

ORIGINAL

ABSTRACT

“ENTERPRISE LEVEL DATA MANAGEMENT”

A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise, the system including background data characterization functionality characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, the at least one access metric being selected from data access permissions and actual data access history and near real time data matching functionality selecting the data of interest by considering only data elements which have the at least one content characteristic thereof and the at least one access metric thereof from among the multiplicity of data elements.

Refer to Figure 1

CLAIMS

1. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise, the system comprising:
 - background data characterization functionality characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, said at least one access metric being selected from data access permissions and actual data access history; and
 - near real time data matching functionality selecting the data of interest by considering only data elements which have said at least one content characteristic thereof and said at least one access metric thereof from among said multiplicity of data elements.

2. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 1 and wherein said near real time data matching functionality comprises background field of search definition and searching functionality operative to define a field of search in accordance with said at least one access metric and to search within said field of search based on said at least one content characteristic.

3. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 1 and wherein said near real time data matching functionality comprises background field of search definition and searching functionality operative to define a field of search in accordance with said at least one access metric multiple times and to search within said field of search multiple times, wherein said at least one access metric is different at least some of said multiple times.

4. A system for identifying data of interest from among a multiplicity of data elements according to claim 1 and wherein said at least one access metric is a dynamic metric which changes over time during operation of the enterprise.

5. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 2 and also comprising:

automatic field of search redefinition and search functionality operative to redefine said field of search in accordance with said at least one access metric multiple times and search within said field of search multiple times, wherein said at least one access metric is different at least some of said multiple times.

6. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 5 and wherein said automatic field of search redefinition and search functionality is operative to search only within those portions of said field of search that have been modified or added as the result of redefining said field of search in accordance with changes in said at least one access metric.

7. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 5 and also comprising data element status monitoring functionality, noting the current status of data elements that have been modified, added or removed in accordance with changes in said at least one access metric.

8. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 2 and wherein said searching is prioritized at least in accordance with at least one access metric related prioritization characteristic.

9. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 8 and wherein results of said searching are ordered at least in accordance with at least one access metric related prioritization characteristic.

10. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 1 and wherein said near real time data matching functionality includes:

searching functionality for searching for data elements which have said at least one content characteristic thereof; and

identification functionality operative separately from said searching for data elements which have said at least one content characteristic thereof, identifying data elements from among said multiplicity of data elements in accordance with said at least one access metric; and

combining functionality, combining results of said searching and said identifying.

11. A system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 10 and wherein said searching and said identifying are performed by separate entities.

12. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise, the method comprising:

characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, said at least one access metric being selected from data access permissions and actual data access history; and

selecting the data of interest by considering only data elements which have said at least one content characteristic thereof and said at least one access metric thereof from among said multiplicity of data elements.

13. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 12 and wherein said considering comprises:

defining a field of search in accordance with said at least one access metric and searching within said field of search based on said at least one content characteristic.

14. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 12 and wherein said considering comprises:

defining a field of search in accordance with said at least one access metric multiple times and searching within said field of search multiple times, wherein said at least one access metric is different at least some of said multiple times.

15. A method for identifying data of interest from among a multiplicity of data elements according to claim 12 and wherein said at least one access metric is a dynamic metric which changes over time during operation of the enterprise.

16. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 13 and also comprising:

automatically redefining said field of search in accordance with said at least one access metric multiple times and searching within said field of search multiple times, wherein said at least one access metric is different at least some of said multiple times.

17. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 16 and wherein said searching comprises searching only within those portions of said field of search that have been modified or added as the result of redefining said field of search in accordance with changes in said at least one access metric.

18. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 16 and also comprising noting the current status of data elements that have been modified, added or removed in accordance with changes in said at least one access metric.

19. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 12 and

wherein said searching is prioritized at least in accordance with at least one access metric related prioritization characteristic.

20. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 19 and wherein results of said searching are ordered at least in accordance with at least one access metric related prioritization characteristic.

21. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 12 and wherein said selecting the data of interest by considering only data elements which have said at least one content characteristic thereof and said at least one access metric thereof from among said multiplicity of data elements includes:


searching for data elements which have said at least one content characteristic thereof; and

separately from said searching for data elements which have said at least one content characteristic thereof, identifying data elements from among said multiplicity of data elements in accordance with said at least one access metric; and

combining results of said searching and said identifying.

22. A method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise according to claim 21 and wherein said searching and said identifying are performed by separate entities.

