

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 January 2010 (07.01.2010)

(10) International Publication Number
WO 2010/000014 A1

- (51) **International Patent Classification:**
G06Q 50/00 (2006.01) G06F 17/30 (2006.01)
G06F 17/20 (2006.01)
- (21) **International Application Number:**
PCT/AU2009/000134
- (22) **International Filing Date:**
6 February 2009 (06.02.2009)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
2008903403 2 July 2008 (02.07.2008) AU
- (71) **Applicant (for all designated States except US):** PACIFIC KNOWLEDGE SYSTEMS PTY. LTD. [AU/AU]; Suite 111, Bay 16, Australian Technology Park, Eveleigh, New South Wales 1430 (AU).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** PETERS, Lindsay [AU/AU]; 46 Epping Avenue, Eastwood, New South Wales 2122 (AU). LAVERS, Timothy [AU/AU]; 7 Liddle Street, Woonona, New South Wales 2517 (AU).
- (74) **Agent:** GRIFFITH HACK; Level 29, Northpoint, 100 Miller Street, North Sydney, New South Wales 2060 (AU).

(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, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

(54) **Title:** METHOD AND SYSTEM FOR GENERATING TEXT

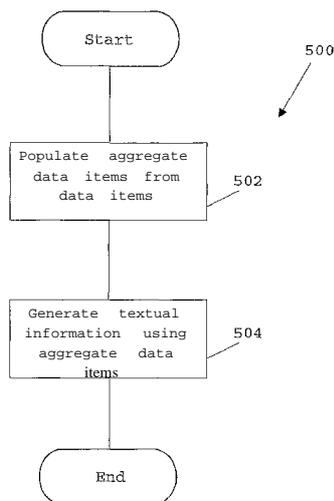


Figure 2

(57) **Abstract:** A method of generating information from a plurality of data items, the method comprising the steps of populating an aggregate data item with at least one of the plurality of data items (502) and generating the information using the aggregate data item (504).

WO 2010/000014 A1

METHOD AND SYSTEM FOR GENERATING TEXTField of the Invention

5 The present invention generally relates to a method and system for generating text, and particularly but not exclusively to a method and system for generating syntactically correct text for a test report.

10 Background of the Invention

 Complicated investigations commonly performed by professionals, such as medical practitioners, often require a large number of tests. The interpretation of
15 the test results is often difficult and requires the skill of an expert or expert system. The expert or expert system will generate text for inclusion in a report containing a useful analysis and interpretation of the test results, sometimes in a highly condensed form, to be
20 forwarded to the professional. To date, the Knowledge Bases of expert systems have been built in domains in which tests are relatively independent of each other. For example, a thyroid reporting Knowledge Base principally considers TSH, FT3 and FT4 results. Other patient
25 demographic data such as age and sex also generally needs to be taken into account. Reports generated by these Knowledge Bases refer to these individual tests and their values, as well as providing a diagnosis. Typically in these domains, there are less than 20 tests to consider,
30 plus patient demographic data like age and sex, plus clinical notes provided by the medical practitioner. These test typically do not interact.

 Specific rules are written for each type of
35 investigation. For example, for a thyroid panel of tests, the comment may be generated "Consistent with primary hypothyroidism" if the TSH test result is elevated.

 Traditional clinical domains have just a few
40 attributes, however, newer clinical domains with potentially hundreds or even thousands of possible investigations, the application of specific rules to each type of investigation becomes infeasible. For example, the medical practitioner may request a number of food
45 allergy tests such as peanut, soya, milk, wheat and egg. If soya and milk return very high positive values 25.3 and 30.1, and the other tests are negative, the pathologist will want the report sent back to the doctor to include a comment like:

- 2 -

"Very high results were detected for milk (30.1) and soya (24.3) ."

The rule giving this comment is:

5 10 <= milk <= 50, indicating a very high result,
 and
 10 <= soya <= 50, indicating another very high
 result, and
10 milk > soya, indicating that the milk value
 should be before the soya in the report, and
 peanut = 0, and
 wheat = 0, and
 egg = 0

15 In this simple example with just 5 allergens tested,
the number of combinations of the above comment is $2^5 = 32$
(when neglecting order of importance) . Corresponding to
each combination of test results there is a different
rule .

20 It is clearly not practical to separately define each
of the 32 possible combinations of this comment and
corresponding rules even for this simple comment - and
real-world examples are far more complex than this.

25 In the case of an allergy Knowledge Base there are
literally hundreds of possible tests that can be performed
in an investigation, each measuring the same chemical
(IgE) , with the value of each test indicating the
30 patient's response to a particular allergen.

 In cases where there are hundreds of tests in an
investigation it is impossible for an expert to define all
the possible interactions between the test results and
35 provide the multitude of comment variations that an
accurate report would require. In these cases,
computerised expert systems containing thousands of rules
would be essential, however, computational complexity of
generating a report that takes into account potentially
40 hundreds of associated tests is beyond the capability of
traditional expert systems. For example, if there were
four hundred tests and each test had only a binary output,
such as "positive" or "negative" , then there would be 2^{400}
possible combinations of test results,- each combination
45 requiring an unique reporting text previously generated
and stored on a computer system. This does not even begin
to account for possible interactions between the tests
which greatly complicates the situation. This is not
feasible when there are hundreds or more test results.

The variety of cases and their corresponding reports even with a few attributes can be huge, and even more so when the patient's historical information and clinical notes are also taken into account, and where the test results
5 have continuous values, rather than binary values.

Summary of Invention

According to a first aspect of the invention there is
10 provided a method of generating information from a plurality of data items, the method comprising the steps of:

populating an aggregate data item with at least one of the plurality of data items; and
15 generating the information using the aggregate data item.

In an embodiment, the method comprises the prestep of
20 populating each of a plurality of elements of a predefined structure with a corresponding one of the data values, the structure relating each of the plurality of elements to the aggregate data item. The method may comprise
populating the aggregate data item with at least one of
the plurality of data items in accordance with the
25 structure.

In an embodiment, the method comprises working
through the structure to determine one or more
characteristics of the aggregate data item.
30

In an embodiment, the structure relates a plurality of aggregate data items, and the method comprises working
through the structure to determine one or more
characteristics of each of the aggregate data items.
35

In an embodiment, the information comprises textual information. The information may be clinical decision support information. Alternatively, the information
40 comprises a machine instruction.

In an embodiment, the step of generating the information comprises the step of generating information using a Knowledge Base or Decision Support System.

45 In an embodiment, the step of populating the aggregate data item comprises a step of receiving the plurality of data items.

In an embodiment, the textual information may be

syntactically and/or grammatically correct . The textual information may be human readable.

5 In an embodiment, the textual information may form at least part of a report . The report may be associated with one or more test results.

10 In an embodiment, each data item may correspond to one of the test results. The test may comprise any one of an allergy test, a leukemia test, a pathology test, a blood test, and any other type of medical test. A data item may correspond to any other information such as gender, age, demographic information or clinical symptoms.

15 In an embodiment, the aggregate data item comprises data items which are related.

20 In an embodiment, the step of populating the aggregate data item comprises the step of populating the aggregate data item by applying a rule to at least one of the plurality of data items, or another aggregate data item. The rule may be a domain specific rule. The rule may alternatively be a case specific rule. The rule may be to populate the aggregate data item with one or more of the data items. The rule may be to populate the aggregate data item with one or more of the data items that are above a threshold value. The rule may be to populate the aggregate data item with one or more of the data items that are below a threshold value .

30

In an embodiment, the method comprises:

- (a) populating one or more aggregate data items with at least one of the plurality of data items;
- 35 (b) populating one or more further aggregate data items with data items from the one or more aggregate data items by applying one or more rules to the one or more aggregate data items; and
- 40 (c) generating the information using the one or more further aggregate data items.

45 In an embodiment, each of the plurality of data items are associated with an identifier and a value. Each of the plurality of data items may comprise the identifier and the value. The identifier may be associated with a name or label for the data item.

In an embodiment the step of generating information comprises the step of including in the information the

names or labels of the data items populating the aggregate data item. The step of generating the information may comprise the step of including in the information the values associated with the data items populating the aggregate data item. The step of generating the information may comprise the step of determining the order of the names or labels in the information.

In an embodiment, the step of generating information comprises determining one or more characteristics of the aggregate data item. The step of determining characteristics may comprise one or more of determining the number of data items comprising the aggregate data item, if the aggregate data item is empty, and if the aggregate data item comprises a specific data item, if an aggregate data item does not contain a specific data item, and if aggregate data items share data items. One of the characteristics may be a value.

