

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
1 December 2005 (01.12.2005)

PCT

(10) International Publication Number
WO 2005/114493 A1

(51) International Patent Classification⁷: **G06F 17/30**

(21) International Application Number:
PCT/US2005/017977

(22) International Filing Date: 19 May 2005 (19.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/573,147 21 May 2004 (21.05.2004) US

(71) Applicant (for all designated States except US): **COM-
PUTER ASSOCIATES THINK, INC.** [US/US]; One
Computer Associates Plaza, Islandia, NY 11749 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BETTS, Christo-
pher** [AU/AU]; 10 Ramage Rd., Mt. Dandenong Victo-
ria 3767 (AU). **ROGERS, Tony** [AU/AU]; P.O. Box 2232,
Rowville, VIC 3178 (AU).

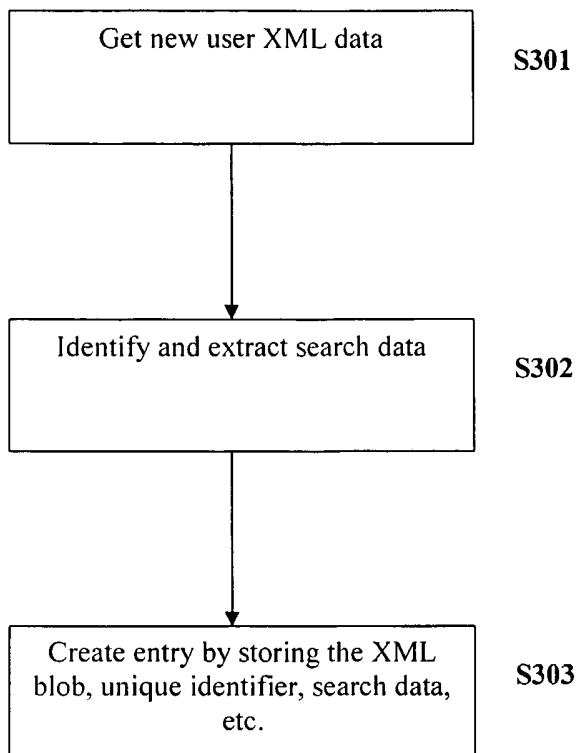
(74) Agents: **PEDERSEN, Luke, K.** et al.; Baker Botts L.L.P.,
2001 Ross Avenue, Suite 600, Dallas, TX 75201-2980
(US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ,
OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL,
SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,
VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO,

[Continued on next page]

(54) Title: METHODS AND SYSTEMS FOR DATA STORAGE



(57) Abstract: A method for storing data includes storing a user entry in a data repository, wherein the user entry comprises a unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.



SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *with international search report*

METHODS AND SYSTEMS FOR DATA STORAGE

5

BACKGROUND

REFERENCE TO RELATED APPLICATION

The present disclosure is based on and claims the benefit of Provisional application
10 Serial No. 60/573,147 filed May 21, 2004, the entire contents of which are herein
incorporated by reference.

1. TECHNICAL FIELD

The present disclosure relates generally to system performance and, more particularly,
to methods and systems for data storage.

15

2. DESCRIPTION OF THE RELATED ART

Web services provide a way for automated resources to be accessed by computer
systems through the Internet. Computer system(s) as referred to herein may include(s)
individual computers, servers, computing resources, networks, or combinations thereof, etc.
20 Computer systems using standards-based web services can use Extensible Markup Language
("XML") to communicate with each other. XML is language that uses a human readable
format to tag data for web services.

The basic security requirements of web services can be the same as they are for
traditional programs. For example, a user's identity can be authenticated before they are able
25 to access web services. The Services Provisioning Markup Language (SPML) is the web
services standard that defines how users can be stored, recovered, and how attributes are to be
associated with those users.

One of the problems that web services often experience is slow performance. This results because the human readable XML data format is particularly verbose and can be computationally expensive to parse. Slow performance can be especially troublesome in user data repositories, which frequently have to cope with high access rates. For example, a large organization's administrative system might log on thousands of users around 9:00 am every day. If each request requires significant computational time, the load on the system would dramatically increase, slowing down system performance.

