



(51) International Patent Classification:

G06F 16/28 (2019.01) G06F 16/23 (2019.01)
G06F 16/21 (2019.01)

(21) International Application Number:

PCT/MY2019/050136

(22) International Filing Date:

27 December 2019 (27.12.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PI 2018003063 31 December 2018 (31.12.2018) MY

(71) Applicant: **MIMOS BERHAD** [MY/MY]; Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

(72) Inventors: **MAT NOR, Fazli Bin**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **M.BALAKRISHNAN, Nathan A/L**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **MOHD FOWZI, Muhammad Hazwan Bin**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **MOHD FAWZI, Mohd Farid Bin**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

MD ZIN, Wan Zawawi Bin; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

(74) Agent: **MIRANDAH ASIA (MALAYSIA) SDN BHD**; Suite 3B-19-3, Plaza Sentral, Jalan Stesen Sentral 5, Kuala Lumpur, 50470 (MY).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, 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,

(54) Title: A SYSTEM AND METHOD FOR IMPACT ANALYSIS OF CHANGE REQUEST THAT AFFECTS DATABASE STRUCTURE THROUGH CLASSIFICATION AND KEYWORD SIMILARITY ANALYSIS

100

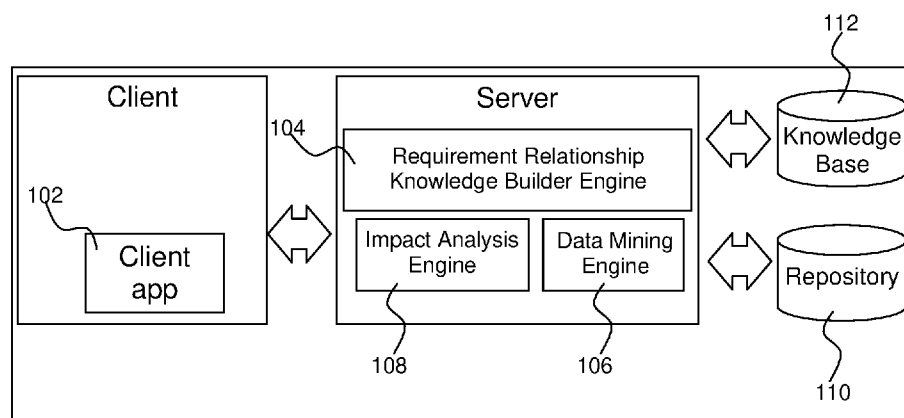


FIG. 1.0

(57) Abstract: The present invention relates to a system and method for impact analysis of change request that affects database structure through classification and keyword similarity analysis. The present invention comprises a client application (102) at client side for input interaction between user and server enabling user to configure an input interaction with the server or repository (110); a server side comprising a Requirement Relationship Knowledge Builder Engine (104); a Data Mining Engine (106) at least one repository (110); and at least one Knowledge base storage (112) of Domain Knowledge Base and Requirement Book (RB) Knowledge Base. The server side further comprising an Impact Analysis Engine (108) for analysing requirements that are impacted of change request based on input from Requirement Relationship Knowledge Builder Engine (104) and Data Mining Engine (106).



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, 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))*
- *of inventorship (Rule 4.17(iv))*

Published:

- *with international search report (Art. 21(3))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

A SYSTEM AND METHOD FOR IMPACT ANALYSIS OF CHANGE REQUEST THAT AFFECTS DATABASE STRUCTURE THROUGH CLASSIFICATION AND KEYWORD SIMILARITY ANALYSIS

5 FIELD OF INVENTION

The present invention relates to a system and method for impact analysis of change request that affects database structure through classification and keyword similarity analysis. In particular, the present invention utilizes classification, tagged words and keyword similarity in detecting impact to non related requirements when a change request is proposed.

10

BACKGROUND ART

Currently, computing systems or machine dependent systems are built based on multiple requirements which are interrelated in nature. A change to a particular requirement would have direct and indirect impact to other requirements of the system. Detecting direct impacts are relatively easy, while indirect impacts might not have related visibility to the change request. Direct and indirect impact may cause malfunction of the system.

15

United States Patent No. 8,768,902 B2 (hereinafter referred to as the US 902 B2 Patent) entitled "Unified concurrent changes to data, schema and application" having a filing date of 11 June 2010; (Assignee: Microsoft Corporation) relates to unified concurrent changes to data, schema, and application whereby changes are managed in a multi-user database application environment. Proposed changes are collected, including changes to data, schema or application descriptions. In the US 902 B2 Patent, user's proposed changes are made according to its dependencies and submitted in a single transaction. For example, a proposed change to a data value may be inconsistent with a change which removes a data element containing the data value. In the US 902 B2 Patent, user may propose changes direct to database objects. Further, the invention in the US 902 B2 includes a diff manager that can produce diffgrams from user gestures to detect inconsistencies whereby diffgrams represent proposed changes to observed data, schema, or application values.

25

30

United States Patent Application Publication No. 2008/0147704 A1 (hereinafter referred to as the US 704 A1 Publication) entitled "Systems and Methods for Propagation of Database Schema Changes", having a filing date of 13 December 2006 (Applicant: Godwin et. al). The US 704 A1 Publication relates to propagation of database schema changes. The US 704 A1 Publication discloses a usability tool which identifies changes in the schema of a database.

35

The usability tool identifies source code dependencies that are affected by the changes in database schema whereby the tool informs a user of the affected dependencies and may also automatically update existing source code dependencies corresponding to the changed schema. The system further comprises a source code manager to determine dependencies affected by the change in schema, the dependencies comprising computer code references to at least some of the tables or columns of data in the database.

United States Patent Application Publication No. 2011/0119288 A1 (hereinafter referred to as the US 288 A1 Publication) entitled "Detecting and Applying Database Schema Changes to Reports", having a filing date of 17 November 2009 (Applicant: Sinha). The US 288 A1 Publication relates to a system and method to detect database schema changes whereby database schema changes are detected via triggers. The changes are also detected by comparing database schemas and determining modifications to database tables and database columns. The invention in the US 288 A1 Publication utilizes only database information which includes tables and columns only to detect the changes. Further, the invention as disclosed in US 288 A1 Publication provides for the result of analysis to be directly applied to the end process.

