(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2012 (13.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/170050 A1

(51) International Patent Classification:
G06F 7/00 (2006.01)

(21) International Application Number:
PCT/US2011/046021

(22) International Filing Date:
29 July 2011 (29.07.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/153,420 4 June 2011 (04.06.2011) US

(71) Applicant *(for all designated States except US)*: **RECOMMIND, INC.** [US/US]; 650 California Street, 12th Floor, San Francisco, California 94108 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **TENNANT, Robert** [CA/US]; C/o, Recommind, INC., 650 California Street, 12th Floor, San Francisco, California 94108 (US).

(74) Agents: **SCHELLING, Myrna M.** et al.; Carr & Ferrell LLP, 120 Constitution Drive, Menlo Park, California 94025 (US).

(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, RO, RS, RU, 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, 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, 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))*

(54) **Title**: SYSTEMS AND METHODS FOR PUBLISHING DATASETS

(57) **Abstract**: Systems and methods for publishing datasets are provided herein. According to some embodiments, methods for publishing datasets may include receiving a request to publish a dataset to at least one of an internal environment located within a secured zone and an external environment located outside the secured zone, the request comprising at least one selection criteria, selecting the dataset based upon the at least one selection criteria, the dataset being selected from an index of collected datasets, and responsive to the request, publishing the dataset to at least one of the internal environment and the external environment.

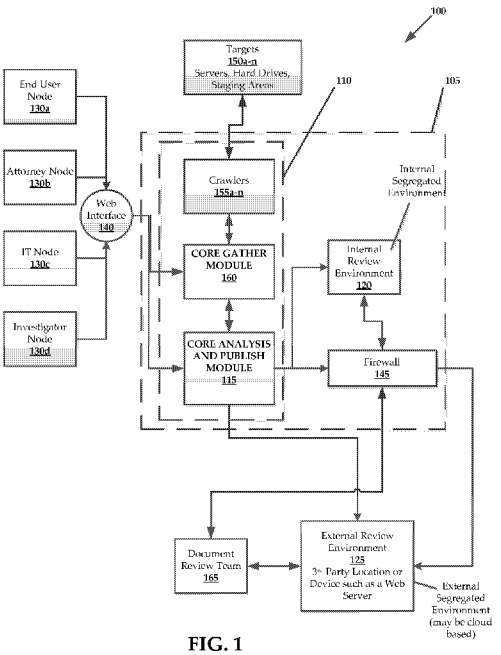

FIG. 1

WO 2012/170050 A1

SYSTEMS AND METHODS FOR PUBLISHING DATASETS

FIELD OF THE TECHNOLOGY

[0001] The present technology relates generally to systems and methods for information review and analysis, and more specifically, but not by way of limitation, to systems and methods for publishing datasets.

BACKGROUND

[0002] According to some embodiments, the systems and methods provided herein may be configured to index and/or collect and publish datasets. As background, electronic discovery, regulatory and investigative processes involve sensitive document review and analysis processes which often involve numerous time consuming and error prone steps associated with the gathering of relevant datasets in a legally defensible manner.

[0003] Dataset collection, review, and analysis often begins with the filing of a lawsuit, commencement of an investigation or other significant event that requires the review and analysis of certain types of datasets. These datasets may include many divers of similar data formats associated with user created files, electronic mail, audio, video, and the like. Datasets may be collected and analyzed from a variety of document sources such as email systems and repositories, file shares, storage media, third party web sites/services, and so forth. Because document review and analysis processes rely heavily on manual human action for certain steps, such as and the publishing of the collected and processed data to document review location, these

processes are often error prone and time consuming. For example, datasets may be mistakenly omitted (i.e., not collected or improperly reviewed), leading to costly errors.

[0004] In some instances, the collected datasets may be published to a virtual segregated document review environment or other secure document review construct that provides a secure location for the review and analysis of the collected datasets. These segregated document review environments provide a location (either secured or unsecured) where dataset review personnel may analyze and evaluate datasets that have been identified and selected for review.

SUMMARY OF THE TECHNOLOGY

[0005] According to some embodiments, the present technology may be directed to methods for publishing datasets. The methods may comprise: (a) receiving a request to publish a dataset to at least one of an internal environment located within a secured zone and an external environment located outside the secured zone, the request comprising at least one selection criteria; (b) selecting the dataset based upon the at least one selection criteria, the dataset being selected from an index of collected datasets; and (c) responsive to the request, publishing the dataset to at least one of the internal environment and the external environment.

