US007315861B2

(12) **United States Patent**
Seibel et al.

(10) **Patent No.: US 7,315,861 B2**
(45) **Date of Patent: *Jan. 1, 2008**

(54) **TEXT MINING SYSTEM FOR WEB-BASED BUSINESS INTELLIGENCE**

(75) Inventors: **John C. Seibel**, Austin, TX (US); **Yu Feng**, Maplewood, NJ (US); **Robert L. Foster**, Austin, TX (US)

(73) Assignee: **ReachForce, Inc.**

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **11/178,721**

(22) Filed: **Jul. 11, 2005**

(65) **Prior Publication Data**

US 2006/0004731 A1 Jan. 5, 2006

Related U.S. Application Data

(63) Continuation of application No. 09/865,735, filed on May 24, 2001, now Pat. No. 7,003,517.

(60) Provisional application No. 60/206,772, filed on May 24, 2000.

(51) **Int. Cl.**
G06F 17/30 (2006.01)

(52) **U.S. Cl.** **707/10**; 707/3; 707/104.1; 707/204; 705/10; 705/14; 709/206; 709/219

(58) **Field of Classification Search** 707/3, 707/10, 104, 1, 204; 705/10, 14; 709/203, 709/219, 206

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,914,586 A 4/1990 Swinehart et al. 364/200
5,619,648 A 4/1997 Canale et al. 395/200.01
5,630,121 A 5/1997 Braden-Harder et al. ... 395/604
5,649,114 A 7/1997 Deaton et al. 395/214
5,659,469 A 8/1997 Deaton et al. 395/214
5,742,816 A 4/1998 Barr et al. 395/615
5,787,422 A 7/1998 Tukey et al. 707/5
5,809,481 A 9/1998 Baron et al. 705/14
5,897,622 A 4/1999 Blinn et al. 705/26

(Continued)

FOREIGN PATENT DOCUMENTS

EP 1118952 1/2001

(Continued)

OTHER PUBLICATIONS

Paul Dean, "Browsable OLAP Apps on SQL Server Analysis Services," Intelligent Enterprise Magazine, product review, 5 pages, May 7, 2001.

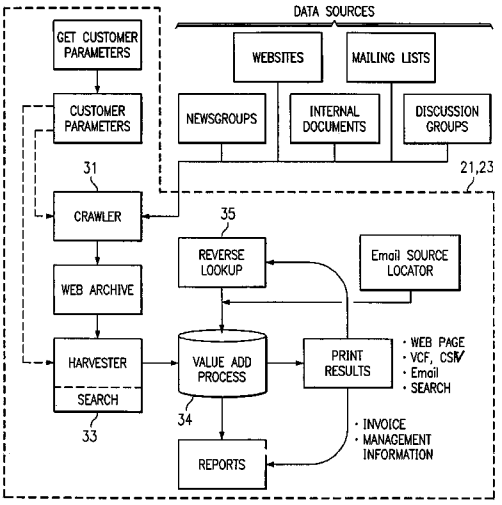
(Continued)

Primary Examiner—Leslie Wong

(57) **ABSTRACT**

A text mining system for collecting business intelligence about a client, as well as for identifying prospective customers of the client, for use in a lead generation system accessible by the client via the Internet. The text mining system has various components, including a data acquisition process that extracts textual data from various Internet sources, a database for storing the extracted data, a text mining server that executes query-based searches of the database, and an output repository. A web server provides client access to the repository, and to the mining server.

11 Claims, 11 Drawing Sheets



IPR2025-00623
UIPATH EXHIBIT 1013

U.S. PATENT DOCUMENTS

5,924,068	A	7/1999	Richard et al.	704/260
5,924,105	A	7/1999	Punch, III et al.	707/513
5,931,907	A	8/1999	Davies et al.	709/218
5,974,398	A	10/1999	Hanson et al.	705/14
5,986,690	A	11/1999	Hendricks	348/7
5,987,247	A	11/1999	Lau	395/702
5,999,927	A	12/1999	Tukey et al.	707/5
6,006,242	A	12/1999	Poole et al.	707/531
6,026,433	A	2/2000	D'Arlach et al.	709/217
6,029,141	A	2/2000	Bezos et al.	705/27
6,029,164	A	2/2000	Birrell et al.	707/3
6,029,174	A	2/2000	Sprenger et al.	707/103
6,029,195	A	2/2000	Herz	709/219
6,034,970	A	3/2000	Levac et al.	370/466
6,055,510	A	4/2000	Henrick et al.	705/14
6,058,398	A	5/2000	Lee	707/104
6,058,418	A	5/2000	Kobata	709/221
6,078,891	A	6/2000	Riordan et al.	705/10
6,105,055	A	8/2000	Pizano et al.	709/204
6,119,101	A	9/2000	Peckover	705/26
6,134,548	A	10/2000	Gottsman et al.	707/5
6,145,003	A	11/2000	Sanu et al.	709/225
6,148,289	A	11/2000	Virdy	705/1
6,151,582	A	11/2000	Huang et al.	705/8
6,151,601	A	11/2000	Papierniak et al.	707/10
6,154,766	A	11/2000	Yost et al.	709/201
6,170,011	B1	1/2001	Macleod Beck et al.	709/224
6,199,081	B1	3/2001	Meyerzon et al.	707/513
6,202,210	B1	3/2001	Ludtke	725/20
6,205,432	B1	3/2001	Gabbard et al.	705/14
6,212,178	B1	4/2001	Beck et al.	370/352
6,226,623	B1	5/2001	Schein et al.	705/35
6,233,575	B1	5/2001	Agrawal et al.	707/6
6,236,975	B1	5/2001	Boe et al.	705/7
6,249,764	B1	6/2001	Kamae et al.	704/270.1
6,256,623	B1	7/2001	Jones	707/3
6,262,987	B1	7/2001	Mogul	370/400
6,263,334	B1	7/2001	Fayyad et al.	707/5
6,282,548	B1	8/2001	Burner et al.	707/104
6,289,342	B1	9/2001	Lawrence et al.	707/7
6,332,154	B2	12/2001	Beck et al.	709/204
6,338,066	B1	1/2002	Martin et al.	707/10
6,345,288	B1	2/2002	Reed et al.	709/201
6,363,377	B1	3/2002	Kravets et al.	707/4
6,377,993	B1	4/2002	Brandt et al.	709/227
6,381,599	B1	4/2002	Jones et al.	707/5
6,393,465	B2	5/2002	Leeds	709/207
6,401,091	B1	6/2002	Butler et al.	707/10
6,401,118	B1	6/2002	Thomas	709/224
6,405,197	B2	6/2002	Gilmour	707/5
6,430,545	B1	8/2002	Honarvar et al.	706/47
6,430,624	B1	8/2002	Jamtgaard et al.	709/246
6,434,544	B1	8/2002	Bakalash et al.	707/2
6,434,548	B1	8/2002	Emens et al.	707/3
6,438,543	B1	8/2002	Kazi et al.	707/5
6,460,038	B1	10/2002	Khan et al.	707/10
6,460,069	B1	10/2002	Berlin et al.	709/201
6,473,756	B1	10/2002	Ballard	707/6
6,477,536	B1	11/2002	Pasumansky et al.	707/102
6,480,842	B1	11/2002	Agassi et al.	707/4
6,480,885	B1	11/2002	Olivier	709/207
6,490,582	B1	12/2002	Fayyad et al.	707/6
6,490,620	B1	12/2002	Ditmer et al.	709/224
6,493,703	B1	12/2002	Knight et al.	707/3
6,510,432	B1	1/2003	Doyle	707/10
6,516,337	B1	2/2003	Tripp et al.	709/202
6,519,571	B1	2/2003	Guheen et al.	705/14
6,523,021	B1	2/2003	Monberg et al.	707/2
6,529,909	B1	3/2003	Bowman-Amuah	707/10
6,546,416	B1	4/2003	Kirsch	709/206
6,555,738	B2	4/2003	Hughes et al.	84/609
6,557,008	B1	4/2003	Temple, III et al.	707/104.1
6,564,209	B1	5/2003	Dempski et al.	707/3
6,567,797	B1	5/2003	Schuetze et al.	707/2
6,567,803	B1	5/2003	Ramasamy et al.	707/4
6,571,234	B1	5/2003	Knight et al.	707/3
6,574,619	B1	6/2003	Reddy et al.	707/2
6,578,009	B1	6/2003	Shinozaki	705/10
6,581,054	B1	6/2003	Bogrett	707/4
6,598,054	B2	7/2003	Schuetze et al.	707/103
6,606,644	B1	8/2003	Ford et al.	709/203
6,609,124	B2	8/2003	Chow et al.	707/5
6,611,839	B1	8/2003	Nwabueze	707/101
6,615,184	B1	9/2003	Hicks	705/26
6,621,505	B1	9/2003	Beauchamp et al.	345/764
6,625,598	B1	9/2003	Kraffert	707/3
6,651,048	B1	11/2003	Agrawal et al.	707/2
6,651,055	B1	11/2003	Kilmer et al.	707/3
6,651,065	B2	11/2003	Brown et al.	707/10
6,665,658	B1	12/2003	DaCosta et al.	707/3
6,668,259	B1	12/2003	Ventura et al.	707/103
6,677,963	B1	1/2004	Mani et al.	345/764
6,684,207	B1	1/2004	Greenfield et al.	707/3
6,684,218	B1	1/2004	Santos et al.	707/102
6,691,105	B1	2/2004	Virdy	707/3
6,700,575	B1	3/2004	Bovarnick et al.	345/440
6,700,590	B1	3/2004	DeMesa et al.	345/744
6,714,979	B1	3/2004	Brandt et al.	709/225
6,721,689	B2	4/2004	Markle et al.	702/188
6,732,161	B1	5/2004	Hess et al.	709/219
6,757,689	B2	6/2004	Battas et al.	707/101
6,763,353	B2	7/2004	Li et al.	707/4
6,769,009	B1	7/2004	Reisman	709/201
6,769,010	B1	7/2004	Knapp et al.	709/203
6,772,196	B1	8/2004	Kirsch et al.	709/206
6,795,830	B1	9/2004	Banerjee et al.	707/200
6,799,221	B1	9/2004	Kenner et al.	709/245
6,804,704	B1	10/2004	Bates et al.	709/217
6,845,370	B2	1/2005	Burkey et al.	707/3
6,868,395	B1	3/2005	Szlam et al.	705/27
6,920,502	B2	7/2005	Araujo et al.	709/229
2001/0020242	A1	9/2001	Gupta et al.	707/501.1
2001/0032092	A1	10/2001	Calver	705/1
2001/0034663	A1	10/2001	Teveler et al.	705/26
2001/0052003	A1	12/2001	Seki et al.	709/219
2001/0054004	A1	12/2001	Powers	705/14
2001/0056366	A1	12/2001	Naismith	705/10
2002/0016735	A1	2/2002	Runge et al.	705/14
2002/0032603	A1	3/2002	Yeiser	705/14
2002/0035501	A1	3/2002	Handel et al.	705/10
2002/0035568	A1	3/2002	Benthin et al.	707/102
2002/0038299	A1	3/2002	Zemik et al.	707/3
2002/0046138	A1	4/2002	Fitzpatrick et al.	705/27
2002/0049622	A1	4/2002	Lettich et al.	705/7
2002/0072982	A1	6/2002	Barton et al.	705/26
2002/0073058	A1	6/2002	Kremer et al.	707/1
2002/0087387	A1	7/2002	Calver et al.	705/10
2002/0107701	A1	8/2002	Batty et al.	705/1
2002/0116484	A1	8/2002	Podracky	709/223
2002/0143870	A1	10/2002	Rau	709/204
2002/0161685	A1	10/2002	Dwinnell	705/36
2002/0178166	A1	11/2002	Hsia	707/100
2003/0028896	A1	2/2003	Swart et al.	725/127
2003/0040845	A1	2/2003	Spool et al.	700/291
2003/0065805	A1	4/2003	Bames, Jr.	709/231
2003/0083922	A1	5/2003	Reed et al.	705/9
2003/0120502	A1	6/2003	Robb et al.	705/1
2003/0139975	A1	7/2003	Perkowski	705/26
2004/0002887	A1	1/2004	Fliess et al.	705/9
2005/0021611	A1	1/2005	Knapp et al.	709/203

2005/0044280 A1 2/2005 Reisman 710/1

FOREIGN PATENT DOCUMENTS

EP	1 162 558	A1	12/2001
EP	1555626	A2	7/2005
WO	95/30201		11/1995
WO	98/21679		5/1998
WO	98/49641		11/1998
WO	9901826		1/1999
WO	99/66446		12/1999
WO	00/23929		4/2000
WO	01/22692		3/2001
WO	0177935		10/2001

OTHER PUBLICATIONS

Journyx and IBM team to deliver enterprise project and time tracking software, article, 3 pages, Apr. 5, 1999.

