

## DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number:	<b>PCT/AU2016/000235</b>
International filing date:	<b>30 June 2016 (30.06.2016)</b>
Document type:	<b>Certified copy of priority document</b>
Document details:	Country/Office: <b>AU</b>
	Number: <b>2015902540</b>
	Filing date: <b>30 June 2015 (30.06.2015)</b>
Date of receipt at the International Bureau:	<b>05 August 2016 (05.08.2016)</b>

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a),(b) or (b-bis)



Australian Government

PCT/AU2016/000235

**Patent Office  
Canberra**

I, KELLY YULE, DELEGATE OF THE COMMISSIONER, hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2015902540 for a patent by HEALTH LANGUAGE ANALYTICS PTY LTD as filed on 30 June 2015.

**WITNESS** my hand this  
TWENTY NINTH day of July 2016

A handwritten signature in black ink, appearing to read 'K. Yule', written over a large, loopy flourish.

KELLY YULE  
DELEGATE OF THE COMMISSIONER



# FRAMEWORKS AND METHODOLOGIES FOR ENABLING CATEGORISATION OF CLINICAL REPORT DATA

## FIELD OF THE INVENTION

[0001] The present invention relates to frameworks and methodologies for enabling categorisation of clinical report data. Embodiments of the invention have been particularly developed to assist categorisation of clinical report data in a streamlined manner based on a pre-existing set of classification codes. This, in some embodiments, enables the discovery and extraction of meaningful patterns from unstructured clinical reports. While some embodiments will be described herein with particular reference to that application, it will be appreciated that the invention is not limited to such a field of use, and is applicable in broader contexts.

## BACKGROUND

[0002] Any discussion of the background art throughout the specification should in no way be considered as an admission that such art is widely known or forms part of common general knowledge in the field.

[0003] A huge quantity of clinical record data is produced each and every day. Even for a relatively small medical facility, the task of categorising and organising clinical record data can be unwieldy. Furthermore, due to complexities of clinical language, text or keyword based searching of electronic records is limited in its efficacy.

## SUMMARY OF THE INVENTION

[0004] It is an object of the present invention to overcome or ameliorate at least one of the disadvantages of the prior art, or to provide a useful alternative.

[0005] One embodiment provides a computer implemented method for enabling searching of clinical records, the method including:

[0006] A computer implemented method for enabling categorisation of clinical records, the method including:

[0007] identifying a set of user-generated text contained in a user interface object in a user interface environment;

[0008] processing the set of user-generated text based on a natural language processing engine, wherein the natural language processing engine leverages a clinical terminology knowledge base;

[0009] based on the processing, identifying a set of one or more suggested classification codes, wherein the classification codes are defined by a pre-existing clinical code structure;

[0010] causing display of a user interface component in the user interface environment, wherein the user interface component: (i) displays the identified set of one or more suggested classification codes; and (ii) enables a user to selectively confirm/reject one or more of the set of identified codes;

[0011] receiving input representative of the one or more of the set of identified codes selectively confirmed by the user; and

[0012] providing a signal thereby to cause association of a clinical record including the set of user-generated text with the set of identified codes selectively confirmed by the user.

[0013] One embodiment provides a method wherein the user interface component is additionally configured to provide a search interface that enables a user to identify and selectively confirm one or more further codes, wherein the a signal causes association of a clinical record including the set of user-generated text with the set of identified codes selectively confirmed by the user and the one or more further codes.

[0014] One embodiment provides a method wherein, in response to an event whereby a user identifies and selectively confirms one or more further codes, a trigger is initiated thereby to cause updating of the natural language processing engine and/or clinical terminology knowledge base.

[0015] One embodiment provides a method including causing rendering of a graphical output that associates each identified suggested clinical classification code with a portion of the user-generated text.

[0016] One embodiment provides a method including causing rendering of a graphical output that associates each identified suggested clinical classification code with additional clinical code context.

[0017] One embodiment provides a method wherein the identification of suggested clinical classification codes is configured to identify a disease stage.

[0018] One embodiment provides a method wherein the user interface component is provided by a software application separate to a further software application in which the set of user generated text is contained.

[0019] One embodiment provides a computer program product for performing a method as described herein.

[0020] One embodiment provides a non-transitory carrier medium for carrying computer executable code that, when executed on a processor, causes the processor to perform a method as described herein.

[0021] One embodiment provides a system configured for performing a method as described herein.

[0022] Reference throughout this specification to "one embodiment", "some embodiments" or "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment", "in some embodiments" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to one of ordinary skill in the art from this disclosure, in one or more embodiments.