[0006] According to additional exemplary embodiments, the present technology may be directed to an early case assessment system for publishing datasets. These systems may include: (a) a memory for storing executable instructions for publishing datasets; (b) a processor for executing the instructions stored in memory, the executable instructions comprising: (1) a core gather module that receives requests to publish a dataset to at least one of an internal environment located within a secured zone of the early case assessment system and an external environment located outside the secured zone, the request comprising at least one selection criteria; and (2) a core analysis module communicatively coupled to the core gather module that (i) selects the dataset based upon a request received by the core gather module, the dataset assembly module selecting the dataset from an index of collected datasets based upon the at least one selection criteria; and (ii) publishes the dataset to at least one of the internal environment and the external environment based upon a request received by the core gather module.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1 is a schematic diagram of an exemplary architecture of a system for publishing datasets. The system may be utilized to practice aspects of the present technology.

[0008] FIG. 2 is an exemplary flow diagram of a method for publishing datasets.

[0009] FIG. 3 is a block diagram of an exemplary computing system that may be utilized to practice aspects of the present disclosure.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0010] While this technology is susceptible of embodiment in many different forms, there is shown in the figures and will herein be described in detail several specific embodiments with the understanding that the present disclosure is to be considered as an exemplification of the principles of the technology and is not intended to limit the technology to the embodiments illustrated.

[0011] Generally speaking, the systems provided herein may include electronic dataset review systems that identify, process, and publish datasets (e.g., documents, electronic mail, and so forth) from a variety of data sources such as an enterprise email server, document repositories, and the like. All datasets identified and/or collected comprise a corpus of reviewable datasets that are available for analysis and/or review. According to some embodiments, each dataset may be crawled and processed by the system according to guidelines established for the system by one or more dataset review administrators.

[0012] The systems and methods provided herein may automatically and continuously review datasets as they become available. It will be understood that the term "review" may be understood to include additional functions such as analyze, organize, enrich, prioritize, code, batch, and so forth. Datasets may be evaluated for keywords, concepts or other metadata that define the content or subject matter of the dataset and potential relevance of individual pieces of data. For example, the system may determine keywords in the body of a particular electronic mail communication. The subject matter of the email communication may be inferred from the occurrence of particular words or phrases or combinations thereof within the dataset. In some embodiments, the system may utilize any one of a number of algorithms to process and

evaluate the subject matter and content of the datasets and their associated metadata to create an index of collected datasets according to the parameters set by the user. For example, email communications may be parsed for terms, dates, subject matter, email addresses, and the like. These terms are then associated with the email communication in the index such that searching the index for a term that is included in the email will yield the email as a search result.

[0013] Upon the occurrence of a triggering event (e.g., commencement of a lawsuit, investigation or other similar event), system users may search the index of datasets for relevant documents according to criteria. Criteria may include keyword searching, filtering according to parameters, and so forth. Once relevant data have been determined from the index and/or original location, the system may copy the relevant datasets from index and publish the same to a document review location.

[0014] According to some embodiments, the document review location may include a virtual segregated document review environment, a secure document review location within the system, and/or a third party document review location. For example, the relevant documents may be communicated to the computing system of the user, the computing system of the user's legal counsel, to regulatory bodies, or the relevant documents may be retained for review and analysis within the system itself.

[0015] It is noteworthy to mention that the more frequently a dataset is communicated between computing systems, the more likely that the dataset may be subject to unwanted exposure to third parties. Therefore, sensitive datasets may be retained within a secure zone of the system, or secure communications protocols may be established for accessing the datasets if they are published outside the secure zone of the system.

[0016] FIG. 1 illustrates an exemplary architecture **100** that may include an early case assessment system, hereinafter "system **105**" that, in turn, includes a data collection system **110**, a core analysis and publish module **115**, an internal review environment **120**, and an external review environment **125**, can be utilized as a part of the early case assessment system, system **105**, or as an external review environment to which data is published from the early case assessment system **105**. It is noteworthy that the system **105** may include additional modules, engines, or components, and still fall within the scope of the present technology. As used herein, the term "module" may also refer to any of an application-specific integrated circuit ("ASIC"), an electronic circuit, a processor (shared, dedicated, or group) that executes one or more software or firmware programs, a combinational logic circuit, and/or other suitable components that provide the described functionality. In other embodiments, individual modules of the system **105** may include separately configured web servers.

[0017] Generally speaking, the system **105** may communicatively couple with any one of a plurality of client devices such as an end user node **130a**, an attorney node **130b**, an IT node **130c**, and an investigator node **130d**. While these nodes represent descriptive or exemplary end user client devices, one of ordinary skill in the art will appreciate that many other types of client device nodes may likewise be utilized in accordance with the present technology. It is further noteworthy to mention that each of the nodes **130a-d** may include an end user computing system such as computing system **300**, which will be described in greater detail with reference to FIG. 3.