Optio Software, Inc. NEWS: Optio Software and Syntax.net Reseller Partnership Offers a Robust Solution to Provider and Deliver Customized Documents to Support E-Business and Extend the Reach of the Global Enterprise, 2 pages, Dec. 20, 1999.

Andreas Geyer-Schultz et al., "A customer purchase incidence model applied to recommender services" WEBKDD 2001 Mining Log data across all customer touch points, third international workshop, p. 1-11, Aug. 26, 2001.

Elprin, Nick et al., An Analysis of Database-Driven Mail Servers, LISA XVII, pp. 15-22, 2003.

Warlick, David, "Searching the Internet: Part III", Raw Materials for the Mind: Teaching & Learning in Information & Technology Rich Schools, ISBN 0-9667432-0-2, Mar. 18, 1999.

Pervasive Solution Sheet "Harvesting Unstructured Data", 5 pages, 2003.

Key Building Blocks for Knowledge Management Solutions, "IBM Intelligent Miner for Text" 2 pages, 1999.

Wood, David, "Metadata Searches of Unstructured Textural Content," Tucana Plugged in Software white Paper, 4 pages, Sep. 26, 2002.

Watson, Ian, "A Case Based Reasoning Application for Engineering Sales Support Using Introspective Reasoning," 2000 American Association for Artificial Intelligence, 6 pages, 2000.

Beantree, "Enterprise Business Application Architecture" Enterprise Business Components Whitepaper, 5 pages, Sep. 1999.

Lee et al., "An enterprise intelligence system integrating WWW intranet resource" IEEE Xplore Release 1.8, pp. 28-35 with abstract, 1999.

Mathur, Srita, "Creating Unique Customer Experiences: The New Business Model of Cross-Enterprise Integration" IEEE Xplore Release 1.8, pp. 76-81 with abstract, 2000.

80-20 Software, "End Email and File Chaos," 80-20 Retriever Enterprise Edition, 4 pages, 2003.

An InsumaGmbH White Paper, "OASIS Distributed Search Engine," pp. 1-11, no date.

Adomavicius et al., "*Using Data Mining Methods to Build Customer Profiles*", IEEE 2001 computer, pp. 74-82, 2001.

Griffin et al., "*Enterprise Customer Relationship Management*", DM review, 15 pages, Dec. 1999.

Parkhomenko et al., "*Personalization Using Hybrid Data Mining Approaches in E-Business Applications*", Amer. assoc. for Artificial Intellignece, 7 pages, 2002.

Weiss, Gary M., "*Data Mining in Telecommunications*", 13 pages, no date.

Delen et al., "*An Integrated Toolkit for Enterprise Modeling and Analysis*", Proceedings of the 1999 winter Simulation Conference, pp. 289-297, 1999.

Murtagh, Fionn, "Distributed Information Search and Retrieval for Astronomical Resource Discovery and Data Mining", Library and Information Services in Astronomy III, ASP Conference Series, vol. 153, 1998, pp. 51-60, 1998.

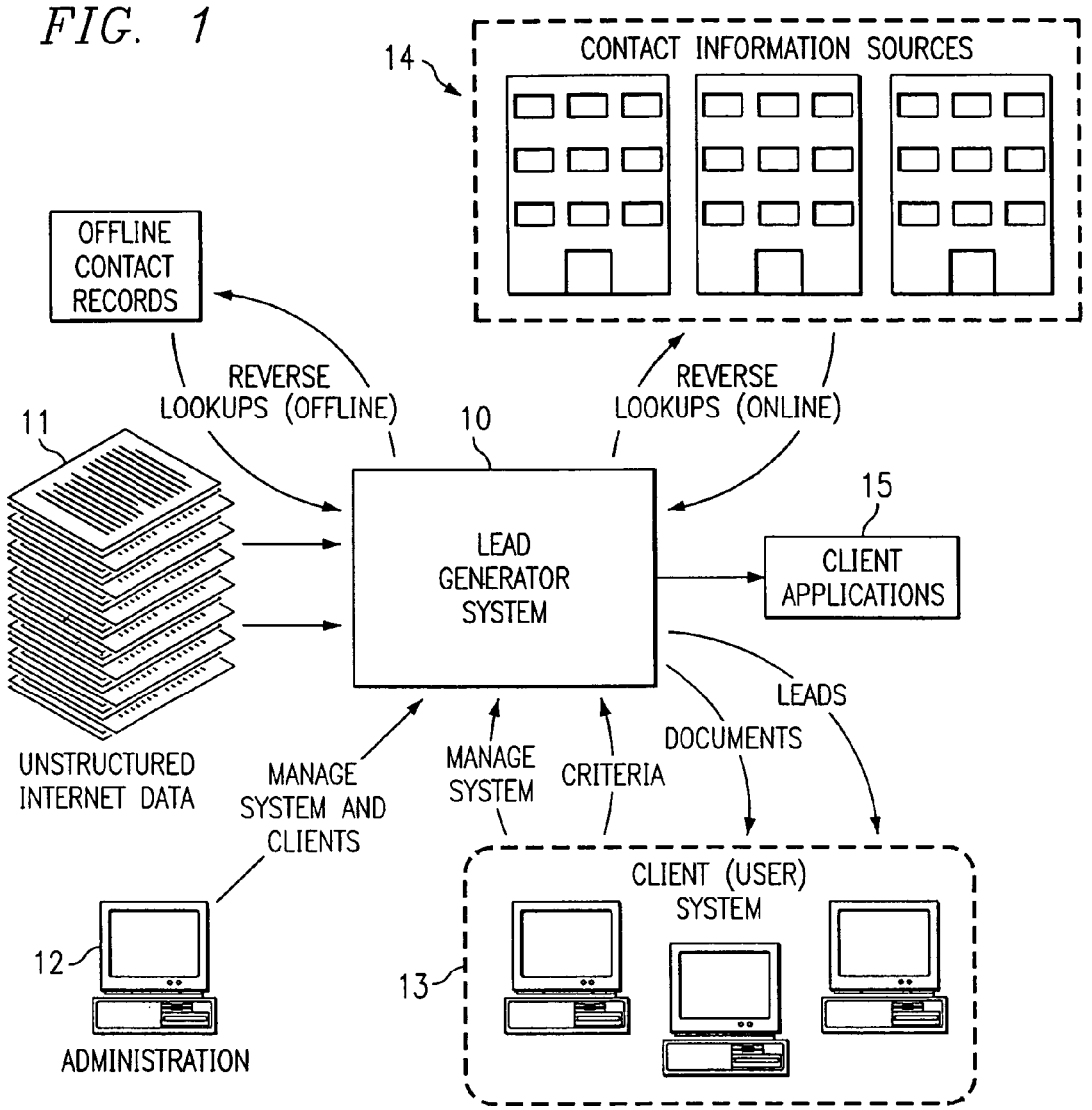
Schwartz, Michael F. et al., "Applying an information gathering architecture to Netfind: a white pages tool for a changing and growing internet", IEEE/A CM Transactions on Networking (TON), vol. 2, Issue 5, Oct. 1994, pp. 426-439.

Gravano, Luis et al., "GIOSS: text-source discovery over the Internet", ACM Transactions on Database Systems (TODS), vol. 24, Issue 2, Jun. 1999, pp. 229-264.

Mouri, T. et al., "Extracting new topic contents from hidden web sites", International Conference on Information Technology: Coding and Computing 2004, pp. 314-319, 2004.

IPR2025-00623
UIPATH EXHIBIT 1013

FIG. 1


14
CONTACT INFORMATION SOURCES
OFFLINE CONTACT RECORDS
REVERSE LOOKUPS (OFFLINE)
REVERSE LOOKUPS (ONLINE)
11
10
15
LEAD GENERATOR SYSTEM
CLIENT APPLICATIONS
UNSTRUCTURED INTERNET DATA
MANAGE SYSTEM AND CLIENTS
MANAGE SYSTEM
CRITERIA
DOCUMENTS
LEADS
CLIENT (USER) SYSTEM
12
13
ADMINISTRATION

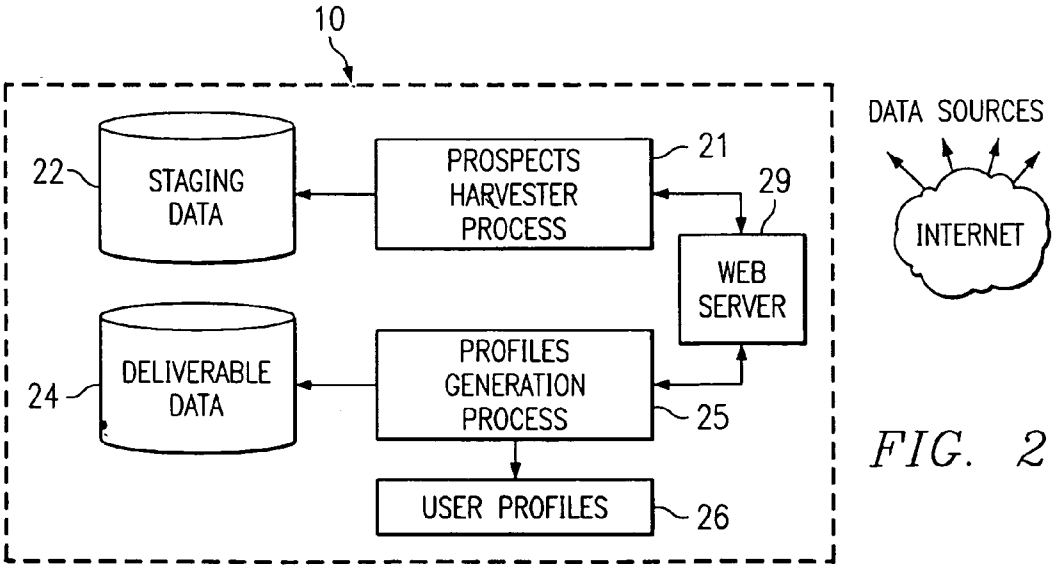

10
22
STAGING DATA
PROSPECTS HARVESTER PROCESS
21
29
WEB SERVER
DATA SOURCES
INTERNET
24
DELIVERABLE DATA
PROFILES GENERATION PROCESS
25
USER PROFILES
26