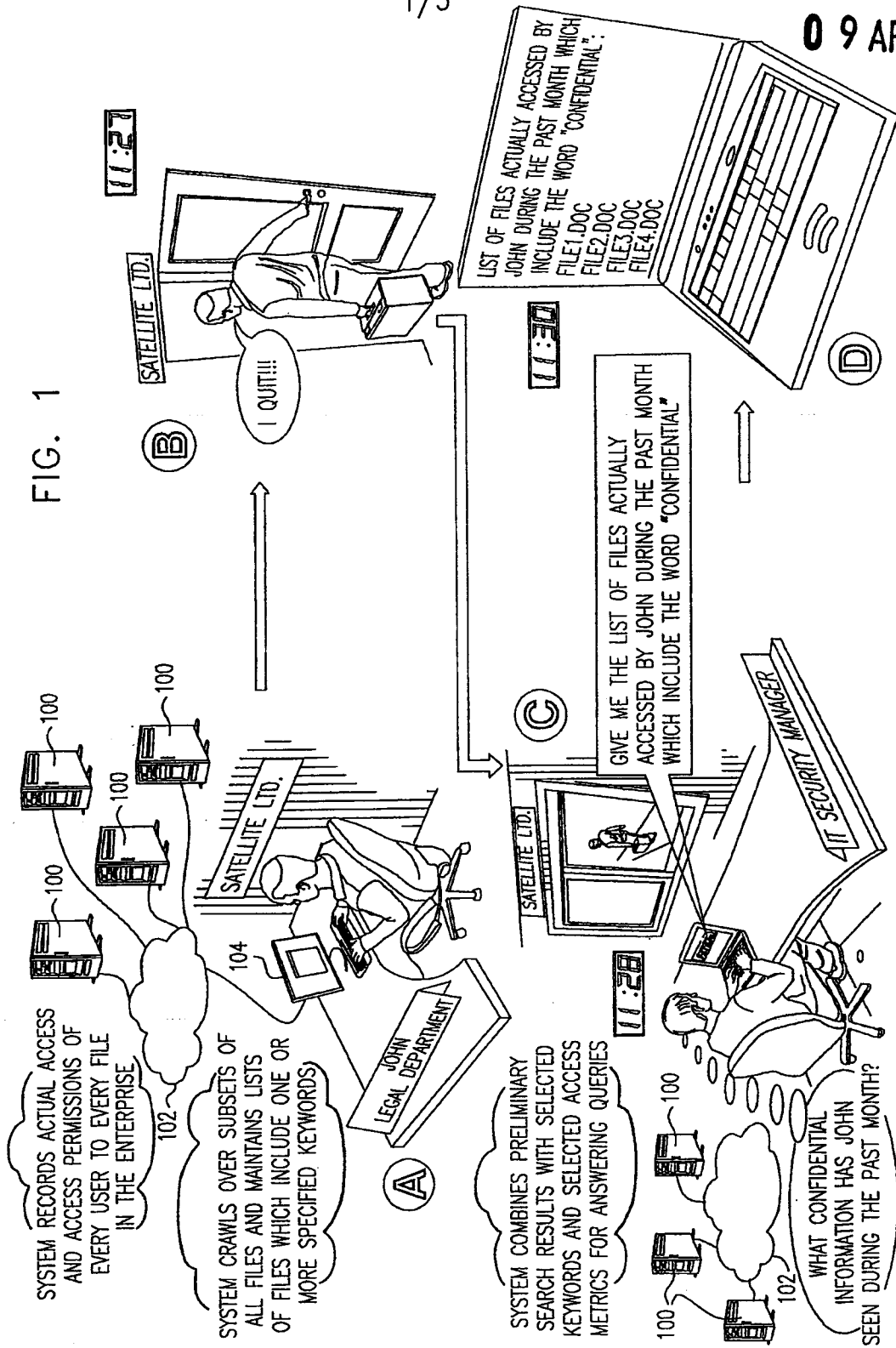
Dated this 9th day of April 2012


of Anand and Anand Advocates
Agents for the Applicants

1/5

09 APR 2012

FIG. 1



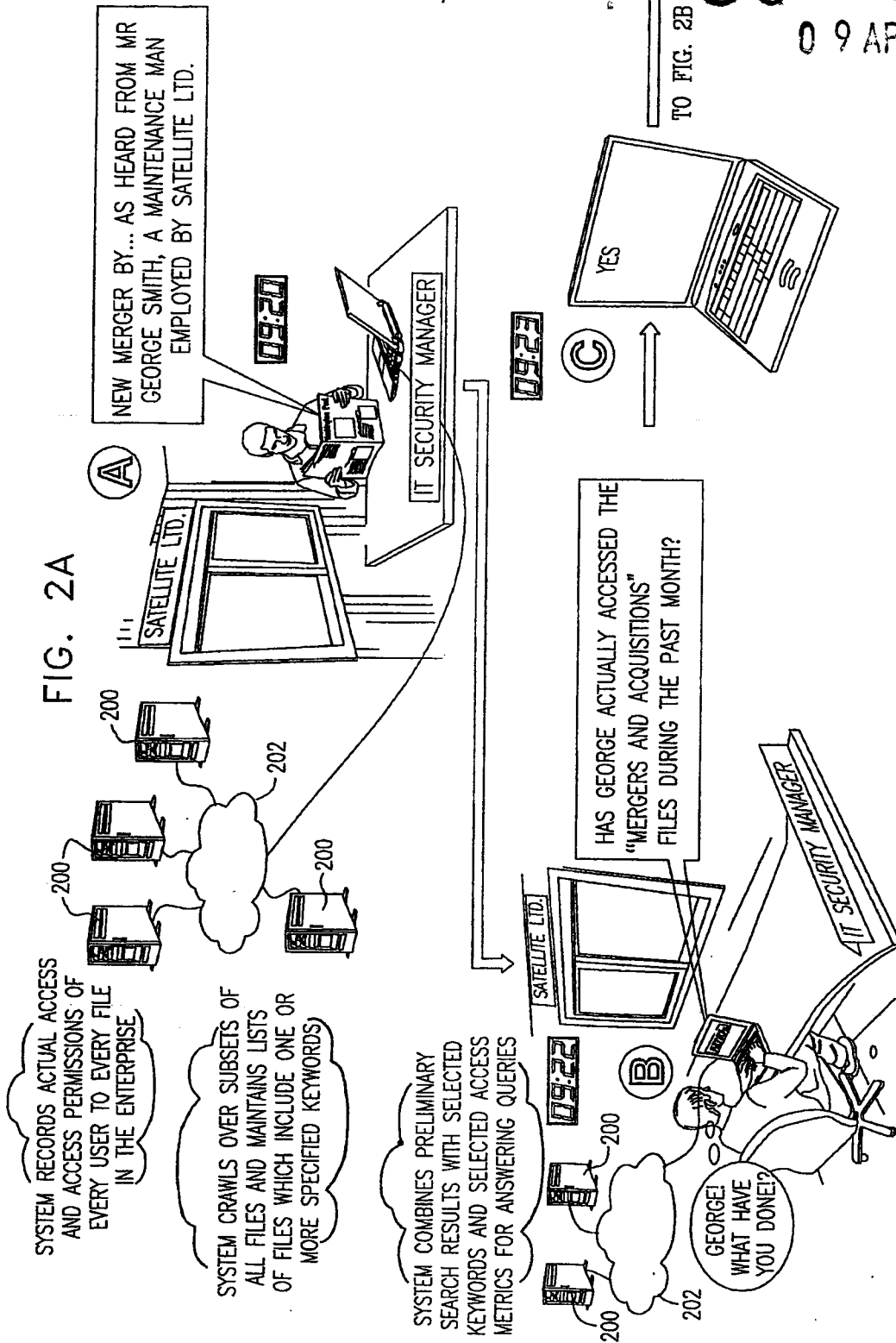
Shanker

ORIGINAL

3035 12

09 APR 2012

2/5



Shanker

Archana Shanker
of Anand and Anand Advocates