[0023] As used herein, unless otherwise specified the use of the ordinal adjectives "first", "second", "third", etc., to describe a common object, merely indicate that different instances of like objects are being referred to, and are not intended to imply that the objects so described must be in a given sequence, either temporally, spatially, in ranking, or in any other manner.

[0024] In the claims below and the description herein, any one of the terms comprising, comprised of or which comprises is an open term that means including at least the elements/features that follow, but not excluding others. Thus, the term comprising, when used in the claims, should not be interpreted as being limitative to the means or elements or steps listed thereafter. For example, the scope of the expression a device comprising A and B should not be limited to devices consisting only of elements A and B. Any one of the terms including or which includes or that includes as used herein is also an open term that also means including at least the elements/features that follow the term, but not excluding others. Thus, including is synonymous with and means comprising.

[0025] As used herein, the term "exemplary" is used in the sense of providing examples, as opposed to indicating quality. That is, an "exemplary embodiment" is an embodiment provided as an example, as opposed to necessarily being an embodiment of exemplary quality.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0026] Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

[0027] FIG. 1 schematically illustrates a framework according to one embodiment.

[0028] FIG. 2 illustrates a method according to one embodiment.

[0029] FIG. 3 illustrates an exemplary client-server framework.

[0030] FIG. 4 illustrates a search interface according to one embodiment.

## DETAILED DESCRIPTION

[0031] Described herein are frameworks and methodologies for enabling categorisation of clinical report data. Embodiments of the invention have been particularly developed to assist categorisation of clinical report data in a streamlined manner based on a pre-existing set of classification codes. This, in some embodiments, enables the discovery and extraction of meaningful patterns from unstructured clinical reports. While some embodiments will be described herein with particular reference to that application, it will be appreciated that the invention is not limited to such a field of use, and is applicable in broader contexts.

[0032] In overview, the technology is described by reference to an abstractor component, configured to provide automation of abstracting and coding of clinical text (for example clinical reports, which may include notes and the like). This makes use of a natural language processing engine (for example a text analytics engine) that operates in conjunction with a leverages a clinical terminology knowledge base (which provides an expandable intimate knowledge of clinical terminology above and beyond a general purpose natural language processing engine). This provides a concept-based abstracting of clinical records that can ignore the idiosyncrasies of clinicians' writing styles and language usage to accurately code records. The abstractor component may be configured to operate within an existing software environment, directly interact with an existing software environment (for example via an API), or indirectly interact with an existing software environment (for example via operating system level functionalities).

[0033] Various embodiments provide computer implemented methods for enabling categorisation of clinical records. Such methods include identifying a set of user-generated text contained in a user interface object in a user interface environment. For example, in some cases this is text contained in an active object, such as a specific data field rendered by a given software application, an active window in third party software, a referenced file, or the like. In some cases the set of user-generated text includes all text in the object; in other cases it is a reduced selection (for example a user-designated selection, which is for instance made via a highlighting function). The methods then include processing the set of user-generated text based on a natural language processing engine, the natural language processing engine leveraging a clinical terminology knowledge base. This knowledge base provides rules and/or relationships to assist in the processing of known medical/clinical terms and expressions based on known word

morphology, sentence parsing, spelling correction, acronym and abbreviation resolution examples. The knowledge based is preferably updated over time (for example via a lexicon management system) thereby to enable identification of and accounting for additional clinician idiosyncratic authoring styles, and the like.

[0034] Based on the processing, a set of one or more suggested classification codes are identified. The classification codes are defined by a pre-existing clinical code structure, such as SNOMED CT, ICD, LOINC, and ICD-O3. In some embodiments multiple code structures are used, and in some cases a user is enabled to select one or more of a set of available code structures (for example via check-boxes or the like). This selection may occur following identification of codes via a user interface component, as discussed below.

[0035] Following identification of a set of one or more codes, the methods include causing display of a user interface component in the user interface environment. This may in some embodiments be a pop-up window, which is triggered by an external software component. The user interface component: (i) displays the identified set of one or more suggested classification codes; and (ii) enables a user to selectively confirm/reject one or more of the set of identified codes. For example, in a preferred embodiment the user interface component displays, for each identified code, some or all of the following aspects of information:

- A section of the code structure to which an identified code belongs.
- A disease class associated with an identified code.
- A disease stage associated with an identified code. For example, in one embodiment the natural language processing engine is configured to make inferences about the data, for example; inferring a stage of cancer diagnosis.
- A disease behaviour associated with an identified code.
- A code-structure-specific description associated with an identified code.
- A portion of text that triggered the identification of an identified code.