In an embodiment, the step of generating the information comprises the step of including in the information a determined characteristic of the aggregate data item.

In an embodiment, the step of populating the aggregate data item may comprise populating the aggregate data item with one or more other aggregate data items. The aggregate data items and each of the one or more aggregate data items may be associated with an aggregate identifier. The aggregate identifiers may each be associated with aggregate names. The step of generating the information may comprise including the aggregate names. The step of including the aggregate names in the information may include the step of determining the order of the aggregate names in the information.

In an embodiment, the step of populating the aggregate data item comprises populating the aggregate data item includes the step of operating on two other aggregate data items. The step of operating may comprise one or more of difference, union and intersection of the two other aggregate data items.

In an embodiment, the step of populating an aggregate data item comprises the step of determining which data items comprising another aggregate data item have values in a particular range or from a particular discrete set.

In an embodiment, the step of generating information comprise applying one or more rules to the aggregate data item. The one or more rules may form at least part of a ripple down rules knowledge system.

5

According to a second aspect of the invention there is provided a system for generating information from a plurality of data items, the system comprising:

10 an aggregate data item populator for populating an aggregate data item with at least one of the plurality of data items; and

an information generator for generating the information using the aggregate data item.

15 In an embodiment, the information generator is a textual information generator for generating textual information. Alternatively, the information generator is a machine instruction generator for generating a machine instruction.

20

In an embodiment, the system comprising a Knowledge Base or Decision Support System.

25 In an embodiment, the system comprises a data item receiver for receiving the plurality of data items.

The information generator may be arranged to generate syntactically and/or grammatically correct textual information. The information generator may be arranged to 30 generate human readable textual information. The information generator may be arranged to generate coded textual information. The coded textual information may be a machine instruction.

35 In an embodiment, the information generator may be arranged to generate textual information forming at least part of a report .

40 In an embodiment, the aggregate data populator may be arranged to populate the aggregate data item by applying a rule to at least one of the plurality of data items.

In an embodiment, the information generator is arranged to include in the information a name or label 45 associated with a data item populating the aggregate data item. The information generator may be arranged to include in the information the value associated with a data item populating the aggregate data item. The information generator may be arranged to determine the

order of the names or labels in the text.

5 In an embodiment, the information generator is arranged to determine the characteristics of the aggregate data item. The information generator may be arranged to include one or more of determining the number of data items comprising the aggregate data item, if the aggregate data item is empty, and if the aggregate data item includes a specific data item.

10

In an embodiment, the information generator is arranged to include in the information about the determined characteristics of the aggregate data item.

15

In an embodiment, the aggregate data populator is arranged to populate the aggregate data item with one or more other aggregate data items. The aggregate data populator may be arranged to include in the text a aggregate data item name associated with an aggregate data item. The aggregate data populator may be arranged to determine the order of the aggregate data item names in the text.

20

25 In an embodiment, the aggregate data item populator is arranged to operate on two other aggregate data items.

In an embodiment, the aggregate data item populator is arranged to determine which data items comprising another aggregate data item have values in a particular range.

30

In an embodiment, the information generator is arranged to apply one or more rules to the aggregate data item.

35

In an embodiment, the information generator is arranged to consider local characteristics where the test was performed.

40

According to a third aspect of the invention, there is provided a method of generating information from a plurality of data items, the method comprising the steps of:

45 evaluating an outcome of one or more rules using one or more aggregate data items each comprising one or more of the data items; and
generating the information according to the outcome.

In an embodiment, the information comprises textual

information. Alternatively, the information comprises a machine instruction.

5 In an embodiment, the step of generating the information comprises the step of generating information using a Knowledge Base or Decision Support System.

10 In an embodiment, the step of evaluating the outcome of one or more rules comprises using characteristics of an aggregate data item as a basis for one or more of the rules .

15 According to a fourth aspect of the invention, there is provided a system for generating information from a plurality of data items, the system comprising:
an evaluator for evaluating an outcome of one or more rules using one or more aggregate data items each comprising one or more of the data items; and
20 an information generator for generating the information according to the outcome.

25 In an embodiment, the information generator is a textual information generator for generating textual information. Alternatively, the information generator is a machine instruction generator for generating a machine instruction .

30 In an embodiment, the system comprising a Knowledge Base or Decision Support System.

35 According to a fifth aspect of the invention, there is provided a method of generating information, the method comprising the steps of:
receiving a conceptual representation of information including an interpretive portion, the interpretive portion representing an operation on an aggregate data item comprising a plurality of data items; and
40 generating the information from the interpretive portion.

45 In an embodiment, the information is textual information. Alternatively, the information is a machine instruction.

In an embodiment, the step of generating the information includes the step of including in the information one or more names or labels associated with each of the data items. The step of generating the information may include the step of including in the

information a collective name for the plurality of the data items. The step of including the one or more names or labels may include the step of integrating the information with a literal portion of the conceptual
5 representation of the text.

In an embodiment, the conceptual representation is pseudo text.

10 In an embodiment, the step of generating the information may include the step of generating syntactically and/or grammatically correct textual information.

15 According to a sixth aspect of the invention, there is provided a system for generating information, the system comprising the steps of:

a receiver for receiving a conceptual representation of information including an interpretive portion, the
20 interpretive portion representing an operation on an aggregate data item comprising a plurality of data items; and

an information generator for generating the information from the interpretive portion.

25 In an embodiment, the information is textual information, in which case the information generator is a textual information generator. Alternatively, the information is a machine instruction, in which case the
30 information generator is a machine instruction generator.

In an embodiment, the generator is arranged to include in the information one or more names or labels associated with each of the data items. The generator may
35 be arranged to include in the information a collective name for the plurality of the data items. The generator may be arranged to integrate the information with a literal portion of the conceptual representation of the text.

40 In an embodiment, the generator may be arranged to generate syntactically and/or grammatically correct textual information.

45 In accordance with a seventh aspect, the present invention provides a computer program comprising instructions for controlling a computer to implement a method in accordance with the first aspect of the invention.

In accordance with an eighth aspect, the present invention provides a computer readable medium providing a computer program in accordance with the seventh aspect of the invention.

In accordance with a ninth aspect, the present invention provides a computer program comprising instructions for controlling a computer to implement a method in accordance with the third aspect of the invention.

In accordance with a tenth aspect, the present invention provides a computer readable medium providing a computer program in accordance with the ninth aspect of the invention.

In accordance with a eleventh aspect, the present invention provides a computer program comprising instructions for controlling a computer to implement a method in accordance with the fifth aspect of the invention.

In accordance with a twelfth aspect, the present invention provides a computer readable medium providing a computer program in accordance with the eleventh aspect of the invention.

The term "server" in this specification is intended to encompass any combination of hardware and software that performs services for connected clients in part of a client-server architecture. The client and a server may be separate software running on a single piece of hardware or a plurality of connected pieces of hardware.

Brief description of the Figures

In order to achieve a better understanding of the nature of the present invention, embodiments of a method and system for generating textual information will now be described, by way of example only, with reference to the accompanying figures in which:

Figure 1 is a block diagram of one embodiment of a system for generating textual information;

Figure 2 is a flow diagram of one embodiment of a method of generating textual information;

Figure 3 is a block diagram of another embodiment of a system for generating textual information;

Figure 4 is a flow diagram of another method of generating textual information;

Figure 5 is a flow diagram of another method of generating textual information; and

5 Figure 6 is a schematic representation of one embodiment of a hierarchical relationship for data items and aggregate data items.

Detailed Description of embodiments of the invention

10

In an embodiment of the invention a device receives a plurality of data items, each data item corresponding to the result of one of a plurality of tests. Typically, the plurality of tests are used in an investigation of a patient's conditions, such as does the patient have a particular form of disease or allergy. In this embodiment, the device includes information in the form of a predefined structure relating the types of data items and aggregate data items, allowing the device to populate the predefined aggregate data items with one or more of the received data items by applying various rules which processes the received data. In this embodiment, each data item in one of the aggregate data items are relevant biomarkers for a particular disease or allergy, for example.

25

Further aggregate data items may then be populated by other rules acting on the aggregate data items. The further aggregate data items may, for example, include data items that have a significant value. Further rules are then applied to the further aggregate data items. An example rule may include determining whether the number of significant data items in a further aggregate data item exceed a threshold value. The outcome of the rule may indicate a positive test result, in which case appropriate text reporting a positive, or otherwise, test result is generated. The text may be generated in a flexible case-by-case basis, without requiring a rule for each case, through use of aggregate data items.

