



- (51) **International Patent Classification:**
G06F 17/30 (2006.01)
- (21) **International Application Number:**
PCT/EP2014/058624
- (22) **International Filing Date:**
28 April 2014 (28.04.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** HEWLETT-PACKARD DEVELOPMENT COMPANY L.P. [US/US]; 11445 Compaq Center Drive West, Houston, Texas 77070 (US).
- (72) **Inventor; and**
- (71) **Applicant (for US only):** VAQUERO GONZALEZ, Luis Miguel [ES/GB]; Longdown Avenue, Stoke Gifford, Bristol Avon BS34 8QZ (GB).
- (74) **Agents:** SWINDELL & PEARSON LTD et al.; 48 Friar Gate, Derby Derbyshire DE1 1GY (GB).

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.

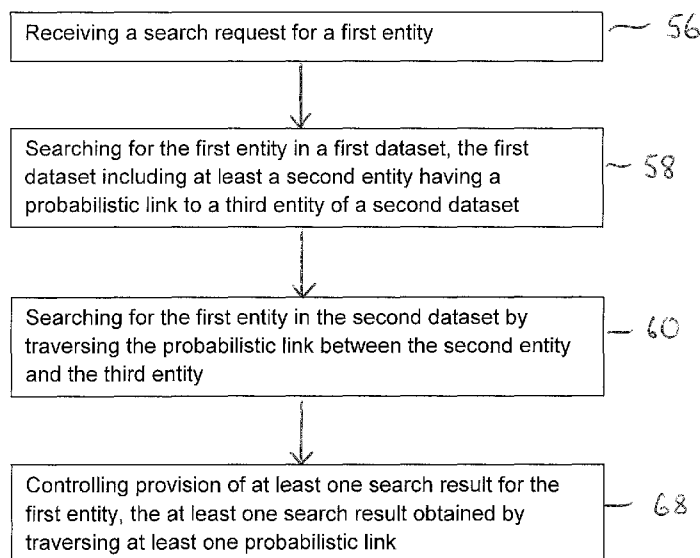
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, 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).

Declarations under Rule 4.17:

- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, — as to the identity of the inventor (Rule 4.17(i))

[Continued on next page]

- (54) **Title:** SEARCH DATASETS HAVING PROBABILISTIC LINKS THERE BETWEEN



(57) **Abstract:** A method to search a plurality of datasets having probabilistic links there between in which a search request for a first entity is received. The first entity is searched for in a first dataset, the first dataset including at least a second entity having a probabilistic link to a third entity of a second dataset. The first entity is searched for in the second dataset by traversing the probabilistic link between the second entity and the third entity. Provision of at least one search result for the first entity is controlled, the at least one search result obtained by traversing at least one probabilistic link.

FIG. 4



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

Published:

— *with international search report (Art. 21(3))*

[0001] TITLE

[0002] Search datasets having probabilistic links there between

5 [0003] BACKGROUND

[0004] Datasets are arranged to store information in a structured manner. For example, a dataset for a social network may store various details concerning a user, such as, user name, user date of birth, user interests, user workplace and so on. Datasets may, at least partially, store substantially the same entities but using different wording. For example, in a first dataset, a user's interests may include 'playing musical instruments', whereas, in a second dataset, the same user's interests may include 'musician'. Such entities may be connected via probabilistic links that indicate a less than certain relationship between the entities.

[0005] BRIEF DESCRIPTION

[0006] Reference will now be made by way of example only to the accompanying drawings in which:

[0007] Fig. 1 illustrates a schematic diagram of an apparatus according to an example;

25 [0008] Fig. 2 illustrates a multiplex graph according to an example;

[0009] Fig. 3 illustrates a flow diagram of a method according to an example;

[0010] Fig. 4 illustrates a flow diagram of another method according to an example; and

[0011] Fig. 5 illustrates a schematic diagram of a controller according to an example.

[0012] DETAILED DESCRIPTION

5

[0013] A dataset may be represented by a graph that includes a plurality of nodes (that represent entities) and connections between the nodes (that represent non-probabilistic links between the entities, and which may also be referred to as 'edges'). A multiplex graph includes a plurality of such graphs and may include connections between entities in different graphs (that represent probabilistic links between those entities).

10

[0014] A user may wish to search a multiplex graph for an entity. One approach is to view the multiplex graph on a display and visually inspect the multiplex graph for the entity. However, the multiplex graph may be relatively large and may take the user a relatively long time to find the entity.

15

[0015] As described in the following paragraphs, there is provided a method and an apparatus in which a search request for an entity is processed by traversing at least one probabilistic link between entities, and in which at least one search result is provided to the user.

20

[0016] A probabilistic link defines a relationship between two entities in different datasets, and provides an indication of the certainty of the relationship between the two entities. For example, a probabilistic link may indicate a 50% certainty that two entities in different datasets are related to one another.

25

[0017] In some examples, an apparatus may create a probabilistic link between entities by comparing at least one characteristic of the entities, (for example, entity name, entity attributes, structure connected to the

30

entity) to determine the probability that they are related. The created probabilistic link may then be stored in a memory of the apparatus.

[0018] Fig. 1 illustrates a schematic diagram of an apparatus 10 including a controller 12, input apparatus 14, and output apparatus 16. The apparatus 10 may also be referred to as a "computer apparatus", a "computer" or a "data storage apparatus". In some examples, the apparatus 10 may be a single machine where the input apparatus 14 and the output apparatus 16 are connected to the controller 12 via a wired or a wireless link, and the controller 12, the input apparatus 14 and the output apparatus 16 are located in close proximity to one another (for example, in the same room as one another). In other examples, the apparatus 10 may be a distributed apparatus where the input apparatus 14 and the output apparatus 16 are located remotely from the controller 12 (for example, in a different room, in a different building, in a different city, or in a different country).