FIG. 2

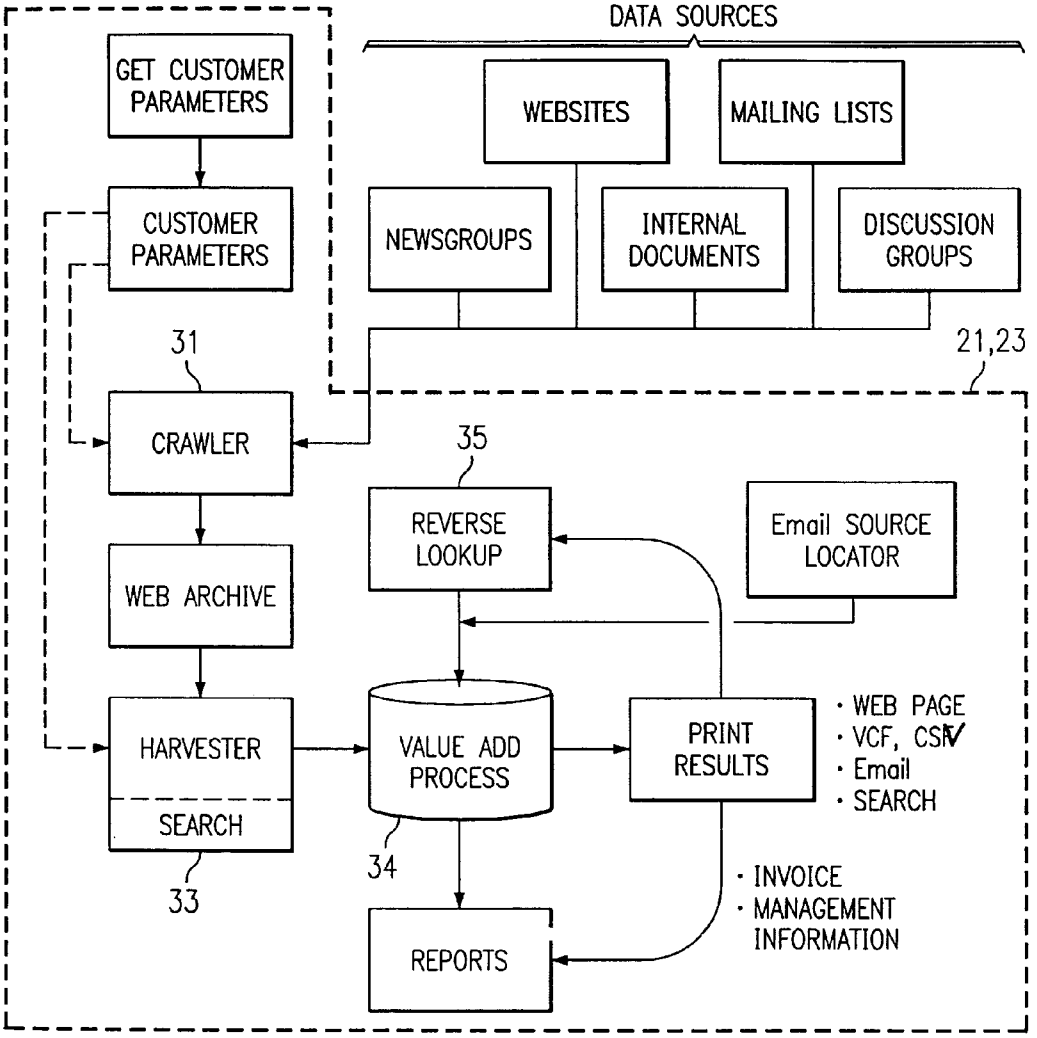

DATA SOURCES
GET CUSTOMER PARAMETERS
WEBSITES
MAILING LISTS
CUSTOMER PARAMETERS
NEWSGROUPS
INTERNAL DOCUMENTS
DISCUSSION GROUPS
31
21,23
CRAWLER
35
REVERSE LOOKUP
Email SOURCE LOCATOR
WEB ARCHIVE
HARVESTER
SEARCH
VALUE ADD PROCESS
PRINT RESULTS
· WEB PAGE
· VCF, CSV
· Email
· SEARCH
33
34
· INVOICE
· MANAGEMENT INFORMATION
REPORTS

FIG. 3

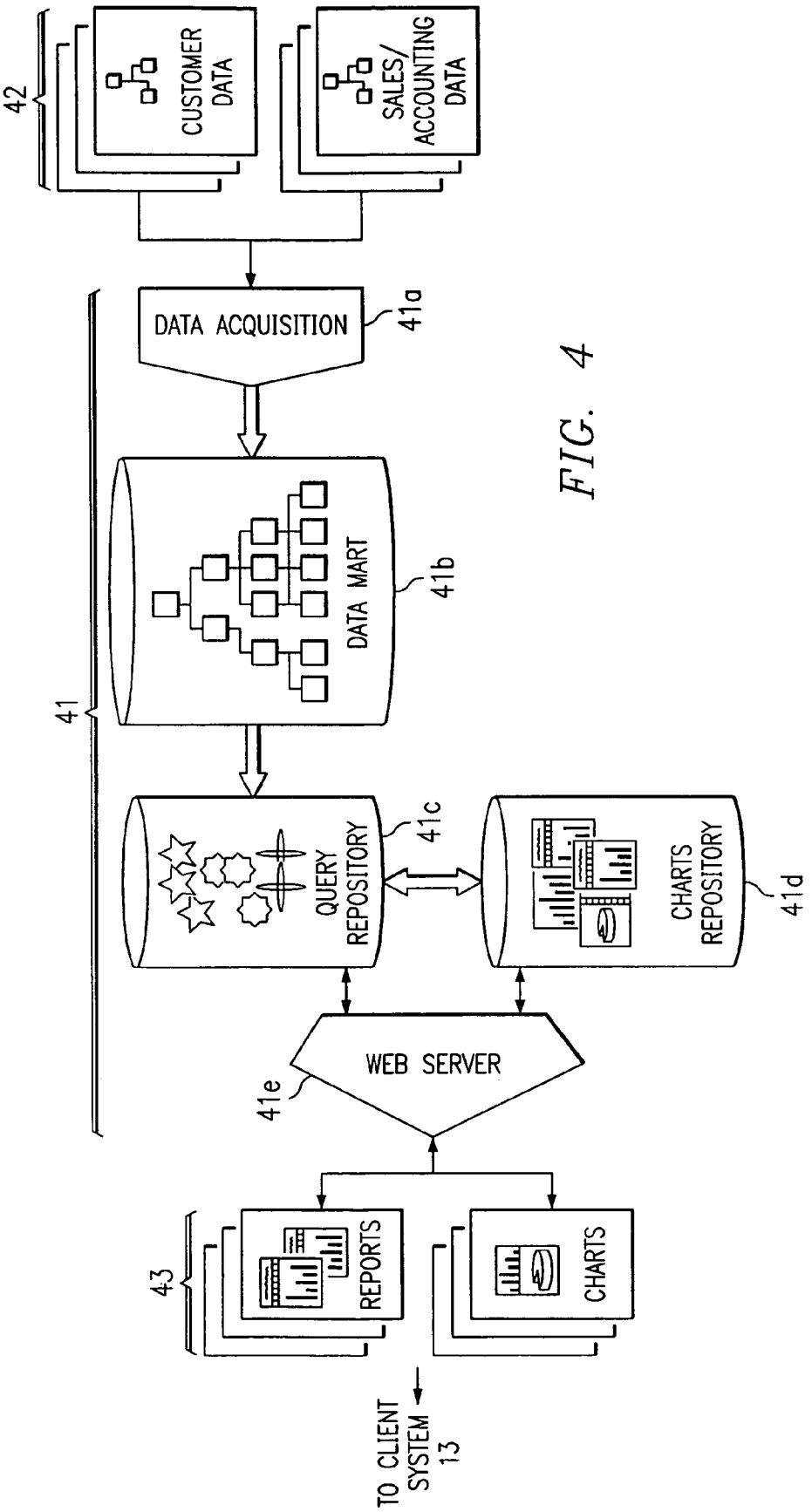
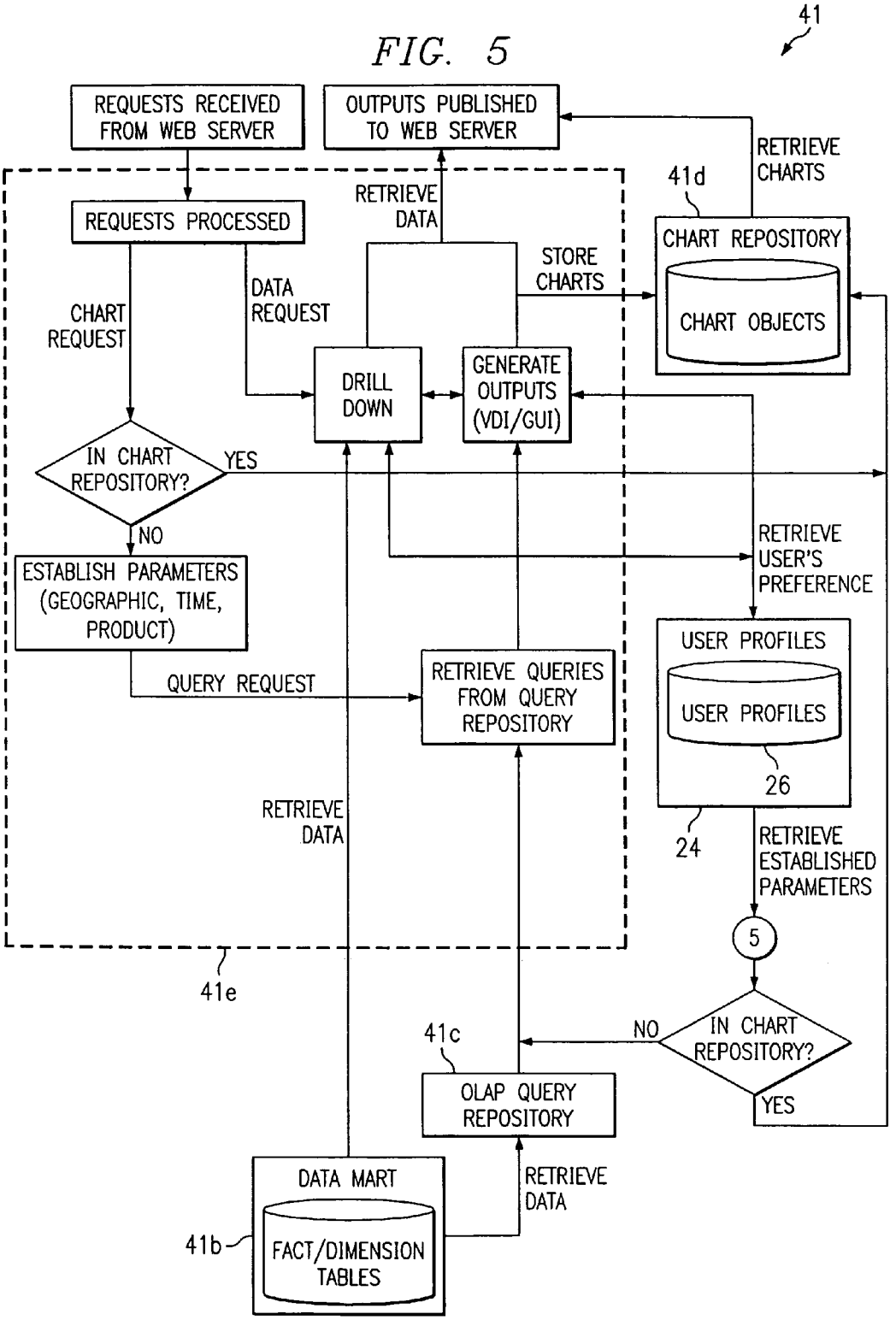

42
CUSTOMER DATA
SALES/ ACCOUNTING DATA
DATA ACQUISITION
41a
41
DATA MART
41b
QUERY REPOSITORY
41c
CHARTS REPOSITORY
41d
WEB SERVER
41e
43
REPORTS
CHARTS
TO CLIENT SYSTEM 13

FIG. 4

FIG. 5


41
REQUESTS RECEIVED FROM WEB SERVER
OUTPUTS PUBLISHED TO WEB SERVER
RETRIEVE CHARTS
41d
REQUESTS PROCESSED
RETRIEVE DATA
CHART REPOSITORY
STORE CHARTS
CHART OBJECTS
CHART REQUEST
DATA REQUEST
DRILL DOWN
GENERATE OUTPUTS (VDI/GUI)
IN CHART REPOSITORY?
YES
NO
RETRIEVE USER'S PREFERENCE
ESTABLISH PARAMETERS (GEOGRAPHIC, TIME, PRODUCT)
USER PROFILES
USER PROFILES
QUERY REQUEST
RETRIEVE QUERIES FROM QUERY REPOSITORY
26
RETRIEVE DATA
24
RETRIEVE ESTABLISHED PARAMETERS
5
41e
41c
NO
IN CHART REPOSITORY?
OLAP QUERY REPOSITORY
YES
DATA MART
RETRIEVE DATA
41b
FACT/DIMENSION TABLES

