

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

(Chapter II of the Patent Cooperation Treaty)

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference P01042	FOR FURTHER ACTION		See Form PCT/IPEA/416
International application No. PCT/IB15/51239	International filing date (day/month/year) 18 February 2015 (18.02.2015)	Priority date (day/month/year)	
International Patent Classification (IPC) or national classification and IPC IPC: G06N 5/04(2006.01) USPC: 706/46,707/E17.033,707/E17.008,707/752,382/321			
Applicant SERVICIOS DIGITALES WEBDOX SPA			
1. This report is the international preliminary examination report, established by this International Preliminary Examining Authority under Article 35 and transmitted to the applicant according to Article 36. 2. This REPORT consists of a total of <u>5</u> sheets, including this cover sheet. 3. This report is also accompanied by ANNEXES, comprising: a. ☒ (sent to the applicant and to the International Bureau) a total of <u>7</u> sheets, as follows: ☒ sheets of the description, claims and/or drawings which have been amended and/or sheets containing rectifications authorized by this Authority, unless those sheets were superseded or cancelled, and any accompanying letters (see Rules 46.5, 66.8, 70.16, 91.2, and Section 607 of the Administrative Instructions). ☐ sheets containing rectifications, where the decision was made by this Authority not to take them into account because they were not authorized by or notified to this Authority at the time when this Authority began to draw up this report, and any accompanying letters (Rules 66.4bis, 70.2(e), 70.16 and 91.2). ☐ superseded sheets and any accompanying letters, where this Authority either considers that the superseding sheets contain an amendment that goes beyond the disclosure in the international application as filed, or the superseding sheets were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in item 4 of Box No.I and the Supplemental Box (see Rule 70.16(b)). b. ☐ (sent to the International Bureau only) a total of (indicate type and number of electronic carrier(s)) _____ containing a sequence listing, in electronic form only, as indicated in the Supplemental Box Relating to Sequence Listing (see paragraph 3bis of Annex C of the Administrative Instructions).			
4. This report contains indications relating to the following items: ☒ Box No. I Basis of the report ☐ Box No. II Priority ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability ☐ Box No. IV Lack of unity of invention ☒ Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement ☐ Box No. VI Certain documents cited ☐ Box No. VII Certain defects in the international application ☐ Box No. VIII Certain observations on the international application			
Date of submission of the demand 07 December 2015 (07.12.2015)		Date of completion of this report 03 May 2016 (03 05 2016)	
Name and mailing address of the IPEA/ US Mail Stop PCT, Attn: IPEA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201		Authorized officer Scott Baderman Telephone No. 11111	

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.

PCT/IB15/51239

Box No. 1 Basis of the report1. With regard to the **language**, this report is based on:

- ☒ the international application in the language in which it was filed.
- ☐ a translation of the international application into English which is the language of a translation furnished for the purposes of:
- ☐ international search (Rules 12.3(a) and 23.1(b)).
- ☐ publication of the international application (Rule 12.4(a)).
- ☐ international preliminary examination (Rules 55.2(a) and/or 55.3(a) and (b)).

2. With regard to the **elements** of the international application, this report is based on (*replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report*):

- ☐ the international application as originally filed/furnished
- ☒ the description:
 pages 1-15 as originally filed/furnished
 pages* NONE received by this Authority on _____
 pages* NONE received by this Authority on _____
- ☒ the claims:
 pages NONE as originally filed/furnished
 pages* NONE as amended (together with any statement) under Article 19
 pages* 16-18 received by this Authority on 07 December 2015
- ☒ the drawings:
 pages 1/9 - 9/9 as originally filed/furnished
 pages* NONE received by this Authority on _____
 pages* NONE received by this Authority on _____
- ☐ a sequence listing - see Supplemental Box Relating to Sequence Listing.

3. ☐ The amendments have resulted in the cancellation of:

- ☐ the description, pages _____
- ☐ the claims, Nos. _____
- ☐ the drawings, sheets/figs _____
- ☐ the sequence listing (*specify*): _____

4. ☐ This report has been established as if (some of) the amendments annexed to this report and listed below had not been made, since either they are considered to go beyond the disclosure as filed, or they were not accompanied by a letter indicating the basis for the amendments in the application as filed, as indicated in the Supplemental Box (Rules 70.2(c) and (c-bis)):

- ☐ the description, pages _____
- ☐ the claims, Nos. _____
- ☐ the drawings, sheets/figs _____
- ☐ the sequence listing (*specify*): _____

5. ☐ This report has been established:

- ☐ taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rules 66.1(d-bis) and 70.2(e)).
- ☐ without taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rules 66.4bis) and 70.2(e)).

