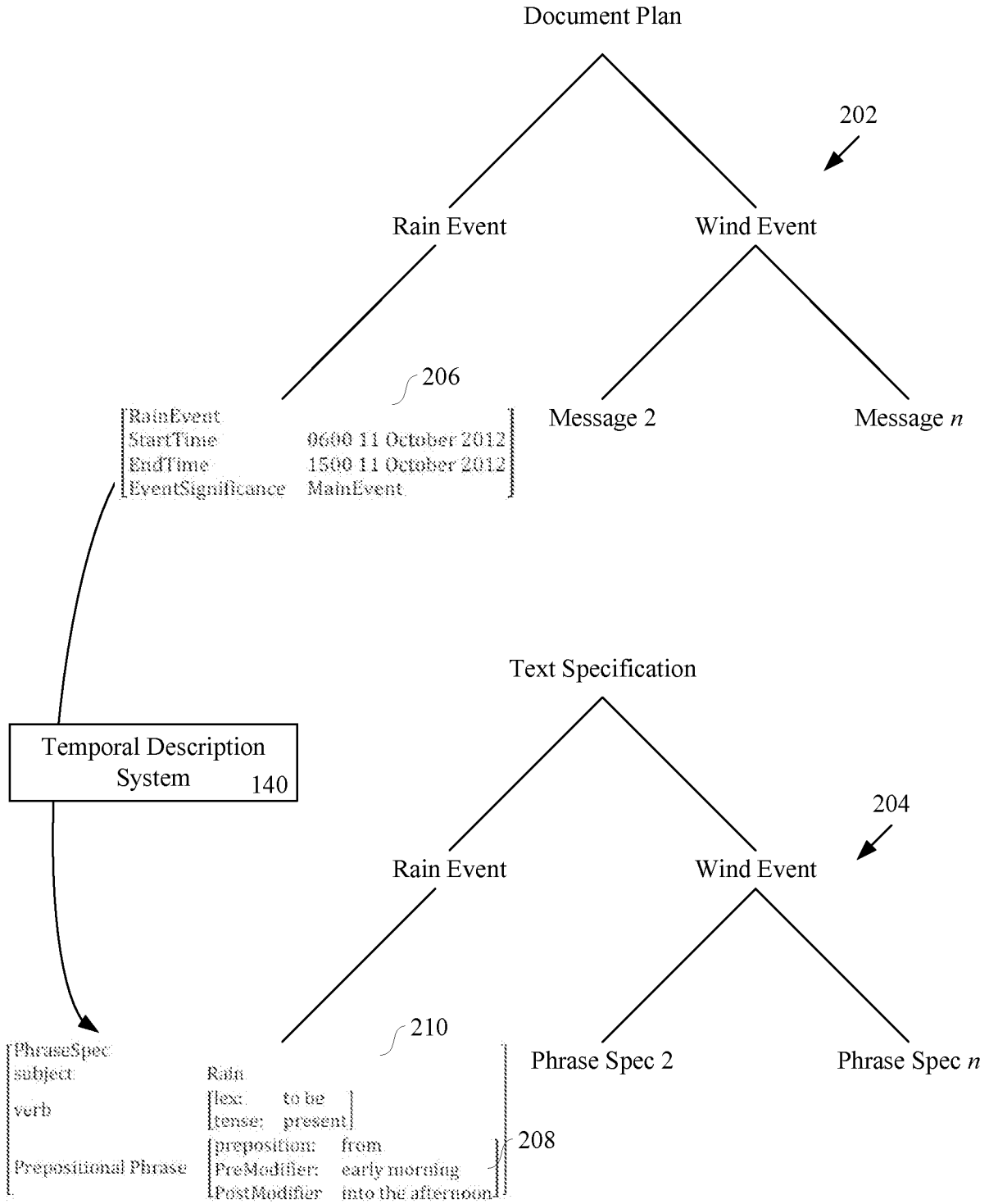


FIG. 2

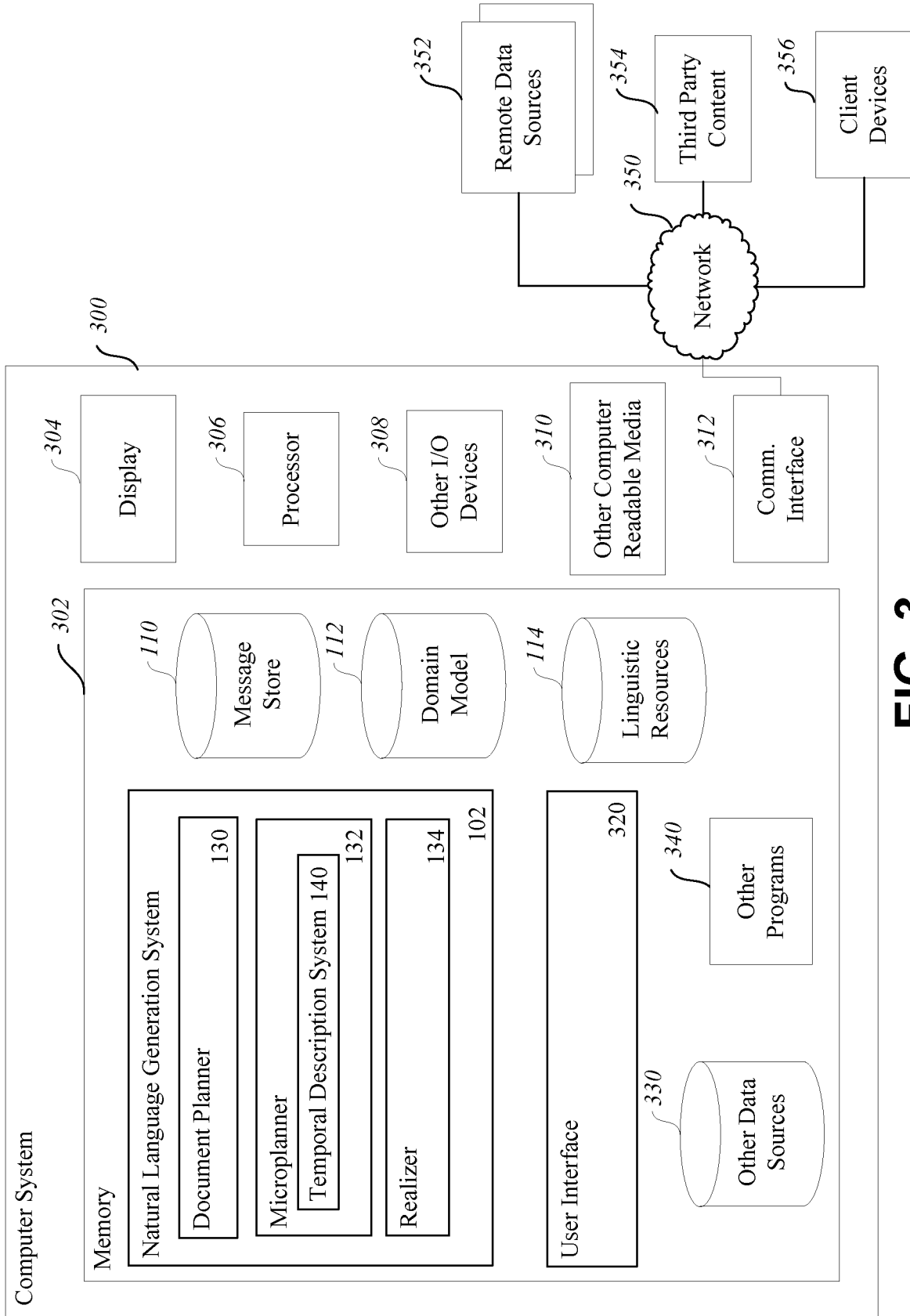
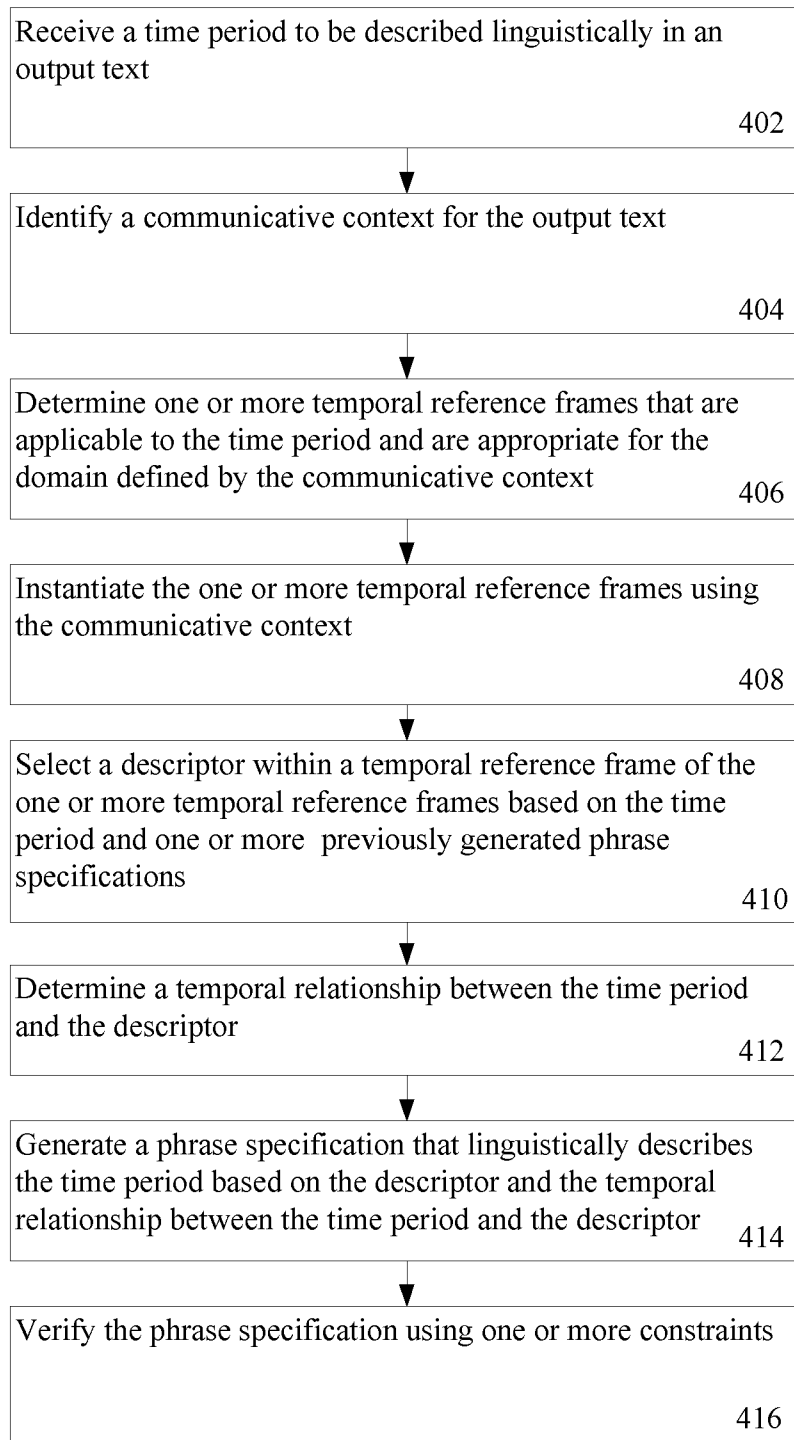


FIG. 3

**FIG. 4**

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2012/056514

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06Q10/06 G06F17/20
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06Q G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2006/178868 A1 (BILLEREY-MOSIER ROGER [US]) 10 August 2006 (2006-08-10) abstract paragraph [0007] - paragraph [0011] paragraph [0020] paragraph [0022] paragraph [0028] paragraph [0030] - paragraph [0033] paragraph [0036] - paragraph [0038] paragraph [0048] - paragraph [0050] -----	1-30
A	US 2002/026306 A1 (BANGALORE SRINIVAS [US] ET AL) 28 February 2002 (2002-02-28) the whole document -----	1-30

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 June 2013

Date of mailing of the international search report

26/06/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Moltenbrey, Michael

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2012/056514

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006178868	A1	10-08-2006	NONE

US 2002026306	A1	28-02-2002	US 7562005 B1 14-07-2009
			US 2002026306 A1 28-02-2002
			US 2005267751 A1 01-12-2005