[0018] Each of the nodes **130a-d** may communicatively couple with the system **105** via any of a private and/or public communications network such as the Internet.

[0019] According to some embodiments, the system **105** may include a cloud based computing environment that collects, processes, analyzes, and publishes datasets. In

general, a cloud-based computing environment is a resource that typically combines the computational power of a large grouping of processors and/or that combines the storage capacity of a large grouping of computer memories or storage devices. For example, systems that provide a cloud resource may be utilized exclusively by their owners, such as Google™ or Yahoo! ™; or such systems may be accessible to outside users who deploy applications within the computing infrastructure to obtain the benefit of large computational or storage resources.

[0020] The cloud may be formed, for example, by a network of web servers s with each server (or at least a plurality thereof) providing processor and/or storage resources. These servers may manage workloads provided by multiple users (e.g., cloud resource customers or other users). Typically, each user places workload demands upon the cloud that vary in real-time, sometimes dramatically. The nature and extent of these variations typically depend on the type of business associated with the user.

[0021] The system **105** may generate a variety of graphical user interfaces that allow end users to interact with the system **105**. For example, the system may generate a web interface **140** that allows end users to enter search queries for searching indexes, establish dataset collection criteria, or automatically obtain and publish datasets to at least one of an internal document review location, such as an internal review environment **120**, and an external review environment **125**. It is noteworthy to mention that the external review environment may be located outside a firewall **145** of the system **105**. As such, accessing datasets published to the internal review environment **120** may require the use of security credentials (e.g., secure communication protocols) to gain access through the firewall **145**.

[0022] It will be understood that components or modules behind the firewall **145** of the system **105** may be referred to as being located within a "secure zone" of the system

105. For example, the internal review environment **120** may be understood as being located within the secure zone because it is located behind the firewall **145**, or other suitable security devices, systems, or processes. In contrast, the external review environment **125** may be understood to be outside the secure zone of the system **105** because the external review environment **125** is located outside the firewall **145**.

[0023] The web interface **140** may be accessed by a web browser application of any of the nodes **130a-d**, or may be accessed by an application resident on a mobile device such as a smartphone.

[0024] Data collection software or modules that may communicatively couple with the data collection module **110** may be installed on, and reside on the computing systems of the client (e.g., the entity for which datasets are to be collected), such as targets **150a-n**. In other embodiments, the data collection module **110** may utilize one or more crawlers **155a-n** that may communicatively couple with the targets **150a-n** to receive datasets therefrom. The communication of datasets from targets **150a-n** to the data collection module **110** may occur automatically and continuously over time (e.g., synchronously), or may occur in batches (e.g., asynchronously). The term "target" may be understood to include any data source such as an email repository, storage media, and so forth.

[0025] As datasets are located by the crawlers **155a-n**, the core analysis and publish module **115** of the data collection module **110** may create an index of the received datasets and associated meta-data and then apply one or more algorithms to process and evaluate the subject matter and content of the datasets and their associated metadata and to create an index of collected datasets, as described above. For example, these evaluative algorithms may determine information such as keywords, dates, times,

and other pertinent information included in the dataset. The core analysis and publish module **115** may also analyze the metadata of the dataset.

[0026] The core gather module **160** may establish and maintain individual indexes for each of targets **150a-n**, or may utilize one index for each client that includes information for each target **150a-n** of the client.

[0027] Additionally, the core gather module **160** may store the collected datasets in a storage medium (not shown) communicatively coupled with the system **105**.

[0028] Upon the occurrence of a triggering event, a user may interact with the system **105** via their client device, such as user node **130a**. That is, users may access the system **105** via the web interface **140** utilizing their user node **130a**. Users may review, analyze and code relevant datasets by first searching or analyzing filtered datasets from an index or a combination of indexes.

[0029] For example, the end user may desire to locate all datasets that mention a particular keyword or group of keywords or relate to a particular concept. The core gather module **160** of the system **105** may compare the query to the index to locate and obtain relevant datasets that correspond to received queries.

[0030] In some embodiments, the core analysis and publish module **115** may make a copy of the index or indexes and select datasets before obtaining the datasets for publishing. Additionally, the core gather module **160** may lock each selected dataset to prevent tampering or modification to the datasets. Moreover, the core gather module **160** may utilize any other necessary protective or security related procedures to ensure that the datasets are obtained and maintained in a legally defensible manner.

[0031] After selecting the desired datasets, the core analysis and publish module **115** may prepare the datasets for export or publishing to a document review or analysis location by encapsulating the datasets into files and combining the datasets into

batches. In other embodiments, datasets may be published in their raw or native format.