6. ☒ With regard to top-up searches (Rules 66.1ter and 70.2(f)):

- ☒ A top-up search was carried out by this Authority on 03 May 2016 (03.05.2016)
- ☒ Additional relevant documents have been discovered during the top-up search.
- ☐ No top-up search was carried out by this Authority because it would serve no useful purpose.

7. ☐ Supplementary international search report(s) from Authority(ies) _____ has/have been received and taken into account in establishing this report (Rule 45bis.8(b) and (c)).

* If item 4 applies, some or all of those sheets may be marked "superseded."

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/IB15/51239**Box No. V Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement****1. Statement**

Novelty (N)	Claims <u>NONE</u>	YES
	Claims <u>1-4</u>	NO
Inventive Step (IS)	Claims <u>NONE</u>	YES
	Claims <u>1-4</u>	NO
Industrial Applicability (IA)	Claims <u>NONE</u>	YES
	Claims <u>NONE</u>	NO

2. Citations and Explanations (Rule 70.7)

Please See Continuation Sheet

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/IB15/51239

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.
Continuation of:

V. 2. Citations and Explanations:

Claims 1-4 lack an inventive step under PCT Article 33(3) as being obvious over US 2007/0206884 (Kato) in view of US 2010/0046842 (Cornwell) in view of US 2011/0196870 (Schmidtler).

Regarding claim 1, Kato discloses a method of automatic document characterization (para [0032]; "The document type determination section 4") which receives a given input unstructured document (para [0027]; "The input document image 100 may be image data obtained by scanning a paper document using a scanner (not shown).") and results in the automatic assignment of one or more document classes or categories to which the contents relates to (para [0032]; "The document type determination section 4 refers to the feature information of each document type that is stored in the document information DB 5 to specify the document type that matches most closely the result of layout analysis and the result of character recognition. As determination of a document type is also a well-known technique, further description is not provided herein."), an automatic determination of the document issue date (para [0055]; The list of features of item values that are likely to be extraction targets may be similar to the category information as illustrated in Fig. 7. The name of the categories (e.g. "date" and "personal name") that correspond to the features of an item value may be used as information to find a character string indicating an item name

Supplemental Box

from the vicinity of an item value"), and an automatic determination of a list of the names of natural or legal persons found in the text (para [0055]; "The list of features of item values that are likely to be extraction targets may be similar to the category information as illustrated in Fig. 7. The name of the categories (e.g. "date" and "personal name") that correspond to the features of an item value may be used as information to find a character string indicating an item name from the vicinity of an item value") and other relevant information mentioned in the text, which is related with a document class or multiclass (para [0057];

"Specifically, the extraction item proposal section 7 retrieves a character string having the feature that is stored in the candidate item DB 8 from the character recognition results."), faster, more complete and more accurate than the manual characterization or manual description made by legal technical personnel, wherein this speed and accuracy comparison is feasible when implementing the process as a computing system (para [0062]; The above-described electronic filing device may be realized by executing a program that describes the functions or processing of the above functional modules on a general-purpose computer."), the method comprises the steps of: a. Upon receiving a digital document, from a user connected from his own computer application (para [0027]; "Further, a document image 100 that is created by application software may be input to the electronic filing device instead of or in addition to the scanned document image 100. "), process it with an OCR application (para [0030]; " The OCR section 3 performs OCR on the image of the text area "); b. Then, execute an automatic document characterization process within the application and process the text of the document received from the user (para [0032]; The document type determination section 4 refers to the feature information of each document type that is stored in the document information DB 5 to specify the document type that matches most closely the result of layout analysis and the result of character recognition. As determination of a document type is also a well-known technique, further description is not provided herein."), wherein this involves, first the refinement of the text using a dictionary of valid words, then the revision of occurrence of big number of different logic rules, that are associated to classes, by recognizing any of many different language expressions and set of keywords, resulting in a score for each of the rules that allow determining which classes apply through a means of specification certain logic text-related rules (para [0038]; " If the natural language analysis section 63 finds a character string that completely matches the item name received from the item name extraction section 61, it can determine the character string to be an item name of an extraction target. If the natural language analysis section 63 does not find the one that completely matches, it can replace each word that composes the item name with a synonym to form a variation of an item name and search for a character string that matches the variation from the character recognition results. If the natural language analysis section 63 finds a character string that completely matches the variation, it can determine the character string to be an item name of an extraction target. It is possible that plural character strings may match plural different variations might be found from the character recognition results. In such a case, the string with highest similarity to the item name of an extraction target (which can be calculated from the similarity of words composing an item name and words composing the variation) may be selected from among the variations found from the character recognition results. It is also possible that an item name might appear as a different character string due to declension or change in the part of speech of a word composing an item name. In such a case, the natural language analysis section 63 may perform morphological analysis, syntactic analysis, and semantic analysis on each character