The SPML standard is an XML based framework for exchanging user, resource and service provisioning information between systems. SPML is used with directory technology for user storage and incorporates the Directory Service Markup Language (DSML) protocol for some elements of user access. However, the SPML architecture does not allow storage of entries in XML in a data repository, such as a relational database or directory. Thus, some form of translation is required for each SPML request that is to be stored in the various data repositories. For example, a SPML server might have to assemble the data in order to create SPML responses. This assembly step can be slow, as it requires a directory or database to be interrogated, possibly multiple times, and the data then collated and assembled into XML messages.

XML signatures are sensitive to the exact format of the XML data. Thus, in order to convert the XML data to a standard signature form, the XML data would have to undergo a canonicalization process. This process requires a system to perform additional parsing and processing work, which can slow system performance.

Accordingly, a need exists for techniques that overcome the disadvantages of conventional data storage techniques. It would be beneficial to have a method and system for optimum data storage of user data and timely and efficient processing of that data.

5 **SUMMARY**

A method for storing data, according to an embodiment of the present disclosure, includes storing a user entry in a data repository, wherein the user entry comprises a unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.

10 A system for storing data, according to an embodiment of the present disclosure, includes means for storing a user entry in a data repository, wherein the user entry comprises a unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.

A computer storage medium including computer executable code for storing data, 15 according to an embodiment of the present disclosure, includes code for storing a user entry in a data repository, wherein the user entry comprises a unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.

BRIEF DESCRIPTION OF THE DRAWINGS

20 A more complete appreciation of the present disclosure and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

Figure 1 shows a block diagram of an exemplary computer system capable of implementing the method and system of the present application;

Figure 2 shows a block diagram illustrating the format of a user entry, according to an embodiment of the present disclosure;

5 Figure 3 shows a flow chart illustrating the execution of a add user request, according to an embodiment of the present disclosure;

Figure 4 shows a flow chart illustrating the execution of a modify user request, according to an embodiment of the present disclosure;

Figure 5 shows a flow chart illustrating the execution of a delete user request,
10 according to an embodiment of the present disclosure; and

Figure 6 shows a flow chart illustrating the execution of a search user request, according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

15 The present disclosure provides tools (in the form of methodologies, apparatuses, and systems) for storing data. The tools may be embodied in one or more computer programs stored on a computer readable medium or program storage device and/or transmitted via a computer network or other transmission medium.

20 The following exemplary embodiments are set forth to aid in an understanding of the subject matter of this disclosure, but are not intended, and should not be construed, to limit in any way the claims which follow thereafter. Therefore, while specific terminology is employed for the sake of clarity in describing some exemplary embodiments, the present disclosure is not intended to be limited to the specific terminology so selected, and it is to be

understood that each specific element includes all technical equivalents which operate in a similar manner.

Figure 1 shows an example of a computer system **100** which may implement the methods and systems of the present disclosure. The systems and methods of the present disclosure may be implemented in the form of a software application running on a computer system, for example, a mainframe, personal computer (PC), handheld computer, server, etc. The software application may be stored on a recording media locally accessible by the computer system, for example, floppy disk, compact disk, hard disk, etc., or may be remote from the computer system and accessible via a hard wired or wireless connection to a network, for example, a local area network, or the Internet.

The computer system **100** can include a central processing unit (CPU) **102**, program and data storage devices **104**, a printer interface **106**, a display unit **108**, a (LAN) local area network data transmission controller **110**, a LAN interface **112**, a network controller **114**, an internal bus **116**, and one or more input devices **118** (for example, a keyboard, mouse etc.). As shown, the system **100** may be connected to a database **120**, via a link **122**.

The specific embodiments described herein are illustrative, and many variations can be introduced on these embodiments without departing from the spirit of the disclosure or from the scope of the appended claims. Elements and/or features of different illustrative embodiments may be combined with each other and/or substituted for each other within the scope of this disclosure and appended claims.

According to an embodiment of the present disclosure, user data can be stored in an "SPML XML" format in a data repository in order to increase system performance. By storing the user data in XML format in a data repository, such as a directory, significant

speed increases can be achieved for SPML requests and responses. For example, if the request is to “read a user”, the response would return a single text string retrieved from the directory without the XML parsing that would typically be required for the response to successfully execute. Significant performance increases can be achieved in the case where
5 multiple user records are being returned at the same time. In that case, data can be read in a single directory operation that returns all user data XML strings at one time, rather than collating into XML the details of many different users.

Figure 2 is a block diagram illustrating the format of a user entry, according to an embodiment of the present disclosure. User data, such as SPML user data, can be stored as a
10 user entry **201** containing an XML blob (contains the raw XML data) **208** as a single valued string attribute, a unique identifier **202**, accompanying search data fields to facilitate fast searching **203-206**, and a signature **207**.