Agent for the Applicant

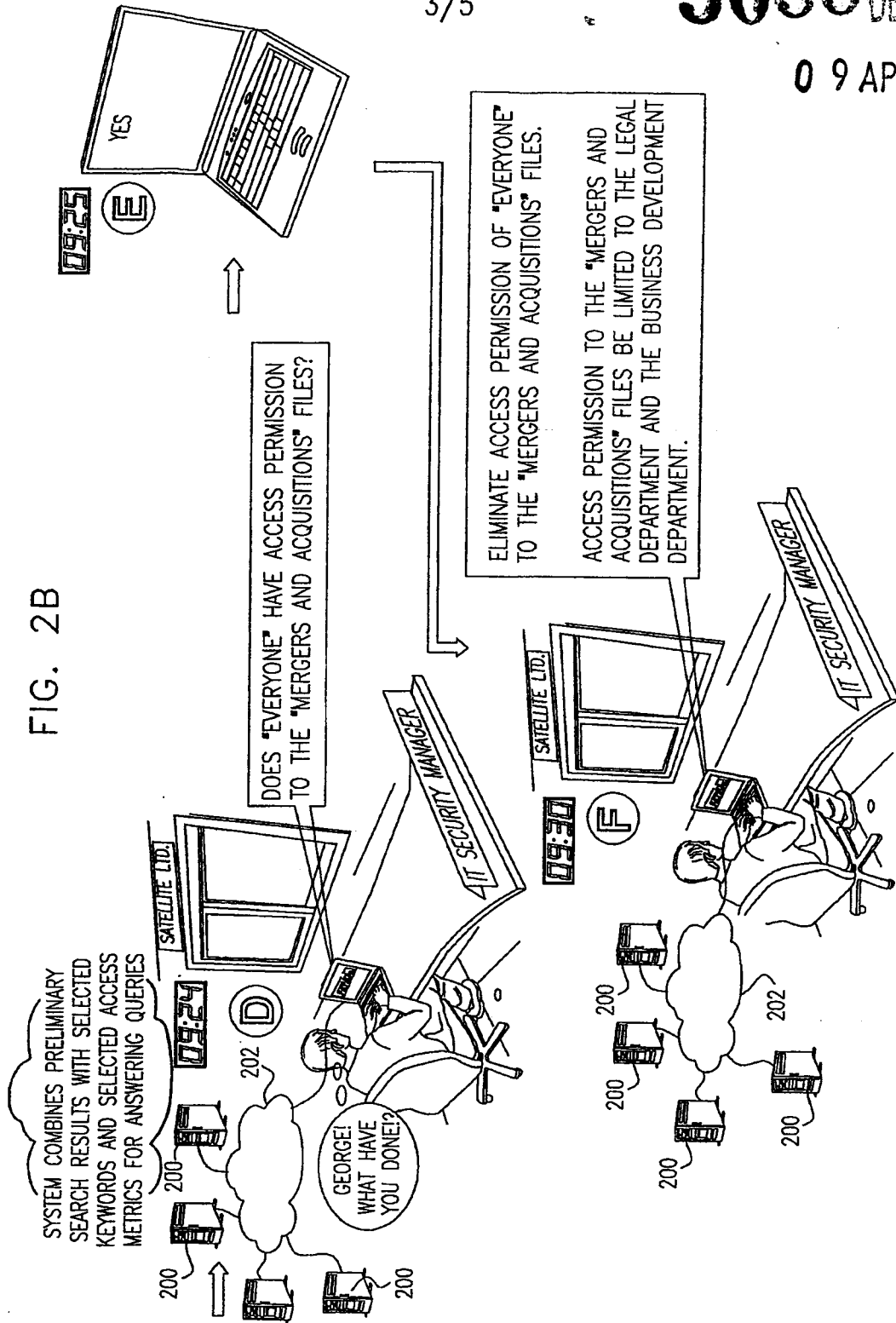
ORIGINAL

3035 DELWP 12

09 APR 2012

3/5

FIG. 2B



Shanker

Archana Shanker
of Anand and Anand Advocates
Agent for the Applicant

3035 DELNP 12

FIG. 3

09 APR 2012

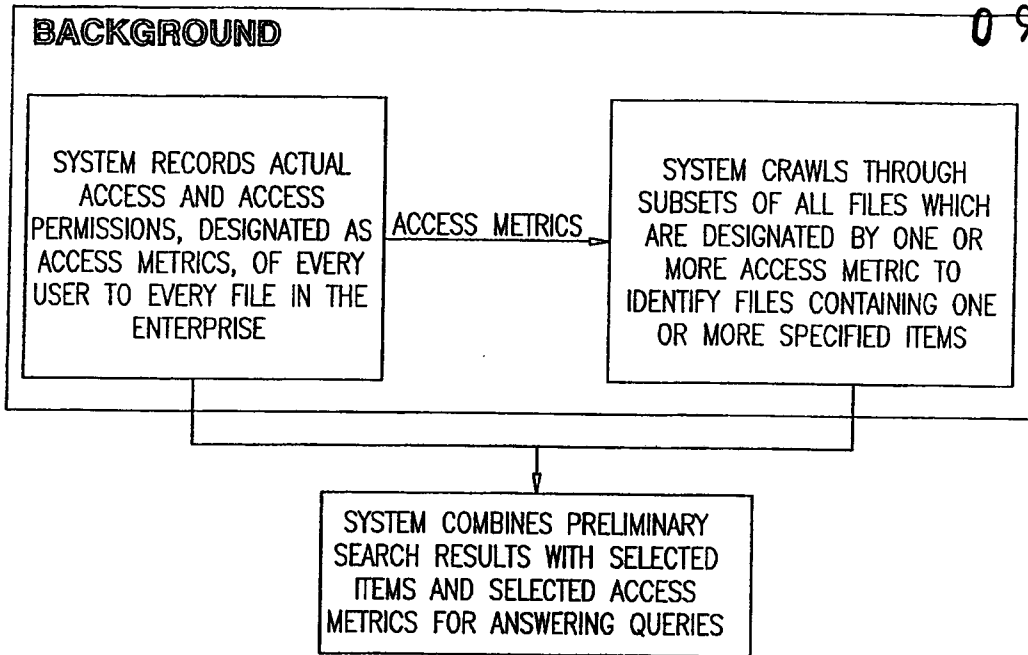