Supplemental Box

string in the character recognition results and extract a character string with high semantic similarity to an item name of an extraction target.");

c. Recognizing names of peoples (para [0048]; The category 'personal name' is an arrangement of a string that is generally known as a surname and a string that is generally known as a given name in a prescribed order, which depends on language.) and organizations (para [0048]; The category 'organization name' is a name of an organization such as a company or a department.") in the document contents by using a wide variety of different possible verbalizations of these names and according to the text (para [0037]; "During this analysis, the natural language analysis section 63 refers to the item name dictionary DB 64. The item name dictionary DB 64 stores information of synonyms and quasi-synonyms of the words that are likely to be contained in the item name of an extraction target"), determine which of these names are relevant to the contents and which only references through a detection means (para [0039]; "The item name extraction section 61 supplies the item name that is extracted from the character recognition results and the information of the position (or existence range) of the character string that is relevant to the item name in the document image 100 to the item value extraction section 62°); d. Recognizing and extracting relevant information based on the class or multiclass definition made and the set of rules defined for that class or multiclass, through the detection means (para [0042]; The information item extraction section 6 uses the above functional modules 61 to 64 to extract a value of each information item of an extraction target from the character recognition results in accordance with the document type of the document image 100.); e. Recognizing relevant dates in the document and assigning a score to each of them according to the preceding or following text, associated with different keywords and language expressions that used to refer to document issue dates, taking into account the previously defined class, and by using the defined rules of the same class, through the detection means, related to the class previously identified and the set of rules used to classify as such (para [0041]; "A character string that is distant from the item name "acquired date" is less likely to contain the value "acquired date", even when it is at the "lower right" of the item name. It is therefore necessary to determine whether or not the character string is in "close proximity" to the item name. If a field of an item name and a field of an item value are divided by a ruled line, it can be determined that the character strings in those fields would be in close proximity to each other when the field of the item name and the field of the item value are in close proximity to each other, as "drafted date" and "name of fixed asset" in the example of Fig. 1. On the other hand, if an item name and an item value are not divided by a ruled line such as "acquired date", a character string within a given threshold distance from the area of the item name may be determined to be a character string in "close proximity", for example. The threshold distance may be variable according to the character size of the item name so that the range of "close proximity" is made wider when the character of an item name is larger."); and f.

Reviewing the document contents, recognizing different text patterns as a combination of keywords, synonyms or equivalent terms, within the contents of the document through detection means (para [0037]; "During this analysis, the natural language analysis section 63 refers to the item name dictionary DB 64. The item name dictionary DB 64 stores information of synonyms and quasi-synonyms of the words that are likely to be contained in the item name of an extraction target, such as, for example, information that "year and month", "year, month and da/1 and "date" are quasi-synonyms. The item name dictionary DB 64 may also

Supplemental Box

store a numerical value indicating a similarity of quasisynonyms").

Kato does not explicitly teach that application is a Web application. Conwell teaches using a web application (para [0300]; The computers are networked, shown schematically by link 50. This connection can be by any known networking arrangement, including the internet and/or wireless links (WiFi, WiMax, Bluetooth, etc), Software in at least certain of the computers includes a peer-to-peer (P2P) client, which makes that computer's resources available to other computers on the network, and reciprocally enables that computer to employ resources of the other computers."). It would have been obvious to a person of ordinary skill in the art to use the web (internet) application as taught by Conwell with the document characterization of Kato. The motivation for doing so would have been by Conwell in (para [0300]; "which makes that computer's resources available to other computers on the network, and reciprocally enables that computer to employ resources of the other computers").