With reference to the above-mentioned disclosures, there is indeed a need for an improved system and method for impact analysis of change request that affects database structure.

SUMMARY OF INVENTION

The present invention relates to a system and method for impact analysis of change request that affects database structure through classification and keyword similarity analysis. In particular, the present invention utilizes classification, tagged words and keyword similarity in
5 detecting impact to non related requirements when a change request is proposed. The present invention incorporates a system and methodology of classification and keywords similarity analysis that improves impact analysis on requirement dependencies, whenever there is occurrence of changes to database structure. The impact analysis on system requirement of the present invention improves the visibility of the related changes when a
10 request for change request is required in a database structure.

One aspect of the invention provides a system (100) for impact analysis of change request that affects database structure through classification and keyword similarity analysis. The system comprising a client application (102) at client side for input interaction between user
15 and server enabling the user to configure an input interaction with the server or repository (110); a server side comprising a Requirement Relationship Knowledge Builder Engine (104) for gathering information from Requirement Book Knowledge Base, Domain Knowledge Base and Named Entity Recognition repository to generate direct tagged keyword; a Data Mining Engine (106) for building classification, indirect tagged keyword and related word
20 based on direct tagged keyword generated by the Requirement Relationship Knowledge Builder Engine (104); at least one repository (110) for storing named entity relationship and engine data classification, tagged keyword and related words; and at least one Knowledge base storage (112) of Domain Knowledge Base and Requirement Book Knowledge Base. The server side further comprising an Impact Analysis Engine (108) for analysing
25 requirements that are impacted of change request based on input from the Requirement Relationship Knowledge Builder Engine (104) and Data Mining Engine (106).

Another aspect of the invention provides that the Domain Knowledge Base includes information related to keyword's domain classification and its related words.

30 A further aspect of the invention provides that the Requirement Book Knowledge Base includes relationship information between requirement, algorithm, database table and database column and column description; wherein each requirement comprises at least one algorithm; wherein each algorithm comprises at least one database table and wherein each
35 database table comprises at least one column and its description.

Yet another aspect of the invention provides that the Domain Knowledge Base and Requirement Book Knowledge Base is used as an input to all processes and engine.

Still another aspect of the invention provides that the repository (110) includes Named Entity
5 Recognition repository and related word repository.

Another aspect of the invention provides a method (200) for impact analysis of change request that affects database structure through classification and keyword similarity. The method comprising steps of generating direct tagged keyword through Requirement
10 Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base, Domain Knowledge Base and repository (202); executing data mining process through Data Mining Engine by using the direct tagged keywords to build classification, indirect tagged keyword and related word (204); storing result of classification and related word in the repository (206); initiating impact analysis process when a change
15 request is made (208); initiating analysis to identify related database objects requirement that would be impacted due to the change request by retrieving related requirements from Requirement Book Knowledge Base (210); obtaining direct tagged keywords, $D\{1...M\}$ and indirect tagged keywords, $ID\{1...X\}$ related to identified related database objects requirement by sending query to the related word repository (212); performing impact analysis to
20 determine highest indirect tagged keyword based on similarity analysis (214); and generating impact analysis report based on the highest indirect tagged keyword identified, $HID\{1,...,n\}$ to notify indirect or other requirements that are impacted of change request (216). The step for obtaining direct tagged keywords, $D\{1...M\}$ and indirect tagged keywords, $ID\{1...X\}$ related to identified related database objects requirement by sending query to the
25 related word repository (212) further comprises steps of (400) selecting at least one indirectly tagged keyword, IDx from a list of indirectly tagged keywords, $ID\{1...X\}$ which was retrieved from the related word repository (402); determining if the indirectly tagged keyword, IDx is higher value than the value of the indirectly tagged keywords, $ID\{1...X\}$ (404); returning set of list of highest indirect keyword if the indirectly tagged keyword, IDx is higher
30 value than the value of the indirectly tagged keywords, $ID\{1...X\}$ (406); if the indirectly tagged keyword, IDx is lower value than the value of the indirectly tagged keywords, $ID\{1...X\}$ obtaining classification $CID(1...Y)$ from each indirectly impacted tagged keyword, IDx (408); retrieving related classifications, $CID(1...Z)$ for each direct tagged keyword, Dm from the related word repository (410); retrieving indirect related words, $WID[1...J]$ from the
35 related word repository for each $CIDy$ (412); retrieving direct related words, $WD[1...L]$ from the related word repository for each CDz (414); comparing related word for each indirect $WID[1...J]$, with each direct related work of $WD[1...L]$ using cosine similarity (416); and

obtaining highest similarity average value as part of relevant indirect keyword which traces back to indirect related requirements from results obtained (418).

5 A further aspect of the invention provides that the step of generating direct tagged keyword through Requirement Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base and Named Entity Recognition repository (202) further comprises steps of (300) generating each requirement related column by using Requirement Book Knowledge Base (302, 302a); tagging each Requirement Book Knowledge Base column with keywords using Named Entity Recognition repository (304, 304a); and
10 generating classification and related words for each tagged keywords of each Requirement Book using Data Mining Engine (306).

Still another aspect of the invention provides that indirect or other requirements that are impacted of change request are retrieved from requirement knowledge base using HID
15 which traces back to classification, tagged keyword and related word repository.

A further aspect of the invention provides that the step of generating each requirement related column by using Requirement Book Knowledge Base (302, 302a) further comprises providing a set of related columns for each requirement.
20

Still another aspect of the invention provides that the step of comparing related word for each indirect WID[1....J], with each direct related work of WD[1....L] using cosine similarity (416) further comprises determining similarity of words from set of related words until all related words have been selected and computed for similarity.
25

The present invention consist of features and a combination of parts hereinafter fully described and illustrated in the accompanying drawings, it being understood that various changes in the details may be made without departing from the scope of the invention or sacrificing any of the advantages of the present invention.

