



- (51) **International Patent Classification:**
G06F 17/30 (2006.01)
- (21) **International Application Number:**
PCT/IB20 12/050821
- (22) **International Filing Date:**
23 February 2012 (23.02.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/036,799 28 February 2011 (28.02.2011) US
- (71) **Applicant (for all designated States except US):** **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (71) **Applicants (for MG only):** **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, North Harbour, Portsmouth Hampshire P06 3AU (GB). **IBM Japan Limited** [JP/JP]; 19-21 Nihonbashi, Hakozaiki-cho, Chuo-ku, Tokyo 103-85 10 (JP).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** **PANDIT, Sushain** [IN/US]; IBM Corporation, M/P 906-4B-007, 11501 Burnet Road, Austin, Texas 78758-3400 (US). **OBER-**

HOFER, Martin [DE/DE]; IBM Deutschland RD GmbH, Schoenaicher Strasse 220, D-71032 Boeblingen (DE). **WOLFSON, Charles, Daniel** [US/US]; IBM Corporation, M/P 906 4A-1 1, 11501 Burnet Road, Austin, Texas 78758-3400 (US). **BYRNE, Brian, Patrick** [IE/US]; IBM Corporation, Bldg 906, 11501 Burnet Road, Austin, Texas 78758-3400 (US).

(74) **Agent:** **ROBERTS, Scott**; IBM United Kingdom Limited, Intellectual Property Law Hursley Park, Winchester Hampshire S021 2JN (GB).

(81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on nextpage]

(54) **Title:** GENERATING A SEMANTIC GRAPH RELATING INFORMATION ASSETS

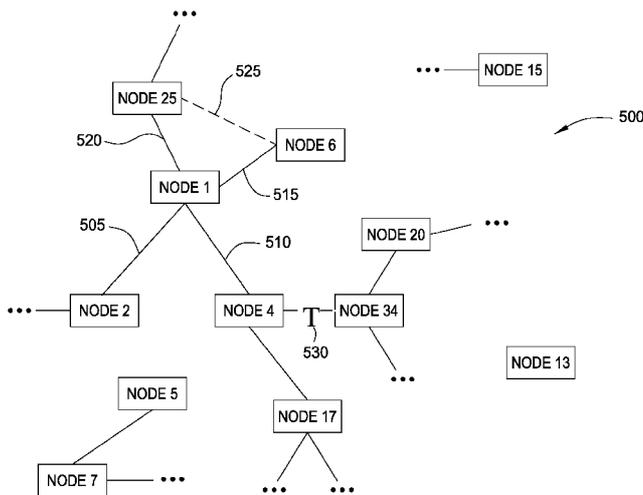


FIG. 5A

(57) **Abstract:** Embodiments of the invention provide an approach for creating, evolving and using a weighted semantic graph to manage and potentially identify certain information assets within an enterprise. The semantic graph may be generated by monitoring users navigating through search results which provide a set of information assets responsive to a search query. By recording the navigation path taken by many users, relationships between information assets may be identified. Further, once generated, the semantic graph may be used to present users with in indication of related information assets as part of the search results. Further still, the semantic graph may also be used to identify information assert "hubs" as well as information assets that may provide low utility to individuals within the enterprise.

W^o 2012/117318 A1

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

Published:

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

GENERATING A SEMANTIC GRAPH RELATING INFORMATION ASSETS

Technical Field

5

Embodiments of the invention are directed to techniques for managing information assets in enterprise environments. More specifically, embodiments of the invention provide an information asset management tool configured to capture and utilize crowd wisdom in order to identify and share relationships about information assets within the enterprise.

10

Background Art

Organizations commonly have to manage very large-scale information systems, which can include a very large number of both interrelated and independent information assets. While information assets can vary in nature, examples include structured systems such as traditional relational databases as well as unstructured systems such as content repositories and document stores. The degree of formality with which these systems are monitored, registered and managed can vary extensively within a large enterprise. An enterprise typically manages mission-critical information systems. However, such systems are frequently outnumbered by unmanaged systems (for example, spreadsheets, local database and document stores, as well as other information assets). More generally, it is not unusual for a large enterprise to manage thousands of distinct information repositories along with a (sometimes unknown) number of ad-hoc data stores and local working environments, which can themselves also, number in the thousands. As noted, the information assets of a given enterprise are frequently not independent of one another. For example, one information asset may store data extracts from another information asset. Similarly, information assets can share processing states during data integration (or during extract, transform, and load (ETL) processes), or provide related information repositories which store equivalent information segmented by line of business, and so on. Accordingly, in many cases, understanding the relationships between assets may be as important as understanding the assets themselves.

30

US Patent Application number 20100070448 (Omoigui, N, "System and method for knowledge retrieval, management, delivery and presentation", 24 June 2002) teaches searching methods using feedback, threshold information, and semantic weighting.

5 US Patent Application number 20060235841 (Betz, et al. "Page rank for the semantic web query", 14 April 2005) teach a searching method which provides weighted search results and weighing of semantic links, and semantic link graphing.

Therefore, there is a need in the art to address the aforementioned problem.

Summary of Invention

10

Viewed from a first aspect, the present invention provides a computer-implemented method for identifying relationships between a plurality of information assets managed by an enterprise, the method comprising: monitoring a plurality of users navigating paths through respective search results provided by an enterprise information asset management
15 application executed on a computing system having at least a processor and a memory; and generating, from the monitoring, a semantic graph capturing a set of relationships between information assets, of the plurality, based on a sequence of information assets selected by the plurality of the respective search results.

20

Advantageously preferred embodiments of the present invention may be used to help predict when a monitored metric indicates that an impending failure is likely to occur in order to allow a corrective action to be performed. One preferred embodiment of the present invention includes a method for identifying relationships between a plurality of information assets managed by an enterprise. The method may generally include monitoring a plurality
25 of users navigating paths through respective search results provided by an enterprise information asset management application executed on a computing system having at least a processor and a memory and generating, from the monitoring, a semantic graph capturing a set of relationships between information assets, of the plurality, based on a sequence of information assets selected by the plurality of the respective search results.

30

In a particular embodiment, the step of monitoring the plurality of users navigating paths through respective search results may include monitoring at least a first user selecting a first

information asset in the respective search results, and subsequently accessing a second information asset identified in the respective search results, and also include storing an indication of a relationship between the first and second information assets

5 Viewed from a second aspect, the present invention provides a system, comprising: a processor; and a memory storing an enterprise information asset management application program, which, when executed on the processor, performs an operation for identifying relationships between a plurality of information assets managed by an enterprise, the system comprising: a monitor component for monitoring a plurality of users navigating paths
10 through respective search results provided by the enterprise information asset management application, and a generate component for generating, responsive to the monitoring, a semantic graph capturing a set of relationships between information assets, of the plurality, based on a sequence of information assets selected by the plurality of the respective search results.

15 Additional embodiments include a computer-readable storage medium storing an application, which, when executed on a processor, performs the above recited method as well as a system having a processor and a memory storing an enterprise information asset management application program, which, when executed on the processor, performs the
20 above recited method.

Viewed from a further aspect, the present invention provides a computer program product for identifying relationships between a plurality of information assets managed by an enterprise, the computer program product comprising: a computer readable storage medium
25 readable by a processing circuit and storing instructions for execution by the processing circuit for performing a method for performing the steps of the invention.

Viewed from a further aspect, the present invention provides a computer program stored on a computer readable medium and loadable into the internal memory of a digital computer,
30 comprising software code portions, when said program is run on a computer, for performing the steps of the invention.

Brief Description of the Drawings

The present invention will now be described, by way of example only, with reference to preferred embodiments, as illustrated in the following figures:

5

Figure 1 illustrates an example of a distributed infrastructure having a variety of information assets, according to a preferred embodiment of the present invention;

10

Figure 2 illustrates an example computing system used to provide an enterprise information technology management tool, according to a preferred embodiment of the present invention;

15

Figure 3 illustrates a method for monitoring user interactions with search results in order to identify and share relationships about information assets within an enterprise, according to a preferred embodiment of the present invention;

20

Figure 4 illustrates a method for generating a semantic graph capturing relationships between information assets, according to a preferred embodiment of the present invention.