[0019] In some examples, the apparatus 10 may be a module. As used herein, 'module' refers to a unit or apparatus that excludes certain parts/components that would be added by an end manufacturer or a user. For example, where the apparatus 10 is a module, the apparatus 10 may comprise the controller 12 and the remaining components (namely, the input apparatus 14 and the output apparatus 16) may be added by an end manufacturer.

[0020] The implementation of the controller 12 can be in hardware alone (for example, a circuit, a processor and so on), have certain aspects in software including firmware alone or can be a combination of hardware and software (including firmware).

[0021] The controller 12 may be implemented using instructions that enable hardware functionality, for example, by using executable computer program instructions 22 (machine readable instructions) in a general-

purpose or special-purpose processor 18 that may be stored on a computer readable storage medium 20 (disk, memory and so on) to be executed by such a processor 18.

5 [0022] The processor 18 is configured to read from and write to the memory 20. The processor 18 may also comprise an output interface via which data and/or commands are output by the processor 18 and an input interface via which data and/or commands are input to the processor 18.

10 [0023] The memory 20 stores a computer program 22 comprising computer program instructions that control the operation of the apparatus 10 when loaded into the processor 18. The computer program instructions 22 provide the logic and routines that enables the apparatus 10 to perform the methods illustrated in Figs. 3, 4 and described in the following paragraphs.
15 The processor 18 by reading the memory 20 is able to load and execute the computer program 22.

[0024] The computer program 22 may arrive at the apparatus 10 via any suitable delivery mechanism 24. The delivery mechanism 24 may be, for
20 example, a non-transitory computer-readable storage medium, a computer program product, a memory device, a record medium such as a compact disc read-only memory (CD-ROM) or digital versatile disc (DVD), an article of manufacture that tangibly embodies the computer program 22. The delivery mechanism 24 may be a signal configured to reliably transfer the
25 computer program 22. The apparatus 10 may propagate or transmit the computer program 22 as a computer data signal.

[0025] The memory 20 stores a plurality of datasets 26. The plurality of datasets 26 may be formed from a plurality of separate databases (that is,
30 the plurality of datasets 26 is provided by a plurality of separate database files). In other examples, the plurality of datasets 26 may be formed from separate data matrices within a single database (that is, the plurality of

5

datasets 26 is provided within a single database file). The memory 20 may receive new datasets which are then stored in the memory 20.

5 [0026] The input apparatus 14 may comprise any suitable apparatus for enabling a user to provide an input signal to the controller 12. For example, the input apparatus 14 may include at least one of a keyboard, a keypad, a computer mouse, and a touch screen display. The controller 12 is arranged to receive input signals from the input apparatus 14.

10 [0027] The output apparatus 16 may comprise any suitable apparatus for providing information to a user. For example, the output apparatus 16 may include at least one display (such as a liquid crystal display or a light emitting diode display). The controller 12 is arranged to control the output apparatus 16 to provide information to the user.

15

[0028] Fig. 2 illustrates a multiplex graph 28 for a plurality of datasets 26 according to an example. The plurality of datasets 26 includes a first dataset 261 for 'IT Support', a second dataset 262 for 'Operations', a third dataset 263 for 'Service Management', and a fourth dataset 264 for
20 'Service Marketing'.

[0029] The first dataset 261 is represented by a first graph 30 and includes the entities: host 32 and log 34. The log entities 34 are connected to the host entities 32 via non-probabilistic links (that is, a link having a probability of
25 100%, in other words, a certain relationship). As illustrated in Fig. 2, the non-probabilistic links are indicated by solid lines.

[0030] The second dataset 262 is represented by a second graph 36 and includes the entities: host 32, chassis 38, disk image 40 and host
30 monitoring metrics 42. The disk image entities 40 and the host monitoring metric entities 42 are connected to the host entities 32 via non-probabilistic

links. The chassis entity 38 interconnects the host entities 32 via non-probabilistic links.

5 [0031] The third dataset 263 is represented by a third graph 44 and includes the entities: host 32, service 46, service description 48, and service monitoring metrics 50. The service entities 46 are connected to one another and to host entities 32, service description entities 48 and to service monitoring metric entities 50 via non-probabilistic links.

10 [0032] The fourth dataset 264 is represented by a fourth graph 52 and includes the entities: service 46, service description 48 and service SLA 54. The service entities 46 are connected to one another and to the service description entities 48 and to service SLA entities 54 via non-probabilistic links.

15 [0033] The host entities 32 of the first graph 30 are connected to the host entities 32 of the second graph 36 via probabilistic links (that is, a link having a probability greater than 0% and less than 100%, in other words, an uncertain relationship). As illustrated in Fig. 2, the probabilistic links are
20 indicated by dashed lines. The host entities 32 of the second graph 30 are connected to the host entities 32 of the third graph 44 via probabilistic links. The service entities 46 of the third graph 44 are connected to the service entities 46 of the fourth graph 52 via probabilistic links. The service description entities 48 of the third graph 44 and connected to the
25 service description entities 48 of the fourth graph 52 via probabilistic links.

[0034] The first, second, third and fourth graphs 30, 36, 44, and 52 share some entities. For example, the first, second and third graphs comprise host entities 32. However, the shared entities may have at least one
30 difference in the different graphs. For example, the host entities 32 may have a different entity name in each of the first, second and third graphs 30, 36 and 44.

[0035] The operation of the apparatus 10 is described in the following paragraphs with reference to Fig. 3.

5 [0036] At block 56, the controller 12 receives a search request for a first entity. For example, a user may operate the input apparatus 14 to provide a search request for a service description entity 48 to the controller 12.