BRIEF DESCRIPTION OF ACCOMPANYING DRAWINGS

To further clarify various aspects of some embodiments of the present invention, a more particular description of the invention will be rendered by references to specific embodiments thereof, which are illustrated in the appended drawings. It is appreciated that these drawings depict only typical embodiments of the invention and are therefore not to be considered
5 limiting of its scope. The invention will be described and explained with additional specificity and detail through the accompanying drawings.

FIG. 1.0 is a block diagram illustrating system of the present invention.

FIG. 2.0 is a flowchart illustrating general methodology of the present invention.

FIG. 3.0 illustrates a schematic diagram of a Requirement Relation Knowledge Builder
10 Engine of the present invention.

FIG. 4.0 is a flowchart illustrating the steps involved in selecting part of relevant information indirect keyword which traces back to indirect related requirements of the present invention.

FIG. 5.0 illustrates an example of a Highest Indirect Keyword.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention relates to a system and method to improve impact analysis of change request that affects database structure through classification and keyword similarity analysis. In particular, the present invention utilizes classification, tagged words and keyword similarity in detecting impact to non related requirements when a change request is proposed.

5 Hereinafter, this specification will describe the present invention according to the preferred embodiments. It is to be understood that limiting the description to the preferred embodiments of the invention is merely to facilitate discussion of the present invention and it is envisioned without departing from the scope of the appended claims.

Reference is first made to FIG. 1.0 which is a block diagram illustrating the system of the present invention. As illustrated in FIG. 1.0, the system (100) for impact analysis of change request that affects database structure through classification and keyword similarity analysis of the present invention comprising a client application (102) at client side for input interaction between user and server enabling the user to configure an input interaction with the server or repository (110) and a server side. The configuration includes uploading of related requirement documents or initiate execution of process. The server side comprising

10 a Requirement Relationship Knowledge Builder Engine (104) for gathering information from Requirement Book Knowledge Base, Domain Knowledge Base and Named Entity Recognition repository to generate direct tagged keyword whereby the Requirement Relationship Knowledge Builder Engine enables generation of direct tagged keyword; a Data Mining Engine (106) for building classification, indirect tagged keyword and related word based on direct tagged keyword generated by the Requirement Relationship Knowledge Builder Engine (104); an Impact Analysis Engine (108) for analysing requirements that are impacted of change request based on input from the Requirement Relationship Knowledge Builder Engine (104) and Data Mining Engine (106); at least one repository (110) for storing

15 named entity relationship and engine data classification, tagged keyword and related words; and at least one Knowledge base storage (112) of Domain Knowledge Base and Requirement Book Knowledge Base. The Domain Knowledge Base includes information related to keyword's domain classification and its related words while the Requirement Book Knowledge Base includes relationship information between requirement, algorithm, database table and database column and column description; wherein each requirement comprises at least one algorithm; wherein each algorithm comprises at least one database table and wherein each database table comprises at least one column and its description. The Domain Knowledge Base and Requirement Book Knowledge Base is used as an input to all processes and engine. The repository (110) includes Named Entity Recognition repository and related word repository.

20
25
30
35

As illustrated in FIG. 1.0, the present invention is initiated with requirement knowledge building through the Requirement Relationship Knowledge Builder Engine (104) which gathers the information from the Domain Knowledge Base and Requirement Book Knowledge Base. The repository specifically a Named Entity Recognition repository generate direct tagged keyword meaning keywords that are related to the requirement. Subsequently, it is followed by using the direct tagged keywords to build classification, indirect tagged keyword and related word through Data Mining Engine (106). Indirect tagged keywords are keywords indirectly to requirement. The result of classification and related word is then stored in the repository (110). As soon as a change request is made, an analysis is done by the Impact Analysis Engine (108) to identify the related database objects that would be impacted due to the change request by retrieving related requirements from the Requirement Book Knowledge Base (112). A further query is made to the repository (110) to obtain both direct and indirect tagged keyword related to the identified related requirement. Subsequently, impact analysis will be performed to determine the highest indirect tagged keyword based on similarity analysis. Finally, based on the highest indirect tagged keyword identified, a report is generated to notify the indirect or other requirements that are impacted of the change request. The methodology will be described in detail in the following paragraphs.

Reference is now made to FIG. 2.0 which is a flowchart illustrating the general methodology of the present invention. As illustrated in FIG. 2.0, the method (200) for impact analysis of change request that affects database structure through classification and keyword similarity analysis is divided into two phases, Phase A and Phase B. Phase A is a pre-requisite of Phase B. Phase A provides the steps involve for constructing the required knowledge base for the execution of Phase B. Phase B provides building blocks of the entire similarity impact analysis process for Impact Analysis Engine. The method (200) as illustrated in FIG. 2.0 comprising steps of first generating direct tagged keyword through Requirement Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base and repository (202). It is followed by executing data mining process through Data Mining Engine by using direct tagged keywords to build classification, indirect tagged keyword and related word (204). The result of classification and related word is stored in the repository (206). Similarity impact analysis process is initiated when a change request is made (208). Subsequently, analysis is initiated to identify related database objects requirement that would be impacted due to the change request by retrieving related requirements from Requirement Book Knowledge Base (210). Direct tagged keywords, $D\{1...M\}$ and indirect tagged keywords, $ID\{1...X\}$ related to identified related database objects requirement are obtained by sending query to the related word repository (212).

Thereafter, impact analysis is performed to determine highest indirect tagged keyword based on similarity analysis (214); and subsequently generating impact analysis report based on highest indirect tagged keyword identified, $HID\{1,...,n\}$ to notify indirect or other requirements that are impacted of change request (216). All indirect or other requirements that are impacted of change request are retrieved from Requirement Book Knowledge Base (112) using Human Interface Device, HID which traces back to classification, tagged keyword and related word repository.

Reference is now made to FIG. 3.0 which illustrates a schematic diagram of the Requirement Relation Knowledge Builder Engine of the present invention. As illustrated in FIG. 3.0, the process in the Requirement Relation Knowledge Builder Engine in generating requirement knowledge building through Requirement Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base and Named Entity Recognition repository (202) further comprises steps of (300) first generating each requirement related column by using Requirement Book Knowledge Base (302, 302a). Generating each requirement related column by using Requirement Book Knowledge Base (302, 302a) further comprises providing a set of related columns for each requirement. Once the related columns are generated, each requirement would have a set of related columns. For example, requirement RB1 would have $\{C1, D1\}$ as its related columns and its description.