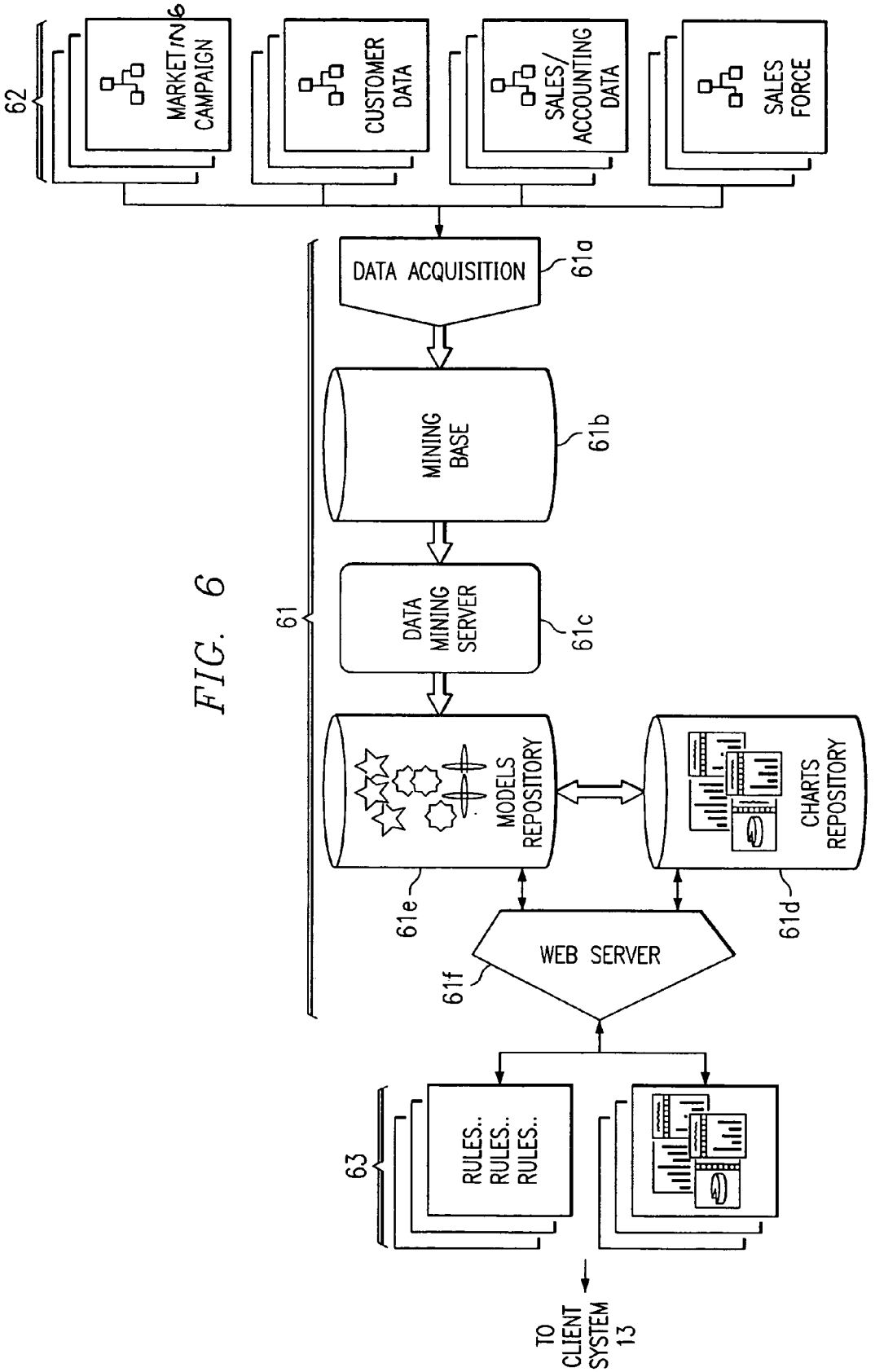
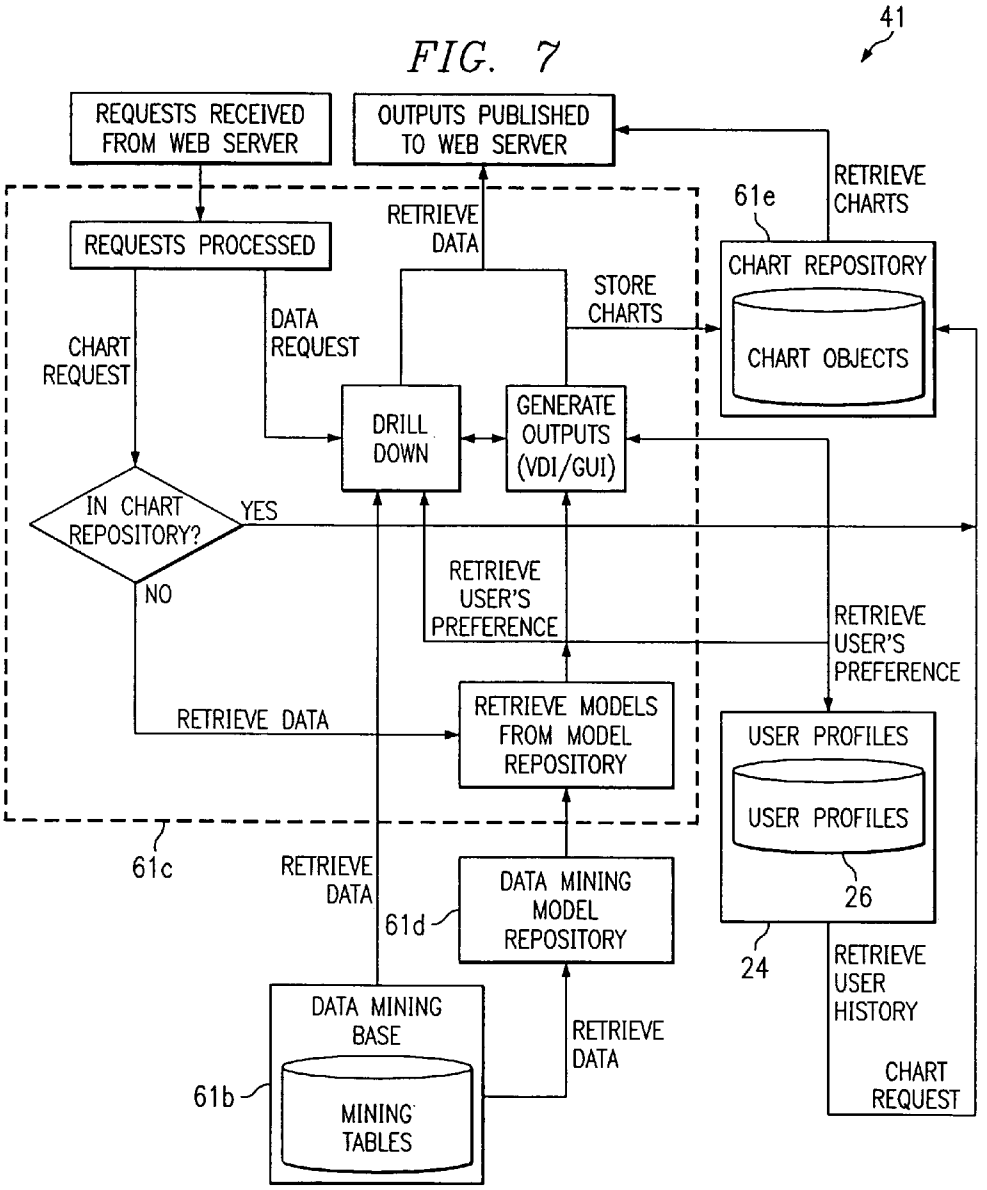


FIG. 6

IPR2025-00623
UIPATH EXHIBIT 1013

FIG. 7


41
REQUESTS RECEIVED FROM WEB SERVER
OUTPUTS PUBLISHED TO WEB SERVER
RETRIEVE CHARTS
61e
REQUESTS PROCESSED
RETRIEVE DATA
CHART REPOSITORY
STORE CHARTS
CHART OBJECTS
CHART REQUEST
DATA REQUEST
DRILL DOWN
GENERATE OUTPUTS (VDI/GUI)
IN CHART REPOSITORY?
YES
NO
RETRIEVE USER'S PREFERENCE
RETRIEVE USER'S PREFERENCE
RETRIEVE DATA
RETRIEVE MODELS FROM MODEL REPOSITORY
USER PROFILES
USER PROFILES
26
61c
RETRIEVE DATA
DATA MINING MODEL REPOSITORY
61d
24
RETRIEVE USER HISTORY
DATA MINING BASE
RETRIEVE DATA
61b
MINING TABLES
CHART REQUEST