40

Figure 1 is a block diagram of one embodiment of a system for generating textual information from a plurality of data items and is generally indicated by the numeral 1. The system 1 may comprise any system able to process information, and in this embodiment may be described as a computer system 1 including a computer program residing on computer readable medium 2 comprising instructions for controlling a central processor 4 of the system, the instructions being to implement a method 500 for

45

generating textual information from a plurality of data items. A flow diagram of the method 500 is shown in figure 2.

5 In an alternative embodiment, one or more machine instructions rather than textual information are generated, and the components of the system 1 are modified accordingly. It is to be understood that the term "textual information" is to be read more broadly to
10 encompass this alternative embodiment where appropriate hereafter.

The computer readable medium 2 includes a non volatile memory 2 in the form of a hard drive disk 2
15 connected to the processor 4 by a suitable bus 6 such as SCSI. In some embodiments the non volatile memory includes FLASH memory, a CD, DVD, or a USB Flash memory unit, for example.

20 In some embodiments such as that shown in figure 3, the system 3 is an embedded system. The components in figure 3 similar to those of the system 1 of figure 1 are similarly numbered. The embedded system in this embodiment forms part of an instrument 5 for carrying out
25 a test such as a medical test. It will be appreciated that any suitable architecture, such as terminal/mainframe, could be used and not only those illustrated.

30 In the embodiments shown in figures 1 and 3 the hard drive 2 holds one or more data items 8 representing the results from one or more test from an investigation or case. The system 1 is arranged to process the data items 8. The data items 8 may be associated with a medical test
35 such as an allergy test, a leukemia test, a pathology test, a blood test, or any other type of test. The hard drive 2 may alternatively hold data items relating to any expert domain such as one or more of fraud detection, real estate validation, bone mineral density reporting, medical
40 alerts, genomic, molecular, and allergy reporting, allergy reporting and the systems 1,3 and methods 500 described here may be arranged to process such data 8.

In the embodiment of figure 1, the system 1 has a
45 data receiver 10 for receiving the data items 8 subsequently stored on the hard drive 2. In an embodiment where the tests have been done remotely from the system 1, for example at a remote site 12, the system 1 may be arranged for connecting to a network 14 to which the

- 13 -

remote site 12 is also connected. The network 14 may be a wide area network such as the internet, although it will be appreciated that the remote site 12 may be far closer, for example, a room adjacent the system 1 in which case
5 the network 14 may be a local area or wireless network such as WiFi or WLAN. Alternatively, in cases as shown in figure 3 where the system 3 is part of a test instrument 5, the data receiver 10 may act as an interface between the processor 4 and the data source 22 such as a sample
10 testing apparatus of the system 3 that performs the physical, chemical or biological test on a sample or other analysis.

The processor 4 is programmed as an aggregate data
15 item populator for populating an aggregate data item 24 with at least one of the plurality of data items 8 stored on the hard drive 2. The aggregate data item 24 is in one embodiment, a type of data structure of a memory 20 for processing by the processor 4. The data items 8 may also
20 be stored in the memory 20. The memory 20 in this embodiment comprises one or more of CPU registers, on-die SRAM caches, external caches, DRAM and/or, paging systems, virtual memory or swap space on the hard drive 2, or any other type of memory. However, embodiments may have
25 additional or less memory types as suitable.

The processor 4 is programmed to be a textual information generator for generating the textual information using the aggregate data item 24. The textual
30 information generator 4 is arranged to store the textual information 26 in the memory 20. The textual information 26, in this embodiment, represents human readable text that is syntactically and/or grammatically correct. The output of the system 1,3 is the textual information,
35 preferably in a human readable form such as one or more of text printed to a monitor or screen 28, text printed by a printer 30 onto a paper report 32, and an email or other type of electronic message 34 sent over the network 14 to a physician or surgeon's computer 32, for example. The
40 textual information may be some other decision support outcome derived from the data items 8. In one embodiment, a SMS gateway 34 is instructed by the system 1 to send an electronic message, such as an SMS or email, including the textual information 26 in human readable form to a
45 handheld electronic device 36. The device 36 may be a mobile telephone or Blackberry. In this embodiment, the system 1 is arranged to send instructions to send an SMS to a handheld mobile device 36. This is advantageous when a test result is abnormal and requires immediate follow

up .

The processor 4 acting as the aggregate data populater is programmed to populate the aggregate data item 24 by applying one or more rules to at least one of the plurality of data items. The rules may form at least part of a ripple down rules knowledge system, as disclosed in the specification of the applicant's US patent 6,553,361 which is incorporated herein by way of reference. The collection of rules is a Knowledge Base that is built up by an expert as described in the US specification. The rules may be domain specific. For example, the rules may be specific to the domain of allergy testing, or the domain of leukemia testing. In some other instances, however, the rule is a rule specific to the case, that is a rule specific to a set of related test results / data items 8. In this case, the system 1 is a Knowledge Base or Decision Support System.

In one case, the data items 8 stored on the hard drive 2 have associated name or label parts and value parts as follows:

milk, 25;
soya, 30; and
peanut, 0 .

Each of the data items milk, soya and peanut are associated with an identifier (milk, soya and peanut) and a value (25, 30 and 0). In this embodiment, each of the data items comprise the identifier and the value. The identifier is, in this example, a name or label for the data item that can be used for generating the textual information 26. An aggregate data item 24 having a name or label *very high food allergens* may be populated from the above data items 8 by a rule such as:

If milk > 25 then include milk in *very high food allergens* AND If soya > 25 then include soya in *very high food allergens* AND If peanut > 25 then include soya in *very high food allergens* .

Alternatively, an aggregate data item 24 having a name or label *very high food allergens* may be populated from the above data items 8 by a rule such as:

very high food allergens is food allergens in range (25, 100)

The processor 4 is also programmed as an evaluator for evaluating the outcome of the one or more rules, as exemplified above, using one or more aggregate data items, such as 24. The textual information generator 4, in the above example, generated textual information for the report 32, for example, according to the outcome of the

rules .

It will be appreciated that the processor 4 may test each data item 8 in turn for inclusion in the aggregate data item 24 .

In this case, one representation of *very high food allergens* would be:
milk, 25; and
soya, 30.

The textual information generator 4 may be arranged to include in the textual information 26 the name or label associated with a data item 8 populating the aggregate data item 26. For example, the processor 4 may be asked to form the textual information:

Very high results were found for *very high food allergens* .

The processor 4 would generate textual information representing the text :

Very high results were found for soya and milk.

The textual information generator 4 has determined that soya has a higher value than milk and thus the best way to present this text is to order the names or labels in the text so that soya is superior. Also, the generator 4 has determined that an *and* should be placed between *soya* and *milk* because there are only two items in this aggregate data item 24. If there was a third item in the aggregate data item 24, such as honey with a value of 26, then the generator 4 would know that one grammatically correct text to generate would be :

Very high results were found for soya, honey and milk.

The textual information generator is arranged, as required, to include in the textual information 26 the value associated with a data item populating the aggregate data item. For example, the above text may instead be

Very high results were found for soya (30) , honey (26) and milk (25) .

The above are examples of one commonly required ordering, but there may be others in different circumstances .

In some embodiments, the information generated does not generate textual information, but rather one or more

machine instructions. In this case, the system includes a machine instruction generator. The machine instruction can control workflow. For example, if the test results show that no allergens were detected, then the machine instruction may cause the system to automatically send a report without it being checked by a human evaluator, for example. Alternatively, the machine instruction may cause additional tests to be carried out on held samples before the report is generated.

The system 1,3 may include a receiver 36 for receiving a conceptual representation of a text. The receiver 36 in this embodiment includes means such as a keyboard connected to the CPU for entry of the conceptual representation of the text by an operator 36. Using the above example, the conceptual representation entered by the operator 38 is in the form of pseudo text:

Very high results were found for very high food allergens .

The pseudotext in this example is compact, informal and represents a high-level description of the text desired by the operator 38, but importantly omits details intended for the system 1,3 to evaluate. It is a natural language description of the computational details. Pseudotext is easier for humans to formulate and read than a more technical description of the desired text which may be achieved using programming or scripting languages.