The method continues with tagging each Requirement Book Knowledge Base column with keywords using Named Entity Recognition repository (304, 304a). The tagging process comprises of extracting the entity name from the column's description and tagged the column with the entity name. For example, C1 columns has description as "*This column is related to employee salary based on their position*", the tagging process then extract the entity name from the description which tagged the column as "Payslip", "Staff", "Money" and "Company". At the end of the process, each Requirement Book would have set of related tagged keywords based on its set of columns. For example, RB1 would have $\{K1, K2,...,Kn\}$. Subsequently, classification and related words are generated for each tagged keywords of each Requirement Book using Data Mining Engine (306). Tagged keywords classification is generated by using Naive Bayes algorithm which uses Requirement Book Knowledge Base as its features reference of the tagged keywords. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, tagged keyword "Staff" will be classified as "Person" if it has features characteristics as a person such as gender, race, address and others. On the other hand, in the same process, related words of tagged keywords are generated using Latent Semantic

Analysis to discover a new keyword based on meaning behind the words based on certain topic or domain. For example, tagged keyword "Staff" will generate related words such as "Employee", "Personnel", "Team" and others. In the end of the process, all requirements and related result which include tagged keywords, classification and related words are stored in the repository as an input for the next process.

Reference is now made to FIG. 4.0 which is a flowchart illustrating the steps involved in selecting part of relevant information indirect keyword which traces back to indirect related requirements of the present invention. As illustrated in FIG. 4.0, the step of obtaining direct tagged keywords, $D\{1...M\}$ and indirect tagged keywords, $ID\{1...X\}$ related to identified related database objects requirement by sending query to the related word repository (212) further comprises steps of (400) first selecting at least one indirectly tagged keyword, IDx from a list of indirectly tagged keywords, $ID\{1...X\}$ which was retrieved from the related word repository (402). It is further determined if the indirectly tagged keyword, IDx is higher value than the value of the indirectly tagged keywords, $ID\{1...X\}$ (404). The set of list of highest indirect keyword is returned if the indirectly tagged keyword, IDx is higher value than the value of the indirectly tagged keywords, $ID\{1...X\}$ (406). If the indirectly tagged keyword, IDx is lower value than the value of the indirectly tagged keywords, $ID\{1...X\}$, proceeding to obtain classification $CID\{1...Y\}$ from each indirectly impacted tagged keyword, IDx (408). Thereafter, related classifications, $CID\{1...Z\}$ are retrieved for each direct tagged keyword, Dm from the related word repository (410); and indirect related words, $WID[1...J]$ are retrieved from the related word repository for each $CIDy$ (412). Subsequently, direct related words, $WD[1...L]$ are retrieved from the related word repository for each CDz (414a) and each related word is compared for each indirect $WID[1...J]$, with each direct related work of $WD[1...L]$ using cosine similarity (416). The highest similarity average value is obtained as part of relevant indirect keyword which traces back to indirect related requirements from results obtained (418). In comparing related word for each indirect $WID[1...J]$, with each direct related work of $WD[1...L]$ using cosine similarity (416) further comprises determining similarity of words from set of related words until all related words have been selected and computed for similarity.

Reference is now made to FIG. 5.0 which illustrates an example of the Highest Indirect Keyword. As illustrated in FIG. 5.0, in this example, the related words of 2 classification $\{CIDWy, CDWz\}$ is represented as $WID[1..J]$ and $WD[1..L]$. For example, if the 2 classification are "Human Resource" and "Finance", $CIDWy$ would represent "Database" and $CDWz$ would represent "Finance". Each classification has its own related words. For example, $CIDWy$ which represent "Human Resource" has related words such as

“Employee”, “Interview” and “Job”, WID1 would represent “Employee”, WID2 would represent “Interview” and WID3 would represent “Job”. Whereby CDWz which represent “Finance” has related words such as “Salary”, “Billing” and “Profit”, WD1 would represent “Salary”, WD2 would represent “Billing” and WD3 would represent “Profit”. The process then
5 select first classification related words, WIDn i.e WID1 from the set of related words {WID1, WID2, WID3}. The process continues by calculating the similarity of WID1 with WIDm i.e. {WD1, WD2, WD3} by using the cosine similarity until all the set of related words from WD have been selected and calculated the similarity. Numbers such as 0.8, 0.7, 0.6 and etc represent the similarity between WIDm and WID1. For example, the similarity between the
10 first set of related words WID i.e. WID1 and the classification’s related words i.e. WD1 is shown as 0.8. The steps of selecting the related words WIDn and calculating the similarity of WIDm and WIDn are repeated until there are no more related words to be selected. Once these steps are completed, the process returns the set of the highest indirect keywords HIDn for each keyword.

15 The present invention incorporates a system and methodology of classification and keywords similarity analysis that improves impact analysis on requirement dependencies, whenever there is occurrence of changes to database structure. The impact analysis on system requirement of the present invention improves the visibility of the related changes
20 when a request for change request is required in a database structure.

Unless the context requires otherwise or specifically stated to the contrary, integers, steps or elements of the invention recited herein as singular integers, steps or elements clearly encompass both singular and plural forms of the recited integers, steps or elements.

25 Throughout this specification, unless the context requires otherwise, the word “comprise”, or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated step or element or integer or group of steps or elements or integers, but not the exclusion of any other step or element or integer or group of steps, elements or integers. Thus, in the context of this specification, the term “comprising” is used in an inclusive sense
30 and thus should be understood as meaning “including principally, but not necessarily solely”.