Figure 5A illustrates an example of semantic graph generated/evolved by monitoring users navigating through a paginated list of search results using the methods of Figures 3 and 4, according to a preferred embodiment of the present invention;

25

Figure 5B illustrates an example of information assets presented as query results, according to a preferred embodiment of the present invention;

30

Figure 6 illustrates a method for presenting a user navigating search results with information about asset relationships identified using a semantic graph, according to a preferred embodiment of the present invention;

Figure 7 illustrates a method for identifying information assets within an enterprise, according to a preferred embodiment of the present invention.

DETAILED DESCRIPTION

Some enterprise-level information management systems provide a search interface which allow users to access information asset descriptions, data models, and/or other similar objects residing in a database or (or stored as metadata). Often however, the results presented using such tools result in significant ambiguity and confusion for the user, particularly for large result-sets. For example, a query result which includes a large count of returned assets can require many pages to be displayed. Further, such results may be interrelated in a variety of interesting ways not accounted for in the result set. As a result, end-users who are unaware as to how the information assets are organized can get lost in a large number of results and are forced to simply jump from one page to the other, hoping to find to the particular related assets they need. Further, when a user accesses one element of a search result, the interface does not provide any visibility to additional related assets without performing a second explicit search.

Embodiments of the invention provide techniques for creating, evolving, and accessing a weighted semantic graph to manage (and potentially identify) related information assets within an enterprise. As used herein, "information assets" broadly refers to any form of information technology used by a given enterprise (for example, application servers, databases (and underlying tables and columns) data models, functions, jobs, scripts, ETL tools, network services, other server systems and applications, networking devices, appliance systems, etc. In one embodiment, an enterprise information management system may be configured with the appropriate tools (and metadata) needed to search for a specified set of information assets. For example, a web-server/application server/database system may be used to provide a search interface which relies on keywords. When a user searches for a given keyword (or words), information assets which match the keyword may be returned and presented to the user.

Once presented, the management tool monitors how a user interacts with the results, for example, by monitoring a user clicking through a sequence of results. For example, if a user clicks "information asset 1" and then "information asset 2" (within a configurable time window) the management tool registers a relationship in a weighted graph between the two.

When the user subsequently clicks on "information asset 3," a relationship between "information asset 2" and "information asset 3" is captured. Over time, if additional users click through the same path on search results, the weight between "information asset 1" and "information asset 2" is increased. If the weight assigned to a link reaches a user-
5 configurable weight, then search results which include these two elements are ordered to be next to one another. When a user selects one asset, data related to or describing that asset may be presented. Further, the search interface may inform a user of what information assets are related to a selected one. In one embodiment, the search interface may access the semantic graph and identify information assets related to the selected one. Doing so may
10 help inform the user about relationships to assets of which they were unaware - or may help them find related assets without having to hunt for them in a large set of results.

In one embodiment, the semantic graph captures relationships between assets based on user behavior, allowing the relative importance of information assets to be identified. The
15 stronger the edge weights to an "information asset" in the semantic graph, the more important that asset is to users within the enterprise. As noted, in large organizations, managing information assets is a significant problem; in particular, where some informally managed assets or ad-hoc systems become widely used. Accordingly in one embodiment, the relative weights of the links to such an asset may be used to identify informal systems
20 within the enterprise that are targets for to bring within the enterprises' formal information management structures. Conversely, a managed asset having few links in the semantic graph (or links with low weights) indicates the asset may be underused. Such assets may be identified as candidates for being shut down, or may be investigated to learn why the given asset is underutilized.

25
In the following, reference is made to embodiments of the invention. However, it should be understood that the invention is not limited to specific described embodiments. Instead, any combination of the following features and elements, whether related to different
embodiments or not, is contemplated to implement and practice the invention. Furthermore,
30 although embodiments of the invention may achieve advantages over other possible solutions and/or over the prior art, whether or not a particular advantage is achieved by a given embodiment is not limiting of the invention. Thus, the following aspects, features,

embodiments and advantages are merely illustrative and are not considered elements or limitations of the appended claims except where explicitly recited in a claim(s). Likewise, reference to "the invention" shall not be construed as a generalization of any inventive subject matter disclosed herein and shall not be considered to be an element or limitation of the appended claims except where explicitly recited in a claim(s).

As will be appreciated by one skilled in the art, aspects of the present invention may be embodied as a system, method or computer program product. Accordingly, aspects of the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system." Furthermore, aspects of the present invention may take the form of a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable signal medium or a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, or store a program for use by or in connection with an instruction execution system, apparatus or device.

A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier

wave. Such a propagated signal may take any of a variety of forms, including, but not limited to, electro-magnetic, optical, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that can communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device.

Program code embodied on a computer readable medium may be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc., or any suitable combination of the foregoing.

Computer program code for carrying out operations for aspects of the present invention may be written in any combination of one or more programming languages, including an object oriented programming language such as Java®, Smalltalk, C++ or the like and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Aspects of the present invention are described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the

computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

5 These computer program instructions may also be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

10 The computer program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other devices to cause a series of operational steps to be performed on the computer, other programmable apparatus or other devices to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the
15 functions/acts specified in the flowchart and/or block diagram block or blocks.

Embodiments of the invention may be provided to end users through a cloud computing infrastructure. Cloud computing generally refers to the provision of scalable computing resources as a service over a network. More formally, cloud computing may be defined as a
20 computing capability that provides an abstraction between the computing resource and its underlying technical architecture (for example, servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. Thus, cloud computing allows a user to access virtual computing
25 resources (for example, storage, data, applications, and even complete virtualized computing systems) in "the cloud," without regard for the underlying physical systems (or locations of those systems) used to provide the computing resources.

Typically, cloud computing resources are provided to a user on a pay-per-use basis, where
30 users are charged only for the computing resources actually used (for example, an amount of storage space consumed by a user or a number of virtualized systems instantiated by the user). A user can access any of the resources that reside in the cloud at any time, and from

anywhere across the Internet. In context of the present invention, a management tool describing a variety of information assets deployed to a computing cloud (whether the cloud itself is provided by the enterprise or a third party). For example, the monitoring application could monitor cloud based database systems, virtual machines, and a variety of other server applications.

The flowchart and block diagrams in the Figures illustrate the architecture, functionality and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

Further, particular embodiments of the invention are described a particular example management tool used to monitor a collection of information assets. However, it should be understood that the techniques described herein may be adapted to a variety of purposes in addition to the uses for predictive modeling analysis. Accordingly, references to the specific embodiment are included to be illustrative and not limiting.

Figure 1 illustrates an example of a distributed infrastructure 100 having a variety of information assets, according to one preferred embodiment of the present invention. As shown, the computing infrastructure 100 includes an enterprise information technology management system 105 and information assets 130, each connected to a communications network 120. The information assets 130 generally represent a collection of information technology resources managed by a given enterprise. In this particular example, the

information assets 130 include structured systems 1101-2 (for example, relational database systems storing enterprise data) server systems 1151-2 (for example, web servers, other web-based services and/or applications, or unstructured data/document storage such as a content management server), and unstructured information assets 125 (for example, spreadsheets, local databases or document stores, as well as other information assets).

As described in greater detail below, the management system 105 may be configured with the appropriate hardware and software applications needed to provide an interface to the information assets 130 managed by the enterprise. Such hardware and software applications may be configured to monitor users interacting with search results, and based on the monitoring, create and evolve a semantic graph representing the relationships between information assets 130 in the enterprise. For example, assume a search result provides a list of information assets 130 responsive to a search query. In such a case, when a user accesses successive items in the search results, a semantic graph is updated to create weighted relationships between elements in the search results. Thus, if a user first clicks on a search result listing database 1101 and then on an unstructured asset 125, the management tool registers between these two information assets as a weighted edge in a semantic graph between these two information assets.