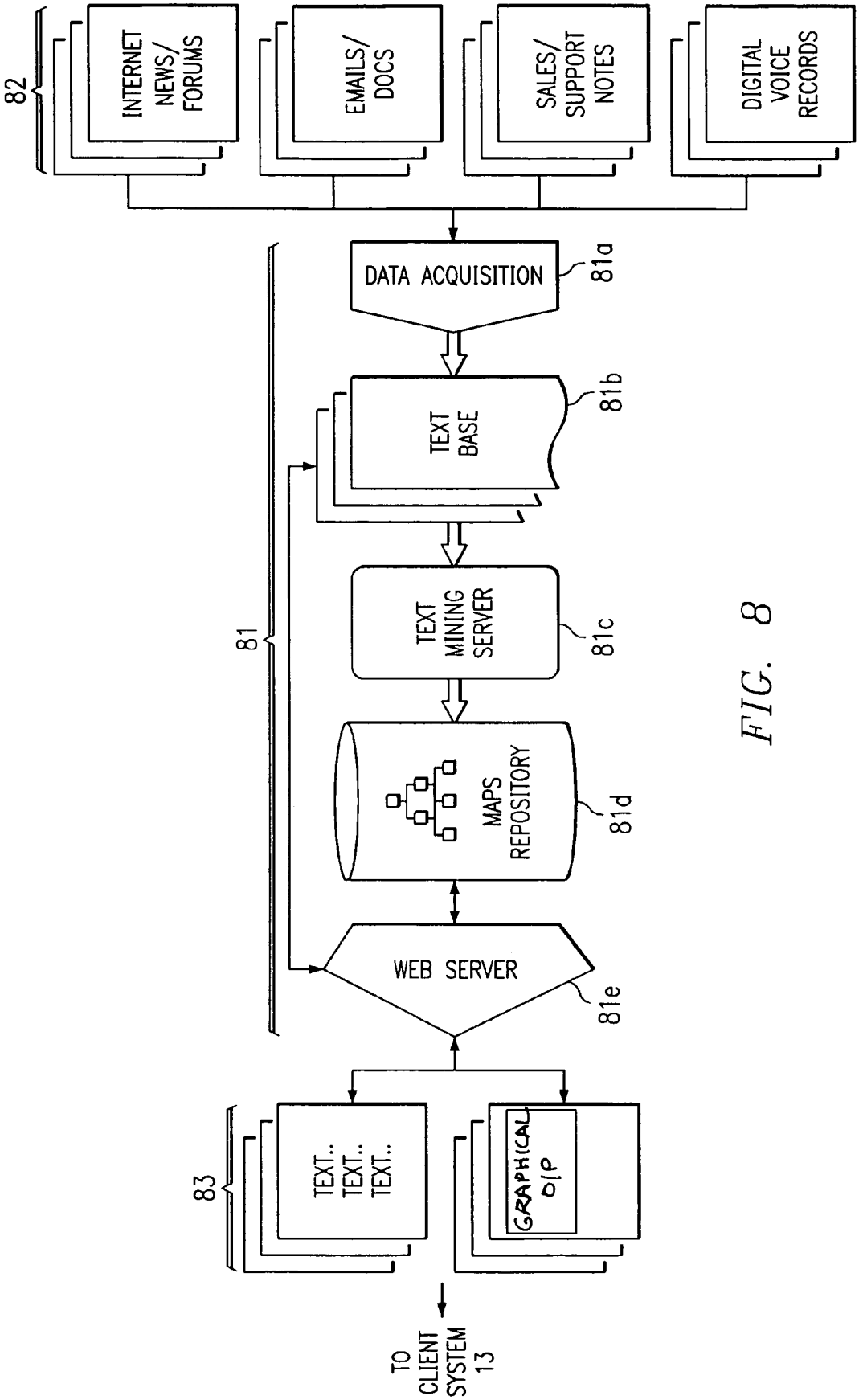


FIG. 8

IPR2025-00623
UIPATH EXHIBIT 1013

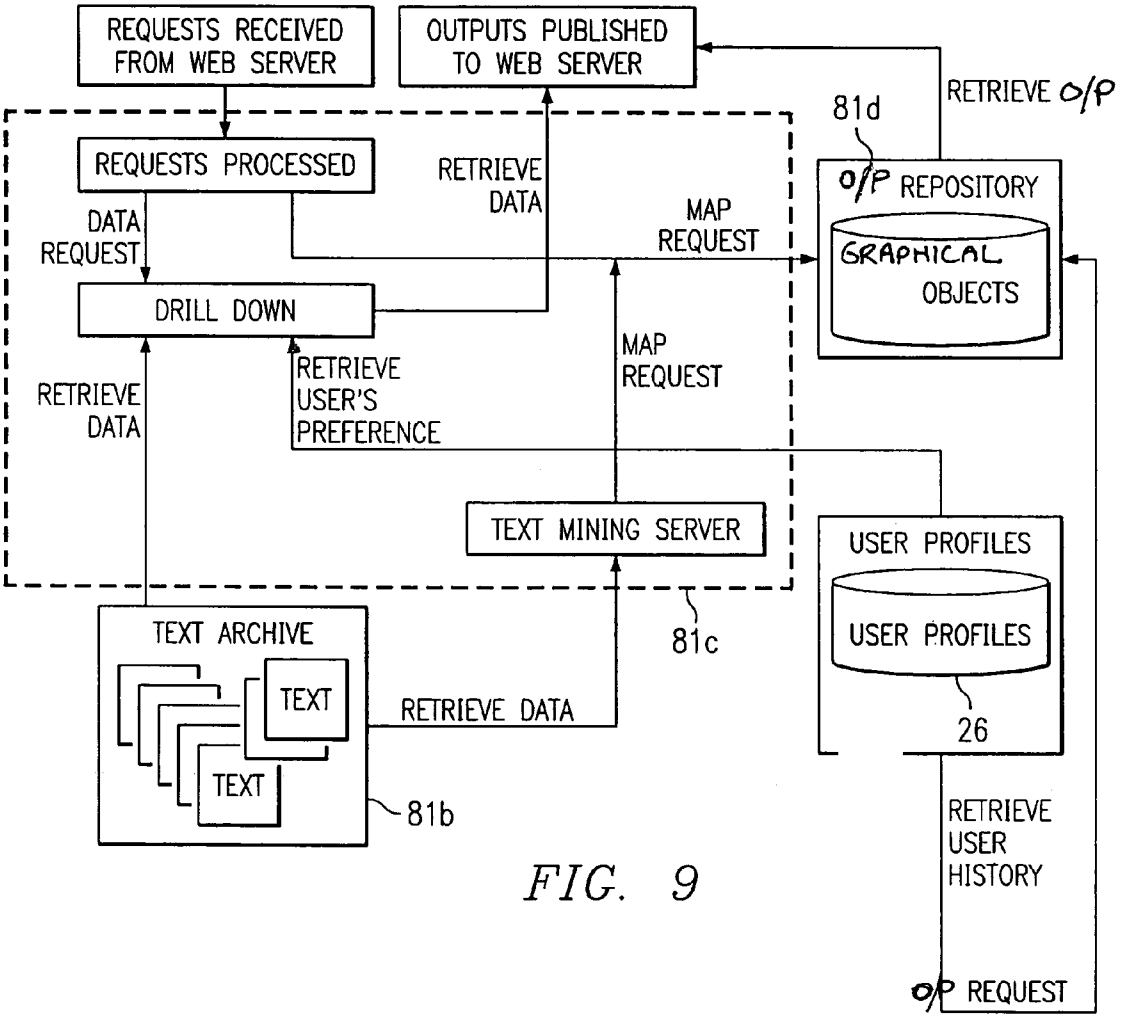


FIG. 9

IPR2025-00623
UIPATH EXHIBIT 1013

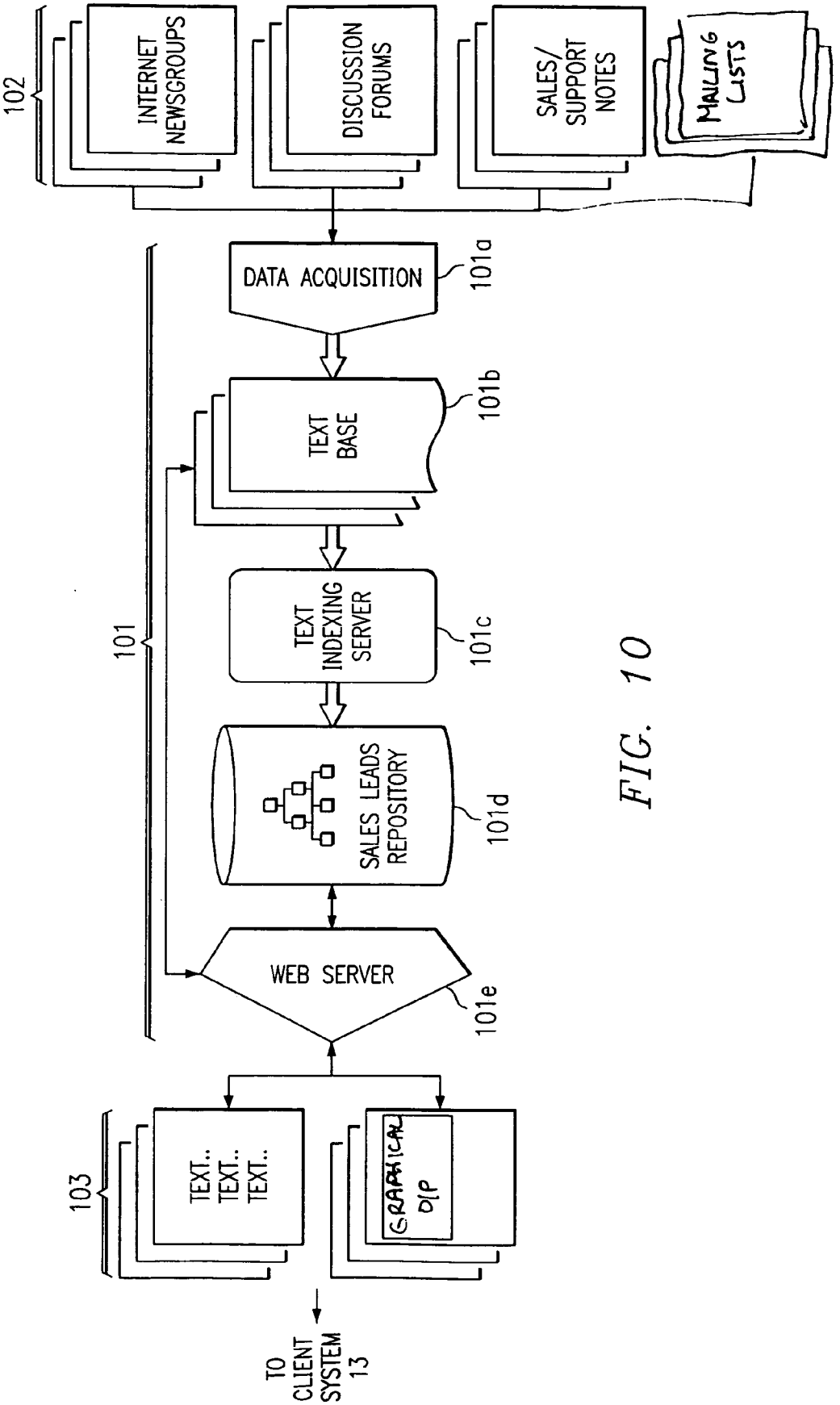
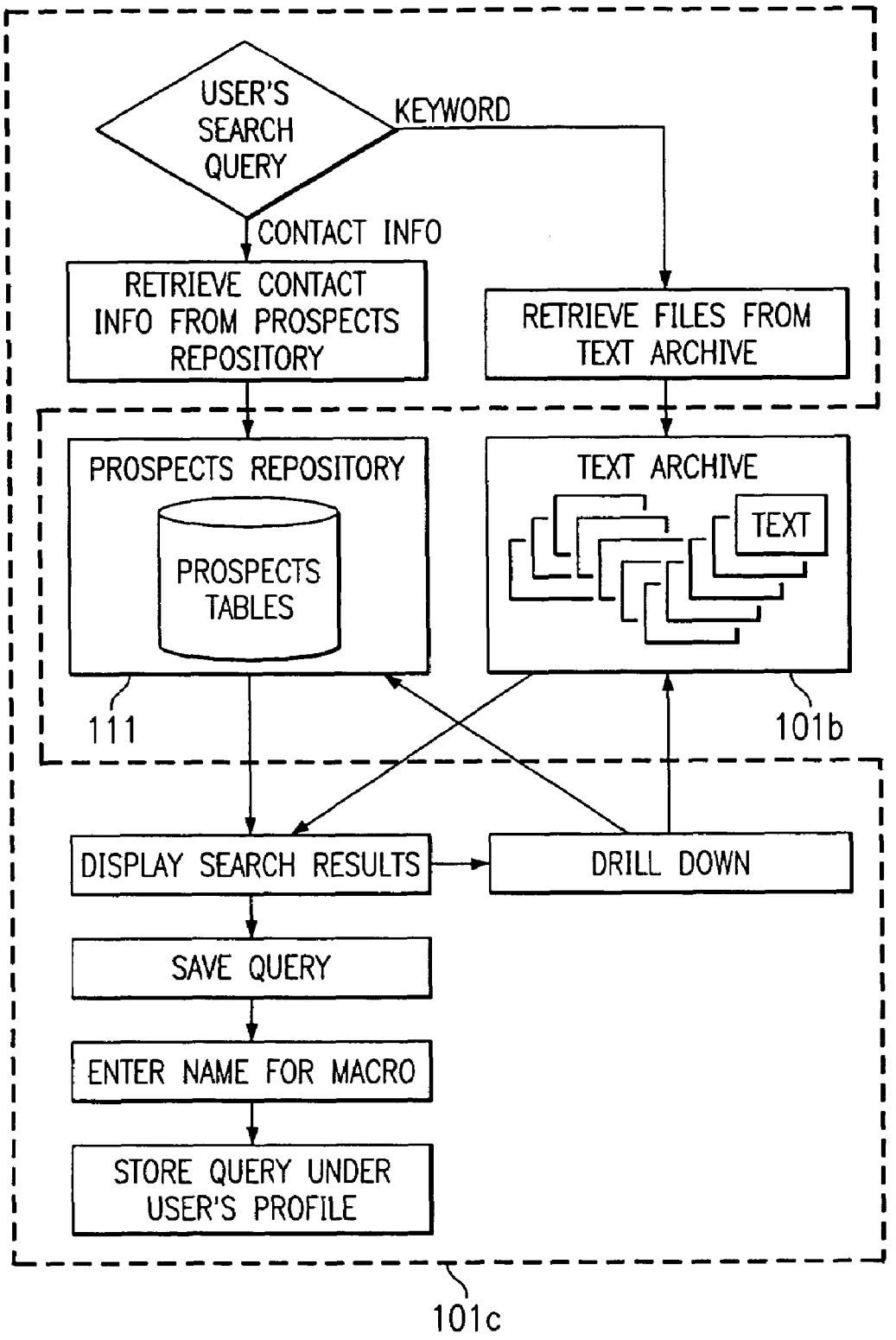