[0032] When the core analysis and publish module **115** has prepared the datasets for publishing, the core analysis and publish module **115** may publish the datasets to at least one of the internal review environment **120** and the external review environment **125**.

[0033] According to some embodiments, the core analysis and publish module **115** may publish the datasets to a document review location via any one of a number of methods. For example, the core analysis and publish module **115** may publish the datasets by uploading the datasets via a file transfer protocol ("FTP"), or other computer networking protocol to one of the aforementioned locations. In other embodiments, the core analysis and publish module **115** may publish the datasets to a third party location, such as the external review environment **125**. In additional embodiments, the core analysis and publish module **115** may stream the datasets to the desired review location, either in batches or as individual datasets.

[0034] As stated previously, the core analysis and publish module **115** may publish the datasets to an external review environment **125** that exists beyond or outside of the firewall **145** of the system **105**. Exemplary external nodes may include, but are not limited to, a web server or storage media associated with a user's web server or a third party web-based destination per the user's selection, and so forth.

[0035] It is noteworthy to mention that, because the system **105** indexes datasets on a periodic or continual basis, requests to obtain and publish datasets may result in the system **105** automatically obtaining datasets from the corpus of available datasets without need for further evaluation of the datasets. That is, the system **105** may pre-analyze and indexes the datasets prior to the occurrence of a triggering event such that

upon the occurrence of a triggering event, the datasets are easily and seamlessly obtained (should it meet certain defined parameters or criteria) and published to the appropriate document review location(s).

[0036] Assuming datasets have been published to an external review environment **125**, a document review team **165** may access the datasets within the environment or may transfer the datasets from the environment for use on a local computing system. It will be understood that the document review team **165** may include a plurality of individual computing systems. The document review team **165** may access the external review environment **125** utilizing secure communications protocols, although in some embodiments, the document review team **165** may access the external review environment **125** without the requirement of a secure communications path. This scenario is more common when the datasets do not include sensitive information such as credit card data, social security numbers, or health record data.

[0037] Alternatively, users may access datasets via the internal review environment **120** from a web based interface, such as web based interface **140**, or by way of an additional review application (not shown) that may reside within a cloud based computing environment.

[0038] It will be understood that according to some embodiments, the external review environment may include a cloud based computing environment that is configured to store selected and published datasets. The datasets may be stored in the cloud computing environment in virtual storage media or physical storage media that are distributed across a plurality of web servers.

[0039] FIG. 2 illustrates a flowchart of an exemplary method **200** for publishing a dataset. The method may begin with the step **205** of crawling (or otherwise parsing, evaluating, analyzing, and so forth) the datasets from one or more data sources or

targets. As datasets are crawled, the method may include a step **210** of indexing the datasets. Indexing of datasets may include applying statistical or analytical algorithms or other mechanisms for determining the subject matter and/or content of the crawled datasets. Keywords, dates, times, and other descriptive information relative to the datasets may be stored in the index and associated with the datasets from which they were extracted.

[0040] Upon the occurrence of a triggering event (such as the commencement of a lawsuit), the method may include a step **215** of the system receiving a request to publish a dataset to at least one of an internal environment located within a secured zone and an external environment located outside the secured zone of the system.

[0041] Next, the method may include the step **220** of selecting or obtaining the dataset based upon the at least one selection criteria. It is noteworthy to mention that the dataset may be selected from the index of collected datasets that is maintained by the system.

[0042] After the selection of datasets, the method may include a step **225** of publishing the dataset to at least one of the internal environment and the external environment. The method **200** may then return to step **205** in anticipation of an additional triggering event.

[0043] It will be understood that in some embodiments, the method may include additional or fewer steps. For example, other exemplary methods may include only the steps of receiving requests, selecting datasets, and publishing the selected datasets.

[0044] It is noteworthy to mention that in alternative embodiments, the system may collect and/or store datasets from one or more targets before the datasets are indexed by the system.

[0045] FIG. 3 illustrates an exemplary computing system 300 that may be used to implement an embodiment of the present technology. The computing system 300 of FIG. 3 includes one or more processors 310 and main memory 320. Main memory 320 stores, in part, instructions and data for execution by processor 310. Main memory 320 can store the executable code when the system 300 is in operation. The system 300 of FIG. 3 may further include a mass storage device 330, portable storage devices 340, output devices 350, user input devices 360, a graphics display 370, and other peripheral devices 380.

[0046] The components shown in FIG. 3 are depicted as being connected via a single bus 390. The components may be connected through one or more data transport means. Processor unit 310 and main memory 320 may be connected via a local microprocessor bus, and the mass storage device 330, peripheral device(s) 380, portable storage device(s) 340, and graphics display system 370 may be connected via one or more input/output (I/O) buses.