Over time, if additional users navigate the same path through search results, the weight between database 1101 and unstructured asset 125 increases. That is, edges between information assets in the semantic graph are weighted according to the paths that previous users have followed while navigating search results. If the weight assigned to an edge between two information assets reaches a user-configurable threshold, then management system 105 may create a permanent semantic association between them. In one embodiment, however, newly created edges in the semantic graph may need to be confirmed by an administrator prior to a permanent semantic association being created. Further, edges between nodes in the semantic graph may include multiple weights, distinguished based on context. For example, context may be relative to an organizational role (for example, a business user versus an application developer). Of course, other contexts could be used.

Once a semantic relationship is established between two nodes (i.e., between two information assets 130), the management system 105 may be configured to enhance search results by presenting users with an indication of related information assets 130, based on the semantic graph. As an example, assume a user searches for assets that relate to "customer name." Typically, customer related information will be stored in multiple information assets across the enterprise. In such a case, in addition to returning the list of assets that support "customer name," the management system 105 also assists the user in understanding related information assets 130 for their query. Thus, when a user selects a result of the "customer name" query (for example, "Customer Mart"), they are presented with additional information assets that are related to that selected asset (perhaps "Customer ODS," "Customer Name Standardization Routines," and so on), where the additional information assets are identified identify using the semantic graph. Further, when a user related information asset, the semantic relationship between two assets may be reinforced, for example, by increasing the weight between the "Customer Mart" information asset and the "CustomerODS" information asset.

Additionally, metadata describing information assets 130 in the semantic graph may be mapped to a domain ontology. Doing so allows additional inter-asset relationships to be identified. For example, nodes in the semantic graph lacking edges (meaning the information asset represented by such a node does not have a relationship to any other asserts) may be classified according to the domain ontology and associated with other assets. Of course, the management system 105 may allow an administrator to manually associate one asset with another in the semantic graph. In one embodiment, the edge weights may be updated dynamically, as a user navigates through search results. Alternatively, the semantic graph may be updated periodically after a certain time window or after gathering a minimum amount of asset relationship data.

Figure 2 illustrates an example of the management system 105 used to provide an enterprise information technology management tool 221, according to one preferred embodiment of the present invention. As shown, the management system 105 includes, without limitation, a central processing unit (CPU) 205, a network interface 215, an interconnect 220, a memory 225, and storage 230. The computer system 200 may also include an I/O device interface

210 connecting I/O devices 212 (for example, keyboard, display and mouse devices) to the computer system 200.

5 In general, the CPU 205 retrieves and executes programming instructions stored in the memory 225. Similarly, the CPU 205 stores and retrieves application data residing in the memory 225. The interconnect 220 provides a communication path for transmitting programming instructions and application data between the CPU 205, I/O devices interface 210, storage 230, network interface 215, and memory 225. CPU 205 is included to be representative of a single CPU, multiple CPUs, a CPU having multiple processing cores, and
10 the like. And the memory 225 is generally included to be representative of a random access memory. The storage 230 may be a hard disk drive or solid state storage device (SSD). Further, although shown as a single unit, the storage 230 may be a combination of fixed and/or removable storage devices, such as fixed disc drives, floppy disc drives, tape drives, removable memory cards, optical storage, network attached storage (NAS), or a storage
15 area-network (SAN).

Illustratively, the memory 225 stores an enterprise IT management tool 221, which itself includes a query tool 222, a graphical user interface 224, a monitoring component 226 and a semantic graph generator 228. And storage 230 stores a semantic graph 232, an ontology
20 definition 234 and an information asset database 236. In one embodiment, the management tool 221 provides a software application configured to allow users to search for information assets, monitor users navigating search results, and build a semantic graph 232. Thus, the semantic graph 232 provides a representation of the relationships among information assets stored by the information asset database 236 learned by observing user behavior.

25 The monitoring component 226 may monitor a user navigating through search results and update or evolve the semantic graph 232 based on user navigation. When a user jumps from one information asset to another listed in search results, the monitoring component 226 can record (or update) an association between a first accessed element from the search results
30 and a second accessed element. In one embodiment, for such an association to be recorded, the user may need to jump from the first information asset to the second information asset within a specified amount of time. While the time may be configured as a matter of

preference, specific values may be set on a case-by-case basis. For example, assume a user queries for assets and is presented with a paginated result-set spanning hundreds of pages. As the user scans through the first page, hoping to find the wanted assets, the user notices a potential match and clicks on it for details. After going through the asset details, the user
5 (reasoning from his experience or domain knowledge) decides that to view a related asset. Following this, the user returns back to the paginated results and starts scanning through pages. Within the next few seconds (say less than ten seconds), the user spots another asset and clicks on it. And after doing so, spends between 2-5 minutes reading the details of this second asset and then, returns to the paginated results. In this scenario, two different
10 thresholds for time-periods are may be set. The first one is an upper threshold (desirable behavior is that the user 'quickly' clicks another related asset). However, the second one is a lower threshold (desirable behavior is that the user spends 'considerable' time reading the second asset). If instead, the user returns back to the paginated results too quickly (as specified by the second threshold), then the monitoring component 226 may determine that
15 there is no relationship between the first and second asset. Of course, one of ordinary skill in the art will recognize that the threshold (or thresholds) may be set in a variety of ways.

Once a specified number of users have traversed the same path between information assets, i.e., have accessed the first and second information assets in search results using the same
20 sequence in a manner that satisfies the thresholds selected for a particular case, the semantic graph generator 228 may add a permanent semantic association between two information assets in the semantic graph 232.

Once generated, the management tool 221 may present users with an indication of
25 relationships between information assets reflected in the semantic graph 232 when users navigate search results. To that end, graphical user interface 224 may allow users to specify query terms used to by the query tool 222 to search the information asset database 236. For example, the interface 224 may provide a text box which allows users to specify keywords (or other attributes) of a set of information assets a user is looking for. Once executed, the
30 results of the query are presented in the interface 224. For example, the interface 224 may present a list of information assets sorted by name, type, location, category, etc., or ranked according relevance. In one embodiment, for example, the most relevant information assets

may be determined as the assets in the results with the greatest number of edges in the semantic graph 232. Further, when a user selects one information asset, the interface may access the semantic graph 232 to identify edges to other, related information assets, and present the user with an indication of the related assets.

5

The ontology 234 may be used to infer additional relationships between information assets and further evolve the semantic graph 232. For example, the information asset database 236 may include a variety of descriptive metadata about a given information asset, for example, a name, a set of semantic tags, keywords, summaries, etc. In one embodiment, the ontology 234 may be used to normalize the information asset metadata to a common form specified by the particular domain ontology. Doing so allows related information assets to be identified based on the normalized terms from the ontology 234. For example, the semantic graph generator 228 may add new edges to reflect relationships identified using the ontology 234, perform transitive closures, sub or super-relationship inference, etc.

10

Figure 3 illustrates a method 300 for monitoring user interactions with search results in order to identify and share relationships about information assets within an enterprise, according to one preferred embodiment of the present invention. As shown, the method 300 begins at step 305 where the management tool receives a user query. As noted, such a query may include a set of user specified keywords describing the information assets a user is looking for. However, such a query could also be specified using attributes of the sought after information assets, specified, for example, by asset type, class, or category, or combinations of both. At step 310, the management tool identifies a set of information assets responsive to the query, and the results are then presented to a user at step 315. As noted, the graphical interface provided by the management tool may present a list of information assets sorted by name, type, location, category, etc.

15

20

25

At step 320, the management tool may begin to monitor a user interacting with the list of search results presented at step 315. At step 325, should the user access successive records in the result set within a specified time, then at step 330, the management tool records an indication of an association between two such assets. The monitoring at step 320 continues so long as a user navigates a given set of search results. Further method 300 may be

30

repeated for new searches by the same user as well as other users. Doing so allows the wisdom across multiple users in an enterprise to be captured over time. For example, the management tool may present users with related information assets when presenting search results.