FIG. 10

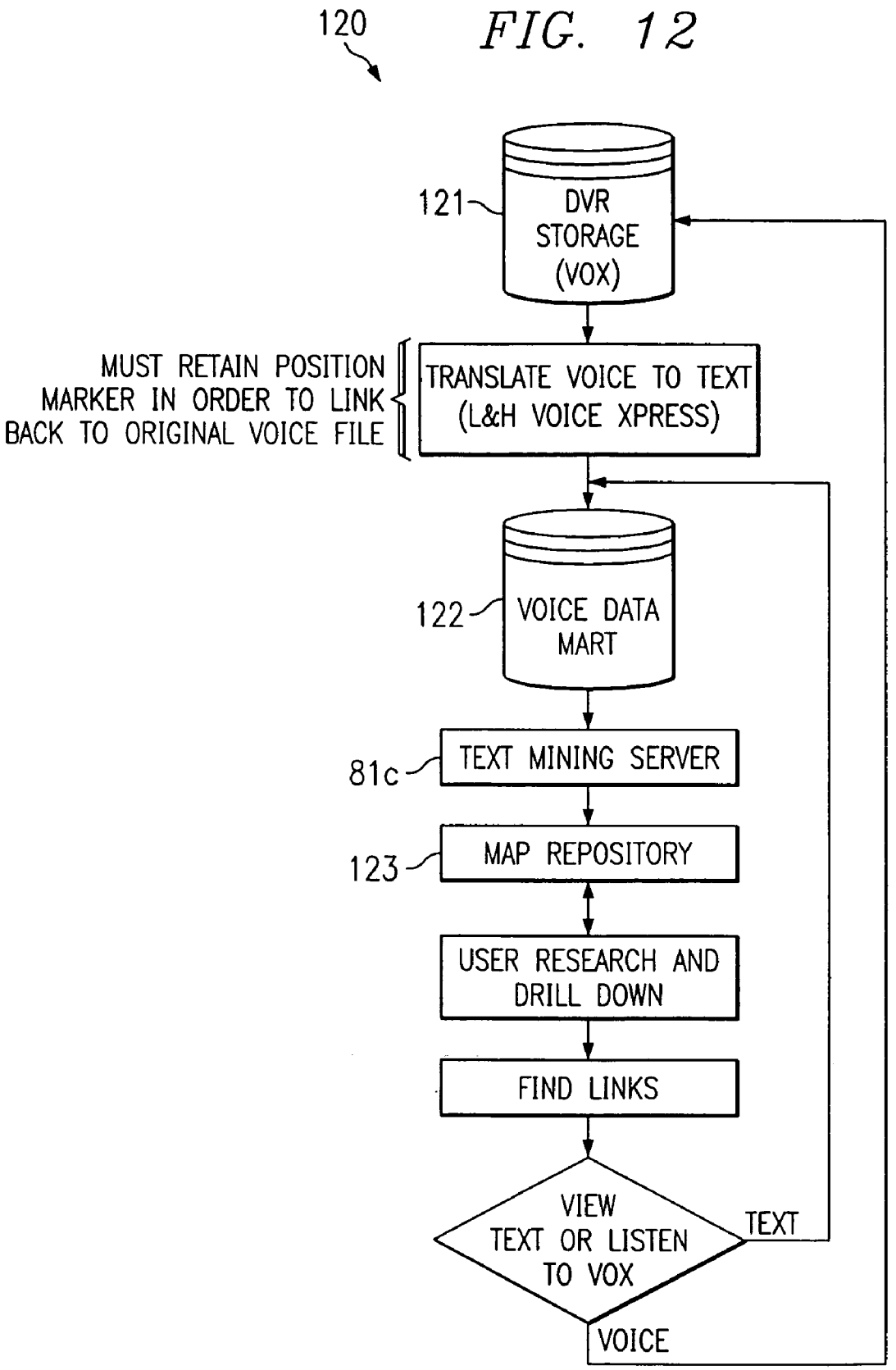
IPR2025-00623
UIPATH EXHIBIT 1013

FIG. 11



IPR2025-00623
UIPATH EXHIBIT 1013

FIG. 12



IPR2025-00623
UIPATH EXHIBIT 1013

TEXT MINING SYSTEM FOR WEB-BASED BUSINESS INTELLIGENCE

RELATED PATENT APPLICATIONS

This application is a continuation of U.S. patent application Ser. No. 09/865,735, filed May 24, 2001, now U.S. Pat. No. 7,003,517 and entitled "Text Mining System for Web-Based Business Intelligence," which claims the benefit of U.S. Provisional Application No. 60/206,772, filed May 24, 2000 and entitled "Web-Based Customer Lead Generator". Additionally, the following patent applications are each conversions from the foregoing provisional filing: patent application Ser. No. 09/862,832 entitled "Web-Based Customer Lead Generator System" and filed May 21, 2001; patent application Ser. No. 09/862,814 entitled "Web-Based Customer Prospects Harvester System" and filed May 21, 2001; patent application Ser. No. 09/865,802 entitled "Database Server System for Web-Based Business Intelligence" and filed May 24, 2001; patent application Ser. No. 09/865,804 entitled "Data Mining System for Web-Based Business Intelligence" and filed May 24, 2001; patent application Ser. No. 09/865,805 entitled "Text Indexing System for Web-Based Business Intelligence" and filed May 24, 2001.

TECHNICAL FIELD OF THE INVENTION

This invention relates to electronic commerce, and more particularly to business intelligence software tools for acquiring leads for prospective customers, using Internet data sources.

BACKGROUND OF THE INVENTION

Most small and medium sized companies face similar challenges in developing successful marketing and sales campaigns. These challenges include locating qualified prospects who are making immediate buying decisions. It is desirable to personalize marketing and sales information to match those prospects, and to deliver the marketing and sales information in a timely and compelling manner. Other challenges are to assess current customers to determine which customer profile produces the highest net revenue, then to use those profiles to maximize prospecting results. Further challenges are to monitor the sales cycle for opportunities and inefficiencies, and to relate those findings to net revenue numbers.

Today's corporations are experiencing exponential growth to the extent that the volume and variety of business information collected and accumulated is overwhelming. Further, this information is found in disparate locations and formats. Finally, even if the individual data bases and information sources are successfully tapped, the output and reports may be little more than spreadsheets, pie charts and bar charts that do not directly relate the exposed business intelligence to the companies' processes, expenses, and to its net revenues.

With the growth of the Internet, one trend in developing marketing and sales campaigns is to gather customer information by accessing Internet data sources. Internet data intelligence and data mining products face specific challenges. First, they tend to be designed for use by technicians, and are not flexible or intuitive in their operation; secondly, the technologies behind the various engines are changing rapidly to take advantage of advances in hardware and software, and finally, the results of their harvesting and mining are not typically related to a specific department goals and objectives.

SUMMARY OF THE INVENTION

One aspect of the invention is a text mining system for collecting business intelligence about a client, as well as for identifying prospective customers of the client. The text mining system is used in a lead generation system accessible by the client via the Internet.

The mining system has various components, including a data acquisition process that extracts textual data from various Internet sources, a database for storing the extracted data, a text mining server that executes query-based searches of the database, and an output repository. A web server provides client access to the repository, and to the mining server.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** illustrates the operating environment for a web based lead generator system in accordance with the invention.

FIG. **2** illustrates the various functional elements of the lead generator system.

FIG. **3** illustrates the various data sources and a first embodiment of the prospects harvester.

FIGS. **4** and **5** illustrate a database server system, which may be used within the lead generation system of FIGS. **1** and **2**.

FIGS. **6** and **7** illustrate a data mining system, which may be used within the lead generation system of FIGS. **1** and **2**.

FIGS. **8** and **9** illustrate a text mining system, which may be used within the lead generation system of FIGS. **1** and **2**.

FIGS. **10** and **11** illustrate a text indexing system, which may be used within the lead generation system of FIGS. **1** and **2**.

FIG. **12** illustrates a digital voice recording mining system, which may be used within the lead generation system of FIGS. **1** and **2**.

DETAILED DESCRIPTION OF THE INVENTION

Lead Generator System Overview

FIG. **1** illustrates the operating environment for a web-based customer lead generation system **10** in accordance with the invention. System **10** is in communication, via the Internet, with unstructured data sources **11**, an administrator **12**, client systems **13**, reverse look-up sources **14**, and client applications **15**.

The users of system **10** may be any business entity that desires to conduct more effective marketing campaigns. These users may be direct marketers who wish to maximizing the effectiveness of direct sales calls, or e-commerce web site who wish to build audiences.

In general, system **10** may be described as a web-based Application Service Provider (ASP) data collection tool. The general purpose of system **10** is to analyze a client's marketing and sales cycle in order to reveal inefficiencies and opportunities, then to relate those discoveries to net revenue estimates. Part of the latter process is proactively harvesting prequalified leads from external and internal data sources. As explained below, system **10** implements an automated process of vertical industry intelligence building that involves automated reverse lookup of contact information using an email address and key phrase highlighting based on business rules and search criteria.

More specifically, system **10** performs the following tasks:

IPR2025-00623
UIPATH EXHIBIT 1013

Uses client-provided criteria to search Internet postings for prospects who are discussing products or services that are related to the client's business offerings

Selects those prospects matching the client's criteria

Pushes the harvested prospect contact information to the client, with a link to the original document that verifies the prospects interest

Automatically opens or generates personalized sales scripts and direct marketing materials that appeal to the prospects' stated or implied interests

Examines internal sales and marketing materials, and by applying data and text mining analytical tools, generates profiles of the client's most profitable customers

Cross-references and matches the customer profiles with harvested leads to facilitate more efficient harvesting and sales presentations

In the audience building environment, requests permission to contact the prospect to offer discounts on services or products that are directly or indirectly related to the conversation topic, or to direct the prospect to a commerce source.

System **10** provides open access to its web site. A firewall (not shown) is used to prevent access to client records and the entire database server. Further details of system security are discussed below in connection with FIG. **5**.

Consistent with the ASP architecture of system **10**, interactions between client system **13** and system **10** will typically be by means of Internet access, such as by a web portal. Authorized client personnel will be able to create and modify profiles that will be used to search designated web sites and other selected sources for relevant prospects.

Client system **11** may be any computer station or network of computers having data communication to lead generator system **10**. Each client system **11** is programmed such that each client has the following capabilities: a master user account and multiple sub user accounts, a user activity log in the system database, the ability to customize and personalize the workspace; configurable, tiered user access; online signup, configuration and modification, sales territory configuration and representation, goals and target establishment, and online reporting comparing goals to target (e.g., expense/revenue; budget/actual).

Administration system **14** performs such tasks as account activation, security administration, performance monitoring and reporting, assignment of master userid and licensing limits (user seats, access, etc.), billing limits and profile, account termination and lockout, and a help system and client communication.

System **10** interfaces with various client applications **15**. For example, system **10** may interface with commercially available enterprise resource planning (ERP), sales force automation (SFA), call center, e-commerce, data warehousing, and custom and legacy applications.

Lead Generator System Architecture

FIG. **2** illustrates the various functional elements of lead generator system **10**. In the embodiment of FIG. **2**, the above described functions of system **10** are partitioned between two distinct processes.

A prospects harvester process **21** uses a combination of external data sources, client internal data sources and user-parameter extraction interfaces, in conjunction with a search, recognition and retrieval system, to harvest contact information from the web and return it to a staging data base **22**. In general, process **21** collects business intelligence data from both inside the client's organization and outside the organization. The information collected can be either structured data as in corporate databases/spreadsheet files or unstructured data as in textual files.

Process **21** may be further programmed to validate and enhance the data, utilizing a system of lookup, reverse lookup and comparative methodologies that maximize the value of the contact information. Process **21** may be used to elicit the prospect's permission to be contacted. The prospect's name and email address are linked to and delivered with ancillary information to facilitate both a more efficient sales call and a tailored e-commerce sales process. The related information may include the prospect's email address, Web site address and other contact information. In addition, prospects are linked to timely documents on the Internet that verify and highlight the reason(s) that they are in fact a viable prospect. For example, process **21** may link the contact data, via the Internet, to a related document wherein the contact's comments and questions verify the high level value of the contact to the user of this system (the client).

A profiles generation process **25** analyzes the user's in-house files and records related to the user's existing customers to identify and group those customers into profile categories based on the customer's buying patterns and purchasing volumes. The patterns and purchasing volumes of the existing customers are overlaid on the salient contact information previously harvested to allow the aggregation of the revenue-based leads into prioritized demand generation sets. Process **25** uses an analysis engine and both data and text mining engines to mine a company's internal client records, digital voice records, accounting records, contact management information and other internal files. It creates a profile of the most profitable customers, reveals additional prospecting opportunities, and enables sales cycle improvements. Profiles include items such as purchasing criteria, buying cycles and trends, cross-selling and up-selling opportunities, and effort to expense/revenue correlations. The resulting profiles are then overlaid on the data obtained by process **21** to facilitate more accurate revenue projections and to enhance the sales and marketing process. The client may add certain value judgments (rankings) in a table that is linked to a unique lead id that can subsequently be analyzed by data mining or OLAP analytical tools. The results are stored in the deliverable database **24**.

Profiles generation process **25** can be used to create a user (client) profiles database **26**, which stores profiles of the client and its customers. As explained below, this database **26** may be accessed during various data and text mining processes to better identify prospective customers of the client.

Web server **29** provides the interface between the client systems **13** and the lead generation system **10**. As explained below, it may route different types of requests to different sub processes within system **10**. The various web servers described below in connection with FIGS. **4-11** may be implemented as separate servers in communication with a front end server **29**. Alternatively, the server functions could be integrated or partitioned in other ways.

Data Sources

FIG. **3** provides additional detail of the data sources of FIGS. **1** and **2**. Access to data sources may be provided by various text mining tools, such as by the crawler process **31** or **41** of FIGS. **3** and **4**.

One data source is newsgroups, such as USENET. To access discussion documents from USENET newsgroups such as "news.giganews.com", NNTP protocol is used by the crawler process to talk to USENET news server such as "news.giganews.com." Most of the news servers only

IPR2025-00623
UIPATH EXHIBIT 1013

archive news articles for a limited period (giganews.com archives news articles for two weeks), it is necessary for the iNet Crawler to incrementally download and archive these newsgroups periodically in a scheduled sequence. This aspect of crawler process **31** is controlled by user-specified parameters such as news server name, IP address, newsgroup name and download frequency, etc.

