



(51) International Patent Classification:
G06F 19/00 (2011.01)

(21) International Application Number:
PCT/IB2011/053509

(22) International Filing Date:
5 August 2011 (05.08.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/370,832 5 August 2010 (05.08.2010) US

(71) Applicant (for all designated States except US): **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL];
Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LEE, Michael, Chun-chieh** [US/US]; c/o High Tech Campus Building 44, NL-5656 AE Eindhoven (NL). **COHEN-SOLAL,**

Eric [FR/US]; c/o High Tech Campus Building 44, NL-5656 AE Eindhoven (NL).

(74) Agents: **VAN VELZEN, Maaïke, M.** et al.; High Tech Campus 44, NL-5600 AE Eindhoven (NL).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,

[Continued on next page]

(54) Title: REPORT AUTHORING

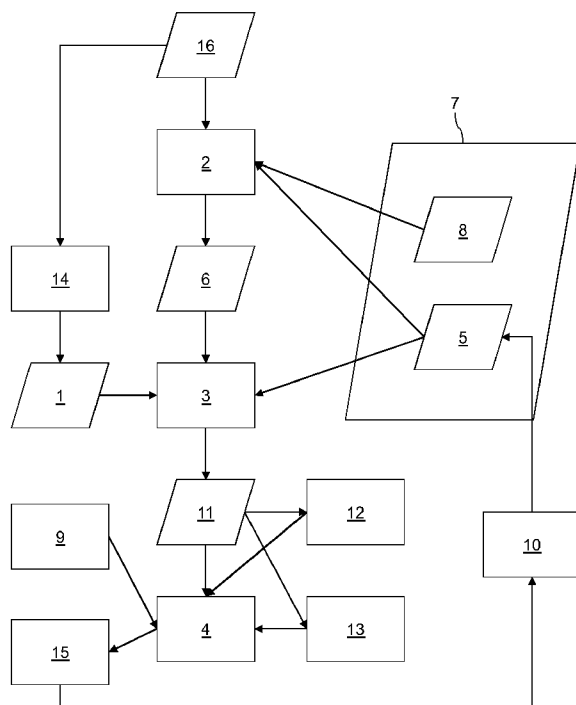


Fig. 1

(57) Abstract: A system for aiding report authoring is disclosed. The system comprises a set of associations (1), wherein an association links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a frequent co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain. The system comprises a concept term extractor (2) for extracting at least one first concept term (6) relating to a particular subject from a data record (7) associated with the particular subject, the data record (7) comprising an at least partially completed report (5) relating to the particular subject. The system comprises a missing concept term finder (3) for finding at least one second concept term (11) linked to the at least one first concept term (6) based on the set of associations (1), wherein the at least one second concept term (11) is missing from the at least partially completed report (5).



TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

Report authoring

FIELD OF THE INVENTION

The invention relates to aiding report authoring.

BACKGROUND OF THE INVENTION

5 A Radiology Information System (RIS) integrated with Picture Archiving and Communication System (PACS) capabilities helps hospitals and imaging centers to optimize their workflow and manage images and information circulation throughout the facility to deliver efficient patient care and services. The different steps involved in this process start with a physician ordering a radiology procedure, a patient going to the radiology department, 10 the imaging procedure being performed, a radiologist reading the images and producing a report to be sent to the referring physician.

The report that is produced by the radiologist typically summarizes relevant aspects of the patient history and the clinical question being posed, and then indicates what regions of the radiology image were examined and what findings there were, if any. The 15 findings and any conclusions or impression that can be drawn from the findings are typically given in separate sections of the report. Broadly speaking, the ordering physician expects the findings and impression to address the clinical question that he or she has originally posed. If the findings and impression inadequately address the clinical question, it may be necessary for the ordering physician to contact the radiologist and ask for clarification.

20 Geis, J. (2007), "Medical Imaging Informatics: How It Improves Radiology Practice Today," Journal of Digital Imaging 20(2): 99 discloses that radiologists are under pressure to add more value to medical imaging-to provide more educated, accurate, useful, and efficient interpretations in the face of increasingly large and complex imaging studies and to communicate this information quickly and in the most useful manner. The radiology 25 department and radiologist both need to be better, faster, and cheaper. Moreover, the radiologist synthesizes what he sees on the images, clinical data, and his medical knowledge to produce the interpretation. This should be a cohesive, accurate, helpful discussion that adds value to the images.

Therefore, there is a need for tools that can help them deliver the desired quality in reporting and still address the local constraints of each radiology institution. There is a need for consistency between radiologists in the way they report on similar cases (inter-radiologist consistency) as well as a need for consistency over time for each individual radiologist (intra-radiologist consistency). Moreover, the nature of their work imposes time constraints and reading quotas increasing the number of potentially incomplete and, consequently, potentially inaccurate reports. This inaccuracy will possibly lead to unnecessary interactions with referring physicians requesting further clarification, causing interruption and decrease of efficiency that can be avoided.

SUMMARY OF THE INVENTION

It would be advantageous to have an improved aid for report authoring. To better address this concern, a first aspect of the invention provides a system comprising

- a set of associations, wherein an association links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a usual co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain;
- a concept term extractor for extracting at least one first concept term relating to a particular subject from a data record associated with the particular subject, the data record comprising an at least partially completed report relating to the particular subject;
- a missing concept term finder for finding at least one second concept term linked to the at least one first concept term based on the set of associations, wherein the at least one second concept term is missing from the at least partially completed report;
- an indicator for indicating the found at least one second concept term to a user.