5

Figure 4 illustrates a method 400 for generating a semantic graph representing relationships between information assets, according to one preferred embodiment of the present invention. As shown, the method 400 begins at step 405, where the semantic graph generator identifies information assets to include in the semantic graph. While this may include all of the assets in a given enterprise, it can also be limited to assets of a user-specified type, category or designation. For example, a semantic graph could be generated using only the "production" or "development" information assets of an enterprise.

10

At step 410, the semantic graph generator may build an initial semantic graph from known relationships between information assets. For example, using the method 300, over time, a series of relationships between information is identified while monitoring user behavior. In one embodiment, each such relationship is used to create an edge in the semantic graph. Further, each such edge may be assigned a weight, increased each time the path between the two particular assets is followed by a user. Once the weight on a given edge passes a user-specified threshold, a meaningful relationship between the nodes connected by the edge is presumed to have been identified. That is, the two information assets connected by the edge are related in a manner that is meaningful to users, as reflected in the history of multiple users navigating different search results. As another example the graph generator may create edges based on the relative strengths of the edges exiting a given node. For example, if one node includes many edges connected to other nodes, but includes one edge having a weight that exceeds all the others by a configurable threshold, then this edge may be modified into a permanent semantic association, while others are discarded.

20

25

30

Once generated, the edges in the semantic graph may be used to enhance the presentation of search results to a user. At step 415, the graph generator may evolve or update the semantic graph based on additional recorded feedback of users interacting with search results. That is, the semantic graph may continue to evolve over time, as new information assets are

identified and new certain paths are reinforced in response to users interacting with search results.

5 The graph generator may perform optional steps 420 and 425 to modify the semantic graph generated at steps 410 and 415. At step 420, the graph generator may identify nodes on the semantic graph without any edges to other nodes (or no edges that exceed a user-configurable threshold). Once identified, the graph generator may prompt an administrator to specify any desired edges from the identified nodes to other nodes in the semantic graph (or to prune such isolated nodes from the semantic graph). At step 425, the graph generator
10 may identify or modify nodes in the semantic graph based on a domain ontology. As noted, such ontology may allow related information assets to be identified from the normalized terms from the ontology. For example, the semantic graph generator may add, modify or remove edges to the semantic graph in order to reflect relationships identified using the ontology.

15 Figure 5A illustrates a portion of a semantic graph 500 generated and evolved by monitoring users navigating through a paginated list of search results using the methods of Figures 3 and 4, according to one preferred embodiment of the present invention. As shown, the graph 500 includes multiple nodes. Each node represents an information asset of an enterprise related to others by edges. In this example, an edge 505 connects node 1 to node 2, an edge 510 connects node 1 to node 4, an edge 515 connects node 1 to node 6, and an edge 520 connects
20 node 1 to node 25. Other edges in the semantic graph connect nodes as illustrated in Figure 5A. An edge connecting two nodes in graph 500 represents a relationship between those nodes created by the semantic graph generator in performing the method 400, as discussed above. Each node may have an associated weight indicating the relative strengths of the
25 learned relationship between nodes.

30 Figure 5B illustrates an example of information assets presented as query results 550, according to one preferred embodiment of the present invention. As shown, query results 550 are presented as an ordered list of information assets (labeled as elements e_1 through e_36). The ordered lists results 550 are also divided into six pages, and may be presented using any suitable graphical user interface. Assume for this example that the numerical

values of the nodes of semantic graph 500 of Figure 5A correspond to the numeric values in elements of list 550.

As noted, while the user interacts with the search results 550, the relationships between nodes in the semantic graph 500 may be used to update semantic graph 500 as well as present the user with an indication of related information assets. For example, assume the user presented with search results 550 first accesses node 1. In response, the interface may be updated to indicate that nodes 2, 4, 6 and 25 are related to node 1. Doing so may encourage the user to favor the related assets when navigating the search results, as well as simply help the user understand what information assets are related to a selected one.

Further, the semantic graph may be updated based on the path a user takes while navigating search results 550. For example, assume the user presented with search results 550 first accesses element e_4 (corresponding to node 4 of graph 500) and then accesses element e_34 (corresponding to node 34 of graph 500). In such a case, the semantic graph 500 may be updated with an edge 530 between node 4 and node 34. The weight assigned to edge 530 may be increased as other users navigate between the information assets associated with nodes 4 and 34. Once the weight assigned to edge 530 exceeds a user-specified threshold, then edge 530 may be updated as a permanent semantic link relating nodes 4 and 34.

Similarly, as noted above, domain ontology or other logical inferences may be used to update and modify the semantic graph 500. For example, assume a transitive closure operation is performed on semantic graph 500. In such a case, as node 1 is related to node 25 (by edge 520) and as node 1 is related to edge node 6 (by edge 515), an edge 525 may be added to the semantic graph 500. That is, as node 1 is related to both node 6 and node 25, node 6 and node 25 may be related to one another. In one embodiment, the transitive relationship between node 6 and node 25 (represented by edge 525) may need to be confirmed by an administrator prior to a permanent semantic association being created.

Figure 6 illustrates a method 600 for presenting a user navigating search results with information about asset relationships identified using a semantic graph, according to one preferred embodiment of the present invention. As shown, the method 600 begins at step

605, where the enterprise information management tool receives a user query. As noted, such a query may include a set of user specified keywords describing the information assets a user is looking for. At step 610, the management tool identifies a set of information assets responsive to the query and the results are then presented to the user.

5

Once displayed, the management tool waits for the user to select an element from the search results. At step 615, the user selects one of the information assets presented in the search results. At step 620, the management tool identifies assets in the semantic graph related to the one selected at step 625 and updates the interface to present the user with an indication of the related information assets. In one embodiment, related information assets may be identified, in part, based on contextual information about the user (for example, user role, etc.) submitting the query. That is, the related information assets shown to a given user may be tailored based on the relevance of certain assets to that user.

10

15

Optionally, at step 630, the management tool may remove unrelated assets from the display as well as present the user with annotations provided by other users characterizing the relationships between information assets. The interface may also allow the user to provide their own annotations or comments regarding the relationship between one information asset and another.

20

In addition to monitoring users interacting with search results to identify relationships between information assets and to using the identified relationships to enhance subsequent user-navigation of search results, the identified relationships may be used in other ways to manage the information assets of a given enterprise. For example, nodes in the semantic graph having a relatively large number of relationships to other nodes may be readily identified as relatively "important" information assets, as users have navigated to/from such an information asset from many others. By evaluating the information assets that form such "hubs" within the semantic graph, an enterprise can identify assets that should be formally administered and managed or given priority when allocating resources. That is, the common wisdom captured by monitoring users interacting with search results also surfaces assets which currently unmanaged assets. Conversely, some information assets may have relatively few links to other assets (or none at all). In the case of information assets

25
30

managed by the enterprise to provide data to certain users, the low utility of such "underperforming" assets may be identified from the semantic graph. Such information assets may be sunset to reduce operating costs in the appropriate case. Thus, the crowd wisdom captured in the semantic graph allows information governance structures within the enterprise to make decisions, such as reconciling relative asset importance to improve operating costs, improve data quality for an information asset which is used often but not formally managed within the enterprise (thus having low data quality or service). The crowd wisdom captured in the semantic graph may also be used to improve compliance issues. For example, rarely used assets which are not properly managed containing sensitive data might expose sensitive data to the risk of non-compliance with legal requirements.

Figure 7 illustrates a method 700 for identifying information assets within an enterprise, according to one preferred embodiment of the present invention. As shown, the method 700 begins at step 705 where the management tool retrieves the current semantic graph for a given set of information assets within the enterprise. As noted, the semantic graph may be generated according to the method 400 of Figure 4.