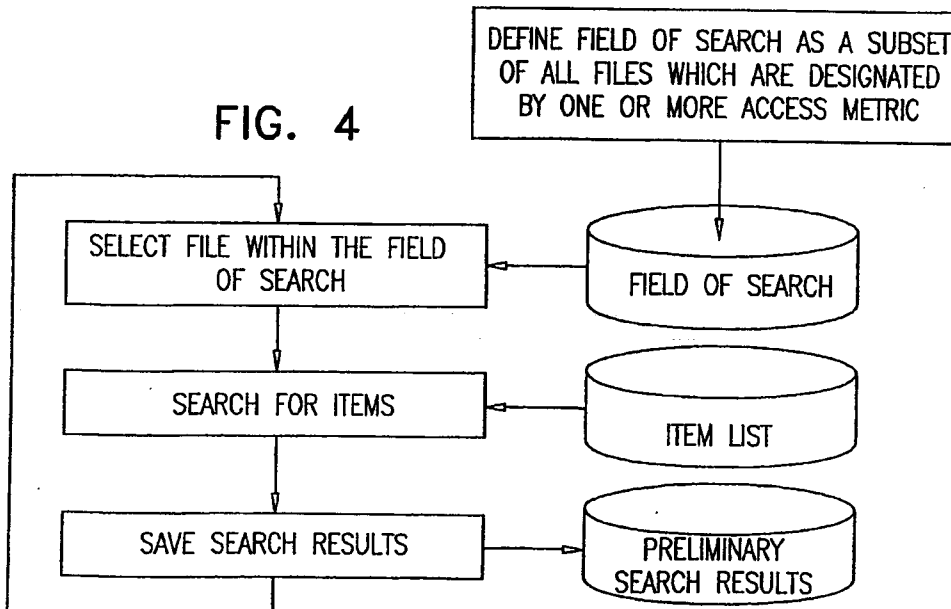


FIG. 4



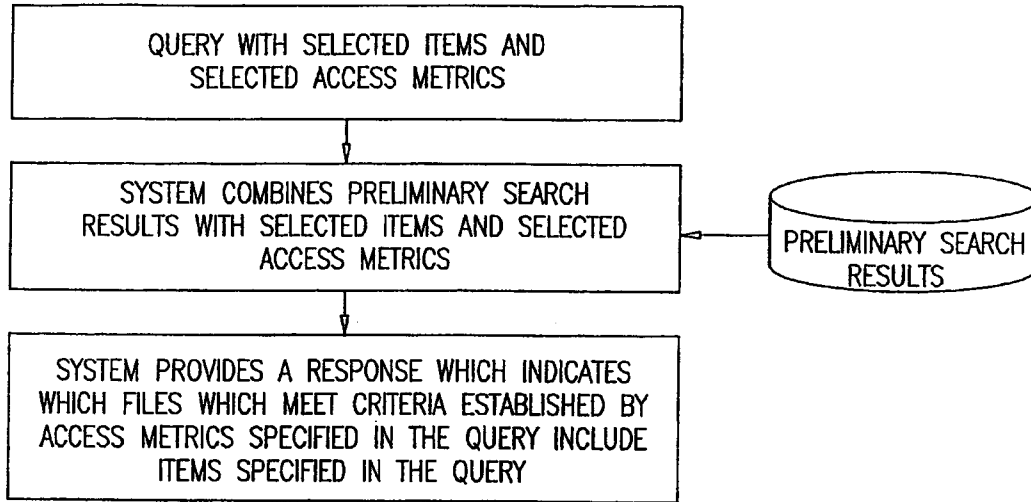
Shanker

ORIGINAL

3035 DELIMP 12

09 APR 2012

FIG. 5



Shanker

Archana Shanker
of Anand and Anand Advocates
Agent for the Applicant

ENTERPRISE LEVEL DATA MANAGEMENT

REFERENCE TO RELATED APPLICATIONS

Reference is made to U.S. Provisional Patent Application Serial No. 61/240,726, filed September 9, 2009 and entitled USE OF ACCESS METRIC IN LARGE SCALE DATA MANIPULATION, the disclosure of which is hereby incorporated by reference and priority of which is hereby claimed pursuant to 37 CFR 1.78(a) (4) and (5)(i).