Kato does not explicitly teach using rules with their calibrated coefficients. Conwell teaches using weighting the metadata, (para [0195], [0201]; "wavelet-domain transform coefficient histograms, and dominant wavelet coefficients; frequency domain transfer coefficient histograms and dominant frequency coefficients"; "(The weighting can be linear or exponential.) Since metadata from similar images is weighted more than metadata from dissimilar images, the resulting set of metadata is tailored to more likely correspond"). It would have been obvious to a person of ordinary skill in the art to use the weighting as taught by Conwell with the document characterization of Kato. The motivation for doing so would have been as Conwell states (para [0201]) that the weighting system makes it more likely that similar results are together.

Kato does not explicitly teach legal documents and automatic detection of essential or legal-imperative rules of each legal document, determining if the legal document is correctly classified leaving aside only high-accurate statistic detection of the document class. However, Schmidtler discloses legal documents and automatic detection of essential or legal-imperative rules of each legal document, determining if the legal document is correctly classified leaving aside only high-accurate statistic detection of the document class (paragraphs 0007-0013). It would have been obvious to one of ordinary skill in the art to perform classification on of legal documents taught by Schmidtler with Kato-Conwell in order to accurately classify documents.

Regarding claim 2, Kato, Conwell, and Schmidtler teaches the method of claim 1, Kato further teaches wherein the rules are defined using a configuration application that allows the user to specify a set of keywords or synonyms, number of occurrences in the contents, and/or usage within certain phrases, thus defining the rules (para [0059]; " Fig. 9 illustrates an example of a display for editing an extraction item that is presented to a user. This display contains a list of pairs of candidates for item names and candidates for item values that are extracted from the document image 100. Accompanying each displayed pair is a check box 90 which a user can use to select, separately for each pair, whether or not that set of items should be designated as an extraction target.").

Regarding claim 3, Kato, Conwell, and Schmidtler teaches the method of claim 1, Conwell further teaches wherein the characterization process is calibrated using automatic specification that defines a weight coefficient and value of the elements on each rule (para [0191]-[0204]; "weighted set of metadata").

Regarding claim 4, Kato, Conwell, and Schmidtler teaches the method of claim 1, Kato further teaches wherein valid language words

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY

International application No.
PCT/IB15/51239

Supplemental Box

and its synonyms list to be used in refining the text in a given document in digital format resulted from an optical character recognition process (para [0030], [0060] -10061); "The OCR section 3 performs OCR on the image of the text area in the preprocessed image to thereby obtain text information from the text area. As is well known, the OCR processing recognizes each character contained in the text area. It is also possible to locate the position of each character in the document image 100. The description on the OCR processing is not detailed herein because it may be performed using well-known techniques."; "A character string that indicates a document name or heading may be extracted from the document image 100 that contains extraction item information using a conventional technique, may be stored in the document information OB 5 as feature information indicating the type of document to which the extraction item information is applied. The information of a positional relationship of each area acquired from the document image 100 by the layout analysis section 2 may be stored in the document information DB 5 as feature information corresponding to the type of the document."; "After the extraction item information is created and stored in the document information OB 5, the electronic filing device may automatically extract required information items from an image of the same type of document using the created information.').

Claims 1-4 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry

----- NEW CITATIONS -----

US 2011/0196870 A1 (Schmidtler et al.) 11 August 2011

CLAIMS

1. A method of automatic document characterization which receives a given input unstructured document, specifically legal documents, and results in the automatic assignment of one or more document classes or categories to which the contents
5 relates to, an automatic detection of essential or legal-imperative rules of each legal document, determining if the legal document is correctly classified leaving aside an only high-accurate statistic detection of the document class, an automatic determination of the document issue date, and an automatic determination of a list of the names of natural or legal persons found in the text and other relevant information
10 mentioned in the text, which is related with a document class or multiclass, faster, more complete and more accurate than the manual characterization or manual description made by legal technical personnel, wherein this speed and accuracy comparison is feasible when implementing the process as a computing system, the method comprises the steps of:
- 15 a. Upon receiving a digital document, from a user connected from his own computer in a Web application, process it with an OCR application;
- b. Then, execute an automatic document characterization process within the application and process the text of the document received from the user, wherein this involves, first the refinement of the text using a dictionary of valid words, then the
20 revision of occurrence of big number of different logic essential rules with their calibrated coefficients, that are associated to classes, by recognizing any of many different language expressions and set of keywords, resulting in a score for each of

the essential rules that allow determining which classes apply through a means of specification certain logic text-related rules;

c. Recognizing names of peoples and organizations in the document contents by using a wide variety of different possible verbalizations of these names and according
5 to the text, determine which of these names are relevant to the contents and which only references through a detection means;.