At step 710, the management tool may traverse the nodes of the semantic graph to identify information asset "hubs," i.e., information assets with a specified minimum number of edges to other information assets. Once identified, an enterprise may determine the operational state of such assets. For example, some information assets may begin as an ad-hoc local database used by a relatively small group of individuals. However, if such an information asset proves useful, the original group of individuals may share its existence with others outside of the initial group. Subsequently, as other users search for the local database and navigate to/from this asset in search results, edges to the local database form in the semantic graph. Once an edge count for a given node exceeds the threshold, the enterprise can identify the asset in the semantic graph as one to consider to formalize within the enterprise. Similarly, an enterprise may evaluate an information asset "hub" to determine whether it should be divided into smaller information assets. Doing so may result in information assets having more specialized functions than the information asset "hub," thereby reducing operational dependency on the information asset "hub" itself.

At step 710, the management tool may traverse the nodes of the semantic graph to identify information assets with few (or no) edges to others within the semantic graph. Doing so allows information assets having an edge count that falls below a specified threshold to be identified. As noted, a node having few edges in the semantic graph may indicate that the information asset corresponding to that node may be of limited utility to the enterprise. At step 720, the nodes (and corresponding information assets) identified at step 715 and 720 may be presented to a user.

Advantageously, embodiments of the invention provide an information asset management tool configured to capture and utilize crowd wisdom in order to identify and share relationships about information assets within the enterprise. As described above, a semantic graph may be generated by monitoring users navigating through search results which provide a set of information assets responsive to a search query. By recording the navigation path taken by many users, relationships between information assets may be identified. Further, once generated, the semantic graph may be used to present users with in indication of related information assets as part of the search results. Further still, the semantic graph may also be used to identify information assert "hubs" as well as information assets that may provide low utility to individuals within the enterprise.

While the foregoing is directed to embodiments of the present invention, other and further embodiments of the invention may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.

CLAIMS

1. A computer-implemented method for identifying relationships between a plurality of information assets managed by an enterprise, the method comprising:
5 monitoring a plurality of users navigating paths through respective search results provided by an enterprise information asset management application executed on a computing system having at least a processor and a memory; and
generating, from the monitoring, a semantic graph capturing a set of relationships
between information assets, of the plurality, based on a sequence of information assets
10 selected by the plurality of the respective search results.
2. The computer-implemented method of claim 1, wherein each respective search result identifies a subset of the plurality of information assets managed by the enterprise.
- 15 3. The computer-implemented method of either of claims 1 or 2, wherein the semantic graph includes a plurality of nodes, each corresponding to one of the plurality of information assets and includes a plurality of weighted edges, each indicating a relationship strength between two information assets connected by a respective edge.
- 20 4. The computer-implemented method of any of claims 1 to 3, wherein the method further comprises:
identifying at least a first node in the semantic graph having zero edges to other nodes; and
prompting an administrator to assign one or more edges to the first node.
25
5. The computer-implemented method of any of claims 1 to 4, wherein the information assets include at least one of a database system, a web service and a server application.
- 30 6. The computer-implemented method of any of claims 1 to 5, wherein monitoring the plurality of users navigating paths through respective search results comprises:

monitoring at least a first user selecting a first information asset in the respective search results, and subsequently accessing a second information asset identified in the respective search results; and

storing an indication of a relationship between the first and second information assets.

7. The computer-implemented method of claim 6, further comprising, assigning a weight to the relationship between the respective first and second information assets.

8. The computer-implemented method of claim 7, wherein the assigned weight is increased each time one of the plurality of users navigates from the first information asset in the respective search results to the second information asset identified in the respective search results, and further comprising, upon determining the assigned weight exceeds a specified threshold, updating the semantic graph with an edge connecting nodes in the semantic graph corresponding to the first and second information assets.

9. The computer-implemented method of any of claims 1 to 8, updating the semantic graph according to a domain ontology associated with the plurality of information assets.

10. A system, comprising:

a processor; and

a memory storing an enterprise information asset management application program, which, when executed on the processor, performs an operation for identifying relationships between a plurality of information assets managed by an enterprise, the system comprising:

a monitor component for monitoring a plurality of users navigating paths through respective search results provided by the enterprise information asset management application, and

a generate component for generating, responsive to the monitoring, a semantic graph capturing a set of relationships between information assets, of the plurality, based on a sequence of information assets selected by the plurality of the respective search results.

11. The system of claim 10, wherein each respective search result identifies a subset of the plurality of information assets managed by the enterprise.

5 12. The system of either of claims 10 or 11, wherein the semantic graph includes a plurality of nodes, each corresponding to one of the plurality of information assets and includes a plurality of weighted edges, each indicating a relationship strength between two information assets connected by a respective edge.

10 13. The system of any of claims 10 to 12, wherein the system further comprises:
an identify component for identifying at least a first node in the semantic graph having zero edges to other nodes; and
a prompt component for prompting an administrator to assign one or more edges to the first node.

15 14. The system of any of claims 10 to 13, wherein the information assets include at least one of a database system, a web service and a server application.

20 15. The system of any of claims 10 to 14, wherein the monitor component is further operable for:
monitoring at least a first user selecting a first information asset in the respective search results, and subsequently accessing a second information asset identified in the respective search results;
storing an indication of a relationship between the first and second information assets; and
25 assigning a weight to the relationship between the respective first and second information assets.

30 16. The system of claim 15, wherein the assigned weight is increased each time one of the plurality of users navigates from the first information asset in the respective search results to the second information asset identified in the respective search results, and , wherein the system further comprises, responsive to determining that the assigned weight exceeds a specified threshold, an update component for updating the semantic graph with an

edge connecting nodes in the semantic graph corresponding to the first and second information assets.

5 17. A computer program product for identifying relationships between a plurality of information assets managed by an enterprise, the computer program product comprising:
a computer readable storage medium readable by a processing circuit and storing instructions for execution by the processing circuit for performing a method according to any of claims 1 to 9.

10 18. A computer program stored on a computer readable medium and loadable into the internal memory of a digital computer, comprising software code portions, when said program is run on a computer, for performing the method of any of claims 1 to 9.