Reference is also made to the following patents and patent applications, owned by assignee, the disclosures of which are hereby incorporated by reference, which are believed to relate to subject matter similar to the subject matter of the present application:

U.S. Patent Nos. 7,555,482 and 7,606,801;

U.S. Published Patent Application Nos. 2007/0244899, 2008/0271157, 2009/0100058, 2009/0265780 and 2009/0119298; and

U.S. Patent Application No. 12/498,675.

FIELD OF THE INVENTION

The present invention relates to data management generally and more particularly enterprise level data management.

BACKGROUND OF THE INVENTION

The following patent publications and articles are believed to represent the current state of the art:

U.S. Patent Nos.: 7,031,984; 6,338,082; 6,928,439; 7,555,482; 7,606,801; 6,393,468; 5,899,991; 7,068,592 and 5,465,387.

U.S. Published Patent Application Nos.: 2003/0051026; 2004/0249847; 2004/0186809; 2005/0108206; 2005/0278334; 2005/0203881; 2005/0120054; 2005/0086529; 2006/0064313; 2006/0184530; 2006/0277184; 2006/0184459 and 2007/0203872.

SUMMARY OF THE INVENTION

The present invention provides improved systems and methodologies for data management.

There is thus provided in accordance with a preferred embodiment of the present invention a system for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise, the system including background data characterization functionality characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, the at least one access metric being selected from data access permissions and actual data access history and near real time data matching functionality selecting the data of interest by considering only data elements which have the at least one content characteristic thereof and the at least one access metric thereof from among the multiplicity of data elements.

Preferably, the near real time data matching functionality includes background field of search definition and searching functionality operative to define a field of search in accordance with the at least one access metric and to search within the field of search based on the at least one content characteristic.

In accordance with a preferred embodiment of the present invention the near real time data matching functionality includes background field of search definition and searching functionality operative to define a field of search in accordance with the at least one access metric multiple times and to search within the field of search multiple times, wherein the at least one access metric is different at least some of the multiple times.

Preferably, the at least one access metric is a dynamic metric which changes over time during operation of the enterprise.

In accordance with a preferred embodiment of the present invention the system also includes automatic field of search redefinition and search functionality operative to redefine the field of search in accordance with the at least one access metric multiple times and search within the field of search multiple times, wherein the at least one access metric is different at least some of the multiple times. Additionally, the

automatic field of search redefinition and search functionality is operative to search only within those portions of the field of search that have been modified or added as the result of redefining the field of search in accordance with changes in the at least one access metric. Alternatively or additionally, the system also includes data element status monitoring functionality, noting the current status of data elements that have been modified, added or removed in accordance with changes in the at least one access metric.

Preferably, the searching is prioritized at least in accordance with at least one access metric related prioritization characteristic. Additionally, results of the searching are ordered at least in accordance with at least one access metric related prioritization characteristic.

In accordance with a preferred embodiment of the present invention the near real time data matching functionality includes searching functionality for searching for data elements which have the at least one content characteristic thereof and identification functionality operative separately from the searching for data elements which have the at least one content characteristic thereof, identifying data elements from among the multiplicity of data elements in accordance with the at least one access metric and combining functionality, combining results of the searching and the identifying. Additionally, the searching and the identifying are performed by separate entities.

There is also provided in accordance with another preferred embodiment of the present invention a method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise, the method including characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, the at least one access metric being selected from data access permissions and actual data access history and selecting the data of interest by considering only data elements which have the at least one content characteristic thereof and the at least one access metric thereof from among the multiplicity of data elements.

Preferably, the considering includes defining a field of search in accordance with the at least one access metric and searching within the field of search based on the at least one content characteristic.

In accordance with a preferred embodiment of the present invention the considering includes defining a field of search in accordance with the at least one access metric multiple times and searching within the field of search multiple times, wherein the at least one access metric is different at least some of the multiple times.

Preferably, the at least one access metric is a dynamic metric which changes over time during operation of the enterprise.

In accordance with a preferred embodiment of the present invention the method also includes automatically redefining the field of search in accordance with the at least one access metric multiple times and searching within the field of search multiple times, wherein the at least one access metric is different at least some of the multiple times. Additionally, the searching includes searching only within those portions of the field of search that have been modified or added as the result of redefining the field of search in accordance with changes in the at least one access metric. Alternatively or additionally, the method also includes noting the current status of data elements that have been modified, added or removed in accordance with changes in the at least one access metric.