An arbitrary number of search data fields can be extracted from the user records when they are first written and stored as separate directory attributes. The search data fields can be
15 used to index the data, allowing a system to promptly process data for fast retrieval. At least one arbitrary search field can be provided, for example, arbitrary search data#1 **203**, arbitrary search data#2 **204**, and arbitrary search data #N **205**.

At least one predetermined search data **206** can also be provided containing generic searchable attributes of the form “name=value”. The predetermined search data **206** can be
20 used, for example, if the searchable user attributes are not all known in advance, or if generic data needs to be searched.

A signature 207 can also be stored in a user entry 201. By storing a raw XML version of a signature 207 in a user entry 201, a system can simply return the signed data with an intact signature when requested, without requiring additional processing work.

Requests can be of four basic types: “add a user”, “modify a user”, “delete a user”,
5 and “search for a user(s)” based on a search filter. Each request will be described in further detail below.

Figure 3 is a flow chart illustrating the execution of a add user request, according to an embodiment of the present disclosure. The XML data is first extracted from a query, such as an SPML query (Step S301). The extracted XML data is then parsed in order to identify
10 and extract the search data (Step S302). Finally, a user entry is created in a data repository, such as a directory or database, containing the following data: an XML blob, the unique identifier, and the search data containing the attributes required for searching (Step S303).

Figure 4 is a flow chart illustrating the execution of a modify user request, according to an embodiment of the present disclosure. The original entry can be read (Step S401) and
15 the XML data parsed in order to identify the data values that are going to be modified (Step S402). The attribute modifications can then be made to the XML structure by replacing the old identified data values with the new modified data values in the structure (Step S403). The modified XML can then be written back to the data repository, such as the directory, (Step S404) along with the changed search attributes, if any (Step S405). If the complete
20 replacement of an existing user entry is executed, the entire existing user entry will be deleted in accordance with the “delete user” operation and the new entry will be added in accordance with the “add user” operation.

Figure 5 is a flow chart illustrating the execution of a delete user request, according to an embodiment of the present disclosure. The XML data can be parsed to extract the unique identifier used to identify the user entry in the data repository (Step S501). No further parsing may be required, once the data has been parsed enough for the unique identifier to be
5 extracted. The user entry can then be deleted from the data repository (Step S502).

Figure 6 is a flow chart illustrating the execution of a search user request, according to an embodiment of the present disclosure. A user query can be parsed to extract a search filter (Step S601) and then the data repository can be searched based on the search filter (Step S602). In many cases, the search filter can be used unaltered. However, the search filter can
10 be modified for any attributes that use the predetermined search data. If the search filter is returning a user entry, or a set of user entries, then the data repository can be sent the search filter with a request to return the raw XML data of any matching entries. The processing that can be required for the SPML response is to concatenate all the XML entries and send them back to the original requesting user without parsing. If the search filter is only returning a
15 subset of attributes, then an XML query can be manually constructed, and the attributes included in the XML query. The results obtained from the data repository can then be organized and output (Step S603).

Numerous additional modifications and variations of the present disclosure are possible in view of the above-teachings. It is therefore to be understood that within the scope
20 of the appended claims, the present disclosure may be practiced other than as specifically described herein.

What is claimed is:

1. A method for storing data, comprising:
storing a user entry in a data repository, wherein the user entry comprises a unique
5 identifier and a string of data, wherein the string of data comprises an XML data of the user
entry as a single valued string.
2. The method of claim 1, wherein the data repository is a directory or database.
- 10 3. The method of claim 1, wherein the user entry is an SPML user entry.
4. The method of claim 1, wherein the user entry further comprises at least one
arbitrary search term.
- 15 5. The method of claim 1, wherein the user entry further comprises at least one
predetermined search term.
6. The method of claim 1, wherein the user entry further comprises a signature.
- 20 7. The method of claim 1, wherein the user entry is added to the data repository by
extracting XML data from a query; identifying and extracting search data; and creating the
user entry by storing the unique identifier, the XML data and the search data.

8. The method of claim 1, wherein the user entry is modified by reading the user entry; parsing through the XML data to identify which values are going to be modified; replacing the old identified values with new modified values; and storing the modified XML
5 data in the data repository.