The system thus provides an indication of missing second concept terms which are associated with first concept terms relating to the subject. By providing this indication of missing second concept terms, the user is reminded to include a comment involving the second concept term in the report. This way, it is easier to author a complete report, because any potentially missing concept terms are indicated to the user. Consequently the number of incomplete reports may be reduced. Moreover, when the report is read by a consumer of the report, such as a referring physician, fewer interactions with the author of the report are necessary because the report contains the information about the relevant concept terms. Alternatively, when the concept was already discussed in the report using different terminology, the user may be reminded to rephrase the report to use the second concept term.

This improves the consistency of the terminology used in the reports. The user of the system may comprise an author of the at least partially completed report.

The knowledge domain may comprise a medical knowledge domain, the subject may comprise a patient, and the report may comprise a medical diagnosis or treatment report. In an effort to help ensure that medical reports adequately address the clinical question posed by e.g. a referring physician, the system may provide means of verifying that a radiologist or physician completes the report using the terminology expected based on the clinical context.

The concept extractor may be arranged for extracting the at least one first concept term from the at least partially completed report. The first concept term(s) may be present in the at least partially completed report, for example, in an introductory section in which the context of the report is set forth. Such sections may include a clinical history description and/or a description of the questions asked by a referring person.

The concept extractor may be arranged for extracting the at least one first concept term from a first section of the at least partially completed report. Moreover, the set of associations may comprise associations linking first concept terms from the first section to second concept terms belonging to a different second section of the reports in the knowledge domain. This way, the concepts which need discussion in the second section of the report, based on the first section of the report, may be identified.

The concept extractor may be arranged for extracting the at least one first concept term from medical history information represented by the data record. Medical history information may provide clues to what concept terms should be discussed in the remainder of the report.

The at least one second concept term may be associated with a particular section of the report, and the indicator may be arranged for indicating with which section the at least one second concept term is associated. This makes the suggestion of missing concept terms more specific, because the missing concept terms are presented in the context of the section in which they should appear.

The system may comprise a report input for receiving a completed report, and wherein the at least partially completed report is the completed report. The completed report may thus be checked for missing concept terms. This way, the completed report may be verified. If any missing concept terms are identified, the completed report and the missing concept terms may be presented to the author (or to another author), who may then consider providing additional text involving the missing concept.

The system may comprise a report creation tool for enabling a user to create at least part of the report. The indicator may be arranged for indicating the found at least one second concept term during a process of creating the at least part of the report. For example, the indicator may be part of the report creation tool. After entering a word, the indicator may
5 update the indication of missing concept terms. This helps the user to check the completeness of the report during creating the report, which helps creating reports having the necessary information and may avoid any iterations of correcting or extending the report.

The at least one second medical concept term may comprise an anatomical region term. The system may comprise an image viewer configured to display a view of an
10 anatomical region of the subject based on the anatomical region term, wherein the anatomical region corresponds to the anatomical region term. This way, the number of manual interactions is reduced, because the relevant anatomical region is visualized automatically.

The system may comprise a second concept term classifier for classifying the at least one second concept term by type of concept, and wherein the indicator is arranged for
15 indicating the type of concept for a second concept term. This provides a more organized view of the concepts.

The associations may have different strengths. The system may comprise a second concept term sorter for sorting the second concept terms found by the missing concept term finder, based on the strengths of the associations by which the second concept terms
20 were found. This way, the concept terms which are most likely to be relevant for discussion in the report may be indicated as such to the user.

The system may comprise an association generator for creating an association based on co-occurrence frequencies of first and second concept terms in a collection of reports. This way, the set of associations can be generated automatically. Moreover, the set of
25 associations can be updated taking into account new reports as they are added to the collection.

The system may be incorporated in a workstation, for example a medical workstation.

In another aspect, the invention provides a method of aiding report authoring,
30 comprising

- extracting at least one first concept term relating to a particular subject from a data record associated with the particular subject, the data record comprising an at least partially completed report relating to the particular subject;

- finding at least one second concept term linked to the at least one first concept term based on a set of associations, wherein the at least one second concept term is missing from the at least partially completed report, wherein an association of the set of associations links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a usual co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain; and
- indicating the found at least one second concept term to a user.

In another aspect, the invention provides a computer program product comprising instructions for causing a processor system to perform the method set forth.

It will be appreciated by those skilled in the art that two or more of the above-mentioned embodiments, implementations, and/or aspects of the invention may be combined in any way deemed useful.

Modifications and variations of the workstation, the method, and/or the computer program product, which correspond to the described modifications and variations of the system, can be carried out by a person skilled in the art on the basis of the present description.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter. In the drawings,

- Fig. 1 is a diagram of a system for aiding report authoring;
- Fig. 2 is a flowchart of a method of aiding report authoring; and
- Fig. 3 is a diagram of a system for aiding report authoring.

DETAILED DESCRIPTION OF EMBODIMENTS

Fig. 1 illustrates a system for aiding report authoring. The system may in particular be used as an aid in authoring reports in a particular knowledge domain. Examples of such a knowledge domain include radiology, cardiology, oncology. It is possible to tailor the system for a narrower knowledge domain such as lung x-ray diagnosis reports, or for a broader knowledge domain, such as medical reports in general. Other, non-medical knowledge domains may also be applied. The Figure shows an exemplary embodiment with a number of features. These features are only described by means of example; most of the features are optional or can be implemented in different ways.