- A portion of contextual text including a sub-portion of text that triggered the identification of an identified code. In some embodiments an author is enabled to modify this portion of text in the interface, optionally causing re-processing and re-population of user interface data, and trigger automated updating of the source clinical report text to account for the modification.

[0036] This content is preferably displayed in a user-friendly tabulation, which is in some embodiments able to subsequently be stored (or linked to) a file or data set defining a clinical record defined from the source text.

[0037] The methods also include receiving input representative of the one or more of the set of identified codes selectively confirmed by the user (for example by confirm/reject check boxes provided via the user interface). This triggers the providing of a signal thereby to cause association of a clinical record including the set of user-generated text with the set of identified codes selectively confirmed by the user. For example, the association may be by way of added textual information, metadata, database level relationships, and so on.

[0038] In some embodiments, the user interface component is additionally configured to provide a search interface that enables a user to identify and selectively confirm one or more further codes (which are also associated with the record). For example, an author may notice one or more code descriptors (such as particular diseases) missing, and search for those. This may, in some cases, encourage the author to modify the text, thereby to include material that had been inadvertently omitted (or to correct spelling mistakes and the like).

[0039] In some cases, in response to an event whereby a user identifies and selectively confirms one or more further codes, a trigger is initiated thereby to cause updating of the natural language processing engine and/or clinical terminology knowledge base. This may be a manual or automated process, whereby the text is analysed thereby to determine whether there is content that *should* have caused the automated identification of a manually added clinical code, this resulting in updating of the processing engine and/or clinical terminology knowledge base such that the relevant code is identified in a corresponding situation in future.

[0040] It should be appreciated that, in preferred embodiments, the textual language processing engine is configured to identify discrete portions of the text, and , identify a set of one or more suggested classification codes for each discrete portion (as opposed to processing the document as a single whole textual entity). In some embodiments association of the clinical record with the set of identified codes includes defining data that enables association of each discrete portion with its respective classification codes (such that a user is able to identify a particular sub-portion of text that resulted in a given code association).

[0041] FIG. 1 illustrates a framework according to one embodiment. In this example, various user interface objects are schematically illustrated as being rendered by an exemplary clinical record system 100. These objects are illustrated as being rendered in a common software application, however in further embodiments they are rendered via multiple different software applications (for example using plug-ins and the like).

[0042] A user authors text for inclusion in (or defining) a clinical record in object 101. Then, by way of a trigger object (such as a "submit" or "code" button), the text data is provided to a record analysis subsystem 120. Subsystem 120 includes natural language processing logic 121 (which leverages an expandable predefined clinical terminology knowledge base), and a repository of clinical code data (which includes textual description data associated with each of a plurality of clinical code identifiers for one or more clinical coding schema). A record processing engine is configured to coordinate the processing of text data retrieved from system 100, and provide data representative of identified codes for rendering in a user interface object 103.

[0043] User interface object 103 provides data representative of identified codes, in this case including, for each identified code, a code descriptor and one or more artefacts of code context data (for example retrieved from code data 123 and/or extracted from the source text from object 101). It also provides confirm/reject controls (which may take the form of checkboxes), thereby to enable user selective confirm/reject command inputs.

[0044] A button 103 provides a user with access to a search interface, which enables text-based searching of data 123 thereby to identify one or more further clinical codes manually. In some cases each interface also leverages logic 121 and engine 122 thereby to expand a search string to assist in identification of codes whilst taking into account morphology, spelling, acronyms, and many other language variables.

[0045] In the example of FIG. 1, records are stored in a database 110 via subsystem 120. However, in another example, interface 103 is configured to modify text in object 101 (thereby to append clinical codes), or otherwise modify operation of a pre-existing software component that causes recording of a clinical record based on text in object 101 (for example by metadata or database-level association).

[0046] FIG. 2 illustrates a method according to one embodiment. A user defined clinical record text at 201, and then a trigger event is caused at 202 (for example subject to a user interaction following completion of text authoring). Natural language processing of the text occurs at 203, resulting in identification of one or more predicted/suggested clinical codes at 204. The clinical codes are presented by way of a user interface at 205, which enables user review and selective confirmation. Optionally, at 206, a user accesses a search interface to manually identify and add one or more further codes. User confirmation of codes is received at 207, and the codes and record stored at 208. This may include adding the codes to the text content of a clinical record, and storing that clinical record in a database.