Another data source is web-Based discussion forums. The crawler process follows the hyper links on a web-based discussion forum, traverse these links to user or design specified depths and subsequently access and retrieve discussion documents. Unless the discussion documents are archived historically on the web site, the crawler process will download and archive a copy for each of the individual documents in a file repository. If the discussion forum is membership-based, the crawler process will act on behalf of the authorized user to logon to the site automatically in order to retrieve documents. This function of the crawler process is controlled by user specified parameters such as a discussion forum's URL, starting page, the number of traversal levels and crawling frequency.

A third data source is Internet-based or facilitated mailing lists wherein individuals send to a centralized location emails that are then viewed and/or responded to by members of a particular group. Once a suitable list has been identified a subscription request is initiated. Once approved, these emails are sent to a mail server where they are downloaded, stored in system **10** and then processed in a fashion similar to documents harvested from other sources. The system stores in a database the filters, original URL and approval information to ensure only authorized messages are actually processed by system **10**.

A fourth data source is corporations' internal documents. These internal documents may include sales notes, customer support notes and knowledge base. The crawler process accesses corporations' internal documents from their Intranet through Unix/Windows file system or alternately be able to access their internal documents by riding in the databases through an ODBC connection. If internal documents are password-protected, crawler process **31** acts on behalf of the authorized user to logon to the file systems or databases and be able to subsequently retrieve documents. This function of the crawler process is controlled by user-specified parameters such as directory path and database ODBC path, starting file id and ending file id, and access frequency. Other internal sources are customer information, sales records, accounting records, and digitally recorded correspondence such as e-mail files or digital voice records.

A fifth data source is web pages from Internet web sites. This function of the crawler process is similar to the functionality associated with web-discussion-forums. Searches are controlled by user-specified parameters such as web site URL, starting page, the number of traversal levels and crawling frequency.

Database Server System

FIGS. **4** and **5** illustrate a database server system **41**, which may be used within system **10** of FIGS. **1** and **2**. FIG. **4** illustrates the elements of system **41** and FIG. **5** is a data flow diagram. Specifically, system **41** could be used to implement the profiles generation process **25**, which collects profile data about the client.

The input data **42** can be the client's sales data, customer-contact data, customer purchase data and account data etc. Various data sources for customer data can be contact management software packages such as ACT, MarketForce, Goldmine, and Remedy. Various data sources for accounting data are Great Plains, Solomon and other accounting packages typically found in small and medium-sized businesses. If the client has ERP (enterprise resource planning) systems (such as JD Edwards, PeopleSoft and SAP) installed, the data sources for customer and accounting data will be extracted from ERP customer and accounting modules. This data is typically structured and stored in flat files or relational databases. System **41** is typically an OLAP (On-line analytic processing) type server-based system. It has five major components. A data acquisition component **41***a* collects and extracts data from different data sources, applying appropriate transformation, aggregation and cleansing to the data collected. This component consists of predefined data conversions to accomplish most commonly used data transformations, for as many different types of data sources as possible. For data sources not covered by these predefined conversions, custom conversions need to be developed. The tools for data acquisition may be commercially available tools, such as Data Junction, ETI*EXTRACT, or equivalents. Open standards and APIs will permit employing the tool that affords the most efficient data acquisition and migration based on the organizational architecture.

Data mart **41***b* captures and stores an enterprise's sales information. The sales data collected from data acquisition component **41***a* are "sliced and diced" into multidimensional tables by time dimension, region dimension, product dimension and customer dimension, etc. The general design of the data mart follows data warehouse/data mart Star-Schema methodology. The total number of dimension tables and fact tables will vary from customer to customer, but data mart **41***b* is designed to accommodate the data collected from the majority of commonly used software packages such as PeopleSoft or Great Plains.

Various commercially available software packages, such as Cognos, Brio, Informatica, may be used to design and deploy data mart **41***b*. The Data Mart can reside in DB2, Oracle, Sybase, MS SQL server, P.SQL or similar database application. Data mart **41***b* stores sales and accounting fact and dimension tables that will accommodate the data extracted from the majority of industry accounting and customer contact software packages.

A Predefined Query Repository Component **41***c* is the central storage for predefined queries. These predefined queries are parameterized macros/business rules that extract information from fact tables or dimension tables in the data mart **41***b*. The results of these queries are delivered as business charts (such as bar charts or pie charts) in a web browser environment to the end users. Charts in the same category are bounded with the same predefined query using different parameters. (i.e. quarterly revenue charts are all associated with the same predefined quarterly revenue query, the parameters passed are the specific region, the specific year and the specific quarter). These queries are stored in either flat file format or as a text field in a relational database.

A Business Intelligence Charts Repository Component **41***d* serves two purposes in the database server system **41**. A first purpose is to improve the performance of chart retrieval process. The chart repository **41***d* captures and stores the most frequently visited charts in a central location. When an end user requests a chart, system **41** first queries the chart repository **41***d* to see if there is an existing chart. If there is a preexisting chart, server **41***e* pulls that chart directly from the repository. If there is no preexisting chart, server **41***e* runs the corresponding predefined query from the query repository **41***c* in order to extract data from data mart **41***b* and subsequently feed the data to the requested chart. A second purpose is to allow chart sharing, collaboration and

IPR2025-00623
UIPATH EXHIBIT 1013

distribution among the end users. Because charts are treated as objects in the chart repository, users can bookmark a chart just like bookmarking a regular URL in a web browser. They can also send and receive charts as an email attachment. In addition, users may logon to system **41** to collaboratively make decisions from different physical locations. These users can also place the comments on an existing chart for collaboration.

Another component of system **41** is the Web Server component **41***e*, which has a number of subcomponents. A web server subcomponent (such as Microsoft IIS or Apache server or any other commercially available web servers) serves HTTP requests. A database server subcomponent (such as Tango, Cold Fusion or PHP) provides database drill-down functionality. An application server subcomponent routes different information requests to different other servers. For example, sales revenue chart requests will be routed to the database system **41**; customer profile requests will be routed to a Data Mining server, and competition information requests will be routed to a Text Mining server. The latter two systems are discussed below. Another subcomponent of server **41***e* is the chart server, which receives requests from the application server. It either runs queries against data mart **41***b*, using query repository **41***c*, or retrieves charts from chart repository **41***c*.

As output **43**, database server system **41** delivers business intelligence about an organization's sales performance as charts over the Internet or corporate Intranet. Users can pick and choose charts by regions, by quarters, by products, by companies and even by different chart styles. Users can drill-down on these charts to reveal the underlying data sources, get detailed information charts or detailed raw data. All charts are drill-down enabled allowing users to navigate and explore information either vertically or horizontally. Pie charts, bar charts, map views and data views are delivered via the Internet or Intranet.

As an example of operation of system **41**, gross revenue analysis of worldwide sales may be contained in predefined queries that are stored in the query repository **41***c*. Gross revenue queries accept region and/or time period as parameters and extract data from the Data Mart **41***b* and send them to the web server **41***e*. Web server **41***e* transforms the raw data into charts and publishes them on the web.

Data Mining System

FIGS. **6** and **7** illustrate a data mining system **61**, which may be used within system **10** of FIGS. **1** and **2**. FIG. **6** illustrates the elements of system **61** and FIG. **7** is a data flow diagram. Specifically, system **61** could be used to implement the profiles process **25**, which collects profile data about the client.

Data sources **62** for system **61** are the Data Mart **41***b*, e.g., data from the tables that reside in Data Mart **41***b*, as well as data collected from marketing campaigns or sales promotions.

For data coming from the Data Mart **41***b*, data acquisition process **61***a* between Mining Base **61***b* and Data Mart **41***b* extract/transfer and format/transform data from tables in the Data Mart **41***b* into Data Mining base **61***b*. For data collected from sales and marketing events, data acquisition process **61***a* may be used to extract and transform this kind of data and store it in the Data Mining base **61***b*.

Data Mining base **61***b* is the central data store for the data for data mining system **61**. The data it stores is specifically prepared and formatted for data mining purposes. The Data Mining base **61***b* is a separate data repository from the Data Mart **41***b*, even though some of the data it stores is extracted from Data Mart's tables. The Data Mining base **61***b* can reside in DB2, Oracle, Sybase, MS SQL server, P.SQL or similar database application.

Chart repository **61***d* contains data mining outputs. The most frequently used decision tree charts are stored in the chart repository **61***d* for rapid retrieval.

Customer purchasing behavior analysis is accomplished by using predefined Data Mining models that are stored in a model repository **61***e*. Unlike the predefined queries of system **41**, these predefined models are industry-specific and business-specific models that address a particular business problem. Third party data mining tools such as IBM Intelligent Miner and Clementine, and various integrated development environments (IDEs) may be used to explore and develop these data mining models until the results are satisfactory. Then the models are exported from the IDE into standalone modules (in C or C++) and integrated into model repository **61***e* by using data mining APIs.

Data mining server **61***c* supplies data for the models, using data from database **61***c*. FIG. **7** illustrates the data paths and functions associated with server **61***c*. Various tools and applications that may be used to implement server **61***c* include VDI, EspressChart, and a data mining GUI.

The outputs of server **61***e* may include various options, such as decision trees, Rule Sets, and charts. By default, all the outputs have drill-down capability to allow users to interactively navigate and explore information in either a vertical or horizontal direction. Views may also be varied, such as by influencing factor. For example, in bar charts, bars may represent factors that influence customer purchasing (decision-making) or purchasing behavior. The height of the bars may represent the impact on the actual customer purchase amount, so that the higher the bar is the more important the influencing factor is on customers' purchasing behavior. Decision trees offer a unique way to deliver business intelligence on customers' purchasing behavior. A decision tree consists of tree nodes, paths and node notations. Each individual node in a decision tree represents an influencing. A path is the route from root node (upper most level) to any other node in the tree. Each path represents a unique purchasing behavior that leads to a particular group of customers with an average purchase amount. This provides a quick and easy way for on-line users to identify where the valued customers are and what the most important factors are when customer are making purchase decisions. This also facilitates tailored marketing campaigns and delivery of sales presentations that focus on the product features or functions that matter most to a particular customer group. Rules Sets are plain-English descriptions of the decision tree. A single rule in the RuleSet is associated with a particular path in the decision tree. Rules that lead to the same destination node are grouped into a RuleSet. RuleSet views allow users to look at the same information presented in a decision tree from a different angle. When users drill down deep enough on any chart, they will reach the last drill-down level that is data view. A data view is a table view of the underlying data that supports the data mining results. Data Views are dynamically linked with Data Mining base **61***b* and Data Mart **41***b* through web server **61***f*.

Web server **61***f*, which may be the same as database server **41***e*, provides Internet access to the output of mining server **61***c*. Existing outputs may be directly accessed from storage in charts repository **61***d*. Or requests may be directed to models repository **61***e*. Consistent with the application service architecture of lead generation system **10**, access by the client to web server **61***f* is via the Internet and the client's web browser.

IPR2025-00623
UIPATH EXHIBIT 1013

Text Mining System

FIGS. **8** and **9** illustrate a text mining system **81**, which may be used within system **10** of FIGS. **1** and **2**. FIG. **8** illustrates the elements of system **81** and FIG. **9** is a data flow diagram. As indicated in FIG. **8**, the source data **82** for system **81** may be either external and internal data sources. Thus, system **81** may be used to implement both the prospects system and profiles system of FIG. **2**.

The source data **82** for text mining system **81** falls into two main categories, which can be mined to provide business intelligence. Internal documents contain business information about sales, marketing, and human resources. External sources consist primarily of the public domain in the Internet. Newsgroups, discussion forums, mailing lists and general web sites provide information on technology trends, competitive information, and customer concerns.

More specifically, the source data **82** for text mining system **81** is from five major sources. Web Sites: on-line discussion groups, forums and general web sites. Internet News Group: Internet newsgroups for special interests such as alt.ecommerce and microsoft.software.interdev. For some of the active newsgroups, hundreds of news articles may be harvested on a weekly basis. Internet Mailing Lists: mailing lists for special interests, such as e-commerce mailing list, company product support mailing list or Internet marketing mailing list. For some of the active mailing lists, hundreds of news articles will be harvested on a weekly basis. Corporate textual files: internal documents such as emails, customer support notes sales notes, and digital voice records.