10 [0037] At block 58, the controller 12 searches for the first entity in a first dataset, the first dataset including at least a second entity having a probabilistic link to a third entity of a second dataset. The controller 12 may search the first dataset for the first entity by traversing the non-probabilistic links between the entities.

15 [0038] For example, the controller 12 may search for a service description entity 48 in the first dataset 261. The first dataset 261 includes a host entity 32 that has a probabilistic link to a host entity 32 in the second dataset 262. The first dataset 261 does not include a service description entity 48 and consequently, the method moves to block 60.

20

[0039] At block 60, the controller 12 may determine whether the probabilistic link between the second entity of the first dataset and the third entity of the second dataset has a probability greater than a threshold probability value.

25 [0040] For example, the threshold probability value may be 50% and at block 60, the controller 12 determines whether the probabilistic link between the host entities in the first and second datasets 261, 262 is equal to or greater than 50%.

30 [0041] In some examples, the threshold probability value may be selected by a user. For example, the user may operate the input apparatus 14 to

8

select a threshold probability value. The controller 12 may then store the selected threshold probability value in the memory 20.

5 [0042] If the probability of the probabilistic link is equal to or greater than the threshold probability value, the method moves to block 62. If the probability of the probabilistic link is less than the threshold probability value, the method returns to block 60 and the controller 12 determines whether another probabilistic link from the first dataset has a probability equal to or greater than the threshold probability value.

10

[0043] At block 62, the controller 12 searches for the first entity in the second dataset by traversing the probabilistic link between the second entity and the third entity. The controller 12 may search the second dataset for the first entity by traversing the non-probabilistic links between the entities.

15

[0044] For example, the controller 12 may search for a service description entity 48 in the second dataset 262. The second dataset 262 includes a host entity 32 that has a probabilistic link to a host entity 32 in the third dataset 263. The second dataset 262 does not include a service description entity 48 and consequently, the method moves to block 64.

20

[0045] At block 64, the controller 12 may determine whether searching has already been performed in a dataset and prevents searching in that dataset where the dataset has already been searched.

25

[0046] For example, where the controller 12 is unable to find a service description entity 48 in the second dataset 262, the controller 12 determines that searching may be continued in a different dataset by traversing a probabilistic link. The second dataset 262 is connected to the first dataset 261 and to the third dataset 263 via probabilistic links. The controller 12 determines that searching has already been performed in the

30

first dataset 261 and consequently, returns to block 60 to determine whether a probabilistic link to the third dataset 263 may be traversed.

5 [0047] In some examples, the controller 12 may prevent searching in a dataset where searching has already been performed by traversing the multiplex graph 28 in a single direction (for example, from the left to the right as illustrated in Fig. 2). In other examples, the controller 12 may store the datasets where searching has already been performed for received search requests in the memory 20 and may consult the memory 10 20 to determine whether searching may be performed in a dataset.

[0048] The controller 12 may then perform block 62 for the remaining datasets (where they are connected via probabilistic links). For example, the controller 12 may search for a service description entity 48 in the third 15 dataset 263 and in the fourth dataset 264.

[0049] At block 66, the controller 12 determines a plurality of search results for the first entity having different associated probabilities. For example, the controller 12 may determine that the two service description entities 48 in 20 the third dataset 263, and the two service description entities 48 in the fourth dataset 264 are search results for a search request for a service description entity.

25 [0050] The probability of the path traversed in the multiplex graph to provide a search result may be calculated by multiplying the probability of all of the connections between entities that are traversed between the starting entity and the search result.

30 [0051] In some examples, a set of per-query trees is used to store the probability of the links used to get to that entity. A tree stores the various paths between the starting entity and the search result entity as a plurality of branches. Since there may be several probabilistic paths to connect

10

any two entities, a tree may be used to store the probabilistic links traversed in those paths. When a probabilistic link is traversed, the probability value is added to the tree. For the branches in the tree, the probability of the path traversed to answer a query is calculated by multiplying the probability of all of the leafs in that branch of the tree from the root (starting entity) to the last leaf (the search result entity).

[0052] In some examples, the controller 12 may determine a single search result for the first entity at block 66. For example, a single search result may be determined where the multiplex graph 28 includes a single instance of the first entity. By way of another example, a single search result may be determined where the first entity is found in a dataset, and the search does not proceed to other datasets because the probabilistic links to those datasets have probabilities less than a threshold probability value.

[0053] At block 68, the controller 12 controls the provision of at least one search result for the first entity, the at least one search result obtained by traversing at least one probabilistic link. For example, the controller 12 may control the provision of the four service description entities 48 in order of highest probability. In some examples, the controller 12 may control provision of N search results, where N is an integer and is a subset of the plurality of search results determined in block 66. N may have a value of one, or may have a value greater than one.

[0054] At block 70, the controller 12 receives a user input signal from the input apparatus 14 to select a search result. For example, the user may operate the input apparatus 14 to select the search result having the highest probability. Where the user selects a search result not having the highest probability, the method may move to block 72.

[0055] At block 72, the controller 12 may change the multiplex graph 28 so that the probability of at least one probabilistic link is changed. For example, where the user selects a search result not having the highest probability, the controller 12 may lower the probability of at least one probabilistic link in the path to the entity that has the highest probability. By way of another example, where the user selects a search result not having the highest probability, the controller 12 may increase the probability of at least one probabilistic link in the path to the entity selected by the user.

