

DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number:	PCT/AU2016/000235
International filing date:	30 June 2016 (30.06.2016)
Document type:	Certified copy of priority document
Document details:	Country/Office: AU
	Number: 2015902542
	Filing date: 30 June 2015 (30.06.2015)
Date of receipt at the International Bureau:	05 August 2016 (05.08.2016)

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. 2015902542 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

KELLY YULE
DELEGATE OF THE COMMISSIONER



FRAMEWORKS AND METHODOLOGIES FOR ENABLING SEARCHING OF CLINICAL REPORT DATA

FIELD OF THE INVENTION

[0001] The present invention relates to frameworks and methodologies for enabling searching of clinical report data. Embodiments of the invention have been particularly developed to assist in 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] maintaining access to a database that includes data representative of clinical records, wherein each clinical record is pre-processed based upon a natural language processing engine thereby to associate the record with one or more clinical codes, wherein the clinical codes are defined in a clinical code hierarchical structure;

[0007] receiving a query from an interface that is configured enable user input of query parameters, wherein the query parameters include text-based concept-defining data;

[0008] processing the text-based concept-defining data based on a predefined protocol thereby to identify one or more clinical codes associated the concept-defining data; and

[0009] identifying one or more clinical records that are associated with the same clinical codes as the concept-defining data.

[0010] One embodiment provides a computer implemented method including, additionally associating with the concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure.

[0011] One embodiment provides a computer implemented method wherein the interface is configured to enable the user to select whether or not to perform the process of additionally associating with the concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure.

[0012] One embodiment provides a computer implemented method wherein there is a plurality of clinical code hierarchical structures.

[0013] One embodiment provides a computer implemented method wherein the interface is configured to enable the user to select one or more of the plurality of clinical code hierarchical structures for a given query.

[0014] One embodiment provides a computer program product for performing a method as described herein.

[0015] 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.

[0016] One embodiment provides a system configured for performing a method as described herein.

[0017] 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.

[0018] 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.

[0019] 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.

[0020] 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

[0021] Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

[0022] FIG. 1 schematically illustrates a framework according to one embodiment.

[0023] FIG. 2 illustrates a method according to one embodiment.

[0024] FIG. 3 illustrates an exemplary client-server framework.

[0025] FIG. 4 illustrates a search interface according to one embodiment.

DETAILED DESCRIPTION

[0026] Described herein are frameworks and methodologies for enabling searching of clinical report data. Embodiments of the invention have been particularly developed to assist in 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.

[0027] Various embodiments provide a computer implemented methods for enabling searching of clinical records. For example, this in some cases includes providing a search engine that is configured to access clinical record data provided through one or more clinical systems (or other record generation means). The methods include maintaining access to a database that includes data representative of clinical records. For example, this may be a locally or remotely hosted database, and in some embodiments includes multiple distributed databases.

[0028] Each clinical record is pre-processed based upon a natural language processing engine, thereby to associate the record with one or more clinical codes. These clinical codes are, at least in some examples, defined in a clinical code hierarchical structure. An example of a clinical code hieratical structure used in some embodiments is SNOMED CT. Other code structures include ICD, LOINC, and ICD-O3. In some embodiments there is a plurality of clinical code hierarchical structures, and a user is enabled to select one or more of these to be used for the purpose of a given query.

[0029] The natural language processing engine preferably accounts factors such as word morphology, sentence parsing, spelling correction, synonym, metonym, acronym and abbreviation resolution, document structure thereby to enable mapping of records to codes in spite of a wide range of language variations that may be observed in clinical records. The natural language processing engine is configured to be updated (and re-run over some or all of the records) over time, thereby to improve its effectiveness.

[0030] A search interface is provided, for example in a manner enabling rendering of a query-defining interface at a client terminal. Such an interface configured to enable user input of query parameters. These query parameters include text-based concept-defining data, for example "breast cancer" or the like.