CLAIMS

1. A system (100) for impact analysis of change request that affects database structure through classification and keyword similarity analysis comprising:
a client application (102) at client side for input interaction between user and server,
5 enabling the user to configure an input interaction with the server or repository (110);
a server side comprising:
a Requirement Relationship Knowledge Builder Engine (104) for gathering
information from Requirement Book Knowledge Base, Domain Knowledge
Base and Named Entity Recognition repository to generate direct tagged
10 keyword;
a Data Mining Engine (106) for building classification, indirect tagged keyword
and related word based on direct tagged keyword generated by the
Requirement Relationship Knowledge Builder Engine (104);
at least one repository (110) for storing named entity relationship and engine
15 data classification, tagged keyword and related words; and
at least one Knowledge base storage (112) of Domain Knowledge Base and
Requirement Book Knowledge Base,
characterized in that
the server side further comprising:
20 an Impact Analysis Engine (108) for analysing requirements that are
impacted by change request based on input from the Requirement
Relationship Knowledge Builder Engine (104) and Data Mining Engine
(106).
- 25 2. The system (100) according to Claim 1, wherein the Domain Knowledge Base
includes information related to keyword's domain classification and its related words.
3. The system (100) according to Claim 1, wherein the Requirement Book Knowledge
Base includes relationship information between requirement, algorithm, database
30 table and database column and column description; wherein each requirement
comprises at least one algorithm; wherein each algorithm comprises at least one
database table; and wherein each database table comprises at least one column and
its description.
- 35 4. The system (100) according to Claim 1, wherein the Domain Knowledge Base and
Requirement Book Knowledge Base is used as an input to all processes and engine.

5. The system (100) according to Claim 1, wherein the repository (110) includes Named Entity Recognition repository and related word repository.

6. A method (200) for impact analysis of change request that affects database structure through classification and keyword similarity analysis comprising steps of:

generating direct tagged keyword through Requirement Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base, Domain Knowledge Base and repository (202);

executing data mining process through Data Mining Engine by using the direct tagged keywords to build classification, indirect tagged keyword and related word (204);

storing result of classification and related word in the repository (206);

initiating impact analysis process when a change request is made (208);

initiating analysis to identify related database objects requirement that would be impacted due to the change request by retrieving related requirements from the Requirement Book Knowledge Base (210);

obtaining direct tagged keywords ($D\{1...M\}$) and indirect tagged keywords ($ID\{1...X\}$) related to identified related database objects requirement by sending query to the related word repository (212);

performing impact analysis to determine highest indirect tagged keyword based on similarity analysis (214); and

generating impact analysis report based on the highest indirect tagged keyword identified, $HID\{1,...,n\}$ to notify indirect or other requirements that are impacted of change request (216),

characterized in that

obtaining direct tagged keywords, $D\{1...M\}$ and indirect tagged keywords, $ID\{1...X\}$ related to identified related database objects requirement by sending query to the related word repository (212) further comprises steps of (400):

selecting at least one indirectly tagged keyword, IDx from a list of the indirectly tagged keywords, $ID\{1...X\}$ which was retrieved from the related word repository (402);

determining if the indirectly tagged keyword, IDx is higher value than value of the indirectly tagged keywords, $ID\{1...X\}$ (404);

returning set of list of highest indirect keyword if the indirectly tagged keyword, IDx is higher value than the value of the indirectly tagged keywords, $ID\{1...X\}$ (406);

if the indirectly tagged keyword, ID_x is lower value than the value of the indirectly tagged keywords, ID{1...X}:

obtaining classification CID (1...Y) from each indirectly impacted tagged keyword, ID_x (408);

retrieving related classifications, CID (1...Z) for each direct tagged keyword, D_m from the related word repository (410);

retrieving indirect related words, WID[1...J] from the related word repository for each CID_y (412);

retrieving direct related words, WD[1...L] from the related word repository for each CD_z (414);

comparing related word for each indirect WID[1...J], with each direct related word of WD[1...L] using cosine similarity (416); and

obtaining highest similarity average value as part of relevant indirect keyword which traces back to indirect related requirements from results obtained (418).

7. The method (200) according to Claim 6, wherein generating direct tagged keyword through Requirement Relationship Knowledge Builder Engine that gathers information from Requirement Book Knowledge Base and Named Entity Recognition repository (202) further comprises steps of (300):

generating each requirement related column by using the Requirement Book Knowledge Base (302, 302a);

tagging each Requirement Book Knowledge Base column with keywords using the Named Entity Recognition repository (304, 304a); and

generating classification and related words for each tagged keywords of each Requirement Book using Data Mining Engine (306).

8. The method (200) according to Claim 6, wherein indirect or other requirements that are impacted of change request are retrieved from Requirement Book Knowledge Base using Human Interface Device, HID which traces back to classification, tagged keyword and related word repository.

9. The method (200) according to Claim 7, wherein generating each requirement related column by using Requirement Book Knowledge Base (302, 302a) further comprises providing a set of related columns for each requirement.

10. The method (200) according to Claim 6, wherein comparing related word for each indirect WID[1....J], with each direct related work of WD[1....L] using cosine similarity (416) further comprises determining similarity of words from set of related words until all related words have been selected and computed for similarity.