The system may comprise a set of associations 1. Each association in the set may link at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain. Associations may indicate that the at least one first concept term frequently co-occurs with the at least one second concept term in reports in the knowledge domain. In other words, reports containing the first concept term frequently also contain the second concept term. The set of associations 1 may be prepared in advance. It is also possible to update the set of associations 1 by analyzing a collection of reports, as will be described hereinafter. The concept terms which are used in the knowledge domain may be organized in a set of concept terms 16. The associations may also indicate that when a data record 7 contains a first concept term, the report frequently contains a second concept term.

The system may further comprise a concept term extractor 2 for extracting at least one first concept term 6 from a data record 7. The data record 7 may be part of a database with data records, wherein each data record may contain information relating to a particular subject, such as a patient. The data record 7 may be understood to also comprise any subject-related data files which are currently being edited, such as a partially completed diagnostic report which is loaded in a report editor. Accordingly, the data record 7 may comprise an at least partially completed report 5 relating to the particular subject. The concept term extractor may be arranged for extracting concept terms matching one of the concept terms in the set of concept terms 16. In particular, the concept term extractor 2 may be arranged for extracting any concept terms which are used as first concept terms in the set of associations 1.

The system may further comprise a missing concept term finder 3, for finding any concept terms which are missing in the at least partially completed report, but which would be expected to appear in the report based on the associations and the concept terms extracted from the data record 7. To this end, the missing concept term finder 3 may identify one or more concept terms which are linked to the at least one first concept term 6 extracted by the concept term extractor 2 by at least one of the associations 1. Moreover, the missing concept term finder 3 may check whether any of these one or more concept terms are missing from the at least partially completed report 5. The missing concept terms are outputted by the missing concept term finder as the at least one second concept term 11.

The system may further comprise an indicator 4 for indicating the found at least one second concept term 11 to a user. This indication can be performed in numerous ways. For example, the at least one second concept term 11 may be shown on a display in a

list form. This may enable the user to update the at least partially completed report 5 with information including the at least one second concept term 11.

As discussed above, in medical applications, the knowledge domain may comprise a medical knowledge domain, the subject may comprise a patient, and the report may comprise a medical diagnosis or treatment report.

The concept term extractor 2 may be arranged for extracting the at least one first concept term 6 from the at least partially completed report 5. Consequently, it is possible that the concept term extractor 2 only evaluates the at least partially completed report, and not the remainder of the data record 7. This may be the case where the reports are self-contained, e.g., they contain the clinical history, any questions posed by the referring physician, the findings, and the diagnosis. These items may be organized in different sections of the report. Another division into sections may also be provided.

Such a systematic sectioning of reports may be used by the system. The concept term extractor 2 may be arranged for extracting the at least one first concept term 6 from a first section of the at least partially completed report 5. The set of associations 1 may comprise associations linking first concept terms belonging to the first section to second concept terms belonging to a different second section of the reports in the knowledge domain. Consequently, the missing concept term finder 3 may then find missing concepts in the second section of the report based on the concept terms extracted from the first section of the report. For example, the first section may comprise clinical history information and/or a diagnostic question posed, and the second section may comprise clinical findings and/or diagnosis.

The at least one second concept term 11 may be associated with a particular section of the report, in the sense that the at least one second concept term frequently appears in that particular section of reports in the knowledge domain. In such a case, the indicator 4 may be arranged for indicating with which section the at least one second concept term is associated. For example, the at least one second concept term may be displayed while the user is editing the relevant section of the report. When the user is editing another section, the second concept terms associated with that other section may be displayed. For example, while editing the “Findings” section, the second concept terms associated with the “Findings” section may be displayed; while editing the “Impressions” section, the second concept terms associated with the “Impressions” section may be displayed.

The concept term extractor 2 may be arranged for extracting the at least one first concept term 6 from medical history information 8 represented by the data record. This medical history information 8 may include past reports, medicine use, lab reports, and more.

The system may further comprise a user interface element 9 for enabling a user to indicate that the report 5 is complete, and wherein the indicator 4 is arranged for indicating the found at least one second concept term 11 in response to a user indication that the report 5 is complete. The concept term extractor and the missing concept term finder may also be triggered by the user indication, however they may also work in the background while the report is being written. The indication of missing concept terms to the user may be given only when the report 5 is considered complete, to avoid any biased reporting by the user.

The system may further comprise a report creation tool 10 for enabling a user to create the report 5 or at least part thereof. For example, an editor may be provided enabling a user to type in the (sections of the) report. The report creation tool 10 may also be based on speech recognition. The tool 10 may enable a user to dictate the report and extract the at least one first concept term 6 from the dictated text. Moreover, the missing concept term finder 3 may be arranged for checking whether the relevant concept terms are missing from the dictated text. The indicator 4 may be arranged for indicating the found at least one second concept term 11 during a process of creating the at least part of the report 5. Immediate feedback may guide a user to the relevant concepts.

The at least one second concept term 11 may comprise an anatomical region term, such as lungs or heart. The system may further comprise an image viewer 15 configured to indicate a view of the anatomical region corresponding to the anatomical region term. In particular, a medical image of the patient may be shown with an indication of the anatomical region. For example, the anatomical region may be indicated by highlighting or by means of an arrow.

The system may comprise a second concept term classifier 12 for classifying the at least one second concept term 11 by type of concept. Examples of types of concepts may include anatomical regions, diseases, findings, diagnosis. The indicator 4 may be arranged for indicating the type of concept for a second concept term 11. For example, the type of concept may be displayed next to the concept term. Alternatively, the second concept terms may be ordered in groups, each group corresponding to a concept type.