The conceptual representation includes an interpretive portion, which in this case is:
very high food allergens

The interpretive portion represents an operation on the aggregate data item with the name *very high food allergens*. The textual information generator 4 generates the textual information 26 from the interpretive portion as described elsewhere in this document. The generator 4 is arranged to include in the textual information 26 one or more names or labels associated with each of the data items 8. The generator 4 may be arranged to include in the textual information 26 a collective name for the plurality of the data items. The generator may be arranged to integrate the textual information 26 with a literal portion of the conceptual representation of the text, which in this case is:

Very high results were found for soya, honey and milk.

In the embodiments shown in figures 1 and 3, the textual information generator 4 is arranged to determine the characteristics of the aggregate data item 24. For example, the textual information generator may be arranged to include one or more of determining the number of data items comprising the aggregate data item, if the aggregate data item is empty, and if the aggregate data item includes a specific data item. These are examples of operations on the aggregate data item. For example, textual information 26 is generated from pseudotext such as:

Very high results were found for *number of very high food allergens* food allergens.
which becomes :
Very high results were found for 3 food allergens.

Thus, the textual information generator 4 is arranged to include in the textual information 26 information about the determined characteristics of the aggregate data item. *Number of* is a type of operation acting on the aggregate data item *very high food allergens*.

The aggregate data populator 4 may be arranged to populate the aggregate data item with one or more other aggregate data items. The aggregate data item may comprise data items which are related, for example, all foods to which a patient is found to be highly allergic to. The aggregate data populator may be arranged to include in the text a aggregate data item name associated with an aggregate data item. The aggregate data populator may be arranged to determine the order of the aggregate data item names in the text.

In an embodiment, the aggregate data item populator 4 is arranged to operate on two other aggregate data items. For example, one aggregate data item may be the very high result food allergens, and the other may be food allergens of interest. The populator 4 may then generate a new aggregate data item, for example the very high result food allergens of interest, by taking the intersection of the two aggregate data items. Other possible operators include difference, union and intersection. In another embodiment, the aggregate data item populator 4 is arranged to determine which data items comprising another aggregate data item have values in a particular range.

In an embodiment, the step of generating the textual information 26 comprises the step of including in the

textual information 26 information about the determined characteristics of the aggregate data item 24.

In some embodiments, the textual information generator 4 is arranged to apply one or more rules to the aggregate data items 24 to control program flow. An example logical test associated with such a rule is:
If number of Moderate foods > 1 AND if number of Symptoms > 1 AND number of very high foods + number of foods = 0

It will be appreciated that aggregate data items can in turn be treated as data items for generating textual information 26 and evaluating rules. Populating the aggregate data 24 item may include populating the aggregate data item 24 with one or more other aggregate data items, each of which may have an associated aggregate identifier in the form of a name or label. Populating the aggregate data item may be achieved by operating on two other aggregate data items, or through the application of more general rules such as determining which data items comprising another aggregate data item have values in a particular range. The aggregate name or label can then be used in the textual information 26 just as for the case of using data item names in the textual information 26. Again, the order of the aggregate names in the textual information 26 may be determined by the textual information generator 4.

In a lengthy report consisting of several report sections (each with an optional heading), the order in which these report sections is presented is an important factor for the professional or end user. That is, the end user wants to see the most important report sections near the top of the report. However, what makes one report section more important than another depends on the particular case that is being interpreted. It is therefore advantageous to order specified report sections using rules that operate on the data in each case. The placement of some other report sections must be fixed, for example a summary report section that is always at the top of the report. Hence the user may be able to define a mixture of both fixed and variable report section orderings.

Allergy reporting is a domain where variable report section ordering may be required. There will be at least five separate report sections - corresponding to the comments on the pollen, food, mite, mould and animal

allergen test results. If the food allergy test results are the most significant for a given patient, then the food report section should come before the other four, and so on. The report section corresponding to the least significant test results should be positioned after the others. Furthermore, there are fixed report sections, namely the summary report section which is at the top of the report, and a recommendations report section which is typically at the bottom of the report .

10

Consequently, the system provides means for the operator 38 to define a "derived data item" for each variable report section, using the rules syntax, which assigns a value corresponding to the desired report section ordering. In the allergy example above, there would be five derived data items, say "pollen_order" , "food_order" , "mite_order" , "mould_order" and "animal_order" . Pollen_order would be defined as the highest value of any pollen data item, and similarly for the others. The derived data item "pollen_order" is associated with the pollen report section. For each case, the values of the five derived data items will be calculated, and the corresponding report sections will be ordered according to these values. For example, if the case had data items and values :

25

```
Grass = 50, birch = 20, (pollen)
wheat = 5, soya = 15, (food)
mould = 2,
mite = 1,
cat = 62, dog = 49 (animal)
```

30

then the report sections would be in the following order: Animal, pollen, food, mould, mite.

35

In some embodiments, the system may provide at least one of the following:

40

- A RippleDown rule system as the underlying technology to manage the very large Knowledge Bases required.
- Facilities to generate coded information that are machine instructions, such as to control a workflow engine, which for example, controls laboratory workflow such as autovalidation and reflexive testing, using coded outputs from the Knowledge Base .
- Natural language syntax for building rule conditions
- Insertion of variables into comments that are evaluated by the specific case that is

45

interpreted. Variables may be defined using aggregate data items .

Example 1

5 A first example application is a leukaemia report knowledge base where diagnosis is performed using hundreds of tests whose values are determined by a micro array of hundreds of protein expression or gene expression markers. An expert may build a diagnosis and report Knowledge Base
10 that identifies the subsets of relevant markers, the diagnosis corresponding to this pattern, and comments for those significant subsets in a textual report to the referring medical practitioner.

15 The array test results are provided as inputs to the Knowledge Base as a plurality of data items and value pairs. The data items are, in this example, labelled CD1 to CD100 to identify them, indicating 100 elements to the array. Real-world examples may contain several hundreds of
20 markers.

In this example, a value for one of the data values of less than 50 means that there is no expression of the antibody corresponding to that marker for that patient
25 sample. A value greater than 50 is possibly significant (depending on the values of other markers) . A value higher than 100 for a marker indicates a significant expression.

The diagnosis of a particular variety of leukaemia can be deduced from the values of specified sub-sets of
30 the 100 data values.

For example, a diagnosis of B-cell Chronic Lymphocytic Leukaemia (B-CLL) can be deduced from the
35 significant expression of at least 2 of CD1, CD2, CD3, CD4 and CD5 . This diagnosis is supported by the significant expression of any of CD6, CD7, CD8, CD9 and CD10 although these are not diagnostic of B-CLL in themselves.

40 Alternatively, a diagnosis of Acute Myeloid Leukaemia (AML) can be deduced from the significant expression of at least 2 of CD11, CD12, CD13, CD14 and CD15 . This diagnosis is supported by the significant expression of any of CD16, CD17, CD18, CD19 and CD20 although these are not
45 diagnostic of AML in themselves .

Five aggregate data items are populated with received data items as specified by an example structure:

- 21 -

1. "BCLL Diagnostic" populated by the data items CD1, CD2, CD3, CD4, CD5
2. "BCLL Supporting" populated by the data items CD6, CD7, CD8, CD9, CD10
- 5 3. "AML Diagnostic" populated by the data items CD11, CD12, CD13, CD14, CD15
4. "AML Supporting" populated by the data items CD16, CD17, CD18, CD19, CD20
- 10 5. "Leukaemia" populated by the aggregate data items BCLL Diagnostic, AML Diagnostic, BCLL Supporting, AML Supporting

A schematic of one embodiment of a structure giving the hierarchical relationship for these data items and aggregate data items is shown in Figure 6. In some 15 embodiments, once the lower levels of the structure is populated, the value or characteristics at the upper levels are calculated. The structure may be stored in the memory 20 or hard drive 2 (or other data storage unit) of the device 1, for example, and interpreted by the CPU 4.

The following ranges may also be defined:

- "Undetected" defined as the constant 50
- "High" defined as the constant 100

25

To represent the significant data items in each set, further aggregate data items are populated by applying the following rules:

- 30 1. "Significant BCLL Diagnostic" populated by the rule "BCLL Diagnostic in range [> High] "
2. "Significant BCLL Supporting" populated by the rule "BCLL Diagnostic in range [> Undetected] "
3. "Significant AML Diagnostic" populated by the 35 rule "AML Diagnostic in range [> High] "
4. "Significant AML Supporting" populated by the rule "AML Diagnostic in range [> Undetected] "

40 A BCLL diagnostic comment is given by the following pseudo text:

"Pan-B cell antigen expression with <Significant BCLL Diagnostic, co-expressed with Significant BCLL Supporting>, typical of B-Cell Chronic Lymphocytic 45 Leukaemia (B-CLL) ."