[0047] In some embodiments, methods and functionalities considered herein are implemented by way of a client-server framework, as illustrated in FIG. 3. In overview, a web server 302 provides a web interface 303. This web interface is accessed by the parties by way of client terminals 304. In overview, users access interface 303 over the Internet by way of client terminals 304, which in various embodiments include the likes of personal computers, PDAs, cellular telephones, gaming consoles, and other Internet enabled devices.

[0048] Server 303 includes a processor 305 coupled to a memory module 306 and a communications interface 307, such as an Internet connection, modem, Ethernet port, wireless network card, serial port, or the like. In other embodiments distributed resources are used. For example, in one embodiment server 302 includes a plurality of distributed servers having respective storage, processing and communications resources. Memory module 306 includes software instructions 308, which are executable on processor 305.

[0049] Server 302 is coupled to a database 310. In further embodiments the database leverages memory module 306.

[0050] In some embodiments web interface 303 includes a website. The term "website" should be read broadly to cover substantially any source of information accessible over the Internet or another communications network (such as WAN, LAN or WLAN) via a browser application running on a client terminal. In some embodiments, a website is a source of information made available by a server and accessible over the Internet by a web-browser application running on a client terminal. The web-browser application downloads code, such as HTML code, from the server. This code is executable through the web-browser on the client terminal for providing a graphical and often interactive representation of the website on the client terminal. By way of the web-browser application, a user of the client terminal is able to navigate between and throughout various web pages provided by the website, and access various functionalities that are provided.

[0051] Although some embodiments make use of a website/browser-based implementation, in other embodiments proprietary software methods are implemented as an alternative. For example, in such embodiments client terminals 304 maintain software instructions for a computer program product that essentially provides access to a portal via which framework 100 is accessed (for instance via an iPhone app or the like).

[0052] In general terms, each terminal 304 includes a processor 311 coupled to a memory module 313 and a communications interface 312, such as an internet connection, modem, Ethernet port, serial port, or the like. Memory module 313 includes software instructions 314, which are executable on processor 311. These software instructions allow terminal 304 to execute a software application, such as a proprietary application or web browser application and thereby render on-screen a user interface and allow communication with server 302. This user interface allows for the creation, viewing and administration of profiles, access to the internal communications interface, and various other functionalities.

[0053] Unless specifically stated otherwise, as apparent from the following discussions, it is appreciated that throughout the specification discussions utilizing terms such as "processing," "computing," "calculating," "determining", "analyzing" or the like, refer to the action and/or processes of a computer or computing system, or similar electronic computing device, that manipulate and/or transform data represented as physical, such as electronic, quantities into other data similarly represented as physical quantities.

[0054] In a similar manner, the term "processor" may refer to any device or portion of a device that processes electronic data, e.g., from registers and/or memory to transform that electronic data into other electronic data that, e.g., may be stored in registers and/or memory. A "computer" or a "computing machine" or a "computing platform" may include one or more processors.

[0055] The methodologies described herein are, in one embodiment, performable by one or more processors that accept computer-readable (also called machine-readable) code containing a set of instructions that when executed by one or more of the processors carry out at least one of the methods described herein. Any processor capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken are included. Thus, one example is a typical processing system that includes one or more processors. Each processor may include one or more of a CPU, a graphics processing unit, and a programmable DSP unit. The processing system further may include a memory subsystem including main RAM and/or a static RAM, and/or ROM. A bus subsystem may be included for communicating between the components. The processing system further may be a distributed processing system with processors coupled by a network. If the processing system requires a display, such a display may be included, e.g., a liquid crystal display (LCD) or a cathode ray tube (CRT) display. If manual data entry is required, the processing system also includes an input device such as one or more of an alphanumeric input unit such as a keyboard, a pointing control device such as a mouse, and so forth. The term memory unit as used herein, if clear from the context and unless explicitly stated otherwise, also encompasses a storage system such as a disk drive unit. The processing system in some configurations may include a sound output device, and a network interface device. The memory subsystem thus includes a computer-readable carrier medium that carries computer-readable code (e.g., software) including a set of instructions to cause performing, when executed by one or more processors, one of more of the methods described herein. Note that when the method includes several elements, e.g., several steps, no ordering of such elements is implied, unless specifically stated. The software may reside in the hard disk, or may also reside, completely or at least partially, within the RAM and/or within the processor during execution thereof by the computer system. Thus, the memory and the processor also constitute computer-readable carrier medium carrying computer-readable code.

[0056] Furthermore, a computer-readable carrier medium may form, or be included in a computer program product.