9. The method of claim 8, further comprising storing the modified search data in the data repository.

10 10. The method of claim 1, wherein the user entry is deleted by extracting the unique identifier for the user entry; and deleting the user entry from the data repository based on the unique identifier.

11. The method of claim 1, wherein the user entry is searched by parsing a user query
15 in order to extract a search filter; searching the data repository based on the search filter; and organizing and outputting the results obtained from the data repository.

12. The method of claim 11, wherein the search filter is unaltered or modified for any attributes that use a predetermined search term.

20

13. The method of claim 11, wherein the XML data for one or more user entries requested is concatenated and sent back to the requesting user.

14. A system for storing data, comprising:

a system for storing a user entry in a data repository, wherein the user entry comprises a unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.

5

15. The system of claim 14, wherein the data repository is a directory or database.

16. The system of claim 14, wherein the user entry is an SPML user entry.

10 17. The system of claim 14, wherein the user entry further comprises at least one arbitrary search term.

18. The system of claim 14, wherein the user entry further comprises at least one predetermined search term.

15

19. The system of claim 14, wherein the user entry further comprises a signature.

20 20. The system of claim 14, wherein the user entry is added to the data repository by extracting XML data from a query; identifying and extracting search data; and creating the user entry by storing the unique identifier, the XML data and the search data.

21. The method of claim 14, wherein the user entry is modified by reading the user entry; parsing through the XML data to identify which values are going to be modified;

replacing the old identified values with new modified values; and storing the modified XML data in the data repository.

22. The system of claim 21, further comprising a system for storing the modified
5 search data in the data repository.

23. The system of claim 14, wherein the user entry is deleted by extracting the unique identifier for the user entry; and deleting the user entry from the data repository based on the unique identifier.

10

24. The system of claim 14, wherein the user entry is searched by parsing a user query in order to extract a search filter; searching the data repository based on the search filter; and organizing and outputting the results obtained from the data repository.

15 25. The system of claim 24, wherein the search filter is unaltered or modified for any attributes that use a predetermined search term.

26. The system of claim 24, wherein the XML data for one or more user entries requested is concatenated and sent back to the requesting user.

20

27. A computer storage medium including computer executable code for storing data, comprising:

code for storing a user entry in a data repository, wherein the user entry comprises a

unique identifier and a string of data, wherein the string of data comprises an XML data of the user entry as a single valued string.

28. The computer storage medium of claim 27, wherein the data repository is a
5 directory or database.

29. The computer storage medium of claim 27, wherein the user entry is an SPML
user entry.

10 30. The computer storage medium of claim 27, wherein the user entry further
comprises at least one arbitrary search term.

31. The computer storage medium of claim 27, wherein the user entry further
comprises at least one predetermined search term.
15

32. The computer storage medium of claim 27, wherein the user entry further
comprises a signature.

33. The computer storage medium of claim 27, wherein the user entry is added to the
20 data repository by extracting XML data from a query; identifying and extracting search data;
and creating the user entry by storing the unique identifier, the XML data and the search data.

34. The computer storage medium of claim 27, wherein the user entry is modified by

reading the user entry; parsing through the XML data to identify which values are going to be modified; replacing the old identified values with new modified values; and storing the modified XML data in the data repository.

5 35. The computer storage medium of claim 34, further comprising code for storing the modified search data in the data repository.

 36. The computer storage medium of claim 27, wherein the user entry is deleted by extracting the unique identifier for the user entry; and deleting the user entry from the data
10 repository based on the unique identifier.

 37. The computer storage medium of claim 27, wherein the user entry is searched by parsing a user query in order to extract a search filter; searching the data repository based on the search filter; and organizing and outputting the results obtained from the data repository.
15

 38. The computer storage medium of claim 37, wherein the search filter is unaltered or modified for any attributes that use a predetermined search term.

 39. The computer storage medium of claim 37, wherein the XML data for one or
20 more user entries requested is concatenated and sent back to the requesting user.