[0056] The methods and apparatus described herein may provide several advantages. Firstly, the method is advantageous in that it enables a multiplex graph having probabilistic links to be searched, and for a probabilistic search result (or results) to be provided to the user. Secondly, since the controller 12 determines whether searching has already been performed in a dataset, the method may advantageously prevent the search request from infinitely looping around the multiplex graph. Consequently, the method may be reliable and deliver at least one result relatively quickly. Thirdly, the method may advantageously use the user's feedback to change the probabilities within the multiplex graph and thereby improve the multiplex graph.

[0057] The blocks illustrated in the Figs. 3 and 4 may represent steps in a method and/or sections of code in the computer program 22. The illustration of a particular order to the blocks does not necessarily imply that there is a required or preferred order for the blocks and the order and arrangement of the block may be varied in some examples. Furthermore, it may be possible for some blocks to be omitted in some examples. For example, Fig. 4 illustrates a flow diagram of another method according to an example. The method illustrated in Fig. 4 includes blocks 56, 58, 60 and 68 and omits blocks 62, 64, 66, 70 and 72.

[0058] Although examples have been described in the preceding paragraphs, it should be appreciated that modifications to the examples given can be made without departing from the scope as claimed.

5 [0059] Although the processor 18 is illustrated as a single component it may be implemented as one or more separate components some or all of which may be integrated/removable and/or may provide permanent/semi-permanent/ dynamic/cached storage.

10 [0060] Although the memory 20 is illustrated as a single component it may be implemented as one or more separate components some or all of which may be integrated/removable and/or may provide permanent/semi-permanent/ dynamic/cached storage.

15 [0061] Fig. 5 illustrates a schematic diagram of a controller 12 according to an example. The controller 12 includes a first search engine 74, a second search engine 76, a search result determination module 78, a search result provision module 80, a user input module 82, a dataset controller module 84, a first dataset 86, and a second dataset 88 having at least one
20 probabilistic link 90 to the first dataset 86. The modules illustrated in Fig. 5 may be provided by machine readable instructions (that is, they are software modules).

[0062] The first search engine 74 is to perform blocks 58 and 60. The second
25 search engine 76 is to perform blocks 62 and 64. The search result determination module 78 is to perform block 66. The search result provision module 80 is to perform block 68. The user input module 82 is to perform block 70. The dataset controller module 84 is to perform block 72.

30 [0063] The controller 12 illustrated in Fig. 5 is advantageous in that the controller 12 includes the first search engine 74 for querying the first dataset 86, and the second search engine 76 for querying the second

13

dataset 88 (and may include further search engines for querying further datasets). Consequently, the controller 12 may query the first and second datasets 86, 88 in parallel with one another (in other words, the first and second datasets 86, 88 may be queried concurrently for different search requests). This may increase the speed at which a multiplex graph may be searched for an entity.

5

[0064] Features described in the preceding description may be used in combinations other than the combinations explicitly described.

10

[0065] Although functions have been described with reference to certain features, those functions may be performable by other features whether described or not.

15

[0066] Although features have been described with reference to certain examples, those features may also be present in other examples whether described or not.

20

[0067] Whilst endeavoring in the foregoing specification to draw attention to those features believed to be of particular importance it should be understood that the Applicant claims protection in respect of any patentable feature or combination of features hereinbefore referred to and/or shown in the drawings whether or not particular emphasis has been placed thereon.

25

[0068] I/we claim:

30

CLAIMS

1. A method to search a plurality of datasets having probabilistic links there between, the method comprising:

5 receiving a search request for a first entity;
 searching for the first entity in a first dataset, the first dataset including at least a second entity having a probabilistic link to a third entity of a second dataset;

 searching for the first entity in the second dataset by traversing the
10 probabilistic link between the second entity and the third entity; and

 controlling provision of at least one search result for the first entity, the at least one search result obtained by traversing at least one probabilistic link.

2. The method as claimed in claim 1, further comprising determining a
15 plurality of search results for the first entity having different associated probabilities, and wherein controlling provision of at least one search result for the first entity comprises providing the plurality of search results in order of highest probability.

20 3. The method as claimed in claim 1, further comprising determining whether the probabilistic link between the second entity and the third entity has a probability greater than a threshold probability value; and wherein traversing the probabilistic link occurs when the probabilistic link has a probability greater than the threshold probability value.

25 4. The method as claimed in claim 1, further comprising receiving a user input signal to select a search result, wherein if the selected search result does not have the highest probability of the at least one search result for the first entity, the method further comprises changing the probability of at least one
30 probabilistic link.

5. The method as claimed in claim 1, further comprising determining whether searching has already been performed in a dataset; and preventing searching in that dataset where that dataset has already been searched.

5 6. The method as claimed in claim 1, wherein the searching in the first dataset is performed by a first search engine, and the searching in the second dataset is performed by a second search engine, different to the first search engine.

10 7. The method as claimed in claim 1, wherein the probabilistic link defines a relationship between the second entity and the third entity, and provides an indication of the certainty of the relationship between the second entity and the third entity.

15 8. An apparatus to search a plurality of datasets having probabilistic links there between, the apparatus comprising:

a controller to:

receive a search request for a first entity;

search for the first entity in a first dataset, the first dataset including at

20 least a second entity having a probabilistic link to a third entity of a second dataset, the probabilistic link defining a relationship between the second entity and the third entity and providing an indication of the certainty of the relationship between the second entity and the third entity;

25 search for the first entity in the second dataset by traversing the probabilistic link between the second entity and the third entity; and

control provision of at least one search result for the first entity, the at least one search result obtained by traversing at least one probabilistic link.

30 9. The apparatus as claimed in claim 8, wherein the controller is to determine a plurality of search results for the first entity having different associated probabilities, and to provide the plurality of search results in order of highest

16

probability when controlling provision of at least one search result for the first entity.