100

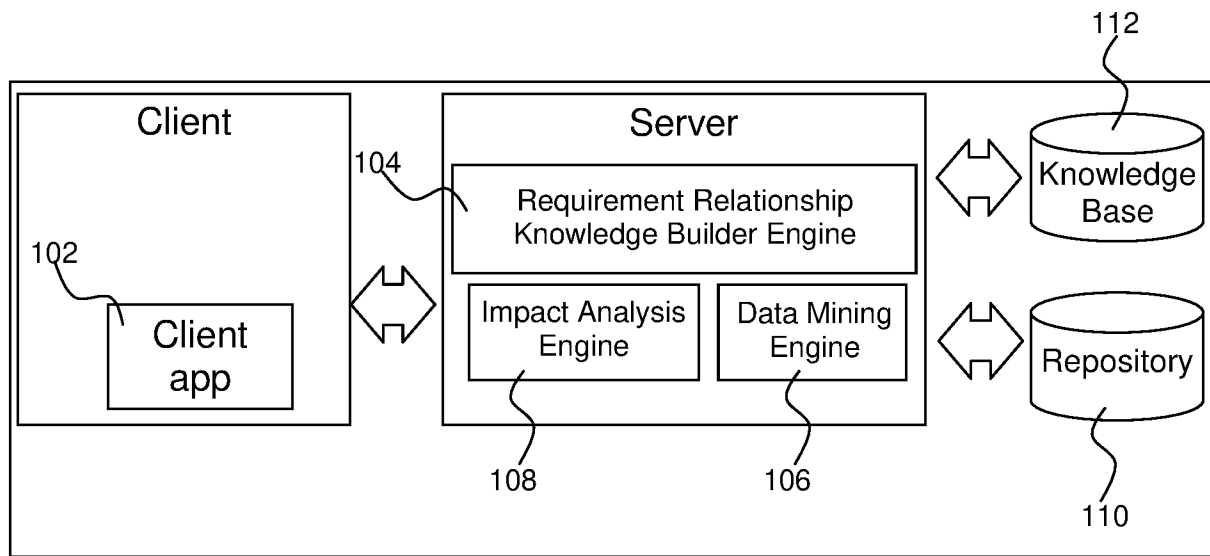


FIG. 1.0

200

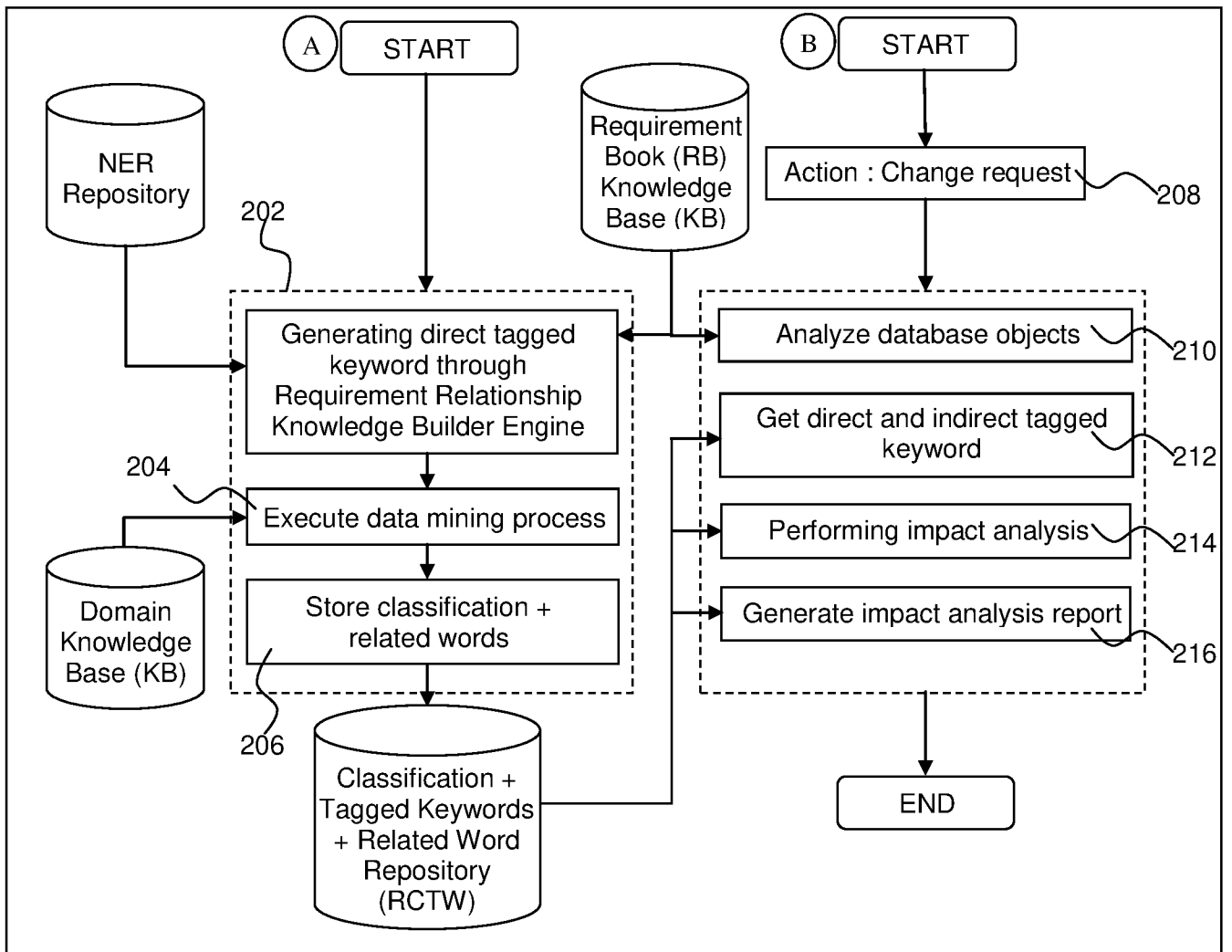


Fig. 2.0

300

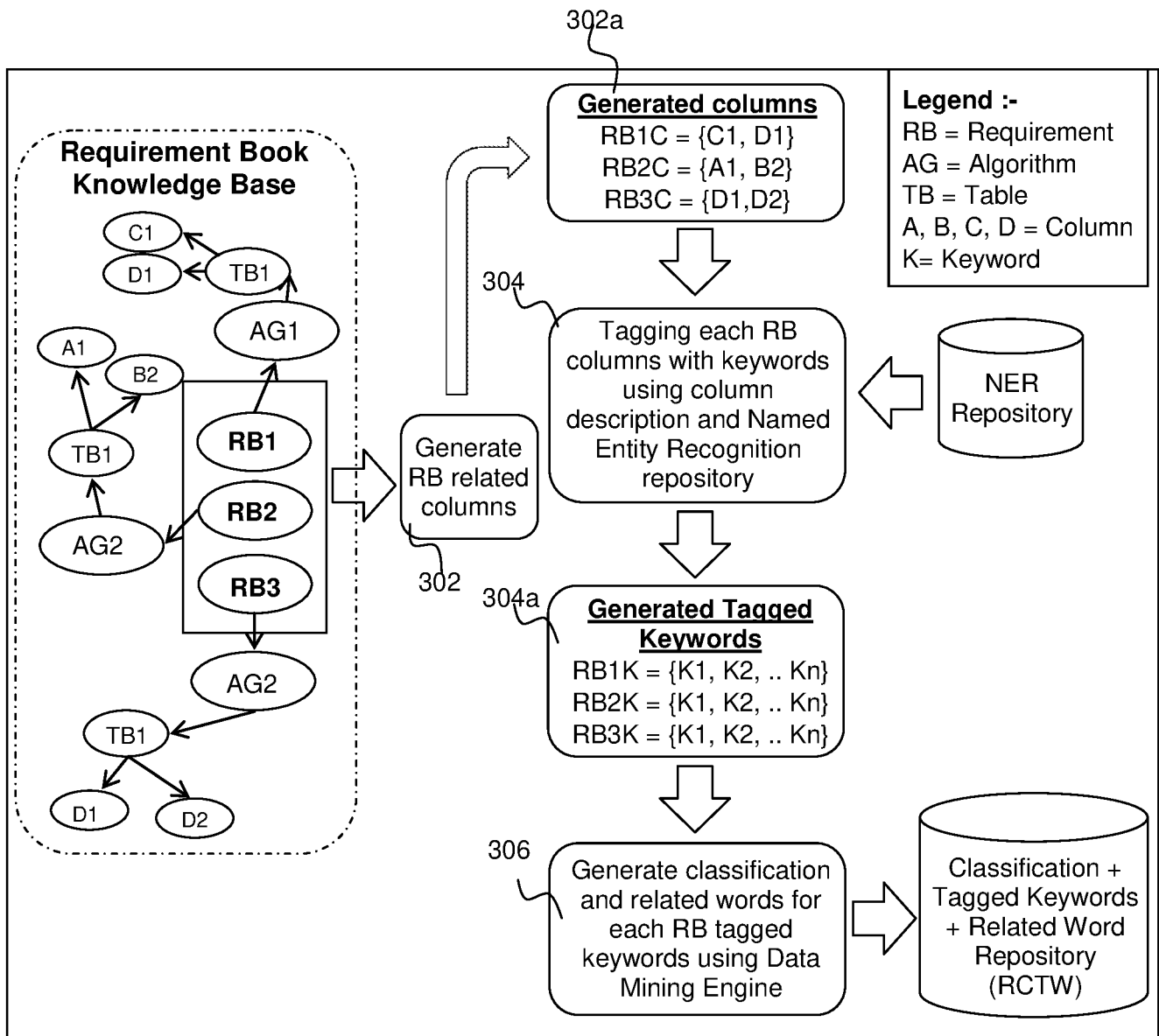


FIG. 3.0

400

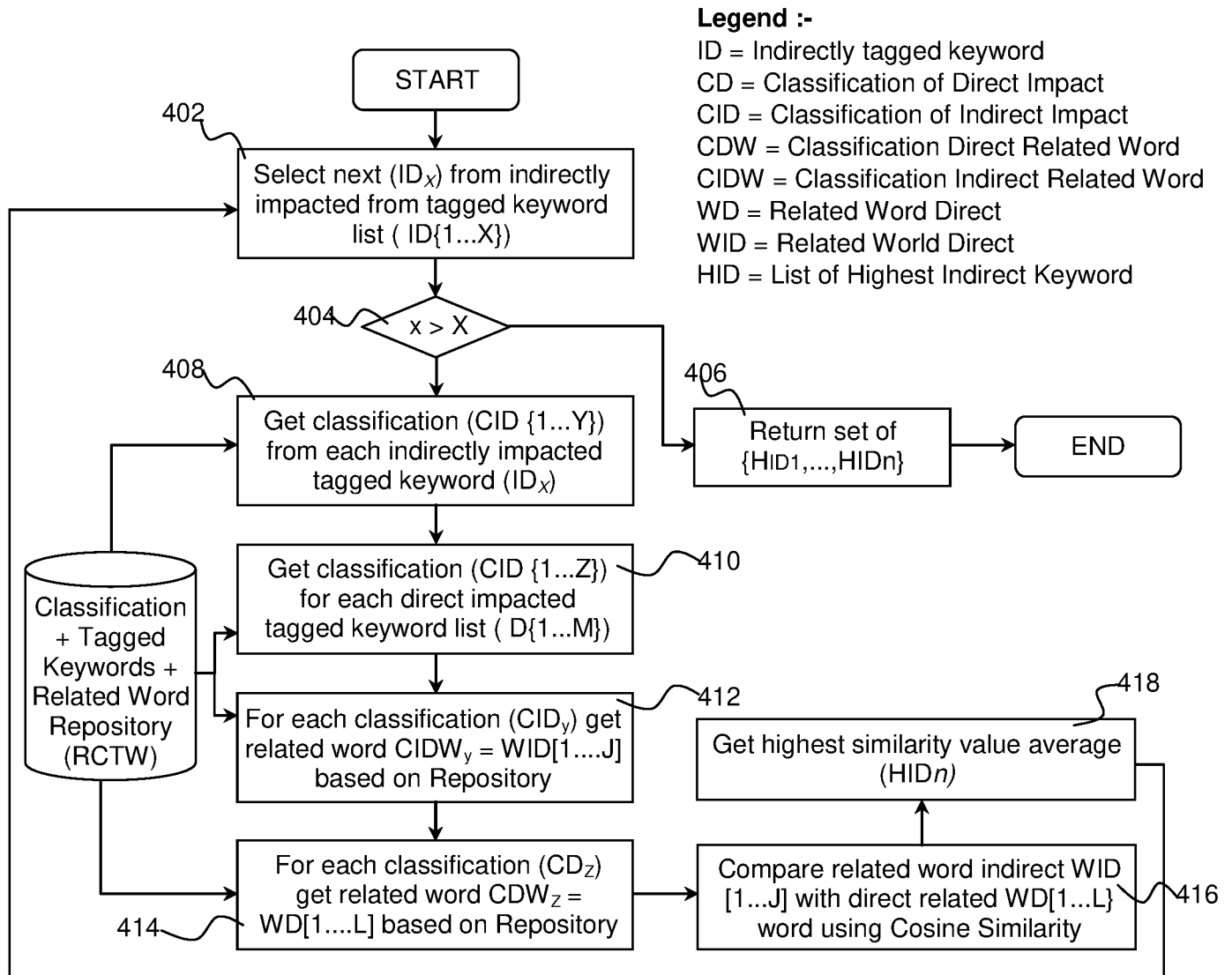


FIG. 4.0

500

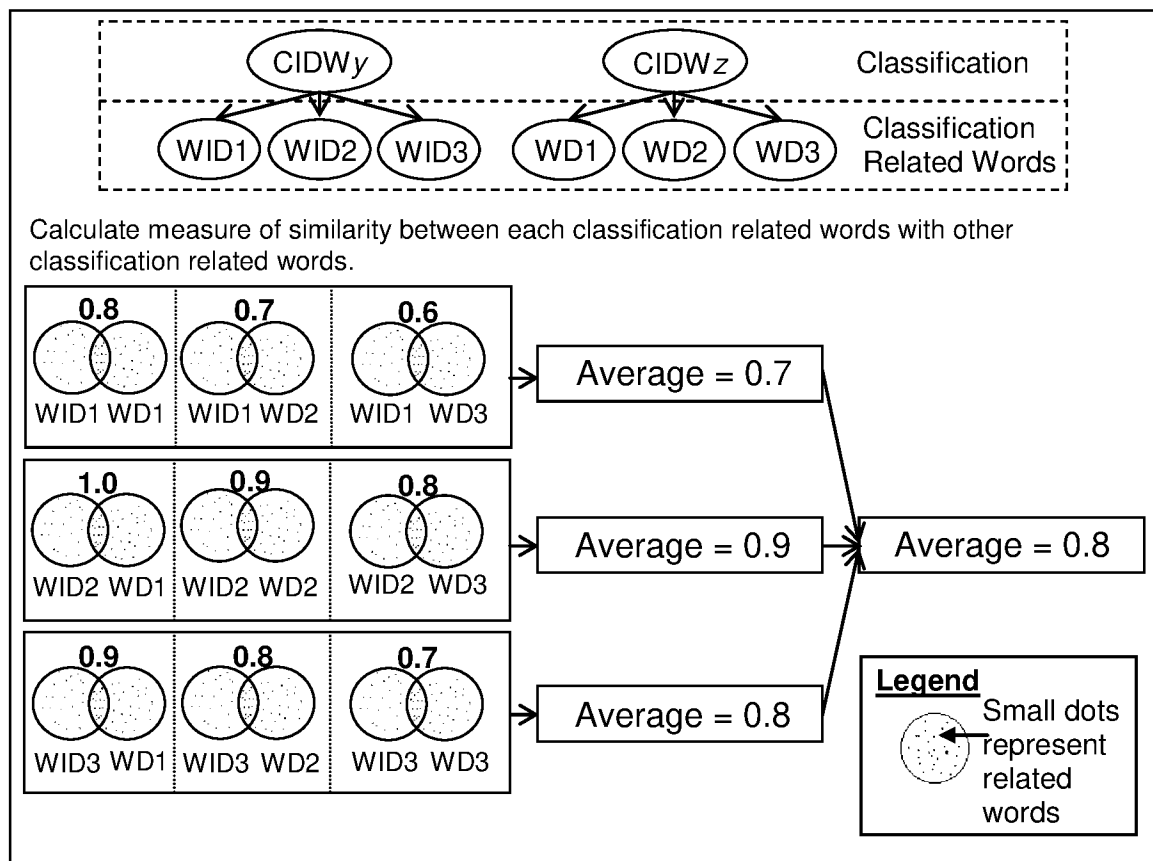


FIG. 5.0

INTERNATIONAL SEARCH REPORT

International application No.
PCT/MY2019/050136**A. CLASSIFICATION OF SUBJECT MATTER****G06F 16/28(2019.01)i, G06F 16/21(2019.01)i, G06F 16/23(2019.01)i**

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 16/28; G06F 11/36; G06F 17/30; G06F 9/44; G06F 16/21; G06F 16/23

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: change, impact, analyze, database, relationship

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012-0159433 A1 (AZADEH AHADIAN et al.) 21 June 2012 Paragraphs [0010]-[0012], [0024], [0034]-[0035]; and figure 3	1-10