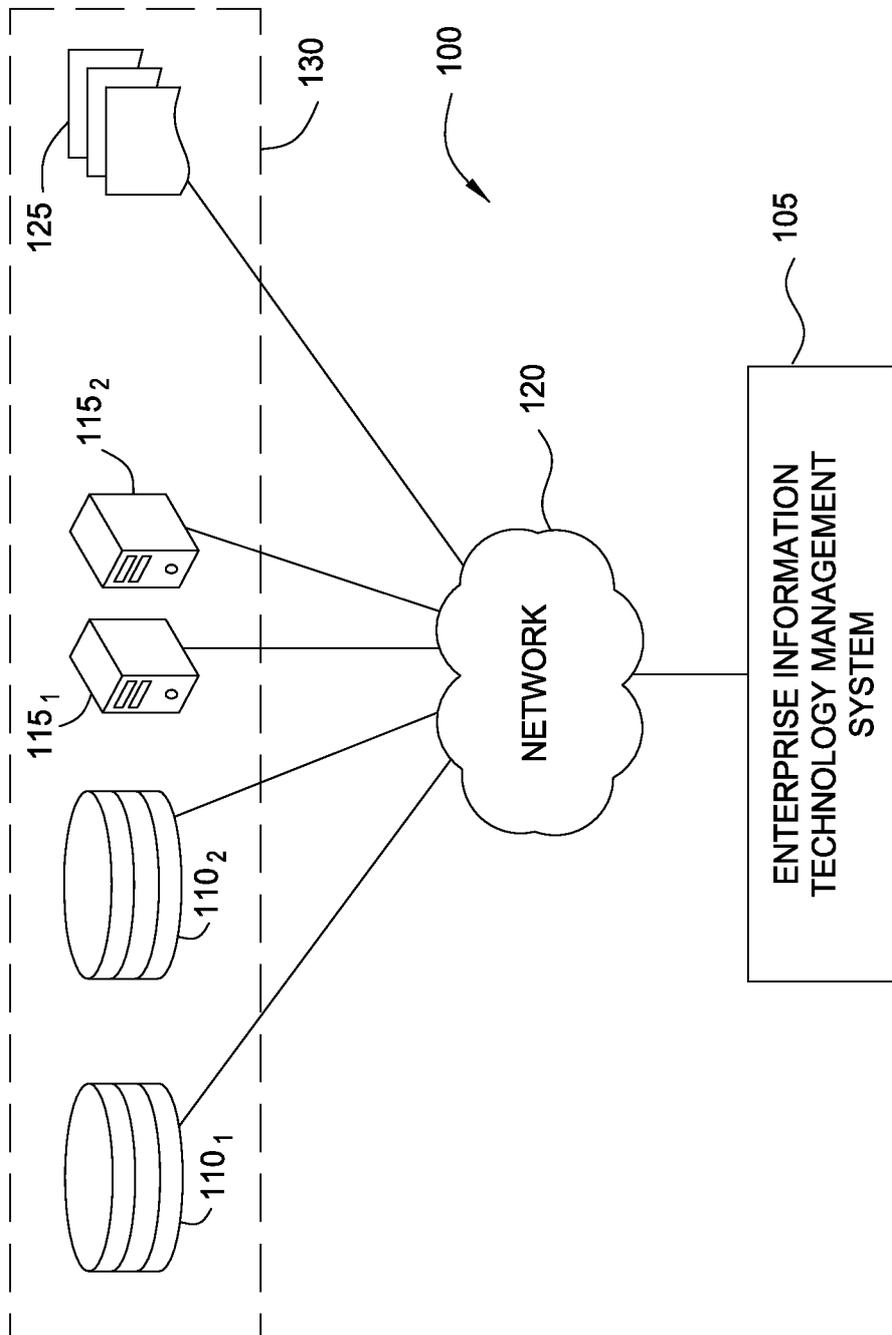


FIG. 1

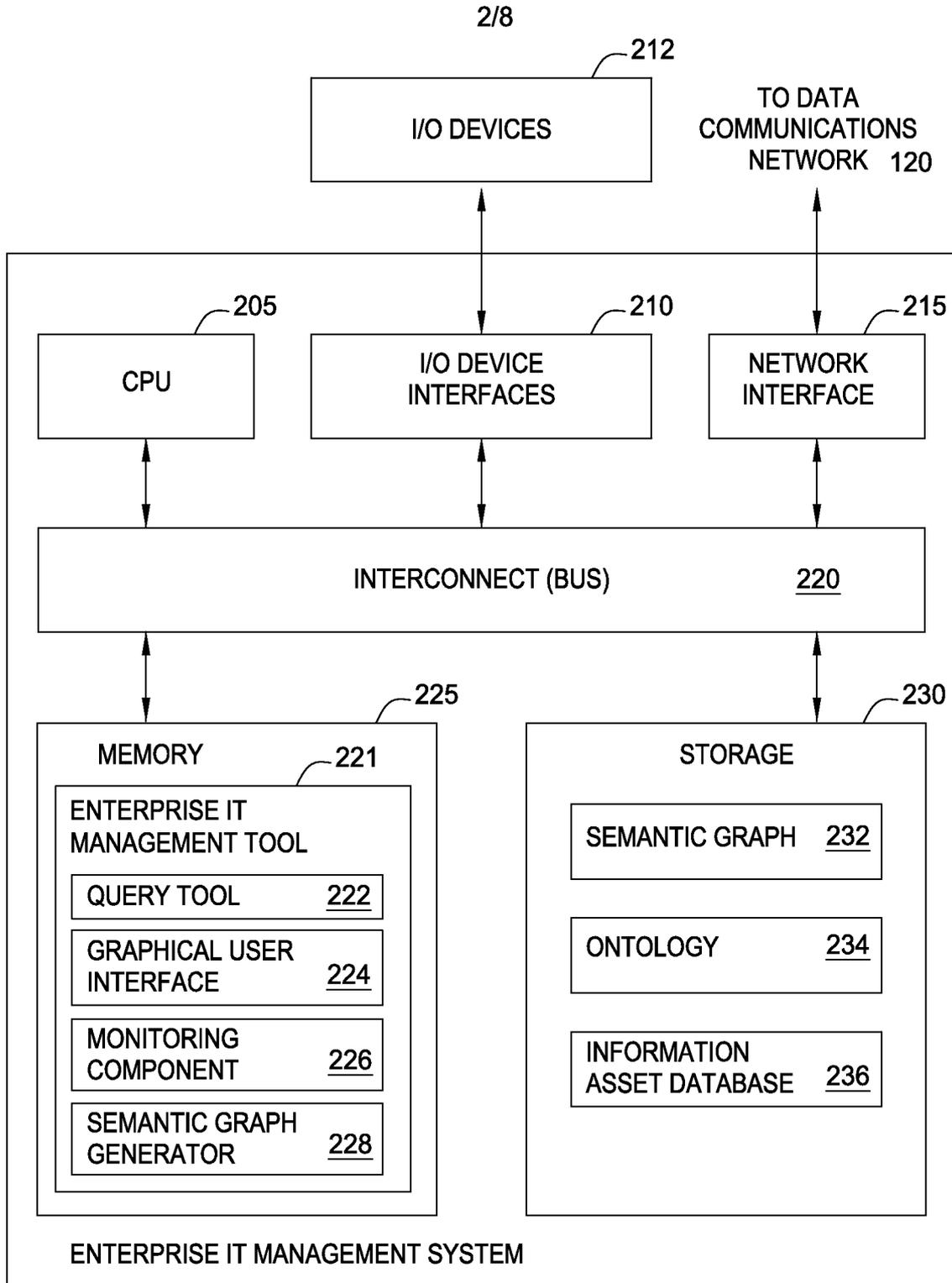


FIG. 2

105

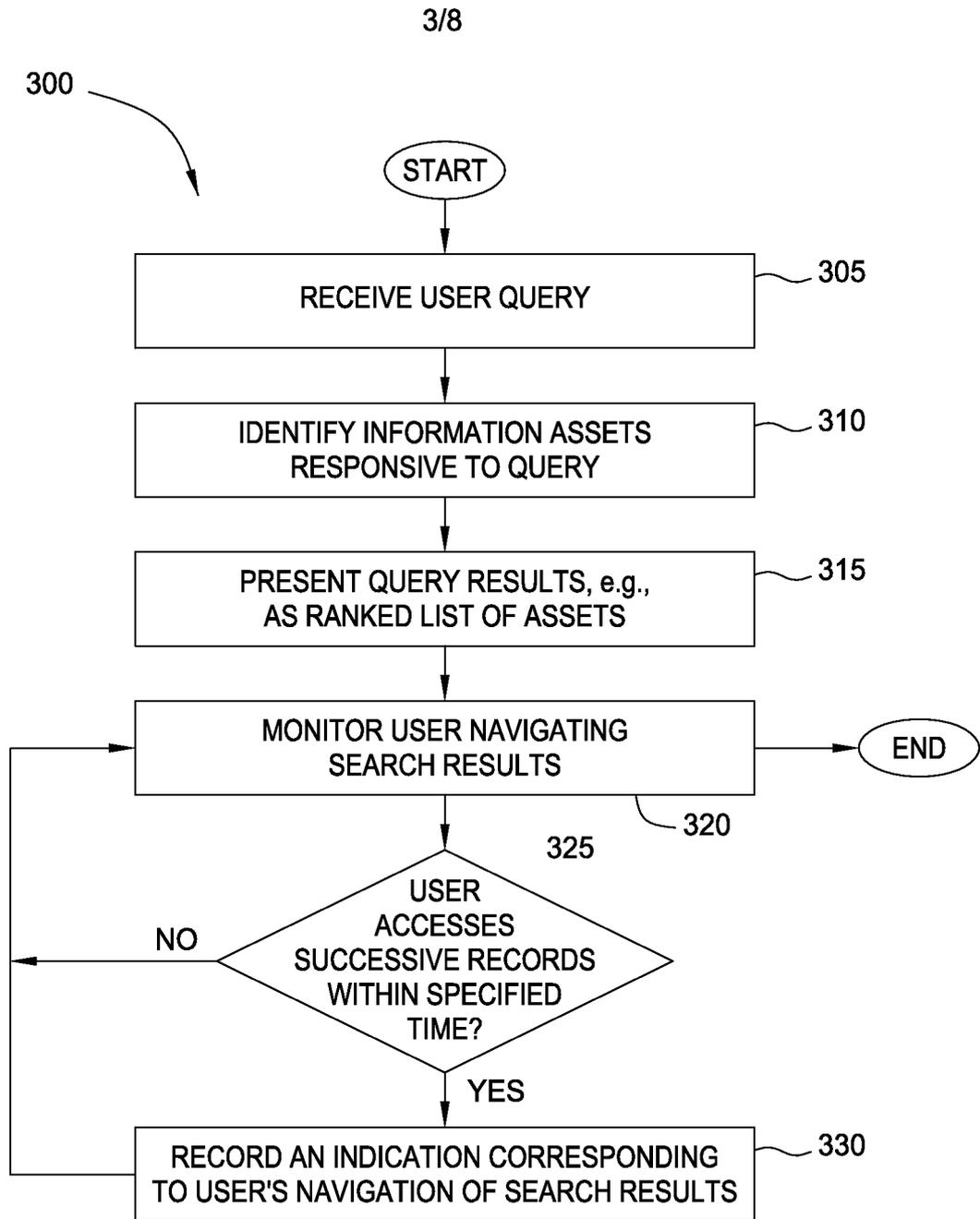


FIG. 3