5 10. The apparatus as claimed in claim 8, wherein the controller is to determine whether the probabilistic link between the second entity and the third entity has a probability greater than a threshold probability value; and wherein traversing the probabilistic link occurs when the probabilistic link has a probability greater than the threshold probability value.

10 11. The apparatus as claimed in claim 8, wherein the controller is to receive a user input signal to select a search result, wherein if the selected search result does not have the highest probability of the at least one search result for the first entity, the controller is to change the probability of at least one probabilistic link.

15

12. The apparatus as claimed in claim 8, wherein the controller is to determine whether searching has already been performed in a dataset; and to prevent searching in that dataset where that dataset has already been searched.

20

13. The apparatus as claimed in claim 8, wherein the controller comprises a first search engine and a second search engine, the first search engine being different to the second search engine, the controller is to search in the first dataset using the first search engine, and the controller is to search in the
25 second dataset using the second search engine.

14. A non-transitory computer-readable storage medium encoded with instructions that, when performed by a processor, cause performance of a method in which:

30

a search request for a first entity is received;

17

the first entity is searched for in a first dataset, the first dataset including at least a second entity having a probabilistic link to a third entity of a second dataset;

5 the first entity is searched for in the second dataset by traversing the probabilistic link between the second entity and the third entity;

provision of at least one search result for the first entity is controlled, the at least one search result obtained by traversing at least one probabilistic link.

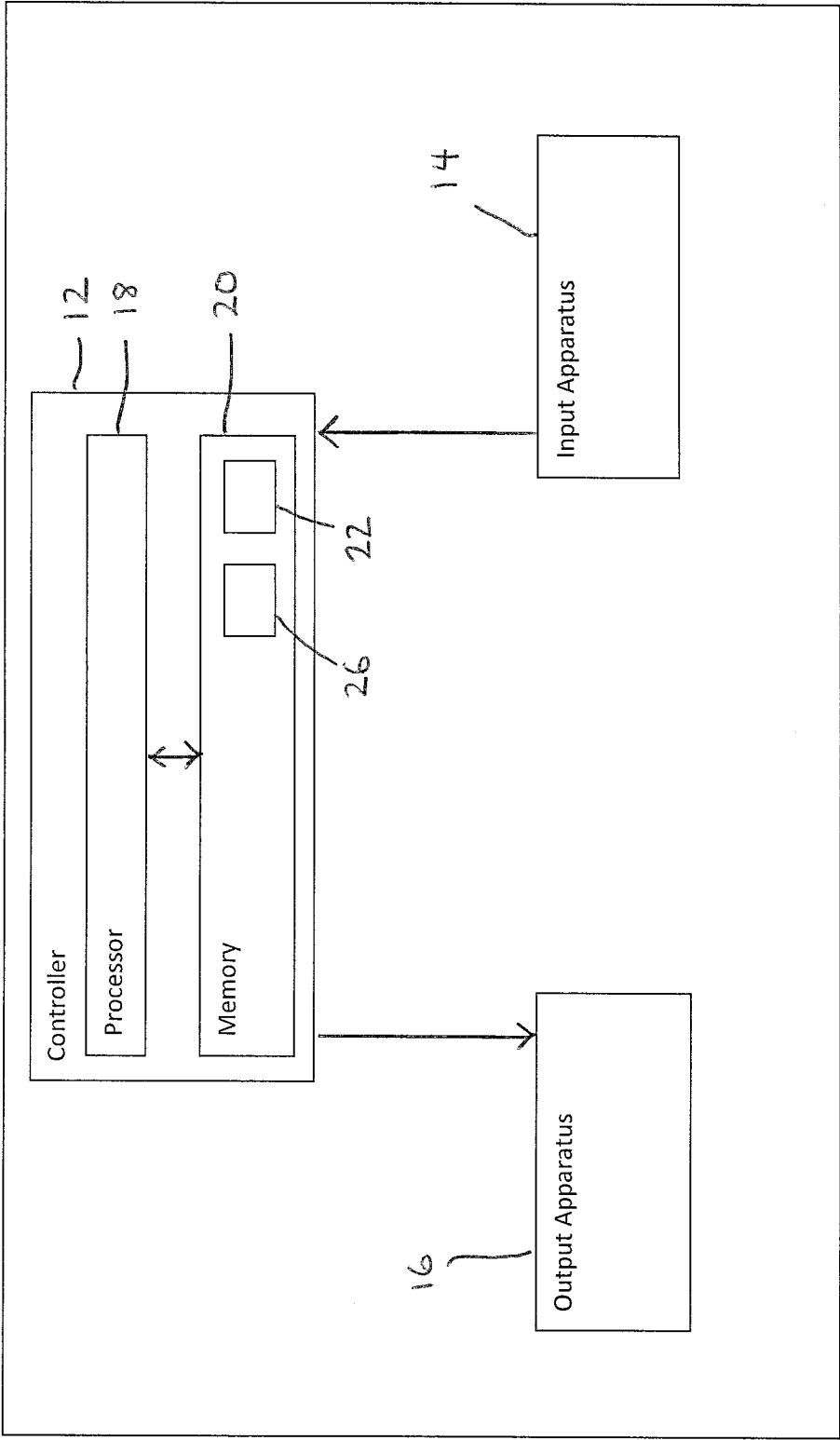
10 15. The non-transitory computer-readable storage medium as claimed in claim 14, encoded with instructions that, when performed by a processor, cause performance of a method in which:

a plurality of search results for the first entity having different associated probabilities is determined, and the plurality of search results is
15 provided in order of highest probability.