[0047] Mass storage device 330, which may be implemented with a magnetic disk drive or an optical disk drive, is a non-volatile storage device for storing data and instructions for use by processor unit 310. Mass storage device 330 can store the system software for implementing embodiments of the present technology for purposes of loading that software into main memory 320.

[0048] Portable storage device 340 operates in conjunction with a portable non-volatile storage media, such as a floppy disk, compact disk or digital video disc, to input and output data and code to and from the computer system 300 of FIG. 3. The system software for implementing embodiments of the present technology may be stored on such a portable media and input to the computer system 300 via the portable storage device 340.

[0049] User input devices **360** provide a portion of a user interface. User input devices **360** may include an alphanumeric keypad, such as a keyboard, for inputting alphanumeric and other information, or a pointing device, such as a mouse, a trackball, stylus, or cursor direction keys. Additionally, the system **300** as shown in FIG. 3 includes output devices **350**. Suitable output devices include speakers, printers, network interfaces, and monitors.

[0050] Graphics display system **370** may include a liquid crystal display (LCD) or other suitable display device. Graphics display system **370** receives textual and graphical information, and processes the information for output to the display device.

[0051] Peripheral devices **380** may include any type of computer support device to add additional functionality to the computer system. Peripheral device(s) **380** may include a modem or a router.

[0052] The components contained in the computer system **300** of FIG. 3 are those typically found in computer systems that may be suitable for use with embodiments of the present technology and are intended to represent a broad category of such computer components that are well known in the art. Thus, the computer system **300** of FIG. 3 can be a personal computer, hand held computing system, telephone, mobile computing system, workstation, server, minicomputer, mainframe computer, or any other computing system. The computer can also include different bus configurations, networked platforms, multi-processor platforms, etc. Various operating systems can be used including UNIX, Linux, Windows, Macintosh OS, Palm OS, and other suitable operating systems.

[0053] Some of the above-described functions may be composed of instructions that are stored on storage media (e.g., computer-readable media). The instructions may be retrieved and executed by the processor. Some examples of storage media are memory

devices, tapes, disks, and the like. The instructions are operational when executed by the processor to direct the processor to operate in accord with the technology. Those skilled in the art are familiar with instructions, processor(s), and storage media.

[0054] It is noteworthy that any hardware platform suitable for performing the processing described herein is suitable for use with the technology. The terms "computer-readable storage media" and "computer-readable storage media" as used herein refer to any media or media that participate in providing instructions to a CPU for execution. Such media can take many forms, including, but not limited to, non-volatile media, volatile media and transmission media. Non-volatile media include, for example, optical or magnetic disks, such as a fixed disk. Volatile media include dynamic memory, such as system RAM. Transmission media include coaxial cables, copper wire and fiber optics, among others, including the wires that comprise one embodiment of a bus. Transmission media can also take the form of acoustic or light waves, such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, a hard disk, magnetic tape, any other magnetic media, a CD-ROM disk, digital video disk (DVD), any other optical media, any other physical media with patterns of marks or holes, a RAM, a PROM, an EPROM, an EEPROM, a FLASHEPROM, any other memory chip or data exchange adapter, a carrier wave, or any other media from which a computer can read.

[0055] Various forms of computer-readable media may be involved in carrying one or more sequences of one or more instructions to a CPU for execution. A bus carries the data to system RAM, from which a CPU retrieves and executes the instructions. The instructions received by system RAM can optionally be stored on a fixed disk either before or after execution by a CPU.

[0056] The above description is illustrative and not restrictive. Many variations of the technology will become apparent to those of skill in the art upon review of this disclosure. The scope of the technology should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

CLAIMS

What is claimed is:

1. A method for publishing datasets, comprising:

receiving a request to publish a dataset to at least one of an internal environment located within a secured zone and an external environment located outside the secured zone, the request comprising at least one selection criteria;

selecting the dataset based upon the at least one selection criteria, the dataset being selected from an index of collected datasets; and

responsive to the request, publishing the dataset to at least one of the internal environment and the external environment.

2. The method according to claim 1, further comprising:

collecting datasets from a plurality of data sources; and

indexing the collected datasets to create the index of collected datasets; and

storing the collected datasets in a storage medium.

3. The method according to claim 1, further comprising, responsive to the request, creating a copy of the index of collected datasets and selecting the dataset from the copy of the index of collected datasets.