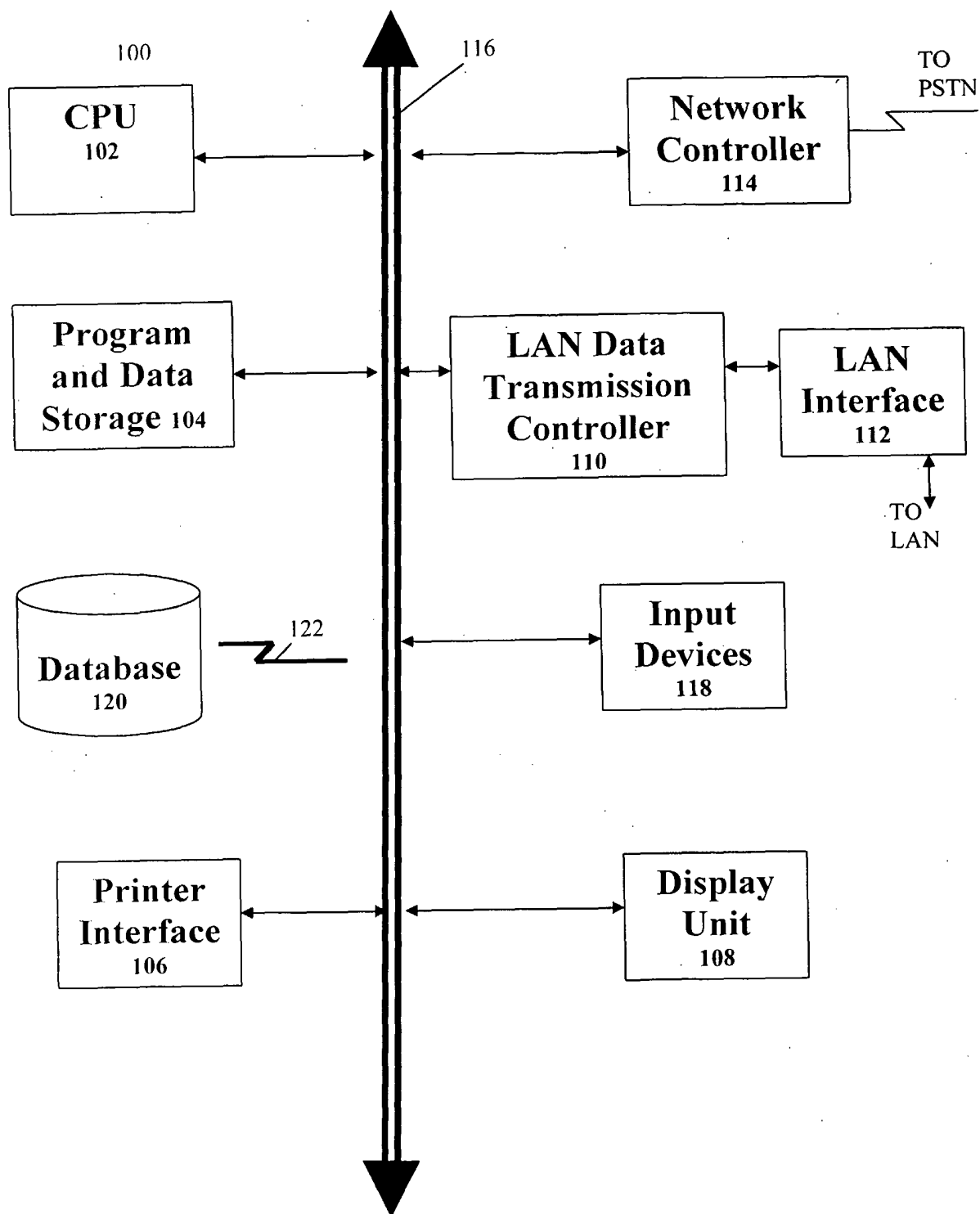
FIGURE 1

FIGURE 2

User Entry 201

Unique Identifier 202	Arbitrary Search Data#1 203	Arbitrary Search Data #2 204	Arbitrary Search Data #N 205	Predetermined Search Data 206	Signature 207	XML BLOB 208
--------------------------	-----------------------------------	------------------------------------	------------------------------------	-------------------------------------	------------------	-----------------