d. Recognizing and extracting relevant information based on the class or multiclass definition made and the set of essential rules defined for that class or multiclass, through the detection means;

10 e. Recognizing relevant dates in the document and assigning a score to each of them according to the preceding or following text, associated with different keywords and language expressions that used to refer to document issue dates, taking into account the previously defined class, and by using the defined essential rules of the same class, through the detection means, related to the class previously identified and the set of
15 essential rules used to classify as such; and

f. Reviewing the document contents, recognizing different text patterns as a combination of keywords, synonyms or equivalent terms, within the contents of the document through detection means, establishing a legal-imperative automatic revision of each legal document, using the essential rules.

20

2. The method of claim 1 wherein the rules are defined using a configuration application that allows the user to specify a set of keywords or synonyms, number of

occurrences in the contents, and/or usage within certain phrases, thus defining the rules.

3. The method of claim 1 wherein the characterization process is calibrated using automatic specification that defines a weight coefficient and value of the elements on
5 each rule, essentials and not essentials.

4. The method of claim 1 wherein valid language words and its synonyms list to be used in refining the text in a given document in digital format resulted from an optical character recognition process.



SILVA & CIA
ABOGADOS
Propiedad Intelectual - Tecnología

December 07th, 2015

VIA e-PCT
USPTO

REQUEST OF IPRP, ACCORDING TO CHAPTER II PCT

REF: PCT/IB2015/051239

Our Ref: P01042

Mandatory: Silva & Cia.

Title: **DOCUMENT CHARACTERIZATION METHOD**

The present declaration corresponds to a modifications and clarifications according to Article 34 of PCT.

1.- Amendments

Please find the followings amendments to the set of claims. The underlined words corresponds to the added content to the claims:

Claim 1:

1. A method of automatic document characterization which receives a given input unstructured document, specifically legal documents, and results in the automatic assignment of one or more document classes or categories to which the contents relates to, an automatic detection of essential or legal-imperative rules of each legal document, determining if the legal document is correctly classified leaving aside an only high-accurate statistic detection of the document class, an automatic determination of the document issue date, and an automatic determination of a list of the names of natural or legal persons found in the text and other relevant information mentioned in the text, which is related with a document class or multiclass, faster, more complete and more accurate than the manual characterization or manual description made by legal technical personnel, wherein this speed and accuracy comparison is feasible when implementing the process as a computing system, the method comprises the steps of:

- a. Upon receiving a digital document, from a user connected from his own computer in a Web application, process it with an OCR application;
- b. Then, execute an automatic document characterization process within the application and process the text of the document received from the user, wherein this involves, first the refinement of the text using a dictionary of valid words, then the revision of occurrence of big number of different logic essential rules with their calibrated coefficients, that are associated to classes, by recognizing any of many different language expressions and set of keywords, resulting in a score for each of the essential rules that allow determining which classes apply through a means of specification certain logic text-related rules;



SILVA & CIA
 ABOGADOS
 PROPIEDAD INTELECTUAL - Tecnología

- c. Recognizing names of peoples and organizations in the document contents by using a wide variety of different possible verbalizations of these names and according to the text, determine which of these names are relevant to the contents and which only references through a detection means;
- d. Recognizing and extracting relevant information based on the class or multiclass definition made and the set of essential rules defined for that class or multiclass, through the detection means;
- e. Recognizing relevant dates in the document and assigning a score to each of them according to the preceding or following text, associated with different keywords and language expressions that used to refer to document issue dates, taking into account the previously defined class, and by using the defined essential rules of the same class, through the detection means, related to the class previously identified and the set of essential rules used to classify as such; and
- f. Reviewing the document contents, recognizing different text patterns as a combination of keywords, synonyms or equivalent terms, within the contents of the document through detection means, establishing a legal-imperative automatic revision of each legal document, using the essential rules.

Claim 3:

The method of claim 1 wherein the characterization process is calibrated using automatic specification that defines a weight coefficient and value of the elements on each rule, essentials and not essentials.