The associations of the set of associations 1 may have different strengths. The system may further comprise a second concept term sorter 13 for sorting the at least one second concept term 11 found by the missing concept term finder 3. The sorting allows the

user's attention to be drawn to the most pertinent missing concept terms. The sorting may be based on the strengths of the associations by which the second concept terms 11 were found.

The system may further comprise an association generator 14 for creating an association based on co-occurrence frequencies of first and second concept terms in a collection of reports. These co-occurrence frequencies may be used to estimate a probability that a report containing a first concept term also contains a second concept term. When this probability is high enough, the association linking these concept terms may be stored in the set of associations 1. This will be further elucidated hereinafter. Other ways to obtain the set of associations 1 are also possible. For example, the associations could be manually created. The associations could also be derived from co-occurrence frequencies of first and second concept terms in a collection of textbooks or scholarly articles on the domain, for example by data-mining for co-occurrence frequencies in medical journals. This could be performed in addition to or instead of analyzing the collection of reports.

The system set forth may be implemented as a workstation loaded with a suitable application program.

Fig. 2 illustrates a method of aiding report authoring. The method may comprise step 201 of extracting at least one first concept term relating to a particular subject from a data record associated with the particular subject, the data record comprising an at least partially completed report relating to the particular subject. The method may further comprise step 202 of finding at least one second concept term linked to the at least one first concept term based on a set of associations, wherein the at least one second concept term is missing from the at least partially completed report, wherein an association of the set of associations links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a frequent co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain. The method further comprises step 203 of indicating the found at least one second concept term to a user. The method may be modified or extended, e.g. based on the functionality described herein in respect of the system. The method may be implemented as a computer program product comprising instructions for causing a processor system to perform the method.

In addition to radiology reports, the system may be used in other medical fields for which reports are routinely generated. Examples include but are not limited to cardiology, oncology, pathology, or surgery. As a concrete example for cardiology, a report for which "pericarditis" is the diagnosis may be reasonably expected to mention "friction

rub” as one of the findings. A report lacking this term may be flagged, as it is either lacking a term that the reader might expect (i.e. the finding should be mentioned as being present or absent) or the terminology actually used in the report is somehow non-standard, and a more consistent use of terminology is appropriate.

5 One of the applications of the techniques described herein is to provide a functionality to discover and suggest relevant or missing medical concept terms to be possibly added to the report by looking at prior reports, being a source of a massive amount of medical practices information. The approach to create a set of associations 1 may comprise extracting key associations or correlations between the different pieces of information
10 contained in prior reports. The key correlations that are used in this approach may comprise correlations between the patient clinical history (body location(s), symptoms, signs, reason(s) for the exam, prior knowledge) and radiologists’ findings, reported anatomical regions, impressions and diagnosis. Other significant correlations between different findings within the report or between findings and diagnosis can be identified as well and used.

15 An example of such a correlation is the following. Consider a CT scan for a patient complaining of “tinnitus”. Analysis of reports may show that it is customary for a radiologist to describe the patient’s “auditory canals”. In some cases, this may be to report no problems in that area, while in others a specific pathology is reported. However, a physician who has ordered this CT scan may reasonably expect to receive the CT report and find some
20 indication of the status of the patient’s “auditory canals”.

 The indication of missing concept terms may be used in a scenario where the radiologist is about to submit the patient case report and asks the system to perform a checking analysis (post-reporting checking). It is also envisioned to suggest possibly relevant medical terms earlier, during the report generation process itself as the radiologist is typing or
25 dictating a report (with speech-to-text technology), or even earlier during the reading, interpretation process where recommendations can be suggested. As an example, giving recommendations to a resident in training can offer access to knowledge the resident does not already know or possess.

 Following are brief descriptions of features in the context of an example of an
30 envisioned scenario:

 Post reporting checking: Suggesting medical terms or concepts to improve the clinical quality of the report when the report is completed. In this scenario the objective is to provide relevant, missing terms to address consistency, completeness and accuracy of each produced report. Given the clinical history information (body location(s), symptoms, signs,

reason(s) for the exam, prior knowledge) for the current patient study, the system may suggest medical concept terms that are typically associated with this particular patient history but are not found in the final report. The radiologist can then read the suggestions and add terms to the report if necessary. The missing terms can be medical concepts related to findings, anatomical regions, impressions and diagnosis. In particular, consistency is addressed not necessarily by establishing a set of rules on reporting practices (from a given institution for example) that physicians have to comply to. It is done by providing a tool that seamlessly suggests medical terms without having to remember which rules apply to this particular patient case.

Reporting assistance: Suggesting medical terms during the creation of the patient report. As the physician is generating the report, medical terms may be displayed which are known to be associated with the available clinical history terms. At the beginning of the reporting, a list of terms is suggested. The radiologist can decide to look at these terms or not. As he is typing or dictating the report, more input terms become available to the system, allowing for more context-focused suggestions of medical terms (correlations between findings or between findings and diagnosis are used). It is also possible to provide an auto-completion function as the simple typing of the beginning of a word narrows down dramatically the possible terms to be proposed. In such a case, the system might have more influence in the selection of terms but offers the possibility to further clarify the description and, perhaps, to hint the radiologist to trigger a course of investigation not considered yet.

Region recommender through smart sorting during reading: The radiologist often knows which image regions to investigate to look for patterns or specific, abnormal visual characteristics, given the clinical history. One of the purposes may be to promote the most probable region terms, based on what has been reported in previous patient studies for a similar clinical history, and consequently avoid regions being overlooked. It also promotes unseen discovery of new links between for example a symptom and an anatomical region, that was not necessarily taught in medical textbooks or training.