[0057] In alternative embodiments, the one or more processors operate as a standalone device or may be connected, e.g., networked to other processor(s), in a networked deployment, the one or more processors may operate in the capacity of a server or a user machine in server-user network environment, or as a peer machine in a peer-to-peer or distributed network environment. The one or more processors may form a personal computer (PC), a tablet PC, a set-top box (STB), a Personal Digital Assistant (PDA), a cellular telephone, a web appliance, a network router, switch or bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine.

[0058] Note that while diagrams only show a single processor and a single memory that carries the computer-readable code, those in the art will understand that many of the components described above are included, but not explicitly shown or described in order not to obscure the inventive aspect. For example, while only a single machine is illustrated, the term "machine" shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

[0059] Thus, one embodiment of each of the methods described herein is in the form of a computer-readable carrier medium carrying a set of instructions, e.g., a computer program that is for execution on one or more processors, e.g., one or more processors that are part of web server arrangement. Thus, as will be appreciated by those skilled in the art, embodiments of the present invention may be embodied as a method, an apparatus such as a special purpose apparatus, an apparatus such as a data processing system, or a computer-readable carrier medium, e.g., a computer program product. The computer-readable carrier medium carries computer readable code including a set of instructions that when executed on one or more processors cause the processor or processors to implement a method. Accordingly, aspects of the present invention may take the form of a method, an entirely hardware embodiment, an entirely software embodiment or an embodiment combining software and hardware aspects. Furthermore, the present invention may take the form of carrier medium (e.g., a computer program product on a computer-readable storage medium) carrying computer-readable program code embodied in the medium.

[0060] The software may further be transmitted or received over a network via a network interface device. While the carrier medium is shown in an exemplary

embodiment to be a single medium, the term "carrier medium" should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term "carrier medium" shall also be taken to include any medium that is capable of storing, encoding or carrying a set of instructions for execution by one or more of the processors and that cause the one or more processors to perform any one or more of the methodologies of the present invention. A carrier medium may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical, magnetic disks, and magneto-optical disks. Volatile media includes dynamic memory, such as main memory. Transmission media includes coaxial cables, copper wire and fiber optics, including the wires that comprise a bus subsystem. Transmission media also may also take the form of acoustic or light waves, such as those generated during radio wave and infrared data communications. For example, the term "carrier medium" shall accordingly be taken to include, but not be limited to, solid-state memories, a computer product embodied in optical and magnetic media; a medium bearing a propagated signal detectable by at least one processor of one or more processors and representing a set of instructions that, when executed, implement a method; and a transmission medium in a network bearing a propagated signal detectable by at least one processor of the one or more processors and representing the set of instructions.

[0061] It will be understood that the steps of methods discussed are performed in one embodiment by an appropriate processor (or processors) of a processing (i.e., computer) system executing instructions (computer-readable code) stored in storage. It will also be understood that the invention is not limited to any particular implementation or programming technique and that the invention may be implemented using any appropriate techniques for implementing the functionality described herein. The invention is not limited to any particular programming language or operating system.

[0062] It should be appreciated that in the above description of exemplary embodiments of the invention, various features of the invention are sometimes grouped together in a single embodiment, FIG., or description thereof for the purpose of streamlining the disclosure and aiding in the understanding of one or more of the various inventive aspects. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed invention requires more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive aspects lie in less

than all features of a single foregoing disclosed embodiment. Thus, the claims following the Detailed Description are hereby expressly incorporated into this Detailed Description, with each claim standing on its own as a separate embodiment of this invention.

[0063] Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention, and form different embodiments, as would be understood by those skilled in the art. For example, in the following claims, any of the claimed embodiments can be used in any combination.

[0064] Furthermore, some of the embodiments are described herein as a method or combination of elements of a method that can be implemented by a processor of a computer system or by other means of carrying out the function. Thus, a processor with the necessary instructions for carrying out such a method or element of a method forms a means for carrying out the method or element of a method. Furthermore, an element described herein of an apparatus embodiment is an example of a means for carrying out the function performed by the element for the purpose of carrying out the invention.

[0065] In the description provided herein, numerous specific details are set forth. However, it is understood that embodiments of the invention may be practiced without these specific details. In other instances, well-known methods, structures and techniques have not been shown in detail in order not to obscure an understanding of this description.

[0066] Similarly, it is to be noticed that the term coupled, when used in the claims, should not be interpreted as being limited to direct connections only. The terms "coupled" and "connected," along with their derivatives, may be used. It should be understood that these terms are not intended as synonyms for each other. Thus, the scope of the expression a device A coupled to a device B should not be limited to devices or systems wherein an output of device A is directly connected to an input of device B. It means that there exists a path between an output of A and an input of B which may be a path including other devices or means. "Coupled" may mean that two or more elements are either in direct physical or electrical contact, or that two or more elements are not in direct contact with each other but yet still co-operate or interact with each other.