4. The method according to claim 1, wherein publishing includes at least one of:

encapsulating the dataset into one or more files and transmitting the one or more files to at least one of the internal environment and the external environment;

grouping the dataset into one or more batches and transmitting the one or more files to at least one of the internal environment and the external environment; and

streaming the dataset to at least one of the internal environment and the external environment.

5. An early case assessment system for publishing datasets, comprising:

a memory for storing executable instructions for publishing datasets;

a processor for executing the instructions stored in memory, the executable instructions comprising:

a core gather module that receives requests to publish a dataset to at least one of an internal environment located within a secured zone of the early case assessment system and an external environment located outside the secured zone, the request comprising at least one selection criteria; and

a core analysis module communicatively coupled to the core gather module that:

selects the dataset based upon a request received by the core gather module, the dataset assembly module selecting the dataset being from an index of collected datasets based upon the at least one selection criteria; and

publishes the dataset to at least one of the internal environment and the external environment based upon a request received by the core gather module.

6. The early case assessment system according to claim 5, wherein the core gather module collects datasets from a plurality of data sources and the core analysis module indexes the collected datasets to create the index of collected datasets and stores the collected datasets in a storage medium.

7. The early case assessment system according to claim 5, wherein the core analysis module creates a copy of the index of collected datasets and selects the dataset from the copy of the index of collected datasets based upon receipt of a request to publish a dataset by the core gather module.

8. The early case assessment system according to claim 1, wherein the external environment that is located outside the secured zone includes a cloud computing system communicatively couplable to the publishing module of the system.