20

25

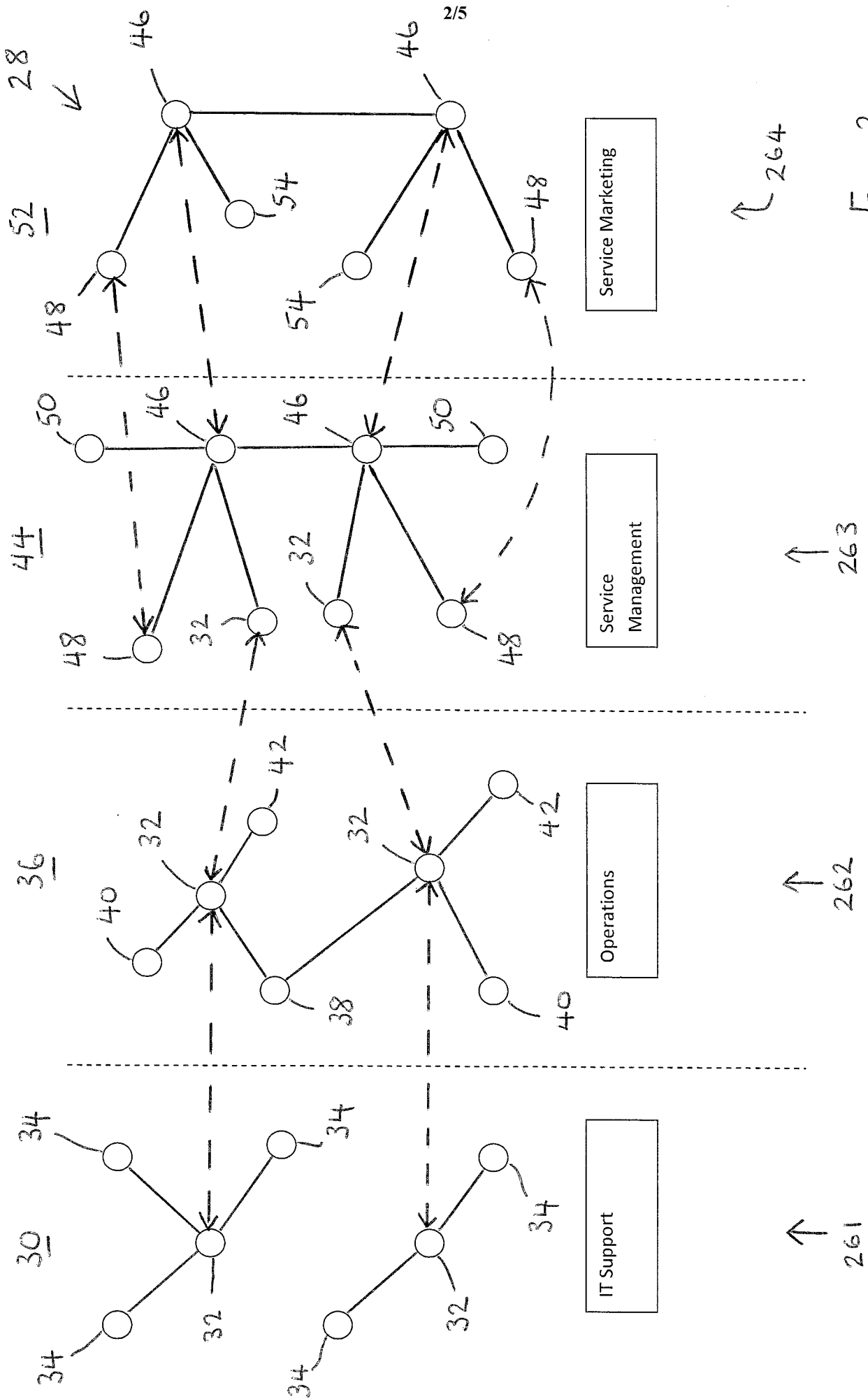
30



24

10

Fig. 1



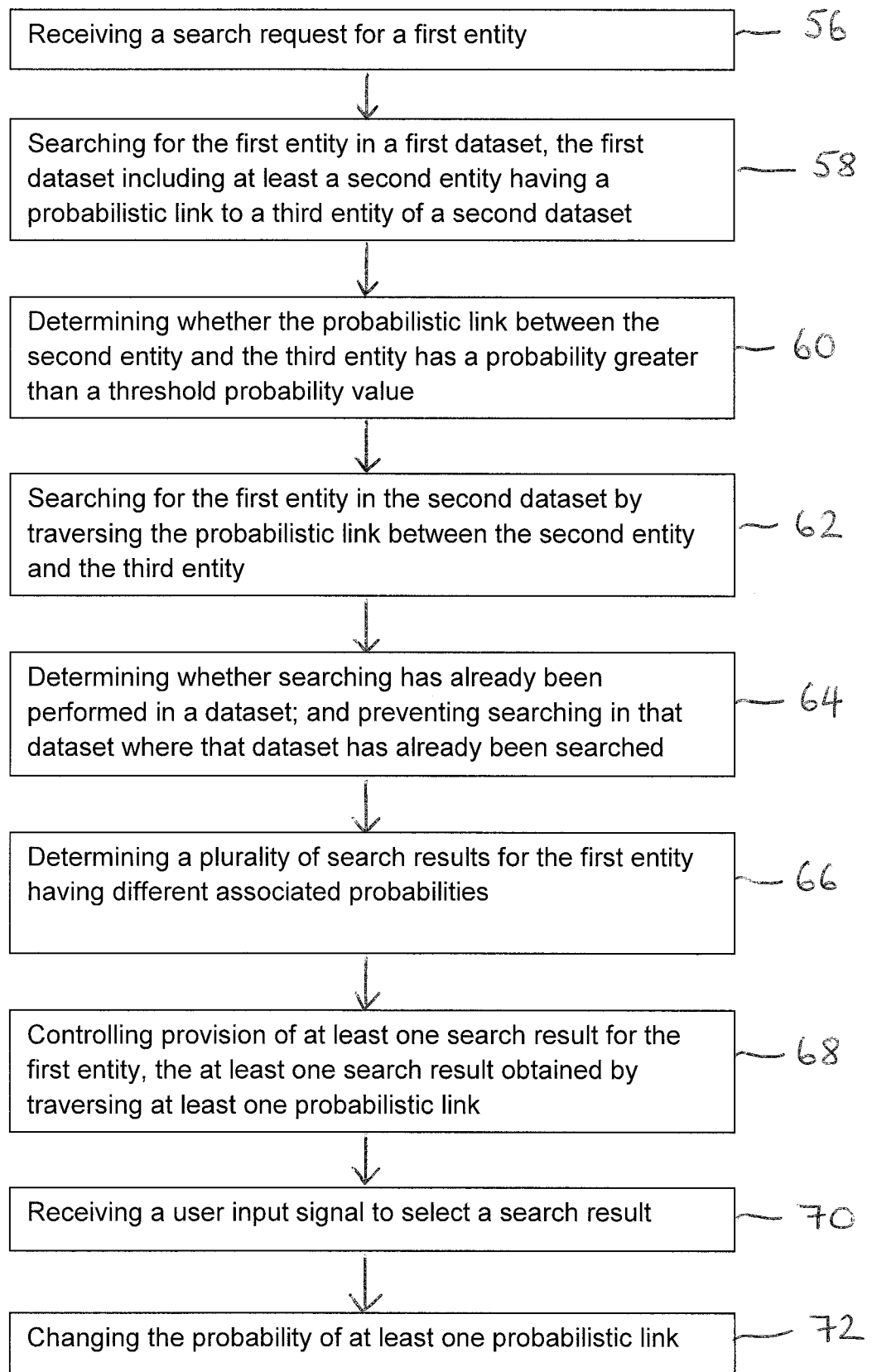


FIG. 3

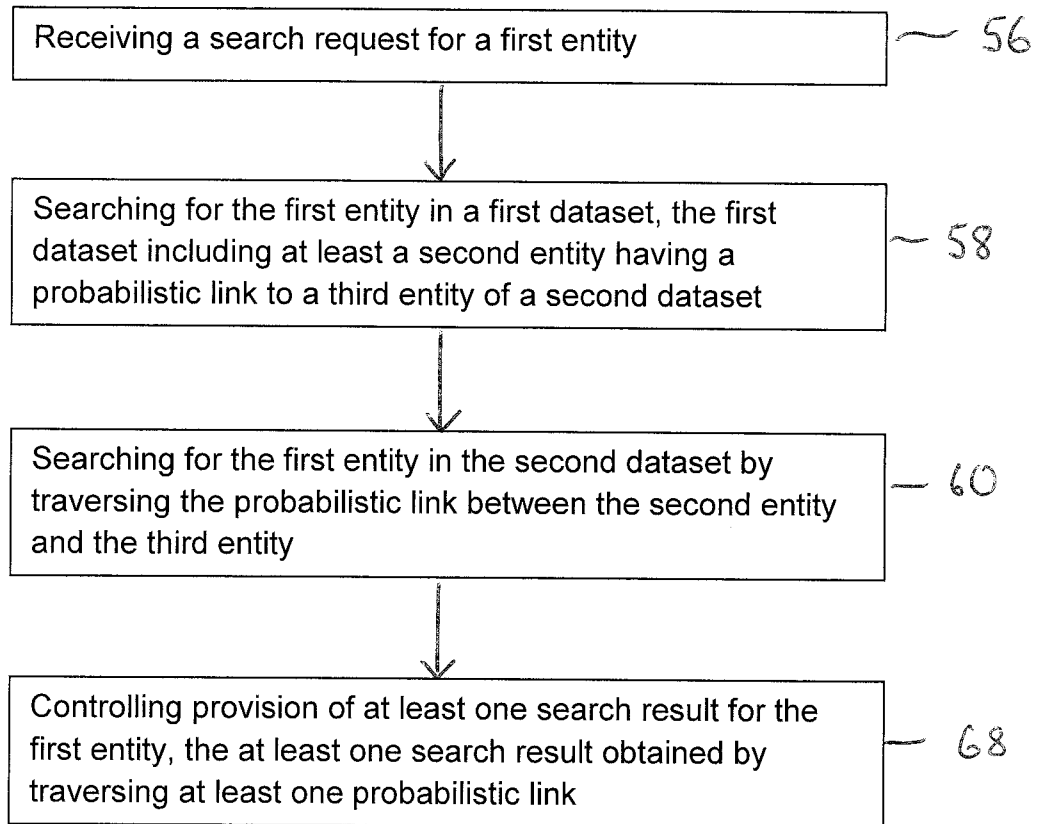


Fig. 4

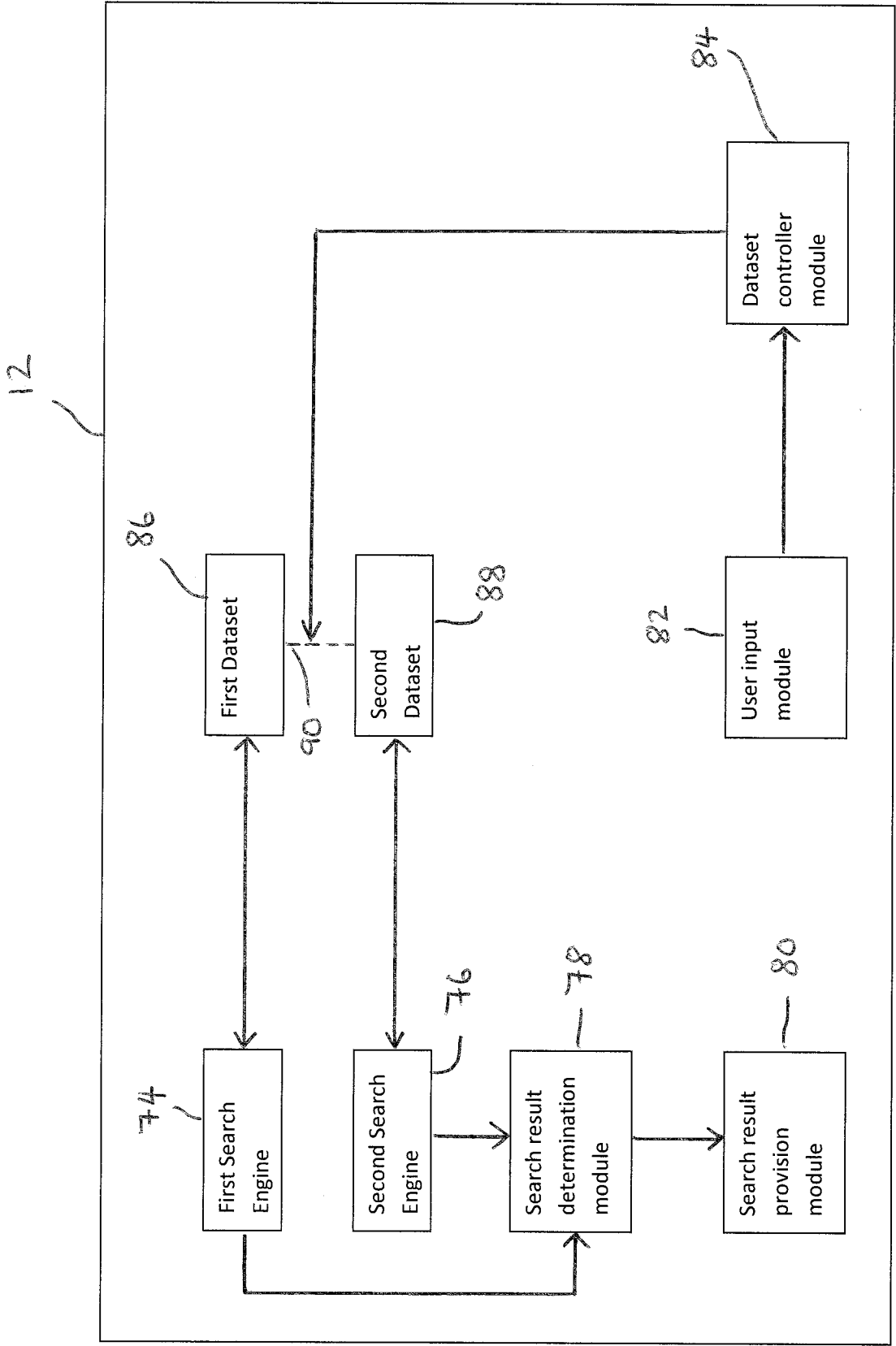


FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/058624

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F17/30
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

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

Electronic data base consulted during the international search (name of data base and, where 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	Anish Das Sarma ET AL: "UNCERTAINTY IN DATA INTEGRATION" In: "Managing and Mining Uncertain Data (Advances in Database Systems)", 11 February 2009 (2009-02-11), Springer, XP055159845, ISBN: 978-0-38-709689-6 pages 185-221, page 188, paragraph 1 - page 194, last paragraph; figures 7.1, 7.2, 7.3 page 195, paragraph 1 - paragraph 4 page 197, paragraph 1 - page 201, last paragraph ----- -/--	1-15