The variable "<Significant BCLL Diagnostic" is an instruction to list the names and values of the significant BCLL data items, and similarly for the

variable <Significant BCLL Supporting> . In this embodiment, the listed names and values are ordered in terms of decreasing data item value so that the most significant attributes are listed first.

5

A BCLL diagnostic rule triggers the generation of the BCLL diagnostic comment as follows:

"number of Significant BCLL Diagnostic >= 2", and
"number of Significant BCLL Supporting >= 1"

10

That is, the comment is generated if there are 2 or more data items in the set Significant BCLL Diagnostic, and 1 or more data items in the set Significant BCLL Supporting.

15

As a second comment example, the AML diagnostic comment is given by the following pseudo text:

20

"Consistent with AML antigen expression based on positive <Significant AML Diagnostic as names >, co-expressed with <Significant AML Supporting as names>. Query possible M2 classification."

25

The data item ^Significant AML Diagnostic as names >" is an instruction to the Knowledge Base to list just the names of the significant BCLL data items, and similarly for the variable Significant AML Supporting as names>.

30

In this embodiment, the listed names and values are ordered in terms of decreasing data item value so that the most significant data items listed first, even though the values will not be shown for this comment .

35

The AML diagnostic rule triggering the generation of the AML comment may be:

"number of Significant AML Diagnostic >= 2", and
"number of Significant AML Supporting >= 1"

40

The comment is given if there are 2 or more data items in the set Significant AML Diagnostic. This in turn means that there are 2 or more data items in the set AML Diagnostic which have values greater than 100, and at least 1 data item in the set BCLL Supporting which has a value greater than 50 .

45

Consider the results of the testing for a sample from patient "A" as follows:

CD1	0
CD2	95
CD3	190
CD4	150
CD5	260
CD6	0
CD7	90
CD8	60
CD9	0
CD10	15
CD11	30
CD12	60
CD14	20
CD15	0
CD16	0
CD17	0
CD18	50
CD19	50
CD20	40
CD26 to CD100	0

5

These results are sent to an embodiment of the Knowledge Base which evaluates the aggregate data items and evaluates expressions as follows:

- 10
- Significant BCLL Diagnostic evaluates to "CD5, CD3 and CD4"
 - Significant BCLL Supporting evaluates to "CD7 and CD8"
- 15
- Both Significant AML Diagnostic and Significant AML Supporting evaluate to null.

20 The Knowledge Base then makes an interpretation according to the rules defined above. The BCLL rule is applicable in this case as there are 3 elements in the Significant BCLL Diagnostic set, and 2 elements in the Significant BCLL Supporting set.

25 The Knowledge Base evaluates the variables in the BCLL comment "<Significant BCLL Diagnostio" and "<Significant BCLL Supporting>" then gives the evaluated comment :

5 ^Pan-B cell antigen expression with CD5 (260) , CD3 (190) and CD4 (150) , co-expressed with CD7 (90) and CD8 (60) , typical of B-Cell Chronic Lymphocytic Leukaemia (B-CLL) ."

For a second example, consider the test results for patient "B" as follows :

CD1	0
CD2	5
CD3	30
CD4	40
CD5	60
CD6	0
CD7	20
CD8	40
CD9	0
CD10	15
CD11	130
CD12	160
CD14	20
CD15	0
CD16	0
CD17	0
CD18	55
CD19	100
CD20	60
CD26 to CD100	0

10

These results are sent to the Knowledge Base which firstly evaluates the aggregate data items as follows:

- 15 • Both Significant BCLL Diagnostic and Significant BCLL Supporting evaluate to null .
- Significant AML Diagnostic evaluates to "CD12 and CD11"
- 20 • Significant AML Supporting evaluates to "CD19, CD20 and CD18"

The Knowledge Base then makes an interpretation according to the rules above. The AML rule is applicable in this case as there are 2 elements in the Significant AML Diagnostic set, and 3 elements in the Significant AML Supporting set.

25

The Knowledge Base then gives the comment:

"Consistent with AML antigen expression based on positive CD12 and CD11, co-expressed with CD19, CD20 and CD18. Query possible M2 classification."

Example 2

Another example application is an allergy report knowledge base where there are potentially 500 or more IgE tests that can be performed. The task of the allergy expert is to advise the referring doctor which subset of the tests performed have significant result values for the patient, which test values may not be significant, which tests need to be followed up, and which further tests, out of the 500 possible tests, should be also be performed.

One example solution:

- From the total collection of possible data item names, to group those data item names into aggregate data items based on domain-specific rules, e.g. significant pollen attributes, and case-specific rules.
- To use any of the characteristics of an aggregate data item as the basis for further rules and / or comments. For example, give a particular comment if the number of elements of an aggregate data item is above a certain number, or if the set includes a particular element
- To use one or more aggregate data item as variables in a comment, for example, "The dog, cat and peanut allergies are significant" where the phrase "dog, cat and peanut" is an evaluation of the aggregate data item consisting of allergens that are significant for this case. The generic form of the comment may be *The {SignificantAllergens} allergies are significant* where the set {SignificantAllergens} is itself defined by rules.
- To optionally include the values of the data item in the comment, e.g. The dog (102.3), cat (56.4) and 'peanut (43.5) allergies are significant.
- To appropriately order the data items in an aggregate data item which appears in a comment, e.g. in terms of decreasing attribute value in the case so that the most significant attributes appear first in the comment.
- To automatically format the data items into a naturally constructed sentence that is consistent with the rest of the report. For example if 3

attributes are significant the format of the set may be "dog, cat and peanut allergies" whereas if only 2 attributes are significant the format of the set may be "dog and cat allergies"

- 5 • To be able to define a grouping of data items based on a previous grouping of data items, e.g. the new aggregate data item could be the difference, union, or intersection of one set from another, or any set operation. This allows the definition of a hierarchy of sets. For 10 example the difference between the set "appropriate tests" and the set "ordered tests" could identify the set of appropriate tests which have not yet been ordered.
- 15 • To be able to define a comment that uses either individual data items or an aggregate data item containing those attributes as appropriate. For example, to use the term "food allergy" rather than "peanut, soy, milk, egg and peach allergy" 20 if the list of individual data items would be too long for the comment for that particular case.
- Similarly, to be able to define a comment that uses a super aggregate data item name rather than subset names as appropriate. E.g. to use the term 25 "inhalant allergy" rather than "pollen, animal, mould, ...allergies" if the list of individual aggregate data items would be too long for the comment for that particular case.

30 Pseudo text

The following table defines some of the features of the pseudo text discussed above, where:

35 s, t, x, y, z refer to data items, either primary or derived data items .

The letters X, Y, Z refer specifically to aggregate data items (a type of derived data item) .

The letters a, b refer to numbers or named constants.

The letter p refers to a boolean expressions.

40

Syntax	Definition	Examples
{ }	{x, y, z ...} an aggregate data item consisting of the specified data items.	Aggregate data item <i>food</i> defined as {peanut, egg, milk}

in range	<p>$\{x, y, z \dots\}$ in range $[a, b]$</p> <p>an aggregate data item consisting of those specified data items where the value of that data item is in the specified range including endpoints a and b.</p> <p>$\{x, y, z \dots\}$ in range (a, b)</p> <p>as above, but the values a and b are not themselves included in the range. Mixed open and closed ranges are also possible, that is</p> <p>$\{x, y, z \dots\}$ in range $(a, b]$</p> <p>and so on</p>	<p>Aggregate data item <i>HighFood</i> defined as</p> <p><i>{peanut, egg, milk} in range [20, 50]</i></p> <p>or equivalently</p> <p><i>food in range [20, 50]</i></p> <p>where <i>food</i> is defined as above.</p>
Number of	<p>Number of $\{x, y, z \}$</p> <p>Number of X</p> <p>The number of data items in an aggregate data item</p>	<p><i>Number of {peanut, egg, milk} in range [20, 50]</i></p> <p><i>Number of HighFood</i></p>
Is empty	<p>X is empty</p> <p>True if there are no data items in an aggregate data item</p>	<i>HighFood is empty</i>
Sum of	<p>sum of $\{x, y, z, \dots\}$</p> <p>sum of X</p> <p>The arithmetic sum of the values of the data items in an aggregate data item</p>	<p><i>sum of {peanut, egg, milk} in range [20, 50]</i></p> <p><i>sum of HighFood</i></p>
Union	<p>$\{s, t, \dots\}$ union $\{x, y, \dots\}$</p> <p>X union Y</p> <p>This is the set union operation</p>	<i>Food union animal</i>
in	<p>$\{s, t, \dots\}$ in $\{x, y, \dots\}$</p> <p>X in Y</p> <p>This is the set intersection operation</p>	<i>HighAllergens in food</i>
Not in	<p>$\{s, t, \dots\}$ not in $\{x, y, \dots\}$</p> <p>X not in Y</p> <p>This is the set difference operation</p>	<i>HighAllergens not in food</i>
Includes	<p>X includes x</p> <p>True if the aggregate data item X contains the data item x</p>	<i>HighFood includes peanut</i>
Does not include	<p>X does not include x</p> <p>Negation of "includes"</p>	<i>HighFood does not include peanut</i>