In accordance with a preferred embodiment of the present invention the searching is prioritized at least in accordance with at least one access metric related prioritization characteristic. Additionally, results of the searching are ordered at least in accordance with at least one access metric related prioritization characteristic.

Preferably, the selecting the data of interest by considering only data elements which have the at least one content characteristic thereof and the at least one access metric thereof from among the multiplicity of data elements includes searching for data elements which have the at least one content characteristic thereof, separately from the searching for data elements which have the at least one content characteristic thereof, identifying data elements from among the multiplicity of data elements in accordance with the at least one access metric and combining results of the searching and the identifying. Additionally, the searching and the identifying are performed by separate entities.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description, taken in conjunction with the drawings in which:

Fig. 1 is a simplified pictorial illustration of one example of operation of the system and methodology of the present invention;

Figs. 2A and 2B are a simplified pictorial illustration of another example of operation of the system and methodology of the present invention;

Fig. 3 is a simplified block diagram illustration of the system and methodology of the present invention;

Fig. 4 is a simplified block diagram illustration of functionality for background characterization of data at least by at least one content characteristic thereof and at least one access metric thereof, useful in the system and methodology of Fig. 3; and

Fig. 5 is a simplified block diagram illustration of functionality for selecting data of interest from among a multiplicity of data elements by considering only data elements which are characterized by a given content characteristic and a given access metric thereof.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Reference is now made to Fig. 1, which is a simplified pictorial illustration of one example of operation of the system and methodology of the present invention. As seen in Fig. 1, there is provided a system and method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise.

Two essential functions are performed:

characterizing the data of interest at least by at least one content characteristic thereof and at least one access metric thereof, the at least one access metric being selected from data access permissions and actual data access history; and

selecting the data of interest by considering only data elements which have the at least one content characteristic thereof and the at least one access metric thereof from among the multiplicity of data elements.

In the example of Fig. 1, an enterprise typically has multiple servers 100 which may be located in disparate locations and are connected by a network 102. The network is typically connected to many clients, of which client 104, a computer which is used by a user named John, is an example. The servers 100 typically contain many files, which are typically numbered in the thousands, hundreds of thousands or millions. John typically has access permissions to some but not all of the files of the enterprise and in any given period, such as a week, month or year, actually accesses some but not all of the files of the enterprise for which he has access permission.

As illustrated pictorially at stage A in Fig. 1, the system and methodology of the present invention operates in the background to record actual access and access permissions of every user to every file in the enterprise. It is appreciated that the scope of activities of the system and methodology of the present invention may be restricted to exclude certain users and certain files.

The system and methodology of the present invention also operates in the background to crawl over subsets of all files in the enterprise and to maintain lists of files which include one or more specified item such as a text or non-text item, a string and one or more specified keywords.

Preferably, subsets of all files are selected in accordance with access permission metrics. For example, for personnel having access permission to legal department files, the subset for crawling is the legal department files.

The specified text item or items may be selected by a manager as being appropriate for each subset. Thus, for example for the subject of legal department files, keywords such as "confidential" "lawsuit" and "judgment" may be appropriate. In other contexts, strings of various types, such as sequences of numbers or non-textual characters, may be employed. The set of items may be updated from time to time by an authorized manager.

Returning to the example of Fig. 1, it is seen that at stage B, John abruptly terminates his employment at the enterprise. In accordance with company policy, as seen at stage C, the IT Security Manager immediately queries the system to indicate what files marked "Confidential" John had actually accessed during the month previous to termination of his employment. The IT Security Manager receives a response to his query in near real time, typically within a minute.

It is a particular feature of the present invention that due to the background operation of the system and methodology of the present invention whereby the history of actual access of every user to every file in the enterprise is recorded and lists of files which include specified items are maintained, the query of the IT Security Manager can be responded to in near real time. The system and methodology of the present invention achieves this near real time response by combining currently available actual access and access permissions information with preliminary search result information.

Reference is now made to Figs. 2A and 2B, which are a simplified pictorial illustration of another example of operation of the system and methodology of the present invention. As seen in Figs. 2A and 2B, there is provided a system and method for identifying data of interest from among a multiplicity of data elements residing on multiple platforms in an enterprise. The same two essential functions described hereinabove with reference to Fig. 1 are performed by the system and functionality of the present invention.

As in the example of Fig. 1, an enterprise typically has multiple servers 200 which may be located in disparate locations and are connected by a network 202.

The network is typically connected to many clients. The servers 200 typically contain many files, which are typically numbered in the thousands, hundreds of thousands or millions.