The set of associations 1 may be created based on an analysis of prior reports. Such analysis may comprise the step of extraction of relevant pieces of information from radiology reports. Radiology reports are artifacts and end result of the radiology reading exercise. A radiology report contains the clinical history of the patient (body location(s), symptoms, signs, reason(s) for the exam, prior knowledge) and the radiologist's findings, relevant anatomy locations and conclusion with current diagnosis for the patient. The report

can vary from well structured (form) to less structured (most of the cases today) with sections containing free text.

In the case of reports with free text, existing natural language processing (NLP) techniques can be used to automatically extract the relevant pieces of information in the form of medical concepts (or terms).

As an illustrative example of clinical history: "Patient with history of neurofibromatosis and hypothalamic astrocytoma. Patient with severe headaches on the right side where the shunt is located." In this case, the system may identify four key clinical history terms: astrocytoma, headache, shunt and neurofibromatosis. From the same report the system may extract medical concepts (terms) from both the findings and conclusion sections: basal ganglia, dentate nucleus, frontal lobe, hypothalamus, nerve, optic chiasm, septum pellucidum, soft tissue, third ventricle.

The analysis of prior reports may further comprise identification of the associations, correlations of medical terms. In this step, relevant statistics may be collected.

These statistics may comprise the frequency of finding a given medical concept within all the reports and/or the co-occurrence frequency between two different types of medical concepts (a co-occurrence being defined as two different terms both appearing at least once in a single report). For example, the frequency with which a concept term appears in a clinical history may be determined as a term frequency, and the frequency with which a first concept term in a clinical history co-occurs with a second concept term in a finding/diagnosis may be determined as a co-occurrence frequency. Accordingly, the counting of the co-occurrences may be restricted based on the location of the terms in the report. For example, a co-occurrence between term A and term B may be counted only if term A appears in the "History" section of the report and term B appears only in the "Findings" section of the report. However, this is not a limitation. The counting of the co-occurrences may be weighted by whether a positive finding was detected. Moreover, the counting of the co-occurrences may be weighted by the seniority of the reading radiologist.

The analysis of prior reports may further comprise selection of significant associations, correlations of medical terms. Some of the identified correlations may be insignificant or just random. In an example embodiment, a Fisher's exact test is applied and the obtained p-value is used to sort the list of associations based on their significance, ranging from the most significant to the most probably random ones at the bottom of the list. Other statistical tests, such as Chi-square tests, are equally applicable.

As an example, consider a set of 800 reports, in which term A appears at least once in every 70 reports, and term B appears at least once in every 10 reports. If these terms appeared randomly in the reports (no real meaningful association), then on average one would expect about 2 out of the 800 reports to contain both term A and B. If, instead, one observes 8, or 9, or 10 reports with both terms, then there may be a good chance that this is a real meaningful association. A variety of well known statistical tests are available to discover this situation in a rigorous manner. The choice of a specific test is not meant to limit the invention.

The analysis of prior reports may further comprise estimation of conditional probabilities linking medical terms. In an example embodiment, the significant associations are found using conditional probabilities. The probabilities of interest here may include:

$P(\text{finding term} \mid \text{clinical history term})$, and/or

$P(\text{diagnostic term} \mid \text{clinical history term})$.

More generally, the conditional probability may be formulated as:

$P(\text{second concept term} \mid \text{first concept term})$.

A conditional probability may be estimated by taking the ratio of the co-occurrence frequency and the first concept term frequency using the Bayes' equation as follows:

$P(\text{second concept term} \mid \text{first concept term}) =$

$P(\text{first concept term AND second concept term}) / P(\text{first concept term})$.

The co-occurrence frequency may be used to estimate $P(\text{first concept term AND second concept term})$. The occurrence frequency of the first concept term may be used to estimate $P(\text{first concept term})$. The same principle may be applied to compute $P(\text{finding term} \mid \text{clinical history term})$ and/or $P(\text{diagnostic term} \mid \text{clinical history term})$.

It is also possible to estimate multi-term conditional probabilities, such as $P(\text{finding terms} \mid \text{clinical history term a, clinical history term b})$ or higher order. For simplicity purposes, the rest of the description will use the single-term probabilities but a higher order probability can be applied in a similar manner.

When analyzing a current report, recommendations may be generated using the set of associations 1. At the same time, the set of associations 1 may be updated by the association generator 14, for example by updating the conditional probabilities discussed above.

Several different usage schemes may be implemented for the indication of a suggestion of missing concept terms.

In a first usage scheme, post-reporting checking is used. Missing concept terms are indicated after the report has been completed. If missing concept terms are found, the user may decide to update the completed report. For example, for each first concept term (or a combination of first concept terms), the system may use the pre-computed list of significant associations to find a relevant second concept term or terms. A threshold may be determined in advance or chosen by the user (e.g. interaction through a slider mechanism) to limit the list of associations. The completed report may then be analyzed to extract the contained concept terms using NLP techniques. This list of terms from the report is compared to the terms corresponding to the significant associations to only display the missing terms as suggestions. An explanation of why a particular term is suggested may also be indicated to the user, for example by showing the first concept term and the strength of the association.

The displayed list of suggested terms can be sorted based on the conditional probabilities calculated, or based on a strength measure of the association computed in a different way. The post-reporting checking can be done right after completion of the report or off-line. In the latter case, the flagged reports having missing terms can be brought later to the attention of the radiologist or selected for peer review.