[0031] A query processing engine is configured to process the text-based concept-defining data based on a predefined protocol thereby to identify one or more clinical codes associated the concept-defining data. For example, factors such as word morphology, sentence parsing, spelling correction, acronym and abbreviation resolution are handled by such a protocol, which is configured to map various instances of concept-defining data to codes, even if there is no direct textual relationship between a code descriptor and the concept-defining data itself. In some embodiments the concept-defining data is defined by a text-based set of search terms (for example a search string), which only one representation of the semantics of a concept, being a concept that is in effect searched as described herein. For example, the concept-defining data provides a single representation of a concept that has an non-enumerable set of representations

[0032] In overview, the query processing engine is configured to identify one or more clinical records that are associated with the same clinical codes as the concept represented by the instance of the concept-defining data. In this regard, there is a dual-level expansion process: firstly expansion in relation to the text of a clinical record thereby to associate it with codes based on natural language processing; and secondly expansion at the concept-defining data end to identify relevant codes. For example, a report may include Term A, which is processed and determined to match Term B associated with Code X, and a query may include Term C, which is processed and also associated with Code X. In such an example, the record containing Term A is identified based on a search for Term C, on the basis that they both match Term B of Code X. Examples of Term A, B and C might be "Lentigo Maligna", "melanoma" and "skin cancer".

[0033] In some embodiments the method additionally includes associating with the concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure. For example, the search interface is configured to enable the user to select whether or not to perform the process of additionally associating with the

concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure. An example is "breast cancer" which has the sub-type "infiltrating ductal carcinoma"

[0034] FIG. 1 illustrates a framework according to one embodiment. A plurality of clinical records defining systems 100, including an exemplary client terminal 100', define clinical record data. This data is provided to a central clinical record database 110. Systems 100 may include various software platforms used to generate clinical report content. A record analysis subsystem includes a record processing engine 132, which is configured to maintain access to database 110, and process records therein based on natural language processing logic 131. More specifically, this processing enables each record to be associated with one or more clinical codes, based on clinical code data 133. Record processing engine 132 is preferably configured to process newly added records, and to re-process existing records following updates to logic 131 and/or data 133.

[0035] A search engine subsystem 130 is configured to process queries submitted from client terminals 150, including an exemplary client terminal 150'. The client terminals in some embodiments provide search interfaces via proprietary software, or in some cases via web-browser delivered interfaces.

[0036] FIG. 2 illustrates exemplary methods 200 and 210. Method 200 is a method for natural language processing of clinical record data thereby to enable association with codes (such as SNOMED CT codes), whereas method 210 is a query processing method.

[0037] In relation to method 100, functional block 201 represents a process including identifying a clinical record for processing. For example this may be a newly submitted record, or a record queued for periodic re-processing (for example based on updated processing logic). Functional block 202 represents a process including application of natural language processing logic, which enables association of the record to codes (such as SNOMED CT codes), whilst resolving amongst other things word morphology, sentence parsing, spelling correction, acronym and abbreviation resolution. Code associations are stored in the database at 204.

[0038] In relation to method 210, functional block 211 includes receiving a query. The query is processed thereby to identify codes (such as SNOMED CT codes) at 212, again

resolving issues such as word morphology, sentence parsing, spelling correction, acronym and abbreviation resolution. This enables translation between terms used by a query submitting user, and terms actually present in code descriptors. As represented by functional block 213, is the user a selected an option for an “expended” query, then sub-tree codes are also determined. Functional block 214 represents a process including identifying those records associated with the same codes (and, where relevant, sub-tree codes).

[0039] FIG. 4 illustrates an exemplary query interface. In this example, a user inputs a concept-defining data in a “terminology code or description” field, and may also input excluded concepts. The user also selects codes to be considered (in this case being ICD-10 and/or SNOMED), and optionally selects whether a sub-tree search is to be performed. The user also selects whether “any” or “all” type search logic is to be applied. Various options for search refining/combining are also provided. In the illustrated embodiment, search results are displayed along with a window that displays content of a selected search result.

[0040] It will be appreciated that technology described herein is readily configured to assist in discovering and extracting meaningful patterns (analytics) in patient data within unstructured clinical reports. These are able to provide fast search speed and precision extraction, using combinations of multiple search algorithms and language processing techniques to maximise the coverage and reliability of results, may co-ordinate with any search engine to multiply the value of locally developed expertise, and can retrieve reports in tandem with structured retrieval queries to collect variable data types into the one location.