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

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

"E" earlier 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

23 December 2014

Date of mailing of the international search report

09/01/2015

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

Herri, Edmond

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2014/058624

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	XIN LUNA DONG ET AL: "Data integration with uncertainty", THE VLDB JOURNAL THE INTERNATIONAL JOURNAL ON VERY LARGE DATA BASES, 23 September 2007 (2007-09-23), pages 687-698, XP055159863, ISSN: 1066-8888 page 687 - page 692 page 695, left-hand column, paragraph 5 - page 696, left-hand column, paragraph 4 -----	1-15
A	Hans-Peter Kriegel ET AL: "PROBABILISTIC JOIN QUERIES IN UNCERTAIN DATABASES" In: "Managing and Mining Uncertain Data (Advances in Database Systems)", 11 February 2009 (2009-02-11), Springer, XP055159833, ISBN: 978-0-38-709689-6 pages 257-298, abstract page 263, last paragraph - page 274, last paragraph page 277 - page 282 -----	1-15
A	MERAY ET AL: "Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number", JOURNAL OF CLINICAL EPIDEMIOLOGY, PERGAMON, GB, vol. 60, no. 9, 7 August 2007 (2007-08-07), pages 883.e1-883.e11, XP022216976, ISSN: 0895-4356, DOI: 10.1016/J.JCLINEPI.2006.11.021 abstract page 883, right-hand column, last paragraph - page 885, left-hand column, paragraph 4 page 886, left-hand column, paragraph 3 - page 887, right-hand column, paragraph 3 -----	1-15
A	Lise Getoor ET AL: "Probabilistic Relational Models", 10 July 2010 (2010-07-10), XP055159783, Retrieved from the Internet: URL:http://wayback.archive.org/web/20100710112953/http://www.seas.upenn.edu/~taskar/pubs/prms-sr107.pdf [retrieved on 2014-12-19] abstract page 131, paragraph 1 - page 137, paragraph 1 page 163 -----	1-15