Other types of correlations can be identified, such as those between two findings concepts or between a finding concept and a diagnostic concept. These concepts were produced by the radiologist and bring key contextual information that could be used to attach more weight to the medical concept terms and narrow down the list of suggested terms. As an example, if a suggested term is a finding “a” and if another finding “b” is already in the report, a high probability $P(\text{finding term “a”} \mid \text{finding term “b”})$ may promote the finding ‘a’ towards the top of the list of suggested terms.

In a second usage scheme, reporting assistance is provided by suggesting medical terms during the creation of the report. This scenario differs from the previous one as the suggested terms are offered on the fly when the radiologist is typing or dictating the report. This gives the doctor the possibility to go back to images for further scrutiny. This functionality can be used, for example, for training purposes, either for resident physicians or for continuous medical education.

In a third usage scheme, a region recommender based on image viewer helps the user to quickly find relevant image portions corresponding to the missing second concept terms. Based on the first concept terms, for example extracted from a clinical history leading the imaging exam, the user may be presented with a list of regions using the method described above. The interaction with the image may comprise the use of image processing

techniques to segment the image (mesh model) and an atlas to map each region with the corresponding region concepts. As a result, upon clicking on one of the suggested regions in the list (the list may be sorted by relevance using the conditional probabilities), the corresponding region may be highlighted.

Fig. 3 illustrates another example of a system for aiding report authoring. The components of this system are to a large extent similar to the system depicted in Fig. 1. Prior reports from a database 301 may be analyzed using an extraction subsystem 302 which converts the raw text into extracted report terms 303. This extraction subsystem 302 may be similar to the concept term extractor 2, and arranged for being applied to the prior reports from database 301. The correlation discovery subsystem 304 looks over the extracted report terms 303 from a large number of reports in the database and identifies the significant associations. These can be stored conceptually in a database of significant correlations 305, similar to the set of associations 1. The correlation discovery subsystem 304 may be similar to the association generator 14. When a new report 306 (similar to the at least partially completed report 5) is being considered, the same or a different extraction subsystem 307 (which may be similar to the concept term extractor 2) is executed to convert the raw report 306 (e.g. the patient history) into extracted report terms 308. These extracted report terms 308 may be similar to the at least one first concept term 6. These extracted report terms 308 are entered into a recommendation subsystem 309 (which may be similar to the missing concept finder 3) which uses the discovered term associations or correlations 305 to produce a list of new recommended terms 310, which may be similar to the at least one second concept term 11. These new recommended terms 310 may be displayed via display subsystem 311 (which may be similar to the indicator 4) by applying one or more of the functionalities described above. Note that the various subsystems and databases mentioned in this description may be part of a single physical computer system or of multiple physical computer systems.

It will be appreciated that the invention also applies to computer programs, particularly computer programs on or in a carrier, adapted to put the invention into practice. The program may be in the form of a source code, an object code, a code intermediate source and object code such as in a partially compiled form, or in any other form suitable for use in the implementation of the method according to the invention. It will also be appreciated that such a program may have many different architectural designs. For example, a program code implementing the functionality of the method or system according to the invention may be sub-divided into one or more sub-routines. Many different ways of distributing the

functionality among these sub-routines will be apparent to the skilled person. The sub-routines may be stored together in one executable file to form a self-contained program. Such an executable file may comprise computer-executable instructions, for example, processor instructions and/or interpreter instructions (e.g. Java interpreter instructions). Alternatively, one or more or all of the sub-routines may be stored in at least one external library file and linked with a main program either statically or dynamically, e.g. at run-time. The main program contains at least one call to at least one of the sub-routines. The sub-routines may also comprise function calls to each other. An embodiment relating to a computer program product comprises computer-executable instructions corresponding to each processing step of at least one of the methods set forth herein. These instructions may be sub-divided into sub-routines and/or stored in one or more files that may be linked statically or dynamically. Another embodiment relating to a computer program product comprises computer-executable instructions corresponding to each means of at least one of the systems and/or products set forth herein. These instructions may be sub-divided into sub-routines and/or stored in one or more files that may be linked statically or dynamically.

The carrier of a computer program may be any entity or device capable of carrying the program. For example, the carrier may include a storage medium, such as a ROM, for example, a CD ROM or a semiconductor ROM, or a magnetic recording medium, for example, a hard disk. Furthermore, the carrier may be a transmissible carrier such as an electric or optical signal, which may be conveyed via electric or optical cable or by radio or other means. When the program is embodied in such a signal, the carrier may be constituted by such a cable or other device or means. Alternatively, the carrier may be an integrated circuit in which the program is embedded, the integrated circuit being adapted to perform, or used in the performance of, the relevant method.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. Use of the verb "comprise" and its conjugations does not exclude the presence of elements or steps other than those stated in a claim. The article "a" or "an" preceding an element does not exclude the presence of a plurality of such elements. The invention may be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In the device claim enumerating several means, several of these means may be embodied by one and the same item of hardware. The mere fact that certain

measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

CLAIMS:

1. A system for aiding report authoring, comprising
 - a set of associations (1), wherein an association links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a frequent co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain;
 - a concept term extractor (2) for extracting at least one first concept term (6) relating to a particular subject from a data record (7) associated with the particular subject, the data record (7) comprising an at least partially completed report (5) relating to the particular subject;
 - a missing concept term finder (3) for finding at least one second concept term (11) linked to the at least one first concept term (6) based on the set of associations (1), wherein the at least one second concept term (11) is missing from the at least partially completed report (5);
 - an indicator (4) for indicating the found at least one second concept term (11) to a user.
2. The system according to claim 1, wherein the knowledge domain comprises a medical knowledge domain, the subject comprises a patient, and the report comprises a medical diagnosis or treatment report.
3. The system according to claim 1, wherein the concept term extractor (2) is arranged for extracting the at least one first concept term (6) from the at least partially completed report (5).
4. The system according to claim 3, wherein the concept term extractor (2) is arranged for extracting the at least one first concept term (6) from a first section of the at least partially completed report (5), and wherein the set of associations (1) comprises associations linking first concept terms belonging to the first section to second concept terms belonging to a different second section of the reports in the knowledge domain.