As illustrated pictorially at stage A in Fig. 2A, similarly to Fig. 1, the system and methodology of the present invention operates in the background to record actual access and access permissions of every user to every file in the enterprise. It is appreciated that the scope of activities of the system and methodology of the present invention may be restricted to exclude certain users and certain files.

The system and methodology of the present invention also operates in the background to crawl over subsets of all files in the enterprise and to maintain lists of files which include one or more specified items.

Preferably, subsets of all files are selected in accordance with access permission metrics. For example, for personnel having access permission to legal department files, the subset for crawling is the legal department files.

The specified items may be selected by a manager as being appropriate for each subset. Thus, for example for the subject of legal department files, keywords such as "merger" "acquisition" and "buyout" may be appropriate. The set of keywords may be updated from time to time by an authorized manager.

Returning to the example of Figs. 2A and 2B, it is seen that at stage A in Fig. 2A, an IT Security Manager becomes aware of a leak of company information to the press. The IT Security Manager queries the system as follows:

1. Did the person to whom the leak is attributed actually access the computer files relating to mergers and acquisitions?
2. Are the computer files relating to mergers and acquisitions available to "everyone" in the enterprise?

The IT Security Manager receives a response to his query in near real time, typically within a minute.

It is a particular feature of the present invention that due to the background operation of the system and methodology of the present invention whereby the history of actual access and access permissions of every user to every file in the enterprise is recorded and lists of files which include specified items are maintained, the query of the IT Security Manager can be responded to in near real time. The system and

methodology of the present invention achieves this near real time response by combining currently available actual access and access permissions information with preliminary search result information.

On the basis of the response to his queries, the IT Security Manager immediately orders elimination of the access permission of "everyone" to the merger and acquisition files and orders that the access permission to the mergers and acquisitions files be henceforth limited to the legal department and the business development department.

Reference is now made to Fig. 3, which is a simplified block diagram illustration of the system and methodology of the present invention. As seen in Fig. 3 and described hereinabove in Figs. 1, 2A and 2B with reference to two specific examples, the system and methodology of the present invention includes the following functionality which takes place in the background:

Actual access of every user to every file in the enterprise is recorded and stored in a database. Access permissions of every user to every file in the enterprise are recorded and stored in a database. This functionality is embodied in a system, commercially available under the trademark DatAdvantage by an affiliate of the assignee of the present invention, Varonis Systems Inc. of New York, NY and is described in U.S. Patent 7,606,801 and in U.S. Published Patent Application 2009/0265780 of the present assignee, the disclosures of which are hereby incorporated by reference. Access permissions and/or actual access are together designated as access metrics and may be used to designate subsets of all of the files in the enterprise.

Crawling through files which are designated by one or more access metrics to identify files containing one or more specified items.

Upon receipt of a query, which could include a request for a report, the system combines information relating to actual access and/or access permissions with preliminary search result information, such as that provided by the crawling functionality described hereinabove, to provide a response which indicates which files which meet criteria established by access metrics specified in the query include items specified in the query.

Reference is now made to Fig. 4, which is a simplified block diagram illustration of crawling functionality for background characterization of data at least by

at least one content characteristic thereof and at least one access metric thereof, useful in the system and methodology of Fig. 3.

The crawling functionality of Fig. 4 includes an initial step of defining a field of search for crawling in accordance with one or more access metric. The access metric is defined by one or both of access permissions and actual access and may change over time. For example the field of search may be: all files to which personnel of the legal department have access permission and have been accessed at least once within the last one year.

The system examines each file in the defined field of search for the presence of at least one item from among a collection of items stored in a item list database. Identification of files containing at least one item in the collection is stored in a preliminary search results database.

Reference is now made to Fig. 5, which is a simplified block diagram illustration of functionality for selecting data of interest from among a multiplicity of data elements by considering only data elements which are characterized by a given content characteristic and a given access metric thereof. The functionality of Fig. 5 takes place in response to a query which selects one or more items from among the specified items and one or more selected access metrics from among the access metrics used to define the field of search.

All of the files whose identification appear in the preliminary search results database are examined:

1. to ascertain which files include the selected items set forth in the query; and
2. to ascertain which files meet criteria established by the selected access metrics set forth in the query.

The foregoing two examinations may be conducted in any suitable order.

Files that both include the selected items set forth in the query and meet criteria established by the selected access metrics set forth in the query are reported in a response to the query.

It will be appreciated by persons skilled in the art that the present invention is not limited by what has been particularly shown and described hereinabove. Rather the scope of the present invention includes both combinations and

subcombinations of the various features described hereinabove as well as modifications thereof which would occur to persons skilled in the art upon reading the foregoing description and which are not in the prior art.