[0041] The use of a concept based search mechanism enables understanding of the idiosyncrasies of clinicians’ writing styles and language usage, thereby to accurately collect desired data. In some embodiments the technology is configured to extract the appropriate content for either display in a user-friendly tabulation or export to a CSV file for subsequent processing by combining a variety of facts in a report. Furthermore, it is in some embodiments set up to produce tailored reports to support each organisation’s specific needs.

[0042] In some embodiments efficacy is improved by updating natural language processing logic by tuning it to local content in a process of training it to understand the

writing vagaries of local authors. Natural language engine resolving amongst other things word morphology, sentence parsing, spelling correction, acronym and abbreviation resolution.

[0043] Concept processing may include identifying any text equivalent in a concept-defining data to ICD, LOINC, O3 and SNOMED CT codes (and in some cases proprietary dictionaries).

[0044] Extractions may include disease, disorders, procedures, with special attention to cancer details for primaries, metastases and nodes. Document analytics may be performed, including classification of documents from their contents such as for tumour stream or report purpose.

[0045] In some embodiments external reuse of extractions & inferences is provided, for example including buttons to export extracted content to CVS files for delivery to researchers or search collections.

[0046] In some embodiments workflow ease is assisted by providing a search history, batch processing and results export so as to assemble a workflow and operate it entirely automatically thus supporting periodic reporting.

[0047] In some embodiments a mechanism is provided to capture content that is not processed initially can be used to lodge that content into the knowledge repository so it is not missed again.

[0048] Some embodiments provide automatic despatch of files to registered users when a record of interest is created in the system.

[0049] Embodiments may, by way of example, be applied in the following environments:

- Patient Studies (for example to recruit a common cohort or phenotypes for clinical trials; to target particular disease classes for research; to alert staff when a pertinent record is created in the storage system).

- Auditing Clinical Records (for example to check the accuracy of clinical coding across disease classes; to semi-automate codification for billing; to identify and understand the case mix of a particular organisation).
- Research (for example to answer ad-hoc questions about the distributions of disease morbidities).
- Assessing Training & Report Writing (for example to investigate report reliability and readability; to provide feedback to staff on the nature of their written composition; to tutor and evaluate the work of trainees; to compile activity reports for professional associations for registration; to investigate more effective systematic methods for preparing reports).

[0050] In some embodiments, methods and functionalities considered herein are implemented by way of a server, 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.

[0051] 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.

[0052] Server 302 is coupled to a database 310. In further embodiments the database leverages memory module 306.

[0053] 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.

[0054] 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).

[0055] 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.

[0056] 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.

[0057] 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.

[0058] 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.

[0059] Furthermore, a computer-readable carrier medium may form, or be included in a computer program product.

[0060] 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.

[0061] 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.

[0062] 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.

[0063] 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 included, 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] 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 searching of clinical records the method including:

maintaining access to a database that includes data representative of clinical records, wherein each clinical record is pre-processed based upon a natural language processing engine thereby to associate the record with one or more clinical codes, wherein the clinical codes are defined in a clinical code hierarchical structure;

receiving a query from an interface that is configured enable user input of query parameters, wherein the query parameters include text-based concept-defining data;

processing the text-based concept-defining data based on a predefined protocol thereby to identify one or more clinical codes associated with the concept-defining data; and

identifying one or more clinical records that are associated with the same clinical codes as the concept-defining data.

2. A method according to claim 1 including additionally associating with the concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure.
3. A method according to claim 2 wherein the interface is configured to enable the user to select whether or not to perform the process of additionally associating with the concept-defining data one or more further clinical codes, wherein the one or more further clinical codes are sub-tree codes to the identified one or more clinical codes in the hierarchical structure.
4. A method according to any preceding claim wherein there is a plurality of clinical code hierarchical structures.

5. A method according to claim 4 wherein the interface is configured to enable the user to select one or more of the plurality of clinical code hierarchical structures for a given query.
6. A method according to any preceding claim wherein the concept-defining data provides a single representation of a concept that has a non-enumerable set of representations
7. A computer system configured to perform a method according to any one of claims 1 to 6.
8. A computer program configured to perform a method according to any one of claims 1 to 6.
9. 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.