5. The system according to claim 1, wherein the concept term extractor (2) is arranged for extracting the at least one first concept term (6) from medical history information (8) represented by the data record (7).

5

6. The system according to claim 1, wherein the at least one second concept term (11) is associated with a particular section of the report, and the indicator (4) is arranged for indicating with which section the at least one second concept term (11) is associated.

10 7. The system according to claim 1, wherein the system comprises a user interface element (9) for enabling a user to indicate that the report (5) is complete, and wherein the indicator (4) is arranged for indicating the found at least one second concept term (11) in response to a user indication that the report (5) is complete.

15 8. The system according to claim 1, comprising a report creation tool (10) for enabling a user to create at least part of the report (5), and wherein the indicator (4) is arranged for indicating the found at least one second concept term (11) during a process of creating the at least part of the report (5).

20 9. The system according to claim 1, wherein the at least one second concept term (11) comprises an anatomical region term, and wherein the system further comprises an image viewer (15) configured to indicate a view of an anatomical region of the subject based on the anatomical region term, wherein the anatomical region corresponds to the anatomical region term.

25

10. The system according to claim 1, further comprising a second concept term classifier (12) for classifying the at least one second concept term (11) by type of concept, and wherein the indicator (4) is arranged for indicating the type of concept for a second concept term (11).

30

11. The system according to claim 1, wherein the associations of the set of associations (1) have different strengths, and further comprising a second concept term sorter (13) for sorting the at least one second concept term (11) found by the missing concept term

finder (3), based on the strengths of the associations by which the at least one second concept term (11) was found.

12. The system according to claim 1, further comprising an association generator (14) for creating an association based on co-occurrence frequencies of first and second concept terms in a collection of reports.

13. A workstation comprising the system according to claim 1.

14. A method of aiding report authoring, comprising
- extracting (201) at least one first concept term relating to a particular subject from a data record associated with the particular subject, the data record comprising an at least partially completed report relating to the particular subject;

- finding (202) at least one second concept term linked to the at least one first concept term based on a set of associations, wherein the at least one second concept term is missing from the at least partially completed report, wherein an association of the set of associations links at least one first concept term in a knowledge domain with at least one second concept term in the knowledge domain, wherein the association indicates a frequent co-occurrence of the at least one first concept term and the at least one second concept term in reports in the knowledge domain; and

- indicating (203) the found at least one second concept term to a user.

15. A computer program product comprising instructions for causing a processor system to perform the method according to claim 14.