For data acquisition **81***a* from web sites, user-interactive web crawlers are used to collect textual information. Users can specify the URLs, the depth and the frequency of web crawling. The information gathered by the web crawlers is stored in a central repository, the text archive **81***b*. For data acquisition from newsgroups, a news collector contacts the news server to download and transform news articles in an html format and deposit them in text archive **81***b*. Users can specify the newsgroups names, the frequency of downloads and the display format of the news articles to news collector. For data acquisition from Internet mailing lists, a mailing list collector automatically receives, sorts and formats email messages from the subscribed mailing lists and deposit them into text archive **81***b*. Users can specify the mailing list names and address and the display format of the mail messages. For data acquisition from client text files, internal documents are sorted, collected and stored in the Text Archive **81***b*. The files stored in Text Archive **81***b* can be either physical copies or dynamic pointers to the original files.

The Text Archive **81***b* is the central data store for all the textual information for mining. The textual information it stores is specially formatted and indexed for text mining purpose. The Text Archive **81***b* supports a wide variety of file formats, such plain text, html, MS Word and Acrobat.

Text Mining Server **81***c* operates on the Text Archive **81***b*. Tools and applications used by server **81***c* may include ThemeScape and a Text Mining GUI **81***c*. A repository **81***d* stores text mining outputs. Web server **81***e* is the front end interface to the client system **13**, permitting the client to access database **81***b*, using an on-line search executed by server **81***c* or server **81***e*.

The outputs of system **81** may include various options. Map views and simple query views may be delivered over the Internet or Intranet. By default, all the outputs have drill-down capability to allow users to reach the original documents. HTML links will be retained to permit further lateral or horizontal navigation. Keywords will be highlighted or otherwise pointed to in order to facilitate rapid location of the relevant areas of text when a document is located through a keyword search. For example, Map Views are the outputs produced by ThemeScape. Textual information is presented on a topological map on which similar "themes" are grouped together to form "mountains." On-line users can search or drill down on the map to get the original files. Simple query views are similar to the interfaces of most of the Internet search engines offered (such as Yahoo, Excite and HotBot). It allows on-line users to query the Text Archive **81***b* for keywords or key phrases or search on different groups of textual information collected over time.

A typical user session using text-mining system **81** might follow the following steps. It is assumed that the user is connected to server **81***e* via the Internet and a web browser, as illustrated in FIG. **1**. In the example of this description, server **81***e* is in communication with server **81***c*, which is implemented using ThemeScape software.

1. Compile list of data sources (Newsgroups, Discussion Groups, etc)
2. Start ThemeScape Publisher or comparable application
3. Select "File"
4. Select "Map Manager" or comparable function
5. Verify that server and email blocks are correctly set. If not, insert proper information.
6. Enter password.
7. Press "Connect" button
8. Select "New"
9. Enter a name for the new map
10. If duplicating another maps settings, use drop down box to select the map name.
11. Select "Next"
12. Select "Add Source"
13. Enter a Source Description
14. Source Type remains "World Wide Web (WWW)"
15. Enter the URL to the site to be mined.
16. Add additional URLs, if desired.
17. Set "Harvest Depth." Parameters range from 1 level to 20 levels.
18. Set "Filters" if appropriate. These include Extensions, Inclusions, Exclusions, Document Length and Rations.
19. Set Advanced Settings, if appropriate. These include Parsing Settings, Harvest Paths, Domains, and Security and their sub-settings.
20. Repeat steps 14 through 20 for each additional URL to be mined.
21. Select "Advanced Settings" if desired. These include Summarization Settings, Stopwords, and Punctuation.
22. Select "Finish" once ready to harvest the sites.
23. The software downloads and mines (collectively known as harvesting) the documents and creates a topographical map.
24. Once the map has been created, it can be opened and searched.

Access to User Profiles Database

As explained above in connection with FIG. **2**, the profiles generation process **25** may be used to generate a profiles database **26**. This database **26** stores information about the client and its customers that may be used to better identify prospective customers.

Referring again to FIGS. **5**, **7** and **9**, various mining processes used to implement system **10** may access and use the date stored in database **26**. For example, as illustrated in FIG. **5**, the database server **41***e* of database server system **41** may access database **24** to determine user preferences in formulating queries and presenting outputs. As illustrated in

IPR2025-00623
UIPATH EXHIBIT 1013

FIG. **7**, the data mining server **61***c* of data mining system **61** may access database **24** for similar purposes. Likewise, as illustrated in FIG. **9**, the text mining server **81***c* of system **81** may access database **24** to determine preferences in formulating queries, especially during query drill downs.

Text Indexing System

FIGS. **10** and **11** illustrate a text indexing system **101**, which may be used within system **10** of FIGS. **1** and **2**. FIG. **10** illustrates the elements of system **101** and FIG. **11** is a data flow diagram. Like system **81**, system **101** may be used to implement either the prospects process **21** or profiles process **25** of FIG. **2**.

Text mining system **81** and text indexing system **101** are two different systems for organizing mass textual information. Text mining system **81** identifies and extracts key phrases, major topics, and major themes from a mass amount of documents. The text mining system **81** is suitable for those on-line users who want to perform thorough research on the document collection. Text indexing system **101** is similar to text mining system **81** but is simpler and faster. It only identifies and extracts syntax information such as key words/key phrases. It provides a simple and fast alternative to users who just want to perform keyword searches.

The data sources **102** for Text Indexing system **101** are similar to those described above for Text Mining system **81**. For data acquisition **101***a*, various software may be used. These include web crawlers and mailing list collecting agents. These are similar to those described above in connection with Text Mining system **81**.

The text archive **101***b* is the central data store for all the textual information for indexing. The textual information it stores is specially formatted and indexed for text mining or indexing purpose. The Text archive **101***b* supports a wide variety of file formats, such plain text, html, MS Word and Acrobat. Text archive **101***b* may be the same text archive as used in system **81**.

Server **101***c* indexes the document collection in a multi-dimensional fashion. It indexes documents not only on keywords/key phases but also on contact information associated within the documents. In other words, the server **101***c* allows on-line users to perform cross-reference search on both keywords and contact information. As an example, when users perform a keyword search on a collection of documents, the text indexing server returns a list of hits that consist of relevance (who-when-what), hyperlink, summary, timestamp, and contact information. Alternately, when users perform contact information search on a collection of documents, the text indexing server **101***c* yields a list of documents associated with that individual.

Using Text Indexing Server **101***c*, users may navigate documents easily and quickly and find information such as "who is interested in what and when."

Contact information and links to the associated documents are migrated into a Sales Prospects repository **101***d* (a relational database). This contact information can be exported into normal contact management software from the repository **101***d*.

The outputs **103** of system **101** are varied. Simple Query Views may be delivered to the client over the Internet or Intranet. By default, all the outputs have drill-down capability to allow users to reach the original documents. The Query Views may be similar to the interfaces of commonly used Internet search engines offered, such as Yahoo, Excite and HotBot. It allows on-line users to query the Text Archive **101***b* for keywords/key phrases and contact information search on different groups of textual information collected over time.

FIG. **11** illustrates the operation of text indexing server **115**, which may be used to integrate queries from both text database **101***b* and another database **111** that stores information about prospective customers. For example, database **111** might be any one of the databases **26**, **41***b*, **61***b*, or **81***b* of FIG. **2**, **4**, **6**, or **8**. Server **115** accepts query parameters from the client, which may specify both contact parameters and keywords for searching database **111** and database **101***b*, respectively. The search results are then targeted toward a particular category of prospects. FIG. **11** also illustrates how server **115** may be used to store, identify, and reuse queries. The queries for a particular client may be stored in user profiles database **26**.

Digital Voice Recording Mining System

FIG. **12** illustrates a digital voice recording mining system **120**. System **12** may be used to implement the prospects process **21** of FIG. **2**, or it may be integrated into the text mining system of FIGS. **8** and **9**.

Digital Voice Records (DVR) are increasing in use as companies move to sell and market over increasing boundaries, improve customer relations and provide a variety of support functions through call centers and third-party vendors. Present technology allows calls to be recalled through date-time stamps and a variety of other positional indicators but there are no means to analyze the content and context of the massive amount of this audio media.

System **120** uses speech-to-text translation capability to convert the digitally recorded voices, most often Vox or Wave (wav) format, into machine-readable text. A positional locator is created in the header file to facilitate direct linking back to the voice record, if needed. Accuracy of the recording on the receiving end is enhanced through training of the voice engine; an acceptable margin of error is expected on the incoming voice. The text files are stored in a Data Mart **122** where they may be mined using a search engine. Search engines such as ThemeScape are especially suitable in that they do more than simply count words and index frequently occurring phrases; they find "themes" by examining where words appear in the subject, text and individual sentence structure.

A typical user session of system **120** might follow the following steps: Call is either received or initiated. Depending on state law, the parties are advised that the call may be recorded for quality control purposes. Call is digitally recorded using existing technology from providers such as IDigiVoice. Vox or Wave (voice) files **121** are translated using speech-to-text conversion programs. Text files are stored in logical areas in Data Mart **122**, for mining with a search engine. Maps or similar visual/graphical representations are placed in a Map or Image Repository **123**. Users search maps using the search engines browser plug-in. When the user finds documents to review, the user is prompted to select "voice" or "text." If text, the original document/file in the Data Mart is displayed in the browser window. If voice, the positional indicator is pumped to the Digital Voice Record application that locates, calls and then plays to voice file segment.

Referring again to FIG. **8**, the voice data mart **122** may be one of the data sources for text mining system **81**. Text mining server **81***c* is programmed to execute the functions of FIG. **12** as well as the other functions described above in connection with FIGS. **8** and **9**. Similarly, the text in Data Mart **120** could be indexed using server **101***c* of FIGS. **10** and **11**. In today's technological environment, the DVR

IPR2025-00623
UIPATH EXHIBIT 1013

storage **121** would originate from internal storage of the client, but Internet retrieval is also a possibility.

OTHER EMBODIMENTS

Although the present invention has been described in detail, it should be understood that various changes, substitutions, and alterations can be made hereto without departing from the spirit and scope of the invention as defined by the appended claims.

What is claimed is:

1. A text mining system that permits an application service provider to provide data identifying prospective customers of a product or service provided by a product/service provider, comprising:

a data acquisition process for extracting text documents from unstructured Internet discussion sites selected from the group of: newsgroups, discussion forums, and mailing lists;

wherein the data acquisition process operates by intruding into the Internet discussion site independently of, and without activation by the participants of the discussion site;

wherein each discussion site comprises a collection of text documents generated by participants of the discussion site and simultaneously received by multiple participants of the discussion site;

an archive database for storing the extracted text documents;

a text mining server for receiving queries from the product/service provider, each query representing a request for information about participants who are discussing a specified product or service;

wherein the text mining server responds to each query by accessing the archive database;

wherein the text mining server is further operable to analyze data in the archive database, to provide the product/service provider with text mining outputs representing themes extracted from the data, and at least one link to at least one document stored in the archive database that contains information identifying a prospective customer who is discussing the product or service; and

a web server for providing access to the text mining server via a web browser and the Internet, such that the product/service provider may interact with the text mining server to make the requests on line via the Internet and to receive the outputs generated by the text mining server;

wherein the at least one link enables the product/service provider to access the at least one text document from the archive database via the web browser.

2. The text mining system of claim **1**, wherein the text documents are items of digitally recorded correspondence.

3. The text mining system of claim **1**, wherein the text documents are digital voice records.

4. The text mining system of claim **1**, wherein the data acquisition process is implemented with a web crawler.

5. The text mining system of claim **1**, wherein text sources are further selected from Internet voice-to-text files.

6. The text-mining system of claim **1**, wherein the data acquisition process is further operable to perform an on-line search for contact information about the prospective customer.

7. The text-mining system of claim **1**, wherein the data acquisition process is further operable to perform an off-line search for contact information about the prospective customer.

8. The text-mining system of claim **1**, further comprising a mailer process for contacting the prospective customer.

9. The text-mining system of claim **1**, wherein the mailer process requests permission for further contact to the prospective customer.

10. The text-mining system of claim **1**, wherein the mailer process provides a link to sales materials.

11. The text-mining system of claim **1**, further comprising a database for storing query criteria of the product/service provider.

* * * * *

IPR2025-00623
UIPATH EXHIBIT 1013