A	US 2011-0282860 A1 (DAVID W. BAARMAN et al.) 17 November 2011 Paragraphs [0032], [0043], [0099]; and figure 4	1-10
A	US 2010-0153908 A1 (SANTONU SARKAR et al.) 17 June 2010 Paragraphs [0007], [0037], [0057]; and figure 2	1-10
A	US 2016-0125057 A1 (AB INITIO TECHNOLOGY LLC) 05 May 2016 Paragraphs [0017], [0031], [0049]; and figure 1	1-10
A	KR 10-2018-0119857 A (HYUNDAI MOTOR COMPANY) 05 November 2018 Paragraphs [0018], [0123]; and figure 1	1-10



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

"D" document cited by the applicant in the international application

"E" earlier application or patent 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

21 April 2020 (21.04.2020)

Date of mailing of the international search report

21 April 2020 (21.04.2020)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea



Facsimile No. +82-42-481-8578

Authorized officer

BYUN, Sung Cheal

Telephone No. +82-42-481-8262



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/MY2019/050136

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0159433 A1	21/06/2012	CN 101727320 A US 2010-0114962 A1 US 8458654 B2 US 8499281 B2	09/06/2010 06/05/2010 04/06/2013 30/07/2013
US 2011-0282860 A1	17/11/2011	CN 102884530 A JP 2013-526747 A JP 5810452 B2 KR 10-2013-0083838 A WO 2011-146391 A2	16/01/2013 24/06/2013 11/11/2015 23/07/2013 24/11/2011
US 2010-0153908 A1	17/06/2010	EP 2199901 A1 US 8352914 B2	23/06/2010 08/01/2013
US 2016-0125057 A1	05/05/2016	CN 107135663 A EP 3215935 A1 JP 2018-501538 A KR 10-2017-0081239 A US 10176234 B2 US 2019-0095390 A1 WO 2016-073735 A1	05/09/2017 13/09/2017 18/01/2018 11/07/2017 08/01/2019 28/03/2019 12/05/2016
KR 10-2018-0119857 A	05/11/2018	US 2018-0314519 A1	01/11/2018