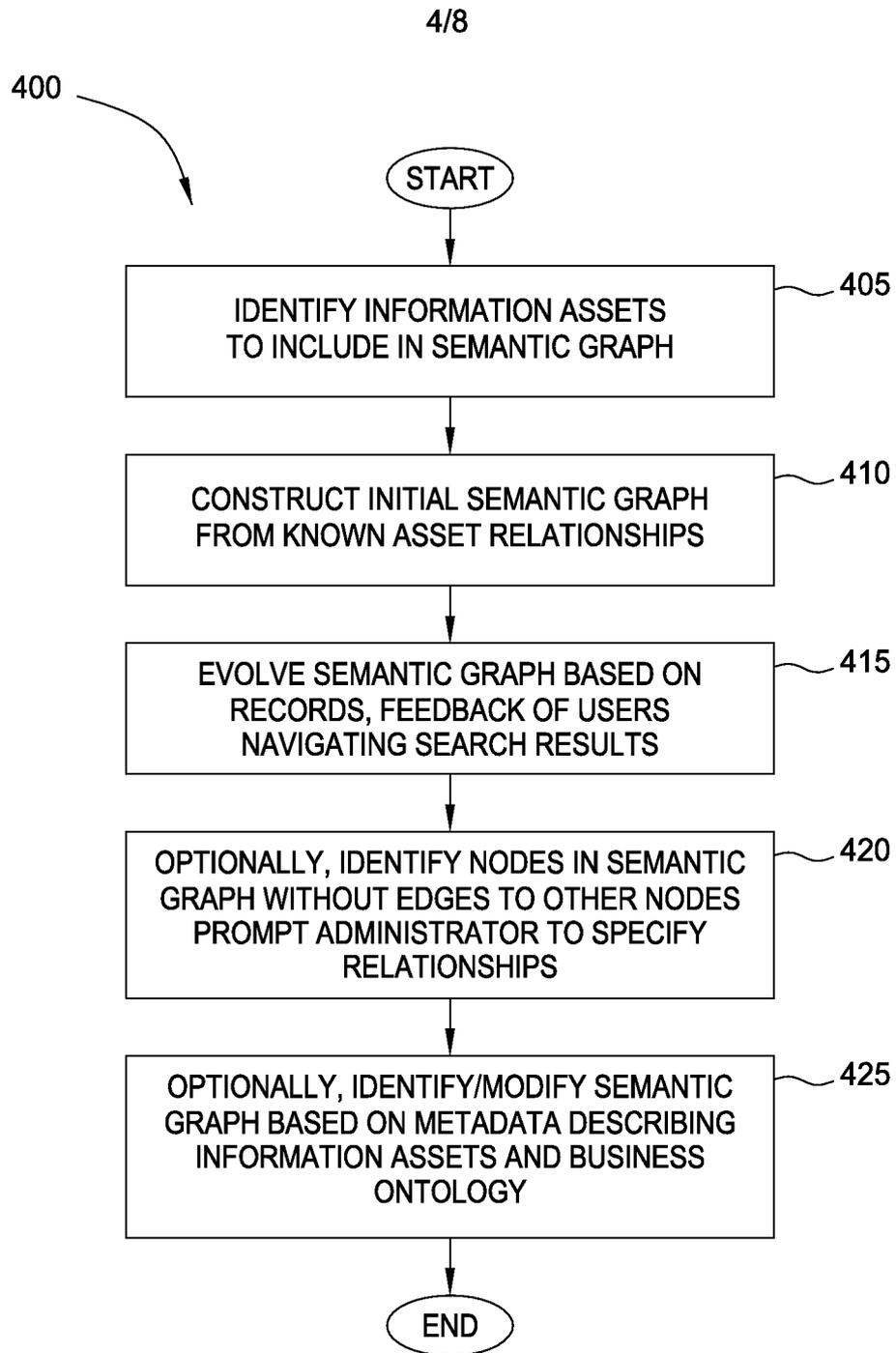


FIG. 4

PAGE 1	PAGE 2	PAGE 3	PAGE 4	PAGE 5	PAGE 6
e_1	e_7	e_13	e_19	e_25	e_31
.
.
.	e_34
.
e_6	e_12	e_18	e_24	e_30	e_36