For	X for p Defines an aggregate data item X subject to a condition p	Aggregate data item <i>AllergensForInfants</i> defined as {peanut, egg, milk} for (age < 2)
As groups from	{X, Y, Z...} as groups from {S, T...} Represents data items in a comment using the specified aggregate names	<i>{allergens} as groups from {food, pollen, mould, mite,}</i> will appear in a comment as "food and pollen" for example, rather than as constituent data item names "peanut (20.0), egg (15.1) and grass (6.4)"
As names	{x, y, z} as names X as names Represents an aggregate data item in a comment without the corresponding data item values	<i>Food as names</i> will appear in a comment as "peanut and egg" rather than "peanut (20.0) and egg (15.1)"

Now that embodiments have been described, it will be appreciated that some embodiments may have some of the following advantages :

- 5 • It is possible to process a large number of text results comprising one investigation.
- The report presents the data items in an appropriate order, with a linguistically natural syntax.
- 10 • The number of specific report variations are essentially infinite due to the number of possible subsets, and the number of possible orderings within each subset .
- The number of specific rule conditions that determine a particular report are also
- 15 essentially infinite due to the number of patterns in the attributes in the case, and in their values .
- The expert is able to build and maintain the
- 20 knowledge base with a manageable number of rules.
- An expert system is provided that can manage large numbers of attributes and the correspondingly large number of report variations .

25 It will be appreciated that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described.

30 The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Claims

- 1 . A method of generating information from a plurality
of data items, the method comprising the steps of:
5 populating an aggregate data item with at least
 one of the plurality of data items; and
 generating the information using the aggregate
 data item.
- 10 2 . A method defined by claim 1 comprising the prestep of
 populating each of a plurality of elements of a
 predefined structure with a corresponding one of the
 data values, the structure relating each of the
 plurality of elements to the aggregate data item.
- 15 3 . A method defined by claim 2 comprising populating the
 aggregate data item with at least one of the
 plurality of data items in accordance with the
 structure .
- 4 . A method defined by claim 1 wherein the information
 comprises textual information.
- 20 5 . A method defined by claim 1 wherein the information
 comprises a machine instruction.
- 6 . A method defined by any one of the preceding claims
 wherein the step of generating the information
 comprises the step of generating information using a
25 Knowledge Base or Decision Support System.
- 7 . A method defined by any one of the preceding claims
 wherein the step of populating the aggregate data
 item comprises a step of receiving the plurality of
 data items .
- 30 8 . A method defined by claim 4 wherein the textual
 information is syntactically and/or grammatically
 correct .
- 9 . A method defined by either claim 4 or 8 wherein the
 textual information is human readable.
- 35 10 . A method defined by any one of claims 4 , 8 and 9
 wherein the textual information forms at least part
 of a report .
- 11 . A method defined by any one of the preceding claims
 wherein the step of populating the aggregate data

item comprises the step of populating the aggregate data item by applying a rule to at least one of the plurality of data items, or another aggregate data item.

- 5 12. A method defined by any one of the preceding claims wherein each of the plurality of data items are associated with an identifier and a value.
13. A method defined by claim 12 wherein the identifier may be associated with a name or label for the data
10 item.
14. A method defined by claim 13 wherein the step of generating the information comprises the step of including in the information the names or labels of the data items populating the aggregate data item.
- 15 15. A method defined by any one of claims 12 to 14 wherein the step of generating the information comprises the step of including in the information the values associated with the data items populating the aggregate data item.
- 20 16. A method defined by either claim 14 or 15 wherein the step of generating information comprises the step of determining the order of the names or labels in the information.
- 25 17. A method defined by any one of the preceding claims wherein, the step of generating the information comprises determining characteristics of the aggregate data item.
- 30 18. A method defined by any one of the preceding claims wherein the step of generating the information comprises the step of including in the information about the determined characteristics of the aggregate data item.
- 35 19. A method defined by any one of the preceding claims wherein the step of populating the aggregate data item may comprise populating the aggregate data item with one or more other aggregate data items .
- 40 20. A method defined by any one of the preceding claims wherein the step of generating the information comprises applying one or more rules to the aggregate data item.

21. A method defined by claim 20 wherein the one or more rules may form at least part of a ripple down rules knowledge system.
22. A method defined by any one or more preceding claims comprising:
- 5
- (a) populating one or more aggregate data items with at least one of the plurality of data items;
- (b) populating one or more further aggregate data items with data items from the one or more aggregate data items by applying one or more rules to the one or more aggregate data items; and
- 10
- (c) generating the information using the one or more further aggregate data items .
- 15 23. A system for generating information from a plurality of data items, the system comprising:
an aggregate data item populator for populating an aggregate data item with at least one of the plurality of data items; and
- 20 an information generator for generating the information using the aggregate data item.
24. A method of generating information from a plurality of data items, the method comprising the steps of:
evaluating an outcome of one or more rules using one or more aggregate data items each comprising one or more of the data items; and
- 25 generating the information according to the outcome .
25. A system for generating information from a plurality of data items, the system comprising:
- 30 an evaluator for evaluating an outcome of one or more rules using one or more aggregate data items each comprising one or more of the data items; and
an information generator for generating the information according to the outcome.
- 35
26. A method of generating information, the method comprising the steps of:
receiving a conceptual representation of information including an interpretive portion, the interpretive portion representing an operation on an aggregate data item comprising a plurality of data items; and
- 40

generating the information from the interpretive portion.

27. A system for generating information, the system comprising:
- 5 a receiver for receiving a conceptual representation of information including an interpretive portion, the interpretive portion representing an operation on an aggregate data item comprising a plurality of data items; and
- 10 an information generator for generating the information from the interpretive portion.
28. A computer program comprising instructions for controlling a computer to implement a method in accordance with the method defined by claim 1.
- 15 29. A computer readable medium providing a computer program in accordance with the computer program of claim 28.
30. A computer program comprising instructions for controlling a computer to implement a method in
- 20 accordance with the method defined by claim 24.
31. A computer readable medium providing a computer program in accordance with the computer program defined by claim 30.
32. A computer program comprising instructions for
- 25 controlling a computer to implement a method in accordance with the method defined by claim 26.
33. A computer readable medium providing a computer program in accordance with the computer program defined by claim 32.
- 30 34. A method substantially as herein described with reference to the accompanying figures.
35. A system substantially as herein described with reference to the accompanying figures.