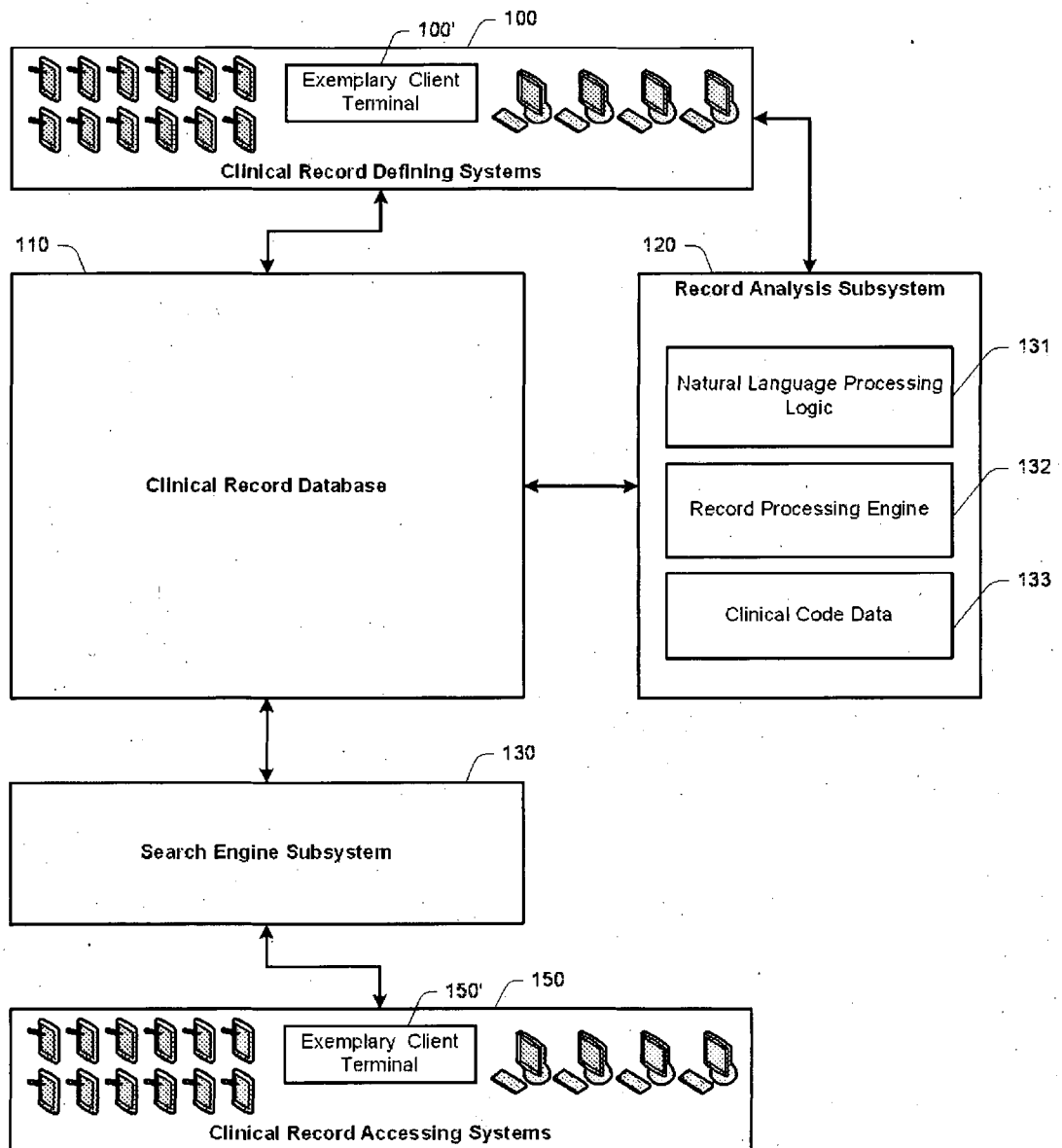


FIG. 1

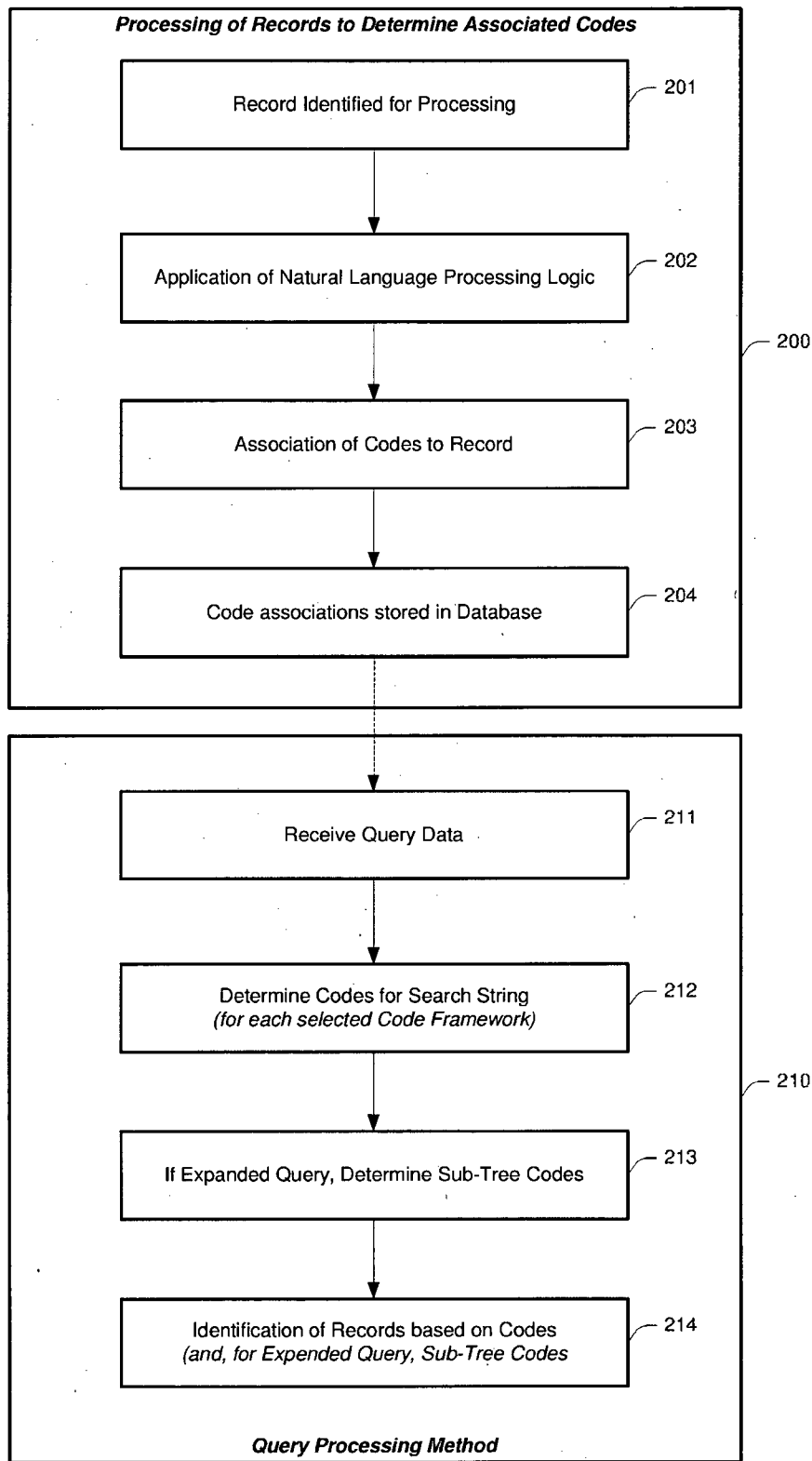


FIG. 2

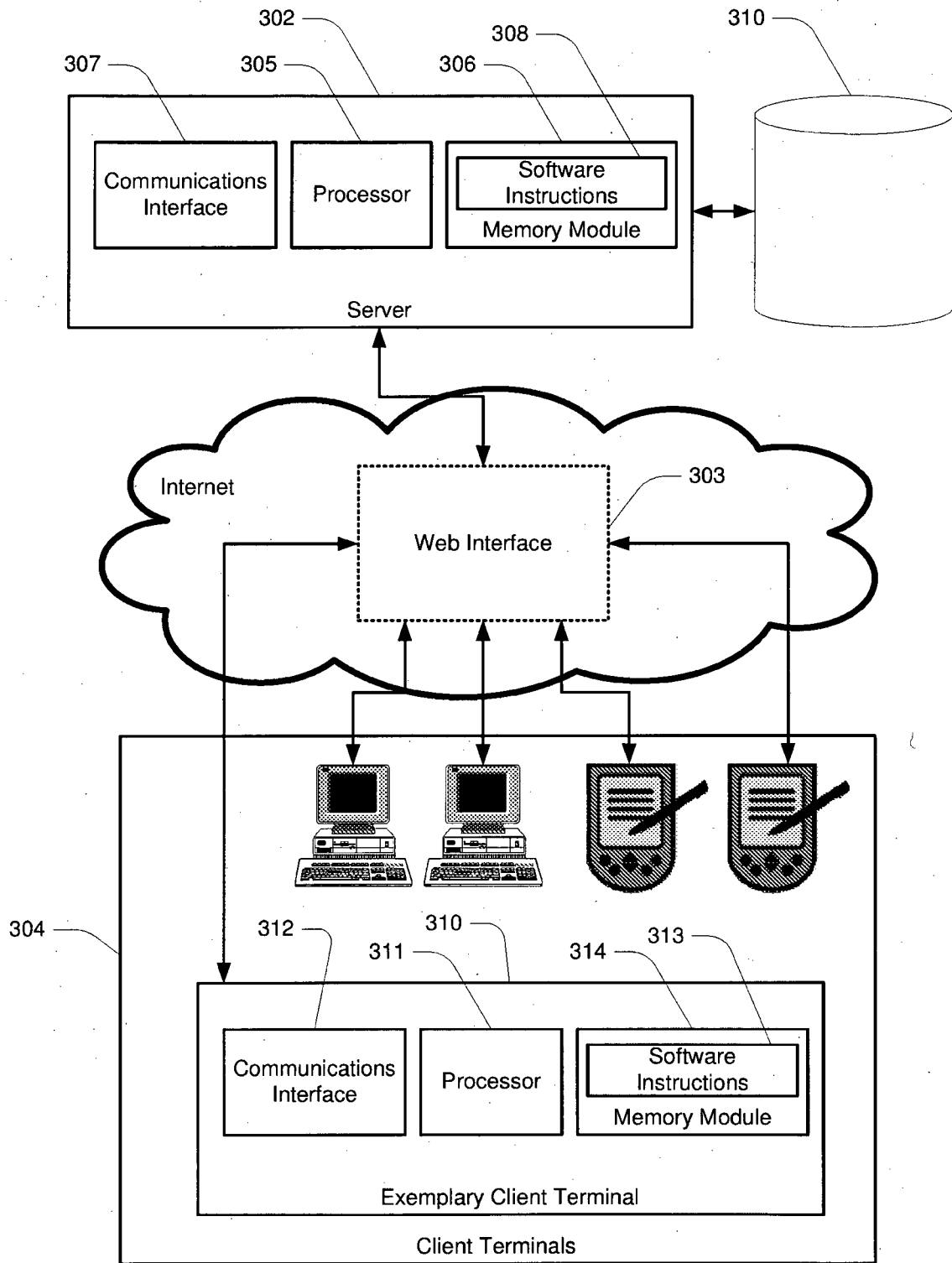


FIG. 3

The screenshot displays the CitiusSearch web application interface. At the top, there are navigation tabs for 'Keywords / Phrases', 'Analytics Engine', and 'CITIUSSEARCH'. The main content area is divided into several sections:

- Search Results:** A table with columns 'File Name' and 'Size'. It lists four results with IDs 10, 11, 12, and 13, each with a size of 1. The descriptions include clinical indicators for breast cancer, such as 'erb ERBB2 with Patient Clinical indicators: invasive ductal carcinoma left breast at 2-3 o'clock. 2nd lesion adjacent' and 'p53 ERBB2R Clinical indicators: Night breast inflammatory carcinoma, left axillary nodal carcinoma. CT - CHEST'.
- Context of Sample 10:** A text box providing details for the first result, including 'erb ERBB2 with Patient Clinical indicators: invasive ductal carcinoma left breast at 2-3 o'clock. 2nd lesion adjacent'.
- Search Tools:** A sidebar on the right contains 'Lexicon Annotations' (with sub-items 'abbreviation' and 'acronym'), 'Analysis for Sentences', and 'Export Page Size' (with options 'Less All Results' and 'Less More Results').
- Footer:** A 'Search History' section at the bottom shows a list of recent searches and their associated dates.

FIG. 4