[0067] Thus, while there has been described what are believed to be the preferred embodiments of the invention, those skilled in the art will recognize that other and further



modifications may be made thereto without departing from the spirit of the invention, and it is intended to claim all such changes and modifications as falling within the scope of the invention. For example, any formulas given above are merely representative of procedures that may be used. Functionality may be added or deleted from the block diagrams and operations may be interchanged among functional blocks. Steps may be added or deleted to methods described within the scope of the present invention.

**CLAIMS:**

1. A computer implemented method for enabling categorisation of clinical records, the method including:
  - identifying a set of user-generated text contained a user interface object in a user interface environment;
  - processing the set of user-generated text based on a natural language processing engine, wherein the natural language processing engine leverages a clinical terminology knowledge base;
  - based on the processing, identifying a set of one or more suggested classification codes, wherein the classification codes are defined by a pre-existing clinical code structure;
  - causing display of a user interface component in the user interface environment, wherein the user interface component: (i) displays the identified set of one or more suggested classification codes; and (ii) enables a user to selectively confirm/reject one or more of the set of identified codes;
  - receiving input representative of the one or more of the set of identified codes selectively confirmed by the user; and
  - providing a signal thereby to cause association of a clinical record including the set of user-generated text with the set of identified codes selectively confirmed by the user.
2. A method according to claim 1 wherein the user interface component is additionally configured to provide a search interface that enables a user to identify and selectively confirm one or more further codes, wherein the a signal causes association of a clinical record including the set of user-generated text with the set of identified codes selectively confirmed by the user and the one or more further codes.
3. A method according to claim 2 wherein, in response to an event whereby a user identifies and selectively confirms one or more further codes, a trigger is initiated thereby to cause updating of the natural language processing engine and/or clinical terminology knowledge base.

4. A method according to any preceding claim including causing rendering of a graphical output that associates each identified suggested clinical classification code with a portion of the user-generated text.
5. A method according to any preceding claim including causing rendering of a graphical output that associates each identified suggested clinical classification code with additional clinical code context.
6. A method according to any preceding claim wherein the identification of suggested clinical classification codes is configured to identify a disease stage and any other features of the disease such as behaviour and extent.
7. A method according to any preceding claim wherein the user interface component is provided by a software application separate to a further software application in which the set of user generated text is contained.
8. A method according to any preceding claim wherein the textual language processing engine is configured to identify discrete portions of the text, and , identify a set of one or more suggested classification codes for each discrete portion.
9. A method according to claim 8 wherein association of the clinical record with the set of identified codes includes defining data that enables association of each discrete portion with its respective classification codes.
10. A computer system configured to perform a method according to any one of claims 1 to 6.
11. A computer program configured to perform a method according to any one of claims 1 to 6.
12. A non-transitory carrier medium carrying computer executable code that, when executed on a processor, causes the processor to perform a method according to any one of claims 1 to 6.