550

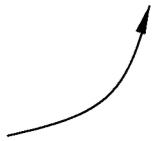


FIG. 5B

7/8

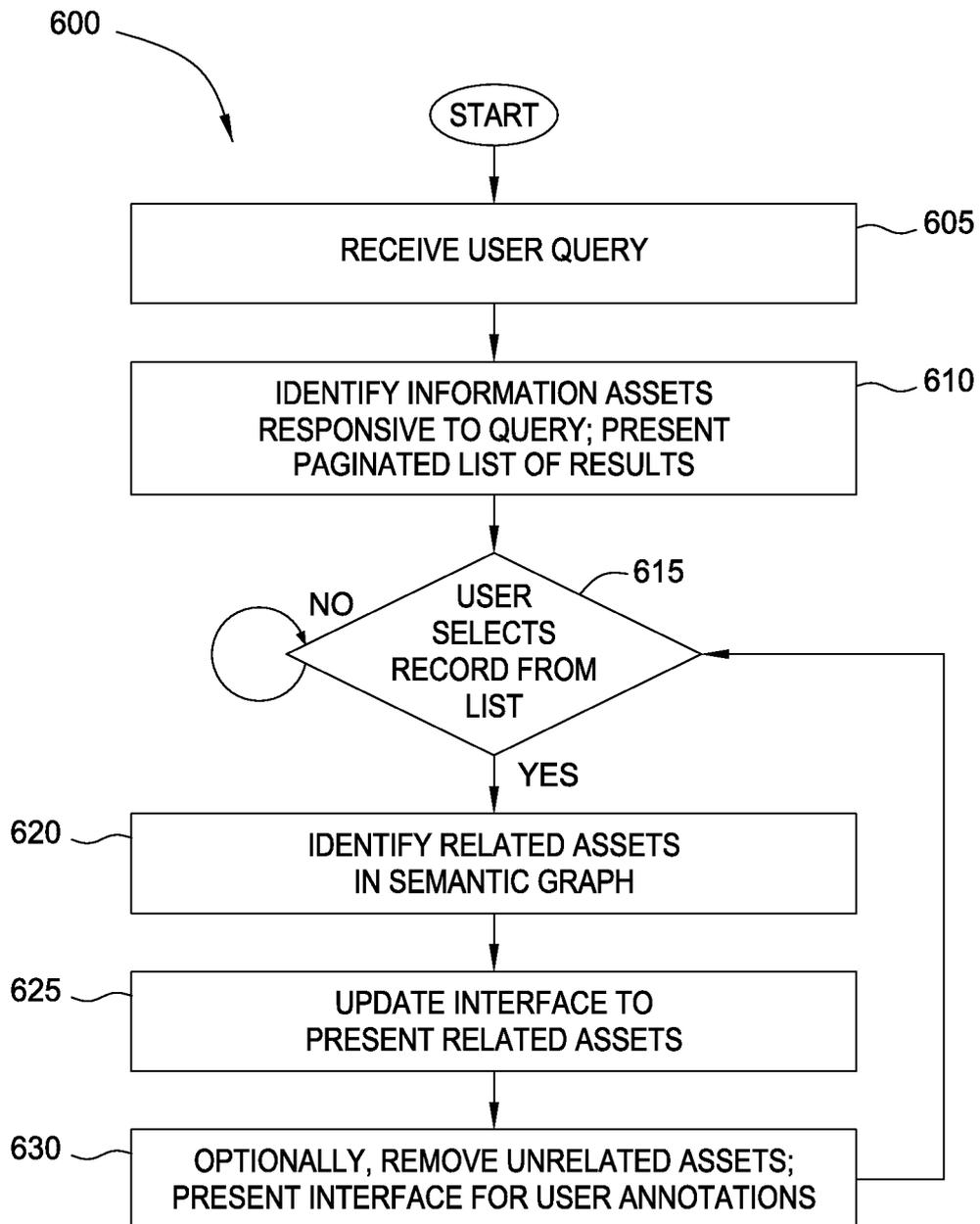


FIG. 6

8/8

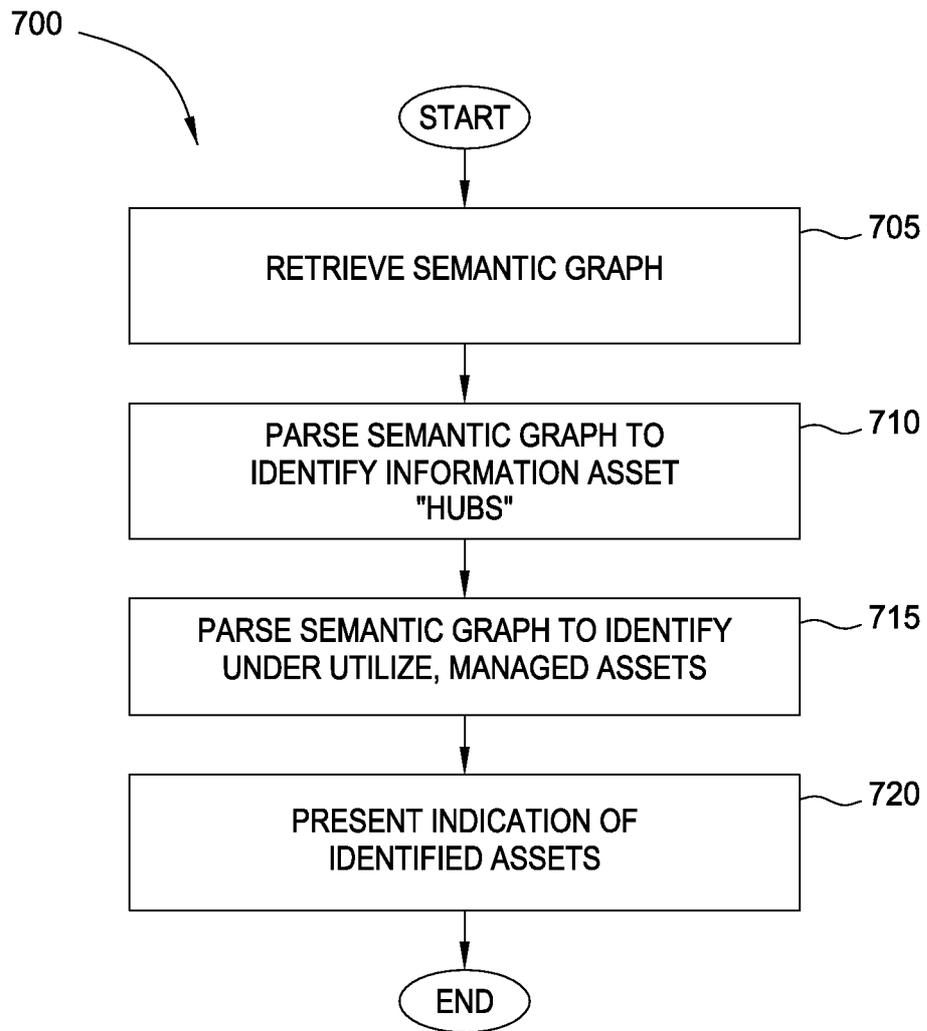


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2012/050821

<p>A. CLASSIFICATION OF SUBJECT MATTER IntCl. G06F1 7/30 (2006.01) i</p> <p>According to International Patent Classification (IPC) or to both national classification and IPC</p>														
<p>B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IntCl. G06F1 7/30</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2012 Registered utility model specifications of Japan 1996-2012 Published registered utility model applications of Japan 1994-2012</p> <p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)</p>														
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td align="center">Y</td> <td>Nobuto INOBUCHI, et al., The Desktop Search Using Relativity of Files Obtained from Access History, IPSJ SIG Technical Reports, 2006.07.12, Vol.2006 No. 77, p.141-146, 2006-NBS-140(1)</td> <td align="center">1-li 3</td> </tr> <tr> <td align="center">Y</td> <td>JP 2010-176198 A (NEC Corporation) 2010.08.12, [0002] - [0004], [0042] - [0043] (No Family)</td> <td align="center">1-li 3</td> </tr> <tr> <td align="center">Y</td> <td>Masashi YAMADA, et al., Extracting relations among files by user's history, IEICE Technical Report, 2000.02.22, Vol.99 No. 649, p.89-94, PRMU99_231-245</td> <td align="center">1-18</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	Y	Nobuto INOBUCHI, et al., The Desktop Search Using Relativity of Files Obtained from Access History, IPSJ SIG Technical Reports, 2006.07.12, Vol.2006 No. 77, p.141-146, 2006-NBS-140(1)	1-li 3	Y	JP 2010-176198 A (NEC Corporation) 2010.08.12, [0002] - [0004], [0042] - [0043] (No Family)	1-li 3	Y	Masashi YAMADA, et al., Extracting relations among files by user's history, IEICE Technical Report, 2000.02.22, Vol.99 No. 649, p.89-94, PRMU99_231-245	1-18
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
Y	Nobuto INOBUCHI, et al., The Desktop Search Using Relativity of Files Obtained from Access History, IPSJ SIG Technical Reports, 2006.07.12, Vol.2006 No. 77, p.141-146, 2006-NBS-140(1)	1-li 3												
Y	JP 2010-176198 A (NEC Corporation) 2010.08.12, [0002] - [0004], [0042] - [0043] (No Family)	1-li 3												
Y	Masashi YAMADA, et al., Extracting relations among files by user's history, IEICE Technical Report, 2000.02.22, Vol.99 No. 649, p.89-94, PRMU99_231-245	1-18												
<p><input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.</p>														
<p>* 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</p>														
<p>Date of the actual completion of the international search 14.05.2012</p>		<p>Date of mailing of the international search report 22.05.2012</p>												
<p>Name and mailing address of the ISA/JP Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan</p>		<p>Authorized officer Jun IWATA Telephone No. +81-3-358 1-1 101 Ext. 3599</p> <table border="1"> <tr> <td align="center">5M</td> <td align="center">4052</td> </tr> </table>	5M	4052										
5M	4052													