FIGURE 3

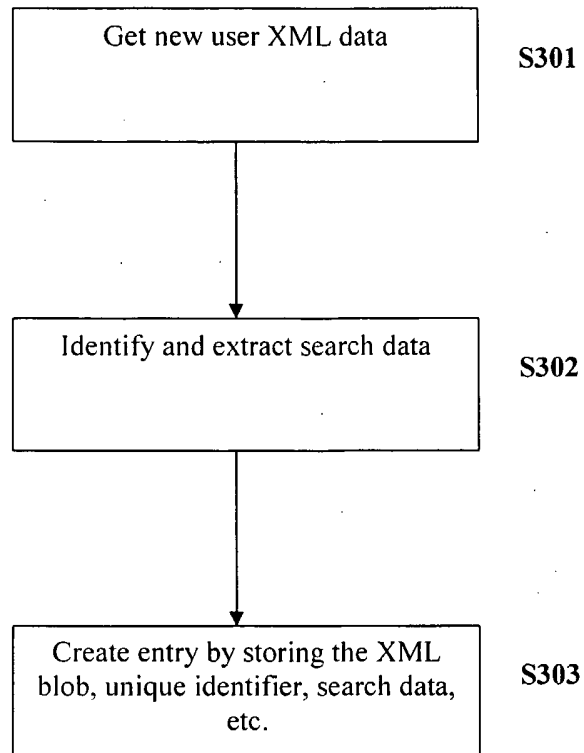


FIGURE 4

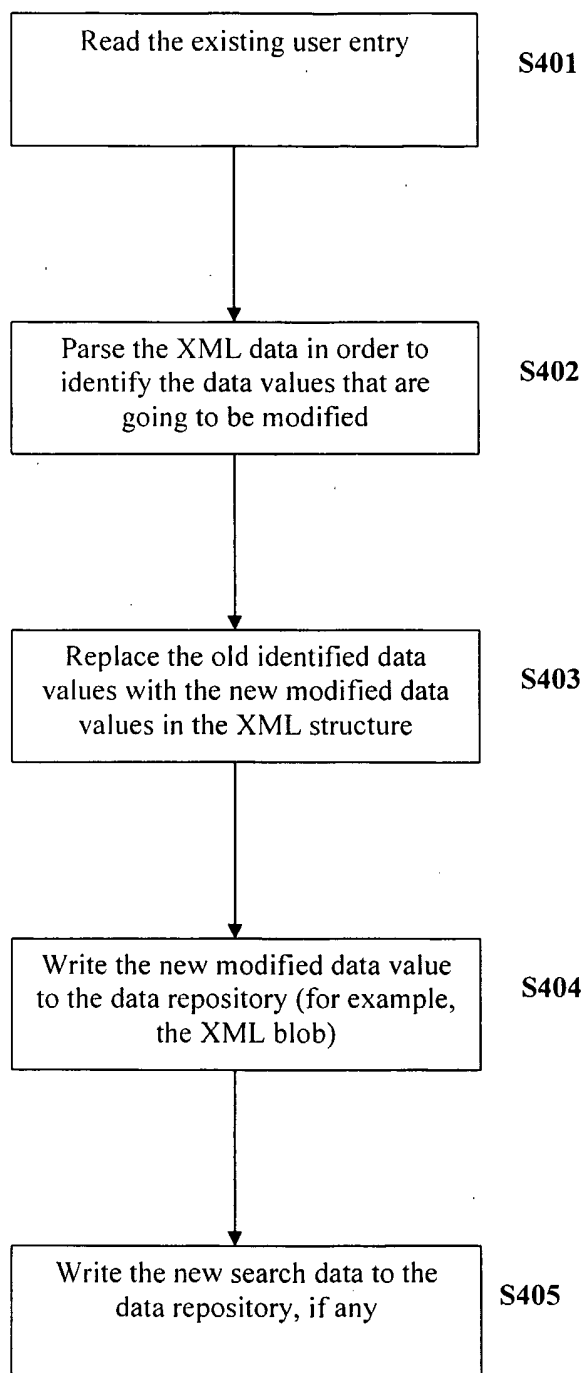


FIGURE 5

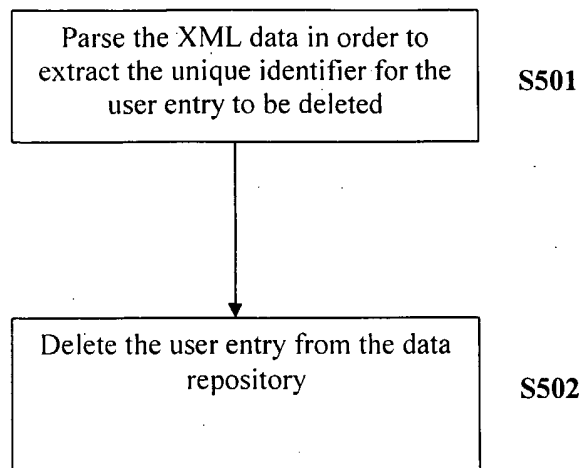
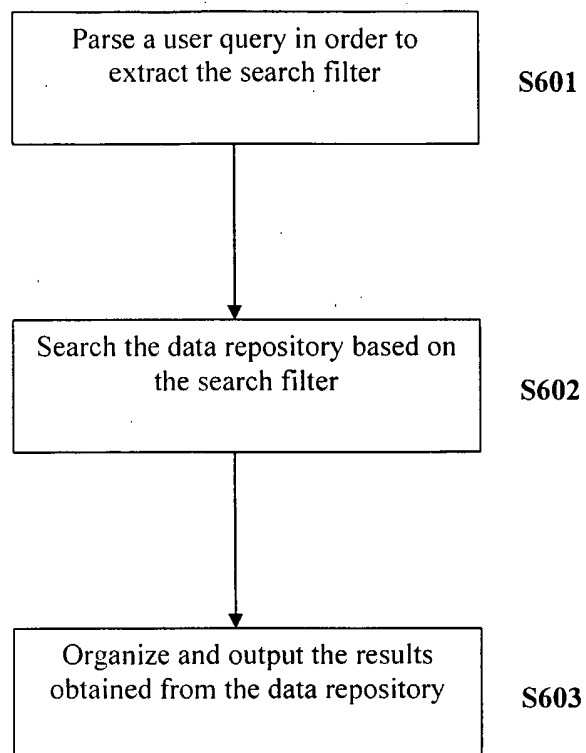


FIGURE 6



INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/017977

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EP0-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2002/123993 A1 (CHAU HOANG K ET AL) 5 September 2002 (2002-09-05) abstract paragraph '0052! - paragraph '0090! paragraph '0209! - paragraph '0399! figures 3,4	1-39
X	WO 02/21339 A (ORACLE CORPORATION; ORACLE INTERNATIONAL CORPORATION) 14 March 2002 (2002-03-14) abstract page 7, line 8 - line 28 page 31 - page 34 figures 4,5	1-39

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

☒ Patent family members are listed in annex.

* Special categories of cited documents:

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

E earlier document 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.

G document member of the same patent family

Date of the actual completion of the international search

17 August 2005

Date of mailing of the international search report

26/08/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Dumitrescu, C

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2005/017977

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BANERJEE S ET AL: "Oracle8i-the XML enabled data management system" DATA ENGINEERING, 2000. PROCEEDINGS. 16TH INTERNATIONAL CONFERENCE ON SAN DIEGO, CA, USA 29 FEB.-3 MARCH 2000, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 29 February 2000 (2000-02-29), pages 561-568, XP010378753 ISBN: 0-7695-0506-6 page 563 - page 565 -----	1-39