1/3

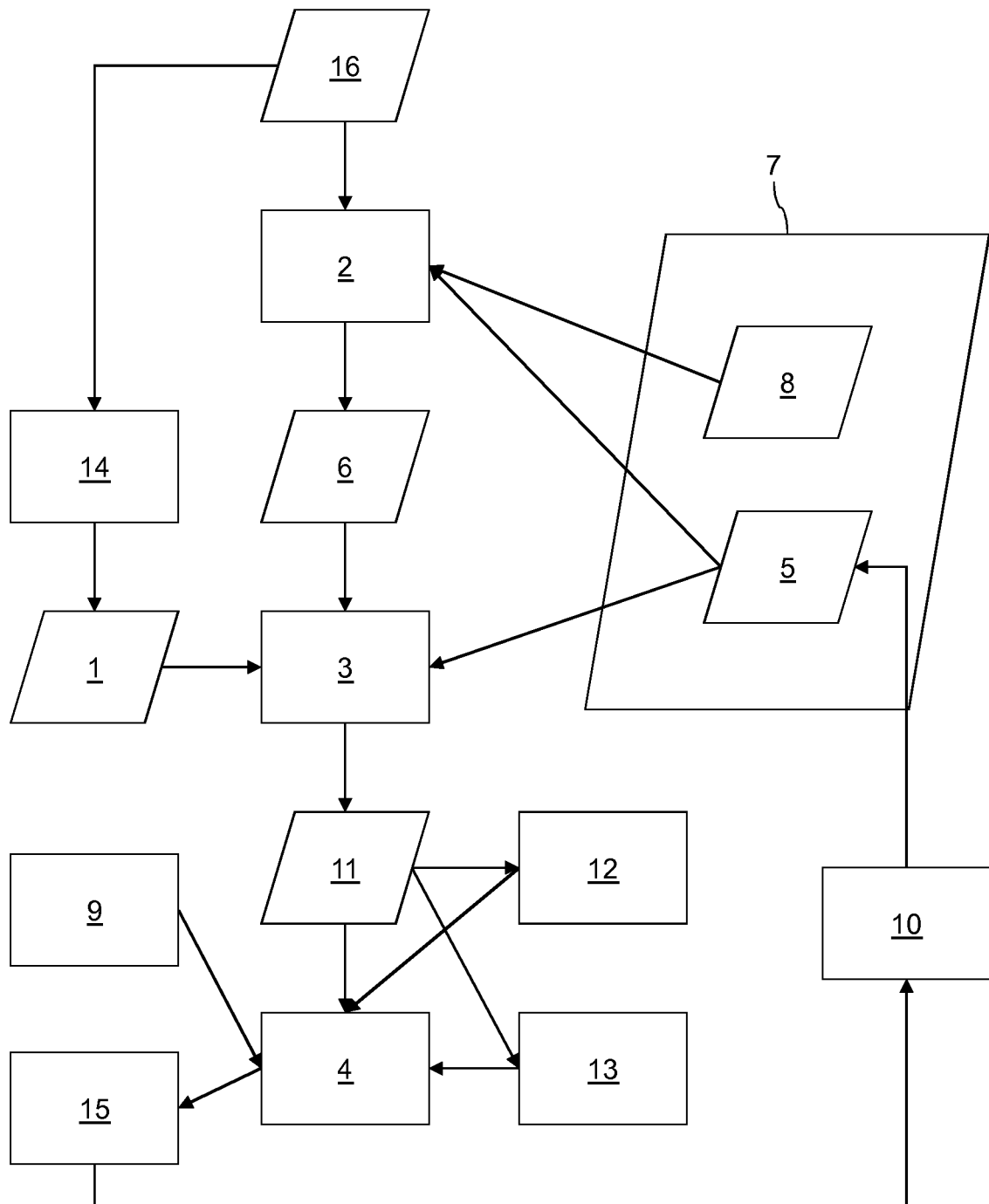


Fig. 1

2/3

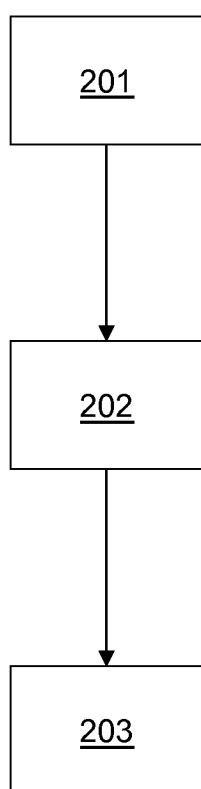


Fig. 2

3/3

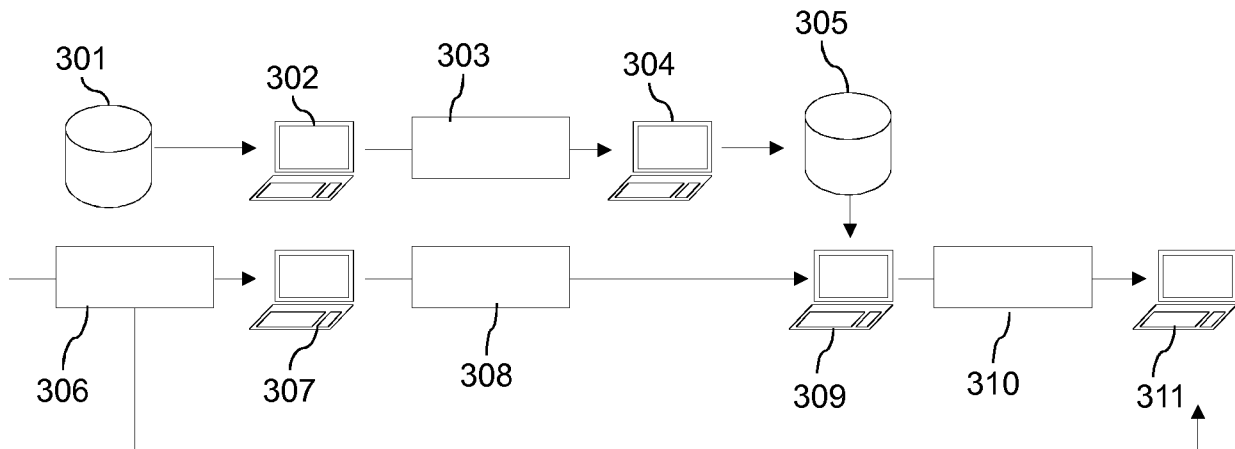


Fig. 3

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2011/053509

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F19/00

ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005/065813 A1 (MISHELEVICH DAVID J [US] ET AL) 24 March 2005 (2005-03-24) paragraph [0009] - paragraph [0015] paragraph [0042] paragraph [0051] - paragraph [0064] paragraph [0082] - paragraph [0083] paragraph [0106] - paragraph [0110] paragraph [0118] - paragraph [0122] paragraph [0130] - paragraph [0149] figures 1-25 -----	1-15
X	US 6 047 259 A (CAMPBELL SCOTT DOUGLAS [US] ET AL) 4 April 2000 (2000-04-04) column 1, line 48 - column 3, line 18 column 15, line 53 - column 18, line 16 figures 1-14 ----- -/-	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

19 December 2011

Date of mailing of the international search report

29/12/2011

Name and mailing address of the ISA/

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

Authorized officer

Hernández Marugán, J

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2011/053509

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EP 2 120 170 A1 (ALGOTEC SYSTEMS LTD [IL]) 18 November 2009 (2009-11-18) paragraph [0108] paragraph [0154] -----</p>	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2011/053509

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005065813	A1	24-03-2005	NONE
US 6047259	A	04-04-2000	CA 2227188 A1 30-06-1999 US 6047259 A 04-04-2000
EP 2120170	A1	18-11-2009	NONE