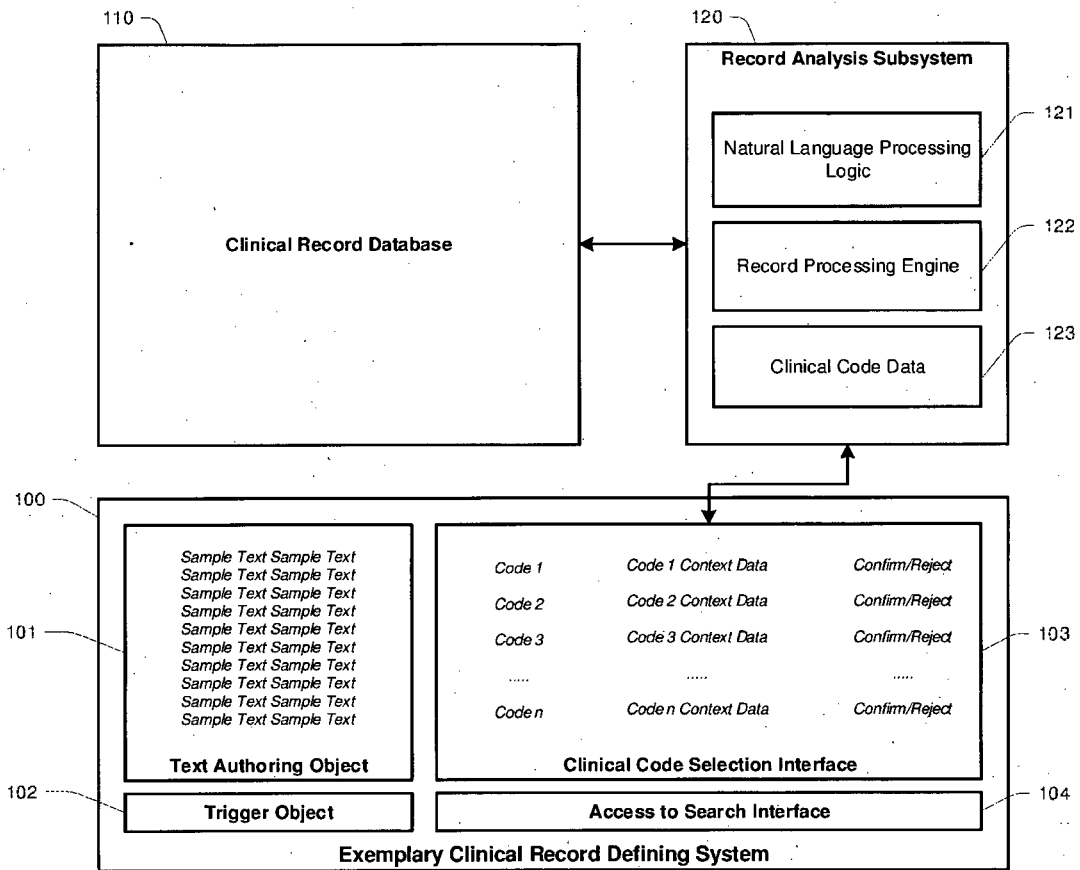
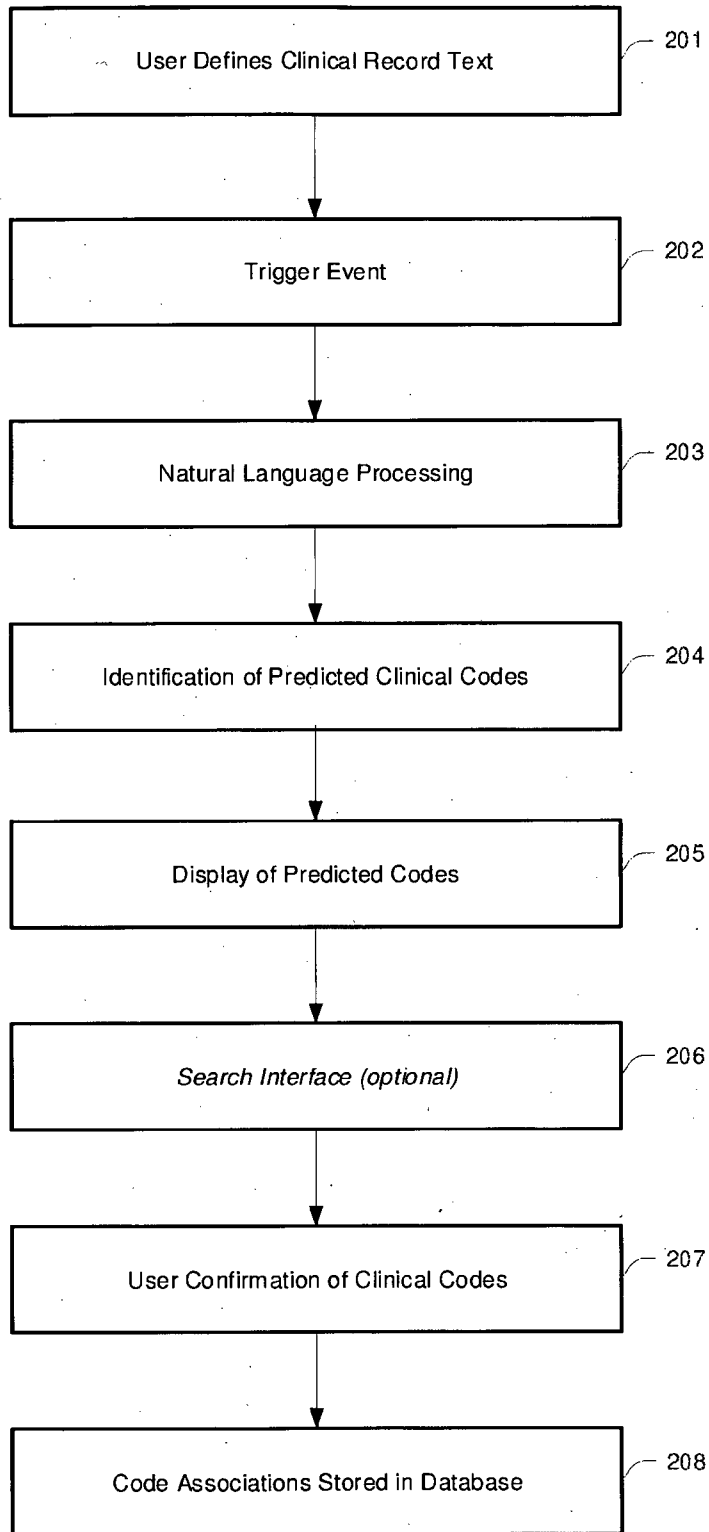