WO 2012/170050 PCT/US2011/046021

SYSTEMS AND METHODS FOR PUBLISHING DATASETS

Inventor: Robert Tennant

Attorney Docket: PA5610PCT

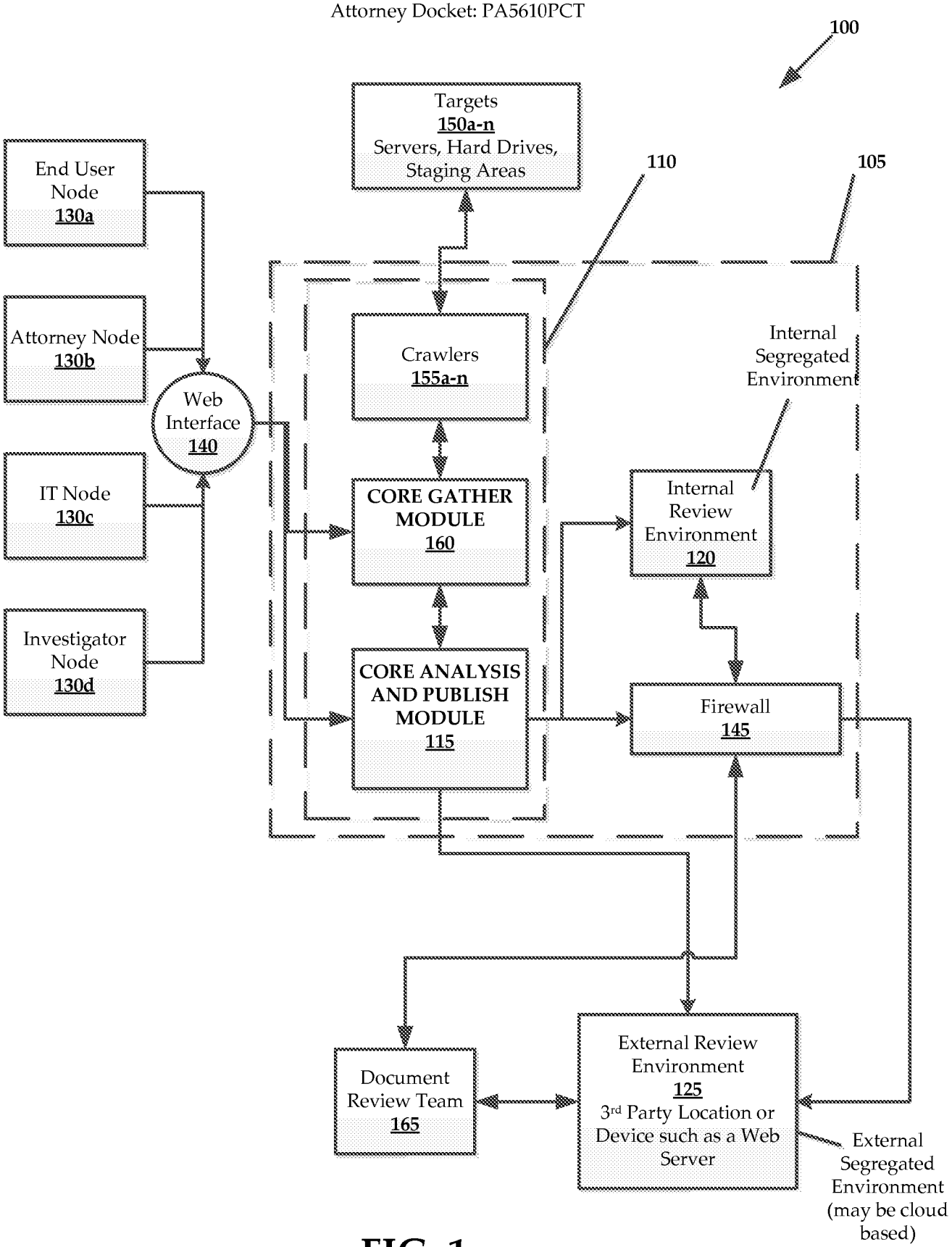

FIG. 1

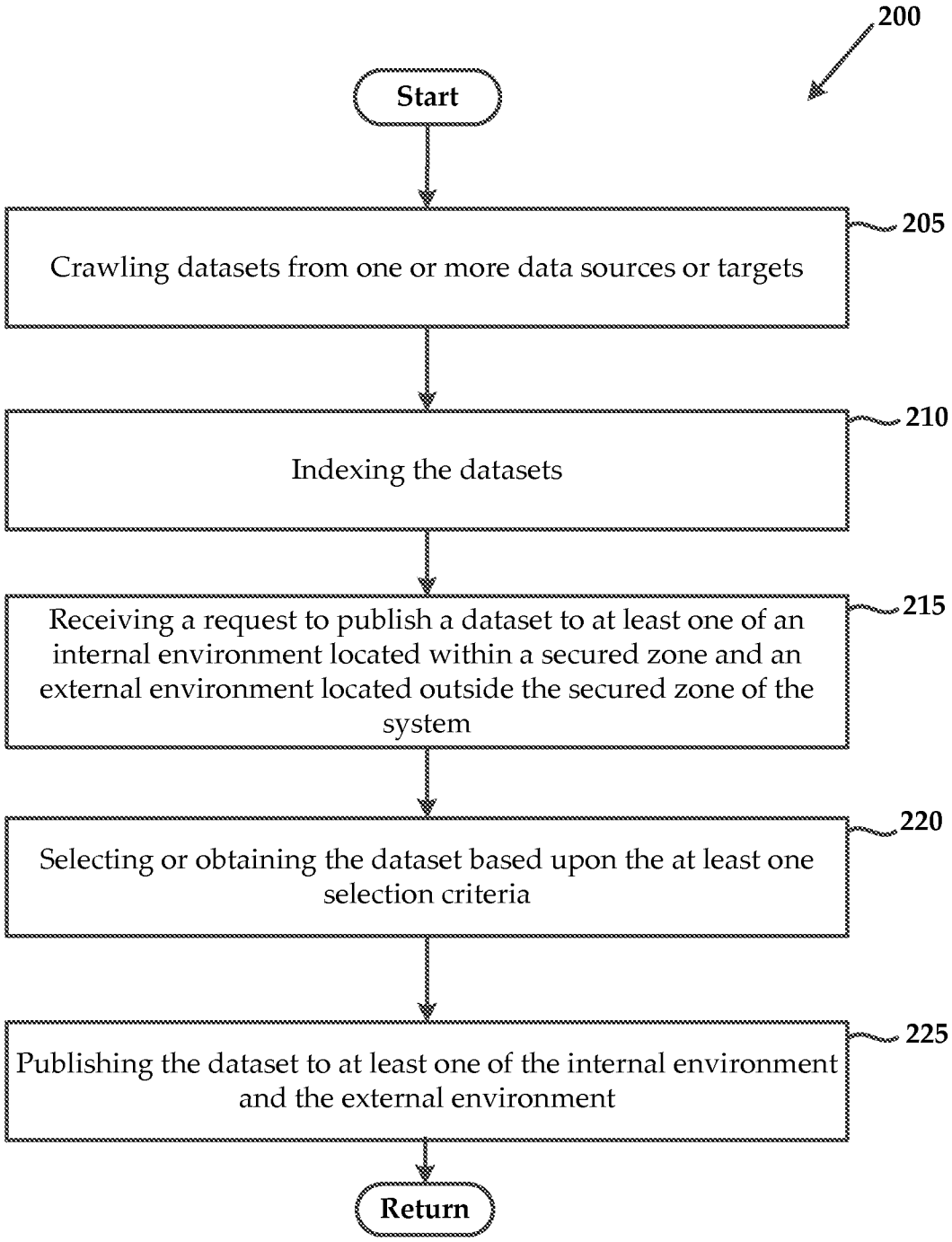
200
Start
Crawling datasets from one or more data sources or targets
205
Indexing the datasets
210
Receiving a request to publish a dataset to at least one of an internal environment located within a secured zone and an external environment located outside the secured zone of the system
215
Selecting or obtaining the dataset based upon the at least one selection criteria
220
Publishing the dataset to at least one of the internal environment and the external environment
225
Return