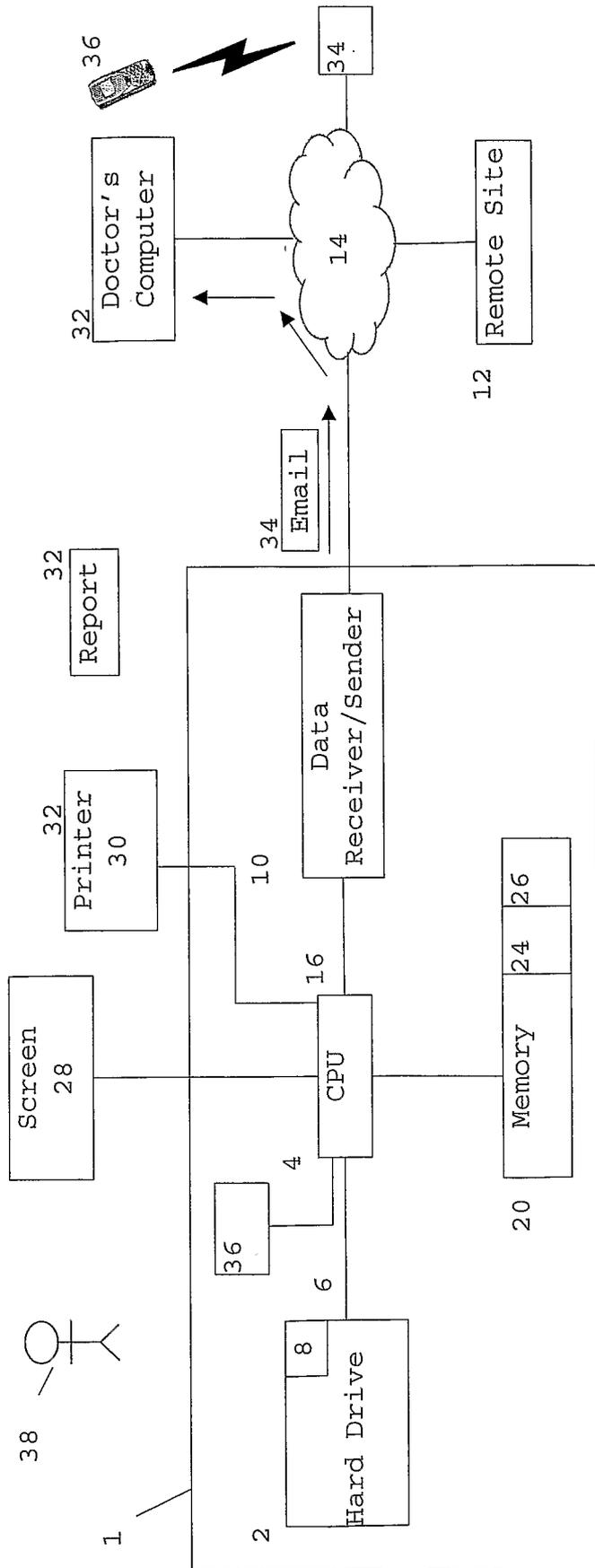


Figure 1

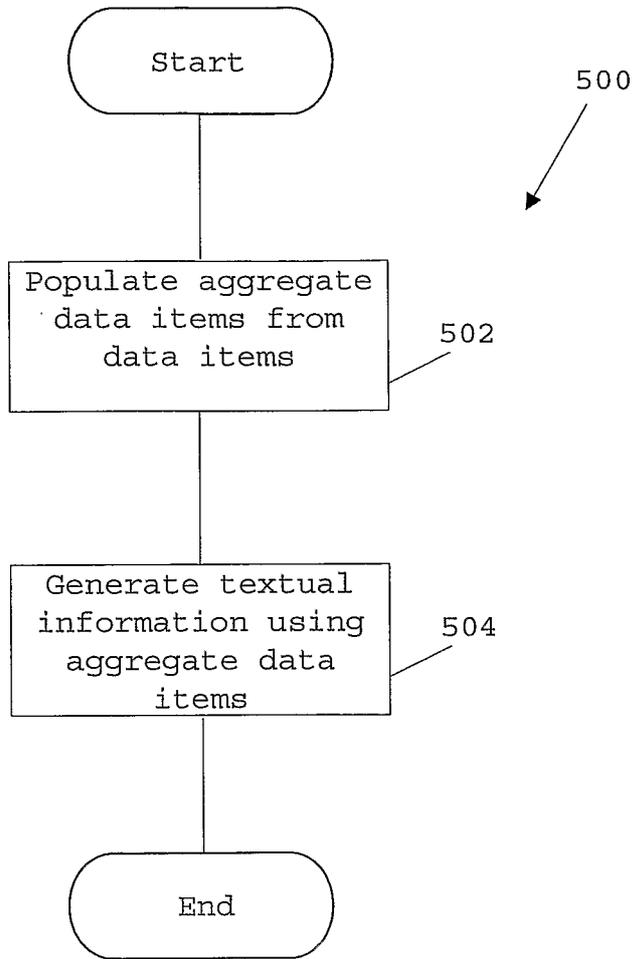


Figure 2

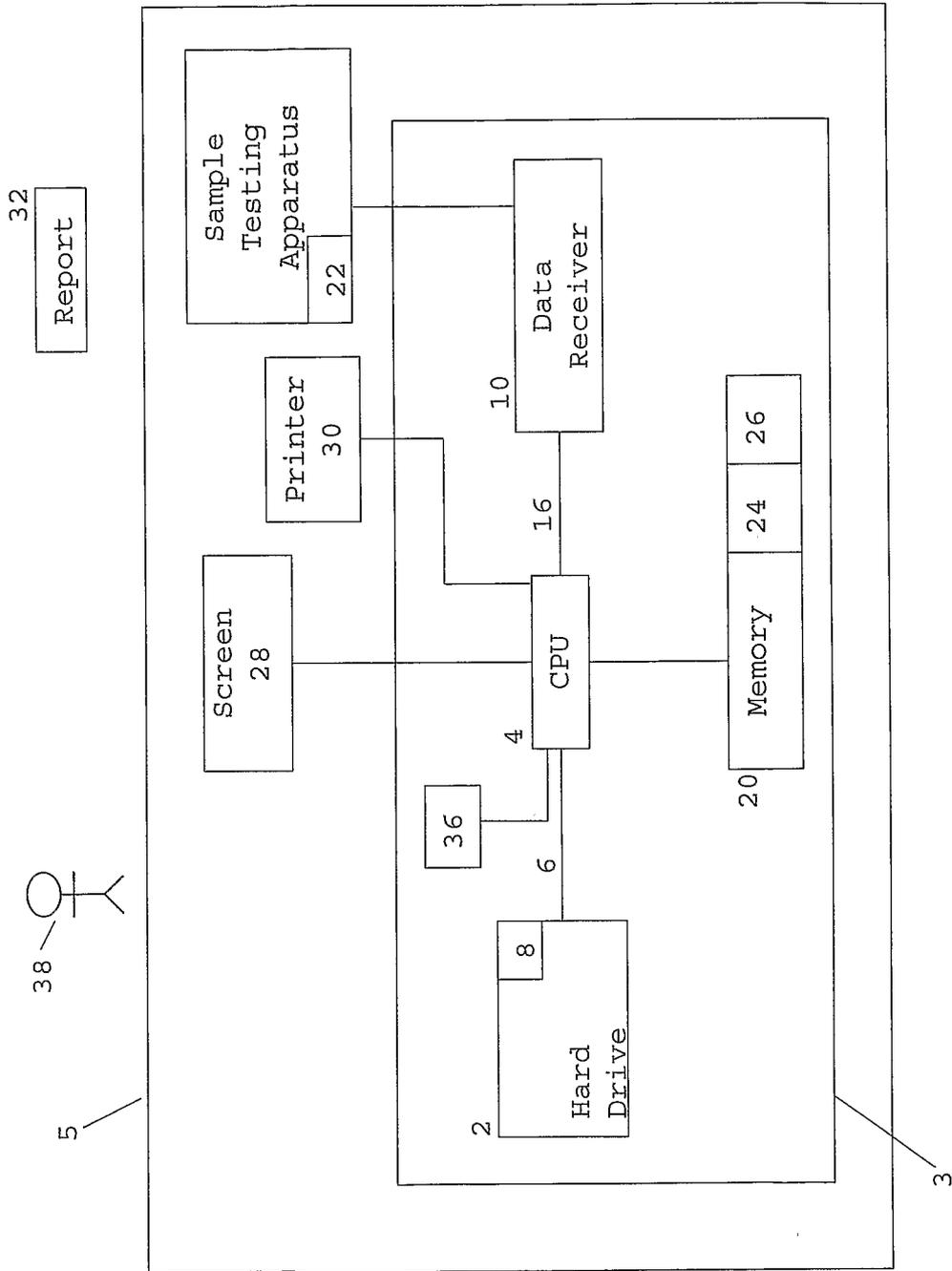


Figure 3

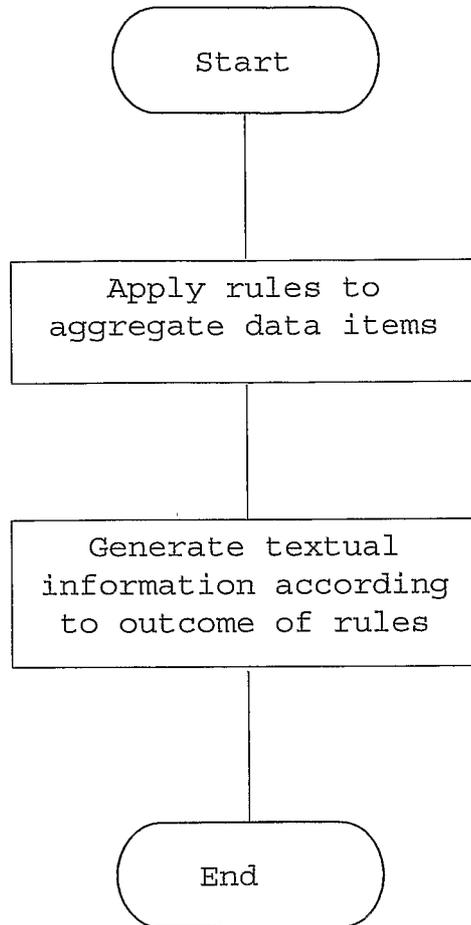


Figure 4

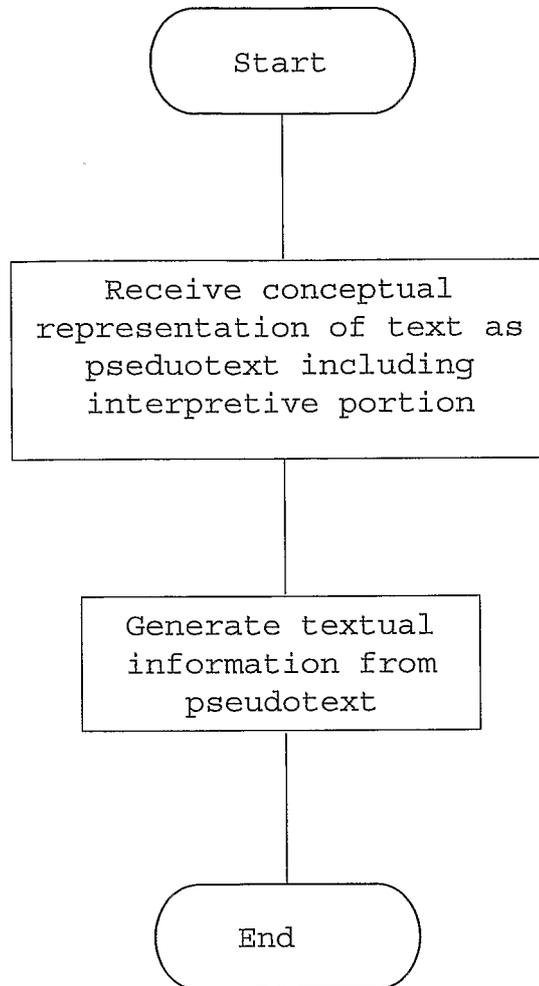


Figure 5

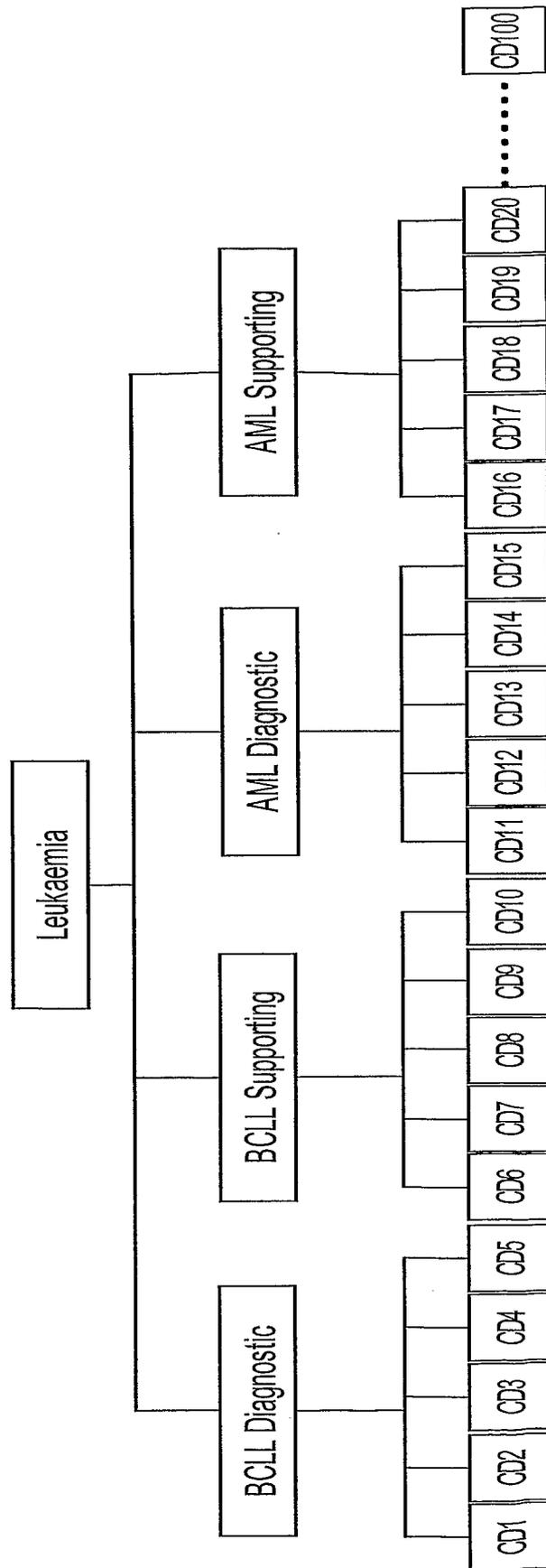


Figure 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2009/000134

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

G06Q 50/00 (2006.01) G06F 17/20 (2006.01) G06F 17/30 (2006.01)

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)

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)

WPI, Esp@cenet, WIPO and USPTO Web patent database, PatentLens, Google etc. "text, report, data, database etc."

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
x	US 6004276 (WRIGHT et al.) 21 December 1999 Column 79 line 13 to column 80 line 57 in particular.	1-23,28,29
x	US 2002/0188896 A1 (FELTEAU et al.) 12 December 2002 Figures 1, 7A, 7B, 8A to 8D, paragraphs 46 and 62 in particular.	1-23,28,29
x	US 2008/0040162 A1 (BRICE) 14 February 2008 Figures 5A to C and paragraph 28 in particular.	1-23,28,29
x	US 6735593 B1 (WILLIAMS) 11 May 2004 Column 11 lines 53 to 59, column 22 line 45, column 28 lines 16 to 26, column 29 lines 14 to 22 and column 38 line 65 in particular.	1-4,7-17,19-23,28,29

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	"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	
"E" earlier application or patent but published on or after the international filing date	"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	
"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)	"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 documents, such combination being obvious to a person skilled in the art	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 05 May 2009	Date of mailing of the international search report 08 MAY 2009
--------------------------------------------------------------------------	--------------------------------------------------------------------------

Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaaustralia.gov.au Facsimile No. +61 2 6283 7999	Authorized officer PETER THONG AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No : +61 2 6283 2128
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU2009/000134

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. 1 Claims Nos.: 34,35
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

The claims do not comply with Rule 6.2(a) because they rely on references to the description and/or drawings.