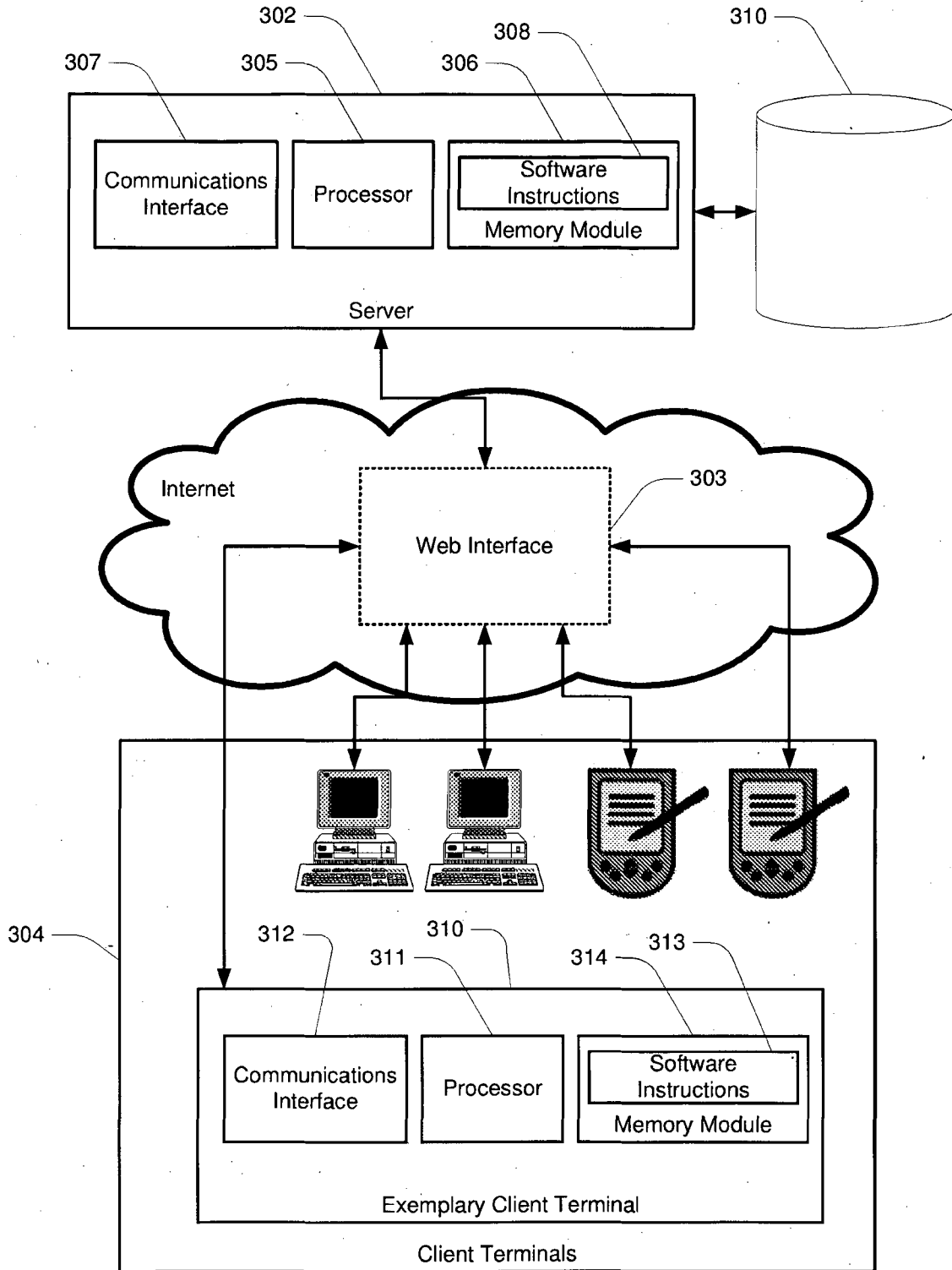


FIG. 1



**FIG. 2**



**FIG. 3**