2.- Arguments

Our invention defines different text-content rules as essential rules and non-essential rules (essential rules are described in the description, page 9 and in figure 2b), which are extremely useful in the preferred embodiment, legal documents. Legal documents are positioned on the realm of non structured documents, in difference of Kato (US 2007/0206884 A1) preferred embodiment, related with business documents like debit notes, that are contained in the sphere of structured and semi-structured documents.

Legal documents (like contracts, corporate matters, amongst others) have some content requirements depending on each document class. The essentiality of a rule is a distinctive characteristic related to required sections of each document class and a law-imperative content that might exist in a specific document. *(In some classes, a rule may be considered as essential, in which case, the coefficient should be high and even the absence of this rule might lead to a penalty regarding the document class in which this rule is considered essential/description, page 2).* The absence of an essential rule also allows the classifier to tell the user why a certain class is discarded for a specific document. The aforementioned aspect allows establishing that a document type determination is not only a statistic similarity between a document and another,



SILVA & CIA
ABOGADOS
 PROPIEDAD INTELECTUAL - Tecnología

but also this document must contain a law obligation or customary rule that should appear in a semantic, syntactic and morphological way.

Following are a few examples to illustrate this concept.

- A. For example, in a **Real State Lease Agreement** some essential elements exist (essential rules) that shall compose the legal document as naturally is. A first case of an essential rule in this type of document is the "*description of premises*", that mentions the details of premises location and the address of these. A semantic expression of this rule it can be "*the landlord leases to tenant the premises located at [ADDRESS], [CITY], [STATE], and described more particularly as follows: [LEGAL DESCRIPTION]*"; or other equivalent semantic expressions that could describe this element. Second, another essential rule in this kind of legal document is the "*payment of the rental by the tenant*", described in a semantic expressions as "*the tenant's rental to be paid to the landlord in an amount of [PRICE] shall have cleared in the landlord/agent's bank account on or before the first week of every consecutive month*". Notwithstanding, some different semantic expressions exist for the same element or rule.
- B. On the other hand, in a **Mortgage Agreement** one example of an essential rule is the "*description of property subject to lien*". A semantic expression of this rule it can be "*Now, therefore, in consideration of the premises and the sum hereinabove set forth, and to secure the payment of the secured indebtedness as defined herein, Mortgagor has granted, bargained, sold and conveyed and by these presents does grant, bargain, sell and convey unto Mortgagee property situated in [CITY, STATE/PROVINCE] more particularly described below*"; or other equivalent semantic expressions that may illustrate that rule.

The use of rules and the calculation of each rule's score in a particular document is also available as a report for an expert user in the preferred embodiment of this invention. This report provides an insight to the correctness of the rules defined and coefficients used, thus allowing corrections and refinements that may result in classification performance improvement. Also, this report allows the user to define if a specific rule, in terms of an essential rule, must be enhanced in relation to specific modifications of normatives interrelated, or should be improved in consideration to the words or phrases that might be used required by specific regulations.

Additionally, as described in the "Document class-related characterization" section of the application description, the characterization process defined in this invention is able to dynamically search specific attributes in a document, in a step by step process, depending on the rules detection (essentials and not essentials) and document class determination. This allows the determination of a much wider set of attributes from a document than Kato, extracting specific information or conclusions (made by a combination of rules) in attention of a specific matching of rules and coefficients related. Moreover, the system disclosed by Kato could detect



SILVA & CIA
ABOGADOS
PROPIEDAD INTELECTUAL - Tecnología

the attributes of a document that is a specific legal file in the field of statistic, but not in the law imperatives terms, adding the attribute associated that is not a specific legal document because it is missing some specific essential rules.

3.- Inventive Step

This invention differs from other by including the consideration of essential rules, which is relevant for legal documents, therefore, improving other simpler classification methods, only based in keyword occurrence, improving the data extraction in consideration of the essential rule detection. This feature, with the provided rule occurrence score report, allows handling a simple class determination, but also analyzing legal documents from the legal correctness point of view, behaving as a revision method, as well as a characterization method, Therefore, this deference will be not obvious for an expert in the art.

4. Conclusions

The present invention, especially considering the preferred embodiment in the realm of legal documents, goes beyond previously known document classification methods, by considering content-related requirements of the documents, and therefore allowing not only a complex classification and characterization (like a summary of each legal document), but also a legal-imperative automatic revision, using in this invention the essential rules mentioned above.

Eduardo Navarro
Agente de Patente