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See Supplemental Box.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1 to 23, 28 and 29.

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU2009/000134

Supplemental Box - Continuation of Box No. III

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

In assessing whether there is more than one invention claimed, I have given consideration to those features which can be considered to potentially distinguish the claimed combination of features from the prior art. Where different claims have different distinguishing features they define different inventions.

This International Searching Authority has found that there are different inventions as follows:

- Claims 1 to 23, 28 and 29 are directed to an arrangement for generating information that involves the population of aggregate data item with data item/s and generation of information using the aggregate data item which arrangement comprises a first distinguishing feature.
- Claims 24, 25, 30 and 31 are directed to an arrangement for generating information that involves an evaluator evaluating an outcome of rules using aggregate data items and generation of information according to the outcome which arrangement comprises a second distinguishing feature.
- Claims 26, 27, 32 and 33 are directed to an arrangement for generating information that involves the reception of conceptual representation of information including an interpretive portion representing an operation on an aggregate data item and the generation of information from the interpretive portion which arrangement comprises a third distinguishing feature.

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

Each of the abovementioned groups of claims has a different distinguishing feature and they do not share any feature which could satisfy the requirement for being a special technical feature. Because there is no common special technical feature it follows that there is no technical relationship between the identified inventions. Therefore the claims do not satisfy the requirement of unity of invention *a priori*.

INTERNATIONAL SEARCH REPORT

International application No.

Information on patent family members.

PCT/AU2009/000134

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
US	6004276	AU	64436/98	WO	9838910		
US	2002/0188896	EP	1435060	WO	02098822		
US	2008/0040162	NONE					
US	6735593	AU	11742/00	EP	1145150	GB	2360108
		WO	0029980				

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX