



- (51) **International Patent Classification:**
G06N 5/02 (2006.01) G06F 17/30 (2006.01)
- (21) **International Application Number:**
PCT/MY2014/000125
- (22) **International Filing Date:**
28 May 2014 (28.05.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PI 2013701827 27 September 2013 (27.09.2013) MY
- (71) **Applicant:** MIMOS BERHAD [MY/MY]; Technology Park Malaysia, 57000 Kuala Lumpur (MY).
- (72) **Inventors:** KOW, Weng Onn; c/o MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY). DICKSON, Lukose; c/o MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY).
- (74) **Agent:** H. A. RASHID, Ahmad Fadzlee; Ipvolusi Sdn Bhd, A-3-3A, Centrio Pantai Hillpark, No. 1, Jalan Pantai Murni, 59200 Kuala Lumpur (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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, 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, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

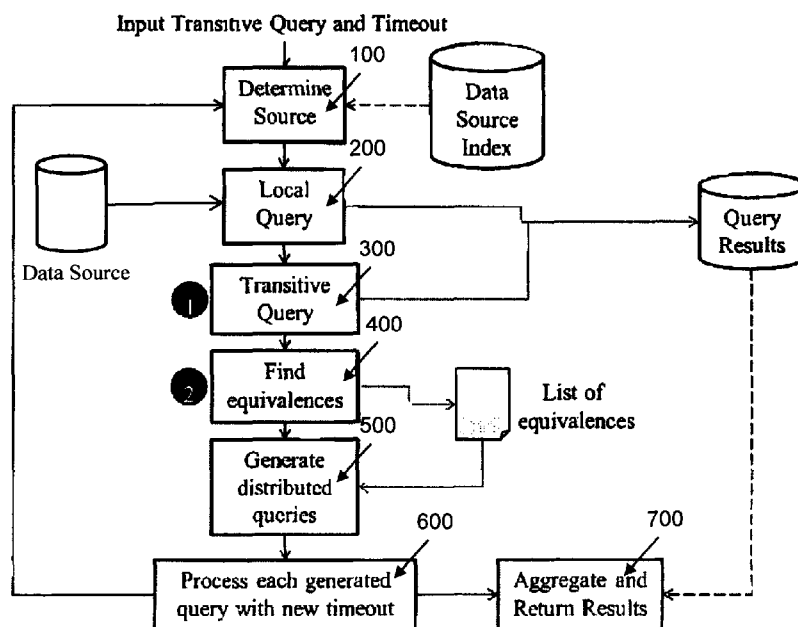
(54) **Title:** METHOD FOR PERFORMING DISTRIBUTED REASONING OVER LINKED DATA

FIG. 1

(57) **Abstract:** The present invention relates to a method for reasoning knowledge that is distributed across a plurality of linked knowledge bases. It handles queries in two ways. First, it handles a query by determining whether or not a fact is true. Secondly, it handles a query by returning query results that are not directly explicit within the triples of the queries knowledge bases which are the subjects, predicates and objects.

**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))*
- *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))*

METHOD FOR PERFORMING DISTRIBUTED REASONING OVER LINKED DATA**FIELD OF INVENTION**

The present invention relates to a method for reasoning knowledge that is distributed across a plurality of linked knowledge bases.

BACKGROUND OF THE INVENTION

In recent times, there are plenty of information sources that everyone is capable to obtain information from all over the world on the world wide web. A knowledge base, also known as kb, is a special database for knowledge management. It is an information repository that provides a means for information to be collected, organised, shared, searched and utilized. A knowledge base can either be machine-readable or intended for human use. Human-readable knowledge bases are used to enable people to retrieve and use the knowledge they contain. A human-readable knowledge base can also be coupled with a machine-readable one.

Machine-readable knowledge bases store knowledge in a computer-readable form which usually has automated deductive reasoning applied to them. These knowledge bases contain a set of data, often in the form of rules that describe the knowledge in a logically consistent manner, whereby an ontology is used to define the structure of the stored data. One of the examples is the Web Ontology Language (OWL), which is a family of knowledge representation language for authoring ontologies. The languages are characterised by formal semantics and Resource Description Framework (RDF) / Extensible Markup Language (XML) – based serializations for the semantic web.

A search engine is usually used to locate information in the system or to browse through a classification scheme. Search engines often provide compilation and retrieval services. Due to the growing amount of data contained within the world wide web and other information sources such as these knowledge bases, it is difficult for users to obtain all information they need when only a single search request is used. Apart from that, reasoning on a distributed set of knowledge bases requires all the information such as the rules and facts to be stored in the same location before the inferencing for new facts can be performed.

This can lead to several issues on how to decide which knowledge bases to bring together and limited results are achieved as results are returned only from the knowledge bases that were specified to be federated. Therefore, there is a need to provide a method that addresses the above mentioned drawbacks of the existing
5 method for reasoning knowledge.

SUMMARY OF INVENTION

The present invention relates to a method for reasoning knowledge that is distributed across a plurality of linked knowledge bases. The method is characterised
10 by the steps of receiving an initial query from user; identifying an appropriate knowledge base to answer the query by a data source index; sending the query to the appropriate knowledge base for a local query; storing the results from the local query in a computer memory; performing a transitive reasoning within the same knowledge base if the query involves a transitive property; discovering equivalences
15 of concepts and relations to the answers of the query in other knowledge bases; generating new queries from the discovered equivalences; distributing queries to knowledge bases to perform reasoning until a timeout is reached or if there are no more answers found from the knowledge bases; aggregating results from all knowledge bases; and returning the results to user.

20

Preferably, the step of performing a transitive reasoning within the same knowledge base if the query involves a transitive property includes replacing the subject of the initial query with the result received from the knowledge base to form a new query; and finding results of the new query within the same knowledge base.

25

Preferably, the step of discovering equivalences of concepts and relations to the answers in other knowledge bases includes finding explicit equivalent subjects, objects and predicates of the results of the queries; and finding implicit equivalent subjects, objects and predicates of the results of the queries.

30

Preferably, the explicit equivalent subjects, objects and predicates of the results of the queries are found by using Web Ontology Language (OWL).

Preferably, the implicit equivalent subjects, objects and predicates of the
35 results of the queries are found by using functional and inverse functional properties.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention.

5

FIG. 1 illustrates a flow chart of a method for reasoning knowledge according to an embodiment of the present invention.

FIG. 2 illustrates a flow chart of a sub-process for transitive reasoning.

10

FIG. 3 illustrates a flow chart of a sub-process for equivalences discovering.

FIG. 4 illustrates the aggregated results of an example of a query across a plurality of linked knowledge bases.

15

DESCRIPTION OF THE PREFERRED EMBODIMENT

A preferred embodiment of the present invention will be described herein below with reference to the accompanying drawings. In the following description, well known functions or constructions are not described in detail since they would obscure the description with unnecessary detail.

20

Reference is made initially to **FIG. 1**, which illustrates a flow chart of a method for reasoning knowledge according to an embodiment of the present invention. The method which is used in a query system enables the inference of knowledge using rules and facts that are distributed across a plurality of linked knowledge bases. It handles queries in two ways. First, it handles a query by determining whether or not a fact is true. Secondly, it handles a query by returning query results that are not directly explicit within the triples of the queries knowledge bases which are the subjects, predicates and objects.

25

30

Initially, when a query is received, a data source index is used to determine which data source or knowledge base that contains the answer to the initial query as in step **100**. The data source index stores a list of namespaces which is used to identify an appropriate knowledge base based on a query. After the appropriate knowledge base is determined, the query is sent to the knowledge base for a local

35

query as in step 200. The results from the local query are then stored in a computer memory until the final aggregated results are returned as in step 700.

If the query involves a transitive property, the system then performs a transitive reasoning within the same knowledge base and retrieves more results as in step 300. For a transitive property, if "a" has ancestor "b" and "b" has ancestor "c", we can infer that "a" also has ancestor "c". Transitive properties are usually used for hierarchy relations such as subclass and relations such as "part of" or "located in" predicates. However, if the query has no transitive property, the system directly returns the results to the user without any further queries. After a query with a transitive property has completed its transitive reasoning, the equivalences of the subject of the query are looked up as in step 400. The subject is replaced with the results of the query to generate a set of new queries as in step 500. The entire process is repeated for each new query from step 100 with the identification of which knowledge base that could answer the new query. The process repeats until a timeout is reached or if there are no more answers to the query found from the knowledge base, upon which the aggregated results are returned to the user as in step 600 and step 700.

Step 300 is further described in FIG. 2 which illustrates a flow chart of a sub-process for the transitive reasoning. Initially, from the original query in step 100, any query clause that contains a transitive property undergoes the replacement of the subject with the result as in step 310. An example of a query having a transitive property is "PJ is located in X?", wherein X is the answer that the system needs to find. Since "is located in" is a transitive predicate, as explained previously, the subject i.e. "PJ" is replaced by the result of the query i.e. "Selangor" which creates a new question i.e. "Selangor is located in X?", wherein the system repeats and continues the query performing and subject replacing to create new queries until the knowledge base stops returning answers to the queries as in steps 310 to 330. Each time a query runs, the timeout is checked as in step 320. If the query exceeds the timeout, then whatever results found within the timeout are returned.

When a knowledge base gets answers for the sent queries as shown in FIG. 2, the system continues to look for equivalences or similar concepts to the answers in another knowledge base as in step 400. From the previous query example, the query

"PJ is located in X?" returns the result "Selangor." The system will then find if "Selangor" exists in any other knowledge base. If the system finds that "Selangor" in the first knowledge base is the same as "Selangor" in the second knowledge base, the query that is created from the transitive reasoning i.e. "Selangor is located in X?" is then sent to the second knowledge base to get more results.

Referring now to **FIG. 3**, there is shown a flow chart of a sub-process for finding explicit and implicit equivalences of the results from the queries. Explicit equivalences are found by using existing relations such Web Ontology Language (OWL) to find the equivalent subjects, objects and predicates as in steps **410** to **430**. OWL has predicates called "equivalentClass" and "sameAs" to indicate that a concept is equivalent to another. The previous result from the query example of "PJ is located in X?" is used to further explain how the equivalences are found by using OWL. The first knowledge base might have a triple of results which are "kb1: Selangor", "owl:sameAs" and "kb2: Selangor", wherein kb1 is a first knowledge base, owl is web ontology language and kb2 is a second knowledge base. The triple of results means that "Selangor" in the first knowledge base is equivalent to "Selangor" in the second knowledge base.

On the other hand, implicit equivalences are found by using functional and inverse functional properties as in steps **440** to **450**. A functional property is a constraint that says a property can only have one object value for each unique subject. For example, a country can only have one capital city. In another example, a person can only have one biological mother. Inference can be performed by the following rule whereby if "a" is a property of "b" and "a" is a property of "c", it is said that "b" is equivalent to "c." An example of a functional property, the first statement says "Malaysia's capital city is KL" and the second statement says "Malaysia's capital city is Kuala Lumpur." It can be deduced that "KL" is the same as "Kuala Lumpur".

Similarly, an inverse functional property is a functional property in reverse where the subject is constrained. The rule of an inverse property states that if "a" has a property of "b" and "c" has a property of "b", then it is said that "a" is equivalent to "c." For example, in a statement it is said that "ferum has an atomic weight of 55.845" and in another statement, it is said that "iron has an atomic weight of 55.845." Thus, it can be deduced that ferum is the same as iron.

After finding the explicit and implicit equivalences of the results, these concepts are then replaced to generate a set of new queries as in step 500. The other knowledge bases will generate more queries and the entire process is repeated again for each new query from step 100 by determining the right knowledge base that could answer the new queries. The process repeats until a new timeout is reached or if there are no more answers to the query found from the knowledge base, upon which the aggregated results are returned to the user as in step 600 and step 700.

Table 1 and Table 2 show another detailed example on finding the equivalent subjects, objects and predicates of the results of the queries according to an embodiment of the present invention, wherein kb1 is a first knowledge base, kb2 is a second knowledge base, owl is web ontology language and rdf is resource description framework.

Table 1

Subject	Predicate	Object
kb1: KLCC	kb1: locatedIn	kb1: KL
kb1: KL	kb1: locatedIn	kb1: Malaysia
kb1: Malaysia	owl: sameAs	kb2: Malaysia
kb1: Malaysia	kb1: hasCapital	kb1: KL
kb1: Malaysia	kb1: hasCapital	kb2: KualaLumpur
kb1: locatedIn	owl: equivalenPproperty	kb2: hasLocation
kb1: hasCapital	rdf: type	owl: FunctionalProperty

Table 2

Subject	Predicate	Object
kb2: KualaLumpur	kb2: hasLocation	kb2: KlangValley
kb2: Malaysia	kb2: hasLocation	kb2: SouthEastAsia

The query "KLCC located in X?" can be answered according to the steps as described earlier on, wherein X is the answer that we need to find. Initially, when the query "KLCC located in X?" is received, a data source index is used to determine

which knowledge base contains the answer to this query as in step **100**. This query is sent to the chosen knowledge base for a local query as in step **200**. The answer "KL" from the local query is stored in a computer memory until the final aggregated results are returned as in step **700**.

5

Since this query involves a transitive property, the system performs the transitive reasoning within the same knowledge base as in step **300** and replaces the subject of the query which in this case is "KLCC" with the result i.e. "KL". The system then creates a new question i.e. "KL is located in X?" and repeats the steps from step **310** to step **330** until the knowledge base stops returning answers. From this step, the system retrieves more result ie "Malaysia". After the knowledge base gets answers for the queries i.e. "KL" and "Malaysia", the system continues to look for equivalences or similar concepts to these answers in another knowledge base as in step **400**. The system finds out if "KL" and "Malaysia" exist in any other knowledge base besides the first knowledge base. From **Table 1**, it shows that the system finds that "Malaysia" exists in a second knowledge base.

The explicit equivalences are found by using OWL to find the equivalent subjects, objects and predicates as in steps **410** to **430**. From **Table 1**, it shows that OWL is used to relate that "Malaysia" in the first knowledge base is the same as "Malaysia" from the second knowledge base. The queries are then sent to the second knowledge base to get more results. It is also computed that the predicate "located in" from the first knowledge base is the same as the predicate "has location" from the second knowledge base. On the other hand, implicit equivalences are found as in steps **440** to **450** by using a functional property. By using "has capital" as the predicate, the first knowledge base returns a result of "KL" while the second knowledge base returns a result of "Kuala Lumpur." By using the functional property, it is deduced that "KL" is the same as "Kuala Lumpur."

These concepts are then replaced to generate a set of new queries as in step **500**. In **Table 2**, it shows that the second knowledge base generates two new queries which are "Kuala Lumpur has location X?" and "Malaysia has location X?", wherein X is the answer that we need to find. The entire process is repeated again for each new query from step **100** with the identification of the appropriate knowledge base that could answer the new queries. The process repeats until a timeout is

reached or if there are no more answers to the queries found from the knowledge bases, upon which the aggregated results are returned to the user as in step 600 and step 700.

5 **Table 2** shows that the results returned from the second knowledge bases are “Klang Valley” and “Malaysia.” The final aggregated results from the first knowledge base are “KL” and “Malaysia”, while the aggregated results from the second knowledge base are “Kuala Lumpur”, “Malaysia”, “Klang Valley” and “South east Asia.” The final result of the query is shown in **FIG. 4**.

10

While embodiments of the invention have been illustrated and described, it is not intended that these embodiments illustrate and describe all possible forms of the invention. Rather, the words used in the specifications are words of description rather than limitation and various changes may be made without departing from the scope
15 of the invention.

CLAIMS

1. A method for reasoning knowledge that is distributed across a plurality of linked knowledge bases is characterised by the steps of:
 - a) receiving an initial query from user;
 - 5 b) identifying an appropriate knowledge base to answer the query by a data source index;
 - c) sending the query to the appropriate knowledge base for a local query;
 - d) storing the results from the local query in a computer memory;
 - 10 e) performing a transitive reasoning within the same knowledge base if the query involves a transitive property;
 - f) discovering equivalences of concepts and relations to the answers of the query in other knowledge bases;
 - g) generating new queries from the discovered equivalences;
 - 15 h) distributing queries to knowledge bases to perform reasoning until a timeout is reached or if there are no more answers found from the knowledge bases;
 - i) aggregating results from all knowledge bases; and
 - j) returning the results to user.
- 20 2. The method as claimed in claim 1, wherein the step of performing a transitive reasoning within the same knowledge base if the query involves a transitive property includes:
 - 25 a) replacing the subject of the initial query with the result received from the knowledge base to form a new query; and
 - b) finding results of the new query within the same knowledge base.
3. The method as claimed in claim 1, wherein the step of discovering equivalences of concepts and relations to the answers in other knowledge bases includes:
 - 30 a) finding explicit equivalent subjects, objects and predicates of the results of the queries; and
 - b) finding implicit equivalent subjects, objects and predicates of the results of the queries.

4. The method as claimed in claim 3, wherein the explicit equivalent subjects, objects and predicates of the results of the queries are found by using Web Ontology Language (OWL).
- 5 5. The method as claimed in claim 3, wherein the implicit equivalent subjects, objects and predicates of the results of the queries are found by using functional and inverse functional properties.

1/4

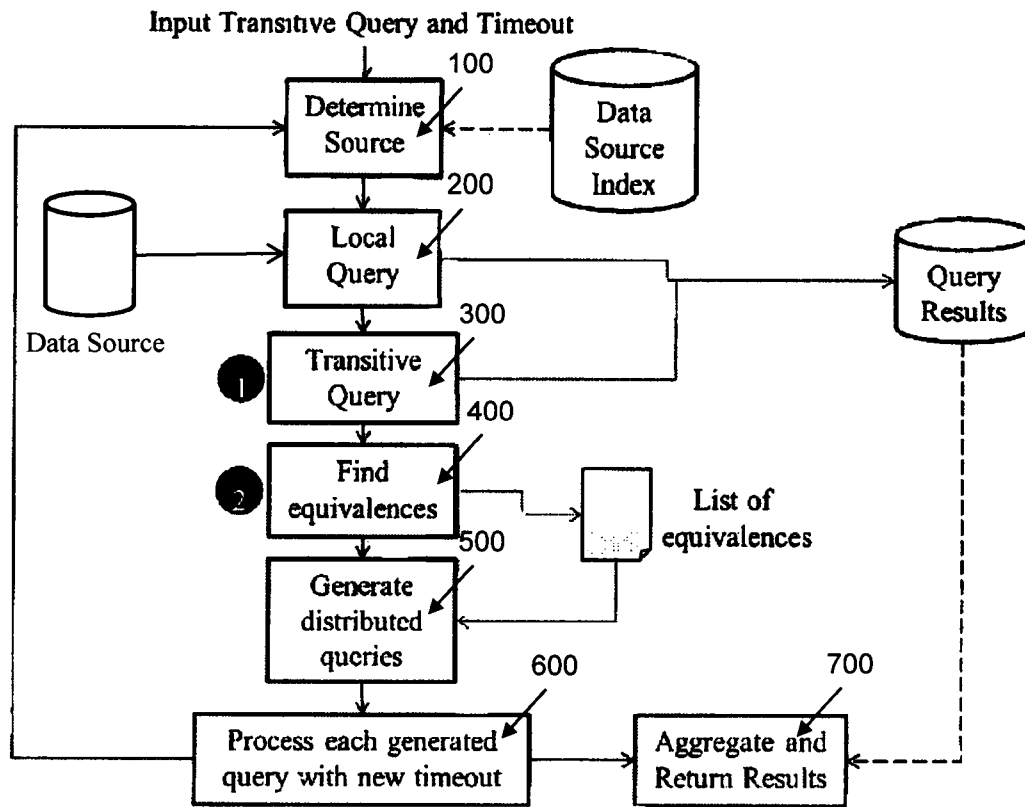


FIG. 1

2/4

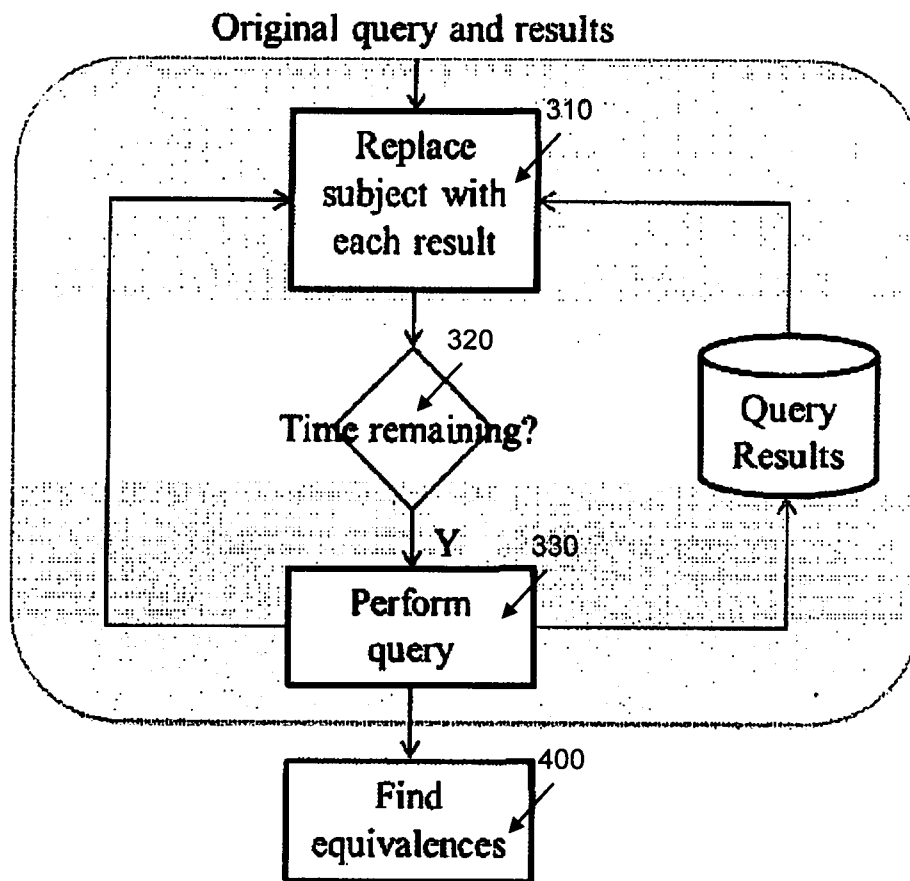


FIG. 2

3/4

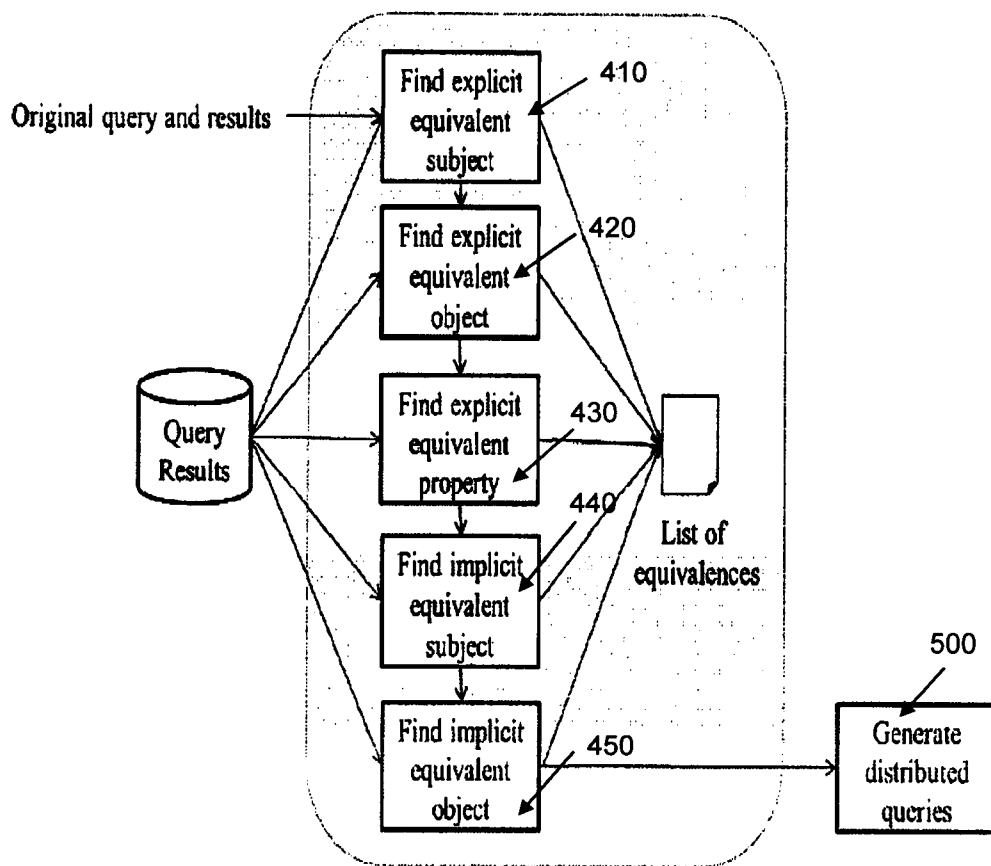


FIG. 3

4/4

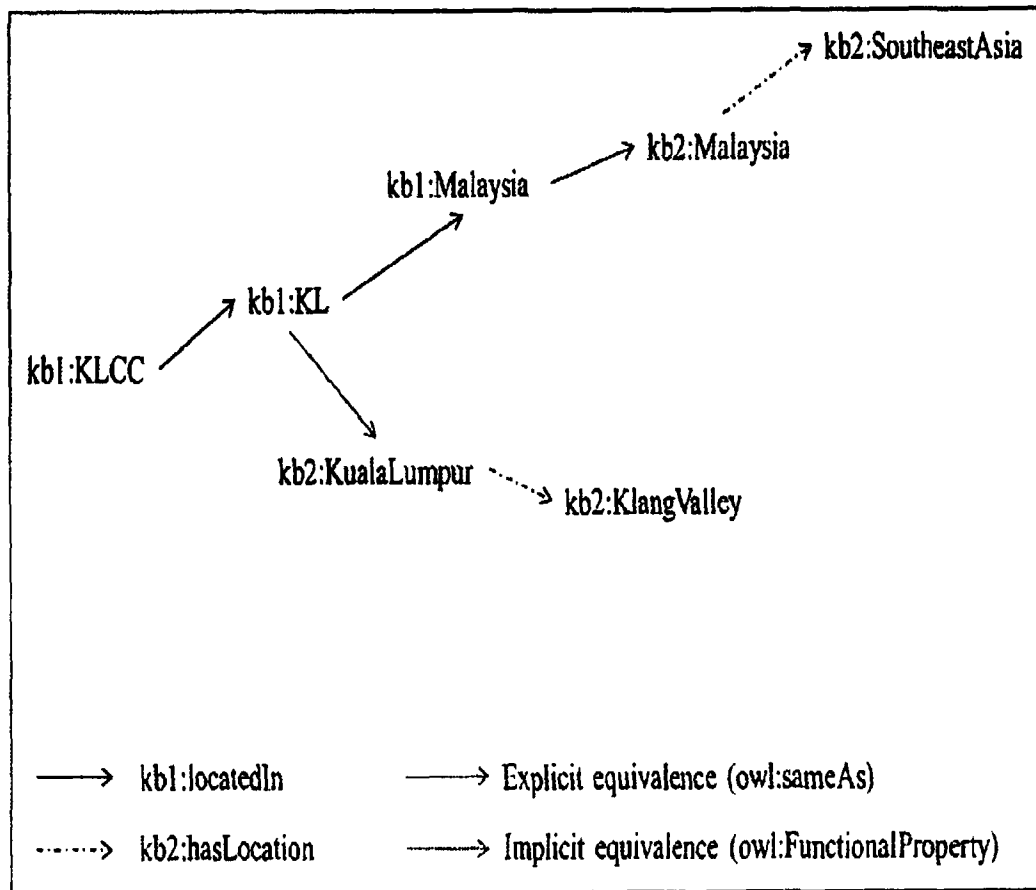


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/MY2014/000125

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06N5/02
ADD. G06F17/30

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)

G06N 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 practicable, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ARENS Y ET AL: "Query reformulation for dynamic information integration", JOURNAL OF INTELLIGENT INFORMATION SYSTEMS: ARTIFICIALINTELLIGENCE AND DATABASE TECHNOLOGIES, KLUWER ACADEMIC PUBLISHERS, AMSTERDAM, NL, vol. 6, no. 2-3, 30 June 1996 (1996-06-30), pages 99-130, XP002312827, ISSN: 0925-9902, DOI: 10.1007/BF00122124 the whole document ----- -/--	1-5



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 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

5 March 2015

Date of mailing of the international search report

16/03/2015

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

Valencia, Erika

INTERNATIONAL SEARCH REPORT

International application No
PCT/MY2014/000125

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YINGJIE LI ET AL: "Using Reformulation Trees to Optimize Queries over Distributed Heterogeneous Sources", 7 November 2010 (2010-11-07), THE SEMANTIC WEB - ISWC 2010, SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PAGE(S) 502 - 517, XP019159316, ISBN: 978-3-642-17745-3 the whole document	1-5
A	----- US 2012/246154 A1 (DUAN SONGYUN [US] ET AL) 27 September 2012 (2012-09-27) abstract	1-5
A	----- BIRTE GLIMM ET AL: "SPARQL beyond Subgraph Matching", 7 November 2010 (2010-11-07), THE SEMANTIC WEB - ISWC 2010, SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PAGE(S) 241 - 256, XP019159298, ISBN: 978-3-642-17745-3 paragraph [0007]	1-5
A	----- Gregory Todd Williams: "Supporting Identity Reasoning in SPARQL Using Bloom Filters", CEUR-WS 2008, 31 December 2008 (2008-12-31), XP055174324, Retrieved from the Internet: URL:http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-350/paper8.pdf [retrieved on 2015-03-05] paragraph [0001] -----	5

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/MY2014/000125

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012246154 A1	27-09-2012	US 2012246154 A1	27-09-2012
		WO 2012129149 A2	27-09-2012