FIG. 2

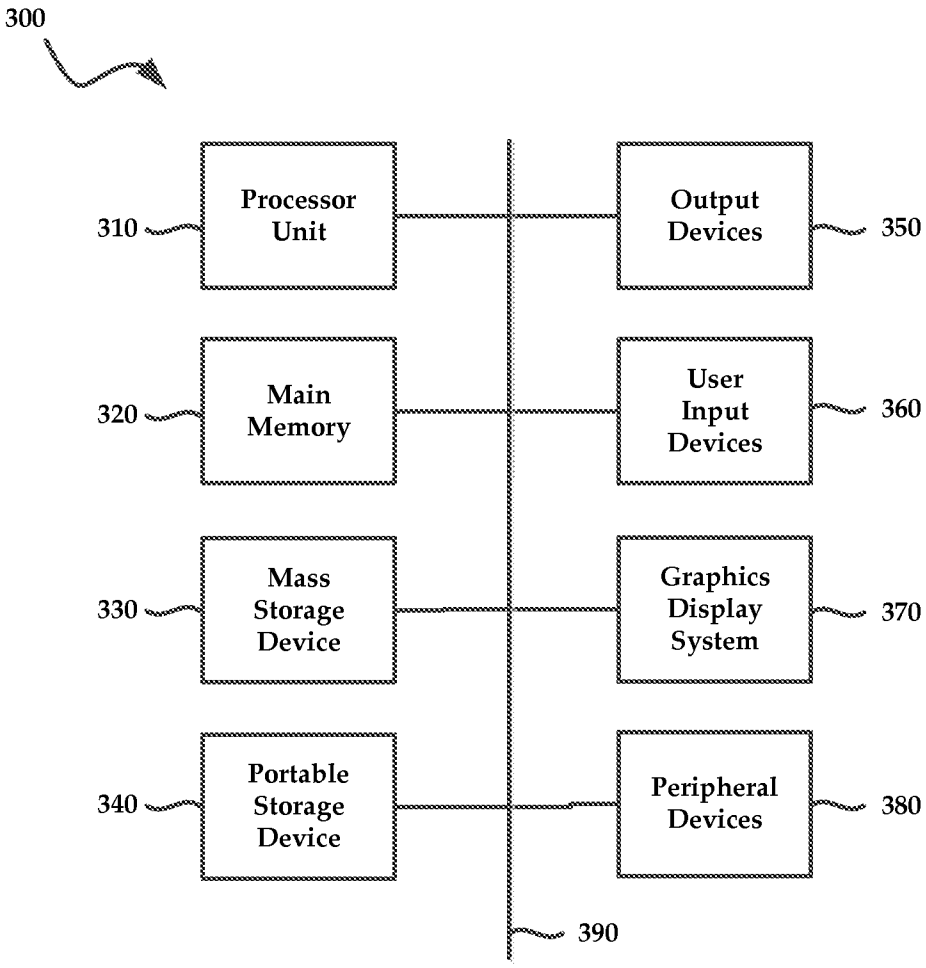
300
Processor Unit
310
Main Memory
320
Mass Storage Device
330
Portable Storage Device
340
Output Devices
350
User Input Devices
360
Graphics Display System
370
Peripheral Devices
380
390

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/46021

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 7/00 (2011.01)
USPC - 707/603
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)
USPC: 707/603

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 706/20, 933; 707/603, 705, 710 (keyword limited - see terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB, USPT, USOC, EPAB, JPAB); GOOGLE; Google Scholar
Search Terms: publish, data, dataset, request, collect, selection, internal, external, secure, zone, index, criteria, parameter, source, group

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/0205153 A1 (Lamm) 12 August 2010 (12.08.2010), entire document, especially; abstract, para. [0005]-[0011], [0038], [0061]	1 - 8
Y	US 2010/0306524 A1 (Runkis et al.) 02 December 2010 (02.12.2010), entire document, especially; abstract, para. [0007]-[0009], [0025]-[0027], [0041], [0046], [0065]	1 - 8
A	US 2007/0073695 A1 (Conlan et al.) 29 March 2007 (29.03.2007), entire document	1 - 8

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

* 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	Date of mailing of the international search report
28 November 2011 (28.11.2011)	16 DEC 2011
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774