X	DAYEN I: "Storing XML in Relational Databases" -, 20 June 2001 (2001-06-20), XP002275971 the whole document -----	1-39
X	US 6 510 434 B1 (ANDERSON DEWEY C ET AL) 21 January 2003 (2003-01-21) abstract column 2, line 36 - column 3, line 43 column 8, line 50 - column 11, line 4 -----	1-39
X	GUIDO N: "XML DATA PROCESSING AND RELATIONAL DATABASE SYSTEMS" XML EUROPE. CONFERENCE PROCEEDINGS. PROCEEDINGS OF XML EUROPE, 26 April 1999 (1999-04-26), pages 713-719, XP001006043 the whole document -----	1-39
A	OASIS COMITEE SPECIFICATION: "SERVICE PROVISIONING MARKUP LANGUAGE (SPML) Version 1.0" -, 'Online! 3 June 2003 (2003-06-03), pages 1-75, XP002340624 - Retrieved from the Internet: URL: http://www.oasis-open.org/committees/download.php/2395/cs-pstc-spml-core-1.0.pdf > 'retrieved on 2005-08-15! the whole document -----	1-39

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US2005/017977

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002123993 A1	05-09-2002	US 2002133484 A1	19-09-2002
		US 2003014397 A1	16-01-2003
		US 2002156772 A1	24-10-2002
WO 0221339 A	14-03-2002	US 2002078068 A1	20-06-2002
		US 2002078094 A1	20-06-2002
		US 2002099687 A1	25-07-2002
		AU 9069301 A	22-03-2002
		CA 2421214 A1	14-03-2002
		EP 1337937 A2	27-08-2003
		JP 2004519755 T	02-07-2004
		WO 0221339 A2	14-03-2002
		US 2005055358 A1	10-03-2005
US 6510434 B1	21-01-2003	NONE	